Journal of High School Science

Reaction simulation for the efficient synthesis of Active Pharmaceutical Ingredients

Jain P

Submitted: July 9, 2023, Revised: version 1, August 19, 2023
Accepted: September 22, 2023

Abstract

Reaction simulation utilizing computational methods can save reactant material and reaction time, decreasing the cost and time necessary for the mass manufacture of active pharmaceutical ingredients (APIs). To simulate the synthesis of APIs, specifically Metoprolol, a route plan and reaction model was designed, and a reaction optimization was performed. To create a reaction model, kinetic models and reaction simulations were used to calculate the yield at various conditions and plotted on a three-dimensional graph. Machine learning algorithms were used to determine the optimal parameters that would maximize the yield of Metoprolol. Reaction simulation can also be applied to other APIs to improve the design and synthesis process, and the utilization of this specific methodology can save time and costs in both industrial and academic applications.

Keywords

Synthesis, Reaction modeling, Route planning, Kinetic model, Reaction simulation, Machine learning, Yield maximization, Active Pharmaceutical Ingredient, Arrhenius equation, Activation energy

Prisha Jain, Irvington High School, 41800 Blacow Rd, Fremont, CA, USA.
prisha.noshini.jain@gmail.com

Introduction

The synthesis of drug molecules requires many specific and potentially complex steps. Designing a route plan to the product can be difficult because different transformations in the synthesis have different optimal parameters. This is a crucial process as efficient steps can lower production costs and improve the purity and yield of the final product. Determining the optimal synthetic steps enables the mass production of APIs, which is necessary for drug manufacturing (1). The success of this synthetic process requires optimization of each synthetic step to ensure high yields and purity of the product. Reaction modeling and optimization are one way to ensure optimal process parameters.

Reaction modeling is essential to understanding reactions by determining how changing an experiment's parameters affects its yield (2). Reactions can be simulated using kinetic modeling, which utilizes computational methods and rate laws to predict the yield of chemical reactions at different reaction conditions. This differs from other simulation methods, such as statistical methods, as it is based on chemical relationships and physio-chemical information instead of mathematical relationships between data values. Kinetic modeling can therefore simulate reactions quickly and efficiently outside of the lab and can even simulate reactions with factors outside experimental constraints which is not possible using statistical methods (3,4). This saves lab materials and time as fewer experiments are necessary to run, which also reduces costs for labs and pharmaceutical companies.

Optimizing the synthetic steps in the formulation of a drug ensures high yields and purity, which are essential to reduce the cost of drugs (5). Reaction optimization is possible by employing various techniques, including but not limited to one factor at a time (OFAT) and design of experiments (DoE). OFAT optimization iteratively performs experiments by controlling all variables except for the one being optimized (7). However, DoE determines the yield of a chemical reaction based on all the factors for that reaction simultaneously. DoE is more efficient than OFAT optimization since it creates predictions considering all factors allowing for better reaction prediction (6,8). The empirical models can be fitted to data gathered from experimentation to understand how changing multiple variables affect reaction output. These optimization techniques save time and money in the lab (7). Additionally, optimal conditions are usually determined by extensive experimentation in the lab, but by computationally optimizing reactions, lab materials and time are saved, reducing product costs.

This paper aims to optimize the final step in the synthesis of Metoprolol, a beta blocker drug manufactured by AstraZeneca (2), by determining the reaction conditions necessary to deliver the highest product yield. Additionally, a route plan will be created to demonstrate a viable route to Metoprolol from simple starting materials. To optimize the reaction, MATLAB will first be used to simulate the reaction at various temperatures, reaction times, and initial reactant concentrations. This model will then be used in conjunction with a machine learning algorithm to determine the optimal parameters to achieve the highest product yield. This research will

determine what conditions will optimize the chemical synthesis of Metoprolol in order to increase its yield. This could lower manufacturing costs for pharmaceutical companies and decrease the price of essential medication for patients.

Route Plan

The route plan for Metoprolol represents a series of possible synthetic steps to synthesize Metoprolol starting from nitrobenzene (Figure 1). Metoprolol is a beta blocker drug manufactured by AstraZeneca (to treat hypertension and angina. Determining a viable route is essential to optimize the reaction, as each transformation can be optimized to maximize yield. However, this paper will focus on the final transformation for which we have kinetic data. Simulating the reaction will reduce the time spent in the lab and money spent on reactants, as fewer experiments will be necessary. Additionally, machine learning techniques alongside reaction modeling help chemists determine the best yield possible without being physically in the lab.

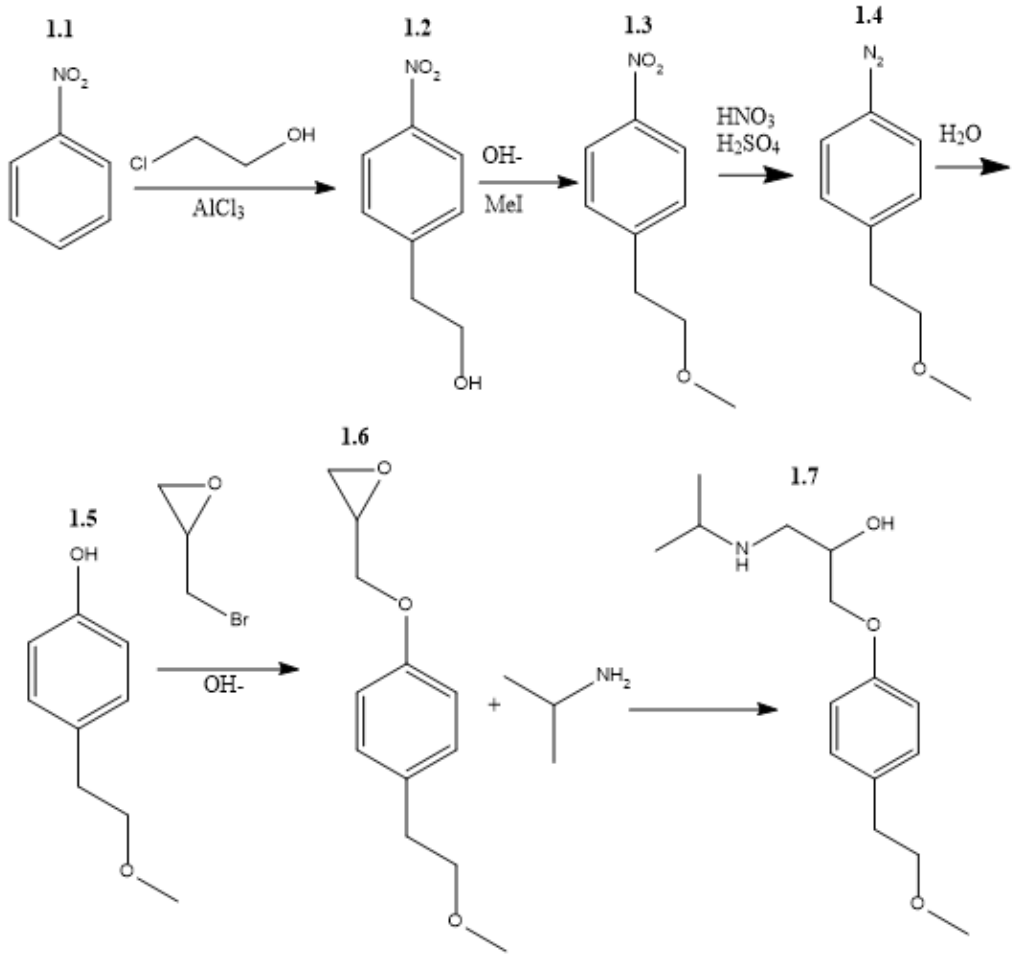

Figure 1. A route plan to Metoprolol from simple, cheap starting materials.

The pricing on each transformation was determined using Sigma Aldrich and for the reaction 0.52 mol of each molecule was used to produce 100g of Metoprolol. The initial cost of the nitrobenzene starting material is $29.90 per mol. The first step of the route plan uses Friedel crafts alkylation to produce 1.2 from nitrobenzene, which would cost $134.00 (9). To make 1.3 from 1.2, the Williamson Ether method is used, and the reaction costs $104.76 (10). Making 1.4 from 1.3 uses diazotization and costs $39.46 (11). Next, to synthesize 1.5 starting from 1.4, hydration is used, and the reaction costs $0.58 (12). To make 1.6 from 1.5, nucleophilic substitution is used, and it costs $545.93 (13). The last transformation results in the synthesis of Metoprolol and costs $2.90 (14). The total price to synthesize 100g Metoprolol using the proposed route plan is $446.77, assuming a 100% yield for each transformation.

Reaction Simulation and Modeling

To simulate the final reaction (Figure 2) in the route plan described previously yielding Metoprolol, a range of initial conditions are inputted into MATLAB, which uses ordinary differential equations (ODEs) to calculate the yield of the product. These ODEs are determined by rate laws and take the initial concentration of the reactants as parameters to determine the final concentrations of all the compounds in the equations. The ODEs that represent the change in concentration are highlighted below in equations 1-4. The final concentration of Metoprolol is then converted into a percentage based on the initial concentration of the reactants and displayed on a three-dimensional graph.

$$Equation\,1: \frac{d[1.1]}{dt} = K(1) * [1.1] * [1.2] - K(2) * [1.1] * [1.3]$$

$$Equation\,2: \frac{d[1.2]}{dt} = -K(1) * [1.1] * [1.2]$$

$$Equation\,3: \frac{d[1.3]}{dt} = K(1) * [1.1] * [1.2] - K(2) * [1.1] * [1.3]$$

$$Equation\,4: \frac{d[1.4]}{dt} = K(2) * [1.1] * [1.3]$$

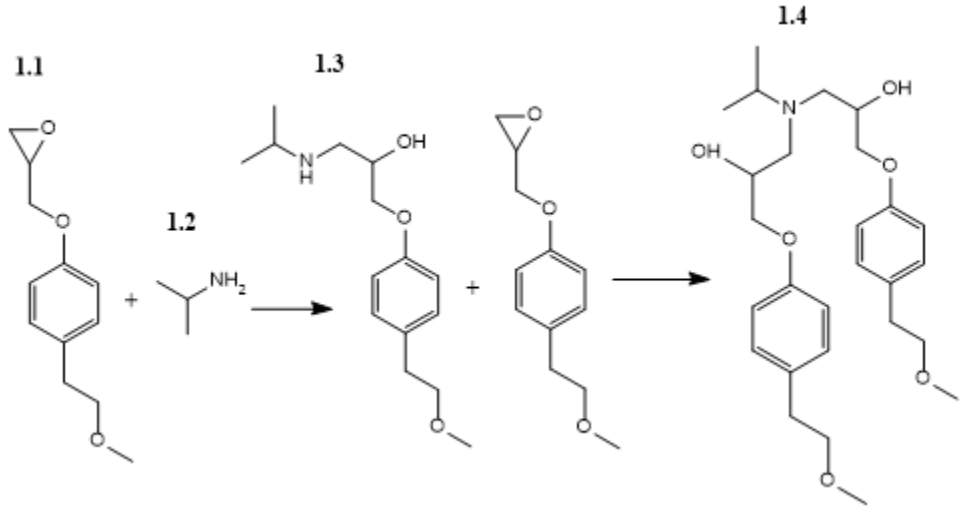

Figure 2. The final transformation for the synthesis of Metoprolol, as well as the undesired side-product 1.4

The epoxide starting material, 1.1, has a starting concentration varying from one to five molars. Each permutation of this variable is individually plotted on the graph to represent how changing this condition affects the reaction yield. The other reactant, isopropylamine (molecule 1.2), has a constant initial yield of four molars to demonstrate how the yield of Metoprolol changes as the reactant equivalents are changed. The varying concentrations of molecule 1.1 are represented by the y-axis on Figure 3 while the concentration of molecule 1.2 is held constant. The temperature and reaction time varies from 380 to 470 degrees Kelvin and 4 to 22 minutes, respectively. These values are used to determine the K-value of the reaction using the re-parameterized Arrhenius equation as show in Equation 5. Then the K-value is used in the ordinary differential equations as part of the rate law. Each time the temperature or reaction time changes, a new K-value is calculated, and the corresponding yield of Metoprolol is determined. The activation energy for the last step in the reaction is 75,000 kJ/mol (14) . The final reaction in this route plan is not reversible so ODEs could be used to determine yield without considering equilibrium constants. The yield is then displayed on a three-dimensional graph with the parameters representing the location of the dot and the color representing the yield (Figure 3). Reaction 1 represents the formation of Metoprolol, molecule 1.3, from molecules 1.1 and 1.2 and Reaction 2 represents the formation of the side product, molecule 1.4. For each temperature the reaction has been simulated it, the rate constant K was calculated (Table 1).

$$Equation\,5: k = k_{ref} * e^{\left[\frac{-E_a}{R}\left(\frac{1}{T}-\frac{1}{T_{ref}}\right)\right]}$$

Table 1. The rate constant for each temperature simulated

		Temperature (K)									
Reaction		383.15	393.15	403.15	413.15	423.15	433.15	443.15	453.15	463.15	473.15
	1	0.013	0.024	0.041	0.069	0.113	0.182	0.286	0.441	0.668	0.994
	2	0.001	0.001	0.003	0.004	0.007	0.012	0.019	0.030	0.046	0.069

The yield experiences a steep decline once the molarity of molecule 1.1 exceeds four because the molarity of molecule 1.2 is held constant at four. If there is an excess of molecule 1.1, it will react further with Metoprolol to produce molecule 1.4, an unintended side product (see Figure 2 for a diagram of the side reaction). Limiting the excess of molecule 1.1 prevents this further reaction and therefore increases the percentage of Metoprolol present in the final product, increasing its yield. Additionally, the yield increases as the reaction time increases since more reaction time increases the number of collisions, which results in the formation of more product. However, the temperature is not optimal at either extreme as the highest yield tends to be when the temperature is in the middle of both bounds, around 430 degrees Kelvin. For this reaction, high temperatures decrease reactivity making it necessary to create reaction models to determine the ideal temperature for the synthesis of Metoprolol.

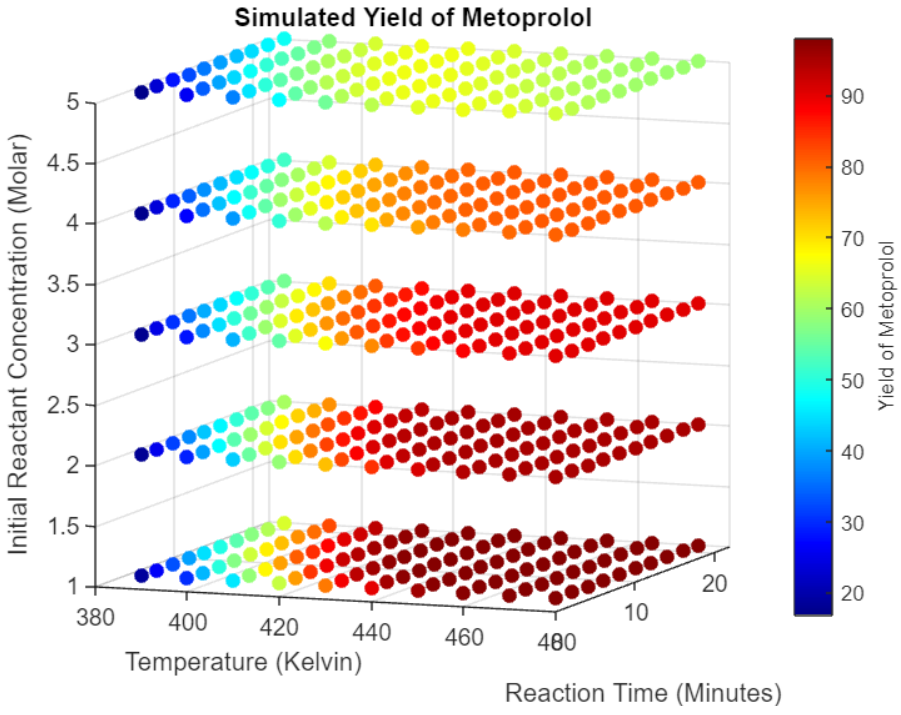

Figure 3. Simulated yield of Metoprolol

The ability to change the kinetic parameters and determine hundreds of data points in just a few minutes makes it cheaper to simulate reactions rather than run them in a lab, as no lab materials are necessary. Additionally, simulating reactions is faster than experimentation as there is no wait time for the reaction to go to completion or set up time necessary. The data points collected from MATLAB can also be plotted as a two or three dimensional figure depending on the number of kinetic parameters making it easy for chemists to visualize the data and determine which conditions lead to the most favorable output.

Reaction Optimization

Machine learning involves using algorithms and models that can analyze and learn from data, allowing them to recognize patterns, make predictions, and improve their performance over time. It can optimize black-box problems where the algorithm does not understand the underlying chemical processes within the optimization procedure. This is why it is sometimes necessary to input specific upper and lower bounds so that the optimization is reasonable and practical. These optimization algorithms can be applied to chemistry and specifically reaction optimization to maximize the yield of chemical reactions.

The genetic algorithm was used to determine the optimal reaction parameters for which the yield of Metoprolol is the highest. Similar to the reaction modeling section, ODEs were used to calculate the yield for the inputted parameters selected by the genetic algorithm between the lower and upper bounds. These bounds were the same as the ranges for the kinetic simulation. The genetic algorithm (was specifically chosen because other optimization functions available on MATLAB, such as fminsearch which uses Simplex, result in optimal parameters outside of the desired bounds, which cannot be conducted practically and safely. The simple genetic algorithm works in 6 stages: initialization, evaluation, selection, crossover, mutation, and replacement. This process repeats until the optimal parameters have been found which is usually when the evaluation is not higher in the succeeding round. In the selection stage, the highest-performing parameters from the evaluation stage are used to generate a new set of parameters. Then in the crossover stage, each variable has a 50% chance of existing in the new set of parameters. For the mutation stage, there is a probability of a random variable mutating. These steps represent one cycle of the genetic algorithm which continues repeating until the optimum set of parameters has been reached. For a detailed description of genetic algorithm and its other associated applications, refer to work by Taylor et al (14).

Optimizing a reaction using machine learning is much more efficient than determining the yield for every set of conditions possible and results in more precise optimal condition identification. On a larger scale, machine learning optimization can determine the maximum yield with bounds that allow millions of possible combinations. Running an experiment for each of these conditions or even simulating it could take a long time, but algorithms such as the genetic algorithm are able to do the same process in just minutes. Computationally optimizing reactions is a more effective way to determine the optimal parameters of a reaction compared to testing each set of conditions experimentally.

Conclusion

The route plan proposed was a series of possible transformations to synthesize Metoprolol from nitrobenzene. Using the proposed route plan, it costs $446.77 to synthesize 100g of Metoprolol. Creating this route plan helps to determine a series of possible synthetic transformations that could be optimized to maximize the yield of Metoprolol.

To simulate the final transformation of Metoprolol, kinetic models and reaction simulation were used, and they proved to be more efficient and cheaper than actual experimentation. Reaction models determined the yield by utilizing ordinary differential equations and plotted it on a scatter plot to show how different kinetic parameters affect the yield of Metoprolol. This helps chemists understand and visualize the reaction, particularly how the value of each factor affects the yield.

Machine learning algorithms were then used in tandem with reaction simulation to increase the efficiency and precision of the outputs and determine the optimal parameters that maximize the yield of Metoprolol. This methodology saves the time it would have taken to simulate hundreds of reactions and will also save the reactant material and chemist time that would have been necessary to run these reactions in the lab. Optimization using computational methods achieves the same goals as experimental optimization at the fraction of the cost and time; therefore, reaction simulation in this manner can positively affect the design and synthesis of other active pharmaceutical ingredients.

References

1. Campos, K.R., Coleman, P.J., Alvarez, J.C., Dreher, S.D., Garbaccio, R.M., Terrett, N.K., et al., *The Importance of Synthetic Chemistry in the Pharmaceutical Industry*. Drug Discovery Today, 2019. **24**(11): p. 2213-2220. https://doi.org/10.1126/science.aat0805

2. Taylor, C.J., Pomberger, A., Felton, K.C., Grainger, R., Barecka, M., Chamberlain, T.W., et al., *A Brief Introduction to Chemical Reaction Optimization*. Chemical Reviews. 2023. **123**(6): p. 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798

3. Zaborenko, N., Bedore, M.W., Jamison, T.F., Jensen, K.F., *Kinetic and scale-up investigations of epoxide aminolysis in microreactors at high temperatures and pressures.* Organic Process Research & Development, 2011. **15**(1): p. 131-139. http://doi.org/10.1021/op100252m

4. Tufvesson, P., Lima-Ramos, J., Al Haque, N., Gernaey, K.V., Woodley, J.M., *Advances in the process development of biocatalytic processes*. Organic Process Research & Development, 2013. **17**(10): p. 1233-1238. https://doi.org/10.1021/op4001675

5. Petrikovics, I., Jayanna, P., Childress, J., Budai, M., Martin, S., Kuzmitcheva, G., et al., *Optimization of Liposomal Lipid Composition for a New, Reactive Sulfur Donor, and In Vivo*

Efficacy Studies on Mice to Antagonize Cyanide Intoxication. Journal of Drug Delivery, 2011. P.1-13. https://doi.org/10.1155/2011/928626

6. Aggarwal, V.K., Staubitz, Owen M., *Optimization of the mizoroki− heck reaction using design of experiment (DoE)*. Organic Process Research & Development, 2006. **10**(1): p.64-69. https://doi.org/10.1021/op058013q

7. Lendrem, D.W., Lendrem, B.C, Woods, D., Rowland-Jones, R., Burke, M., Chatfield, M., et al., *Lost in space: design of experiments and scientific exploration in a Hogarth universe*. Drug Discovery Today, 2015. **20**(11): p. 1365-1371. https://doi.org/10.1016/j.drudis.2015.09.015

8. Clayton, A., Manson, J.A, Taylor, C.J., Chamberlain, T.W., Taylor, B.A, Clemens, G., et al., *Algorithms for the Self-Optimization of Chemical Reactions*. Reaction Chemistry and Engineering, 2019. **4**: p. 1545-1554. https://doi.org/10.1039/C9RE00209J

9. Rueping, M, Nachtsheim, B.J., *A Review of New Developments in the Friedel-Crafts Alkylation – From Green Chemistry to Asymmetric Catalysis*. Beilstein Journal of Organic Chemistry, 2010, **6**(6). https://doi.org/10.3762/bjoc.6.6

10. Diamanti, A., Ganase, Z., Grant, E., Armstrong, A., Piccione, P.M., Rea, A.M., et al., *Mechanism, Kinetics and Selectivity of a Williamson Ether Synthesis: Elucidation Under Different Reaction Conditions*. Reaction Chemistry and Engineering, 2021, **8**: p. 1195-1211. https://doi.org/10.1039/D0RE00437E

11. Placidi, M.P., Villaraza, A.J, Natrajan, L.S., Sykes, D., Kenwright, A.M., Faulkner, S., *Synthesis and Spectroscopic Studies on Azo-Dye Derivatives of Polymetallic Lanthanide Complexes: Using Diazotization to Link Metal Complexes Together*. Journal of the American Chemical Society, 2009, **131**(29): p. 9916-9917. https://doi.org/10.1021/ja904362f

12. Chanda, A., Fokin, V.V., *Organic Synthesis "On Water"*. Journal of the American Chemical Society, 2009, **109**(2): p. 725-784. https://doi.org/10.1021/cr800448q

13. Li, J., Beak, P., *Endocyclic Restriction Test: Experimental Evaluation of the Transition Structure Geometry of a Nucleophilic Substitution at Trivalent Phosphorus*. Journal of the American Chemical Society, 1992, **114**(23): p. 9206-9207. https://doi.org/10.1021/ja00049a073

14. Taylor, C., Seki, H., Dannheim, F.M., Willis, M.J., Clemens, G., Taylor, B.A., *An Automated, Computational Approaches to Kinetic Model and Parameter Determination*, Reaction Chemistry and Engineering, **6**(8): p. 1404-1411. https://doi.org/10.1039/d1re00098e