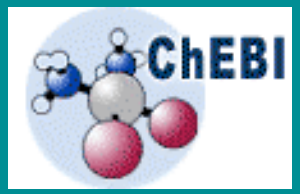
ChEBI, an Open-access Chemistry Resource for the Life Sciences:

Facilities for On-line Submission and Curation

Marcus Ennis
European Bioinformatics Institute
Cambridge
UK

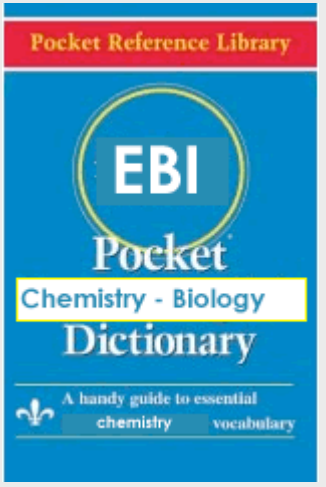


EMBL-EBI

Biocuration 2010
12 October, 2010

What is ChEBI?

- **Ch**emical **E**ntities of **B**iological **I**nterest
- Freely available
- Focused on ‘small’ chemical entities (no proteins or nucleic acids)
- Illustrated dictionary of chemical nomenclature
- Provides chemical ontology
- High quality, manual curation
- Public release: First Monday of every month
  (latest release: 22,000 manually curated entries; 620,000 total entries).
- Data made available via
  - Public web application
  - Flat-file binary table dumps
  - Oracle table dumps
  - Generic SQL table dumps
  - MDL SD file export
- Query via Web services



Access ChEBI at http://www.ebi.ac.uk/chebi/

ChEBI entry view (i)

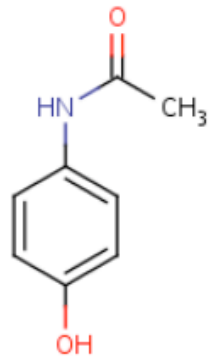
EBI > Databases > Small Molecules > ChEBI > Main

Quick search here! GO

Search for ★★★ only

paracetamol (CHEBI:46195)

Main | Automatic Xrefs



Molfile

ChEBI Name	**paracetamol**
ChEBI ID	**CHEBI:46195**
Definition	A derivative of phenol which has an acetamido substituent located *para* to the phenolic -OH group.
Last Modified	29 June 2010
Stars	★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:383982, CHEBI:116450, CHEBI:46191, CHEBI:2386

Image
Applet

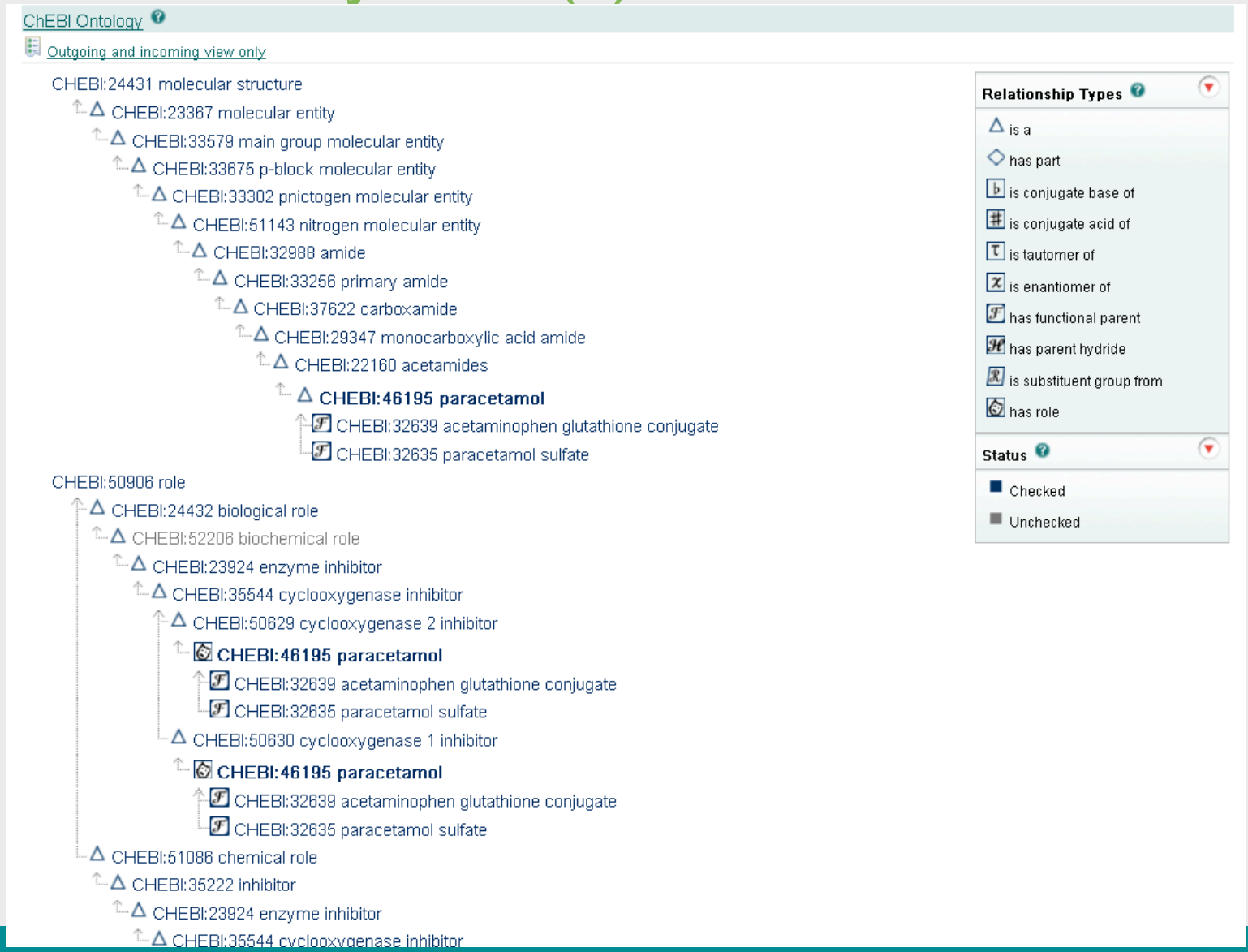
more structures >>

InChI	InChI=1/C8H9NO2/c1-6(10)9-7-2-4-8(11)5-3-7/h2-5,11H,1H3,(H,9,10)/f/h9H
InChIKey	InChIKey=RZVAJINKPMORJF-BGGKNDAXCW
SMILES	CC(=O)Nc1ccc(O)cc1

Formula	Source
C8H9NO2	KEGG COMPOUND

Charge	0
Mass	151.16260

ChEBI entry view (ii)



ChEBI entry view (iii)

IUPAC Name
N-(4-hydroxyphenyl)acetamide

INNs	Sources
paracetamol	KEGG DRUG
paracetamol	ChEBI
paracétamol	ChEBI
paracetamolum	ChemIDplus

Synonyms	Sources
4'-hydroxyacetanilide	ChemIDplus
4-(Acetylamino)phenol	ChemIDplus
4-acetamidophenol	NIST Chemistry WebBook
Acenol	ChemIDplus
acetaminofén	ChemIDplus
Acetaminofen	ChemIDplus
Acetaminophen	KEGG COMPOUND
acétaminophène	ChEBI
APAP	DrugBank
N-acetyl-*p*-aminophenol	ChEBI
p-acetamidophenol	NIST Chemistry WebBook
p-acetaminophenol	NIST Chemistry WebBook
p-Acetylaminophenol	ChemIDplus
p-hydroxyacetanilide	NIST Chemistry WebBook
p-hydroxyphenolacetamide	NIST Chemistry WebBook

Brand Name	Source
Tylenol	KEGG DRUG

Database Links	Databases
C06804	KEGG COMPOUND
D00217	KEGG DRUG
DB00316	DrugBank
TYL	PDBeChem
116450	ChEMBL

ChEBI entry view (iv)

Registry Numbers	Types	Sources
103-90-2	CAS Registry Number	KEGG COMPOUND
103-90-2	CAS Registry Number	NIST Chemistry WebBook
103-90-2	CAS Registry Number	ChemIDplus
2208089	Beilstein Registry Number	Beilstein

Citations

Jorgensen WL,Duffy EM (2000)
Prediction of drug solubility from Monte Carlo simulations.
Bioorganic & medicinal chemistry letters **10**, *1155-8 (Source: ChEMBL)* [PubMed:10866370]
[show Abstract]

Lombardo F,Shalaeva MY,Tupper KA,Gao F,Abraham MH (2000)
ElogPoct: a tool for lipophilicity determination in drug discovery.
Journal of medicinal chemistry **43**, *2922-8 (Source: ChEMBL)* [PubMed:10956200]
[show Abstract]

Colmenarejo G,Alvarez-Pedraglio A,Lavandera JL (2001)
Cheminformatic models to predict binding affinities to human serum albumin.
Journal of medicinal chemistry **44**, *4370-8 (Source: ChEMBL)* [PubMed:11728183]
[show Abstract]

Lombardo F,Obach RS,Shalaeva MY,Gao F (2002)
Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data.
Journal of medicinal chemistry **45**, *2867-76 (Source: ChEMBL)* [PubMed:12061889]
[show Abstract]

Sharma V,Duffel MW (2002)
Comparative molecular field analysis of substrates for an aryl sulfotransferase based on catalytic mechanism and protein homology modeling.
Journal of medicinal chemistry **45**, *5514-22 (Source: ChEMBL)* [PubMed:12459019]
[show Abstract]

Lombardo F,Shalaeva MY,Tupper KA,Gao F (2001)
ElogD(oct): a tool for lipophilicity determination in drug discovery. 2. Basic and neutral compounds.
Journal of medicinal chemistry **44**, *2490-7 (Source: ChEMBL)* [PubMed:11448232]
[show Abstract]

Lombardo F,Obach RS,Shalaeva MY,Gao F (2004)
Prediction of human volume of distribution values for neutral and basic drugs. 2. Extended data set and leave-class-out statistics.

The submission tool

- Web-based software utility to enable direct user submissions
- Needed to alleviate backlog of user requests
- An invitation to the community to participate in the growth and development of ChEBI
- Secure (submitter registration required)
- Simple stepwise procedure
- Instant generation of ChEBI ID and inclusion in next monthly public release

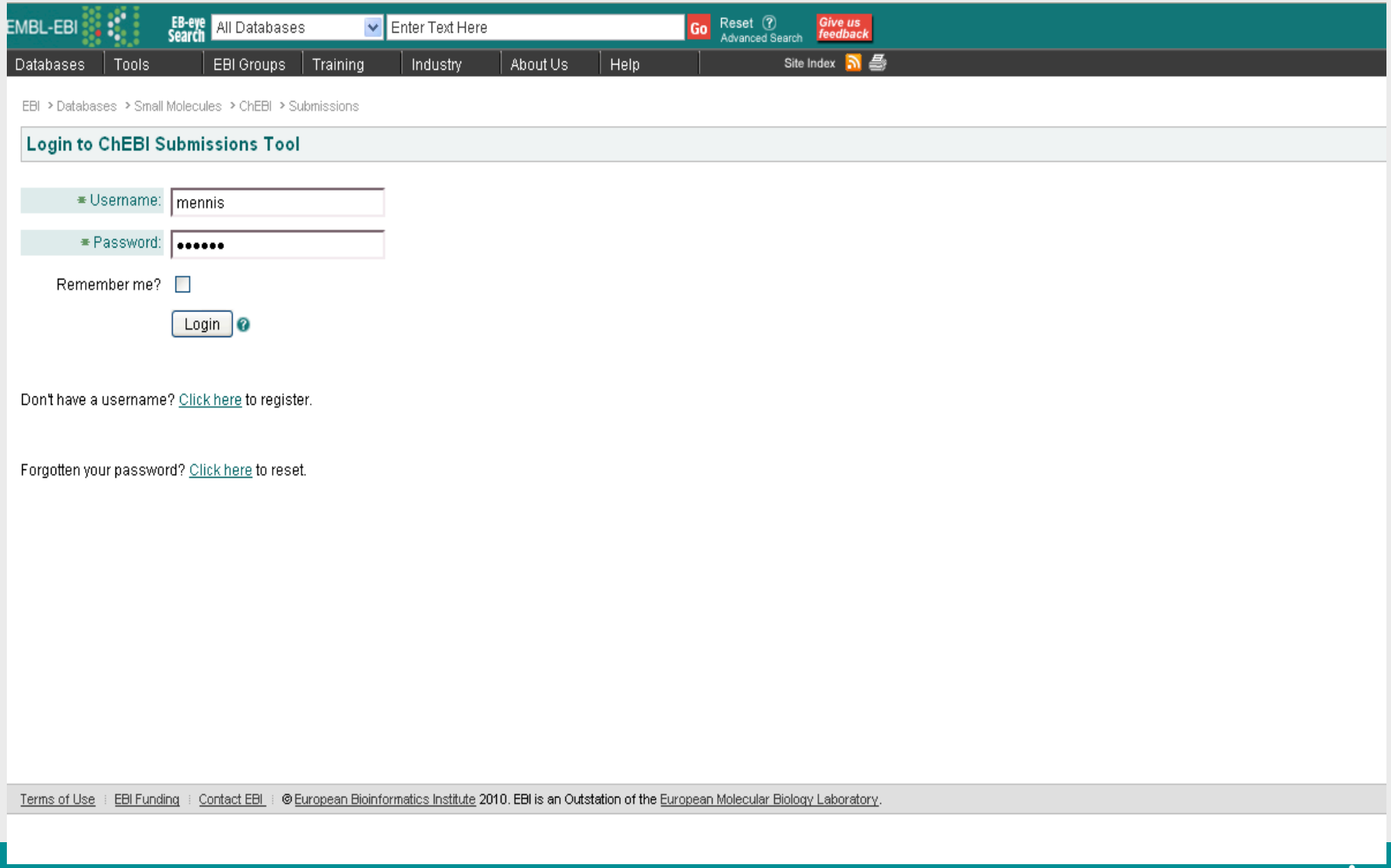
Access at https://www.ebi.ac.uk/chebi/submissions/ or via ChEBI home page

Submission validations

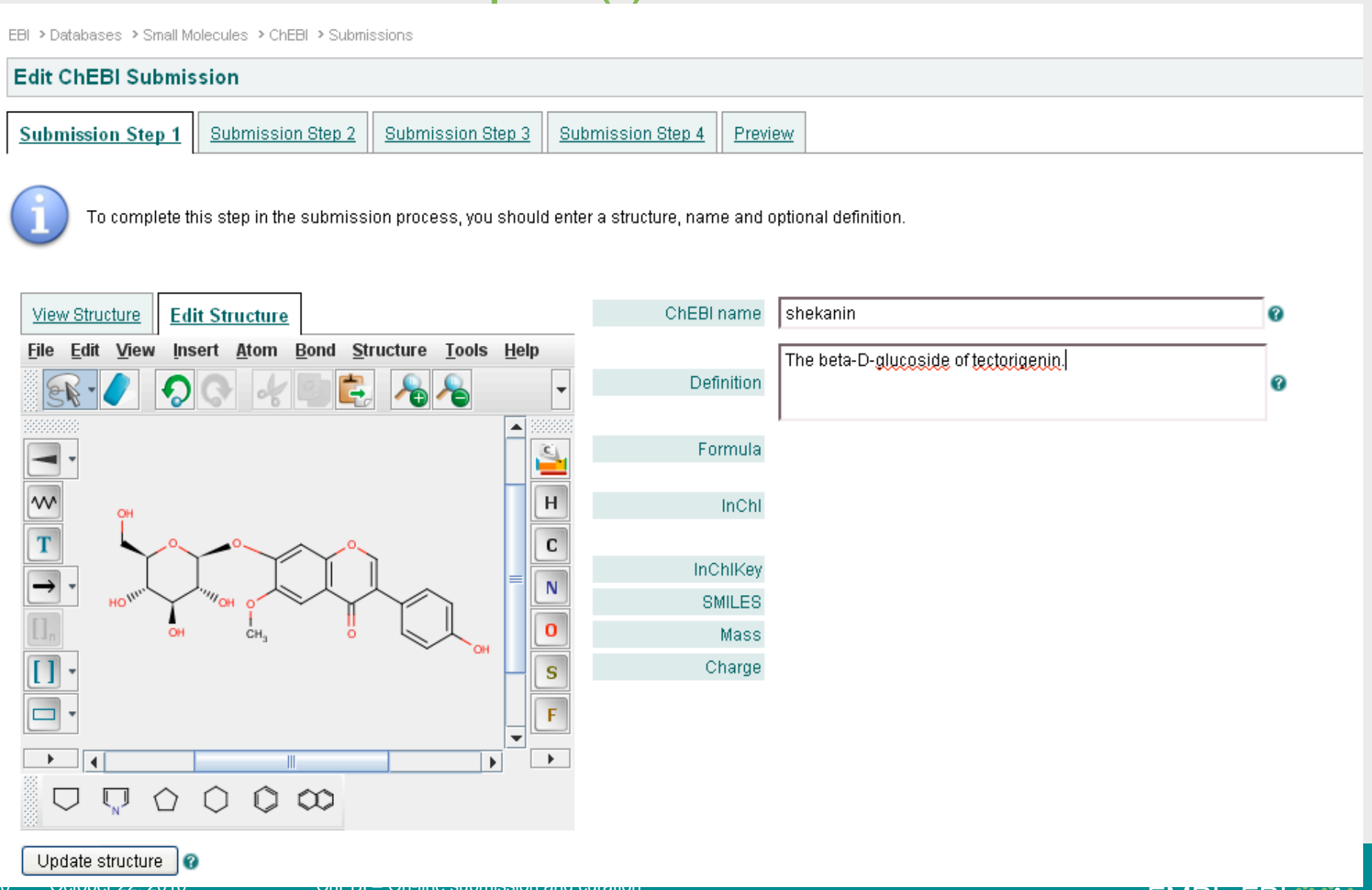
Built-in integrity checker will check for:

- Uniqueness of name
- Uniqueness of chemical structure
- Correctness of ontology relationships
  - No non-allowed cycles
  - Relationships only between entities of relevant types
- Correctness of xref format
  - CAS Registry Number
  - DrugBank
  - KEGG LIGAND/DRUG/GLYCAN
  - Lipid Maps

Login page


EMBL-EBI
EB-eye Search
All Databases
Enter Text Here
Go
Reset
Advanced Search
Give us feedback
Databases
Tools
EBI Groups
Training
Industry
About Us
Help
Site Index
EBI > Databases > Small Molecules > ChEBI > Submissions
Login to ChEBI Submissions Tool
Username:
mennis
Password:
Remember me?
Login
Don't have a username? Click here to register.
Forgotten your password? Click here to reset.
Terms of Use
EBI Funding
Contact EBI
© European Bioinformatics Institute 2010. EBI is an Outstation of the European Molecular Biology Laboratory.

Submission Step 1 (i)



Submission Step 1(ii)

EBI > Databases > Small Molecules > ChEBI > Submissions

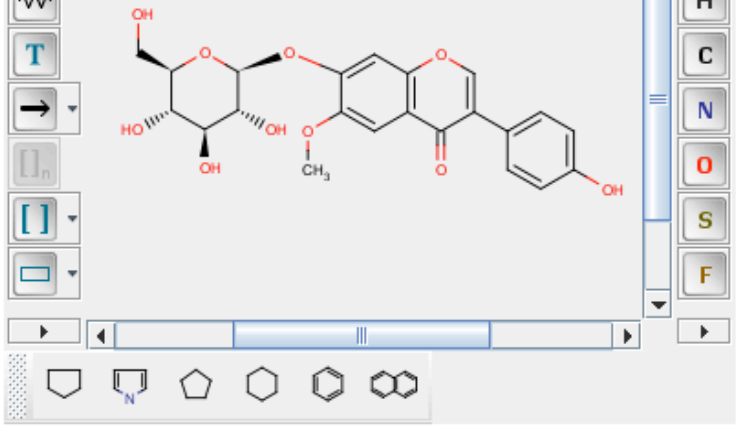
Edit ChEBI Submission

Submission Step 1 | Submission Step 2 | Submission Step 3 | Submission Step 4 | Preview

To complete this step in the submission process, you should enter a structure, name and optional definition.

View Structure | Edit Structure

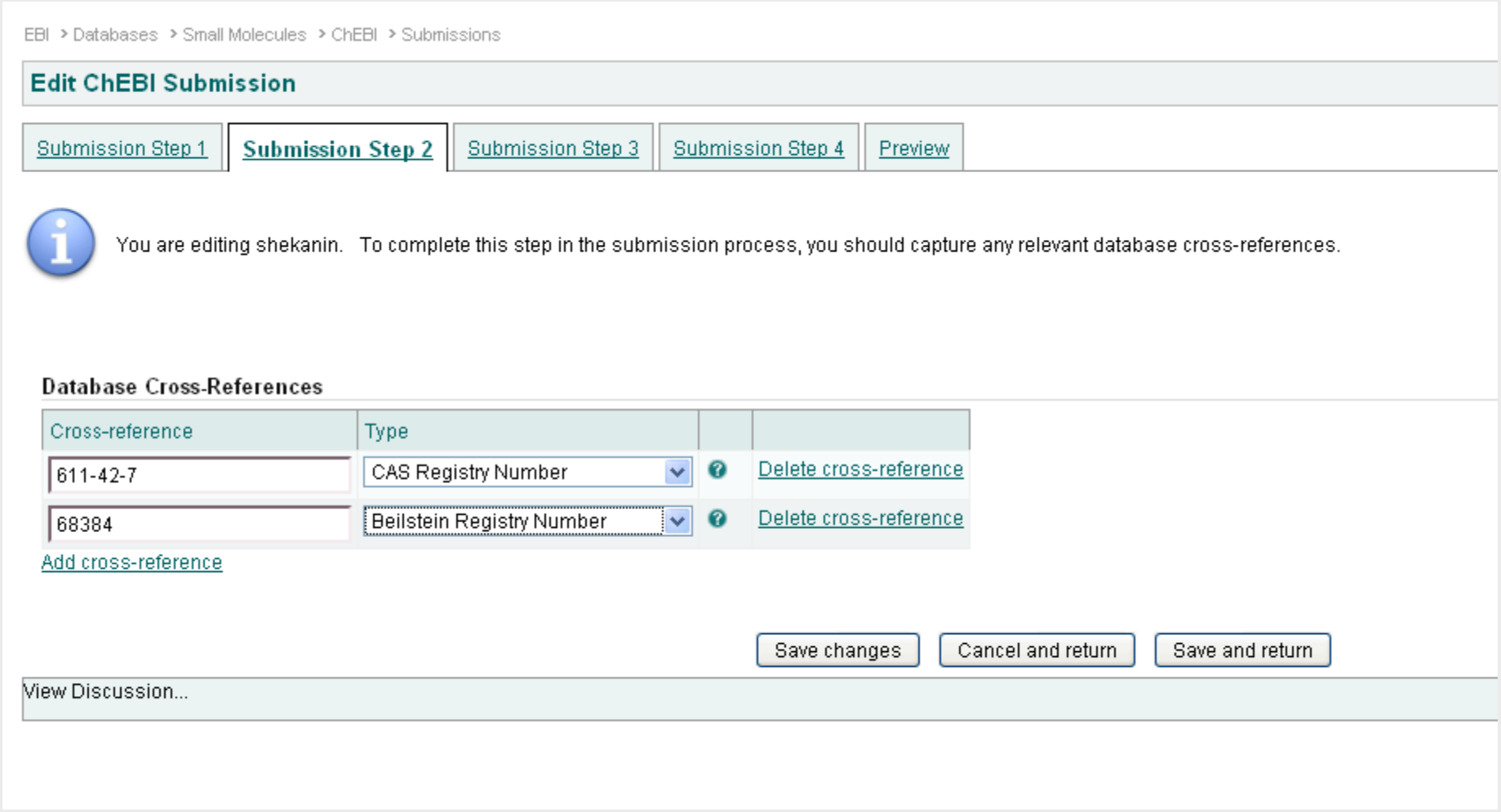
File Edit View Insert Atom Bond Structure Tools Help



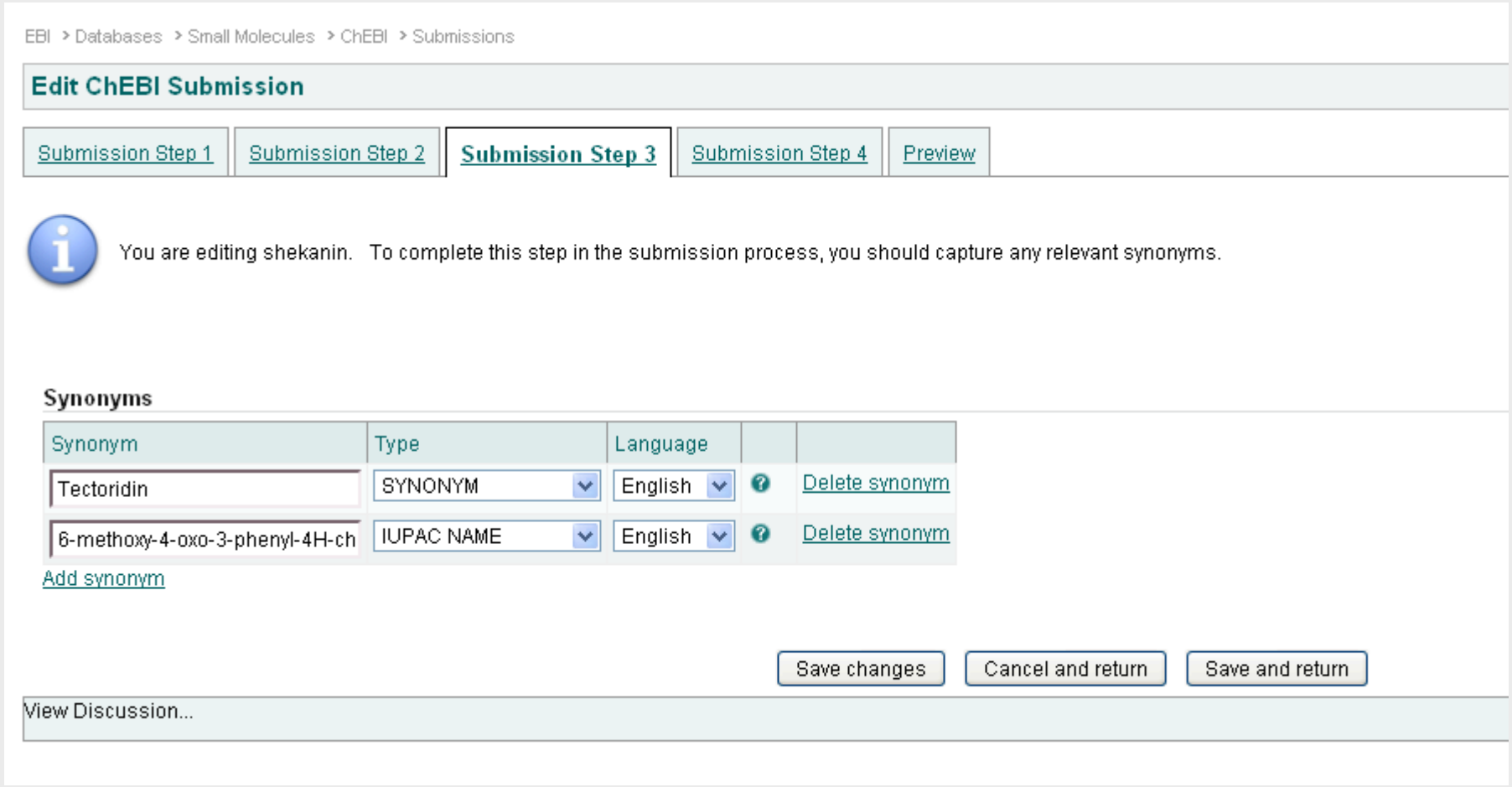
Update structure

ChEBI name	shekanin
Definition	The beta-D-glucoside of tectorigenin.
Formula	C22H22O10
InChI	InChI=1/C22H22O10 /c1-29-15-6-12-14(30-9-13(18(12)25)10-2-4-11(24)5-3-10) 7-16(15)31-22-21(28)20(27)19(26)17(8-23)32-22
InChIKey	InChIKey=OZBAVEKZGSOMOJ-MIUGBVLSBI
SMILES	COc1cc2c(cc1O[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)occ(-c1ccc(C
Mass	446.40410
Charge	0

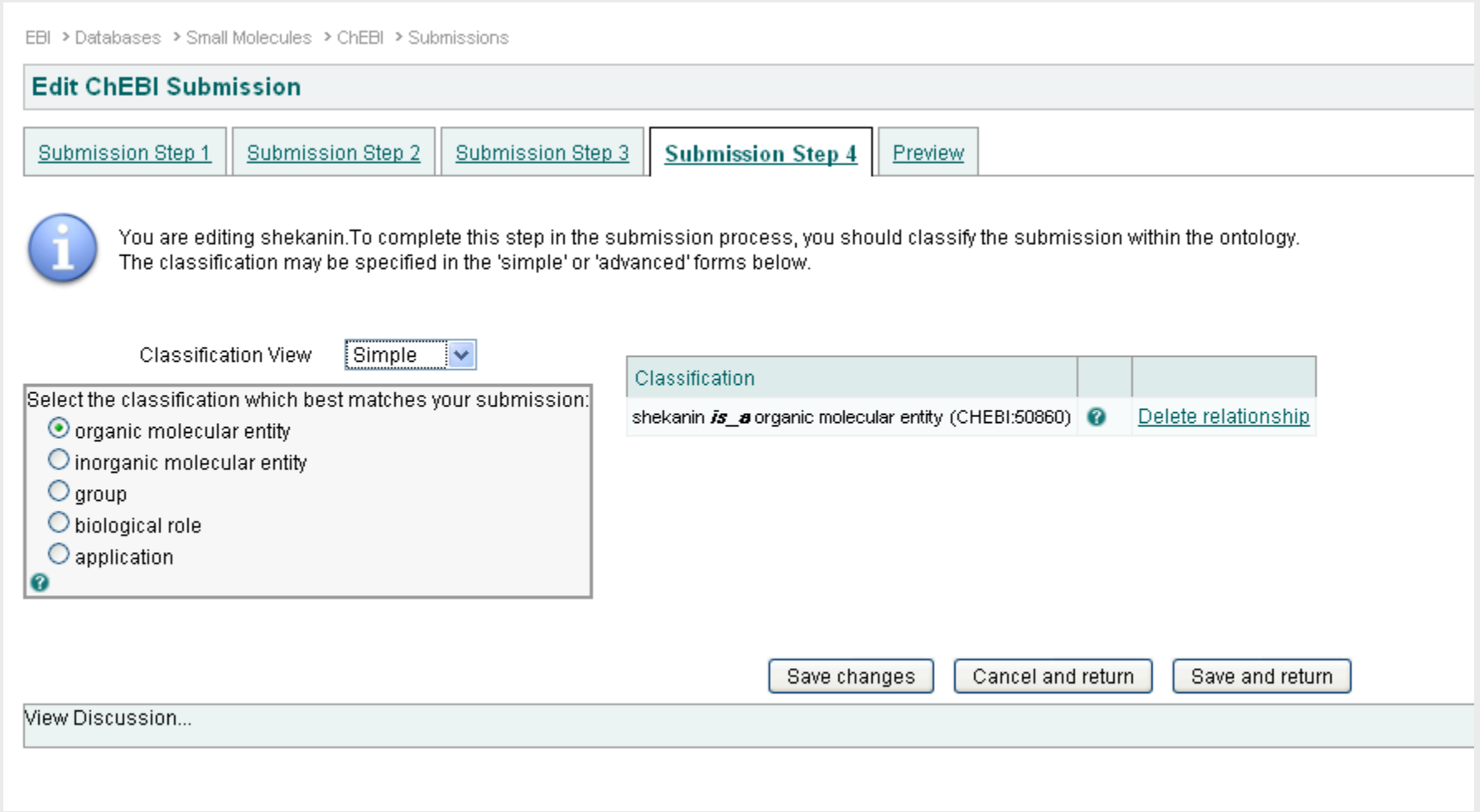
Submission Step 2



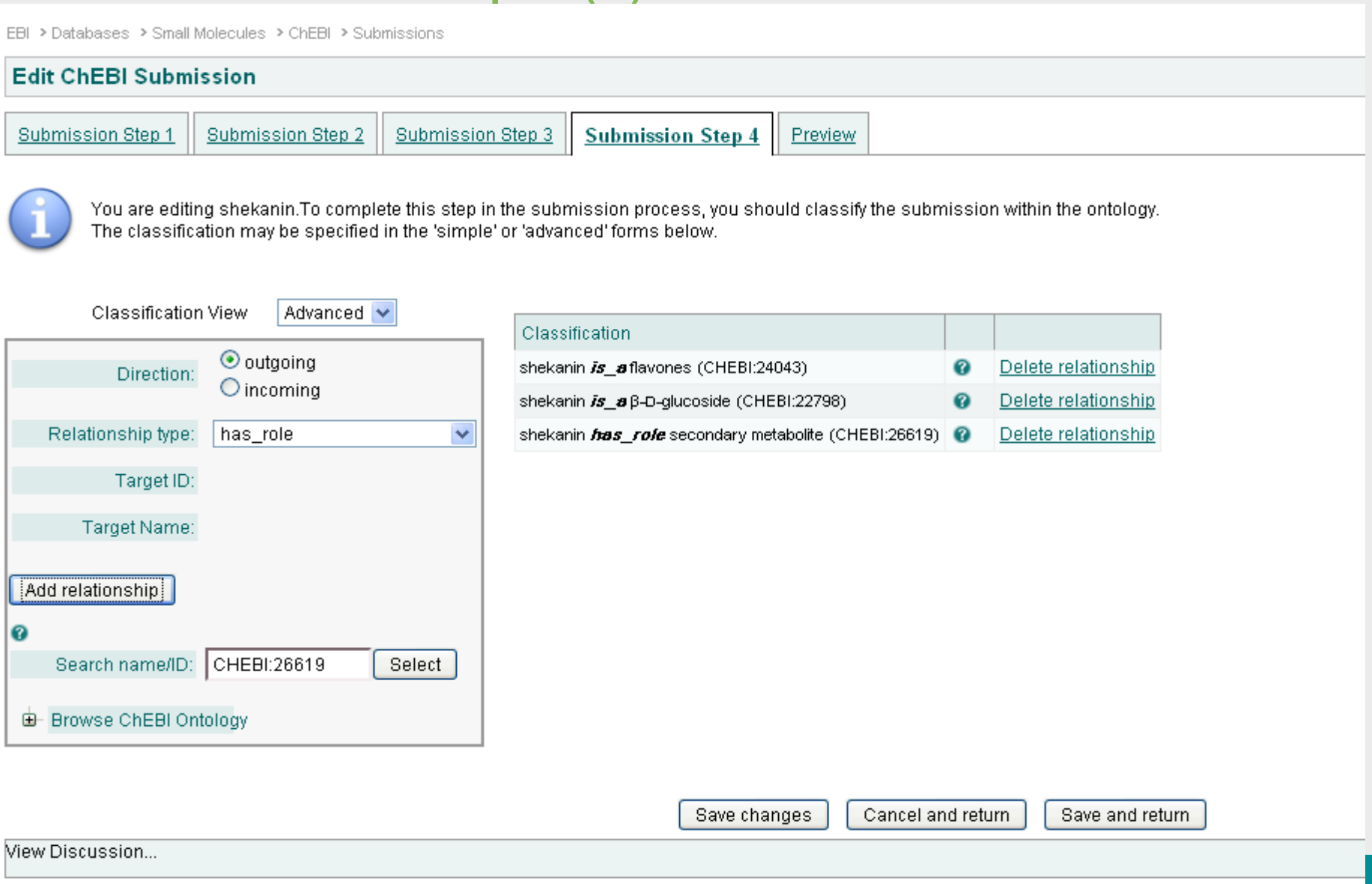
Submission Step 3



Submission Step 4(i)



Submission Step 4(ii)



Submission preview(i)

EBI > Databases > Small Molecules > ChEBI > Submissions Help | Logou

Edit ChEBI Submission

Submission Step 1 | Submission Step 2 | Submission Step 3 | Submission Step 4 | **Preview**

Submission of shekanin.

Review your submission and, if there are no changes required, submit by clicking 'submit'. Once you have submitted you will not be able to alter this entry, ChEBI annotators will review this submission and if it is deemed appropriate it will be included in a ChEBI release.

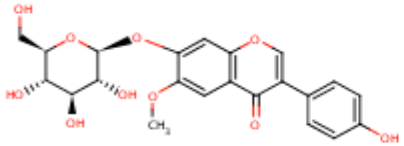
Final Submission

Date submission to be active

Please enter the six characters you can see in the image on the left, in sequence.

Can't read the image? Get a new code...

Submit Entry to ChEBI



ChEBI name	shekanin
Definition	The beta-D-glucoside of tectorigenin.
Formula	C22H22O10
InChI	InChI=1/C22H22O10 /c1-29-15-6-12-14(30-9-13(18(12)25)10-2-4-11(24)5-3-10) 7-16(15)31-22-21(28)20(27)19(26)17(8-23)32-22
InChIKey	InChIKey=OZBAVEKZGSOMOJ-MIUGBVLSBI

Submission preview(ii)

SMILES	COc1cc2c(cc1O[C@@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@H]1O)occ(-c1ccc(O)cc1)c2=O
Mass	446.40410
Charge	0

Cross-reference	Type
611-42-7	CAS Registry Number
68384	Beilstein Registry Number

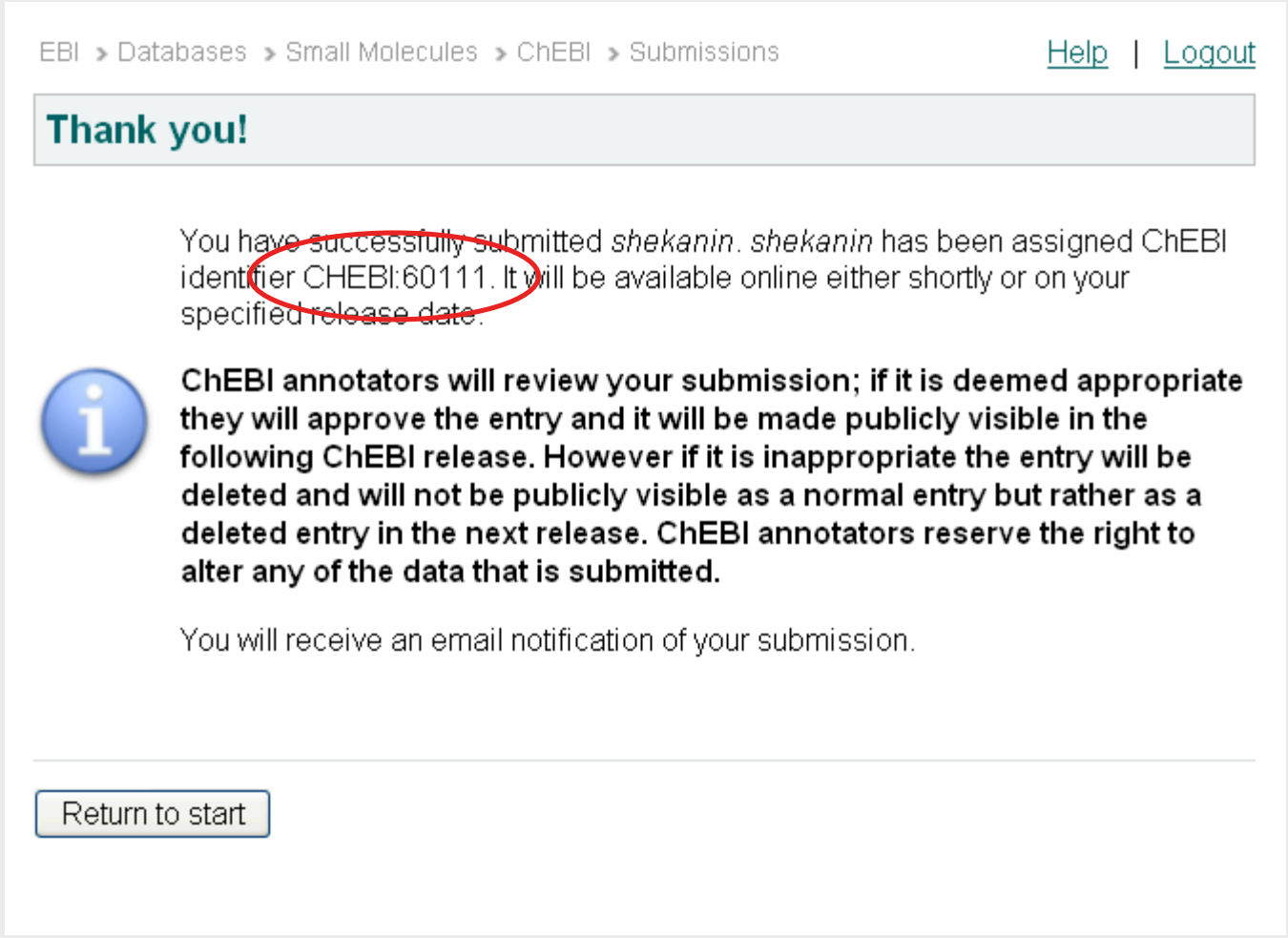
Synonym	Type	Language
Tectoridin	Type: SYNONYM Source: SUBMITTER	English
6-methoxy-4-oxo-3-phenyl-4H-chromen-7-yl beta-D-glucopyranoside	Type: IUPAC NAME Source: SUBMITTER	English

Classification
shekanin ***is_a*** flavones (CHEBI:24043)
shekanin ***is_a*** β-D-glucoside (CHEBI:22798)
shekanin ***has_role*** secondary metabolite (CHEBI:26619)

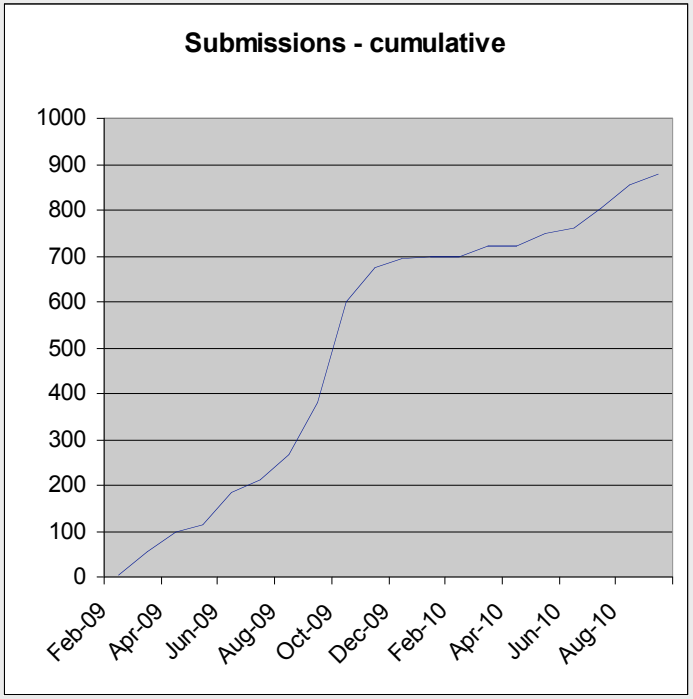
Save changes | Cancel and return | Save and return

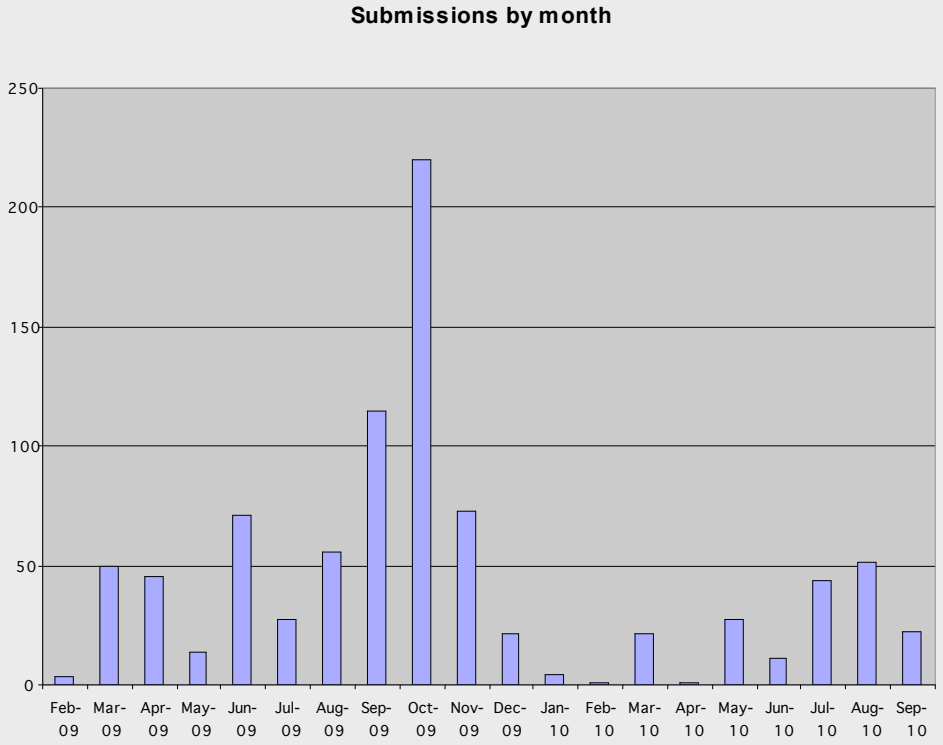
View Discussion...

Submission confirmation



ChEBI submissions Feb 2009 – Sep 2010

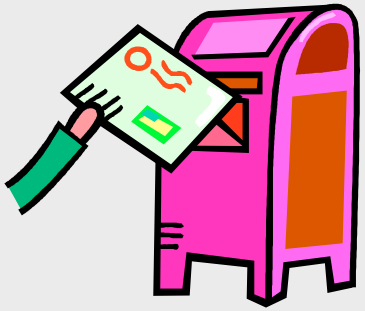




(Data as at September 9th 2010)

ChEBI submitters

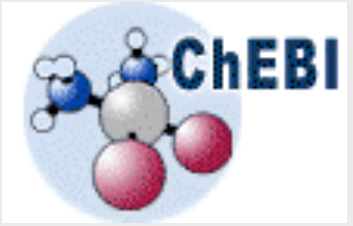
- Reactome
- Gene Ontology Consortium
- RSC Journals
- Rutgers Univ.
- Cornell Univ.
- Carleton Univ.
- Stanford Univ.
- SIB
- IntAct
- Rhea
- Array Express
- Manchester Centre for Integrative Systems Biology
- EML Research, Heidelberg
- Dept of Genetics, Cambridge Univ.



Submission tool - Future plans

- Bulk submissions (by invitation only) - imminent
- Modifications to existing ChEBI entries - ???

The ChEBI team



Curators

Marcus Ennis
Steve Turner
Zara Josephs
Nico Adams
Gareth Owen

Developers

Paula de Matos (team coordinator)
Janna Hastings
Adriano Dekker
Kenneth Haug

Group Leader

Christoph Steinbeck

www.ebi.ac.uk/chebi

ChEBI is funded by the European Commission under SLING, contract number 226073 within the Research Infrastructure Action of the FP7 "Capacities Specific" Programme.