



Diversité Moléculaire:

Application au Criblage Virtuel et Corrélation avec des Propriétés Physico-chimiques

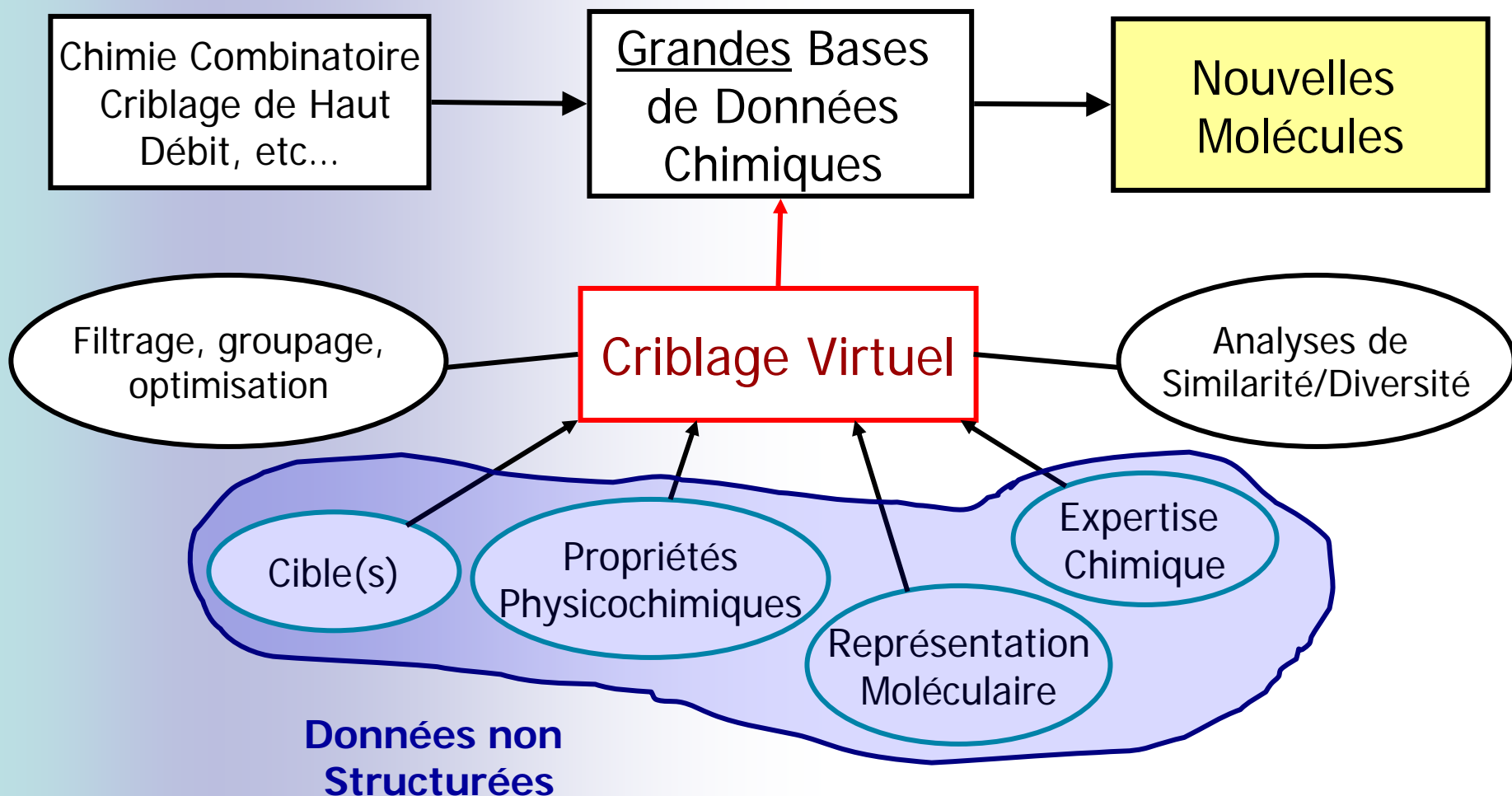
Présenté par: Ana Maldonado

Responsable: Prof. Bo-Tao Fan

Plan de la Présentation

- Problématique
- Similarité et Diversité Moléculaires
- Représentation Moléculaire
- Mesures de Similarité
- Système de Poids
- Présentation et analyse des résultats
- Conclusion et Perspectives

Criblage Virtuel et Nouvelles Molécules



Similarité et Diversité Moléculaires (I)

- Concept flou et relatif:



⇒ besoin de définitions quantitatives



- En chimie:

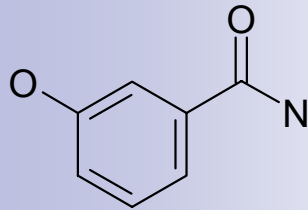
- *Méthode scientifique* d'Aristote
- *Table périodique* de Mendeleev
- *Principe de similarité des propriétés*



⇒ structures similaires → propriétés similaires

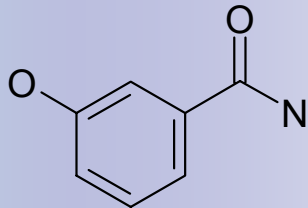
Similarité et Diversité Moléculaires (II)

- Trois composants principaux:
 - La représentation moléculaire → descripteurs

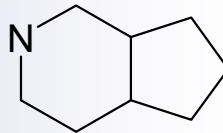


0, A, ☺, ¥, Bla bla ...

- Les mesures de similarité → indices



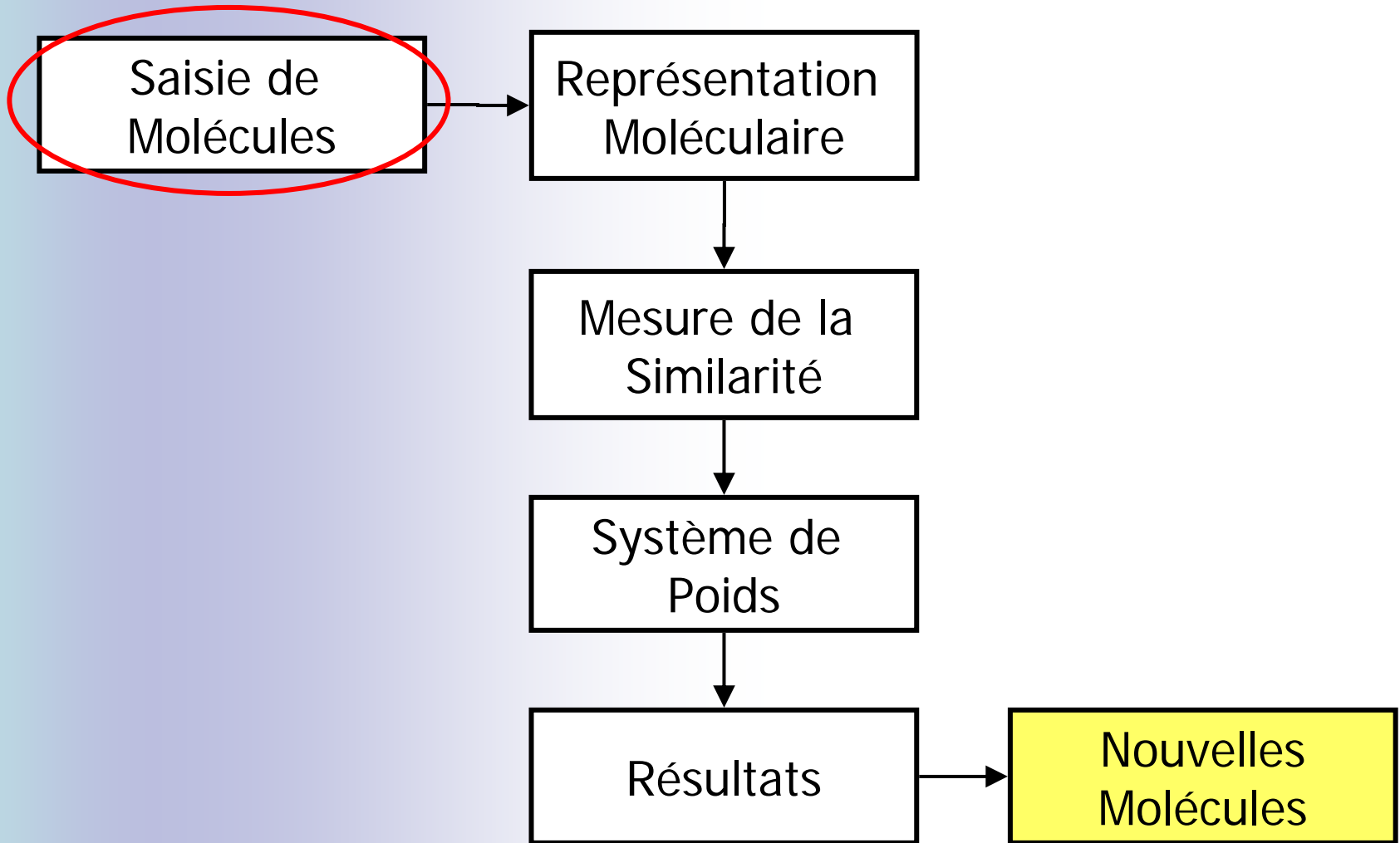
et



Se ressemblent? Oui? Non?
Beaucoup? Pas du tout?

- Système de poids

Aperçu de l'outil



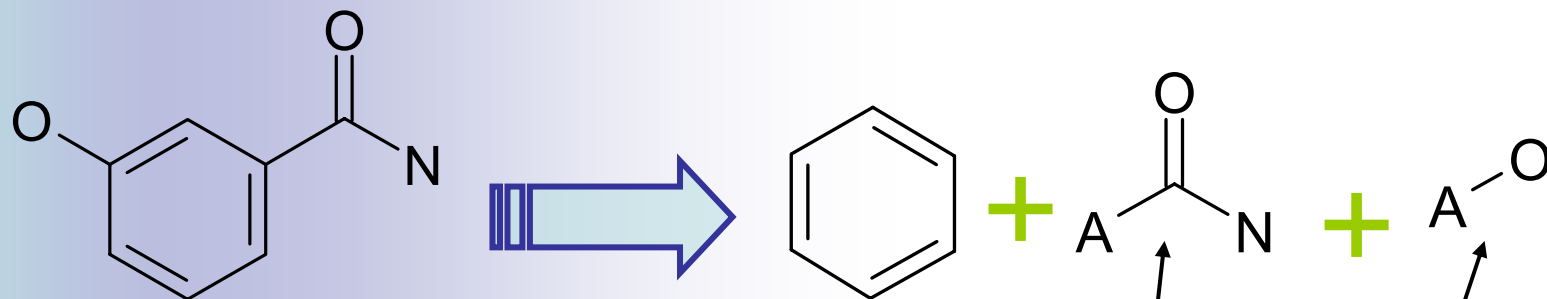
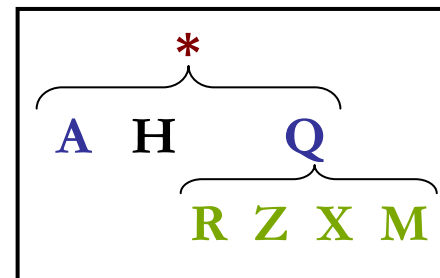
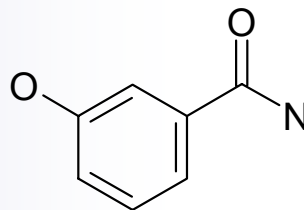
Descripteurs Structuraux

- Notre choix: approche de sous-structures

Molécule Test : 3-Hydroxy-benzamide

Masse moléculaire : 137,14

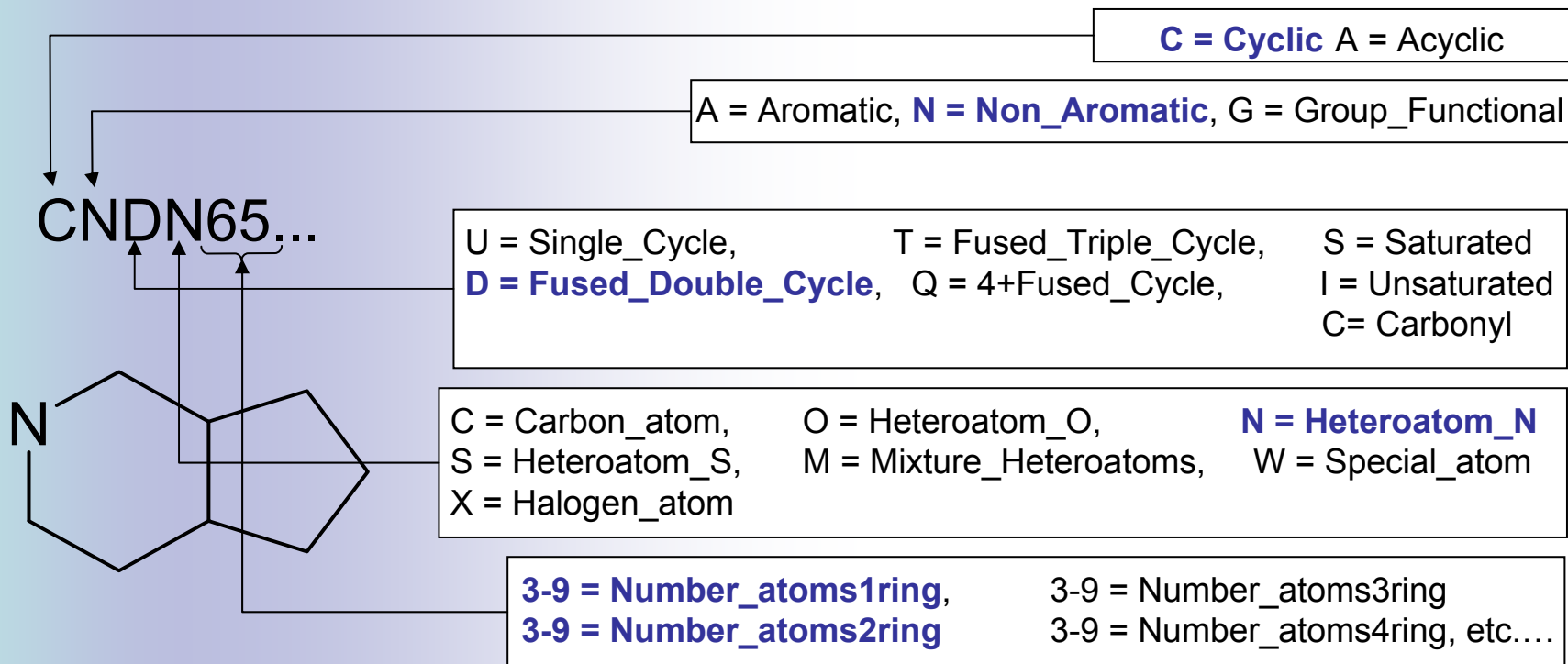
Formule : $C_7H_7NO_2$



ResultVector :

< 3-Hydroxy-benzamide; <<CAUC6-54, AGCM-29c, AGSO-24 >;3>>

« Nomenclature » des fragments

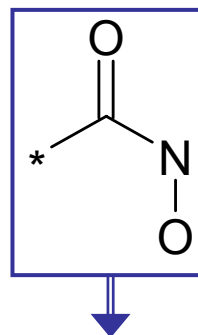


- Nom de fichier \Rightarrow combinaison de caractères et de chiffres pour décrire les propriétés des fragments

Structurer l'information

- Notre choix: fragments/sous-structures en XML

```
<Index>
<File name="AGCM-29c.mol">
  <Keys>
    <Key name="FID" value="29c"/>
    <Key name="FAtomSum" value="5"/>
    <Key name="FRing" value="0"/>
    <Key name="FGF" value="N-hydroxy-formamide"/>
  </Keys>
  <Properties>
    <Property name="HBondAD" value="1"/>
    <Property name="PotPCharged" value="0"/>
    <Property name="PotNCharged" value="1"/>
    <Property name="HydPhi" value="1"/>
    <Property name="Aromat" value="0"/>
    <Property name="Polar" value="1"/>
    <Property name="HydPho" value="0"/>
  </Properties>
</File>
....
</Index>
```

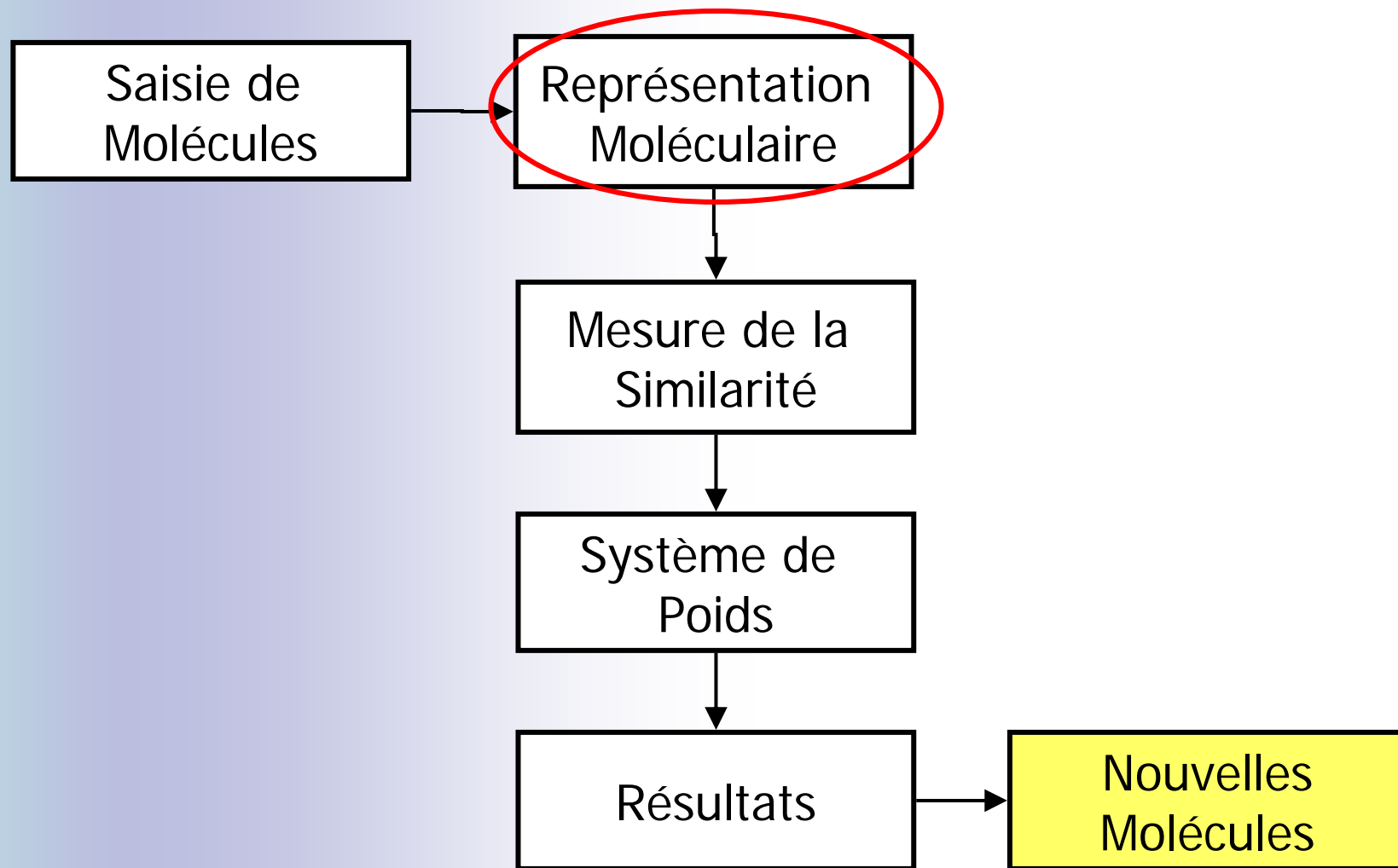


AGCM-29c.mol

-ISIS- 05200314222D

```
5 4 0 0 0 0 0 0 0 0999 V2000
-1.6083 1.0958 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
-1.6083 0.2667 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
-2.3208 1.5047 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
-2.3208 2.3366 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0
-3.0333 1.0958 0.0000 * 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
3 4 2 0 0 0 0
3 5 1 0 0 0 0
3 1 1 0 0 0 0
M END
```

Aperçu de l'outil



Mesures de Similarité (I)

- Indices de Similarité

- Multiples indices \Rightarrow Multiples mesures

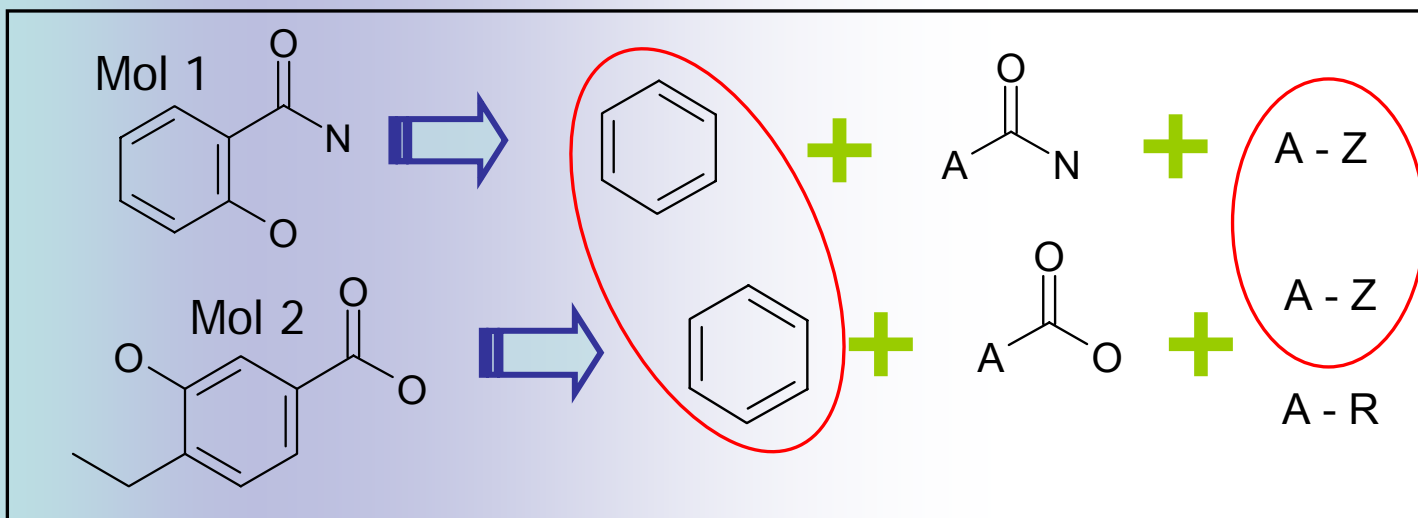
Où les variables sont:

a = fragments de la molécule 1

b = fragments de la molécule 2

c = fragments communs entre 1 et 2

Index	Équation
Tanimoto	$St = c / (a + b - c)$



a	b	c
3	4	2

$$Sim = c / (a + b - c)$$

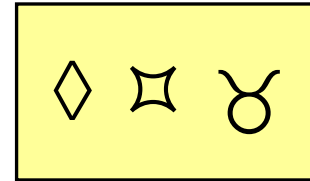
$$Sim = 2 / (3 + 4 - 2)$$

$$Sim = 0,4$$

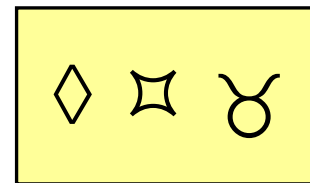
Mesures de Similarité (II)

- 3 niveaux de comparaison

- 1^{er} niveau: on utilise des informations structurelles

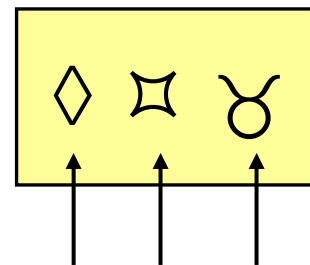


- 2nd niveau: on prend en compte les propriétés physico-chimiques, leur importance ainsi que des poids pour les fragments



+
HBondAD
PotPCharged
HydPhi
Aromat

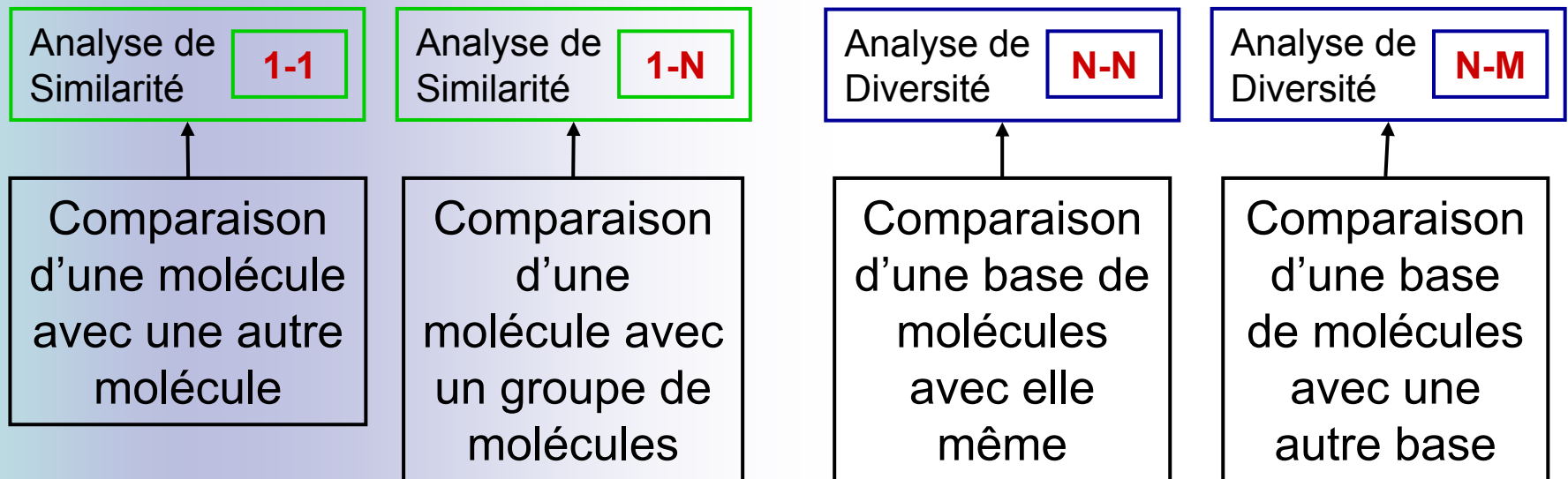
- 3^{ième} niveau: on utilise en plus l'information correspondant aux positions des fragments par rapport à la molécule cible



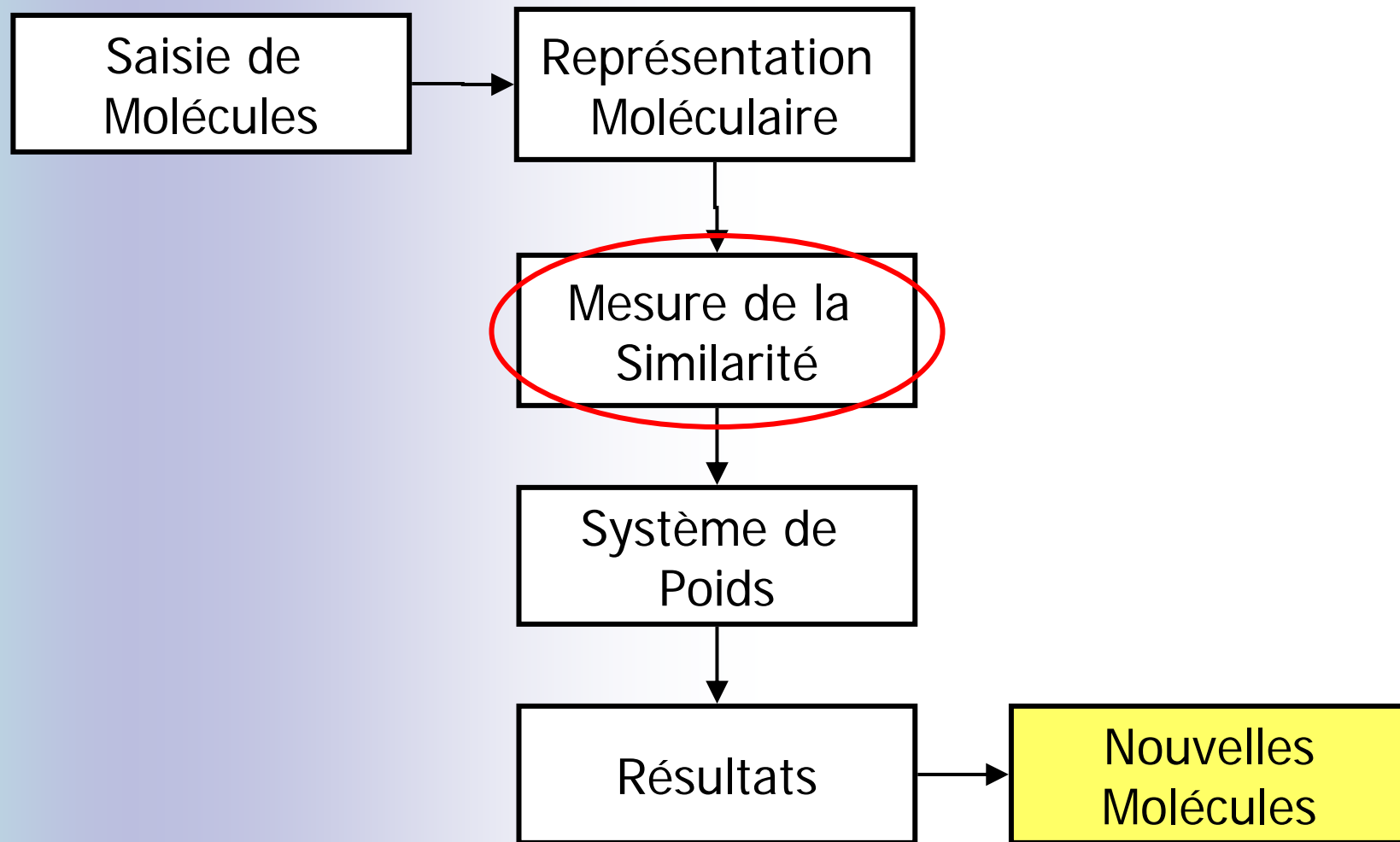
+
HBondAD
PotPCharged
HydPhi
Aromat

Mesures de Similarité (III)

- Différentes analyses de Similarité et de Diversité



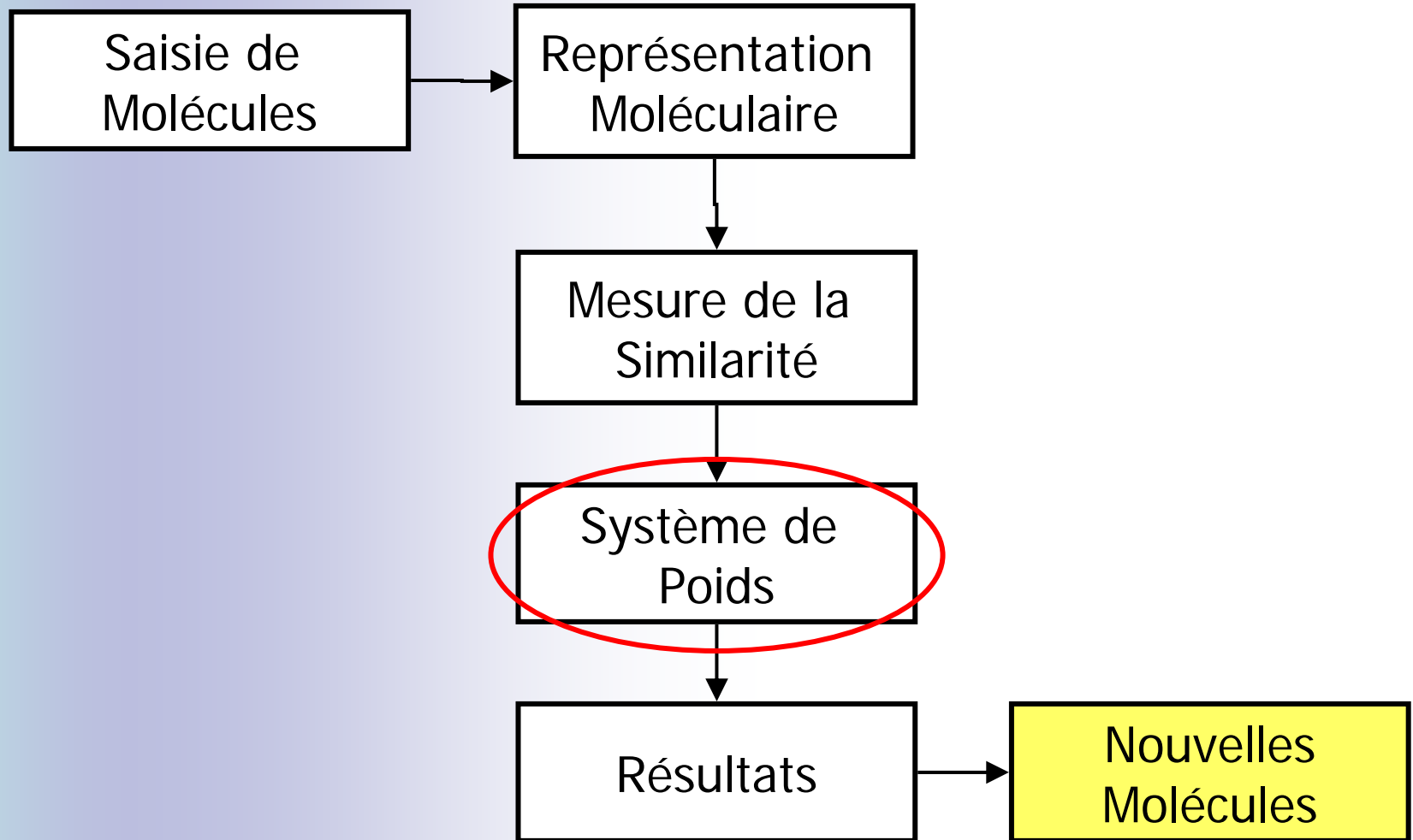
Aperçu de l'outil



Systeme de Poids

- Pourquoi ?
 - Parce que tous les chimistes n'ont pas les mêmes besoins
- Comment ?
 - En donnant différents niveaux d'importance à chaque tâche
- Et si on paramétrait ?
 - ⇒ Un outil → multiples usages

Aperçu de l'outil



Présentation et analyse des résultats

Version 1.2.3

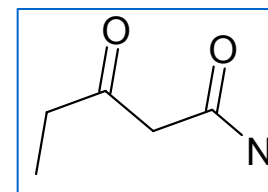
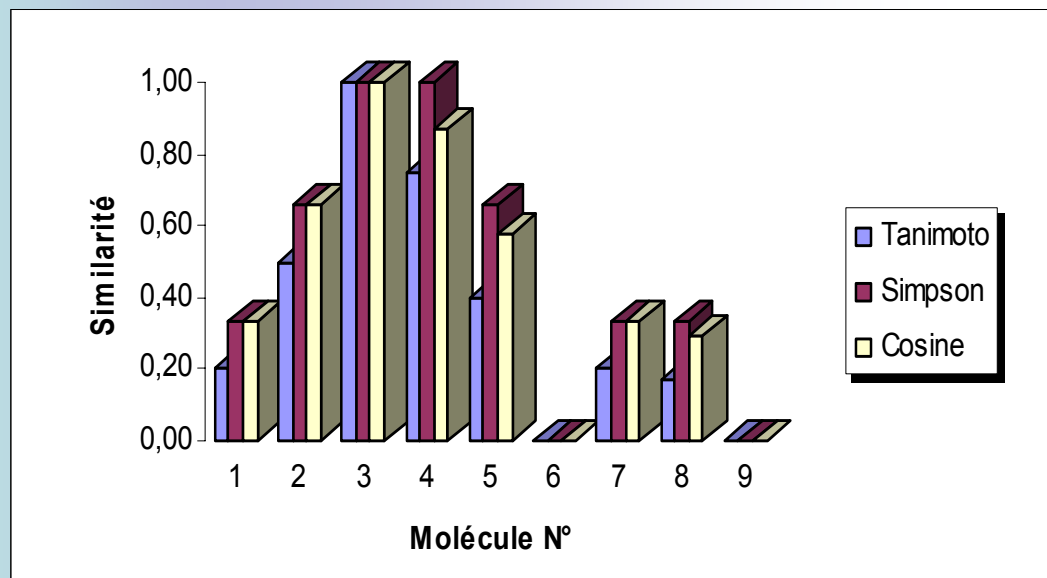
- Similarité entre une molécule cible et une base de molécules (analyse 1-N)

Molécule	Tanimoto	Simpson	Cosine
1	0.20	0.33	0.33
2	0.50	0.66	0.66
3	1.00	1.00	1.00
4	0.75	1.00	0.87
5	0.40	0.66	0.58
6	0.00	0.00	0.00
7	0.20	0.33	0.33
8	0.17	0.33	0.29
9	0.00	0.00	0.00

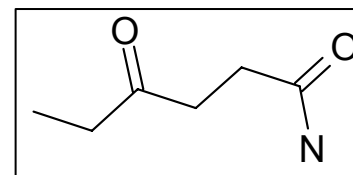
Présentation et analyse des résultats

Version 1.2.3

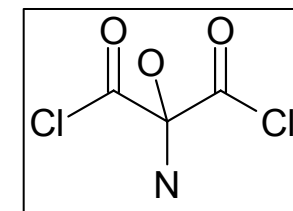
- Graphique de similarité



Molécule 3 (cible)



Molécule 4



Molécule 9

- Les applications de l'outil
 - Chimie combinatoire
 - Chimie médicinale
 - Et à priori toute base de données chimique...

Conclusions

- Outil de criblage virtuel
 - Nouvelle extension de la notion de diversité
 - * Informations structurelles
 - * Propriétés physicochimiques
- Nomenclature des fragments
- Hiérarchie des atomes génériques
- Utilisation des langages de marquage (XML)
 - Structuration, exploitation et échange de données chimiques complexes
- Différents niveaux de comparaison
- Différentes mesures de similarité et diversité
- Possibilité de paramétrer en mettant des poids

Merci de votre attention !

Questions ?

Laboratoire ITODYS, CNRS - UMR 7086
Université Paris 7 – Denis Diderot