

Open Cyclo. database

## RÉUTILISATION DES DONNÉES EN CHIMIE : EXEMPLE PRATIQUE AVEC LA BASE DE DONNÉES OPENCYCLODB.ORG

Prof. Adlane SAYEDE

Demi-journée d'étude UPEC — 07/11/2024

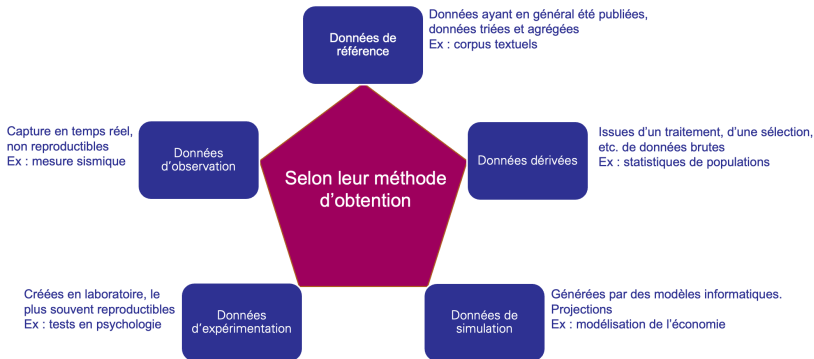
1. La donnée de la recherche
2. Détermination de constantes d'association entre des cyclodextrines et ses invités
  - Problématique
  - Création de l'ensemble de données
  - Enrichissement des données
  - Utilisation des données
3. OpenCycloDB.org
  - Le portail
4. Réutilisation de données

- La donnée de la recherche est un élément clé dans le processus de production de connaissances scientifiques. Elle peut être définie comme toute information collectée, créée ou utilisée pour soutenir une recherche spécifique.
- Les données de recherche peuvent prendre différentes formes, telles que des enregistrements audio, des images, des vidéos, des bases de données, des questionnaires, des notes de terrain, etc.

- La gestion efficace des données de recherche est essentielle pour garantir l'intégrité et la qualité des résultats de recherche. Cela implique de conserver les données de manière organisée, sécurisée et accessible à long terme.
- Une bonne gestion des données de recherche peut également faciliter la collaboration entre chercheurs, permettre la réutilisation des données pour d'autres recherches, et assurer la conformité aux exigences des organismes de financement et des revues scientifiques.

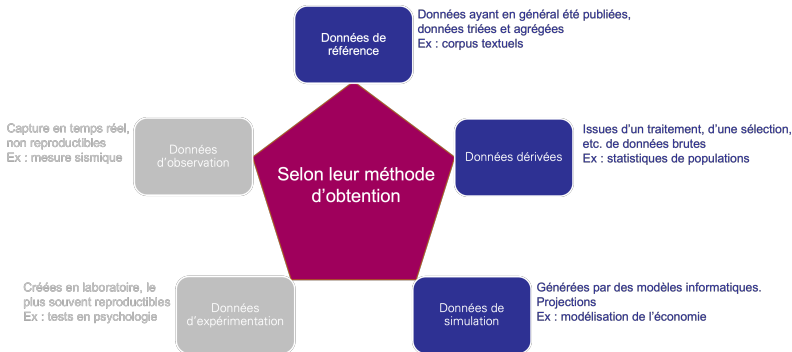
- La gestion des données de recherche peut être complexe en raison de la diversité des types de données et des méthodes de collecte, ainsi que des exigences de confidentialité et de sécurité.
- De plus, il peut y avoir des défis liés au partage et à la réutilisation des données, tels que la propriété intellectuelle, la citation appropriée et la protection de la vie privée.

## TYPOLOGIE DES DONNÉES DE LA RECHERCHE



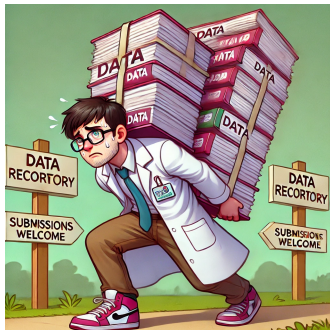
La typologie des données de recherche selon leur méthode d'obtention peut être classifiée en différentes catégories, chacune correspondant à un type de collecte ou de génération des données.

## TYPOLOGIE DES DONNÉES DE LA RECHERCHE



La typologie des données de recherche selon leur méthode d'obtention peut être classifiée en différentes catégories, chacune correspondant à un type de collecte ou de génération des données.

## DIFFUSION DES DONNÉES DE LA RECHERCHE



La diffusion des données de la recherche est une étape clé pour maximiser l'impact et la visibilité des résultats de la recherche. Les chercheurs peuvent diffuser leurs données de recherche en les **publiant dans des revues scientifiques (data paper)** et/ou en les déposant dans des dépôts de données en libre accès.





Faciliter la découverte des données par les humains et les systèmes informatiques. Pour cela, les (méta)données doivent disposer d'un identifiant unique et pérenne (DOI), et être enregistrées ou indexées dans un dispositif permettant leur recherche.



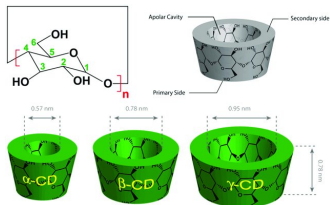
Stocker durablement les données et les métadonnées et faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint) et d'utilisation (licence).



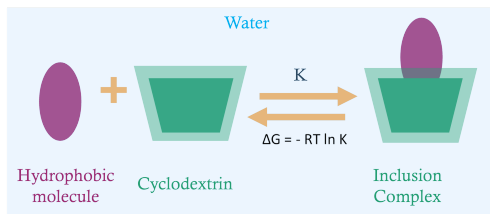
Les données et leurs métadonnées doivent être formatées et structurées (CSV, JSON, XML,...) de manière à pouvoir être lues, interprétées et utilisées facilement par différents systèmes et applications, sans nécessiter de conversion complexe.



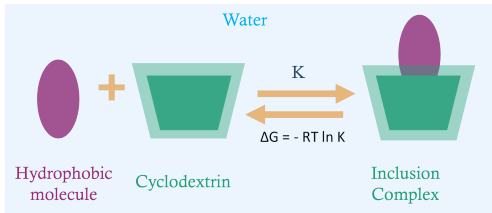
La réutilisation (libre, conditionnelle ou payante) doit être facilitée par l'utilisation de standards communs, grâce à des bases de données rassemblant des données claires, vérifiées et bien décrites, directement (ré)utilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction/transparence de la science).



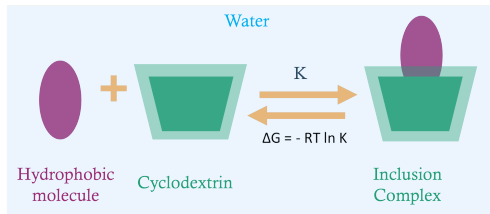
Les cyclodextrines (CD) sont des molécules en forme d'anneau constituées de glucoses liés de manière cyclique. Elles appartiennent à une classe de composés appelés oligosaccharides et **sont principalement constituées de 6, 7 ou 8 unités de glucose.**



Leur structure présente une cavité interne hydrophobe (apolaire) et une surface externe hydrophile (polaire). Cette configuration permet aux cyclodextrines de capturer d'autres molécules dans leur cavité, un phénomène appelé "complexation".



Cette capacité de complexation est utilisée pour améliorer la solubilité de composés organiques insolubles, en particulier dans les domaines pharmaceutique, alimentaire et cosmétique. Par exemple, les cyclodextrines sont utilisées pour encapsuler des médicaments afin d'en améliorer la biodisponibilité.



L'énergie libre de complexation est essentielle pour évaluer l'interaction hôte-invité et déterminer par exemple l'adéquation d'un complexe médicament-CDs pour la formulation.

- Sa détermination expérimentale est souvent chronophage.
- Les simulations moléculaires offrent une alternative, mais sont limitées par la capacité de calcul.
- L'utilisation de l'IA offre des perspectives prometteuses, mais son efficacité repose sur la **disponibilité de données fiables et bien structurées**.



Les données proviennent de différents articles/sources publiés entre 1963 et 2021. Elles sont sous forme de **tableau** :

1880 Chemical Reviews, 1998, Vol. 98, No. 5

Rekharsky and Inoue

Table 1. Complex Stability Constant (log *K*), Standard Free Energy ( $\Delta G^\circ$ ), Enthalpy ( $\Delta H^\circ$ ), and Entropy Changes ( $T\Delta S^\circ$ ) for 1:1 Inclusion Complexation of Various Guests with Natural  $\alpha$ -,  $\beta$ -, and  $\gamma$ -Cyclodextrins (1 $\alpha$ , 1 $\beta$ , and 1 $\gamma$ )

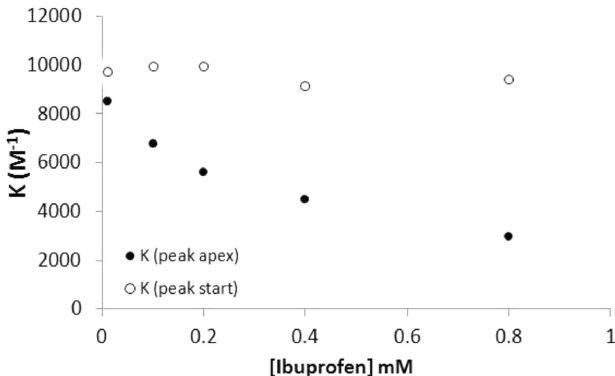
host	guest	solvent	<i>T</i> /K	log <i>K</i>	$\Delta G^\circ$ / kJ mol <sup>-1</sup>	$\Delta H^\circ$ / kJ mol <sup>-1</sup>	$T\Delta S^\circ$ / kJ mol <sup>-1</sup>	method <sup>a</sup>	ref
1 $\alpha$	acetic acid	H <sub>2</sub> O	298	0.96 ± 0.01	-5.5 ± 0.1	-11.2 ± 0.1	-5.6 ± 1.3	pot	172
1 $\alpha$	acetic acid	H <sub>2</sub> O	298	3.8 ± 1.2	-21.8 ± 6.7	-5.0 ± 0.4	16.2	cal	192 <sup>b</sup>
1 $\alpha$	acetonitrile	H <sub>2</sub> O (pH 3.8–4.5)	298	0.75 ± 0.01	-4.3 ± 0.1	-10 ± 1	-6 ± 1	pot	235 <sup>c</sup>
1 $\alpha$	acetonitrile	H <sub>2</sub> O (0.05 M HCl)	298	0.72	-4.1	-14	-10	uv	236 <sup>d</sup>
1 $\alpha$	<i>N</i> -acetyl-L-leucinamide	H <sub>2</sub> O	298	1.30 ± 0.08	-7.4 ± 0.5	-6.1 ± 0.8	1.3 ± 1.3	cal	201
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O (pH 7.2)	298		-14.2	-14.6	-0.4	cal	21
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O-CH <sub>3</sub> CN (98.5:1.5; pH 9.0)	298	3.09 ± 0.03	-17.6	-5.0 ± 1.7	12.5 ± 2.5	kin	125
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O	298	2.15 ± 0.04	-12.2 ± 0.2	-14 ± 3	-2 ± 2	con	133
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O (pH 8.5)	298	2.15	-12.2	-13.4 ± 0.4	-1.2	cal	81
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O	298		-11.6	-14.3	-2.7	cal	188
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O (pH 8.50)	298		-12.3 ± 0.2	-13.5 ± 0.3	-0.4 ± 0.4	cal	237
1 $\alpha$	1-adamantanecarboxylate	H <sub>2</sub> O	298	2.16 ± 0.05	-12.3 ± 0.3	-23 ± 3	-11 ± 2	pot	133
1 $\alpha$	1-adamantylammonium	H <sub>2</sub> O	298	2.43 ± 0.01	-13.9 ± 0.1	-20.3 ± 1.4	-6.3 ± 1.4	pot	238
1 $\alpha$	1-adamantylammonium	H <sub>2</sub> O (pH < 2.5)	298	1.697 ± 0.007	-9.68 ± 0.04	-15.3 ± 0.9	-5.6 ± 0.9	uv	238
1 $\alpha$	adipate (dianion)	H <sub>2</sub> O (pH 9.5)	298	2.13 ± 0.01	-12.16 ± 0.04	-15.16 ± 0.04	-3.00 ± 0.06	cal	239
1 $\alpha$	hydrogen adipate (monoanion)	H <sub>2</sub> O	298	2.18 ± 0.01	-12.4 ± 0.1	-31.4 ± 1.3	-19.0 ± 1.4	pot	72
1 $\alpha$	adiphenine-HCl	H <sub>2</sub> O	298	1.06 ± 0.03	-6.1	-25.5 ± 1.6	-19.5	cal	74
1 $\alpha$	2-aminobenzoic acid	H <sub>2</sub> O	298	5.0 ± 1.3	-28.5 ± 7.5	-11.3 ± 0.4	26.2	cal	192
1 $\alpha$	3-aminobenzoic acid	H <sub>2</sub> O	298		-9.9	-32.5 ± 0.3	-22.6 ± 0.8	cd	194
1 $\alpha$	4-aminobenzoic acid	H <sub>2</sub> O	298		-16.0	-43.6 ± 0.8	-27.6 ± 2.5	cd	194
1 $\alpha$	4-aminobenzoic acid	H <sub>2</sub> O	298	2.8 ± 0.1	-15.9 ± 0.4	-49 ± 2	-33	cal	192
1 $\alpha$	6-aminohexanoic acid	H <sub>2</sub> O (pH 7.0)	298	1.34	-7.6 ± 0.7	-1.5 ± 0.3	6.0 ± 1.0	cal	35
1 $\alpha$	3-(aminomethyl)propryl	H <sub>2</sub> O	298	0.69 ± 0.10	-3.9 ± 0.6	-33 ± 8	-29 ± 8	esr	114
1 $\alpha$	2-aminooctanoic acid	H <sub>2</sub> O (pH 7.8)	298	2.80 ± 0.02	-16.0 ± 0.2	-17.7 ± 0.3	-1.7 ± 0.5	cal	35
1 $\alpha$	8-aminooctanoic acid	H <sub>2</sub> O (pH 7.1)	298	1.88 ± 0.02	-10.7 ± 0.1	-14.8 ± 0.4	-4.1 ± 0.5	cal	35
1 $\alpha$	4-aminophenol	H <sub>2</sub> O (pH 4.0)	298	1.24 ± 0.02	-7.1 ± 0.3	-20.1 ± 0.5	-13.1 ± 0.6	uv	191
1 $\alpha$	11-aminoundecanoic acid	H <sub>2</sub> O (pH 6.7)	298	3.34 ± 0.04	-19.0 ± 0.2	-26.6 ± 0.5	-7.6 ± 0.7	cal	35
1 $\alpha$	<i>d</i> -amphetamine	H <sub>2</sub> O (pH 11.0)	298	1.38	-7.9	-12.6 ± 2.1	-4.7	uv/cal	202

Les données proviennent de différents articles/sources publiés entre 1963 et 2021. Elles sont sous forme de **texte** :

0.999) is conclusive evidence for the 1:1 plus 1:2 complexation. The estimates of association constants  $K_1$  and  $K_2$  were found to be  $844 \pm 12$  and  $52 \pm 4 M^{-1}$ , respectively.

In conclusion, the appropriate Scatchard equation for the formation of 1:1 and 1:2 substrate to cyclodextrin complexes was used to estimate the association constants,  $K_1$  and  $K_2$ . This method requires only valid data for the concentration of free or bound substrate species. The technique of ion-selective electrodes is particularly suitable for this type of analysis since it provides measurement of the free ion in the presence of bound ion.

Les données proviennent de différents articles/sources publiés entre 1963 et 2021. Elles sont sous forme de [figure](#) :



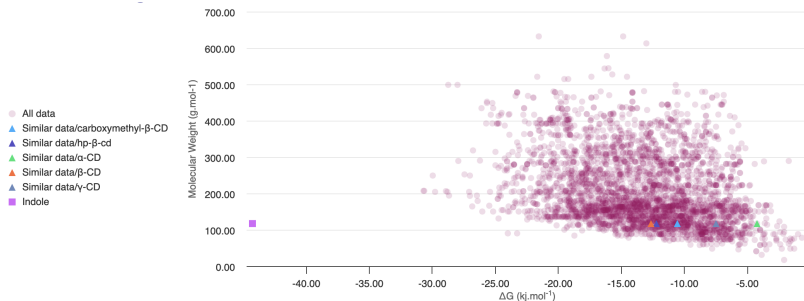
- Pour les tableaux, nous avons eu recours à la [bibliothèque Camelot](#).
- Pour les figures, nous avons employé des techniques de reconnaissance d'image telles que l'[OCR \(Reconnaissance Optique de Caractères\)](#) avec [Tesseract](#), ainsi que des méthodes de [segmentation d'image](#) pour identifier et extraire les légendes et les annotations.

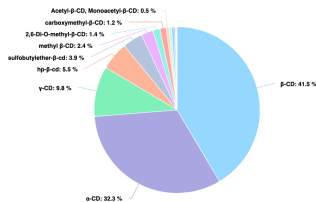
	Host	Guest	pH	T	DeltaG	Error
0	alpha-cyclodextrin	acetic acid	7.0	298.0	-21.8	6.7
1	alpha-cyclodextrin	acetonitrile	3.8	298.0	-4.3	0.1
2	alpha-cyclodextrin	N-acetyl-l-leucinamide	7.0	298.0	-7.4	0.5
3	alpha-cyclodextrin	1-adamantanecarboxylate	7.2	298.0	-14.2	0.0
4	alpha-cyclodextrin	1-adamantanecarboxylate	7.0	298.0	-12.2	0.2

Les données brutes ont ensuite été enrichies avec des descripteurs physico-chimiques pour permettre une analyse plus poussée et une utilisation dans des modèles de machine learning (ML).

	Molecular Formula	Molecular Weight	Canonical SMILES	Isomeric SMILES	IUPAC Name	XLogP	TPSA	Complexity	Charge
0	C2H4O2	60.05	CC(=O)O	CC(=O)O	acetic acid	-0.2	37.3	31	0
1	C2H3N	41.05	CC#N	CC#N	acetonitrile	0.0	23.8	29	0
2	C8H16N2O2	172.22	CC(C)CC(C(=O)NC(=O)C)N	CC(C)C[C@@H](C(=O)(2S)-N-acetyl-2-amino-4-methylpentanamide)NC(=O)C)N		0.1	72.2	178	0
3	C11H15O2-	179.24	C1C2CC3CC1CC(C2)(C3)C(=O)[O-]	C1C2CC3CC1CC(C2)(C3)C(=O)[O-]	adamantane-1-carboxylate	3.3	40.1	210	-1
4	C11H15O2-	179.24	C1C2CC3CC1CC(C2)(C3)C(=O)[O-]	C1C2CC3CC1CC(C2)(C3)C(=O)[O-]	adamantane-1-carboxylate	3.3	40.1	210	-1

Les valeurs extraites présentant des écarts significatifs par rapport aux autres données ont été exclues de l'ensemble de données. Par exemple, pour la molécule indole molécule,  $\Delta G = -44,3 \pm 0,3$  kJ/mol (J. Chem. Soc. Perkin Trans., 2 (1973), pp. 2081-2085), ce qui est considérée comme valeur aberrante par rapport aux molécules présentant des caractéristiques physico-chimiques similaires.





Un total de 3459 données a été collecté pour divers types de cyclodextrines. L'ensemble de données, comprenant les données brutes et enrichies, a été publié<sup>1</sup> sous la licence "Open Database License v1.0". Les fichiers de scripts associés ont été diffusés<sup>2</sup> sous la licence "BSD 3-Clause New License".

1. Curated Dataset of Association Constants Between a Cyclodextrin and a Guest for Machine Learning. DOI: [10.5281/zenodo.7575539](https://doi.org/10.5281/zenodo.7575539)

2. Curated Dataset of Association Constants Between a Cyclodextrin and a Guest for Machine Learning: Raw Data and Generation Script. DOI: [10.5281/zenodo.7575579](https://doi.org/10.5281/zenodo.7575579)

Chemical Data Collections 45 (2023) 101022



Contents lists available at [ScienceDirect](#)

Chemical Data Collections

journal homepage: [www.elsevier.com/locate/cdc](http://www.elsevier.com/locate/cdc)



Data Article

## Curated dataset of association constants between a cyclodextrin and a guest for machine learning

Gökhan Tahlı<sup>a,b,\*</sup>, Fabien Delorme<sup>a</sup>, Daniel Le Berre<sup>a</sup>, Éric Monflier<sup>b</sup>,  
Adlane Sayede<sup>b</sup>, Sébastien Tilloy<sup>b</sup>

<sup>a</sup> University of Artois, CNRS, Centre de Recherche en Informatique de Lens (CRIL), F-62300 Lens, France.

<sup>b</sup> University of Artois, CNRS, Centrale Lille, University of Lille, UMR 8181, Unité de Catalyse et Chimie du Solide (UCCS), rue Jean Souvraz, SP 18, F-62307 Lens Cedex, France



### ARTICLE INFO

**Keywords:**  
Machine learning  
Association constant  
Cyclodextrin

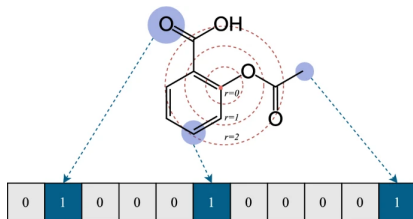
### ABSTRACT

Determining the association constant between a cyclodextrin and a guest molecule is an important task for various applications in various industrial and academic fields. However, such a task is time consuming, tedious and requires samples of both molecules. A significant number of association constants and relevant data is available from the literature. The availability of data makes the use of machine learning techniques to predict association constants possible. However, such data is mainly available from tables in articles or appendices. It is necessary to make them available in a computer friendly format and to curate them. Furthermore, the raw data need to be enriched with physicochemical information about each molecule and when such information does not allow to discriminate molecules, some additional data is needed. We present a dataset built from data gathered from the literature. The dataset contains both the original raw data from the articles and the enriched ones. We also provide the scripts used to curate and enrich the raw data.

La procédure détaillée de collecte et de traitement des données a également été publiée dans un "data paper".






- En ML, les formules brutes et noms chimiques sont insuffisants. Ils ne décrivent pas la structure complète ni les liaisons atomiques.
- La notation SMILES rend les molécules exploitables pour les algorithmes de ML.
- Les empreintes moléculaires (fingerprints) transforment cette notation linéaire en un vecteur binaire ou en une série de bits (0 et 1). Exemple de l'acide acétylsalicylique (aspirine) :



CC(=O)OC1=CC=CC=C1C(=O)O

- La notation SMILES standard ne capture pas ces différences spatiales. Cette limite complique la précision des modèles en ML nécessitant la stéréochimie.
- Les isomères partagent la même formule brute, mais diffèrent par leur arrangement spatial. Les stéréoisomères, comme le butan-2-ol (2R et 2S), ont des configurations distinctes.

Molécule	Isomeric SMILES	Representation	TPSA (Å <sup>2</sup> )	MW (g/mol)	Complexité $\gamma$
(2S)-butan-2-ol	<chem>CC[C@@H](C)O</chem>		20.23	74.12	17.61
(2R)-butan-2-ol	<chem>CC[C@H](C)O</chem>		20.23	74.12	17.61
butan-2-ol	<chem>CCC(C)O</chem>		20.23	74.12	17.61

Représentations du butan-2-ol et ses stéréoisomères.

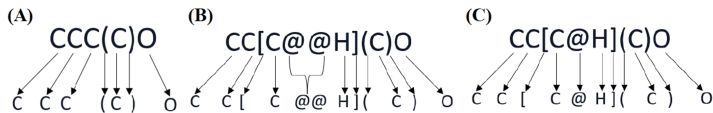
De manière analogue au traitement du langage naturel (NLP), les molécules sont considérées comme des phrases et leurs sous-structures comme des mots :

- Les chaînes SMILES sont divisées en fragments plus petits, chaque fragment étant composé de plusieurs caractères, ce qui permet de capturer les sous-structures moléculaires.



Création de fragments pour le (2R)-butan-2-ol

- En entraînant un modèle de réseau de neurones, on apprend à encoder des représentations significatives et informatives pour différentes structures moléculaires.



Molécule	Isomeric SMILES	2D Depiction	IsoSym0	IsoSym1	IsoSym2	...	IsoSym299
(2S)-butan-2-ol	<chem>CC[C@H](C)O</chem>		-0.05	-0.94	0.17	...	0.09
(2R)-butan-2-ol	<chem>CC[C@@H](C)O</chem>		-0.07	-0.97	0.11	...	0.08
butan-2-ol	<chem>CCC(C)O</chem>		0.05	-0.69	0.41	...	0.23

## Stereoisomers Are Not Machine Learning's Best Friends

Gökhan Tahl,\* Fabien Delorme, Daniel Le Berre, Éric Monflier, Adlane Sayede, and Sébastien Tilloy\*

 Cite This: *J. Chem. Inf. Model.* 2024, 64, 5451–5469 Read Online

ACCESS |

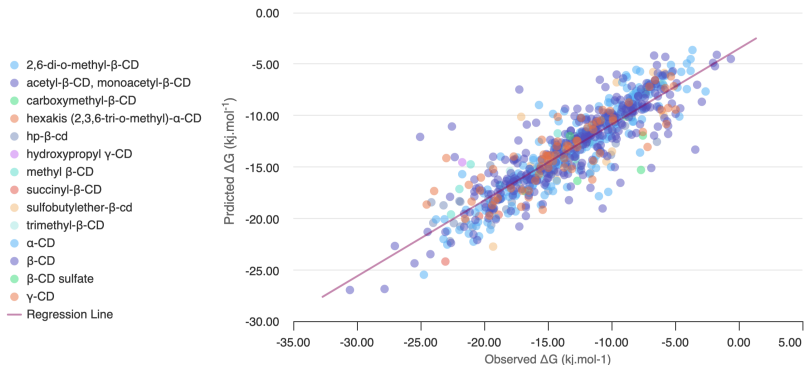
 Metrics & More Article Recommendations Supporting Information

**ABSTRACT:** This study addresses the challenge of accurately identifying stereoisomers in cheminformatics, which originates from our objective to apply machine learning to predict the association constant between cyclodextrin and a guest. Identifying stereoisomers is indeed crucial for machine learning applications. Current tools offer various molecular descriptors, including their textual representation as Isomeric SMILES that can distinguish stereoisomers. However, such representation is text-based and does not have a fixed size, so a conversion is needed to make it usable to machine learning approaches. Word embedding techniques can be used to solve this problem. Mol2vec, a word embedding approach for molecules, offers such a conversion. Unfortunately, it cannot distinguish between stereoisomers due to its inability to capture the spatial configuration of molecular structures. This study proposes several approaches that use word embedding techniques to handle molecular discrimination using stereochemical information on molecules or considering Isomeric SMILES notation as a text in Natural Language Processing. Our aim is to generate a distinct vector for each unique molecule, correctly identifying stereoisomer information in cheminformatics. The proposed approaches are then compared to our original machine learning task: predicting the association constant between cyclodextrin and a guest molecule.



La procédure détaillée d'enrichissement des données a également été publiée.

L'ensemble de données a été utilisé pour développer un modèle ML capable de prédire l'énergie libre de complexation ( $\Delta G$ ) entre des molécules organiques invitées et des cyclodextrines.



Energies de complexation prédites vs. observées.

- Offrir une base de données centralisée et accessible pour les constantes d'associations des cyclodextrines.

[Search](#) [Data](#) [Predictor](#) [Upload](#) [Citing](#) [About](#) [Validator](#)

## Welcome to the Open Cyclodextrin databases



## Open Cyclo. database

Total entries

**3459**

$\alpha$ -Cyclodextrin

**1128**

$\beta$ -Cyclodextrin

**1666**

$\gamma$ -Cyclodextrin

**343**

- OpenCycloDB permet de retrouver rapidement des données expérimentales sur les cyclodextrines.
- Données classées par type de cyclodextrine, type de molécule invitée, propriétés thermodynamiques, etc.

Quickly find information from OCdb sources either by :

Text Search Structure Search

Chemical Name, Smiles or InChI

Examples of Chemical Identifiers >

Chemical Name	SMILES	(Standard) InChI	(Standard) InChIKey
furan   formaldehyde	<chem>c1ccoc1   CC(C)=O</chem>	InChI=1S/C4H4O/c1-2-4-5-3-1/h1-4H	InChIKey=YLQBMQCUJZJEEH-UHFFFAOYSA-N

Number of entries for 2-[4-(2-methylpropyl)phenyl]propanoic acid (👁) is 1

Show  entries Search:

Host	Guest	$\Delta G$	Error	K	T	pH	Method	DOI
$\beta$ -CD	ibuprofen	-28.90	1.30	116383.0	None	2.0		Menard, F. A.; Dedhiya, M. G.; Rhodes, C. T. Drug Dev. Ind. Pharm. 1990, 16, 91-113.



- Utilisation d'algorithmes ML (LightGBM) pour prédire les  $\Delta G$ , y compris pour des molécules absentes de la base de données.

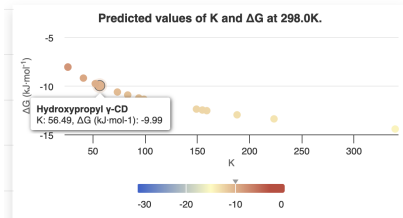


IA Predictor



ML prediction for 1H-Indole

Show 50 entries Search:


Host	$\Delta G$ (kJ.mol <sup>-1</sup> )	K	T	pH
2,6-Di-O-methyl- $\beta$ -CD	-11.38	98.67	298.0	7.0
Acetyl- $\beta$ -CD, Monoacetyl- $\beta$ -CD	-14.44	339.49	298.0	7.0
Hexakis (2,3,6-tri-O-methyl)- $\alpha$ -CD	-9.21	41.09	298.0	7.0




- Les utilisateurs peuvent enrichir la base de données en ajoutant leurs propres résultats expérimentaux.
- Normes de soumission pour garantir la qualité, la fiabilité et l'uniformité des données ajoutées.

Single data  Multiple data 

Guest

Host  [Add a New Cyclodextrin Type](#) 

Method  [Add a New Method](#) 


T(K)  (Will be set to 298K if not specified)



pH  (Will be set to 7 if not specified)



$\Delta G$  (kJ.mol<sup>-1</sup>)  [Uncertainty](#)

K  [Uncertainty](#)

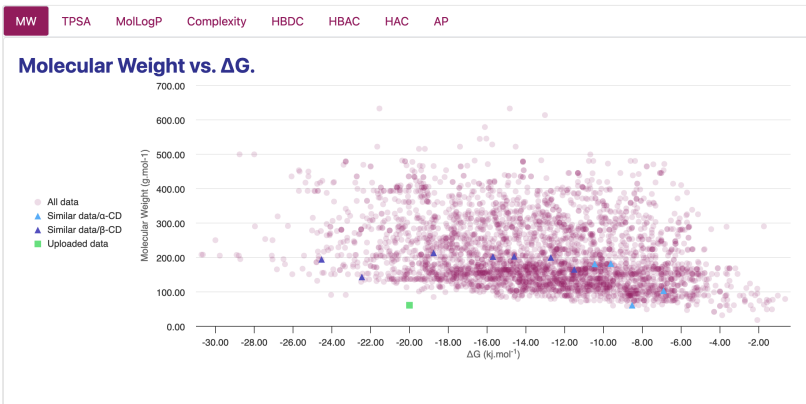
DOI  (Only Published data can be uploaded)

Email address   (For the confirmation of your entry)

Captcha  Copy the captcha here! [Reload Captcha](#) 

[Upload](#)  [Reset](#) 

- Les utilisateurs peuvent enrichir la base de données en ajoutant leurs propres résultats expérimentaux.
- Normes de soumission pour garantir la qualité, la fiabilité et l'uniformité des données ajoutées.



# 4



Open Cyclo. database

Identifiant  
pérenne (DOI)



Open Cyclo. database

Licence Open  
Database  
License v1.0

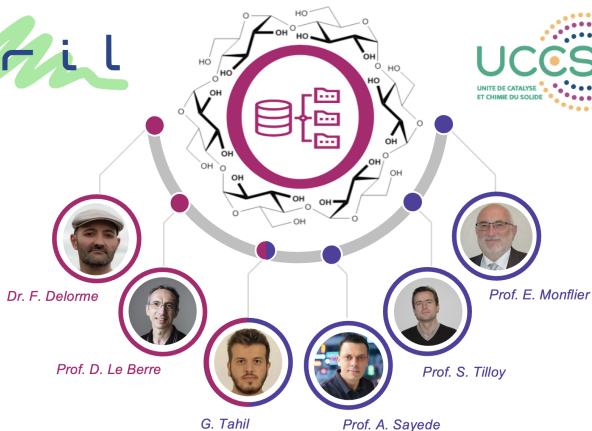


Données en  
fichier CSV  
formaté et  
structuré ;



Données de  
haute qualité :  
claires, vérifiées  
et bien  
documentées  
(ReadMe,  
DataPaper, site  
web, etc.)

- **OpenCycloDB** : Une base de données conçue pour soutenir la recherche sur les cyclodextrines en facilitant l'accès, la gestion et la réutilisation des données.
- **Adoption des principes FAIR** : Mise à disposition de données accessibles, réutilisables et bien structurées pour encourager la transparence scientifique.
- **Pour la communauté scientifique** :
  - Accès simplifié aux données fiables pour des études approfondies et reproductibles.
  - Prédiction améliorées pour des propriétés comme l'énergie de complexation, grâce aux algorithmes de machine learning.
  - Promotion de la collaboration pour enrichir la base de données et faciliter de nouvelles applications scientifiques.
- **Un pas vers la science ouverte** : En partageant ces données, nous espérons modestement contribuer à l'avancement des connaissances et à une recherche plus collaborative.



**Prof. Adlane SAYEDE**

UCCS — UMR 8181

<https://pro.univ-lille.fr/adlane-sayede>

