Artificial Intelligence in Organic Synthesis

Newsletter-1 | April 2022

About the author

Rajesh is a group Leader in Organic Synthesis Dept. He has around 14 years of industrial experience in Pharma, Agro & Fine chemicals. He received his Ph.D in Organic chemistry from Osaka University. He is associated with JRF for the past one year.

Artificial Intelligence in Organic Synthesis

Designing a multistep organic synthesis needs an extensive knowledge base including thousands of reaction types, pattern-recognition skills, visualization in case of chiral targets, and sometimes a human instinct or even creative inspiration. The planning process is quite tedious and involves finding the matching reaction types; considering potential reactivity differences, non-selectivity's, by-products; and often taking clues from the literature for similar reaction precedents.
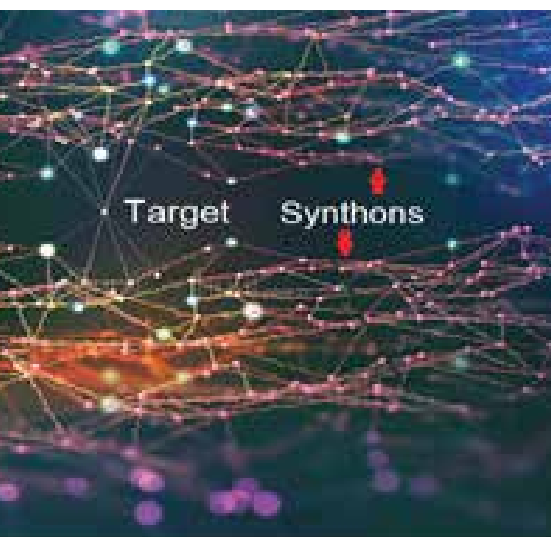
The approach to ease the process is to form a database network, which started with **chemical abstracts** the largest well organized scientific data available till now.

The other promising technique was retrosynthetic analysis, for which Prof. Corey won the Nobel Prize for Chemistry in 1990[1]. But the process is not molecule specific and hence resulted in a combinatorial explosion resulting in too many synthons for a molecule. A database or retrosynthetic analysis thus remains inadequate when it comes to the ever-changing requirements of streamlined & high throughput organic processes as the number of possible routes to consider is very large.

Semi-automated tools like Sci-Finder & Reaxsys, to some extent simplified the process, but still not sufficient to fulfill the requirements for automated predictions. These complexities made scientists to restructure scientific data in a more intelligent way paving way for the futuristic word called Artificial Intelligence (AI).

1

JRF GLOBAL
Pioneering Solutions since 1977 - Responsibly

AI in early days involves result generation from the data sets and diagnosed according to their nearest plausibility. This resulted too many options without a definite solution for the problem. With technical advancements, even today's **AI** is still narrow. Applying rules to new challenges requires a huge amount of new data training, rule creation and time. Renewed interest in AI, driven by improved computing power, data availability and algorithms, is overturning the limited success previously obtained.

Traditional models like Random Forest or Gaussian Processes, which have been applied successfully in the past has been substituted with novel techniques like deep neural networks (DNNs), convolutional neural network (CNNs) or recurrent neural network (RNNs) and recognized as valuable additions. Scientists are working in interrelating the technical data sets with the available network programmes to properly connect the required links.

Access to high-quality & reliable data in sufficient amounts is essential to generate productive AI for chemistry. It is imperative to ensure that all scientific discoveries are registered in data bases with absolute integrity & high accessibility.
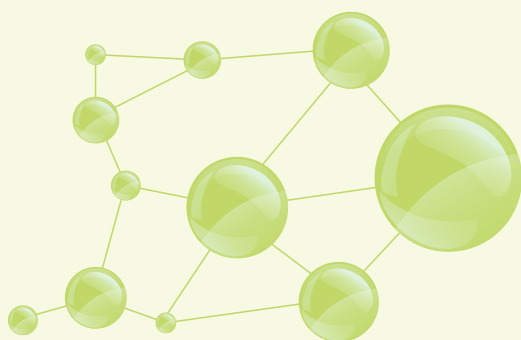
Some Milestones

Software's like ROBIA[2] & Drug Guru[3] can predict synthesis routes based on coded rules and molecular modeling calculations, generating possible transition states, intermediates. Kowalik et al.,[4]. have developed a promising approach using the Network of Organic Chemistry (NOC) and an enumeration algorithm of possible synthetic routes. Ahneman et al.,[5] from Princeton University used machine learning technique where Spartan calculations can be used as input for a random forest model to predict yields of multidimensional chemical data.

Synthia (Chematica)[6] is an in-silico chemical intelligence software development, led by Bartosz A et, al., that uses algorithms and a collective database to predict synthesis pathways for molecules. According to the team, it would probably be more like 50,000 reaction rules before the machine began to produce sensible pathways. IBM RXN for Chemistry used a simplified molecular-input line-entry system, or SMILES. SMILES represents a molecule as a sequence of characters, and this helps the software to overcome the problem of image correlation with databases.

Other software's such as MANIFOLD, SPAYA[7], SYLVIA & CHEM PLANNER were developed that were able to address a broad challenge but remains inadequate for complete automation in route prediction for a required target.

The first step in **AI** incorporation for chemical synthesis have been brilliant, but tedious with both time and labor intensive. Complete digitization of chemical inventions is found to be the key aspects towards further developments.

JRF GLOBAL
Pioneering Solutions since 1977 - Responsibly

Future of AI –
What chemists are looking at?

Mining the best synthetic route from available databases is nothing but a Needle from a Haystack!!!!

Practically, chemists can expect a proper rule or algorithm that predicts possible synthetic routes within, parallelly suggesting synthetic routes that meet our requirements. It is possible only when humans co-work with AI, rather than using them as tools. Unfortunately, we are far from realizing this vision, but keeping the options open for a breakthrough.

## Some of the key questions that needs to be answered are,

- Can **AI** execute the same decisions towards preferred reaction types made by scientists for many reasons (Considering Yield, Purity, Raw materials, and scalability)?

- Is all data available to be extracted and used as algorithm?

- Most of the data sets available are tailor made and it does not provide a generalized predictions for a given target.

- Does **AI** require a supporting software which can generalize a given data set for a wide variety of other conversions?

- When searching for the discovery of new compounds as a search problem, we need an effective representation of the problem. A blind search without specific commands that would explore the whole search space is not the right option.

JRF GLOBAL
Pioneering Solutions since 1977 - Responsibly

Even though AI is gaining enormous interests, experts agree that the commercial issues in this area are quite high.Every scientist should possess a good statistical knowledge along with software skills to operate the program effectively or should be able to collaborate with an IT guy providing all technical support. This again is going to be tedious.

A uniform system of naming, labelling & schematic representation of reaction systems are required so that the machine can be able to track all the relevant information's without any compromise. Parallelly, discoveries that have been incorrectly classified or published, are to be separated to improve the integrity of data output.
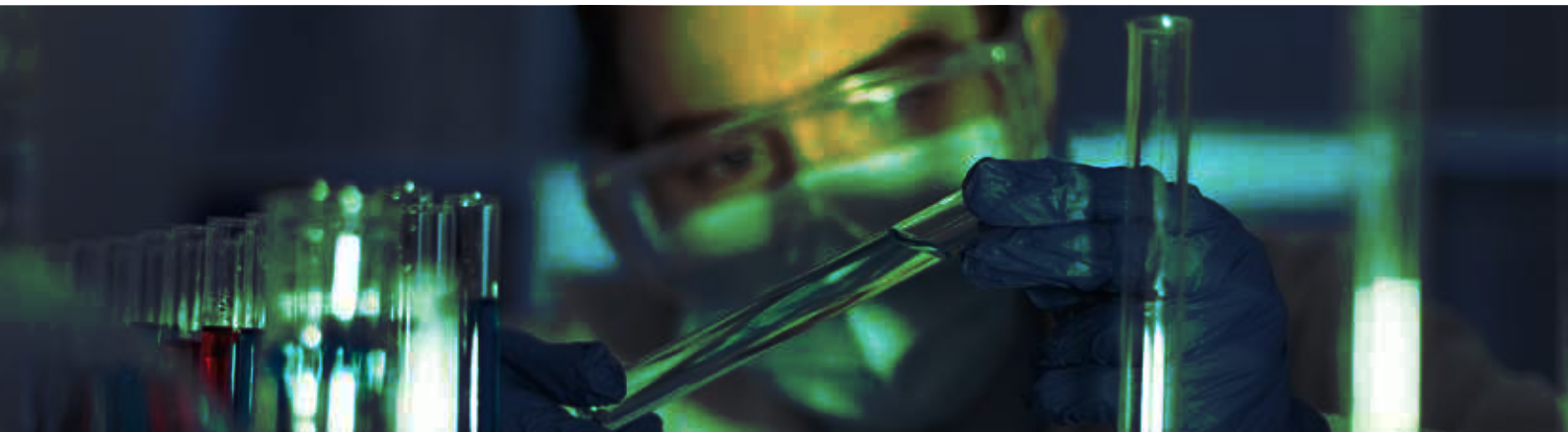
Conclusions

AI will not replace chemists at least in the short term but AI clearly appears as the future of chemistry. As a dark side, people feel that relying on AI will result in a shortage of work for humans and may make people resourceless (human intelligence) and dependent. But as every coin has two faces, so is the case with science. It is thus our choice to make science a boon for humanity or to convert it into a curse.

References

1. Science **1969**, *(166)*, 3902, 178-192

2. J Chem Inf Model. **2006**, *46(2)*, 606-14.

3. Bioorg Med Chem. **2006** Oct, *14(20)*, 7011-22.

4. *Angew Chem Int Ed Engl* **2012**, *51(32)*:7928–7932.

5. Science **2018**, 360, 186–190.

6. Chem **2018**, 390–398.

7. www.spaya.ai

JRF GLOBAL
Pioneering Solutions since 1977 - Responsibly