SureChem

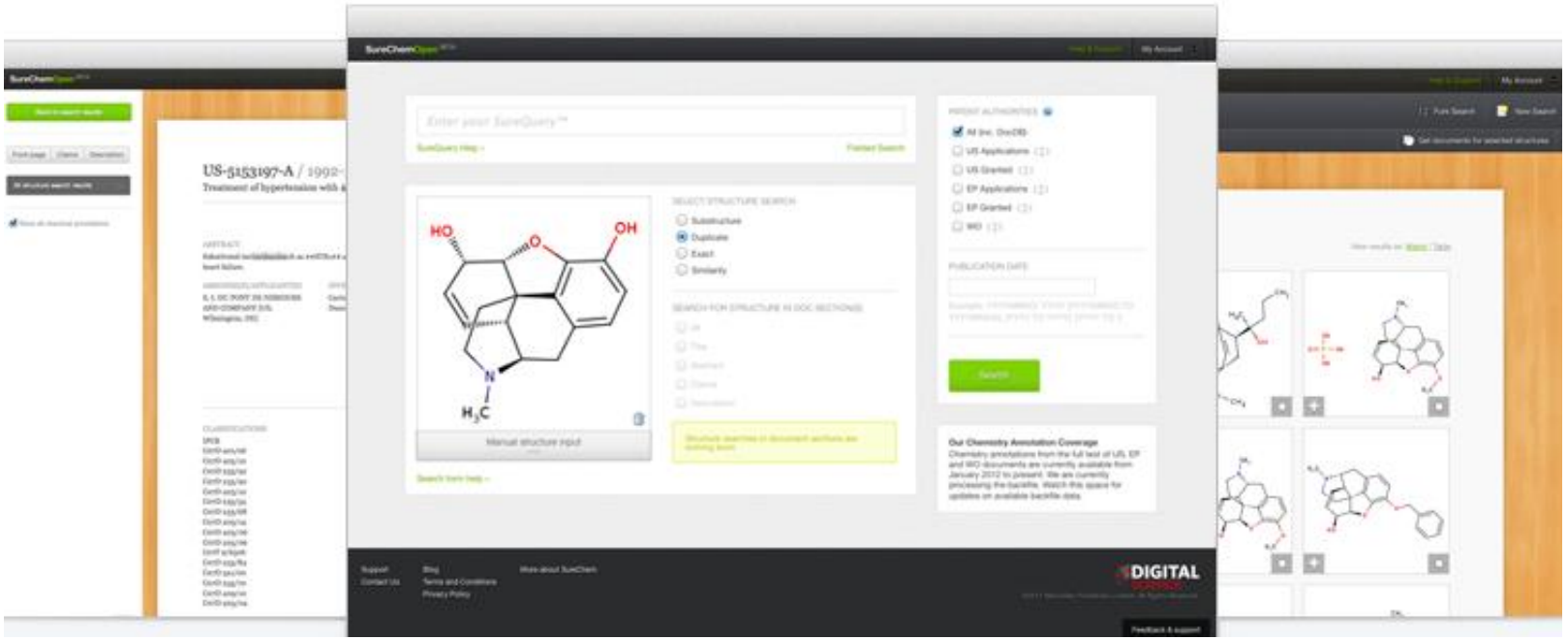
Integrating patent chemistry with public and private non-patent research resources

Andrew Hinton, PhD
Christopher Southan, PhD
Evan Bolton, PhD
Nicko Goncharoff

ICIC 2012
17 October

Patent chemistry made easy and accessible

We're integrating patent chemistry into the scientific community and giving customers control over data



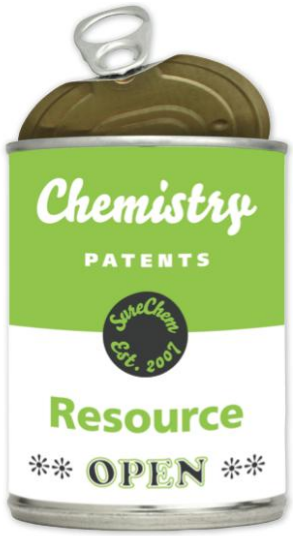
SureChem Data Collection

Database of automatically mined structure data from text and images

- 20M annotated US, EP, WO full text records and Japan patent abstracts
- 12.8M unique chemical structures
- MEDLINE – 19M abstracts (upcoming)

SureChemOpen

- Free resource for researchers
- Enables linking to public and proprietary content



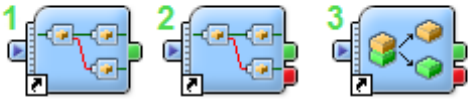
Science is better when data is opened up
Welcome to **SureChemOpen**

SureChemPro

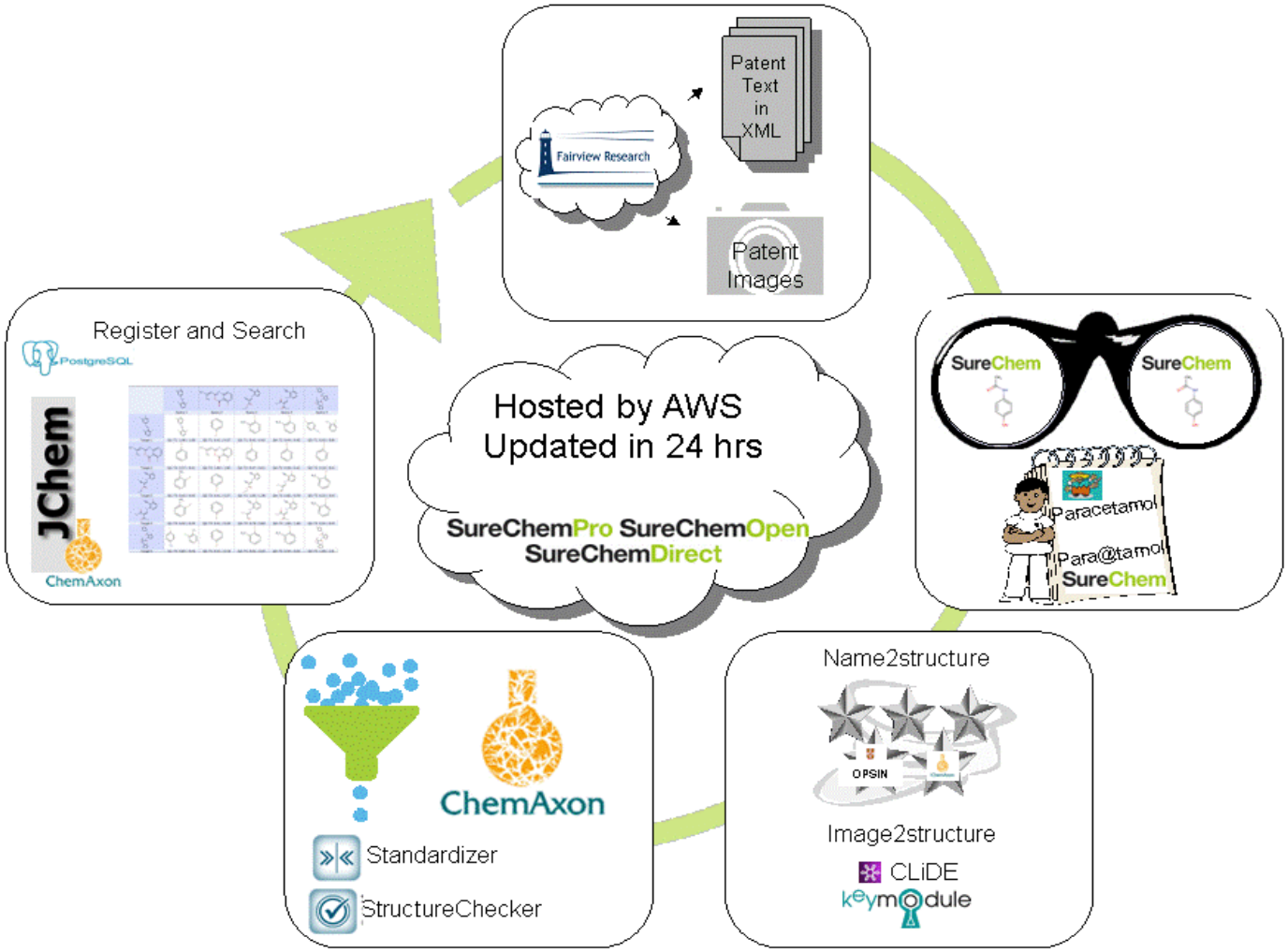
- Professional search needs
- Data export, alerts, patent family search, chemical relevance filters…

SureChemDirect

- API or Data Feed access to chemistry & full text
- Integrate with internal databases & workflows



Chemistry Mining Workflow



Public Patent Chemistry – A Changing Landscape

SureChem Depositing All* Structures into PubChem – Q4 2012

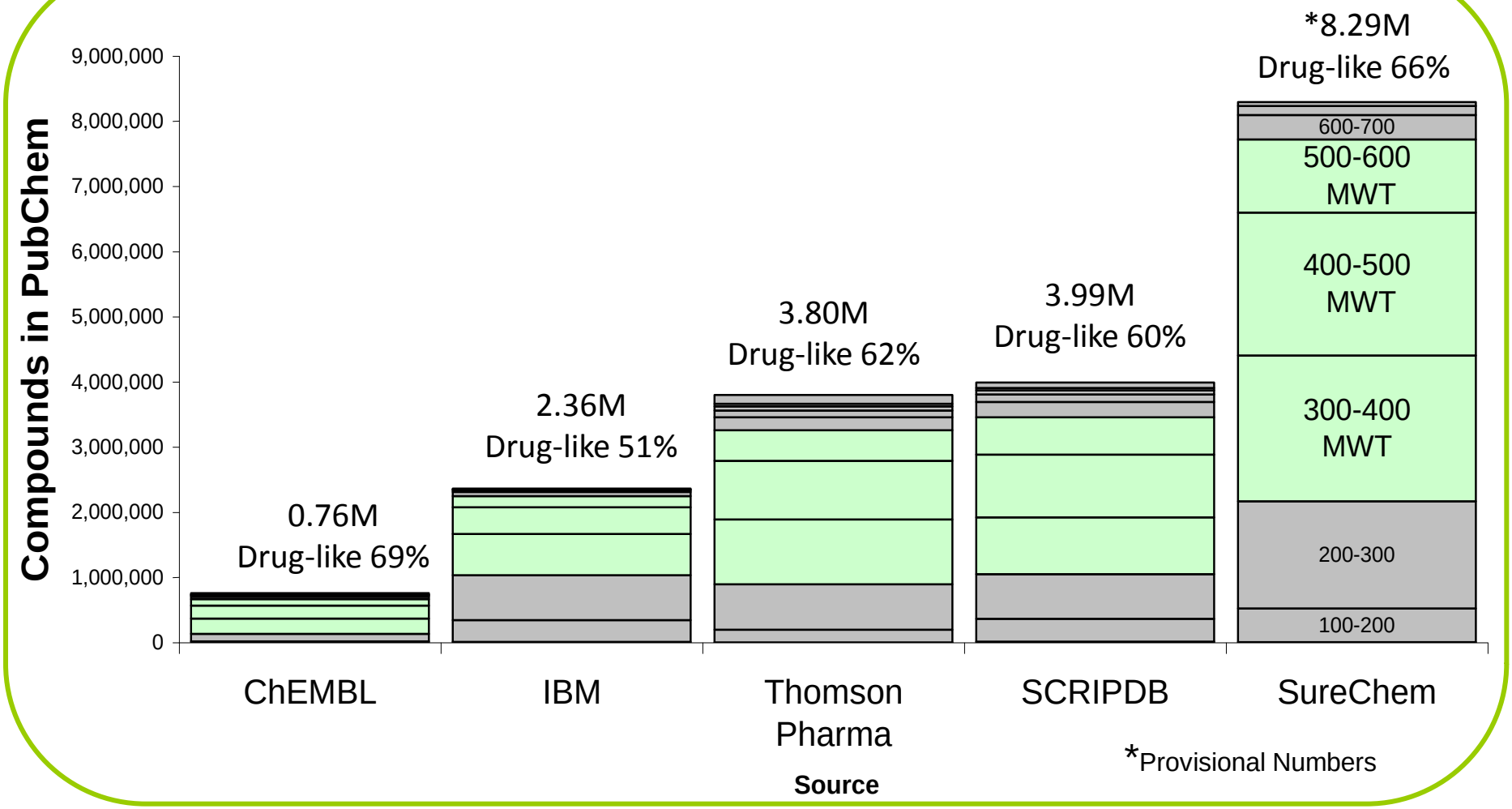
SureChem

PubChem

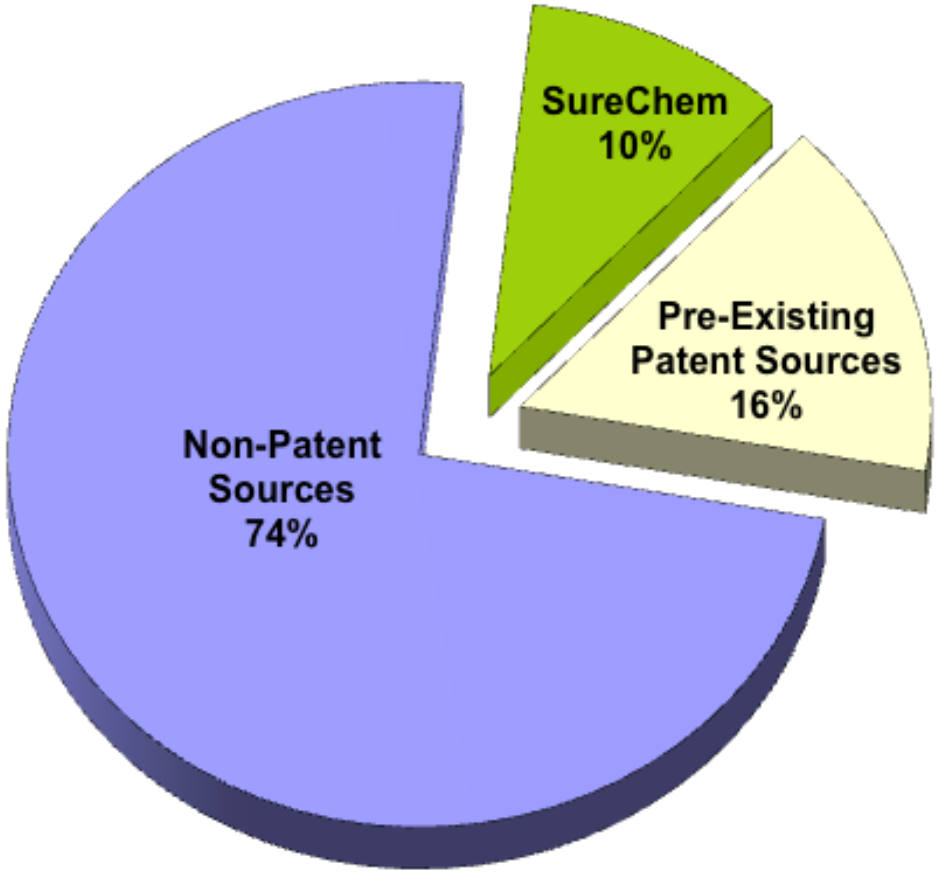
- 1976 to present
- Deposition of structures only
- Currently ‘on hold’
- Will link to patents in **SureChem**Open

* After filtering of fragments and highly common chemistry

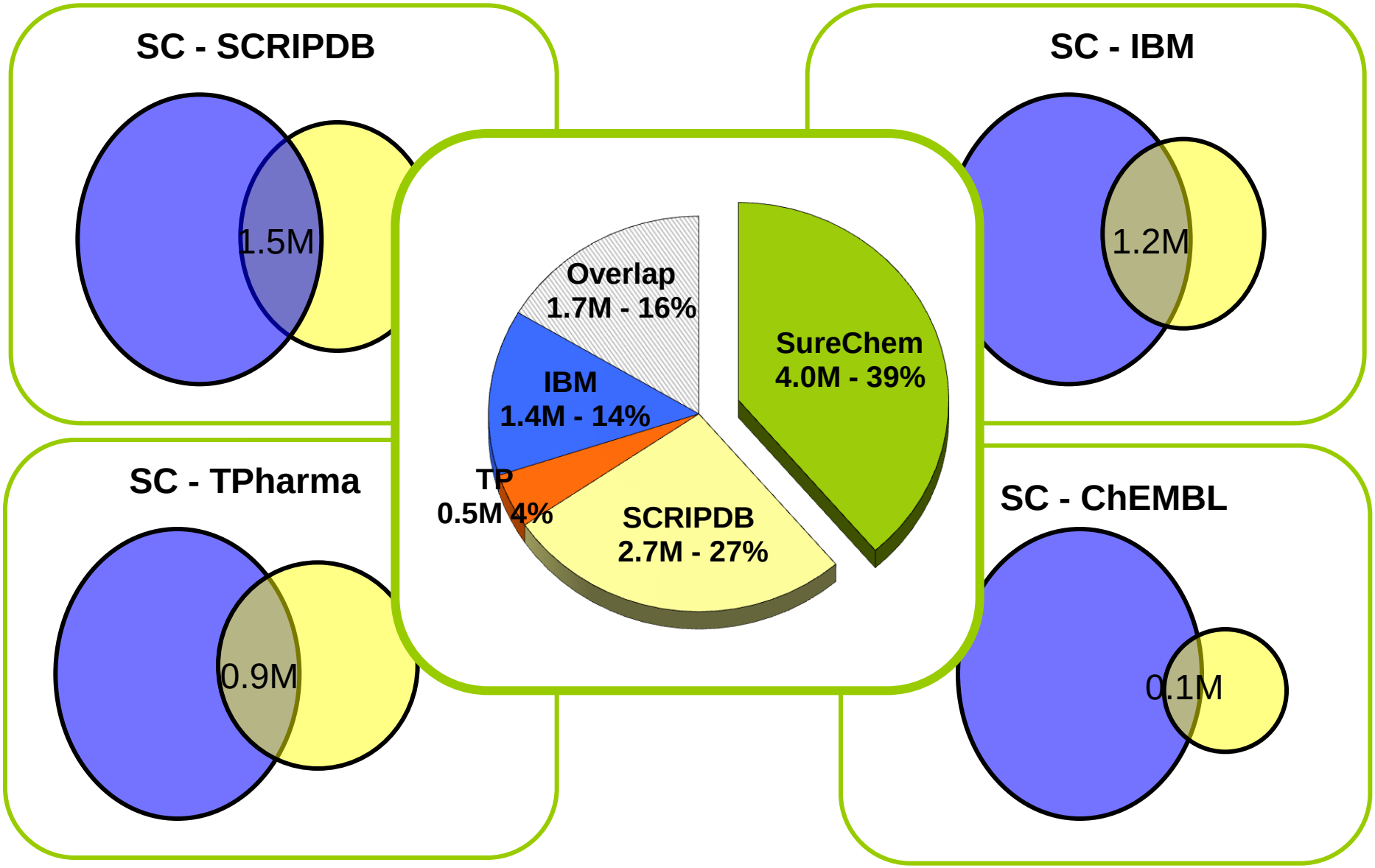
Compounds Derived from Patents and Literature found in PubChem By Molecular Weight Range (MWT) and Source



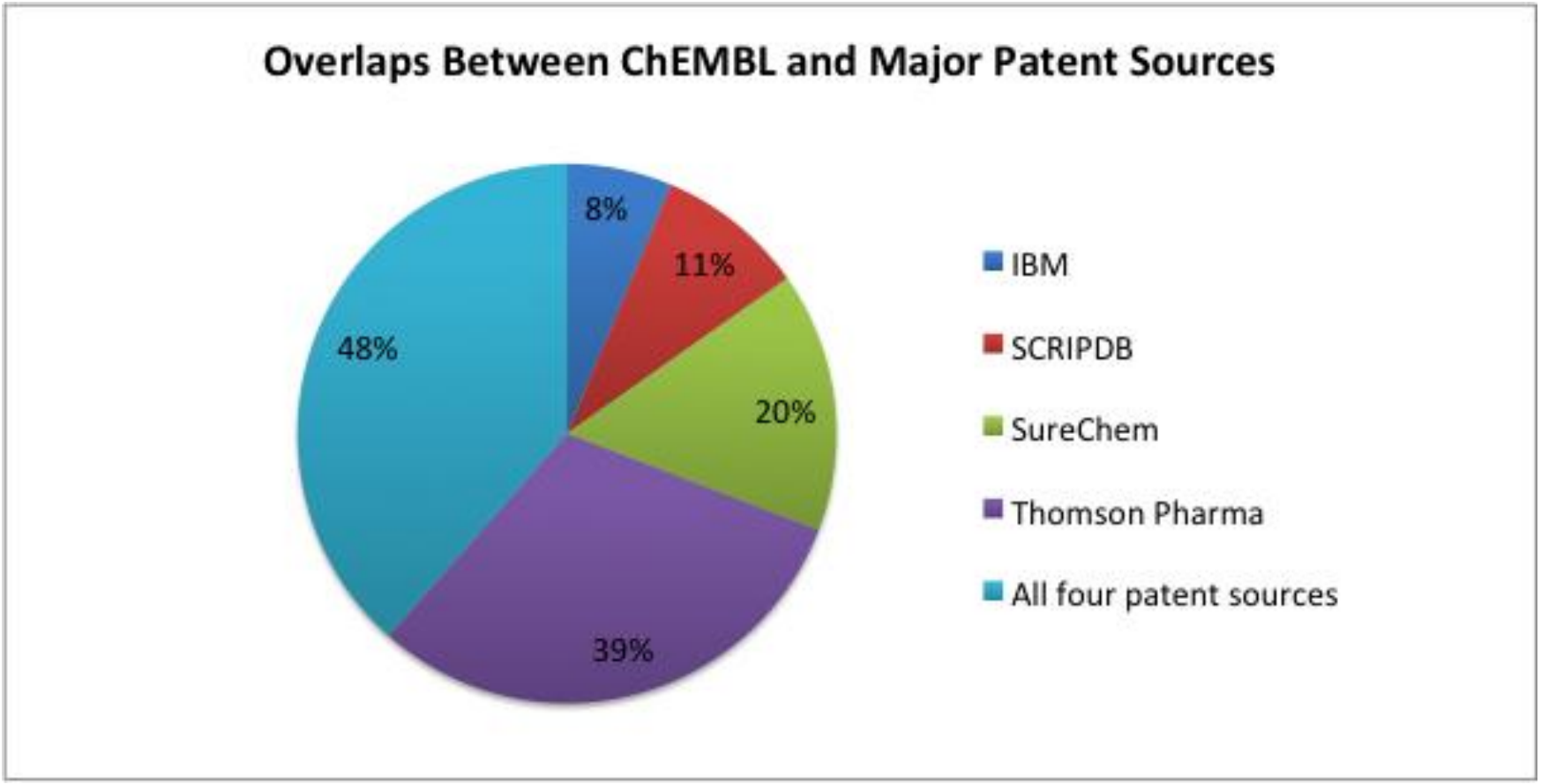
SureChem Deposition Pushes PubChem to 40 Million Compounds



Uniques and Overlaps

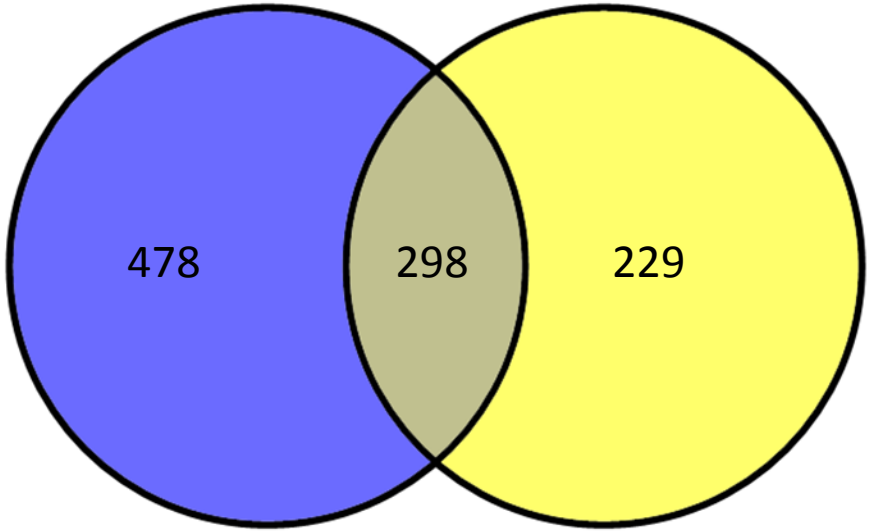


ChEMBL overlaps with Patent Sources in PubChem



Intersects – Patent Document View (2 Examples – SC & IBM)

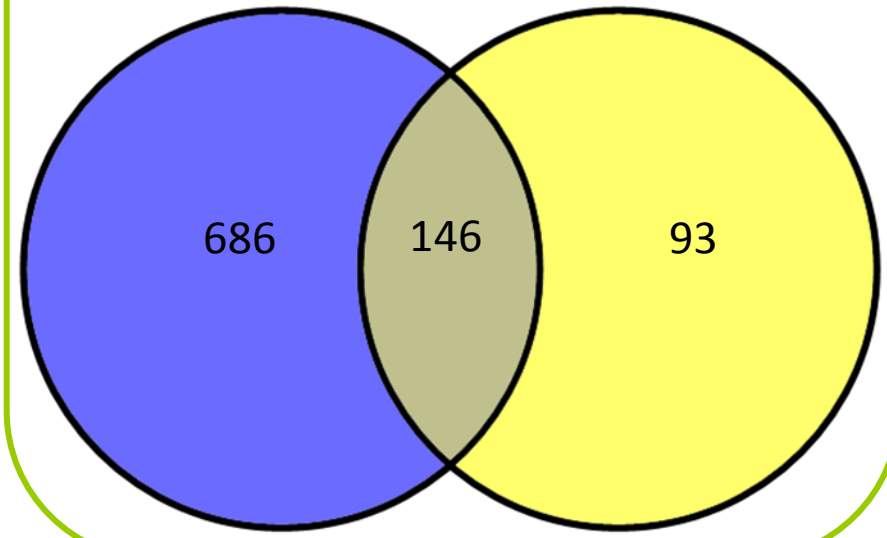
SureChem Total: 776 IBM Total : 527



WO-1994018188-A1
4-hydroxy-benzopyran-2-ones and 4-hydroxy-cycloalkyl[b]pyran-2-ones HIV protease inhibitors, **Upjohn**

US583593, Inhibitors of squalene synthetase and protein farnesyltransferase. **Abbott**

SureChem Total: 832 IBM Total: 239



Identifying Relevant Chemistry - IC_{50}

US-20120035195-A1 BACE2, Hoffman LaRoche

US 20120035195A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2012/0035195 A1**

Banner et al. (43) **Pub. Date: Feb. 9, 2012**

(54) **1,4,5,6-TETRAHYDRO-PYRIMIDIN-2-YLAMINE COMPOUNDS**

Publication Classification

(51) **Int. Cl.**

US 2012/0035195 A1

26

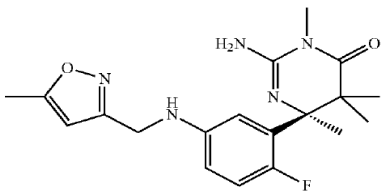
-continued

Example	IC_{50} [μM]
104	0.004[a]
105	0.012[a]
106	0.002[a]
107	0.002[a]

Example 104

(S)-2-Amino-6-{2-fluoro-5-[(5-methyl-isoxazol-3-ylmethyl)-amino]-phenyl}-3,5,5,6-tetramethyl-5,6-dihydro-3H-pyrimidin-4-one

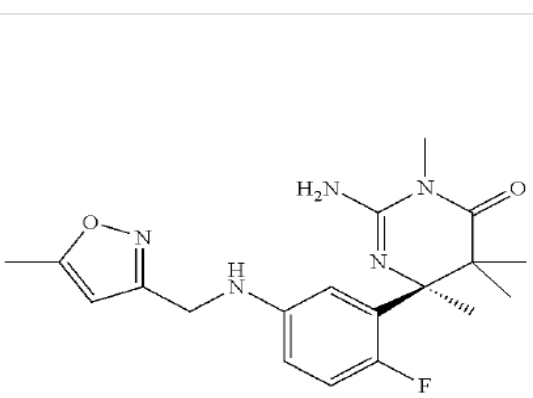
[0828]



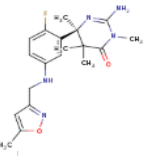
[0829] The reductive amination of (S)-2-amino-6-(5-amino-2-fluoro-phenyl)-3,5,5,6-tetramethyl-5,6-dihydro-3H-pyrimidin-4-one (intermediate J) and 5-methyl-isoxazole-3-carbaldehyde yielded the title compound as a colorless solid. MS (ESI): m/z=374.3 $[M+H]^+$.

https://open.surechem.com/en/document/US-20120035195-A1/

Yahoo! Google Maps YouTube Wikipedia News (104) Popular



STRUCTURE(S) EXTRACTED FROM THIS IMAGE



Chemical name:
(6S)-2-amino-6-(2-fluoro-5-[(5-methyl-1,2-oxazol-3-yl)methyl]amino}phenyl)-3,5,5,6-tetramethyl-3,4,5,6-tetrahydropyrimidin-4-one

SMILES:
CN1C(N)=N[C@]
(C)(C2=CC(NCC3=NOC(C)=C3)=CC=C2F)C(C)(C)C1=O

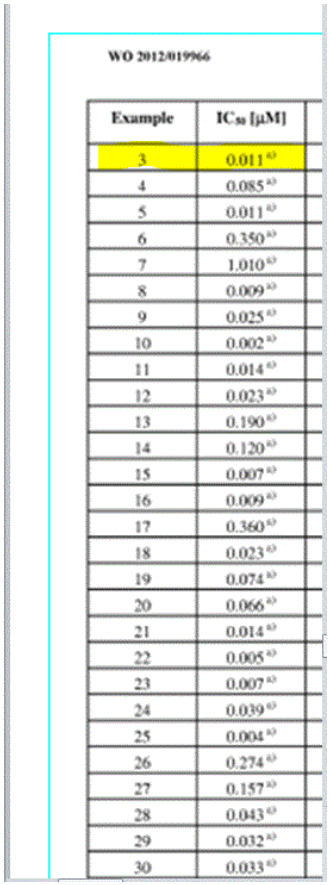
InChI:
InChI=1S/C19H24FN5O2
/c1-11-8-13(24-27-11)10-22-12-6-7-15(20)14(9-12)19(4)18(2,3)16(25

The reductive amination of (S)-2-amino-6-(5-amino-2-fluoro-phenyl)-3,5,5,6-tetramethyl-5,6-dihydro-3H-pyrimidin-4-one (intermediate J) and 5-methyl-isoxazole-3-carbaldehyde yielded the title compound as a colorless solid. MS (ESI): m/z=374.3 [M+H]+.

Example 105(S)-2-Amino-6-(2-fluoro-5-(pyridin-2-ylamino)phenyl)-3,5,5,6-tetramethyl-5,6-dihydropyrimidin-4(3H)-one

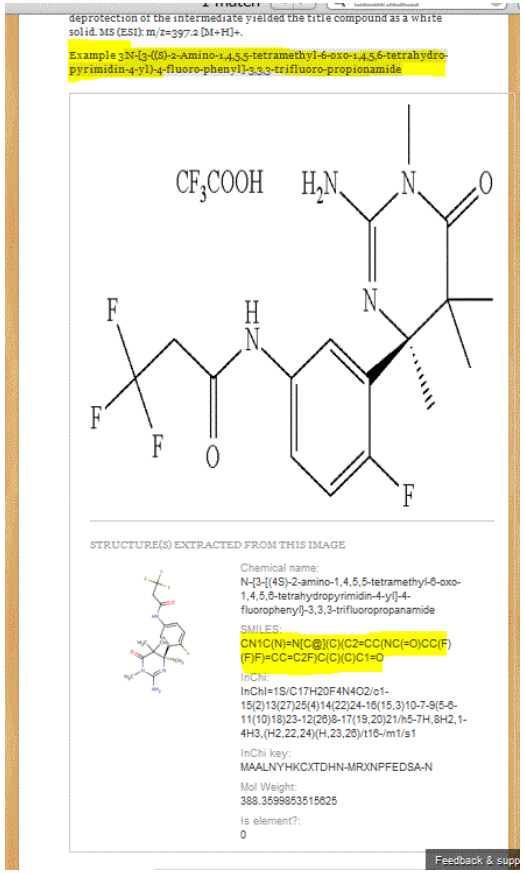
Structures with IC_{50} Values

US-20120035195-A1



WO 2012/019966

Example	IC_{50} [μM]
3	0.011 a)
4	0.085 a)
5	0.011 a)
6	0.350 a)
7	1.010 a)
8	0.009 a)
9	0.025 a)
10	0.002 a)
11	0.014 a)
12	0.023 a)
13	0.190 a)
14	0.120 a)
15	0.007 a)
16	0.009 a)
17	0.360 a)
18	0.023 a)
19	0.074 a)
20	0.066 a)
21	0.014 a)
22	0.005 a)
23	0.007 a)
24	0.039 a)
25	0.004 a)
26	0.274 a)
27	0.157 a)
28	0.043 a)
29	0.032 a)
30	0.033 a)



deprotection of the intermediate yielded the title compound as a white solid. MS (ESI): m/z=397.2 [M+H]+.

Example 33

N-[3-[(3S)-2-Amino-1,4,5,5-tetramethyl-6-oxo-1,4,5,6-tetrahydropyrimidin-4-yl]-4-fluoro-phenyl]-3,3,3-trifluoro-propionamide

STRUCTURE(S) EXTRACTED FROM THIS IMAGE

Chemical name:
N-[3-[(4S)-2-amino-1,4,5,5-tetramethyl-6-oxo-1,4,5,6-tetrahydropyrimidin-4-yl]-4-fluorophenyl]-3,3,3-trifluoropropanamide

SMILES:
CN1C(N)=N[C@](C)(C2=CC(NC(=O)CC(F)(F)F)=CC=C2F)C(C)(C)C1=O

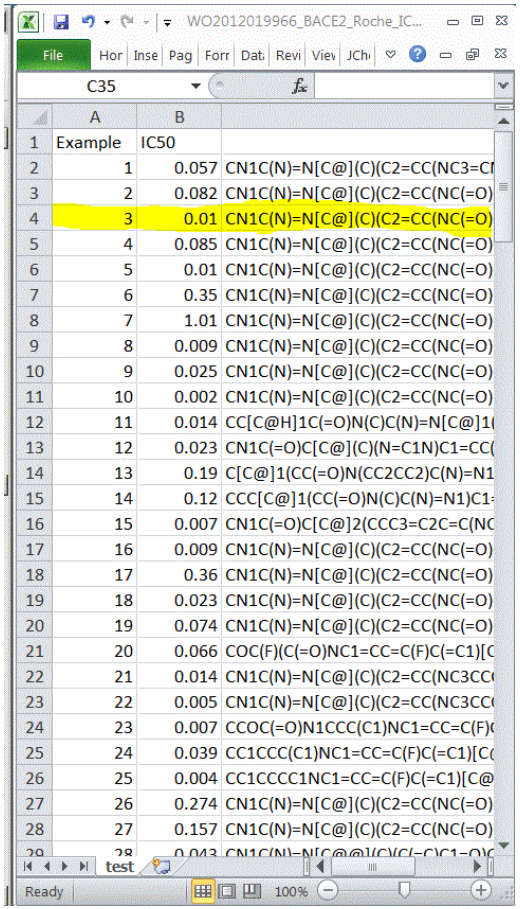
InChI:
InChI=1S/C17H20F4N4O2/c1-15(2)13(27)25(4)14(22)24-16(15,3)10-7-9(5-6-11(10)18)23-12(26)8-17(19,20)21/h5-7H,8H2,1-4H3,(H2,22,24)(H,23,26)/t16-/m1/s1

InChI key:
MAALNYHKCXTDHN-MRXNPFEDSA-N

Mol Weight:
388.3599653515625

Is element?:
0

Feedback & supp



WO2012019966_BACE2_Roche_IC...

File | Hor | Inse | Pag | Forr | Dat | Revi | Viev | JCh

C35

	A	B	
1	Example	IC50	
2	1	0.057	CN1C(N)=N[C@](C)(C2=CC(NC3=Cl
3	2	0.082	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
4	3	0.01	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
5	4	0.085	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
6	5	0.01	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
7	6	0.35	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
8	7	1.01	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
9	8	0.009	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
10	9	0.025	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
11	10	0.002	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
12	11	0.014	CC[C@H]1C(=O)N(C)C(N)=N[C@]1(
13	12	0.023	CN1C(=O)C[C@](C)(N=C1N)C1=CC(
14	13	0.19	C[C@]1(CC(=O)N(CC2CC2)C(N)=N1
15	14	0.12	CCC[C@]1(CC(=O)N(C)C(N)=N1)C1=
16	15	0.007	CN1C(=O)C[C@]2(CCC3=C2C=C(NC
17	16	0.009	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
18	17	0.36	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
19	18	0.023	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
20	19	0.074	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
21	20	0.066	COC(F)(C(=O)NC1=CC=C(F)C(=C1)[C
22	21	0.014	CN1C(N)=N[C@](C)(C2=CC(NC3CC(
23	22	0.005	CN1C(N)=N[C@](C)(C2=CC(NC3CC(
24	23	0.007	CCOC(=O)N1CCC(C1)NC1=CC=C(F)(
25	24	0.039	CC1CCC(C1)NC1=CC=C(F)C(=C1)[C(
26	25	0.004	CC1CCCC1NC1=CC=C(F)C(=C1)[C@
27	26	0.274	CN1C(N)=N[C@](C)(C2=CC(NC(=O)
28	27	0.157	CN1C(N)=N[C@](C)(C2=CC(NC(=O)

test

Ready | 100%

PDF → SureChemOpen → Excel

Search IC_{50} Structures in PubChem

File | Home | Insert | Page Layout | Formulas

C25 | CC1CCC(C1

	A	B	
1	Example	BACE2 IC50	SMILES
2	1	0.057	CN1C(N)=N[C@](C)(C
3	2	0.082	CN1C(N)=N[C@](C)(C
4	3	0.01	CN1C(N)=N[C@](C)(C
5	4	0.085	CN1C(N)=N[C@](C)(C
6	5	0.01	CN1C(N)=N[C@](C)(C
7	6	0.35	CN1C(N)=N[C@](C)(C
8	7	1.01	CN1C(N)=N[C@](C)(C
9	8	0.009	CN1C(N)=N[C@](C)(C
10	9	0.025	CN1C(N)=N[C@](C)(C

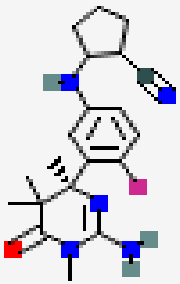
search

PubChem

Results: 12

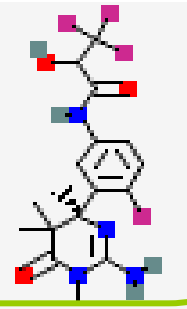
) Summary, 20 per page, Sorted by Default order

1.



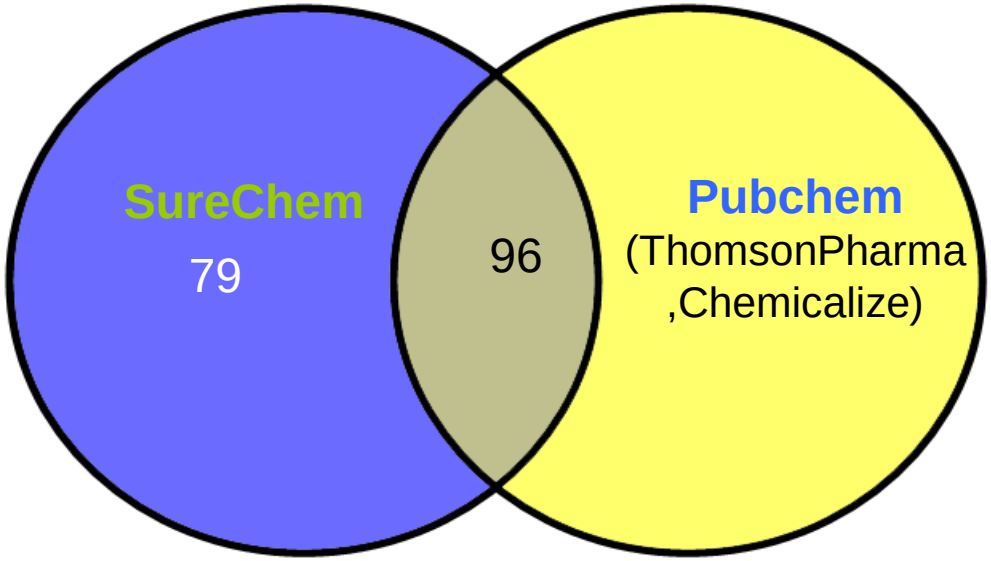
MW: 371.451743 g/mol MF: $C_{20}H_{26}FN_5O$
IUPAC name: 2-[3-[(4S)-2-amino-1,4,5,5-tetramethyl-6-oxopyrimidin-4-
CID: 56649992

2.



MW: 404.359313 g/mol MF: $C_{17}H_{20}F_4N_4O_3$
IUPAC name: N-[3-[(4S)-2-amino-1,4,5,5-tetramethyl-6-oxopyrimidin-4-
CID: 56649994

SureChem Unique Contribution



Stage	No. of Structures
Available from SureChem (SC)	1848
Pre-Exist in PubChem	669
Pre-Exist – not from IC_{50} table	573
Pre-Exist – from IC_{50} table	96 (12 from TP + 84 via chemicalize.org)
Unique-SC with IC_{50}	79
Unique-SC – beyond IC_{50} table	1100

SureChem Chemical Relevance Filtering

- Frequency counts of chemicals within patents
- Additional molecular property filtering and structural alerts
- Structural identification of “Likely Exemplars”
- Natural Language Processing – based indexing of Exemplified Compounds

Automated indexing of Exemplified Compounds in text

concentrated in vacuo to remove residual DCM. The white powder (61 mg, 12%) was then dried at 60° C. under high vacuum for 1 hour. LRMS (APCI+): 100% purity, 220 nm, m/z 422 (M+1); H NMR (400 MHz, DMSO-d6) δ 13.79 (s, 1H), 12.49 (s, 1H), 11.84 (s, 1H), 8.39 (d, J=5.5 Hz, 1H), 7.99 (d, J=12.1 Hz, 1H), 7.79 (s, 1H), 7.54 (m, 2H), 7.35 (m, 4H), 7.29 (m, 1H), 6.51 (d, J=5.5 Hz, 1H), 3.84 (s, 2H); F NMR (376 MHz, DMSO-d6) δ −129.2 (m).

Example 2

Preparation of N-(4-(1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorophenyl)-N-(4-fluorophenyl)cyclopropane-1,1-dicarboxamide

Step A: Preparation of N-(4-(1-(4-methoxybenzyl)-1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorophenyl)-N-(4-fluorophenyl)cyclopropane-1,1-dicarboxamide : To a stirred mixture of 4-(1-(4-methoxybenzyl)-1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorobenzenamine (73 mg, 0.20 mmol; obtained from Example 1, Step D) and ((4-fluorophenyl)carbamoyl)cyclopropanecarboxylic acid (49 mg, 0.220 mmol; prepared from cyclopropane-1,1-dicarboxylic acid and 4-fluoroaniline using the methods of WO 2005/030140 and by Shih and Rankin, Synth. Comm. 1996, 26(4), 833-836) in DMA (2 mL) was added N1-((ethylimino)methylene)-N3,N3-dimethylpropane-1,3-diamine hydrochloride (EDCI) (77 mg, 0.400 mmol). The reaction was stirred for 1 hour at room temperature. The reaction was diluted with EtOAc (10 mL) and water (10 mL). The phases were separated, and the organic phase washed with water (3×10 mL), brine (10 mL), dried (Na 2 SO 4), filtered, and concentrated in vacuo. The crude was purified by preparative TLC eluting with 3% MeOH/DCM. The product was obtained as a waxy solid (42 mg, 33%). H NMR (400 MHz, CDCl 3) δ 9.97 (s, 1H), 8.36 (d, J=5 Hz, 1H), 8.20 (s, 1H), 7.77 (m, 2H), 7.46 (m, 2H), 7.26 (m, 4H), 7.06 (m, 2H), 6.83 (d, J=9 Hz, 2H), 6.40 (d, J=5 Hz, 1H), 5.62 (s, 2H), 3.76 (s, 3H), 1.79 (m, 2H), 1.62 (m, 2H, overlaps with water).

Step B: Preparation of N-(4-(1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorophenyl)-N-(4-fluorophenyl)cyclopropane-1,1-dicarboxamide : Prepared according to the procedure for Example 1, Step F, substituting (4-(1-(4-methoxybenzyl)-1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorophenyl)-N-(4-fluorophenyl)cyclopropane-1,1-dicarboxamide (0.040 g, 0.0702 mmol) for 1-(4-(1-(4-methoxybenzyl)-1H-pyrazolo[3,4-b]pyridin-4-yloxy)-3-fluorophenyl)-3-(2-phenylacetyl)thiourea . The product was obtained as a white powder (7 mg, 20%). LRMS (ESI+): 94% purity, 220 nm, m/z 450 (M+1) detected; H NMR (MeOD, 400 MHz) δ 8.34 (d, J=5 Hz, 1H), 7.85 (m, 2H), 7.56 (m, 2H), 7.42 (m, 1H), 7.35 (m, 1H), 7.06 (m, 2H), 6.49 (d, J=5 Hz, 1H), 1.64 (s, 4H).

Example 3

Preparation of N-(3-fluoro-4-(1-methyl-1H-pyrazolo[3,4-b]pyridin-4-yloxy)phenyl)-N-(4-fluorophenyl)cyclopropane-1,1-dicarboxamide

Step A: Preparation of 4-(2-fluoro-4-nitrophenoxy)-1H-pyrazolo[3,4-b]pyridine : A stirred mixture of 1-(4-methoxybenzyl)-4-(2-fluoro-4-nitrophenoxy)-1H-pyrazolo[3,4-b]pyridine (27.6 g, 70.0 mmol; obtained from Example 1, Step C) and TFA (53.9 mL, 700 mmol) was heated to reflux for 18 hours under N 2 . The reaction was allowed to cool to room temperature, and then concentrated in vacuo using toluene (4×100 mL) to azeotrope residual TFA. The residue was diluted with EtOAc (200 mL) and carefully neutralized (pH=8-9) with saturated aqueous NaHCO 3 (100 mL). The biphasic suspension was stirred at room temperature for 30 minutes. The suspension was filtered. The resulting solid was dried by toluene azeotrope (2×200 mL) to obtain the product (18.7 g, 97%). H NMR (DMSO-d6, 400 MHz) δ 13.85 (br s, 1H), 8.40 (m, 2H), 8.15 (m, 1H), 7.91 (s, 1H), 7.66 (m, 1H), 6.65 (m, 1H).

Step B: Preparation of 4-(2-fluoro-4-nitrophenoxy)-1-methyl-1H-pyrazolo[3,4-b]pyridine : A similar pyrazole alkylation protocol was utilized by Lynch, B. et al. Can. J. Chem. 1988, 66, 420-428. To a stirred mixture of 4-(2-fluoro-4-nitrophenoxy)-1H-pyrazolo[3,4-b]pyridine (0.250 g, 0.912 mmol) absolute EtOH (0.5 mL), and a 1.5 M sodium ethoxide-ethanol solution (1.22 mL, 1.82 mmol; prepared from absolute EtOH and Na metal) at 0° C. under N 2 was added iodomethane (0.114 mL, 1.82 mmol). The suspension was allowed to warm to room temperature slowly as the ice melted, and stirring was continued for 18 hours at room temperature. The reaction was concentrated in vacuo, suspended in DCM and loaded onto a preparative TLC plate, eluting with 3% MeOH/DCM to separate the two pyrazole regioisomers. The desired 1-methyl isomer was obtained as a white solid (49 mg, 19%). H NMR (400 MHz, CDCl 3) δ 8.44 (d, J=5 Hz, 1H), 8.17 (m, 2H), 7.85 (s, 1H), 7.41 (m, 1H), 6.49 (d, J=5 Hz, 1H), 4.18 (s, 3H).

Step C: Preparation of 3-fluoro-4-(1-methyl-1H-pyrazolo[3,4-b]pyridin-4-yloxy)benzenamine : Prepared according to the procedure of Example 1, Step D, substituting 4-(2-fluoro-4-nitrophenoxy)-1-methyl-1H-pyrazolo[3,4-b]pyridine (49 mg, 0.17 mmol; obtained from Example 3, Step B) for 1-(4-methoxybenzyl)-4-(2-fluoro-4-nitrophenoxy)-1H-pyrazolo[3,4-b]pyridine . Yield: 22 mg, 42%. The product was used in the next step without purification. H NMR (400 MHz, CDCl 3) δ 8.34 (d, J=6 Hz, 1H), 7.71 (s, 1H), 7.04 (m, 1H), 6.55 (m, 1H), 6.49 (m, 1H), 6.42 (d, J=6 Hz, 1H), 4.13 (s, 3H), 3.86 (br s, 2H).

Conclusion

SureChem deposition into PubChem:

- Significantly expands public patent chemistry scope
- Contributes unique and timely MedChem-relevant data
- Enables open drug discovery and chemical biology
- Advances progress toward a more open, federated chemical information network

SureChem

PubChem

SOFTWARE THAT UNDERSTANDS SCIENCE

We are an innovative technology company developing software and apps that change the way science is done.

OUR STORY >

SOFTWARE FOR SCIENCE

Browse our range of innovative products

Browse

BioRAFT

Integrate your laboratory safety and regulatory compliance

readcube

Organise and discover the research relevant to you

figshare

Publish all of your data in a citable, searchable and sharable manner

Dr. Eli Lewis explains how he has benefitted from using LabGuru at Ben-Gurion University

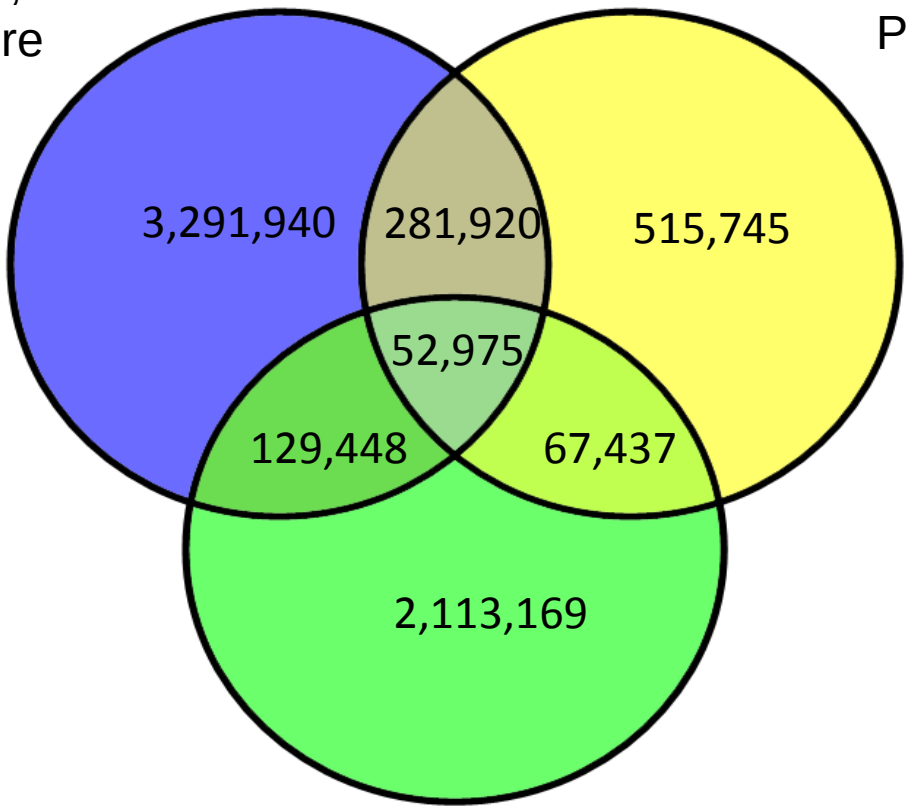
LabGuru in practice

Patent & Literature Sources in PubChem

The Big Three

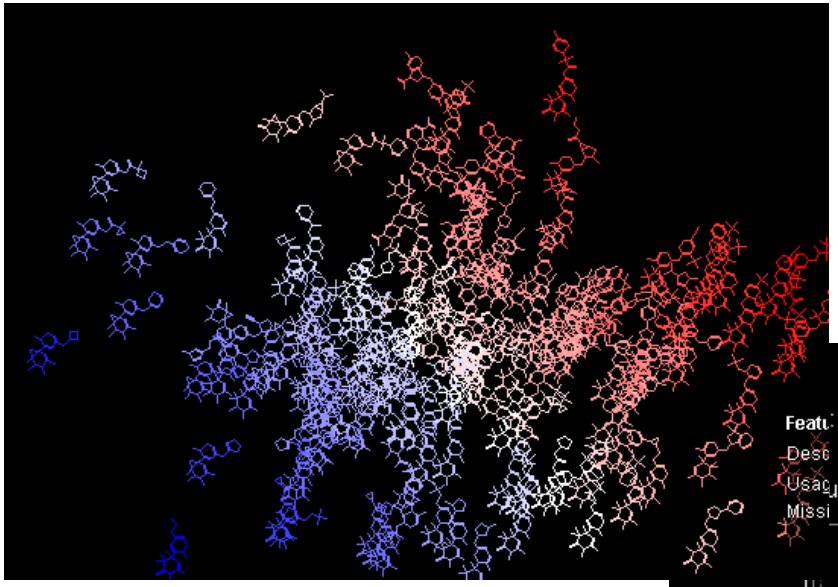
Thomson Pharma,
patents and literature
3,756,283
41% lead-like

ChEMBL +
PubMed + Journals
918,077
45% lead-like

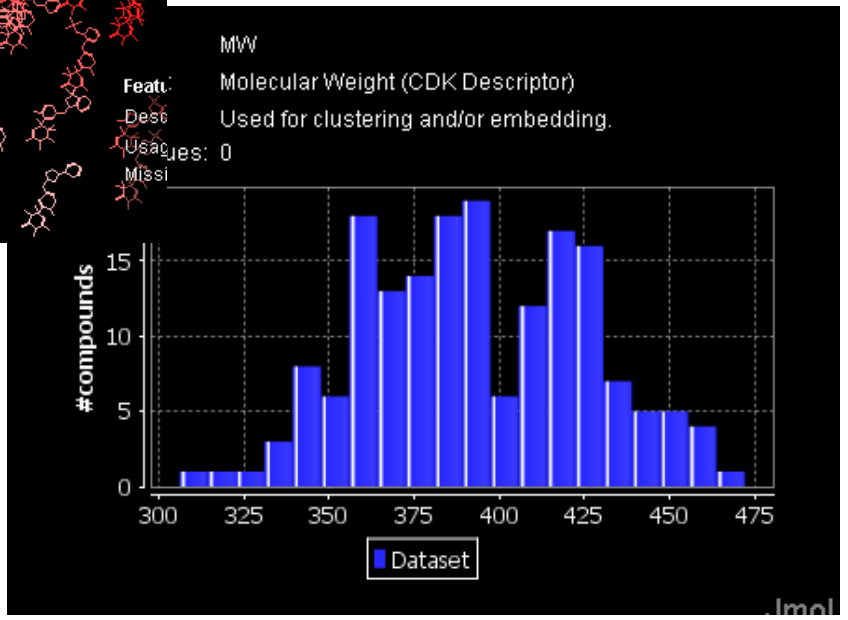


IBM, pre-2000 patents **2,369,481** 32% lead-like

Identifying Relevant Chemistry



Patent
US-20120035195-A1



http://opentox.informatik.uni-freiburg.de/ches-mapper/