

Introduction to Retrosynthesis: Strategies and Approaches

Haoyu Wang^{1*}, Guanghao Wei²

¹St. John's School, Vancouver, BC V6K 2J1, Canada

²Hongwen School Qingdao Campus, Qingdao, 266100, China

Abstract—Chemical synthesis is a powerful tool for human being as it provides many useful products, especially those that do not exist or only exist in a small quantity in nature. Since it is often hard to design and visualize the reactions of organic synthesis, retrosynthesis is often used to achieve a desired product. This review is an introduction of the concepts of retrosynthetic analysis and how it can be established, including the basic steps and strategies. It also discuss the guiding principles of retrosynthesis (simplification, complexity, yield) along with some possible problems and coping strategies in the analyses, including functional group disconnections, fine tuning, and illogical disconnections.

1. Introduction

Chemical synthesis, the process that converts one or more compounds into different compounds via execution of a series of reactions, is a powerful tool for human being. It provides humans with many useful products that are important in many important fields, such as pharmaceutical, construction, agriculture, and higher-technology industries. Many products that do not exist in nature or are not easy to obtain in large scale from nature must be synthesized chemically. Taxol (see Figure 1), for instance, is a drug commonly used today in treating various types of cancers. Originally isolated from the bark of the Pacific yew trees, Taxol can only be found in very small quantities in these trees and yet the extraction of

Taxol may result in considerable ecological impacts to environment. This led scientists to attempt to produce it by chemical synthesis. After around two decades of investigation, this drug was first successfully synthesized in 1994 by a research group led by Robert A. Holton, although its synthesis was quite expensive and complicated with very minimum yield [1,2]. A few years later, scientists identified 10-deacetylbacatin III, a compound that was remarkably similar to Taxol in structure and can be readily converted to Taxol, in the needles of the European yew evergreen. The benefit of using this compound is that it is easier to extract with little impacts on the environment [3]. Without chemical synthesis, important drugs such as Taxol would not be available at such a low cost today.

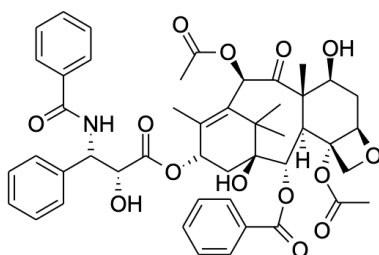


Figure 1. Structure of Taxol (also known as Paclitaxel) [Owner-drawn]

Compared with inorganic syntheses, organic syntheses are usually much harder to visualize and design, due to the large complex structures of many organic compounds. Hence, an organic synthesis usually requires much more thoughts and strategies than an inorganic synthesis. Retrosynthesis is one of the most innovative designing strategies scientists use to achieve a desired product. Retrosynthetic analysis was first coherently brought up by American organic chemist Elias James Corey in 1957, when he began to shape his ideas of backtracking to formulate this methodology. Around that time, the most common way to formulate new synthetic

pathways was to start with simple and readily available substances to produce new compounds with a series of reactions that lead to the desired target molecules. Corey, however, devised this strategy where he started with the compound he wanted to synthesize, and broke it into pieces that are simpler until he reached the starting materials. Before this, other scientists have fragments of insight to retrosynthetic analysis, but none were as complete as Corey's, which gradually became the norm of organic syntheses. Consequently, Corey was awarded the 1990 Nobel Prize for Chemistry for his methodology of retrosynthesis [4].

* Corresponding author. Email: Wanghaochen0519@gmail.com

2. Guiding Principles of Retrosynthesis

A retrosynthetic analysis starts with the target molecule(s) (T.M.) — the final product(s) that are desired — and works backwards to reach molecules that are easier to access. Each step of the retrosynthetic process is represented with a retrosynthetic arrow (different from the arrow used in forward syntheses). The steps in this process chosen may or may not be possible, but plausible reactions can be predicted with existing knowledge. For example, if one wants to synthesize a molecule (target molecule) from a given starting substance, one can devise a retrosynthetic pathway such as where each step brings us closer from to, that can act as a guide for writing a forward synthetic pathway. This method is especially effective because one does not necessarily need to know the details of a forward reaction to be able to envision a retrosynthetic disconnection. For example, if one wants to synthesize an ester from an olefin, instead of thinking of how to produce an ester from an olefin, focus can be put on how to design reactions that can generate the olefin. Even though there are thousands of possibilities in each retrosynthetic step, some solutions are better than others, with only a few being the best. In order to devise the most logical and practical retrosynthesis, chemists look for three main things in each step of the process: Simplification, complexity, and yield. A retrosynthetic step that involves the disconnection of ten bonds simultaneously, although largely simplifying, is extremely unlikely in a practical viewpoint. On the contrary, a step that requires almost no effort but does not simplify the target molecule is essentially pointless.

2.1 Simplification

The first goal of a retrosynthetic step is simplification. In order to turn the usually very complicated target molecule into simple and readily accessible starting materials, each step of the retrosynthetic process should simplify the molecule as much as possible to resemble the starting materials. An example of a step that simplifies the target molecule may be disconnecting a cyclic structure into

linear pieces or disconnecting a functional group that is absent in the starting materials.

2.2 Minimal Complexity

The second goal of a retrosynthetic step is minimal complexity. Although a step can be envisioned where every bond in the target molecule is disconnected to obtain the separate organic, such a step is certainly impossible in a forward synthesis where it would require every bond to be formed in an exact fashion simultaneously. Disconnecting one bond is generally easier than disconnecting two or more bonds. If the forward reaction is considered — a unimolecular process occurs more easily than a bimolecular process because the unimolecular process is easier to arrange geometrically for a successful collision than the bimolecular process. Similarly, a bimolecular process is much more likely to happen than a trimolecular process. This is why most existing elementary steps are unimolecular and bimolecular, and very infrequently does it involve more molecules. Therefore, although breaking more bonds at the same time is much simpler, these processes are less likely to occur practically. In general, one should use one bond disconnection as frequently as possible and should rarely consider simultaneous disconnections of more than two bonds.

2.3 “Retrosynthesis” of a Star

The first two goals can be demonstrated with a non-chemistry example. If the target “molecule” is a star, then some of the possible retrosynthetic pathways are shown in Figure 2. Each of the five pathways shows a possible way to break apart a star. Path A involves cutting the ten corners of the star to produce ten straight pieces. Path B shows taking away the two middle pieces to produce two straight pieces and two remaining linear pieces. Path C involves gluing five pieces surrounding the star to form a square. Path D shows taking away one of the middle pieces to produce one straight piece and the remainder of the star. Finally, path E involves cutting five non-adjacent corners to produce five V shapes.

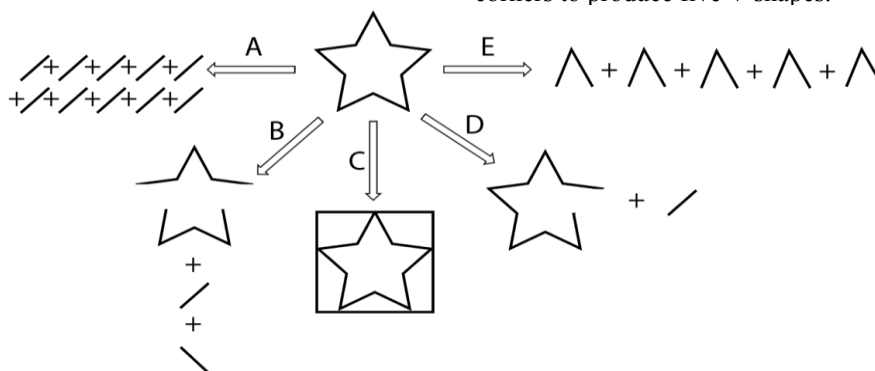


Figure 2. Retrosynthesis of a star [Owner-drawn]

By applying the guiding principles, the most simplifying retrosynthetic step is A, but it's also the most difficult step. In a forward synthesis of the star, ten pieces need to be simultaneously put together, which is much harder to do than the other pathways shown. The next

most simplifying step here is E. It is not as simplifying as step A and V shapes are a bit more complicated to deal with than straight pieces. The complexity of this step is still very high as it involves the interaction of five “molecules”. Step D is the easiest step here, as it only

involves disconnecting one piece from the target molecule. However, notice that much of the star structure is still there, so this step is not very simplifying. Step B is similar to step D, but it involves the disconnection of an extra piece, which results in a more difficult step, but is not as difficult as path A or E. However, this step successfully eliminates a large piece of the central structure of the star, so it is much simpler than step D. Out of these four paths (A, B, D, E), step B achieves both simplification and minimum complexity, so it is most likely the best practical approach.

Path C is the most unique step here, as it is adding more to the target molecule instead of taking pieces away from it. This step is not conventional, but it is certainly plausible. In the forward reaction, one begins with the square and trims the five surrounding pieces away from it. A square is technically much simpler than a star, as it only involves four pieces instead of ten. With suitable technology (something that is much quicker than a scissor, for instance), this step may also be practical. Another example in Figure 3 shows a molecule that is much more complicated. Longifolene contains a tricyclic system, making it a very challenging synthesis task.

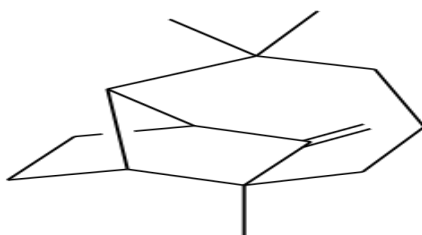


Figure 3. Structure of longifolene [Owner-drawn]

In order to reduce the complexity of this molecule as much as possible, the “common atom” approach can be utilized. Begin by marking all the atoms on all three cycles,

as shown in Figure 4. A simplifying step will most likely involve disconnecting one of the bonds that are between two of the common atoms.

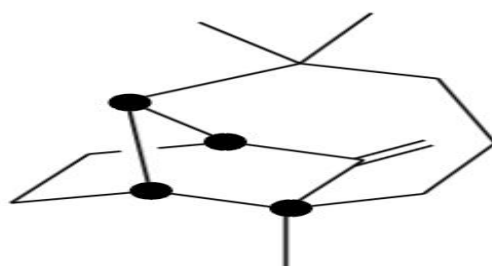


Figure 4. Common atoms of longifolene [Owner-drawn]

In 1964, the total synthesis of longifolene was proposed and published by Corey’s research group. The strategy that was identified in the paper involved breaking the bond between two of the common atoms as shown in

Figure 5. If the bond disconnection didn’t involve any of the common atoms, then the complicated tricyclic system would remain, and the problem is not really simplified [5].

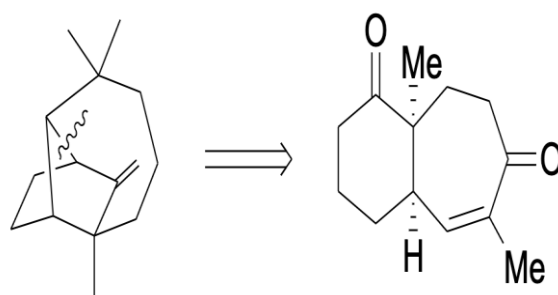


Figure 5. Proposed bond disconnection [Owner-drawn]

2.4 Yield

The third goal of a retrosynthetic analysis is yield. A synthesis process that is successful but only converts 1% of the starting material to the product is generally impractical. It is always beneficial to have a forward synthesis that results in as much product as possible. The closer a process gets to 100% yield, the better it is.

Therefore, the “maximum convergency” strategy can be adopted. This is a divide and conquer strategy where one attempts to divide the complex target molecule into equal and smaller molecules as early as possible, hence reducing the overall number of steps that must be taken. Consider a hypothetical target molecule. One can disconnect the molecule linearly, or convergently. If one disconnects this molecule linearly, the process would be If one disconnects

this molecule convergently, the process would be each step here is very easy to perform, as it only involves the disconnection of one bond. Assume that each step has an 80% yield, the overall yield of the two processes can be compared. In the linear process, there are five steps in total. Therefore, the product obtained at the end is of the starting materials. In the convergent process, the forward reaction scheme may be. Here, in the first two synthesis gives of and each. Therefore, of the starting materials are remaining, more than what can be obtained in the linear process. Notice that this example also demonstrates how simplification is always better, as and are much simpler than.

3. Basic Steps of Retrosynthesis

As shown previously, each step in a retrosynthetic analysis involves one or more disconnection of the bonds

in the structure. One can only disconnect bonds because it is not very practical to slice the atoms themselves because they occupy very limited volumes. Up to this point, it has not been thought about what happens to the molecule when a bond is disconnected. Since each bond is equivalent to a pair of electrons, it is important to think about what happens to those electrons after disconnections.

3.1 Disconnection

As the consequence of a disconnection, one of the three things may happen: both electrons go to the left synthon (A hypothetical fragment of a compound that is a building block for the target molecule), both electrons go to the right synthon, or the electrons are split between the two synthons. The three ways can be illustrated with the compound (Aspirin - a common pain reliever) below (see Figure 6).

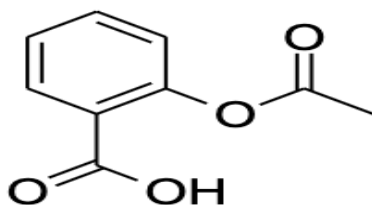


Figure 6. Structure of Acetylsalicylic acid (Aspirin) [Owner-drawn]

If the bond is disconnected as shown in Figure 7, either one of the three retrosynthetic steps shown in Figure 8 must happen.

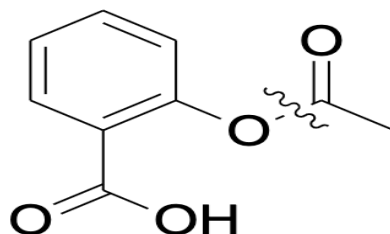


Figure 7. Disconnection of the bond [Owner-drawn]

Note that the curved arrow in the first two scenarios represents the transfer of an electron pair, and the curved “half-arrows” (known as “fish-hooks”) in the last scenario represent the transfer of one electron. In the first retrosynthetic step, both electrons are transferred to the right synthon, so the carbon of the right synthon bears a negative charge; to balance that charge, the left synthon must have a positive charge on the oxygen. In the second step, both electrons are transferred to the left synthon, so the oxygen of the left synthon bears a negative charge, and the carbon of the right synthon has a positive charge. In

the last case, the electron is split between the left and the right synthons, forming two free radicals. Free radicals are generally highly reactive and are as stable as organic ions, so they will not be considered here. If the first two pathways are compared, the second pathway is more likely to happen because the synthons are more stable than the synthons in the first pathway. This is because oxygen is more electronegative than carbon, which makes it more suitable to hold the negative charge; in addition, the carbon carrying the negative charge in the first pathway is overloaded with electrons.

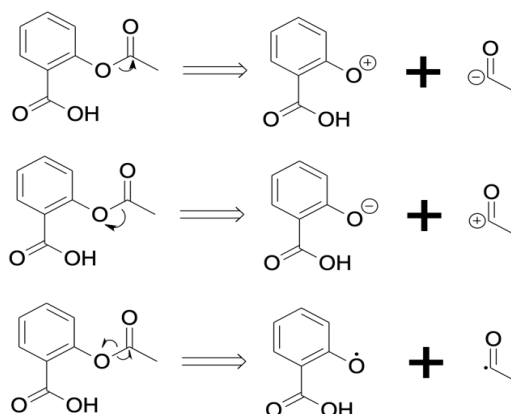


Figure 8. Three possibilities of a bond disconnection [Owner-drawn]

3.2 Synthetic Equivalents

In a practical setting, these idealized synthons are only hypothetical since they are too unstable to exist. Therefore, scientists use synthetic equivalents in place of these synthons, which are much more stable and share the same electrophilic or nucleophilic functions as the hypothetical synthons. These synthetic equivalents are the compounds

that will be used in a scientific laboratory. For the example above, the synthetic equivalents of the two synthons are shown in Figure 9. The negative synthon is replaced with a benzoic acid, which has a hydrogen bond, creating the partial negative charge on the oxygen; the positive synthon is replaced with acetic anhydride (replaced with a carboxylic acid group), which creates a partial positive charge on the carbon due to the polar bond.

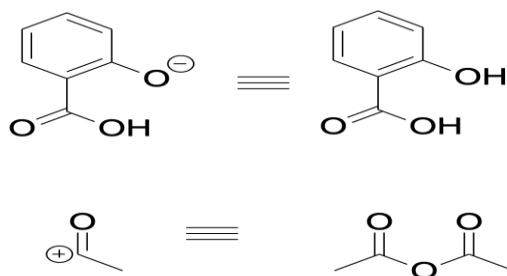


Figure 9. Synthetic equivalents of the synthons [Owner-drawn]

3.3 Strategies

3.3.1 Functional groups

In the process of retrosynthesis, there are lots of choices to disconnect a bond, but only one or two bonds will be broken. This problem is known as chemoselectivity. Chemoselectivity works when there are more than one functional group that can take a reaction with the reagent and only one functional group can be chosen [6]. If there are more than one functional group existing in the target molecule, it should be considered whether the functional group should be broken or just the order of breaking. We

must choose which bond needs to be broken first. Essentially, we should break the bond that minimizes the chemoselectivity problem. Taking ICI-D7114 (potential anti-obesity drug) as an example, there are many positions that we may consider as a bond disconnection. In this case, the bond in position (d) is where the disconnection will most likely to happen. (a) and (b) can lead to great chemoselectivity problems because it is difficult to alkylate the phenol due to the basic nitrogen atom. As to the disconnection at (c), the alkylation of is favorable in the presence of group, so it is better than (d). In the forward synthesis, we want the most reactive compound to be introduced in the last step because it reacts with many reactants.

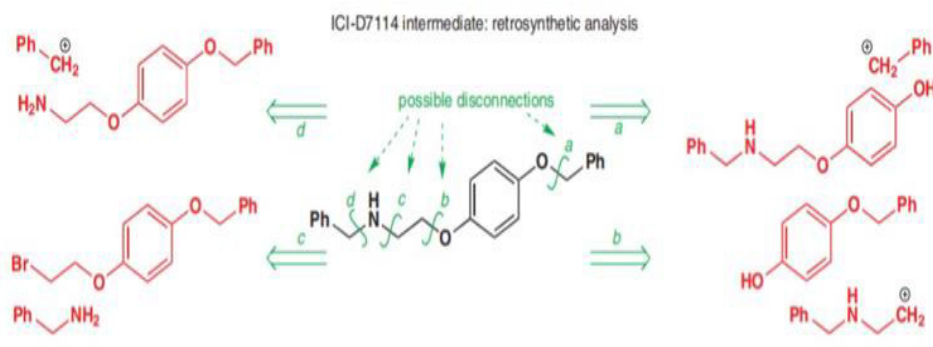


Figure 10. Retrosynthesis of ICI-D7114.

3.4 Functional group interconversion (FGI)

Figure 10 shows the compound was an intermediate in the synthesis of the potential anti-obesity drug ICI-D7114. The problem of chemoselectivity can be avoided by a tool called “Functional Group Interconversion”. In the case of a multistep synthesis, this method is very

beneficial. This technique works by converting one functional group to another. But it has nothing to do with a disconnection because the functional group is only being converted without adding or reducing another functional group. The reason we want to carry out a functional group interconversion is due to chemoselectivity as we have previously discussed. For example, the retrosynthesis of an antihypertensive drug Ofornine is shown below [7].

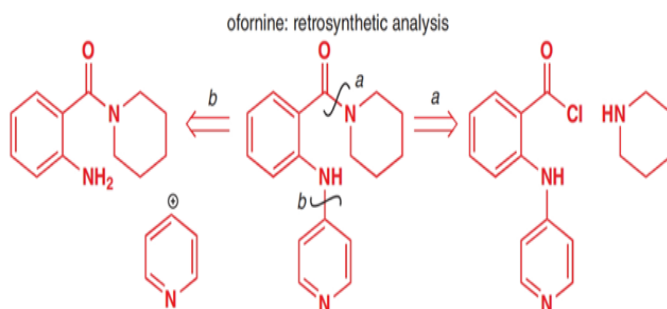


Figure 11. Retrosynthetic analysis of Ofornine

To start a retrosynthesis, we must choose a more favorable path. There the figure 11 describes the retrosynthetic analysis of ofornine. After the disconnection of (a), we need to pay attention to the acrylic chloride group because it poses chemoselectivity problem, therefore the amine group cannot form. Hence, we want to go back to carboxylic acid, which then forms the acyl chloride. Once it is converted to acyl chloride, there is no more chemoselectivity problem, so the amine group can be easily disconnected. Now, we can look at the forward synthesis [8]. It is always important to write the forward synthetic process after the retrosynthesis analysis

to ensure the reactions are reliable. the forward synthesis is shown in Figure 12. The chemoselectivity problems we mentioned have been avoided so the synthesis proceeds properly. First step is the halogenation of the secondary amine, where the electronegative chlorine atom acts as a leaving group and takes the electrons from the C—Cl sigma bond. Meanwhile, one of the N-H bonds is broken and the proton bonds with chlorine to form hydrogen chloride. Finally, the negative charge on nitrogen and positive charge on carbon lead to a bond. After converting to acyl chloride, it can react with the amine to form the target molecule.

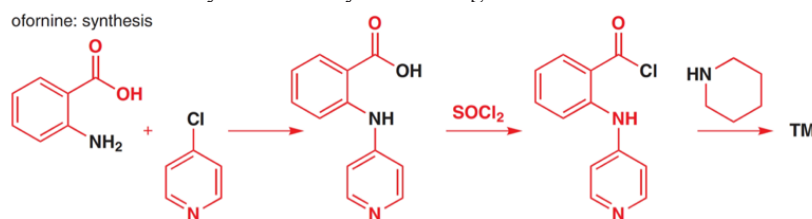


Figure 12. Synthesis of Ofornine

Functional group addition (FGA) is a process of adding a new functional group to the molecule. The simplest example is a hexane (), which can be converted to a hexene () by dehydrogenation with nickel as a catalyst. In this process, an alkene functional group is added to the molecule. On the contrary, a functional group removal (FGR) removes a functional group from the original

molecule. For example, the cyclic ketone can be obtained from , where the halogen (Br) functional group is removed from in this process. However, in the opposite direction, a cyclohexanone can be converted to a bromocyclohexone in the presence of a base and bromine (). Therefore, FGA and FGR are two mutual processes. A two functional group disconnection is generally better than one

functional group disconnection. Consider the functional groups created in the process. Normally many functional groups can be formed after one or two bonds are broken, and these functional groups may have two main options: they can be either divided into two molecules or placed in a single molecule. The two options lead to different results in terms of the yield of reaction and the type of product.

Hence, we need to care about the choice of disconnection that make the synthetic process plausible. Below is an example of a retrosynthesis of an alcohol (see Figure 13), and the mechanism for both primary and secondary alcohol (see Figure 14) [9]. We need to choose a reliable disconnection for the ether group around the oxygen atom, but both sides of the oxygen atom can be made a disconnection. If the disconnection is made at position (a), we can see that there are two hydroxyl groups existing in one molecule simultaneously and an alkene group in another synthon. This is a chemoselectivity problem. Each of the two hydroxyl groups has a probability to experience the alkylation. However, because they are classified as primary and secondary alcohols (joined to two carbon atoms) respectively, they must have different electro-positivity to the positive charge on another synthon. We know that secondary alcohols can form more stable carbocations than

secondary alcohols can do, but since there is no carbocation formed, we should consider the electron donating effect—the alkyl group acts as an electron donating group, which pushes the electron density a little bit to the carbon atom that is attached to the oxygen atom. Therefore, the more electron density that is available to the oxygen atom, the less positive it will be when the positive charge moves to it. To balance the total charge and electrons in the total system, the C-H bond must be broken with two electrons going to oxygen atom as a lone pair. During this process, the electron transfer is not as efficient as the electron transfer of the primary alcohol because there is a smaller demand of electron for the secondary carbon. In addition, the phenyl group attached also poses an effect. Phenyl group has a delocalized electron ring which is formed from combination of p orbitals. Because this part is close to the secondary carbon, it has a huge electron density that pushes towards the oxygen, leading to the low efficiency of electron transfer. These two factors therefore produce a major product and a minor product. If we use the path based on disconnection (b), the two hydroxyl groups are separated. Only one option of alkylation is left to produce the target molecule, which is reliable and avoids any chemoselectivity problem [10].

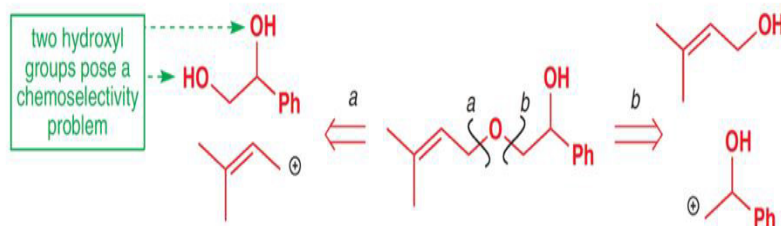


Figure 13. Retrosynthesis with one functional group disconnection or two

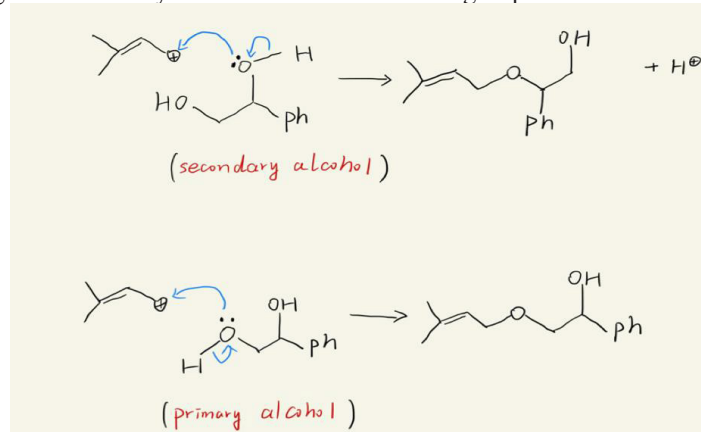


Figure 14. Mechanism of electrophilic addition.[Owner-drawn]

3.5 Electrocyclic Disconnection

Electrocyclic disconnection is the cutting of a cyclic molecule into pieces. In a cyclohexane molecule which contains 6 carbon atoms (C_6H_{12}), we can cut down a piece of two carbons from the molecule (4+2 disconnection) or we can cut the molecule into two identical propane (3+3 disconnection). We can also break one bond instead of two, to form a straight-chain hexane. In any of the processes, the number of carbon atoms remains the same. In a molecule with double bonds like cyclohex-1,3-ene, the

single bond between carbons 5 and 6 can be disconnected by an electron transfer from the double bond to an adjacent single bond. All these are described as cyclic disconnection for simplifying the target molecule.

3.6 Illogical Disconnection

The question around an illogical disconnection is due to wrong distribution of charges. We know the three possible options when we break a C-C sigma bond, but in a more complex molecule, we need to care of the unfavorable

interaction between like charges as well. In Figure 15, there is an 1,3 disconnection to form two synthons. The left synthon corresponds to an alcohol after it is converted

to its resonance form, while the corresponding starting material of the right synthon contains an electronegative halogen.

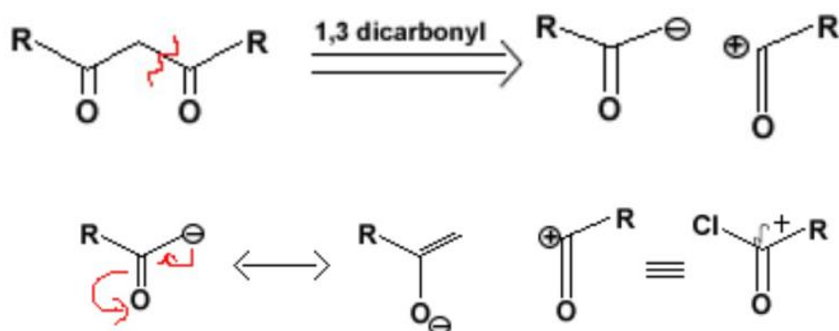


Figure 15. Favorable interaction of two synthons [Owner-drawn]

If the positive and negative charges change positions (see Figure 16) they cannot take the same reaction to be formed. The halogen can only leave a positive charge remaining, but never a negative charge.

On the other hand, the positive charged carbon atom at the end and slightly positive carbon on the double bond have like charges, leading to unfavorable interactions.

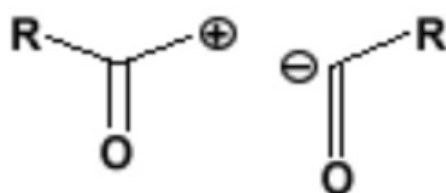


Figure 16. Synthons with swapped charges [Owner-drawn]

4. Conclusion

The intention of this review is to describe the function and usage of retrosynthetic analysis. In general, retrosynthesis has become an important tool for the development of chemistry. For every molecule, retrosynthesis can be used to predict starting materials and reliable reactions. Through retrosynthetic analysis, not only can we artificially synthesize substances that do not exist in nature, but also create more efficient and higher output routes for already existing pathways, which may greatly reduce the cost and time.

Reference

- Holton, Robert A., et al. (1994) First total synthesis of taxol. 1. Functionalization of the B ring. *Journal of the American Chemical Society*, 116.4: 1597-1598.
- Holton, Robert A., et al. (1994) First total synthesis of taxol. 2. Completion of the C and D rings. *Journal of the American Chemical Society*, 116.4: 1599-1600.
- Zocher, Rainer, et al. (1996) Biosynthesis of Taxol: Enzymatic Acetylation of 10-Deacetylbaccatin-III to Baccatin-III in Crude Extracts from Roots of *Taxus baccata*. *Biochemical and biophysical research communications*, 229.1: 16-20.
- Shampo, M.A., Kyle, R.A., and Steensma, D.P. (2013) Elias James Corey—Nobel Prize for Retrosynthetic Analysis. *Stamp Vignette on Medical Science*, 88: e7.
- [5] Corey, E. J., et al. (1964) Total synthesis of longifolene. *Journal of the American Chemical Society*, 86.3: 478-485.
- Clayden, J. and Greeves, N. et al. (2001) Multiple step syntheses: avoid chemoselectivity problems. *Organic Chemistry*. [e-book]. 2nd ed. OX2 6DP: Oxford University Press, Great Clarendon Street, Oxford. pp. 698-699
- Clayden, J. and Greeves, N. et al. (2001) Functional group interconversion. *Organic Chemistry*. [e-book]. 2nd ed. OX2 6DP: Oxford University Press, Great Clarendon Street, Oxford. pp. 699-702
- Clayden, J. and Greeves, N. et al. (2001) Functional group interconversion. *Organic Chemistry*. [e-book]. 2nd ed. OX2 6DP: Oxford University Press, Great Clarendon Street, Oxford. pp. 699-702
- Clayden, J. and Greeves, N. et al. (2001) *Organic Chemistry*. [e-book]. 2nd ed. OX2 6DP: Oxford University Press, Great Clarendon Street, Oxford. pp. 702-706
- Clayden, J. and Greeves, N. et al. (2001) *Organic Chemistry*. [e-book]. 2nd ed. OX2 6DP: Oxford University Press, Great Clarendon Street, Oxford. pp. 703