

Chloé-Agathe Azencott

Dossier de qualification aux fonctions de maître de conférence

Table des matières

CV	2
Activités d'enseignement	4
Activités de recherche	5
Publications et interventions lors de colloques	8
Liste des pièces jointes	9

Informations personnelles

État civil Chloé-Agathe Azencott
Coordonnées MPI for Intelligent Systems +49 7071 601 582
Professionnelles AG Karsten Borgwardt
Spemannstr. 38
72072 Tübingen – ALLEMAGNE
Coordonnées Königsbergerstr. 17 +49 176 3223 5372
Personnelles 72076 Tübingen – ALLEMAGNE
Email cazencott@tuebingen.mpg.de Site web <http://www.cazencott.info>
Née le 7 janvier 1985 à Digne (04) Nationalité Française

Situation actuelle

depuis 2011 **Chercheur postdoctorante**, *Machine Learning and Computational Biology (MLCB) Research Group*, Max Planck Institute for Developmental Biology & Max Planck Institute for Intelligent Systems, Tübingen (Allemagne).

Formation universitaire

2010–2005 **Doctorat en Informatique**, *University of California Irvine (UCI)*, États-Unis.
« Statistical data mining and machine learning for chemoinformatics and drug discovery »
Soutenue : 31 août 2010 à UCI
Directeur : Pierre Baldi (UCI–Informatique et sciences de l'information)
Jury de thèse : Padhraic Smyth (UCI–Informatique et sciences de l'information)
Sheryl Tsai (UCI–Biologie moléculaire et biochimie)
Jury d'avancement de thèse : Padhraic Smyth
Dennis Kibler (UCI–Informatique et sciences de l'information)
David Eppstein (UCI–Informatique et sciences de l'information)
Hartmut Lücke (UCI–Biologie moléculaire et biochimie)

2004–2005 **Master Mathématique et Informatique**, *École Nationale Supérieure des Télécommunications de Bretagne*, Brest (28), Mention Bien.
Spécialité Logiciel et méthodes formelles
Stage de recherche : « Implémentation et validation de noyaux pour séquences biologiques »
Directeur : Jean-Philippe Vert (École des Mines de Paris)

2002–2005 **Diplôme d'Ingénieur**, *École Nationale Supérieure des Télécommunications de Bretagne*, Brest (28).
Option Informatique des télécommunications

1999–2002 **Classes Préparatoires aux Grandes Écoles**, *Lycée Louis-le-Grand*, Paris (75).
Maths et Physique (MP)

1999 **Baccalauréat**, *Lycée Pierre-Gilles de Gennes*, Digne-les-Bains (04), Mention AB.
Spécialité Mathématiques

Expérience professionnelle: Enseignement

2011–2013 **Encadrement de stages de M2**, *Universität Tübingen*, Allemagne.
– 2011–2012 : Valeri Velkov, « Mining correlated loci at a genome-wide scale »
– 2012–2013 : Fabian Aicheler, « Disease status prediction based on SNP annotation »

- 2012 **Cours magistral**, *Universität Tübingen*, Allemagne.
 - Hiver 2012 : Fouille de données en bioinformatique
- 2012 **Monitorat**, *Universität Tübingen*, Allemagne.
 - Été 2012 : Séminaire de bioinformatique
- 2006-2008 **Monitorat**, *University of California Irvine*, États-Unis.
 - Été 2008 : Introduction aux probabilités et aux statistiques pour l'informatique
 - Été 2007 : Statistiques
 - Été 2006 : Introduction à l'intelligence artificielle

Expérience professionnelle: Recherche

- depuis 2011 **Chargée de recherche**, *Max Planck Institute for Developmental Biology & Max Planck Institute for Intelligent Systems*, Tübingen (Allemagne).
 - Groupe « Machine Learning and Computational Biology » dirigé par Karsten Borgwardt
 - Méthodes statistiques pour études d'associations pan-génomiques multilocus
 - Prédiction d'associations gène-maladie à l'aide de graphes
 - Intégration de données biologiques partiellement disponibles et partiellement étiquetées
- 2005–2010 **Graduate Student Researcher**, *University of California Irvine*, (États-Unis).
 - « Institute for Genomics and Bioinformatics » dirigé par Pierre Baldi
 - Prédiction de propriétés biologiques, chimiques et physiques de molécules (noyaux pour graphes et machines à vecteurs de supports)
 - Criblage haut-débit virtuel (méthodes à noyaux, réseaux de neurones)
 - Prédiction de réactions chimiques (machines à vecteurs de supports, réseaux de neurones)
 - Docking moléculaire
- Été 2009 **Stage de Recherche**, *IBM R&D Tel-Aviv*, (Israël).
 - Groupe « Machine Learning and Data Mining » dirigé par Michal Rosen-Zvi
 - Analyse statistique de données SNP pour le projet européen HyperGene
- 2005–2010 **Stage de Master Recherche**, *École des Mines de Paris*, Fontainebleau (78).
 - Centre for Computational Biology dirigé par Jean-Philippe Vert
 - Implémentation et validation de noyaux pour séquences biologiques
 - Interface web pour le test de noyaux

Prix et Récompenses

- 2011–2013 : Bourse de recherche post-doctorale Alexander von Humboldt
- 2009–2010 : Bourse de doctorat IBM
- 2009 : Bourse d'excellence scientifique CINF-Symyx
- 2007: Premier prix du défi Agnostic Learning vs. Prior Knowledge pour le jeu de données HIVA et présentation invitée au workshop correspondant à IJCNN (International Joint Conference on Neural Networks)
- 2005–2008 : Bourse de doctorat Ted & Janice Smith

Services professionnels

- Relectures pour des revues : Bioinformatics, Journal of Chemical Information and Modeling, Molecular BioSystems, Neurocomputing, Neural Networks
- Relectures pour des colloques internationaux : NETTAB2011, MLSB12, ICML13
- Participation à l'organisation de l'école d'été européenne de bioinformatique 2013

Activités d'enseignement

1. Introduction à l'intelligence artificielle (UCI, Été 2006)

Objectifs du module Ce module, adressé à des étudiants en informatique (niveau équivalent L2), était une introduction à un certain nombre de techniques et d'algorithmes développés en intelligence artificielle : heuristiques, jeux, logique des prédicats, incertitude, classification bayésienne, éléments d'apprentissage statistique. Le module s'appuyait sur le livre *Artificial Intelligence: a Modern Approach* de Russell et Norvig.

Effectifs ~ 50 étudiants.

Participation

- Une séance de travaux dirigés par semaine : récapitulation des points principaux abordés pendant le cours magistral, notamment grâce à des problèmes que j'ai conçus pour préparer à l'examen ;
- Plusieurs permanences d'une heure (« *office hours* ») par semaine pour assister les étudiants individuellement ;
- Correction des fiches d'exercices ;
- Correction (avec l'enseignante) des examens.

2. Statistiques (UCI, Été 2007)

Objectifs du module Ce module, adressé à des étudiants en informatique ou mathématiques (niveau équivalent L1 ou L2) était une introduction à l'inférence statistique : intervalles de confiance, test d'hypothèses, valeurs p, lois classiques, régression, corrélation. Le but du module était de préparer les étudiants à comprendre et analyser les études statistiques qu'ils rencontrent, que ce soit dans d'autres cours, dans des articles scientifiques, ou dans la presse, et à conduire leurs propres études statistiques.

Effectifs ~ 110 étudiants.

Participation Mon rôle de « *reader* » pour ce module a été de corriger les copies des étudiants (six fiches d'exercices et trois examens), ainsi que d'élaborer de courts rapports sur les acquis et lacunes apparaissant dans ces copies.

3. Probabilités et statistiques en informatique (UCI, Été 2008)

Objectifs du module Ce module introductif, adressé à des étudiants en informatique ou mathématiques (niveau équivalent L1 ou L2) avait pour but de présenter aux étudiants les concepts basiques d'inférence statistique et de probabilité, en particulier théorie des probabilités, probabilités conditionnelles, variables aléatoires, distributions classiques, inférence statistique (estimation, tests, et régression). Le module s'appuyait sur *Probability and Statistics for Engineering and the Sciences* de Devore.

Effectifs ~ 35 étudiants.

Participation

- Une séance de travaux dirigés par semaine : récapitulation des points principaux abordés pendant les cours magistraux à travers des problèmes pris dans le manuel ou développés par moi ;
- Au moins trois permanences d'une heure par semaine pour assister les étudiants individuellement ;
- Correction des fiches d'exercices.

4. Fouille de données en bioinformatique (Tübingen, Hiver 2012)

Objectifs du module Ce module, adressé à des étudiants en bioinformatique (niveau équivalent M1 ou M2), avait pour but de les familiariser avec les bases de la fouille de données (classification, régression, regroupements, sélection de variables) et leurs applications en bioinformatique (sélection

de gènes à partir de données d'expression, prédiction de la fonction de protéines, découverte de motifs dans des réseaux biologiques).

Effectifs ~ 25 étudiants.

Participation

- Deux cours magistraux : Fouille de graphes et fouille de motifs fréquents¹, et Fouille de graphes en chémoinformatique² ;
- Conception et rédaction de la partie de l'examen portant sur les cours assurés ;
- Conception et rédaction d'un des projets proposés aux étudiants (Classification de graphes) ;
- Assistance aux étudiants dans la réalisation de leurs projets ;

5. Séminaire de bioinformatique (Tübingen, Été 2012)

Objectifs du module Ce module, adressé à des étudiants en bioinformatique (niveau équivalent M1 ou M2), avait pour but de les familiariser avec la lecture d'articles scientifiques en leur faisant lire et exposer un article récemment publié dans une conférence de bioinformatique (ISMB ou RECOMB).

Effectifs ~ 15 étudiants.

Participation

- Encadrement suivi de deux élèves et évaluation de leurs rapports écrits ;
- Évaluation des exposés oraux des élèves.

6. Encadrement de travaux de Master

Encadrement dans le groupe « Machine Learning and Computational Biology » du stage de Master Recherche de Valeri Velkov, sur le sujet « Mining correlated loci at a genome-wide scale ».

Encadrement (depuis le 1er octobre 2012) dans le groupe « Machine Learning and Computational Biology » du stage de Master Recherche de Fabian Aicheler, sur le sujet « Disease status prediction based on SNP annotations ».

Activités de recherche

Mes travaux de recherche s'inscrivent dans le cadre du développement de techniques d'apprentissage automatique et statistique pour les sciences de la vie. Je m'intéresse particulièrement à la recherche thérapeutique : comment extraire de quantités toujours plus larges de données biologiques, chimiques, médicales et pharmaceutiques des informations qui facilitent la découverte et le développement de nouvelles thérapies, depuis la compréhension des mécanismes biologiques sous-jacents jusqu'à la découverte de médicaments candidats et leur synthèse.

1. Noyaux pour séquences biologiques

Pendant mon projet de recherche en M2, j'ai implémenté des noyaux pour séquences et développé une interface web permettant d'entraîner et de tester des machines à vecteur de support utilisant ces noyaux. J'ai ensuite utilisé cette interface pour évaluer différentes méthodes de classification de protéines à signaux et montré que le noyau d'alignement local permet d'obtenir des résultats similaires à l'état de l'art.

1. transparents disponibles sur : http://agbs.kyb.tuebingen.mpg.de/wikis/bg/DIB12_3.pdf

2. transparents disponibles sur : http://agbs.kyb.tuebingen.mpg.de/wikis/bg/DIB12_8.pdf

2. Chémoinformatique et recherche médicamenteuse

Durant mon doctorat à UCI, dirigé par Pierre Baldi, j'ai travaillé pendant cinq ans sur des projets d'apprentissage automatique dont le but était de faciliter le développement de nouveaux médicaments, grâce à des modèles prédictifs permettant de quantifier les propriétés d'une molécule telles que sa toxicité ou son activité thérapeutique.

2.1. Représentation de petites molécules La plupart des médicaments sont des molécules de faible masse moléculaire, composées de quelques douzaines d'atomes, par opposition aux macromolécules telles que les protéines sur lesquelles elles agissent et peuvent en comporter plusieurs milliers. Pour construire des modèles prédictifs utiles dans la recherche médicamenteuse, il est donc essentiel de développer des représentations riches et informatives de telles molécules. La première étape de mon travail de doctorat a donc été de développer et d'étudier des descripteurs pour petites molécules dérivés systématiquement de représentations structurales variant d'une à quatre dimensions. En particulier, j'ai utilisé ces différentes représentations en combinaison avec des machines à vecteurs de support pour plusieurs problèmes de classification et de régression et j'ai montré qu'en pratique, les représentations dites 2D dérivées du graphe moléculaire sont souvent les plus informatives [5].

2.2. Criblage virtuel haut-débit Je me suis ensuite intéressée à la prédiction d'activité thérapeutique et au criblage virtuel : comment exploiter au mieux un criblage exploratoire en laboratoire pour trier des substances non encore testées en fonction de leur activité envers la même cible. J'ai ainsi élaboré avec Joshua Swamidass un nouvel algorithme à base de réseaux de neurones, IRV (Influence Relevance Voter), qui est adapté à ce problème et plus performant que l'état de l'art (permettant de retrouver jusqu'à trois fois plus de composés actifs parmi ceux dans le premier pourcent de la liste triée retournée par l'algorithme) [4].

L'évaluation d'IRV m'a amenée à étudier la quantification de la capacité d'un algorithme à la « reconnaissance précoce », autrement dit à présenter les substances les plus actives au tout début de la liste triée afin de permettre de concentrer les efforts expérimentaux sur un faible nombre de composés. J'ai ainsi développé CROC, une extension des courbes ROC permettant la quantification, la visualisation, et l'optimisation de la reconnaissance précoce [3].

2.3. Prédiction de réactions chimiques Je me suis aussi intéressée à une autre étape cruciale et coûteuse du *pipeline* pharmaceutique : la synthèse moléculaire. Plus précisément, j'ai étudié la prédiction du déroulement de réactions chimiques. Avec Matthew Kayala, nous nous sommes éloignés de l'approche système-expert généralement adoptée, et avons construit et rendu public un jeu d'apprentissage de mécanismes élémentaires. Nous avons ensuite dérivé des travaux décrits en 2.1 des descripteurs adaptés à la classification des atomes d'une molécule selon leur appartenance à un site actif et à trier ces sites actifs par favorabilité. Ce travail nous a permis de construire le premier prédicteur de réactions chimiques capable de généraliser à des types de réactions non incluses dans son jeu d'apprentissage [2].

3. Découverte d'associations pangénomiques

3.1. Annotations de marqueurs génétiques J'ai commencé à m'intéresser à un aspect complémentaire de la recherche thérapeutique pendant le stage de recherche que j'ai effectué au sein de l'équipe d'apprentissage automatique d'IBM R&D à Tel-Aviv en 2009, sous la direction de Michal Rosen-Zvi. À cette occasion, je me suis familiarisée avec les études d'association pangénomiques (*Genome-Wide Association Studies*, ou GWAS) et les techniques statistiques traditionnellement employées dans leur cadre. Ces études ont pour but d'identifier quels loci génétiques (souvent

appelés SNPs pour *Single Nucleotide Polymorphisms*, ou polymorphismes d'un seul nucléotide) sont individuellement associés avec un phénotype particulier. En pratique, elles ne permettent le plus souvent pas d'expliquer la totalité de variabilité phénotypique que l'on sait d'origine génétique. Afin de surmonter ces limitations, j'ai travaillé, dans le cadre du projet européen HyperGene, sur l'utilisation d'annotations de SNPs pour dériver la probabilité *a priori* de leurs effets et augmenter ainsi la puissance statistique des études GWA.

L'amélioration récente des connaissances biologiques sur les caractéristiques de mutations délétères permet désormais d'envisager de développer des représentations plus riches de SNPs qui intègrent ces connaissances. C'est le sujet du stage de Master Recherche de Fabian Aicheler que j'encadre depuis octobre 2012.

J'ai en effet reçu pour poursuivre mes travaux en génétique statistique une bourse de la fondation Humboldt et travaille depuis mars 2011 avec Karsten Borgwardt à l'Institut Max Planck de Tübingen sur des méthodes permettant de mieux exploiter les données GWA. La prise en compte des effets joints de multiples SNPs multiples, une autre piste à explorer pour améliorer l'interprétation de données GWA, est le cadre dans lequel s'inscrivent majoritairement mes travaux actuels.

3.2. Effets de deux loci génétiques Un obstacle majeur aux études multi-locus est la simple quantité de variables à prendre en compte ; une étude GWA peut concerner plusieurs millions de loci pour plusieurs milliers ou dizaines de milliers d'individus. Un simple modèle linéaire de type $y = \alpha x_1 + \beta x_2 + \gamma x_1 x_2 + \delta$ (où y est le phénotype et x_1, x_2 deux loci susceptibles de l'expliquer conjointement) peut donc requérir d'évaluer de l'ordre de 10^{12} à 10^{14} paires de loci (x_1, x_2) , ce qui a un coût prohibitif sur une seule machine. Cependant, ces évaluations se prêtent particulièrement bien au paradigme GPGPU (*General-Purpose computing on Graphic Processing Units*), qui permet d'exploiter les capacités de parallélisation d'opérations simples des cartes graphiques utilisées pour l'affichage vidéo des ordinateurs modernes. J'ai participé au développement de GLIDE, un outil qui permet d'explorer exhaustivement un jeu de données sur carte graphique. GLIDE permet d'explorer exhaustivement les paires formées par 500 000 SNPs pour 5 000 individus en seulement 3 jours sur un carte graphique unique (contre plus d'un an sur un seul CPU) [1].

L'étude des effets joints de deux SNPs sur un phénotype requiert à mon sens de mieux comprendre les corrélations entre paires de SNPs au niveau de populations entières et indépendamment de phénotypes. C'était le but du stage de Master Recherche de Valeri Velkov que j'ai encadré, et qui a conclu que l'hypothèse que la coévolution de certains gènes contribue à la formation de corrélations entres positions éloignées du génomes est statistiquement raisonnable.

3.3. Effets de réseaux de loci génétiques Si des techniques de parallélisation intensive permettent d'évaluer systématiquement les effets joints de paires de loci, l'étude exhaustive de groupes de plus de deux loci est pour l'instant hors de portée à moins d'exploiter des connaissances préalables pour réduire drastiquement le nombre de tels groupes. Plutôt que de se restreindre à des loci appartenant par exemple à quelques voies de signalisation pré-sélectionnées comme étant susceptibles de jouer un rôle dans le phénotype examiné, je propose de découvrir automatiquement un réseau de loci qui soit maximalelement associé avec le phénotype. Ce réseau est défini comme un sous-réseau d'un réseau de loci pré-établi sur la base, par exemple, d'un réseau d'interactions entre protéines. En termes d'apprentissage automatique, il s'agit d'un problème de sélection de variables sous contraintes de parcimonie et de connectivité. Mes premiers efforts dans cette direction ont conduit au développement de SConES (*Selection of Connected Explanatory SNPs*, actuellement soumis à publication, présenté au workshop MLCB en décembre 2012 [8]), qui résout ce problème efficacement dans le cas d'un modèle additif grâce à une reformulation flot-max/coupe-min.

Publications et interventions lors de colloques

Articles dans des revues à comité de lecture

Tony Kam-Thong, Chloé-Agathe Azencott, Lawrence Cayton, Benno Pütz, André Altmann, Nazanin Karbalai, Philipp G. Sämann, Bernhard Schölkopf, Betram Müller-Myhsok, and Karsten M. Borgwardt. **GLIDE: GPU-based Linear Regression for the Detection of Epistasis**. *Human Heredity*, 73:220–236, 2012.

<http://content.karger.com/ProdukteDB/produkte.asp?doi=10.1159/000341885>.

Matthew A. Kayala, Chloé-Agathe Azencott, Jonathan H. Chen, and Pierre Baldi. **Learning to Predict Chemical Reactions**. *Journal of Chemical Information and Modeling*, 51(9):2209–2222, 2011.

<http://pubs.acs.org/doi/abs/10.1021/ci200207y>.

S. Joshua Swamidass, Chloé-Agathe Azencott, Kenny Daily, and Pierre Baldi. **A CROC Stronger Than ROC: Measuring, Visualizing and Optimizing Early Retrieval**. *Bioinformatics*, 26(10):1348–1356, 2010.

<http://bioinformatics.oxfordjournals.org/cgi/content/full/26/10/1348>.

S. Joshua Swamidass, Chloé-Agathe Azencott, Ting-Wan Lin, Hugo Gramajo, Sheryl Tsai, and Pierre Baldi. **The Influence Relevance Voter: an Accurate and Interpretable Virtual High Throughput Screening Method**. *Journal of Chemical Information and Modeling*, 49(4):756–766, 2009.

<http://pubs.acs.org/doi/abs/10.1021/ci8004379>.

Chloé-Agathe Azencott, Alexandre Ksikes, S. Joshua Swamidass, Jonathan H. Chen, Liva Ralaivola, and Pierre Baldi. **One- to Four-Dimensional Kernels for Virtual Screening and the Prediction of Physical, Chemical and Biological Properties**. *Journal of Chemical Information and Modeling*, 47(3):965–974, 2007.

<http://pubs.acs.org/doi/abs/10.1021/ci600397p>.

Chapitre d'ouvrage

Chloé-Agathe Azencott and Pierre Baldi. **Virtual High-Throughput Screening With Two-Dimensional Kernels**, In *Hands-On Pattern Recognition: Challenges in Machine Learning*, volume 1, pages 131–146. Microtome, 2011.

<http://www.clopinet.com/ChallengeBook.html>.

Thèse de doctorat

Chloé-Agathe Azencott. **Statistical Machine Learning and Data Mining for Chemoinformatics and Drug Discovery**. PhD thesis, University of California, Irvine, 2010. ProQuest/UMI, AAT 3422105.

<http://gradworks.umi.com/34/22/3422105.html>.

Communications dans des colloques internationaux avec comité de lecture

Chloé-Agathe Azencott, Dominik Grimm, Yoshinobu Kawahara, and Karsten M. Borgwardt. **Efficiently Mapping Phenotypes to Networks of Genetic Loci**. Machine Learning and Computational Biology Workshop (poster), 2012.

<http://www.mlcb.org/>.

Chloé-Agathe Azencott, Matthew A. Kayala, and Pierre Baldi. **PropOrb: A Frontier Molecular Orbital Interaction Proposer**. 239th American Chemistry Society National

Meeting, 2010.

<http://acscinf.org/docs/publications/CIB/2010/1/mtgs/cinfprogram.asp?nth=239>.

Chloé-Agathe Azencott, Matthew A. Kayala, and Pierre Baldi. **Combining Quantitative Data and Qualitative Knowledge to Score Reaction Energies**. 237th American Chemistry Society National Meeting, 2009.

<http://acscinf.org/docs/meetings/237nm/237cinfabstracts.htm>.

Chloé-Agathe Azencott, S. Joshua Swamidass, and Pierre Baldi. **Performance Prediction of the Influence Relevance Voter**. The Learning Workshop (poster), 2009.

<http://snowbird.djvuzone.org/2009/posters.html#a>.

Chloé-Agathe Azencott, S. Joshua Swamidass, and Pierre Baldi. **Virtual High-Throughput Screening and Early Recognition**. Women in Machine Learning Workshop (poster), 2009.

http://acsweb.ucsd.edu/~tnarayan/past_workshops/2009/index.php.

Chloé-Agathe Azencott and Pierre Baldi. **Virtual High-Throughput Screening with Two-Dimensional Kernels**. Agnostic Learning vs. Prior Knowledge Workshop, International Joint Conference on Neural Networks, 2007.

<http://www.agnostic.inf.ethz.ch/>.

Chloé-Agathe Azencott and Pierre Baldi. **Kernels for Predictive Regression - Physical, Biological and Chemical Properties of Small Molecules**. Workshop for Women in Machine Learning, 2006.

<http://researchvm-9.cs.rutgers.edu/2006/wiml06/slides/>.

Des informations complémentaires (notamment textes des articles et de la thèse, présentations et posters) sont disponibles sur la page que je maintiens à <http://cazencott.info>.

Liste des pièces jointes

Documents administratifs

1. Copie du diplôme de doctorat, délivré par University of California Irvine ;
2. Traduction en français du diplôme ;
3. À défaut du rapport de soutenance (inexistant à University of California Irvine), une copie de la page de garde de la thèse de doctorat, signée par les membres du jury de thèse ;
4. Traduction en français de cette dernière.

Exemplaires des travaux de recherche

1. Copie de **GLIDE: GPU-based Linear Regression for the Detection of Epistasis**, Tony Kam-Thong, Chloé-Agathe Azencott, Lawrence Cayton, Benno Pütz, André Altmann, et al., *Human Heredity*, 2012.
2. Copie de **A CROC Stronger Than ROC: Measuring, Visualizing and Optimizing Early Retrieval**, S. Joshua Swamidass, Chloé-Agathe Azencott, Kenny Daily, and Pierre Baldi, *Bioinformatics*, 2010.
3. Copie de **One- to Four-Dimensional Kernels for Virtual Screening and the Prediction of Physical, Chemical and Biological Properties.**, Chloé-Agathe Azencott, Alexandre Ksikes, S. Joshua Swamidass, Jonathan H. Chen, Liva Ralaivola, and Pierre Baldi, *Journal of Chemical Information and Modeling*, 2007.