J Comput Aided Mol Des (2009) 23:669–675
DOI 10.1007/s10822-009-9290-3

OpenCDLig: a free web application for sharing resources about cyclodextrin/ligand complexes

Roberto Esposito · Giuseppe Ermondi · Giulia Caron

Received: 15 April 2009 / Accepted: 29 May 2009 / Published online: 17 June 2009

Abstract OpenCDLig (freely available at https://kdd.di.unito.it/casmedchem) implements a repository of complexes formed by cyclodextrins (CDs) and ligands (*L*) with related experimental data and make it available by an easy to use web interface. At the present time, the application supports two kinds of data: K, the association constant of a CD/*L* complex in [M^{-1}] and 3D structures in mol2 format. OpenCDLig is meant to become a community maintained source of experimental resources related to cyclodextrin complexes.

Keywords 3D structures · Cyclodextrins · Complexes · Database · Ligand · Ruby on Rails · Stability constants · Web application

Introduction

The widespread employ of cyclodextrins in chemistry and related fields

Cyclodextrins (CDs) are a class of biocompounds [1] of great interest not only in chemistry but also in many other fields mainly because of their inclusion complex forming capability. This important property is described in quantitative terms by the equilibrium binding constant (often called association constant and/or stability constant). For 1:1 complexation reaction between a CD and a ligand (*L*) the equilibrium binding constant is denoted $K_{1:1}$ and defined in Eq. 1

$$K_{1:1} = \frac{[L]_{\text{complex}}}{[L]_{\text{free}} \cdot [\text{CD}]_{\text{free}}} \quad (1)$$

where $[L]_{\text{complex}}$ represents the concentration of the ligand in the complex, $[L]_{\text{free}}$ is the free ligand complexation and $[\text{CD}]_{\text{free}}$ represents the concentration of the free cyclodextrin [2].

Due to the molecular complexation phenomenon, CDs (which also exhibit a very low toxicity) are widely used in many industrial products, technologies and analytical methods [3]. As an example of the usefulness of CD complexation, let's consider pharmaceutical research and development where this phenomenon enables drugs to increase solubility, reduce bitterness, enhance stability, and decrease tissue irritation upon dosing [2].

The need for OpenCDLig

Since their widespread usage, there is a lot of CDs related data reported in textbooks or print journals covering various fields of expertise (more than 30,000 publications about cyclodextrins were reported until today and the number is increasing in an exponential way [3]). Unfortunately, this wealth of knowledge is not replicated anywhere in any public repository and is thus unavailable to search engines like Google, MSN, etc.

We believe that making this knowledge available in electronic form to a wide community of scientists could greatly help the pace of the research discovery [4] and in particular a fully searchable free and open source Web-enabled tool for sharing resources of certified scientific

R. Esposito
Dipartimento di Informatica, Corso Svizzera, 185, 10149 Torino, Italy

G. Ermondi · G. Caron (✉)
Dipartimento di Scienza e Tecnologia del Farmaco, CASMedChem Laboratory, via P. Giuria, 9, 10125 Torino, Italy
e-mail: giulia.caron@unito.it

value is expected to significantly improve the future of CD-related research. Potential applications of such a tool include: (a) analysis of ligands (CDs) which bind a specific CD (ligand) to discover chemical features that correlate with complex formation skills (=discover of QSAR models), (b) identification of the best experimental technique to measure the stability of a given complex, (c) download of reliable starting 3D complex structures for in silico studies, (d) retrieval of CD-related literature and (e) publication of data of quality not amenable to journal reports (e.g. raw data).

OpenCDLig was thus created in line with this scenario and for its design and set up a team was put together which includes a medicinal chemist, a theoretical chemist specialized in molecular modeling and a computer scientist.

The potential interest in the OpenCDLig project is witnessed by the existence of recently born associations like "Association Réseau de recherche en cyclodextrines" (http://www.chimie.ens.fr/Glycoscience/Francais/RRCD/RRCD.html), the "Associazione Italiana di Chimica e Tecnologia delle Ciclodestrine" (www.cdtec.unito.it), the "Society of Cyclodextrins, Japan" (http://wwwsoc.nii.ac.jp/scdj/index-e.html) and the European Cyclodextrin Society (http://www.eurocdsoc.com/). It is then clear that there exists a network of scientists with common interests and exigencies that would benefit from the initiative.

For the sake of comprehensiveness, it must be mentioned that one database appeared in 2008 focused on CDs, named Cyclodextrin Database (http://www.cyclodextrin.net/, no reference can be given since no scientific paper about this resource has been published until now). This is a commercial database related to CycloLab, one of the most important worldwide vendor of CDs. Given its commercial nature, Cyclodextrin Database was evidently not built for collaborative purposes.

Finally, the attitude towards working with different disciplinary and towards resource sharing is exponentially increasing in scientific research. OpenCDLig is an open source chemoinformatics project that aims at catching the attention of people with different scientific backgrounds and expertise but similar research interests and, if successful, it will have positive effects on the collaboration among researchers.

Methods

Software architecture

OpenCDLig has been implemented using the Ruby on Rails (RoR) web application framework. Like other frameworks, RoR allows easy web development by leveraging the Model-View-Controller paradigm and by providing a solid object relational mapping library. RoR has received much attention over the recent years due to its clean design and the extensive use of meta-programming techniques [5].

Such techniques nicely combine with the dynamic features of the Ruby scripting language [6] and allow a fast paced development. A characteristic peculiar to the RoR approach is the *convention-over-configuration* approach: instead of providing the user with a set of configuration files to be tweaked, the framework defines a number of sensible conventions that, if followed, allows a project to be built with almost zero configuration effort. Although being a relatively young project, its open license allowed a very active community to rapidly develop. Today is one of the most successful web development project on the market. For instance RoR is the technology behind Twitter (http://twitter.com/), Yellow Pages (http://www.yellowpages.com/), git-hub (http://github.com/), and more (see http://rubyonrails.org/applications for a list of the most successful).

At the present time, we use MySQL, the standard de-facto among open source database management systems (DBMS). MySQL is open source, fast and features rich. The object relational mapping library built into RoR let it possible to write the entire application without directly referencing the database server of choice. As a consequence, OpenCDLig is (almost) database independent and there is no real technical barriers preventing one to switch to a different SQL server.

A summary of the pieces of software that are important to OpenCDLig is reported in Table 1. Figure 1 reports a slightly simplified version of the entity relationship diagram [7] of the database implemented in OpenCDLig. The entity CDLig plays the central role in the application by storing all the pieces of information required to model complexes. As reported in the figure, we allow CDLigs to have any number of cyclodextrins and any number of ligands (in most cases the number of cyclodextrins and the number of ligands is one). Also, zero or more experiments can be associated to each CDLig.

Experiments are the core datum of the application, being the most precious piece of information and the one about which most searches will be performed. The current implementation supports either specifying quantitative results such as log K constants, or uploading mol2 files describing the 3D geometry of complexes. Each experiment refers either to a scientific publication or to a private communication by researchers with proved experience in the field.

Property of data is preserved throughout the application: each experiment and each CDLig belongs to the user who first created it. Owners retain all rights over their data; they can modify the entries, and are always cited along with them. Also, data owners may ask to remove an entry from the database at any time.

Table 1 Libraries and applications used in the development and deployment of OpenCDLig

Tool	Version	License	Description
Ruby	1.8.6	Open source (GPL)	The scripting language underlying the whole application
Rails	2.2.2	Open source (MIT)	Web application framework
MySQL	5.0.26	Open source (GPL)	SQL server
Apache	2.2.3	Open source (Apache)	HTTP server used for load balancing
Mongrel Cluster	1.0.2	Open source (GPL)	HTTP server cluster used for running application instances
JME		Proprietary (free for non-commercial use)	Java applet for drawing molecules (Ertl 1998)
Jmol	11.4	Open source (LGPL)	3D molecular viewer
OpenBabel	2.2.0	Open source (GPL)	Chemistry file formats translator

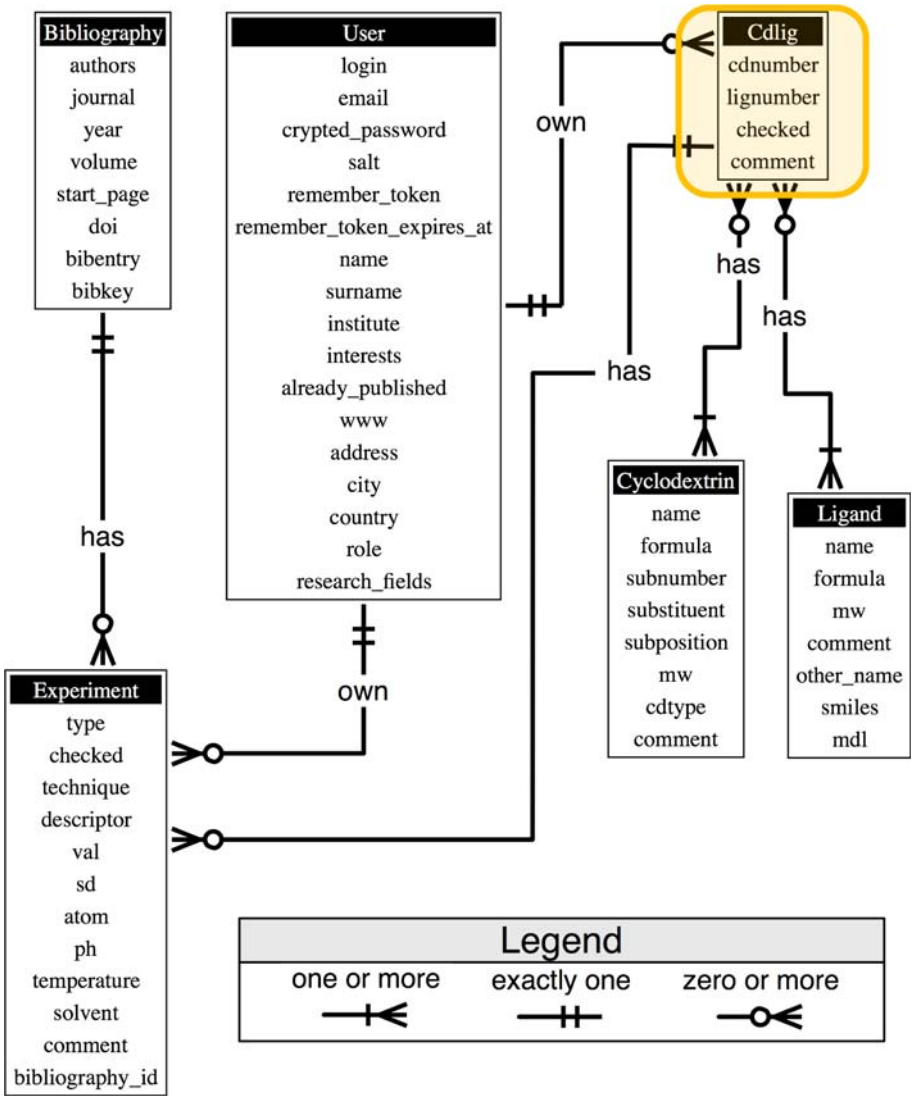

Fig. 1 OpenCDLig database schema. The entity CDLig which plays the central role in the application is in *orange*

User levels

At the first visit at OpenCDLig, users are asked to sign up and, after inserting some relevant piece of information, they are authorized to log in. At first, users are recognized as viewers by the software. Users with such role can access to all public sections of the application, they can search the database, drill down the results and export them in CSV

format. The other two roles supported by the system are "normal" and "admin". Users with role normal have all privileges reserved to viewers. Additionally, they can insert new pieces of data and be recognized as data owners. Finally, users with role admin have full privileges: they can delete and modify any inserted datum; they can ban misbehaving users, raise or lower the privileges, and validate pieces of data.

Administrators decide the roles of each user of the application. When a new user registers on the system, OpenCDLig administrators check her publication records. If she happens to have already published in open access repositories recognized internationally (e.g. Medline, PubMed Central), then she will be asked to actively participated to the initiative and, on agreement, her role will be set to "normal". Consequently, she will be allowed to insert experiments and complexes in the application.

Data peer review

Since the normal status is granted upon check of relatively mild conditions, we expect practically any user of the application to become a 'normal' user. This implies that, in order to guarantee the quality of the inserted resources, some kind of data validation needs to take place. The validation process is carried on by a group of users selected based on the volume and quality of their scientific production. The "Associazione Italiana Chimica e Tecnologia delle Ciclodestrine" (http://www.cdtec.unito.it/) agreed to carry on the selection process by proposing researchers among the ones that are believed to be capable and willing to work as data reviewers. More importantly, this reviewing process will be hopefully integrated and supplemented by the continual scrutiny provided by the community of OpenCDLig users.

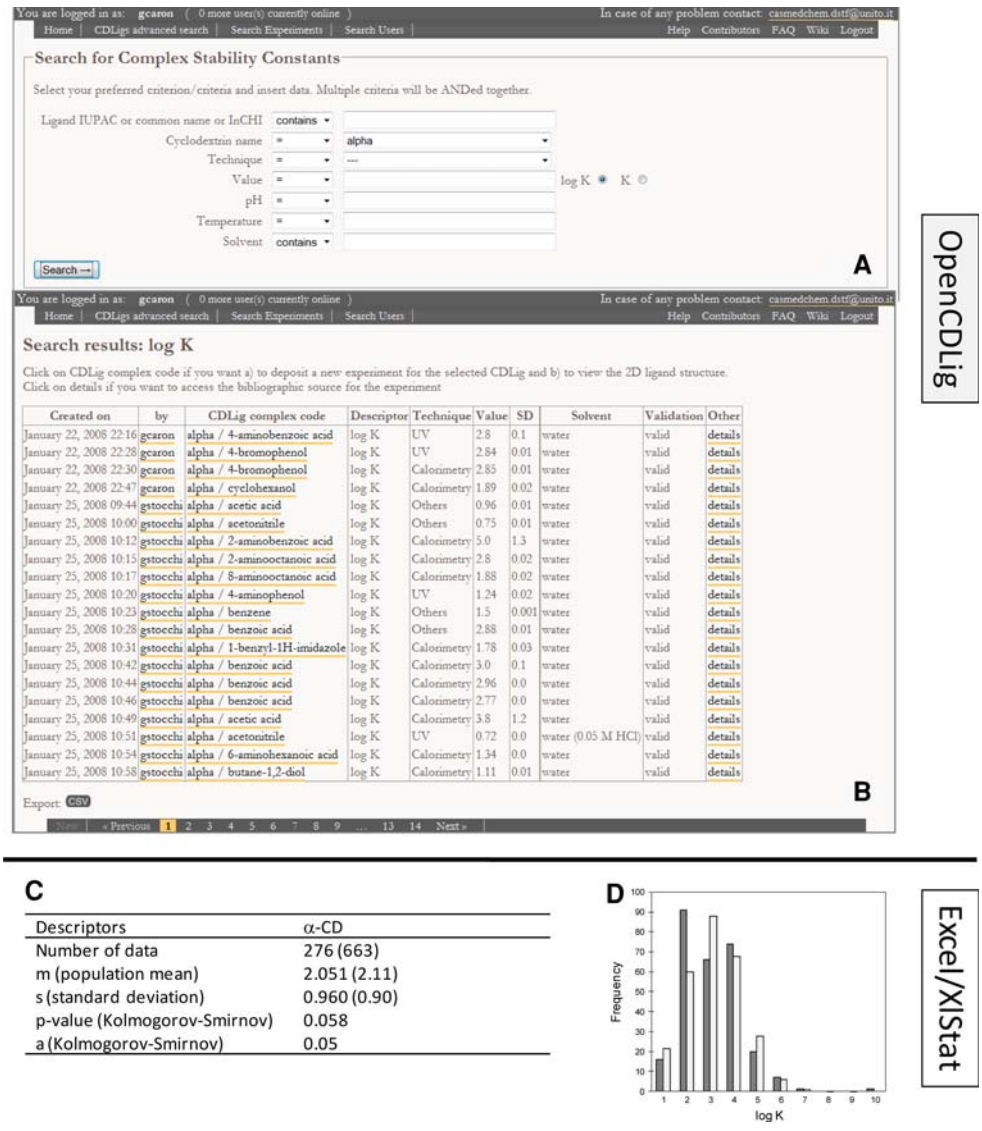

Fig. 2 Schematic representation of the procedure to search (OpenCDLig), extract (OpenCDLig) and analyze (Excel and XlStat) stability constants values expressed as log K for complexes formed by α-CD used by testers to validate the application: **a** the query; **b** experiments retrieved by OpenCDLig and export command; **c** statistical parameters for retrieved log K values calculated with XlStat (www.xlstat.com, version 2008.7.01) on an Excel spreadsheet are in line with literature data since **a** in *parentheses* data from Connors [16] shows similar mean and standard deviation with data retrieved from OpenCDLig despite the number of considered experiments and **b** Kolmogorov Smirnov test indicates that a normal distribution is highly probable; **d** the *plot* shows frequency distribution (*dark gray*) and normal distribution (*light gray*) of log K values and graphically confirms the results of the Kolmogorov Smirnov test

Data deposit policies

As a way to simplify the data reviewing process, we require that any experiment datum be accompanied by a bibliographic reference. The reference, by providing additional information over the experimental setup and results, will permit the assessment of the accuracy of the given piece of information. In particular cases, e.g. if the reference is not available (e.g. the reference manuscript is in course of preparation), the OpenCDLig policy requires that user who inserted the datum to be a scientist with provable publication records.

Results and discussion

The application

From the point of view of user interaction, OpenCDLig is organized around the concepts of "complexes" (CDLigs) and "experiments" (Fig. 1).

A CDLig is defined by a cyclodextrin (CD, e.g. α-CD), a ligand (e.g. benzene) and by the stoichiometry (e.g. one CD contains one ligand, 1:1).

The CD can be native (=not substituted) or derivatized (=substituted). The application describes each CD using the following attributes: (chemical) name (possibly assigned by the user following the indications present in the literature [8]); the type (α, β, γ, δ); the chemical formula (e.g. $C_{42}H_{70}O_{35}$ for β-CD); the molecular weight (MW); and the number, type and positions of substituents.

The application allows one to specify the chemical structure of the ligand using the well-known JME graphical editor (Table 1). Upon submission the graphical representation of the ligand structure will be converted into a SMILES [9] and an InCHI code (which guarantees the uniqueness of the code and supports for Google searches, for more details see http://www.iupac.org/inchi/) and stored in the database. Specifying the IUPAC name (=the systematic name), the common name (also called generic name, e.g. aspirin), the chemical formula and the molecular weight completes the description of the ligand.

An experiment is the procedure responsible for the production of scientific information. The application allows each user to associate one or more experiments to a single complex. In the present version, OpenCDLig database supports two types of experiments: (a) the numerical values of K in [M^{-1}] and their standard deviation ($K_{1:1}$ is defined in Eq. 1, and in OpenCDLig it is characterised by the experimental technique used to make measurement, e.g. calorimetry, the pH and the temperature at which the determination was performed, the solvent used and the bibliographic source), and (b) the 3D structures of complexes (in turn defined by the technique used to obtain the structure, e.g. NMR, molecular modeling, etc, the mol2 file and the bibliographic source).

The two series of data (K and 3D structures) are devoid of any relationship and thus it is possible to introduce K data and not 3D structures and vice versa. The introduction of both of them is permitted as well.

OpenCDLig provides an extensive support for user-defined queries. Queries can be posed to search for particular complexes, for experiments or about the user who introduced the data. Many criteria can be used to search for the complexes. Retrieved complexes are listed in a summary table, with the option to drill it down to see more details of a single CDLig. (e.g. the 2D structure of the ligand). The experiment search is performed by several criteria (an example is shown in Fig. 2a). Retrieved stability constants (K) experiments (Fig. 2b) can be downloaded using the comma-separated values (CSV) format and analyzed by common statistical tools (Fig. 2c, d). For 3D structures experiments, retrieved complexes can be

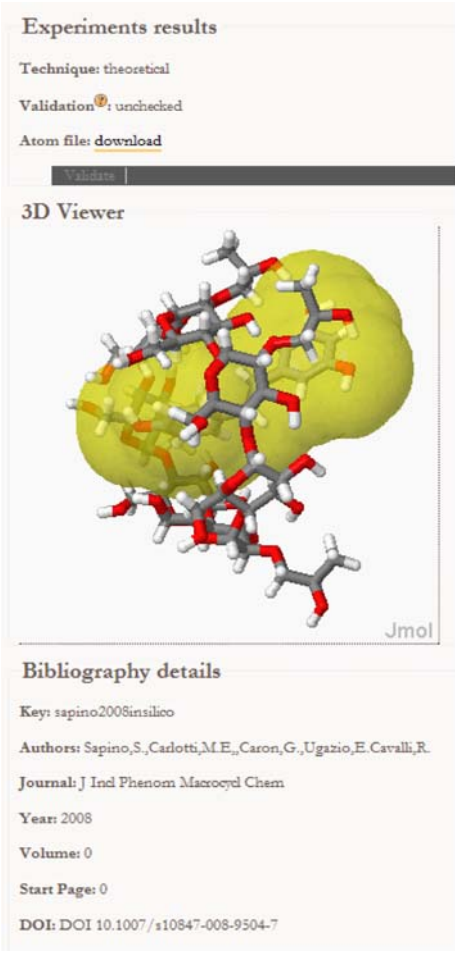

Fig. 3 The result of the retrieval procedure for the query "resveratrol" as a ligand name in the Search Experiment section. The Van Der Waals molecular surface of the ligand is shown in *yellow*

visualized on-line through Jmol (Fig. 3) and/or can be downloaded in .mol2 format and manipulated by most common molecular modeling packages.

The ability to search the database according to user-defined criteria is an important tool also when the final goal of the user is to insert new data. In facts, each experiment needs to refer to the complex to which it belongs. Then, the very first operation that needs to be done to insert a novel experimental datum is to search the database for the target complex and navigate to its details page. Only from here, the user has the opportunity of inserting the results of his experiment. The CDLig search procedure is available (a) in a simple form from the home page (Fig. 4a, b) in an advanced form from a dedicated page.

In case the CDLig cannot be found, the user needs to create the required complex by specifying a ligand and a CD which, in turn, can either be already present or created ex-novo (Fig 4b, c, respectively).

To deposit stability constant data, the user will be required to provide all the attributes that define the experiment (see above) (Fig. 4d) whereas to deposit 3D structures, he will be asked to upload a file in the mol2 format. For both kind of experiments, the insertion of the full bibliographic source is mandatory.

Surveying and testing

Extensive surveying of under- and post-graduated students and academic staff was conducted at the Dipartimento di Scienza e Tecnologia del Farmaco. Oral interviews were undertaken and revealed a great interest in the initiative and curiosity about a project not commonly addressed in medicinal/pharmaceutical chemistry departments. People recruited for the survey were also asked to test the application in practice using the Wiki site as a tutorial. In particular, most log K data listed in the well-known paper by Rekharsky et al. [10] and in some Zia's papers [11, 12] were introduced in the database, searched, retrieved and analyzed to compare their distribution with published results [13]. Figure 2 schematically shows how testers operated to validate OpenCDLig using stability constants of α-CD as an example. To date, about 400 complexes, and 580 experiments were introduced by testers.

During the testing, the retrieval of the IUPAC name was individuated as the longest and most complicated step. Moreover, users showed almost no knowledge of the InCHI format and confirms that the use of IUPAC and common compound names is still universal, despite efforts to introduce identifiers and line notation [14].

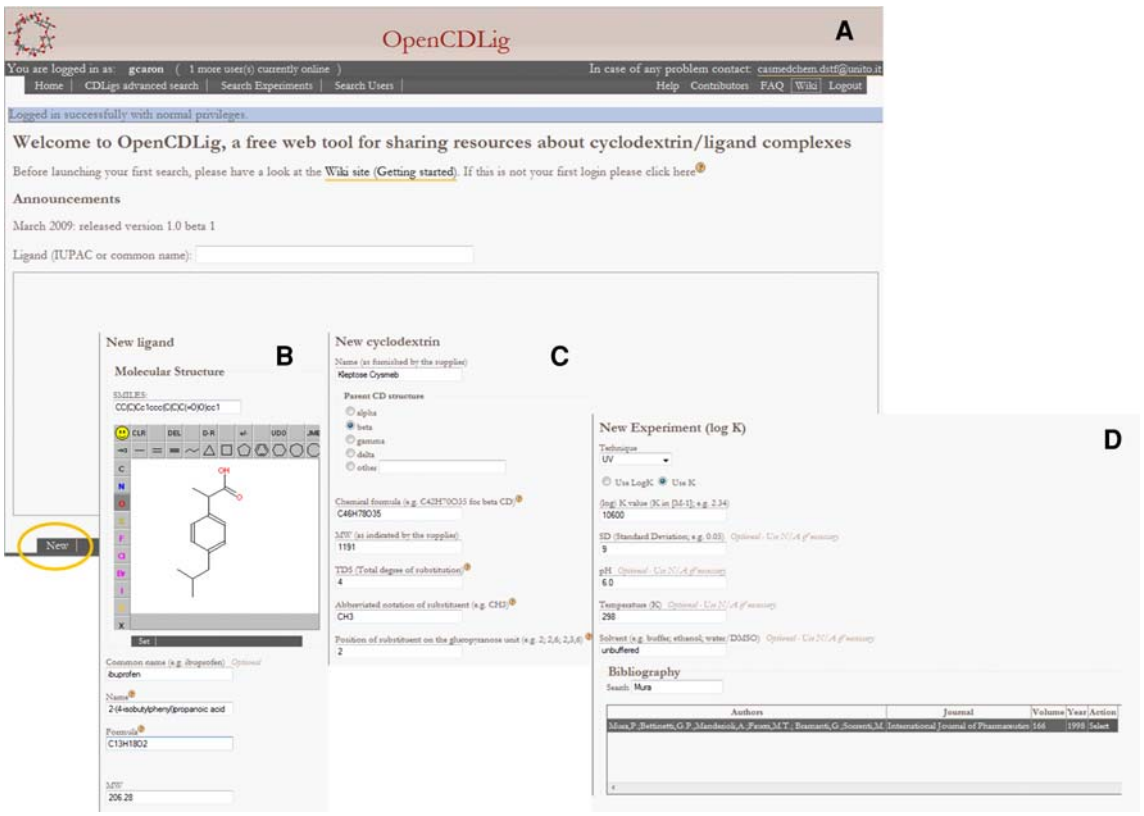

Fig. 4 Screenshot of some OpenCDLig tools: **a** the home page; **b** the creation of a new ligand; **c** the creation of a new cyclodextrin; **d** the creation of a new experiment

Finally, the volunteers were asked to deposit a 3D structure, retrieve and visualize it with OpenCDLig. The result of the retrieval procedure for the query "resveratrol" as a ligand name is shown in Fig. 3 which shows the complex between Resveratrol and HP-β-cyclodextrin as calculated by a docking procedure described in the literature [15].

Conclusion and perspectives

We believe that a collaborative platform allowing direct deposition of CD complexes data by experimentalists is the best approach to collect records for a shared and freely accessible database.

From our side (and thus before feedback of people involved in the field), we believe in the growth of the initiative and thus we are planning to extend the type of experiments to be included in the database; in particular we look with particular interest towards thermodynamic, solubility and toxicity data. Work is also in progress to include more host-guest systems in the database.

One of the driving forces of the Internet is the work of millions of users that share data and related information in a free and collaborating environment. One of the most used paradigms in this context is based on "social tagging" (also known as "Folksonomy") technologies where the users collaborate in describing pieces of information by assigning meaningful labels to them. In such systems the meaning of unintelligible objects is partly uncovered by the network of labels that are assigned to that objects and to similar ones. The additional information provided by these semantic labels allows: better results to be provided to user searches, building recommendation systems, building a stronger sense of community, and, ultimately, it allows a better user experience. How such kind of technology can be adapted to work with OpenCDLig is not immediately clear since most pieces of data are already described using rich descriptions. However, qualitative descriptions of the data and its possible usage could be easily understood as a nice feature currently lacking in the system.

Acknowledgments This work was financially supported by the University of Turin.

References

1. Oakes JV, Shewmaker CK, Stalker DM (1991) Production of cyclodextrins, a novel carbohydrate, in the tubers of transgenic potato plants. Biotechnology 9:982–986
2. Stella VJ, Rajewski RA (1997) Cyclodextrins: their future in drug formulation and delivery. Pharm Res 14:556–567
3. Szejtli J (1998) Introduction and general overview of cyclodextrin chemistry. Chem Rev 98:1743–1753
4. Agrafiotis DK, Bandyopadhyay D, Wegner JK, van Vlijmen H (2007) Recent advances in cheminformatics. J Chem Inf Model 47:1279–1293
5. Geer D (2006) Will software developers ride ruby on rails to success? IEEE Comput 39:18–20
6. Thomas D (2005) Programming Ruby. The Pragmatic Programmers LLC, Raleigh
7. Silberschatz A, Korth HF, Sudarshan S (2002) Database system concepts. McGraw-Hill, New York
8. Thompson DO (1997) Cyclodextrins-enabling excipients: their present and future use in pharmaceuticals. Crit Rev Ther Drug Carrier Syst 14:1–104
9. Weininger D (1988) SMILES, a chemical language and information system 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
10. Rekharsky MV, Inoue Y (1998) Complexation thermodynamics of cyclodextrins. Chem Rev 98:1875–1917
11. Zia V, Rajewski RA, Stella VJ (2000) Thermodynamics of binding of neutral molecules to sulfobutyl ether á-cyclodextrins (SBE-á-CDs): the effect of total degree of substitution. Pharm Res 17:936–941
12. Zia V, Rajewski RA, Stella VJ (2001) Effect of cyclodextrin charge on complexation of neutral and charged substrates: comparison of (SBE) 7M-β-CD to HP-β-CD. Pharm Res 18:667–673
13. Houk KN, Leach AG, Kim S, Zhang X (2003) Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. Angew Chem Int Ed 42:4872–4897
14. Sayle R (2009) Foreign language translation of chemical nomenclature by computer. J Chem Inf Model 49:519–530
15. Sapino S, Carlotti ME, Caron G, Ugazio E, Cavalli R (2009) In silico design, photostability and biological properties of the complex resveratrol/hydroxypropyl-β-cyclodextrin. J Incl Phenom Macrocycl Chem 63:171–180
16. Connors KA (1997) The stability of cyclodextrin complexes in solution. Chem Rev 97:1325–1357