The Analysis and Presentation of Chemical Structural Information

Kerry G. Stanley – Science IP

ACS 229th National Meeting
San Diego, California

Topics for discussion

- Why structural analysis
- Challenges with structural analysis
- Using the R-group Analysis Table Tool in STN Express
- Mining the R-group analysis table
- Additional examples

Why Structural Analysis

- Test the landscape – what analogs have been disclosed
- Analyze a specific document – cpds described generically as A -x- B -y- C
- Identify chemistry associated with an organization
- Present compounds in an easy to compare table – create an SAR table

What are the challenges with structural analysis

- Can they be sorted?
  - MF
  - Properties
  - Ring Identifiers

E3 14140155 46.150.18/RID

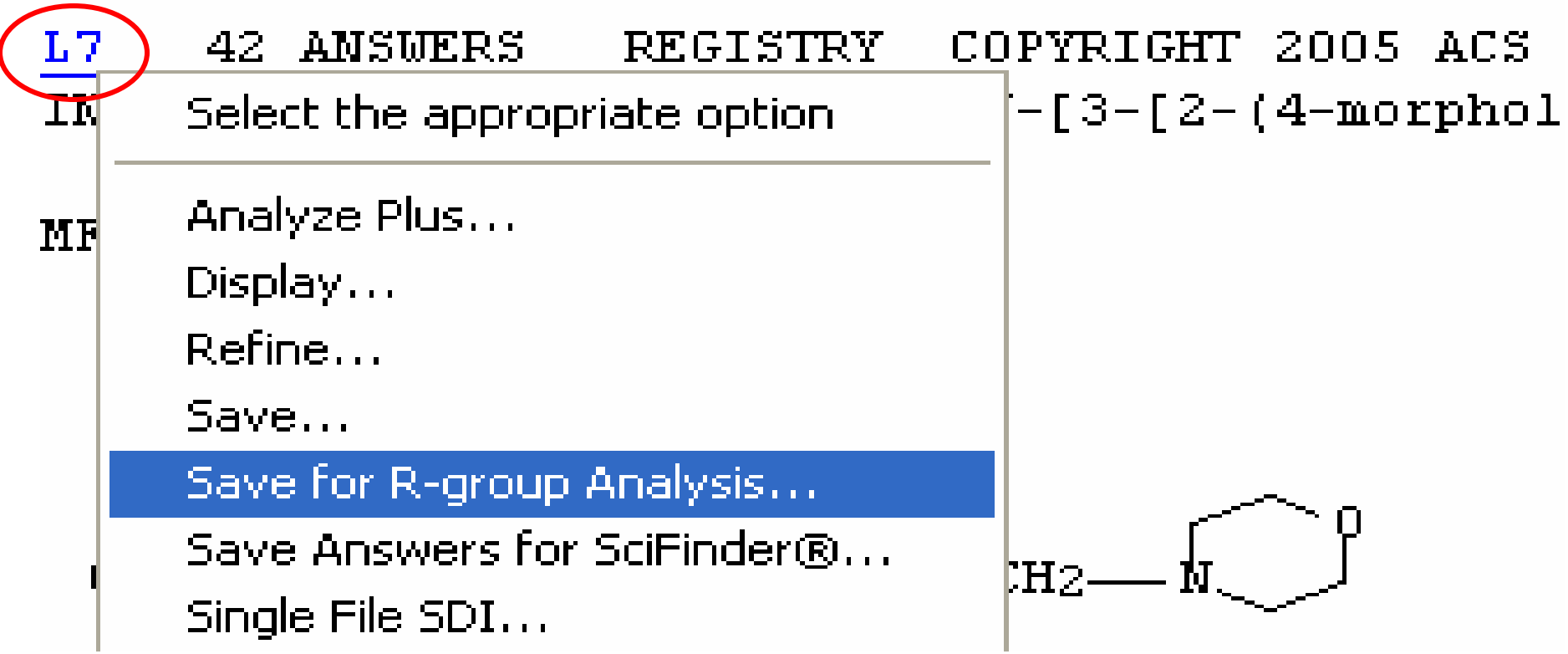
Benzene

- Can we readily identify the substances disclosed within a document
- Do we understand the nomenclature

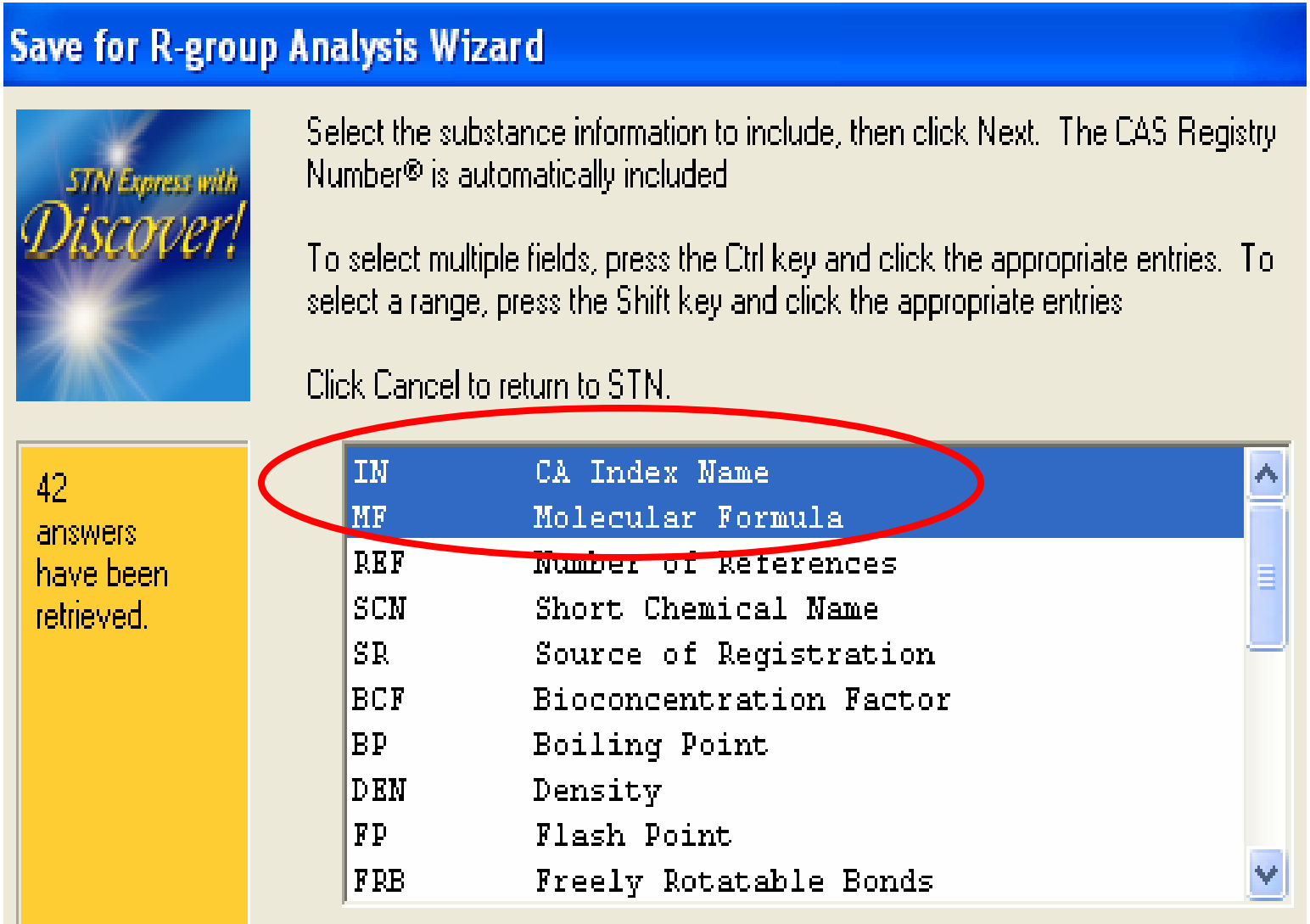
What are the basics of using STN Express R-group Analysis

- Create a REGISTRY File L-numbered set
- Save R-group data
- Generate a “common” structure
  - Software defined
  - User defined
- Creates columns for each R-group
- Export to MS Word or Excel

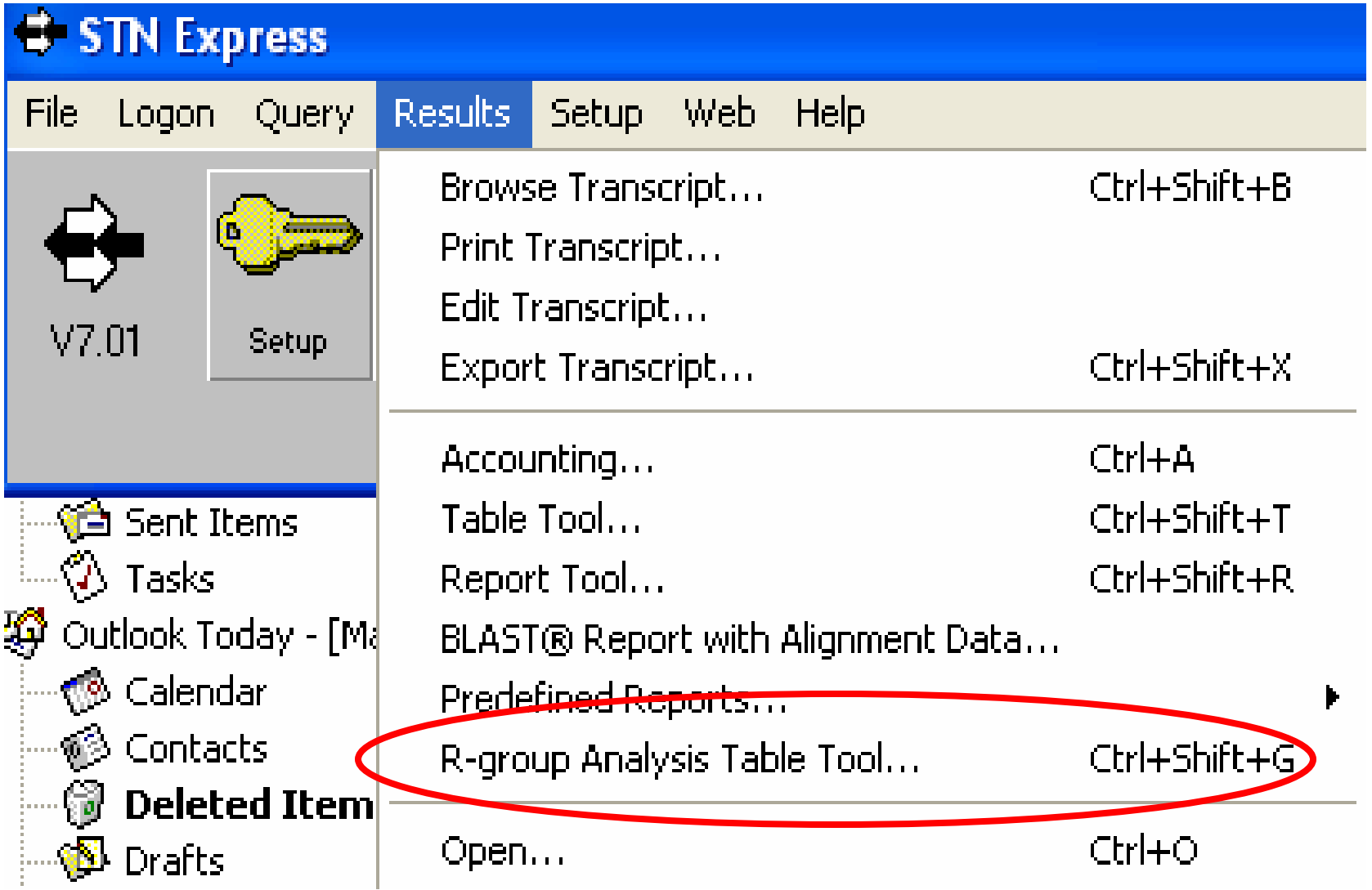
Save R-group Analysis data from any L-numbered set



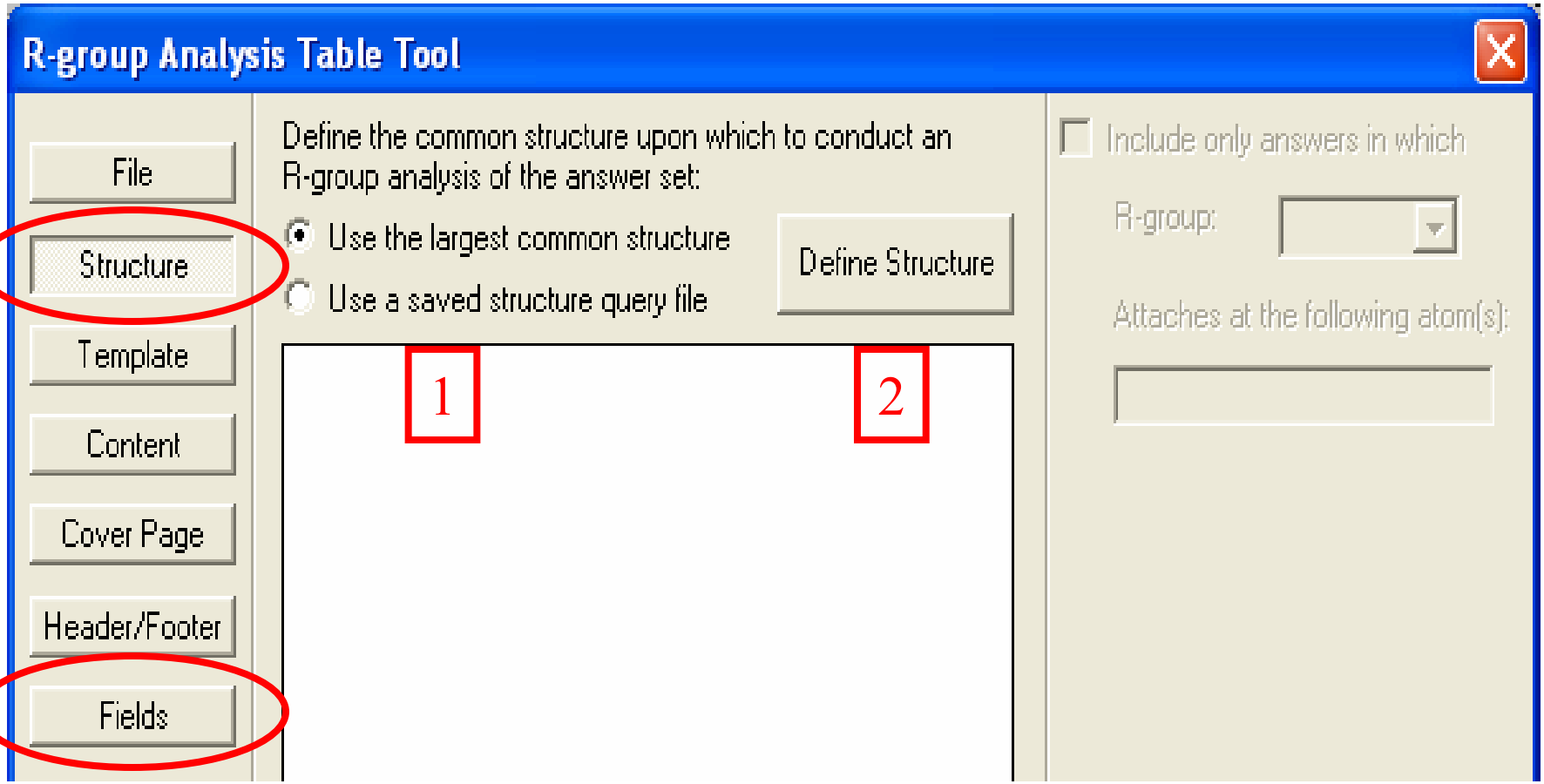
Choose additional data to include in the R-group table



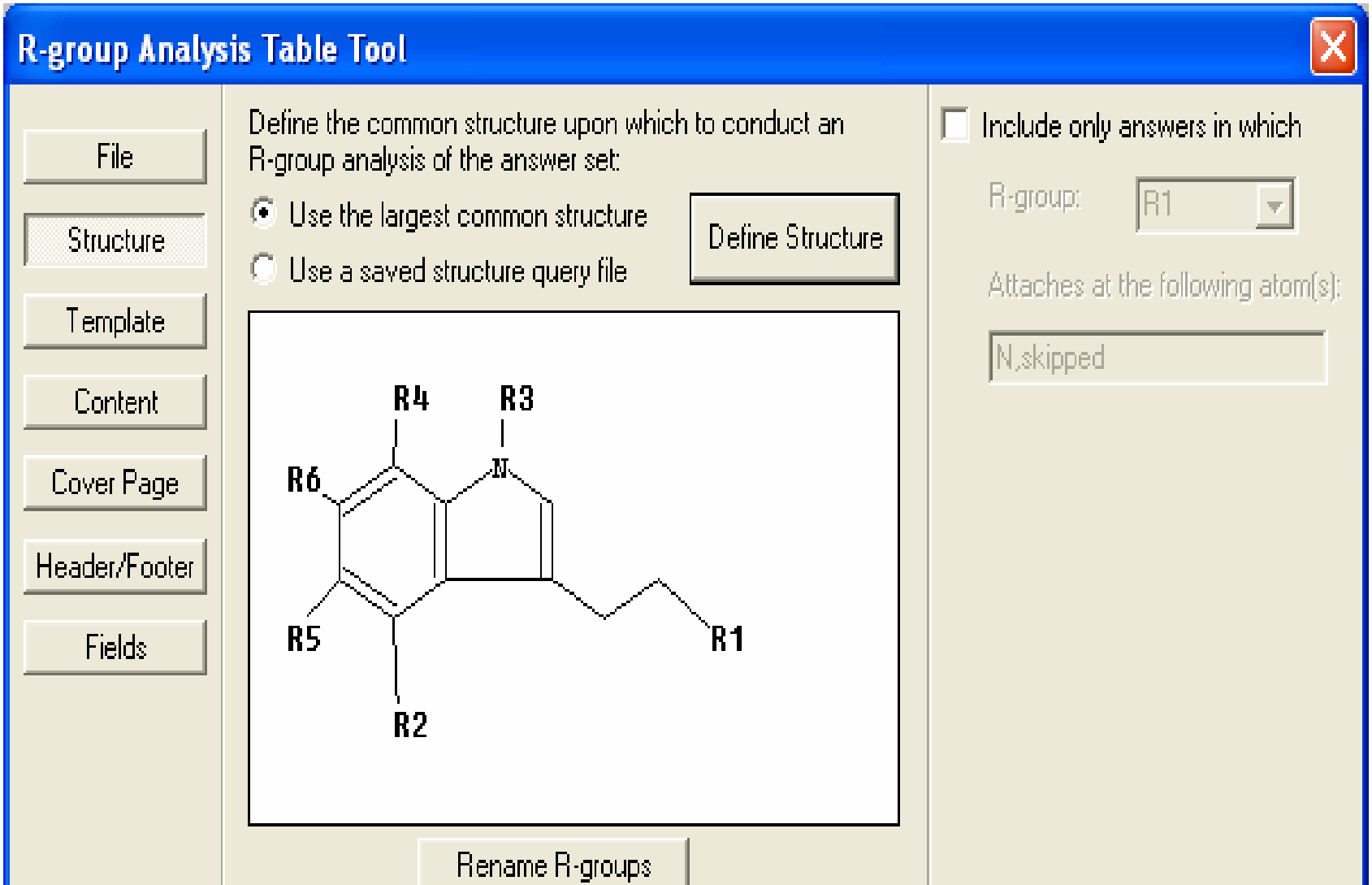
Select the R-group Analysis choice from the Results menu



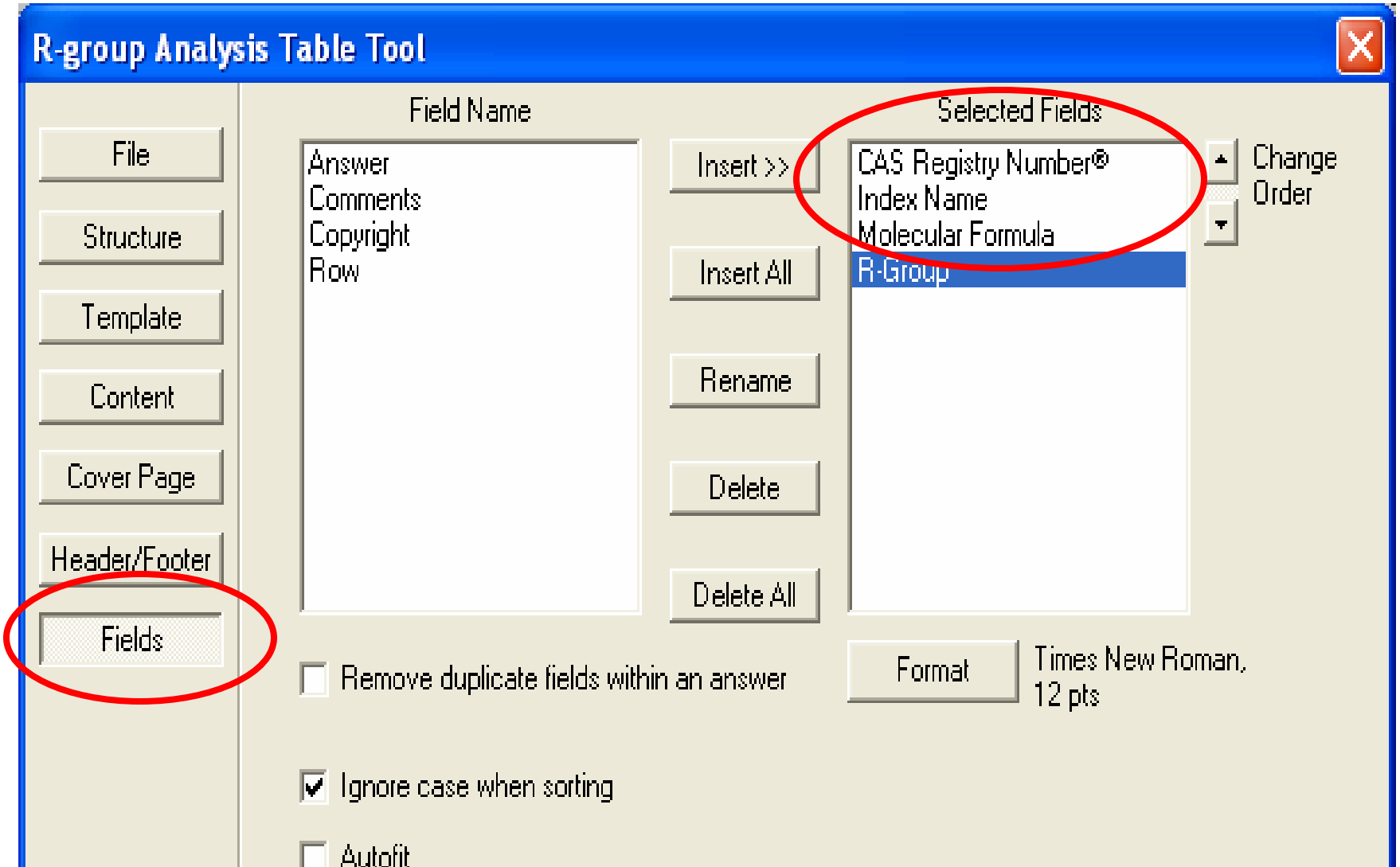
Follow the steps of the R-group Analysis wizard



On analysis of the data the software developed a common structure



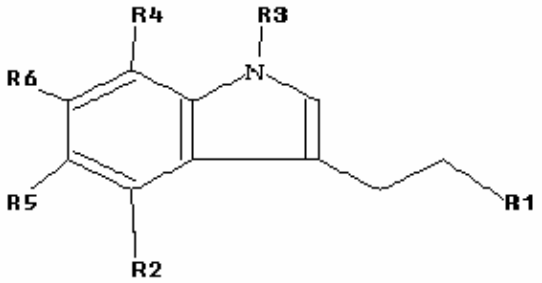
Additional data may be selected for the table

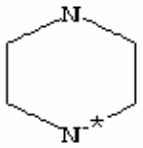


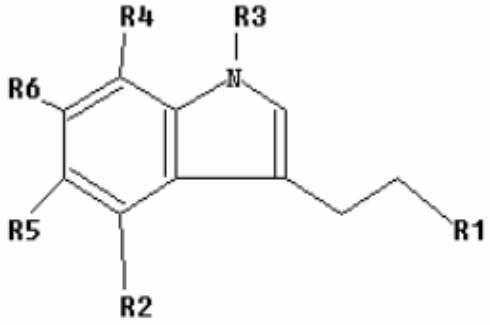
Express table may also be saved to RTF or XLS format

STN Online and Results - [Table Output - ACS Rgp analysis.tbl]

File Edit Preferences! Web Window Help



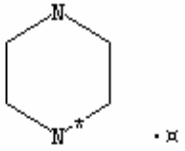
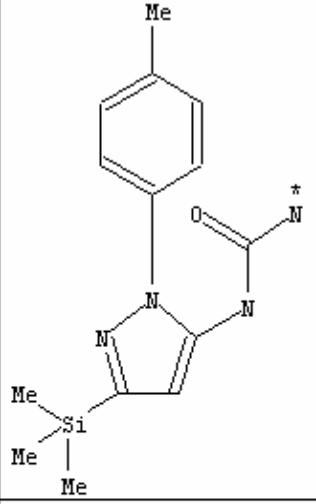
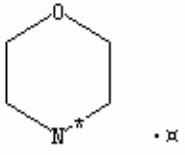
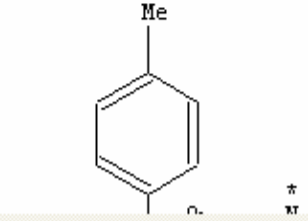
CAS Registry Number®	Index Name	Molecular Formula	R1	R2	R3	R4
813449-83-1 REGISTRY	Urea, N-[1-(4-methylphenyl)-3-(trimethylsilyl)-1H-pyrazol-5-yl]-N'-[3-[2-(1-piperazinyl)ethyl]-1H-indol-5-yl]-(9CI)	C28 H37 N7 O Si		H	H	H



¶

Project·XX·–·Indoles,·nitrogen·containing·heterocyclic·ring·substitution·on·3-ethyl·group¶

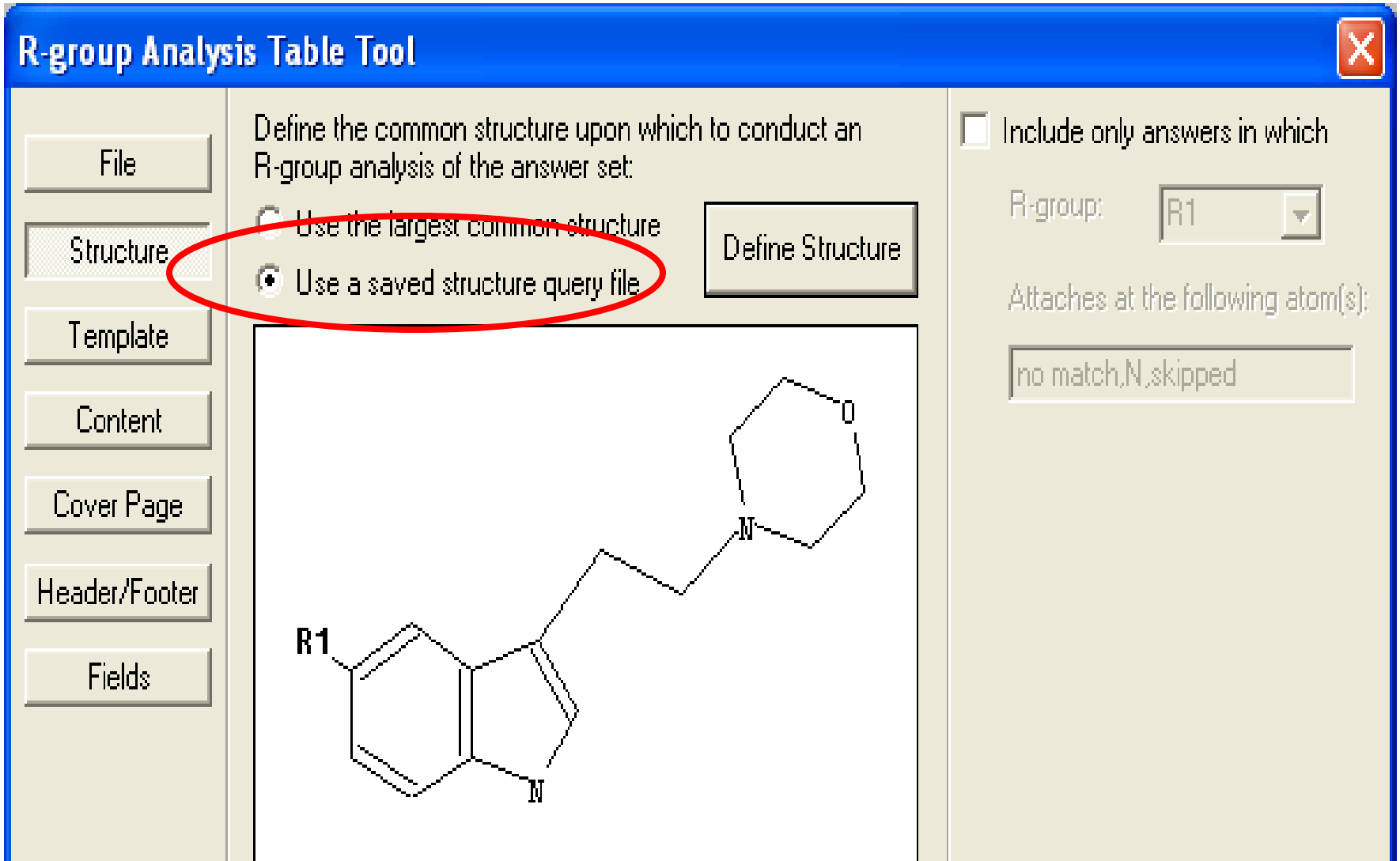
¶

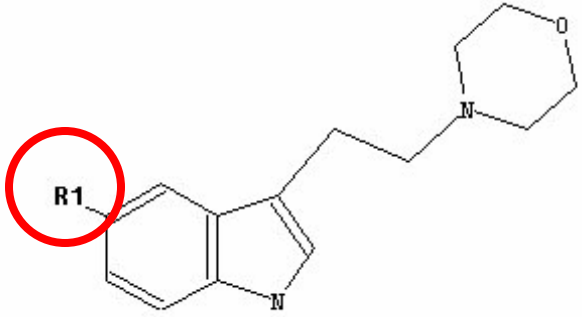
CAS· Registry· Number® ·¤	Index·Name ·¤	Molecular· Formula ·¤	R1 · – · heterocyclic· ring¤	R2 ·¤	R3 ·¤	R4 ·¤	R5 ·¤
813449-83-1···¤	Urea, ·N-[1-(4-methylphenyl)-3-(trimethylsilyl)-1H-pyrazol-5-yl]-N'-[3-[2-(1-piperazinyl)ethyl]-1H-indol-5-yl]- ·(9CI) ·¤	C28·H37·N7·O·Si·¤	 ·¤	H·¤	H·¤	H·¤	
813449-82-0···¤	Urea, ·N-[1-(4-methylphenyl)-3-(trimethylsilyl)-1H-pyrazol-5-yl]-N'-[3-[2-(4-morpholinyl)ethyl]-1H-indol-5-yl]- ·(9CI) ·¤	C28·H36·N6·O2·Si·¤	 ·¤	H·¤	H·¤	H·¤	

Can an R-group analysis table be mined?

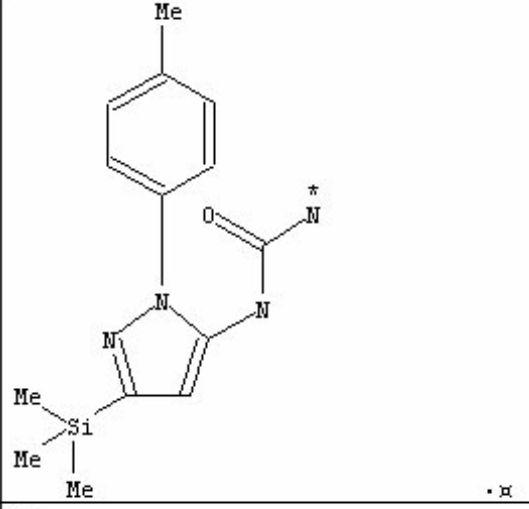
- Can build as many user drawn common structures as desired
  - Rebuild the table
  - Any non-matched compounds are retained as "no match"
- Export to Word or Excel and sort – all "no matches" together

Use any Express structure file to define the common structure





Project XX – Indoles, morpholinyl ring substitution on 3-ethyl group

CAS Registry Number®	Index Name	Molecular Formula	R1
813449-83-1	Urea, N-[1-(4-methylphenyl)-3-(trimethylsilyl)-1H-pyrazol-5-yl]-N'-[3-[2-(1-piperazinyl)ethyl]-1H-indol-5-yl]- (9CI)	C28 H37 N7 O Si	no match
813449-82-0	Urea, N-[1-(4-methylphenyl)-3-(trimethylsilyl)-1H-pyrazol-5-yl]-N'-[3-[2-(4-morpholinyl)ethyl]-1H-indol-5-yl]- (9CI)	C28 H36 N6 O2 Si	
753021-42-0	1H-Indol-5-amine, 3-[2-(4-morpholinyl)ethyl]- (9CI)	C14 H19 N3 O	N

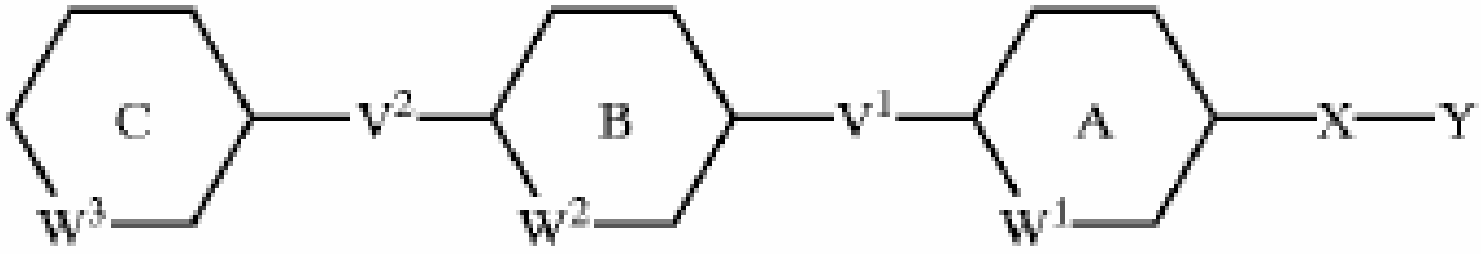
Why Structural Analysis

- Test the landscape – what analogs have been disclosed
- Analyze a specific document – cpds described generically as A -x- B -y- C
- Identify chemistry associated with an organization
- Present compounds in an easy to compare table – create an SAR table

What is claimed?

What is claimed is:

1. A compound of the formula (I):



I

wherein A ring, B ring and C ring are each independently optionally substituted aromatic carbocycle or optionally substituted 5- or 6-membered heterocycle which may fuse with benzene ring,

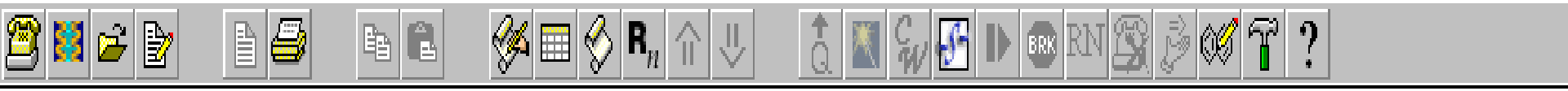
How do we identify the substances of this patent

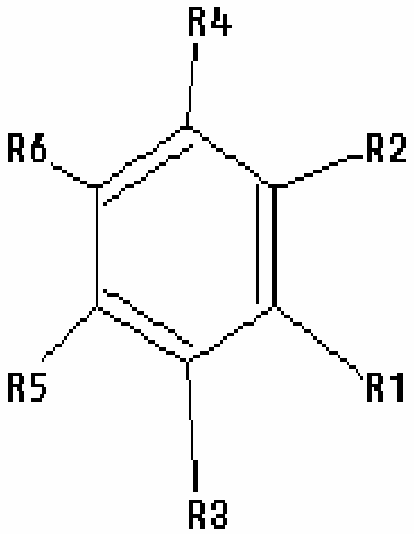
- Read the patent!!
  - 75 claims
  - Many, many tables
  - Hundreds of substances – yet few structures
- Thank the analyst!!
  - Identify the indexed substances
  - Analyze using R-group analysis

Identifying the substances associated with a document

- Retrieve the record for the document in CAplus
- TRANSFER all CAS RN's to the REGISTRY
- Limit the retrieval if desired
  - RANGE of CAS RN's
  - Structural or non-structural information
- R-group analysis

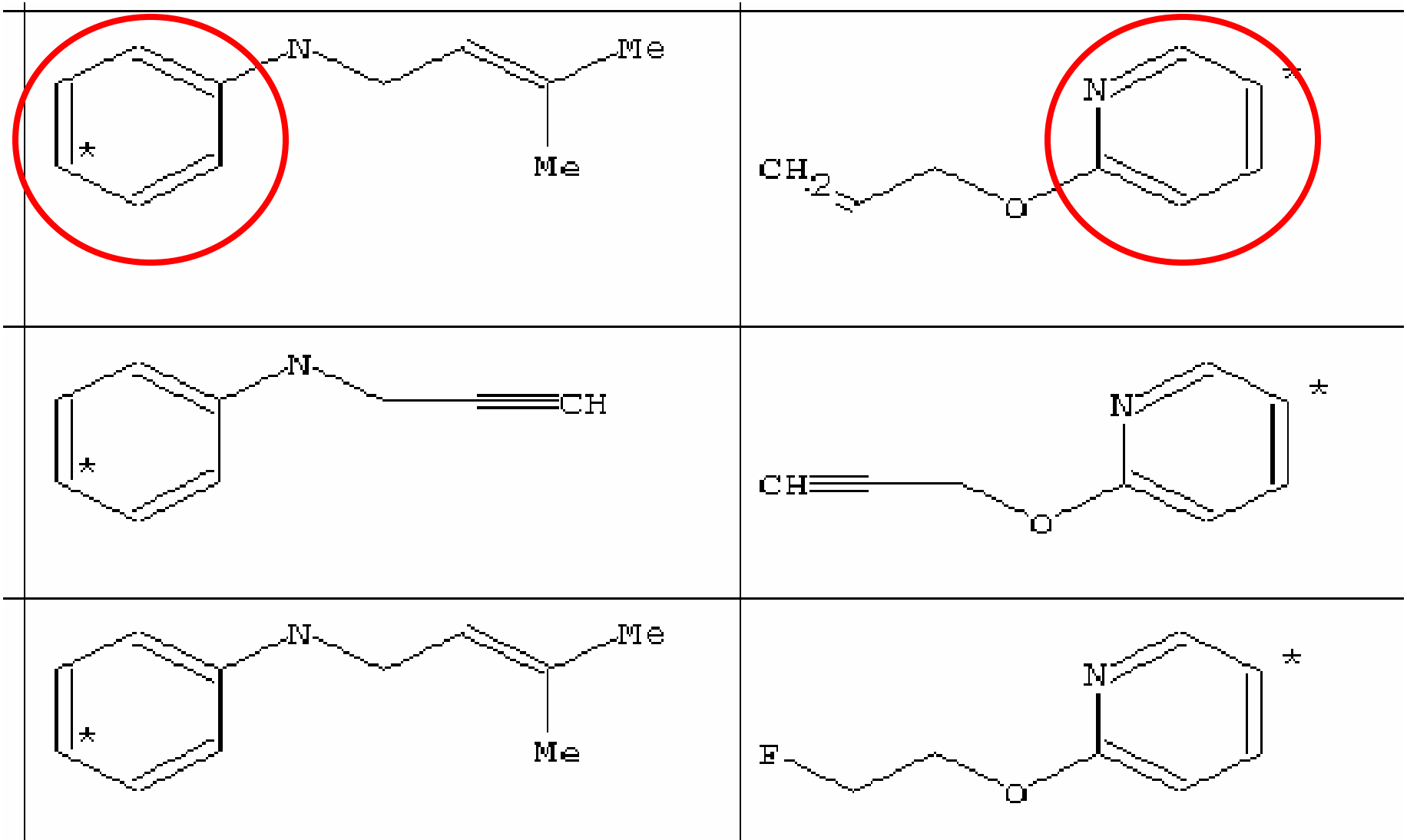
Common structure is benzene!!



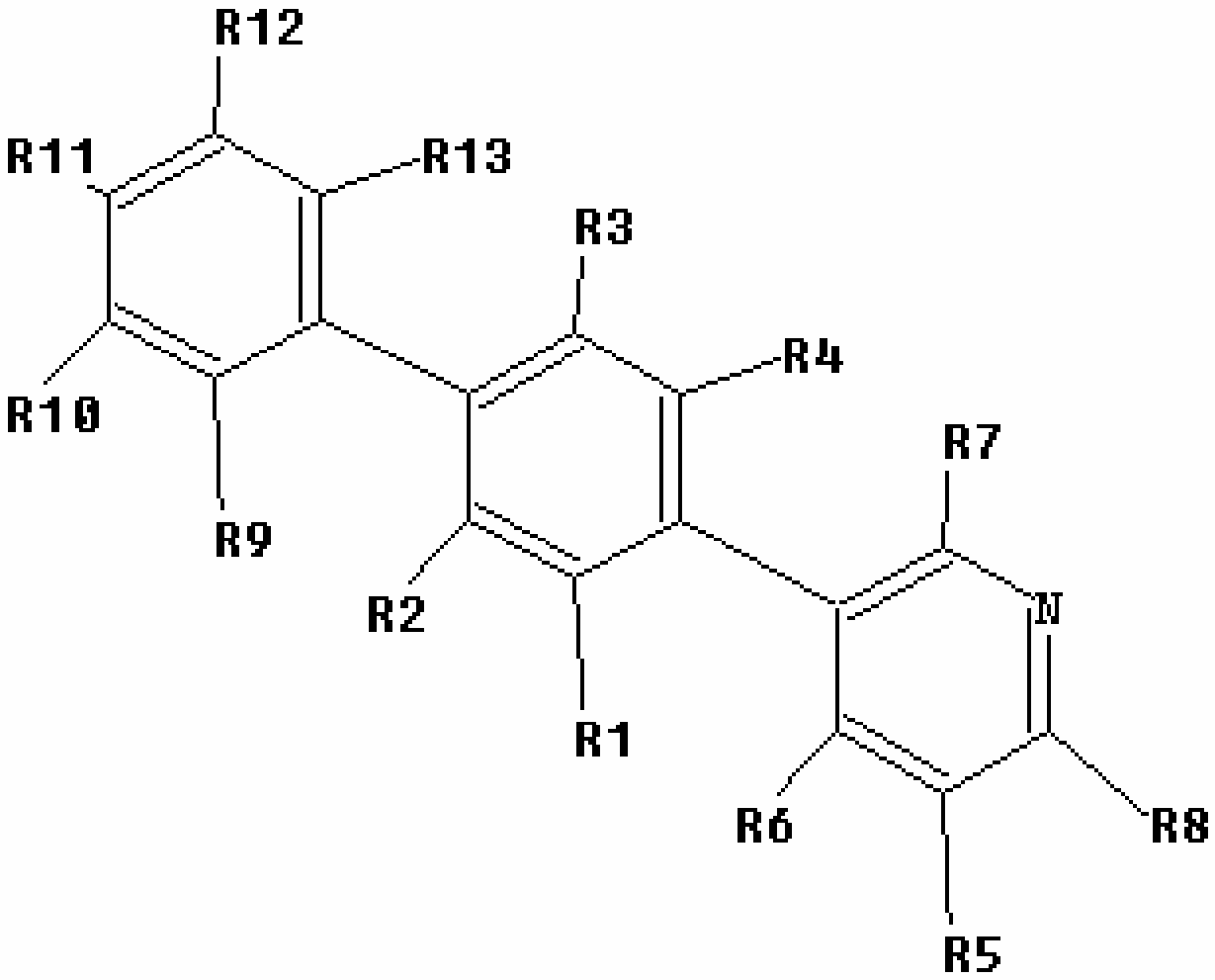


CAS Registry Number®	Index Name	R1	R2	R3

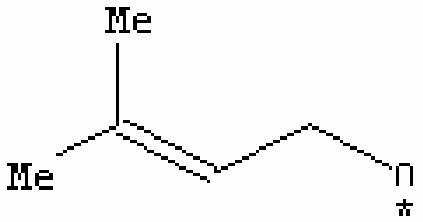
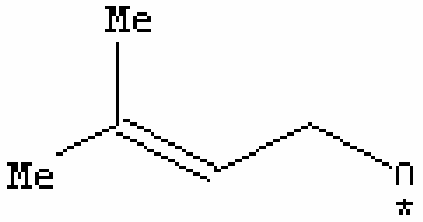
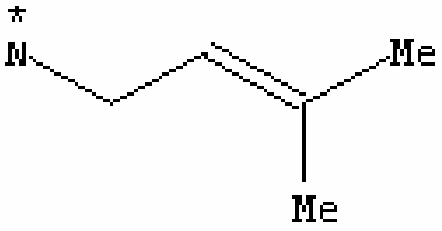
Representative R3 and R4 substituents



Use a customized “common structure”



Substitution patterns at R8 and R11

R8·¤	R9·¤	R10·¤	R11·¤
	H·¤	H·¤	
	H·¤	H·¤	

Why Structural Analysis

- Test the landscape – what analogs have been disclosed
- Analyze a specific document – cpds described generically as A -x- B -y- C
- Identify chemistry associated with an organization
- Present compounds in an easy to compare table – create an SAR table

Identifying the class of substances

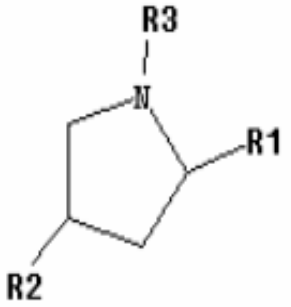
- Structure search
- Nomenclature search, e.g. ?PRIL/CNS or ?PRILAT/CNS

L-numbered set of substances in the REGISTRY file

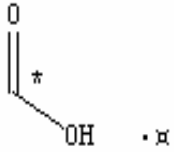
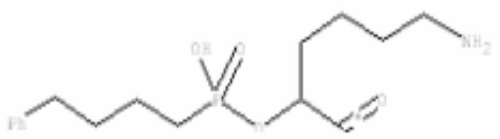
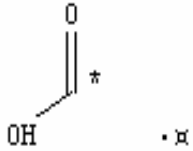
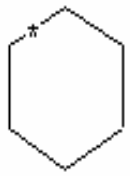
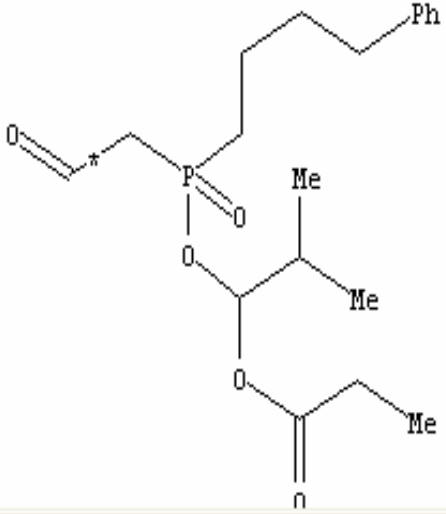
Identifying the substances associated with an organization

- S the L# in CAplus
- ANALYZE PA (optional)
- S desired PA terms
- TRANSFER HIT RN's back to the REGISTRY file

L-numbered set of substances associated with an organization



Proline·containing·ACE·Inhibitors·associated·with·company·YY

CAS· Registry· Number®	Index·Name	R1	R2	R3
111223-26-8· REGISTRY	L-Proline,·1-[(2S)-6-amino-2-[[hydroxy(4-phenylbutyl)phosphinyl]oxy]-1-oxohexyl]-·(9CI)		H	
98048-97- REGISTRY	L-Proline,·4-cyclohexyl-1-[[(R)-[(1S)-2-methyl-1-(1-oxopropoxy)propoxy](4-phenylbutyl)phosphinyl]acetyl]-,·(4S)-·(9CI)			

Summary of the presentation

- Why structural analysis
- Challenges with structural analysis
- Using the R-group Analysis Table Tool in STN Express
- Mining the R-group analysis table
- Additional examples

Thank you

- Colleagues
- Clients
- Tony Trippe
- ACS CINF organizers
- YOU