Aus der Medizinischen Klinik
mit Schwerpunkt Hämatologie und Onkologie

der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Xenobiotika in der Biomedizin: Wirkung und Stoffwechsel

zur Erlangung des akademischen Grades
Doctor medicinae dentariae (Dr. med. dent.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Saskia Maria Preißner

aus Berlin

Gutachter: 1. Prof. Dr. med. P. T. Daniel

2. Prof. Dr. P. Wrede

3. Priv.-Doz. Dr. med. C. Gärtner

Datum der Promotion: 9.9.2011

Inhalt

ZUSAMMENFASSUNG DER PUBLIKATIONSPROMOTION

Xenobiotika in der Biomedizin: Wirkung und Stoffwechsel

von Saskia Preißner
Arbeitsgruppe Prof. Peter Daniel
Medizinische Klinik mit Schwerpunkt Hämatologie, Onkologie und Tumoronkologie
Charité – Universitätsmedizin Berlin

Zusammenfassung

Abstract

Als Xenobiotika werden Substanzen betrachtet, die nicht durch den Stoffwechsel des menschlichen Körpers synthetisiert werden, sondern beispielsweise als Medikamente oder Lebensmittelzusätze aufgenommen werden. Diese Arbeit beschäftigt sich mit drei Aspekten der Xenobiotika: Metabolismus, Interaktionen mit Zielmolekülen der Tumortherapie und einer speziellen Rezeptorbindung.

Obwohl der größere Teil der Medikamente von Naturstoffen abgeleitet ist, stellt der Abbau dieser Substanzen in der Regel ein Problem für den menschlichen Organismus dar. Die Familie der Cytochrom P450 Enzyme umfasst 57 Vertreter, die für die Oxygenierung der Xenobiotika sorgen, wodurch über eine folgende Konjugation die Ausscheidung ermöglicht wird. Einen Zusammenhang zwischen Struktur eines Medikaments und dem abbauenden Cytochrom herzustellen, ist ein schwieriger, aktueller Forschungsgegenstand. Um diesem Ziel näher zu kommen, mussten zunächst alle gegenwärtig verfügbaren Informationen zum Stoffwechsel von Medikamenten über Textmining – einer spezifischen, automatisierten Analyse von online verfügbaren Literaturdatenbanken – gesammelt und in einer Datenbank mit Web-Schnittstelle zugänglich gemacht werden. Die in diesem Zusammenhang gefundenen Informationen zu häufigen Mutationen wurden auf die selbst erstellten Strukturmodellen projiziert, um ein Hilfsmittel zu schaffen, das dem Mediziner eine Optimierung der Medikamentierung bzgl. möglicher Probleme – insbesondere bei multimorbiden Patienten – ermöglicht.

Für die optimale Wirkung von Medikamenten ist neben dem Metabolismus von entscheidender Bedeutung, in welche Signalwege die jeweiligen Zielmoleküle involviert sind und wo diese Zielmoleküle exprimiert werden. Deshalb ist es von großer Bedeutung, eine umfassende Aufstellung von Interaktionen zwischen Medikamenten und ihren Zielmolekülen vorzunehmen, was wiederum nur durch automatisches Textmining möglich ist. Diese Relationen auf biochemische Pathways zu projizieren und mit Expressionsdaten verschiedener Gewebe zu kombinieren, ermöglicht eine rationalere Therapie. Im Kontext mit Strukturmodellen für die Zielmoleküle, Mutationsdaten und der

Chemosensitivität der verschiedenen Tumortypen wurde ein universelles Werkzeug geschaffen, das einen ersten Schritt in Richtung individualisierte Krebstherapie ermöglicht.

Xenobiotika sind wichtige Komponenten verschiedener Lebensmittel. Bestimmte Nahrungsbestandteile werden vom menschlichen Körper bereits bei der Aufnahme als angenehm oder unangenehm empfunden, was evolutionär wichtig war, um giftige oder kohlenhydratreiche Nahrungsmittel schnell zu identifizieren. Heute spielen Zucker und Süßstoffe in der Zahnmedizin eine wichtige Rolle, da sie extrem kariogen, aber auch kariesprotektiv wirken können. Durch die Verwendung von Süßstoffen können Diabetiker Nahrungsmittel nach Belieben süßen und dabei trotzdem ihre Diät einhalten. Erstaunlicherweise gab es bisher keine online verfügbare, umfassende Sammlung von Informationen über natürliche und künstliche Süßstoffe. Der Fokus dieser Arbeit lag deshalb auf der Zusammenstellung weiter reichender Informationen wie Süßkraft, Nährwert, physikochemische Eigenschaften, 3D Strukturen und therapeutischer Effekt. Mit Hilfe dieser zusammengetragenen Informationen kann man Ähnlichkeitssuchen durchführen und neue, kalorienfreie, Karies inhibierende Zucker entdecken.

1. Einleitung

Cytochrom P450

Die Familie der Cytochrom P450 Enzyme (CYPs) befindet sich seit Jahrzehnten im Fokus der pharmazeutischen Forschung, was durch mehr als 100.000 Artikel in PubMed eindrucksvoll belegt wird. Der Metabolismus der meisten Medikamente und Xenobiotika wird durch CYPs realisiert.

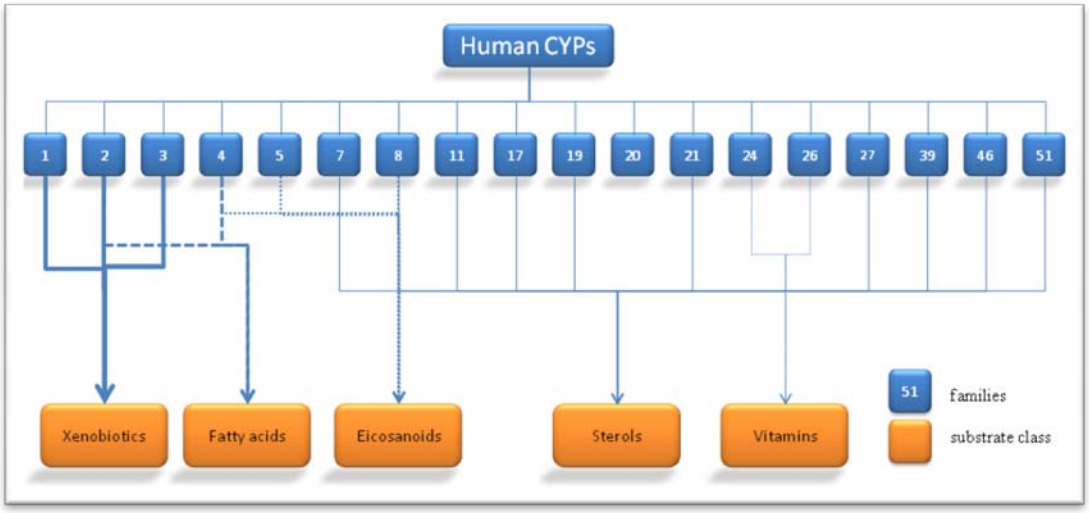


Abbildung 1 - 18 humane CYP Familien. Die Pfeile führen zu den Haupt-Substratklassen

Im Rahmen des Human Genome Project (1) konnten 57 CYPs identifiziert werden, die diverse Aufgaben im Medikamentenstoffwechsel übernehmen. Einige Medikamente haben einen inhibitorischen oder induktiven Effekt auf bestimmte CYPs, was von großer Bedeutung für die Dosierung und Kombination unterschiedlicher Medikamente miteinander ist.

Kombiniert man zum Beispiel ein Medikament, das induktiv auf ein CYP wirkt, mit einem weiteren Medikament, das ebenfalls über dieses CYP abgebaut wird, so muss die Dosis beider Medikamente erhöht werden, um den gleichen therapeutischen Effekt zu erzielen, denn die Medikamente werden schneller eliminiert als in einer Monotherapie.

Bei der Kombination eines Inhibitors mit einem Substrat hingegen, sollte die Dosis niedriger gewählt werden, weil die Medikamente länger im Organismus verbleiben, als in Monotherapie. Ohne die Kenntnis dieser Zusammenhänge, kann es zu schwerwiegenden Nebenwirkungen kommen.

Die Informationen über den Metabolismus sind über das Internet und die wissenschaftliche Literatur verstreut und damit nicht direkt für die individuelle Medikamentierung abrufbar.

Die Arbeitsgruppe um David Nelson hat humane CYPs aufgelistet, mit Alignments versehen und eine Nomenklatur erarbeitet (2). Flockhart erstellte eine Tabelle mit typischen CYP-Medikamenten-Interaktionen (3), die unerwünschten Nebenwirkungen bei der Kombination mehrerer Medikamente vorbeugen soll, denn besonders in einer alternden Gesellschaft ist Übermedikamentierung ein Thema von zentraler Bedeutung. Rendic und DiCarlo (4,5) stellten eine umfassende Sammlung über typische Reaktionen der CYPs und daraus folgenden Medikamenteninteraktionen zusammen.

Neben der gegenseitigen Beeinflussung der Medikamente untereinander, können auch bestimmte Nahrungsmittel und Getränke die Aktivität der CYPs beeinflussen. Solche Informationen hat die University of Maryland für ihre Patienten anschaulich zusammengefasst (http://www.umm.edu/adam/drug_checker.htm).

Ein weiteres wichtiges Forschungsfeld ist die interindividuelle Diversität der CYPs in verschiedenen ethnischen Gruppen. Aus pharmakogenomischen Studien weiß man beispielsweise, dass afro-amerikanische Patienten höhere Dosen Antidepressiva benötigen, als Kaukasier (6). Zu den 2.000 bekannten Mutationen gibt es vielfältige Informationen in Publikationen und im Internet. Trotzdem scheitern klinische Studien häufig an bis dahin nicht bekannten ethnischen Besonderheiten im Stoffwechsel der betreffenden Substanzen.

Alle derzeit verfügbaren 3D-Strukturen der Protein Data Bank (7) werden in ‚CYPED' (8,9) präsentiert und regelmäßig aktualisiert.

Das Center for Molecular Design sammelte Informationen zu den Sequenzen von CYPs verschiedener Spezies, die in ihrer ‚Knowledgebase' zusammengefasst sind (http://cpd.ibmh.msk.su).

Trotz der guten Informationslage ist es nach wie vor schwierig für Ärzte, Medikamente in Hinblick auf ihren Metabolismus aufeinander abzustimmen und so Wechselwirkungen zu minimieren.

Cancer Resource

Die multiplen Interaktionen zwischen Medikamenten und ihren Zielmolekülen werden immer interessanter für verschiedene Bereiche der Forschung und Medizin. Daher haben sich verschiedene Datenbanken etabliert (21) und es ist schwierig, die richtigen Informationen überhaupt zu finden. CancerResource widmet sich der Komplexität von Tumorerkrankungen, indem Informationen über sämtliche Medikamenten-Zielmolekül-Interaktionen, experimentelle Daten und unterstützende Informationen zusammengetragen wurden. Desweiteren ist es nun möglich, eigene Daten (Patientendaten) hochzuladen und weitere Analysen durchzuführen, um die individuellen Eigenschaften des jeweiligen Tumors bei der Therapie berücksichtigen zu können.

Süßstoffe

Zucker war zu Zeiten schwieriger Importbedingungen ein Luxusartikel und schon früh beschäftigten sich Wissenschaftler mit der Synthese von künstlichen Süßstoffen. Constantin Fahlberg war es schließlich, der 1885 den ersten künstlichen Süßstoff synthetisieren konnte (10). Saccharin war nicht nur extrem süß, sondern auch für ärmere Gesellschaftsschichten erschwinglich. Es begann ein regelrechter Krieg zwischen Zucker- und Süßstoffindustrie (11). Dies könnte einer der Gründe dafür sein, dass behauptet wurde, Saccharin sei kanzerogen (12). Erst viel später konnte bewiesen werden, dass es in normalen Mengen weder toxisch noch kanzerogen ist (13). Heutzutage ist der Austausch von Zucker und anderen Kohlenhydraten durch künstliche Süßstoffe gebräuchlich (14), um Karies vorzubeugen (15,16) und Diäten zu unterstützen (17,18). Der Rezeptor für süßen Geschmack ist ein Heterodimer aus zwei Transmembranproteinen und hat verschiedene Bindungsstellen. Bisher konnte er nicht kristallisiert werden und ist daher nicht in der Protein Data Bank (7) vorhanden. Seine Struktur ist hoch interessant, denn sowohl kleine Kohlenhydratstrukturen, als auch Moleküle, die die Größe von Proteinen besitzen, binden an den Rezeptor und aktivieren ihn (19). Virtuelle Modelle können helfen, die verschiedenen Bindungsmechanismen zu simulieren und so zu verstehen, wie süßer Geschmack erzeugt wird (20) und basierend auf diesem Wissen neue Süßstoffe zu entwickeln.

Zielstellung

Analyse und Vermeidung von Medikamenten-Interaktionen

Ziel des Forschungsprojektes war es, eine Cytochrom P450-Plattform für Wissenschaftler und Ärzte aufzubauen. Dabei ist es wichtig, dass Wissenschaftler aktuelle Publikationen und Zulassungsinformationen nachlesen, aber auch vorhandene Modelle oder Sequenzen herunterladen können. Desweiteren sollte es die Möglichkeit geben, aktuelle Forschungsergebnisse online zu stellen und sich mit anderen Wissenschaftlern und Ärzten auszutauschen. In die Datenbank sollten zur besseren Nutzbarkeit für Ärzte Handelsnamen und verschiedene Darreichungsformen und Dosierungen integriert werden, um eine klinisch relevante Aussage zu Interaktionen der Medikamente miteinander treffen zu können. Auch Alter, Größe, Gewicht, Geschlecht, Ethnizität und Begleiterkrankungen des Patienten spielen beim Medikamentenmetabolismus eine entscheidende Rolle. Die Patientendaten aus den CYP-Chips (Genotypisierung) sollten ebenfalls in der Datenbank gesammelt werden, um von den wichtigsten CYPs eine genaue Vorstellung ihrer Diversität und der damit verbundenen Folgen für den Medikamentenstoffwechsel zu haben. Die bereits entdeckten Mutationen sollten nach Häufigkeit und Ethnizität sortiert werden und beim Nachschlagen einzelner Medikamente als Warnung angezeigt werden, sodass der verschreibende Arzt einen Anhaltspunkt hat, ob eine Bestimmung der CYPs sinnvoll wäre. Insbesondere die weit verbreitete Polypragmasie (Kombination diverser Medikamente) erfordert einen wissensbasierten, rationalen Ansatz für die konzertierte Therapie der zunehmenden Multimorbidität. Im Mittelpunkt steht letztendlich der Patient, der vor Übermedikamentierung, Unwirksamkeit und gefährlichen Wechselwirkungen durch die Ergebnisse der Datensammlung dieses Projektes bewahrt werden soll.

Individualisierte Krebstherapie

Es liegen zunehmend individuelle Patientendaten zur Expression oder sogar zum Genom des Tumorgewebes vor. Bisher ist es sehr schwierig, aus der Fülle dieser Informationen konkrete therapeutische Schlussfolgerungen zu ziehen. Die Zielstellung besteht also darin, mit einer Web-basierten Applikation automatisiert folgende drei Schritte zu realisieren:

1. für die Expressionsdaten eines Tumors die ähnlichste Tumorzelllinie ermitteln,
2. die effizienteste Therapie für diese Zelllinie ermitteln,
3. individuelle Veränderungen der Zielmoleküle (Knockouts, Mutationen) und ihre funktionelle Auswirkung für die Therapie abschätzen.

Datenbank SuperSweet

Ziel des Projektes war es, alle verfügbaren Informationen über Kohlenhydrate und Süßstoffe in einer Datenbank zu vereinen und ausgehend von dieser Plattform weitere Forschung über neue Süßstoffe und die verschiedenen Rezeptorbindungsmechanismen (Abbildung 2) für die wissenschaftliche Gemeinschaft zu ermöglichen.

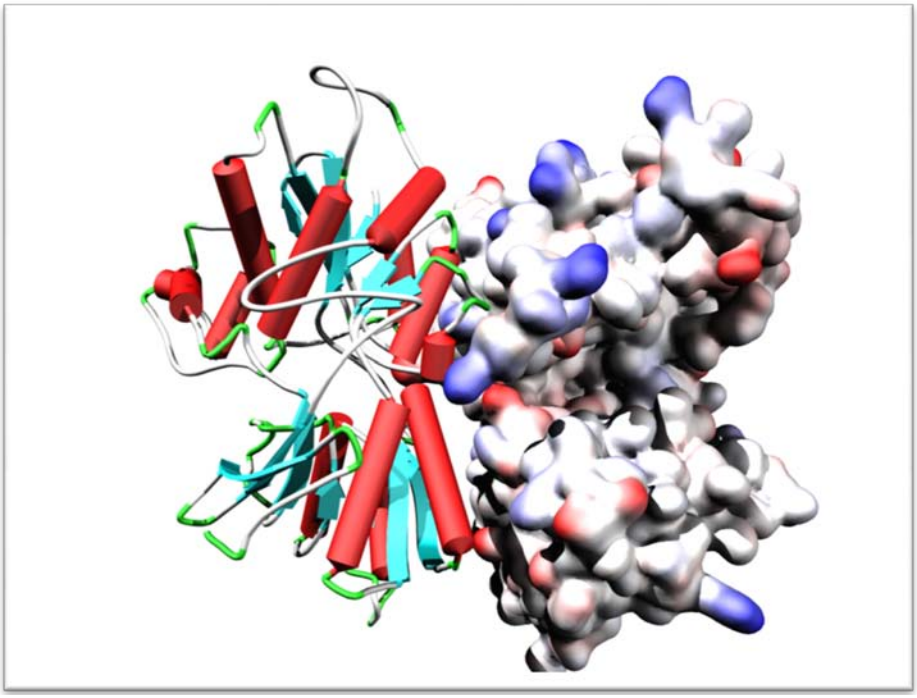


Abbildung 2 *Homology-modellierter Süßstoffrezeptor. Dimer aus je 2 Domänen.*

2. Methoden

Textmining

Zu Beginn eines jeden Projektes wurde zunächst ein Textmining durchgeführt.

SuperCYP

Hierbei wurde nach Medikamenten und deren Synonymen in Verbindung mit einem oder mehreren der 57 CYPs gesucht. Die Ergebnisse wurden manuell bearbeitet. Zunächst wurden die Abstracts nach Relevanz sortiert, bei Bedarf wird der Volltext überprüft. Den Medikamenten wurden die CYPs nach Interaktion (Substrat, Inhibitor, Induktor) zugeordnet. Anschließend wurden die neuen Daten in der Datenbank implementiert. Insgesamt wurden 1170 Medikamente mit über 3800 Interaktionen gefunden.

Ein weiteres Textmining wurde durchgeführt, um die verschiedenen bekannten Polymorphismen und Mutationen an den einzelnen CYPs zu finden. Hierfür wurden die Informationen zum Allel, der Änderung an den Nukleotiden, dem getesteten Medikament, der Testmethode, häufig betroffenen Populationen und der Konsequenz auf die Aktivität des Enzyms. Auf diese Weise wurden über 2000 Polymorphismen gefunden. Da sich die Datenlage durch ständige Forschung und Einführung neuer Medikamente häufig ändert, wurde ein Aktualisierungsintervall von 6 Monaten gewählt.

SuperSweet

Verschiedene Datenbanken wie PubChem (22), die PDB (7) und die Monosaccharid-Datenbank (http://www.monosaccharidedb.org/) wurden genutzt, um einen Basisdatensatz von Zuckern zu erstellen. Die Süßstoffe wurden größtenteils durch manuelle Literaturrecherche gefunden, da hierzu keine Datenbanken existierten. Auf dieser Grundlage wurde nun automatisiert nach Namen, Synonymen und Strukturen sämtlicher Zucker und Süßstoffe gesucht, um weitere Daten zu sammeln. Auf diese Weise wurden mehr als 8000 süße Moleküle gefunden. Für weitere Ähnlichkeitssuchen, wurden die Strukturen nach chemischen Eigenschaften geclustert und eine Gruppierung vorgenommen. Tabelle 1 zeigt diese Gruppierung innerhalb der Süßstoffdatenbank mit einzelnen Beispielen.

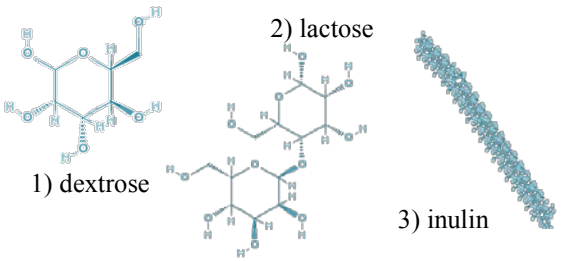
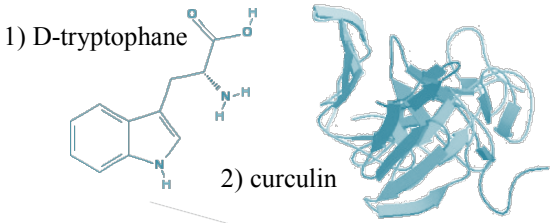
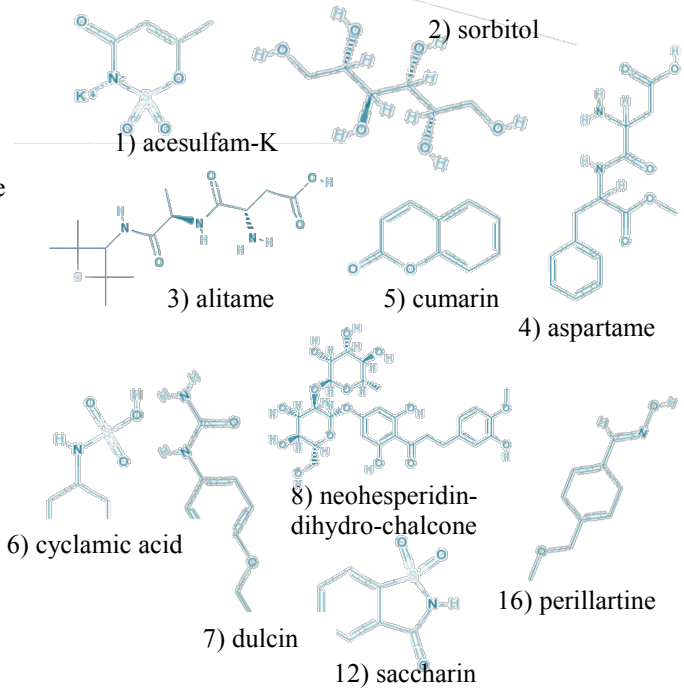
Hauptklasse	Subklasse	Beispiele
Kohlenhydrate	1) monosaccharides 2) disaccharides 3) polysaccharide 4) sugar alcohols	
Peptide	1) proteins 2) amino acids	
kleine Moleküle	1) acesulfames 2) alditols 3) alitames 4) aspartames 5) cumarins 6) cyclamat-like 7) dulcines 8) flavonoids 9) guanidine 10) carbonate 11) nucleotides 12) saccharines 13) saponins 14) steviol 15) glycosides 16) terpens 17) others	

Tabelle 1 - Organisation der Süßstoffe nach chemischer Struktur. In der ersten Spalte finden sich drei Hauptklassen, in der zweiten Spalte sämtliche Subklassen und in der dritten Spalte gibt es einige ausgewählte Strukturen von Zuckern und Süßstoffen.

CancerResource

In den letzten Jahrzehnten wurde die Krebsforschung durch viele einzelne Studien vorangetrieben. Zahlreiche Datenbanken haben es sich bereits zur Aufgabe gemacht, diese Informationen zusammenzutragen (23-26). Nach Inspektion dieser Datenbanken konnte festgestellt werden, dass diese nur wenige Übereinstimmungen aufwiesen, sodass ein Zusammentragen dieser Daten sinnvoll war. Das durchgeführte Textmining beinhaltete zunächst die

Suche nach Medikamenten und deren Synonymen im Zusammenhang mit Krebs. Einige Filtermaßnahmen ergaben letztendlich 8000 Publikationen, die für die Datenbank manuell nach Relevanz evaluiert wurden.

Nahrungsmittel

Es ist bekannt, dass verschiedene Nahrungsmittel einen Einfluss auf CYPs haben. Broccoli und Grillfleisch induzieren das CYP 1A2, während die Sternfrucht und Grapefruit Inhibitoren von CYP 3A4 sind. Durch solche weithin unbekannten, meist zufällig entdeckten Wechselwirkungen kann die Verweildauer der Medikamente im Organismus entweder stark verlängert oder extrem verkürzt werden, was wiederum zu gefährlichen Nebenwirkungen führen könnte. Das Wissen darum kann man sich aber auch zu Nutze machen, indem man Kombinationen teurer oder schwer verträglicher Medikamente niedriger dosiert und auf kritische Nahrungsmittel hinweist. Daher wurden Daten über Nahrungsmittel-CYP-Interaktionen über ein weiteres Textmining-Tool erhoben, manuell aufbereitet und in die Datenbank integriert, wodurch eine hohe Praxisrelevanz entsteht.

Medikamenten-Cocktail Optimierung

Die WHO-Klassifikation der Medikamente ermöglicht eine Sortierung nach Indikationsgruppen. Viele Patienten nehmen mehr als drei Medikamente täglich ein, die meist über gleiche CYPs abgebaut werden. Dadurch kann nicht mehr gewährleistet werden, dass die Medikamente die gewünschte Wirkung zeigen. Es kann außerdem zu Unwirksamkeit oder ernsthaften Nebenwirkungen kommen. Durch geschickte Kombination von Medikamenten, die über unterschiedliche CYPs abgebaut werden, kann dies verhindert werden. Während die theoretische Optimierung der Medikamente nach WHO-Gruppen bereits bewerkstelligt wurde, muss eine klinische Optimierung noch vorgenommen werden. Es soll am Ende möglich sein, typische Medikamente für Krankheitsbilder wie Bluthochdruck, Herzinsuffizienz und Diabetes miteinander zu kombinieren und dabei möglichst unterschiedliche CYPs zu nutzen. Abbildung 3 und 4 zeigen,

welche CYPs bei gleichzeitig verabreichten Medikamenten überlastet werden und welche medikamentösen Alternativen es gibt.

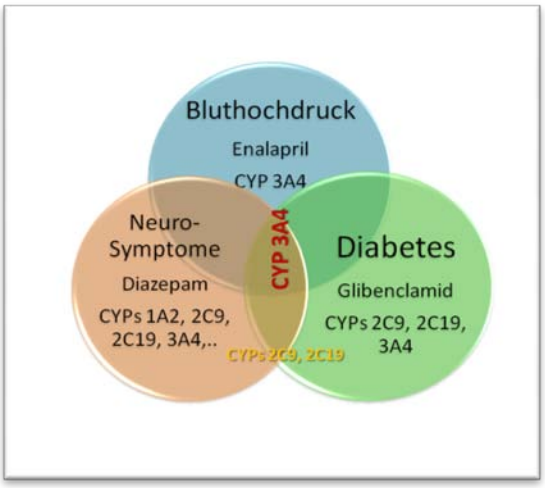


Abbildung 3 *Typische Medikamentierung bei Diabetikern mit neurologischen Problemen. Kommt Bluthochdruck hinzu, gibt es ernste Arzneimittel-Wechselwirkungen, wenn der Metabolismus außer Acht gelassen wird.*

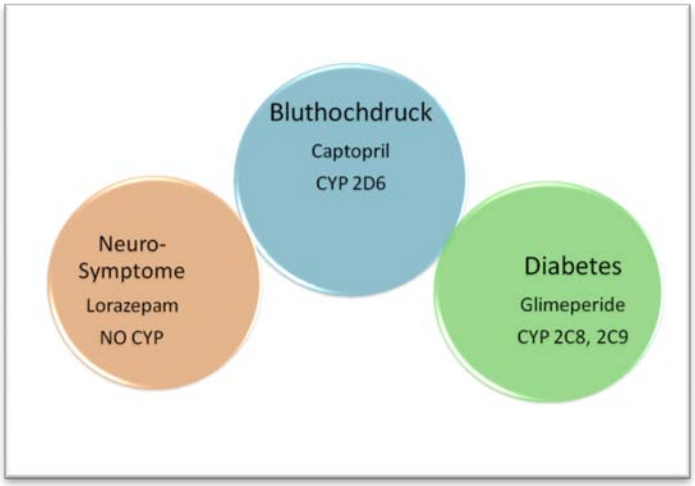


Abbildung 4 *Durch eine vergleichende Analyse innerhalb der jeweiligen Indikationsgruppen lassen sich Alternativen finden, die unerwünschte Arzneimittel-Interaktionen vermeiden helfen.*

CYP-Genotypisierung

Die individuelle CYP-Typisierung wird zwar noch erprobt und bisher nur in seltenen Fällen von den Krankenkassen übernommen, dies wird sich aber aller Voraussicht nach mit dem wachsenden Wissen über die vielen Interaktionen und das große Einsparungspotential bei der Dosierung in naher Zukunft ändern. In der Psychiatrie wird ein Chip für das CYP 2D6 bereits häufig verwendet, da die häufig verschriebenen trizyklische Antidepressiva ausnahmslos über dieses CYP abgebaut werden. Hinzu kommt, dass vom CYP 2D6 über 70 verschiedene

Varianten mit großen Unterschieden in der Metabolisierungpotenz bekannt sind. Tabelle 2 zeigt Ergebnisse der DNA-Microarray Genotypisierung von 4532 psychiatrischen Patienten aus drei staatlichen Krankenhäusern in Kentucky (27-30).

CYP	Genotyp	Häufigkeit
2D6	EM	78.8%
	IM	12.1%
	PM	7.6%
	UM	1.5%

Tabelle 2 *Ergebnisse der CYP 2D6 Genotypisierung (EM – extensive, IM – intermediate, PM – poor, UM - ultrarapid metabolizer).*

Hier konnte gezeigt werden, dass 7,6% aller Patienten „poor metabolizer“ waren. Bei diesen Patienten waren mindestens zwei der 27 Allele, die mit dem CYP 2D6 in Verbindung gebracht werden, inaktiv. 1,5% aller Patienten waren „ultrarapid metabolizer“ mit hoher enzymatischer Aktivität des CYPs. Diese Klassifizierung erfordert eine Anpassung der Dosis der entsprechenden Medikamente je nach Metabolisierungstyp. In der SuperCYP Datenbank wurden alle bekannten Polymorphismen und Mutationen zusammengetragen, sodass Kliniker die Möglichkeit haben, sich die Häufigkeit bestimmter Mutationen an den CYPs online anzusehen und gegebenenfalls eine Genotypisierung anzuordnen.

Homology Modelling

Generell kann man davon ausgehen, dass die globale Faltung von verwandten (homologen) Proteinen erhalten bleibt. Auf dieser Basis wird für nicht bekannte Proteinstrukturen durch Austausch der Seitenketten ein erstes Modell generiert. Sollten im Alignment des bekannten und unbekannten Proteins Lücken entstehen, wird mittels Spezial-Software (31) die entsprechende Region ersetzt.

CYP-Enzyme

Obwohl die Gruppe der Cytochrom P450 Enzyme eine zentrale Rolle im Medikamentenstoffwechsel einnimmt, wurden längst nicht alle Strukturen kristallografisch entschlüsselt. Tabelle 3 zeigt alle momentan verfügbaren CYP-Strukturen der Protein Data Bank. Die fehlenden Strukturen wurden mit Hilfe des Homology Modelling gebaut. Diese theoretischen Modelle entstehen durch die

Suche nach ähnlichen Aminosäuresequenzen, deren Sekundär- und Tertiärstruktur bereits experimentell bestimmt wurden. Teilweise wurden die Enzyme aus verschiedenen Elementen manuell zusammengesetzt.

CYP	PDB-ID
1A2	2hi4
2A6	1z10, 1z11, 1fdu, 2fdv, 2fdw, 2fdy, 2pg5, 2pg6, 2pg7
2A13	2p85
2C8	1pq2, 2nni, 2vn0
2C9	1og2, 1og5, 1r9o
2D6	2f9q
2R1	3c6g
3A4	1w0e, 1w0f, 1w0g, 1tqn, 2j0d, 2v0m
8A1	2iag

Tabelle 3 *– kristallisierte 3D Strukturen der CYP-Enzyme der PDB*

Sweet receptor

Der Rezeptor für süßen Geschmack wurde mit Hilfe von Homology Modelling gebaut. Hierfür wurde das Template des metabotrophischen Glutamatrezeptors (mGluR1) genutzt und mit Hilfe des Programms Modeller modifiziert. Mit MUSCLE wurde ein multiples Sequenzalignment erstellt, das auf der Website heruntergeladen werden kann.

Zielmoleküle

Inzwischen sind etwa 70.000 Strukturen in der Proteindatenbank strukturell aufgeklärt (7). Membranproteine, wie der Sweet receptor, sind dabei unterrepräsentiert, aber die meisten Faltungen globulärer Proteine sind bekannt, sodass man etwa 30% der Zielmoleküle in der PDB findet oder ein gutes Arbeitsmodell auf der Basis von Ähnlichkeitsmodellierung generieren kann.

Ähnlichkeitssuchen von organischen Kleinstrukturen

Um neue potenzielle Süßstoffe zu detektieren, wurden 2D Ähnlichkeitssuchen durchgeführt. Hierfür wurden die strukturellen Fingerprints (Bit-Vektoren, die chemische Eigenschaften von Molekülen kodiert) (32) der Moleküle automatisiert jeweils paarweise miteinander verglichen. Als Maß für die Ähnlichkeit wurde der Tanimoto-Koeffizient herangezogen, der ausdrückt, wie viele chemisch-topologische Eigenschaften übereinstimmen bzw. abweichen.

3. Ergebnisse

Datenbank zum Medikamenten-Stoffwechsel

Daten über CYPs wurden aus der wissenschaftlichen Literatur und verschiedenen Web-Ressourcen extrahiert. PubMed Abstracts wurden über eine automatische Suche nach relevanten Artikeln mit Hilfe spezieller Schlüsselwörter herausgesucht. Gesucht wurde weiterhin nach WHO-Medikamenten und deren Synonymen. Anschließend wurden die gefundenen Publikationen von einem Team von Wissenschaftlern manuell nach Informationen zum Metabolismus durchsucht. Jedem Medikament wurden CYPs zugeordnet, die im Metabolismus eine Rolle als Substrat, Induktor oder Inhibitor spielen und die zugehörige PubMed-Referenz wurde integriert. Desweiteren wurden CYP-Mutationen auf dem Protein-Level mit Hilfe des Textmining-Tools von Winnenburg und Schroeder (33) herausgesucht und nach Einfluss auf den Medikamentenstoffwechsel sortiert. Die Daten wurden in der SuperCYP Database (34) implementiert und Anwender-freundlich aufbereitet. Es ist möglich, nach einzelnen Medikamenten und ihrem Stoffwechsel zu suchen. Man kann aber auch mehrere Medikamente zusammenstellen und überprüfen, ob und wie diese sich potenziell gegenseitig beeinflussen. Außerdem kann man beispielsweise alle Medikamente anwählen, die von einem einzelnen CYP abgebaut werden. Zudem wurden auf Basis der bekannten CYP-Strukturen (Abbildung 5) theoretische 3D-Modelle der 48 fehlenden CYPs gebaut, die man für vergleichende Analysen nutzen kann. Es ist auch möglich, einzelne CYPs auszuwählen und zu überlagern, um Ähnlichkeiten und Unterschiede herauszufiltern.

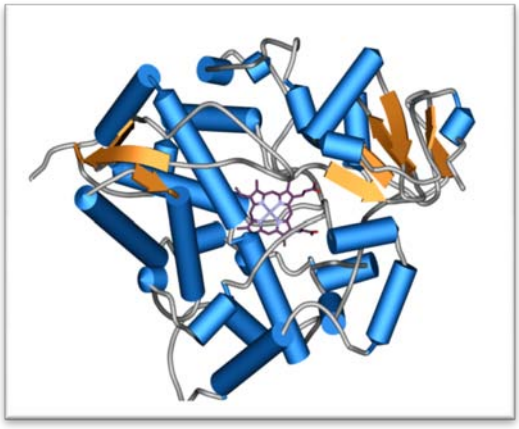


Abbildung 5 *3D-Modell des CYPs 3A4. Helices als Zylinder, β-Faltblätter als Pfeile dargestellt.*

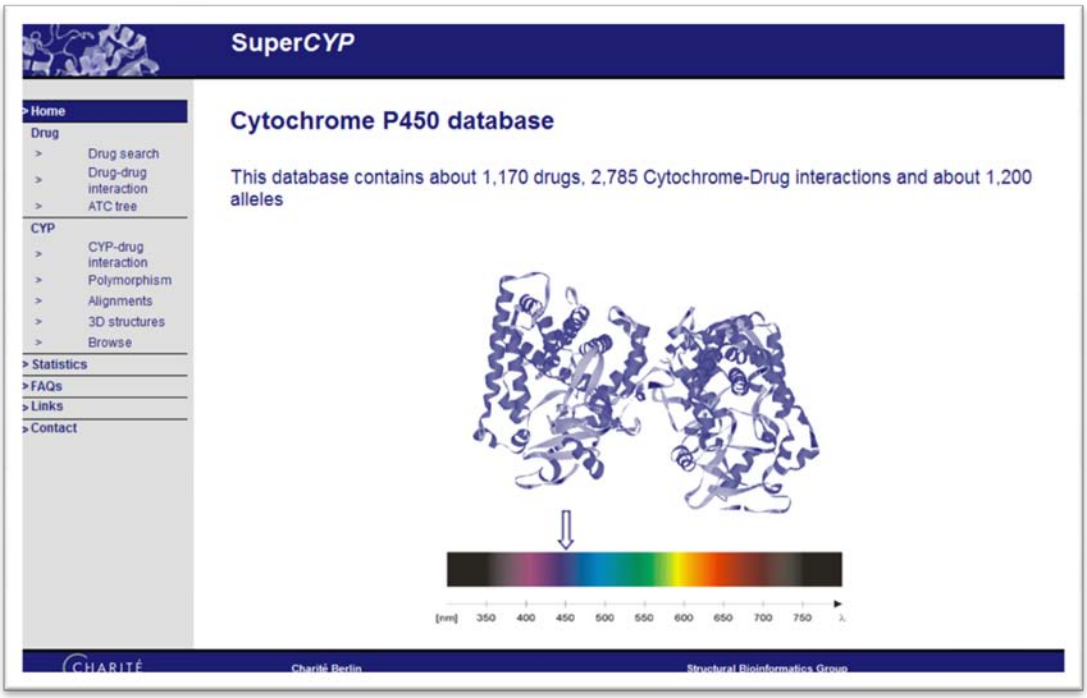


Abbildung 6 *Startseite von SuperCYP*

Mit 1.170 Medikamenten, 3.800 Interaktionen, 1.000 SNPs und 1.200 Protein Mutationen ist diese Datenbank die größte online verfügbare CYP-Informationsquelle im Internet und ist frei verfügbar unter http://bioinformatics.charite.de/supercyp.

Web-basierte, individualisierte Tumortherapie

Die erste in der Zielstellung definierte Teilaufgabe wurde gelöst, indem die Expressionsdaten des National Cancer Institute Panels (NCI-60) (36) in eine eigene Datenbank übernommen wurden. Der Vergleich der Expressionsdaten zwischen Tumor und Zelllinien wurde auf therapeutische Zielmoleküle begrenzt, da die sonst übliche Einbeziehung aller Proteine zu einer Überdeckung der wirklich relevanten Zielmoleküle führt. Die Auswahl dieser Targets war ein längerer, iterativer Prozess unter Einbeziehung verschiedener, interdisziplinärer Partner.

Zur Ermittlung effizienter Substanzen gibt es umfangreiche Daten (über zehn Millionen Werte zur Wachstumsinhibition) zur Chemosensitivität von Tumorzellen. Diese wurden erschlossen und in einen zellulären Fingerprint (37,38) übersetzt. Dadurch ist es möglich, schnell und automatisch in 50.000

Substanzen zu suchen, für die die Wirkung auf mehr als 60 Tumorzellen bekannt ist.

Zur Ermittlung der Auswirkung von stark beschädigten (mutierten) oder nicht mehr vorhandenen Zielmolekülen ist es wichtig, deren dreidimensionale Struktur und ihre Einbindung in verschiedene Signalwege zu berücksichtigen. Dafür wurden die Proteinstrukturen aus der PDB (7) extrahiert und weitere mittels Homology Modelling generiert. Der Signalwegskontext wurde hergestellt, indem Tumor-relevante Pathways aus der KEGG-Datenbank (39) extrahiert wurden und ein Mapping der Zielmoleküle und Medikamente durchgeführt wurde. Eine interaktive graphische Darstellung ermöglicht die Nutzung der Website als Explorationstool (40).

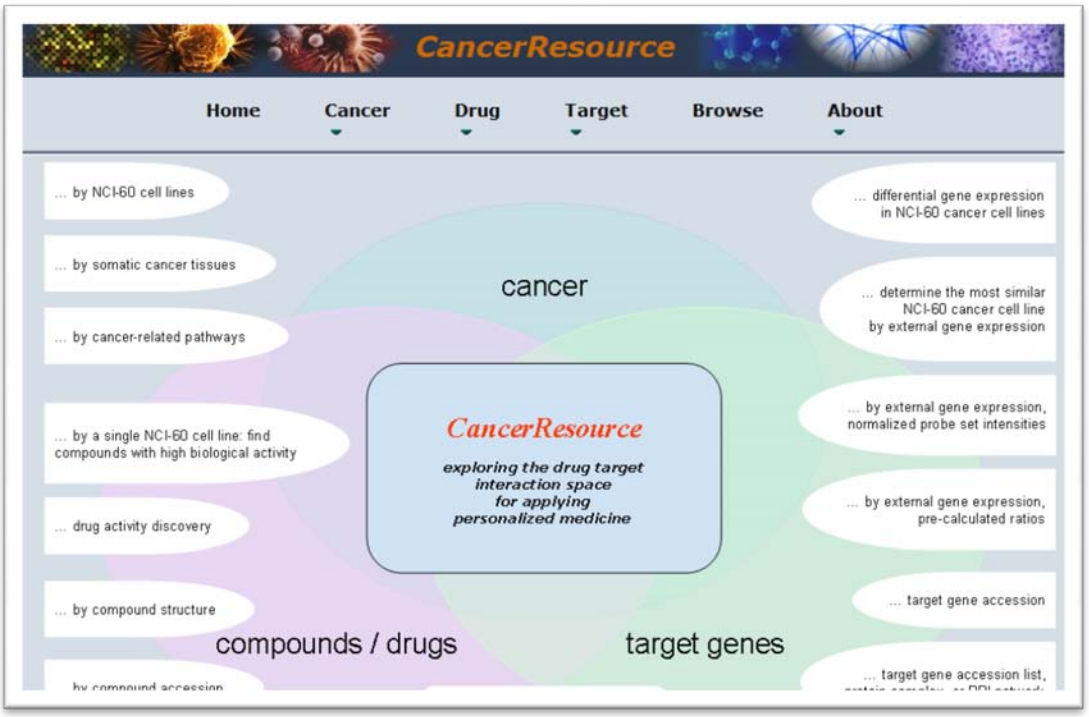


Abbildung 7 *Startseite von CancerResource*

Die Datenbank ist für die akademische Nutzung frei im Internet verfügbar unter: http://bioinformatics.charite.de/cancerresource/.

Süßstoff-Entwicklung

Die SuperSweet Datenbank (35) wurde in erster Linie für Wissenschaftler, insbesondere Ernährungsforscher, entwickelt. Momentan beherbergt die Datenbank 8000 süße Kohlenhydrate, Proteine, D-Aminosäuren und künstliche Süßstoffe, die größtenteils aus der Literaturrecherche und bereits existierenden Datenquellen wie PubChem und PDB gewonnen wurden. Durch Ähnlichkeitssuchen konnten weitere süße oder potenziell süße Substanzen gefunden werden. Neben den allgemeinen physikalischen und chemischen Informationen über die einzelnen Substanzen, gibt es zusätzliche Informationen wie die Anzahl der Kalorien, 3D Struktur, therapeutische Wirkung und Süßkraft. Der Rezeptor für süßen Geschmack wurde mit Hilfe von Homology Modelling gebaut und wird auf der Website zum Download angeboten. Es gibt viele verschiedene Möglichkeiten der Nutzung. Man kann im Datensatz Zucker und Süßstoffe nach sämtlichen oben genannten Eigenschaften suchen, man kann aber auch ein Molekül hochladen oder zeichnen und nach diesen und ähnlichen suchen oder man nutzt den Sweet-Tree, in dem die süßen Moleküle nach Proteinen, Peptiden, Flavonoiden usw. gruppiert sind.

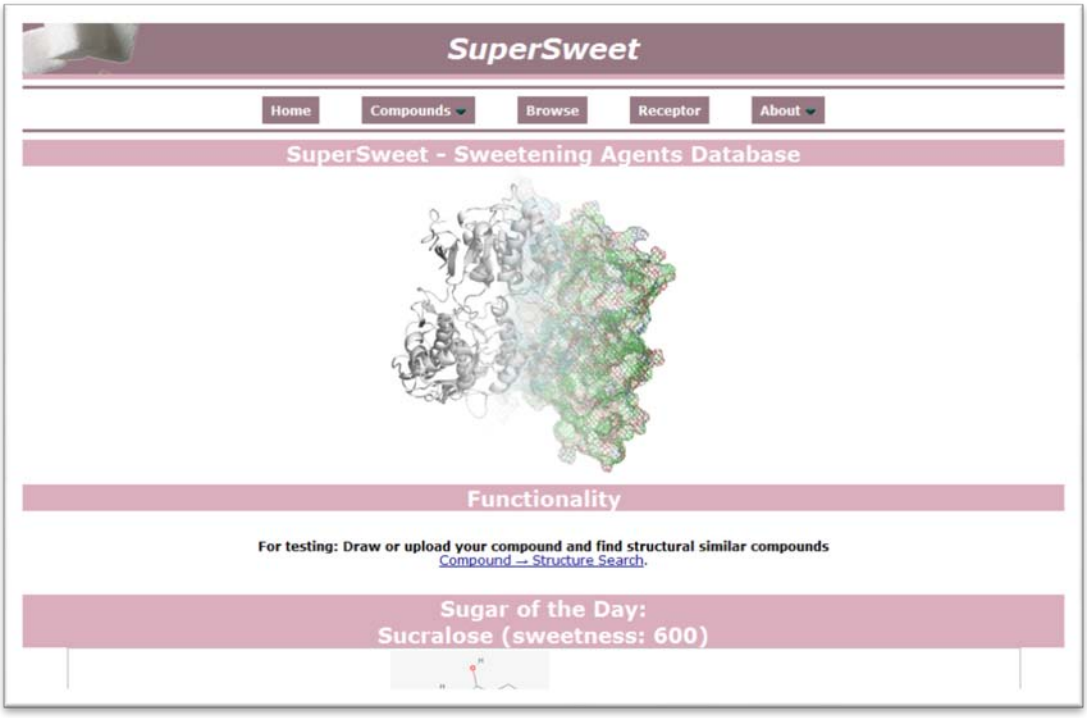


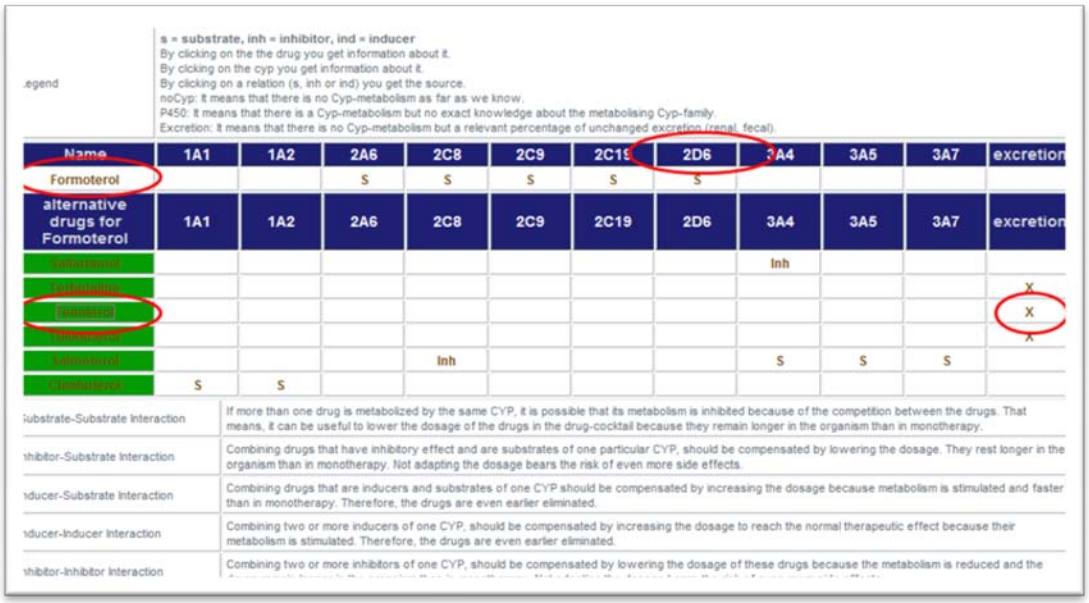
Abbildung 8 *- Startseite von SuperSweet*

Die Datenbank ist für die akademische Nutzung frei im Internet verfügbar unter: http://bioinformatics.charite.de/sweet/.

4. Diskussion

SuperCYP

Mit der Erstellung der SuperCYP wurde in zweierlei Hinsicht eine Lücke geschlossen: die Extraktion und Aufbereitung des Wissens um den Metabolismus von Medikamenten einschließlich der relevanten Informationen über Mutationen, sowie die Erarbeitung eines Tools, um alternative Medikamente bei Überlastung eines CYPs vorschlagen zu können.



egend

s = substrate, inh = inhibitor, ind = inducer
By clicking on the the drug you get information about it.
By clicking on the cyp you get information about it.
By clicking on a relation (s, inh or ind) you get the source.
noCyp: It means that there is no Cyp-metabolism as far as we know.
P450: It means that there is a Cyp-metabolism but no exact knowledge about the metabolising Cyp-family.
Excretion: It means that there is no Cyp-metabolism but a relevant percentage of unchanged excretion (renal, fecal).

Name	1A1	1A2	2A6	2C8	2C9	2C19	2D6	3A4	3A5	3A7	excretion
Formoterol			S	S	S	S	S				

alternative drugs for Formoterol	1A1	1A2	2A6	2C8	2C9	2C19	2D6	3A4	3A5	3A7	excretion
[illegible]								Inh			
[illegible]											X
[illegible]											X
[illegible]											X
[illegible]				Inh				S	S	S	
[illegible]	S	S									

ubstrate-Substrate Interaction	If more than one drug is metabolized by the same CYP, it is possible that its metabolism is inhibited because of the competition between the drugs. That means, it can be useful to lower the dosage of the drugs in the drug-cocktail because they remain longer in the organism than in monotherapy.
hibitor-Substrate Interaction	Combining drugs that have inhibitory effect and are substrates of one particular CYP, should be compensated by lowering the dosage. They rest longer in the organism than in monotherapy. Not adapting the dosage bears the risk of even more side effects.
ducer-Substrate Interaction	Combining drugs that are inducers and substrates of one CYP should be compensated by increasing the dosage because metabolism is stimulated and faster than in monotherapy. Therefore, the drugs are even earlier eliminated.
ducer-Inducer Interaction	Combining two or more inducers of one CYP, should be compensated by increasing the dosage to reach the normal therapeutic effect because their metabolism is stimulated. Therefore, the drugs are even earlier eliminated.
hibitor-Inhibitor Interaction	Combining two or more inhibitors of one CYP, should be compensated by lowering the dosage of these drugs because the metabolism is reduced and the

Abbildung 9 *Auszug der Website mit Vorschlägen zu alternativen Medikamenten: Das häufig verschriebene Medikament gegen Asthma Formoterol wird von CYP 2D6 metabolisiert. Dieses könnte bei Problemen im Metabolismus aufgrund von Polymorphismen durch Fenoterol ersetzt werden, das nur über die Nieren ausgeschieden wird.*

Bisherige Programme, die in Apotheken eingesetzt werden, benutzen nur das Wissen über behördlich gemeldete, meist schwerwiegende Probleme, ohne die möglichen Wechselwirkungen über bestimmte CYPs aus einer aktuellen Datenbank zu berücksichtigen. Direkt nach der Veröffentlichung der Publikation zeigte sich großes Interesse seitens der Industrie, einiger Kliniker und anderer akademischer Gruppen an der SuperCYP Datenbank, sodass die weitere Forschung auf diesem Gebiet wichtig und lohnend ist.

Um die gesammelten Informationen nicht nur der Forschung, sondern auch der medizinischen Praxis zuzuführen, wäre die Entwicklung einer Applikation für mobile Geräte (41), die Ärzte bei sich haben, sinnvoll und lohnend – etwas

entsprechendes existiert bisher nicht, lediglich Informationen aus Beipackzetteln werden berücksichtigt (42,43).

Cancer Resource

In diesem Projekt wurden im Wesentlichen drei Ziele verfolgt. Die Ermittlung einer ähnlichen Tumorzelllinie zu einer Patientenzelllinie ist ein komplexes Problem, denn es gibt keine eindeutige Herangehensweise, die dieses Problem löst. Entgegen anderen Verfahren wurde hier nicht das ganze Genom oder Proteom in den Vergleich einbezogen, sondern sich auf Krebs-relevante Targets zu fokussieren. Dabei spielen folgende biologische Prozesse (bzw. die daran beteiligten Proteine) eine entscheidende Rolle: Zellteilung, Apoptose, DNA-Reparatur, Signal-Kinasen und diverse medizinisch bereits validierte Targets. Natürlich steht ein Beleg für die Richtigkeit dieses Herangehens aus, aber das Problem des Rauschens bei der Verwendung des Gesamt-Proteoms wurde bereits erkannt, weshalb dieser Ansatz viel internationales Interesse findet. Im Rahmen des 1000 Genom Projektes werden die in CancerResource gesammelten Informationen verwendet.

Der zweite Schritt besteht in der Ermittlung einer effizienten Therapie für die ermittelten Zelllinien. Dazu wurde der Ansatz des zellulären Fingerprints implementiert, der Vorteile gegenüber der vom National Cancer Institute (NCI) vorgeschlagenen Methode (44) aufweist. Auf diesem Weg ist es möglich, Substanzen zu ermitteln, die auf die Zielmolekülausstattung der Tumorzellen des Patienten abgestimmt sind – ein erster Schritt in Richtung individualisierte Therapie.

Eine wirklich individuelle Therapie zu entwickeln, die die Veränderungen der Zielmoleküle berücksichtigt, ist im Augenblick noch eine Arbeit, an der verschiedene Spezialisten zusammenarbeiten (http://www.collabrx.com/). Die Nutzung von Web-Servern ist dabei eine wesentliche Komponente und CancerResource stellt verschiedene Tools dafür zur Verfügung. Ein anderer wichtiger Web-Dienst ist „PolyPhen“ (45), weil hier automatisch eine Struktur-basierte Vorhersage für die wahrscheinlichen Auswirkungen gefundener

Mutationen gegeben wird. In Zukunft werden verschiedene Web-Server automatisch über entsprechende Schnittstellen Informationen austauschen, sodass die verschiedenen Informationsquellen virtuell zusammengeführt werden können.

SuperSweet

Zucker als Kalorienquelle, Karies- und Diabetes-Auslöser wird zunehmend durch Süßstoffe ersetzt, die mithin eine wachsende Rolle in der Ernährung spielen. Bisher fehlte eine öffentlich zugängliche, umfassende Ressource, die diese Aspekte mit chemischen und Strukturinformationen verbindet. SuperSweet ermöglicht den Nutzer-freundlichen, Web-basierten Zugang zu den aktuell zugelassenen oder in der Forschung befindlichen Süßstoffen. Dieser umfangreichere Datenbestand wird eine verbesserte Vorhersage von neuen Süßstoffen ermöglichen, denn jetzt ist Entwicklung verbesserter Pharmakophore möglich und umfangreiche Datenbanken mit Millionen von verfügbaren Substanzen (z.B. PubChem) ermöglichen eine schnelle Testung solcher Vorhersagen.

Das entwickelte Modell des Süßstoffrezeptors wird ebenfalls in dieser Richtung neue Vorhersagen durch virtuelle Dockingexperimente, wie sie teilweise bereits auf der Seite hinterlegt sind, ermöglichen.

5. Literaturverzeichnis

1. Ingelman-Sundberg, M. (2005) The human genome project and novel aspects of cytochrome P450 research. *Toxicol Appl Pharmacol*, **207**, 52-56.
2. Nelson, D.R. (2002) Mining databases for cytochrome P450 genes. *Methods Enzymol*, **357**, 3-15.
3. Flockhart, D.A. and Oesterheld, J.R. (2000) Cytochrome P450-mediated drug interactions. *Child Adolesc Psychiatr Clin N Am*, **9**, 43-76.
4. Rendic, S. (2002) Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev*, **34**, 83-448.
5. Rendic, S. and Di Carlo, F.J. (1997) Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab Rev*, **29**, 413-580.
6. Bakare, M.O. (2008) Effective therapeutic dosage of antipsychotic medications in patients with psychotic symptoms: Is there a racial difference? *BMC Res Notes*, **1**, 25.
7. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.
8. Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S. and Pleiss, J. (2007) The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics*, **23**, 2015-2017.
9. Sirim, D., Wagner, F., Lisitsa, A. and Pleiss, J. (2009) The cytochrome P450 engineering database: Integration of biochemical properties. *BMC Biochem*, **10**, 27.
10. Merki, C. (1993) *Zucker gegen Saccharin - About the history of artificial sweeteners*. Campus, Frankfurt/Main / New York.
11. de la Pena, C. Artificial sweetener as a historical window to culturally situated health. *Ann N Y Acad Sci*, **1190**, 159-165.
12. Hicks, R.M., Wakefield, J.S. and Chowaniec, J. (1973) Letter: Co-carcinogenic action of saccharin in the chemical induction of bladder cancer. *Nature*, **243**, 347-349.
13. Weihrauch, M.R. and Diehl, V. (2004) Artificial sweeteners--do they bear a carcinogenic risk? *Ann Oncol*, **15**, 1460-1465.
14. (2006) Artificial sweeteners: no calories...sweet! *FDA Consum*, **40**, 27-28.
15. Edgar, W.M. (1998) Sugar substitutes, chewing gum and dental caries--a review. *Br Dent J*, **184**, 29-32.
16. Hayes, C. (2001) The effect of non-cariogenic sweeteners on the prevention of dental caries: a review of the evidence. *J Dent Educ*, **65**, 1106-1109.
17. Benton, D. (2005) Can artificial sweeteners help control body weight and prevent obesity? *Nutr Res Rev*, **18**, 63-76.
18. Drewnowski, A. (1999) Intense sweeteners and energy density of foods: implications for weight control. *Eur J Clin Nutr*, **53**, 757-763.
19. Temussi, P. (2006) The history of sweet taste: not exactly a piece of cake. *J Mol Recognit*, **19**, 188-199.
20. Morini, G., Bassoli, A. and Temussi, P.A. (2005) From small sweeteners to sweet proteins: anatomy of the binding sites of the human T1R2_T1R3 receptor. *J Med Chem*, **48**, 5520-5529.
21. Sy, S.K., Johnston, J.A., Class, C.F., Roberts, E.A., Kalow, W. and Tang, B.K. (2002) Mulitvariate cluster analysis of a human hepatic cytochrome P450 database. *Eur J Clin Pharmacol*, **58**, 559-562.
22. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J. and Bryant, S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*, **37**, W623-633.
23. Davis, A.P., Murphy, C.G., Saraceni-Richards, C.A., Rosenstein, M.C., Wiegers, T.C. and Mattingly, C.J. (2009) Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res*, **37**, D786-792.
24. Hernandez-Boussard, T., Whirl-Carrillo, M., Hebert, J.M., Gong, L., Owen, R., Gong, M., Gor, W., Liu, F., Truong, C., Whaley, R. *et al.* (2008) The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge. *Nucleic Acids Res*, **36**, D913-918.
25. Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B. and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*, **36**, D901-906.

26. Zhu, F., Han, B., Kumar, P., Liu, X., Ma, X., Wei, X., Huang, L., Guo, Y., Han, L., Zheng, C. *et al.* Update of TTD: Therapeutic Target Database. *Nucleic Acids Res*, **38**, D787-791.
27. de Leon, J. (2009) The future (or lack of future) of personalized prescription in psychiatry. *Pharmacol Res*, **59**, 81-89.
28. Kirchheiner, J., Brosen, K., Dahl, M.L., Gram, L.F., Kasper, S., Roots, I., Sjoqvist, F., Spina, E. and Brockmoller, J. (2001) CYP2D6 and CYP2C19 genotype-based dose recommendations for antidepressants: a first step towards subpopulation-specific dosages. *Acta psychiatrica Scandinavica*, **104**, 173-192.
29. Roots, I., Laschinski, G., Arjomand-Nahad, F., Kirchheiner, J., Schwarz, D., Brockmoller, J., Cascorbi, I. and Gerloff, T. (2007) Genotype and phenotype relationship in drug metabolism. *Ernst Schering Research Foundation workshop*, 81-100.
30. Seeringer, A. and Kirchheiner, J. (2008) Pharmacogenetics-guided dose modifications of antidepressants. *Clinics in laboratory medicine*, **28**, 619-626.
31. Hildebrand, P.W., Goede, A., Bauer, R.A., Gruening, B., Ismer, J., Michalsky, E. and Preissner, R. (2009) SuperLooper--a prediction server for the modeling of loops in globular and membrane proteins. *Nucleic Acids Res*, **37**, W571-574.
32. Willett, P. Similarity searching using 2D structural fingerprints. *Methods Mol Biol*, **672**, 133-158.
33. Winnenburg, R., Plake, C. and Schroeder, M. (2009) Improved mutation tagging with gene identifiers applied to membrane protein stability prediction. *BMC Bioinformatics*, **10 Suppl 8**, S3.
34. Preissner, S., Kroll, K., Dunkel, M., Senger, C., Goldsobel, G., Kuzman, D., Guenther, S., Winnenburg, R., Schroeder, M. and Preissner, R. SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res*, **38**, D237-243.
35. Ahmed, J., Preissner, S., Dunkel, M., Worth, C.L., Eckert, A. and Preissner, R. SuperSweet--a resource on natural and artificial sweetening agents. *Nucleic Acids Res*.
36. Shoemaker, R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nature reviews*, **6**, 813-823.
37. Thimm, M., Goede, A., Hougardy, S. and Preissner, R. (2004) Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *Journal of chemical information and computer sciences*, **44**, 1816-1822.
38. Fullbeck, M., Dunkel, M., Hossbach, J., Daniel, P.T. and Preissner, R. (2009) Cellular fingerprints: a novel approach using large-scale cancer cell line data for the identification of potential anticancer agents. *Chemical biology & drug design*, **74**, 439-448.
39. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*, **34**, D354-357.
40. Ahmed, J., Meinel, T., Dunkel, M., Murgueitio, M.S., Adams, R., Blasse, C., Eckert, A., Preissner, S. and Preissner, R. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res*.
41. Rappaport, H. (2006) Epocrates Rx a favourite tool. *Canadian family physician Medecin de famille canadien*, **52**, 1208.
42. Fox, B.I., Felkey, B.G., Berger, B.A., Krueger, K.P. and Rainer, R.K., Jr. (2007) Use of personal digital assistants for documentation of pharmacists' interventions: a literature review. *Am J Health Syst Pharm*, **64**, 1516-1525.
43. Perkins, N.A., Murphy, J.E., Malone, D.C. and Armstrong, E.P. (2006) Performance of drug-drug interaction software for personal digital assistants. *The Annals of pharmacotherapy*, **40**, 850-855.
44. Paull, K.D., Shoemaker, R.H., Hodes, L., Monks, A., Scudiero, D.A., Rubinstein, L., Plowman, J. and Boyd, M.R. (1989) Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *Journal of the National Cancer Institute*, **81**, 1088-1092.
45. Flanagan, S.E., Patch, A.M. and Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic testing and molecular biomarkers*, **14**, 533-537.

Anteilserklärung

Publikation 1

Autoren: **Saskia Preissner**, Katharina Kroll, Mathias Dunkel, Christian Senger, Gady Goldsobel, Daniel Kuzman, Stefan Guenther, Rainer Winnenburg, Michael Schroeder, Robert Preissner

Titel: SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions.

Zeitschrift: Nucleic Acids Research (PMID: 19934256)

Anteil: 50 Prozent

Beitrag im Einzelnen: Mitarbeit an der Erstellung des Konzepts, Entwicklung der Suchkriterien für das Textmining, Mitarbeit an der Auswertung und Sortierung der Textmining-Ergebnisse, individuelle Suche nach Medikamentendaten in der Roten Liste und auf den Seiten der FDA, Entwicklung des Druck-Cocktail-Konzeptes, 3D-Homology-Modelling aller nicht verfügbaren CYP-Strukturen, Mitarbeit an der Oberflächengestaltung der Website, Erstellung der Publikation in Absprache mit RP.

Publikation 2

Autoren: Jessica Ahmed, **Saskia Preissner**, Mathias Dunkel, Catherine L. Worth, Andreas Eckert and Robert Preissner

Titel: SuperSweet – a resource on natural and artificial sweetening agents

Zeitschrift: Nucleic Acids Research (PMID: 20952410)

Anteil: 40 Prozent

Beitrag im Einzelnen: Mitarbeit an der Erstellung des Konzepts, Erstellung eines Rohdatensatzes, auf dessen Basis Ähnlichkeitssuchen und Texmining durchgeführt wurden, manuelle Evaluation, Sortierung und Gruppierung der Süßstoffe und Zucker und Ergänzung fehlender Daten inklusive 3D-Modelle der Zucker und Süßstoffe, Mitarbeit an der Oberflächengestaltung der Website, Mitarbeit an der Erstellung der Publikation in Absprache mit JA und RP.

Publikation 3

Autoren: Jessica Ahmed, Thomas Meinel, Mathias Dunkel, Manuela Murgueitio, Robert Adams, Corinna Blasse, Andreas Eckert, **Saskia Preissner**, Robert Preissner

Titel: CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge.

Zeitschrift: Nucleic Acids Research (PMID: 20952398)

Anteil: 10 Prozent

Beitrag im Einzelnen: Homology Modelling von etwa 400 Zielmolekülen.

Published online 24 November 2009 *Nucleic Acids Research, 2010, Vol. 38, Database issue* ***D237–D243***
doi:10.1093/nar/gkp970

SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions

Saskia Preissner[1], Katharina Kroll[1], Mathias Dunkel[1], Christian Senger[2], Gady Goldsobel[1], Daniel Kuzman[1], Stefan Guenther[1], Rainer Winnenburg[3], Michael Schroeder[3] and Robert Preissner[1,*]

[1]Structural Bioinformatics Group, Institute of Physiology, Charité—University Medicine Berlin, Arnimallee 22, 14197 Berlin, [2]Department of Internal Medicine VI, Clinical Pharmacology and Pharmacoepidemiology, University of Heidelberg, Heidelberg and [3]Biotechnology Center, Technische Universität Dresden, Dresden, Germany

Received August 18, 2009; Revised September 30, 2009; Accepted October 14, 2009

ABSTRACT

Much of the information on the Cytochrome P450 enzymes (CYPs) is spread across literature and the internet. Aggregating knowledge about CYPs into one database makes the search more efficient. Text mining on 57 CYPs and drugs led to a mass of papers, which were screened manually for facts about metabolism, SNPs and their effects on drug degradation. Information was put into a database, which enables the user not only to look up a particular CYP and all metabolized drugs, but also to check tolerability of drug-cocktails and to find alternative combinations, to use metabolic pathways more efficiently. The SuperCYP database contains 1170 drugs with more than 3800 interactions including references. Approximately 2000 SNPs and mutations are listed and ordered according to their effect on expression and/or activity. SuperCYP (http://bioinformatics.charite.de/supercyp) is a comprehensive resource focused on CYPs and drug metabolism. Homology-modeled structures of the CYPs can be downloaded in PDB format and related drugs are available as MOL-files. Within the resource, CYPs can be aligned with each other, drug-cocktails can be 'mixed', SNPs, protein point mutations, and their effects can be viewed and corresponding PubMed IDs are given. SuperCYP is meant to be a platform and a starting point for scientists and health professionals for furthering their research.

INTRODUCTION

The family of Cytochrome P450 enzymes has been the focus of pharmaceutical research for decades, as evidenced by the more than 100 000 articles in PubMed. Drug metabolism is a complex biochemical network, which consists of many different parts and reactions in the human organism. Some drugs are excreted unchanged in urine and feces without passing any metabolic treatment in the liver, but most of the drugs have a multi-step metabolism, which is mainly associated with Cytochrome P450 (CYP). These enzymes belong to the family of monooxygenases, which reach maximum of absorption at 450 nm. This work focuses on human CYPs (Table 1).

The chemical reaction is: $R–H + O_2 + NADPH + H^+ \rightarrow R–OH + H_2O + NADP^+$ (1). CYPs catalyze a large amount of chemical reactions, such as alcohol oxidations, dehydrogenation and isomerizations. One of the most difficult tasks of medical science is to find combinations of drugs that do not affect each other's metabolic pathways. The Human Genome Project identified 57 human CYPs (2), ordered into 18 families and 43 subfamilies by sequence similarities. In general, the human CYPs share the same fold, but because of its spacious binding site CYP 3A4 is capable of metabolizing at least 422 drugs (Figure 1).

Many drugs also inhibit or induce the activity of CYPs, which is important to health professionals trying to dose medicines. If a drug induces a CYP that is also active in another drug's metabolism, the dosage of the first drug must be enhanced to achieve a therapeutic effect. In case of inhibition of a CYP, the dosage of the drug can be reduced, which also lowers side effects (3). Information on CYP-structures (4), binding sites (5,6), interactions

*To whom correspondence should be addressed. Tel: +49 30 8445 1649; Fax: +49 30 8445 1551; Email: robert.preissner@charite.de

Table 1. CYP-overview

CYP	Tissue sites	Localization	Typical reaction
1A1	Lung, several extrahepatic sites, peripher blood cells	ER	Benzopyrene 3-hydroxylation
1A2	Liver	ER	Caffeine N3-demethylation
1B1	Many extrahepatic sites, incl. lung and kidney	ER	17β-Estradiol 4-hydroxylation
2A6	Liver, lung and several extrahepatic sites	ER	Coumarin 7-hydroxylation
2A7	Only information is identification in human genome	ER	
2A13	Nasal tissue	ER	Activation of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
2B6	Liver, lung	ER	(*S*)-Mephenytoin N-demethylation
2C8	Liver	ER	Taxol 6α-hydroxylation
2C9	Liver	ER	Tobutamine methyl hydroxylation
2C18	Liver	ER	
2C19	Liver	ER	(*S*)-Mephenytoin 4′-hydroxylation
2D6	Liver	ER	Debrisoquine 4-hydroxylation
2E1	Liver, lung and other tissues	ER	Chlorzoxazone 6-hydroxylation
2F1	Lung	ER	3-Methylindole activation
2J2	Lung	ER	Arachidonic acid oxidation
2R1	Only information is identification in human genome		
2S1	Lung	ER	
2U1	Only information is identification in human genome		
2W1	Only information is identification in human genome		
3A4	Liver, small intestine	ER	Testosterone 6β-hydroxylation
3A5	Liver, lung	ER	Testosterone 6β-hydroxylation
3A7	Ffetal liver	ER	Testosterone 6β-hydroxylation
3A43	mRNA detected in gonads	ER	
4A11	Liver	ER	Fetty acid ω-hydroxylation
4A22	Only information is identification in human genome	ER	
4B1	Lung	ER	Lauric acid ω-hydroxylation
4F2	Liver	ER	Leukotriene B4 ω-hydroxylation
4F3	Neutrophils	ER	Leukotriene B4 ω-hydroxylation
4F8	Seminal vesicles	ER	Prostaglandin ω-2 hydroxylation
4F11	Liver	ER	
4F12	Liver	ER	Arachidonic acid ω-,ω-2-hydroxylation
4F22	Only information is identification in human genome		
4V2	Only information is identification in human genome		
4X1	Only information is identification in human genome		
4Z1	Only information is identification in human genome		CYP 4A20
5A1	Platelets	ER	Thromboxane A2 synthetase reaction
7A1	Liver	ER	Cholesterol 7α-hydroxylation
7B1	Brain	ER	Dehydroepiandrosterone 7α-hydroxylation
8A1	Aorta, others	ER	Prostacyclin synthase reaction
8B1	Liver	ER	7α-hydroxyprogesterone 12-hydroxylation
11A1	Adrenals, other steroidogenic tissues	Mitochondrium	Cholesterol side-chain clevage
11B1	Adrenals	Mitochondrium	11-Deoxycortisol 11-hydroxylation
11B2	Adrenals	Mitochondrium	Corticosterone 18-hydroxylation
17A1	Steroidogenic tissue	ER	Steroid 17α-hydroxylation
19A1	Steroidogenic tissue, adipose, brain	ER	Androgen aromatization
20A1	Only information is identification in human genome		
21A2	Steroidogenic tissue	ER	17-Hydroxyprogesterone 21-hydroxylation
24A1	Kidney	Mitochondrium	25-Hydroxyvitamin D3 24-hydroxylation
26A1	Several	ER	Retinoic acid 4-hydroxylation
26B1	Brain	ER	Retinoic acid 4-hydroxylation
26C1	Only information is identification in human genome	ER	
27A1	Liver	Mitochondrium	Sterol 27-hydroxylation
27B1	Kidney	Mitochondrium	Vitamin D3 1-hydroxylation
27C1	Only information is identification in human genome		
39A1	Liver	ER	24-Hydroxycholesterol 7-hydroxylase
46A1	Brain	ER	Cholesterol 24-hydroxylation
51A1	Liver, testes	ER	Lanosterol 14α-demethylation

and different genotypes (7,8) must be combined to allow reduce side effects and to determine correct dosages of medicine. David Nelson and colleagues listed human CYPs with alignments (9) and also address nomenclature. Flockhart created a CYP–drug-interaction table with the intention of avoiding undesired side effects when prescribing more than one drug (10), a key issue in an aging society. In 1997, Rendic and Di Carlo (11) compiled a comprehensive collection of data on CYP reactions and drug interactions. From the patients' perspective, it is useful to know what kind of food or additional drugs they should avoid when taking their medications. Drug interactions such as these can be checked at the University of Maryland

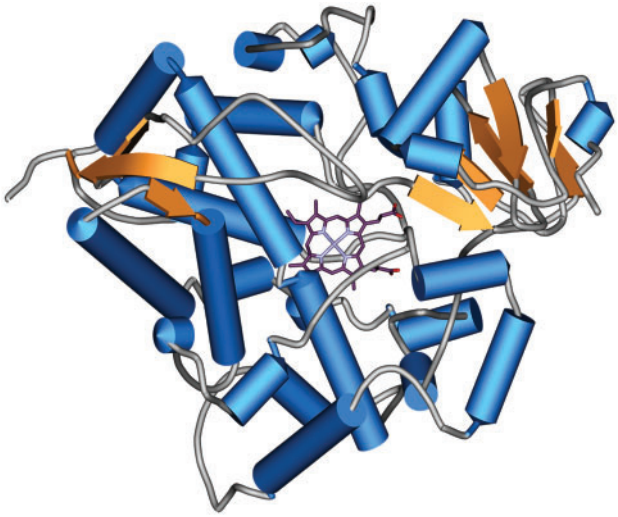


Figure 1. Plot of the 3D structure of CYP 3A4. Helices are shown as cylinders (light blue) and beta-sheets as orange arrows. The heme moiety is depicted in stick representation. The majority of the structure is made up of an alpha-helical arrangement, but N- and C-terminal small beta-domains can be found.

(http://www.umm.edu/adam/drug_checker.htm). Another field of interest is the diversity of human CYPs in different ethnic groups. Pharmacogenomic studies reported that African–American patients are precribed higher dosages of antipsychotic medications compared to Caucasians (12). Information on mutations in CYPs is widely spanned accross the internet. An up-to-date resource on available 3D CYP structures from the Protein Data Bank (13) is presented by CYPED (14). The Center for Molecular Design collected exhaustive information on the CYPs of many species in their Knowledgebase (http://cpd.ibmh.msk.su/), including sequences and links to the Nomenclature Committee (15). Despite the large amount of information on CYPs, optimizing multiple drug prescriptions using CYP metabolism is still complicated (16). To overcome these problems, SuperCYP aims at providing a user-friendly platform enabling health professionals to optimize drug cocktails regarding the degree of CYP capacity utilization. Furthermore, a number of scientific issues were addressed: comprehensive information about mutations influencing the metabolism of drugs (17,18), racial differences (19), comparative analysis of drug binding sites to explain their promiscuousness. To this end, multiple sequence alignments are required and will be the basis of homology-built 3D structures of all human CYPs (20,21).

MATERIALS AND METHODS

Information on CYPs was collected from scientific literature (22) and various web resources: e.g. Nelsons Homepage (23), Flockharts Interaction Table (http://www.medicine.iupui.edu/Flockhart/table.htm), University of Maryland's Drug Checker, PubChem (24), PDB (13). Some information was gathered from FDA-files. Abstracts of PubMed were automatically filtered for relevant articles using specific keywords. The abstracts were screened for WHO-drugs and their synonyms, as was a set of human CYPs with synonyms. A team of scientists manually processed the papers found in PubMed. Each drug was attributed to those CYPs that are involved in drug metabolism as substrate, inhibitor or inducer. In addition, mutations on the protein level were retrieved utilizing the mutation/gene association text mining system described by Winneburg and Schroeder (25). The tool retrieved ~550 distinct mutations from more than 400 scientific journal abstracts. In total, 450 new amino acid substitutions were retrieved in addition to those 500 already obtained from PubMed. The approach was adapted to gather information on alleles, the change of enzyme/transcription activity and populations. As with the first literature screening, the automated predictions were manually verified before being included in the database, and in all cases PubMed-IDs as references are provided to track details regarding the given information.

SuperCYP is designed as a relational database on a MySQL server. For chemical functionality, the MyChem package is included, which aims to provide a complete set of functions for handling chemical data within MySQL. Most of the functions used by MyChem depend on Open Babel (26). For displaying 3D structures, Jmol—an open-source Java viewer for chemical structures in 3D—is used. For visual inspection of the alignments, JALVIEW is installed. ChemSketch was applied as a built-in molecule editor, which allows users to screen using self-edited molecules. The website is built with PHP and javascript, web access is enabled via Apache Webserver 2.2.

DATABASE

The SuperCYP website was developed as a user-friendly platform for researchers and health professionals. The navigation bar on the left side offers 'FAQs' or Frequently Asked Questions, for first-time users.

'Drug search' enables the user to search for a drug and find information on its metabolism. 'Get Information' leads to a table listing CYPs involved in the metabolism of the drug. Here there is also a description of possible consequences and after clicking on the drug name on the results page 'Drug search' enables the user to get information for compounds by means of the CAS-number or name.

The 'ATC tree' is the WHO classification system that classifies drugs into different groups according to anatomic site of action, their therapeutical effect and chemical structure. It is the basis for drug alternative drug recommendations. In a Java applet, the user finds a drop-down tree with major and minor branches of classification. All drugs affiliated with a minor branch are listed in a table and information on the CYP metabolism is provided.

'Drug–drug interaction' is the main feature of the database. It allows users to enter names of several different drugs and to check interactions between these drugs, but they also receive alternative drug options.

As an example, Omeprazol, a proton pump inhibitor, and Nebivolol, a beta-blocker, interact on the CYP level. After selecting the drugs, the database provides detailed

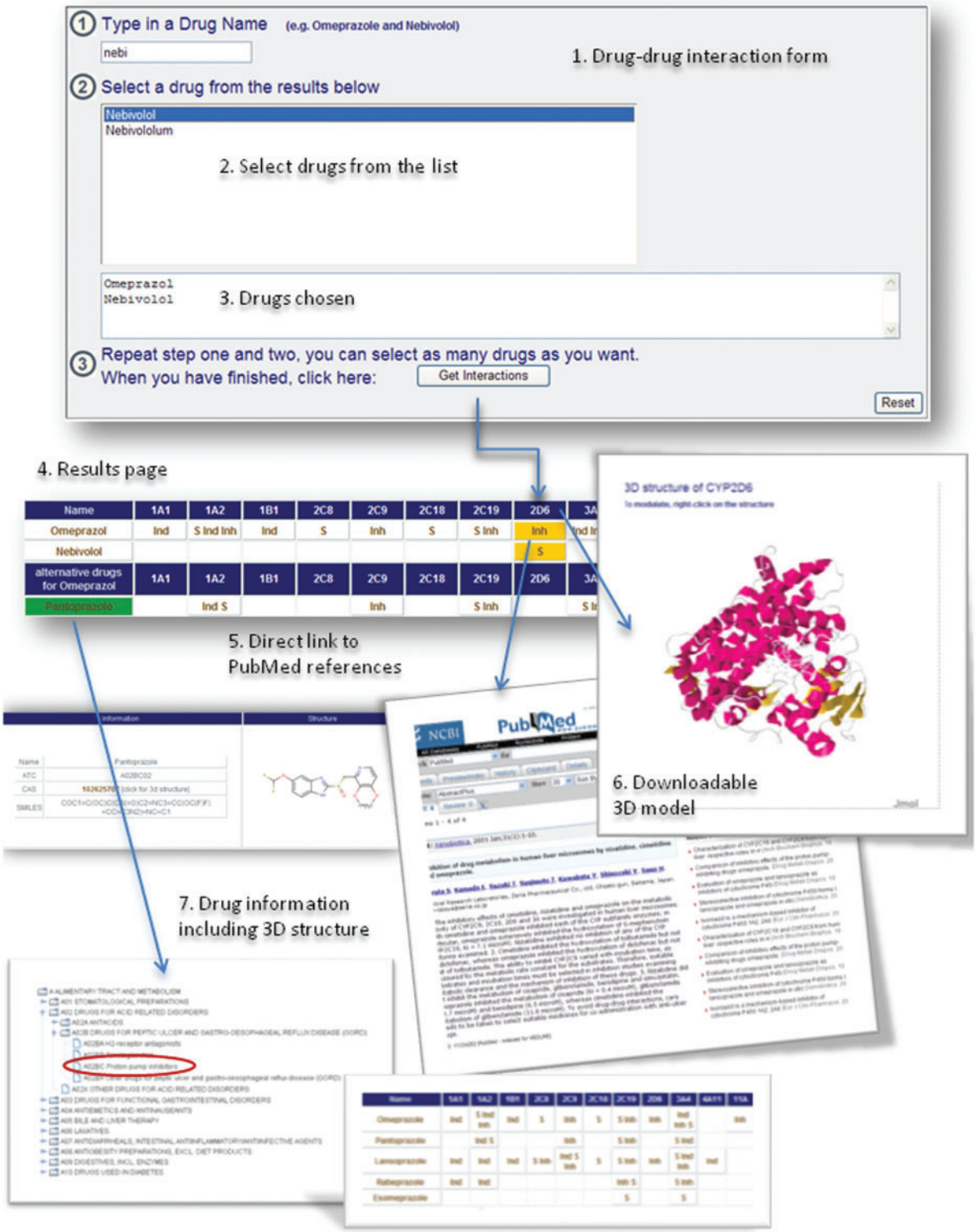


Figure 2. Queries and results of the SuperCyp web-interface explaining the various possibilities of the 'Drug–drug-interaction' option with the help of two example drugs: Omeprazole and Nebivolol.

information on drug structures and ATC group plus CAS numbers (Figure 2).

The successive 'results' page warns that Omeprazol has an inhibitory effect on CYP 2D6, whereas Nebivolol is a substrate. The colored background of the table illustrates this dual use of the CYP metabolism pathway. To avoid this and to optimize the drug composition, Omeprazol can be substituted with other drugs from the

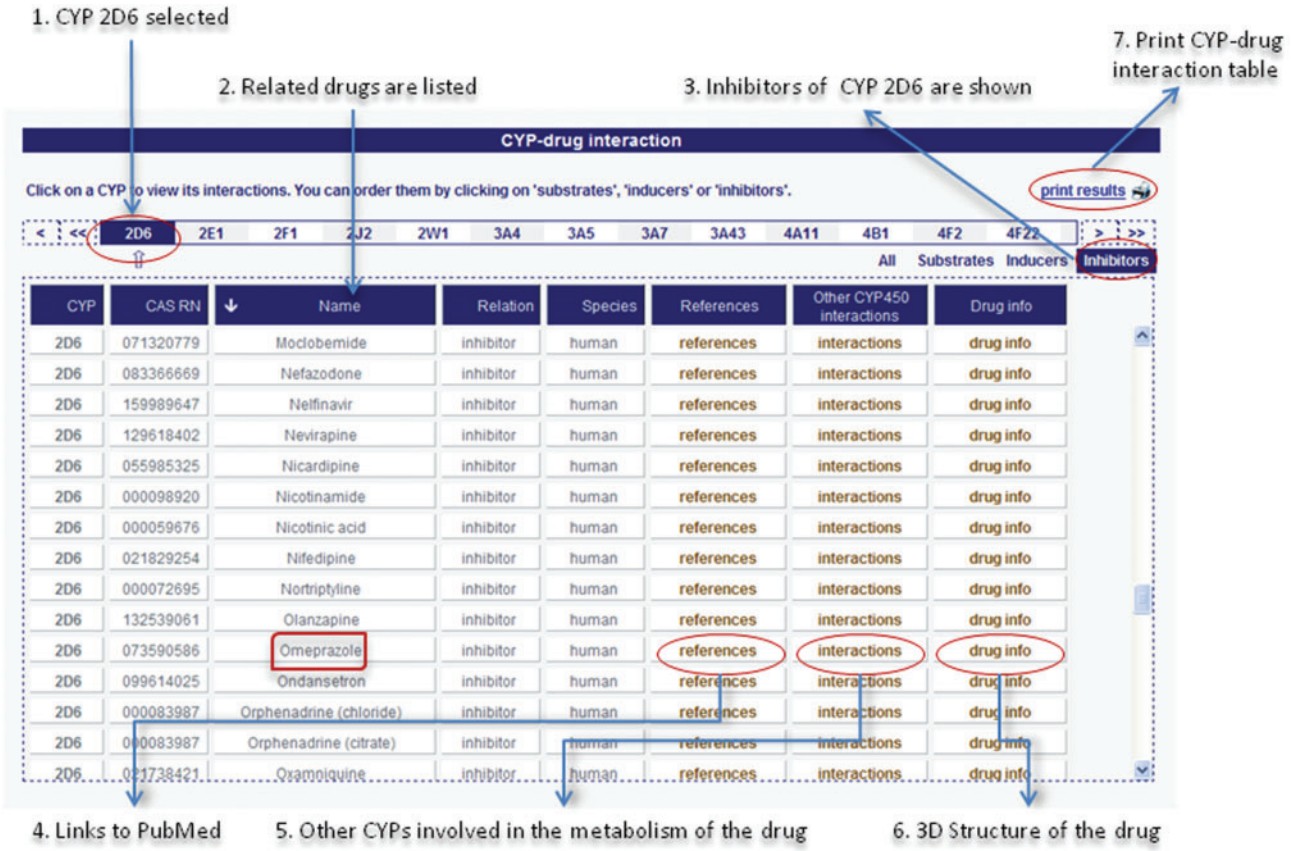


Figure 3. Results of the SuperCyp web-interface for CYP 2D6 explaining the functionality of the 'CYP–drug interaction' table.

same ATC-group, for example Pantoprazol, achieving a comparable effect, but using another pathway.

The proposal of Pantoprazol is derived from the assumption that it does not interact with CYP 2D6. All data on the proposed drugs are provided, and the reference to related publications is given.

The 'CYP–Drug-interaction' allows users to browse substrates, inducers and inhibitors of a certain CYP (Figure 3).

After the user has selected a CYP from the task menu, all known relations with drugs are listed in a table. Then users can specify the relation and focus on substrates, inducers or inhibitors. Respective drugs are given in a table and combined with further information on the particular drug and all CYP interactions. References are linked to PubMed and other scientific websites or articles. The 'Drug Info' button is linked to the SuperDrug Database (27), which provides a large number of more specific information on the particular drug.

'Polymorphism' shows single nucleotide polymorphisms for a particular CYP. All known alleles (15,28) are shown and if there is a decrease or increase in activity or expression, this information is provided. Nucleotide changes and their effects, as well as enzyme activity and assay type are given with corresponding references. Some mutation entries address the protein level directly, in which cases information on SNPs may be missing. However, it is desirable to include protein data, as they provide valuable insights into structure-function relationships.

Example: For CYP11B2, which encodes the enzyme aldosterone synthase (P450aldo), no SNPs were retrieved through keyword searches. However, our mutation/gene association text mining system found 54 protein mutations in 41 PubMed abstracts, which were then added to the database. Among those, the substitution of the highly conserved arginine at position 384 by proline reportedly led to a complete loss of function of this enzyme as part of the autosomal recessively inherited disorder CMO-I deficiency in male Caucasians.

'Alignments' uses a structure-based alignment program to match the amino acid sequence of all CYPs. It is possible to create a multiple sequence alignment from any number of sequences or to align them with external sequences by uploading a file or entering a sequence in FASTA format. Users may also draft a convenient output with Jalview.

'Three-dimensional structures' displays protein structures of human CYPs. Existing structures were extracted from the PDB. Theoretical models were generated with Swiss-Model (29) or built manually. All structures are downloadable as PDB-files and more information on the CYP is given in the box on the right side.

Clicking the 'Browse'-button leads to a Java applet, where all CYPs are listed in a drop-down tree, ordered

by main families and subfamilies. Each CYP is viewable as a model and further information on its interactions is provided.

RESULTS

Comprehensive data on the 57 human CYPs are stored in the SuperCYP database. For all CYPs, the sequences can be viewed and aligned. Around 1000 SNPs and more than 1200 protein mutations are listed and ordered by their effect on expression and/or activity.

The Protein Data Bank (13,30) provides several crystal structures of Cytochome P450 enzymes. Table 2 lists those nine CYPs that are represented in PDB, while families 4, 5, 7, 11, 17, 19, 20, 21, 24, 26, 27, 39, 46 and 51 are not. Based on the known structures, the SuperCYP database provides theoretical 3D models of the 48 missing human CYPs (Table 2).

With 1170 drugs and ~3800 interactions, SuperCYP provides the largest number of CYP relations and corresponding information available online. Additionally, checking the tolerability of drug-cocktails and finding alternative uses of metabolic pathways has been made more efficient with the database.

DISCUSSION

The degradation of compounds by CYPs plays an important role in drug–drug interactions that are responsible for harmful adverse effects, e.g. deadly acute renal failure (31). Detailed knowledge about CYP–drug relations is therefore essential for recognizing incompatible drug combinations and to allow individualized therapies. Besides all-inclusive information about drugs, CYPs and their relations, descriptive data such as known CYP-mutations, their phenotypic effects, or the structural information about the CYPs and drugs will enable systematic approaches to quantitative structure–activity relationships. SuperCYP indicates potentially perilous drug–drug interactions and proposes alternative drugs improving mixture or dosage. In conclusion, the server provides a comprehensive resource on CYPs as well as a discovery tool for analysis of drug degradation and drug-cocktail optimization.

Table 2. Overview of the coverage of the CYP-classes with experimentally determined structures

CYP	PDB-ID
1A2	2hi4
2A6	1z10, 1z11, 1fdu, 2fdv, 2fdw, 2fdy, 2pg5, 2pg6, 2pg7
2A13	2p85
2C8	1pq2, 2nni, 2vn0
2C9	1og2, 1og5, 1r9o
2D6	2f9q
2R1	3c6g
3A4	1w0e, 1w0f, 1w0g, 1tqn, 2j0d, 2v0m
8A1	2iag

AVAILABILITY AND REQUIREMENTS

SuperCYP is available at http://bioinformatics.charite.de/supercyp and can be obtained via a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License. To access all features of the website the latest version of Java should be installed. The database will be updated every 6 months.

ACKNOWLEDGEMENTS

We would like to thank Sarah Preissner for her assistance in completing the structural models of the CYPs.

FUNDING

Deutsche Forschungsgemeinschaft (SFB 449), Investitionsbank Berlin (IBB), International Research Training Group (IRTG) Berlin–Boston–Kyoto and Deutsche Krebshilfe, Bundesministerium für Bildung und Forschung (BMBF) and European Union (EU). Funding for open access charge: DFG SFB 449, BMBF.

Conflict of interest statement. None declared.

REFERENCES

1. Lepesheva,G.I., Hargrove,T.Y., Kleshchenko,Y., Nes,W.D., Villalta,F. and Waterman,M.R. (2008) CYP51: a major drug target in the Cytochrome P450 superfamily. *Lipids*, **43**, 1117–1125.
2. Ingelman-Sundberg,M. (2005) The human genome project and novel aspects of cytochrome P450 research. *Toxicol. Appl. Pharmacol.*, **207**, 52–56.
3. Lu,C., Hatsis,P., Berg,C., Lee,F.W. and Balani,S.K. (2008) Prediction of pharmacokinetic drug-drug interactions using human hepatocyte suspension in plasma and cytochrome P450 phenotypic data. II. In vitro-in vivo correlation with ketoconazole. *Drug Metab. Dispos.*, **36**, 1255–1260.
4. Keseru,G.M. (1998) Application of theoretical and synthetic models to cytochrome P450 catalysed metabolic reactions. *Acta Pharm. Hung.*, **68**, 65–69.
5. Fernando,H., Halpert,J.R. and Davydov,D.R. (2006) Resolution of multiple substrate binding sites in cytochrome P450 3A4: the stoichiometry of the enzyme-substrate complexes probed by FRET and Job's titration. *Biochemistry*, **45**, 4199–4209.
6. Dell,H., Wong,S.C., Phillips,I.R. and Shephard,E.A. (1998) Identification of regulatory protein-binding sites in cytochrome P450 genes by means of the gel-retardation assay. *Methods Mol. Biol.*, **107**, 381–394.
7. Bertilsson,L. (1995) Geographical/interracial differences in polymorphic drug oxidation. Current state of knowledge of cytochromes P450 (CYP) 2D6 and 2C19. *Clin. Pharmacokinet*, **29**, 192–209.
8. Bathum,L., Andersen-Ranberg,K., Boldsen,J., Brosen,K. and Jeune,B. (1998) Genotypes for the cytochrome P450 enzymes CYP2D6 and CYP2C19 in human longevitY. Role of CYP2D6 and CYP2C19 in longevity. *Eur. J. Clin. Pharmacol.*, **54**, 427–430.
9. Nelson,D.R. (2002) Mining databases for cytochrome P450 genes. *Methods Enzymol.*, **357**, 3–15.
10. Flockhart,D.A. and Oesterheld,J.R. (2000) Cytochrome P450-mediated drug interactions. *Child Adolesc. Psychiatr. Clin. N. Am.*, **9**, 43–76.
11. Rendic,S. and Di Carlo,F.J. (1997) Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab. Rev.*, **29**, 413–580.
12. Bakare,M.O. (2008) Effective therapeutic dosage of antipsychotic medications in patients with psychotic symptoms: Is there a racial difference? *BMC Res. Notes*, **1**, 25.

13. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
14. Fischer,M., Knoll,M., Sirim,D., Wagner,F., Funke,S. and Pleiss,J. (2007) The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics*, **23**, 2015–2017.
15. Sim,S.C. and Ingelman-Sundberg,M. (2006) The human cytochrome P450 Allele Nomenclature Committee Web site: submission criteria, procedures, and objectives. *Methods Mol. Biol.*, **320**, 183–191.
16. Radhakrishnan,M.L. and Tidor,B. (2008) Optimal drug cocktail design: methods for targeting molecular ensembles and insights from theoretical model systems. *J. Chem. Inf. Model*, **48**, 1055–1073.
17. van der Weide,J. and Steijns,L.S. (1999) Cytochrome P450 enzyme system: genetic polymorphisms and impact on clinical pharmacology. *Ann. Clin. Biochem.*, **36(Pt 6)**, 722–729.
18. Giacomini,K.M., Brett,C.M., Altman,R.B., Benowitz,N.L., Dolan,M.E., Flockhart,D.A., Johnson,J.A., Hayes,D.F., Klein,T., Krauss,R.M. *et al.* (2007) The pharmacogenetics research network: from SNP discovery to clinical drug response. *Clin. Pharmacol. Ther.*, **81**, 328–345.
19. Jaja,C., Burke,W., Thummel,K., Edwards,K. and Veenstra,D.L. (2008) Cytochrome p450 enzyme polymorphism frequency in indigenous and native American populations: a systematic review. *Community Genet.*, **11**, 141–149.
20. Lewis,D.F., Ito,Y. and Goldfarb,P.S. (2006) Structural modelling of the human drug-metabolizing cytochromes P450. *Curr. Med. Chem.*, **13**, 2645–2652.
21. Michalsky,E., Goede,A. and Preissner,R. (2003) Loops In Proteins (LIP)—a comprehensive loop database for homology modelling. *Protein Eng.*, **16**, 979–985.
22. Montellano,P.R.O.d. (2004) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edn. Springer, Netherlands.
23. Nelson,D.R. (2006) Cytochrome P450 nomenclature, 2004. *Methods Mol. Biol.*, **320**, 1–10.
24. Geer,L.Y., Marchler-Bauer,A., Geer,R.C., Han,L., He,J., He,S., Liu,C., Shi,W. and Bryant,S.H. (2009) The NCBI BioSystems database. *Nucleic Acids Res.*
25. Winnenburg,R., Plake,C. and Schroeder,M. (2008) Mutation tagging with gene identifiers applied to membrane protein stability prediction. *Proceedings of the ECCB 2008 Workshop: Annotation, Interpretation and Management of Mutations (AIMM).*
26. Guha,R., Howard,M.T., Hutchison,G.R., Murray-Rust,P., Rzepa,H., Steinbeck,C., Wegner,J. and Willighagen,E.L. (2006) The Blue Obelisk-interoperability in chemical informatics. *J. Chem. Inf. Model*, **46**, 991–998.
27. Goede,A., Dunkel,M., Mester,N., Frommel,C. and Preissner,R. (2005) SuperDrug: a conformational drug database. *Bioinformatics*, **21**, 1751–1753.
28. Lee,S.J., Perera,L., Coulter,S.J., Mohrenweiser,H.W., Jetten,A. and Goldstein,J.A. (2007) The discovery of new coding alleles of human CYP26A1 that are potentially defective in the metabolism of all-trans retinoic acid and their assessment in a recombinant cDNA expression system. *Pharmacogenet. Genomics*, **17**, 169–180.
29. Arnold,K., Bordoli,L., Kopp,J. and Schwede,T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.
30. Sussman,J.L., Lin,D., Jiang,J., Manning,N.O., Prilusky,J., Ritter,O. and Abola,E.E. (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 1078–1084.
31. Kusus,M., Stapleton,D.D., Lertora,J.J., Simon,E.E. and Dreisbach,A.W. (2000) Rhabdomyolysis and acute renal failure in a cardiac transplant recipient due to multiple drug interactions. *Am. J. Med. Sci.*, **320**, 394–397.

Nucleic Acids Research Advance Access published October 15, 2010

Nucleic Acids Research, 2010, 1–8
doi:10.1093/nar/gkq910

CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge

Jessica Ahmed, Thomas Meinel, Mathias Dunkel, Manuela S. Murgueitio, Robert Adams, Corinna Blasse, Andreas Eckert, Saskia Preissner and Robert Preissner*

Charité – University Medicine Berlin, Institute for Physiology, Structural Bioinformatics Group, Thielallee 71, 14195 Berlin, Germany

Received August 12, 2010; Revised September 21, 2010; Accepted September 23, 2010

ABSTRACT

During the development of methods for cancer diagnosis and treatment, a vast amount of information is generated. Novel cancer target proteins have been identified and many compounds that activate or inhibit cancer-relevant target genes have been developed. This knowledge is based on an immense number of experimentally validated compound–target interactions in the literature, and excerpts from literature text mining are spread over numerous data sources. Our own analysis shows that the overlap between important existing repositories such as Comparative Toxicogenomics Database (CTD), Therapeutic Target Database (TTD), Pharmacogenomics Knowledge Base (PharmGKB) and DrugBank as well as between our own literature mining for cancer-annotated entries is surprisingly small. In order to provide an easy overview of interaction data, it is essential to integrate this information into a single, comprehensive data repository. Here, we present CancerResource, a database that integrates cancer-relevant relationships of compounds and targets from (i) our own literature mining and (ii) external resources complemented with (iii) essential experimental and supporting information on genes and cellular effects. In order to facilitate an overview of existing and supporting information, a series of novel information connections have been established. CancerResource addresses the spectrum of research on compound–target interactions in natural sciences as well as in individualized medicine; CancerResource is available at: http://bioinformatics.charite.de/cancerresource/.

INTRODUCTION

Drug–protein interactions, or more generally, compound–target interactions, are becoming increasingly available for several layers of information according to the different interests in biological, physical or pharmacological research. Consequently, a broad set of data resources have been established and it is therefore not easy for biological, chemical or pharmaceutical scientists to deal with the often widespread and vast amounts of data. However, it is straightforward to use the capability of the Internet (1)—this includes up-to-date techniques like Web Services (2) to access existing repositories—for discovering compound–target interactions or determining the druggability of genes.

CancerResource addresses the complexity of cancer by covering not only a large but specific set of compound–target interactions, experimental data and supporting information but also by allowing individual data to be processed for advanced analyses. This article describes the database content and access to stored data together with the usage of provided tools and tool combinations toward workflows.

CANCER LITERATURE AND TEXT MINING

In the past three decades, huge effort was spent on research into cancer by an overwhelming number of

*To whom correspondence should be addressed. Tel: +49 30 450540755; Fax: +49 30 450 540 955; Email: robert.preissner@charite.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

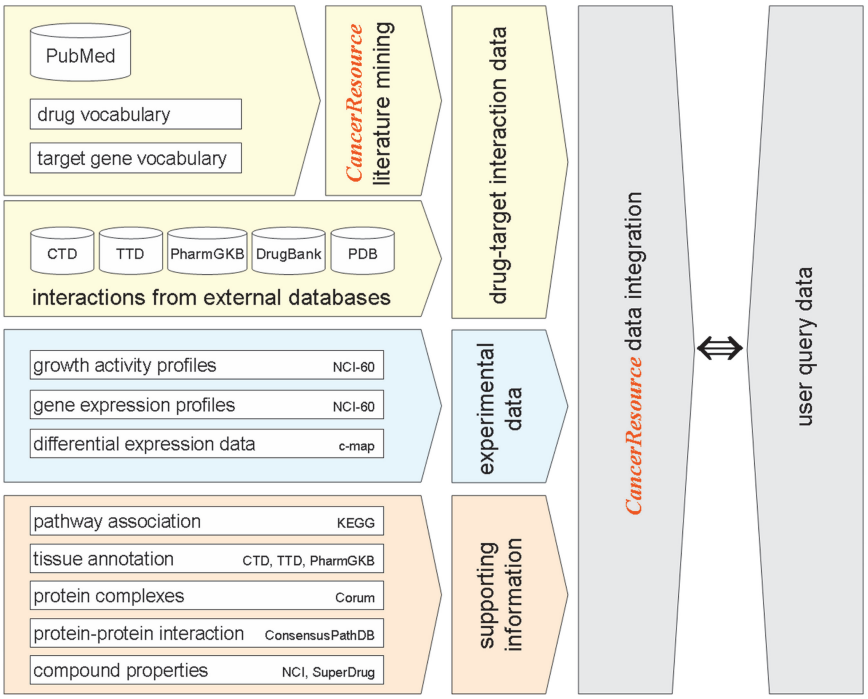


Figure 1. Data integration in CancerResource with three levels of data information: compound–target interaction data- PubMed (14), CTD (8), TTD (10), PharmGKB (9), DrugBank (11), PDB (18) -, experimental data - DTP at NCI (12), Connectivity Map (31) - and supporting information. - KEGG (29), CORUM (24), ConsensusPathDB (23), SuperDrug (15). CancerResource can be explored with queries or user-defined external data.

single studies. Literature on genetic disorders in cancer is extracted and made available in the Online Mendelian Inheritance in Man (OMIM) (3) database. Multiple data collections have arisen from the available repertoire of knowledge on cancer by text mining. They are often specialized like the Catalogue Of Somatic Mutations in Cancer (COSMIC) (4), a web resource on mutations in cancer genes that are detected in somatic tissues but also in cultured tissue samples. Cancer-relevant genes have been intensively studied and a fundamental model of cancer was established by Hanahan and Weinberg (5). On the other hand, the druggable genome (6) is, independent from the kind of disease, a set of proteins that are regarded as possible drug target candidates. Genome-scale targeting was identified originally by literature mining but has been successively developed by adding other information resources (7). The overlap between both perspectives, cancer-specific genes and druggable genome entities, forms the theoretical background of the CancerResource approach to the target genes; practically, target genes are derived from literature mining approaches.

Existing repositories like the Comparative Toxicogenomics Database (CTD) (8), the Pharmacogenomics Knowledge Base (PharmGKB) (9), the Therapeutic Target Database (TTD) (10) and the DrugBank (11) provide rich information on interactions of drugs (or drug-like compounds) with target genes or proteins. After inspecting cancer-relevant compound–target interactions, we found, surprisingly, that the data sets of these resources are more or less disjunct (Table 1) even when the results of the CancerResource literature mining are considered. This analysis indicated that there is need for integrating compound–target interactions from external data sets into one source and hence stimulated the creation of the CancerResource.

Table 1. Numbers of known interactions in external databases and from the CancerResource literature text mining

Resource	References	Numbers of compound–target relationships		Uniqueness
		Redundant	Unique	Degree (%)
External databases				
CTD	(8)	3875	3748	96
PharmGKB	(9)	1307	1158	88
TTD	(10)	282	163	58
DrugBank	(11)	4949	4763	96
CancerResource				
Literature mining	(this article)	1122	992	88
CancerResource				
Full data integration	(this article)	11 585	10 824	93

Data from CTD, PharmGKB and TTD are filtered according to cancer-related disease annotations, data for DrugBank are unfiltered. Relationships unique to each approach include the CancerResource literature mining result. The full integration result is presented additionally. The degree of uniqueness reveals that the data sets are more or less disjunct.

Cancer is often studied using somatic tissues, which are cultured for research as tissue samples of various cancer types and established as human standard cell lines. This inhomogeneous spectrum of cancers is well characterized and analyzed in large experimental studies investigating gene expression or cell growth activity under the influence of chemicals (12). This compound set of the National

Cancer Institute (NCI) is a rich resource for knowledge and research on gene dysfunctions in cancer. A data integration tool like CancerResource demands extended functionality. It is obvious that, similar to other compound–target interaction resources presented in a toxicological perspective (13), additional data such as experimental results and further supporting information enhance the knowledge of interactions together with features like: relationship of genes in pathways, druggability of the genes in the interactome, capability for user-defined data analyses and data mining and curation.

DATA INTEGRATION PROCEDURES AND METHODS

Compound–target gene interactions: CancerResource text mining

Compound–target relationships were automatically detected by own literature text mining over 19 million PubMed (14) abstracts using our vocabularies for drugs and targets. The drug vocabulary was generated from compounds having a cancer-related classification with respect to the Anatomical Therapeutic Chemical (ATC) Classification system via SuperDrug (15) or if the compound and its synonymous name are in the NCI compound set. The cancer relationship of a gene was determined from annotations in cancer-related pathways (see sub-section 'KEGG pathways') and the Gene Ontology (GO) (16). Abstracts, titles and Medical Subject Headings (MeSH) terms were converted into a text index using the LingPipe (http://alias-i.com/lingpipe/index.html) and the Lucene software packages (17). Both vocabularies were searched against each indexed abstract and the result was scored by an own rule-based validation algorithm. After this automatic procedure and a subsequent ranking revealing about 8000 publications, a manual revision of the hits followed resulting in about 900 highly significant publications of direct interactions.

Compound–target gene interactions: more data

Important interaction resources are integrated in CancerResource: CTD, TTD and PharmGKB. Sub-sets of cancer-specific interactions are filtered out according to the cancer vocabulary that is inherent in the three resources. The cancer-specific vocabulary is searchable and consists of more than 400 redundant cancer expressions. These are grouped into about 30 (mostly tissue-related) categories. To explore the impact of a particular drug on genes that are not just connected with cancer we integrated cancer-unspecific information on interactions provided by DrugBank (11). For ligands that are entries in the NCI compound set ligand–protein interactions from the Protein Data Bank (PDB) (18) were integrated into CancerResource.

PubMed references are extracted for identified compound–target relations to be cited in the web interface (if available; otherwise the relation is referenced by linking to the data resource by the resource's identifier).

(Drug-like) compounds and target genes

Core information of compounds and drugs was collected from different databases like the Developmental Therapeutics Program (DTP) at the NCI (12), PubChem (19), SuperTarget (20) and SuperDrug (15). CancerResource contains more than 40 000 cancer-relevant compounds.

The current set of target genes or proteins with cancer relevance are confirmed by the own text mining and complemented by genes extracted from existing interaction databases. The drug association is generally given (and searchable) at gene level and, if available, additionally at protein level. Core information on proteins and genes is based on UniProt (21) and Ensembl (22). Supporting information on cancer-relevant genes or proteins are provided by ten of thousands protein–protein interactions from ConsensusPathDB (23), affiliation of proteins to more than a thousand protein complexes from the Comprehensive Resource of Mammalian protein complexes (CORUM) (24); hundreds of gene mutations in NCI-60 tissue samples from COSMIC (4) and information by Web Service requests or virtual data links to iHOP, Reactome, Pfam and SYSTERS (25–28), see also Supplementary Table S1.

KEGG pathways

To put compound–target relations into a cellular context, we analyzed KEGG (signaling) pathways (29) according to their relevance in cancer emergence and cancer development. Forty four KEGG pathways were integrated into the CancerResource environment. This set comprises cancer-specific pathways, pathways related to cell-cycle regulation, replication, immune response and drug metabolism. Pathway maps are dynamically retrieved via Web Service from KEGG facultative with highlighted expression data if gene expression is computed online before. KEGG genes were excerpted from the set of analyzed pathways and used in the gene vocabulary for the text mining.

Cancer cell lines

Sixty human cancer cell lines of the NCI (NCI-60 set) were selected with respect to the availability of expression data as well as data of changes in biological activity by compound treatment. (Human cancer cell lines and cancer types are described in the Supplementary Data.)

Biological activity profiles: cellular fingerprints

Biological activity profiles indicate the influence of compounds on the growth rates of human cancer cell lines, wherein a GI-50 value indicates the compound concentration that induces 50% growth inhibition after treatment. More than 40 000 biological activity profiles are obtained for each compound. All activity profiles are translated into cellular fingerprints which allow the fast computation (30) of profile differences.

Gene expression data

Expression data of NCI-60 cancer cell lines were retrieved from DTP at the NCI and re-calculated to be comparable to external data sets in three steps (see Supplementary Data). Over the whole microarray data set (Affymetrix U133A chips), we introduced (i) the median normalization on Affymetrix probe set expression and (ii) compared normalized expression values of each probe set across NCI-60 tissue samples by introducing the relative abundance over all 60 cancer cell lines. Expression intensities of probe sets are ignored if they are associated with multiple genes. For each gene that is, according to Ensembl, associated with multiple probe sets (iii) the average of respective expression intensities is calculated.

Differential expression of genes after treatment

The Connectivity Map (31) provides differential expression data for five human cancer cell lines from the NCI-60 set before and after treatment with more than 200 compounds. Data correspond to the Gene Expression Omnibus (GEO) (32) data set GSE5258 and ratios are retrieved by Web Service from the GenomeMatrix repository (33), see also a detailed description in (34).

RESULTS

Currently, CancerResource comprises more than 10 800 non-redundant compound–target relations. More than 6000 (56%) are associated with cancer and over 4700 relations from DrugBank that do not have a disease specification. However, integration of DrugBank data enables high-quality searches for alternative targeting, which is, in the context of pharmacogenomic research, also known as drug repositioning. The CancerResource literature text mining revealed 992 new compound–target interactions (Table 1), which are ~16% of the cancer-related drug–target interactions or ~10% of all unique interactions in CancerResource. This ostensibly low number is owing to mining abstract texts only. Even after integration of our text mining results, the degree of uniqueness for the CancerResource is still ~90%, which indicates that all four text mining strategies with focus on cancer are obviously different to each other. In the whole CancerResource interaction data set, 2392 cancer-related target genes from CancerResource text mining, CTD, PharmGKB and TTD and additionally 995 genes from DrugBank cover 30% of the druggable genome (7); additionally 728 cancer-related genes not present in the druggable genome are found having compound–target interactions. (More issues and numbers on integrated data can be inspected in Supplementary Table S2.)

The integration of the set of more than 40 000 NCI compounds, dedicated as experimental drugs, extends the set of Food and Drug Administration (FDA) approved drugs by a factor of 100. It enriches CancerResource as an information resource for better understanding cancer and its treatment with drugs with a huge experimental background.

MUTUAL ACCESS TO COMPOUNDS AND GENES

CancerResource provides the referencing to interaction literature by links to citations in PubMed. In the web interface, such relations can be accessed by a drug, a target or a cancer feature; each of the three subjects can be used to query the web tool. Both molecular instances, target genes and drugs, can be mutually accessed (see Figure 2a). Respective web pages are organized into three parts that describe in detail (i) the relevance of a drug or a gene to cancer, (ii) compound–target interactions and (iii) supporting information.

At several sites in the web tool, interaction matrices of compounds and target genes provide information on single drugs targeting multiple genes (ambiguity) as well as multiple drugs targeting a single gene (redundancy). Such information on alternative targeting, which is helpful for the potential repositioning of compounds, is amplified through the integration of non-specific interaction information by DrugBank entries.

ACCESS TO EXPERIMENTAL DATA

Compound–target interaction information is more valuable by integrating experimental and supporting information. Therefore, CancerResource provides experimental data in addition to the information on interactions. Thereby, data stored in CancerResource can be compared with the user's own data. Several ways for accessing the web tool are presented in this section.

Similarity of compounds by structure or biological activity

The influence of a compound on the growth of cancer cell lines is a frequently used approach for the characterization and development of drugs. The biological activity of two compounds across the 60 NCI cell lines can be compared by a similarity measure, the Tanimoto coefficient of cellular fingerprints (30); this comparison by biological characteristics of a compound is a strong feature of the CancerResource web tool that complements the comparison of compounds by 2D structures. Here, the Tanimoto coefficient of structural fingerprints (35) enables the comparison of 2D similarities independently from the biological activity of a compound. CancerResource suggests thereby substitutability, alternative compound applications and support thereby drug research and drug treatment. Similar compounds are searchable by a given activity profile or the profile of a particular compound (query options are given in Figure 2b). Moreover, activity profiles can be found for a given compound structure.

Most active drugs against a cancer cell line

Alternatively to the compound characteristic defined by all cancer cell lines CancerResource enables the searching for compounds that are most biologically active against a single cell line (second part in Figure 3). In clinical medicine, one of the most successful approach to treat cancer is the growth inhibition of the cancer tissue. Therefore, CancerResource implies a module to find

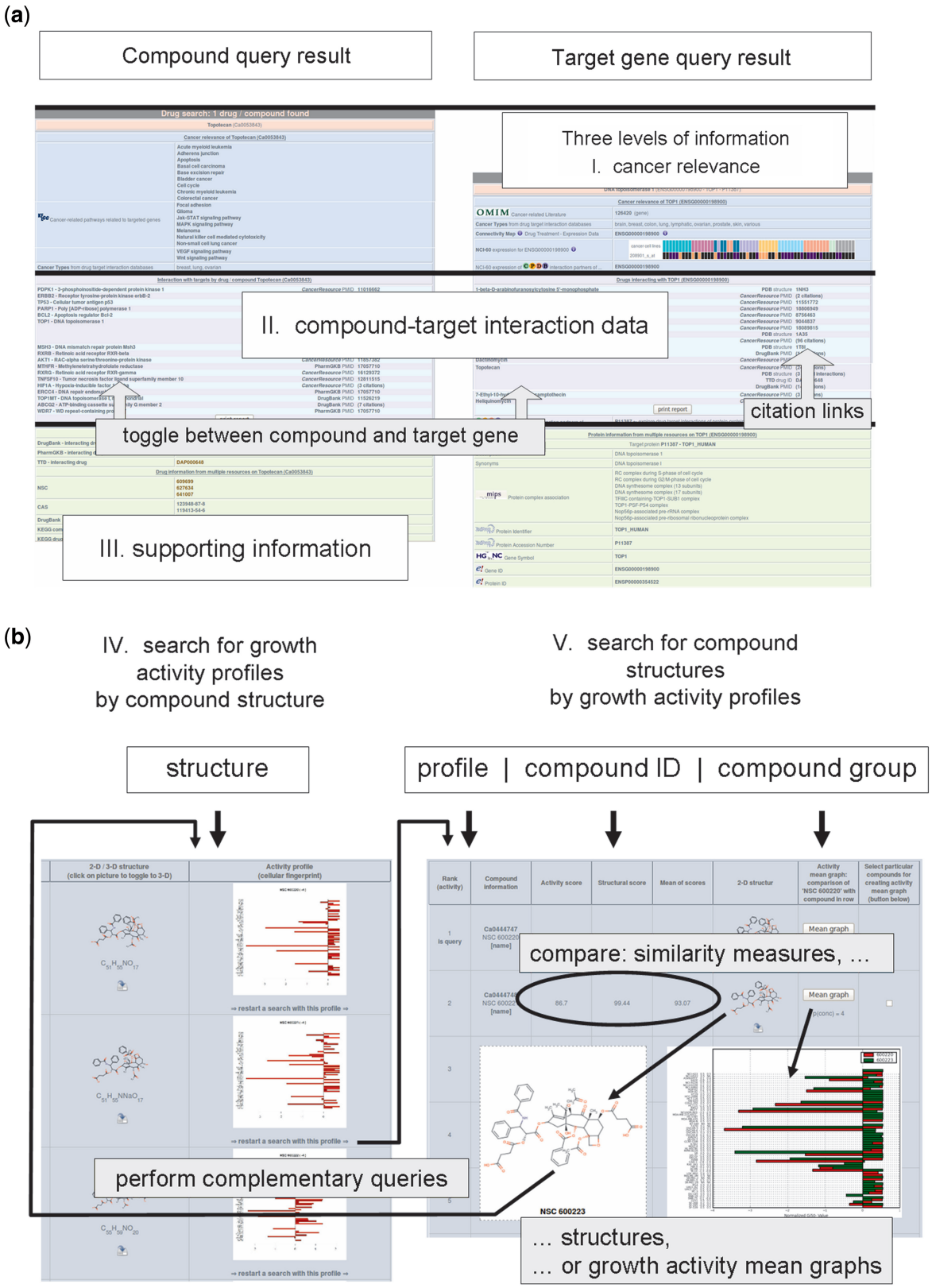


Figure 2. Knowledge retrieval in CancerResource: (**a**) Access to detailed compound/drug and target gene information in CancerResource. A similar layout for both information layers, compounds (left) and target genes (right), comprises three information sections: (i) cancer relevance of a target gene or a compound with KEGG cancer pathways, involved somatic cancer types, information on expression across cancer cell lines; (ii) information on interactions with a toggle option between compound and target gene, source of information and link to the original literature source; (iii) specific compound or gene information. (**b**) Access to complementary information on growth activity across NCI-60 human cancer cell lines and structures of acting compounds. A search by compound structures (iv) reveals similar structures and associated growth activity profiles. The search by activity profiles (v) enables the user to compare structure formulas, activity profiles (pairwise mean graphs) and similarity measures for both growth activity and structures. Complementary queries can be performed by structures after downloading or by implemented links for a profile.

most effective compounds (that are inducing highest inhibition) against a single cell line.

Gene expression data of NCI-60 cancer cell lines

In CancerResource, gene expression data are available for about 4000 genes (see 'Results' section) and 60 NCI-60 cancer cell lines in both dimensions: genes are described and can be compared by expression profiles, the arrays of expression values across cell lines; NCI-60 cell lines are described and can be compared by a profile across genes. Relative abundances (data are calculated online if external data are uploaded) are displayed in the web tool by an array of colored boxes, each corresponding to a single gene. The blue/black/yellow color scheme is used for lower/non-significant/higher expression relatively to the average across all cell lines.

Several entry points for expression data with respect to genes are enabled: genes are searchable by the affiliation of genes to KEGG pathways, affiliation to protein–protein interaction data and for genes with low or high relative abundance in a couple of cancer cell lines. Resulting expression profiles over all 60 cell lines are characteristics for genes. They can be ranked by similarity (Pearson's correlation) if a gene is selected as center; protein–protein interactions and expression profile similarity are combined features here.

Furthermore, the NCI-60 cell line closest to a user-defined expression set (chip experiment; expression data sets are compared by Pearson correlation) can be searched both for genes (or probe sets) of a whole microarray or for selected probe sets or genes.

Differential gene expression

CancerResource allows the genome-wide online validation of two microarray chip experiments by computation of differential expression via ratios. Either external data are compared to a NCI-60 cell line or two external data sets can be compared to each other. Normalization for a subset of genes is regarded as a positive selection feature. It is enabled in CancerResource, which hence supports tumor/normal tissue comparisons or drug-treatment/control experiments. Alternatively, pre-calculated ratios associated with Ensembl gene IDs can be uploaded to enable the import of results from other experiment types (e.g. data collected using other micro-array platforms, next generation sequencing, protein chips, etc).

Ratios for differential expression are displayed in the web tool by the green/black/red color scheme (down/non-significant/up). Arrays of colored boxes are arranged according to the affiliation of respective genes to chromosomes or KEGG pathways. For the latter, differentially expressed genes are analyzed in order to estimate the over-representation in a pathway. This is calculated by a *P*-value using the hypergeometric function and distribution, see details in (36).

Connectivity map

The Connectivity Map (31) was intended to aid the discovery of functional connections among diseases, genetic perturbation and drug action. The influence of more than 200 compounds on differential gene expression was determined for the whole genome of five cancer cell lines. Two query options in CancerResource provide access to expression profiles (i) for the influences of the set of compounds in the five cell lines on a single gene and (ii) for the influence of a single compound on all genes in a single cell line. The visualization, again by arrays of colored boxes, is restricted to target genes that possess interactions integrated in CancerResource.

Direct and indirect knowledge on compound–target interaction

In the Connectivity Map data set, the influence of a row of compounds on genes is experimentally studied by differential expression, which is indirect knowledge about gene targeting (but no about cause-and-effect relationships). The simultaneous comparison (Supplementary Figure S1) with compound–target interactions from the literature mining ('direct knowledge') facilitates considerations about druggability and targeting of genes.

PROPOSED WORKFLOWS

CancerResource facilitates complex searches by the implementation of several ways of accessing the data. Two workflows are demonstrating suggested research use cases.

Finding alternative, most effective drugs for a (somatic) tissue similar to a cancer cell line

An external tissue sample can be identified as most similar to a single NCI-60 cell line by expression profiles across genes or probe sets. Figure 3 explains how the most similar ('best') NCI-60 cell line can be determined with differentially expressed genes by calculation of Pearson correlations between the upload data and all 60 tissues samples (see above). In the next step, the most effective drugs will be determined for this cell line (which is basing on the growth inhibition of a compound is measured for all NCI-60 cancer cell lines and is described above). Finally, for the identified compounds the tool displays the genes they target including the alternative targeting.

Finding alternative compound–target gene interactions for differentially expressed genes via pathway information

KEGG (signaling) pathways elucidate the context of genes according to functionality. To visualize the differential regulation of genes in a pathway, colored pathway maps are dynamically generated in CancerResource. The workflow starts with the loading of expression data (Supplementary Figure S2), which is possible in multiple forms. The data are re-calculated and displayed as an array of colored boxes for each KEGG pathway; overrepresentation analyses are available for each pathway and for both up- and down-regulated genes; the pathway map generation can be started from here to display integrated expression of genes, either for a single gene or for all genes in the pathway. Finally, drug information is available (via the pathway map) and,

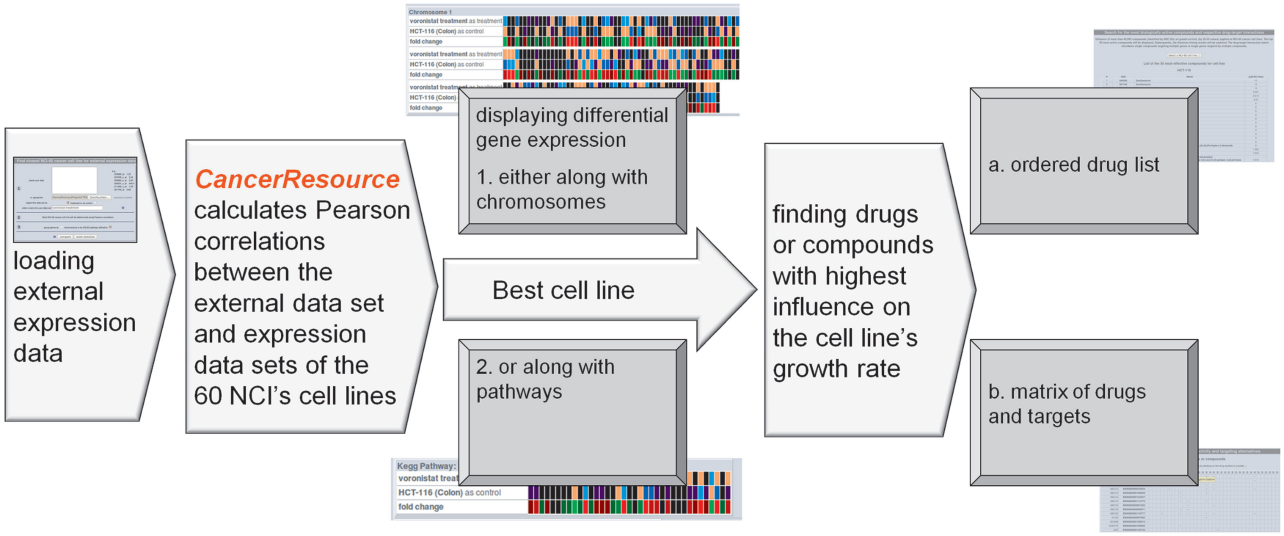


Figure 3. Finding the most similar cell line from the NCI-60 set and, subsequently, compounds or drugs having the highest influence on that cell line (this workflow includes two user interactions).

subsequently, the compound–target matrix for alternative targeting. The integration of dynamically assigned pathway maps makes CancerResource into a systems biology approach.

CONCLUSIONS AND FUTURE DIRECTIONS

The feedback by many scientists shows that there is a need for specialized resources that not only cover a specific set of interaction data but also deliver tools that are specialized for the further analysis of the respective data set. CancerResource tries to cover both levels of scientific work, the support of scientists who try to develop novel drugs and the medic who is reliant on advice for the development of individualized therapy approaches.

Cancers, even of the same tissue type, are extremely divergent in terms of gene alterations. Individual therapy will be made possible by understanding single nucleotide polymorphisms (SNPs), complete or partial gene deletions, copy number variations, gene aberrations or gene fusions. All of those issues may cause substantial dysfunctions of defected genes that have influence on gene regulation in the whole cell of an individual. Additionally to those integration issues, new data integration concepts will be required or are planned to be integrated in CancerResource for coping with personalized therapies. The literature mining will be extended to full text mining, manual upload of single relationships and enhanced specificity in cancer annotations. Expression data will be comparable for platforms other than Affymetrix U133A. Large studies performed on the basis of new techniques (Next Generation Sequencing; e.g. Genetics of 1000 Tumors) are highly interesting objectives to be made available in CancerResource. Updating of data is projected to occur once a year.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors like to thank Björn Grüning for the development of the initial database and Ilona Gawronska and Ulrike Schmidt for manual validation of cancer-relevant drug–target relations.

FUNDING

The International Research Training Group IRTG (Genomics and Systems Biology of Molecular Networks) Deutsche Krebshilfe; Deutsche Forschungsgemeinschaft DFG; Federal Ministry of Education and Research BMBF (MedSys Project); European Union EU (SynSys Project). Funding for open access charge: DFG, BMBF, EU.

Conflict of interest statement. None declared.

REFERENCES

1. Wishart,D.S. (2007) Discovering drug targets through the web. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, **2**, 9–17.
2. Meinel,T. and Herwig,R. (2010) SOAP/WSDL-based Web Services for Biomedicine. In Lazakidou,A. (ed.), *Web-Based Applications in Healthcare and Biomedicine*. Springer, New York, USA, pp. 101–116.
3. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **37**, D793–796.
4. Forbes,S.A., Tang,G., Bindal,N., Bamford,S., Dawson,E., Cole,C., Kok,C.Y., Jia,M., Ewing,R., Menzies,A. *et al.* (2010) COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.*, **38**, D652–D657.
5. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.

6. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.
7. Sophic Alliance. (2010) *The Integrated Druggable Genome Database*. Sophic Systems Alliance Inc, Rockville, MD, USA.
8. Davis,A.P., Murphy,C.G., Saraceni-Richards,C.A., Rosenstein,M.C., Wiegers,T.C. and Mattingly,C.J. (2009) Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res.*, **37**, D786–D792.
9. Hernandez-Boussard,T., Whirl-Carrillo,M., Hebert,J.M., Gong,L., Owen,R., Gong,M., Gor,W., Liu,F., Truong,C., Whaley,R. *et al.* (2008) The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge. *Nucleic Acids Res.*, **36**, D913–D918.
10. Zhu,F., Han,B., Kumar,P., Liu,X., Ma,X., Wei,X., Huang,L., Guo,Y., Han,L., Zheng,C. *et al.* (2010) Update of TTD: Therapeutic Target Database. *Nucleic Acids Res.*, **38**, D787–D791.
11. Wishart,D.S., Knox,C., Guo,A.C., Cheng,D., Shrivastava,S., Tzur,D., Gautam,B. and Hassanali,M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
12. Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, **6**, 813–823.
13. Mattingly,C.J. (2009) Chemical databases for environmental health and clinical research. *Toxicol. Lett.*, **186**, 62–65.
14. McEntyre,J. and Lipman,D. (2001) PubMed: bridging the information gap. *Can. Med. Assoc. J.*, **164**, 1317–1319.
15. Goede,A., Dunkel,M., Mester,N., Frommel,C. and Preissner,R. (2005) SuperDrug: a conformational drug database. *Bioinformatics*, **21**, 1751–1753.
16. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
17. McCandless,M., Hatcher,E. and Gospodnetić,O. (2010) *Lucene in Action*, 2nd edn. Manning Publications Co, Greenwich, CT, USA.
18. Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
19. Wang,Y., Bolton,E., Dracheva,S., Karapetyan,K., Shoemaker,B.A., Suzek,T.O., Wang,J., Xiao,J., Zhang,J. and Bryant,S.H. (2010) An overview of the PubChem BioAssay resource. *Nucleic Acids Res.*, **38**, D255–D266.
20. Gunther,S., Kuhn,M., Dunkel,M., Campillos,M., Senger,C., Petsalaki,E., Ahmed,J., Urdiales,E.G., Gewiess,A., Jensen,L.J. *et al.* (2008) SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, **36**, D919–D922.
21. UniProtConsortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
22. Flicek,P., Aken,B.L., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S. *et al.* (2010) Ensembl's 10th year. *Nucleic Acids Res.*, **38**, D557–D562.
23. Kamburov,A., Wierling,C., Lehrach,H. and Herwig,R. (2009) ConsensusPathDB: a database for integrating human functional interaction networks. *Nucleic Acids Res.*, **37**, D623–D628.
24. Ruepp,A., Brauner,B., Dunger-Kaltenbach,I., Frishman,G., Montrone,C., Stransky,M., Waegele,B., Schmidt,T., Doudieu,O.N., Stumpflen,V. *et al.* (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.*, **36**, D646–D650.
25. Hoffmann,R. and Valencia,A. (2004) A gene network for navigating the literature. *Nat. Genet.*, **36**, 664.
26. Matthews,L., Gopinath,G., Gillespie,M., Caudy,M., Croft,D., de Bono,B., Garapati,P., Hemish,J., Hermjakob,H., Jassal,B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
27. Meinel,T., Krause,A., Luz,H., Vingron,M. and Staub,E. (2005) The SYSTERS Protein Family Database in 2005. *Nucleic Acids Res.*, **33**, D226–D229.
28. Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Pollington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Forslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
29. Kanehisa,M., Goto,S., Hattori,M., Aoki-Kinoshita,K.F., Itoh,M., Kawashima,S., Katayama,T., Araki,M. and Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
30. Fullbeck,M., Dunkel,M., Hossbach,J., Daniel,P.T. and Preissner,R. (2009) Cellular fingerprints: a novel approach using large-scale cancer cell line data for the identification of potential anticancer agents. *Chem. Biol. Drug Des.*, **74**, 439–448.
31. Lamb,J., Crawford,E.D., Peck,D., Modell,J.W., Blat,I.C., Wrobel,M.J., Lerner,J., Brunet,J.P., Subramanian,A., Ross,K.N. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929–1935.
32. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
33. Hewelt,A., Ben Kahla,A., Hennig,S., Nagel,A., Himmelbauer,H., Zehetner,G., Haas,S., Vingron,M., Yaspo,M.L. and Lehrach,H. (2002) *Human Genome Meeting HGM2002*. Shanghai, China, Abstract 23.
34. Meinel,T., Mueller,M.S., Ahmed,J., Yildirimman,R., Dunkel,M., Herwig,R. and Preissner,R. (2010) SOAP/WSDL-based web services for biomedicine: demonstrating the technique with the CancerResource. In Bamidis,P.D. and Pallikarakis,N. (eds), *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*. Springer, Berlin Heidelberg, Chalkidiki, Greece, pp. 835–838.
35. Thimm,M., Goede,A., Hougardy,S. and Preissner,R. (2004) Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J. Chem. Inf. Comput. Sci.*, **44**, 1816–1822.
36. Masseroli,M. and Tagliasacchi,M. (2010) Web resources for gene list analysis in biomedicine. In Lazakidou,A. (ed.), *Web-Based Applications in Healthcare and Biomedicine*. Springer, New York, USA, pp. 117–141.

Nucleic Acids Research Advance Access published October 14, 2010

Nucleic Acids Research, 2010, 1–6
doi:10.1093/nar/gkq917

SuperSweet—a resource on natural and artificial sweetening agents

Jessica Ahmed, Saskia Preissner, Mathias Dunkel, Catherine L. Worth, Andreas Eckert and Robert Preissner*

Charité – Universitätsmedizin Berlin, Institute of Physiology, Structural Bioinformatics Group, Lindenberger Weg 80, 13125 Berlin, Germany

Received August 15, 2010; Revised September 21, 2010; Accepted September 24, 2010

ABSTRACT

A vast number of sweet tasting molecules are known, encompassing small compounds, carbohydrates, D-amino acids and large proteins. Carbohydrates play a particularly big role in human diet. The replacement of sugars in food with artificial sweeteners is common and is a general approach to prevent cavities, obesity and associated diseases such as diabetes and hyperlipidemia. Knowledge about the molecular basis of taste may reveal new strategies to overcome diet-induced diseases. In this context, the design of safe, low-calorie sweeteners is particularly important. Here, we provide a comprehensive collection of carbohydrates, artificial sweeteners and other sweet tasting agents like proteins and peptides. Additionally, structural information and properties such as number of calories, therapeutic annotations and a sweetness-index are stored in SuperSweet. Currently, the database consists of more than 8000 sweet molecules. Moreover, the database provides a modeled 3D structure of the sweet taste receptor and binding poses of the small sweet molecules. These binding poses provide hints for the design of new sweeteners. A user-friendly graphical interface allows similarity searching, visualization of docked sweeteners into the receptor etc. A sweetener classification tree and browsing features allow quick requests to be made to the database. The database is freely available at: http://bioinformatics.charite.de/sweet/.

INTRODUCTION

There are three major compounds of life: proteins; lipids and carbohydrates. The perception of sweet taste, mainly associated with advantageous food, has had an important evolutionary influence on different physiological regulation mechanisms. During human development, sugar was always luxury. In 1885 Constantin Fahlberg produced the first artificial sweetener, saccharin, and the scientific establishment was surprised by its extreme sweetness (1). Significant to this discovery was the fact that sweet taste became affordable to poor people. Following the commercial success of artificial sweeteners, a battle between the sugar and sweetener industries began (2). Saccharin was claimed to be carcinogenic in rats (3). However, it was later shown that saccharin is neither toxic nor carcinogenic in normal amounts (4), yet its reputation remains tarnished. Today, the replacement of sugar and other carbohydrates with artificial sweeteners in food is common (5) and is a general approach to prevent cavities (6,7), obesity and associated diseases such as diabetes and hyperlipidemia (8,9).

Currently, the sweet taste receptor, which is a heterodimer of two transmembrane proteins (T1R2 and T1R3) and has several different binding sites, has not been crystallized and is therefore unavailable in the Protein Data Bank (PDB) (10). Such a structure is crucial to elucidating how both small sweeteners and molecules as large as proteins bind and activate the sweet taste receptor (11). In the meantime, modeling studies can provide vital clues to these mechanisms (12). The understanding of compounds binding to the receptor is of relevance not only for the development of new artificial sweeteners but also for improving our understanding of known sweet molecules and what makes them 'sweet'.

The first publicly available carbohydrate database was CarbBank (13), where users are able to search for carbohydrate structures, sub-structures and non-carbohydrate substituents. Wilhelm von der Lieth established the SweetDB (14), a web-based interface for glycoscientists, which was the basis for further carbohydrate tools collected in the Glycosciences portal (15) and the Glycome-DB (16) that comprises 35 000 carbohydrate sequences with a variety of query options.

*To whom correspondence should be addressed. Tel: +49 30 450 540 755; Fax: +49 30 450 540 955; Email: robert.preissner@charite.de

There are also a number of databases that deal with glycans. GlycoBase (17) and GlycoEctractor (18) are databases that assist with interpreting high-performance liquid chromatography-glycan profiles. Tyrian Diagnostics used text-mining to develop the GlycoSuiteDB (19), which stores over 7650 glycan structures extracted from 740 papers. The Glycoconjugate Data Bank (20) provides a special tool for N-glycan primary structure verification. The connection to metabolic pathways is provided by KEGG-Glycan (21).

Although, there are a number of resources available with relation to carbohydrates, they are lacking with respect to sweetness and sweeteners. SuperSweet aims to integrate knowledge about the structure of sugars and sweetening agents with receptor binding poses, chemical properties and additional information like sweetness, approval, origin, therapeutic effect and metabolism.

THE DATABASE

The SuperSweet database was developed for researchers and dieticians and offers a user-friendly interface with helpful examples and FAQs. Currently, SuperSweet comprises more than 8000 carbohydrates, proteins, D-amino acids and artificial (synthesized) sweeteners, which were retrieved from the literature and different pre-existing data sources like Pubchem (22) and the PDB. Similarity searches extended and completed the sweetening agent data set. Besides information about the physicochemical properties of the sweet compounds, the database also offers information about the number of calories, the 3D structure, therapeutic annotations and, if detectable, the sweetness of the molecule. Structural information is available and displayed for each sweet molecule and sweet protein in the database. Moreover, the domain containing the small molecule active site of the sweet receptor was homology modeled and provided in SuperSweet (Figure 1). The small molecules were docked into the modeled binding site and the poses are also stored in the database.

There are different options for browsing through the database and for retrieving the data. First, the data can be retrieved by name, physicochemical properties or properties such as calories and sweetness. Secondly, the user is allowed to upload or draw a molecule using the Marvin Sketch plugin (http://www.chemaxon.com). The query structure is compared with the entries of SuperSweet and the results presented in a table comprising molecules and a Tanimoto coefficient expressing their similarity. Thirdly, a sweet-tree is available in SuperSweet that allows the fast and easy selection of a group of sweet tasting molecules (Table 1). Here, the user can find for example, all sweet tasting proteins, or peptides or small molecules like Flavonoids. Finally, SuperSweet offers a browse section, which provides an easy way to access the SuperSweet entries by choosing different categories of molecules based on properties.

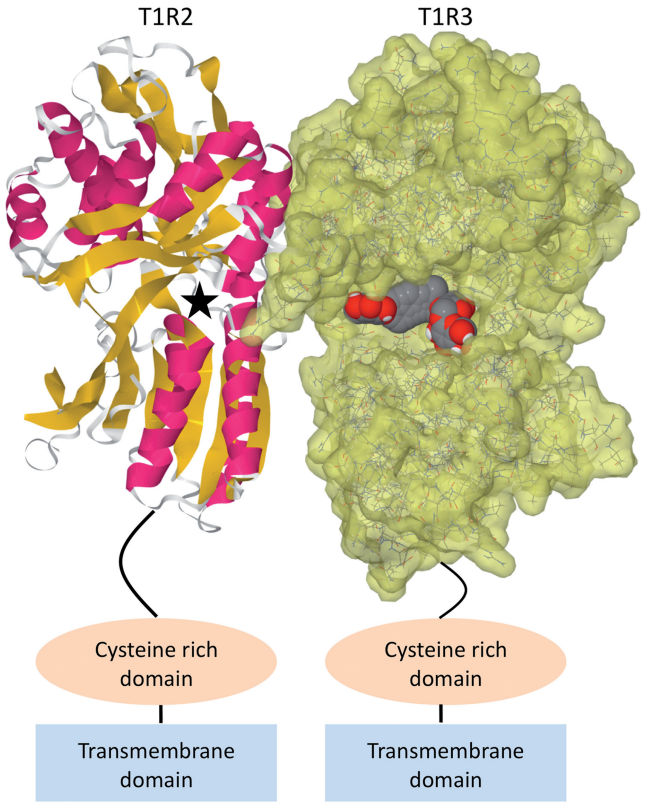


Figure 1. Homology model of the sweet taste receptor with the sweetening agent Stevioside docked. The T1R2 protomer is displayed in cartoon format and the T1R3 protomer is displayed in wireframe format with a solvent accessible surface rendered (1.2Å probe radius) in transparent yellow. The stevioside molecule is displayed in spacefill format and colored according to atom type. Stevioside was docked into the open protomer, T1R3, as the closed protomer, T1R2, is too small to host large sweeteners. The pocket in the closed protomer is situated on the opposite side to that with stevioside bound—the approximate position is indicated by a black star. The cysteine-rich and transmembrane domains are shown schematically and are not to scale.

MATERIALS AND METHODS

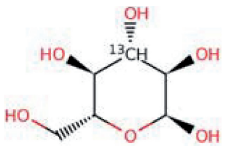
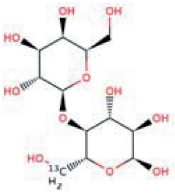
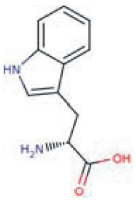
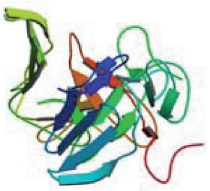
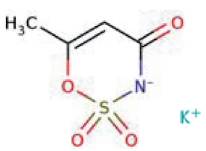
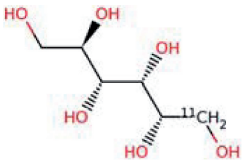
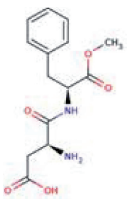
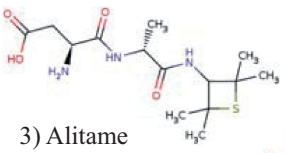
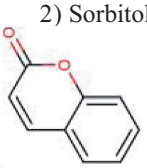
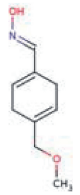
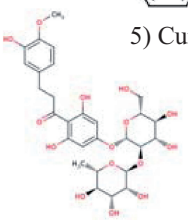
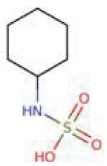
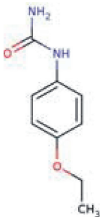
Data acquisition

The sweet tasting molecules were extracted from the literature and publicly available databases like Pubchem, the PDB and MonoSaccharideDB and were filtered using different terms like 'sweetening agents'. In the next step the data set was extended by using similarity search methods.

Homology modeling of the sweet taste receptor

A model of the large extracellular domain of the sweet taste receptor, which contains two binding sites for small molecular-weight sweeteners, was built using homology modeling. The sweet taste receptor is a class C (or metabotropic) G-protein coupled receptor and exists as a heterodimer (consisting of T1R2 and T1R3) (Figure 1). PSI-BLAST searches (23) of the PDB revealed that T1R2 and T1R3 share ~25% sequence identity with the metabotropic glutamate receptors and are in close proximity to one another on the phylogentic tree of class C

Table 1. Organization of the Sweet-tree

Main class	Subclass	Examples
Carbohydrates	1) Monosaccharides, 2) Disaccharides, 3) Polysaccharide, 4) Sugar alcohols	 1) Dextrose  2) Lactose  3) Inulin
Peptides	1) Amino acids, 2) Proteins	 1) D-Tryptophane  2) Curculin
Small molecules	1) Acesulfames, 2) Alditols, 3) Alitames/aclames, 4) Aspartames, 5) Cumarins, 6) Cyclamat-like 7) Dulcines, 8) Flavonoids, 9) Guanidine 10) Carbonate, 11) Nucleotides, 12) Saccharines, 13) Saponins, 14) Steviol 15) Glycosides, 16) Terpens, 17) Others	 1) Acesulfam-K  2) Sorbitol  4) Aspartame  3) Alitame  5) Cumarin  16) Perillartine  8) Neohesperidin-dihydrochalcone  6) Cyclamic acid  7) Dulcin  12) Saccharin

The first column presents the three main classes: carbohydrates, peptides and small molecules. The second column shows the subclasses of the main class. The third column shows some examples of the subclasses

GPCRs (24). As we wanted to use the homology model of the sweet taste receptor for docking studies, it was important that we chose a template structure that is in an active (open-closed) conformation, preferably with a natural ligand bound. Accordingly, an active form of metabotropic glutamate receptor 1 (mGluR1) was used as the template (PDB code: 1EWK) (25), which also had the highest sequence coverage and the highest resolution (2.2Å) compared to other crystal structures of activated glutamate receptors. A multiple sequence alignment was created using MUSCLE (26). The alignment can be downloaded from the SuperSweet website. Homology modeling was carried out using Modeller (27) in Accelrys Discovery Studio 2.5. T1R2 was built using chain A of 1EWK (closed form) and T1R3 using chain B (open form) (12). The large insertions in T1R2 and T1R3 compared to the template

were removed from the final model. Lastly, side-chain clashes were removed and the structure was minimized by carrying out 100 steps of both steepest descent and conjugant gradient minimization.

Generation of the binding poses of the small molecules

Docking of the small compounds into the homology modeled receptor was done using the docking program GOLD 4.1.1 (28). In order to define the binding site of the sweet taste receptor, the template structure (mGluR1 containing a glutamate bound to each chain) was superimposed onto the homology model of the sweet taste receptor and the glutamate molecules copied over to the homology model. The binding sites of the sweet taste receptor were then defined by using the glutamate molecules as reference ligands; all atoms within 5Å of the glutamate molecule formed the binding sites for the docking experiments. For each small molecule, 100 docking runs were performed. A previous docking study showed that the sweet taste receptor's active site in the closed protomer is too small to host some of the larger synthetic sweeteners and is only able to host four compounds out of those tested: saccharin, alitame, aspartame and 6-Cl-tryptophan (12). Experimental work has shown that aspartame and neotame bind to the T1R2 subunit (29). In accordance with these findings, we therefore docked molecules with a molecular weight >400 kDa into T1R3 (open form) and all other molecules into the pockets of both T1R2 and T1R3. The resulting docking poses were then ranked using the GoldScore fitness function. The best scoring docking pose for each molecule can be viewed using a Jmol applet and the respective structure files are also available for download.

Conformer generation

For the small sweetening molecules, conformers were generated using the Accelrys tool. For each small molecule 20 conformers are stored and available for download on the website (30).

Similarity search

For the similarity search in the SuperSweet database, we implemented a bit vector 'structural fingerprint', which encodes the chemical and topological characteristics of a molecule. The fingerprint was pre-calculated for the small molecules of SuperSweet and is also calculated for the query structure, in order to compare it to the database entries. Open Babel implements four different fingerprints (FP2, FP3, FP4 and MACCS). Fingerprint 2 (FP2) is widely used for the comparison of small molecules and is path-based and indexes linear molecules up to seven atoms. However, this fingerprint is not ideal for use in SuperSweet due to its inability to distinguish between different ring structures and therefore between carbohydrates. To overcome this problem, we implemented a combinatorial fingerprint of fingerprint 2 and fingerprint 4. Fingerprint 4 is based on a set of SMARTS patterns and also considers functional groups. For similarity searching the Tanimoto coefficient is used, which gives values in the range of zero (fingerprints have no bits in common) to identical (all bits the same).

Server

SuperSweet is designed as a relational database on a MySQL server. Additionally, the MyChem package (http://mychem.sourceforge.net/) is installed to provide a complete set of functions for handling chemical data within MySQL. Most of the functions used by MyChem depend upon Open Babel (http://openbabel.sourceforge.net/). The structural fingerprint is implemented in Open Babel. To allow the upload or drawing of a query structure, the Marvin Sketch plugin (http://www.chemaxon.com) was installed. For the visualization of the 3D structures Jmol (http://www.jmol.org/) was installed. The website is built with PHP and web access is enabled via Apache HTTP Server 2.2.

EXAMPLE OF USE

Searching for a natural sweetening agent (search field 'Origin') with molecular weight between 800 and 900 and sweetness above 200 returns Stevioside. Steviol glycosides like Rebaudioside A or Rubusoside are non-calorific sweeteners that are found, for instance, in sweet Chinese tea (*Rubus suavissimus*) and *Stevia rebaudiana* (31). These compounds are of research interest because advantageous effects were observed regarding cancer and blood pressure (32). These effects seem to be the result of binding to other membrane proteins (33). Clicking on the protein icon in the results table shows the docking pose of Stevioside to the sweet receptor (see Figure 1). The structure of Stevioside, the computed conformers, the best docking pose and the modeled receptor structure are downloadable from the website.

More information on Steviol glycosides can be found using the 'Sweet-tree' or by performing a search using the field 'Compound name' on the 'Property search' page. Clicking on the similarity search icon delivers the top 10 similar compounds.

CONCLUSION AND FURTHER DIRECTIONS

SuperSweet compiles information on natural and artificial sweetening agents including their properties such as 3D structure, origin, sweetness, approval, calories etc. and provides hypotheses on their binding to the receptor.

Homology modeling provides a useful means of generating 3D conformations of proteins where experimental structures are not available. For this work, we generated models of the sweet taste receptor using mGluR1. The sequence identity between the receptors is rather low and is within the twilight zone of protein sequence alignments, which makes homology modeling more difficult (34). The quality of our homology model may also be affected by the fact that mGluR1 exists as a homodimer, whereas the sweet taste receptor is a heterodimer. These facts have implications for our docking experiment results due to the strong dependence of docking results on the accuracy of protein structure,

especially in the binding site (35). Although GOLD has consistently been shown to be among the best performing docking algorithms in terms of the accuracy of docking poses, it is less able to distinguish the most native-like pose from all of the generated docking poses (36–38). In SuperSweet we have only made the highest scoring pose available for each docked compound and therefore the accuracy of these docking solutions should be considered in light of the aforementioned limitations in homology models and *in silico* docking.

Unlike the metabotropic glutamate receptors, the sweet taste receptor is predicted to have multiple binding sites (29): (i) two cavities which correspond to the Glu hosting cavities of mGluR1; (ii) a secondary binding site on the surface of the receptor for sweet proteins that corresponds to the wedge model (39) and (iii) overlapping binding sites in the seven transmembrane helix domain for the agonist cyclamate (40) and the inverse agonist lactisole (41). For this work, we only performed docking experiments to the binding sites corresponding to the Glu hosting cavities of mGluR1. These two sites differ in size due to the receptor model being in an open-closed conformation; the binding site in T1R3 is much larger as it is in an open conformation, whereas T1R2 is in a closed conformation. Therefore, we only docked compounds with a molecular weight >400 kDa into T1R3 (open form) and all other compounds into the pockets of both T1R2 and T1R3. In addition, the existence of other binding sites or alternative binding mechanisms cannot be excluded (12). Compared to mGluR1, the additional diversity of compounds binding to the sweet taste receptor and the existence of additional binding sites therefore adds further complexity to *in silico* docking experiments to the sweet taste receptor.

One of the future goals of SuperSweet is the integration of sugars and sweetening agents into biochemical pathway maps (including PubMed references) to better understand their different ways of metabolism and their impact on metabolic diseases and to foresee possible risk factors. After improving the similarity search and inclusion of pharmacorphore searching to find new putative sweetening agents, a sweetness prediction tool is planned to be implemented. We plan to perform additional text-mining in order to obtain information on the number of calories, sweetness and therapeutic effects of sweet compounds where missing. Docking poses of the sweet proteins to the sweet taste receptor are also planned to be integrated. Another interesting aspect would be the comparison of sweet taste perception with characteristics of sour and bitter taste perception, which is a problem in the development of artificial sweeteners.

AVAILABILITY

The SuperSweet database is freely available under the url: http://bioinformatics.charite.de/sweet/ and will be updated regularly.

ACKNOWLEDGEMENTS

The authors thank D. Kuzman and A. Chefai for manual curation of the sweetener properties and A.L. Wölke for help with the 'Sweet tree'.

FUNDING

Deutsche Forschungsgemeinschaft (SFB 449); Investitionsbank Berlin (IBB); Deutsche Krebshilfe; Bundesministerium für Bildung und Forschung (BMBF); European Union (EU). Funding for open access charge: DFG, BMBF and EU.

Conflict of interest statement. None declared.

REFERENCES

1. Merki,C. (1993) *Zucker gegen saccharin - zur geschichte der künstlichen süssstoffe*. Campus, Frankfurt/Main/New York.
2. de la Pena,C. (2010) Artificial sweetener as a historical window to culturally situated health. *Ann. NY Acad. Sci.*, **1190**, 159–165.
3. Hicks,R.M., Wakefield,J.S. and Chowaniec,J. (1973) Letter: Co-carcinogenic action of saccharin in the chemical induction of bladder cancer. *Nature*, **243**, 347–349.
4. Weihrauch,M.R. and Diehl,V. (2004) Artificial sweeteners—do they bear a carcinogenic risk? *Ann. Oncol.*, **15**, 1460–1465.
5. (2006) Artificial sweeteners: no calories . . . sweet! *FDA Consum.*, **40**, 27–28.
6. Edgar,W.M. (1998) Sugar substitutes, chewing gum and dental caries—a review. *Br. Dent. J.*, **184**, 29–32.
7. Hayes,C. (2001) The effect of non-cariogenic sweeteners on the prevention of dental caries: a review of the evidence. *J. Dent. Educ.*, **65**, 1106–1109.
8. Benton,D. (2005) Can artificial sweeteners help control body weight and prevent obesity? *Nutr. Res. Rev.*, **18**, 63–76.
9. Drewnowski,A. (1999) Intense sweeteners and energy density of foods: implications for weight control. *Eur. J. Clin. Nutr.*, **53**, 757–763.
10. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
11. Temussi,P. (2007) The sweet taste receptor: a single receptor with multiple sites and modes of interaction. *Adv. Food Nutr. Res.*, **53**, 199–239.
12. Morini,G., Bassoli,A. and Temussi,P.A. (2005) From small sweeteners to sweet proteins: anatomy of the binding sites of the human T1R2_T1R3 receptor. *J. Med. Chem.*, **48**, 5520–5529.
13. Doubet,S. and Albersheim,P. (1992) CarbBank. *Glycobiology*, **2**, 505.
14. Loss,A., Bunsmann,P., Bohne,A., Loss,A., Schwarzer,E., Lang,E. and von der Lieth,C.W. (2002) SWEET-DB: an attempt to create annotated data collections for carbohydrates. *Nucleic Acids Res.*, **30**, 405–408.
15. Lutteke,T., Bohne-Lang,A., Loss,A., Goetz,T., Frank,M. and von der Lieth,C.W. (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology*, **16**, 71R–81R.
16. Ranzinger,R., Frank,M., von der Lieth,C.W. and Herget,S. (2009) Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. *Glycobiology*, **19**, 1563–1567.
17. Campbell,M.P., Royle,L., Radcliffe,C.M., Dwek,R.A. and Rudd,P.M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*, **24**, 1214–1216.
18. Artemenko,N.V., Campbell,M.P. and Rudd,P.M. (2010) GlycoExtractor: a web-based interface for high throughput processing of HPLC-glycan data. *J. Proteome Res.*, **9**, 2037–2041.
19. Cooper,C.A., Joshi,H.J., Harrison,M.J., Wilkins,M.R. and Packer,N.H. (2003) GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources 2003 update. *Nucleic Acids Res.*, **31**, 511–513.

20. Nakahara,T., Hashimoto,R., Nakagawa,H., Monde,K., Miura,N. and Nishimura,S. (2008) Glycoconjugate Data Bank: Structures—an annotated glycan structure database and N-glycan primary structure verification service. *Nucleic Acids Res.*, **36**, D368–D371.
21. Hashimoto,K., Kawano,S., Goto,S., Aoki-Kinoshita,K.F., Kawashima,M. and Kanehisa,M. (2005) A global representation of the carbohydrate structures: a tool for the analysis of glycan. *Genome Inform.*, **16**, 214–222.
22. Wang,Y., Xiao,J., Suzek,T.O., Zhang,J., Wang,J. and Bryant,S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–W633.
23. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
24. Fredriksson,R. and Schioth,H.B. (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.*, **67**, 1414–1425.
25. Kunishima,N., Shimada,Y., Tsuji,Y., Sato,T., Yamamoto,M., Kumasaka,T., Nakanishi,S., Jingami,H. and Morikawa,K. (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature*, **407**, 971–977.
26. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
27. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
28. Verdonk,M.L., Cole,J.C., Hartshorn,M.J., Murray,C.W. and Taylor,R.D. (2003) Improved protein-ligand docking using GOLD. *Proteins*, **52**, 609–623.
29. Xu,H., Staszewski,L., Tang,H., Adler,E., Zoller,M. and Li,X. (2004) Different functional roles of T1R subunits in the heteromeric taste receptors. *Proc. Natl Acad. Sci. USA*, **101**, 14258–14263.
30. Gunther,S., Senger,C., Michalsky,E., Goede,A. and Preissner,R. (2006) Representation of target-bound drugs by computed conformers: implications for conformational libraries. *BMC Bioinformatics*, **7**, 293.
31. Geuns,J.M. (2003) Stevioside. *Phytochemistry*, **64**, 913–921.
32. Takasaki,M., Konoshima,T., Kozuka,M., Tokuda,H., Takayasu,J., Nishino,H., Miyakoshi,M., Mizutani,K. and Lee,K.H. (2009) Cancer preventive agents. Part 8: chemopreventive effects of stevioside and related compounds. *Bioorg. Med. Chem.*, **17**, 600–605.
33. Chatsudthipong,V. and Muanprasat,C. (2009) Stevioside and related compounds: therapeutic benefits beyond sweetness. *Pharmacol. Ther.*, **121**, 41–54.
34. Rost,B. (1999) Twilight zone of protein sequence alignments. *Protein Eng.*, **12**, 85–94.
35. Bordogna,A., Pandini,A. and Bonati,L. Predicting the accuracy of protein-ligand docking on homology models. *J. Comput. Chem.*, [Epub ahead of print, 6 July 2010]; PMID: 20607693.
36. Kellenberger,E., Rodrigo,J., Muller,P. and Rognan,D. (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*, **57**, 225–242.
37. Kontoyianni,M., McClellan,L.M. and Sokol,G.S. (2004) Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.*, **47**, 558–565.
38. Plewczynski,D., Lazniewski,M., Augustyniak,R. and Ginalski,K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.*, [Epub ahead of print, 1 September 2010]; PMID: 20812323.
39. Temussi,P.A. (2002) Why are sweet proteins sweet? Interaction of brazzein, monellin and thaumatin with the T1R2-T1R3 receptor. *FEBS Lett.*, **526**, 1–4.
40. Jiang,P., Cui,M., Zhao,B., Snyder,L.A., Benard,L.M., Osman,R., Max,M. and Margolskee,R.F. (2005) Identification of the cyclamate interaction site within the transmembrane domain of the human sweet taste receptor subunit T1R3. *J. Biol. Chem.*, **280**, 34296–34305.
41. Jiang,P., Cui,M., Zhao,B., Liu,Z., Snyder,L.A., Benard,L.M., Osman,R., Margolskee,R.F. and Max,M. (2005) Lactisole interacts with the transmembrane domains of human T1R3 to inhibit sweet taste. *J. Biol. Chem.*, **280**, 15238–15246.

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Saskia Preissner, Katharina Kroll, Mathias Dunkel, Christian Senger, Gady Goldsobel, Daniel Kuzman, Stefan Guenther, Rainer Winnenburg, Michael Schroeder, Robert Preissner

SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions.

Zeitschrift: Nucleic Acids Research

Impact Factor: 7,479

Jessica Ahmed, Saskia Preissner, Mathias Dunkel, Catherine L. Worth, Andreas Eckert and Robert Preissner

SuperSweet – a resource on natural and artificial sweetening agents

Zeitschrift: Nucleic Acids Research

Impact Factor: 7,479

Jessica Ahmed, Thomas Meinel, Mathias Dunkel, Manuela Murgueitio, Robert Adams, Corinna Blasse, Andreas Eckert, Saskia Preissner, Robert Preissner: *CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge.*

Zeitschrift: Nucleic Acids Research

Impact Factor: 7,479

Selbständigkeitserklärung

„Ich, Saskia Preißner, erkläre, dass ich die vorgelegte Dissertation mit dem Thema: „Xenobiotika in der Biomedizin – Wirkung und Stoffwechsel“ selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.“

Datum Unterschrift

Danksagung

Zunächst möchte ich mich bei Professor Dr. Peter Daniel bedanken, der es mir erst möglich machte, an verschiedenen interessanten wissenschaftlichen Projekten teilzunehmen und dadurch tief in die Welt der Forschung einzutauchen.

Jessica Ahmed möchte ich für ihre Zusammenarbeit beim SuperSweet Projekt, ihre konstruktive Anleitung beim CancerResource Projekt und für die unzähligen lustigen Arbeitsgruppensitzungen danken.

Mathias Dunkel danke ich für seine unermüdliche Geduld beim Programmieren, Umschreiben und Gestalten der Datenbanken und für den hervorragenden morgendlichen Kaffee.

Ein großer Dank gilt auch Andreas Eckert und Christian Senger, da sie oft aus der Ferne möglich gemacht haben, was vorher gar nicht möglich war.

Catherine Worth danke ich für die Zusammenarbeit im Rahmen des Sweet-Projektes und dafür, dass sie bei unseren auf Englisch abgehaltenen Sitzungen nie über uns gelacht hat, sondern uns viele neue Worte beigebracht hat.

Stefan Guenther danke ich für seine Skepsis und seine kritischen Nachfragen im Rahmen des SuperCYP-Projektes, das dadurch immer besser geworden ist.

Ich danke auch Ilona Gawronska, die sich zusammen mit mir durch den Wald der Zielmolekül-Strukturen gekämpft hat.

Daniel Kuzman und Dorothée Pfitzner danke ich für elf bittersüße Semester Studium und die vielen gemeinsamen Lern- und Forschungsabende, in denen wir vor allem viel fürs Leben gelernt haben.

Herrn Dr. Kostka danke ich dafür, dass er mir gezeigt hat, was Zahnmedizin neben Bohren und Füllen noch alles sein kann, wodurch mir der klinische Abschnitt des Studiums große Freude bereitet hat.

Zu guter Letzt möchte ich meiner Familie danken. Meine kleinen Geschwister Sarah, Sally und Robert haben mich zwar des Öfteren vom Lernen abgehalten, aber mir dafür viel Freude und Inspiration zurückgegeben. Meine Eltern haben mich immer unterstützt, mir den Rücken freigehalten, an mich geglaubt und mich zu der Person gemacht, die ich bin. Die Dankbarkeit dafür lässt sich nicht in Worte fassen.