

## KAPITEL 1 Einleitung

Die Beziehung zwischen Sequenzinformation und Funktion (SAR) ist entscheidend für das gezielte *de novo* Design von Proteinen. Dabei kann die Struktur der Proteine ein hilfreiches Zwischenergebnis darstellen (Abbildung 1).

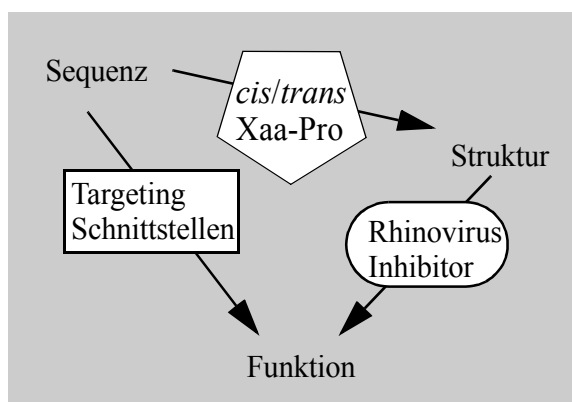


ABBILDUNG 1: Der Zusammenhang zwischen Sequenz und Struktur/Funktion wird an den drei angegebenen Beispielen in dieser Arbeit behandelt.

Die Kodierung der Aminosäuren in Zahlen, die von mathematischen Verfahren verarbeitet werden, ist für die SAR ein zentrales Problem. Auf der einen Seite soll sie die charakteristische Information enthalten, auf der anderen Seite soll für statistische Verfahren die Anzahl der Parameter möglichst gering gehalten werden. Die Kodierung, die die vollständige Sequenzinformation beinhaltet, ist die verteilte Kodierung, die jeder Aminosäureposition in der Sequenz 20 Zahlen zuordnet, entsprechend den 20 natürlich vorkommenden Aminosäuren. Diese Kodierung hat jedoch den Nachteil, daß in der Regel mehr Parameter verwendet werden als nötig sind, um das Problem korrekt zu beschreiben. Andere Kodierungen, die z.B. auf physikochemische Eigenschaften zurückzuführen sind, reduzieren diesen Datenraum, indem Relationen zwischen den einzelnen Aminosäuren vorgegeben werden. Es ist aber nicht immer klar, welche Beziehungen für das gegebene Problem wichtig sind. In der vorliegenden Arbeit wird ein Algorithmus vorgestellt, der eine Kodierung mit sehr wenigen Parametern verwendet, ohne sich auf eine bestimmte Relation zu beschränken. Es wird lediglich vorgegeben, wieviele Rela-

tionen verwendet werden sollen. So wird eine problemspezifische Kodierung entwickelt, die anschließend mit physikochemischen Eigenschaften verglichen werden kann.

Dieser Algorithmus wird zur Vorhersage von Schnittstellensequenzen humaner sekretorischer Proteine verwendet. Er ist über das World Wide Web (WWW) allgemein zugänglich.

Inwieweit sich diese Methode auch zur Vorhersage struktureller Eigenschaften eignet, wird am Beispiel der Konformationsvorhersage von Peptidylprolylbindungen untersucht. Einige dieser Bindungen werden von bestimmten Enzymen isomerisiert. Da Enzyme lokale Sequenzen erkennen, sollte es möglich sein, diese Sequenzen mit Hilfe von Klassifikationsverfahren, wie dem hier vorgestellten, zu analysieren.

Es wurde eine weitere Methode entwickelt, die, basierend auf bekannten Protein-Protein-Komplexen, die charakteristischen Eigenschaften von Aminosäurewechselwirkungen berücksichtigt. Hierbei werden aus bekannten dreidimensionalen Strukturen Eigenschaften abgeleitet, mit denen die gezielte Vorhersage von Bindungspartnern möglich ist. Daneben wird gezeigt, wie auch ohne Ausnutzung von strukturellen Kenntnissen zellprotektive Peptide entwickelt werden können. Dazu haben sich bereits mehrfach Künstliche Neuronale Netze (KNN) in Verbindung mit der Simulierten Molekularen Evolution (SME) bewährt [111].

Grundlage aller Untersuchungen sind problemspezifische Datenbanken, die repräsentativ für das gegebene Problem sein sollen. Diese werden mit statistischen und bioinformatischen Methoden analysiert, um die für das Problem wesentlichen Eigenschaften zu charakterisieren. Mit dem Design und Testen neuer Proteine können die gewonnenen Hypothesen überprüft werden.

## 1.a Schnittstellen humaner Signalpeptide

Viele Proteine, die *in vivo* im Cytosol synthetisiert werden, müssen die Membran des Endoplasmatischen Retikulums (ER) passieren. Eine Signalsequenz am N-Terminus des Proteins bestimmt, ob ein solcher Transport stattfindet. Es gibt zwei mögliche Wege, die ER-Membran zu passieren, den cotranslationalen und den posttranslationalen Weg. Beide Wege führen das Protein zum Translokationsapparat, dem Translokon, der einen Transport durch die Membran ermöglicht. Das Signalpeptid wird nach dem Transport für die meisten Proteine nicht mehr benötigt und auf der luminalen Seite durch einen spezifischen Komplex, den Signalpeptidase

---

Komplex [85] abgespalten. Mutationen in der unmittelbaren Umgebung der Schnittstellenregion können zu schweren Defekten führen [3, 94, 123].



ABBILDUNG 2: Schematische Darstellung eines Proteins mit einem Signalpeptid. Am N-Terminus befindet sich die N-Region (N) mit einer negativen Ladung. Danach folgt die H-Region (H) mit hydrophoben Aminosäuren und der C-Region (C), der Schnittstellenregion. Danach folgt das mature Protein.

Signalpeptide besitzen spezifische Eigenschaften, die zur Entwicklung neuer Signalpeptide verwendet werden können [40]. Sie bestehen aus drei Bereichen: einer negativ geladenen Region am N-Terminus (N-Region), einer stark hydrophoben Region (H-Region) und der Schnittstellenregion (C-Region) (Abbildung 2). Trotz dieser funktionellen Ähnlichkeiten besitzen Signalsequenzen eine sehr geringe Homologie von weniger als 20% [34]. In den letzten Jahren wurde versucht, die Targetingsignale von Proteinen vor allem mit Künstlichen Neuronalen Netzen vorherzusagen. Diese Vorgehensweise hat sich in Zusammenhang mit wachsenden Datenmengen als vorteilhaft erwiesen, so daß jetzt zuverlässige Werkzeuge zur Vorhersage von ER-Signalpeptiden, mitochondrialen Targetingsignalen und Chloroplasten Signalpeptiden vorliegen. Eine Vorhersage von Schnittstellen ist bis jetzt noch nicht vollständig möglich [14].

In dieser Arbeit wird eine Methode vorgestellt, mit deren Hilfe es möglich ist, die Schnittstellen von humanen sekretorischen Proteinen zu identifizieren.

## 1.b Konformation der Peptidylprolylbindung

Die Struktur von Proteinen kann eine nützliche Zwischenstufe zur Bestimmung der Sequenz-Aktivitäts-Relation darstellen. Die dreidimensionale Struktur hängt auch von der *cis/trans*-Isomerie der Peptidbindung ab. Eine Analyse aller bekannten 3D-Strukturen von Proteinen ergab, daß etwa 25% aller Proteine mindestens eine *cis*-Peptidgruppe enthalten. Insgesamt liegen nur 0,36% aller Peptidbindungen in einer *cis*-Konformation vor [117]. Von diesen 0,36% wiederum sind über 90% Xaa-Pro-Gruppen (, wobei Xaa für eine beliebige Aminosäure und Pro für Prolin steht). Prolin spielt somit bei der Entstehung einer *cis*-Peptidgruppe eine wichtige Rolle.

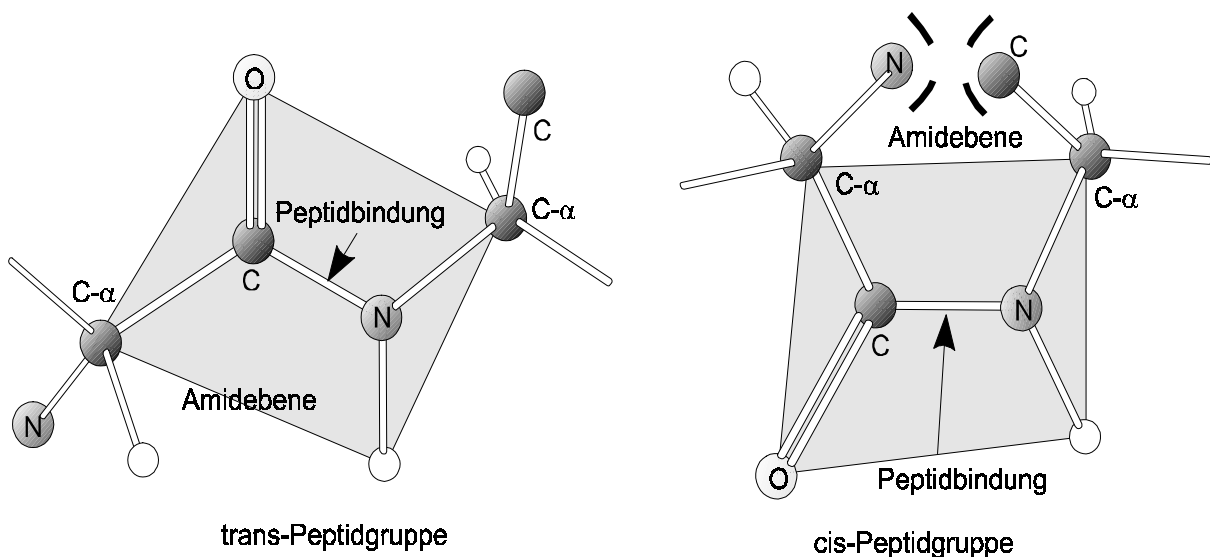


ABBILDUNG 3: Beschreibung der *cis*- und *trans*- Isomere einer Peptidgruppe. Weiße Kreise: Wasserstoffatome; alle anderen Atome sind beschriftet. Die sterische Hinderung der *cis*-Konformation ist durch gestrichelte Linien gekennzeichnet. O steht für Sauerstoff, C für Kohlenstoff, N für Stickstoff.

Der unterschiedliche Aufbau von *cis*- und *trans*-Peptidisomeren ist in Abbildung 3 dargestellt. Es sind nur zwei Diederwinkel frei beweglich, nämlich der  $\phi$  und der  $\psi$ -Winkel (Abbildung 4). Der  $\omega$ -Winkel nimmt im allgemeinen Werte im Bereich von  $0^\circ$  (*cis*) oder  $180^\circ$  (*trans*) an. Daraus resultiert eine ebene Konformation der Amidbindung.

Bestimmte Enzyme, die Peptidyl-prolyl-isomerasen (PPIasen), sind in der Lage die Umwandlung von *trans* nach *cis* zu katalysieren. Wenn diese Enzyme sequenzspezifisch arbeiten und das Signal für die Isomerisierung lokal kodiert ist, sollte es möglich sein, das Problem der Konformationseinstellung der Xaa-Pro-Bindung unabhängig von der Faltung des Gesamtproteins zu betrachten. Deshalb wurde die Konformationsvorhersage als Beispiel für die Sequenz-Struktur-Beziehung gewählt. Weiterhin ist Prolin an der Ausbildung von  $\beta$ -II-Faltblattstrukturen beteiligt, die, wie im Fall des Fibrinogen-Rezeptorantagonisten, erfolgreich zum Entwurf von Peptidomimetika benutzt wurden [7].

### 1.c Peptid Docking

Manchmal ist die Struktur des Proteins bekannt, für das ein Bindungspartner gesucht wird. Dies trifft auch auf das Schnupfenvirus, das humane Rhinovirus HRV-14, zu. Rossmann und Mitarbeiter haben 1985 die Struktur eines kompletten Rhinovirus Serotyp 14 aufgeklärt [100]. Das Rhinovirus ist Mitglied der Familie der Picornaviren. Sie sind kleine ikosaedrische RNA-

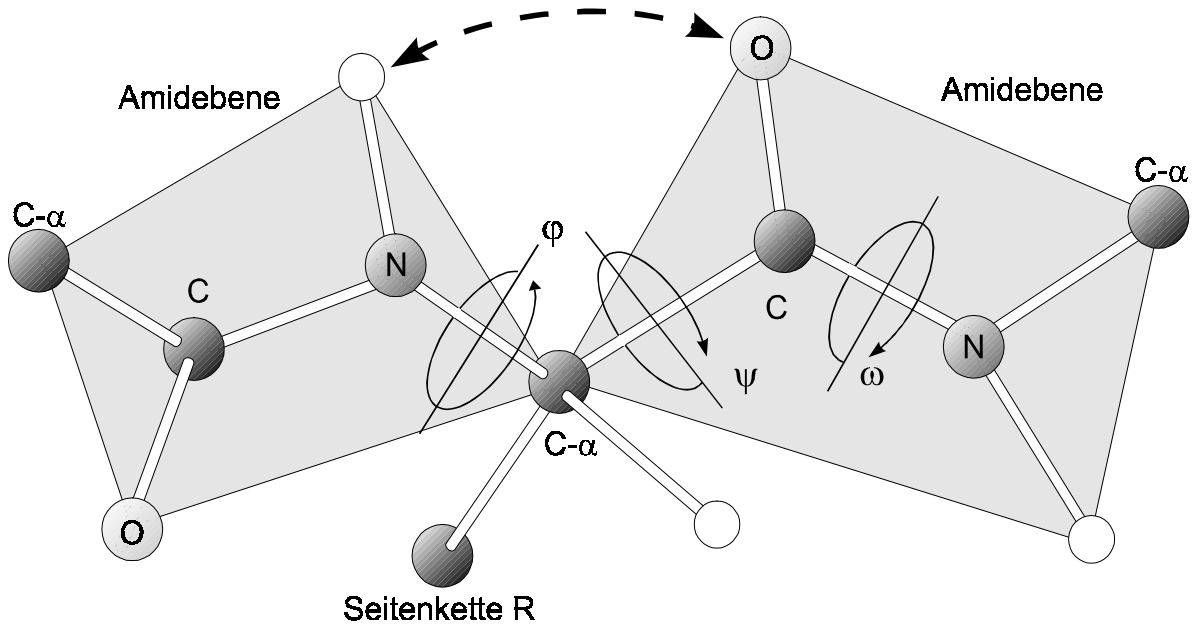


ABBILDUNG 4: Teil einer Polypeptidkette mit den Torsionsfreiheitsgraden einer Peptideinheit. Die einzelnen freien Drehungen des Proteinrückgrats sind entlang der  $C\alpha$ -N-Bindungen ( $\Phi$ ) und der  $C\alpha$ -C-Bindung ( $\Psi$ ) möglich. Die Torsionswinkel betragen alle  $180^\circ$  in der gezeigten Konformation und erhöhen sich im Uhrzeigersinn, wenn man aus Richtung des zentralen  $C\alpha$ -Atoms schaut. Die gestrichelte Linie zeigt die Wechselwirkung des Amidprotons mit dem räumlich benachbarten Sauerstoffatom. Weiße Kreise: Wasserstoffatome.

Viren und die Hauptursache für den gewöhnlichen Schnupfen. Die Hülle des Picornavirus besteht aus 60 Kopien von je vier Proteinen, VP1, VP2, VP3 und VP4. Ein 2,5 nm tiefer und bis zu 3,5 nm breiter Canyon [98, 99] umgibt die fünfzählige Symmetrieachse der Virusoberfläche. Bindungsstellen für neutralisierende Antikörper, deren Fab-Fragmente einen Durchmesser von ungefähr 3,5 nm haben, liegen an den hochvariablen Regionen, die durch die Loops zwischen den  $\beta$ -Ketten gebildet werden (Abbildung 5). Diese Stellen befinden sich am oberen Rand des Canyons, der relativ unabhängig von der Aminosäurezusammensetzung zu sein scheint. Die tieferliegenden Regionen sind weniger variabel und stellen die Erkennungsregion für das interzelluläre Adhäsions Molekül 1 (ICAM-1, [28]), den Zellrezeptor, dar.

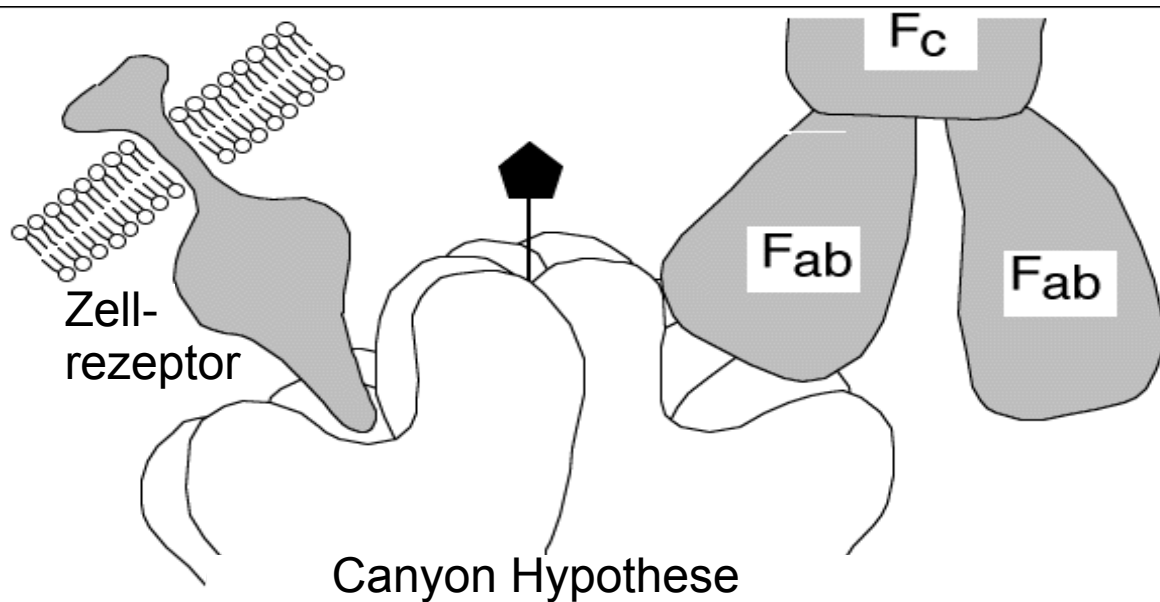


ABBILDUNG 5: Modell der Canyonregion des Rhinovirus 14 (weiß) mit einem Antikörpermolekül (rechts) und dem Zellrezeptor nach [100]. Die fünfzählige Rotationsachse ist ebenfalls eingezeichnet.

Einen Schutz der Zelle vor dem Virus kann zum einen durch die Blockade des Zellrezeptors erreicht werden. Eine weitere Möglichkeit bietet die Blockierung der weniger variablen tiefen Canyonregion des Virus. Es wurde hier eine Methode entwickelt, um, ausgehend von der Struktur der Canyonregion, Peptide zu entwerfen, die eine zellprotektive Eigenschaft besitzen. Dazu wurden bekannte Protein-Protein-Komplexe auf charakteristische Aminosäure-Bindungsmuster zwischen den Bindungspartnern untersucht. Die daraus abgeleiteten Wahrscheinlichkeiten für Wechselwirkungspartner wurden beim Einpassen von Peptiden in die Canyonregion berücksichtigt. Es wurde ein Peptid mit wahrscheinlichen Aminosäuren und eines mit weniger wahrscheinlichen Aminosäuren vorgestellt. Diese Peptide wurden auf ihre zellprotektiven Fähigkeiten untersucht.

Daneben wurden, ausgehend von Peptiden mit nachgewiesenen zellprotektiven Effekten, fünf weitere Sequenzen generiert und ebenfalls biochemisch untersucht. Diese wurden mit der bereits mehrfach bewährten Simulierten Molekularen Evolution [104] generiert.