

2. Material und Methoden

2. 1 Stochastische globale Optimierung und Modellierung im internen Koordinatenraum

2. 1. 1 Interne Koordinaten

Für sämtliche Modellierungsexperimente wurde das Software-Paket ICM (Internal Coordinate Mechanics, Molsoft LLC, La Jolla; Abagyan et al., 1994a; Abagyan et al., 1994b) verwendet. Dieses behandelt die Molekülgeometrie im Raum der internen Koordinaten, d. h. im Gegensatz zum klassischen kartesischen Koordinatenraum wird die Position aller Atome im Molekül nur durch einen Satz von Winkeln und Bindungslängen beschrieben, die in Verbindung mit einer hierarchischen Nummerierung eine eindeutige Festlegung sowohl der Molekülkonformation als auch der Lage des gesamten Moleküls im Raum erlaubt. Die Unterteilung der Freiheitsgrade in „harte“ (Bindungslängen, Bindungswinkel) und „weiche“ (Torsionswinkel, Planarwinkel) ermöglicht hierbei eine enorme Beschleunigung der

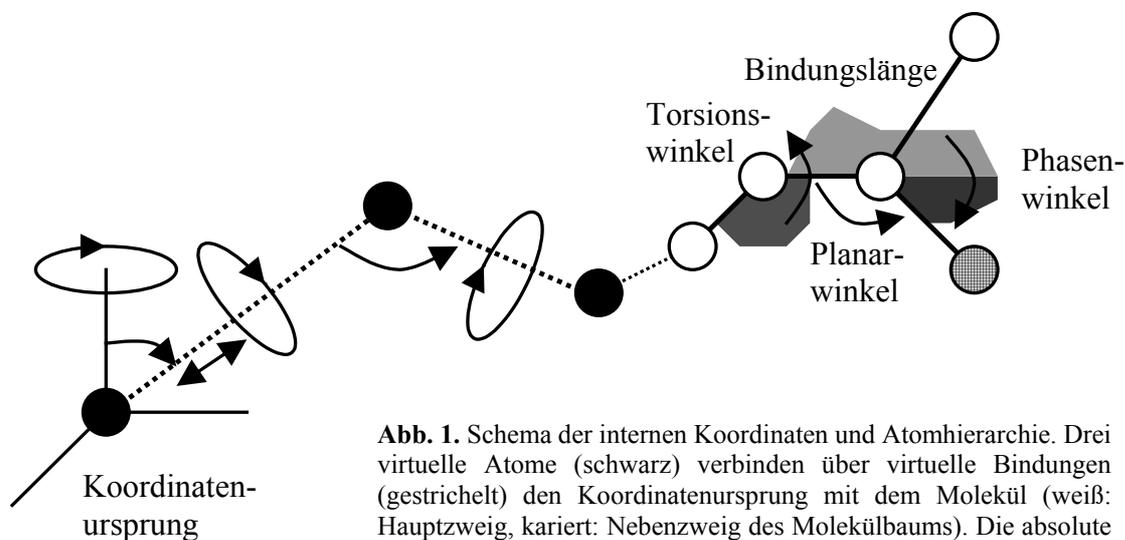


Abb. 1. Schema der internen Koordinaten und Atomhierarchie. Drei virtuelle Atome (schwarz) verbinden über virtuelle Bindungen (gestrichelt) den Koordinatenursprung mit dem Molekül (weiß: Hauptzweig, kariert: Nebenzweig des Molekülbaums). Die absolute Lage des Moleküls wird durch sechs mit den virtuellen Atomen verbundenen Variablen definiert.

Simulationen, wobei harte Freiheitsgrade fixiert und schnelle Wasserstoff-Rotationen unberücksichtigt bleiben. Diese Darstellungsweise hat sowohl bei stochastischen Modellierungsmethoden als auch bei der klassischen Newtonschen Moleküldynamik Verbreitung gefunden, da sie die Sampling-Effizienz erhöht bzw. eine Verlängerung des minimalen Zeitintervalls erlaubt. Eine konformationelle Änderung kann so durch die

Änderung nur eines Winkels beschrieben werden, während im kartesischen Raum die Neuberechnung vieler Atomkoordinaten notwendig ist.

Abb. 1 illustriert das Schema der internen Koordinaten und der Atomhierarchie. Drei virtuelle Atome verbinden das eigentliche Molekül mit dem Koordinatenursprung, die Veränderung ihrer „Bindungs“längen und Torsionswinkel erlaubt die vollständige Beschreibung der räumlichen Bewegung des gesamten Moleküls bzw. der relativen Bewegung mehrerer Moleküle zueinander. Jedes Atom ist durch die Positionen der Vorgängeratome und drei Parameter (Bindungslänge, Planarwinkel, Dihedralwinkel) eindeutig bestimmt. Alle Atome werden entsprechend ihrer chemischen Eigenschaften einem Atomtyp zugeordnet und zusammen mit den zu ihnen gehörenden Winkeln und Bindungslängen in einer Liste zusammengestellt, die den Molekül-„Baum“ bildet. Dieser wiederum verweist auf weitere Listen, in denen die einzelnen Wechselwirkungspotentiale und –parameter abgelegt sind (elektrostatische Wechselwirkung, H-Brückenparameter, Torsionspotentiale usw., siehe Abschnitt 2).

2. 1. 2 *Optimal-Bias Monte-Carlo Minimierung*

Konventionelle Monte-Carlo Methoden benötigen als Gleichgewichtsbedingung die Erzeugung eines Boltzmann-Ensembles, d. h. für die Konvergenz um ein lokales Energieminimum sind bereits mehrere Berechnungen nötig. Ist das Ziel der Simulation aber die Identifizierung nur der möglichen energetisch tiefsten Konformationen eines freien oder gebundenen Peptids, also in der Auffindung des globalen Minimums der vieldimensionalen Energielandschaft, ist es hinreichend, die lokalen Minima durch die Anwendung effizienter Energieminimierungsschemata zu orten. In der Regel wurde hierzu das Verfahren des konjugierten Gradienten benutzt. Diese stochastische globale Optimierung erster Ordnung wird weiterhin beschleunigt durch Anwendung experimentell ermittelter Präferenzen (Bias, (Abagyan et al., 1994a)). Für jede Aminosäure gibt es dabei eine kontinuierliche Wahrscheinlichkeitsfunktion, die die Verteilung der entsprechenden Rotamer-Zustände für korrelierte Seitenketten-Torsionswinkel und Rückgrat-Winkelpaare bei nativen Proteinen beschreibt. Die Verteilungsfunktionen wurden aus einem Datensatz von repräsentativen dreidimensionalen Strukturen abgeleitet (für ϕ - ψ -Paare aus 191 Domänen mit Sequenzähnlichkeit $<30\%$ und Auflösung $<2,4 \text{ \AA}$; für korrelierte χ -Winkel aus 166 Domänen mit $<50\%$ Sequenzähnlichkeit und Auflösung $<2,0 \text{ \AA}$). Untersuchungen zur optimalen stochastischen Ratestrategie, die darauf gerichtet ist, die durchschnittliche Anzahl der

Rateversuche zu minimieren, haben dabei gezeigt, daß die relativen Frequenzen der Rateschritte proportional zur Quadratwurzel der ermittelten Wahrscheinlichkeitsverteilungen sind. Diese Implimentierung des Bias wird deshalb auch Optimal-Bias genannt und stützt sich auf der Annahme, daß die lokale Energielandschaft, die durch sequentielle, lokale Wechselwirkungen entsteht, durch nicht-lokale Interaktion von ungefähr derselben Größenordnung verzerrt wird. Nach jedem Zufallsschritt ist aber diese nicht-lokale Umgebung durch andere Wechselwirkungspartner verändert und daher ist jeder Rateschritt unabhängig von seinem Vorgänger. Für eine vollständige Beweisführung des Optimal-Bias siehe (Abagyan et al., 1999; Zhou et al., 1999).

2. 1. 3 Klassifizierung der Zufallsschritte und Aufzeichnung der Simulation

Je nach Aufgabenstellung kann von dem Programm einer von drei Typen von Zufallsschritten angesprochen werden:

- Zonaler Schritt: Singuläre Winkelveränderung oder gebiaster kollektiver Schritt (bei gekoppelten Winkelpaaren)
- Translation/Rotation des gesamten Moleküls (pseudo-Brownsche Bewegung)
- Loop-Schritt: Veränderung symmetrisch abhängiger Winkel (bei Loop-Vorhersage/ β -Reißverschluß-Schritt)

Diese Schritte können *statisch* oder *dynamisch* sein, wobei bei letzteren die Sekundärstruktur mit in Betracht gezogen wird (z. B. bei β -Reißverschluß-Schritten, bei denen ein enger Turn und seine H-Brücken erkannt wird und mit einer gewissen Wahrscheinlichkeit symmetrische Schritte zur Schließung der β -Struktur versucht werden; (Abagyan et al., 1999)). Letzteres Beispiel gehört auch zur Gruppe der *verteilten* Zufallsschritte, da im Gegensatz zu *lokalen* Schritten mehrere Gruppen von Variablen verändert werden. Gebiaste Monte-Carlo-Schritte sind in der Regel *korrelierte* Schritte, hier werden die zu verändernden Variablen konzertiert geändert. Darüber hinaus können die Zufallsschritte *kontinuierlich* oder *diskret* sein. Die größte Bedeutung kommt in der Regel kontinuierlichen, korrelierten Schritten, wie bei den zonalen und Loop-Schritten, zu, da sie die Konvergenz für Peptidsimulationen stark beschleunigen.

Während der Simulation besteht die Möglichkeit, die Trajektorie aufzunehmen. Dafür werden für jeden nach der lokalen Minimierung akzeptierten Zufallsschritt die Koordinaten sowie die Energiewerte in eine binäre Datei ausgeschrieben. Zur Rückkopplung mit bereits untersuchten Konformationen während der Simulation wird ein sog. konformationeller Stapel angelegt, der

energiehierarchisch definiert ist (Abagyan et al., 1992). Hierfür wird zunächst ein Ähnlichkeitskriterium bestimmt, das vorgibt, innerhalb welcher Grenzen und für welche Variablen eine Konformation strukturell einem Cluster zugeordnet wird. Der Distanzparameter kann dabei im Torsionsraum oder im Cartesischen Koordinatensystem festgelegt werden. Typischerweise wird als Ähnlichkeitskriterium der anguläre RMSD

$$d_{AB} = \sqrt{\frac{1}{n_{V_c}} \sum_{i=1}^{n_{V_c}} \min\left((\theta_i^A - \theta_i^B)^2, (2\pi - \theta_i^A + \theta_i^B)^2\right)} \quad (1)$$

verwandt, wobei n_{V_c} die Anzahl der für den Strukturvergleich betrachteten Winkel θ der Konformationen A und B bezeichnet. In der Regel werden dies die ϕ - und ψ -Winkel des Rückgrates sein. Wird nun eine Konformation während der Simulation akzeptiert, wird diese mit allen anderen bereits im Stapel vorhandenen Konformationen sowohl in energetischer als auch in struktureller Hinsicht verglichen und entsprechend eingeordnet. Ist die betreffende Konformation nicht ähnlich und energetisch höher als die anderen Konformationen, besetzt sie, so noch vorhanden, eine freie Stapelposition. Trifft dagegen das Ähnlichkeitskriterium gegenüber einer vorhandenen Konformation zu, muß die neue eine tiefere Energie besitzen, um die alte zu ersetzen, andernfalls wird ein Zähler inkrementiert, der angibt, wie oft dieses Cluster besucht wurde. Ist eine neugefundene Konformation eines Clusters energietiefer als eine andere Konformation in einem anderen, bisher energiegünstigeren Cluster, wird die Hierarchie neu geordnet. Der Konformationsstapel ist also während der Simulation einer quasi-biologischen Evolution unterworfen. Er kann einerseits zur Analyse und Visualisierung energetisch günstiger Konformationsfamilien verwandt werden, die bei entsprechender Parameterwahl eine hinreichende Diversität erkennen lassen. Andererseits dient er zur Steuerung der Simulation bei folgenden problematischen Situationen und damit zur Vermeidung unnötigen Samplings:

(1) Zu viele sukzessive erfolglose Versuche. In unergiebigem Bereichen der Energielandschaft kann es vorkommen, daß es zu viele Zufallsschritte gibt, die nicht akzeptiert werden bei einer bestimmten Temperatur. Diese Versuche werden durch einen Zähler aufgezeichnet und nachdem dieser einen vorher festgelegten Wert überschritten hat, wird die Blockierung der Simulation durch einen von drei möglichen Ausweichmechanismen aufgelöst (Temperaturverdoppelung, Konformationsausbildung innerhalb des am wenigsten besuchten Clusters, Randomisierung der freien Variablen). Für die beschriebene Situation hat sich die Verdoppelung der Temperatur als optimal erwiesen; diese wird zurückgesetzt auf den vorgegebenen Wert, sollte eine neue Konformation gefunden werden. Ebenso wird der Zähler zurückgesetzt, so daß nur die jüngere Samplinggeschichte bei diesem Szenario von Interesse

ist.

(2) Zu viele Besuche derselben Konformation. Es werden neue Konformationen gefunden, aber diese ähneln einer vorhandenen und sind energetisch höher. Diese Szenario tritt auf, wenn die Simulation in einem engen Bereich des Konformationsraumes steckengeblieben ist. Zur Detektion dieses Effektes wird ein Zähler inkrementiert, der die erfolglosen Besuche für jedes Konformationscluster aufzeichnet. Überschreitet dieser einen vorgegebenen Wert, greift einer der drei oben erwähnten Ausweichmechanismen ein, ersetzt dagegen eine neue Konformation die in einem Cluster vorhandene, wird er auf 0 zurückgesetzt. Für die Peptidfaltung hat sich die Randomisierung aller Torsionswinkel mit einer Maximalamplitude von 30° bewährt, beim Peptiddocking dagegen die Verdoppelung der Temperatur.

(3) Keine Änderung des Konformationsstapels. Bewegt sich die Simulation in einem Hochenergie-Bereich, tritt keine Änderung des Konformationsstapels auf. Weder ersetzt eine neue Konformationsfamilie eine bereits vorhandene, noch wird eine neue Stapelposition angelegt (das geht natürlich nur bis zur Erreichung der Maximalzahl von Stapelpositionen), noch wird ein Besuch eines bereits vorhandenen Clusters registriert. Dem Stapel sollten aber nicht unbegrenzt neue Positionen zugefügt werden können, da ja alle neuen Konformationen gegen die vorhandenen Positionen getestet werden müssen. Deshalb werden alle Versuche gezählt, die in keiner Veränderung des Stapels münden und nach Erreichen der vorgegebenen Maximalzahl (in der Regel 50) wird die am wenigsten besuchte Konformation aus dem Stapel angenommen, von wo aus die Simulation fortfährt.

2. 1. 4 Energiepotentiale und Helicitätsunterdrückung

Eine exakte Einschätzung der freien Energie für jede Versuchskonformation ist notwendig, um die native Konformation (die die energietiefste sein sollte) aus einer enorm großen Anzahl von möglichen Konformationen herauszufiltern. Die Grundlage der Energiepotentiale bildet in dem Programm ICM das ECEPP/3-Kraftfeld (Nemethy et al., 1992), das für alle Atomtypen *in vacuo* definiert ist, ergänzt durch Terme für die Solvatationsenergie und den entropischen Beitrag, der durch das Einfrieren der Freiheitsgrade der Seitenketten entsteht. Die einzelnen Terme sind durch folgende Gleichungen gegeben:

- Van-der-Waals-Energie
$$E_{vw} = \sum_{i,j} \left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{CB_{ij}}{d_{ij}^6} \right) \quad (2)$$

- Wasserstoff-Brückenenergie
$$E_{hbond} = \sum_{i,j} \left(\frac{A'_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}} \right) \quad (3)$$

(i,j Indices von Atomen, die durch mehr als zwei kovalente Bindungen und weniger als 7,5 Å (bzw. 3 Å bei H-Brücken) voneinander getrennt sind; d_{ij} interatomarer Abstand; A, B, A', D Parameter für die zwei Atomtypen; C=0,5 für durch drei Bindungen getrenntes Atompaar, C=1 für alle anderen Fälle)

- Torsionsenergie
$$E_{torsion} = \sum_{torsions} U_m \cos(n\phi - \phi_0) \quad (4)$$

(U_m individuelle Torsionsbarrieren)

- Elektrostatische Energie
$$E_{electr} = \sum_{i,j} \frac{332q_i q_j}{\epsilon d_{ij}} \quad (5)$$

(q elektrische Ladungen; ϵ Distanz-abhängige Dielektrizitätskonstante [$\epsilon=4 \cdot d_{ij}$])

- Solvatationsenergie
$$E_{solv} = \sum_i \sigma_i a_i \quad (6)$$

(a_i individuelle Solvent-zugängliche Oberfläche des i-ten Atoms; σ_i individuelle Solvatationsenergie-Dichte aus Wasser-Vakuum-Transferenergien)

- Entropischer Beitrag
$$E_{entropy} = -RT \sum_{residue\rho} \frac{\Delta S_{\rho}^{\max} a_{\rho}}{a_{\rho}^{\max}} \quad (7)$$

(S^{\max} maximale Entropie beim Begraben der Seitenkette einer individuellen Aminosäure; a/a^{\max} relative Solvent-zugängliche Oberfläche, bezogen auf die maximale Oberfläche in einem gestreckten Dreier-Peptid Gly-Xxx-Gly).

- Distanz-Restraint

$$E_p = \frac{0.25W_p(d^2 - D_u^2)^2}{D_u^2}; d > D_u \quad (8)$$

$$E_p = \frac{0.25W_p(d^2 - D_L^2)^2}{D_L^2}; d < D_L$$

(0 falls d zwischen D_L und D_U ; wobei d die aktuelle Distanz, D_L und D_U unteres bzw. oberes Limit, W_p das Gewicht ist)

- Tether-Restraint („Gummiband“)
$$E_T = W_T d^2 \quad (9)$$

(mit d als Abstand zwischen dem freien Atom und dem fixen Zielatom, W_T als Gewicht)

- Multidimensionaler Variablen-Restraint (Ellipsoidalzonen)

$$E_v = \frac{U_v (\delta_v^2 - 1)(2\delta_v^2 - 3F_v^2 + 1)}{(1 - F_v^2)^3}; F_v^2 < \delta_v^2 < 1$$

$$E_v = U_v; \delta_v^2 \leq F_v^2 \quad (10)$$

$$E_v = 0; \delta_v^2 \geq 1$$

$$\text{mit } \delta_v^2(\theta_1^v, \dots, \theta_n^v) = \sum_{i=1}^n \frac{(\theta_i^v - \theta_{0,i}^v)^2}{(\Delta_i^v)^2}$$

(δ ist die normalisierte Distanz von n Variablen θ von ihrem jeweiligen Wert im Zentrum der Energiemulde; Δ ist der Abstand vom Zentrum, wo die Energiefunktion 0 wird; F_v ist der relative Anteil der Muldendimension Δ_i mit flachem Boden, U_v ist die Tiefe der Energiemulde)

Die Solvent-zugängliche Oberfläche wird mit einem schnellen Shrake-Rupley-Algorithmus ermittelt (Shrake et al., 1973), der auf der Zählung der nicht-verborgenen Oberflächenpunkte basiert, durch die jedes Schweratom repräsentiert wird. Der entropische Beitrag besitzt keine analytische Ableitung und wird daher nicht bei der Minimierung berücksichtigt. Die Form der Funktion für multidimensionale Zonen wird sowohl für das Sampling mit optimalen Bias (als Wahrscheinlichkeitsfunktion) als auch als eigentliche Energiefunktion (als Straf- oder Penaltyfunktion) genutzt. Die letzten drei Energiefunktionen sind zusätzliche Terme, die benötigt werden z. B. zur Emulation von Disulfid- und Lactambindungen, zur Regularisierung von Kristallstrukturen und zum Docking. Die Gesamtenergie ergibt sich aus der Aufsummierung aller Energieterme. Die Ableitungen der einzelnen Terme sind hier nicht aufgeführt, es wird auf die Literatur verwiesen (Abagyan et al., 1994b).

Es wurde gezeigt, daß das ECEPP/3-Feld die Herausbildung von α -Helices begünstigt. Zur Kompensation dieses Effektes liegt zusätzlich auf den ψ -Winkeln ein weiches Torsionspotential von der Form: $1 + \cos(90 + \psi)$, wobei dieses Potential bei den β -verzweigten Aminosäuren Valin, Isoleucin und Threonin eine Amplitude von 1 kcal/mol hat und bei allen anderen Residuen 0,5 kcal/mol beträgt.

2. 1. 5 Parametersätze

Die Atomtypen sind definiert durch den van-der-Waals Typ, Wasserstoffbindungs-Typ, Hydratationstyp, das Atomgewicht und die Oberfläche. Van-der-Waals Parameter sind

entnommen aus dem ECEPP/3 Kraftfeld (Nemethy et al., 1992) und sind bestimmt durch Polarisierbarkeit, effektive Elektronenanzahl, Potentialtiefe und Gleichgewichtsabstand für die Distanz gleicher Atome sowie den Radius. Ebenfalls aus ECEPP sind die Wasserstoffbindungsparameter übernommen (Nemethy et al., 1983). Die Hydratationstypen sind gegeben durch experimentell bestimmte Solvatationsenergie-Dichten für den Transfer von Vakuum-Wasser bzw. Wasser-Oktanol (Wesson et al., 1992). Torsionsparameter sind größtenteils übernommen aus (Momany et al., 1975). Andere Parametersätze bestimmen die Bindungswinkel- und Improperwinkel-Deformationstypen und Bindungsstreckung; da diese für gewöhnlich aber fixiert sind, wird auf sie nicht eingegangen. Die Parameter für die multidimensionalen Zonen gekoppelter Variablen, die das Sampling mit dem optimalen Bias steuern (Abagyan et al., 1994a) bestimmen die Muldentiefe, den flachen Anteil, die relative Wahrscheinlichkeit in Bezug auf andere Zonen derselben Variablen und die Variablenwerte für das Muldenzentrum. Aus den Wahrscheinlichkeiten wurden auch die Entropiebeiträge der Seitenkettenvergrabung für 300 K ermittelt (Abagyan et al., 1994a). Diese Werte finden sich wieder in der Aminosäure-Bibliothek, die für jeden Residuentypus den Molekülbaum vorgibt und für jedes Atom auf die entsprechenden Parameter verweist. Neben der Ladung bzw. Formalladung jedes Atoms und den Werten für Bindungslänge und -winkel wird außerdem die Definition der internen Variablen darin gegeben sowie deren Fixierungsstatus. Zudem werden dort Gruppierungen polarer Atome wie direkt gebundener N-H und C-O Paare definiert, damit auch bei Überschreitung der Cutoff-Distanz für einzelne Partner die Gruppe als elektrostatisches Ensemble behandelt wird. Folgende Distanzrestraint-Werte für Disulfid- bzw. Lactambindungen wurden verwendet: $S\gamma$ - $S\gamma$ Abstand 2.04 Å/10 kcal/mol Gewicht, $S\gamma$ - $C\beta$ 3.052 Å/5 kcal/mol, $C\beta$ - $C\beta$ 3.855 Å/3 kcal/mol; Lactambindungen: C^o -N 1.325Å/80 kcal/mol, C^o -NH 2.085 Å/65 kcal/mol, O-N 2.261/65 kcal/mol, O-NH 2.171 Å/65 kcal/mol, C^{o-1} -N 2.41 Å/65 kcal/mol, C^o - C^{n-1} 2.42 Å/65 kcal/mol, C^{o-1} - C^{n-1} 3.78 Å/65 kcal/mol; Indices o and o-1 bezeichnen das Carbonyl-Kohlenstoffatom und dessen vorhergehendes Kohlenstoffatom im sauren Lactambindungspartner, Index n-1 das dem Stickstoffatom vorausgehende Kohlenstoffatom im basischen Partner.

2. 1. 6 Allgemeiner Ablauf einer Simulation

Das zu simulierende Molekül wird entweder aus Aminosäure-Monomerbausteinen aus der Residuenbibliothek aufgebaut oder durch Regularisierung aus einem kartesischen Koordinatensatz gewonnen (siehe unten). Eine Simulation startet jeweils mit einer

Zufallskonformation, d. h. jeder freien Torsionsvariable wird zufällig ein Wert zwischen 180° und -180° zugeordnet. Aus der Gesamtheit der freien Variablen werden zwei Subsets ausgewählt, die einerseits für die Zufallsbewegungen in Frage kommen und andererseits während der lokalen Minimierung nach jedem Zufallsschritt berücksichtigt werden. Die globale Optimierung läuft dann nach folgendem Schema ab:

- Eine Variable oder eine Gruppe gekoppelter Variablen wird zufällig ausgewählt und innerhalb einer gesetzten Grenze zufällig (für singuläre Variablen, Bereich in der Regel von -180° bis 180°) oder zwischen Rotamerzustands-Zonen (für den optimalen Bias bei gekoppelten Variablen) entsprechend ihrer Verteilungswahrscheinlichkeiten verändert.
- Die freien Variablen des Moleküls werden einer vollständigen Energieminimierung mit allen Termen, die eine analytische Ableitung besitzen, unterzogen. Die Minimierung endet nach einer vorher festgesetzten Anzahl von Schritten bzw. vorher bei Erreichen eines gewissen Toleranzgradienten. Für das minimierte Molekül wird die Solvatationsenergie und die Seitenkettenentropie berechnet und zur minimierten Energie addiert.
- Selektion der Konformation nach dem Metropolis-Kriterium (Metropolis et al., 1953):

$$\Delta E = E_{c+1} - E_c \leq 0$$

bzw.

$$\rho < e^{-\frac{\Delta E}{RT}}; \rho \in (0,1) \quad (11)$$

d. h. die Konformation wird akzeptiert, wenn ihre Energie kleiner ist als die der vorherigen oder wenn ihre temperaturabhängige Boltzmann-Wahrscheinlichkeit größer ist als als eine Zufallszahl zwischen 0 und 1. Die Temperatur kann dabei erhöht werden, wenn eine problematische Situation (siehe oben) eintritt; die Wahrscheinlichkeit der Akzeptanz einer Hochenergie-Struktur steigt dadurch.

Falls die Konformation angenommen wurde, wird ihre Ähnlichkeit zu den im Konformationsstapel befindlichen Strukturen berechnet (siehe Gl. 1) und entsprechend dieses Wertes und der Energie eingeordnet. Hier geschieht eine mögliche Rückkopplung mit der Simulationgeschichte: der Zähler eines Konformationsclusters wird erhöht, ein neuer Slot belegt, die repräsentative Konformation in einem Cluster ersetzt oder die Hierarchie neu geordnet. Die neue Konformation wird auch als Ausgang für einen neuen Monte-Carlo-Schritt genommen. Hat die Konformation das Metropolitkriterium nicht erfüllt, wird der Zähler der abgelehnten Strukturen erhöht, ohne den Stapel zu verändern, und die zuletzt angenommene Konformation wiederhergestellt. Danach beginnt die Abfolge der Schritte erneut, bis ein Terminationskriterium erfüllt ist (in der Regel die Maximalzahl an Funktionsaufrufen, d. h. Monte-Carlo-Schritten plus Minimierungsschritten).

2. 1. 6. 1 *Faltungsexperimente*

Bei den Faltungsexperimenten wurden die Peptide aus den Monomer-Bausteinen der Residuenbibliothek aufgebaut und ungeladene Termini (acetylierter N-Terminus und amidierter C-Terminus) zugefügt. Alle Standard-Variablen, also ϕ -, ψ - und ω -Dihedralwinkel des Rückgrates sowie χ -Winkel der Seitenketten wurden sowohl für die Monte-Carlo-Bewegungen (mit Ausnahme der Torsionswinkel der an polare Seitenkettenatome gebundenen Protonen) als auch für die Minimierungen freigegeben, hingegen wurden die Variablen der virtuellen Atome, die die relative Lage des Peptids bestimmen, fixiert, da sie unerheblich für ein monomeres Molekül sind. Allen gekoppelten Variablen wurden die entsprechenden Rotamer-Präferenzen zugeordnet und die freien Winkel mit Ausnahme von ω randomisiert. Als ähnlich wurden Strukturen definiert, deren ϕ - und ψ -Winkel einen angulären RMSD von kleiner als 30° aufwiesen. Der Toleranzgradient für die Minimierung betrug 0,01 und die Temperatur 600 K. Die maximale Anzahl der Monte-Carlo-Schritte betrug immer $4 \cdot 10^7$, aber die Höchstzahl an Minimierungsschritten war abhängig von der Anzahl der Freiheitsgrade und berechnete sich erfahrungsgemäß zu $170 + 3 \cdot n$, wobei n die Anzahl der ϕ -, ψ - und Seitenkettentorsionen mit Ausnahme der Methylgruppenrotationen ist. Die Blockierungssituation bei Überbesuchung einer Konformationsfamilie wurde mit Randomisierung der freien Variablen mit einer Maximalamplitude von 30° aufgelöst, bei anderen Blockierungen wurde die Temperatur verdoppelt. Die ECEPP/3-Vakuumenergieterme sowie Solvations- und Entropieterme wurden während der Simulation berücksichtigt, wobei die van-der-Waals-Energie bei $7,5 \text{ \AA}$ abgeschnitten wurde. β -Reißverschlußbewegungen, die nach der Identifizierung eines von vier möglichen H-Bindungsmustern des Rückgrates eine Weiterführung der β -Struktur durch korrelierte Veränderung der benachbarten ϕ - ψ -Paare versuchen (Abagyan et al., 1999), wurde eine doppelte Wahrscheinlichkeit gegenüber anderen nichtdistributiven Bewegungen zugewiesen. Ein Konformationsstapel mit 100 Positionen wurde für jede Rechnung angelegt; insgesamt wurde mit dem gleichen Setup für jedes Peptid fünf Rechnungen durchgeführt, von denen eine mit einer vollständigen Trajektorie aufgezeichnet wurde.

2. 1. 6. 2 *Dockingexperimente*

In einem ersten Schritt wurden die kartesischen Koordinatensätze der Kristallstrukturen der Antikörper bzw. ihrer Komplexe (PDB Codes 1bog, 1cfs, 1cfq, 1hh6, 1hi6, 1e4w, 1tet) in

interne Koordinaten überführt in einer Regularisierungs-Prozedur, die eine idealisierte Kovalenzgeometrie mit größtmöglicher Ähnlichkeit zu den kartesischen Koordinaten herstellt. Wassermoleküle wurden alle entfernt aus der Struktur und entsprechend der Sequenz des zu regularisierenden Objektes eine ausgestreckte Polymerkette aus der Monomerbibliothek erstellt, die reguläre Geometrie gewährleistet und alle Atome enthält (auch Wasserstoffatome, die in der Regel bei Kristallstrukturen fehlen). Die identischen Schweratome wurden dann mittels Tether-Restraints (siehe Gl. 9) auf die entsprechenden Atome in der PDB-Struktur gezogen und minimiert unter Beibehaltung der idealen Kovalenzgeometrie. Methyl-Gruppen wurden anschließend rotiert zur Vermeidung von van-der-Waals-Zusammenstößen und eine kombinierte Geometrie/Energie-Optimierung (bis zu 50 Iterationen mit energieabhängigen Tether-Gewichten) durchgeführt, die auch die polaren Wasserstoffatome richtig orientiert. Abschließend wurde eine freie Minimierung ohne Restraints durchgeführt. Für das Docking von u2-pep wurden die Seitenketten-Torsionswinkel der Residuen in der Bindungstasche des nativen CB4-1 Antikörpers in einer kurzen Monte-Carlo-Rechnung (360000 Funktionsaufrufe, 600 K) optimiert.

Um die Bindungsstelle der Antikörper herum, die durch die sechs CDR's gut charakterisiert ist, wurde (nach Entfernen der jeweiligen Bindungspartner) auf ein Raumgitter (0,5 Å Gitterweite, Dimension ca. 31*24*32Å) die Werte eines Satzes von fünf Energiepotentialen projiziert (van-der-Waals-Energie für eine Schweratom- bzw. Wasserstoffsonde, elektrostatisches Potential, H-Bindungsenergie und hydrophobes Potential (Totrov et al., 1997; Totrov et al., 1999)). Das elektrostatische Potential wurde dabei unter -20 kcal/mol und über 20 kcal/mol abgeschnitten, als Maximalwert für die van-der-Waals-Potentiale wurde ein weicher Wert von 1,5 kcal/mol gewählt, der eine gewisse sterische Überlappung zwischen Peptid und den Potentialen erlaubt. Eine weitere Unschärfe der Bindungstasche wurde dadurch erzielt, daß allen Potentialwerten der Minimalwert der jeweils umliegende Gitterpunkte zugeordnet wurde, was effektiv eine Aufweitung der Bindungstasche unter Beibehaltung ihrer Geometrie bedeutet. Die zu dockenden Peptide wurden zunächst aus Monomerbausteinen mit ungeladenen Termini zusammengesetzt, ihre freien Variablen mit maximal 180° randomisiert und der Massenschwerpunkt in die Mitte der Potentialgitter-Box verlegt, um unnötiges Sampling am Simulationsbeginn zu vermeiden. Des Weiteren wurde der Entfernung des Peptids vom Koordinatenursprung des Systems ein ellipsoider Restraint zugeordnet (Muldentiefe -200 kcal/mol, Muldenweite 20 Å, 50% flacher Anteil in Gl. 10), um ein Wegdriften des Peptids aus der Potentialbox zu verhindern. Das Ähnlichkeitskriterium der ϕ - ψ -Winkel (Gl. 1) wurde auf 50° festgelegt, um im Konformationsstapel 100

nichtähnliche Strukturen zu sichern. ω -Winkel wurden in der Minimierung berücksichtigt, sie wurden aber nicht für das Sampling einbezogen, dessen Limit bei 10^8 Funktionsaufrufen festgelegt wurde. Neben den eigentlichen Peptidvariablen wurden aber auch die sechs mit der absoluten Lage des Peptids verknüpften virtuellen Variablen stochastisch verändert und minimiert, wobei für die Translation ein Maximalwert von 3 Å Amplitude galt (entsprechend für die Rotation 3 Å/Trägheitsradius). Die Anzahl der Minimierungsschritte war wiederum abhängig von der Länge des Peptids, in der Regel wurden nach jedem Zufallsschritt 250-300 Iterationen mit konjugiertem Gradienten erlaubt. Die im Metropolis-Kriterium verwendete Temperatur betrug wiederum 600 K, die wiederum verdoppelt werden konnte bei den problematischen Situationen 2 und 3 (siehe oben). Neben den ECEPP-Energietermen wurden für Minimierung und Evaluation die Gitterpotentiale sowie die entsprechenden Zwangsterme für den Abstand und eventuell für Disulfid- bzw. Lactambrückenbindungen verwendet (Gl. 8). Nach der eigentlichen Dockingsimulation an den weichen Gitterpotentialen wurden die Strukturen aus dem Konformationsstapel wieder dem Vollatom-Antikörpermodell hinzugefügt und eine vollständige Energieminimierung durchgeführt, wobei die Seitenketten der Antikörper-Bindungstasche flexibel gehalten wurden. Der Konformationsstapel wurde dann entsprechend neu sortiert, wobei auch der Beitrag der Seitenketten-Entropie und der Solvation berücksichtigt wurde. Die bestrangierte Struktur wurde anschließend verfeinert, indem die Seitenketten des Peptids zusammen mit den flexibel gehaltenen Seitenketten der Antikörper-Bindungstasche (als 6 Å-Umgebung des Peptids definiert) des Vollatom-Modells in einer nochmaligen kurzen Monte-Carlo Simulation (2 Mill. Funktionsaufrufe) konformationell durchsucht wurden, während die Rückgratkonformation des Peptids fixiert blieb, aber leichte Positions- und Konformationsänderungen während der Minimierungsschritte annehmen konnte.

2. 1. 7 Konformationsanalyse

Die Darstellung der Trajektorie einer Simulation bedeutet, einen vieldimensionalen Verlauf zwischen sehr vielen lokalen Minima, die sehr verschiedene Konformationen darstellen, auf wenige Dimensionen zu reduzieren. Mit einem geeigneten Distanzkriterium zwischen allen Paaren von n_{ref} Referenzstrukturen kann man in einem maximal $(n_{\text{ref}}-1)$ -dimensionalen Raum von Prinzipalkomponenten eine Verteilung dieser Strukturen darstellen und alle anderen Konformationen daran koordinieren (Gower, 1966; Abagyan et al., 1992). Es handelt sich daher um ein Problem der Distanzgeometrie, die aus einem Satz von Abständen ein Satz von

Koordinaten gewinnen will. Als Abstandskriterium bietet sich der anguläre RMSD (Gl. 1) an, obwohl die Periodizität von 2π eine exakte Lösung nicht zuläßt, die streng genommen nur im euklidischen Raum erhältlich ist. Der nichteuklidische Fehler ist aber als gering einzuschätzen. Für die paarweisen Abstände (anguläre RMSD der ϕ - ψ -Werte) der n_{ref} Referenzstrukturen d_{ik} ($i, k=1, \dots, n_{\text{ref}}$) kann man die Koordinaten x_{ij} ($j=1, \dots, n_p; n_p \leq n_{\text{ref}}-1$) im n_p -dimensionalen euklidischen Raum folgendermaßen finden.

$$(1) \text{ Definition der Matrix } a_{ik} \text{ mit } a_{ik} = -\frac{1}{2} d_{ik}^2 \quad (12)$$

$$(2) \text{ Transformation in Matrix } b_{ik}: b_{ik} = a_{ik} - \langle a_{ik} \rangle_i - \langle a_{ik} \rangle_k + \langle a_{ik} \rangle_{ik} \quad (13)$$

(3) Auffinden aller positiven Eigenwerte λ_j von b_{ik} und der korrespondierenden Eigenvektoren $[x_{ij}]$ mit $i=1, \dots, n_{\text{ref}}$ und $j=1, \dots, n_p$

(4) Skalieren der Eigenvektoren, so daß gilt:

$$\mathbf{X}^T \mathbf{X} = \mathbf{\Lambda}, \text{ wobei } \mathbf{X} = \text{mat}([x_{ij}]) \text{ und } \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{n_p}) \quad (14)$$

Der größte Eigenwert repräsentiert die größte Varianz entlang seiner Vektorachse, die drei größten Eigenwerte können demnach für eine dreidimensionale Darstellung verwendet werden; sie machen in der Regel ca. 80% der Gesamtvarianz aus, wobei der nichteuklidische Fehler mit weniger als 5% beobachtet wurde. Als Referenzstrukturen wurden solche verwendet, die in der jeweiligen Trajektorie der Simulation den Maximalgehalt an den Sekundärstrukturelementen α -Helix, β -Brücke, β -Faltblatt, π -Helix und 3_{10} -Helix aufwiesen. Die Koordinaten aller anderen Strukturen c der Trajektorie mit den Abständen d_{ci} zu den n_{ref} Referenzstrukturen berechnen sich dann in Relation zu deren Koordinaten x_{ij} als

$$x_{cj} = \frac{\sum_{i=1}^{n_{\text{ref}}} x_{ij} (b_{ii} - d_{ci}^2)}{2\lambda_j} \quad \text{mit } j=1, \dots, n_p \quad (15)$$

Für die normalisierte statistische Wichtung w_i einer Konformation i gegenüber dem Ensemble eines Konformationsstapels bzw. den Strukturen einer Trajektorie wurde die Boltzmann-Gleichung verwendet, wie von anderen Autoren vorgeschlagen (Zimmerman et al., 1977):

$$w_i = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_{i=1}^n e^{-\frac{\Delta G_i}{RT}}} \quad (16)$$

Mit diesem Gewicht wurden dann die entsprechenden Eigenschaften berechnet, z. B. den Gehalt an Sekundärstrukturelementen $\langle q \rangle$:

$$\langle q \rangle = \sum_{i=1}^n w_i \frac{n_i^{SE}}{l_i} \quad (17)$$

wobei n_i^{SE} die Anzahl der in das jeweilige Sekundärstrukturmotiv SE einbezogenen Aminosäuren und l_i die Länge des gesamten Peptids ist. Die statistisch gemittelten Eigenschaften wurden immer für 298 K berechnet.

β -Turns wurden identifiziert zwischen den Residuen i und $i+3$, wenn deren $C\alpha$ -Atome weniger als 7 Å voneinander entfernt waren und die beiden zentralen Aminosäuren nicht helical waren. Die Kriterien für die Bestimmung der β -Turntypen wurden aus dem PROMOTIF-Programm entnommen (Hutchinson et al., 1996), die sich wiederum nach veröffentlichten Werten richten (Venkatachalam, 1968; Richardson, 1981). Als Toleranz von den idealen Werten wurden dabei $\pm 30^\circ$ zugelassen. Für die Identifizierung von Helices und β -Brücken- und Faltblattabschnitten wurde ein modifizierter DSSP-Algorithmus (Kabsch et al., 1983) eingesetzt, der in ICM implementiert ist.

2. 2 Design cyclischer Peptide

Für das Design von cyclischen Peptiden ausgehend von einem antikörpergebundenen Templatpeptid wurde die Sequenz zunächst aus Monomerbausteinen aufgebaut. Die zum Templat identischen Atome wurden dann mit Tether-Restraints (siehe Gl. 9) mit einer Kraftkonstante von 1 kcal/mol/Å² auf die entsprechende Koordinaten des Templats gezogen und die ideale Bindungsgeometrie anschließend wiederhergestellt. Die entsprechende Cyclisierung wurde dann durch Zuweisung von Distanz-Zwangstermen (Gl. 8) emuliert, wobei gleichzeitig die überflüssigen Atome eliminiert wurde ($H^{S\gamma}$ bei Disulfid-Brücken, $O^{C\gamma/\delta}$ und $H^{N\zeta}$ bei Lactambrücken). Unter Einbeziehung der Distanzterme und der Tether-Restraints wurde dann eine vollständige Energieminimierung der Vakuum-Energierterme in Gegenwart des Vollatom-Antikörpermodells durchgeführt. Das entsprechende Konstrukt wurde angenommen, wenn keine van-der-Waals Zusammenstöße beobachtet wurden und die Penalty-Funktionen (Straf-Funktionen) von Tethers und Cyclisierungstermen Werte von kleiner als 0,5 kcal/mol hatten.

2.3 Computerressourcen

Für die Rechnungen wurden entweder Prozessoren vom Typ R10000 (180 MHz IP27/175 MHz IP32) oder vom Typ Pentium II (400 MHz) bzw. Pentium III (600/700 MHz) verwendet. Faltungsrechnungen benötigten in der Regel ca. 100 h pro Simulation, Dockingrechnungen bis 150 h. Für eine Verfeinerungsrechnung der gedockten Strukturen sind ca. 60 h zu veranschlagen.

2.4 Bindungsstudien

2.4.1 *Kompetitions-ELISA*

Von einer Lösung von 0,1 µg/ml p24-Peptid SALSEGATQDLNTMLNTV in 0,1 M Natriumcarbonatpuffer (pH 9,5) wurden 50 µl in jede Vertiefung einer 96-Löcher MaxiSorb-Mikrotiterplatte (Nunc, Roskilde, Dänemark) gebracht und über Nacht bei 4°C inkubiert. Nach dreimaligem Waschen mit T-PBS-Puffer (1,6 mM NaH₂PO₄, 9,2 mM Na₂HPO₄, 0,15 M NaCl, 0,1% Tween-20, pH 7,8) wurde CB4-1 Antikörper (mit Meerrettich-Peroxidase markiert, 0,8 µl/ml) mit den in einer geometrischen Reihe verdünnten Peptiden (Ausgangskonzentrationen: 1 µM bei h-pep, 1 mM bei Übergangspeptiden bzw. cyclischen Analoga) in T-PBS-Puffer mit 6% Gelifundol S (Biotest, Dreieich) in einem Gesamtvolumen von 50 µl zugegeben und über Nacht bei 4°C inkubiert. Nach erneutem dreimaligen Waschen mit T-PBS-Puffer wurde die Aktivität des gebundenen Enzyms bestimmt, indem 5,5 mM o-Phenylendiaminehydrochlorid (Fluka, Buchs, Schweiz) und 8,5 mM H₂O₂ in 0,1 M Citratpuffer (pH 5,0) zugegeben wurde. Die Reaktion wurde nach 10 Min. abgebrochen durch Zusatz von 0,05 M Natriumsulfit in 1 M Schwefelsäure. Die Absorption wurde bei 490 nm bzw. 620 nm als Referenz an einem ELISA-Reader (BioluminTM960, Molecular Dynamics, Sunnyvale, USA) gemessen. Die Ermittlung der Inhibitionskonstanten erfolgte nach (Friguet et al., 1985) über die Beziehung

$$K_I = c \frac{E_0}{E_0 - E} \quad (17)$$

(c-Konzentration, E-Absorption, E₀-Absorption ohne Peptid) mit Hilfe der Software GraphPad Prism[®] (GraphPad Software Inc., San Diego).

2. 4. 2 Substitutionsanalysen

Substitutionsanalysen wurden auf Cellulose-Membranen Whatman 50 (Whatman, Maidstone, UK) durch robotergestützte Spot-Synthese hergestellt (Abimed GmbH, Langenfeld; Software DIGEN, Jerini Biotech GmbH, Berlin) nach einem Standardprotokoll (Frank, 1992; Kramer et al., 1998). Die cellulosegebundenen Peptide wurden N-terminal acetyliert (bis auf die head-to-sidechain verbrückten Lactamcyclen) mit 10% Acetanhydrid/20% Diisopropylethylamin in Dimethylformamid (DMF). Disulfid-cyclisierte Peptide wurden auf der Membran erzeugt durch Inkubation mit 10% Dimethylsulfoxid in Phosphatpuffer (10 mM, pH 7,5) über Nacht bei Raumtemperatur. Zur Erzeugung von Lactamcyclen wurden bei der Spotsynthese für Aspartat/Glutamat und Lysin spezielle Seitenketten-Schutzgruppen eingesetzt (2-Phenylisopropylester bei Asp/Glu; N ζ -alkyliert mit 4-Monomethoxytrityl bei Lys). Diese wurden selektiv abgespalten mit 1% Trifluoressigsäure/5% Triisobutylsilan in Dichlormethan (DCM) (6x10 Min.). Die Membranen wurden zweimal mit Dimethylformamid, anschließend einmal mit 5% Triethylamin in DMF und wieder zweimal mit DMF gewaschen. Die Cyclisierung erfolgte in einer Aktivierungslösung (0,05 M o-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluroniumhexafluorophosphat (HATU), 0,1 M Diisopropylethylamin in DMF) über 2x3 Stunden bei Raumtemperatur. Schließlich wurden die restlichen Schutzgruppen mit 95% Trifluoressigsäure in DCM entsprechend dem Standardprotokoll abgespalten.

Vor der Inkubation der Membranen wurden diese 10 Min. mit Methanol und anschließend dreimal mit TBS-Puffer gewaschen (8 ‰ m/m NaCl, 0,2 ‰ m/m KCl, 6,1 ‰ m/m Trishydroxymethylaminomethan, pH 8,0), gefolgt von einer Blockierung mit Blockierungsreagens (Sigma-Genosys, Cambridge, USA) für 2 h in TBS mit 5 % Sucrose. Nach 10 minütigem Waschen mit T-TBS (TBS-Puffer mit 0,05 % Tween 20) wurde der Antikörper (1 μ g/ml in Blockierungspuffer) zugegeben und die Membran über Nacht bei 4°C inkubiert. Anschließend wurde wieder 3x5 Min. mit T-TBS gewaschen und ein Peroxidase-markierter anti-Maus-Antikörper (A-9309, Sigma, Deisenhofen, Germany), 1:1000 in Blockierungspuffer, zugesetzt und 2 h bei Raumtemperatur inkubiert. Nach nochmaligem Waschen 3x10 Min. mit T-TBS wurde die Membran mit einem Chemolumineszenz-System (Pierce, Rockford, Illinois) entwickelt und die Bindung mit einem LumiImager™ (Boehringer Mannheim GmbH, Mannheim) detektiert.