

I. Motivation

"As practitioners of organic synthesis can appreciate, visual contact with a given target molecule is primordial in the design of a synthetic strategy." -Hanessian

"The first taste is with the eyes" -Sophocles

Academic Research

Emphasis on creativity and learning

Industrial Research

Emphasis on time and economic constraints

"The academic sector can boast a plethora of trophies in achieving the highest summits of complex natural products synthesis. There is a great deal of personal pride and satisfaction in these Herculean feats, despite the sometimes arduous climb to the summit. Synthetic chemists are often individualistic and reluctant to abandon approaches conceived on paper, even in the face of adversity in the laboratory." -Hanessian

Are Computers Necessary?

"That there is a need for such an application is made apparent by the fact that a complete, logic-centered synthetic analysis of a complex organic structure often requires so much time, even of the most skilled chemist, as to endanger or remove the feasibility of this approach." E.J. Corey, 1969

II. Introduction

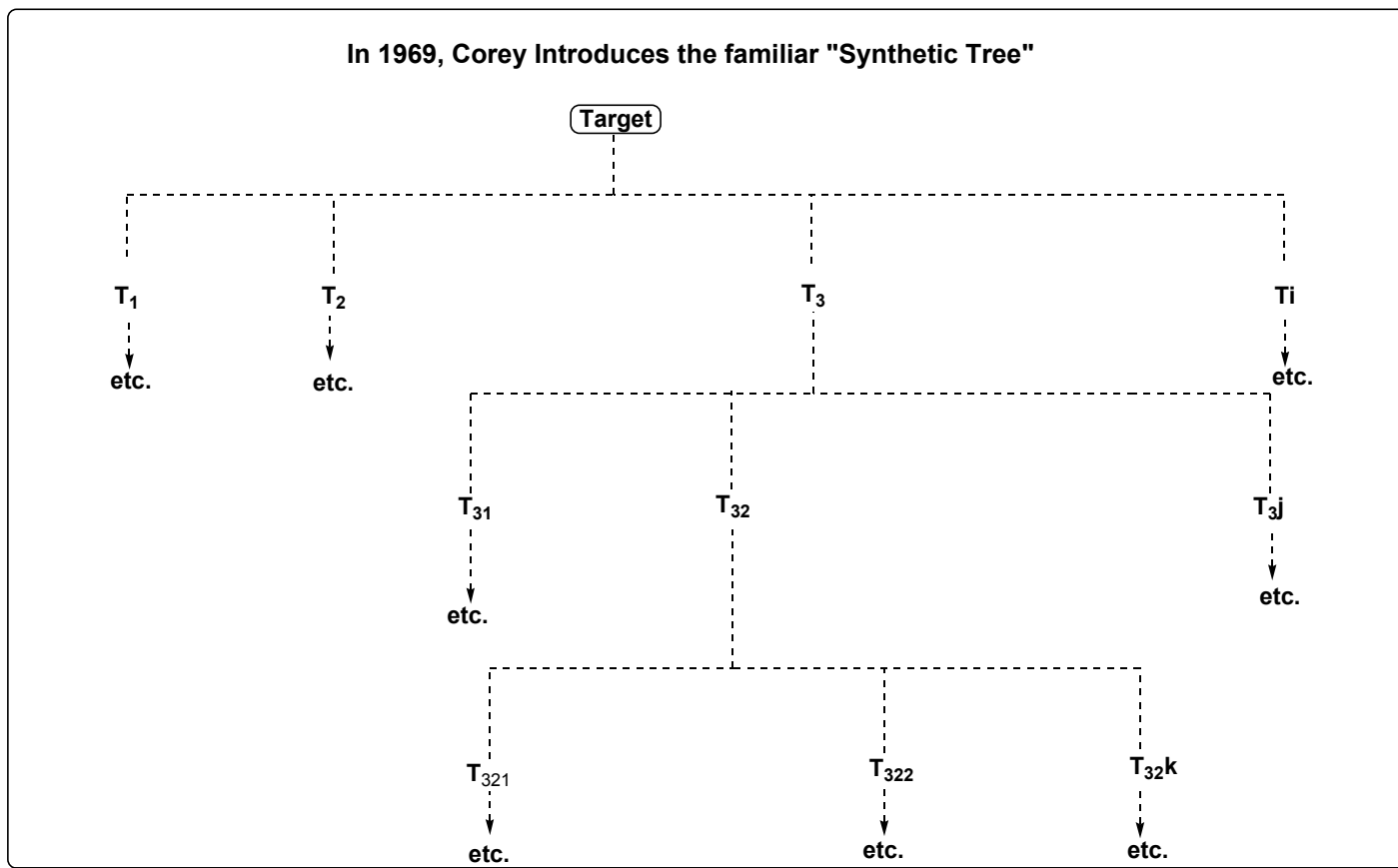
This presentation will attempt to survey both the history and current usage of computers in developing retrosynthetic strategies

programs/concepts discussed:

- | | | |
|------------|---------------|------------|
| 1. OCSS | 7. SESAM | 13. TRESOR |
| 2. LHASA | 8. SECS | |
| 3. SYNCHEM | 9. SST | |
| 4. SYNGEN | 10. SYNSUP-MB | |
| 5. WODCA | 11. KOSP | |
| 6. CHIRON | 12. LILITH | |

III. OCSS

- In 1969, the first major effort in developing a computer-based method for synthetic planning was reported by Corey and Wipke.
- The Program OCSS was the predecessor of the more well known LHASA program which is currently in its 18th version.



As any synthetic chemist can recognize, this tree could quickly (and likely) become far too large to efficiently interpret

III. OCSS

In 1969, Corey states that a successful program must be interactive and be capable of the following:

1. Generate trees which are limited in size, but incorporate as many useful pathways as possible
2. Allow for interruption by the chemist to re-direct the analysis at any time
3. That the depth of the search or analysis be decided by the chemist
4. That the evaluation of the various pathways be done by the chemist, but that the machine order the output structures in a way tantamount to preliminary evaluation

Thus the "logic-centered" part of the analysis is performed by the computer, while the more complex "information-centered" portion is left to the chemist

Where (How) Does Analysis Begin?

define structural features within the target which are of synthetic interest

- i. chains, rings, appendages
- ii. functional groups
- iii. asymmetric centers, and groups attached thereto
- iv. chemical reactivity, sensitivity, instability



Reduce in molecular complexity

Done through combinations of the following:

- i. scission of rings
- ii. disconnections of chains, appendages
- iii. removal of functionality
- iv. Modification or removal of sites of high chemical reactivity, instability
- v. simplification of stereochemistry, removal of asymmetric centers

Subgoals aid in simplification without themselves being simplifiers

- i. Functional Group Interconversion/introduction
- ii. introduction of groups for stereochemical or regiochemical control
- iii. internal rearrangement to modify rings, chains, func. Groups

As Corey states, *mechanism* is the most powerful technique for establishing a link between the perceived molecular features and the operations needed to simplify them. We will see that both mechanism based disconnections and functional group disconnections can be efficiently applied.

III. OCSS

1. Notation

Vocabulary consists of atoms (C, H, O, N, S, P, X), charges, and bond orders

Molecules are represented as graphs, with atoms being the nodes and bonds being the branches

2. Perception

The perception module contains algorithms to recognize Functional groups, rings, appendages, symmetry, stereochemistry, etc., as well as their relationships to each other

Electronic descriptor algorithms can be made to recognize electronic group properties (n or π electron withdrawing, donating, etc.)

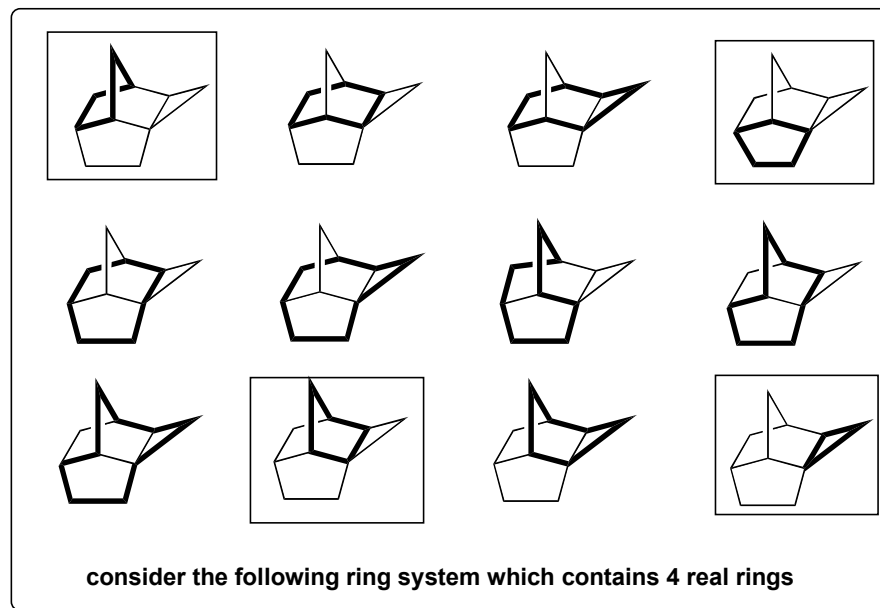
example:

Ring Perception Algorithm to Find all Cycles in a Chemical Graph.

1. algorithm arbitrarily chooses an atom as the origin and a path grows out along the molecular network, until the path doubles back on itself.
2. If the ring does not duplicate an already recorded one, it is placed in the ring list
3. If when all paths from the origin have been traversed all atoms in the structure have not been covered, then the structure consists of more than one fragment.
4. A new origin is chosen in the next fragment, and the process is repeated

The number of chemical rings is then given by $n_{\text{Realrings}} = n_b - n_a + n_f$, where n_b = number of bonds, n_a = number of atoms, and n_f = number of fragments in the structure.

Corey realized that a "real ring" is easily recognized by a chemist, and that all of the other "pseudo rings" are simply combinations of one or more real rings.



III. OCSS

3. Strategy and Control: Heuristics and The Human Element

Once the program has identified all perceptions (functional groups (with spacial orientation), rings, appendages, etc) the task of developing a strategy and goals begins

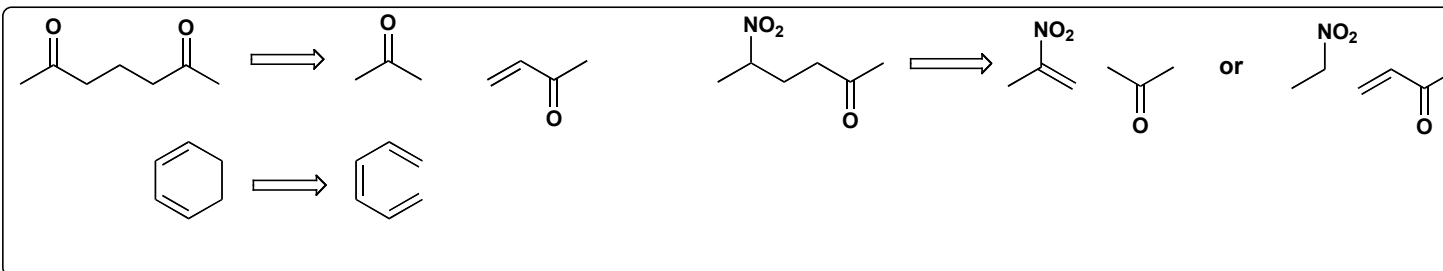
A set of fundamental Heuristics was developed by Corey by choosing the most general and powerful principles/reactions available in organic synthesis at the time. (which he admitted were largely incomplete at the time)

"The effectiveness of the heuristics referred to above is one of the critical factors in the performance and quality of a computer program for synthetic analysis." E.J. Corey

The Heuristics lead to the development of goals by the strategy module, and directly or indirectly to commands in the manipulation module.

Heuristic (Corey Definition): *noun*, meaning a "rule of thumb" which may lead by a shortcut to the solution of a problem, or may lead to a blind alley.

Heuristics can be categorized according to whether they relate primarily to functional groups, molecular skeleton, appendage groups, geometry (stereochemistry), or various combinations.



A transformation which serves more than one goal simultaneously will have a higher priority

4. Manipulations

The manipulation module performs the symbolic chemical transformations to create precursor structures.

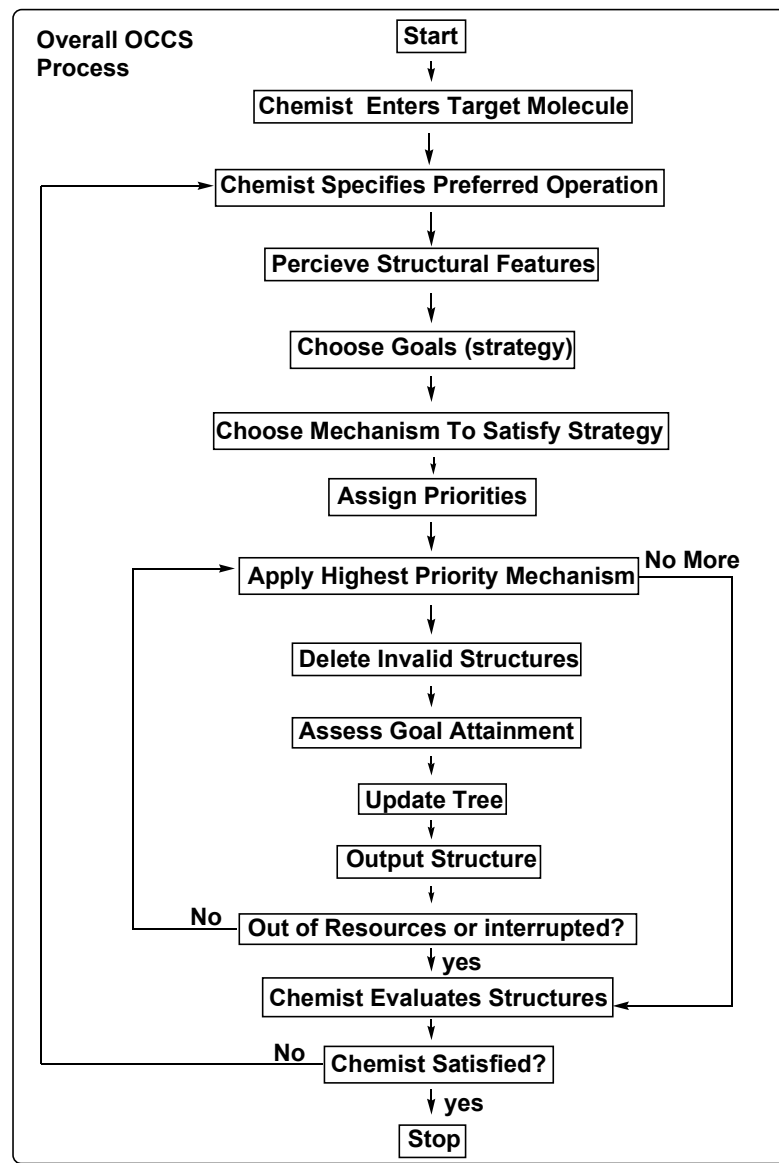
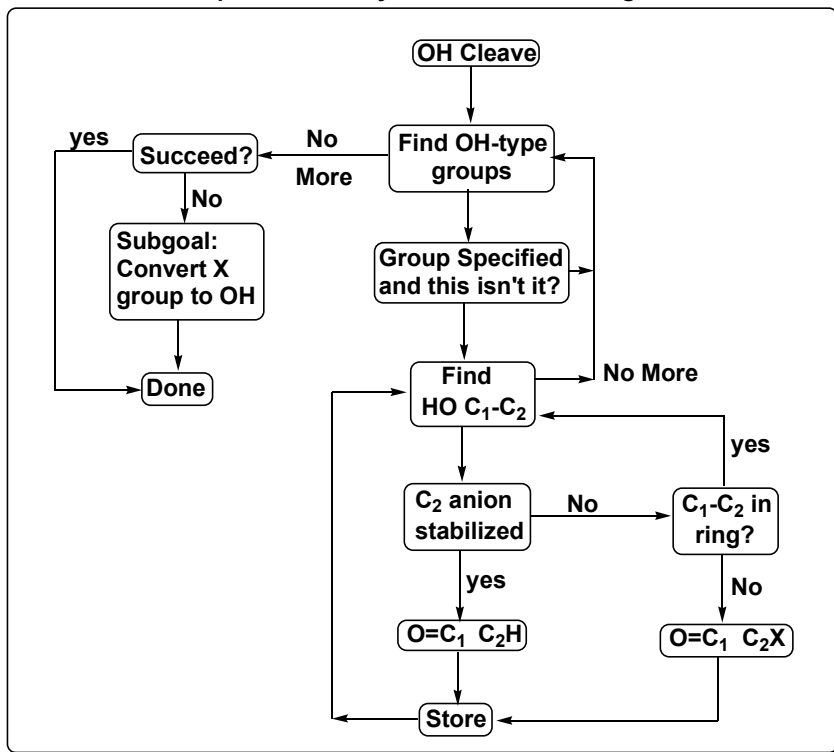
Two kinds of symbolic transformations are used *Symbolic mechanism*, and *symbolic functional group modification*

symbolic mechanism: may or may not actually be a complete chemical reactions

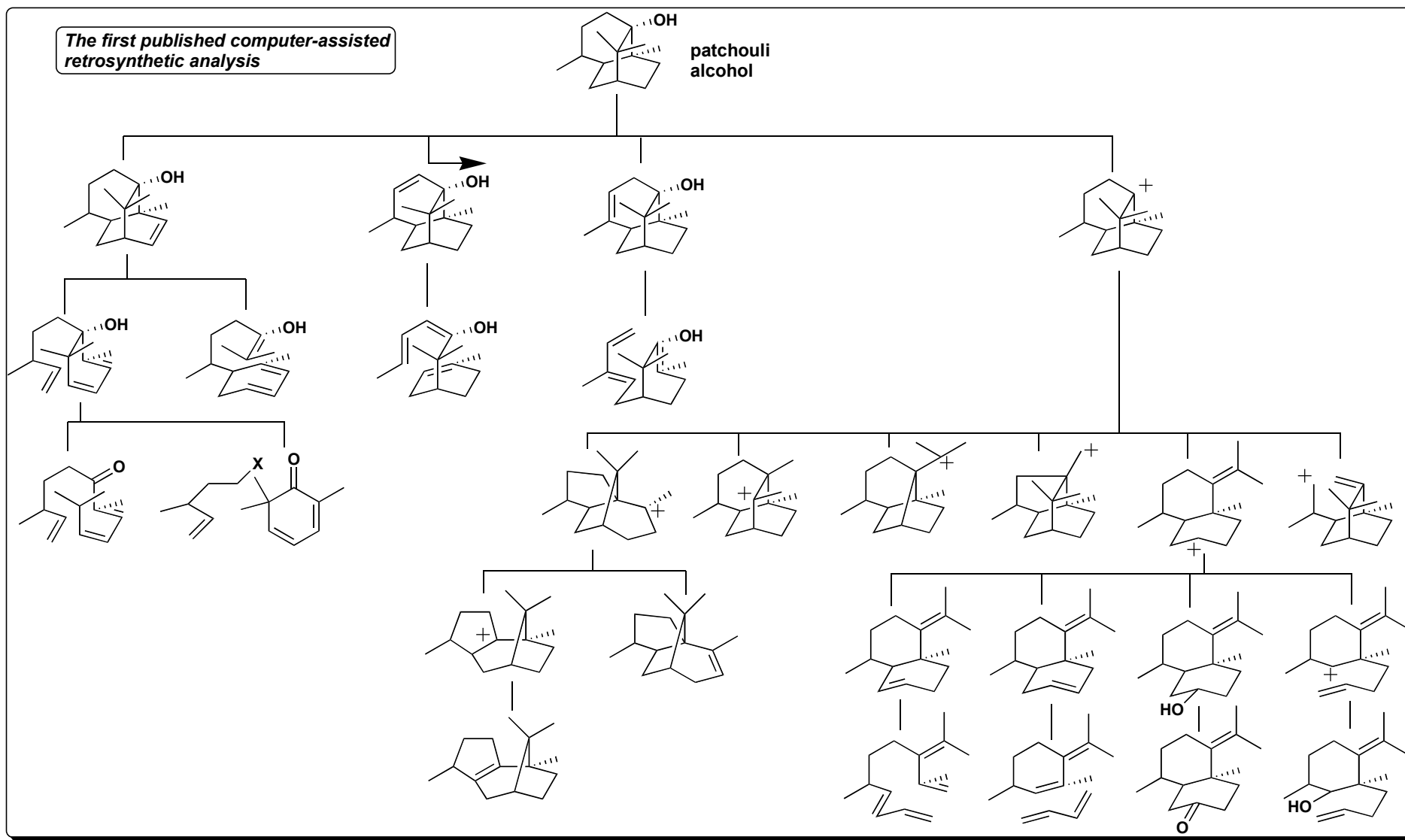
symbolic functional group modifications: results in the exchange, introduction, or removal of functional groups, but does not itself affect the skeletal connections in the molecule

III. OCSS Processing Scheme

Here is a Representative cycle for an OH-cleavage mechanism



III. OCSS in Action: Retrosynthetic Analysis of Patchouli alcohol



IV. Logic and Heuristics Applied to Synthetic Analysis (LHASA)

OCSS evolved into the more powerful and well-known program LHASA, which is in its 18th version

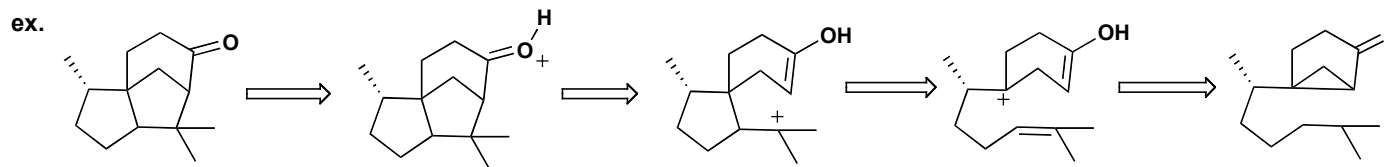
As of 1994, over 2000 applicable reactions were included in its database

The following Strategies are used by LHASA for retrosynthetic analysis (concurrent use of several strategies can be powerful):

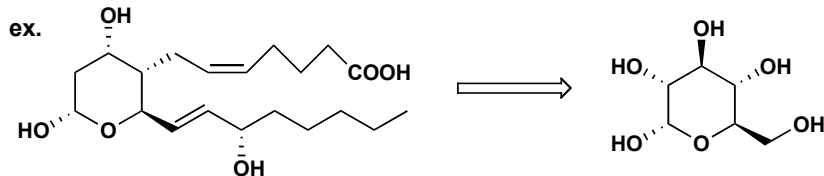
1. **Transform-based strategy:** Identification of powerful simplifying transformation, retron need not be present because program will look-ahead for applications

ex. Diels-Alder	Sigmatropic rearrangements
Robinson Annuulation	Photocyclizations
Cation π -cyclization	Diastereoselective π additions
Aldol cyclization	
Radical π -cyclization	

2. **Mechanistic Transforms:** Target is converted into a reactive intermediate and other intermediates of synthetic value can be generated

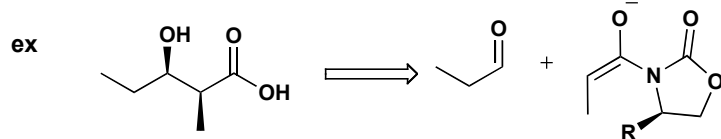


3. **Structure-goal (S-goal):** The identification of a potential starting material, building block, retron-containing subunit, or initiating chiral element



4. **Topological Strategies:** Identification of one or more bonds which can lead to major simplifications

5. **Stereochemical Strategies:** Stereoselective reactions, or steric based arguments are used to reduce stereocomplexity



IV. Logic and Heuristics Applied to Synthetic Analysis (LHASA)

6. **Functional group-oriented strategies:** one or more functional groups in a specific arrangement leads to logical disconnection. Functional group inter conversion/removal as well as functional group protection can be considered.

These 6 Strategies form the basis for nearly all retrosynthetic analysis programs

Differences between programs primarily occur because:

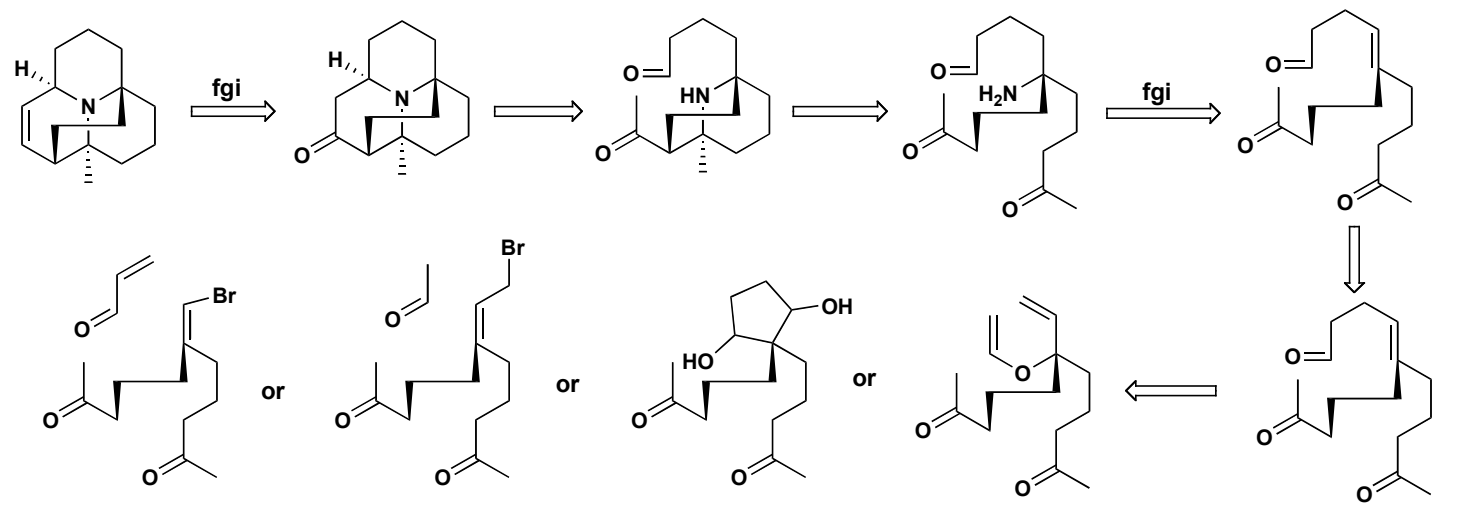
- Size of transformation databases differ (new reactions constantly be discovered)
- Different reactions given more priorities than others (based on historical success)
- ability to "long range" planning

One Key Note:

The **chemist** chooses what strategies and tactics are to be tried.

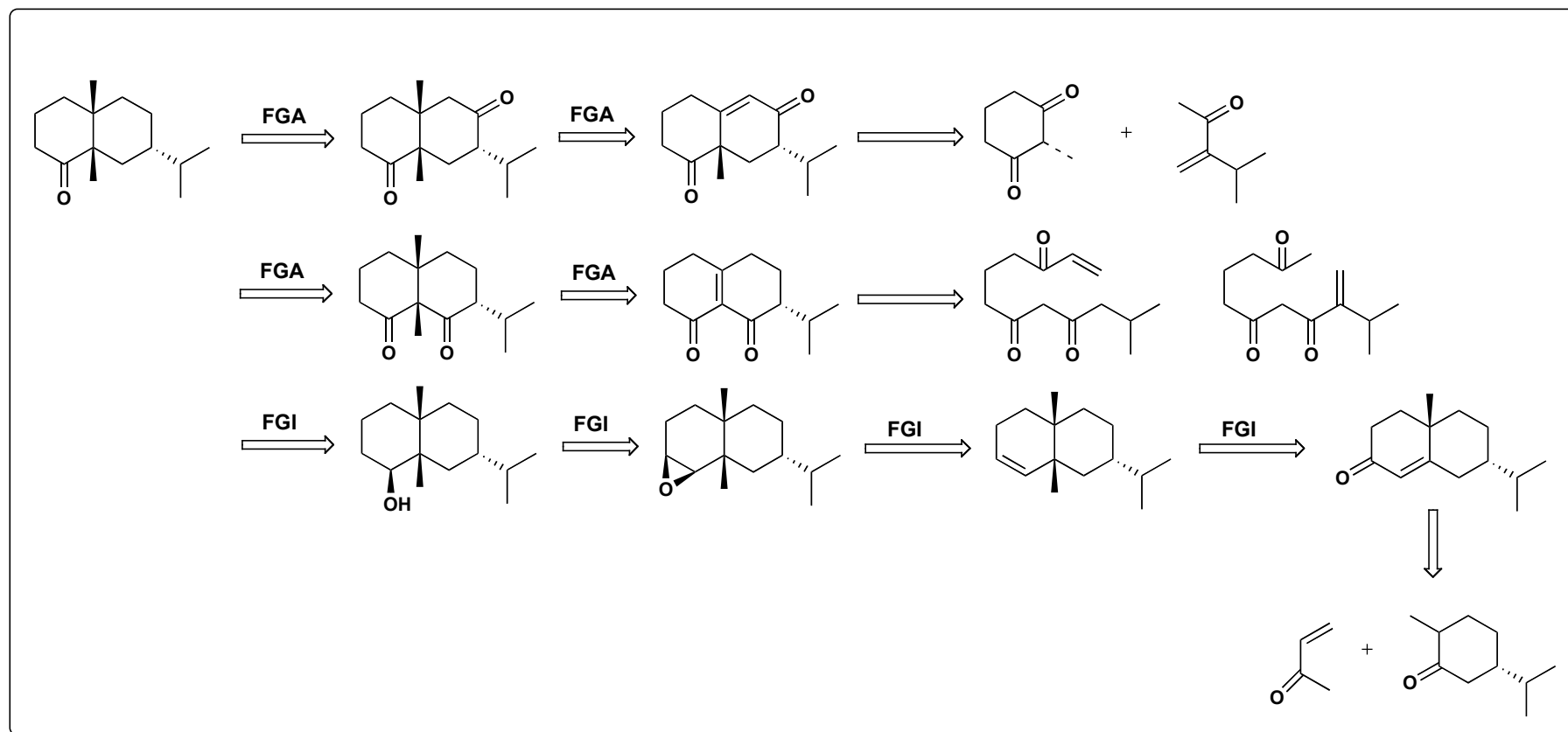
"LHASA WAS NOT DESIGNED TO INVENT CHEMISTRY THAT HAS NEVER BEEN PERFORMED IN THE LABORATORY." E.J. Corey

Functional-group oriented search in LHASA applied to Porantherine



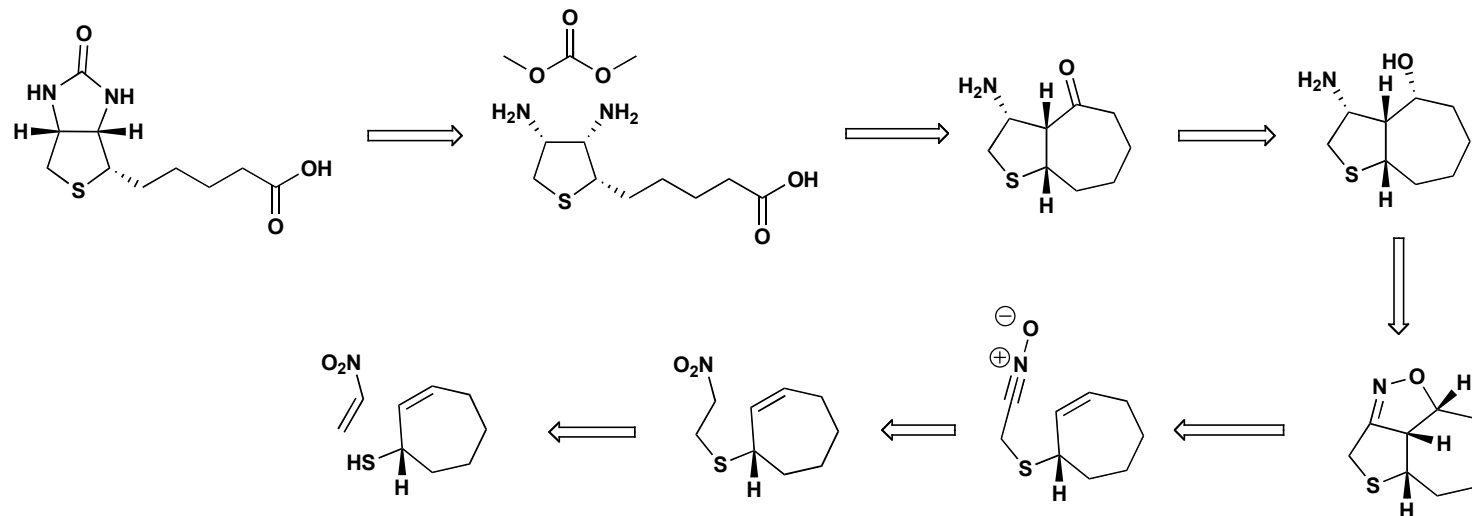
IV. Logic and Heuristics Applied to Synthetic Analysis (LHASA)

Transform-based (Robinson Annulation) search in LHASA applied to Valeranone

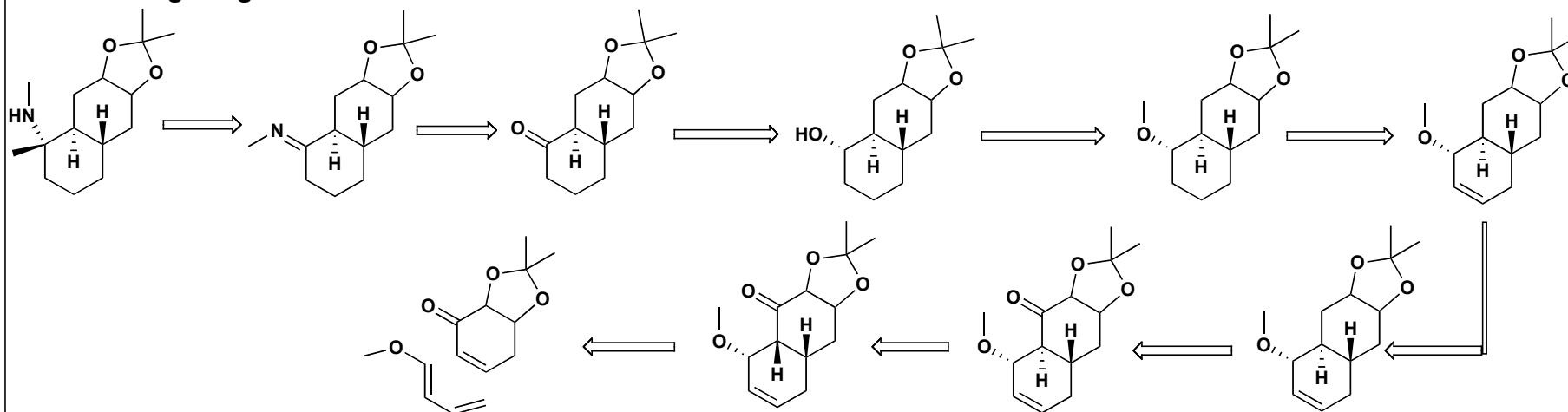


IV. Logic and Heuristics Applied to Synthetic Analysis (LHASA)

LHASA retrosynthetic analysis of biotin



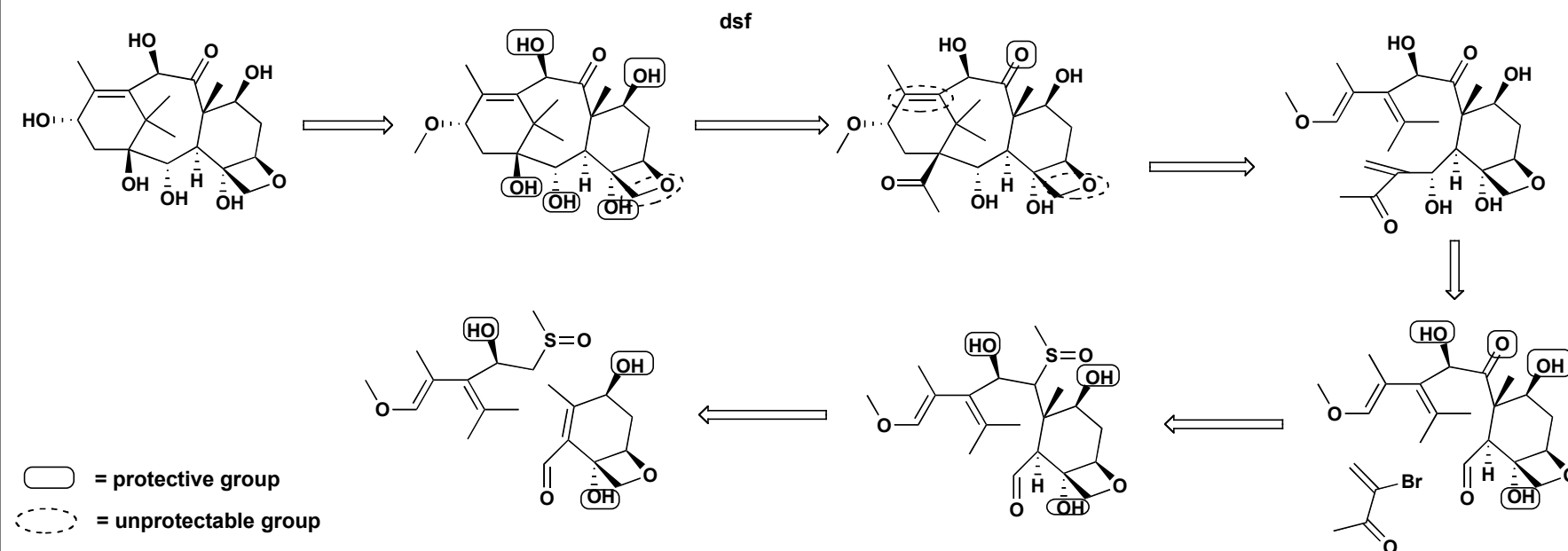
LHASA "long-range" Diels-Alder Transform



IV. Logic and Heuristics Applied to Synthetic Analysis (LHASA)

LHASA Partial Retrosynthetic Analysis of Taxol Alcohol

Take note of the programs familiarity with reaction conditions which allows it to specify which groups need to be (or can be) protected for each individual transformation



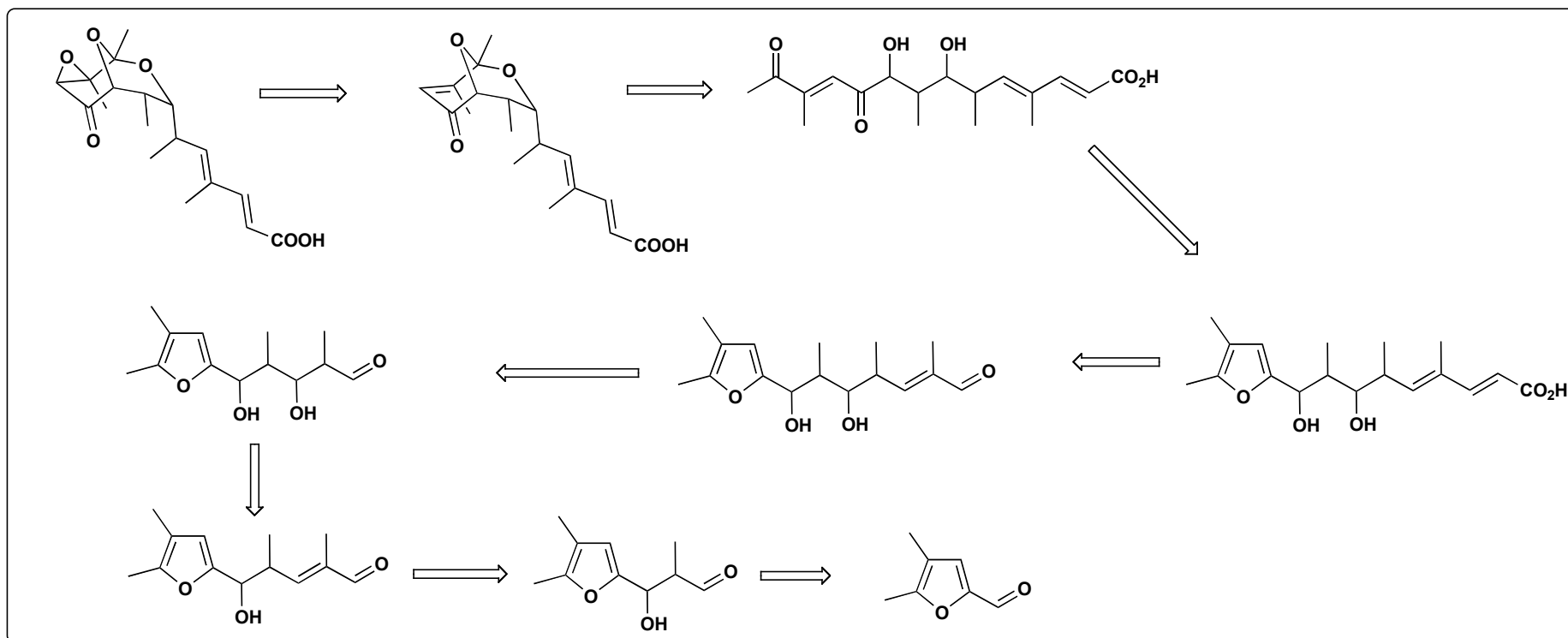
V. SYNCHEM

- SYNCHEM* is also a Heuristic search program for retrosynthetic analysis. It was introduced by Gelernter (1977) and co-workers shortly after LHASA.
- originally contained the aldrich catalog of starting materials (~3000 compounds) and could not deal with stereochemistry
- Newer Versions contain over 5000 compounds, over 1000 reaction schemes, and can handle stereochemistry

An important note:

SYNCHEM, unlike LHASA, was developed to be self-guided and not dependent on the chemists suggestions.

An early Retrosynthesis of tirandamycic acid



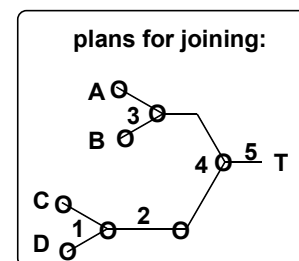
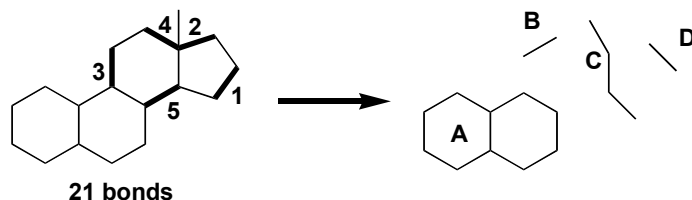
VI. SYNGEN

- The concept for the SYNGEN program was outlined by Hendrickson in 1971
- The focus of the program was on skeletal construction, because the best and shortest syntheses consists of only construction reactions.

bondset: a bondset in a target skeleton is a set on bonds λ which need to be constructed.

"Synthesis itself is a skeletal concept." -Hendrickson

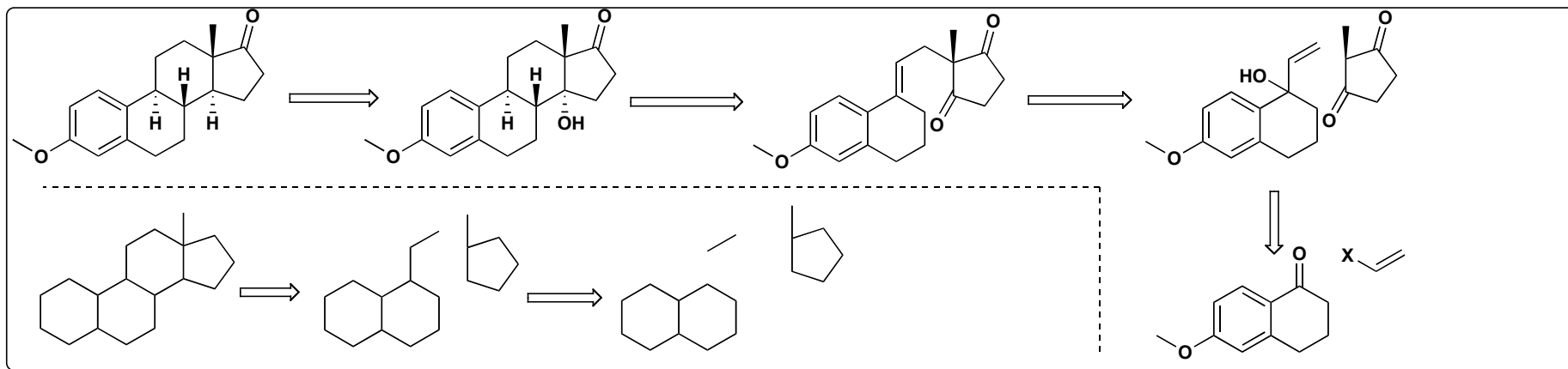
consider the following skeletal disconnection:



The number of possible bondsets is the number of ways to dissect the skeleton: for a target of b bonds, there are $b!/(b-\lambda)!$ ways.

Constructing the above skeleton ($b=21$) from pieces averaging 3 carbons, would require $\lambda = 9$ and there would be 100,000,000,000 routes, if one carbon units are used there would be 6×10^{23} assembly plans.

Using a complicated algorithm, SYNGEN will strive for convergent assemblies of fairly large pieces, using starting materials in its database (over 6000 compounds in 1990)

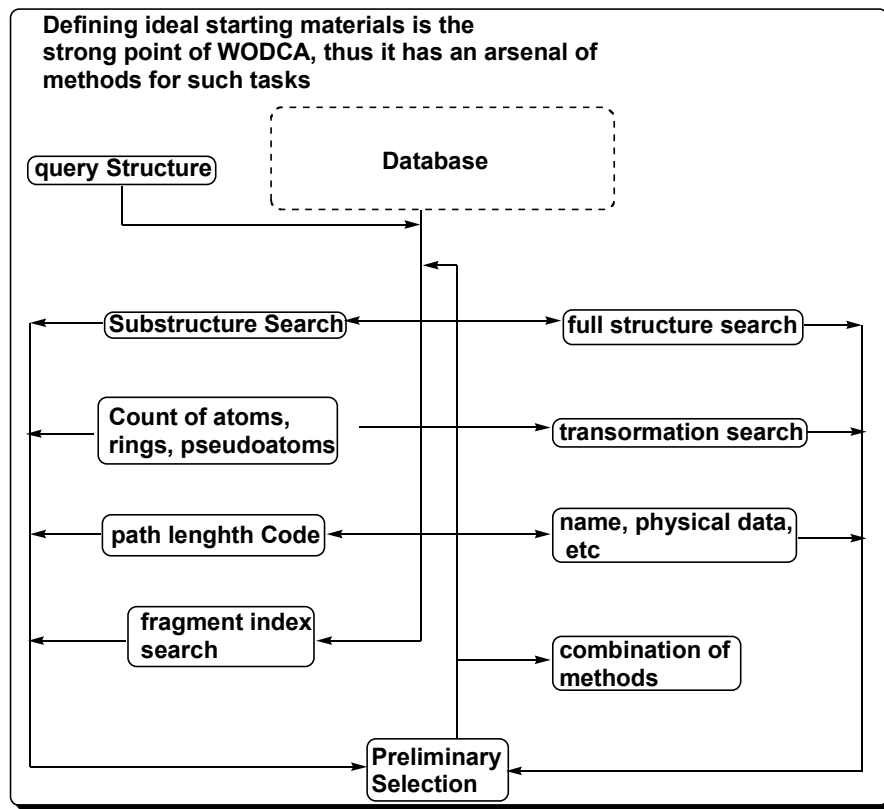


VII. Workbench for the Organization of Data for Chemical Applications (WODCA)

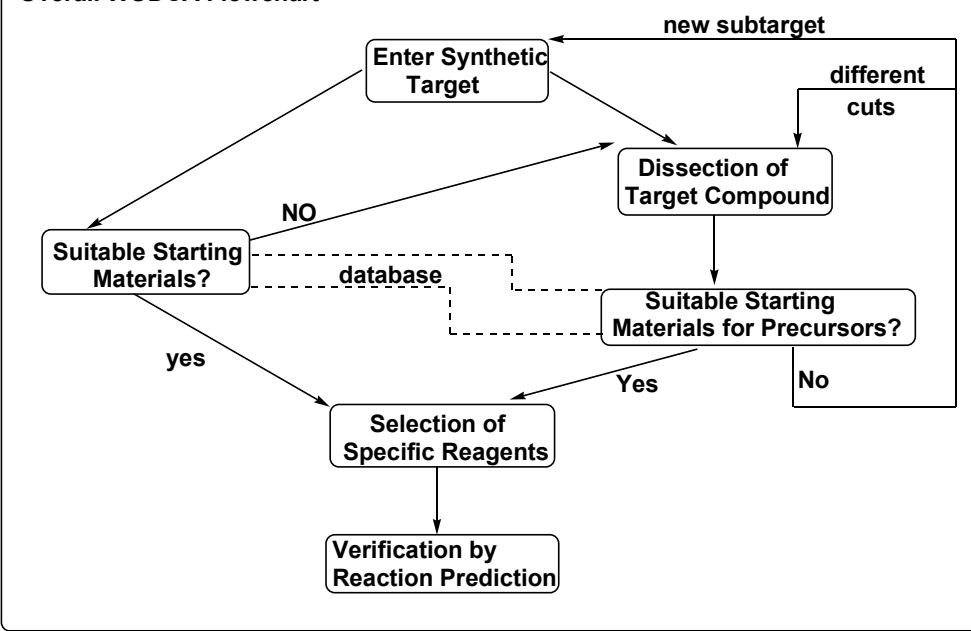
- First described by Gastieger, and Ihlenfeldt in 1995, WODCA seeks to improve upon the "first Generation" Programs
- It is part of a large (>8000 compound) database, and can be linked to the CHIRON (more to come) database of 2000 chiral compounds

The Three Core Operations are:

1. Search for starting materials
2. Search for synthesis precursors
3. Prediction of reactions

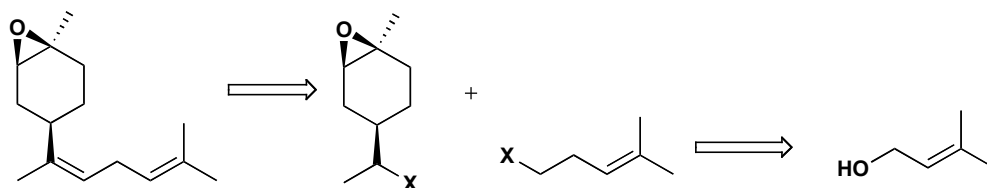


Overall WODCA Flowchart

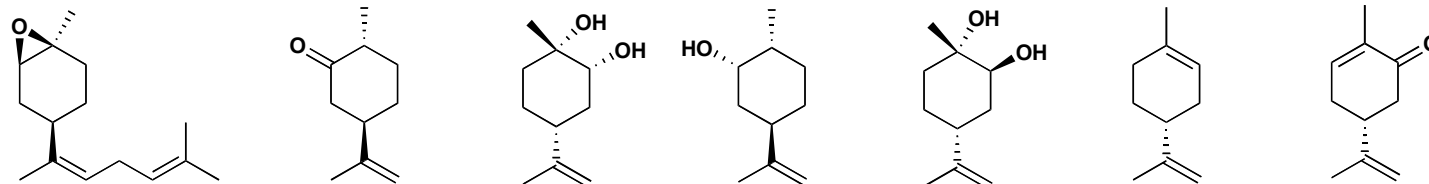


VII. Workbench for the Organization of Data for Chemical Applications (WODCA)

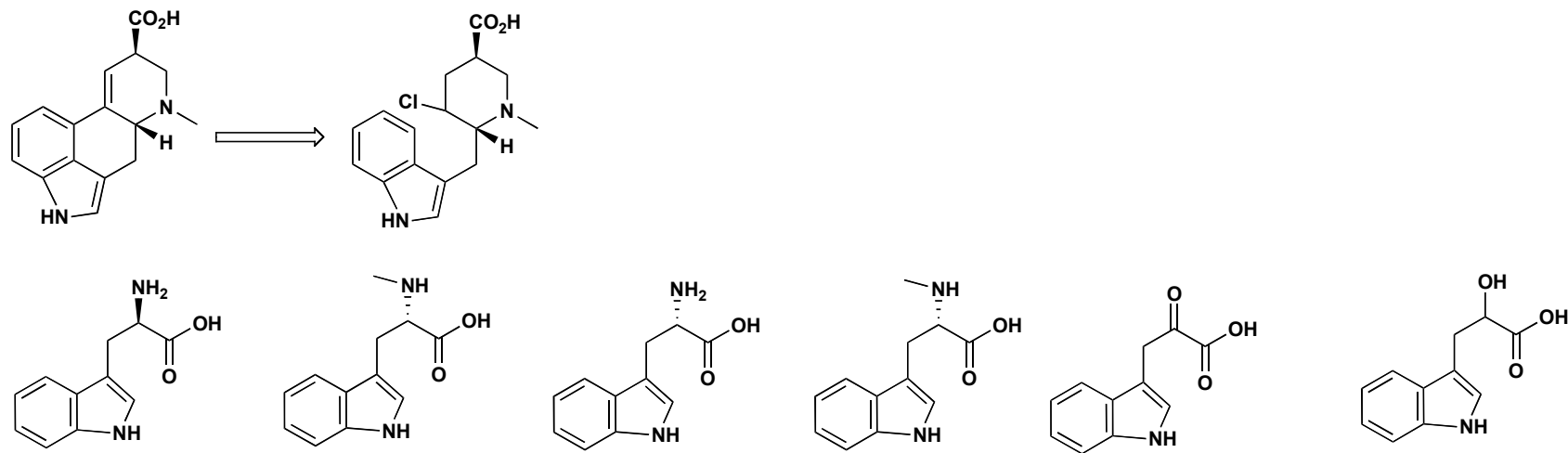
WODCA disconnection of α -bisabolene



WODCA/CHIRON-suggested precursors to α -bisabolene



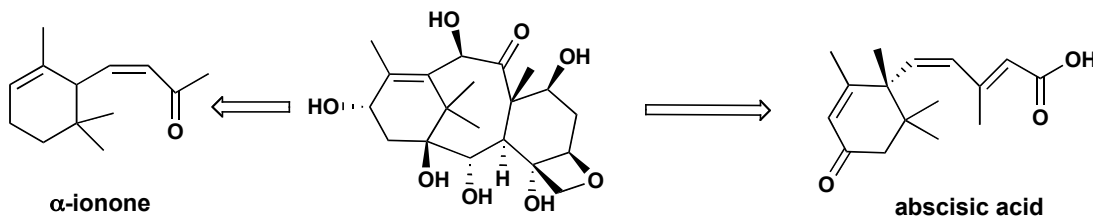
WODCA Search for lysergic acid precursors



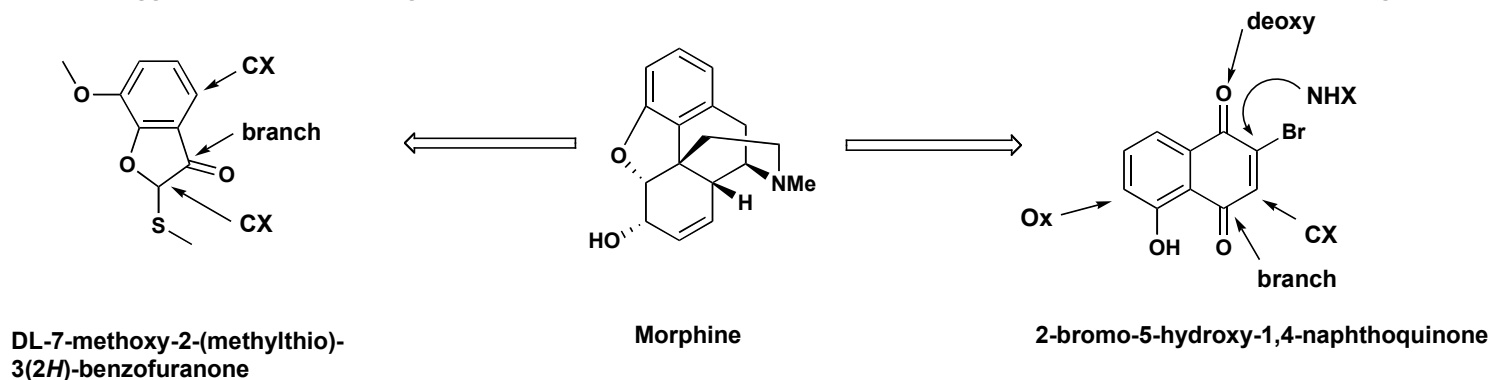
VIII. Chiral Synthon (CHIRON)

1. Developed by Hannessian as a program that could recognize chiral substructures in a target molecule, as well as their access from the chiral pool.
2. Currently in 5th edition, with a database of over 200,000 compounds including all commercially available compounds, over 5000 handpicked literature compounds, and over 1000 medicinally active compounds.

CHIRON suggests α -ionone and abscisic acid as precursors for Taxol alcohol based on skeletal overlaps

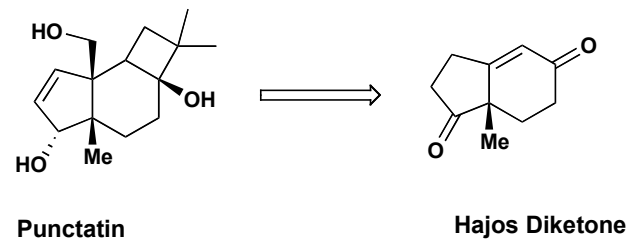
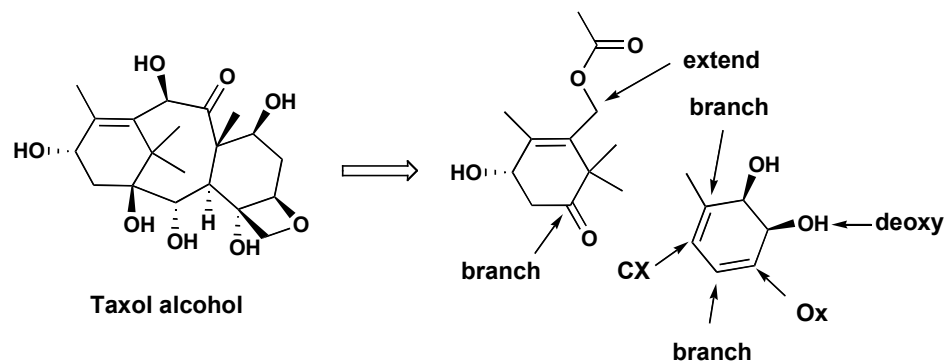
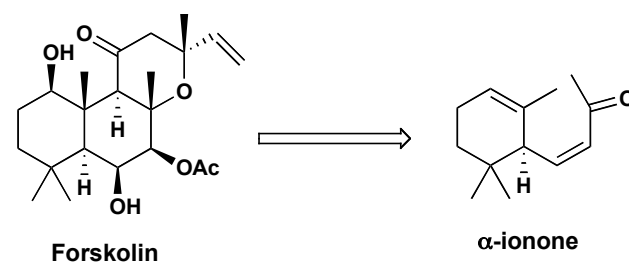
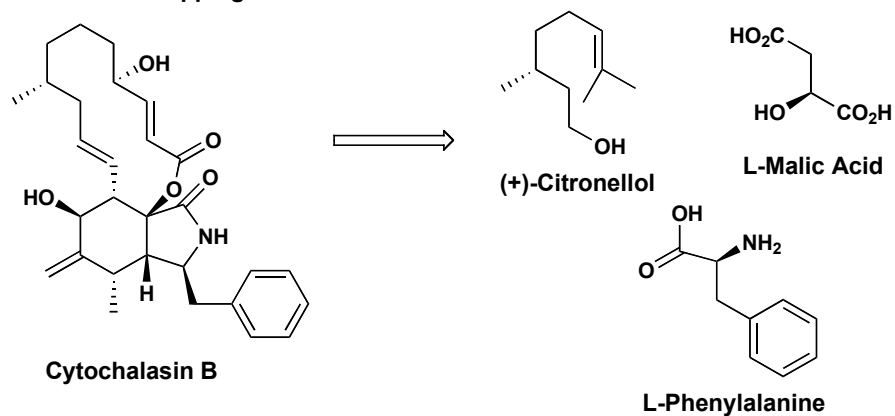


CHIRON can suggest unobvious starting materials as well as the transformations needed to convert them into the desired target



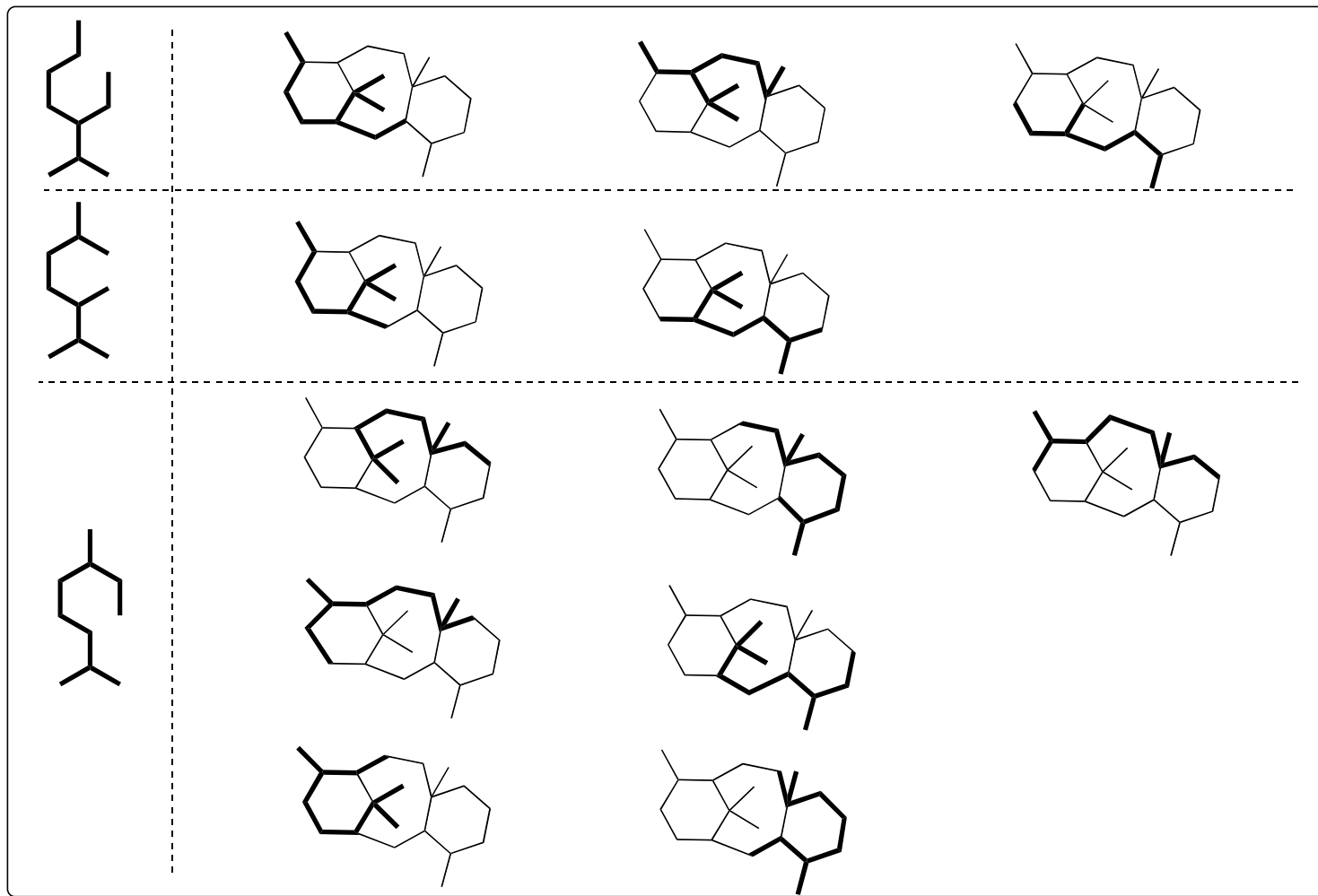
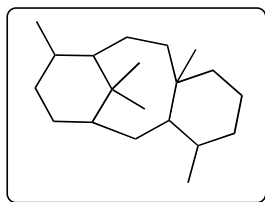
VIII. Chiral Synthon (CHIRON)

CHIRON has a Rapid Scanning of Best Combinations (RSBC) option that displays groups of precursors that match different parts of the molecule without themselves overlapping



IX. Search for Starting Materials (SESAM)

- Developed by Barone and Chanon as a tool for identifying synthons based on skeletal overlaps with potential starting materials,
- well-suited to terpene skeleton recognition
- program does not consider functional groups



X. Simulation and Evaluation of Chemical Synthesis (SECS)

- Developed by Wipke, a heuristic program developed after--and similar to--LHASA
- Significant effort was placed on stereochemistry, topology, and energy-minimization

XI. Starting Material Selection Strategies (SST)

- Also developed by Wipke, SST seeks to match starting materials to a given target by pattern recognition.
- Can generate pathways based on three ideas:
 1. constructive synthesis: SM can be directly incorporated in target
 2. degradative synthesis: Significant modifications of the SM are needed for incorporation
 3. remote relationship synthesis: several bond-forming/bond-cleaving operations must be performed
- routes are assigned scores for the ability of the SM to be efficiently mapped onto the target

XII. SYNSUP-MB

- Developed by Sumitomo Chemical Co., a heuristic program with a 2500 reaction database. 22,000 reactions can be simulated in about 1 hour on moderately complex molecules (5-10 fg's, multiple stereocenters)
- user places constraints on routes (max number of steps, etc) and a search is conducted without the input of the chemist.

XIII. Knowledge base-Oriented System for Synthesis Planning (KOSP)

- Developed by Satoh and Funatsu, retrosynthesis is planned based on reaction databases
- disconnections are made to place leaving groups at strategic bonds, and the process is repeated.

XIV. LILITH

- Heuristic program developed by Sello and co-workers.

XV. TRESOR

- Heuristic program developed by Moll, introduced in 1994.

Wipke, W.T.; Ouchi, G.I.; Krishnan, S. *Artif. Intell.* **1978**, *11*, 1-2, 173-193
Satoh, K.; Funatsu, K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 2, 316-325
Barone, R.; Chanon, M. *Eur. J. Org. Chem.* **1998**, *18*, 7, 1409-1412

Wipke, W.T.; Rogers, D. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 2, 71-81
Sello, G. J. *Chem. Inf. Comput. Sci.* **1994**, *34*, 120-129
Moll, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 117-119