



HAL
open science

Découverte automatique de motifs chimiques associés à un profil d'activité: application pour la recherche de structures d'alertes en mutagénicité

Jean-Philippe Metivier, Alban Lepailleur, Aleksey Buzmakov, Guillaume Poezevara, Bruno Crémilleux, Sergei O. Kuznetsov, Jérémie Le Goff, Valentin Lemière, Amedeo Napoli, Ronan Bureau, et al.

► To cite this version:

Jean-Philippe Metivier, Alban Lepailleur, Aleksey Buzmakov, Guillaume Poezevara, Bruno Crémilleux, et al.. Découverte automatique de motifs chimiques associés à un profil d'activité: application pour la recherche de structures d'alertes en mutagénicité. 7èmes Journées de la Société Française de Chémoinformatique, Oct 2015, Nice, France. 2015. hal-01627308

HAL Id: hal-01627308

<https://hal.science/hal-01627308>

Submitted on 6 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Découverte automatique de motifs chimiques associés à un profil d'activité :
application pour la recherche de structures d'alertes en mutagénicité**

Jean-Philippe Métivier [1,2], Alban Lepailleur [2], Aleksey Buzmakov [3,4], Guillaume Poezevara [1,2], Bruno Crémilleux [1], Sergei O. Kuznetsov [4], Jérémie Le Goff [5], Valentin Lemière [1], Amedeo Napoli [3], Ronan Bureau [2], Bertrand Cuissart [1]

[1] UNICAEN, GREYC, UMR CNRS 6072, F-14032 Caen, France

[2] UNICAEN, CERMN, UPRES EA 4258, FR CNRS 3038, F-14032 Caen, France

[3] Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), University of Lorraine, Nancy, France

[4] National Research University Higher School of Economics (HSE), Moscow, Russia

[5] ADn'tox, Caen, France

bertrand.cuissart@unicaen.fr

Cette étude est dédiée à l'introduction d'une nouvelle méthode extrayant automatiquement de potentielles structures d'alerte à partir d'un ensemble de molécules. Ces structures à surveiller peuvent être utilisées pour extraire des connaissances chimiques ou pour anticiper l'activité des nouvelles molécules. Le calcul de ces structures d'alerte résulte d'un workflow [1] intégrant des outils de fouille de graphes, de mesure de contraste (ici le taux de croissance) et d'analyse de concepts formels.

Les outils de fouille de graphes permettent d'extraire des fragments moléculaires qui sont combinés pour obtenir des motifs. Grâce au taux de croissance, qui est une mesure de contraste bien connue en fouille de données, il est possible de déterminer la corrélation entre les occurrences d'un motif et une activité donnée. Un motif ayant un fort taux de croissance est qualifié d'émergent et caractérise fortement une classe. Afin de filtrer les motifs émergents, on utilise la notion de stabilité, introduite dans le domaine de l'analyse de concepts formels. En effet, la stabilité permet de conserver seulement les motifs les plus robustes et ainsi de réduire drastiquement le nombre de motifs à analyser. Les motifs résultants du processus de sélection sont ainsi émergents et stables (ou SEPs pour stable emerging patterns).

Cette méthode a été utilisée sur un jeu de données public moléculaires associé à la mutagénicité [2]. Les résultats expérimentaux montrent l'efficacité de cette méthode qui permet d'extraire un ensemble de taille raisonnable de structures d'alerte fortement corrélées à la mutagénicité. Il est possible d'explorer ces structures d'alerte à l'adresse suivante : https://chemoinfo.greyc.fr/2014_Metivier/

L'analyse experte de ces structures d'alertes à montrer la pertinence de cette approche pour la découverte de connaissances chimiques. De plus, cette méthode étant générique et flexible, elle peut être exploitée pour l'étude d'autres profils d'activité tel que des affinités ligand-récepteur.

[1] Métivier, JP; Lepailleur, A; Buzmakov, A; Poezevara, G; Crémilleux, B; Kuznetsov, S; Le Goff, J; Napoli, A; Bureau, R; Cuissart, B. Discovering Structural Alerts for Mutagenicity Using Stable Emerging Molecular Patterns, *J. Chem. Inf. Model.* 2015, 55, 925–940

[2] Hansen, K; Mika, S; Schroeter, T; Sutter, A; ter Laak, A; Steger-Hartmann, T; Heinrich, N; Muller, KR. Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. *J. Chem. Inf. Model.* 2009, 49, 2077–2081