

**Institut für Experimentelle Genetik
GSF-Forschungszentrum für Umwelt und Gesundheit,
Neuherberg**



**Phylogenetische und bioinformatische Untersuchung der
17 β -Hydroxysteroiddehydrogenasen:
Struktur, Funktion und Evolution einer komplexen
Proteinfamilie**

Dipl.-Biochem. Rainer Breitling

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weiherstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Alfons Gierl

Prüfer der Dissertation: 1. Priv.-Doz. Dr. Jerzy Adamski
2. Univ.-Prof. Dr. Rudolf Balling
3. Univ.-Prof. Dr. Siegfried Scherer

Die Dissertation wurde am 18.12.2000 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weiherstephan für Ernährung, Landnutzung und Umwelt am 24.04.2001 angenommen.

Inhaltsverzeichnis

Abstract	5
Anmerkungen zur Nomenklatur	6
1 Einleitung	7
1.1 Möglichkeiten der Bioinformatik zur Funktionsbestimmung von Proteinen	7
1.2 Zielsetzung	12
2 Verwendete Computerprogramme und Datenbanken	15
2.1 Computerprogramme	15
2.2 Datenbanken	19
3 Pilotstudien	21
3.1 Pilotstudie I: Paired-Box-Proteine	21
3.1.1 Problemstellung	21
3.1.2 Hintergrund	22
3.1.3 Ergebnisse und Diskussion	25
3.2 Pilotstudie II: Cholin-Acetyl-Transferasen	34
3.2.1 Problemstellung	34
3.2.2 Hintergrund	34
3.2.3 Ergebnisse und Diskussion	35
4 Die 17β-Hydroxysteroid-Dehydrogenasen	39
4.1 Ein strukturbasierter Stammbaum der HSD-Familie	43
4.1.1 Problemstellung	43
4.1.2 Ergebnisse	43
4.1.3 Diskussion	56
4.2 Die 17 β -HSD Typ 7	59
4.2.1 Problemstellung	59
4.2.2 Hintergrund	59
4.2.3 Ergebnisse	60
4.2.4 Diskussion	67
4.3 Die 17 β -HSD Typ 4	81
4.3.1 Problemstellung	81
4.3.2 Hintergrund	81
4.3.3 Ergebnisse	83

Inhaltsverzeichnis

4.3.4	Diskussion	87
4.4	Die 17 β -HSD Typ 1	90
4.4.1	Problemstellung	90
4.4.2	Hintergrund	90
4.4.3	Ergebnisse und Diskussion	91
4.5	Die 17 β -HSD Typ 5	94
4.5.1	Problemstellung	94
4.5.2	Hintergrund	95
4.5.3	Ergebnisse und Diskussion	99
5	Zusammenfassung	109
	Epilog	111
	Glossar	113
	Referenzen	117
	Literaturverzeichnis	117
	Eigene Veröffentlichungen	127
	Abbildungsverzeichnis	129
	Tabellenverzeichnis	131
	Anhang: Quellcode des STATRAS2-Programms	133
	Danksagung	141
	Lebenslauf	143

Abstract

This work describes the integrative application of various bioinformatic approaches to the family of 17β -hydroxysteroid dehydrogenases. This protein family is involved in the activation and inactivation of steroid hormones by reduction and oxidation at position 17 of the steroid backbone. Several family members are known to be involved in human disorders (e. g. pseudohermaphroditism or Zellweger-like syndrome) and are implicated in cancerogenesis and tumor proliferation.

A reliable evolutionary scenario for the large family of HSD-related proteins was constructed by combining structure-based and sequence-based comparisons into a single phylogenetic tree. The tree indicates a fundamental subdivision of the 17β -HSDs into two major groups. By determining the root of the phylogenetic tree it was possible to evaluate the time course of evolutionary events leading to the diversity 17β -HSDs.

A comprehensive bioinformatic study was performed to assess the physiological function of human 17β -HSD type 7. Integrating expression data, phylogenetics and structural analyses it was possible to demonstrate that this estrogenic enzyme probably has an ancestral function in cholesterol metabolism, namely the reduction of 3-ketosteroids. Adjacent steps in the pathway are catalysed by proteins that are mutated in developmental defects (CHILD syndrome and CDPX2 syndrome), making 17β -HSD7 a likely candidate for similar disorders.

The combination of protein structure and statistical evaluation enabled the identification of unexpected conserved structural elements in 17β -HSD type 5 and its relatives. A 3D-model of the enzyme allowed the explanation of the inhibitory capacity of phytoestrogens acting on 17β -HSD5. The computer algorithm developed for that purpose is universally applicable and might be used for the structural analysis of other large protein families.

The techniques necessary for the evolutionary analysis of large sequence datasets were established and validated in a pilot study on paired-box proteins. It was possible to find the ancestors of the paired-box family among Tc1-transposase proteins. It is shown that paired-box proteins originated at the beginning of metazoan evolution by a single fusion event between a transposase DNA binding domain and a homeobox protein. The fusion event was followed by a rapid diversification of the protein family.

Anmerkungen zur Nomenklatur

Die vorliegende Arbeit behandelt die verschiedensten biologischen Makromoleküle: Proteine, cDNAs und Genomfragmente; Enzyme und Transkriptionsfaktoren; aus Menschen, Mäusen, Fliegen, Würmern und anderen Organismen. Für jede dieser Molekülklassen gibt es spezielle Nomenklatorsysteme. Um den vergleichenden Überblick zu erleichtern, wurden möglichst einheitliche, vereinfachte Bezeichnungen verwendet. Wo keine Missverständnisse zu befürchten sind, wurde z. B. nicht zwischen Protein, cDNA und dazugehörigem Gen unterschieden. Auch orthologe Sequenzen aus verschiedenen Organismen wurden, wo es möglich war, mit denselben Namen bezeichnet. Enzyme sind in der Regel nach ihrer zuerst beschriebenen katalytischen Aktivität benannt. Die dazugehörigen Abkürzungen sind die gleichen, wie in den vorab veröffentlichten Artikeln.

Alle Proteine, von denen man annimmt, dass sie einen historischen Vorläufer mit den meisten 17β -HSDs teilen, werden in der vorliegenden Arbeit als HSD-Familie bzw. HSD-Proteine bezeichnet. Damit enthält die HSD-Familie im wesentlichen die gleichen Proteine wie die erweiterte *short chain*-Dehydrogenase/Reduktase-Familie (SDR- oder SCAD-Familie) *sensu* Jörnvall et al. (1999). Nicht alle HSD-Proteine sind jedoch gegenüber Hydroxysteroiden als Oxidoreduktasen aktiv. Die Untergruppe der HSD-Familie, die diese Aktivität besitzt, wird in dieser Arbeit als „HSD-Proteine im engeren Sinne“ bezeichnet. Die Einteilung der 17β -Hydroxysteroiddehydrogenasen (17β -HSDs) erfolgt nach Peltoketo et al. (1999b) und Adamski & Jakob (2000, im Druck).

Die Nomenklatur und Abgrenzung der Aldo-Keto-Reduktasen erfolgt nach Jez et al. (1997) und Jez & Penning (2000).

1 Einleitung

„Nullus est liber tam malus,
ut non aliqua parte prosit“
PLINIUS DER ÄLTERE¹

Exponentiell wachsende Sequenzdatenbanken, neu entwickelte Hochdurchsatz-Experimente und verfeinerte Techniken der Röntgenkristallografie liefern seit einigen Jahren eine Fülle interessanter Daten mit immer noch wachsender Geschwindigkeit. Das volle Potential dieser Informationsflut erschließt sich erst durch eine bioinformatische Analyse (Vukmirovic & Tilghman, 2000). Während sich viele der dazu notwendigen Algorithmen automatisieren lassen, muss die Integration der Informationen individuell erfolgen, denn sie erfordert je nach Fragestellung spezielle biologische Hintergrundkenntnisse (Fuchs, 2000). In der vorliegenden Arbeit wird am Beispiel der 17β -Hydroxysteroiddehydrogenasen (17β -HSDs) vorgeführt, wie sich evolutionsbiologische, strukturelle und genetische Daten in einer bioinformatischen Analyse erfolgreich verbinden lassen. Erst in dieser Kombination ergeben sich neue Erkenntnisse, die für das weitere experimentelle Vorgehen unerlässlich sind.

1.1 Möglichkeiten der Bioinformatik zur Funktionsbestimmung von Proteinen

Proteinsequenzen enthalten Informationen über die Proteinfunktion

Die Aufgaben der Bioinformatik lassen sich grob in zwei große, überlappende Bereiche einteilen: Einen „technischen“ und einen „wissenschaftlichen“. Der „technische“ Bereich umfasst die Versuche, die immer schneller akkumulierenden biologischen Datenmengen zu ordnen, zu kommentieren und in aufgearbeiteter Form zugänglich zu machen. Der „wissenschaftliche“ Bereich, um den es in dieser Arbeit geht, versucht, die vorhandenen Daten zu interpretieren, zu verbinden und daraus neue Arbeitshypothesen für die biomedizinische Forschung zu entwickeln. Besondere Bedeutung hat in diesem Zusammenhang die Vorhersage von möglichen Funktionen für bisher ungenügend charakterisierte

¹ zitiert in: Plinius der Jüngere, Epistulae, Buch 3, Brief 5.10.

1 Einleitung

Proteine (Eisenberg et al., 2000), denn die stetig wachsende Datenflut macht eine umfassende, individuelle Untersuchung aller neu entdeckten Proteine unmöglich.

Ausgangspunkt einer („wissenschaftlichen“) bioinformatischen Arbeit ist in der Regel die Sequenz eines Proteins, die häufig auf der Grundlage einer kodierenden DNA-Sequenz vorhergesagt wurde. Die Proteinsequenz kann bereits genutzt werden, um biophysikalische Parameter zu berechnen, z. B. Molekulargewicht, isoelektrischen Punkt, Sekundärstrukturen, Transmembranbereiche oder Peptidaseverdau-Fragmente. Eine differenzierte Funktionsvorhersage ist aber alleine mit diesen Methoden nicht möglich.

Der nächste Schritt ist üblicherweise der Vergleich mit bereits bekannten Sequenzen. Schon damit ist in vielen Fällen eine weitgehende Funktionsbestimmung möglich, nämlich dann, wenn ähnliche Proteine bereits genauer untersucht wurden. Sequenzvergleiche erlauben die Zuordnung zu Proteinklassen (z. B. Yona et al., 2000) und können auch konservierte Motive (Lokalisatonsignale, Modifizierungsstellen) erkennen, die ebenfalls Aufschluss über die mögliche Funktion geben können.

Der gleichzeitige Vergleich von mehreren verwandten Sequenzen (Alignment) liefert zusätzliche Erkenntnisse, z. B. über die funktionelle Relevanz einzelner Aminosäuren. Er liefert aber auch Informationen über die Evolutionsgeschichte der Sequenzen (Zuckerlandl & Pauling, 1965; Doolittle, 1981): Bereits eine einfache BLAST-Suche (Altschul et al., 1997) macht quantitative Aussagen über die Ähnlichkeit von Sequenzen, die anhand des E-Scores auch hierarchisch geordnet sind: Weiter entfernte Verwandte haben höhere E-Scores. Ein exakter Stammbaum soll jedoch detailliertere Angaben über die Verwandtschaftsbeziehungen machen und dadurch ein differenzierteres Verständnis der Proteine ermöglichen. Einen solchen phylogenetischen Stammbaum zu rekonstruieren ist eine anspruchsvolle bioinformatische Aufgabe. Eine einfache Rechnung veranschaulicht die Komplexität der Fragestellung: Für drei Proteine (oder Organismen) gibt es drei mögliche Stammbäume mit festgelegter Wurzel, für vier Proteine bereits 15, und ganz allgemein für N Proteine:

$$X = \prod_{i=1}^{N-2} (2i + 1) = \frac{(2N - 3)!}{2^{N-2}(N - 2)!}$$

Das bedeutet, dass bereits für 60 Proteine die Zahl der möglichen Stammbäume die Zahl der Atome im Universum (etwa 10^{80}) weit übersteigt. Und nur einer dieser Stammbäume ist korrekt, beschreibt also den tatsächlichen historischen Ablauf der Evolution. Die verschiedenen Algorithmen können über diesen „wahren“ Stammbaum zwar nichts aussagen, aber sie versuchen, den unter bestimmten evolutionsbiologischen Annahmen „besten“ (z. B. kürzesten oder wahrscheinlichsten) Stammbaum herauszufinden. Aber auch diese Aufgabe ist nicht analytisch zu lösen, denn die Suche nach dem „besten“ Stammbaum gehört in die gleiche Klasse von Problemen, wie das klassische „Problem des

1.1 Möglichkeiten der Bioinformatik zur Funktionsbestimmung von Proteinen

Handlungsreisenden“, die NP-kompletten Probleme (Korostensky & Gonnet, 2000). Das bedeutet, um mit Sicherheit den „besten“ Stammbaum zu finden, müssten sämtliche Möglichkeiten durchprobiert werden. Es ist leicht einzusehen, dass das mit den heutigen Computertechniken nicht zu bewältigen ist. Die bisher implementierten Algorithmen beschränken sich stattdessen darauf, eine signifikante Untermenge der Stammbäume zu analysieren und die erhaltenen Stammbäume evtl. schrittweise zu optimieren (Parsimony und Maximum-Likelihood-Methode), oder sie machen die vereinfachte Annahme, dass sich ein sinnvoller Stammbaum bereits dann ergibt, wenn jeweils die ähnlichsten Sequenzen nächste Verwandte sind (Neighbour-Joining). Letzteres Verfahren kann auch genutzt werden, um Proteine nach Ähnlichkeitskriterien zu sortieren, die nicht auf den Sequenzen beruhen, z. B. nach Übereinstimmungen der 3D-Struktur (Holm & Sander 1996; und Kap. 4.1 der vorliegenden Arbeit).

Warum phylogenetische Forschung?

Stellte man diese Frage einem Ahnenforscher oder Anthropologen, so wäre die Antwort klar: Die Kenntnis der eigenen Vorfahren gehört zu den Grundbedürfnissen des Menschen – und ihre Verwandtschaftsbeziehungen sind häufig das einzige, was wir von ihnen wissen. Die „elementaren Strukturen der Verwandtschaft“ (Lévi-Strauss, 1947) bestimmen aber nicht nur die Art, wie wir von unseren Ahnen denken, sondern durchziehen als gliederndes, ordnendes, beherrschendes Prinzip unsere gesamte Weltsicht. Verwandtschaftsgrade liefern die Koordinaten, anhand derer wir unsere Mitmenschen einordnen, sie kehren wieder in der Taxonomie unserer belebten und unbelebten Umwelt und sind selbst in den diffizilen Hierarchien juristischer Regelwerke oder musikalischer Kompositionen noch zu spüren.

Für den einzelnen Naturwissenschaftler mögen diese Motivationen bei der phylogenetischen Forschung ebenfalls eine Rolle spielen. Die ästhetische Wirkung eines feinstrukturierten Stammbaums ist sicher nicht zu unterschätzen, egal ob es sich um die verwickelte Stammesgeschichte einer obskuren Insektengattung handelt oder um die Rekonstruktion des Ursprungs der Pax-Familie. Es sind jedoch deutlich profanere Gründe, die die Stammbaum-Forschung zu einem unentbehrlichen Werkzeug der Evolutionsbiologie machen. Zunächst einmal kondensiert ein Stammbaum eine Fülle von Informationen. Erst durch ihn werden Strukturen sichtbar, die in den biologischen Primärdaten nicht zu erkennen wären. Zwei Organismen (oder Makromoleküle) werden dann als verwandt erkannt, wenn sie gemeinsame Eigenschaften haben, und ein guter Stammbaum ermöglicht es, diese Eigenschaften ohne weitere Vorkenntnisse abzulesen. Aus diesem Grunde bietet sich die phylogenetische Einordnung auch als Grundlage für eine möglichst informative Nomenklatur an (de Queiroz & Gauthier, 1990, 1992).

Ein Stammbaum ist aber gleichzeitig eine historische Beschreibung, und deshalb kann es auch immer nur einen (korrekten) geben. Was hilft nun diese historische Information, wenn sich doch die Vergangenheit auf immer dem

1 Einleitung

naturwissenschaftlichen Zugriff entzieht? Ein Stammbaum erzählt in komprimierter Form, wie biologische Einheiten auseinander hervorgegangen sind. Und weil alles Leben aus Leben entstand und jede Funktion aus einer bestehenden Funktion, erzählt der Stammbaum auch, welche Möglichkeiten bereits erprobt wurden, welche Wege die Natur schon gegangen ist. Er sagt, was möglich ist. Ein kompletter und korrekter Stammbaum legt sehr genau fest, wie die Vorläufer der heutigen biologischen Vielfalt ausgesehen haben, und jeder dieser (hypothetischen) Vorläufer kann Gegenstand wissenschaftlicher Überprüfung werden. Zum Beispiel ist es heute möglich, postulierte Vorläufermoleküle mit gentechnischen Methoden nachzubauen und so einen Einblick in die Funktionsweise heutiger Proteine zu bekommen (Chang & Donoghue, 2000). Das bedeutet aber auch, dass die Rekonstruktion von Stammbäumen ganz praktische Konsequenzen hat: Ihr primäres Ziel ist es nicht, metaphysisch „wahre“ Aussagen über die Entwicklungsgeschichte zu machen, sondern überprüfbare Vorschläge, wie die Geschichte abgelaufen sein kann. Deshalb wurde im Folgenden auf die experimentelle Prüfbarkeit der Resultate besonderer Wert gelegt.

Hochdurchsatz-Experimente liefern weitere Informationen

Eine Reihe weiterer bioinformatischer Verfahren nutzt die Daten, die in groß angelegten Experimenten (*high throughput*- oder Hochdurchsatz-Experimenten) gewonnen und öffentlich zugänglich gemacht werden, u. a. im Rahmen des humanen Genomprojekts und den ergänzenden Untersuchungen in anderen Organismen. Zusätzliche Informationen über die Funktion eines Proteins liefern z. B. organismenübergreifende Vergleiche, insbesondere Vergleiche von kompletten Genomen. Für jedes Protein lässt sich ein „phylogenetisches Profil“ angeben, das sagt, in welchen Genomen orthologe Proteine vorkommen und in welchen nicht. Je größer die Zahl komplett sequenzierter Genome ist, desto informativer ist ein solches „phylogenetisches Profil“, und ähnliche Profile verschiedener Proteine zeigen unter Umständen an, dass sie an einer gemeinsamen Aufgabe beteiligt sind (Pellegrini et al., 1999). Umfassende Vergleiche der kompletten Genome zahlreicher Organismen können die Rekonstruktion von Proteinfunktionen und ganzen Biosynthesewegen ermöglichen, wie Smit & Mushegian (2000) am Beispiel der Isoprenoid-Biosynthese zeigen konnten (siehe auch Lange et al., 2000; Boucher & Doolittle, 2000).

Bei der „Stein von Rosetta“-Methode wird eine funktionelle Interaktion von Proteinen (z. B. ihre Beteiligung am selben Biosyntheseweg) durch das Auftreten von Proteinfusionen vorhergesagt (Enright et al., 1999; und Kap. 4.3 der vorliegenden Arbeit). Natürlich können auch die Daten von genomweiten molekularbiologischen Interaktions-Screens (z. B. Hefe-*Two-Hybrid-System*) vergleichbare Erkenntnisse liefern, wie sie für die Hefe *S. cerevisiae* (Uetz et al., 2000; Ito et al., 2000) und im kleineren Maßstab für *C. elegans* (Walhout et al., 2000) bereits durchgeführt wurden.

Weitere Einblicke in die Funktion eines uncharakterisierten Proteins kann

1.1 Möglichkeiten der Bioinformatik zur Funktionsbestimmung von Proteinen

auch seine gezielte Ausschaltung in geeigneten Modellorganismen liefern. Umfassende Projekte, in denen sämtliche vorhergesagten Gene einzeln deletiert werden, sind für die Hefe *S. cerevisiae* und den Nematoden *C. elegans* bereits fast abgeschlossen (Liu et al., 1999; Smith et al., 1996).

Eine andere Möglichkeit, die Funktion eines Proteins näher zu bestimmen, ist die Untersuchung seines Expressionsmusters („Expressions-Profiling“) – z. B. auch in K.-o.-Organismen. Bei der Hefe *S. cerevisiae*, von der schon viele gut charakterisierte Mutanten existieren, ist dieses Verfahren besonders ergiebig. Die „Expressions-Profile“ sämtlicher Hefe-Gene wurden von Hughes et al. (2000) unter zahlreichen Bedingungen bestimmt und verglichen. Koregulierte Expression führte zur Vorhersage von verschiedenen neuen Enzymfunktionen, von denen acht auch experimentell bestätigt wurden (Hughes et al., 2000). Ähnliche Experimente sind auch für andere Organismen möglich, deren Genom vollständig bekannt ist (z. B. *C. elegans*: Hill et al., 2000).

Aber auch für humane Proteine ist ein Expressions-Profiling möglich. Dafür stehen verschiedene Daten zur Verfügung.

1. Die SAGE-Technik (Velculescu et al., 1995) bestimmt nahezu quantitativ die Expressionsniveaus aller Gene in einer Probe (mit gewissen technisch bedingten Einschränkungen, siehe Kap. 4.2). Dabei ist es nicht erforderlich, die betreffenden Gene vorher zu kennen. Der Vergleich der Expressionsmuster in verschiedenen Geweben oder Zelllinien kann Aufschluss über die Funktion einzelner Gene geben. Z. B. wurden auf diese Weise mehrere Gene identifiziert, die spezifisch nur in Tumorzellen, nicht aber in normalem Gewebe exprimiert werden (Velculescu et al., 1999). Einige dieser Gene waren bereits bekannt (z. B. Metastasin und Survivin), andere waren dagegen nur als ESTs mit bisher unbekannter Funktion beobachtet worden.
2. Microarray-Experimente machen ebenfalls quantitative Aussagen über die Expression von zahlreichen Genen (Lockhart & Winzler, 2000). Sie beschränken sich aber auf Gene, die bereits teilweise sequenziert sind, und betrachten (beim Menschen) nur eine vorher ausgewählte Untermenge der bekannten Gene. Schwierigkeiten bereitet bei dieser Methode die Bestimmung absoluter Expressionslevel, während es relativ zuverlässig möglich ist, Veränderungen festzustellen oder Expressionsmuster zu vergleichen. Zusätzlich können Kreuzhybridisierungen verschiedener mRNAs und die geringe Sensitivität die Aussagefähigkeit der Daten beeinträchtigen. Koregulation in Microarray-Experimenten weist ebenfalls auf die Beteiligung an einem gemeinsamen zellulären Prozess hin (z. B. Kap. 4.2: Stearoyl-Desaturase als Enzym des Cholesterinmetabolismus).
3. Auch das Auftreten von ESTs in verschiedenen Geweben kann Hinweise auf die Funktion eines Proteins geben (Okubo et al., 1992). Nachteile

1 Einleitung

dieses Verfahrens sind die ungleichmäßige Repräsentation von verschiedenen Geweben in den EST-Datenbanken, sowie die fehlende Normalisierung der Daten.

Wenn nahe Verwandte eines Proteins bereits kristallisiert wurden, lassen sich zusätzliche Detailinformationen über die Funktionsweise eines neuen Enzyms auch durch homologiebasierte Modellierung seiner dreidimensionalen Struktur gewinnen (Peitsch et al., 1992).

Es stehen also zahlreiche Möglichkeiten zur Verfügung, ein neues Protein auf bioinformatischem Wege zu charakterisieren. Schwierigkeiten bereitet aber immer noch die Integration aller verfügbaren Daten zu einem biologisch sinnvollen Gesamtbild (Fuchs, 2000; Marcotte et al., 1999). Die derzeit im Aufbau befindlichen Datenbanken stellen die Informationen aus verschiedenen Ansätzen vergleichend nebeneinander (Costanzo et al., 2000; Ball et al., 2000; Blake et al., 2000), aber die individuelle Interpretation erfordert immer noch zu viel biologisches Hintergrundwissen und verlangt zu viele Einzelentscheidungen über Analyseparameter und Signifikanzen, als dass sie vollständig automatisierbar wäre. Teilweise sind die Informationen, die sich mit einer einzelnen Methode gewinnen lassen, unvollständig oder nur wenig signifikant und gewinnen erst durch die Kombination mit ganz anderen Daten eine unerwartete Bedeutung (siehe z. B. Kap. 4.2). Die Integration der Daten ist also weiterhin der entscheidende Schritt zur Entwicklung von zuverlässigen Arbeitshypothesen.

1.2 Zielsetzung

Wie eine solche integrierende bioinformatische Analyse aussehen könnte, sollte in der vorliegenden Arbeit am Beispiel der 17β -Hydroxysteroiddehydrogenasen (17β -HSDs) gezeigt werden. Durch Kombination eines möglichst breiten Spektrums von Einzelinformationen sollten neue Proteinfunktionen beschrieben und die bekannten 17β -HSDs in ihren physiologischen Zusammenhang eingeordnet werden.

17β -HSDs katalysieren die Aktivierung und Inaktivierung von Steroidhormonen (Östradiol, Testosteron) durch Umsetzung an der Position 17 des Steroidgerüsts. Ihre Aktivität ist für die normale Sexualphysiologie notwendig, sie sind aber auch an der Pathogenese von verschiedenen Krankheiten beteiligt, insbesondere an der Entstehung und Proliferation hormonabhängiger Tumore (Labrie et al., 2000; Henderson & Feigelson, 2000). Verschiedene Isoenzyme mit breitem Substratspektrum wirken dabei in komplexer Weise zusammen. Auch pharmakologische und diätetische Eingriffe wirken sich in der Regel auf mehrere Isoenzyme gleichzeitig aus. Die 17β -HSDs sind Mitglieder einer alten, großen und sehr komplexen Proteinfamilie (Jörnvall, 1999), deren Funktion und Entstehung weitgehend ungeklärt ist (Jörnvall et al., 1999). In der

vorliegenden Arbeit sollten phylogenetische Untersuchungen und verschiedene bioinformatische Verfahren kombiniert werden, um ein genaueres Verständnis der physiologischen Bedeutung der 17β -HSDs zu erreichen.

1. Als Grundlage dieser Analyse sollte zunächst die Entwicklungsgeschichte der HSD-Familie im Detail beschrieben werden. Zu diesem Zweck sollten mit einem neu zu entwickelnden Verfahren Proteinsequenz- und Röntgenstrukturdaten zu einem konsistenten Evolutionsszenario kombiniert werden. Vor dem Hintergrund dieses neuen Szenarios sollten dann ausgewählte 17β -HSDs genauer betrachtet werden, mit dem Ziel, durch den Einsatz bioinformatischer Methoden die physiologische Funktion einzelner Enzyme genauer zu definieren.
2. Die dazu notwendigen Methoden wurden vorab in zwei Pilotstudien an besser charakterisierten Datensätzen validiert. In der ersten Pilotstudie sollte die frühe Evolutionsgeschichte und Herkunft der Paired-Box-Proteine bestimmt werden. Dabei sollten gleichzeitig die Verfahren getestet werden, die zur Analyse großer Datensätze notwendig sind. In der zweiten Pilotstudie sollte am Beispiel der Cholin-Acetyl-Transferase-Familie die Möglichkeit geprüft werden, aus stammesgeschichtlichen Daten unbekannte enzymatische Aktivitäten abzulesen.
3. Schwerpunkt der Arbeit sollte die funktionelle Charakterisierung der humanen 17β -HSD Typ 7 sein. Dieses Östradiol-produzierende Protein war bereits in Nagetieren untersucht worden. Seine Bedeutung für die Humanphysiologie war jedoch unklar. Mit Hilfe von Expressions-Profiling, Molekülmodellierung, vergleichender Genom- und Promotor-Analyse sowie mit phylogenetischen Methoden, sollte die Position der 17β -HSD7 im humanen Stoffwechsel bestimmt werden.
4. An zwei weiteren 17β -HSD-Isoenzymen (Typ 1 und 4) wurden einzelne Aspekte genauer untersucht, die sich im Rahmen der Arbeit an 17β -HSD7 als wichtig erwiesen hatten: Zum einen sollte durch vergleichende Genanalyse die Regulation der human 17β -HSD1 untersucht werden, deren physiologische Funktion und Gewebeverteilung große Ähnlichkeiten mit 17β -HSD7 zeigt. Zum anderen sollte am Beispiel der 17β -HSD1 durch eine umfassende evolutionäre Betrachtung herausgefunden werden, wie sich die Hydroxysteroid-Dehydrogenase-Aktivität bei einem Enzym entwickelt hat, das ursprünglich eine ganz andere physiologische Funktion (nämlich bei der β -Oxidation von Fettsäuren) hatte. Anlass dazu war die Beobachtung, dass 17β -HSD7 vermutlich an der Cholesterinbiosynthese beteiligt ist.
5. Am Beispiel der 17β -HSD Typ 5 sollte schließlich ein Verfahren entwickelt werden, das es ermöglicht, die in Protein-Alignments enthaltenen Informationen mit dreidimensionalen Strukturdaten zu verbinden. Die

1 Einleitung

so gewonnenen Ergebnisse sollten dann zur Interpretation von *in vitro*-Inhibitionsstudien verwendet werden.

2 Verwendete Computerprogramme und Datenbanken

2.1 Computerprogramme

Parsimony- und Neighbour-Joining-Methoden zur phylogenetischen Analyse:

PHYLIP – PHYLogeny Inference Package

<http://evolution.genetics.washington.edu/phylip.html>

Felsenstein J (1989): PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.

Dieses Programmpaket enthält Implementierungen von verschiedenen Algorithmen zur phylogenetischen Analyse von Protein- und Nukleotidsequenzen. Als Eingabe dient in den meisten Fällen ein Sequenzalignment, aus dem sich zum Beispiel Distanzmatrizen (`protdist`) oder Stammbäume nach dem Neighbour-joining- oder Parsimony-Verfahren berechnen lassen (`neighbour`, `fitch`, `protpars`).

Maximum-Likelihood-Verfahren:

Tree-Puzzle 4.0.2

<http://www.tree-puzzle.de/>

Strimmer & von Haeseler (1996): Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13: 964–965

Strimmer K & von Haeseler A (1997): Likelihood mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci.* 94: 6815–6819.

Dieses Programmpaket implementiert effektive Algorithmen zur Berechnung von Maximum likelihood-Stammbäumen, zum statistischen Vergleich von alternativen Stammbäumen, sowie zum Likelihood-Mapping, einem Verfahren, den phylogenetischen Informationsgehalt eines Alignments grafisch abzuschätzen.

Redundanz-Filterung und Sekundärstruktur-Vorhersage:

JPred – Consensus Secondary Structure Prediction Server

<http://jura.ebi.ac.uk:8888/>

Cuff JA, Clamp ME, Siddiqui AS, Finlay M & Barton GJ (1998): JPred:

2 Verwendete Computerprogramme und Datenbanken

a consensus secondary structure prediction server. *Bioinformatics* 14: 892–893. Der JPred-Server kombiniert verschiedene Verfahren zur Sekundärstrukturvorhersage und berechnet einen (gewichteten) Konsensus der verschiedenen Vorhersagemethoden. Als Eingabe dient entweder ein fertiges Proteinalignment oder eine Einzelsequenz, für die dann vor der Vorhersage automatisch ein Alignment mit verwandten Sequenzen erstellt wird. Ein Vergleich der Vorhersageergebnisse mit bekannten Röntgenkristallstrukturen zeigte die generell hohe Zuverlässigkeit des Konsensus-Algorithmus. Der Server lässt sich außerdem nutzen, um eine Redundanzfilterung eines Datensatzes vorzunehmen, bei der ähnliche Sequenzen solange entfernt werden, bis keine paarweisen Ähnlichkeiten >75% beobachtet werden.

Vergleich von Protein-Primärstrukturen:

Gapped-Blast und PSI-Blast

<http://www.ncbi.nlm.nih.gov/BLAST/>

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25: 3389–3402.

Diese Algorithmen dienen zur schnellen Suche nach ähnlichen Proteinen (oder Nukleotidsequenzen) in großen Datenbanken. Sie beruhen auf der Suche nach kurzen übereinstimmenden Sequenzen (“*local alignments*”). Nur wenn eine signifikante Übereinstimmung („Treffer“) in einem kleinen Sequenzbereich entdeckt wird, wird der Vergleich auf die ganze Sequenz ausgedehnt, um den übereinstimmenden Sequenzbereich zu identifizieren. Die Wahrscheinlichkeit, in den großen Sequenzdatenbanken irgendeine (in einem kleinen Bereich) übereinstimmende Sequenz zu entdecken, ist relativ hoch. Es lässt sich jedoch abschätzen, mit welcher Wahrscheinlichkeit eine bestimmte Übereinstimmung per Zufall in einer Datenbank gegebener Größe beobachtet wird. Diese Wahrscheinlichkeit wird als E-Score bezeichnet und erlaubt eine schnelle Beurteilung der Qualität eines Treffers, zumindest dann, wenn die Zusammensetzung der Sequenz nicht deutlich von einer Zufallssequenz abweicht – in poly-Prolin-Regionen oder Repeat-Elementen wäre der E-Score dagegen kaum aussagekräftig. Beim PSI-Blast-Verfahren werden nach dem ersten Suchschritt alle gefundenen Treffer zu einer Matrix vereinigt, die die Aminosäurehäufigkeiten an jeder Position beschreibt. Diese Matrix dient dann in weiteren Blast-Schritten als Eingabe, anstelle einer einzelnen Proteinsequenz.

Vorhersage von Transmembran-Bereichen:

TMHMM – Hidden Markov Modell-basierte Vorhersage

<http://www.cbs.dtu.dk/services/TMHMM-2.0>

Sonnhammer ELL, von Heijne G & Krogh A (1998): A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. of the Sixth Conference on Intelligent Systems for Molecular Biology.* 175–182.

Die Vorhersage von Transmembranbereichen in Proteinen beruht bei diesem Programm nicht auf der Bestimmung hydrophober Sequenzbereiche, wie bei klassischen Verfahren (z. B. Engelman et al., 1986), sondern auf einem Hidden-Markov-Modell, das mit den Sequenzen bekannter Transmembran-Helices trainiert wurde. Dadurch wird die Spezifität der Vorhersagen deutlich verbessert und die Zahl falsch-positiver Vorhersagen reduziert. Zusätzlich wird die mutmaßliche Topologie der Transmembranbereiche vorhergesagt.

Proteinalignments

ClustalW

<http://www2.ebi.ac.uk/clustalw>

Thompson JD, Higgins DG & Gibson TJ (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22: 4673–4680.

Dieses Programm erlaubt die Berechnung von Sequenzalignments zahlreicher Proteine. Dabei werden zuerst die Proteine paarweise verglichen und dann iterativ zu Clustern zusammengefasst und “aligned”. Dabei sind verschiedene Aminosäure-Ersetzungsmodelle auswählbar, sowie verschiedene “gap penalties”, die angeben, wie das Einführen von Lücken im Alignment behandelt werden soll. Trotz dieser Flexibilität ist der Algorithmus gerade bei weitläufig verwandten Proteinen oft überfordert, wie der Vergleich von ClustalW-Ergebnissen und Röntgenstruktur-basierten Alignments zeigt. Das liegt zum einen an der Schwierigkeit, die Lage von Insertionen in der Proteinsequenz korrekt vorherzusagen, zum anderen aber auch an dem grundlegenden Algorithmus, der nicht versucht, ein global-optimiertes Alignment zu berechnen, sondern schrittweise einzelne Cluster “aligned”, so dass das Alignmentergebnis entscheidend von den jeweils im Datensatz enthaltenen Sequenzen abhängen kann.

Vergleich von Protein-Tertiärstrukturen:

DALI-Server

<http://www2.ebi.ac.uk/dali/>

Holm L & Sander C (1995): Dali: a network tool protein structure comparison. Trends Biochem. Sci. 20: 478–480.

Dieses Programm berechnet paarweise Vergleiche von Röntgenstrukturen anhand der dreidimensionalen Koordinaten ihres Proteinrückgrats. Dabei versucht es, lokale Übereinstimmungen zu optimieren und äquivalente Aminosäuren in den verglichenen Strukturen zu identifizieren. Das Ergebnis ist ein allgemeiner Ähnlichkeitsparameter (Z-Score) und die mittlere quadratische Abweichung der Koordinaten (*root mean square*), sowie ein Sequenzalignment der verglichenen Proteine.

Statistische Analyse von Proteinalignments:

STATRAS2

Quellcode s. Anhang

Das für diese Arbeit erstellte Programm STATRAS2 zur dreidimensionalen Darstellung von statistischen Parametern eines Proteinalignments wird in Kap. 4.5 (S. 94) näher beschrieben. Es dient dazu anhand eines umfangreichen Proteinalignments zu quantifizieren, welche Aminosäurepositionen innerhalb einer Proteinfamilie konserviert sind. Die erhaltenen Daten werden zur besseren Visualisierung mit den Röntgenstrukturdaten eines Familienmitglieds kombiniert, so dass auf einen Blick deutlich wird, welche räumlichen Beziehungen zwischen den konservierten Aminosäuren bestehen.

Homologie-basiertes Modellieren von Proteinstrukturen:

Swiss-Model – automated comparative protein modeling server und SwissPDBViewer

<http://www.expasy.ch/swissmod/SWISS-MODEL.html>

Peitsch MC (1996): ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochem. Soc. Trans. 24: 274–279.

Guex N & Peitsch MC (1997): Swiss-Model and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714–2723.

Der Swiss-Model-Server erlaubt die fast-automatische Erzeugung von dreidimensionalen Proteinmodellen anhand der Röntgenstrukturen von homologen Proteinen. Als Grundlage dient ein Sequenzalignment. Die Sequenz des zu modellierenden Proteins wird dann auf die Struktur ihres Homologen „aufgefädelt“. Lücken und Insertionen werden anschließend separat behandelt. Das erste, grobe Modell wird dann schrittweise durch lokale Anpassungen und molekular

2 *Verwendete Computerprogramme und Datenbanken*

laddynamische Berechnungen optimiert und schließlich die Seitenketten eingefügt. Mit Hilfe des SwissPDBViewers lassen sich das Ausgangsalignment manuell anpassen und die Qualität der Modellstruktur überprüfen. Außerdem wird das Modell automatisch mit dem WhatCheck-Programm (Hooft et al., 1996) auf Abweichungen von den röntgenkristallografisch beobachteten Strukturparametern überprüft. Diese Information kann ebenfalls genutzt werden, um die Qualität des Modells abzuschätzen.

Grafische Darstellung von Proteinstrukturen:

MolScript v2.1.2 und MolAuto

<http://www.avatar.se/molscript/>

Kraulis PJ (1991): MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallography* 24: 946–950.

Raster3D v2.3

<http://www.bmsc.washington.edu/raster3d/raster3d.html>

Merritt EA & Bacon DJ(1997): Raster3D: Photorealistic Molecular Graphics. *Meth. Enzymol.* 277: 505–524.

Die Kombination dieser beiden Programme ermöglicht die flexible Darstellung von Proteinstrukturen in Fotoqualität.

Vorhersage von Promotor-Bereichen in genomischen Sequenzen:

PromoterInspector

<http://genomatix.gsf.de/cgi-bin/>

[promoterinspector/promoterinspector.pl](http://genomatix.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl)

Scherf M, Klingenhoff A & Werner T (2000): Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* 297: 599–606.

Dieser Algorithmus entdeckt Promotoren in großen genomischen Sequenzen mit hoher Zuverlässigkeit und einer sehr geringen Zahl von falsch-positiven Vorhersagen.

Modellierung von konservierten Promotor-Modulen:

GEMS-Launcher (einschließlich MatInspector, ModelInspector und FastM)

<http://genomatix.gsf.de>

Quandt K, Frech K, Karas H, Wingender E & Werner T (1995): MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl. Acids Res.* 23: 4878–4884.

Klingenhoff A, Frech K, Quandt K & Werner T (1999): Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. *Bioinformatics* 15: 180–186.

Frech K, Danescu-Mayer J & Werner T (1997): A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J. Mol. Biol.* 270: 674–687.

Dieses Programmpaket erlaubt die komfortable Analyse von Nukleotidsequenzen. Unter anderem ist es möglich, Transkriptionsfaktor-Bindungsstellen vorherzusagen und Modelle für das kombinierte Auftreten von verschiedenen Transkriptionsfaktoren in festgelegter Orientierung und bestimmtem Abstand zu definieren. Diese Promotor-Modelle können dann gegen große Sequenzdatenbanken auf ihre Spezifität überprüft werden.

2.2 Datenbanken

Protein-Primärstrukturen:

Swiss-Prot und TrEMBL

<http://www.expasy.ch/sprot/>

Bairoch A & Apweiler R (2000): The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* 28: 45–48.

Eine ausführlich annotierte Datenbank von Proteinsequenzen, die im Vergleich zu den Datenbanken des National Center for Biotechnology Information intensiver gewartet wird. Der Umfang der Datenbank ist viel kleiner und enthält kaum redundante Sequenzen, dafür fehlen aber auch Einträge für viele Proteine. Die Daten sind durch umfangreiche Querverweise auf andere Datenbanken erschlossen.

Protein-Tertiärstrukturen:

PDB – Protein Data Bank

<http://www.rcsb.org/pdb/>

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000): The Protein Data Bank. *Nucl. Acids Res.* 28: 235–242.

Datenbank aller dreidimensionalen Proteinstrukturen (12 777 Strukturen am 18. Okt. 2000), die meisten davon Röntgenkristallstrukturen, aber auch eine wachsende Zahl von NMR-Daten.

Protein-Tertiärstruktur-Vergleiche:

FSSP - Fold classification based on structure-structure alignment of proteins

<http://www2.ebi.ac.uk/dali/fssp/fssp.html>

Holm L & Sander C (1996): The FSSP database: fold classification based on structure-structure alignments of proteins. *Nucl. Acids Res.* 24: 206–209.

Auf dem DALI-Programm (siehe oben) aufbauende Datenbank von bereits fertig berechneten Proteinstruktur-Vergleichen, einschließlich einer hierarchischen Klassifikation der Strukturen.

cDNA-Sequenzen, *expressed sequence tags* und genomische Sequenzen:

GenBank database (einschließlich dbEST und High Throughput Genomic [HTG] Sequences)

<http://www.ncbi.nlm.nih.gov>

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA & Wheeler DL (2000): GenBank. *Nucl. Acids Res.* 28: 15–18.

Boguski MS, Lowe TM & Tolstoshev CM (1993): dbEST – database for “expressed sequence tags”. *Nat. Genet.* 4: 332–333.

Ouellette BF & Boguski MS (1997): Database divisions and homology search files: a guide for the perplexed. *Genome Res.* 7: 952–955.

Die Datenbank des National Center for Biotechnology Information enthält die größte Menge an öffentlich zugänglichen Nukleotid-Sequenzen. Spezialisierte Unterdateien enthalten die Sequenzen von *expressed sequence tags* und die vorläufigen Sequenzierergebnisse der großen Genom-Sequenzierprojekte. Die Daten sind durch eine komfortable Oberfläche (Entrez) zugänglich und sind vielfältig untereinander und mit anderen Datenbanken vernetzt.

2 *Verwendete Computerprogramme und Datenbanken*

Serial Analysis of Gene Expression:

SAGEmap

<http://www.ncbi.nlm.nih.gov/SAGE/>

Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul SF (2000): SAGEmap: a public gene expression resource. *Genome Res.* 10: 1051–1060.

Diese Datenbank enthält die veröffentlichten Ergebnisse von SAGE-Experimenten, bei denen die Expressionslevel von Genen quantitativ bestimmt werden. Es ist möglich, die Expression einzelner Gene in allen Experimenten vergleichend zu ermitteln.

Microarray-Experimente:

Stanford Microarray Database

<http://genome-www4.stanford.edu/MicroArray/SMD/>

Eisen MB, Spellman PT, Brown PO & Botstein D (1998): Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* 95: 14863–14868.

Diese Datenbank ermöglicht den Zugriff auf eine wachsende Anzahl von Microarray-Daten (insgesamt 13 Experimente im Jahr 2000). Durch die Verwendung des von Eisen et al. (1998) beschriebenen Clustering-Algorithmus kann nicht nur das Expressionsniveau einzelner Gene bestimmt werden, sondern auch die Identifizierung koregulierter Proteine ist in einem einzigen Schritt möglich. Von einzelnen Experimenten sind außerdem die detaillierten Primärdaten verfügbar.

Transkriptionsfaktor-Bindungsstellen:

TRANSFAC – The Transcription Factor Database

<http://transfac.gbf.de/TRANSFAC/index.html>

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I & Schacherer F (2000): TRANSFAC: an integrated system for gene expression regulation. *Nucl. Acids Res.* 28: 316–319.

Diese Datenbank enthält die Beschreibung von experimentell charakterisierten Transkriptionsfaktor-Bindungsmotiven in Form von Matrizen, die die Nukleotidpräferenz an den einzelnen Positionen der Bindungsstelle beschreiben und vom PromoterInspector (siehe oben) zur Vorhersage von Transkriptionsfaktor-Bindungsstellen genutzt werden können.

Eukaryontische Promotor-Sequenzen:

EPD – Eukaryotic Promoter Databank Release 63

<http://www.epd.isb-sib.ch/>

Perier RC, Praz V, Junier T, Bonnard C & Bucher P (2000): The eukaryotic promoter database (EPD). *Nucl. Acid Res.* 28: 302–303.

Annotierte, nicht redundante Datenbank experimentell charakterisierter eukaryontischer Promotoren.

Aktuelle Nomenklatur der Aldo-Keto-Reduktase-Gene:

HGNC Gene Family Nomenclature

<http://www.gene.ucl.ac.uk/nomenclature/AKR.shtml>

Jez JM, Flynn TG & Penning TM (1997): A new nomenclature for the aldo-keto reductase superfamily. *Biochem. Pharmacol.* 64: 639–647.

Datenbank der offiziell empfohlenen Nomenklatur der AKR-Familie.

3 Pilotstudien

3.1 Pilotstudie I: Paired-Box-Proteine – Entwicklung eines komplexen Evolutionsszenarios für eine sehr große Proteinfamilie¹

3.1.1 Problemstellung

Besondere Schwierigkeiten bei der phylogenetischen Analyse der Hydroxysteroid-Dehydrogenasen entstehen durch die Größe dieser Proteinfamilie. Der Rechenaufwand für die meisten Algorithmen, die ja auf paarweisen Sequenzvergleichen beruhen, steigt mit der Größe der Datensätze überproportional an. Große Datensätze lassen sich zudem schwerer handhaben und enthalten oft redundante Sequenzen: Fragmente, Splicevarianten und identische Sequenzen unter verschiedenen Namen. Außerdem sind oft sehr ähnliche Sequenzen aus nahe verwandten Organismen enthalten, die den Rechenaufwand erhöhen, ohne zur phylogenetischen Information entscheidend beizutragen. Eine detaillierte Bearbeitung mit den zur Verfügung stehenden Computern erfordert deshalb eine sinnvolle Reduktion des Datensatzes, die am Besten automatisiert erfolgt. Dabei sollten redundante Sequenzen entfernt werden, ohne dadurch phylogenetisch relevante Informationen zu verlieren.

Ein geeigneter Datensatz zur Entwicklung und Validierung der notwendigen Methoden zum Umgang mit großen Datensätzen sollte vier Kriterien erfüllen:

- Er sollte etwa so viele Proteine enthalten wie die HSD-Familie.
- Die Funktion der Proteine sollte weitgehend bekannt sein.
- Ihre Evolution sollte schon untersucht worden sein, damit eine unabhängige Kontrolle der hier erhaltenen Ergebnisse möglich ist.
- Und die untersuchte Fragestellung sollte nicht nur methodisch, sondern auch evolutionsbiologisch interessant sein.

¹ Teilweise veröffentlicht in: Breitling R & Gerber JK (2000): Origin of the paired domain. *Dev. Genes Evol.* 210: 644–650.

3 Pilotstudien

Für diese Pilotstudie wurde deshalb die Familie der Paired-Box-Proteine untersucht. Diese Proteinfamilie enthält derzeit einschließlich der verwandten Transposasen etwa 1000 Mitglieder in den öffentlichen Datenbanken. Ihr Umfang ist damit vergleichbar mit dem der HSD-Familie. Die Paired-Box-Familie ist funktionell gut charakterisiert und es liegen bereits einige detaillierte phylogenetische Untersuchungen vor, die zur Validierung der hier eingesetzten Verfahren geeignet waren.

Ähnlich wie bei der HSD-Familie ist auch in der Paired-Box-Familie die frühe Differenzierung in Unterfamilien besonders aufschlussreich, weil sie Einblick in die Funktion der ältesten Proteine geben kann. Vor allem ist die Frage interessant, wann und wie die Paired-Box-Familie entstanden ist. In der Pilotstudie wurde deshalb versucht, durch die Kombination verschiedener phylogenetischer Analyseverfahren ein evolutionäres Szenario für die Paired-Box-Familie zu entwerfen und gleichzeitig Methoden zu etablieren, die den Umgang mit dem großen HSD-Datensatz erleichtern würden. Außerdem wurde von dem Ergebnis erwartet, dass es einige Unstimmigkeiten in den bisher veröffentlichten Evolutionsszenarien der Paired-Box-Familie klären könnte.

3.1.2 Hintergrund

Funktion der Pax-Proteine

Paired-Box-Proteine (Pax-Proteine) sind Transkriptionsfaktoren, die als regulatorische Faktoren an der Organ-Differenzierung beteiligt sind (Balling et al., 1996; Dahl et al., 1997; Mansouri et al., 1999). Ihre Funktion ist teilweise hoch konserviert. Das klassische Beispiel ist Pax6, das „*master control gene*“ der Augenentwicklung (Gehring & Ikeo, 1999). Ektopische Expression von Pax6 aus Mäusen oder Tintenfischen (*Loligo*) induziert überzählige Augen in *Drosophila* (Halder et al., 1995; Tomarev et al., 1997). Die Pax-Proteine werden bei Wirbeltieren in verschiedenen segmentierten Strukturen (Wirbelsäule, Nieren, Muskelanlagen) exprimiert und kommen mit Ausnahme von Pax1 und Pax9 alle im entstehenden Nervensystem vor (Tab. 3.1, S. 23).

Struktur und Einteilung der Pax-Proteine

Das charakteristische Element der Pax-Proteine ist die Paired-Box, eine hoch konservierte DNA-Bindungsdomäne von etwa 128 Aminosäuren. Sie besteht aus zwei Untereinheiten, der N-terminalen PAI-Domäne und der C-terminalen RED-Domäne, die beide ein Helix-Turn-Helix-Motiv enthalten. Beide Untereinheiten können unabhängig an DNA binden (Czerny et al., 1993; Epstein et al., 1994; Vogan et al., 1996, 1997; Kozmik et al., 1997). Beim Paired-Protein in *Drosophila* ist die N-terminale Untereinheit sogar ausreichend für die volle biologische Funktion des ganzen Proteins (Bertuccioli et al., 1996). Die C-terminale Untereinheit ist etwas weniger konserviert und ist vielleicht für das Protein-Targeting notwendig (Poleev et al., 1997). Neben der Paired-Box enthalten die meisten Pax-Proteine noch eine zweite DNA-Bindungsdomäne aus der Homöobox-Familie. Die Anwesenheit und Struktur dieser Homöo-Domäne und eines weiteren kurzen Motivs, des Oktapeptids, diente zur Klassifizierung der Pax-Proteine in fünf

3.1 Pilotstudie I: Paired-Box-Proteine

Tabelle 3.1: Hauptfunktionen der humanen Paired-Box-Proteine.

Protein	Hauptfunktionsort	humanes Syndrom	Referenzen
Pax1	Skelett	Spina bifida	Koseki et al. (1993); Wilm et al. (1998)
Pax2	Nieren, Augen	Renal-Coloboma-Syndrom	Sanyanusin et al. (1995); Torres et al. (1996)
Pax3	Myoblasten, Neuralleiste	Waardenburg-Syndrom Typ 1	Tassabehji et al. (1992); Goulding et al. (1994); Bober et al. (1994); Mansouri (1998); Tremblay et al. (1998)
Pax4	β -Zellen des Pankreas	unbekannt	Sosa-Pineda et al. (1997);
Pax5	B-Zellen	unbekannt	Urbanek et al. (1994)
Pax6	Augen, α -Zellen des Pankreas	Aniridia	Gehring (1996); St-Onge et al. (1997)
Pax7	Myoblasten, Neuralleiste	unbekannt	Mansouri (1998)
Pax8	Schilddrüse	Schilddrüsen-Dysplasie	Macchia et al. (1998)
Pax9	Skelett, Zähne, Kiementaschenderivate	Oligodontia	Peters et al. (1998); Stockton et al. (2000)

große Untergruppen, die nach ihren wichtigsten Mitgliedern benannt werden: Pax1/9/*meso*, PaxD/3/7/*paired/gooseberry*, Pax4/6/*eyeless*, PaxB/2/5/8/*sparkling* und PaxA/*neuro* (Strachnan & Read, 1994; Mansouri et al., 1996; Callaerts et al., 1997; Balczarek et al., 1997; Miller et al., 2000). Pax-Proteine kommen nur bei Metazoen vor und sind bei Vertebraten und *Drosophila* besonders genau untersucht. Ihre hoch-konservierte Funktion im gesamten Tierreich hat die Suche nach weiteren Homologen aus niederen Organismen veranlasst. Pax-Homologe wurden z. B. aus *Hydra* (PaxA und PaxB), Korallen (PaxA, PaxB, PaxC und PaxD) und Schwämmen (spongePax, ein PaxB/2/5/8/*sparkling*-Mitglied) isoliert (Sun et al., 1997; Hoshiyama et al., 1998; Miller et al., 2000). Mehr als 100 Pax-Protein-Sequenzen sind derzeit in den öffentlichen Datenbanken enthalten.

Kürzlich haben Galliot & Miller (2000) ein Evolutions-Szenario für die Pax-Proteine vorgestellt (Abb. 3.1A, S. 24): Ein PaxA-artiger Vorläufer, der nur aus der Paired-Box bestand, fusionierte in zwei unabhängigen Entwicklungslinien mit Homöobox-Proteinen. Die beiden Fusionsproteine entwickelten sich dann zur PaxB/2/5/8/*sparkling*-Familie und zur PaxC+1/9+3/7+4/6-Familiengruppe. Das erste Fusionsereignis soll vor der Differenzierung der Schwämme (vor etwa 900×10^6 Jahren) stattgefunden haben, das zweite vor der Auftrennung von Cnidariern und Triploblasten (vor etwa 700×10^6 Jahren). Im dazugehörigen Protein-Stammbaum (Abb. 3.2, S. 25) bildet PaxC die Schwestergruppe zu Pax1/9/*meso*, Pax3/7/*gooseberry/paired* und Pax4/6/*eyeless*. Ein ähnliches Szenario wurde auch von Catmull et al. (1998) vorge-

Evolution der Pax-Proteine

3 Pilotstudien

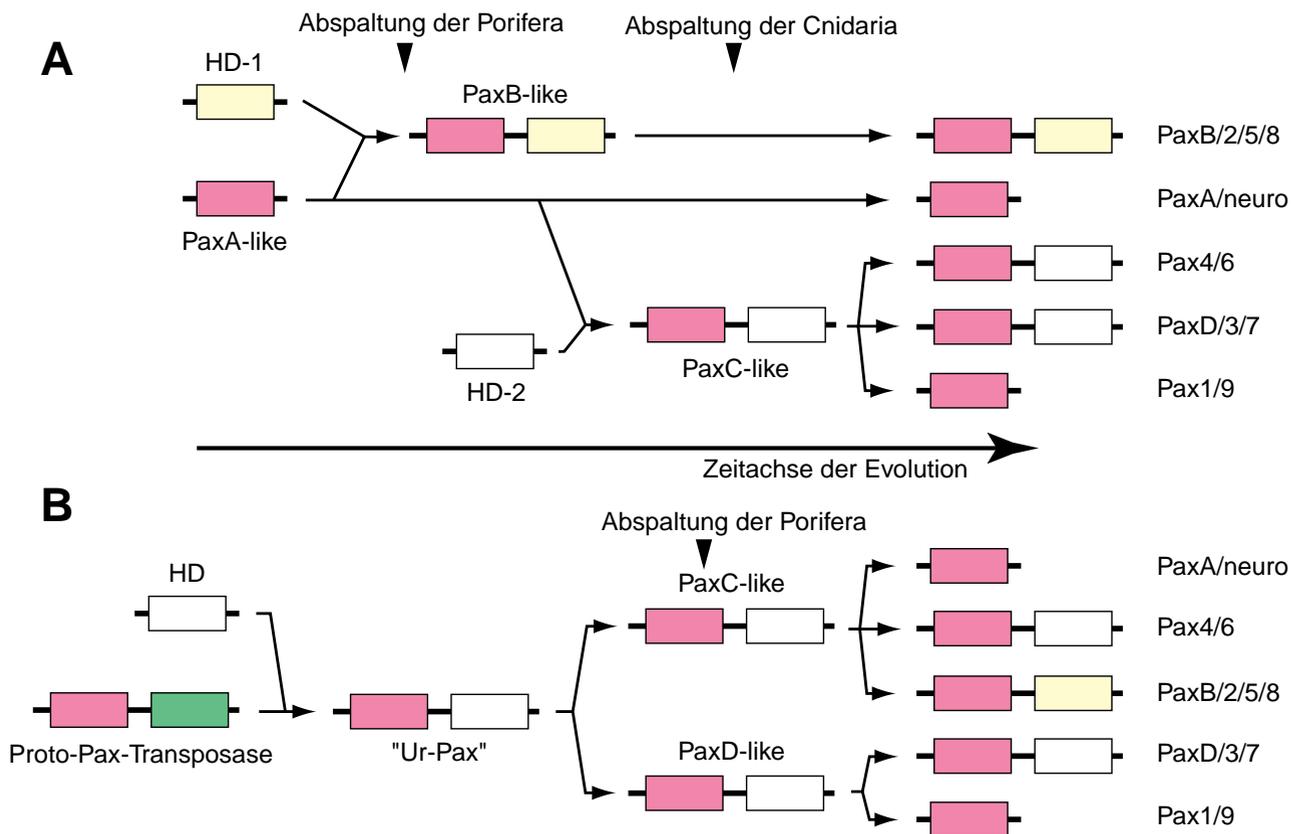


Abbildung 3.1: Alternative Evolutions-Szenarien für die Pax-Familie. **A** Szenario nach Galliot und Miller (2000). **B** Neues Szenario nach dem Einschluss der Transposasen in die evolutionäre Analyse. Weiß (HD): Homöodomäne, gelb: unvollständige Homöodomäne wie in der PaxB/2/5/8-Unterfamilie, rot: DNA-Bindungsdomäne, grün: katalytische Domäne der Proto-Pax-Transposase.

schlagen. Diesem Entwicklungsmodell widerspricht der Vorschlag von Hoshiyama et al. (1998), der allerdings vor der Beschreibung von PaxC entworfen wurde (Abb. 3.2, S. 25). In ihrem Stammbaum bilden Pax1/9/*meso* und Pax3/7/*gooseberry* die Schwestergruppe aller anderen Pax-Proteine. Die Wurzel des Stammbaums wurde in diesem Fall durch die Analyse der Homöodomänen von Pax3/7/*gooseberry/paired*, PaxB/2/5/8/*sparkling* und Pax4/6/*eyeless* festgestellt.

Woher stammen die Pax-Proteine?

Beide Modelle können nicht erklären, woher die ursprüngliche Paired-Box eigentlich stammt. Während nämlich Homöo-Domänen auch in Pflanzen (etwa 70 bekannte Sequenzen) und Hefen (10 bekannte Sequenzen) vorkommen (Kappen, 2000), ist die Paired-Box auf die Metazoa beschränkt. Ist sie dort gewissermaßen plötzlich „aus dem Nichts“ entstanden oder ist sie in den anderen Gruppen verlorengegangen? Die Beantwortung dieser Frage könnte helfen, die Funktion von Pax-Proteinen in den ursprünglichsten Tieren zu klären. Dieses Wissen liefert den Hintergrund für die Beantwortung zentraler Fragen der

3.1 Pilotstudie I: Paired-Box-Proteine

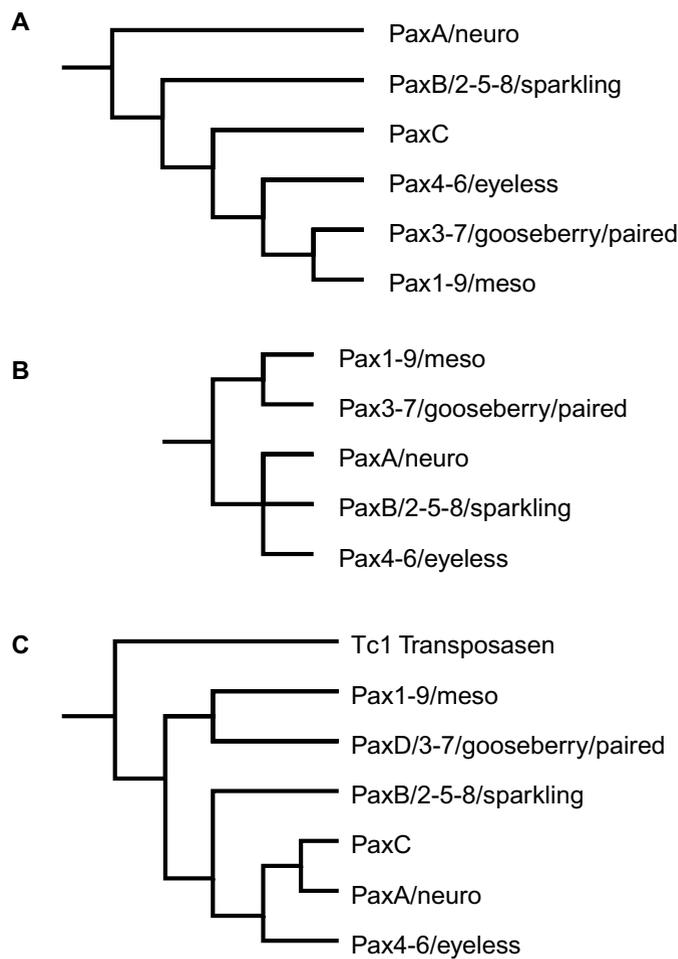


Abbildung 3.2: Stammbäume der Pax-Familie. **A** nach dem Szenario von Galliot und Miller (2000) und Catmull et al. (1998). **B** nach Hoshiyama et al. (1998). **C** nach der neuen Analyse mit Transposasen als Outgroup.

Entwicklungsbiologie, z. B.:

- Welche Entwicklungshomologien sind hinter der Vielfalt der heutigen Körperbaupläne verborgen?
- Wie sind die Verschiedenheiten ursprünglich entstanden?
- Oder auch: Hat das erste Pax6 bereits die Entwicklung eines Proto-Auges kontrolliert?

3.1.3 Ergebnisse und Diskussion

Die Evolution einer Proteinfamilie kann nur dann zuverlässig geklärt werden, wenn die Wurzel ihres Stammbaums korrekt bestimmt wird. Zu diesem Zweck

Bestimmung einer Outgroup der Pax-Proteine

3 Pilotstudien

hatten Hoshiyama et al. (1998) für die Pax-Proteine die Homöo-Domäne verwendet. Damals war aber das PaxA-Protein unbekannt, dem eine Homöodomäne fehlt und das eine zentrale Rolle im Evolutionsszenario von Galliot & Miller (2000) spielt. Um diese Neuentdeckung angemessen zu berücksichtigen, musste jetzt ein neuer Weg gefunden werden, den nächsten gemeinsamen Verwandten aller Pax-Proteine zu bestimmen. Zu diesem Zweck wurde in der vorliegenden Arbeit ein Vergleich der Röntgenstrukturen von *Drosophila*-Paired (1PDN, Xu et al., 1995) und humanem Pax6 (6PAX, Xu et al., 1999) mit allen bekannten Proteinstrukturen durchgeführt, der zeigte, dass die N-terminale Subdomäne (PAI-Domäne) der Pax-Proteine auffallende Übereinstimmung mit der Tc3-Transposase von *C. elegans* (1TC3, van Pouderooyen et al., 1997) zeigt. Der DALI-Z-Score für Pax6 und Tc3 beträgt 6.5. Andere Helix-Turn-Helix-Proteine, z. B. die Homöodomäne von *Drosophila*-Engrailed (1ENH, Kissinger et al., 1990) haben Z-Scores unter 5. Alle diese Scores sind relativ niedrig (Mitglieder der HSD-Familie haben untereinander DALI-Scores von über 20 und verschiedene Strukturen desselben Proteins haben sogar Werte über 30), sie zeigen aber trotzdem, dass die Strukturen in entscheidenden Merkmalen übereinstimmen und dass die Ähnlichkeit zwischen Paired-Domäne und Transposasen größer ist als die Ähnlichkeit zwischen Paired-Domäne und anderen Helix-Turn-Helix-Domänen. Über Ähnlichkeiten zwischen der Paired-Box und Transposase-DNA-Bindungsdomänen wurde bereits früher berichtet (Ivics et al., 1996). Die strukturellen Übereinstimmungen werden auch von van Pouderooyen et al. (1997) bemerkt, die eine Superposition der beiden Kristallstrukturen zeigen. Transposasen sind alte, weitverbreitete Proteine und kommen in Eukaryonten und Prokaryonten vor, sie sind also gute Kandidaten als Paired-Box-Vorläufer.

PSI-Blast Suche nach dem nächsten Verwandten

Durch PSI-Blast-Suchen wurde eine Gruppe von Transposasen aus *C. elegans* identifiziert, die eine engere Verwandtschaft mit den Pax-Proteinen als mit den meisten anderen Transposasen zeigten. Die DNA-Bindungsdomäne dieser *C. elegans*-Transposasen (Proteine K03H6.3, W04G5.1, F26H9.3, F49C5.8 und C27H2.1) zeigt größere Übereinstimmung mit vielen Pax-Proteinen (z. B. *Hydra magnapapillata* Pax2/5/8, E-Score (E) = 9×10^{-5} ; *Phallusia mammilata* Pax6, $E = 3 \times 10^{-4}$; oder *Paracentrotus lividus* Pax1/9, $E = 6 \times 10^{-4}$), und mit der DNA-Bindungsdomäne von Bmmar1, einer Transposase aus *Bombyx mori* ($E = 2 \times 10^{-27}$). Im Vergleich dazu sind die E-Scores von anderen Transposasen schlechter als 1×10^{-3} (z. B. *Anopheles albimanus*-Transposase AAB02109, $E = 9 \times 10^{-3}$). Aufgrund des Blast-Ergebnisses wurde daher angenommen, dass die Transposasen aus *C. elegans* und *Bombyx mori* „molekulare Fossilien“ sein könnten, sozusagen Proto-Pax-Transposasen, die aus der Zeit vor der Fusion mit der Homöo-Box stammen. Die ersten Pax-Proteine wären dann vor etwa einer Milliarde Jahren durch ein Fusionsereignis entstanden, bei dem die katalytische Domäne der Proto-Pax-Transposasen verloren ging und ihre DNA-Bindungsdomäne mit einer Homöodomäne fusionierte. Der auf diese Weise entstandene Hybrid-Transkriptionsfaktor hätte neue Re-

3.1 Pilotstudie I: Paired-Box-Proteine

gulationsmöglichkeiten eröffnet und seine zwei DNA-Bindungsdomänen differenzierten sich in der Folgezeit zu den verschiedenen Pax-Typen. Im Gegensatz zur Paired-Domäne sind Homöodomänen sehr alte Proteine, die in allen Eukaryonten-Gruppen vorkommen (Kappen, 2000), und Transposasen sind noch weiter verbreitet. Die Voraussetzungen für das vorgeschlagene Szenario sind also gegeben.

Dieses Szenario setzt voraus, dass die DNA-Bindungsdomäne der Transposasen beide Untereinheiten der Paired-Box (PAI- und RED-Domäne) enthält. Die Proteinsequenz zeigt keine signifikante Ähnlichkeit im C-terminalen Bereich. Eine Konsensus-basierte Vorhersage der Sekundärstruktur der Proto-Pax-Transposasen zeigte aber trotzdem zwei Helix-Turn-Helix-Motive, die den beiden Untereinheiten der Paired-Box homolog sind, entsprechend den Ergebnissen von Ivics et al. (1996). Also ist wahrscheinlich die ganze Paired-Domäne der Transposase-DNA-Bindungsdomäne homolog. Die beiden DNA-bindenden Subdomänen (PAI und RED) haben sich aber unabhängig voneinander differenziert, wobei sich die RED-Domäne fast bis zur Unkenntlichkeit von der Ursprungssequenz entfernte.

Die Beobachtung, dass die Paired-Box von der DNA-Bindungsdomäne einer Proto-Pax-Transposase abstammt, ermöglichte es, Transposase-Sequenzen als Outgroup zu verwenden, um die Wurzel des Protein-Stammbaums festzulegen. Dabei war es nicht notwendig, genau zu wissen, welche Transposase der Proto-Pax-Transposase am ähnlichsten ist. Es genügt die Tatsache, dass die Pax-Proteine monophyletisch sind und dass ihr nächster gemeinsamer Verwandter eine Transposase ist. Alternativ wäre z. B. auch möglich, dass einige der Proto-Pax-Transposasen sekundär aus Pax-Proteinen entstanden sind. Dann wären die Pax-Proteine paraphyletisch und die Transposasen könnten nicht mehr als ganzes als Outgroup dienen.

Für die phylogenetische Analyse mussten nun Proteinsequenzen von Pax-Proteinen und Pax-ähnlichen Transposasen gesammelt werden. Eine PSI-Blast-Suche fand mehr als 250 Datenbank-Einträge von Pax-Proteinen, die etwa 100 verschiedenen Proteinen entsprechen. Um die Analyse zu erleichtern, wurde dieser Datensatz mit dem JPred-Server weiter reduziert, so dass er keine Sequenzpaare mit mehr als 75% Identität enthielt (Redundanz-Filterung). Für die so erhaltenen 55 Sequenzen wurde ein vorläufiger Stammbaum berechnet, der zeigte, dass der reduzierte Datensatz eine repräsentative Auswahl der Pax-Proteine enthält: Herausgefiltert wurden ausschließlich Splice-Varianten und Mutanten, sowie Sequenzen aus nahe verwandten Organismen. Dabei ist offensichtlich keine phylogenetische Information verlorengegangen, denn der erhaltene Stammbaum entspricht weitgehend dem von Miller et al. (2000) und Gröger et al. (2000), die beide ebenfalls Stammbäume für die gesamte Pax-Familie präsentieren. Die Filterung hat die Datenmenge auf etwa ein Fünftel reduziert, die Berechnungen lassen sich somit mehr als 20mal schneller durchführen, ohne dass die Qualität des Resultats leidet. Für die Pax-ähnlichen Transposasen wurde ebenfalls ein solcher Redundanz-gefilterter Datensatz zusammenge-

**Sekundärstruktur-
Vorhersage**

**Transposasen als
Outgroup**

**Redundanz-
Filterung reduziert
die Rechenzeit**

3 Pilotstudien

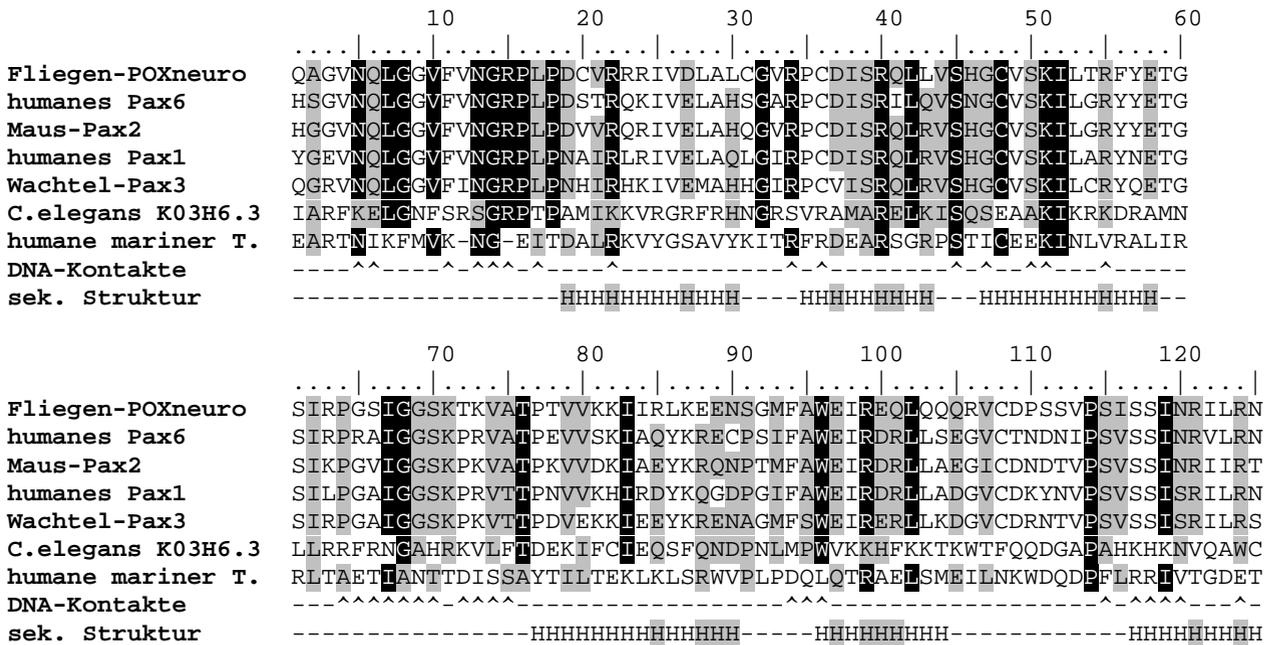


Abbildung 3.3: Proteinalignment von Pax-Proteinen und Transposasen. Je ein Mitglied aus jeder Untergruppe der Pax-Proteine wird gezeigt, außerdem je ein Mitglied der Tc1- und Mariner-Transposasen. Identische konservierte Aminosäuren sind schwarz hinterlegt, ähnliche Aminosäuren sind grau schattiert. Unter dem Alignment sind die DNA-bindenden Reste des humanen Pax6, sowie seine Sekundärstrukturelemente markiert, wie sie sich aus der Röntgen-Kristallstruktur ablesen lassen.

stellt, der Mitglieder der Tc1- und der Mariner-Familien enthielt (Plasterk et al., 1999). Alignments für beide Datensätze wurden berechnet und zu einem Gesamtalignment zusammengefasst. Positionen, die in mehr als 50% der Proteine Lücken enthielten, wurden entfernt (Abb. 3.3, S. 28). Das verbleibende Alignment hatte eine Länge von 125 Aminosäuren.

Aufteilung der Pax-Familie

Die phylogenetische Analyse des Komplet-Alignments mit Parsimony- und Neighbour joining-Algorithmen teilte die Paired-Box-Familie zuverlässig in fünf große Gruppen ein, die den klassischen Unterfamilien (Pax1/9/*meso*, PaxD/3/7/*paired/gooseberry*, Pax4/6/*eyeless*, PaxB/2/5/8/*sparkling* und PaxA/*neuro*) entsprechen (Abb. 3.4 und 3.5, S. 29 und 30.). Die interne Topologie dieser Untergruppen entspricht weitgehend den auf anderen Wegen ermittelten Verwandtschaftsbeziehungen der Organismen. Eine Ausnahme ist die Pax4/6/*eyeless*-Unterfamilie. Diese Unterfamilie ist so hoch konserviert (die Pax6-Paired-Box von Mensch und Maus ist identisch), dass kaum phylogenetisch informative Positionen vorhanden sind, so dass die festgestellten Verwandtschaftsbeziehungen in diesem Teil des Stammbaums nicht ganz zuverlässig sind. Die Position von *eyegone* aus *Drosophila* ist ebenfalls unzuver-

3.1 Pilotstudie I: Paired-Box-Proteine

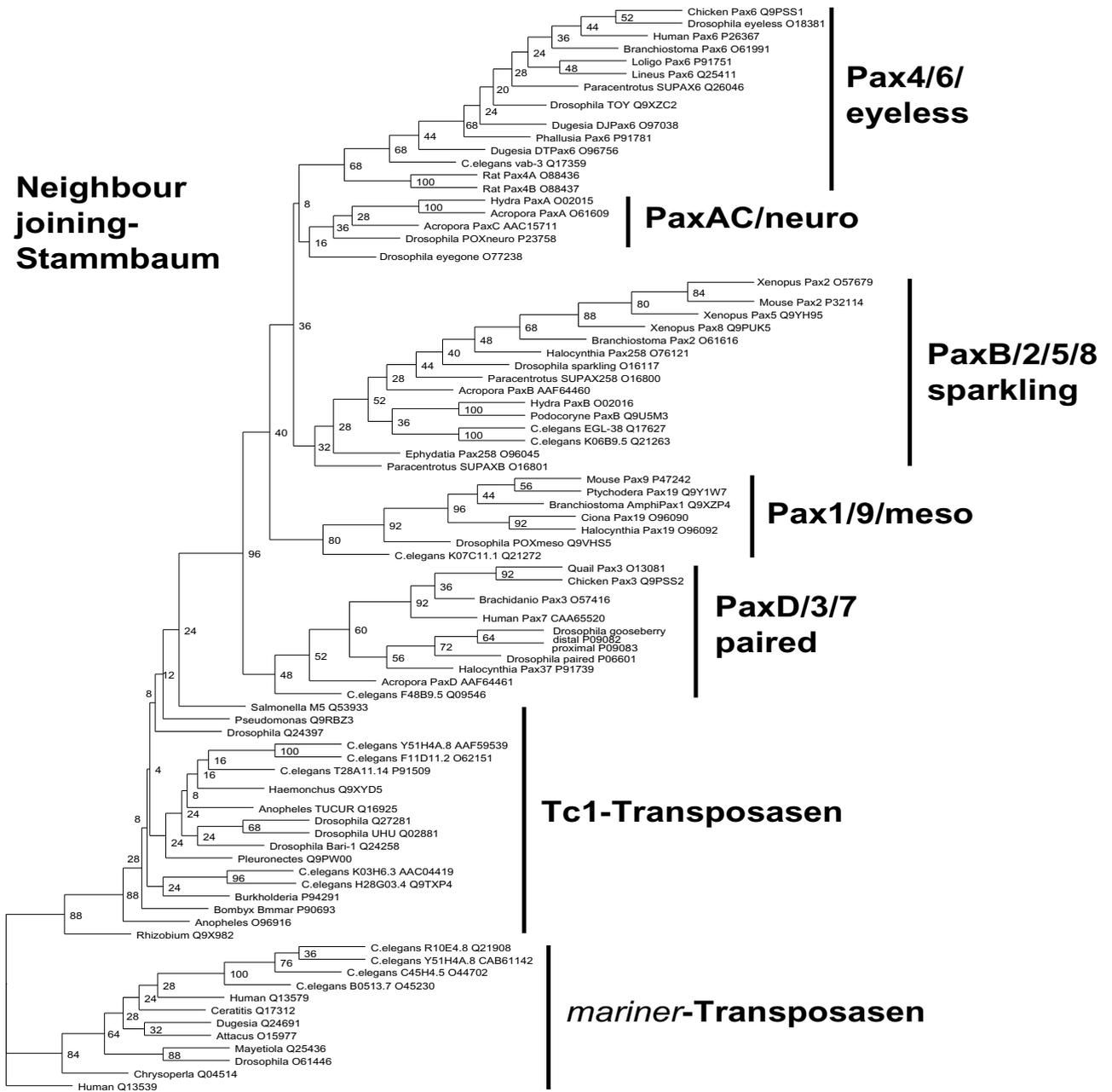


Abbildung 3.4: Neighbour-joining-Stammbaum der Pax-Domänen und der DNA-Bindungsdomänen aus Transposasen. Die Länge der Äste und die Ziffern geben den *bootstrap support* für 25 Pseudoreplikate an, Werte über 50 kennzeichnen gut abgesicherte Dichotomien. Die großen monophyletischen Gruppen sind beschriftet.

3 Pilotstudien

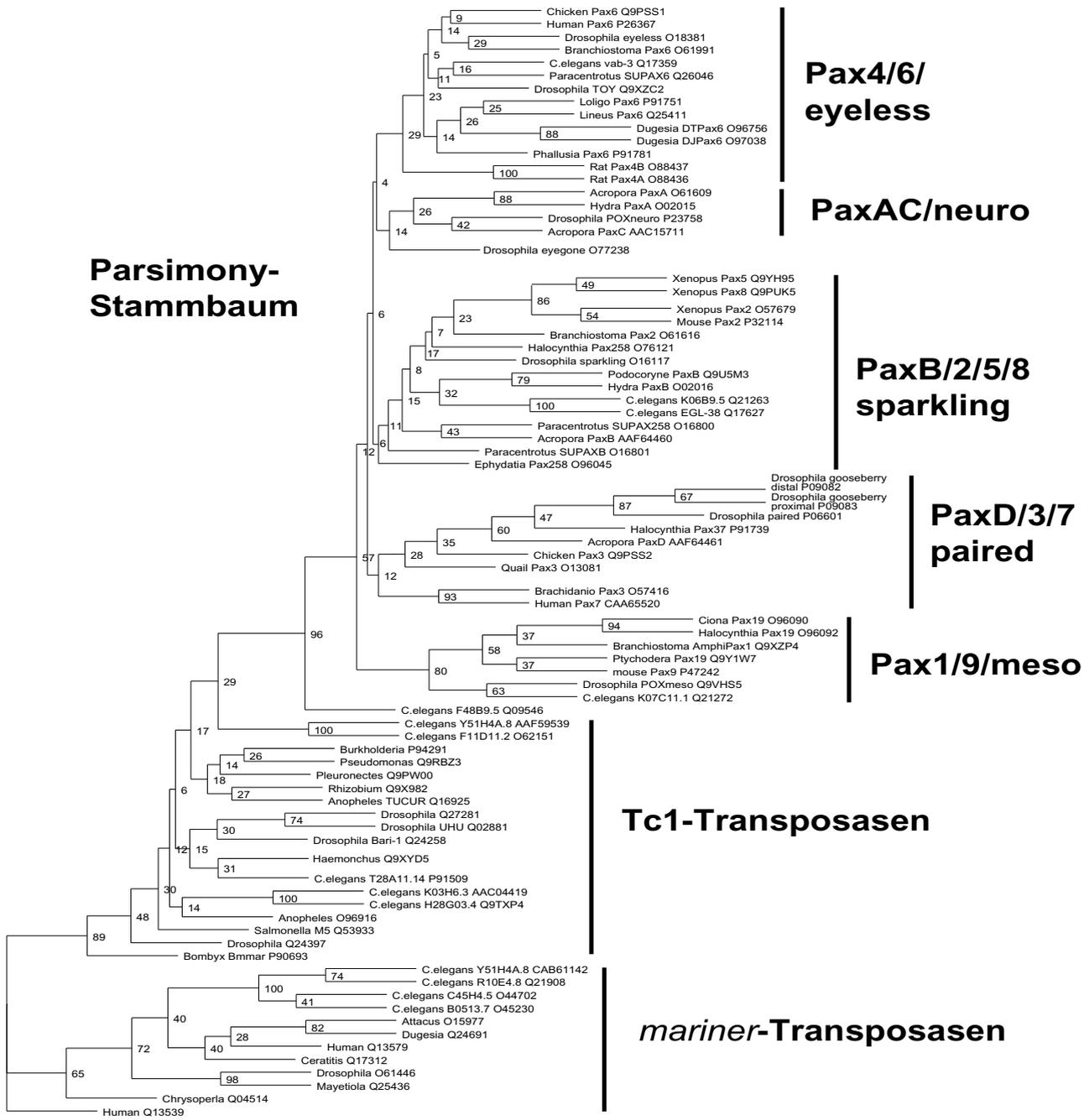


Abbildung 3.5: Parsimony-Stammbaum der Pax-Domänen und der DNA-Bindungsdomänen aus Transposasen. Die Länge der Äste und die Ziffern geben den *bootstrap support* für 100 Pseudoreplikate an, Werte über 50 kennzeichnen gut abgesicherte Dichotomien. Die großen monophyletischen Gruppen sind beschriftet. Die Topologie dieser großen Gruppen entspricht der Topologie im Neighbour-Joining-Stammbaum, nur die Positionen der PaxD/3/7/*paired* und der Pax1/9/*meso*-Gruppe sind vertauscht.

3.1 Pilotstudie I: Paired-Box-Proteine

lässig, weil dieses Protein nur einen Teil der Paired-Box enthält (Jun et al., 1998).

In beiden Stammbäumen ist PaxC Bestandteil der PaxA/*neuro*-Gruppe, obwohl PaxC eine Homöobox enthält, die anderen nicht. Mit der Maximum-likelihood-Methode werden die Unterfamilien der Pax-Proteine zwar auch erkannt, die innere Topologie ist aber nicht aufgeklärt, und auch das Verhältnis zu den Transposasen ließ sich bei dieser Methode nicht feststellen.

Parsimony- und Neighbour joining-Stammbaum stimmen darin überein, dass die Pax-Familie ein monophyletischer Bestandteil der Tc1-Transposase-Familie ist. Es ist zwar nicht klar, welche der untersuchten Transposasen der nächste Verwandte der Pax-Proteine ist, aber die generelle Verwandtschaft ist eindeutig. Die vermuteten Proto-Pax-Transposasen aus *C. elegans* und *Bombyx mori* tauchen in den Stammbäumen nicht als Schwestergruppe der Pax-Proteine auf, der signifikant höhere Blast-Score ist möglicherweise auf zufällige Ähnlichkeiten zurückzuführen, die beim gleichzeitigen Vergleich vieler Sequenzen weniger Einfluss haben. Ein anderer Grund könnte sein, dass die phylogenetischen Zusammenhänge innerhalb der Transposase-Familie allgemein nur sehr schlecht aufgeklärt sind. Die genauen Verwandtschaftsbeziehungen einzelner Transposasen ist nicht mit Sicherheit festzustellen, wie auch Plasterk et al. (1999) berichten.

Bezüglich der Position der Stammbaum-Wurzel liefern die beiden Algorithmen verschiedene Ergebnisse: Entweder trennt sich zuerst die PaxD/3/7/*gooseberry/paired* von den anderen Pax-Proteinen und dann die Pax1/9/*meso*-Familie, oder umgekehrt. Um diesen Sachverhalt zu klären, wurde das Alignment einer *four-cluster likelihood*-Analyse nach Strimmer & von Haeseler (1997) unterzogen (Abb. 3.6, S. 32). Ein großer Teil der Vektoren (18.3%) befindet sich im zentralen, nicht-aufgelösten Bereich des Diagramms, entsprechend der teilweise unklaren Topologie des Stammbaums. Eine signifikante Mehrheit der Datenpunkte (37.2%) spricht aber trotzdem für eine einzige, vollständig geklärte Topologie. Diese Topologie besagt allerdings, dass weder Neighbour-joining- noch Parsimony-Stammbaum die frühe Aufspaltung der Pax-Proteine korrekt wiedergeben, sondern dass vielmehr eine basale Dichotomie der Pax-Familie zwischen (PaxD/3/7/*gooseberry/paired* plus Pax1/9/*meso*) und den restlichen Pax-Proteinen besteht.

Dieses Ergebnis widerspricht dem Szenario von Galliot & Miller (2000) und Miller et al. (2000), das auf der Annahme beruht, dass PaxA dem Vorläufer aller Pax-Proteine am nächsten steht. Die hier verwendete Methode verzichtet auf solche *a priori*-Annahmen und verwendet stattdessen den Vergleich mit den Transposase-Sequenzen, um den zeitlichen Ablauf der Pax-Evolution zu bestimmen. Es ergibt sich ein Szenario (Abb. 3.1B, S. 24), in dem nur ein einziges Fusionsereignis stattgefunden hat – anstelle von zwei „*homeobox capturing events*“ bei Galliot & Miller (2000). Das zuerst in einem sehr frühen Metazoon entstandene Pax-Protein bestand aus der DNA-Bindungsdomäne einer Transposase und einer Homöodomäne. Dieses Ur-Pax wurde bereits sehr

Position von PaxC

Stellung der Pax-Proteine innerhalb der Transposasen

Frühe Aufspaltung der Pax-Familie

Evolutions-Szenario für die Pax-Familie

3 Pilotstudien

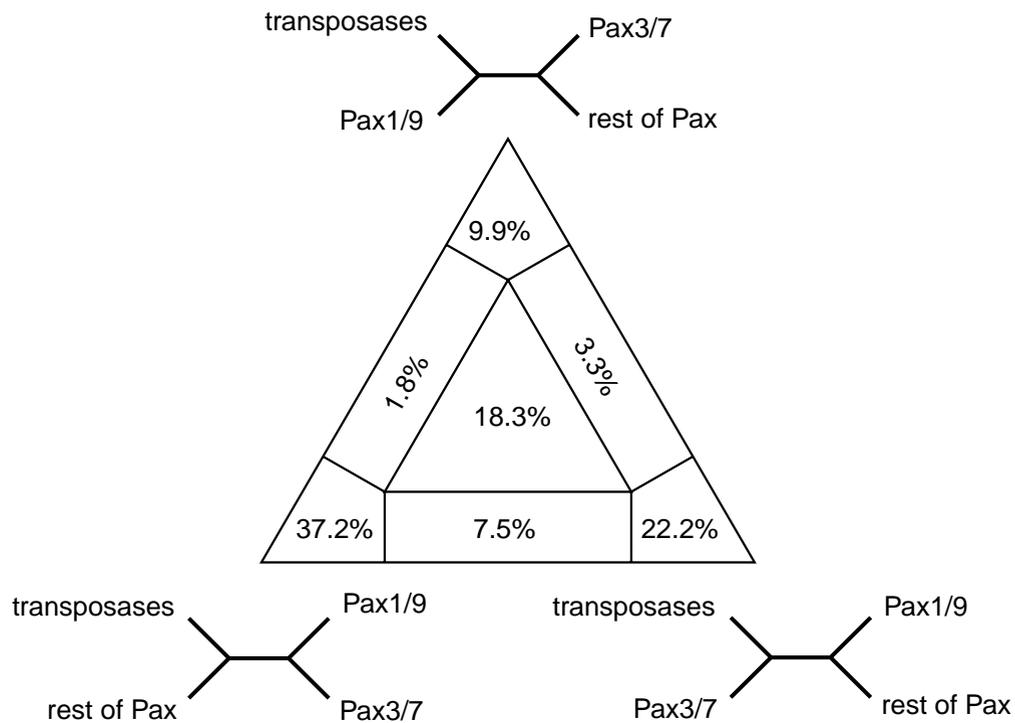


Abbildung 3.6: *Four-cluster likelihood mapping* von Pax-Proteinen und Transposasen, durchgeführt nach Strimmer & von Haeseler (1997). Die Besetzung der sieben Attraktor-Bereiche ist in Prozent angegeben. An der Ecken des Dreiecks sind die drei alternativen Stammbaum-Topologien gezeigt. Die Topologie oben entspricht dem Parsimony-Stammbaum, die rechts unten dem Neighbour-Joining-Stammbaum. Gegenüber diesen Alternativen ist die Topologie in der linken unteren Ecke aber mit 37.2% deutlich bevorzugt und wird auch von unabhängigen Daten (Homöodomänensequenz) unterstützt (siehe z. B. Galliot & Miller, 2000; Hoshiyama et al., 1998).

früh, nämlich vor der Divergenz der Schwämme (Porifera) dupliziert. Das bedeutet, dass bereits diese wenig differenzierten Organismen mindestens zwei Pax-Proteine besaßen, obwohl ihnen Nervenzellen und komplexe Sinnesorgane fehlen, deren Entwicklung bei höheren Tieren von Pax-Proteinen reguliert wird. Diese frühe Vielfalt der Pax-Familie ist unerwartet und unerklärt. Sie muss aber bei der Interpretation der Expressionsmuster von Pax-Proteinen in primitiven Organismen (z. B. Gröger et al., 2000) und bei Spekulationen über die ursprüngliche Funktion der Pax-Familie berücksichtigt werden. Die einmalige Entstehung der Pax-Familie durch ein Fusionsereignis ganz am Beginn der Metazoen-Evolution ist sicher einer der Meilensteine der Stammesgeschichte.

Schlussfolgerungen

Folgende Punkte sollten bei der anschließenden Analyse der HSD-Familie berücksichtigt werden:

1. Die Analyse von großen Datenmengen kann erheblich erleichtert wer-

3.1 Pilotstudie I: Paired-Box-Proteine

den, wenn die Zahl der betrachteten Sequenzen durch Redundanz-Filterung verringert wird. Die Analyse der Pax-Familie hat gezeigt, dass dabei die Qualität der phylogenetischen Rekonstruktion erhalten bleibt und keine Informationsverluste auftreten.

2. Die Aufklärung der Entstehung der Pax-Familie war erst möglich, als eine geeignete Outgroup (die Transposasen) gefunden wurde. Das sollte bei der Analyse der HSD-Familie berücksichtigt werden. Die zuverlässige Identifizierung der Transposasen als Outgroup war nur durch den Tertiärstrukturvergleich möglich.
3. Die Aufteilung in Unterfamilien ist mit der Maximum-Likelihood-Methode besonders zuverlässig. Um die Feinstruktur der einzelnen Unterfamilien und ihre Beziehungen untereinander zu klären, sind aber andere Methoden (Parsimony, Neighbour-Joining) notwendig.

3.2 Pilotstudie II: Cholin-Acetyl-Transferasen – Vorhersage von ancestraler Substratspezifität durch phylogenetische Untersuchung²

3.2.1 Problemstellung

Eine der häufigsten Aufgaben bei der phylogenetischen Analyse von Proteinfamilien ist die Vorhersage bisher unbekannter Proteinfunktionen. Besonders interessant ist die Möglichkeit, die Funktion der gemeinsamen „Vorfahren“ von Enzymen abzuleiten, die heute ganz verschiedene katalytische Aktivitäten besitzen. Als Modellsystem wurden die humanen Cholin-Acetyltransferasen (ChAT) und ihre Verwandten ausgewählt, deren physiologischen Funktionen bereits detailliert untersucht wurden. Im Gegensatz zu den Hydroxysteroid-Dehydrogenase (HSD)-Verwandten, für die zum größten Teil nur die cDNA-Sequenzen bekannt sind, sind viele Mitglieder der ChAT-Familie auch *in vitro* enzymatisch charakterisiert worden. Die Familie ist außerdem viel kleiner als die HSD-Familie, so dass es leichter war, an diesem Beispiel einige der Konzepte zu erarbeiten, die für die Analyse der größeren Proteinfamilie nötig waren. Es sollte versucht werden, die Evolutionsgeschichte der ChAT-Verwandten herauszuarbeiten, die Ergebnisse mit den Resultaten der biochemischen Analyse zu vergleichen, die Aktivität von evolutionären Zwischenstufen herzuleiten und davon ausgehend die enzymatischen Funktionen der heutigen Enzyme einzuordnen.

3.2.2 Hintergrund

Hintergrund für das Interesse an der Evolution und Substratspezifität der humanen Cholin-Acetyltransferasen (ChAT) und ihrer Verwandten ist der kürzlich erfolgte Nachweis, dass Granulosa-Luteinzellen des humanen Ovars m-Cholinozeptoren (Muscarinrezeptoren) exprimieren (Fritz et al., 1999). Muscarin-Agonisten führen zur schnellen Proliferation humaner Granulosa-Zellkulturen (Fritz et al., 1999). Mit immunhistochemischen Methoden konnte gezeigt werden, dass das Ovar keine ChAT-positiven Nervenzellen enthält; die Produktion des Liganden für die ACh-Rezeptor muss also lokal erfolgen. Eine solche ACh-Produktion außerhalb des Nervensystems wurde kürzlich in verschiedenen Geweben nachgewiesen, u. a. in Epithelien, Muskeln, T-Lympho-

² Teilweise veröffentlicht in: Fritz S, Wessler I, Breitling R, Rossmannith W, Ojeda SR, Dissen GA, Amsterdam A & Mayerhofer A (2001, im Druck): Expression of muscarinic receptor types in the primate ovary and evidence for non-neuronal acetylcholine synthesis. J. Clin. Endocrinol. Metabol.

3.2 Pilotstudie II: Cholin-Acetyl-Transferasen

zyten und der Placenta (Wessler et al., 1998; Kawashima et al., 2000; Wessler et al., 1999). In diesen Geweben kann ACh über die ebenfalls nachgewiesenen ACh-Rezeptoren parakrine Effekte auslösen, z. B. auf Proliferation, Differenzierung und Immunfunktionen (Lauder & Schambra, 1999; Grando et al., 1993; Kawashima et al., 1998).

Durch HPLC-Analyse wurde gezeigt, dass die Granulosa-Luteinzellen tatsächlich signifikante ACh-Konzentrationen enthalten (etwa $10 \text{ pmol}/10^6$ Zellen). Diese Zellen zeigen außerdem mit anti-ChAT-Antikörper ein deutliches Signal in der immunhistochemischen Analyse. Allerdings konnte mit RT-PCR weder in den Granulosa-Zellen noch im humanen Ovar eine Expression von ChAT nachgewiesen werden. Stattdessen wurden große Mengen des Transkripts des homologen Enzyms Carnitin-Acetyltransferase (CRAT) gefunden. Dieses Enzym ist *in vitro* ebenfalls in der Lage ACh zu produzieren. ChAT und CRAT haben zwar nur 40% identische Aminosäuren, sind also nicht hoch konserviert, die Ähnlichkeit ist aber groß genug, um evtl. zu einer Kreuzreaktion des ChAT-Antikörpers zu führen. Das könnte den Widerspruch zwischen RT-PCR und immunhistochemischem Befund erklären (Fritz et al., 2001). Die nachfolgend beschriebene phylogenetische Analyse sollte vor allem Aufschluss darüber geben, ob ovarielle CRAT tatsächlich die Funktion einer ChAT übernehmen kann und folglich für die lokale ACh-Synthese zuständig ist.

ACh-Synthese im Ovar

3.2.3 Ergebnisse und Diskussion

Um einen ersten Überblick über die Verwandtschaft der ChAT- und CRAT-Proteine zu erhalten, wurden durch eine PSI-BLAST-Suche und anschließendes automatisiertes Entfernen redundanter Sequenzen 33 homologe Proteinsequenzen zusammengetragen. Dieser Datensatz diente als Grundlage für eine Maximum-Likelihood (ML)-Analyse zur Bestimmung der Verwandtschaftsverhältnisse. Wie bei der Untersuchung der Paired-Box-Familie gezeigt wurde, ist die ML-Methode im Gegensatz zu den Parsimony- und Neighbour-Joining-Verfahren besonders geeignet, größere Proteinklassen zuverlässig gegeneinander abzugrenzen. Dagegen reicht ihr Auflösungsvermögen oft nicht aus, um den genauen Evolutionsablauf innerhalb der Untergruppen zu bestimmen. Für die erste Betrachtung bot sich dieses Verfahren also an. Das Ergebnis zeigt, dass die Verwandtschaft der ChAT/CRAT-Proteine in fünf große Gruppen zerfällt, von denen vier auch in Säugetieren vorkommen (Abb. 3.7, S. 36). Drei Gruppen bestehen aus Carnitin-Palmitoyltransferasen, eine Gruppe enthält die Carnitin-Octanoyltransferasen, und die fünfte Gruppe wird von den ChAT- und CRAT-Enzymen gebildet, die gegen die anderen Gruppen klar abgegrenzt sind und ihrerseits in drei distinkte Untergruppen aufgeteilt werden können, von denen eine nur in Wirbellosen gefunden wurde.

Aufteilung der ChAT/CRAT-Familie in Untergruppen

Um die genauen Verwandtschaftsverhältnisse innerhalb der fünften Gruppe zu bestimmen, wurden ihre Mitglieder einer weiteren phylogenetischen Analyse mit Parsimony-, Neighbour-Joining und Maximum-Likelihood-Methoden

Analyse der ChAT-Untergruppe

3 Pilotstudien

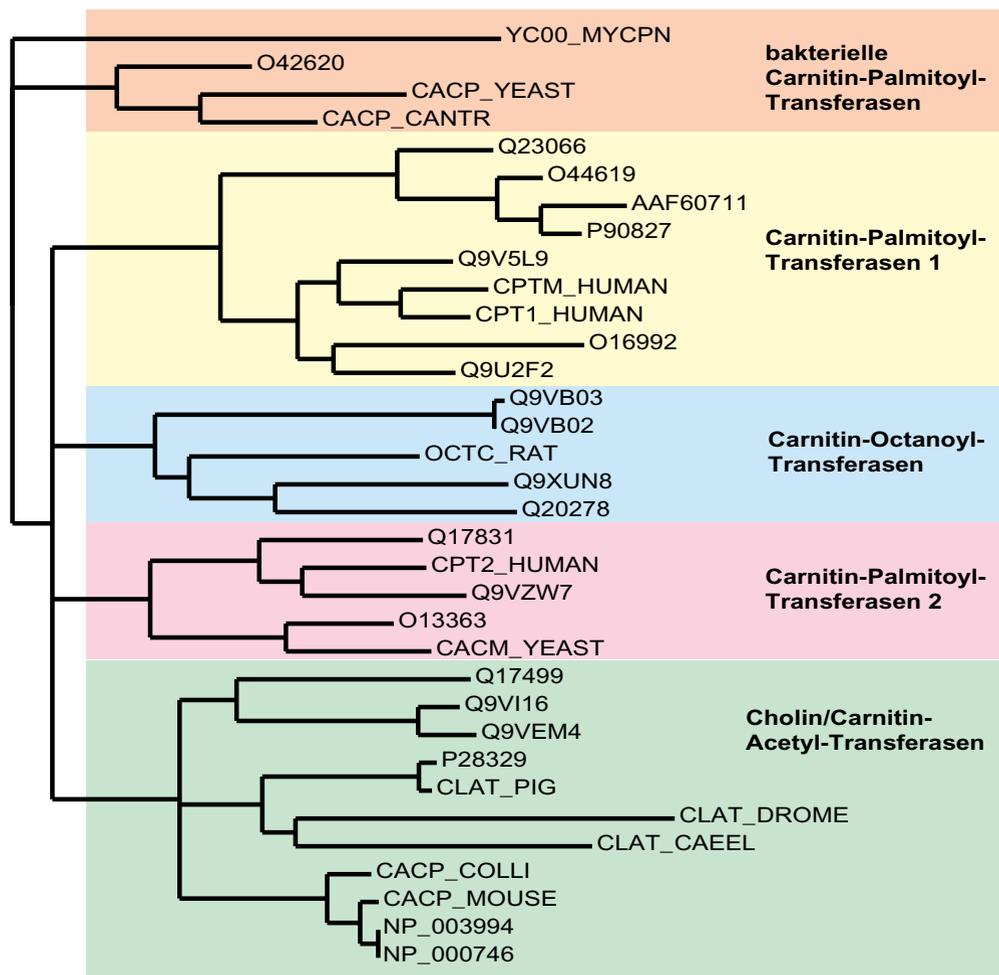


Abbildung 3.7: *Maximum likelihood*-Stammbaum der ChAT/CRAT-Familie. Die Proteine sind unter ihrer Swiss-Prot-Datenbanknummer aufgeführt.

unterzogen 3.8, S. 37). Die erhaltenen Stammbäume stimmen in allen Details überein, nur die Stellung der *Drosophila*-ChAT als Schwestergruppe entweder des *C. elegans*- oder der Vertebraten-Enzyme ist unklar. Das entspricht der allgemeinen Unsicherheit, ob die Ecdysozoa (*sensu* Aguinaldo et al., 1997) wirklich eine monophyletische Einheit bilden (Wägele et al., 1999).

Was lässt sich aus dem erhaltenen Konsensus-Stammbaum (Abb. 3.8, S. 37) für die Substratspezifitäten der hypothetischen Vorläufer-Proteine der humanen ChAT/CRAT-Proteine folgern? Der Vergleich mit den Outgroup-Enzymen zeigt, dass als alkoholisches Substrat des Stammenzym vermutlich Carnitin gedient hat. Die Verwendung von Acetat als Carboxyl-Substrat ist dagegen eine neue Entwicklung innerhalb der ChAT/CRAT-Gruppe, gewissermaßen ihre Synapomorphie (*sensu* Hennig, 1982). Das Stammenzym war also eine CRAT. Dieses Enzym wurde dann schon im gemeinsamen Vorläufer von Protosto-

Herleitung ancestraler Substrat- Spezifitäten

3.2 Pilotstudie II: Cholin-Acetyl-Transferasen

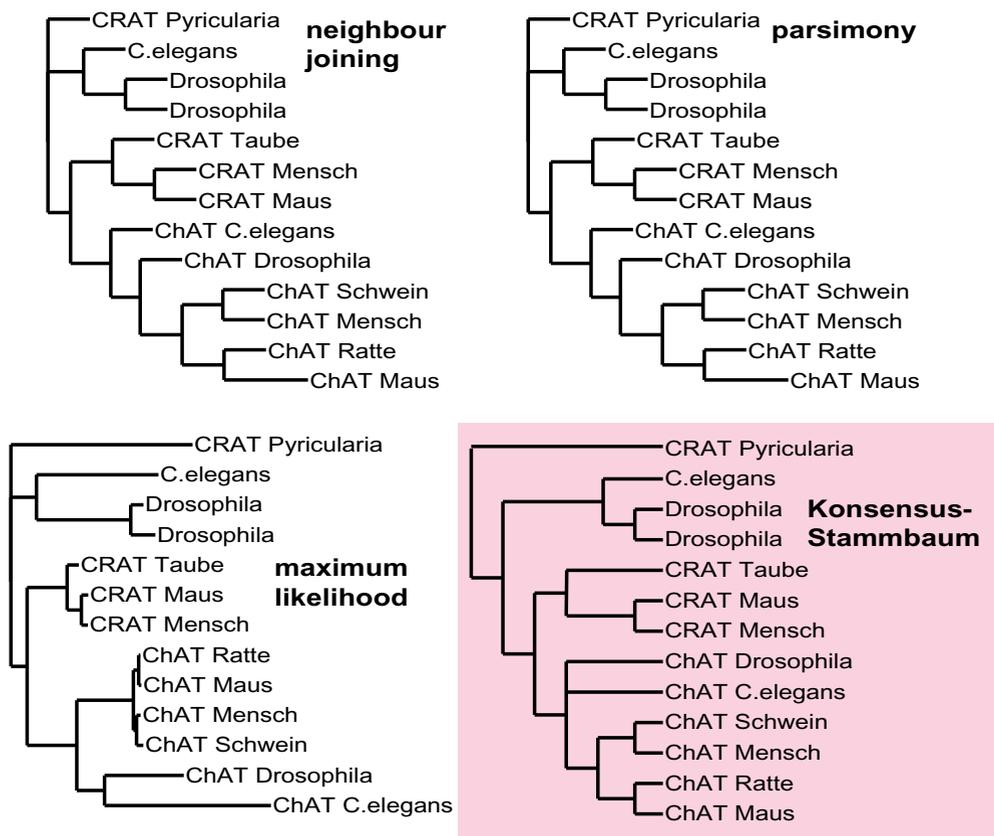


Abbildung 3.8: Stammbäume der ChAT/CRAT-Unterfamilie, nach verschiedenen Verfahren aus dem Protein-Sequenzalignment berechnet. Außerdem ist der Konsensus der drei Stammbäume angegeben. CRAT = Carnitin-Acetyl-Transferase. ChAT = Cholin-Acetyl-Transferase. Proteine ohne bekannte Substratspezifität sind nur unter dem Namen des Herkunftsorganismus angeführt. Die Carnitin-Acetyl-Transferase des Pilzes *Pyricularia (Magnaporthe) grisea* dient als Outgroup.

men und Deuterostomen dupliziert und die eine Kopie kurz darauf erneut verdoppelt – ebenfalls vor der Aufspaltung von Protostomen und Deuterostomen. Von den so entstandenen Untergruppen ging jeweils eine bei Vertebraten und Ecdysozoa verloren (sofern die endgültige Fertigstellung der Genomsequenzen nicht noch unerwartete neue Mitglieder der Proteinfamilie zutage bringt). Wenn man nun annimmt, dass sowohl Vertebraten wie auch Ecdysozoa beide Enzymfunktionen (ChAT und CRAT) besitzen, dann ist die Funktion der Ecdysozoa-spezifischen Enzymgruppe vermutlich die einer CRAT. Für den gemeinsamen Vorläufer der Vertebraten-CRATs und der ChAT-Enzyme ist daher ebenfalls die Verwendung eines Carnitin-Substrats anzunehmen, während die Spezialisierung auf die Acetylierung von Cholin erst mit der erneuten Aufspaltung der Proteinfamilie eintrat. Dieses Szenario erfordert die geringste Zahl

3 Pilotstudien

von Evolutionsereignissen, ist also die sparsamste Hypothese.

Datierung der Evolutionsereignisse

Für die Interpretation der ACh-Produktion im humanen Ovar folgt allerdings, dass die Annahme, die lokale ACh-Synthese sei auf CRAT zurückzuführen, durch die evolutionsbiologische Analyse nicht unterstützt wird. ChAT-Proteine sind zwar ursprünglich aus CRAT-Vorläufern entstanden, aber die Differenzierung der Enzymfunktionen fand bereits vor etwa 670×10^6 Jahren statt, nämlich vor der Auftrennung von Proto- und Deuterostomen (Doolittle et al., 1996). Der Verlust redundanter Enzyme in *C. elegans*, *Drosophila* und den Vertebraten bedeutet auch, dass sich die Substrat-Spezifität von ChAT- und CRAT *in vivo* vermutlich nicht überlappt. Das trotzdem eine Erweiterung der Substratspezifität von CRAT im humanen Ovar aufgetreten sein kann, wird damit natürlich nicht ausgeschlossen. Diese Frage muss aber auf anderem Wege geklärt werden.

Schlussfolgerungen

An diesem Beispiel lassen sich zwei Phänomene illustrieren, die uns auch in der HSD-Familie wieder begegnen werden:

1. Substratspezifitäten sind konservativ. Nahe verwandte Proteine haben gleiche oder vergleichbare Substratspezifitäten. Unter der Annahme, dass auch die evolutionären Vorläufer funktionelle Enzyme waren, lassen sich ancestrale Substratspezifitäten rekonstruieren, wenn die Funktionen einer ausreichenden Anzahl rezenter Proteine bekannt ist.
2. Duplizierte Enzyme entwickeln entweder neue Substratspezifitäten oder gehen verloren. Die Redundanz von Enzymfunktionen ist gering. Sowohl Vertebraten wie Invertebraten besitzen je eine Cholin-Acetyltransferase und eine Carnitin-Acetyltransferase; alle weiteren Kopien, die sich aus der phylogenetischen Rekonstruktion ableiten lassen, sind wieder verloren gegangen.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Die 17 β -HSDs sind eine kleine Gruppe von Enzymen, die an der Position 17 des Steroid-Grundgerüsts als Oxidoreduktasen aktiv werden (Tab. 4.1, S. 40). 17 β -HSDs wandeln die aktiven und inaktiven Formen der Steroidhormone (Östrogene und Androgene) ineinander um (Tab. 4.2, S. 41). Die meisten 17 β -HSDs bevorzugen *in vivo* eine Richtung der Umsetzung, wirken also entweder reduktiv oder oxidativ. Dadurch lassen sich die 17 β -HSDs in zwei Gruppen einteilen: vorwiegend aktivierende und vorwiegend inaktivierende Enzyme. Alle bekannten 17 β -HSDs verwenden Nikotinamid-Dinukleotid-Kofaktoren, je nach bevorzugter Katalyserichtung entweder NAD oder NADPH. Die 17 β -HSDs unterscheiden sich weiterhin durch ihre kinetischen Parameter, Substratspezifität, Gewebeverteilung und subzelluläre Lokalisation. Zehn Isoenzyme sind bereits genauer charakterisiert worden (Tab. 4.1, S. 40); die Existenz von weiteren Formen wird aufgrund enzymatischer Messungen erwartet (Peltoke et al., 1999b).

Neben der physiologischen Rolle im Steroidmetabolismus, werden den 17 β -HSDs auch Einflüsse auf die Entstehung und Proliferation hormonabhängiger Tumore nachgesagt (Duax et al., 2000; Labrie et al., 2000). Weil 17 β -HSDs den letzten Schritt bei der Aktivierung von Androgenen und Östrogenen katalysieren, kann eine Verschiebung des Verhältnisses von reduktiver (aktivierender) und oxidativer (inaktivierender) 17 β -HSD-Aktivität die aktiven Hormonkonzentrationen in den Zielgeweben verändern. In diesem Zusammenhang ist insbesondere der Einfluss von Östradiol auf die Entwicklung von Brustkrebs eingehend untersucht worden (Dickson & Lippman, 1987; Poutanen et al., 1995). Die Entwicklung von gewebe- und enzymespezifischen Inhibitoren der 17 β -HSDs ist also ein attraktives Ziel des rationellen Arzneimittel-Designs.

Die bisher beschriebenen 17 β -HSDs gehören mit Ausnahme der 17 β -HSD5 zu einer einzigen Proteinfamilie, den short-chain Dehydrogenasen/Reduktasen (SDR: Persson et al., 1991; Jörnvall et al., 1995). Diese Proteinfamilie ist sehr groß und komplex (Jörnvall et al., 1999). Eine PSI-Blast-Suche zeigt, dass allein das Genom von *Drosophila* über 50 Mitglieder enthält, von denen aber erst 7 enzymatisch untersucht wurden. Tabelle 4.3 (S. 42) gibt einen Überblick über die HSD-Proteine, die bereits in Modellorganismen näher untersucht wurden. Keines der in den Modellorganismen untersuch-

Die 17 β -HSD-Isoenzyme

**17 β -HSDs und Tumorbio-
logie**

Die HSD-Familie

**HSD-Proteine in
Modell-Organismen**

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Tabelle 4.1: Gewebeverteilung der 17 β -HSDs der Säugetiere. Die angegebene Gewebeverteilung ist nicht vollständig, sondern beschränkt sich auf die Organe, in denen die betreffende 17 β -HSD vermutlich eine wichtige Rolle spielt

Name	Lokalisation	Gewebeverteilung	Referenzen
HSD1	Cytosol	Ovar, Placenta, Brust	Peltoketo et al. (1988); Peltoketo et al. (1999a)
HSD2	Mikrosomen	Placenta, Leber, Gastrointestinal-Trakt	Wu et al. (1993); Mustonen et al. (1998)
HSD3	Mikrosomen	Testis	Geissler et al. (1994); Andersson et al. (1996)
HSD4	Peroxisomen	ubiquitär	Adamski et al. (1995); de Launoit & Adamski (1999)
HSD5	Cytosol	Leber, Prostata, Brust, Uterus	Qin et al. (1993); Penning et al. 2000
HSD6	Membranen	Prostata, Leber	Biswas & Russell (1997)[nicht im Menschen bekannt, siehe auch HSD9]
HSD7	Membranen	Ovar, Placenta, Brust, Leber	Krazeisen et al. (1999); Nokelainen et al. (2000); Breitling et al. (2000)
HSD8	Unbekannt	Leber, Niere, Ovar, Testes	Ando et al. (1996); Fomitcheva et al. (1998)
HSD9	Mikrosomen	Leber	Su et al. (1999)[nicht im Menschen bekannt, ist das Maus-Orthologe der Ratten-HSD6]
HSD10	Mitochondrien	Gehirn, Prostata	He et al. (1999, 2000a, 2000b)

ten HSD-Proteine wird als 17 β -HSD bezeichnet, aber einige von ihnen sind sehr nahe mit den 17 β -HSDs des Menschen verwandt: Das Fox2-Protein der Hefe ist funktionell und strukturell der nächste Verwandte der humanen 17 β -HSD4, das Erg27-Protein ist der nächste Verwandte der 17 β -HSD7. Die Bedeutung dieser Ähnlichkeiten wird unten eingehender diskutiert. Das Scully-Protein der Taufliege ist das Orthologe der 17 β -HSD10. Scully wird nicht nur im Gehirn exprimiert, sondern vor allem in Ovar und Testes, sowie in deren embryonalen Vorläufern, also genau in den Geweben, in denen auch die reduktiven 17 β -HSDs des Menschen ihre wichtigste Rolle spielen (Torroja et al., 1998). Manche Funktionen der HSD-Familie sind hochkonserviert, z. B. die der Nukleotid-Zucker-Epimerasen oder der 3-Oxoacyl-[Acyl-Carrier-Protein]-Reduktasen. Andere sind vermutlich spezifische Anpassung der einzelnen Organismen. Z. B. war die Entstehung der Alkohol-Dehydrogenase in *Drosophila* eine Voraussetzung, um vergärendes Obst als neue ökologische Nische zu

Tabelle 4.2: Funktionen der 17β -HSDs der Säugetiere. Die angegebenen Funktionen der Enzyme sind aus der Gewebeverteilung, Substratspezifität und pathophysiologischen Daten hergeleitet. Dabei ist nur die Hauptfunktion abgeben, bei der es sich nicht immer um die 17β -HSD-Aktivität handelt. Zum Beispiel scheint 17β -HSD10 bevorzugt als 3α -HSD zu wirken und einige der oxidativen Proteine mit breiter Substratspezifität (—) sind vermutlich nicht auf Steroid-Substrate beschränkt.

Name	Substratspezifität	Richtung	Funktion
HSD1	Östron	red.	Östradiol-Produktion
HSD2	—	oxid.	Inaktivierung von Steroidhormonen
HSD3	Androstendion	red.	Testosteron-Produktion
HSD4	—	oxid.	β -Oxidation von Fettsäuren / Steroid-Inaktivierung
HSD5	Östron, Testosteron u. a.	red./ox.	Elimination von Androgenen (Prostata), Erzeugung eines pro-östrogenen Zustands (Brust)
HSD6	5α -Androstan- 3α , 17β -diol	oxid.	Inaktivierung von Dihydrotestosteron
HSD7	Östron, Zymosteron?	red.	Östradiol-Produktion, Cholesterinbiosynthese
HSD8	—	oxid.	Inaktivierung von Steroidhormonen
HSD9	—	oxid.	Inaktivierung von Steroidhormonen
HSD10	5α -Androstan- 3α , 17β -diol	red.	Produktion von Dihydrotestosteron

erschließen (Ashburner, 1998). Auch die 17β -HSDs der Säugetiere sind wahrscheinlich aus Vorläufern entstanden, die andere Substrate hatten. Wie im Folgenden gezeigt wird, kann ein besseres Verständnis der Verwandtschaftsbeziehungen innerhalb der HSD-Familie helfen, die heutige Funktion der 17β -HSDs besser zu verstehen.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Tabelle 4.3: Charakterisierte HSD-Proteine in komplett sequenzierten Genomen

<i>Escherichia coli</i> (insgesamt sind 21 von 4289 Proteinen HSD-Mitglieder)	
AAC75881	2-Deoxy-D-gluconat-5-Reduktase
AAC74691	NAD-abhängige 7 α -Hydroxysteroiddehydrogenase
AAC74177	3-Oxoacyl-[Acyl-Carrier-Protein]-Reduktase
AAC73697	2,3-Dihydro-2,3-dihydroxybenzoat-Dehydrogenase
AAC75747	Glucitol (Sorbitol)-6-phosphat-Dehydrogenase
AAC74370	Enoyl-[Acyl-Carrier-Protein]-Reduktase
AAC76793	dTDP-Glucose-4,6-Dehydratase
AAC75102	—“—
AAC75594	2,3-Dihydroxy-2,3-dihydrophenylpropionat-Dehydrogenase
AAC73846	UDP-Galactose-4-Epimerase
<i>Saccharomyces cerevisiae</i> (22 von 5885 Proteinen)	
NP_012934	Multifunktionelles Protein der peroxisomalen β -Oxidation (Fox2p)
NP_012142	1-Acyldihydroxyacetonphosphat-Reduktase (Ayr1p)
NP_014197	Peroxisomale 2,4-Dienoyl-CoA-Reduktase (Sps19p)
NP_012868	3-Oxoacyl-[Acyl-Carrier-Protein]-Reduktase (Oar1p)
NP_009824	3-Ketosphinganin-Reduktase (Tsc10p)
NP_013201	3-Ketosterol-Reduktase (Erg27p)
NP_011514	C-3-Sterol-Reduktase (Erg26p)
NP_009575	UDP-Glucose-4-Epimerase (Gal10p)
<i>Drosophila melanogaster</i> (55 von etwa 13500 Proteinen)	
AAF55546	Photorezeptor-Dehydrogenase (Pdh Genprodukt)
AAF48012	Antennen-spezifische Dehydrogenase (antdh Genprodukt)
AAF55877	Fettkörper-Protein (Fbp2 Genprodukt)
AAF53404	Alkohol-Dehydrogenase-related (Adhr Genprodukt)
AAF53403	Alkohol-Dehydrogenase (Adh Genprodukt)
AAF46376	Sepiapterin-Reduktase (Sptr Genprodukt)
AAF48797	3-Hydroxyacyl-CoA-Dehydrogenase (Scully [scu] Genprodukt)

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

4.1.1 Problemstellung

Die Familie der HSDs umfasst derzeit mehr als 1000 Mitglieder, von denen nur der kleinste Teil enzymatisch charakterisiert ist (Jörnvall, 1999). Für ein besseres Verständnis ihrer physiologischen Funktionen wäre es hilfreich, die komplexe Datenmenge in kleinere, homogene Gruppen zu gliedern, also Untergruppen zu definieren, wie es für die AKR-Proteine erfolgreich durchgeführt wurde (Jez et al., 1997). In den Vorstudien an der Paired-Box-Familie hatte sich gezeigt, dass der Versuch, die gesamte Sequenzmasse auf einmal zu bearbeiten, wenig Aussicht auf Erfolg haben würde. Andererseits ist die Bearbeitung kleiner Unterfamilien, wie z. B. in Baker (1996, 1998), zwar für spezielle Fragestellungen (z. B. 17β -HSD7-Analyse (Kap. 4.2)) sinnvoll, nicht aber für den Aufbau eines evolutionären Gesamtszenarios und auch die willkürliche Beschränkung auf wenige ausgewählte Proteine (z. B. Baker, 1991) ist zu diesem Zweck ungeeignet – ebenso wie die Verwendung des bei der Analyse der Paired-Box-Familie erfolgreich eingesetzten Redundanzfilters –, denn die HSD-Familie hat sich bereits sehr früh vielfach aufgespalten. Sie ist deshalb viel feiner gegliedert als die Paired-Box-Familie, die wie oben gezeigt in fünf klar abgegrenzte Unterfamilien zerfällt. Die frühe Aufspaltung in viele unabhängige Entwicklungslinien hat dazu geführt, dass heute zwar noch einige kurze Sequenzelemente (glycinreiches Kofaktorbindungsmotiv, NAG-Motiv, aktives Zentrum) hochkonserviert sind, die restliche Sequenz dagegen so mutationsgesättigt ist, dass das phylogenetische Signal in der Primärsequenz kaum noch zu erkennen ist (Homoplasieproblem). Es wurde jedoch angenommen, dass die Tertiärstruktur höher konserviert ist, so dass sich in den Ähnlichkeiten der Proteinstruktur auch die frühen Aufspaltungsprozesse noch widerspiegeln sollten. Für die erste Analyse wurde deshalb die Tatsache genutzt, dass bereits 21 Mitglieder der HSD-Familie kristallisiert wurden, von denen 19 schon in den öffentlichen Datenbanken zugänglich sind (Duax et al., 2000; und Daten aus der Protein Data Bank). Diese Strukturdaten wurden verwendet, um ein Evolutionsszenario für die bereits kristallisierten HSD-Verwandten zu entwickeln. Anschließend wurden die anderen, noch nicht kristallisierten 17β -HSDs in dieses Grundschema eingebunden. Dabei wurde der Schwerpunkt nicht auf die Vollständigkeit des Datensatzes gelegt, sondern auf die mögliche Klärung evolutionsbiologischer Fragen.

4.1.2 Ergebnisse

Mit Hilfe der FSSP-Datenbank wurden die Proteine identifiziert, deren Röntgenstrukturen eine signifikante Ähnlichkeit mit den Mitgliedern der HSD-Familie zeigen. Insgesamt zeigen 19 Proteine in den öffentlichen Datenbanken

**Suche nach
HSD-verwandten
Kristall-Strukturen**

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

eine so weitgehende Übereinstimmung mit der einzigen kristallisierten 17 β -HSD (17 β -HSD1, 1FDS, DALI-Score > 15), dass ein gemeinsamer Vorläufer angenommen werden kann. Mehrere weitere Dehydrogenasen zeigen ebenfalls lokale Strukturähnlichkeiten zur HSD-Familie, insbesondere im Bereich der Kofaktor-Bindungsstelle (Rossmann-Faltung, DALI-Score etwa 5 bis 10); diese Übereinstimmung ist aber vermutlich auf konvergente Entwicklung zurückzuführen, denn die Gesamtstruktur dieser Proteine unterscheidet sich erheblich von der der HSD-Proteine und auch in der Sequenz des aktiven Zentrums sind keine Ähnlichkeiten zu erkennen. Von einigen der identifizierten HSD-Verwandten existieren mehrere Röntgenstrukturen, z. B. von verschiedenen Mutanten oder von Enzym-Ligand-Komplexen. Für die folgenden Analysen wurde jeweils eine repräsentative Röntgenstruktur mit hoher Auflösung und Qualität ausgewählt (Tab. 4.4, S. 44). Die Koordinaten dieser Röntgenstrukturen wurden aus der Protein Data Bank bezogen.

Tabelle 4.4: Kristallstrukturen von Mitgliedern der HSD-Familie

Dateiname	Protein	Spezies
1ybv	1,3,8-Hydroxynaphthalin-Reduktase	<i>Magnaporthe</i>
1fmc	7 α -Hydroxysteroiddehydrogenase	<i>Escherichia</i>
1cyd	Carbonyl-Reduktase	<i>Mus</i>
2ae1	Tropinon-Reduktase II	<i>Datura</i>
1eny	Enoyl-Acyl-Carrier-Protein-Reduktase	<i>Mycobacterium</i>
1oaa	Sepiapterin-Reduktase	<i>Mus</i>
1fds	17 β -Hydroxysteroiddehydrogenase Typ 1	<i>Homo</i>
1b16	Alkoholdehydrogenase	<i>Drosophila</i>
1dhr	Dihydropteridin-Reduktase	<i>Rattus</i>
1xel	UDP-Galactose-4-Epimerase	<i>Escherichia</i>
1qrr	Sulfolipidbiosyntheseprotein	<i>Arabidopsis</i>
1bxk	dTDP-Glucose-4,6-Dehydratase	<i>Escherichia</i>
1db3	GDP-Mannose-4,6-Dehydratase	<i>Escherichia</i>
1gfs	GDP-Fucosesynthase	<i>Escherichia</i>
1bdb	cis-Biphenyl-2,3-dihydrodiol-2,3-Dehydrogenase	<i>Pseudomonas</i>
1hdc	3 α ,20 β -Hydroxysteroiddehydrogenase	<i>Streptomyces</i>
1ae1	Tropinon-Reduktase I	<i>Datura</i>
1enp	Enoyl-Acyl-Carrier-Protein-Reduktase	<i>Brassica</i>
1dfg	Enoyl-Acyl-Carrier-Protein-Reduktase	<i>Escherichia</i>

Paarweiser Vergleich der Strukturen

Alle 19 Strukturen wurden dann mit Hilfe des DALI-Algorithmus paarweise verglichen. Dieser Algorithmus liefert nicht nur eine optimierte Überlagerung der Strukturen, sondern gleichzeitig ein strukturbasiertes Alignment der Primärstrukturen (Abb. 4.1, S. 45). Im Gegensatz zu den üblichen Proteinalignment-Algorithmen werden auf diese Weise auch in weit divergierten Proteinen die Reste erkannt, die tatsächlich äquivalente Positionen in

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

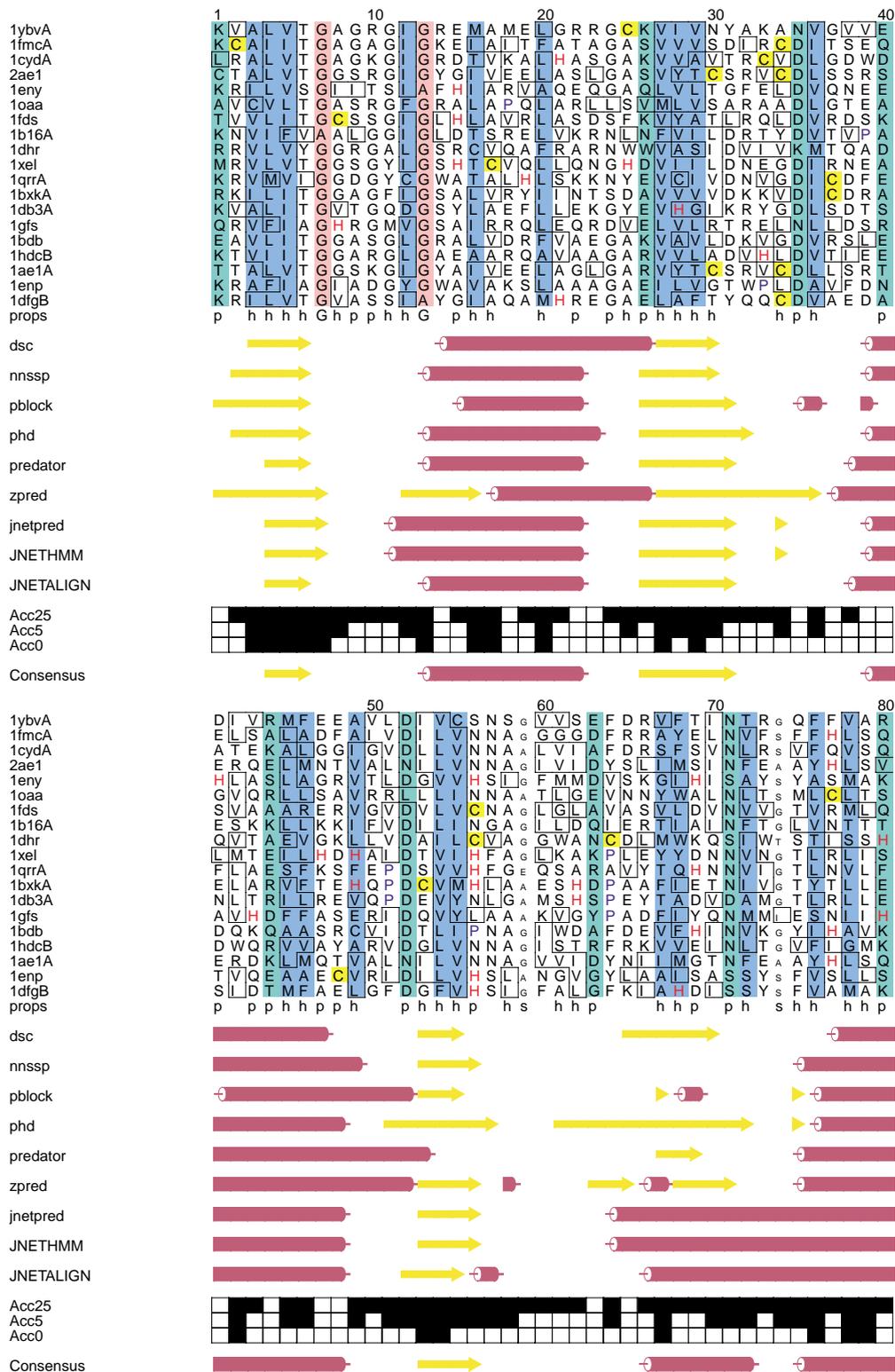


Abbildung 4.1: Alignment der Proteinsequenzen von kristallisierten HSD-Proteinen. Konservierte Reste sind farblich gekennzeichnet. Die mit dem JPred-Server vorhergesagte Sekundärstruktur ist ebenfalls angegeben (α -Helix rot, β -Faltblatt gelb). Die letzte Zeile zeigt den Konsensus der Einzelalgorithmen. Das glycinreiche Kofaktorbindungsmotiv liegt zwischen Position 6 und 13, das NAG-Motiv zwischen 57 und 59, das aktive Zentrum zwischen 103 und 107.

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

der dreidimensionalen Struktur einnehmen. Die so erhaltenen paarweisen Sequenzalignments können anschließend kombiniert werden und als zuverlässige Grundlage für das Alignment weiterer (nicht kristallisierter) Familienmitglieder dienen. Auch als Grundlage für phylogenetische Analysen eignet sich das auf diese Weise gewonnene Alignment.

Im weiteren wurden insgesamt fünf Protein-Stammbäume berechnet: Die ersten zwei basierten auf dem Vergleich der dreidimensionalen Koordinaten der Proteinstrukturen. Drei weitere wurden mit Sequenz-basierten Methoden aus dem Proteinalignments berechnet, das der Strukturvergleich geliefert hatte. Für die zwei Ähnlichkeits-Stammbäume der Kristallstrukturen wurden zwei verschiedene Ähnlichkeitsmaße verwendet, die durch den DALI-Algorithmus geliefert werden: Zum einen der Z-Score, der die Qualität der Überlagerung und die Länge der übereinstimmenden Bereiche berücksichtigt (Tab. 4.5, S. 48), und zweitens das quadratische Mittel der Differenz der Koordinaten in den überlagerten Strukturbereichen (*root mean square*, Tab. 4.6, S. 48). Für beide Werte wurde jeweils eine Distanzmatrix erstellt, die die Ergebnisse aller paarweisen Sequenzvergleiche zusammenfasste. Die Matrix für die Z-Werte wurde dann invertiert, damit Proteine mit der höchsten Ähnlichkeit die niedrigsten Werte erhielten. Die so erhaltenen Matrizen wurden dann als Eingabe für den Neighbour-Joining-Algorithmus nach Fitch & Margoliash (1967) benutzt (Abb. 4.3, S. 49f.).

**Struktur-
Stammbäume**

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Tabelle 4.5: Distanzmatrix der kristallisierten HSD-Proteine. Als Ähnlichkeitsmaß dient der vom DALI-Algorithmus berechnete Z-Score. Gezeigt ist die Matrix vor der Invertierung (s. Text). Je größer die Werte, desto ähnlicher sind die Strukturen. Die Proteine sind mit den Namen der für die Berechnung verwendeten Kristallstrukturdateien bezeichnet. Die Zuordnung zu den Enzymnamen ist in Tab. 4.4 (S. 44) angegeben.

	1b16	1xel	1ae1	1bdb	1eny	1bxk	1cyd	1enp	1db3	1dfg	1dhr	1fds	1gfs	1hdc	1oaa	1qrr	1ybv	2ae1	1fmc
1b16	47.9	19.6	22.5	22.2	21.5	19.7	23.9	19.9	17.8	20.2	19.2	21.2	16.6	24.2	20.5	16.8	23.8	24.1	23.4
1xel	19.6	63.2	17.4	18.5	17.2	35.5	18.0	16.5	30.9	15.7	16.1	20.0	31.3	20.0	17.3	34.7	18.6	18.9	18.4
1ae1	22.5	17.4	46.6	30.6	27.1	17.6	31.2	27.0	17.7	30.1	20.7	25.8	16.8	32.7	25.3	17.0	32.2	35.4	30.6
1bdb	22.2	18.5	30.6	48.2	25.2	18.1	29.8	24.5	18.4	25.2	19.9	25.1	16.2	31.3	23.6	17.5	29.6	29.2	30.0
1eny	21.5	17.2	27.1	25.2	49.9	16.7	26.9	30.0	16.3	33.6	19.3	22.6	15.2	27.2	21.2	13.3	27.8	28.3	27.4
1bxk	19.7	35.5	17.6	18.1	16.7	59.6	18.1	15.4	31.6	15.4	16.0	18.9	29.5	19.2	17.4	30.0	18.1	17.9	17.9
1cyd	23.9	18.0	31.2	29.8	26.9	18.1	46.6	26.0	18.4	25.6	21.5	26.9	17.2	31.5	27.2	17.1	31.4	32.3	32.5
1enp	19.9	16.5	27.0	24.5	30.0	15.4	26.0	52.5	16.2	31.5	18.0	23.0	15.4	26.7	19.7	15.5	26.6	28.8	26.5
1db3	17.8	30.9	17.7	18.4	16.3	31.6	18.4	16.2	57.1	16.3	16.8	18.4	29.4	19.2	14.3	25.9	18.1	17.6	18.7
1dfg	20.2	15.7	30.1	25.2	33.8	15.4	25.6	31.5	16.3	48.4	17.6	21.4	15.6	28.2	20.1	12.2	27.6	29.3	28.5
1dhr	19.2	16.1	20.7	19.9	19.3	16.0	21.5	18.0	16.8	17.6	45.7	19.6	15.5	22.0	18.8	15.6	22.1	20.4	21.2
1fds	21.2	20.0	25.8	25.1	22.6	18.9	26.9	23.0	18.4	21.4	19.6	48.5	17.0	25.9	22.5	15.8	24.8	26.3	25.5
1gfs	16.6	31.3	16.8	16.2	15.2	29.5	17.2	15.4	29.4	15.6	15.5	17.0	55.6	17.4	15.6	25.9	16.6	16.9	16.4
1hdc	24.2	20.0	32.7	32.4	27.2	19.2	31.5	26.7	19.2	28.2	22.0	25.9	17.4	48.1	24.7	18.2	32.6	29.9	32.1
1oaa	20.5	17.3	25.3	23.6	21.2	17.4	27.2	19.7	14.3	20.1	18.8	22.5	15.6	24.7	48.1	12.8	26.4	27.0	26.9
1qrr	16.8	34.7	17.0	17.5	13.3	30.0	17.1	15.5	25.9	12.2	15.6	15.8	25.9	18.2	12.8	73.6	18.2	16.5	17.6
1ybv	23.8	18.6	32.2	29.6	27.8	18.1	31.4	26.6	18.1	27.6	22.1	24.8	16.6	32.6	26.4	18.2	50.2	32.5	36.4
2ae1	24.1	18.9	35.4	29.2	28.3	17.9	32.3	28.8	17.6	29.3	20.4	26.3	16.9	29.9	27.0	16.5	30.3	48.1	30.8
1fmc	23.4	18.4	30.6	30.0	27.4	17.9	32.5	26.5	18.7	28.5	21.2	25.5	16.4	32.1	26.9	17.6	36.4	30.8	48.8

Tabelle 4.6: Distanzmatrix der kristallisierten HSD-Proteine. Als Ähnlichkeitsmaß dient die mittlere quadratische Abweichung der Koordinaten (*root mean square*). Identische Strukturen haben eine Abweichung von null.

	1b16	1xel	1ae1	1bdb	1eny	1bxk	1cyd	1enp	1db3	1dfg	1dhr	1fds	1gfs	1hdc	1oaa	1qrr	1ybv	2ae1	1fmc
1b16	0.0	2.3	2.5	2.4	2.8	3.1	2.7	3.0	2.9	2.7	3.0	2.4	2.1	2.5	2.5	3.2	2.4	2.8	2.3
1xel	2.3	0.0	3.0	2.9	2.9	2.1	2.8	3.1	2.6	3.1	2.9	2.9	2.3	2.8	2.8	2.6	2.9	2.7	2.7
1ae1	2.5	3.0	0.0	1.8	2.1	2.8	1.4	1.9	2.6	2.0	2.7	2.1	2.8	1.8	2.5	2.8	1.7	0.8	1.4
1bdb	2.4	2.9	1.8	0.0	2.6	2.8	1.9	3.1	2.8	2.4	2.8	2.3	2.5	1.6	2.5	4.9	1.6	1.8	1.7
1eny	2.8	2.9	2.1	2.6	0.0	3.4	2.3	1.8	3.0	1.7	3.2	2.8	2.7	2.2	2.9	3.2	2.3	2.2	2.1
1bxk	3.1	2.1	2.8	2.8	3.4	0.0	3.0	3.4	2.2	3.3	3.1	3.2	2.2	2.7	2.9	2.8	3.1	3.0	3.0
1cyd	2.7	2.8	1.4	1.9	2.3	3.0	0.0	2.1	2.6	2.4	2.8	2.3	2.7	2.1	2.3	2.8	2.0	1.5	1.6
1enp	3.0	3.1	1.9	3.1	1.8	3.4	2.2	0.0	3.0	1.6	2.8	2.7	2.5	2.3	3.0	4.2	2.1	2.3	2.0
1db3	2.6	2.6	2.6	2.8	3.0	2.2	2.6	3.0	0.0	3.0	2.9	3.1	2.5	2.3	2.9	2.5	2.7	2.8	2.8
1dfg	2.7	3.1	2.0	2.4	1.7	3.3	2.4	1.6	3.0	0.0	3.3	2.7	2.7	2.2	2.7	3.3	2.0	2.0	2.0
1dhr	3.0	2.9	2.7	2.8	3.2	3.1	2.8	3.0	2.9	3.3	0.0	2.7	2.9	2.5	2.9	2.9	2.5	2.9	2.7
1fds	2.4	2.9	2.1	2.3	2.8	3.2	2.3	2.7	3.1	2.7	2.7	0.0	3.1	2.7	2.6	3.4	2.9	2.1	2.5
1gfs	2.1	2.3	2.8	2.5	2.7	2.2	2.7	2.5	2.5	2.7	2.9	3.1	0.0	2.7	3.0	2.9	2.9	3.0	3.0
1hdc	2.5	2.8	1.8	1.6	2.2	2.7	2.1	2.3	2.3	2.2	2.5	2.7	2.7	0.0	2.2	2.5	1.8	1.9	1.7
1oaa	2.5	2.8	2.5	2.5	2.9	3.1	2.3	3.0	2.9	2.7	2.9	2.6	2.4	2.2	0.0	3.2	2.5	2.5	2.2
1qrr	3.2	2.6	2.8	4.9	3.2	2.8	2.8	4.2	2.5	3.3	2.9	3.4	2.9	2.5	3.2	0.0	2.9	3.0	2.8
1ybv	2.4	2.9	1.7	1.6	2.3	3.1	2.0	2.1	2.7	2.0	2.5	2.9	2.9	1.8	2.5	2.9	0.0	1.9	1.4
2ae1	2.8	2.7	0.8	1.8	2.2	3.0	1.5	2.3	2.8	2.0	2.9	2.1	3.0	1.9	2.5	3.0	1.9	0.0	1.5
1fmc	2.3	2.7	1.4	1.7	2.1	3.0	1.6	2.0	2.8	2.0	2.7	2.5	3.0	1.9	2.2	2.8	1.4	1.5	0.0

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

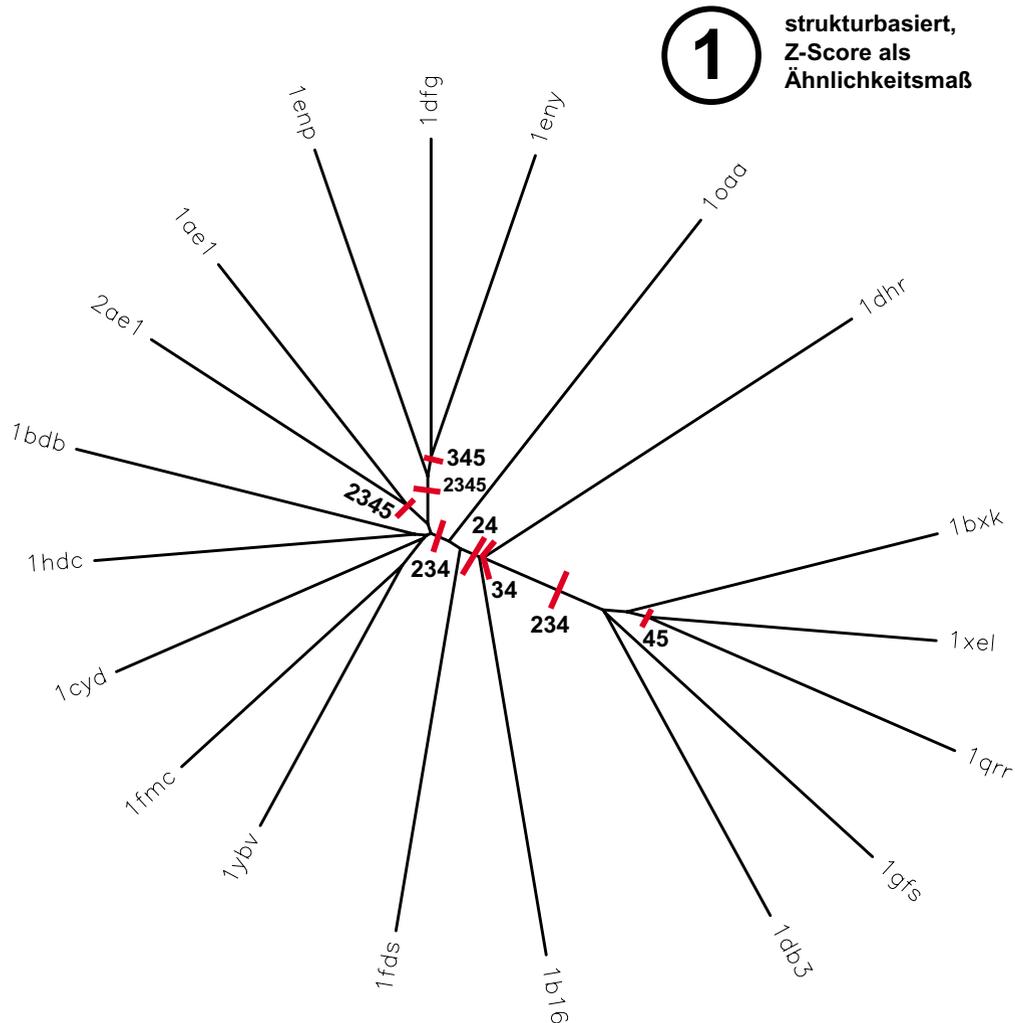


Abbildung 4.3: (bis Abbildung 4.7) Struktur-basierte Stammbäume der HSD-Familie. Die ersten zwei Bäume sind aus dem direkten Vergleich der dreidimensionalen Strukturen berechnet, die drei anderen basieren auf dem Sequenz-Alignment, das sich aus dem Strukturvergleich ergibt. Die Ziffern geben an, in welchen anderen Stammbäumen die rot markierte Dichotomie ebenfalls beobachtet wird. Die Proteine sind mit den Namen der für die Berechnung verwendeten Kristallstrukturdateien bezeichnet. Die Zuordnung zu den Enzymnamen ist in Abb. 4.9 (S. 55) angegeben. Dort ist der Konsensus-Stammbaum gezeigt, der sich aus den fünf Einzelstammbäumen ergibt.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

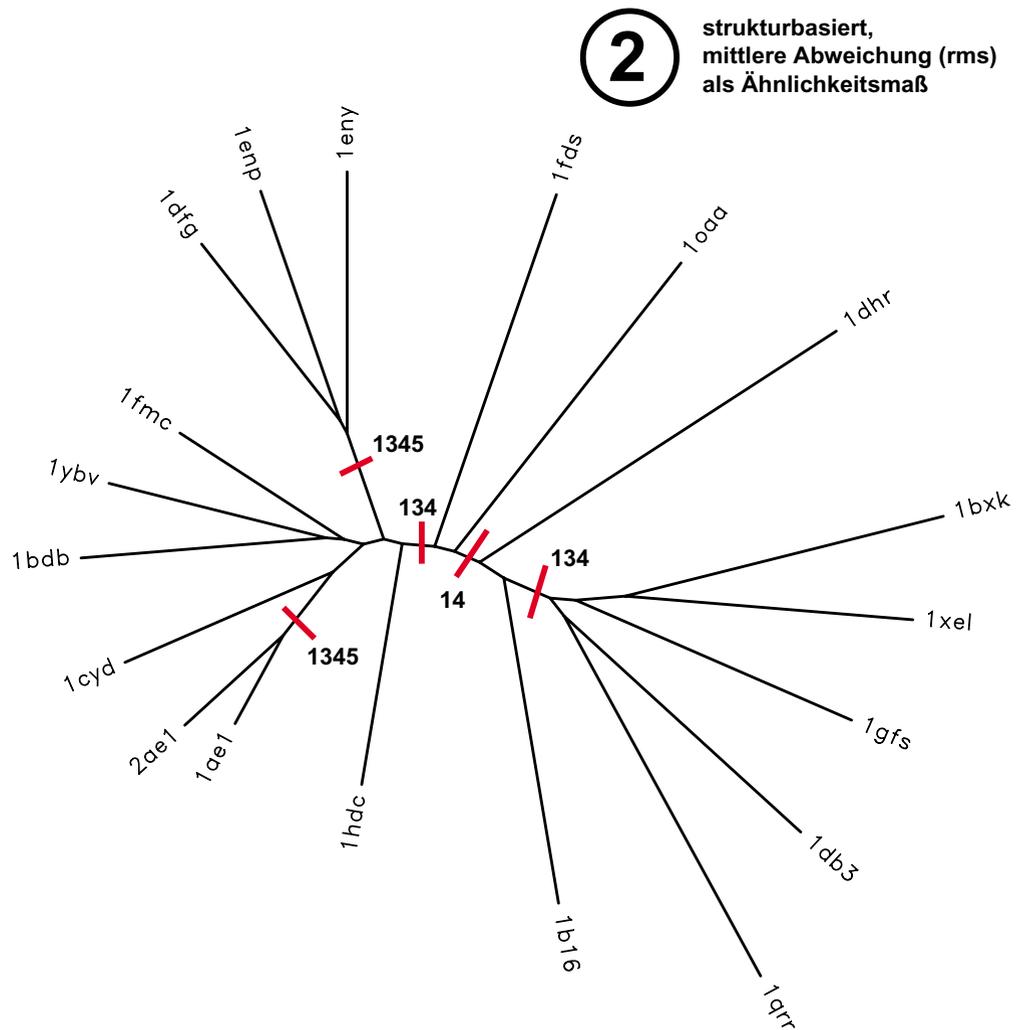


Abbildung 4.4: ... Fortsetzung von Abb. 4.3

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

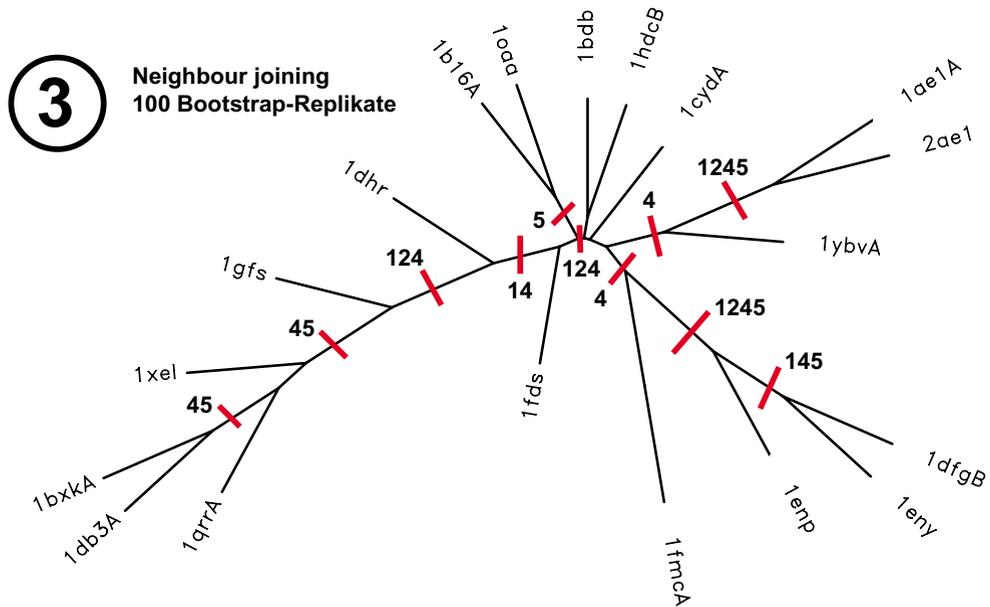


Abbildung 4.5: ... Fortsetzung von Abb. 4.3

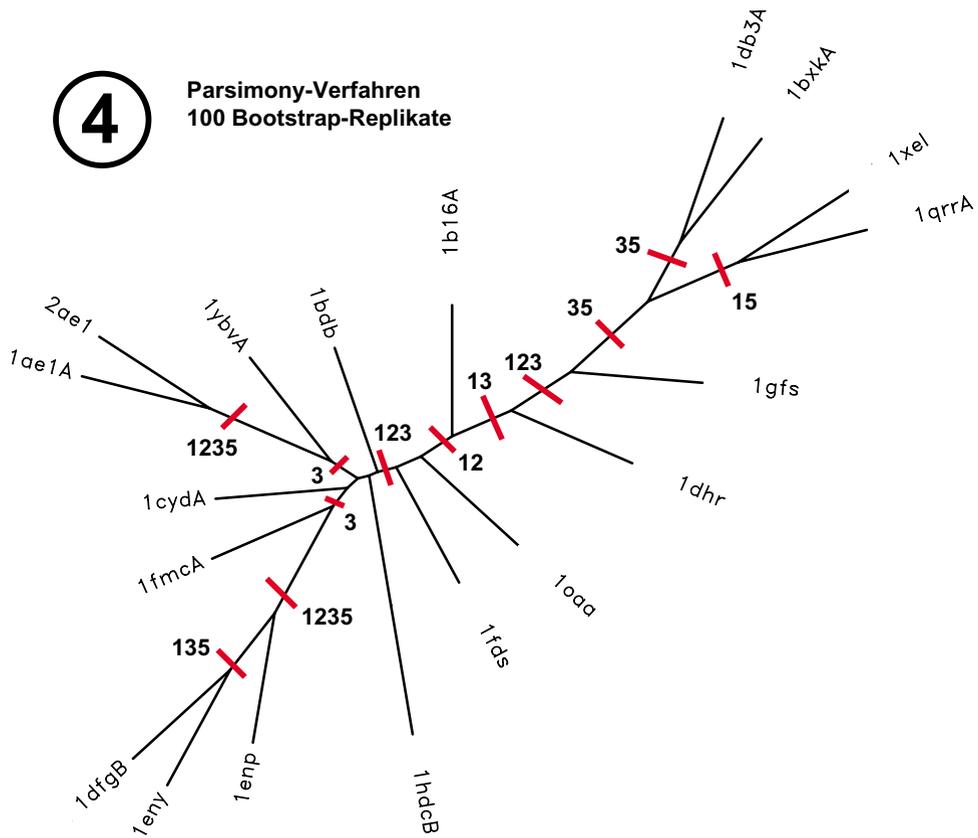


Abbildung 4.6: ... Fortsetzung von Abb. 4.3

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

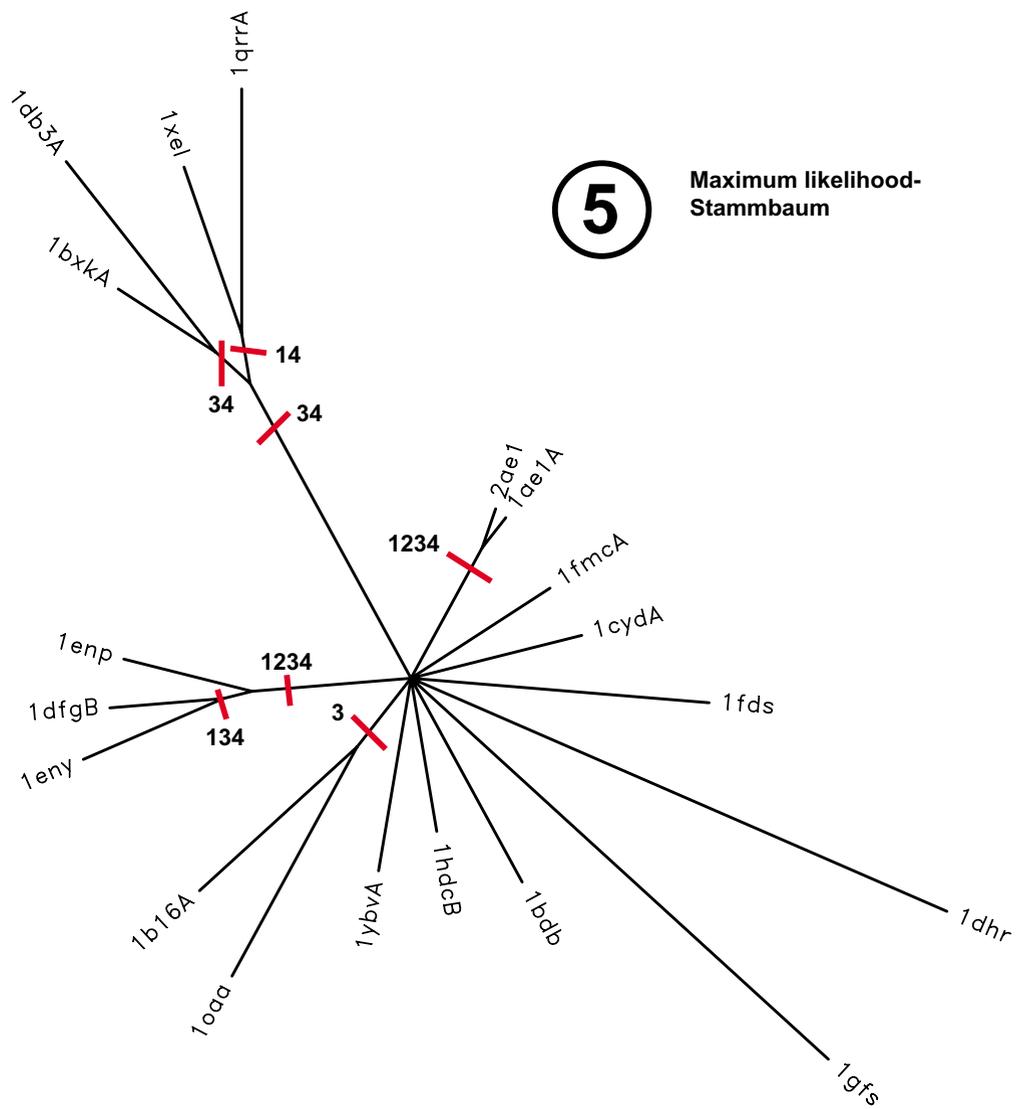


Abbildung 4.7: ... Fortsetzung von Abb. 4.3

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

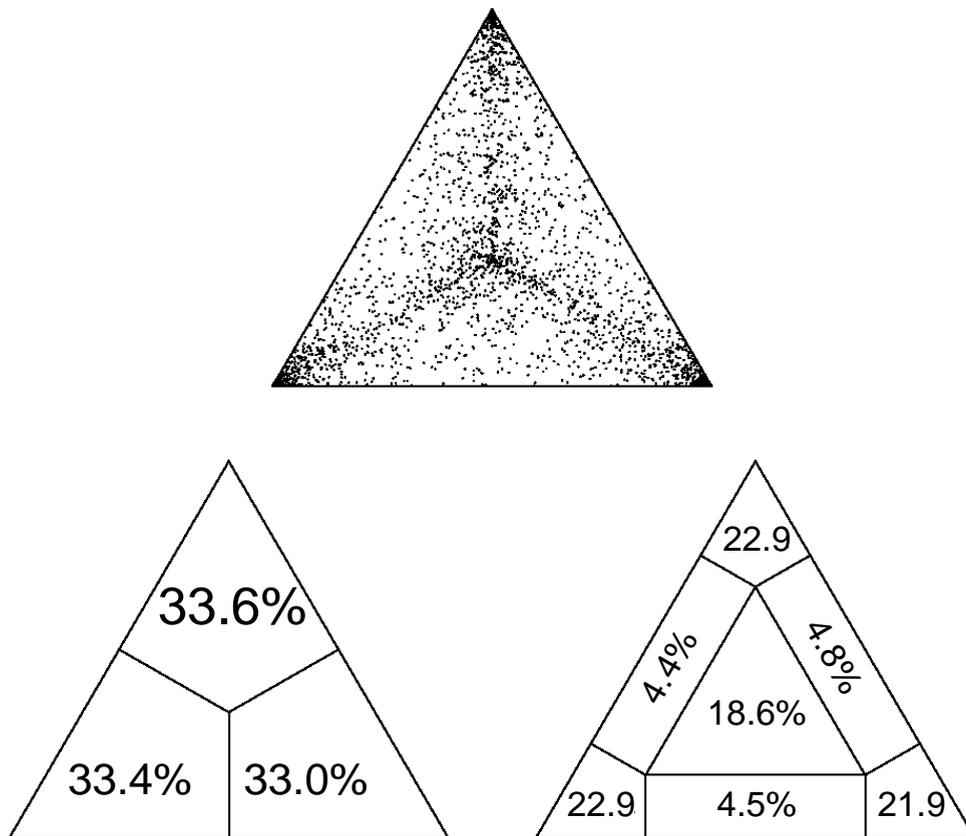


Abbildung 4.8: Phylogenetischer Gehalt des Struktur-basierten Alignments, bestimmt durch Maximum-Likelihood-Mapping. Der größte Teil der Vektoren verteilt sich auf die Eckbereiche des Diagramms ($22.9+22.9+21.9=67.7\%$), nur ein kleiner Teil (18.6%) liegt in Mitte. Das bedeutet, dass der Datensatz geeignet ist, die phylogenetischen Beziehungen der Sequenzen wenigstens teilweise aufzuklären.

Für die Sequenz-basierten Methoden wurde zuerst das kombinierte Alignment der 19 Proteinsequenzen auf seinen phylogenetischen Gehalt geprüft (durch Maximum-Likelihood-Mapping nach Strimmer & von Haeseler, 1997). Dieses grafische Verfahren ermöglicht es, festzustellen, ob der Datensatz für eine phylogenetische Rekonstruktion geeignet ist, oder nicht. Das Ergebnis zeigt, dass sich ein großer Teil der Wahrscheinlichkeitsvektoren (67.7%) auf die drei Ecken des Diagramms verteilt, die einer voll-aufgelösten Topologie des Stammbaums entsprechen (Abb. 4.8, S. 53). Ein nicht zu vernachlässigender Teil (18.6%) befindet sich jedoch in der Mitte des Diagramms, die eine sternförmige, nicht-aufgelöste Topologie repräsentiert. Das bedeutet, dass sich für den betrachteten Datensatz vermutlich nur Teile des Stammbaums zuverlässig aufklären lassen. Die Divergenz der Sequenzen und das „evolutionäre Rau-

**Phylogenetischer
Gehalt des
Protein-Alignments**

4 Die 17β -Hydroxysteroid-Dehydrogenasen

schen“ sind zu hoch, um an allen Stellen den Ablauf der Evolution vollständig zu rekonstruieren. Bereits ein teilweise aufgeklärter, dafür aber besonders verlässlicher Stammbaum wäre aber ein großer Gewinn für das weitere Vorgehen.

Sequenz-basierte Stammbäume

Als nächstes wurden mit drei unabhängigen sequenzbasierten Methoden (Neighbour-joining [NJ], Parsimony und Maximum-Likelihood [ML]) Stammbäume für die 19 kristallisierten HSD-Mitglieder berechnet (Abb. 4.5 bis 4.7, S. 51ff.). In allen drei Methoden wurden Standard-Parameter verwendet und bei den NJ- und Parsimony-Berechnungen wurden jeweils 100 Bootstrap-Pseudoreplikate analysiert, um die Qualität der Verzweigungen abzuschätzen (Felsenstein, 1985).

Stammbaum-Vergleiche

Der Vergleich der fünf Stammbäume lieferte eine Reihe von interessanten Resultaten (Abb. 4.3 bis 4.7, S. 49ff.): Alle Stammbäume, mit Ausnahme des ML-Stammbaumes, sind hochaufgelöst (15 bis 16 Dichotomien, verglichen mit der maximal möglichen Anzahl $n - 3 = 16$). Das ist mehr als angesichts des geringen phylogenetischen Gehalts erwartet werden kann, so dass vermutlich eine Reihe der Dichotomien nicht statistisch abgesichert ist. Der geringer aufgelöste ML-Stammbaum enthält dagegen fast ausschließlich Dichotomien, die sich bereits aus den Enzymaktivitäten mit großer Wahrscheinlichkeit vorhersagen ließen: Enoyl-Acyl-Carrierprotein-Reduktasen, Tropinon-Reduktasen und Nukleotid-Zucker-Dehydratasen bilden hier jeweils monophyletische Gruppen. Die größte Übereinstimmung untereinander zeigen NJ- und Parsimony-Methode, und die sequenzbasierten Verfahren liefern insgesamt homogenere Resultate als die beiden strukturbasierten Methoden. Von diesen hat wiederum die Z-Score-Variante die größere Auflösung und die größere Übereinstimmung mit den sequenzbasierten Verfahren. Voraussetzung für die hohe Qualität der sequenzbasierten Stammbäume war sicher die Verwendung des strukturbasierten Alignments.

Konsensus-Stammbaum

Aus den fünf Stammbäumen wurde schließlich ein Mehrheitsregel-Konsensus-Stammbaum abgeleitet, der 10 Dichotomien enthält, die jeweils von mindestens drei der verwendeten Methoden unterstützt werden (Abb. 4.9, S. 55). Der Stammbaum ist, wie nach dem Maximum-Likelihood-Mapping-Ergebnis zu erwarten, nicht vollständig aufgeklärt, sondern enthält eine Reihe von Polytomien. Er zeigt die größte Übereinstimmung mit dem Parsimony-basierten Stammbaum, mit dem er alle 10 Dichotomien gemeinsam hat. Die geringste Übereinstimmung hat der Mittelwert-basierte Stammbaum. Dieses Ähnlichkeitsmaß scheint also für eine phylogenetische Analyse ungeeignet zu sein, wahrscheinlich weil es die Länge der verglichenen Bereiche nicht berücksichtigt.

Einordnung der anderen 17β -HSDs

Als nächster Schritt wurde versucht, die weiteren 17β -HSDs in den Konsensus-Stammbaum einzuordnen. Dazu wurden durch eine PSI-Blast-Suche und anschließende Redundanzfilterung eine repräsentative Auswahl naher Verwandter der jeweils neu einzuordnenden 17β -HSD bestimmt. Für diese Sequenzen wurde dann ein Alignment bestimmt und dem der Proteinstrukturen hinzugefügt. Positionen, die in mehreren Sequenzen Lücken aufwiesen, wur-

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

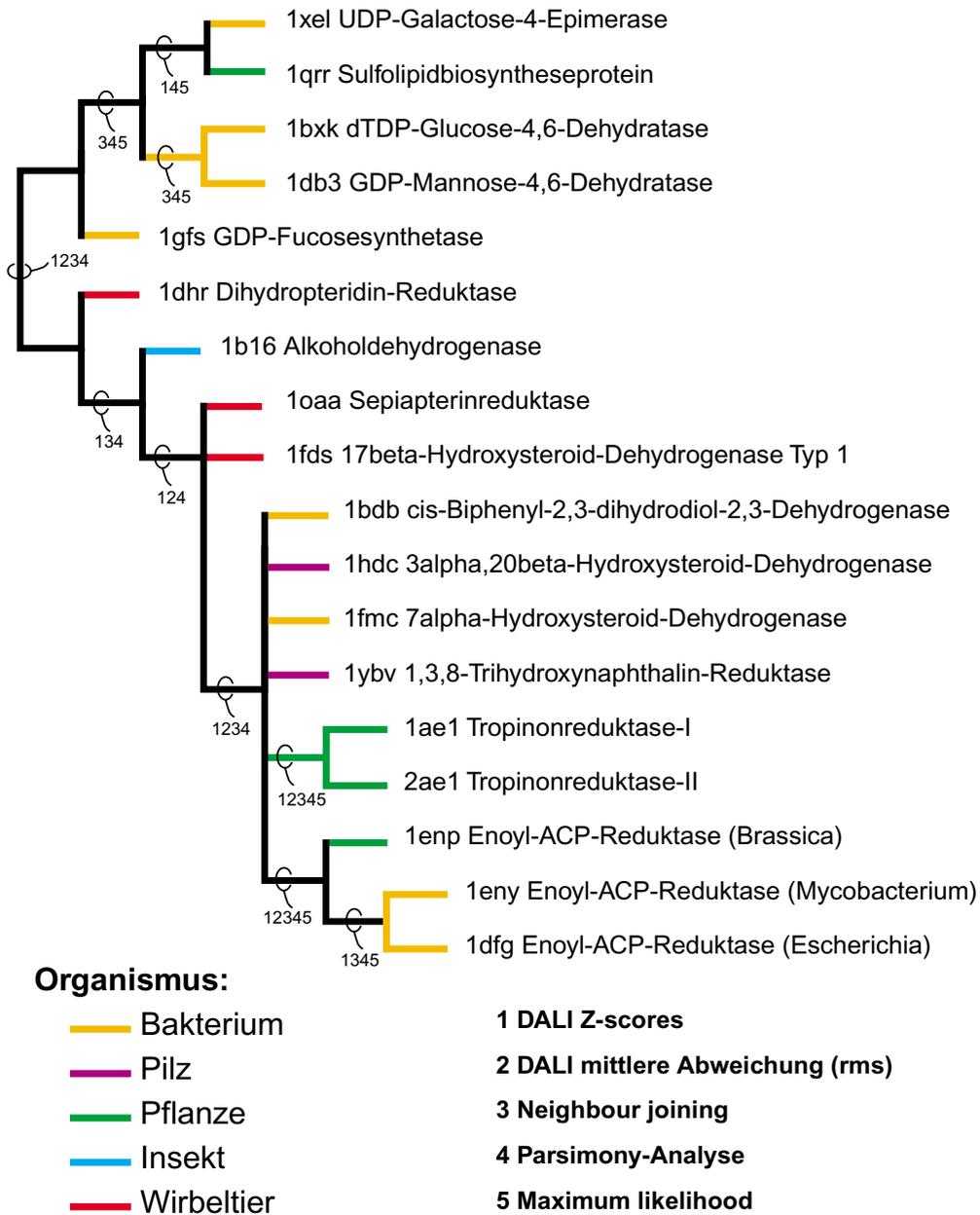


Abbildung 4.9: Konsensus-Stammbaum der HSD-Familie. Die Ziffern geben an, in welchen Einzelstammbäumen die jeweils markierte Dichotomie ebenfalls beobachtet wird.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

den aus dem Alignment entfernt. Das ist nötig, weil die Algorithmen eine korrekte Zuordnung homologer Positionen voraussetzen. Es ist aber nicht sinnvoll, eine Lücke als Homologes einer Aminosäure zu bezeichnen. Anschließend wurde ein Stammbaum für das kombinierte Alignment bestimmt, und zwar mit dem Parsimony-Algorithmus, der in der vorangegangenen Analyse die größte Übereinstimmung mit dem Konsensus-Stammbaum geliefert hatte. Die Hinzunahme von nahen Verwandten der eigentlich interessierenden 17 β -HSD sollte verhindern, dass zufällige Sequenzähnlichkeiten einen zu großen Einfluss auf das Ergebnis der phylogenetischen Analyse nehmen - die Erweiterung der Stichprobe hatte also den Zweck, phylogenetisches Signal und Rauschen zuverlässiger zu trennen.

17 β -HSDs treten in zwei Gruppen auf

Das Resultat der Untersuchung war, dass sich die bekannten 17 β -HSDs tatsächlich in zwei große Gruppen einteilen lassen: 17 β -HSD1, 17 β -HSD2, 17 β -HSD6, 17 β -HSD7 und 17 β -HSD9 stehen den Mitgliedern einer schlecht aufgelösten Gruppe um die 7 α -HSD (1FMC), bestehend aus 17 β -HSD3, 17 β -HSD4, 17 β -HSD8 und 17 β -HSD10 gegenüber (Abb. 4.10, S. 57). Sämtliche 17 β -HSDs haben zwar nähere Verwandte, mit anderer Reaktivität und Substratspezifität, ihre Selektivität ist also auf konvergente Evolution zurückzuführen, aber die Grundeigenschaft, die diese konvergente Entwicklung möglich gemacht hat, scheint nur an zwei Stellen des Stammbaums unabhängig voneinander entstanden zu sein. Die Zweiteilung der 17 β -HSDs korreliert nicht mit der Substratspezifität oder mit Einteilung in aktivierende und inaktivierende Enzyme.

4.1.3 Diskussion

Qualität des Stammbaums

Der Konsensus-Stammbaum bestätigt erneut die frühe und weitgehende Divergenz der HSD-Familie (Jörnvall et al., 1999). Er macht aber auch klar, dass eine zuverlässige Rekonstruktion aller Evolutionsereignisse, insbesondere der frühen Aufspaltungen, nicht möglich sein wird. Gerade das Segment des Stammbaums, das die meisten HSD-Enzyme *im engeren Sinne* enthält, fällt in einer einzigen Polytomie zusammen. Andererseits sind einige der frühen Evolutionschritte deutlich festzustellen, z. B. die frühe Divergenz der *Drosophila*-Alkohol-Dehydrogenasen, die sich später in zahlreiche Isoenzyme aufgegliedert haben (Jörnvall et al., 1999). Auch die Aufteilung der Hydroxysteroid-Dehydrogenasen im engeren Sinne in (mindestens) zwei Gruppen (17 β -HSD1 vs. 3 α /20 β -HSD und 7 α -HSD) ist schon an diesem kleinen Stammbaum abzulesen.

Zeitliche Einordnung der Differenzierung

Bemerkenswert ist auch die Möglichkeit, die Wurzel des Stammbaums mit ziemlicher Sicherheit zu bestimmen, und zwar zwischen den Proteinen, die Nukleotid-Zucker umsetzen (der erweiterten SDR-Familie nach Jörnvall et al., 1999), und den eigentlichen Dehydrogenasen bzw. Oxidoreduktasen. Das Vorkommen der "extended SDR"-Proteine in den komplett sequenzierten Genomen von Archaea, die keine anderen HSD-Mitglieder enthalten, mag dar-

4.1 Ein strukturbasierter Stammbaum der HSD-Familie

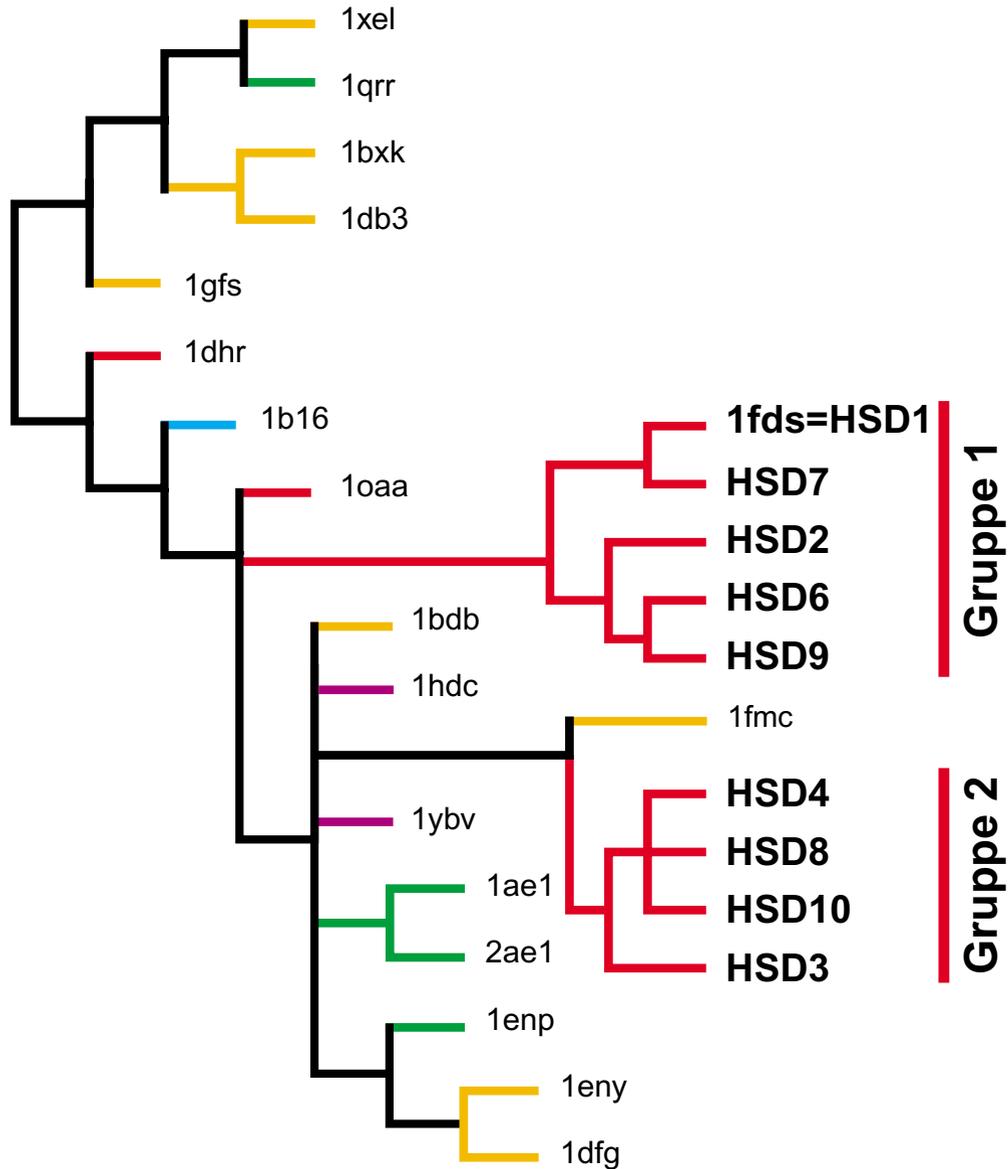


Abbildung 4.10: Konsensus-Stammbaum der HSD-Familie einschließlich der 17β -HSDs. Die Abkürzungen der Proteinstrukturen und die Farbgebung entsprechen Abb. 4.9. Die zwei im Text beschriebenen Gruppen von 17β -HSD-Enzymen sind markiert.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

auf hinweisen, dass es sich beim Umsatz von Zucker-Nukleotiden um die ursprüngliche Funktion der ganzen Proteinfamilie handelt (Jörnvall et al., 1999). Die Identifizierung der mRNA-bindenden Ribonuklease CSP41 aus Spinat als Mitglied der HSD-Familie wurde sogar als Hinweis interpretiert, dass der Ursprung der Proteinfamilie bis in die Übergangszeit zwischen RNA- und DNA-Welt (*sensu* Gesteland & Atkins 1993) zurückreicht (Baker et al., 1998). Um diese Frage abschließend zu klären, ist aber wenigstens die Aufklärung der Struktur und des katalytischen Mechanismus dieser Ribonuklease notwendig.

Zweiteilung der 17 β -Familie

Die neu entdeckte Zweiteilung der 17 β -HSDs ist viel tiefer in der Evolution anzusetzen als die bereits sehr früh beschriebene Einteilung in eine 17 β -HSD1- und eine 17 β -HSD2-Gruppe (Krozowski, 1994). Sie beruht auf der Analyse einer großen Zahl von Enzymen mit unterschiedlichster Substratspezifität. Dadurch war es möglich, die Polytomien aufzulösen, die bisherige, auf HSDs im engeren Sinne beschränkte Stammbäume kennzeichneten (z. B. Baker, 1996). Für die Interpretation war wie bei der Analyse der Paired-Box-Familie entscheidend, dass die Wurzel des Stammbaums zuverlässig bestimmt werden konnte. Damit war die Richtung der evolutionären Entwicklung klar zu erkennen. Durch die Zweiteilung wird in Zukunft auch die weitere, detaillierte Untersuchung der Evolution von Substratspezifitäten und Reaktivitäten in der HSD-Familie erleichtert. Sie kann sich jetzt jeweils auf die Betrachtung einer der Untergruppen beschränken, die übrigens durch ihre Einbindung in den Gesamtstammbaum eine definierte Wurzel besitzen. Das wird die Interpretation in Zukunft zusätzlich vereinfachen, denn dadurch ist die Richtung evolutionärer Trends immer klar festzustellen. Die Umsetzung an der Position 17 des Steroidgerüsts ist mehrmals unabhängig (konvergent) an verschiedenen Stellen der Entwicklungsgeschichte entstanden. Innerhalb der beiden oben definierten Protein-Gruppen (Abb. 4.10, S. 57) handelt es sich bei der 17 β -HSD-Aktivität jedoch um eine „latente Homologie“ (*sensu* de Beer, 1971), d. h. in den beiden Gruppen scheint eine „Neigung“ oder Tendenz zu bestehen, diese Eigenschaft zu entwickeln. Besonders vielversprechend scheint der Versuch zu sein, die strukturelle Grundlage für diese Neigung zur Ausprägung einer 17 β -HSD-Aktivität gerade in diesen beiden Bereichen des HSD-Stammbaums zu bestimmen, z. B. mit Hilfe der für die 17 β -HSD5 entwickelten strukturanalytischen Verfahren (Kap. 4.5, S. 94).

4.2 Die 17 β -HSD Typ 7 – Funktionsvorhersage, Modeling, Regulation der Genexpression¹

4.2.1 Problemstellung

17 β -HSD7 wurde als neues Östradiol-produzierendes Enzym im Ovar von Ratten und Mäusen identifiziert (Nokelainen et al., 1998). Ihr humanes Orthologes wurde daraufhin in unserer Arbeitsgruppe kloniert (Krazeisen et al., 1999). Das Enzym wird in Nagetieren vorwiegend im Corpus luteum exprimiert und wird für den Anstieg der Östradiolproduktion während der Schwangerschaft verantwortlich gemacht (Nokelainen et al., 2000). Diese Funktion wird jedoch beim Menschen der placentalen 17 β -HSD1 zugeschrieben, und es ist unklar, weshalb sich ein zweites hochkonserviertes Protein mit sehr ähnlicher Substratspezifität (Peltoketo et al., 1999a) und weitgehend überlappender Gewebeverteilung überhaupt in der Evolution erhalten hat. Mit Hilfe der in den Vorstudien entwickelten bioinformatischen Techniken sollten alle vorhandenen Daten integriert werden, um ein möglichst detailliertes Bild der physiologischen Funktion von 17 β -HSD7 beim Menschen zu gewinnen.

4.2.2 Hintergrund

Das 17 β -HSD7-Protein der Ratte wurde bereits 1996 kloniert (Duan et al., 1996). Es wurde aber nicht sofort als Dehydrogenase erkannt, obwohl es alle typischen Motive der HSD-Familie enthält, die von Krozowski (1994) und Persson et al. (1991) beschrieben wurden. Stattdessen wurde Ratten-17 β -HSD7 als Ovar-spezifisches Prolactin-Rezeptor-assoziiertes 32kDa-Protein charakterisiert. Das Protein bindet spezifisch an die cytoplasmatische Domäne der kurzen Form des Prolactin-Rezeptors im Corpus luteum, wie durch Koimmunpräzipitation gezeigt werden konnte (Duan et al., 1996). 17 β -HSD7 wird bevorzugt in den Theka-Zellen trächtiger Ratten exprimiert. Die Expression beginnt etwa am Tag 11 p.c. anzusteigen und erreicht zwischen Tag 14 und 15 p.c. ihr Maximum, um dann bis zur Geburt wieder abzufallen (Parmer et al., 1992; Duan et al., 1997). Der Anstieg der 17 β -HSD7-Expression korreliert mit der Östradiol-induzierten Hypertrophie der Lutealzellen in der Mitte der Schwangerschaft.

17 β -HSD7 in der Ratte

1998 wurde aus einer Brustepithelzelllinie (HC11) das 17 β -HSD7-Protein der Maus kloniert und gezeigt, dass es sich um eine 17 β -HSD handelt, die

... in der Maus

¹ Teilweise veröffentlicht in: Breitling R, Krazeisen A, Möller G & Adamski J (2000, im Druck): 17 β -hydroxysteroid dehydrogenase type 7 – an ancient 3-ketosteroid reductase of cholesterologenesis. Mol. Cell. Endocrinol.

Krazeisen A, Breitling R, Imai K, Fritz S, Möller G & Adamski J (1999): Determination of cDNA, gene structure and chromosomal localization of the novel human 17 β -hydroxysteroid dehydrogenase type 7. FEBS Lett. 460: 373-379.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

bevorzugt die Reduktion von Östron zu Östradiol katalysiert (Nokelainen et al., 1998). Das Protein wird ebenfalls in großen Mengen ab der Mitte der Schwangerschaft im Ovar exprimiert. Zusätzlich werden geringere mRNA-Mengen auch in anderen Geweben (Placenta, Brustdrüse, Leber, Niere, Testis) und in den Ovarien nicht-schwangerer Mäuse gefunden. Auch bei Mäusen ist der Prolactin-Rezeptor spezifisch mit einem Protein assoziiert, dessen pI-Wert und Molekulargewicht denen der 17 β -HSD7 ähnlich sind (Clarke et al., 1994).

... im Menschen

Die humane 17 β -HSD7 wurde aus einer ovariellen cDNA-Bibliothek kloniert (Krazeisen et al., 1999). Die Genstruktur und Teile der Intronsequenz des auf Chromosom 10p11.2 liegenden Gens wurden durch teilweise Sequenzierung eines P1-abgeleiteten künstlichen Chromosoms (PAC) bestimmt. Mit RT-PCR konnte gezeigt werden, dass die humane 17 β -HSD7 wie die Proteine der Nagetiere in Ovar, Brust und Placenta, sowie in steroidproduzierenden Geweben des Mannes (Testis, Prostata) exprimiert wird. Außerdem wurde eine sehr hohe Expression in HepG2-Hepatoma-Zellen beobachtet (Krazeisen et al., 1999).

4.2.3 Ergebnisse

in silico-Northern Blot

Um eine erste Vorstellung von der möglichen Funktion der 17 β -HSD7 des Menschen zu erhalten, wurde ein Expressions-Profiling („*in silico Northern blot*“) durchgeführt. Dazu wurde die Datenbank der humanen *expressed sequence tags* (ESTs) nach der cDNA der 17 β -HSD7 durchsucht. Die Annotierung der identifizierten ESTs wurde verwendet, um das Ursprungsgewebe festzustellen. Das gleiche Verfahren wurde auch auf die cDNAs von 17 β -HSD1 und -HSD2 angewendet, deren Expressionsmuster beim Menschen genauer charakterisiert ist (Miettinen et al., 1996; Mustonen et al., 1998; Takeyama et al., 2000). Insgesamt wurden 32 EST gefunden, die von der 17 β -HSD7-cDNA stammen und sich einem Gewebe zuordnen lassen (Abb. 4.11, S. 73). Die beobachtete Gewebeverteilung ist der von 17 β -HSD1 sehr ähnlich. Für beide Proteine gibt es zahlreiche ESTs aus Placenta, Uterus, Ovar und Brust, im Gegensatz zu 17 β -HSD2, die wie erwartet zum größten Teil in Colon gefunden wurde. Mehrere 17 β -HSD7-ESTs stammten aus Hormon-abhängigen Tumoren (z. B. Glioblastom, Adenokarzinom des Endometriums und Brustkrebs). Eine überraschend große Zahl von 17 β -HSD7-ESTs (mehr als 50%) stammte aus Nervengewebe und aus fötaler Leber, einem Gewebe, das als Östradiol-inaktivierend beschrieben wurde (Mustonen et al., 1998).

SAGE-Analyse

Zur weiteren Charakterisierung der 17 β -HSD7-Expression wurde die SAGEmap-Datenbank herangezogen. Diese Datenbank enthält die Ergebnisse von *serial analysis of gene expression* (SAGE)-Experimenten (Velculescu et al., 1995; Zhang et al., 1997) in verschiedenen Zelllinien. Die Ergebnisse sollten zwar die Expression quantitativ beschreiben, bei gering exprimierten Genen überwiegen aber die methodenspezifischen Fehler, besonders Sequenzierungsungenauigkeiten, so dass keine exakte Quantifizierung möglich ist. Stattdessen

4.2 Die 17 β -HSD Typ 7

wurden die Daten von 69 SAGE-Bibliotheken (aus 69 verschiedenen Zelllinien oder Experimenten) zusammengefasst, um eine Vorstellung von der Verbreitung und Höhe der Expression zu erhalten. Das Ergebnis zeigt, dass HSD7-SAGE-*Tags* in einer großen Zahl von SAGE-Bibliotheken gefunden werden (Tab. 4.7, S. 61). Die Zahl der *Tags* ist zwar relativ niedrig, besonders im Vergleich mit dem *house keeping*-Gen Glycerinaldehyd-Phosphat-Dehydrogenase. Sie liegt aber innerhalb der 17 β -HSD-Gruppe im oberen Bereich. Auch die Verteilung der 17 β -HSD7 auf die verschiedenen Bibliotheken ist breiter gestreut als die Expression der 17 β -HSDs Typ 1 bis 3 – ähnlich wie die der ubiquitären 17 β -HSD4. 17 β -HSD7, die ursprünglich als Ovar-spezifisch beschrieben wurde, ist also beim Menschen viel weiter verbreitet.

Tabelle 4.7: SAGE-*Tag*-Verteilung von humanen HSDs und Enzymen der Cholesterin-Biosynthese. Die Sequenz der verwendeten *tags* ist angegeben. Zum Vergleich ist auch das *housekeeping*-Gen Glycerinaldehydphosphat-Dehydrogenase aufgeführt, das mit Abstand die stärkste Expression zeigt und in allen 69 Bibliotheken vertreten ist. Die Enzyme der Cholesterinbiosynthese sind ebenfalls in zahlreichen Bibliotheken repräsentiert. Die 17 β -HSDs haben deutlich geringere Expressionslevel und werden nur in wenigen Bibliotheken gefunden, wobei die Expression der ubiquitären 17 β -HSD4 und der 17 β -HSD7 fast das Niveau der Cholesterinbiosynthese-Enzyme erreicht.

Gen	<i>tag</i>	Zahl der <i>tags</i>	Zahl der Bibliotheken	<i>tags</i> pro Bibliothek
17 β -HSD1	CACCGGGAAG	24	15	0.35
17 β -HSD2	CCCAGAGCTC	16	7	0.23
17 β -HSD3	CTGGCTGAAG	13	11	0.19
17 β -HSD4	TATCATTATT	47	24	0.68
17 β -HSD7	GATTTCTACT	40	23	0.58
HMG-CoA-Reduktase	GAAGTGGCAG	52	34	0.75
Isopentenylidiphosphat-Isomerase	TGACACCCAC	99	43	1.43
<i>tattered</i> -Homolog	GGCAGGCACA	146	47	2.12
<i>bare patches</i> -Homolog	GAAGTGGATT	87	42	1.26
7-DHC-Reduktase	TCGTTTCCTT	86	37	1.25
Glycerinaldehydphosphat-Dehydrogenase	TACCATCAAT	6930	69	100.4
Gesamt		2 902 360	69	42 063

Um herauszufinden, was die Funktion der ubiquitären 17 β -HSD7-Expression sein könnte, wurde nach homologen Proteinen in anderen Organismen gesucht. Ein PSI-Blast identifizierte ein Hefeprotein, die 3-Ketosteroid-Dehydrogenase Erg27p, als nächsten Verwandten der 17 β -HSDs der Säuge-

Identifizierung des nächsten Verwandten

4 Die 17β -Hydroxysteroid-Dehydrogenasen

tiere. Dieses Protein katalysiert die Umsetzung von Zymosteron zu Zymosterol, eine der letzten Reaktionen der Ergosterol- und Cholesterol-Biosynthese (Abb. 4.12, S. 74). Der Blast-Algorithmus ist jedoch nicht für die Analyse von Verwandtschaftsbeziehungen entworfen worden: Da er auf lokalen Alignments basiert, ist es möglich, dass zufällige Ähnlichkeiten (z. B. in der hochkonservierten Nukleotid-Bindungsregion) zu großen Einfluss auf das Ergebnis nehmen. Um solche zufälligen Effekte auszuschließen, wurde die Proteinsequenz in überlappende Blöcke von 80 Aminosäuren zerlegt und für jeden Block eine getrennte PSI-Blast-Suche durchgeführt. Die ersten Blöcke zeigen, wie erwartet, Ähnlichkeiten zu einer großen Zahl von Dehydrogenasen der HSD-Familie, jedoch nicht zu anderen 17β -HSDs. Weiter C-terminal liegende Bereiche haben weniger Ähnlichkeit mit anderen HSD-Proteinen, wie schon Jörnvall et al. (1995) für die ganze Protein-Familie feststellten. Nach dem Block von Aminosäure 121–200, also nach dem aktiven Zentrum, zeigen nur noch Erg27p und seine Homologen aus anderen Hefearten Ähnlichkeit mit 17β -HSD7, und zwar bis zum Block 161-240.

Phylogenetische Analyse

Um die Verwandtschaftsbeziehungen genauer zu charakterisieren, wurden die Sequenzen von 17β -HSD7, Erg27p und weiteren HSD-Proteinen einer phylogenetischen Analyse unterzogen. Die Auswahl der Proteine erfolgte nach dem Ergebnis der blockweisen PSI-Blast-Suche, wobei Proteine gewählt wurden, die in mindestens zwei nicht-überlappenden Bereichen signifikante Ähnlichkeiten mit den 17β -HSD7-Proteinen zeigten. Aus der großen Familie der Protochlorophyllid-Reduktasen wurde je ein repräsentatives Mitglied aus Cyanobakterien, Grünalgen, Mono- und Dikotyledonen ausgewählt. Die Wurzel des Stammbaums wurde durch den Vergleich mit anderen 17β -HSDs bestimmt. Für die ausgewählten Sequenzen wurde ein Alignment berechnet. Positionen mit Lücken in mehr als 50% der Sequenzen wurden entfernt. Die Analyse zeigte, dass 17β -HSD7 und Erg27p tatsächlich eine monophyletische Gruppen bilden, die klar von den anderen HSD-Mitgliedern abgegrenzt ist (Abb. 4.13, S. 74).

Keine 17β -HSD7 in *C. elegans* und *Drosophila*

Bei der Suche nach weiteren Mitgliedern der 17β -HSD7/Erg27p-Familie wurden auch die kompletten Genome von *C. elegans*, *Drosophila* und *Arabidopsis* durchsucht. In keinem dieser Organismen gibt es ein 17β -HSD7-Orthologes, obwohl sie alle eine große Zahl von HSD-Proteinen enthalten.

Transmembran-Helix-Vorhersage

Duan et al. (1996) haben für 17β -HSD7 eine Transmembran-Helix im C-terminalen Bereich vorhergesagt. Eine Analyse der Transmembran-Topologie mit einem Hidden-Markov-Modell zeigt, dass Vertebraten- 17β -HSD7 und die Hefe-Erg27p-Homologen auch diese Eigenschaft gemeinsam haben (Abb. 4.14, S. 75). Bei beiden Proteinen befindet sich ein hydrophober, helikaler Bereich ungefähr zwischen Aminosäure 240 und 260, der bei den entfernteren Verwandten, z. B. den Protochlorophyllid-Reduktasen, fehlt.

Identifizierung von gemeinsamen Sequenz-Elementen

Das Alignment von 17β -HSD7 und Erg27p zeigte, dass die beiden Proteine im Vergleich mit anderen HSD-Mitgliedern einen charakteristischen Einschub von etwa 20 Aminosäuren nach dem NAG-Motiv (Aminosäure 90-92 des hu-

manen Proteins) zeigen. Dieser Einschub wurde bei der phylogenetischen Analyse nicht berücksichtigt, weil sich an dieser Stelle in allen anderen Proteinen Lücken befanden. Er ist also, wie die hydrophobe Helix, ein unabhängiges Indiz für die gemeinsame Herkunft von 17 β -HSD7 und Erg27p.

Um die Funktion der beiden charakteristischen Motive (hydrophobe Helix und 20-Aminosäure-Einschub) zu verstehen, wurde ein dreidimensionales Modell der 17 β -HSD7 berechnet. Die Ergebnisse der 17 β -HSD5-Analyse (siehe unten, Kap. 4.5) hatten gezeigt, dass die geringe Sequenzähnlichkeit zwischen 17 β -HSD7 und den kristallisierten HSD-Proteinen die exakte Modellierung atomarer Details verhindern würde. Stattdessen wurde die globale Struktur und die relative Lage der einzelnen Strukturelemente betrachtet. Als Template diente die Struktur der humanen 17 β -HSD1 (1FDS, Sawicki et al., 1999), die, wie oben gezeigt, zur gleichen Untergruppe der HSD-Familie gehört. Dass sie außerdem das gleiche Substrat und den gleichen Kofaktor benutzt und dieselbe Reaktion katalysiert, ist für die weitere Interpretation natürlich ein zusätzlicher Vorteil. Für die Modell-Berechnung wurde zuerst mit dem JPred-Programm die Sekundärstruktur der HSD7 vorhergesagt. Der Algorithmus identifizierte die α -Helix und β -Faltblatt-Elemente, die den konservierten Kern der HSD-Familie ausmachen. Die Sekundärstrukturvorhersage und das Alignment aller bekannten HSD-Strukturen wurden genutzt, um äquivalente Aminosäuren in 17 β -HSD7 und 17 β -HSD1 zu identifizieren (Abb. 4.15, S. 76). Das Alignment der beiden Sequenzen wurde dann als Grundlage für die Modellierung genutzt. Die Positionierung einzelner Lücken und Einschübe in der Sequenz wurde manuell so optimiert, dass die Zahl der Aminosäure-Zusammenstöße im Modell möglichst gering war. Das vorläufig beste Modell hatte eine Packungsqualität von $Z = -2.570$ und entspricht damit einer niedrig auflösenden Röntgenstruktur (Abb. 4.16, S. 77). Die geringe Sequenzhomologie und die großen, nicht homologen Einschübe in der Sequenz verhinderten eine weitere Verbesserung der Modell-Qualität.

Entscheidend ist aber, dass sich die relative Anordnung der charakteristischen Strukturelemente bereits in diesem groben Modell erkennen lässt. Man kann sehen, dass der 20-Aminosäure-Einschub in der 17 β -HSD7 vermutlich in Kontakt zum Substrat steht. Unter Umständen hat er eine ähnliche Funktion wie die flexiblen Substratbindungs-Schleifen der AKR-Familie (Kap. 4.5.2) oder die charakteristische Insertion peroxisomaler Dehydrogenasen (Kap. 4.3.3). In diesen Fällen sind die zusätzlichen Protein-Schleifen für die Substratspezifität verantwortlich. Das bedeutet, dass auch Erg27p und 17 β -HSD7 ähnliche Substrate umsetzen könnten.

Die hydrophobe Helix ist in dem dreidimensionalen Strukturmodell Teil des konservierten (α/β)-Motivs der HSD-Strukturen. Sie ist also ganz sicher keine Transmembran-Helix, wie ursprünglich vorhergesagt worden war (Duan et al., 1996), denn dazu müsste angenommen werden, dass die 17 β -HSD7 eine grundsätzlich andere Konformation besitzt als die bisher kristallisierten HSD-Proteine. Die exponierte Lage der Helix könnte aber trotzdem die Bindung an

Sekundär-Struktur-Vorhersage und 3D-Modellierung

Identifizierung der Substrat-Bindungs-Region

Die hydrophobe Helix ist keine Transmembran-Helix

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

die Zellmembran vermitteln.

Nachdem sich gezeigt hatte, dass 17 β -HSD7 beim Menschen ein ganz anderes Expressionsmuster hat als es bisher bei Nagetieren gefunden wurde (Nokelainen et al., 1998, 2000), wurde versucht, die Grundlage dieser Expression durch eine Promotor-Analyse aufzuklären. Dafür wurden zunächst mögliche Transkriptions-Faktor-Bindungsstellen im Promotor-Bereich der humanen 17 β -HSD7 bestimmt. Die Analyse identifizierte zwar eine ganze Reihe von möglichen Bindungsstellen, z. B. für die gehirnspezifischen Transkriptionsfaktoren Ikaros und *neuron-restrictive silencer* (NRS) oder für den *nuclear factor of activated T-cells* (NF-AT). Diese Elemente treten aber viel zu häufig auf, unter anderem auch in nicht regulatorischen Sequenzen, als dass sie für die Spezifität des Promotors verantwortlich gemacht werden könnten. Diese wird vermutlich durch das Zusammenspiel mehrerer Transkriptionsfaktoren in charakteristischer Anordnung (Funktionelle Module nach Klingenhoff et al., 1999) bewirkt.

Um solche funktionellen Transkriptionsfaktor-Module zu identifizieren, eignet sich besonders ein vergleichender Ansatz, bei dem die Gemeinsamkeiten in koregulierten Promotoren bestimmt werden (Klingenhoff et al., 1999). Um koregulierte Proteine zu identifizieren, wurden zuerst die Ergebnisse von cDNA-Microarray-Experimenten (Brown & Botstein, 1999) untersucht. Als Maß für die Ähnlichkeit des Expressionsmusters zweier Gene X und Y unter N verschiedenen Bedingungen diente der Korrelationskoeffizient nach Eisen et al. (1998):

$$S(X, Y) = \frac{1}{N} \sum_{i=1, N} \frac{X_i - X_{offset}}{\Phi_X} \frac{Y_i - Y_{offset}}{\Phi_Y}$$

$$\text{mit } \Phi_G = \sqrt{\sum_{i=1, N} \frac{(G_i - G_{offset})^2}{N}},$$

N = Zahl der Datenpunkte,

$$G_i = \log \frac{\text{Fluoreszenzsignal der Probe}}{\text{Fluoreszenzsignal der Referenz}}$$

und G_{offset} = Mittelwert der Beobachtungen.

Die 17 β -HSD7 war nur in einem Microarray-Experiment untersucht worden und zeigte dort keine signifikante Koregulation mit anderen Genen. Andere 17 β -HSDs (Typ 1 bis 3) zeigten in anderen Experimenten ebenfalls keine Koregulation. Da aufgrund der Ähnlichkeit von 17 β -HSD7 und Erg27p angenommen wurde, dass 17 β -HSD7 am Cholesterin-Metabolismus beteiligt sein könnte, wurden auch Proteine der Cholesterin-Biosynthese untersucht. Von diesen zeigte ein Protein (die Isopentenylidiphosphat-Dimethylallyl-Isomerase Typ 1 [IDI1]) eine signifikante Koregulation mit mehreren anderen Proteinen (Abb. 4.17, S. 78). In einem Experiment, das die Reaktion von Fibroblasten auf die Zugabe von Serum untersuchte (Iyer et al., 1999), wurde die Expression

Transkriptionsfaktor-Bindungsstellen

Microarray-Experimente zur Vorhersage von Promotor-Modulen

Bestimmung koregulierter Gene

von IDI1 nach etwa 8 Stunden abgeschaltet. Dieses Verhalten korrelierte stark mit dem des ribosomalen Proteins L39 (RPL39), einer STE-20-verwandten Serin/Threonin-Kinase (OSR1) und eines StAR-artigen Proteins (MLN64). Außerdem korrelierte die IDI1-Expression erwartungsgemäß mit der Expression von anderen Proteinen der Cholesterin-Biosynthese (z. B. Lanosterol-14 α -Demethylase, Squalene-Epoxidase, Farnesyl-Diphosphat-Farnesyltransferase 1, Stearoyl-CoA-Desaturase) und von Insulin-induziertem Protein (INSIG1). In Anwesenheit des Proteinsynthese-Inhibitors Cycloheximid wird die Expression der ersten vier Gene (IDI1, RPL39, OSR1, MLN64) nach Serumzugabe sofort und stark hochreguliert, im Gegensatz zu den anderen Proteinen der Cholesterin-Biosynthese, deren Expression unbeeinflusst bleibt. In einem zweiten Experiment, das die Gen-Expression in 60 verschiedenen Tumorzelllinien unterschiedlichster Herkunft untersucht (Ross et al., 2000), korreliert die Expression von IDI1 ebenfalls signifikant mit der Expression von RPL39, OSR1 und MLN64, während die Korrelation mit den Proteinen der Cholesterin-Biosynthese in diesem Experiment viel schwächer ist. 17 β -HSD7 war in diesen beiden Datensätzen (Iyer et al., 1999; Ross et al., 2000) *nicht* enthalten.

Um die molekulare Grundlage der Koregulation festzustellen wurde die Gen-Struktur und genomische Sequenz der koregulierten Proteine bestimmt. Dafür wurden die Sequenzen der cDNA mit den vorläufigen genomischen Sequenzen der GenBank verglichen. Für IDI1 wurden auch unveröffentlichte Sequenzier-Ergebnisse (J.Adamski et al., persönliche Mitteilung) herangezogen, um eine Lücke in der Datenbanksequenz zu schließen. Zum Vergleich wurde durch Datenbankrecherchen auch die komplette genomische Sequenz der 17 β -HSD7 bestimmt. Die vier Gene befinden sich auf vier verschiedenen Chromosomen und ihre Größe und Struktur unterscheidet sich erheblich (Abb. 4.18, S. 79 und Tab. 4.8, S. 65).

Bestimmung der Gen-Strukturen

Tabelle 4.8: Genomische Struktur und Lokalisation von 17 β -HSD7 und der koregulierten Proteine, die das IDI1-Promotor-Modell enthalten (nach vorläufigen Daten des humanen Genomprojekts).

Name	Lokalisation	Struktur	Gen-Länge	cDNA-Länge ^a
IDI1	Chr. 10	5 Exons	8 134 bp	1 807 bp
RPL39	Chr. 19	3 Exons	5 105 bp	371 bp
OSR1	Chr. 3	18 Exons	89 917 bp	4 519 bp
MLN64	Chr. 17	13 Exons	23 927 bp	2 020 bp
HSD7	Chr. 10	9 Exons	22 060 bp	1 434 bp

^a nach EST-Daten abgeschätzt.

Durch Promotor-Analyse mit dem PromoterInspector (Scherf et al., 2000) wurde in den genomischen Sequenzen je ein Polymerase-II-Promotor in unmittelbarer Nachbarschaft zum erwarteten Transkriptions-Startpunkt von IDI1,

Vorhersage der Transkriptions-Startpunkte

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

OSR1 und MLN64 vorhergesagt. Dieser Kontext-basierende Algorithmus sagt weniger als einen Promotor pro 10 000 bp vorher und das Verhältnis von korrekten zu falschen Vorhersagen ist besser als 1 zu 2, viel besser als bei allen anderen getesteten Algorithmen (Scherf et al., 2000). Das bedeutet, dass die vorhergesagten Promotoren mit großer Wahrscheinlichkeit die echten Transkriptions-Startpunkte kennzeichnen. Die weitere Analyse beschränkte sich deshalb auf den vorhergesagten Promotorbereich und je 200 zusätzliche Basenpaare auf beiden Seiten. Für RPL39 wurde kein Promotor vorhergesagt. Deshalb wurden je 500 bp auf beiden Seiten des Anfangs von Exon 1 in die Analyse einbezogen. Durch bloße Sequenzvergleiche war jedoch keine signifikante Ähnlichkeit zwischen den vier koregulierten Promotoren festzustellen.

Um gemeinsame funktionelle Module festzustellen, wurden mit dem MatInspector (Quandt et al., 1995) mögliche Transkriptionsfaktor-Bindungsstellen in allen vier Promotoren bestimmt und die Transkriptionsfaktor-Familien identifiziert, die in allen Promotoren binden können. Dazu gehörten u. a. die *cAMP response element binding protein* (CREBP)-Familie, die *sterol regulatory element binding protein* (E-BOX/SREBP)-Familie, und die SP1- und NF-kappaB-Familien. Aus diesen Elementen konnte ein modulares Promotor-Modell zusammengestellt werden, das allen vier Promotoren gemeinsam ist (IDI1-Promotor-Modell: Abb. 4.18, S. 79 und Tab. 4.9, S. 66). Es besteht aus zwei Teil-Modulen in variablem Abstand und enthält Bindungsstellen für die CREBP-Familie, die SP1-Familie und zweimal für die EBOX/SREBP-Familie.

Tabelle 4.9: Beschreibung des gemeinsamen Promotor-Modells von IDI1, RPL39, MLN64 und OSR1. Die Nomenklatur der Teilmodule entspricht der Benennung der Transkriptionsfaktor-Matrizen in der MatInspector-Datenbank (<http://genomatix.gsf.de>). Dabei gibt das V\$-Präfix an, dass es sich um Transkriptionsfaktorfamilien aus Vertebraten handelt, während die darauf folgende Abkürzung die einzelnen Familien bezeichnet: CREB = *cAMP response element binding protein*-Familie, EBOX = *E-BOX/sterol regulatory element binding protein*-Familie, SP1F = SP1-Familie.

Teil-Modul A	V\$CREB (+Strang, Core-Ähnlichkeit >0.744, Matrix-Ähnlichkeit >0.841 55 bis 65 bp Abstand V\$EBOX (beide Stränge, C>0.847, M>0.846)
Teil-Modul B	V\$EBOX (beide Stränge, C>0.796, M>0.840) 15 bis 27 bp Abstand V\$SP1F (+Strang, C=1.000, M>0.885)
IDI1-Promotor-Modell	Teil-Modul A (beide Richtungen, Schwellenwert>84) 95 bis 250 bp Abstand Teil-Modul B (beide Richtungen, Schwellenwert>91)

Das Promotor-Modell ist spezifisch für die vier koregulierten Promotoren. Es fehlt im Promotor von IDI2, einer duplizierten Form von IDI1 (Ma-

Entwicklung eines Promotor-Modells

Überprüfung der Spezifität

suda et al., in Vorbereitung), und in allen 1375 eukaryontischen Promotoren in der EPD-Datenbank, die insgesamt 825 000 bp enthält. Eine Untersuchung von mehr als 800 Millionen Basenpaaren humaner Sequenzen zeigte, dass das Modell etwa alle 4 Millionen Basenpaare einmal auftritt. Es ist also hochspezifisch. Die einzelnen Elemente können auch zwei regulatorische Eigenschaften der Promotoren erklären: Sie werden durch einen Überschuss an Cholesterol ausgeschaltet (Sterol-responsives Element), in Anwesenheit des Proteinsynthese-Inhibitors Cycloheximid werden sie schnell hochreguliert (cAMP-responsives Element).

Bei der Überprüfung humaner Sequenzen (EMBL Release 63: 824 675 648 bp, 113 725 Sequenzen) wurden insgesamt 190 Treffer erzielt. In 26 Fällen wurde ein vermutlicher Transkriptions-Startpunkt in der Nähe des IDI1-Promotor-Modells gefunden (Tab. 4.10, S. 68). Einer davon war der Beginn des 17 β -HSD7-Gens. Die Wahrscheinlichkeit, das Modell zufällig im 17 β -HSD7-Gen (22 060 bp) zu finden, beträgt $(22\,060 \times 190 / 824\,675\,648 = 1:197)$, die Wahrscheinlichkeit eines Treffers in unmittelbarer Nähe des Transkriptions-Startpunktes ist etwa 20mal geringer (weniger als 1:4000). Da einige Sequenzen, in denen das IDI1-Promotor-Modell gefunden wurde, mehrfach in der Datenbank vertreten sind, ist die tatsächliche Wahrscheinlichkeit sogar noch etwas geringer. Die Anwesenheit des koregulierten IDI1-Modells im 17 β -HSD7-Promotor ist also hochsignifikant. Außer dem 17 β -HSD7-Promotor enthält auch der Promotor des C14-Sterolreduktase-Homologen TM7SF2 das IDI1-Modell.

17 β -HSD7 enthält ebenfalls das IDI1-Promotor-Modell

4.2.4 Diskussion

Die bioinformatische Untersuchung der humanen 17 β -HSD7 zeigte einige überraschende Resultate.

Die Expression von 17 β -HSD7 in der Leber war unerwartet, denn dieses Organ ist für die Inaktivierung von Östradiol verantwortlich (Mustonen et al., 1998). Das Ergebnis entspricht aber den Beobachtungen von Krazeisen et al. (1999) und von Schwabe et al. (2000). Die Anwesenheit von pro-östrogener 17 β -HSD7 in der Leber kann zur lokalen Östradiol-Produktion notwendig sein. Möglicherweise setzt 17 β -HSD7 dort aber auch ein anderes, unidentifiziertes Substrat um. Außer der spezifischen ovariellen Expression, die bei der Ratte beobachtet wurde, findet sich für die humane 17 β -HSD7 kräftige Expression in Leber und neuronalem Gewebe. Die SAGE-Analyse weist sogar auf eine fast ubiquitäre Expression hin, ähnlich wie bei der ubiquitären 17 β -HSD4. 17 β -HSD7 ist also beim Menschen viel weiter verbreitet und keineswegs so spezifisch exprimiert wie es bei den Nagetieren der Fall zu sein scheint.

Ubiquitäre Expression der 17 β -HSD7

Der nächste Verwandte der 17 β -HSD7 ist ein Protein der Ergosterol-Biosynthese aus Hefen. Die von diesem Protein katalysierte Reaktion, die Reduktion von 3-Ketosterolen (z. B. Zymosteron) zu 3 β -Hydroxysterolen (z. B. Zymosterol), ist der von 17 β -HSD7 katalysierten Reduktion von Östron zu

Funktion bei der Cholesterinsynthese

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Tabelle 4.10: Genomische Sequenzen, die das IDI1-Promotor-Modell in der Nähe ihres vorhergesagten Transkriptionsstarts enthalten. Einige Gene sind in mehreren verschiedenen Einträgen in der Datenbank vorhanden (HADH-SC, TRX, CTSD, XB, EDMD), und einige Gene enthalten das Promotormodell zweimal (TPT1 und Guaninnucleotid-bindendes Regulatorprotein).

Accession-Nr.	Gename	Ähnlichkeit (%)
AF096303	C14-Sterolreduktase-Homolog (TM7SF2)	90
AF162759	17 β -Hydroxysteroiddehydrogenase (HSD17B7)	93
AF026853	L-3-Hydroxyacyl-CoA-Dehydrogenase (HADHSC)	92
AF001903	ebenso	92
AF001904	ebenso	92
D45370	Spezifisches Protein des Fettgewebes	91
AF030246	peroxisomale Dienoyl-CoA-Isomerase (ECH1)	89
AF010400	RNA Helicase p68 (HUMP68)	90
AF026853	jagged 2	92
AF084048	Thioredoxin (TRX)	92
X70286	ebenso	92
AF118569	Angiotensin-Converting-Enzym (DCP1)	93
AF155195	cGMP-spezifische Phosphodiesterase (PDE5A)	92
AL031228	RING1 related-Protein	91
AJ249275	Methylentetrahydrofolat-Reductase (MTHFR)	91
AJ400717	translational kontrolliertes Tumorprotein (TPT1) (zwei Treffer)	90 und 91
L12980	Cathepsin D (CTSD)	95
S74689	ebenso	95
AL049547	Tenascin XB-1 (XB)	90
U52697	ebenso	90
U89337	ebenso	90
X86810	Emerin (EDMD)	92
L44140	ebenso	92
M60156	Guaninnucleotid-bindendes Regulatorprotein (zwei Treffer)	93 und 94
L19686	Macrophagenmigration-inhibierender Factor (MIF)	90
AJ239325	Integrin β 2	91

Östradiol sehr ähnlich. Sie erfolgt allerdings am anderen Ende des Steroidgerüsts (Position 3 statt Position 17) (Abb. 4.12, S. 74). Eine solche Flexibilität wird aber auch bei der humanen 17 β -HSD5 und ihren Verwandten beobachtet (Kap. 4.5) – dort setzt sogar dasselbe Enzym Steroide an Position 3, 17 und 20 um. Es ist also nicht ausgeschlossen, dass auch 17 β -HSD7 ursprünglich als 3-Ketosteroid-Reduktase aktiv war, also als Enzym der Cholesterinbiosynthese. Die 17-Ketosteroid-Reduktase-Aktivität hätte sich dann erst später zusätzlich entwickelt. Ein vergleichbarer Vorgang wird in Kap. 4.3 (S. 81) für die 17 β -HSD4 beschrieben, die in erster Linie ein Protein der peroxisomalen β -Oxidation ist. Eine Reihe zusätzlicher Beobachtungen sprechen für diese Theorie:

1. Die Cholesterol-auxotrophen Organismen *C. elegans* und *Drosophila* besitzen keine 17 β -HSD7/Erg27p-homologen Proteine. Angesichts der deutlichen Verwandtschaft der Enzyme von Hefe und Mensch ist es unwahrscheinlich, dass die homologen Proteine in Invertebraten so weit divergiert sind, dass sie nicht mehr als solche erkannt werden. *Arabidopsis*, der ebenfalls ein 17 β -HSD7/Erg27p-Homologes fehlt, nutzt bei der Sterol-Synthese einen anderen Reaktionsweg als Wirbeltiere und Hefen.
2. Die Proteine von Mensch und Hefe haben als gemeinsames Charakteristikum eine 20-Aminosäuren-Insertion, die laut Modellierungs-Ergebnis vermutlich an der Substratbindung beteiligt ist. Eine solche Konservierung ist nur dann erklärbar, wenn tatsächlich sehr ähnliche Substrate umgesetzt werden. Ein solcher Effekt wurde z. B. bei den 17 β -HSD4-Enzymen beobachtet, wo die An- und Abwesenheit einer kleinen Insertion mit der Präferenz für lang- und kurzkettige Fettsäuren korreliert (siehe Kap. 4.3.3, S. 85).
3. 17 β -HSD7 und Erg27p haben außerdem eine hydrophobe Helix gemeinsam, die ursprünglich für eine Transmembran-Helix gehalten wurde (Abb. 4.14, S. 75). Die Struktur-Modellierung zeigt jedoch, dass es sich eher um eine oberflächliche Helix handelt, die den Kontakt mit der Zellmembran vermittelt. Das Modell zeigt außerdem, dass sich die vorhergesagte Substratbindungstasche in Richtung dieser Helix (und damit in Richtung der Zellmembran) öffnet. Die Beteiligung am Metabolismus von Zellmembran-Komponenten wie Cholesterin ist also sicher für beide Proteine möglich.
4. Besonders hohe Expressionslevel von 17 β -HSD7 wurden in fötaler Leber, differenzierenden Neuronen und Placenta gefunden. Das sind alles Gewebe mit sehr aktiver *de novo*-Cholesterinsynthese (Jurevics et al., 1997; Morell & Jurevics, 1996; Feingold et al., 1983). Selbst die hohe Expression im Corpus luteum könnte auf die cholesterogene Wirkung von 17 β -HSD7 zurückzuführen sein, denn eine der Hauptfunktionen von Östradiol im Gelbkörper der Ratte ist die Stimulation der *de*

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

novo-Cholesterinsynthese (Azhar et al., 1989). Cholesterinbiosynthese-Inhibitoren wie Simvastatin können die Hormonproduktion der Follikelzellen blockieren (Spicer et al., 1996). Auch die Expression in anderen steroidogenen Geweben könnte auf deren erhöhten Cholesterinbedarf zurückzuführen sein.

Die Promotor-Analyse als zusätzliches Indiz

Besonders auffallend ist die Identifizierung eines gemeinsamen Promotor-Modells in den Promotoren von IDI1, einem Schlüsselenzym der Cholesterin-Biosynthese (Hahn et al., 1996), und 17 β -HSD7. Um zu beurteilen, welche Bedeutung dieses Modell haben könnte, muss zunächst seine Qualität festgestellt werden. Bei der Suche in den großen genomischen Datenbanken und in der Datenbank eukaryontischer Promotor-Sequenzen zeigte sich, dass das IDI1-Promotor-Modell sehr spezifisch ist. Es ist jedoch sehr schwer, exakt zu berechnen, mit welcher Wahrscheinlichkeit sich spezifische Promotor-Modelle auch für Sequenzen entwerfen lassen, die nicht gemeinsam reguliert werden. Deshalb ist es wichtig, festzustellen, ob die beobachtete Koexpression von IDI1, OSR1, RPL39 und MLN64 wirklich durch homologe Promotorelemente bewirkt wird. Wäre dies nicht der Fall, so wäre auch das beschriebene IDI1-Promotor-Modell ein Artefakt und seine Anwesenheit im 17 β -HSD7-Promotor hätte nichts zu bedeuten.

Welche anderen Gründe könnten die in Microarray-Experimenten beobachtete Koexpression bewirken?

1. Wenn die Expressionslevel zweier Gene nur wenig variieren, z. B. in allen Experimenten sehr niedrig sind, dann ist auch der Korrelationskoeffizient sehr hoch (etwa 1). Zufällige, gleichgerichtete Schwankungen können dann als signifikante Koexpression erscheinen. Für die Koexpression von IDI1, OSR1, RPL39 und MLN64 kann dieser Effekt nicht verantwortlich sein, denn im Serumantwort-Experiment (Iyer et al., 1999) sind die Veränderungen der Expressionslevel sehr deutlich. Der Zeitverlauf der Expression dient hier als zusätzliche Kontrolle, dass es sich um eine regulierte Antwort und nicht um zufällige Schwankungen handelt.
2. Wenn Proteine auf benachbarten Genorten kodiert werden, die in Tumoren regelmäßig gemeinsam amplifiziert sind (Amplicon), dann sind sie in Tumorzelllinien oft koexprimiert. Ein Beispiel ist MLN64, das in Brustkrebs-Microarrays mit ERBB2 und GRB7 gemeinsam exprimiert wird (Perou et al., 2000). Die Existenz dieses Amplicons war bereits in anderen Experimenten festgestellt worden (Kishi et al., 1997; Moog-Lutz et al., 1997). Die hier untersuchten Gene liegen jedoch alle auf verschiedenen Chromosomen. Ein Amplicon-Effekt kann also ausgeschlossen werden.
3. Gene der Ergosterol-Biosynthese in Hefe (Hughes et al., 2000) und der Cholesterin-Biosynthese in Fibroblasten (Iyer et al., 1999) bilden in

Microarray-Experimenten prominente Cluster. Das bedeutet nicht notwendig, dass sie den gleichen Regulationsmechanismus benutzen, sondern kann auch darauf zurückzuführen sein, dass sie alle auf Änderungen der Sterol-Konzentration im Medium in der gleichen Richtung reagieren. Für die Koexpression von IDI1, OSR1, RPL39 und MLN64 ist das als Erklärung aber unwahrscheinlich, denn sie sind auch in 60 Tumorzelllinien koexprimiert, die unter standardisierten Bedingungen kultiviert wurden (Ross et al., 2000).

Bei Proteinen, die in Microarray-Experimenten als koexprimiert erkannt werden, liegt es nahe, nach einer gemeinsamen Funktion zu suchen. Z. B. wurde während der Untersuchung der IDI1-Expression festgestellt, dass IDI1 in Fibroblasten mit Stearoyl-CoA-Desaturase kovariiert. Erst kürzlich wurde dann gezeigt, dass letzteres Protein tatsächlich überraschende regulatorische Funktionen in der Cholesterinester-Synthese hat (Miyazaki et al., 2000). Was könnte die gemeinsame Funktion von IDI1, OSR1, RPL39 und MLN64 (und damit auch von 17 β -HSD7) sein?

RPL39 ist ein Ribosomen-Bestandteil und sollte abhängig von der Translationsrate konstitutiv exprimiert sein. In Mais und *Candida albicans* variiert die RPL39-Expression im Verlauf des Zellzyklus (Delbrück et al., 1997; Dresselhaus et al., 1999). IDI1 ist ein Schlüsselenzym der Cholesterin-Biosynthese (Hahn et al., 1996). MLN64 ist ein Homologes des *sterol acute regulatory* (StAR)-Proteins und stimuliert wie dieses die Steroidogenese (Watari et al., 1997). OSR1 ist eine Serin/Threonin-Kinase, die als Antwort auf oxidativen Stress exprimiert wird (OSR1 = *oxidative stress responsive protein 1*). Solche Proteine sind in Hefen mit Proteinen der Ergosterol-Biosynthese koreguliert (Bammert & Fostel, 2000). Für 17 β -HSD7 wird in der vorliegenden Arbeit eine Funktion in der Cholesterinbiosynthese angenommen. Gemeinsam ist allen diesen Proteinen also eine Beteiligung am Steroidmetabolismus. Steroidmetabolisierende Enzyme sind in Pflanzen (Schrick et al., 2000) und Säugetieren (Reviews in Moebius et al., 2000; Roux et al., 2000; Kelley, 2000) an embryonalen Entwicklungsprozessen beteiligt. Intermediärmetabolite der Cholesterinbiosynthese sind wichtige Signalmoleküle in der Embryonalentwicklung, bei Apoptose, Proteinmetabolismus und Zellproliferation (Edwards & Ericsson, 1999). Die Koregulation einer solchen scheinbar willkürlichen Auswahl von Steroid-metabolisierenden Enzymen könnte also darauf hinweisen, dass auch sie neben der Cholesterolbiosynthese an einem anderen Signalweg beteiligt sind.

Die bioinformatische Analyse hat also gezeigt, dass 17 β -HSD7 möglicherweise eine zentrale Rolle bei der Cholesterinbiosynthese spielt, und zwar bei der Reduktion von 3-Ketosterolen wie Zymosteron. Unmittelbar anschließende Reaktionen werden von Δ^8 - Δ^7 -Sterolisomerase und 3 β -Hydroxysteroiddehydrogenase katalysiert (Abb. 4.19, S. 80), deren Mutation zu schweren Entwicklungsstörungen führt: Chondrodysplasie, Ichthyosis, Katarakt und weitgehenden Extremitätendefekten (Derry et al., 1999; Liu et al., 1999; König et

Funktion der Koregulation

17- β -HSD7 und Entwicklungsdefekte

4 Die 17β -Hydroxysteroid-Dehydrogenasen

al., 2000). Es ist interessant zu bemerken, das nur ein weiteres Cholesterin-Biosynthese-Enzym ebenfalls das IDI1-Promotor-Modell enthält, nämlich die C14-Sterolreduktase (TM7SF2), und auch Defekte in der von diesem Enzym katalysierten Reaktion führen zu ganz ähnlichen (aber deutlich schwereren) Entwicklungsstörungen (Greenberg-Syndrom: Kelley, 2000). Damit ist 17β -HSD7 nicht nur an der ovariellen Östradiolsynthese beteiligt, sondern auch ein Kandidaten-Gen für angeborene Missbildungen des Menschen. Unter Umständen wurde diese Funktion nur deshalb noch nicht entdeckt, weil 17β -HSD7 anders als $\Delta 8$ - $\Delta 7$ -Sterolisomerase und 3β -Hydroxysteroiddehydrogenase nicht X-chromosomal vererbt wird, so dass sie nicht den gleichen auffälligen Phänotypus bewirkt, wie er bei *bare patches*-Mäusen oder CHILD-Patienten durch die X-Inaktivierung verursacht wird.

Tabelle 4.11: Enzyme der Cholesterin-Biosynthese in Hefe und Mensch

Enzym	Hefehomolog	Homologie Hefe-Mensch
1 C3-Steroldehydrogenase (C4-Decarboxylase)	Erg26p	38%
2 3-Ketosteroiddehydrogenase (17β -HSD7?)	Erg27p	22%
3 Sterol- $\Delta 8$ - $\Delta 7$ -Isomerase	Erg2p	30%

Tabelle 4.12: Defekte der Cholesterinbiosynthese bei Maus und Mensch

Aktivität	Maus	Mensch
1 C3-Steroldehydrogenase	bare patches, striated	CHILD-Syndrom
2 3-Ketosteroiddehydrogenase	unbekannt	17β -HSD7-Defizienz?
3 Sterol- $\Delta 8$ - $\Delta 7$ -Isomerase	tattered	Conradi-Hünemann-Happle-Syndrom (CDPX2)

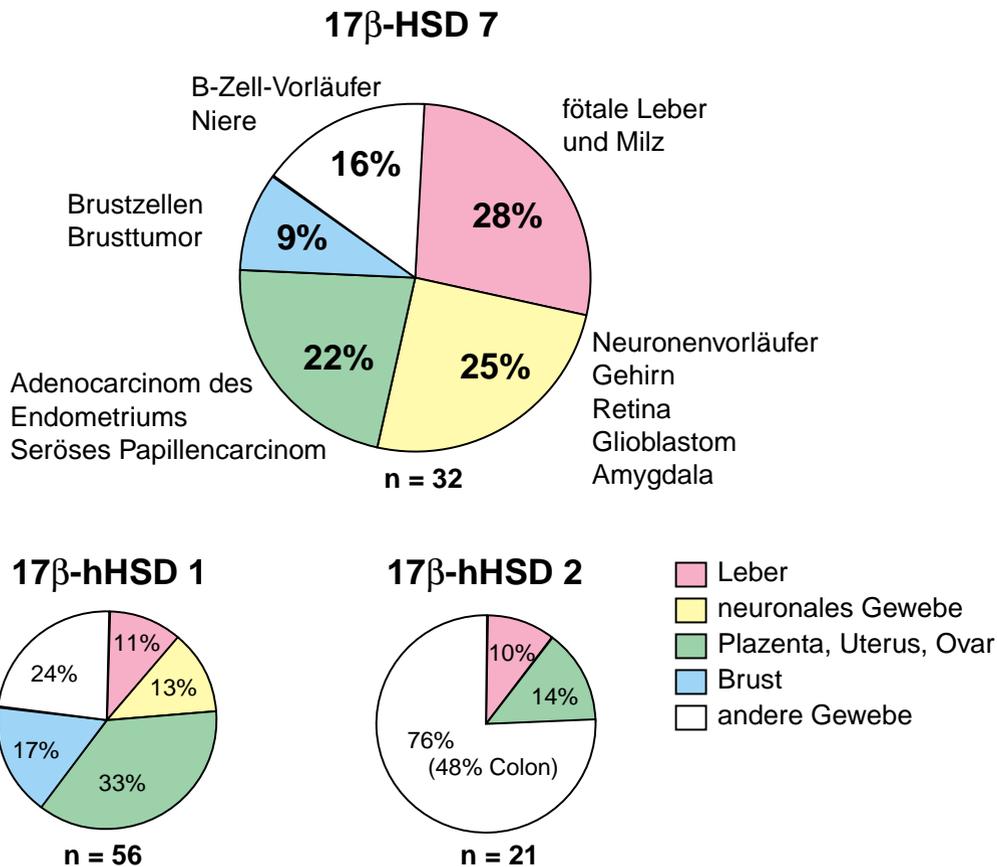


Abbildung 4.11: *in silico* Northern Blot von 17 β -HSD Typ 1, 2 und 7. Die Herkunft und relative Häufigkeit von ESTs der 17 β -HSD-Isoenzyme gibt Aufschluss über ihre Funktion. Die Expressionsmuster der östradiolproduzierenden 17 β -HSD1 und 7 sind sehr ähnlich, während die steroidinaktivierende 17 β -HSD2 ein ganz anderes Muster zeigt und überwiegend im Colon gefunden wird (vergleiche Mustonen et al., 1998). 17 β -HSD7 zeigt eine auffallend hohe Expression in Leber und neuralem Gewebe.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

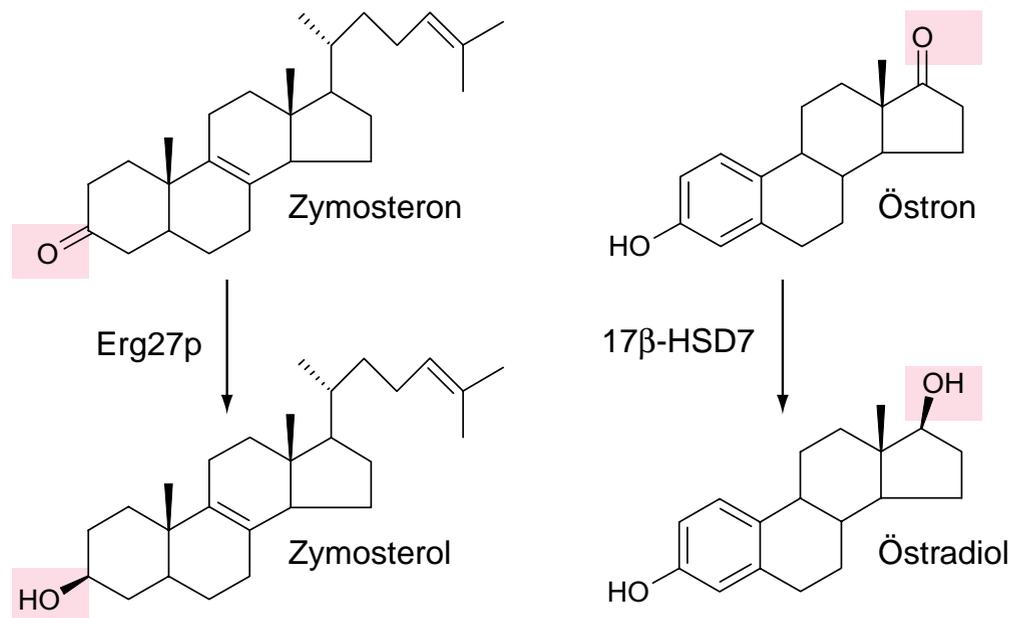


Abbildung 4.12: Reaktionen, die von 17 β -HSD7 und seinem Hefe-Homologen Erg27p katalysiert werden.

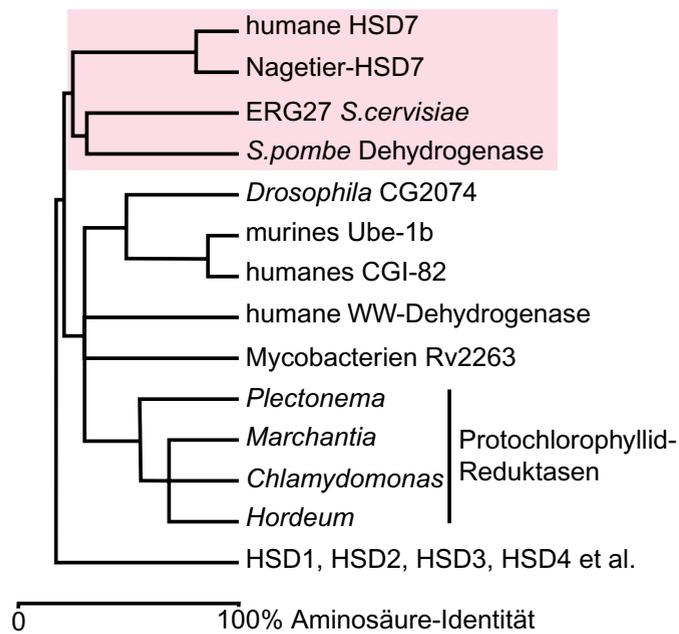


Abbildung 4.13: Stammbaum der 17 β -HSD7 und ihrer nächsten Verwandten. Die 17 β -HSD7 der Säugetiere bildet eine monophyletische Einheit mit den ERG27-Proteinen der Hefen. Entferntere Verwandte sind die Protochlorophyllid-Reduktasen der photosynthetischen Organismen und eine androgen-induzierte Dehydrogenase aus Testis (CGI-82).

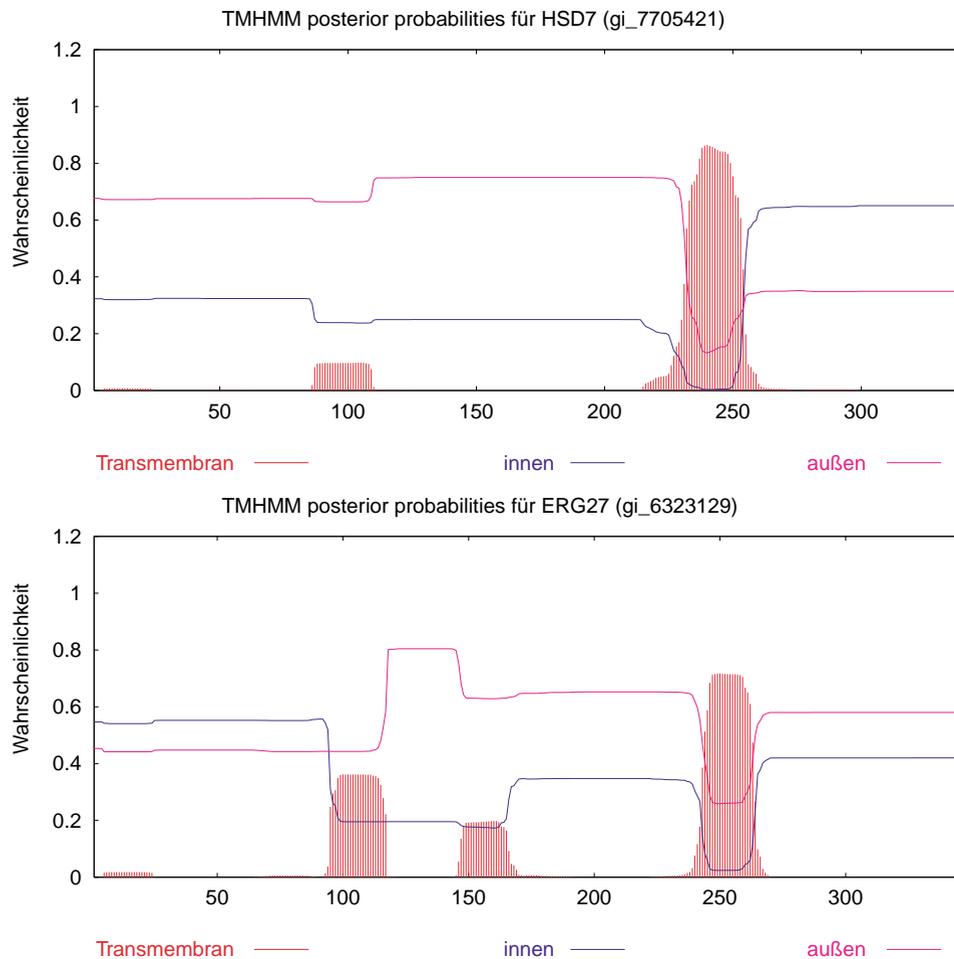


Abbildung 4.14: Vorhersage von Transmembranbereichen in humaner 17 β -HSD7 (oben) und *S. cerevisiae*-ERG27p (unten). Beide Proteine haben einen hydrophoben, helikalen Bereich um Aminosäure 250. Auch die konservierte Insertion (um Aminosäure 100) wird als hydrophober Bereich erkannt. Diese beiden Merkmale unterscheiden die 17 β -HSD und die ERG27-Proteine von allen anderen untersuchten Dehydrogenasen.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen



Abbildung 4.15: Alignment der Proteinsequenzen von 17 β -HSD Typ 1 und 7. Die konservierten Bereiche der HSD-Familie, sowie die charakteristischen Merkmale der 17 β -HSD7 (konservierte Insertion und hydrophobe Helix) sind markiert. Außerdem ist die vorhergesagte Sekundärstruktur der 17 β -HSD7 angegeben (α -Helices rosa, β -Faltblätter gelb), die genutzt wurden, um das Alignment der Proteine zu optimieren. Der C-Terminus der 17 β -HSD1 zeigt keine Homologie mit 17 β -HSD7 und wurde deshalb für das Alignment und die Modellierung nicht berücksichtigt.

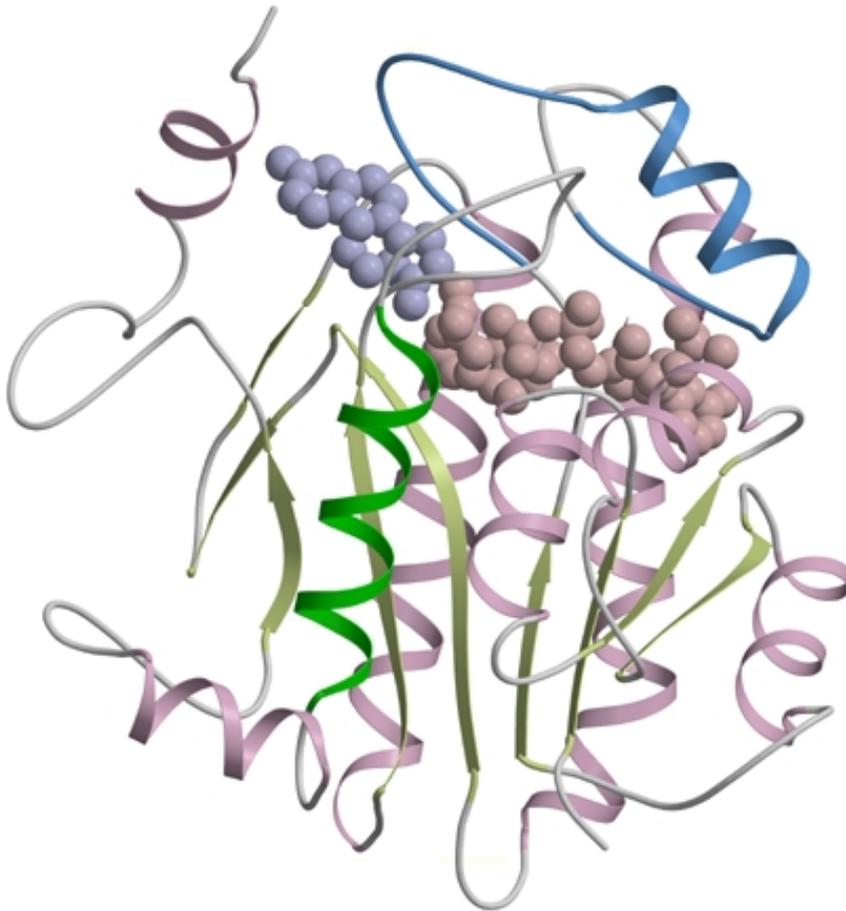


Abbildung 4.16: Homologie-basiertes Modell der Proteinstruktur von 17β -HSD7. Die hydrophobe Helix ist grün markiert, die konservierte 20 Aminosäuren-Insertion blau. Als Template diente die Röntgenstruktur der 17β -HSD1 (1FDS), der auch die Positionen des Kofaktors (rosa) und des Substrats (hellblau) entnommen wurden. Die hydrophobe Helix ist offensichtlich keine Transmembranelix, könnte jedoch die Assoziation an die Membran vermitteln. Die konservierte Insertion steht in Kontakt zum Substrat und scheint eine Substratbindungstasche zu bilden, die sich in Richtung der hydrophoben Helix öffnet.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

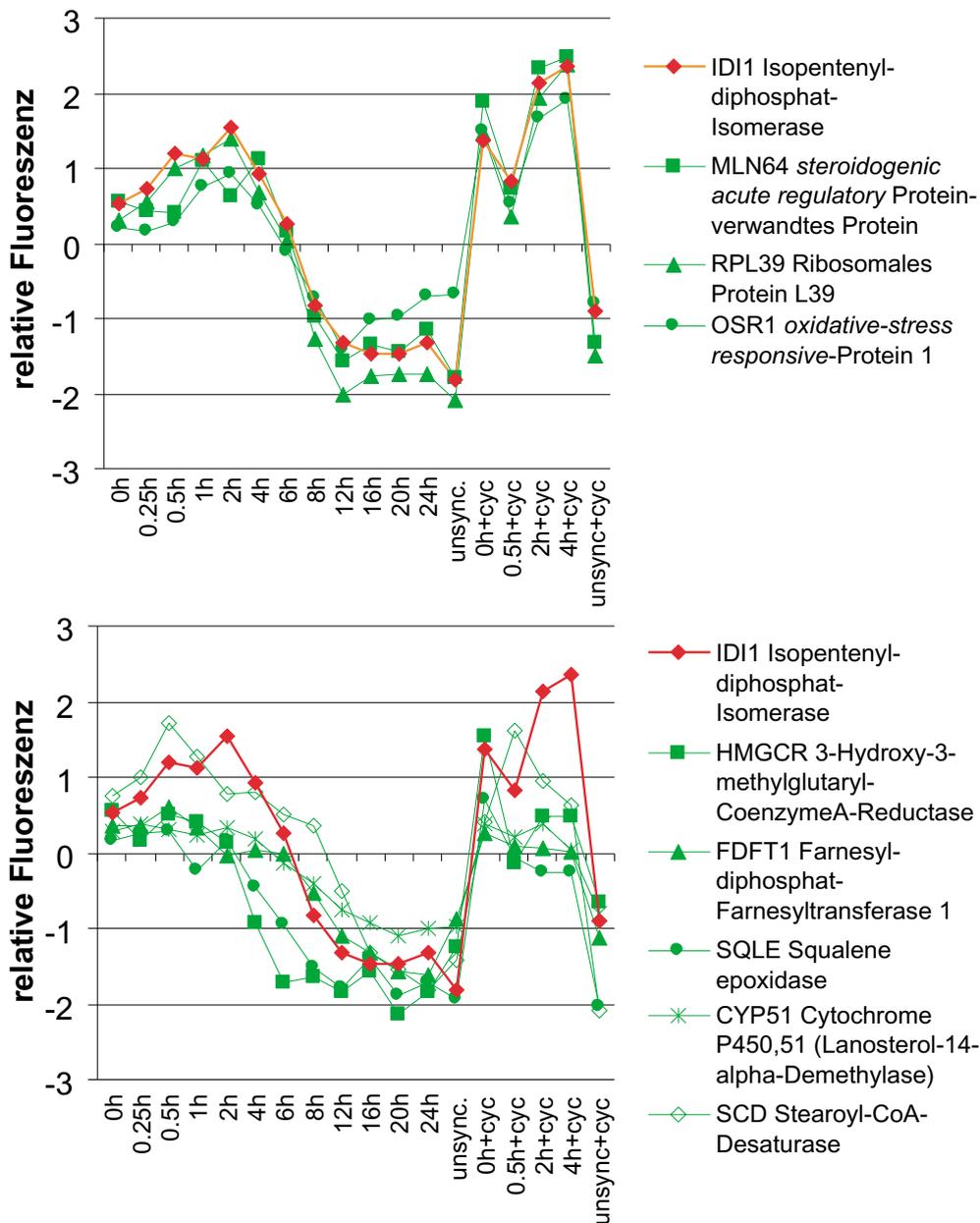


Abbildung 4.17: Expression von IDI1 in Fibroblasten nach Daten von Iyer et al. (1999). Außerdem sind die Expressionsmuster der koregulierten Proteine MLN64, OSR1 und RPL39 gezeigt, sowie von Proteinen des Cholesterol-Metabolismus. +cyc=Experimente mit Cycloheximid-Behandlung, unsync=unbehandelte Kontrollzellen (ohne Serumentzug). Die Zeitangaben beziehen sich auf den Zeitpunkt der Serumzugabe nach 3-tägigem „Aushungern“ der Zellen. Das obere Diagramm zeigt die strenge Koregulation von IDI1, MLN64, OSR1 und RPL39, das untere die weniger deutliche Koregulation der cholesterometabolisierenden Enzyme. Ein besonders auffälliger Unterschied ist der kurzfristige Anstieg der Expression von IDI1, MLN64, OSR1 und RPL39 während der frühen Phase der Serumantwort. Dieser Anstieg ist unter dem Einfluss von Cycloheximid (also in Abwesenheit von neuer Proteinsynthese) besonders ausgeprägt.

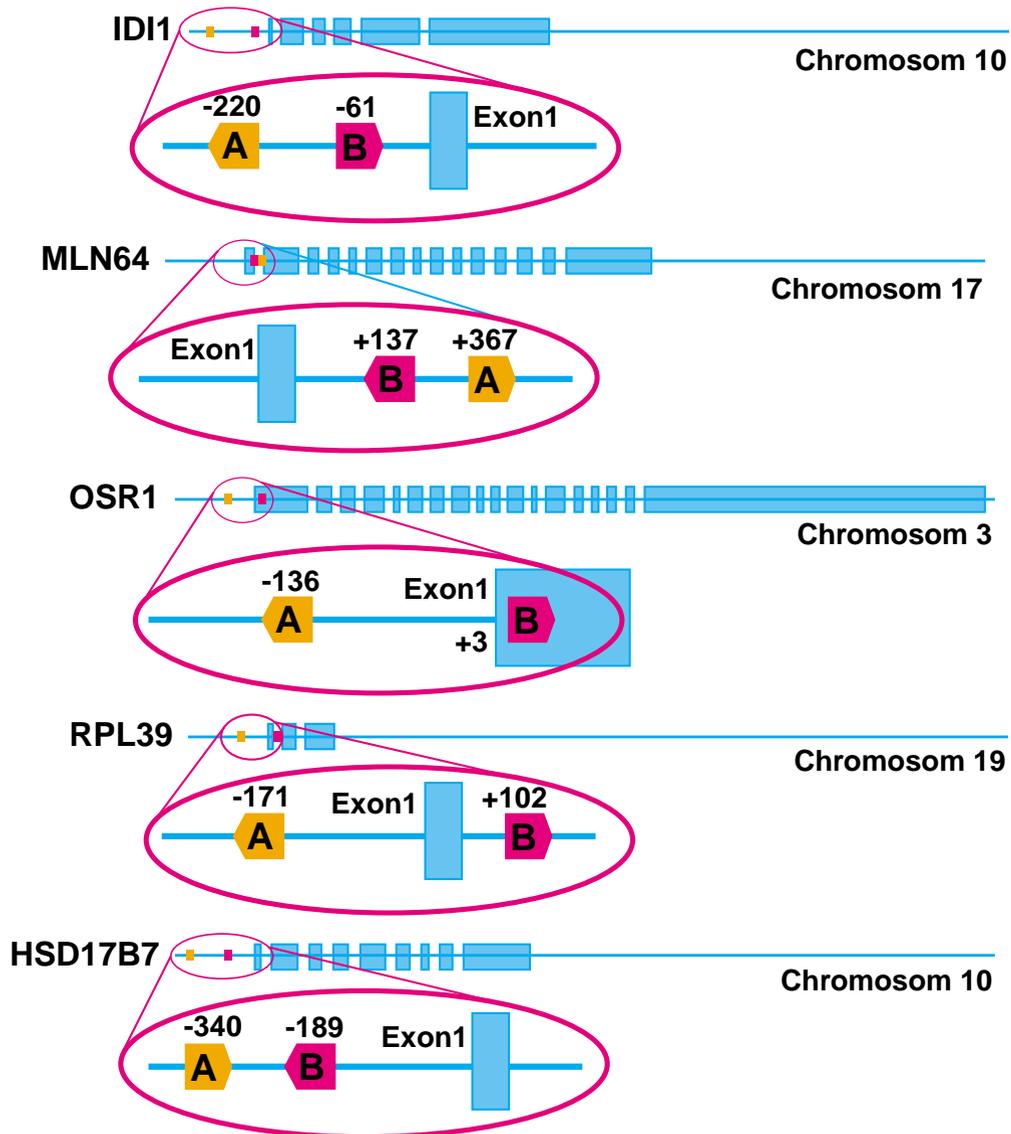


Abbildung 4.18: IDI1-Promotor-Modell und Genstrukturen für IDI1, MLN64, OSR1, RPL39 und 17 β -HSD7 (nach vorläufigen Ergebnissen des humanen Genomprojekts). Die dicken blauen Balken zeigen die Größe und Lage der Exons. Die Größe der Introns ist nicht maßstabsgerecht eingezeichnet. Die Lage der Teilmodule ist relativ zum Translationsstart angegeben (in Basenpaaren). Die Orientierung der Module ist durch die Pfeilspitze angezeigt.

4 Die 17β -Hydroxysteroid-Dehydrogenasen

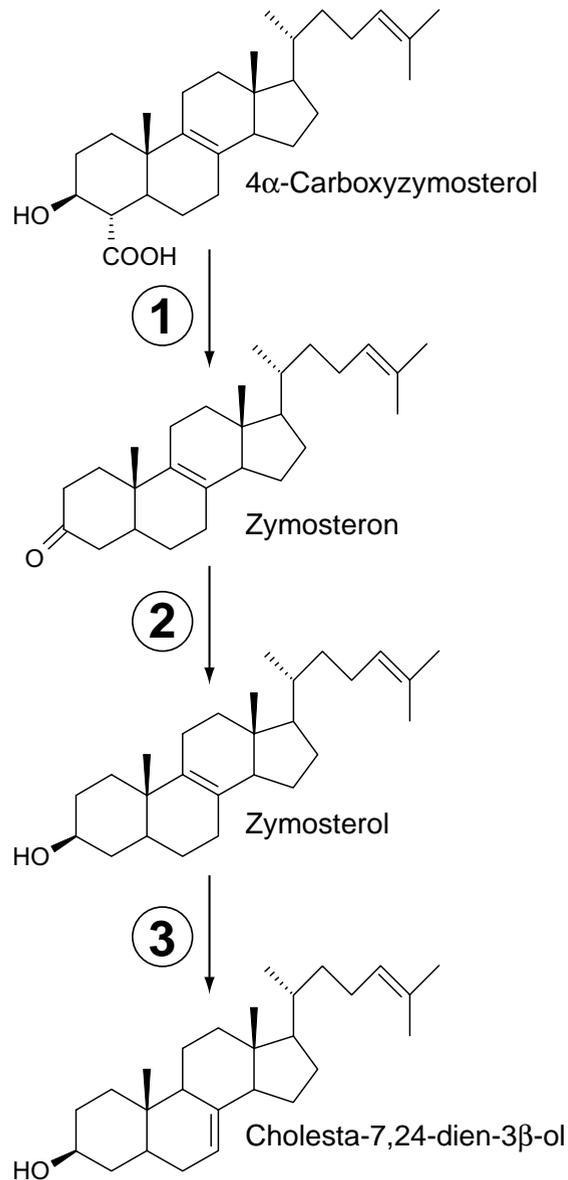


Abbildung 4.19: Reaktionen der Cholesterin-Biosynthese, die von (1) 3 β -Hydroxysteroiddehydrogenase, (2) 3-Ketosteroid-Reduktase und (3) Δ^8 - Δ^7 -Sterolisomerase katalysiert werden. Die Reaktionen (1) und (3) sind in CDPX2- bzw. CHILD-Patienten gestört. Reaktion (2) wird möglicherweise von 17 β -HSD7 katalysiert.

4.3 Die 17 β -HSD Typ 4 – Funktionsbestimmung durch genetische und phylogenetische Analyse²

4.3.1 Problemstellung

Im vorigen Abschnitt wurde festgestellt, dass die 17 β -HSD7 nicht primär für die Östradiolsynthese verantwortlich ist, sondern vermutlich für die Cholesterinbiosynthese. Einen ganz ähnlichen Fall findet man bei der 17 β -HSD4: Das Enzym ist nicht nur als Östradioldehydrogenase aktiv, sondern hat auch eine Schlüsselrolle in der peroxisomalen β -Oxidation langkettiger Fettsäuren. Es ist fraglich, welche physiologische Rolle die Inaktivierung von Östradiol durch 17 β -HSD4 beim Menschen tatsächlich spielt. Im Folgenden sollte durch Sequenzanalysen auf Protein- und DNA-Ebene die Evolution der Multifunktionalität von 17 β -HSD4 untersucht werden.

4.3.2 Hintergrund

Die 17 β -HSD4 wurde ursprünglich als Östradiol-Dehydrogenase aus dem Schweineuterus beschrieben (Adamski et al., 1992, Leenders et al., 1994). Orthologe Proteine wurden anschließend in vielen weiteren Wirbeltieren gefunden und kloniert (Mensch: Adamski et al., 1995; Maus: Normand et al., 1995; Ratte: Corton et al., 1996; Huhn: Kobayashi et al., 1997; Meerschweinchen: Caira et al., 1998). 17 β -HSD4 enthält neben der Dehydrogenase-Domäne zwei weitere Proteindomänen: Eine Hydratase- und eine Lipid-Carrier-Domäne. Hydratase und Dehydrogenase katalysieren aufeinanderfolgende Reaktionsschritte bei der peroxisomalen β -Oxidation von Fettsäuren (Adamski et al., 1997) (Abb. 4.20, S. 82). Die Lipid-Carrier-Domäne kommt nicht nur als Bestandteil der 17 β -HSD4 vor, sondern auch als Fusionsdomäne der peroxisomalen Thiolase (SCPx), von dem sie posttranslational abgespalten werden kann, und als Bestandteil nicht-peroxisomaler Proteine, z. B. des Stomatin-artigen Proteins (Seidel & Prohaska, 1998). Ihre Funktion ist unklar. Sie könnte am Transfer von Fettsäuren aus der peroxisomalen Membran beteiligt sein oder am Transport von einem katalytischen Zentrum zum nächsten (Seedorf et al., 2000). Es ist bekannt, dass die Lipid-Carrier-Domäne mit verschiedenen Komponenten des β -Oxidationswegs (Acyl-CoA-Oxidase, Bifunktionelles Protein, Thiolase: Wouters et al., 1998) interagiert und in ihrem Inneren einen hydrophoben Tunnel enthält, der für die Lipidbindung geeignet ist (Choinowski et al., 2000; Garcia et al., 2000). Der N-Terminus der Lipid-Carrier-Domäne von SCPx vermittelt die Bindung an Membranen (Huang et al., 1999).

Klonierung und Funktion der 17 β -HSD4

² Teilweise veröffentlicht in: Breitling R, Marijanović Z, Perović D & Adamski J (2000, im Druck): Evolution of 17 β -HSD type 4, a multifunctional protein of β -oxidation. Mol. Cell. Endocrinol.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

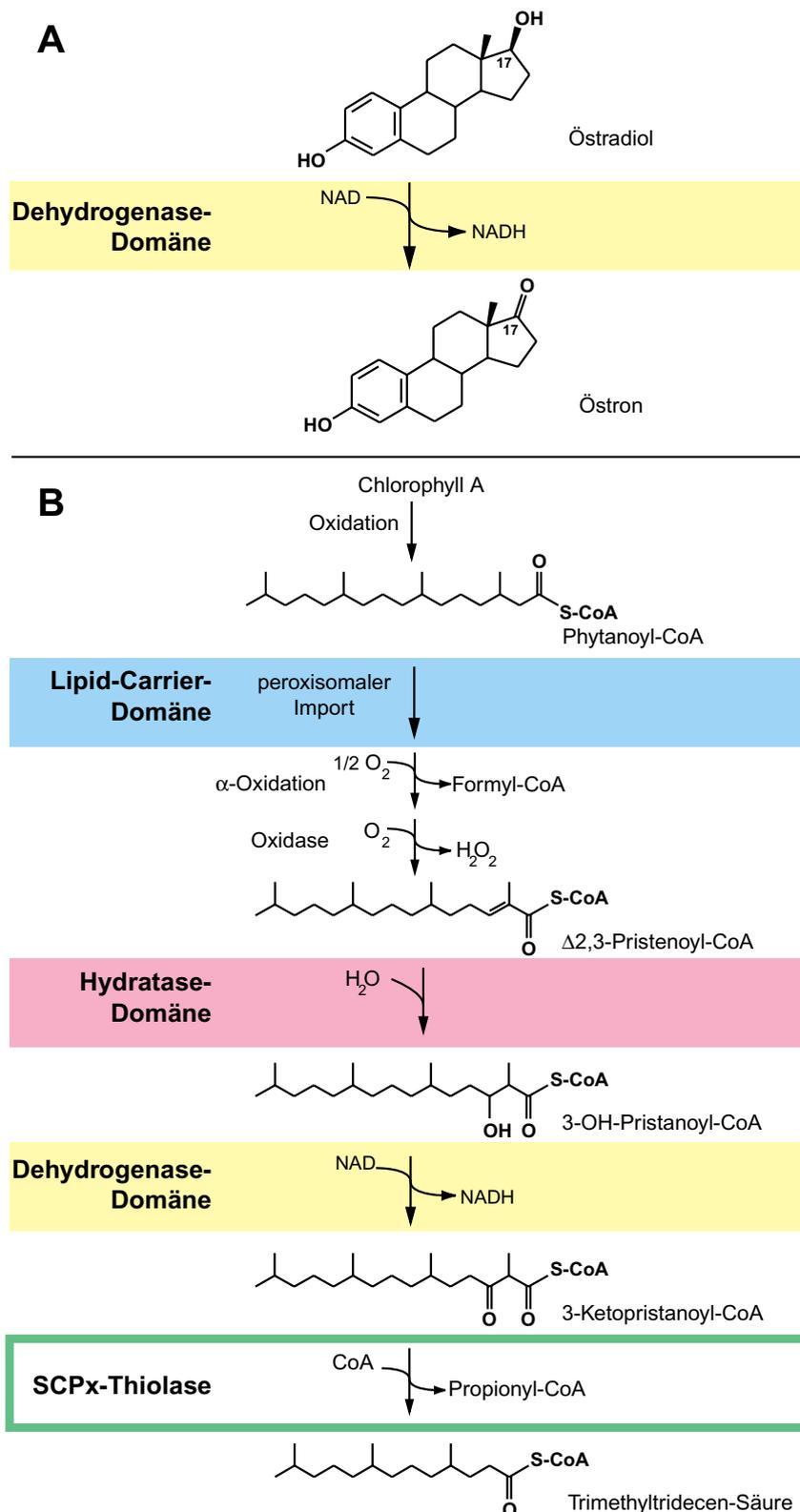


Abbildung 4.20: Reaktionen der peroxisomalen β -Oxidation. Die von 17 β -HSD4 katalysierten Schritte sind hervorgehoben. Außerdem ist die Thiolase-Reaktion markiert, die vom SCPx-Protein katalysiert wird, dessen SCP-Domäne der Lipid-Carrier-Domäne der 17 β -HSD4 homolog ist.

Die Aminosäure-Sequenz von 17 β -HSD4 zeigt keine auffällige Übereinstimmung mit den Sequenzen der anderen 17 β -HSDs, sondern ist vielmehr sehr ähnlich aufgebaut wie das Hefeprotein Fox2p. Dieses Protein ist ebenfalls ein multifunktionelles Protein der peroxisomalen β -Oxidation (Hiltunen et al., 1992). 17 β -HSD4 wurde aufgrund seiner Multifunktionalität auch unter anderen Namen beschrieben: Hydroxyacyl-CoA-Dehydrogenase/Hydratase, Multifunktionelles Protein 2 (MFP2) oder D-Bifunktionelles Protein (DBP) (Dieuaide-Noubhani et al., 1996; Jiang et al., 1996; Novikov et al., 1997). Bei den Untersuchungen, die zu diesen Bezeichnungen führten, stand die Beteiligung von 17 β -HSD4 an der peroxisomalen β -Oxidation von Pristansäure und Gallensäure-Intermediaten im Vordergrund. 17 β -HSD4 ist ein peroxisomales Protein, das fast ubiquitär exprimiert wird (de Launoit & Adamski, 1999).

Inaktivierende Mutationen des HSD17B4-Gens führen beim Menschen zu schweren Entwicklungsstörungen, ähnlich wie beim Zellweger-Syndrom (Suzuki et al., 1997; van Grunsven et al., 1998; van Grunsven et al., 1999). Bei der Maus führt eine Inaktivierung des 17 β -HSD4-Gens zu verringertem Wachstum und eingeschränkter Fertilität der Männchen (Baes et al., 2000). Die K.o.-Mäuse zeigen schwere Defekte in der peroxisomalen β -Oxidation und akkumulieren langkettige Fettsäuren und Gallensäure-Vorläufer. Die Inaktivierung der murinen 17 β -HSD4 ist jedoch nicht letal und der Phänotyp ist deutlich schwächer als bei den menschlichen Patienten. Neben ihrer Rolle in der β -Oxidation und bei der Inaktivierung von Steroidhormonen, wird der 17 β -HSD4 auch eine Beteiligung an der Neuronenwanderung im sich entwickelnden Gehirn zugeschrieben (Itoh et al., 1999).

Orthologe der der 17 β -HSD4 kommen auch in verschiedenen Hefen vor, wo sie ursprünglich als multifunktionelle Proteine der peroxisomalen β -Oxidation (Hydratase-Dehydrogenase-Epimerase) beschrieben wurden (*Candida tropicalis*: Nuttley et al., 1988; *Saccharomyces cerevisiae*: Hiltunen et al., 1992; *Neurospora crassa*: Fossa et al., 1995; *Yarrowia lipolytica*: Smith et al., 2000). Diese Proteine werden als Fox2p bezeichnet. *C. tropicalis*-Fox2p kann Östradiol umsetzen, aber mit deutlich geringerer Effizienz als seine Fettsäuresubstrate (Qin et al., 2000a). Ein weiteres Homologes wurde im arbuskulären Mykorrhizapilz *Glomus mosseae* gefunden, wo es an der Interaktion mit Rhizobakterien beteiligt ist (Requena et al., 1999). Der jüngste Zugang zur 17 β -HSD4/Fox2p-Gruppe ist das homologe Protein aus dem primitiven Eukaryonten *Dictyostelium discoideum* (M. Maeda, persönliche Mitteilung), das an der interzellulären Kommunikation dieses fakultativen Mehrzellers beteiligt zu sein scheint.

**Defekte der
17 β -HSD4**

**17 β -HSD4 in
primitiven
Eukaryonten**

4.3.3 Ergebnisse

Die Domänenzusammensetzung der 17 β -HSD4/Fox2p-Familie ist erstaunlich variabel: Fünf verschiedene Muster wurden identifiziert (Abb. 4.21, S. 84). Hefe-Fox2-Proteine haben eine duplizierte Dehydrogenase-Domäne, die an

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

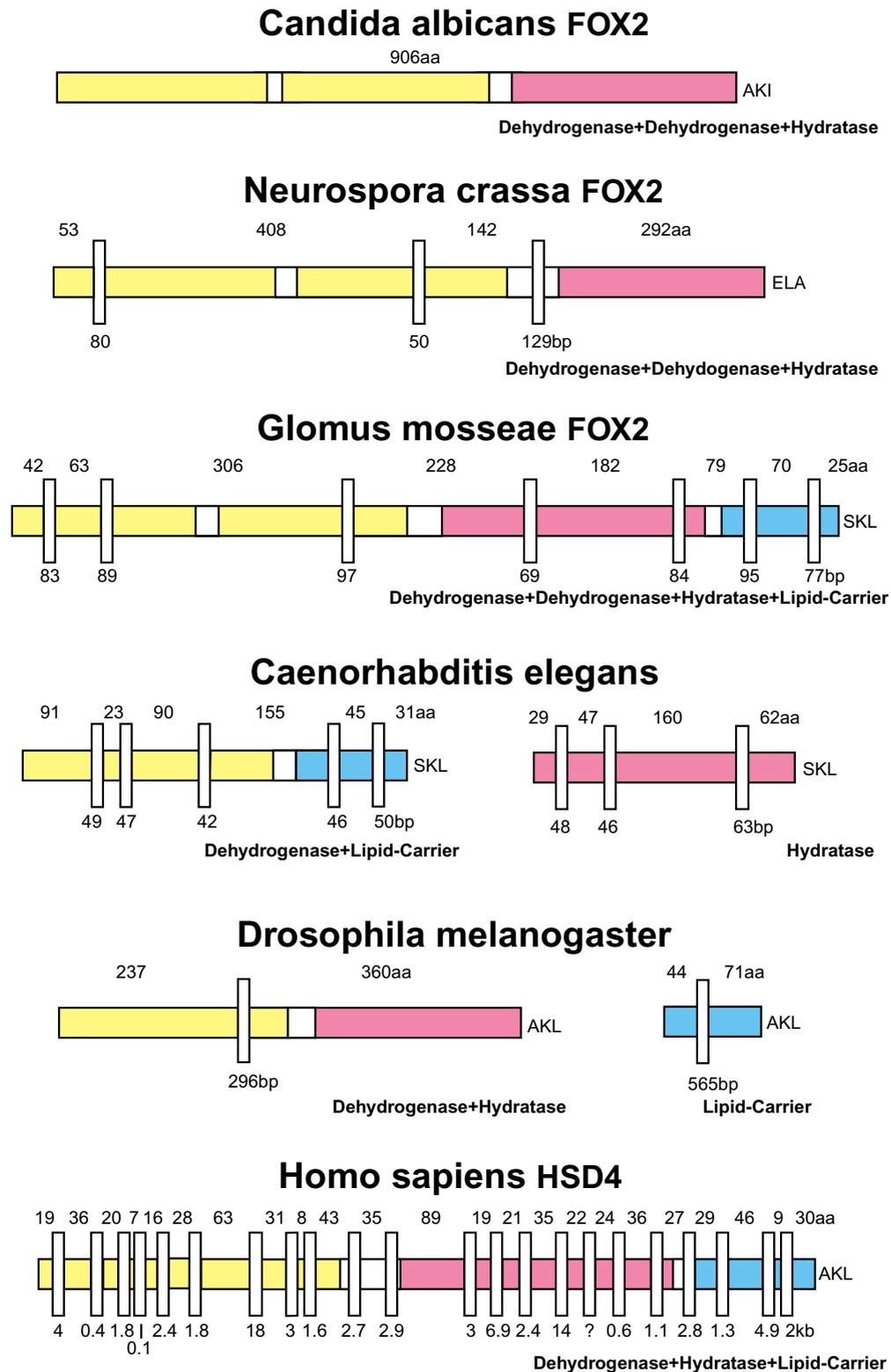


Abbildung 4.21: Genstrukturen von 17 β -HSD4 und ihren Orthologen. Die verschiedenen Domänen sind farblich gekennzeichnet (vgl. Abb. 4.20, S. 82). Die Größe der Introns (schmale weiße Balken) ist nicht maßstabsgetreu wiedergegeben. Die Sequenzen der Signalpeptide für den peroxisomalen Import sind an den C-Termini angegeben.

4.3 Die 17 β -HSD Typ 4

eine Hydratase-Domäne fusioniert ist. Bei *Glomus mosseae* ist zusätzlich eine Lipid-Carrier-Domäne angehängt. Die Proteine der Vertebraten enthalten nur eine einzige Dehydrogenase-Domäne, die wie bei *G. mosseae* mit einer Hydratase- und einer Lipid-Carrier-Domäne fusioniert ist. Bei *C. elegans* und *Drosophila* sind die Domänen umgeordnet: Entweder die Hydratase (bei *C. elegans*) oder die Lipid-Carrier-Domäne (bei *Drosophila*) liegen als getrennte Polypeptide vor. In diesen Fällen haben beide Teil-Proteine ein peroxisomales Targeting-Signal am C-Terminus. Beim 17 β -HSD4-Homologen in *Dictyostelium* fehlt wie bei *C. elegans* die Hydratase-Domäne. Ob sie als eigenes Polypeptid vorkommt, ist aber nicht bekannt.

Für alle drei Domänen wurden unabhängig voneinander Protein-Stammbäume berechnet. Diese Untersuchung sollte zeigen, in welcher Reihenfolge die Domänen fusioniert und aufgeteilt wurden (Abb. 4.22, S. 86). Protein-Stammbäume, Domänenzusammensetzung und allgemeine Kenntnisse der Verwandtschaft der Organismen wurden zu einem Evolutions-Szenario kombiniert (Abb. 4.23, S. 89). Dieses Szenario zeigt verschiedene Schritte bei der Entstehung der Multifunktionalität. Ursprünglich lagen alle Domänen als eigene Proteine vor – eine Konfiguration die sich bei Mycobakterien bis heute erhalten hat. Bei Eukaryonten kam es dagegen frühzeitig zu Fusionen der drei Domänen. In Hefen wurde außerdem vor der Aufspaltung in Ascomycota und Zygomycota (vor 600×10^6 Jahren, Redecker et al., 2000) die Dehydrogenase-Domäne dupliziert. Dass es sich tatsächlich um eine einmalige Verdopplung handelt und nicht um zwei unabhängige Ereignisse, zeigt der Stammbaum, in dem die C-terminale und die N-terminale Kopie unabhängige Cluster bilden. Der Duplikation folgte eine schnelle Diversifizierung der beiden Kopien. Während die C-terminale Kopie ihre Präferenz für langkettige Substrate beibehielt (wie die Wirbeltier-Homologen), spezialisierte sich die zweite Kopie auf kurzkettige Fettsäuren (Qin et al., 1999). Nur bei *Glomus mosseae*, wo die C-terminale Kopie durch eine Mutation im aktiven Zentrum inaktiviert ist, ist die zweite Kopie höher konserviert. Das äußert sich in der basaleren Stellung im Protein-Stammbaum. In diesem Fall enthält die zweite Domäne die charakteristische Insertion, die die peroxisomalen Dehydrogenasen kennzeichnet (Qin et al., 2000a). Diese Insertion bildet eine Schleife in der Proteinstruktur und ermöglicht vermutlich die Bindung von Methyl-verzweigten Fettsäuren. In der zweiten Dehydrogenase-Domäne der anderen Hefen fehlt diese Schleife. Den Fusionsschritten folgte in mehreren Fällen eine sekundäre Aufteilung in separate Polypeptide, so dass es zu neuen Kombinationen der Domänen kam. Das es sich um sekundäre Umordnungen und nicht um unabhängige Fusionen handelt, zeigt die Verwandtschaft aller beteiligten Domänen zu den jeweils homologen Domänen der kompletten Multifunktions-Proteine.

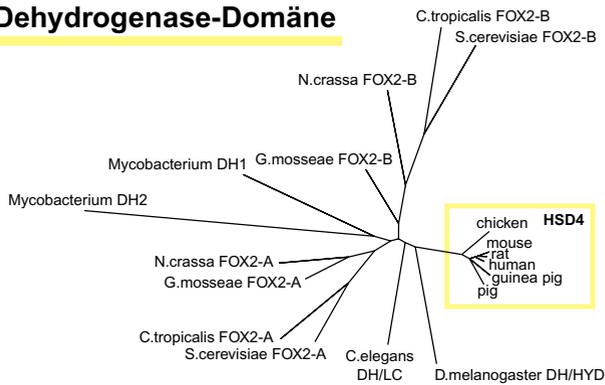
Die Lipid-Carrier-Domäne ist bei diesen Umordnungen der mobilste Teil der 17 β -HSD4-Verwandten. Die Lipid-Carrier-Domäne der 17 β -HSD4-Proteine ist bei Wirbeltieren viel weniger konserviert als die Domäne der Thiolase (SCPx/SCP2). Möglicherweise ist die Lipid-Carrier-Funktion in den 17 β -

**Domänen-
Stammbäume**

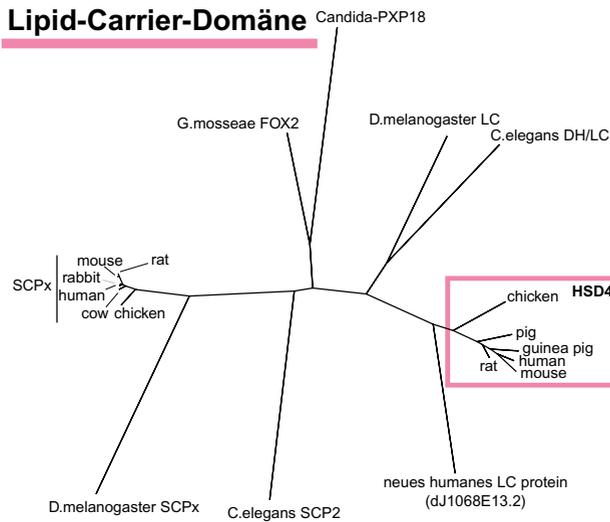
**Die Lipid-Carrier-
Domäne**

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

Dehydrogenase-Domäne



Lipid-Carrier-Domäne



Hydratase-Domäne

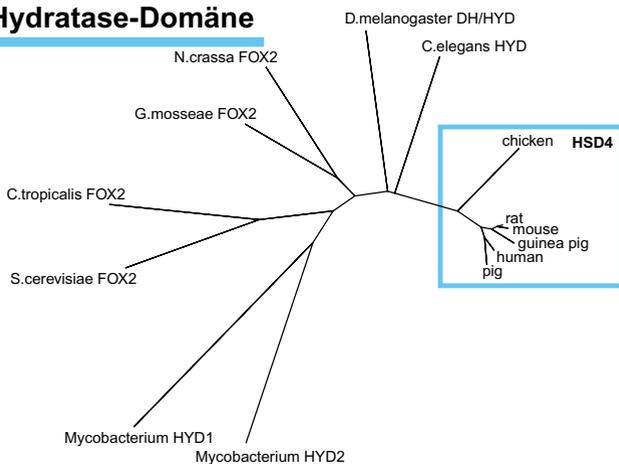


Abbildung 4.22: Stammbäume der Proteindomänen von 17 β -HSD4-Verwandten. **A** Dehydrogenase-Domäne. **B** Hydratase-Domäne. **C** Lipid-Carrier-Domäne. Die Stammbäume wurden nach der Neighbour-Joining-Methode berechnet. Parsimony-Stammbäume zeigen fast die gleiche Topologie.

HSD4-Proteinen redundant und unterliegt nicht der gleichen, streng konservativen Selektion wie bei SCPx/SCP2. Eine kürzlich beim Menschen beschriebene, separate Lipid-Carrier-Domäne (GenBank-Eintrag dJ1068E13.2 auf Chromosom 20p11.21-12.3) stammt wohl aus der 17 β -HSD4, denn sie ist der 17 β -HSD4-Domäne deutlich ähnlicher als der SCPx-Domäne. Die Lipid-Carrier-Domäne in *Glomus mosseae* zeigt größere Ähnlichkeit zum peroxisomalen Lipid-Carrier Pxp18p der Hefen, als zu den Lipid-Carrier-Domänen von SCPx oder 17 β -HSD4. Das spricht dafür, dass diese Fusion unabhängig von der Fusion in den Metazoen erfolgte. Das bedeutet aber auch, dass die Fox2-Proteine der Ascomycota vermutlich nie eine Lipid-Carrier-Domäne enthalten haben.

Die separate Hydratase von *C. elegans* enthält die konservierten katalytischen Aminosäurereste (entsprechend Glutamat 336 und Aspartat 510 des humanen Proteins, Qin et al., 2000b). Auch der Rest der Sequenz ist deutlich kon-

serviert. Sie enthält außerdem ein funktionelles Peroxisomen-Targeting-Signal. Das spricht dafür, dass die *C. elegans*-Hydratase ein katalytisch aktives Protein ist.

Bei Datenbanksuchen wurden die genomischen Sequenzen von sechs Mitgliedern der 17 β HSD4/Fox2p-Familie identifiziert. Durch Vergleich mit den cDNA-Sequenzen wurden die Exon/Intron-Grenzen in den sechs Genen bestimmt. *C. albicans*- und *S. cerevisiae*-Fox2 haben keine Introns, die anderen vier Gene haben zwischen zwei (*Drosophila*) und 23 (Mensch) Introns. Die Größen der Introns beim Menschen sind sehr unterschiedlich, liegen aber in der Größenordnung von einigen Tausend Basenpaaren pro Intron. In den anderen Organismen ist die Introngröße stark reduziert und liegt im Allgemeinen unter 100 Basenpaaren. Die Lage der Exon-Intron-Grenzen ist sehr variabel. Sie liegen in allen untersuchten Arten an verschiedenen Stellen. Durch Vergleiche auf Proteinebene wurden die Grenzen der Proteindomänen bestimmt: Funktionelle Module (Domänen) sind durch weniger konservierte Verbindungsstücke getrennt. Überraschenderweise liegen keine konservierten Introns an den Domänengrenzen, obwohl die evolutionäre Variabilität der Domänenzusammensetzung diese Vermutung nahegelegt hatte.

Bestimmung der Gen-Strukturen

4.3.4 Diskussion

Die Evolution der Multifunktionalität von 17 β -HSD4 zeigt eine erstaunliche Flexibilität der Proteinzusammensetzung. Trotz der hohen funktionellen Konservierung treten regelmäßig Fusions-, Aufspaltungs- und Umordnungsprozesse auf. Diese ständige Neuzusammenstellung der drei Domänen ohne Beeinträchtigung der Funktion zeigt, dass die Multifunktionalität keine physiologische Notwendigkeit ist. Vermutlich können die Domänen auch als getrennte Polypeptide interagieren. Die Fusion hat dann vor allem den Zweck, eine gemeinsame Regulation zu ermöglichen. Diese Art der Koregulation ist aber nur dann sinnvoll, wenn die Proteine keine zusätzlichen Funktionen in anderen Stoffwechselwegen erfüllen müssen. Diese Überlegung macht es unwahrscheinlich, dass 17 β -HSD4 eine physiologisch bedeutsame Funktion als Östradiol-Dehydrogenase entwickeln konnte.

Flexible Domänen-Struktur

Die variable Lage der Exon/Intron-Grenzen korreliert nicht mit der Position der Interdomänenbereiche. Diese Beobachtung ist unerwartet, denn es wurde angenommen, dass die hochvariable Zusammensetzung der Domänen, die in mehreren unabhängigen Ereignissen fusioniert, auseinandergeteilt und wieder neu kombiniert wurden, durch konservierte Introns an den Domänengrenzen ermöglicht würde. Für das Fehlen solcher Interdomänen-Introns gibt es zwei Erklärungen: Entweder sind die Introns nach den Umordnungsereignissen verloren gegangen, so dass die heute vorkommenden Proteine ein fixiertes Abbild des ursprünglichen dynamischen Prozesses bieten. Oder die Neukombination der Proteindomänen verläuft nach einem unbekanntem Alternativ-Mechanismus, der auch ohne Introns eine ausreichende Wahrscheinlichkeit für

Exon-Intron-Grenzen

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

funktionelle Neukombinationen bietet.

Funktion des humanen Enzyms

Die Konservierung der kompletten Multidomänen-Struktur beim Menschen (und anderen Vertebraten) zeigt, dass auch hier die gemeinsame Aktivität der drei Domänen die Hauptfunktion darstellt: Humane 17 β -HSD4 ist also in erster Linie ein Enzym der peroxisomalen β -Oxidation. Die Inaktivierung von Östradiol spielt dagegen *in vivo* nur eine untergeordnete Rolle. Das erklärt die Beobachtung, dass Östradiol-Dehydrogenase-Aktivität und 17 β -HSD4-Expression in Zelllinien und Gewebeproben nicht korrelieren (Miittinen et al., 1999). Auch die ubiquitäre Expression der 17 β -HSD4 und der Phänotyp der 17 β -HSD4-K.-o.-Mäuse können auf diese Weise erklärt werden. Die östradiol-katabolisierende Wirkung der 17 β -HSD4 ist also nur eine sekundäre Eigenschaft, vermutlich ein Ergebnis der verbreiterten Substratspezifität der Wirbeltier-17 β -HSD4, die neben Fettsäuren auch Gallensäure-Vorläufer (also Steroide) metabolisiert.

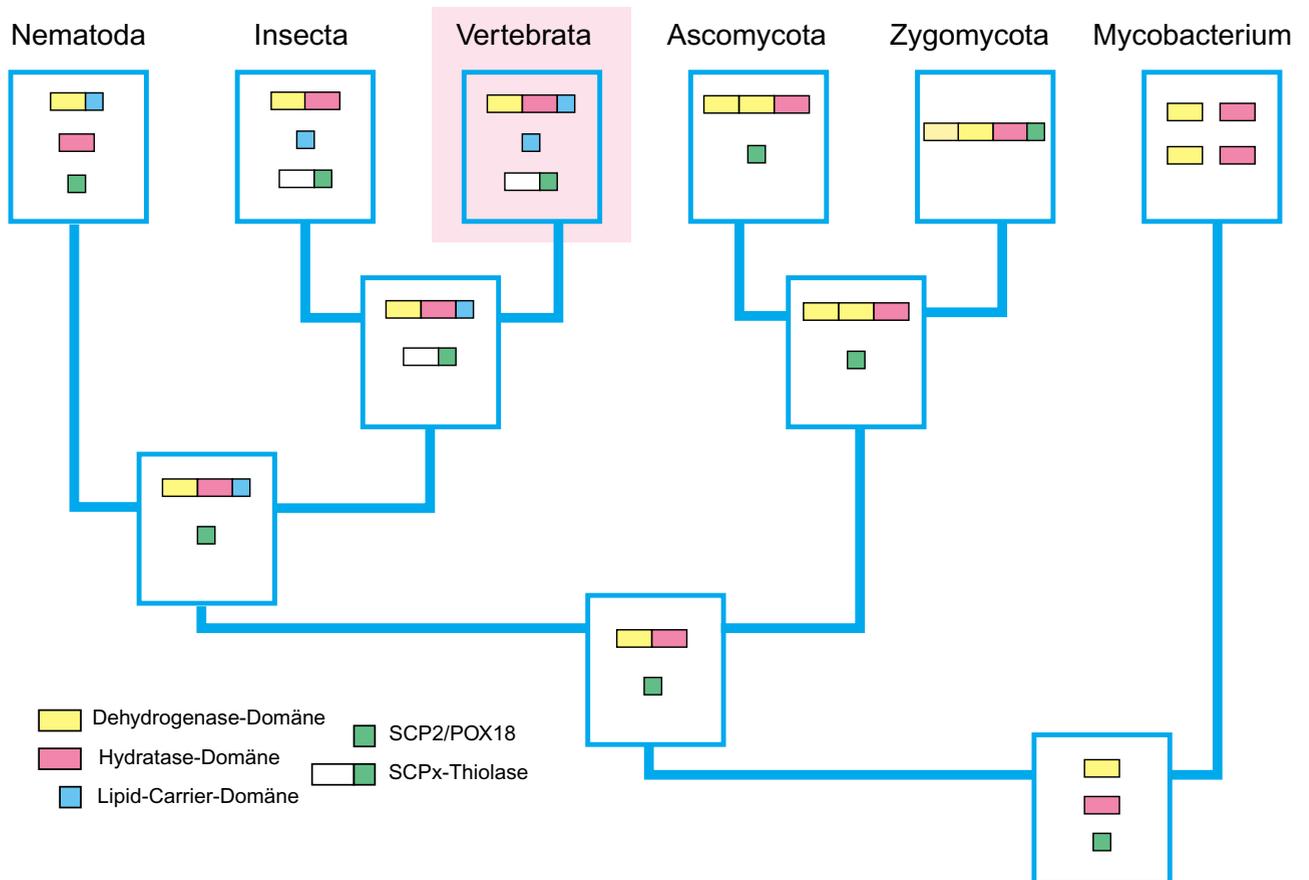


Abbildung 4.23: Evolutionsszenario der 17 β -HSD4-Verwandten. Die Topologie des Stammbaums für die Eukaryonten entspricht den phylogenetischen Verhältnissen, die sich aus molekularen und morphologischen Analysen ableiten lassen (Baldauf et al., 2000). An den Verzweigungspunkten dieses Stammbaums ist die rekonstruierte Domänenzusammensetzung bei den hypothetischen Vorläuferorganismen angegeben, die sich aus den Einzelstammbäumen der Domänen als sparsamste Hypothese ergibt. Mycobakterien (ganz rechts) besitzen kein Homologes der Lipid-Carrier-Domäne.

4.4 Die 17 β -HSD Typ 1 – interspezifischer Vergleich der genomischen Struktur

4.4.1 Problemstellung

Bei der Untersuchung der 17 β -HSD7 wurde deutlich, dass dieses Enzym möglicherweise eine Funktion bei der Cholesterinbiosynthese besitzt. Andererseits wurde aber berichtet, dass 17 β -HSD7 und 17 β -HSD1 zumindest bei Nagetieren gemeinsam für die Östradiolbiosynthese verantwortlich sind (Nokelainen et al., 2000) und sich dabei funktionell ergänzen. Ein solches Zusammenwirken der beiden östrogenen 17 β -HSDs ist beim Menschen jedoch – wie oben (Kap. 4.2, S. 59) diskutiert – unwahrscheinlich. Es ist also möglich, dass sich die Aufgaben von 17 β -HSD7 und 17 β -HSD1 in Menschen und Nagetieren deutlich unterscheiden. Als neuer Schritt zur genaueren Analyse der physiologischen Funktion von 17 β -HSD1 in Menschen und Nagetieren wurde eine vergleichende Analyse des 17 β -HSD1-Gens in Mäusen, Ratten und Menschen durchgeführt. Die bisherigen Untersuchungen betrafen seit langem getrennte Organismengruppen und konnten sich deshalb auf Vergleiche von hochkonservierten Proteinsequenzen beschränken. Für die 17 β -HSD1 sollten dagegen Entwicklungen untersucht werden, die vor relativ kurzer Zeit (nämlich innerhalb der Evolution der Säugetiere) stattgefunden haben. Zu diesem Zweck ist der Vergleich auf DNA-Ebene, insbesondere in den schnell evolvierenden intronischen Sequenzen, vorzuziehen. Aufgabe war es, die Evolution der drei 17 β -HSD1-Gene zu beschreiben, ihre Entstehung in einem entwicklungsgeschichtlichen Zusammenhang zu datieren und, wenn möglich, charakteristische Unterschiede zu definieren, die den endokrinologischen Differenzen von Menschen und Nagetieren zugrundeliegen könnten.

4.4.2 Hintergrund

Gen-Strukturen

Die 17 β -HSD1-Gene in Mensch, Maus und Ratte haben eine kompakte Struktur aus sechs Exons und fünf Introns, mit einer Gesamtlänge von etwa 2.2 kb. Das menschliche Gen ist in Tandem-Anordnung dupliziert, wobei der Abstand zwischen den beiden Kopien etwa 3 kb beträgt (Luu-The et al., 1990). Das Gen der Ratte (Akinola et al., 1998) und der Maus (Gloeckner et al., unveröffentlicht) liegen dagegen als Einzelkopien vor.

Klonierung und Funktion

Beim Menschen ist 17 β -HSD1 das vorherrschende östradiol-produzierende Enzym in Ovar, Brust und Placenta. Aus Placenta-Gewebe wurde sie 1988 auch zuerst kloniert (Peltoketo et al., 1988; Luu-The et al., 1989). In der Placenta von Nagetieren wird 17 β -HSD1 dagegen nicht exprimiert (Akinola et al., 1997; Nokelainen et al., 1996). Auch im Ovar der Nagetiere beschränkt sich die Expression von 17 β -HSD1 auf die Granulosa-Zellen des sich entwickelnden Follikels und hört während der Luteinisierung auf, während die ovarielle Expression beim Menschen bis in die Lutealphase erhalten bleibt (Zhang et al.,

1996). Diese Unterschiede sind besonders von Interesse, weil sich das Nagetierovar während der Schwangerschaft zum Hauptproduzenten von Östradiol entwickelt, wohingegen diese Aufgabe beim Menschen von der Placenta übernommen wird (Peltoketo et al., 1999a). Die ovarielle Östradiolproduktion muss also bei Nagetieren von einem zweiten Enzym übernommen werden, während der endokrinologische Befund beim Menschen allein durch die Aktivitätsverteilung von 17 β -HSD1 erklärt werden kann.

Die regulatorischen Elemente des humanen Gens sind genauer untersucht worden (Peltoketo et al., 1999a). Sie bestehen aus einem Kernpromotor (Position -78 bis $+9$ relativ zum Transkriptionsstart), einem Silencer (-391 bis -79) und einem Enhancer (-661 bis -392). Der Kernpromotor enthält eine klassische TATA-Box, einen GC-reichen Abschnitt und AP-2-, Sp1- und Sp3-Bindungsstellen, deren Wechselspiel mit der Regulation der 17 β -HSD1-Expression im Syncytiotrophoblast in Zusammenhang gebracht wird. Im Silencer-Bereich wurden GATA-Bindungsstellen nachgewiesen, die für die Beschränkung der 17 β -HSD1-Expression in der Placenta verantwortlich gemacht werden. Der Enhancer wirkt zellspezifisch und ist in Choriokarzinoma und 17 β -HSD1-positiven Brustkrebszellen aktiv (Piao et al., 1995). Er scheint aus mehreren essentiellen Untereinheiten zu bestehen, unter anderem einem Retinsäure-responsivem Element, das für die Retinsäure-abhängige Stimulation der 17 β -HSD1-Expression in verschiedenen Zelllinien verantwortlich sein könnte. Wie die verschiedenen Elemente zusammenwirken, um eine zeitlich und räumlich genau regulierte Expression von 17 β -HSD1 zu bewirken, ist unbekannt.

**Der
HSD17B1-Promotor**

4.4.3 Ergebnisse und Diskussion

Der Vergleich des Ratten- und Maus-Gens der 17 β -HSD1 zeigt eine Gesamtähnlichkeit der beiden Strukturen von etwa 89% (Abb. 4.24, S. 92). Die Ähnlichkeit der Exonsequenzen ist mit 92% deutlich höher als die der Intronsequenzen (83%). Dieser Unterschied wird für ein proteinkodierendes Gen erwartet. Beim Vergleich der beiden humanen Genkopien fällt dagegen auf, dass die Ähnlichkeit sowohl in Exon- wie in Intronbereichen gleichmäßig 89% beträgt; ein Hinweis darauf, dass es sich bei der ersten Kopie um ein nicht-translatiertes (und vermutlich nicht transkribiertes) Pseudogen handelt, und nicht um eine Androgen-spezifische Isoform, wie ursprünglich vermutet (Luu-The et al., 1990).

**Interspezifischer
Gen-Vergleich**

Die Sequenz-Ähnlichkeit zwischen Mensch- und Nagetier-Genen beträgt nur etwa 66% (Exons 71%, Introns 60%), die Genstruktur ist dagegen hochkonserviert. Unter der Voraussetzung, dass die „molekulare Uhr“ sowohl in Menschen wie auch in Nagetieren regelmäßig „tickt“, also konservierte Punktmutationen mit einer gleichmäßigen Rate auftreten, und mit der zusätzlichen Annahme, dass Nukleotidaustausche bei Nagetieren wegen der kürzeren Generationsdauer etwa dreimal schneller akkumulieren (Wu & Li, 1985), können

**Datierung der
humanen
Duplikation**

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

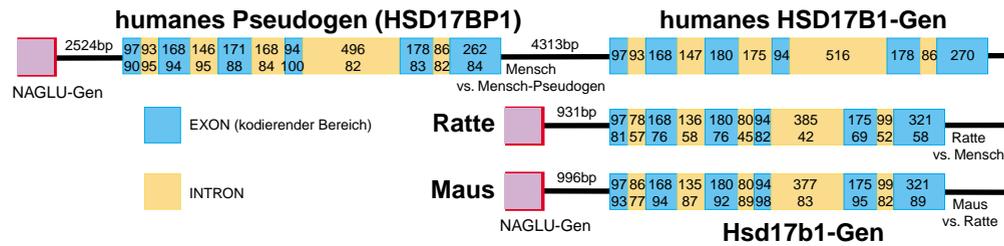


Abbildung 4.24: Vergleich der Genstrukturen der 17 β -HSD1 in Menschen und Nagetieren. Die obere Zahl gibt jeweils die Länge des Exons bzw. Introns an, die untere die Ähnlichkeit in Prozent für den jeweils rechts angegebenen Vergleich. Die Länge der intergenischen Regionen ist nicht maßstabsgetreu wiedergegeben. NAGLU = N-Acetyl-Glucosaminidase, HSD17B1 und HSD17BP1 = humanes Gen und Pseudogen der 17 β -HSD1, Hsd17b1 = 17 β -HSD1-Gen der Nagetiere.

diese Informationen verwendet werden, um den ungefähren Zeitpunkt der Gen-duplikation beim Menschen abzuschätzen:

1. Die Trennung von Mäusen und Ratten erfolgte vor etwa $20\text{--}29 \times 10^6$ Jahren (O'hUigin & Li, 1992).
 2. Die Zahl der Nukleotidaustausche in den intronischen Sequenzen beträgt etwa 17% zwischen Maus und Ratte und etwa 11% zwischen den beiden humanen Kopien (entsprechend $3 \times 11\% = 33\%$ „Nagetieräquivalenten“).
- ⇒ Die Duplikation ist also etwa $33\%/17\% = 1.9$ mal früher aufgetreten als die Divergenz der beiden Nagetierlinien, d.h. vor ungefähr $39\text{--}56 \times 10^6$ Jahren.

Dieser Zeitpunkt liegt überraschend früh, ganz am Anfang der Primatenevolution, noch vor der Divergenz von Alt- und Neuweltaffen vor etwa $25\text{--}34 \times 10^6$ Jahren (Sibley & Ahlquist, 1987). Für diese Datierung spricht außerdem, dass die beiden Kopien von mitduplizierten Alu-Repeats flankiert werden. Die Ausbreitung von Alu-Repeats in den Primaten-Genomen fand vor etwa $30\text{--}50 \times 10^6$ Jahren statt, und nur wenige Insertionsereignisse werden später datiert (Britten, 1994). Es ist aber nicht klar, ob eine Alu-Element-Retrotransposition an der Duplikation beteiligt war. Ein solcher Mechanismus ist jedoch eher unwahrscheinlich, denn die genomische Lokalisation hat sich durch die Duplikation im Vergleich zu den Nagetieren nicht geändert – in beiden Fällen liegt das HSD17B1-Gen in unmittelbarer Nachbarschaft zum α -N-Acetylglucosaminidase (NAGLU)-Gen – während eine Transposon-vermittelte Tandemduplikation ein intermolekulares Ereignis ist, also zur Integration an einem neuen Gen-Ort führen würde. Integrationen und Duplikationen sind im

4.4 Die 17β -HSD Typ 1

Gegensatz zu den meisten Punktmutationen irreversibel und deshalb ein entscheidender Schrittmacher der Primatenevolution (Hamdi et al., 1999). Erfolgt eine solche Duplikationen intermolekular, so ist es z. B. möglich, dass ganze Transkriptionseinheiten unter die Kontrolle eines neuen Promotors geraten. Das scheint aber bei der 17β -HSD1-Duplikation nicht der Fall zu sein.

Das zellspezifische Enhancer-Element der 17β -HSD1 ist auch im Promotorbereich des Pseudogens enthalten und unterscheidet sich nur durch 5 Nukleotidaustausche (Peltoketo et al., 1999a). Die Promotoren der Nagetiergene sind davon deutlich verschieden; die für die spezifische Regulation des humanen Promotors verantwortlich gemachten Elemente sind dort nicht konserviert. Es ist aber schwierig, ein Muster hinter den Unterschieden zwischen Menschen und Nagetieren zu erkennen, das die Expressionsunterschiede erklären könnte, zumal der hochkonservierte Pseudogenpromotor ja offensichtlich nicht genutzt wird. Ein Promotor-Modell wie für die 17β -HSD7 (Kap. 4.2) konnte mit den verfügbaren Daten nicht entwickelt werden.

Es deutet sich in diesem Ergebnis schon an, dass über die evolutionäre Konservierung von Promotor-Sequenzen und regulatorischen Effekten noch viel zu wenig bekannt ist. Ein auffälliger Unterschied zwischen Maus- und Mensch-Promotor soll aber trotzdem abschließend erwähnt werden: Der Promotor des Maus- 17β -HSD1-Gens enthält eine Reihe von größeren Insertionen, von denen eine zum Auftreten einer (hypothetischen) Östrogen-Rezeptor-Bindungsstelle etwa 400 bp vor dem Translationsstart führt, also etwa in dem Bereich, der dem humanen Enhancer entspricht.

Promotor-Vergleich

Schlussfolgerung

4.5 Die 17 β -HSD Typ 5 – Integration von phylogenetischer und struktureller Information zur Vorhersage und Analyse der Ligandenbindung³

4.5.1 Problemstellung

Eine große Zahl von Mitgliedern der HSD-Familie ist bereits röntgenkristallografisch untersucht worden, ihre Tertiärstruktur ist also im atomaren Detail bekannt. Es fehlen jedoch Röntgenstrukturen der meisten 17 β -HSDs. Die einzige Ausnahme ist die 17 β -HSD1, von der einige Kristallstrukturen vorliegen. Die Bestimmung von Proteinstrukturen ist immer noch sehr zeitaufwendig und nicht immer erfolgreich. Sie kann deshalb in absehbarer Zeit nicht für jede 17 β -HSD durchgeführt werden. Die Kenntnis der dreidimensionalen Strukturen ist aber ein entscheidender Vorteil bei der Analyse des katalytischen Mechanismus und auch beim strukturbasierten Design neuer Therapeutika. Prinzipiell sollte es möglich sein, die Röntgenstrukturen von homologen Proteinen zu nutzen, um auch die Strukturen der 17 β -HSDs zu modellieren. Die geringe Sequenzkonservierung zwischen 17 β -HSD-Proteinen und ihren kristallisierten Verwandten macht die Homologie-basierte Modellierung aber sehr schwierig. Deshalb wurde zuerst die humane 17 β -HSD5 (HSD17B5=AKR1C3) betrachtet. Sie ist das einzige bisher bekannte 17 β -HSD-Protein, das nicht zu den short-chain Alkoholdehydrogenasen gehört, sondern zu den Aldo-Keto-Reduktasen, einer kleineren und gut untersuchten Proteinfamilie. Ein naher Verwandter der 17 β -HSD5, die 3 α -Reduktase der Ratte (AKR1C9; >70% Sequenzübereinstimmung), ist kristallisiert worden, und weitere nahe Verwandte im Menschen (AKR1C1, 2, 4; 84–88% Sequenzübereinstimmung) sind enzymatisch genau charakterisiert. Diese Daten erleichtern die Untersuchung der 17 β -HSD5. Es wurde damit begonnen, einen Zusammenhang zwischen Proteinstrukturelementen und der Sequenzkonservierung innerhalb der AKR-Überfamilie herzustellen. Auf diese Weise wurden die hoch konservierten Bereiche der Struktur identifiziert, die für eine erfolgreiche Homologie-Modellierung ausschlaggebend sind. Der dafür entwickelte Algorithmus ist universell verwendbar und kann dazu dienen, auch in anderen Proteinfamilien die dreidimensionalen Muster der Sequenzkonservierung zu untersuchen. Anschließend wurde auf dieser Grundlage ein dreidimensionales Modell der 17 β -HSD5 entwickelt und zur Interpretation von *in vitro*-Inhibitionsstudien genutzt.

³ Teilweise veröffentlicht in: Krazeisen A, Breitling R, Möller G & Adamski J (2000, im Druck): Phytoestrogens inhibit human 17 β -hydroxysteroid dehydrogenase type 5. Mol. Cell. Endocrinol.

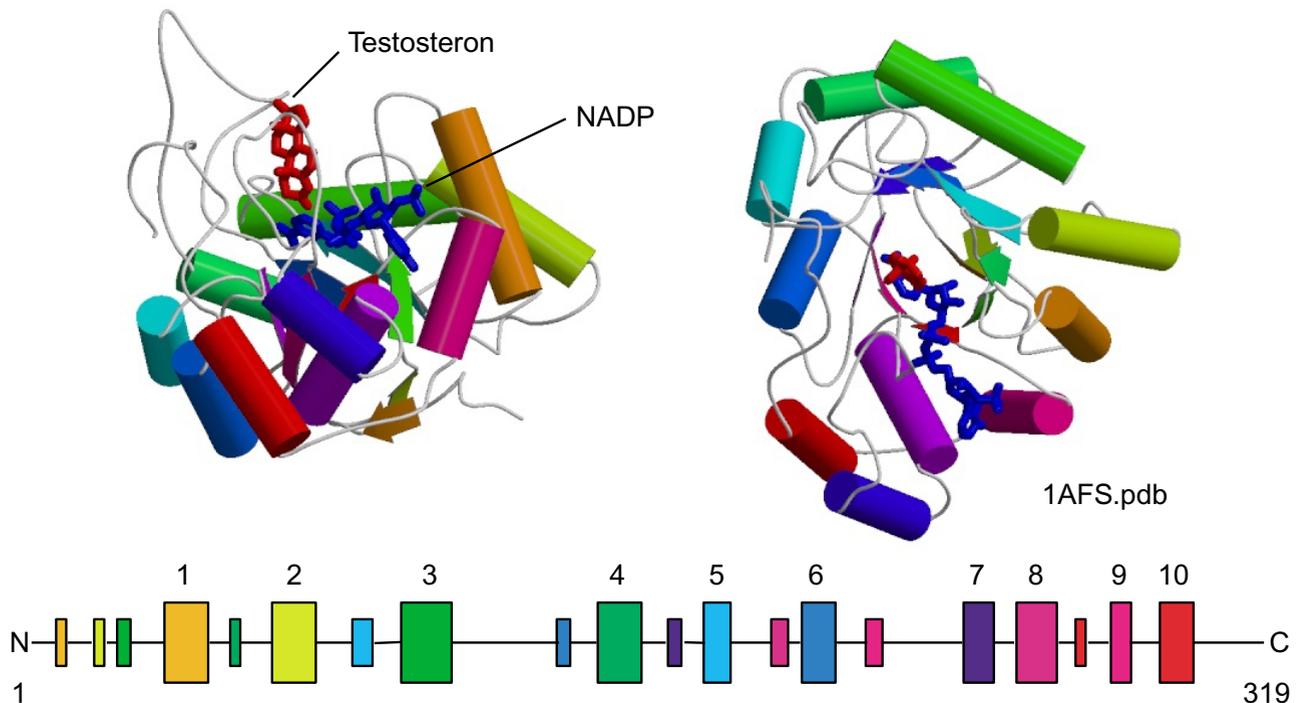


Abbildung 4.25: Struktur der AKR-Proteine. Die konservierten Elemente des $(\alpha/\beta)_8$ -barrel sind farblich gekennzeichnet.

4.5.2 Hintergrund

Aldo-Keto-Reduktasen

Die Aldo-Keto-Reduktase-Überfamilie ist eine der drei großen Proteinklassen, denen die Mehrzahl der bekannten Oxido-Reduktase-Enzymen angehört. Sie kommen in allen Organismenreichen vor, ihre Anzahl pro Genom ist jedoch im Allgemeinen deutlich geringer als die der HSD-Mitglieder. 105 AKR-Enzyme sind bereits genauer charakterisiert worden, und von weiteren 125 potentiellen Familien-Mitgliedern sind die cDNA-Sequenzen bekannt. Typische Mitglieder der AKR-Familie sind monomere Proteine von etwa 320 Aminosäuren Länge (Molekulargewicht etwa 37 kDa), die NAD(P)(H) nutzen, um eine Vielzahl von Substraten umzusetzen. Das aktive Zentrum der Proteine enthält eine konservierte katalytische Tetrade aus Tyrosin, Lysin, Asparat und Histidin.

Die dreidimensionale Struktur von neun Mitgliedern der AKR-Überfamilie ist im atomaren Detail bekannt. Sie alle besitzen eine hochkonservierte dreidimensionale Struktur, die durch eine ring- oder fassförmige Anordnung von je 8 α -Helices und 8 β -Faltblättern gekennzeichnet ist (Abb. 4.25, S. 95). Diese Grundstruktur, die als $(\alpha/\beta)_8$ -barrel oder Triosephosphat-Isomerase (TIM)-barrel bezeichnet wird, ist auch aus anderen Proteinfamilien bekannt und stellt

Die AKR-Familie

3D-Strukturen

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

eines der häufigsten Proteinfaltungsmotive dar. Das aktive Zentrum befindet sich in der Mitte dieser Struktur. Der Kofaktor bindet in allen AKRs in der gleichen ausgestreckten anti-Konformation oberhalb des (α/β)₈-barrels. Der Nicotinamid-Ring ist dabei in allen Strukturen in der gleichen Weise relativ zum aktiven Zentrum angeordnet, so dass die Stereochemie des Hydridtransfers erhalten bleibt. Die Substratbindungstasche wird von Aminosäureresten aus mehreren Proteinschleifen oberhalb des (α/β)₈-barrels gemeinsam gebildet.

Nomenklatur

Die aktuelle Nomenklatur der AKR-Überfamilie beruht auf einer hierarchischen Einteilung in verschiedenen Untergruppen, auf der Grundlage ihrer Aminosäuresequenz-Ähnlichkeit. Der komplette Name besteht aus dem Kürzel AKR, einer arabischen Ziffer, die die Proteinfamilie bezeichnet, einem Buchstaben, der die Unterfamilie angibt, und einer weiteren Ziffer, die die individuelle Protein-Sequenz identifiziert (Jez et al., 1997; <http://www.med.upenn.edu/akr/>).

Die humane 17 β -HSD5

Katalytische Aktivität

Die humane 17 β -HSD5 (AKR1C3) wurde unter zahlreichen Namen beschrieben, u. a.:

- 3 α -HSD Typ 2 der humanen Leber (E.C. 1.1.1.213),
- Dihydrodiol-Dehydrogenase X (E.C. 1.3.1.20),
- Prostaglandin-F-Synthase (E.C. 1.1.1.188)
- und 17 β -HSD Typ 5 (E.C. 1.1.1.62).

Diese Vielfalt deutet schon an, dass es sich bei der 17 β -HSD5 nicht um ein einfaches Enzym mit klar abgegrenzter Substratspezifität handelt. Die Eigenschaft, eine Vielzahl von verschiedenen Substraten, wenn auch mit unterschiedlicher Effizienz umsetzen zu können, teilt die 17 β -HSD5 mit anderen Mitgliedern der AKR-Familie und auch mit anderen 17 β -HSDs, z. B. 17 β -HSD4.

Isoenzyme

17 β -HSD5 ist eines von vier Isoenzymen (AKR1C1 bis C4), die als 3-, 17- oder 20-Ketosteroid-Dehydrogenasen und als 3 α -, 17 β - und 20 α -Hydroxysteroid-Oxidoreduktasen aktiv sind, und sich nur durch ihre katalytische Effizienz und Gewebeverteilung unterscheiden (Penning et al., 2000; Nishizawa et al., 2000). Alle vier Isoenzyme haben eine hochkonservierte Exon-Intron-Struktur und befinden sich in enger Nachbarschaft auf Chromosom 10 (Nishizawa et al., 2000). Die wichtigsten Funktionen der vier Isoenzyme, wie sie sich aus dem Vergleich von enzymatischer Aktivität und Gewebeverteilung ergeben, sind in der Tabelle 4.13 (S. 97) zusammengefasst.

Funktion der Isoenzyme

Alle vier Isoenzyme kommen in großer Menge in der Leber vor, wo

Tabelle 4.13: Verteilung und Funktion der AKR1C-Isoenzyme des Menschen

Name	Hauptvorkommen	Identität mit HSD5	Funktion
AKR1C1	Testis, Gehirn	86%	Anxiolyse durch Modulation des GABA A-Rezeptors durch Synthese von 3 α -Hydroxy-5 α -pregnan-20-on.
AKR1C2	Prostata, Gehirn	88%	ebenso
AKR1C3	Prostata	100%	Elimination von aktiven Androgenen durch Oxidation von Testosteron und 5 α -Dihydroxytestosteron.
	Brustdrüse		Erzeugung eines pro-östrogenen Zustands durch Synthese von Östradiol, Testosteron und 20 α -Hydroxyprogesteron.
	Uterus		Geburtsvorbereitung durch Produktion des Geburtshormons 3 α -Androstandiol.
AKR1C4	leberspezifisch	84%	Schutz gegen Steroidhormon-Überschuss durch Bildung von 5 α /5 β -Tetrahydroxysteroiden, Produktion des Gallensäurevorläufers 5 β -Cholestan-3 α ,7 α -diol.

sie durch ihre breite Substratspezifität zum Abbau von Xenobiotika beitragen können. In der Prostata hat 17 β -HSD5 vorwiegend 3 α - und 20 β -Aktivität und wirkt damit anti-androgen durch die Inaktivierung von 5 α -Dihydroxytestosteron. Die 17 β -Aktivität, die zur intrakrinen Produktion von Testosteron beitragen könnte, ist bei gesunden Männern vermutlich zu vernachlässigen (Penning et al., 2000), verglichen mit den durch 17 β -HSD5 produzierten Testosteron-Mengen aus den Leydig-Zellen. In der Brustdrüse kann 17 β -HSD5 durch die Kombination seiner verschiedenen reduktiven Aktivitäten zur Entstehung eines pro-östrogenen Zustands beitragen. Inwieweit die Aktivität von 17 β -HSD5 aber tatsächlich zum östrogenabhängigen Wachstum beitragen kann, wird jedoch von ihrer Expression im Verhältnis zu den beiden anderen pro-östrogenen 17 β -HSDs (17 β -HSD1 und 17 β -HSD7) abhängen.

Da der Kofaktor in allen AKR-Mitgliedern starr und gleichförmig orientiert ist, kommt die Reaktivität an den verschiedenen Positionen des Steroidgerüsts vermutlich durch die unterschiedliche Orientierung des Substrats zustande. 17 β -HSD5 muss also in der Lage sein, ihr Substrat in mindestens drei deutlich verschiedenen Orientierungen zu binden, und zwar in derselben Substratbindungstasche, denn die Röntgenstrukturen zahlreicher naher Verwandten machen die Existenz mehrerer unabhängiger Bindungsstellen äußerst unwahrscheinlich.

Struktur und Substrat-Bindung

Phytoöstrogene als 17 β -HSD5-Inhibitoren

17 β -HSD5 als *drug target*

Die anti-androgene Wirkung von 17 β -HSD5 in der Prostata und ihre östrogene Wirkung in der Brustdrüse machen dieses Enzym zu einem potentiellen Angriffspunkt für Medikamente zur Behandlung von hormon-abhängigen Tumoren, aber auch zu einem möglichen pharmakologischen Rezeptor für Phytoöstrogene.

Vorkommen der Phytoöstrogene

Diese Pflanzenmetabolite, die chemisch sehr vielfältig sind (Flavonoide, Isoflavonoide, Lignane und andere), werden unter anderem mit der krebsvorbeugenden Wirkung einer vegetarischen und/oder Soja-reichen Diät in Verbindung gebracht. Phytoöstrogene kommen in allen grünen Pflanzen in unterschiedlicher Menge und Zusammensetzung vor, in besonders hoher Konzentration in Leguminosen und speziell in Sojabohnen. In den Pflanzen erfüllen sie eine Vielzahl von physiologischen Funktionen, unter anderem sind sie an der Interaktion mit Parasiten und Symbionten beteiligt, sie schützen die Pflanzen aber auch vor ultravioletter Strahlung (van Rhijn & Vanderleyden, 1995; Albrecht et al., 1999; van de Sande & Bisseling, 1997; Stapleton & Walbot, 1994).

Epidemiologie

In epidemiologischen Studien konnte gezeigt werden, dass in Japan und einigen anderen asiatischen Ländern die Sterblichkeit durch Prostatakarzinom deutlich niedriger ist, als in westlichen Industrieländern. Diese geringere Sterblichkeit korreliert mit hohen Plasmakonzentrationen von Phytoöstrogenen wie dem Isoflavon Genistein (Adlercreutz et al., 1993b). Die Phytoöstrogene könnten in diesen Fällen als natürliche Chemotherapeutika wirken, und das Flavonoid Quercetin befindet sich bereits in der ersten klinischen Erprobung zur Krebstherapie (Ferry et al., 1996).

Labor-Experimente

Von 26 experimentellen Studien zur Auswirkung einer Soja-haltigen Ernährung bei Nagetieren zeigten 17 (65%) einen positiven, schützenden Effekt gegenüber verschiedenen experimentell erzeugten Tumoren (Messina et al., 1994). In keiner Untersuchung wurden dagegen negative Wirkungen beobachtet. Kontrollierte Studien zeigten, dass bereits relativ geringe zusätzliche Phytoöstrogenmengen (45 mg Isoflavon pro Tag) die Zykluslänge erhöhen. Das führt über die Lebenszeit betrachtet zu einer verringerten Östrogen-Belastung und sollte damit zu einer Senkung des Krebsrisikos beitragen. Diese Korrelation wird tatsächlich bei japanischen Frauen beobachtet, deren im Vergleich zu Europäerinnen um 4-6 Tage verlängerter Zyklus mit einer verringerten Brustkrebshäufigkeit einhergeht (Cassidy et al., 1994). Für das Isoflavonoid Genistein wird *in vitro* eine anti-proliferative Wirkung auf zahlreiche Krebszelllinien beobachtet, und Genistein und sein Vorläufer Biochanin A lösen Apoptose von Tumorzellen aus (Yanagihara et al., 1993).

Wirkungsweise

Verschiedene Mechanismen wurden für die chemoprotektive Wirkung der Phytoöstrogene vorgeschlagen:

- Phytoöstrogene erhöhen die Synthese von hormonbindendem Globulin im Serum und verringern dadurch die Menge an freiem Östradiol (Ad-

lercreutz et al., 1992).

- Lignane inhibieren die Aromatase und können dadurch die Umsetzung von Androstandion zu Östron, bzw. von Testosterone zu Östradiol hemmen (Adlercreutz et al., 1993a).
- Genistein hemmt Protein-Tyrosinkinase (Adlercreutz et al., 1995).
- Genistein und andere Flavonoide haben anti-oxidative Wirkungen (Wei et al., 1995; Pietta 2000).
- Zahlreiche Phytoöstrogene hemmen 17 β -HSD-Enzyme und können dadurch das Verhältnis von aktiven und inaktiven Steroidhormonen verschieben (Krazeisen et al., 2000).
- Negativ ist zu vermerken, dass Phytoöstrogene wie die Bioflavonoide Quercetin und Genistein die DNA-Topoisomerase II hemmen und dadurch zu Chromosomentranslokationen führen können. Eben solche Translokationen werden bei frühkindlichen und Chemotherapie-induzierten Leukämien beobachtet (Strick et al., 2000). Diese Tatsache und die epidemiologische Beobachtung, dass frühkindliche Leukämien in ostasiatischen Ländern vermehrt auftreten (Parkin et al., 1988; Gurney et al., 1997), machen die Phytoöstrogene zu potentiellen Karzinogenen.

Um die Wirkungen der Phytoöstrogene auf die 17 β -HSD5 zu untersuchen, wurden *in vitro*-Inhibitionsmessungen an rekombinantem Enzym durchgeführt (Krazeisen et al., 2000). Um die Ergebnisse zusammenfassend interpretieren zu können und evtl. für ein rationelles Arzneimittel-Design zu verwenden, wurden die folgenden Analysen zur dreidimensionalen Struktur der 17 β -HSD5 durchgeführt.

Aufgabenstellung

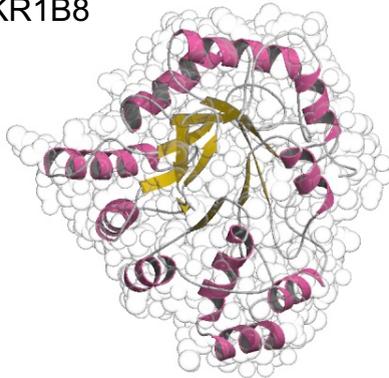
4.5.3 Ergebnisse und Diskussion

Der Vergleich der bereits röntgenkristallografisch untersuchten Verwandten der 17 β -HSD5 zeigte, dass alle diese Proteine einen hochkonservierten Strukturkern besitzen, das (α/β)₈-barrel (Abb. 4.26, S. 100). Diejenigen Abschnitte des Polypeptids, die die konservierten (α/β)-Elemente miteinander verbinden, sind dagegen sehr flexibel. Es ist aber klar, dass die Ähnlichkeit in diesen Bereichen mit dem Grad der Sequenzhomologie zunimmt: z. B. sind die Strukturen der Aldehyd-Reduktase aus Rattenlinsen (AKR1B4) und von murinem FR-1 (AKR1B8) fast identisch; diese Proteine gehören der gleichen Unterfamilie (AKR1B) an. Weitere Mitglieder der gleichen Proteinfamilie (AKR1), nämlich humane Aldose-Reduktase (AKR1A1) und Ratten 3 α -Hydroxysteroiddehydrogenase (AKR1C9) zeigen ebenfalls deutliche Übereinstimmung, selbst in den Bereichen außerhalb der (α/β)₈-Struktur. Die Mitglieder anderer Familien (2,5-Diketo- Δ -gluconsäure-Reduktase [AKR5C] und Shaker-Kanal

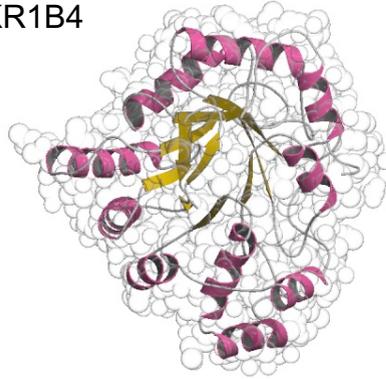
Vergleich der AKR-Protein-Strukturen

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

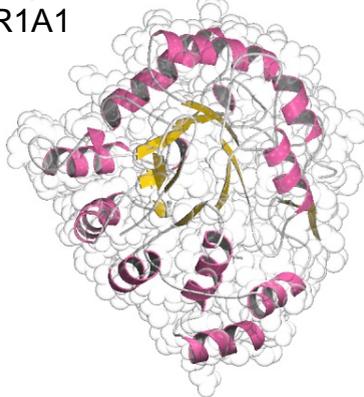
FR-1
AKR1B8



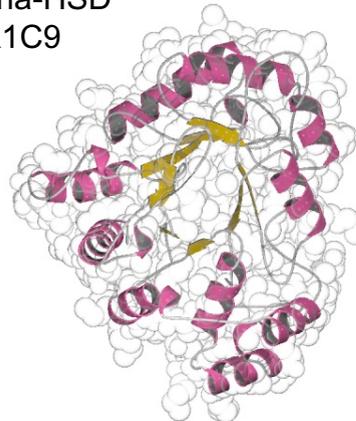
Aldose-Reduktase
AKR1B4



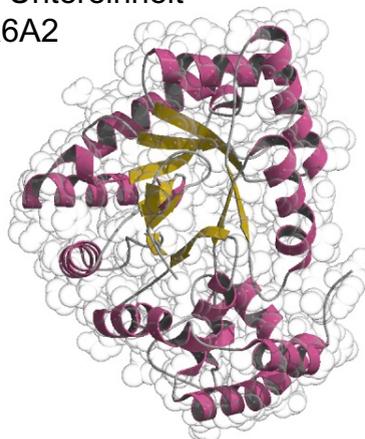
Aldehyd-Reduktase
AKR1A1



3 α -HSD
AKR1C9



Kaliumkanal-
beta-Untereinheit
AKR6A2



Diketogluconsäure-
Reduktase
AKR5C

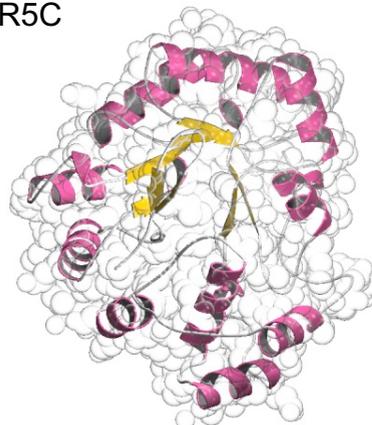


Abbildung 4.26: Vergleich der Strukturen von AKR-Proteinen. Für die Darstellung wurde eine Überlagerung der röntgenkristallografischen Daten neu berechnet, so dass die Ähnlichkeiten der Strukturen besonders deutlich werden. Dargestellt sind die Kristallstrukturen von Aldehyd-Reduktase, FR-1, Aldose-Reduktase, 3 α -Hydroxysteroiddehydrogenase, 2,5-Diketo- Δ -gluconsäure-Reduktase und der Shaker-Kanal β -Untereinheit. Nähere Erläuterungen im Text.

β -Untereinheit [AKR6A2]) sind dagegen bereits deutlich verschieden. Aber selbst bei diesen Strukturen bleibt die relative Anordnung der konservierten Strukturelemente zueinander erhalten. Es schien deshalb möglich, die humane 17 β -HSD5 (AKR1C3) aufgrund ihrer hohen Homologie zur bereits kristallisierten 3 α -HSD (AKR1C9) zuverlässig zu modellieren.

Der Vergleich der Strukturen des 3 α -HSD-Monomers und seiner binären und tertiären Komplexe mit Substrat und/oder Kofaktor zeigte jedoch, dass die substratbindenden Schleifen und selbst der katalytische Tyrosinrest des aktiven Zentrums sehr flexibel sind; ein Befund der angesichts der hohen Substratplastizität der 17 β -HSD5 nicht überrascht, aber die Generierung eines im atomaren Detail zuverlässigen Molekülmodells sehr erschwert (Abb. 4.27, S. 105). Selbst wenn es gelänge, ein genaues Modell zu erzeugen, wäre damit für das Verständnis der Substrat- oder Inhibitorbindung unter Umständen wenig gewonnen, weil jede der in Frage kommenden Substanzen durch ihre Bindung individuell eine passende Verformung der Struktur induzieren könnte.

Deshalb wurde versucht, zuerst die Proteinelemente zu definieren, die evtl. unabhängig vom Liganden eine Beurteilung der Bindungseigenschaften ermöglichen. Dazu wurde ein Computeralgorithmus entwickelt (STATRAS2), der ausgehend von einem Alignment verwandter Proteinsequenzen die Konservierung jeder einzelnen Position in der Sequenz berechnet und die so erhaltenen Daten in grafischer, leicht interpretierbarer Form ausgibt. Als Maß für die empirische Konservierung an der Position i diente die statistische Energie ΔG_i^{stat} , die wie folgt definiert ist:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x} \right)^2},$$

wobei kT^* eine willkürlich gewählte Energieeinheit ist und x die zwanzig möglichen Aminosäuren bezeichnet. P_i^x und P_{MSA}^x sind die Wahrscheinlichkeiten eine bestimmte Aminosäure x an der Position i bzw. im gesamten Alignment zu finden. Das bedeutet, dass die Konservierung an Position i proportional ist zur Abweichung der Aminosäurehäufigkeiten an dieser Position von der mittleren Häufigkeitsverteilung im gesamten Alignment, oder, anders formuliert: Eine Position ist um so weniger konserviert, je mehr sich ihre Aminosäureverteilung einer Zufallsverteilung nähert. Das Ergebnis der Berechnungen wird von dem zu diesem Zweck entwickelten Programm in einer EXCEL-lesbaren Tabelle ausgegeben, die die Aminosäurehäufigkeiten an den einzelnen Positionen und die ΔG^{stat} -Werte enthält. Die ΔG^{stat} -Werte werden außerdem in eine Proteinkoordinaten-Datei im PDB-Format eingetragen, so dass sich die relative Anordnung der konservierten Reste innerhalb der dreidimensionalen Struktur leicht erkennen lässt.

Die Verwendung dieses Verfahrens setzt voraus, dass das Alignment einen möglichst umfangreichen und gleichmäßig verteilten Ausschnitt aus der gesamten Proteinfamilie enthält (repräsentative Stichprobe), so dass die Hinzunahme weiterer Sequenzen die Häufigkeitsverteilung an den einzelnen Positio-

**Struktur-
Änderungen durch
Substrat-Bindung**

**Statistische
Analyse von
konservierten
Strukturen**

Protein-Alignment

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

nen nicht signifikant beeinflussen würde. Für die Berechnungen wurde deshalb ein Alignment zugrunde gelegt, das Mitglieder aller Teilfamilien der AKR-Überfamilie enthielt. Dieses Alignment wurde durch eine PSI-Blast-Suche gewonnen, deren Ergebnis einer Redundanzfilterung unterzogen wurde, um den überproportionalen Einfluss besonders häufig vertretener Proteine zu vermeiden.

3D-Darstellung der Konservierung

In der dreidimensionalen Darstellung wird deutlich, dass nicht nur die Reste im katalytischen Zentrum hochkonserviert sind, sondern dass sich konservierte Elemente auch an mehreren anderen Stellen der Proteine befinden (Abb. 4.28, S. 106). Zwei solche Elemente befinden sich an den beiden Enden der Kofaktorbindungstasche. Sie sind vermutlich an der immer in gleicher Weise erfolgenden Bindung des Kofaktors in seiner ausgestreckten anti-Konformation und an seiner passenden Orientierung zu den katalytisch wirksamen Aminosäureseitenketten beteiligt. Eine weitere Anhäufung hochkonservierter Reste befindet sich an der Oberfläche des Proteins zwischen Helix $\alpha 3$ und Faltblatt $\beta 4$ (Nomenklatur nach Penning et al., 1997). Diese Reste sind weit von den Substrat- und Kofaktorbindungsstellen entfernt und scheinen keinen zentralen Platz in der $(\alpha/\beta)_8$ -Struktur einzunehmen. Es ist jedoch möglich, dass sie an den allosterischen Umordnungsprozessen beteiligt sind, die mit der Ligandenbindung einhergehen. Die Substratbindungstasche ist erwartungsgemäß nicht von konservierten Resten umgeben. Das erklärt sich sicher durch die Substratvielfalt der in die Analyse einbezogenen Enzyme, im Vergleich zur Einförmigkeit des verwendeten Kofaktors.

3D-Modellierung

Ein molekulares Modell der 17 β -HSD5 wurde anschließend durch automatische homologiebasierte Modellierung (Swiss-Model) erzeugt (Abb. 4.29, S. 107). Die vorhergesagte Struktur ist der der Ratten-3 α -HSD sehr ähnlich. Die bereits festgestellte Flexibilität der Substratbindungstasche machte eine Vorhersage von Inhibitoraffinitäten z. B. durch molecular docking-Algorithmen sinnlos, da diese Verfahren einen starren Rezeptor voraussetzen. Einige der beobachteten Inhibitorwirkungen konnten aber trotzdem erklärt werden.

Angriffspunkt der Inhibitoren

1. Im ternären Komplex von Ratten-3 α -HSD, NADP und Testosteron liegt die 3 α -Hydroxygruppe des Steroids zwischen den konservierten Tyrosin- und Histidin-Resten der katalytischen Tetrade und bildet mit diesen beiden Seitenketten Wasserstoffbrückenbindungen (Bennett et al., 1997). Im binären Komplex ohne Testosteron ist dieselbe Position durch ein fest gebundenes Wassermolekül besetzt (Bennett et al., 1996). Auch mehrere Inhibitoren von anderen AKR-Enzymen binden mit einem Sauerstoffatom an dieser Position (El-Kabbani et al., 1998). Es wird deshalb vorgeschlagen, dass die Hydroxygruppe in Position 7, die den größten Einfluss auf die inhibitorische Potenz der Flavonoide hat, ebenfalls an derselben Position im aktiven Zentrum bindet.

pH-Abhängigkeit der Inhibition

2. Auch die Carboxygruppe der Glycyrrhetinsäure sollte die Carbonylgruppe des endogenen Substrats nachahmen können und im aktiven Zen-

trum binden. Diese Art der Bindung ist auch für die pharmakologisch wichtigen Inhibitoren Tolrestat (an Aldehydreduktase AKR1B4) und Zopolrestat (an Aldosereduktase AKR1B1) gezeigt worden (El-Kabbani et al., 1998). Im Falle der Glycyrrhetinsäure-Wirkung auf 17 β -HSD5 kann durch dieses Modell auch erklärt werden, warum die Inhibition der oxidativen Reaktion deutlich geringer ausfällt als die der reduktiven Reaktion. Ein solcher Unterschied wird für einen kompetitiven Inhibitor eigentlich nicht erwartet. Die oxidative Umsetzung wurde jedoch in einem basischeren Milieu gemessen als die reduktive. Wahrscheinlich führt die Erhöhung des pH-Wertes zu einer teilweisen Deprotonierung des Histidin-Restes im aktiven Zentrum. Die daraufhin freiwerdende negative Ladung verhindert dann die Bindung der ebenfalls negativ geladenen Carboxylgruppe des Inhibitors.

3. Die Orientierung der Inhibitoren innerhalb des Enzyms ist aus den vorhandenen Daten nicht abzuleiten. Die östrogene Aktivität der Phytoöstrogene führte ursprünglich zu der Vermutung, dass sie als Substratanaloga wirken und an der Steroidbindungsstelle kompetitiv angreifen. Die Inhibitorwirkung der Flavonoide nimmt jedoch mit dem Hydroxylierungsgrad systematisch zu; die am stärksten hydroxylierten Substanzen sind unabhängig von der betrachteten Reaktion die besten Inhibitoren. Das ist mit der Vorstellung einer Bindung in der apolaren Substratbindungsstelle nur schwer zu vereinbaren; es weist im Gegenteil auf eine Bindung in der mit hydrophilen Resten ausgekleideten hochkonservierten Kofaktor-Bindungsstelle hin. Andererseits ist aber auch denkbar, dass durch die Inhibitorbindung eine Strukturveränderung der Substratbindungsschleifen induziert wird, durch die zusätzliche hydrophile Reste in diesem Bereich zur Bindung an die Hydroxylgruppen der Inhibitoren bereitgestellt werden. Wenn die Phytoöstrogene tatsächlich an der konservierten Kofaktor-Bindungsstelle angreifen, dann wird ihre Nützlichkeit als Leitsubstanzen für das rationale Design spezifischer Inhibitoren in Frage gestellt - es ist zu erwarten, dass Phytoöstrogen-Derivate immer ein breites Spektrum von verschiedenen Dehydrogenasen hemmen. Dass die Phytoöstrogene nicht nur als Steroidanaloga wirken können, wird ja durch ihre direkte Inhibitor-Wirkung auf Topoisomerasen und Protein-kinasen gezeigt.

Orientierung der Inhibitoren

Folgende Schlussfolgerungen ergeben sich aus der Betrachtung der 17 β -HSD5:

Schlussfolgerungen

1. Die Modellierungsergebnisse für die 17 β -HSD5 reichen nicht aus, um die Vorhersage von Substratspezifitäten oder das Design von Inhibitoren zu ermöglichen, obwohl in diesem Fall die dreidimensionale Struktur eines sehr nahen Verwandten bekannt ist. Dieses Problem wird bei der Betrachtung der anderen 17 β -HSD-Proteine in verschärfter Form auftreten.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

2. Das Modellieren eines konservierten Proteingrundgerüsts ist aber auch bei geringer Sequenzhomologie aussichtsreich. Voraussetzung ist aber ein korrektes Sequenzalignment. Wie einzelne Proteinelemente relativ zueinander angeordnet sind, kann aber auch in den Bereichen festgestellt werden, die außerhalb der hochkonservierten Kernstruktur liegen.
3. Die statistische Analyse von Proteinalignments in Verbindung mit Röntgenstrukturdaten mit dem neu entwickelten Programm STATRAS2 liefert unvermutete Erkenntnisse über das dreidimensionale Muster der Sequenzkonservierung. Davon sind auch Aufschlüsse über konservierte allosterische Effekte (z. B. nach Substratbindung) zu erwarten. Der Algorithmus sollte zu diesem Zweck aber noch erweitert werden, um auch die statistische Kopplung von Sequenzpositionen untersuchen zu können, wie dies Lockless & Ranganathan (1999) für die PDZ- und POZ-Domänen vorgeführt haben. Die Einbeziehung der auf diese Weise identifizierten „*pathways of energetic connectivity*“ kann auch die Qualität von Homologie-basierten Proteinmodellen erhöhen. Auch eine Einbindung des STATRAS2-Algorithmus in das derzeit an der University of California, San Francisco, entwickelte Chimera-Paket zur Molekülmodellierung würde sich in diesem Zusammenhang anbieten (Huang et al., 2000).
4. Diese Verbindung von dreidimensionaler Information und statistischer Analyse ermöglicht auch dann die Interpretation von Proteinmodellen, wenn keine atomare Auflösung erreicht werden kann.
5. Das Programm STATRAS2 kann natürlich auch für die Untersuchung anderer Proteinfamilien verwendet werden. Voraussetzung sind eine bekannte dreidimensionale Struktur und ein Sequenzalignment einer repräsentativen Stichprobe der Proteinfamilie.

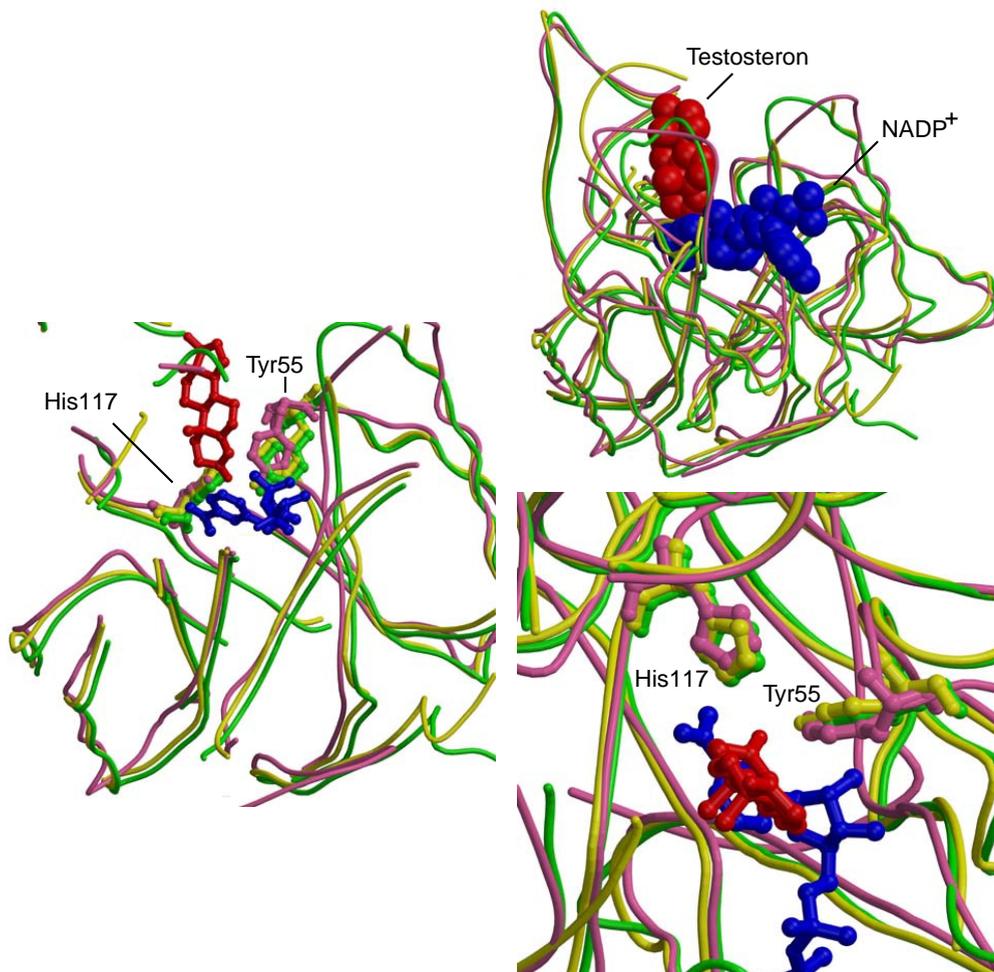


Abbildung 4.27: Strukturänderungen der 3α -HSD nach Ligandenbindung. Hierfür wurde die optimale Überlagerung von Apoenzym (rosa), binärem Komplex mit Kofaktor (gelb) und ternärem Komplex mit Kofaktor und Substrat (grün) berechnet. rot: Testosteron, blau: NADP. Die weit ausladenden Bewegungen der Substratbindungsschleifen und die Umorientierung der katalytischen Reste sind deutlich zu sehen.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

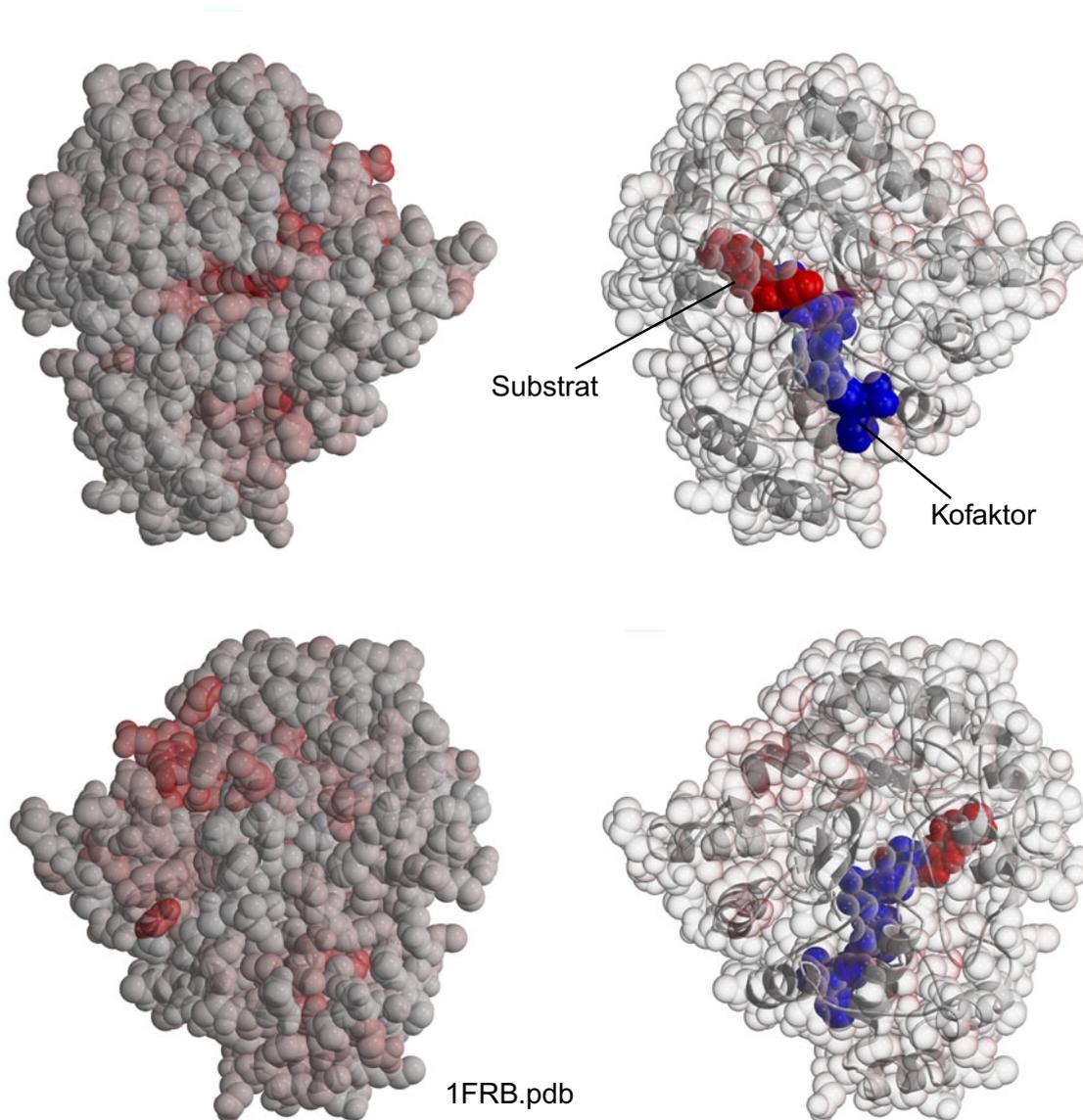


Abbildung 4.28: Sequenzkonservierung in der AKR-Familie. An diesem Beispiel werden die Möglichkeiten des in dieser Arbeit entwickelten STATRAS2-Programms demonstriert. Das Ausmaß der Sequenzkonservierung (ΔG^{stat}) ist durch die Stärke der roten Schattierung dargestellt. In der transparenten Darstellung (rechts) sind Ligand und Kofaktor wie in Abb. 4.27 gefärbt. Die obere Reihe zeigt die „Vorderansicht“ des Enzyms, mit dem konservierten aktiven Zentrum in der Mitte. Auch das distale Ende der Kofaktorbinderungsstelle ist hoch konserviert (roter Bereich unterhalb des aktiven Zentrums). In der „Rückansicht“ (untere Reihe) ist links oben ein großer konservierter Bereich zu erkennen, der weit von den Ligandenbindungsstellen und dem aktiven Zentrum entfernt ist. Er ist möglicherweise an allosterischen Umordnungen beteiligt.

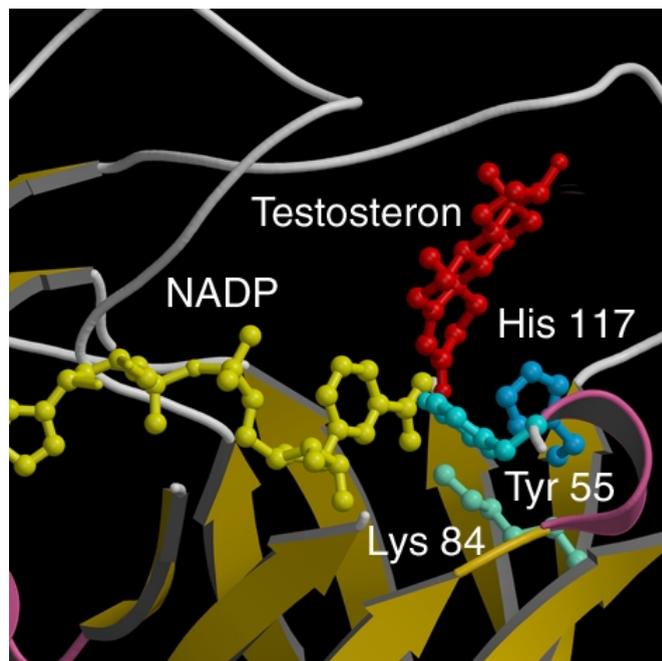


Abbildung 4.29: Modell des aktiven Zentrums der humanen 17β -HSD5. Die katalytisch aktiven Reste sind gekennzeichnet.

4 Die 17 β -Hydroxysteroid-Dehydrogenasen

5 Zusammenfassung

Diese Arbeit demonstriert die kombinierte Anwendung unterschiedlicher bioinformatischer Verfahren auf die Familie der 17β -Hydroxysteroid-Dehydrogenasen. Diese Protein-Familie ist an der Aktivierung und Inaktivierung von Steroid-Hormonen beteiligt, die sie an der Position 17 des Steroid-Grundgerüsts reduzieren bzw. oxidieren. Verschiedene 17β -HSDs sind an humanen Erkrankungen beteiligt, z. B. an Pseudohermaphroditismus (17β -HSD3) oder Zellweger-artigem Syndrom (17β -HSD4). Außerdem wird den 17β -HSDs eine Rolle bei der Entstehung und Proliferation von verschiedenen Tumoren zugeschrieben.

In dieser Arbeit wird ein neues Evolutionsszenario für die 17β -HSDs und ihre Verwandten beschrieben. Durch Vereinigung von Proteinstrukturvergleichen und Sequenzvergleichen zu einem einheitlichen Modell war es möglich, einen zuverlässigen Stammbaum der Proteinfamilie zu erstellen. In diesem Stammbaum zeigte sich eine frühe, grundlegende Teilung der 17β -HSDs in zwei Hauptklassen. Durch Feststellung der Stammbaumwurzel war es möglich den zeitlichen Verlauf der Evolutionsereignisse festzustellen, die zur Entwicklung der Proteinvielfalt in der HSD-Familie geführt haben.

Eine umfassende bioinformatische Studie wurde durchgeführt, um die physiologische Funktion der 17β -HSD Typ 7 festzustellen. Durch Kombination von Expressionsdaten, phylogenetischer Rekonstruktion und Strukturanalysen war es möglich, zu zeigen, dass dieses Östrogen-synthetisierende Enzym wahrscheinlich zuerst an der Cholesterin-Biosynthese beteiligt war, und zwar bei der Reduktion von 3-Ketosteroiden. 17β -HSD7 ist beim Menschen fast ubiquitär exprimiert und zeigt signifikante Ähnlichkeit mit der 3-Ketosteroid-Reduktase Erg27p aus Hefen. Das Protein fehlt dagegen in Cholesterin-auxotrophen Organismen. Zusätzliches Gewicht erhält diese Vermutung durch die Entdeckung eines spezifischen Promotormoduls, das 17β -HSD7 mit Proteinen des Cholesterinmetabolismus (IDI1 und MLN64) gemeinsam hat. Anschließende Reaktionen des Cholesterin-Synthesewegs werden von Proteinen katalysiert, deren Mutation zu den Entwicklungsdefekten CHILD-Syndrom und CDPX2-Syndrom führt. Damit ist auch 17β -HSD7 ein Kandidat für ähnliche Erkrankungen.

Durch die Integration von Strukturdaten mit der statistischen Auswertung von Proteinalignments wurden unerwartete konservierte Elemente in der dreidimensionalen Struktur von 17β -HSD Typ 5 und ihren Verwandten entdeckt.

5 Zusammenfassung

Ein 3D-Modell des Enzyms ermöglichte es, Struktur-Funktions-Beziehungen von Phytoöstrogenen aufzuklären, die als Inhibitoren der 17β -HSD5 wirken. Außerdem war es möglich die pH-Abhängigkeit der Inhibition durch Glycyrrhetin-Säure zu erklären.

Die notwendigen Methoden zur Evolutionsanalyse großer Datensätze wurden in einer Pilotstudie entwickelt und getestet. Dazu wurde die Evolution der Paired-Box (Pax)-Familie untersucht. Es konnte gezeigt werden, dass der nächste Verwandte dieser Proteine eine Tc1-Transposase ist. Das erste Pax-Protein entstand am Anfang der Metazoen-Evolution durch ein einmaliges Fusionsereignis, bei dem sich die DNA-Bindungsdomäne einer Transposase mit einer Homöodomäne zu einem neuen Transkriptionsfaktor verband. Diesem Fusionsereignis folgte dann eine schnelle Aufspaltung der Familie in mehrere Untergruppen. Durch die Entdeckung der Verwandtschaft mit den Transposasen ist es möglich, nicht nur die Einteilung der Pax-Proteine festzustellen, sondern auch den zeitlichen Ablauf ihrer Evolution zu bestimmen.

Epilog

Nachdem zu Beginn der Sechzigerjahre mit der Entwicklung der „*New Synthesis*“ die Entstehungsphase der modernen Evolutionstheorie abgeschlossen war, fasste der amerikanische Genetiker Theodosius Dobzhansky das Selbstbewusstsein der jungen Wissenschaft in einem klassisch gewordenen Zitat zusammen: „*Nothing in biology makes sense, except in the light of evolution*“ (Dobzhansky, 1973). Heute, am Beginn des post-genomischen Zeitalters könnte man versucht sein hinzuzufügen: „*And little in biology makes sense, except through the eyes of bioinformatics.*“

Evolutionsbiologie und Bioinformatik zeigen ihre gemeinsame Stärke bei der Synthese großer, unübersichtlicher Datenmengen zu einheitlichen Modellen von großer Informationsdichte, Erklärungskraft und „Kühnheit“ im Sinne der Popperschen Erkenntnistheorie (Popper, 1972). Je abstrakter und tiefschürfender aber die Analyse wird und je weiter sich die Modelle entwickeln, desto deutlicher wird die verführerische Schwäche der beiden Ansätze: Aus wissenschaftlichen Modellen werden unverbindliche Spekulationen, deren empirische Überprüfung und Falsifizierbarkeit immer mehr in den Hintergrund gerät. In der vorliegenden Arbeit wurde versucht, dieser Gefahr durch den ständigen Bezug zur experimentellen Laborarbeit zu entgehen; inwieweit das gelungen ist, möge der Leser beurteilen.

Epilog

Glossar

Anxiolyse: Aufhebung von Angstsymptomen.

Bootstrap: Computer-gestütztes Verfahren zur Beurteilung der Zuverlässigkeit einer statistischen Abschätzung, z. B. eines aus Sequenzdaten berechneten Stammbaums. Die Methode beruht auf der Auswertung einer großen Zahl von Pseudoreplikaten (>50) des betrachteten Datensatzes, die jeweils eine zufällig ausgewählte Untermenge des Ausgangsdatsatzes enthalten (Felsenstein, 1985; Efron, 1996).

Deuterostomia: Taxon, das alle Nachkommen des jüngsten gemeinsamen Vorfahren von Echinodermen (z. B. Seeigeln) und Wirbeltieren enthält und durch den Beginn der Gastrulation am zukünftigen Anus gekennzeichnet ist.

Dichotomie: logische Aufteilung einer Klasse in genau zwei nicht überlappende Unterklassen. Da die divergierende Evolution von Proteinen durch aufeinanderfolgende Duplikationen und Artbildungsereignisse bestimmt wird, muss der vollständig geklärte Stammbaum einer Proteinfamilie eine hierarchische Abfolge von Dichotomien repräsentieren.

Ecdysozoa: Taxon, das alle Protostomia enthält, die regelmäßige Häutungen ihrer α -Chitin-reichen Cuticula durchlaufen (Aguinaldo et al., 1997). Beinhaltet unter anderem Nematoden (z. B. *C. elegans*) und Arthropoden (z. B. *Drosophila*).

Homologe: Zuerst definiert von Richard Owen (1843) als “homologue: the same organ in different organisms under every variety of form and function”. Heute meist Strukturen, die sich auf einen einzigen, gemeinsamen stammesgeschichtlichen Vorläufer zurückführen lassen, z. B. die Vorderextremitäten der Säugetiere und die Flügel der Vögel. In der vorliegenden Arbeit auch für Proteine verwendet, für die die Herkunft von einem gemeinsamen Vorläufer wahrscheinlich ist.

Homologie: Vielfältiges evolutionsbiologisches Konzept (Haszprunar, 1991; Donoghue, 1992; Hall, 1994). Homologie bezeichnet heute im Allgemeinen eine biologische Ähnlichkeit, die sich auf gemeinsame Abstammung zurückführen lässt.

Homoplasie: Ähnlichkeit, die sich nicht notwendig auf gemeinsame Abstammung zurückführen lässt.

latente Homologie: Von de Beer (1971) eingeführte Bezeichnung für die in bestimmten Taxa bestehende Neigung, gemeinsame Eigenschaften scheinbar unabhängig voneinander zu entwickeln, z. B. ist das soziale Verhalten von Ameisen, Bienen und Wespen eine „latente Homologie“ der Hymenoptera, d. h. in dieser Gruppe scheint eine Neigung zur Eusozialität zu bestehen.

Maximum likelihood: Verfahren zur Stammbaumrekonstruktion, bei dem versucht wird, den Stammbaum zu finden, der unter vorgegebenen Evolutionsbedingungen mit der größten Wahrscheinlichkeit zur beobachteten Verteilung von Merkmalen (z. B. Aminosäuresequenzen) geführt hätte.

Metazoa: Vielzellige Tiere.

monophyletisch: Taxon, das alle Nachkommen eines evolutionären Vorläufers umfasst (Gegensatz paraphyletisch).

Neighbour joining: Verfahren zur Stammbaumrekonstruktion, bei dem nacheinander die jeweils ähnlichsten Taxa (*neighbours*) zusammengefasst werden. Für die Analyse von n Taxa wird zuerst eine $n \times n$ -Matrix der paarweisen Ähnlichkeiten erstellt (nach einem vorgegebenen Evolutionsmodell). Die ähnlichsten Taxa werden dann zusammengefasst, die Matrix nach einer vorgegebenen Regel auf eine $(n - 1) \times (n - 1)$ -Matrix reduziert und das Verfahren solange wiederholt, bis der komplette Stammbaum aufgebaut ist.

Orthologe: Homologe, die sich aufgrund eines Artbildungsereignisses auseinanderentwickelt haben (und nicht etwa durch Genduplikation und darauf folgende Divergenz).

Outgroup: Für die Stammbaumrekonstruktion verwendetes Konzept, bei dem ein entfernter Verwandter der betrachteten Gruppe in die Analyse einbezogen wird, der mit keinem ihrer Mitglieder näher verwandt ist, als diese untereinander.

paraphyletisch: Taxon, das nur einen Teil der Nachkommen eines evolutionären Vorläufers umfasst, während andere Nachkommen ausgeschlossen werden, obwohl ihre Verwandtschaft zu einzelnen Mitgliedern des paraphyletischen Taxons enger ist, als die Verwandtschaftsbeziehungen innerhalb des Taxons. Z. B. sind die 17β -HSDs innerhalb der HSD-Familie paraphyletisch, denn zu den Nachkommen ihres jüngsten gemeinsamen Vorläufers gehören auch zahlreiche Proteine, die keine 17β -HSD-Aktivität besitzen.

Parsimony: Verfahren der Stammbaumrekonstruktion, bei dem der Stammbaum gesucht wird, der bei einem vorgegebenen Evolutionsmodell die wenigsten Evolutionsschritte benötigt, um die beobachtete Merkmalsverteilung (z. B. Aminosäuresequenz) zu erklären.

polyphyletisch: Taxon, das aus mehreren para- oder monophyletischen Taxa zusammengesetzt ist und sich nicht auf einen (nur diesen) gemeinsamen Vorläufer zurückführen lässt. Z. B. sind die „17 β -HSD-Enzyme im engeren Sinne“ polyphyletisch, denn zu ihnen gehören Proteine aus zwei nicht verwandten Familien, der AKR- und der HSD-Familie.

Polytomie: logische Aufteilung einer Klasse in mehr als zwei gleichwertige, nicht überlappende Unterklassen.

Protostomia: vermutlich paraphyletisches Taxon innerhalb der Metazoa, das durch Beginn der Gastrulation im zukünftigen Mundbereich gekennzeichnet ist (u. a. Arthropoden, Mollusken, Nematoden).

Schwestergruppen: Zwei Taxa, die untereinander näher verwandt sind, als mit allen anderen betrachteten Taxa.

Synapomorphie: syn. *taxic homology*. Diagnostisches (abgeleitetes) Merkmal eines monophyletischen Taxons.

Taxon: Einheit der Klassifizierung (Taxonomie), z. B. *Homo sapiens*, Primates oder Mammalia. Hier auch auf Proteine angewendet, z. B. HSD-Familie.

Z-Score: Ähnlichkeitsmaß, das die Übereinstimmung von zwei Proteinstrukturen angibt. Je besser sich die Koordinaten zur Deckung bringen lassen und je länger die vergleichbaren Bereiche sind, desto höher ist der Z-Score.

Glossar

Referenzen

Literaturverzeichnis

- Adamski J & Jakob F (2000, im Druck): A guide to 17 β -hydroxysteroid dehydrogenases. *Mol. Cell. Endocrinol.*
- Adamski J, Husen B, Marks F & Jungblut PW (1992): Purification and properties of oestradiol 17 β -dehydrogenase extracted from cytoplasmic vesicles of porcine endometrial cells. *Biochem. J.* 288: 375–381.
- Adamski J, Normand T, Leenders F, Monte D, Begue A, Stehelin D, Jungblut PW & de Launoit Y (1995): Molecular cloning of a novel widely expressed human 80 kDa 17 β -hydroxysteroid dehydrogenase IV. *Biochem. J.* 311: 437–443.
- Adamski J, Leenders F, Carstensen JF, Kaufmann M, Markus MM, Husen B, Tesdorpf JG, Seedorf U, de Launoit Y & Jakob F (1997): Steroids, fatty acyl-CoA, and sterols are substrates of 80-kDa multifunctional protein. *Steroids* 62: 159–163.
- Adlercreutz H, Mousavi Y, Clark J, Hockerstedt K, Hamalainen E, Wahala K, Mäkelä T & Hase T (1992): Dietary phytoestrogens and cancer: in vitro and in vivo studies. *J. Steroid Biochem. Mol. Biol.* 41: 331–337.
- Adlercreutz H, Bannwart C, Wahala K, Mäkelä T, Brunow G, Hase T, Arosemena PJ, Kellis JT jr. & Vickery LE (1993a): Inhibition of human aromatase by mammalian lignans and isoflavonoid phytoestrogens. *J. Steroid Biochem. Mol. Biol.* 44: 147–153.
- Adlercreutz H, Markkanen H & Watanabe S (1993b): Plasma concentrations of phyto-oestrogens in Japanese men. *Lancet* 342: 1209–1210.
- Adlercreutz CH, Goldin BR, Gorbach SL, Hockerstedt KA, Watanabe S, Hamalainen EK, Markkanen MH, Mäkelä TH, Wahala KT & Adlercreutz T (1995): Soybean phytoestrogen intake and cancer risk. *J. Nutr.* 125: 757S–770S.
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA & Lake JA (1997): Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387: 489–493.
- Akinola LA, Poutanen M, Peltoketo H, Vihko R & Vihko P (1998): Characterization of rat 17 β -hydroxysteroid dehydrogenase type 1 gene and mRNA transcripts. *Gene* 208: 229–238.
- Albrecht H, Yoder JI & Phillips DA (1999): Flavonoids promote haustoria formation in the root parasite *Triphysaria versicolor*. *Plant Physiol.* 119: 585–592.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25: 3389–3402.
- Andersson S, Geissler WM, Wu L, Davis DL, Grumbach MM, New MI, Schwarz HP, Blethen SL, Mendonca BB, Bloise W, Witchel SF, Cutler GB jr., Griffin JE, Wilson JD & Russell DW (1996): Molecular genetics and pathophysiology of 17 β -hydroxysteroid dehydrogenase 3 deficiency. *J. Clin. Endocrinol. Metabol.* 81: 130–136.
- Ando A, Kikuti YY, Shigenari A, Kawata H, Okamoto N, Shiina T, Chen L, Ikemura T, Abe K, Kimura M & Inoko H (1996): cDNA cloning of the human homologues of the mouse Ke4 and Ke6 genes at the centromeric end of the human MHC region. *Genomics* 35: 600–602.
- Ashburner M (1998): Speculations on the subject of alcohol dehydrogenase and its properties in *Drosophila* and other flies. *BioEssays* 20: 949–954.
- Azhar S, Khan I & Gibori G (1989): The influence of estradiol on cholesterol processing by the corpus luteum. *Biol. Reprod.* 40: 961–971.
- Baes M, Huyghe S, Carmeliet P, Declercq PE, Collen D, Mannaerts GP & Van Veldhoven PP (2000): Inactivation of the peroxisomal multifunctional protein-2 in mice impedes the degradation of not only 2-methyl-branched fatty acids and bile acid intermediates but also of very long chain fatty acids. *J. Biol. Chem.* 275: 16329–16336.
- Baker ME (1991): Genealogy of regulation of human sex and adrenal function, prostaglandin action, snapdragon and petunia flower colors, antibiotics, and nitrogen fixation: functional diversity from two ancestral dehydrogenases. *Steroids* 56: 354–360.
- Baker ME (1996): Unusual evolution of 11 β - and 17 β -hydroxysteroid and retinol dehydrogenases. *BioEssays* 18: 63–70.
- Baker ME (1998): Evolution of mammalian 11 β - and 17 β -hydroxysteroid dehydrogenases type 2 and retinol dehydrogenases from ancestors in *C. elegans* and evidence for horizontal transfer of eukaryote dehydrogenase to *E. coli*. *J. Steroid Biochem. Mol. Biol.* 66: 355–363.
- Baker ME, Grundy WN & Elkan CP (1998): Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. *Biochem. Biophys. Res. Com.* 248: 250–254.
- Balczarek KA, Lai ZC & Kumar S (1997): Evolution and functional diversification of the paired box (Pax)

Referenzen

- DNA-binding domains. *Mol. Biol. Evol.* 14: 829–842.
- Baldauf SL, Roger AJ, Wenk-Siefert I & Doolittle WF (2000): A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972–977.
- Ball CA, Dolinski K, Dwight SS, Harris MA, Issel-Tarver L, Kasarskis A, Scafe CR, Sherlock G, Blinkey G, Jin H, Kaloper M, Orr SD, Schroeder M, Weng S, Zhu Y, Botstein D & Cherry JM (2000): Integrating functional genomic information into the *Saccharomyces Genome Database*. *Nucl. Acids Res.* 28: 77–80.
- Balling R, Helwig U, Nadeau J, Neubüser A, Schmahl W & Imai K (1996): Pax genes and skeletal development. *Ann. N.Y. Acad. Sci.* 785: 27–33.
- Bammert GF & Fostel JM (2000): Genome-wide expression patterns in *Saccharomyces cerevisiae*: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol. *Antimicrob. Agents Chemother.* 44: 1255–1265.
- Bennett MJ, Schlegel BP, Jez JM, Penning TM & Lewis M (1996): Structure of 3 α -hydroxysteroid/dihydrodiol dehydrogenase complexed with NADP⁺. *Biochemistry* 35: 10702–10711.
- Bennett MJ, Albert RH, Jez LM, Ma H, Penning TM & Lewis M (1997): Steroid recognition and regulation of hormone action: crystal structure of testosterone and NADP⁺ bound to 3 α -hydroxysteroid/dihydrodiol dehydrogenase. *Structure* 5: 799–812.
- Bertuccioli C, Fasano L, Jun S, Wand S, Sheng G & Desplan C (1996): In vivo requirements for the paired domain and homeodomain of the paired segmentation gene product. *Development* 122: 2673–2685.
- Biswas MG & Russell DW (1997): Expression cloning and characterization of oxidative 17 β - and 3 α -hydroxysteroid dehydrogenases from rat and human prostate. *J. Biol. Chem.* 272: 15959–15966.
- Blake JA, Eppig JT, Richardson JE, Davison MT & the Mouse Genome Database Group (2000): The Mouse Genome Database (MGD): expanding genetic and genomic resources for the laboratory mouse. *Nucl. Acids Res.* 28: 108–111.
- Bober E, Franz T, Arnold HH, Gruss P & Tremblay P (1994): Pax-3 is required for the development of limb muscles: a possible role for migration of dermomyotomal muscle progenitor cells. *Development* 120: 603–612.
- Boucher Y & Doolittle WF (2000): The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.* 37: 703–716.
- Breitling R & Gerber JK (2000): Origin of the paired domain. *Dev. Genes Evol.* 210: 644–650.
- Breitling R, Krazeisen A, Möller G & Adamski J (2000, im Druck): 17 β -hydroxysteroid dehydrogenase type 7 – an ancient 3-ketosteroid reductase of cholesterologenesis. *Mol. Cell. Endocrinol.*
- Breitling R, Marijanović Z, Perović D & Adamski J (2000, im Druck): Evolution of 17 β -HSD type 4, a multifunctional protein of β -oxidation. *Mol. Cell. Endocrinol.*
- Britten RJ (1994): Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. *Proc. Natl. Acad. Sci. USA.* 91: 6148–6150.
- Brown PO & Botstein D (1999): Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21: 33–37.
- Caira F, Clemencet MC, Cherkaoui-Malki M, Dieuaide-Noubhani M, Pacot C, van Veldhoven PP & Latruffe N (1998): Differential regulation by a peroxisome proliferator of the different multifunctional proteins in guinea pig: cDNA cloning of the guinea pig D-specific multifunctional protein 2. *Biochem. J.* 330: 1361–1368.
- Cassidy A, Bingham S & Setchell KD (1994): Biological effects of a diet of soy protein rich in isoflavones on the menstrual cycle of premenopausal women. *Am. J. Clin. Nutr.* 60: 333–340.
- Catmull J, Hayward DC, McIntyre NE, Reece-Hoyes JS, Mastro R, Callaerts P, Ball EE & Miller DJ (1998): Pax-6 origins – implications from the structure of two coral pax genes. *Dev. Genes Evol.* 208: 352–356.
- Chang BSW & Donoghue MJ (2000): Recreating ancestral proteins. *Trends Ecol. Evol.* 15: 109–114.
- Choinowski T, Hauser H & Piontek K (2000): Structure of sterol carrier protein 2 at 1.8 Å resolution reveals a hydrophobic tunnel suitable for lipid binding. *Biochemistry* 39: 1897–1902.
- Clarke DL, Young KH & Linzer DIH (1993): in Hsueh AJW & Schomburg DW (Hrsg.): Ovarian cell interactions: genes to physiology. New York: Springer: 110–124.
- Corton JC, Bocos C, Moreno ES, Merritt A, Marsman DS, Sausen PJ, Cattley RC & Gustafsson JÅ (1996): Rat 17 β -hydroxysteroid dehydrogenase type IV is a novel peroxisome proliferator-inducible gene. *Mol. Pharmacol.* 50: 1157–1166.
- Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P, Lengieza C, Lew-Smith JE, Lingner C, Roberg-Perez KJ, Tillberg M, Brooks JE & JI Garrels (2000): The Yeast Proteome Database (YPD) and *Caenorhabditis elegans* Proteome Database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucl. Acids Res.* 28: 73–76.
- Czerny T, Schaffner G & Busslinger M (1993): DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.* 7: 2048–2061.
- Dahl E, Koseki H, Balling R (1997): Pax genes and organogenesis. *BioEssays* 19: 755–765.
- de Beer, G (1971): Homology: an unsolved problem. Oxford Biol. Readers No. 11. London, Oxford Univ. Press.

- de Launoit Y, Adamski J (1999): Unique multifunctional HSD17B4 gene product: 17 β -hydroxysteroid dehydrogenase 4 and D-3-hydroxyacyl-coenzyme A dehydrogenase/hydratase involved in Zellweger syndrome. *J. Mol. Endocrinol.* 22: 227–240.
- Delbrück S, Sonneborn A, Gerads M, Grablowitz AH & Ernst JF (1997): Characterization and regulation of the genes encoding ribosomal proteins L39 and S7 of the human pathogen *Candida albicans*. *Yeast* 13: 1199–1210.
- de Queiroz K & Gauthier J (1990): Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. *Syst. Zool.* 39: 307–322.
- de Queiroz K & Gauthier J (1992): Phylogenetic taxonomy. *Annu. Rev. Ecol. Syst.* 23: 449–480.
- Derry JM, Gormally E, Means GD, Zhao W, Meindl A, Kelley RI, Boyd Y & Herman GE (1999): Mutations in a $\Delta 8$ - $\Delta 7$ sterol isomerase in the tattered mouse and X-linked dominant chondrodysplasia punctata. *Nat. Genet.* 22: 286–290.
- Dickson RB & Lippman ME (1987): Estrogenic regulation of growth and polypeptide growth factor secretion in human breast carcinoma. *Endocr. Rev.* 8: 29–43.
- Dieuaide-Noubhani M, Novikov D, Baumgart E, Vanhoooren JC, Franssen M, Goethals M, Vandekerckhove J, van Veldhoven PP & Mannaerts (1996): Further characterization of the peroxisomal 3-hydroxyacyl-CoA dehydrogenase from rat liver. Relationship between the different dehydrogenases and evidence that fatty acids and the C27 bile acids di- and trihydroxycoprostanic acids are metabolized by separate multifunctional proteins. *Eur. J. Biochem.* 240: 660–666.
- Dobzhansky, T (1973): Nothing in biology makes sense except in the light of evolution. *Amer. Biol. Teacher* 35: 125–129.
- Donoghue, MJ (1992): Homology. in Keller EF & Lloyd EA (Hrsg.): *Keywords in evolutionary biology*. Cambridge, Harvard Univ. Press: 170–177.
- Doolittle RF (1981): Similar amino acid sequences: chance or common ancestry? *Science* 214: 149–159.
- Doolittle RF, Feng DF, Tsang S, Cho G & Little E (1996): Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 271: 470–477.
- Dresselhaus T, Cordts S, Heuer S, Sauter M, Lorz H & Kranz E (1999): Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. *Mol. Gen. Genet.* 261: 416–427.
- Duan WR, Linzer DIH & Gibori G (1996): Cloning and characterization of an ovarian-specific protein that associates with the short form of the prolactin receptor. *J. Biol. Chem.* 271: 15602–15607.
- Duan WR, Parmer TG, Albarracin CT, Zhong L & Gibori G (1997): PRAP, a prolactin receptor associated protein: its gene expression and regulation in the corpus luteum. *Endocrinology* 138: 3216–3221.
- Duax WL, Ghosh D & Pletnev V (2000): Steroid dehydrogenase structures, mechanism of action, and disease. *Vitam. Horm.* 58: 121–148.
- Edwards PA & Ericsson J (1999): Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway. *Annu. Rev. Biochem.* 68: 157–185.
- Efron B, Halloran E & Holmes S (1996): Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA.* 93: 7085–7090.
- Eisen MB, Spellman PT, Brown PO & Botstein D (1998): Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA.* 95: 14863–14868.
- Eisenberg D, Marcotte EM, Xenarios I & Yeates TO (2000): Protein function in the post-genomic era. *Nature* 405: 823–826.
- El-Kabbani O, Wilson DK, Petrash M & Quiocho FA (1998): Structural features of the aldose reductase and aldehyde reductase inhibitor-binding sites. *Mol. Vis.* 4: 19.
- Engelman DM, Steitz TA & Goldman A (1986): Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15: 321–353.
- Enright AJ, Iliopoulos I, Kyprides NC & Ouzounis CA (1999): Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86–90.
- Epstein J, Cai J, Glaser T, Jepeal L & Maas R (1994): Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.* 269: 8355–8361.
- Feingold KR, Wiley T, Moser AH, Lear SR & Wiley MH (1983): De novo cholesterologenesis in pregnancy. *J. Lab. Clin. Med.* 101: 256–263.
- Felsenstein J (1985): Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.
- Ferry DR, Smith A, Malkhandi J, Fyfe DW, deTakats PG, Anderson D, Baker J & Kerr DJ (1996): Phase I clinical trial of the flavonoid quercetin: pharmacokinetics and evidence for in vivo tyrosine kinase inhibition. *Clin. Cancer Res.* 2: 659–668.
- Fitch WM & Margoliash E (1967): The construction of phylogenetic trees. *Science* 155: 279–284.
- Fomitcheva J, Baker ME, Anderson E, Lee GY & Aziz N (1998): Characterization of Ke6, a new 17 β -hydroxysteroid dehydrogenase, and its expression in gonadal tissues. *J. Biol. Chem.* 273: 22664–22671.
- Fossa A, Beyer A, Pfitzner E, Wenzel B & Kunau WH (1995): Molecular cloning, sequencing and sequence analysis of the fox-2 gene of *Neurospora crassa* encoding the multifunctional beta-oxidation protein. *Mol. Gen. Genet.* 247: 95–104.
- Fritz S, Fohr KJ, Boddien S, Berg U, Brucker C & Mayerhofer A (1999): Functional and molecular characterization of a muscarinic receptor type and evidence

Referenzen

- for expression of choline-acetyltransferase and vesicular acetylcholine transporter in human granulosa-luteal cells. *J. Clin. Endocrinol. Metabol.* 84: 1744–1750.
- Fritz S, Wessler I, Breitling R, Rossmanith W, Ojeda SR, Dissen GA, Amsterdam A & Mayerhofer A (2001, im Druck): Expression of muscarinic receptor types in the primate ovary and evidence for non-neuronal acetylcholine synthesis. *J. Clin. Endocrinol. Metabol.*
- Fuchs R (2000): Analyse this... or: intelligent help for the rest of us. *Bioinformatics* 16: 491–493.
- Gachotte D, Sen SE, Eckstein J, Barbuch R, Krieger M, Ray BD & Bard M (1999): Characterization of the *Saccharomyces cerevisiae* ERG27 gene encoding the 3-keto reductase involved in C-4 sterol demethylation. *Proc. Natl. Acad. Sci. USA* 96: 12655–12660.
- Galliot I & Miller I (2000): Origin of anterior patterning. How old is our head? *Trends Genet.* 16: 1–5.
- Garcia FL, Szyperski T, Dyer JH, Choinowski T, Seedorf U, Hauser H & Wüthrich K (2000): NMR structure of the sterol carrier protein-2: implications for the biological role. *J. Mol. Biol.* 295: 595–603.
- Gehring WJ (1996): The master control gene for morphogenesis and evolution of the eye. *Genes Cells* 1: 11–15.
- Gehring WJ & Ikeo K (1999): Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet.* 15: 371–377.
- Geissler WM, Davis DL, Wu L, Bradshaw KD, Patel S, Mendonca BB, Eliston KO, Wilson JD, Russell DW & Andersson S (1994): Male pseudohermaphroditism caused by mutations of testicular 17 β -hydroxysteroid dehydrogenase type 3. *Nat. Genet.* 7: 34–39.
- Gesteland RF & Atkins JF (Hrsg.)(1993): The RNA world. Cold Spring Harbor Laboratory Monograph 24.
- Goulding M, Lumsden A & Paquette AJ (1994): Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. *Development* 120: 957–971.
- Grando SA, Crosby AM, Zelickson BD & Dahl MV (1993): Agarose gel keratinocyte outgrowth system as a model for skin re-epithelization: requirement of endogenous acetylcholine for outgrowth initiation. *J. Invest. Dermatol.* 101: 804–810.
- Gröger H, Callaerts P, Gehring W & Schmid V (2000): Characterization and expression analysis of an ancestor-type Pax gene in the hydrozoan jellyfish *Podocoryne carnea*. *Mech. Dev.* 94: 157–169.
- Grundy WN, Bailey TL, Elkan CP, Baker ME (1997): Hidden Markov model analysis of motifs in steroid dehydrogenases and their homologs. *Biochem. Biophys. Res. Comm.* 231: 760–766.
- Gurney JG, Ross JA, Wall DA, Bleyer WA, Severson RK & Robison LL (1997): Infant cancer in the U.S.: histology-specific incidence and trends, 1973 to 1992. *J. Pediatr. Hematol. Oncol.* 19: 428–432.
- Hahn FM, Xuan JW, Chambers AF & Poulter CD (1996): Human isopentenyl diphosphate: dimethylallyl diphosphate isomerase: overproduction, purification, and characterization. *Arch. Biochem. Biophys.* 332: 30–34.
- Halder G, Callaerts P & Gehring WJ (1995): New perspectives on eye evolution. *Curr. Opin. Genet. Dev.* 5: 602–609.
- Hall BK (Hrsg.)(1994): Homology: the hierarchical basis of comparative biology. San Diego. Academic Press.
- Hamdi H, Nishio H, Zielinski R & Dugaiczky A (1999): Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. *J. Mol. Biol.* 289: 861–871.
- Haszprunar G (1991): The types of homology and their significance for evolutionary biology and phylogenetics. *J. Evol. Biol.* 5: 13–24.
- He XY, Merz G, Mehta P, Schulz H & Yang SY (1999): Human brain short chain L-3-hydroxyacyl coenzyme A dehydrogenase is a single-domain multifunctional enzyme. Characterization of a novel 17 β -hydroxysteroid dehydrogenase. *J. Biol. Chem.* 274: 15014–15019.
- He XY, Yang YZ, Schulz H & Yang SY (2000a): Intrinsic alcohol dehydrogenase and hydroxysteroid dehydrogenase activities of human mitochondrial short-chain L-3-hydroxyacyl-CoA dehydrogenase. *Biochem. J.* 345: 139–143.
- He XY, Merz G, Yang YZ, Pullakart R, Mehta P, Schulz H & Yang SY (2000b): Function of human brain short chain L-3-hydroxyacyl coenzyme A dehydrogenase in androgen metabolism. *Biochem. Biophys. Acta* 1484: 267–277.
- Henderson BE & Feigelson HS (2000): Hormonal carcinogenesis. *Carcinogenesis* 21: 427–433.
- Hennig, W (1982): Phylogenetische Systematik. Pareys Studentexte 34. Berlin, Parey.
- Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G & Brown EL (2000): Genomic analysis of gene expression in *C. elegans*. *Science* 290: 809–812.
- Hiltunen JK, Wenzel B, Beyer A, Erdmann R, Fosså & Kunau WH (1992): Peroxisomal multifunctional β -oxidation protein of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 267: 6646–6653.
- Holm L & Sander C (1996): The FSSP database: fold classification based on structure-structure alignments of proteins. *Nucl. Acids Res.* 24: 206–209.
- Hooft RWW, Vriend G, Sander C & Abola EE (1996): Errors in protein structures. *Nature*: 381: 272.
- Hoshiyama D, Suga H, Iwabe N, Koyanagi M, Nikoh N, Kuma K, Matsuda F, Honjo T & Miyata T (1998): Sponge Pax cDNA related to Pax-2/5/8 and ancient gene duplications in the pax family. *J. Mol. Evol.* 47: 640–648.
- Huang CC, Novak WR, Babbitt PC, Jewett AI, Ferrin TE & Klein TE (2000): Integrated tools for structural and sequence alignment and analysis. *Pac. Symp. Biocomp.* 2000: 230–241.

- Huang H, Ball JM, Billheimer JT & Schroeder F (1999): The sterol carrier protein-2 amino terminus: a membrane interaction domain. *Biochemistry* 38: 13231–13243.
- Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburttu K, Simon J, Bard M & Friend SH (2000): Functional discovery via a compendium of expression profiles. *Cell* 102: 109–126.
- Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S & Sakaki Y (2000): Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* 97: 1143–1147.
- Itoh M, Suzuki Y & Takashima S (1999): A novel peroxisomal enzyme, D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein: its expression in the developing human brain. *Microsc. Res. Tech.* 45: 383–388.
- Ivics Z, Izsvák Z, Minter A & Hackett PB (1996): Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl. Acad. Sci. USA* 93: 5008–5013.
- Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J jr., Boguski MS, Lashkari D, Shalon D, Botstein D & Brown PO (1999): The transcriptional program in the response of human fibroblasts to serum. *Science* 283: 83–87.
- Jez JM, Bennett MJ, Schlegel BP, Lewis M & Penning TM (1997): Comparative anatomy of the aldo-keto reductase superfamily. *Biochem. J.* 326: 625–636.
- Jez JM & Penning TM (2000): The aldo-keto reductase superfamily: an update. Version 2.0. 14. August 2000.
- Jiang LL, Miyazawa S & Hashimoto T (1996): Purification and properties of rat D-3-hydroxyacyl-CoA dehydratase: D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein. *J. Biochem. (Tokyo)* 120: 633–641.
- Jörnvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffrey J & Ghosh D (1995): Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 34: 6003–6013.
- Jörnvall H (1999): Multiplicity and complexity of SDR and MDR enzymes. *Adv. Exp. Med. Biol.* 463: 359–364.
- Jörnvall H, Hoog JO & Persson B (1999): SDR and MDR: complete genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* 445: 261–264.
- Jun S, Wallen RV, Goriely A, Kalionis B & Desplan C (1998): *Lune/eye gone*, a Pax-like protein, uses a partial paired domain and a homeodomain for DNA recognition. *Proc. Natl. Acad. Sci. USA* 95: 13720–13725.
- Jurevics HA, Kidwai FZ & Morell P (1997): Sources of cholesterol during development of the rat fetus and fetal organs. *J. Lipid Res.* 38: 723–733.
- Kappen C (2000): Analysis of a complete homeobox gene repertoire: implications for the evolution of diversity. *Proc. Natl. Acad. Sci. USA* 97: 4481–4486.
- Kawashima K, Fujii T, Watanabe Y & Misawa H (1998): Acetylcholine synthesis and muscarinic receptor subtype mRNA expression in T-lymphocytes. *Life Sci.* 62: 1701–1705.
- Kawashima K & Fujii T (2000): Extraneuronal cholinergic system in lymphocytes. *Pharmacol. Ther.* 86: 29–48.
- Kelley RI (2000): Inborn errors of cholesterol biosynthesis. *Adv. Pediatr.* 47: 1–53.
- Kishi T, Sasaki H, Akiyama N, Ishizuka T, Sakamoto H, Aizawa S, Sugimura T & Terada M (1997): Molecular cloning of human GRB-7 co-amplified with CAB1 and c-ERBB-2 in primary gastric cancer. *Biochem. Biophys. Res. Com.* 232: 5–9.
- Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB & Pabo CO (1990): Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 63: 579–590.
- Klingenhoff A, Frech K, Quandt K & Werner T (1999): Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. *Bioinformatics* 15: 180–186.
- Kobayashi K, Kobayashi H, Ueda M & Honda Y (1997): Expression of 17 β -hydroxysteroid dehydrogenase type IV in chick retinal pigment epithelium. *Exp. Eye Res.* 64: 719–726.
- König A, Happel R, Bornholdt D, Engel H & Grzeschik KH (2000): Mutations in the NSDHL gene, encoding a 3 β -hydroxysteroid dehydrogenase, cause CHLD syndrome. *Am. J. Med. Genet.* 90: 339–346.
- Korostensky C & Gonnet GH (2000): Using traveling salesman problem algorithms for evolutionary tree construction. *Bioinformatics* 16: 619–627.
- Koseki H, Wallin J, Wilting J, Mizutani Y, Kispert A, Ebensperger C, Hermann BG, Christ B & Balling R (1993): A role for Pax-1 as a mediator of notochordal signals during dorsoventral specification of vertebrates. *Development* 119: 649–660.
- Kozmik Z, Czerny T & Busslinger M (1997): Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *EMBO J.* 16: 6793–6803.
- Krazeisen A, Breitling R, Imai K, Fritz S, Möller G & Adamski J (1999): Determination of cDNA, gene structure and chromosomal localization of the novel human 17 β -hydroxysteroid dehydrogenase type 7. *FEBS Lett.* 460: 373–379.
- Krazeisen A, Breitling R, Möller G & Adamski J (2000, im Druck): Phytoestrogens inhibit human 17 β -hydroxysteroid dehydrogenase type 5. *Mol. Cell. Endocrinol.*

Referenzen

- Krozowski Z (1994): The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *J. Steroid Biochem. Mol. Biol.* 51: 125–130.
- Labrie F, Luu-The V, Lin SX, Simard J, Labrie C, El-Alfy M, Pelletier G & Belanger A (2000): Intracrinology: role of the family of 17β -hydroxysteroid dehydrogenases in human physiology and disease. *J. Mol. Endocrinol.* 25: 1–16.
- Lange BM, Rujan T, Martin W & Croteau R (2000): Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. USA.* 97: 13172–13177.
- Lauder JM & Schambra UB (1999): Morphogenetic roles of acetylcholine. *Environ. Health Perspect.* 107: 65–69.
- Leenders F, Adamski J, Husen B, Thole HH & Jungblut PW (1994): Molecular cloning and amino acid sequence of the porcine 17β -estradiol dehydrogenase. *Eur. J. Biochem.* 222: 221–227.
- Lévi-Strauss, C (1947): Les structures élémentaires de la parenté. Paris. Presses Universitaire de France.
- Liu LX, Spoerke JM, Mulligan EL, Chen J, Reardon B, Westlund B, Sun L, Abel K, Armstrong B, Hardiman G, King J, McCague L, Basson M, Clover R & Johnson CD (1999): High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* 9: 859–867.
- Liu XY, Dangel AW, Kelley RI, Zhao W, Denny P, Botcherby M, Cattanaach B, Peters J, Hunsicker PR, Mallon AM, Strivens MA, Bate R, Miller W, Rhodes M, Brown SD & Herman GE (1999): The gene mutated in bare patches and striated mice encodes a novel 3β -hydroxysteroid dehydrogenase. *Nat. Genet.* 22: 182–187.
- Lockhart DJ & Winzeler EA (2000): Genomics, gene expression and DNA arrays. *Nature* 405: 827–836.
- Lockless SW & Ranganathan R (1999): Evolutionary conserved pathways of energetic connectivity in protein families. *Science* 286: 295–299.
- Luu-The V, Labrie C, Zhao HF, Couët J, Lachance Y, Simard J, Leblanc G, Côté J, Bérubé D, Gagné R & Labrie F (1989): Characterization of cDNAs for human estradiol 17β -dehydrogenase and assignment of the gene to chromosome 17: evidence of two mRNA species with distinct 5'-termini in human placenta. *Mol. Endocrinol.* 3: 1301–1309.
- Luu-The V, Labrie C, Simard J, Lachance Y, Zhao HF, Couët J, Leblanc G & Labrie F (1990): Structure of two in tandem human 17β -hydroxysteroid dehydrogenase genes. *Mol. Endocrinol.* 4: 268–275.
- Macchia PE, Lapi P, Krude H, Pirro MT, Missero C, Chiovato L, Souabni A, Baserga M, Tassi V, Pinchera A, Fenzi G, Gruters A, Busslinger M & Di Lauro R (1998): PAX8 mutations associated with congenital hypothyroidism caused by thyroid dysgenesis. *Nat. Genet.* 19: 83–86.
- Mansouri A (1998): The role of Pax3 and Pax7 in development and cancer. *Crit. Rev. Oncog.* 9: 141–149.
- Mansouri A, Goudreau G & Gruss P (1999): Pax genes and their role in organogenesis. *Cancer Res.* 59: 1707s–1709s.
- Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO & Eisenberg D (1999): A combined algorithm for genome-wide prediction of protein function. *Nature* 402: 83–86.
- Messina MJ, Persky V, Setchell KD & Barnes S (1994): Soy intake and cancer risk: a review of the in vitro and in vivo data. *Nutr. Cancer* 21: 113–131.
- Miettinen MM, Mustonen MV, Poutanen MH, Isomaa VV & Vihko RK (1996): Human 17β -hydroxysteroid dehydrogenase type 1 and type 2 isoenzymes have opposite activities in cultured cells and characteristic cell- and tissue specific expression. *Biochem. J.* 314: 839–845.
- Miettinen M, Mustonen M, Poutanen M, Isomaa V, Wickman M, Soderqvist G, Vihko R & Vihko P (1999): 17β -hydroxysteroid dehydrogenases in normal human mammary epithelial cells and breast tissue. *Breast Cancer Res. Treat.* 57: 175–182.
- Miller DJ, Hayward DC, Reece-Hoyes JS, Scholten I, Catmull J, Gehring WJ, Callaerts P, Larsen JE & Ball EE (2000): Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. *Proc. Natl. Acad. Sci. USA.* 97: 4475–4480.
- Miyazaki M, Kim YC, Gray-Keller MP, Attie AD & Ntambi JM (2000): The biosynthesis of hepatic cholesterol esters and triglycerides is impaired in mice with a disruption of the gene for stearoyl-CoA desaturase 1. *J. Biol. Chem.* 275: 30132–30138.
- Moebius FF, Fitzky BU & Glossmann I (2000): Genetic defects in postsqualene cholesterol biosynthesis. *Trends Endocrinol. Metab.* 11: 106–114.
- Moog-Lutz C, Tomasetto C, Regnier CH, Wendling C, Lutz Y, Muller D, Chenard MP, Bassett P & Rio MC (1997): MLN64 exhibits homology with the steroidogenic acute regulatory protein (STAR) and is over-expressed in human breast carcinomas. *Int. J. Cancer* 71: 183–191.
- Morell P & Jurevics H (1996): Origin of cholesterol in myelin. *Neurochem. Res.* 21: 463–470.
- Mustonen MV, Poutanen MH, Kellomumpu S, de Lannoit Y, Isomaa VV, Vihko RK & Vihko PT (1998): Mouse 17β -hydroxysteroid dehydrogenase type 2 mRNA is predominantly expressed in hepatocytes and in surface epithelial cells of the gastrointestinal and urinary tracts. *J. Mol. Endocrinol.* 20: 67–74.
- Nishizawa M, Nakajima T, Yasuda K, Kanzaki H, Sasaguri Y, Watanabe K & Ito S (2000): Close kinship of human 20α -hydroxysteroid dehydrogenase gene with three aldo-keto reductase genes. *Genes Cells.* 5: 111–125.
- Nokelainen P, Puranen T, Peltoketo H, Orava M, Vihko P & Vihko R (1996): Molecular cloning of mouse 17β -hydroxysteroid dehydrogenase type 1 and characterization of enzyme activity. *Eur. J. Biochem.* 236: 482–490.
- Nokelainen P, Peltoketo H, Vihko R & Vihko P (1998): Expression cloning of a novel estrogenic mouse

- 17 β -hydroxysteroid dehydrogenase/17-ketosteroid reductase (m17HSD7), previously described as a prolactin receptor-associated protein (PRAP) in rat. *Mol. Endocrinol.* 12: 1048–1059.
- Nokelainen P, Peltoketo H, Mustonen M & Vihko P (2000): Expression of mouse 17 β -hydroxysteroid dehydrogenase/17-ketosteroid reductase type 7 in ovary, uterus, and placenta: localization from implantation to late pregnancy. *Endocrinology* 141: 772–778.
- Normand T, Husen B, Leenders F, Pelczar H, Baert JL, Begue A, Flourens AC, Adamski J & de Launoit Y (1995): Molecular characterization of mouse 17 β -hydroxysteroid dehydrogenase IV. *J. Steroid Biochem. Mol. Biol.* 55: 541–548.
- Novikov D, Dieuaide-Noubhani M, Vermeersch JR, Fournier B, Mannaerts GP & van Veldhoven PP (1997): The human peroxisomal multifunctional protein involved in bile acid synthesis: activity measurement, deficiency in Zellweger syndrome and chromosome mapping. *Biochem. Biophys. Acta* 1360: 229–240.
- Nuttley WM, Aitchison JD & Rachubinski RA (1988): cDNA cloning and primary structure determination of the peroxisomal trifunctional enzyme hydratase-dehydrogenase-epimerase from the yeast *Candida tropicalis* pK233. *Gene* 69: 171–180.
- Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y & Matsubara K (1992): Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 2: 173–179.
- O'huigin C & Li WH (1992): The molecular clock ticks regularly in murid rodents and hamsters. *J. Mol. Evol.* 35: 377–384.
- Owen, R (1843): Lectures on the comparative anatomy and physiology of the invertebrate animals. London, Longman, Brown, Green, & Longmans: 379.
- Parkin DM, Stiller CA, Draper GJ, Bieber CA, Terracini B & Young JI (Hrsg.) (1988): International incidence of childhood cancer, IARC scientific publications, No. 87. Lyon. W.H.O.
- Parmer TG, McLean MP, Duan WR, Nelson SE, Albarraçin CT, Khan I & Gibori G (1992): Hormonal and immunological characterization of the 32 kilodalton ovarian-specific protein. *Endocrinology* 131: 2213–2221.
- Peitsch MC, Wilkins MR, Tonella L, Sanchez JC, Appel RD & Hochstrasser DF (1997): Large-scale protein modelling and integration with the SWISS-PROT and SWISS-2DPAGE databases: the example of *Escherichia coli*. *Electrophoresis* 18: 498–501.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D & Yeates TO (1999): Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA.* 96: 4285–4288.
- Peltoketo H, Isomaa V, Mäentausta O & Vihko R (1988): Complete amino acid sequence of human placental 17 β -hydroxysteroid dehydrogenase deduced from cDNA. *FEBS Lett.* 239: 73–77.
- Peltoketo H, Nokelainen P, Piao YS, Vihko R & Vihko P (1999a): Two 17 β -hydroxysteroid dehydrogenases (17HSDs) of estradiol biosynthesis: 17HSD type 1 and type 7. *J. Steroid. Biochem. Mol. Biol.* 69: 431–439.
- Peltoketo H, Luu-The V, Simard J & Adamski J (1999b): 17 β -hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. *J. Mol. Endocrinol.* 23: 1–11.
- Penning TM, Bennett MJ, Smith-Hoog S, Schlegel BP, Jez JM & Lewis M (1997): Structure and function of 3 α -hydroxysteroid dehydrogenase. *Steroids* 62: 101–111.
- Penning TM, Burczynski ME, Jez JM, Hung CF, Lin HK, Ma H, Moore M, Palackal N & Ratnam K (2000): Human 3 α -hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem. J.* 351: 67–77.
- Peters H, Neubüser A, Kratochwil K & Balling R (1998): Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev.* 12: 2735–2747.
- Persson B, Krook M & Jörnvall H (1991): Characteristics of short-chain alcohol dehydrogenases and related enzymes. *Eur. J. Biochem.* 200: 537–543.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO & Botstein D (2000): Molecular portraits of human breast tumours. *Nature* 406: 747–752.
- Piao YS, Peltoketo H, Vihko P & Vihko R (1997): The proximal promoter region of the gene encoding human 17 β -hydroxysteroid dehydrogenase type 1 contains GATA, AP-2, and Sp1 response elements: analysis of promoter function in choriocarcinoma cells. *Endocrinology* 138: 3417–3425.
- Pietta PG (2000): Flavonoids as antioxidants. *J. Nat. Prod.* 63: 1035–1042.
- Plasterk RHA, Izsvák Z & Ivics Z (1999): Resident aliens – the Tc1/mariner superfamily of transposable elements. *Trends Genet.* 15: 326–332.
- Poleev A, Okladnova O, Musti AM, Schneider S, Royer-Pokora B & Plachov D (1997): Determination of functional domains of the human transcription factor PAX8 responsible for its nuclear localization and transactivating potential. *Eur. J. Biochem.* 247: 860–869.
- Popper KR (1972): *Objective Knowledge*. Oxford. Clarendon.
- Poutanen M, Isomaa V, Peltoketo H & Vihko R (1995): Regulation of oestrogen action: role of 17 β -hydroxysteroid dehydrogenases. *Ann. Med.* 27: 675–682.
- Qin KN, New MI & Cheng KC (1993): Molecular cloning of multiple cDNAs encoding human enzymes

Referenzen

- structurally related to 3 α -hydroxysteroid dehydrogenase. *J. Steroid Biochem. Mol. Biol.* 46: 637–679.
- Qin YM, Poutanen MH, Helander HM, Kvist AP, Siivari KM, Schmitz W, Conzelmann E, Hellman U & Hiltunen JK (1997): Peroxisomal multifunctional enzyme of β -oxidation metabolizing D-3-hydroxyacyl-CoA esters in rat liver: molecular cloning, expression and characterization. *Biochem. J.* 321: 21–28.
- Qin YM, Haapalainen AM, Conry D, Cuebas DA, Hiltunen JK & Novikov DK (1997): Recombinant 2-enoyl-CoA hydratase derived from rat peroxisomal enzyme 2: role of the hydratase reaction in bile acid synthesis. *Biochem. J.* 328: 377–382.
- Qin YM, Marttila MS, Haapalainen AM, Siivari KM, Glumoff T & Hiltunen JK (1999): Yeast peroxisomal multifunctional enzyme: (3R)-hydroxyacyl CoA dehydrogenase domains A and B are required for optimal growth on oleic acid. *J. Biol. Chem.* 274: 28619–28625.
- Qin YM, Poutanen MH & Novikov DK (2000a): Substrate specificities of peroxisomal members of short-chain alcohol dehydrogenase superfamily: expression and characterization of dehydrogenase part of *Candida tropicalis* multifunctional enzyme. *J. Lipid Res.* 41: 93–98.
- Qin YM, Haapalainen AM, Kilpelainen SH, Marttila MS, Koski MK, Glumoff T, Novikov DK & Hiltunen JK (2000b): Human peroxisomal multifunctional enzyme type 2. Site directed mutagenesis studies show the importance of two protic residues for 2-enoyl CoA hydratase 2 activity. *J. Biol. Chem.* 275: 4965–4972.
- Quandt K, Frech K, Karas H, Wingender E & Werner T (1995): MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl. Acids Res.* 23: 4878–4884.
- Redecker D, Kodner R, Graham LE (2000): Glomalean fungi from the Ordovician. *Science* 289: 1920–1921.
- Requena N, Fuller P & Franken P (1999): Molecular characterization of GmFOX2, an evolutionary highly conserved gene from the mycorrhizal fungus *Glomus mosseae*, down-regulated during interaction with rhizobacteria. *Mol. Plant Microbe Interact.* 12: 934–942.
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D & Brown PO (2000): Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24: 227–235.
- Roux C, Wolf C, Mulliez N, Gaoua W, Cormier V, Chevy F & Citadelle D (2000): Role of cholesterol in embryonic development. *Am. J. Clin. Nutr.* 71: 1270S–1279S.
- Sanyanusin P, Schimmenti LA, McNoe LA, Ward TA, Pierpont ME, Sullivan MJ, Dobyns WB & Eccles MR (1995): Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat. Genet.* 9: 358–364.
- Sawicki MW, Erman M, Pouranen T, Vihko P & Ghosh D (1999): Structure of the ternary complex of human 17 β -hydroxysteroid dehydrogenase type 1 with 3-hydroxyestra-1,3,5,7-tetraen-17-one (equilin) and NADP⁺. *Proc. Natl. Acad. Sci. USA* 96: 840–845.
- Scherf M, Klingenhoff A & Werner T (2000): Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* 297: 599–606.
- Schrick K, Mayer U, Horrichs A, Kuhnt C, Bellini C, Dangl J, Schmidt J & Jürgens G (2000): FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. *Genes Dev.* 14: 1471–1484.
- Schwabe I, Husen B & Einspanier A (2000, im Druck): Expression of the estradiol-synthesizing 17 β -hydroxysteroid dehydrogenases type 1 and type 7 in the nonhuman primate *Callithrix jacchus*. *Mol. Cell. Endocrinol.*
- Seedorf U, Ellinghaus P & Roch Nofer J (2000): Sterol carrier protein-2. *Biochim. Biophys. Acta* 1486: 45–54.
- Seidel G & Prohaska R (1998): Molecular cloning of hSLP-1, a novel human brain-specific member of the band 7/MEC-2 family similar to *Caenorhabditis elegans* UNC-24. *Gene* 225: 23–29.
- Sibley CG & Ahlquist JE (1987): DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. *J. Mol. Evol.* 26: 99–121.
- Smit A & Mushegian A (2000): Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome Res.* 10: 1468–1484.
- Smith JJ, Brown TW, Eitzen GA & Rachubinski RA (2000): Regulation of peroxisome size and number by fatty acid beta-oxidation in the yeast *Yarrowia lipolytica*. *J. Biol. Chem.* 275: 20168–20178.
- Smith V, Chou KN, Lashkari D, Botstein D & Brown PO (1996): Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* 274: 2069–2074.
- Sosa-Pineda B, Chowdhury K, Torres M, Oliver G & Gruss P (1997): The Pax4 gene is essential for differentiation of insulin-producing beta-cells in the mammalian pancreas. *Nature* 386: 399–402.
- Spicer LJ, Hamilton TD & Keefer BE (1996): Insulin-like growth factor I enhancement of steroidogenesis by bovine granulosa cells and thecal cells: dependence on de novo cholesterol synthesis. *J. Endocrinol.* 151: 365–373.
- Stapleton AE & Walbot V (1994): Flavonoids can protect maize DNA from the induction of ultraviolet radiation damage. *Plant Physiol.* 105: 881–889.
- Stockton DW, Das P, Goldenberg M, D'Souza RN & Patel PI (2000): Mutation of PAX9 is associated with oligodontia. *Nat. Genet.* 24: 18–19.

- Stolowich N, Frolov A, Petrescu AD, Scott AI, Billheimer JT & Schroeder F (1999): Holo-sterol carrier protein-2. ^{13}C NMR investigation of cholesterol and fatty acid binding sites. *J. Biol. Chem.* 274: 35425–35433.
- St-Onge L, Sosa-Pineda B, Chowdhury K, Mansouri A & Gruss P (1997): Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature* 387: 406–409.
- Strick R, Strissel PL, Borgers S, Smith SL & Rowley JD (2000): Dietary bioflavonoids induce cleavage in the MLL gene and may contribute to infant leukemia. *Proc. Natl. Acad. Sci. USA.* 97: 4790–4795.
- Strimmer K & von Haeseler A (1997): Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA.* 94: 6815–6819.
- Su J, Lin M & Napoli JL (1999): Complementary deoxyribonucleic acid cloning and enzymatic characterization of a novel $17\beta/3\alpha$ -hydroxysteroid/retinoid short chain dehydrogenase/reductase. *Endocrinology* 140: 5275–5284.
- Sun H, Rodin A, Zhou Y, Dickinson DP, Harper DE, Hewett-Emmett D & Li WH (1997): Evolution of paired domains: isolation and sequencing of jellyfish and hydra Pax genes related to Pax-5 and Pax-6. *Proc. Natl. Acad. Sci. USA.* 94: 5156–5161.
- Suzuki Y, Jiang LL, Souri M, Miyazawa S, Fukada S, Zhang Z, Une M, Shimozawa N, Kondo N, Orii T & Hashimoto T (1997): D-3-hydroxyacyl-CoA dehydratase/D-3-hydroxyacyl-CoA dehydrogenase bifunctional protein deficiency: A newly identified peroxisomal disorder. *Am. J. Hum. Genet.* 61: 1153–1162.
- Takeyama J, Suzuki T, Hirasawa G, Muramatsu Y, Nagura H, Iinuma K, Nakamura J, Kimura KI, Yoshihama M, Harada N, Andersson S & Sasano H (2000): 17β -hydroxysteroid dehydrogenase type 1 and 2 expression in the human fetus. *J. Clin. Endocrinol. Metabol.* 85: 410–416.
- Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P & Strachan T (1992): Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature.* 355: 635–636.
- Tomarev SI, Callaerts P, Kos L, Zinovieva R, Halder G, Gehring W & Piatyogorsky J (1997): Squid Pax-6 and eye development. *Proc. Natl. Acad. Sci. USA.* 94: 2421–2426.
- Torres M, Gomez-Pardo E & Gruss P (1996): Pax2 contributes to inner ear patterning and optic nerve trajectory. *Development* 122: 3381–3391.
- Torroja L, Ortuna-Sahagun D, Ferrus A, Hammerle B & Barbas JA (1998): scully, an essential gene of *Drosophila*, is homologous to mammalian mitochondrial type II L-3-hydroxyacyl-CoA dehydrogenase/amyloid-beta peptide-binding protein. *J. Cell Biol.* 141: 1009–1017.
- Tremblay P, Dietrich S, Mericskay M, Schubert FR, Li Z & Paulin D (1998): A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. *Dev. Biol.* 203: 49–61.
- Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamar G, Yang M, Johnston M, Fields S & Rothberg JM (2000): A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623–627.
- Urbanek P, Wang ZQ, Fetka I, Wagner EF & Busslinger M (1994): Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* 79: 901–912.
- Van den Sande K & Bisseling T (1997): Signalling in symbiotic root nodule formation. *Essays Biochem.* 32: 127–142.
- Van Grunsven EG, van Berkel E, Ijlst L, de Klerk JB, Adamski J, Lemond H, Clayton PT, Cuebas DA & Wanders RJ (1998): Peroxisomal D-hydroxyacyl-CoA dehydrogenase deficiency: resolution of the enzyme defect and its molecular basis in bifunctional protein deficiency. *Proc. Natl. Acad. Sci. USA.* 95: 2128–2133.
- Van Grunsven EG, van Berkel E, Mooijer PA, Watkins PA, Moser HW, Suzuki Y, Jiang LL, Hashimoto T, Hoefler G, Adamski J & Wanders RJ (1999): Peroxisomal bifunctional protein deficiency revisited: resolution of its true enzymatic and molecular basis. *Am. J. Hum. Genet.* 64: 99–107.
- Van Pouderoyen G, Ketting RF, Perakis A, Plasterk RH & Sixma TK (1997): Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J.* 16: 6044–6054.
- Van Rhijn P & Vanderleyden J (1995): The Rhizobium-plant symbiosis. *Microbiol. Rev.* 59: 124–142.
- Velculescu VE, Zhang L, Vogelstein B & Kinzler KW (1995): Serial analysis of gene expression. *Science* 270: 484–487.
- Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM, Cook BP, Dufault MR, Ferguson AT, Gao Y, He TC, Hermeking H, Hiraldo SK, Hwang PM, Lopez MA, Luderer HF, Mathews B, Petroziello JM, Polyak K, Zavel L, Zhang W, Zhang X, Zhou W, Haluska FG, Jen J, Sukumar S, Landes GM, Riggins GJ, Vogelstein B & Kinzler KW (2000): Analysis of human transcriptomes. *Nat. Genet.* 23: 387–388.
- Vogan KJ, Underhill DA & Gros P (1996): An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol. Cell Biol.* 16: 6677–6686.
- Vogan KJ & Gros P (1997): The C-terminal subdomain makes an important contribution to the DNA binding activity of the Pax-3 paired domain. *J. Biol. Chem.* 272: 28289–28295.
- Vukmirovic OG & Tilghman SM (2000): Exploring genome space. *Nature* 405: 820–822.

Referenzen

- Wägele JW, Erikson T, Lockhart P, Misof B (1999): The Ecdysozoa: Artifact or monophylum? *J. Zool. Syst. Evol. Research* 37: 211–223.
- Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N & Vidal M (2000): Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287: 116–122.
- Watari H, Arakane F, Moog-Lutz C, Kallen CB, Tomasetto C, Gerton GL, Rio MC, Baker ME & Strauss JF 3rd (1997): MLN64 contains a domain with homology to the steroidogenic acute regulatory protein (StAR) that stimulates steroidogenesis. *Proc. Natl. Acad. Sci. USA*. 94: 8462–8467.
- Wei H, Bowen R, Cai Q, Barnes S & Wang Y (1995): Antioxidant effects of the soybean isoflavone genistein. *Proc. Soc. Exp. Biol. Med.* 208: 124–130.
- Wessler I, Kirkpatrick CJ & Racke K (1998): Non-neuronal acetylcholine, a locally acting molecule, widely distributed in biological systems: expression and function in humans. *Pharmacol. Ther.* 77: 59–79.
- Wessler I, Kirkpatrick CJ & Racke K (1999): The cholinergic ‘pitfall’: acetylcholine, a universal cell molecule in biological systems, including humans. *Clin. Exp. Pharmacol. Physiol.* 26: 198–205.
- Wilm B, Dahl E, Peters H, Balling R & Imai K (1998): Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proc. Natl. Acad. Sci. USA*. 95: 8692–8697.
- Wouters FS, Bastiaens PIH, Wirtz KWA & Jovin TM (1998): FRET microscopy demonstrates molecular association of non-specific lipid transfer protein (nsL-TP) with fatty acid oxidation enzymes in peroxisomes. *EMBO J.* 17: 7179–7189.
- Wu CI & Li WH (1985): Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA*. 82: 1741–1745.
- Wu L, Einstein M, Geissler WM, Chan HK, Elliston KO & Andersson S (1993): Expression cloning and characterization of human 17 β -hydroxysteroid dehydrogenase type 2, a microsomal enzyme possessing 20 α -hydroxysteroid dehydrogenase activity. *J. Biol. Chem.* 268: 12964–12969.
- Xu W, Rould MA, Jun S, Desplan & Pabo CO (1995): Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80: 639–650.
- Xu HE, Rould MA, Xu W, Epstein JA, Maas RL & Pabo CO (1999): Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding.
- Yanagihara K, Ito A, Toge T & Numoto M (1993): Antiproliferative effects of isoflavones on human cancer cell lines established from the gastrointestinal tract. *Cancer Res.* 53: 5815–5821.
- Yona G, Linial N & Linial M (2000): ProtoMap: automatic classification of protein sequences and hierarchy of protein families. *Nucl. Acids Res.* 28: 49–55.
- Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B & Kinzler KW (1997): Gene expression in normal and cancer cells. *Science* 276: 1268–1272.
- Zhang Y, Word RA, Fesmire S, Carr BR & Rainey WE (1996): Human ovarian expression of 17 β -hydroxysteroid dehydrogenase types 1, 2, and 3. *J. Clin. Endocrinol. Metab.* 81: 3594–3598.
- Zuckerkindl E & Pauling L (1965): Molecules as documents of evolutionary history. *J. Theor. Biol.* 8: 357–366.

Eigene Veröffentlichungen

Zeitschriftenartikel

Breitling R, Krazeisen A, Möller G, Adamski J (2000, im Druck): 17β -hydroxysteroid dehydrogenase type 7 — an ancient 3-ketosteroid reductase of cholesterologenesis. *J. Steroid Biochem. Mol. Biol.*

Breitling R, Marijanovic Z, Perovic D, Adamski J (2000, im Druck): Evolution of 17β -HSD type 4, a multifunctional protein of β -oxidation. *J. Steroid Biochem. Mol. Biol.*

Krazeisen A, **Breitling R**, Möller G, Adamski J (2000, im Druck): Phytoestrogens inhibit human 17β -hydroxysteroid dehydrogenase type 5. *J. Steroid Biochem. Mol. Biol.*

Fritz S, Wessler I, **Breitling R**, Rossmannith W, Ojeda SR, Dissen GA, Amsterdam A, Mayerhofer A (2000, im Druck): Expression of muscarinic receptor types in the primate ovary and evidence for non-neuronal acetylcholine synthesis. *J. Clin. Endocrinol. Metabol.*

Breitling R & Gerber JK (2000): Origin of the paired domain. *Dev. Genes Evol.* **210**: 644–650.

Krazeisen A, **Breitling R**, Imai K, Fritz S, Möller G, Adamski J (1999): Determination of cDNA, gene structure and chromosomal localization of the novel human 17β -hydroxysteroid dehydrogenase type 7. *FEBS Lett.* **460**: 373–379.

Datenbankeinträge

Protein	Spezies	Sequenztyp	Zugangsnummer
17β -HSD Typ 7	Mensch	Protein	NP_057455
			P56937
		mRNA	AAF14537
			AF098786
			NM_016371
Teile des Gens	AF162759		
	bis AF162767		
Muscarinrezeptor M5R	Macaca	Teile der mRNA	AF285237
		Protein	AAG00509
Isopentenyl-Isomerase	Mensch	zwei Genkopien	AF291755

Referenzen

Abbildungsverzeichnis

3.1	Alternative Evolutions-Szenarien für die Pax-Familie	24
3.2	Stammbäume der Pax-Familie	25
3.3	Proteinalignment von Pax-Proteinen und Transposasen	28
3.4	Neighbour-joining-Stammbaum der Pax-Domänen und der DNA-Bindungsdomänen aus Transposasen	29
3.5	Parsimony-Stammbaum der Pax-Domänen und der DNA-Bindungsdomänen aus Transposasen	30
3.6	<i>Four-cluster likelihood mapping</i> von Pax-Proteinen und Transposasen	32
3.7	<i>Maximum likelihood</i> -Stammbaum der ChAT/CRAT-Familie	36
3.8	Stammbäume der ChAT/CRAT-Unterfamilie	37
4.1	Alignment der Proteinsequenzen von kristallisierten HSD-Proteinen	45
4.2	Alignment der Proteinsequenzen von kristallisierten HSD-Proteinen-Teil2	46
4.3	Struktur-basierte Stammbäume der HSD-Familie 1	49
4.4	Struktur-basierte Stammbäume der HSD-Familie 2	50
4.5	Struktur-basierte Stammbäume der HSD-Familie 3	51
4.6	Struktur-basierte Stammbäume der HSD-Familie 4	51
4.7	Struktur-basierte Stammbäume der HSD-Familie 5	52
4.8	Phylogenetischer Gehalt des Struktur-basierten Alignments	53
4.9	Konsensus-Stammbaum der HSD-Familie	55
4.10	Konsensus-Stammbaum der HSD-Familie einschließlich der 17β -HSDs	57
4.11	<i>in silico</i> Northern Blot von 17β -HSD Typ 1, 2 und 7	73
4.12	Reaktionen, die von 17β -HSD7 und seinem Hefe-Homologen Erg27p katalysiert werden	74
4.13	Stammbaum der 17β -HSD7 und ihrer nächsten Verwandten	74
4.14	Vorhersage von Transmembranbereichen	75
4.15	Alignment der Proteinsequenzen von 17β -HSD1 und 7	76
4.16	Modell der Proteinstruktur von 17β -HSD7	77
4.17	Expression von IDI1 in Fibroblasten	78
4.18	IDI1-Promotor-Modell und Genstruktur für IDI1, MLN64, OSR1, RPL39 und 17β -HSD7	79

Abbildungsverzeichnis

4.19	Reaktionen der Cholesterin-Biosynthese	80
4.20	Reaktionen der peroxisomalen β -Oxidation	82
4.21	Genstrukturen von 17β -HSD4 und ihren Orthologen	84
4.22	Stammbäume der Proteindomänen von 17β -HSD4-Verwandten	86
4.23	Evolutionsszenario der 17β -HSD4-Verwandten	89
4.24	Vergleich der Genstrukturen der 17β -HSD1 in Menschen und Nagetieren	92
4.25	Struktur der AKR-Proteine	95
4.26	Vergleich der Strukturen von AKR-Proteinen	100
4.27	Strukturänderungen der 3α -HSD nach Ligandenbindung	105
4.28	Sequenzkonservierung in der AKR-Familie	106
4.29	Modell des aktiven Zentrums der 17β -HSD5	107

Tabellenverzeichnis

3.1	Hauptfunktionen der humanen Paired-Box-Proteine.	23
4.1	Gewebeverteilung der 17β -HSDs der Säugetiere	40
4.2	Funktionen der 17β -HSDs der Säugetiere.	41
4.3	Charakterisierte HSD-Proteine in komplett sequenzierten Ge- nomen	42
4.4	Kristallstrukturen von Mitgliedern der HSD-Familie	44
4.5	Distanzmatrix der kristallisierten HSD-Proteine 1	48
4.6	Distanzmatrix der kristallisierten HSD-Proteine 2	48
4.7	SAGE- <i>Tag</i> -Verteilung von humanen HSDs und Enzymen der Cholesterin-Biosynthese	61
4.8	Genomische Struktur und Lokalisation von 17β -HSD7 und der koregulierten Proteine, die das IDI1-Promotor-Modell enthal- ten (nach vorläufigen Daten des humanen Genomprojekts). . .	65
4.9	Beschreibung des gemeinsamen Promotor-Modells von IDI1, RPL39, MLN64 und OSR1	66
4.10	Genomische Sequenzen, die das IDI1-Promotor-Modell in der Nähe ihres vorhergesagten Transkriptionsstarts enthalten. Ei- nige Gene sind in mehreren verschiedenen Einträgen in der Datenbank vorhanden (HADHSC, TRX, CTSD, XB, EDMD), und einige Gene enthalten das Promotormodell zweimal (TPT1 und Guaninnucleotid-bindendes Regulatorprotein).	68
4.11	Enzyme der Cholesterin-Biosynthese in Hefe und Mensch . . .	72
4.12	Defekte der Cholesterinbiosynthese bei Maus und Mensch . . .	72
4.13	Verteilung und Funktion der AKR1C-Isoenzyme des Menschen	97

Tabellenverzeichnis

Anhang: Quellcode des STATRAS2-Programms

Quellcode

```
'STATRAS2.BAS
'(c) 2000 by Rainer Breitling
'Programm zur statistischen Auswertung von Proteinalignments
'Die Ergebnisse werden in Tabellenform ausgegeben
'und ausserdem in die Temperatur-Position (TempFactor) einer
'.PDB-Datei eingetragen.

DECLARE FUNCTION fak# (x AS INTEGER)
DECLARE FUNCTION faklog# (x AS INTEGER)
DECLARE FUNCTION zahl$ (b#) 'Wandelt eine Zahl in einen String der Laenge 6
                             'mit zwei Nachkommastellen und Dezimalpunkt um
DECLARE FUNCTION code$ (a%) 'Gibt den Ein-Buchstaben-Code einer Aminosaeu-
re an,
                             'nach alphabetischer Reihenfolge
DECLARE FUNCTION aanumber! (code AS ANY) 'Gibt die Nummer (in alphabetischer
                                         'Reihenfolge) zu einer Aminosaeure an

CONST defaultdatei = "input.gde" 'Alignment-Datei
CONST defaultpfad = "C:\Rainer"

TYPE PDBZeile                                     'Format einer "ATOM"-Zeile in .PDB-Datei
  ID AS STRING * 6                               'Identifizier
  serial AS STRING * 5
  sp1 AS STRING * 1 'sp sind nicht genutzte "Luecken"
  atom AS STRING * 4
  altloc AS STRING * 1
  resName AS STRING * 3
  sp2 AS STRING * 1
  chainID AS STRING * 1
  resSeq AS STRING * 4
  iCode AS STRING * 1
  sp3 AS STRING * 3
  x AS STRING * 8
  y AS STRING * 8
  z AS STRING * 8
  occupancy AS STRING * 6
  tempFactor AS STRING * 6
  sp4 AS STRING * 6
  segID AS STRING * 4
  element AS STRING * 2
  charge AS STRING * 2
  carret AS STRING * 1 'Abschliessender Carriage-Return
END TYPE

DIM dateilaenge AS INTEGER
DIM max.sequenzlaenge AS INTEGER 'Laenge der laengsten Sequenz im Alignment
DIM max.namenlaenge AS INTEGER 'Laenge des laengsten Sequenzna-
mens im Alignment
DIM sequenz.ident AS STRING 'ein Sequenzname
```



```

        max.namenlaenge = LEN(sequenz.ident)
    END IF
    LOOP UNTIL EOF(1) OR ASC(workstring) = ASC("%")
    DO WHILE NOT EOF(1)
        INPUT #1, workstring
        IF LEFT$(workstring, 1) = "%" THEN

            'ausgegeben werden die laufende Nummer, der Name und der Beginn der Se-
            quenz,
            'ausserdem ihre Laenge.
            'Die erste Sequenz wird gespeichert. Sie muss der Sequenz der .pdb-Datei
            'entsprechen, in der die statistischen Daten eingetragen werden.

            PRINT USING format$; datensatz; sequenz.ident; sequenz$; LEN(sequenz$)
            IF datensatz = 1 THEN template = sequenz$
            datensatz = datensatz + 1
            IF LEN(sequenz$) > max.sequenzlaenge THEN max.sequenzlaenge = LEN(sequenz$)
            IF LEN(sequenz.ident) > max.namenlaenge THEN max.namenlaenge = LEN(sequenz.ident)
            sequenz.ident = workstring
            sequenz$ = ""
        ELSE
            sequenz$ = sequenz$ + workstring
        END IF
    LOOP
    PRINT USING format$; datensatz; sequenz.ident; sequenz$; LEN(sequenz$)
    PRINT "Zahl der gelesenen Datensaeetze: "; datensatz
    PRINT "Die laengste Sequenz ist"; max.sequenzlaenge; "Aminosaeuren lang,";
    PRINT " der laengste Identifier hat"; max.namenlaenge; "Zeichen."
ELSE
    PRINT "Datei ist leer!"
END IF
CLOSE #1

'Nachdem die Alignmentdatei einmal gelesen wurde, um Laenge und Anzahl der
'Sequenzen zu bestimmen, werden die Felder definiert, in denen die
'statistischen Daten gespeichert werden:
'Aminosaeuren=Haeufigkeit der 20 AA und der Luecken im ganzen Alignment
'Aminomatrix=Haeufigkeit der 20 AA und der Luecken fuer jede Positi-
on im Alignment
'deltaG=Statistische Konservierung fuer jede Position

DIM identifier(1 TO datensatz) AS STRING
DIM aminosaeuren(1 TO 21) AS INTEGER
DIM aminomatrix(1 TO max.sequenzlaenge, 1 TO 21) AS INTEGER
DIM deltaG(1 TO max.sequenzlaenge) AS DOUBLE

'Initialisierung

FOR i = 1 TO max.sequenzlaenge
    deltaG(i) = 0
    FOR ii = 1 TO 21
        aminosaeuren(ii) = 0
        aminomatrix(i, ii) = 0
    NEXT
NEXT

FOR i = 1 TO datensatz
    identifier(i) = ""
NEXT

'Sequentielles Lesen der Sequenzen und Zaehlen der Aminosaeuren

OPEN dateiname FOR INPUT AS #1
DO
    INPUT #1, workstring
    IF LEFT$(workstring, 1) = "%" THEN

```

Anhang: Quellcode des STATRAS2-Programms

```
        sequenz.ident = workstring
    END IF
    LOOP UNTIL EOF(1) OR LEFT$(workstring, 1) = "%"
    FOR i = 1 TO datensatz
    DO
        INPUT #1, workstring
        IF LEFT$(workstring, 1) = "%" OR EOF(1) THEN
            FOR b% = 1 TO max.sequenzlaenge
                c$ = MID$(sequenz$, b%, 1)
                aminosaeuren(aanumber(c$)) = aminosaeuren(aanumber(c$)) + 1
                aminomatrix(b%, aanumber(c$)) = aminomatrix(b%, aanumber(c$)) + 1
            NEXT
            identifier(i) = sequenz.ident
            sequenz.ident = workstring
            sequenz$ = ""
        ELSE
            sequenz$ = sequenz$ + workstring
        END IF
    LOOP UNTIL LEFT$(workstring, 1) = "%" OR EOF(1)
NEXT
CLOSE #1
```

'Schreibt die globale Aminosaeurehaeufigkeit in Datei mit Trennzeichen &

```
OPEN "output1.txt" FOR OUTPUT AS #1
FOR ii% = 1 TO 21
    PRINT code$(ii%); ":"; aminosaeuren(ii%)
    WRITE #1, "&" + code$(ii%) + "&" + STR$(aminosaeuren(ii%)) + "&"
NEXT
CLOSE #1
```

'Schreibt die lokale Aminosaeurehaeufigkeit in Datei

```
OPEN "output2.txt" FOR OUTPUT AS #1
workstring = "&"
FOR ii% = 1 TO 21
    workstring = workstring + code$(ii%) + "&"
NEXT
WRITE #1, workstring
FOR ii% = 1 TO max.sequenzlaenge
    workstring = "&"
    FOR iii% = 1 TO 21
        workstring = workstring + STR$(aminomatrix(ii%, iii%)) + "&"
    NEXT
    WRITE #1, workstring
NEXT
CLOSE #1
```

'deltaG-Statistik, berechnet die Konservierung der einzelnen Positionen
'nach der Formel in Science.

```
aasumme = 0
FOR i% = 1 TO 20
    aasumme = aasumme + aminosaeuren(i%)
NEXT

FOR i% = 1 TO max.sequenzlaenge
    deltaG(i%) = 0
    zahlersaeuren% = 0
    FOR ii% = 1 TO 20
        zahlersaeuren% = zahlersaeuren% + aminomatrix(i%, ii%)
    NEXT
    IF zahlersaeuren% = 0 THEN
```

```

    zahlersaeuren% = 1
    PRINT "Fehler im Alignment? (Gap-only positions entfernen!)"
    END IF

'Stirling-Naeherung fuer Fakultaelen N>150 benutzen.

    FOR ii% = 1 TO 20
        IF zahlersaeuren% < 150 THEN
            k# = (fak#(zahlersaeuren%) / fak#(zahlersaeuren% - aminoma-
trix(i%, ii%)))
            k# = k# / fak#(aminomatrix(i%, ii%))
        ELSEIF aminomatrix(i%, ii%) < 150 THEN
            k# = EXP(faklog#(zahlersaeuren%) - faklog#(zahlersaeuren% - aminoma-
trix(i%, ii%)))
            k# = k# / fak(aminomatrix(i%, ii%))
        ELSE
            k# = EXP(faklog#(zahlersaeuren%) - (faklog#(zahlersaeuren% - aminoma-
trix(i%, ii%))
                    + faklog#(aminomatrix(i%, ii%))))
        END IF
        b# = aminosaeuren(ii%) / aassumme 'Eigentlich: Wahrscheinlichkeit in Daten-
bank,
                                'nicht im Alignment!!
        a# = k# * b# ^ aminomatrix(i%, ii%) * (1 - b#) ^ (zahlersaeuren% - amino-
matrix(i%, ii%))
        zc# = a# / b#
        deltaG(i%) = deltaG(i%) + LOG(zc#) * LOG(zc#)
    NEXT
    deltaG(i%) = SQR(deltaG(i%))
NEXT

'Statistische Konservierung in Datei schreiben, parallel zur ersten Sequenz
'des Alignments (Template)

OPEN "output3.txt" FOR OUTPUT AS #1
    FOR i% = 1 TO max.sequenzlaenge
        WRITE #1, "&" + MID$(template, i%, 1) + "&" + STR$(deltaG(i%)) + "&"
    NEXT
CLOSE #1

'PDB-File modifizieren, d. h. TempFactor der Atome auf die deltaG-Werte setzen
'und Heteroatome auf null. Es wird nicht geprueft, ob die Sequenz auch stimmt,
'sondern nur der Reihe nach abgearbeitet.

OPEN "c:\rainer\lFRB.pdb" FOR BINARY AS #1
seqpos = 0
dbpos = 1

maximum# = 0
FOR i% = 1 TO max.sequenzlaenge
    IF maximum# < deltaG(i%) THEN maximum# = deltaG(i%)
NEXT

DO WHILE NOT EOF(1)
    DO
        GET #1, dbpos, zeile
        IF zeile.ID = "HETATM" THEN
            zeile.tempFactor = " 0.00"
            PUT #1, dbpos, zeile
        END IF
        dbpos = dbpos + 81
    LOOP UNTIL (zeile.ID = "ATOM ") OR EOF(1)
    IF zeile.ID = "ATOM " THEN
        IF zeile.resSeq <> altezeile.resSeq THEN

```

Anhang: Quellcode des STATRAS2-Programms

```
        IF zeile.chainID <> altezeile.chainID THEN 'von vorne anfan-
gen, bei jeder
                                                    'neuen Kette
            seqpos = 0
        END IF
        DO
            seqpos = seqpos + 1
            LOOP UNTIL MID$(template, seqpos, 1) <> "-" OR seq-
pos > max.sequenzlaenge
            altezeile = zeile
        END IF
        umkehrwert# = -(deltaG(seqpos) - maximum#)
        zeile.tempFactor = zahl(umkehrwert#)
        PUT #1, dbpos - 81, zeile 'Zeilen sind einschliesslich Return 81 By-
tes lang.
        END IF
    LOOP
    CLOSE #1

abschied:
PRINT "Vielen Dank fuer die gute Zusammenarbeit."
END

fehlerbehandlung:
SELECT CASE ERR
    CASE 53:
        PRINT "Datei nicht gefunden!"
        dateiexist = 0
        RESUME NEXT
    CASE 76:
        PRINT "Pfad nicht gefunden!"
        pfadexist = 0
        RESUME NEXT
    CASE ELSE
        PRINT "Fehlercode: "; ERR; ". Programm beendet."
        END
END SELECT

'Gibt zu einem Zeichen die AA-Nummer zurueck, die als Index in den
'Arrays verwendet wird.
,
FUNCTION aanumber (code$)
    SELECT CASE LEFT$(UCASE$(code$), 1)
        CASE "A": aanumber = 1
        CASE "C": aanumber = 2
        CASE "D": aanumber = 3
        CASE "E": aanumber = 4
        CASE "F": aanumber = 5
        CASE "G": aanumber = 6
        CASE "H": aanumber = 7
        CASE "I": aanumber = 8
        CASE "K": aanumber = 9
        CASE "L": aanumber = 10
        CASE "M": aanumber = 11
        CASE "N": aanumber = 12
        CASE "P": aanumber = 13
        CASE "Q": aanumber = 14
        CASE "R": aanumber = 15
        CASE "S": aanumber = 16
        CASE "T": aanumber = 17
        CASE "V": aanumber = 18
        CASE "W": aanumber = 19
        CASE "Y": aanumber = 20
        CASE ELSE: aanumber = 21
    END SELECT
END FUNCTION
```

'Gibt zu einer Index-Nummer die passende Aminosaeure (One-letter code) zurueck.

```
,  
,  
FUNCTION code$ (a%)  
  SELECT CASE a%  
    CASE 1: code$ = "A"  
    CASE 2: code$ = "C"  
    CASE 3: code$ = "D"  
    CASE 4: code$ = "E"  
    CASE 5: code$ = "F"  
    CASE 6: code$ = "G"  
    CASE 7: code$ = "H"  
    CASE 8: code$ = "I"  
    CASE 9: code$ = "K"  
    CASE 10: code$ = "L"  
    CASE 11: code$ = "M"  
    CASE 12: code$ = "N"  
    CASE 13: code$ = "P"  
    CASE 14: code$ = "Q"  
    CASE 15: code$ = "R"  
    CASE 16: code$ = "S"  
    CASE 17: code$ = "T"  
    CASE 18: code$ = "V"  
    CASE 19: code$ = "W"  
    CASE 20: code$ = "Y"  
    CASE ELSE: code$ = "-"  
  END SELECT  
END FUNCTION
```

'Berechnet fuer Ganzzahlen kleiner 170 die Fakulaet. Als Fehlerwert wird 0 zurueckgegeben.

```
,  
,  
FUNCTION fak# (x AS INTEGER)  
  IF x < 170 AND x >= 0 THEN  
    y# = 1  
    FOR i% = 1 TO x  
      y# = y# * i%  
    NEXT  
    fak# = y#  
  ELSE  
    fak# = 0  
  END IF  
END FUNCTION
```

'Berechnet den Logarithmus der Fakultaet einer Ganzzahl fuer die Verwendung in der Stirling-Naeherung

```
,  
,  
FUNCTION faklog# (x AS INTEGER)  
  faklog# = x * LOG(x) - x + LOG(SQR(2 * 3.1415927# * x)) - LOG(2)  
END FUNCTION
```

'Macht aus einer Double-Zahl einen String der Laenge 6 mit zwei Nachkommastellen und Dezimalpunkt.

```
,  
,  
FUNCTION zahl$ (b#)  
  vorkomma% = FIX(b#)  
  nachkomma% = FIX(100 * (b# - FIX(b#)))  
  IF nachkomma% >= 10 THEN  
    zahl$ = RIGHT$(" " + STR$(vorkomma%) + "." + RIGHT$(STR$(nachkomma%), 2), 6)  
  ELSE  
    zahl$ = RIGHT$(" " + STR$(vorkomma%) + ".0" + RIGHT$(STR$(nachkomma%), 1), 6)  
  END IF  
END FUNCTION
```

Anhang: Quellcode des STATRAS2-Programms

Beispiel einer Eingabedatei im GDE-Format:

```
%Sequenz1  
MAHLIPLAQRLANMPALA  
%Sequenz2  
MVRLLPLAQKVIN--KAV  
%Sequenz3  
MIKVLQLVWRAAQMPVLI
```

Beispielzeile der modifizierten PDB-Datei:

```
ATOM      18  CA  THR  A   3           18.986  10.932  61.250  1.00  14.14           C  
                                         ^^^^^  
                                         deltaG-Wert
```

Danksagung

Mein ganz besonderer Dank geht an den Betreuer meiner Doktorarbeit, PD Dr. Jerzy Adamski, dessen Engagement, Großzügigkeit und verständnisvolle Unterstützung zum erfolgreichen Abschluss der Arbeit entscheidend beigetragen haben.

Prof. Dr. Rudi Balling (Institut für Säugetiergenetik) und Dr. Martin Hrabé-de Angelis (Institut für Experimentelle Genetik) danke ich für die Möglichkeit an ihren Instituten zu arbeiten.

Dr. Josef-Karl Gerber hat meine Arbeit jederzeit mit kritischer Aufmerksamkeit begleitet und unterstützt. Die Pilotstudie über die „Evolution der Paired Box“ geht auf seine Anregung zurück und wäre ohne seine Hilfe nicht möglich gewesen.

Danken möchte ich auch allen anderen Kollegen in der Arbeitsgruppe „Molekulare Endokrinologie“, für die freundschaftliche Atmosphäre und ausgeprägte Diskussionsbereitschaft. Ganz besonders verpflichtet bin ich aber den Kollegen, die mir ihre noch unveröffentlichten Daten zur Analyse zugänglich machten:

- Christian-Johannes Gloeckner, der die Struktur des murinen Hsd17b1-Gens bestimmte,
- Kim Masuda (San Diego), Brigitte Keller und Jerzy Adamski, die die humanen IDI1- und IDI2-Gene charakterisierten,
- Antje Krazeisen, die die Inhibition der humanen 17 β -HSD5 untersuchte,
- Zrinka Marijanović, die die Expression der 17 β -HSD7 in Zellkulturexperimenten analysierte,
- und schließlich ganz besonders Daniela Laubner, die bei ihren mRNA-*in situ*-Hybridisierungen an Mäuseembryonen die Rolle der 17 β -HSD7 in der frühen Embryonalentwicklung entdeckte.

Ohne ihre Unterstützung wäre diese Arbeit nicht möglich gewesen.

Ein letztes, ganz entscheidendes Dankeschön geht schließlich an meine Familie, die meine „Forschungen“ nicht erst während meiner Münchener Zeit ausdauernd interessiert, engagiert und hilfsbereit unterstützt hat.

Danksagung

Lebenslauf

Name: Rainer Breitling
Geburtsdatum: 25. Juni 1971
Geburtsort: Laatzen

Januar 1998 bis heute Wissenschaftlicher Mitarbeiter am Institut für experimentelle Genetik, GSF Forschungszentrum für Umwelt und Gesundheit, Neuherberg.

Dezember 1998 Forschungsaufenthalt am Institut für Onkologie, University Medical School, Palermo, Italien.

Juli bis Dezember 1997 Diplomarbeit am Institut für Säugetiergenetik, GSF Forschungszentrum für Umwelt und Gesundheit, Neuherberg. Thema: „Expression und Reinigung von *sonic hedgehog*“.

Oktober bis Dezember 1996 *Research assistant* im Department of Plant Biology, University of California, Berkeley, USA.

Juli 1992 bis Juni 1997 Studium der Biochemie an der Universität Hannover.

Januar 1991 bis März 1992 Zivildienst in den Städtischen Kliniken, Delmenhorst.

Juni 1990 Abitur an der St. Ursula-Schule, Hannover.