Patent Cheminformatics

Identification of key compounds in patents

3rd Strasbourg Summer School on Chemoinformatics
Strasbourg, France, 25-29 June 2012

Dr. Sorel Muresan

AstraZeneca R&D Mölndal

AstraZeneca

Outlook

Patents, brief into

Sources for accessing full text patents

Compound extraction from patents

Key compound prediction

What is a patent?

A patent application is an agreement between inventor and state, allowing an inventor a monopoly over their invention for a limited time. In the EU, applicants are required to disclose their inventions in a manner sufficiently clear and complete for them to be carried out by a person skilled in the art. In the United States, inventors are additionally required to include the 'best mode' of making or practicing the invention.

Patents are very interesting documents

Three reasons why life scientists should read patents more frequently

1. Some information appears earlier in patents than in scientific journals
2. Patents may contain sound data that never appear in the literature.
3. Patents are a source of hard-to-get information from commercial suppliers.

Seeber, F. *Nat. Protocols 2007*

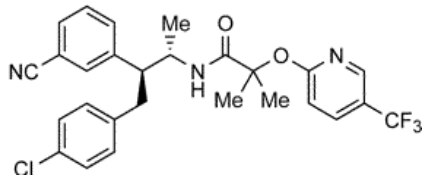
Patents as pharmaceutical data source

Complementary between journals and patents

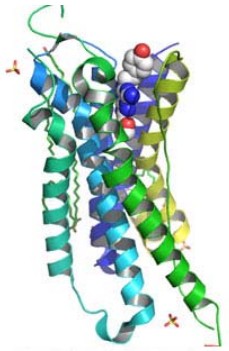
"In certain fields, new advances are disclosed in patents long before they are published in peer-reviewed journals." *Grubb. W.P.*

"Novel Cannabinoid-1 Receptor Inverse Agonist for the Treatment of Obesity"

MERCK Research Laboratories



modulates → CNR1



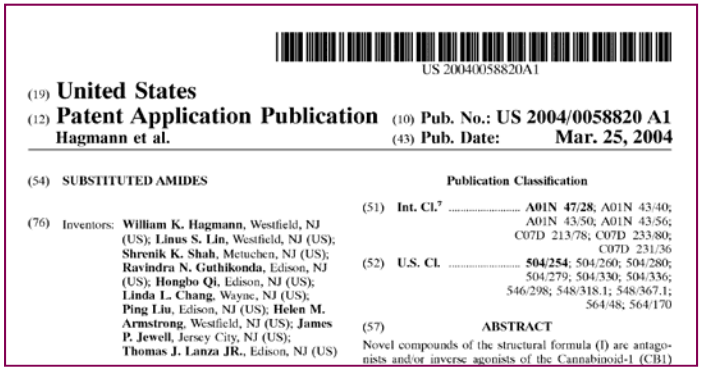
Patent application
Nov 2002

Patent publication
Mar 2004

Journal publication
Dec 2006

~18 months **2.5 years**

USPTO patents (PN: US20040058820)



US 20040058820A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0058820 A1**

Hagmann et al. (43) **Pub. Date: Mar. 25, 2004**

(54) **SUBSTITUTED AMIDES**

Publication Classification

(76) Inventors: **William K. Hagmann**, Westfield, NJ (US); **Linus S. Lin**, Westfield, NJ (US); **Shrenik K. Shah**, Metuchen, NJ (US); **Ravindra N. Guthikonda**, Edison, NJ (US); **Hongbo Qi**, Edison, NJ (US); **Linda L. Chang**, Wayne, NJ (US); **Ping Liu**, Edison, NJ (US); **Helen M. Armstrong**, Westfield, NJ (US); **James P. Jewell**, Jersey City, NJ (US); **Thomas J. Lanza JR.**, Edison, NJ (US)

(51) **Int. Cl.**[7] **A01N 47/28**; A01N 43/40; A01N 43/50; A01N 43/56; C07D 213/78; C07D 233/80; C07D 231/36

(52) **U.S. Cl.** **504/254**; 504/260; 504/280; 504/279; 504/330; 504/336; 546/298; 548/318.1; 548/367.1; 564/48; 564/170

(57) **ABSTRACT**

Novel compounds of the structural formula (I) are antagonists and/or inverse agonists of the Cannabinoid-1 (CB1)

Journal of **Medicinal Chemistry** **(PMID: 17181138)**

Discovery of *N*-[(1*S*,2*S*)-3-(4-Chlorophenyl)-2- (3-cyanophenyl)-1-methylpropyl]-2-methyl-2- {[5-(trifluoromethyl)pyridin-2-yl]oxy} propanamide (MK-0364), a Novel, Acyclic Cannabinoid-1 Receptor Inverse Agonist for the Treatment of Obesity

Linus S. Lin,*[‡] Thomas J. Lanza, Jr.,[‡] James P. Jewell,[‡] Ping Liu,[‡] Shrenik K. Shah,[‡] Hongbo Qi,[‡] Xinchun Tong,[‡] Junying Wang,[‡] Suoyu S. Xu,[‡] Tung M. Fong,[§] Chun-Pyn Shen,[§] Julie Lao,[§] Jing Chen Xiao,[§] Lauren P. Shearman,[∞] D. Sloan Stribling,[∞] Kimberly Rosko,[∞] Alison Strack,[∞] Donald J. Marsh,[§] Yue Feng,[§] Sanjeev Kumar,[‡] Koppara Samuel,[‡] Wenji Yin,[‡] Lex H. T. Van der Ploeg,[§#] Mark T. Goulet,[‡#] and William K. Hagmann[‡]

Abstract | Supporting Info

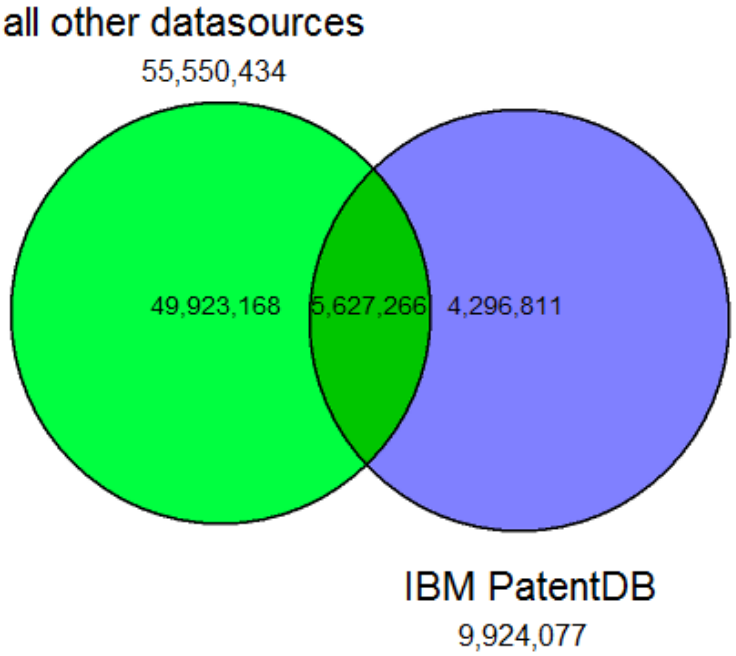
Full Text HTML

Hi-Res PDF [85 KB]

PDF w/ Links [163 KB]

Unique chemistry from patents

Data from AstraZeneca's Chemistry Connect



~6% of compound structures exemplified in patents were also published in journal articles

Muresan, S. et al. *DDT* 2011
Southan, C. et al *J.Cheminfo.* 2009

Anatomy of a patent

Front page - contains a wealth of information about the patent

Detailed description of the invention - the heart of a patent application. It generally describes one or more preferred embodiments of the invention in enough detail to enable someone of ordinary skill in the art to make or use the invention without having to resort to undue experimentation

Claims - the most important part of the patent application. Define the scope of patent protection afforded to the owner of a patent.

C.P. Miller, M.J. Evans, The Chemist's Companion Guide to Patent Law, 2011

United States Patent [19]

Edwards et al.

[11] **Patent Number: 4,935,437**

[45] **Date of Patent: Jun. 19, 1990**

[54] **(SUBSTITUTED ARALKYL) HETEROCYCLIC COMPOUNDS**

[75] Inventors: **Philip N. Edwards**, Bramhall; **Michael S. Large**, Stoke-on-Trent, both of England

[73] Assignee: **Imperial Chemical Industries plc**, London, England

[21] Appl. No.: **204,743**

[22] Filed: **Jun. 10, 1988**

[30] **Foreign Application Priority Data**

Jun. 16, 1987 [GB] United Kingdom 8714013

[51] **Int. Cl.**[5] **C07D 249/08; A61K 31/41**

[52] **U.S. Cl.** **514/383;** 514/236.2; 514/326; 514/422; 544/132; 546/210; 548/518; 548/262.2

[58] **Field of Search** 548/262; 514/383, 236.2, 514/326; 544/132; 546/210

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,226,878	10/1980	Iizuka et al.	548/335
4,271,170	6/1981	Tanouchi et al.	546/284
4,713,387	12/1987	Watanabe et al.	514/332
4,749,713	6/1988	Bowman et al.	514/341
4,755,526	7/1988	Hirsch et al.	514/399

FOREIGN PATENT DOCUMENTS

0165778 12/1985 European Pat. Off. .

OTHER PUBLICATIONS

Iizuka et al., J. Med. Chem., 1981, 24, 1139–1148.

Tanouchi et al., J. Med. Chem., 1981, 24, 1149–1155.

Chem. Abstr., 99, 139652b, Mitsubishi Chemical.

J. Org. Chem., 1987, 52(5), 946–8, (Eq. to Chem. Abstr., 106, 102640).

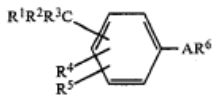
Chem. Abstr., 106, 50050u.

Chemical Abstracts, vol. 107, No. 21, Nov. 23, 1987, Columbus, Ohio, U.S.A., Wickings, E. J.; Middleton, M. C., "Non-Steroidal Inhibition of Granulosa Cell Aromatase Activity in Vitro", p. 88, col. 2, Abstract No. 191 136j.

Primary Examiner—Glennon H. Hollrah
Assistant Examiner—Patricia L. Morris
Attorney, Agent, or Firm—Cushman, Darby & Cushman

[57] **ABSTRACT**

A (substituted-aralkyl)heterocyclic compound of the formula I



I

wherein R^1 is an azido, carbamoyl, cyano, formyl, hydroxy or nitro radical, a 1–6C 1-hydroxyalkyl, alkoxy, alkylcarbamoyl, alkylthio, alkylsulphinyl or alkylsulphonyl radical, a 2-cyanoethyl radical, optionally bearing one to four 1–6C alkyl substituents, or a 2–6C alkanoyl, halogenoalkanoyl, alkanoyloxy, alkanoylamino, dialkylcarbamoyl or alkoxycarbonyl radical; R^2 and R^3, which may be the same or different, are each a hydrogen atom, a 1–6C alkyl, dueterioalkyl or halogenoalkyl radical, or a phenyl or phenyl(1–6C alkyl) radical, in each of which the phenyl may optionally bear one or more substituents; or R^2 and R^3, together with the carbon atom to which they are attached, may form a 3- to 6-membered ring; or $R^1R^2R^3C$- is a 1,1-dicyanoethyl or trifluoromethylsulphonyl radical; R^4 is a hydrogen or halogen atom, a cyano or nitro radical or a 1–6C alkyl or halogenoalkyl radical; R^5 has any of the values defined above for the group $R^1R^2R^3C$ but is not necessarily the same as $R^1R^2R^3C$, or has any of the values defined above for R^4 but is not necesarily the same as R^4, or is a carbamoyl, 1-pyrrolidinyl-carbonyl, piperidinocarbonyl, morpholinocarbonyl or nitro radical, a 1–6C alkoxy or halogenoalkoxy radical or a 2–6C alkanoyl or alkoxy-carbonyl radical; A is a methylene or ethylene radical optionally bearing one or more substituents selected from deuterium and halogen atoms, carbamoyl, cyano and hydroxy radicals, 1–6C alkyl and alkoxy radicals, and 2–6C alkanoyloxy radicals provided that when A is linked to R^6 through a nitrogen atom thereof, it may not bear a hydroxy, alkoxy or alkanoyloxy substituent on the carbon atom adjacent to such nitrogen atoms; and R^6 is a 1H-1,2,4-triazol-1-yl, 4H-1,2,4-triazol-4-yl, 1H-imidazol-1-yl, 5-cyano-1H-imidazol-1-yl, 3-pyridyl or 5-pyrimidinyl radical, or a 1H-imidazol-1-yl radical, bearing at the 5-position thereof a 1–6C alkyl substituent which is itself optionally substituted by one or more carbamoyl, cyano, hydroxy or 2–6C alkoxycarbonyl radicals; and provided that when R^2, R^3, R^4 and R^5 are hydrogen, A is a methylene radical and R^6 is a 3-pyridyl radical, R^1 is not a cyano, hydroxy or hydroxymethyl radical, and when R^1 is a hydroxy radical, R^3, R^4 and R^5 are hydrogen, A is a methylene radical and R^6 is 3-pyridyl, R^2 is not a methyl or a 2-chloro-1-methylethyl radical, and provided that when R^1 is a methoxycarbonyl radical, R^2, R^3, R^4 and R^5 are hydrogen and A is a methylene radical, R_1 is not a 1H-imidazol-1-yl radical; and the pharmaceutically acceptable acid addition salts thereof.

9 Claims, No Drawings

EXAMPLE 66

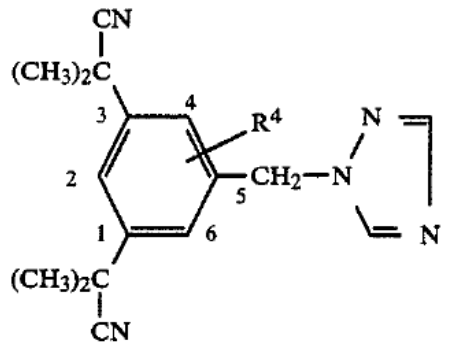
A mixture of 2-[2-bromo-5-(1H-1,2,4-triazol-1-ylmethyl)-phenyl]-2-methylpropiononitrile (0.15 g), dimethylformamide (2 ml) and cuprous cyanide (0.09 g) was stirred and heated under reflux for 8 h. The cooled mixture was treated with aqueous potassium cyanide solut
utes,
The
dryn
purif
by v
5-(1H
pion
159°
A
phen
triaz
stirre
wate
extra
duce
mato
rome
triaz
pion
mp 1
Th
taine

EXAMPLE 69

A 20% (w/v) solution of sodium nitrite in water was added dropwise in a stirred mixture of 4 amino-1-[3,5-bis (1-cyano-1-methylethyl)benzyl]-1H-1,2,4-triazolium bromide and 2N aqueous hydrochloric acid (10 ml), until a slight excess of nitrite was present. The solution

EXAMPLES 49–52

The process described in Example, 1 was repeated, using the appropriate 2- or 4-substituted 2,2-(5-methyl-1,3-phenylene)di(2-methylpropiononitrile) as starting material, to give the following compounds:



Ex	R^4	Position of substitution	Mp.	Footnote
49	NO_2	4	—	1,2
50	Br	4	83–86	3
51	Br	2	128–131	3
52	CN	4	35–37	4

EXAMPLE 53

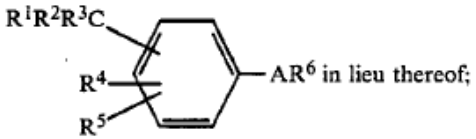
A mixture of 2,2′-(5-chlorodideuteriomethyl-1,3-phenylene)-di(2-trideuteriomethyl-3,3,3-trideuteriopropiononitrile) (0.65 g), dimethylformamide (5 ml) and sodium triazole (0.45 g) was stirred at room temperature for 18 h. The mixture was diluted with water (30 ml) and extracted with ethyl acetate, and the extract was dried and evaporated to dryness under reduced pressure. The residue was purified by flash chromatography, using ethyl acetate as eluant, to give 2,2′-[5-dideuterio-(1H-1,2,4-triazol-1-yl)methyl-1,3-phenylene]-di(2-trideuteriomethyl-3,3,3-trideuteriopropiononitrile), mp 82°–83° after crystallisation from ethyl acetate/cyclohexane.

The starting material from the above process may be prepared as follows:

The process used to prepare methyl 3,5-bis(1-cyano-1-methylethyl)benzoate, described in the later part of Example 8, was repeated, using trideuterioiodomethane instead of iodomethane, to give methyl 3,5-bis[1-cyano-2,2,2-trideuterio-1-(trideuteriomethyl)ethyl]-benzoate, m.p. 83°–84°.

We claim:

1. A (substituted-aralkyl)heterocyclic compound of the formula I



wherein R^1 is an azido, carbamoyl, cyano, formy
droxy or nitro radical, a 1–6C 1-hydroxyalkyl, a
alkylcarbamoyl, alkylthio, alkylsulphinyl or all
phonyl radical, a 2-cyanoethyl radical, optionally
ing one to four 1–6C alkyl substituents, or a 2–6C
oyl, halogenoalkanoyl, alkanoyloxy, alkanoyla
dialkylcarbamoyl or alkoxycarbonyl radical; R^2 a
which may be the same or different, are each a
alkyl, deuterioalkyl or halogenoalkyl radica
$R^1R^2R^3C$-is a 1,1-dicyanoethyl or trifluorometh
phonyl radical; R^4 is a hydrogen or halogen at
cyano or nitro radical or a 1–6C alkyl or halogen
radical; R^5 has any of the values defined above f
group $R^1R^2R^3C$, or has any of the values defined
for R^4, or is a carbamoyl, 1-pyrrolidinyl-carl
piperidinocarbonyl, morpholinocarbonyl or nitro
cal, a 1–6C alkoxy or halogenoalkoxy radical or a
alkanoyl or alkoxy-carbonyl radical; A is a meth
or ethylene radical optionally bearing one or mor
stituents selected from the group consisting of
rium and halogen atoms, carbamoyl, cyano an
droxy radicals, 1–6C alkyl and alkoxy radicals, and
alkanoyloxy radicals provided that when A is lin
R^6 through a nitrogen atom thereof, it may not
hydroxy, alkoxy or alkanoyloxy substituent on th

bon atom adjacent to such nitrogen atoms; and R^6 is a 1H-1,2,4-triazol-1-yl or 4H-1,2,4-triazol-4 yl; and the pharmaceutically acceptable acid addition salts thereof.

2. A compound as claimed in claim 1 wherein R^1 is an

ycarbonyl, butoxycarbonyl, isobutoxycarbonyl, sec-butoxycarbonyl, tert-butoxycarbonyl or pentyloxycarbonyl radical; R^2 and R^3, which may be the same or different, are each a methyl, ethyl, propyl, isopropyl, butyl, isobutyl, sec-butyl, tert-butyl, pentyl, neopentyl, hexyl, trideuteriomethyl, mono-, di or tri-chloromethyl, mono-, di- or trifluoromethyl, 2,2,2-trichloro- or trifluoro-ethyl, 1,2,2-trichloro-or trifluoro-ethyl, pentafluoroethyl, 2,2,3,3,3-pentafluoropropyl, 2,2-dichloro-3,3,3-trifluoropropyl, 4,4,4-trifluorobutyl, 5,5,5-trifluoropentyl or 6,6,6-trifluorohexyl radical, or $R^1R^2R^3C$ is a 1,1-dicyanoethyl or trifluoromethylsulphonyl radical; R^4 is a hydrogen atom, a cyano or nitro radical, or a 1–6C alkyl or halogenoalkyl radical as defined above; R^5 has any of the values defined above for the group $R^1R^2R^3C$, or has any of the values defined above for R^4, or is a 1–6C alkoxy or a 2 6C alkanoyl or alkoxycarbonyl radical as defined above, or a carbamoyl, 1-pyrrolidinylcarbonyl, piperidinocarbonyl, morpholinocarbonyl or nitro radical, a fluorine, chlorine, bromine or iodine atom, or a mono-, di- or tri-chloromethoxy, mono-, di- or trifluoromethoxy, bromomethoxy, iodomethoxy, 2,2,2-trichloro- or trifluoro-ethoxy, 1,2,2-trichloro- or trifluoro-ethoxy, pentafluoroethoxy, 2,2,3,3,3-pentafluoropropoxy, 2,2-dichloro-3,3,3-trifluoropropoxy, 4,4,4-trifluorobutoxy, 5,5,5-trifluoropentyloxy or 6,6,6-trifluorohexyloxy radical; A is an ethylidene, propylidene, butylidene, 1- or 2-methylethylene, 1,2-dimethylethylene, dideuteriomethylene, difluoromethylene, hydroxymethylene, cyanomethylene or carbamoylmethylene radical, or a 1-hydroxyethylene radical (in which C-1 of the ethylene is linked to the benzene ring) radicals.

3. A compound as claimed in claim 1 which is a hydrochloride, hydrobromide, sulphate, nitrate, phosphate or toluene-p-sulphonate.

4. A compound as claimed in claim 1, 2 or 3 wherein R^1 is a carbamoyl, cyano, hydroxy, 1-hydroxyethyl, methylthio, methylsulphinyl, methylsulphonyl or acetyl radical R^2 and R^3, which may be the same or different, are each a methyl, ethyl, trideuteriomethyl or fluoromethyl radical; R^4 is a hydrogen, fluorine or bromine atom or a cyano, nitro, isopropyl or chloromethyl radical; R^5 is a 1-cyano-1-methylethyl, 1,1-dimethyl-2-oxopropyl, 1-carbamoyl-1-methylethyl, 1-cyano-1-trideuteriomethyl-2,2,2-trideuterioethyl, 1-cyano-2-fluoro-1-(fluoromethyl)ethyl, 1-methyl-1-(methylsulphonyl)-ethyl, 1-cyano-1-ethylpropyl, carbamoyl, 1-piperidinocarbonyl, 1-morpholinocarbonyl, acetyl or methoxycarbonyl radical; A is a methylene, ethylene, ethylidene, isopropylidene, dideuteriomethylene, hydroxymethylene, cyanomethylene, fluoromethylene or difluoromethylene radical, or a 1-hydroxyethylene radical in which the carbon atom bearing the hydroxy substituent is bonded to the benzene ring; and R^6 is a 1H-1,2,4-triazol-1-yl, 4H-1,2,4-triazol-4-yl.

5. A compound as claimed in claim 1 wherein R^1 is a cyano radical, R^5 is a radical of the formula $R^1R^2R^3C$ wherein R^1 is a cyano or hydroxy radical, and R^6 is a 1H-1,2,4-triazol-1-yl radical.

6. A compound as claimed in claim 5 wherein R^2 and R^3, both in the group $R^1R^2R^3C$ and in R^5, are methyl or trideuteriomethyl radicals, and A is a methylene or dideuteriomethylene radical.

7. A compound as claimed in claim 1 which is 2,2′-[5-(1H-1,2,4-triazol-1-ylmethyl)-1,3-phenylene]di(2-methylpropiononitrile), 2-[3-(1-hydroxy-1-methylethyl)-5-(1H-1,2,4-triazol-1-ylmethyl)phenyl]-2-methylpropiononitrile, 2,2′-[5-dideuterio(1H-1,2,4-triazol-1-yl)-methyl-1,3-phenylene]di(2-trideuteriomethyl-3,3,3-trideuteriopropiononitrile) or 2,2′-[5-dideuterio(1H-1,2,4-triazol-1-yl)methyl-1,3-phenylene]di(2-methylpropiononitrile).

8. A pharmaceutical or veterinary composition which comprises an effective amount of a compound as claimed in claim 1 together with a pharmaceutically or veterinarily acceptable diluent or carrier.

9. A method of treating steroid hormone-dependent tumours which comprises administering to a host in need of such treatment an effective amount of a compound as claimed in claim 1.

* * * * *

Searching for patents

Official public portals

Europäisches Patentamt
European Patent Office
Office européen des brevets

uspto.GOV
The United States Patent and Trademark Office
an agency of the Department of Commerce

WIPO
Encouraging Creativity and Innovation
WORLD INTELLECTUAL PROPERTY ORGANIZATION

Open full-text

FPQ
IP Research Tools, Data, & Communities

Google patents

Structure -> Patents & Patent -> Structure

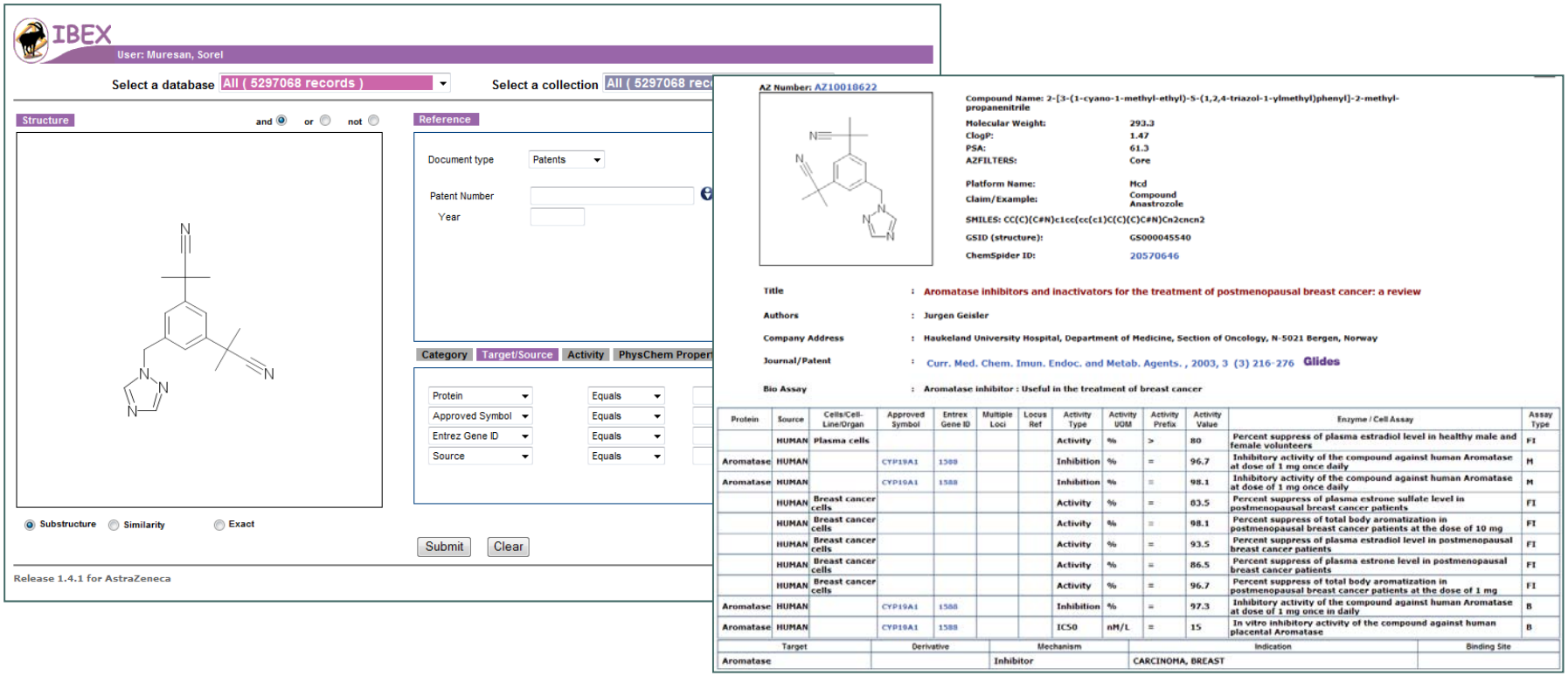
SureChemOpen

PubChem

ChemSpider
The free chemical database

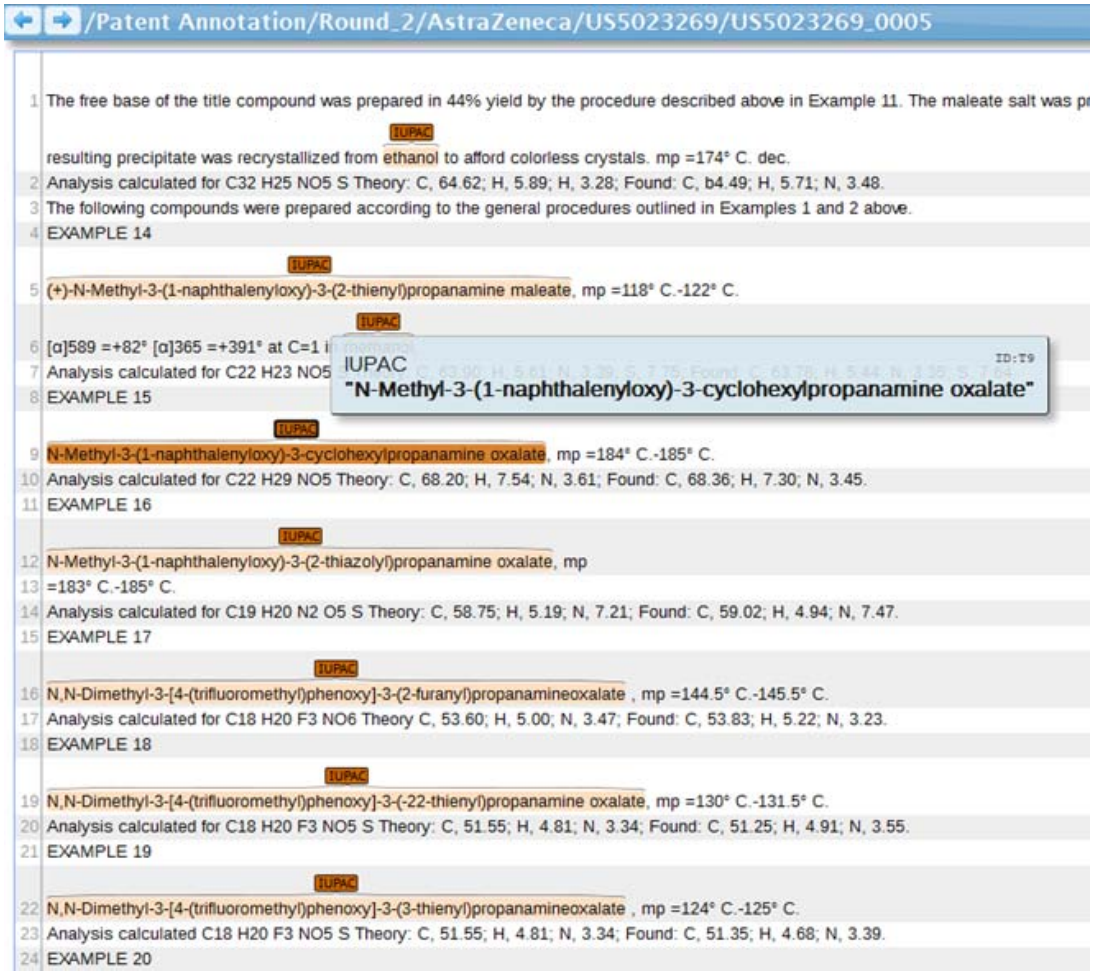
Manually extracted SAR data (commercial)

- GOSTAR (*GVKBIO Online Structure Activity Relationship Database*) is a comprehensive database that captures explicit relationships between the three entities of publications, compounds and sequences.
- It includes 2.6 million compounds linked to 3,500 sequences with 12.5M SAR points extracted from 43,000 patents and 67,000 articles from 125 journals



Annotate patents manually

Brat – rapid annotation tool (http://brat.nlplab.org/index.html)



Annotate patents manually

Brat – rapid annotation tool (http://brat.nlplab.org/index.html)

EXAMPLE 15

N-Methyl-3-(1-naphthalenyloxy)-3-cyclohexyl-propanamine oxalate, mp =184° C.–185° C.
Analysis calculated for $C_{22}H_{29}NO_5$ Theory: C, 68.20; H, 7.54; N, 3.61; Found: C, 68.36; H, 7.30; N, 3.45.

8 EXAMPLE 15

IUPAC

9 N-Methyl-3-(1-naphthalenyloxy)-3-cyclohexylpropanamine oxalate, mp =184° C.-185° C.

10 Analysis calculated for C22 H29 NO5 Theory: C, 68.20; H, 7.54; N, 3.61; Found: C, 68.36; H, 7.30; N, 3.45.

```
T6  M 567 634 (+)-N-Methyl-3-(1-naphthalenyloxy)-3-(2-thienyl)propanamine maleate
T8  M 693 701 methanol
T9  M 841 903 N-Methyl-3-(1-naphthalenyloxy)-3-cyclohexylpropanamine oxalate
T11 M 1043 1108 N-Methyl-3-(1-naphthalenyloxy)-3-(2-thiazolyl)propanamine oxalate
T13 M 1252 1327 N,N-Dimethyl-3-[4-(trifluoromethyl)phenoxy]-3-(2-furanyl)propanamineoxalate
T16 M 1474 1552 N,N-Dimethyl-3-[4-(trifluoromethyl)phenoxy]-3-(-22-thienyl)propanamine oxalate
T18 M 1699 1774 N,N-Dimethyl-3-[4-(trifluoromethyl)phenoxy]-3-(3-thienyl)propanamineoxalate
T21 M 1916 1987 N-Methyl-3-[4-(trifluoromethyl)phenoxy-3-(2-thienyl)propanamine oxalate
```

Name-2-structure

OPSIN – http://opsin.ch.cam.ac.uk/

OPSIN: Open Parser for Systematic IUPAC nomenclature

University of Cambridge › Department of Chemistry › Unilever Centre for Molecular Science Informatics

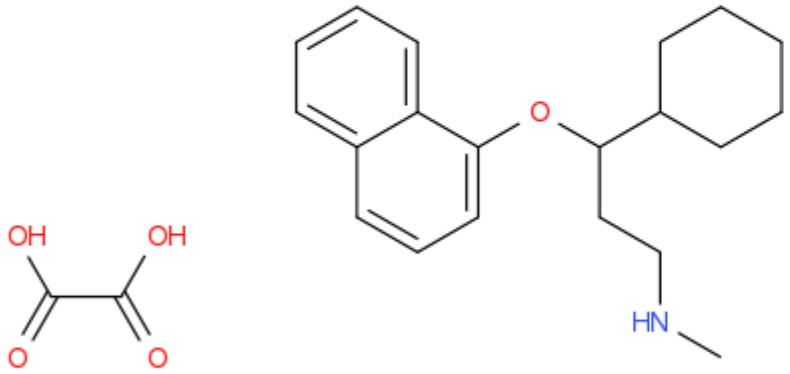
Enter a chemical name into the box and then click submit. If the name can be interpreted, a depiction, a SMILES string, its InChI and its CML will be returned.

N-Methyl-3-(1-naphthalenyloxy)-3-cyclohexylpropanamine oxalate [Submit]

Updated 30/5/12: Minor bug fixes and minor vocabulary improvements

A paper describing OPSIN is now available from JCIM. If you have found OPSIN useful in your work citing it would be very much appreciated.

Depiction courtesy of the Indigo Toolkit



InChI:

InChI=1/C20H27NO.C2H2O4/c1-21-15-14-19(17-9-3-2-4-10-17)22-20-13-7-11-16-8-5-6-12-18(16)20;3-1(4)2(5)6/h5-8,11-13,17,19,21H,2-4,9-10,14-15H2,1H3;(H,3,4)(H,5,6)/f/h;3,5H

SMILES:

C(C(=O)O)(=O)O.CNCCC(C1CCCCC1)OC1=CC=CC2=CC=CC=C12

Chemical Named Entity Recognition (CNER)

tion (10) **Pub. No.: US 2008/0038386 A1**
(43) **Pub. Date: Feb. 14, 2008**

(52) **U.S. Cl.** **424/755**; 514/514; 424/769; 424/760; 424/773; 514/617

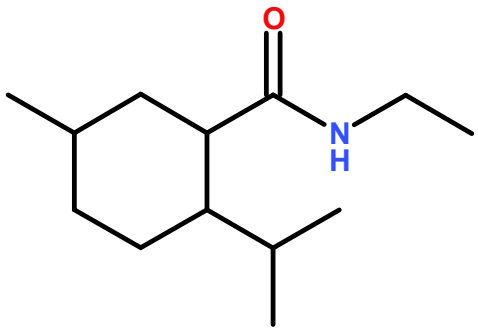
(57) **ABSTRACT**

The present invention is directed to a sensate composition consisting of a liquid cooling composition, and at least one warming, tingling, or pungent sensate ingredient of either synthetic or natural origin which provides enhanced warming, tingling, or pungent properties. The liquid cooling sensate of the invention is a eutectic mixture of 2-Isopropyl-N,2,3-trimethylbutyramide and N-Ethyl-p-menthane-3-carboxamide, or at least one cooling agent selected from the group consisting of N-(2-hydroxyethyl)-2,3-dimethyl-2-isopropylbutanamide, N-(3-ethoxypropyl)-2,3-dimethyl-2-isopropylbutanamide, N-(3-isopropoxypropyl)-2,3-dimethyl-2-isopropylbutanamide, N-(3-butoxypropyl)-2,3-dimethyl-2-isopropylbutanamide,N-Ethyl-2,2-diisopropylbutanamide, N-(1,1-dimethyl-2-hydroxyethyl)-2,2-diethylbutanamide, N-(2-ethoxyethyl)-2,3-dimethyl-2-isopropylbutanamide and N-(3-methoxypropyl)-2,3-dimethyl-2-isopropylbutanamide. The warming, tingling, or pungent sensate of the invention consists of at least one component that is an isothiocyanate and/or an amide or a natural product that contains at least one warming, tingling, or pungent principle that is an isothiocyanate and/or an amide. The present invention is further directed to a method of using the sensate composition in a food, pharmaceutical or personal care product.

N-Ethyl-p-menthane-3-carboxamide

Name-to-Structure
software

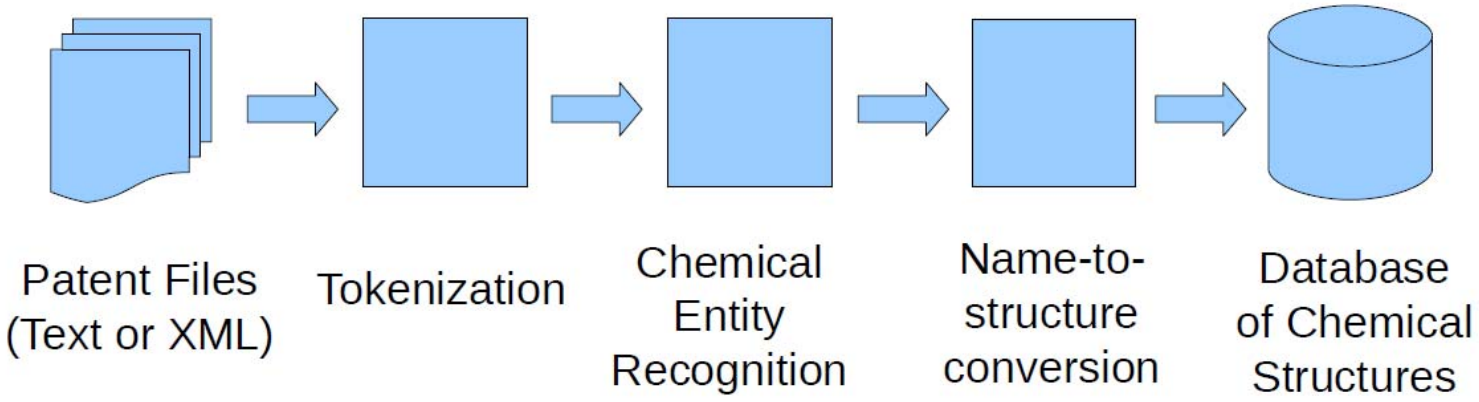
C(C)NC(=O)C1CC(CCC1C(C)C)C



WS3 cooling agent (CAS 39711-79-0)

Compound extraction via chemical text mining

Traditional text mining pipeline



- Determining the start and end of IUPAClike names in free text is a tricky business.
- Chemical names can contain whitespace, hyphens, commas, parenthesis, brackets, braces, apostrophes, superscripts, greek characters, digits and periods.
- This is made harder still by typos, OCR errors, hyphenation, linefeeds, XML tags, line and page numbers and similar noise.

OCR Errors: Compound Names

- lH-ben zimidazole → 1H-benzimidazole
- 4- (2-ADAMANTYLCARBAM0YL) -5-TERT-BUTYL-PYRAZOL-1-YL] BENZOIC ACID → 4-(2-adamantylcarbamoyl)-5-tert-butyl-pyrazol-1-yl]benzoic acid
- triphenylposhine → triphenylphosphine

Text Mining Conversion by Name Class

Class	Category	Names		n2s_3 (%)	None (%)
M	Molecule	7,262,798	81.4	64.8	77.1	7.8
D	Dictionary	26,876	38.1	45.1	3.5	38.5
R	Registry number	304,064	0	0	0	100
C	CAS number	47,815	0	0	0	100
E	Element	836				
P	Fragment	2,663,677				
A	Atom fragment	96				
Y	Polymer	295	0	44.1	22.7	36.9
G	Generic	1263	2.6	6.3	0.5	91.9
N	Noise	104	32.7	24	19.2	52.9
	Total	**10,307,824**	**76.3**	**61.0**	**54.3**	**14.1**

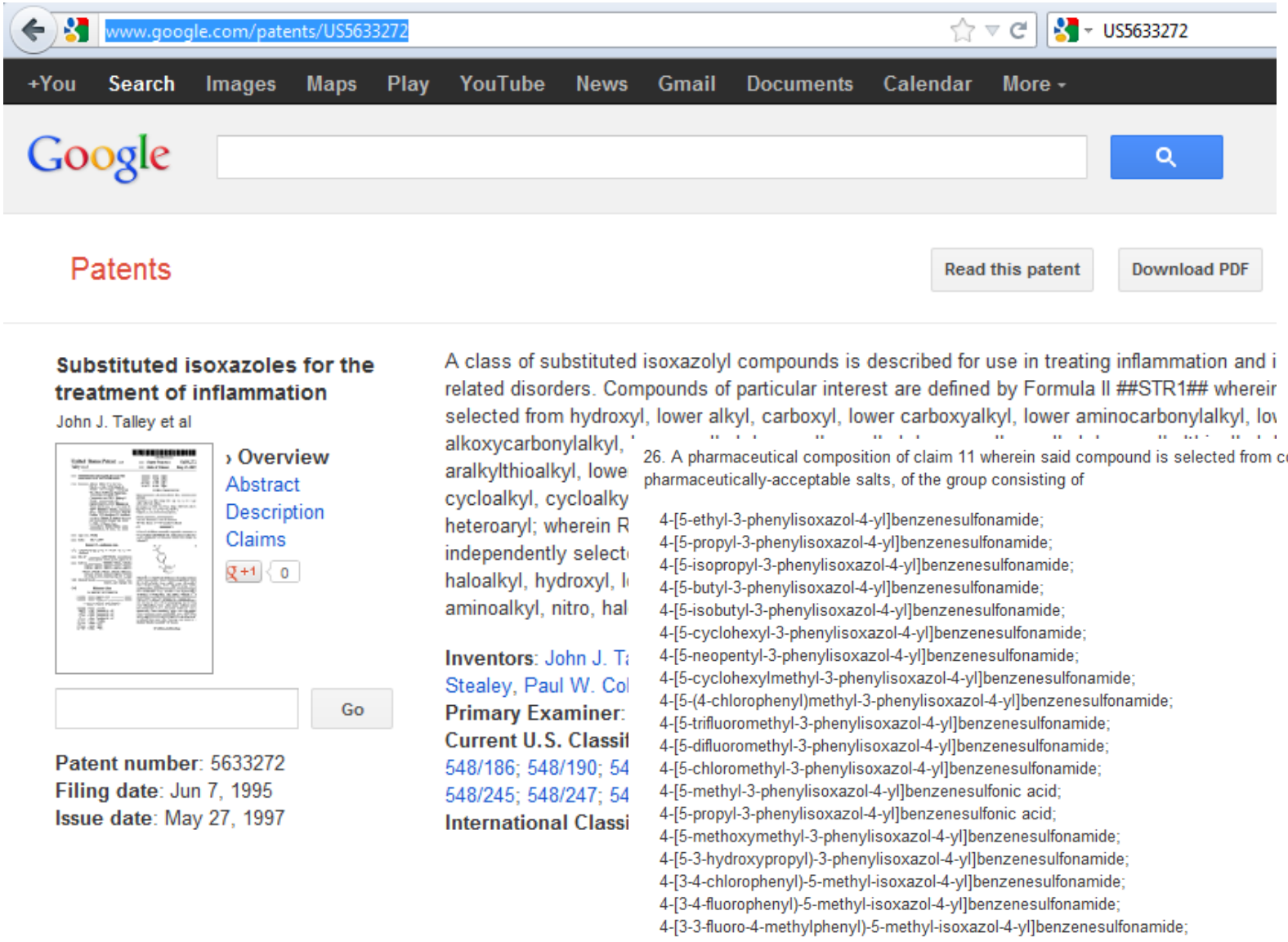
ChemAxon 5.5 converts 60%

NCI/CADD Chemical Identifier Resolver converts 48%

Muresan, S. *ChemAxon UGM 2011*

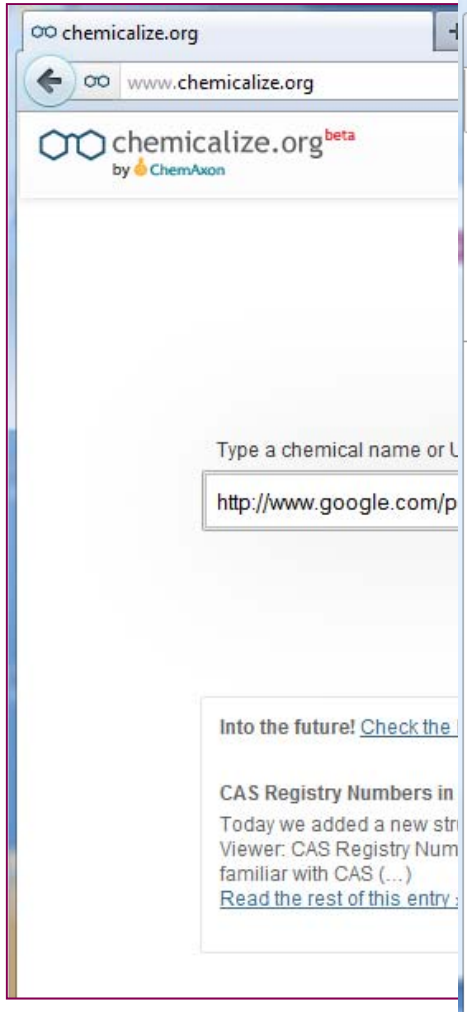
Extraction compounds from US20100221398

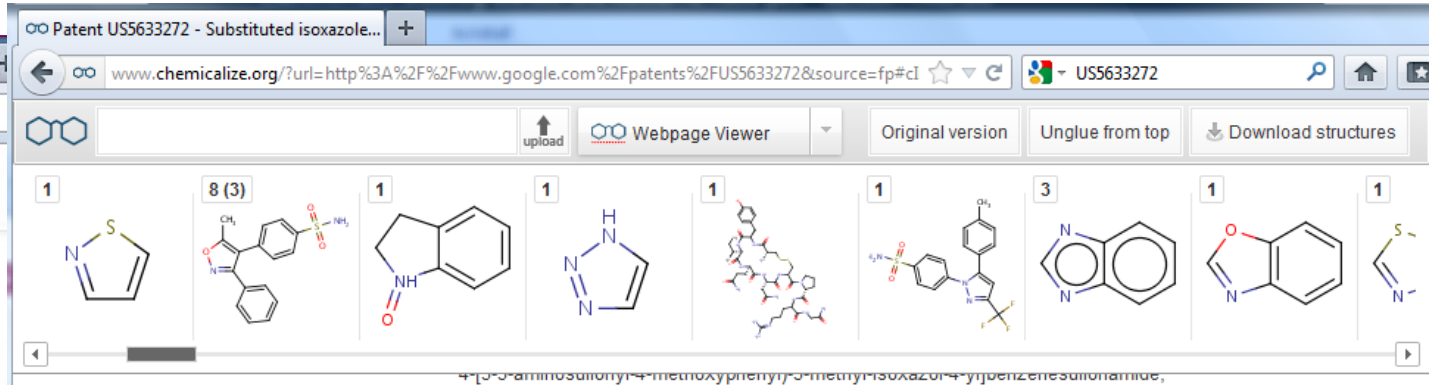
Google Patents + Chemicalize.org (ChemAxon)



Extraction compounds from US20100221398

Google Patents + Chemicalize.org (ChemAxon)





4-[3-3-chloro-4-methylphenyl)-5-methyl-isoxazol-4-yl]benzenesulfonamide;
4-[3-3-fluorophenyl)-5-methyl-isoxazol-4-yl]benzenesulfonamide;
4-[5-hydroxymethyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
[4-[4-(aminosulfonyl)phenyl]-3-phenylisoxazol-5-yl]carboxylic acid;
4-[5-hydroxy-3-phenyl-4-isoxazolyl]benzenesulfonamide;
4-[3-methyl-5-phenyl-isoxazol-4-yl]benzenesulfonamide;
4-[5-methyl-3-phenyl-isoxazol-4-yl]benzenesulfonamide;
4-[3-(3-fluoro-4-methoxyphenyl)-5-methyl-isoxazol-4-yl]benzenesulfonamide;
[3-(3-chloro-4-methoxyphenyl)-4-[4-(methylsulfonyl)phenyl]isoxazol-5-yl]acetic acid;
5-methyl-4-[4-(methylsulfonyl)phenyl]-3-phenyl-isoxazole;
3-(3-chloro-4-methoxyphenyl)-5-methyl-4-[4-(methylsulfonyl)phenyl]isoxazole
[4-[4-(aminosulfonyl)phenyl]-3-phenyl-isoxazol-5-yl]acetic acid;
[4-[4-(aminosulfonyl)phenyl]-3-phenyl-isoxazol-5-yl]propanoic acid;
ethyl [4-[4-(aminosulfonyl)phenyl]-3-phenyl-isoxazol-5-yl]propanoate;
[3-(3-fluoro-4-methoxyphenyl)-4-[4-(methylsulfonyl)phenyl]isoxazol-5-yl]acetic acid; and
[4-[4-(aminosulfonyl)phenyl]-3-(3-fluoro-4-methoxyphenyl)isoxazol-5-yl]propanoic acid.

25. The compound of claim 4 which is 4-[5-hydroxymethyl-3-phenylisoxazol-4-yl]benzenesulfonamide, or a pharmaceutically-acceptable salt thereof.

26. A pharmaceutical composition of claim 11 wherein said compound is selected from compounds, or their pharmaceutically-acceptable salts, of the group consisting of

4-[5-ethyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-propyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-isopropyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-butyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-isobutyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-cyclohexyl-3-phenylisoxazol-4-yl]benzenesulfonamide;
4-[5-neopentyl-3-phenylisoxazol-4-yl]benzenesulfonamide;

What is a key compound?

Drug candidate

Compound(s) with optimal physicochemical properties

Compouns(s) with the most suitable pharmacokinetic profile

The most biologically active tool or probe

Key compounds from patents

”Old school” techniques

Look for clues in text: “most preferred” wording in claims; crystal form info; scale of synthesis

Structural information alone

Frequency of group (FOG) analysis of exemplified compounds

Structures and SAR data

Work out SAR using biological data and structures

Key compound prediction from patents

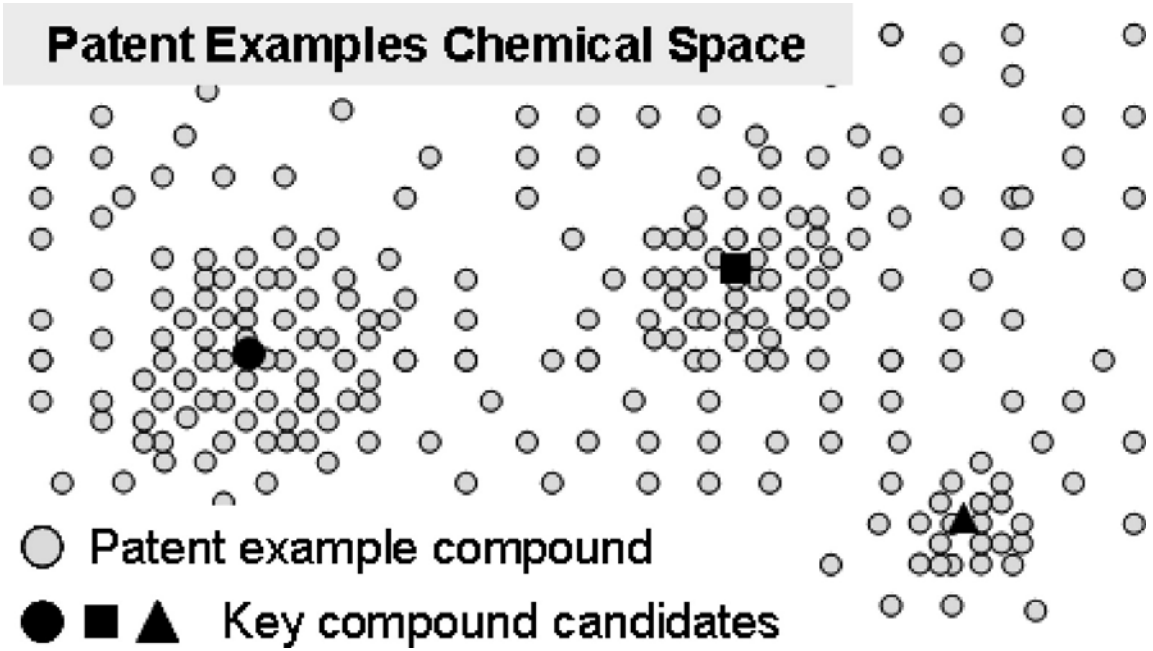


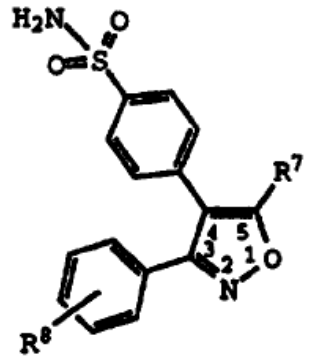
Figure 2. Graphical image of patent example compounds in chemical space. Each gray circle represents an example compound. The black circle, square, and triangle represent key compound candidates.

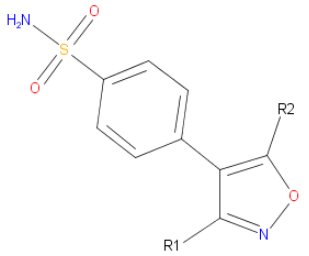
Theory: Chemists carry out extensive SAR around key compounds. Cluster examples and look for centres of densely populated regions

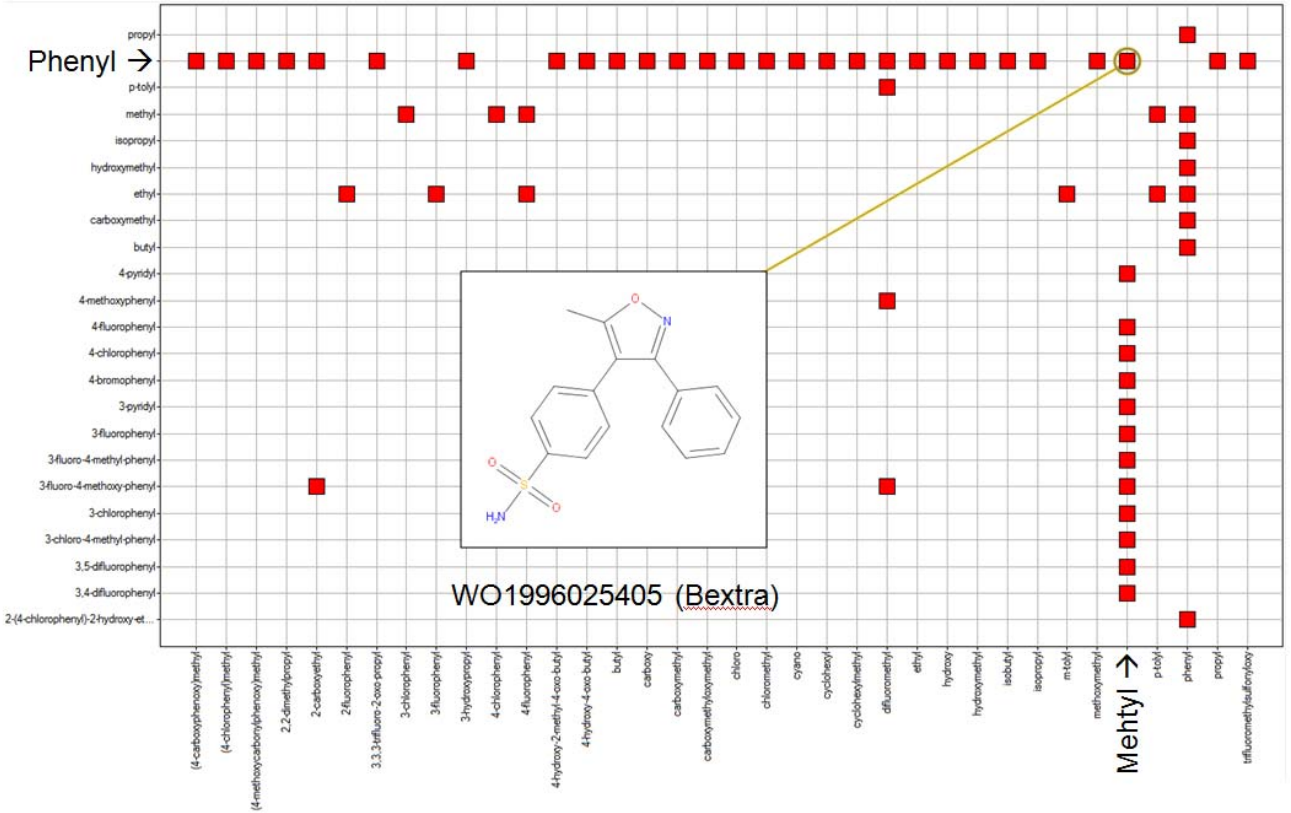
Kazunari Hattori; Hiroaki Wakabayashi; Kenta Tamaki; *J. Chem. Inf. Model.* 2008, 48, 135-142.
DOI: 10.1021/ci7002686

Key compound prediction from patents

From **WO1996025405** the earliest patent which claims it, can you work out the structure of Bextra (Valdecoxib), the Pfizer NSAID?

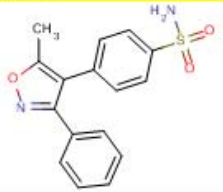
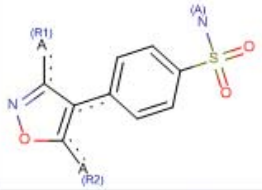
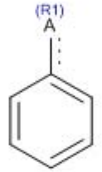
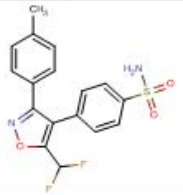
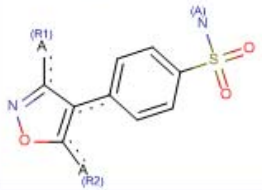
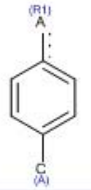
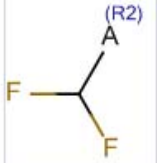
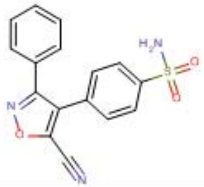
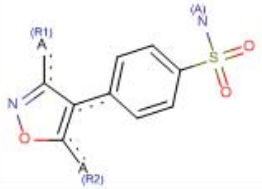
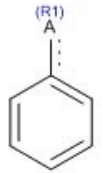
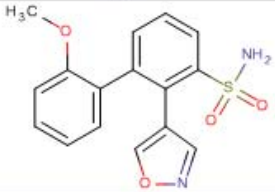
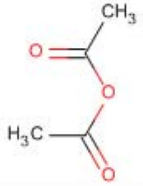




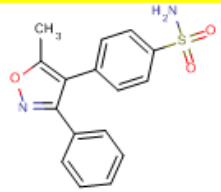
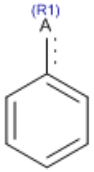
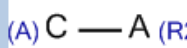
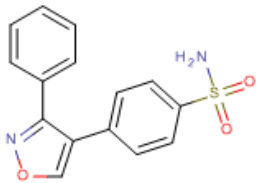
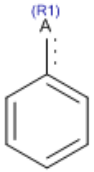
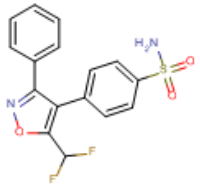
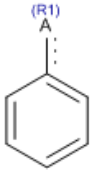
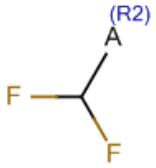
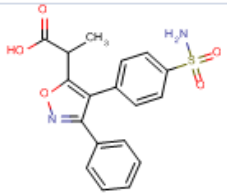
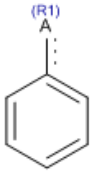
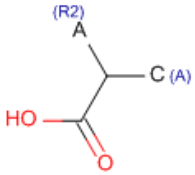
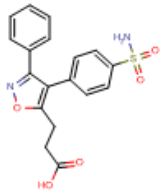
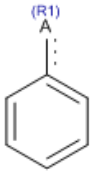
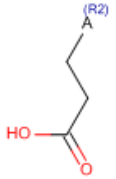


74 exemplified cmpds

R-group decomposition (Free-Wilson like)

A	B	C	D	E	F
cmpdID	**cmpdSmiles**	**error**	**coreSmiles**	**R1**	**R2**
200000516_BEXTRA		None			
208014805		None			
206771831		None			
256502606		Unable to map core	None		
200271239		Unable to map core	None		

FOG ranking

cmpdID	smiles	score	R1	R1_count	R2	R2_count
200000516_BEXTRA		**54**		**34**		**20**
256227569		38		34		4
200004495		38		34		4
220904851		36		34		2
207700988		36		34		2

Key compound prediction from patents

Frequency of group (FOG) analysis

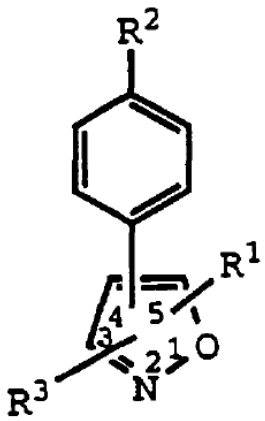
1) Extract compounds from patents
   - manual curation (GVKBIO)
   - text mining (SureChemOpen, OSCAR)

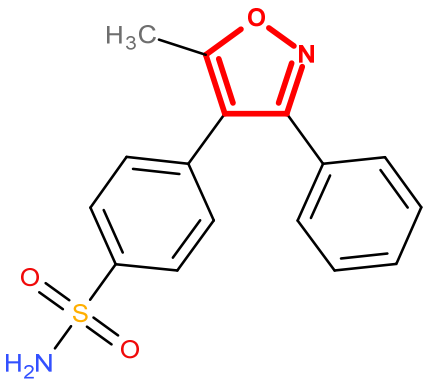
2) Define core
   - manually (e.g. from patent Markush)
   - automated core perception

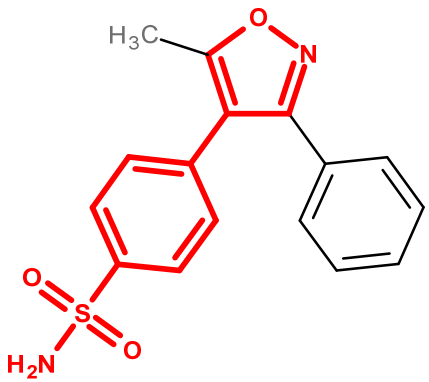
3) R-group decomposition (ChemAxon's JChem)

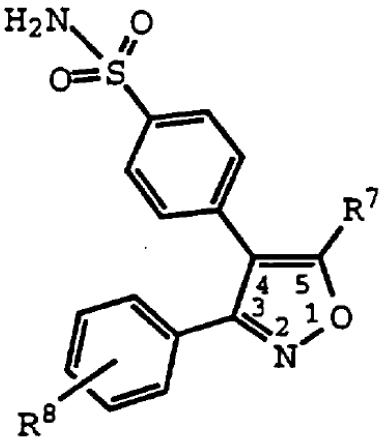
4) Rank compounds based on FOG (Spotfire, EXCEL)

WO1996025405 - Bextra



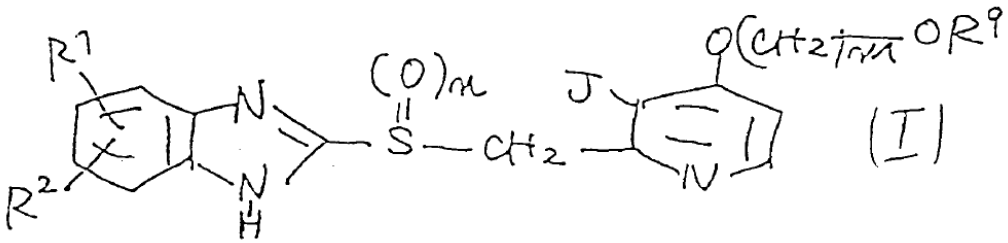


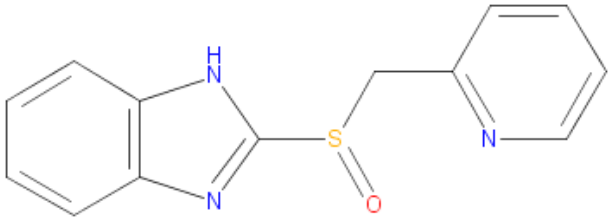


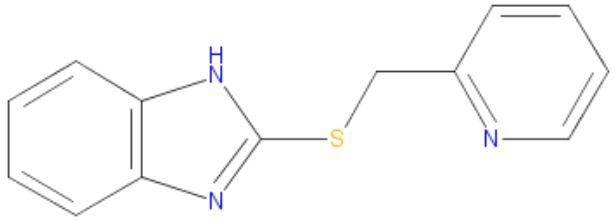


Source	#compounds	Bextra exists	Bextra ranked
GVKBIO	74	Y	1 (broad core) 1 (narrow core)
SureChem	501	Y	1 (broad core) 1 (narrow core)

EP268956 - Aciphex

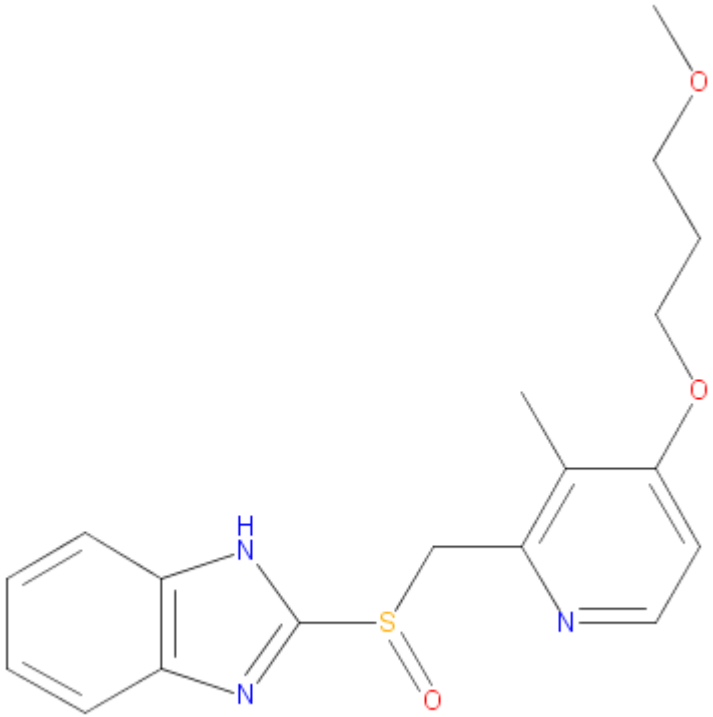


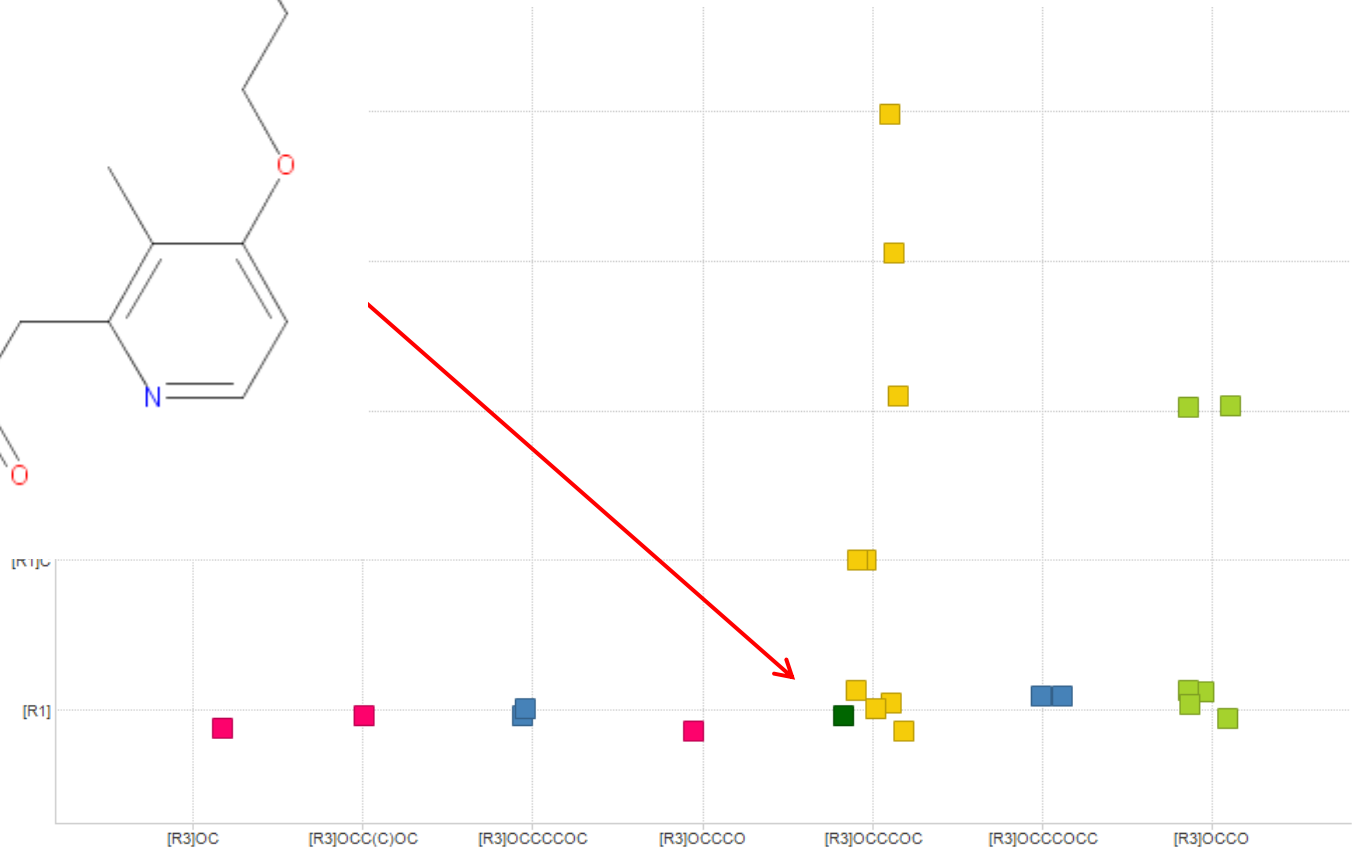




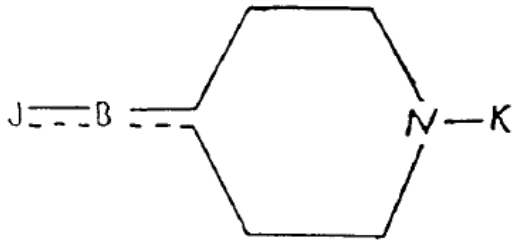
Source	#compounds	Aciphex exists	Aciphex ranked
GVKBIO	27	Y	2 (core1) 1 (core2)
SureChem	168	Y	1 (core1) 1 (core2)

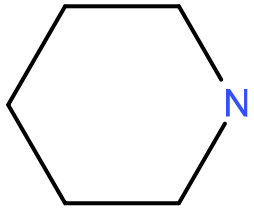
EP268956 - Aciphex

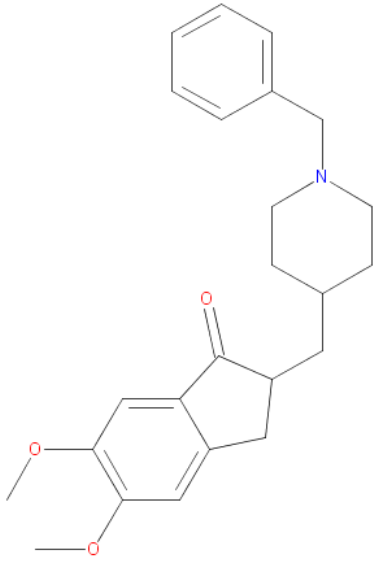




EP296749 - Aricept

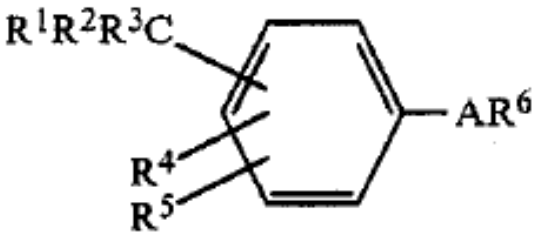


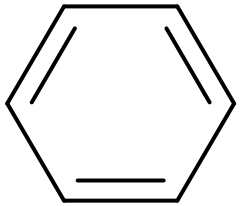


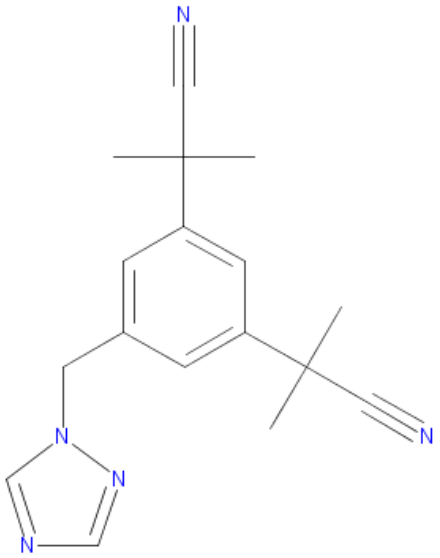


Source	#compounds	Aricept exists	Aricept ranked
GVKBIO	76	Y	1
SureChem	108	Y	1

EP296749 - Arimidex

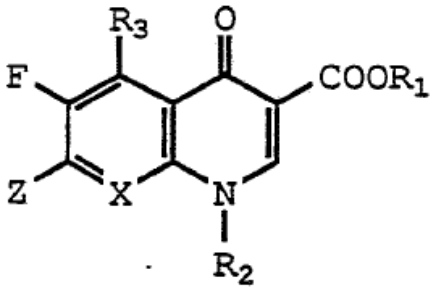


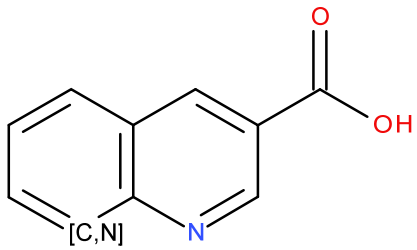


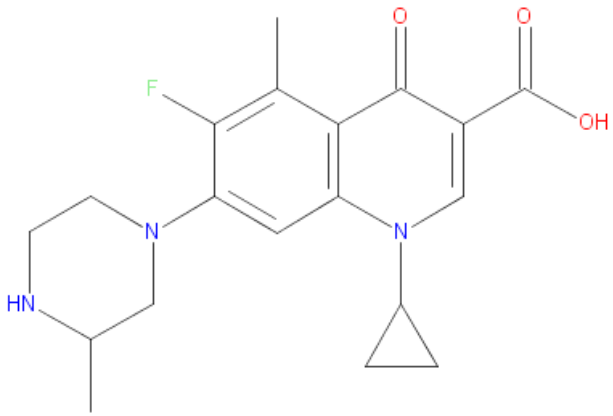


Source	#compounds	Arimidex exists	Arimidex ranked
GVKBIO	70	Y	1
SureChem	267	Y	1

WO1989006649 - Raxar



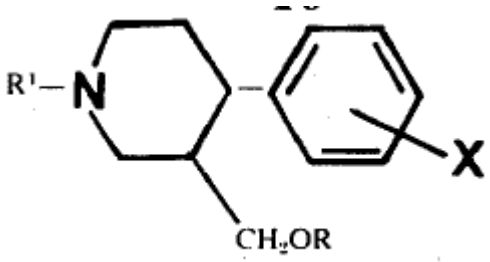


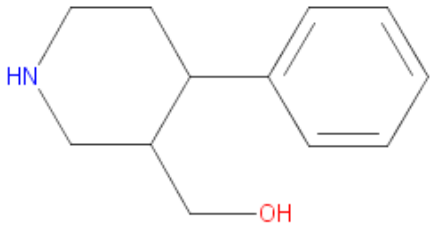


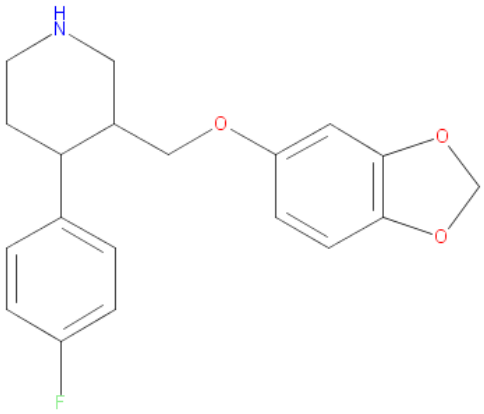
X is CH, CF, CCl, or N.

Source	#compounds	Raxar exists	Raxar ranked
GVKBIO	102	Y	5
SureChem	0	N	n/a

US3912743 – PAXIL

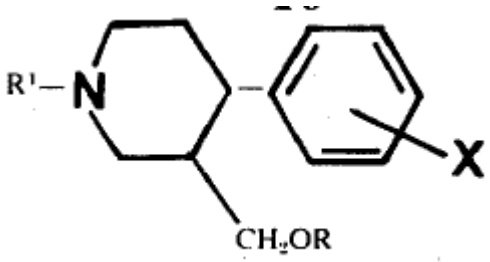


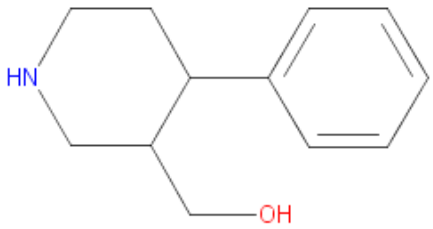


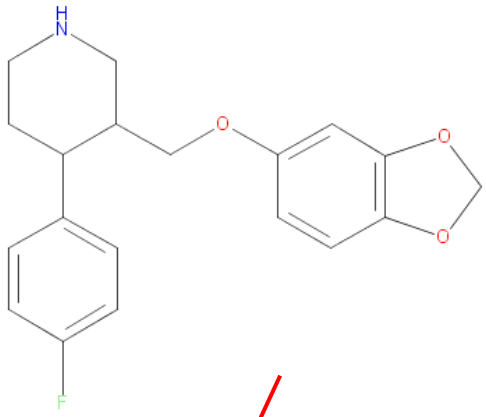


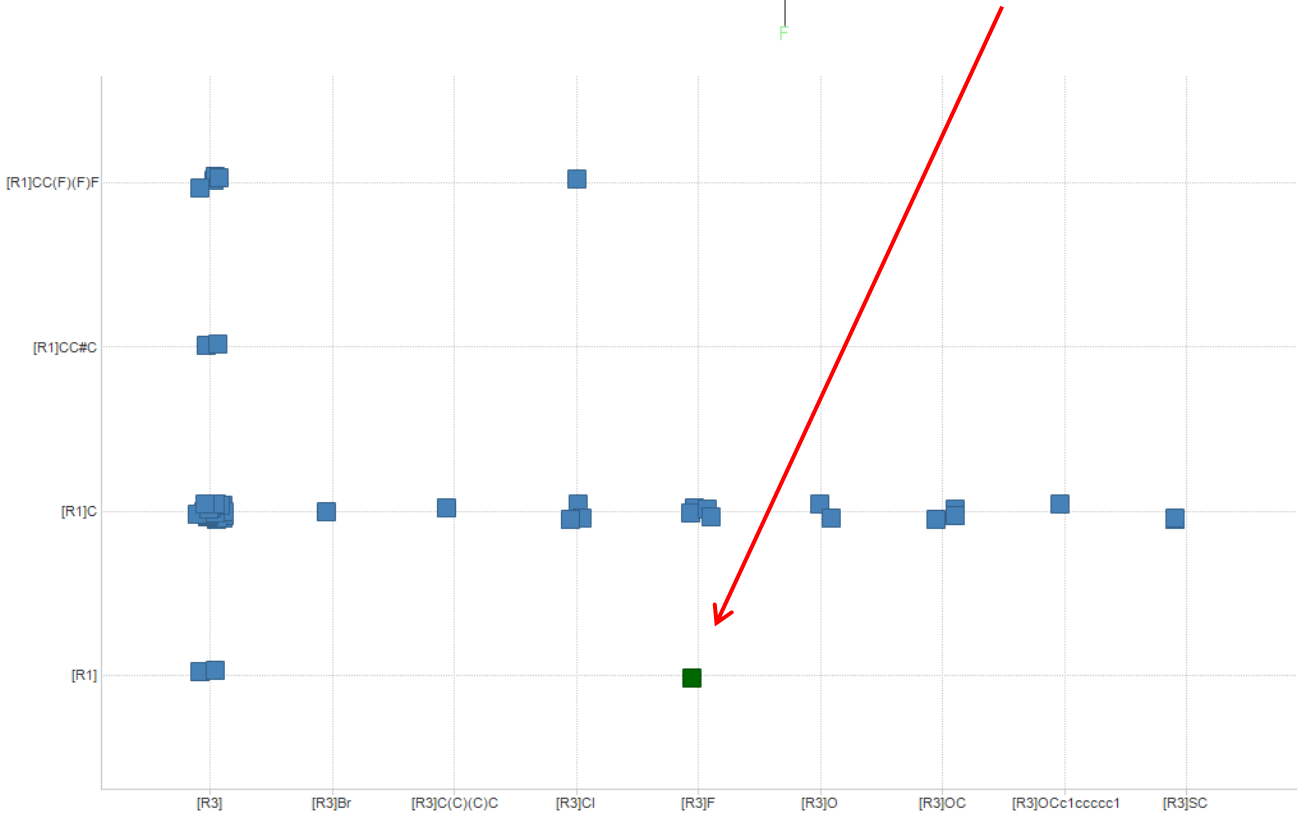
Source	#compounds	Paxil exists	Paxil ranked
GVKBIO	60	Y	54

US3912743 – PAXIL





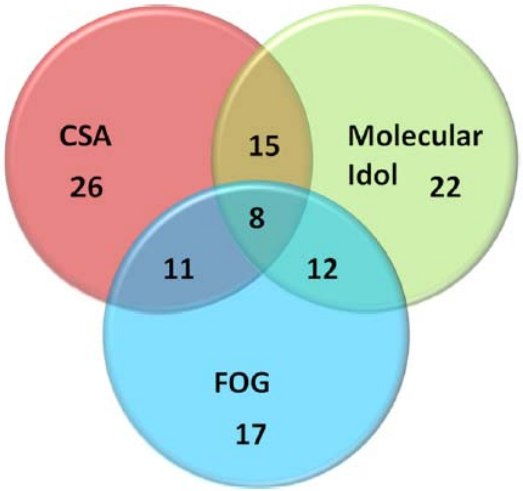




Key compound prediction

48 drug patents, with more than 10 compounds, including the drug

Method	Key compound as the first ranked compound	Key compound within the first 5 ranked compounds
Cluster seed analysis	11 (23%)	26 (54%)
Molecular Idol	5 (10%)	22 (46%)
Frequency of group analysis	11 (23%)	17 (35%)



JCIM paper

http://dx.doi.org/10.1021/ci3001293

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Exploiting Structural Information in Patent Specifications for Key Compound Prediction

Christian Tyrchan,*[,†] Jonas Boström,[†] Fabrizio Giordanetto,[†] Jon Winter,[§] and Sorel Muresan[‡]

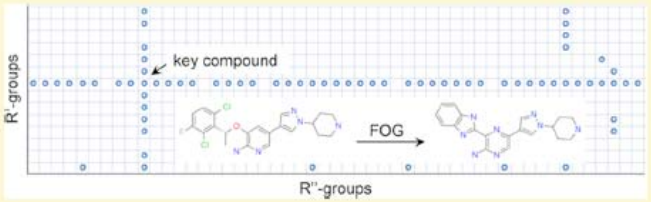
[†]AstraZeneca R&D, CVGI iMed, Pepparedsleden 1, S-431 83 Mölndal, Sweden

[‡]AstraZeneca R&D, Discovery Sciences Computational Sciences, Pepparedsleden 1, S-431 83 Mölndal, Sweden

[§]AstraZeneca R&D, Oncology iMed, Alderley Park, Macclesfield, Cheshire, SK10 4TG, United Kingdom

Supporting Information

ABSTRACT: Patent specifications are one of many information sources needed to progress drug discovery projects. Understanding compound prior art and novelty checking, validation of biological assays, and identification of new starting points for chemical explorations are a few areas where patent analysis is an important component. Cheminformatics methods can be used to facilitate the identification of so-called key compounds in patent specifications. Such methods, relying on



Summary

- Unique chemistry from patents (8% out of 55M parent structures in AstraZeneca's Chemistry Connect)
- Public sources for full text patent access and open source software for chemical text mining
- Frequency of group analysis is a simple process to visualize and rank patent compounds

Acknowledgements

- Jon Winter
- Christian Tyrchan
- Jonas Boström
- Mats Eriksson
- Paul Xie
- Chris Southan

- GVKBIO
- IBM
- SureChem

- Next Move Software
- ChemAxon
- Unilever Centre for Molecular Science Informatics