

Sonu Kumar

Dr. sc. Hum.

Database Development and Analysis of Carbohydrate-Protein Interactions

Promotionsfach: DKFZ (Deutsch Krebsforschungszentrum)

Doktorvater: PD. Dr. Reinhard Schwartz-Albiez

Nachdem das humane Genomprojekt vollendet war, wurden für das humane Genom nur 25000 Gene insgesamt identifiziert. Es rief in der wissenschaftlichen Gemeinschaft eine gewisse Verwunderung aus, wie man aus dieser relativ kleinen Anzahl menschlicher Gene die biologische Komplexität des menschlichen Körpers erklären könne. Eine Antwort auf diese Frage liegt sicherlich darin, dass die Komplexität durch post-translationale Modifikationen der translatierten Proteine, wie zum Beispiel der Glykosylierung, mit verursacht wird. Glykoproteine können unterschiedliche Funktionen ausüben, wie zum Beispiel im Immunsystem die Erkennung von pathogenen Keimen und die Kommunikation zwischen verschiedenen Elementen des Immunapparates, aber auch die gezielte Anheftung („Homing“) der Lymphozyten an das infizierte Gewebe oder die transmembranöse Signalübertragung über glykosylierte Rezeptoren.

In der experimentellen Forschung ist durch moderne Methoden besonders im letzten Jahrzehnt eine enorme Datenfülle auch in der Glykobiologie angefallen, die einer Behandlung durch Computertechnologie hinsichtlich der Speicherung und der weiteren Verarbeitung bedarf. In diesem Zusammenhang spielen Datenbanken eine zentrale Rolle in der Bioinformatik, die dem Forscher die Möglichkeit eröffnen, auf eine große Bandbreite biologisch relevanter Daten zurückzugreifen.

In der vorliegenden Arbeit haben wir Daten, die in sogenannten Kohlenhydrat-Micro-Arrays (Glycan-Arrays) generiert wurden, in einer neuen Datenbank so zusammengefasst und strukturiert, dass sie dem Benutzer ohne allzu große Kenntnis der Computertechnik zugänglich sind und von ihm weiter verarbeitet werden können.

Um Probleme zu überwinden, die bei der Übertragung von Glycan Array Ergebnissen in die Datenbank entstehen, wie zum Beispiel das Fehlen einer allgemein anerkannt standardisierten Darstellung von Kohlenhydraten und bestimmter datentechnischer Werkzeuge, sowie unterschiedlicher Identifizierungsnummerierung der selben Glykane (IDS) in verschiedenen

Arbeitsgruppen, haben wir zunächst eine Glykan Datenbank (Glycan Database) etabliert für die Glykane, die in den Glycan Arrays verwendet werden und eine zweite für die experimentellen Glycan Array Ergebnisse (Glycan Array Database). Beide Datenbanken wurden sodann miteinander verbunden. Um eine standardisierte Darstellung der Kohlenhydrate zu ermöglichen, wurden die Kohlenhydratstrukturen in für den Rechner lesbare Formate konvertiert (GlycoCT und GlydeII) und dann in der „Glycan Database“ gespeichert. Beide Datenbanken sind nun mit einer benutzerfreundlichen Oberfläche ausgestattet, um auch verschiedene Funktionen mit den vorhandenen Array-Daten ausführen zu können, wie z.B. Suchfunktionen, interaktive Möglichkeiten zur Generierung von Affinitätsgraphiken aus vorhandenen Array Daten und Optionen zur Konvertierung von Glykanstrukturen in Computer-lesbare Formate. Beide Datenbanken sind unter EuroGlycoArrayDB (<http://glycoarray.dkfz.org/glycoarray/>) zusammengefasst.

Es wurde eine weitere Datenbank, GlycoCD, erstellt, welche Informationen zur Struktur und Funktion sogenannter CD (cluster-defined)-Antigene aus dem Glykanbereich enthält. Dies betrifft sowohl solche Antigene mit Kohlenhydratnatur als auch solche, welche Kohlenhydrate erkennen (Lektine). Als CD Antigene sind solche Zelloberflächenantigene nomenklatorisch erfasst, die von monoklonalen Antikörpern erkannt werden. Die neue Datenbank (<http://glycosciences.de/glycocd/index.php>) erlaubt einfachen Zugang zu den aktuell vorhandenen Glykan relevanten CD-Antigenen. Die Datenbank ist zudem mit anderen Datenbanken zu Kohlenhydratstrukturen und MicroArray Daten vernetzt.

In der vorliegenden Arbeit wurden außerdem an einem Beispiel gezeigt, wie Ergebnisse aus Glykan-Arrays zur dreidimensionalen Darstellung von Protein-Kohlenhydrat Interaktionen angewendet werden können. Zu diesem Zweck wurde das human Galektin-8 (Gal-8) Lektin genommen, einem Mitglied der Galektin-Proteinfamilie, welche eine hohe Affinität zur Bindung von b-Galaktosiden vereint. Das sogenannte „Tandem-Repeat“ Gal-8 besteht aus einer N- und C-terminalen Glykanbindungsregion (N-CRD und C-CRD, CRD=carbohydrate recognition domain), welche durch ein Verbindungspeptid unterschiedlicher Länge miteinander verbunden sind. Trotz ihrer grundsätzlichen strukturellen Ähnlichkeit erkennen beide Erkennungsregionen unterschiedliche Kohlenhydratstrukturen. Die molekularen Bedingungen der N-Domäne für die hoch-affine Bindung von sulfatierten und sialylierten Kohlenhydraten wurde kürzlich mittels kristallographischer Analyse des Protein-Glykan Komplexes mit mehreren unterschiedlichen Glykanen beschrieben. Dagegen wurden die Bindungsspezifitäten der C-Domäne für Oligosaccharide, die über experimentelle Ergebnisse bekannt sind, bisher nicht in Bezug auf die dreidimensionale Struktur des Gesamtkomplexes

erklärt. In dieser Arbeit wurde nun auf Grund der vorliegenden experimentellen Glykan-Array Daten und der kürzlich veröffentlichten kristallographischen Analyse der einzelnen Gal-8C-Domäne eine „molecular dynamics“ (MD) Simulation zur Aufklärung der molekularen Bedingungen für die Komplexbildung mit einzelnen Kohlenhydratsequenzen durchgeführt. Die terminale β -Galaktose der Disaccharide (LacNAII, LacNAI und Laktose) und der internen β -Galaktose der Blutgruppenantigene A und B (BGA und BGB), sowie längerer Oligosaccharidketten (polyLacNAc und Lacto-N-tetraose) interagierten mit den konservierten Aminosäuren H53, R57, N66, W73 und E76. Lacto-N-Neotetraose und PolyLacNAc, sowie BGA und BGB passten sich sehr gut in die Bindungstasche der C-Domäne ein und interagierten mit den Aminosäuren N39, D41, N130 und E128 der C-Domäne. BGA und BGB zeigten eine wesentlich höhere Affinität zur C-Domäne als LacNAc und Laktose, bedingt durch stärkere Interaktionen mittels Wasserstoffbrücken und Wasser vermittelte Wasserstoffbrücken mit der α 1-2 gebundenen Fucose.

Unsere Ergebnisse der MD Simulationen ermöglichen eine genaue molekulare Erklärung der experimentellen Ergebnisse zu den Bindungsspezifitäten der Gal8C Domäne, auch im Vergleich zu denen der N-Domäne.