

Andreas Bohne
Dr. sc. hum.

Ein wissensbasiertes System zur schnellen Erzeugung von 3D-Strukturen biologisch relevanter N-Glykane sowie ihrer mimikrierenden Glykoclusterstrukturen

Geboren in Alfeld (Leine)

Studiengang der Fachrichtung Informatik
Vordiplom an der Universität Hildesheim
Diplom an der Universität Hildesheim

Promotionsfach: Medizinische Biometrie und Informatik
Doktorvater: Prof. Dr. T. Wetter

Die vorliegende Arbeit beschreibt die Konzeption, Entwicklung und Implementierung eines wissensbasierten Systems zur schnellen, computergestützten Generierung der dreidimensionalen Struktur von N-Glykanen, einer biologisch besonders relevanten und strukturell sehr anspruchsvollen Klasse von komplexen Kohlenhydraten.

Das Wissen über die dreidimensionale Struktur ist von entscheidender Bedeutung bei biologisch-medizinischen Vorgängen, da Bindeprozesse nach dem Schlüssel-Schloß-Prinzip ablaufen. Ziel der Arbeit ist es, Informationen über die konformative Vielfalt der N-Glykane so aufzubereiten und zu strukturieren, daß das gesamte Verfahren so schnell abläuft, daß Benutzer bei einer Anfrage über das Web nicht länger als wenige Minuten auf die Antwort warten müssen. Die Grundlage der Wissensbasis ist die Beschreibung des flexibelsten Teils von Glykanen - der glykosidischen Bindungen - in Form von Tabellen der realistischen Torsionswinkel und deren relativer Häufigkeitspopulation. Die Torsionswinkel wurden in dieser Arbeit mittels Langzeit-Molekular-Dynamik-Simulationen bei hohen Temperaturen (1000 K) unter Einbeziehung eines Lösungsmittelterms (Wasser) in die Berechnung ermittelt. Aus früheren Arbeiten war bekannt, daß die Betrachtung der konformativen Vielfalt von Polysacchariden sehr schnell zu einer kombinatorischen Explosion der zu generierenden und zu bewertenden Strukturen führt. Aus diesem Grunde wurden die Simulationen mit größeren Substrukturen (Tri- bis Pentasaccharide) durchgeführt, die auch den Grad der Verzweigung mit einbeziehen. Dafür wurden etwa 2000 N-Glykane aus der CarbBank extrahiert. Um alle in der CarbBank enthaltenen Strukturen beschreiben zu können, waren 225 Teilstrukturen von Tri-, Tetra- und Pentasacchariden notwendig. Das konformative Verhalten aller 225 Teilstrukturen wurde simuliert, und die so erhaltenen Rohdaten im Hinblick auf die Flexibilität bezüglich der glykosidischen Bindungen analysiert und in Form von entsprechend ihrer Population gewichteten Torsionswinkeln in der Wissensbasis abgelegt.

Die Abfrage der Wissensbasis erfolgt über die Eingabe der Struktur eines N-Glykan durch den Benutzer, der Zerlegung der Struktur in die entsprechenden Tri-, Tetra- und Pentasaccharide, der Abfrage und Einstellung der abgelegten Winkelwerte in der Reihenfolge abnehmender Po-

pulation und der Berechnung des Energieinhalts der konstruierten N-Glykane. Dem Benutzer zugänglich gemacht werden die n Strukturen mit dem geringsten Energieinhalt. Ferner wird aus den so erzeugten Strukturen der Vorschlag für ein Glykocluster abgeleitet, das die räumlichen Bedingungen des entsprechenden N-Glykane abbilden kann.

Ein Vergleich der generierten räumlichen Strukturen mit verfügbaren experimentellen Strukturen, abgeleitet aus NMR- und Röntgenstruktur-Untersuchungen, ergab, daß die Schwankungsbreite der mittels des wissensbasierten Ansatzes abgeleiteten Torsionswinkel für die glykosidischen Bindungen den aus experimentellen Daten gewonnenen Werten sehr ähnliche sind.

Um das in dieser Arbeit entwickelte Verfahren zur schnellen Generierung von N-Glykanen allen interessierten Wissenschaftlern zugänglich zu machen, wurde der Algorithmus als C-Programm implementiert und kann über eine WWW-Seite mit Daten versorgt werden. Die Datenausgabe erfolgt ebenfalls über WWW-Seiten, die vom Benutzer abgerufen werden können.

Der vorgestellte Ansatz läßt sich einfach auf andere Klassen von Kohlenhydraten erweitern (O-Glykane, Lipopolysaccharide, Ganglioside etc.), deren Glykoanteil strukturell weniger komplex ist als die hier behandelten N-Glykane.

Das Programm ist als Web-Tool auf dem WWW-Server der Arbeitsgruppe „Molecular Modeling“ des Deutschen Krebsforschungszentrums unter der Web-Adresse <http://www.dkfz.de/spec/glydict/> zu erreichen.