



PRESSEINFORMATION

FORSCHUNG

Luise Dirscherl (Leitung)

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
dirscherl@lmu.de

Infoservice:
+49 (0)89 2180 - 3423

Geschwister-Scholl-Platz 1
80539 München
presse@lmu.de
www.lmu.de

Schneller, genauer, empfindlicher:

Neue Methode verbessert Sequenzanalysen entscheidend

München, 27 Dezember 2011 – Blitzschnell und dabei hoch empfindlich: Mit HHblits steht der Proteinforschung ein neues Werkzeug zur Verfügung, das die Analyse von Proteineigenschaften deutlich verbessern kann. LMU-Bioinformatiker um Dr. Johannes Söding vom Genzentrum der LMU entwickelten das neue Verfahren, mit dem verwandte Proteine durch die Ähnlichkeit ihrer Sequenzen - also der Abfolge ihrer Aminosäuren - schneller und genauer aufgespürt werden können als bisher - bei bis zu doppelter Empfindlichkeit. Aus den Eigenschaften bekannter Proteine oder Proteinuntereinheiten kann dann auf die Eigenschaften verwandter Proteine rückgeschlossen werden - einschließlich der Vorhersage ihrer Funktion und Struktur. „Diese Vorhersagen ermöglichen in allen Bereichen der auf molekularem Niveau forschenden Lebenswissenschaften oft ein gezielteres experimentelles Vorgehen“, sagt Söding, der auch dem Exzellenzcluster „Center for Integrated Protein Science Munich“ (CIPSM) angehört. (Nature Methods, 25.12.2011)

Proteine sind an allen Lebensvorgängen in der Zelle entscheidend beteiligt. Welche konkrete Aufgabe ein Protein übernimmt, beruht zum einen auf seiner Sequenz, also der Abfolge der 20 Aminosäuren, aus denen es aufgebaut ist. Zum anderen spielt die dreidimensionale Struktur, in die Proteine gefaltet sind, eine wichtige Rolle. Um die Funktion eines Proteins vorherzusagen, wird die Abfolge seiner Bausteine - der Aminosäuren - mit derjenigen anderer Proteine mit bereits bekannter Funktion und Struktur verglichen. Die Vergleichssequenzen sind in Datenbanken gespeichert, in denen auch Eigenschaften und Funktionen bereits bekannter Proteine hinterlegt sind. „Solche Sequenzanalysen sind ein fundamentales Werkzeug der Bioinformatik“, erklärt Söding.

Spezielle Suchprogramme bewerten die Ähnlichkeit der Sequenzen, indem diese so untereinander angeordnet werden, dass sich möglichst ähnliche

Kommunikation und Presse

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
dirschler@lmu.de

Infoservice:
+49 (0)89 2180 - 3423

Aminosäuren in der gleichen Spalte dieses sogenannten Sequenz-Alignments befinden. „Fast noch wichtiger als die Suche nach paarweise Sequenzähnlichkeiten ist die Erstellung sogenannter multipler Sequenz-Alignments, das heißt es werden ähnliche Sequenzen vieler verwandter Proteine gesucht und zu einem großen Alignment zusammengebaut“, sagt Söding. Da die Struktur und Funktionen von Proteinen meist konserviert sind – das bedeutet sie bleiben erhalten, auch wenn sich die Sequenz im Lauf der Evolution durch Mutationen ändert – sind multiple Sequenz-Alignments heutzutage die wichtigste Grundlage für die Vorhersage ihrer Struktur und molekularen Funktionen.

Seit fast 15 Jahren ist PSI-BLAST das meistverwendete Programm zur Suche von Proteinsequenzen, da es hohe Geschwindigkeit mit großer Empfindlichkeit und Genauigkeit verbindet. Nun ist es Södings Team gelungen, mit HHblits eine Methode zu entwickeln, die PSI-BLAST in allen Bereichen deutlich übertrifft. Dies erreichten die Wissenschaftler zum einen, indem sie sowohl die zu analysierenden Sequenzen als auch die Vergleichssequenzen in den zu durchsuchenden Datenbanken in sogenannte Hidden-Markov-Modelle (HMM) umwandeln. HMMs sind statistische Modelle der Aminosäuresequenzen, die auch die im Sequenzalignment ablesbaren Wahrscheinlichkeiten von Mutationen berücksichtigen – so wird die Suche empfindlicher und genauer. Zum anderen gelang es den Bioinformatikern, die zu durchsuchende Datenmenge durch einen Vorfilter zu reduzieren, ohne dabei merklich an Empfindlichkeit einzubüßen. Der Trick: Ähnliche Datenbanksequenzen werden zunächst zu multiplen Sequenz-Alignments zusammengefasst. Deren Spalten werden jeweils durch einen von 219 „Buchstaben“ beschrieben, wobei ähnliche Spalten durch die gleichen Buchstaben repräsentiert werden. Dadurch kann jedes Alignment durch eine Sequenz aus diesen 219 Buchstaben angenähert werden, und der langsame paarweise Vergleich von Sequenz-Alignments lässt sich durch den viel schnelleren Vergleich dieser Sequenzen ersetzen, was die Suche um den Faktor 2500 beschleunigt. „Insgesamt ermöglicht HHblits, häufiger und genauer als bisher die Funktion und Struktur unbekannter Proteine vorherzusagen“, betont Söding, der die Methode nun noch weiter verbessern möchte, unter anderem durch die Einbeziehung von Strukturinformationen. (göd)

Publikation:

„HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment“;
M. Remmert, A. Biegert, A. Hauser, J. Söding;
Nature Methods, 25. 12.2011;
doi: 10.1038/NMETH.1818

Ansprechpartner:

Dr. Johannes Söding
Genzentrum München und Department of Biochemistry
Tel.: 089 / 2180 - 76742
Fax: 089 / 2180 - 76999
E-Mail: soeding@genzentrum.lmu.de
Web: www.soeding.genzentrum.lmu.de/

Kommunikation und Presse

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
[dirtscherl@lmu.de](mailto:dirscherl@lmu.de)

Infoservice:
+49 (0)89 2180 - 3423