Development of Technology for Transforming Analysis Of Patent Information

Agenda

Introduction

Evaluation Of Markush Structure Enumeration Technology

Q/ A

anton.fliri@pfizer.com

Challenges

How To Maximize Productivity In Light Of Increasing Complexity Of R&D Paradigms

Challenges-cont.

Information Overload

- Patent Examination Systems
- Scientist

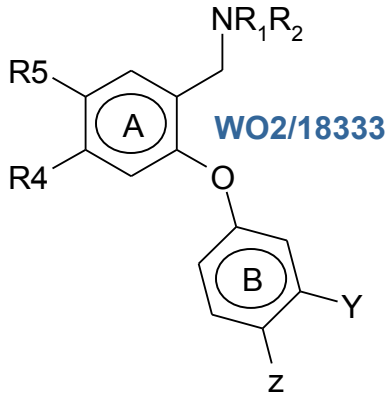
The Burden Of Proof That Inventions Are Novel And Patentable is Shifting To Inventors

Needed

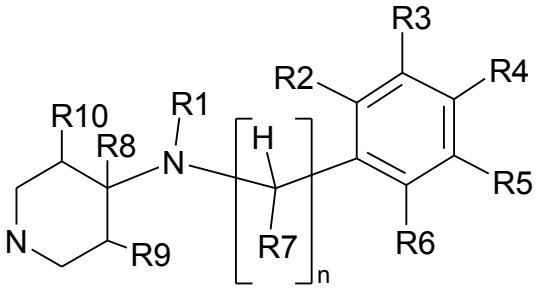
Methods For Increasing The Efficiency Of Translating Internal And External Know-how Into High Value IP

The Problem: Markush Structures

The Packaging Obscures The Content



WO2/18333



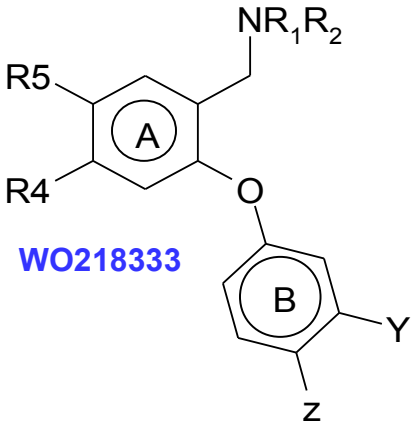
WO52858

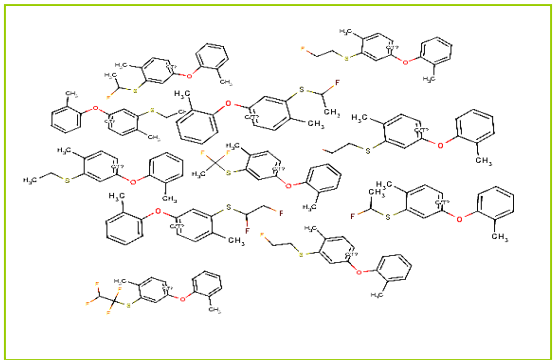
1. How Similar Is The Content?
2. What Properties Has The Content?

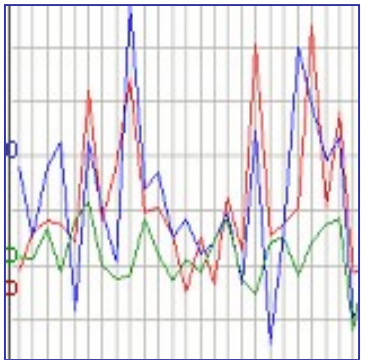
The Outcome Of This Analysis Determines The Quality Of IP And The Security Of Investments

Collaborations With INPI and Decript:

Enumeration Of Markush Structures In The MMS Data Base



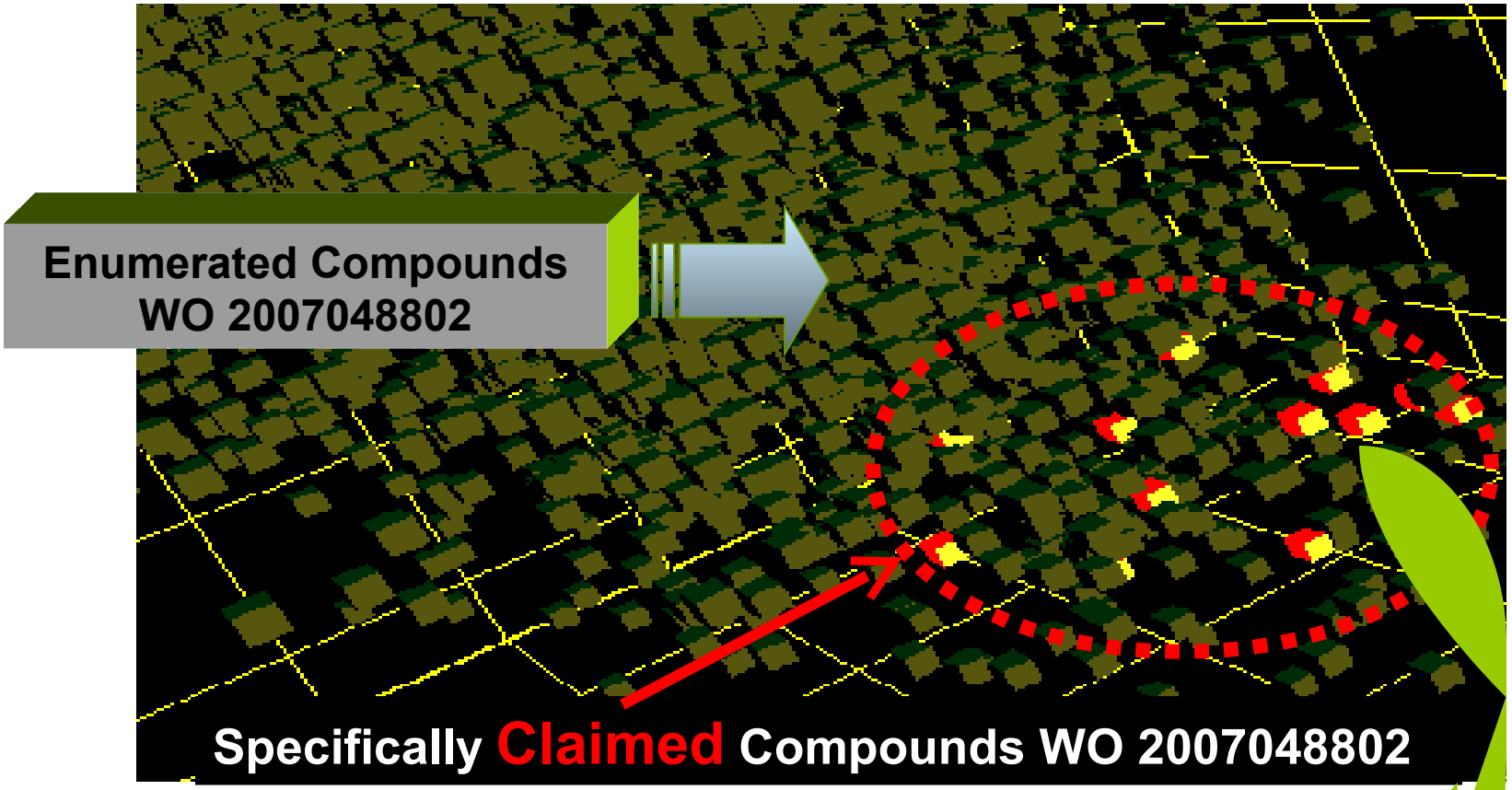




Enumeration Output is Converted into

- Structure Fingerprints (Atom Pair)
- Molecular Property Fingerprints

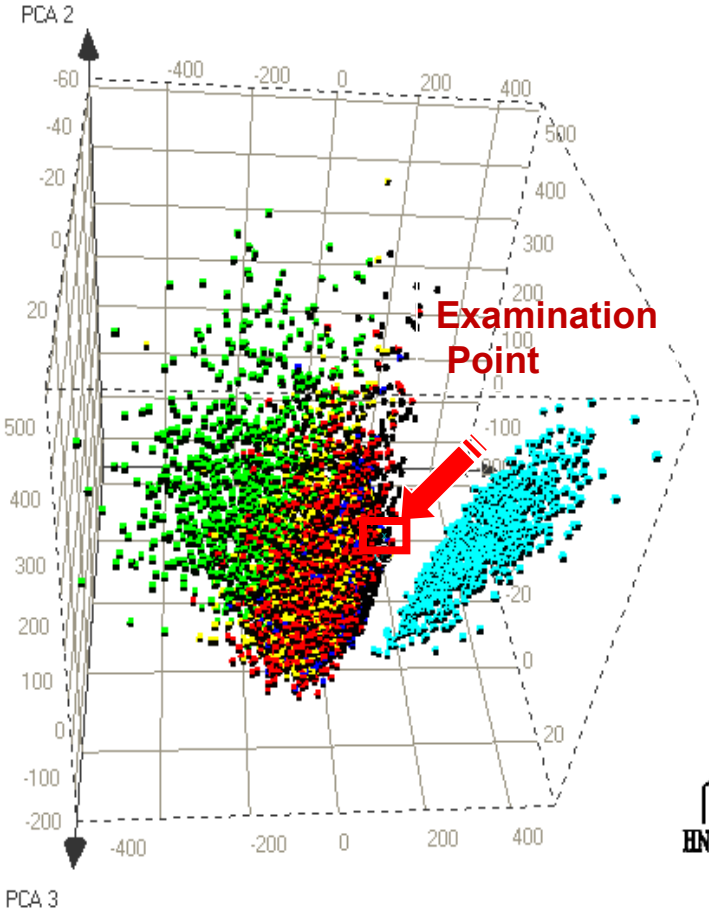
Does The Enumeration Of MMS Data Capture The MKST Space Occupied By Exemplified Compounds?

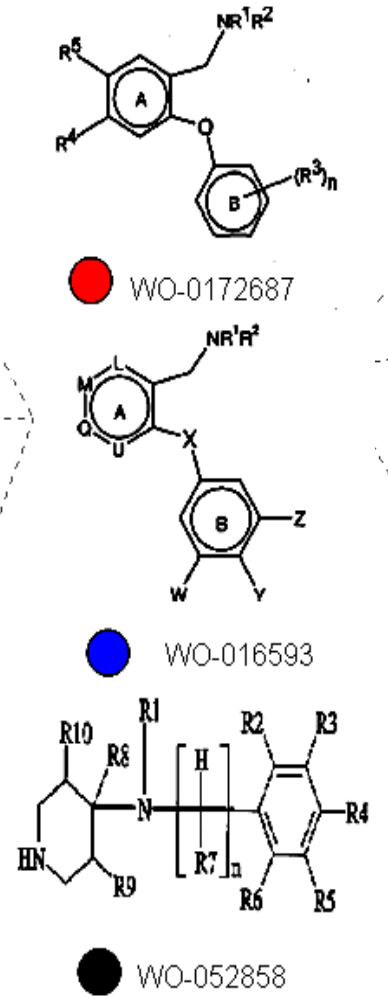


Enumeration Adequately Samples The Claimed MKST Space

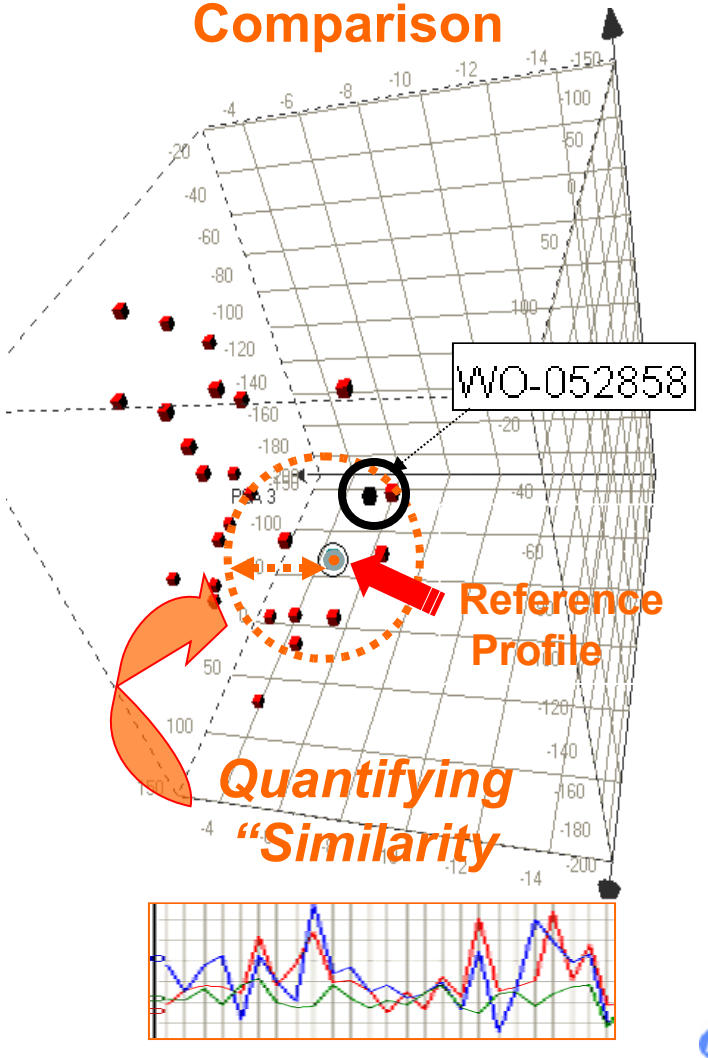
Analysis Of Molecular Fingerprint Similarity: Determining MKST Overlap

Principal Component Analysis

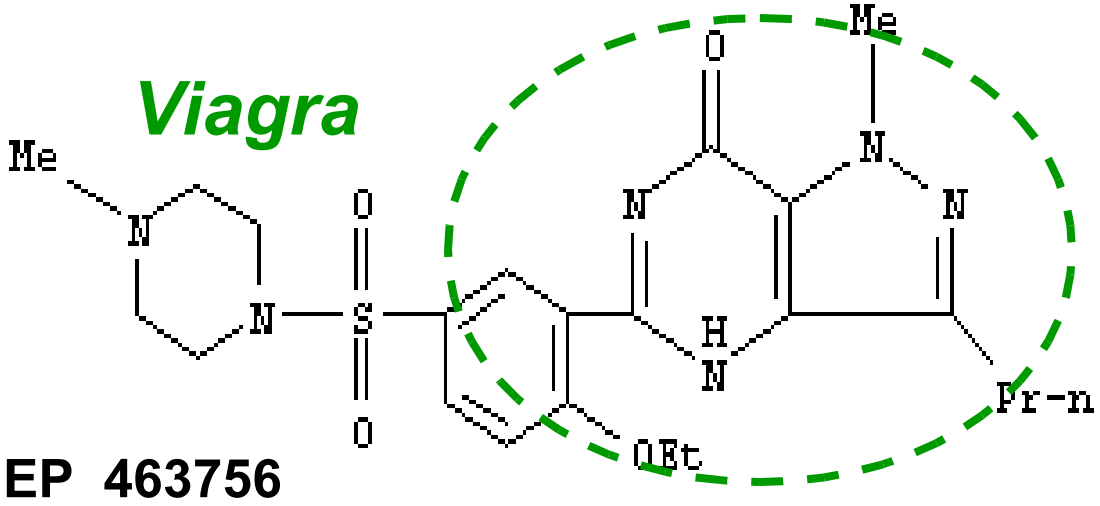




Fingerprint Profile Comparison

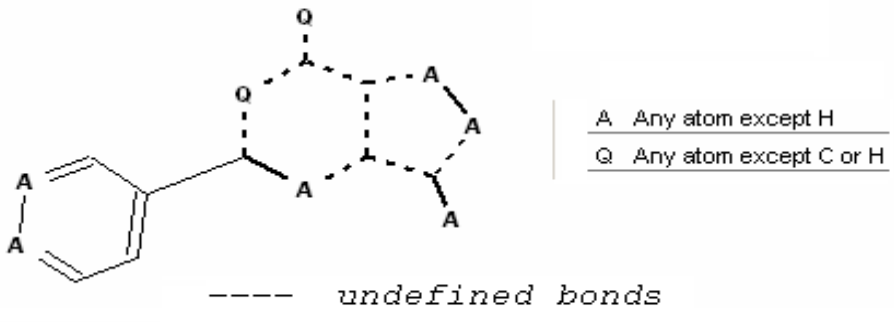


Application: Who Is Working With Pfizer Structures?



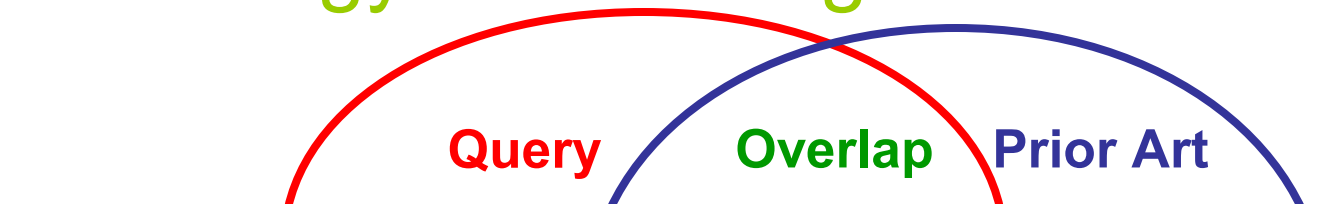
EP 463756

MKST "like" Query using SCIFINDER (CAS)



- **~ 400 patents found in the CAS data base**
- **Which one of these 400 documents is of interest**
- **354 patents were mapped to the MMS database and enumerated ~ 4000,000 Cpds**

Strategy for Ranking Patent Documents



#	Patent	70-79%	60-69%	50-59%	40-49%	Comment
1	WO2005037257		25	816	19272	Similar
2	WO2004005324		3	134	5445	Similar
3	US20050239867		3	98	2476	Similar
4	WO2005041879		2	363	12735	Similar
5	WO2002030470		2	223	15332	Similar
6	WO2006092691		1	113	6040	Similar
7	EP1176147			202	12016	Lower similarity
8	US20030040514			200	18486	Lower similarity

Determining The Chemical Structure Similarity Of Enumeration Results Identifies MKST Overlap And The Structure Similarity Of Compounds Residing In Overlapping MKST Regions

Comparison Of Ranking Results Specifically Claimed V. Enumerated Compounds

Ranking Of Documents Using Enumeration Results

#	Patent	70-79%	60-69%	50-59%	40-49%	Comment
1	WO2005037257		25	816	19272	Similar
2	WO2004005324		3	134	5445	Similar
3	US20050239867		3	98	2476	Similar
4	WO2005041879		2	363	12735	Similar
5	WO2002030470		2	223	15332	Similar
6	WO2006092691		1	113	6040	Similar
7	EP1176147			202	12016	Lower similarity
8	US20030040514			200	18486	Lower similarity

How relevant are these patents?

Ranking Of Enumeration Results Identifies Documents Containing Structures With Higher Similarity

Ranking Of Documents Using Specifically Claimed Compounds

#	Patent	70-79%	60-69%	50-59%	40-49%	Comment
1	WO2002013798			10	144	Lower similarity
2	WO2005105096			9	354	Lower similarity
3	EP1199070			4	40	Lower similarity
4	WO2002024679			2	153	Lower similarity
5	WO2002049651			1	18	Lower similarity
6	DE10063224			1	18	Lower similarity
7	WO2003053975			1	10	Lower similarity
8	US20060128729				326	Worth Inspecting

Analysis of Ranking Results

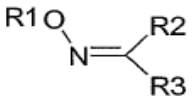
Result # 1:

WO2005037257

Ranking identifies that this reference contains direct structure analogs of query despite the language of the multiple page claim and dissimilar MKST

This prior art reference would likely have been missed in a conventional analysis

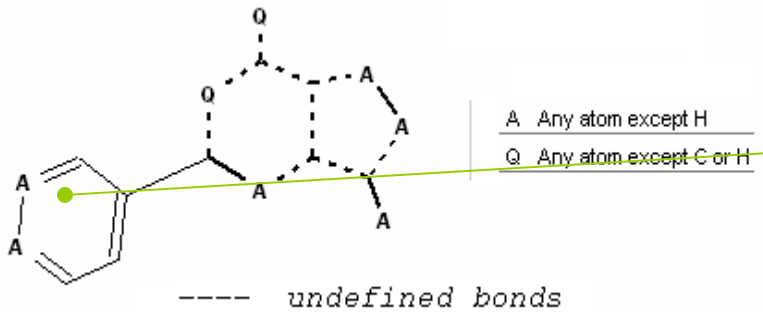
3. Verbindungen der allgemeinen Formel A2

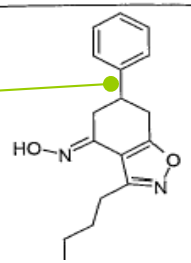


A2

worin

- R1, R2 und R3 gewählt sind aus der Gruppe, die besteht aus Wasserstoff, unsubstituiertem oder substituiertem, geradkettigem oder verzweigtem C_1- bis C_{12}- Alkyl, C_2- bis C_{12}-Alkenyl und C_2- bis C_{12}-Alkinyl, Hydroxy, Thiol, C_1- bis C_{12}-Alkoxy, C_1- bis C_{12}-Alkylthio, unsubstituiertem oder substituiertem, unkondensiertem oder kondensiertem, gegebenenfalls ein oder mehrere Heteroatome aus der Gruppe N, O, P und S enthaltendem Aryl und Cycloalkyl, unsubstituiertem oder substituiertem Amino, unsubstituiertem oder substituiertem Carbonyl, unsubstituiertem oder substituiertem Thiocarbonyl und unsubstituiertem oder substituiertem Imino; und
- die heteroaromatischen oder heterocyclischen Reste über ein C-Atom oder ein Heteroatom mit der Grundstruktur der allgemeinen Formel A2 verbunden sind
- und Tautomere, Stereoisomere der Verbindungen der allgemeinen Formel A2 und

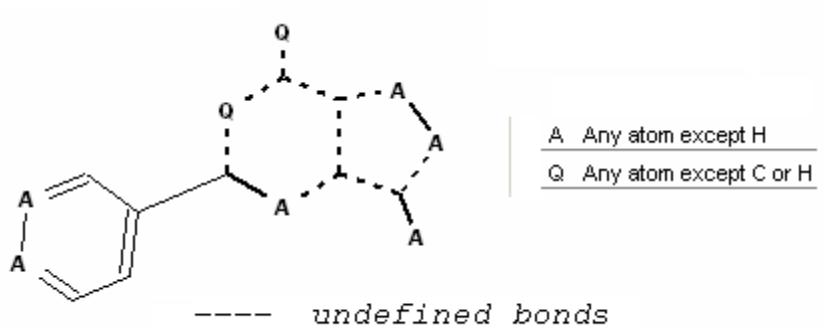


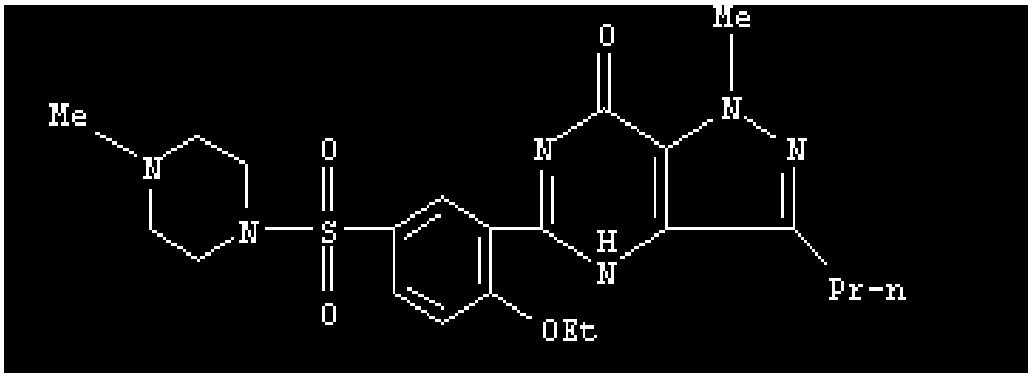
Compound ID.	Structure
A2.001	

Analysis of Ranking Results

Result 2 :

WO2004005324 claims Viagra

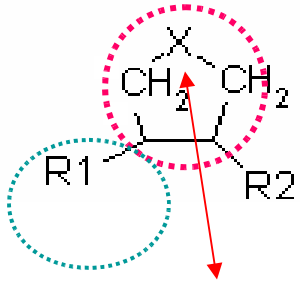




Case Study 2

QUERY

Enumeration of 1400 patents

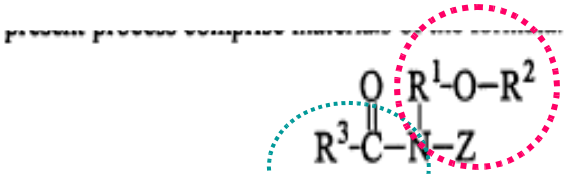


X= CH2. O. SO2. N-R3. C=O

R1= HN-C=O-R4

R2 = C1-C6 alkyl.

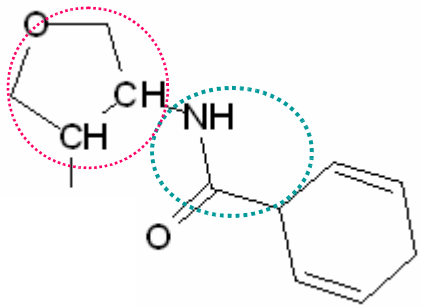
Ranking Result #2 WO9522519



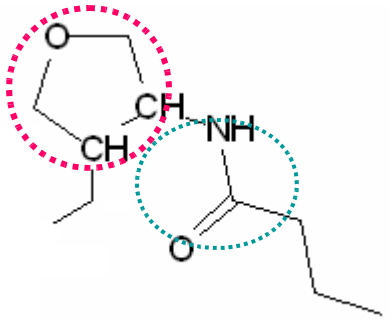
(II)

wherein in formulas (I) and (II): R^3 is C_5-C_{31} hydrocarbyl, preferably C_9-C_{17} hydrocarbyl, including straight-chain and branched-chain alkyl and alkenyl, or mixtures thereof; R^1 is C_2-C_8 hydrocarbyl including straight-chain, branched-chain and cyclic (including aryl), and is preferably C_2-C_4 alkylene, i.e., $-CH_2CH_2-$, -

QUERY ⇩ *Enumeration*



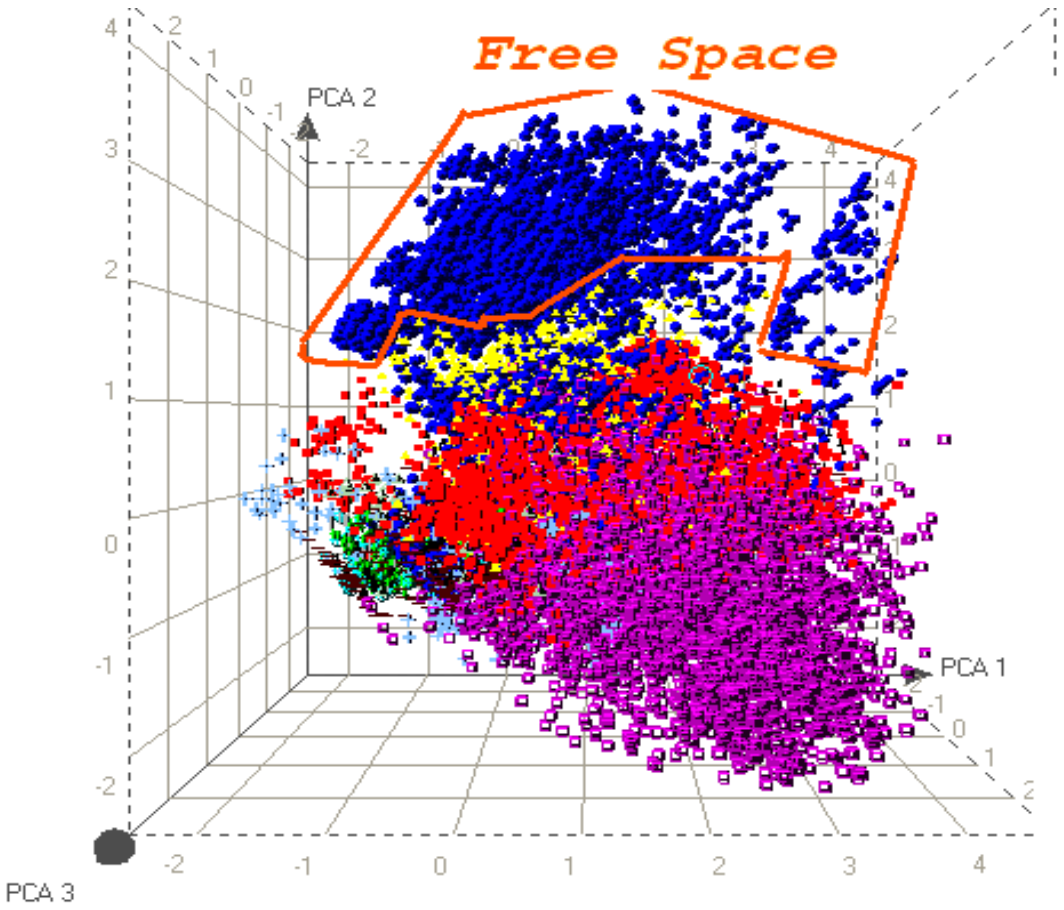
WO9522519 ⇩ *Enumeration*



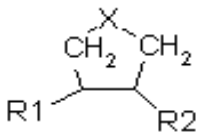
This Reference Would Have Likely Been Missed

“Rough” Check of “FTO Space” : Analysis of 1400 Patents

PCA analysis > 400000 compounds
Provided by enumeration of the
Of top 10 ranking patents



Querry



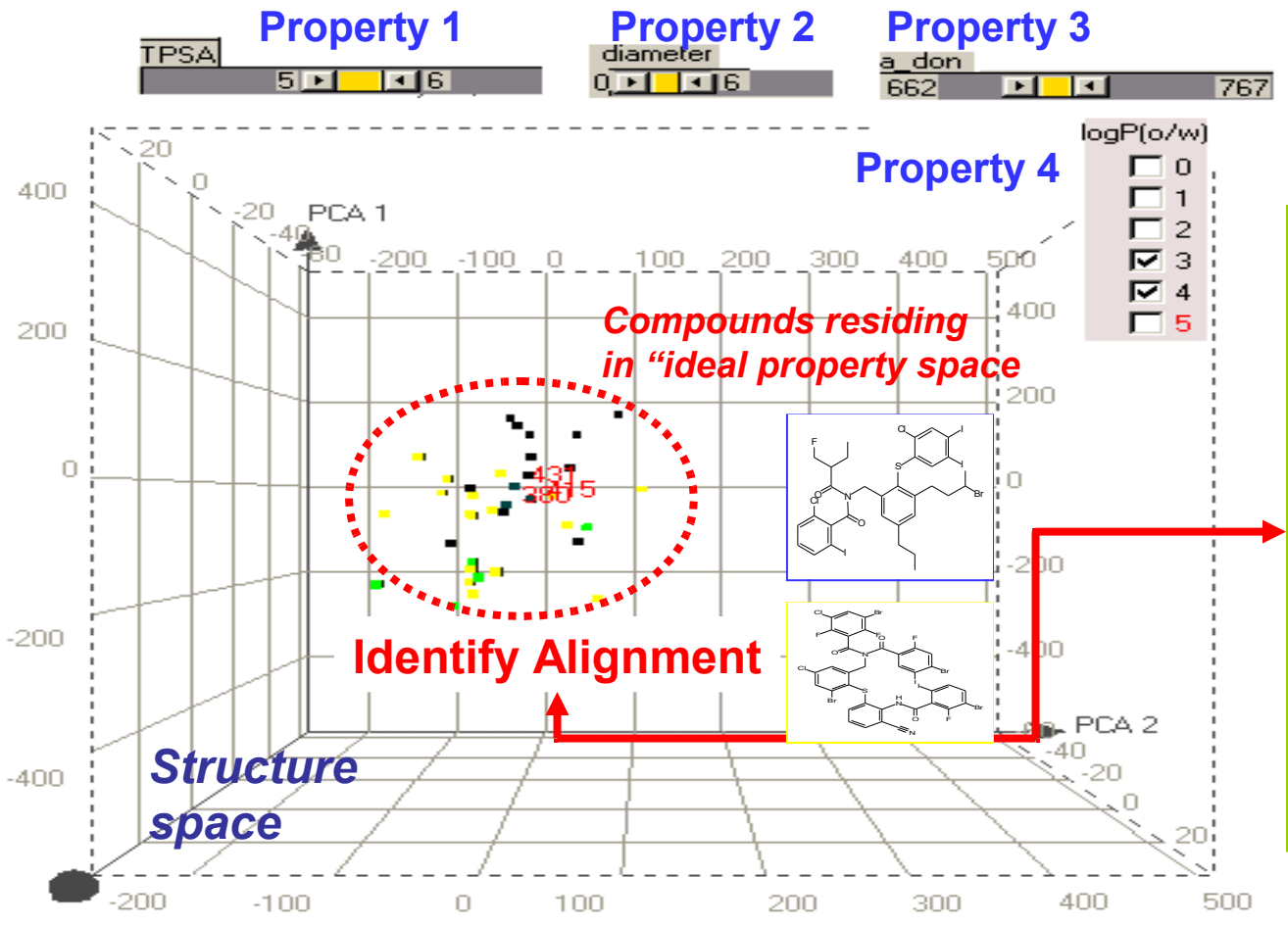
X= CH2. O. SO2. N-R3. C=O

R1= HN-C=O-R4

R2 = C1-C6 alkyl.

Idea Generation Support

Literature Search → **Identify Scaffolds Of Interest** → **Enumerate And Calculate Molecular Properties**



1. *Alignment Of Scaffolds In Ideal Property Space Can Be Used To Create Models*
2. *Create New MKST Based On Property Model,*
3. *Re-indexing of MKST In Ideal Property Space and claim level enumeration*
4. *Evaluate FTO*

Summary

The Technology:

1. Is Useful For Identifying Relevant Prior Art Documents Buried In Large Number Of Hits In A Patent Search
2. Enables Scientists To Compare Properties Of The Contents Of Markush Structures

Summary

Technology enables scientists to increase the efficiency in identifying those sections of the chemical space that are associated with the “best molecular properties” by being able to use the knowledge encoded in form of MKST in patent data bases.

Project Boundaries

- Precise FTO analysis requires manual indexing
- Development of indexing tools rendering MKST enumerable station which will enable trained scientists to produce “ claim level” accuracy of enumerations
- This precision analysis will likely always require a certain degree of human supervision since it involves the concise interpretation of patent claims which includes provisos and examiner imposed restrictions
- This service is currently performed for Pfizer by Decript

Acknowledgements

Pfizer

- Christopher Kibbey
- Scott Mente
- Matthias Nolte
- Meihua Tu
- Garth Butterfield
- Jacob Schwarz
- Gregory Kaufmann
- Bruce Lefker
- Martin Petterson

INPI

- Serge Chambaud
- Isabelle Leclerque
- Jaques Feldmar
- Philipp Borne

Decript

- Erwan Moysan
- Pierre Benichou
  Jean Killian
- Matthias Nolte
  Victoria Willis
- Staff of Awat
- Staff of Eilix