Chemistry update
OCTOBER 2023
www.hoddereducation.co.uk/chemistryreview

# Artificial intelligence and chemistry


QUARDIA INC./STOCK.ADOBE.COM

Artificial intelligence has rapidly risen to fame, thanks to the likes of ChatGPT and Google's Bard, and it is already changing the world of chemistry. **Joseph Lloyd** explains

'Artificial' is rarely perceived positively: artificial light can disrupt our natural sleep cycles, and food companies are quick to tell us their products do not use artificial colourings or flavourings. What about artificial intelligence? Could it be revolutionary? Before we can say for certain, we need to know what artificial intelligence (AI) is.

While computers follow instructions, AI allows machines to perform actions that typically require human intelligence, such as learning and problem solving. Just as humans have applied intelligence to every element of modern life, AI could be applied widely too. AI can use machine learning, natural language processing, robotics or deep learning — or even a combination of several technologies — to perform and enhance our approach to problems and tasks. The world of chemistry is no exception.

## AI is changing chemistry

Students of chemistry learn of the infinite possibilities that come with the bonds and interactions between atoms and molecules. The more we learn, the more we realise the impossibility of humans ever comprehending and using the vast library of options.

This is particularly accurate when considering proteins, where the number and sequence of amino acids in the primary structure is vital to secondary and tertiary structures. One slight change to the primary structure can affect hydrogen bonding, sulfur–sulfur bonds and other intramolecular interactions, which go on to deeply disrupt the shape and therefore the action of the protein. In among all the possible combinations, there is naturally a wealth of opportunities for scientific advancement. Detailed knowledge of protein structure enables drug companies to design molecules to bind with target proteins in the fight against diseases. Proteins can become vaccines and therapies, tools for the bioremediation of pollution or catalysts for synthesis.

Science wants to be able to synthesise proteins capable of these feats, yet the complexity of proteins stands in the way. Humans have been limited to tweaking known protein structures, rather than building tailored molecules from the ground up, known as *de novo* protein design. This is where AI steps in.

## Future potential

Prior to implementing AI, a great deal of work had already been done in documenting all the protein structures that had been determined experimentally. The Protein Data Bank was established in 1971. It is an open-source database that holds information on more than 100 000 protein structures. That is only a fraction of the billions of known protein structures. Determining each structure in detail takes months of excruciating effort and, even then, the atomic detail of the structure can remain elusive. Additionally, predicting how a protein folds has been an unresolved scientific question for over half a century.

Using the huge quantity of high-quality data in the Protein Data Bank allowed AI, in the form of AlphaFold, to achieve something remarkable. The AlphaFold protein-structure-prediction tool predicted

Next page

Chemistry update
OCTOBER 2022
www.hoddereducation.co.uk/chemistryreview

the structures of nearly all known proteins with incredible accuracy. Because of AlphaFold and the systems it inspired, the database of known protein structures has grown to millions of records, which scientists can now use and investigate.

## Problems

The immediate problem in the wider field of chemistry is not overwhelming progress, but rather slow and frustrating development. AI is only as good as the data that feeds it. Protein chemistry has data in abundance, so the world of proteins has already enjoyed great leaps forwards thanks to AI. Unfortunately, this is a rare example.

In other areas of chemistry, accumulating even 5000 relevant records is difficult. This is the amount of data that is needed to allow AI to outperform existing computational models. Those records must also include examples of negative or obscure outcomes, which are more rarely reported, and they must be openly accessible to those feeding the AI machine. Without this, AI is liable to make mistakes.

## Where are we heading?

AI is likely to follow in a similar path to the advent of computers. With computers, scientists were able to speed-up and automate mundane, repetitive tasks and focus on the bigger questions. AI is already doing this. While it took humans 50 years to determine 100 000 structures, AI predicted billions in a fraction of that time. AI is unlikely to replace chemists. Instead, it will become another tool at their disposal.

Before this is possible, more comprehensive databases are needed to feed AI machines. High-quality data will enable AI to revolutionise other areas of chemistry in the same way it has revolutionised protein chemistry. Progress will become rapid and scientists will have to wrestle with new moral and technical questions.

### Questions

1 Suggest other areas of chemistry where AI could be used, apart from protein design.
2 Explain how hydrogen bonds influence the structure of proteins.

### Answers

1 All reasonable suggestions including:
- Drug discovery: suggesting molecules that could have a therapeutic effect based on their likely interactions with living organisms.
- Chemical synthesis: suggesting how to synthesise new molecules based on known reactions.
- Materials chemistry: creating new materials with specific properties.
- Atmospheric chemistry: using historic data to make predictions of climate change.
- Data interpretation: AI could process vast quantities of data and identify patterns.

2 The following points should be outlined:
- Hydrogen bonds form between the carbonyl oxygen of one amino acid residue and the amino hydrogen of another.
- They form because the lone pairs on an oxygen atom have a delta negative charge and the hydrogen attached to a nitrogen atom has a delta positive charge.
- The lack of rotation around peptide bonds, and the planar arrangement of the atoms around these bonds, restricts the way in which protein chains can fold to form secondary structures (alpha helices and beta pleated sheets). These secondary structures are held in place by hydrogen bonds.
- Amino acids have different R groups (side chains), some of which can also form hydrogen bonds with other parts of the protein, stabilising the overall tertiary structure.