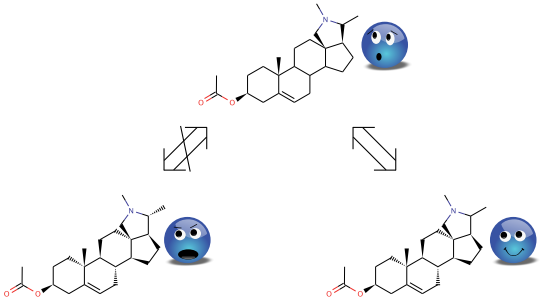
Recommandations pour la normalisation de structures de la Chimiothérapie Nationale

F. Ruggiu, G. Marcou, D. Horvath, A. Varnek
Laboratoire d'Infochimie, Université de Strasbourg
E. Grand
Laboratoire des Glucides, Université de Picardie
F. Massicot
Université de Reims-Champagne-Ardenne

Ce document résume des recommandations pour représenter les structures des substances référencées dans la Chimiothèque Nationale afin de diminuer le nombre d'ambiguïtés qui nuisent au traitement automatique des données et à l'interprétation des résultats de criblage.



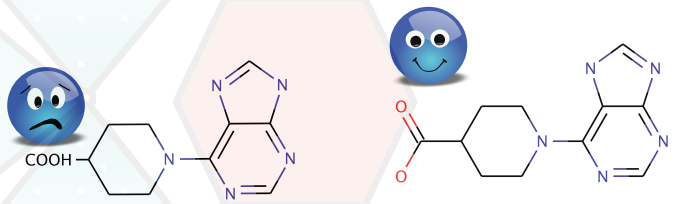
15 Juin 2012

Format et identifiant

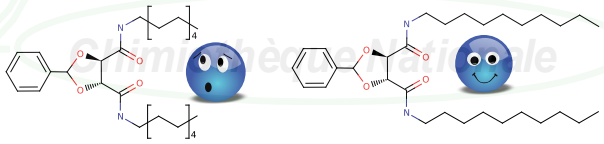
Vérifier avant de transmettre des structures que les identifiant sont conformes à la Chimiothérapie Nationale.

Utiliser le format SDF v2000. La quatrième ligne du fichier doit se terminer par V2000. Certains outils comme ChemDraw sauvegardent automatiquement en SDF v3000 si des éléments non supportés par le format v2000 sont introduits dans le dessin. Dans ce cas, il faut soit refaire une sauvegarde en éliminant ces composants (souvent des groupements chimiques, des labels, des couleurs, des formes géométriques), soit utiliser un outil tiers pour faire la conversion.

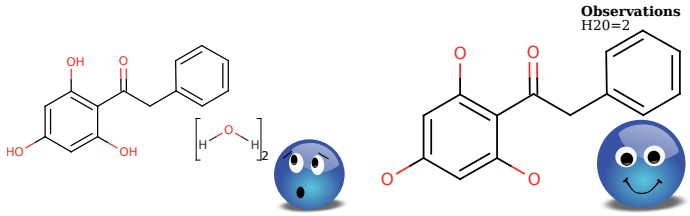
Groupes implicites/explicites



Ne pas utiliser un représentation implicite de groupes chimiques

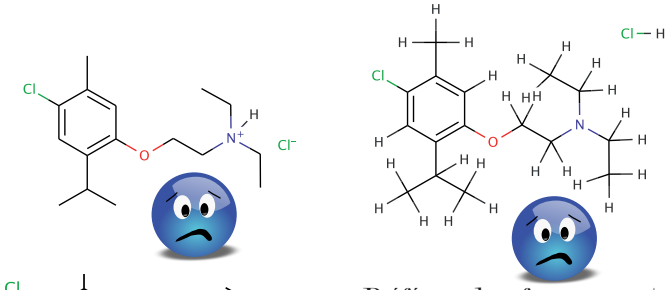


Ne pas utiliser de notation raccourcies.



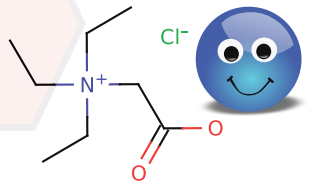
Ne pas représenter de molécule d'eau pour un hydrate et préciser le degré d'hydratation *d* dans le champs `Observations` en utilisant la syntaxe `H2O=`*d*.

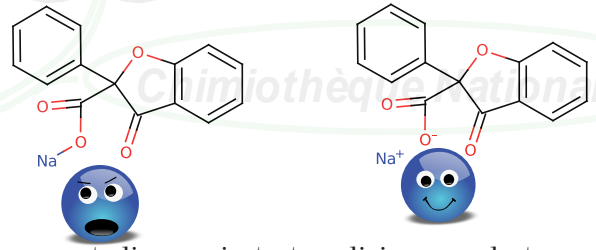
Hydrogènes et charges formelles



Préférer des formes neutres pour toutes les espèces d'une substance si la neutralisation est permise par échange de proton. Ne pas représenter les hydrogène explicitement. Les hydrogènes *implicites* se déduisent de l'état des valences et de la charge formelle.

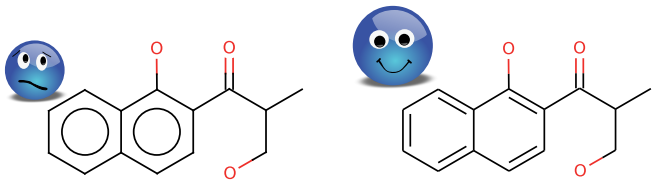
Les charges formelles sont figurées quand elles sont permanentes. Les contre-ions nécessaires ne sont pas neutralisés.





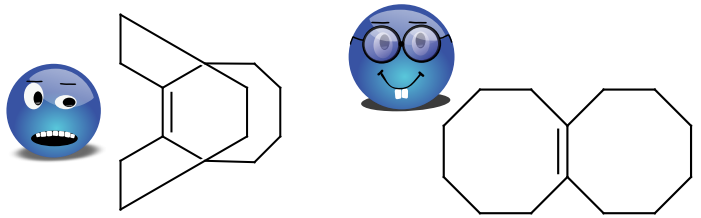
Ne pas neutraliser en ajoutant une liaison convalente.

Aromatisations



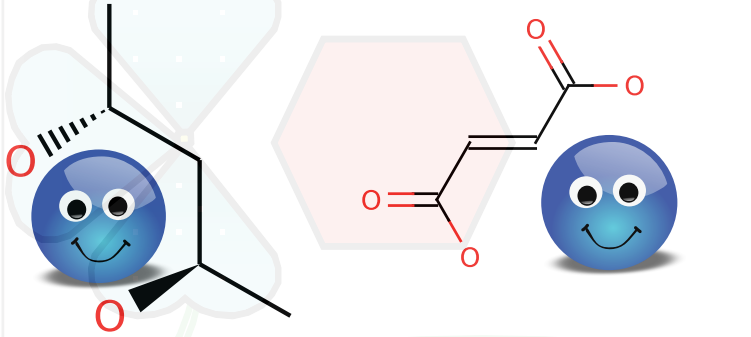
Les formes kekulisés des cycles sont toujours préférées.

Croisements

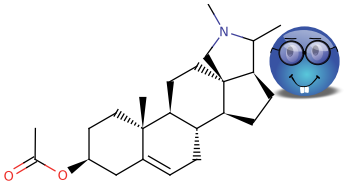


Ne pas croiser les liaisons sans nécessité..

Stéréochimie

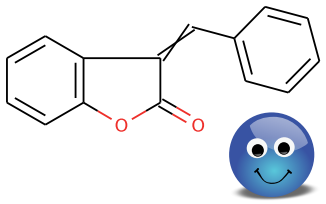


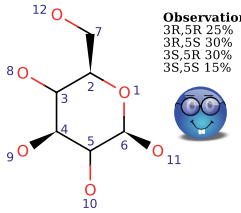
La configuration absolue des centres stéréogènes est spécifiée par l'orientation des liaisons qui en partent. Le « coin »d'une liaison stéréo doit être du côté de l'atome asymétrique.



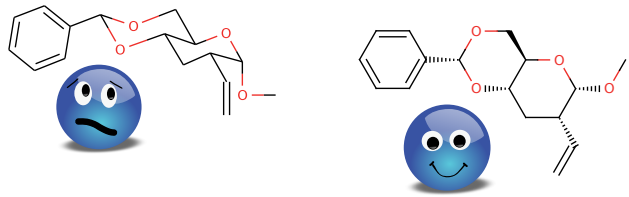
Si la configuration d'un ou plusieurs centres stéréogènes n'est pas spécifiée, cela indique que la substance est stéréochimiquement pure mais que le stéréoisomère isolé n'est pas connu.

Si la stéréoisomérie d'une double liaison est figurée par une double liaison Cis/Trans cela signifie qu'un stéréoisomère est isolé mais il n'est pas connu duquel il s'agit.

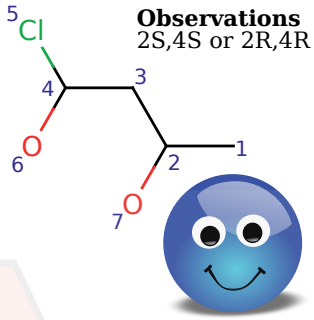


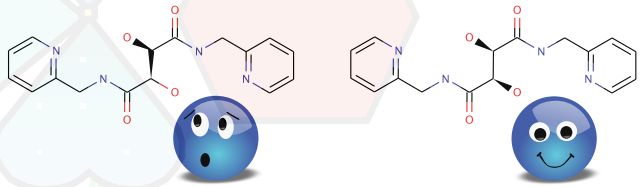


Si plusieurs stéréoisomères sont présent dans la solution: utiliser une numérotation des atomes pour spécifier les proportions des stéréoisomères dans le champs `Observations` de la base de données. Les racémiques sont un cas particulier de cette situation.

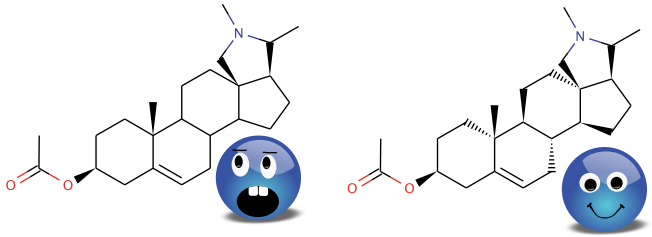


Si la substance contient le composé dessiné ou son énantiomère, le couple énantiomère est précisé dans le champs `Observations`.



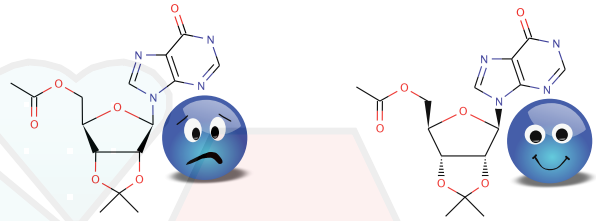


Deux atomes asymétriques ne doivent pas être reliés par une liaison stéréo.



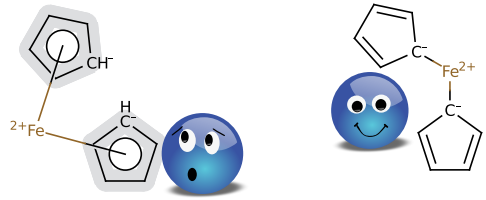
Spécifier la stéréochimie de tous les centres asymétriques explicitement, surtout si d'usage elle est implicite comme, par exemple, pour les chassis gonane, estane, androstane, pregane et leurs alcènes.

Il ne faut pas utiliser de représentations en pseudo-perspectives pour spécifier la position des groupes.



Ne pas utiliser de représentations spécifiques (Haworth, Fisher,...) pour spécifier la stéréochimie.

Organo-métalliques



Préférer des représentations utilisant des liaisons covalentes dans le respect des règles de Lewis. Seules les liaisons conventionelles (simple, double, triple et aromatique) sont supportées par tous les formats électroniques utilisés dans les bases de données.