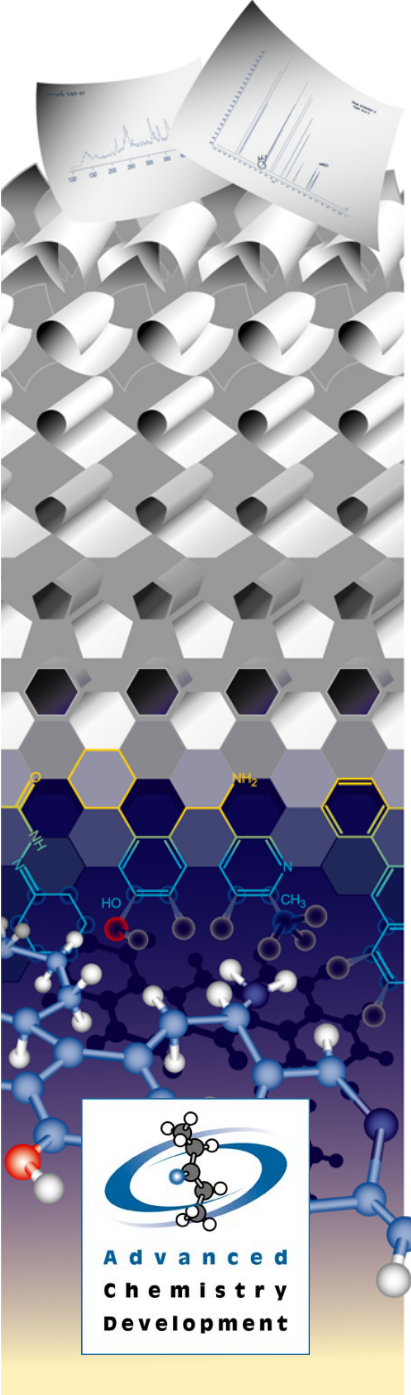


Structured Solutions

Streamlining Separations
ACD/Labs Chromatographic Software Tools

Mike McBrien
Advanced Chemistry Development

HPLC 2003 June 18/03

Advanced Chemistry Development

Advanced Chemistry Development

Integrating chemical structures with analytical chemistry

Structured Solutions

Spectroscopy • Chromatography • PhysChem • Naming • Drawing and Databasing • Enterprise Solutions

Outline

- Method Development Suite
  - Application Databasing
  - Prediction of Retention Times
  - Optimization
  - New in 7.0
- ChromGenius
  - High-throughput method selection
- "Mobile Chemistry"
  - Column Selector for the Palm

Computer-assisted method development

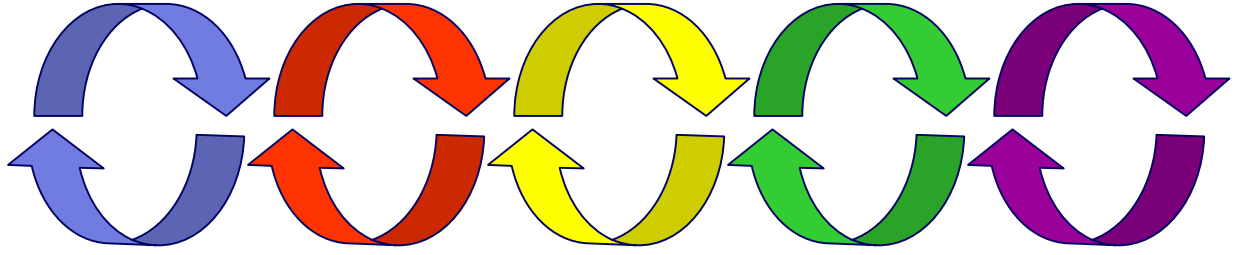
1. Application Databases
2. Best starting point based on structure
3. Prediction of modifications
4. Experimental optimization
5. Archival of successful separations

The Chromatographic Method Development Cycle

Modify:

- Column
- Mobile Phase
- Temperature
- Gradient
- pH
- Type of chromatography???

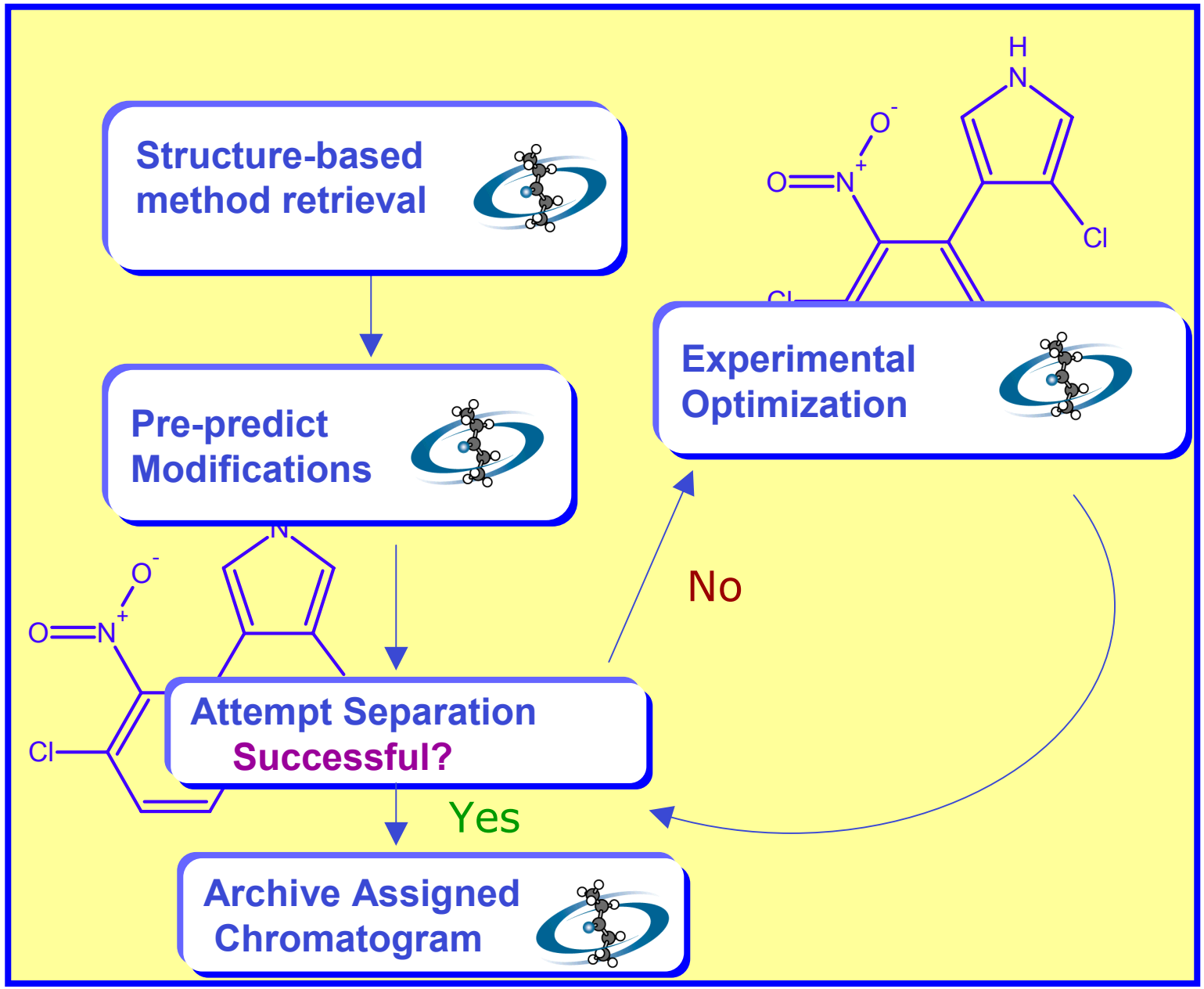
Starting Point



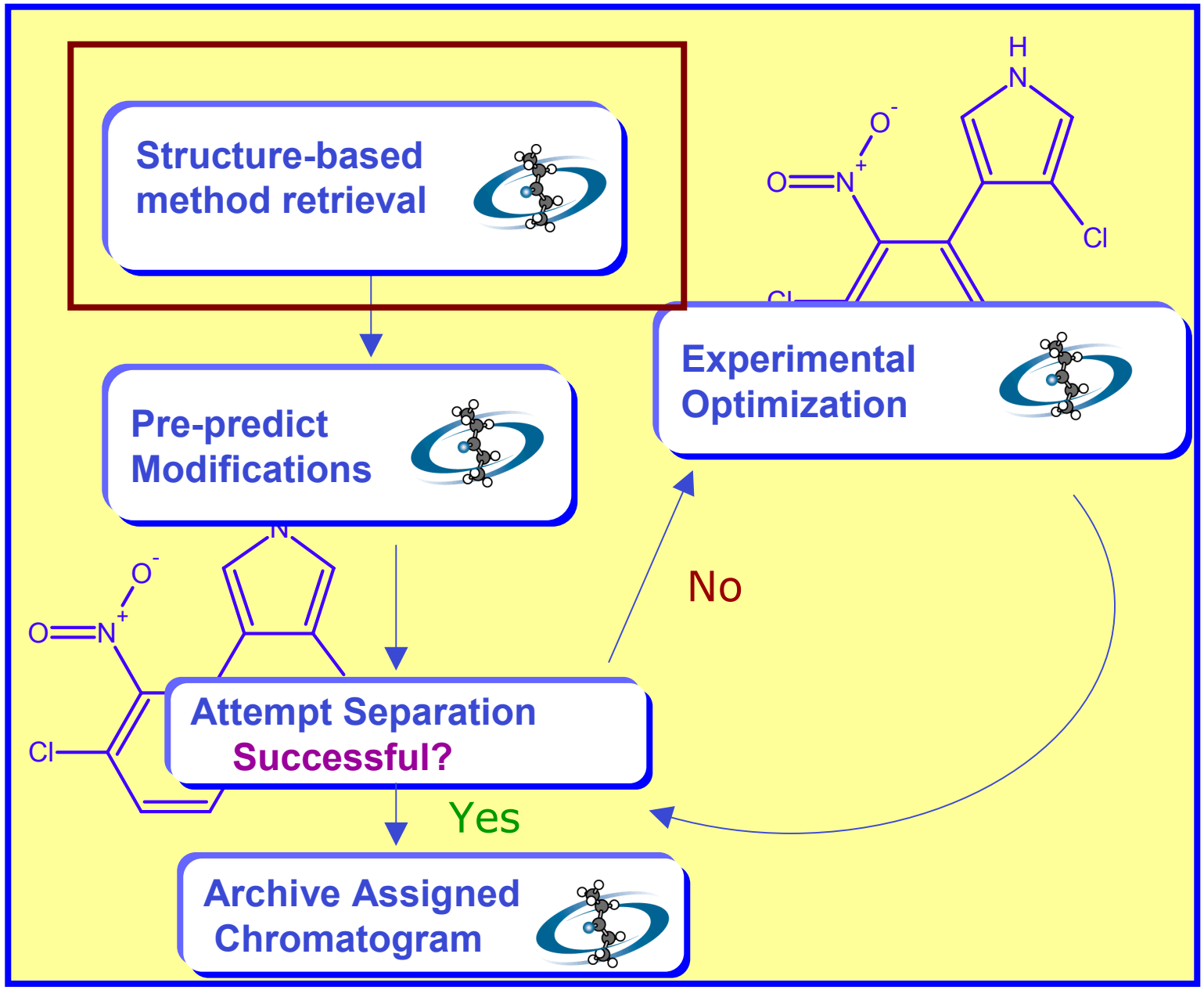
Suitable Solution ?

So what can we do to streamline this complex task?

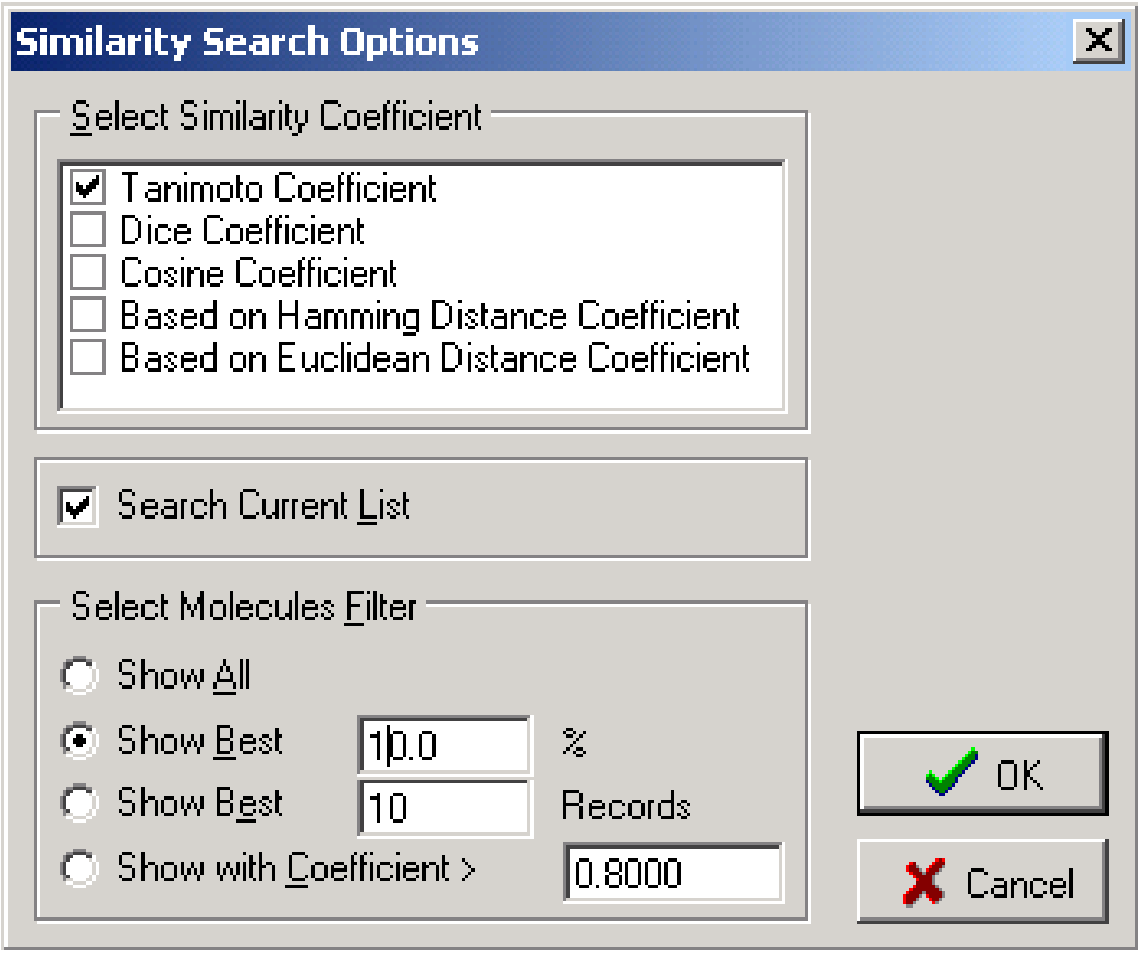
"Structured" Method Development



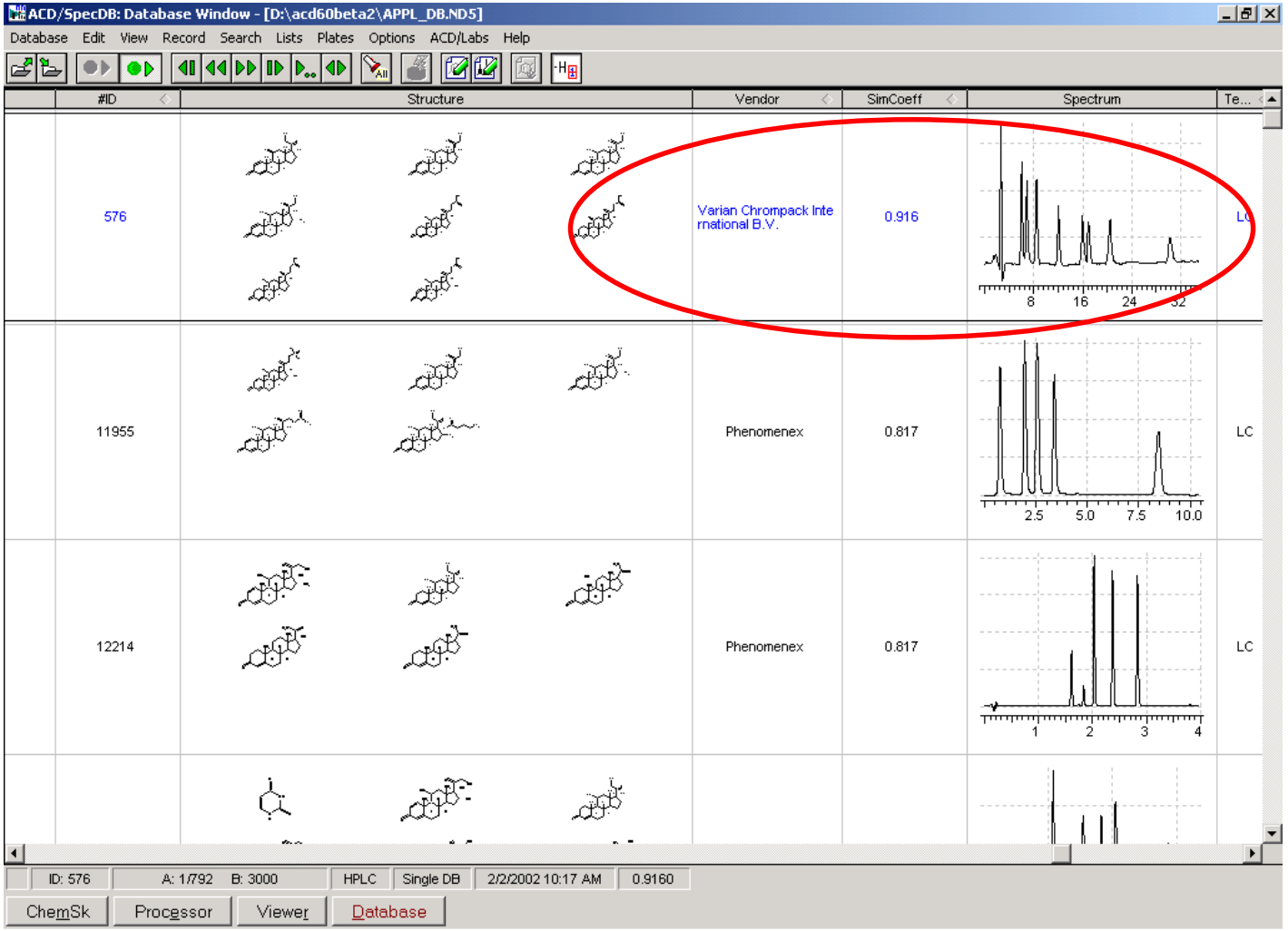
"Structured" Method Development



Structure Similarity Search



Retrieving a Method



ACD Chrom AppDB Collaborators

Chromatography column vendors

- 17 collaborators
- Years of research
- ~4810 applications/20000 structures currently in DB
- Regular updates
- HPLC, Chiral, CE and GC

Thermo Hypersil

astec

Agilent Technologies
Innovating the HP Way

REGIS TECHNOLOGIES, INC.

GL Sciences Inc.

HAMILTON

Alltech

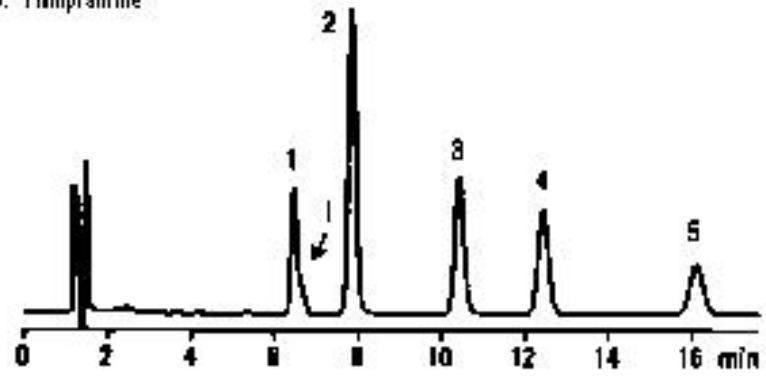
phenomenex

VARIAN

Structured Solutions

Method Development and Chemical Structures – Prediction Mode

Sample:
I. Impurity
1. Doxepin
2. Imipramine
3. Nortriptyline
4. Amitriptyline
5. Trimipramine

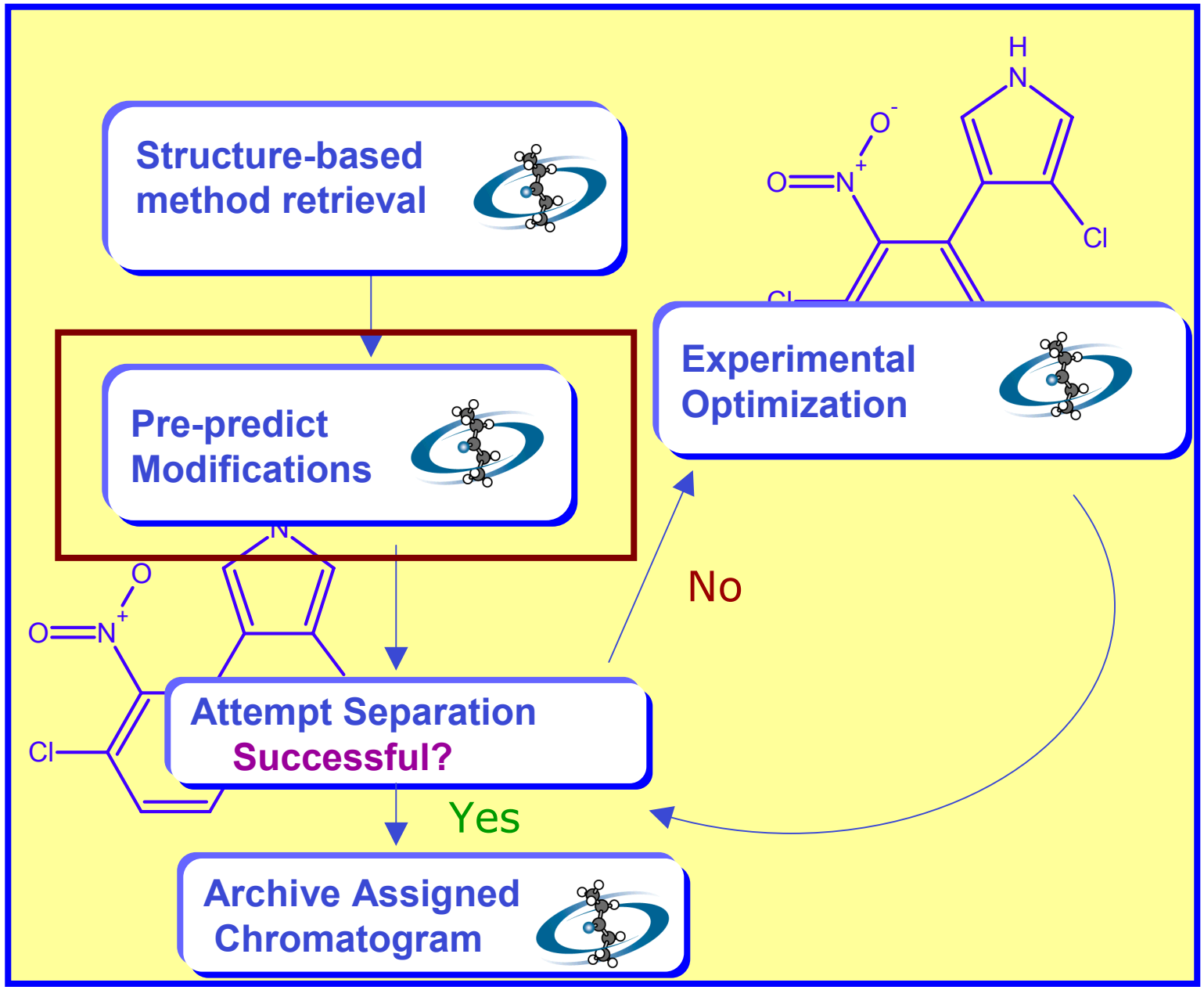


Column:	ZORBAX Extend-C18 4.6 x 150 mm
Mobile Phase:	75% Methanol / 25% 50 mM Pyrrolidine Buffer, pH 11.5
Flow Rate:	0.5 mL/min.
Flow Rate:	Panel A: 1.5 mL/min.; Panel B: 1.0 mL/min.
Temperature:	40 °C
Detection:	UV at 215 nm

Basic drugs can often be separated in their charged form at low pH with StableBond or at mid-range pH with Eclipse XDB or Bonus-RP columns. With Extend C18, you can separate at high pH to improve solubility, improve retention, or obtain different selectivity.

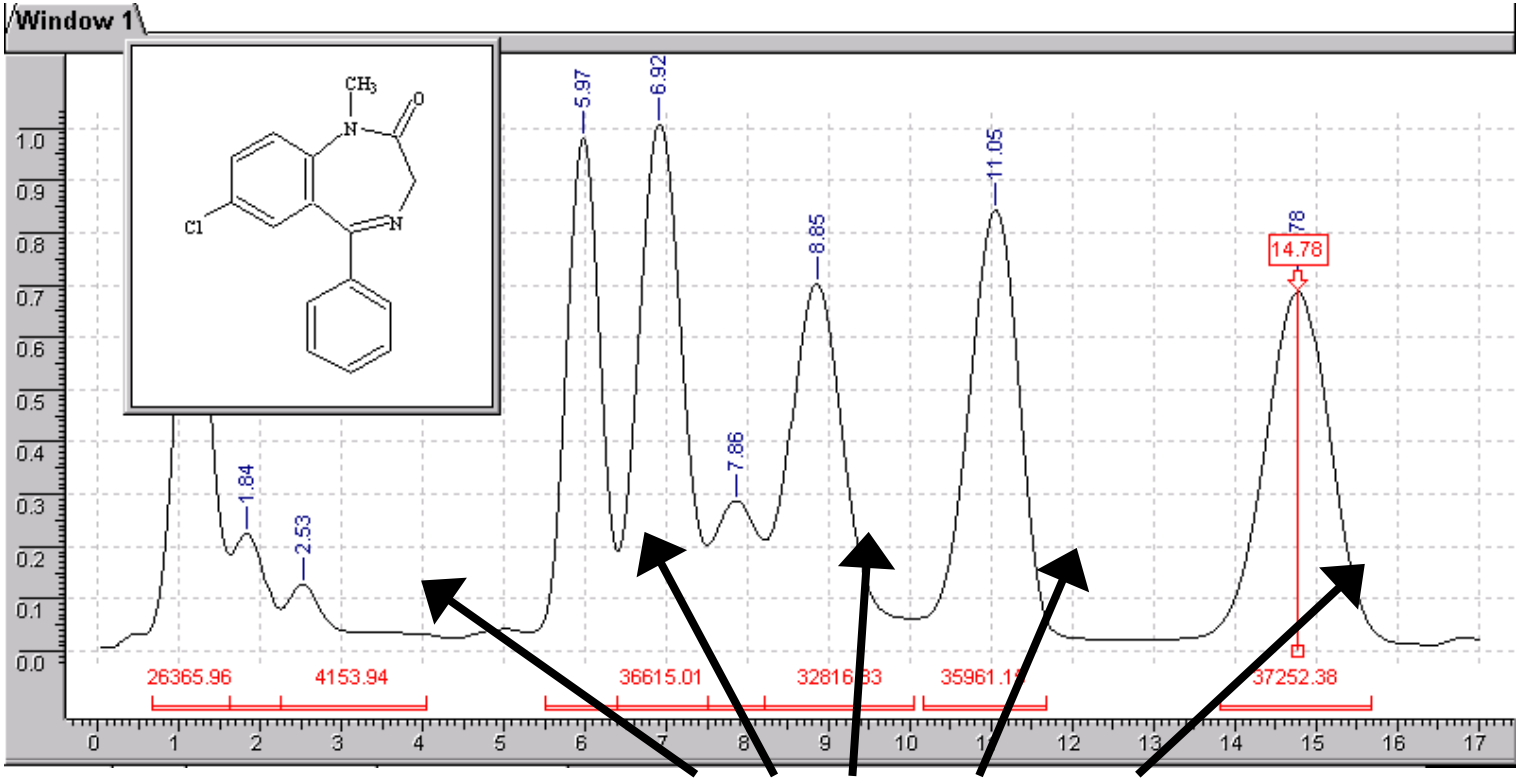
- How effective will the retrieved method be?

"Structured" Method Development



LC Simulator
Prediction of tRs

Given an experimental set:



Where will my new compound elute?

LC Simulator and tR Prediction

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

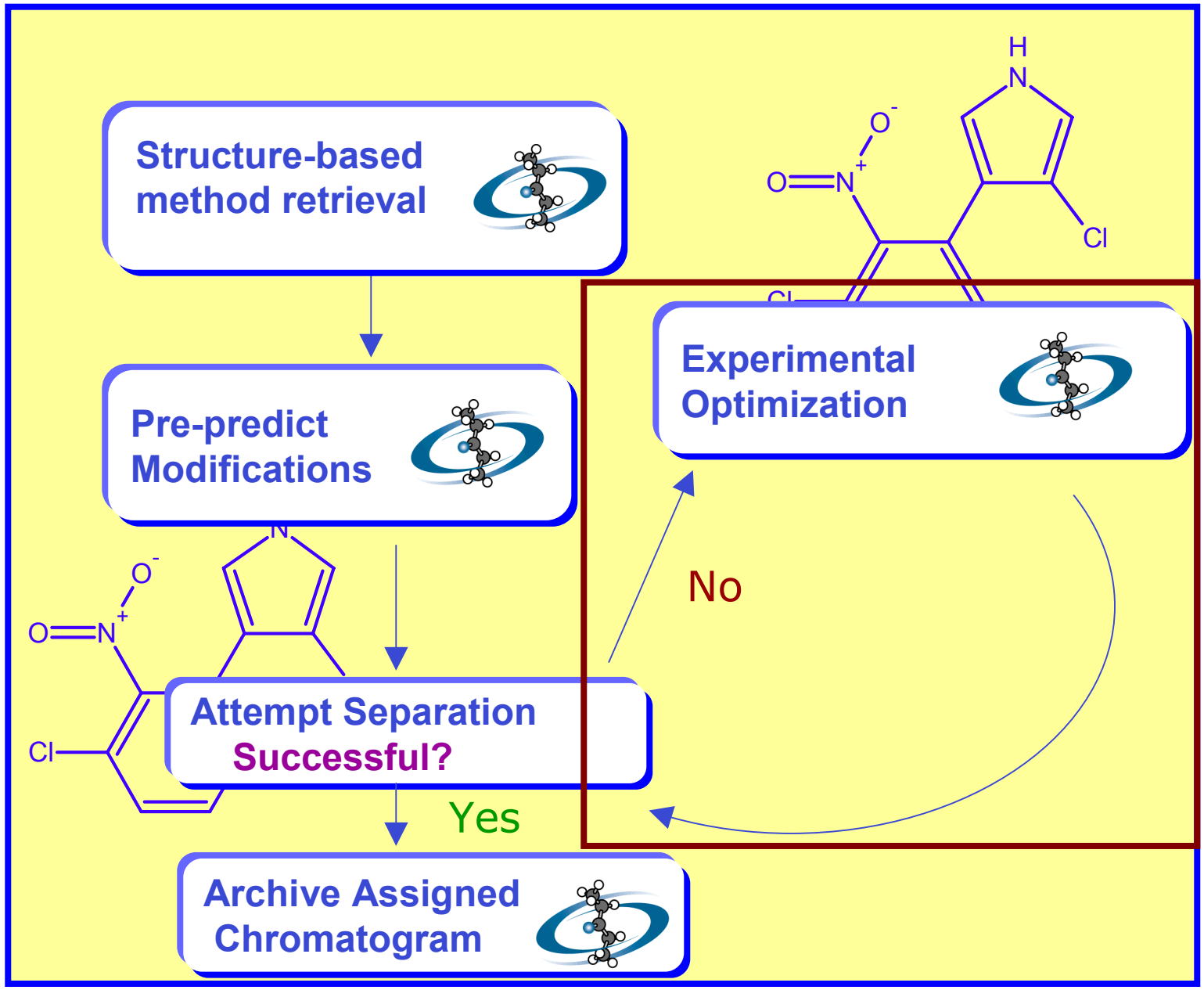
LC Simulator uses experimental retention times for old compounds as a "training set" to predict retention times for new compounds.

Benefits of Applying Prediction Mode

Non-viable separations?

1. Reject method and find another
2. Modify strength of solvent
3. Modify/introduce gradient
4. Modify pH
   - Resolution and/or robustness.

"Structured" Method Development

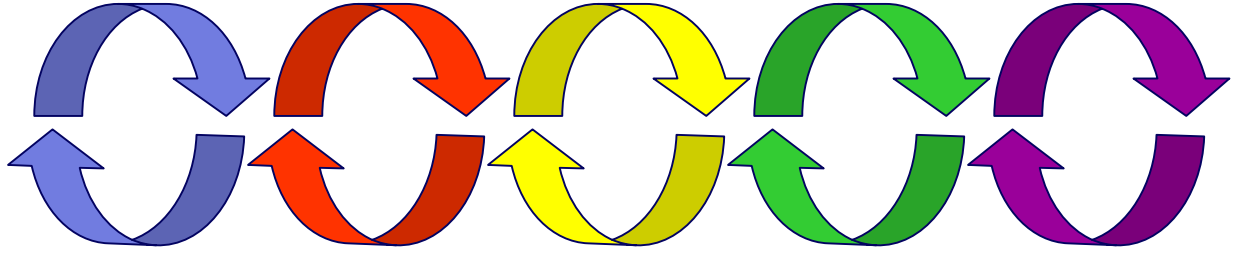


The Chromatographic Method Development Cycle

Modify:

- Column
- Mobile Phase
- Temperature
- Gradient
- pH
- Type of chromatography???

Starting Point



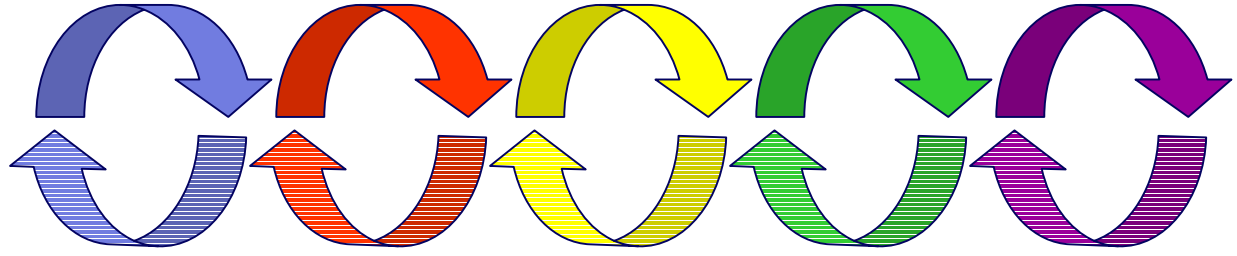
Suitable Solution ?

The Chromatographic Method Development Cycle

Modify:

- Column
- Mobile Phase
- Temperature
- Gradient
- pH

Starting Point



Suitable Solution

Computer-assisted Method Optimization

The goal is to reduce the amount of trial and error by using computer modeling to point to better experimental conditions based on a few runs.

Optimization Mode
Parameters

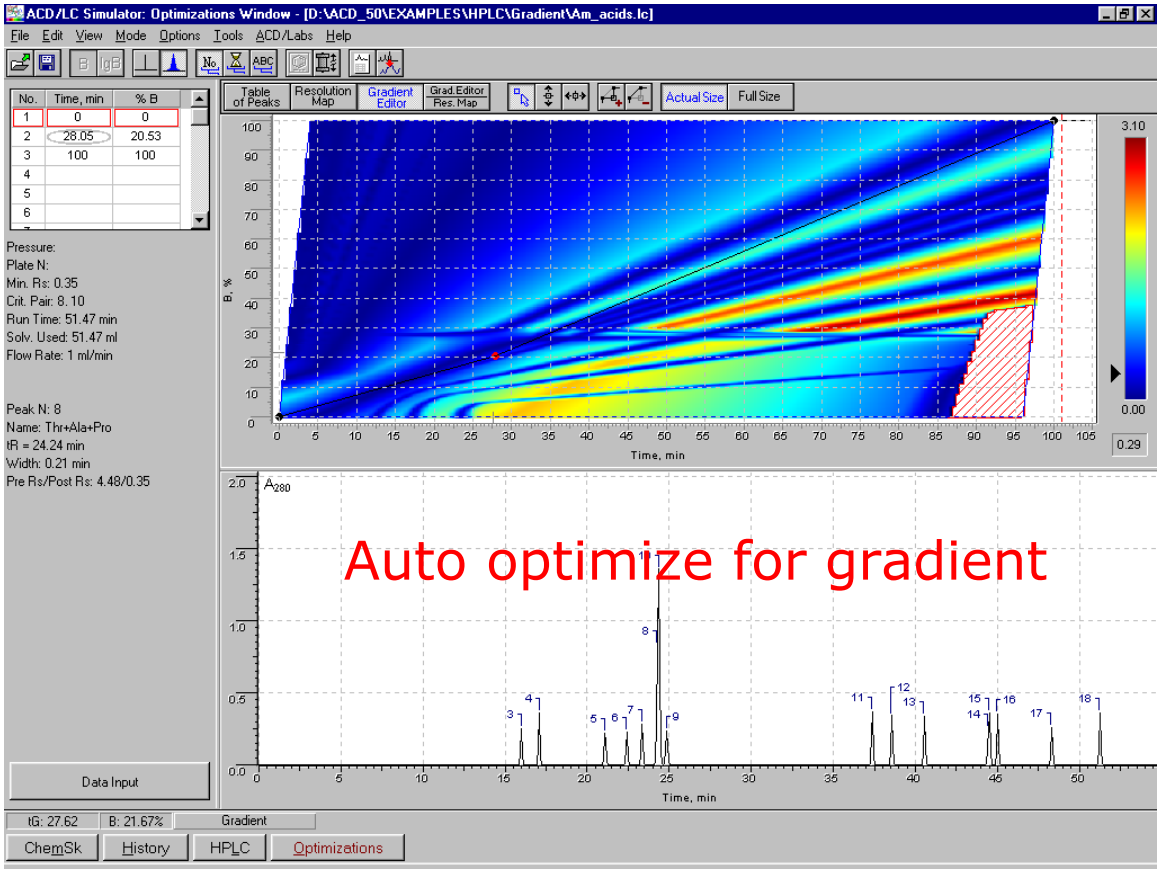
Modify:

- Temperature
- NP isocratic
- RP isocratic
- Gradient
- Salt concentration
- pH

…and 2D modes

Gradient Optimization

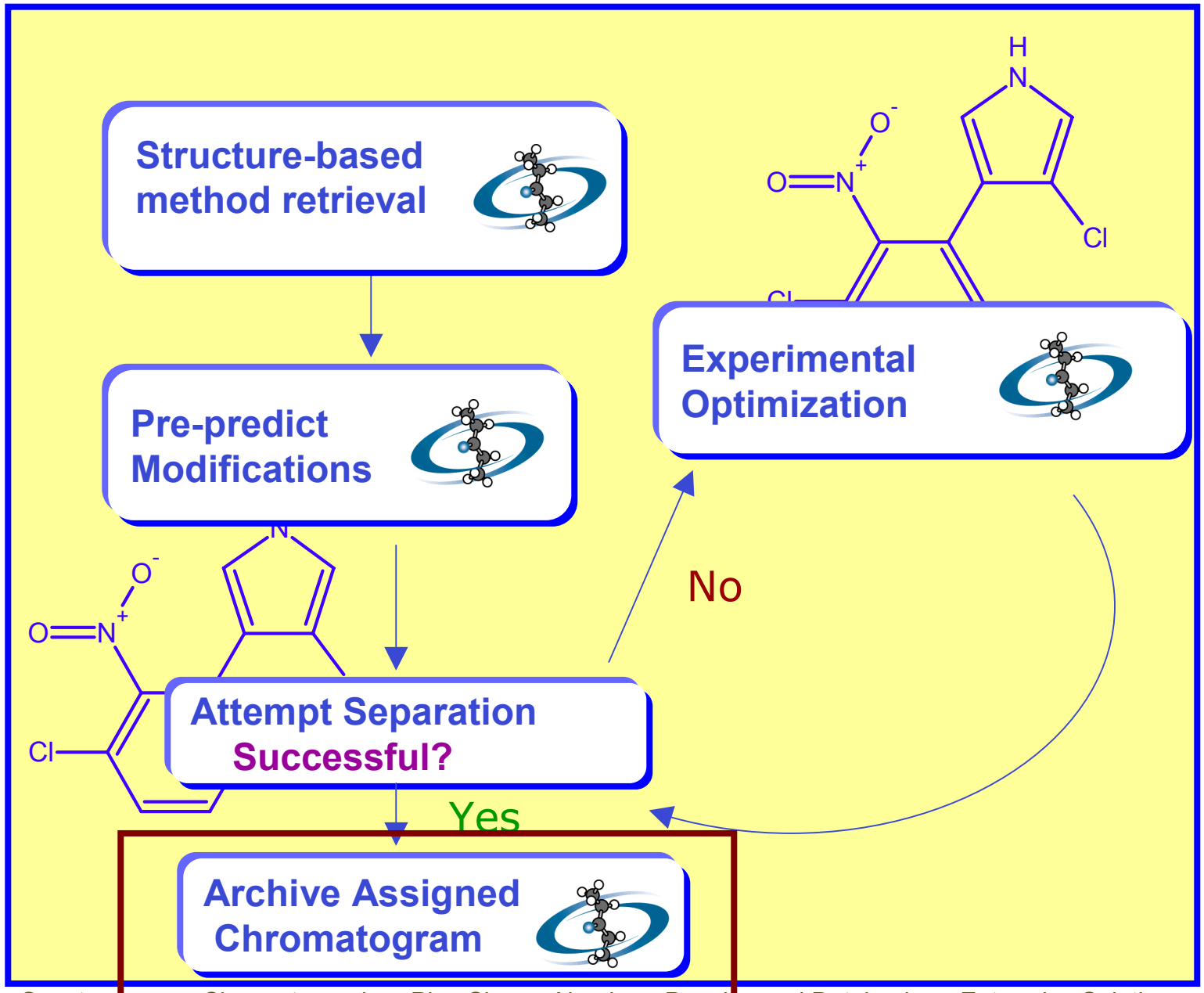
- 2D Gradient Optimization



Experimental Optimization in MDS

- Advantages of computer simulation:
  - Faster development;
  - Shorter runs;
  - increase robustness;
  - Pinpoint and alleviate validation issues.

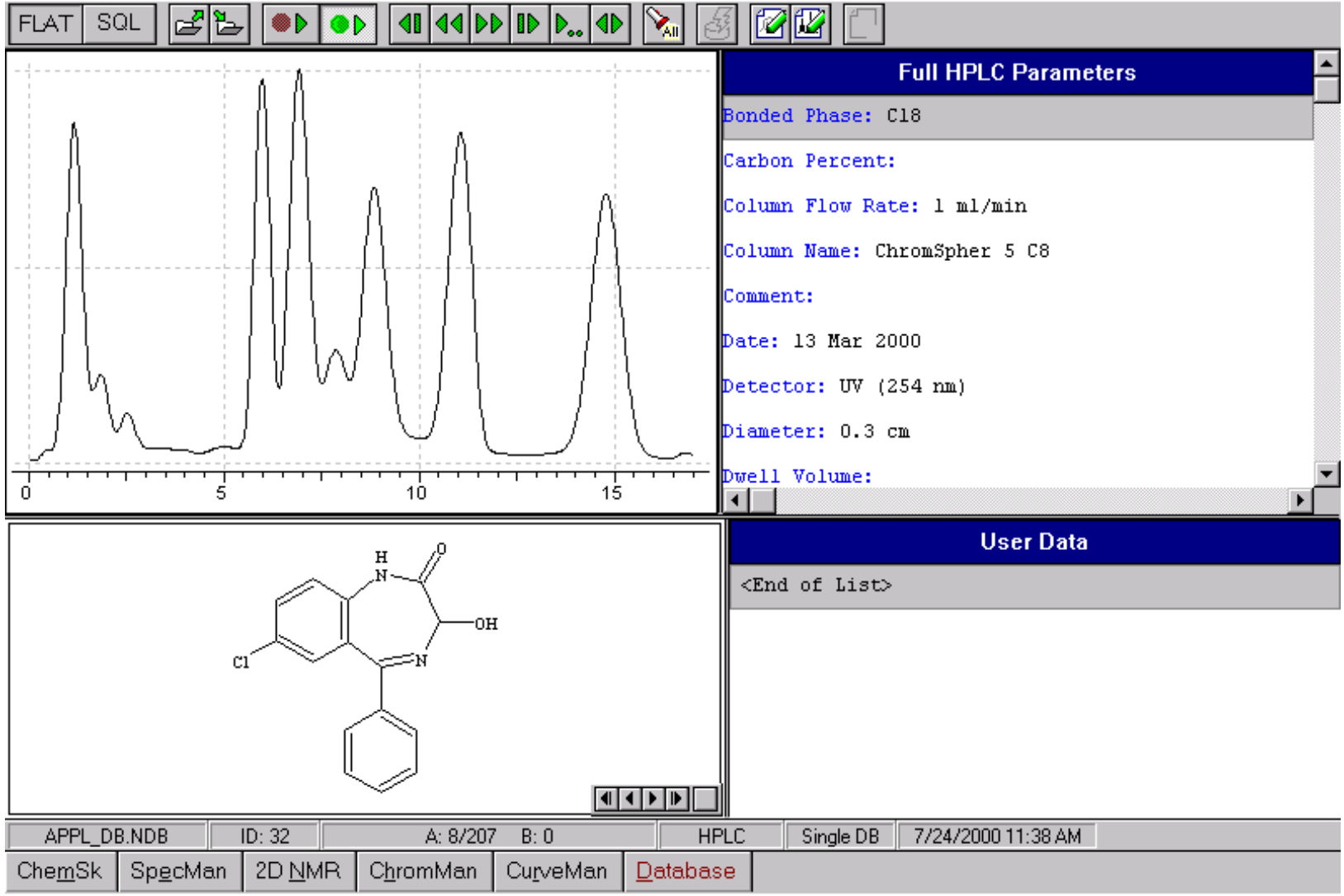
"Structured" Method Development



Archiving Organizational Applications

- Same interface and DB engine as Chrom Applications DB.
- Emphasis on “favoured” methods
- Format support:
  - Millennium^{32}/Empower integration
  - ChemStation Plus
  - TotalChrom
  - Chromeleon
  - Others through NetCDF

Archiving Organizational Applications



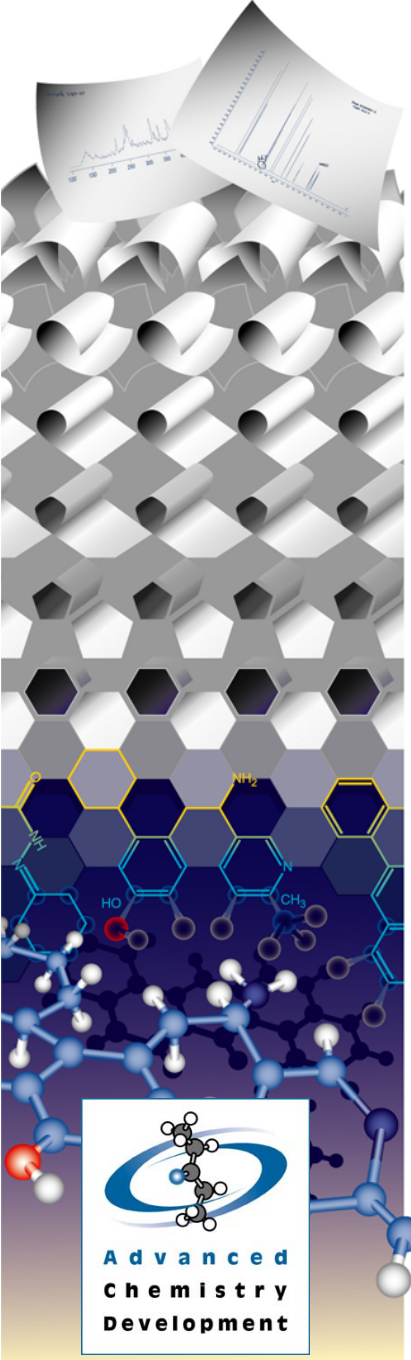
Spectroscopy • Chromatography • PhysChem • Naming • Drawing and Databasing • Enterprise Solutions

Why Use MDS?

- Less time to develop methods
- Faster separations
- More robust separations
- Ease of validation

What's New in 7.0

- Support for more compounds
- More collaborators and applications (17/4810)
- Peak matching tools
- Support for Chromeleon and TotalChrom
- Complex gradient generation tool
- More!



Structured Solutions

ACD/Column Selector for Palm

New for Version 7.0

Advanced Chemistry Development

Spectroscopy • Chromatography • PhysChem • Naming • Drawing and Databasing • Enterprise Solutions

A Chromatographic Column Knowledge Base

- A scientific column classification system
- 9 years' research at AstraZeneca
- 135 characterized columns
- A tool for making use of these characterizations.

What is Column Difference Factor?

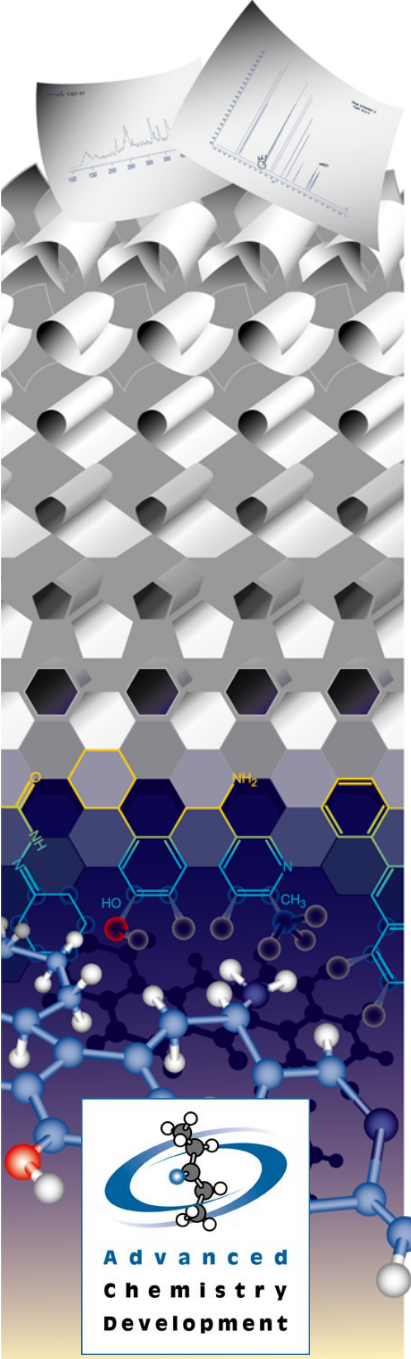
- The "sum" of the differences:
  - Euclidean Distance between the two columns, considering six factors.
- Each term has equal weighting in the default settings.
- CDFs are customizable.
  - Coefficients between 0 and 1 can be introduced depending on user knowledge of the analytes.

Why Use Column Selector?

- You do a search in the AppDB and the column that you need is not at hand:
  - Find the most equivalent, available column.
- The column that you initially tried is not giving you resolution, despite optimization:
  - Find an orthogonal column to try next.
- You have two compounds that differ in a readily discernable fashion, and want to exploit the difference:
  - Find a column with a higher specific coefficient.

Column Selector is Included with ACD/MDS 7.0

- Files are contained in the ACD/Labs directory.
- Run Web Updater to get the latest version.
- or, download the free version from www.acdlabs.com/columnselector/
- Palm emulators enable PC users to use Column Selector!



Structured Solutions

ChromGenius

Spectroscopy • Chromatography • PhysChem • Naming • Drawing and Databasing • Enterprise Solutions

How are high-throughput chromatographers different?

- Faced with hundreds, or thousands of samples.
- No time for method development
- FAST run times are necessary
- Always gradients
- A few standard, or generic, methods
- Almost always LCMS instruments

Generic Methods Currently

Methods Applied without Targeting

- Designed to be widely applicable
- Fast LCMS gradients
- A few methods should cover most anticipated structural diversity
  - Inject all samples under each method and observe the results;
  - or, examine the results and re-inject the samples that didn't work.

Regardless of the untargeted approach applied, throughput suffers.

The Situation Today
Analytical Scale

- Some compounds elute too early:
  - MW verification works; No purity estimation
  - Minimal resolution; bad purity estimation
- Some compounds - "brick dust":
  - Instrument downtime!
- Verification on MW:
  - Sometimes the weight is right, but the compound is not.

The Situation Today
Preparatory Scale

- Compounds elute too early:
  - Resolution insufficient; impure samples = bad screens
  - Elution solvent is not volatile; throughput suffers
- Some compounds elute too late:
  - Run times are too long = low throughput
- Peak shapes are not good:
  - Resolution insufficient = bad screens

Samples are often thrown away due to an inability to purify…

Advanced Chemistry Development

Increased throughput, better information, better quality samples.

Structured Solutions

The Prediction of Chromatography: A Great Challenge

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

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

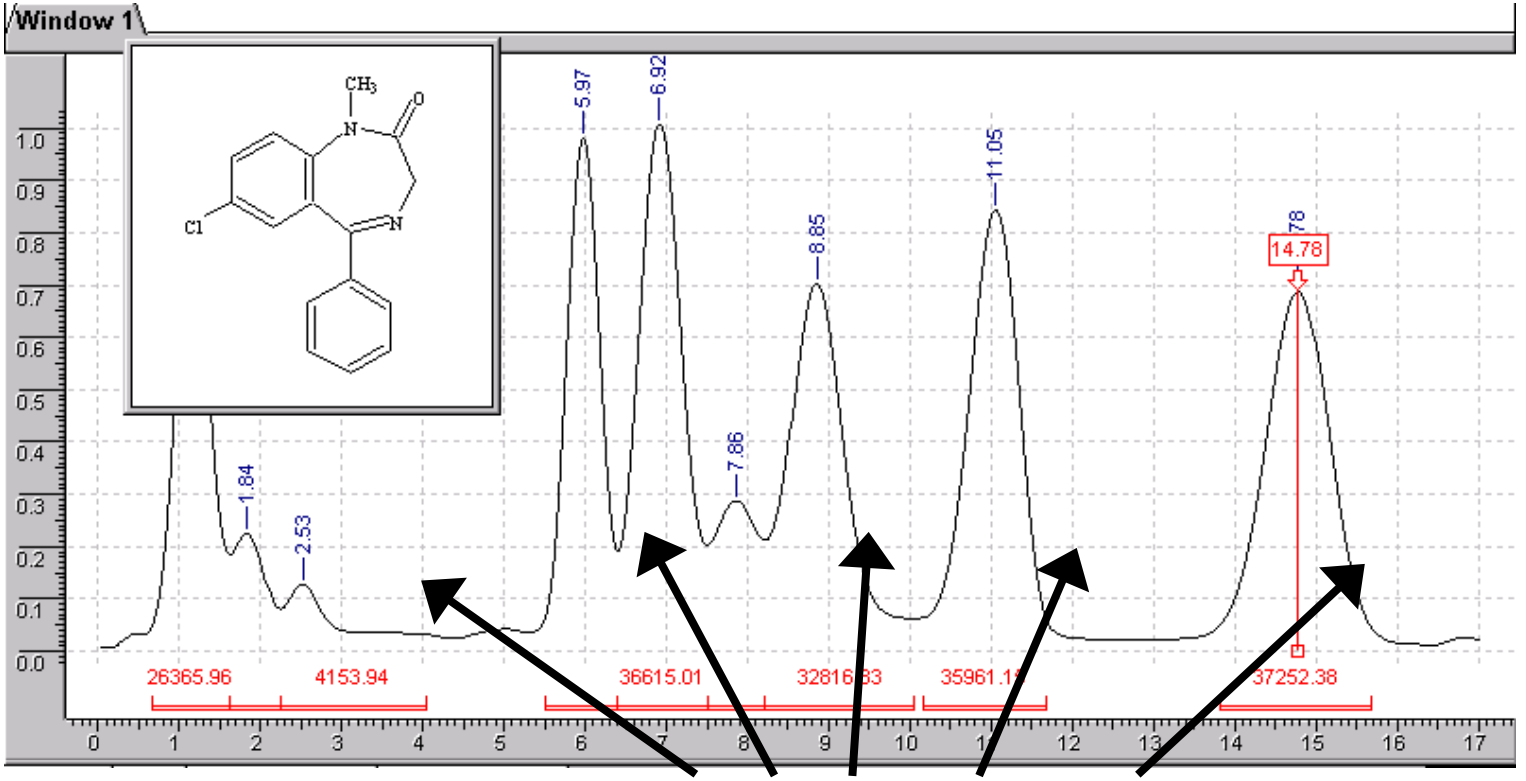
LC Simulator and t_R Prediction.

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

LC Simulator uses experimental retention times for old compounds as a "training set" to predict retention times for new compounds.

LC Simulator Backbone
Prediction of t_Rs

Given an experimental set:



Where will my new compound elute?

What are the limitations of this approach?

- Accuracy is limited by:
  - Similarity of training set to new compounds
  - Number of compounds in training set
  - pH issues
  - Gradients are not explicitly modeled

How can ChromGenius do better?

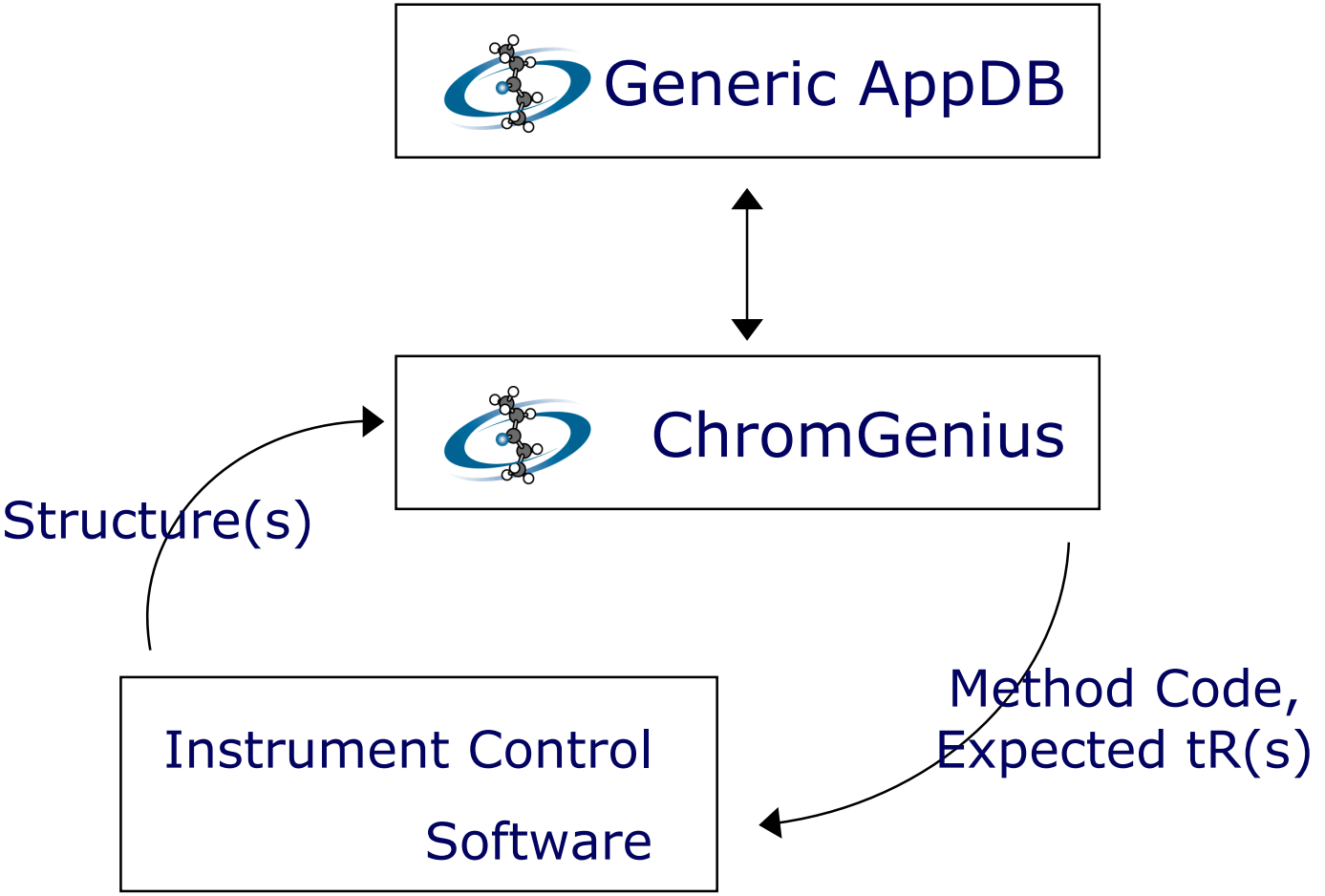
- ChromGenius chooses the most relevant training set from *multiple* chromatograms.
- More compounds:
  - Better characterization of separation
  - More PhysChem terms
- and more relevant compounds:
  - Similar retention mechanism
  - Inherent gradient compensation

Physicochemical Properties

- LogD – hydrophobicity/pKa
- LogP – neutral form hydrophobicity
- PSA – Polar Surface Area
- NA – Number of hydrogen bond acceptors
- ND – Hydrogen bond donors
- BP – Boiling Point
- MW – Molecular Weight
- MV – Molar Volume

ChromGenius is also designed to correct for gradients, and for solvent content/pH issues.

Automated selection of methods
The Process



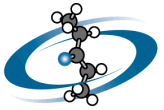
Automated selection of methods
The Process

Instrument Control Software

Automated selection of methods
The Process

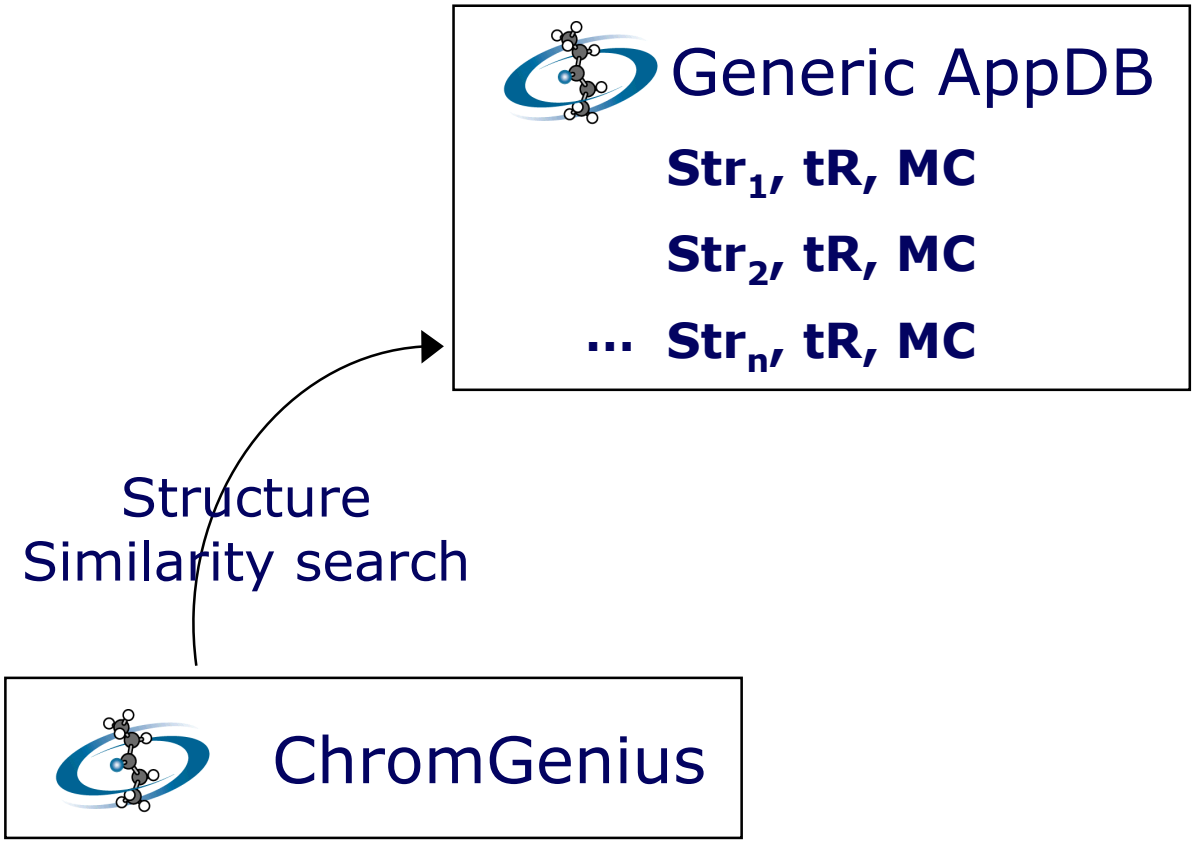
Instrument Control Software

Structures

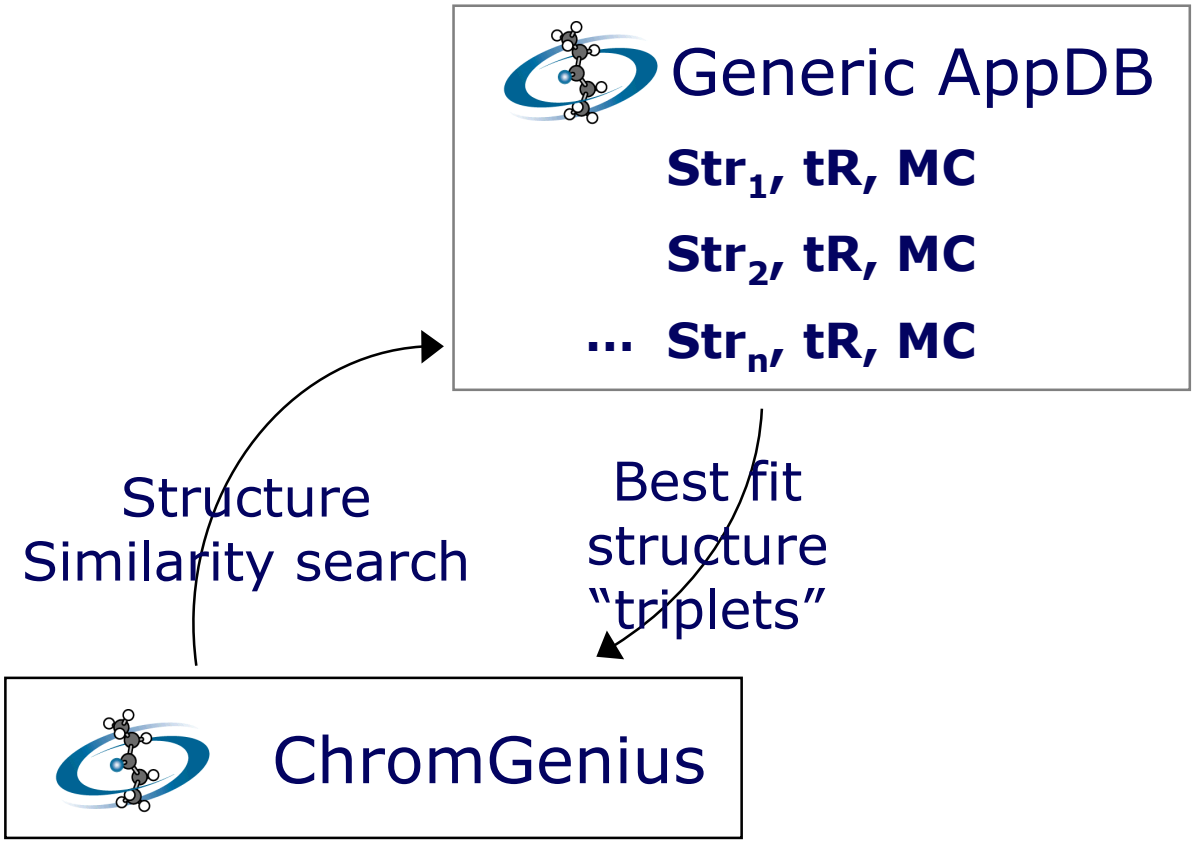


ChromGenius

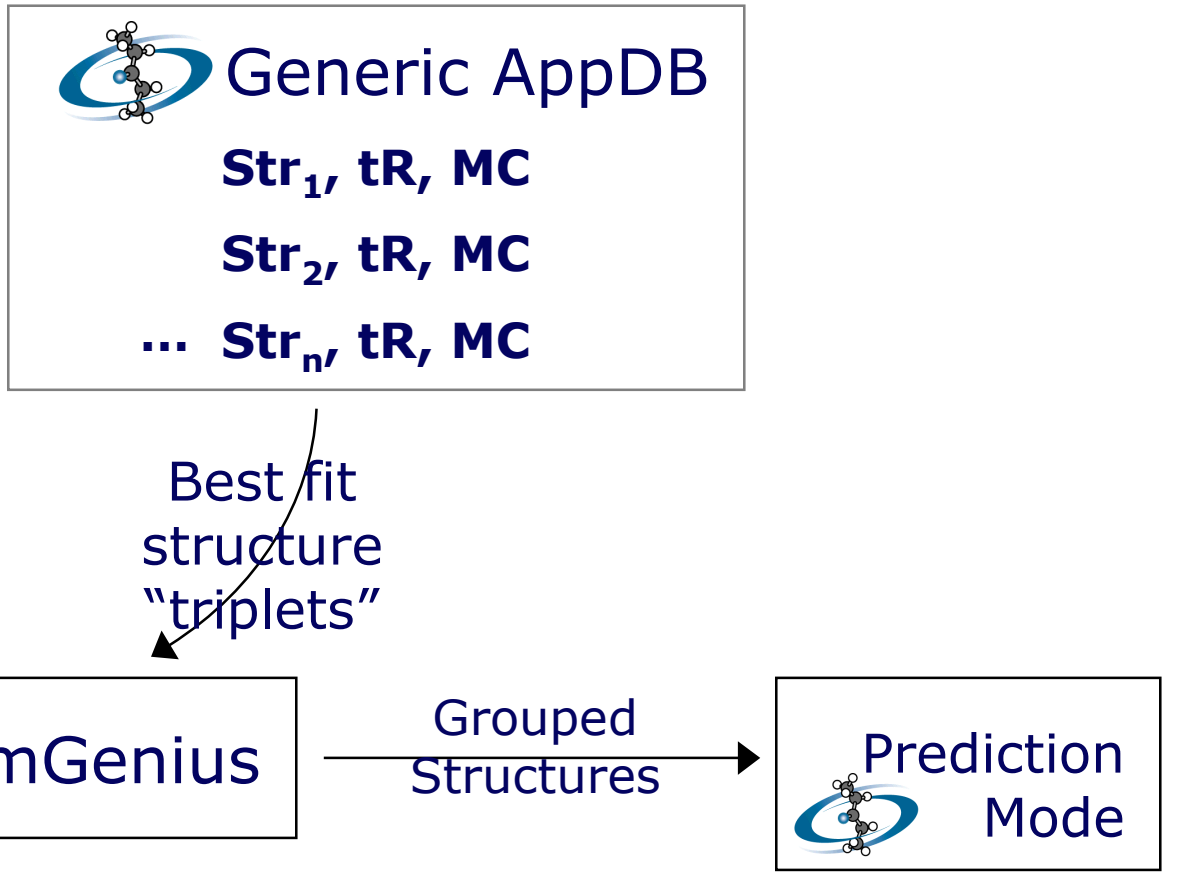
Automated selection of methods
The Process



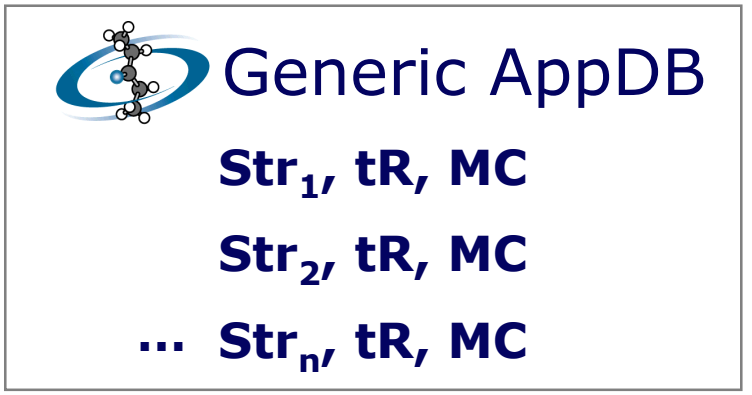
Automated selection of methods
The Process

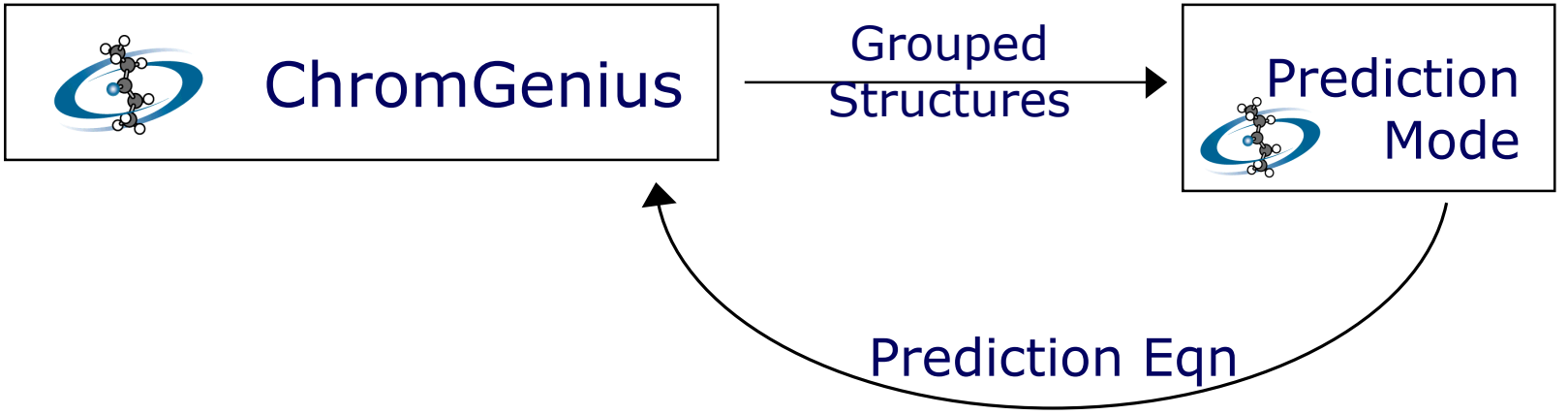


Automated selection of methods
The Process

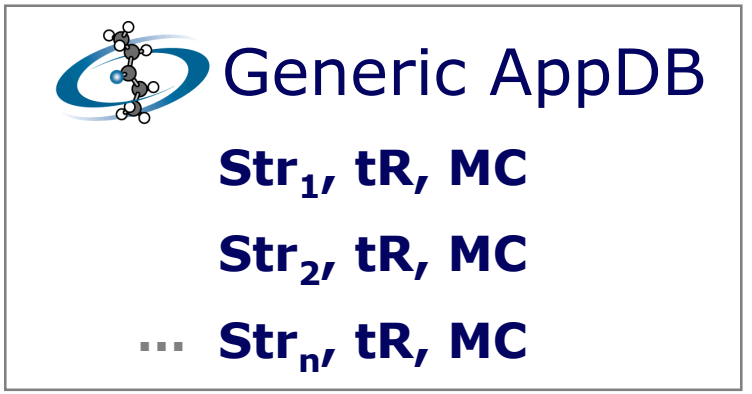


Automated selection of methods
The Process





Automated selection of methods
The Process



Instrument Control Software

ChromGenius

Grouped Structures

Prediction Mode

Prediction Eqn is applied to new structures; best method & ? outputted

Prediction Eqn

“A Federation of Local Models”

"A Federation of Local Models"

- Training sets are unique for each compound, and for each method
- The number of potential models is impressive; assuming 32 compound training set:
  - N choose 32 (N is the number of experimental compounds available)
  - ~(2000)^32 different models!

An Example

- Structure is input to the system
- The 32 most similar structures are chosen (Dice Coefficient)
- PhysChem is predicted for these compounds
- Fit prediction equation
- Predicted PhysChem for new compounds is plugged into equation
- Predicted retention time for compound is output

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

An Early Application

Collaboration with Specs

- 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:

- 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:

- 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

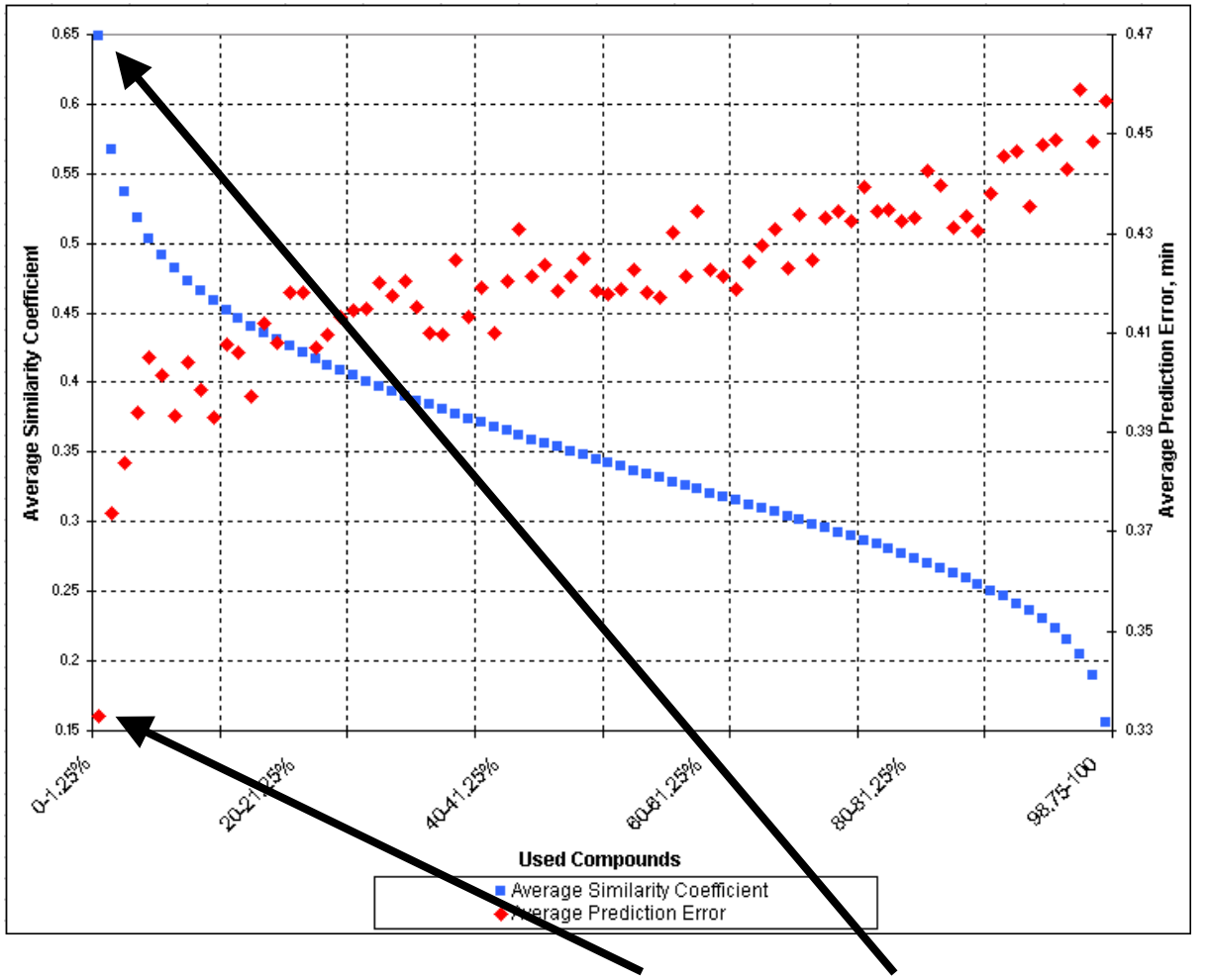
ChromGenius Calculation Settings

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

Structure Similarity and Accuracy – 2006 compounds

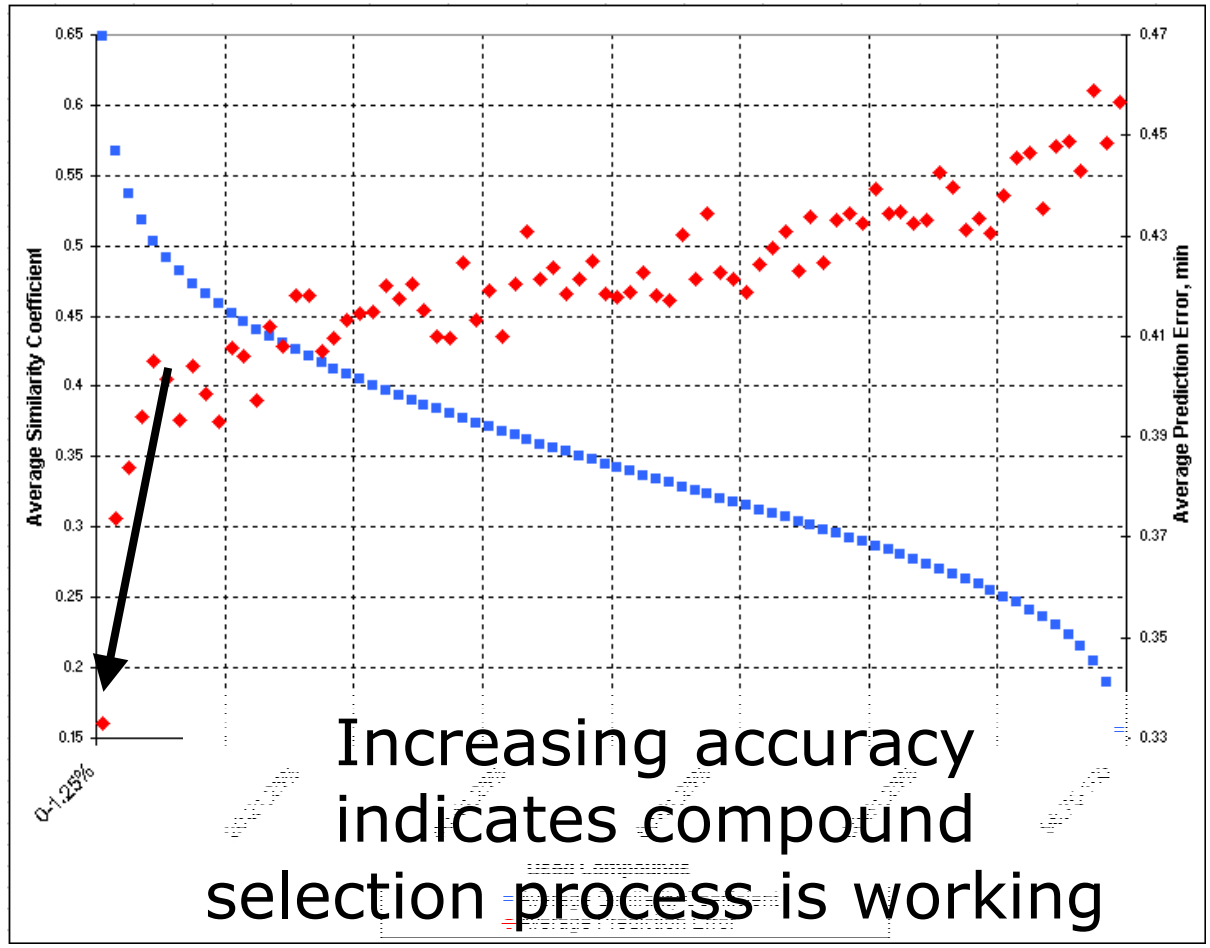
Each red point is the average of 2006 predictions

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

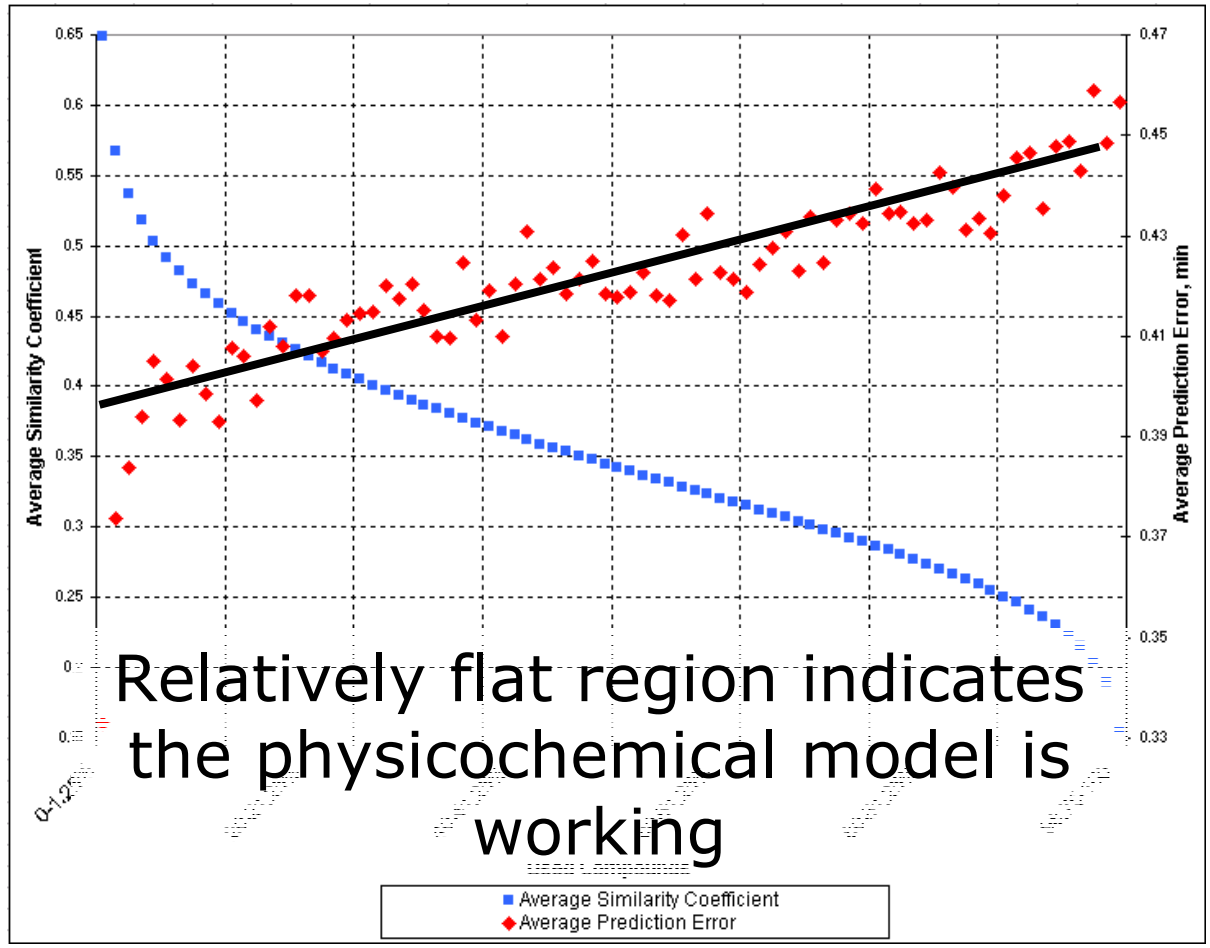


These points represent CG working under normal conditions

Structure Similarity and Accuracy – 2006 compounds



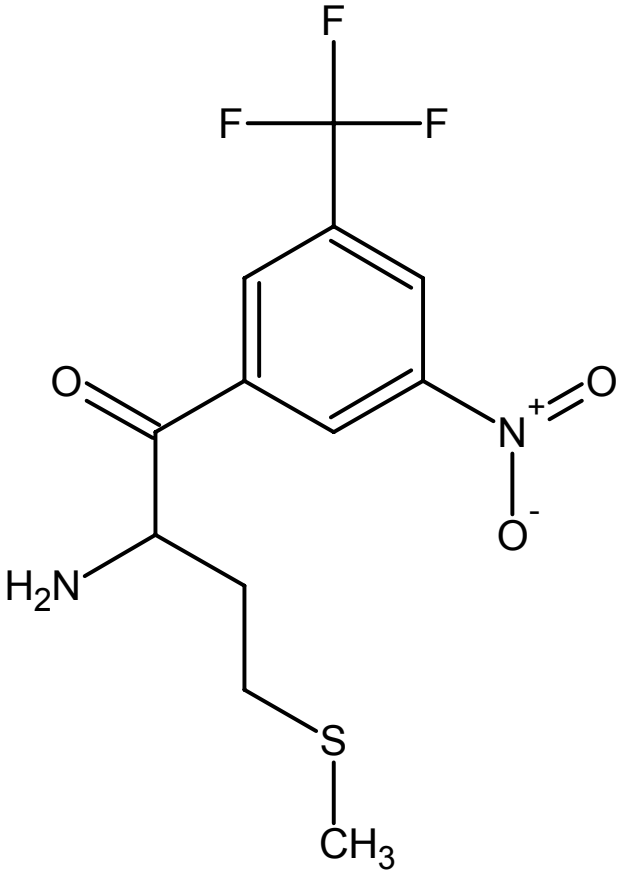
Structure Similarity and Accuracy – 2006 compounds



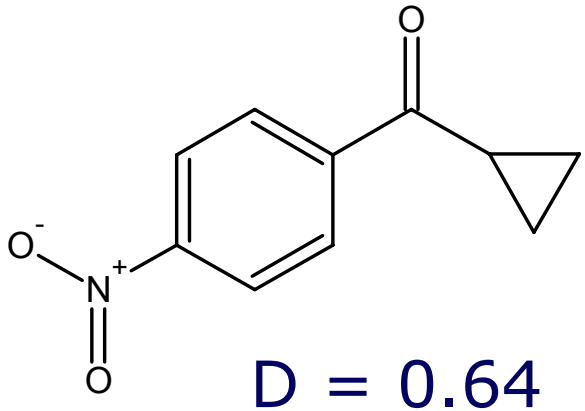
Relatively flat region indicates the physicochemical model is working

How Similar is Dice Coeff. = 0.65?

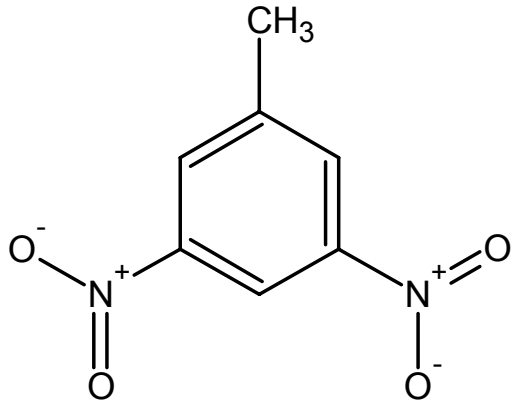
Target molecule



Hits



D = 0.64



D = 0.66

"How Accurate are the Predictions?

- More than accurate enough for method selection.
- Specs Method 1:
  - 7 minute run; fast gradient with flow rate change
  - 2007 compound test (included with CG)
  - VERY diverse compounds
  - 20 compounds: error more than 40%
  - 4 compounds: error more than 60%

Even with this very challenging dataset, 99.8 % of compounds were of sufficient accuracy for method selection.

Generic Methods and Structure Verification

- Structures *cannot* be determined by retention times in chrom

But…

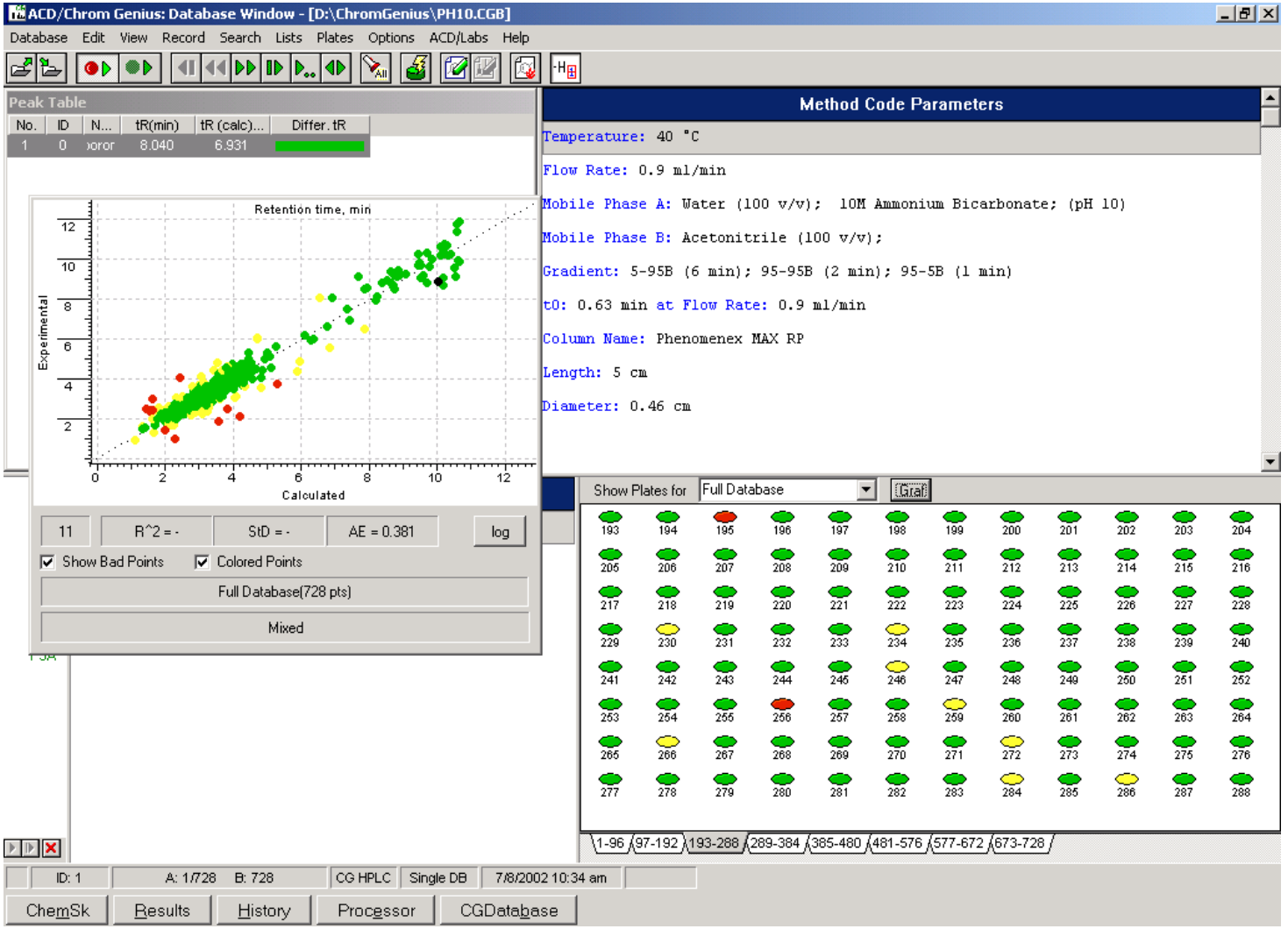
- t_R *can* be used to help confirm chemical structures.

ChromGenius can operate as a high-confidence predictor of retention time for verifying chemical structures, supplementing MW data.

The Verification Process

1. Train the system with the new retention times
2. ChromGenius performs a “leave one out” study on all compounds
3. Predicted and experimental are compared
4. Incorrect compounds *may* have different retention times: try HNMR?

ChromGenius and Structure Verification



Spectroscopy • Chromatography • PhysChem • Naming • Drawing and Databasing • Enterprise Solutions

A Preliminary Test of Structure Verification

Celltech Method 1

- Leave one out study of 654 data points
- 7 questionable retention time predictions
  - 3 correct
  - 2 incorrect structures
  - 2 typos

…predicted retention times *can* be used to assist with structure verification.

The Future

- Wider retention time spread
- Greater examination of Specs Method 2
- 130,000 (!) archived retention times for Specs Method 1
- SFC, GC and Normal Phase examples
- Semiprep scale prediction?
- More!

Acknowledgements

- Eduard Kolovanov; ACD/Labs
- Vadim Tashlitsky; ACD/Labs
- The ACD/Labs development team
- Robin "PalmGuy" Martin; ACD/Labs
- Mel Euerby and Patrik Petersson; AZ
- Richard Taylor, Celltech
- Andy de Laet, Specs
- Wibo van Scheppingen, Specs
- Peter Maas, Specs

Downloads from ACD/Labs:

- This presentation (in one week)
- Posters from HPLC2003 (in one week)
  - MAP peak matching
  - Prediction of Retention Times on Mixed Retention Mode Mobile Phases
- Freeware (now):
  - ACD/Column Selector for Palm
    - www.acdlabs.com/columnselector
  - ACD/ChemSketch 5.0
    - www.acdlabs.com/downloads