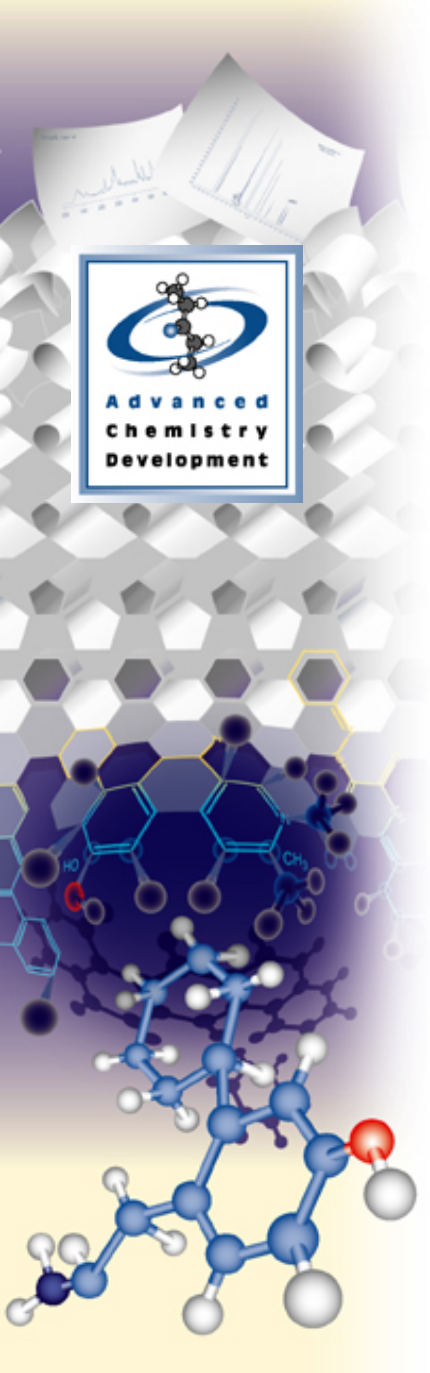



# ACD/ChromGenius:

# Automated Method Selection for LCMS

RTP Seminar
June 3, 2004

**Advanced Chemistry Development, Inc. (ACD/Labs)**

Visionary Software for Scientists

# Introduction

ChromGenius

- Who is the ideal candidate for automated method selection?
- What is the usual approach?
- What are the limitations?
- How does ChromGenius work?
- What's new in 8.0?
- Wrapping up

# The ideal candidate:

ChromGenius

- Hundreds to thousands of samples
- No time for method development
- FAST run times are necessary
- Always gradients
- A few standard, or generic, methods
- Restricted column choices
- Almost always LCMS

# What is the typical approach in HT Labs?

- Generic methods – non-targeted approach
- Fast LCMS gradients
- A few methods are used to cover anticipated structural diversity
  - Inject samples under each method and observe the results
  - Or perhaps examine those results and re-inject the samples that didn't work.

ChromGenius

# Challenges associated with Generic Methods

ChromGenius

- Problem compounds:
  - Instrument downtime!
- Some compounds elute too early:
  - MW verification, but no purity estimation
  - Minimal resolution, but bad purity estimation
- Scale-up from an “okay” method can be tricky at best

# So Let's Predict the Best Method to Use!

- Design several complementary methods
- Use structure(s) available to choose
- Target:
  - Reasonable k'
  - Resolution from expected impurities

ChromGenius

# The Prediction of Chromatography: A Great Challenge

ChromGenius

- Complex retention mechanisms
- Fast gradient conditions
- pH issues when organic solvents are present

So how do you go about predicting chromatography under these conditions?

# LC Simulator and $t_R$ Prediction.

- LC Simulator uses predicted physicochemical parameters to model retention behavior in separations.
  - LogP
  - pKa
  - LogD
  - MW, MR, MV

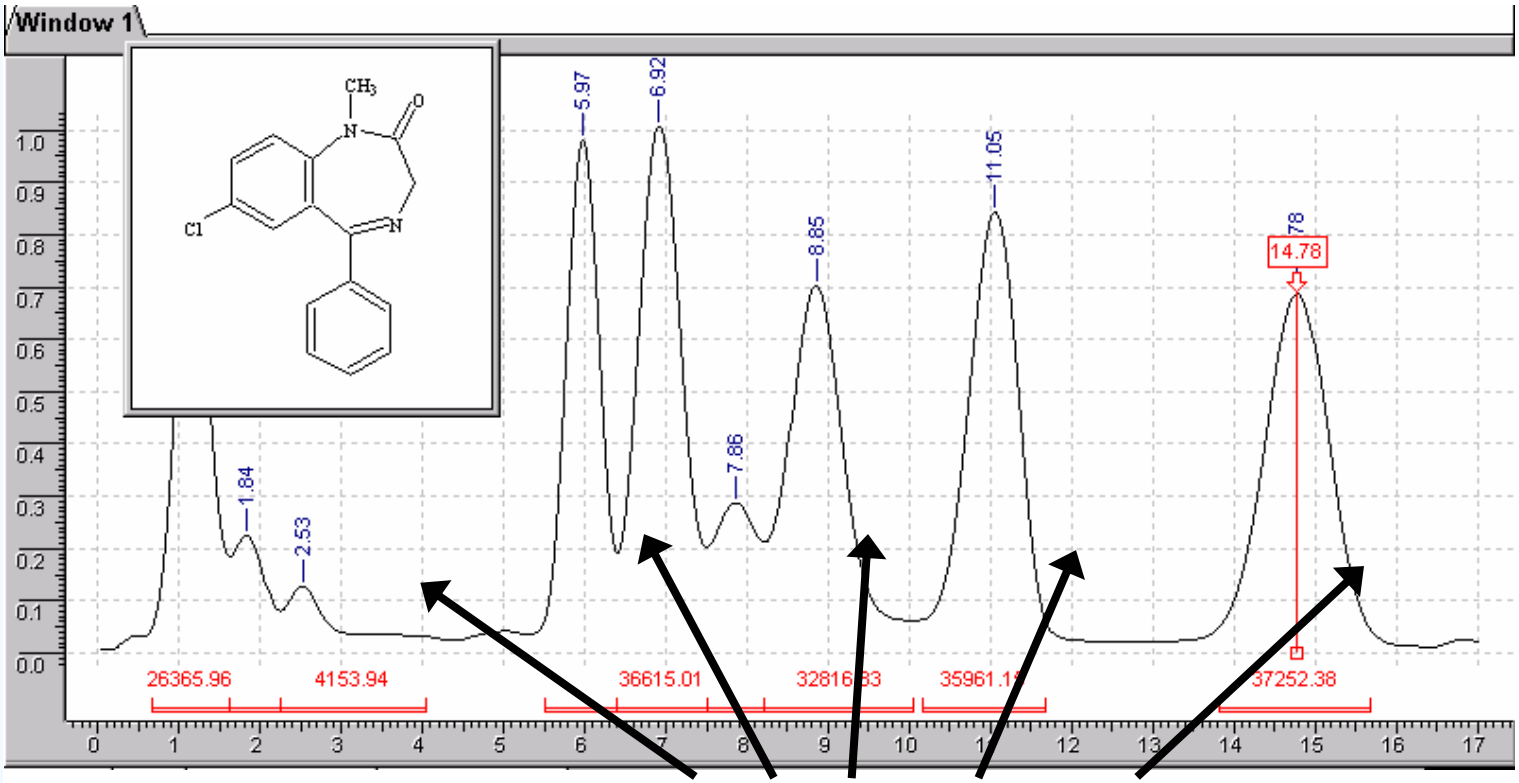
LC simulator uses actual chromatographic data to train the prediction.

ChromGenius

# LC Simulator Backbone

## Prediction of $t_R$s

Given an experimental set:




Where will my new compound elute?

ChromGenius

# What are the limitations of this approach?

ChromGenius

- Similarity of training set to new compounds
- Number of compounds in training set
- pH issues
- Gradients are not explicitly modeled

# How can ChromGenius help?

- ChromGenius chooses the most relevant training set from *multiple* chromatograms.
- More compounds:
  - Better characterization of separation
  - More PhysChem terms
- and more relevant compounds:
  - Similar retention mechanism
  - Inherent gradient compensation

ChromGenius

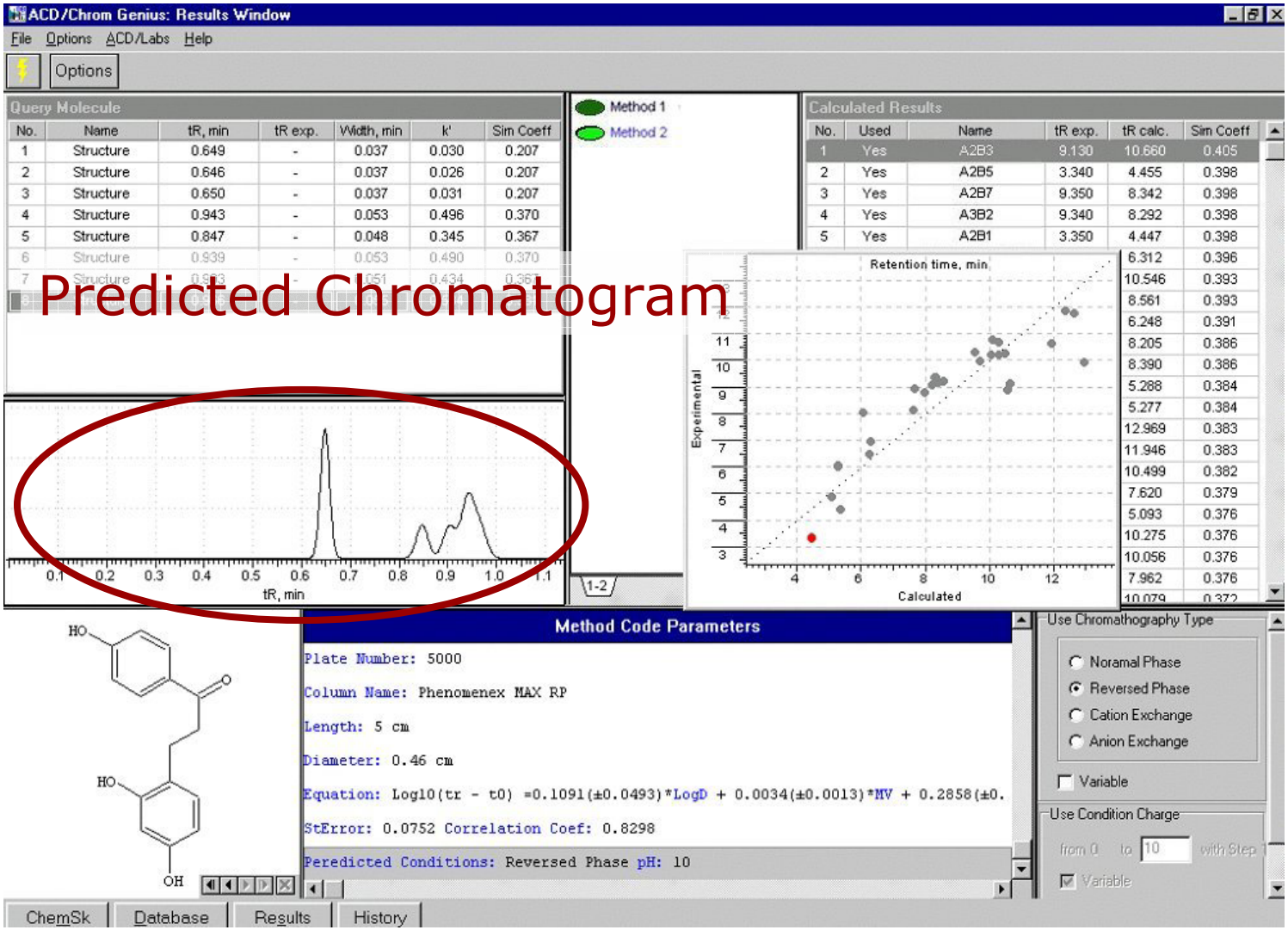
# Structure Similarity and Accuracy

- Accuracy of prediction goes up with similarity to training set
- "Relevance" of the training compounds has a bearing on the accuracy:
  - Processes that are NOT modeled are "constants"
  - Similar compounds elute at similar elution times – inherent gradient correction

ChromGenius

# Generic Methods and ChromGenius
# **Under the surface**

ChromGenius

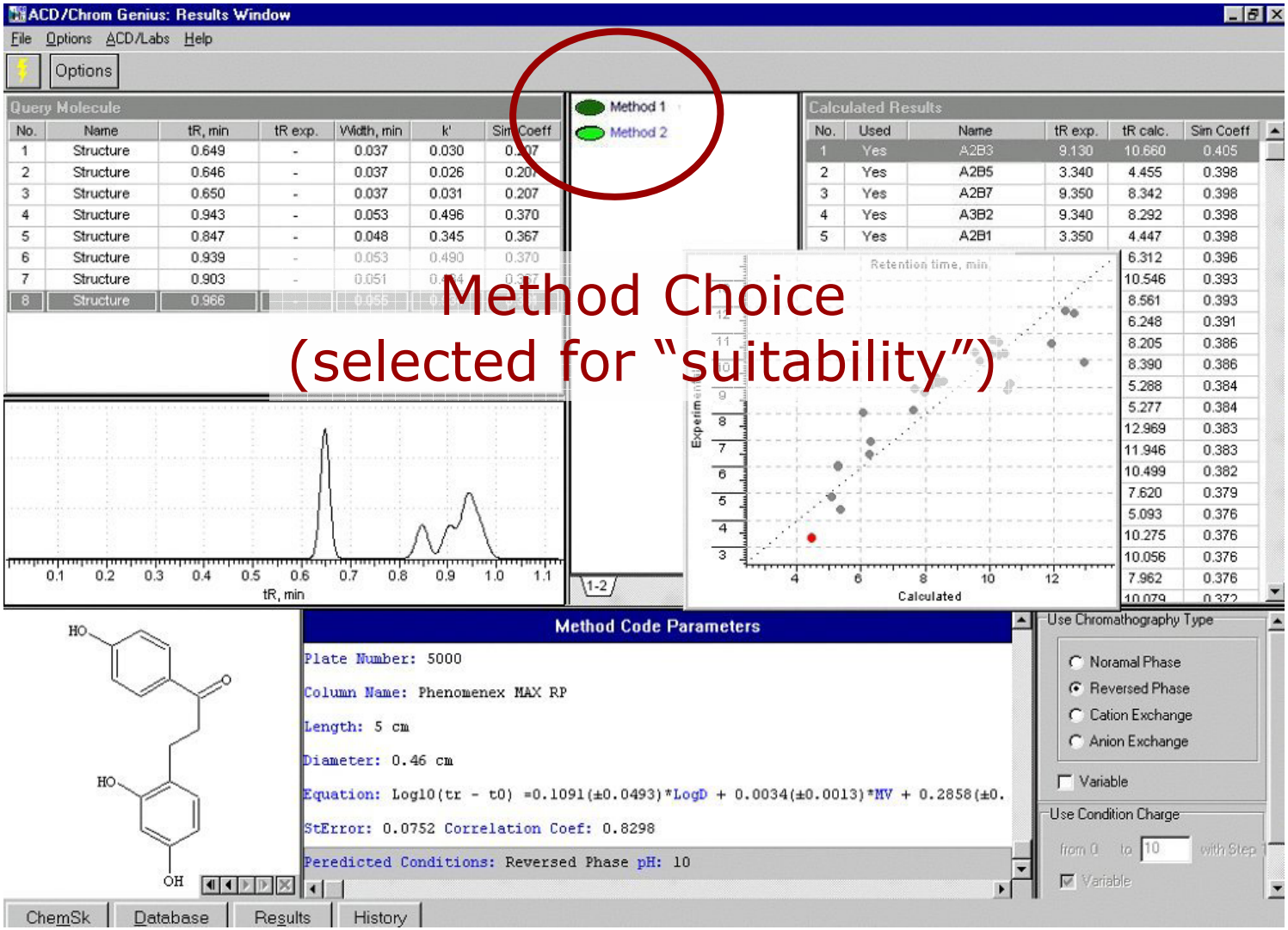



**Visionary Software for Scientists**

**Spectroscopy • Chromatography • PhysChem • Chemical Naming • Drawing • Databasing • Enterprise Solutions**

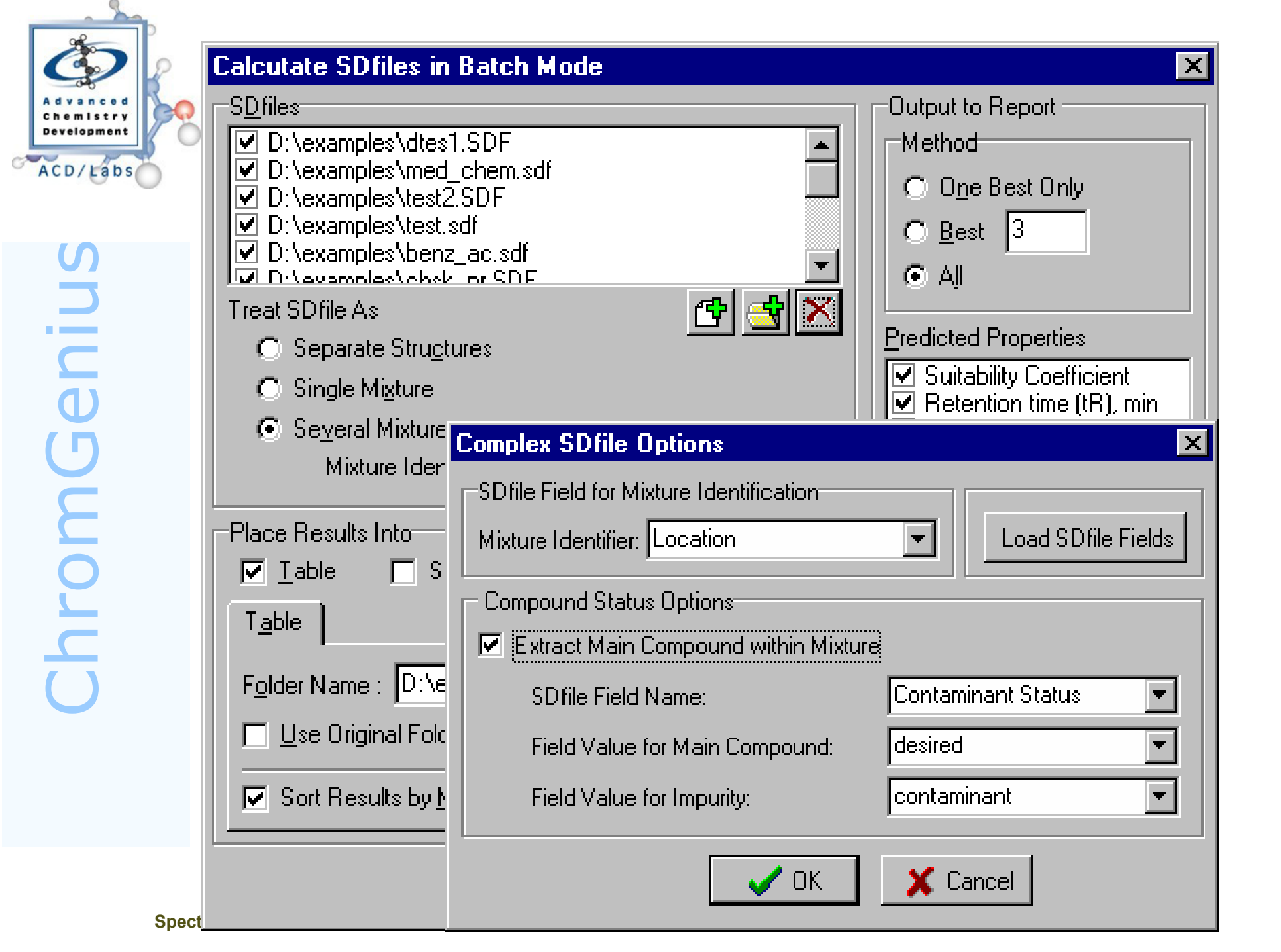
# Selection of Methods

ChromGenius




**Visionary Software for Scientists**

**Spectroscopy • Chromatography • PhysChem • Chemical Naming • Drawing • Databasing • Enterprise Solutions**


Advanced Chemistry Development
ACD/Labs
ChromGenius
Calculate SDfiles in Batch Mode
SDfiles
D:\examples\dtes1.SDF
D:\examples\med_chem.sdf
D:\examples\test2.SDF
D:\examples\test.sdf
D:\examples\benz_ac.sdf
Treat SDfile As
Separate Structures
Single Mixture
Several Mixture
Mixture Ider
Output to Report
Method
One Best Only
Best
3
All
Predicted Properties
Suitability Coefficient
Retention time (tR), min
Place Results Into
Table
Table
Folder Name :
Use Original Fold
Sort Results by
Complex SDfile Options
SDfile Field for Mixture Identification
Mixture Identifier:
Location
Load SDfile Fields
Compound Status Options
Extract Main Compound within Mixture
SDfile Field Name:
Contaminant Status
Field Value for Main Compound:
desired
Field Value for Impurity:
contaminant
OK
Cancel

# The Basis of Method Selection

ChromGenius

- Minimum k’
- Maximum retention time
- Resolution from expected impurities
- Favored methods versus unfavored

# An Early Application
# **Collaboration with Specs**

ChromGenius

- LCMS structure verification and purity measurement
- One high-throughput method
- 100,000 diverse drug-like and "building block" compounds per year
- ONE chromatographic method
- 20% of compounds elute too soon for purity measurement

The challenge is to locate the 'fast eluters' prior to the run so that a different method can be applied.

# Structure Similarity and Accuracy – 2006 compounds

## Specs Method 1 Experimental Conditions:

ChromGenius

- Waters 2690 Separation Module
- Waters Symmetry® C18 (2.1 x 50 mm), 3.5 µm, 100 Å
- Column temperature: 30°C
- Flow: 0.7 – 1.0 ml/min.
- Mobile phase: gradient elution with water/acetonitrile/5% formic acid in water
  - start with the ratio 89:10:1 going to 0:99:1 in 4 minutes
  - 2 minutes isocratic with 0:99:1
  - 1 minute stabilization at initial conditions (89:10:1)
  - After 5 min the flow is increased to 1.0 ml/min (!)
  - Total run time: 7 minutes

# Structure Similarity and Accuracy – 2006 compounds

## Specs Method 1 Experimental Conditions:

ChromGenius

- Waters 2690 Separation Module
- Waters Symmetry® C18 (2.1 x 50 mm), 3.5 µm, 100 Å
- Column temperature: 30°C
- **Flow: 0.7 – 1.0 ml/min.**
- Mobile phase: gradient elution with water/acetonitrile/5% formic acid in water
  - start with the ratio 89:10:1 going to 0:99:1 in **4 minutes**
  - 2 minutes isocratic with 0:99:1
  - 1 minute stabilization at initial conditions (89:10:1)
  - After 5 min the flow is increased to 1.0 ml/min (!)
  - Total run time: 7 minutes

# Structure Similarity and Accuracy – 2006 compounds

ChromGenius

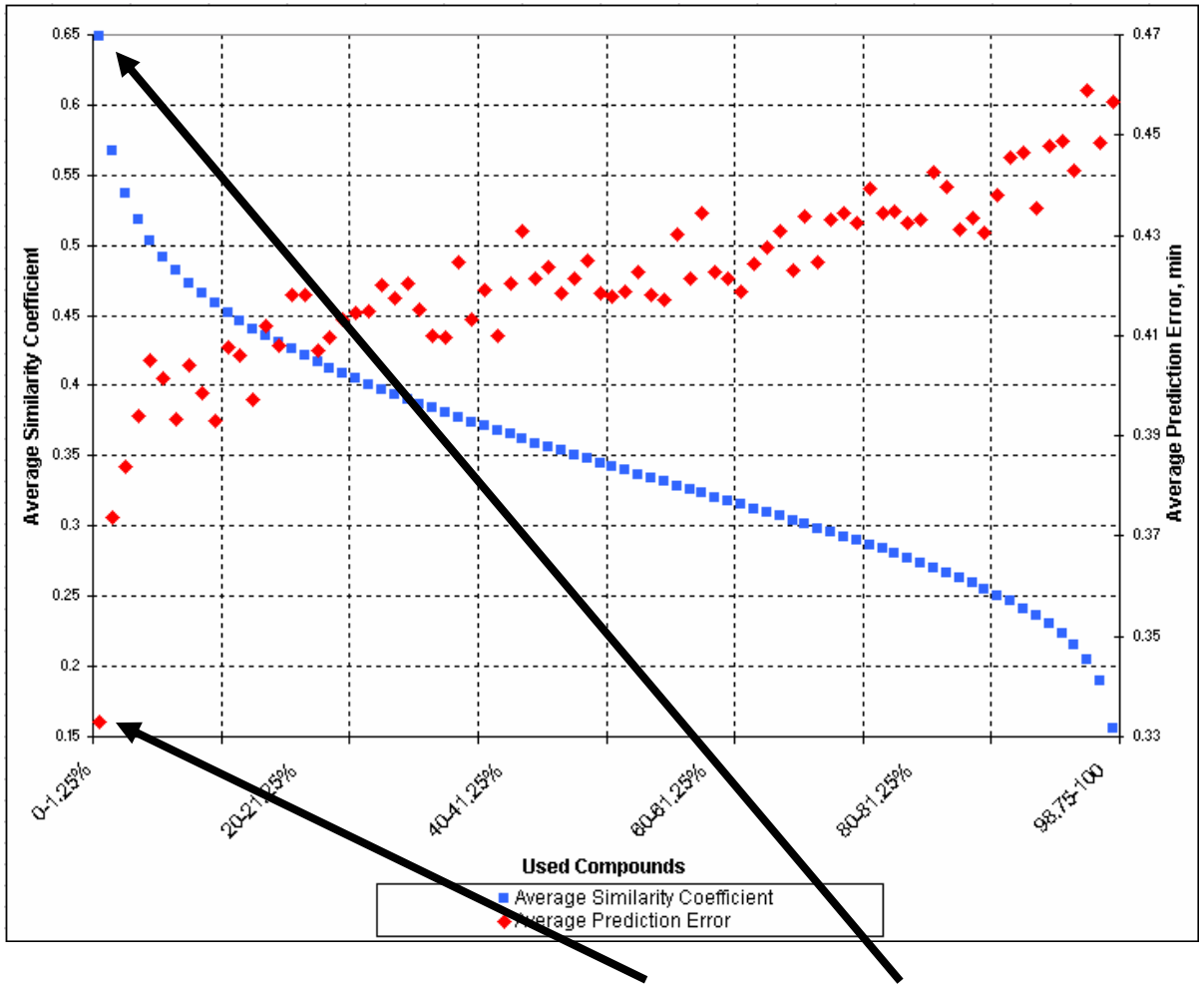
## ChromGenius Calculation Settings

- Dice Coefficient Similarity Search
- 25 structure training sets
- pH = 2.88

# Structure Similarity and Accuracy – 2006 compounds

Each red point is the average of 2006 predictions

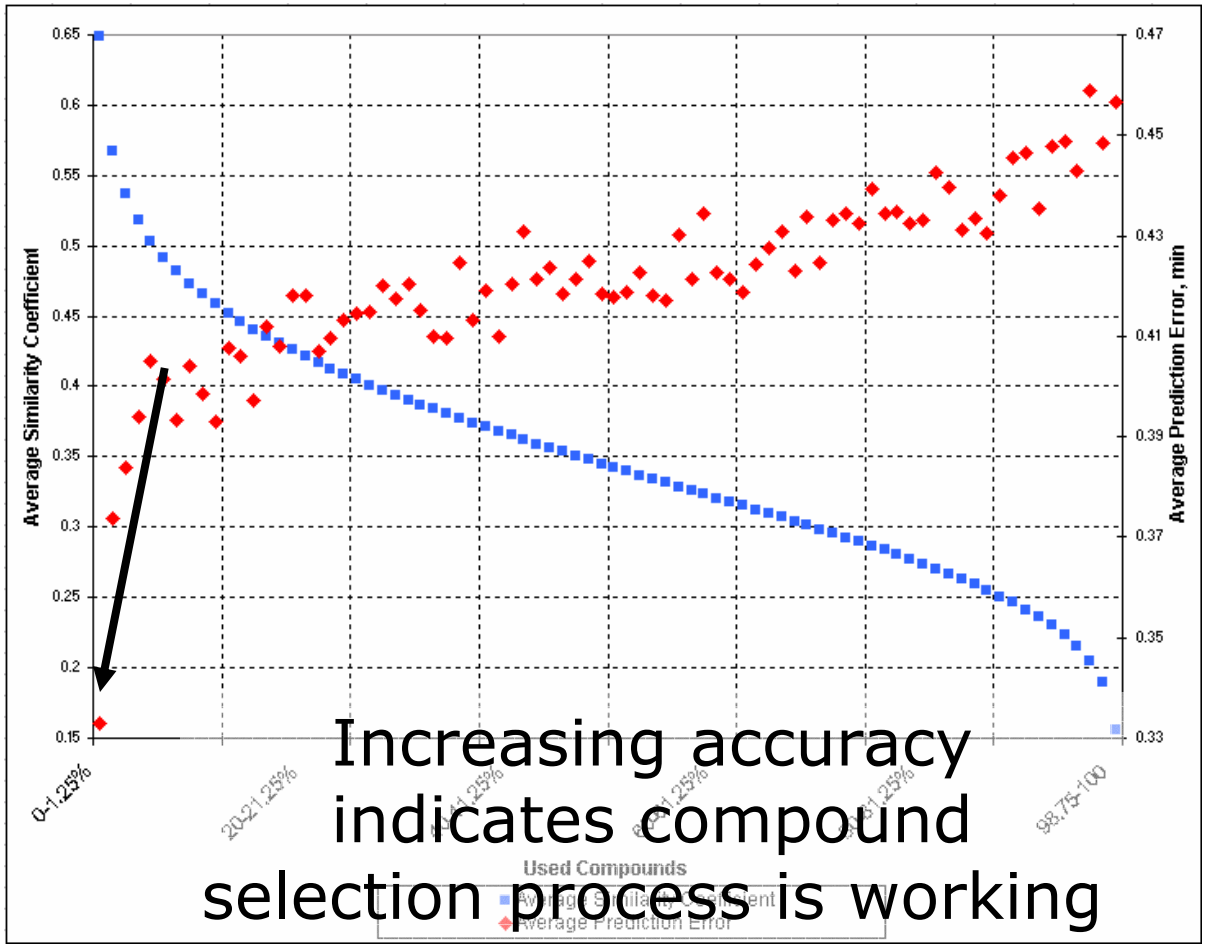
Average error goes from 9 to 13% as similarity goes from 0.65 to 0.15




These points represent normal working conditions

# Structure Similarity and Accuracy – 2006 compounds

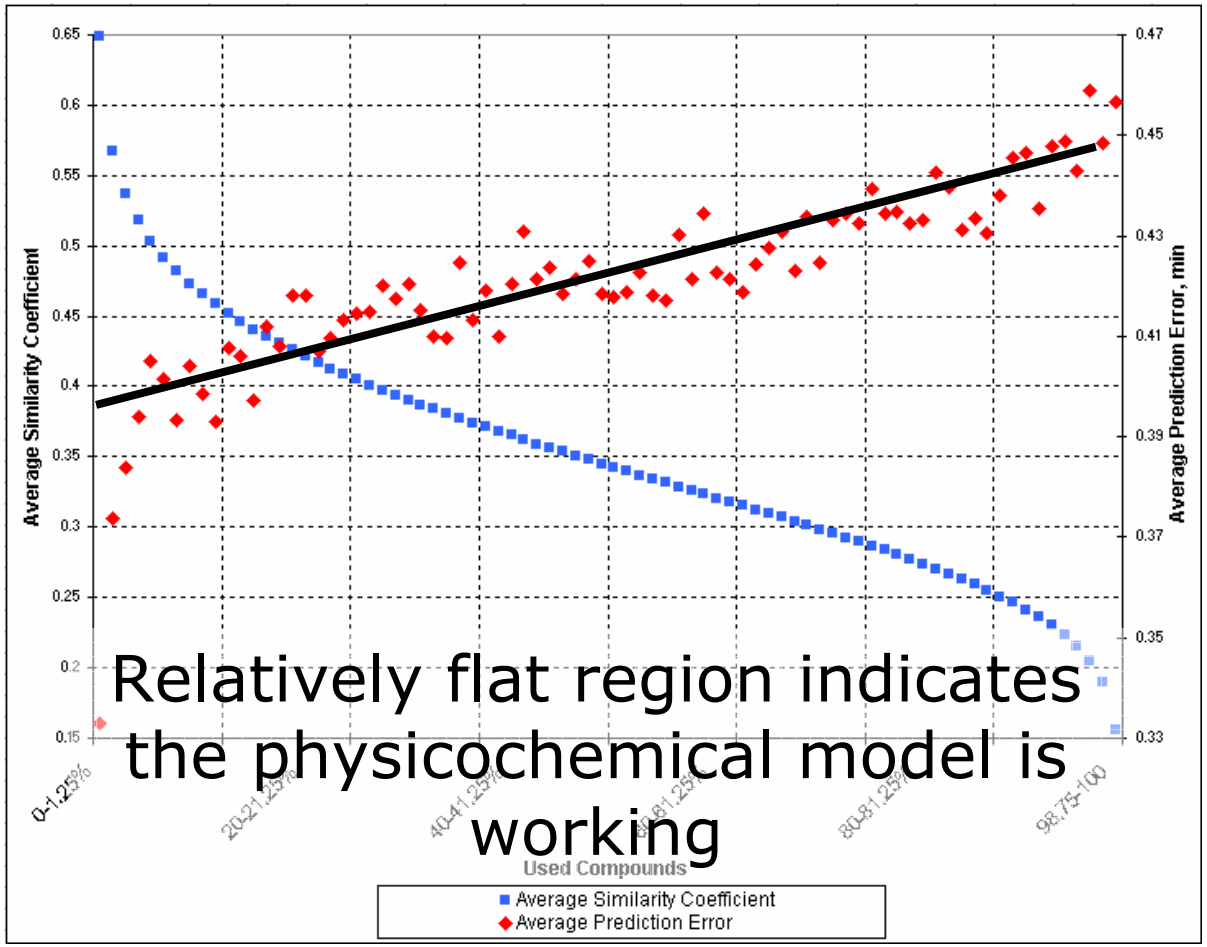
ChromGenius




Increasing accuracy indicates compound selection process is working

# Structure Similarity and Accuracy – 2006 compounds

ChromGenius

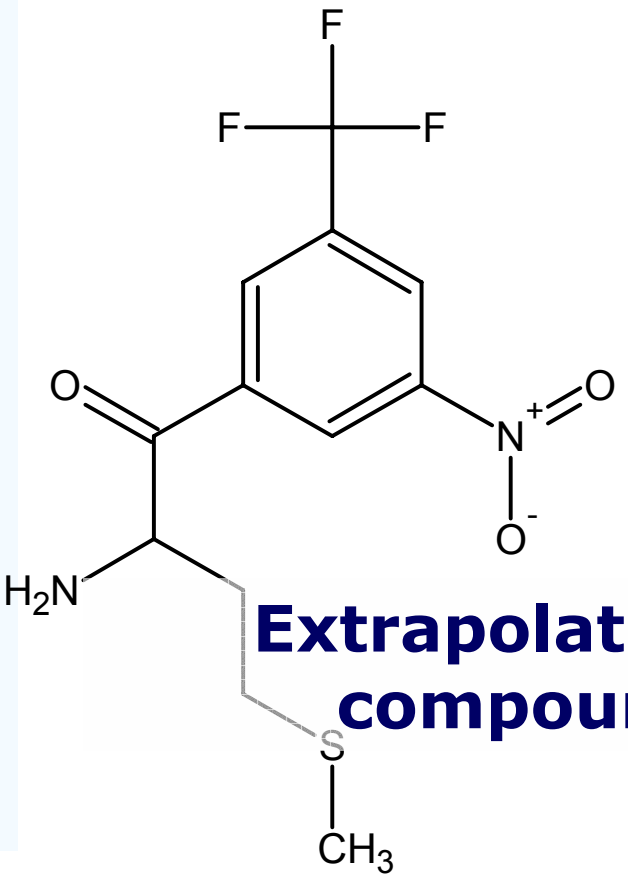



Relatively flat region indicates the physicochemical model is working

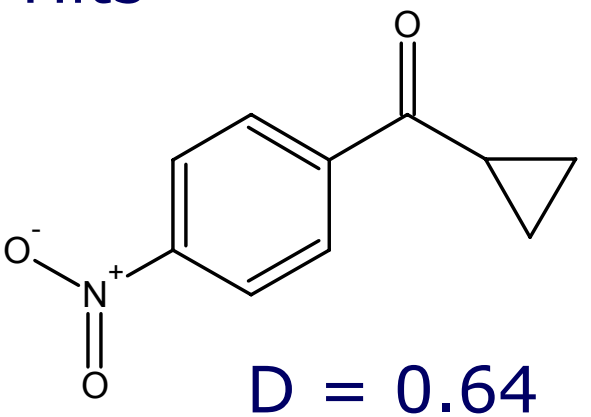
# How Similar is Dice Coeff. = 0.65?

ChromGenius

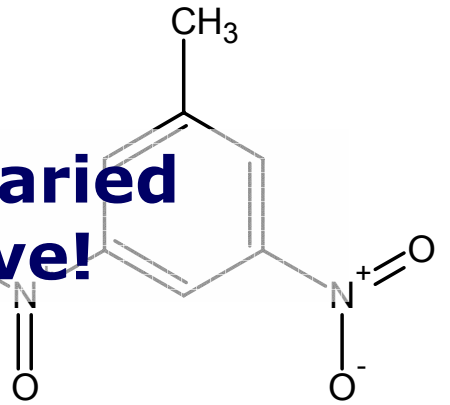
Target molecule




Hits




D = 0.64




D = 0.66

**Extrapolation to new, varied compounds is effective!**

# Better Analytical Methods Mean Better Purity

- Decisions are based on initial data:
  - Use/purify/reject
  - Better estimations of purity lead to better decisions
- Analytical methods are normally scaled up for preparatory work:
  - Better analytical methods are easier to scale up

…users have better libraries with higher confidence in their measurements.

ChromGenius

# Tiered Methods

- Are all methods equally desirable?

ChromGenius

# TFA and Generic Methods

- TFA is great for chromatography.
- TFA is bad for drug discovery:
  - Cytotoxicity at 10 nM concentrations
  - chemical stability issues
- TFA should be used as an additive on the prep scale only when absolutely necessary.
  - ChromGenius is designed with "tiered" methods.

ChromGenius

# ChromGenius – Batch versus Single Sample

ChromGenius

- ChromGenius is designed to support combichem as well as walk-ups
- Plate data can be input with spreadsheets of method choices as output; or
- Users input sample information manually and view expected chromatograms

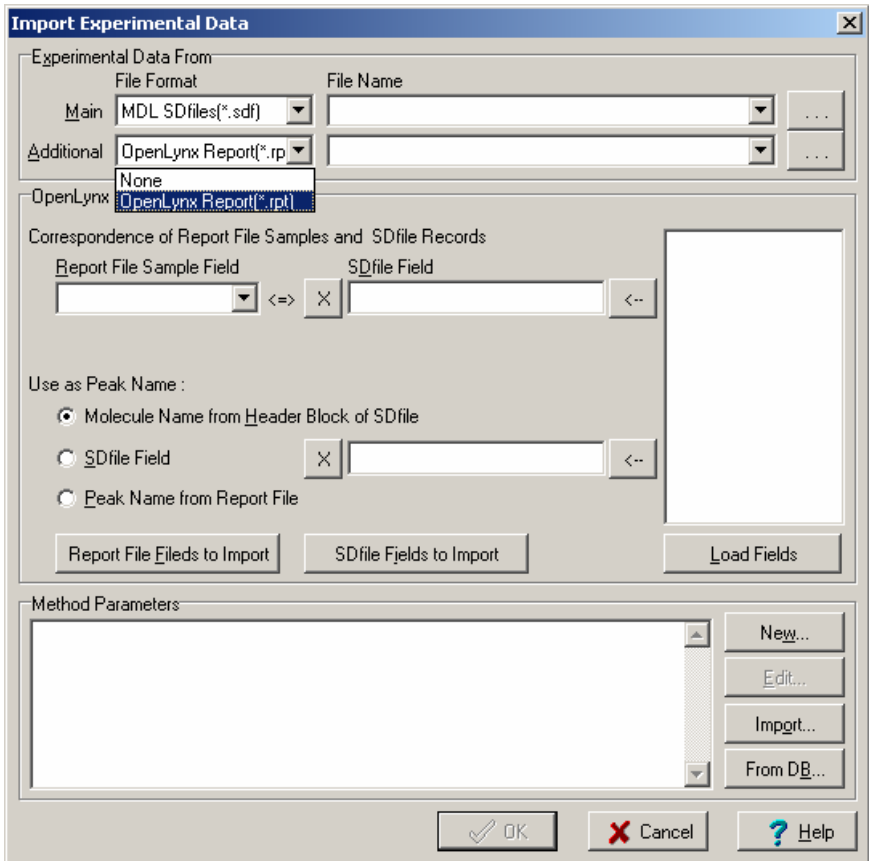
# Creating the Database
# **Amassing a Knowledge Base**

ChromGenius

- The methods we ship are excellent
- Your methods may be preferred
- ChrGen databases can be created easily:
  - ChromManager database conversion
  - SDF import

# Creating the Database
# **OpenLynx Data Conversion**

- Automatic .SDF/.RPT retention time extraction utility




ChromGenius

# Structure-based Method Selection

- Inappropriate methods may mean:
  - Costly reruns
    - Instrument/analyst time wasted
    - Delays in the project
  - Inaccurate purity estimations
    - Incorrect decisions
    - Missed candidates?
- ChromGenius can help increase throughput and sample quality

ChromGenius

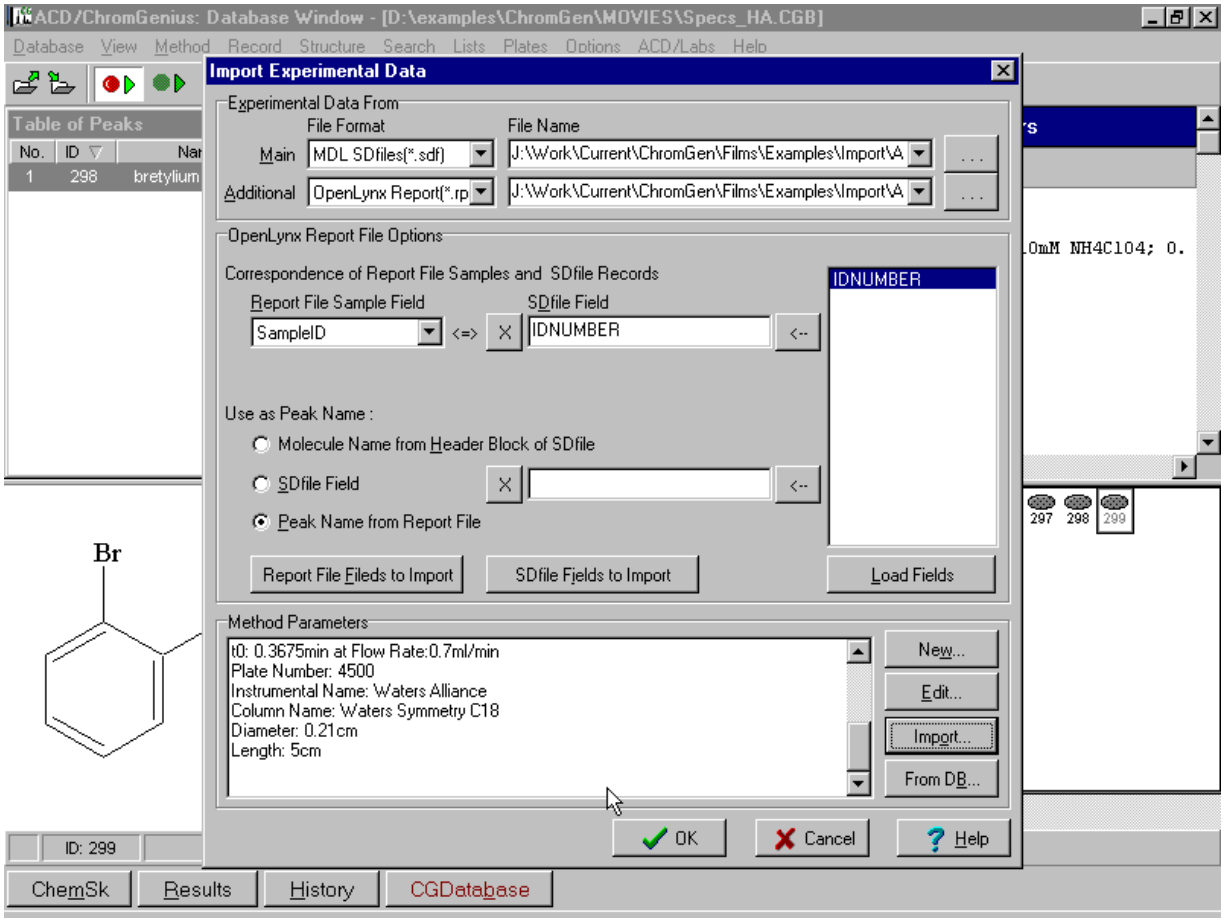
# What's new in v8.0?

ChromGenius

- Faster calculations
- Enhanced import options
- Improved plate view and report

# Can import RT from MassLynx .rpt

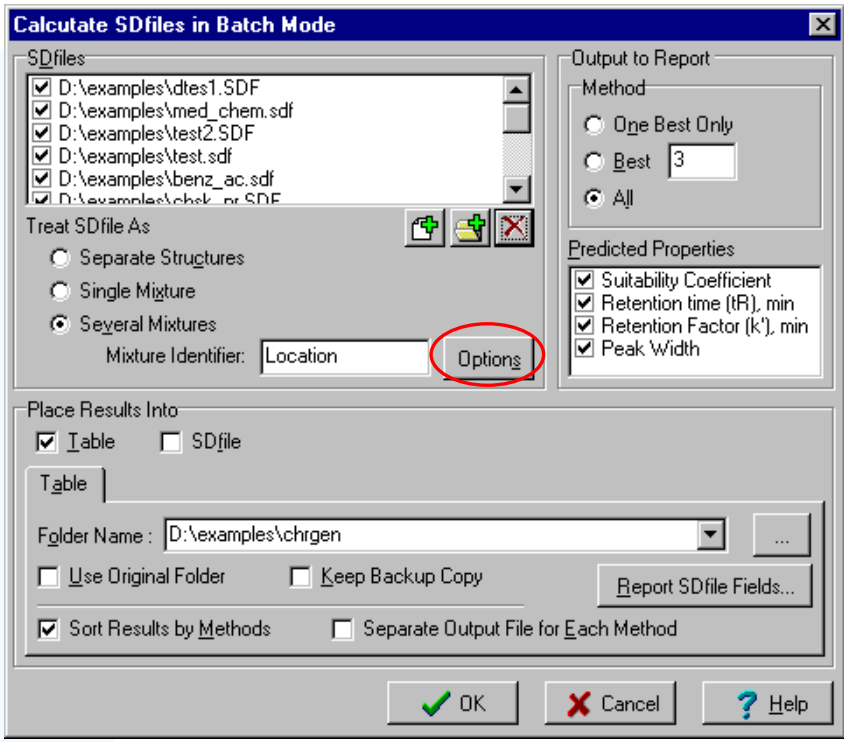
ChromGenius

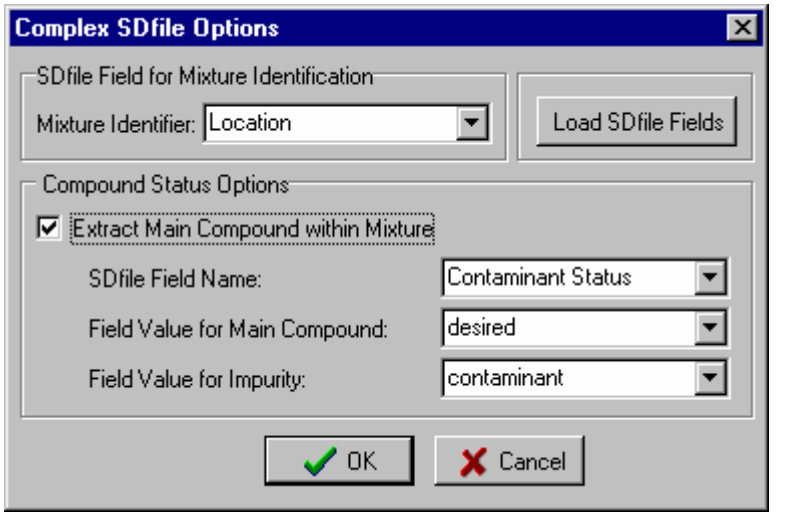



**Visionary Software for Scientists**

**Spectroscopy • Chromatography • PhysChem • Chemical Naming • Drawing • Databasing • Enterprise Solutions**

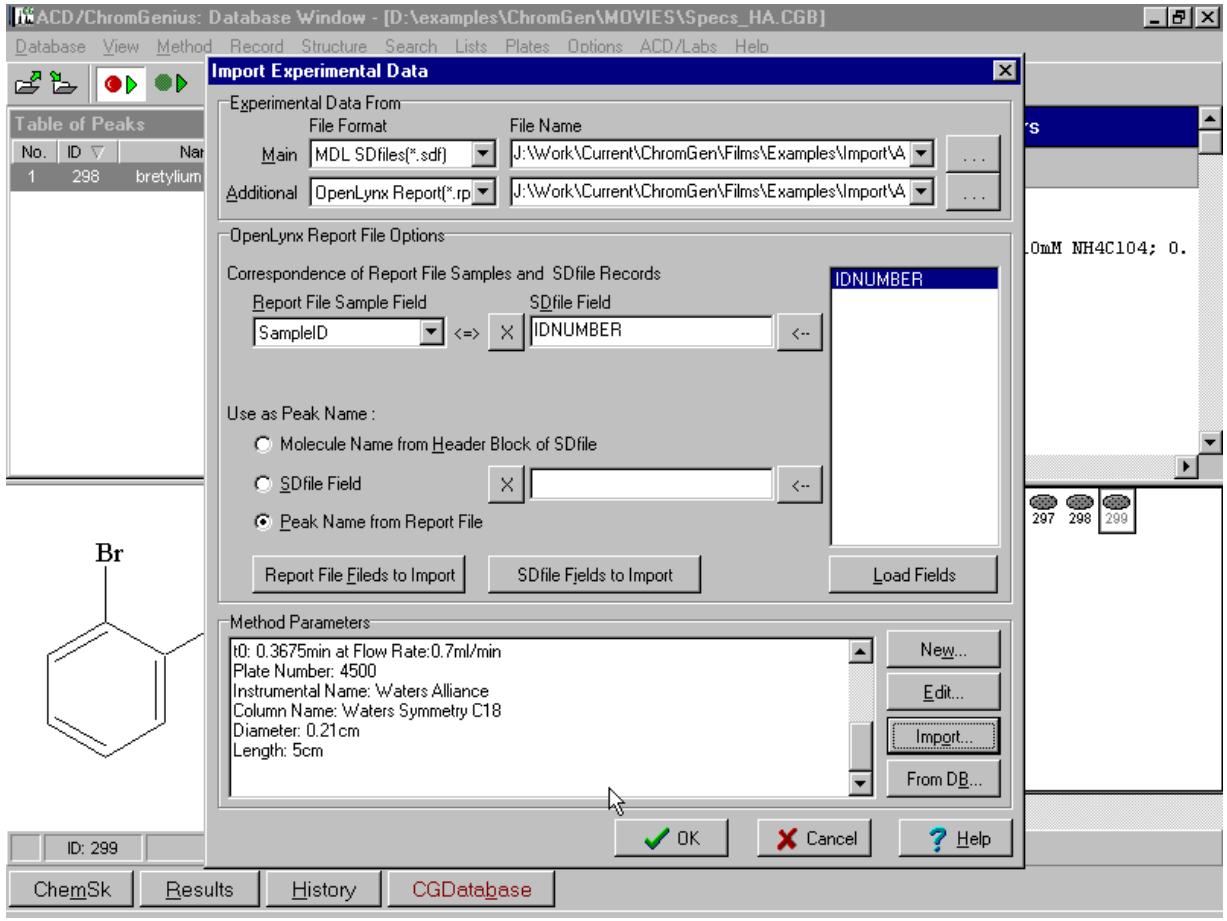
# New SDF calculation options



# More SDF options

ChromGenius

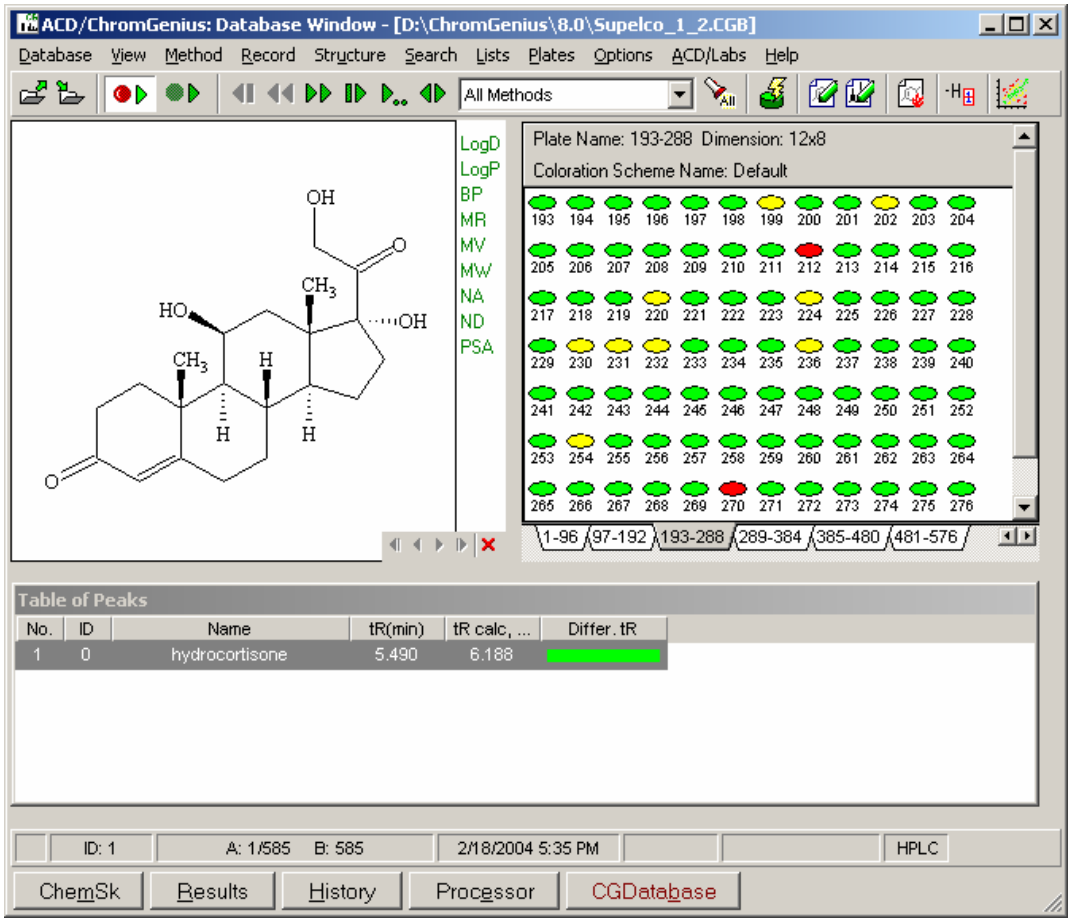



**Visionary Software for Scientists**

**Spectroscopy • Chromatography • PhysChem • Chemical Naming • Drawing • Databasing • Enterprise Solutions**

# Enhanced plate options

ChromGenius




**Visionary Software for Scientists**

**Spectroscopy • Chromatography • PhysChem • Chemical Naming • Drawing • Databasing • Enterprise Solutions**

# Acknowledgements


- Michael McBrien, ACD/Labs
- Rhiannon Jones, ACD/Labs
- Eduard Kolovanov, ACD/Labs