SureChem

SureChem: Open Access to Current, Comprehensive Chemical Patent Search

ICIC 2007

Nicko Goncharoff
Reel Two, Inc.

New Technologies Enable a New Paradigm

- Automated chemical name/structure extraction - USPTO, EPO, WO
- SureChem database
  - 8+ million patents
  - 10+ million structures
- Technology that enables comprehensive and current chemical patent coverage
- Also enables a forward-looking, Open Access-based business model – the first in chemical patent search

Good for business, good for science

"Flying Blind"

- Researchers & information specialists lack clarity on the IP protection around target compounds
- They need:
  - *Structures* from full text of all major patent databases
  - Data on their desktop or database for batch searching/analysis
  - Updates within 24 hours of new patent issuance
  - View of structures in context
- Without it patent chemistry data will remain an essential yet underutilized resource[1]

1. Zhou Y, Zhou B, Chen K, Yan SF, King FJ, Jiang S, Winzeler EA, *Large-Scale Annotation of Small-Molecule Libraries Using Public Databases*. J Chem Inf Model. 2007 Jul 3.

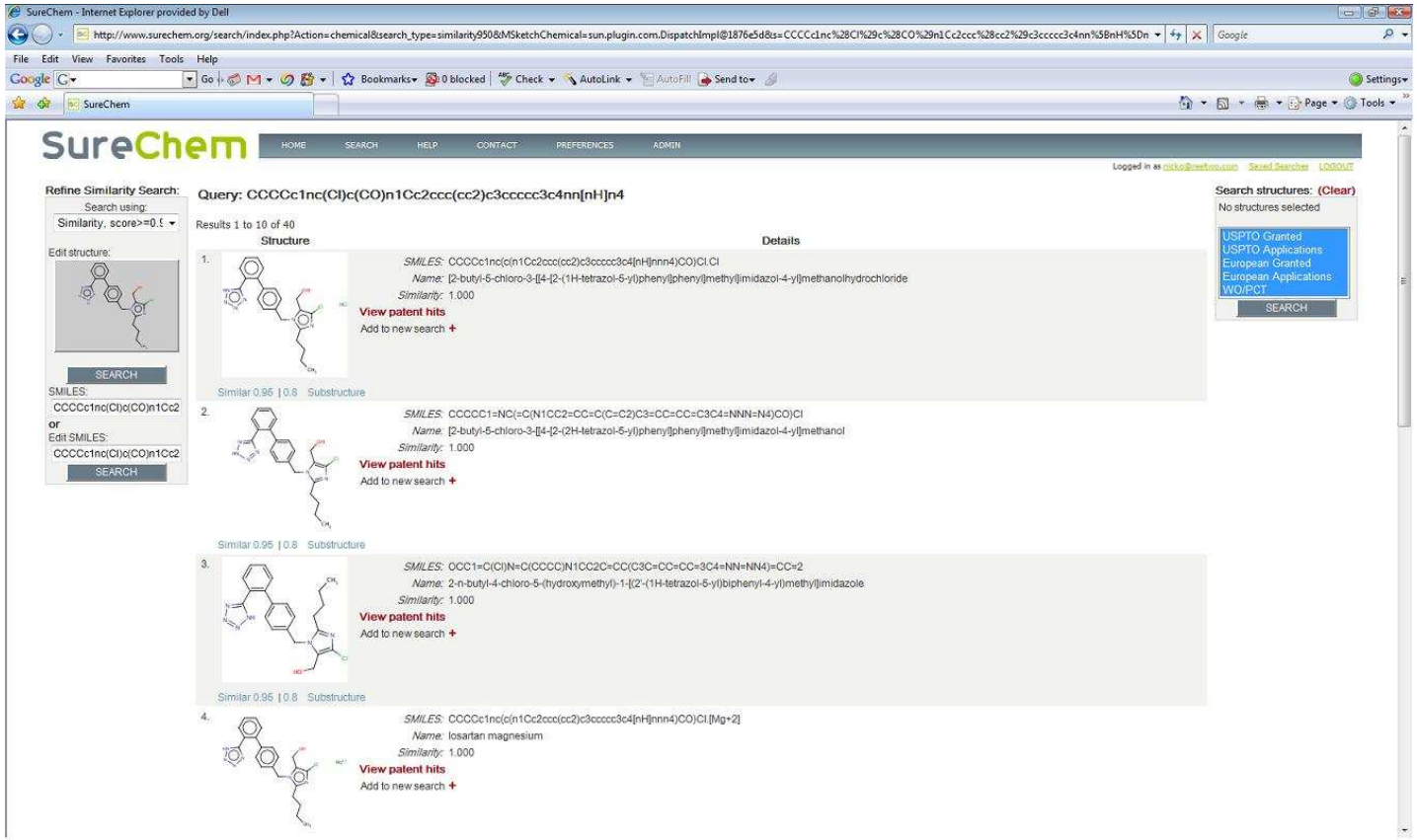
And What Would You Like?

- I want my data!
  - Chemical structure data can be exported to desktop or deployed on in-house databases
- I want it now!
  - Can deliver new structure data within 24 hours of patent issuance
- I want it my way!
  - SureChem Online (Portal)/In-House (Database)
- I want it for free (or for a good price)!
  - OK!

SureChem – How It's Done

- Chemical names are extracted from full text patents and converted to structures
  - Extensive data cleaning, error correction, text formatting performed on patent documents
  - Entity extractor heuristics for systematic names, drug names
  - Multiple name-structure conversion tools ensure highest possible conversion rates
  - Image data available for USPTO documents from 2001 onward

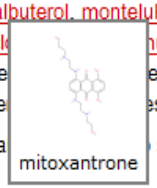
SureChem Online – Search Portal



SureChem Portal – Structure Context

12. A pharmaceutical composition comprising a compound according to claim 1 and a pharmaceutically acceptable carrier, adjuvant, or vehicle.

13. The composition of claim 12, further comprising an additional therapeutic agent selected mechlorethamine, chlorambucil, cyclophosphamide, melphalan, ifosfamide, methotrexate, 6-mercaptopurine, 5-fluorouracil, cytarabile, gemcitabine, vinblastine, vincristine, vinorelbine, paclitaxel, etoposide, irinotecan, topotecan, doxorubicin, bleomycin, mitomycin, carmustine, lomustine, cisplatin, carboplatin, asparaginase, tamoxifen, leuprolide, flutamide, megestrol, imatinib mesylate, adriamycin, dexamethasone, cyclophosphamide, donepezil (Aricept®), rivastigmine (Excelon®)), L-DOPAlcarbidopa, entacapone, ropinrole, pramipexole, bromocriptine, pergolide, trihexephendyl, amantadine, glatiramer (Copaxone®), mitoxantrone, albuterol, montelukast (Singulair®), zyprexa, risperdal, seroquel, haloperidol, a corticosteroid, a TNF blocker, IL-1 RA, azathioprine, cyclophosphamide, sulfasalazine, cycl…nus, rapamycin, mycophenolate mofetil, an interferon, cyclophosphamide, azathioprine, sulfasalazine, an acetylcholinesterase inhibitor, a monoamine oxidase…erferon, an anti-convulsant, an ion channel blocker, riluzole, a beta-blocker, an acetyl cholinesterase inhibitor, a diuretic, a nitrate, a calcium channel blocke…estyramine, or gamma globulin.



14. A method of inhibiting FLT-3 or c-KIT kinase activity in a biologica… selected from a cell culture, saliva, urine, feces, semen, tears, or extracts thereof, comprising the step of contacting said biological sample with:

How Well Does it Work?

- SureChem outperformed in-house chemical extraction efforts of several major pharma firms
- SureChem database v1.0 matched 55-65% of structures in benchmark set of 435 patents. v2.0 matches 73%
- Structure conversion – micro and macro-averages
- Validation
  - Name-structure tool rate of agreement
  - Manual QA
  - Community curation
- Future releases = further improvement

What Users Like

- Confidence they’re scanning the entire patent
- Data is theirs to keep
- Flexible search options
- Can get full database in house
- Immediate updates
- Rational, predictable pricing (Pro)

Next!.....Other Chemistry Challenges

Next 3 months

- Synonym linkage and expansion
- Enhanced Tautomer searches
- Markush element search
- Linking structure data between patent families

6-18 months

- Extraction of structures from images
- Linking structures with genes, proteins, disease names, biological data
- Expanding to other data sources – journals, clinical trial data, etc.
- Markush Analysis

SureChem Open Access

- SureChem will be freely available to all users
- Access all patent databases, structure search and query refinement tools
- Similar functionality to current paid services
- ‘Pro’ version will also offer:
  - Secure connection (https://)
  - Data export
  - Alert services
  - User collaboration, data viewing/merging tools

Why Open Access?

- Good for science, good for business
  - Contributes to scientific community – helps chemistry open access catch up to biology data and informatics
  - Raises awareness of our Pro and Database products
  - Takes advantage of community curation, collaboration, cleansing of data
  - We enjoy our jobs much more!

Building a Better Web with a Spider

- ChemSpider
  - Single point access to 15-20 million structures
  - Searches span patents, public databases, supplier databases, journals
  - www.chemspider.com
- In discussions with other partners for Open Access and paid services

Where To Find Us

- www.surechem.org