





**Automatically adding chemistry to Web pages
& PDF's, predicted structure data & search.**






Alex Allardyce
ChemAxon

aspirin  

-  Webpage Viewer
-  Properties Viewer »
-  Chem Search »
-  Web Search »
-  Document Viewer

What it's for

1. Identify and see chemical structures on webpages, discrete text and PDF files
2. Calculate and predict structure properties
3. Structure search structures chemicalize has already met
4. Search the web using chemical and non chemical terms

 Webpage Viewer
 Properties Viewer »
 Chem Search »
 Web Search »
 Document Viewer

Let's see it – chemicalize a webpage

W Penicillin - Wikipedia, th x

www.chemicalize.org?url=http://en.wikipedia.org/wiki/Penicillin

upload Webpage Viewer Original version Unglue from top Download structures

106 (3) 1 23 (3) 6 7 8 (3) 2 2 2 3 (2) 2

Article Discussion Read Edit View history Search

Penicillin

From Wikipedia, the free encyclopedia

For the Japanese band, see [Penicillin \(band\)](#).

Penicillin (sometimes abbreviated **PCN** or **pen**) is a group of antibiotics derived from *Penicillium* fungi.^[1] They include [penicillin G](#), [procaine penicillin](#), [benzathine penicillin](#), and [penicillin V](#). Penicillin antibiotics are historically significant because they are the first drugs that were effective against many previously serious diseases such as [streptococci](#). Penicillins are still widely used today, though many t antibiotics and are used in the treatment of bacterial infections ca.

Penicillin staphylococci and penicillins are beta-lactam sensitive, organisms.

Penicillin core structure. The "R" is the variable group.

Contents [hide]	
1	Medical uses
2	Adverse effects
3	Mechanism of action
4	Structure
5	Biosynthesis
6	Production
7	History
7.1	Discovery
7.2	Medical application
7.3	Mass production
7.4	Unethical experimentation
7.5	Developments from penicillin
8	See also
9	References

NEW

... chemicalize a PDF file

The screenshot shows a web browser window with the URL www.chemicalize.org/document/#u=http%3A%2F%2Fwww.chemaxon.com%2Fwp-content%2Fuploads%2F2011%2F09%2F1758-2946-3-S1-O3.pdf. The page features a navigation bar with chemical structures and names: isatin, Mu₂, N-ethyl isatin, indole, oxide, isatin hydrazone, 3-(methylene)in..., EtOH, and bromoethane. The main content area includes a 'RESEARCH ARTICLE' header with an 'Open Access' badge, the title 'Simple isatin derivatives as free radical scavengers: Synthesis, biological evaluation and structure-activity relationship', authors 'Gang Chen¹, Ye Wang², Xiaojun Sun¹, and Jianyun Sun²', and an abstract. The abstract text is partially obscured by a chemical structure tooltip for isatin. The background text of the abstract is: 'To develop more potent small molecule radical scavengers, a series of *N*-substituted isatin derivatives was synthesized and screened. All these compounds showed a protective effect on the apoptosis of PC12 cells induced by H₂O₂. The *N*-ethyl isatin was found with the most potent activity of 69.7% protective effect. Structure-activity relationship analyses showed the bioactivity of *N*-alkyl isatins decline as the increasing of the chain of the alkyl group, furthermore odd-even effect was found in the activity, which is interesting for further investigation.'

Abstract

To develop more potent small molecule radical scavengers, a series of *N*-substituted isatin derivatives was synthesized and screened. All these compounds showed a protective effect on the apoptosis of PC12 cells induced by H₂O₂. The *N*-ethyl isatin was found with the most potent activity of 69.7% protective effect. Structure-activity relationship analyses showed the bioactivity of *N*-alkyl isatins decline as the increasing of the chain of the alkyl group, furthermore odd-even effect was found in the activity, which is interesting for further investigation.

Background

Oxidative stress has been implicated as a major role in the onset and progression of a vast variety of clinical abnormalities including neurodegenerative disorders. Free radicals play important roles in many physiological and pathological conditions [1]. To reveal the mechanism of oxidative stress and its relationship with the onset of neurodegenerative diseases, we synthesized a series of *N*-substituted isatin derivatives and investigated their antioxidant properties [13-15]. Isatin has a distinct and discontinuous distribution in rat brain and other tissues; the highest concentrations in the brain are found in the hippocampus and cerebellum [10]. Many isatin derivatives, such as isatin hydrazone, isatin Mannich bases, isatin based Mannich bases and 3-substituted isatin Mannich bases have been reported [16-18].



... do sg with discrete text

chemicalize.org^{beta} by ChemAxon

Home Blog Extend Feedback & Contact Changelog About

draw upload Webpage Viewer

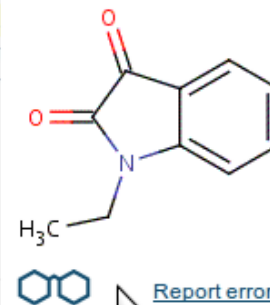
Original

To develop more potent small molecules with enhanced free radical scavenger properties, a series of N-substituted isatin derivatives was synthesized, and the cytoprotective effect on the apoptosis of PC12 cells induced by H2O2 was screened. All these compounds were found to be active, and N-ethyl isatin was found with the most potent activity of 69.7% protective effect on PC12 cells. Structure-activity relationship analyses showed the bioactivity of Nalkyl isatins decline as the increasing of the chain of the alkyl group, furthermore odd-even effect was found in the activity, which is interesting for further investigation.

chemicalized

To develop more potent small molecules with enhanced free radical scavenger properties, a series of N-substituted **isatin** derivatives was synthesized, and the cytoprotective effect on the apoptosis of PC12 cells induced by H2O2 was screened. All these compounds were found to be active, and **N-ethyl isatin** was found with the most potent activity of 69.7% protective effect on PC12 cells. Structure-activity relationship analyses showed the bioactivity of Nalkyl **isati** derivatives decline as the increasing of the chain of the alkyl group, furthermore odd-even effect was found in the activity, which is interesting for further investigation.

N-ethyl isatin



Useful to check missing structures lost because of html tags etc (copy paste through textpad)



... predict structure properties

penicillin - chemicalize.org

www.chemicalize.org/structure/#1mol=Penicillin

chemicalize.org^{beta} by ChemAxon **Loaded in 421ms**

Home Blog Extend Feedback & Contact Changelog About

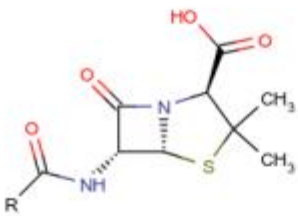
Penicillin   Properties Viewer

Discuss our future plans! Extending the datagap with QSPR / ADMET c...

JChem for Excel Chemistry in Excel - Structure handling and visualizing capabilities within a Microsoft Excel® environment. [Try it now!](#)

Manage calculations Layout: Medicinal Chemist Open All Close All Download results

Molecule



logP

logP: -0.91



Polar Surface Area

Polar surface area: 86.71

Molecular Surface Area

Solvent accessible surface area: 340.40

Similar Structures

0.977  0.977 

Names and Identifiers

Common names: penicillin

Smiles: CC1(C)S[C@@H]2[C@H](NC(=O)C(=O)N2[C@H]1C(=O)O

Elemental Analysis

Formula: C₉H₁₁N₂O₄S
Isotope formula: C₉H₁₁N₂O₄S
Composition: C (44.44%), H (4.56%), N (11.52%), O (26.31%), S (13.18%)
Isotope composition: C (44.44%), H (4.56%), N (11.52%), O (26.31%), S (13.18%)
Mass: 243.26
Exact mass: 243.04395254

Topology Analysis

Simple Ring Counts Path and distance

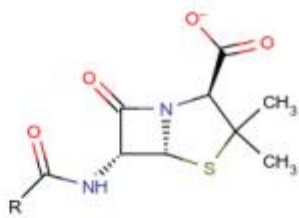
Atom count: 27
Bond count: 29
Cyclomatic number: 2
Chain atom count: 10
Chain bond count: 10
Asymmetric atom count: 3
Rotatable bond count: 2

Webpages

Penicillin - Wikipedia, the free encyclopedia


Major Microspecies

Major microspecies at pH=7.4:



Geometry

Dreiding energy = 89.66 kcal/mol
Volume = 225.31 Å³
Minimal projection area = 39.26 Å²
Min z length = 12.21 Å
Maximal projection area = 66.81 Å²
Max z length = 6.95 Å



Specialties 1- input box

Type a chemical name or URL to begin

penici|



penicillin



Penicillamine



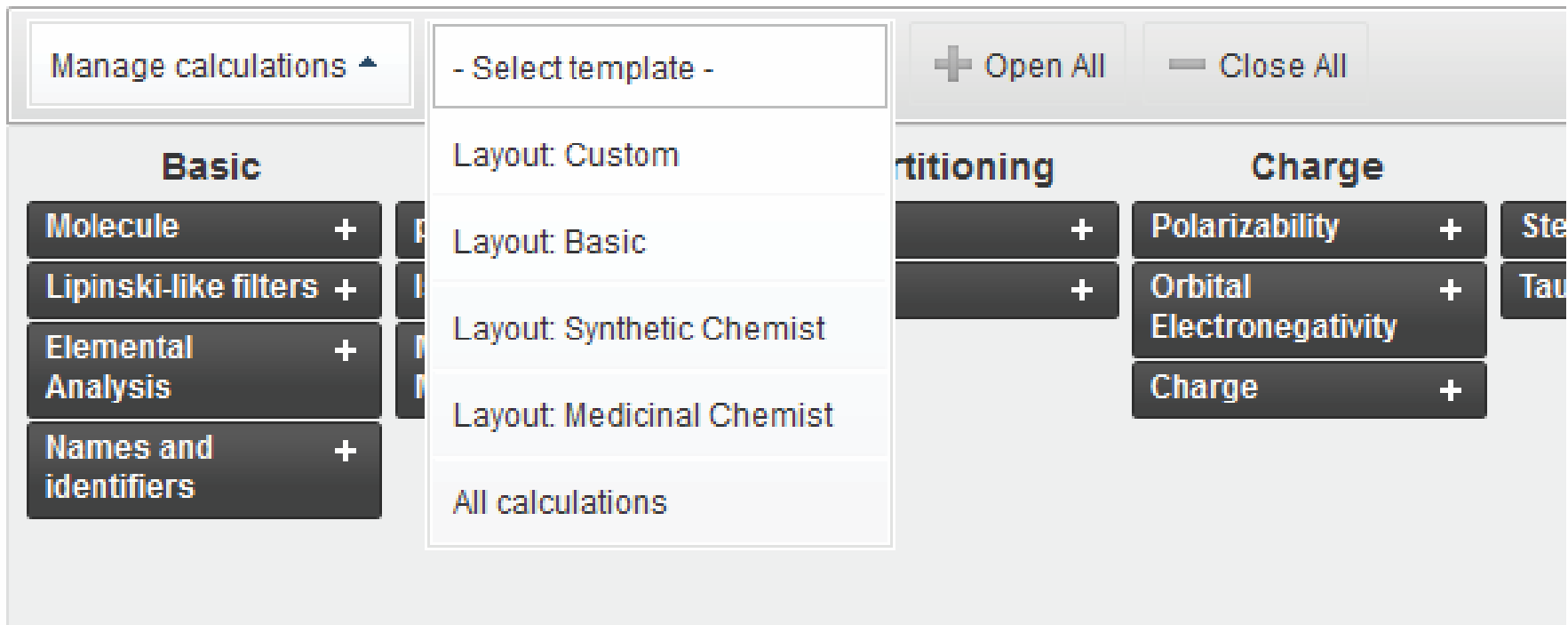
Penicillin G



Penicillin V

Structure name
autosuggest with
visualizations

Specialties 2 – configure views

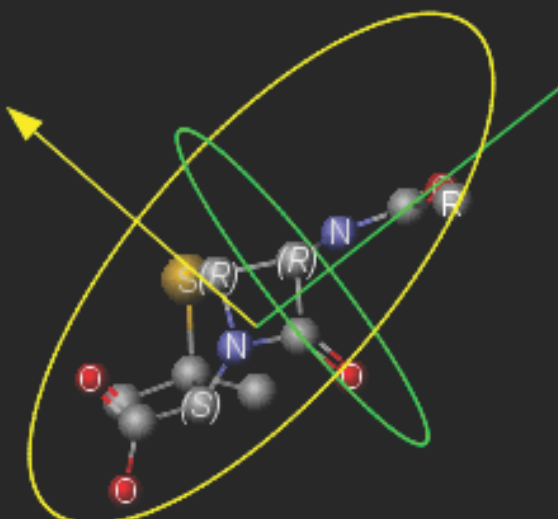


The screenshot displays the 'Manage calculations' interface. On the left, a 'Basic' section contains four expandable items: 'Molecule', 'Lipinski-like filters', 'Elemental Analysis', and 'Names and identifiers', each with a '+' icon. A central dropdown menu is open, showing options: '- Select template -', 'Layout: Custom', 'Layout: Basic', 'Layout: Synthetic Chemist', 'Layout: Medicinal Chemist', and 'All calculations'. On the right, there are two buttons: '+ Open All' and '- Close All'. Below these, there are sections for 'Partitioning' and 'Charge'. The 'Partitioning' section has two expandable items with '+' icons. The 'Charge' section has three expandable items: 'Polarizability', 'Orbital Electronegativity', and 'Charge', each with a '+' icon. To the right of the 'Charge' section, parts of other sections are visible, including 'Ste' and 'Tau'.

Template layouts, manage layouts, drag boxes.
 Session remembered on next visit

Specialties 3 – viewing structures

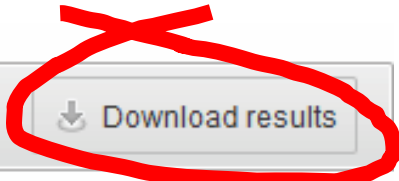
MarvinView - Geometry Block... ✕



- Transform ▶
- Animation ▶
- Display ▶
- Colors ▶
- Stereo ▶
- Implicit Hydrogens ▶
- Misc ▶
- Edit ▶
- Structure ▶
- Save Selection... Ctrl-S
- Save All... Ctrl+Shift-S
- Print ▶
- Open in MarvinView2D
- Help ▶

Opens in MarvinView

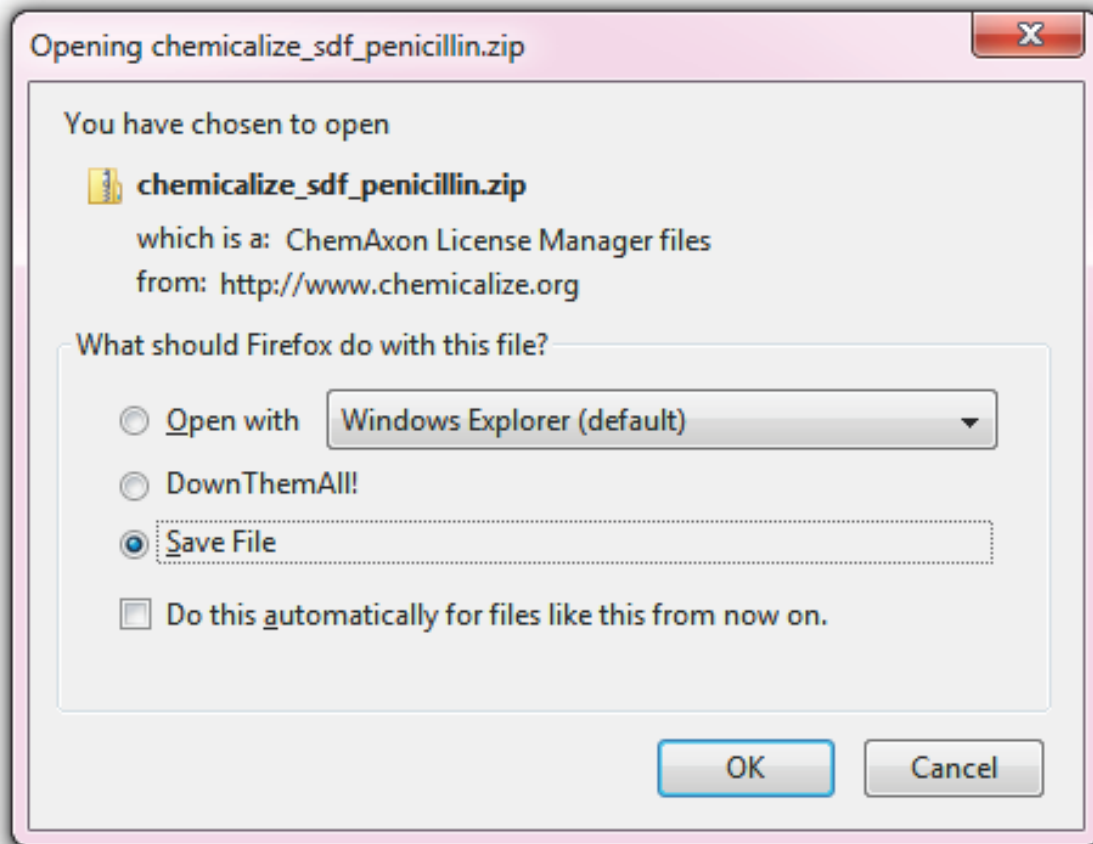
Specialties 4 - download










Download results

Structure and all
generated data

or all structures in a
document if using
webpage or document
viewer



How about structure search

-  All structures chemicalized are stored and can be structure searched
-  Uses ChemAxon's structure search: JChem WebServices:
 -  substructure, similarity, exact
 -  canonicalization of queries
-  Database (Oct 17, 2011):
 -  177,000 structures, (swelling soon)
 -  283,000 names extracted

Lets create a query

Type a chemical name or URL to begin

structureId:1210138318996

draw upload

File Edit View Insert Atom Bond Structure Tools Help

100%

Paste Ctrl-V
Select All Ctrl-A
Group... Ctrl-G
Attribute
Check Structure Ctrl-R
Transformation
Format...
Help Contents
About Marvin Sketch

Webpage Viewer
Properties Viewer »
Chem Search »
Web Search »
Document Viewer

C
N
O
S
F
P
Cl

Input via;
MarvinSketch structure editor, name, SMILES, InChI, ...
upload query file, drag structure to input box

Query building

Very powerful query options

(MDL and Daylight logic can be combined)

Periodic System ✕

Periodic Table **Advanced**

Description

Generic query atoms

A	Q	M	X
AH	QH	MH	XH

Atom query properties

.H+	.v+	.X+	.R+	.r+	.rb+	.s+	.h+	.D+	.u
.H-	.v-	.X-	.R-	.r-	.rb-	.s-	.h-	.D-	.a/A

Periodic Table Groups

G1	G2	G3	G4	G5	G6	G7	G8	G9
G10	G11	G12	G13	G14	G15	G16	G17	G18

Special nodes

LP	Pol	*
----	-----	---

alkyl ▾

R-groups

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16
R17	R18	R19	R20	R21	R22	R23	R24	R25	R26	R27	R28	R29	R30	R31	R32

Custom Property

Type:

Value:

Query results 1 – result pool

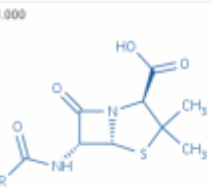
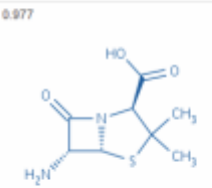
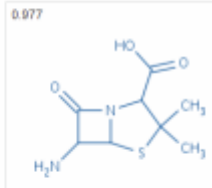
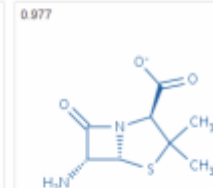
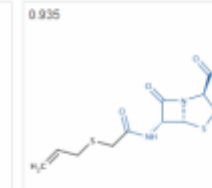
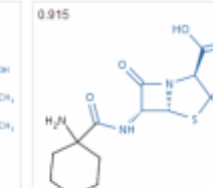
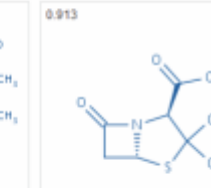
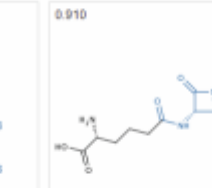
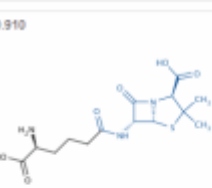
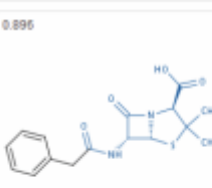
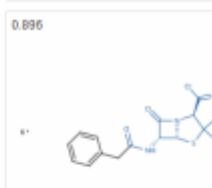
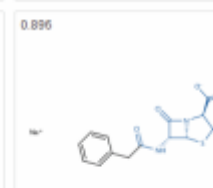
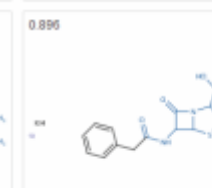
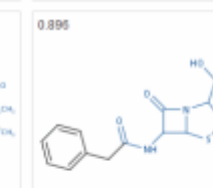
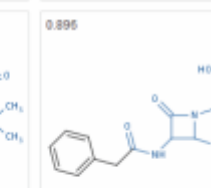
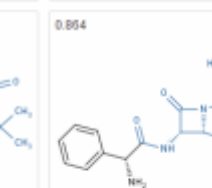
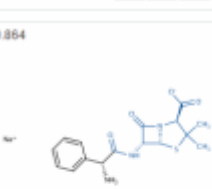
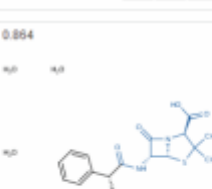
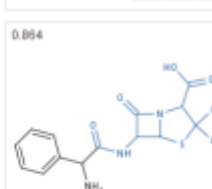
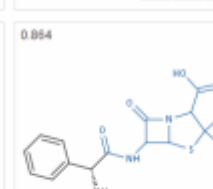
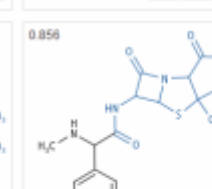
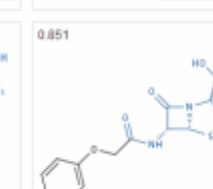
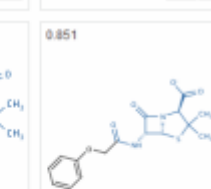
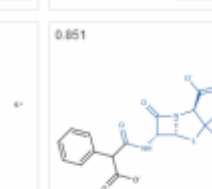
chemicalize.org^{beta} by ChemAxon

Home Blog Extend Feedback & Contact Changelog About

penicillin

Instant JChem Create, explore and share chemical and non chemical data in local and remote databases, on your desktop. [Try it now!](#)

235 structures found:

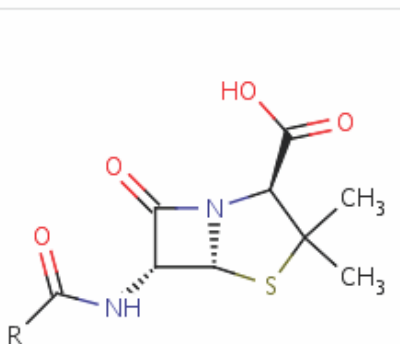
<p>1.000</p>  <p>penicillin</p> <p>Calculate Search</p> <p>Pages: 2055 SSS EKA SIM</p>	<p>0.977</p>  <p>aminopenicillanic acid</p> <p>Calculate Search</p> <p>Pages: 74 SSS EKA SIM</p>	<p>0.977</p>  <p>penin</p> <p>Calculate Search</p> <p>Pages: 34 SSS EKA SIM</p>	<p>0.977</p>  <p>penicillin G</p> <p>Calculate Search</p> <p>Pages: 3 SSS EKA SIM</p>	<p>0.935</p>  <p>penicillin AT</p> <p>Calculate Search</p> <p>Pages: 5 SSS EKA SIM</p>	<p>0.915</p>  <p>cycloillin</p> <p>Calculate Search</p> <p>Pages: 69 SSS EKA SIM</p>	<p>0.913</p>  <p>penicilanic acid</p> <p>Calculate Search</p> <p>Pages: 15 SSS EKA SIM</p>	<p>0.910</p>  <p>penicillin N</p> <p>Calculate Search</p> <p>Pages: 15 SSS EKA SIM</p>
<p>0.910</p>  <p>isopenicillin N</p> <p>Calculate Search</p> <p>Pages: 12 SSS EKA SIM</p>	<p>0.896</p>  <p>penicillin G</p> <p>Calculate Search</p> <p>Pages: 498 SSS EKA SIM</p>	<p>0.896</p>  <p>penicillin G potassium</p> <p>Calculate Search</p> <p>Pages: 28 SSS EKA SIM</p>	<p>0.896</p>  <p>penicillin G sodium</p> <p>Calculate Search</p> <p>Pages: 37 SSS EKA SIM</p>	<p>0.896</p>  <p>cillin</p> <p>Calculate Search</p> <p>Pages: 2 SSS EKA SIM</p>	<p>0.896</p>  <p>phenacillin</p> <p>Calculate Search</p> <p>Pages: 8 SSS EKA SIM</p>	<p>0.896</p>  <p>phenacillin</p> <p>Calculate Search</p> <p>Pages: 3 SSS EKA SIM</p>	<p>0.864</p>  <p>ampicillin</p> <p>Calculate Search</p> <p>Pages: 973 SSS EKA SIM</p>
<p>0.864</p>  <p>ampicillin sodium</p> <p>Calculate Search</p> <p>Pages: 26 SSS EKA SIM</p>	<p>0.864</p>  <p>principin</p> <p>Calculate Search</p> <p>Pages: 45 SSS EKA SIM</p>	<p>0.864</p>  <p>azim</p> <p>Calculate Search</p> <p>Pages: 255 SSS EKA SIM</p>	<p>0.864</p>  <p>ampicillin</p> <p>Calculate Search</p> <p>Pages: 2 SSS EKA SIM</p>	<p>0.856</p>  <p>ampicillin</p> <p>Calculate Search</p> <p>Pages: 1 SSS EKA SIM</p>	<p>0.851</p>  <p>phenoxymethylpenicillin</p> <p>Calculate Search</p> <p>Pages: 581 SSS EKA SIM</p>	<p>0.851</p>  <p>penicillin vk</p> <p>Calculate Search</p> <p>Pages: 19 SSS EKA SIM</p>	<p>0.851</p>  <p>gepotin</p> <p>Calculate Search</p> <p>Pages: 99 SSS EKA SIM</p>

%age similar - query highlighted (blue) in hits

Query result 2 – discrete structure

[« Back to search](#)

Structure found on 2055 pages:



[Calculate](#)

[Search](#)

SSS

EXA

SIM

Structure appears as
penicillin

[show all related structures >](#)

Diverse selection

[show all related links >](#)

[Penicillin - Wikipedia](#)

4 Sep 2011 - original page

sv.wikipedia.org/w/index.php?title=Penicillin

+ show structures on this page (9) + more results from sv.wikipedia.org

[Penicilin - Wikipedia, slobodna enciklopedija - ??????????, ????????? ??????????](#)

4 Sep 2011 - original page

sh.wikipedia.org/w/index.php?title=Penicilin

+ show structures on this page (22) + more results from sh.wikipedia.org

[Rheumatic fever - Wikipedia, the free encyclopedia](#)

4 Sep 2011 - original page

en.wikipedia.org/wiki/Rheumatic_fever

+ show structures on this page (10) + more results from en.wikipedia.org

[Penicillium chrysogenum Wikipedia](#)

4 Sep 2011 - original page

de.wikipedia.org/wiki/Penicillium_chrysogenum

+ show structures on this page (1) + more results from de.wikipedia.org

[User contributions for Mcstrother - Wikimedia Commons](#)

4 Sep 2011 - original page

commons.wikimedia.org/wiki/Special:Contributions/Mcstrother

+ show structures on this page (4) + more results from commons.wikimedia.org

[Staphylococcus aureus - Wikipedia, the free encyclopedia](#)

2 Sep 2011 - original page

en.wikipedia.org/wiki/Staphylococcus_aureus

+ show structures on this page (37) + more results from en.wikipedia.org

[Neuroprotection with Beta-Lactam Compounds - Johns Hopkins University](#)

1 Sep 2011 - original page

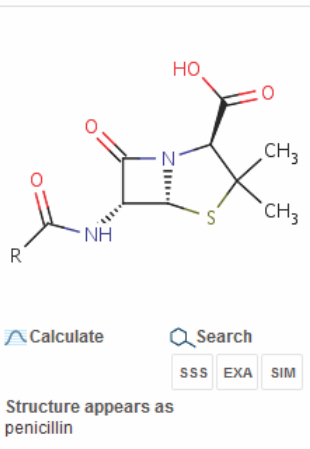
freepatentsonline.com/y2007/0238717.html

+ show structures on this page (84) + more results from freepatentsonline.com

Web pages we
have met that
contain this
structure

Query result 2 – discrete structure

« Back to search Structure found on 2055 pages:

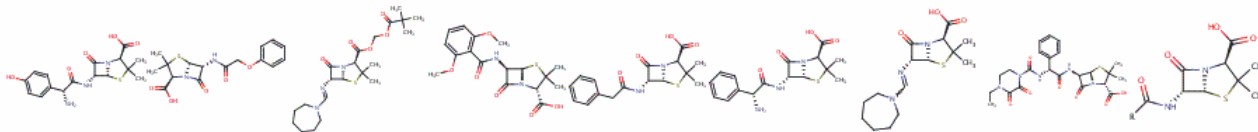


hide all related structures <
 Diverse selection

show all related links >

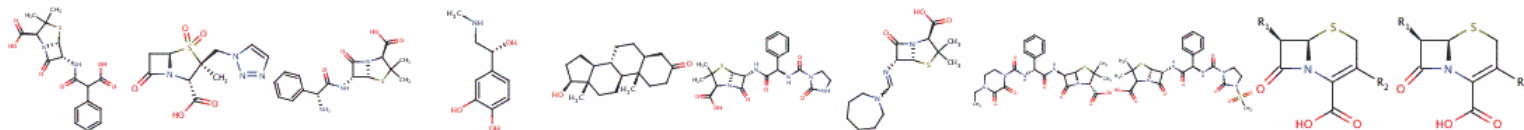
Penicillin - Wikipedia

4 Sep 2011 - original page
sv.wikipedia.org/w/index.php?title=Penicillin
- hide structures on this page (9) + more results from sv.wikipedia.org



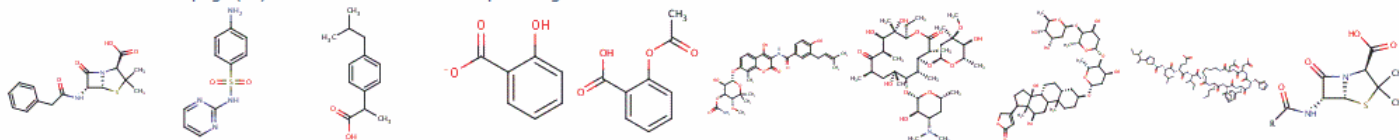
Penicilin - Wikipedia, slobodna enciklopedija - ??????????, ????????? ??????????...

4 Sep 2011 - original page
sh.wikipedia.org/w/index.php?title=Penicilin
- hide structures on this page (22) + more results from sh.wikipedia.org



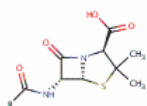
Rheumatic fever - Wikipedia, the free encyclopedia

4 Sep 2011 - original page
en.wikipedia.org/wiki/Rheumatic_fever
- hide structures on this page (10) + more results from en.wikipedia.org



Penicillium chrysogenum ^{GB} Wikipedia

4 Sep 2011 - original page
de.wikipedia.org/wiki/Penicillium_chrysogenum
- hide structures on this page (1) + more results from de.wikipedia.org



User contributions for Mcstrother - Wikimedia Commons

4 Sep 2011 - original page

Same result but
showing structures on
the pages

Web search – what's the difference

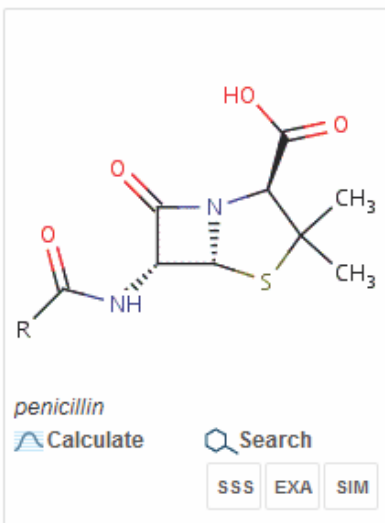
- We don't search the chemicalize structure database we send text search to Google

Generated query: "penicillin" OR "CC1(C)S[C@@H]2[C@H](NC([*])=O)C(=O)N2[C@H]1C(O)=O" endothelial

- You can combine chemical names with non chemical text
- It's a bit slower to complete because we chemicalize the Google results in the background
- Searches much wider space than chemicalize structure db, so more results ;)

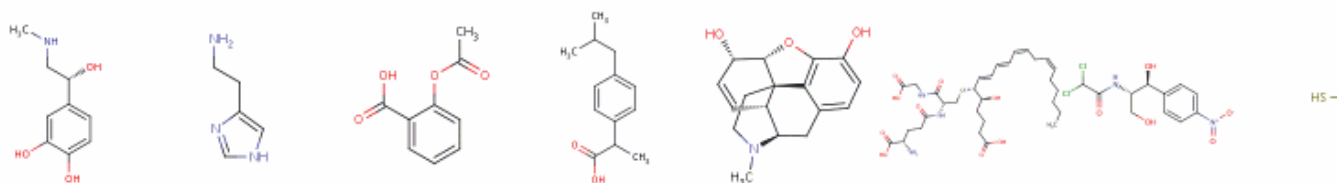
WebSearch result

81800 results found:



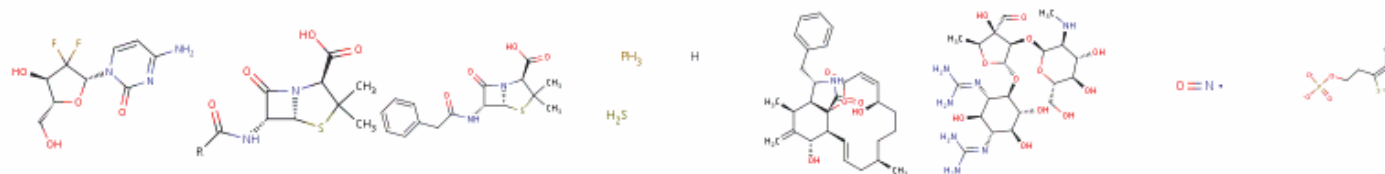
[Anaphylaxis - Wikipedia, the free encyclopedia](#)

The importance of TNF- α is most noted in the activation of the endothelium. they are exposed to one or more allergens, especially penicillin and insect stings. ... [en.wikipedia.org](#) (15)



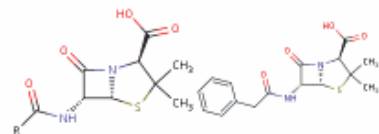
[Penicillin G-induced microbicidal activity of endothelial cells cultured ...](#)

Penicillin G-induced microbicidal activity of endothelial cells cultured on gelfoam blocks. Zhang B, Centra M, Cao GL, Taylor RM, Ratych RE, Rosen GM. ... [www.ncbi.nlm.nih.gov](#) (8)



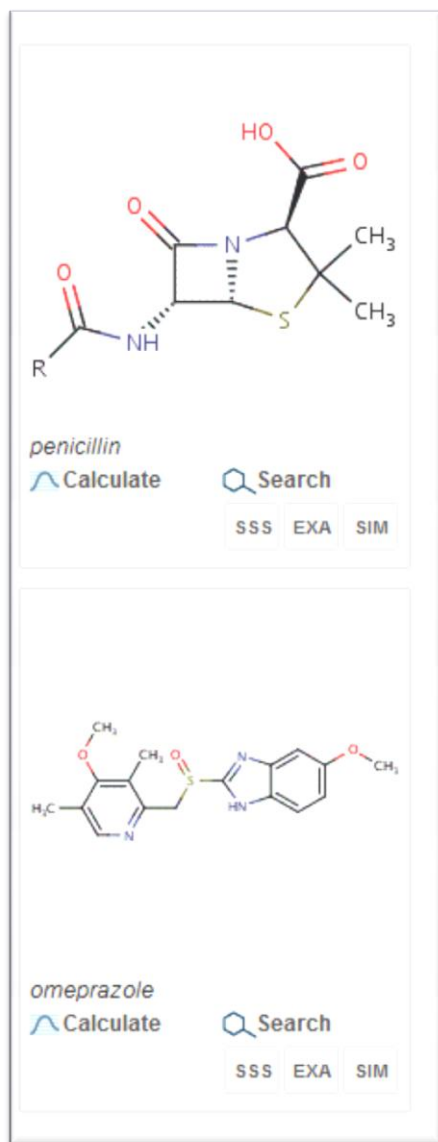
[Penicillin G-Induced Microbicidal Activity of Endothelial Cells ...](#)

Penicillin G, included in the endothelial cell growth medium, was found to be a critical factor in the bactericidal activity demonstrated by. Gelfoam blocks laden ... [www.istor.org](#) (2)



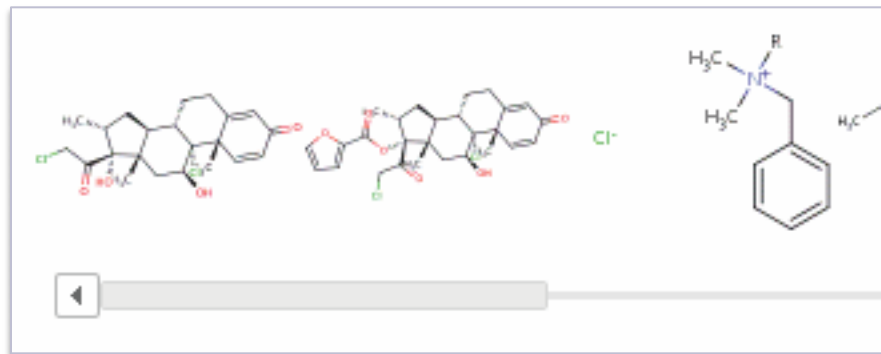
Structures shown
by default

Specialties - snippets



penicillin
Calculate Search
SSS EXA SIM

omeprazole
Calculate Search
SSS EXA SIM



Scroll bar to navigate chemistry
on pages

penicillin site:www.freepatentsonline.com

Google search features supported

Multiple structures
in same query

Why did we do this

History

- Free academic package and FreeWeb licensing since 2005
- Marvin free for all desktops (since the beginning)
- Open support forum developed to allow support for free users (no login to see all threads)
- Historically we are Java based so already very 'webby'

So, why did we do this

- There is a lot of content on the web (understatement)
- Is useful + increase visibility/utility of chemical structures
- Creates user interest in this type of functionality and so demand for chemistry and ‘clever’ content for publishers
- Lets us develop directly with end users:
 - Functionality/feature development
 - GUI usability
 - Crowd sourced bug fixing “Report Error” for naming.
- Pushing state of the art
 - Browser tech (svg, chunking, reducing calls)
 - ChemAxon tech (on the web, must be superfast, finalise features)
- We love cheminformatics “cheminfomaniacs”

ChemAxon bits

- [Marvin](#): structure editor, viewer, image generation
- [Name <> structure, Document to Structure](#): parsing, dictionaries, correction and lexing IUPAC names
- [JChem Base](#), [JChem Web Services](#), [Standardizer](#), [MCES](#): structure database, duplicate checking, structure search, web services layer, canonicalization, hit highlighting
- [Calculator Plugins](#): structure based predictions like pKa, logP, logD, charge, HBDA, tautomer, stereoisomers, etc.
Notable combined predictions yield argument results – like “Lipinski-likeness” etc

Patent based use case – Sourcing novel bio activity data

1. Search for new and relevant patents at freepatents-online.com, with target related text search – ie “alzheimer’s”
 2. Copy URL’s for relevant novel patent(s) and paste into chemicalize.org (web page viewer)
 3. Review chemicalize structure TOC to ‘see’ the chemistry in the patent
 4. If relevant, download all structures in the document as SDF
 5. Upload all structures to PubChem and run as single query
 6. Review/capture bioactivity data output
- If too many hits (5) then filter non drug like from the SDF and re-run

Other general use cases

- Wanted to know the logP of...
- What are the structures for known drugs
(http://en.wikipedia.org/wiki/List_of_drugs)
- Seeing structures in relation to a chemical name
- All wikipedia pages with a “chembox” have been indexed by chemicalize.org so can be searched by structure search (sub structure, similar, exact)
- See all similar structures (and names) for any similar structure
: sildenafil = viagra, lodenafil, aildenafil, udenafil ...
- Draw a structure and see it's name
- Automatically chemicalize my blog (WordPress plugin)

Thank you and thanks to



Andras Stracz

Site implementation



Daniel Bonniot, Eufrozina Hoffmann

Document & Name to structure



Alex Allardyce, Ferenc Csizmadia

Features, project management, idiot and advanced testing



Zsolt Kocsmarszky

Design



Roland Molnar

JChem Web Services

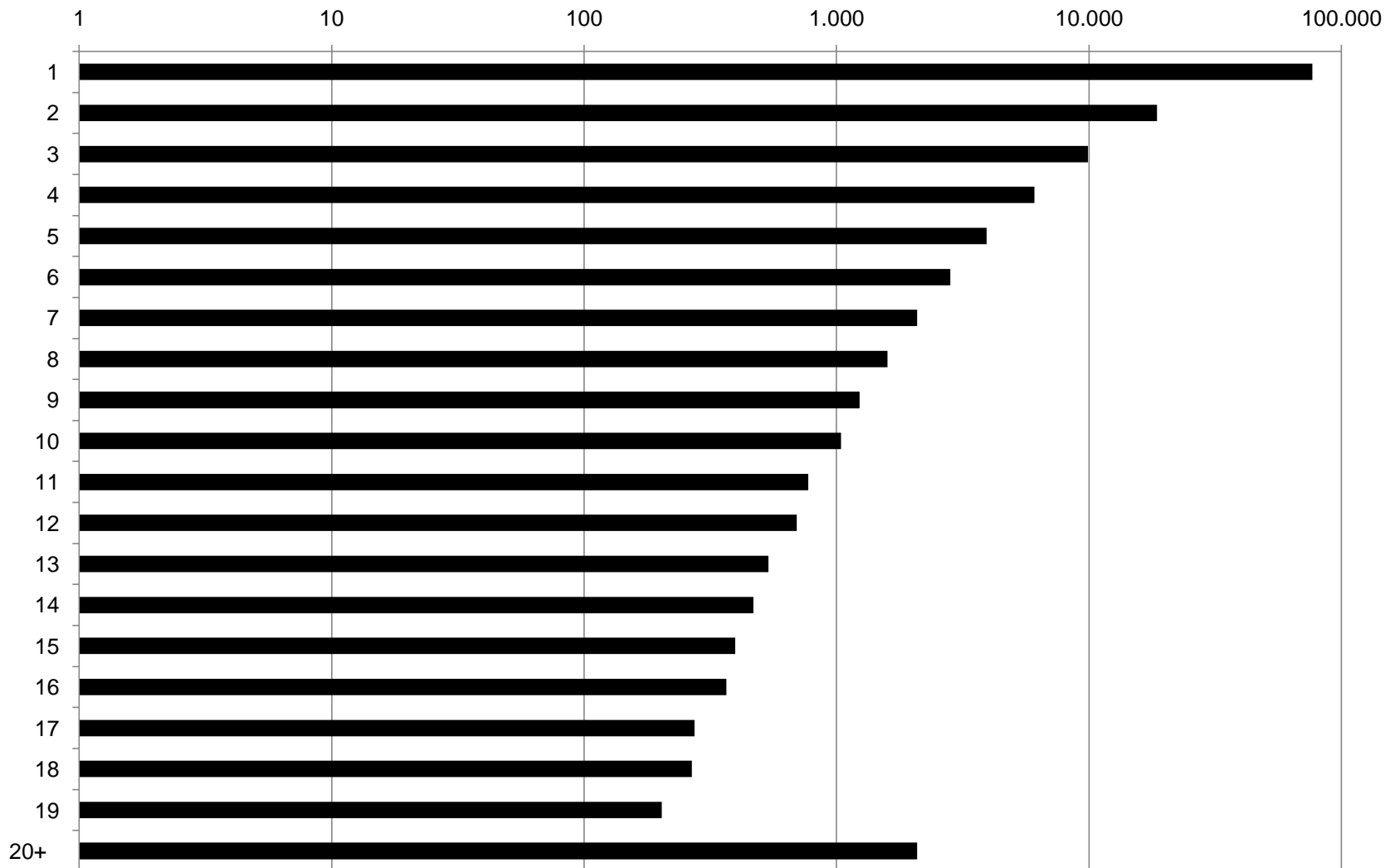
Stats: Raw numbers

(Apr 1, 2010 – Sept 17, 2011)

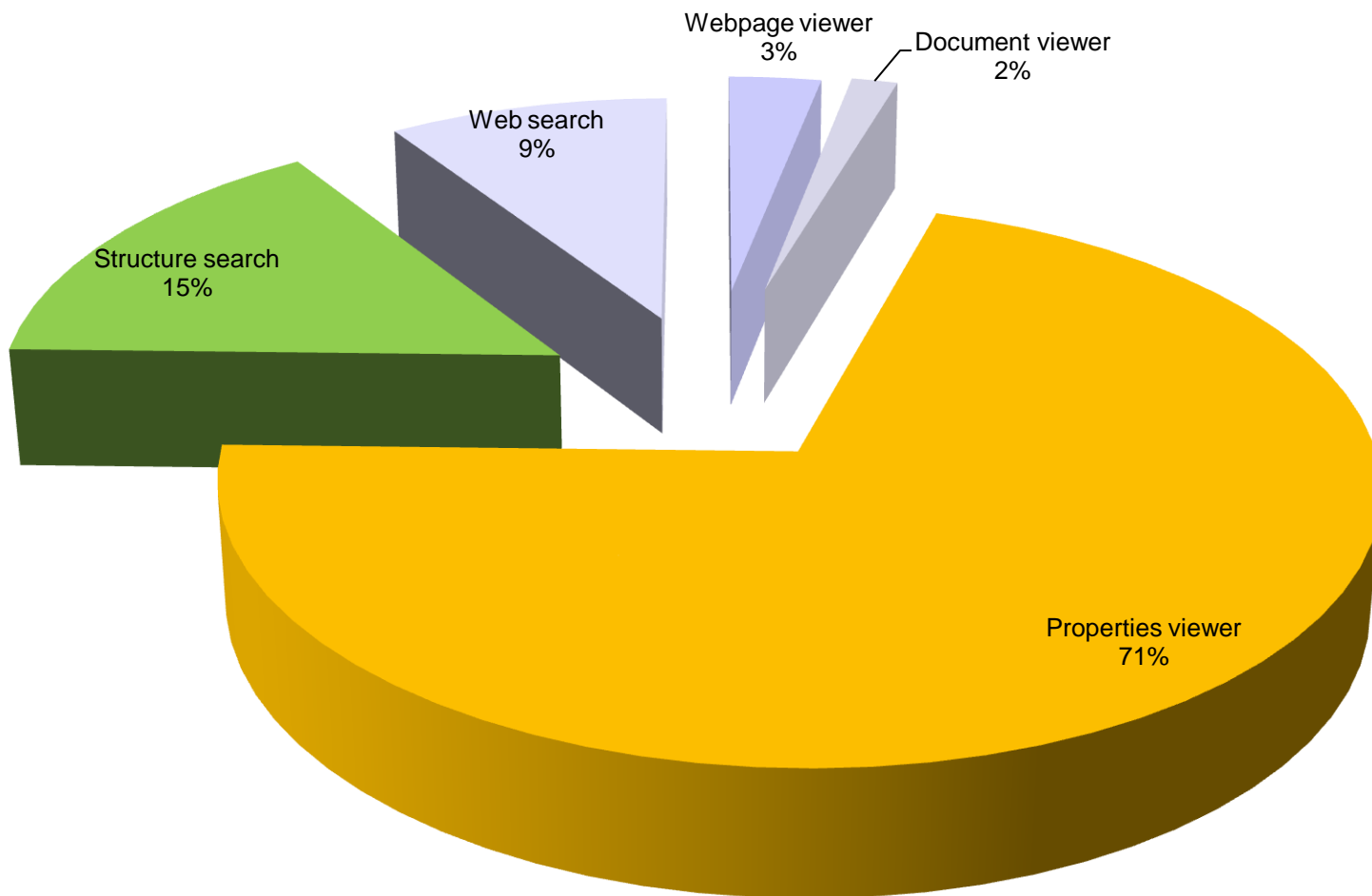
- URL's visited: 400,000
- domains visited: 11,400
- URL's/domain: 35
- total number of chemical names:
 - 4,686,000 (11.71 names/page)
- unique names extracted: 283,000
- structures extracted: 177,000
- unique visitors: 87,100
- avg visitors/day: 437
- avg/longest time on site: 5.37 / 28:41

How busy are they?

Sessions / page views



What features are they using?



Top domains

