

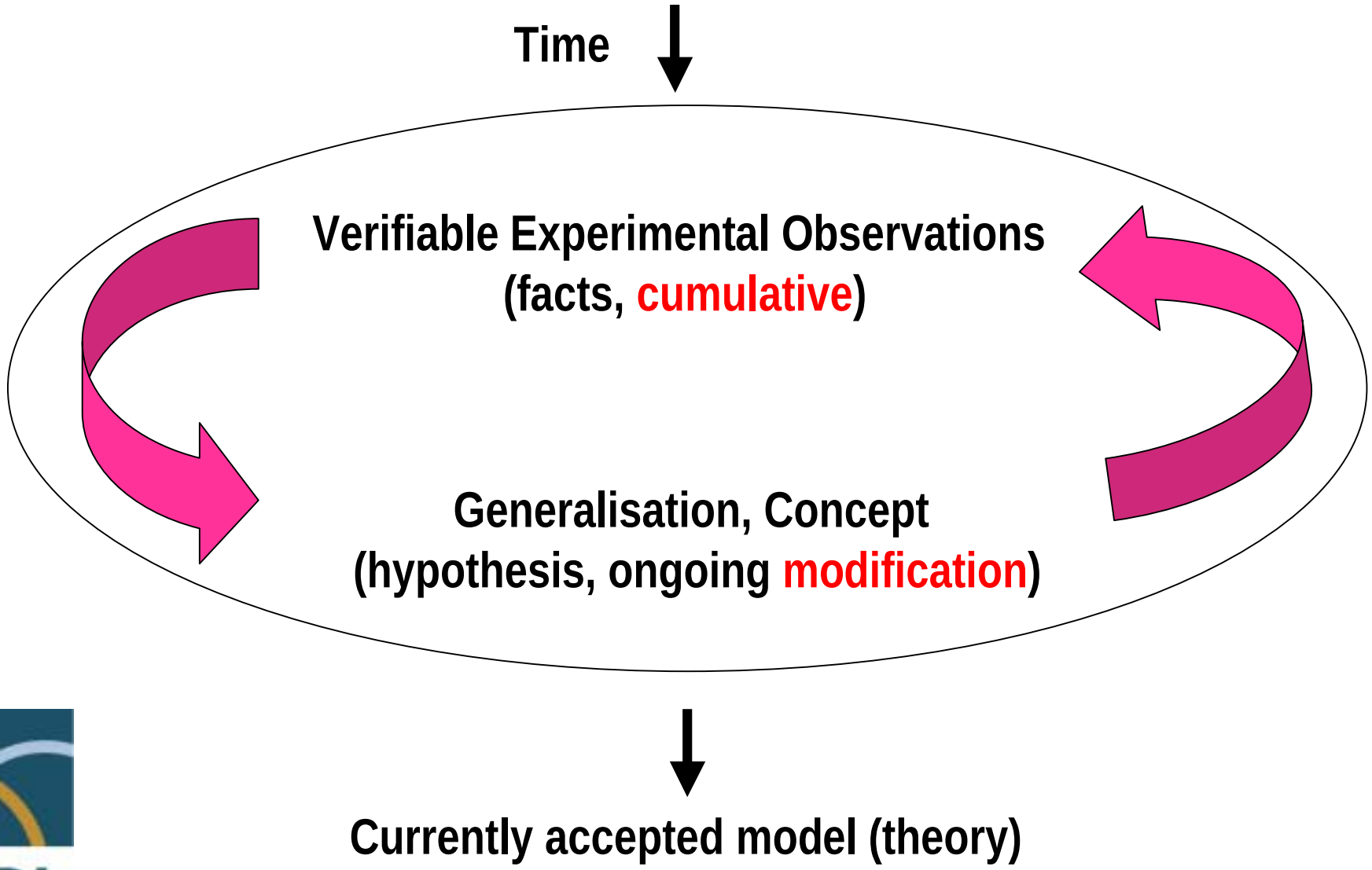


# Improvement of the bioactivity profiling process using literature and patent databases

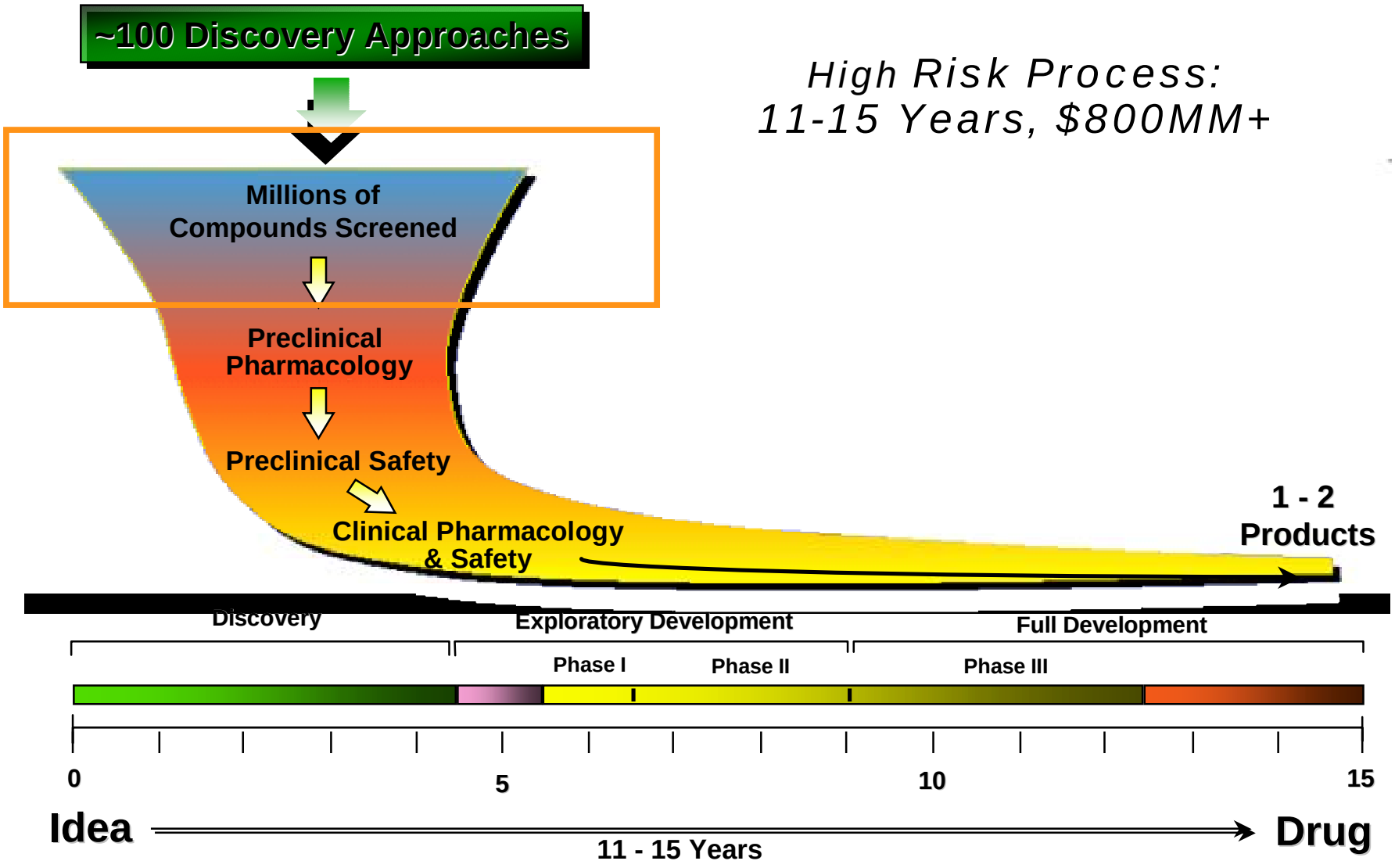
International Chemical Information Conference 2005

Presented by: Prof. Dr. Alexander Lawson
Title: Director Research & Development
Date: October 2005

# The Scientific Method



# Drug Discovery




*High Risk Process:*
*11-15 Years, $800MM+*

Source: Pfizer

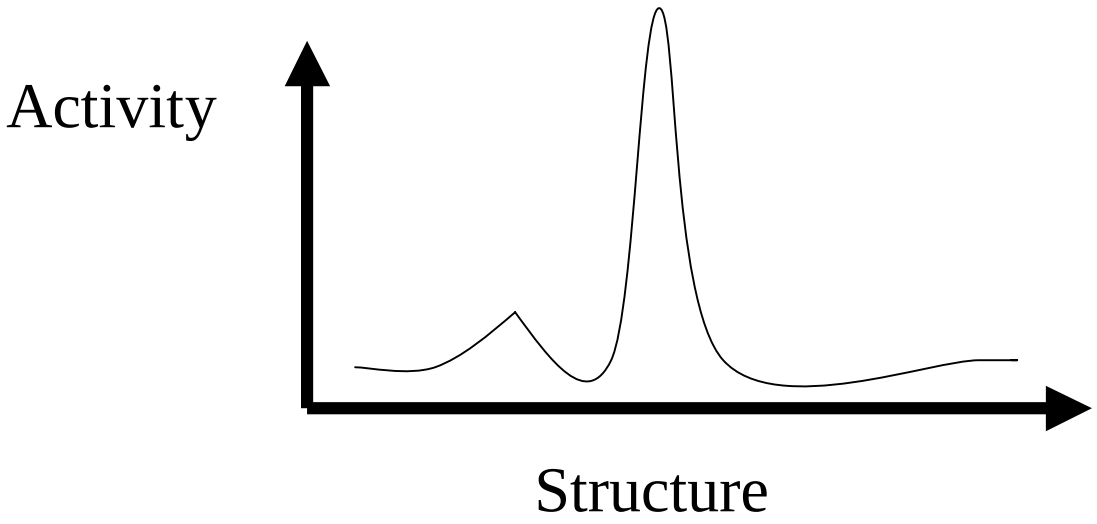
# Drug Discovery Process

- High throughput screening of millions of compounds
  - Demands strong efforts to get or synthesize the test compounds
  - Generates huge amounts of measured data, which have to be analyzed – difficult, error-sensitive, time-consuming, ineffective, cost-intensive

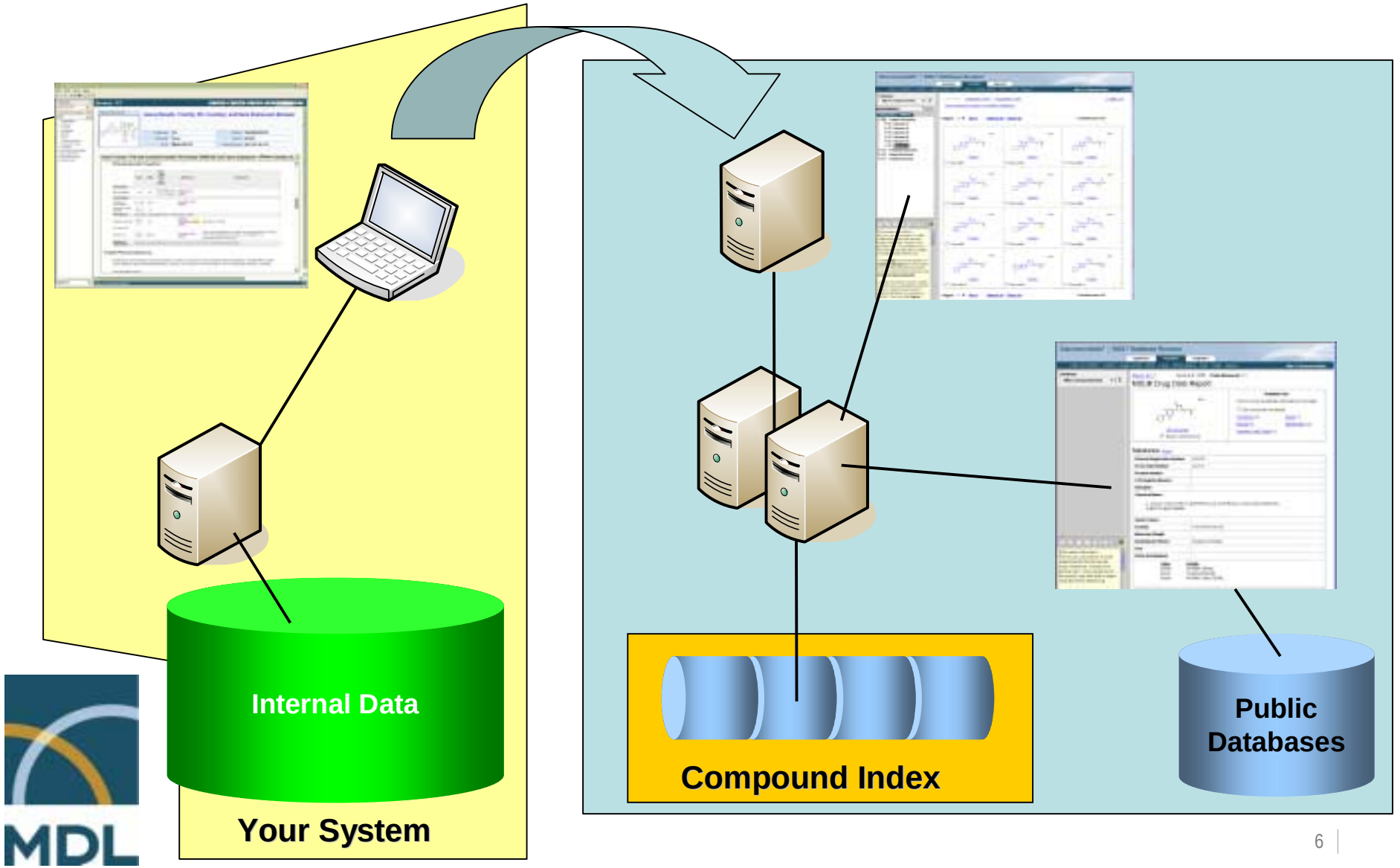
- An early pre-selection of test candidates by analyzing published measured data (internal and external) allows
  - Determination of more focussed libraries
  - Higher quality leads
  - Shorter drug development times
  - Less resources
  - Saving of time and money

# Structure / Activity Relationships

- ➔ Availability of structure and activity data
- ➔ Tabular focus for interpretation
  - (Quantifying structure / structure relationships)
  - (Quantifying activity / activity relationships)



# Combination of internal & external data



# Databases in the Drug Design Process

- **Bibliographic "Abstracting & Indexing" databases**
  - Provide qualitative information
  - Demand browsing of the primary literature to find the data
  - Demand manual data extraction for analysis
    - Time-consuming
    - Expensive

- **Factual databases**
  - Provide a direct access to measured data having a higher value than calculated data (cost-effective)
  - Don't need the time-consuming browsing of primary literature
  - Generate smaller hit sets with more relevant result sets
  - Often allow automatic data export to SAR tables
    - Time-savings
    - Cost-savings

# Databases in the Drug Design Process

Question:

- How can factual databases help in the drug design process?
  - Which commercially available databases provide measured pharmacological, metabolism and toxicological data?
  - Which data in detail do they provide?
  - How can their data be used effectively (e.g. in form of SAR tables) to make the drug design and screening process easier?

# Data used in the drug design process

- **Pharmacological data**
  - Effect / Action / Activity
  - Effective concentrations EC..
  - Inhibition coefficients IC..
  - Protein-ligand binding constants - Ki value
  - Drug-drug-interactions
  - Adverse Effects
  - etc.
- **Metabolism Data**
  - Metabolites, Metabolic pathways
- **Toxicological Data**
  - Toxic effects
  - Lethal doses LD..
- **Other substance data**
  - LogP, pKa

# Measured data in databases

| Database | Owner | Pharmacol. Data (EC, IC, Ki ) | Meta-bolism Data | Tox Data (LD) | pKa, LogP |
|---|---|---|---|---|---|
| **CrossFire Beilstein*** | Beilstein Institute | X | X | X | X |
| **Drug Data Report*** | Prous | X | | X | X |
| **Patent Chemistry Database*** | MDL | X | X | X | X |
| **PHAR** | PJB | X | | | X |
| **Metabolite Database*** | MDL | X | X | | |
| **Toxicity Database*** | MDL | | | X | |
| **xPharm*** | Elsevier | X | X | X | X |

*Note: Databases are available by licence from Elsevier MDL

# CrossFire Beilstein

Huge pool of measured data for the drug design process

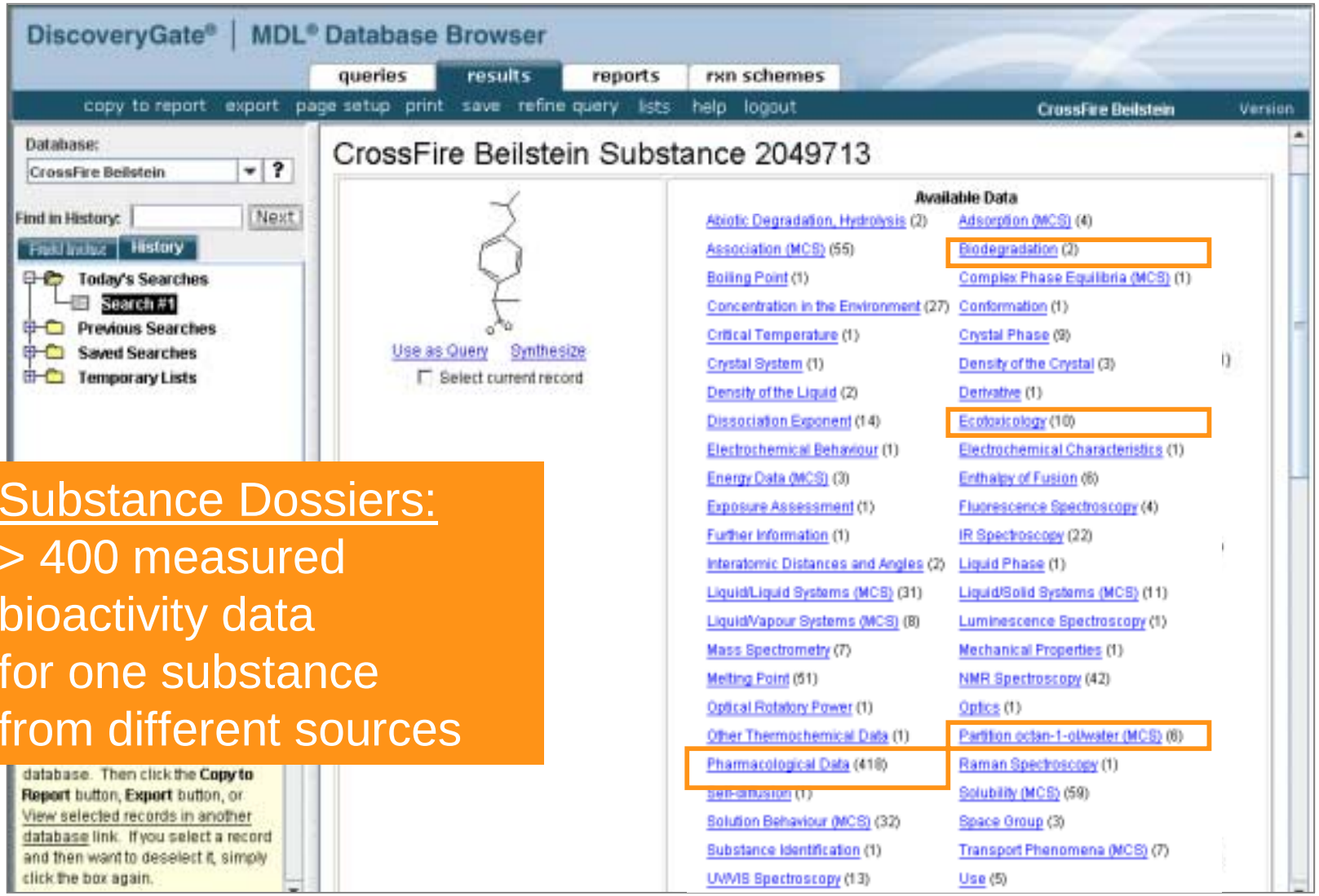
- Pharmacological data
  incl. >26,000 protein-ligand binding constants
- Metabolism Data (metabolites)
- (Eco-)Toxicological Data (broad spectrum of toxicological areas)
- Isolation from natural products

- Substance Dossiers: Substance records accumulate data from multiple sources

- Searchable via DiscoveryGate/DatabaseBrowser or CrossFire Commander

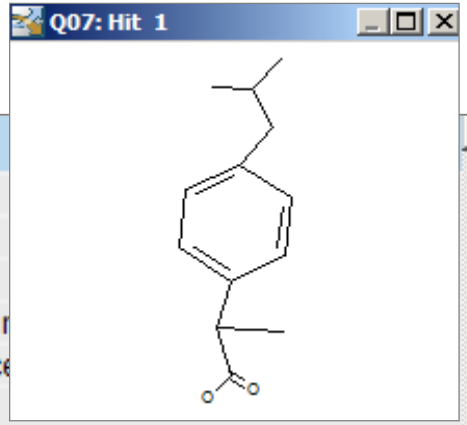
- Export of structures and their measured data to structure-activity-relationship (SAR) tables using CrossFire Commander

# Substance Dossiers in CrossFire Beilstein




**Substance Dossiers:**
> 400 measured bioactivity data for one substance from different sources

# Example
# Ibuprofen in CrossFire Beilstein




| Pharmacological Data 68 of 414 | |
|---|---|
| Effect | anti-inflammatory |
| Species or Test-System | murine B cell hybridoma MH60/BSF-2 |
| Concentration | 100 µmol/l |
| Method | 3 independent tests; 1E4 cells incub. in supplemented RPMI-1640 r |
| | rhIL-6 (1.2 unit/ml, 50 µl) at 37 deg C with 5 percent CO2 for 48 h; ce |
| | density using MTT method |
| Further Details | interleukin-6 (IL-6) growth-dependent cell line; untreated control; IC50 calc. for IL-6 bioactivity; |
| | inhibition percent of bioactivity determined as percent of untreated control; IL-6 bioactivity measured |
| | by proliferation of cells in the presence of rhIL-6 |
| Type | IC50 |
| Value of Type | > 100 µmol/l |
| Results | no inhibitory effect on IL-6 bioactivity |
| Ref. 1 | 6334522, Original Document ; Journal; Kang, Bo-Seong; Chung, Eun-Yong; Yun, Yeo-Pyo; Lee, |
| | Myung Koo; Lee, Yong Rok; Lee, Ki-Sung; Min, Kyung Rak; Kim, Youngsoo; BPBLEO; Biol. Pharm. |
| | Bull.; EN; 24; 6; 2001; 701 - 703. |

# MDL® Patent Chemistry Database

- Chemical Reactions and substance data from organic chemistry and life sciences patent publications (WO, EP, US since 1976/78)
  - Many measured bioactivity data* from research institutes and universities; especially interesting:

    > 16.000 Ki values (protein-ligand-binding constants)

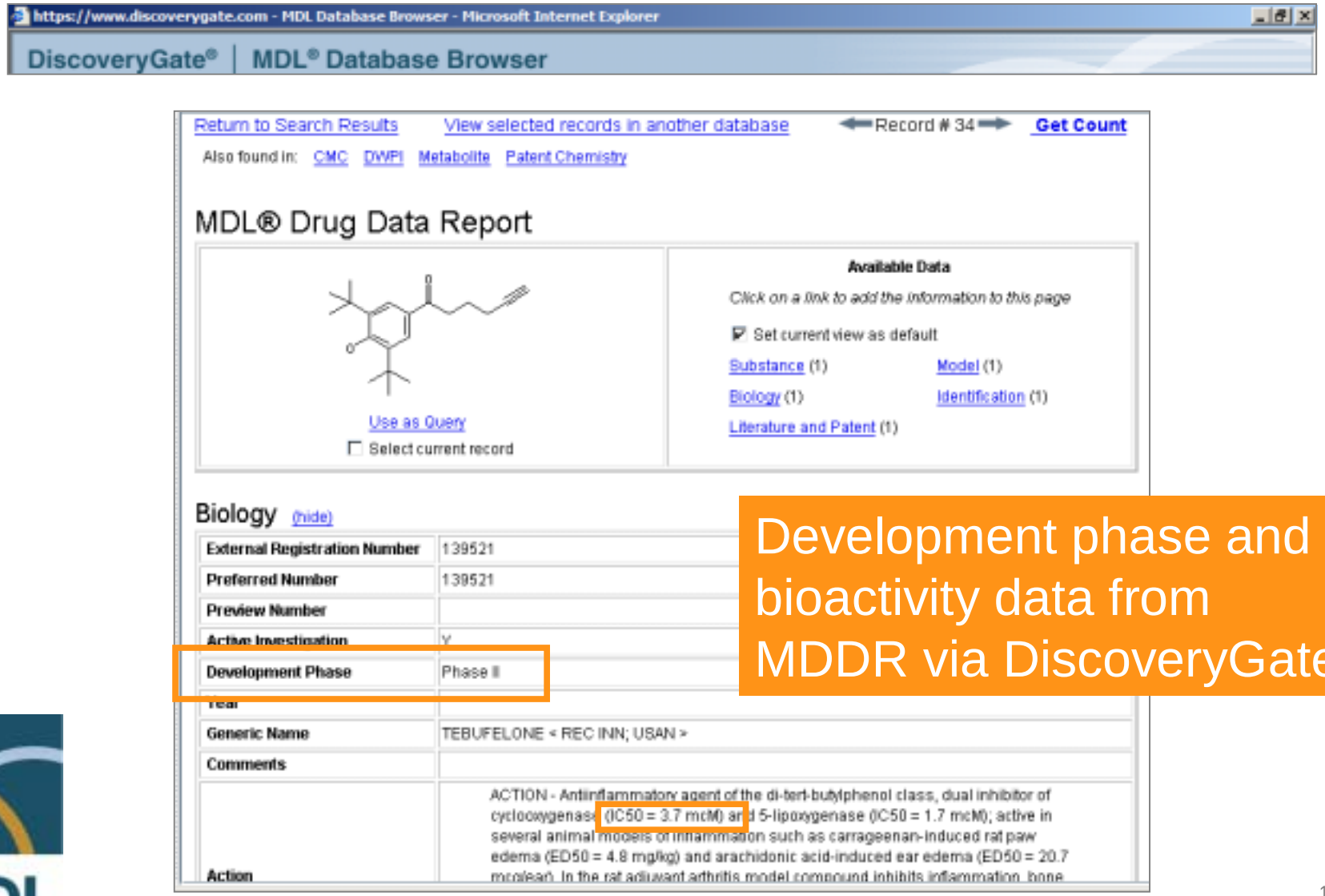
  - Pharma, agro, cosmetics
  - Metabolism* data (metabolites)
- Data available earlier than in journal publications
- Searchable via DiscoveryGate or CrossFire Commander 7.0 SP1 or SP2
- Export to SAR tables using CrossFire Commander

* from Dec. 2003 onwards

# MDL® Drug Data Report (MDDR)

- Current bioactivity data for newly launched and developmental drugs
  - Generic names
  - Action
    - Textual description incl. measured data (e.g. EC-, IC-, LD/LC-values)
  - Classification of pharmacological effects
  - Development Phase, e.g. pre-clinical, launched
  - Trademark
  - Developing company, organization
- Searchable 3D models
- Data owned by Prous Science

# Data from development drugs




Development phase and bioactivity data from MDDR via DiscoveryGate

# PHAR (Pharmaprojects)

- Drug marketing, R&D, and licensing information about pharmaceutical products under development in the world's major markets since 1980.
  - Development status
  - Pharmacokinetics
  - Pharmacological Activity Code
  - Therapeutic Classification Codes
  - logP
  - etc.

# Pharmacokinetic data for developmental products

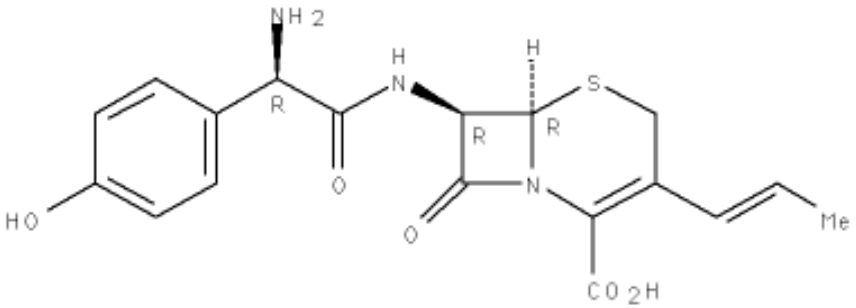
PHK

| Model | Parameter | Values | Units |
|---|---|---|---|
| Human (capsules) | t1/2 | 1.3 | hr |
| Human (capsules) | Vd | 0.23 | l/kg |
| Human (capsules) | Cl | 3 | ml/min/kg |
| Human (500 mg capsules) | Tmax | 1.5 | hr |
| Human (500 mg capsules) | Cmax | 10.5 | microg/ml |

LN

| Therapy (CC) | Pharmacology (PHCD) | Status (DSTC) |
|---|---|---|
| J1D1 | SY-CW-AN | L |

LCDAT 20030902: CH : Country statuses updated




Taken from PHAR - access via STN

# MDL® Metabolite Database

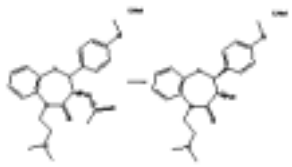
- **Transformation (Scheme)**
  - Parent structure
  - Substrate structure
  - Metabolite structure
  - Activity
  - Species
  - Enzymes (Name, Classification)
  - Isoenzymes
  - Excretion

- **Metabolic scheme diagram**
  (all metabolic transformations from a common parent compound in one scheme)

# Metabolic Pathways and Pharmacokinetic information

## MDL® Metabolite Database

MDL® Metabolite Database



Use as Query

Select current record

**Available Data**

*Click on a link to add the information to this page*

Set current view as default

View results as **transformations:**

Transformation (1)
Reference (66)
Species (66)
Enzyme (66)

View results as **molecules:**

Parent (1)
Species (81)
Substrate (1)
Metabolite (1)

Molecule Results -- Parent (hide)

| | CAS Number |
|---|---|
| | 42399-41-7 |
| | 33286- |

| Number of Species | Species | Route | Excretion | Pharmacokinetics | First Pass Metabolism |
|---|---|---|---|---|---|
| 1 | Rabbit (Diabetic) | Oral | Plasma | Y | |
| 2 | Rabbit | Oral | Plasma | Y | |
| 3 | Rabbit (Diabetic) | Oral | Urine | N | |
| 4 | Rabbit | Oral | Urine | N | |

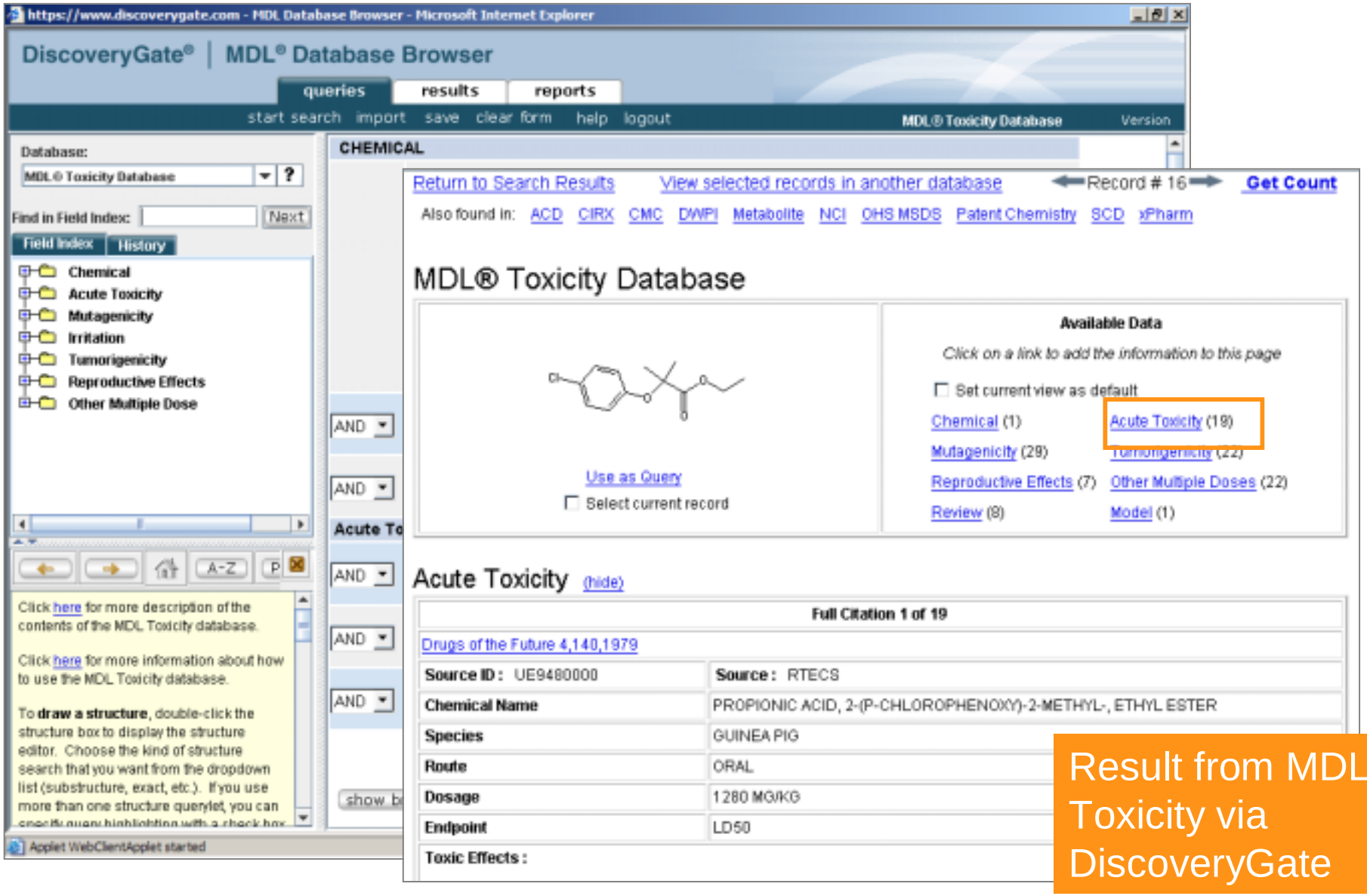
| Number of Species | Total Activity | Parent Activity | Time (hrs) |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 3 | | 2.26 | 0-24 |
| 4 | | 2.12 | 0-24 |

| Species Full Citation 15 of 81 |
|---|
| Zhao, X.-J.; Jones, D. R.; Wang, Y.-H.; Grimm, S. W.; Hall, S. D.; Xenobiotica [XENOBH] 2002, 32 (10), 863. |

# MDL® Toxicity Database

- Complete contents of the Registry of Toxic Effects of Chemical Substances (RTECS) database from the US EPA and the CCRIS database from the National Cancer Institute
  - Acute Toxicity
  - Irritation
  - Mutagenicity
  - Tumorigenicity
  - Reproductive Effects
- Data from GeneTox
- Searchable 3D structures

# Lethal doses from multiple species



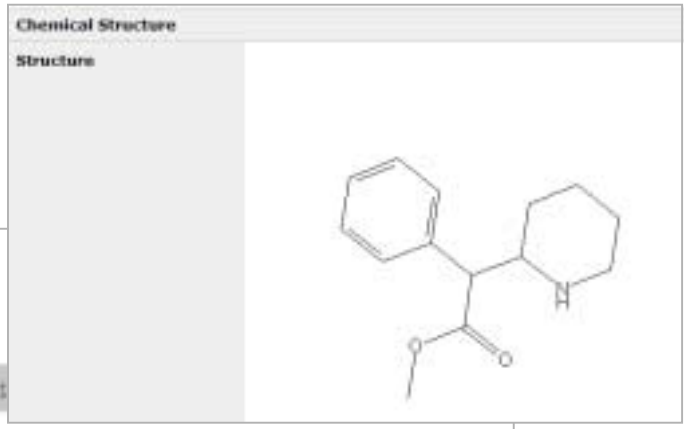

Result from MDL® Toxicity via DiscoveryGate

# xPharm®

- A comprehensive set of records linking four pharmacological topics:
  - Agents (compounds)
  - Targets
  - Disorders
  - Principles

- Available Data
  - Pharmacokinetics
  - Pre-clinical Data
  - Drug-Drug-Interactions
  - Adverse Effects
  - Absorption and Distribution
  - Metabolism
  - Toxicity

# xPharm®




DiscoveryGate | xPharm®

Home | Contents | Search | Agents | Targets | Disorders | Principles

Quick Search | within All xPharm | Go | Search Tips | About | Logout

## Methylphenidate

Annette Fleckenstein

Click here to cite this

**Introduction**

Methylphenidat
as a schedule II

**Nomenclature**

| |
|---|
| Name of the Clinical Form |
| Synonyms |
| Chemical Names |
| CAS Number |

**Ionization Constant**

| | Value | Salt | Conditions | Reference | Comments |
|---|---|---|---|---|---|
| pKa | 8.9 | | hydrochloride salt | Windholz and Budavari (1983) | |

**Human Pharmacokinetics**

**Pharmacokinetic Properties**

| | Value | Units | Pr a Rou Ad | | |
|---|---|---|---|---|---|
| **Absorption** | | | | | |
| Bioavailability | approximately 30 | % | p.o. | | |
| **Distribution** | | | | | |
| Volume of Distribution | approximately 20 | l/kg | p.o. | | |
| Plasma Protein Binding | approximately 15 | % | p.o. | | |
| **Metabolism** | | | | | |
| Plasma Half-Life | 2.4 | hrs | p.o. | Wargin et al (1983) | as assessed in children |
| Bio Half-Life | 2.1 | hrs | p.o. | Wargin et al (1983) | as assessed in adults |
| Clearance | 10.2 | l/hrs/kg | p.o. | Wargin et al (1983) | |
| **Routes of Elimination** | urine (80–90%): major metabolite is ritalinic acid; less than 1% of unchanged methylphenidate is excreted in the urine; feces (1–3%). The predominant route of metabolism is via de-esterification Faraj et al (1974) | | | | |

| | | | | |
|---|---|---|---|---|
| **Mouse** | | | | |
| LD50 | 150 | mg/kg | | s.c. |
| LD50 | 40 | mg/kg | | i.v. |
| **Rabbit** | | | | |
| LD50 | 900 | mg/kg | | p.o. |
| LD50 | 170 | mg/kg | | s.c. |
| LD50 | 30 | mg/kg | | i.v. |

# Databases in the Drug Design Process

Question:

- How can factual databases help in the drug design process?
  - Which commercially available databases provide measured pharmacological, metabolism and toxicological data?
  - Which data in detail do they provide?
  - How can their data be used effectively (e.g. in form of SAR tables) to make the drug design and screening process easier?

# Structure-Activity-Relationship Analysis using SAR tables

SAR tables play an important role in the pre-selection of potential winner candidates, but only factual databases allow an automatic creation of SAR tables from the data source

Analysis Process:

- Exporting structures and their activity into a SAR table
- Sorting SAR table by Value of effect, e.g. increasing EC50 value
- Looking at highest effect values
- Analysis, which scaffolds or substituents might be mainly responsible for high activity values

# Availability of SAR table functionality

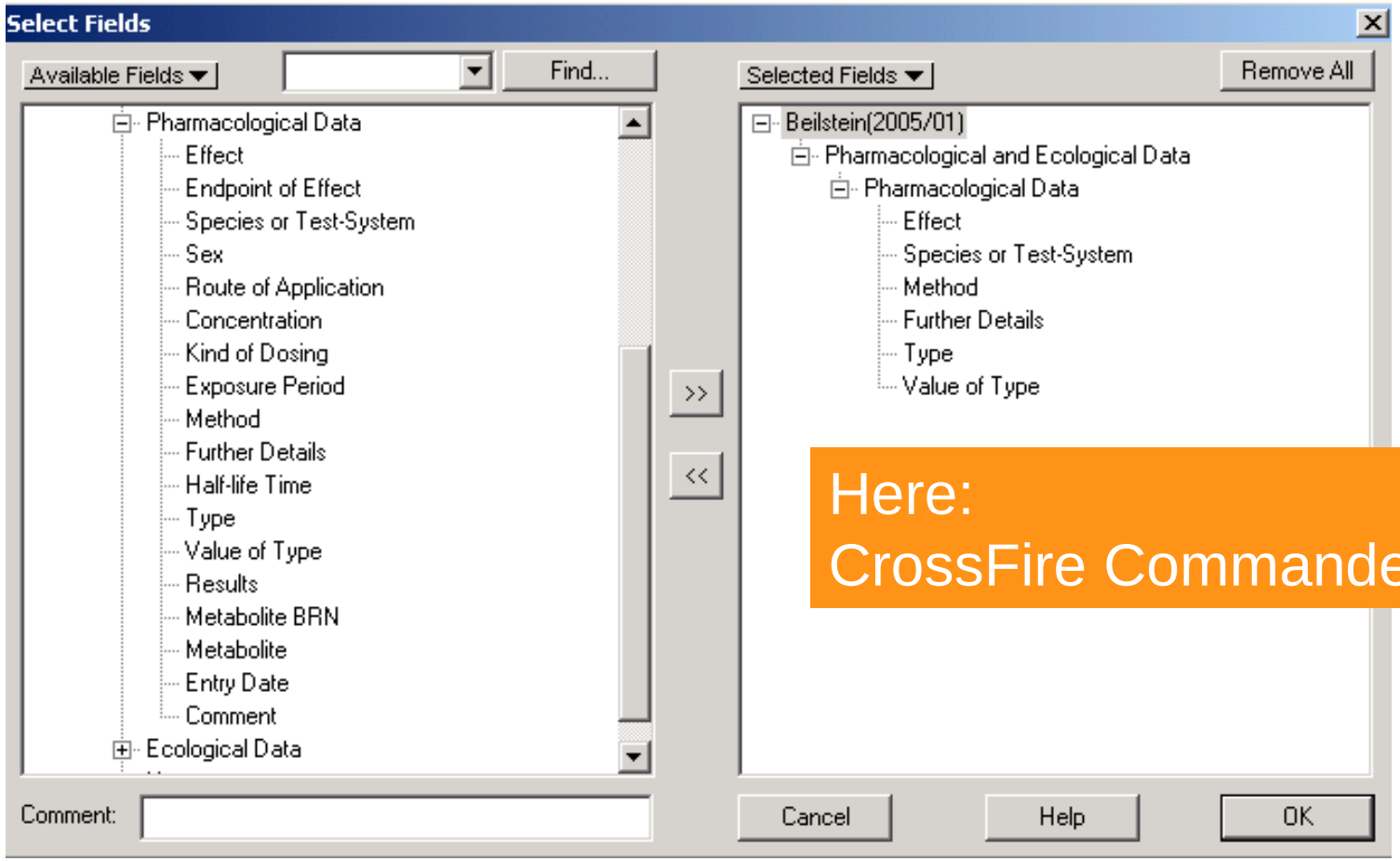
- CrossFire Commander
  - Export to SAR tables available for CrossFire Beilstein and MDL® Patent Chemistry Database
  - Many predefined export formats available (xls, doc, html, xml, sdf, tab-delimited)

- DiscoveryGate/DatabaseBrowser
  - Selected MDL databases
  - Export to SDfiles (sdf), which can be imported to ISIS FOR EXCEL to create SAR tables

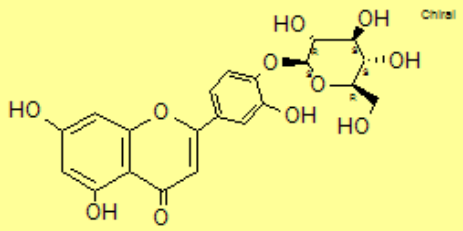
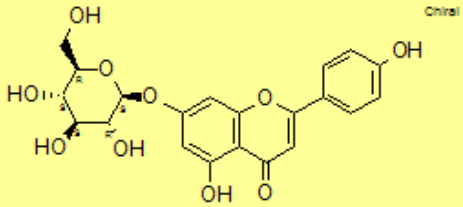
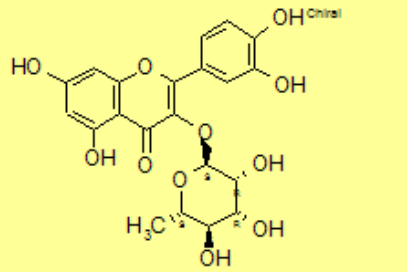
- High flexibility to create own export formats by selection of data fields
- High flexibility in modifying table layout using EXCEL, ISIS FOR EXCEL

# Export Bioactivity data to SAR table

## Define your own SAR-Tables



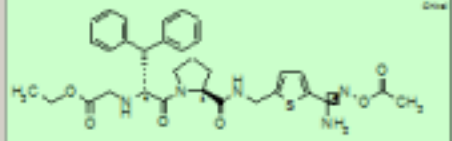
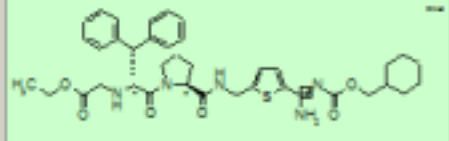
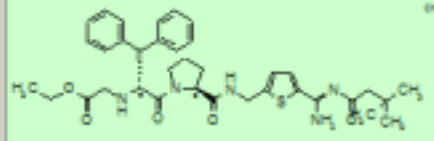
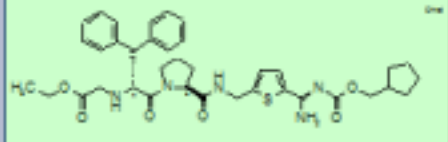
# IC50 values of anti-inflammatory agents

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Structure | BRN | Effect | Species | Type | Value | Citation |
| 2 |  | 65285 | anti-inflammatory | Y16 cells | IC50 | 3.7 µmol/l | 6215841; 2000/06/30; Journal; Park, Keun Yeong; Lee, Seung-Ho; Min, Bo-Kyung; Lee, Kyong-Soon; Choi, Jae-Sue; Chung, See Ryun; Min, Kyung Rak; Kim, Youngsoo; PLMEAA; Planta Med.; EN; 65; 5; 1999; 457 - 459; |
| 3 |  | 65669 | anti-inflammatory | Y16 cells | IC50 | 14.2 µmol/l | 6215841; 2000/06/30; Journal; Park, Keun Yeong; Lee, Seung-Ho; Min, Bo-Kyung; Lee, Kyong-Soon; Choi, Jae-Sue; Chung, See Ryun; Min, Kyung Rak; Kim, Youngsoo; PLMEAA; Planta Med.; EN; 65; 5; 1999; 457 - 459; |
| 4 |  | 68135 | anti-inflammatory | Y16 cells | IC50 | > 50 µmol/l | 6215841; 2000/06/30; Journal; Park, Keun Yeong; Lee, Seung-Ho; Min, Bo-Kyung; Lee, Kyong-Soon; Choi, Jae-Sue; Chung, See Ryun; Min, Kyung Rak; Kim, Youngsoo; PLMEAA; Planta Med.; EN; 65; 5; 1999; 457 - 459; |
| |  | | | | | | 6215841; 2000/06/30; Journal; |

Sheet1 | Sheet2 | Sheet3

Export from CrossFire Beilstein to MS EXCEL via CrossFire Commander

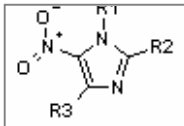
# Protein-ligand-binding constants

| | A | B | P | Q | R | AA | AG | AP |
|---|---|---|---|---|---|---|---|---|
| 1 | **Structure** | **Compound Regno** | **Action** | **Para meter** | **Value µmol/l** | **Protein** | **Lcation in Patent** | **Citation** |
| 2 |  | 4170422 | trypsin; inhibition of | Ki | 21.333 | trypsin | Page 61-62 | 248933; Patent; 2005/01/29; 2005/01/31; LG LIFE SCIENCES LTD.; WO2004/2985; WO; A1; KR; 2004/01/08; 2004; WO2003-KR1166; 2003/06/13; 2003; |
| 3 |  | 4170433 | trypsin; inhibition of | Ki | 21.019 | trypsin | Page 61-62 | 248933; Patent; 2005/01/29; 2005/01/31; LG LIFE SCIENCES LTD.; WO2004/2985; WO; A1; KR; 2004/01/08; 2004; WO2003-KR1166; 2003/06/13; 2003; |
| 4 |  | 4170450 | trypsin; inhibition of | Ki | 1 | trypsin | Page 61-62 | 248933; Patent; 2005/01/29; 2005/01/31; LG LIFE SCIENCES LTD.; WO2004/2985; WO; A1; KR; 2004/01/08; 2004; WO2003-KR1166; 2003/06/13; 2003; |
| 5 |  | 4170448 | trypsin; inhibition of | Ki | 1 | tryp | | |

Sheet1 | Sheet2 | Sheet3

MDL® Patent Chemistry Database: > 16,000 Ki values

# Analysis of SAR tables




| ID | R1 | R2 | R3 | Bioactivty |
|---|---|---|---|---|
| 000101 | -CH3 | -CH2OCONH2 | -CH3 | 5.33 |
| 000102 | -CH3 | -p-F-Phenyl | -CH3 | 4.371 |
| 000103 | -CH2CH2OH | -CH3 | -CH3 | 3.935 |
| 000104 | -CH3 | -CH3 | -CH3 | 3.685 |
| 000105 | -CH3 | -CH3 | -CH2CH2OH | 4.966 |
| 000106 | -CH3 | -CH2OH | -CH2CH2OH | 4.002 |
| 000107 | -CH3 | -p-F-Phenyl | -CH2CH2OH | 4.537 |
| 000108 | -CH3 | -CH3 | -(CH2)2OCO(CH3)3 | 4.388 |
| 000109 | -CH3 | -CH3 | -(CH2)2OCOCH3 | 4.909 |
| 000110 | -CH3 | -CH3 | -Br | 5.467 |
| 000111 | -CH3 | -CH3 | -F | 4.04 |
| 000112 | -CH3 | -H | -F | 4.605 |
| 000113 | -(CH2)2OCO-3,4,5-Trimethoxy-Phenyl | -CH3 | -H | 4.833 |
| 000114 | -(CH2)2OCO-3,4,5-Trimethoxy-Phenyl | -p-F-Phenyl | -H | 5.132 |
| 000115 | -(CH2)2OCO-3,4,5 Trimethoxy-Phenyl | -CH3 | -H | 5.432 |
| 000116 | -CH3 | -CH2OCO-3,4,5-Triacetoxy-Phenyl | -H | 5.267 |
| 000117 | -CH3 | -CH3 | -H | 5.672 |
| 000118 | -CH3 | -CH2CH2OH | -H | |
| 000119 | -CH3 | -CH2OH | -H | 4.992 |
| 000120 | -CH3 | -CH2OCONH2 | -H | 5.699 |
| 000121 | -CH3 | -p-F-Phenyl | -H | 5.946 |
| 000122 | -CH3 | -CH2CH2OH | -CH3 | |
| 000123 | -CH2CH2OH | -CH3 | -H | 5.631 |

# Summary

- Factual databases offer a huge amount of measured data useful in the drug design process
- Data can be merged and compared with customer internal data via SDfiles
- Analysis of SAR tables from internal and external data
  - allows a faster and more effective recognition of new test candidates (focus on winners)
  - increases the quality of leads entering the drug-development process

**Usage of factual databases and SAR tables reduces drug development times and costs**