Scientists from the iMolecule group at Skoltech Center for Computational and Data-Intensive Science and Engineering (CDISE) developed BiteNet, a machine learning (ML) algorithm that helps find drug binding sites, i.e. potential drug targets, in proteins. BiteNet can analyze 1,000 protein structures in 1.5 minutes and find optimal spots for drug molecules to attach. The research was published in Communications Biology.

Proteins, the molecules that control most biological processes, are typically the common targets for drugs. To produce a therapeutic effect, drugs should attach to proteins at specific sites called binding sites. The protein's ability to bind to a drug is determined by the site's amino acid sequence and spatial structure. Binding sites are real "hot spots" in pharmacology. The more binding sites are known, the more opportunities there are for creating more effective and safer drugs.

Skoltech CDISE assistant professor Petr Popov and Ph.D. student Igor Kozlovskii developed a new computational approach for spatio-temporal detection of binding sites in proteins by applying deep learning algorithms and computer vision to protein structures treated as 3-D images. With this new technology, one can detect even elusive sites: for instance, scientists managed to detect binding sites concealed in experimental atomic structures or formed by several protein molecules for the ion channel, G protein-coupled receptor, and the epithelial growth factor, one of the most important drug targets.

Petr Popov, the study lead and assistant professor at Skoltech, says, "The human genome consists of nearly 20,000 proteins, and very few among them are associated with a pharmacological target. Our approach allows searching the protein for binding sites for drug-like compounds, thus expanding the array of possible pharmacological targets. Additionally, initial structure-based drug discovery strongly depends on the choice of the protein's atomic structure. Working on a structure with the binding site barred for the drug or missing altogether can fail. Our method enables analyzing a large number of structures in a protein and finding the most suitable one for a specific stage."

According to Igor Kozlovskii, the first author of the paper, BiteNet outperforms its counterparts both in speed and accuracy: "BiteNet is based on the computer vision, we treat protein structures as images, and binding sites as objects to detect on this images. It takes about 0.1 seconds to analyze one spatial structure and 1.5 minutes to evaluate 1,000 protein structures of about 2,000 atoms."

More information: Communications Biology (2020). DOI: 10.1038/s42003-020-01350-0

Provided by Skolkovo Institute of Science and Technology