

Section III

Cheminformatics - Basics: Searching for Chemicals on the Internet (*PubChem, eMolecules*)

Oleg Ursu, Dan C. Fara and Tudor I. Oprea

Objective

Introduction to mining PubChem database for chemicals structures and biological activities associated with the chemical structures.

1. PubChem

[PubChem](#) provides information on the biological activities of small molecules. It is a component of NIH's [Molecular Libraries Roadmap Initiative](#). PubChem includes substance information, compound structures, and bioactivity data in three primary databases, [PCSubstance](#), [PCCompound](#), and [PCBioAssay](#), respectively.

- PCSubstance contains more than 38 million records. You can [check the count](#) of substance records as of today.
- PCCompound contains more than 18 million unique structures. You can [check the count](#) of compound records as of today.
- PCBioAssay contains more than 680 bioassays. Each bioassay contains a various number of data points. You can [check the count](#) of bioassay records as of today.

The Substance/Compound database, where possible, provides links to bioassay description, literature, references, and assay data points. The BioAssay database also includes links back to the Substance/Compound database. PubChem is integrated with [Entrez](#), NCBI's primary search engine, and also provides compound neighboring, sub/superstructure, similarity structure, bioactivity data, and other searching features.

PubChem contains [substance](#) and [bioassay](#) information from a multitude of depositors. You can check the PubChem [data source status](#) as of today.

How to

- Go to <http://pubchem.ncbi.nlm.nih.gov/>
- Under PubChem Text Search section select PubChem Compound
- Type beta-estradiol and click Go

beta-estradiol - PubChem Compound Results - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/pubs/entrez/compound/term-beta-estradiol

NCBI PubChem

Search PubChem Compound for beta-estradiol

Limits Preview/Index History Clipboard Details

Display Summary Share 20 Sort by Send to

All: 3 BioAssay: 2 Protein3D: 1 Rule of 5: 3

Items 1 - 3 of 3

One page

Related Structures, Assays, Literature, Other Links

1: CID: 5757

Follisyclin; Oestradiol; nchembio775-comp2 ...
IUPAC: (8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decahydrocyclopenta[a]phenanthrene-3,17-diol
MW: 272.38196 | MF: C18H24O2

2: CID: 5317221

estradiol; beta-Estradiol; Spectrum_001048 ...
IUPAC: (13S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decahydrocyclopenta[a]phenanthrene-3,17-diol
MW: 272.382 | MF: C18H24O2

3: CID: 450

estradiol; beta-Estradiol; estradiol-17beta ...
IUPAC: 13-methyl-6,7,8,9,11,12,14,15,16,17-decahydrocyclopenta[a]phenanthrene-3,17-diol
MW: 272.38196 | MF: C18H24O2

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

- Click on the first CID (compound ID link [5757](#))

CID 5757 - PubChem Compound Summary - Mozilla Firefox

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?CID=5757

NCBI PubChem

Information on biological activities of small molecules

HOME SEARCH SITE MAP PubMed Entrez Structure GenBank PubChem Help

Search PubChem Compound

Compound Summary:

CID: 5757

BioActivity: Summary

All: 89 Links
Active: 6 Links
Inactive: 76 Links
Inconclusive: 6 Links

Protein Structures: 18 Links

Protein Sequences: 40 Links

NLM Toxicology:

Link1, Link2

Substances:

All: 152 Links
Same: 71 Links
Mixture: 81 Links

Related Compounds:

Same, Connectivity: 33 Links
Same, Stereochemistry: 4 Links
Same, Isotopes: 28 Links
Same, Any Tautomers: 35 Links

Similar Compounds: 1185 Links

Structure Search

MeSH Synonyms Properties Descriptors Category Exports

Medical Subject Annotations: (Total:2) Display: Next 1 | All

Estradiol

Generally refers to the 17-beta-isomer of estradiol, an aromatized C18 steroid with hydroxyl group at 3-beta- and 17-beta-position. Estradiol-17-beta is the most potent form of mammalian estrogenic steroids. In humans, it is produced primarily by the cyclic ovaries and the PLACENTA. It is also produced by the adipose tissue of men and postmenopausal women. The 17-alpha-isomer of estradiol binds weakly to estrogen receptors (RECEPTORS, ESTROGEN) and exhibits little estrogenic activity in estrogen-responsive tissues. Various isomers can be synthesized.

- The top of page opened contains the chemical structure of the Estradiol, Bioactivity Summary (assays in which the Estradiol was tested, found active, inactive and inconclusive), Protein Structures complexes with Estradiol, Substances associated with the chemical structure of Estradiol, Related Compounds and Similar compounds links to all tautomers, isomers, and similar chemical structures deposited in PubChem. For example, clicking on [Similar Compounds](#) will return a list of steroid like structures deposited in PubChem.

The screenshot displays the PubChem website interface in a Mozilla Firefox browser. The search results are for 'Similar Compounds for PubChem Compound (Select 5757)'. The search bar shows 'PubChem Compound' and 'for'. The results are displayed in a table with columns for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The table lists four items, each with a chemical structure, name, and identifiers.

Item	Chemical Structure	Name	Identifiers
1		Follicle-stimulating hormone (FSH)	PubChem CID: 5757
2		beta-Estradiol-d2; E4260_SIGMA; 1,3,5(10)-Estratriene-2,4-d2-3,17beta-diol	PubChem CID: 16219317
3		491187_ALDRICH; 17beta-Estradiol-16,16,17-d3	PubChem CID: 16213507
4		ZINC03978002	PubChem CID: 11870431

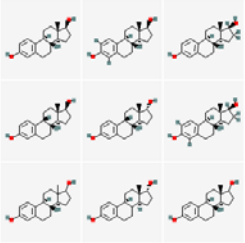
This list can be downloaded (select PubChem Download from display combo box) as SDF file which contains all the identification data as well as some chemical properties of the structures. It is also possible to take a quick look at all assays in which all of these structures were tested by selecting PubChemBioActivity Summary (see below).

PubChem BioAssay Summary - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay.cgi?query=33&status=done®id=634703838171703776

PubChem BioAssay Summary

BioActivity Summary:



Total BioAssay Count: 275

Total Substance Count: 66

Total Compound Count: 1185
(9 structures shown)

Revise Substance Selection:

- Select Active
- Add Active
- Add Tested
- Add Similar Substances

Revise BioAssay Selection:

- Select Active
- Add Active
- Add Tested
- Select BioAssays

Structure Activity Analysis

BioAssay Data Table

Structure Clustering

Selected BioAssays to Entrez

Result Display Option

Group Results By: Compound

Re-Show the Page

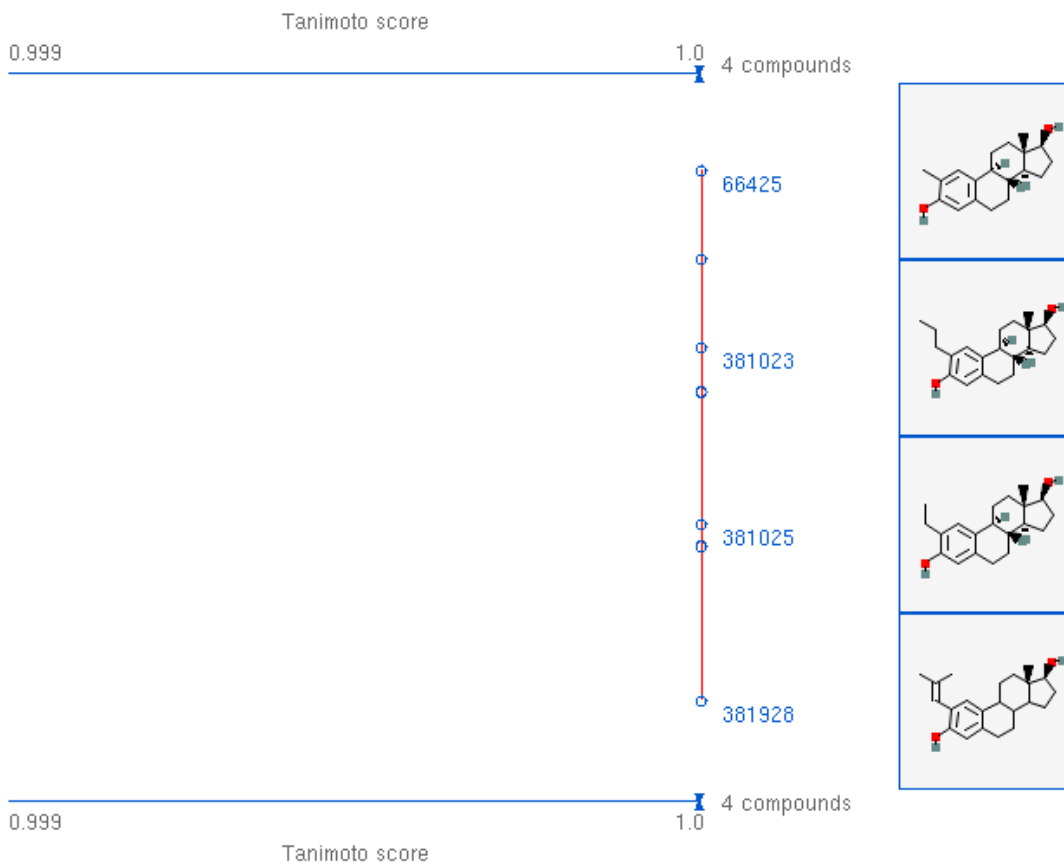
Total Pages: 3 Display: 100 Go To Page 1

#	ATD	Active	Inactive	Discrepant	Tested	Outcome Method	Name
1	161	10	13		23		NCI Yeast Anticancer Drug Screen. Data for the sgs1 mgf1 strain
2	354	5			9	Other	NCTR Estrogen Receptor Binding Database (NCTREB)
3	175	9	14		23		NCI Yeast Anticancer Drug Screen. Data for the mlh1 rad18 strain
4	165	9	14		23		NCI Yeast Anticancer Drug Screen. Data for the dln2 rad14 strain
5	167	8	15		23		NCI Yeast Anticancer Drug Screen. Data for the bub2 strain
6	155	8	15		23		NCI Yeast Anticancer Drug Screen. Data for the rad50 strain
7	157	7	16		23		NCI Yeast Anticancer Drug Screen. Data for the mac2-1 strain
8	85	5	13		18	Confirmatory	NCI human tumor cell line growth inhibition assay. Data for the MDA-MB-432 Breast cell line
9	87	4	14		18		NCI human tumor cell line growth inhibition assay. Data for the MDA-MB-432 Breast cell line

The page displays all assay in which chemical compounds similar to Estradiol were tested; assays in which any of the compounds were found to be active are listed first.

- Select only the assay in which at least one compound is active and then click Structure Activity Analysis. The page displayed contains summary information only for compounds and assays but one can easily notice that there are 4 compounds CIDs (66413, 5468270, 5468334, 388008) found active in multiple assays. The Compound Cluster tree displayed in the left area of the page is a useful tool for analyzing structure activity relationships. Click on the tree node of interest to display the compounds in a new web page.





All compounds in the cluster have a hydrophobic substituent at aromatic ring, 3 compounds are active in 2 assays, and 2 compounds are active in 4 assays.

- Return to the page where Estradiol is displayed (CID 5757) scroll to the bottom of the page and click Journal Publisher links.



Substance Category: [2](#)

[Biological Properties: 28 Links](#)
[Journal Publishers: 6 Links](#)
[Metabolic Pathways: 3 Links](#)
[Molecular Libraries Screening Center Network: 2 Links](#)
[NIH Substance Repository: 1 Link](#)
[Protein 3D Structures: 19 Links](#)
[Substance Vendors: 7 Links](#)
[Theoretical Properties: 1 Link](#)
[Toxicology: 3 Links](#)

ASN1	Display	XML	Display	SDF	Display
	Save		Save		Save

5757[scid] AND *Journal Publishers[SourceCategory] - PubChem Substance Results - Mozilla Firefox

Search: PubChem Substance for 5757[scid] AND *Journal Publishers[SourceCategory] Go Clear Save Search

Display: Summary Show 20 Sort by Rule of 5.6

Items 1 - 6 of 6

1: SID: 17436666

CID: 5757; nchembio660-comp1; (8S,9S,13S,14S,17S)-13-Methyl-6,7,8,9,11,12,14,15,16,17-decalhydrocyclopenta[α]phenanthrene-3,17-diol

Source: *Nature Chemical Biology*(nchembio660-comp1)

IUPAC: (8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decalhydrocyclopenta[α]phenanthrene-3,17-diol

MW: 272.38196 | MF: C18H24O2

2: SID: 14799367

CID: 5757

Source: *Thomson Pharma*(00013519)

IUPAC: (8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decalhydrocyclopenta[α]phenanthrene-3,17-diol

MW: 272.38196 | MF: C18H24O2

3: SID: 14799365

CID: 5757

Source: *Thomson Pharma*(00006586)

IUPAC: (8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decalhydrocyclopenta[α]phenanthrene-3,17-diol

MW: 272.38196 | MF: C18H24O2

4: SID: 12012600

CID: 5757; estradiol Estrace; 17beta-Estradiol ...

Source: *Prens Science Drugs of the Future*(70490)

IUPAC: (8R,9S,13S,14S,17S)-13-methyl-6,7,8,9,11,12,14,15,16,17-decalhydrocyclopenta[α]phenanthrene-3,17-diol

MW: 272.38196 | MF: C18H24O2

- Click on the last [link](#) to open the paper published in Nature Chemical Biology.

Virtual and biomolecular screening converge on a selective agonist for GPR30 - Nature Chemical Biology - Mozilla Firefox

http://www.nature.com/nchembio/journal/v2/n4/abs/nchembio775.html

nature.com Jump to main content Jump to navigation

REGISTER MY ACCOUNT SUBSCRIBE E-ALERT SIGN UP

PUBLICATIONS A-Z INDEX BROWSE BY SUBJECT SEARCH This journal go ADVANCED SEARCH

The ISME Journal First issue FREE online now!

Journal home > Archive > Letter > Abstract

Letter

Nature Chemical Biology 2, 207-212 (2006)
doi:10.1038/nchembio775

Virtual and biomolecular screening converge on a selective agonist for GPR30

Cristian G Bolaga^{1,2}, Chetana M Revankar^{2,3,7}, Susan M Young³, Bruce G Edwards^{3,4}, Jeffrey B Arterburn⁵, Alexander S Kiselyov⁶, Matthew A Parker⁶, Sergey E Tkachenko⁶, Nikolay P Savchuck⁶, Larry A Sklar^{4,5}, Tudor I Oprea⁴ and Eric R Prossnitz^{2,4}

Estrogen is a hormone critical in the development, normal physiology and pathophysiology¹ of numerous human tissues². The effects of estrogen have traditionally been solely ascribed to estrogen receptor α (ERα) and more recently ERβ, members of the soluble, nuclear ligand-activated family of transcription factors³. We have recently shown that the seven-transmembrane G protein-coupled receptor GPR30 binds estrogen with high affinity and resides in the endoplasmic reticulum, where it activates multiple intracellular signaling pathways⁴. To differentiate between the functions of ERα or ERβ and GPR30, we used a combination of virtual and biomolecular screening to isolate compounds that selectively bind to GPR30. Here we describe the identification of the first GPR30-specific agonist, G-1 (1), capable of activating GPR30 in a complex environment of classical and new estrogen receptors. The development of compounds specific to estrogen receptor family members provides the opportunity to increase our understanding of these receptors and their contribution to estrogen biology.

1. Division of Biocomputing, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
2. Department of Cell Biology & Physiology, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
3. Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
4. Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
5. Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
6. Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA;
7. Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico 87131, USA

Full text
Download PDF
Send to a friend
Report citation
Rights and permissions
Order commercial reprints
Bookmark in Connotea
Figures, schemas & tables
Supplementary info

SEE ALSO
News and Views by Weatherman

SEARCH PUBLISHED FOR
Cristian G Bolaga
Chetana M Revankar
Susan M Young
Bruce G Edwards
Jeffrey B Arterburn
Alexander S Kiselyov
more authors of this article

The paper describes how virtual screening can be applied to discover new compounds that bind selectively to GPR30.

2. eMolecules

[eMolecules](#) is one of the world's most comprehensive openly accessible search engine for chemical structures.

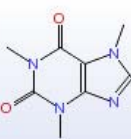
- The database contains around 8 million unique chemical structures from 19 million sources
- It allows unparalleled search speed—typically just seconds.
- The search can be done by drawing chemical structures or substructures using common industry tools — ISIS/Draw, ChemDraw, ChemSketch and JME.
- It has more than 4 million commercially available screening compounds and hundreds of thousands of building blocks and intermediates.
- It includes reference links to many prominent sources of public data for spectra, physical properties and biological data, including NIST WebBook, National Cancer Institute, DrugBank and PubChem.

The basic search can be done by (i) drawing chemical structures, and (ii) text, as can be seen below.

eMolecules

Find Suppliers and Information for over 7 Million Chemicals!

Draw a Structure



Search by Chemical Structure

[Draw Structure](#)

Drawing Tool:

☒ JME (default)

☐ ISIS Draw [\(read this\)](#)

☐ ChemDraw [\(read this\)](#)

Advanced Options:

☐ Commercial Compounds Only

☐ Specific Supplier [Choose](#)

or

Text Search

Search Named Chemicals

[Search](#)

Formats: Example:

- Trade Name [Advil](#)
- INN [Ibuprofen](#)
- CAS Num [15687-27-1](#)
- SMILES [S=C=NC](#)

[more info...](#)

Advanced Options:

☐ Commercial Compounds Only

☐ Specific Supplier [Choose](#)

Or, browse by category:

Purines	Betalactams	Isothiocyanates
Sulfonamides	Phosphorus Compounds	Carbohydrates
Triazines	Isonitriles	Benzodiazepines

- Search the eMolecules database for beta estradiol
- Click on the PubChem Supplier's ID (10468968)

Are you re-directed to the same compound as before (CID 5757)? What is the difference between these two compounds?