Introductory to Organic Synthesis and Retrosynthesis Analysis

Yutong Ye

Abstract

Retrosynthesis is an indispensable and auxiliary technique used by chemists in assisting the organic synthesis of the complex desired product. This study aims to teach high school students how to do retrosynthesis analysis for organic molecules by considering the disconnection strategies, and basic approaches, for instance, fine-tuning & protecting functional groups and fundamental principles when doing the analysis. It also involves some techniques that could apply to a specific pattern, such as deoxygenation. Two complex molecules are given as samples, and their retrosynthesis will be exploited.

Keywords: retrosynthesis analysis, disconnection, dioxygenation pattern

1. Introduction

As human civilization developed, organic compounds has became more widespread and crucial in people's everyday life. As time passes, more and more naturally occurring molecules have been used up by humans, and the drugs isolated from plants through phytoextraction and animals are not even touching our demand. However, the turning point of this circumstance occurred in 1824, when a German chemist, Friedrich Wohler was evaporating a solution of ammonium cyanate. This compound can be used to treat skin problems such as eczema and callus which means it can be used in specified hand creams. Its other purpose is to remove dead tissues in the wound to assist in healing. This remarkable discovery broke the original theory which states that organic compounds could only be found in animals and plants, Whole the legendary started a new era in the field of organic chemistry [1]. Later, he inspired other organic chemists, and as time passed, more organic compounds gained their synthesis route.

Nonetheless, there is still a problem of how to plan and choose the proper reaction route for the desired molecule. With this issue, E.J. Corey introduced a method called retrosynthesis analysis in his book *"Logic of Chemical Synthesis",* depending on that he won the Nobel Chemistry price in 1990 [2]. The idea was widespread in the chemical community and became the most fundamental and idiomatic method for organic compound synthesis. Paring up with organic synthesis, they have extensive and crucial uses such as making chemical drugs, military weaponry, and cosmetics purposes. Surveys show that 80 per cent of medicine was prepared through those methods. While chemists are analyzing their retrosynthesis, they should cover several features for that route to ensure it is feasible [3]. For instance, the reaction steps must be as few as possible, as fast the reaction rate as possible, whether the desired product will have a high yield or not and if the starting materials are cost-effective and easy to access [4].

2. Background Information

We have talked about retrosynthesis analysis for so long, but how does it work? Retrosynthesis analysis works by predicting what the latest step could be in synthesizing the targeted molecule, which is the desired product. Disconnections could be made to break the target molecule aiming to simplify it into a smaller precursor, then do the disconnections on our precursor [5]. After a certain number of disconnections, we should get the basic building block called synthons [2]. They are the most simplified product we could get in retrosynthesis analysis. Now the paper will give a deeper explanation for making disconnections.

3. Disconnection

Remember that disconnections are aimed to make the molecule simpler, so disconnections could be made in two ways, one is obeying actual reactions, and the other is to disconnect in a way that chemist thinks helps simplify the molecule. Although both strategies are allowed, a detail should be noticed by both, that is, specific signs should be applied for different purposes, as can be seen in figure 1.

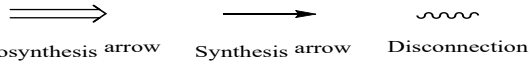

Figure 1 symbols for retrosynthesis analysis

3.1 *Understand Strategy Through Triangle*

As figure 2 showed, A triangle can be used as a sample to help us understand strategic retrosynthesis, and this is when we disconnect bonds as we like if it helps simplify the target molecule into precursors; yet sometimes, the forward reaction may not succeed.

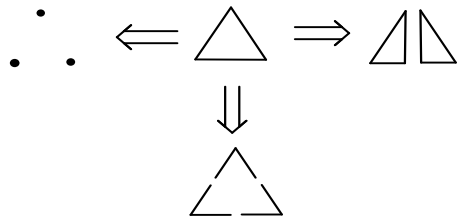

Figure 2 triangle retrosynthesis

The triangle can be disconnected into three edges, three small dots and two smaller triangles adn more. For the first two conditions, we are simplifying the shape by disconnecting edges and sides while thinking of the shape in 2-dimension. For the other condition, we consider the triangle as a 3-dimension project that can be made from two tiny building blocks, which may not seem as simplifying as the previous ones. Whereas little blocks will be easier to find than larger ones. Remember, the aim is not only to simplify the molecule but also to get starting materials that are easy to access and cheap. Hence, the third one is also an effective disconnection.

3.2 Strategic C-C Bond Disconnections

3.2.1 Decalin:

Only dealing with c-c bond disconnections is very similar to the shapes we covered before. Now, using a real organic compound as an example to demonstrate strategic retrosynthesis. Figure 3 is a molecule called decalin, a two-ring system.

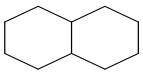

Figure 3 Decalin

How many and which bonds we disconnect could give off very different precursors. Some are simpler than others. How simple a precursor is, is a pretty subjective opinion, whereas leaving a long chain behind would be more strategic under some conditions. Which structure is more concise depends on the chemist who is doing the retrosynthesis analysis. Also, remind that there is a very principle in strategic retrosynthesis that states that the number of atoms of each element is conserved during the disconnections. Detail will be told with the examples below in figure 4.

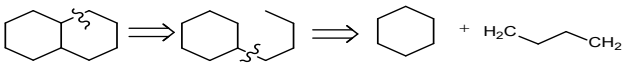

Figure 4 strategic disconnection

Let us consider disconnecting two bonds of decalin in Figure 4, after disconnections, we get a cyclohexane and a four carbon fragment which is not a butane as we said before, chemists call them by their carbon number for convenience. And is this case it is a [6 + 4] disconnection. If to think carefully, the forward synthesis will only involve reactions between the two carbons on the side of the ring and the four carbon fragments, for it is a [2 + 4] cycloaddition. The term cycloaddition will be exploit later in the paper.

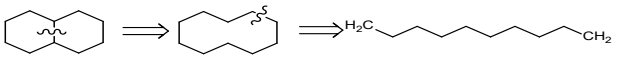

Figure 5 strategic disconnection

As can be seen in figure 5, here we are still disconnecting two bonds; however disconnections in different positions, the result is very different from the one we get in Figure 4, where a ten is generated, which is a long chain. It may be better than leaving a ring behind because further disconnection could be applied to the ring which adds extra steps.

3.2.*2 Common Atom Approach*

Common atoms are carbon atoms that are shared between multiple rings, and the disconnection of bonds linking those atoms is a method called common atom approach. It is a universal way that significantly simplifies complex ring systems and is classified as strategic disconnection (figure 6).

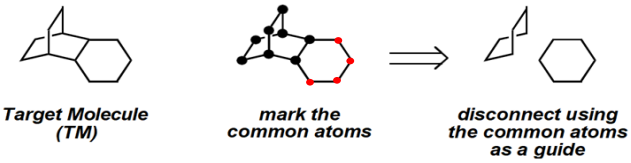

Figure 6 common atom approach

In this case, the common atoms of the target molecule are marked in bold black dots. Even though the bold red atoms are also shared between several rings if see the structure as a whole ten carbon ring, they are not considered in the approach because making changes and disconnecting them does not simplify the structure, the most complicated part is still there, not doing the job.

3.3 Synthons & Synthetic Equivalents

3.3.1 Synthons

Precursors are partly simplified molecules that result from the disconnection of the target molecule, and synthons are the simplest precursor, usually obtained from several disconnections. The term ‘synthon’ is first introduced by E.J.Corey, describing them as synthetic building blocks. [2] Synthons appear in several states, usually positive and negative ions, but can also be free radicals. Therefore it is obvious that it is not the natural occurring reagents for the reaction but hypothetical reagents. Figure 7 shows what synthons are, there are three circumstances for the

electron to move in the targeted molecule, resulting in disparate pairs of synthons.

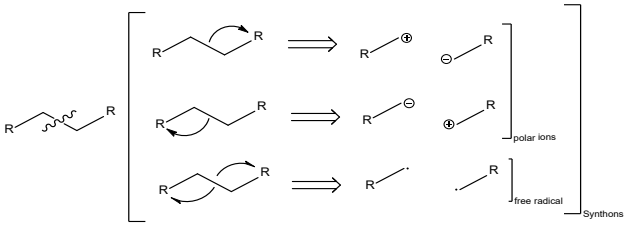

Figure 7 types of synthons

3.3.*2 Positive & Negative Synthons*

As said before, synthons can have diverse charges, fragments with a positive charge are considered as acceptors (of pairs of electrons) synthons, which acts like electrophiles. Vice versa, fragments with a negative charge are considered to be donor synthons that acts like a nucleophile, one classic synthetic equivalent for it would be the Grignard reagent.

3.3.*3 Synthetic Equivalents*

The actual reagent in the forwarding reaction corresponding to the synthons are called synthetic equivalents. In other words, they are sources of synthons. See Figure 8.

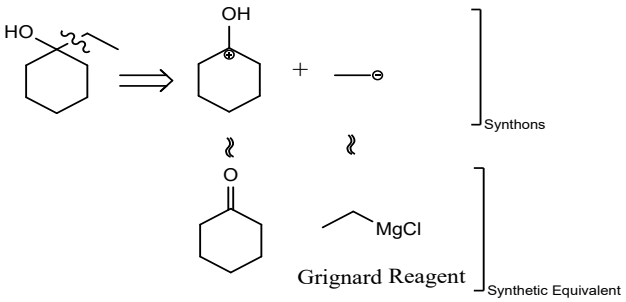

Figure 8 demonstration of synthons and synthetic equivalents

Here the target molecule is symbolized as the letter "E" by imagining E's last letter, we could notice the disconnection should be made in order to get the precursor " D ". Yet D is not simplified enough to be able to be our synthons so that further disconnection could be made to get the synthons, in this case, A, B and C. The rest is easy, we only need to figure out how to react them together to get the target molecule, E. (figure 9)

$$\mathbf{E} \Longrightarrow \mathbf{D} \Longrightarrow \mathbf{A+B+C}$$

Figure 9 process of retrosynthesis

However, when we are dealing with those disconnection routes, we want to avoid routes which pose a chemoselective issue when synthesizing the molecule [6,7]. There may be a trend where a functional group is aimed to react with another functional group which may not get us the desired product.

3.3.4 Grignard Reagent

A standard, occurring synthetic equivalent is the foremost resource for an alkane, an R group anion. It is an organometallic reagent between an organic and inorganic compound. The forward reaction of carbonyl reduction using the Grignard reagent is shown below in figure 10. The reaction is back up with water or acid.

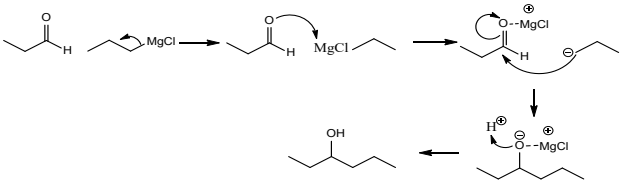

Figure 10 use of Grignard reagent in carbonyl reduction

3.4 Types of Disconnection

3.4.1 One Functional Group Disconnection

The functional group plays a crucial role in deciding and planning the route for disconnection. Out of an abundant number of various functional groups, alcohol is the most classic and commonly used [8]. Since it is easy to get from other functional groups, for example, adding a Grignard reagent to the carbonyl. It also would be converted into other functional groups through plain reactions, figure 11 has a wide range for alcohol's conversion into other functional groups. For example, we could oxidise alcohol to aldehyde by adding potassium dichromate (VI). Once aldehyde is formed will be oxidised to carboxylic acid.

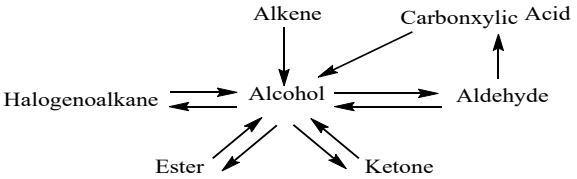

Figure 11 convertions between functional groups [9]

There are an abundant number of reactions with different functional groups, even dealing with one functional group there are still many routes for disconnection and each one will give varies synthons with different yields at desperate reaction rate. Take the molecule 2-butanol as an example, see Figure 12.

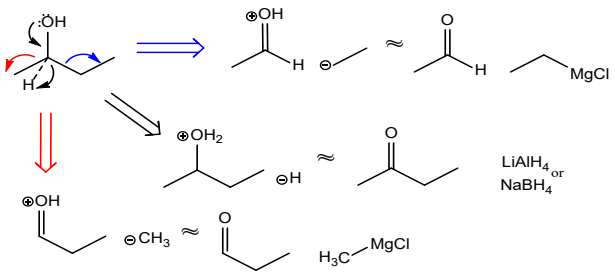

Figure 12 differences for the disconnection

When doing disconnection about functional group or heteroatom, start transferring electrons from the lone bond, in this case since there is an alcohol functional group, start moving electrons from the oxygen's lone pair would be feasible. Then the carbon would have been connected to five bonds, that not possible, so one of it's bond should be broken, then there comes the decision making, three routes are available for the future retrosynthesis. This process of moving electrons around the compound directed by curly (half) arrow is called electron pushing or arrow pushing, which is very common (figure 13).

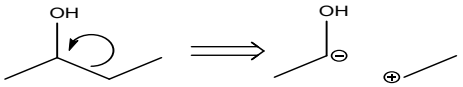

Figure 13 unique arrow pushing

The disconnection is possible but not as common as starting from lone pair, less feasible deconstruction.

3.4.2 *Two Functional Group Disconnection*

The aim of two functional group disconnection is to simplify molecule by separating them apart from each other. For functional groups that are attached at different places in the carbon chain or ring, different strategies would be applied to deconstruct the molecule.

3.4.3 *Electrocyclic Disconnection*

The electron flows in the ring will guide the retrosynthesis strategy, a renowned example of it is the Diels-Alder Cycloaddition. See Figure 14. During the 4+2 cycloaddition retrosynthesis, two synthons are produced, however unlike universal synthons they don't carry any charge. This is another type of synthons whose synthetic equivalent is themselves, or in other words they don't need them since they are actual occurring and are involved in forward synthesis.

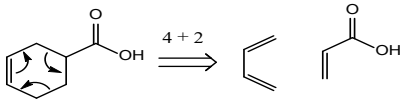

Figure 14 reverse of cycloaddition

3.4.4 *Illogical Disconnection*

A very unique type of disconnection is illogical disconnection, as it sounds, illogical disconnection does not obey the original concept of disconnections, that is, not breaking any bonds but form bonds instead (figure 15).

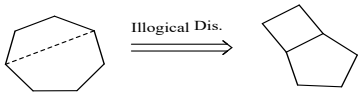

Figure 15 example of illogical disconnection of cycloheptane

4. Functional Group Assisting

4.1 *Fine Tuning*

When dealing with functional groups, problems will occur such as having no idea where to start the retrosynthesis disconnection. Then we need the technique to get us on the right track, fine tuning would be a suitable approach to start with, there are three different types of fine tuning, FGI, FGA and FGR. These are little tools which helps to plan route for disconnection by doing minor changes to the functional group of an organic compound.

4.1.1 *Functional Group Interconvertion (FGI)*

FGI works by converting the functional group in a molecule into another (figure 16).

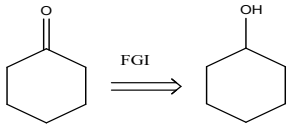

Figure 16 FGI of cyclohexanone

FGI can also help the forward synthesis by protecting a group during the synthesis. If the starting material and the targeted molecule have a same reactive functional group in common. It would react with other reagents, then the possibility of it to occur in the targeted molecule is negligible. So we need a way to let it not be substituted or changed during the forward synthesis by a substance called protecting group.

4.1.2 *Functional Group Addition (FGA):*

FGA is to add a functional group to the target molecule (figure 17).

Figure 17 FGA of butane

4.1.3 Functional Group Removal (FGR):

Is removing a functional group from the origin molecule early in the retrosynthesis, since that group is quite hard to

get at the start of the forward synthesis (figure 18).

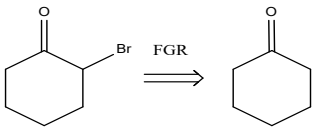

Figure 18 FGR is to remove a functional group

4.2 Protecting Group:

As known, functional groups are important in the molecule functioning and guiding retrosynthesis disconnection. If during the synthesis functional group that is desired in the targeted molecule is destroyed or substituted by other functional group, that will pose problems for instance form a compound that's not functioning, side reactions, low rate and yield, or even not synthesizing anymore, none of those effects is favourable in organic chemistry. Solutions can be carried out by using some reagent which adhere to that functional group in early stages and in disguise for the whole process and then converted back into the functional group in late steps of the synthesis [10]. These reagent of preventing the group to be changed in organic chemistry is called protecting group. Two classic example is protecting hydroxyl group and carbonyl group.

4.2.1 *Hydroxyl Group Protection:*

For this protection, a TBDMS-Cl is used, where under acidic conditions binds with the oxygen atom, protecting the oxygen from other reactions, and at last converted back to alcohol group using TBAF (figure 19).

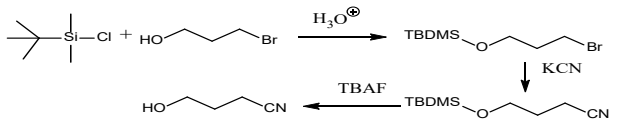

Figure 19 protecting hydroxyl group

4.2.2 Carbonyl Group Protection:

Carbonyl is a quite reactive group which if we want it to present in the target molecule we need to use a protecting group named ethyleneglycol. An example is shown below in figure 20.

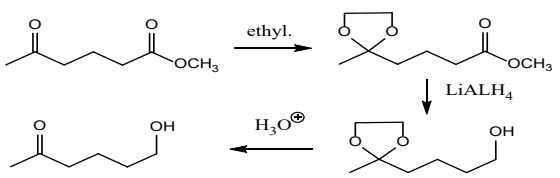

Figure 20 using ethylene glycol to protect carbonyl group

Ketone is more reactive than ester group, yet in this case the final product have the ketone remained and the ester turned into hydroxyl, so quite obvious that the ketone is protected by a protecting group named ethylene glycol, then convert back under acidic conditions.

So in retrosynthesis analysis, if a relatively less reactive functional group in starting material is converted into another in the target molecule while the other functional groups not converting, chemists would notice that a protecting group should be used when doing forward synthesis.

5. Retrosynthetic Analysis Guiding Principles

5.1 *Guideline 1: Focus on Simplifying*

The main aim of retrosynthesis is to disconnect the target molecule to maximize the simplification of its structure. In order to initiate the synthesis with a cheap and easy accessible starting material. The very first step is to lock the most complex structure in the targeted molecule, disconnecting that structure will simplify the whole molecule the most.

5.2 Guideline 2: Maximizing Convergence:

In a series of organic block, disconnection in the middle of the chain is preferred as it generates the synthons using less steps than the linear retrosynthesis. Having extra steps like the linear one would generate low yield of the target molecule, which small amount would be produced since each of the extra steps have disparate yields which lowers the total yield of the reaction (figure 21).

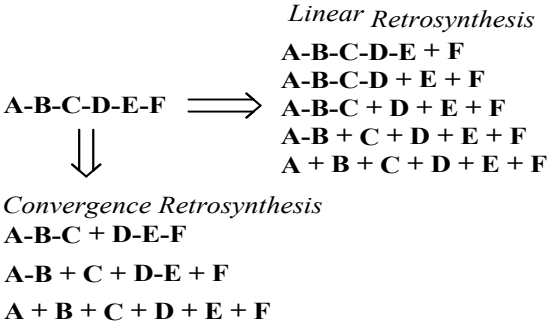

Figure 21 difference between convergent and linear disconnection

In this case, we disconnect convergently in the middle, then we get A-B-C and D-E-F as our precursors, further disconnections could be made, whereas if we can get A-B in real life commercially, then the further disconnection don't need to be applied.

5.3 Guideline 3: Minimizing Fine Tuning:

Even though that fine tuning is a helpful tool in planning the retrosynthesis route, minimizing fine tuning is strongly recommended in retrosynthesis analysis as it adds extra

steps into the reaction which lowers the total rate and percentage yield. We should focus on a bigger picture than those small changes.

5.4 Guideline 4: Two Group Disconnection > One Group Disconnection

Because both of the two functional groups are involved in the disconnection where one group does not. Maximizes the participation of the whole structure, also activating both the functional group for further disconnection.

5.5 *Guideline 5: Look For Symmetric In Compound*

Symmetrical structure in the targeted molecule means the two side could be deconstructed in the same way which reduces unnecessary steps and reduce the diversity of starting materials. It would do two disconnections in one single step. The pentane-1,5-diyl diacetate has a line of symmetry on the middle carbon, highlighted in pink. The two ends are seen as ester, disconnection on the bond connecting the oxygen and the carbonyl group is also reflected on the right side, this reduces steps needed in retrosynthesis (figure 22).

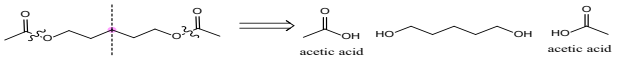

Figure 22 exploiting symmetry in molecules

6. Dioxygenation Pattern

It is the pattern for disconnecting two functional groups separated by differnet number of carbons in-between. For different places we have varied ways to disconnect them. Where taking the targeted molecule's functional groups to their canonical functional groups then a common disconnection could be applied. The set of patterns lists from 1,2 to 1,6 dioxygenation, the 1,3 and 1,5 dioxygenation are often known as logical disconnections which they direct get the synthons with no extra reaction. While the others are illogical.

6.1 *1,2 Dioxygenation Pattern:*

When the two funtional groups are just near to each other in the targeted molecule, best way is to use them together [11]. The two carbons connecting functional groups can be both in ketone or alcohol oxidation level, where we treat them by FGI into their canonical structure which is shown below in figure 23, 1,2 alcohol. Then remove the two hydroxyl functional groups and we get an olefin.

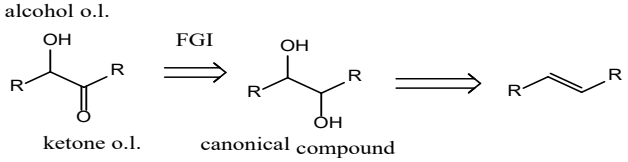

Figure 23 disconnecting 1,2 functional group

The olefin could be converted back to the canonical compound reacting with osmium tetroxide (OsO_4) as shown below in figure 24. Reaction where the osmium tetroxide is involved in reaction with alkene is called dihdroxylation.

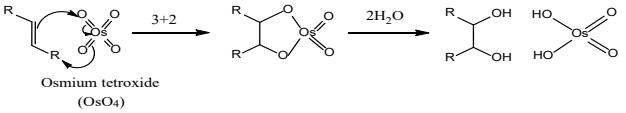

Figure 24 Dihydroxylation

6.2 1,3 Dioxygenation Pattern:

Compounds that have carbonyl or hydroxyl group attached to carbon atoms separated by one carbon atom. Firstly should be converted by FGI into the canonical compound where carbonyl on one carbon and an alcohol group on the other. Disconnection on the 2,3 bond leads to the formation of two carbonyl functional groups (an enolate of ketone and an aldehyde). Aldol reaction is where we get the enolates back to the canonincal compound in forward reaction (figure 25).

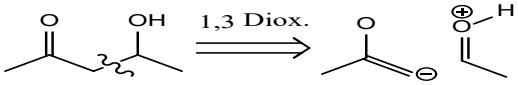

Figure 25 through 1,3 disconnection

6.3 1,4 Dioxygenation Pattern:

The canonical structure for the 1,4 dioxygenation pattern is a symmetrical 1, 4 dicarbonyl molecule. With breaking the middle bond we can get a negative carbonyl and a postive carbonyl in both ways, the negative ion is the enolate for ketone which can be get from ketone deprotonation. The positive ion has a synthetic equivalent of alkyl halide (figure 26).

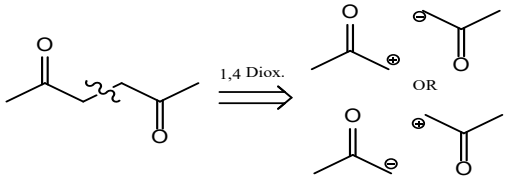

Figure 26 1,4 functional group disconnecting

6.4 1,5 Dioxygenation Pattern:

The first step of 1,5 dioxygenation disconnection is to convert the TM into 1,5 pattern's specified canonical structure, which is just like the 1,4 pattern, a dicarbonyl on the carbons responsible for the position. We could disconnect either bond a or bond b, both gives similar synthons. An α-β unsaturated structure and an enolate for carbonyl group. We can get the synthetic equivalents back to the canonical structure through Micheal Reaction (figure 27).

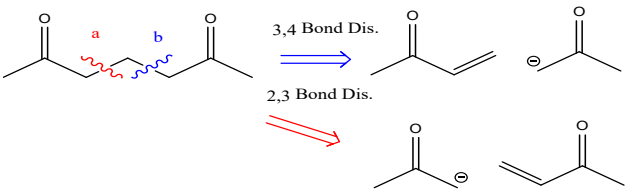

Figure 27 two routes for 1,5 disconnection

6.5 1,6 Dioxygenation Pattern:

1,6 dioxygenation retrosynthesis is an illogical disconnection where the two carbonyl group were joined to form a cyclohexane. We should first convert the target molecule to a 1,6 dicarbonyl structure by FGI which is the canonical structure for 1,6 dioxygenation disconnection. The reverse reaction could be done through ozonolysis (figure 28, 29).

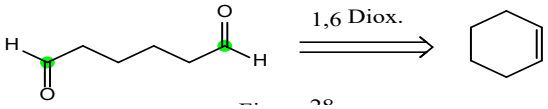

Figure 28 1,5 functional group disconnection

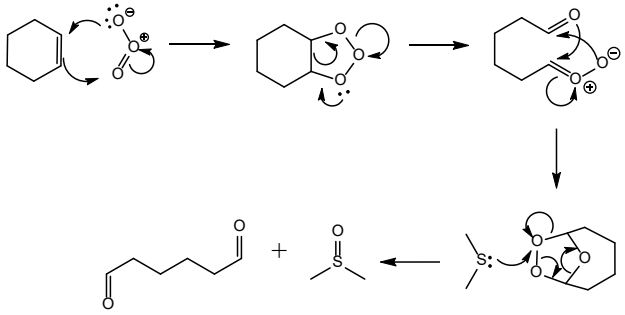

Figure 29 ozonolysis

7. Problem Practice

7.1 *β hydroxyl ketone:*

β-hydroxyl-ketone is an organic compound with different relative locations between functional groups. The ketone and the alcohol group is a 1,3 relationship marked in red, and the other is 1,2 relationship numbered in blue. In this case we could consider disconnections of 1,2 and 1,3 dioxygenation pattern (figure 30)

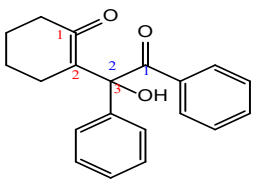

Figure 30 β hydroxyl ketone molecule

Therefore two ways can be applied for the retrosynthesis analysis. Disconnection a is to disconnect abide to 1,2 dioxygenation pattern, where the electron pair is removed to the carbon bonded to alcohol group, resulting in an alcohol anion and a carbonyl cation whose synthetic equivalent is a acid halide. Disconnection a poses a problem which is to find synthetic equivalent for unstable and unknown synthon, so not feasible (figure 31).

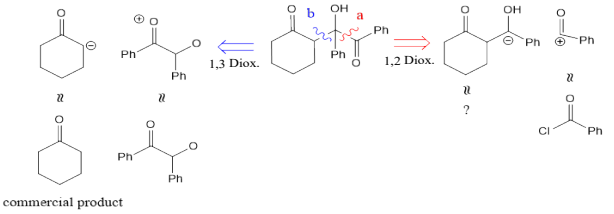

Figure 31 two retrosynthesis route

With disconnection b we are following the 1,3 dioxygenation pattern where results in a cyclohexanone as synthetic equivalent, a commercial compound, and the other compound obeys the 1,2 dioxygenation patter, therefore gets us an alkene.

7.2 Sitagliptin:

The first thing we should notice when doing retrosynthesis is to identify the dioxygenation pattern in the target molecule if any. In stagliptin, between the two cycles there are two carbon atoms with an alcohol and a ketone oxidation level, therefore obeys the 1,3 dioxygenation pattern. Now we should convert that part into the corresponding canonical structure of that specific patter. In this case we convert the amide functional group into a carbonyl through FGI. Then we could identify that there is an amide group in between of the converted carbonyl and the cycle, through one functional group gets us molecule 3 and 4. This is also a convergent disconnection where the molecule is broken in half therefore each two precursors have their own synthesis route. Now it is quite obvious that molecule 3 has a dicarbonyl structure which also belongs to canonical structures for 1,3 disconnections, therefore helps us to get molecule 7 and 8. For molecule 4, it is has a multicyclic structure with many heteroatoms (N) attached inside, so through heterocyclic disconnection we should be able to get molecule 5 and 6. That is the brief retrosynthesis analysis for the drug stagliptin (figure 32).

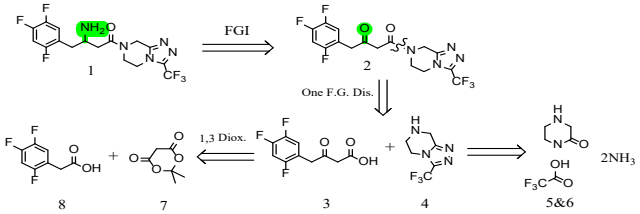

Figure 32 Retrosynthesis for Stagliptin molecule [12]

8. Conclusion

So when dealing with a complicated target molecule we first should try to find if there's any carbons abiding to the dioxygenation pattern or any other pattern, and initiate the disconnection from there. Or if the complex part of the molecule has a multicyclic structure you may use common atom approach to simplified there. Sometimes the cyclic structure have some heteroatoms involved in the rings, therefore consider about the oxidation level of the carbon atoms in it, that will give you an idea of how the precursors may be. And remember retrosynthesis analysis is all about making the problem simple and getting a cost-effective starting material, high yield, high reaction rate should also be considered.

Reference

[1] HanMei.Mao ZhengQin.Ji Organic Chemistry

[2] Logic of Chemical Reactions by E.J.Corey & X.M.Cheng (1995)

[3] Yu.Mo LinJia.Zhu Organic Synthesis and Retrosynthesis Analysis of Organic Compounds in Life Sciences of the New Century

[4] FengYan. Han Application of Retrosynthesis Analysis in Organic Chemistry Education

[5] Wikipedia Retrosynthesis Analysis

[6] James Law et.al:"Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation", Journal of Chemical Information and Modelling (ACS JCIM) Publication Date (Web): February 6, 2009;

[7] Organic Chemistry (Second Edition) by Clayden Jonathan, Greeves Nick & Warren Stuart.

[8] Wang Ming Design of Organic Synthesis Route : Retrosynthesis Analysis

[9] Edexcel International A Level Student Book 2 by Curtis Cliff, Murgatroyd Jason & Scott David

[10] Baidu.baike Protecting group

[11] Designing Organic Syntheses by Warren Stuart

[12] C.K. Savile, J.M. Janey, E.C. Mundorff, J.C. Moore, S. Tam, W. R. Jarvis, J.C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G.W. Huisman, G.J. Hughes, Science 2021, 329, 305-309;