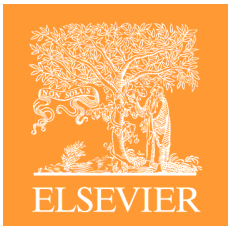# MDL Patent Chemistry Database

Positioning, Content, Competitive Advantages
[Confidential]

Presented by: Eva Seip
**Title:**   Senior Product Manager
Date: 27-October-2004

MDL® Patent Chemistry Database

What is it?

- A new structure-searchable CrossFire database
  - mainly for end-users, but with interesting features for information professionals also
  - indexing chemical reactions, substances and substance related information from Chemistry and LifeScience patents (World, U.S. and European) since 1976
  - addressing the following business critical issues:
    - **More effective synthesis planning**
    - **Better bioactivity profiling**
    - **Easy patent relevance check**
- Integrated with other data sources on DiscoveryGate soon (mid 2005)
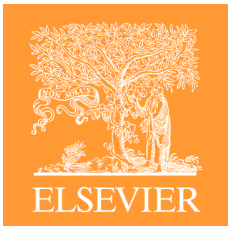
Content and Coverage

- **Covered International Patent Classes [IPC]**

    C07   Organic Chemistry*

    A61K   Drugs [Medicinal, Dental, Cosmetic Preparations]

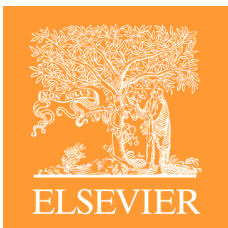   **A01N**   Biocides [Agrochemicals, Disinfectants, etc.]

    C09B   Dyes [can be pharmacological active]

    * Polymers indexed from 2004 onwards, when mentioned
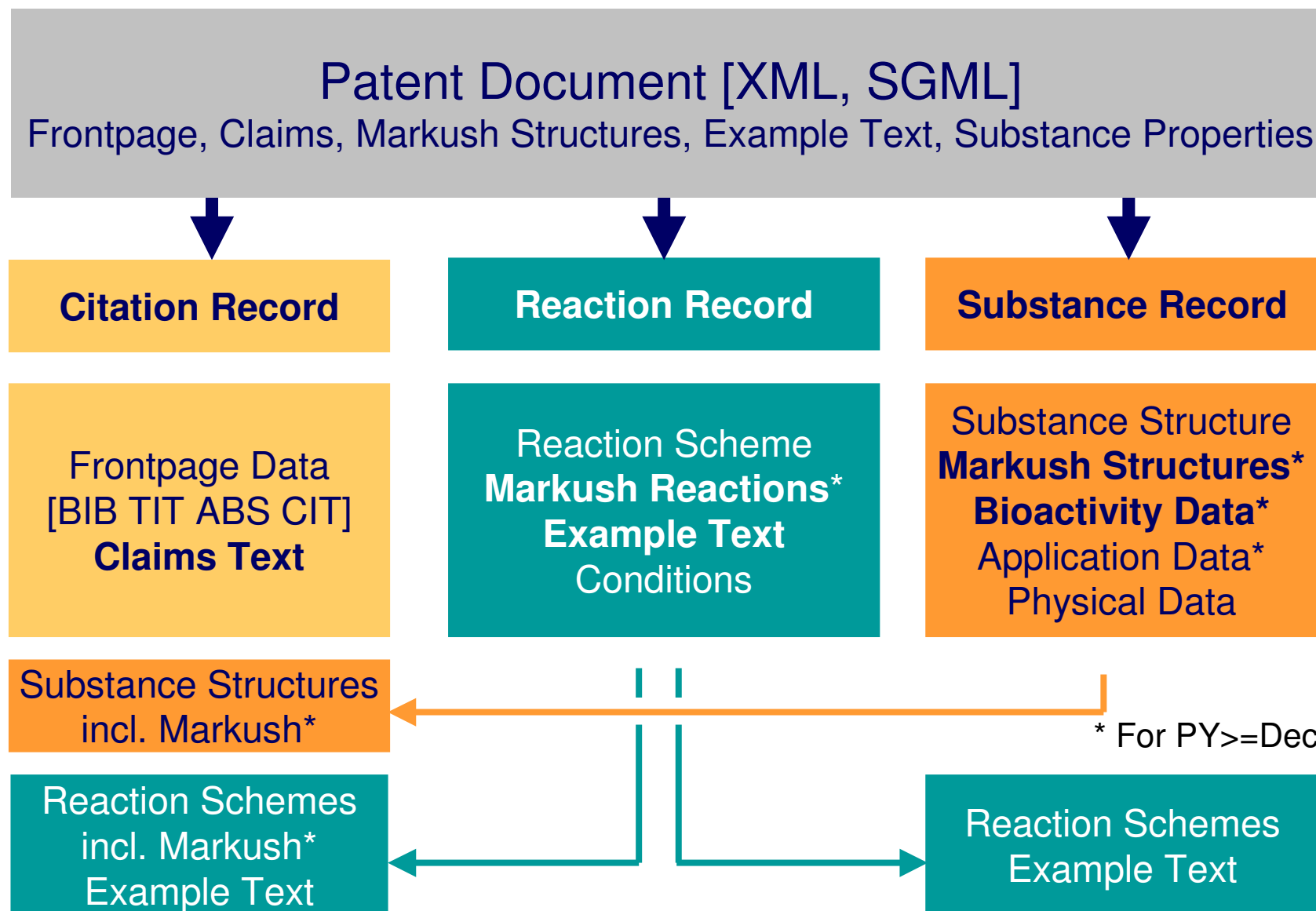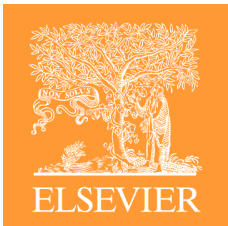   in these 4 IPC's (e. g. formulations)

- **Covered Patent Agencies**

    - World (WO), United States (US), European (EP) patents
      from Publication Year 1976 onwards

Database Design – Three "Contexts"

Patent Document [XML, SGML]
Frontpage, Claims, Markush Structures, Example Text, Substance Properties

Citation Record	Reaction Record	Substance Record
Frontpage Data [BIB TIT ABS CIT] **Claims Text**	Reaction Scheme **Markush Reactions*** **Example Text** Conditions	Substance Structure **Markush Structures*** **Bioactivity Data*** Application Data* Physical Data

Substance Structures incl. Markush*

Reaction Schemes incl. Markush* Example Text

Reaction Schemes Example Text

* For PY>=Dec 2003

Competitive Advantages - Citations

Database Content:

At release: approx. 340,000 Patent Citations

Anticipated Growth: 35,000 Patents / Year

Your Benefits: Easy Relevance Check

- **Full claims text** together with Main Markush structures and main Markush Reactions*
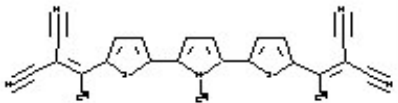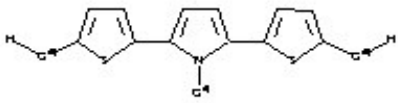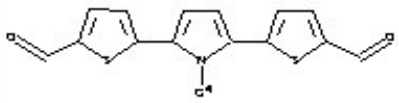
- **Expanded Markush Structure** with full substituent list

- Defined **compounds related to Markush** structure

- Easy relevance check without time consuming access to original patent document

* On release, available for patents published since December 2003

Claim & Markush Structures to explore fast scope of patent

Title / Abstract / Claims

Title	1-substituted 2,5-dithienyl pyrrole derivatives and film-forming materials
Abstract	A 1-substituted 2,5-dithienylpyrrole derivative having the following formula (I). in which R is hydrogen, a substituted or non-substituted alkyl group, or a substituted or non-substituted aromatic group, Y is hydrogen or cyano group, it can be involved the case that one of Ys may be hydrogen and the other may be cyano group, and n is an integer of 1 to 3. The derivative is used for forming films.
Claims	What is claimed is:
	1. A 1-substituted 2,5-dithienylpyrrole derivative having the following formula (I).
	[Figure]
	in which R, is hydrogen, a substituted or non-substituted alkyl group, or a substituted or non-substituted aromatic group, Y is hydrogen or cyano group, provided that one of Ys may be hydrogen and the other may be cyano group, and n is an integer of 1 to 3.
Language	English
Number of pages	8
Manually excerpted	yes

Claim Text

Markush Structures

Markush PRN   19, 48, 49, 50

PRN=19   PRN=48   PRN=49

Markush* Structure Display

* On release, available for patents published since December 2003

Highlight: Expanded Markush Structure

Substance Characterization

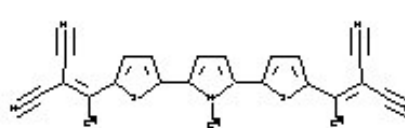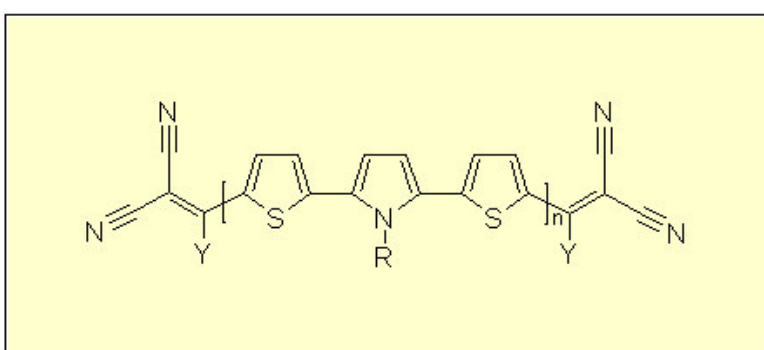|	
Patent Compound Registry Number	**305**
Substance Type	*Markush*
	organic compound
Entry Date (YYYY/MM/DD)	2003/10/16
Update Date (YYYY/MM/DD)	2003/10/16
Referencing Compounds	*click here*

PRN=19

Compressed MARKUSH: Click to expand details

Markush Viewer

To Report   Print   Help   OK

Label	Value	Size	Attributes	Substitution	Frequency
Y	H				
	CN				
R	H				
	alkyl	1-30C	os		
	aromatic group		os		
	formyl				
	acyl				
	alkoxycarbonyl				
	alkenyl				
n	1-3				

Generic Symbols as written in Full-text

* On release, available for patents published since December 2003

Defined Substances related to Markush - Easy relevance check

Substance Characterization

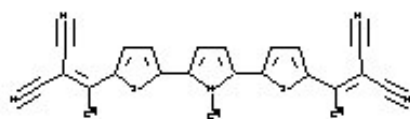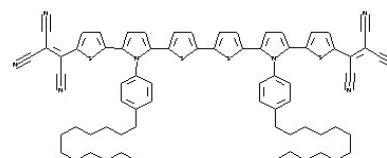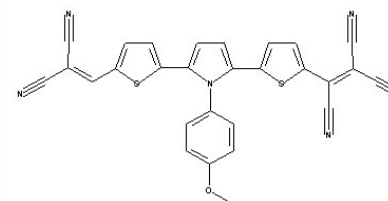|	
Patent Compound Registry Number	305
Substance Type	*Markush*
	organic compound
Entry Date (YYYY/MM/DD)	2003/10/16
Update Date (YYYY/MM/DD)	2003/10/16
Referencing Compounds	*click here*

PRN=19

Compressed MARKUSH: Click to expand details

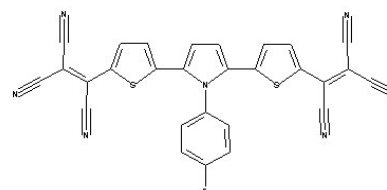**Representatives of the given Markush Structure:**

Hit 1 PRN=9 C70H70N8S4

Hit 2 PRN=11 C28H14N6OS2
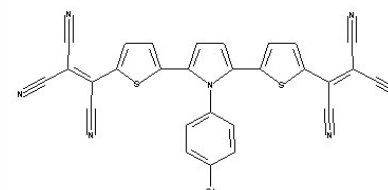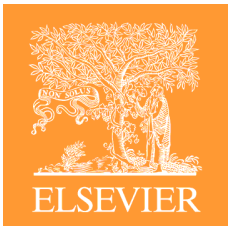
Hit 3 PRN=13 C28H10FN7S2

Hit 4 PRN=14 C28H10ClN7S2

* On release, available for patents published since December 2003

Competitive Advantages - Reactions

Approx. 1.5 M reactions (structure-searchable; 1976 onwards)
Strong growth anticipated: ~ 500,000 reactions/year

**Your Benefit: More Effective Synthesis Planning**

- **Complete Reaction Descriptions**
  Immediately reproduce chemistry without the need to order the patent document

- **InfoChem ClassCodes**
  Link to Similar Reactions in all other MDL Databases

- **Markush Reaction Display***
  Explore the scope of patent coverage and reaction scope

* On release, available for patents published since December 2003

Experimental Text
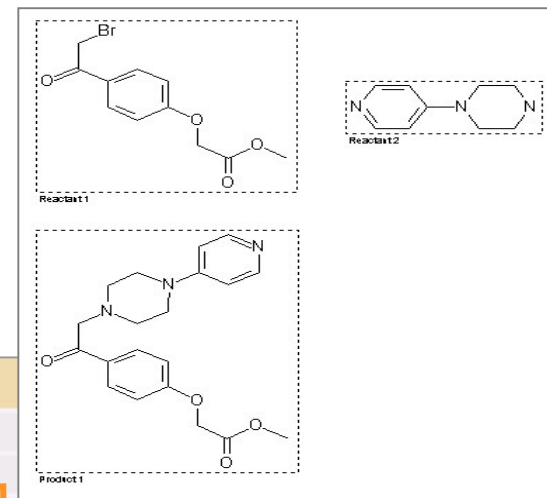
Reaction Details

Reaction Classification	Preparation
Example Text	EXAMPLE 1
	methyl 4-[2-[4-(4-pyridyl)piperazin-1-yl]acetyl]phenoxyacetate

A solution of methyl 4-bromoacetylphenoxyacetate (4.3 g) in acetonitrile (50 ml) was added dropwise over 40 minutes to a stirred solution of 1-(4-pyridyl)piperazine (4.9 g) in acetonitrile (100 ml).

Stirring was continued for a further 1.5 hours, then the solution was filtered and the filtrate evaporated in vacuo.

The solid residue was triturated with water (50 ml), then dried and suspended in methylene chloride (50 ml).

The suspension was then filtered and the filtrate concentrated to a small volume.

Purification by flash chromatography on neutral alumina eluding first with dichloromethane, then 0. 5percent v/v methanol/dichloromethane and finally 1percent v/v methanol/dichloromethane gave the title compound, 1.93 g, as a solid: m.p. 150.deg.-152.deg. C.; NMR(d6DMSO) 88.14(2H,d), 7.98(2H,d), 7.03 (2H,d), 6.78(2H,d), 4.90(2H,s), 3.83 (2H,s), 3.72(3H,s), 3.34(4H,bt), 2.65(4H,bt); m/e 370 (M+H)+; calculated for C20H23N3O4: C, 65.0; H, 6.3; N, 11.4. found: C, 65.2; H, 6.4; N, 11.3percent.
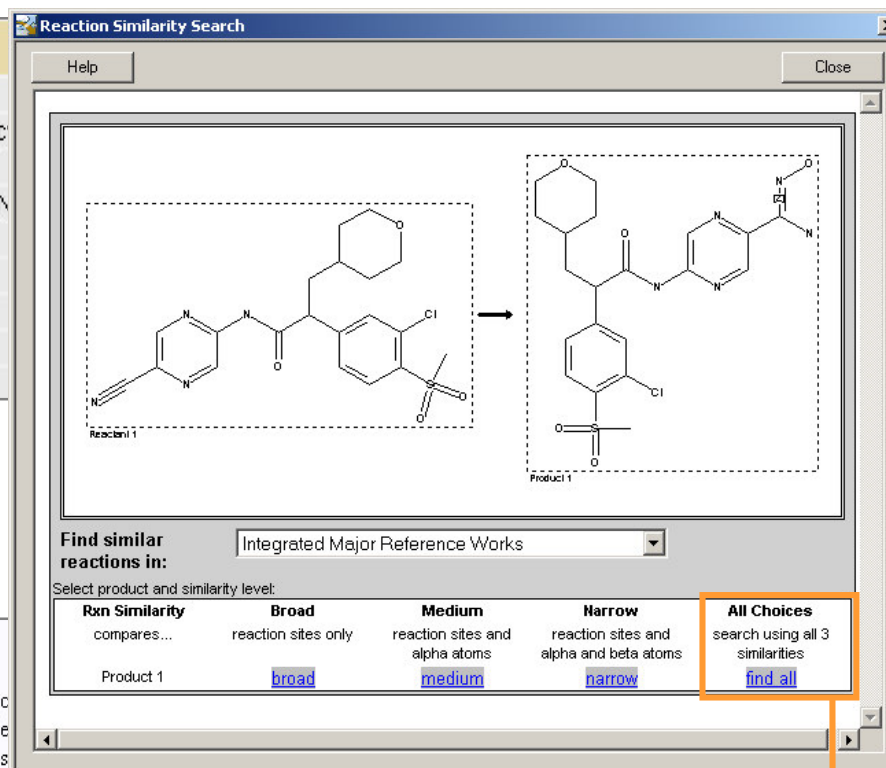
Easy relevance check
- **Example Number**
- **Example Title**
- **Example Text**
- **Spectral Data (Product)**

Find Similar Reactions in (other) db's

Reaction Identification

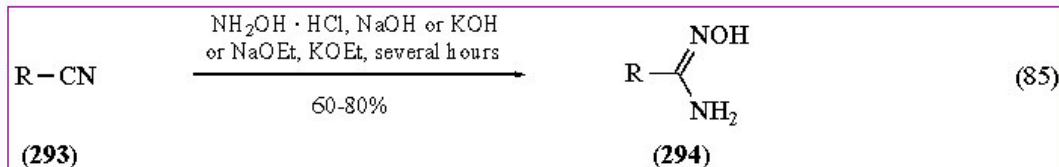|	
Reaction RN	4966189
Reactant PRN	3259400 2-(3-Chloro-4-methanesulfonylphenyl)-N-(5-c... (tetrahydro-pyran-4-yl)-propionamide
Product PRN	3412781 2-(3-chloro-4-methanesulfonylphenyl)-N-[5-(N... pyrazin-2-yl]-3-(tetrahydropyran-4-yl)propionamide
Reaction Specification	full reaction
Entry Date (YYYY/MM/DD)	2004/03/28
Update Date (YYYY/MM/DD)	2004/03/28
Find Similar Reactions	click here

Reaction Similarity Search

Find similar reactions in: Integrated Major Reference Works

Select product and similarity level:

Rxn Similarity compares...	Broad reaction sites only	Medium reaction sites and alpha atoms	Narrow reaction sites and alpha and beta atoms	All Choices search using all 3 similarities
Product 1	broad	medium	narrow	find all

5.21.2.4.1 From nitriles

Nitriles (**293**) are the most frequently employed starting materials for this class of c... hydrochloride with sodium carbonate, sodium or potassium hydroxide, or sodium e... C to give (**294**) (Equation (85)) [8]. To avoid the separation of the product from s... anhydrous methanol or ethanol is used [130]. The highest yields are obtained when 15% excess of hydroxylamine in butanol is used and the mixture is left for 48 h at 60 °C. The product separates as a practically pure crystalline material [189].Diamidoximes (**297**) can, in principle, be prepared according to the same method from dinitriles, such as cyanogen (**295**) with hydroxylamine or alternatively, from its addition compounds with aniline, for example, diphenyloxamidine (**296**), which is treated with hydroxylamine hydrochloride (Scheme 34) [16] [17][26]. The best yields and the purest products are obtained if gaseous cyanogen is led directly in an aqueous hydroxylamine solution at 0 °C [62CVR155]. Polyoxyimidic amides [172] and polymers containing the oxyimidic amide function have been prepared from polyacrylonitrile of low molecular weight with a slight excess of hydroxylamine [158] [164].
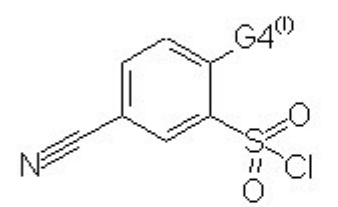
$$R-CN \xrightarrow[\text{60-80\%}]{\substack{NH_2OH \cdot HCl, \ NaOH \text{ or } KOH \\ \text{or } NaOEt, KOEt, \text{ several hours}}} R-C{\overset{\displaystyle NOH}{\underset{\displaystyle NH_2}{\|}}} \qquad (85)$$

(**293**)   (**294**)

Markush Reaction Display

Competitive Advantages - Compounds

- **1.6 million compounds** with data since 1976

- **Strong growth anticipated:** ~800,000 Compounds/Year

Your Benefit: Better Bioactivity Profiling

- **Numerical bioactivity data**

- Indexing of compounds not covered elsewhere:
  Indexing of Prophetic Compounds since 1976

  - could be made analogously to given methods

  - have a structure, but no data associated

  - normally only accessible via Markush Structure Search

  - 20-25 % more compounds than in other databases

Competitive Advantages - Compounds

Prophetic Compounds

Synthesis of methyl-(S)-2-(2-naphthalenesulfonylamino)-3-(4-hydroxyphenyl)propionate

[0029] Methyltyrosinate(1g, 4.32mmol) was suspended in 5ml of methylene chloride, pyridine(1.4ml, 17.28mmol) was added thereto, and the reaction mixture was allowed to stand until the mixture became thoroughly transparent. After the reaction solution became transparent, trimethylsilylchloride(1.1ml, 8.64mmol) was slowly added thereto at room temperature. After 1 hour, 2-naphthalenesulfonylchloride(1.08g, 4.75mmol) was added and the resulting mixture was stirred for 24 hours. To this mixture was added 3N HCl solution, which was then stirred for about 2 hours. The organic layer was separated, the aqueous layer was extracted with methylene chloride, and the organic layers were combined. The combined organic layer was dried over Na_2SO_4 and the organic solvent was removed by distillation under reduced pressure. The resulting residue was recrystallized from toluene to give 1.29g(Yield 79%) of the title compound as a white solid.

Compounds with data (e.g. yield, spectra)

^{1}H NMR(CDCl$_3$, 300MHz) δ 8.33(s, 1H), 7.91(m, 3H), 7.68(m, 3H), 6.90(d, J=8.4Hz, 2H), 6.62(d, J=8.4Hz, 2H), 5.18(d, J=9.1Hz, 1H), 4.92(br.s, 1H), 4.21(dt, J=15, 6.0Hz, 1H), 3.37(s, 3H), 3.00(dd, J=13.5, 5.7Hz, 1H), 2.93(dd, J=14.3, 6.5Hz, 1H).

The following compounds could be prepared according to the similar procedure as Preparation 2 above

Methyl-(S)-2-(p-toluenesulfonylamino)-3-(4-hydroxyphenyl)propionate.
Methyl-(S)-2-(benzenesulfonylamino)-3-(4-hydroxyphenyl)propionate.
Methyl-(S)-2-(benzothiazole-2-sulfonylamino)-3-(4-hydroxyphenyl)propionate,
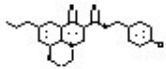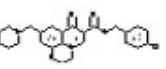
Prophetic Compounds

tial -

Numerical Bioactivity Data

Bioactivity Data 1 of 2	
Effect	inhibition of human P38 kinase alpha
Class of Effect	Pharmacology
Type	IC50
Value of Type (mole conc.-unit)	1.98 µmol/l
Species (Scientific Name)	PHAS-I (phosphorylated heat and acid st... inducible)
Named Method	In Vitro Assay: PHAS-I
Method Details	96 well plates; biotinylated PHAS-I was used as substrate (c=1.5 µM); activated human p38 kinase alpha (c=0.3 µM); gamma 32P-ATP (activity 1.2 µCi per 50 µl); title comp. prediluted in DMSO (1 percent final conc.); incubation either for 1 hour or overnigth at 30 degC; capture of biotinylated PHAS-I with 32P incorporated using high capacity streptavidin coated filter place; scintilation detection.
Location in Patent	Table 1
Ref. 1	Frontpage/Claim: 192, Fulltext: LitLink; G.D. Searle and Company; Publ.: US6335336 B1 (2002/01/01), Appl.: US09561423 (2000/04/28)
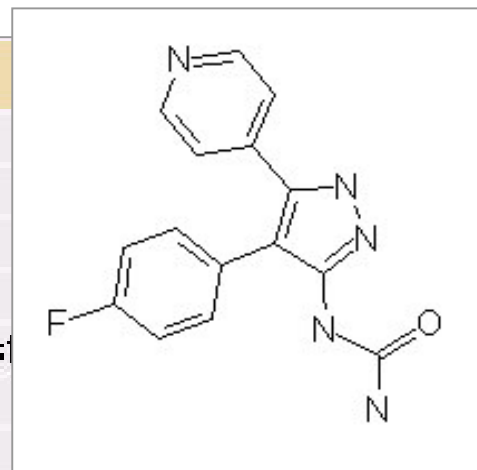
Location in Patent "Page Number"

* From Publication Year 2004 onwards
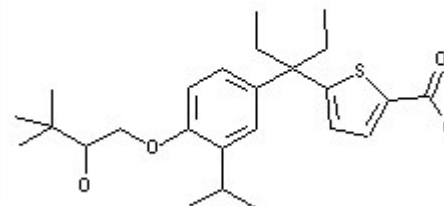
Export to Structure-Activity-Table (SAR)

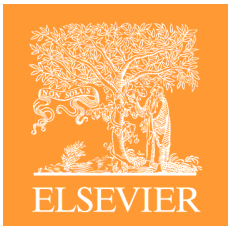| Structure	Compound RegNo	Molecular Formula	Class of Effect	Effect	Type	Value (muMol/l)	Species (Scientific Name)	Location in Patent	Citation
	179361	C22H21ClN2O3	Pharmacology	antiviral	IC50	1.8	HCMV polymerase	Page column 7-8	253; Patent; Pharmacia and Upjohn Company; Publ.: US6340680; B1; (2002/01/22), Appl.: US2000-672472; (2000/09/28)
	179360	C22H21ClN2O4	Pharmacology	antiviral	IC50	0.31	HCMV polymerase	Page column 7-8	253; Patent; Pharmacia and Upjohn Company; Publ.: US6340680; B1; (2002/01/22), Appl.: US2000-672472; (2000/09/28)
	179359	C22H17ClN2O4	Pharmacology	antiviral	IC50	1.2	HCMV polymerase	Page column 7-8	253; Patent; Pharmacia and Upjohn Company; Publ.: US6340680; B1; (2002/01/22), Appl.: US2000-672472; (2000/09/28)
	179351	C24H24ClN3O4	Pharmacology	antiviral	IC50	0.48	HCMV polymerase	Page column 7-8	253; Patent; Pharmacia and Upjohn Company; Publ.: US6340680; B1; (2002/01/22), Appl.: US2000-672472; (2000/09/28)

Predefined Export Forms

Substance Profile

NMR	
Nucleus	1H
Signals given	yes
Solvent PRN	502 chloroform-d1
Location in Patent	Page 231; 236
Original String	H-NMR (ppm, [CDC13]) : 7.69 (1H, d, 4.0 Hz), 7.08 (1H, s), 6.99 (1H, d, 6 Hz), 4.08 (1H, d, 8.0 Hz), 3.84 (1H, t, 8.0 Hz), 3.73 (1H, d, 8.0 Hz), 3. 25 Hz), 1.02 (9H, s), 0.72 (6H, t, 7.0 Hz).

Ref. 1	Frontpage/Claim: 240650, Fulltext: *LitLink* ; Patent; ELI LILLY AND COMPANY; Publ.: WO2003/101978 A1 (2003 /12/11), Appl.: WO2003-US14539 (2003/05/22)

Bioactivity Data 1 of 2	
Class of Effect	Pharmacology
Effect	osteocalcin activation
Type	EC50
Value of Type (mole conc.-unit)	0.07535 μmol/l
Cell Line / Test System	osteoblast-like cell line RG-15 (ROS 17/2.8)
Method Name	OCN (osteocalcin) Promoter Assay
Method Details	rat osteoblast-like cell line; cell were trypsinized (0.25percent trypsin) and plated into white opaque 96-well plates; after 24 h cells treated with title comp., after 48 h treatment cells were lysed and assayed for luciferase activity using Luciferase Reporter Gene Assay kit
Location in Patent	Page 353-364; 368-376

Ref. 1	Frontpage/Claim: 240650, Fulltext: *LitLink* ; Patent; ELI LILLY AND COMPANY; Publ.: WO2003/101978 A1 (2003 /12/11), Appl.: WO2003-US14539 (2003/05/22)

Application Data	
Area of Use	Pharmaceuticals
Use	Drug acting on osteoporosis
	useful for treating or preventing disease states responsive to Vitamin D receptor ligands; displays the desirable cell differentiation and antiproliferative effects of Vitamin D receptor ligand with reduced calcium mobilization (calcemic) effects
Preferred Administration Form	Emulsion, Ointment, Tablet, Capsule, Pill, Powder, Lozenge, Syrup, Aerosol
Preferred Route of Application	oral
Preferred Dosage (nonnumeric)	0.0001 - 50 mg/(kg*d)
Formulation Given	yes

A substance record is a substance profile accumulating multiple properties and reactions (from different patents)

Technical Details

- **CrossFire format database for use with**
  - CrossFire Direct
  - in-house CrossFire Server [delivery via tape, FTP]
  - DiscoveryGate (coming soon)
- **DiscoveryGate implementation**
  - DatabaseBrowser (coming soon)
  - Compound Index (coming soon)
- **Updates / Timeliness**
  - Bi-weekly updates on CrossFire Direct and DatabaseBrowser
  - Semi-automatically indexed patents with frontpage, claims, many reactions and physical data already 2 weeks after publication
  - Fully indexed record incl. Markush structures/reactions and bioactivity data a few weeks later
- **Strong Yearly Growth anticipated:**
  - + 500,000 reactions
  - + 800,000 compounds
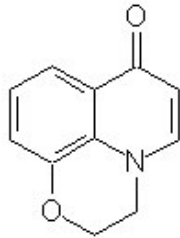 - + 35,000 patent citations

Alert Service – to stay up-to-date

Alert Conditions

Competitive Advantages - Summary

Aspect	Advantage
Reactions	**More Effective Synthesis Planning** ▪ 1.5 Million Structure-Searchable Reactions ▪ Strong growth anticipated: + 500,000 reactions/year ▪ Complete Experimental Section ▪ InfoChem ClassCodes [Find Similar Reactions] ▪ Markush Reaction Display* ▪ Location in Patent* * From Publication Year 2004 onwards
Substances	**Better Bioactivity Profiling** ▪ 1.6 Million Structure-Searchable Substances ▪ Strong growth anticipated: + 800,000/Year ▪ Prophetic Compounds ▪ Substance Profiles: Properties & reactions accumulated in one record ▪ Numerical Bioactivity Data [Structure-Activity-Relationship-Table]* ▪ Related Markush Structure* ▪ Compound Identifier in Patent* & Location in Patent*
Citations	**Easy Relevance Check** ▪ Full Claims Text – searchable together with structures/reactions ▪ Markush Structure/ Reaction Display* [Expanded Form: Substituent List] ▪ Defined Substances related to a Markush structure* ["Representatives"]

Competitive Advantages - Summary

Aspect	Advantage
Timeliness	▪ Bi-weekly Update ▪ Semi-automatically indexed patents with frontpage, claims, many compounds, reactions and phys. data already 2 weeks after publication ▪ Fully indexed record incl. Markush structures, Markush reactions and bioactivity data a few weeks later ▪ Alert Service [Export/Import Alert Profiles]
User-Interface	**Better Integration** ▪ Integration into DiscoveryGate (soon) ▪ MDL Standard Look & Feel ▪ Find Similar Reactions in all MDL databases [Reaction Class Codes] ▪ Grouping & Sorting ▪ Search & Hitset History ▪ Easy to Use: Find Field Name; Predefined Search Forms ▪ Expert Search Mode: Command Language ▪ Sophisticated Export/Report Formats [List and Table Views]
Costs	Fixed User Fee [All-You-Can-Eat] ▪ Subscription; Alert included in Fixed Fee