

„Es ist eben durchaus und für immer unbegreiflich, daß es einer Anzahl von Kohlenstoff-, Wasserstoff-, Stickstoff-, Sauerstoffatomen nicht sollte gleichgültig sein, wie sie liegen und sich bewegen, wie sie lagen und sich bewegten, wie sie liegen und sich bewegen werden.“  
Emil Dubois-Reymond, „Über die Grenzen des Naturerkennens“, 1872

## 1. Einleitung

### 1. 1 Biologischer Hintergrund

#### 1. 1. 1 Antikörper und ihre Rolle im Immunsystem

Jeder Organismus braucht zum Überleben in seiner Umwelt ein Immunsystem, das ihn vor dem Angriff von Krankheitserregern schützt und degradiertes körpereigenes Material abbaut. Bei höheren Organismen wie Wirbeltieren wird diese Aufgabe durch zwei verschiedene Mechanismen sichergestellt: T-Lymphozyten ermöglichen den Schutz vor intrazellulären Infektionen, während auf extrazellulärer Ebene die von B-Lymphozyten produzierten Antikörper Fremdorganismen, körperfremde Substanzen und Toxine erkennen und deren Vernichtung bewirken. Diese als Antigene bezeichneten Stoffe sind jedoch prinzipiell aus den gleichen molekularen Bausteinen aufgebaut wie das körpereigene Material. Es ist daher einerseits von vitaler Wichtigkeit, daß Antikörper hochspezifisch zwischen körperfremden und –eigenen Makromolekülen unterscheiden können, aber andererseits so adaptiv sind, daß sie diverse (und potentiell unbekannte) Erreger erkennen und binden können. Diese sich grundsätzlich widersprechenden Anforderungen wurden von der Natur durch die besondere räumliche Struktur der Antikörper sowie durch den Prozeß ihrer Produktion und Reifung durch klonale Selektion im Organismus (Burnet, 1959) in perfekter Funktionalität zusammengeführt.

Antikörper, auch als Immunoglobuline bezeichnet, sind Glykoproteine mit einer Y-förmigen dreidimensionalen Struktur. Sie bestehen aus vier Polypeptid-Ketten, die durch Disulfid-Brücken und nichtkovalente Wechselwirkungen verbunden sind: je zwei identische schwere Ketten (H-Ketten, 50 bis 60 kDa) und leichte Ketten (L-Ketten, ca. 25 kDa). Es werden acht konstante (keine sequentiellen Unterschiede innerhalb einer Klasse von Immunoglobulinen) und vier variable Domänen unterschieden, die je von achtsträngigen antiparallelen  $\beta$ -Faltblättern gebildet werden und etwa 110 Aminosäuren umfassen. Bei den meisten Säugetieren werden fünf Immunoglobulin-Isotypen an Hand von Sequenzvergleichen der konstanten Domänen unterschieden: IgM, IgD, IgG, IgE und IgA. Der „Stamm“ eines Antikörpers, der auch als  $F_c$ -Fragment bezeichnet wird, besteht aus konstanten Domänen und

wird von den schweren Ketten gebildet. Die beiden „Äste“, die F<sub>ab</sub>-Fragmente genannt werden, werden gemeinsam von schweren und leichten Ketten geformt. Der variable Teil eines F<sub>ab</sub>-Fragmentes, als F<sub>v</sub>-Fragment bezeichnet, formt am N-terminalen Teil die eigentliche Antigen-Bindungsstelle oder Paratop, das aus je drei Loops der schweren Kette (V<sub>L</sub>) und der leichten Kette (V<sub>H</sub>) gebildet wird. Diese sechs hypervariablen Regionen werden auch CDR's (*complementarity determining regions*) genannt. Die Aminosäure-Zusammensetzung in den CDR's bestimmt die räumliche und elektronische Struktur der Antigen-Bindungsstelle und damit ihre molekulare Spezifität, d. h. welches Epitop mit einer biologisch relevanten Bindungsaffinität erkannt wird (Chothia et al., 1989; Mian et al., 1991). Die Bindung eines Antigens erfolgt ausschließlich über nichtkovalente Interaktionen.

Von einem körperfremden Protein, einem Antigen, wird im Verlauf der Bindung an einen Antikörper nur ein kleiner Teil erkannt. Dieser als Epitop oder antigene Determinante bezeichnete Proteinabschnitt kann in der Primärstruktur aus einem Segment bestehen, was als lineares bzw. kontinuierliches Epitop bezeichnet wird; entsprechend findet sich ein diskontinuierliches bzw. Konformationsepitop auf verschiedenen Abschnitten der Aminosäuresequenz, die durch die Faltung des Antigen-Proteins aber räumlich eng benachbart sind und deshalb zusammen vom Paratop erkannt werden (Barlow et al., 1986; Van Regenmortel, 1987). In den meisten Fällen befindet sich die von B-Zellen erkannte Epitopregion auf der Proteinoberfläche des antigenen Moleküls, es können allerdings auch in Ausnahmefällen durch Prozessierung der Antigene und damit verbundene Auffaltung Sequenzen, die sich vormals im unzugänglichen Kernbereich des Proteins befanden, gebunden werden. Im Prozeß einer humoralen Immunantwort wird jedoch meist das vollständige Antigen von membranständigen Antikörpern der B-Zelle gebunden und anschließend internalisiert. Im Proteasom wird dieses Fremdprotein dann proteolytisch zerlegt und die Fragmente zusammen mit MHC-Molekülen (*major histocompatibility complex*) an die Zelloberfläche transportiert, wo diese die Peptidabschnitte präsentieren. Peptidbeladene MHC-Moleküle der Klasse II binden wiederum epitopspezifisch T-Helferzellen. In einer Folgekaskade wird die Produktion verschiedener Cytokine ausgelöst, was zur Differenzierung der gebundenen B-Zelle in Antikörper-sezernierende Zellen führt. Wird ein Antigenfragment dagegen im Verlauf einer zellulären Immunantwort, also etwa bei der viralen Infektion einer normalen Zelle, von einem MHC-Protein der Klasse I präsentiert, führt dies in der Folge zur Bindung an T-Killerzellen oder cytotoxische T-Zellen, was die Vernichtung der infizierten Zelle auslöst.

### ***1. 1. 2 Polyspezifität und Kreuzreaktivität von Antikörpern***

Die hohe Spezifität der Wechselwirkung von Antikörpern, die durch die besondere Struktur jeder *antibody combining site* (ACS) determiniert ist, macht diese zu wertvollen Werkzeugen der medizinischen Diagnostik und Forschung. Besonders seit der Einführung der Hybridomtechnik (Köhler et al., 1975) zur Erzeugung monoklonaler Antikörper hat der Nachweis und das Design biologisch aktiver Moleküle und Substanzen enorme Fortschritte gemacht (Cook et al., 1988; Banting et al., 1989; Saragovi et al., 1991). Die meisten Antikörper erkennen diskontinuierliche Epitope, die eine native Antigen-Konformation für eine Bindung voraussetzen; ca. 10% aller Antikörper allerdings binden lineare Epitope (Pellequer et al., 1991). Durch Immunisierung mit synthetisch hergestellten Peptiden der entsprechenden Sequenz kann man sogenannte Anti-Peptidantikörper gewinnen (Walter, 1986), die dann auch mit hoher Affinität an das ursprüngliche Antigen binden. Darüber hinaus wurde beobachtet, daß auch Peptide, die Teile eines diskontinuierlichen Epitops darstellen, mit geringer Affinität von den entsprechenden Antikörpern gebunden werden (Gao et al., 1996; Reineke et al., 1996). Dieses als Kreuzreaktivität zwischen vollständigem Protein und abgeleitetem Peptid bezeichnete Phänomen wurde zunächst für ein vereinzelt Phänomen von geringer biologischer Relevanz gehalten. Generell war über lange Zeit die vorherrschende Meinung, daß hochaffine monoklonale Antikörper auch monospezifisch sind, also nur ein bestimmtes Epitop zu erkennen. Die Möglichkeit, große Peptidpools zu synthetisieren und untersuchen, hat inzwischen jedoch zu einer Reihe von Studien geführt, in denen festgestellt wurde, daß viele monoklonale Antikörper polyspezifisch mehrere Epitope binden können, d. h. daß sie sequentiell und strukturell unterschiedliche Peptide mit dergleichen Paratopregion erkennen (Duggan et al., 1988; Pinilla et al., 1995; Abakushin et al., 1996; Appel et al., 1996; Kramer et al., 1997; Borisova et al., 1999). Aus struktureller Sicht wurde am Beispiel des Antikörpers CB4-1 festgestellt, daß die Bindung verschiedener Peptidsequenzen, die in unterschiedlichen Konformationen vorliegen, durch unterschiedliche Seitenkettenkontakte mit den CDR's vermittelt wird (Keitel et al., 1997). Darüber hinaus konnte für denselben Antikörper gezeigt werden, daß zwei sequentiell nicht verwandte Epitope ineinander durch sukzessiven Austausch einzelner Positionen so ineinander überführt werden können, daß bei jedem Zwischenschritt eine Bindung zu dem Paratop beibehalten wird (Hoffmüller et al., 2000). Diese Erkenntnis hat weitreichende Implikationen für die Theorie der Evolution von

Protein-Protein-Wechselwirkungen als auch für die Untersuchung möglicher Entstehungswege von Autoimmunkrankheiten (Olds et al., 1991; Ordonez et al., 1995).

### ***1. 1. 3 Molekulare Bibliotheken als Mittel zur Aufklärung von Protein-Peptid-Wechselwirkungen***

Peptide dienen oft als Leitstrukturen im Prozeß der Entwicklung biologisch aktiver Substanzen. Die Entwicklung synthetischer Peptidbibliotheken seit der Mitte der 80er Jahre (Geysen et al., 1986; Houghten et al., 1991; Lam et al., 1991; Schneider-Mergener et al., 1994) sowie die Einführung biologisch generierter Peptidrepertoires auf Phagenoberflächen (Cwirla et al., 1990; Devlin et al., 1990; McCafferty et al., 1990; Scott et al., 1990; Scott, 1992) bedeuteten einen großen Beitrag zur systematischen hochaufgelösten Untersuchung der Wechselwirkungen von Peptiden mit ihren Bindungspartnern. Diese Methoden haben in der Diagnostik und dem Wirkstoffdesign bereits breite Verwendung gefunden (Edgington, 1993; Schneider-Mergener et al., 1994; Houghten, 1995). Inzwischen gibt es eine Vielzahl von Bibliotheksansätzen, mit denen aus einem enormen Sequenzraum (ein Peptid der Länge  $N$  hat  $N^{20}$  mögliche Varianten) durch unterschiedliche Adressiervarianten bindende Peptide identifiziert werden können – z. B. Zufallsbibliotheken, kombinatorische Bibliotheken mit zwei (Geysen et al., 1986) oder drei (Kramer et al., 1995) festgelegten Positionen oder *positional scanning* Bibliotheken (Pinilla et al., 1993; Schneider-Mergener et al., 1996). Von besonderer Bedeutung für die detaillierte Aufklärung des Beitrags einzelner Aminosäuren zur Bindung ist die Substitutionsanalyse, die auch, wie die oben erwähnten Bibliotheken, mittels Spot-Synthese (Frank, 1992) auf kontinuierliche Cellulose-Träger halbautomatisch aufgebracht werden kann. Hierbei werden alle Positionen in einem Peptid systematisch durch alle 20 (bzw. 19 unter Vermeidung von Cystein wegen dessen leichter Oxidierbarkeit) natürlichen Aminosäuren ausgetauscht, so daß man für ein Peptid der Länge  $N$  ein Repertoire von  $N \cdot 19$  (bzw.  $N \cdot 18$ ) Einzelsubstitutionen der nativen Sequenz erhält. Nach der Inkubation einer solchen Membran mit dem entsprechenden Targetmolekül und der anschließenden Visualisierung (Stigler et al., 1995) ist an Hand des Musters der Substitutionsmatrix sofort die Wichtigkeit jeder Aminosäure an den verschiedenen Positionen in der Sequenz ablesbar. Je nach der Zugehörigkeit der entsprechenden Aminosäuren an Schlüsselpositionen (*key residues*) zu Gruppen physikochemisch ähnlicher Residuen kann man darauf schließen, ob die Wichtigkeit einer bestimmte Aminosäure an einer bestimmten Position strukturell, sterisch und/oder elektronisch determiniert ist. Diese auch mathematisch faßbare relative Wichtigkeit einer einzelnen Aminosäure (Volkmer-Engert et al., 1995) kann jedoch summarische Effekte

aus strukturellen Faktoren der Bindung an den Antikörper und dynamischen Einflüssen der einzelnen Aminosäuren bei der Ausbildung einer bestimmten Vorzugskonformation des ungebundenen Peptids enthalten. Diese Effekte sind nicht ohne weiteres auseinander zu halten, da man stets nur Einzelaustausche betrachtet, aber keine Auswirkungen konzertierter Austausch sieht, die etwa intrapeptidische Wechselwirkungen unterscheiden lassen würden. In der Tat wurde z. B. für das C-terminale Leucin im Epitop GATPEDLNQKL, das an den anti-p24 (HIV-1) Antikörper CB4-1 bindet, nur ein sehr schwacher und unspezifischer Kontakt zur Paratop-Region in der Röntgenkristallstruktur gefunden (Keitel et al., 1997), obwohl es als essentiell in der Substitutionsanalyse erscheint.

## 1. 2 Struktureller Hintergrund

### *1. 2. 1 Peptide als Bindungspartner von Proteinen*

Peptide spielen bei einer Vielzahl biologischer Prozesse eine essentielle Rolle. Wie schon beschrieben, kommt ihnen eine Schlüsselposition in verschiedenen Wegen der Immunantwort in höheren Organismen zu. Darüber hinaus haben sie bedeutende regulatorische Funktionen bei der Steuerung von Körperfunktionen u. a. als Neuropeptide (van Hagen et al., 1999), im Hormonhaushalt (Pellegrini et al., 1999) und der Spermatogenese (Fraser, 1998). Paralytisch und hämolytisch wirkende Peptide wurden beschrieben (Subbalakshmi et al., 2000). Peptide aus der Familie der VIP-Peptide wurden charakterisiert als Wachstumsfaktoren für fötale und Tumorzellen (Gozes et al., 1999b; Hill et al., 1999) und während der embryonalen Gehirnentwicklung (Rayan et al., 1991; Gozes et al., 1999a), sie sind bei der Schmerzwahrnehmung involviert (Dickinson et al., 1999), bei der Atmungsregulierung (Maggi et al., 1995) usw. Ihre Beteiligung bei einer Vielzahl von Krankheiten, u. a. Asthma (Boomsma et al., 1992), wird diskutiert. Für viele dieser Peptide gibt es hochspezifische Rezeptoren auf der Zellmembranoberfläche (Pellegrini et al., 1999) einerseits und andererseits Transportproteine wie OppA, die völlig sequenzunabhängig kurze Peptide binden und umschließen (Tame et al., 1994). Viele niedere Organismen produzieren amphiphile Peptide und Peptomere, die unspezifisch membranzerstörend wirken, zur antibiotischen Abwehr, was auch medizinisch genutzt wird (Stachelhaus et al., 1996). In vielen Signaltransduktionswegen werden von kleinen Proteindomänen, wie WW- und SH3-Domänen, kurze, prolinhaltige Peptidabschnitte erkannt (Kay et al., 2000). Eine große Zahl von giftigen Tieren und Pflanzen produzieren cyclische und makrocyclische Peptide, deren hochtoxische Wirkung auf der

Blockierung von Ionenkanälen und auf Enzyminhibition beruhen (Favel et al., 1989; Myers et al., 1993; Lipkind et al., 1997).

Die erstaunliche Diversität der biologischen Funktionen beruht auf der Fähigkeit von Peptiden, abhängig von ihrer Sequenz eine Vielzahl von Konformationen und Strukturmotiven auszubilden. Antikörper binden bevorzugt Peptide mit  $\beta$ -Turns (Wilson et al., 1994), können aber im Gegensatz zu anderen Peptid-bindenden Proteinen, wie MHC, SH2- und SH3-Domänen oder Calmodulin, auch andere Konformationen erkennen (Stanfield et al., 1995). So gibt es unter den 38 bekannten dreidimensionalen Strukturen von Antikörper-Peptid-Komplexen auch Bindermotive mit kurzen  $3_{10}$ -Helices und  $\alpha$ -Helices, die über die ganze Peptidlänge reichen. In beladenen MHC-Molekülen liegen die Peptide in gestreckter Form vor, was die Exposition von Ankerpositionen für die T-Zellrezeptorbindung ermöglicht. Wie auch die PDZ-Domänen binden MHC-Moleküle der Klasse I nur Peptide mit freien C-Termini, die über Salzbrücken in speziellen Taschen fixiert werden. Von Peptiden der VIP-Familie wiederum wird angenommen, daß sie überwiegend in  $\alpha$ -helicaler Form an ihren Rezeptor binden (Nicole et al., 2000), ebenso liegen die Liganden von Calmodulin als  $\alpha$ -Helices vor (Wilkinson, 1996). Die Wichtigkeit von repetitiven Prolin-Motiven wird bei Bindern von WW- und SH3-Domänen sowie Profilin in der Herausbildung von Polyprolin-Helices reflektiert, die eine Versteifung der kurzen Peptide und somit die Bindung an die  $\beta$ -Faltblätter der Domänen ermöglicht (Wilkinson, 1996). Eine enorme konformationelle Stabilisierung wird bei cyclischen und makrocyclischen Peptiden erreicht, bei denen bis zu vier Disulfidbrücken eine „Knottin“-Struktur erzeugen. Letztere Peptidgruppe kann bereits als Mini-Proteine angesprochen werden, da sie im zeitlichen Mittel eine hochdefinierte Sekundär- und Tertiärstruktur besitzen (Craik et al., 1999).

Antikörper-Peptid-Wechselwirkungen sind gut untersuchte Modellsysteme für Protein-Protein-Interaktionen im allgemeinen (Van Regenmortel, 1994). Eine Untersuchung an 68 monoklonalen IgG-Antikörpern und ihren linearen Epitopen (Dong, 1998) hat gezeigt, daß ein Peptidepitop vier bis zwölf, im Durchschnitt acht Aminosäuren lang ist. Durchschnittlich 70% der Residuen in einem Epitop sind essentiell für die Bindung, wobei unter den Schlüsselaminosäuren Tryptophan, Prolin, Asparaginsäure und Phenylalanin besonders häufig auftreten. Bis auf Phenylalanin sind diese Aminosäuren auch in allen Epitopen signifikant überrepräsentiert. Das gehäufte Vorkommen von Prolin bestätigt die bevorzugte Erkennung von  $\beta$ -Turns. Auf der Seite der Antikörper besteht etwa die Hälfte der 15-20 Aminosäuren in der Paratop-Region aus aromatischen Residuen (Braden et al., 1995) und in der Regel sind 5 oder alle 6 CDRs beteiligt, wobei meist die CDR3 der schweren Kette von besonders hoher

Diversität und damit auch Wichtigkeit ist. Bei der Bindung zwischen Antikörper und Antigen bzw. Peptid kann es zu konformationellen Änderungen bis hin zu einer geringen gegenseitigen Verschiebung der Domänen kommen. Der dabei zu Grunde liegende Mechanismus ist noch nicht aufgeklärt, besonders die Theorien des „induzierten Fit“ (Bhat et al., 1990) bzw. des „Bindungstrichters“ (Kumar et al., 2000) werden diskutiert.

### ***1. 2. 2 Faltungszustand von Peptiden***

Auf Grund ihrer großen inhärenten Flexibilität besitzen Peptide in der Regel in wässriger Lösung keine ausgeprägte Sekundärstruktur. In Abhängigkeit von der Peptidlänge können bestimmte Sequenzen jedoch definierte Konformationspopulationen im zeitlichen Mittel stärker besuchen, eine Tatsache, die zur Untersuchung von Peptiden als Modellsysteme für frühe Stadien der Proteinfaltung benutzt wird (Dyson et al., 1991, 1993). Bekannt ist, daß bereits das Tripeptid Arg-Gly-Asp bevorzugt eine  $\beta$ -Turnstruktur einnimmt (Kang et al., 2000). Bei NMR-Untersuchungen an immunogenen Peptiden wurde gezeigt, daß auch das Tetrapeptid Tyr-Pro-Tyr-Asp eine ausgeprägte  $\beta$ -Turnstruktur ausbildet, wobei die dritte Position dafür entscheidend ist (Dyson et al., 1988a). Von den häufiger vorkommenden Sekundärstrukturmotiven werden für Peptide besonders  $\alpha$ -Helices,  $\beta$ -Faltblätter und -Haarnadeln sowie  $\beta$ -Turns angegeben. Helices und Turns falten dabei schneller und sind wahrscheinlicher, da sie durch lokale Wechselwirkungen gebildet werden (Lauer et al., 1998). Stabile  $\alpha$ -Helices werden insbesondere durch interhelicale Wechselwirkungen in helicalen Haarnadeln beobachtet (Fezoui et al., 1995). Obwohl die Interaktionen in Helices und Turns lokaler Natur sind, werden sie doch meist durch fernwirkende Wechselwirkungen stabilisiert. Ein aus einem helicalen Abschnitt eines nativen Proteins abgeleitetes 19-meres Peptid zeigte ein Ensemble schnell ineinander übergehender turn-artiger Strukturen, eine sog. nascente Helix, die nur durch weitere nicht-lokale Wechselwirkungen stabilisiert werden konnte (Dyson et al., 1988b). Es wird vermutet, daß multiple Turnstrukturen und nascente Helices sehr häufig bei kurzen Peptiden vorkommen (Lauer et al., 1998). Andererseits sind stabile  $\alpha$ -Helices bei kurzen, monomeren Peptiden beobachtet worden: Bereits ein 10-meres Peptid kann bei Raumtemperatur 51% helical sein, ein 17-meres war es zu 100% (Krstenansky et al., 1989). Untersuchungen an Decapeptid-Hairpins haben ergeben, daß sie stabile antiparallele Faltblätter bilden können, wobei in Abhängigkeit von der Sequenz im Turnbereich unterschiedliche Faltblatt-Paarungen zustande kommen (De Alba et al., 1999). Ein in Wasser monomeres 16-meres Peptid, das aus Protein G abgeleitet wurde, nimmt eine

Konformationsfamilie ein, die bis zu 40% native  $\beta$ -Haarnadelstruktur enthält (Lauer et al., 1998). Es ist auch beobachtet worden, daß ein aus BPTI abgeleitetes 20-meres Peptid abhängig von der elektrostatischen Umgebung sowohl eine  $\alpha$ -Helix als auch  $\beta$ -Faltblattstrukturen herausbilden kann (Nakazawa et al., 1999). Bekannt ist das 20-meres Betanova, das ein dreisträngiges  $\beta$ -Faltblatt aufweist (Kortemme et al., 1998). Darüber hinaus sind eine Reihe von Peptiden beschrieben worden, die gemischte  $\alpha/\beta$ -Architektur besitzen (Struthers et al., 1997). Ein erstaunliches  $\alpha/\beta$ -Peptid ist kürzlich vorgestellt worden, das bei einer Länge von 20 Aminosäuren ohne Disulfidbrücken zu 100% gefaltet ist und eine Schmelztemperatur von 61°C besitzt (Andersen et al., 2001). Auch die durch Disulfidbrücken stabilisierten Knottine haben in der Regel gemischte Struktur motive, die eine wohldefinierte Tertiärstruktur bilden. Charybdotoxin z. B. kann durch eine kompakte  $\beta\alpha\beta\beta$ -Topologie beschrieben werden (Imperiali et al., 1999). Hier ist der Übergang zwischen Peptid und Protein erreicht.

Für die Herausbildung einer Vorzugskonformation einer Peptidkette ist einerseits die physikochemische Umgebung entscheidend und andererseits die Kombination aus Aminosäurezusammensetzung und Länge der Kette. Die Zusammensetzung ist dabei maßgebend für stabilisierende oder schwächende Kontakte, während die Kettenlänge die Entropie des Systems bestimmt bzw. ab einer gewissen Länge auch erst eine Tertiärstruktur ermöglicht. Ein interessanter Fall sind ionische selbstkomplementäre Peptide, die in Abhängigkeit von pH-Wert und Temperatur reversibel zwischen stabilen  $\alpha$ -Helices und  $\beta$ -Faltblättern alternieren können (Altman et al., 2000), was als Testmodell für die Amyloid-Bildung diskutiert wird. Für kleine Peptide kann der Gewinn an freier Energie, der sich durch Faltung ergibt, so minimal sein, daß er durch die thermische Energie der Umgebung ständig aufgehoben wird. Andererseits kann eine Konformation mit geringer Population bevorzugt von dem entsprechenden Rezeptor erkannt werden, was zu einer erhöhten Population ebendieser Konformation führt, d. h. die Wechselwirkung mit einem Partner führt zu einer Konformationsstabilisierung (Lauer et al., 1998; Tsai et al., 1999a). Ob es sich dabei wie bei der Proteinfaltung um einen internen Interaktionspartner handelt oder um ein anderes Molekül wie im Falle der MHC-Bindung (MHC-Moleküle sind instabil im unbeladenem Zustand), spielt keine Rolle. Bei jeder Bindung aber handelt es sich wie bei der Proteinfaltung um die Einfrierung einer Konformation.

Im allgemeinen kann gesagt werden, daß abgeleitete Peptide, die die gleiche biologische Aktivität wie das parentale Protein haben, als wohlstrukturiert gelten können (Pincus, 1992), was für die im Hauptteil besprochenen Antikörper-Peptid-Wechselwirkungen von großer

Bedeutung ist. Durch das Fehlen von stabilisierenden lang-reichweitigen Wechselwirkungen kommt es jedoch häufig bei Peptiden zu Aggregation und Oligomerisierung. Daran sind insbesondere hydrophobe Kontakte beteiligt.

### ***1. 2. 3 Peptid-Design und Konformationsstabilisierung***

Das zunehmende Verständnis der Wechselwirkungen innerhalb von Proteinen und bei der nichtkovalenten Bindung von Liganden hat in den letzten Jahren zu erfolgreichen Versuchen des Peptid- und Proteindesigns geführt. Ziele sind einerseits die Konformationsstabilisierung erwünschter Strukturen und andererseits das molekulare Mimikry, d. h. das Nachbilden von räumlichen Anordnungen größerer Einheiten in kleinen Liganden. Die gezielte de novo Erzeugung von  $\alpha$ -helicalen Motiven, wie beim bereits erwähnten  $\alpha\alpha$ -Peptid (Fezoui et al., 1995) scheint dabei leichter erreichbar zu sein als das Design stabiler, monomerer  $\beta$ -Strukturen, wie im Betanova-Protein (Kortemme et al., 1998). Die gut untersuchte Häufigkeitsverteilung der einzelnen Aminosäuren in Helices dient hier als wissensbasierte Vorgabe zur Herausbildung gezielter Eigenschaften. Von besonderem Interesse sind das Gesamt-Dipolmoment und die Verteilung von polaren und hydrophoben Wechselwirkungen entlang der Helixachse. Die Verteilung geladener Residuen verhindert Dimerisierung, während C- und N-terminale Capping-Motive die Länge der Helix definieren. Bei  $\beta$ -Hairpins wird besonderes Augenmerk auf die Turn-Sequenz gelegt, die erheblich die Stabilität der Sekundärstruktur beeinflussen kann (Imperiali et al., 1999). Der I' $\beta$ -Turn in Betanova dient z. B. nicht der Faltblattbildung, die wesentlich durch Seitenketten-Wechselwirkung vorangetrieben wird, während der II' $\beta$ -Turn mit einer DPro-Gly Sequenz in einem ähnlichen Peptid (Schenck et al., 1998) essentiell für die Gesamtstruktur ist. Der Einsatz von LPro im Turn führt hier bereits zu einem Verlust der Sekundärstruktur.

Für das rationale Design von Peptidstrukturen werden häufig neben D-Aminosäuren auch andere ungewöhnliche Aminosäuren, wie Aminoisobuttersäure, Ornithin oder Norleucin, sowie  $\beta$ -Aminosäuren und N-alkylierte Aminosäuren eingesetzt, wobei man im letzteren Fall Peptomere bzw. Peptoide erhält. Neben einer durch natürliche Residuen nicht erreichbaren Spezifität zeichnen sich solche modifizierten Peptide durch eine erhöhte Resistenz gegenüber enzymatischem Abbau durch Proteasen aus. Peptide mit D-Aminosäuren kommen auch natürlich vor; sie lassen sich im Labor aus kombinatorischen D-Bibliotheken gewinnen (Dooley et al., 1994), durch Spiegelbild-Phagendisplay (Schumacher et al., 1996) oder rationelles Design (Jameson et al., 1994). D-Enantiomere und Retro-Inverso-Isomere von

natürlichen Epitop-Peptiden wurden erfolgreich in der Impfstoffentwicklung und der Immundiagnostik eingesetzt (Guichard et al., 1994; Van Regenmortel et al., 1998).

Wenn ein Protein ein anderes Protein bindet, besitzt dieser Partner bereits die räumliche Struktur, für die die Bindungsstelle spezifisch ist. Ist der Bindungspartner allerdings ein Peptid, muß es sowohl in der Nähe der Bindungsstelle sein als auch bei Annäherung die Konformation ausbilden, mit der es erkannt wird. Obwohl die weitreichenden elektrostatischen Kräfte, die von der Bindungsstelle ausgehen, die Ausbildung dieser Konformation erleichtern, ist die Fixierung aus einem großen Ensemble zeitlich nebeneinander vorliegender Strukturen ein Grund dafür, daß Peptide in der Regel mit einer deutlich geringeren Affinität an Proteine binden. Bereits seit längerer Zeit ist bekannt, daß die Bindungsenergie eines Peptids direkt proportional zur freien Energie ist, die nötig ist, um es in die Konformation zu falten, mit der es letztendlich an das Rezeptormolekül bindet (Sachs et al., 1972). Eine Möglichkeit, die Bindungsaffinität bei gleicher Spezifität zu erhöhen ist daher, von vornherein die bindende Konformation zu stabilisieren. Mehrere Methoden sind über die Jahre entwickelt worden, die es ermöglichen, verschiedene Struktur motive in ihren Konformationen zu restringieren (Kessler, 1990). Die Ausbildung einer Disulfid-Bindung durch zwei Cysteine bzw. Cystein-Analoga ist auch in der Natur die bevorzugte Methode, da sie durch einfache Oxidation und nur mit natürlichen Bausteinen zustande kommt. Allerdings sind Disulfid-Bindungen sehr anfällig für Reduktion und die aliphatische Kette der Cysteine und die Einfachbindung der Disulfidbrücke ermöglicht immer noch eine gewisse Flexibilität, die jedoch auch erwünscht sein kann, wenn z. B. die Geometrie der Bindungsstelle nicht hochaufgelöst bekannt ist. Chemisch stabiler sind Säureamidbindungen in Rückrat-Cyclisierungen (Gilon et al., 1991) und Pseudo-Peptidbindungen zwischen zwei Seitenketten oder einer Seitenkette und einem Terminus. Der partielle Doppelbindungscharakter dieser Verbrückung führt darüber hinaus zu einer weiteren Versteifung der Struktur. Eine interessante neuere Methode der Konformationsstabilisierung besteht darin, Wasserstoffbrückenbindungen, die für die jeweiligen Sekundärstruktur motive bestimmend sind, durch kovalente Bindungen zu ersetzen. So kann durch die Einführung einer Hydrazon-Bindung ( $N=N=CH-R$ ) ein Peptid sowohl in  $\alpha$ -helicaler Form (Cabezas et al., 1999) als auch in Turn-Konformation stabilisiert werden. Letzteres ist besonders für Antikörperbindung von Peptiden interessant, da sie dort bevorzugt als Loops erkannt werden. In der Tat gibt es auch seit kurzer Zeit die Kristallstruktur des Komplexes eines IgG1-Antikörpers mit einem aus HIV-1 gp120 abgeleiteten 12-meren Peptid, das durch Hydrazon-Verbrückung die Konformation des nativen Loops mimikriert (Stanfield et al., 1999). Mehrfach intern

verbrückte Peptide erhöhen zwar weiter die Rigidität gewünschter Sekundärstrukturen, sie sind aber wegen der erhöhten Komplexität synthetisch nur sehr schwer zugänglich. Hier bereits erzielte Ergebnisse gehen von bereits aus der Natur bekannten kompakten Molekülen wie den Knottinen aus; die gewünschte neue Spezifität wird dann durch Substitution von Sequenzabschnitten erzielt, die für die eigentliche Sekundärstruktur von geringer Bedeutung sind. So wurde kürzlich durch Einfügung einer 7-meren immunodominanten „Gastsequenz“ aus dem Herpes Simplex-Virus in die Struktur des  $\alpha$ -Conotoxins ein immunogenes chimäres Miniprotein erzeugt (Mezö et al., 2000).

### **1. 3 Methodischer Hintergrund**

#### ***1. 3. 1 Bedeutung und Methoden der Strukturvorhersage bei Peptiden und Proteinen***

Die Aufklärung der detaillierten dreidimensionalen Struktur von Peptiden bzw. Proteinen in monomerer und gebundener Form ist eine wichtige Voraussetzung zum Verständnis der Wechselwirkungen innerhalb und zwischen biologisch aktiven Molekülen und damit zur Aufklärung von Wirkmechanismen und der Entwicklung neuer Wirkstoffe. Peptide werden dabei als synthetisch gut zugängliche Stoffklasse mit enormer kombinatorischer Breite als Leitstrukturen für die Arzneistoff- (Li et al., 1995; Wrighton et al., 1996) und für die Impfstoffentwicklung (Brown, 1994) geschätzt. Neben den traditionellen physikalischen Verfahren zur Strukturaufklärung – Röntgenkristallstrukturanalyse und Kernmagnetresonanzspektroskopie – haben sich in den letzten Jahren, begünstigt durch zunehmende Verfügbarkeit starker Computerressourcen, zunehmend theoretische Methoden der Strukturvorhersage etabliert (Böhm, 1996).

Während Proteine in der Regel eine definierte Hauptkonformation einnehmen und somit die dreidimensionale Struktur, die man aus einer Röntgenkristall- oder NMR-Analyse entnehmen kann, auch als die in einem biologischen System vorliegende Struktur ansehen werden kann, liegen Peptide auf Grund ihrer inhärenten Flexibilität unter physiologischen Bedingungen als Ensemble mehr oder weniger definierter, sich ständig ineinander umwandelnder Konformationen vor. Bei der Bindung von Peptiden an Proteine wird dann aus dieser Mischung die „rezeptorgebundene“ oder biologisch aktive Konformation (oder Konformationen) selektiert, die aber in dem konformationellen Ensemble nicht das höchste statistische Gewicht aufweisen müssen (Nikiforovich, 1994). Darüber hinaus besitzen Peptide auch im gebundenen Zustand noch eine gewisse Flexibilität, besonders in Abschnitten, die nicht

unmittelbar an der Bindung beteiligt sind oder auch in den C- bzw. N-terminalen Bereichen. Dies begründet auch, warum physikalische Methoden der Strukturaufklärung nur eingeschränkt geeignet sind, die ganze Realität der Peptidbindung abzubilden. Bei der Röntgenstrukturanalyse wird nur diejenige Konformation des gebundenen Peptids erfaßt, die auch zur Kristallisation des Komplexes führt. Andere Peptidkonformationen, die eventuell auch zu einer Bindung führen, aber nicht im gebundenen Zustand auskristallisieren oder aber im Kristall vorkommen, aber ein geringes statistisches Gewicht haben, werden nicht detektiert. Das mag ein Grund dafür sein, warum Schlüsselaminosäuren, die sich bei Bindungsstudien wie Substitutionsanalysen als essentiell herausstellen, in der Komplexstruktur nicht immer befriedigend durch Rezeptorkontakte erklären lassen (Keitel et al., 1997). Darüber hinaus besteht eine gewisse Wahrscheinlichkeit, daß die Kristallpackung einzelne Kontakte lokal verzerrt und zu einer falschen Interpretation der Struktur führt. Nicht zuletzt führt darüber hinaus eine im gebundene Zustand verbleibende Flexibilität in Teilen des Peptids zu fehlender Elektronendichte und somit zum „Verschwinden“ der entsprechenden Bereiche in der 3D-Struktur.

Bei der Interpretation von NMR-Strukturen treten dagegen andere Unsicherheiten auf. Die Zeitauflösung der NMR-Methoden liegt physikalisch bedingt in einem Bereich von  $10^{-5}$  bis  $10^{-3}$  s. Die Umwandlungsrate der verschiedenen Konformere eines Peptids ist dagegen bedeutend höher, sofern man cis/trans-Isomerisierungen von Peptidbindungen außer acht läßt. Für den als Normalität anzunehmenden Fall, daß ein Peptid keine im zeitlichen Mittel hochdominante Konformation annimmt, wird man also durch die Kernmagnetresonanz immer eine durchschnittliche Konformation über alle in Lösung vorliegenden Strukturen erhalten, die mit den reellen Konformeren nicht verwandt sein muß (Nikiforovich, 1994). Für Komplexstrukturen ist darüber hinaus auch die Größe des Systems ein limitierender Faktor, da die Komplexität und die Überlappung der Signalmuster trotz der stürmischen Entwicklung neuer Pulssequenz-Programme und mehrdimensionaler NMR-Experimente in den letzten Jahren die Messung großer Proteine erheblich erschwert.

Die letzten Jahre haben mit der rapiden Entwicklung der Computertechnik große Fortschritte bei der Modellierung von Peptiden und Proteinen *in silico* erlebt. Im Mittelpunkt der Forschungen stand und steht dabei die *ab initio* Vorhersage des Faltungszustandes von Proteinen aus der Primärsequenz einerseits und die Strukturbestimmung von Komplexen andererseits, da diese beiden Aspekte unmittelbar mit der biologischen Funktionalität verbunden sind. Die allermeisten globulären Proteine besitzen eine einzige, kompakte Struktur, die dem globalen Minimum der freien Energiefunktion entspricht (Anfinsen et al.,

1975). Die Herausforderung beim „Faltungsproblem“ der theoretischen Biologie besteht darin, in einem vieldimensionalen Raum durch realistische Einschätzung der freien Energie von möglicherweise Milliarden Konformeren mit hinreichender Genauigkeit und in vertretbarer Zeit dieses Minimum des nativen Faltungszustandes zu identifizieren. Es gibt eine Reihe interessanter alternativer Ansätze, die Erzeugung und Evaluierung von so vielen Konformationen zu umgehen, z. B. in der Strukturvorhersage durch wissensbasierte Abstandspotentiale (Crippen, 1991; Jones et al., 1992; Wall et al., 1999), aber die meisten Arbeiten befassen sich doch mit der Optimierung der Konformationserzeugung und Energieabschätzung. Das Faltungsproblem findet seine Entsprechung im „Dockingproblem“, wenn einer der beteiligten Partner ein flexibles Molekül ohne native Struktur ist. In diesem Fall kann man den Docking-Vorgang als Faltung des flexiblen Moleküls im Potential des Bindungspartners auffassen. Ein darüber hinausgehendes Problem besteht beim Docking darin, daß sich die Konformation des rigiden Partner in Folge der Bindung lokal verändern kann, wobei es nach wie vor eine Kontroverse über die zu Grunde liegenden Mechanismen gibt („Induzierter Fit“ vs. „Bindungstrichter“-Szenarien, siehe Punkt 1. 2. 1).

### **1. 3. 2 Kraftfelder**

Für computergestützte Konformationssimulationen braucht man im wesentlichen zwei Komponenten: ein effizientes Protokoll zur Konformationserzeugung und ein geeignetes Kraftfeld, in dem die Energie dieser Konformationen bestimmt wird. Beide Komponenten sind in mehreren Reviews in den letzten Jahren gründlich besprochen worden (Bowen et al., 1991; Dinur et al., 1991; Leach, 1991; Smith et al., 1991; Scheraga, 1992; Tenette et al., 1996; Wang et al., 2001).

Ein Kraftfeld beschreibt ein Molekül als ein räumlich angeordnetes System von Atompunkten, zwischen denen Kräfte wirken, die durch Formeln beschrieben werden. Diese Kräfte kann man grob in zwei Klassen einteilen: Kräfte zwischen Atomen, die in der Valenzstruktur gebunden sind („bonded“-Wechselwirkungen) und zwischen nicht miteinander verbundenen Atomen („nonbonded“-Wechselwirkungen). Zu ersteren gehören Bindungsdehnungs- (Vibrations-), Bindungswinkelbeugungs- sowie Phasenwinkelkräfte. Die meisten Kraftfelder behandeln die Valenzgeometrie als rigide, so daß Bindungslängen und Bindungswinkel fixiert werden. Die non-bonded-Wechselwirkungen umfassen neben van-der-Waals-Kräften und elektrostatischen Wechselwirkungen meist noch Terme für Wasserstoffbrücken-Bindungsenergie und Torsionspotentiale, die Veränderungen von

Dihedralwinkel beschreiben. Darüber hinaus werden einige Kraftfelder ergänzt durch entropische Beiträge bei der Fixierung von Freiheitsgraden und Terme, die die Oberflächenenergie beschreiben.

Alle diese Kräfte hängen im wesentlichen von der Distanz der wechselwirkenden Atome ab, so daß sich die sie beschreibenden Gleichungen in den Kraftfeldern sehr ähneln. Die Unterschiede zwischen ihnen ergeben sich überwiegend aus den verschiedenen Sätzen von Parametern, wie Atomradien, -ladungen, -solvatation etc. Von den bekannteren Kraftfeldern sind u. a. AMBER (Cornell et al., 1995), CHARMM (Brooks et al., 1983), VFF (Hagler, 1985) resp. CVFF (Ewig et al., 1999), TRIPOS (Clark et al., 1989) sowie GROMOS96 (Stocker et al., 2000) für Protein- und Peptidmodellierungen geeignet. Ein speziell für die Modellierung von Peptidkonformationen in internen Koordinaten optimiertes Kraftfeld ist ECEPP (*Empirical Conformational Energy Program for Peptides*) (Momany et al., 1975; Nemethy et al., 1983; Nemethy et al., 1992). Viele Programme benutzen aber auch gemischte Parametersets aus verschiedenen Kraftfeldern sowie Erweiterungen, die in den ursprünglichen Feldern nicht vorgesehen sind, z. B. weiche van-der-Waals-Potentiale. Wie in einer Untersuchung gezeigt wurde, sind die verschiedenen Kraftfelder für unterschiedliche Aufgabenstellung geeignet (Nikiforovich, 1994). CHARMM, VFF oder TRIPOS sind sehr gut geeignet für Moleküldynamik-Untersuchungen, da flexible Valenzen es den Molekülen erlauben, Energiebarrieren zwischen lokalen Minima zu überwinden. Für globale Optimierungen und extensives Monte-Carlo-Sampling an Peptiden sind jedoch Felder zu wählen, die mit fixer Bindungsgeometrie und speziellen Peptid-Parametrisierungen arbeiten. In dieser Studie wurde auch gezeigt, daß das ECEPP-Kraftfeld besonders für das Peptiddesign gute Ergebnisse gezeitigt hat.

### ***1. 3. 3 Konformationserzeugung und –sampling***

Die große Zahl von Freiheitsgraden macht es bereits für relativ kleine Peptide nahezu unmöglich, alle möglichen Konformationen in einer systematischen Suche zu erzeugen und zu bewerten, so daß diese Art des Sampling speziellen Problemen, z. B. „build-up“-Prozeduren, wo jeweils nur kleine Bereiche eines Peptids gesampelt werden, vorbehalten bleibt. Für längere Peptide und Proteine wächst die Dimensionalität des Konformationsraumes enorm, was als „konformationelle Explosion“ bekannt ist und eine systematische Suche in diesem vieldimensionalen Raum unmöglich macht (das sog. Levinthal-Paradoxon (Levinthal, 1969)). Ein der systematischen Suche verwandtes Verfahren

ist die Stereoalphabet-Methode, bei der alle möglichen Rückgrat-Konformere erzeugt werden, die sich aus allen Kombinationen der lokalen Minima des Ramachandran-Plots für die jeweilige Aminosäure ergeben (Ralston et al., 1974).

Traditionelle Moleküldynamik (für Reviews siehe (Brooks III, 1995; Elber, 1996; Berne et al., 1997; Schlick et al., 1997) beschreibt die Bewegung des Moleküls in der Zeit. Auf jedes Atom mit bestimmter Position, Masse und Geschwindigkeit werden die Newtonschen Bewegungsgleichungen angewandt; über die Geschwindigkeiten wird die kinetische Energie und somit die Temperatur in das System gebracht. Für eine akkurate Beschreibung des Systems, d. h. für die Beibehaltung eines Gleichgewichtszustandes, sind kleine Zeitschritte (im Femtosekundenbereich) unabdingbar, was aber die Anwendung auf größere Systeme sehr schwierig macht, denn ein vollständiges Sampling der Konformationen wird dann nicht mehr erreicht. Die bislang längste Simulation für ein Polypeptid war eine 1 $\mu$ s-Trajektorie der Villin-Kopfdomäne, die einen metastabilen Zustand erreichte (Duan et al., 1998). Der große Vorteil der Moleküldynamik ist das Konformationssampling in einer expliziten Lösungsumgebung, die durch periodische Randbedingungen einem unendlichen Raum angenähert werden kann. Das macht sie auch für die Untersuchung des Faltungszustandes von Peptiden in der Nähe der nativen Konformation zu einem wertvollen Werkzeug; Untersuchungen zur Faltung von Peptiden aus einem denaturierten Zustand heraus hatten dagegen bisher nur begrenzten Erfolg (Brucolleri et al., 1990; Brooks III, 1995; Demchuk et al., 1997; Prevost et al., 1997). Limitierende Faktoren sind hier sowohl das kleine Zeitintervall als auch die hohe Genauigkeit der Energiefunktion, die notwendig ist, um die korrekte Lösung von Abermillionen falsch-positiver unterscheiden zu können. Einige Erweiterungen der Moleküldynamik mit dem Ziel größerer Zeitintervalle sind in letzten Jahren zur Anwendung gekommen. Die Simulation von Molekülen im internen Koordinatenraum (Torsionsraum) erlaubt im Gegensatz zur Behandlung im Cartesischen Raum eine Unterscheidung zwischen „harten“ Freiheitsgraden, wie Bindungslängen und –winkel, und „weichen“ Freiheitsgraden, wie Torsionswinkeln. Erstmals zu Beginn der 1990er Jahre für beliebig verzweigte Makromoleküle eingeführt (Mazur et al., 1991), erlaubt die Fixierung von harten Freiheitsgraden eine Erhöhung des Zeitintervalls auf 2-4 fs und die Unterdrückung schneller Wasserstoff-Rotation nochmals eine siebenfache Beschleunigung des minimalen Zeitschritts (Mazur, 1997). Eine andere Modifikation mit dem Ziel, einen größeren Konformationsraum zu sampeln, besteht bei der gequenchten Moleküldynamik darin, beim ersten Schritt eine sehr hohe Temperatur (bis 1000K) anzuwenden und die nach bestimmten Zeitintervallen erhaltenen Konformationen zu quenchen, d. h. sie einer lokalen

Energieminimierung zu unterwerfen (O'Connor et al., 1992). Schließlich wird bei einer der jüngsten Erweiterungen der Moleküldynamik die Update-Zeit der „langsamen“ Kräfte auf 48 fs oder mehr erhöht, was eine 10-fache Erhöhung der Effizienz zur Folge haben soll (Barth et al., 1998).

Monte-Carlo-Methoden basieren auf der statistischen Mechanik zur Erzeugung eines Konformationsensembles durch Zufallsbewegungen (für Reviews siehe Nemethy et al., 1990; Karplus et al., 1995; Scheraga, 1996; Abagyan, 1997; Berne et al., 1997). Durch einen Zufallsgenerator wird eine oder mehrere freie Variablen ausgewählt und deren Wert innerhalb eines vorgegebenen Intervalls zufällig verändert. Wie bei der Moleküldynamik ist das unmittelbare Ziel und die Gleichgewichtsbedingung einer traditionellen Monte-Carlo-Simulation die Generierung eines Boltzmann-Ensembles, was sie in der Regel für die Anwendung auf die Polypeptid-Faltung und Docking-Probleme zu langsam und damit ungeeignet macht – allein für das Absuchen der Umgebung eines lokalen Minimums sind mehrere Dutzend Simulationsschritte erforderlich. Es sind aber eine Reihe von Modifikationen entwickelt worden, um einen größeren Konformationsraum abzusuchen und die Simulation zu beschleunigen. Die Anwendung des Metropolis-Kriteriums (Metropolis et al., 1953) ist die am häufigsten benutzte Methode, energietiefe Konformationen zu bevorzugen und die Samplinggeschichte mit in die Simulation einzubeziehen. Einen weiteren Effizienzgewinn bedeutet Monte-Carlo-Sampling erster Ordnung, d. h. mit lokaler Energieminimierung nach jedem Zufallsschritt (Monte-Carlo Minimierung (MCM)) (Li et al., 1987). Diese Art ist einem Sampling nullter Ordnung deutlich überlegen, da eine harmonisches Energietal bereits in einem Schritt gefunden werden kann (strenggenommen sind MC-Methoden erster Ordnung bereits globale Optimierungen (siehe unten), da die lokale Gleichgewichtsbedingung eines kanonischen Ensembles verletzt wird). Als Minimierungstechnik kann auch das „simulated annealing“ (Kirkpatrick et al., 1983) während des Samplings eingesetzt werden und man spricht in dem Fall von Monte-Carlo Simulated Annealing (MCSA); jedoch ist die Konvergenz hier schlechter als bei MCM (Nayeem et al., 1991). Als weitere abgeleitete Methoden sind u. a. noch zu nennen: elektrostatisch getriebenes Monte-Carlo-Sampling (EDMC, (Ripoll et al., 1988)), restriktives Monte-Carlo-Sampling (Holm et al., 1991), hochgerichtetes Monte-Carlo-Sampling (high directional MC (HDMC), (Shin et al., 1991)) und erweitertes MC mit skalierten kollektiven Variablen (extended scaled collective variable MC (ESCV), (Kidera, 1995)). Daneben existieren aber auch hybride Methoden zwischen Moleküldynamik und Monte-Carlo, d. h. nach einem Zufallsschritt folgt ein vorgegebenes Intervall MD-Simulation. Eine weitere Art der Konformationserzeugung,

die derzeit in stetiger Entwicklung ist, ist die Methode des genetischen Algorithmus (GA) (Kawai et al., 1989), auf die aber hier nicht weiter eingegangen werden soll.

### ***1. 3. 3 Stochastische globale Optimierung***

Für die exakte Einschätzung der freien Energie eines Makromoleküls ist es notwendig, die dynamische Natur des gesamten Systems zu betrachten, einschließlich der frei beweglichen Wassermoleküle, der flexiblen Seitengruppen und aller Vibrationsmodi. Das ist im Prinzip durch Moleküldynamik oder klassische Monte-Carlo-Simulation möglich; aber um die native Konformation eines gefalteten Polypeptids oder die gebundene Konformation eines Peptids in einem Komplex (was ja einer extern beeinflussten Faltung entspricht) zu eruieren, müßte man die freie Energie in der lokalen Umgebung jeder von Milliarden möglicher Konformationen mit hoher Genauigkeit bestimmen, d. h. ein Boltzmann-Ensemble von jeder Versuchskonformation bilden. Das ist mit den verfügbaren Computer-Ressourcen zur Zeit nur für sehr kleine Peptide möglich. In den meisten Fällen wird die biologisch aktive Spezies eines Polypeptids aber eine einzigartige Konformation im mittleren Feld des Lösungsmittels sein und nicht ein wirklich dynamisches System bilden, wenn man von einigen sehr flexiblen Seitenketten und Wassermolekülen der Umgebung absieht; anderenfalls wäre eine Spezifität biologischer Wechselwirkungen kaum zu erreichen oder zu erklären. Wenn daher die freie Energie des Lösungsmittels für jede Versuchskonformation durch eine implizite Berechnung der elektrostatischen und Solvatationseffekte angenähert werden kann (Honig et al., 1995) und die Seitenketten-Beweglichkeit durch die jeweilige Solvent-Zugänglichkeit abgeschätzt wird, wird die freie Energie des Systems direkt als Funktion einer einzelnen solvatierten und gefalteten Konformation ausgedrückt, unter der berechtigten Voraussetzung, daß die Schwingungsenergien in verschiedenen gefalteten (oder gebundenen) Zuständen von vergleichbarer Amplitude sind. Die Notwendigkeit der Erzeugung eines kanonischen Ensembles in der Umgebung jeder Versuchskonformation bzw. der Kontinuität der Trajektorie ist somit nicht mehr gegeben, wenn man nur das globale Minimum der Konformationsenergie-Hyperfläche sucht. Die freie Energie kann (als Pseudo-Potential) unter den erwähnten Voraussetzungen als Funktion einer einzelnen Konformation mit hinreichender Genauigkeit bestimmt werden und die native Konformation eines Polypeptids ist demnach das globale Minimum dieser Pseudo-Potentiallandschaft im Raum der freien Torsionswinkel. Das Auffinden dieses Potentialminimums ist das Ziel globaler stochastischer Optimierungsmethoden.

Globale Optimierung an Peptid- und Polypeptidkonformationen ist durch die enorme Dimensionalität des Problems und die geringe Energiedifferenz zwischen dem globalen und den Millionen von lokalen Minima und den kleinen Abstand der Minima im Hyperraum erschwert. Deshalb muß die verwendete Energiefunktion hinreichend genau sein ( $<1\text{kcal/mol}$ ) (auch die Wahrscheinlichkeitsdichte nicht-nativer Zustände wächst exponentiell mit der Energie). Schließlich kann man die Suche nicht nur auf diskrete Rotamer-Bereiche beschränken, da in praktisch allen kompakten Konformationen kleine lokale Rearrangements für eine ideale Packung nötig sind. Im Grunde sind die meisten der oben erwähnten stochastischen Methoden globale Optimierungsprobleme und können anhand der Abhängigkeit von der Simulationsgeschichte weiter klassifiziert werden. Auch MD-Methoden können als Optimierungsprozeduren verstanden werden, sie sind dabei maximal abhängig von der Simulationsvorgeschichte, da sie alle Variablen quasi-kontinuierlich verändern entsprechend der dynamischen Gleichungen. Auch einige MC-Methoden (ESCV, HDMC etc.) gehören in diese Kategorie, da sie mit sehr kleinen Amplituden der Zufallsbewegungen arbeiten. Werden alle freien Variablen bei jedem Schritt zufällig verändert, ist die Simulation unabhängig von der Vorgeschichte. Die meisten gängigen MC-artigen Optimierungsmethoden arbeiten aber mit großen Zufallsbewegungen von einer oder einer Gruppe gekoppelter Variablen. Sie sind damit nur von der unmittelbaren Simulationsgeschichte abhängig. MC-Methoden mit Metropolis-Kriterium sowie MCM gehören in der Regel in diese Gruppe, wie auch MC-basierte Methoden mit Vorinformation (bias-MC (BMC) (Kang et al., 1993; Lee et al., 1996); optimal-bias MC-Minimierung (OBMCM) (Abagyan et al., 1994a)) sowie der genetische Algorithmus.

Die Berechnung der Energie erfolgt bei Protokollen erster Ordnung nach einer lokalen Energieminimierung. Das ist im Vergleich zur Erzeugung eines kanonischen Ensembles der kürzere Weg, ein lokales Minimum in der Energiefläche zu identifizieren, da analytische Ableitungen der Energiefunktion genutzt werden können und außerdem die Energielandschaften bei Proteinfaltungen oft zerklüftet sind. Nach der Einführung von MCM-Methoden 1987 wurde gezeigt, daß eine volle Energieminimierung nach jedem Zufallsschritt deutlich überlegen ist gegenüber keiner oder nur partiellen Minimierung (Abagyan et al., 1992).

Stochastische Optimierungsmethoden lassen sich durch Einfließen von statistischen Vorinformationen erheblich beschleunigen, indem sie die Samplinghäufigkeit gekoppelter Variablen beeinflussen. Das können sowohl a priori Informationen sein, die sequenzunabhängig durch statistische Auswertung von bekannten dreidimensionalen

Strukturen ermittelt wurden (Abagyan et al., 1994a), oder Informationen, die während der Simulation gesammelt werden (Holm et al., 1991). Es ist bekannt, daß Gruppen von gekoppelten Torsionswinkeln (wie  $\phi$ - $\psi$ -Winkelpaare und Seitenketten-Winkel) bestimmte Wertebereiche mit höherer Präferenz annehmen als andere. Diese sog. Rotamer-Bibliotheken sind gut untersucht und beschrieben (Ponder et al., 1987; Dunbrack et al., 1993; Abagyan et al., 1994a) und Verteilungshäufigkeiten der einzelnen Zustände sind für jede Aminosäure als kontinuierliche und diskontinuierliche Funktionen bekannt. Ein Algorithmus, der ohne Vorinformation arbeitet, sucht alle Bereiche im Winkelraum mit gleicher Wahrscheinlichkeit auf. Werden dagegen die Verteilungshäufigkeiten verwendet, um die Samplinghäufigkeit der einzelnen Zustände zu beeinflussen, spricht man von *biased sampling*. Bei einem linearen Bias sind Samplingwahrscheinlichkeit direkt proportional zur Verteilungsfunktion der Rotamerzustände, es werden also hochwahrscheinliche Zustände sehr häufig gesampelt und geringwahrscheinliche sehr selten. Wie gezeigt werden konnte (Li et al., 1987), maximiert diese Art des Bias die Wahrscheinlichkeit einer korrekten Lösung. Die Anzahl der stochastischen Schritte bei diesem linearen Bias ist aber nicht optimal, da in nahezu jedem Peptid und Protein ungewöhnliche Winkelpaare auftreten, die hier nur selten aufgesucht werden. Globale Wechselwirkungen, von gleicher Größenordnung wie lokale Kontakte, aber zufällig verteilt, verzerren lokale Interaktionen und können Winkelwerte erzeugen, die a priori wenig bevorzugt sind. Diese sind auch durch lokale Minimierung nicht immer zu identifizieren. Diese Beobachtung und die Annahme der Unabhängigkeit der einzelnen Zufallsschritte voneinander (der globale Kontext einer Gruppe von Variablen ändert sich ständig und so kann ein Winkelpaar in einer richtigen und in einer falschen Lösung den gleichen Wert haben), führte zur Ableitung des Quadratwurzel-Bias (Abagyan et al., 1999). Hierbei ist die Samplinghäufigkeit der einzelnen Zustände proportional der Quadratwurzel der kontinuierlichen Verteilungsfunktion der Rotamerzustände. Seltene Zustände werden also häufiger besucht als bei einem linearen Bias, aber die Positionen der Extrema sind unverändert. Diese Art des Samplings minimiert die Anzahl der Schritte bis zur Identifizierung der richtigen Lösung.

#### **1. 3. 4 Docking-Methoden**

Docking-Methoden beschäftigen sich mit der Vorhersage der gebundenen Konformation von kleinen organischen Molekülen (Kuntz et al., 1994; Verlinde et al., 1994), Peptiden, Proteinen oder DNA/RNA (Cherfils et al., 1993; Shoichet et al., 1996; Sternberg et al., 1998; Lengauer

et al., 1999) im Komplex mit Proteinen. Bei allen Algorithmen ist zunächst die Kenntnis der exakten dreidimensionalen Gestalt des Rezeptor-Proteins die Voraussetzung. Diese sind entweder aus Röntgenkristall- bzw. NMR-Daten aus der Protein Daten Bank (Berman et al., 2000) zugänglich oder werden durch Homologie-Modelling sequenzverwandter Proteine gewonnen. Die entsprechenden Zielproteine müssen dann dem zu dockenden Molekül präsentiert werden: als Gitter-Potential (vorgeschlagen von Goodford, 1985), als molekulare bzw. wasserzugängliche Oberfläche (z. B. die Connolly-Oberfläche (Connolly, 1983) oder als „Negativbild“ der Oberfläche (vorgeschlagen von (Kuntz et al., 1982)). Der eigentliche Docking-Schritt besteht in der Erzeugung vieler neuer Konformationen des zu dockenden Objektes in Relation zum Rezeptor. Das geschieht durch einfache Positionierung (z. B. in den Programmen GRID (Goodford, 1985) und MCSS/HOOK (Miranker et al., 1991)), durch Oberflächen-Deskriptoren wie „kritische Punkte“ oder Normalmoden (z. B. in LUDI (Böhm, 1992), DOCK (Meng et al., 1992), CLIX (Lawrence et al., 1992)), durch Monte-Carlo-Methoden (in AutoDOCK (Goodsell et al., 1990), BOXSEARCH (Hart et al., 1992), FLEXX (Rarey et al., 1996), ICM (Abagyan et al., 1994b)), durch genetische Algorithmen (in GOLD (Jones et al., 1997)) oder durch Moleküldynamik (Di Nola et al., 1994). Die große Zahl generierter Konformationen wird während oder nach dem eigentlichen Docking evaluiert und sortiert. Eine exakte Berechnung der freien Bindungsenergie wäre das ideale Kriterium, ist aber extrem aufwändig und gegenwärtig auch noch zu ungenau. Viele der gebräuchlichen Docking-Methoden rangieren die gedockten Konformationen nach Oberflächen-Komplementarität zu dem Rezeptor-Protein zusätzlich zu energetischen Kriterien. In letzter Zeit wurden zweistufige Algorithmen eingeführt, bei denen z. B. die Oberflächenkomplementarität erst in einem genetischen Algorithmus und die energetische Kompatibilität anschließend in einer Tabu-Suche ermittelt wird (Hou et al., 1999) oder das zu dockende Molekül zuerst in einem approximierten Feld einer extensiven Konformations-Suche unterzogen und anschließend in einem klassischen Kraftfeld verfeinert wird (Hoffmann et al., 1999). Auch virtuelles Screening vieler zehntausend potentieller Liganden an einem Rezeptor ist so erfolgreich durchgeführt worden (Schapira et al., 2001). Viele Sortier-Schemata sind inzwischen entwickelt worden, die entweder auf Molekularmechanik, empirischen Daten oder einer Kombination beider Methoden beruhen und darauf gerichtet sind, die Bindungsenergie so genau wie möglich emulieren (z. B. (Åqvist et al., 1994; Böhm, 1994; Schapira et al., 1999)). Bisher ist jedoch noch keine dieser Algorithmen zu allgemeiner Anerkennung gelangt, da sie meist einer Parametrisierung innerhalb einer Klasse von Molekülen bedürfen, die nicht unmittelbar auf eine andere Klasse übertragbar ist.

Im allgemeinen kann man Docking-Methoden unterscheiden nach der Art, ob der zu dockende Teil und/oder das Rezeptor-Molekül flexibel oder rigide behandelt werden. *Rigid body*-Docking wurde mit Erfolg bei Protein-Protein-Komplexen angewandt (Lengauer et al., 1999), allerdings sinkt die Genauigkeit, wenn Strukturen der unkomplexierten Partner aneinander gedockt werden. Ursachen hierfür sind einerseits mögliche leichte Gesamtkonformationsveränderungen (*induced fit*) durch die Bindung und andererseits flexible Seitenketten an der Proteinoberfläche, deren Konformation im gebundenen und ungebundenen Zustand nicht identisch sein muß. Durch flexible Behandlung von solchen Seitenketten kann man auch aus apo-Proteinen sehr genaue Komplexstrukturen erhalten (Totrov et al., 1994). Eine vollständige Flexibilisierung auch des Rückgrats ist für Proteine zur Zeit nicht beherrschbar und bereits für relativ kleine Peptide problematisch im Sinne einer daraus erwachsenen „Konformationsexplosion“. Erfolgreiche Docking-Experimente ohne weitere Zwangsbedingungen sind daher überwiegend auf kurze Peptide bis höchstens 8 oder 9 Residuen beschränkt (Caflisch et al., 1992; Rosenfeld et al., 1993; Desmet et al., 1997; Stigler et al., 1999), während für längere oft Orientierungs- oder Distanzlimits bzw. eine eingeschränkte Flexibilität nötig sind (Nakajima et al., 1997; Zhou et al., 1998). Auch ein 26-meres Peptid wurde bereits erfolgreich, allerdings auch hier unter einem vereinfachten Flexibilitätsmodell (Sandak et al., 1998). Der Erfolg des Dockings flexibler Peptide hängt allerdings stark vom Rezeptormodell ab: Für das Docking an Antikörper-Strukturen wurde gefunden (Goodsell et al., 1996), daß Peptide länger als drei Residuen nur korrekt an die Komplexstruktur des F<sub>ab</sub>-Fragmentes gedockt werden konnten, während dies für die native Form des Antikörpers nicht gelang.