

UPDATES

AI can solve the crystallographic phase problem

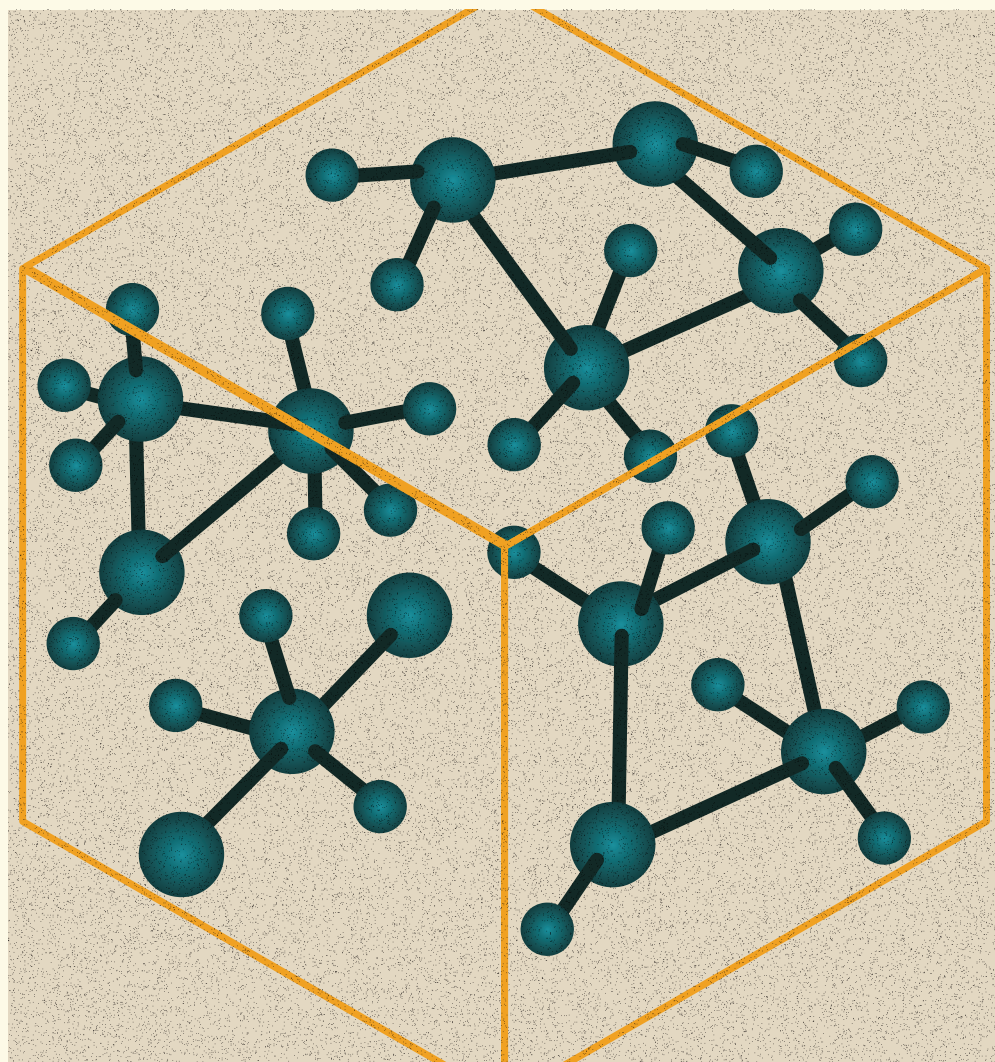
A neural network identifies molecular crystal structures with just a fraction of the data that other methods need.

As x rays scatter off a crystalline sample, detectors measure their intensity but are unable to measure their phase information. But both intensity and phase are necessary for x-ray diffraction to reconstruct the position of each atom in such samples. The phase problem in x-ray crystallography does have some partial solutions, which have made the experimental technique spectacularly successful in condensed-matter physics and biophysics.

Each of the solutions, however, has certain limitations. For direct methods, a complete set of diffraction-peak measurements is plugged into a system of equations that estimates phases based on the probable location of electrons in the crystal lattice. The phase information can also be collected by other means, such as computational microscopy (see the article by Manuel Guizar-Sicairos and Pierre Thibault, *PHYSICS TODAY*, September 2021, page 42). But they all need lots of high-resolution data and often require crystallographers to painstakingly grow high-quality samples. That's true even for molecules that are smaller than proteins, the typical targets of x-ray crystallography.

Now Anders Madsen and Anders Larsen of the University of Copenhagen and Toms Rekis of Goethe University Frankfurt in Germany report that a deep-learning algorithm can determine the crystal structure of various molecules with up to about 50 atoms. Despite working with just 10–20% of the data that are typically required for direct methods, the AI approach generated crisp electron-density maps of a sample's structure.

Madsen and colleagues trained the



THE 3D STRUCTURE of molecules in a crystal can be determined by deep-learning AI. (Adapted from A. S. Larsen, T. Rekis, A. Ø. Madsen, *Science* **385**, 522, 2024.)

AI algorithm with 49 million chemical crystal structures that they artificially generated. Then they tested the algorithm against 2300 real chemical structures that they obtained from the Cambridge Structural Database, an archive of more than 1.25 million 3D crystal structures determined by x-ray diffraction and neutron diffraction. For their proof of principle, the researchers chose molecular crystal structures that have a length scale of about 10 Å and fall into one of the relatively simple space groups, which describe the packing symmetry of a crystal's molecules. The first results were successful: The algorithm produced phase values and correctly identified the 3D structure for each of the 2300 test cases.

The phase problem isn't entirely solved. Many other molecular crystal structures in different space groups have yet to be evaluated, and Madsen and his team are testing the algorithm's effectiveness for more complex space groups and larger crystal sizes. Existing AI tools, such as AlphaFold, are already making exquisite predictions of proteins' 3D structures (see *PHYSICS TODAY*, October 2021, page 14); Madsen and colleagues' algorithm fills a complementary niche for smaller molecules that are difficult to crystallize or for weakly scattering crystals that are unable to yield better than low-resolution diffraction data sets. (A. S. Larsen, T. Rekis, A. Ø. Madsen, *Science* **385**, 522, 2024.)

Alex Lopatka