Introduction of Retrosynthetic Analysis and Guidelines Application of Analyze Molecules

Haoyang Liu, Jianing Zhu

Abstract

Retrosynthesis is vital in breaking down complex target organic compounds based on known reactions. It serves as the reverse process of synthesis, which provides an alternative thinking path to aim for finding possibilities of synthesis reactions. In addition, different synthesis pathways can be discovered in the process of utilizing the retrosynthesis technology. It introduces the history of retrosynthetic synthesis using deoxygenation patterns which are different from disconnections, guidelines, and ordinary. In this work, a complete introduction to retrosynthesis and guidelines for people to apply retrosynthesis are well supported and introduced. The purpose of this paper is to serve as a detailed tutorial and offers specific rules for readers to learn.

1. Introduction

Retrosynthetic analysis is regarded as a very significant component to understand today's synthetic organic chemistry. It is an intellectual tool that organic chemists use to think about how they might be able to construct the molecule. Retrosynthetic analysis is a very recent concept. At first, people didn't have specific plans when proceeding retrosynthetic analysis. People started to process synthesis by planning only in the last fifty or sixty years. The way they perform the process that is fairly through reverse engineering process or deconstructed analysis, which means to break bond in a complex molecule in order to simplify it to the greatest extend in order for us to start analysis the complex molecule in the lab [1]. The complex molecule in retrosynthetic analysis is defined by the term of targeted molecule, a desired molecule that we want to make in organic chemistry. Thinking about which bond should be broken and trying to make the starting material as simple as possible is very important intuitions that a student who studies organic chemistry should have.

2. Background Information

2.1 Understanding retrosynthetic analysis by geometrical graph

Retrosynthetic analysis can be started to understand by talking about how to retrosynthesis a geometrical graph [2]. Because retrosynthetic analysis is fairly turning a complex compound into simplest building blocks. The first step is to think about how to deconstruct a geometrical graph beyond chemistry, for example a square. Therefore, starting on the question of how many ways you can come up with to deconstruct the square.



Figure 1. deconstruct a square

The square can be thought as a complex molecule or targeted molecule in this case. According to Figure 1, it is simplified as four dots, sticks, small triangles, small squares, and right-angles. In chemistry, this process is called on paper plan, which can be understood as a backward reaction. A retrosynthetic arrow is needed in this case. The general reaction of simplifying process is shown by figure 2.

Е ____ D+C+B+А

Figure 2. backward reaction/ retrosynthesis general reaction

As having a targeted molecule E, square, it can be simplified as A, B, C, and D, dots, and sticks shown Figure 1. After planning specifically and getting the simplified result, those starting materials A, B, C, and D can be brought to the lab and be turned to the targeted molecule E by various reactions. This process can be considered as forward reaction as follow in figure 3 (note: different arrows are used to represent forward synthesis and retrosynthesis):

A + B + C + D — Figure 3. forward reaction

A more specific chemical example of cyclopentene and

Bromo cyclopentane are shown in figure 4 [2]:

Forward Synthesis

$$\swarrow$$
 + H-Br \longrightarrow $\overset{Br}{\longleftrightarrow}$ $\overset{Br}{\longleftrightarrow}$ $\overset{Br}{\longleftrightarrow}$ + H-Br

Figure 4. forward ans backward synthesis of cyclopentene

However, in real chemistry, retrosynthetic analysis is usually not as easy as the one step reaction that shown above. It is more likely to include multiple steps. Schematic examples of both retrosynthesis and forward synthesis are shown below in figure 5 and figure 6: Retrosynthesis:

 $E \implies D \implies C \implies B + A$

Figure 5. retrosynthesis analysis for multiple steps reaction

Forward synthesis:

 $A + B \longrightarrow C \longrightarrow D \longrightarrow E$

Figure 6. forward synthesis for multiple steps reaction

2.2 "Common atom" approach

Common atoms are atoms that are common to more than one ring in a molecule. For instance, in the figure 7, the common atoms of 1-hydroxytaxinine molecule are marked as a bold dot. The common atom approach is a technique that helps retrosynthetic analysis by disconnecting bonds that link two or more of those common atoms, which is the most strategic disconnection because it leads to significant simplification. Otherwise, it will be considered as non-strategic disconnections because it does not lead to significant simplification [2].



Figure 7. Disconnection for 1-hydroxytaxinine molecule

3. Disconnection

3.1 Arrow pushing

To comprehend retrosynthesis, familiarizing with basic phenomena and key guidelines are essential. Firstly, instead of only considering the framework, which is the purpose of common atoms, we now have to take the movements of electrons into account during retrosynthesis. One way to demonstrate electron activity is using a technique called Arrow Pushing. Arrow Pushing accurately describes the progression of organic reactions through representing the flow of electrons with arrows superimposing on organic structures [3]. Through this tool, we can picture and imagine the details of a reaction. Retrosynthesis, with its philosophy of dissecting a relatively complicated molecule into a simpler precursor, requires the breaking of bonds disconnections. Essentially, a bond consists of 2 shared electrons, and a disconnection results in consequences of the 2 electrons [1]. For each disconnection, there will be three possibilities of electron movements, each possibility directly affecting the synthons (the outcome of a disconnection). The first circumstance is the left synthon bearing a positive formal charge and the right synthon bearing a negative formal charge (see Figure 8):

R ⊂ R' ⇒ R[⊕] ⊕ R' Figure 8. Disconnection of bond where the positive formal charge falls on the left synthon

The second circumstance is the right synthon bearing a negative formal charge and the left synthon bearing a positive formal charge (see Figure 9):



Figure 9. Disconnection of bond where the positive formal charge falls on the right synthon

The third circumstance is the free radical case (see Figure 10). Free radicals are very reactive, as they desperately want 1 other electron to fill up their orbital:



Figure 10. Disconnection where radicals are formed

Notice how the arrow pushing signifies the electrons. A full, double-headed arrow indicates the transfer of two electrons; a partial, single-headed arrow indicates the transfer of one electron. Understanding the underlying principles of bonds gives us more options for viable disconnections, in turn supplying more choices for the most optimized disconnection.

3.2 Synthons and synthetic equivalent

Next, consider the definition and practicality of synthons and synthetic equivalents. Synthons, retrosynthetically, are constituent units of the target molecule that have the potential of becoming a starting material. Synthons are oftentimes unstable, because they are charged organic species, as described above. Because of that, a synthetic equivalent is needed to substitute the synthon in an actual experiment. Synthetic equivalents are stable molecules that react like the synthons in a realistic condition. For example, MgCl added onto an ethyl carbanion would be the synthetic equivalent of the ethyl carbanion synthon (see Figure 11). AGrignard reagent is polarized so that there is a delta negative on the carbon it's attached to and a delta positive on itself. Grignard reagents inherently act like a carbanion, and that is the reason they match for a synthetic equivalent of the ethyl carbanion:

$$H_3C \xrightarrow{\oplus} \odot CH_3 \approx CIMg$$

Figure 11. Disconnection of butane along with its synthetic equivalent

3.3 Crucial guidelines and tools

Thirdly, it is necessary to be comfortable with some crucial guidelines and tools related to disconnections. There are numerous types of disconnections, but the most important ones are 1' group disconnections, 2' group disconnections, electrocyclic disconnections, and illogical disconnections. In all of these disconnections, there are choices to be made in which bond should be broken and choosing the most effective one matters. For example, 2-butanol (see Figure 12):



Figure 12. Multiple disconnection methods of 2-butanol

Notice how there are 3 different types of disconnections, and all yields individual synthons.

For 1' functional group disconnections, there is only one functional group, so use that to help guide disconnections. For 2' functional group disconnections, the goal should be to separate the two functional groups. For electrocyclic disconnections, the retrosynthesis should be guided by electron flows inside the ring (for example, the Diels-Alder reaction). For Illogical disconnections, it creates a bond instead of disrupting one.

One tool that can aid discover the right disconnection is called Fine Tuning. Fine-tuning does minor changes to the original compound, which would help us find feasible disconnections [1]. There are three types of Fine Tuning: Functional Group Interconversion (FGI), Functional Group Addition (FGA), and Functional Group Removal. FGI trades a functional group for another, as exemplified in Figure 13:



Figure 13. Example of FGI

FGA adds a functional group to the original compound, as exemplified by Figure 14:

Figure 14. Example of FGA

FGR removes a functional group from the original compound as exemplified by Figure 15:



Figure 15. Example of FGR

A molecule that has two or more carbon connecting to the oxygen-containing functional group has a particular pattern of disconnection. The oxygen-containing group can be hydroxyl or carbonyl. If there are n minus 2 (n is less than 6) carbon between the two carbon atoms that are connecting to an oxygen-containing functional group, it obeys the 1, n dioxygenation pattern. The dioxygenation pattern gives clues to the pieces from which the target molecule is formed. In the discussion, the numbering starts from the carbon connected to the oxygen-containing functional group.

4. Guidelines

In the designing process of retrosynthetic analysis, the most significant step is to identify where to make disconnections. Several guidelines have been summarized to help with this hard task.

4.1 Guideline 1: Disconnections based on known, reliable reactions

Retrosynthesis in principle is the reverse process of synthesis. Therefore, the first idea of breaking down the compound should correspond to the reverse of a workable reaction. In figure 16 below, one common example shown is the ester retrosynthesis process. Alcohol adds on to acyl chloride is ester is one workable, well-known reaction so that we could simply write the retrosynthesis reaction based on the synthesis process.



Figure 16. Disconnection of ester

4.2 Guideline 2: Focusing on maximally simplifying disconnections

4.2.1 Simplify the most complex structure

One sample rule to decide the breaking point of a complex compound is to aim to separate the most complicated structure in the compound in order to reach the simplicity goal. In figure 17, the most complex structure is the saturated six-membered rings which should be the first task to target it down. Because of the complicated mechanism of composing the saturated six-membered rings, simpler substrates are needed. The reverse reaction of Path A represents a common process called A [4+2] cycloaddition in order to form the saturated six-membered ring. Therefore, compare to Path B, Path A breaks down the complex structure in one step which meets the goal of simplicity in retrosynthesis steps. Path B is not ideal due to the extra steps needed to break the saturated six-membered ring.



Figure 17. Break down of saturated six-memebred rings

4.2.2 Minimize Fine Tuning

Fine Tuning adds extra steps to the whole process and leads to a low-production rate. Avoiding minimizing finetuning steps is highly recommended when designing retrosynthesis.

4.2.3 Exploit Symmetry when possible

Implying symmetry in process of breaking down complex compounds helps simplify the retrosynthesis because it could reduce the steps and reserve the total amount of materials. It oftentimes insists we find a convergent pathway that serves the purpose of simplicity. [4]

In Figure 18[3], the target molecules 1,3-Propanediol can be simplified by breaking both functional groups of carboxylic acid and alcohol at both ends in one step. In this way, reactions at both sides happened at the same time also meet the goal of simplicity.



Figure 18. Symmetrical breakdown of 1,3-Propamedopl

4.3 Guideline 3: Maximizing Convergence

In synthesis designs, convergence synthesis is always more preferable to linear synthesis because convergence synthesis saves steps and aims at the simplest steps. Extra steps would diminish the total amount of original reactants because each step would have chemical yields which influence the final yields. The convergence retrosynthetic process serves the same purpose and aims to break down the complex compound into similar pieces.

In Figure 19[3], an example of a compound represented by A-B-C-D-E-F introduces the mechanism of the convergence process. When implementing a linear synthesis path, 5 steps needed to be complete; while in the convergent pathway, it designs to break down in the middle to reduce the total steps needed.

Linear synthesis: 5 steps	Convergence synthesis: 3 steps
A-B-C-D-E+F	A–B–C + D–E–F
A–B–C–D+E+F	A–B+C + D–E+F
A–B–C+D+E+F	A+B+C + D+E+F
A–B+C+D+E+F	
A+B+C+D+E+F	

Figure 19. Simple graphical decoposition representation

4.4 Guideline 4:Two-group disconnections are better than one-group disconnections

Two-group disconnections manage both functional groups into one step at the same time to serve the simplicity purpose. It is another way to reduce reaction steps to increase overall yields.



Figure 20. Two-group Disconnection of hydroxyl groups

In Figure 20[2], when we implement one group disconnection presented in path B, two hydroxyl groups will be generated which pose a chemoselectivity problem. In path A which represents a two-group disconnection process, the reagent attacks the styrene oxide structure (Figure 21.), which is the synthetic equivalent to the carbocation on the right side of Fig 20. It completes the two hydroxyl groups at one time which completes the retrosynthesis of two functional groups together.



Figure 21. Synthetic equivalent of carbocation

5. Dioxygenation pattern

A molecule that has two or more carbon attaching to the oxygen-containing functional group, including hydroxyl or carbonyl, has a particular pattern of disconnection. The dioxygenation pattern gives clues to the pieces from which the target molecule is formed. When dioxygenation patterns are discussed, the numbering of carbon starts from the carbon connected to the oxygen-containing functional group. If there is n minus 2 (n is at least 2 and at most 6) carbon between the two carbon atoms connecting to an oxygen-containing functional group, it obeys the 1,n dioxygenation pattern. For each dioxygenation pattern, there is a model procedure to disconnect the target molecule that obeys the pattern. Generally, the first step is to convert the target molecule to its canonical structure which is the standard structure to perform the disconnection as the results of its disconnection have a known pathway to produce the canonical structure as a product of a known reaction by functional group interconversion (FGI). Then, the disconnection of the canonical structure will lead to the fragments which are the reactants forming the target molecule. Therefore, by carrying out this dioxygenation pattern analysis, the formation pathway of the target molecule is obtained.

5.1 1,2 dioxygenation pattern

For a diol, dicarbonyl or hydroxy carbonyl target molecule that obeys the 1,2 dioxygenation pattern, the first step is to convert the molecule to its canonical structure 1,2-diol by FGI. Then the diol is disconnected to an alkene (See figure 22) as alkenes produce diols when added to osmium tetroxide (OsO4) to give diolate species that hydrolyze to cis-diols. The net process that produces diols from alkenes in the presence of OsO4 is called dihydroxylation (See figure 23).



Figure 22. 1,2 dioxygenation pattern disconnection



5.2 1,3 dioxygenation pattern

The canonical structure of a molecule that obeys the 1,3 dioxygenation pattern is a hydroxy ketone or an alpha-beta unsaturated ketone which is interconvertible with the hydroxy ketone. After the functional group interconversion of the target molecule, the 2,3 bond of the canonical structure is disconnected, and then ketone and aldehyde would be produced (See figure 24). To produce the canonical structure, the carbonyl is treated with a base to form an elonate. The elonate reacts with the carbonyl, for which the process is called aldol condensation (aldol reaction). If the aldol addition takes place, the product is a hydroxy ketone, or an aldol condensation takes place, the product would be an alpha-beta unsaturated ketone (See figure 25).



Figure 24. 1,3 dioxygenation pattern disconnection



Figure 25. Aldol reaction [6]

5.3 1,4 dioxygenation pattern

The canonical structure for the 1,4 dioxygenation pattern is a 1,4-dicarbonyl. Similarly, the target molecule is converted. The canonical structure is disconnected by 2 group disconnection, breaking the 2,3 bond (See figure 26). As the result of disconnection, the two conventional parts are methyl carbonyl and halogenated methyl ketone which can form the canonical structure by a series of reactions. The first step is to treat the methyl ketone with a base to produce an enolate (the enolisation). Then the alkylation reaction (See figure 27), in which the enolate reacts with the halogenated methyl ketone to produce the 1,4-dicarbonyl, takes place.



Figure 26. 1,4 dioxygenation pattern disconnection



Figure 27. Alkylation

5.4 1,5 dioxygenation pattern

1,5-dicarbonyl is the canonical structure for the 1,5 dioxygenation pattern. The first step is the FGI of the target molecule. Next, the 2,3 bond or 3,4 bond in the canonical structure is disconnected, producing a methyl carbonyl and an alpha beta unsaturated system (See figure 28). To form the 1,5-dicarbonyl, the enolisation of the methyl carbonyl takes place first, forming enolate. Second, the enolate reacts with the alpha-beta unsaturated system to produce the 1,5-dicarbonyl, which is termed Michael addition (See figure 29).



Figure 28. 1,5 dioxygenation pattern disconnection

The Robinson Annulation, Part 1: Michael Reaction



Figure 29. Micheal Addition [7]

5.5 1,6 dioxygenation pattern

To find the pathway of synthesis of a molecule that obeys the 1,6 dioxygenation pattern, it is a prioritized step to convert the molecule to the 1,6-dicarbonyl which has a canonical structure. The illogical disconnection of 1,6-dicarbonyl forms a carbon-carbon bond and produces a 6-carbon ring in the molecule (See figure 30). The ozonolysis of the product of the illogical disconnection produces 1,6-dicarbonyl which leads to the synthesis of the target molecule (See figure 31).



Figure 30. 1,6 dioxygenation pattern illogical disconnection



6. Extension

Here are two different syntheses which are planned by performing retrosynthetic analysis for a randomly chosen target molecule.

synthesis 1

First, by identifying the 1,4 dioxigenation pattern, the ring is disconnected to a straight chain. Second, the chain is disconnected into two parts according to the 1,5 dioxigenation pattern. As a result of the disconnection of the 2,3 bond, a piece that follows a 1,3 dioxygenation

pattern is produced. The piece undergoes FGI to perform the disconnection of the canonical structure. Therefore, two other smaller molecules are produced. Overall, three pieces are produced through the retrosynthesis process.



Figure 32. 1,4 dioxigenation pattern disconnection method

For this method, first identify that there is a 1,3 diox pattern. To disconnect the bond, first FGI the ketone into alcohol to get the canonical structure necessary for the 1,3 diox pattern. After the disconnection, we can then add an OR group (OEt in this case) onto the newly formed

carbonyl, because it's a better leaving group. Next, there is a 1,5 diox pattern. By disconnecting that creates a 1,4 diox pattern, which is the last step of the retrosynthesis. Thus, we get the final starting materials (see Figure 34).



Figure 33. 1,3 dioxigenation pattern disconnection method

7. Conclusion

Generally, the first step to perform the retrosynthesis of a complex target molecule is the analysis of the oxidation levels of each carbon atom. By analysing the oxidation levels, we can convert the bonds to electronegative atoms to bonds to oxygen. Therefore, we can find the possible dioxygenation patterns which indicate the type of disconnection performed. The result of disconnection will lead to simpler molecules that can be further disconnected. In addition, the dioxygenation pattern guides the pathway of the synthesis of the target molecule which is a purpose of the retrosynthetic analysis as it provides known reactions that can produce the target molecule from disconnected parts.

Retrosynthetic analysis is a crucial part to understand organic chemistry. In organic chemistry, we might encounter problems that are too difficult to us to solve. Learning retrosynthetic analysis enables us to think about problems from different perspectives critically, from forward to backward and then reversely. Having such a chemical intuition can help us to jump out of the dilemma and view problems with a different sight when we encounter harder chemical problems in the future. Thinking retrosynthetically can help us to deconstruct those complex and unlearned problems into something simpler that we are more familiar with. This way of thinking not only applies to our future chemistry study but also to our school lives.

Acknowledgement

All the authors contributed equally to this work and should be considered co-first authors.

Reference

[1] Loudon, G. M., & amp; Parise, J. (2021). Organic Chemistry. MacMillan Learning.

[2] Stoltz, B. (n.d.). Chem 41c quiz 7 - stoltz2.caltech.edu. Retrieved January 14, 2022, from https://stoltz2.caltech. edu/classes/chem41c/spring2011/quizzes/Quiz%207%20 Answer%20key%20S2011.pdf

[3] Clayden, J., Greeves, N., & Warren, S. G. (2012). Organic Chemistry. Oxford University Press.

[4] Retrosynthetic analysis and synthetic planning. (n.d.). Retrieved January 14, 2022, from https://profiles.uonbi.ac.ke/ andakala/files/sch_302_retrosynthetic_analysis_and_ synthetic_planning.pdf

[5] Periodic table. Ptable. (n.d.). Retrieved January 14, 2022,

from https://ptable.com/

[6] Organic Reactions Wiki. (2020, September 13). Aldol condensation. Chemistry LibreTexts. Retrieved January 14, 2022, from https://chem.libretexts.org/Bookshelves/Organic_ Chemistry/Supplemental_Modules_(Organic_Chemistry)/ Reactions/Organic_Reactions/Aldol_Condensation

[7] says:, W. Z., says:, J. K., says:, J. A., says:, A., says:, K. S., & amp; says:, M. (2020, December 16). The Robinson annulation. Master Organic Chemistry. Retrieved January 14, 2022, from https://www.masterorganicchemistry. com/2018/12/10/the-robinson-annulation/

[8] Ozonolysis. Oz. (2001, October 24). Retrieved January 14, 2022, from http://ursula.chem.yale.edu/~chem220/chem220js/ STUDYAIDS/ozonolysis/Oz.html#Overview