Researchers develop advanced software to speed up discovery of new polymers

May 15 2023

Credit: Unsplash/CC0 Public Domain

A program to advance the discovery of new polymers has been developed by a team of interdisciplinary researchers across King's Faculty of Natural, Mathematical and Engineering Sciences. The

software, called PySoftK, uses AI to identify new polymer materials, which could be used across a wide range of applications including in medical technology, pharmaceuticals, energy storage and more. The software is described in a paper available on the *chemRxiv* pre-print server, and PySoftK is available on [GitHub](#).

Professor Chris Lorenz from the Department of Physics and lead researcher on the technology said, "PySoftK will allow us to accelerate the development of novel polymers for a whole range of applications, from using polymers with embedded nanoparticles to stitch human tissue together, to improving energy storage methods. These materials will help form a building block to tackle large scale challenges that we face in [health care](#), in developing biodegradable home and personal care products and in creating more environmentally friendly energy storage systems."

Polymers are [large molecules](#) made up of smaller repeating molecules called monomers, which bond together in a chain-like fashion to form a long polymer molecule. Polymers can be naturally occurring, such as proteins and DNA, or they can be synthetic, such as plastics and synthetic fibers. This new software development could change the way we investigate the relationship between the [chemical structure](#) and function of new polymeric materials, by providing a robust dataset for researchers to train artificial intelligence (AI) to identify desirable polymer properties.

Synthetic polymers can be designed to interact with changes in their environment or make use of certain properties. For example, Gore-Tex, a polymer used in clothing was developed as an improvement to nylon, a traditional polymer. While both materials are waterproof, Gore-Tex is also be breathable, because it has been designed with a particular chemical property to perform a specific function. This is known as a designer polymer.

Other areas where designer polymers are used, include medical ointments, paints, coatings, food packaging, biomedical imaging and energy storage. Designer polymers have the potential to have a wide range of different functions due to their underlying physical and chemical properties, which originate from the type and arrangement of monomers that build the polymer.

To advance our discovery of these types of materials, high-performance computers (HPCs) are used to simulate and predict the behavior of polymers, which then informs researchers how best to build polymers with the desired properties for fulfilling certain tasks.

Over the past several decades, molecular scale simulations, computer simulations representing 3D structures of molecules, have improved our understanding of the relationship between chemical structure and function in increasingly complex polymers. However, more recent advances in computing power and computational algorithms have enabled scientists to investigate more complex systems and provide more accurate predictions using molecular-scale simulations at speed. This can lead to faster and more cost-effective design of materials, as less time is devoted to rounds of experimentation.

Lorenz suggests, "Normally, maintaining a large, diverse and accurate molecular database can be a hugely costly and time intensive process, as researchers race to label and categorize models correctly."

By offering a set of tools and programming modules to automate the process of curating, modeling and creating libraries of polymers, PySoftK facilitates the generation of large databases on which to train future machine learning (ML) and deep learning (DL) models. This allows researchers to move their focus away from exhaustive library maintenance and onto discovering new materials."

Dr. Alejandro Santana Bonilla, research software engineer within the Faculty of Natural, Mathematical and Engineering Sciences, and one of the lead researchers on the project said, "The software package is versatile, flexible, and easy to install. It can generate a wide range of polymer topologies and perform library generation in a fully parallelized manner, making it highly efficient,"

Researchers hope that these models will be the driving force of new designer polymer development. PySoftK could also play a significant role for researchers in nano- and bio-technology, who are searching for new functional materials. But without reliable data to train the AI, they risk making inaccurate predictions.

"Ultimately, the integration of molecular scale simulations and machine learning into the rational design process for designer polymers will be vital to furthering our understanding between the chemical structure and function of complex materials," says Lorenz.

  More information: Alejandro Santana-Bonilla et al, Modular Software for Generating and Modelling Diverse Polymer Databases, *chemRxiv* (2023). DOI: 10.26434/chemrxiv-2023-j01nk

Provided by King's College London