

## Basics of Retrosynthetic Analysis

Jiarui Yang

The Affiliated High School of Peking University, Beijing 100080, China.

yangjiarui@i.pkuschool.edu.cn

---

### Abstract

Retrosynthetic analysis is a problem solving method for designing a synthetic pathway. Retrosynthetic analysis is a series of transformations that eventually simplifies the synthetic target molecule into numerous simple and commercially available compounds called starting materials. Retrosynthesis, the reverse process of a chemical synthesis, is accomplished by simplification and transformation depending on the existing chemical reactions. After each step of retrosynthetic analysis, the simplified precursor structure is alternatively became the target molecule. In this paper, the author will outlines several retrosynthetic analysis strategies and summarizes basics of retrosynthesis with simple examples to illustrate the strategies' possible application. In general, the retrosynthetic analysis can be simplified into four groups: common atom approach, one group disconnection, two groups disconnection, and illogical disconnection. A good retrosynthetic analysis should have good mechanism, achieve the greatest possibility of simplification, employ maximum convergence, and result in recognizable starting materials. Retrosynthetic analysis can lead to many synthetic routes and explores the optimal route for industrial manufacturing.

### Keywords

Retrosynthetic analysis, Starting material, Disconnection, Simplification, Functional groups.

---

### 1. Introduction

Retrosynthesis analysis is a technique used in the process of designing synthetic reactions. The concept of retrosynthesis analysis is first proposed by E.J. Corey in his book *The Logic of Chemical Synthesis* in 1964<sup>[1]</sup>. Different from common synthesis, it doesn't follow the usual pathway of choosing starting material and converting those identified material into the target molecule. Retrosynthesis analysis is the reverse process of common synthesis. The analysis starts from the target molecule (TGT) and transforms them into simpler precursor structure and further into simple or commercially available starting materials based on both preexisting principles and structural analysis. Each structure derived antithetically from a target the itself becomes a target molecule for a further analysis<sup>[2]</sup>. The ultimate goal is to simplify a complex organic molecule into several simpler molecules. If there exist more than one reasonable synthetic routes, the final optimal synthetic route should be determined through comparison in terms of environmental acceptability, safety, efficiency, economical acceptability, reproducibility, etc<sup>[3]</sup>. Retrosynthesis can be employed in drug synthesis and natural compound synthesis. Some natural compounds can be extremely costly in industrial manufacture if extract from plant materials or separate from their isomers. Manual synthesis would instead create bountiful compounds with high purity. In medicine, retrosynthesis help scientists to achieve their desirable target ligands that combine more compactly with receptors in human body. Despite of its easy operability and wide applicability, few of the researchers summarize its basic approaches and rules employed in retrosynthetic analysis. In this article, the author concludes the

most commonly used analytic methods in retrosynthetic analysis with detailed examples of how it can be employed.

## 2. Types of Disconnection

### 2.1 Retrosynthesis of polycyclic compounds

#### 2.1.1 Common Atom approach

In order to deconstruct a polycyclic target molecule, the approach called "disconnection of common atom" is utilized. Common atoms are atoms that serve the role of connection in the polycyclic molecule and exist in more than one ring. As the common atoms play a pivotal role in integrating the molecule, the most ideal strategy to simplify the molecule is to disconnect the bonds between two common atoms. The common atom approach involves two steps: first, identify the common atoms; second, disconnect using common atom as a guide.

#### 2.1.2 Example 1

The common atoms in this atom is identified and marked using bold dots. After breaking different bonds between the common atom, the ring system is simplified into a series of two connecting carbon rings (presented in Figure 1). One thing worth noticing about is that not all the disconnections between common atoms lead to simpler intermediates. The fourth route results into a 10 membered carbon ring with bridge structure, which is undesirable from a topological standpoint due to its large ring <sup>[4]</sup>.

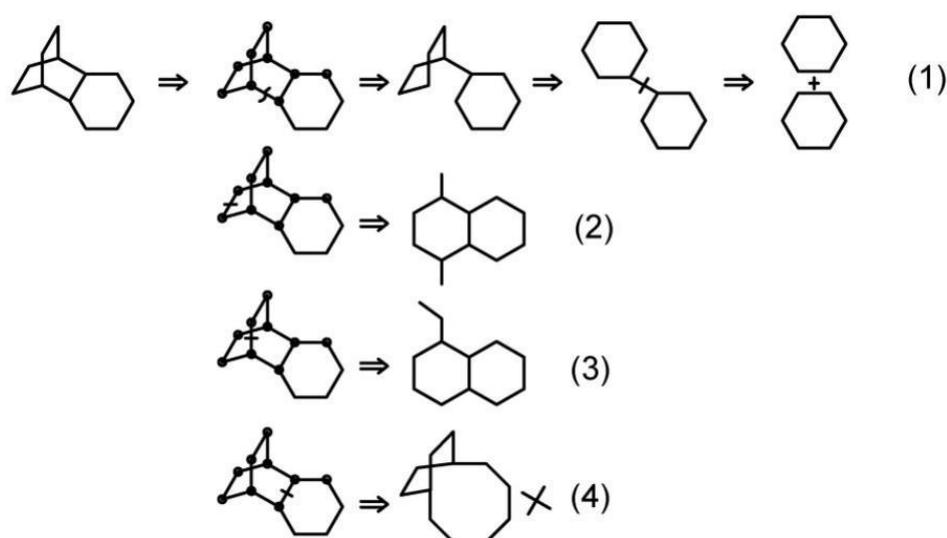


Figure 1. Example 1 of common carbon approach. Four possible ways of disconnecting a polycyclic compound

#### 2.1.3 Example 2

Another example, the disconnection of Longifolene, is given in Figure 2. Chemically, longifolene is a tricyclic sesquiterpene. To better simplify the complex ring system of Longifolene, the common atoms should be defined as ring-member atoms that are bonded to more than two rings <sup>[5]</sup>. The atoms which are common to every ring is more strategic to disconnect than atoms which are only common to some of the rings. The "common atoms" are numbered 1-5 in the picture down below <sup>[1]</sup>. A ring system is considered as less complex, when the number of the rings in the system is reduced and there is no large carbon ring. By disconnecting the bonds connecting two common atoms (1-2, 2-3, 3-5, 1-4, 4-5) respectively, a simplified ring system is therefore derived <sup>[5]</sup>. Among these structures, the desirable intermediate structure in synthesis may be hidden. Further evaluation of synthetic route and synthetic reactions will be taken into consideration.

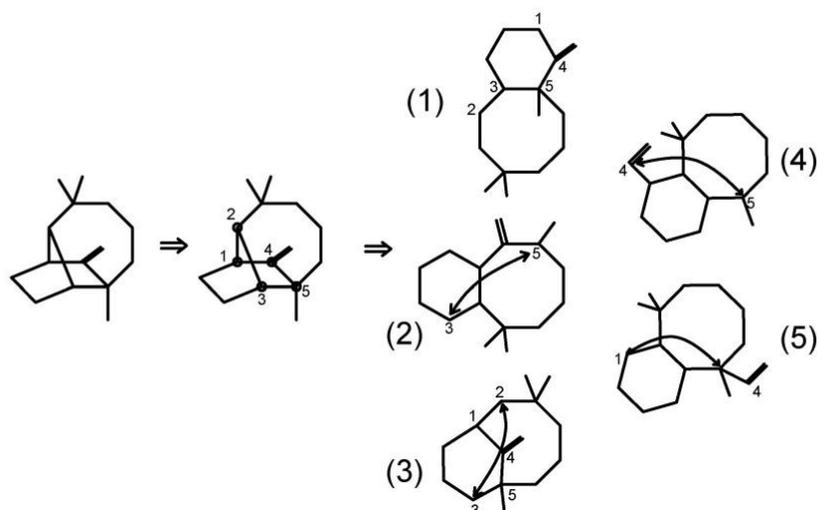


Figure 2. Example 2 of common carbon approach. Different ways of disconnecting Longifolene.

## 2.2 One group disconnection

### 2.2.1 Disconnection of simple Alcohols

The alcohol group will provide lone pair of electrons to stabilize the cation. Therefore, to disconnect simple alcohols, we simply need to determine disconnecting which bond will form the most stable anion (see Figure 3).

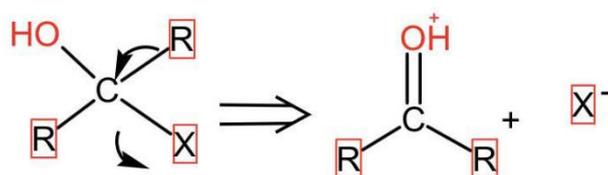


Figure 3. The mechanism of disconnecting simple alcohols.

### 2.2.2 Example of Alcohol disconnection

As presented in Figure 4, cyanide group is consisted of a carbon atom triple bonded to a nitrogen atom. In inorganic cyanides, the group  $C\equiv N$  is often presented in the form of an anion. After disconnection, cyanide will be an optimal ion. The carbon cation is stabilized the lone pair electrons on oxygen.

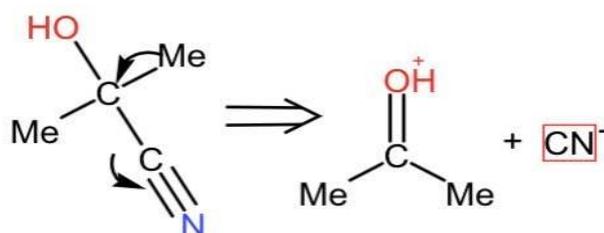


Figure 4. Example of disconnecting simple alcohols. Disconnection of 2-hydroxy-2-methylpropanenitrile

### 2.2.3 Disconnection of simple Olefins

Olefins can be made by the dehydration of alcohols [6]. Thus, in the process of designing the retrosynthesis of olefins, adding water across the double bond could be a promising starting point. The water addition across the double bond often result in two possible alcohols. Only one of them is a useful disconnection [4].

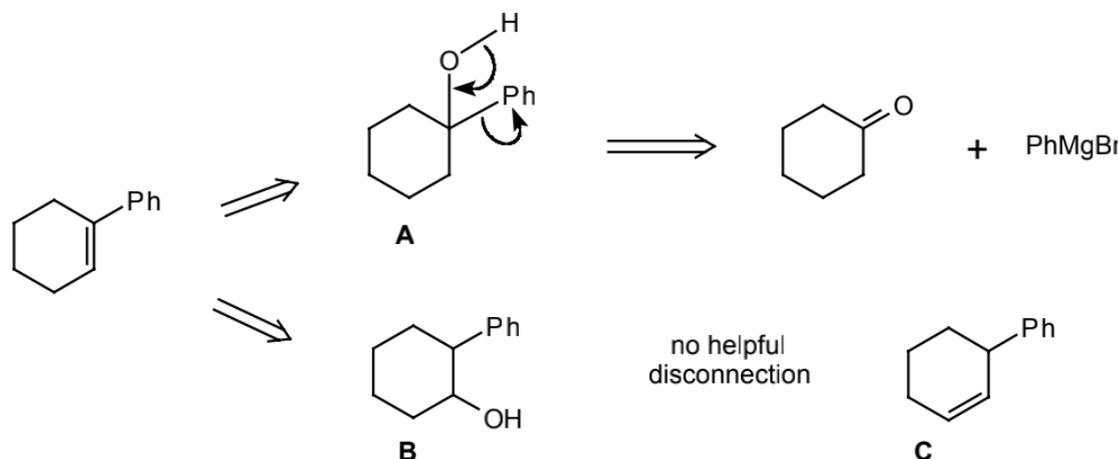


Figure 5. Example 1 of Olefin disconnection. Disconnection of (cyclohex-2-en-1-yl) benzene.[8]

### 2.2.4 Example 1 of Olefin disconnection

The addition of alcohol at the same side of Ph group helps break the bond between Ph and the carbon ring, whereas the addition on the opposite side leads to a meaningless disconnection (see Figure 5) [4].

### 2.2.5 Example 2 of Olefin disconnection

Other than through alcohol additions, another way to disconnect olefins is through direct breakage of the double bond [4]. This immediate disconnection is a relatively shorter route (shown in Figure 6).

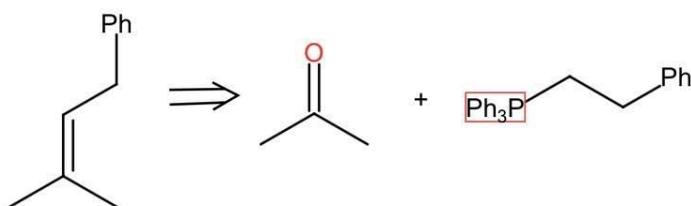


Figure 6. Example 2 of Olefin disconnection. Disconnection of (3-methylbut-2-en-yl) benzene. [4]

## 2.3 Disconnection of Simple Ketones and Acids

### 2.3.1 Disconnection of Simple Ketones

Figure 7 shows that ketone is made by first returning to alcohols [1].

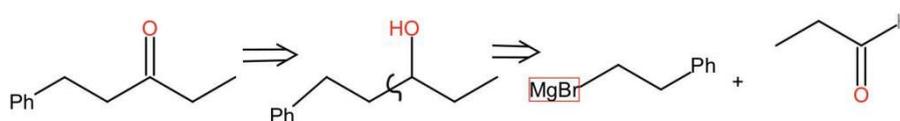


Figure 7. Example of disconnection of simple Ketones. Disconnection of 1-phenylpentan-3-one.

### 2.3.2 Disconnection of Simple Acids

As acid itself is a hydroxyl compound, it doesn't have to return to the parent alcohols. It can be simply disconnect in the same way as alcohols do <sup>[6]</sup>.



Figure 8. Example of disconnection of simple acids.

### 2.4 Illogical Disconnection--Fine Tuning:

Usually, the logical way in consider of simplifying a molecule is to disconnect the bond. However, a strategy of retrosynthesis is called "illogical disconnection". Rather than breaking bonds, illogical disconnection will simplify the atom through connecting the two parts of the atom. The example of linear diadehyde in Figure 9 is listed below <sup>[4,7]</sup>. Other than separating the molecule in the middle, the illogical disconnection joins the two carbonyl groups to a double bond. In this way, the dialdehyde is transformed back to a cyclohexene. This conversion is actually easy to achieve in the forward direction in real life. One product of the reaction cyclohexane oxidation is linear dialdehyde, as the addition of oxygen will cause the double bond to rupture.

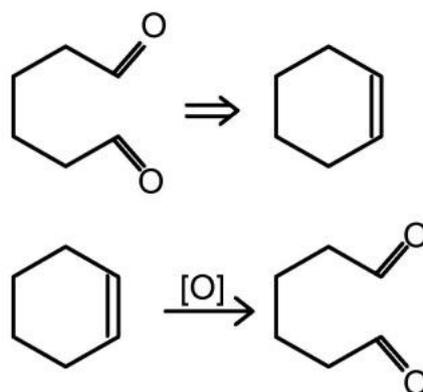


Figure 9. Example of illogical disconnection. Conversion between hexanedial and cyclohexene.

#### 2.4.1 FGI -- Functional Group Interconversion

Functional group interconversion will convert its functional group to another under the premise of not changing the main molecular skeleton of the target molecule. It doesn't simplify the problem directly, but shift the problem and allows for more possibility. Through functional group interconversion, the target molecule can be converted into molecule that is easier to prepare.

As shown in Figure 10, the acetal functional group is originated from alcohols and carbonyl compound. Through functional group interconversion, we can change it back to alcohols and carbonyl.

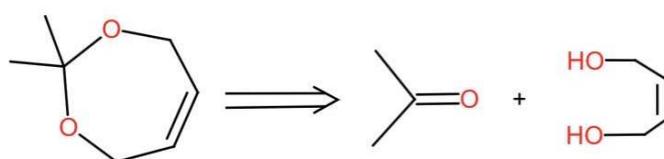


Figure 10. Example of Functional group interconversion. IUPAC name: 2,2-dimethyl-4,7-dihydro-2H-1,3-dioxepine.

### 2.4.2 FGA -- Functional Group Addition

The addition of functional group allows the molecule to undergo disconnections which are previously impractical. Because functional groups undergo same type of reactions regardless of the compound of which they are a part, the addition of functional groups will allow the compound to have new and desirable properties.

### 2.4.3 FGR -- Functional Group Removal

The two functional group disconnection of bromoketone is hard to imagine. Functional group removal is an ideal way to simplify the molecule and facilitate further synthetic analysis. The removal of the bromine group allows people to only consider the ketone (see Figure 11).

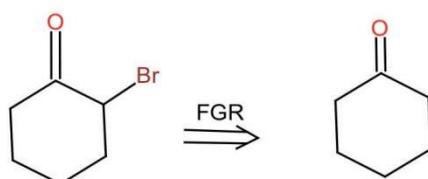


Figure 11. Example of Functional Group Removal. IUPAC name: 2-bromocyclohexan-1-one.

## 2.5 Retrosynthesis of Vitamin E

Vitamin E consists of eight fat-soluble compounds, including four tocopherols and four tocotrienols. The structure of alpha-tocopherol is composed of two parts, one chromane double ring, and a hydrophobic side chain. The hydroxyl group on the chromane double ring release hydrogen atom to reduce active oxygen species, and the side chain allows the vitamin E compound to penetrate into cell membranes<sup>[8]</sup>. Among the eight forms of vitamin E, alpha-tocopherol is the biologically active form. In this article, alpha-tocopherol is taken as an example to illustrate the retrosynthesis process of vitamin E. Most of the alpha-tocopherol manufactured industrially is synthesized from 2,3,5-trimethyl-hydroquinone and isophytol with the catalysis of hydrogen chloride gas and iron. In retrosynthesis, we can just disconnect the side chain from the carbon ring, as shown in *Figure 12*<sup>[9]</sup>. Through functional group inter conversion, the vitamin E compound is separated into 2,3,5-trimethyl-hydroquinone and isophytol<sup>[9]</sup>. The two compounds can then be simplified into starting materials.

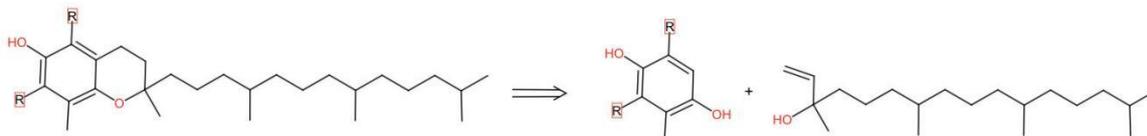


Figure 12. Retrosynthesis of Vitamin C.

## 2.6 Basic Principles of Retrosynthesis

1. Break chemical bonds based on existing chemical reactions. When breaking chemical bonds that are formed between different atoms, electronegativity should be considered<sup>[3]</sup>.
2. Follow the maximum possible simplification principle. The bond break should happen in the center of the molecule, at the branch chain, or used the symmetry of the molecule<sup>[2]</sup>.

Example: One kind of disconnection is to focus on the carboxylic ester group and disconnect it to a carboxylic acid group and a methyl group (*see Figure 13*). Although this disconnection follows existing chemical reactions, it doesn't simplify the problem for the carbon ring, olefin, and carbonyl are still existing. In contrast, the second approach greatly simplify this molecule, as it breaks the

molecule into smaller pieces. Typically, smaller compounds are more available than larger compounds.

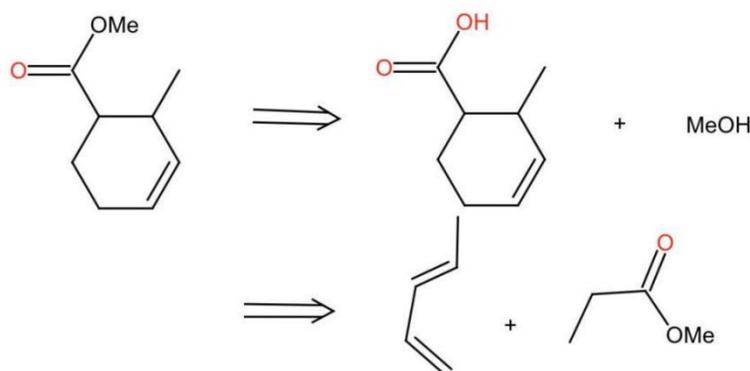


Figure 13. Example of disconnecting compounds into smaller starting material. Methyl 2-methylcyclohex-3-ene-carboxylate

3. Break the carbon—heteroatom bond. When the organic molecule contains C-X bonds, people typically choose to break the C-X bond, especially when the heteroatom is oxygen, nitrogen, and sulfur. The bond to connect heteroatom is usually easy to form; therefore, after being broken, the bond will form again <sup>[3]</sup>.

4. Employ maximum convergence. Choose the cutting scheme with fewer synthesis steps, high reaction yield, and easy availability of raw materials, if there exist several possible cutting scheme <sup>[1]</sup>.

5. In the case of functional groups, the bond connecting the functional group with the rest of the molecule should be severed.

### 3. Conclusion

The basic analysis methods for retrosynthesis can be generally concluded into four kinds: common atom approach for polycyclic compounds; one group disconnection; two groups disconnection; and illogical disconnection (fine tuning). A good retrosynthetic analysis should have good mechanism, achieve the greatest possibility of simplification, employ maximum convergence, and result in recognizable starting materials. In this article, the author briefly summarizes and introduces those methods and provides simple examples to help readers to understand. It is vital to notice that this is only a very general introduction and that retrosynthetic analysis is far more complicated. Even though some retrosynthetic routes might not be practical, retrosynthetic analysis is one important problem-solving strategy which guide the search of a possible synthetic route on behalf imagination. Not only can retrosynthetic analysis help us find more suitable raw materials and choose better synthetic routes, but also, with the help of computer technology, the method can be used to guide the Green Synthesis, Economic Synthesis, and rapid R & D of target compounds. Retrosynthetic analysis is expected to play an important role in drug development for the prevention and treatment of new coronaviruses, especially in the current epidemic.

### Acknowledgments

Special Thanks to Professor Brian M. Stoltz for offering valuable instructions and providing suggestions in the process of writing. I also acknowledge the support and encouragement of Teaching Assistant Yuqing Zheng.

## References

- [1] Corey, E. J. "Robert Robinson Lecture. Retrosynthetic Thinking—Essentials and Examples." *Chem. Soc. Rev.*, vol. 17, no. 0, 1988, pp. 111–133, 10.1039/cs9881700111. Accessed 12 Aug. 2020.
- [2] Corey, E J, and Xue-Min Cheng. *The Logic of Chemical Synthesis*. New Delhi, Wiley, 1995.
- [3] Law, James, et al. "Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation." *Journal of Chemical Information and Modeling*, vol. 49, no. 3, 6 Feb. 2009, pp. 593–602, 10.1021/ci800228y. Accessed 12 Aug. 2020.
- [4] Warren, Stuart G. *Designing Organic Syntheses : A Programmed Introduction to the Synthron Approach*. Chichester ; New York, Wiley, 1978.
- [5] Corey, E. J. "Total Synthesis of Longifolene." *Journal of the American Chemical Society*, vol. 86, no. 3, Feb. 1964, pp. 478–485, 10.1021/ja01057a039. Accessed 12 Aug. 2020.
- [6] "Functional Group Names, Properties, and Reactions | Boundless Chemistry." *Courses. Lumenlearning. Com*, courses.lumenlearning.com/boundless-chemistry/chapter/functional-group-names-properties-and-reactions. Accessed 12 Aug. 2020.
- [7] Simpson, Jeffrey H. *Organic Structure Determination Using 2-D NMR Spectroscopy : A Problem-Based Approach*. London, Elsevier/Ap, 2012.
- [8] FUJISHIMA, T., et al. "ChemInform Abstract: Synthesis of Vitamin E Analogues: Possible Active Forms of Vitamin E." *ChemInform*, vol. 27, no. 19, 5 Aug. 2010, p. no-no, 10.1002/chin.199619259. Accessed 17 May 2019.
- [9] Brigelius-Flohé R, Traber MG (July 1999). "Vitamin E: function and metabolism". *FASEB Journal*. 13 (10): 1145–55. doi:10.1096/fasebj.13.10.1145. PMID 10385606.