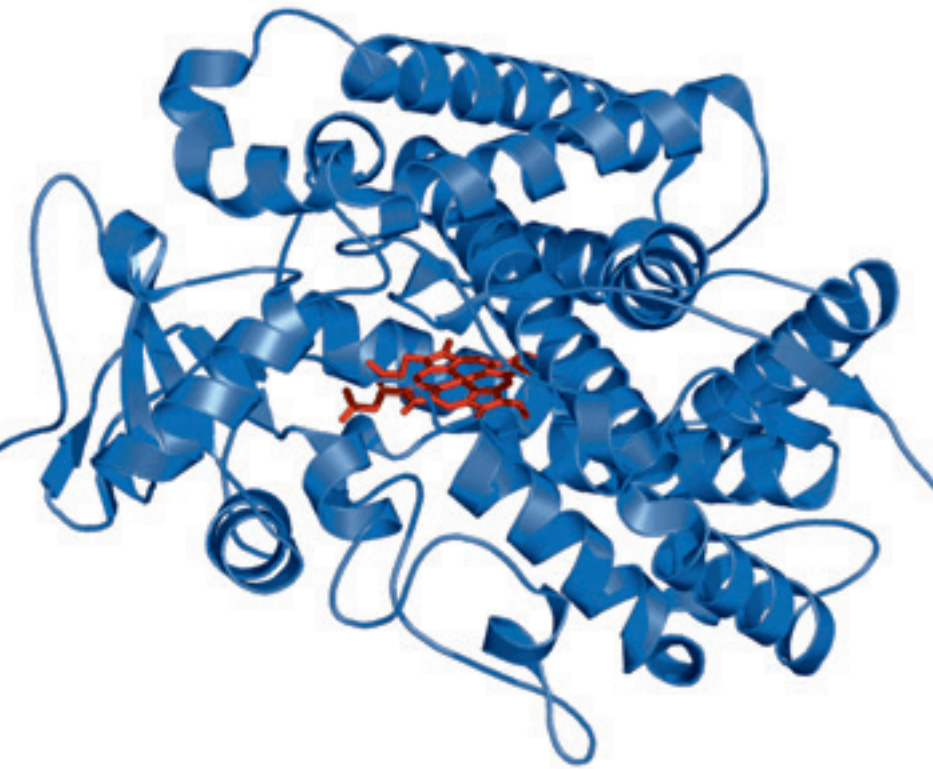


Von der Sequenz zur Funktion

Datenbanken und Modellierung



Enzyme spielen in der Medizin eine wichtige Rolle, da je nach der vorhandenen Enzymausstattung die individuelle Reaktion auf Medikamente äußerst unterschiedlich ausfallen kann und von Mensch zu Mensch variiert. Wird zuviel eines Enzyms produziert, das am Abbau eines bestimmten Medikaments beteiligt ist, kann dieses Medikament unwirksam werden. Ist hingegen das Enzym in zu geringen Mengen vorhanden, wird das Medikament zu langsam abgebaut.

Ein zu langsamer Abbau hat dabei den gleichen Effekt wie eine Überdosierung und kann zu Nebenwirkungen führen. Ein Fehlen

des Enzyms kann ebenfalls die Bildung giftiger Nebenprodukte bewirken, da der eigentliche Abbauweg nicht funktioniert. Aus der großen Enzymklasse der P450 Monooxygenasen sind die drei wichtigsten am Abbau von 30 Prozent aller in der Leber verarbeiteten Arzneimitteln beteiligt. Ein großer Schritt für die Entwicklung maßgeschneiderter Medikamente ist daher das Verständnis, wie sich die unterschiedlichen Variationen in der DNA-Sequenz eines Gens, die entweder zu einem veränderten Enzym oder zu einer Änderung der Menge an produziertem Enzym führen können, auf die Eigenschaften von P450 Monooxygenasen auswirken.

1. Einleitung

Die Enzymklasse der P450 Monooxygenasen spielt eine wichtige Rolle bei der Entgiftung und der Biosynthese. Monooxygenasen kommen in allen Lebewesen vor; so wurden drei P450 Monooxygenasen in der Hefe *Saccharomyces cerevisiae*, 18 in dem Bodenbakterium *Streptomyces coelicolor*, 80 in dem Fadenwurm *Caenorhabditis elegans* und 257 in der Pflanze *Arabidopsis thaliana* gefunden. Erstaunlicherweise enthalten die Bakterien *Escherichia coli* und *Salmonella typhimurium* keine P450 Monooxygenasen. Das menschliche Genom kodiert für 60 verschiedene P450 Monooxygenasen. 15 dieser Enzyme sind für den Abbau von Fremdstoffen wie etwa Medikamenten, die normalerweise nicht im Körper vorkommen, zuständig; 14 sind am Steroidstoffwechsel beteiligt, vier P450 Monooxygenasen oxidieren fettlösliche Vitamine und neun Enzyme sind am Stoffwechsel von Fettsäuren beteiligt. Die Funktion der restlichen P450 Monooxygenasen ist zurzeit noch nicht bekannt. Die Hauptaufgabe dieser Enzyme ist es, wasserunlösliche Substanzen in wasserlösliche Produkte umzuwandeln. Dies trifft auch auf Medikamente zu, die dadurch unwirksam und aus dem Körper ausgeschieden werden können. Die wichtigsten an Abbau, Aktivierung oder Umbau von Medikamenten beteiligten P450 Monooxygenasen kommen in der Leber vor, aber auch in anderen Geweben befindet sich eine Vielzahl dieser Enzyme. Allerdings hat nicht jeder Mensch die gleiche Ausstattung an P450 Monooxygenasen, sondern es existieren erblich bedingte Variationen. Diese werden verursacht durch so genannte *Single Nucleotide Polymorphisms* (SNPs). SNPs sind Variationen in der DNA-Sequenz eines Gens, die entweder zu einem veränderten Enzym oder zu einer Änderung der Menge an produziertem Enzym führen können.

2. P450 Datenbank

Die Funktion der meisten humanen P450 Monooxygenasen ist noch nicht bekannt. Die Vorhersage der Eigenschaften dieser Enzyme und ihrer SNPs bezüglich Menge und biochemischer Eigenschaften, und somit die Vorhersage ihrer Aktivität gegenüber neuen Wirkstoffen ist das Ziel der P450 Datenbank. Durch systematisches Sammeln und Vergleichen der Daten von

P450 Monooxygenasen mit bekannten und unbekanntem Eigenschaften können die Funktionen der einzelnen Bereiche des Proteins entschlüsselt werden. Obwohl die Proteinsequenzen von P450 Monooxygenasen nur geringe Ähnlichkeit aufweisen, haben die Proteine eine ähnliche Struktur und Funktion, wodurch sie direkt verglichen werden können. Der systematische Vergleich aller bekannten Sequenzen, Strukturen und Eigenschaften soll daher zum Verständnis dieser Enzyme führen.

Die P450 Datenbank ist eine familienspezifische Datenbank, sie bündelt und organisiert große Mengen an Sequenz-, Struktur- und Annotationsinformationen und ist daher ein nützliches Werkzeug, um Beziehungen zwischen Proteinsequenzen und Proteinstrukturen zu untersuchen. Ein Protein ist ein Polymer bestehend aus einer Kette von Aminosäuren, die eine definierte räumliche Struktur einnehmen. Die räumliche Struktur eines Proteins bedingt seine Funktion und Wirkungsweise. Verwandte, also aus einem gemeinsamen Vorgängerprotein hervorgegangene Proteine weisen neben einer ähnlichen Struktur auch eine ähnliche Proteinsequenz auf. Je nach dem Verwandtschaftsgrad wurden die Proteinsequenzen daher in homologe Familien und übergeordnet in Superfamilien eingeteilt. Die Benennung von P450 erfolgt nach ihrem Verwandtschaftsgrad und deckt sich mit der Einteilung der Enzyme in der P450 Datenbank: Der Name eines P450 beginnt mit CYP (*C*ytocrom *P*450) gefolgt von einer Zahl, die die Superfamilie bestimmt, eines Buchstabens, der die homologe Familie bezeichnet und einer weiteren Zahl, die das Enzym identifiziert (z.B. CYP2C9, CYP3A4). Die P450 Datenbank enthält 3538 Proteineinträge, wobei von 128 Einträgen Proteinstrukturen bekannt sind. Sie lassen sich in 934 homologe Familien und 462 Superfamilien einteilen. Hiervon sind 218 Proteine aufgeteilt in 43 homologe Familien und 20 Superfamilien humane P450 Monooxygenasen.

ZUSAMMENFASSUNG

Die Enzymklasse der P450 Monooxygenasen spielt eine wichtige Rolle beim Ab- und Umbau von Fremdstoffen wie Giftstoffen oder Medikamenten. Obwohl alle Vertreter dieser Klasse den gleichen Reaktionstypus katalysieren, variieren die einzelnen Enzyme erheblich in ihrem Substrat- und Produktspektrum. Alleine im menschlichen Genom gibt es 60 verschiedene P450 Gene. An unserem Institut wird durch Methoden der Bioinformatik und durch computergestützte Proteinmodellierung untersucht, welche Bereiche in den P450 Monooxygenasen für die Erkennung der Fremdstoffe verantwortlich sind und welche Produkte dabei jeweils entstehen. Mit einem solchen Modell lässt sich erklären, wie eine unterschiedliche Zusammensetzung von P450 Monooxygenasen bei zwei Patienten zu unterschiedlichen Abbauraten desselben Medikaments führen kann. Ziel ist es, die Abbauraten und die Abbauprodukte für jeden beliebigen Fremdstoff individuell vorhersagen zu können.



01

Phylogenetischer Baum der homologen Familien 2A, 2B, 2C, 2D, 2E und 2F. Humane P450 wurden in rot hervorgehoben, homologe Familien wurde durch Einrahmungen gekennzeichnet.

Eine interessante Familie ist die Superfamilie 2. Mitglieder dieser Superfamilie sind beteiligt am Abbau von Nikotin (CYP2A6, CYP2B6), Morphinen (CYP2B1), AT1-Antagonisten (Cyp2C9), Protonenpumpenhemmern, Antiepileptika und Antidepressiva (CYP2C19), β -Blockern (CYP2D6) sowie Anästhetika, Ethanol und Paracetamol (CYP2E1). Um Proteinsequenzen dieser Superfamilie zu untersuchen, werden verschiedene Analysemethoden angewandt.

Zur Analyse der gesammelten Sequenzdaten wurden für jede homologe Familie und jede Superfamilie Sequenzvergleiche erstellt. Bei einem Sequenzvergleich werden Proteinsequenzen nach ihrer Ähnlichkeit angeordnet, wobei die Eigenschaften ihrer Aminosäuren (hydrophob, polar, geladen) eine wichtige Rolle spielen. Von besonderer Bedeutung ist hierbei die Beobachtung, dass die Eigenschaften von Aminosäuren an bestimmten Positionen im Protein für die Faltung und Aktivität essentiell sind. Anhand eines Sequenzvergleichs können daher Aminosäuren, die in allen Proteinsequenzen einer Familie gleich sind, identifiziert werden. Solche konservierten Aminosäuren sind häufig entweder für die Funktion oder die Struktur eines Proteins essentiell. Sequenzvergleiche werden als Grundlage für eine Vielzahl von Analysemethoden eingesetzt, wie für die Berechnung von phylogenetischen Bäumen, welche die Verwandtschaftsbeziehungen zwischen einzelnen Proteinen aufgrund ihrer

Sequenzähnlichkeit darstellen. In (01) ist ein phylogenetischer Baum ausgesuchter Proteine aus sechs homologen Familien der Superfamilie 2 dargestellt. Der Baum gliedert sich für die homologen Familien 2A, 2B, 2D und 2F deutlich in vier Äste, wobei auch Unterteilungen innerhalb einer homologen Familie sichtbar sind (2A, 2B und 2D). Die homologen Familien 2C und 2E hingegen

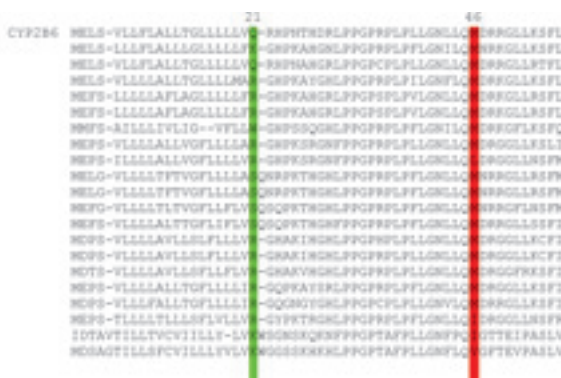
zeigen keine deutliche Trennung, da die Familie 2C streut. In diesem Fall wäre zu überlegen, ob die Einordnung der Proteine

2C22, 2C30, 2C117 und 2C33 aus diesem Beispiel in die homologe Familie 2C korrekt ist.

Zur Analyse der Auswirkungen von Polymorphismen werden alle verfügbaren Informationen über bekannte SNPs in der Datenbank gesammelt und ausgewertet: welche Position wurde mutiert, welche Aminosäure wurde ausgetauscht, und welche Auswirkungen auf die Aktivität und die Substratspezifität des Enzyms sind bekannt. Allerdings führt nicht jeder Austausch notwendigerweise zu einer Änderung der Eigenschaften. Gewisse Variationen in der Proteinsequenz werden toleriert, andere hingegen führen zur Inaktivierung des Enzyms. Um diese Variationen innerhalb einer Familie zu untersuchen, wurden Konservierungsanalysen durchgeführt. Hierbei werden die Eigenschaften der Aminosäuren, ihre Position im Protein und ihre Konservierung berücksichtigt. Im Alignment der Familie CYP2B erkennt man, dass die Aminosäure Glutamin an Position 21 des humanen CYP2B6 Enzyms nur in einer weiteren Sequenz der Familie vorkommt, und damit nicht konserviert ist (02). Die Aminosäure Methionin an Position 46 hingegen ist in fast allen Sequenzen der Familie 2B erhalten und damit hoch konserviert. Experimentelle Daten zu Mutationen an diesen Positionen in CYP2B6 zeigen, dass eine Mutation des hoch konservierten Methionins zu einem Verlust der Enzymfunktion führt, während das Enzym tolerant gegenüber einer Mutation an der nicht konservierten Position 21 ist. Im Vergleich mit P450/SNPs mit bekannten Eigenschaften lassen sich Rückschlüsse auf die Eigenschaften neu entdeckter SNPs ziehen. Das Ziel dieser Untersuchungen ist, die Eigenschaften unbekannter SNPs vorherzusagen und somit die Grundlage zu schaffen, dass individuell für jeden Patienten das richtige Medikament in der richtigen Dosierung verabreicht werden kann.

3. Modellierung

Um die Funktionsweise von Enzymen zu verstehen sind Informationen über ihre dreidimensionale Proteinstruktur nötig. Zur Zeit sind etwa 28.500 Proteinstrukturen in der öffentlich zugänglichen Datenbank, der *Protein Data Base* (PDB) (www.pdb.org) zugänglich. Seit Anfang der neunziger Jahre des letzten Jahrhun-



02

Ausschnitt aus dem Sequenzalignment der Familie 2B. Die Positionen 21 und 46 sind farbig markiert.

derts, als die erste Kristallstruktur eines bakteriellen P450 Enzyms publiziert wurde, konnten die Strukturen von 13 verschiedenen P450 Monooxygenasen aufgelöst werden. Zunächst wurden vor allem die Strukturen bakterieller Enzyme, in den letzten Jahren jedoch auch die von höheren Organismen aufgeklärt (03). Trotz unterschiedlicher Sequenzen zeigen die bisher bekannten P450 Monooxygenasen eine große Ähnlichkeit. Da die Kristallisation von Proteinen, die für die Ermittlung von Proteinstrukturen benötigt wird, sehr aufwendig und schwierig ist, gibt es, im Verhältnis zu den bekannten Sequenzen, nur eine relativ geringe Anzahl aufgelöster Proteinstrukturen. Allerdings kann die Struktur eines Proteins auch auf dem Computer modelliert werden. Dabei wird vorausgesetzt, dass Proteine mit ähnlicher Sequenz eine ähnliche Struktur besitzen. Die Sequenzidentität zwischen einem Protein mit bekannter Struktur und dem zu modellierenden Protein sollte 50 Prozent oder mehr betragen, um ein vertrauenswürdigeres Modell zu erhalten.

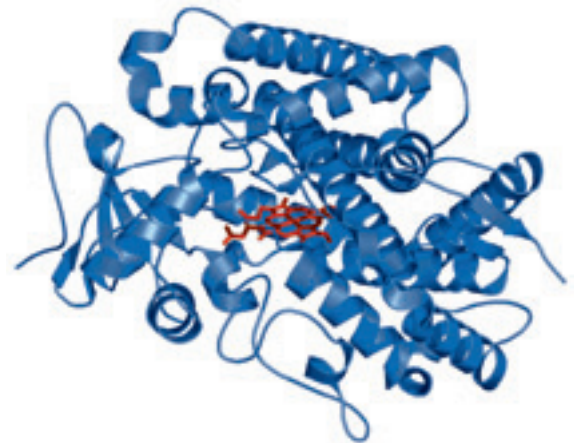
Mit Hilfe der Methoden der Bioinformatik und der computergestützten molekularen Modellierung soll der Zusammenhang zwischen Sequenz, Struktur und biochemischen Eigenschaften der P450 Monooxygenasen untersucht werden. Auf dem Computer wird modelliert, wie ein Enzym sein Substrat in der Bindungsstelle bindet, und welche Aminosäuren für die biochemischen Eigenschaften verantwortlich sind. Wechselwirkungen zwischen Substrat und Enzym können mit der Methode des Dockings untersucht werden. Mit Hilfe dieser Methode können große Substratbibliotheken virtuell in Enzymstrukturen gedockt werden und deren Wechselwirkungen untersucht werden (*virtual screening*). Da für die Erkennung von Substraten durch Enzyme sowohl die Struktur als auch die Flexibilität eine wichtige Rolle spielen, müssen bei der Modellierung die Bewegungen des gesamten Systems simuliert werden. Ein typisches Enzym-Substrat-System besteht zusammen mit dem Lösungsmittel aus 50.000 bis 100.000 Atomen. Um solch große Systeme simulieren zu können, wurde 2002 am Institut für Technische Biochemie ein Höchstleistungscomputer aus 256 PC-CPU's mit einer Spitzenleistung von 785 GFlops aufgebaut (04).

Durch Modellierung von P450 Enzym-Substrat-Komplexen auf dem Computer wird die Selektivität verschiedener P450 gegenüber Substraten und der Einfluss einzelner Aminosäuren untersucht. Aber nicht nur die Substratspezifität, sondern auch die Selektivität der entstehenden Produkte kann modelliert und verstanden werden. Dies ist ein wichtiger Gesichtspunkt bei der Entwicklung von Medikamenten, da unterschiedliche Reaktionsprodukte verschiedene, teils toxische Eigenschaften haben können. So konnte die unterschiedliche Regioselektivität von CYP2B6 gegenüber Cyclophosphamid und Ifosfamid, Wirkstoffen aus der Krebstherapie, durch Modellierung verstanden werden. Durch den Austausch von Aminosäuren werden die Auswirkungen von SNPs auf die Selektivität einzelner P450 untersucht. Die computergestützte Modellierung ist damit eine Brücke zwischen Proteinstrukturen und ihren Funktionen.

• Sandra Barth
Stephan Tatzel
Jürgen Pleiss

SUMMARY

The enzyme class of P450 monooxygenases plays a major role in the metabolism of xenobiotics like toxic agents or drugs. Although all members of this class catalyse the same reaction type, their substrate and product spectrum varies considerably. In the human genome 60 different P450 genes have been identified. At our institute we investigate by bioinformatics and computer-aided protein modelling methods which binding sites in the protein mediate the recognition of xenobiotics and the formation of product. Such a model explains how for the same drug the different compositions of P450 monooxygenases in two patients lead to a different degradation rate. This model will then be applied to predict the individual degradation rates and degradation products for any xenobiotic.



03

*Struktur des P450 CYP2B4 aus *Oryctolagus cuniculus* (Hase). Im aktiven Zentrum ist das Häm, die prosthetische Gruppe des Enzyms, rot dargestellt.*



04

BioCORTEX Ein PC-Cluster bestehend aus 256 AMD MP1800+ Prozessoren.

DIE AUTOREN

**SANDRA BARTH**

Geboren 1972, studierte nach einer Ausbildung zur chemisch-technischen Assistentin an der Universität Karlsruhe Chemieingenieurwesen, anschließend an der Universität Stuttgart Technische Biologie, wo sie am Institut für Technische Biochemie ihre Doktorarbeit über die systematische Analyse von Sequenz- und Strukturdaten von Proteinfamilien anfertigte. Seit 2005 ist sie in der Forschungs- und Entwicklungsabteilung von Philips Medical Systems tätig.

**STEPHAN TATZEL**

Geboren 1975, studierte Chemie an der Universität Stuttgart und schloss sein Studium mit einer Diplomarbeit zur Untersuchung von Enzym-Substrat-Wechselwirkungen von Cytochrom P450 Monoxygenasen mittels Dockingsimulationen ab. Seit 2003 promoviert er am Institut für Technische Biochemie auf dem Gebiet der molekularen Modellierung von humanen P450 Monoxygenasen.

**JÜRGEN PLEISS**

Geboren 1959, studierte Physik an den Universitäten Stuttgart und Tübingen und promovierte am Max-Planck Institut für Biologie in Tübingen. Nach einer dreijährigen Tätigkeit als Product Manager bei Biostructure (Strasbourg) baute er seit 1994 am Institut das Gebiet der Bioinformatik in Forschung und Lehre auf und habilitierte sich im Jahr 2001. Seine Arbeitsgruppe beschäftigt sich mit der molekularen Modellierung von Proteineigenschaften und Proteindesign.

Kontakt

Universität Stuttgart
Institut für Technische Biochemie
Allmandring 31
70569 Stuttgart
Tel. 0711/685-3191
Fax 0711/685-4569
E-Mail: itbjpl@itb.uni-stuttgart.de