

Poster 07: MISATO: Datensatz für das maschinelle Lernen von Protein-Ligand-Komplexen für die strukturbasierte Arzneimittelfindung

Grzegorz Popowicz¹, Till Siebenmorgen¹, Filipe Menezes¹, Sabrina Benassou², André Santos Dias Mourao¹, Radoslaw Kitel³, Pietro Lio⁴, Stefan Kesselheim², Marie Piraud⁵, Fabian Theis⁵, Michael Sattler¹

¹Molecular Targets and Therapeutics Center, Institute of Structural Biology, Helmholtz Munich, Neuherberg, Germany

²Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany

³Faculty of Chemistry, Jagiellonian University, Krakow, Poland

⁴Computer Laboratory, Cambridge University, Cambridge, UK

⁵Helmholtz AI, Helmholtz Munich, Neuherberg, Germany

Die MISATO-Datenbank stellt eine umfangreiche Sammlung von über 19.000 maschinell lesbaren Protein-Ligand-Komplexen dar, die speziell für struktur-basiertes Wirkstoffdesign mittels künstlicher Intelligenz (KI) entwickelt wurde. Die Datenbank basiert auf experimentellen Protein-Ligand-Strukturen der PDBbind-Datenbank, die mithilfe semi-empirischer quantenchemischer Methoden systematisch korrigiert und verfeinert wurden. Etwa 20 Prozent der Strukturen erforderten umfangreiche Anpassungen hinsichtlich Protonierungszuständen, Atomsorten und geometrischer Verzerrungen.

Zusätzlich zur statischen Strukturaufbereitung umfasst MISATO dynamische Informationen aus Molekulardynamik (MD)-Simulationen in explizitem Lösungsmittel mit einer Gesamtzeit von über 170 Mikrosekunden für knapp 17.000 Komplexe. Dabei konnten wichtige Einblicke in die Dynamik der Bindetaschen, die Bildung von transienten („kryptischen“) Bindestellen und das Anpassungsverhalten der Liganden gewonnen werden.

MISATO liefert umfangreiche quantenchemische Ligand-Parameter (Ionisationspotenziale, Elektronenaffinitäten, Polarisierungen) sowie aus MD-Simulationen abgeleitete Deskriptoren wie Root-Mean-Square(RMS)-Fluktuationen, Anpassungsfähigkeit der Bindestelle und geschätzte Bindungsenergien. Diese umfassenden Deskriptoren bilden eine wichtige Grundlage für die Entwicklung von KI-Modellen zur Vorhersage von quantenchemischen Eigenschaften, Proteinflexibilität und relativen Bindungsaffinitäten.

Die experimentelle Validierung der Datenbank erfolgte durch vergleichende Messungen der Protein-Flexibilität mittels NMR-Spektroskopie sowie durch Korrelation mit kristallographischen B-Faktoren. MISATO stellt eine validierte und innovative Ressource dar, die auch als Grundlage für das TPM-Patent (WO2024115584A1) diente und inzwischen kommerziell von Khumbu.AI genutzt wird. Die MISATO-Datenbank ist öffentlich zugänglich und ermöglicht die Entwicklung neuer KI-gestützter Ansätze im strukturbasierten Wirkstoffdesign.