



# 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

E3 14140155 46.150.18/RID

- Ring Identifiers

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

L7

42 ANSWERS      REGISTRY      COPYRIGHT 2005 ACS

IR      Select the appropriate option

MF      Analyze Plus...

Display...

Refine...

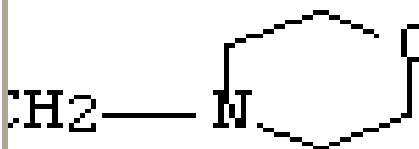
Save...

Save for R-group Analysis...

Save Answers for SciFinder®...

Single File SDI...

-[3-[2-(4-morphol



# Choose additional data to include in the R-group table

## Save for R-group Analysis Wizard



Select the substance information to include, then click Next. The CAS Registry Number® is automatically included

To select multiple fields, press the Ctrl key and click the appropriate entries. To select a range, press the Shift key and click the appropriate entries

Click Cancel to return to STN.

42  
answers  
have been  
retrieved.

IN	CA Index Name
MF	Molecular Formula
REF	Number of References
SCN	Short Chemical Name
SR	Source of Registration
BCF	Bioconcentration Factor
BP	Boiling Point
DEN	Density
FP	Flash Point
FRB	Freely Rotatable Bonds

# Select the R-group Analysis choice from the Results menu

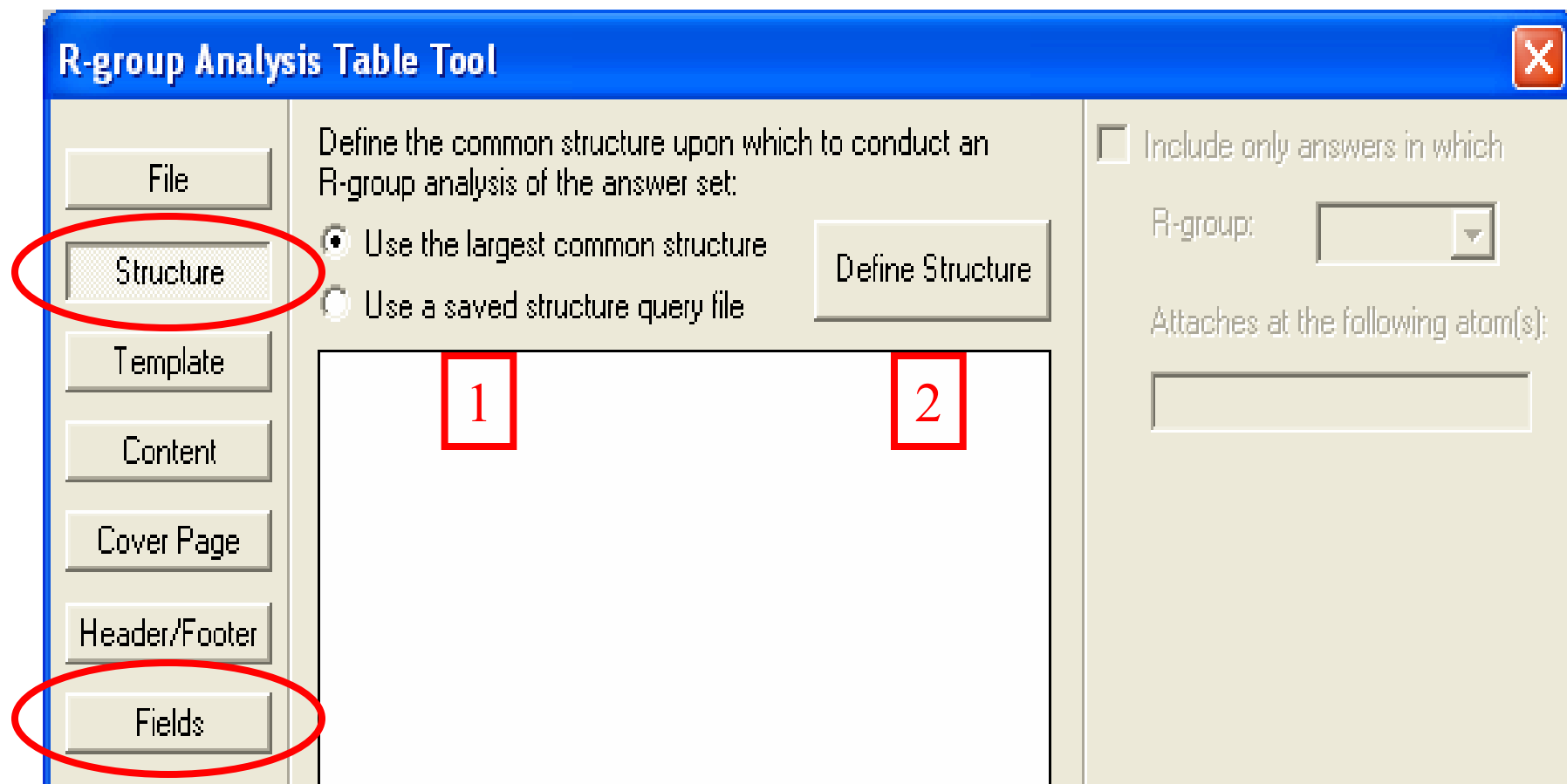
The screenshot shows the STN Express application window. The title bar reads "STN Express". The menu bar includes "File", "Logon", "Query", "Results", "Setup", "Web", and "Help". The "Results" menu is open, displaying the following options:

- Browse Transcript... (Ctrl+Shift+B)
- Print Transcript...
- Edit Transcript...
- Export Transcript... (Ctrl+Shift+X)
- Accounting... (Ctrl+A)
- Table Tool... (Ctrl+Shift+T)
- Report Tool... (Ctrl+Shift+R)
- BLAST® Report with Alignment Data...
- Predefined Reports...
- R-group Analysis Table Tool... (Ctrl+Shift+G)**
- Open... (Ctrl+O)

The "R-group Analysis Table Tool..." option is circled in red. In the background, the "Setup" menu item is highlighted with a yellow key icon, and the "Outlook Today" taskbar icon is visible.



# Follow the steps of the R-group Analysis wizard

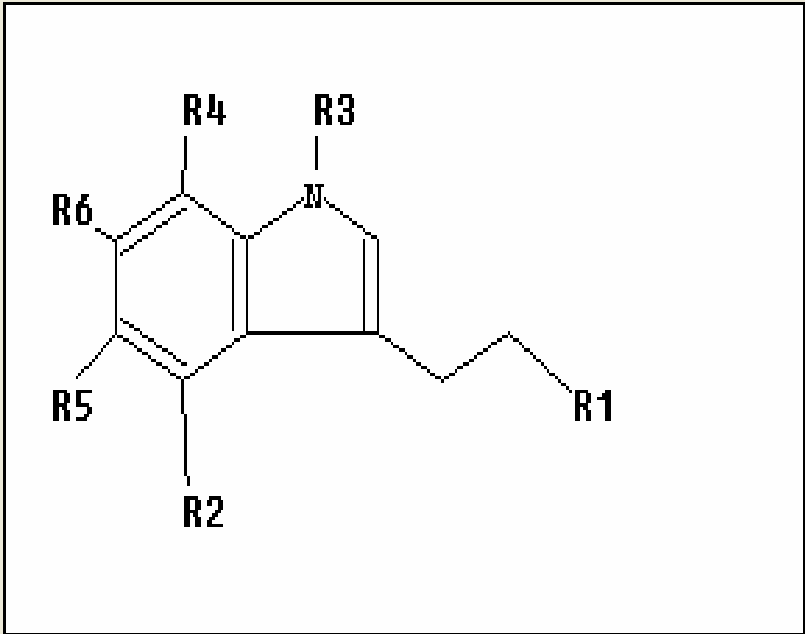


# On analysis of the data the software developed a common structure

**R-group Analysis Table Tool**

Define the common structure upon which to conduct an R-group analysis of the answer set:

- Use the largest common structure
- Use a saved structure query file



Include only answers in which

R-group:

Attaches at the following atom(s):

File

Structure

Template

Content

Cover Page

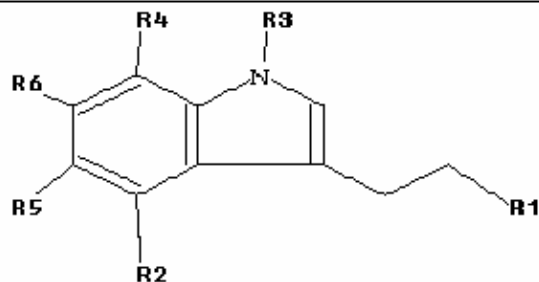
Header/Footer

Fields

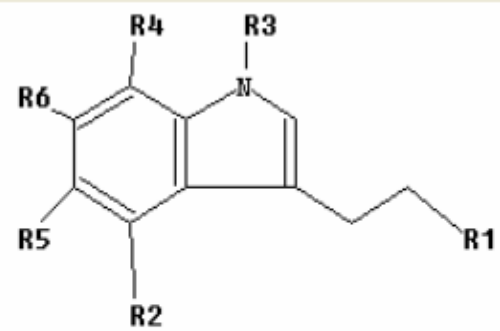
# Additional data may be selected for the table

The screenshot shows the 'R-group Analysis Table Tool' window. On the left side, there is a vertical menu with buttons for 'File', 'Structure', 'Template', 'Content', 'Cover Page', 'Header/Footer', and 'Fields'. The 'Fields' button is circled in red. The main area is divided into two columns: 'Field Name' and 'Selected Fields'. The 'Field Name' column contains a list of fields: 'Answer', 'Comments', 'Copyright', and 'Row'. The 'Selected Fields' column contains a list of fields: 'CAS Registry Number', 'Index Name', 'Molecular Formula', and 'R-Group'. The 'Selected Fields' list is circled in red, and the 'R-Group' item is highlighted in blue. Between the two columns are buttons for 'Insert >>', 'Insert All', 'Rename', 'Delete', and 'Delete All'. To the right of the 'Selected Fields' list are 'Change Order' buttons (up and down arrows). At the bottom, there are checkboxes for 'Remove duplicate fields within an answer', 'Ignore case when sorting', and 'Autofit'. A 'Format' button is also present, showing the current font settings as 'Times New Roman, 12 pts'.

# Express table may also be saved to RTF or XLS format



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)	C <sub>28</sub> H <sub>37</sub> N <sub>7</sub> 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-piperaziny)ethyl]-1H-indol-5-yl]- (9CD)	C <sub>28</sub> H <sub>37</sub> N <sub>7</sub> O <sub>2</sub> 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]- (9CD)	C <sub>28</sub> H <sub>36</sub> N <sub>6</sub> O <sub>2</sub> 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

**R-group Analysis Table Tool**

Define the common structure upon which to conduct an R-group analysis of the answer set:

- Use the largest common structure
- Use a saved structure query file

Define Structure

File  
Structure  
Template  
Content  
Cover Page  
Header/Footer  
Fields

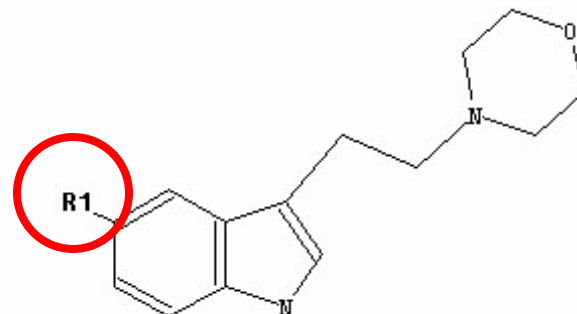
Include only answers in which

R-group: R1

Attaches at the following atom(s):

no match,N,skipped

The diagram shows a chemical structure with a benzimidazole core. A substituent labeled 'R1' is attached to the 6-position of the benzene ring. At the 2-position of the imidazole ring, there is a propyl chain that is connected to the nitrogen atom of a morpholine ring.



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)	C <sub>28</sub> H <sub>37</sub> N <sub>7</sub> O <sub>2</sub> 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)	C <sub>28</sub> H <sub>36</sub> N <sub>6</sub> O <sub>2</sub> Si	
753021-42-0	1H-Indol-5-amine, 3-[2-(4-morpholinyl)ethyl]- (9CI)	C <sub>14</sub> H <sub>19</sub> N <sub>3</sub> O	N



# Why Structural Analysis

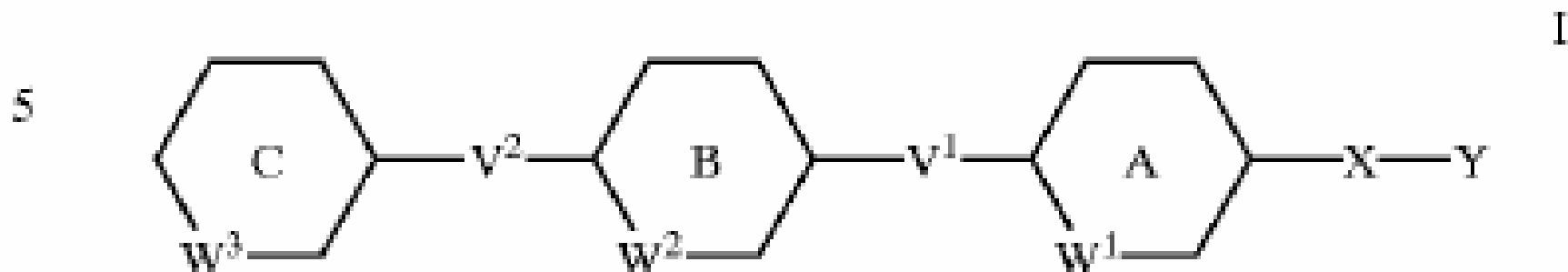
- Test the landscape – what analogs have been disclosed
- Analyze a specific document – cpds described generically as A -x- B -y- C
- Identify chemistry associated with an organization
- Present compounds in an easy to compare table – create an SAR table



# What is claimed?

What is claimed is:

1. A compound of the formula (I):



10 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

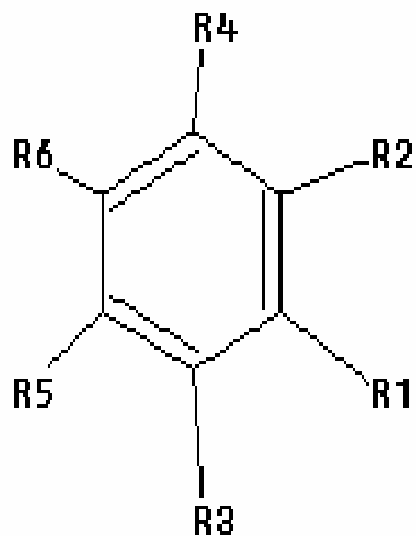


# Identifying the substances associated with a document

- 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®

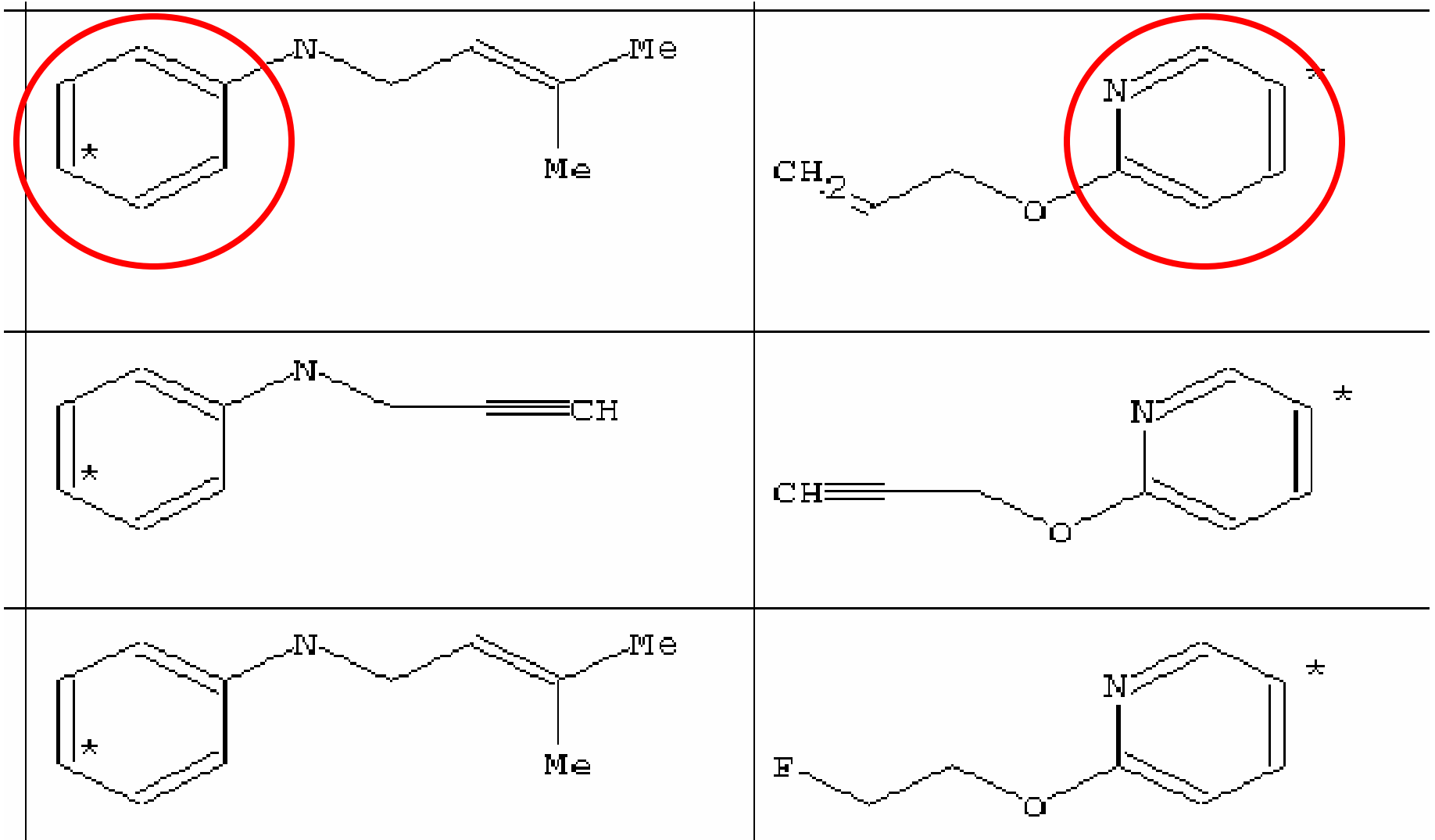
Index Name

R1

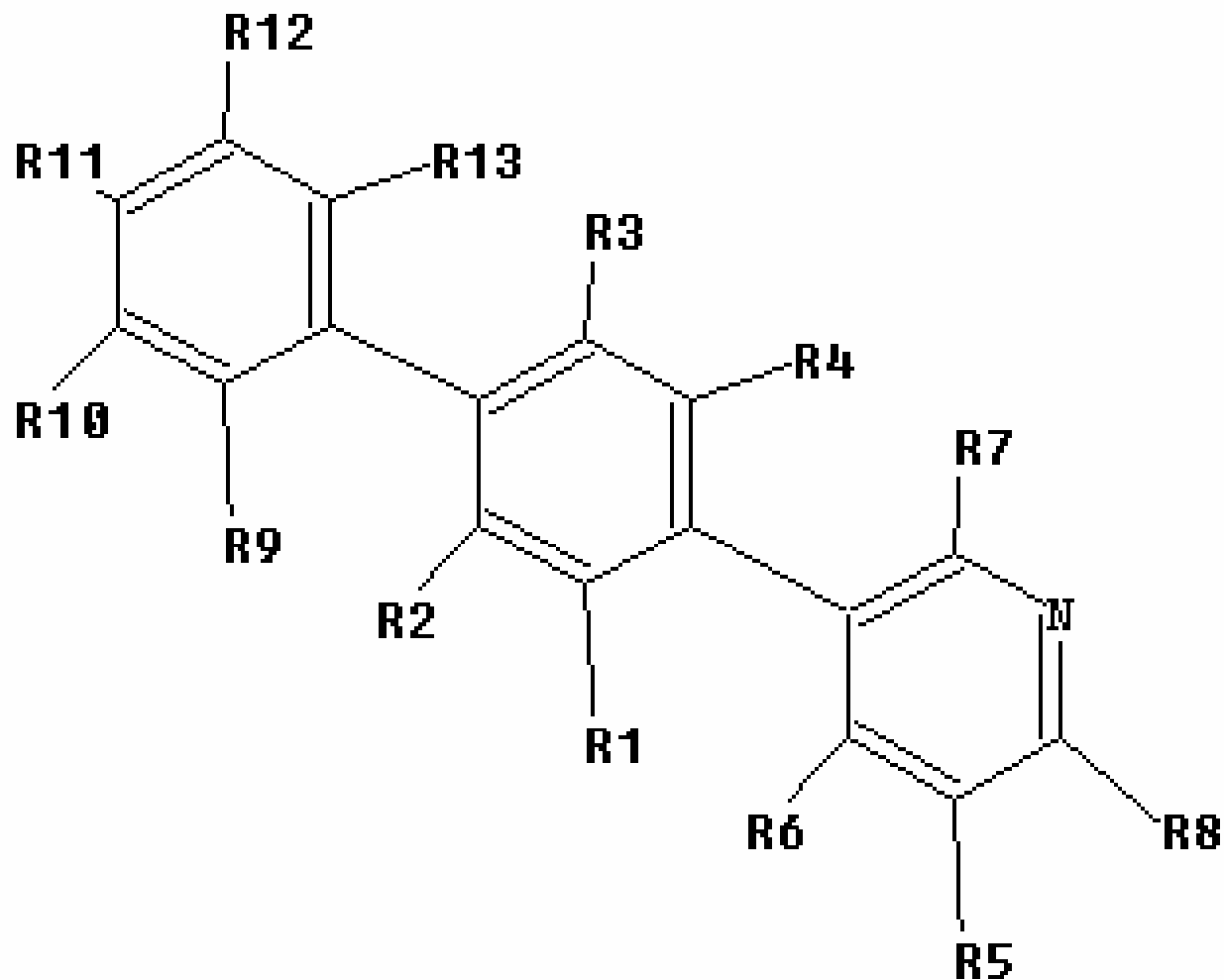
R2

R3

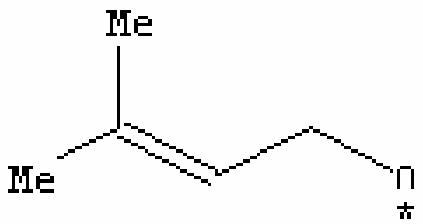

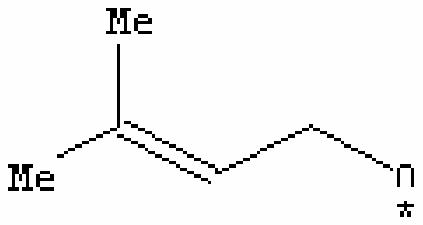
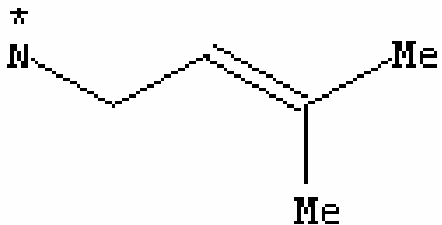
# Representative R3 and R4 substituents



Use a customized “common structure”



# Substitution patterns at R8 and R11

R8. $\alpha$	R9. $\alpha$	R10. $\alpha$	R11. $\alpha$
 <p>Chemical structure showing a double bond between two carbons. The left carbon is substituted with two methyl groups (Me). The right carbon is substituted with a hydrogen atom (H) and a propyl chain. The propyl chain ends in a methyl group (Me) and a hydrogen atom (H) marked with an asterisk (*).</p>	H. $\alpha$	H. $\alpha$	 <p>Chemical structure showing a benzene ring (Ph) attached to a propyl chain. The propyl chain starts with a hydrogen atom (H) marked with an asterisk (*). The chain continues to a methyl group (Me) and ends with a hydrogen atom (H) marked with an asterisk (*).</p>
 <p>Chemical structure showing a double bond between two carbons. The left carbon is substituted with two methyl groups (Me). The right carbon is substituted with a hydrogen atom (H) and a propyl chain. The propyl chain ends in a methyl group (Me) and a hydrogen atom (H) marked with an asterisk (*).</p>	H. $\alpha$	H. $\alpha$	 <p>Chemical structure showing a double bond between two carbons. The left carbon is substituted with a methyl group (Me) and a propyl chain. The propyl chain starts with a hydrogen atom (H) marked with an asterisk (*). The right carbon is substituted with two methyl groups (Me). The propyl chain ends with a methyl group (Me) and a hydrogen atom (H) marked with an asterisk (*).</p>



# 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

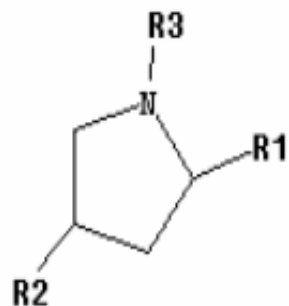


# Identifying the substances associated with an organization

- 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

March 13, 2005

ACS 229<sup>th</sup> National Meeting, San Diego, California

Copyright © 2005 American Chemical Society.

Copying or further distribution requires permission of CAS.

