

Künstliche Intelligenz ermöglicht Design neuartiger Proteine

Künstliche Intelligenz (KI) hat neue Möglichkeiten für das Design von Proteinen geschaffen, die zur Lösung medizinischer oder ökologischer Probleme maßgeschneidert sind. Ein Forschungsteam der Universität Bayreuth unter der Leitung von Prof. Dr. Birte Höcker hat jetzt ein computerbasiertes Modell zur Verarbeitung natürlicher Sprachen erfolgreich auf die Proteinforschung angewendet. Das Modell ProtGPT2 entwirft völlig selbständig neue Proteine, die zu stabilen Faltungen fähig sind und definierte Funktionen in größeren molekularen Zusammenhängen übernehmen könnten. In „Nature Communications“ werden das Modell und seine Potenziale wissenschaftlich beschrieben.

Natürliche Sprachen und Proteine ähneln sich in ihrem Aufbau: Aminosäuren ordnen sich in einer Vielzahl von Kombinationen zu Strukturen an, die im lebenden Organismus spezifische Funktionen haben – ähnlich wie Wörter in unterschiedlichen Kombinationen Sätze bilden, die bestimmte Sachverhalte ausdrücken. In den letzten Jahren wurden daher zahlreiche Ansätze entwickelt, um Prinzipien und Prozesse, welche die computergestützte Verarbeitung natürlicher Sprache steuern, in der Proteinforschung zu nutzen. „Die Verarbeitung natürlicher Sprache – das Natural Language Processing – hat aufgrund neuer KI-Technologien außerordentliche Fortschritte gemacht. Modelle der Sprachverarbeitung versetzen Maschinen heute in die Lage, sinnvolle Sätze nicht nur zu verstehen, sondern auch selbst zu erzeugen. Ein solches Modell war der Ausgangspunkt unserer Forschungsarbeiten. Mit detaillierten Informationen, die etwa 50 Millionen Sequenzen natürlicher Proteine betreffen, hat meine Mitarbeiterin Noelia Ferruz das Modell trainiert und in die Lage versetzt, selbstständig Proteinsequenzen zu erzeugen. Es versteht jetzt die Sprache der Proteine und kann sie kreativ anwenden. Diese kreativen Anwendungen folgen, wie wir festgestellt haben, den grundlegenden Bauprinzipien natürlicher Proteine“, sagt Prof. Dr. Birte Höcker, Leiterin der Arbeitsgruppe für Proteindesign an der Universität Bayreuth.

Das auf die Proteinentwicklung übertragene Sprachverarbeitungsmodell trägt den Namen „ProtGPT2“. Es kann nun genutzt werden, um Proteine zu entwerfen, die durch Faltung stabile Strukturen annehmen und in diesem Zustand dauerhaft funktionstüchtig sind. Zudem haben die Bayreuther Biochemiker*innen durch komplexe Untersuchungen herausgefunden, dass das Modell sogar Proteine kreieren kann, die in der Natur nicht vorkommen und in der Geschichte der Evolution womöglich noch nie existiert haben. Diese Einblicke in die unermesslich weite Welt möglicher Proteine öffnen die Tür zu einer innovativen Forschung, die bisher unbekannte Proteine auf neuartige Weise erzeugt. ProtGPT2 hat überdies einen weiteren Vorteil: Die meisten Proteine, die bisher *de novo* entworfen wurden, haben idealisierte Strukturen. Bevor sie angewendet werden können, sind in der Regel aufwändige Funktionalisierungsprozesse erforderlich, beispielsweise das Einfügen von Erweiterungen und Hohlräumen. Erst aufgrund dieser Funktionalisierung können sie mit ihrer Umgebung interagieren und in größeren Systemzusammenhängen genau definierte Funktionen übernehmen. ProtGPT2 erzeugt dagegen Proteine, die von Hause aus derart ausdifferenzierte Strukturen besitzen und so in ihrer jeweiligen Umgebung bereits einsatzfähig sind.

„Unser neues Modell ist ein weiterer eindrucksvoller Beleg für die systemische Affinität von Proteindesign und der Verarbeitung natürlicher Sprache. Künstliche Intelligenz eröffnet hochinteressante und vielversprechende Möglichkeiten, Methoden der Sprachverarbeitung für die Herstellung maßgeschneiderter Proteine zu nutzen. An der Universität Bayreuth wollen wir auf

diesem Weg dazu beitragen, innovative Lösungen für biomedizinische, pharmazeutische oder ökologische Probleme zu entwickeln“, sagt Prof. Dr. Birte Höcker.

Veröffentlichung:

Noelia Ferruz, Steffen Schmidt, Birte Höcker: ProtGPT2 is a deep unsupervised language model for protein design. Nature Communications (2022), DOI: <https://doi.org/10.1038/s41467-022-32007-7>

Die Erstautorin Dr. Noelia Ferruz war bis vor kurzem Mitglied der Arbeitsgruppe von Prof. Dr. Birte Höcker an der Universität Bayreuth. Sie forscht derzeit an der Universität Girona in Spanien.

Zum Verhältnis von Proteindesign und Modellen natürlicher Sprache vgl. den kürzlich erschienenen Review-Artikel: Noelia Ferruz, Birte Höcker: Controllable protein design with language models. Nature Machine Intelligence (2022), DOI: <https://doi.org/10.1038/s42256-022-00499-z>