J Comput Aided Mol Des (2010) 24:713–717
DOI 10.1007/s10822-010-9368-y

# Cyclodextrin knowledgebase a web-based service managing CD-ligand complexation data

**Eszter Hazai · Istvan Hazai · Laszlo Demko · Sandor Kovacs · David Malik · Peter Akli · Peter Hari · Julianna Szeman · Eva Fenyvesi · Edina Benes · Lajos Szente · Zsolt Bikadi**

Received: 12 February 2010 / Accepted: 17 May 2010 / Published online: 3 June 2010


**Abstract** Cyclodextrins are cyclic oligosaccharides that are able to form water-soluble inclusion complexes with small molecules. Because of their complexing ability, they are widely applied in food, pharmaceutical and chemical industries. In this paper we describe the development of a free web-service, Cyclodextrin KnowledgeBase: (http://www.cyclodextrin.net). The database contains four modules: the Publication, Interaction, Chirality and Analysis Modules. In the Publication Module, almost 50,000 publication details are collected that can be retrieved by text search. In the Interaction and Chirality Modules relevant literature data on cyclodextrin complexation and chiral recognition are collected that can be retrieved by both text and structural searches. Moreover, in the Analysis Module, the geometries of small molecule-cyclodextrin complexes can be predicted using molecular docking tools in order to explore the structures and interaction energies of the inclusion complexes. Complex geometry prediction is made possible by the built-in database of 95 cyclodextrin derivatives, where the 3D structures as well as the partial charges are calculated and stored for further utilization. The use of the database is demonstrated by several examples.

E. Hazai · I. Hazai · L. Demko · S. Kovacs · D. Malik · Z. Bikadi (✉)
Virtua Drug Research and Development Ltd., Csalogany u. 4/C, 1015 Budapest, Hungary
e-mail: Zsolt.Bikadi@virtuadrug.com

P. Akli · P. Hari
Delta Informatika Zrt., Szentendrei út 39-53, 1033 Budapest, Hungary

J. Szeman · E. Fenyvesi · E. Benes · L. Szente
CycloLab Cyclodextrin Research and Development Laboratory Ltd., P.O. Box 435, 1525 Budapest, Hungary

**Keywords** Cyclodextrin · Interaction database · In silico prediction · Literature database · Chiral recognition

## Introduction

Cyclodextrins (CDs) are compounds consisting of six, seven or eight glucopyranose units ($\alpha$-CD, $\beta$-CD and $\gamma$-CD, respectively) and because of their inclusion complex forming properties, advantageous changes in the chemical and physical properties of the guest molecules can be achieved. Thus, they are widely used in numerous fields (analytical chemistry, environmental protection, separation science, food industry, etc.). The most important application of cyclodextrins proved to be their pharmaceutical application [1–4].

In the past decades a huge amount of experimental data has been accumulated resulting in a high number of publications. For example, according to ISI Citation Index in the period of 2003–2008, more than 600 papers were published yearly containing the word cyclodextrin in the title. As a consequence, some research groups built CD-ligand databases for *in-house* use. For example, Chari et al. built a CD-ligand database (containing 13 different CDs and 350 ligands with 642 unique entries describing experimental conditions and results of complexation) in order to use it in computational studies of cyclodextrin complexation [5]. Since the database is not publicly available, obviously it is not intended for general use. Thus, there is a strong demand to have a publicly available and easily usable complex knowledgebase platform that gathers relevant information on cyclodextrin complexation. Nowadays, the internet has become the first choice to

gather information and a high number of resources are available which provide rapid access to a vast amount of data [6]. Recently, two web applications have been implemented to search and handle CD related data. OpenCDLig (https://kdd.di.unito.it/casmedchem/) has recently been published which is intended to be a community maintained source of resources [7]. This approach is very promising, considering the fact that a large number of chemical and biological databases available on the web are developed and maintained by means of collaboration of scientists. The idea that many motivated scientists concurrently contribute and maintain a database that would speed up data accumulation has really been justified. To support this notion, NMR ShiftDB can be cited as a good example [8]. It is a collection of organic structures with their NMR (nuclear magnetic resonance) spectral data and in 2004 contained data on about 10,000 chemical structures, whereas according to its homepage to date it contains 35,239 spectra submitted by 2,142 registered users. Due to its novelty, however, OpenCDLig now has only several hundred entries (434 complexes and 626 experiments) and thus, the search of the literature is rather limited. In addition, it is unlikely that individuals are going to submit older (maybe a few decades old) results to this database, which would be needed for a comprehensive data set. Consequently, it is highly desirable to have a dedicated CD complex database containing published scientific literature (papers, books, patents, etc.). The Cyclodextrin Database was built to meet this demand. It was introduced in 2007, but has not yet been reported in the scientific literature. Here we report the Cyclodextrin KnowledgeBase, an expanded version of Cyclodextrin Database web-based system freely accessible at http://www.cyclodextrin.net. In May 2010 it contains more than 15,000 CD-ligand complex entries, which provides structure searching (both similarity and substructure) for cyclodextrin complexation and chiral recognition. Moreover, almost 50,000 publication details covering the period between 1891 and 2010 are collected in a unique publication database. Additionally, a built-in CD-ligand docking module enables the exploration of ligand—cyclodextrin geometry. To the authors' knowledge, this is the most comprehensive database on cyclodextrin complexation available on the web.

## Methods

### KnowledgeBase design and implementation

Cyclodextrin KnowledgeBase has the user's web browser on the client side. These html-based web interfaces were tested using Internet Explorer 8 and Mozilla Firefox 3.5.

The core of the Cyclodextrin KnowledgeBase web application is our integrating PHP software connected to a MySQL relational database, where the different tasks are automatically managed by daemons running on our servers. The MySQL database plays a central role in storing all pieces of information required for the application to fully function, e.g. molecular data of cyclodextrins, ligands, etc. The input data is read from the database and output data is directed into the database. Besides, the powerful web application Zend Framework is incorporated. It allows easy, fast and robust web development by leveraging the Model-View-Controller (MVC) paradigm. It also provides components for the MVC, Row Data Gateway or Table Data Gateway design patterns.

For chemical structure drawing Marvin Sketch (Chemaxon, http://www.chemaxon.com) is provided for users to draw a small molecule structure—or, alternatively, to upload a ligand structure file in most popular formats (e.g. mol; mol2; pdb). For file format conversion Open Babel 2.2.3 (www.openbabel.org) is used. Two-dimensional (2D) and three-dimensional (3D) molecular structures are represented by Marvin View. Generated by molecular docking, CD-ligand complexes are rendered by VMD [9]. Jchem engine, a collection of programs and toolkits for many tasks (Chemaxon, http://www.chemaxon.com) is applied for structure search by means of calculating dissimilarity coefficients of molecular descriptors using Tanimoto metrics. In the Analysis Module, docking calculations are performed using AutoDock 4.2 software with built in scoring function, allowing docking of flexible ligands to cyclodextrins. The calculations are performed using a distance-dependent dielectric function to model solvent. The first rank result based on the calculated free energy of binding is automatically presented as the final result.

When using the Analysis Module, each in silico experiment belongs to and is only accessible by the user who created it.

The inventory of incorporated software is given in Table 1.

### Cyclodextrin structures

Three dimensional structures of $\alpha$-CD, $\beta$-CD, $\gamma$-CD and 95 cyclodextrin derivatives built based on the crystal structures of parent cyclodextrins are offered for molecular docking calculations. The optimized structures and partial charges of cyclodextrins were calculated at a given pH using molecular mechanics (MMFF94) and semi-empirical quantum mechanical (PM6) methods. These structures are incorporated as an integral part of the Analysis Module.

## Results

### Data content and usage

Cyclodextrin KnowledgeBase contains four modules: Publication, Interaction, Chirality databases and an Analysis

**Table 1** Applications used in the Cyclodextrin KnowledgeBase

| Tool | Description |
|---|---|
| Apache | HTTP server for load balancing |
| MySQL | SQL Server |
| Babel | File format converter in Chemistry |
| Jmol | 3D molecular viewer |
| Chemaxon Marvin | Ligand structure sketching and editing |
| Chemaxon Jchem | Ligand physico-chemical property calculation |
| Autodock | Software for molecular docking calculation |
| MOPAC2009 | Semiempierical quantum chemical software for partial charge calculation |

Module. The Publication Module (Fig. 1a) contains 47,235 entries including references to peer-reviewed scientific papers, patents and conference abstracts. Using this module, a text search in the database can be carried out by keywords in the title, abstract, by author(s) and year of publication. In addition, for many entries a link to the publication is given. Furthermore, where possible a link to the KnowledgeBase's Interaction Module is also provided.

The Interaction Module (Fig. 1b) contains 10,847 entries. It provides structural search within the database being the most straightforward technique of data mining [10]. The chemical structure of the compound of interest can be drawn in and either substructure (the retrieval of all molecules that contain the query substructure) or similarity search (finding the compounds that are most similar to the query) can be carried out. The result output provides the list of interactions that meet the query. An interaction entry contains the structure of the complexing ligand, with its empirical formula and molecular mass, the name and structure of the interacting cyclodextrin with its physical properties (formula, average molecular mass) as well as -where available- the binding constant and experimental details such as the method of determination including the temperature and pH (Fig. 1c). This module provides a direct link to the publication source as well.

Besides the interest focusing on cyclodextrins' ability to form inclusion complexes, great attention has been paid to their chiral recognition ability as well. The Chirality Module, containing 5,123 entries provides thermodynamic data of interactions with respect to the separation of enantiomers. The architecture of the module is very similar to the Interaction Module. The technique of searching and the results output is identical to that described above.

By the Analysis Module cyclodextrin complex geometry prediction to cyclodextrin hosts can easily be carried out. A structural database of 95 cyclodextrin hosts has been built in into the Analysis Module (see Methods). This module was built in because it has been recognized that in silico tools are used to explore the geometries and interaction energies of the cyclodextrin inclusion complexes (e.g. [11–13]). This module offers our web-based interface that handles all aspects of the molecular docking procedure and the DockingServer that enables the calculation of complex geometries at an atomic level as well [14]. The query compound can be drawn in (or imported) into a Marvin applet and then the structure is automatically set up for docking calculations as described in our previous work for Cycloserver [15]. After selecting (up to 5) cyclodextrin molecules, docking calculations are performed, resulting in docking energies, docking geometries and interaction surfaces of the calculated complexes. Three dimensional figures of the calculated ligand-cyclodextrin complexes are generated thus they can be visually inspected in detail. In addition, calculated docking energies and contact surfaces are also given. The Analysis Module is especially useful when no data is available in the literature concerning the complexation of the compound of interest.

## Sample uses of Cyclodextrin KnowledgeBase

A possible utilization of Cyclodextrin KnowledgeBase may be demonstrated by the following examples:

### Example 1 Coumestrol

Coumestrol (3,9-dihydroxybenzofurano[3,2-c]chromen-6-one) (see Fig. 2A) is an estrogenic and antioxidant agent with low solubility properties. The chemical shape of coumestrol orients its two hydroxy groups in the same position as the two hydroxy groups in estradiol, allowing it to inhibit the activity of certain enzymes. To improve solubility, an adequate option seems to be to form a cyclodextrin complex. A text search in the Publication Module according to its name gave three results. One of them reported $^{1}$H NMR and SEM (scanning electron microscopy) experiments and molecular modeling calculations of inclusion complexation with $\beta$-CD [16].

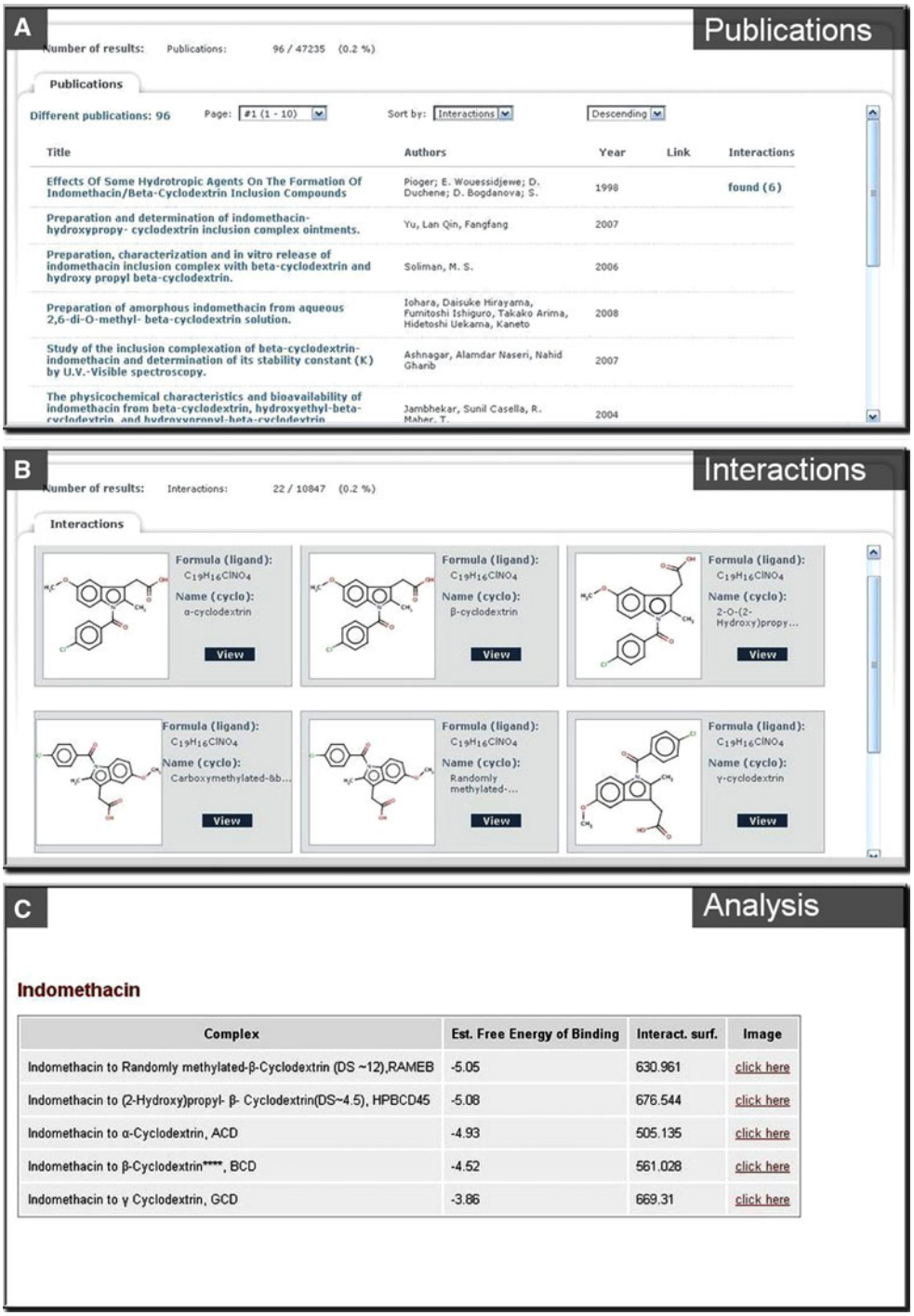


**Fig. 1** Result outputs of 3 approaches by cyclodextrin knowledgebase **a** publication **b** interaction **c** analysis modules

Example 2 Estradiol analogues

Estradiol (17$\beta$)-estra-1, 3, 5(10)- triene-3, 17-diol) (Fig. 2B) is the predominant sex hormone present in females and is contained in a number of marketed drugs. The goal of this search was to find data dealing with complexation of this compound as well as of compounds possessing identical skeleton. It was anticipated that the Publication Module will provide a vast amount of entries, and indeed, using search in the abstract field gave 180 results, and even search in the title field gave 81 results. Thus, conducting a structural search is a more straightforward approach to find data pertaining to analogues. A substructure search in the Chirality Module as well as a similarity search at 70% similarity threshold gave no results. At lower levels of similarity threshold, however, positive results (37 and 246 at 60 and 50%, respectively) were obtained, but none of them provided compounds with the sterane skeleton. A similar approach in the Interaction

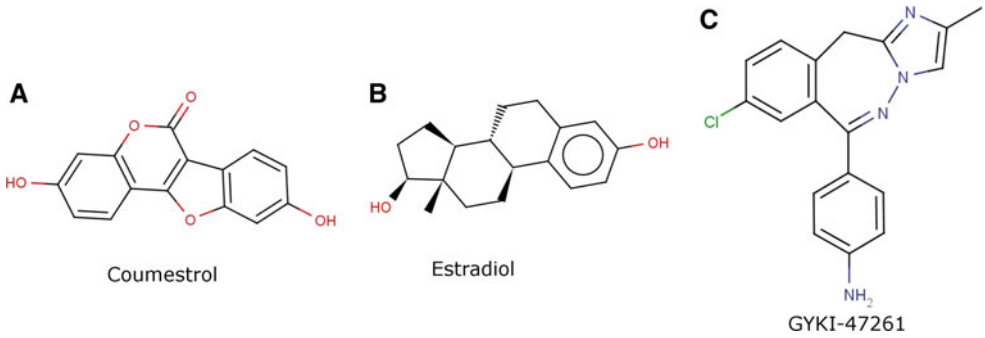


**Fig. 2** Chemical structures of the investigated compounds

Module provided better results. Namely, substructure search gave 35 matches, 9 matches of which belonged to the parent compound (i.e. to estradiol) thus indicating the fact that in this set of data 9 entries correspond to the parent compound. Similarity search at 90 and 80% similarity thresholds gave 26 and 37 matches, respectively, and all of them possessed a sterane skeleton. Although at lower similarity thresholds more hits were obtained, the number of compounds with sterane skeleton did not increase significantly (e.g. at 50% threshold, only 44 out of 300 matches offered a compound with a sterane skeleton).

Example 3 A 2,3-benzodiazepine derivative

The goal of this search was to acquire information concerning complexation with CDs of an AMPA antagonist (6-4-aminophenyl)8-chloro-2-methyl-11H-imidazo[1,2c][2, 3]benzodiazepine (GYKI-47261) (see Fig. 2C). No hits were found in any of the Modules (Publication, Interaction, Chirality) and therefore, the Analysis Module was used in order to get a hint regarding possible ligand-cyclodextrin complex geometry. The results of molecular docking with five most widely used CD structures have shown that the annellated ring system has penetrated into the cyclodextrin cavity forming hydrophobic interaction (see Fig. 2C). Besides that, the heterocyclic nitrogens are able to form hydrogen bonds with cyclodextrins as well (Fig. 3). Based on the calculated binding energy and contact surface data, it seems likely that substituted BCDs (RAMEB or DIMEB) will form a complex with this compound.

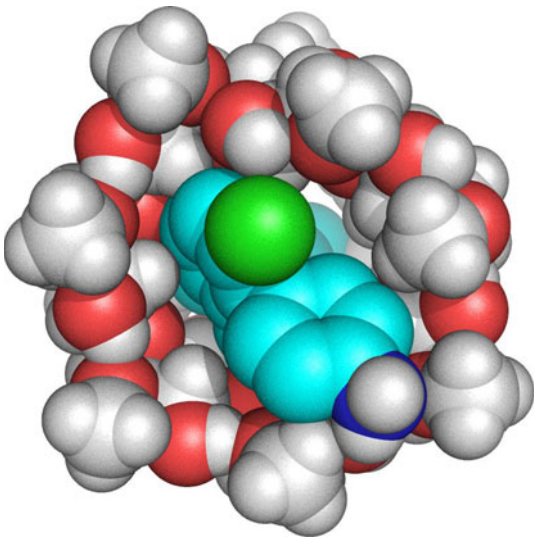

**Fig. 3** Predicted structure of GYKI-47261- DIMEB complex

## Conclusions

As a conclusion, Cyclodextrin KnowledgeBase complex web-service can be used as a robust and integrated system. It enables a fast checking of published data as well as a structural search of the available literature concerning CD-ligand complexes combined with an easy to use prediction tool for in silico study of cyclodextrin complexation.

**Acknowledgment** The financial supports of Jedlik Ányos grant 00180/2007, NKFP_A3-2008-0211 are acknowledged.

## References

1. Szejtli J (1994) Med Res Rev 14(3):353
2. Davis ME, Brewster ME (2004) Nat Rev Drug Discov 3(12):1023
3. Del Valle EMM (2004) Process Biochem 39(9):1033
4. Loftsson T, Jarho P, Masson M, Jarvinen T (2005) Expert Opin Drug Deliv 2(2):335
5. Chari R, Qureshi F, Moschera J, Tarantino R, Kalonia D (2009) Pharm Res 26(1):161
6. Jonsdottir SO, Jorgensen FS, Brunak S (2005) Bioinformatics 21(10):2145
7. Esposito R, Ermondi G, Caron G (2009) J Comput Aided Mol Des 23(9):669
8. Steinbeck C, Kuhn S (2004) Phytochemistry 65(19):2711
9. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14(1):33
10. Miller MA (2002) Nat Rev Drug Discov 1(3):220
11. Ermondi G, Cillis GI, Caron G (2007) J Incl Phenom Macrocycl Chem 57(1–4):355
12. Garcia-Zubiri IX, Gonzalez-Gaitano G, Isasi JR (2007) J Incl Phenom Macrocycl Chem 57(1–4):265
13. Zhou HW, Lai WP, Zhang ZQ, Li WK, Cheung HY (2009) J Comput Aided Mol Des 23(3):153
14. Bikadi Z, Hazai E (2009) Journal of Cheminform 1(1):15
15. Bikadi Z, Hazai I, Demko L, Hari P, Hazai E (2010) Lett Drug Des Discov 7(3):200
16. Franco C, Schwingel L, Lula I, Sinisterra RD, Koester LS, Bassani VL (2009) Int J Pharm 369(1–2):5