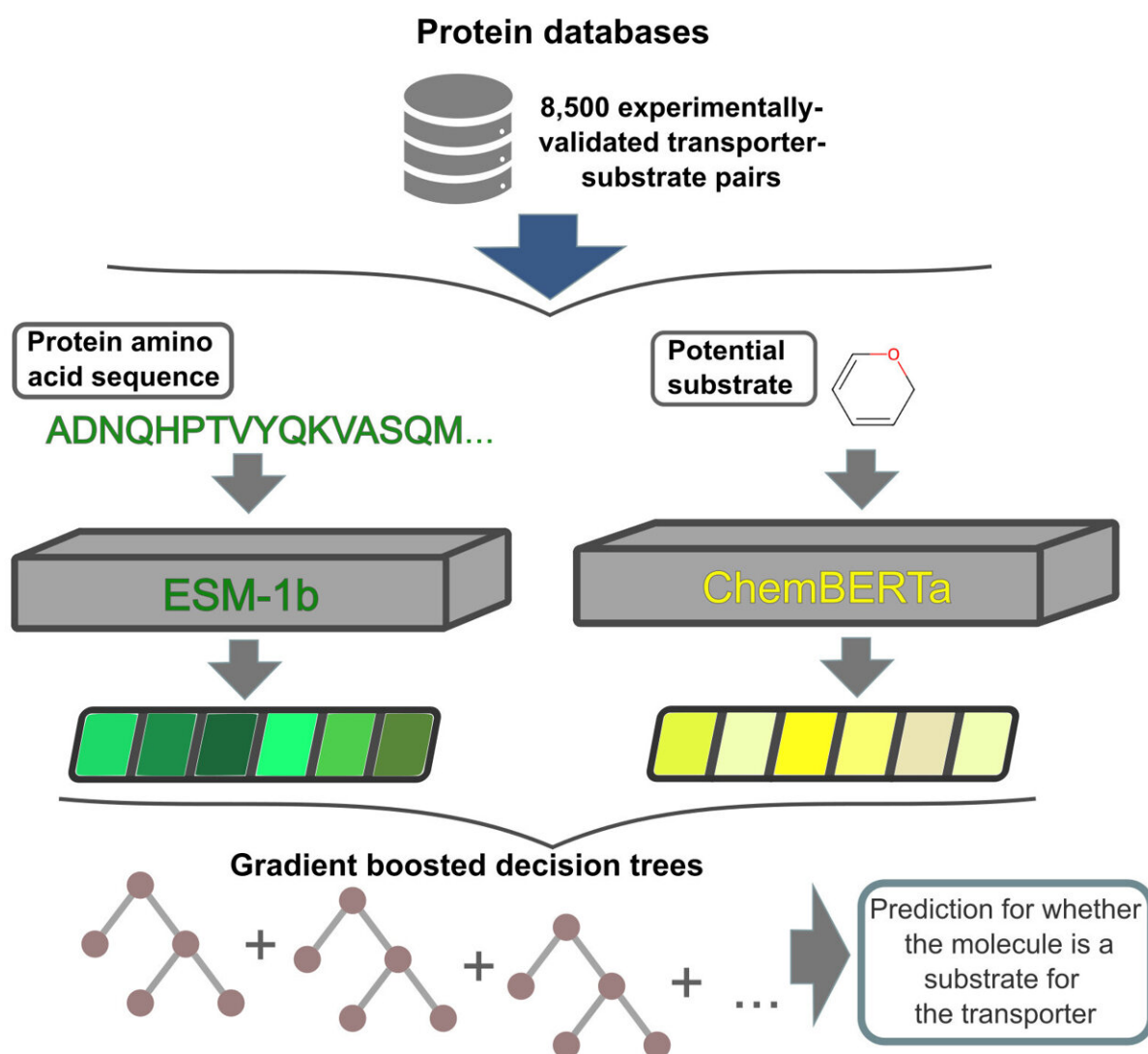


A new AI model can predict substrate movement into and out of cells

September 26 2024



Schematic diagram of the prediction process: Protein databases provide a dataset with 8,500 experimentally validated transporter-substrate pairs to train the model (top). Transport proteins comprise a sequence of amino acids, which have been converted into vectors by a deep learning model (center left, in different shades of green). Information about potential substrates is also converted into numerical vectors (center right, in different shades of yellow). These vectors train a so-called gradient boosting model (ensemble of multiple decision trees) to predict whether the molecule is a substrate for a specific transport protein (bottom). Credit: HHU/Alexander Kroll

Transport proteins are responsible for the ongoing movement of substrates into and out of a biological cell. However, it is difficult to determine which substrates a specific protein can transport.

Bioinformaticians at Heinrich Heine University Düsseldorf (HHU) have developed a model—called SPOT—that can predict this with a high degree of accuracy using artificial intelligence (AI).

The researchers present [their approach](#), which can be used with arbitrary [transport proteins](#), in the journal *PLOS Biology*.

Substrates in biological cells must be continuously transported inwards and outwards across the cell membrane to ensure the survival of the cells and enable them to perform their function. However, not all substrates that move through the body should be allowed to enter the cells. And some of these transport processes must be controllable so that they only occur at a particular time or under specific conditions in order to trigger a cell function.

The role of these active and specialized transport channels is assumed by so-called transport proteins (transporters), a wide variety of which are integrated into the cell membranes. A transport protein includes a large number of individual amino acids, which together form a complex three-

dimensional structure.

Each transporter is tailored to a specific molecule—the so-called [substrate](#)—or a small group of substrates. But which, exactly? Researchers are constantly searching for matching transporter-substrate pairs.

Professor Dr. Martin Lercher, from the research group for Computational Cell Biology and corresponding author of the study, says, "Determining which substrates match which transporters experimentally is difficult. Even determining the three-dimensional structure of a transporter—from which it may be possible to identify the substrates—is a challenge, as the proteins become unstable as soon as they are isolated from the [cell membrane](#)."

"We have chosen a different—AI-based—approach," says Dr. Alexander Kroll, lead author of the study and postdoc in the research group of Professor Lercher. "Our method—which is called SPOT—used more than 8,500 transporter-substrate pairs, which have already been experimentally validated, as a training dataset for a deep learning model."

To enable a computer to process the transporter proteins and substrate molecules, the bioinformaticians in Düsseldorf first convert the [protein sequences](#) and substrate molecules into numerical vectors, which can be processed by AI models. After completion of the learning process, the vector for a new transporter and those for potentially suitable substrates can be entered into the AI system. The model then predicts how likely it is that certain substrates will match the transporter.

Kroll explains, "We have validated our trained model using an independent test dataset where we also already knew the transporter-substrate pairs. SPOT predicts with an accuracy above 92% whether an

arbitrary molecule is a substrate for a specific transporter."

SPOT thus suggests highly promising substrate candidates. "This enables us to limit the search scope for experimenters to a significant degree, which in turn speeds up the process of identifying which substrate is a definite match for a transporter in the laboratory," says Professor Lercher, explaining the link between bioinformatic prediction and experimental verification.

Kroll adds, "And this applies for any arbitrary transport protein, not just for limited classes of similar proteins, as is the case in other approaches to date."

There are various potential application areas for the [model](#).

Lercher notes, "In biotechnology, [metabolic pathways](#) can be modified to enable the manufacture of specific products such as biofuels, or drugs can be tailored to transporters to facilitate their entry into precisely those cells in which they are meant to have an effect."

More information: Alexander Kroll et al, SPOT: A machine learning model that predicts specific substrates for transport proteins, *PLOS Biology* (2024). [DOI: 10.1371/journal.pbio.3002807](https://doi.org/10.1371/journal.pbio.3002807)

Provided by Heinrich-Heine University Duesseldorf

Citation: A new AI model can predict substrate movement into and out of cells (2024, September 26) retrieved 27 September 2024 from <https://phys.org/news/2024-09-ai-substrate-movement-cells.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.