Flexibilität ist alles: Wie maschinelles Lernen beim Bau neuer Proteine hilft

Ein Forschungsteam des Heidelberger Instituts für Theoretische Studien (HITS) und des Max-Planck-Instituts für Polymerforschung (MPIP) hat ein Computermodell entwickelt, das lernt, Proteine mit sehr flexiblen Strukturen zu bauen – sogar mit Mustern, die in natürlichen Proteinen ungewöhnlich sind. Das Modell wurde jetzt auf der International Conference on Machine Learning (ICML) präsentiert. Es markiert einen Schritt hin zum Ziel, neue Proteine für Anwendungen in der Biotechnologie, in Therapien und in der Umweltforschung zu konstruieren.

Proteine kommen in der Natur in einer großen Vielfalt vor, und als molekulare Werkzeuge erfüllen sie unterschiedlichste Aufgaben im Organismus. Dennoch kann die Entwicklung neuer, in der Natur nicht beobachteter Proteine zu neuen Funktionen führen und Probleme in Medizin oder Materialwissenschaft lösen. Gerade in den letzten Jahren begann ein „goldenes Zeitalter" des De-novo-Designs von Proteinen: Methoden des maschinellen Lernens ermöglichten eine bislang nicht dagewesene Genauigkeit der Modellierung. Dadurch können Forschende Proteinstrukturen mit spezifischen funktionellen Eigenschaften entwerfen, die nie zuvor beobachtet wurden. Dies ist besonders interessant für biotechnologische Anwendungen, die Entwicklung von Therapien und Probleme im Bereich der Nachhaltigkeit, Nachhaltigkeitsprobleme, wie zum Beispiel den Abbau von Plastik.

Eines der wichtigsten Merkmale funktioneller Proteine – große Biomoleküle mit komplexen Strukturen – ist ihre strukturelle Flexibilität: Sie wackeln, zappeln und verändern ihre Form. Beim derzeitigen Protein-Design fehlt dieses wichtige Merkmal weitgehend. Für ein Forschungsteam des Heidelberger Instituts für Theoretische Studien (HITS) und des Max-Planck-Instituts für Polymerforschung (MPIP) war dies der Ausgangspunkt, sich mit der Frage zu befassen, ob man Proteine mit einer maßgeschneiderten Flexibilität von Grund auf entwerfen kann. Die Ergebnisse ihrer Arbeit präsentierten sie jetzt auf der International Conference on Machine Learning (ICML) in Vancouver, Kanada.

Schnell und effizient: Ein System aus zwei Bausteinen

„Wir wollten ein Modell entwickeln, das lernt, Proteine so zu erzeugen, dass ihre Strukturen an einer bestimmten Stelle in einem bestimmten Maße flexibel sind", so Erstautor Vsevolod Viliuga (MPIP). Deshalb erarbeiteten sie ein System, in dem flexible Proteinstrukturen generiert werden können. Dieses System besteht aus zwei Bausteinen: Erstens einem neuronalen Netzwerk, darauf trainiert, die Flexibilität von sogenannten „Backbones" vorherzusagen – eine Reihe von Atomen, die als gefaltete Kette das „Rückgrat" der Makromoleküle bilden. Und zweitens einem generativen Modell für Proteinstrukturen.

„Natürliche Proteine erfüllen ihre Aufgaben so hervorragend, weil sie überall, wo sie gebraucht werden, flexibel einsetzbar sind", sagt Ko-Autor Leif Seute (HITS). „Wir können jetzt neue Proteine erzeugen, die diese Schlüsseleigenschaft nachahmen." HITS-Gruppenleiter Jan Stühmer ergänzt: „Das ist eine Erweiterung des „Geometric Algebra Flow Matching", kurz: GAFL, das wir letztes Jahr entwickelt haben." GAFL ist dreimal schneller als vergleichbare Modelle und erreicht nicht nur hohe Werte in der Designfähigkeit – die Ergebnisse sind ebenfalls natürlichen Proteinen in verschiedenen Kriterien ähnlicher.

Das Team konnte zeigen, dass das Modell Proteine mit den gewünschten Flexibilitätsmustern erzeugen kann, sogar für Muster, die in natürlichen Proteinen ungewöhnlich sind. Frauke Gräter, Teamleiterin auf Seiten des MPIP, stellt zusammenfassend fest: „Diese Arbeit ist ein Schritt in Richtung des anspruchsvollen Ziels, neue Proteine für Anwendungen zu entwerfen, bei denen Flexibilität erforderlich ist, wie beispielsweise bei Katalysatoren wie Enzymen."

Publikation:

Vsevolod Viliuga,Leif Seute, Nicolas Wolf, Simon Wagner, Arne Elofsson, Jan Stühmer, Frauke Gräter: Flexibility-conditioned protein structure design with flow matching.
https://icml.cc/virtual/2025/poster/46289

Die Arbeit wurde von der Klaus Tschira Stiftung (HITS Lab) gefördert.

Pressemitteilung

23.07.2025
Quelle: HITS gGmbH Heidelberger Institut für Theoretische Studien

Weitere Informationen

- Heidelberger Institut für Theoretische Studien (HITS)