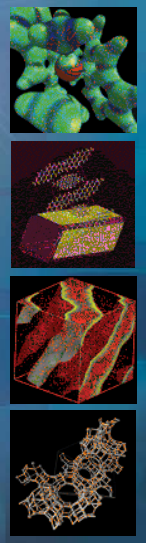


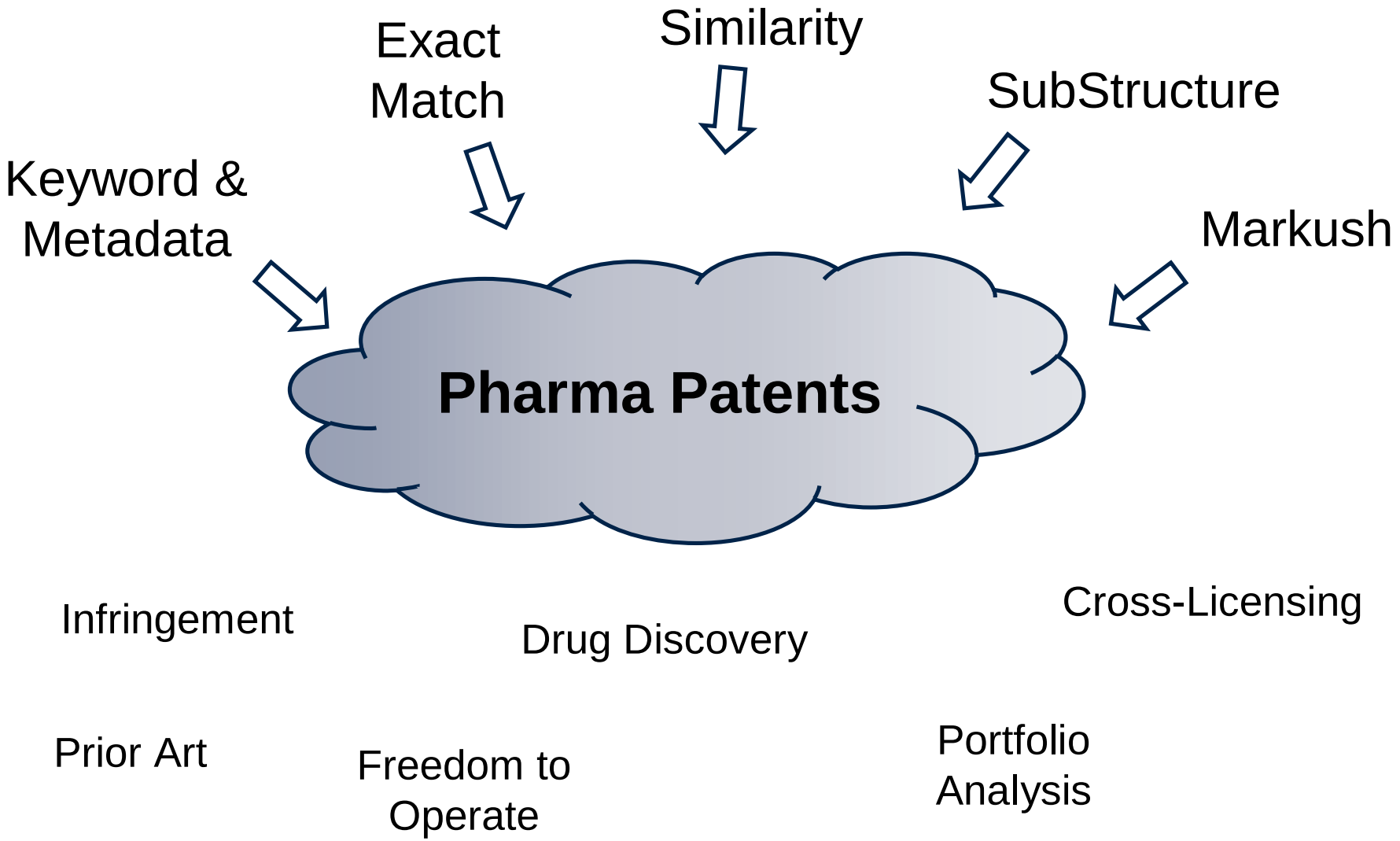
Characterizing Pharmaceutical Patents Using a Combined Text and Chemical Analysis Approach

Brian Bartell (Notiora Inc)
Robert Brown, Eddy Vande Water,
Andrew LeBeau (Accelrys, Inc)

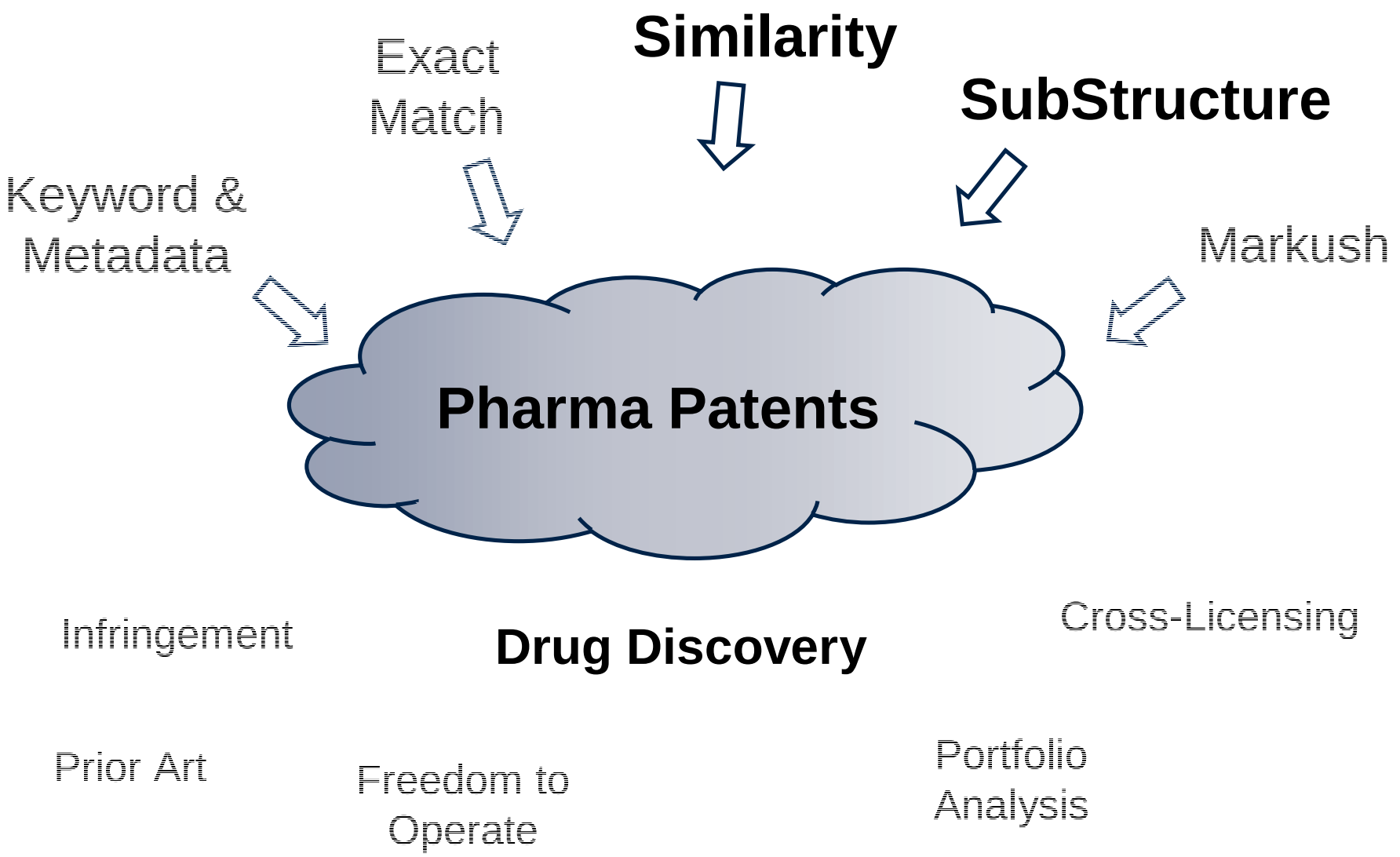
Overview

- **Structure-Based Patent Search**
  - "Best match" addresses varied problems in drug development
- **Q: Can we match a target to a patent as a whole?**
  - and is it interesting?
- **A: Aggregate Asymmetric Search**
  - Fast, Novel, Interesting
- **UI for Fast Patent Qualification**
- **Examples**

Patent Search



Patent Search

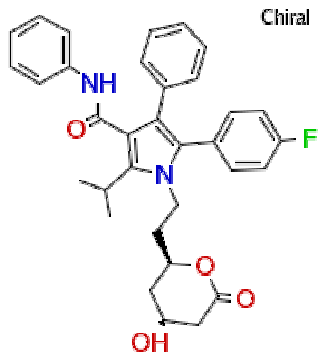


Structure Search

- **Why find patents with related structures?**
  - Similar Property Principal
    - "structurally similar molecules are expected to exhibit similar properties or biological activities"
      - Mark Johnson and Gerry Maggiora (Eds.) *Concepts and Applications of Molecular Similarity*. Wiley, New York, 1990
  - Lead Generation
    - Find structures that are likely to bind
  - Lead Optimization
    - Refine structures to make them bind better

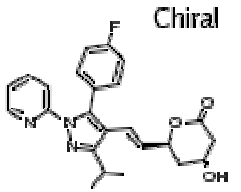
Structure Search

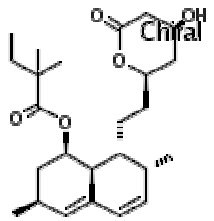
Target

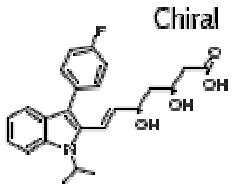


Match?

Candidate Patent





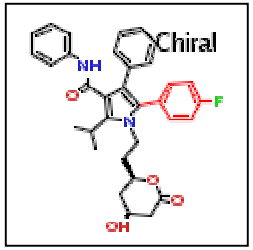


Search is “Best Match”

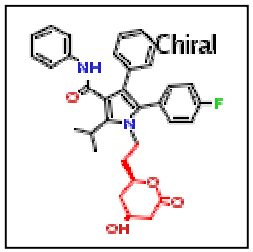
Notiora

accelrys®

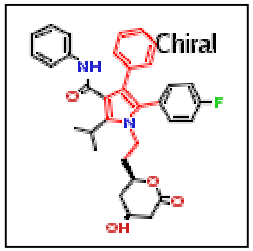
Match?



0.3

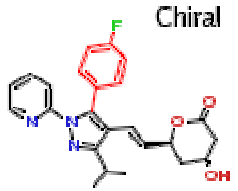


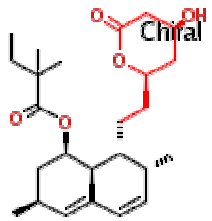
0.25

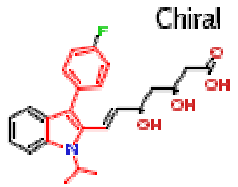


0.42

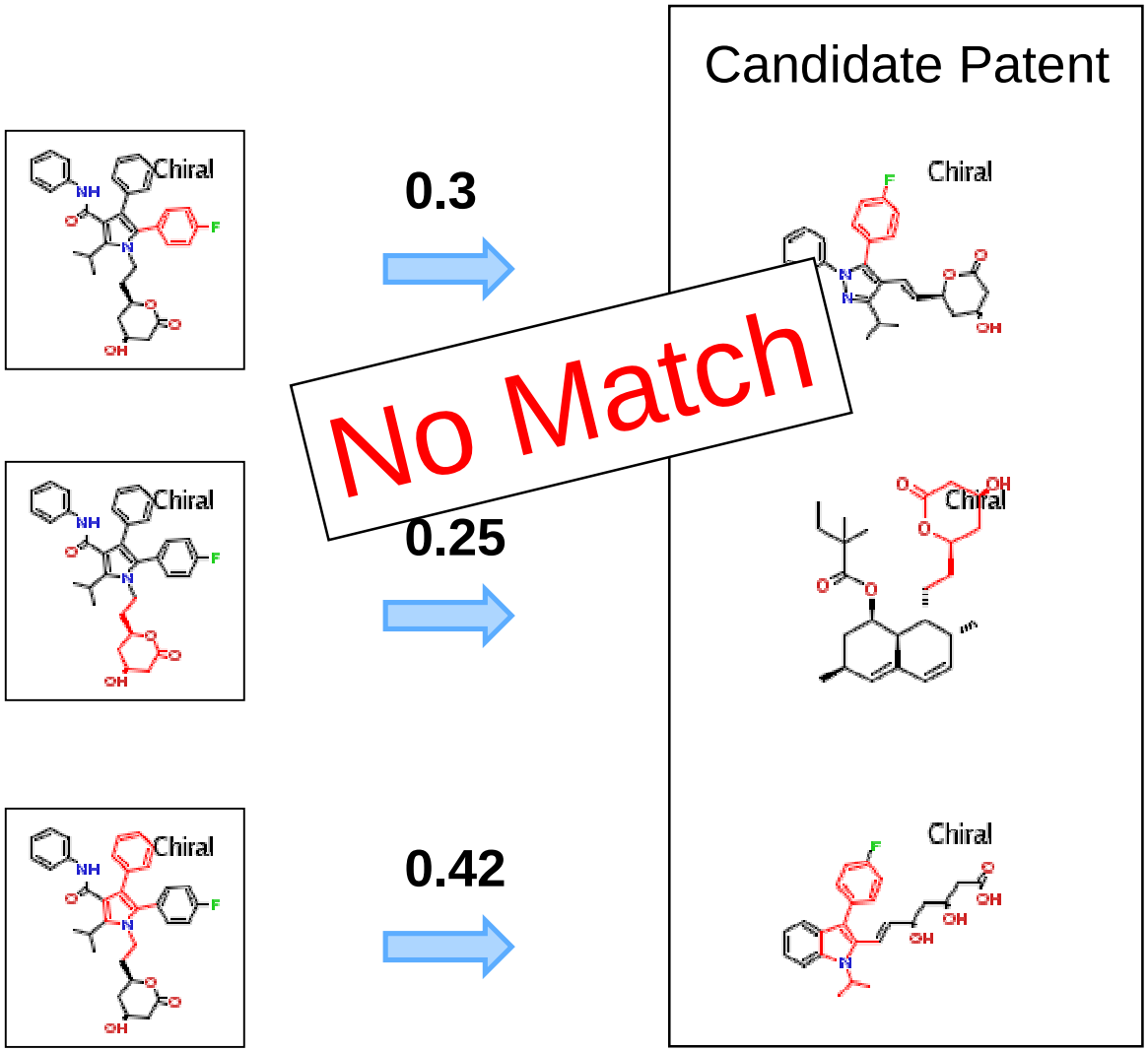
Candidate Patent







Search is “Best Match”



What's Missing?

- **Patent is a useful locus**
  - Structures therein will be related by a use or research program

- **Hypothesis**
  - Patents having structures that overlap different fragments of the target may be interesting.

- **Novelty**
  - Traditional Structure Search doesn't find these patents

Aggregate Asymmetric Search

- **Goal: Find patents where…**
  - No single structure is similar/substructure to target
  - Parts from multiple structures cover the target
- **This identifies patents with fragments in target**

Aggregate Asymmetric Search

- **Goal: Find patents where…**
  - No single structure is similar/substructure to target
  - Parts from multiple structures cover the target
- **This identifies patents with fragments in target**
- **Challenge:**
  - Combinatorics

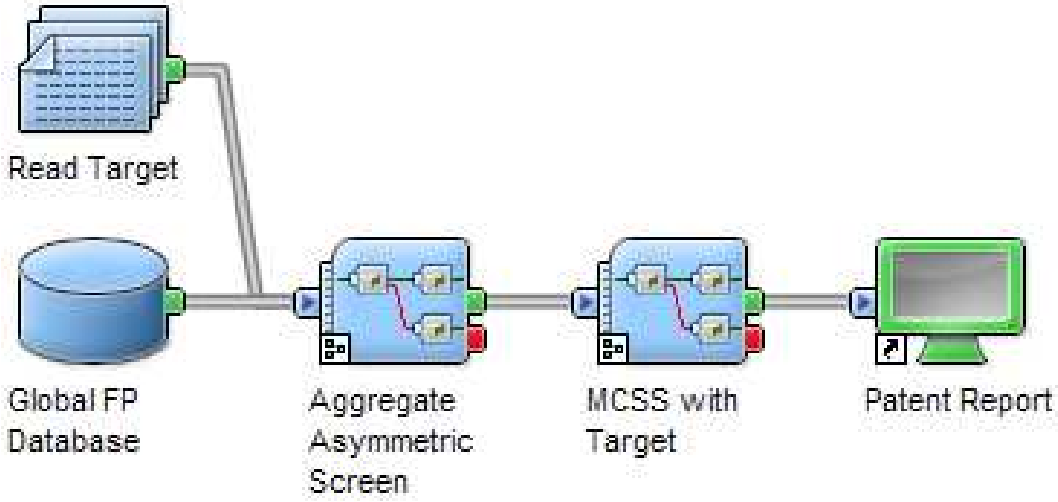
Aggregate Asymmetric Search

- **Goal: Find patents where…**
  - No single structure is similar/substructure to target
  - Parts from multiple structures cover the target
- **This identifies patents with fragments in target**
- **Challenge:**
  - Combinatorics
- **Solution:**
  - Patent-wide Fingerprints (fast screen)
  - Pairwise Maximal Common SubStructure (MCSS)

Processing Overview

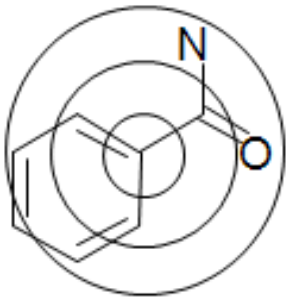
One Time Pre-Process

Per Target

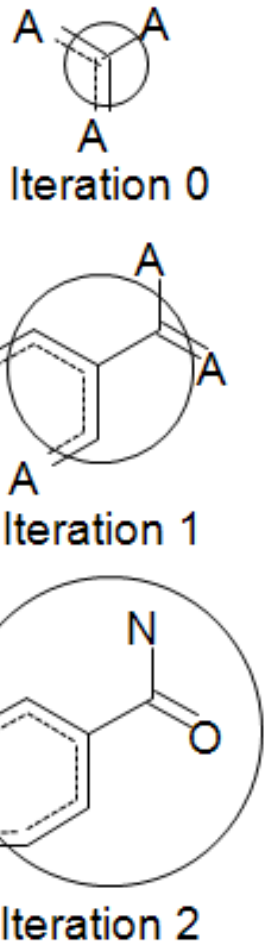


Extended Connectivity FPs

- **Record (bond-type,atom-type) codes for each neighbor**
- **Sort to avoid order dependency**
- **Apply hashing function to map to a single number in the 2^{32} address space (~4 billion bits)**
- **Chance of collisions is *extremely* low**

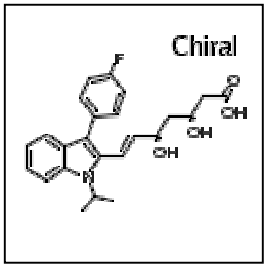


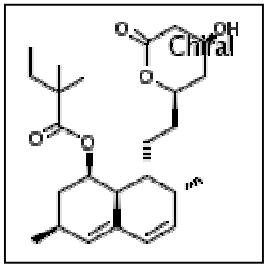
Each iteration adds bits that represent larger and larger structures

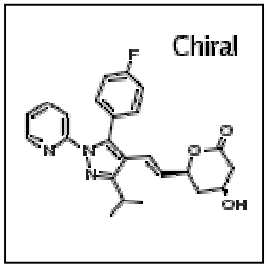


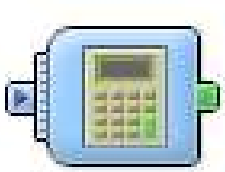
Aggregate Patent Fingerprint

Candidate Patent

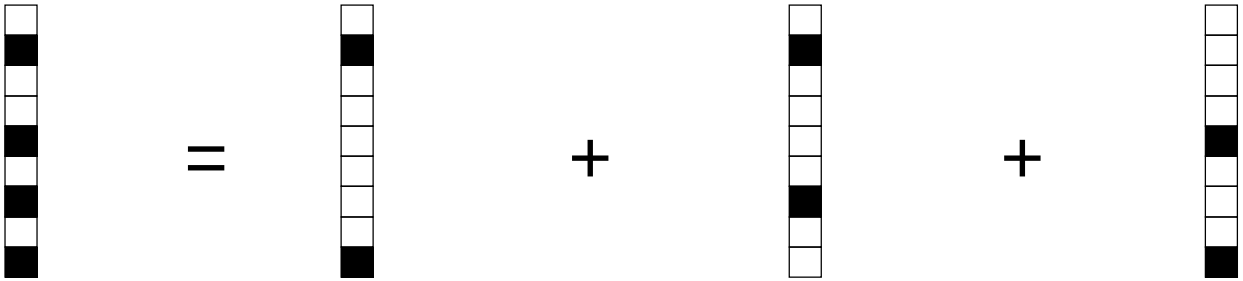






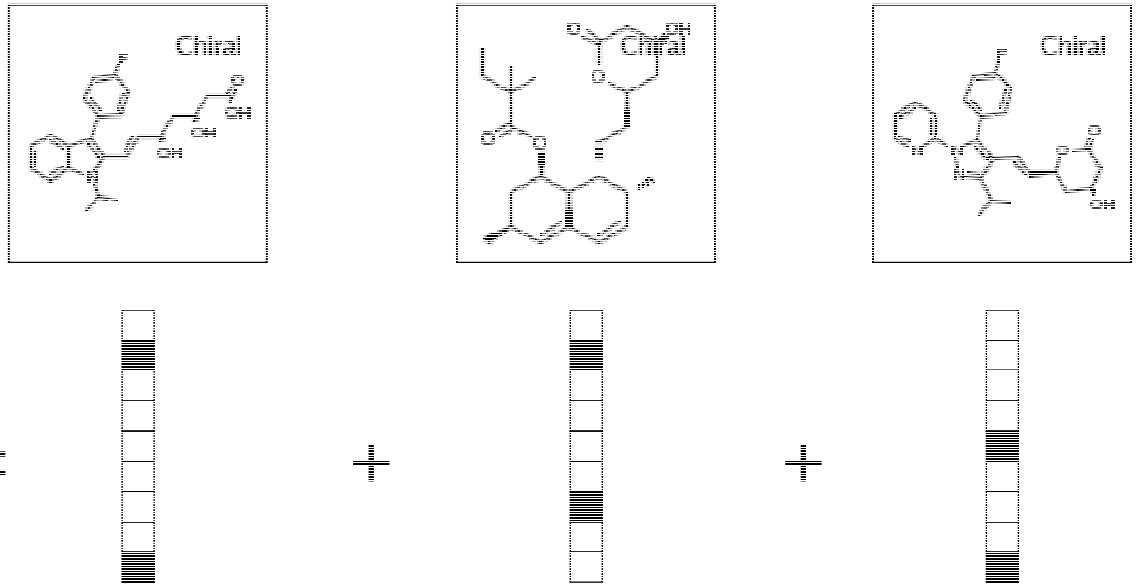


Aggregate Patent Fingerprint

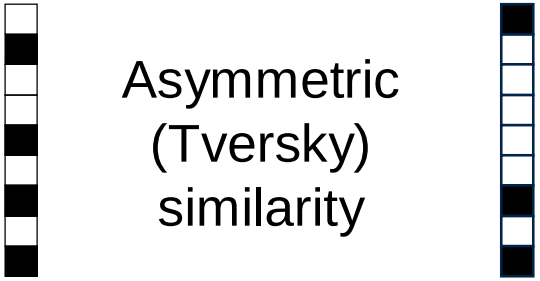


Aggregate Asymmetric Screen

Candidate Patent

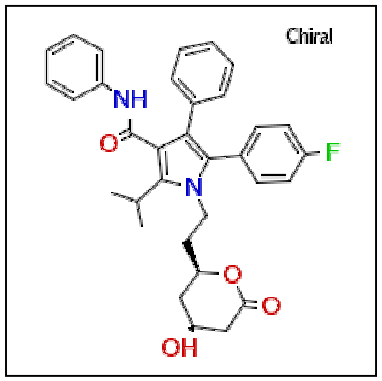


Aggregate Asymmetric Screen

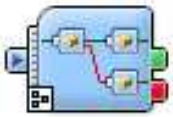


Asymmetric (Tversky) similarity

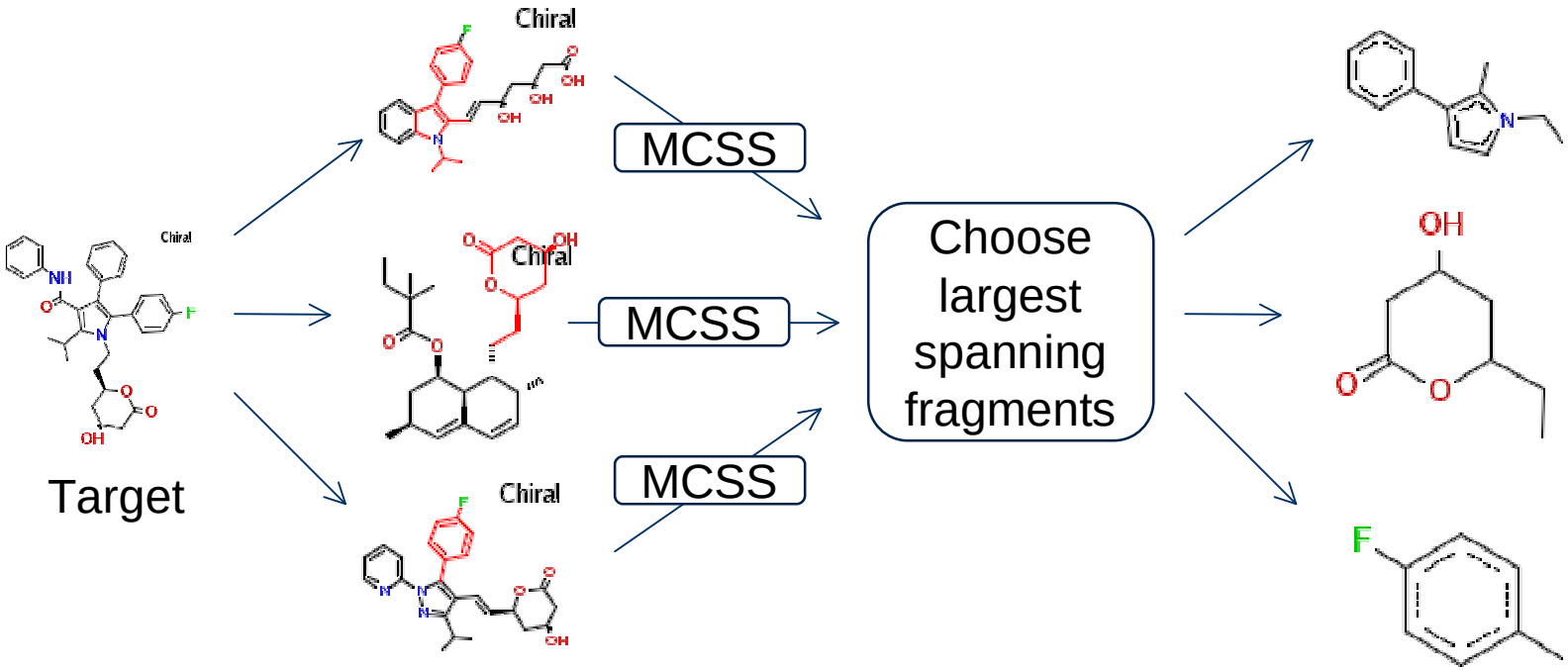
Target



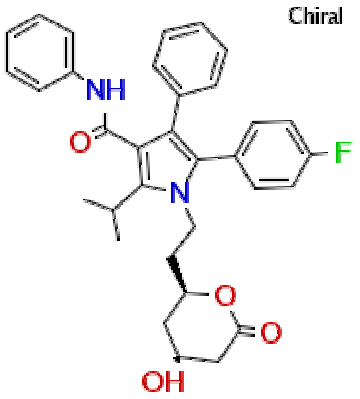
Target Coverage via MCSS



- **Pairwise MCSS for each candidate Patent**
  - Calculate all common substructures with target
  - Choose largest fragments that together cover the target
- **Isolates structures and fragments that cover target**



Example - Lipitor

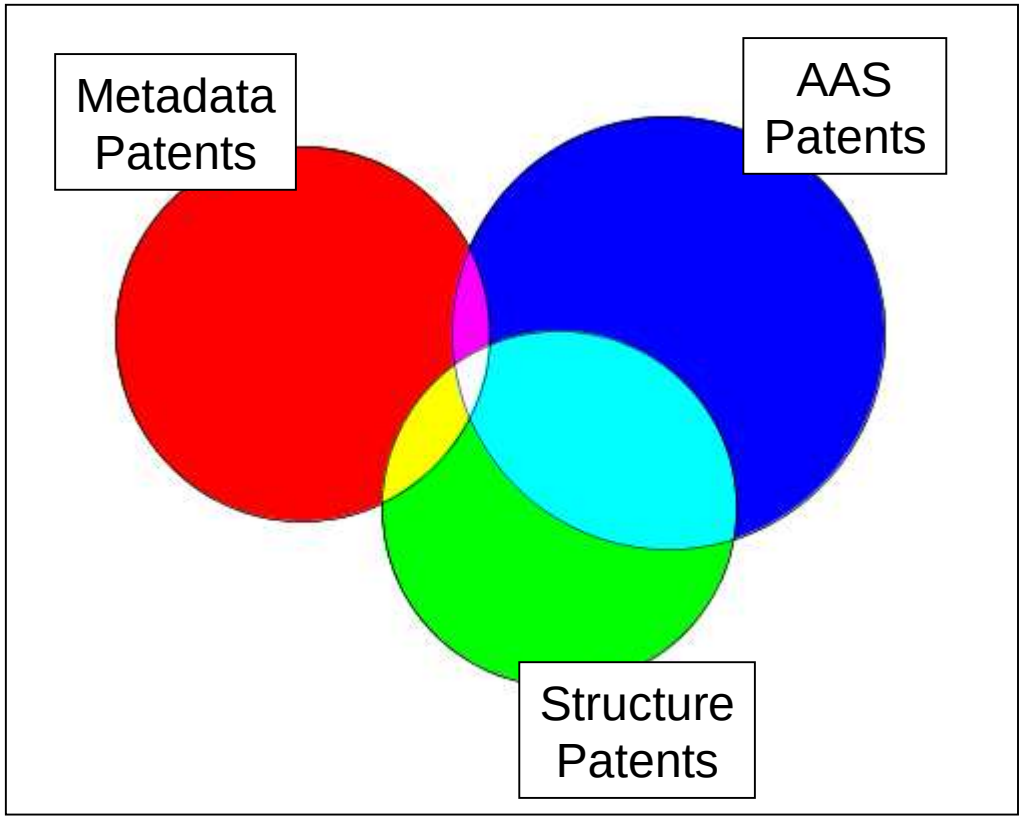


Name	**Lipitor / Atorvastatin**
IUPAC	(3*R*,5*R*)-7-[2-(4-fluorophenyl)-3-phenyl-4-(phenylcarbamoyl)-5-(propan-2-yl)-1*H*-pyrrol-1-yl]-3,5-dihydroxyheptanoic acid
Pat No	US 4,681,893 – Pfizer – 1987
US CCL	514/422,423 ; 548/517,537

- **Compare with ~ 30,000 US patents**
  - Selected from large pharma via AN/
  - Pre- and post-dating lipitor
  - From variety of patent categories

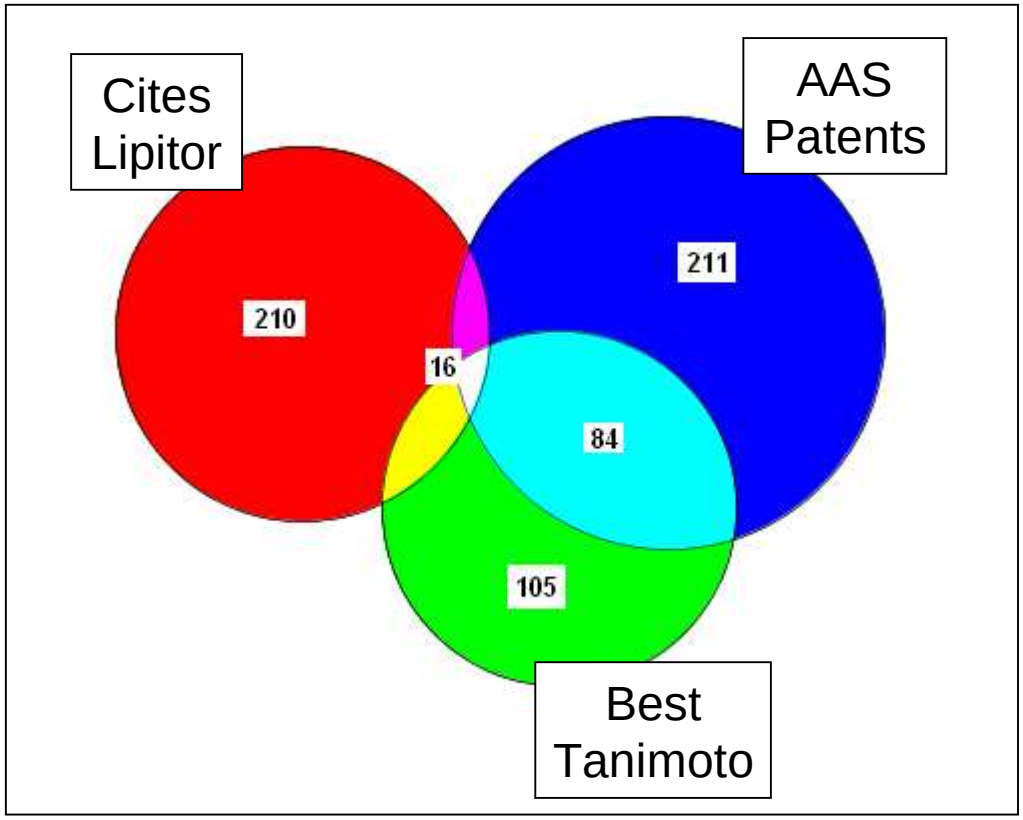
Novel?

- **Does AAS find patents you would otherwise not find?**
  - How do the AAS patents compare to:
    - Structure search
    - Keyword search
    - Citation walking
    - Patent class



Novel

- **AAS finds patents you would not otherwise find**
  - Finds novel patents relative to other searches
  - Promotes some patents with low “Best” single structure similarity
  - Finds patents with multiple fragments covering target



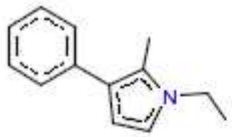
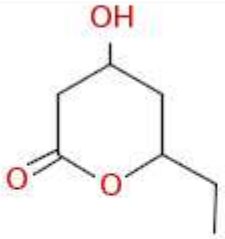
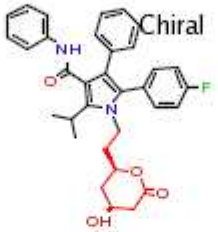
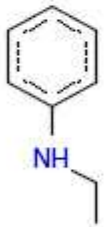
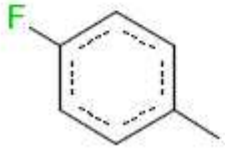
Relevance?

- **Top 3 AAS patents for Lipitor are:**

Patent	Assignee	Date	Novel?	Title
6,306,891	Merck	2001	**CCL**	HIV integrase inhibitors
5,470,845	BMS	1995	**Yes**	Methods of using .alpha.-phosphonosulfonate squalene synthetase inhibitors including the treatment of atherosclerosis and hypercholesterolemia
7,030,112	BMS	2006	**Yes**	Pyrrolopyridazine compounds and methods of use thereof for the treatment of proliferative disorders

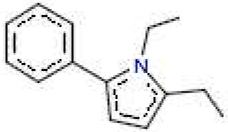
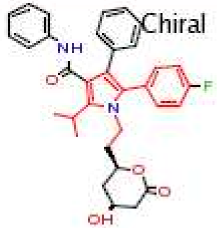
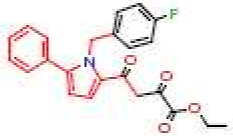
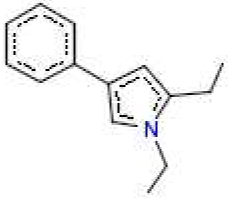
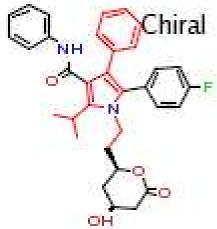
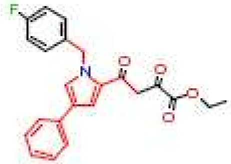
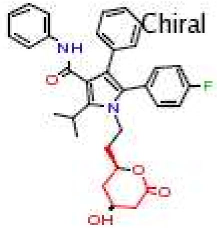
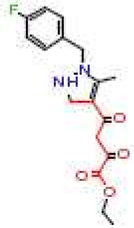
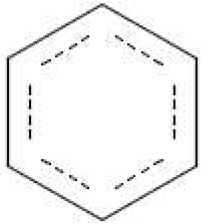
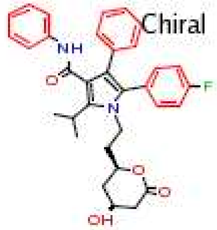
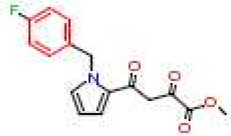
Relevance

- **Patent #2: hypercholesterolemia statin patent (BMS)**
  - Does not share citations, classification, named structures.
  - Most similar single structure is 0.42 tversky
  - Obviously relevant by disorder

Common Fragment	Lipitor w/Fragment	Patent Structures
		
		
		
		

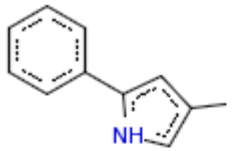
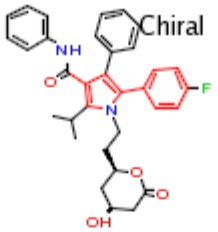
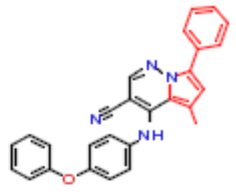
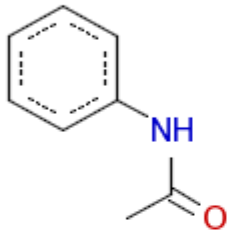
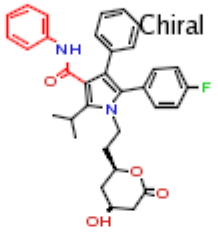
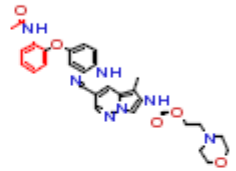
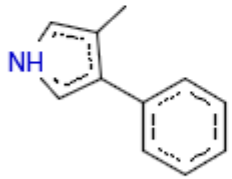
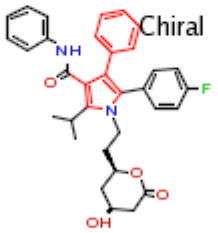
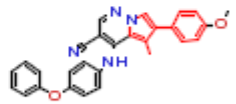
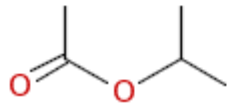
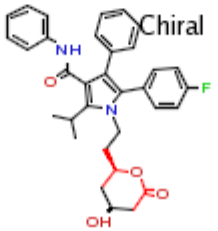
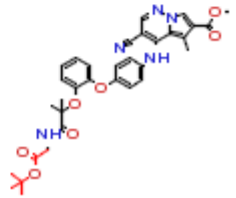
Relevance

- **Patent #1: HIV integrase inhibitors (Merck)**
  - Does not share citations or named structures
  - Same primary US CCL: 514/423

Common Fragment	Lipitor w/Fragment	Patent Structures
		
		
		
		

Relevance

- **Patent #3: Proliferative disorders (BMS)**
  - Does not share citations, classification, named structures

Common Fragment	Lipitor w/Fragment	Patent Structures
		
		
		
		

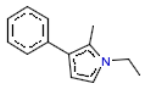
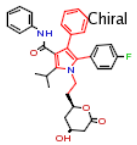
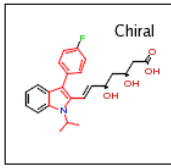
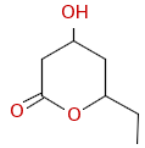
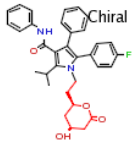
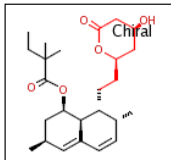
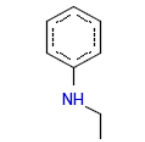
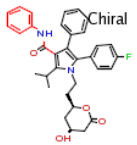
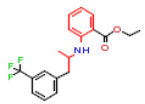
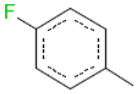
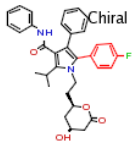
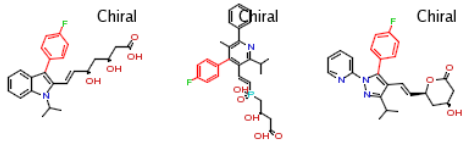
UI for Assessing AAS Patents

- Displays each matching patent and the largest spanning fragments
- Scientist quickly qualifies patents for further investigation
- Applicable also to general structure search results

US 5,470,845: **Methods of using .alpha.-phosphonosulfonate squalene synthetase inhibitors including the treatment of atherosclerosis and hypercholesterolemia**

1995/11/28: Bristol-Myers Squibb Company (Princeton, NJ): Magnin, David R.; Biller, Scott A.; ... Sulsky, Richard B.

.alpha.-Phosphonosulfonate compounds are provided which inhibit the enzyme squalene synthetase and thereby inhibit cholesterol biosynthesis. These compounds have the formula ##STR1## wherein R^2 is OR^5 or R^5a ; R^3 and R^5 are independently H, alkyl, arylalkyl, aryl or cyc... ...cloheteroalkyl, cycloheteroalkylalkyl; as further defined above; including pharmaceutically acceptable salts and or prodrug esters of the phosphonic (phosphinic) and/or sulfonic acids.

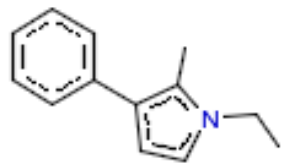
Common Fragment	Lipitor w/Fragment	Patent Structures containing Fragment
		 Structure Name: **fluvastatin** Structure in Context: • *Claim #13:* ...uctase inhibitor is pravastatin, lovastatin, simvastatin, velostatin, **#STR#**, rivastatin, compactin, SDZ-63,370 (San...
		 Structure Name: **simvastatin** Structure in Context: • *Claim #13:* ...2 wherein the HMG CoA reductase inhibitor is pravastatin, lovastatin, **#STR#**, velostatin, fluvastatin, rivastatin, c...
		 Structure Name: **ethanol 2-[[1-methyl-2-[3-(trifluoromethyl) phenyl]ethyl]amino]benzoate** Structure in Context: • *Claim #13:* ...minosalicylic acid, probucol, hydroxypropylmethyl cellulose, LS-2904, **#STR#** (ester). 14. A method for inhibiting a...
		

UI for Assessing AAS Patents

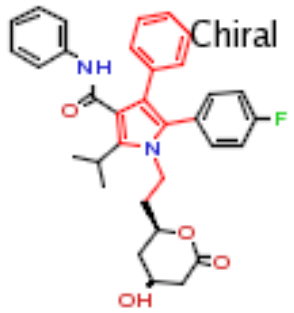
Notiora

accelrys®

Common Fragment



Lipitor w/Fragment

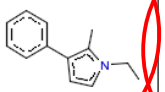
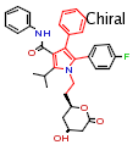
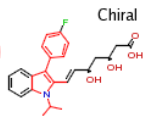
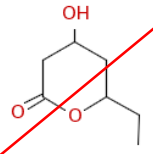
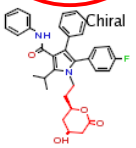
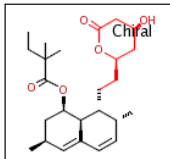
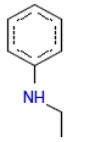
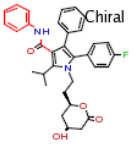
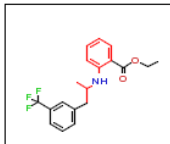
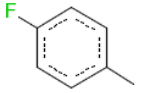
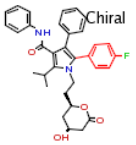
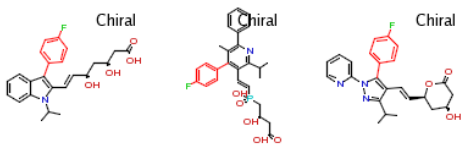


US 5,470,845: Methods
the treatment of atheros
1995/11/28: Bristol-Myers

US 5,470,845: Methods of using .alpha.-phosphonosulfonate squalene synthetase inhibitors including the treatment of atherosclerosis and hypercholesterolemia

1995/11/28: Bristol-Myers Squibb Company (Princeton, NJ): Magnin, David R.; Biller, Scott A.; ... Sulsky, Richard B.

.alpha.-Phosphonosulfonate compounds are provided which inhibit the enzyme squalene synthetase and thereby inhibit cholesterol biosynthesis. These compounds have the formula ##STR1## wherein R^2 is OR^5 or R^5a ; R^3 and R^5 are independently H, alkyl, arylalkyl, aryl or cyc... ...cloheteroalkyl, cycloheteroalkylalkyl; as further defined above; including pharmaceutically acceptable salts and or prodrug esters of the phosphonic (phosphinic) and/or sulfonic acids.

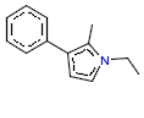
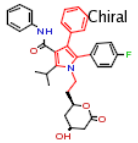
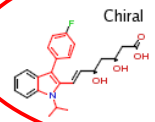
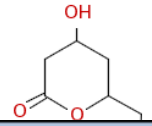
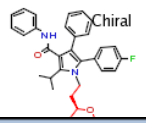
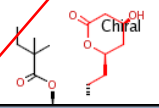
Common Fragment	Lipitor w/Fragment	Patent Structures containing Fragment
		 Structure Name: **fluvastatin** Structure in Context: • *Claim #13:* ...uctase inhibitor is pravastatin, lovastatin, simvastatin, velostatin, **#STR#**, rivastatin, compactin, SDZ-63,370 (San...
		 Structure Name: **simvastatin** Structure in Context: • *Claim #13:* ...2 wherein the HMG CoA reductase inhibitor is pravastatin, lovastatin, **#STR#**, velostatin, fluvastatin, rivastatin, c...
		 Structure Name: **ethanol 2-[[1-methyl-2-[3-(trifluoromethyl) phenyl]ethyl]amino]benzoate** Structure in Context: • *Claim #13:* ...minosalicylic acid, probucol, hydroxypropylmethyl cellulose, LS-2904, **#STR#** (ester). 14. A method for inhibiting a...
		

UI for Assessing AAS Patents

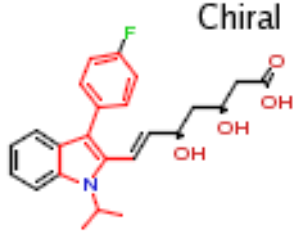
US 5,470,845: **Methods of using .alpha.-phosphonosulfonate squalene synthetase inhibitors including the treatment of atherosclerosis and hypercholesterolemia**

1995/11/28: Bristol-Myers Squibb Company (Princeton, NJ): Magnin, David R.; Biller, Scott A.; ... Sulsky, Richard B.

.alpha.-Phosphonosulfonate compounds are provided which inhibit the enzyme squalene synthetase and thereby inhibit cholesterol biosynthesis. These compounds have the formula ##STR1## wherein R^2 is OR^5 or R^5a ; R^3 and R^5 are independently H, alkyl, arylalkyl, aryl or cyc... ...cloheteroalkyl, cycloheteroalkylalkyl; as further defined above; including pharmaceutically acceptable salts and or prodrug esters of the phosphonic (phosphinic) and/or sulfonic acids.

Common Fragment	Lipitor w/Fragment	Patent Structures containing Fragment
		 Structure Name: **fluvastatin** Structure in Context: • *Claim #13:* ...uctase inhibitor is pravastatin, lovastatin, simvastatin, velostatin, **#STR#**, rivastatin, compactin, SDZ-63,370 (San...
		 Structure Name: **simvastatin** Structure in Context: • *Claim #13:* ...2 wherein the HMG CoA reductase inhibitor is pravastatin, lovastatin, **#STR#**, velostatin, fluvastatin, rivastatin, c...

Patent Structures containing Fragment



Structure Name: **fluvastatin**

Structure in Context:

- *Claim #13:* ...uctase inhibitor is pravastatin, lovastatin, simvastatin, velostatin, **#STR#**, rivastatin, compactin, SDZ-63,370 (San...

Summary

- **Q: Can we match a target to a patent as a whole?**
  - and is it interesting? Is it novel?

- **A: Yes, yes, and yes**
  - Aggregate Asymmetric Search
  - Compliments typical structure-based search
  - Finds patents with only multiple fragments covering target

- **UI for Assessing AAS Patents**

- **Future Work**
  - Literature, Corporate Documents

brian@notiora.com