Retrosynthetic Analysis

[References:
(1) Nobel Lecture of Prof. E.J.Corey (1990)
(2) Material from the website of Centre for Molecular and Biomolecular Informatics, University of Nijmegen, Toernooiveld1, P.O.Box 9010, 6500 GL Nijmegen
(3) Lecture notes of Prof. Gareth Rowlands, University of Sussex at http://www.sussex.ac.uk/users/kaff6, retrosynthesis 2001.
(4) "Retrosynthetic Analysis" – Lecture delivered by Dr. S. Chandrasekharan (Chairman, Dep. of Organic Chemistry, IISc., Bangalore) at U. C. College, Aluva on 7-11-2001.
(5) Willis C and Wills M. Organic Synthesis – Oxford Chemistry Primer series No.31. Oxford University Press (1995).
(6) Carruthers W. Modern methods of Organic Synthesis. 3rd edn. Cambridge (1996).
(7) House H.O. Modern Synthetic Reactions. 2nd edn. Benjamin/Cummings (1972).
(8) Norman R.O.C. Principles of Organic Synthesis. 2nd edn. Chapman & Hall (1978).
(9) Ireland R.E. Organic Synthesis. PHI (EEE) (1975)]

Importance of Organic Synthesis: A very large number of compounds with unimaginably complicated structures are produced by plants, animals or microorganisms. Examples are antibiotics, alkaloids, rubber, chlorophyll, steroids, proteins, carbohydrates, fats, vitamins, dyes and perfumes. These are called natural products. The chemist isolates, purifies, analyses and determines the structures of these compounds. Then people find various applications for these substances such as in medicine, plastics, paints, textiles, electronics etc. Then the synthesis of these compounds becomes important for the following reasons:

(1) Synthesis from simple starting materials of *known* structure using *familiar* reactions with *predictable* **regioselectivity** (in which part of the molecule will the reaction take place?) and **stereochemistry** (what will be the spatial arrangement of the atoms and groups?) and demonstrating that the product obtained is *identical with the one in nature* becomes the *ultimate proof of the structure* of the compound.

(2) The natural product is available only in very small quantities and will be very costly. Synthesis provides a means for the cheap mass production of the compound and makes it available for research and use for the benefit of man, for example in medicines.

(3) Structural variations made in the natural product may provide molecules which are more active, more useful or with lesser side effects.

(4) Syntheses of complicated molecules require good planning derived from a deep knowledge of the various reactions and their mechanism in great detail. It is extremely challenging and provides immense *intellectual satisfaction* to the chemist as a test and *proof of his ability*.

Till the 1950s, most syntheses were developed by selecting an appropriate starting material after a trial and error search of a few *commercially available* starting materials having *structural resemblance* to the molecule to be synthesized. Attempts were then made to convert them to the required product through a handful of well-known reactions using common laboratory reagents. The total synthesis of structurally complex compounds is a challenging undertaking, in intellectual as well as practical respects. Whereas simple compounds can usually be made by synthetic routes comprising of a few reaction steps (say two to five), complicated molecules may require a lengthy sequence of reactions, usually more than twenty. Most such multi-step syntheses are executed, or at least attempted, according to a plan designed beforehand on paper or blackboard. How do chemists arrive at such synthetic plans? Traditionally, synthesis design was based upon associative thinking processes, the most important of which were:

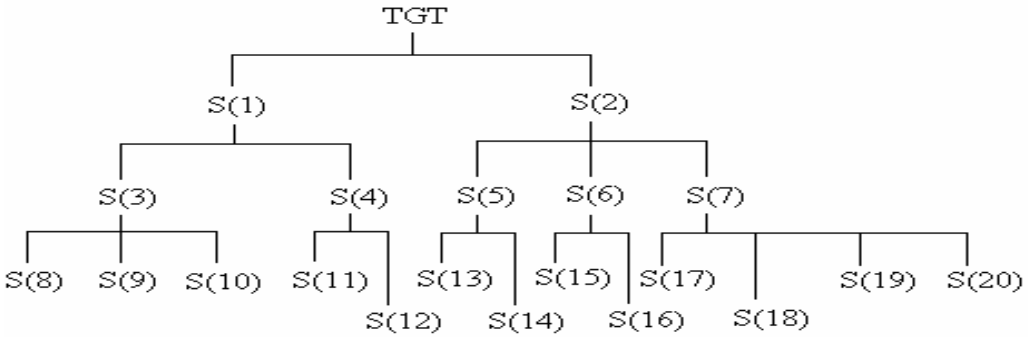
- Association with existing syntheses of similar compounds
- Association with known starting material(s)
- Association with a hypothetical advanced intermediate

The last of these represents an attempt to reduce the complexity of the design problem, but the selection of a suitable intermediary structure is still a highly intuitive process. The associative approach to synthesis design becomes less practical as the complexity of the problem, and hence the number of steps required, increases. There had been no systematic planning or visualization of the synthetic steps required to reach the target material *economically*.

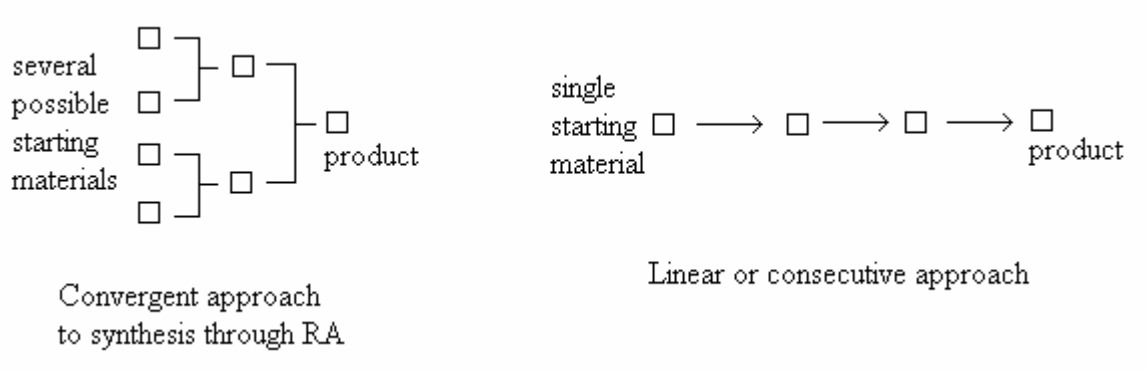
But a look at some of the complex molecules that were synthesized in the last century (given on the previous page) will indicate the difficulty in recognizing any available starting material. This called for a more systematic method for recognizing simpler molecules from which the required product could be synthesized in a small number of steps. Such a systematic method was developed by Prof. Elias J. Corey of the Harvard University and was called **Retrosynthetic Analysis** (RA). The method was so logical and efficient that not only students of chemistry, but even machines (computers) could be taught (programmed) to find out reasonable synthetic routes to obtain complex molecules. A few of these programs such as LHASA (Logic and Heuristics (= empirical rules) Applied to Synthetic Analysis), AIPHOS (Artificial Intelligence for Planning and Handling Organic Synthesis), COMPASS (Computer Assisted Organic Synthesis), OSET (Organic Synthesis Exploration Tool), SECS (Simulation and Evaluation of Chemical Synthesis), IGOR (Interactive Generation of Organic Reactions) and CHIRON (Chiral Synthon) are now commercially available.

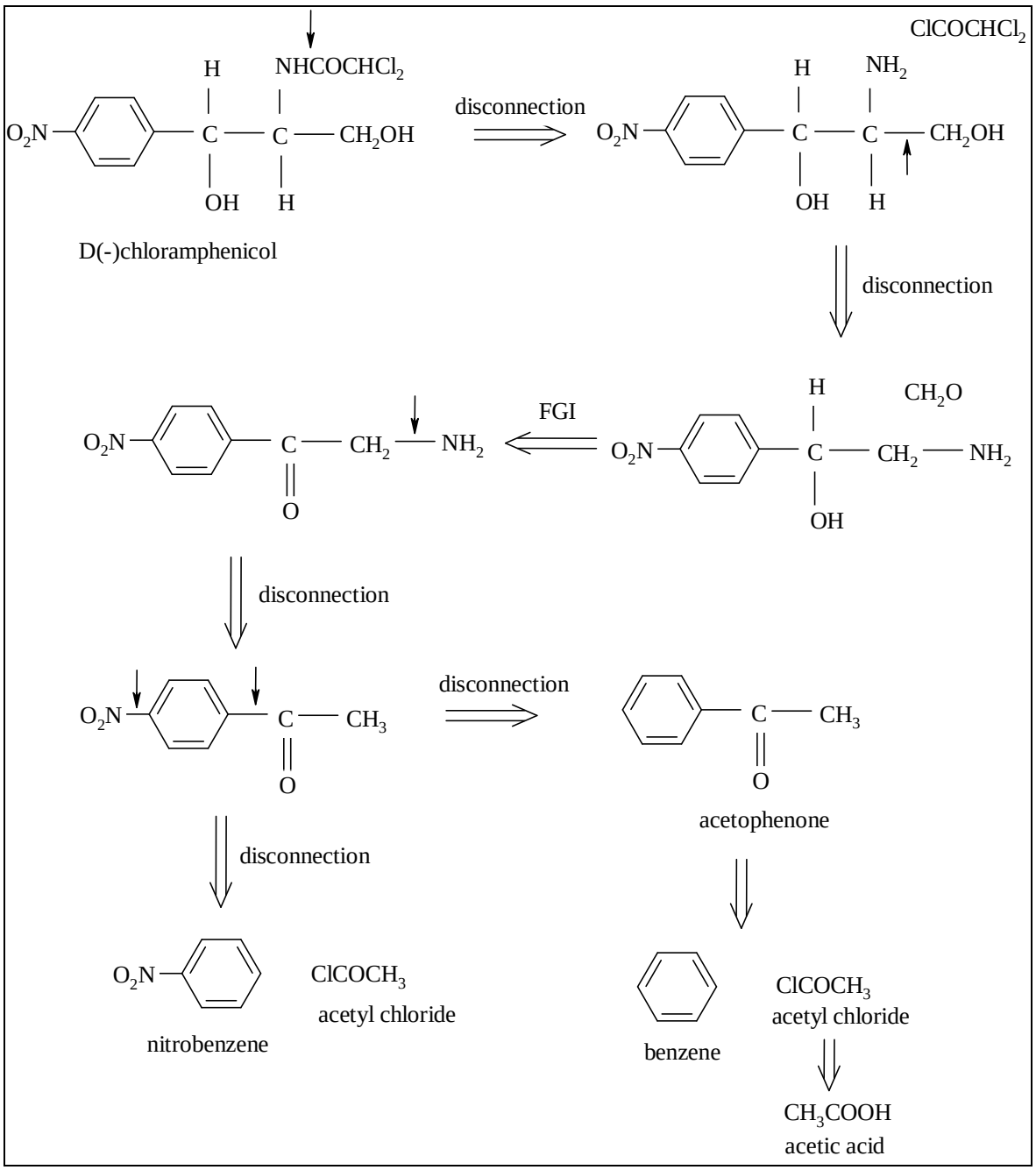
Retrosynthetic (or Antithetic) Analysis and the disconnection approach:
In this procedure, the **target structure** (TGT) is subjected to a **deconstruction** (= disconnection) process which corresponds to the *reverse of a synthetic reaction*, so as to convert the target structure to simpler *precursor structures* (**synthons**) without any assumptions regarding the starting materials. Each of the precursors so generated is then examined in the same way, and *the process is repeated until simple or commercially available structures result*. In other words, **"retrosynthetic analysis (or antithetic analysis) is a problem-solving technique for transforming the structure of a synthetic target (TGT) molecule to a sequence of progressively simpler structures along a pathway which ultimately leads to simple or commercially available starting materials for a chemical synthesis."**

The transformation of a molecule to a synthetic precursor is accomplished by the application of a **transform**, which *is the exact reverse of a synthetic reaction* to the target structure. It is not always necessary that the transformation be realizable in the laboratory, but *a synthetic reaction should be available in the reverse direction*. Each structure derived antithetically from a TGT then itself becomes a TGT for further analysis. Repetition of this process eventually produces a tree of intermediates having chemical structures as nodes and pathways from bottom to top corresponding to possible synthetic routes to the TGT. Such trees, (called EXTGT trees since they grow out from the TGT) can be quite complex since a high degree of branching is possible at each node and the vertical pathways can include many steps.



It is possible that a transform generates a precursor actually consisting of two or more fragments, as in the case of a convergent step; these fragments can each be treated in the same way as single precursors. For clarity, multiple fragments are not shown. Further analysis can generate deeper levels of precursors. Each precursor generated can then be checked for availability, thus defining an endpoint for that line of analysis. The final result, *a complete retrosynthetic tree, will contain all possible syntheses of the given target,* reasonable and unreasonable, efficient and cumbersome ones. Of course, such a tree would be unmanageably large both for man and computer, even when the number of precursor levels is limited. The **combinatorial explosion**, as this phenomenon is called, effectively prohibits the use of retrosynthetic analysis in such an unconstrained way. To keep the size of the retrosynthetic tree under control, a selection of transforms to be considered must be made. The guiding principles for this selection are called **strategies**. Another advantage of RA is that it may lead to the possibility of identifying *several different starting materials*, or *several different routes* to the synthesis of the same compound. Trying out all these alternate paths simultaneously may achieve a successful synthesis in a shorter time. The several precursor pieces from which the target may be assembled are called **synthons**. The synthons may be separately prepared and assembled together to get the target. This approach is called a **convergent synthesis** as compared to the usual **linear** or **consecutive synthesis**. This serves to reduce the time required to achieve the synthesis, and also improves the yield. For example, if each step gives 80% yield, The linear approach $A \rightarrow B \rightarrow C \rightarrow D$ gives D in an overall yield of 51%, while the convergent scheme $V \rightarrow W$, $X \rightarrow Y$, $W + Y \rightarrow Z$ gives Z in an overall yield of 64%.



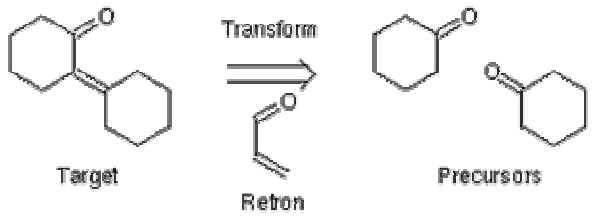


A sample retrosynthetic analysis of chloramphenicol

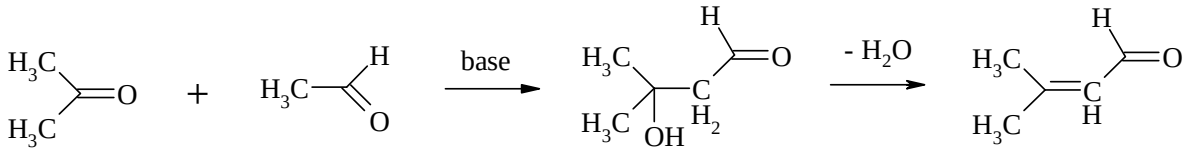
The terminology used with RA, as opposed to synthesis, is summarized below:

Terms for	Synthetic	Retrosynthetic
Starting structure	Starting material	Target (TGT)
Steps	Reactions	Transforms or disconnection
Steps indicated by	$\rightarrow$	$\Rightarrow$
Structural feature required	Functional group	Retron
Product after the step	Intermediate	Precursor or Synthon
Ending structure	Desired product	Probable starting materials

The **terminology** may be illustrated by the following examples:
(1)

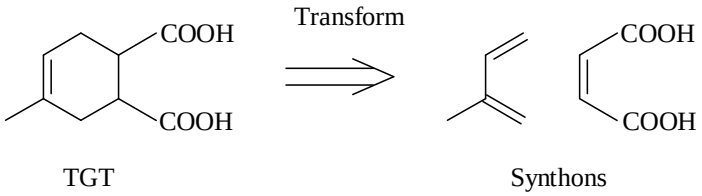


Explanation: The target is the molecule we want to synthesise. In RA, we are trying to find out possible starting materials (precursors) from which the target can be prepared using suitable reactions. On examining the target molecule carefully, we find that it contains a structural feature, the **C=C–C=O** group. Knowledge of synthetic reactions tells us that such a structural feature (or the 'retron') can be produced by an **aldol condensation** between two carbonyl compounds:



Note that the product of the aldol condensation between propanone and ethanal contains the C=C–C=O functionality. Thus whenever we come across such a structural feature, it provides us the possibility of synthesizing that part of the molecule through an aldol condensation from simpler molecules. Thinking back (retro = backward), it is an ideal position to *split the target molecule* (the 'transform' or 'disconnection') into *simpler units* (the 'precursors' or 'synthons') that may be joined together by an aldol reaction in the synthetic step. Therefore this structural feature is the '**aldol retron**'. In the above example, we thus come across the possibility of creating the target from two molecules of cyclohexanone. (But the target molecule cannot be actually split into two cyclohexanone molecules; it is only a mental exercise). Also note the arrow used to indicate the transform. The –C(OH)–CH_2–C=O group is also an aldol retron.

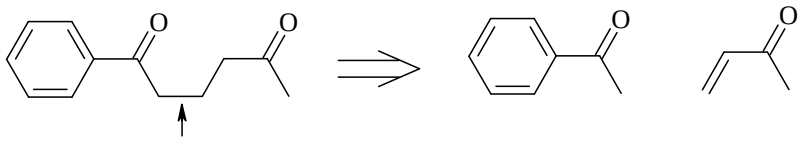
(2) The **Diels-Alder retron**



A cyclohexene ring is a Diels-Alder retron since it indicates a possibility of synthesizing it using a DA reaction between a suitable diene and a dienophile. The Diels-Alder reaction is one of the most useful synthetic reactions ever devised and is a Nobel prize winner. Since a *cyclohexane* ring can be obtained by reduction of a cyclohexene ring, it may also be produced through a DA reaction.

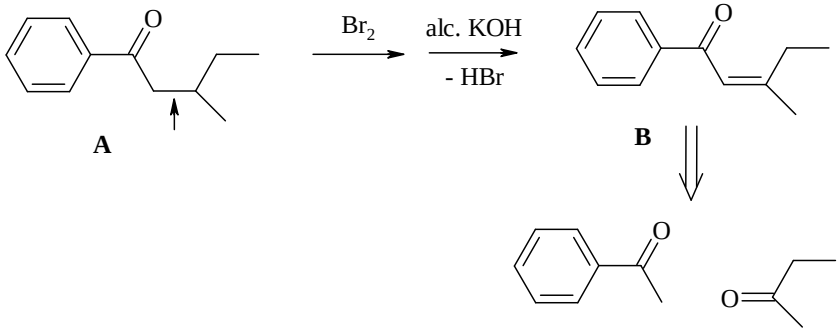
(3) The **Michael retron**

The addition of an α-hydrogen atom adjacent to a carbonyl group to an alkene, thus forming a new C-C bond is known as Michael addition. Consider the following molecule:



Can we synthesise it from simpler molecules? Keeping the Michael addition in mind, we see that we can 'disconnect' the molecule at the bond indicated by the arrow to get the precursors shown in the transform. The precursors can be 'connected' together and the desired C-C bond obtained through a Michael addition. Thus a single C-C bond α-β to a carbonyl group is identified as a Michael retron. Of course, there may be other methods also to obtain the same result.

Thus *remembering the (named) reactions and rearrangements* can help us in identifying retrons in the TGT, which will in turn help us locate positions where suitable disconnections can be made. Sometimes, when a suitable retron cannot be identified, even a suitable *modification of the molecule through a synthetic step* may be attempted to *generate* the retron. An example is given below:



Suppose we want to attempt a disconnection at the bond indicated by the arrow in molecule **A**. We see that a suitable retron for this disconnection is not immediately obvious. But if the molecule had been an enone with structure as in **B**, we could easily attempt the indicated transform since the conjugated enone is the aldol retron. **A** contains only the ketone part of the enone system, and therefore has only a **partial retron**. Hence a few extra synthetic steps which will easily convert **A** to **B** (with a **full retron**) are very helpful to identify suitable precursors. After joining the precursors, the extra double bond in **B** can be reduced to obtain **A**.

Assignment 1:
Referring to the reactions you have studied, identify the retrons for acyclic and cyclic alkenes, alcohols, ketones and lactones. Also suggest suitable reagents for achieving these synthetic steps.

Retrosynthetic analysis will only lead to useful results if it is directed towards some **goal**. The **basic goal** *is to generate precursors that correspond to available starting materials*. In very complex molecules, it may not be possible to immediately derive such simple starting materials. The basic goal, then, becomes the *generation of precursors that are easier to synthesize than the original target*. Stated differently, *retrosynthetic analysis is directed towards molecular simplification*. Corey has formulated five main types of strategies that lead to the desired simplification. These will be treated briefly, each illustrated by a sample analysis:

(1) **Functional-group based strategies**

Identification of functional groups such as carbonyl, hydroxyl, double bonds etc. can be useful in identifying suitable points for disconnection as discussed in the earlier examples. Functional groups in the target structure may direct the transform search in several ways:

- Removal of reactive and masked functionality
- Disconnection based on the location of functional groups
- Reconnection of functional groups to form rings retrosynthetically

The reconnective strategy is constrained by strategic rules. Clearly, it is not practical to attempt every possible reconnection.

(2) **Topological strategies**

The disconnection of specific, so-called **"strategic" bonds** can lead to major molecular simplification. These are identified based on the topology (= shape) of the target structure. There are several types of strategic bonds:

- Bonds in (poly)cyclic ring systems
- Bonds in (poly)fused ring systems
- Pairs of bonds in ring systems
  (disconnection by intramolecular cycloaddition transforms)
- Bonds connecting chains to rings
- Bonds connecting chains to other chains
- Bonds connecting chains to functional groups

Heuristics (empirical rules) have been devised to select these types of bonds from any target structure. For example, it may not be a good idea to open up an aromatic ring system. Positions where branching occurs in a chain are usually ideal locations for a disconnection. It is also possible to identify rings which should be disassembled early in the retrosynthetic process, or rings which should be kept intact during these stages.

(3) **Transform-based strategies**

A very useful guidance for retrosynthetic analysis can be provided by the application of a powerfully simplifying transform -- corresponding to a reaction effecting a considerable increase in complexity. Very often such an application is suggested by the presence of (functionalized) rings of specific sizes in the target molecule. Some powerfully simplifying transforms are:

- Diels-Alder
- Hetero Diels-Alder
- Robinson annulation
- Birch reduction
- Internal ene reaction
- Halolactonization

Examples were discussed earlier.

(4) **Structure-goal strategies**

The analysis can also be directed towards a particular (sub)structure. Such a (sub)structure can be a:

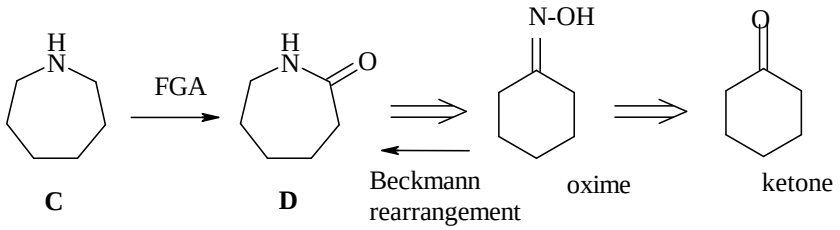
- Starting material
- Chiral building block
- Retron-containing structure

An analysis directed towards such a structure-goal does not need to be purely retrosynthetic. It can even be synthetic, but probably the most efficient search would be a bidirectional one. An example is given on page 8 in the conversion of **A** to **B** aimed at producing a retron-containaing structure.

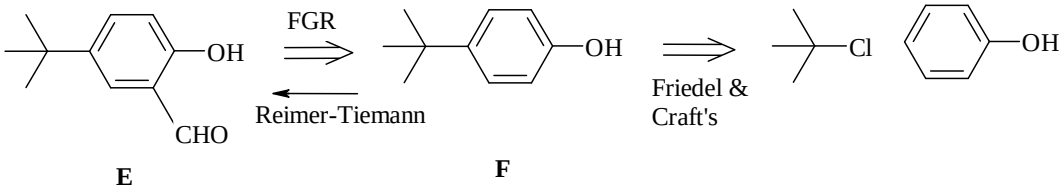
(5) **Stereochemical strategies**

Here the focus is on removal of stereocenters under stereocontrol. Stereocontrol can be achieved through either mechanistic control or substrate control. Reconnections that move stereocenters from chains (where they are difficult to introduce) into rings (where introduction is usually much easier) can also be considered stereochemically strategic. Suitable stereoselective reactions must be available to reverse-engineer the stereochemistry at that centre.

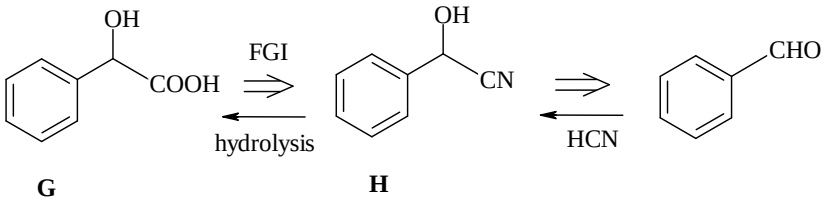
Clearly, the sequence from **A** to **B** on page 8 is non-simplifying, but these steps are needed to pave the way for the goal transform. Thus producing the structural feature present in **B** is a **subgoal**. The term "subgoal" is used because such a step is subordinate to the application of a goal transform. Without a goal transform 'in mind', there is really no good reason to apply a subgoal transform. Subgoal transforms which manipulate functional groups are very common: **functional group addition** (FGA), **functional group removal** (FGR), **functional group interchange** (FGI), and **functional group transposition** (FGT) are frequently employed. But in fact *any transform which assists in 'setting up' the retron for a goal transform can be thought of as a subgoal transform*. Examples of each type are given below:



The above is an example of functional group addition. We wish to synthesise **C** (which is the goal). If a keto group is introduced into the ring as in **D**, the ring system can be generated by a Beckmann rearrangement of the oxime which in turn can be obtained from the ketone. Therefore production of **D** is a subgoal, and it is generated from **C** through a functional group addition (**FGA**). From **D**, **C** can be obtained by reduction. Another example for FGA is the transform at the top of page 7.



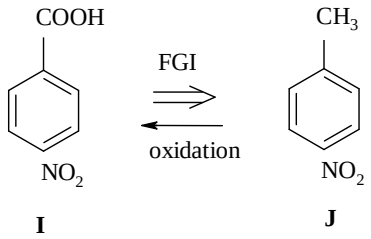
Given above is an example of a functional group removal (**FGR**) subgoal transform. How to create molecule **E**? We know that **F** can be easily obtained by Friedel and Craft's reaction between 2-chloro-2-methyl propane and phenol in presence of anhydrous $AlCl_3$. Therefore removal of the aldehyde group will help in obtaining a transform to create the carbon skeleton of **E**. This functional group may be introduced later through a Reimer-Tiemann reaction.



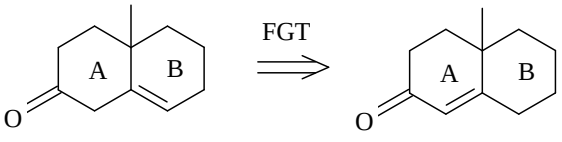
Given above is an example of a functional group interchange (**FGI**) transform. How to create molecule **G**? If instead of the acid group we had the cyano group as in **H**, it will be a cyanhydrin which can be easily obtained by the action of HCN on benzaldehyde. The cyano group on hydrolysis will then give the required acid functionality. Therefore it is advisable to change the functional group in **G** to that in **H** although simpler fragments are not produced.

Assignment 2:
Explain the following terms: (1) Antithetic analysis (2) Retron (3) Partial Retron (4) Synthon (5) Transform (6) Target molecule (7) Disconnection (8) Convergent synthesis (9) Linear synthesis (10) Goal (11) Subgoal (12) Combinatorial explosion.

Another example for FGI is given below:



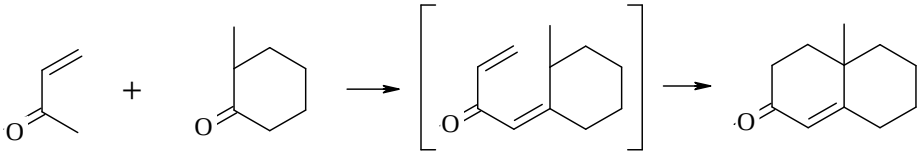
Nitration of benzoic acid is not easy and gives very poor yields. But if the acid group is exchanged for a methyl group, it can be easily nitrated in the 4-position and then the methyl group oxidized to get the acid function. The following is an example for a functional group transposition transform:



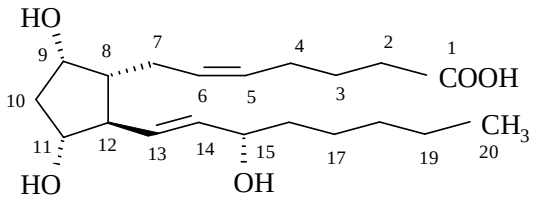
K **L**

It is the non-conjugated molecule **K** which we want to synthesise. But the *transposition* (changing the position) of the double bond to get the conjugated enone **L** (remember the aldol retron?) immediately gives us the possibility of creating the ring A from ring B using an aldol condensation. The double bond can later be isomerised to the required position in **K**.

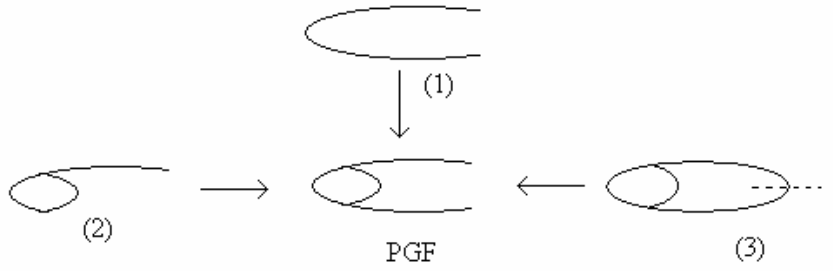
There is an added bonus if the subgoal transform is itself also simplifying; the combined sequence will probably have a high merit. This is clearly the case with *tactical combinations*. A **tactical combination** is the retrosynthetic equivalent of a standard reaction sequence. The steps in a tactical combination follow naturally one after the other because *each step sets up the retron for the next step*. Tactical combinations are most powerful when they have a high content of simplifying steps. The best-known tactical combination is probably the **Robinson annulation**: an aldol transform followed by a Michael transform. In fact, it is so common that it is often treated as a single-step process:



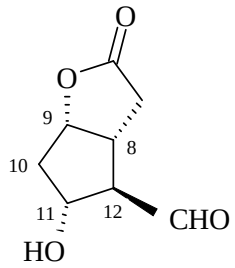
Retrosynthetic analysis of Prostaglandin $F_{2\alpha}$



Examination of the structure of PGF2α reveals a cyclopentane ring with two tails attached in adjacent positions on the ring. The tails are *trans* to each other. All the three hydroxyl groups are *α* to the ring system, and lie on the same side of the ring as the tail with the COOH group. The double bond on this tail is *cis* while that on the other tail is *trans*. Three main approaches have been taken: (1) formation of the cyclopentane ring from acyclic precursors, (2) starting with a preformed ring, and (3) starting from bicyclic precursors.



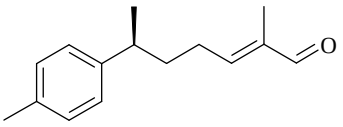
The conjugated enol in positions 13-15 is an aldol retron and may be expected to be produced from an aldol condensation. The side chain at position 8 and the OH at position 9 are on the same side of the ring; this stereochemistry may be achieved if both are produced simultaneously by the cleavage of a ring structure attached to the cyclopentane ring at the 8,9 positions. Similar will be the case with the stereochemistry at the 8,11 positions or the 9,11 positions. Thus Prof. E. J. Corey used the following synthon for his synthesis. The synthon was synthesized from cyclopentadiene through suitable reactions.



The Corey lactone aldehyde

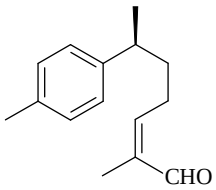
Assignment 3:
Give the expansions of and explain the following terms using suitable examples:
(1) FGA (2) FGR (3) FGI (4) FGT (5) TGT.

Retrosynthetic analysis of Nuciferal:



(E)-(S)-(+)-Nuciferal

or



Nuciferal is a sesquiterpene component of the essential oil obtained from the wood of *Torreya nucifera*. Its structure and stereochemistry are similar to that of curcumene and turmerone found in the essential oils of *Curcuma longa* (turmeric, a ª Ä) and *Zingiber* (C © ⇓). The structure of nuciferal is given above.

Assignment 4:
Examine the structure of nuciferal and write down the retrosynthetic analysis scheme for it.