

# 1. Einleitung

Die ungeheure Vielfalt des Lebens, wie wir sie heute vorfinden, ist Ausdruck der Anpassung an verschiedenste ökologische Nischen. Jedoch findet diese Vielfalt bei erster Betrachtung keine Entsprechung auf molekularer Ebene. Allen lebenden Zellen ist gemein, dass die genetische Information auf der Ebene von Nukleinsäuren organisiert ist und durch den Prozess der Transkription und der Translation ausgedrückt wird. Die Endprodukte dieser zwei Prozesse, die Proteine, sind der Hauptbestandteil der Zellen. Ihre räumlich-strukturelle Ausdehnung und ihre chemischen Eigenschaften befähigen sie zu unterschiedlichsten Aufgaben wie z.B. Reaktionskatalysen, Speicherung und Transport von Stoffen, zellstrukturenbenden Funktionen oder der Regulation ihrer eigenen Synthese. Die Abfolge der Grundbausteine der Proteine, der Aminosäuren, wird durch die Sequenzabfolge der Basen in der DNA festgelegt. Seitdem in den 90er Jahren des letzten Jahrhunderts große Anstrengungen unternommen wurden, Hochdurchsatz DNA-Sequenzierautomaten zu entwickeln (Übersicht bei Meldrum, 2000), ist die Zahl der auf Basenabfolge entschlüsselten Genome enorm angestiegen (Sterky und Lundeberg, 2000). Dies gipfelte in der Sequenzierung des gesamten menschlichen Genoms (Venter *et al.*, 2001; International Human Genome Consortium, 2001). Diese Fülle von neu hinzugewonnenen Informationen erfordert die Entwicklung von neuen bioinformatischen Methoden, um die Daten einzuordnen (eine Übersicht findet sich bei Dandekar und Sauerborn, 2002). Insbesondere in der pharmakologischen Forschung werden große Hoffnungen auf neue Erkenntnisse, die aus den Genomsequenzen abgeleitet werden können, gesetzt (Ashton *et al.*, 2002). So lassen sich durch Sequenzvergleiche krankheitsrelevante Gene identifizieren und genetische Ursachen oder Prädispositionen für Krankheiten festmachen. Allerdings ist zum tieferen Verständnis zellulärer Prozesse das alleinige Wissen über die Aminosäureabfolge der beteiligten Proteine nicht hinreichend. Sämtliche Prozesse des Lebens unterliegen physikalischen Gesetzen und werden durch die Interaktion von Molekülen gesteuert. Diese Interaktionen finden an den Oberflächen dieser Biomoleküle statt und setzen in der Regel eine geordnete Raumstruktur der interagierenden Komponenten voraus. Ziel der molekularbiologischen Forschung ist es, diese Wechselwirkungen im atomaren Detail zu verstehen und zu modulieren. Diese Arbeit will einen methodischen Beitrag dazu leisten, Wechselwirkungen zwischen Peptiden und Proteinen durch Anwendung eines evolutiven Ansatzes zu studieren und auf Fragestellungen der Immunbiologie, Biotechnologie und molekularer Medizin anzuwenden.

Die nachfolgenden Abschnitte sollen einen einführenden Überblick zu immunologischen Fragestellungen von Protein-Protein Interaktionen geben (Abschnitt 1.1), Konzepte, Möglichkeiten und Grenzen der Modulation biomolekularer Wechselwirkungen aufzeigen (Abschnitt 1.2) und die gegenwärtig technischen Möglichkeiten ausloten, um Zugang zu Proteinen zu erhalten, die vorgegebenen Ansprüchen an die Bindung eines Zielproteins genügen (Abschnitt 1.3).

## 1.1 Antikörper-Antigen Wechselwirkungen

Ein typisches Beispiel für spezifische Interaktionen zwischen Proteinen ist die Bindung von Antikörpern an Antigene. Das Immunsystem der Vertebraten hat im wesentlichen zwei zentrale Aufgaben. Es dient zur spezifischen Abwehr von Fremdstoffen wie z.B. Viren oder Bakterien. Weiterhin dient es zur Eliminierung von veränderten körpereigenen Strukturen wie infizierten Zellen (s.o.) oder Tumorzellen (Kleining und Sitte, 1992). Für diese Aufgabe ist es speziell ausgerüstet mit den T-Lymphozyten und den B-Lymphozyten sowie den membrangebundenen und löslichen Produkten der B-Lymphozyten, den Antikörpern. Fremdmoleküle sind Antigene, die vom Immunsystem als solche erkannt werden. Dazu dienen hauptsächlich die T-Zellrezeptoren auf den T-Lymphozyten und die membranständigen oder löslichen Antikörper der B-Lymphozyten. Nahezu jede in den Organismus eindringende Fremdzelle bzw. jedes Makromolekül wird vom Immunsystem als solches erkannt und wenn möglich ausgeschaltet. Die Antikörper, die daran einen entscheidenden Anteil haben, sind aus zwei leichten und zwei schweren durch Disulfidbrücken verknüpften Ketten aufgebaut. Jede der Ketten besitzt eine variable Domäne mit drei hypervariablen Schleifen. Diese hypervariablen Bereiche sind für die spezifische Erkennung der Fremdmoleküle verantwortlich (Roitt, 1991). Dabei wird von den Immunglobulinen nicht das gesamte Makromolekül als Fremdstoff erkannt, sondern nur eine kleine Teilstruktur, das Epitop. Dieses Epitop wird von der Bindungsregion des Immunglobulins, dem Paratop, gebunden. Jedes Paratop bindet hoch spezifisch und selektiv nur eine antigene Determinante eines Makromoleküls. Da ein Makromolekül mehrere solcher antigenen Determinanten (Epitope) auf seiner Oberfläche besitzt, werden als Reaktion auf das Eindringen dieses Makromoleküls viele B-Zellen und ebenso viele Immunglobuline induziert.

Die Kenntnisse über solche antigenen Determinanten können von großer Bedeutung sein. Aus ihnen lassen sich vielfältigste Informationen ableiten. So können sie dafür benutzt werden, neue effektivere Impfstoffe gegen pathogene Viren, Bakterien oder Eukaryonten zu entwickeln (Langedijk *et al.*, 2001; van Oirschot, 1999). Sind die antigenen Determinanten nicht bekannt, so muss das gesamte Antigen zur Immunisierung verwendet werden (z.B. abgeschwächte Lebendimpfstoffe). Ebenso kann bei Kenntnis der Epitope eine Infektion effektiv durch immunochemische Verfahren nachgewiesen werden, möglicherweise sogar zwischen unterschiedlichen Stämmen eines Pathogens unterschieden werden (sofern unterschiedliche Epitope bekannt sind). Dies wiederum könnte zu einer gezielteren und dadurch effizienteren Therapie führen. Weiterhin ließen sich durch die Bestimmung von Autoantigenen neue Therapien bei Autoimmunkrankheiten entwickeln (Mennuni *et al.*, 1997).

Um solche antigenen Determinanten oder Epitope zu kartieren, werden in der Regel Peptidbibliotheken benutzt. Dazu wurden verschiedene Verfahren etabliert. Besonders häufig wird dabei auf chemische Synthesen zurückgegriffen. Das zu untersuchende Antigen wird dabei in geordneten, sich überlappenden, Peptiden auf einer festen Matrix synthetisiert. Anschließend kann nach Inkubation der Peptide mit den entsprechenden Antikörpern ein immunochemischer Nachweis des oder die Peptide nachweisen, die von den Antikörpern gebunden werden (eine Übersicht findet

sich bei Geysen *et al.*, 1984 und bei Reineke *et al.*, 1999). Oft werden auch evolutive Anreicherungsmethoden angewendet. Am häufigsten wird dabei das *Phage Display* (vergl. 1.3.1) Verfahren benutzt. Dabei wird die für das Antigen kodierende DNA fragmentiert und diese Fragmente werden dann in den Sequenzkontext eines Phagenhüllproteins gesetzt. Anschließend werden die *Phagen*, die ein Epitop präsentieren, von den entsprechenden Antikörpern gebunden und können angereichert werden. Die Sequenz des Epitops wird dann indirekt durch Sequenzierung der *Phagen*-DNA bestimmt. Als Beispiele hierfür seien die Bestimmung eines autoimmun-Epitopes, das bei der Typ1 Diabetes relevant ist (Farilla *et al.*, 2002, vergl. auch Abschnitt 1.3.1), sowie die Kartierung von Hepatitis C-Virus Epitopen (Pereboeva *et al.*, 1998 und 2000) genannt. Weiterhin konnten einzelne Epitope erfolgreich auf Bakterienzellen präsentiert und durch die entsprechenden Antikörper nachgewiesen werden (vergl. Abschnitt 1.3.2).

## *1.2 Herstellung und Isolierung von Polypeptiden mit vorgegebenen Interaktionseigenschaften*

Im Lauf der Evolution hat die Natur viele funktionell verschiedene Proteine geschaffen. Durch iterative Runden von Mutation und Selektion sind hochspezielle und präzise Werkzeuge der Zellen entstanden, die sich untereinander durch unterschiedlichste Interaktionen beeinflussen. Wenn dieses natürliche Wechselspiel zwischen den Makromolekülen aus irgendwelchen Anlässen gestört ist, hat dies für das betroffene Individuum oft fatale Folgen. So führt die Typ1 Diabetes, eine Autoimmunerkrankung, durch Fehlsteuerung des Immunsystems zur Zerstörung der  $\beta$ -Zellen der Langerhansschen Inseln im endokrinen Pankreas. In diesen Inseln wird das Insulin, ein Peptidhormon, das an den membranständigen Insulinrezeptor bindet, gebildet. Aufgrund dieser Wechselwirkung wird eine Signalkaskade induziert, die zur vermehrten Protein-, Glycogen- und Speicherfettsynthese führt. Durch diese vermehrte Synthese wird der Blutzuckerspiegel gesenkt und reguliert. Dass diese Regulation tatsächlich durch die Interaktion von Insulin mit der Zellmembran zusammenhängt, konnte bereits 1969 nachgewiesen werden (Cuatrecasas, 1969). Sofern aufgrund der Zerstörung der  $\beta$ -Zellen der Langerhansschen Inseln kein Insulin gebildet wird, erfolgt keine Regulation des Blutzuckerspiegels und es kommt zur Erkrankung. In diesem Fall ist es notwendig, die Interaktion durch Gabe von Insulin gezielt herbeizuführen. Die Regulation wird somit von außen durch Einbringen des natürlichen Interaktionspartners gesteuert. Es gibt allerdings auch wünschenswerte regulatorische Interaktionen zwischen Molekülen, für die es keine natürlichen Bindungspartner in dem jeweiligen Organismus gibt. So kann man beispielsweise die Signalweiterleitung in den Nervenbahnen durch Zugabe von Stoffen unterbrechen, die Ionenkanäle blockieren (Übersicht bei Miljanich und Ramachandran, 1995). Stoffe, die diese Funktion erfüllen, können aus der Natur isoliert werden. So finden sich in den Venomen von Meeresschnecken Peptide, die solche schmerzrelevanten Kanäle blockieren (Übersicht bei Olivera *et al.*, 1994). Nicht für jedes medizinisch, molekularbiologisch oder biotechnologisch interessante Zielprotein wird man mit

vertretbarem Aufwand einen natürlichen Modulator finden. Aus diesem Grund wird häufig versucht, die gewünschte Funktion, in den oben beschriebenen Beispielen die Fähigkeit zur Interaktion eines Moleküls mit einem vorgegebenen Protein, *de novo* zu erzeugen. Dabei kommen, sofern man proteinbasierte Interaktionspartner sucht, zwei vom konzeptionellen Ansatz her unterschiedliche Verfahren zum Einsatz: Erstens das rationale *protein design* und zweitens das evolutive *Protein Engineering*. Bei dem erstgenannten wird ausgehend von einer dreidimensionalen Kristallstruktur des Zielproteins und des potentiellen Interaktionspartners mittels computerunterstütztem molekularem Modellieren versucht, durch gezielte Austausche in der Aminosäuresequenz des Interaktionspartners, die gewünschte Funktion auf die Ausgangsstruktur zu übertragen. Anschließend werden die Vorhersagen experimentell überprüft. Durch die enormen Fortschritte in der Computertechnik, insbesondere durch schnellere Rechenleistungen, Parallelisierung von Rechenaufgaben und den Einsatz neuer Algorithmen zur Simulation von Moleküldynamik und -faltung, werden die Vorhersagen aller Voraussicht nach immer genauer (eine Übersicht geben Street und Mayo, 1999). Nach wie vor wird die gewünschte Funktionalität allerdings in der Regel nicht bei der ersten Vorhersage erreicht, sondern es müssen vielmehr iterative Näherungsschritte unternommen werden. Somit wird in aufeinanderfolgenden Durchläufen durch einen Zyklus aus Vorhersage und experimenteller Überprüfung die gewünschte Funktion des Ausgangsproteins optimiert.

Ein Problem des rationalen *Protein Engineerings* ist die Voraussetzung der genauen Kenntnis der Strukturen beider Interaktionspartner. Sind diese Informationen nicht vorhanden, so kann keine Optimierung vorgenommen werden. Es gibt aber häufig für Moleküle, für die ein Interaktionspartner gesucht wird, keine verlässlichen Strukturdaten. Weiterhin sind die Vorhersagen über Konsequenzen von Aminosäureaustauschen auf die Tertiärstruktur nur begrenzt möglich, da der Faltungsmechanismus der Proteine nicht hinreichend verstanden ist. Im Extremfall kann der Austausch einer einzigen Aminosäure zum Kollabieren der gesamten Struktur führen.

Einen konzeptionell anderen Ansatz verfolgt die zweitgenannte Methode, das evolutive *Protein Engineering*. Hierbei wird versucht, ein breites Repertoire an möglichen Interaktionsmolekülen herzustellen, die dann anschließend einer Durchmusterung unterzogen werden. Bei dieser Durchmusterung sollen dann solche Varianten isoliert werden, die die gewünschten Eigenschaften aufweisen. In der Regel werden bei solchen Verfahren die Varianten in Organismen, die als Synthesemaschinen dienen, hergestellt. Die Variation der einzelnen Moleküle wird meistens zuvor auf DNA-Ebene mittels PCR-Techniken erreicht. Sofern die gewünschte Eigenschaft die Bindung an ein vorgegebenes Zielmolekül ist (z.B. ein Ionenkanal, s.o.) ist es häufig ausreichend, die Veränderungen der Gensequenz auf bestimmte Regionen zu beschränken. Oft wird nur eine einzige Schleife, die für die Interaktion mit dem vorgegebenen Protein gedacht ist, verändert. Es bietet sich daher an, diese Veränderungen mit randomisierten Oligonukleotiden einzuführen. Nach der Klonierung der varianten Gene und deren Expression steht eine Sammlung von Molekülen mit unterschiedlichsten Eigenschaften zur Verfügung. Die Suche nach demjenigen Molekül, das die gewünschte Eigenschaft - z.B. die Blockierung eines

Ionenkanals - besitzt, kann auf unterschiedliche Weise geschehen. Der Vorteil eines solchen Verfahrens ist, dass die erzeugte Molekülbibliothek auf unterschiedlichste Funktionen hin getestet werden kann. Ist die Bibliothek - aufgrund ihres strukturellen Grundgerüsts (s.u.) - besonders geeignet, Bindemoleküle zu enthalten, die mit Ionenkanälen interagieren, so kann man aus ihr spezifische Bindemoleküle gegen viele unterschiedliche Kanäle isolieren. Die technische Schwierigkeit dieses evolutiven Ansatzes liegt darin, aus einer sehr großen Sammlung von Kandidatenmolekülen diejenigen wenigen in die Hand zu bekommen, die den vorgegebenen Ansprüchen an die Bindung eines Zielproteins genügen.

### 1.2.1 Anforderungen an ein strukturelles Grundgerüst für *Protein Engineering*

In den letzten Jahren wurde eine Reihe von Proteingrundgerüsten (*scaffold's*) wie Antikörperfragmente, Helixbündelproteine oder Lipocaline eingesetzt, von denen durch Variation oberflächenexponierter Reste Molekülbibliotheken hergestellt werden können (Übersicht bei Skerra, 2000). Soll einem solchen Grundgerüst eine neue Funktion zugewiesen werden, ist es immer von Vorteil, die Struktur und die natürliche Funktion zu kennen, um Anhaltspunkte über mögliche zu variierende Bereiche zu erlangen. Weiterhin sind an solche rahmengebende Proteine gewisse Ansprüche bezüglich der Faltungsstabilität zu stellen. Vorzugsweise sollte sich die lokal beschränkte Veränderung der Aminosäuresequenz nicht auf die Faltung des Moleküls und seine Stabilität auswirken. Soweit möglich, ist es von Nutzen, solche Proteine als Grundgerüst zu wählen, deren natürliche Funktion der gesuchten ähnlich ist. Dadurch steigt die Wahrscheinlichkeit, mit wenigen Aminosäureaustauschen die gewünschte Eigenschaft einzuführen. Ein ideales Grundgerüst sollte aus Gründen der experimentellen Überschaubarkeit möglichst klein und als Modulator biomolekularer Wechselwirkungen gut charakterisiert sein.

### 1.2.2 Die Familie der Cystinknoten Proteine

Eine interessante Klasse von Proteinen, die die oben genannten Anforderungen erfüllen, ist die Familie der Cystinknotenproteine. Diese sehr kleinen Proteine von nur wenigen Dutzend Aminosäuren Länge kommen in der Natur häufig vor und interagieren mit einer Vielzahl verschiedener Zielproteine. So werden einige Wachstumsfaktoren, wie NGF, TGF $\eta$  und PDGF zu dieser Klasse gezählt. Diese Peptidhormone erkennen Rezeptoren auf der Zelloberfläche und lösen durch Andocken an diese, wie schon das oben beschriebene Insulin, eine Signalweiterleitung aus (eine Übersicht findet sich bei McDonald und Hendrickson, 1993). Weiterhin gehören zu dieser Klasse die oben erwähnten Conotoxine und Proteinaseinhibitoren. Allen Vertretern dieser Klasse ist die Verknüpfung der Hauptkette durch mindestens drei Disulfidbrücken gemeinsam. Bei den Vertretern der Proteaseinhibitoren und der Neurotoxine ist die Verknüpfung von sechs Cysteinresten charakteristisch. Es wird ein Ring aus zwei Cysteinbrücken zwischen den Resten C#1-C#4 und C#2-C#5 gebildet (die Nummerierung der Reste gibt die Reihenfolge in der Hauptkette an). Anschließend wird dieser Ring von einer dritten Disulfidbrücke (C#3-C#6) durchzogen. Bei den zuvor oben genannten Wachstumsfaktoren wird der Ring durch