

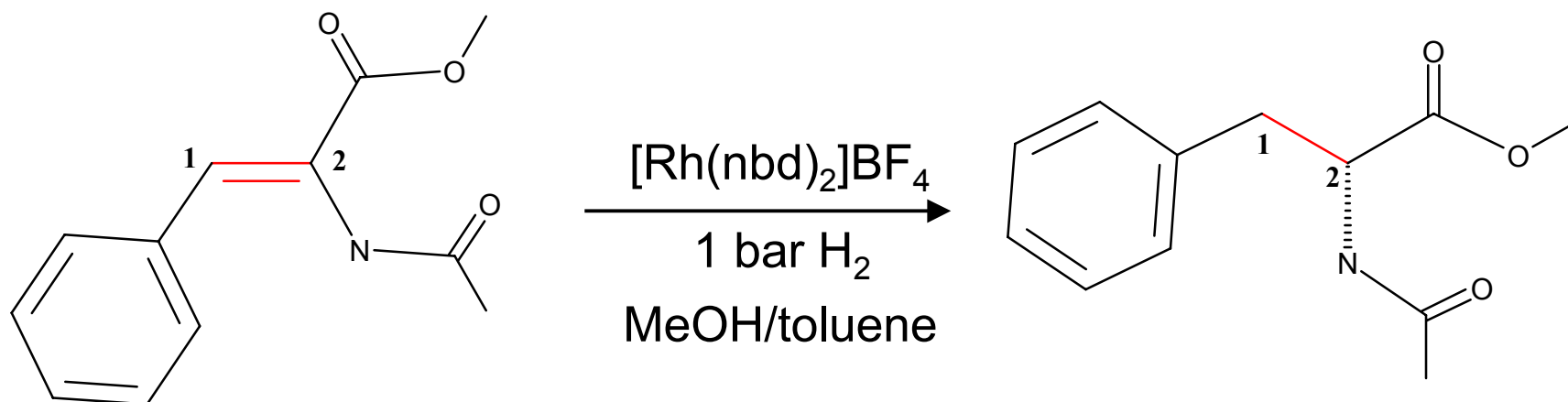
# Solving the Reaction Retrieval Problem



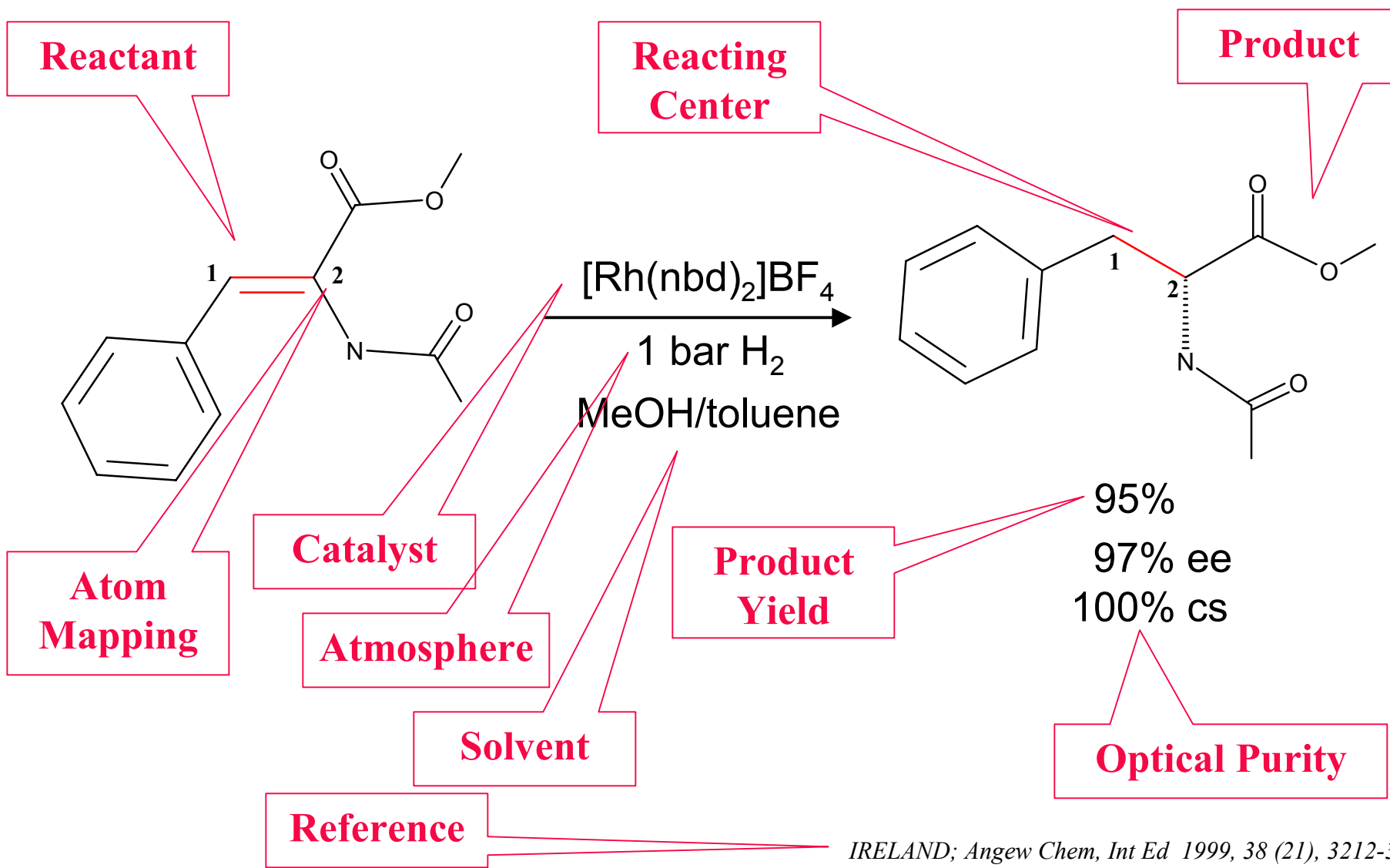
Chemistry Division Special Libraries Association  
San Antonio, June 9 – 14, 2001

Robert W. Snyder

- ▶ Reaction Informatics
- ▶ Reaction Retrieval
- ▶ Common Reaction Searches
- ▶ The Automatic Search



95%  
97% ee  
100% cs



- ▶ Pointer to the primary literature
  - ◆ Soon to include tertiary literature
- ▶ Looking for methodology, not preparation of a specific compound
- ▶ Database search acts as filter to remove nonrelevant examples in search results
  - ◆ “retrieve all” ... “exact reaction”
- ▶ Cluster browsing can help filter nonrelevant examples after the search
  - ◆ Use of reaction classification from InfoChem

- ▶ Focus on preparations
  - ◆ CrossFire Beilstein, ORGSYN
- ▶ Focus on transformations
  - ◆ Theilheimer, RX-JSM, CHC, Reference Library
- ▶ Focus on methodology
  - ◆ ChemInform Reaction Library
  - ◆ Solid-Phase Organic REactions (SPORE)

## CrossFire Beilstein

- ◆ Over 8.5 million preparations
- ◆ Hyperlinked to citations and compounds
- ◆ Extensive compilation of factual data

## ORGSYN

- ◆ model synthetic procedures
- ◆ preparation of common reagents
- ◆ independently tested and optimized procedures
- ◆ explicit experimental details and hazard warnings

## THEILHEIMER

- ◆ high-yield functional group transformations

## RX-JSM

- ◆ new synthetic methods
- ◆ novel functional group and ring chemistry
- ◆ protected group chemistry

## Comprehensive Heterocyclic Chemistry (CHC)

- ◆ synthesis of heterocycles
- ◆ reactions of heterocyclic systems
- ◆ use of heterocycles in the synthesis of non-heterocyclic structures

## The Reference Library of Synthetic Methodology

- ◆ novel methods, selective transformations
- ◆ new applications of known reagents
- ◆ asymmetric synthesis
- ◆ catalysis

## ChemInform Reaction Library

- ◆ novel methodologies
- ◆ heterocyclic chemistry
- ◆ synthesis of organo-element compounds
- ◆ utilization of new catalysts
- ◆ enzymatic processes

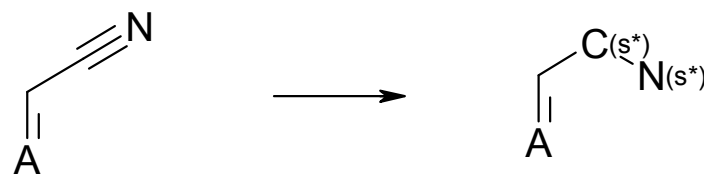
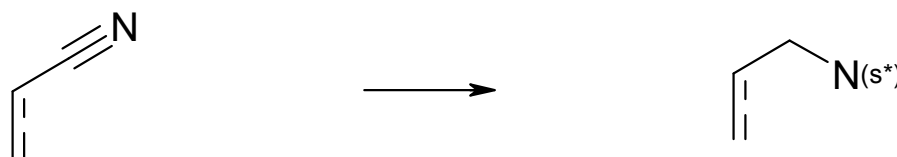
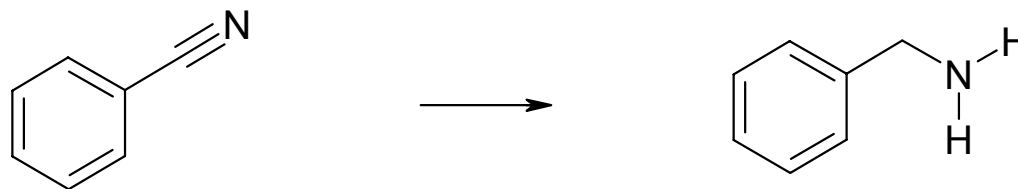
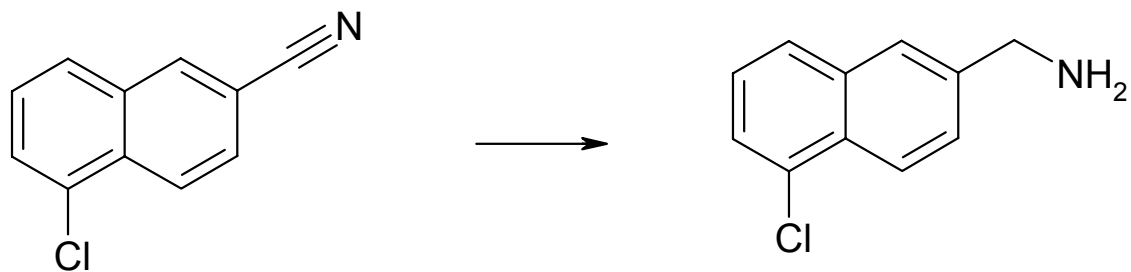
## Solid-Phase Organic Reactions (SPORE)

- ◆ reactions of small organic compounds on solid support
- ◆ coupling to and from solid supports
- ◆ protection/deprotection of functional groups

- ▶ Molecule (SSS)
  - ◆ know (sub)structure of reactant, product, catalyst, solvent, polymer, solid support, or protecting group
- ▶ Reaction (RSS)
  - ◆ know (sub)structure of reactant and product
  - ◆ functional group compatibility
    - know reagent and functional group

- ▶ Data searches
  - ◆ very useful in solid-phase chemistry
  - ◆ usually refine prior search
- ▶ Scheme navigation
  - ◆ previous step
  - ◆ next step
  - ◆ derivative reactions
  - ◆ view entire scheme

# Reaction Search



# Reaction Search

ISIS/Base - [RXNBROWS (GIMLI).DB/ExpertQuery]

File Edit Options Object Database Search List Window Help

QUERY ASSISTANT Clear Data Automap Clear Mapping Helpful Hints SEARCH

Forms Query Browse Update <RXN> 0 of 0  
Search Domain: All

Select type of reaction search: Reaction Substructure (RSS) Expert Query Form

Chemical reaction diagram showing the conversion of a nitrile to an imine. The reactant is a nitrile with a double bond between carbon 2 and nitrogen 4, and a single bond between carbon 2 and carbon 3. The product is an imine with a double bond between carbon 2 and carbon 3, and a single bond between carbon 3 and nitrogen 4. The carbon and nitrogen atoms are labeled with their respective numbers and 's\*' to indicate stereochemistry.

# Reaction Search

ISIS/Base - [RXNBROWS (GIMLI).DB/Summary]

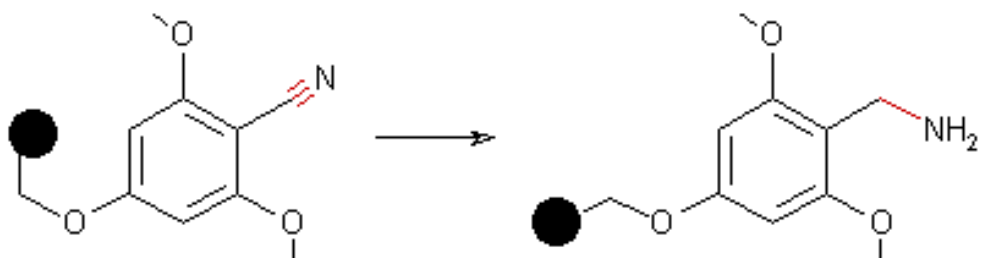
File Edit Options Object Database Search List Window Help

Main List Scheme Deriv.Rxns Examples CLUSTER Cluster # View Scheme

Forms Query Browse Update <RXN> 1 of 80 Search Domain: All

Summary Reference Reagents Conditions Prot. Group Solid Table Classcodes

SPORE 2001.1 RXII 1 of 80 Variation 1 of 1 Path A Step 2 OF 5

Cluster #	Item #	Value:
		

Variation	Literature Reference	Reagents	% Yield
1	SHAO, J.; LI, Y.-H.; VOELTER, W.; Int J Pept Protein Res [JPPC3] 1990, 36, 182-187.	LiAlH4 dry NH3 Et2O	

# Reaction Search

ISIS/Base - [RXNBROWS (GIMLI).DB/ExpertQuery]

File Edit Options Object Database Search List Window Help

QUERY ASSISTANT Clear Data Automap Clear Mapping Helpful Hints SEARCH

Forms Query Browse Update <RXN> 0 of 0  
Search Domain: All

Select type of reaction search: Reaction Substructure (RSS) Expert Query Form

The diagram illustrates a chemical reaction. On the left, a chloroalkene molecule is shown with a chlorine atom (Cl) labeled '5' and a double bond to a carbon atom labeled '1'. This carbon is also bonded to a carbon atom labeled '2', which is further bonded to a nitrogen atom labeled '4'. The nitrogen atom has a double bond to a carbon atom labeled '3'. On the right, the product molecule is shown, which is identical to the reactant but with additional labels: the carbon atom '3' is now labeled '3' and 'C(s\*)', and the nitrogen atom '4' is now labeled '4' and 'N(s\*)'. An arrow points from the reactant to the product.

# Reaction Search

ISIS/Base - [RXNBROWS (GIMLI).DB/Summary]

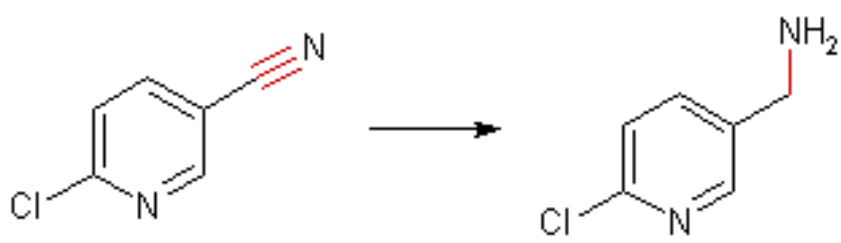
File Edit Options Object Database Search List Window Help

Main List Scheme Deriv.Rxns Examples CLUSTER Cluster # View Scheme

Forms Query Browse Update <RXN> 3 of 5  
Search Domain: All

Summary Reference Reagents Conditions Table Classcodes

RX-JSM 2000.1 RXII 3 of 5 Variation 1 of 1 Path A Step 1 Step

Cluster #	Item #	Value:	
			
Variation	Literature Reference	Reagents	% Yield
1	Bayer Ag; FRG Patent 1989, 3726993 (DE-3726993), 89-061967.	Raney Ni (Damp)	

# Functional Group Compatibility

ISIS/Base - [RXNBROWS (GIMLI).DB/ExpertQuery]

File Edit Options Object Database Search List Window Help

QUERY ASSISTANT Clear Data Automap Clear Mapping Helpful Hints SEARCH

Forms Query Browse Update <RXN> 0 of 0  
Search Domain: All

Select type of reaction search: Reaction Substructure (RSS) Expert Query Form

Select type of molecular search and role of molecule in reaction:

<a href="#">Click to select search type and role of molecule in rxn</a>	Formula
	Reagent

"Li Al H4"

Select field and enter data value:

<a href="#">Click here to select data field</a>
<a href="#">Click here to select data field</a>
<a href="#">Click here to select data field</a>
<a href="#">Click here to select data field</a>

# Functional Group Compatibility

ISIS/Base - [RXNBROWS (GIMLI).DB/Summary]

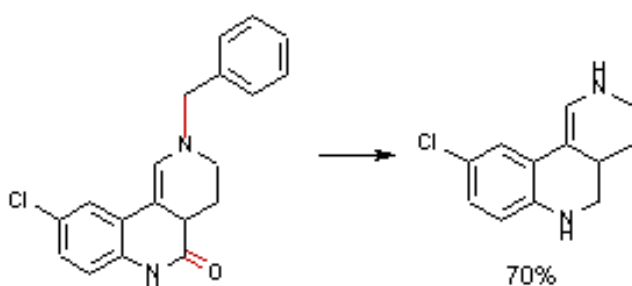
File Edit Options Object Database Search List Window Help

Main List Scheme Deriv.Rxns Examples CLUSTER Cluster # View Scheme

Forms Query Browse Update <RXN> 1 of 166  
Search Domain: All

Summary Reference Reagents Conditions Table Classcodes

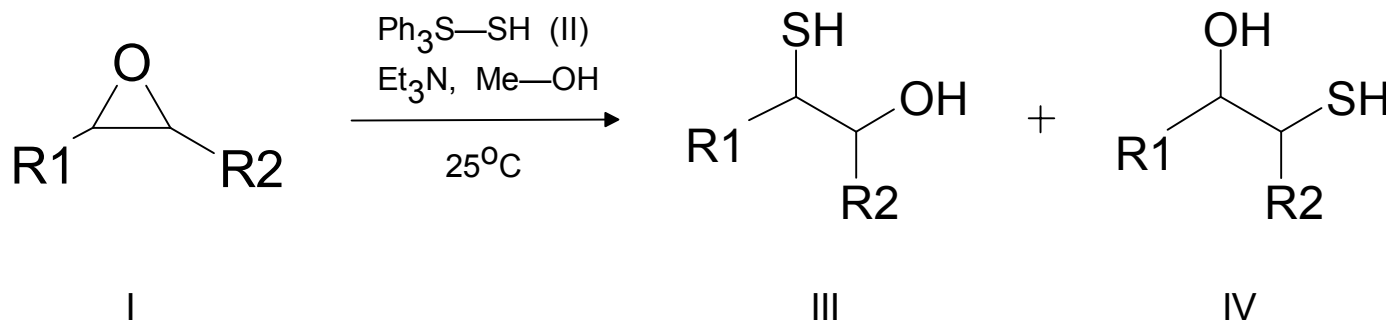
ChemInform RX 2000.1 RXII 1 of 166 Variation 1 of 1 Path A Step 2 OF 2

Cluster #	Item #	Value:
		

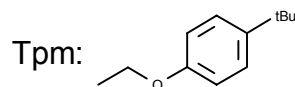
Variation	Literature Reference	Reagents	% Yield
1	DEL GIUDICE, M. R.; MUSTAZZA, C.; FERRETTI, R.; BORIONI, A.; GATTA, F.; J Heterocycl Chem [JHTCAD] 1998, 35 (4), 915-922.	1. LiAlH4  Et2O 2. 2 equiv. Cl-CO-O-Et toluene	70

- ▶ Concise representation of overall synthetic strategy in article
- ▶ Quickly locate starting materials, intermediates, and end products
- ▶ Derivative reactions represented as Markush-like reactions
- ▶ Determine scope and limitations of methodology
- ▶ Ability to copy, paste, print reaction schemes

# Scheme Display Files: Substituent Effects

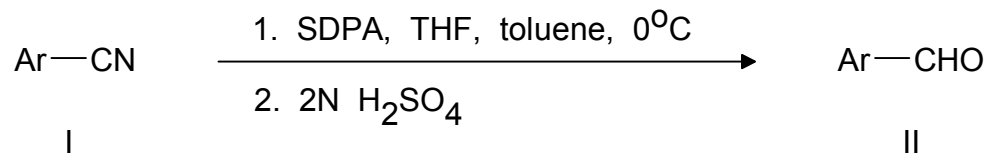


a	R1 : —Ph ; R2 : —H	76%	11%
b	R1 : —(CH <sub>2</sub> ) <sub>5</sub> —Me ; R2 : —H	0%	85%
c	R1 : —Ph ; R2 : —Me	74%	0%
d	R1 : —Tpm ; R2 : —H	0%	81%

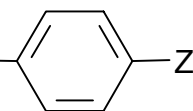


(Brittain, J.; Gareau, Y.; Tetrahedron Lett. (1993) 21, 3363)

# Scheme Display Files: Electronic Effects

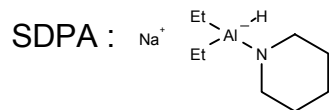


a Ar : —Ph 98%

b Ar : -Z [Z : —O—Me, —Me, —NMe<sub>2</sub>] 97/95/97%

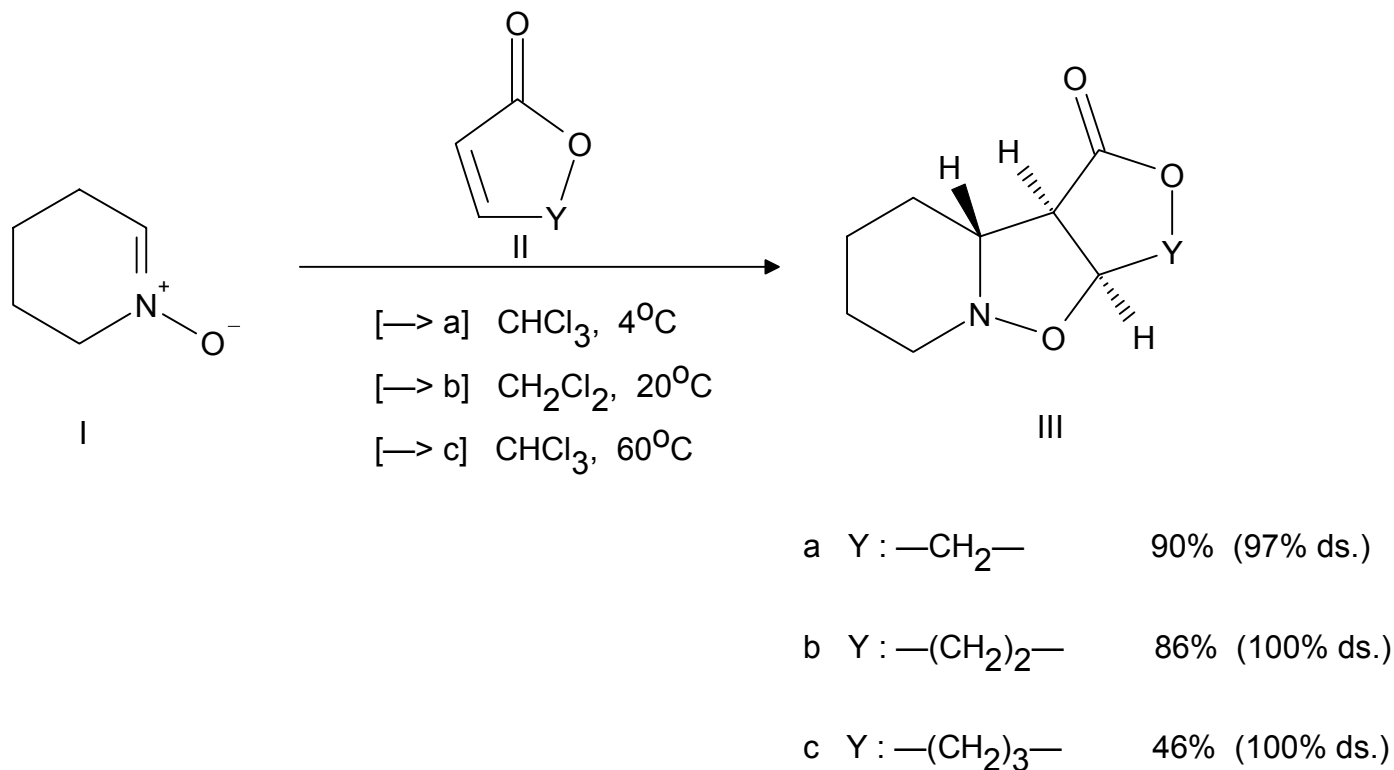
c Ar : -Br 84%

d Ar : -NO<sub>2</sub> 0%

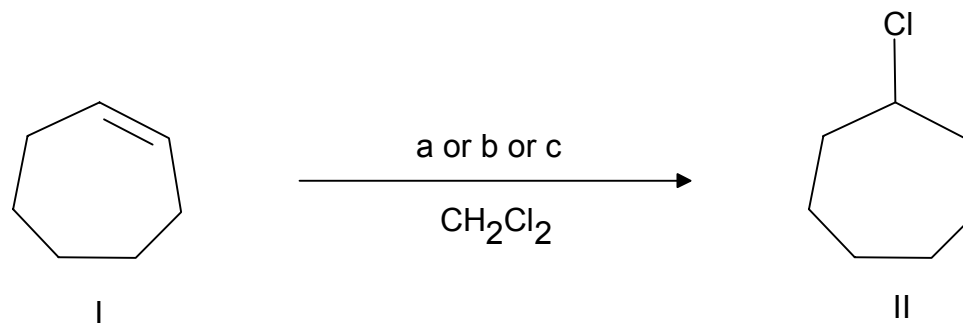


(Yoon, N.; An, D.; Jeong, K.; Bull. Korean Chem Soc. (1993) 2, 302)

# Scheme Display Files: Structural Effects



# Scheme Display Files: Reagent Effects



a : gaseous HCl

b :  $\text{SOCl}_2$ ,  $\text{SiO}_2$

c :  $\text{SOCl}_2$ ,  $\text{Al}_2\text{O}_3$

0%

81%

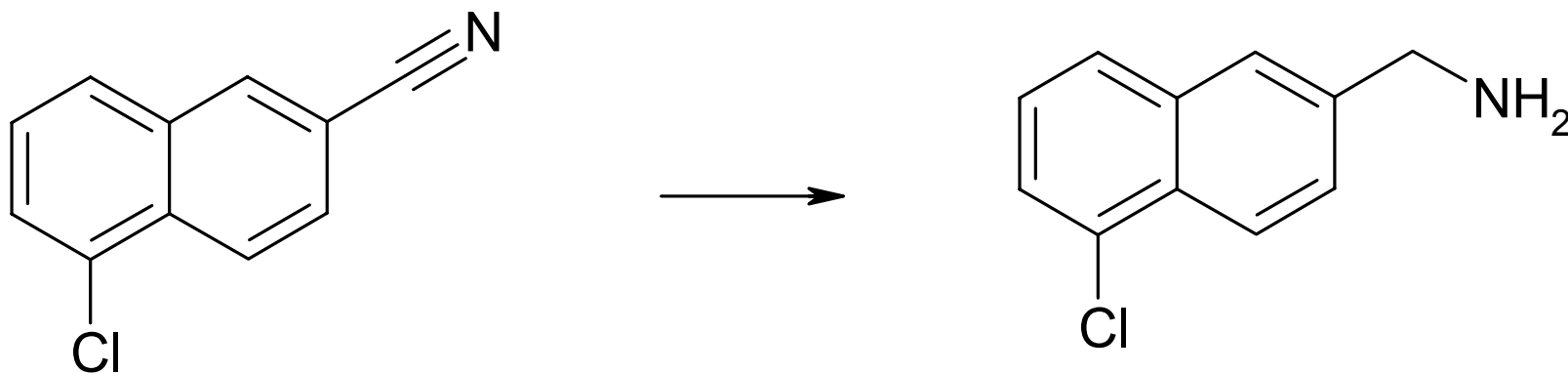
94%

(Kropp, P., et al; J. Am. Chem. Soc. (1993) 8, 3071)

# The Automatic Search

- ▶ Enduser specifies full reaction of interest to the system
- ▶ A series of searches are automatically conducted, each with less specificity, until reaction examples are retrieved from the database
  - ◆ Exact reaction
  - ◆ Same transformation (narrow, medium, broad)
  - ◆ Reaction substructure (RSS)
  - ◆ Similarity (100...0)
- ▶ Emulates the thought process of the experienced searcher

# Automatic Search Example



# Automatic Search Query

ISIS/Base - [RXNBROWS (GIMLI).DB/ExpertQuery]

File Edit Options Object Database Search List Window Help

QUERY ASSISTANT Clear Data Automap Clear Mapping Helpful Hints SEARCH

Forms Query Browse Update <RXN> 0 of 0  
Search Domain: All

Select type of reaction search: Automatic Search Expert Query Form

The diagram illustrates a chemical reaction. On the left is the reactant, 4-chlorobenzonitrile, which consists of a benzene ring with a chlorine atom at the para position and a nitrile group (-C≡N) at the other para position. An arrow points to the right, leading to the product, 4-chlorobenzylamine, which is a benzene ring with a chlorine atom at the para position and a primary amine group (-CH<sub>2</sub>NH<sub>2</sub>) at the other para position.

# Results Retrieved From Automatic Search

ISIS/Base - [RXNBROWS (GIMLI).DB/Summary]

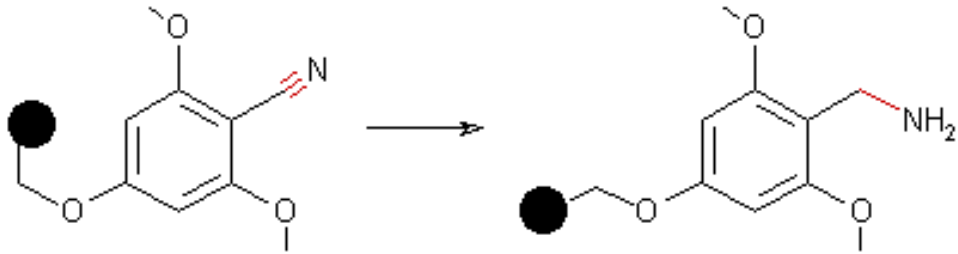
File Edit Options Object Database Search List Window Help

Main List Scheme Deriv.Rms Examples CLUSTER Cluster # View Scheme

Forms Query Browse Update <RXN> 1 of 56 Search Domain: All

Summary Reference Reagents Conditions Prot. Group Solid Table Classcodes

SPORE 2001.1 RXII 1 of 56 Variation 1 of 1 Path A Step 2 OF 5

Cluster #	Item #	Value:
		

Variation	Literature Reference	Reagents	% Yield
1	SHAO, J.; LI, Y.-H.; VOELTER, W.; Int J Pept Protein Res [JPPC3] 1990, 36, 182-187.	LiAlH4 dry NH3 Et2O	

- ▶ Proper reaction retrieval requires both structural and data indexing
- ▶ Reaction informatics can make available a large collection of synthetic chemistry information – but gaining access can be difficult
- ▶ Reaction searching can be very powerful for the experienced searcher
- ▶ Automatic search enables the occasional user to benefit from reaction databases

*Thank you!*

Questions: [bobs@mdli.com](mailto:bobs@mdli.com)