

F. Dioury^c, A. Duprat^{a,b}, C. Ferroud^c, J. Cossy^b, G. Dreyfus^a

^a Laboratoire SIGnaux, Modèles, Apprentissage (SIGMA), ESPCI ParisTech, Paris, ^b Laboratoire de Chimie Organique, ESPCI ParisTech, Paris

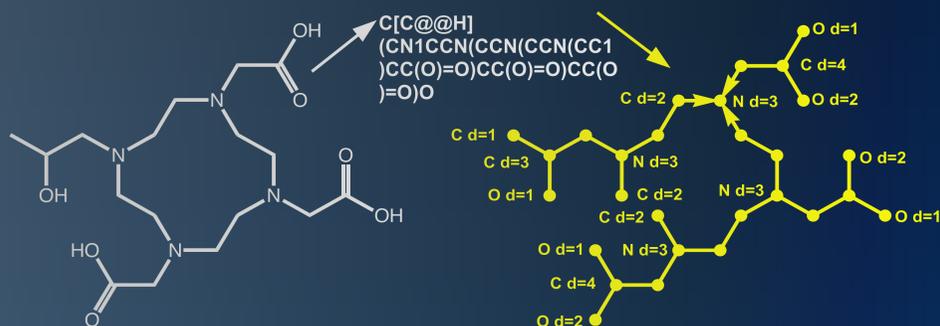
^c Laboratoire de Transformations Chimiques et Pharmaceutiques, CNAM, Paris

INTRODUCTION

Les *graph machines* constituent une nouvelle famille de méthodes de modélisation par apprentissage artificiel ; elles permettent de prédire des propriétés et activités de molécules à partir de leurs structures sans utiliser de descripteurs. Nous présentons l'application de cette méthode à la prédiction des constantes de stabilité de complexes de Gadolinium utilisés comme agents de contraste en imagerie IRM.

MÉTHODE

Transformation des structures en graph machines



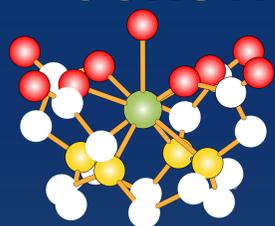
Chaque graph machine est une combinaison de fonctions paramétrées isomorphe du graphe de la molécule, qui associe la valeur à prédire à la structure de la molécule.

Les paramètres identiques pour toutes les machines.

Les paramètres sont estimés par apprentissage statistique.

Les performances sont évaluées par les méthodologies habituelles de l'apprentissage statistique (validation croisée, leave-one-out, jackknife, ...)

CONSTRUCTION DU MODÈLE

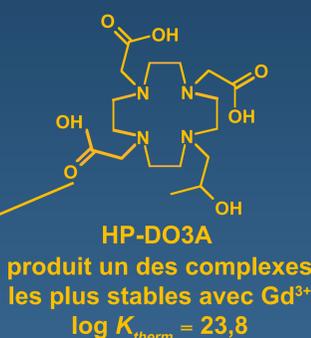
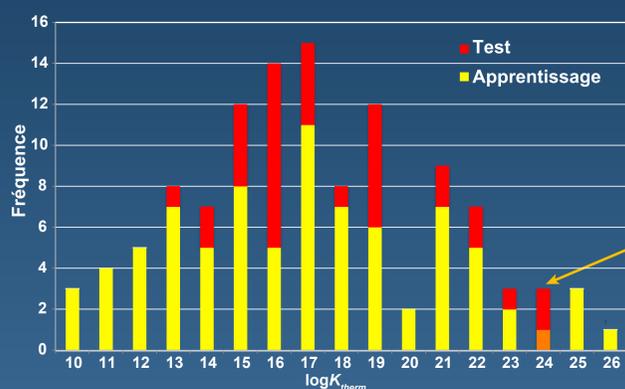


GRAPH
MACHINES

K_{therm} : constante de stabilité
du complexe de gadolinium
correspondant

BASE DE DONNÉES : ligands polyamino-polyacides (carboxyliques, amides, alcools).

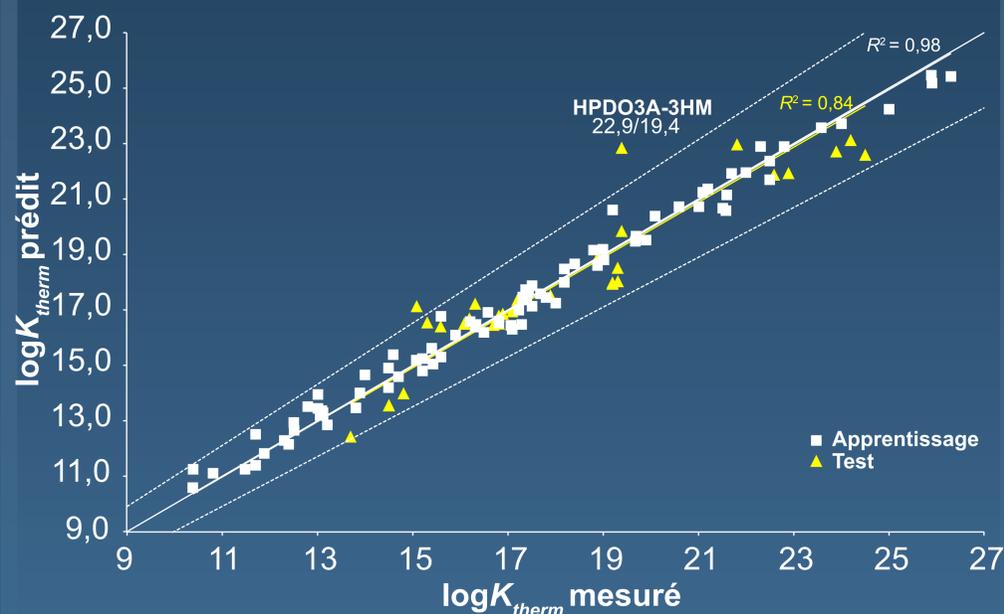
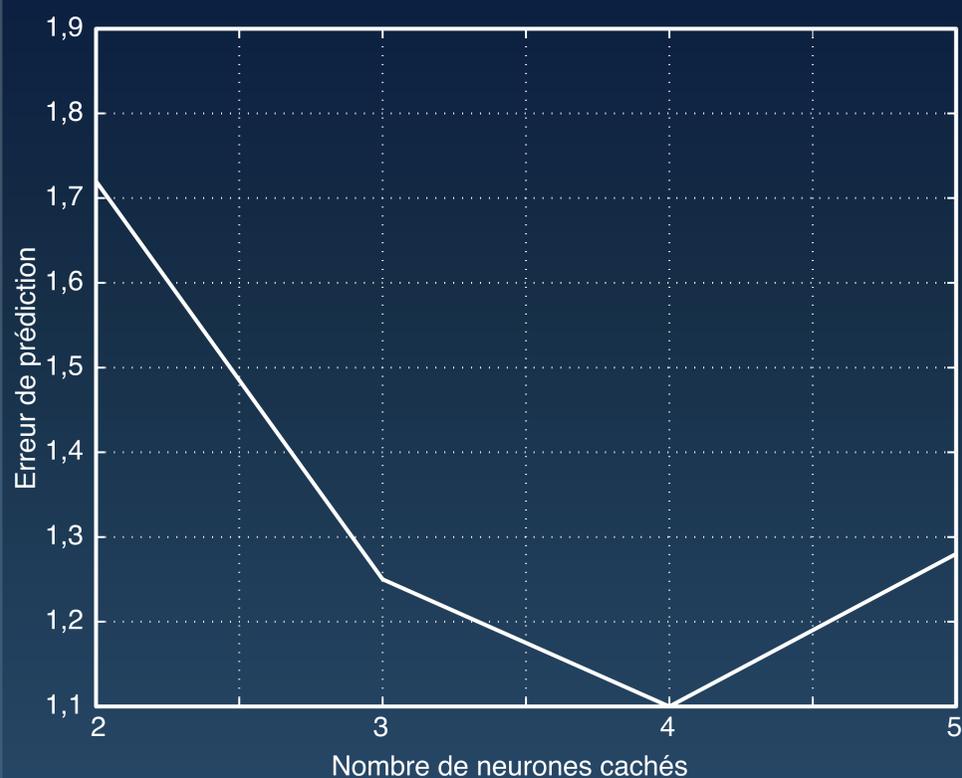
Sélection de 116 molécules munies de valeurs fiables de constantes de stabilité. Partition en un ensemble d'apprentissage (82 ligands) et un ensemble de test (34 ligands).



APPRENTISSAGE ET SÉLECTION DE MODÈLE (leave-one-out, leave-one out virtuel).

TEST sur des molécules nouvelles et aide au choix d'une valeur expérimentale.

RÉSULTATS



CONCLUSIONS ET PERSPECTIVES

- Méthode puissante de codage de graphe moléculaire
- Aucun descripteur moléculaire à calculer et sélectionner
- Chimie combinatoire virtuelle : concevoir de nouvelles molécules actives par prédiction, puis les synthétiser.

Références

- 1 - A. Goulon, T. Picot, A. Duprat, G. Dreyfus, "Predicting activities without computing descriptors: graph machines for QSAR", SAR and QSAR in environmental research, vol. 18, pp. 141 - 153, 2007.
- 2 - A. Goulon, A. Duprat, G. Dreyfus, "Graph machines and their applications to computer-aided drug design; a new approach to learning from structured data", Lecture notes in computer science vol. 4135, pp. 1 - 19, 2006.
- 3 - A. Goulon, A. Duprat, G. Dreyfus, "From Hopfield nets to recursive networks to graph machines: numerical machine learning from structured data", Theoretical Computer Science vol. 344, pp. 298 - 334, 2005.