

ОРГАНИЧЕСКИЙ СИНТЕЗ: ХИМИК-КОМПЬЮТЕР

Белорусский государственный университет информатики и радиоэлектроники,
Минск, Республика Беларусь

Кисель Д. С., Дервянко Э. Г.

Позняк А. А. – канд. физ.-мат. наук, доцент

Представлена краткая историческая справка о развитии программ, помогающих химику-органику при синтезе сложных молекул, а также рассмотрена современная программа *Chematica*, предназначенная для вычислительного планирования оптимальных путей синтеза органических соединений.

Введение

В выцветших фотографиях 1960-х годов лаборатории органической химии выглядят как рай алхимика. Бутылочки различных реагентов стоят на полках; стеклянная посуда цветет на деревянных стеллажах; и ученые, которые деловито синтезируют различные молекулы. В 2018 году лаборатории органической химии могут похвастаться стеной из вытяжных шкафов, сложными аналитическими приборами и еще большим разнообразием реагентов. Но суть того, что делают ученые, все та же. Органические химики обычно планируют свою работу на бумаге, набрасывая шестиугольники и углеродные цепи на странице за страницей, поскольку они думают о последовательности реакций, которые им нужно будет сделать для данной молекулы. Затем они стараются следовать этой последовательности вручную – тщательно перемешивая, фильтруя и дистиллируя, сшивая молекулы так, будто они вышивают платье от кутюр.

Но одна растущая группа химиков теперь пытается освободиться от этой ремесленной работы, создав программу, способная автоматически изготовить любую органическую молекулу (рис. 1).

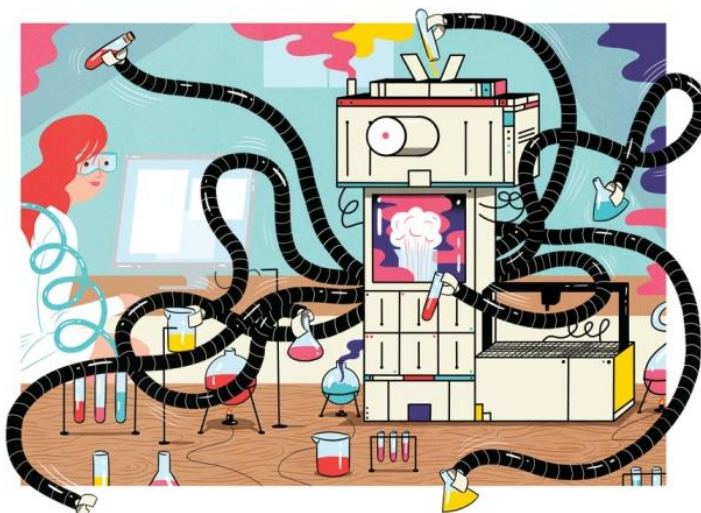


Рис. 1. – Химик-машина, Иллюстрация Райана Сноука [1]

Думает, как химик. Действует, как машина

Химики-органики, планирующие сложный синтез, как правило, используют метод, называемый ретросинтетическим анализом. Они рисуют законченную молекулу, затем выделяют ее, стирают связи, которые легко формировать и оставляют фрагменты молекулы, которые являются стабильными или легкодоступными в плане синтеза. Это позволяет им идентифицировать эти химические кусочки, которые нужны им в качестве сырья, и разработать стратегию для соединения частей в лаборатории. В случае необходимости они могут искать вдохновение из коммерческих базы данных химических веществ и реакций, таких как *SciFinder* или *Reaxys*. Введя молекулярные структуры или реакции в эти базы данных, они дают примеры из литературы. Но даже с помощью онлайн-справочников люди часто терпят неудачу при синтезе, ибо с таким количеством химии, которое заложено в базе данных, никто не может знать, что получится в конечном итоге.

Первой ласточкой в создании программы-синтезатора любой органической молекулы стала разработка химика Гарвардского университета в Кембридже, штат Массачусетс, Элиаса Кори, который формализовал правила ретросинтеза в 1960-х годах. В 70-х Кори создал программное обеспечение под названием *LHASA* (Логика и Эвристика, Примененные к Синтетическому Анализу (ЛЭПСА)), которое могло бы использовать эти правила, чтобы предлагать последовательности шагов к синтезу [2]. Но *LHASA* (ЛЭПСА) и ее преемники никогда не пользовались большой популярностью: либо базы данных включали слишком мало реакций, либо в них было слишком много ошибок, либо алгоритмы не оценивали надлежащим образом совместимость предлагаемых реакций со всеми функциональными группами в молекуле.

Из этого разочарования и родилась *Chematica*. *Chematica* – это программа, которая ищет синтетические пути, ведущие от стартовых химикатов (красный), через последовательности промежуточных продуктов

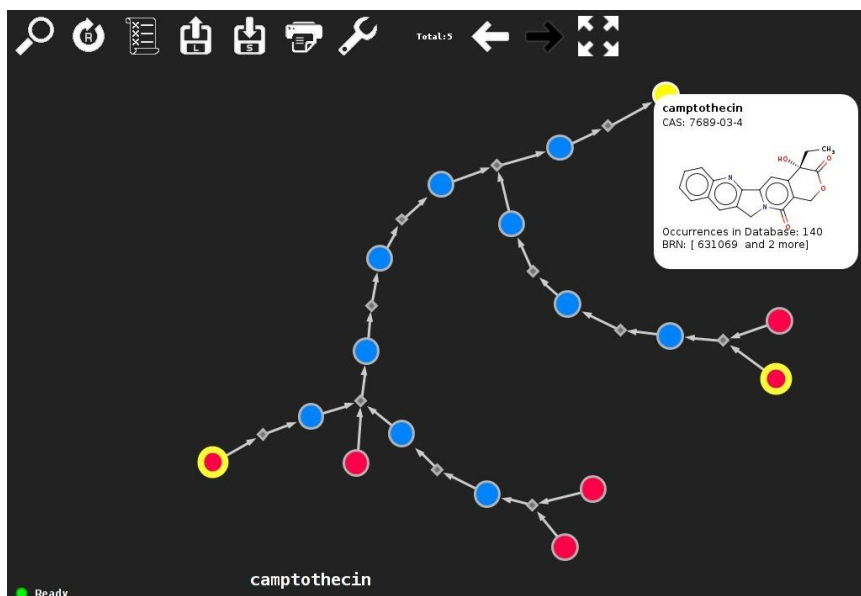


Рис. 2. – Сетевая схема синтеза камптотецина, который является основой нескольких противораковых препаратов [1]

(синий) к целевому соединению (желтый).

Целевым веществом выбранного примера является камптотецин: встречающееся в природе соединение, которое является основой нескольких противораковых препаратов (рис. 2).

Химик может ввести свое место назначения (целевую молекулу) в *Chematica* и искать любое количество путей получения, основанных на стоимости, популярности реагента, доступности и количестве шагов – и все это в считанные секунды. Каждый шаг и продукт, который он производит, получают оценку, основанную на двух уравнениях: функции реакции и химического скоринга (т.е. по очкам). Функция скоринга реакции (*RSF – Reaction Scoring Function*) будет штрафовать ход, если включенная химия сложна для вывода или просто сложна из-за стоимости реагентов или большого количества времени на получение и т. п. (рис. 3). Но функция химического скоринга (*CSF – Chemical Scoring Function*) приписывает оценку, основанную на простоте молекулы, и является ли она известной структурой – чем выше показатель *CSF*, тем более привлекательным является путь [1].



Рис. 3. – Интерфейс программы *Chematica* [1]

Сетевая химия

Программа *Chematica* в целом состоит из двух модулей: **сетевого модуля** и **модуля ретросинтеза**.

Сетевой модуль. Подход *Chematica* к вычислительному синтетическому планированию можно рассмотреть с точки зрения теории сетей, в соответствии с которой химические реакции представлены в виде сетевого графа, синтетическое знание происходит от изучения соединений в нем, а новые синтетические пути обнаруживаются посредством «переустройства» этой сети. Сеть *Chematica* состоит из около 10 миллионов химических веществ, которые «соединены» аналогичным числом реакций, взятых из химической литературы. Благодаря соответствующим алгоритмам поиска, аналогичным тем, которые используются в телекоммуникационной отрасли для поиска сетей связи, *Chematica* может пересекать эту гигантскую сеть, анализировать миллиарды синтетических возможностей, определять оптимальные пути синтеза с учетом пользовательских ограничений и делать поэтому в течение нескольких секунд.

Chematica предлагает семейства сетевых алгоритмов с возможностью разработки синтетических путей, оптимизированных по стоимости или по популярности. «Алгоритмы затрат» определяют путь с минимальными суммарными денежными затратами при сопоставлении результатов с различными экономическими реалиями (т. е. высокими/низкими затратами на рабочие компоненты). «Алгоритмы популярности» ищут путь с максимальной общей популярностью всех веществ, входящих в его состав. В сетевом формализме синтетическая популярность измеряется количеством входящих и исходящих синтетических соединений молекулы, что свидетельствует о легкодоступных, надежных химических веществах. Высокая связь также гарантирует, что химиаты вокруг этих молекул широко используются, следовательно, надежны. Оба этих семейства сетевых алгоритмов спланировать оптимальный синтетический механизм.

Модуль ретросинтеза (*Syntaurus*). *Syntaurus* по структуре имитирует «протоколы», через которые химики изучают химию. Химик изучает специфические реакции и классы реакций, узнает, когда и в каких конкретных условиях основные «движения» химии могут быть использованы для создания молекул. Аналогичным образом, химик-компьютер *Syntaurus* обладает базой знаний, которая включает в себя не только правила реакции, но и условные правила химии, так что она может «разрезать» специфическую связь в контекстно-зависимом режиме (т. е. путем изучения других групп, присутствующих в молекуле) (рис. 4). В значительной степени это способность оценивать контекстную зависимость, которая позволяет *Chematica* создавать жизнеспособные химические вещества [3].

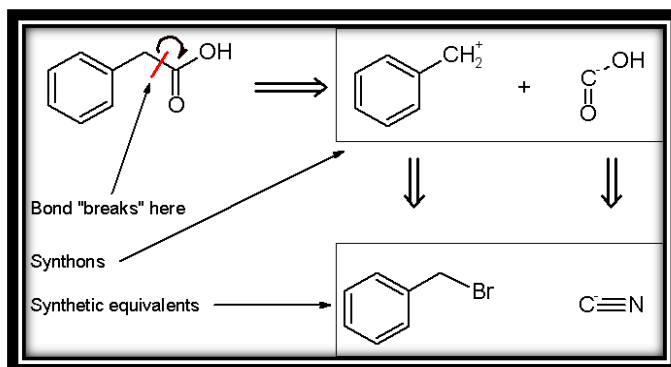


Рис. 4. – Пример ретросинтетического анализа, иллюстрирующий получение фенилуксусной кислоты [3]

Список использованных источников:

1. Peplow, M. Organic synthesis: The robo-chemist / M. Peplow // Nature. – Vol. 512, Iss. 7512. – 07 August 2014. [Electronic resource]. – Mode of access: <https://www.nature.com/news/organic-synthesis-the-robo-chemist-1.15661>. – Date of access: 13.04.2018.
2. Corey, E. J. Computer-assisted synthetic analysis. Methods for machine generation of synthetic intermediates involving multistep look-ahead / E. J. Corey, W. J. Howe, D. A. Pensak // J. Am. Chem. Soc. – 1974. – Vol. 96, № 25. – P. 7724 – 7737.
3. How does Chematica work? // Grzybowski Scientific Inventions. [Electronic resource]. – Mode of access: <http://chematica.net/#/technology/technology/howdoeschematicawork>. – Date of access: 12.04.2018.