ChEBI: Quick tour

Gareth Owen [1]

- Chemical biology
- Beginner
- 0.5 hour

This quick tour provides a brief introduction to ChEBI, the EBI's Chemical Entities of Biological Interest database, which focuses on 'small' chemical compounds. For a more detailed walkthrough of ChEBI, have a look at our ChEBI: the online dictionary for small molecules [2] tutorial.

Updated in March 2015.

Learning objectives:

- Basic understanding of ChEBI Database and how you can use it to access chemical compounds of interest.
- Know where to find out more about ChEBI

What is ChEBI?

ChEBI scope and contents

ChEBI [3], EMBL-EBI's database of Chemical Entities of Biological Interest, is a freely available, manually annotated database of small molecular entities (molecules not encoded by the genome, Figure 1). These could include any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion [4], complex, conformer [5], or anything else that is a separately distinguishable entity.

ChEBI focuses on chemical nomenclature and structures, and provides a wide range of related chemical data such as formulae, links to other databases and an ontology [6] for the chemical space. It aims to bridge the gap between small molecules [7] and the macromolecules with which they interact in living systems.
What data can I find in ChEBI?

The ChEBI [8] database combines chemical nomenclature, structures, synonyms and related chemical information from a number of freely accessible sources. All data are manually annotated to a high standard before public release, using nomenclature, symbolism and terminology endorsed by the International Union of Pure and Applied Chemistry (IUPAC [9]) and the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB [10]).

A major feature of ChEBI is that entries are related to each other using the ChEBI ontology [6]. This represents the meaning of the data in a structured manner by creating relationships between entities and their parents (less specialised terms) and/or children (more specialised terms). ChEBI is probably the only chemistry database to include an ontology. The ChEBI ontology is used by a number of biological ontologies to manage their chemistry-related terms. For more information, see 'The ChEBI ontology [11]' section of this Quick tour.

Data in ChEBI are divided into:

1. Fully annotated ('three star') entries.
2. Data curated elsewhere but not yet checked by ChEBI curators.

You can choose to search for 'three star' entries only, or to search 'All in ChEBI' (three star entries plus data curated elsewhere, Figure 2).
Different types of data

Structural data

If a chemical structure can be drawn for an entity, it will be shown in the top left-hand corner of the main ChEBI [12] display page for the entity (see Figure 3 below). Structures of large or complex entities can be enlarged if desired by hovering the cursor over the structure. By default, the structure is shown as a still image; clicking in the ‘Dynamic applet’ checkbox will open an applet enabling alternative displays of the structure to be selected.
Figure 3 Results of a search for benzophenone, showing the section on structural data.

Immediately underneath the structure, there are direct links to carry out the commands ‘Find compounds which contain this structure’, ‘Find compounds which resemble this structure’, or to ‘Take the structure to the Advanced Search’, which allows the structure to be used in conjunction with text and ontology searches.

To the right of the structure, the recommended ChEBI [8] Name is shown, along with the ChEBI ID (a unique and stable identifier [13]), an ontological definition, the ‘star’ status of the entry, and secondary IDs (i.e. identifiers of records that have been merged with the record; searching for any secondary ID will automatically find it and display the merged record).

If a Wikipedia article is available for the entity, the introductory paragraph of the article is shown in a 'Wikipedia' section underneath the structure, along with a link to the full article. Under this section, the molecular formula, charge, and molecular weight are shown, together with standard InChI [14], InChIKey [15], and SMILES [16] line entry versions of the structure.

Nomenclature

In addition to the recommended ChEBI [8] name (to the right of the structure), various other names for an entity are listed in the lower half of the main display page, together with the resource in which they were found. The synonyms are divided into the following categories:

1. IUPAC recommended name [17]
2. International Non-proprietary Names [18] (INNs) - as designated by the World Health Organisation.
3. Other synonyms
4. Brand names (generally restricted to drug entries)

Where a non-English INN or other synonym [19] is given, the language is indicated by the appropriate flag to the right of the name (Spanish, French and Latin in Figure 4 below).
Figure 4 Results of a search for paracetamol, showing the section on synonyms.

**Links to other databases and registry numbers**

The 'Database Links' section of the main display page (beneath the 'Synonyms' section) provides links to other resources (Figure 5). The links have been selected by ChEBI [8] curators as being particularly relevant to the ChEBI entity.

Some resources (e.g. ChemIDplus [20], NIST Chemistry WebBook [21]) use Chemical Abstracts Service (CAS [22]) registry numbers as their identifiers. In these cases, the links to the resources are provided in the 'Registry Numbers' section. In addition to CAS registry numbers, this section lists Beilstein [23], Reaxys [24] and Gmelin [25] registry numbers, as appropriate, together with the sources of this information.
Figure 5 Results of a search for paracetamol, showing the section on database links and registry numbers.

At the bottom of the 'Database Links' section on the main display page for an entity is a link marked 'View more database links'. Clicking on this opens a separate display page (also accessible by clicking on the 'Automatic Xrefs' tab at the top of the page), where a series of automatically-generated cross-links are provided.

To return to the main display page, simply click on the 'Main' tab at the top of the page.

Citations

The 'Citations' section is found immediately below the 'Registry Numbers' section at the bottom of the main display page for an entity (Figure 6). Here, a list of selected references is provided, together with links for accessing the abstracts or full papers if desired.
Figure 6 Search results for paracetamol, showing the citations section.

**Natural products data**

ChEBI [8] currently contains data for over 5000 natural products. In addition to the chemical structure, nomenclature data, database links, registry numbers and citation information, the vast majority of these entries include detailed information on the source of the compound (species, strain, tissue type, etc, Figure 7). This information, which is fully searchable, is displayed on the main data page, along with links to appropriate taxonomies and ontologies, and is directly linked to references in the primary literature.
The ChEBI ontology

The ChEBI Ontology (1)

The ChEBI ontology [26] is used to classify entities according to their structural and biological properties.

For biological properties, the relationship has_role is used. For example, diclofenac (CHEBI:47381; see Figure 8 below) has_role non-narcotic analgesic and has_role EC 1.14.99.1 (prostaglandin-endoperoxide synthase) inhibitor.

Figure 7 Species and tissue information for avicularin (CHEBI:65460), showing links to appropriate taxonomies and ontologies, and to primary literature ‘evidence’ via PubMed, DOI, etc. Citation details are given where no electronic link is available.
Figure 8 Search results for diclofenac, showing the ChEBI ontology section.

For chemical properties, the is_a relationship is most commonly used (and it is the relationship which will be most familiar to ontologists), but a number of chemistry-specific relationships are also used, including:

- is_enantiomer [27] of
- is_tautomer_of
- is_conjugate_acid_of and is_conjugate_base_of
- has_functional_parent and has_parent_hydride.

Thus diclofenac is_a monocarboxylic acid, is_a secondary amino compound, and is_conjugate_acid_of diclofenac anion.

On the main display page for an entity, the chemical and biological roles and applications that are applicable to the entity are listed with the appropriate definitions immediately below the SMILES [28] string for the entry. Beneath these, the immediate (one-step) structural links are shown. These are divided into two sections: ‘Outgoing’ (i.e. less specific terms) and ‘Incoming’ (i.e. more specific terms).

The ChEBI Ontology (2)
Full details about the ontological classification of an entity can be found by clicking on the 'ChEBI Ontology' tab, situated towards the top of the screen between the 'main' and 'Automatic Xrefs' tabs. Using this view, chemical and biological roles and applications are listed as on the Main display page, while both the names and structures of compounds linked by relationships such as 'is_enantiomer [27]_of', 'is_tautomer_of' are displayed underneath. At the bottom of the page, is_a and has_part relationships are shown in a fully interactive graphical display (Figure 9).
Figure 9 Shows part of the ChEBI ontology tab for spermidine (CHEBI:16610).

Hovering the cursor over any line linking two terms in the graph will display the relationship between the terms (in this case, acetamides is_a monocarboxylic acid amide [29]), while clicking on any name in the graph (in this case, acetamides) will display the definition of the term, together with a link which will take you to the ChEBI [8] entry for that term.

What can I do with ChEBI?

- Find the correct chemical terminology using name, formula or registry numbers, including CAS, Beilstein/Reaxys and Gmelin Registry Numbers

- Visualise chemical structures and use the chemical substructure and similarity search powered by OrChem [30], an open source Oracle [31] chemistry plug-in. The facility allows you to draw or upload a chemical structure and then perform exact, substructure, or similarity searches.

- View the relationships between molecules using the ChEBI [8] ontology [6], either from within a ChEBI entry or using the EBI's Ontology Lookup Service [32].

- Bridge the gap between small molecules [33] and the macromolecules with which they interact. Biological databases such as the UniProt Knowledgebase [34] and Reactome [35] allow you to view cross-references
to all entries featuring a particular chemical.

- Download chemical structures in MDL Molfile [36] format and manipulate them using a Java applet.
- Request and discuss new entries using the ChEBI submission tool or ChEBI's SourceForge discussion forum [37].

Searching and getting data from ChEBI

Searching ChEBI

Quick search

Simply type the term (e.g. cholesterol), formula (e.g. C6H12O6), registry number (e.g. 64-17-5), InChI (IUPAC [9] International Chemical Identifier [13], e.g. InChI=1/H2O/h1H2) or ChEBI identifier (e.g. 30815) into the search box on the ChEBI home page, then click on the 'Search ChEBI' button or press 'Enter' on your keyboard.

Wildcards (*) can be used to search using a partial name; e.g. searching for cholest* will find all those entities which have a name or synonym [19] starting with 'cholest', such as cholesterol and cholesteryl ?-D-glucoside.

Advanced search

From the Advanced Search page (accessed from the menu in the top left-hand corner of any ChEBI page) searches can be performed using several terms at once or restricted to specific fields (e.g. ChEBI name, synonym, formula).

Structure-based searches can be performed by drawing or loading a chemical structure and the search can be further restricted by combining with a term-based search.

Retrieving data from ChEBI

Data download

The entire ChEBI database can be downloaded from ChEBI's ftp site [38] in several formats including SDF [39], Oracle [40] and generic database dumps [41], flat files [42] and the Open Biomedical Ontologies [43] (OBO) format.

Web services

Programmatic access to ChEBI is available through ChEBI's web services page [44].

Submitting data to ChEBI

To request a log-in to submit data to ChEBI, please visit the submissions page [45].

Getting help and support on ChEBI

Support and find out more
For more detailed information about how to use ChEBI [8], see the ChEBI user manual [46].
For information about the ChEBI mailing lists and forums, see our SourceForge page [37].
For other support-related enquiries, please contact the chebi-help [at] ebi.ac.uk (ChEBI support team).

References


Funding

ChEBI is funded by BBSRC, grant agreement number BB/K019783/1 within the "Bioinformatics and Biological Resources Fund".

Contributors

Gareth Owen [1]
EMBL-EBI
Scientific Database Curator

Gareth Owen is a member of the Cheminformatics and Metabolism group at the EBI, where he works as curator and project manager for the ChEBI database. Gareth obtained his PhD in synthetic organic chemistry from Leeds University. He continued practising bench chemistry in a collaborative project with the Biotechnology unit at Sheffield University, synthesising radioactive intermediates that were used as part of an effort to produce morphine from microorganisms. He subsequently moved into the area of cheminformatics, designing and building both reaction and molecule databases for ORAC Ltd and later for Synopsys and Accelrys, before joining EMBL-EBI in 2010.