

## Pressemitteilung

Heinrich-Heine-Universität Düsseldorf

Dr.rer.nat. Arne Claussen

18.07.2023

<http://idw-online.de/de/news817996>

Forschungsergebnisse, Wissenschaftliche Publikationen  
Biologie, Informationstechnik  
überregional



## KI sagt Funktion von Enzymen voraus

**Bioinformatik: Veröffentlichung in Nature Communications** In zellulären Stoffwechselprozessen spielen Enzyme eine entscheidende Rolle. Um diese Prozesse auch quantitativ einschätzen zu können, müssen Forschende die sogenannte „Wechselzahl“ (kurz *k<sub>cat</sub>*) der Enzyme kennen. In der Fachzeitschrift *Nature Communications* beschreibt ein Team von Bioinformatikern der Heinrich-Heine-Universität Düsseldorf (HHU) nun ein Verfahren, um diese Größe bei verschiedenen Enzymen mittels KI-Methoden vorherzusagen.

Enzyme sind wichtige Biokatalysatoren in allen lebenden Zellen. Sie sind im Normalfall große Proteine, die kleinere Moleküle – sogenannte Substrate – an sich binden und diese anschließend in andere Moleküle umwandeln, die „Produkte“. Ohne die Enzyme könnte die Reaktion der Substrate in die Produkte nicht oder nur mit sehr kleiner Rate stattfinden. Die meisten Organismen besitzen Tausende verschiedene Enzyme. In vielen biotechnologischen Prozessen, aber auch im Alltag – von der Reifung von Brotteigen bis hin zu Waschmitteln – finden Enzyme vielfache Anwendungen.

Die maximale Geschwindigkeit, mit der ein bestimmtes Enzym seine Substrate in Produkte umwandeln kann, wird durch die sogenannte Wechselzahl *k<sub>cat</sub>* (englisch „turnover number“) bestimmt. Sie ist ein wichtiger Parameter für die quantitative Erforschung von Enzymaktivitäten und spielt eine entscheidende Rolle für das Verständnis des zellulären Stoffwechsels.

Allerdings ist es zeitaufwendig und teuer, die *k<sub>cat</sub>*-Werte experimentell zu bestimmen; daher sind ihre Werte für die allermeisten Reaktionen nicht bekannt. Die Arbeitsgruppe Computergestützte Zellbiologie an der HHU um Prof. Dr. Martin Lercher entwickelte nun ein neues Verfahren namens TurNuP, um *k<sub>cat</sub>*-Werte von Enzymen mithilfe von KI-Methoden vorherzusagen.

Um ein *k<sub>cat</sub>*-Vorhersagemodell zu trainieren, wurden mithilfe von Deep Learning-Modellen Informationen über die Enzyme und die katalysierten Reaktionen in numerische Vektoren umgewandelt. Diese numerischen Vektoren dienen als Eingabe eines Machine Learning-Modells – eines sogenannten Gradient Boosting-Modells –, das die *k<sub>cat</sub>*-Werte vorhersagt.

Erstautor Alexander Kroll: „TurNuP liefert bessere Ergebnisse als frühere Ansätze und kann sogar erfolgreich auf Enzyme angewendet werden, die nur eine geringe Ähnlichkeit zu den Enzymen im Trainingsdatensatz aufweisen.“ Mit bisherigen Modellen konnten keine sinnvollen Aussagen getroffen werden, wenn die Enzymsequenz nicht mindestens 40 Prozent mit denjenigen der Trainingsenzyme überstimmte. TurNuP dagegen macht schon bei solchen Enzymen sinnvolle Vorhersagen, bei denen die Übereinstimmung zwischen 0 und 40 Prozent liegt.

Prof. Lercher ergänzt: „Wir zeigen in unserer Studie, dass die Vorhersagen durch TurNuP verwendet werden können, um die Konzentrationen von Enzymen in lebenden Zellen deutlich genauer als bisher vorherzusagen.“

Um das Vorhersagemodell möglichst vielen Anwendern leicht zugänglich zu machen, entwickelte das HHU-Team einen benutzerfreundlichen Webserver, mit dem andere Forschende  $k_{cat}$ -Werte von Enzymen vorhersagen lassen können.  
Link zum Webserver: <https://turnup.cs.hhu.de/>

Hintergrund: Machine Learning und Deep Learning

Deep Learning-Modelle bestehen aus vielen miteinander verbundenen Schichten von künstlichen Neuronen, die Muster in den Eingabedaten erkennen und verarbeiten können. Wie die numerischen Eingaben von einem Deep Learning-Modell verarbeitet werden, wird anhand von optimalerweise großen Trainingsdatensätzen erlernt.

Gradient Boosting-Modelle sind eine Methode des Machine Learnings, bei denen viele Entscheidungsbäume erstellt werden. Die Ergebnisse aller Entscheidungsbäume für eine bestimmte Eingabe werden verwendet, um Vorhersagen zu treffen. Ähnlich wie beim Deep Learning werden Trainingsdaten verwendet, um das Modell anzupassen, d.h. um die Entscheidungsbäume zu erstellen.

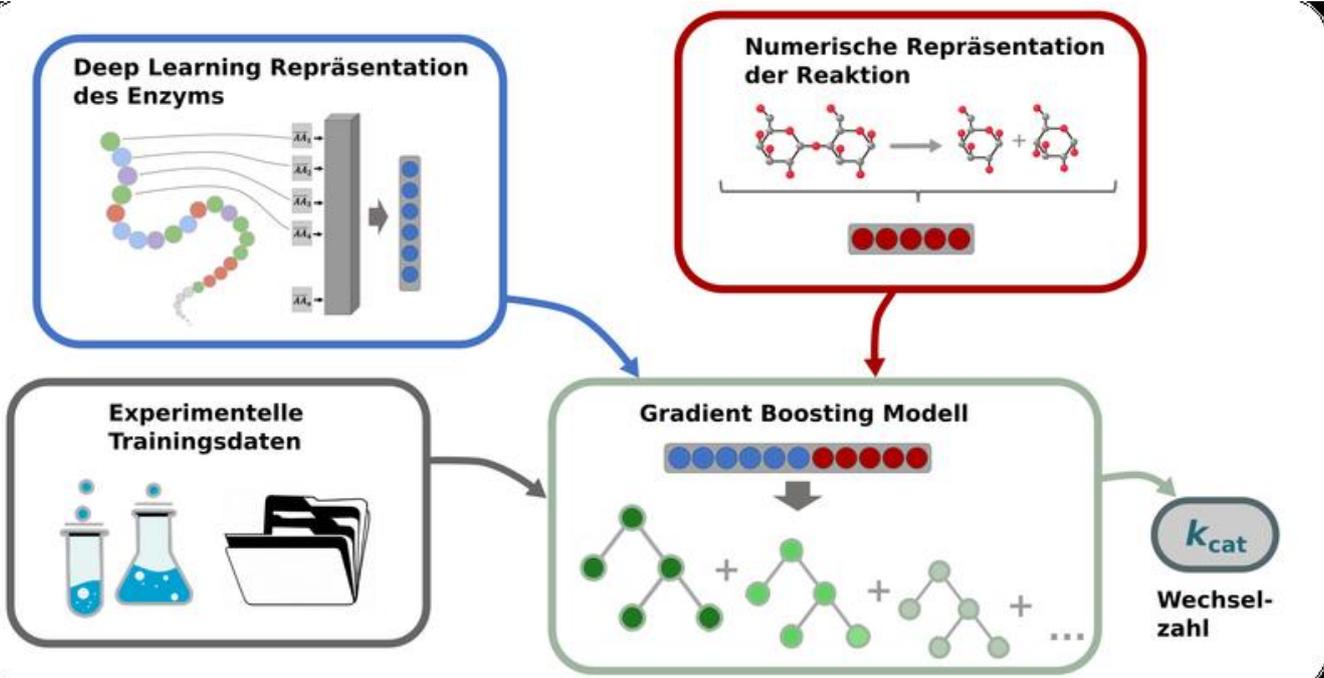
Ausführliche Bildunterschrift:

Schematische Darstellung des Vorhersageprozesses für Wechselzahlen von enzymatischen Reaktionen: Ein Enzym besteht aus Sequenzen von Aminosäuren; diese Sequenzen werden in numerische Vektoren umgewandelt, dargestellt durch graue Quadrate, welche dann von einem Deep Learning-Modell in einen einzigen Vektor transformiert werden (oben links). Informationen über die katalysierten Reaktionen werden ebenfalls in numerische Vektoren umgewandelt (oben rechts). Experimentell bestimmte Wechselzahlen (unten links) werden zum Training eines Gradient Boosting-Modells verwendet, um die Wechselzahl  $k_{cat}$  vorherzusagen (unten rechts). Gradient Boosting-Modelle sind ein Ensemble mehrerer Entscheidungsbäume, dargestellt in unterschiedlichen Grüntönen. (Abbildung: HHU / Alexander Kroll)

Originalpublikation:

Kroll, A., Rousset, Y., Hu, X.P., Liebrand, N. & Lercher, M.J.; Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning; Nature Communications 14, 4139 (2023).

DOI: 10.1038/s41467-023-39840-4



Schematische Darstellung des Vorhersageprozesses für Wechselzahlen von enzymatischen Reaktionen. Ausführliche BU s. Text.  
HHU / Alexander Kroll