

The Overview of Retrosynthetic Analysis: Guidelines and Approaches

Luyi Zhang^{1,a,*}

¹*School of Material Science, Chemistry and Chemical Engineering, Hangzhou Normal University, Hangzhou, 311121, China*

a. 2913168570@qq.com

**corresponding author*

Abstract: Retrosynthetic analysis has, in recent years, become increasingly important, especially in designing total synthetic routes. Since retrosynthetic analysis was proposed, there are many substances whose synthetic routes have been deduced or further optimized, which greatly facilitated organic synthesis. Therefore, in order to further the study of organic synthesis, this paper will commence from the background of retrosynthetic analysis and relevant specific concepts to help novices have a preliminary understanding of retrosynthesis. Furthermore, this paper also illustrates some useful approaches of retrosynthetic analysis and some simple examples are used to better the understanding of retrosynthetic analysis, enabling novices to initially establish a basic knowledge system of retrosynthetic analysis.

Keywords: retrosynthetic analysis, disconnection, guiding principles

1. Introduction

Synthesis, the process of constructing a complex molecule from simpler ones, plays a vital role in organic chemistry. For some complicated molecules, there are infinite possible ways to synthesize and Retrosynthetic analysis is a method to think about the possibilities in a logical and quick way. Moreover, retrosynthesis is conducive to finding the most favorable and efficient synthesis route by comparing them in terms of various reagents and intermediates.[1] Such analysis was pioneered by Elias James Corey and was traditionally carried out by hand initially.[2] The synthetic planning process, as proposed by Corey,[3] advocates for commencing with the desired final product and subsequently working backwards towards the basic starting materials. In 1957, Corey introduced the term 'retrosynthetic analysis' to describe this methodology. Now, with the advancements in science, some computational approaches are used in retrosynthesis leveraging on reaction transformation rules from off-the-shelf and vast datasets, which can provide several feasible ways theoretically. [4,5]

One of the important molecules in retrosynthesis is Paclitaxel, a useful chemotherapy medication to treat several cancers. The supply of the natural products, however, was excessively limited non-renewable and non-environmentally friendly before. Until 1994, the first total synthesis of Paclitaxel was published by Robert A. Holton and his group. [1] Nowadays, other distinguished synthesis routes have come up to synthesize Paclitaxel, most of which utilize retrosynthetic analysis. Except for Paclitaxel, retrosynthesis is also adopted in synthesizing other valuable natural products, not only minimizing the cost but also conserving natural resources, which has a wide range of applications.

Given that the retrosynthesis exerts such a tremendous impact on synthesis, this paper is devoted to delve into primary strategies of retrosynthesis which make retrosynthesis more understandable to a for those newly introduced to the field. First followed are basic concepts.

2. Key concepts

Before learning the retrosynthetic analysis, this part will introduce some basic relevant glossaries below.

(1) Disconnection: An analytical operation that involves the breaking of a chemical bond, thereby converting a molecule into a potential reactant. It also means the reverse of a chemical reaction.

(2) Target molecule: This refers to the molecule whose synthesis is being intended. It is typically denoted as 'TM' and identified by its frame number.

(3) Synthons: These are general fragments, often in the form of ions, produced by a disconnection. Alternatively, some may refer to synthetic equivalents as synthons.

(4) Synthetic equivalents: These are reagents that serve the purpose of a synthon but cannot be used directly due to their instability.

(5) Reagent: These compounds are employed in a synthesis to generate intermediates or to give the target molecule itself. They can also be considered the synthetic equivalents of a synthon."

3. Retrosynthetic analysis

Synthesis is a laboratory technique used to obtain complex compounds what we want in the lab by mixing chemicals together and isolating the products of those reactions. As shown in the Fig.1, suppose synthesis as the process of building blocks, one building block is a star and the other is a circle and you can link them together by some kind of chemical reactions.

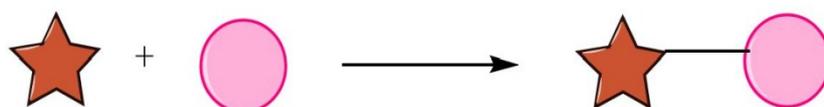


Figure 1: The process of synthesis

Retrosynthesis, on the other hand, is an intellectual approach that breaks down the complex molecules to a sequence of progressively simpler structures step by step in a logical and feasible way until the simple or commercially available starting materials are recognizable.[6]

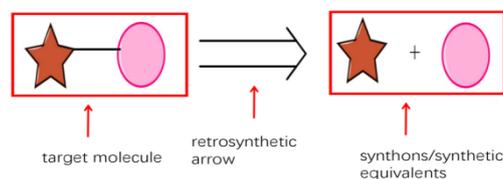


Figure 2: The process of retrosynthesis

As is revealed in the Fig.2, For example, back to this molecule, it can be broken into these two parts and the retrosynthetic analysis is conducted as follows. The left part is a target molecule and would be retro synthetically derived from the star plus circle. A double-lined arrow is used as the retrosynthetic arrow; it is important to distinguish this arrow from the single-lined arrow. Plus, the right part is usually regarded as synthons or synthetic equivalents.

The subsequent section starts from basics. As illustrated in the Fig.3, the first scenario is breaking one bond and consequently getting back to the linear structure. The second situation is breaking two bonds and thus going back to two carbon framework and one Carbon framework. And the third is breaking three bonds and therefore getting three atoms. One important thing you have to figure out is that breaking one bond is very easy and the disconnection of two bonds is also quite good. But its challenging to disconnect three bonds

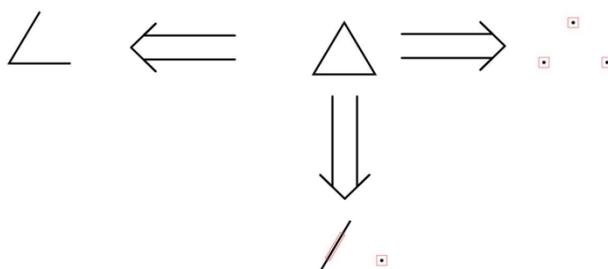


Figure 3: The disconnection of triangle

4. Retrosynthesis approach

4.1. Guiding principles:

4.1.1. Simplification

The first principle of retrosynthesis is simplification. In retrosynthetic analysis, our purpose is to deconstruct the complex compound into more elementary molecules so that we can design a favorable synthesis route. Otherwise, conducting such an analysis would be futile. The example of Fig.4 which complicates the problem can aptly shed light on the principle.

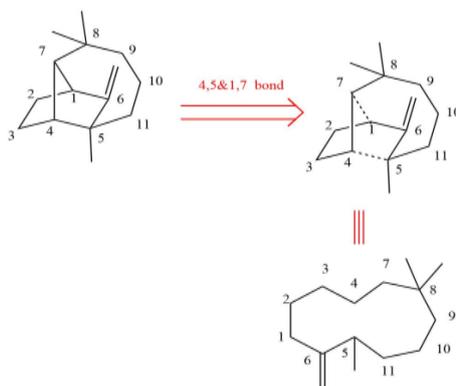


Figure 4: The disconnection of longifolene

4.1.2. Maximize convergency

The second principle is to maximize convergency. When encountering large molecules consisting of several units, it is best to divide the target molecule in half, which allows roughly the same molecular mass and complexity. As shown in the Fig.5, the convergence synthesis was displayed on the left her. Rather than shave off one part at the ends, at a time, we broke the molecule into two pieces. And rather than shave off one part at the ends, at a time, the molecule was broken roughly in half into two roughly equal and smaller fragments, ABC and DEF, which subsequently become the new target molecule. These can be synthesized in a minimum number of steps and then combined. This strategy

effectively reduces the synthesis effort and ultimately increases the overall yield of molecules. In contrast, the linear synthesis on the right requires more steps and has a lower yield.

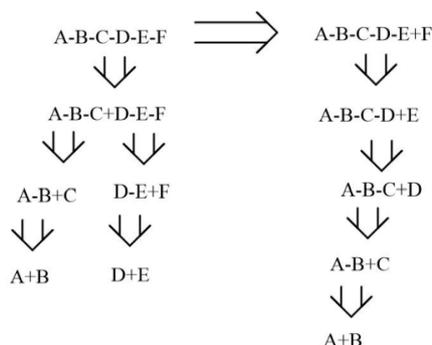


Figure 5: The comparison of two routes

4.1.3. Exploit symmetry when possible

In retrosynthesis analysis, utilizing the symmetry of molecules can help us get same or similar synthons, which enables simultaneous reactions on both sides and can simplify the problem and reduce the number of potential disconnection steps, thereby accelerating the identification of the most efficient synthetic routes. As illustrated in Fig.6, this molecule was broken into two identical parts and the other different synthon using the symmetry. It's apparent to know the route steps are reduced.

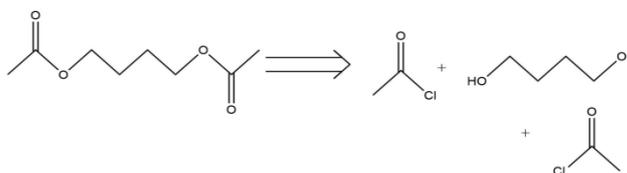


Figure 6: The disconnection of using symmetry

4.2. Fine tuning

(1) FGA: Functional Group Addition. Some functional groups need the addition of a group to the immediate precursor suitable for disconnection.

As is shown in the Fig.7, the carbonyl group is introduced to 1-phenylbutyl-2-ketone.

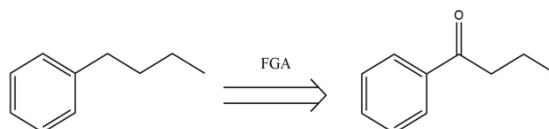


Figure 7: The example of FGA

(2) FGI: Functional Group Interconversion. The operation of writing one functional group for another so that disconnection becomes possible.

As is seen in the Fig.8, the acyl chloride is converted to an ester group.



Figure 8: The example of FGI

(3) FGR: Functional Group Removal. If a functional group is too complicated and figure out how to install it at a later time in the synthesis, particularly if it's in proximity to some other functional group, bringing it back is of great possibility. This transformation involves the removal of a particular functional group during the retrosynthesis. As is seen in the Fig.9, the alcoholic functional group is removed from the 6-hydroxycyclohex-2-enone to give us cyclohexanone.

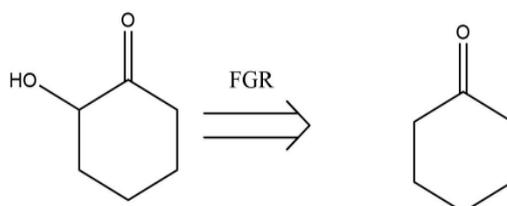


Figure 9: The example of FGR

But fine tuning doesn't actually simply the problem, which just shifts the problem. In some cases, this approach could even exacerbate the complexity of the problem. Consequently, the practice of fine-tuning should be minimized to avoid unnecessary intricacies.

4.3. Consequences of bond breaking

The disconnection of a bond generally yields three outcomes. As illustrated in the Fig.10, the first scenario is that two electrons flow to the right after disconnection, making the left carbon a cation and the one on the right anion. The second possibility is that both two electrons flow to the left, which makes the right carbon electronegative and the left carbon electronegative. And the third case is that one electron goes to the right side and simultaneously one electron goes to the other side, thus getting free radicals.

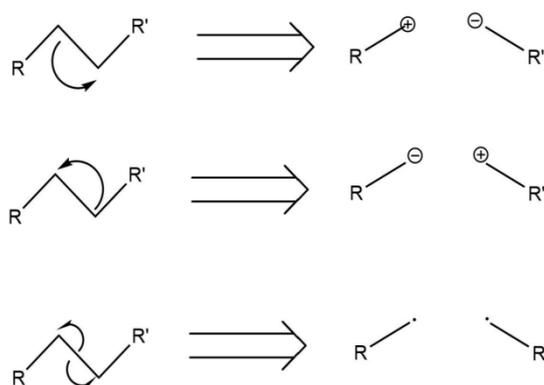


Figure 10: The Consequences of bond breaking

Each broken bond has three different possibilities. However, the more complex the organic compound, the more possibilities there are. What we usually need to do is use the polarity embedded in target molecule to help guide the disconnections and use the same kind of arrow pushing formalism to help us as we go through the disconnections.

4.4. Electrocyclic Disconnection

(1) An electrocyclic disconnection is like a cycle direction in the reverse. For instance, as depicted in the Fig.11, electrons are moving around the six-member ring in order to go back to a four-carbon diene and a two carbon olefin system in the forward direction.

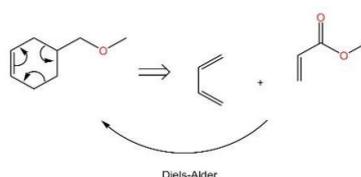


Figure 11: The example of Electrocyclic Disconnection

4.5. Dioxygenation pattern

(1) 1,2 Dioxygenation pattern

As depicted in the Fig.12, the molecules of 1,2 dioxygenation pattern can convert to a olefin through inner conversion.

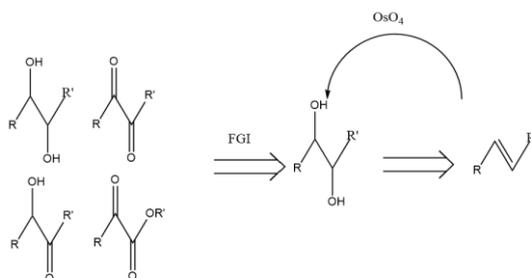


Figure 12: The example of 1,2 Dioxygenation pattern

(3) 1,3 Dioxygenation pattern

In terms of 1,3 dioxygenation pattern, the polarity should be considered. As depicted in the Fig.13, the molecules of 1,3 dioxygenation pattern can convert to a ketone and aldehyde through the moving of charges.

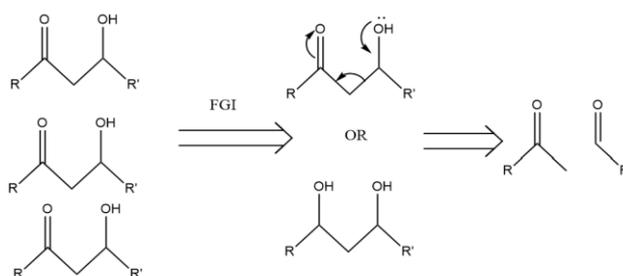


Figure 13: The example of 1,3 Dioxygenation pattern

4.6. Illogical Disconnection

As shown in the Fig.14, This is also 1,6 dioxygenation pattern. In this pattern you could break apart different bonds, internal to that chain. But an alternative strategy is to select carbon atom number one and carbon atom number six, and to reform the bond between them. This is a disconnection, but we're actually connecting a bond,

In this instance, we achieve simplification through bond formation, utilizing a well-known elementary chemical reaction that facilitates the breaking of the carbon double bond.

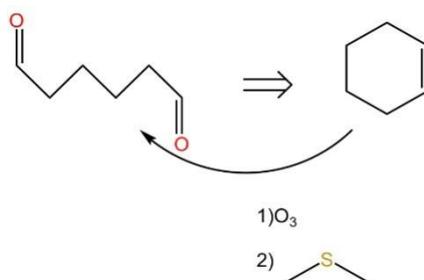


Figure 14: The example of 1,6 Dioxygenation pattern/illogical pattern

5. Example

To further consolidate the previous learning, this paper has given an example to enable us to perform a more integrated and comprehensive retrosynthesis analysis, as depicted in the Fig.15

Analysis: disconnect the α , β -unsaturated aldehyde.

This is a α , β -unsaturated aldehyde, so in the first step, we can disconnect the bond as shown in the figure below and get a 1,6-dicarbonyl compound. Then we can re-connect it into a cyclohexene as we learned above. Obviously, this molecule has the oxygenation pattern of a Diels-Alder adduct if we convert it to a carbonyl compound.[7] Finally, we can get the starting materials.

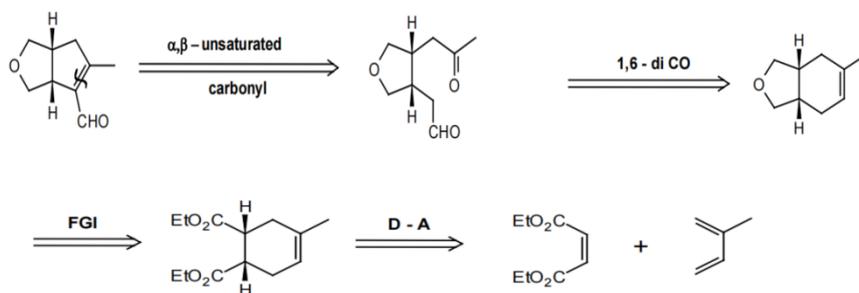


Figure 15: [7]The example of retrosynthesis

6. Conclusion

Based on the results and discussions presented above, the conclusions are obtained as below:

- (1) This paper explores the main guidelines and approaches of retrosynthetic analysis. The point lies in the different types of disconnections. The highlight is judging flow directions of the electrons by the substances itself.
- (2) This paper aims to help novices to establish a basic understanding of retrosynthesis, in the hope of stimulating the thinking and creativity and thus laying a solid foundation of the study of retrosynthesis in the future.

- (3) However, there remain certain deficiencies, including the lack of a detailed principles of the methodologies employed, and the omission of application of AI in this domain,

Acknowledgments

The authors would like to acknowledge the contributions of Professor Brian Stoltz who renders insightful comments and suggestions as well as teaching assistant Jeremy Liu.

References

- [1] *Importance and Impact of Organic Synthesis and Retrosynthesis in the Field of Chemistry*. (2019). Iqvia.com. <https://www.iqvia.com/blogs/2019/11/importance-and-impact-of-organic-synthesis-and-retrosynthesis-in-the-field-of-chemistry>
- [2] Genheden, S., Thakkar, A., Chadimová, V., Reymond, J.-L., Engkvist, O., & Bjerrum, E. (2020). AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of Cheminformatics*, 12(1). <https://doi.org/10.1186/s13321-020-00472-1>
- [3] *The Nobel Prize in Chemistry 1990*. (n.d.). NobelPrize.org. <https://www.nobelprize.org/prizes/chemistry/1990/8823-retrosynthetic-tree/>
- [4] Jiang, Y., Yu, Y., Kong, M., Mei, Y., Yuan, L., Huang, Z., Kuang, K., Wang, Z., Yao, H., Zou, J., Coley, C. W., & Wei, Y. (2022). *Artificial Intelligence for Retrosynthesis Prediction*. *Engineering*. <https://doi.org/10.1016/j.eng.2022.04.021>
- [5] Watson, I. A., Wang, J., & Nicolaou, C. A. (2019). A retrosynthetic analysis algorithm implementation. *Journal of Cheminformatics*, 11(1). <https://doi.org/10.1186/s13321-018-0323-6>
- [6] COREY, E. J. (1989). *ChemInform Abstract: Robert Robinson Lecture. Retrosynthetic Thinking - Essentials and Examples*. *ChemInform*, 20(7). <https://doi.org/10.1002/chin.198907330>
- [7] Warren, S. G. (1978). *Designing organic syntheses: a programmed introduction to the synthon approach*. Wiley.