

Similarity in Chemical and Protein Space: Finding novel starting points for library design

Zusammenfassung der Doktorarbeit von Stefan Wetzel

Chemische Genomik, d.h. die Beeinflussung der Aktivität von Proteinen durch organische Moleküle zur Untersuchung grundlegender biologischer Vorgänge, ist eines der zentralen Forschungsgebiete der Chemischen Biologie. Die hier vorgestellte Arbeit zielte auf die Entwicklung und Anwendung computer-gestützter Verfahren zur Entwicklung von Verbindungsbibliotheken ab. Die entwickelten Methoden basieren auf der Analyse komplementärer chemischer und biologischer Strukturräume zur Kartierung und Erforschung biologisch relevanter Teile des chemischen Strukturraumes sowie der dazugehörigen Zielproteine.

1 Kartographie und Navigation im chemischen Strukturraum

1.1 Das Baumdiagramm der chemischen Gerüststrukturen

Bei dem „Strukturgerüstbaum“ handelt es sich um eine hierarchische Klassifikation von chemischen Strukturen basierend auf chemisch sinnvollen Gerüststrukturen, d.h. den Ringsystemen, Ringverbindenden aliphatischen Ketten sowie allen Doppelbindungen, die von den genannten Strukturen ausgehen. Die schrittweise Zerlegung dieser Gerüststrukturen Ring für Ring, die durch einen in Zusammenarbeit mit Novartis entwickelten Regelsatz gesteuert wird, erzeugt einen Ast von Gerüststrukturen unterschiedlicher Größe. In dieser Hierarchie wird die kleinere Struktur als „Kind“, die größere als „Elter“ bezeichnet. Die Klassifikation vieler Moleküle erzeugt ein baum-ähnliches Gerüstdiagramm. Gerüststrukturen im selben Ast besitzen möglicherweise ähnliche Eigenschaften wie, zum Beispiel, ähnliche Bioaktivität. Der Ausdruck des „Schwinghangelns“ bezeichnet dabei die Vereinfachung von Strukturen von außen nach innen entlang eines Astes bei ähnlicher biologischer Aktivität. „Virtuelle Gerüststrukturen“ repräsentieren keine Moleküle im Datensatz.

1.2 Scaffold Hunter

Strukturbäume stellen den chemischen Strukturraum auf aussagekräftige und intuitive Weise dar. Um ihre Anwendbarkeit zu verbessern, wurde das Computerprogramm „Scaffold Hunter“ in einem gemeinsamen Studentenprojekt mit dem Lehrstuhl für Algorithmen an der Technischen Universität Dortmund entwickelt. Scaffold Hunter visualisiert automatisch die Strukturbäume ausgehend von den Daten in einer Datenbank. Es ermöglicht die Navigation in den Strukturbäumen einschließlich der Anwendung von Filtern, zoomen, Farbverläufen basierend auf molekularen Eigenschaften und der Markierung bestimmter Strukturen. Die entsprechende Datenbank der Gerüststrukturen kann leicht mit Hilfe eines zweiten Programms, des „Scaffold Tree Generator“ erstellt werden. Dieses Programm wurde von Dr. Steffen Renner entwickelt und verarbeitet SD Dateien, ein offenes Standardformat für Molekülstrukturen. Beide Programme können kostenlos unter einer Lizenz für quelloffene Software von der Webseite www.scaffoldhunter.com heruntergeladen werden.

1.3 Identifikation und Besetzung von Lücken im Chemischen Strukturraum

Wie oben beschrieben, stellen virtuellen Gerüststrukturen Lücken im chemischen Strukturraum dar, die vielversprechende Gerüststrukturen in der Nähe zu Gerüststrukturen von aktiven Verbindungen enthalten können. Eine erste Bestätigung dieser Hypothese ergab die Analyse vielversprechende virtuelle Gerüststrukturen aus den Daten in der PubChem Datenbank, die anschließend mit biologisch aktiven Substanzen aus der Literatur verglichen wurden. Zur Identifikation neuer, aktiver Verbindungen wurde der Pyruvatkinase Datensatz in PubChem analysiert. Von 65 vielversprechenden virtuellen Gerüststrukturen wurden vier ausgewählt und Verbindungsbibliotheken basierend auf diesen Strukturen gekauft. Das Testen dieser Verbindungen auf die Beeinflussung der Aktivität von Pyruvatkinase identifizierte acht potente Substanzen, von welchen keine gemäß einer SciFinder Recherche bis dato mit Pyruvatkinase in Verbindung gebracht worden war.

1.4 Die Nutzung von Naturstoffen: γ -Pyrone

Eine offensichtliche Anwendung der Strukturbäume ist der Vergleich verschiedener Verbindungsbibliotheken durch Überlagerung ihrer Strukturbäume. Auf diesem Weg kann eine Bibliotheken mit Eigenschaften einer anderen annotieren werden, beispielsweise mit Information über mögliche Zielproteine. In einer ersten Anwendung wurde der chemische Strukturraum der Naturstoffe mit den Literaturinformationen über Zielproteine aus der WOMBAT Datenbank annotiert. Der Ast der γ -Pyrone zeigte vielversprechende biologische Aktivitäten und war synthetisch zugänglich. Eine Substanzbibliothek höherer γ -Pyrone (mit Gerüststrukturen aus 2-4 Ringen) wurde zusammengestellt und auf Aktivität gegen die Monoamine Oxidasen (MAO) A und B, die signal transducers and activators of transcription (STAT) Proteine sowie die saure Sphingomyelinase getestet. Diese Zielproteine wurden ausgewählt, da für sie Aktivität für mindestens zwei Gerüsttypen des γ -Pyron Astes beschrieben war. Insgesamt identifizierten die Tests eine große Zahl an Inhibitoren mit bemerkenswerter Selektivität – bis zur Isoenzympezifität.

1.5 Ausblick

Dieser Abschnitt beschreibt zukünftige Erweiterungen von Scaffold Hunter, unter anderem Methoden zur Generierung von Strukturbäumen mittels individuell angepasster Regelsätze, einschließlich Biologie-basierter Strukturbäume, die von S. Renner entwickelt wurden. Schließlich wird eine mögliche Erweiterung zur Erzeugung chemisch sinnvoller Naturstofffragmente vorgeschlagen. Solche Fragmente könnten die Nutzung der chemischen Diversität in der Natur mit einem begrenzten synthetischen Aufwand ermöglichen.

2 Erforschung des Proteinstrukturraumes – Proteinstrukturähnlichkeitsclustering

PSSC basiert auf dem Konzept, dass Proteine mit strukturell ähnlichen Bindungsstellen auch ähnlich Substanzen binden. Daher kann die strukturelle Ähnlichkeit genutzt werden, um Gruppen von Zielproteinen zu identifizieren, deren Aktivität mit höherer Wahrscheinlichkeit von Substanzen mit ähnlichen Strukturmotiven beeinflusst wird.

2.1 Methodenentwicklung – automatisiertes PSSC

Der neu entwickelte PSSC Prozeß basiert auf dem strukturellen Vergleich sogenannter Ligandenbindender Kernstrukturen (LBK), d.h. kugelförmigen Ausschnitten aus der Struktur um die Bindungstasche herum, nicht mehr auf ganzen Proteinstrukturen. Daraus resultieren deutlich weniger falsch positive Ergebnisse. Die Kriterien für die Erzeugung der LBKs wurden ausgehend von einer Analyse der katalytischen Zentren im Catalytic Site Atlas (CSA) entwickelt. Ein eigens entwickeltes Computerprogramm extrahierte die LBKs aus den Strukturdaten in der PDB basierend auf der annotierten Bindungstasche aus dem CSA. Die nachfolgenden Strukturüberlagerungen wurden mit DaliLite berechnet, wodurch die Ergebnisse mit denen frühere Analysen vergleichbar sind, da die verwendete FSSP Datenbank ebenfalls auf Dali beruht. Die Ergebnisse des Strukturvergleichs wurden mit einer angepaßten Implementierung des OptiSim Clustering Algorithmus prozessiert. Der Vergleich mit den Daten der SCOP Datenbank, des Goldstandards bezüglich struktureller Ähnlichkeit von Proteinen, ergab eine große Einstimmung der gefundenen Cluster mit der in SCOP definierten Ähnlichkeit. Für 33 Cluster sagte die SCOP Datenbank weniger als die Hälfte aller Cluster Mitglieder vorher. Zur experimentellen Validierung wurden zwei Zielproteine eines Clusters, die Pyruvatkinase (PK) und die Dihydropteroat Synthetase (DHPS) ausgewählt. Eine Bibliothek von 740 Sulfanilamiden, einer bekannten Klasse von DHPS Inhibitoren, wurde auf PK Inhibition getestet. Der Test identifizierte jedoch keine aktiven Moleküle, was einerseits an den gewählten Testbedingungen liegen könnte aber auch an eine eingeschränkten Anwendbarkeit des PSSC Konzept in diesem Fall. Eine abschließende Betrachtung der Anwendbarkeit von PSSC als Konzept ist daher auf Grund dieser Ergebnisse nicht möglich.

Zur PSSC Analyse großer Zahlen von Strukturen wurde die Entwicklung einer schnelleren Methode zum Strukturvergleich basiert auf „protein fingerprints“ begonnen. Kreuzvalidierung zwischen den Ergebnissen dieser Methode und denen von Dali für den Datensatz von 15.000 LBKs zeigte keine gute Übereinstimmung, so dass weiteres Verbesserungspotenzial besteht.

2.1.1 Methodenentwicklung –PSSC mit dynamisierten Proteinstrukturen

Ein wesentlicher Nachteil von PSSC ist die benötigte Struktur des zu untersuchenden Proteins. Obwohl gezeigt wurde, dass PSSC auch mit Homologie-Modellen funktioniert, können Ligand-induzierte Konformationsänderungen („induced fit“) ein Problem darstellen. Zusammen mit B.D. Charette aus Professor Berkowitz Gruppe in Lincoln, Nebraska konnte erfolgreich gezeigt werden, dass Molekulardynamiksimulationen den Übergang von der apo zur gebundenen Struktur berechnen und so die PSSC relevante gebundene Struktur simulieren können.

3 Verschiedene Projekte

Im Rahmen dieser Arbeit wurden auch einige andere Projekte unterstützt wie z.B. die NMR-unterstützte Konformationsanalyse mittels Kraftfeldrechnungen von Makrocyclen und Docking Simulationen von Inhibitorstrukturen zur Erklärung der experimentell beobachteten Selektivitäten. Weitere Arbeiten befaßten sich mit der Auswertung, statistischen Analyse und Qualitätskontrolle von biochemischen Tests sowie der nachfolgenden Optimierung des Experiments.