

Specifying and Searching for Ring System and Functional Group Environments Using Concepts of Molecular Equivalence Analysis

Mark Johnson

Pannanugget Consulting

**Cutting Edge Tool and Concepts in
Molecular Equivalence Analysis**

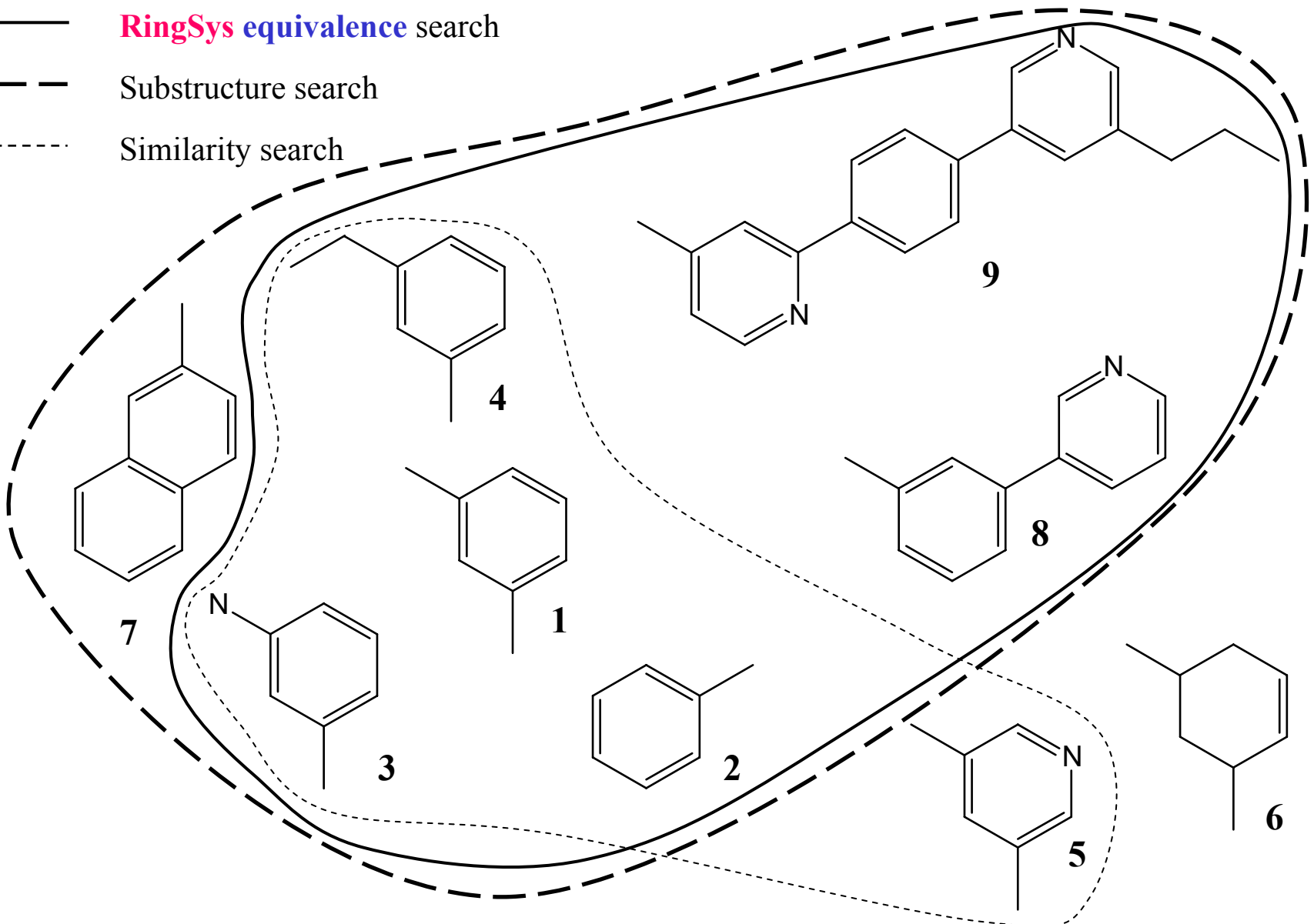
mark@pannanugget.com

Which Compounds have a Ring System Environment Like 1,3-Dimethylbenzene?

— RingSys equivalence search

- - - Substructure search

· · · Similarity search



Executing This Ring-System Equivalence Search in Spotfire

The Structures

Structure Viewer

Structures - Marked Records

Database: BenzeneRngSysFile.db

Ring System	Frequency
30T5	1
515K	1
A14R	1
FY5	3
FY5;30T5	1
FY5;30T5;30T5	1

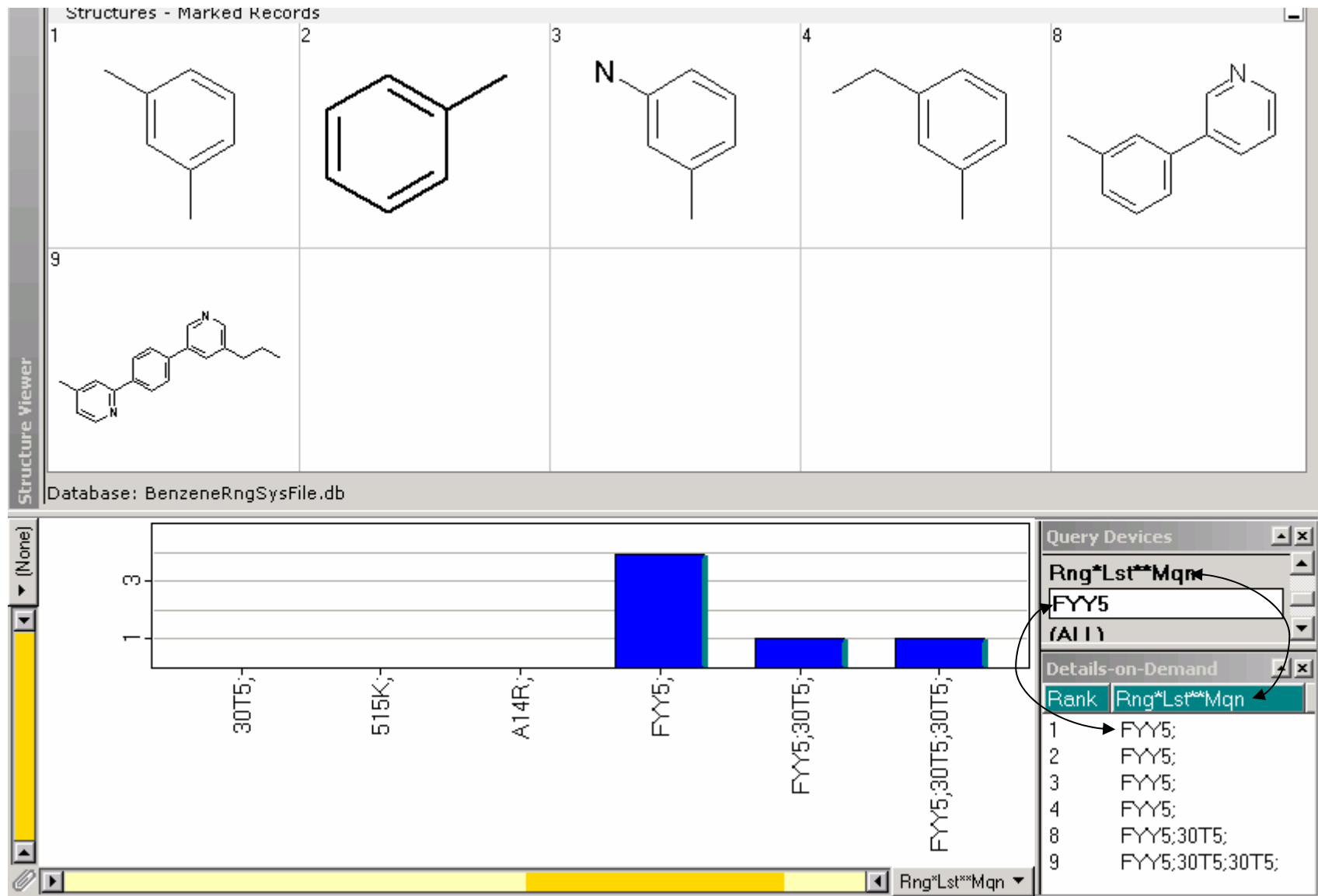
Details-on-Demand

Rank	Rng*Lst**Mqn
1	FY5;
2	FY5;
3	FY5;
4	FY5;
5	30T5;
6	A14R;
7	515K;
8	FY5;30T5;
9	FY5;30T5;30T5;

Rng*Lst**Mqn

Executing This Ring-System Equivalence Search in Spotfire

The 1,3-Dimethylbenzene Query



Formalizing This Approach to Ring-System Equivalence Searching

Develop an algorithm that assigns each ring-system a unique *a priori* identifier.¹

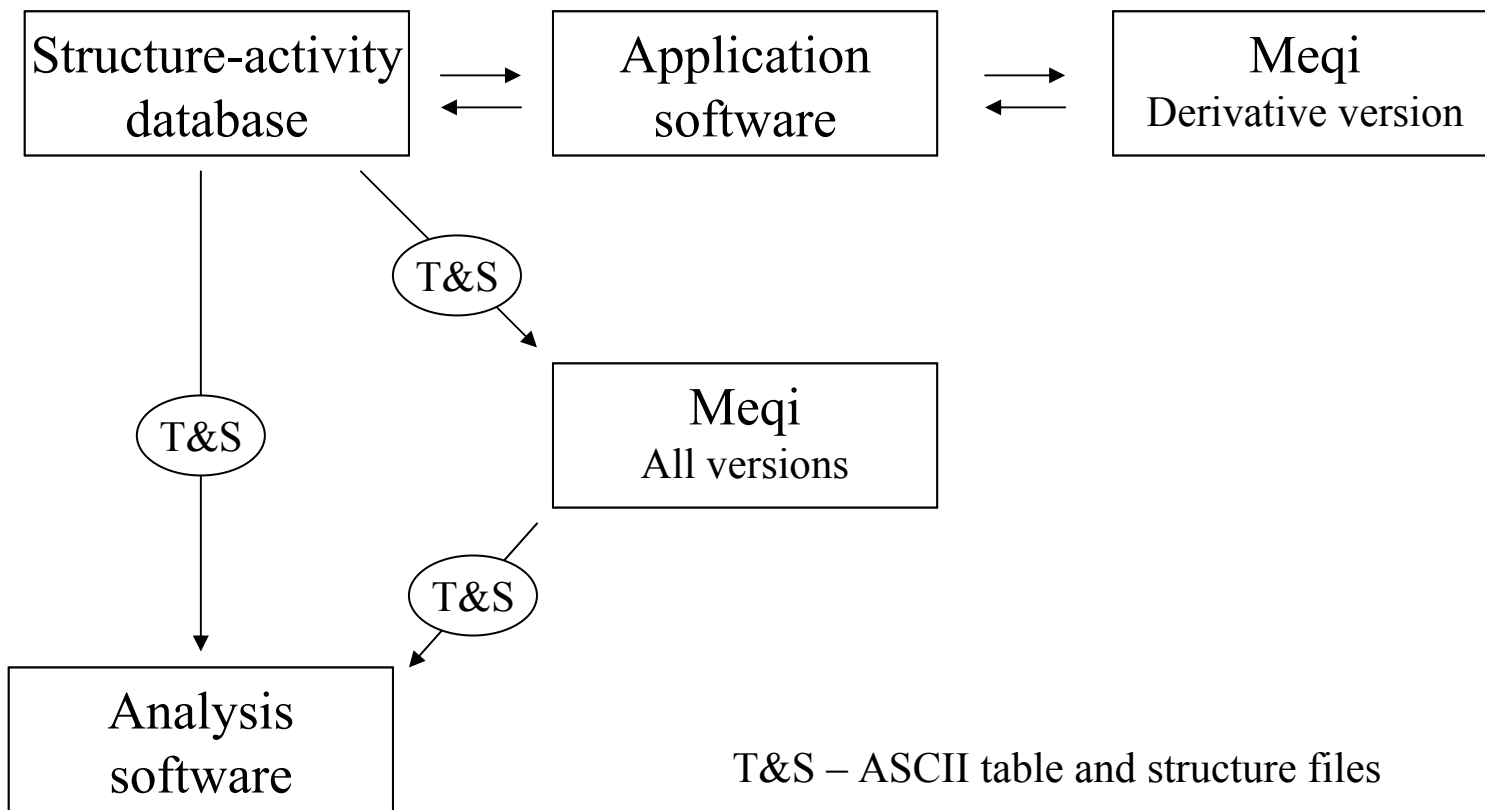
Populate a “ring-system” field with the identifiers of the ring systems for each structure in your database.

Look up or compute the identifier for the query ring system.

Perform a substring search of the ring-system field using the identifier of the query ring system.

¹ Nilakantan et al., JCICS, 30(1990)65 – most chemical graphs
Xu & Johnson, JCICS, 41(2001)181 – labeled pseudographs

Meqi is a command-line driven program that easily interfaces with other software.



MeqiLite operates on up to 15K structures and is available as a free download from www.pannanugget.com.

Meqi is a template-based programming language!

The program generating the Rng*Lst**Mqn index

	A	B	C	D	E	F	G	H	I	J	K	L
1	Meqi_Command_Template Copyright © 2003 by Pannanugget Consulting, L.L.C.											
2	Global_parameters Save As Meqi.csv in the directory from which Meqi is run.											
3	Structure file name	Registry number field	Input table	Output table	Graph output file	Subgraph output file						
4	BenzeneRngSysFile.sdf											
5	Printing frequency	Asterisk replacement	Registry column	Sort Commands?	Maximum Count	Maximum Count						
6	Debug trail	Final graphs	Embedded graphs	Omit header 1								
7	Graph_parameters											
8	Original graph	Chemical graph	Proton donating	Hydrogen reduced	Functional groups - bond separated	Largest component	Min substructure size					
9	Graph prefix	Range of graphs	Ring-augmenting edges									
10	Functional groups - atom separated											
11												
12	Global_variables											
13	Rank	RegistryName	Comment1	Comment2	MaxLayer	Run						
14	Combinatorial_parameters											
15	Combinatorial prefix	Vertex-counting list	Edge-counting list	Meqnum size	Ordering variable	Suppress Suite?						
16												
17	Vertex-labeling list											
18	Edge-labeling list											
19	Combinatorial_table The selected options will generate lists of ring-system meqnums											
20	Subgraphs	Embedding	Components	Attachment type	Topology	Variables						
21	Entire (*)	Unembed (*)	Group (*)	RingSys (*)	Unchanged (*)	Meqnum (Mqn)						
22	CyclicSystem (Cyc)	Layer (Lay)	List (Lst)	SideCh (Sc)	Reduced (Red)	Component (Cct)						
23	RingSystems (Rng)	Linkage (Lnk)	First (Fst)	Template (Tp)	GrpTop (Gpt)	Vertex (Vct)						
24	SideChains (Sch)	Embed (Emb)		AlphaExtend (Ae)	FgdTop (Fgt)	VertexTypes						
25	FunctionalGroups (Fnc)			FuncGrp (Fg)	BkgTop (Bgt)	EdgeTypes (E#X)						
26				AlphaAttach (Aa)	BnrTop (Bnt)	SingleTypes						

Meqi Generated This File Which Was Read into Spotfire.

	A	B
1	Rank	Rng*Lst**Mqn
2	INT	STRING
3	1	FYY5;
4	2	FYY5;
5	3	FYY5;
6	4	FYY5;
7	5	30T5;
8	6	LD7K;
9	7	515K;
10	8	FYY5;30T5;
11	9	FYY5;30T5;30T5
12		
13		

30T5 names to a **particular representation** of pyridine.

Setting up a ring-system equivalence search capability
with Meqi.

Check the indicated options.

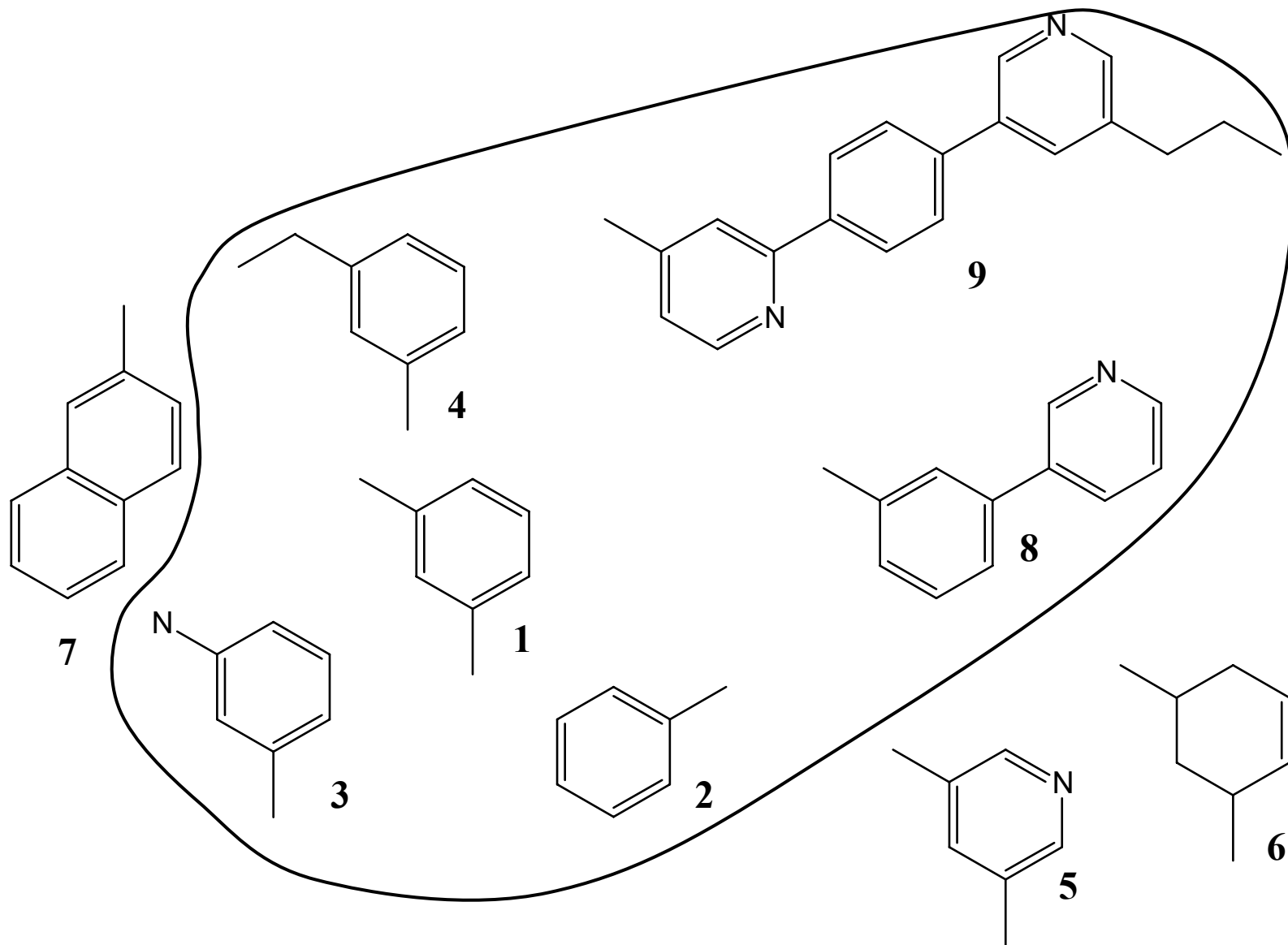
Run your program.

Enter the output table into your database.

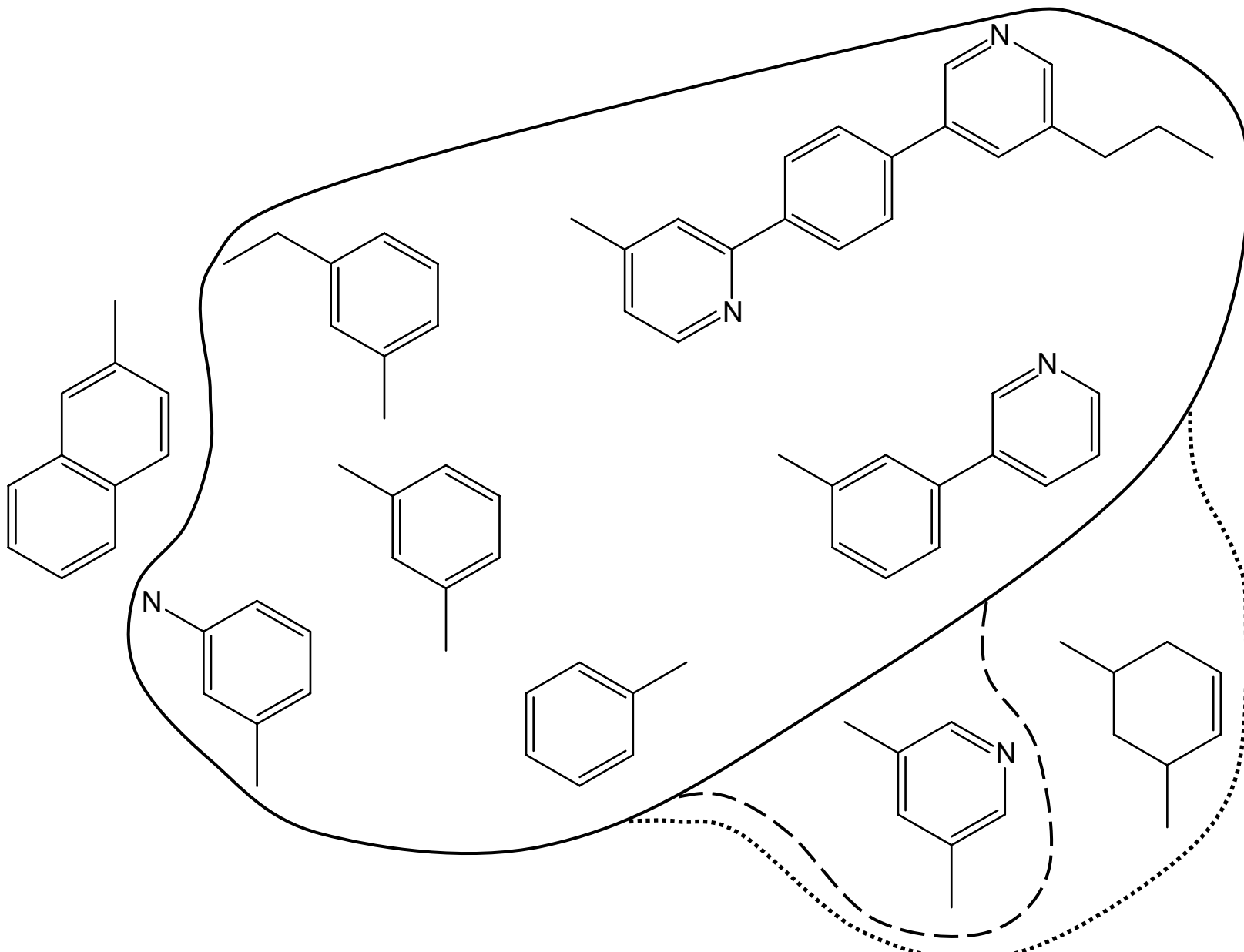
You now use your in-house substring search capability.

Where Do We Go From Here?

Addressing the Shape of the Hit Region



Increasing the Size of the Hit List



The Basic Approach

Different aspects of chemistry are addressed by changing the structure representation.

In molecular equivalence analysis we change the connection-table representation.

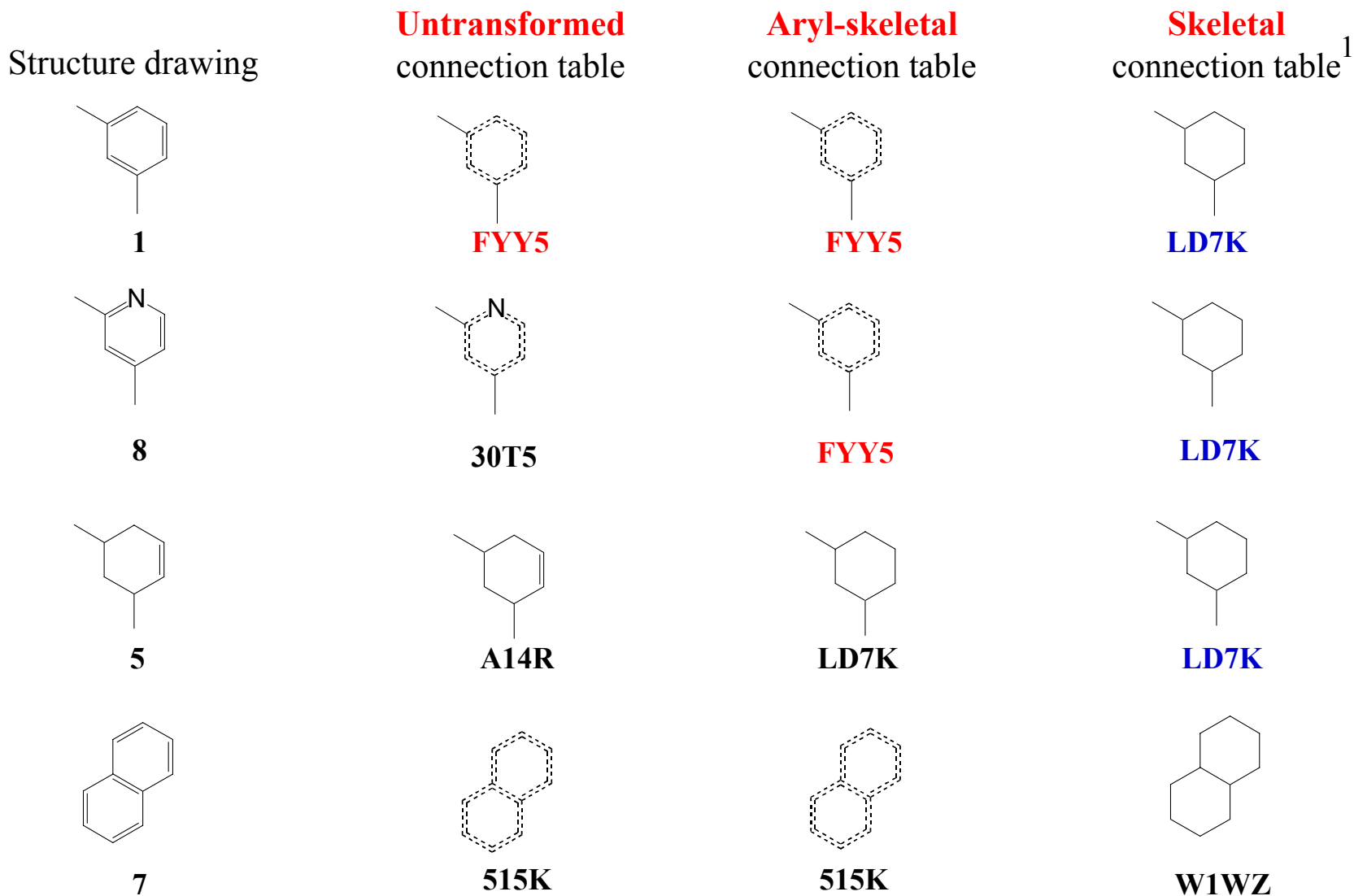
This **does not change** the equivalence features.

It creates a **new** molecular equivalence index.

It generally **changes** the meqnums.

Two Atom-Labeling and Bond-Typing Transformations

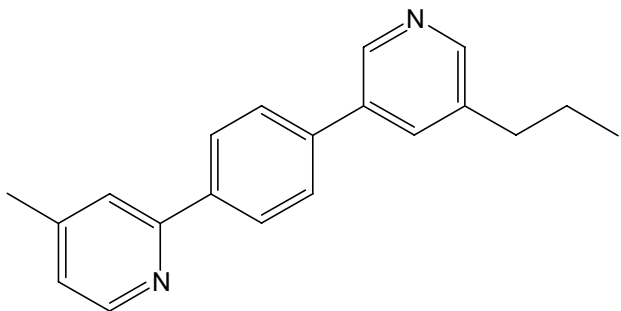
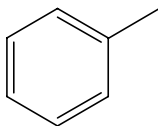
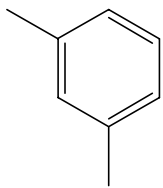
Using Hydrogen-Reduced Representations



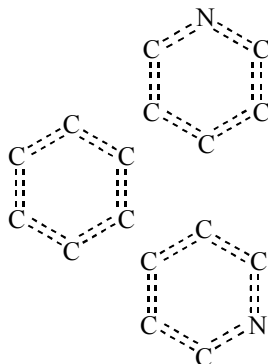
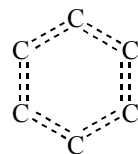
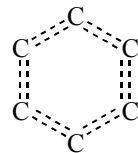
¹

A Preserving-Attachment-Points Transformation

Drawing



Labeled ring
systems



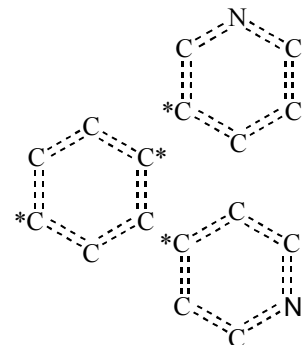
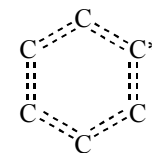
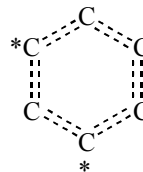
Original
meqnums

FYY5

FYY5

FYY5 30T5 30T5

Preserving
attachment



New
meqnums

DJMD

2DCK

JTR7 6YNH 3VZR

A Preserving Linkage Transformation

Linkage suffix

Drawing

Labeled graph

Ring systems

Meqnums

0 – Isolated

1 – Side chain (SC)

2 – Bridge (B)

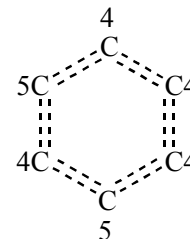
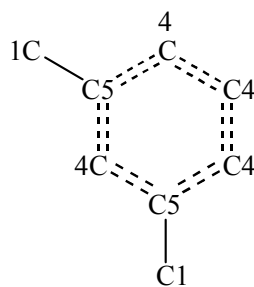
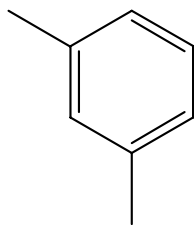
3 – SC + B

4 – Ring (R)

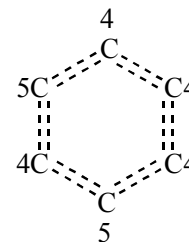
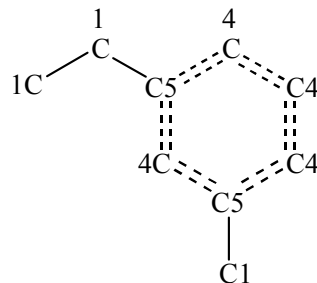
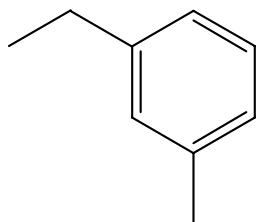
5 – R + SC

6 – R + B

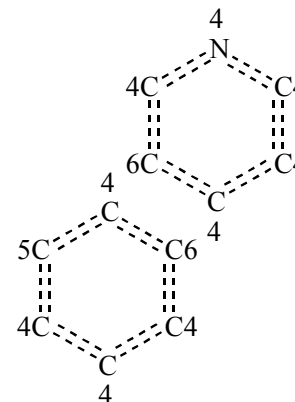
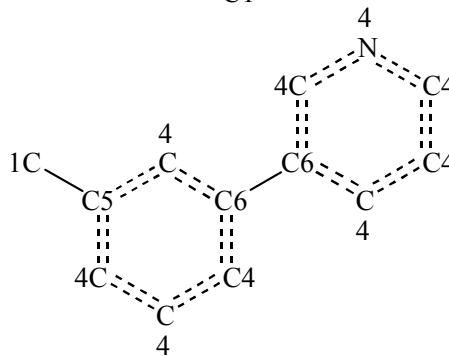
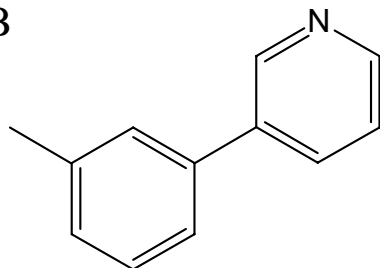
7 – R + SC + B



HN9K



HN9K

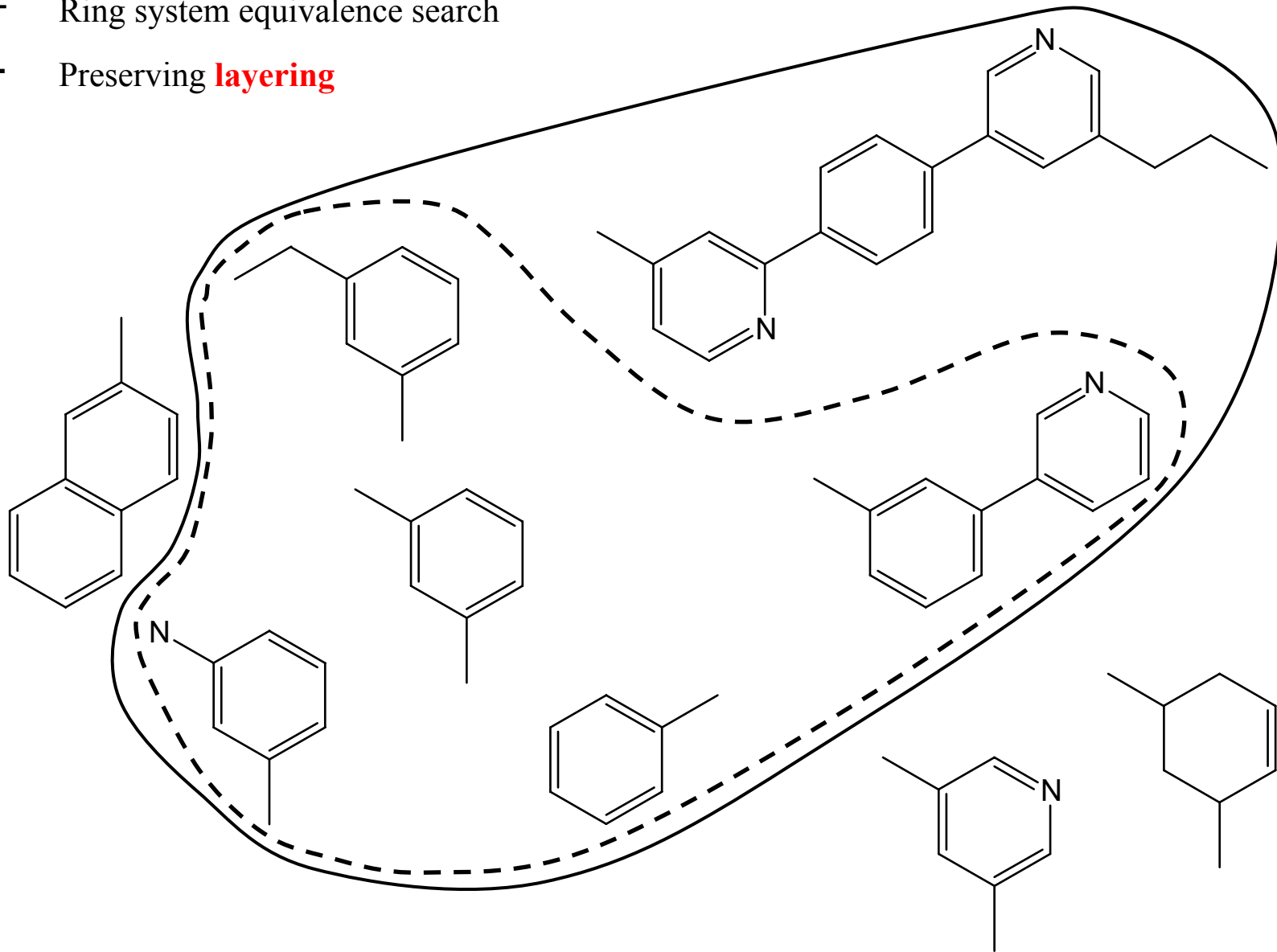


JJK7 GF0M

Preserving the Bond Layers

— Ring system equivalence search

- - - Preserving **layering**

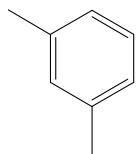


A Preserving Bond-Layers Transformation

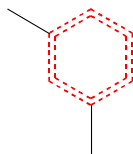
Bond layer
color code



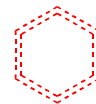
Drawing



Labeled graph

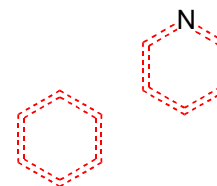
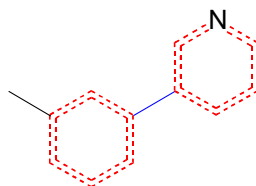
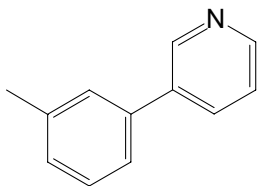


Ring systems

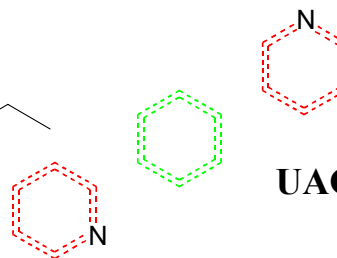
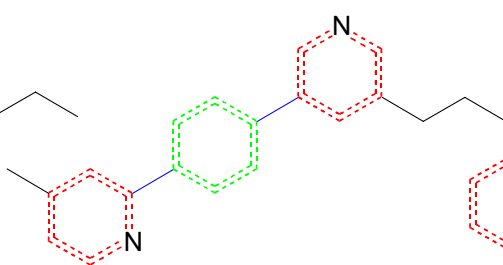
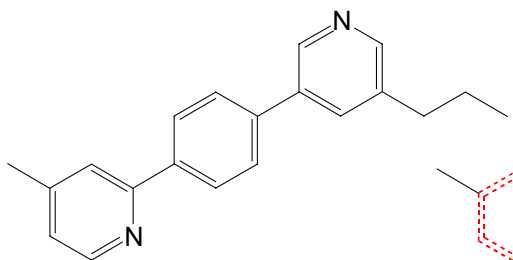


Meqnums

3GNZ



8ML1 3GNZ



UAQV 8ML1 8ML1

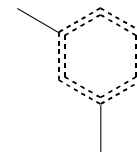
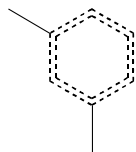
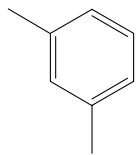
The Alpha-Extended **Extraction**

Drawing

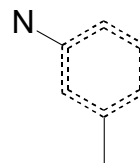
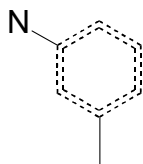
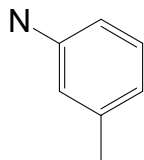
Labeled graph

Alpha-extended
ring systems

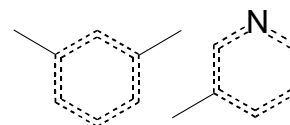
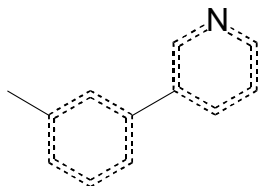
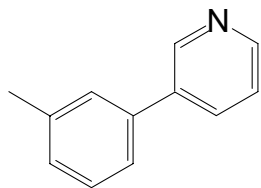
Meqnums



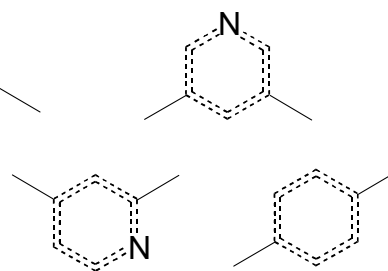
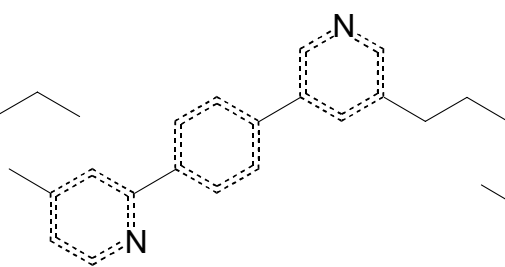
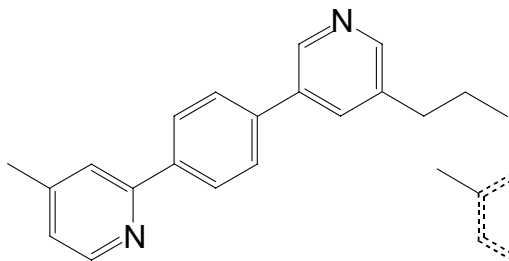
ZFEZ



4KEX



ZFEZ Q8LR



XFCD J0MR CEFK

Equivalence Searching Complements Substructure and Similarity Searching

	Compound number								
Search	1	2	3	4	5	6	7	8	9
Substructure	q	X	X	X			X	X	X
Similarity	q	X	X	X	X				
Ring system equivalence	q	X	X	X				X	X
<i>Using aryl skeletons</i>	q	X	X	X	X			X	X
<i>Using skeletons</i>	q	X	X	X	X	X		X	X
Preserving attachment	q		X	X				X	
Preserving linkage	q		X	X					
Preserving bond layer	q	X	X	X				X	
Preserving alpha atoms (AAs)	q			X				X	
Preserving AAs & attachment	q	No hits. Logical & does not imply \cap !							
Preserving AAs & attachment using <i>aryl skeletons</i>	q				X				

What Chemistry Can Be Ignored or Preserved

The atom labels

The bond types

Attachment points

Linkage

Layering

Alpha-atom augmentation

Any combination of these defines a molecular equivalence index for each **category** of equivalence features.

Categories of Molecular Equivalence Features

Complete Structure

Atom set¹

Bond set¹

Cyclic System²

Ring systems²

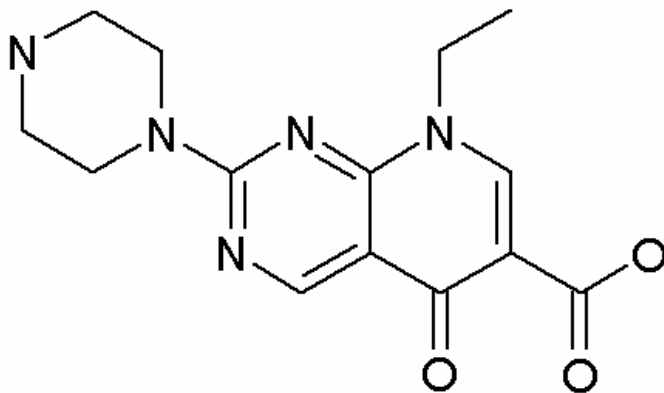
Bridge systems²

Side chains²

Functional groups³

¹ Adamson, Lynch & Town, J. Chem. Soc. (C) (1971) 3702; ² Bemis & Murcko, J. Med. Chem. 39(1996)2887; ³ Xu & Johnson, J. CICS, 42(2002)912;

Database Probing Using the Molecular Equivalence Principle



7613

The case of the antibacterial Pipemidic Acid - One of 200 structures selected at random from the Merck Index

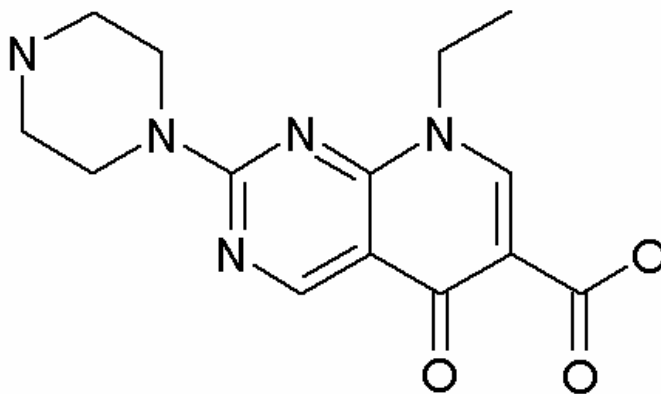
The Molecular Equivalence Principle

2D **property-critical** features **tend** to **co-occur** with molecular equivalence features.

There are only a handful of equivalence-feature **categories**.

Each category is represented by a small number of **features per structure**.

Database Probing Using the Molecular Equivalence Principle



7613

Note the ketone on an “outside” aromatic ring system with an alpha-attached side chain.

Using the Bond-Deleted Functional Groups (BDFGs)

First 10 of 29 hits using the “ketone” equivalence feature

Structure Viewer

Database: MrkSmp.DB

On a side chain

Antibacterial

235 898 1199 1795 2312

2536 2601 2602 2916 3175

01GM;01G... 2923;... 8GDB;4... CC5P;... EBGX;... PV21;6... TK97;Z... YFXT;44X...

Mer... N... B... Fnc*Lst**Mqn Fnc*LstAa*Mqn FncEmbLst**Mqn FncEmbLstAa*Mqn

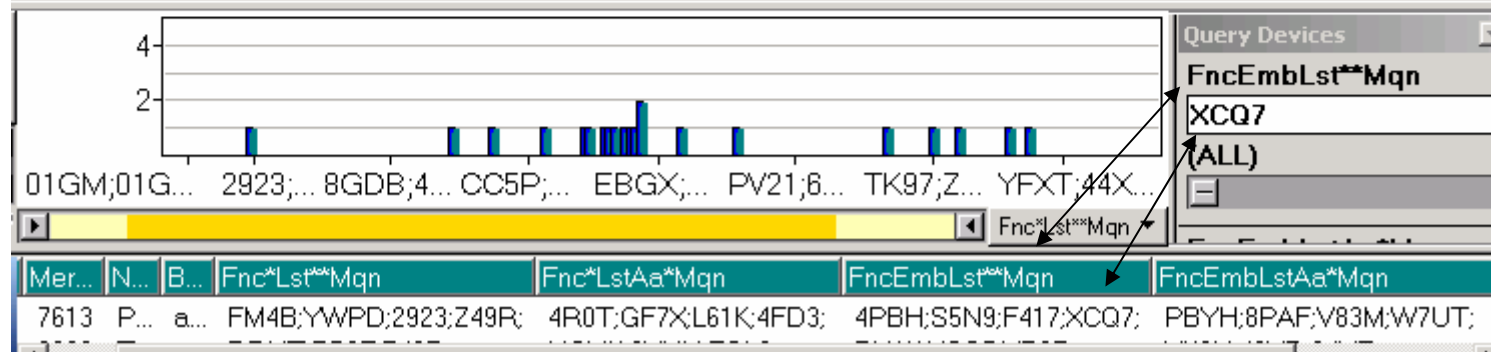
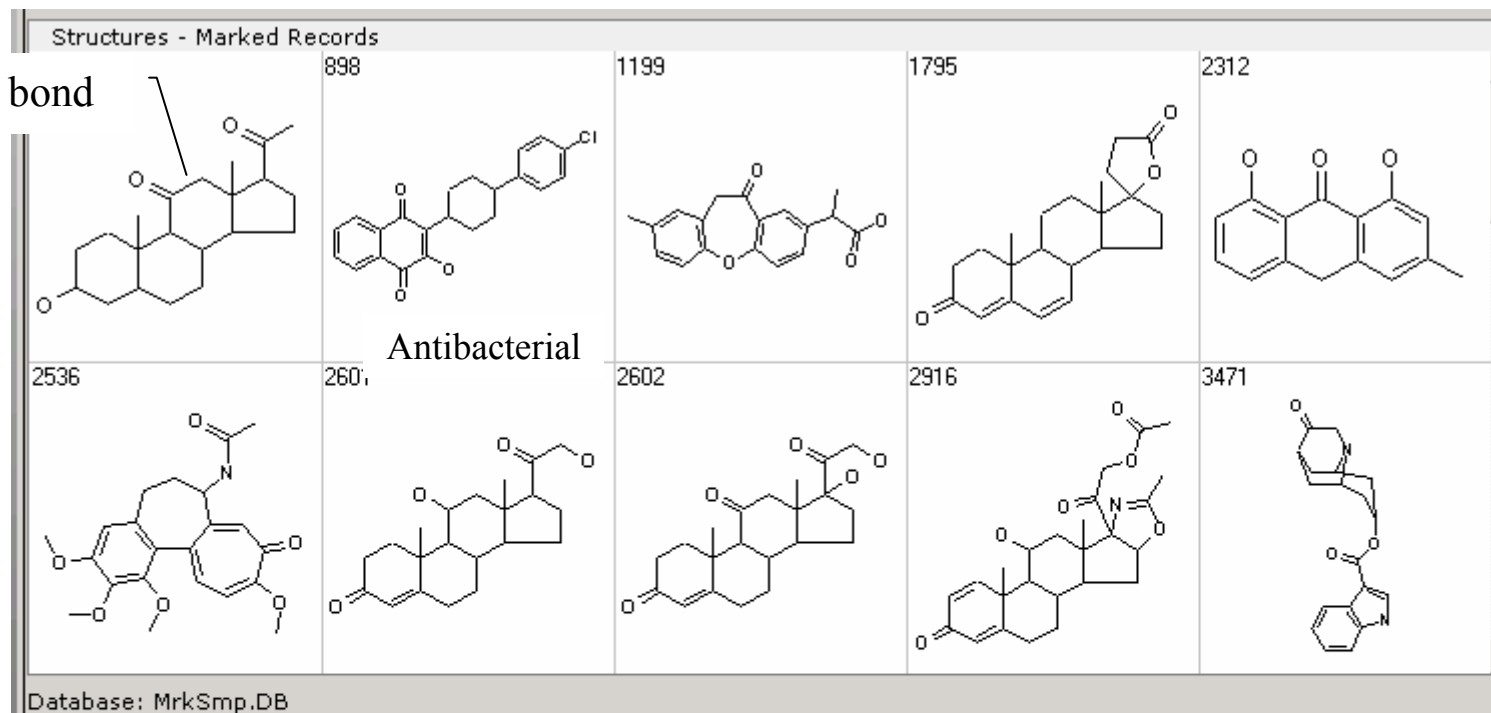
7613 P... a... FM4B;YWPD;2923;Z49R; 4R0T;GF7X;L61K;4FD3; 4PBH;S5N9;F417;XCQ7; PBYH;8PAF;V83M;W7UT;

30 out of 200 records visible (15.00 %), 30 marked

Using BDFGs that Preserve Linkage and Bond Layer

First 10 of 21 hits

Alkyl bond



Using Alpha-augmented BDFGs that Preserve Attachment

All six hits

Structures - Marked Records

898	2536	4919	4931	
7613	Not antibacterial		Not antibacterial	

Database: MrkSmp.DB

Query Devices

Fnc*LstAa*Mqn

4FD3

(ALL)

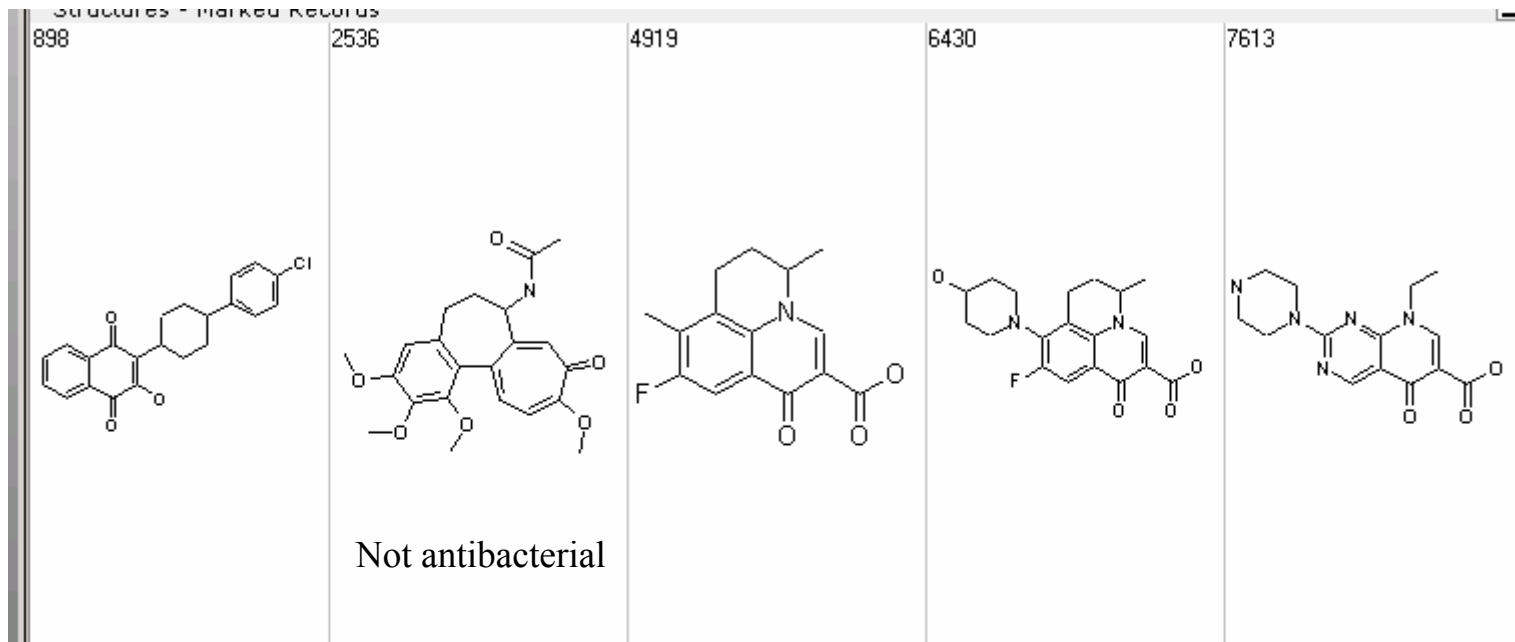
01GM;01G... 2923;... 8GDB;4... CC5P;... EBGX;... PV21;6... TK97;Z... YFXT;44X...

Fnc*Lst**Mqn

Mer...	N...	B...	Fnc*Lst**Mqn	Fnc*LstAa*Mqn	FncEmbLst**Mqn	FncEmbLstAa*Mqn
7613	P...	a...	FM4B;YWPD;2923;Z49R;	4R0T;GF7X;L61K;4FD3;	4PBH;S5N9;F417;XCQ7;	PBYH;8PAF;V83M;W7UT;

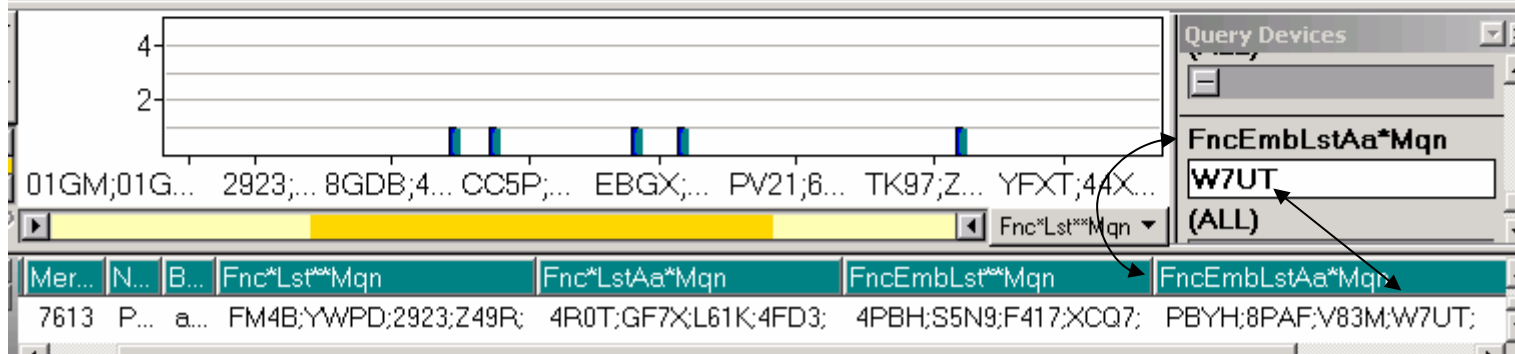
Using Alpha-augmented BDFGs that Preserve Attachment, Linkage and Bond Layer

All five hits



Not antibacterial

Database: MrkSmp.DB



A Whole New Area of 2D Stuff to Explore

A New Class of **Interpretable** Indices

Equivalence searching & database probing

Molecular equivalence analysis

And software to program the indices

Acknowledgements

Yong-jin Xu, Mic Lajiness and Gerry Maggiora

Long Relationship With the CADD Group at Kalamazoo

Upjohn – Pharmacia - Pfizer

Spotfire Inc. (Chris Ahlberg)

Molecular Design, Inc. (Doug Henry)