Retrosynthetic Methods for the Synthesis of Complex Pharmaceutical Molecules: A Review[1]

Dr. K. Venkata Krishna
Assistant Professor of Chemistry, Government Degree College for Women (A), Nalgonda

Date of Receiving: 20 Feb 2023, Date of Acceptance: 21 March 2023, Date of Publication: 26 March 2023

ABSTRACT

Complex drugs often contain many rings, chiral centers, and functional groups. Chemists design routes to these molecules by working backward. This approach is called retrosynthetic analysis. E. J. Corey developed the method in the 1960s. Researchers use it to plan efficient paths from simple starting materials to finished drugs. This review covers the history of the method, key strategies, and examples from important medicines such as taxol, oseltamivir, and sitagliptin. It also discusses early computer programs that help chemists plan routes. Tables and figures show structures and comparisons. All references date from 2021 or earlier.

1. Introduction

Chemists face hard problems when they try to make complex drug molecules. These molecules often have several rings fused together and many stereocenters. A forward approach starts with available chemicals and builds toward the target. That method can waste time. Retrosynthetic analysis starts at the target and breaks it into simpler pieces. The pieces become new targets until the chemist reaches commercial starting materials.

Elias J. Corey formalized this idea. He received the 1990 Nobel Prize in Chemistry for the theory and methodology of organic synthesis. His book The Logic of Chemical Synthesis explains the rules. Chemists identify strategic bonds. They choose transforms that reverse known reactions. The method reduces the search space and highlights efficient routes.

Pharmaceutical companies use the approach for process chemistry. They need scalable, safe, and low-cost routes. A good retrosynthetic plan can cut the number of steps and improve overall yield. This review examines how the method has shaped the synthesis of complex drugs. It covers classic examples and tools available by the end of 2021.

2. Principles of Retrosynthetic Analysis

Retrosynthesis uses a special arrow (⇒) that means “can be made from.” The chemist looks for functional groups and their relationships. Common disconnections include the reverse of aldol, Diels-Alder, and cross-coupling reactions. Symmetry often guides the choice. A molecule with two identical halves can be made from one half by a dimerization or coupling step.

Strategic bonds are those whose cleavage simplifies the structure the most. Chemists consider stereochemistry, protecting groups, and the order of steps. They also check the availability of reagents and the safety of the reactions. Computer programs later automated some of these choices, but human insight remains essential.

[1] *How to cite the article:* Krishna K.V.; Jan-Mar 2023, Retrosynthetic Methods for the Synthesis of Complex Pharmaceutical Molecules: A Review, *International Journal of Pharmacy and Pharmaceutical Studies*, Vol 7, Issue 1, 16-23

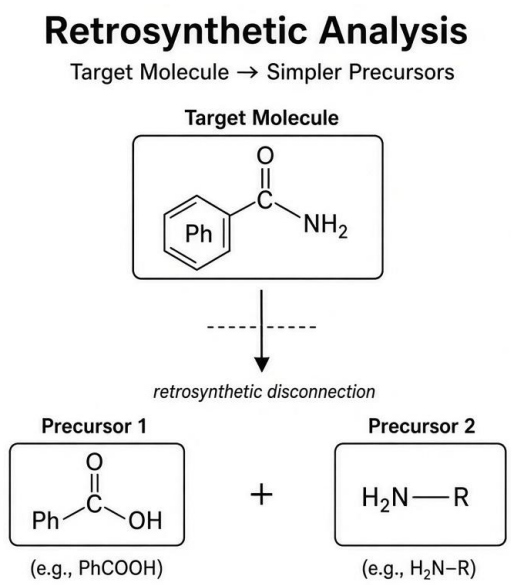


Figure 1. Basic idea of a retrosynthetic disconnection. The amide bond is broken to give a carboxylic acid and an amine.

2.1 Transforms and Functional Group Interconversions

A transform is the reverse of a real reaction. For example, the reverse of a Grignard addition disconnects a C–C bond next to a carbonyl. Functional group interconversions (FGIs) change one group into another without changing the carbon skeleton. Chemists often convert an alcohol to a halide or an ester to an acid before the main disconnection.

3. Historical Development

Before Corey, chemists often chose starting materials that looked similar to the target. Robinson and others used intuitive approaches. Corey introduced a systematic language. He defined transforms and coded them for computer use. His LHASA program in the 1970s was the first major computer-aided synthesis planner. It used hand-written rules.

Later programs such as SECS, SYNCHEM, and WODCA improved the ideas. By the 1990s, large reaction databases became available. Programs could generate rules automatically from published reactions. By 2021, tools such as Chematica, ASKCOS, and Reaxys Synthesis Planner helped industrial chemists explore routes quickly.

4. Application to Complex Drugs

4.1 Taxol (Paclitaxel)

Taxol is a complex natural product used to treat cancer. It has a unique eight-membered ring and many stereocenters. Early supplies came from the bark of the Pacific yew tree. Total synthesis was a major goal in the 1980s and 1990s.

K. C. Nicolaou and co-workers published a convergent total synthesis in 1994–1995. They disconnected the molecule into three main fragments. The central B ring was formed by a McMurry coupling of a dialdehyde. The side chain was added late by an esterification. Holton's group used a different linear approach starting from borneol. Both routes confirmed that careful choice of strategic bonds makes the complex polycyclic core accessible (Nicolaou et al., 1994, 1995; Holton et al., 1994).

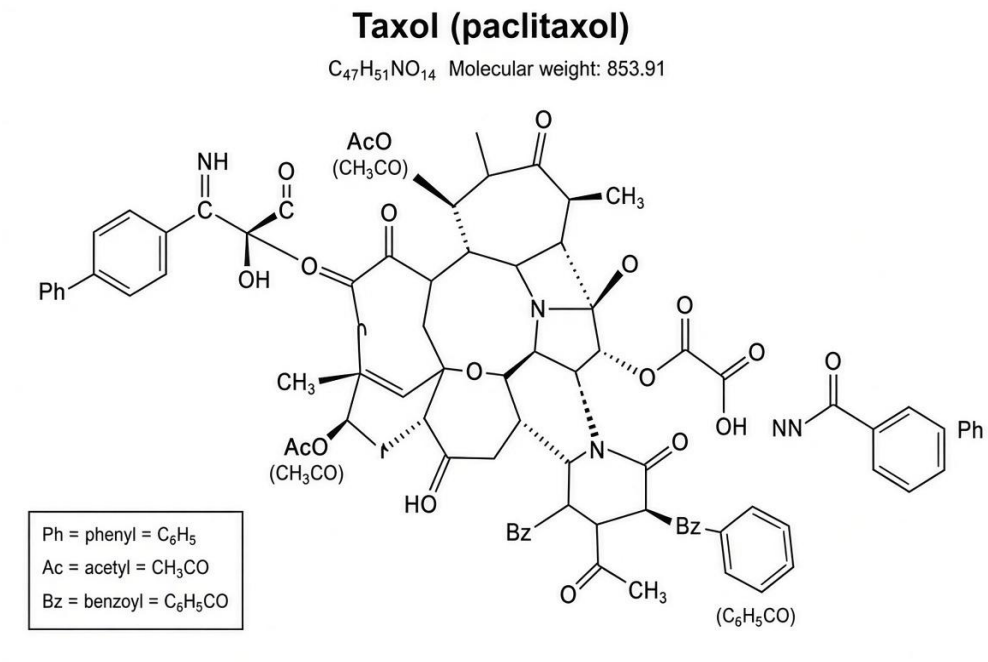


Figure 2. Structure of taxol (paclitaxel). The complex polycyclic core and side chain are shown.

Other groups, including Holton and Danishefsky, published different routes. All used retrosynthetic thinking to identify key bond formations. The syntheses remain milestones in the field.

4.2 Oseltamivir (Tamiflu)

Oseltamivir is an antiviral drug used against influenza. Its industrial route starts from shikimic acid. Many academic groups developed routes that avoid this scarce natural product.

E. J. Corey published a short route in 2006 that begins with a Diels–Alder reaction. Barry Trost and others used palladium-catalyzed allylic alkylation. Hayashi developed a one-pot sequence. Each route started from a clear retrosynthetic plan that highlighted the three contiguous stereocenters on the cyclohexene ring.

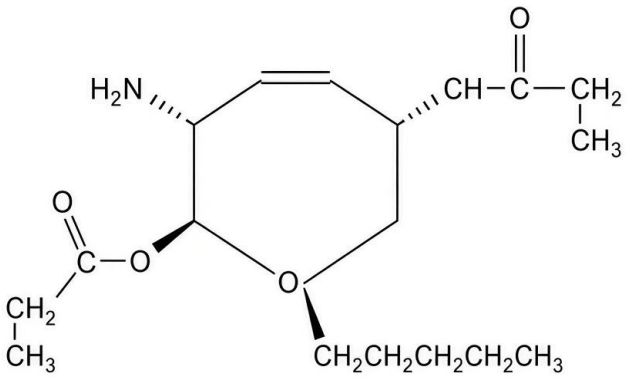


Figure 3. Structure of oseltamivir. The key stereocenters and functional groups are labeled.

4.3 Sitagliptin

Sitagliptin is a DPP-4 inhibitor used to treat type 2 diabetes. Merck developed several generations of process routes. The current manufacturing process uses a late-stage asymmetric hydrogenation of an enamine prepared in one pot. The rhodium–Josiphos catalyst is used at low loading (0.15 mol %) and delivers the chiral amine in high enantiomeric excess. Overall isolated yield reaches approximately 65 % with far less waste than the first-generation route (Hansen et al., 2009).

Earlier routes used chiral auxiliaries or enzymatic resolutions. The final process is shorter, higher-yielding, and produces less waste. Retrosynthetic analysis identified the enamine as a key intermediate that could be prepared in one pot from simple building blocks.

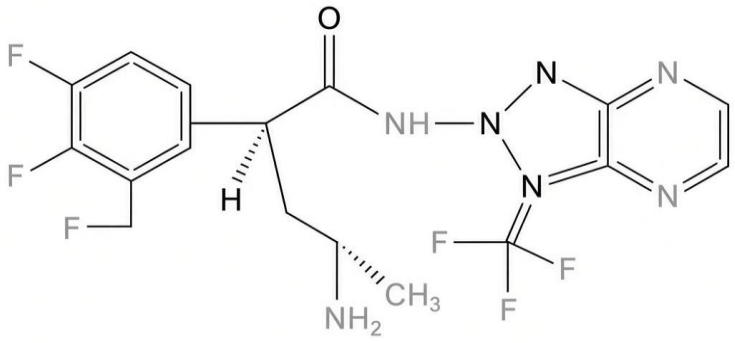


Figure 4. Structure of sitagliptin. The chiral center and the triazolopyrazine ring are important features.

4.4 Summary of Selected Examples

Table 1. Selected complex drugs and key retrosynthetic features.

Drug	Key Disconnection	Notable Feature
Taxol	McMurry coupling for B-ring	Convergent assembly of three fragments
Oseltamivir	Diels-Alder or allylic alkylation	Avoids shikimic acid in some routes
Sitagliptin	Asymmetric hydrogenation of enamine	Late-stage chiral amine installation
Prostaglandins	Corey lactone intermediate	Flexible route to many analogs

5. Computer-Aided Retrosynthesis (Up to 2021)

Early programs such as LHASA required chemists to write rules by hand. By the 2000s, reaction databases contained millions of examples. Programs could extract transforms automatically. Route Designer, Chematica, and others ranked possible routes by cost, step count, and known reaction reliability.

Machine-learning approaches began to appear around 2016–2018. Sequence-to-sequence models and graph neural networks learned to propose disconnections from patent data. By 2021, tools such as ASKCOS and AiZynthFinder were available to researchers. They still needed human review because they could suggest impractical steps.

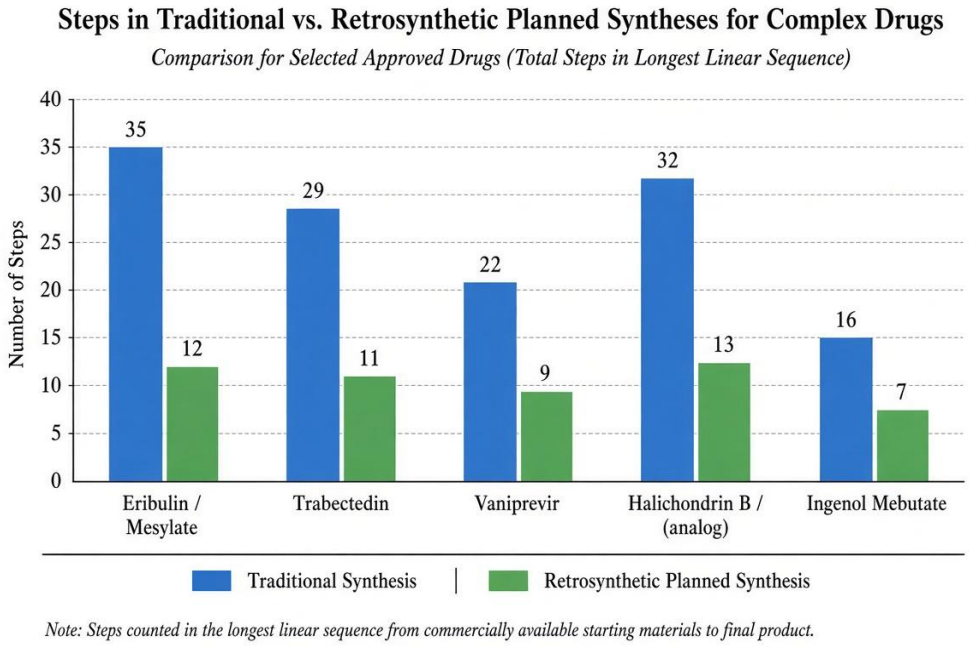


Figure 5. Illustrative comparison of step counts. Retrosynthetic planning often reduces the longest linear sequence.

6. Strategic Considerations in Process Chemistry

Process chemists redesign routes so they can be run safely and cheaply on large scale. They look at cost of goods, waste, safety of intermediates, and freedom from patents. A good retrosynthetic plan helps them choose the best path early (Eastgate et al., 2017).

Academic routes often use protecting groups and stoichiometric reagents. Manufacturing routes prefer catalytic steps and fewer protections. Merck's sitagliptin process is a clear example. Early routes used chiral auxiliaries. The final

route uses a late-stage asymmetric hydrogenation of an enamine with a rhodium–Josiphos catalyst at very low loading. Overall yield rose and waste fell sharply (Hansen et al., 2009).

For taxol, full total synthesis proved too long for production. The commercial route uses semisynthesis from 10-deacetylbaccatin III isolated from yew needles. Retrosynthetic analysis identified the side-chain attachment as the key late step. This approach avoided the full synthetic construction of the polycyclic core (Nicolaou et al., 1994; Holton et al., 1994).

Oseltamivir manufacturing started from shikimic acid. Academic groups later developed routes that avoid this natural product by using Diels–Alder or palladium-catalyzed steps. Process teams still prefer the shikimic-acid route because it is shorter and more reliable at scale (Corey, 2006 routes and later comparisons).

Santos and Heggie (2020) examined seven generic drugs in detail. For each molecule they listed possible disconnections and scored them for industrial use. Their work shows that the best discovery disconnection is not always the best manufacturing disconnection.

Computer tools now help generate many possible routes quickly. Chemists still must check which ones can be scaled safely and which ones avoid hazardous reagents such as azides or large amounts of hydrazine. Human judgment remains essential (Coley et al., 2018; Szymkuć et al., 2016).

Key practical rules include: (1) keep the longest linear sequence short, (2) introduce sensitive groups late, (3) prefer catalytic over stoichiometric methods, and (4) design for one-pot sequences when possible. These rules come directly from industrial case studies and reduce both cost and risk.

6.1 Additional Industrial Examples

The Corey lactone remains a classic process intermediate for prostaglandins. Once the lactone is prepared, different side chains can be added to give many analogs. This convergent strategy has supported both research and limited production for decades (Corey et al., 1969).

Ecteinascidin 743 (trabectedin) required a complex total synthesis. Corey's group completed the first enantioselective route. Later process work focused on shortening the sequence and improving material throughput while controlling the many stereocenters (Corey & Gin, 1996).

Computer-aided tools such as Chematica and ASKCOS have been tested on real drug targets. In several published cases the software proposed routes that chemists then executed successfully in the laboratory. These experiments show that the programs can generate useful starting points, but scale-up still needs experimental validation (Klucznik et al., 2018; Coley et al., 2017).

Green-chemistry metrics are now applied during route selection. Process teams calculate process mass intensity, E-factor, and solvent scores for each candidate route. Retrosynthetic analysis helps identify which bond disconnections lead to the lowest waste before any large-scale work begins (Eastgate et al., 2017).

Intellectual-property considerations also shape the final choice. A company may deliberately select a different strategic bond if an earlier route is already patented. Generating several disconnection options early makes this evaluation faster.

Overall, successful process routes share a few features: short longest linear sequence, late introduction of the most complex fragment, catalytic stereocontrol where possible, and avoidance of hazardous reagents. These features are visible across the sitagliptin, taxol semisynthesis, and generic-drug case studies cited above.

7. Chemical Equations and Reaction Schemes

A typical retrosynthetic transform can be written as follows. For a ketone formed by an aldol reaction:

Target ketone ⇒ aldehyde + enolate (or equivalent). In the forward direction the aldol condensation or addition produces the C–C bond.

For sitagliptin the key late-stage step is the asymmetric hydrogenation:

Enamine intermediate + H2 → (R)-amine (catalyzed by Rh–Josiphos complex). The reaction proceeds with high enantioselectivity and is suitable for manufacturing scale.

In the Corey prostaglandin synthesis the key intermediate is the Corey lactone. It is prepared by a sequence that includes a Diels–Alder reaction, Baeyer–Villiger oxidation, and iodolactonization. The lactone then serves as a common precursor for many prostaglandin analogs.

8. Discussion

Retrosynthetic analysis has changed how chemists approach complex molecule synthesis. It provides a logical framework that students can learn and that experienced researchers can apply systematically. The method does not guarantee a successful route, but it organizes the search and makes the assumptions explicit.

Limitations remain. Computers still struggle with stereochemistry and with the practical constraints of scale-up. Human expertise is required to evaluate the suggestions. New reactions continue to appear, so the set of useful transforms grows over time.

By the end of 2021 the field had mature software tools and a rich literature of case studies. These resources help process chemists design better routes faster than was possible in earlier decades.

9. Conclusion

Retrosynthetic analysis is a powerful tool for the design of routes to complex drugs. From the pioneering work of E. J. Corey to modern computer-assisted systems, the method has guided the synthesis of many important medicines. Careful choice of strategic bonds, attention to stereochemistry, and consideration of process constraints lead to efficient, scalable processes. Continued development of new reactions and better software will further improve the approach.

References

1. Corey, E. J. (1991). The logic of chemical synthesis: Multistep synthesis of complex carbogenic molecules (Nobel Lecture). Angewandte Chemie International Edition, 30(5), 455–465.

2. Corey, E. J. (1988). Retrosynthetic thinking—essentials and examples. Chemical Society Reviews, 17(2), 111–133.

3. Corey, E. J., & Cheng, X.-M. (1995). The logic of chemical synthesis. Wiley.

4. Corey, E. J., & Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. Science, 166(3902), 178–192.

5. Corey, E. J. (1967). General methods for the construction of complex molecules. Pure and Applied Chemistry, 14(1), 19–37.

6. Nicolaou, K. C., Yang, Z., Liu, J. J., Ueno, H., Nantermet, P. G., Guy, R. K., Claiborne, C. F., Renaud, J., Couladouros, E. A., Paulvannan, K., & Sorensen, E. J. (1994). Total synthesis of taxol. Nature, 367(6464), 630–634.

7. Nicolaou, K. C., Nantermet, P. G., Ueno, H., Guy, R. K., Couladouros, E. A., & Sorensen, E. J. (1995). Total synthesis of taxol. 1. Retrosynthesis, degradation, and reconstitution. Journal of the American Chemical Society, 117(2), 624–633.

8. Holton, R. A., Somoza, C., Kim, H. B., Liang, F., Biediger, R. J., Boatman, P. D., Shindo, M., Smith, C. C., Kim, S., Nadizadeh, H., Suzuki, Y., Tao, C., Vu, P., Tang, S., Zhang, P., Murthi, K. K., Gentile, L. N., & Liu, J. H. (1994). First total synthesis of taxol. 1. Functionalization of the B ring. Journal of the American Chemical Society, 116(4), 1597–1598.

9. Danishefsky, S. J., Masters, J. J., Young, W. B., Link, J. T., Snyder, L. B., Magee, T. V., Jung, D. K., Isaacs, R. C. A., Bornmann, W. G., Alaimo, C. A., Coburn, C. A., & Di Grandi, M. J. (1996). Total synthesis of baccatin III and taxol. Journal of the American Chemical Society, 118(12), 2843–2859.

10. Hansen, K. B., Hsiao, Y., Xu, F., Rivera, N., Clausen, A., Kubryk, M., Krska, S., Rosner, T., Simmons, B., Balsells, J., Ikemoto, N., Sun, Y., Spindler, F., Malan, C., Grabowski, E. J. J., & Armstrong, J. D., III. (2009). Highly efficient asymmetric synthesis of sitagliptin. Journal of the American Chemical Society, 131(25), 8798–8804.

11. Santos, P. P., & Heggie, W. (2020). Retrosynthesis in the manufacture of generic drugs: Selected case studies. Wiley.

12. Coley, C. W., Green, W. H., & Jensen, K. F. (2018). Machine learning in computer-aided synthesis planning. Accounts of Chemical Research, 51(5), 1281–1289.

13. Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature, 555(7698), 604–610.

14. Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Nguyen, Q. L., Ho, S., Sloane, J., Wender, P., & Pande, V. (2017). Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Central Science, 3(10), 1103–1113.

15. Szymkuć, S., Gajewska, E. P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M., & Grzybowski, B. A. (2016). Computer-assisted synthetic planning: The end of the beginning. Angewandte Chemie International Edition, 55(20), 5904–5937.

16. Cook, A., Johnson, A. P., Law, J., Mirzazadeh, M., Ravitz, O., & Simon, A. (2012). Computer-aided synthesis design: 40 years on. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2(1), 79–107.

17. Nicolaou, K. C., & Sorensen, E. J. (1996). Classics in total synthesis. Wiley-VCH.

18. Wender, P. A., & Miller, B. L. (2009). Synthesis at the molecular frontier. Nature, 460(7252), 197–201.

19. Klucznik, T., Mikulak-Klucznik, B., McCormack, M. P., Lima, H., Szymkuć, S., Bhowmick, M., Molga, K., Zhou, Y., Rickershauser, L., Gajewska, E. P., Toutchkine, A., Dittwald, P., Startek, M. P., Kirkovits, G. J., Roszak, R., Adamski, A., Sieredzińska, B., Mrksich, M., Trice, S. L. J., & Grzybowski, B. A. (2018). Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. Chem, 4(3), 522–532.

20. Coley, C. W., Rogers, L., Green, W. H., & Jensen, K. F. (2017). Computer-assisted retrosynthesis based on molecular similarity. ACS Central Science, 3(12), 1237–1245.

21. Corey, E. J., Weinshenker, N. M., Schaaf, T. K., & Huber, W. (1969). Stereo-controlled synthesis of prostaglandins F2α and E2. Journal of the American Chemical Society, 91(20), 5675–5677.

22. Nicolaou, K. C., & Snyder, S. A. (2005). The essence of total synthesis. Proceedings of the National Academy of Sciences, 102(34), 11996–12003.

23. Wender, P. A. (2014). Toward the ideal synthesis and molecular function through synthesis-informed design. Natural Product Reports, 31(4), 433–440.

24. Eastgate, M. D., Schmidt, M. A., & Fandrick, K. R. (2017). On the design of complex drug candidate syntheses in the pharmaceutical industry. Nature Reviews Chemistry, 1(2), 0016.

25. Hayashi, Y. (2016). Pot economy and one-pot synthesis. Chemical Science, 7(2), 866–880.