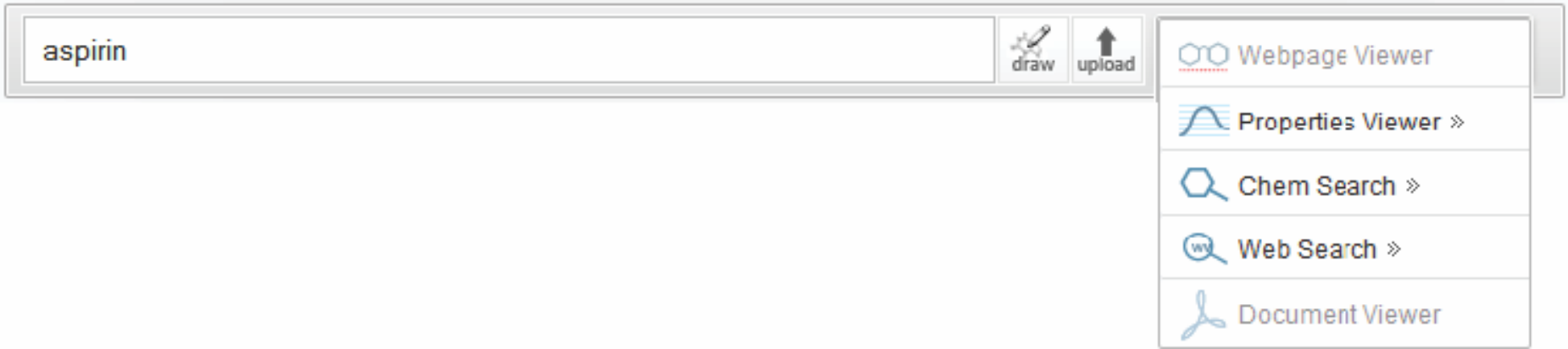
chemicalize.org beta

by ChemAxon

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

Alex Allardyce

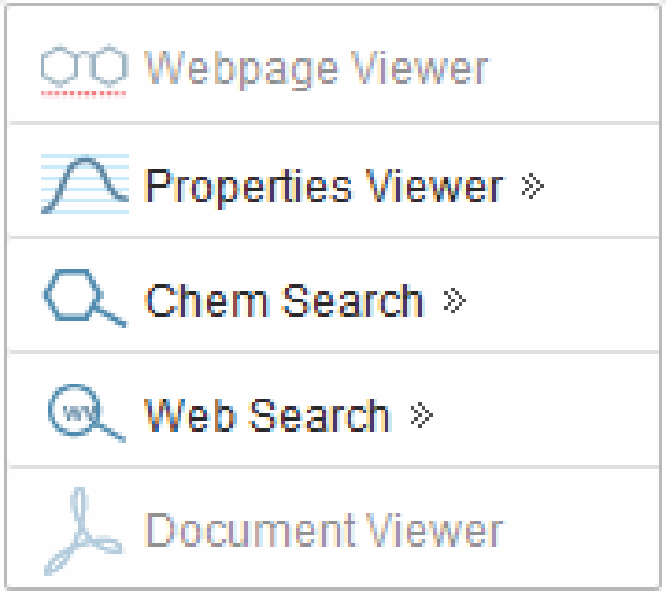
ChemAxon



chemicalize.org beta
by ChemAxon

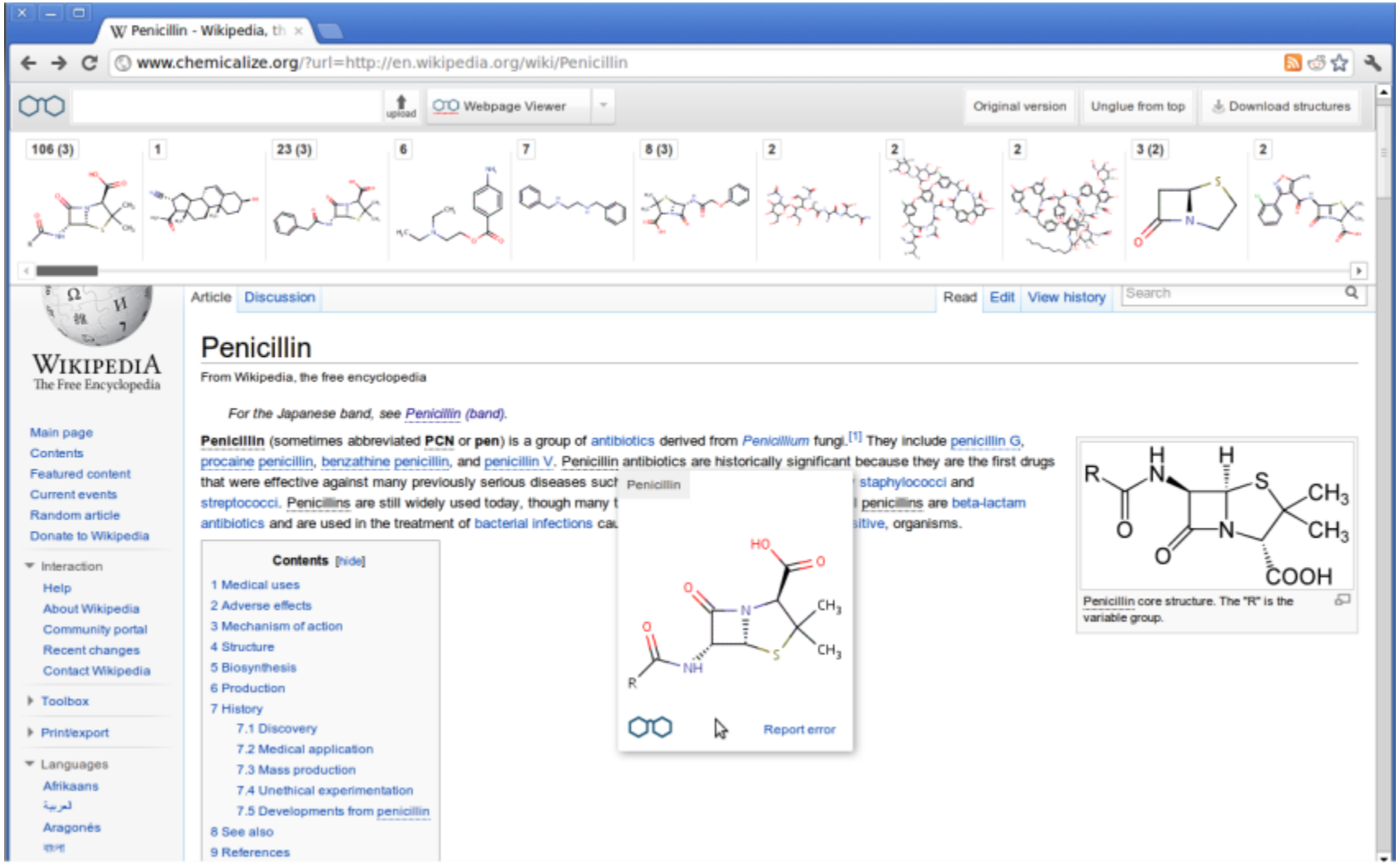
What it's for

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



chemicalize.org beta
by ChemAxon

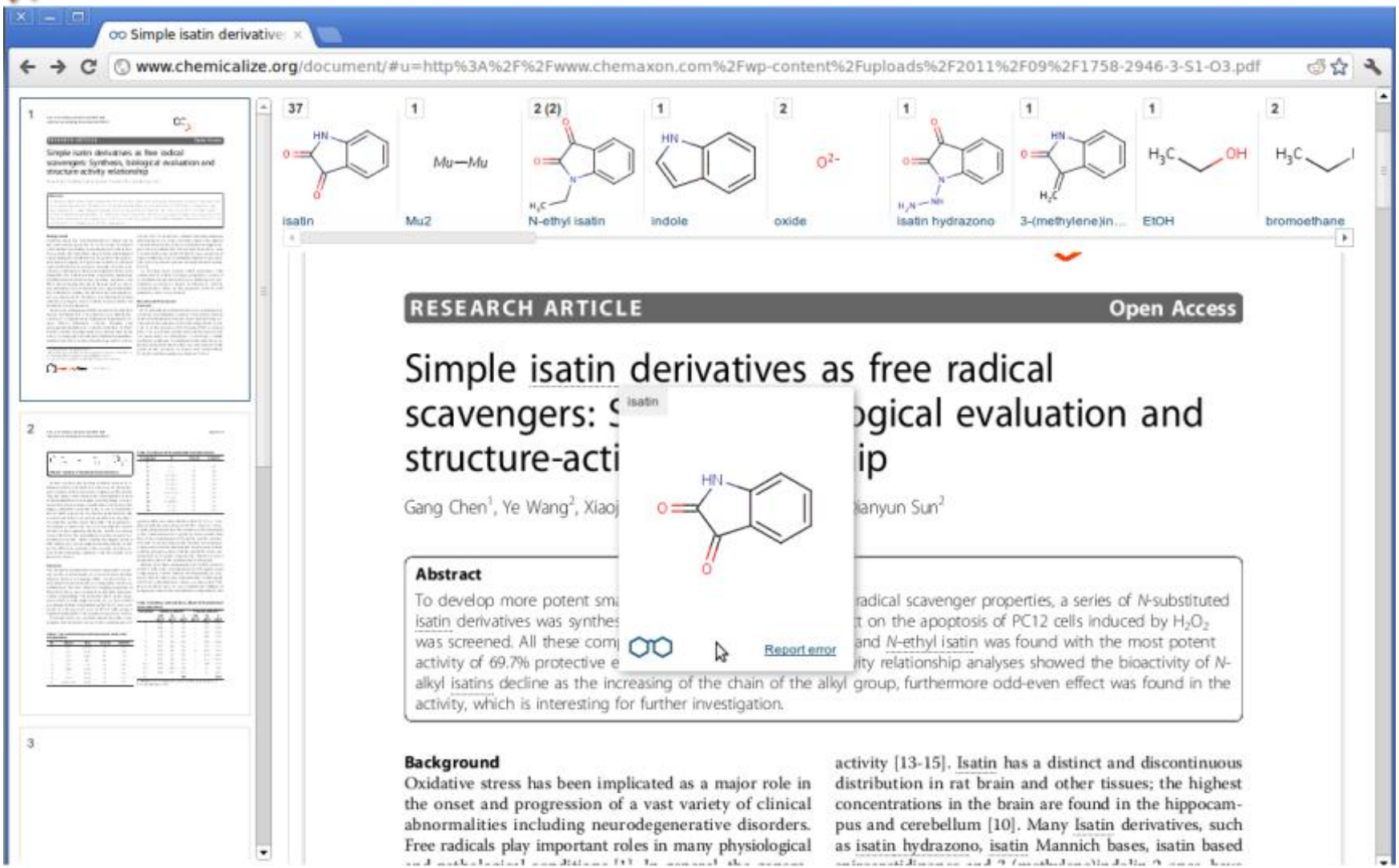
Let's see it – chemicalize a webpage



chemicalize.org beta
by ChemAxon

NEW

… chemicalize a PDF file



chemicalize.org beta
by ChemAxon

NEW

… do sg with discrete text

chemicalize.org beta
by ChemAxon

Home Blog Extend Feedback & Contact Changelog About

draw upload Webpage Viewer

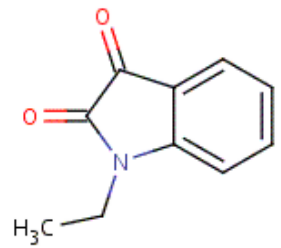
Original

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% prot
PC12 cells. Structure-activity relatio
showed the bioactivity of Nalkyl isati
increasing of the chain of the alkyl g
odd-even effect was found in the activ
interesting for further investigation.

N-ethyl isatin

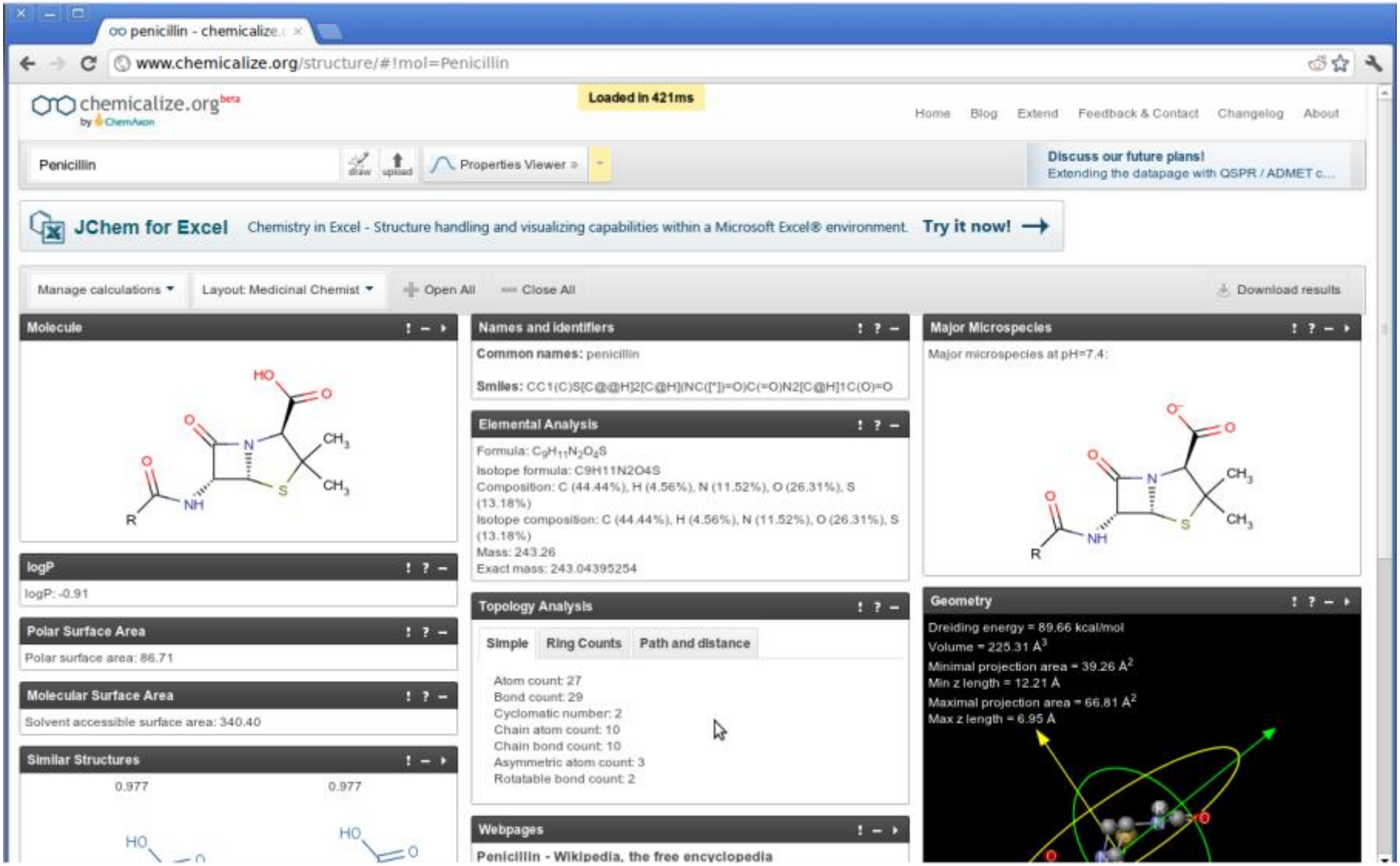


Report error

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

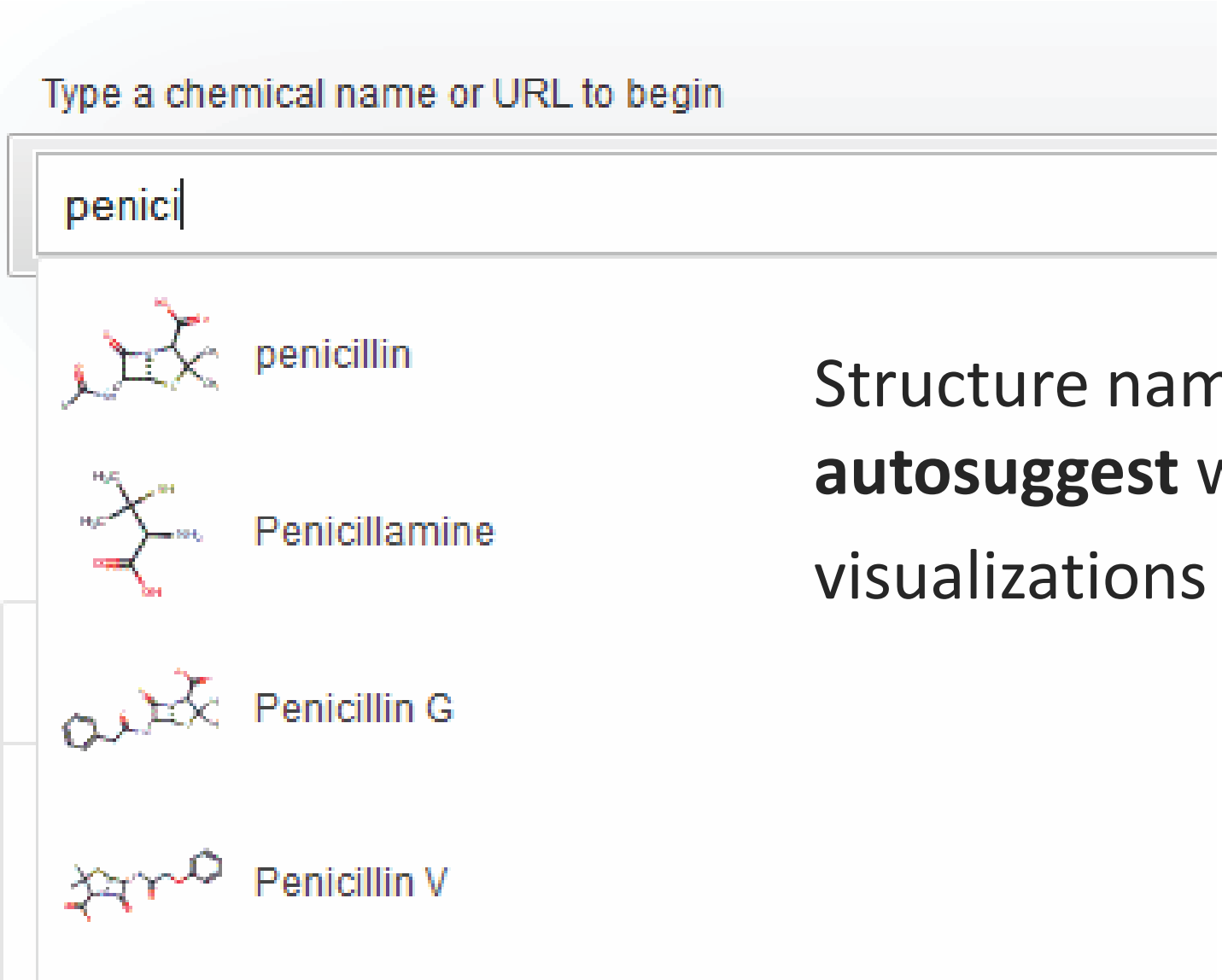
chemicalize.org beta
by ChemAxon

… predict structure properties



chemicalize.org beta
by ChemAxon

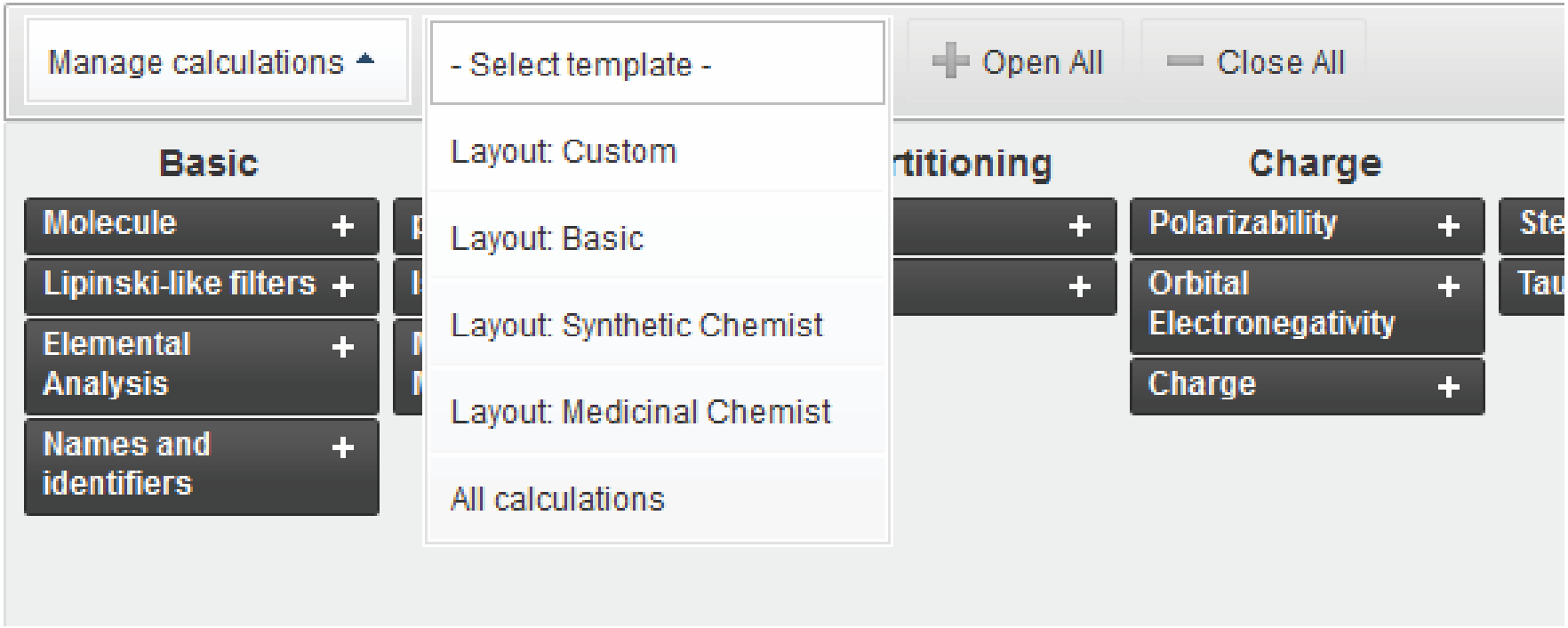
Specialties 1- input box



Structure name **autosuggest** with visualizations

chemicalize.org beta
by ChemAxon

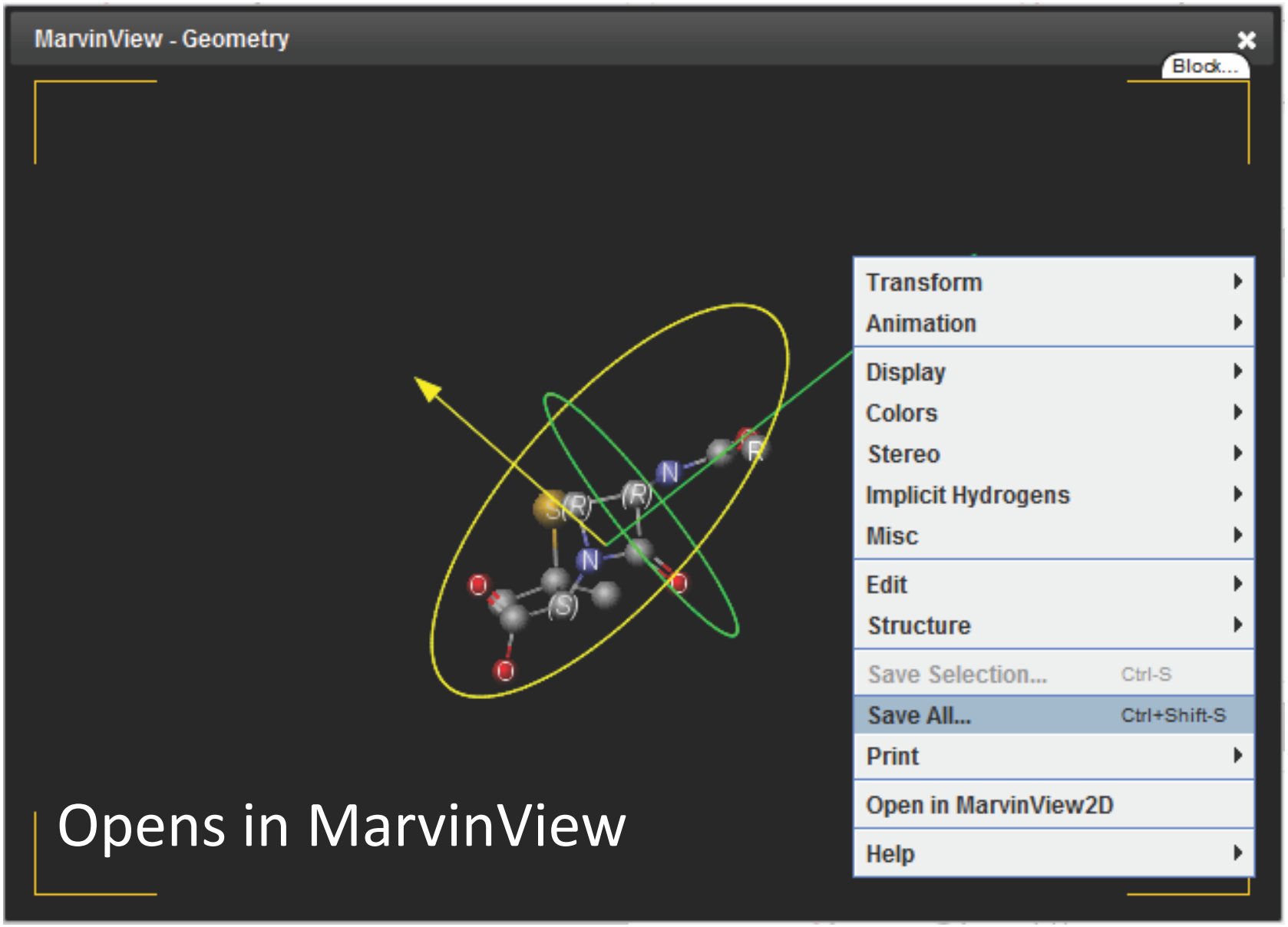
Specialties 2 – configure views



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

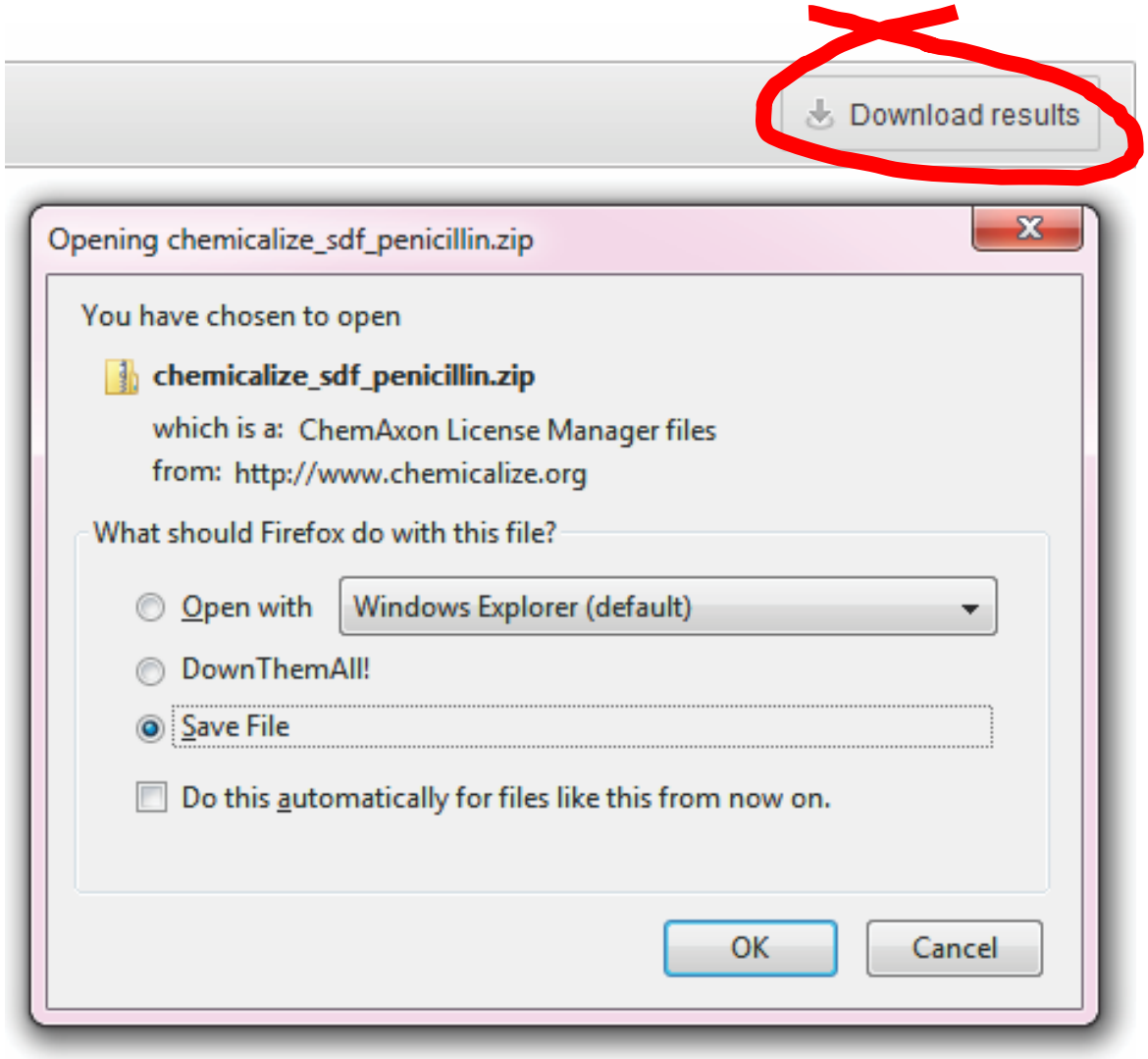
chemicalize.org beta
by ChemAxon

Specialties 3 – viewing structures



chemicalize.org beta
by ChemAxon

Specialties 4 - download



Structure and all generated data

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

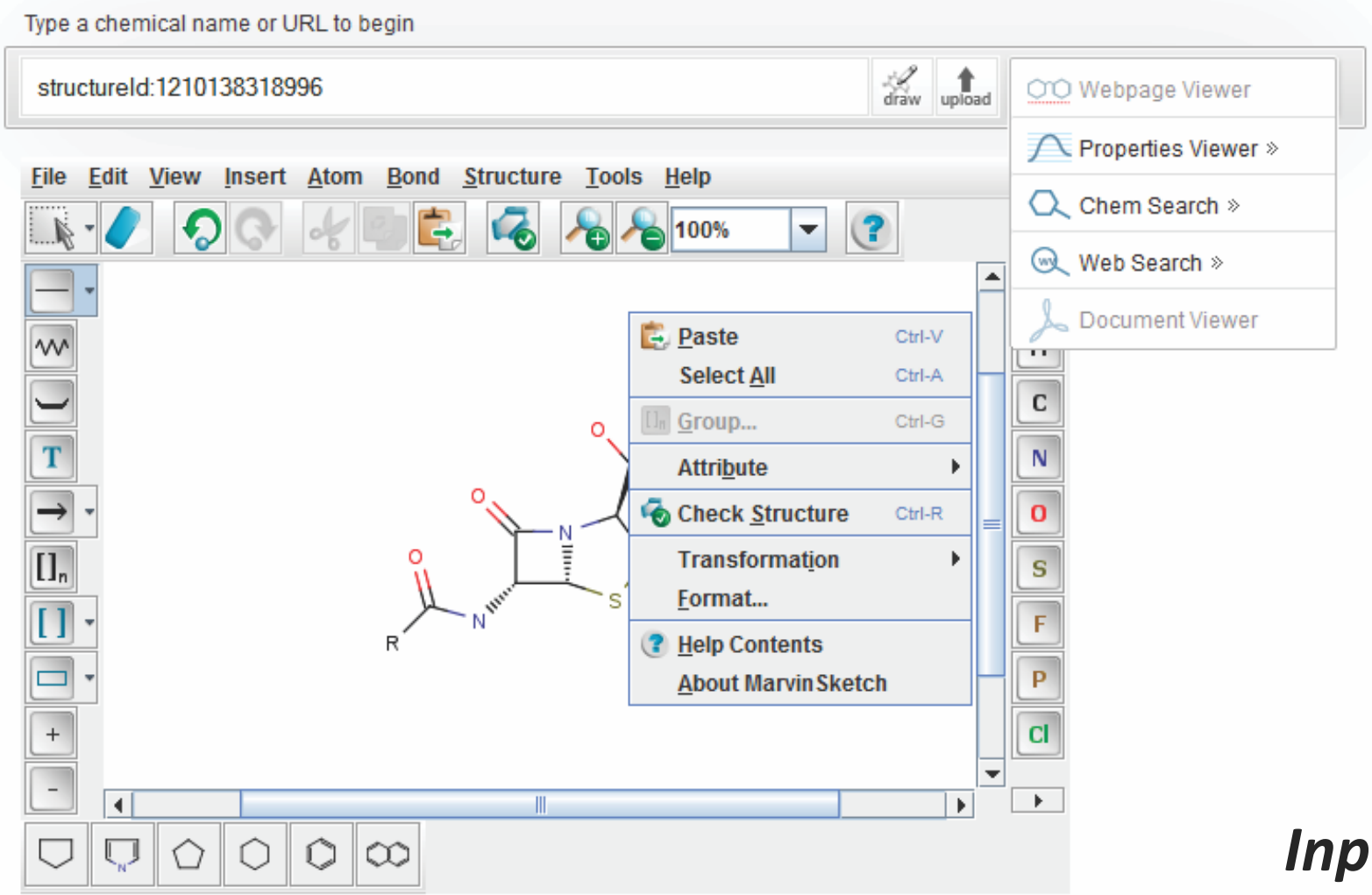
chemicalize.org beta
by ChemAxon

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

chemicalize.org beta
by ChemAxon

Lets create a query



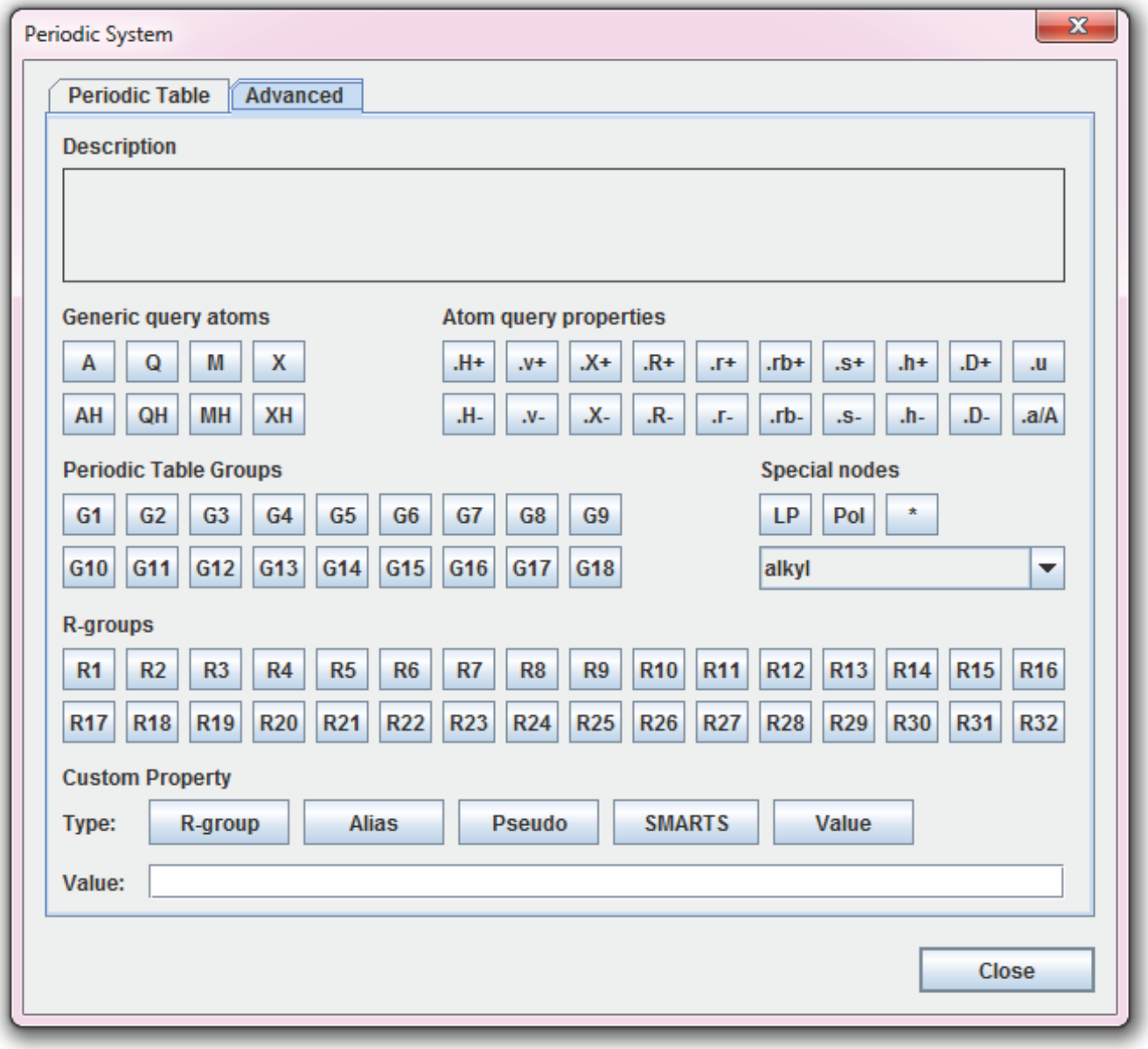
Input via;

MarvinSketch structure editor, name, SMILES, InChI, …

upload query file, drag structure to input box

chemicalize.org beta
by ChemAxon

Query building

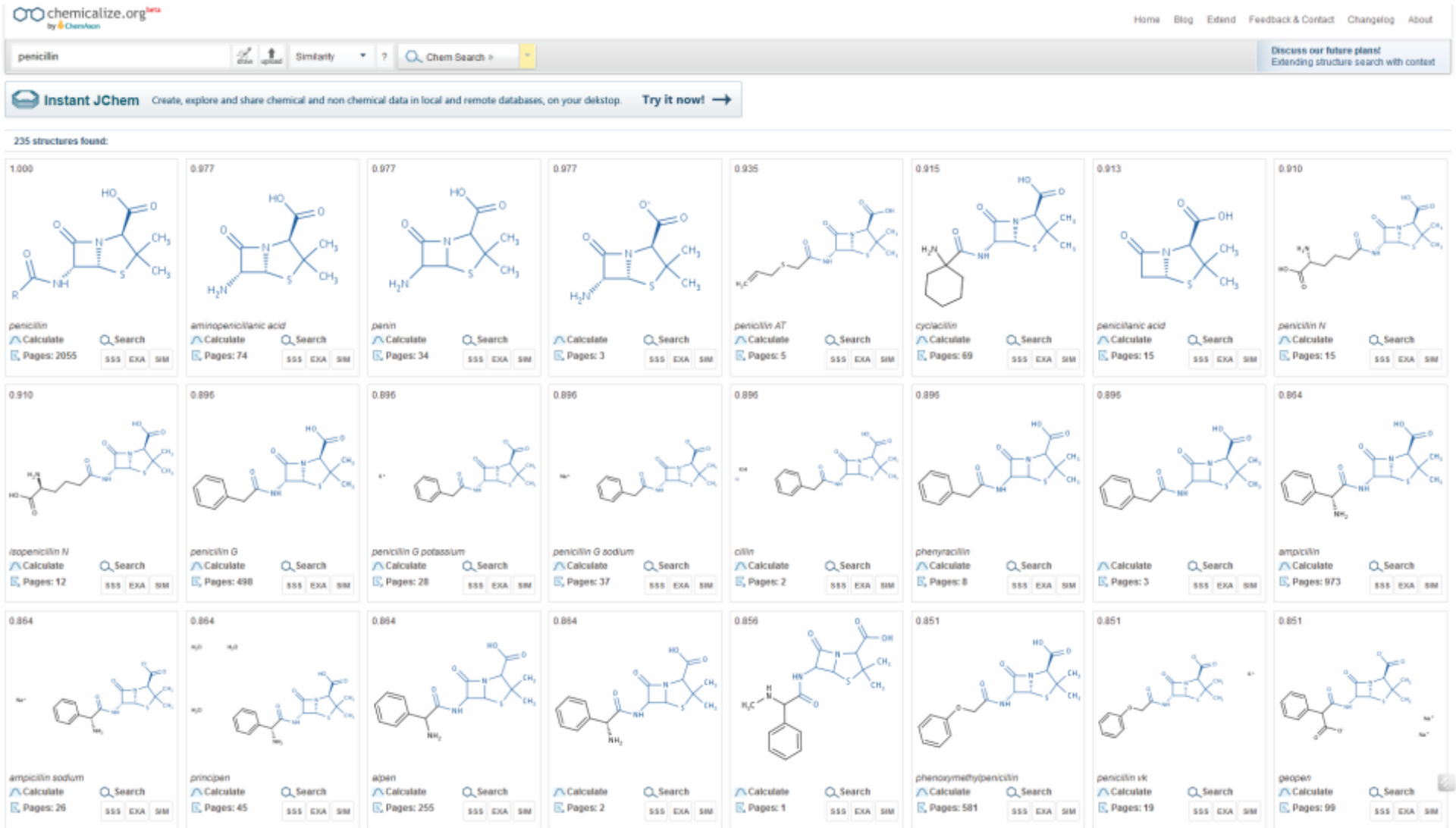


Very powerful query options

(MDL and Daylight logic can be combined)

chemicalize.org beta
by ChemAxon

Query results 1 – result pool

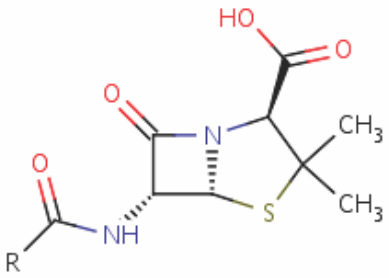


%age similar - query highlighted (blue) in hits

chemicalize.org beta
by ChemAxon

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

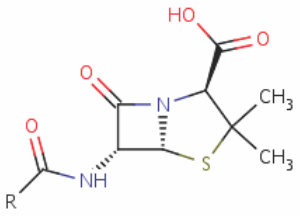
Neuroprotection with Beta-Lactam Compounds - Johns Hopkins University
1 Sep 2011 - original page
freepatentsonline.com/y2007/0238717.html
+ show structures on this page (84) + more results from freepatentsonline.com

Web pages we have met that contain this structure

chemicalize.org beta
by ChemAxon

Query result 2 – discrete structure

« Back to search **Structure found on 2055 pages:**



Calculate Search

SSS EXA SIM

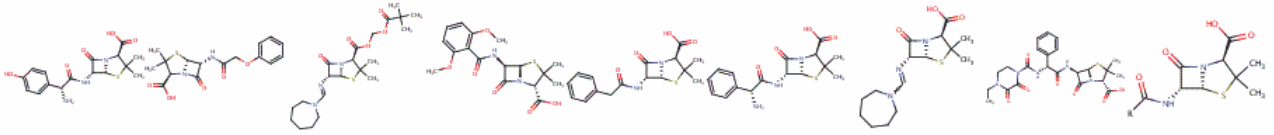
Structure appears as penicillin

hide all related structures ‹

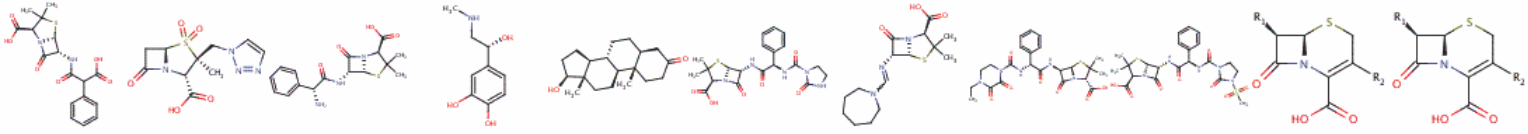
Diverse selection

show all related links ›

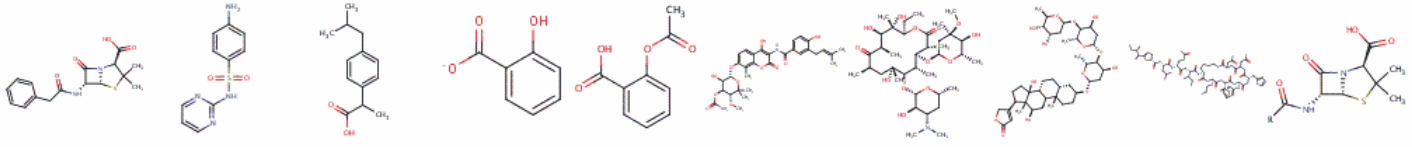
Penicillin - Wikipedia
4 Sep 2011 - original page
sv.wikipedia.org/w/index.php?title=Penicillin
- hide structures on this page (9) + more results from sv.wikipedia.org



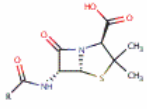
Penicilin - Wikipedia, slobodna enciklopedija - ??????????, ???????? ????????...
4 Sep 2011 - original page
sh.wikipedia.org/w/index.php?title=Penicilin
- hide structures on this page (22) + more results from sh.wikipedia.org



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 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

chemicalize.org beta
by ChemAxon

Web search – what's the difference

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

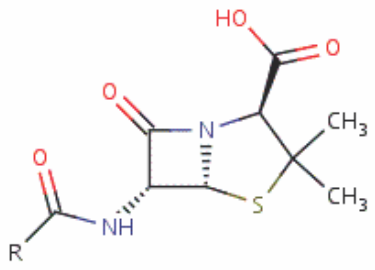
Generated query: "penicillin" OR "CC1(C)S[C@@H]2[C@H](NC([*])=O)C(=O)N2[C@H]1C(O)=O" endothelial

- 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 ;)

chemicalize.org beta
by ChemAxon

WebSearch result

81800 results found:



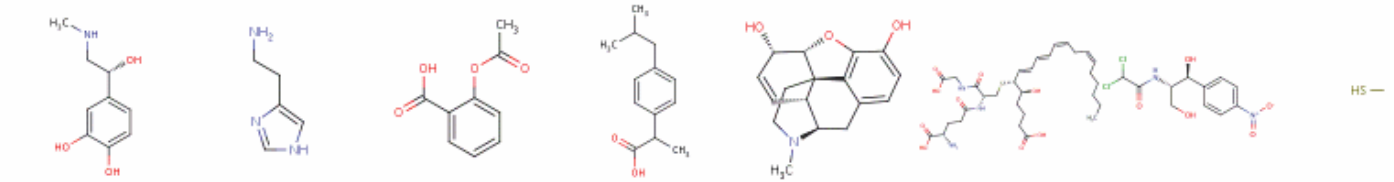
penicillin

Calculate Search

SSS EXA SIM

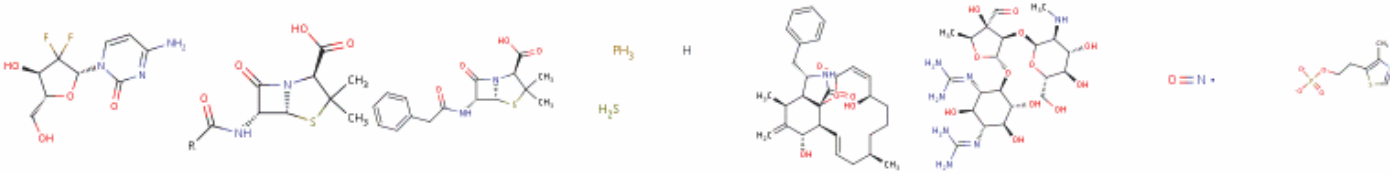
Anaphylaxis - Wikipedia, the free encyclopedia

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)



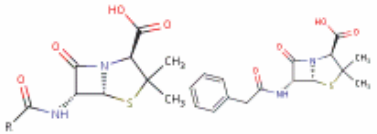
Penicillin G-induced microbicidal activity of endothelial cells cultured ...

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 ...

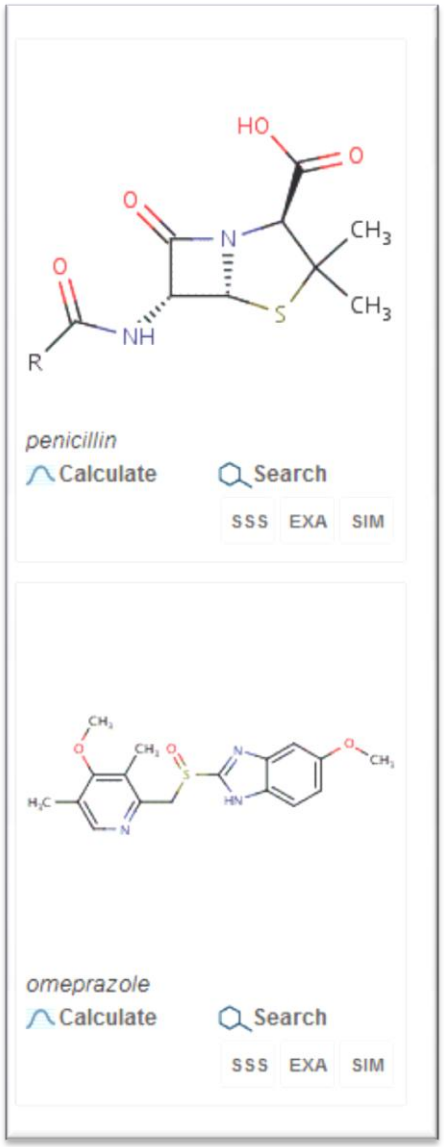
Penicillin G, included in the **endothelial** cell growth medium, was found to be a critical factor in the bactericidal activity demonstrated by. Gelfoam blocks laden ...
www.jstor.org (2)

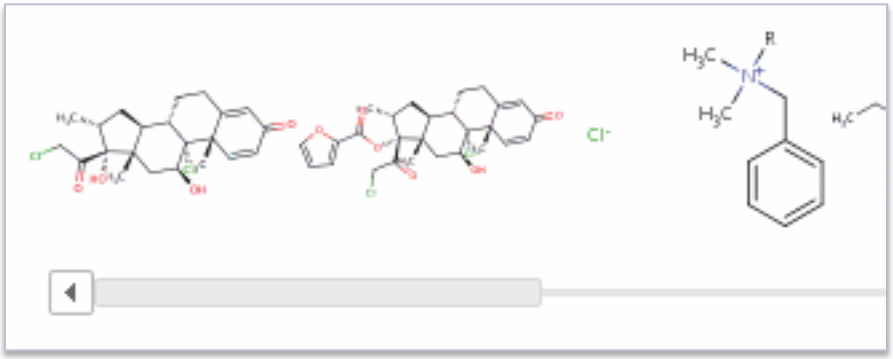


Structures shown by default

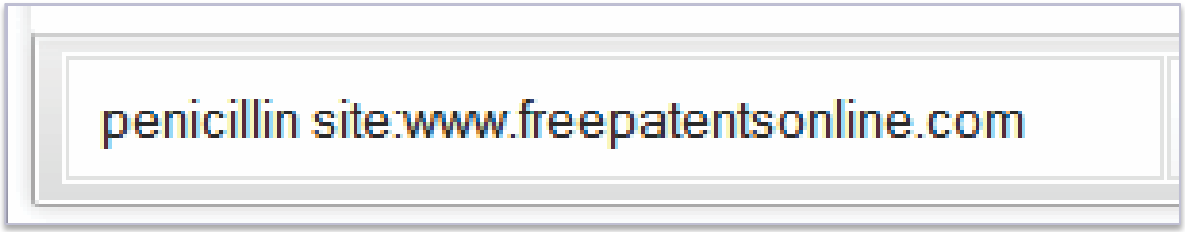
chemicalize.org beta
by ChemAxon

Specialties - snippets





Scroll bar to navigate chemistry on pages



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’

chemicalize.org beta
by ChemAxon

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”

chemicalize.org beta
by ChemAxon

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

chemicalize.org beta
by ChemAxon

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 “alzheimers”
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

chemicalize.org beta
by ChemAxon

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)

chemicalize.org beta
by ChemAxon

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

chemicalize.org beta
by ChemAxon

Stats: Raw numbers

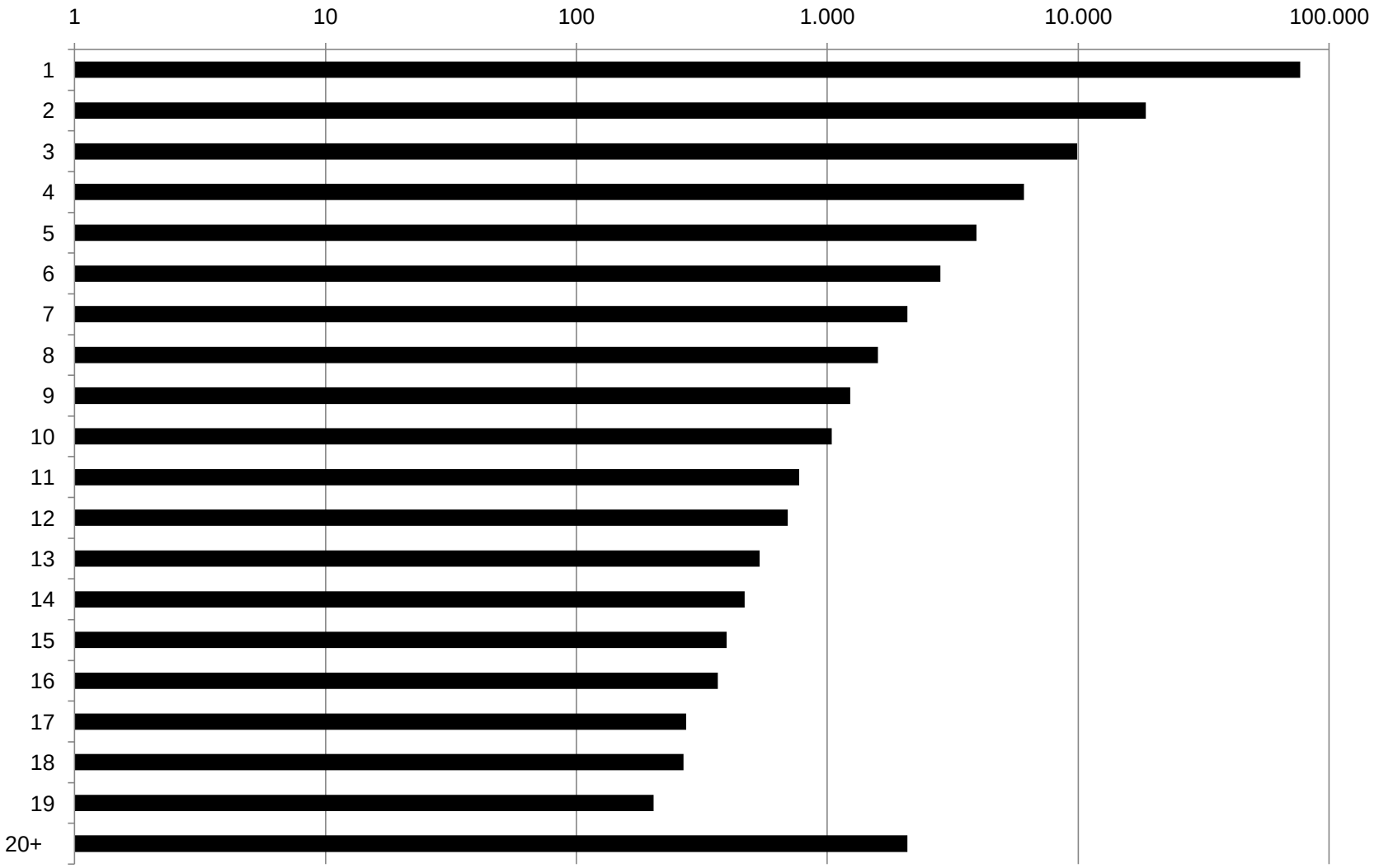
(Apr 1, 2010 – Sept 17, 2011)

- 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

chemicalize.org beta
by ChemAxon

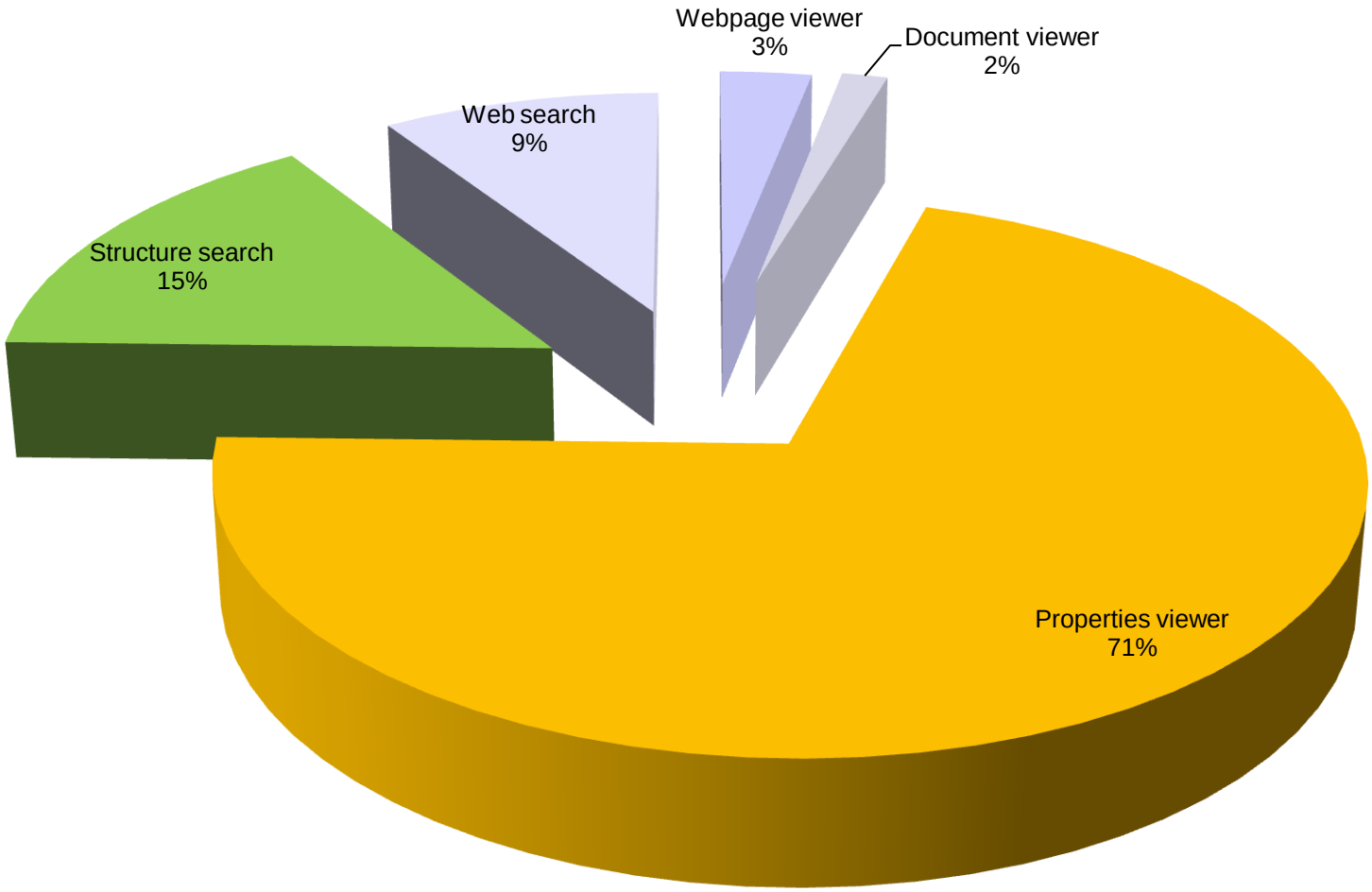
How busy are they?

Sessions / page views



chemicalize.org beta
by ChemAxon

What features are they using?



chemicalize.org beta
by ChemAxon

Top domains