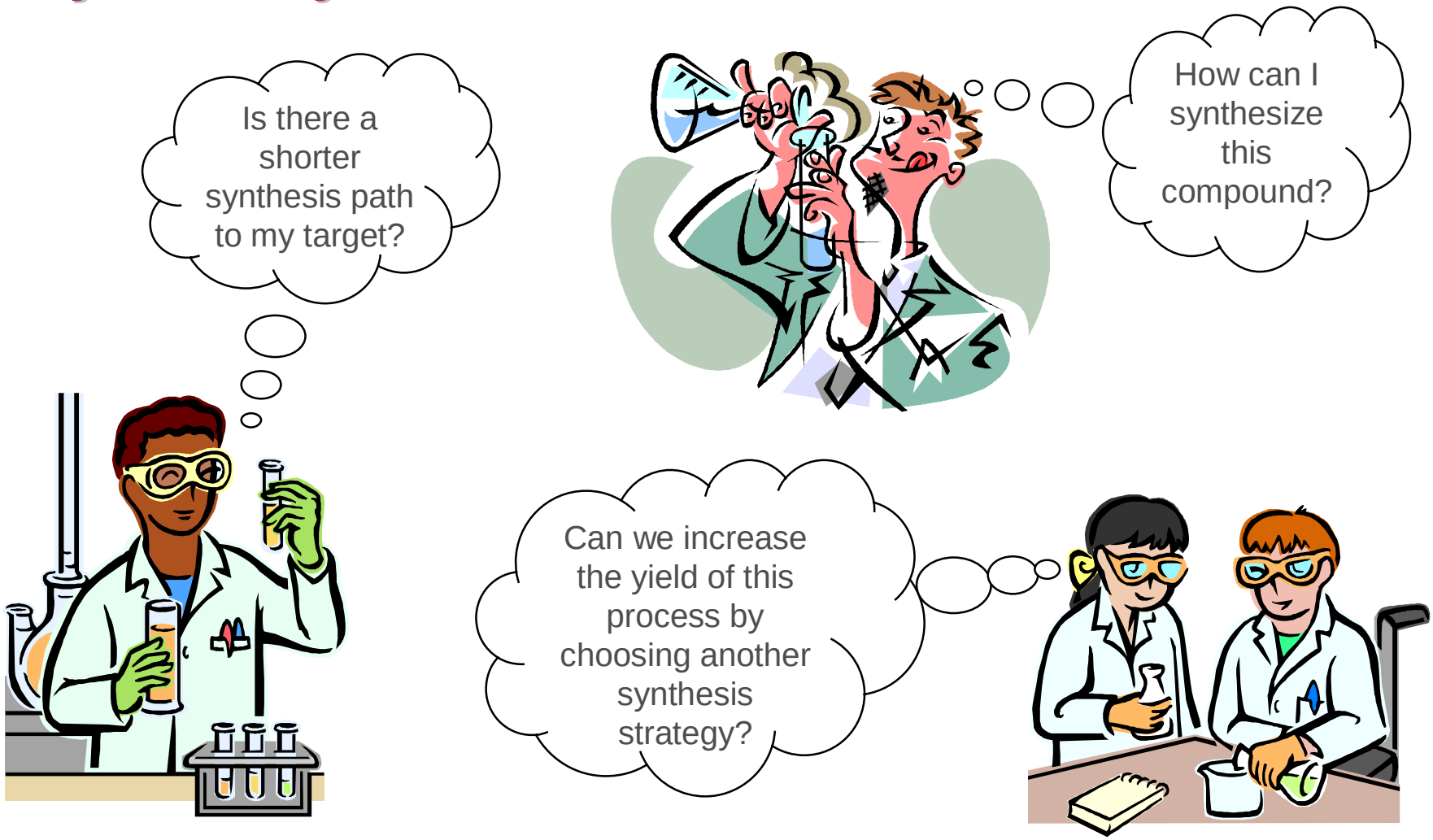
Novel Approach to Retrosynthesis Automatic Generation of Transform Libraries

V. Eigner-Pitto, J. Eiblmaier, H. Kraut, H. Saller and P. Loew.
InfoChem GmbH, Landsberger Strasse 408, Munich, 81241, Germany

Why Retrosynthesis?



Retrosynthesis: Expected Features

Main Challenge: Automatic generation of chemically meaningful precursors and their reasonable validation

1. Evaluation of Alternative Synthesis Pathways

Sorting criteria are necessary for broad representations with many expanding paths:

- Yield
- Type of transformation (e.g. ring-closures or construction reactions preferred, etc.)

2. Support of Multistep Synthesis

Necessary to guide the synthesis from the target to all existing precursors that can be obtained from commercially available substances

3. Organized Visualization

Allows a broad view over the different strategies. A tree view, i.e., gives an organized representation of the synthesis paths to the target

Synthesis-Planning Tools

1. Use of known compounds and published reactions

i.e. MDL *(CrossFire Commander),* Symyx *(Reaction Pathfinder),* InfoChem *(STS)*

- ✓ Handle multistep synthesis
- ✓ Order in synthesis paths - Sorting criteria
- ✓ Shows synthesis options in your database (tree representation)
- × Structure and corresponding reaction(s) must be in database

2. „Intellectual“ retrosynthesis systems

- ✓ New molecules (not in the database)
- ✓ Alternative routes to known molecules
- ✓ Reaction transformation / Synthesis strategy to specify
- ✓ Transform based

„Intellectual“ Retrosynthesis Approaches

2a. Transforms compiled manually from literature

- ✓ accurate, highly specific
- ✗ tedious, expensive, restrictive

2b. Transforms based on enumerated bond changes

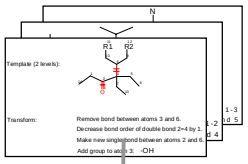
- ✓ fully automated
- ✗ unspecific, combinatorial explosion

Examples:

- LHASA, SECS / CASP ... (Corey, Wipke)
- WODCA, EROS ... (Gasteiger)
- SYNGEN ... (Hendrickson)
- AIPHOS ... (Sasaki)
- .....

RetroSynthesis Concept Overview

1. Pre-processing



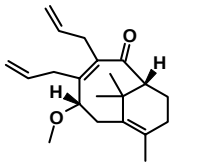
Automatic transform extraction
(IC*MAP*/CLASSIFY)

Reaction database

e. g.
SPRESI
proprietary databases
commercial databases

Transform Library

2. RetroSynthesis



Target molecule

RetroSyn

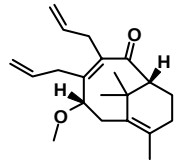
Lookup stored examples

Reaction 1: Precursor(s) 1 — Transform x

Reaction 2: Precursor(s) 2 — Transform y

...

Reaction n: Precursor(s) n — Transform z



Target molecule

Transform x

Transform y

Transform z

Reaction Database Requirements

1. Correct Reaction Mapping

Description of the transformation occuring during the reaction

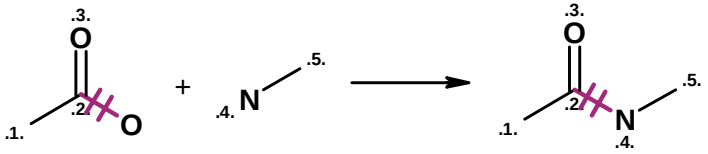
⇒ Determination of reaction centers

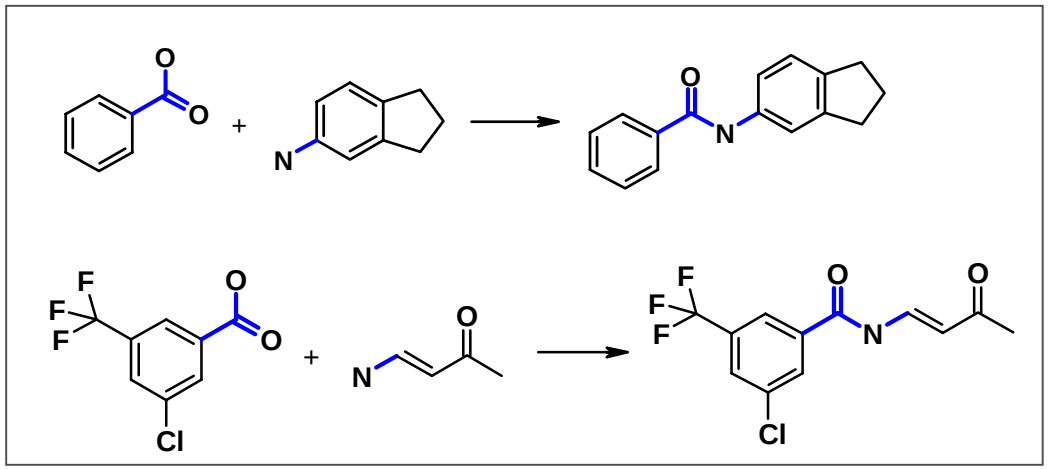
2. Reaction Classification

Automatic identification of the structural environment around the reacting centers

⇒ Determination of unique identifiers (ClassCodes)

Reaction Type **Amination**:

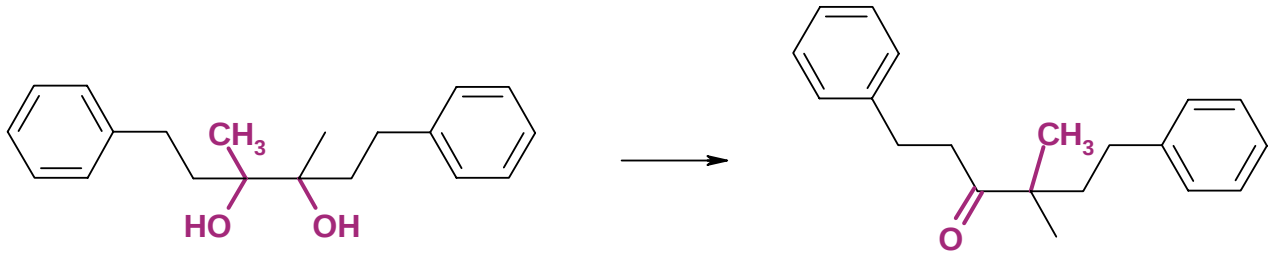


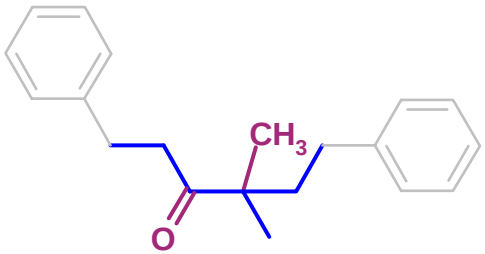


Definition of Template (Retron)

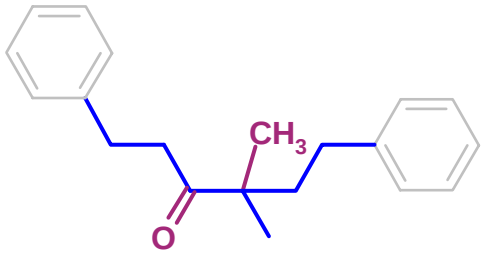
SPRESI Reaction Reg-No.: 3179546

Kita Yasuyuki, Yoshida Yutaka, Mihara Sachiko, Fang Dai-Fei, Higuchi Kazuhiro, Furukawa Akihiro, Fujioka Hiromichi
Efficient pinacol rearrangement mediated by trimethyl orthoformate
Tetrahedron Lett., 38 (1997) 48, 8315-8318

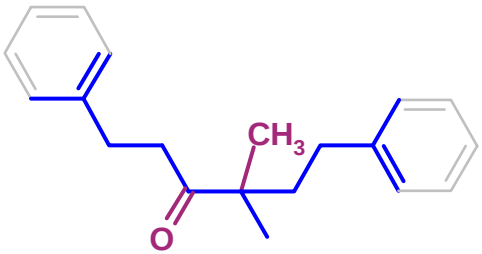




Template (level 2)



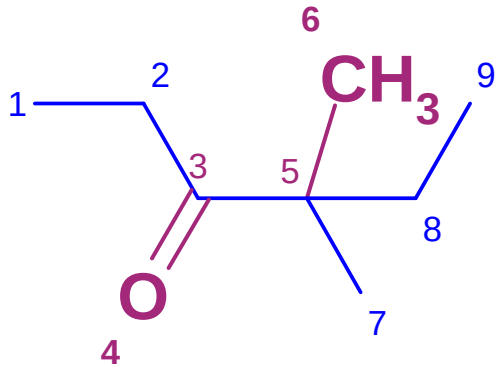
Template (level 3)



Template (level 4)

Definition of Transform

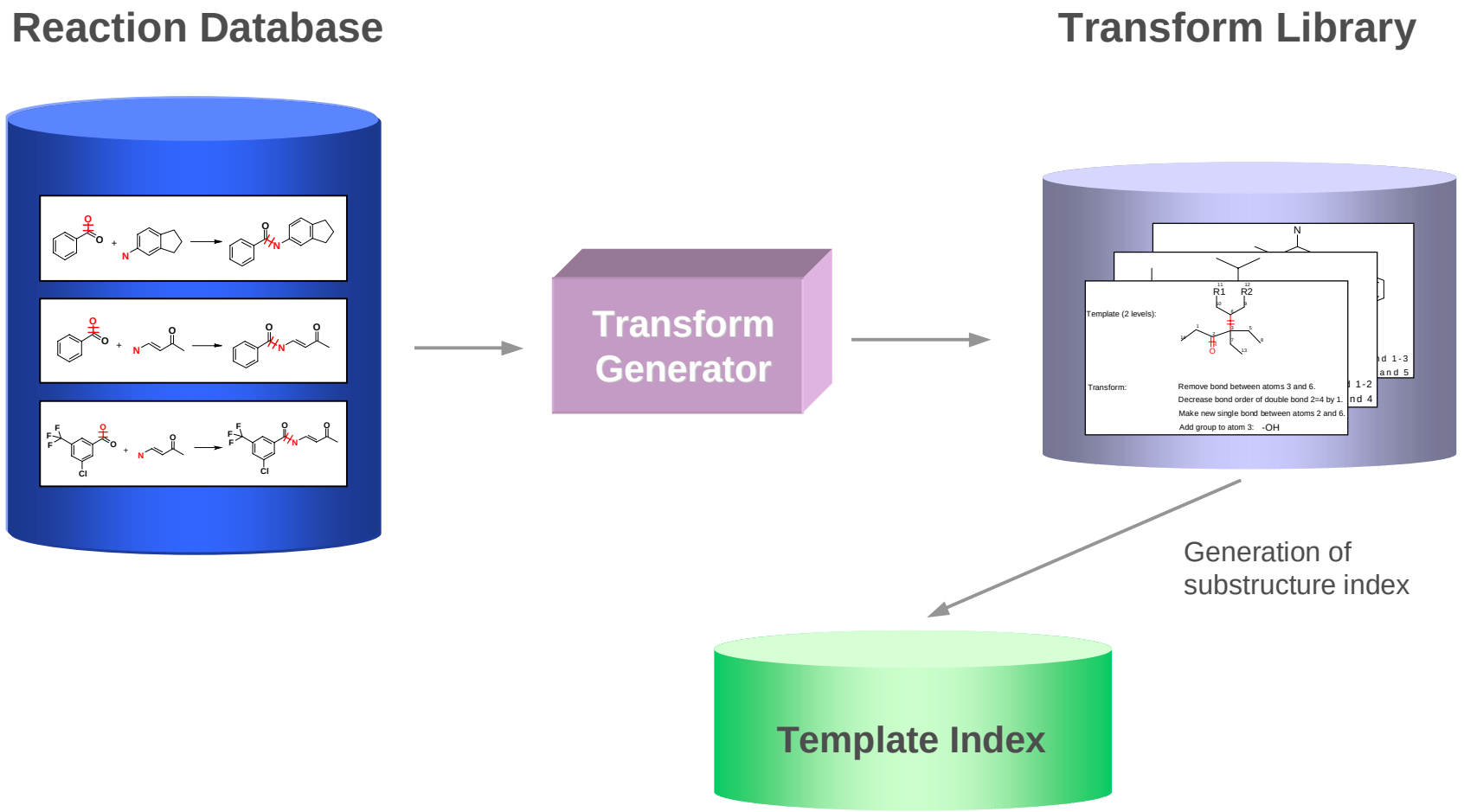
Template:



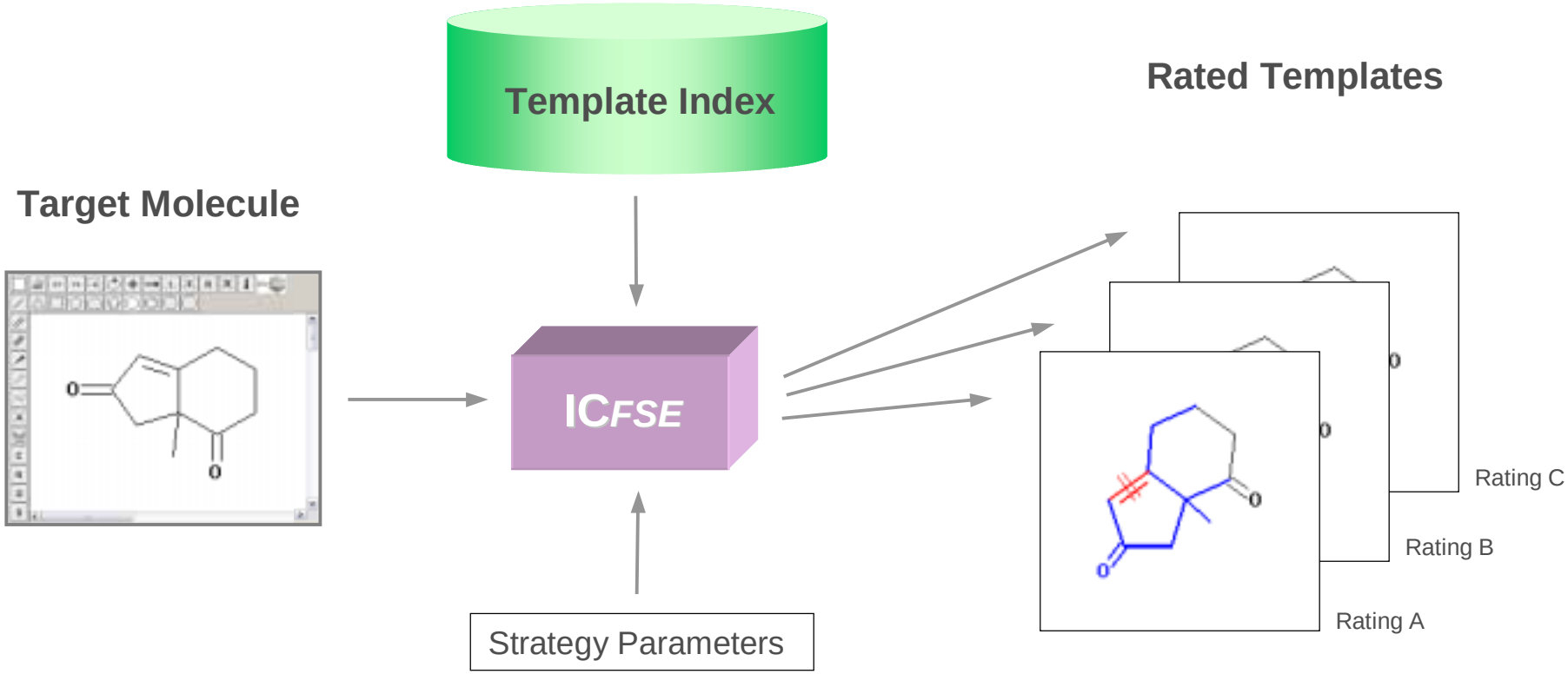
Indications:

- Remove bond between atom 5 and 6
- Decrease bond order of double bond 3=4 by 1
- Make new single bond between atoms 3 and 6
- Add group to atom 5: -OH

Pre-processing Step (Transform Library Building)



Templates Search



Strategy Parameters and Precursors Rating

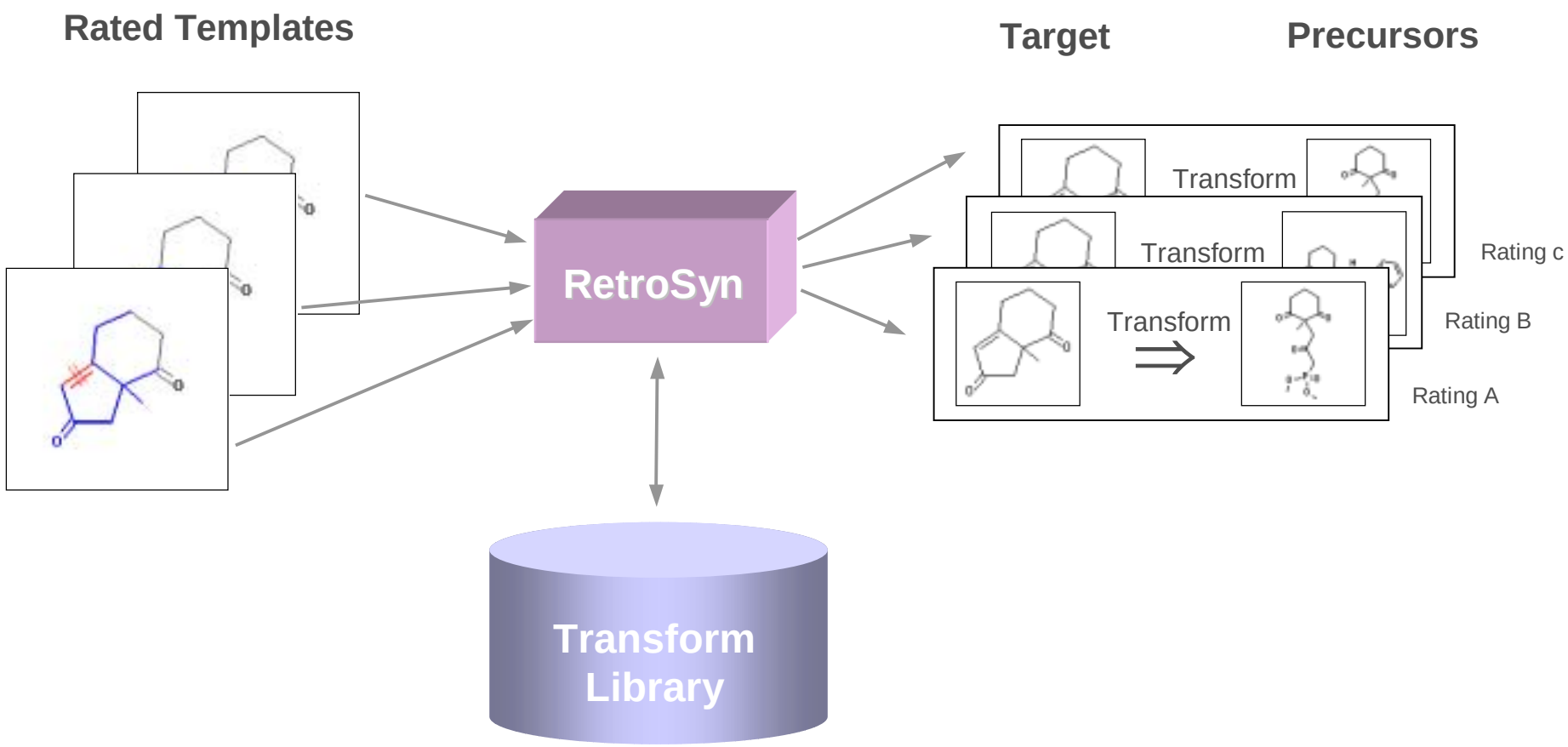
Strategy parameters for precursor-rating consider the following criteria:

- formation of a new carbon-carbon bond
- bond change within a functional group
- formation of a new ring/aromatic bond
- formation of a new bond between two symmetric fragments
- formation of a new bond near the center of the molecule
- number of generated precursors

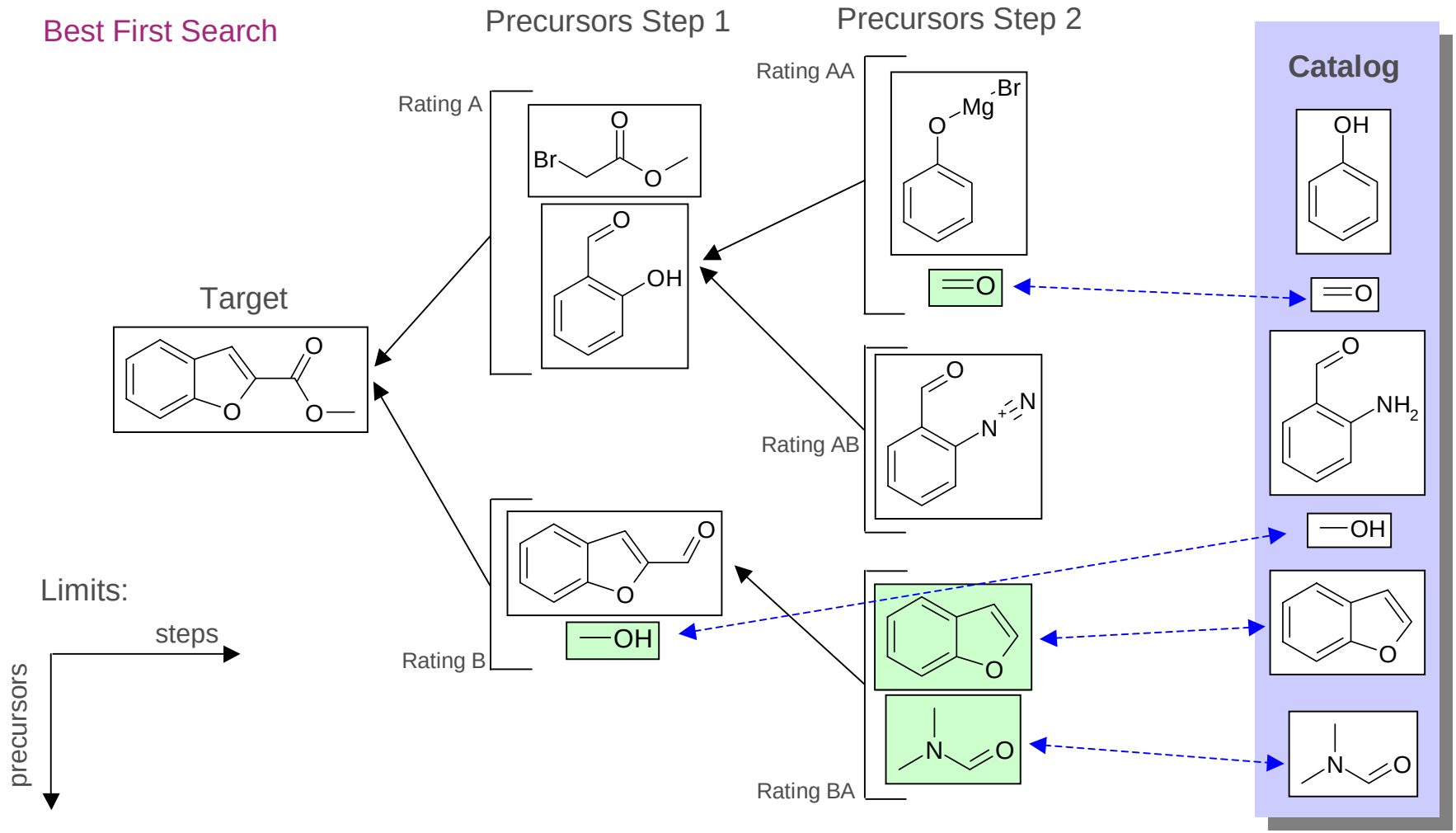
These criteria are weighted depending on the chosen configuration:

- sets of pre-defined strategy parameters are available
- sets may be defined from user

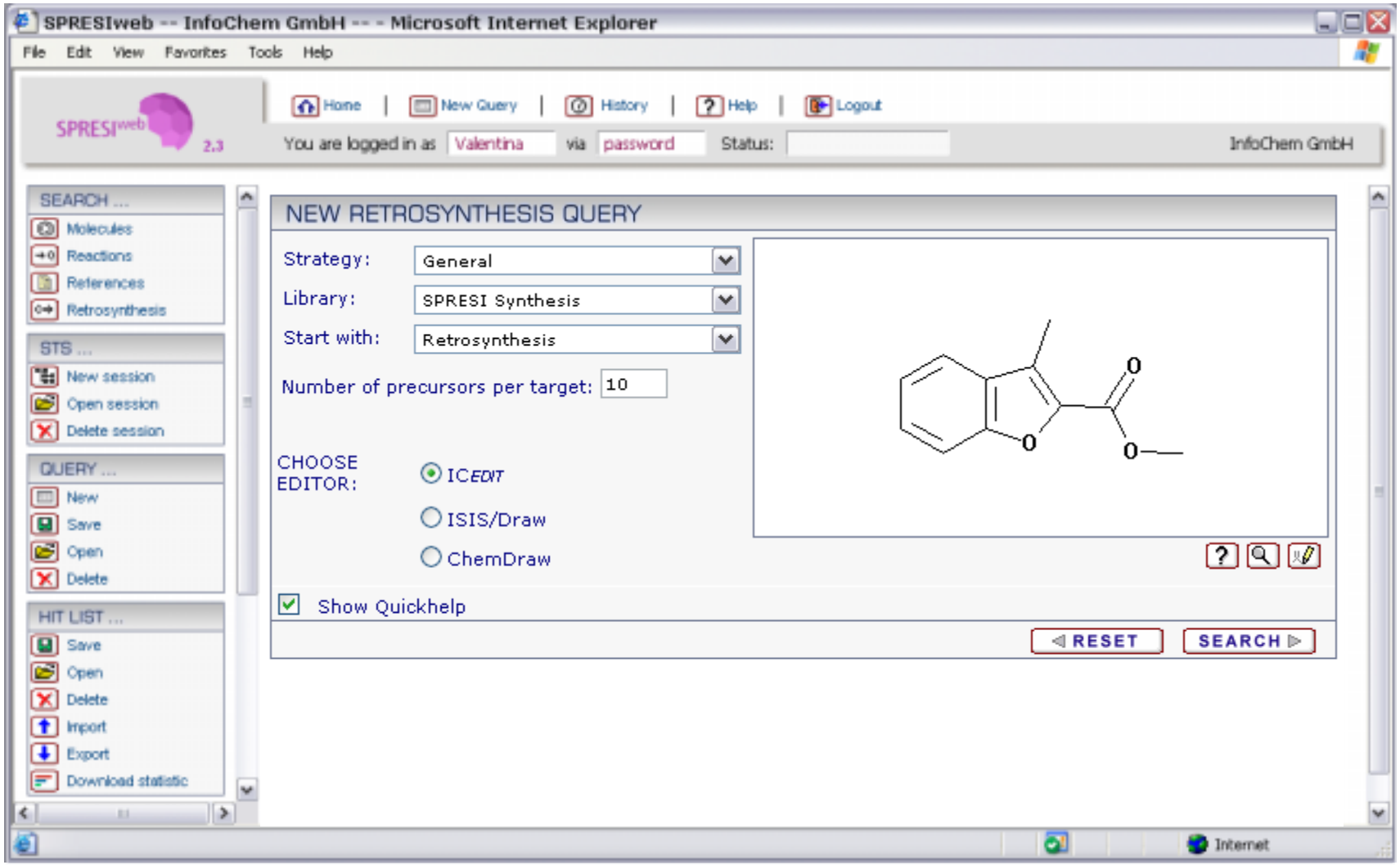
Precursors Generation from Target



Multistep Synthesis



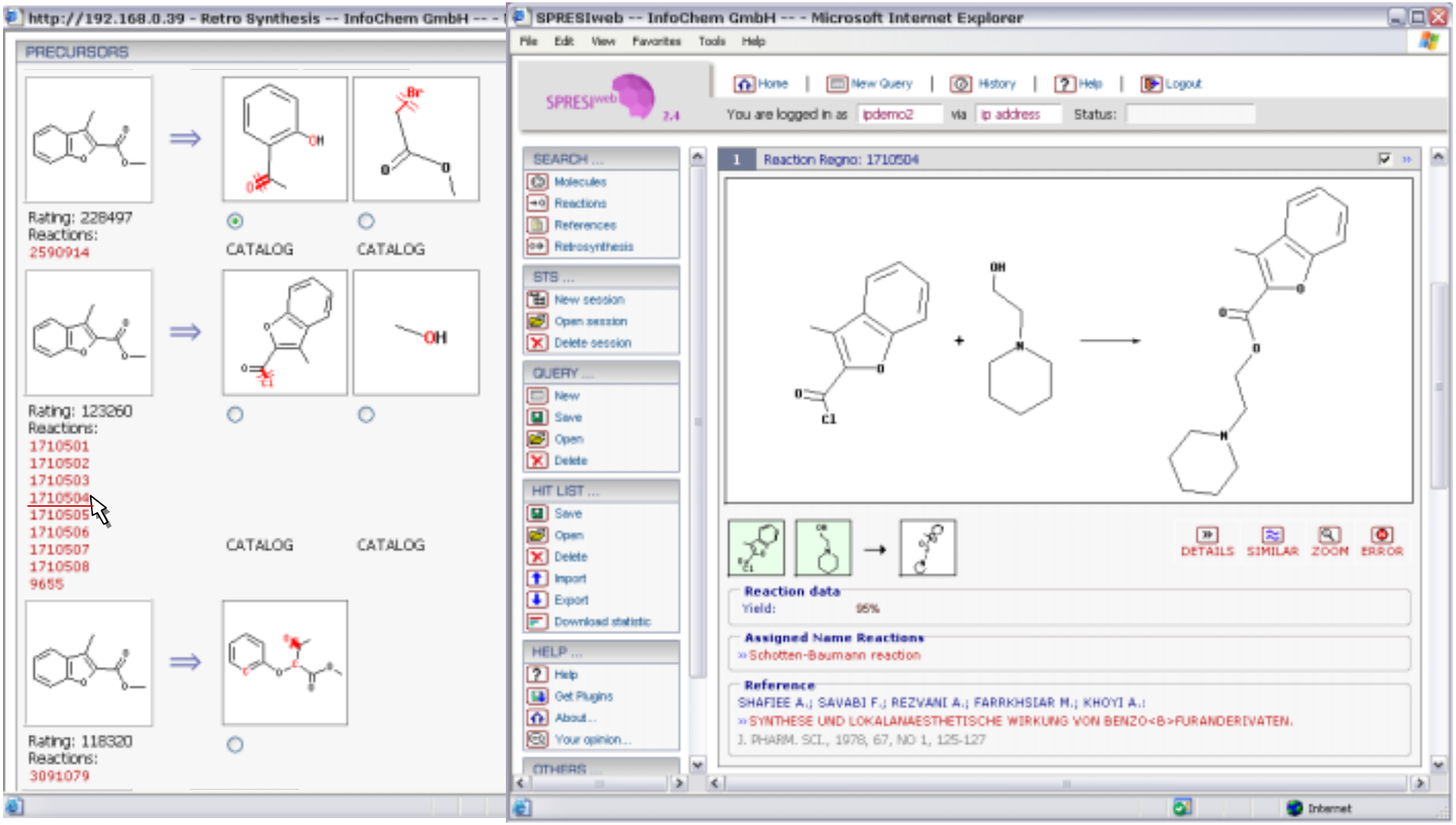
Prototype: Query Form



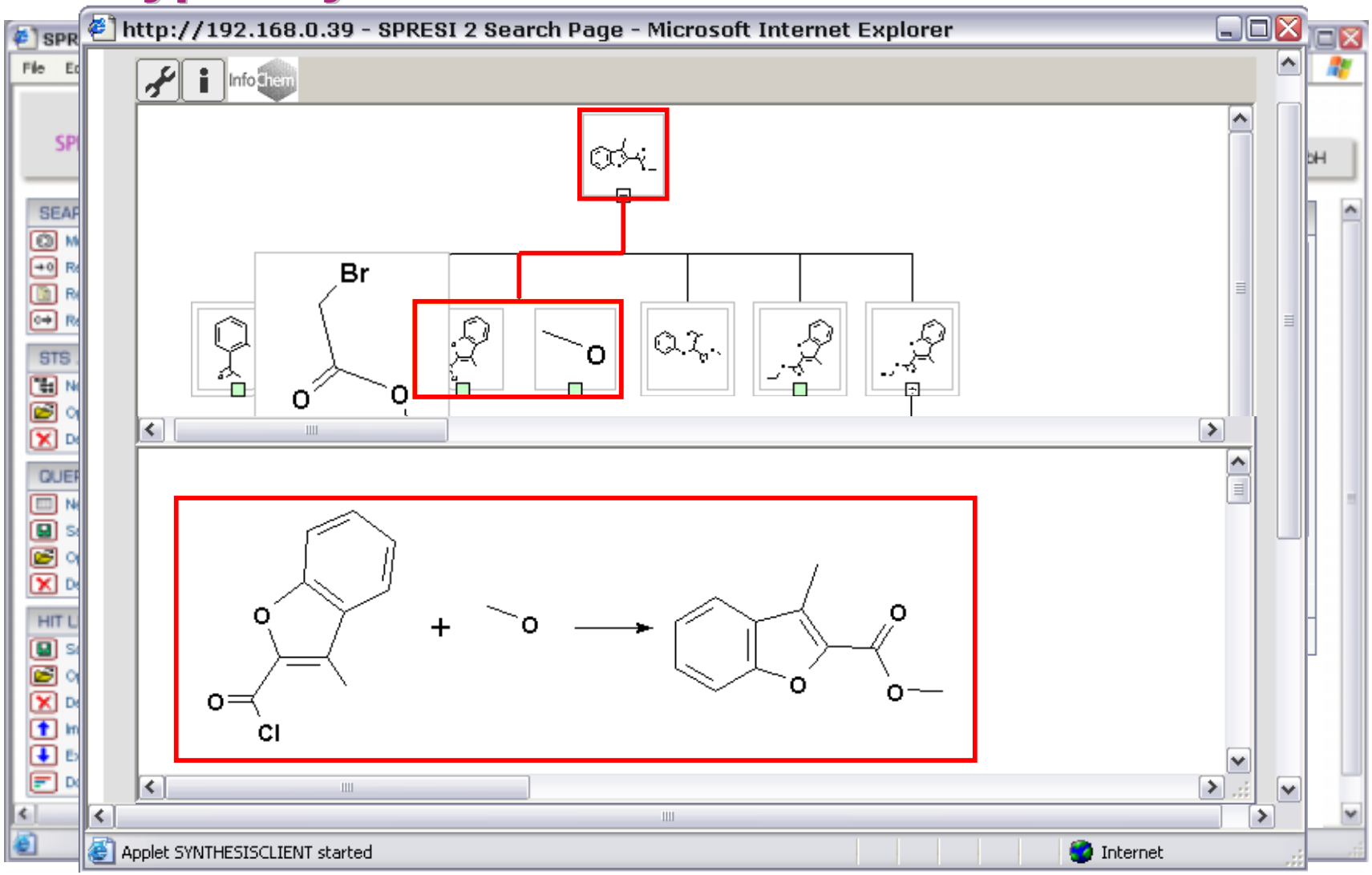
Reaction Databases Used in Prototype

- **SPRESI Synthesis**
  2.5 mio reactions from SPRESI
  reactions with 50% minimum yield
- **Name Reactions**
  140.000 reactions from SPRESI
  all examples for a large set of Name Reactions
- **EROS**
  68.000 reactions from J. Wileys
  „Encylopedia of Reagents for Organic Synthesis“
- **SOS**
  120.000 reactions from Thiemes
  „Science of Synthesis“
- **CAC**
  6.600 reactions from Springers
  „Comprehensive Asymmetric Catalysis“

Prototype: Suggested Precursors



Prototype: Synthesis Tree



Conclusions

- Very large pool of data available (SPRESI 4 million reactions)
- Automatic transform generation from reaction databases
- Suggestions based on real (analogous) reactions
- Verification of suggested synthesis reactions, through link to reaction database / literature
- Availability-check of precursors in catalogs
- System will be integrated in a retrieval system providing common search tools

Thank you!

InfoChem GmbH:

www.infochem.de
www.spresi.de
info@infochem.de