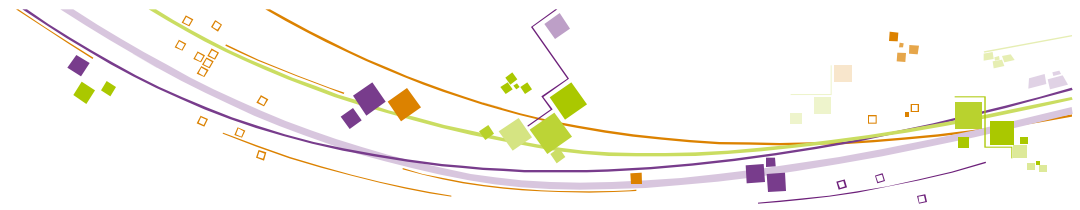


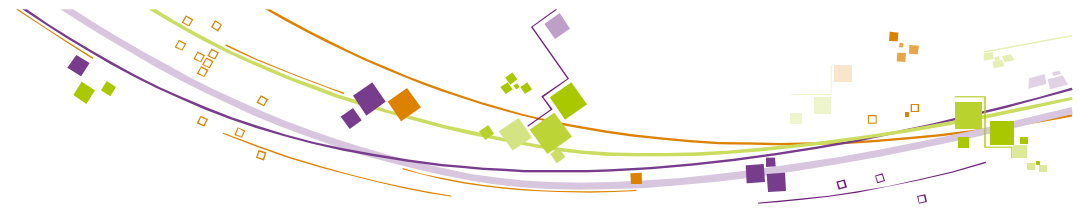
Modélisation des propriétés physico-chimiques des molécules à partir de leur structure

A. Faraj – F. Porcheron – M. Jacquin
IFP Energies nouvelles



Plan

- **Contexte**
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- **Modélisation de la tension de vapeur d'une amine**
 - Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
 - Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
 - Comparaison des résultats des 2 méthodes
- **Planification expérimentale itérative pour les GM**
- **Conclusions**



Plan

■ Contexte

- Captage du CO₂ par des solutions d'amines
- Expérimentation haut débit (EHD)
- Boucle EHD

■ Modélisation de la tension de vapeur d'une amine

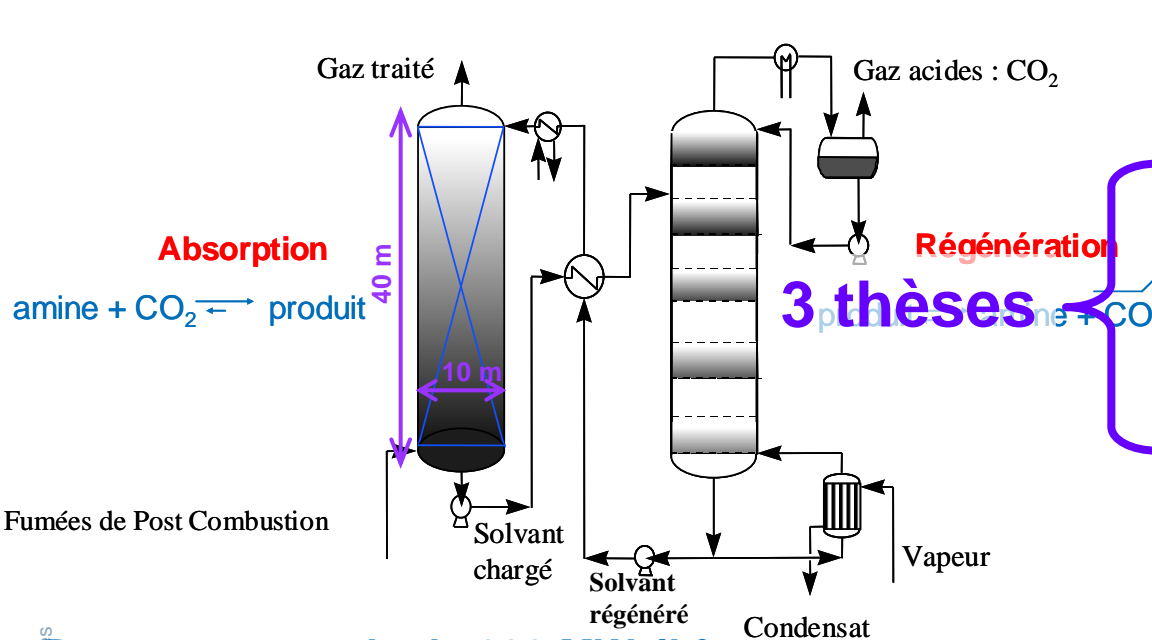
- Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
- Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
- Comparaison des résultats des 2 méthodes

■ Planification expérimentale itérative pour les GM

■ Conclusions

if Traitement de fumées industrielles en postcombustion : captage de CO₂ par solution d'amine

- Ex. : Centrale à charbon T = 40°C – P = 1 bar – %CO₂ = 10%
- Procédé de séparation du CO₂ par un solvant chimique à la MEA (solution la plus viable et utilisée dans le cas du traitement de gaz avec la MDEA)



■ Choix de la "meilleure" amine ...

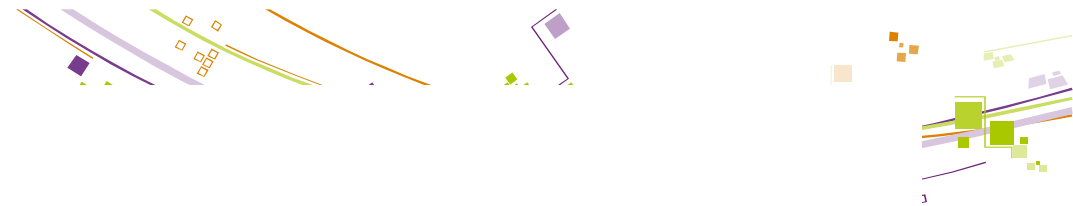
- faible toxicité
- bonne cinétique
- **bonne thermodynamique**
- **très faible tension de vapeur**
- faible dégradation
- faible corrosion
- faible énergie de régénération
- coût raisonnable

- ... dans une base de données de 600 amines

Pour une centrale de 600 MW, il faut
 4 absorbeurs et 2 régénérateurs

Coût actuel 40 à 60 € / tonne de CO₂ capté
 → cible = 20 € / tonne

On capte 90 % du CO₂ (10 % contenu dans les fumées)



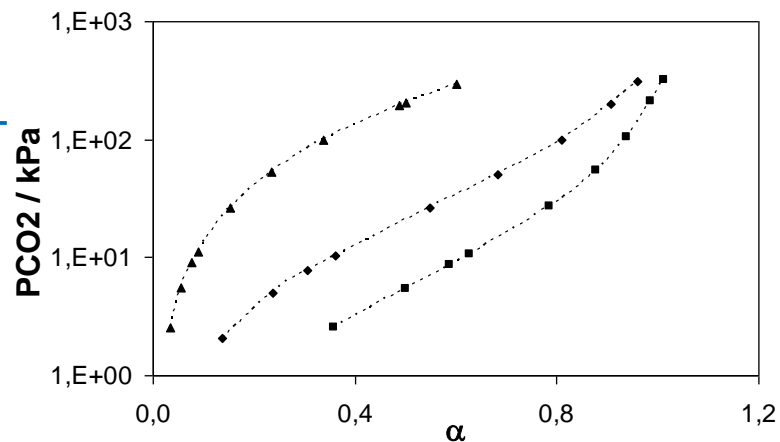
Expérimentation à haut débit (EHD)



- 6 réacteurs en //
- Indépendants
 - T
 - P
 - Injection de gaz : CO₂ [ou NO_x, SO_x, H₂S dans le cas du traitement de gaz naturel]
- Fonctionnement automatique
- Mesures en ligne d'isothermes d'absorption par bilan matière

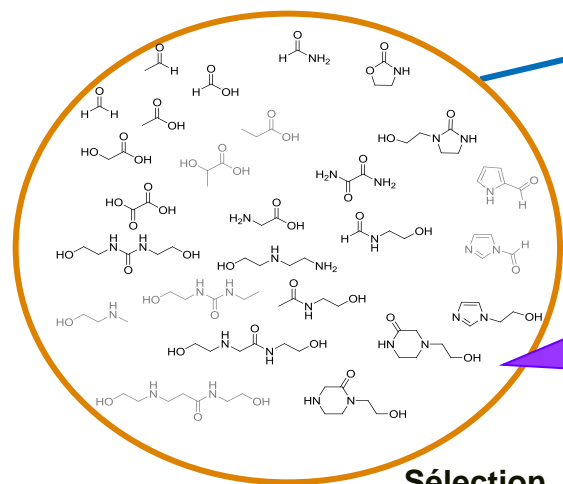
- BbD de 600 amines à tester (500 commerciales + 100 synthétisées à IFPEN)
- Avec les méthodes classiques : 1 isotherme / semaine
- Avec l'EHD : 12 isothermes / semaine
⇒ 140 molécules testées en 2 ans
- Il faudrait encore environ 6 ans pour tester les 460 molécules restantes

⇒ pas encore suffisant !



Boucle EHD

BdD des molécules candidates
500 commerciales + 100 synthétisées



Sélection des expériences à réaliser (expérimentalement)

Screening expérimental

Expérimentation

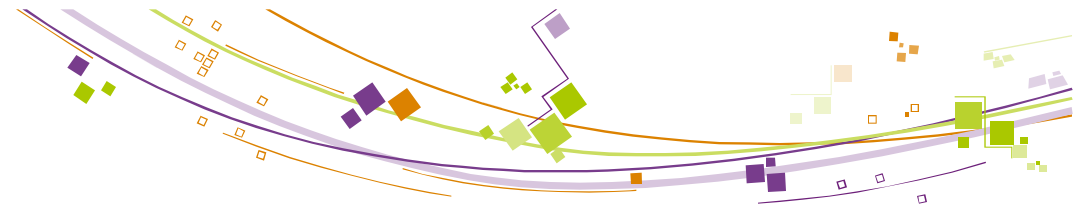
BdD expérimentale

Connaissance chimique

Plans d'expériences

Datamining Modélisation

Relation structure / propriétés

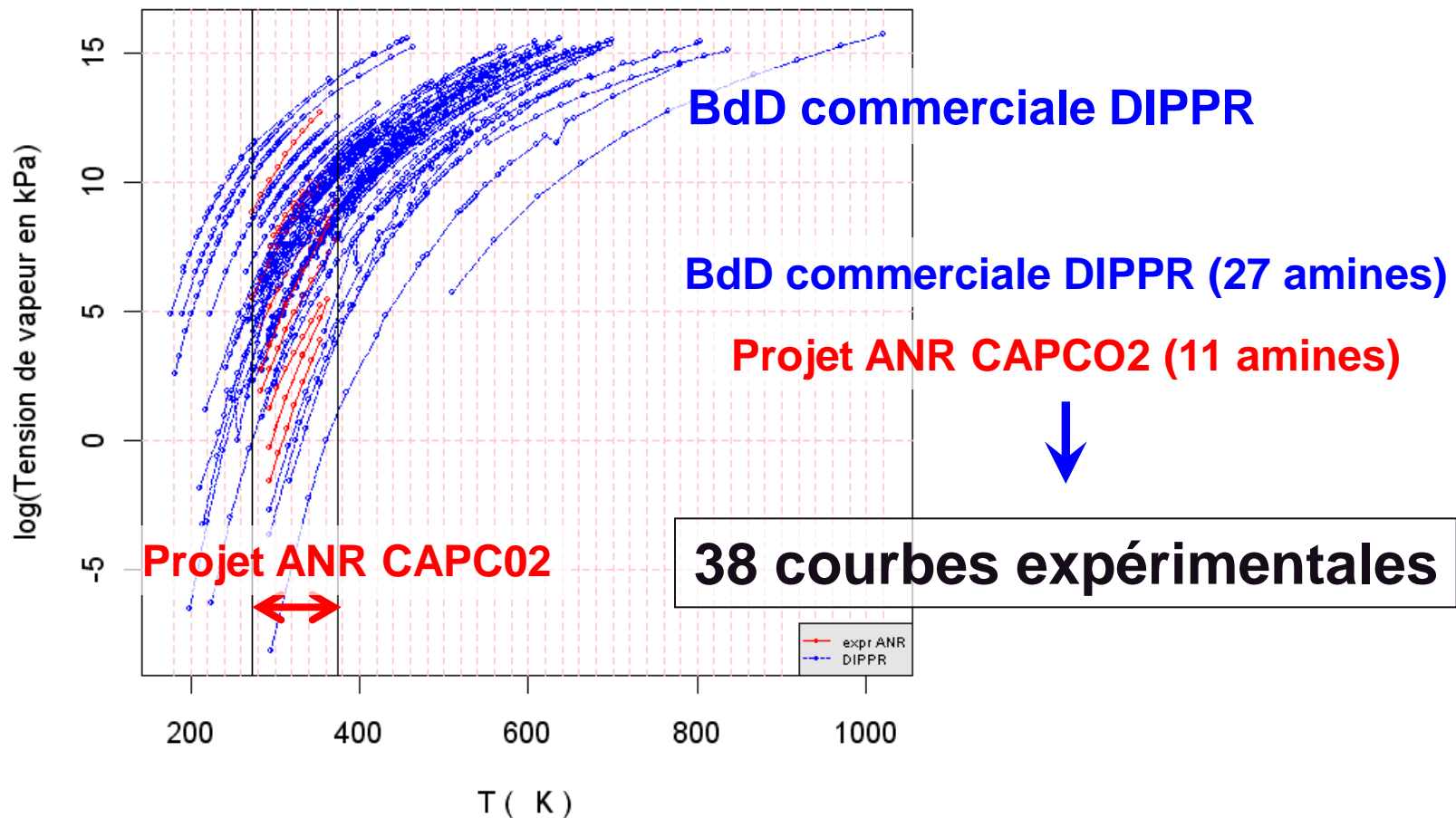


Plan

- **Contexte**
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- **Modélisation de la tension de vapeur d'une amine**
 - Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
 - Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
 - Comparaison des résultats des 2 méthodes
- **Planification expérimentale itérative pour les GM**
- **Conclusions**

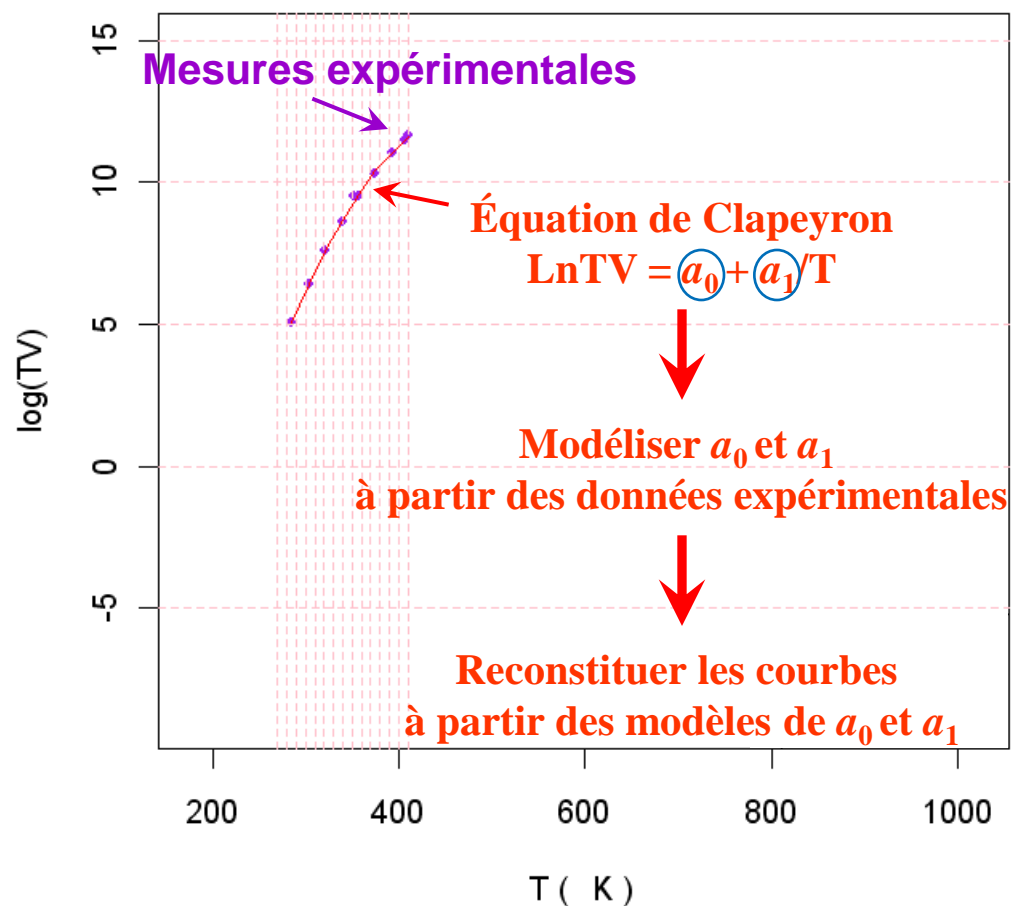
Exemple : modélisation de la courbe de tension de vapeur [Projet ANR CapCO2]

Courbes de tension de vapeur



Exemple : modélisation de la courbe de tension de vapeur [Projet ANR CapCO2]

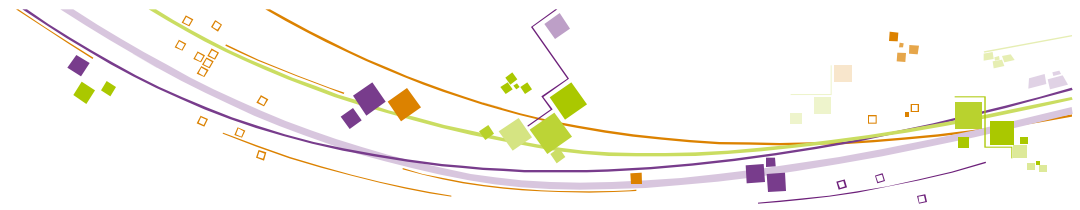
Courbe de tension de vapeur



- Modélisation par *QSAR-MD*
 - Régression PLS
- Modélisation *Graph Machines*
 - Réseau de neurones



Comparaison des résultats des deux modèles



Plan

- Contexte
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- Modélisation de la tension de vapeur d'une amine
 - **Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires**
 - Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
 - Comparaison des résultats des 2 méthodes
- Planification expérimentale itérative pour les GM
- Conclusions

QSAR-MD : Quantitative Structure/ Activity Relationship par descripteurs moléculaires

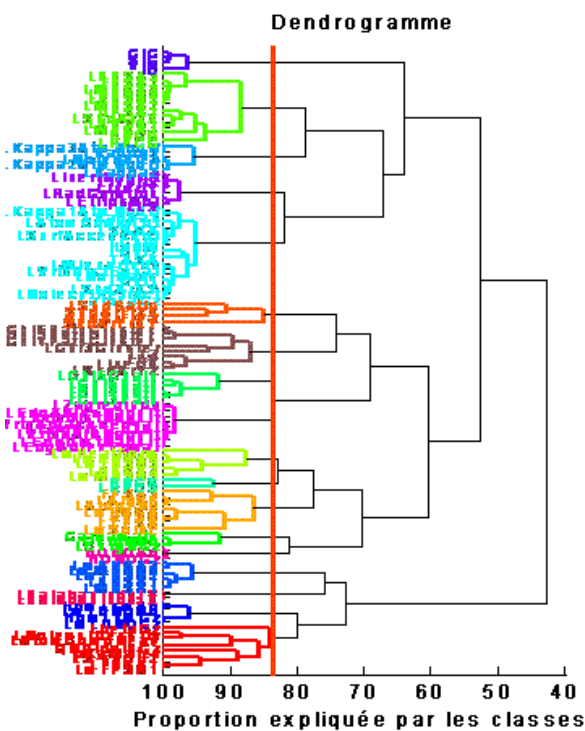
Calcul de 106 descripteurs moléculaires

CAH en 18 classes des descripteurs

Sélection de 45 descripteurs

Propriétés structurales et électroniques des amines

- Charge de l'atome N
- Basicité de la molécule
- Niveau énergétique de orbitales moléculaires
- Moment d'inertie
- Densité moléculaire
- Volume ellipsoïdal
- ...



Modélisation de a_0 et a_1 par régression PLS + leave-one-out

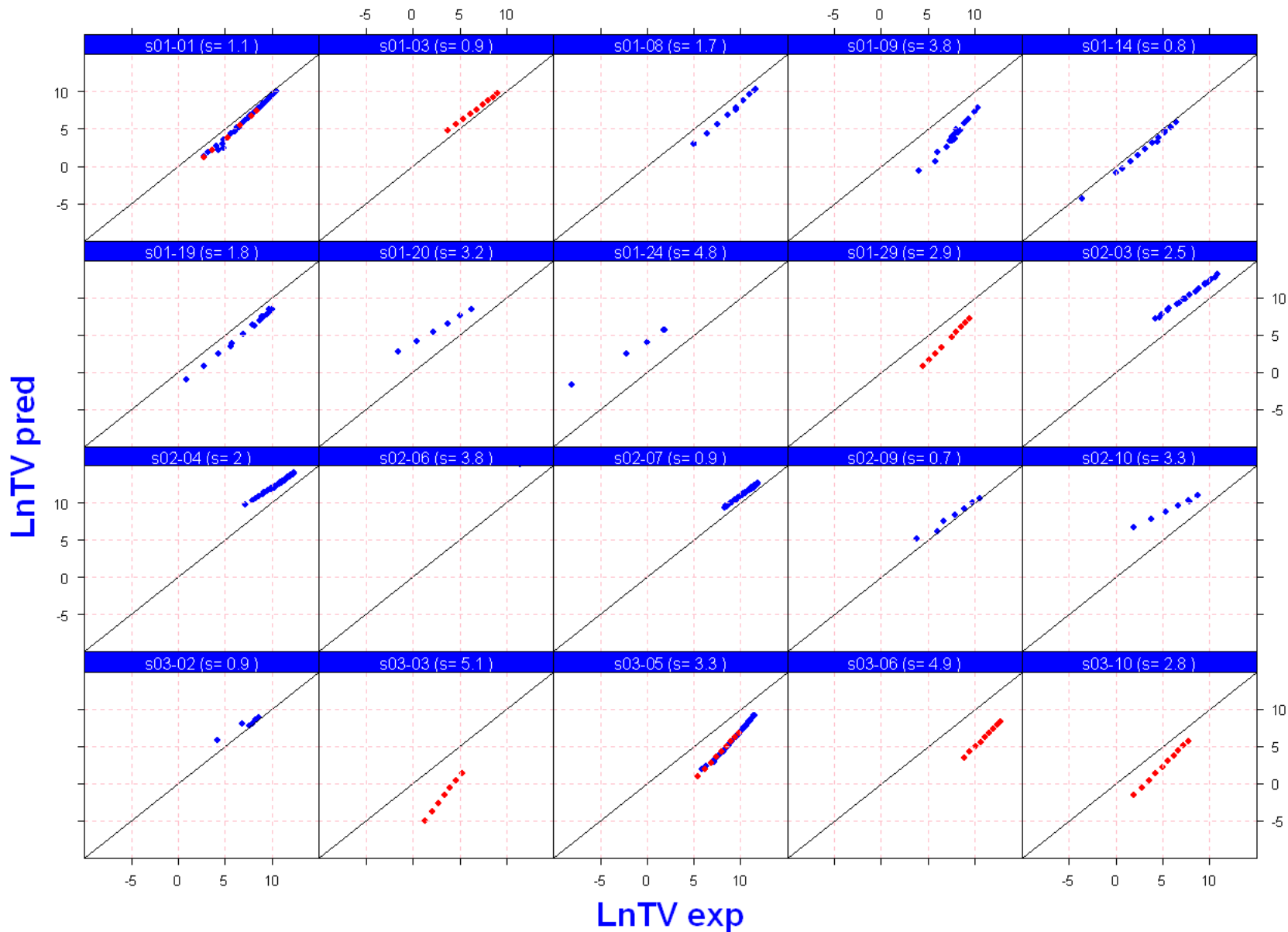
Modèles de a_0 et de a_1 en fonction des 45 descripteurs moléculaires

Reconstitution des courbes de tension de vapeur

$$\hat{a}_0 = \alpha_0^0 + \sum_{j=1,J} \alpha_j^0 X_j$$

$$\hat{a}_1 = \alpha_0^1 + \sum_{j=1,J} \alpha_j^1 X_j$$

Prédiction de la tension de vapeur par la méthode QSAR-MD



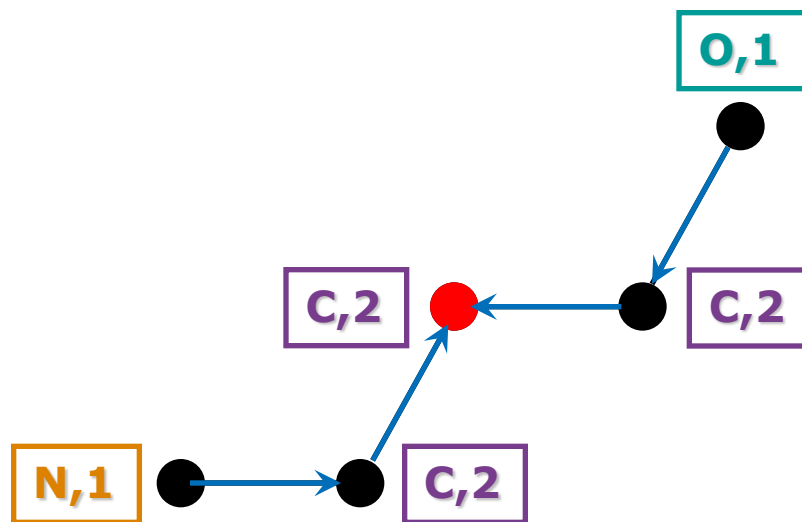


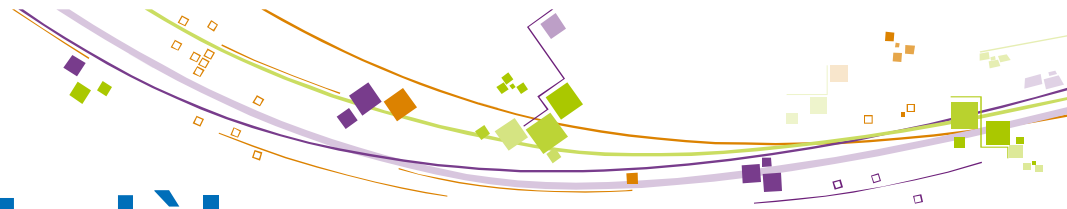
Plan

- Contexte
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- Modélisation de la tension de vapeur d'une amine
 - Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
 - **Méthode des Graph Machines : QSAR-GM**
 - **Codage d'une molécule en graphe**
 - **Construction du modèle GM par apprentissage statistique**
 - Comparaison des résultats des 2 méthodes
- Planification expérimentale itérative pour les GM
- Conclusions

Construction Graphe

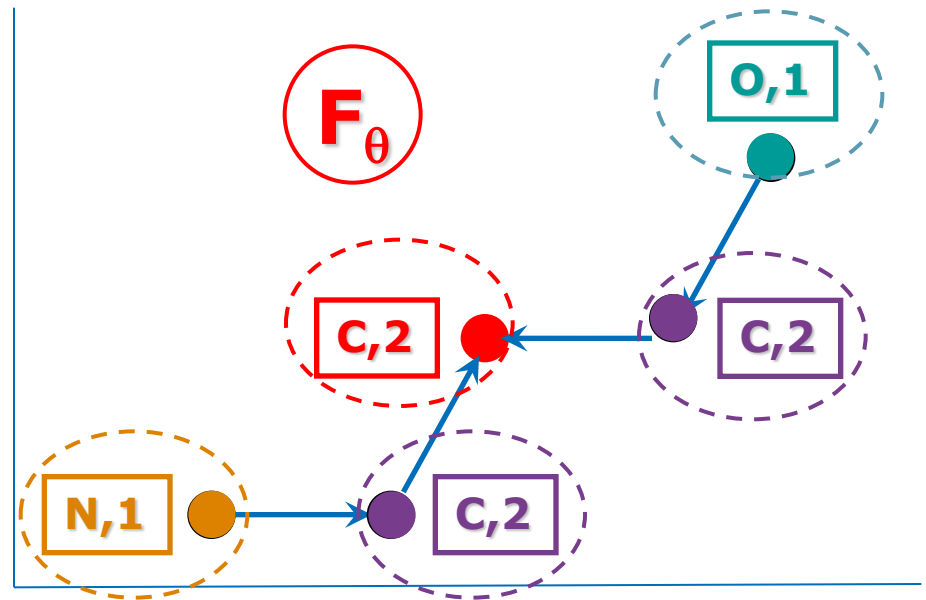
- **Formule chimique d'une molécule :** $\text{H}_2\text{N}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{OH}$
(3-amino-propane-1-ol)
- **Chaque atome non - H = Noeud**
- **Choix d'un noeud central**
- **Chaque liaison chimique = Arête orientée vers le noeud central**
- **Étiquette = Nature de l'atome + Nombre de liaisons**





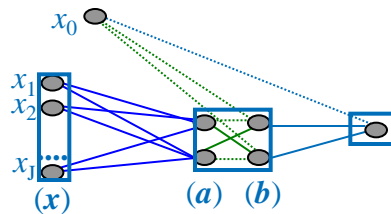
Construction Modèle

- F_θ = fonction "Graph Machine" assignée au noeud central
- Construction de F_θ : à partir d'une fonction élémentaire f_θ



$$F_\theta = f_\theta(f_\theta(f_\theta(0,0,N,1),0,C,2),f_\theta(f_\theta(0,0,O,1),0,C,2),C,2)$$

f_θ est un réseau de neurones



$$f_\theta(x_1, x_2, \dots, x_j) = v_{\ell 0} + \sum_{k=1}^K v_{k\ell} b_k$$

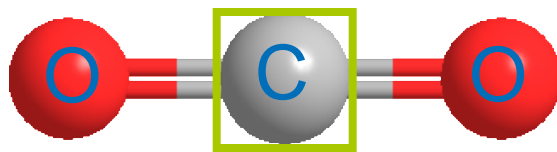
$$a_k = \sum_{j=1}^J u_{jk} x_j$$

$$b_\ell = th\left(w_{\ell 0} + \sum_{k=1}^K w_{k\ell} a_k\right)$$

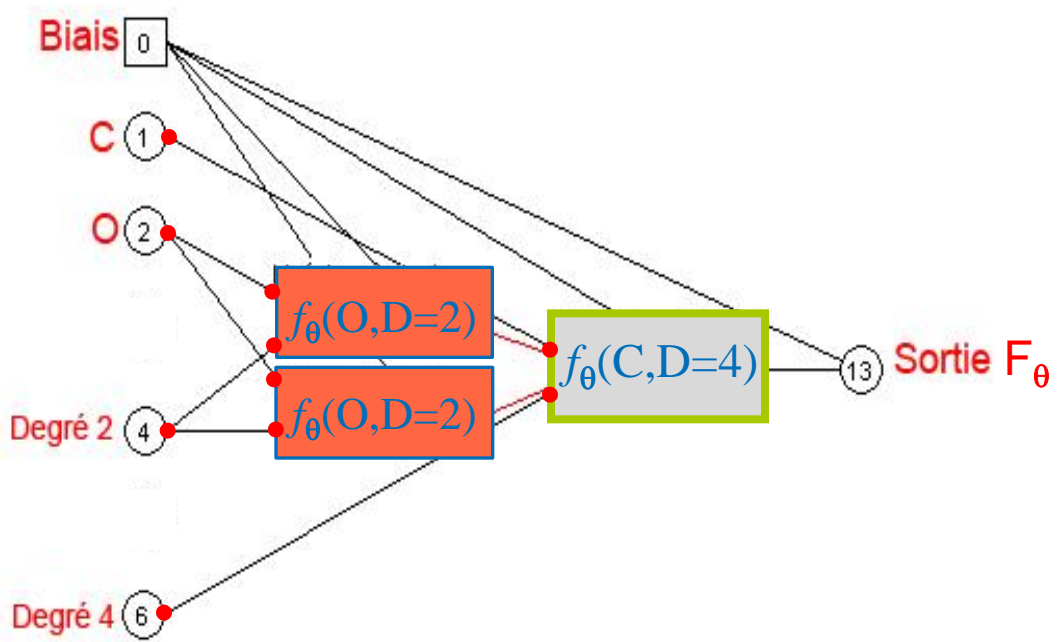
$$th(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Exemples de Graph Machines

Réseau associé à la molécule de dioxyde de carbone

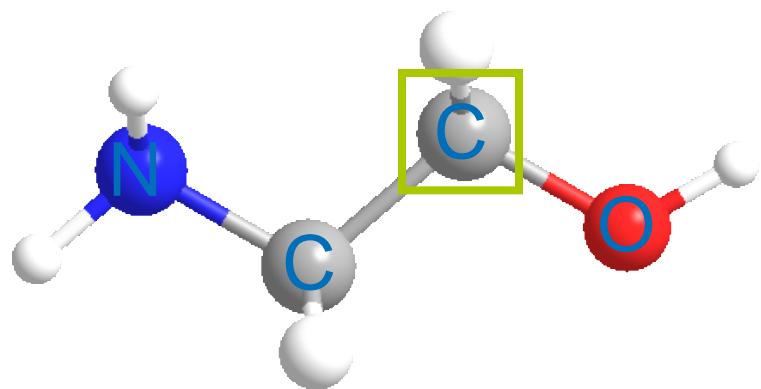


$$F_{\theta} = f_{\theta} (f_{\theta} (0,0,0,0,O,D=2) , f_{\theta} (0,0,0,0,O,D=2) , 0,0,C,D=4)$$

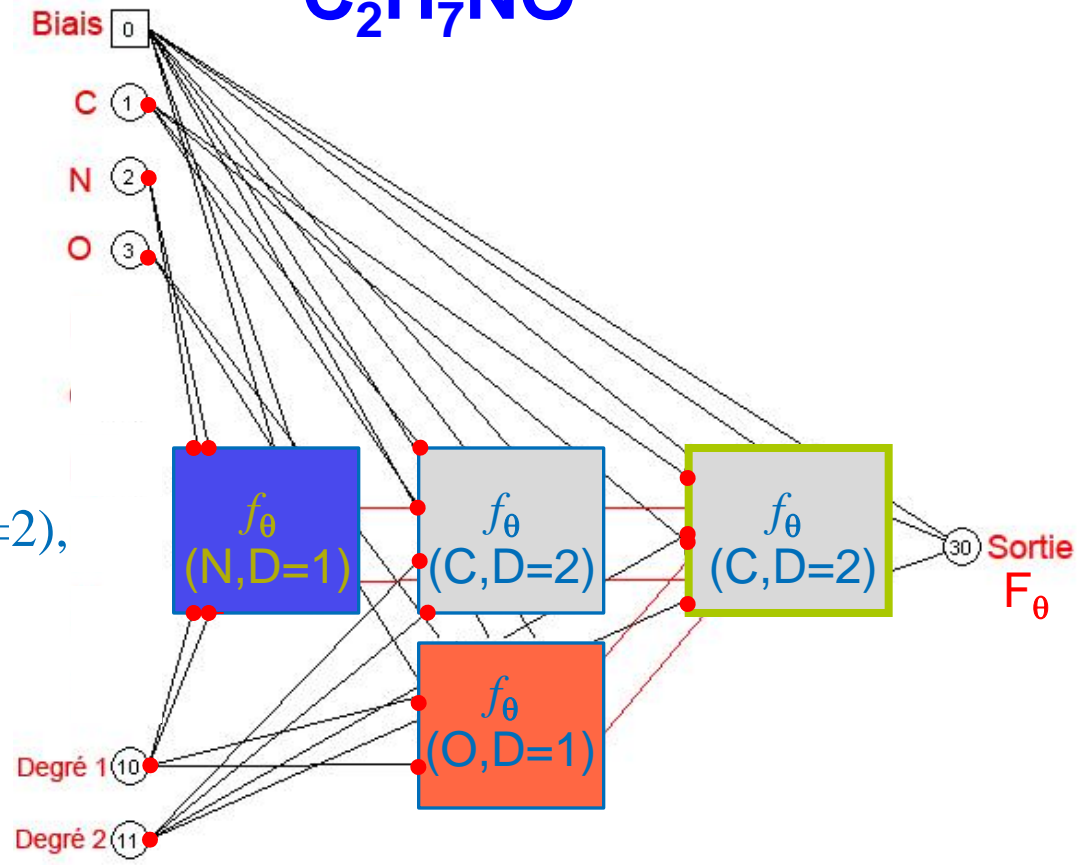


Exemples de Graph Machines

Réseau associé à la monoéthanolamine

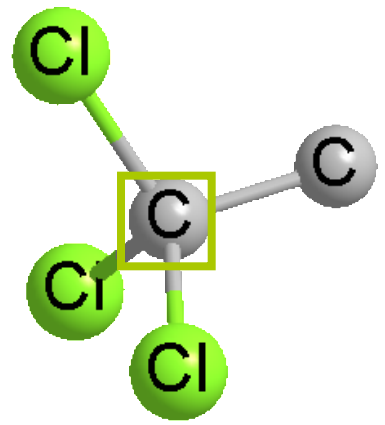
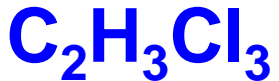


$$F_{\theta} = (f_{\theta}(f_{\theta}(f_{\theta}(0,0,N,D=1), 0,C,D=2), f_{\theta}(0,0,O,D=1), C,D=2))$$

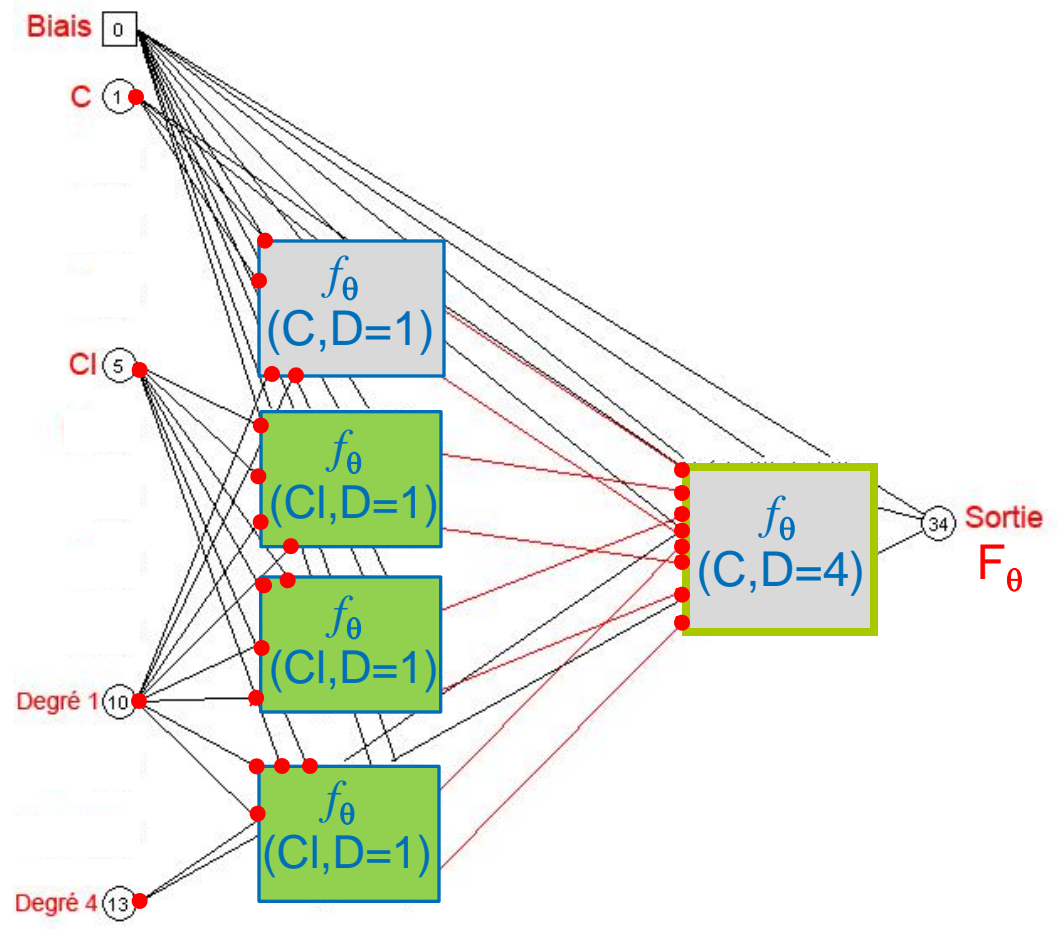


Exemples de Graph Machines

Réseau associé au 1,1,1-trichloroéthane



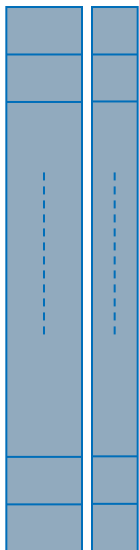
$$F_{\theta} = f_{\theta} (f_{\theta} (0,0,0,0,C,D=1) , \\ f_{\theta} (0,0,0,0,Cl,D=1) , \\ f_{\theta} (0,0,0,0,Cl,D=1) , \\ f_{\theta} (0,0,0,0,Cl,D=1) , \\ C,D=4)$$



Construction du modèle GM par apprentissage

n molécules

smiles y



Réponse :
propriété à
modéliser

= { n molécules + n mesures y_i de la réponse y }



n graphes acycliques orientés



**n fonctions paramétrées $F_\theta(i)$
composées de fonctions f_θ
traduisant la structure des graphes,
 θ est identique pour les n fonctions**



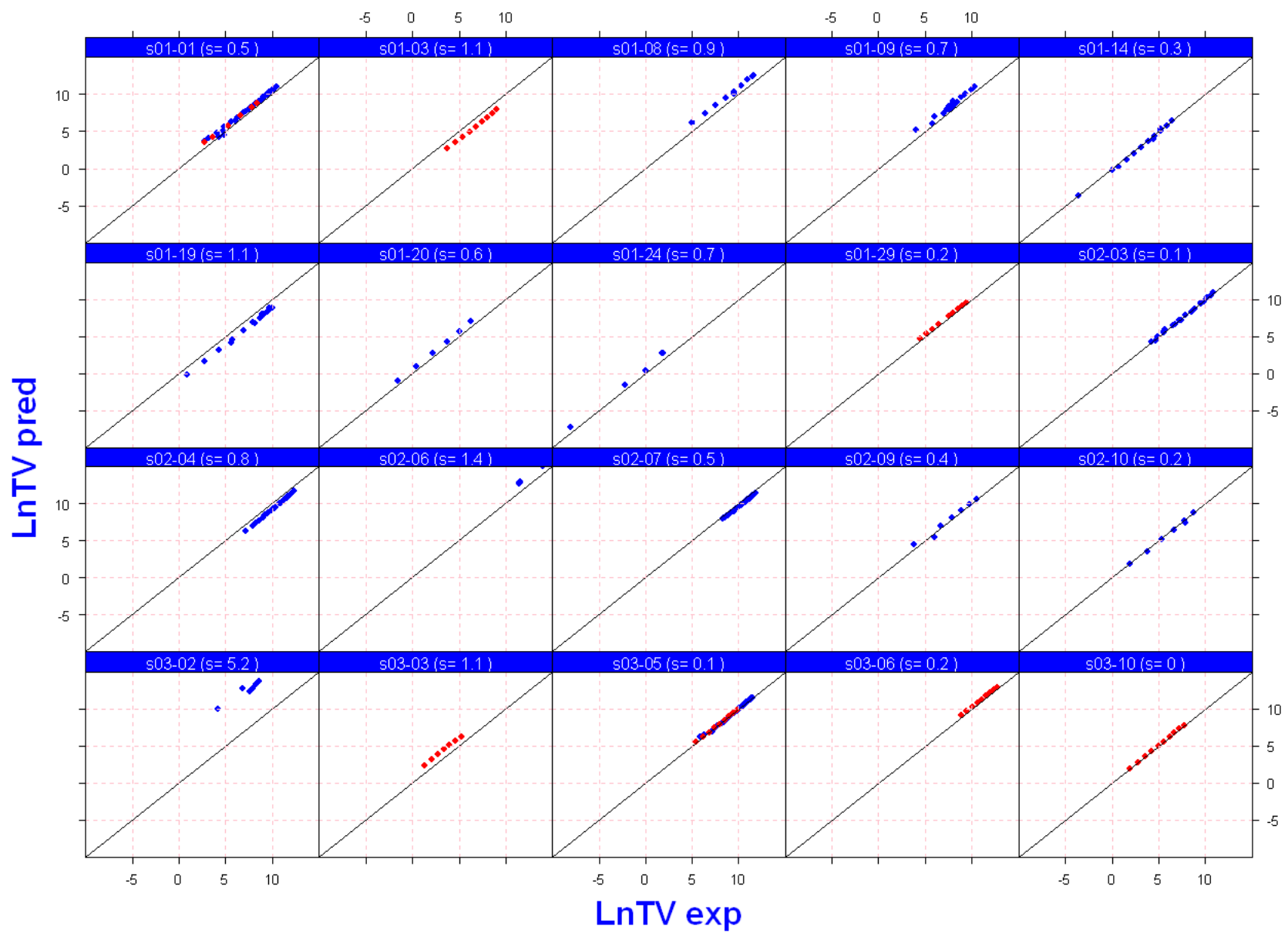
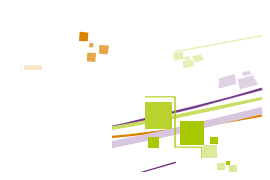
Estimation de θ par minimisation de la fonction de coût

$$J(\theta) = \sum_{i=1}^n \left(y_i - F_\theta^{(i)} \right)^2$$



Solution : $F_{\hat{\theta}}$

Prédiction de la tension de vapeur par la méthode des GM





Plan

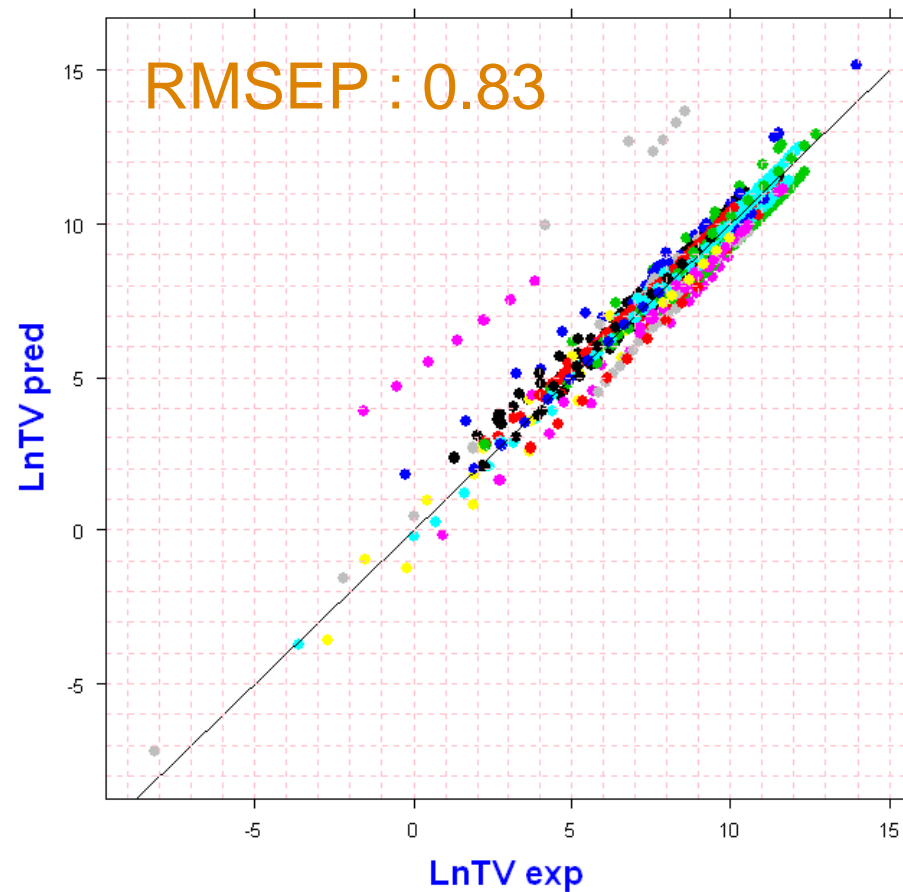
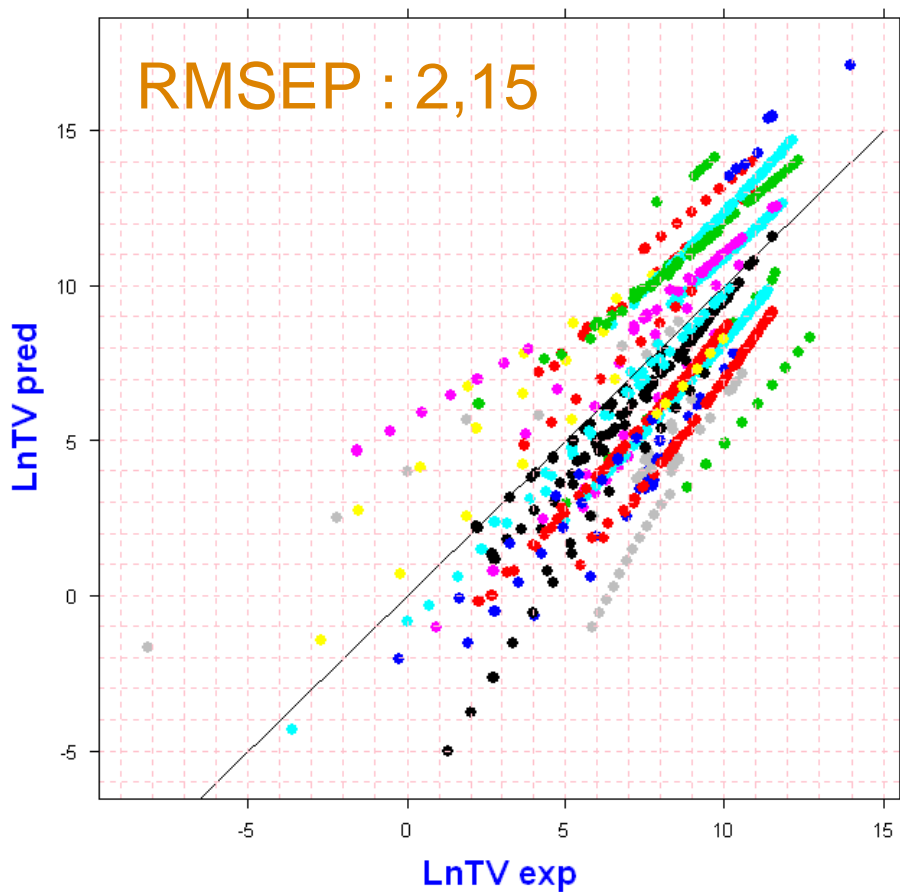
- **Contexte**
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- **Modélisation de la tension de vapeur d'une amine**
 - Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
 - Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
 - **Comparaison des résultats des 2 méthodes**
- **Planification expérimentale itérative pour les GM**
- **Conclusions**

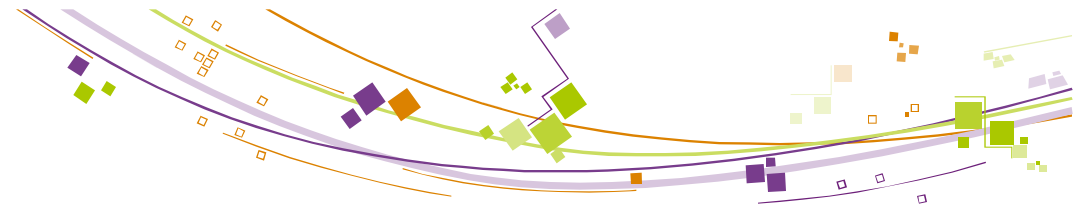


Résultats obtenus par QSAR-MD et GM

Prédiction QSAR-MD par leave-one-out

Prédiction Gaph Machines par leave-one-out





Plan

- Contexte
 - Captage du CO₂ par des solutions d'amines
 - Expérimentation haut débit (EHD)
 - Boucle EHD
- Modélisation de la tension de vapeur d'une amine
 - Méthode QSAR-MD : Quantitative Structure-Activity Relationship par Descripteurs Moléculaires
 - Méthode des Graph Machines : QSAR-GM
 - Codage d'une molécule en graphe
 - Construction du modèle GM par apprentissage statistique
 - Comparaison des résultats des 2 méthodes
- **Planification expérimentale itérative pour les GM**
- Conclusions

Planification expérimentale itérative pour les GM

n molécules

smiles y ← propriété à modéliser

θ^{*1} θ^{*2} θ^{*M}

1		$F_{\theta}^{(1)}$	$F_{\theta^{*1}}^{(1)}$	$F_{\theta^{*2}}^{(1)}$	$F_{\theta^{*M}}^{(1)}$
2	ϵ	$F_{\theta}^{(2)}$	$F_{\theta^{*1}}^{(2)}$	$F_{\theta^{*2}}^{(2)}$	$F_{\theta^{*M}}^{(2)}$
...						
i	ϵ	$F_{\theta}^{(i)}$	$F_{\theta^{*1}}^{(i)}$	$F_{\theta^{*2}}^{(i)}$	$F_{\theta^{*M}}^{(i)}$
...						
$n-1$	ϵ	$F_{\theta}^{(n-1)}$	$F_{\theta^{*1}}^{(n-1)}$	$F_{\theta^{*2}}^{(n-1)}$	$F_{\theta^{*M}}^{(n-1)}$
n		$F_{\theta}^{(n)}$	$F_{\theta^{*1}}^{(n)}$	$F_{\theta^{*2}}^{(n)}$	$F_{\theta^{*M}}^{(n)}$

Définition d'une architecture de RN

f_{θ}

$1 \leq i \leq n$

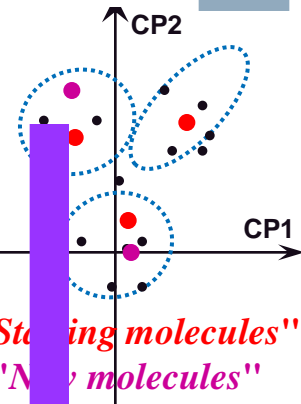
$F_{\hat{\theta}^1}$

$F_{\hat{\theta}^2}$

ACP

Classification

$\sigma_{F_{\hat{\theta}}(x)}$ max



Boucle EHD

"Staying molecules"
"New molecules"



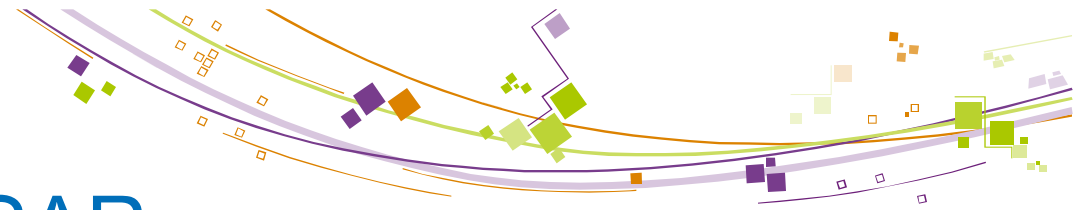
QSAR-MD et GM : 2 approches fondamentalement différentes mais complémentaires

■ QSAR-MD

- données en entrée du modèle = vecteurs des descripteurs moléculaires
- la forme du modèle est identique pour tous les individus
- représentation des individus dans l'espace des descripteurs 😊
- modèle explicite : interprétation des liens entrées/sortie(s) 😊
- méthode lourde : nécessite le calcul des descripteurs et la sélection des plus pertinents 😞
- méthode moins efficace : pouvoir de prédiction moins bon 😞

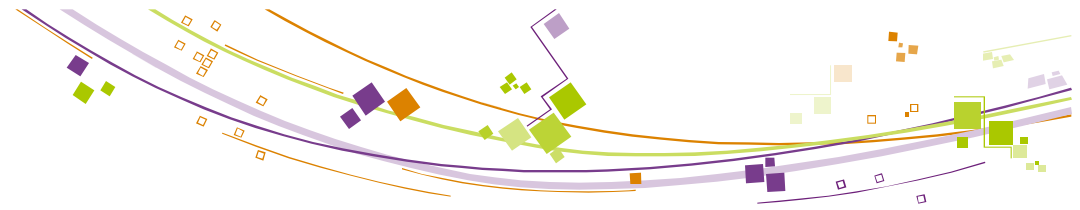
■ QSAR-GM

- données en entrée du modèle = graphes qui traduisent la structure des molécules
- le modèle = fonction particulière pour chaque individu
- pas d'espace de représentation des individus 😞 ⇒ **sinon planification expérimentale** 😊
- modèle *boîte noire* 😞 ⇒ **est-il possible d'expliciter les liens entrées/sortie(s) ?** 😊
- méthode rapide : pas besoin de calcul préalable des descripteurs 😊
- méthode efficace : meilleur pouvoir de prédiction 😊



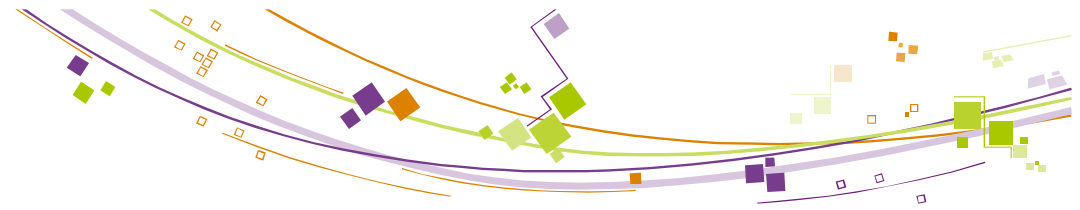
Articles EHD et QSAR

- PORCHERON F., JACQUIN M., SALDANA D., FARAJ A., GOULON A., Graph Machines based-QSAR approach for experimental design and modeling of amines thermodynamic properties: application to CO₂ capture, (en cours no spécial OGST Oil and Gas Science and Technologie consacré à l'Expérimentation Haut Débit, sous presse)
- S. Martin, H. Lepaumier, D. Picq, J. Kittel, T. de Bruin, A. Faraj, P-L. Carrette, New amines for CO₂ capture. IV. Degradation, Corrosion and Quantitative Structure Property Relationship Model, Industrial & Engineering Chemistry Research, Corrosion and QSPR Model, Industrial & Engineering Chemistry Research (2012)
- T. Duerinck, P. Leflaive, I.C. Arik, G. Pirngruber, V. Meynen, P. Cool, J.A. Martens, G.V. Baron, A. Faraj, J.F.M. Denayer, (2011), Experimental and statistical modeling study of low coverage gas adsorption of light alkanes on meso-microporous silica, Chemical Engineering Journal 179 (2012) 52– 62,
- A. Goulon, A. Faraj, G. Pirngruber, M. Jacquin, F. Porcheron, P. Leflaive, P. Martin , G.V. Baron, J.F.M. Denayer, (2011), Novel Graph Machine based QSAR approach for the prediction of the adsorption enthalpies of alkanes on zeolites, Catalysis today, vol.159 n.1 pp:74-83 (2011)
- Porcheron, F.; Gibert, A.; Jacquin, M.; Mougine, P. Faraj, A.; Goulon, A.; Bouillon, P.-A.; Delfort, B.; Le Pennec, D.; Raynal, L., (2010), High Throughput Screening of amine thermodynamic properties applied to post-combustion CO₂ capture process evaluation. *Energy Procedia* 2010, in press. Available at <http://www.ghgt.info/>
- P. Leflaive, G. Pirngruber, A. Faraj, P. Martin, G.V. Baron, J.F.M. Denayer, (2010), Statistical Analysis and Partial Least Square Regression as new tools for modelling and understanding the adsorption properties of zeolites, Microporous and Mesoporous Materials 132 (2010) 246–257



Merci pour votre attention

Des questions ?



Modélisation des propriétés physico-chimiques des molécules à partir de leur structure

A. Faraj – F. Porcheron – M. Jacquin
IFP Energies nouvelles