

Neue Rechenverfahren für die Proteinstrukturbestimmung mit Hilfe der magnetischen Kernspinresonanz



Peter Güntert

Diss. ETH Nr. 10135

**Neue Rechenverfahren für die
Proteinstrukturbestimmung mit Hilfe der
magnetischen Kernspinresonanz**

**Abhandlung
zur Erlangung des Titels
Doktor der Naturwissenschaften
der
Eidgenössischen Technischen Hochschule Zürich**

**vorgelegt von
Peter Güntert
Dipl. Phys. ETH**

**geboren am 29. März 1964
von Schleithem (Schaffhausen)**

**Angenommen auf Antrag von
Prof. Dr. Kurt Wüthrich, Referent
Prof. Dr. Wilfred van Gunsteren, Korreferent
1993**

Copyright © 1993 Peter Güntert
All rights reserved.

Vorwort

Die hervorragende Leitung meiner Doktorarbeit durch Herrn Professor Kurt Wüthrich und die außergewöhnlich gute Zusammenarbeit mit meinen Kollegen im Institut für Molekularbiologie und Biophysik haben diese Arbeit ermöglicht.

Herrn Professor Kurt Wüthrich danke ich für die Möglichkeit, diese Arbeit in seiner Forschungsgruppe durchzuführen. Er hat mir jederzeit große Freiheit und weitreichende Unterstützung gewährt.

Herrn Professor Wilfred van Gunsteren danke ich für das durch die Übernahme des Korreferats bezeugte Interesse an meiner Arbeit.

Ganz besonders danke ich Dr. Martin Billeter für die große Hilfe, Unterstützung und auch Geduld, die er mir während der ganzen Zeit, die wir im gleichen Büro hier am Institut für Molekularbiologie und Biophysik verbrachten, entgegenbrachte.

Große Freude bereitete mir jederzeit die Zusammenarbeit mit Frau Dr. Yan-Qiu Qian, Volker Dötsch, Peter Luginbühl und Dr. Thomas Szyperski. Ich danke ihnen sehr.

Ein besonderer Dank gebührt auch Rudolf Baumann, der dafür sorgte, daß die meisten Computer-Probleme gar nicht erst aufzutauchen wagten.

Für die fruchtbare Zusammenarbeit bin ich Dr. Kurt Berndt, Dr. Werner Braun, Dr. Craig Eccles, Dr. John Mertz, Dr. Leon Orbons, Prof. Gottfried Otting, Niklaus Schaefer, Dr. Gerhard Wider und Dr. Tai-he Xia zu großem Dank verpflichtet.

Für all die Hilfe, Zusammenarbeit und interessanten Diskussionen, ohne die meine Arbeit nicht möglich gewesen wäre, danke ich allen Kollegen am Institut für Molekularbiologie und Biophysik.

Dr. Martin Billeter, Volker Dötsch und Dr. Thomas Szyperski danke ich auch für ihr sorgfältiges Durchlesen und Verbessern des Manuskripts der vorliegenden Abhandlung.

Meinen Eltern danke ich ganz herzlich dafür, daß sie meine Ausbildung ermöglicht haben.

Inhaltsverzeichnis

<i>Kurzfassung</i>	7
<i>Summary</i>	9
<i>Einführung</i>	11
<i>Produktoperatorrechnungen in Mathematica</i>	18
<i>Verarbeitung von Kernresonanzdaten mit dem Programm PROSA</i>	27
Lineare Vorhersage	28
Automatische Phasenkorrektur	30
Basislinienkorrektur	33
Gebrauch des Programms PROSA	35
PROSA-Anweisungen	39
PROSA-Programmvariablen	49
PROSA-Standardmakros	52
<i>Konformationseinschränkungen aus Kern-Overhausereffektspektren</i>	56
Intensitätsbestimmung bei überlappenden Signalen	56
Messung skalarer Spin-Spin-Kopplungskonstanten durch inverse Fourier-Transformation von In-Phase-Multipletts	58
Umwandlung von Signalintensitäten in Distanzschranken	61
Signalzuordnungen aufgrund dreidimensionaler Strukturen	63
<i>Datenformate für Strukturrechnungen</i>	66
Bausteinbibliothek	66
Aminosäuresequenz	73
Konformationen, beschrieben durch kartesische Koordinaten	74
Konformationen, beschrieben durch Diederwinkel	75
Skalare Spin-Spin Kopplungskonstanten	76
Stereospezifisch zugeordnete diastereotope Gruppen	76
Obere und untere Distanzschranken	77

Diederwinkelschränkungen	78
Bearbeitung von Datensätzen mit dem Programm COFIMA	79
<i>Stereospezifische Zuordnungen</i>	<i>84</i>
Einfluß stereospezifischer Zuordnungen auf Lösungsstrukturen	84
Bestimmung stereospezifischer Zuordnungen	94
<i>Berechnung dreidimensionaler Proteinstrukturen mit dem Programm DIANA</i>	<i>104</i>
Die variable Zielfunktion	106
Eingabe- und Ausgabedaten des Programms DIANA	109
Irrelevante und zu restriktive Einschränkungen	109
Diastereotope Gruppen ohne stereospezifische Zuordnung	110
Van der Waals Abstoßung der Atome	112
Berechnung der Atomkoordinaten aus Diederwinkelwerten	114
Minimierung der Zielfunktion	117
Optimierung der Strukturrechnung mit DIANA	118
Interaktive Eingaben für DIANA	119
Standardminimierungsparameter für DIANA	124
Beispiel einer Strukturrechnung mit DIANA	124
<i>Konvergenzbeschleunigung in Strukturrechnungen mit dem Programm DIANA</i>	<i>132</i>
Redundante Diederwinkelschränkungen (REDAC)	133
Ausmaß der Konvergenzbeschleunigung	134
Grundlage der Konvergenzbeschleunigung	139
Gebrauch der REDAC-Strategie mit dem Programm DIANA	140
<i>Anwendungen</i>	<i>143</i>
Analyse und Vergleich von Proteinstrukturen	143
Antennapedia(C39S) Homeodomäne	144
Hirudin(1–51)	149
<i>Literaturverzeichnis</i>	<i>158</i>
<i>Abkürzungen</i>	<i>169</i>
<i>Lebenslauf</i>	<i>170</i>

Kurzfassung

Diese Abhandlung führt neue Rechenmethoden ein zur Ermittlung der dreidimensionalen Struktur von Proteinen in Lösung aufgrund von Daten, wie sie die magnetische Kernspinresonanzspektroskopie (Ernst *et al.*, 1987) liefert. Um das vor einigen Jahren dafür eingeführte Verfahren (Wüthrich, 1986) auf kleine Proteine effizient oder auf größere Proteine überhaupt anwenden zu können, waren und sind auf verschiedenen Stufen — bei den Kernresonanzexperimenten, der Herstellung und Auswertung der Spektren und den Strukturrechnungen — Neuerungen und Verbesserungen nötig. Die im Rahmen dieser Dissertation entwickelten Verfahren werden hier in der Reihenfolge, in der sie während einer Proteinstrukturbestimmung zum Einsatz kommen, vorgestellt und an Hand von praktischen Anwendungen erläutert. Der Schwerpunkt liegt bei der Berechnung dreidimensionaler Proteinstrukturen mittels Distanzgeometrierechnungen.

Dem Entwurf neuer Experimente in der hochauflösenden Kernspinresonanzspektroskopie dient der Produktoperatorformalismus, dessen erstmalige vollständige Implementation in *Mathematica*, einer Programmiersprache für symbolisches Rechnen, die Simulation von Kernspinresonanzexperimenten mit Spin-1/2-Systemen wesentlich erleichtert (Güntert *et al.*, 1993).

Das neue Programm PROSA zur Verarbeitung mehrdimensionaler Kernresonanzdaten wird vorgestellt, in dem die gebräuchlichen Verarbeitungsschritte wie lineare Vorhersage, digitale Filterung, Fourier-Transformation und automatische Phasen- und Basislinienkorrektur effizient und flexibel implementiert sind (Güntert *et al.*, 1992). Dank einer zuverlässigen Routine zur automatischen Phasenkorrektur wird keine Grafik benötigt, so daß das Programm ohne weiteres auch auf Hochleistungsrechnern, für die es besonders optimiert ist, laufen kann. Ebenfalls neu ist das FLATT Verfahren zur Basislinienkorrektur (Güntert & Wüthrich, 1992).

Zur Intensitätsbestimmung von Kern-Overhausereffekten wird eine Methode vorgeschlagen, die gut aufgelösten Signalen Referenzlinienformen entnimmt, um mit ihrer Hilfe die Intensitäten überlappender Signale zu finden (Eccles *et al.*, 1991). Die Messung skalarer Spin-Spin-Kopplungskonstanten durch inverse Fourier-Transformation von In-Phase-Multipletts wird einge-

führt (Szyperski *et al.*, 1992a). Überlappende Resonanzfrequenzen erschweren sowohl die Resonanzzuordnung als auch Intensitätsbestimmung von Kern-Overhausereffekten. Ein Verfahren zur Auflösung von Zweideutigkeiten in der Resonanzzuordnung, implementiert im Programm ASNO, verwendet dazu eine vorläufige dreidimensionale Struktur (Berndt *et al.*, 1993).

Das Programm DIANA wurde für die Berechnung der dreidimensionalen Struktur von Proteinen aufgrund geometrischer Einschränkungen, d. h. oberer Distanzschranken aus Kern-Overhausereffekten, unterer sterischer Distanzschranken und Diederwinkeleinschränkungen aus skalaren Spin-Spin Kopplungskonstanten, entwickelt, wobei die einzigen Freiheitsgrade des Moleküls die drehbaren Diederwinkel sind. Eine Startkonformation mit zufälligen Einstellungen der Diederwinkel wird schrittweise an die Einschränkungen anpaßt (Güntert *et al.*, 1991a). Weil es sich dabei um ein kompliziertes Minimierungsproblem in einem hochdimensionalen Raum handelt, treten oft lokale Minima auf. Ein wirkungsvolles Verfahren, die Häufigkeit ihres Auftretens zu verringern und damit die Konvergenz zu beschleunigen, ist die Verwendung redundanter Diederwinkeleinschränkungen (Güntert & Wüthrich, 1991).

Eine Reihe von Hilfsprogrammen umgeben das zentrale Programm DIANA: CALIBA zur Umwandlung von Signalintensitäten in obere Distanzschranken, HABAS (Güntert *et al.*, 1989) und GLOMSA zur Bestimmung stereospezifischer Zuordnungen, deren Einfluß auf die Genauigkeit der berechneten Struktur diskutiert wird, und COFIMA zur Bearbeitung und Formatumwandlung verschiedener Datentypen (Güntert, 1992).

Die Strukturbestimmungen von zwei Proteinen — der *Antennapedia* (C39S) Homeodomäne (Güntert *et al.*, 1991b) und von Hirudin(1–51) (Szyperski *et al.*, 1993b) — illustrieren die Anwendung und Nützlichkeit der vorgestellten neuen Methoden zur Strukturbestimmung aufgrund von Kernresonanzdaten in Lösung.

Summary

This dissertation introduces new computational methods for the determination of three-dimensional protein structures from nuclear magnetic resonance data (Ernst *et al.*, 1987). The method introduced for this purpose a few years ago (Wüthrich, 1986) needs innovations and improvements at different stages — the nuclear magnetic resonance experiments, the processing and evaluation of spectra, and the structure calculations — in order to apply it efficiently to small proteins, or to larger proteins. The new methods developed during the dissertation are presented in the order in which they are employed in the course of a protein structure determination, and their use is explained by means of practical applications. The emphasis lies on the calculation of three-dimensional protein structures with distance geometry calculations.

The product operator formalism is an important tool to develop new schemes for nuclear magnetic resonance experiments. Its first complete implementation in *Mathematica*, a programming language designed for symbolic computation, simplifies the simulation of nuclear magnetic resonance experiments with spin-1/2 systems significantly (Güntert *et al.*, 1993).

The new program PROSA for the processing of multi-dimensional nuclear magnetic resonance data is presented. It provides efficient and flexible implementations of the common data processing steps such as linear prediction, digital filtering, Fourier transformation, and automatic phase and baseline correction (Güntert *et al.*, 1992). Due to a reliable automatic phase correction method no graphics display is needed, such that the program can operate on supercomputers for which it is especially optimized. New is also the FLATT method for baseline correction (Güntert & Wüthrich, 1992).

For the determination of the intensity of nuclear Overhauser effects a method is proposed that uses reference line-shapes from well-separated peaks in order to find the intensities of overlapping peaks (Eccles *et al.*, 1991). The measurement of scalar spin-spin coupling constants by inverse Fourier transformation of in-phase multiplets is introduced (Szyperski *et al.*, 1992a). Overlapping resonance frequencies make both the assignment and the quantification of nuclear Overhauser effects difficult. A method to resolve ambiguities in resonance assignments, implemented in the program ASNO, uses a preliminary three-dimensional structure (Berndt *et al.*, 1993).

The distance geometry program DIANA was developed for the calculation of three-dimensional protein structures on the basis of geometric restraints, *i. e.*, upper distance bounds from nuclear Overhauser effects, steric lower distance bounds, and dihedral angle constraints from scalar spin-spin coupling constants. The only degrees of freedom of the molecule are its rotatable dihedral angles. A start conformation with random values of the dihedral angles is adapted to the constraints in a step-wise fashion (Güntert *et al.*, 1991a). Local minima are abundant for such a complex optimization problem in a high-dimensional space. The use of redundant dihedral angle constraints is an effective method to improve the convergence by reducing the occurrence of local minima (Güntert & Wüthrich, 1991).

Several auxiliary programs surround the central program DIANA: CALIBA converts signal intensities into upper distance bounds, HABAS and GLOMSA determine stereospecific assignments, whose influence on the precision of the calculated structures is discussed, and COFIMA is used to manipulate different types of data that is relevant for the structure calculation (Güntert, 1992).

The structure determinations of two proteins — the *Antennapedia(C39S)* homeodomain (Güntert *et al.*, 1991b) and Hirudin(1–51) (Szyperski *et al.*, 1992b) — illustrate the application and advantages of the new methods for the structure determination on the basis of nuclear magnetic resonance data in solution.

Einführung

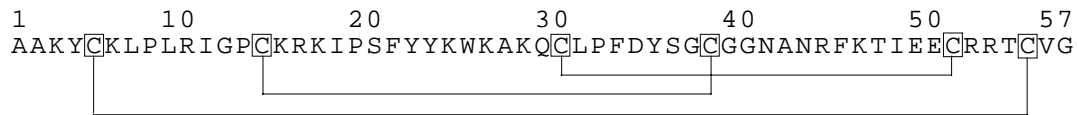
Proteine sind eigentliche »Moleküle des Lebens«, die in allen Lebewesen eine Vielzahl von Funktionen wahrnehmen (Dickerson & Geis, 1971; Stryer, 1985). Insbesondere können sie in der Form von Enzymen als meist hochspezifische und hochwirksame Katalysatoren auftreten — eine Aufgabe, die in Lebewesen fast ausschließlich Proteinen vorbehalten ist. Proteine können aber auch auf andere Weise an grundlegenden Vorgängen des Lebens teilhaben, zum Beispiel indem sie als Inhibitoren die Wirkung von Enzymen steuern, indem sie spezifisch an die Erbsubstanz binden und damit deren Übersetzung oder Vervielfältigung regeln oder indem sie als Bausteine für den Aufbau von Geweben dienen.

Die überragende Bedeutung der Proteine für das Leben kommt auch darin zum Ausdruck, daß die genetische Information, die im Erbgut bewahrt wird, in allererster Linie aus den Bauplänen der Proteine besteht — Bauplänen, die außerordentlich kompakt gehalten werden können, weil sowohl die Bausteine eines Proteins als auch deren Verknüpfungen untereinander standardisiert sind: aus nur 20 verschiedenen Bausteinen, den 20 Standardamino­säuren, lassen sich mit Hilfe einer einzigen Verknüpfung, der Peptidbindung, alle Proteine aufbauen (Schulz & Schirmer, 1979). Der Bauplan eines Proteins besteht deshalb lediglich aus der Folge der Aminosäuren, die, der Reihe nach durch Peptidbindungen miteinander verknüpft, eine lineare Kette bilden — die »Primärstruktur« des Proteins. Grundsätzlich bestimmt bei natürlich vorkommenden Proteinen diese Primärstruktur, welche Faltung die Kette im dreidimensionalen Raum einnimmt und damit, welche Funktion das Protein besitzen wird, da sich die Funktion auf regelmäßige, sich wiederholende Faltungselemente, die Sekundärstruktur, und vor allem die Faltung der Kette als Ganzes, die Tertiärstruktur, zurückführen läßt (Bild 1). In der Natur führt also ein direkter Weg von der genetischen Information, dem Bauplan, über die Herstellung der Polypeptidkette zur dreidimensionalen Struktur und damit zur Funktion des Proteins.

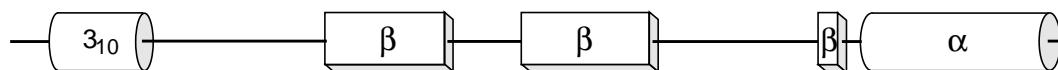
Der Wissenschaft ist dagegen dieser direkte Weg zur Struktur und Funktion der Proteine bis heute durch einen tiefen Einschnitt in unserem Verständnis verwehrt: der Zusammenhang zwischen dem Bauplan und der Faltung

eines Proteins ist so ungenügend bekannt, daß die — leicht erhältliche — Kenntnis der Aminosäuresequenz keine verlässliche Vorhersage der Faltung oder gar der Funktion erlaubt. Statt dessen begibt sich die Wissenschaft auf

Primärstruktur



Sekundärstruktur



Tertiärstruktur

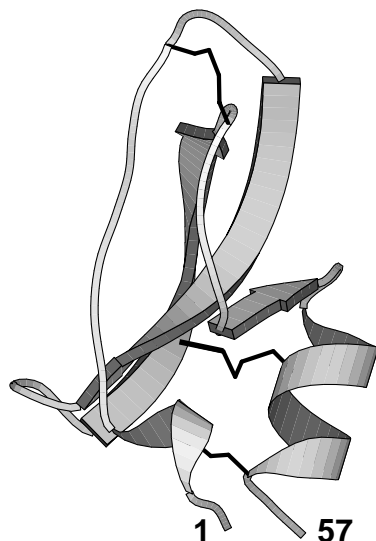


Bild 1. Primär-, Sekundär- und Tertiärstruktur des Proteins Toxin K der Schwarzen Mamba (Wüthrich et al., 1993). In der Primärstruktur sind die Disulfidbrücken durch Linien wiedergegeben. In der Sekundärstruktur bezeichnet »3₁₀« eine 3₁₀-Helix, »β« einen Strang eines β-Blatts und »α« eine α-Helix. In der schematischen Darstellung der Tertiärstruktur mit Hilfe des Programms MOLSCRIPT (Kraulis, 1991) sind Helices durch spiralenförmige und β-Stränge durch pfeilförmige Bänder dargestellt.

den Umweg der Proteinstrukturaufklärung mittels physikalischer Methoden, wofür zur Zeit zwei Verfahren im Gebrauch sind: die magnetische Kernresonanzspektroskopie (Abraham, 1961; Ernst *et al.*, 1987), die im Mittelpunkt dieser Abhandlung stehen wird, und die Beugung von Röntgenstrahlen an Proteineinkristallen (Glusker & Trueblood, 1985), die 1959 die erste hochaufgelöste dreidimensionale Struktur eines Proteins — Myoglobin — lieferte (Kendrew *et al.*, 1960; Kendrew, 1963). Die Röntgenkristallografie ist auch heute noch das einzige Verfahren zur Aufklärung der Faltung von Proteinen mit mehr als etwa 200 Aminosäureresten. Ihre Anwendung ist aber durch die Einschränkung auf Einkristalle oft behindert. Man kann einerseits nie völlig gewiß sein, ob die Faltung im Einkristall und im natürlichen Wirkungsgebiet der allermeisten Proteine — der Lösung — gleich ist, und andererseits ist es oft

langwierig oder gar unmöglich, Einkristalle eines Proteins zu züchten, weil gezielte Verfahren dafür fehlen.

Die Strukturaufklärung aufgrund der magnetischen Kernresonanz (Wüthrich, 1986, 1988), die für bis etwa 200 Aminosäurereste große Proteine verwendet werden kann, bietet dagegen den großen Vorteil, daß die Proteine in Lösung, also in annähernd natürlicher Umgebung untersucht werden. Im Vergleich zur Röntgenstrukturanalyse handelt es sich um eine junge Technik: 1985 wurde als erste dreidimensionale Proteinstruktur in Lösung die Faltung des Proteinaseinhibitors IIa aus Bullensamen bestimmt (Williamson *et al.*, 1985). Seither hat sich die Methode stürmisch entwickelt — 1988 erschien für Tendamistat die erste hochaufgelöste Lösungsstruktur eines Proteins (Kline *et al.*, 1988); in den Jahren 1990 und 1991 wurden die Lösungsstrukturen von 23 bzw. 38 verschiedenen biologischen Makromolekülen veröffentlicht (Hendrickson & Wüthrich, 1991, 1992) — und ist heute dabei, zum in Forschung und Industrie weitverbreiteten Standardverfahren zu werden. Zusätzlich zum statischen Bild der dreidimensionalen Struktur kann die magnetische Kernresonanzspektroskopie Einblick in dynamische Vorgänge im Protein auf verschiedenen Zeitskalen im Bereich von 10^{-12} – 10^4 s gewähren, und ist damit auch Ausgangspunkt für die theoretische Auseinandersetzung mit der Proteindynamik (McCammon & Harvey, 1987; Brooks *et al.*, 1988).

Die magnetische Kernresonanzspektroskopie beruht auf der Wechselwirkung zwischen den magnetischen Momenten der Atomkerne und äußeren Magnetfeldern (Bloch, 1946). Die magnetischen Momente — in biologischen Makromolekülen in erster Linie der Spin-1/2 Kerne ^1H , ^{13}C und ^{15}N — sind empfindliche Sensoren für ihre lokale Umgebung, die in großer Zahl in einem Protein auftreten, sehr gut lokalisiert sind und nur schwach untereinander und mit ihrer Umgebung wechselwirken: drei Eigenschaften, die es erlauben, reichhaltige, spezifische und leicht interpretierbare Information über ein Molekül zu gewinnen (Ernst, 1992). Magnetische Dipol-Dipol Wechselwirkungen zwischen räumlich benachbarten Kernspins liefern Angaben über interatomare Abstände (Solomon, 1955), und skalare J -Kopplungen zwischen Kernspins, die durch drei kovalente Bindungen voneinander entfernt sind, erlauben die Einschränkung von Diederwinkeln (Karplus, 1963).

Mit Hilfe der magnetischen Kernresonanzspektroskopie lassen sich daher genügend viele Parameter messen, um die dreidimensionale Struktur eines Proteins weitgehend festzulegen — trotz der hohen Anzahl von Freiheitsgraden eines Makromoleküls, die gleich der Anzahl seiner drehbaren Diederwinkel ist, wenn die Bindungslängen und -winkel konstant gehalten werden. Die magnetische Kernresonanzspektroskopie ist allerdings, wie auch die Röntgenstrukturanalyse, kein »Mikroskop mit atomarer Auflösung«, welches unmittelbar ein direktes Bild der Proteinkonformation liefern würde. Vielmehr muß die

Faltung aus den gemessenen Kernresonanzdaten *berechnet* werden. Weil bereits ein kleines Protein etwa tausend Atome und mehrere Hundert frei drehbare Diederwinkel besitzt, die durch bis zu tausend Kernresonanzmeßwerte eingeschränkt werden (Bild 2), müssen große Datenmengen bewältigt und aufwendige Berechnungen durchgeführt werden — Aufgaben, die nur mit Hilfe von leistungsfähigen Rechnern angegangen werden können.

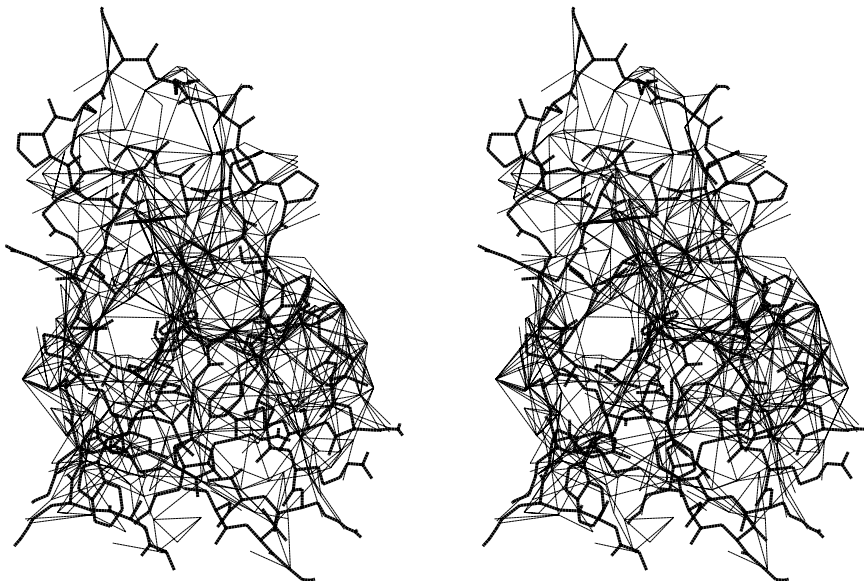


Bild 2. Stereobild der Lösungsstruktur des Proteins Toxin K der Schwarzen Mamba. Die kovalenten Bindungen zwischen Nichtwasserstoffatomen sind durch dicke, die 809 für die Strukturrechnung verwendete oberen Distanzschranken durch dünne Linien dargestellt (Wüthrich et al., 1993).

Der Ablauf einer Lösungsstrukturbestimmung mit Hilfe der magnetischen Kernresonanzspektroskopie ist im Bild 3 schematisch dargestellt. Er beinhaltet die Präparation des Proteins, die Durchführung der Kernresonanzexperimente, das Prozessieren der gemessenen Rohdaten, die Zuordnung der Resonanzlinien zu einzelnen Protonen, das Sammeln von konformationellen Einschränkungen und schließlich die Berechnung und Verfeinerung der dreidimensionalen Struktur. Die Herstellung und Reinigung des Proteins für Kernspinresonanzexperimente wird weitgehend dadurch bestimmt, daß die Kernspinresonanzspektroskopie eine recht unempfindliche Methode ist, die rund 0,5 ml Lösung mit einer Proteinkonzentration im millimolaren Bereich verlangt, wo schon kleinere Mengen an Verunreinigungen durch zusätzliche Resonanzlinien die Auswertung der Spektren beträchtlich erschweren können. Die Daten für die Strukturbestimmung kleiner Proteine — bis etwa 100 Aminosäurereste — können mit zweidimensionalen, homonuklearen ^1H Kernspinresonanzexperimenten gewonnen werden (Wüthrich, 1986), während für

größere Proteine auch höherdimensionale, heteronukleare (^1H , ^{13}C , ^{15}N) Experimente notwendig sind, um die Schwierigkeiten durch zunehmende Linienverbreiterung und -überlappung zu überwinden (Clore & Gronenborn, 1991). In einem mehrdimensionalen Kernspinresonanzexperiment zeichnet das Spektrometer ein Signal in Abhängigkeit von mehreren Zeitkoordinaten auf. Durch Prozessieren dieser Rohdaten entstehen mehrdimensionale Spektren — Darstellungen der Signale im Frequenzbereich (Ernst *et al.*, 1987). Ihre Aus-

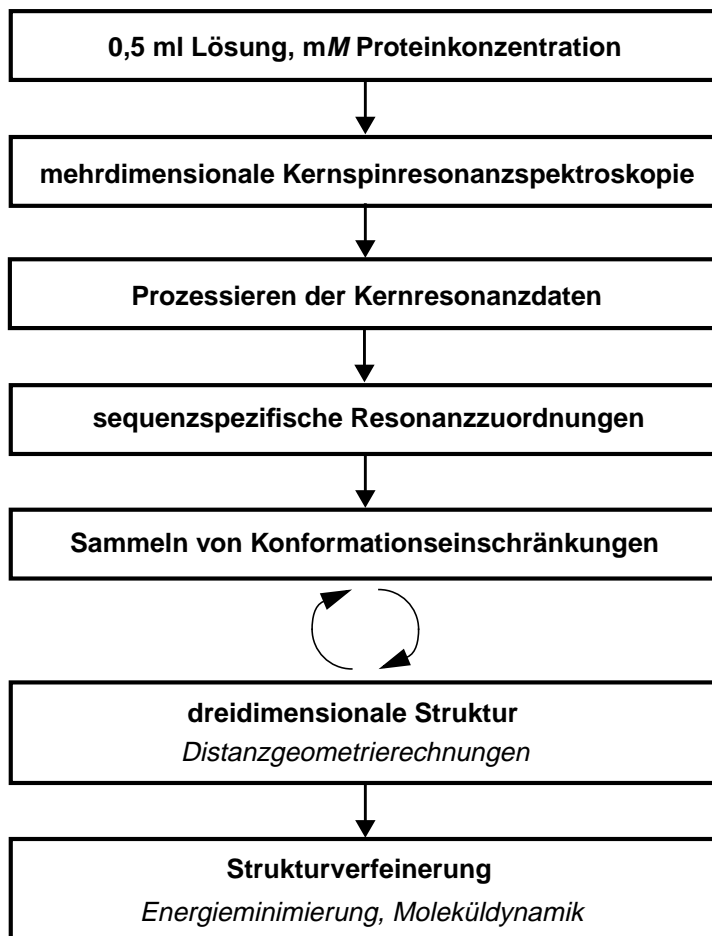


Bild 3. Ablauf einer Bestimmung der dreidimensionalen Struktur eines Proteins in Lösung mit Hilfe der magnetischen Kernresonanzspektroskopie (nach Wüthrich *et al.*, 1993).

wertung beginnt mit der sequentiellen Resonanzzuordnung (Wüthrich *et al.*, 1982), die einen eindeutigen Zusammenhang zwischen Resonanzfrequenzen im Spektrum und einzelnen Atomen im Proteinmolekül herstellt. Auf dieser Grundlage wird es möglich, geometrische Einschränkungen der Konformation zu sammeln, mit deren Hilfe anschließend die dreidimensionale Struktur des Proteins berechnet werden kann. Das Sammeln von Konformationseinschränkungen und die Strukturrechnung bilden einen iterativen Prozeß, der in Bild

4 ausführlicher dargestellt ist, in dem die Kenntnis einer vorläufigen dreidimensionalen Struktur die Bestimmung zusätzlicher Einschränkungen erlaubt, welche wiederum genauere dreidimensionale Strukturen liefern, usw. Der letzte Schritt in einer Proteinstrukturermittlung ist die Verfeinerung der aufgrund rein geometrischer Daten berechneten Struktur durch Energieminimierung oder Moleküldynamikrechnungen, die eine Konformationsenergiefunktion verwenden (Allen & Tildesley, 1987; McCammon & Harvey, 1987).

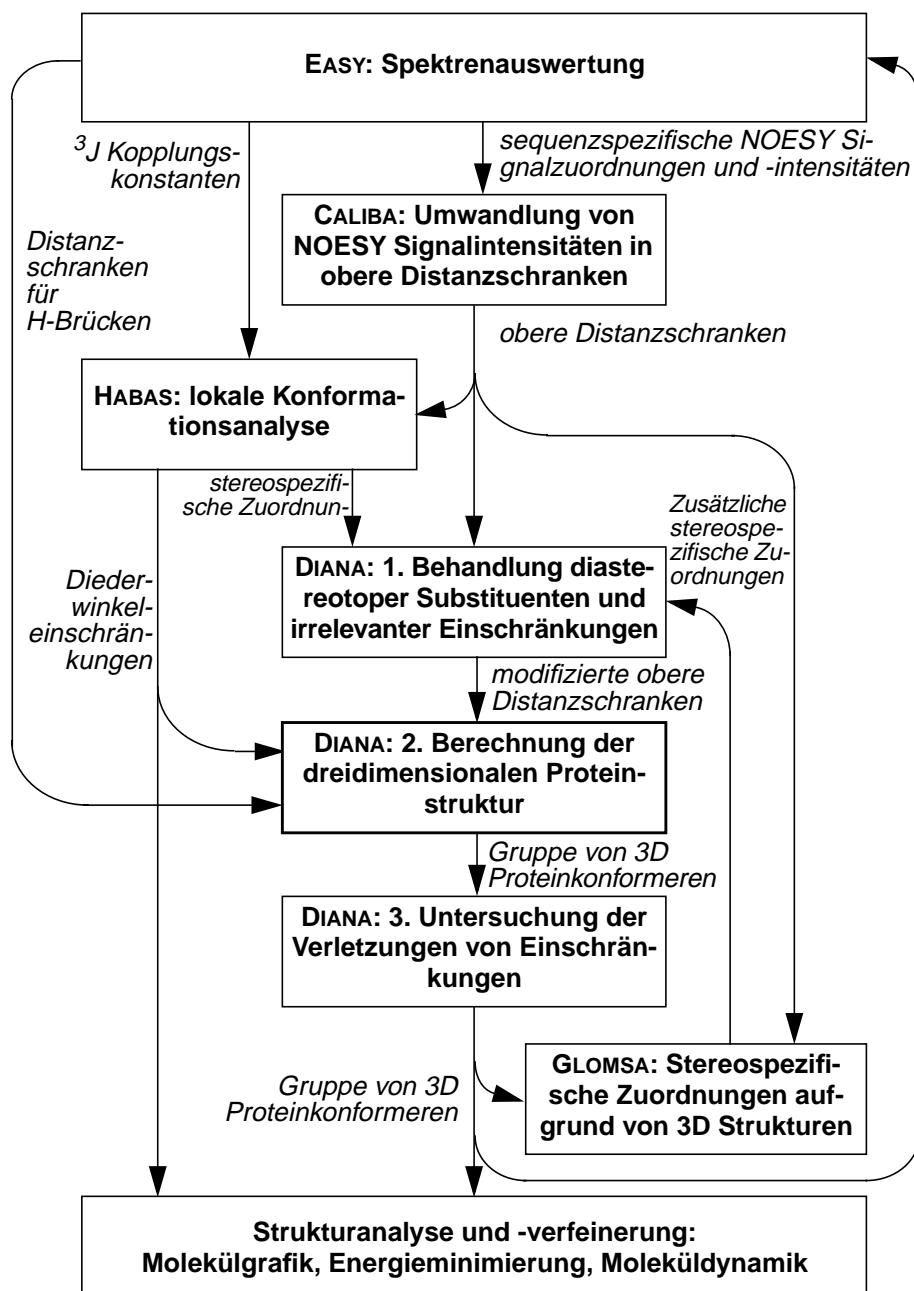


Bild 4. Flußdiagramm für die Berechnung einer dreidimensionalen Proteinstruktur aufgrund von Kernresonanzdaten mit Hilfe des Programms DIANA und Unterstützungsprogrammen (Güntert et al., 1991b).

Eine Strukturrechnung, die die in dieser Abhandlung aufgezeigten Methoden verwendet, läuft wie in Bild 4 gezeigt ab. Die Auswertung der Kernresonanzspektren mit Hilfe des interaktiven Programms EASY (Eccles *et al.*, 1991) liefert sequentielle Resonanzzuordnungen, Intensitäten von Kern-Overhauser-effekten und Werte skalarer Spin-Spin-Kopplungskonstanten, die die Eingabe der Programme HABAS (Güntert *et al.*, 1989) und CALIBA (Güntert *et al.*, 1991a) bilden. In den nächsten Schritten entstehen mit Hilfe des Programms CALIBA aus Intensitäten von Kern-Overhausereffekten obere Schranken für interatomare Distanzen sowie — durch eine Analyse der lokalen Distanzschranken und Spin-Spin-Kopplungskonstanten mit Hilfe des Programms HABAS — Diederwinkelschränkungen und stereospezifische Zuordnungen. Das zentrale Programm DIANA (Güntert *et al.*, 1991a) führt die eigentliche Strukturrechnung durch. Ausgehend von Startkonformationen mit zufälligen Diederwinkelwerten, minimiert das Programm eine Zielfunktion, die ein Maß für die Übereinstimmung der Struktur mit den experimentellen Daten bildet. Falls ein genügend großer, widerspruchsfreier Satz von experimentellen Daten vorliegt, liefert das Programm ein Bündel konvergierter »guter« Konformere, welche weiter verwendet werden. Vor der Strukturrechnung behandelt das Programm DIANA Distanzschranken zu Gruppen von diastereotopen Substituenten ohne stereospezifische Zuordnung so, daß sie mit beiden möglichen stereospezifischen Zuordnungen verträglich werden, und streicht irrelevante, d.h. die Konformation nicht einschränkende Distanzschranken. Nach der Strukturrechnung analysiert es die berechneten Konformere, indem es Verletzungen der Einschränkungen, Wasserstoffbrücken und Abweichungen der Atompositionen aufzeigt. Die berechnete, vorläufige Struktur kann auf zwei Arten iterativ der Verbesserung der Datenauswertung dienen: einerseits bestimmt das Programm GLOMSA (Güntert *et al.*, 1991a) durch einen Vergleich der konvergierten Konformere mit den experimentellen oberen Distanzschranken zusätzliche stereospezifische Zuordnungen, andererseits kann das Programm ASNO (Berndt *et al.*, 1993) Mehrdeutigkeiten bei der Zuordnung von Kern-Overhausereffekten auflösen und dadurch — via die Programme EASY und CALIBA — die Anzahl oberer Distanzschranken für die Strukturrechnung erhöhen. Um eine qualitativ hochwertige Lösungsstruktur eines Proteins zu erhalten, sind im allgemeinen mehrere Strukturrechnungen mit dem Programm DIANA und schrittweise verfeinerten Einschränkungsdatensätzen nötig. Die im letzten DIANA-Lauf ausgewählten »guten« Konformere werden durch Energieminimierung oder in Moleküldynamikrechnungen verfeinert und stellen die Lösungsstruktur des Proteins dar.

Produktoperatorrechnungen in Mathematica

Für ein schwach gekoppeltes Spinsystem läßt sich die Zeitevolution der Dichtematrix σ unter dem Einfluß eines zeitunabhängigen Hamiltonoperators H (Messiah, 1979, Bd. 1, S. 296ff.),

$$\sigma(t + \tau) = e^{-iH\tau} \sigma(t) e^{iH\tau} \quad (\hbar \equiv 1), \quad [1]$$

in einem gepulsten Kernresonanzexperiment analytisch berechnen, wenn man den Hamiltonoperator H und die Dichtematrix σ in der Basis der kartesischen Produktoperatoren ausdrückt (Sørensen *et al.*, 1983; Ernst *et al.*, 1987). Die Hamiltonoperatoren H_f für die freie Präzession unter der Wirkung chemischer Verschiebungen $\omega_k \omega_b \dots$ und skalarer Spin-Spin Kopplungen $J_{mn} J_{pq} \dots$ und H_p für einen Radiofrequenzpuls mit Drehwinkel β und Phase ϕ sind durch

$$H_f = \sum_k \omega_k I_{kz} + \sum_{k < l} 2\pi J_{kl} I_{kz} I_{lz} \quad [2]$$

und

$$H_p \tau = \beta \sum_k [I_{kx} \cos \phi + I_{ky} \sin \phi] \quad [3]$$

gegeben. $I_{k\alpha}$ bezeichnet einen Spinoperator für die α -Komponente des Spins k ($\alpha = x, y, z; k = 1, 2, \dots$). Weil sowohl die Terme von H_f als auch H_p untereinander vertauschen (nicht aber H_f mit H_p), reduzieren sich Berechnungen im Rahmen des Produktoperatorformalismus' auf die wiederholte Anwendung einfacher Regeln, welche die Zeitentwicklung von Produktoperatoren unter dem Einfluß der einzelnen Terme der Hamiltonoperatoren [2] und [3] beschreiben und ohne weiteres aus Gl. [1] und den Eigenschaften von Spin-1/2 Operatoren (Messiah, 1979, Bd. 2, S. 49f.) folgen. Dabei wird die Relaxation (Abragam, 1961; McConnell, 1987) vernachlässigt und eine unendlich kurze Pulsdauer angenommen. Nach dem Schema

Wechselwirkung: $H\tau$

$\sigma(t) \rightarrow \sigma(t + \tau)$
--

faßt die folgende Tabelle die Regeln für Spin-1/2 Teilchen von Sørensen *et al.* (1983) zusammen, wobei Spinoperatoren, für die keine Regel angegeben ist, wie zum Beispiel I_{kz} bei chemischen Verschiebungen, unverändert bleiben und im Fall der skalaren Spin-Spin Kopplung die Regeln für die Produktoperatoren $2I_{k\alpha}I_{k\beta}$ Vorrang vor den Regeln für einzelne Spinoperatoren haben.

Produktoperatorregeln für Spin 1/2

Chem. Verschiebung: $\omega_k \tau I_{kz}$		skalare Spin-Spin Kopplung: $2\pi J_{kl} \tau I_{kz} I_{lz}$	
$I_{kx} \rightarrow I_{kx} \cos \omega_k \tau + I_{ky} \sin \omega_k \tau$		$2I_{k\alpha}I_{l\beta} \rightarrow 2I_{k\alpha}I_{l\beta} \quad (\alpha, \beta \neq z)$	
$I_{ky} \rightarrow I_{ky} \cos \omega_k \tau - I_{kx} \sin \omega_k \tau$		$2I_{kx}I_{lz} \rightarrow 2I_{kx}I_{lz} \cos \pi J_{kl} \tau + I_{ky} \sin \pi J_{kl} \tau$	
		$2I_{ky}I_{lz} \rightarrow 2I_{ky}I_{lz} \cos \pi J_{kl} \tau - I_{kx} \sin \pi J_{kl} \tau$	
		$I_{kx} \rightarrow I_{kx} \cos \pi J_{kl} \tau + 2I_{ky}I_{lz} \sin \pi J_{kl} \tau$	
		$I_{ky} \rightarrow I_{ky} \cos \pi J_{kl} \tau - 2I_{kx}I_{lz} \sin \pi J_{kl} \tau$	
Puls mit Drehwinkel β und Phase ϕ : $\beta [I_{kx} \cos \phi + I_{ky} \sin \phi]$			
$I_{kz} \rightarrow I_{kz} \cos \beta + I_{kx} \sin \beta \sin \phi - I_{ky} \sin \beta \cos \phi$			
$I_{kx} \rightarrow -I_{kz} \sin \beta \sin \phi + I_{kx} (\cos \beta \sin^2 \phi + \cos^2 \phi) + I_{ky} \sin^2 (\beta/2) \sin 2\phi$			
$I_{ky} \rightarrow I_{kz} \sin \beta \cos \phi + I_{kx} \sin^2 (\beta/2) \sin 2\phi + I_{ky} (\cos \beta \cos^2 \phi + \sin^2 \phi)$			
x -Puls: βI_{kx}	y -Puls: βI_{ky}	z -Puls: βI_{kz}	
$I_{kz} \rightarrow I_{kz} \cos \beta - I_{ky} \sin \beta$	$I_{kz} \rightarrow I_{kz} \cos \beta + I_{kx} \sin \beta$	$I_{kx} \rightarrow I_{kx} \cos \beta + I_{ky} \sin \beta$	
$I_{ky} \rightarrow I_{ky} \cos \beta + I_{kz} \sin \beta$	$I_{kx} \rightarrow I_{kx} \cos \beta - I_{kz} \sin \beta$	$I_{ky} \rightarrow I_{ky} \cos \beta - I_{kx} \sin \beta$	

Der Produktoperatorformalismus verbindet die exakte quantenmechanische Behandlung mit einer anschaulichen klassischen Beschreibung und war deshalb eine Grundlage für die Entwicklung vieler Kernresonanzexperimente (Ernst *et al.*, 1987; Kessler *et al.*, 1988). Natürlich kann die Zeitentwicklung gemäß den Gl. [1] auch numerisch berechnet werden (Widmer & Wüthrich, 1986), die Anschaulichkeit geht dabei aber weitgehend verloren. Rechnungen im Produktoperatorformalismus sind zwar nicht schwierig, aber mühsam, weil im allgemeinen im Verlauf der Rechnung eine hohe Anzahl von Termen auftritt. Unterstützung durch einen leistungsfähigen Rechner ist daher sehr erwünscht.

Vor kurzem erschien eine teilweise Implementation des Produktoperator-

formalismus' in *Mathematica* (Shriver, 1991), einer speziell für symbolische Berechnungen geschaffenen Programmiersprache (Wolfram, 1988). Den vollständigen Satz von Regeln für die Zeitentwicklung von Produktoperatoren für Spin-1/2 Kerne enthält das flexible *Mathematica* Programm POMA (»Produktoperatorformalismus in *Mathematica*«; Güntert *et al.*, 1993), das beliebige, schwach gekoppelte Spinsysteme unter dem Einfluß freier Präzession, selektiver und unselektiver Pulse behandeln kann. Phasenzyklen (Ernst *et al.*, 1987) sind ebenfalls vorgesehen. Die Komplexität der Pulssequenzen und Spinsysteme, die untersucht werden können, ist in der Praxis nur durch die verfügbare Rechenleistung begrenzt. Ein Beispiel einer komplexen heteronuklearen Pulssequenz, die mit Hilfe von POMA entworfen wurde, geben Szyperski *et al.* (1993b).

Der folgende, vollständige Programmtext von POMA läuft unter den *Mathematica* Versionen 1.2 und 2.0, allerdings ist die Rechengeschwindigkeit etwa 2,5 mal niedriger mit der Version 2.0. Die Konventionen von Sørensen *et al.* (1983) wurden verwendet, mit der Ausnahme, daß Normierungsfaktoren vor Produktoperatoren weggelassen wurden.¹ Gleichungsnummern, zum Beispiel (*13*), beziehen sich auf die Veröffentlichung von Sørensen *et al.* (1983).

(* Freie Präzession ----- *)

```
MemberQ[allspins,_] ^= True
delay[t_,couples_:{},spins_:allspins][sigma_] := Block[{r,c},
  WriteString["stdout", "delay[" ,InputForm[t],",",couples,",",spins,"]"];
  r = Timing[sigma // shift[t,spins] // coupling[t,couples] // simplifysigma // normalize];(*13*)
  c = r[[1]] /. Second->1; Print[" "]; Print[" (" ,terms[r[[2]]], " terms, ",N[c,3], " s CPU time)"];
  r[[2]]]
```

(* Chemische Verschiebungen ----- *)

```
shift[t_,spins_:allspins][sigma_] := sigma /. {
  spin[n_,x] -> spin[n,x] Cos[w[n]t] + spin[n,y] Sin[w[n]t] /; MemberQ[spins,n],(*14*)
  spin[n_,y] -> spin[n,y] Cos[w[n]t] - spin[n,x] Sin[w[n]t] /; MemberQ[spins,n]}(*15*)
```

(* Schwache skalare Spin-Spin Kopplungen ----- *)

```
Attributes[j] = {Orderless}
coupling[t_,couples_:{}}[sigma_] := sigma /; couples=={}
coupling[t_,{{m1_,m2_}}][sigma_] := normalize[sigma] /. {
  spin[n1_,k] spin[n2_,l] -> spin[n1,k] spin[n2,l] /;
  MemberQ[{x,y},k] && MemberQ[{x,y},l] && {m1,m2}=={n1,n2},
  spin[n1_,x] spin[n2_,z] -> spin[n1,x] spin[n2,z] Cos[Pi j[m1,m2]t] +
  spin[n1,y] Sin[Pi j[m1,m2]t] /; {m1,m2}=={n1,n2} || {m1,m2}=={n2,n1},(*19*)
```

¹ So ist zum Beispiel der Produktoperator für »longitudinale Zweispinordnung« $2I_{1z}I_{2z}$ bei Sørensen *et al.* (1983) und in der Tabelle auf Seite 19, aber $I_{1z}I_{2z}$ in POMA.

```

spin[n1_,y] spin[n2_,z] := spin[n1,y] spin[n2,z] Cos[Pi j[m1,m2]t] -
  spin[n1,x] Sin[Pi j[m1,m2]t] /; {m1,m2}=={n1,n2} || {m1,m2}=={n2,n1},(*20*)
spin[n_,x] := spin[n,x] Cos[Pi j[m1,m2]t] + spin[n,y] spin[m2,z] Sin[Pi j[m1,m2]t] /;
  m1==n,(*17*)
spin[n_,x] := spin[n,x] Cos[Pi j[m1,m2]t] + spin[n,y] spin[m1,z] Sin[Pi j[m1,m2]t] /; m2==n,
spin[n_,y] := spin[n,y] Cos[Pi j[m1,m2]t] - spin[n,x] spin[m2,z] Sin[Pi j[m1,m2]t] /;
  m1==n,(*18*)
spin[n_,y] := spin[n,y] Cos[Pi j[m1,m2]t] - spin[n,x] spin[m1,z] Sin[Pi j[m1,m2]t] /; m2==n}
coupling[t_,couples_][sigma_] := sigma // coupling[t,{First[couples]}] // coupling[t,Rest[couples]]

```

(* Radiofrequenzpulse ----- *)

```

pulse[b_,p_,spins_:allspins][sigma_] := Block[{r,c},
  WriteString["stdout","pulse[\",InputForm[b],\",\",InputForm[p],\",\",spins,\"]"];
  r = Timing[normalize[pulses[b,p,spins][sigma]]];
  c = r[[1]] /. Second->1; Print[""]; Print[" (\",terms[r[[2]],\", terms, \",N[c,3],\", s CPU time)\"];
  r[[2]]]
pulses[b_,p_List,spins_][sigma_List] := Block[{i},
  Table[pulses[b,p[[i]],spins][sigma[[i]],{i,1,Length[p]}]]
pulses[b_,p_List,spins_][sigma_] := Block[{i},
  Table[pulses[b,p[[i]],spins][sigma],{i,1,Length[p]}]]
pulses[b_, p_,spins_] := pulses[b Pi/180,p,spins] /; NumberQ[b] && b!=0
pulses[b_, p_,spins_] := pulses[b,p Pi/180,spins] /; NumberQ[p] && p!=0
pulses[b_,-p_,spins_] := pulses[-b,p,spins]
pulses[b_, p_,spins_][sigma_] := sigma /. {
  spin[n_,k_] := spin[n,k] /; p==k,
  spin[n_,z] := spin[n,z] Cos[b] - spin[n,y] Sin[b] /; p==x && MemberQ[spins,n],(*24*)
  spin[n_,y] := spin[n,y] Cos[b] + spin[n,z] Sin[b] /; p==x && MemberQ[spins,n],(*25*)
  spin[n_,z] := spin[n,z] Cos[b] + spin[n,x] Sin[b] /; p==y && MemberQ[spins,n], (*26*)
  spin[n_,x] := spin[n,x] Cos[b] - spin[n,z] Sin[b] /; p==y && MemberQ[spins,n], (*27*)
  spin[n_,x] := spin[n,x] Cos[b] + spin[n,y] Sin[b] /; p==z && MemberQ[spins,n], (*28*)
  spin[n_,y] := spin[n,y] Cos[b] - spin[n,x] Sin[b] /; p==z && MemberQ[spins,n], (*29*)
  spin[n_,z] := spin[n,z] Cos[b] + spin[n,x] Sin[b] Sin[p] -
    spin[n,y] Sin[b] Cos[p] /; MemberQ[spins,n],(*35*)
  spin[n_,x] := -spin[n,z] Sin[b] Sin[p] + spin[n,x] (Cos[b] Sin[p]^2 + Cos[p]^2) +
    spin[n,y] Sin[b/2]^2 Sin[2p] /; MemberQ[spins,n],(*36*)
  spin[n_,y] := spin[n,z] Sin[b] Cos[p] + spin[n,x] Sin[b/2]^2 Sin[2p] +
    spin[n,y] (Cos[b] Cos[p]^2 + Sin[p]^2) /; MemberQ[spins,n](*)37*)
} /. {Cos[-u_] := Cos[u], Sin[-u_] := -Sin[u]}

```

(* Empfänger ----- *)

```

receiver[p_,spins_:allspins][sigma_] := Block[{r,c},
  WriteString["stdout","receiver[\",InputForm[p],\",\",spins,\"]"];
  r = Timing[normalize[receivers[p,spins][sigma]]];
  c = r[[1]] /. Second->1; Print[""]; Print[" (\",terms[r[[2]],\", terms, \",N[c,3],\", s CPU time)\"];
  r[[2]]]
receivers[p_List,spins_][sigma_List] := Block[{i},
  Sum[receivers[p[[i]],spins][sigma[[i]],{i,1,Length[p]}]/Length[p]]
receivers[p_List,spins_][sigma_] := Block[{i},
  Sum[receivers[p[[i]],spins][sigma],{i,1,Length[p]}]/Length[p]]
receivers[p_,spins_][sigma_List] := Block[{i},
  Sum[receivers[p,spins][sigma[[i]],{i,1,Length[sigma]}]/Length[sigma]]
receivers[x,spins_][sigma_] := sigma

```

```

receivers[-x,spins_][sigma_] := sigma // pulses[180,z,spins]
receivers[y,spins_][sigma_] := sigma // pulses[-90,z,spins]
receivers[-y,spins_][sigma_] := sigma // pulses[90,z,spins]
receivers[p_,spins_][sigma_] := sigma // pulses[-p,z,spins]

```

(* Beobachtbare Terme ----- *)

```

transverse[k_] := MemberQ[{x,y},k]
tfree[u_] := !MatchQ[u,a_. spin[_]?transverse]]
observable[sigma_] := observable[allspins][sigma] /; !FreeQ[sigma,spin[_]] || sigma==0
observable[spins_][sigma_] := Block[{r,c},WriteString["stdout","observable[" ,spins,"]"];
  r = Timing[normalize[observables[spins][sigma]];
  c = r[[1]] /. Second->1; Print[" "]; Print[" (" ,terms[r[[2]]], " terms, ",N[c,3], " s CPU time)"];
  r[[2]]]
observables[spins_][0] := 0
observables[spins_][sigma_List]:=Block[{i},
  Table[Apply[Plus,observables[spins][sigma[[i]]],{i,1,Length[sigma]}]]
observables[spins_][sigma_] := Block[{detected,qloc},detected[n_] := MemberQ[spins,n];
  Cases[Apply[List,1+normalize[qloc sigma]],a_?tfree spin[n_?detected,k_?transverse]]
  /. qloc->1 // trigsimplify]

```

(* Dichtematrix auf Standardform bringen ----- *)

```

spinlist[sigma_] := sigma // If[Head[#]==Plus,Apply[List,#],{#}]& //
  Map[If[Head[#]==Times,Apply[List,#],{#}]&,#]& //
  Map[Cases[#,spin[_]]&,#]& // Map[{# /. List->Times}&,#]& // Union
normalize[sigma_List] := Map[normalize,sigma]
normalize[sigma_] := Block[{e,p}, e=Expand[sigma]; p=spinlist[e]; WriteString["stdout","."];
  (Map[Coefficient[e,#]&,p] /. spin[_] -> 0).p]

```

(* Trigonometrische Vereinfachungen ----- *)

```

trigsimplify[u_] := Expand[u] //. {
  c_. Sin[a_]^2 + c_. Cos[a_]^2 := c,
  c_. Cos[a_]^2 + d_. Sin[a_]^2 := c Cos[2a] /; c+d==0,
  Sin[a_] Cos[a_] := 1/2 Sin[2 a],
  c_. Sin[a_] Cos[b_] + c_. Cos[a_] Sin[b_] := c Sin[a+b],
  c_. Sin[a_] Cos[b_] + d_. Cos[a_] Sin[b_] := c Sin[a-b] /; c+d == 0,
  c_. Cos[a_] Cos[b_] + d_. Sin[a_] Sin[b_] := c Cos[a+b] /; c+d == 0,
  c_. Cos[a_] Cos[b_] + c_. Sin[a_] Sin[b_] := c Cos[a-b]}
simplifysigma[sigma_List] := Map[simplifysigma,sigma]
simplifysigma[sigma_] := Apply[Plus,trigsimplify[Apply[List,1+normalize[sigma]]]]-1

```

(* Ausgabeformate ----- *)

```

nucleus[_] = I
terms[sigma_List] := Block[{i},Sum[terms[sigma[[i]]],{i,1,Length[sigma]}]]
terms[sigma_] := Length[1+sigma]-1
spinfree[u_] := FreeQ[u,spin[_]]
spinop[u_] := !FreeQ[u,spin[_]]
times[u_] := FreeQ[u,j[_]] && FreeQ[u,w[_]] && FreeQ[u,_Integer] &&
  FreeQ[u,_Real] && FreeQ[u,Pi]

```



```

sort[sigma_] := (Print[TableForm[
  Complement[Apply[List, 1 + Apply[Plus, sigma /. w[n_]->Aw[n]] // {
    Sin[a_?times b_] := AA[a,b,Sin], Cos[a_?times b_] := AA[a,b,Cos],
    a_Times + b_Times := Select[a,spinfree] (Select[a,spinop] + Select[b,spinop]) /;
    Select[a,spinfree]==Select[b,spinfree]},{1}]]];
sigma;)
Format[AA[a_,b_Plus,c_] := SequenceForm[c,“(b,”) “a,”)”]
Format[AA[a_,1,c_] := SequenceForm[c,“(a,”)”]
Format[AA[a_,b_,c_] := SequenceForm[c,“(b,”) “a,”)”]
Format[Aw[n_] := SequenceForm[“w”,n]
Format[w[n_] := SequenceForm[“w”,n]
Format[j[n1_,n2_] := SequenceForm[“J”,n1,n2]
Format[spin[n_,k_] := SequenceForm[nucleus[n],n,k]

```

Es ist offensichtlich, wie elegant und kompakt sich die Regeln für Produktoperatoren in *Mathematica* formulieren lassen. Außerdem können die Ergebnisse mit den leistungsfähigen *Mathematica*-Werkzeugen für die Manipulation mathematischer Ausdrücke weiter behandelt werden.

In der POMA Implementation des Produktoperatorformalismus' hat ein Spinoperator $I_{k\alpha}$ für die α -Komponente des Spins k ($\alpha = x, y, z; k = 1, 2, \dots$) in der Eingabe die Form **spin**[k, α] und in der Ausgabe $X_{k\alpha}$, wobei der Buchstabe X mit der Anweisung **nucleus**[k] = X festgelegt werden kann. Freie Präzession der Zeitdauer t , während der die Spins in der Menge $\{k, l, \dots\}$ unter dem Einfluß chemischer Verschiebungen mit den Kreisfrequenzen $\omega_k, \omega_l, \dots$ stehen und skalare Spin-Spin Kopplungen J_{mn}, J_{pq}, \dots zwischen den Spinpaaren in der Menge $\{mn, pq, \dots\}$ bestehen, wird mit der Anweisung

$$\mathbf{delay}[t, \{mn, pq, \dots\}, \{k, l, \dots\}] \quad [4]$$

eingegeben. Falls die Angabe der Mengen $\{k, l, \dots\}$ und $\{mn, pq, \dots\}$ fehlt, wird angenommen, daß alle Spins unter der Wirkung chemischer Verschiebungen stehen und keine skalaren Kopplungen vorhanden sind.

Einem Radiofrequenzpuls mit einem Drehwinkel β , der auf die Spins in der Menge $\{k, l, \dots\}$ wirkt und im Phasenzklus die Phasenwerte $\{\phi_1, \phi_2, \dots\}$ annimmt, entspricht die Anweisung

$$\mathbf{pulse}[\beta, \{\phi_1, \phi_2, \dots\}, \{k, l, \dots\}]. \quad [5]$$

Wird für einen Dreh- oder Phasenwinkel eine reine Zahl gegeben, so wird der Wert in Grad, sonst in Radiant interpretiert; den speziellen Phasenwerten $\phi = x, y, -x, -y$ entsprechen der Reihe nach Winkel von 0, 90, 180 und 270°. Falls die Phase in allen Schritten des Phasenzklus' gleich ist — $\{\phi, \phi, \dots\}$ — so kann die Phasenliste durch eine einzige Phase ϕ abgekürzt werden. Wenn die Menge $\{k, l, \dots\}$ nicht angegeben wird, ist der Puls nicht-selektiv und wirkt auf alle Spins. Zum Beispiel kann ein »harter« 90° x -Puls in Kurzform als **pul-**

se[90, x] geschrieben werden. Der Empfängerphasenzyklus $\{\psi_1, \psi_2, \dots\}$ wird mit der Anweisung

receiver $\{\{\psi_1, \psi_2, \dots\}\}$ [6]

festgelegt. Eine Empfängerphase $\psi \neq x$ wird durch einen Puls mit Drehwinkel $\Delta\psi$ um die z -Achse vor der Signalaufnahme simuliert, wobei $\Delta\psi$ die Phasendifferenz bezüglich der x -Achse bezeichnet.

Das Programm arbeitet die Pulssequenz für jeden einzelnen Schritt des Phasenzyklus' ab, zählt die Ergebnisse zusammen und normiert die Summe, indem es sie durch die Länge des Phasenzyklus' teilt. Das Ergebnis kann mit Hilfe der Anweisung

observable $\{\{k, l, \dots\}\}$ [7]

nach »beobachtbaren Termen« durchsucht werden, wobei $\{k, l, \dots\}$ die Menge der transversalen Spins, die während der Aufnahme interessieren, bezeichnet.

POMA Anweisungen

spin $[k, \alpha]$	Spinoperator $I_{k\alpha}$ ($\alpha = x, y, z; k = 1, 2, \dots$)
nucleus $[k] = X$	Bezeichnung von Spin k auf X setzen
delay $[t, \{mn, pq, \dots\}]$	freie Präzession der Zeitdauer t mit skalaren Kopplungen J_{mn}, J_{pq}, \dots
delay $[t, \{mn, pq, \dots\}, \{k, l, \dots\}]$	freie Präzession mit chemischen Verschiebungen nur für die Spins $\{k, l, \dots\}$
pulse $[\beta, \phi]$	nicht-selektiver Radiofrequenzpuls mit Drehwinkel β und Phase ϕ
pulse $[\beta, \{\phi_1, \phi_2, \dots\}]$	nicht-selektiver Radiofrequenzpuls mit Drehwinkel β und Phasenzyklus $\{\phi_1, \phi_2, \dots\}$
pulse $[\beta, \{\phi_1, \phi_2, \dots\}, \{k, l, \dots\}]$	Selektiver Radiofrequenzpuls für die Spins $\{k, l, \dots\}$
receiver $\{\{\psi_1, \psi_2, \dots\}\}$	Empfängerphasenzyklus $\{\psi_1, \psi_2, \dots\}$
observable	alle beobachtbaren Terme
observable $\{\{k, l, \dots\}\}$	beobachtbare Terme der Spins $\{k, l, \dots\}$
sort	sortieren gemäß Sørensen <i>et al.</i> (1983)

Schließlich ersetzt die Anweisung **sort** die in *Mathematica* normale, alphabetische Reihenfolge der Terme durch die Konvention von Sørensen *et al.* (1983), d. h. die Reihenfolge »chemische Verschiebungsterme — skalare Kopplungsterme — Spinoperatoren«.

Für die Simulation eines gepulsten Kernresonanzexperiments verknüpft man die entsprechenden Anweisungen in der Reihenfolge der Pulssequenz durch den *Mathematica* Postfixoperator »//«¹. Eine Pulssequenz

$$\beta_1 - t_1 - \beta_2 - \dots - \beta_n - \text{Aufnahme} \quad [8]$$

wird in POMA durch

$$\begin{aligned} &\mathbf{pulse}[\beta_1, \dots] // \mathbf{delay}[t_1, \dots] // \mathbf{pulse}[\beta_2, \dots] // \dots \\ & // \mathbf{pulse}[\beta_n, \dots] // \mathbf{receiver}[\dots] \end{aligned} \quad [9]$$

dargestellt. Im allgemeinen wird eine Pulssequenz auf den thermodynamischen Gleichgewichtszustand des Spinsystems angewandt, d. h. auf I_{kz} oder, in *Mathematica* Eingabeformat, **spin**[*k*, *z*], und man behält am Ende lediglich die beobachtbaren Terme, indem man den **observable** Operator anwendet.

Die Pulssequenz des 3-quantengefilterten COSY Experiments (Müller *et al.*, 1986) ist ein einfaches Beispiel für die Anwendung von POMA:

$$90^\circ_\phi - t_1 - 90^\circ_\phi - 90^\circ_x - \text{Aufnahme}_\psi \quad [10]$$

Der Phasenzklus ist $\phi = 0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3$ und $\psi = 3(x, -x)$. In POMA wird die Pulssequenz durch die folgende Reihe von Anweisungen dargestellt:

$$\begin{aligned} &\mathbf{spin}[1, z] // \mathbf{pulse}[90, \{0, \text{Pi}/3, 2\text{Pi}/3, \text{Pi}, 4\text{Pi}/3, 5\text{Pi}/3\}] // \\ & \quad \mathbf{delay}[t1, \{\{1, 2\}, \{1, 3\}\}] // \\ & \quad \mathbf{pulse}[90, \{0, \text{Pi}/3, 2\text{Pi}/3, \text{Pi}, 4\text{Pi}/3, 5\text{Pi}/3\}] // \\ & \quad \mathbf{pulse}[90, x] // \\ & \quad \mathbf{receiver}[\{x, -x, x, -x, x, -x\}] // \\ & \quad \mathbf{observable} // \mathbf{Simplify} // \mathbf{sort} \end{aligned} \quad [11]$$

Das Experiment beginnt also im thermodynamischen Gleichgewicht, alle Pulse sind nicht-selektive 90°-Pulse, und während der Entwicklungszeit t_1 ist Spin 1 mit den Spins 2 und 3 skalar gekoppelt, nicht aber Spin 2 mit Spin 3. **Simplify** ist ein eingebauter *Mathematica* Operator für algebraische Vereinfachungen.

¹ »x// f« besitzt in *Mathematica* die Bedeutung: »Wende den Operator *f* auf den Ausdruck *x* an«, und ist gleichbedeutend mit *f*[*x*].

Für die Pulssequenz [11] liefert POMA nach 12 s Rechenzeit auf einem Sun-4 Rechner mit der *Mathematica* Version 1.2 das Ergebnis

$$-\left(\sin[\omega_1 t_1] \sin[\pi J_{12} t_1] \sin[\pi J_{13} t_1] \right. \\ \left. (I_{1z} I_{2z} I_{3x} + I_{1z} I_{2x} I_{3z} + I_{1x} I_{2z} I_{3z})\right)/4. \quad [12]$$

In der Schreibweise von Sørensen *et al.* (1983) entspricht dies

$$-\sin(\omega_1 t_1) \sin(\pi J_{12} t_1) \sin(\pi J_{13} t_1) (I_{1z} I_{2z} I_{3x} + I_{1z} I_{2x} I_{3z} + I_{1x} I_{2z} I_{3z})/4. [13]$$

Obschon das Ergebnis einfach ist, entstanden nach dem zweiten Puls vorübergehend 124 Terme, weshalb die Rechnung »von Hand« recht mühsam wäre.

Verarbeitung von Kernresonanzdaten mit dem Programm PROSA

Ein gepulstes mehrdimensionales Kernspinresonanzexperiment zeichnet ein Signal in Abhängigkeit von mehreren Zeitkoordinaten auf. Ausgewertet werden aber mehrdimensionale Spektren, d. h. Darstellungen der Signale im Frequenzbereich (Ernst *et al.*, 1987). Die Berechnung der Spektren — der Wechsel vom Zeit- in den Frequenzbereich — erfolgt normalerweise aus Gründen der Robustheit und Effizienz mit Hilfe der diskreten Fourier-Transformation (Press *et al.*, 1986), obwohl auch andere Methoden verfügbar sind (Marple, 1987; Stephenson, 1988). Hilfsverfahren verbessern das Aussehen der Spektren und erleichtern ihre Auswertung, indem sie Artefakte, die den experimentellen Daten innewohnen oder durch die diskrete Fourier-Transformation entstehen können, unterdrücken oder verringern. Dazu gehören unter anderem: lineare Vorhersage von Datenpunkten im Zeitbereich und Einsatz von Fensterfunktionen, um das abrupte Abschneiden der Zeitbereichsdaten am Ende des Meßintervalls zu mildern, Phasenkorrektur, um Spektren mit rein absorptiven Signalen zu gewinnen, Unterdrückung von Restsignalen des Lösungsmittels, und Basislinienkorrektur zum Glätten der Basislinie.

Dieses Kapitel behandelt das Programm PROSA (Güntert *et al.*, 1992), das die üblichen Verfahren zur Prozessierung von Kernresonanzdaten effizient und flexibel zur Verfügung stellt. Die hohe Verarbeitungsgeschwindigkeit des Programms PROSA beruht darauf, daß keine Zwischenergebnisse extern abgespeichert oder Grafikdarstellungen verwendet werden. Letzteres erlaubt auch ohne weiteres, weil das Programm in Standard FORTRAN-77 geschrieben ist, die Verwendung verschiedener Rechnertypen¹, darunter auch Hochleistungsrechner, für die alle zeitaufwendigen Routinen vektorisiert worden sind. Im Gegensatz dazu bedürfen kommerzielle Programme zur Prozessierung von Kernresonanzdaten, wie zum Beispiel FELIX (Hare Research, Inc., 1992) oder SYBYL/NMR TRIAD (TRIPOS Associates, Inc., 1992), einer Grafikarbeitsstation oder gar, wie zum Beispiel UXNMR (Bruker Analytische Meßtechnik GmbH,

¹ Zur Zeit läuft das Programm auf dem NEC SX-3 Hochleistungsrechner des *Centro Svizzero di Calcolo Scientifico*, sowie auf Convex C2, Silicon Graphics Iris und Sun-4 Rechnern.

1991), eines spezialisierten Rechners und sind daher in der Praxis nicht auf Hochleistungsrechnern verwendbar.

Das Vermeiden von Grafikdarstellungen und des externen Speicherns von Zwischenergebnissen vereinfacht den Aufbau des Programms PROSA stark: Zu Beginn liest das Programm alle Zeitbereichsdaten in den Arbeitsspeicher des Rechners, bearbeitet sie, und gibt schließlich das fertige Spektrum aus. Der Arbeitsspeicher des Rechners muß genügend groß sein, um jederzeit den gesamten Datensatz sowie eine (zweidimensionale) Ebene daraus aufnehmen zu können.

Lineare Vorhersage

Das Programm PROSA verwendet die lineare Vorhersage (nicht zu verwechseln mit linearer Extrapolation) von Datenpunkten s_k im Zeitbereich, um die Effekte der diskreten Fourier-Transformation bei abgeschnittenen Zeitbereichsdaten — vor allem Linienverbreiterung und Oszillationen am Rand der Signale im Spektrum — zu mildern (Stephenson, 1988; Olejniczak & Eaton, 1990; Zhu & Bax, 1990). Die lineare Vorhersage ist eine Methode, die auf der Annahme beruht, daß sich der Datenpunkt s_k als Linearkombination der m vorhergehenden Datenpunkte s_{k-m}, \dots, s_{k-1} darstellen läßt:

$$s_k = \sum_{l=1}^m a_l s_{k-l}. \quad [14]$$

Für eine in Zeitschritten Δt abgetastete Überlagerung von höchstens m gedämpften Schwingungen mit Amplituden A_α , Phasen ϕ_α , Frequenzen ω_α und Dämpfungen Γ_α ,

$$s_k = \sum_{\alpha} A_{\alpha} e^{i\phi_{\alpha}} e^{-(\Gamma_{\alpha} + i\omega_{\alpha})k\Delta t}, \quad [15]$$

ist diese Annahme erfüllt, und die Nullstellen z_1, \dots, z_m des Polynoms

$$1 - \sum_{l=1}^m a_l z^l \quad [16]$$

sind mit den Frequenzen und Dämpfungen durch $z_{\alpha} = e^{(\Gamma_{\alpha} + i\omega_{\alpha})\Delta t}$ verknüpft. Die Koeffizienten a_1, \dots, a_m werden durch Anwendung von Gleichung [14] auf die gemessenen Daten mit Hilfe der Singulärwertzerlegung (Press *et al.*, 1986) bestimmt (Kumaresan & Tufts, 1982; Barkhuijsen *et al.*, 1987), wobei die Anzahl m der Koeffizienten genügend hoch in muß, um alle Signale einzuschließen, aber höchstens halb so groß wie die Anzahl gemessener, komplexer Datenpunkte sein darf. Wegen des Rauschens können auch bei exponentiell abfallenden Signalen Nullstellen z_{α} außerhalb des Einheitskreises auftreten,

welche zu einem exponentiellen Anstieg ($\Gamma_a < 0$) des vorhergesagten Signals führen würden. Das Programm PROSA stellt ein stabiles Signal durch Spiegeln der außerhalb liegenden Nullstellen am Einheitskreis sicher: $z_\alpha \rightarrow z_\alpha / |z_\alpha|$ (Press *et al.*, 1986).

Weil nur zweidimensionale Ebenen eines mehrdimensionalen Spektrums auf einem Bildschirm oder auf Papier ohne weiteres dargestellt werden können, erschweren Signale mit breiten oder oszillierenden Linienformen in Richtungen senkrecht zur Betrachtungsebene die Auswertung von drei- und vor allem vierdimensionalen Spektren beträchtlich. Die lineare Vorhersage, welche zu schmaleren und glatteren Linienformen führt, kann in diesen Fällen die Interpretation der Spektren wesentlich erleichtern.

Als Beispiel zeigt Bild 5 Gebiete eines vierdimensionalen [^{13}C - ^{13}C]-korrelierten [^1H - ^1H]-NOESY Spektrums (Clare *et al.*, 1991; Zuiderweg *et al.*, 1991) des Komplexes, der vom gleichförmig ^{13}C -markierten Protein Cyclophilin und dem unmarkierten Immunosuppressivum Cyclosporin A gebildet wird. Die

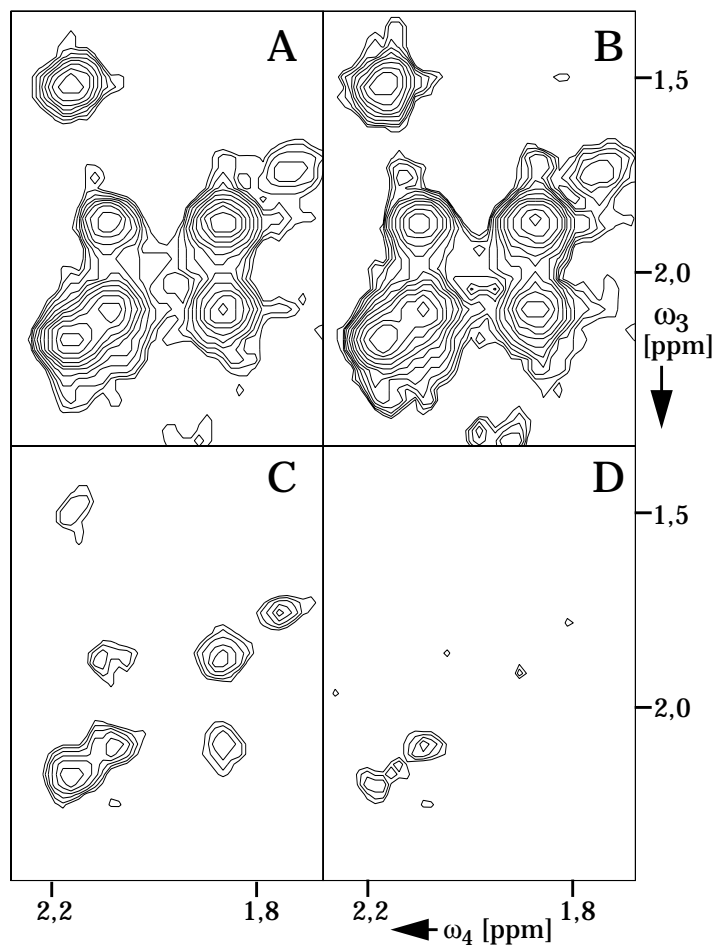


Bild 5. Auswirkungen der linearen Vorhersage, illustriert anhand des Gebiets $\omega_3(^1\text{H}) = 1,34\text{--}2,44$ ppm, $\omega_4(^1\text{H}) = 1,68\text{--}2,29$ ppm aus einem 4D [^{13}C - ^{13}C]-korrelierten [^1H - ^1H]-NOESY Spektrum des Komplexes zwischen gleichförmig ^{13}C -markiertem Cyclophilin und Cyclosporin A. (A) Die [^1H - ^1H]-Ebene bei $\omega_1(^{13}\text{C}) = \omega_2(^{13}\text{C}) = 40,5$ ppm im ohne lineare Vorhersage erhaltenen Spektrum. (B) Dieselbe Ebene wie in A im mit linearer Vorhersage erhaltenen Spektrum. (C) Die [^1H - ^1H]-Ebene bei $\omega_1(^{13}\text{C}) = 39,4$ ppm, $\omega_2(^{13}\text{C}) = 41,6$ ppm im ohne lineare Vorhersage erhaltenen Spektrum. (D) Dieselbe Ebene wie in C im mit linearer Vorhersage erhaltenen Spektrum. Die in B und D gezeigten Spektren wurden berechnet wie in der Tabelle auf S. 36 angegeben, die in A und C gezeigten Spektren durch Auffüllen der Zeitbereichsdaten mit Nullen anstatt linearer Vorhersage.

Messungen wurden auf einem Bruker AMX 500 Spektrometer bei einer Konzentration des Komplexes von 1 mM in $^2\text{H}_2\text{O}$ bei $p^2\text{H}$ 6,0 und einer Temperatur von 32°C durchgeführt (C. Spitzfaden, G. Wider & K. Wüthrich, unveröffentlicht). Die vollständige Sequenz von PROSA-Anweisungen zur Berechnung des Spektrums wird in der Tabelle auf S. 36 gegeben. Bild 5A und B zeigen eine $[\text{}^1\text{H}-\text{}^1\text{H}]$ -Region aus der sechsten von jeweils 16 Ebenen entlang ω_1 und ω_2 , die mehrere Signale enthält, deren Maxima bezüglich der beiden ^{13}C Dimensionen in diese Ebene fallen. Nur unwesentliche Abweichungen treten zwischen den Spektren auf, die durch Ergänzen der Zeitbereichsdaten mit Nullen (A) bzw. mit linearer Vorhersage (B) in beiden ^{13}C Dimensionen berechnet wurden. Im Gegensatz dazu enthält die gleiche $[\text{}^1\text{H}-\text{}^1\text{H}]$ -Region aus der siebenten und fünften Ebene entlang ω_1 und ω_2 , d. h. neben den ^{13}C Resonanzfrequenzen der gezeigten Signale, im mit linearer Vorhersage berechneten Spektrum (D) eine bedeutend geringere Zahl von Signalen als sonst (C). Die meisten »Signale« in Bild 5C sind eigentlich Ausläufer der breiten Resonanzlinien, die ihr Maximum in der $[\text{}^1\text{H}-\text{}^1\text{H}]$ -Ebene von Bild 5A haben.

Verzerrungen der Basislinie sind in Kernresonanzexperimenten in erster Linie die Folge von Fehlern in der Messung oder Behandlung der ersten Datenpunkte eines freien Induktionszerfalls (Otting *et al.*, 1986). Sie können in PROSA durch rückwärtsgerichtete lineare Vorhersage unterdrückt werden, die die verfälschten ersten Datenpunkte des Signals wiederherstellt (Marion & Bax, 1989). Im Vergleich zu Basislinienkorrekturmethoden im Frequenzbereich (Pearson, 1977; Dietrich *et al.*, 1991; Güntert & Wüthrich, 1992) bietet diese Methode den Vorteil, daß das Glätten der Basislinie am Anfang anstatt am Ende der Datenprozessierung erfolgt, wodurch andere Prozessierungsschritte, die eine flache Basislinie voraussetzen, bessere Ergebnisse liefern können.

Automatische Phasenkorrektur

Verschiedene Algorithmen zur automatischen Phasenkorrektur von eindimensionalen (Ernst, 1969; Neff *et al.*, 1977; Gladden & Elliott, 1986; Brown *et al.*, 1989; Nelson & Brown, 1989; Heuer, 1991) und zweidimensionalen (Cieslar *et al.*, 1988; Hoffman *et al.*, 1992) Kernresonanzspektren wurden bereits vorgeschlagen. Die neue automatische Phasenkorrekturroutine, die für das Programm PROSA entwickelt wurde (Güntert *et al.*, 1992), bestimmt die konstante (ϕ_0) und lineare (ϕ_1) Phasenkorrektur, indem sie zunächst eindimensionale Schnitte des Absolutwertspektrums nach starken, isolierten Signalen absucht und anschließend die Summe

$$S(\phi_0, \phi_1) = \sum_p (\text{Re} \hat{I}_p)^2 - (\text{Im} \hat{I}_p)^2 \quad [17]$$

über die Differenz der quadrierten Real- und Imaginärteile der normierten Signalintensitäten

$$\hat{I}_p = \frac{I_p}{|I_p|} e^{-i(\phi_0 + \phi_1 \omega_p)} \quad [18]$$

im phasenkorrigierten Spektrum maximiert. I_p bezeichnet das (komplexe) Integral über das Signal p im nicht phasenkorrigierten Spektrum, und ω_p steht für die auf das Intervall $[0, 1]$ normierte Position des Signals p . Das Maximum von $S(\phi_0, \phi_1)$ wird erreicht, wenn die lineare Phasenkorrektur ϕ_1 so gewählt wird, daß die Funktion

$$s(\beta) = \sum_p \left(\frac{I_p}{|I_p|} \right)^2 e^{-2i\beta\omega_p} \quad [19]$$

ihren maximalen Absolutwert bei $\beta = \phi_1$ annimmt und für die konstante Phasenkorrektur gilt

$$\phi_0 = \arg(s(\phi_1)) . \quad [20]$$

Weil im allgemeinen die Funktion $s(\beta)$ mehrere lokale Maxima besitzt, bestimmt PROSA den Wert von ϕ_1 durch Absuchen eines eindimensionalen Gitters mit der Maschenweite 1° . Das phasenkorrigierte Spektrum $\hat{s}_1, \dots, \hat{s}_n$ mit rein absorptiven Linienformen ergibt sich aus dem ursprünglichen Spektrum s_1, \dots, s_n gemäß der Beziehung

$$\hat{s}_k = s_k e^{-i\left(\phi_0 + \phi_1 \frac{k-1}{n-1}\right)} . \quad [21]$$

Das Verfahren kann auch für Spektren, die sowohl positive als auch negative Signale enthalten, gebraucht werden.

Um Signalpositionen in *eindimensionalen* Schnitten des Absolutwertspektrums zu bestimmen, sucht das Programm zunächst alle lokalen Maxima, die mindestens κ Mal über der Rauschintensität des Absolutwertspektrums liegen. Der Wert von κ hängt vom Signal-zu-Rauschen Verhältnis des Spektrums ab. Für jedes so gefundene lokale Maximum wird das Signalgebiet durch die ersten Datenpunkte links und rechts vom Maximum begrenzt, deren Intensität 10% der Maximalintensität oder die zweifache Rauschintensität unterschreitet. Überlappende Signale werden ausgeschlossen, indem die gleiche Bedingung an die Intensität für Gebiete der Breite $u/4$ zur Linken und Rechts des Signalgebiets gestellt wird, wobei u die vorgegebene Maximalbreite des Signalgebiets bezeichnet. Die vorgegebene maximale Anzahl v von Signalen aus verschiedenen eindimensionalen Schnitten, die die gleichen Frequenzkoordinaten beinhalten, verhindert, daß die Phasenkorrektur durch eine Häufung von Signalen in einem schmalen Frequenzband dominiert wird. Das Programm verwendet in einem solchen Fall nur die jeweils v stärksten Signale,

die eine Frequenzkoordinate gemeinsam haben, zur Phasenkorrektur. Üblicherweise werden einige hundert (eindimensionale) Signale für die Phasenkorrektur berücksichtigt; das Verfahren ist dadurch robust gegenüber Instabilitäten, die auftreten können, wenn nur wenige Signale miteinbezogen werden.

Die automatische Phasenkorrektur in PROSA erzeugt rein absorptive Liniensformen zuverlässig, unabhängig von der Anzahl Datenpunkte in der betrachteten Dimension, wie eindimensionale Schnitte in allen vier Richtungen durch ein vierdimensionales [^{13}C - ^{13}C]-korreliertes [^1H - ^1H]-NOESY Spektrum des Komplexes zwischen gleichförmig ^{13}C -markiertem Cyclophilin und Cyclosporin A (vgl. S. 29) in Bild 6 illustrieren. Weiter zeigt das Bild, daß sich auch in Spektren, die sowohl positive als auch negative Signale enthalten, die Phase einwandfrei korrigieren läßt.

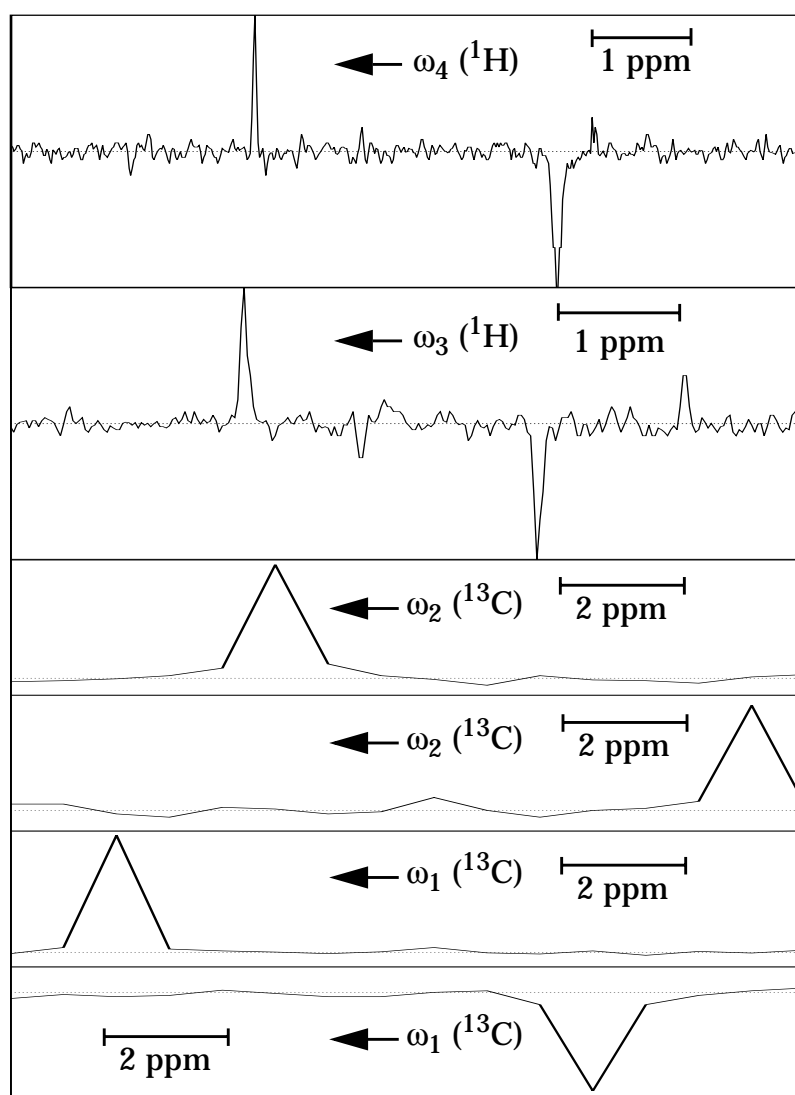


Bild 6. Schnitte durch ein 4D [^{13}C - ^{13}C]-korreliertes [^1H - ^1H]-NOESY Spektrum des Komplexes zwischen gleichförmig ^{13}C -markiertem Cyclophilin und Cyclosporin A, die das Ergebnis der automatischen Phasenkorrektur in allen vier Dimensionen illustrieren. Negative Signale sind gefaltet. In allen Beispielen wird die volle spektrale Breite gezeigt. Die Nulllinie ist gepunktet.

Basislinienkorrektur

Verzerrungen der Basislinie in Kernresonanzspektren können eine Vielzahl von Ursachen haben (vgl. z. B. Stejskal & Schaefer, 1974; Hoult *et al.*, 1983; Otting *et al.*, 1986; Ernst *et al.*, 1987; Freeman, R., 1988). Die Techniken zur Verringerung von Basislinienfehlern können in drei Kategorien eingeteilt werden: Vorkehrungen während des Experiments, z. B. optimale Phasenzyklen (Ernst *et al.*, 1987) oder überhöhte Abtastfrequenz während der Datenaufnahme (Wider, 1990); Verbesserungen im Zeitbereich, z. B. durch korrekte Gewichtung des ersten Datenpunkts (Otting *et al.*, 1986) bzw. lineare Vorhersage der ersten Datenpunkte des freien Induktionszerfalls (Marion & Bax, 1989); und Korrekturen im Frequenzbereich, d. h. in den Spektren, normalerweise durch Anpassung und Subtraktion einer geeigneten Funktion, die die Verzerrung der Basislinie wiedergibt. Dies kann ein Polynom (Bruker Analytische Meßtechnik GmbH, 1991; Dietrich *et al.*, 1991) oder eine Linearkombination der trigonometrischen Funktionen, die dem Anfang des freien Induktionszerfall entsprechen (Henrichs *et al.*, 1986; Cavanagh & Rance, 1990), sein. Im Programm PROSA sind Methoden der beiden letzteren Klassen erhältlich: Basislinienkorrektur im Zeitbereich durch lineare Vorhersage (vgl. S. 30) und im Frequenzbereich mit dem FLATT Verfahren (Güntert & Wüthrich, 1992).

Die FLATT Methode verbessert die Basislinie jedes eindimensionalen Schnitts eines mehrdimensionalen Spektrums getrennt, indem zunächst signalfreie Gebiete »reiner Basislinie«, identifiziert werden, an die anschließend eine geeignete Funktion angepaßt und vom gesamten eindimensionalen Spektrum subtrahiert wird. Die Suche von Gebieten reiner Basislinie beruht darauf, daß ein kurzes, zusammenhängendes Stück des Spektrums, das klein verglichen mit der gesamten spektralen Breite, aber größer als die Breite der Resonanzlinien ist, nur dann gut durch eine Gerade wiedergegeben werden kann, wenn es frei von Signalen ist. Für ein eindimensionales Spektrum von N Datenpunkten mit den Intensitäten s_1, \dots, s_N gibt die mittlere quadratische Abweichung

$$\chi_k^2 = \min_{a,b} \frac{1}{2n+1} \sum_{l=-n}^n (s_{k+l} - a - bl)^2 \quad [22]$$

für ein Gebiet von $2n+1$ Datenpunkten, zentriert um den Datenpunkt k ($k = n+1, \dots, N-n$), die Güte der Anpassung an eine Gerade an. Der Parameter n wird je nach Spektrum festgelegt; bei ^1H Spektren wird n typischerweise so gesetzt, daß $2n+1$ Datenpunkte etwa 75 Hz entsprechen. Das Minimum in Gleichung [22] wird durch

$$a = \frac{1}{2n+1} \sum_{l=-n}^n s_{k+l} \quad \text{und} \quad b = \frac{3}{n(n+1)(2n+1)} \sum_{l=-n}^n l s_{k+l} \quad [23]$$

erreicht. In den Randgebieten $k \leq n$ und $k > N - n$ wird χ_k^2 auf den χ^2 -Wert des nächsten Datenpunktes im Definitionsbereich von Gleichung [22] gesetzt. Um Gebiete reiner Basislinie festzulegen, werden geglättete Werte verwendet:

$$\chi_k^2 = \min(\chi_{k-n/3}^2, \dots, \chi_{k+n/3}^2), \quad [24]$$

die für reine Basisliniengebiete eine minimale Weite sicherstellen und sie weiter in die Nähe von Signalen ausdehnen, als es mit χ_k^2 aus Gleichung [22] der Fall wäre. Ein Datenpunkt k liegt genau dann in einem Gebiet reiner Basislinie, wenn $\chi_k^2 \leq \tau \chi_c^2$ ist, wobei τ ein vorgegebener Parameter (üblicherweise $\tau = 4$) und der Schwellenwert¹ χ_c^2 in Analogie zum Median so festgelegt ist, daß für einen Drittel der Datenpunkte $\chi_k^2 \leq \chi_c^2$ und für die restlichen zwei Drittel $\chi_k^2 > \chi_c^2$ gilt. Diese Wahl von χ_c^2 stellt sicher, daß für $\tau \geq 1$ die gefundenen Gebiete reiner Basislinie immer mindestens einen Drittel der Datenpunkte umfassen. Zur Anpassung an die Basislinienverzerrung kann eine Linearkombination beliebiger, in FORTRAN-77 darstellbarer Funktionen verwendet werden — in der Praxis meist eine Funktion, die einer Korrektur der ersten Datenpunkte im Zeitbereich entspricht. Die Anpassung der Funktion an die Intensitäten in den Gebieten reiner Basislinie verwendet die Standardmethode der Singulärwertzerlegung (Press *et al.*, 1986).

Bild 7 zeigt anhand von eindimensionalen Schnitten durch ein zweidimensionales NOESY Spektrum des Toxins K der Schwarzen Mamba, *Dendroaspis polylepis polylepis*, eines Proteins mit 57 Aminosäureresten (Berndt *et al.*, 1993), wie erfolgreich die FLATT Methode selbst bei Spektren mit vielen Resonanzlinien die Basislinie zu glätten versteht. Die Messungen wurden auf einem Bruker AM 600 Spektrometer unter den folgenden experimentellen Bedingungen durchgeführt: 15mM Lösung des Proteins in 95% H₂O/5% D₂O; pH 4,6; Temperatur 35°C; ¹H Frequenz 600 MHz; Mischzeit 40 ms; 920 und 2048 aufgenommene Datenpunkte in t_1 und t_2 ; Kosinusfilter in beiden Dimensionen.

Weil größere Basislinienverzerrungen in der Aufnahmedimension (ω_2) auftreten, wurde die Basislinie zuerst in dieser Richtung mit $n = 20$ in den Gl. [22–24], anschließend in der indirekt gemessenen Dimension (ω_1) mit $n = 10$ korrigiert. An die Basislinienverzerrungen wurde eine Funktion, die einer Änderung der ersten vier Datenpunkte im Zeitbereich entspricht, angepaßt, deren Subtraktion vom Original- das basislinienkorrigierte Spektrum lieferte. Eine solche Funktion kann Sprünge der Basislinie, die auf der Diagonalen durch starke Resonanzlinien mit dispersivem Anteil verursacht werden, nicht korrekt wiedergeben. Daher wurde die Basislinienkorrektur für die Gebiete

¹ Ursprünglich diente das Minimum von χ_k^2 als Schwellenwert χ_c^2 (Güntert & Wüthrich, 1992).

auf beiden Seiten der Diagonale unabhängig voneinander ausgeführt (Güntert & Wüthrich, 1992). Die von FLATT festgestellten Gebiete reiner Basislinie in Bild 7B und D zeigen, wie zuverlässig der Algorithmus zwischen Gebieten reiner Basislinie und Signalen unterscheiden kann. Nach der Basislinienkorrektur weist das Spektrum in fast allen Gebieten reiner Basislinie im Rahmen des Rauschens verschwindende Intensität auf.

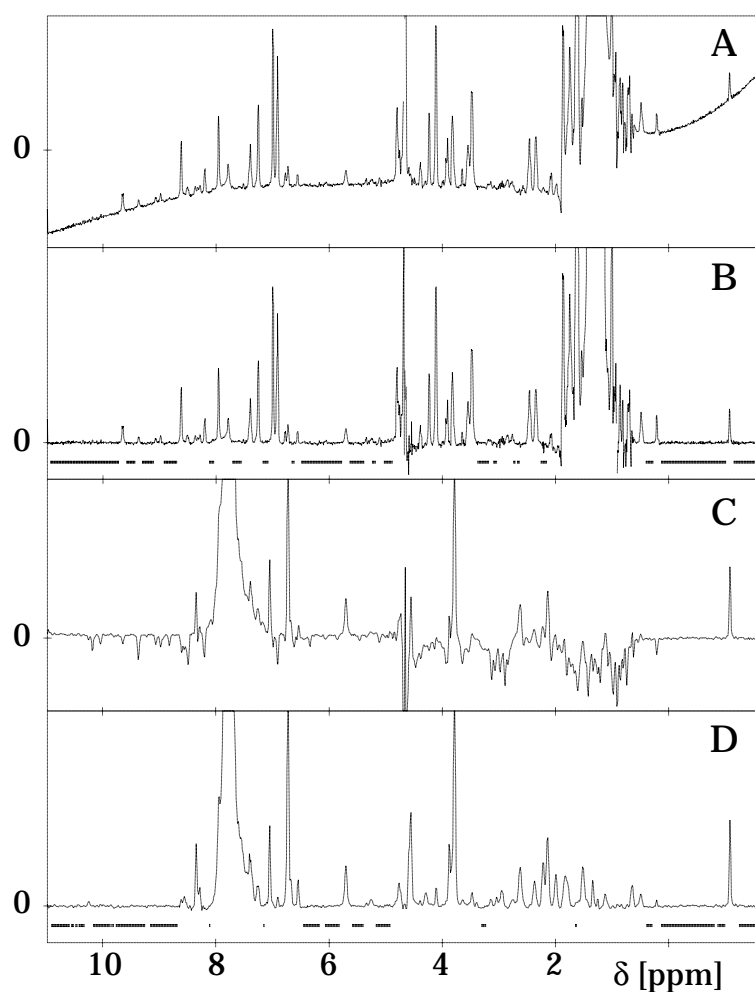


Bild 7. Schnitte durch ein NOESY Spektrum von Toxin K; gezeigt wird jeweils die gesamte spektrale Weite von 12,62 ppm:

(A) Schnitt bei $\omega_1 = 1,36$ ppm vor der Basislinienkorrektur:

(B) Der gleiche Schnitt nach Basislinienkorrektur in beiden Dimensionen.

(C) Schnitt bei $\omega_2 = 7,76$ ppm vor der Basislinienkorrektur:

(D) Der gleiche Schnitt nach Basislinienkorrektur in beiden Dimensionen. Die Verbesserung gegenüber C beruht in erster Linie auf der Basislinienkorrektur in der (ω_2 -) Richtung senkrecht zum Schnitt.

Horizontale Linien in B und D zeigen die gefundenen Gebiete reiner Basislinie an.

Im Programm PROSA können signalfreie Gebiete außer mit FLATT auch durch numerisches Differenzieren des geglätteten Spektrums erkannt werden (Dietrich *et al.*, 1991).

Gebrauch des Programms PROSA

Die folgende Tabelle enthält als Beispiel die vollständige Sequenz von PROSA-Anweisungen zur Berechnung eines vierdimensionalen [^{13}C - ^{13}C]-korrelier-

ten [¹H-¹H]-NOESY Spektrums des Komplexes zwischen gleichförmig ¹³C-markiertem Cyclophilin und unmarkiertem Cyclosporin A (vgl. S. 29):

Sequenz von PROSA-Anweisungen zur Berechnung eines 4D Spektrums

PROSA-Anweisung	aktiv	Erläuterung
read <i>fn</i> 256c 110c 8c 6c	alle	liest die Datei <i>fn</i> (256 x 110 x 8 x 6 komplexe Punkte)
multiply 0.01	ω_4	multipliziert alle Datenpunkte mit 0,01 (Skalierung)
window gauss 6.7 2.0	ω_4	legt eine Lorentz-Gauß Fensterfunktion an
multiply 0.5 1	ω_4	halbiert den ersten Punkt jedes freien Induktionszerfalls
ft 512	ω_4	füllt (mit 0) auf 512 Punkte auf und fouriertransformiert
dimension 2	$\omega_4 \rightarrow \omega_3$	transponiert, um Dimension 2 (ω_3) zu aktivieren
window cos	ω_3	legt eine Kosinus Fensterfunktion an
multiply -1 2 \$n 2	ω_3	wechselt das Vorzeichen jedes zweiten Punktes (»States-TPPI«)
ft 256	ω_3	füllt (mit 0) auf 256 Punkte auf und fouriertransformiert
dimension 1	$\omega_3 \rightarrow \omega_4$	transponiert, um Dimension 1 (ω_4) zu aktivieren
autophase 10 2 10 20	ω_4	korrigiert die Phase (mit $u = 10$, $\kappa = 10$ und $v = 20$)
re	ω_4	behält nur den Realteil des Spektrums
dimension 2	$\omega_4 \rightarrow \omega_3$	transponiert, um Dimension 2 (ω_3) zu aktivieren
autophase 8 2 10 20	ω_3	korrigiert die Phase (mit $u = 8$, $\kappa = 10$ und $v = 20$)
re	ω_3	behält nur den Realteil des Spektrums
dimension 3	$\omega_3 \rightarrow \omega_2$	transponiert, um Dimension 3 (ω_2) zu aktivieren
ft	ω_2	fouriertransformiert, ohne mit Nullen aufzufüllen
dimension 4	$\omega_2 \rightarrow \omega_1$	transponiert, um Dimension 4 (ω_1) zu aktivieren
predict lpsvd 2 10	ω_1	lineare Vorhersage von 10 Punkten mit 2 Koeffizienten
window cos2	ω_1	legt eine Kosinusquadrat Fensterfunktion an
multiply -1 2 \$n 2	ω_1	wechselt das Vorzeichen jedes zweiten Punktes (»States-TPPI«)
ft	ω_1	fouriertransformiert, ohne mit Nullen aufzufüllen
autophase 4 2 10 50	ω_1	korrigiert die Phase (mit $u = 4$, $\kappa = 10$ und $v = 50$)
re	ω_1	behält nur den Realteil des Spektrums
dimension 3	$\omega_1 \rightarrow \omega_2$	transponiert, um Dimension 3 (ω_2) zu aktivieren
ift	ω_2	wendet eine inverse Fourier-Transformation an
predict lpsvd 3 8	ω_2	lineare Vorhersage von 8 Punkten mit 3 Koeffizienten
window cos2	ω_2	legt eine Kosinusquadrat Fensterfunktion an
multiply -1 2 \$n 2	ω_2	wechselt das Vorzeichen jedes zweiten Punktes (»States-TPPI«)
ft	ω_2	fouriertransformiert, ohne mit Nullen aufzufüllen
autophase 4 2 10 50	ω_2	korrigiert die Phase (mit $u = 4$, $\kappa = 10$ und $v = 50$)
re	ω_2	behält nur den Realteil des Spektrums
dimension 1	$\omega_2 \rightarrow \omega_4$	transponiert, um Dimension 1 (ω_4) zu aktivieren
flatt 10 5.0 3	ω_4	korrigiert die Basislinie (mit $n = 10$, $\tau = 5$ und $m = 3$)
dimension 2	$\omega_4 \rightarrow \omega_3$	transponiert, um Dimension 2 (ω_3) zu aktivieren
flatt 8 5.0 3	ω_3	korrigiert die Basislinie (mit $n = 10$, $\tau = 5$ und $m = 3$)
dimension 1 2 3 4	$\omega_3 \rightarrow \omega_4$	stellt die ursprüngliche Ordnung der Dimensionen her
write easy8 <i>spec</i>	alle	schreibt die Spektrumdatei <i>spec</i> für EASY (8 Bit/Punkt)

In seiner jetzigen Version liest das Programm PROSA serielle Eingabedateien

mit ganzen oder reellen Zahlen (im Sinne der FORTRAN Datentypen *Integer* und *Real*) sowie Submatrixdateien im Format des Programms EASY mit 8 oder 16 Bit pro Datenpunkt (Eccles *et al.*, 1991; C. Bartels und K. Wüthrich, unveröffentlicht). Die Erweiterung auf andere Datenformate ist einfach, weil nur die gemessenen Zeitbereichsdaten, ohne etwaigen Vorspann oder separate Parameterlisten, gelesen werden. Die meisten Anweisungen des Programms PROSA beziehen sich auf eine, die *aktive* Dimension des mehrdimensionalen Datensatzes, die durch Transponieren mit Hilfe der Anweisung **dimension** gewechselt werden kann. In der Tabelle auf S. 36 steht die aktive Dimension jeweils in der mittleren Spalte. Intern behandelt PROSA die Daten stets als zweidimensionale Matrix, wobei die Datenpunkte der aktiven Dimension sequentiell im Speicher des Rechners abgelegt werden.

Jeder reelle FORTRAN-77 Ausdruck kann als Fensterfunktion vor der Fourier-Transformation gebraucht werden; die allgemein gebräuchlichen Fensterfunktionen (DeMarco & Wüthrich, 1976; Ernst *et al.*, 1987) können über den Makro **window** (vgl. S. 54) aufgerufen werden. Allgemein sind *Makros* Dateien, die PROSA-Anweisungen enthalten (vgl. S. 52). Eine wichtige Anweisung ist **multiply**, die es erlaubt, alle oder eine Auswahl von Datenpunkten mit einem konstanten oder von Datenpunkt zu Datenpunkt veränderlichen Faktor zu multiplizieren:

$$\mathbf{multiply} \ f_k \ k_1 \ k_2 \ k_3 \quad [25]$$

multipliziert den Datenpunkt k mit dem Faktor f_k , wobei der Index k von k_1 bis k_2 in Schritten von k_3 läuft¹. Ein Kosinusfenster,

$$f(t) = \cos \frac{\pi t}{2} \quad (0 \leq t \leq 1), \quad [26]$$

wird zum Beispiel durch die PROSA-Anweisung

$$\mathbf{multiply} \ \cos(\$pi/(2*\$n)*(k-1)) \quad [27]$$

implementiert, wobei $\$pi = \pi$ ist, $\$n$ die Anzahl Datenpunkte in der aktiven Dimension bezeichnet und der Index k über alle Datenpunkte in der aktiven Dimension läuft: $k = 1, \dots, \$n$. Die Anweisung **multiply** dient weiter dazu, den gesamten Datensatz zu skalieren (**multiply 0.01** in der Tabelle auf S. 36), dem ersten Datenpunkt des freien Induktionszerfalls vor der diskreten Fourier-Transformation das richtige Gewicht zu geben (Otting *et al.*, 1986; **multiply 0.5 1** in der Tabelle), das Vorzeichen jedes zweiten Datenpunktes zu wechseln,

¹ Alle Datenpunkte werden multipliziert, wenn k_1 , k_2 und k_3 weggelassen werden; wenn nur k_1 gegeben ist, wird nur der Datenpunkt k_1 multipliziert; wenn k_3 fehlt, nimmt das Programm als Schrittweite $k_3 = 1$ an.

wenn die »States-TPPI« Methode zur phasensensitiven Aufnahme (Marion *et al.*, 1989) verwendet wird (**multiply -1 2 \$n 2** in der Tabelle), oder Phasenkorrekturen anzuwenden, die, zum Beispiel durch korrektes Einstellen der Anfangswartezeit in der Pulssequenz (Bax *et al.*, 1991), bereits bekannt sind. Eine vollständige Liste aller PROSA-Anweisungen folgt auf den Seiten 39ff.

Rechenzeiten auf verschiedenen Rechnern für die Berechnung von drei typischen Spektren — eines vierdimensionalen [¹³C–¹³C]-korrelierten [¹H–¹H]-NOESY-Spektrums des Komplexes zwischen gleichförmig ¹³C-markiertem Cyclophilin und Cyclosporin A (vgl. S. 29) als Beispiel für einen großen Datensatz, eines dreidimensionalen HCCH-TOCSY-Spektrums (Bax *et al.*, 1990) des gleichen Komplexes (C. Spitzfaden, G. Wider & K. Wüthrich, unveröffentlicht) als Beispiel für ein übliches heteronukleares dreidimensionales Spektrum und ein homonukleares zweidimensionales NOESY-Spektrum des 434-Repressors(1–63) (Schott, 1992) als typisches Beispiel eines Spektrums, wie es gebraucht wird, um konformationelle Einschränkungen für die Strukturbestimmung eines kleineren Proteins zu gewinnen — gibt die folgende Tabelle an:

Rechenzeiten für die Berechnung mehrdimensionaler Spektren mit PROSA

Eingabegröße [komplexe Punkte]	Rechner	Rechenzeit [s]				
		total	lin. Vorhersage	Fourier Transf.	Phasen-korr.	Basislinienkorr.
256 x 110 x 8 x 6 (512x256x16x16)	NEC SX-3	574	285	18	92	144
512 x 80 x 8 (512 x 256 x 32)	NEC SX-3	39	11	1,2	7	16
	Convex C2	600	258	23	69	179
	Sun-4/690	2093	1052	114	87	644
2048 x 391 (2048 x 512)	NEC SX-3	4,5	–	0,7	0,8	2,0
	Convex C2	80	–	16	10	29
	Iris 4D/25	300	–	77	19	143
	Sun-4/690	343	–	86	22	145

Die Prozessierungsschritte für das vierdimensionale Spektrum stehen in der Tabelle auf S. 36. Fourier-Transformation, automatische Phasen- und Basislinienkorrektur wurden für alle Spektren gebraucht, und in allen ¹³C-Dimensionen wurden die Zeitbereichsdaten durch lineare Vorhersage verlängert. Auf dem NEC SX-3 Hochleistungsrechner betragen die Rechenzeiten weniger als 10 Min., 40 s bzw. 5 s für die drei Spektren. Die zwei- und dreidimensiona-

len Spektren wurden auch auf Convex C2, Silicon Graphics Iris und Sun Rechnern berechnet, wozu zwischen 15 und 76 mal mehr Rechenzeit benötigt wurde. Die Tabelle gibt auch getrennte Rechenzeiten für die wichtigsten Prozessierungsschritte an. Auf allen Rechnern bedürfen die lineare Vorhersage und die Basislinienkorrektur mit 28–50% bzw. 25–45% der meisten Rechenzeit, wogegen die Fourier-Transformation und die automatische Phasenkorrektur mit 16–26% bzw. 4–18% der gesamten Rechenzeit deutlich weniger ins Gewicht fallen. Diese Rechenzeiten bedeuten einen wesentlichen Fortschritt gegenüber früheren Angaben in der Literatur, die für die lineare Vorhersage eines typischen dreidimensionalen Datensatzes mittels Singulärwertzerlegung von 8–32 Std. Rechenzeit auf einem Sun-4 Rechner ausgehen (Zhu & Bax, 1990).

PROSA-Anweisungen

Es gibt im Programm PROSA zwei Sorten von Anweisungen: Allgemeine, die in der Routine zum Lesen der Kommandozeile (vergleichbar mit einer »Shell« im UNIX Betriebssystem) eingebaut und nicht spezifisch für das Programm PROSA sind¹, sowie PROSA-spezifische Anweisungen.

In der folgenden, alphabetischen Liste aller Anweisungen sind wörtliche Eingaben **fett** und nicht-wörtliche Eingaben *kursiv* gedruckt, wobei fakultative Eingaben in eckigen Klammern [...] und fakultative Eingaben, die mehrmals nacheinander wiederholt werden können, in geschweiften Klammern {...} stehen. In Beispielen sind Ausgaben des Programms *kursiv* gedruckt.

Die Spezialzeichen »\$ % { } : \ ' ' #« haben in der PROSA-Kommandozeile spezielle Bedeutungen: »\$ Variable« oder »% Variable« bezeichnen den Wert der *Variablen* (vgl. S. 49); »{ \$ Variable}« oder »{ % Variable}« grenzen den Variablennamen *Variable* vom unmittelbar folgenden Text ab; »Marke:« bezeichnet eine Marke, die zusammen mit einer **goto** Anweisung gebraucht werden kann (vgl. S. 42); »\ Buchstabe« behandelt den *Buchstaben* wörtlich und erlaubt die Aufnahme von Spezialzeichen in normalen Text, »\« am Ende einer Zeile gibt an, daß die Anweisung auf der nächsten Zeile weitergeht; »"Text"« behandelt den *Text* als einen Parameter, selbst wenn er Leerzeichen enthält; »' Text'« behandelt den *Text* ebenfalls als einen Parameter, wobei im Gegensatz zu den Anführungszeichen in »"Text"« die Apostrophe Teile des Texts bleiben. Apostrophe werden für FORTRAN-77 Zeichenkettenkonstanten benötigt. Text zwischen einem Kommentarzeichen »#« und dem Ende der betreffenden Zeile gilt für das Programm als Kommentar und bleibt unbeachtet.

abs

ersetzt die Daten in der aktiven Dimension durch ihren Absolutwert, $s \rightarrow |s|$.

ask *Text Variable {Variable}*

schreibt den *Text* auf die Standardausgabe, liest eine Eingabezeile und weist die darin enthaltenen, durch Leerzeichen voneinander getrennten Zeichenketten der Reihe nach den angegebenen Variablen zu. Diese Anweisung wird normalerweise in Makros verwendet, um interaktive Eingaben zu erfragen. Soll der *Text* Leerzeichen enthalten, müssen ihn Anführungszeichen einschließen.

¹ Dies sind die Anweisungen **ask, break, do, error, eval, exit, goto, help, if, parameter, print, quit, return, set, show, type** und **var**.

Beispiel: **ask "Erster und letzter Punkt:" Anfang Ende**

Erster und letzter Punkt:

12 45

print Bereich = \$Anfang...\$Ende

Bereich = 12...45

autophase *Breite Schwelle Höhe Überlapp* ϕ_1^{\max} {Option}

bestimmt konstante und lineare Phasenkorrekturparameter und führt eine automatische Phasenkorrektur aus (vgl. S. 31ff.). Die Parameter haben folgende Bedeutung:

Breite Maximale halbe Breite von akzeptablen Signalen im Absolutwertspektrum, angegeben in Datenpunkten (Parameter u auf S. 31).

Schwelle Schwellenwert zur Abgrenzung von Signalgebieten, bezogen auf die Rauschintensität: Signalgebiete werden durch die ersten Datenpunkte begrenzt, deren Intensität 10% des Signalmaximums oder *Schwelle* mal die Rauschintensität unterschreitet. Normalerweise ist 2,0 ein guter Wert für diesen Parameter.

Höhe Minimale Intensität akzeptabler Signalmaxima, bezogen auf die Rauschintensität: akzeptable Signalmaxima im Absolutwertspektrum müssen das Produkt aus *Höhe* und Rauschintensität übersteigen (Parameter κ auf S. 31).

Überlapp Maximale Anzahl von akzeptablen Signalen, die eine gemeinsame Frequenzkoordinate haben dürfen (Parameter ν auf S. 31).

ϕ_1^{\max} maximaler Absolutwert der linearen Phasenkorrektur ϕ_1 , d. h. ϕ_1 wird so gewählt, daß $|\phi_1| \leq \phi_1^{\max}$ gilt; $\phi_1^{\max} = 0$ bedeutet, daß nur eine konstante Phasenkorrektur bestimmt wird.

Option

- = **apply** Die Phasenkorrektur wird bestimmt und angewandt.
- = **determine** Die Phasenkorrektur wird bestimmt, aber nicht angewandt.
- = **complex** Signale werden im Real- und Imaginärteil der Daten bezüglich der nicht-aktiven Dimensionen gesucht.
- = **real** Signale werden nur im Realteil der Daten bezüglich der nicht-aktiven Dimensionen gesucht. Diese Option ist vorteilhaft, wenn die Phase in den nicht-aktiven Dimensionen bereits näherungsweise richtig eingestellt ist.
- = **global** Das globale Maximum der Zielfunktion im Bereich $-\phi_1^{\max} \leq \phi_1 \leq \phi_1^{\max}$ wird zur Bestimmung der linearen Phasenkorrektur ϕ_1 verwendet.
- = **local** Das lokale Maximum der Zielfunktion mit dem kleinsten Absolutbetrag $|\phi_1|$ wird zur Bestimmung der linearen Phasenkorrektur ϕ_1 verwendet.
- = **symmetrize** Die Signalregionen werden für die Phasenbestimmung so symmetrisiert, daß das Absolutwertspektrum symmetrisch bezüglich des Signalmaximums wird.
- = **info** Informationen über jedes berücksichtigte Signal werden angezeigt.
- = **equal** Signale erhalten das gleiche Gewicht.
- = **sqrt** Signale werden mit der Quadratwurzel ihrer Intensität gewichtet.
- = **proportional** Signale werden mit ihrer Intensität gewichtet.

Standardmäßig sind die Optionen **apply**, **complex**, **global** und **equal** gesetzt.

Die bestimmten Werte der konstanten und linearen Phasenkorrekturparameter werden den Programmvariablen **phi0** und **phi1** zugewiesen (vgl. S. 51).

break

verläßt eine »do«-Schleife und ist nur innerhalb von Makros erlaubt. Die Ausführung des Makros wird mit der ersten Anweisung, die auf die verlassene Schleife folgt, fortgesetzt.

complex

wandelt $2n$ reelle Datenpunkte in der aktiven Dimension in n komplexe Datenpunkte um, indem aufeinanderfolgende reelle Datenpunkte r_{2k-1} und r_{2k} abwechselnd als Real- und Imaginärteile komplexer Zahlen $z_k = r_{2k-1} + i r_{2k}$ aufgefaßt werden ($k = 1, \dots, n$). Falls die Anzahl reeller Datenpunkte ungerade ist, wird der Imaginärteil des letzten komplexen Datenpunktes auf null gesetzt. Komplexe Daten bleiben unverändert.

complexify

wandelt n reelle Datenpunkte r_k in der aktiven Dimension in n komplexe Datenpunkte $z_k = r_k$ mit verschwindendem Imaginärteil um. Komplexe Daten bleiben unverändert.

conjugate

nimmt das komplex konjugierte der komplexen Daten. Reelle Daten bleiben unverändert.

dimension *aktive Dimension* {*Dimension*}

transponiert die Datenmatrix so, daß die *aktive Dimension* zur aktiven Dimension wird. Falls weitere Dimensionen angegeben werden, wird mittels Transpositionen die gewünschte Ordnung der Dimensionen im Datensatz hergestellt.

do [*Variable Startwert Endwert* [*Schrittweite*]]

Sequenz von Anweisungen

end do

führt innerhalb eines Makros eine Schleife aus; entweder, falls **do** ohne Parameter steht, bedingungslos, d. h. so oft bis innerhalb der Schleife eine der Anweisungen **break**, **exit**, **quit** oder **return** auftritt, oder im Sinne einer FORTRAN-77 »do«-Schleife, wobei die *Variable* und die ganzzahligen Ausdrücke *Startwert*, *Endwert* und *Schrittweite* die übliche Bedeutung haben.

Beispiele:**do**

```
    if ('$Dateiname'.eq.' ') break
    ...
end do
do i 1 10
    print Iteration $i.
end do
```

error *Fehlermeldung* [*Dateiname*] [*Option*]

schreibt den *Text* auf die Standardausgabe oder in die angegebene Ausgabedatei und ruft die in der Programmvariablen **erract** (vgl. S. 51) gesetzte Fehlerbehandlungsroutine auf. Diese Anweisung eignet sich, um Fehler anzuzeigen, die während der Ausführung eines Makros auftreten. Enthält die *Fehlermeldung* Leerzeichen, müssen sie Anführungszeichen einschließen. Wenn die *Option* **rewrite** gesetzt ist, wird eine vorhandene Ausgabedatei überschrieben, anderenfalls wird die *Fehlermeldung* am Ende der Ausgabedatei angefügt. Falls keine Datei mit dem *Dateinamen* existiert, wird eine neue eröffnet.

eval *Variable* = *Ausdruck*

wertet den arithmetischen oder Zeichenkettenausdruck *Ausdruck* entsprechend den Regeln von FORTRAN-77 aus und weist das Ergebnis der Variablen mit dem Namen *Variable* zu. Anders als in FORTRAN-77 müssen Funktionsnamen in Kleinbuchstaben geschrieben werden und der *Ausdruck* darf (außer in Zeichenkettenkonstanten) keine Leerzeichen enthalten.

Beispiele:**set i=7**

```
set Satz=Ein flexibles Programm!
eval j=mod($i,4)**2
eval l=len('$Satz')
show i Satz j l
... Variables:
```

$i = 7$
Satz = *Ein flexibles Programm!*
 $j = 9$
 $l = 23$

exit

kehrt aus einem Makro zur interaktiven Eingabe zurück. Interaktiv eingegeben, verläßt **exit** das Programm.

flatten *Methode* n τ *Funktion* {*Funktion*}

glättet die Basislinie im Frequenzbereich (vgl. S. 33ff.) und verlangt reelle Daten in der aktiven Dimension. Die Parameter haben folgende Bedeutung:

Methode kann die Werte **flatt** und **derivative** annehmen und legt das Verfahren zur Suche von Gebieten reiner Basislinie fest — entweder die FLATT Methode (vgl. S. 33) oder das Verfahren von Dietrich *et al.* (1991), welches auf einer geglätteten Ableitung des Spektrums beruht.

n bezeichnet bei FLATT die halbe Länge des anzupassenden Geradenstücks (vgl. S. 33) und bei der »Ableitungsmethode« die halbe Breite des Glättungsbereichs (Güntert *et al.*, 1992).

τ ist ein Schwellenwert für die Bestimmung von Gebieten reiner Basislinie. Mit $\tau = 1$ erkennt das Programm etwa einen Drittel aller Datenpunkte als reine Basislinie; höhere Werte von τ führen zu größeren Gebieten reiner Basislinie (vgl. S. 33; Güntert *et al.*, 1992).

Funktion bezeichnet einen reellen FORTRAN-77 Ausdruck, in dem **k** für den Index steht, der über alle Datenpunkte in der aktiven Dimension läuft, und wird als Basisfunktion zur Anpassung an die Basislinienverzerrung verwendet: Die Verzerrung der Basislinie wird durch eine Linearkombination der angegebenen Basisfunktionen modelliert.

ft [N] [n_b] [n_e] [ϕ_0] [ϕ_1] {*Option*}

führt in der aktiven Dimension mit reellen Daten eine reelle oder mit komplexen Daten eine komplexe diskrete Fourier-Transformation durch. Die Daten nach der Fourier-Transformation sind immer komplex. Vor der Fourier-Transformation werden die Daten am Ende mit Nullen auf N komplexe Punkte aufgefüllt. N muß eine Potenz von 2 sein. Falls N nicht angegeben wird, füllt das Programm wenn nötig bis zur nächsten Zweierpotenz mit Nullen auf. Von den gesamten Daten im Frequenzbereich behält das Programm auf Wunsch nur den Streifen mit den Datenpunkten n_b, \dots, n_e zurück. Falls Phasenkorrekturparameter ϕ_1 und ϕ_2 angegeben sind, wird die Phase gemäß Gl. [21] korrigiert und es wird nur der Realteil der Daten aufbewahrt. Mit der Option **full** beziehen sich die Phasenkorrekturparameter ϕ_1 und ϕ_2 auf die gesamte spektrale Breite; mit der standardmäßig gesetzten Option **strip** beziehen sie sich dagegen auf den Streifen mit den Datenpunkten n_b, \dots, n_e .

goto *Marke*

springt innerhalb eines Makros zur ersten Zeile, die mit der angegebenen Marke beginnt. Sprünge ins Innere von Schleifen (**do . . . end do**) oder zu bedingt ausgeführten Anweisungen (**if . . . else . . . end if**) sind nicht erlaubt und können zu unvorhersehbaren Ergebnissen führen. Eine Marke kann Buchstaben, Ziffern und das Unterstreichungszeichen »_« enthalten. Falls die Marke nicht ausschließlich aus Ziffern besteht, muß ihr ein Doppelpunkt unmittelbar folgen.

Beispiel: **goto Fortsetzung**

```

. . .
Fortsetzung: print Jetzt bei der Marke Fortsetzung.
goto 40
. . .
40 print Zahlenmarken brauchen keinen Doppelpunkt.

```

help [*Thema*]

hilft mit Informationen zum angegebenen Thema weiter. Falls kein Thema angegeben ist, wird eine Liste aller Themen, zu denen Informationen verfügbar sind, angezeigt. Hilfsinformation zu einem Makro kann direkt in den Makro eingeschlossen werden: **help** *Makro-name* zeigt alle Zeilen des Makros, die mit »##« beginnen, an.

if (*Bedingung*) *Anweisung*

führt wie in einer logischen »if« Anweisung von FORTRAN-77 die gegebene Anweisung aus, wenn das Ergebnis des logischen Ausdruck *Bedingung* wahr ist. Eine Zeile mit einer logischen »if« Anweisung darf nicht mit dem Wort »then« enden. Zusätzlich zu den Möglichkeiten von FORTRAN-77 stehen zwei logische Funktionen zur Verfügung: **exist**(*Variable*) ist genau dann wahr, wenn die *Variable* existiert; **def**(*Variable*) ist genau dann wahr, wenn die *Variable* existiert und einen Wert verschieden von **NULL** hat.

Beispiel: **set i=-56**
if (Si.lt.0) print Si ist negativ.
-56 ist negativ.

if (*Bedingung*) then

Sequenz von Anweisungen

else if (*Bedingung*) then

Sequenz von Anweisungen

else

Sequenz von Anweisungen

end if

führt eine Block-»if« Anweisung aus wie in FORTRAN-77. Zusätzlich zu den Möglichkeiten von FORTRAN-77 stehen zwei logische Funktionen zur Verfügung: **exist**(*Variable*) ist genau dann wahr, wenn die *Variable* existiert; **def**(*Variable*) ist genau dann wahr, wenn die *Variable* existiert und einen Wert verschieden von **NULL** hat.

Beispiel: **if (mod(Si,2).eq.1) then**
print Si ist eine ungerade Zahl.
else if (def(x) .and. exist(y)) then
print Die Variable x ist definiert, und die Variable y existiert.
else if ('Ss'.eq.' ') then
print Die Variable s enthält nur Leerzeichen.
end if

ift [*N*] [*n_b*] [*n_e*]

führt eine inverse komplexe Fourier-Transformation aus. Vor der inversen Fourier-Transformation werden die Daten symmetrisch auf beiden Seiten mit Nullen auf *N* komplexe Punkte aufgefüllt. *N* muß eine Potenz von 2 sein. Falls *N* nicht angegeben wird, füllt das Programm wenn nötig bis zur nächsten Zweierpotenz mit Nullen auf. Von den gesamten Daten im Zeitbereich behält das Programm auf Wunsch nur den Streifen mit den Datenpunkten *n_b*, ..., *n_e* zurück.

multiply *Faktor* [*Startwert* *Endwert* [*Schrittweite*]]

multipliziert die Daten in der aktiven Dimension mit einem konstanten oder variablen Ausdruck *Faktor*; in dem **k** für den Index stehen kann, der mit der angegebenen *Schrittweite* vom *Startwert* zum *Endwert* über die Datenpunkte in der aktiven Dimension läuft (vgl. S. 37). *Faktor* muß ein ganzzahliger, reeller oder (im Falle komplexer Daten) komplexer FORTRAN-77 Ausdruck sein. Wenn *Startwert*, *Endwert* und *Schrittweite* nicht gegeben sind, werden alle Datenpunkte multipliziert. Wenn lediglich ein *Startwert* gegeben ist, wird nur der entsprechende Datenpunkt multipliziert. Der Standardwert für die *Schrittweite* ist 1.

Beispiele: **multiply 0.05** # Daten skalieren
multiply 0.5 1 # 1. Datenpunkt halbieren

multiply -1 2 \$n 2 # Vorzeichen jedes zweiten Datenpunkts wechseln
multiply cos(\$pi/(2*\$n)*(k-1)) # Kosinus Fensterfunktion

parameter *Variable {Variable}*

ändert in einem Makro die Namen der übergebenen Parameter, d. h. die Parameter **p1**, **p2**, ... erhalten der Reihe nach die in der **parameter** Anweisung gegebenen Namen. Die **parameter** Anweisung muß vor allen anderen Anweisungen (außer **var**) am Anfang eines Makros stehen und ist nicht interaktiv verwendbar.

plot *Format Dateiname Basishöhe Faktor n_+ n_-*

x-Größe x-Verschiebung x-Randbreite x-Zahlenabstand x-Markenabstand

y-Größe y-Verschiebung y-Randbreite y-Zahlenabstand y-Markenabstand {Option}

fertigt eine Konturzeichnung des Spektrums an. Die Parameter haben folgende Bedeutung:

Format = **hp7550a** HP-GL Format für den HP 7550A Plotter mit A3 Papier,
 = **hp7550a/a4** HP-GL Format für den HP 7550A Plotter mit A4 Papier,
 = **hp7596a** HP-GL Format für den HP 7596A Plotter,
 = **postscript** *Postscript* Format für A4 Papier.

Dateiname Name der Ausgabedatei. Wenn das Spektrum mehrere Ebenen umfaßt, wird für jede Ebene eine Ausgabedatei mit dem Namen *Dateiname.k* geschrieben.

Basishöhe Höhe der untersten Konturlinie. Standardwert: 5-fache Rauschintensität.

Faktor Faktor zwischen den Höhen benachbarter Konturlinien. Standardwert: $\sqrt{2}$.

n_+ Maximale Anzahl positiver Konturlinien. Standardwert: 12.

n_- Maximale Anzahl negativer Konturlinien. Standardwert: 12

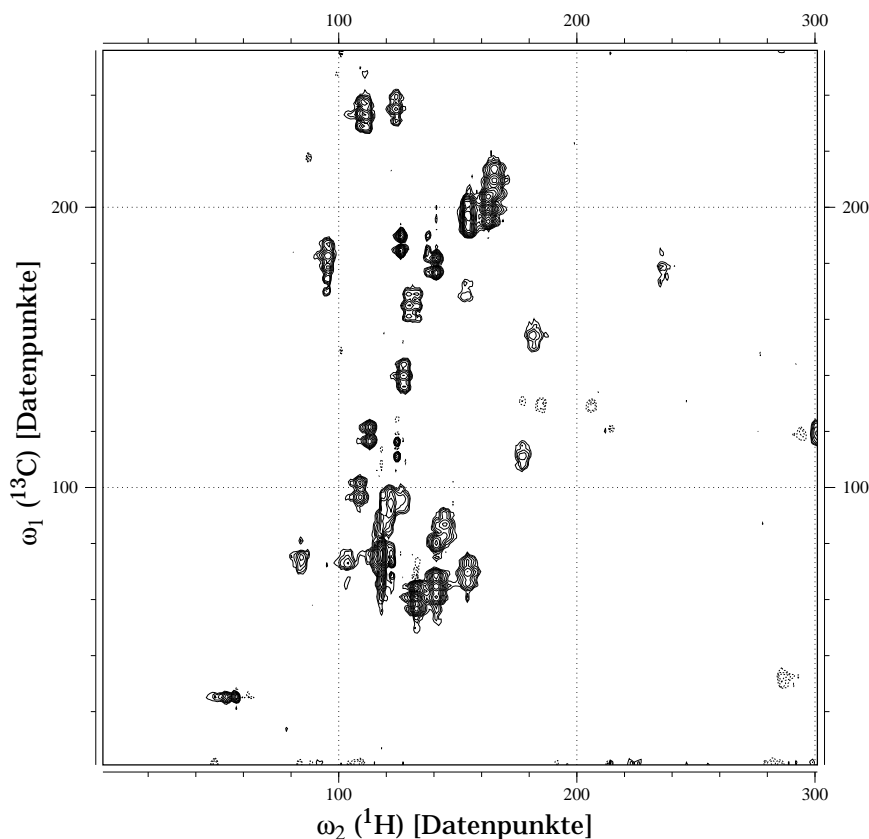


Bild 8. Mit Hilfe der Anweisung **plot** des Programms *PROSA* erzeugte Konturzeichnung eines Ausschnitts aus einem $[^{13}\text{C}-^1\text{H}]$ COSY Spektrum eines Komplexes zwischen der Antennapedia(C39S) Homeodomäne und einem 14 Basenpaaren langen DNS-Stück (Qian et al., 1993). Gezeigt werden Korrelationen zwischen ^{13}C und ^1H Atomen in aromatischen Ringen. Für alle Parameter der **plot** Anweisung wurden die Standardwerte verwendet.

<i>α-Größe</i>	Größe der Konturzeichnung (ohne Rand) in α -Richtung in cm ($\alpha = x, y$).
<i>α-Verschiebung</i>	Verschiebung der Konturzeichnung in cm vom Referenzpunkt in α -Richtung ($\alpha = x, y$).
<i>α-Randbreite</i>	Breite des Rands in α -Richtung in cm ($\alpha = x, y$).
<i>α-Zahlenabstand</i>	Abstand zwischen Zahlen der Beschriftung in α -Richtung ($\alpha = x, y$), gemessen in spektralen Einheiten (ppm, falls das Spektrum kalibriert ist, sonst Punkte). Diese Parameter gibt auch den Gitterabstand an, wenn ein Gitter gezeichnet werden soll.
<i>α-Markenabstand</i>	Abstand zwischen Marken auf den Beschriftungsachsen in α -Richtung ($\alpha = x, y$), gemessen in spektralen Einheiten (ppm, falls das Spektrum kalibriert ist, sonst Punkte).
<i>Option</i>	<ul style="list-style-type: none"> = grid legt ein Gitter auf das Spektrum, = nogrid zeichnet kein Gitter, = margin umgibt das Spektrum mit einem beschrifteten Rand, = nomargin zeichnet keinen Rand, = eject sorgt dafür, daß die fertige Zeichnung ausgeworfen wird, = noeject verhindert, daß die fertige Zeichnung ausgeworfen wird. Standardmäßig sind die Optionen grid , margin und (außer für den HP 7596A Plotter) eject gesetzt.

Die Standardwerte für die Parameter *α -Größe*, *α -Verschiebung*, *α -Randbreite*, *α -Zahlenabstand* und *α -Markenabstand* hängen von gewählten *Format* ab und sind so gewählt, daß die Zeichnung die vorhandene Papiergröße gut ausnutzt. Soll für einen Parameter sein Standardwert verwendet werden, kann dies durch einen Stern »*« angegeben werden.

power

ersetzt die Daten in der aktiven Dimension durch ihr Absolutwertquadrat, $s \rightarrow |s|^2$.

predict Methode m n [k_b] [k_e]

berechnet komplexe Datenpunkte in der aktiven Dimension durch lineare Vorhersage (vgl. S. 28ff.) mit folgenden Parametern:

<i>Methode</i>	muß in der jetzigen Version von PROSA lpsvd sein, d. h. die Koeffizienten der linearen Vorhersage werden mittels Singulärwertzerlegung berechnet.
<i>m</i>	Anzahl der Koeffizienten zur linearen Vorhersage in Gl. [14].
<i>n</i>	Anzahl vorhergesagter Datenpunkte. Wenn n positiv ist, werden n neue Datenpunkte am Ende vorhergesagt; bei negativem n werden die $ n $ ersten Datenpunkte durch linear vorhergesagte ersetzt. Die letztere Möglichkeit wird zur Korrektur von Basislinienverzerrungen verwendet, deren Ursache fehlerhafte Datenpunkte am Anfang der Zeitbereichsdaten sind (vgl. S. 30).
<i>k_b, k_e</i>	bezeichnen den Bereich von Datenpunkten, der zur Bestimmung der Koeffizienten zur linearen Vorhersage in Gl. [14] herangezogen wird. Ohne Angabe von k_b und k_e verwendet das Programm alle vorhandenen Datenpunkte zur Berechnung der Koeffizienten.

print Text [Dateiname] [Option]

schreibt den *Text* auf die Standardausgabe oder in die angegebene Ausgabedatei. Enthält der *Text* Leerzeichen, müssen ihn Anführungszeichen einschließen. Wenn die *Option rewrite* gesetzt ist, wird eine vorhandene Ausgabedatei überschrieben, anderenfalls wird der *Text* am Ende der Ausgabedatei angefügt. Falls keine Datei mit dem *Dateinamen* existiert, wird eine neue eröffnet.

project n

projiziert die Daten entlang der *letzten* Dimension. Falls $n = 0$ ist, ist die Projektion durch den Datenpunkt mit dem größten Absolutbetrag gegeben. Wenn n eine natürliche Zahl ist, berechnet sich die Projektion p der Datenpunkte s_k entlang der letzten Dimension gemäß

$$p = \sqrt[n]{\sum_k \text{sgn } s_k |s_k|^n}. \quad [28]$$

quit

verläßt das Programm PROSA.

re

ersetzt komplexe Daten in der aktiven Dimension durch ihren Realteil. Reelle Daten bleiben unverändert.

read *Format* *Dateiname* { n_k [c]}

read *Format* *Dateiname* **combine** [f_1] [f_2]

liest eine Datei mit Zeit- oder Frequenzbereichsdaten; in der ersten Form werden allenfalls im Programm vorhandene Daten überschrieben; in der zweiten Form wird eine Linearkombination zwischen den im Programm vorhandenen und den Eingabedaten erzeugt. Die Parameter haben folgende Bedeutung:

Format = **real** serielle Datei mit binär gespeicherten reellen Zahlen,
= **integer** serielle Datei mit binär gespeicherten ganzen Zahlen,
= **x32** serielle Datei mit binär gespeicherten ganzen Zahlen, wobei die Reihenfolge der 4 Bytes umgedreht ist,
= **text** Textdatei, die durch listengesteuerte Umwandlung im Sinne von FORTRAN-77 gelesen wird,
= **easy** Submatrix 8- oder 16-Bit Format des Programms EASY (Eccles *et al.*, 1991; C. Bartels und K. Wüthrich, unveröffentlicht).

Dateiname Name der Eingabedatei. Im Fall des *Formats* **easy** werden zwei Dateien gelesen: eine Parameterdatei mit dem Namen *Dateiname.3D.param* und die eigentliche Submatrixdatei *Dateiname.3D.8* bzw. *Dateiname.3D.16*.

n_k Anzahl der reellen oder komplexen Datenpunkte in der Dimension k des einzulesenden Datensatzes. Falls n_k unmittelbar von einem »c« gefolgt wird, werden komplexe, sonst reelle Daten gelesen. Die Angabe von n_k ist bei den *Formaten* **real**, **integer**, **x32** und **text** nötig, im Fall des *Formats* **easy** werden die entsprechenden Größen der EASY Parameterdatei entnommen, wobei die Daten immer reell aufgefaßt werden.

f_1 Gewichtungsfaktor der bereits im Programm vorhandenen Daten. Standardwert: 1.

f_2 Gewichtungsfaktor der einzulesenden Daten. Standardwert: 1.

real

wandelt n komplexe Datenpunkte in der aktiven Dimension in $2n$ reelle Datenpunkte um, indem aufeinanderfolgende Real- und Imaginärteile komplexer Zahlen z_k als reelle Datenpunkte $r_{2k-1} = \text{Re } z_k$ und $r_{2k} = \text{Im } z_k$ aufgefaßt werden ($k = 1, \dots, n$). Reelle Daten bleiben unverändert.

return

verläßt einen Makro und kehrt zum aufrufenden Makro bzw. zur interaktiven Eingabe zurück. Interaktiv eingegeben, verläßt **return** das Programm.

reduce *Bereich* {*Bereich*}

reduziert die Daten auf die angegebenen Bereiche; der erste *Bereich* bezieht sich auf die aktive Dimension, der zweite auf die zweite Dimension usw., wobei höchstens so viele Bereiche vorkommen dürfen, wie der Datensatz Dimensionen hat. Werden weniger Bereiche als Dimensionen angegeben, nimmt das Programm in den verbleibenden Dimensionen alle Datenpunkte mit. Außerhalb der gewählten Bereiche liegende Daten fallen weg. Ein *Bereich* kann in einer der folgenden Formen angegeben werden:

n bezeichnet den Datenpunkt n ,

$m \dots n$ umfaßt die Datenpunkte $m, m+1, \dots, n$,

$m \dots$ umfaßt die Datenpunkte $m, m+1, \dots$ bis zum letzten Datenpunkt,

$\dots n$ beinhaltet die Datenpunkte $1, 2, \dots, n$,

* steht für alle Datenpunkte.

select {*Bereich*}

wählt aus den Daten die angegebenen Bereiche aus; der erste *Bereich* bezieht sich auf die aktive Dimension, der zweite auf die zweite Dimension usw., wobei höchstens so viele Bereiche vorkommen dürfen, wie der Datensatz Dimensionen hat. Werden weniger Bereiche als Dimensionen angegeben, nimmt das Programm in den verbleibenden Dimensionen alle Datenpunkte mit. Außerhalb der gewählten Bereiche liegende Daten bleiben gespeichert, und der gesamte Datensatz kann wieder hergestellt werden, indem **select** ohne Bereichsangaben ausgeführt wird, sofern die Größe der ausgewählten Daten sich nicht geändert hat. Auf den mit **select** ausgewählten Teil der gesamten Daten können sämtliche PROSA-Anweisungen genau so wie auf den vollen Datensatz angewandt werden. Ein *Bereich* kann in einer der folgenden Formen angegeben werden:

n bezeichnet den Datenpunkt n ,
 $m..n$ umfaßt die Datenpunkte $m, m+1, \dots, n$,
 $m..$ umfaßt die Datenpunkte $m, m+1, \dots$ bis zum letzten Datenpunkt,
 $..n$ beinhaltet die Datenpunkte $1, 2, \dots, n$,
* steht für alle Datenpunkte.

n und m in Bereichsangaben beziehen sich immer auf den gesamten Datensatz, d. h. **select** kann nicht rekursiv verwendet werden.

Beispiele:**select 100..200 50..80** # wählt die Punkte 100, ..., 200 in der ersten (aktiven)
und 50, ..., 80 in der zweiten Dimension aus
select * * 20 # wählt die 20. Ebene eines 3D Spektrums aus
select # verwendet wieder alle Daten

set {*Variable*}

set *Variable* = *Wert*

zeigt oder setzt die Werte von Variablen. Wenn keine *Variable* gegeben ist, werden alle Variablen, deren Werte von **NULL** verschieden sind, angezeigt. Sind die Namen von Variablen angegeben, werden die Werte der entsprechenden Variablen angezeigt. Programmvariablen, die vom Benutzer nur gelesen, aber nicht verändert werden dürfen, werden mit *read-only* gekennzeichnet. In der Form **set** *Variable* = *Wert* wird der *Variablen* ein *Wert*, d. h. eine Zeichenkette zugewiesen.

Beispiele:**set i=456**
set j=2 + \$i
set i j
... *Variables*:
 $i = 456$
 $j = 2 + 456$

shift m

verschiebt die Daten in der aktiven Dimension zyklisch um m Datenpunkte nach rechts: $S_1, S_2, \dots, S_n \rightarrow S_{n-m+1}, S_{n-m+2}, \dots, S_n, S_1, S_2, \dots, S_{n-m}$. Eine zyklische Verschiebung um m' Datenpunkte nach links wird durch $m = -m'$ oder $m = n - m'$ erreicht. m ist ein ganzzahliger Ausdruck, der *modulo* n interpretiert wird und in dem k für den Index stehen kann, der über alle Datenpunkte in den nicht-aktiven Dimensionen läuft.

Beispiel: **shift \$n/2-nint(real(\$n-1)/(\$n_2-1)*(k-1)+1)**
verschiebt in einem zweidimensionalen Spektrum, in dem die Diagonale durch die Ecken verläuft, die Diagonale in die Mitte des Spektrums, wo sie mit Hilfe der Anweisung **smooth** anschließend unterdrückt werden kann (Friedrichs *et al.*, 1991).

show {*Variable*}

zeigt die Werte aller oder ausgesuchter *globaler* Variablen an. Wenn keine *Variable* angegeben ist, werden alle globalen Variablen, deren Werte von **NULL** verschieden sind, angezeigt. Sind die Namen von globalen Variablen angegeben, werden die Werte der

entsprechenden Variablen angezeigt. Programmvariablen, die vom Benutzer nur gelesen, aber nicht verändert werden dürfen, werden mit *read-only*, globale Variablen, die durch lokale Variablen gleichen Namens verborgen sind, mit *hidden* gekennzeichnet.

smooth *n* *Funktion* [*m*] {Option}

glättet die Daten durch Bildung des gleitenden Mittelwerts s'_k über die n vorangehenden und n folgenden Datenpunkte s_k , die mit der angegebenen *Funktion* gewichtet werden. In der *Funktion* f_k bezeichnet k den Index, der von $-n$ bis n läuft.

$$s'_k = \frac{f_{-n}s_{k-n} + \dots + f_0s_k + \dots + f_ns_{k+n}}{f_{-n} + \dots + f_0 + \dots + f_n} \quad [29]$$

Die folgenden Optionen sind möglich:

<i>Option</i>	= extrapolate	berechnet die $m \geq n$ Datenpunkte in den Randgebieten durch quadratische Extrapolation der geglätteten Daten,
	= circular	nimmt zum Glätten der Randbereiche periodische Daten an,
	= linear	verwendet zum Glätten der Randbereiche nur die verfügbaren Datenpunkte (z. B. für den 2. Datenpunkt die Datenpunkte 1, 2, 3, ..., $2+n$),
	= replace	ersetzt die Daten durch die geglätteten Daten,
	= subtract	subtrahiert die geglätteten von den ursprünglichen Daten.

Der Parameter m hat nur in Verbindung mit der Option **extrapolate** eine Bedeutung. Standardmäßig sind die Optionen **extrapolate** und **replace** gesetzt.

Beispiel: **smooth 20 cos(0.5*\$pi/19*k) 22 extrapolate subtract**

ist eine Methode, um ein Signal mit Frequenz null, z. B. die Wasserlinie, zu unterdrücken (Marion *et al.*, 1989). Am Rand werden die Daten über mehr Punkte extrapoliert als unbedingt nötig, um zu vermeiden, daß die — oft verfälschten — ersten Datenpunkte das Glätten beeinflussen können.

status [full] [data]

gibt Informationen über die vorhandenen Daten. Wenn die Option **full** gesetzt ist, werden auch der maximale Absolutwert und die Rauschintensität berechnet, angezeigt und den Programmvariablen **max** und **noise** zugewiesen. Die Option **data** veranlaßt, daß die Werte der aller Datenpunkte angezeigt werden, wenn es sich um nicht mehr als 2048 Zahlen handelt.

type *Dateiname*

zeigt den Inhalt des Makros (oder irgendeiner Textdatei) mit dem angegebenen *Dateiname* an. Für Standardmakros (vgl. S. 52ff.) und Dateien im aktuellen Verzeichnis braucht kein Pfad angegeben zu werden.

var *Variable* {*Variable*}

deklariert am Anfang eines Makros lokale Variablen. Im Gegensatz zu globalen Variablen sind lokale Variablen nur innerhalb des deklarierenden Makros, und in Makros sichtbar, die ausgehend vom deklarierenden Makro aufgerufen werden (es sei denn, ein solcher Makro deklarieren seinerseits lokale Variablen gleichen Namens). Die **var** Anweisung muß vor allen anderen Anweisungen (außer **parameter**) am Anfang eines Makros stehen und ist nicht interaktiv verwendbar.

write *Format* *Dateiname* {*Bereich*}

schreibt alle oder einen Teil der vorhandenen Daten in eine Ausgabedatei. Die Parameter haben folgende Bedeutung:

<i>Format</i>	= real	serielle Datei mit binär kodierten reellen Zahlen,
	= integer	serielle Datei mit binär kodierten ganzen Zahlen,
	= x32	serielle Datei mit binär kodierten ganzen Zahlen, wobei die Reihenfolge der 4 Bytes umgedreht ist,
	= text	Textdatei, die im FORTRAN-77 Format (1PE12.4) (Kießling &

	= easy8	Lowes, 1987) geschrieben wird, Submatrix 8-Bit Format des Programms EASY (Eccles <i>et al.</i> , 1991; C. Bartels und K. Wüthrich, unveröffentlicht),
	= easy16	Submatrix 16-Bit Format des Programms EASY.
<i>Dateiname</i>	Name der Ausgabedatei. Im Fall der <i>Formate easy8</i> und <i>easy16</i> werden zwei Dateien, eine Parameterdatei mit dem Namen <i>Dateiname.3D.param</i> und die eigentliche Submatrixdatei <i>Dateiname.3D.8</i> bzw. <i>Dateiname.3D.16</i> geschrie- ben. Eintragungen in der Parameterdatei, die sich auf die Kalibrierung des Spektrums beziehen, werden wenn möglich auf die entsprechenden Werte der Programmvariablen delta_k , w0_k und ppmmax_k (vgl. S. 50/52) ge- setzt, andernfalls wird das Spektrum von EASY als »unkalibriert« behandelt.	
<i>Bereich</i>	Bereich der Daten, der geschrieben wird; der erste <i>Bereich</i> bezieht sich auf die aktive Dimension, der zweite auf die zweite Dimension usw., wobei höch- stens so viele Bereiche, wie der Datensatz Dimensionen hat, vorkommen dürfen. Werden weniger Bereiche als Dimensionen angegeben, nimmt das Programm in den verbleibenden Dimensionen alle Datenpunkte mit. Ein <i>Bereich</i> kann in einer der folgenden Formen angegeben werden:	
	<i>n</i>	bezeichnet den Datenpunkt <i>n</i> ,
	<i>m . . n</i>	umfaßt die Datenpunkte <i>m, m + 1, ..., n</i> ,
	<i>m . .</i>	umfaßt die Datenpunkte <i>m, m + 1, ...</i> bis zum letzten Punkt,
	<i>. . n</i>	beinhaltet die Datenpunkte <i>1, 2, ..., n</i> ,
	*	steht für alle Datenpunkte.

PROSA-Programmvariablen

Die Routine zum Lesen der Kommandozeile im Programm PROSA bietet die Möglichkeit, Variablen, die mit »Shell-Variablen« im UNIX Betriebssystem vergleichbar sind, zu verwenden. Ein *Variablenname* besteht aus bis zu 20 Buchstaben, Ziffern oder Unterstreichungszeichen »_«. Der *Wert* einer Variablen ist stets eine Zeichenkette — auch die Ergebnisse arithmetischer Ausdrücke werden bei der Zuweisung zu einer Variablen in eine Zeichenkette umgewandelt — und wird in der Kommandozeile mit **\$Variable** oder **%Variable** bezeichnet.¹ Wie in FORTRAN-77 (Kießling & Lowes, 1987) lassen sich Teile von Zeichenketten mit **\$Variable(Anfang:Ende)** ansprechen, wobei *Anfang* und *Ende* ganzzahlige Ausdrücke sind, die das erste und letzte Zeichen der Teilkette angeben. Sollen einer Variable Buchstaben, Ziffern oder Unterstreichungszeichen unmittelbar folgen, muß sie in geschweifte Klammern eingeschlossen werden: **{ \$Variable }**. Einige Beispiele:

```

set x=4.6                # Wert der Variablen x setzen
set y=2.0               # Wert der Variablen y setzen
eval Summe=$x+$y       # Ausdruck auswerten
set t=ein Beispiel     # Wert der Variablen t setzen
set x y Summe t        # Werte anzeigen
... Variables:
  x          = 4.6
  y          = 2.0
  Summe      = 6.60000
  t          = ein Beispiel
print Dies ist $t: $x + $y = $Summe # Werte einsetzen
Dies ist ein Beispiel: 4.6 + 2.0 = 6.60000
print Ein zweites $t(5:12)!      # Verwendung einer Teilkette

```

¹ Wenn PROSA-Anweisungen in UNIX-»Shellskripts« vorkommen, ist die Form **%Variable** vorzuziehen, weil sie verhindert, daß PROSA-Variablen durch die UNIX-Shell ausgewertet werden.

Ein zweites Beispiel!

```
set t(5)=Programm           # Zuweisung an eine Teilkette  
print $t oder {$t}paket?   # Verwendung von »{ }«  
ein Programm oder ein Programmpaket?
```

Das Einsetzen der Werte von Variablen erfolgt in der Kommandozeile *von rechts nach links*. Dies ermöglicht zum Beispiel das Ansprechen »indexierter« Variablen in einer Schleife (Annahme: **\$ndim** = 2, **\$n_1** = 2048, **\$n_2** = 512):

```
do i 1 $ndim  
  print Dimension $i: $n_$i Punkte  
end do  
Dimension 1: 2048 Punkte  
Dimension 2: 512 Punkte
```

Typen von Variablen: *Programmvariablen* sind Variablen, die vom Programm PROSA (nicht nur vom Benutzer mittels **eval**, **set** usw.) gesetzt oder gelesen werden. *Schreibgeschützte* Programmvariablen können vom Benutzer nicht explizit verändert werden. Eine *globale Variable* ist immer sichtbar, außer wenn sie durch eine lokale Variable gleichen Namens verborgen wird. Alle Variablen, die nicht in einer **var** Anweisung genannt oder als Parameter an einen Makro übergeben wurden — insbesondere alle Programmvariablen — sind global. Im Gegensatz zu globalen sind *lokale Variablen* nur innerhalb des deklarierenden Makros, und in Makros sichtbar, die ausgehend vom deklarierenden Makro aufgerufen werden (es sei denn, ein solcher Makro deklarieren seinerseits lokale Variablen gleichen Namens). Lokale Variablen müssen in einer **var** Anweisung deklariert oder als Parameter an einen Makro übergeben werden (vgl. S. 52). Die folgende, alphabetische Liste enthält alle Programmvariablen:

check

legt fest, ob PROSA-Anweisungen, die eine Änderung der Daten zur Folge haben, lediglich getestet oder wirklich ausgeführt werden. Anweisungen werden ausgeführt, wenn **check** auf **NULL** gesetzt oder leer ist. Andernfalls, d. h. wenn **check** eine oder mehrere der folgenden Optionen enthält, werden die Anweisungen nur getestet. Die Optionen bestimmen, welche Fehler festgestellt werden:

memory Zu kleiner Daten- oder Arbeitsspeicher löst einen Fehler aus.

file Eingabedateien, die nicht existieren, oder Ausgabedateien, die nicht angelegt werden können, führen zu einem Fehler.

command Alle anderen Fehler werden festgestellt.

Die Option **command** ist immer aktiv. Um zum Beispiel den Speicherbedarf für die Ausführung eines Makros festzustellen, ist es nützlich, die Option **memory** *nicht* zu setzen, und erst nach dem Test die Programmvariablen **usedsize** und **usedwork** abzufragen. Falls während der Ausführung eines Makros eine neue Datei geschrieben und später wieder eingelesen wird, sollte für einen Test die Option **file** nicht gesetzt werden, weil sonst beim Versuch, das Lesen der Datei zu testen, ihr Nichtvorhandensein einen Fehler auslöst, der nicht auftritt, wenn der Makro wirklich ausgeführt wird. Makros werden am Einfachsten mit Hilfe des Standardmakros **job** (vgl. S. 53) vor dem Ausführen getestet.

delta_k

bezeichnet den Zeit- oder Frequenzschritt, d. h. den Abstand benachbarter Datenpunkte, in der Dimension k (in Sekunden im Zeit-, in Hertz im Frequenzbereich). Die Programmvariablen **delta_k** werden beim Einlesen von Daten im Format des Programms EASY (Eccles *et al.*, 1991; C. Bartels und K. Wüthrich, unveröffentlicht) gemäß den Werten in der Parameterdatei gesetzt und bei einer Fourier-Transformation entsprechend geändert.

echo

bestimmt, welche Anweisungen vor ihrer Ausführung auf die Standardausgabe kopiert werden. Die möglichen Werte sind:

NULL PROSA-spezifische Anweisungen (vgl. S. 39) in Makros werden angezeigt.

on Alle PROSA-spezifischen Anweisungen (vgl. S. 39) werden angezeigt.
full Alle Anweisungen werden angezeigt (in Makros mit zugehöriger Zeilennummer).
off Anweisungen werden nicht angezeigt.
Beim Start des Programms wird **echo** = **NULL** gesetzt.

erract

ist eine Programmvariable zur Fehlerbehandlung, deren Wert die Anweisung angibt, die beim Auftreten eines Fehlers *innerhalb eines Makros* ausgeführt wird.¹ Standardmäßig wird **exit** ausgeführt, d. h. das Programm kehrt zur interaktiven Eingabe zurück.

Beispiel: **set erract=chain show ; quit**

hat zur Folge, daß im Falle eines Fehlers die Werte aller globalen Variablen angezeigt werden und anschließend das Programm abgebrochen wird. Diese Art der Fehlerbehandlung kann nützlich sein, wenn PROSA nicht-interaktiv benutzt wird.

icmplx

besitzt die Werte 1 oder 2, je nachdem, ob die Daten in der aktiven Dimension reell oder komplex sind, und ist schreibgeschützt.

m

gibt das Produkt der Anzahl reeller Datenpunkte in den nicht-aktiven Dimensionen an und ist schreibgeschützt.

max

gibt den maximalen Absolutwert der Daten an und ist schreibgeschützt. **max** wird nur mit der Anweisung **status full** berechnet und gesetzt.

maxsize und **maxwork**

bezeichnen die Größe des Daten- und Arbeitsspeichers in Worten und sind schreibgeschützt.

n

gibt die Anzahl reeller oder komplexer Datenpunkte in der aktiven Dimension an und ist schreibgeschützt.

n_k

gibt die Anzahl reeller oder komplexer Datenpunkte in der Dimensionen *k* an und ist schreibgeschützt. **n₁** ist äquivalent zu **n**.

ndim

gibt die Anzahl Dimensionen an und ist schreibgeschützt.

noise

gibt die Rauschintensität an und ist schreibgeschützt. **noise** wird nur mit der Anweisung **status full** berechnet und gesetzt. Als Rauschintensität wird ein Schätzwert für den Median der absoluten Intensitäten der Datenpunkte berechnet.

phi0 und **phi1**

geben die konstante (ϕ_0) und lineare (ϕ_1) Phasenkorrektur an (vgl. Gl. [21]). **phi0** und **phi1** werden von der Anweisung **autophase** berechnet und gesetzt, können aber auch vom Benutzer selbst eingestellt werden.

¹ Fehler, die bei interaktiver Eingabe auftreten, werden angezeigt, und das Programm wird mit der Ausführung der nächsten Anweisung fortgesetzt.

pi

hat den Wert **3.141593** und ist schreibgeschützt.

ppmmax_k

bezeichnet die chemische Verschiebung (in ppm) des ersten Datenpunktes in der Dimension k . Die Programmvariablen **ppmmax_k** werden beim Einlesen von Daten im Format des Programms EASY (Eccles *et al.*, 1991; C. Bartels und K. Wüthrich, unveröffentlicht) gemäß den Werten in der Parameterdatei gesetzt.

timing

ist eine Programmvariable, deren Wert (in s) die Schwelle angibt, ab welcher die Rechenzeit für die Ausführung einer PROSA-Anweisung auf der Standardausgabe angezeigt wird.

usedsize und usedwork

bezeichnen die Größe des *gebrauchten* Daten- und Arbeitsspeichers in Worten. Am Anfang des Programms haben **usedsize** und **usedwork** den Wert 0 und werden mit der Ausführung jeder Anweisung, die mehr Daten- oder Arbeitsspeicher benötigt, als die Werte von **usedsize** und **usedwork** angeben, entsprechend erhöht, können aber auch vom Benutzer explizit geändert werden.

w0_k

bezeichnet die Spektrometerfrequenz (in MHz) in der Dimension k . Die Programmvariablen **w0_k** werden beim Einlesen von Daten im Format des Programms EASY (Eccles *et al.*, 1991; C. Bartels und K. Wüthrich, unveröffentlicht) gemäß den Werten in der Parameterdatei gesetzt.

PROSA-Standardmakros

Makros sind Dateien, die PROSA-Anweisungen enthalten, und durch Angabe ihres Dateinamens, eventuell gefolgt von Parametern, aufgerufen werden. Für Standardmakros und Makros im aktuellen Verzeichnis braucht der Dateiname keinen Pfad zu enthalten. Parameter können an einen Makro übergeben werden; innerhalb des Makros sind sie als lokale Variablen mit den Standardnamen **p1**, **p2** usw. verfügbar. Die Namen lassen sich mit Hilfe der Anweisung **parameter** (vgl. S. 44) abändern. Die Anzahl übergebener Parameter wird in der lokalen Variable **nparam** gespeichert. Makros können innerhalb von Makros aufgerufen werden. Die folgende, alphabetische Liste führt alle Standardmakros auf:

cflatt Methode $n \tau m [n_0 n_b]$

glättet die Basislinie im Frequenzbereich der aktiven Dimension mit dem FLATT-Verfahren (vgl. S. 33ff.). **cflatt** die Anweisung **flatten flatt** $n \tau . . .$ durch (vgl. S. 42), wobei die Basisfunktionen entsprechend dem Parameter *Methode* gewählt werden:

Methode = **cft** verwendet die trigonometrischen Funktionen, die den ersten m Datenpunkten vor der *komplexen* Fourier Transformation entsprechen, als Basisfunktionen;
= **rft** verwendet die trigonometrischen Funktionen, die den ersten m Datenpunkten vor der *reellen* Fourier Transformation entsprechen, als Basisfunktionen;
= **polynom** verwendet ein Polynom vom Grad m .

Für die Methoden **cft** und **rft** wird angenommen, daß die vorhandenen reellen Frequenzbereichsdaten einer Fourier-Transformation von n_0 komplexen Datenpunkten entstammen und einen Streifen bilden, dessen erster Punkt dem Punkt n_b in den ursprünglichen n_0 Datenpunkten entsprechen. Standardmäßig wird angenommen, daß noch alle fourier-transformierten Datenpunkte vorhanden sind, d. h. $n_0 = \mathbf{n}$ und $n_b = 1$.

chain *Anweisung* {Parameter} {; *Anweisung* {Parameter}}

führt eine Reihe von Anweisungen, die, getrennt durch Strichpunkte, auf einer Zeile stehen, nacheinander aus. Die Strichpunkte müssen auf beiden Seiten durch Leerzeichen abgetrennt sein. Dieser Makro ist zum Beispiel nützlich, um zur Fehlerbehandlung die Programmvariable **erract** (vgl. S. 51) zu setzen. Er ist wie folgt aufgebaut:

```
var cmd i
set cmd=
do i 1 $nparam
  if ('$p$i'.eq. ';') then
    $cmd
    set cmd=
  else
    set cmd=$cmd $p$i
  end if
end do
if ('$cmd'.ne. ' ') then
  $cmd
end if
```

checker *Makroname* {Parameter}

untersucht den angegebenen Makro auf Fehler, ohne die eigentlichen Berechnungen durchzuführen. Der Wert der Programmvariablen **check** (vgl. S. 50) bestimmt die Art der Fehler, die entdeckt werden. Standardmäßig, d. h. wenn **check** beim Aufruf des Makros den Wert **NULL** besitzt, wird **check** auf **command file memory** gesetzt, womit alle möglichen Fehler gefunden werden. Tritt kein Fehler auf, werden die benötigten Größen des Daten- und Arbeitsspeichers angezeigt und, wenn diese die vorhandenen Speichergrößen übersteigen, Warnungen ausgegeben.

dummyscal

setzt die Programmvariablen **delta_k**, **ppmmax_k** und **w0_k** (vgl. S. 50/52) auf Standardwerte (**delta_k** = **w0_k** = 1000.0 und **ppmmax_k** = *Anzahl Datenpunkte in Dimension k*) und verhindert damit, daß das Programm EASY ein Spektrum als »unkalibriert« betrachtet.

hilbert

führt in der aktiven Dimension eine Hilbert-Transformation durch (Ernst, 1969). Reelle Daten werden in komplexe Daten verwandelt, die den Kramers-Kroning Relationen genügen und deren Realteil mit den ursprünglichen reellen Daten übereinstimmt.

im

ersetzt komplexe Daten in der aktiven Dimension durch ihren Imaginärteil. Der Realteil geht verloren. Reelle Daten bleiben unverändert.

job *Makroname* {Parameter}

untersucht den angegebenen Makro zunächst auf Fehler, ohne die eigentlichen Berechnungen durchzuführen, und führt ihn, wenn er fehlerfrei ist, anschließend aus. Der Wert der Programmvariablen **check** (vgl. S. 50) bestimmt die Art der Fehler, die entdeckt werden. Standardmäßig, d. h. wenn **check** beim Aufruf des Makros den Wert **NULL** besitzt, wird **check** auf **command file** gesetzt. Tritt nicht bereits vorher ein Fehler auf, werden nach dem Test die benötigten Größen des Daten- und Arbeitsspeichers angezeigt und, wenn diese die vorhandenen Speichergrößen übersteigen, ein entsprechender Fehler ausgelöst. Im Falle eines Fehlers werden die Werte aller globalen Variablen aufgelistet und das Programm abgebrochen. **job** eignet sich besonders, um Makros im nicht-interaktiven Betrieb ablaufen zu lassen.

phase ϕ_0 [ϕ_1]

wendet auf die komplexen Daten in der aktiven Dimension eine konstante (ϕ_0) und lineare

(ϕ_1) Phasenkorrektur an (vgl. Gl. [21]). Die Werte von ϕ_0 und ϕ_1 müssen in Grad angegeben werden. Wenn die Phasenkorrekturparameter bekannt sind, ist es im allgemeinen effizienter, sie bereits bei der Fourier-Transformation anzugeben (vgl. Anweisung **ft**, S. 42), als den Makro **phase** aufzurufen.

reduceppm *Bereich* {*Bereich*}

arbeitet wie die Anweisung **reduce** (vgl. S. 46), außer daß die *Bereiche* in ppm anstatt in Datenpunkten anzugeben sind. Der Makro kann nur gebraucht werden, wenn die Programmvariablen **delta_k**, **ppmmax_k** und **w0_k** (vgl. S. 50/52) gesetzt sind.

savequit {*Dateiname*}

zeigt die Werte aller globalen Variablen an, schreibt die aktuellen Daten im Format **real** (vgl. S. 48) in die Datei mit dem angegebenen *Dateinamen* (Standardname: **savequit.out**) und verläßt das Programm. Der Makro ist als Fehlerbehandlungsroutine für längere Rechnungen nützlich.

Beispiel: **set erract=savequit /tmp/prosa.tmp**

aktiviert **savequit** als Fehlerbehandlungsroutine, wobei die Daten im Falle eines Fehlers in die Datei **/tmp/prosa.tmp** geschrieben werden.

scale *Methode* *Intensität*

skaliert die Daten so, daß im Falle *Methode* = **max** die maximale absolute Intensität bzw. im Falle *Methode* = **noise** die Rauschintensität des Spektrums den Wert *Intensität* annimmt. Standardwerte: 500'000 für die maximale absolute Intensität bzw. 100 für die Rauschintensität.

selectppm {*Bereich*}

arbeitet wie die Anweisung **select** (vgl. S. 47), außer daß die *Bereiche* in ppm anstatt in Datenpunkten anzugeben sind. Der Makro kann nur gebraucht werden, wenn die Programmvariablen **delta_k**, **ppmmax_k** und **w0_k** (vgl. S. 50/52) gesetzt sind.

suppress [*Gewichtung* [*n*]]

unterdrückt Signale der Frequenz null (z. B. die Wasserlinie) durch Subtraktion der geglätteten von den originalen Zeitbereichsdaten mit Hilfe der Anweisung **smooth** (vgl. S. 48). Die geglätteten Daten s'_k berechnen sich aus den Originaldaten gemäß

$$s'_k = \frac{f_{-n}s_{k-n} + \dots + f_0s_k + \dots + f_ns_{k+n}}{f_{-n} + \dots + f_0 + \dots + f_n} \quad [30]$$

Die folgenden *Gewichtungen* sind möglich:

<i>Gewichtung</i>	<i>Gewichtungsfunktion</i>	<i>Name</i>
cos	$f_k = \cos(\pi k / (2(n+1)))$	Kosinus-Gewichtung
gauss	$f_k = e^{-4(k/n)^2}$	Gaußsche Gewichtung
equal	$f_k \equiv 1$	Gleichgewichtung

window *Typ* {*Parameter*}

wendet gebräuchliche Fensterfunktionen an (DeMarco & Wüthrich, 1976; Ernst *et al.*, 1987):

<i>Typ</i>	<i>Parameter</i>	<i>Filterfunktion</i>	<i>Name</i>
cos	–	$\cos(\pi t / 2)$	Kosinus-Fenster
cos2	–	$\cos(\pi t / 2)^2$	Kosinusquadrat-Fenster
exp	<i>L</i>	$e^{-\pi L n \Delta t}$	exponentielle Linienverbreiterung

<i>Typ</i>	<i>Parameter</i>	<i>Filterfunktion</i>	<i>Name</i>
gauss	$L G$	$e^{-\pi L n \Delta t(1-t/2G)}$	Lorentz-Gauß Transformation
hamming	–	$0.54 + 0.46 \cos \pi t$	Hamming-Fenster
hanning	–	$0.5 + 0.5 \cos \pi t$	»Hanning«-Fenster
sin	ϕ	$\sin(\phi - (\phi - \pi)t)$	verschobene »Sinusglocke«
sin2	ϕ	$\sin(\phi - (\phi - \pi)t)^2$	verschobene »Sinusquadratglocke«

Die Symbole in der Tabelle haben folgende Bedeutungen: $t = (k-1)/n$, wobei $k = 1, \dots, n$ über alle n Datenpunkte in der aktiven Dimension läuft; L gibt die Linienverbreiterung in Hz an; G legt das Maximum der Lorentz-Gauß-Filterfunktion bei $t = G$ fest; Δ entspricht dem Wert der Programmvariablen **delta_1** (vgl. S. 50), die den Zeitschritt (in s) von einem Datenpunkt zum nächsten bezeichnet; ϕ steht für die Verschiebung der Sinus- oder Sinusquadratglocke in Grad, so ist zum Beispiel **window sin 90** äquivalent zu **window cos**.

Konformationseinschränkungen aus Kern-Overhausereffektspektren

Die Auswertung mehrdimensionaler Kernresonanzspektren für die Bestimmung der Lösungsstruktur eines Proteins — früher manuell anhand von Konturzeichnungen auf Papier, heute in unserem Labor mit Hilfe des interaktiven Programms EASY (Eccles *et al.*, 1991) und einiger Unterstützungsprogramme durchgeführt — umfaßt zwei grundlegende Schritte: erstens die sequentielle Resonanzzuordnung (Dubs *et al.*, 1979; Wider *et al.*, 1982; Wüthrich *et al.*, 1982; Wüthrich, 1986), für die das Programm EASY eine Vielzahl interaktiver Werkzeuge und einige automatisierte Verfahren anbietet (Eccles, 1990; Eccles *et al.*, 1991). Zweitens das Sammeln von Konformationseinschränkungen für die Strukturrechnung. Dieses Kapitel beschränkt sich auf die Auswertung von Kern-Overhausereffektspektren (NOESY) im Hinblick auf die nachfolgende Strukturrechnung und behandelt die Bestimmung der Stärke von Kern-Overhausereffekten durch Integration der entsprechenden, möglicherweise überlappenden Signale im NOESY-Spektrum (Eccles *et al.*, 1991), die Messung von $^3J_{HN\alpha}$ -Kopplungskonstanten durch inverse Fourier-Transformation von In-Phase Dubletts (Szyperski *et al.*, 1992a), die Umwandlung von NOESY-Signalintensitäten in obere Distanzschranken (Güntert *et al.*, 1991a) und die Zuordnung von Signalen im NOESY-Spektrum aufgrund einer dreidimensionalen Struktur (Berndt *et al.*, 1993).

Intensitätsbestimmung bei überlappenden Signalen

In einem zweidimensionalen [^1H - ^1H]-NOESY-Spektrum (Anil-Kumar *et al.*, 1980) eines Proteins treten oft überlappende Signale auf, deren Volumina nicht einfach durch Summation der Datenpunkte im »Signalgebiet« quantifiziert werden kann. Daher bietet das Programm EASY (Eccles *et al.*, 1991) eine automatisierte Integrationsmethode an, die isolierten Signalen Referenzlinienformen entnimmt, um damit die Volumina überlappender Signale zu bestimmen (Denk *et al.*, 1986; Eccles *et al.*, 1991). Dieses Dekonvolutionsverfahren

beruht darauf, daß alle Signale eines gegebenen Protons entlang einer Richtung im Spektrum die gleiche Linienform aufweisen.

Die Integration der Signale in einem zweidimensionalen NOESY-Spektrum umfaßt drei Schritte: Zunächst müssen für alle Protonen — manuell oder automatisch — Referenzlinienformen entlang der ω_1 - und ω_2 -Richtungen ausgewählt werden, dann gruppiert das Programm die Signale in Überlappcluster, wobei zwei Signale als überlappend gelten, wenn sich die von den Referenzlinienformen definierten, rechteckigen Signalgebiete überschneiden. Schließlich werden die Volumina aller Signale in einem Überlappcluster durch eine lineare Anpassung der Linienformen, die sich aus den entsprechenden ω_1 - und ω_2 -Referenzlinienformen ergeben, an das experimentelle Spektrum bestimmt. Um für ein Signal die zugehörigen Referenzlinienformen finden zu können, muß die Zuordnung der Signale bekannt sein.

Der Benutzer des Programms EASY kann Referenzlinienformen interaktiv auswählen, aber im allgemeinen ist es einfacher und zuverlässiger, diese vom Programm automatisch suchen zu lassen. Zu diesem Zweck wählt das Programm für ein gegebenes Proton aus den isolierten, nicht überlappenden Signalen das stärkste als Referenzlinienform aus.¹ Die Ausdehnung einer Referenzlinienform wird von den ersten Datenpunkten auf beiden Seiten des Signalmaximums festgelegt, deren Intensität eine vorgegebene Schwelle unterschreitet.

Es bezeichne $L_i^r(\omega_i)$ die Referenzlinienform der Resonanz r in der ω_i -Richtung ($i = 1, 2$), $S(\omega_1, \omega_2)$ die Intensität des zweidimensionalen NOESY-Spektrums an der Stelle (ω_1, ω_2) und Γ das Gebiet eines Clusters von $m \geq 1$ überlappenden Signalen, wo das Signal p ($p = 1, \dots, m$) die Resonanz $r_1(p)$ in der ω_1 -Richtung mit der Resonanz $r_2(p)$ in der ω_2 -Richtung verknüpft. Die Volumina V_1, \dots, V_m der Signale, die den Überlappcluster bilden, ergeben sich durch Minimierung der quadratischen Abweichung

$$\chi^2(V_1, \dots, V_p) = \sum_{(\omega_1, \omega_2) \in \Gamma} \left[S(\omega_1, \omega_2) - \sum_{p=1}^m V_p L_1^{r_1(p)}(\omega_1) L_2^{r_2(p)}(\omega_2) \right]^2 \quad [31]$$

zwischen dem experimentellen und dem aufgrund der Referenzlinienformen berechneten Spektrum mit Hilfe von Standardmethoden (Press *et al.*, 1986).

¹ Signale in der Nähe der Diagonalen bleiben unberücksichtigt, weil ihre Linienform oft durch Ausläufer von Signalen auf der Diagonalen verfälscht sind.

Messung skalarer Spin-Spin-Kopplungskonstanten durch inverse Fourier-Transformation von In-Phase-Multipletts

Neben oberen Schranken für interatomare Distanzen, die aus Kern-Overhauserereffekten folgen (vgl. S. 61ff.), sind skalare Kopplungskonstanten 3J zwischen Spins, die durch drei kovalente Bindungen voneinander getrennt sind, wichtige Quellen von Information über die Konformation eines Proteins, weil sie aufgrund der Karplus-Beziehung (Karplus, 1963; vgl. Gl. [37–39]) Einschränkungen von Diederwinkeln liefern. Traditionell werden skalare Kopplungskonstanten durch Messen des Abstands der Komponenten in Anti-Phase-Multipletts bestimmt (Marion & Wüthrich, 1983; Kessler *et al.*, 1988). Obwohl gut geeignet für manuelle Auswertung, weist diese Methode große Nachteile auf, die daher rühren, daß in Proteinen die Linienbreiten der Protonenresonanzen Werte von der gleichen Größenordnung wie die 3J -Kopplungskonstanten annehmen und es daher zu teilweiser Auslöschung der Signalkomponenten unterschiedlichen Vorzeichens kommt. Dies führt zu einer systematischen Überschätzung der Kopplungskonstante (Neuhaus *et al.*, 1985) und zu einer Verringerung der Signalintensität. Die hier beschriebene Methode (Szyperski *et al.*, 1992a) zur Bestimmung von ${}^3J_{HN\alpha}$ -Kopplungskonstanten durch inverse Fourier-Transformation geht dagegen von In-Phase-Signalen aus, für die keine gegenseitige Auslöschung der Multiplett-Komponenten auftritt. Sie bietet den zusätzlichen Vorteil, daß im Rahmen der Strukturbestimmung mit homonuklearen Methoden ohnehin ein [${}^1\text{H}$, ${}^1\text{H}$]-NOESY-Spektrum (Anil-Kumar *et al.*, 1980) — das (außer bei Glycin) In-Phase-Dubletts für die Resonanzen der Amidprotonen enthält — mit gutem Signal-zu-Rauschen Verhältnis aufgenommen wird, um Distanzeinschränkungen zu sammeln, so daß dann sämtliche konformativen Einschränkungen demselben Datensatz entstammen. Für Proteine, die mit ${}^{15}\text{N}$ markiert sind, können die ${}^3J_{HN\alpha}$ -Kopplungskonstanten auch aus [${}^{15}\text{N}$, ${}^1\text{H}$]-COSY-Spektren (Bodenhausen & Ruben, 1980) bestimmt werden, da diese ebenfalls In-Phase-Dubletts enthalten.¹

Außer bei Glycin führt in einem [${}^1\text{H}$, ${}^1\text{H}$]-NOESY-Spektrum die skalare Kopplung des Amid- mit dem α -Proton zu einer einfachen $\cos(\pi J_{HN\alpha} t)$ -Modulation aller Amidprotonensignale während der Datenaufnahme. Im allgemeinen kann in einem NOESY-Spektrum für jedes Amidproton mehr als ein Signal beobachtet werden, so daß die Analyse der Dublett-Feinstruktur nicht wie bei [${}^1\text{H}$, ${}^1\text{H}$]-COSY-Spektren (Rance *et al.*, 1983) durch die Überlappungen im »Fingerabdruck«, d. h. im der Gebiet der Signale zwischen Amid- und α -Protonen,

¹ Dabei ist es wichtig, entweder eine Pulssequenz zu verwenden, die auf rein absorptive Linienformen führt, oder dispersive Anteile bei der Auswertung zu berücksichtigen (Szyperski *et al.*, 1992a).

begrenzt wird. Außerdem lassen sich mehrere Signale aufaddieren, um das Signal-zu-Rauschen Verhältnis für die Bestimmung der Kopplungskonstanten zu verbessern (Bild 9A und B). Um die ${}^3J_{HN\alpha}$ -Kopplungskonstante eines gegebenen Aminosäurerestes zu bestimmen, werden nur jene Datenpunkte zurückbehalten, die über dem Rauschpegel liegen (Bild 9B). Eine lineare Basislinienkorrektur bringt die beiden äußersten Punkte des Signalbereichs auf Null. Nachdem die Basislinie auf beiden Seiten des Signals mit Nullen erweitert wurde, erfolgt eine inverse Fourier-Transformation. Der Nullpunkt des

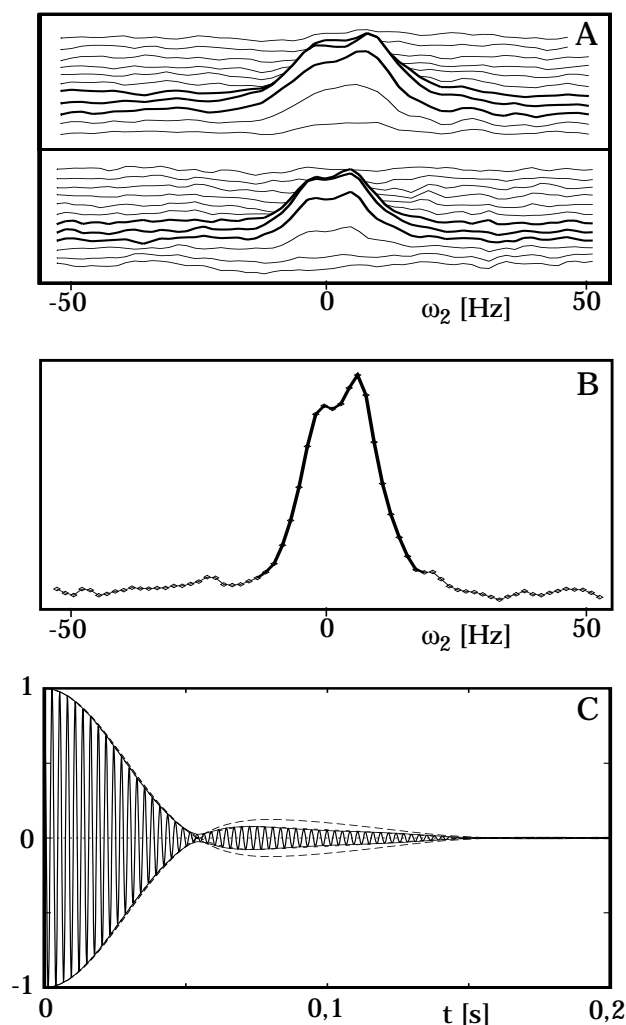


Bild 9. Bestimmung von ${}^3J_{HN\alpha}$ -Kopplungskonstanten durch inverse Fourier-Transformation von In-Phase-Dubletts, gezeigt am Beispiel zweier $[{}^1H-{}^1H]$ -NOESY-Signale zwischen den diastereotopen β -Protonen und dem Amidproton von Cys 22 in Hirudin(1-51):

(A) Ansicht der beiden NOESY-Signale. Die sechs fett gedruckten Zeilen werden zur Kopplungskonstantenbestimmung verwendet.

(B) Summe der ausgewählten Zeilen aus A. Der Signalbereich, der fouriertransformiert wird, ist fett gedruckt.

(C) Signal im Zeitbereich, das durch inverse Fourier-Transformation aus dem Signalbereich in B entstand. Zusätzlich ist die Einhüllende des Signals als durchgezogene Linie, die Einhüllende, die den anfänglichen Schätzwerten von Kopplungskonstante und Linienbreite entspricht, als gestrichelte Linie und die durch nicht-lineare Anpassung erhaltene Einhüllende als gestrichelte Linie, die fast mit der durchgezogenen zusammenfällt, gezeichnet.

Frequenzbereichs wird etwa 1000 Hz vom Signal entfernt gewählt, um die Modulation des Signals im Zeitbereich aufgrund der skalaren Kopplung nicht durch unerwünschte Modulationen zu stören, die auftraten, wenn der Nullpunkt nahe, nicht jedoch genau in der Signalmittle läge. Das erste lokale Mini-

mum der Einhüllenden des Signals im Zeitbereich (Bild 9C) liefert einen Schätzwert $J^{(0)}$ für die Kopplungskonstante, und aus der vollen Breite auf halber Höhe $\Delta\nu_{1/2}$ des Signals im Frequenzbereich ergibt sich ein Schätzwert für die Linienbreite $\Gamma^{(0)} = \Delta\nu_{1/2} - J^{(0)}$. Diese werden nun mit Hilfe des 'Levenberg-Marquardt Algorithmus' (Press et al., 1986) durch eine nicht-lineare Anpassung iterativ verbessert, in dem die quadratische Abweichung der simulierten von der experimentell erhaltenen Einhüllenden des Signals im Zeitbereich minimiert wird. Im allgemeinen läßt sich eine gute Übereinstimmung der beiden Einhüllenden erzielen (Bild 9C). Zur Simulation der Einhüllenden nach der inversen Fourier-Transformation müssen die Effekte der diskreten Fourier-Transformation, die benutzten digitalen Filterfunktionen, die Erweiterung mit Nullen im Zeitbereich und insbesondere die Auswirkungen des Herausschneidens des Signalbereichs (Bild 9B) berücksichtigt werden. Die simulierte Einhüllende muß daher mit Hilfe aller Schritte, die zur experimentellen Einhüllenden geführt haben, erzeugt werden. Für ein In-Phase-Dublett beschreibt

$$s(t) = \cos(\pi J_{HN\alpha}) e^{-\pi\Gamma t}, \quad t \in [0, t_{aq}] \quad [32]$$

unter der Annahme eines einfachen exponentiellen Relaxationsverhaltens die Einhüllende des Signals in der Zeitdomäne. t_{aq} bezeichnet die experimentelle Datenaufnahmezeit. Ausgehend von $s(t)$ in Gl. [32] werden die gleichen Filterfunktionen und die gleiche Erweiterung mit Nullen wie für die experimentellen Daten angewandt. Eine Fourier-Transformation wird durchgeführt und die erhaltene Resonanzlinie genau so wie das experimentelle Signal behandelt. Schließlich liefert eine inverse Fourier-Transformation die simulierte Einhüllende, welche an die experimentelle angepaßt werden kann. Ausgehend von den entsprechenden partiellen Ableitungen von $s(t)$ in Gl. [32] ergeben sich analog die für den Levenberg-Marquardt Algorithmus benötigten partiellen Ableitungen der simulierten Einhüllenden nach der Kopplungskonstanten und der Linienbreite.

Das Verfahren ist im Programm INFIT (Szyperski *et al.*, 1992a) implementiert und benötigt nur wenige Sekunden Rechenzeit pro Kopplungskonstante. Die Auswahl der geeigneten Signale trifft der Benutzer mit Hilfe des interaktiven Programms EASY (Eccles *et al.*, 1991). Die Genauigkeit der Kopplungskonstantenbestimmung ist vergleichbar mit derjenigen, die mit einer Serie von J -modulierten [^{15}N , ^1H]-COSY-Spektren erreicht werden kann (Neri *et al.*, 1990; Billeter *et al.*, 1992).

Umwandlung von Signalintensitäten in Distanzschranken

Der Kern-Overhauserereffekt — die Übertragung von Magnetisierung zwischen Spins aufgrund der Dipol-Dipol Wechselwirkung in einem sich bewegenden Molekül (Solomon, 1955) — hängt von der Distanz r zwischen zwei Spins und der Rotationskorrelationszeit τ_c so ab, daß seine Stärke proportional zu

$$\left\langle \frac{1}{r^6} \right\rangle f(\tau_c) \quad [33]$$

ist. Die Mittelung » $\langle \rangle$ « deutet an, daß für Moleküle, die nicht völlig starr sind, der Abstand zwischen den beiden Spins nicht mehr konstant zu sein braucht und deshalb geeignet gemittelt werden muß. Die übrige Abhängigkeit der Magnetisierungsübertragung von der Bewegung geht über die Funktion $f(\tau_c)$ ein, die in nicht-trivialer Weise den Einfluß globaler und interner Bewegungen des Moleküls beinhaltet (Abragam, 1961; McConnell, 1987; Luginbühl, 1991). Weil globuläre Proteine bis auf Seitenketten an der Oberfläche oder unbestimmte Segmente der Polypeptidkette im allgemeinen relativ starr sind, wird in der Strukturrechnung meistens¹ davon ausgegangen, daß eine starre Konformation existiert, die alle Einschränkungen erfüllt (Wüthrich, 1986), und interne Beweglichkeiten zu einer Verringerung der Kern-Overhauserereffekte führen, so daß ihre Interpretation als *obere* Schranken interatomarer Distanzen eine robuste Strukturbestimmungsmethode ergibt (Braun *et al.*, 1981; Wüthrich, 1986). Auf dieser Grundlage dient das Programm CALIBA (Güntert *et al.*, 1991a) der »Kalibrierung« von Distanzeinschränkungen, d. h. der Umwandlung der Intensitäten von Kern-Overhauserereffekten in obere Distanzschranken.

Das Programm CALIBA verwendet in seiner jetzigen Form zwei Typen von »Kalibrierkurven«, die den Zusammenhang zwischen dem Volumen V eines Signals im NOESY-Spektrum (Anil-Kumar *et al.*, 1980) und der oberen Schranke o für die Distanz zwischen den beiden betroffenen Spins: Entweder

$$V = \frac{C}{o^n}, \quad [34]$$

wo C eine Konstante ist, die von der Skalierung des Spektrums abhängt, und der Exponent n typischerweise 4–6 ist, oder das Modell gleichverteilter Distanzen (Braun *et al.*, 1981),

¹ Eine Ausnahme bilden Moleküldynamikrechnungen, in denen verlangt wird, daß die Einschränkungen nicht zu jedem Zeitpunkt sondern nur im Zeitmittel erfüllt sein müssen (Torda *et al.*, 1989; 1990; 1993). Experimentell zeigen sich interne Bewegungen von Proteinen zum Beispiel auch in den Relaxationszeiten (Lipari & Szabo, 1982; Szyperski *et al.*, 1993a).

$$V = \frac{C}{o-d_0} \int_{d_0}^o \frac{dr}{r^6} = \frac{C}{o-d_0} \left(\frac{1}{d_0^5} - \frac{1}{o^5} \right), \quad [35]$$

wo C wiederum eine Konstante und $d_0 = 1.9 \text{ \AA}$ den kleinsten, sterisch erlaubten Abstand zweier Protonen bezeichnet. Gl. [34] entspricht im Fall $n = 6$ der Annahme eines starren Moleküls; empirisch haben sich auch Exponenten $n < 6$ bewährt, insbesondere, wenn äußere Seitenkettenprotonen im Spiel sind. Gl. [35] entspricht dem Extremfall großer interner Bewegungen, denn es wird angenommen, daß die Distanz zwischen den beiden Spins jeden Wert zwischen d_0 und einer oberen Schranke o mit gleicher Wahrscheinlichkeit annimmt.

Die Eingabe für das Programm CALIBA besteht aus den Signalvolumina eines oder mehrerer NOESY-Spektren im Format des Programms EASY (Eccles *et al.*, 1991). Volumina aus verschiedenen Spektren, die bei unterschiedlichen experimentellen Bedingungen aufgenommen wurden, skaliert das Programm mit vom Benutzer gegebenen Faktoren. Die Volumina von Signalen, die von mehreren Spins mit entarteten chemischen Verschiebungen herrühren, werden durch die Anzahl entarteter Spins dividiert.¹ Beziehen sich mehrere Signale auf die gleiche Distanz, verwendet das Programm entweder das stärkste Signal oder ihren Mittelwert zur Bestimmung der oberen Distanzschranke — je nach Wunsch des Benutzers. Um eine optimale empirische Kalibrierung zu erhalten, können der Typ der Kalibrierkurve — Gl. [34] oder [35] — und die Konstanten C und C für verschiedene Klassen interatomarer Distanzen (Wüthrich, 1986) und für Signale ohne und mit Methylgruppen gesondert festgelegt werden. Um sinnvolle obere Schranken o auch für sehr starke oder sehr schwache Kern-Overhausereffekte zu erhalten, begrenzt sie das Programm auf einen vom Benutzer vorgegebenen Bereich — zum Beispiel $2.4 \leq o \leq 5.0 \text{ \AA}$. Einschränkungen, die aus Resonanzlinien mehrerer Spins mit entarteten chemischen Verschiebungen stammen, bezieht CALIBA auf das entsprechende Pseudoatom und erhöht die obere Schranke um die zugehörige Pseudoatomkorrektur (Wüthrich *et al.*, 1983). Das Programm erzeugt eine Datei mit oberen Distanzschranken (vgl. S. 77), die von den Programmen HABAS (vgl. S. 94ff.; Güntert *et al.*, 1989), DIANA (vgl. S. 104ff.; Güntert *et al.*, 1991a) und GLOMSA (vgl. S. 100ff.; Güntert *et al.*, 1991a) gelesen werden kann.

Um die Wahl der Kalibrierkurven zu optimieren, wenn bereits eine vorläufige Struktur vorliegt, verfügt das Programm CALIBA über die Möglichkeit, die Signalvolumina doppelt-logarithmisch gegen die mittlere entsprechende Distanz im Bündel von Konformeren aufzutragen. Mit Hilfe dieser grafischen Darstellung können die Kalibrierkurven aufgrund *aller* Signale meist besser festgelegt werden als mit wenigen Bezugssignalen, für die der Abstand bekannt ist (oder zu sein scheint).

¹ Dies gilt nicht für Methylgruppen, die gesondert behandelt werden.

Neben interner Bewegung kann auch die »Spindiffusion« (Kalk & Berendsen, 1976; Gordon & Wüthrich, 1978; Wagner & Wüthrich, 1979; Wüthrich, 1986), die daher rührt, daß in einem Molekül nicht isolierte Paare sondern ein ganzes Netz von Spins miteinander wechselwirken, Schwierigkeiten bei der Interpretation von Kern-Overhausereffekten verursachen. Daher treten vor allem in NOESY Spektren mit langen Mischzeiten Signale auf, die nicht auf direkter Magnetisierungsübertragung beruhen, sondern auf dem Fluß von Magnetisierung über benachbarte Spins. Die Spindiffusion kann mittels einer Relaxationsmatrix erfaßt (Solomon, 1955; Macura & Ernst, 1980; Ernst *et al.*, 1987) und in der Strukturverfeinerung berücksichtigt werden (Dobson *et al.*, 1982; Keepers & James, 1984; Boelens *et al.*, 1989; Yip & Case, 1989; Mertz *et al.*, 1991); für die Strukturbestimmung von Proteinen bleibt jedoch die Verwendung kurzer Mischzeiten in NOESY-Spektren vorläufig die einfachste Methode, um Spindiffusionseffekte zu minimieren (Anil-Kumar *et al.*, 1981).

Signalzuordnungen aufgrund dreidimensionaler Strukturen

Die Zuordnung möglichst vieler Signale im NOESY-Spektrum ist ein grundlegender Schritt in einer Lösungsstrukturbestimmung, weil er obere Schranken für interatomarer Distanzen liefert, aus denen sich die dreidimensionale Struktur des Proteins berechnen läßt (Wüthrich, 1986). Träten keine Entartungen der chemischen Verschiebungen von Protonen auf und wären die Signalpositionen immer präzise bestimmt, könnten aufgrund der sequentiellen Resonanzzuordnung alle Signale im NOESY-Spektrum sofort zugeordnet werden. In der Praxis führen die beiden genannten Schwierigkeiten dazu, daß nur ein Teil (ca. 30–50%) der Signale eindeutig zwei Protonen oder Methylgruppen zugeordnet werden können. Diese werden verwendet, um eine grobe dreidimensionale Struktur zu berechnen, die, weil ein Signal im NOESY-Spektrum nur auftreten kann, wenn die wechselwirkenden Spins weniger als etwa 5 Å voneinander entfernt sind, dazu dienen kann, Mehrdeutigkeiten in der Signalzuordnung aufzulösen. Dies wird mit dem Programm ASNO (Berndt *et al.*, 1993) erreicht, das aufgrund der chemischen Verschiebungen mögliche Zuordnungen von Signalen im NOESY-Spektrum eliminiert, wenn der Abstand zwischen den betroffenen Spins einen vom Benutzer festgelegten Wert überschreitet.

Die Eingabe zum Programm ASNO besteht aus einer Liste der chemischen Verschiebungen der Protonen, einer Liste der chemischen Verschiebungskordinaten der Signale im NOESY-Spektrum — beide im Format des Programms EASY (Eccles *et al.*, 1991) — und einem Bündel von Konformeren, wie sie zum

Beispiel mit dem Programm DIANA (Güntert *et al.*, 1991a; vgl. 104ff.) aufgrund eines vorläufigen, noch unvollständigen Satzes von Einschränkungen berechnet wurden. Zusätzlich gibt der Benutzer an: die maximalen Abweichungen $\Delta\omega_1$ und $\Delta\omega_2$ zwischen entsprechenden Koordinaten der Signale und chemischen Verschiebungen der Protonen für mögliche Zuordnungen aufgrund übereinstimmender chemischer Verschiebungen; die maximale Proton-Proton Distanz d_{\max} , für die noch ein Kern-Overhauserereffekt zu erwarten ist; und die minimale Anzahl c_{\min} von Konformeren aus dem Bündel, das die Lösungsstruktur bildet, in denen eine Proton-Proton Distanz kürzer als d_{\max} sein muß, damit die zugehörige Signalzuordnung akzeptiert wird. Für ein Signal p mit den chemischen Verschiebungs koordinaten ω_1^p und ω_2^p bestimmt ASNO zunächst alle aufgrund passender chemischer Verschiebungen mögliche Zuordnungen, d. h. alle Paare (α, β) von Einträgen in der Protonenliste, die die Bedingung

$$\left(\frac{\omega_1^p - \omega^\alpha}{\Delta\omega_1}\right)^2 + \left(\frac{\omega_2^p - \omega^\beta}{\Delta\omega_2}\right)^2 \leq 1 \quad [36]$$

erfüllen. ω^α und ω^β bezeichnen die chemischen Verschiebungen der Protonen. Von den Signalzuordnungen, die sich auf passende chemische Verschiebungen gründen, streicht das Programm anschließend all jene, für die die Distanz zwischen den Protonen α und β in weniger als c_{\min} Konformeren kleiner als d_{\max} ist. Treten Pseudoatome an die Stelle einer Gruppe von Protonen mit entarteten chemischen Verschiebungen, erhöht das Programm ASNO die Grenze d_{\max} um die entsprechende Pseudoatomkorrektur (Wüthrich *et al.*, 1983). Typische, konservative Werte für die benutzerabhängigen Parameter sind: $\Delta\omega_1 = \Delta\omega_2 = 0.02$ ppm, $d_{\max} = 5$ Å und $c_{\min} = 1$ aus 20 Konformeren. Die mit dem Programm ASNO erhaltenen Zuordnungen können im Programm EASY dargestellt, überprüft und gegebenenfalls geändert werden. So neu zugeordnete Kern-Overhauserereffekte führen zu weiteren Distanzeinschränkungen, die im nächsten Durchgang der Strukturrechnung eine genauer bestimmte Struktur ergeben. Dieser Prozeß kann iterativ fortgeführt werden.

Bild 10, das Bündel von je 10 Konformeren des Proteins Toxin K der Schwarzen Mamba zu verschiedenen Zeitpunkten während des iterativen Verfeinerungsprozesses zeigt, illustriert die Auswirkungen die die Zunahme von experimentellen Distanzeinschränkungen auf die Genauigkeit der dreidimensionalen Strukturbestimmung hat. Der erste Datensatz (Bild 10A) umfaßte 322 obere Distanzschranken¹ — rund 40% der endgültigen Zahl, zumeist solche innerhalb eines Aminosäurerests oder zwischen benachbarten Resten —

¹ Irrelevante Distanzschranken (vgl. S. 109), die die Konformation nicht einschränken, sind nicht mitgezählt.

die ausschließlich aufgrund übereinstimmender chemischer Verschiebungen, ohne Kenntnis einer dreidimensionalen Struktur, zugeordnet wurden. Die mittlere Abweichung der Atompositionen von der gemittelten Struktur liegt mit $1,7 \text{ \AA}$ im Bereich einer Grobstruktur, die die globale Faltung korrekt wiedergibt, aber lokal große Abweichungen zeigt (Wüthrich, 1986; 1989b). Dieser Struktur diente im Programm ASNO dazu, Mehrdeutigkeiten in der Signalzuordnung aufzulösen, wodurch sich die Anzahl relevanter oberer Distanzschranken für die Strukturrechnung fast verdoppelte. Die verbesserte Genauigkeit der damit berechneten Struktur zeigt Bild 10B. Weitere Durchgänge der Verfeinerung mit dem Programm ASNO liefern nur noch vergleichsweise geringe Verbesserungen für die Struktur des Rückgrats, die hauptsächlich Gebiete zwischen den regulären Sekundärstrukturelementen betreffen (Bild 10C und D).

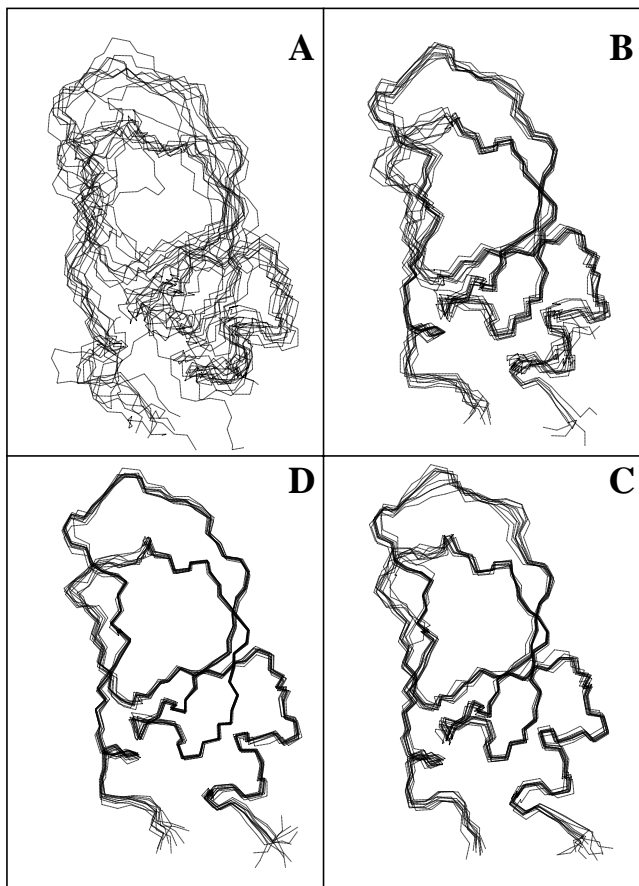


Bild 10. Überlagerung des Rückgrats der jeweils 10 (von 100) besten DIANA-Konformere zu verschiedenen Zeitpunkten während des Verfeinerungsprozesses mit Hilfe des Programms ASNO (Berndt et al., 1993).

(A) $n = 322$ obere Distanzschranken. Mittlere Abweichung der Atompositionen von der gemittelten Struktur: $A = 1.66 \text{ \AA}$.
 (B) $n = 657$, $A = 0.65 \text{ \AA}$.
 (C) $n = 747$, $A = 0.38 \text{ \AA}$.
 (D) Die endgültige Struktur mit $n = 809$ und $A = 0.32 \text{ \AA}$.

Datenformate für Strukturrechnungen

Dieses Kapitel beschreibt Datenformate, die im Verlauf einer Proteinstrukturbestimmung mittels Kernresonanzexperimenten verwendet werden, um die grundlegenden Ein- und Ausgabedaten der in dieser Abhandlung beschriebenen Programme zu speichern, insbesondere die des Programms DIANA und seiner Unterstützungsprogramme (Güntert *et al.*, 1991a; Güntert, 1992). Weiter wird das Programm COFIMA (Güntert, 1992), ein flexibles Werkzeug zur Bearbeitung und Formatumwandlung einiger wichtiger Typen von Daten, eingeführt. Die meisten Programme setzen voraus, daß die angegebenen FORTRAN-Formate (Kießling & Lowes, 1987) der verschiedenen Dateien strikt eingehalten werden und erlauben in den Dateien Kommentare, die mit »#« beginnen und sich bis zum Ende der entsprechenden Zeile erstrecken.

Bausteinbibliothek

Das Programm DIANA (Güntert *et al.*, 1991a) und verschiedene weitere Programme, so zum Beispiel CALIBA (Güntert *et al.*, 1991a), benötigen eine Bausteinbibliothek, in welcher für alle Aminosäurereste Atomtypen, Nomenklatur, Diederwinkeldefinitionen, kovalente Bindungen und die Standardgeometrie gemäß dem ECEPP/2 Kraftfeld (Momany *et al.*, 1975; Némethy *et al.*, 1983) niedergelegt sind.¹ Um mit anderen Programmen verträglich zu sein, enthält die DIANA-Bibliothek, die aus der Bibliothek des Energieminimierungsprogramms FANTOM (Schaumann, 1988; Schaumann *et al.*, 1990; Freiherr von Freyberg-Eisenberg, 1993) hervorging, zusätzliche Informationen, die von den hier beschriebenen Programmen nicht gebraucht und in diesem Kapitel nicht erklärt werden. Weil DIANA keine Endgruppen verwendet, sind auch Bibliothekseinträge für Endgruppen hier nicht relevant, einzig die mit den Schlüsselworten »ATOMTYPES« und »RESIDUE« gekennzeichneten Einträge

¹ Zusätzlich enthält die Standardbausteinbibliothek Einträge für die vier Nukleotide A, C, G und T, deren Standardgeometrie einer Modellstruktur entnommen ist, weil ECEPP diese Größen nur für Aminosäurereste definiert.

werden erläutert.

Der Atomtypeneintrag, der vor den Einträgen für die einzelnen Aminosäurereste stehen muß, beginnt mit einer Kopfzeile im FORTRAN Format (A10,I5), die das Wort »ATOMTYPES« und die Anzahl folgender Atomtypendeklarationen enthält. Jede Atomtypendeklaration besteht aus einer Zeile im FORTRAN Format (5X,A5,F10.2,2I5), die der Reihe nach den Atomtyp, den Radius für das abstoßende Potential in DIANA (Güntert *et al.*, 1991a), ein Code für Wasserstoffbrücken (1 für Wasserstoffatome, die Wasserstoffbrücken bilden können; -1 für Wasserstoffbrückenakzeptoren, z. B. Sauerstoff; 0 für Atome, die nicht an Wasserstoffbrücken teil haben), und einem Code, der angibt, ob entsprechende Atome in DIANA in die Berechnung von Abweichungen der »Schweratome« mit einbezogen werden (Code = 1) oder nicht (Code = 0).

Ein Eintrag für einen Aminosäurerest (oder ein Nukleotid) beginnt mit einer Kopfzeile im FORTRAN Format (A10, A5, 4I5) und enthält, in dieser Reihenfolge, das Wort »RESIDUE«, den Namen des Rests, die Anzahl Diederwinkel- und Atomdeklarationen, und die Nummern des ersten und letzten Atoms, das in der Polypeptidkette zu diesem (und nicht zu einem benachbarten) Rest gehört. Es folgen die Diederwinkeldeklarationen: Zeilen im FORTRAN Format (5X, A5, 20X, 5I5) bestehend aus Winkelname, den Nummern der vier Atome, die den Winkel definieren sowie die Nummer des letzten Atoms in der Polypeptidkette, das von einer Drehung um den Winkel betroffen sein wird (für Diederwinkel des Rückgrats steht statt dessen 0). Atomnummern beziehen sich auf die nachfolgenden Zeilen mit Atomdeklarationen im Format (5X, 2A5, 15X, 3F10.0, 5I5), die der Reihe nach den Atomnamen, -typ und -radius, die kartesischen Koordinaten für eine beliebige Konformation, die die Standardgeometrie erfüllt, vier Atomnummern, die kovalente Bindungen anzeigen (falls weniger als vier kovalente Bindungen vorhanden sind, steht an den entsprechenden Stellen 0) und schließlich die Atomnummer des anderen diastereotopen Substituenten, falls vorhanden (sonst 0). Die Reihenfolge der Diederwinkel- und Atomdeklarationen muß so sein, daß die Menge der Atome, die eine Drehung um einen Diederwinkel bewegt, vom Atom, das dem dritten Atom in der Diederwinkeldefinition folgt bis zum letzten bewegten Atom, dem fünften Atom in der Diederwinkeldefinition (oder bis zum C-Terminus für Rückgratswinkel), reicht.

Die Nomenklatur der Atome folgt weitgehend den entsprechenden IUPAC Regeln (IUPAC-IUB Commission on Biochemical Nomenclature, 1970). Die einzige Ausnahme betrifft das Rückgratamidproton, das mit HN anstatt H heißt. Zusätzlich zu den echten Atomen enthält die Bausteinbibliothek noch Pseudoatome, die den Atomtyp »PSEUD« besitzen, und als ausdehnungslose Bezugspunkte für Distanzeinschränkungen dienen (Wüthrich *et al.*, 1983). Um Verwirrungen in bezug auf die Nomenklatur zu vermeiden, folgt die Liste der

11 H4'	H-ALI	20.86	-4.72	12.14	10	0	0	0	0	0	2 O3'	O-EST	7.61	2.77	12.95	1	3	0	0	0	0
12 O4'	O-EST	21.00	-3.07	10.97	10	13	0	0	0	0	3 P	P-ALI	8.70	3.91	13.22	2	4	5	6	0	0
13 C1'	C-ALI	22.12	-2.27	10.63	12	14	15	30	0	0	4 OP1	O-BYL	8.22	4.82	14.29	3	0	0	0	0	0
14 H1'	H-ALI	22.56	-2.63	9.70	13	0	0	0	0	0	5 OP2	O-BYL	9.06	4.57	11.94	3	0	0	0	0	0
15 N9	N-AMI	21.63	-0.91	10.28	13	16	29	0	0	0	6 O5'	O-EST	9.94	3.05	13.76	3	7	0	0	0	0
16 C8	C-ARO	22.35	0.20	9.89	15	17	18	0	0	0	7 C5'	C-ALI	11.16	3.73	14.12	6	8	9	10	0	0
17 H8	H-ARO	23.42	0.14	9.80	16	0	0	0	0	0	8 H5'	H-ALI	11.54	4.27	13.26	7	0	0	0	0	9
18 N7	N-AMI	21.61	1.25	9.65	16	19	0	0	0	0	9 H5''	H-ALI	10.96	4.42	14.94	7	0	0	0	0	8
19 C5	C-ARO	20.31	0.81	9.91	18	20	29	0	0	0	10 C4'	C-ALI	12.19	2.72	14.60	7	11	12	32	0	0
20 C6	C-ARO	19.08	1.51	9.82	19	21	22	0	0	0	11 H4'	H-ALI	11.74	2.07	15.36	10	0	0	0	0	0
21 O6	O-BYL	18.87	2.68	9.50	20	0	0	0	0	0	12 O4'	O-EST	12.51	1.81	13.50	10	13	0	0	0	0
22 N1	N-AMI	18.00	0.68	10.17	20	23	24	0	0	0	13 C1'	C-ALI	13.79	2.12	12.97	12	14	15	29	0	0
23 H1	H-AMI	17.09	1.08	10.15	22	0	0	0	0	0	14 H1'	H-ALI	14.44	1.26	13.07	13	0	0	0	0	0
24 C2	C-ARO	18.11	-0.64	10.54	22	25	28	0	0	0	15 N1	N-AMI	13.69	2.31	11.46	13	16	27	0	0	0
25 N2	N-AMI	16.96	-1.26	10.84	24	26	27	0	0	0	16 C6	C-ARO	14.77	2.62	10.67	15	17	18	0	0	0
26 H21	H-AMI	16.03	-0.86	10.83	25	0	0	0	27	0	17 H6	H-ARO	15.74	2.74	11.11	16	0	0	0	0	0
27 H22	H-AMI	17.03	-2.23	11.12	25	0	0	0	26	0	18 C5	C-ARO	14.65	2.72	9.34	16	19	23	0	0	0
28 N3	N-AMI	19.27	-1.29	10.62	24	29	0	0	0	0	19 C7	C-ALI	15.82	3.06	8.47	18	20	21	22	0	0
29 C4	C-ARO	20.32	-0.49	10.28	15	19	28	0	0	0	20 H71	H-ALI	16.59	2.30	8.59	19	0	0	0	0	0
30 C2'	C-ALI	23.09	-2.29	11.82	13	31	32	33	0	0	21 H72	H-ALI	15.50	3.09	7.43	19	0	0	0	0	0
31 H2'	H-ALI	23.54	-1.62	12.55	30	0	0	0	32	0	22 H73	H-ALI	16.22	4.03	8.76	19	0	0	0	0	0
32 H2''	H-ALI	23.81	-2.48	11.02	30	0	0	0	31	0	23 C4	C-ARO	13.38	2.51	8.69	18	24	25	0	0	0
33 C3'	C-ALI	22.72	-3.61	12.50	10	30	34	35	0	0	24 O4	O-BYL	13.17	2.59	7.48	23	0	0	0	0	0
34 H3'	H-ALI	22.94	-3.72	13.56	33	0	0	0	0	0	25 N3	N-AMI	12.35	2.21	9.56	23	26	27	0	0	0
35 O3'	O-EST	23.51	-4.63	11.90	33	36	0	0	0	0	26 H3	H-AMI	11.37	2.04	9.12	25	0	0	0	0	0
36 P	P-ALI	25.10	-4.60	12.10	35	0	0	0	0	0	27 C2	C-ARO	12.44	2.09	10.94	15	25	28	0	0	0
RESIDUE	THY	6	35	3	34						28 O2	O-BYL	11.47	1.82	11.62	27	0	0	0	0	0
1 TORZ		2	3	6	0						29 C2'	C-ALI	14.31	3.36	13.70	13	30	31	32	0	0
2 TORA		3	6	7	0						30 H2'	H-ALI	14.59	4.40	13.50	29	0	0	0	0	31
3 TORE		6	7	10	0						31 H2''	H-ALI	15.21	2.75	13.77	29	0	0	0	0	30
4 TORG		7	10	32	0						32 C3'	C-ALI	13.54	3.29	15.02	10	29	33	34	0	0
5 CHI		12	13	15	16	28					33 H3'	H-ALI	13.40	4.30	15.41	32	0	0	0	0	0
6 TORE		10	32	34	35	0					34 O3'	O-EST	14.31	2.51	15.92	32	35	0	0	0	0
1 C3'	C-ALI	6.34	3.14	12.45	2	0	0	0	0	0	35 P	P-ALI	15.73	3.06	16.41	34	0	0	0	0	0

Die ersten drei Buchstaben der Aminosäurenamen entsprechen dem Standard Dreibuchstabencode; das vierte Zeichen gibt Ladungen an (z. B. ARG+) oder unterscheidet zwischen Cystein- (CYS) und Cystinresten (CYSS). Nichtstandard-Bausteine lassen sich jederzeit der Bibliothek anfügen.

Aminosäuresequenz

Die Sequenzdatei definiert die Primärstruktur eines Proteins, identifiziert *cis*-Peptidbindungen und spezielle kovalente Bindungen, d. h. Bindungen, die nicht mit der im Programm DIANA verwendeten Baumstruktur verträglich sind (vgl. S. 114), und deklariert fixierte und drehbare Diederwinkel.

Aminosäurenamen bestehen aus höchstens vier Buchstaben und müssen einem Eintrag in der Bausteinbibliothek entsprechen. Wenn einem Aminosäurenamen ein kleines »c« vorangeht, wird der ω -Diederwinkel der vorangehenden Peptidbindung bei 0° (*cis*) anstatt, wie üblich, bei 180° (*trans*) fixiert. Ein Aminosäurename kann von der Nummer des entsprechenden Rests gefolgt werden; anderenfalls gibt das Programm DIANA dem ersten (N-terminalen) Rest die Nummer 1 und weiteren Resten die um eins erhöhte Nummer des vorangehenden Rests. Abgesehen davon, daß Namen und Nummern durch mindestens ein Leerzeichen oder einen Zeilenanfang voneinander getrennt sein müssen, ist das Format frei. Ein einfaches Beispiel einer Sequenzdatei:

```
# Zweite Helix der Antennapedia Homeodomäne  
ARG+ 29 ARG+ ARG+ ARG+ ARG+ ILE GLU- ILE ALA HIS ALA LEU
```

Diese Datei enthält die Sequenz eines 11 Aminosäuren langen Polypeptids mit den Aminosäurenummern 29, 30, 31, ..., 39, durchgehend *trans* Peptidbindungen und ohne spezielle kovalente Bindungen. Die erste Zeile ist eine Kommentarzeile. Die Deklaration einer speziellen kovalenten Bindung besteht aus dem kleingeschriebenen Schlüsselwort »link«, gefolgt vom ersten Atomnamen, der ersten Aminosäurenummer, dem zweiten Atomnamen und der zweiten Aminosäurenummer. Spezielle kovalente Bindungen in Proteinen sind zum Beispiel Disulfidbrücken oder zyklische Peptidbindungen. Eine Disulfidbrücke zwischen den Resten 3 und 55 wird zum Beispiel angegeben durch:

link SG 3 SG 55

Normalerweise nimmt das Programm DIANA an, daß alle Diederwinkel außer ω (OMEGA) drehbar, d. h. Freiheitsgrade während der Minimierung sind. Um einen sonst fixierten Diederwinkel drehbar zu machen, dient der Eintrag

Winkelname = **free**

Umgekehrt fixieren die beiden Anweisungen

Winkelname = **fixed** oder *Winkelname* = *Wert*

einen Diederwinkel beim Wert in der Ausgangskonformation oder an einer gegebenen Position *Wert* (in Grad). Wenn eine solche Winkeldeklaration in der Sequenzdatei nur für einen Teil der Sequenz gelten soll, wird sie in geschweifte Klammern, die auch geschachtelt auftreten können, eingeschlossen:

{*Winkeldeklaration Teil der Aminosäuresequenz*}

Konformationen, beschrieben durch kartesische Koordinaten

Dateien mit kartesischen Atomkoordinaten beginnen immer mit drei Kommentarzeilen, gefolgt von je einer Zeile pro Atom im FORTRAN-Format (6X, A5, I6, 1X, A5, 3F11.4, 4I6), die der Reihe nach den Atomnamen, die Nummer und den Namen des Aminosäurerests, die *x*, *y* und *z*-Koordinate in Å und — fakultativ — vier Nummern von Atomen, mit denen kovalente Bindungen bestehen (oder 0, falls weniger als vier kovalente Bindungen vorhanden sind), angibt. Eine Beispieldatei mit kartesischen Koordinaten, aber ohne explizit angegebene kovalente Bindungen, wie sie vom Programm DIANA geschrieben wird, sieht

so aus:

```
Structure 001 from DIANA, f=1.843E-01
DIANA UNIX version 2.1
Number of residues:      1 Number of atoms:      29
  1 N          29 ARG+      1.3249      0.0000      0.0000
  2 HN         29 ARG+      1.8841      0.0000      0.8290
  3 CA         29 ARG+      2.0733      0.0000     -1.2455
  4 HA         29 ARG+      1.3311     -0.2065     -2.0166
  5 CB         29 ARG+      2.7313      1.3590     -1.4927
  6 HB2        29 ARG+      2.1094      2.1498     -1.0732
  7 HB3        29 ARG+      3.6911      1.4014     -0.9778
  8 QB         29 ARG+      2.9002      1.7756     -1.0255
  9 CG         29 ARG+      2.9376      1.6042     -2.9887
 10 HG2        29 ARG+      3.2843      2.6251     -3.1491
 11 HG3        29 ARG+      3.7157      0.9405     -3.3658
 12 QG         29 ARG+      3.5000      1.7828     -3.2575
 13 CD         29 ARG+      1.6406      1.3720     -3.7664
 14 HD2        29 ARG+      0.8728      0.9790     -3.1001
 15 HD3        29 ARG+      1.2680      2.3181     -4.1590
 16 QD         29 ARG+      1.0704      1.6486     -3.6296
 17 NE         29 ARG+      1.8792      0.4234     -4.8770
 18 HE         29 ARG+      2.7489     -0.0699     -4.8934
 19 CZ         29 ARG+      1.0028      0.1841     -5.8619
 20 NH1        29 ARG+      1.3044     -0.6935     -6.8286
 21 HH11       29 ARG+      0.6503     -0.8723     -7.5636
 22 HH12       29 ARG+      2.1836     -1.1699     -6.8150
 23 QH1        29 ARG+      1.4170     -1.0211     -7.1893
 24 NH2        29 ARG+     -0.1753      0.8223     -5.8802
 25 HH21       29 ARG+     -0.8294      0.6436     -6.6153
 26 HH22       29 ARG+     -0.4004      1.4772     -5.1588
 27 QH2        29 ARG+     -0.6150      1.0604     -5.8871
 28 C          29 ARG+      3.1528     -1.0838     -1.2161
 29 O          29 ARG+      2.8524     -2.2586     -1.0102
```

Konformationen, beschrieben durch Diederwinkel

Diederwinkeldateien des Programms DIANA bestehen aus Zeilen im FORTRAN-Format (I3, 1X, A5, 4(1X, A5, F9.3)) und enthalten der Reihe nach die Nummer und den Namen des Aminosäurerests und bis zu vier Namen und Werte (in Grad) von Diederwinkeln. Die Nummer und der Name des Aminosäurerests müssen nicht unbedingt wiederholt werden, wenn sie in aufeinanderfolgenden Zeilen gleich sind:

```
# Structure 001 from DIANA, f=4.265E+01
29 ARG+  PHI   -130.949 CHI1  -153.369 CHI2    51.242 CHI3  -127.183
        CHI4  -168.829 PSI   -123.439
30 ARG+  PHI    93.710 CHI1  -36.259 CHI2   -82.801 CHI3    86.123
        CHI4   96.563 PSI   -113.212
31 ARG+  PHI    72.589 CHI1  114.583 CHI2    36.050 CHI3  -170.814
        CHI4    51.485 PSI   132.168
32 ARG+  PHI    44.439 CHI1  -76.452 CHI2   129.358 CHI3  -154.539
        CHI4  -165.137 PSI  -151.520
33 ILE   PHI   -123.429 CHI1  -117.875 CHI22 -158.650 CHI21   49.504
        CHI31   59.630 PSI  -131.114
```

34	GLU-	PHI	-11.771	CHI1	-177.685	CHI2	151.794	CHI3	136.739
		PSI	173.092						
35	ILE	PHI	-156.599	CHI1	-116.335	CHI22	-105.820	CHI21	141.274
		CHI31	-28.062	PSI	132.270				
36	ALA	PHI	-53.903	CHI1	-25.906	PSI	-125.155		
37	HIS	PHI	-140.445	CHI1	129.973	CHI2	-130.746	PSI	-162.362
38	ALA	PHI	-174.388	CHI1	-15.129	PSI	33.330		
39	LEU	PHI	-150.177	CHI1	13.039	CHI2	-89.991	CHI31	-112.141
		CHI32	3.817	PSI	88.705				

Skalare Spin-Spin Kopplungskonstanten

Dateien mit Werten für skalare Spin-Spin Kopplungskonstanten werden vom Programm HABAS (Güntert *et al.*, 1989) gelesen und anhand der Karpluskurven [37–39] in Diederwinkeleinschränkungen übersetzt. Kopplungskonstantendateien bestehen aus Zeilen im FORTRAN-Format (I4, 1X, 3A5, F4.1, F5.1), die nacheinander die Nummer und den Namen des Aminosäurerests, die Atomnamen der beiden gekoppelten Spins, den Wert der Spin-Spin Kopplungskonstante und ihr Fehlergrenzen (beide in Hertz) angeben, wie im folgenden Beispiel:

29	ARG+	HN	HA	4.8	2.0
30	ARG+	HN	HA	6.5	2.0
31	ARG+	HN	HA	5.2	2.0
32	ILE	HN	HA	5.2	2.0
32	ILE	HA	HB	10.5	2.0
33	ASP-	HN	HA	6.3	2.0
33	ASP-	HA	HB2	10.0	2.0
33	ASP-	HA	HB3	6.0	2.0
34	ILE	HN	HA	5.9	2.0
34	ILE	HA	HB	9.5	2.0
35	ALA	HN	HA	4.0	2.0
36	ASN	HN	HA	5.4	2.0

Stereospezifisch zugeordnete diastereotope Gruppen

Diese Datei teilt dem Programm DIANA mit, welche diastereotopen Protonen oder Methylgruppen stereospezifisch zugeordnet sind. Welcher der beiden Substituenten in der *pro-R* Konformation vorliegt, hängt von den kartesischen Koordinaten in der Bausteinbibliothek ab. Die Datei besteht aus Zeilen im FORTRAN-Format (I4, 1X, A5, 1X, 13A5), die der Reihe nach die Nummer und den Namen des Aminosäurerests und die Namen von bis zu 13 diastereotopen Substituenten, die korrekt stereospezifisch zugeordnet sind, angeben:

```
# Stereospezifische Zuordnungen
# durch HABAS
```

```

29 ARG+ HB2
32 ARG+ HB2
34 GLU- HB2
39 LEU HB2
# durch GLOMSA
32 ARG+ HG2
34 GLU- HG2
31 ARG+ HD2
39 LEU QD1
# für Guanidiniumgruppen von Arg
31 ARG+ QH1
32 ARG+ QH1

```

Es genügt, einen diastereotopen Substituenten anzugeben, um für das gesamte Paar die stereospezifische Zuordnung festzulegen. Für Isopropylgruppen müssen die Methylgruppen durch Pseudoatome (z. B. QD1 oder QD2 bei Leu) angegeben werden, und nicht durch die einzelnen Methylprotonen.

Obere und untere Distanzschranken

Dateien für obere und untere Distanzschranken besitzen den gleichen Aufbau: Zeilen mit dem FORTRAN-Format (I3, 1X, 2A5, I4, 1X, 2A5, F7.2, 2X, E8.2), die der Reihe nach die Aminosäurenummer, den Aminosäure- und Atomnamen des ersten Atoms, die Aminosäurenummer, den Aminosäure- und Atomnamen des zweiten Atoms, die Distanzschranke in Å und — fakultativ — einen relativen Gewichtungsfaktor für die Minimierung in DIANA aufführen. Gewichtungsfaktoren sollten positiv sein, ihr Standardwert ist 1. Eine Beispieldatei:

```

29 ARG+ HN      29 ARG+ HB2      2.90
29 ARG+ HN      29 ARG+ HB3      3.00
29 ARG+ HN      29 ARG+ QG       4.33
29 ARG+ HN      30 ARG+ HN       3.30
29 ARG+ HA      29 ARG+ QG       3.87
29 ARG+ HA      32 ARG+ HN       4.00
29 ARG+ QD      33 ILE  QD1      6.80
30 ARG+ HN      30 ARG+ QB       2.99
30 ARG+ HN      31 ARG+ HN       3.40
30 ARG+ HA      30 ARG+ QB       2.72
30 ARG+ HA      30 ARG+ QD       5.80
30 ARG+ HA      33 ILE  HN       3.80  5.00E+00

```

In diesem Beispiel besitzt die letzte Einschränkung relativ zu den anderen das fünffache Gewicht.

Fehlen auf einer Zeile die Nummer und der Name des ersten Aminosäurerests, die auch allein auf einer Zeile stehen dürfen, so werden die entsprechenden Daten der vorhergehenden Zeile verwendet. Äquivalent zur obigen Beispieldatei ist daher folgende vereinfachte Form:

29	ARG+	HN	29	ARG+	HB2	2.90	
		HN	29	ARG+	HB3	3.00	
		HN	29	ARG+	QG	4.33	
		HN	30	ARG+	HN	3.30	
		HA	29	ARG+	QG	3.87	
		HA	32	ARG+	HN	4.00	
		QD	33	ILE	QD1	6.80	
30	ARG+	HN	30	ARG+	QB	2.99	
		HN	31	ARG+	HN	3.40	
		HA	30	ARG+	QB	2.72	
		HA	30	ARG+	QD	5.80	
		HA	33	ILE	HN	3.80	5.00E+00

Diederwinkleinschränkungen

Eine Datei mit Diederwinkleinschränkungen enthält direkte Einschränkungen für einzelne Diederwinkel in der Form eines erlaubten Intervalls $[\phi^{\min}, \phi^{\max}]$ mit $\phi^{\min} < \phi^{\max} < \phi^{\min} + 2\pi$, das nicht zu einem Punkt entarten darf. In der aktuellen Version des Programms DIANA darf höchstens ein Intervall pro Diederwinkel angegeben werden. Die Datei besteht aus Zeilen mit dem FORTRAN-Format (I4, 1X, A5, 1X, A5, 2F8.1, E9.2), die der Reihe nach die Nummer und den Namen des Aminosäurerests, den Diederwinkelnamen, die untere ϕ^{\min} und obere Schranke ϕ^{\max} in Grad und — fakultativ — einen relativen Gewichtungsfaktor für die Minimierung in DIANA aufführen. Gewichtungsfaktoren sollten positiv sein, ihr Standardwert ist 1. Eine Beispieldatei:

32	ARG+	PHI	-55.0	-35.0	
32	ARG+	PSI	-75.0	-15.0	1.00E-01
32	ARG+	CHI1	-155.0	-125.0	
33	ILE	PHI	-65.0	-35.0	
33	ILE	PSI	-85.0	-15.0	
33	ILE	CHI1	-105.0	-35.0	
34	GLU-	PHI	-65.0	-45.0	
34	GLU-	PSI	-85.0	-25.0	
34	GLU-	CHI1	-5.0	125.0	

In diesem Beispiel besitzt die zweite Einschränkung 1/10 des Gewichts der anderen.

Bearbeitung von Datensätzen mit dem Programm COFIMA

Das Programm COFIMA (*»coordinate file manipulation«*) ist ein vielseitiges Programm zur Bearbeitung von Dateien mit kartesischen Koordinaten, Distanz- und Diederwinkeleinschränkungen (Güntert, 1992). Es dient insbesondere zum Umformatieren, Umbenennen, Auflisten, Einfügen und Sortieren von Einträgen in solchen Dateien. Das Programm besteht aus drei Teilen, *cofima* zur Bearbeitung von Koordinaten, *difima* für Distanzschranken und *anco* für Diederwinkel. Viele Anweisungen können auf alle drei Typen von Daten angewandt werden, andere sind spezifisch für einzelne Datentypen.

Im folgenden sind wörtliche Eingaben **fett** und nicht-wörtliche Eingaben *kursiv* gedruckt, wobei fakultative Eingaben in eckigen Klammern [...] stehen. *a*, *a*₁ usw. bezeichnen Atom oder Diederwinkelnamen, die mit einem Buchstaben beginnen müssen und in vielen Fällen die Sonderzeichen »*« und »%« enthalten dürfen, die für eine beliebige Anzahl von Zeichen bzw. genau ein Zeichen stehen. Atom-, Diederwinkel- und Aminosäurenamen sowie Aminosäurenummern kann oft als Verneinung ein Ausrufezeichen »!« vorangestellt werden. Aminosäurenummern sind ganze Zahlen. Die speziellen Aminosäurenamen **FIRST** und **LAST** bezeichnen den ersten bzw. letzten Rest. Viele Anweisungen können anstatt auf alle Reste nur auf einen Bereich von Aminosäureresten, im weiteren mit *Bereich* bezeichnet, angewandt werden, der durch eines oder mehrere der folgenden Elemente angegeben wird:

Aminosäurenummer
Aminosäurenummer.
..Aminosäurenummer
Aminosäurenummer..Aminosäurenummer
@Aminosäurename

So bezeichnet zum Beispiel

12	den Rest 12;
12 20..25	die Reste 12, 20, 21, 22, 23, 24 und 25;
@THR 20..25	die THR-Reste im Segment 20–25;
!@CY*	die Reste, deren Namen nicht mit »CY« beginnt.

Anweisungen für COFIMA lassen sich in Makrodateien, Dateien mit der Namensergänzung *».cfm«*, zusammenfassen, um Routineanwendungen zu erleichtern. Verschiedene Standardmakros gehören zum Programm, so zum Beispiel zur Umwandlung der Nomenklatur zwischen den Programmen DIANA, AMBER (Singh *et al.*, 1986), FANTOM (Schaumann *et al.*, 1990) und X-PLOR (Brünger, 1992) oder zum Einbau der Wasserstoffatome in eine Proteinstruktur.

Alphabetische Liste der COFIMA Anweisungen

angles [~]a₁ a₂ [~]a₃ [[~]a₄] [[~]a₅] *Bereich*

zeigt Bindungswinkel, Diederwinkel oder relative Diederwinkel an und ist nur für Koordinatendateien anwendbar. Bei drei angegebenen Atomnamen werden die Bindungswinkel $a_1-a_2-a_3$, bei vier Atomnamen die Diederwinkel $a_1-a_2-a_3-a_4$ und bei fünf Atomnamen die Differenzen zwischen den Diederwinkeln $a_1-a_2-a_3-a_4$ und $a_1-a_2-a_3-a_5$ berechnet. Normalerweise müssen sich alle vier Atome im gleichen Rest befinden, es sei denn, einem Atomnamen geht eine Tilde »~« voraus. In diesem Fall sucht das Programm die betreffenden Atome über die Liste der chemischen Bindungen, die natürlich vorhanden sein muß; entweder indem sie eingelesen oder mit Hilfe der Anweisung **connect** erzeugt wurde. Zum Beispiel berechnet **angles ~CA C ~N ~CA** den ω Diederwinkel der Peptidbindungen.

ancoma

wechselt zu *ancoma* (*»angle constraint manipulation«*), dem Programmteil zur Bearbeitung von Diederwinkelschranken.

attach a [~]a₁ a₂ a₃ [[~]a₄] *Bindungslänge Bindungswinkel Diederwinkel [Bereich]*

baut Atome in einen Struktur ein und ist nur für Koordinatendateien anwendbar. Das Atom *a* wird so eingebaut, daß die Distanz $a-a_3$, der Bindungswinkel a_2-a_3-a und der Diederwinkel $a_1-a_2-a_3-a$ bzw. die Differenz der Diederwinkel $a_1-a_2-a_3-a$ und $a_1-a_2-a_3-a_4$ (falls a_4 angegeben ist) die in der Kommandozeile angegebenen Werte annehmen. Das folgende Beispiel baut das β -Proton von Threonin in eine Struktur, für die nur die Schweratompositionen vorhanden sind, ein: **attach HB N CA CB OG1 1.09 110.9 123.0 @THR**. Soll ein Atom nicht im gleichen Rest, sondern über die Liste kovalenter Bindungen gesucht werden, kann seinem Namen eine Tilde »~« vorangestellt werden.

bind a₁ *Aminosäurenummer*₁ a₂ *Aminosäurenummer*₂

führt eine kovalente Bindung zwischen dem Atom a_1 im Rest *Aminosäurenummer*₁ und dem Atom a_2 im Rest *Aminosäurenummer*₂ ein und ist nur für Koordinatendateien anwendbar.

break a₁ *Aminosäurenummer*₁ a₂ *Aminosäurenummer*₂

entfernt eine kovalente Bindung zwischen dem Atom a_1 im Rest *Aminosäurenummer*₁ und dem Atom a_2 im Rest *Aminosäurenummer*₂ und ist nur für Koordinatendateien anwendbar.

change @alter *Aminosäurename* @neuer *Aminosäurename* [*Bereich*]

ändert Aminosäurenamen.

cofima

wechselt zu *cofima* (*»coordinate file manipulation«*), dem Programmteil zur Bearbeitung von kartesischen Koordinaten.

connect [a₁=*Radius*₁ . . . a_n=*Radius*_n]

erzeugt kovalente Bindungen innerhalb eines Rests oder zwischen sequentiell benachbarten Resten aufgrund eines Bindungslängenkriteriums und ist nur für Koordinatendateien anwendbar. Falsche Ergebnisse, d. h. zu viele kovalente Bindungen, können auftreten, wenn großer sterische Überlapp vorkommt. Üblicherweise kann die Anweisung ohne Parameter gebraucht werden und ist dann äquivalent zu **connect H*=0.4 C*=0.85 N*=0.8 O*=0.7 S*=1.3 P*=1.2 Q*=-999 LP*=-999 *=0.85** und kann in dieser Form gebraucht werden. Eine kovalente Bindung erhalten Atompaare, für die die Summe der beiden Radien den Abstand zwischen den beiden Atomen überschreitet. Falls auf ein Atom mehrere der Radiusdefinitionen anwendbar sind, wird die am weitesten links stehende verwendet.

constraints [a₁ . . . a_n] [*Bereich*]

zeigt Distanz- oder Diederwinkelschranken an, die die angegebenen Atome oder

Diederwinkel betreffen.

coordinates [$a_1 \dots a_n$] [*Bereich*]

listet Namen, Aminosäurenummer und -name, kartesische Koordinaten und, falls vorhanden, kovalente Bindungen, der angegebenen Atome und ist nur für Koordinatendateien anwendbar.

copy a_1 Aminosäurenummer₁ [a_2] Aminosäurenummer₂ [a_3]

kopiert die kartesischen Koordinaten des Atoms a_1 im Rest Aminosäurenummer₁ zu einem neuen, zusätzlichen Atom a_2 im Rest Aminosäurenummer₂ und ist nur für Koordinatendateien anwendbar. Wird a_2 nicht angegeben, erhält das neue Atom den Namen a_1 . Falls ein dritter Atomname a_3 vorhanden ist, wird das zusätzliche Atom hinter dem Atom a_3 eingefügt, anderenfalls als letztes Atom des Rests.

delete [$a_1 \dots a_n$] [*Bereich*]

löscht Einträge, die die angegebenen Atome oder Diederwinkel betreffen.

difima

wechselt zu *difima* (*»distance constraint file manipulation«*), dem Programmteil zur Bearbeitung von Distanzeinschränkungen.

directory [*Makroname*]

gibt eine Liste der Standardmakros und der Makros im aktuellen Verzeichnis. Das Programm listet Makros, deren Name zum angegebenen *Makronamen* paßt, oder, falls dieser fehlt, alle Makros. *Makroname* kann die Sonderzeichen »*« und »%« enthalten und muß ohne die Namensergänzung ».cfm« angegeben werden. Von jedem Makro werden Kommentarzeilen, d. h. mit »#« beginnende Zeilen, die vor der ersten Anweisung stehen, angezeigt.

disconnect [$a_1 \dots a_n$] [*Bereich*]

löscht kovalente Bindungen, die die angegebenen Atome betreffen, und ist nur für Koordinatendateien anwendbar. Ohne Parameter werden alle kovalenten Bindungen gestrichen.

distances a_1 [*Bereich 1*] a_2 [*Bereich 2*]

berechnet, wenn angewandt auf Koordinatendateien, interatomare Distanzen zwischen den Atomen a_1 der Reste im Bereich *Bereich 1* und den Atomen a_2 der Reste im Bereich *Bereich 2*, oder listet, wenn angewandt auf Distanzschrankendateien, Schranken für den entsprechenden Distanzen. Für Diederwinkeleinschränkungen kann die Anweisung nicht verwendet werden.

end

beendet das Programm COFIMA.

firstpseudo a $a_1 \dots a_n$ a_{n+1} =Korrektur₁ \dots *=Standardkorrektur [*Bereich*]

funktioniert gleich wie die Anweisung **pseudo** (vgl. S. 82), wirkt aber nur auf das *erste* Atom einer Distanzeinschränkung.

help

gibt eine kurze Übersicht der Anweisungen.

insert a $a_1 \dots a_n$ [*Bereich*]

fügt Pseudoatome a im Zentrum der Atome a_1, \dots, a_n ein und ist nur für Koordinatendateien anwendbar. So fügt zum Beispiel **insert QB HB%** ein Pseudoatom QB im Zentrum der β -Protonen ein.

keep [$a_1 \dots a_n$] [*Bereich*]

löscht Einträge, die die angegebenen Atome oder Diederwinkel *nicht* betreffen.

link [*Bindungslänge*] a_1 [*Bereich 1*] a_2 [*Bereich 2*]

erzeugt kovalente Bindungen zwischen den Atomen a_1 der Reste im Bereich *Bereich 1* und

den Atomen a_2 der Reste im Bereich *Bereich 2*, falls ihr Abstand kleiner als die angegebene Bindungslänge (Standardwert 2,5 Å) und ist nur für Koordinatendateien anwendbar. Zum Beispiel verbindet **link SG SG** Atome mit Namen SG, die weniger als 2,5 Å voneinander entfernt sind und erzeugt damit die kovalenten Bindungen für Disulfidbrücken.

list [*Bereich*]

gibt einen Überblick über die Einträge, die Reste im angegebenen Bereich betreffen. Die Anzahl Atome, Distanz- oder Diederwinkeleinschränkungen sowie eine Liste der auftretenden Atom-, Winkel- und Aminosäurenamen wird angezeigt.

pseudo $a a_1 \dots a_n a_{n+1} = \text{Korrektur}_1 \dots * = \text{Standardkorrektur}$ [*Bereich*]

führt Pseudoatomkorrekturen für Distanzeinschränkungen ein, die diastereotope Protonen oder Methylgruppen betreffen und ist nur für obere Distanzschranken anwendbar. Distanzschranken, die die Atome a_1, \dots, a_n betreffen, werden neu auf das Atom a bezogen und um eine Korrektur erhöht. Normalerweise wird die Standardkorrektur (gegeben in Å) verwendet; zu Distanzschranken innerhalb eines Rests, die eines der Atome a_{n+1}, \dots betreffen, wird die entsprechende spezielle Korrektur addiert (Wüthrich *et al.*, 1983).

read *Dateiname*

liest eine Eingabedatei mit kartesischen Koordinaten (im Programmteil *cofima*), Distanz- (in *difima*) oder Diederwinkeleinschränkungen (in *ancoma*). Die folgenden Datenformate der Eingabedatei wird vom Programm selbständig erkannt: für Koordinatendateien die Formate des Programms DIANA mit oder ohne explizite kovalente Bindungen (vgl. S. 74), der Proteindatenbank (mit gewissen Einschränkungen; Bernstein *et al.*, 1977) und des Programms AMBER (Singh *et al.*, 1986); für Distanzschranken das Format des Programms DIANA in beiden Varianten (vgl. S. 77); für Diederwinkeleinschränkungen die Formate der Programme DIANA (vgl. S. 78) und DISMAN (Braun & Gö, 1985).

rename $a_1 a_2$ [*Bereich*]

ändert Atom- oder Diederwinkelnamen.

save

schreibt eine Ausgabedatei im gleichen Format und mit dem gleichen Namen wie die Eingabedatei, die mit **read** gelesen wurde.

secondpseudo $a a_1 \dots a_n a_{n+1} = \text{Korrektur}_1 \dots * = \text{Standardkorrektur}$ [*Bereich*]

funktioniert gleich wie die Anweisung **pseudo** (vgl. S. 82), wirkt aber nur auf das zweite Atom einer Distanzeinschränkung.

sort [$a_1 \dots a_m * a_{m+1} \dots a_n$] [*Bereich*]

sortiert Einträge nach Atom- bzw. Diederwinkelnamen in die angegebene Reihenfolge, wobei »*« für die nicht explizit aufgeführten Namen steht. Ohne Parameter wird in eine Standardreihenfolge sortiert.

type [*Makroname*]

zeigt den Inhalt von Makros an.

writamber *Dateiname*

schreibt eine Koordinatendatei im Format des Programms AMBER (Singh *et al.*, 1986).

writedg *Dateiname*

schreibt eine Koordinatendatei im Format des Programms DIANA (vgl. S. 77).

writedco *Dateiname*

schreibt eine Datei mit Distanzeinschränkungen im vereinfachten Format des Programms DIANA (vgl. S. 78).

writediana *Dateiname*

schreibt eine Datei mit Diederwinkeleinschränkungen im Format des Programms DIANA (vgl. S. 78).

writedisman *Dateiname*

schreibt eine Datei mit Diederwinkeleinschränkungen im Format des Programms DISMAN (Braun & Gö, 1985).

writelongdco *Dateiname*

schreibt eine Datei mit Distanzeinschränkungen im ausführlichen Format des Programms DIANA (vgl. S. 77).

writepdb *Dateiname*

schreibt eine Koordinatendatei im Format der Proteindatenbank (Bernstein *et al.*, 1977).

@Makroname

führt den angegebenen Makro aus. Makros sind Dateien, die COFIMA Anweisungen enthalten; ihr Dateiname muß mit der Ergänzung ».cfm« enden. Diese Ergänzung braucht nicht im *Makronamen* aufgeführt werden. Standardmakros und Makros im aktuellen Verzeichnis können ohne Angabe des Pfades aufgerufen werden.

!Wort

wiederholt die letzte Anweisung, die mit dem angegebenen *Wort* begann.

Stereospezifische Zuordnungen

Das Verfahren zur Lösungsstrukturbestimmung von Proteinen aufgrund von Kernresonanzexperimenten (Wüthrich, 1986) ist vor allem deshalb robust, weil es auf *halbquantitativen* experimentellen Einschränkungen der Konformation beruht, deren Zahl die der Freiheitsgrade, d. h. der drehbaren Bindungen¹, im allgemeinen übersteigt (Wüthrich *et al.*, 1982; Braun & Gö, 1985; Havel & Wüthrich, 1985). Für *quantitative* Distanzmessungen gibt es neben technischen auch grundsätzliche Schwierigkeiten, weil beobachtete Kern-Overhausereffekte nicht nur von der interatomaren Distanz sondern auch von den effektiven Rotationskorrelationszeiten abhängen (Solomon, 1955; Noggle & Schirmer, 1971), die innerhalb eines Proteinmoleküls variieren können (Olejniczak *et al.*, 1984; Wüthrich, 1986). Stereospezifische Zuordnungen, d. h. individuelle Resonanzzuordnungen für diastereotope Protonen, sind deshalb hilfreich, weil mit ihrer Hilfe genauere Strukturen erhalten werden können, ohne daß dafür genauere Distanzmessungen notwendig wären. Das allgemein gebräuchliche Verfahren zur sequentiellen Resonanzzuordnung in Proteinen (Wüthrich *et al.*, 1982; Wüthrich, 1986) beruht allein auf der Konstitution des Proteins, berücksichtigt also nicht die Konformation, und liefert deshalb keine stereospezifischen Zuordnungen für diastereotope Protonen oder Methylgruppen, so daß besondere Methoden für diesen Zweck benötigt werden. Dieses Kapitel untersucht den Einfluß stereospezifischer Zuordnungen auf die Präzision der Lösungsstrukturen und führt zwei rechnerische Methoden zur Bestimmung stereospezifischer Zuordnungen ein.

Einfluß stereospezifischer Zuordnungen auf Lösungsstrukturen

In drei verschiedenen Testrechnungen wird untersucht, wieweit stereospezifische Zuordnungen die Genauigkeit einer Lösungsstruktur beeinflussen:

Die erste Untersuchung (Güntert *et al.*, 1989) geht von einem Datensatz aus, der einer regularisierten Kristallstruktur² des basischen pankreatischen

¹ Bindungslängen und -winkel werden als fest angenommen.

Trypsininhibitors (BPTI) entnommen ist und für den alle möglicherweise aufgrund von Kernresonanzexperimenten erhältlichen stereospezifischen Zuordnungen und individuelle Zuordnungen für aromatische Ringprotonen angenommen werden, und vergleicht die erhaltenen Strukturen mit solchen, die auf einem Datensatz beruhen, der sich nur dadurch unterscheidet, daß er keine individuellen Zuordnungen für diastereotope Gruppen und aromatische Ringe vorsieht.

Die zweite Arbeit (Güntert *et al.*, 1991a) untersucht anhand eines experimentellen Datensatzes für BPTI den Einfluß stereospezifischer Zuordnungen und zeigt, daß im Falle nicht vorhandener stereospezifischer Zuordnungen der Verlust an Genauigkeit in der Strukturrechnung begrenzt werden kann, indem anstelle der üblichen Pseudoatomkorrekturen (Wüthrich *et al.*, 1983) eine verbesserte Behandlung von Einschränkungen mit nicht stereospezifisch zugeordneten diastereotopen Gruppen angewandt wird (Güntert *et al.*, 1991a).

Die dritte, bisher unveröffentlichte Untersuchung beruht ebenfalls auf einem experimentellen Datensatz für BPTI (Berndt, *et al.*, 1992), verwendet aber im Gegensatz zu den beiden vorhergegangenen Arbeiten die heutige Standardstrategie (Güntert & Wüthrich, 1991; vgl. Kapitel *Konvergenzbeschleunigung*) zur Proteinstrukturberechnung mit dem Programm DIANA und kommt daher für Datensätze hoher Qualität den Verhältnissen in der Praxis am nächsten.

Einfluß stereospezifischer Zuordnungen bei simulierten BPTI Datensätzen

Ausgehend von einer regularisierten Kristallstruktur von BPTI (Marquardt *et al.*, 1983) wurden zwei Kernresonanzdatensätze simuliert, die sich nur in stereospezifischen und individuellen Ringprotonenzuordnungen unterscheiden (Güntert *et al.*, 1989): im Datensatz ohne solche Zuordnungen beziehen sich alle Distanzeinschränkungen, die diastereotope oder Ringprotonen betreffen, auf Pseudoatome; im anderen Datensatz sind stereospezifische Zuordnungen für alle β -Methylengruppen, die α -Methylengruppen der Glycine, alle Methylengruppen der Proline und die Isopropylgruppen der Valine und Leucine sowie individuelle Zuordnungen der aromatischen Ringprotonen vorhanden. Distanzschranken wurden erzeugt für alle Proton-Proton Distanzen in der regularisierten Kristallstruktur, die kürzer als 4 Å sind und zwischen Protonen verschiedener Aminosäurereste oder zwischen Rückgratprotonen und äußeren (nicht β -) Seitenkettenprotonen eines Rests auftreten. Für Di-

² Weil die Strukturberechnung im Diederwinkelraum, d. h. mit festen Bindungslängen und -winkeln stattfindet, sind simulierte Datensätze im allgemeinen nur dann widerspruchsfrei, wenn sie einer Struktur entstammen, die die Standardgeometrie der Aminosäurereste einhält. Für Einzelheiten der Regularisierung vgl. Güntert *et al.* (1989).

stanzen innerhalb eines Rests oder zwischen sequentiell benachbarten Resten wurde die kleinstmögliche der oberen Schranken 2,5, 3,0, 3,5 und 4,0 Å gewählt; alle Distanzschranken zwischen weiter auseinanderliegenden Resten erhielten den Wert 4 Å. Wenn Einschränkungen für Gruppen von diastereotopen Substituenten oder Protonen von aromatischen Ringen auf im Schwerpunkt gelegene Pseudoatome bezogen wurden, erhöhten sich die oberen Distanzschranken um eine Pseudoatomkorrektur gemäß Wüthrich *et al.* (1983). Einschränkungen für die Diederwinkel ϕ , ψ und χ^1 lieferte das Programm HABAS (Güntert *et al.*, 1989; vgl. S. 94ff.) aufgrund lokaler Distanzschranken und skalarer Spin-Spin Kopplungen $^3J_{HN\alpha}$ und $^3J_{\alpha\beta}$, die mit Hilfe von Karplus-Beziehungen (Pardi *et al.*, 1984; DeMarco *et al.*, 1978) aus der regularisierten Kristallstruktur abgeleitet wurden (Einzelheiten siehe Güntert *et al.*, 1989). Wie üblich fixierten Distanzeinschränkungen die drei Disulfidbrücken (Williamson *et al.*, 1985).

Für beide Datensätze wurden mit dem Programm DISMAN (Braun & Gö, 1985) Strukturen berechnet und anschließend die Bündel der jeweils vier besten Konformere miteinander verglichen (Güntert *et al.*, 1989). Die folgende Tabelle zeigt, daß die vier besten Konformere beider Gruppen die Einschränkungen gut erfüllen:

Restverletzungen der ohne bzw. mit stereospezifischen Zuordnungen berechneten BPTI Strukturen

Restverletzungen	<i>ohne stereo- spezifische Zuordnungen</i>	<i>mit stereo- spezifischen Zuordnungen</i>
oberer Distanzschranken: mittlere Anzahl > 0,2 Å Maximum	6,1 0,93 Å	3,8 0,37 Å
sterischer Einschränkungen: mittlere Anzahl > 0,2 Å Maximum	6,3 0,58 Å	2,5 0,43 Å
von Diederwinkelschranken: mittlere Anzahl > 5° Maximum	1,6 16,4°	0 < 5°

Die mittlere Anzahl signifikanter Restverletzungen pro Konformer ist in beiden Fällen klein gegenüber der gesamten Anzahl Einschränkungen — 629 obere Distanzschranken im Datensatz ohne, 827 im Datensatz mit stereospe-

zifische Zuordnungen und 123 Diederwinkeleinschränkungen in beiden Datensätzen. Für den stärker einschränkenden Datensatz *mit* stereospezifischen Zuordnungen konvergierte das Programm DISMAN etwas besser, weshalb die Restverletzungen für die entsprechenden Strukturen etwas kleiner sind als für die Rechnungen ohne stereospezifische Zuordnungen. Dieser Effekt ist auf Konvergenzschwierigkeiten des Programms DISMAN zurückzuführen, die umso mehr in Erscheinung treten, je ungenauer ein Datensatz und je komplizierter die Faltung der Polypeptidkette ist, und kann durch Konvergenzbeschleunigung mit Hilfe redundanter Diederwinkeleinschränkungen behoben werden (vgl. S. 132ff.).

Über die Ausdehnung der beiden Strukturbündel und ihre Lage im Konformationsraum bezüglich der als Referenzstruktur dienenden regularisierten Kristallstruktur gibt die folgende Tabelle Auskunft:

Ausdehnung und Lage der ohne bzw. mit stereospezifischen Zuordnungen berechneten BPTI Strukturbündel im Konformationsraum

Mittlere Abweichungen der Atompositionen	<i>ohne</i> stereospezifische Zuordnungen	<i>mit</i> stereospezifischen Zuordnungen
paarweise innerhalb der Bündel:		
Rückgratatome der Reste 3–55	$1,0 \pm 0,3 \text{ \AA}$	$0,5 \pm 0,3 \text{ \AA}$
Schweratome der Reste 3–55	$1,7 \pm 0,2 \text{ \AA}$	$1,0 \pm 0,3 \text{ \AA}$
Rückgratatome 18–35 (β -Blatt)	$0,3 \pm 0,1 \text{ \AA}$	$0,2 \pm 0,1 \text{ \AA}$
Schweratome 18–35 (β -Blatt)	$1,3 \pm 0,3 \text{ \AA}$	$0,9 \pm 0,2 \text{ \AA}$
Rückgratatome 48–55 (α -Helix)	$0,7 \pm 0,3 \text{ \AA}$	$0,2 \pm 0,1 \text{ \AA}$
Schweratome 48–55 (α -Helix)	$1,4 \pm 0,2 \text{ \AA}$	$0,8 \pm 0,1 \text{ \AA}$
zur regularisierten Kristallstruktur:		
Rückgratatome der Reste 3–55	$0,9 \pm 0,2 \text{ \AA}$	$0,4 \pm 0,1 \text{ \AA}$
Schweratome der Reste 3–55	$1,6 \pm 0,3 \text{ \AA}$	$1,1 \pm 0,1 \text{ \AA}$
Rückgratatome 18–35 (β -Blatt)	$0,2 \pm 0,1 \text{ \AA}$	$0,2 \pm 0,1 \text{ \AA}$
Schweratome 18–35 (β -Blatt)	$1,4 \pm 0,3 \text{ \AA}$	$1,1 \pm 0,1 \text{ \AA}$
Rückgratatome 48–55 (α -Helix)	$0,7 \pm 0,3 \text{ \AA}$	$0,2 \pm 0,1 \text{ \AA}$
Schweratome 48–55 (α -Helix)	$1,3 \pm 0,3 \text{ \AA}$	$0,7 \pm 0,1 \text{ \AA}$

Die mittleren paarweisen Abweichungen der Atompositionen (vgl. S. 143) innerhalb der beiden Strukturbündel zeigen, daß stereospezifische Zuordnungen die Genauigkeit der Strukturen bedeutend verbessern können, insbesondere im Bereich des Rückgrats und in der α -Helix, während für das β -Faltblatt nur kleinere Fortschritte auftreten. Die mittleren Abweichungen der Atomp-

sitionen zur regularisierten Kristallstruktur sind durchwegs fast gleich den entsprechenden Werten innerhalb der Bündel, was zeigt, daß die Lage der zwei Strukturbündel im Konformationsraum die regularisierte Kristallstruktur einschließt und daß die Konformere keine *gerichtete* Tendenz zeigen, sich von der Referenzstruktur zu entfernen. Weitere Einzelheiten und Bilder der berechneten Strukturen enthält die Veröffentlichung von Güntert *et al.* (1989).

Diese Testrechnung mit simulierten Datensätzen gibt einen Hinweis darauf, inwiefern die Genauigkeit der berechneten Strukturen durch stereospezifische Zuordnungen *maximal* verbessert werden kann — maximal deshalb, weil es in der Praxis kaum möglich ist, alle hier vorausgesetzten stereospezifischen Zuordnungen experimentell zu bestimmen, sei es aus technischen Gründen oder weil die chemischen Verschiebungen diastereotoper Protonen zufällig entartet sind. Immerhin ist es beeindruckend, daß allein durch stereospezifische Zuordnungen bei sonst gleichen Daten die Ausdehnung des berechneten Strukturbündels etwa halbiert werden kann: die paarweise Abweichung der Rückgratotope der Reste 3–55 sinkt von 1,0 auf 0,5 Å.

Die Testrechnung hat einige Unvollkommenheiten, die die Aussagekraft der Ergebnisse schmälern: einerseits wird der Einfluß stereospezifischer Zuordnungen durch individuelle Zuordnungen der aromatischen Ringprotonen im Datensatz mit stereospezifischen Zuordnungen und Pseudoatomkorrekturen (Wüthrich *et al.*, 1983) im Datensatz ohne stereospezifische Zuordnungen, die häufig größer als nötig sind (vgl. S. 110; Güntert *et al.*, 1991a), zusätzlich verstärkt; andererseits ist die statistische Signifikanz beschränkt, weil beide Strukturbündel aus lediglich vier Konformeren bestehen und wegen Konvergenzschwierigkeiten des Programms DISMAN nicht die gleiche Qualität in bezug auf Restverletzungen der Einschränkungen besitzen. Die folgenden Testrechnungen mit experimentellen Datensätzen vermeiden diese Probleme weitgehend.

Einfluß stereospezifischer Zuordnungen bei experimentellen BPTI Datensätzen und verschiedenen Behandlungen von Distanzstrahlen mit Gruppen diastereotoper Substituenten

Die konservativen Pseudoatomkorrekturen von Wüthrich *et al.* (1983) bewirken für Einschränkungen mit diastereotopen Gruppen ohne stereospezifische Zuordnungen in vielen Fällen einen zu großen Informationsverlust, weil jede Einschränkung für sich behandelt wird und man nicht berücksichtigt, daß oft von einem Proton gleichzeitig Einschränkungen zu *beiden* diastereotopen Protonen vorliegen. Das Programm DIANA verwendet deshalb eine verbesserte Behandlung solcher Einschränkungen (vgl. S. 110; Güntert *et al.*, 1991a; im

Folgenden mit »DIANA« bezeichnet). Um zu testen, inwieweit der Genauigkeitsverlust durch fehlende stereospezifische Zuordnungen mit dieser verbesserten Behandlung wettgemacht werden kann, wurden auf der Grundlage eines experimentellen Datensatzes für BPTI (L. Orbons, P. Güntert und K. Wüthrich, unveröffentlicht; ein Vorläufer des Datensatzes von Berndt *et al.*, 1992) vier Datensätze erzeugt, die sich nur in stereospezifischen Zuordnungen und in der Behandlung von Einschränkungen bei fehlender stereospezifischer Zuordnung unterscheiden (Güntert *et al.*, 1991a):

Experimentelle Datensätze für BPTI zur Untersuchung des Einflusses verschiedener Behandlungen von Einschränkungen mit diastereotopen Gruppen

	I	II	III	IV
Stereospezifische Zuordnungen	42	42	0	0
Individuell zugeordnete NH ₂ Gruppen von Asn	3	3	0	0
Behandlung von Einschränkungen mit Gruppen diastereotoper Substituenten	DIANA	Wüthrich <i>et al.</i> (1983)	DIANA	Wüthrich <i>et al.</i> (1983)
obere Distanzschranken	866	781	924	637

Zusätzlich zu den oberen Distanzschranken, die aus Kern-Overhauser-effekten abgeleitet wurden, enthalten alle vier Datensätze identische Distanzeinschränkungen für die 3 Disulfidbrücken und 11 experimentell bestätigte Wasserstoffbrücken (Williamson *et al.*, 1985) sowie 140 mit dem Programm HABAS bestimmte Diederwinkelschranken. Die unterschiedliche Anzahl oberer Distanzschranken in den vier Datensätzen ist eine direkte Folge der unterschiedlichen Behandlung von Einschränkungen mit Gruppen diastereotoper Substituenten. Datensatz I entspricht dem vollständigen experimentellen Datensatz mit allen experimentell bestimmten stereospezifischen Zuordnungen und der bevorzugten Methode zur Behandlung von Distanzeinschränkungen für prochirale Zentren ohne individuelle Zuordnung; die Datensätze II–IV entsprechen verschiedenen Stufen der Abschwächung der Einschränkungen.

Für jeden Datensatz wurde eine DIANA Strukturrechnung ohne Konvergenzbeschleunigung mit 150 Zufallsstartstrukturen und maximal 9900 Zielfunktionsauswertungen pro Konformer angesetzt. Auf einem Cray X-MP Hochleistungsrechner betrug die gesamte Rechenzeit 18,6 Std. Für die Datensätze I–IV ergaben sich 17, 19, 10 bzw. 6 Konformere guter Qualität, d. h. mit

einem Endwert der Zielfunktion unter 5 \AA^2 . In Übereinstimmung mit der Testrechnung für simulierte Daten (vgl. S. 86) zeigt der Vergleich der Strukturbündel I und III bzw. II und IV, daß stereospezifische Zuordnungen auch bei experimentellen Daten die Genauigkeit der Distanzgeometriestrukturen verbessern, wenngleich die Unterschiede weniger ausgeprägt sind als bei simulierten Daten — vermutlich weil nicht alle diastereotopen Gruppen stereospezifisch zugeordnet werden können:

Ausdehnung der BPTI Strukturbündel I–IV im Konformationsraum

Mittlere paarweise Abweichungen [\AA] der	I	II	III	IV
Rückgratatome der Reste 1–58	1,1	1,2	1,1	1,4
Rückgratatome der Reste 3–55	0,9	0,9	1,0	1,3
Rückgratatome 18–35 (β -Blatt)	0,6	0,6	0,6	0,8
Rückgratatome 48–55 (α -Helix)	0,3	0,3	0,3	0,4
Schweratome der Reste 1–58	1,8	1,9	2,0	2,2
Schweratome der Reste 3–55	1,7	1,8	1,9	2,1
Schweratome 18–35 (β -Blatt)	1,1	1,1	1,3	1,4
Schweratome 48–55 (α -Helix)	1,3	1,2	1,4	1,6

Ferner zeigt die Tabelle, daß die DIANA-Methode zur Behandlung diastereotoper Gruppen vor allem dann die Genauigkeit der Strukturbündel erhöht, wenn keine stereospezifischen Zuordnungen vorhanden sind. Ein Vergleich der Strukturbündel II und III belegt, daß der Informationsverlust durch fehlende stereospezifische Zuordnungen mit dem DIANA-Verfahren zum Teil, aber nicht vollständig wettgemacht werden kann.

Einfluß stereospezifischer Zuordnungen bei experimentellen BPTI Datensätzen und Strukturrechnungen mit Konvergenzbeschleunigung

Die beiden vorhergehenden Testrechnungen beruhen auf Strukturrechnungen mit den Programmen DISMAN oder DIANA ohne Konvergenzbeschleunigung. Die folgende Testrechnung, die der heutigen Praxis der Strukturberechnung mit Konvergenzbeschleunigung aufgrund redundanter Diederwinkelschränkungen entspricht (vgl. S. 132ff.), vergleicht die veröffentlichten DIANA-Strukturen von BPTI (Berndt *et al.*, 1992) mit einem Struk-

turbündel, dessen Berechnung bis auf die Elimination der stereospezifischen Zuordnungen völlig mit der Rechnung von Berndt *et al.* (1992) übereinstimmt. Die Anzahl relevanter oberer Distanzschranken aus Kern-Overhausereffekten beträgt 642 im Originaldatensatz, der 32 stereospezifische Zuordnungen und individuelle Zuordnungen für einen Glutamin- und drei Asparaginreste enthält, und 684 im Datensatz ohne stereospezifische Zuordnungen. Beide Datensätze verwenden die gleichen 115 Diederwinkelschranken und gleiche Distanzschranken zur Fixierung der Disulfidbrücken (Williamson *et al.*, 1985).

Im Gegensatz zu den DIANA Rechnungen ohne Konvergenzbeschleunigung bewirkt die Konvergenzbeschleunigung, daß der Datensatz ohne stereospezifische Zuordnungen Strukturen mit etwas niedrigeren Werten für Zielfunktion und Restverletzungen liefert als der stärker einschränkende Originaldatensatz von Berndt *et al.* (1992) mit stereospezifischen Zuordnungen:

Restverletzungen der mit und ohne stereospezifische Zuordnungen berechneten BPTI Strukturen

Restverletzungen	<i>mit stereo-</i> <i>spezifischen</i> <i>Zuordnungen</i> <i>(Berndt et al.,</i> <i>1992)</i>	<i>ohne stereo-</i> <i>spezifische</i> <i>Zuordnungen</i>
DIANA Zielfunktionswert	$0,40 \pm 0,10 \text{ \AA}^2$	$0,14 \pm 0,07 \text{ \AA}^2$
Mittlere Restverletzungen oberer Distanzschranken:		
Anzahl > 0,2 \AA	0 ± 1	0 ± 0
Summe	$3,0 \pm 0,4 \text{ \AA}$	$1,2 \pm 0,4 \text{ \AA}$
Maximum	$0,22 \pm 0,06 \text{ \AA}$	$0,13 \pm 0,04 \text{ \AA}$
sterischer Einschränkungen:		
Anzahl > 0,2 \AA	0 ± 0	0 ± 0
Summe	$1,5 \pm 0,3 \text{ \AA}$	$0,8 \pm 0,3 \text{ \AA}$
Maximum	$0,12 \pm 0,04 \text{ \AA}$	$0,10 \pm 0,05 \text{ \AA}$
von Diederwinkelschranken:		
Anzahl > 5°	0 ± 0	0 ± 0
Summe	$5,8 \pm 2,0^\circ$	$3,2 \pm 2,0^\circ$
Maximum	$2,2 \pm 1,0^\circ$	$1,2 \pm 0,6^\circ$

Dies unterstreicht erneut, daß das gegenteilige Ergebnis in den beiden vorangehenden Testrechnungen ohne Konvergenzbeschleunigung auf Konver-

genzuschwierigkeiten zurückgeht, die umso mehr zunehmen, je »loser« die Einschränkungen sind, und keine den Datensätzen innewohnende Eigenschaft ist.

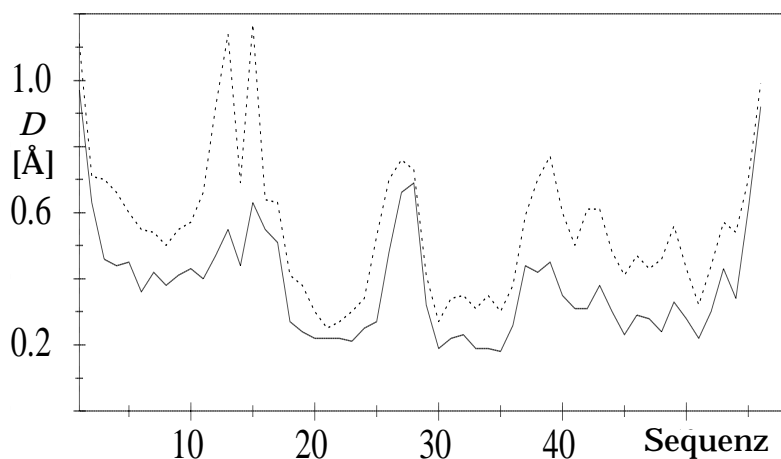


Bild 11. Mittlere Atomverschiebungen D (vgl. S. 144) der Rückgratkatome der einzelnen Aminosäurereste nach globaler Überlagerung der Reste 2–56 der je 20 besten DIANA-Konformere, die aufgrund des experimentellen Datensatzes für BPTI von Berndt et al. (1992) mit (ausgezogene Linie) bzw. ohne (gepunktete Linie) stereospezifische Zuordnungen berechnet wurden.

Die folgende Tabelle enthält die mittleren Abweichungen der Atompositionen in den einzelnen Konformeren relativ zur gemittelten Struktur, die durch Überlagern der Rückgratkatome N, C $^{\alpha}$ und C' in den Resten 2–56 und anschließendes Mitteln der kartesischen Koordinaten berechnet wurde:

Ausdehnung der mit und ohne stereospezifische Zuordnungen berechneten BPTI Strukturbündel im Konformationsraum

Abweichungen der Atompositionen zur gemittelten Struktur	mit stereospezifischen Zuordnungen (Berndt et al., 1992)	ohne stereospezifische Zuordnungen
Rückgratkatome der Reste 2–56	$0,43 \pm 0,08 \text{ \AA}$	$0,63 \pm 0,07 \text{ \AA}$
Schwerkatome der Reste 2–56	$0,96 \pm 0,09 \text{ \AA}$	$1,14 \pm 0,09 \text{ \AA}$
Rückgratkatome 18–35 (β -Blatt)	$0,28 \pm 0,13 \text{ \AA}$	$0,31 \pm 0,12 \text{ \AA}$
Schwerkatome 18–35 (β -Blatt)	$0,68 \pm 0,11 \text{ \AA}$	$0,73 \pm 0,09 \text{ \AA}$
Rückgratkatome 47–56 (α -Helix)	$0,34 \pm 0,11 \text{ \AA}$	$0,32 \pm 0,08 \text{ \AA}$
Schwerkatome 47–56 (α -Helix)	$0,86 \pm 0,11 \text{ \AA}$	$0,88 \pm 0,13 \text{ \AA}$

Die stereospezifischen Zuordnungen verringern die Abweichungen für die Reste 2–56 signifikant — um $0,2 \text{ \AA}$ — sowohl für die Rückgrat- als auch für alle Schweratome, wogegen die Verbesserungen innerhalb der regulären Sekundärstrukturelemente gering sind. Im Einklang damit weisen im Bild 11 die Verschiebungen der Rückgratatompositionen der einzelnen Aminosäurereste nach globaler Überlagerung verhältnismäßig große Unterschiede im Bereich der Reste 3–15 und 38–48 auf. Ein visueller Vergleich der Strukturbündel, die mit und ohne stereospezifische Zuordnungen berechnet wurden, in Bild 12 zeigt ebenfalls deutlich die verbesserte Präzision, mit der die Konformation des Rückgrats bestimmt werden kann, wenn stereospezifische Zuordnungen vorliegen.

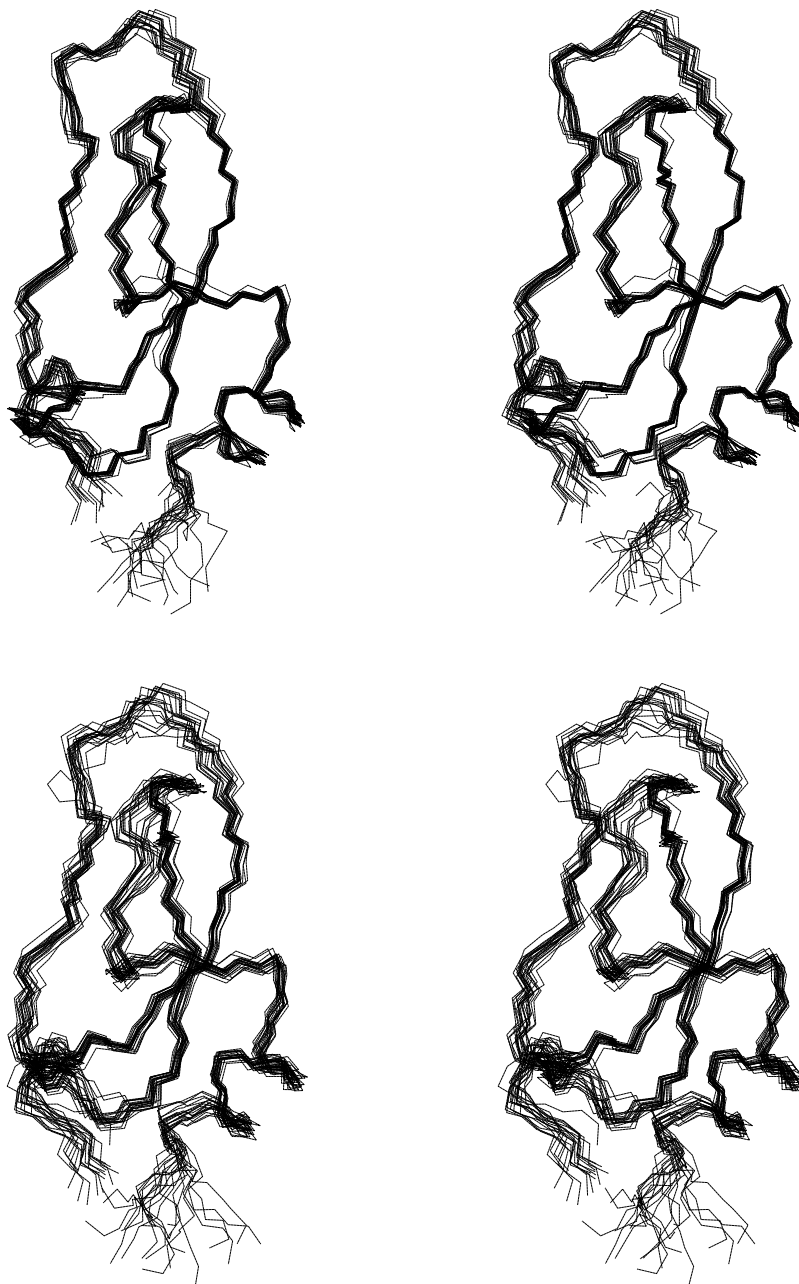


Bild 12. Stereobilder, erzeugt mit dem Programm XAM (Xia, 1992), der je 20 BPTI-Konformere, die aufgrund des experimentellen Datensatzes von Berndt et al. (1992) mit (oben) bzw. ohne (unten) stereospezifische Zuordnungen berechnet wurden. Die Positionen der Atome N, C^α und C' sind entlang des Polypeptidrückgrats verbunden. Deutlich sichtbar ist die genauere Bestimmung der Konformation des Rückgrats, falls stereospezifische Zuordnungen verwendet werden.

Bestimmung stereospezifischer Zuordnungen

Das allgemein gebrauchte Verfahren zur sequentiellen Resonanzzuordnung in Proteinen (Wüthrich *et al.*, 1983; Wüthrich, 1986) liefert keine stereospezifischen Zuordnungen für Gruppen von diastereotopen Substituenten. Kern-Overhausereffekte, die diastereotope Protonen oder Methylgruppen betreffen, können trotzdem für die Strukturrechnung verwendet werden, wenn Pseudoatome als Referenzpunkte für Distanzeinschränkungen dienen (Wüthrich *et al.*, 1983). Dieser Kompromiß verringert die Genauigkeit der experimentellen Einschränkungen und führt, wie die obigen Testrechnungen gezeigt haben, zu weniger genau bestimmten Strukturen. Dieser Abschnitt stellt zwei rechnerische Methoden zur Bestimmung stereospezifischer Zuordnungen vor, die die manuelle Analyse von Spin-Spin Kopplungskonstanten und kurzreichweitigen Kern-Overhausereffekten (Wagner *et al.*, 1987; Hyberts *et al.*, 1987; Arseniev *et al.*, 1988; Kline *et al.*, 1988) sowie den empirischen Vergleich von charakteristischen Distanzen in vorläufigen Konformeren (Kline *et al.*, 1988) auf eine systematische Grundlage stellen und automatisieren. Neben den beiden hier vorzustellenden Verfahren stehen noch andere rechnerische Methoden, insbesondere das Verfahren der »fließenden stereospezifischen Zuordnungen« (Weber *et al.*, 1988), das aber nicht immer zuverlässig ist (Havel, 1991), sowie — für die diastereotopen Methylgruppen von Valin und Leucin — die Methode der biosynthetischen stereospezifischen Markierung mit ^{13}C (Senn *et al.*, 1989; Neri *et al.*, 1989; Szyperski *et al.*, 1992d) zur Verfügung.

Stereospezifische Zuordnungen aufgrund systematischer Analyse der lokalen Einschränkungen mit dem Programm HABAS

Das Programm HABAS (Güntert, 1987; Güntert *et al.*, 1989) untersucht systematisch experimentelle Datensätze, um stereospezifische Zuordnungen für β -Methylenprotonen zu bestimmen. Es wird vor der Strukturrechnung gebraucht und verwendet nur lokale experimentelle Einschränkungen. Für einen gegebenen Aminosäurerest i untersucht HABAS die Einschränkungen, die nur von den Diederwinkeln ϕ_i , ψ_i und χ_i^1 abhängen, d.h. alle sterischen und experimentellen Einschränkungen im Bereich des Dipeptidsegments von Bild 13.

Sterische Einschränkungen werden behandelt, indem jedem Atom ein Radius zugeordnet wird und eine Verletzung genau dann auftritt, wenn die Distanz zwischen zwei Atomen die Summe der beiden Atomradien unterschreitet. Die Atomradien haben die gleichen Werte wie im Programm DIANA (vgl. S. 68). Obere Schranken, die sich aus Kern-Overhausereffekten ergeben, berücksichtigt das Programm HABAS für die Proton-Proton Distanzen

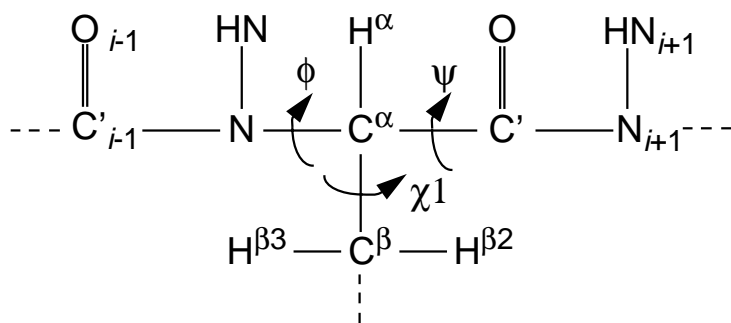


Bild 13. Dipeptidsegment, dessen Konformation vom Programm HABAS systematisch untersucht wird, um die stereospezifische Zuordnung der β -Methylenprotonen zu bestimmen.

$d_{N\alpha}(i, i)$, $d_{N\beta2}(i, i)$, $d_{N\beta3}(i, i)$, $d_{\alpha\beta2}(i, i)$ und $d_{\alpha\beta3}(i, i)$ innerhalb des Rests i , sowie d_{NN} , $d_{\alpha N}$, $d_{\beta2N}$ und $d_{\beta3N}$ zwischen den Resten i und $i + 1$. Zusätzlich verarbeitet das Programm HABAS auch Relationen zwischen den zwei Distanzen $d(A, B')$ und $d(A, B)$ von einem Proton A zu den beiden Protonen B und B' einer Methylengruppe. Eine Relation ist erfüllt, wenn $d(A, B') > d(A, B) + \Delta d$ gilt, wobei Δd ein Parameter ist, der üblicherweise den Wert $\Delta d = 0$ hat. Dies ermöglicht es, relative Intensitäten von Kern-Overhausereffekten zu verwenden, die oft leichter zugänglich sind als Abschätzungen absoluter Distanzen. Experimentelle Werte ${}^3J^{\text{exp}} \pm \Delta {}^3J$ der skalaren Spin-Spin Kopplungskonstanten führt HABAS gemäß den Karplusbeziehungen (Pardi *et al.*, 1984; DeMarco *et al.*, 1978)

$${}^3J_{HN\alpha}(\phi) = 6.4 \cos(\phi - 60^\circ)^2 - 1.4 \cos(\phi - 60^\circ) + 1.9, \quad [37]$$

$${}^3J_{\alpha\beta2}(\chi^1) = 9.5 \cos(\chi^1 - 120^\circ)^2 - 1.6 \cos(\chi^1 - 120^\circ) + 1.8, \quad [38]$$

$${}^3J_{\alpha\beta3}(\chi^1) = 9.5 \cos(\chi^1)^2 - 1.6 \cos\chi^1 + 1.8 \quad [39]$$

in erlaubte Bereiche der Diederwinkel ϕ und χ^1 über.¹

HABAS bestimmt die stereospezifische Zuordnung der β -Protonen, indem es für die zwei möglichen stereospezifischen Zuordnungen — wie in der Eingabe oder umgekehrt — den durch die drei Diederwinkel ϕ , ψ und χ^1 aufgespannten lokalen Konformationsraum in Schritten von $\Delta\phi = \Delta\psi = \Delta\chi^1 = 10^\circ$ systematisch auf erlaubte Konformationen absucht, d.h. für jeden Gitterpunkt prüft, ob alle sterischen und experimentellen Einschränkungen erfüllt sind. Wenn nur für eine der beiden möglichen stereospezifischen Zuordnungen erlaubte Konformationen auftreten, ist die stereospezifische Zuordnung eindeutig bestimmt, anderenfalls kann sie aufgrund der verfügbaren lokalen Daten

¹ Für das β -Methinproton von Valin gilt Gleichung [39], für dasjenige von Isoleucin oder Threonin gilt Gleichung [38].

entweder nicht eindeutig festgelegt werden (wenn erlaubte Konformationen für beide möglichen Zuordnungen auftreten) oder die lokalen Einschränkungen sind inkonsistent (wenn gar keine erlaubten Konformationen erscheinen). Für die Suche nach erlaubten Konformationen wird die ECEPP Standardgeometrie des Segments Ala–Ser–Ala verwendet (Momany *et al.*, 1975; Némethy *et al.*, 1983). Von den $36^3 = 46656$ Gitterpunkten müssen nur 13050, die den sterischen Bedingungen (Ramachandran-Plot) genügen, auf Konsistenz mit den experimentellen Einschränkungen überprüft werden.

Durch das systematische Absuchen des Gitters im lokalen Konformationsraum ermöglicht es das Programm HABAS grundsätzlich, alle aufgrund geometrischer Kriterien durch die lokalen Einschränkungen eindeutig definierten stereospezifischen Zuordnungen zu finden, ohne dabei (mit korrekten Eingabedaten) Fehler zu machen. In der Praxis können die Ergebnisse allerdings durch experimentelle Ungenauigkeiten oder interne Beweglichkeit der Proteine verzerrt werden, weil die »harten« Limiten für Einschränkungen jede Konformation, in der eine Einschränkung auch nur sehr wenig verletzt ist, ausschließen. Die Erfahrung hat gezeigt, daß man bei der Anwendung »harter« Limiten zu falschen Schlußfolgerungen gelangen kann, wenn die Eingabedaten kleine Inkonsistenzen aufweisen oder wenn die kovalente Struktur von der ECEPP Standardgeometrie abweicht. Deshalb besitzt das Programm HABAS auch die Möglichkeit, »weiche« Limiten anzuwenden, bei denen geringe Verletzungen von Einschränkungen, üblicherweise 0,1 Å für Distanzeinschränkungen und Distanzrelationen sowie 0,5 Hz für Kopplungskonstanten, noch toleriert werden.

Das Programm HABAS durchläuft alle Werte des χ^1 Winkels und vermeidet damit die Annahme, daß die erlaubten Werte für χ^1 auf enge Bereiche um die gestaffelten Rotamerpositionen $\chi^1 = -60, 60$ und 180° beschränkt seien, die in vielen empirischen stereospezifischen Zuordnungsverfahren gemacht wird, die lokale Einschränkungen verwenden (Wagner *et al.*, 1987; Arseniev *et al.*, 1988; Kline *et al.*, 1988). Um die Auswirkung der Beschränkung auf die gestaffelten Rotamerpositionen mit dem Programm HABAS zu untersuchen, kann χ^1 auf Bereiche von $\pm 20^\circ$ um diese Positionen eingeschränkt werden. Der Vergleich der Ergebnisse, die mit und ohne die Beschränkung auf die gestaffelten Rotamerpositionen erhalten werden, hat gezeigt, daß diese Einschränkung nur sehr wenige zusätzliche stereospezifische Zuordnungen ermöglicht, aber zu Inkonsistenzen oder sogar falschen stereospezifischen Zuordnungen führen kann (Güntert *et al.*, 1989).

Als ein Nebenprodukt des systematischen Absuchens des lokalen Konformationsraums liefert das Programm HABAS — unabhängig davon, ob eine eindeutige stereospezifische Zuordnung möglich ist oder nicht — erlaubte Intervalle für die Diederwinkel ϕ , ψ und χ^1 , die mit den lokalen experimentel-

len und sterischen Einschränkungen verträglich sind und als Winkeleinschränkungen für die Strukturrechnung mit dem Programm DIANA (vgl. S.104ff.) Verwendung finden.

Drei typische Fälle von HABAS Ergebnissen zeigt die folgende Tabelle anhand von Beispielen, die dem experimentellen Datensatz für Hirudin(1–51) (Szyperski *et al.*, 1992b) entnommen sind. n_e bezeichnet die Anzahl Konformationen, die mit der stereospezifischen Zuordnung der Eingabedaten, n_u die Anzahl Konformationen, die mit der umgekehrten stereospezifischen Zuordnung verträglich sind. Die erlaubten Diederwinkelwerte für die stereospezifische Zuordnung der Eingabedaten sind durch »/«, diejenigen für die umgekehrte Zuordnung durch »\« und diejenigen für beide möglichen Zuordnungen durch »X« gekennzeichnet. Die rechte Spalte enthält die von HABAS erzeugten Diederwinkeleinschränkungen.

Beispiele von HABAS Ergebnissen für Hirudin(1–51)

Aminosäurerest	Erlaubte Diederwinkelwerte						erzeugte Winkeleinschränkung	
	n_e	n_u		-90°	0°	90°		170°
Cys 6	175	0	ϕ :	/				-155... - 85°
			ψ :	/			/	105...195°
			χ^1 :	/				-105... - 75°
Lys 27	816	645	ϕ :	X		X		45...295°
			ψ :	X			X	125...185°
			χ^1 :	/		\	/	5...355°
Ile 29	623	-	ϕ :	/				-155... - 85°
			ψ :	/			/	105...195°
			χ^1 :	/				-105... - 75°

Das erste Beispiel, Cys 6, zeigt einen Aminosäurerest, wo die β -Protonen eindeutig stereospezifisch zugeordnet werden können: da $n_e > 0$ und $n_u = 0$ ist, ist die stereospezifische Zuordnung der Eingabedaten richtig. Lys 27 ist ein Beispiel für den Fall, daß keine eindeutige stereospezifische Zuordnung möglich ist, weil $n_e > 0$ und $n_u > 0$ ist. Isoleucin, Threonin und Valin besitzen nur ein β -Proton; das Problem der stereospezifischen Zuordnung stellt sich also nicht. Das dritte Beispiel, Ile 29, zeigt, daß HABAS auch für diese Aminosäuretypen Diederwinkeleinschränkungen erzeugen kann.

Anzahl von stereospezifischen Zuordnungen, die durch lokale Konformationsanalyse in Proteinen bestimmt werden können

Jede Methode zur stereospezifischen Zuordnung diastereotoper Gruppen ist auf verschiedenen Ebenen limitiert. Die offensichtlichste Beschränkung tritt in der Praxis auf, wenn die chemischen Verschiebungen diastereotoper Protonen zufälligerweise entartet sind, so daß sie nicht einzeln beobachtbar sind, oder wenn Bindungen hohe Rotationsbeweglichkeit aufweisen. Beide Möglichkeiten werden hier nicht explizit behandelt, aber man sollte sich bewußt sein, daß ihr Auftreten die Anzahl möglicher stereospezifischer Zuordnungen gegenüber den Ergebnissen der folgenden Testrechnungen, die entartete chemische Verschiebungen ausschließen und starre Konformationen annehmen, stets verringern.

Mit Verfahren wie HABAS, die ausschließlich lokale Einschränkungen verwenden, kann die stereospezifische Zuordnung grundsätzlich immer festgelegt werden, vorausgesetzt daß eine genügende Menge exakter Daten vorliegt und die Standardgeometrie exakt eingehalten wird. In Kernresonanzuntersuchungen von Proteinen ist es aber zum Beispiel nicht einfach, zuverlässige nicht-triviale untere Schranken für Proton-Proton Distanzen zu bestimmen, obere Distanzschranken genauer als etwa 0,2 Å zu messen, oder skalare Kopplungskonstanten mit weniger als ± 1.0 Hz Fehler festzulegen. Wenn Daten mit der Genauigkeit, die in Kernresonanzmessungen mit Proteinen erhältlich ist, verwendet werden, ist es daher möglich, daß das Auftreten gewisser lokaler Konformationen eine eindeutige stereospezifische Zuordnung nicht zuläßt. Die beiden folgenden Testrechnungen (Güntert *et al.*, 1989) schätzen das Ausmaß eindeutiger stereospezifischer Zuordnungen, das mit realistischen experimentellen Daten erwartet werden kann, ab. Zunächst wurden für die sterisch erlaubten Konformationen (13050 auf einem Gitter von 10° Maschenweite) des Dipeptidsegments in Bild 13 experimentelle Daten verschiedener Genauigkeit simuliert, und das Ausmaß eindeutiger stereospezifischer Zuordnungen bestimmt. In der zweiten Testrechnung wurde das Programm HABAS auf simulierte Datensätze für eine Auswahl hochaufgelöster Kristall- und Lösungsstrukturen von Proteinen angewandt.

Die Ergebnisse in der folgenden Tabelle beruhen auf simulierten Datensätzen, in denen die von HABAS verwendeten Distanzen (vgl. S. 94) durch die kleinstmögliche der oberen Schranken 2,5, 3,0, 3,5 und 4,0 Å, die über der Distanz in der Ausgangskonformation liegt, eingeschränkt sind und skalare Kopplungskonstanten aufgrund der Gleichungen [37–39] berechnet und mit verschiedenen Genauigkeiten $\Delta^3 J$ versehen wurden. n_r bzw. n_f bezeichnet die Anzahl Konformationen, die mit der richtigen bzw. falschen stereospezifischen Zuordnung verträglich sind. Die Kolonne $n_f = 0$ gibt den Anteil der Ausgangs-

konformationen an, für die HABAS eine eindeutige stereospezifische Zuordnung findet; die Kolonnen $n_r \geq 10n_f$ und $n_r \geq 2n_f$ geben die entsprechenden Anteile für weniger strikte aber fehleranfälliger Kriterien zur stereospezifischen Zuordnung an; die letzte Kolonne, $n_f > n_r$, gibt den Anteil der Ausgangskonformationen, deren Daten mit mehr Konformationen verträglich sind, wenn die falsche anstelle der richtigen stereospezifischen Zuordnung angenommen wird (aber in keinem einzigen Fall lieferte HABAS eine eindeutig falsche stereospezifische Zuordnung, d.h. es ist immer $n_r > 0$):

Ergebnisse eines systematischen Tests des Programms HABAS

$\Delta^3 J$	Anteil Ausgangskonformationen mit			
	$n_f = 0$	$n_r \geq 10n_f$	$n_r \geq 2n_f$	$n_f > n_r$
1,0 Hz	73%	74%	82%	12%
2,0 Hz	52%	52%	70%	12%
3,0 Hz	32%	37%	60%	16%

Das wichtigste Ergebnis ist, daß bei weitem nicht immer eine eindeutige stereospezifische Zuordnung möglich ist. So kann sie zum Beispiel bei einer in der Praxis üblichen Genauigkeit¹ der skalaren Kopplungskonstanten von ± 2.0 Hz nur in etwa der Hälfte aller Fälle aufgrund der in HABAS verwendeten lokalen Daten eindeutig festgelegt werden (Güntert *et al.*, 1989). Ein leicht gelockertes Kriterium für stereospezifische Zuordnungen ($n_r \geq 10n_f$ anstatt $n_f = 0$) bringt keine Verbesserung und kann deshalb fallengelassen werden.

Die Ergebnisse in der obigen Tabelle sind nur indirekt auf Proteine anwendbar, weil nicht berücksichtigt wird, daß in nativen Proteinen gewisse lokale Konformationen bevorzugt auftreten. Daher wurden in einem zweiten Test für 13 Kristallstrukturen aus der Proteindatenbank² (Bernstein *et al.*,

¹ Die Werte $\Delta^3 J$ für die Genauigkeit der skalaren Kopplungskonstanten müssen nicht nur Meßungenauigkeiten sondern auch Abweichungen von der theoretischen Form und Fehler in der Kalibrierung der Karpluskurven (Karplus, 1963) berücksichtigen.

² Alle damals in der Proteindatenbank vorhandenen Kristallstrukturen von Proteinen mit 30–99 Aminosäureresten und einer Auflösung von 2,0 Å oder besser (außer Ferredoxin, dessen Struktur mit einer Methode verfeinert wurde, die sie für diese Untersuchung ungeeignet macht): *Crambin* (Code der Proteindatenbank: 1CRN), *L7/L12 50 S ribosomales Protein* (1CTF), *high-potential iron-sulphur protein* (1HIP), *avian pancreatic polypeptide* (1PPT), *Skorpionen-Neurotoxin* (1SN3), *Cytochrome b₅ (oxidiert)* (2B5C), *Erabutoxin b* (2EBX), *Ovomucoid* (dritte Domäne) (2OVO), *Cytochrome c₅₅₁ (reduced)* (451C), BPTI (4PTI), *Rubredoxin (oxidiert)* (1RXN) und *Plastocyanin* (6PCY).

1977) und die Lösungsstruktur von Tendamistat (Kline *et al.*, 1988) simulierte lokale Einschränkungen auf die gleiche Weise wie im obigen systematischen Test mit $\Delta^3 J = 2.0$ Hz erzeugt und als Eingabedaten für HABAS verwendet. Ohne Proline und N- und C-terminale Reste enthalten diese Proteine insgesamt 496 β -Methylengruppen, wovon sich gemäß den Angaben der Proteindatenbank 133 in Helices und 132 in β -Blättern befinden. HABAS lieferte eindeutige stereospezifische Zuordnungen für 42% aller β -Methylengruppen, 38% derjenigen in Helices und signifikant mehr, nämlich 57%, derjenigen in β -Blättern. Schränkt man die erlaubten Werte des Diederwinkels χ^1 auf $\pm 20^\circ$ um die gestaffelten Rotamerpositionen $\chi^1 = -60, 60$ und 180° ein, ergeben sich 49%, d.h. nur unwesentlich mehr, eindeutige stereospezifische Zuordnungen, aber es treten für 10% der β -Methylengruppen Widersprüche oder falsche stereospezifische Zuordnungen auf. Dies unterstreicht erneut, daß die Einschränkung auf die gestaffelten Rotamerpositionen vermieden werden sollte (Güntert *et al.*, 1989).

Stereospezifische Zuordnungen aufgrund vorläufiger dreidimensionaler Strukturen mit dem Programm GLOMSA

Stereospezifische Zuordnungen können unter Einbezug *aller* vorhandenen Einschränkungen gemacht werden, indem man mit einem vorläufigen Datensatz vollständig gefaltete dreidimensionale Strukturen berechnet und anschließend Paare von Distanzeinschränkungen zu diastereotopen Protonen oder Methylgruppen mit den entsprechenden Paaren von Distanzen im Strukturbündel vergleicht. Das Programm GLOMSA (Güntert *et al.*, 1991a) führt eine entsprechende Statistik durch; es benötigt dazu ein Bündel von m dreidimensionalen Proteinstrukturen und die Liste der Distanzeinschränkungen. In einem typischen Fall seien zwei obere Schranken b_1 und b_2 für die Distanzen zwischen einem einzelnen Proton α und den zwei diastereotopen Protonen β_1 und β_2 vorhanden, und es bezeichne d_{kI} die Distanz zwischen den Atomen α und β_k ($k = 1, 2$) im Konformer I ($I = 1, \dots, m$). Dann berechnet GLOMSA die Summe V der Restverletzungen der beiden oberen Distanzschranken in den m Konformeren für beide möglichen stereospezifischen Zuordnungen e und u (« e » steht für diejenige der Eingabedaten, « u » für die umgekehrte Zuordnung):

$$V^e = \sum_{I=1}^m (\theta(d_{1I} - b_1)(d_{1I} - b_1) + \theta(d_{2I} - b_2)(d_{2I} - b_2)) \quad , \quad [40]$$

$$V^u = \sum_{I=1}^m (\theta(d_{1I} - b_2)(d_{1I} - b_2) + \theta(d_{2I} - b_1)(d_{2I} - b_1)) \quad [41]$$

und das Minimum v über die m Konformere der größeren Verletzung der bei-

den Distanzschranken b_1 und b_2 :

$$v^e = \min_{l=1, \dots, m} \max(\theta(d_{1l} - b_1)(d_{1l} - b_1), \theta(d_{2l} - b_2)(d_{2l} - b_2)), \quad [42]$$

$$v^u = \min_{l=1, \dots, m} \max(\theta(d_{1l} - b_2)(d_{1l} - b_2), \theta(d_{2l} - b_1)(d_{2l} - b_1)). \quad [43]$$

θ bezeichnet die Heavysidefunktion, die für positive Argumente den Wert 1 annimmt und sonst verschwindet. Ein auf der minimalen Verletzung v beruhendes Verfahren ergibt eine stereospezifische Zuordnung, wenn $v^e = 0$ und $v^u > 0$ ist oder umgekehrt. Die Erfahrung zeigt allerdings, daß dieses Kriterium recht restriktiv ist und in der Praxis nur wenige stereospezifische Zuordnungen liefert. Ähnliches gilt für ein etwas weniger einschränkendes Verfahren, daß die Größe der Summen der Verletzungen V^e und V^u miteinander vergleicht.

Daher berechnet das Programm GLOMSA die Größen V und v nur zur Information des Benutzers und verwendet zusätzlich den Mittelwert $\overline{\Delta d}$ der Differenzen $\Delta d_l = d_{1l} - d_{2l}$ in den m Konformeren:

$$\overline{\Delta d} = \frac{1}{m} \sum_{l=1}^m \Delta d_l \quad [44]$$

und die maximale Anzahl Konformationen, in denen Δd_l das gleiche Vorzeichen hat:

$$n_{\Delta d} = \begin{cases} s, & \text{falls } s > m/2; \\ s - m, & \text{falls } s \leq m/2; \end{cases} \quad \text{mit } s = \sum_{l=1}^m \theta(\Delta d_l). \quad [45]$$

$n_{\Delta d}$ ist so konstruiert, daß immer $|n_{\Delta d}| \geq m/2$ gilt und ein positives Vorzeichen genau dann auftritt, wenn in der Mehrzahl der Konformere d_{1l} größer als d_{2l} ist. In Grenzfällen können die Vorzeichen von $\overline{\Delta d}$ und $n_{\Delta d}$ verschieden sein.

Um stereospezifische Zuordnungen zu finden, vergleicht GLOMSA die Vorzeichen von $\overline{\Delta d}$ und $n_{\Delta d}$ mit demjenigen von $\Delta b = b_1 - b_2$. Übereinstimmende Vorzeichen bestätigen die stereospezifische Zuordnung e in den Eingabedaten; unterschiedliche Vorzeichen geben die umgekehrte Zuordnung u an. Um Fälle, in denen die vorhandenen Daten nicht klar zwischen den beiden möglichen Zuordnungen zu unterscheiden vermögen auszuschließen, müssen die Größen $|\Delta b|$, $|\overline{\Delta d}|$ und $|n_{\Delta d}|$ vom Benutzer vorgegebene Schwellenwerte überschreiten, damit GLOMSA eine stereospezifische Zuordnung festlegt.

Sind für zwei diastereotope Protonen oder Methylgruppen Paare von oberen Distanzschranken zu mehreren verschiedenen Protonen vorhanden, führt GLOMSA die obige Statistik für jedes Paar getrennt durch und überläßt es dem Benutzer, die Ergebnisse zusammenzufassen, weil ein automatisches Kombinieren möglicherweise vorhandene Widersprüche verschleiern könnte.

Verschiedene Aspekte der Ergebnisse von GLOMSA beleuchtet die folgende

Tabelle anhand von Beispielen, die während der Strukturbestimmung von Hirudin(1–51) (Szyperski *et al.*, 1992b) auftraten. Die Ergebnisse beruhen auf der Untersuchung von $m = 20$ Konformeren und den Schwellenwerten $|\Delta b| \geq 0.4 \text{ \AA}$, $|\overline{\Delta d}| \geq 0.4 \text{ \AA}$ und $|n_{\Delta d}| = 20$, d.h. es werden Paare von Distanzschranken berücksichtigt, wenn sich die beiden Limiten um mindestens $0,4 \text{ \AA}$ unterscheiden, und eine stereospezifische Zuordnung (e , wie in den Eingabedaten, oder u , umgekehrt) wird vorgeschlagen, wenn eine der beiden Distanzen in allen 20 Konformeren kürzer als die andere ist und ihr mittlerer Unterschied mindestens $0,4 \text{ \AA}$ erreicht:

Beispiele von GLOMSA Ergebnissen für Hirudin(1–51)

Gruppe	Schranken zu	b_1 [Å]	b_2 [Å]	V^e [Å]	v^e [Å]	V^u [Å]	v^u [Å]	$n_{\Delta d}$	$\overline{\Delta d}$ [Å]	Zuordnung
Tyr3 HB2/3	Y3 HN	2,8	4,1	0,0	0,0	12,9	0,5	-20	-1,1	e
	Y3 HA	3,5	2,8	0,1	0,0	4,8	0,2	20	0,4	e
	T4 HN	4,2	3,4	0,0	0,0	8,2	0,0	20	1,2	e
	L15 QD2	4,0	4,6	0,1	0,0	3,2	0,0	13	-0,1	
Cys6 HB2/3	C6 HN	3,5	4,1	0,0	0,0	0,5	0,0	-20	-1,2	e
	C6 HA	3,1	2,7	0,6	0,0	5,5	0,2	20	0,3	
	T7 HN	5,0	3,7	0,0	0,0	3,7	0,0	20	1,4	e
	L30 QQD	6,0	7,3	0,0	0,0	0,0	0,0	13	0,2	
Gln11 HB2/3	Q11 HN	3,4	2,9	4,9	0,0	0,1	0,0	-20	-0,8	u
	N12 HN	3,2	4,2	16,7	0,5	0,1	0,0	20	1,2	u
Ser19 HB2/3	S19 HN	3,3	2,6	2,7	0,0	1,1	0,0	-11	-0,1	
	N20 HN	5,0	3,8	0,0	0,0	8,3	0,0	19	0,6	
Leu13 QD1/2	L13 HN	4,7	6,0	7,1	0,0	8,0	0,0	12	0,2	
	L13 HA	3,6	5,5	3,1	0,0	6,5	0,0	11	-0,1	
	L13 HB3	4,6	3,6	1,4	0,0	1,4	0,0	-12	-0,1	
	Q24 HN	5,0	6,0	11,4	0,0	7,8	0,0	12	0,1	
Leu15 QD1/2	L15 HA	6,0	3,4	0,0	0,0	23,5	1,1	20	1,8	e
	L15 HB2	3,7	5,0	0,0	0,0	1,7	0,1	-20	-0,8	e
	L15 HG	2,9	3,4	0,0	0,0	0,0	0,0	-14	0,0	
	T4 HA	4,4	6,0	0,0	0,0	25,3	0,8	-20	-2,0	e
	S19 HA	4,0	4,6	0,0	0,0	0,0	0,0	-18	-0,4	
	V21 HA	6,0	3,8	0,1	0,0	42,7	1,7	20	2,8	e

Für die stereospezifische Zuordnung der β -Methylengruppe von Tyr 3 stehen vier Paare von Distanzeinschränkungen zur Verfügung, von denen die ersten drei klar anzeigen, daß die stereospezifische Zuordnung der Eingabedaten korrekt ist, während das letzte kein eindeutiges Ergebnis liefert. Die Distanzschranken zur QD2 Methylgruppe von Leu 15 dürften sowieso nur dann verwendet werden, wenn die stereospezifische Zuordnung der entsprechenden Isopropylgruppe bekannt ist. Die stereospezifische Zuordnung der β -Methylengruppe von Tyr 3 wird durch die Statistik über die Verletzungen der Distanzschranken bestätigt: Für die Distanzeinschränkungen zum Amidproton innerhalb des Rests tritt in jedem der 20 Konformere eine Verletzung von mindestens 0,5 Å auf, wenn die stereospezifische Zuordnung gegenüber den Eingabedaten umgekehrt wird ($v^u = 0.5$ Å). Die β -Methylengruppe von Gln 11 ist ein Fall, wo die umgekehrte stereospezifische Zuordnung richtig ist. Die β -Methylengruppe von Ser 19 und die Isopropylgruppe von Leu 13 sind zwei Beispiele, bei denen die vorhandenen Daten für eine eindeutige stereospezifische Zuordnung nicht ausreichen. Das Distanzschrankenpaar, das die Isopropylgruppe von Leu 15 mit dem α -Proton von Val 21 verbindet, zeigt, daß mit GLOMSA, anders als mit HABAS, auch nicht-lokale Einschränkungen verwendet werden können, um stereospezifische Zuordnungen festzulegen.

Berechnung dreidimensionaler Proteinstrukturen mit dem Programm DIANA

Bei der Entwicklung der Methode zur Bestimmung dreidimensionaler Proteinstrukturen in Lösung aufgrund von Kernspinresonanzmessungen (Wüthrich, 1989a) war es bereits früh klar, daß die für die Strukturrechnung wichtigsten Kernresonanzdaten ein Netz von Distanzeinschränkungen zwischen räumlich benachbarten Wasserstoffatomen bilden würden (Gordon & Wüthrich, 1978; Dubs *et al.*, 1979; Keller & Wüthrich, 1980; Wüthrich *et al.*, 1982). Deshalb konnten die vorhandenen Verfahren zur Strukturbestimmung aufgrund von Röntgenstreuungsexperimenten an Einkristallen mit Kernresonanzdaten nicht gebraucht werden, und es mußten neue Methoden für die Strukturrechnung entwickelt werden. Zuerst waren dies Distanzgeometriealgorithmen auf der Grundlage der metrischen Matrix (Blumenthal, 1970), die bereits für systematische Untersuchungen von Proteinstrukturen eingesetzt worden waren (Crippen, 1977; Kuntz *et al.*, 1976; Havel *et al.*, 1983), in Kombination mit Einschränkungen zur Beschreibung der kovalenten Struktur, der sterischen Bedingungen und der Kernresonanzdaten (Braun *et al.*, 1981, 1983; Havel & Wüthrich, 1984, 1985). Weitere Arbeiten umfaßten die Entwicklung der Methode der variablen Zielfunktion (Braun & Gö, 1985), interaktiven Modellbau mit Hilfe von Grafikprogrammen (Billeter *et al.*, 1985), eingeschränkte Moleküldynamikrechnungen¹ — entweder direkt (Brünger *et al.*, 1986) oder anschließend an interaktiven Modellbau (Kaptein *et al.*, 1985) oder Distanzgeometrierechnungen (Clore *et al.*, 1985) gebraucht — und einen Ellipsoid-Algorithmus (Billeter *et al.*, 1987). Bis heute durchgesetzt haben sich vor allem zwei Verfahren: Einbettung der Proteinstruktur im dreidimensionalen Raum ausgehend von der metrischen Matrix mittels eines Distanzgeometrieprogramms, zum Beispiel DISGEO (Havel & Wüthrich, 1984) oder DG-II (Havel, 1991), gefolgt von eingeschränkten Moleküldynamikrechnungen (Driscoll *et al.*, 1989; Lee *et al.*, 1989), oder der Algorithmus der variablen Zielfunktion, verwirklicht in den Programmen DISMAN (Braun & Gö, 1985), DADAS (Kohda

¹ D. h. Moleküldynamikrechnungen, in denen die »physikalische« Energiefunktion um Terme zur Beschreibung der experimentellen Einschränkungen erweitert wurde.

et al., 1988) und DIANA (Güntert *et al.*, 1991a), der direkt — ohne Moleküldynamikverfeinerung — akzeptable Strukturen liefert (Kline *et al.*, 1988; Schultze *et al.*, 1988; Zuiderweg *et al.*, 1989) oder durch eingeschränkte Energieminimierung ergänzt werden kann (Billeter *et al.*, 1990a, b; Qian *et al.*, 1989; Widmer *et al.*, 1989; vgl. S. 145ff.).

Nachdem klar war, daß Kernspinresonanzexperimente genügend Daten liefern können, um die dreidimensionale Struktur globulärer Proteine mit atomarer Auflösung zu bestimmen (Havel & Wüthrich, 1984, 1985; Williamson *et al.*, 1985), richtete sich das Interesse auf die Entwicklung von Methoden zur Strukturrechnung, die effizient sind und im erlaubten Gebiet des Konformationsraums möglichst gleichmäßig verteilte Strukturen erzeugen. Dieses Kapitel stellt eine Neuimplementierung des Algorithmus' der variablen Zielfunktion (Braun & Gö, 1985) im Programm DIANA (Güntert *et al.*, 1991a) vor, das effizient im Verbrauch von Rechenzeit und — zusammen mit den Unterstützungsprogrammen CALIBA (Güntert *et al.*, 1991a; vgl. S. 61ff.), HABAS (Güntert *et al.*, 1989; vgl. S. 94ff.) und GLOMSA (Güntert *et al.*, 1991a; vgl. S. 100ff.) — besonders benutzerfreundlich ist (Güntert *et al.*, 1991b).¹

Die Strukturrechnung mit dem Programm DIANA basiert auf der Minimierung einer *variablen Zielfunktion* $T(\phi_1, \dots, \phi_n)$, wobei die n Freiheitsgrade die Diederwinkel ϕ_1, \dots, ϕ_n um drehbare Bindungen der Polypeptidkette sind. Bindungslängen, Bindungswinkel und Chiralitäten sind während der gesamten Rechnung fest und entsprechen der ECEPP Standardgeometrie (Momany *et al.*, 1975; Némethy *et al.*, 1983). Die nicht-negative Zielfunktion ist so definiert (vgl. S. 106), daß sie verschwindet, wenn alle experimentellen Distanz- und Winkeleinschränkungen erfüllt sind und keine sterischen Überlappungen auftreten, und daß $T(\phi_1, \dots, \phi_n) < T(\theta_1, \dots, \theta_n)$ gilt, wenn die Konformation (ϕ_1, \dots, ϕ_n) die Einschränkungen besser erfüllt als $(\theta_1, \dots, \theta_n)$. Aufgabe des Programms ist es, Konformationen (ϕ_1, \dots, ϕ_n) zu finden, die kleine Zielfunktionswerte aufweisen. Um die Gefahr zu verringern, in einem lokalen Minimum mit einem Zielfunktionswert weit über dem des globalen Minimums

¹ In unserer Forschungsgruppe wurden bislang die dreidimensionalen Lösungsstrukturen folgender Proteine mit dem Programm DIANA berechnet (in alphabetischer Reihenfolge): 434-Repressor(1–63) (Schott, 1992); 434-Repressor(1–69) (Neri *et al.*, 1992); α -Neurotoxin der Schwarzen Mamba (Brown & Wüthrich, 1992); Aktivierungsdomäne von Procarboxypeptidase B (Vendrell *et al.*, 1991); *Antennapedia(C39S)* Homeodomäne (Güntert *et al.*, 1991b; vgl. S. 144ff.); *Antennapedia(C39S)* Homeodomäne im Komplex mit einem 14 Basenpaaren großen DNS-Stück (Qian *et al.*, 1993); BPTI (Berndt *et al.*, 1992); Glutaredoxin in oxidiertes (Xia *et al.*, 1992) und reduziertes Form (Sodano *et al.*, 1991); Hirudin(1–51) (Szyperski *et al.*, 1992b; vgl. S. 149ff.); Kardiotoxin CTX IIb (O'Connell *et al.*, 1993); Kunitz-Typ Proteinase Inhibitor einer Seeanemone (Antuch *et al.*, 1993); P22 c2 Repressor(1–76) (G. Otting, P. Sevilla & K. Wüthrich, unveröffentlicht); Pheromon Er-10 (Brown *et al.*, 1993); Toxin K der Schwarzen Mamba (Berndt *et al.*, 1993).

hängenzubleiben, wird die Form der Zielfunktion während der Rechnung verändert. Am Anfang berücksichtigt die Zielfunktion nur Einschränkungen, die bezüglich der Polypeptidsequenz lokal sind, später werden schrittweise Einschränkungen zwischen Atomen, die entlang der Sequenz weiter voneinander entfernt sind, miteinbezogen (Braun & Gö, 1985). Folglich wird in einer Strukturrechnung mit DIANA am Anfang zuerst die lokale Struktur aufgebaut, während sich die globale Faltung der Polypeptidkette erst zum Ende hin ausbildet.

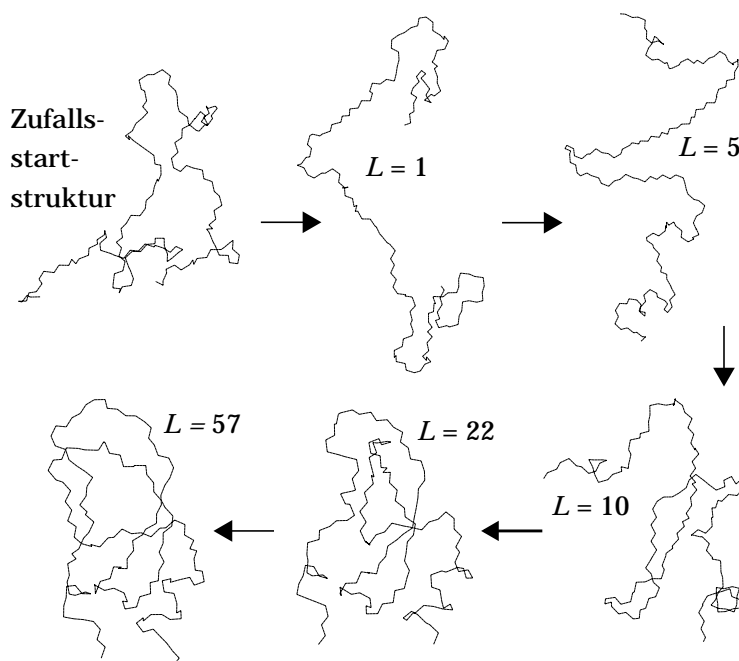


Bild 14. Verlauf einer Proteinstrukturrechnung mit dem Programm DIANA, illustriert anhand einer Startstruktur des Toxins K der Schwarzen Mamba (Wüthrich et al., 1993) mit zufälligen Diederwinkelwerten, den zugehörigen Strukturen bei den Minimierungsweiten (vgl. S. 107) $L = 1, 5, 10, 22$ und der entsprechenden konvergierten Struktur am Ende der Minimierung ($L = 57$). Die Atome N, C^α und C' sind entlang des Rückgrats miteinander verbunden.

Die variable Zielfunktion

Die Zielfunktion berücksichtigt zwei grundsätzlich verschiedene Arten geometrischer Einschränkungen (Wüthrich, 1986; Braun, 1987): einerseits obere oder untere Schranken $c_{\alpha\beta}$ für die Distanz zwischen zwei Atomen α und β und andererseits Einschränkungen für einzelne Diederwinkel a in der Form erlaubter Intervalle $[\phi_a^{\min}, \phi_a^{\max}]$ mit $\phi_a^{\min} < \phi_a^{\max} < \phi_a^{\min} + 2\pi$. Es bezeichne

$$\Gamma_a = \pi - \frac{\phi_a^{\max} - \phi_a^{\min}}{2} \quad [46]$$

die halbe Breite des *verbotenen* Intervalls von Diederwinkelwerten ϕ_a und

$$\Delta_a(\phi_a) = \begin{cases} 0, & \text{falls } \phi_a \in [\phi_a^{\min}, \phi_a^{\max}]; \\ -\Delta_a^{\min}, & \text{falls } \phi_a \notin [\phi_a^{\min}, \phi_a^{\max}] \text{ und } \Delta_a^{\min} \leq \Delta_a^{\max}; \\ \Delta_a^{\max}, & \text{falls } \phi_a \notin [\phi_a^{\min}, \phi_a^{\max}] \text{ und } \Delta_a^{\min} > \Delta_a^{\max}; \end{cases} \quad [47]$$

mit

$$\begin{aligned} \Delta_a^{\min} &= \min \left\{ \left| \hat{\phi}_a^{\min} - \hat{\phi}_a \right|, 2\pi - \left| \hat{\phi}_a^{\min} - \hat{\phi}_a \right| \right\} \\ \Delta_a^{\max} &= \min \left\{ \left| \hat{\phi}_a^{\max} - \hat{\phi}_a \right|, 2\pi - \left| \hat{\phi}_a^{\max} - \hat{\phi}_a \right| \right\} \end{aligned} \quad [48]$$

die (vorzeichenbehaftete) Größe der Diederwinkelverletzung. $\hat{\phi}$ steht für den äquivalenten Wert von ϕ im Intervall $[0, 2\pi[$, der immer durch Addition eines geeigneten ganzzahligen Vielfachen von 2π zu ϕ erhalten werden kann. Das Vorzeichen von Δ ist nur für die Berechnung des Gradienten der Zielfunktion wichtig; es ist positiv, wenn eine kleine Zunahme von ϕ_a auch die Größe der Verletzung vergrößert, sonst negativ. Weiter seien I_o , I_u und I_v die Mengen von Atompaaeren $\alpha\beta$, für die eine obere, untere oder van der Waals (vgl. S. 112) Distanzschranke $o_{\alpha\beta}$, $u_{\alpha\beta}$ oder $v_{\alpha\beta}$ existiert, $I_c \subseteq I_c$ ($c = o, u, v$) Teilmengen davon, I_w die Menge der eingeschränkten Diederwinkel a , w_o , w_u , w_v und w_w Gewichtungsfaktoren für obere, untere und van der Waals Distanzschranken sowie für Winkeleinschränkungen, $d_{\alpha\beta} = d_{\alpha\beta}(\phi_1, \dots, \phi_n)$ die Distanz zwischen den beiden Atomen α und β , $\Theta_o(t) = \max(0, t)$ und $\Theta_u(t) = \Theta_v(t) = \min(0, t)$. Dann definiert

$$T = \sum_{c=o, u, v} w_c \sum_{\alpha\beta \in I_c} \left(\frac{\Theta_c(d_{\alpha\beta}^2 - c_{\alpha\beta}^2)}{2c_{\alpha\beta}} \right)^2 + w_w \sum_{a \in I_w} \left(1 - \frac{1}{2} \left(\frac{\Delta_a}{\Gamma_a} \right)^2 \right) \Delta_a^2 \quad [49]$$

die Zielfunktion (Güntert *et al.*, 1991a). Die Teilmengen $I_c \subseteq I_c$ ($c = o, u, v$) für die im Programm DIANA die Zielfunktion berechnet werden kann, bestehen aus allen Einschränkungen zwischen Atomen α und β in Aminosäureresten, deren Sequenznummern R_α und R_β sich um nicht mehr als eine gegebene *Minimierungsweite* L_c unterscheiden:

$$I_c = \{ \alpha\beta \in I_c \mid |R_\alpha - R_\beta| \leq L_c \} \quad (c = o, u, v). \quad [50]$$

Normalerweise wählt man für die drei Typen von Distanzeinschränkungen die gleiche Minimierungsweite $L = L_o = L_u = L_v$.

Im allgemeinen umfaßt eine Strukturrechnung mit dem Programm DIANA verschiedene *Minimierungsschritte* (nicht zu verwechseln mit einzelnen Iterationen des Minimierungsalgorithmus), d. h. die Minimierung verschiedener Formen der Zielfunktion mit ansteigenden Minimierungsweiten L_c oder anderen Gewichtungsfaktoren w_c . Das Programm bietet dafür »Standardminimie-

rungsparameter« (vgl. S. 124) an, die für Routineanwendungen meistens genügen. Eine optimale Strategie ist nicht bekannt, aber die Erfahrung hat gezeigt, daß es wichtig ist, die Minimierungsweite in kleinen Schritten, ausgehend von $L_c = 0$ oder 1, zu erhöhen und den van der Waals Einschränkungen im Vergleich zu den experimentellen Distanzschranken ein kleines Gewicht von typischerweise $w_v = 0.2 w_o$ zu geben, außer am Ende der Rechnung, wo sterische Überlappungen mit Hilfe höherer Gewichte, üblicherweise $w_v = 2 w_o$, minimiert werden.

Standardminimierungsparameter für eine BPTI-Strukturrechnung (vgl. S. 134ff)
($w_o = w_u = 1$, $w_w = 5 \text{Å}^2$; $N = \text{max. Anzahl Iterationen}$; »H« gibt an, daß Wasserstoffatome in der van der Waals Abstoßung berücksichtigt werden)

Schritt	L	w_v	N	Schritt	L	w_v	N
1	0	0,2	150	16	21	0,2	150
2	1	0,2	150	17	23	0,2	150
3	2	0,2	150	18	25	0,2	150
4	3	0,2	150	19	27	0,2	150
5	4	0,2	150	20	30	0,2	150
6	5	0,2	150	21	32	0,2	150
7	6	0,2	150	22	36	0,2	150
8	7	0,2	150	23	38	0,2	150
9	8	0,2	150	24	41	0,2	150
10	9	0,2	150	25	50	0,2	150
11	11	0,2	150	26	56	0,2	150
12	13	0,2	150	27	58	0,2 H	400
13	15	0,2	150	28	58	0,6 H	400
14	17	0,2	150	29	58	2,0 H	400
15	19	0,2	150				

Die Zielfunktion [49] ist über den ganzen Konformationsraum stetig differenzierbar und so gewählt, daß für alle Typen von Einschränkungen der Beitrag einer kleinen Verletzung δ_c zur Zielfunktion $w_c \delta_c^2$ beträgt ($c = o, u, v, w$). Weil nur das Quadrat der interatomaren Distanzen in die Zielfunktion eingeht, kommt ihre Auswertung ohne die Berechnung von Quadratwurzeln aus und ist daher schnell.

Eingabe- und Ausgabedaten des Programms DIANA

Das Programm DIANA liest verschiedene Eingabedateien sowie interaktiv eingegebene Größen. Nomenklatur, Standardgeometrie, Diederwinkeldeklarationen und Atomradien für die Aminosäurereste und Nukleotide werden der *Bausteinbibliothek* (vgl. S. 66) entnommen. Die *Sequenzdatei* (vgl. S. 73) enthält die Primärstruktur sowie Angaben über fixierte Diederwinkel (z. B. *cis*-Peptidbindungen) und spezielle, nicht mit der Baumstruktur der Diederwinkel verträgliche kovalente Bindungen. Paare diastereotoper Substituenten mit stereospezifischer Zuordnung identifiziert die *Datei mit stereospezifischen Zuordnungen* (vgl. S. 76). Wenn diese Datei fehlt, nimmt das Programm — je nach Wunsch des Benutzers — stereospezifische Zuordnungen für alle oder keine diastereotopen Gruppen an. Obere und untere Distanzschranken sowie Diederwinkeleinschränkungen werden den entsprechenden Dateien entnommen (vgl. S. 77/78). Parameter für die einzelnen Minimierungsschritte können aus einer *Minimierungsparameterdatei* (vgl. S. 120) eingelesen werden, obwohl für Routineanwendungen die vom Programm implizit gewählten Standardminimierungsparameter (vgl. S. 124) fast immer genügen. Startkonformationen können in *Winkel-* (vgl. S. 75) oder *Koordinatendateien* (vgl. S. 74) angegeben oder vom Programm intern mit zufälligen Diederwinkelwerten erzeugt werden.

Die *Ergebnisdatei* gibt die interaktiven Eingaben wieder, protokolliert den Verlauf der Rechnung und führt auf Wunsch Verletzungen von Einschränkungen, die gegebene Schwellen überschreiten, auf. Die *Übersichtsdatei* enthält für alle Typen von Einschränkungen eine vollständige Liste mit Anzahl, Summe und Maximum der Verletzungen in den fertiggestellten Konformeren sowie eine Statistik über die Verletzungen, Wasserstoffbrücken und Abweichungen der Atompositionen in allen Konformeren, deren Zielfunktionsendwert unter einem vorgegebenen Wert liegt. Während der Berechnung eines Bündels von Konformeren wird die Übersichtsdatei nach jedem fertiggestellten Konformer auf den neusten Stand gebracht, damit der Benutzer den Ablauf der Strukturrechnung verfolgen kann. *Winkel-* oder *Koordinatendateien* der berechneten Konformere werden, ebenso wie Dateien mit redundanten Diederwinkeleinschränkungen zur Konvergenzbeschleunigung (vgl. S. 132ff.), vom Programm geschrieben.

Irrelevante und zu restriktive Einschränkungen

Das Programm DIANA eliminiert irrelevante und warnt vor zu restriktiven Distanzeinschränkungen (Güntert *et al.*, 1991a). Eine Distanzschranke ist

irrelevant, wenn die betreffende Distanz konformationsunabhängig ist oder keine Konformation existiert, die die Schranke verletzen würde. Eine untere Distanzschranke ist auch irrelevant, wenn sie kleiner als die intern erzeugte van der Waals Distanzschranke (vgl. S. 112ff.) ist. Es ist nicht immer einfach festzustellen, ob zu einer gegebenen Distanzschranke eine Konformation auftreten kann, die sie verletzt; das Programm DIANA prüft diese Bedingung nur für Distanzen, welche von einem oder zwei Diederwinkel abhängen. Der Wertebereich der Distanz $d_{\alpha\beta} = |\vec{r}_\alpha - \vec{r}_\beta|$ zwischen zwei Atomen α und β , die nur von einem Diederwinkel a abhängt, ist gegeben durch (Billeter *et al.*, 1982)

$$A - B \leq d_{\alpha\beta}^2 \leq A + B \quad [51]$$

mit

$$\begin{aligned} A &= \vec{d}_\alpha^2 + \vec{d}_\beta^2 - 2(\vec{e}_a \cdot \vec{d}_\alpha)(\vec{e}_a \cdot \vec{d}_\beta), \\ B &= 2\sqrt{[\vec{d}_\alpha^2 - (\vec{e}_a \cdot \vec{d}_\alpha)^2][\vec{d}_\beta^2 - (\vec{e}_a \cdot \vec{d}_\beta)^2]}, \\ \vec{d}_\alpha &= \vec{r}_\alpha - \vec{r}_a \quad \text{und} \quad \vec{d}_\beta = \vec{r}_\beta - \vec{r}_a. \end{aligned} \quad [52]$$

\vec{r}_α und \vec{r}_β bezeichnen die Ortsvektoren der Atome α und β , \vec{r}_a den Ortsvektor des Anfangspunktes und \vec{e}_a den Einheitsvektor entlang der drehbaren Bindung a für eine beliebige Konformation. Eine obere Schranke $o_{\alpha\beta}$ ist irrelevant, wenn $o_{\alpha\beta} \geq \sqrt{A + B}$ und zu restriktiv, wenn $o_{\alpha\beta} < \sqrt{A - B}$ ist. Analoges gilt für untere Distanzschranken und Schranken für Distanzen die von zwei Diederwinkeln abhängen.

Diastereotope Gruppen ohne stereospezifische Zuordnung

Weil die sequentielle Resonanzzuordnung in Proteinen (Wüthrich, 1986) diastereotope Substituenten nicht stereospezifisch zuordnen kann und auch andere Verfahren keine vollständigen stereospezifischen Zuordnungen liefern (vgl. S. 84ff.), müssen Strukturrechnungsprogramme Einschränkungen zu diastereotopen Gruppen ohne stereospezifische Zuordnung speziell behandeln. Die einfachste Behandlung führt Pseudoatome β_P im Zentrum der beiden diastereotopen Substituenten β_1 und β_2 ein und bezieht obere Distanzschranken $o_{\alpha\beta_1}$ und $o_{\alpha\beta_2}$ zwischen ihnen und einem Atom α auf das Pseudoatom β_P , indem eine dem ungünstigstem Fall entsprechende Korrektur addiert wird (Wüthrich *et al.*, 1983)

$$o_{\alpha\beta_P} = \min(o_{\alpha\beta_1}, o_{\alpha\beta_2}) + d_{\beta_1\beta_P}. \quad [53]$$

Das Programm DIANA ersetzt diese Behandlung durch eine Kombination von zwei Methoden und begrenzt dadurch den Informationsverlust verglichen mit Gl. [53] wesentlich (Güntert *et al.*, 1991a).

Die erste Methode verwendet ebenfalls Pseudoatome, jedoch mit optimalen Korrekturen, die berücksichtigen, daß oft Distanzschranken $o_{\alpha\beta_1}$ und $o_{\alpha\beta_2}$ von einem Atom α zu *beiden* diastereotopen Substituenten β_1 und β_2 vorhanden sind. An die Stelle von Gl. [53] tritt

$$o_{\alpha\beta_P} = \sqrt{\frac{o_{\alpha\beta_1}^2 + o_{\alpha\beta_2}^2}{2}} - d_{\beta_1\beta_P}^2. \quad [54]$$

Falls nur eine der beiden Distanzschranken, z. B. $o_{\alpha\beta_1}$, vorhanden ist oder die beiden Distanzschranken sich um mehr als den Abstand zwischen den beiden diastereotopen Substituenten unterscheiden, bringt Gl. [54] keinen Fortschritt, und es wird das ursprüngliche Pseudoatomkonzept (Wüthrich *et al.*, 1983) verwendet:

$$o_{\alpha\beta_P} = o_{\alpha\beta_1} + d_{\beta_1\beta_P}. \quad [55]$$

Analog zu Gl. [54] berechnet sich die Distanzschranke zwischen den Pseudoatomen β_P und γ_P zweier Paare diastereotoper Substituenten β_1/β_2 und γ_1/γ_2 gemäß

$$o_{\beta_P\gamma_P} = \sqrt{\frac{o_{\beta_1\gamma_1}^2 + o_{\beta_1\gamma_2}^2 + o_{\beta_2\gamma_1}^2 + o_{\beta_2\gamma_2}^2}{4}} - d_{\beta_1\beta_P}^2 - d_{\gamma_1\gamma_P}^2. \quad [56]$$

Wenn eine oder mehrere der vier Distanzschranken fehlen, erzeugt das Programm mit Hilfe der Dreiecksungleichung redundante Schranken für die entsprechenden Distanzen, damit Gl. [56] trotzdem angewandt werden kann.

Die zweite Methode kommt ohne Pseudoatom aus und legt die gleichen Distanzschranken

$$o'_{\alpha\beta_1} = o'_{\alpha\beta_2} = \min\{\max(o_{\alpha\beta_1}, o_{\alpha\beta_2}), \min(o_{\alpha\beta_1}, o_{\alpha\beta_2}) + d_{\beta_1\beta_2}\} \quad [57]$$

an beiden diastereotopen Substituenten an.

Die beiden Methoden von Gl. [54–56] und [57] sind im allgemeinen nicht äquivalent, und das Programm verwendet, wenn es sinnvoll ist, beide Methoden zugleich. Wenn jedoch nur eine der beiden Distanzschranken $o_{\alpha\beta_1}$ und $o_{\alpha\beta_2}$ vorhanden ist, liefert Gl. [57] keine zusätzliche Information, die die Distanzschranke zum Pseudoatom gemäß Gl. [55] ergänzen würde, und wird deshalb nicht verwendet. Andererseits wird bei fast gleichen Distanzschranken $o_{\alpha\beta_1} \approx o_{\alpha\beta_2}$ keine Distanzschranke zum Pseudoatom erzeugt. Die folgende Tabelle gibt einige Beispiele für die Behandlung diastereotoper Gruppen ohne

stereospezifische Zuordnung im Programm DIANA und mit dem ursprüngliche Pseudoatomkonzept (Wüthrich *et al.*, 1983) und zeigt, daß die geschilderte Methode mit einem bedeutend geringeren Informationsverlust auskommt als traditionelle Pseudoatomkorrekturen (Wüthrich *et al.*, 1983), was sich in der Praxis in besser definierten Strukturen äußert, die allerdings noch nicht die Genauigkeit erreichen, die mit stereospezifischen Zuordnungen möglich ist (vgl. Testrechnungen S. 88ff.):

Beispiele für die Behandlung von oberen Distanzschranken [\AA] zu Methylenprotonen ($d_{\beta_1\beta_0} = 0.88 \text{ \AA}$) ohne stereospezifische Zuordnung

Eingabe	DIANA Behandlung gemäß Gl. [54–57]	nach Wüthrich <i>et al.</i> (1983)
$o_{\alpha\beta_1} = 3.7, o_{\alpha\beta_2} = 3.1$	$o_{\alpha\beta_1} = o_{\alpha\beta_2} = 3.7$ und $o_{\alpha\beta_P} = 3.3$	$o_{\alpha\beta_P} = 5.9$
$o_{\alpha\beta_1} = 4.9, o_{\alpha\beta_2} = 5.0$	$o_{\alpha\beta_1} = o_{\alpha\beta_2} = 5.0$	$o_{\alpha\beta_P} = 4.1$
$o_{\alpha\beta_1} = 3.5$	$o_{\alpha\beta_P} = 4.4$	$o_{\alpha\beta_P} = 4.5$
$o_{\beta_1\gamma_1} = 4.7, o_{\beta_1\gamma_2} = 4.6$	$o_{\beta_1\gamma_j} = 6.5$ und $o_{\beta_P\gamma_P} = 5.5$	$o_{\beta_P\gamma_P} = 6.6$

Im ersten Beispiel in der Tabelle werden beide Methoden zugleich gebraucht; im zweiten Beispiel wird keine Distanzschranke zum Pseudoatom erzeugt, da $o_{\alpha\beta_1} \approx o_{\alpha\beta_2}$ ist; das dritte bzw. vierte Beispiel zeigt Anwendungen von Gl. [55] bzw. [56].

Untere Distanzschranken zu diastereotopen Gruppen ohne stereospezifische Zuordnung können analog behandelt werden.

Van der Waals Abstoßung der Atome

Energieminimierungs- und Moleküldynamikprogramme behandeln die nicht-bindende Wechselwirkung zwischen Atomen im allgemeinen durch ein Lennard-Jones Potential (Momany *et al.*, 1975; Brooks *et al.*, 1983; van Gunsteren *et al.*, 1983; Weiner *et al.*, 1986; Wako & Gō, 1987; Schaumann *et al.*, 1990). Distanzgeometrierechnungen behalten nur den wichtigsten Teil dieser Wechselwirkung — die sterische Abstoßung — bei (Havel & Wüthrich, 1984; Braun & Gō, 1985). Das Programm DIANA (Güntert *et al.*, 1991a) modelliert die van der Waals Abstoßung zweier Atome durch eine untere Schranke für die Distanz zwischen den beiden Atomen, die sich aus der Summe ihrer abstoßenden Radien ergeben, welche in der Bausteinbibliothek aufgeführt sind (vgl. S. 68). So-

lange die Distanz zwischen zwei Atomen größer als die Summe ihrer abstoßenden Radien ist, verschwindet für dieses Atompaar der entsprechende Term der Zielfunktion bei, d. h. die allermeisten Atompaare liefern keinen Beitrag zur Zielfunktion.

Ein einfacher Algorithmus zur Berechnung der van der Waals Abstoßung berechnet mit jeder Auswertung der Zielfunktion alle interatomaren Distanzen und wäre sehr ineffizient. Deshalb nimmt DIANA alle Atompaare mit genügend kleinem interatomarem Abstand in eine Liste möglicher nicht-bindender Wechselwirkungen auf (Verlet, 1967; Allen & Tildesley, 1987), die nur nach einer merklichen Konformationsänderung oder einer gewissen Anzahl von Iterationen in der Minimierung jeweils mit Hilfe eines schnellen Algorithmus', der auf die Berechnung aller interatomarer Distanzen verzichtet (Hockney & Eastwood, 1981; Braun & Gö, 1985), auf den neusten Stand gebracht wird. In einem Proteinmolekül mit $n \approx 1000$ Atomen wird die Liste möglicher nicht-bindender Wechselwirkungen normalerweise weniger als 30'000 von allen $n(n-1)/2 \approx 500'000$ Atompaaren beinhalten, von denen wiederum in einer »guten« Struktur nur einige hundert einen nicht-verschwindenden Beitrag zur Zielfunktion liefern.

Um die Berechnung der van der Waals Abstoßung weiter zu beschleunigen, teilt das Programm DIANA die Liste möglicher nicht-bindender Wechselwirkungen in zwei Teile auf: einen konstanten Teil, der am Anfang einer Strukturrechnung aufgebaut und anschließend nicht mehr verändert wird, bestehend aus allen van der Waals Distanzschranken, die Atome innerhalb eines Rests oder in sequentiell benachbarten Resten betreffen, und einen veränderlichen, konformationsabhängigen Teil, der die restlichen Atompaare betrifft, und in Abständen aktualisiert wird. Aus dem festen Teil der Liste bleiben irrelevante Distanzschranken (vgl. S. 109), insbesondere solche die feste Distanzen wie zum Beispiel Bindungslängen betreffen, ausgeschlossen. Der veränderliche Teil der Liste entsteht, indem das Proteinmolekül in ein kubisches Gitter gesetzt wird, dessen Gitterkonstante g mindestens doppelt so groß wie der größte Atomradius ist (für die Standardbausteinbibliothek, vgl. S. 68, ergibt sich $g = 3.2 \text{ \AA}$), und nur Paare von Atomen in der gleichen oder in benachbarten Gitterzellen, die noch nicht für den festen Teil der Liste berücksichtigt wurden, aufgenommen werden (Hockney & Eastwood, 1981; Braun & Gö, 1985). Damit ist sichergestellt, daß die Liste alle Atompaare enthält, deren van der Waals Abstoßung in der aktuellen Konformation nicht verschwindet. Da kleine Konformationsänderungen die Liste nur wenig ändern, wird sie nur dann auf den neusten Stand gebracht, wenn seit der letzten Aktualisierung ein Diederwinkel um mehr als einen gegebenen Wert (standardmäßig 10°) geändert hat oder mehr als eine gewisse Anzahl Iterationen (normalerweise 100) ausgeführt wurden.

Für die Behandlung von Wasserstoffbrücken, Disulfidbrücken und anderen

speziellen kovalenten Bindungen müssen besondere Vorkehrungen getroffen werden, weil diese Bindungen mit der Baumstruktur der drehbaren Bindungen (Abe *et al.*, 1984), die keine geschlossenen flexiblen Ringe erlaubt, nicht verträglich sind. Daher werden die van der Waals Distanzschranken für solche Bindungen gegenüber der Summe der abstoßenden Radien genügend verringert, um die Bindung zu ermöglichen. Die Bindungslängen und -winkel werden durch explizite Distanzeinschränkungen festgelegt (Williamson *et al.*, 1985). Auf diese Weise lassen sich zum Beispiel auch flexible Prolinringe einführen (In der ECEPP Standardgeometrie (Momany *et al.*, 1975; Némethy *et al.*, 1983) sind Prolinringe normalerweise starr).

Berechnung der Atomkoordinaten aus Diederwinkelwerten

Weil das Programm DIANA im Diederwinkelraum arbeitet, d. h. die Freiheitsgrade Diederwinkel sind, Distanzeinschränkungen aber über die kartesischen Koordinaten der Atome definiert sind, müssen zum Auswerten der Zielfunktion nach jeder Konformationsänderung — d. h. Änderung von Diederwinkelwerten — die Positionen der Atome neu berechnet werden. Zu diesem Zweck wird das Molekül als Baum, bestehend aus starren Einheiten, die durch drehbare Bindungen verbunden sind, aufgefaßt (Abe *et al.*, 1983). Starre Einheiten enthalten keine drehbaren Bindungen und bestehen aus einem oder mehreren Atomen, deren Positionen relativ zueinander unveränderlich sind.

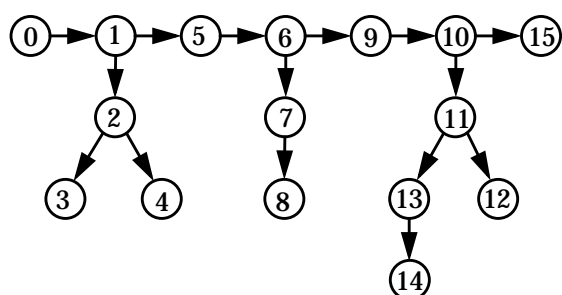


Bild 15. Die vom Programm DIANA verwendete Baumstruktur im Fall des Tripeptids Val-Ser-Ile. Nummerierte Kreise stehen für starre Einheiten, drehbare Bindungen sind durch Pfeile angegeben, die zu dem Teil des Baums zeigen, der bei einer Drehung um die Bindung bewegt wird (Abe *et al.*, 1983).

Die Diederwinkel und starren Einheiten sind ansteigend nummeriert, wenn der Baum in Richtung der Pfeile durchlaufen wird (vgl. Bild 15), die zu dem Teil des Baums zeigen, der bei einer Drehung um die Bindung bewegt wird. Eine starre Einheit und die zu ihr führende drehbare Bindung besitzen jeweils die gleiche Nummer a und eindeutig bestimmte Vorgänger mit der Nummer

$p(a) < a$. Eine starre Einheit a enthält alle Atome, deren Position nur vom Diederwinkel a abhängt, wenn alle Diederwinkel $a' < a$ festgehalten werden. So enthalten zum Beispiel die starren Einheiten 1, 6 und 9 in Bild 15 jeweils die Atome HA, C und CB, die Einheiten 5 und 9 jeweils die Atome O, N, HN und CA (die Peptidbindung ist nicht drehbar; für die Nomenklatur vgl. S. 66ff.).

Beim Einlesen der Bausteinbibliothek (vgl. S. 66ff.) werden die einzelnen Reste (als Ganzes) mit Hilfe der Transformation

$$\dot{r}'_{\alpha} = R(\dot{r}_3 - \dot{r}_1, \dot{r}_2 - \dot{r}_1)(\dot{r}_{\alpha} - \dot{r}_1) \quad [58]$$

in eine Standardorientierung gebracht. \dot{r}_{α} bezeichnet den Ortsvektor des Atoms $\alpha = 1, 2, 3, \dots$ in der Bausteinbibliothek, \dot{r}'_{α} den entsprechenden Ortsvektor in der Standardorientierung des Rests und $R(\dot{u}, \dot{v})$ die Rotationsmatrix, deren Zeilen durch die drei orthonormalen Vektoren

$$\dot{\epsilon}_1 = \frac{\dot{u}}{|\dot{u}|}, \quad \dot{\epsilon}_2 = \frac{\dot{u} \wedge \dot{v}}{|\dot{u} \wedge \dot{v}|} \quad \text{und} \quad \dot{\epsilon}_3 = \dot{\epsilon}_1 \wedge \dot{\epsilon}_2 \quad [59]$$

gegeben sind.

Um die Koordinaten der Atome in der Konformation zu berechnen, die durch die Diederwinkelwerte (ϕ_1, \dots, ϕ_n) bestimmt ist, kann jeweils ein neuer Aminosäurerest an den bereits vorhandenen Teil der Polypeptidkette angehängt werden. Anschließend — bevor der nächste Rest angehängt wird — müssen die gewünschten Diederwinkelwerte eingestellt werden. Dem Andocken eines Rests an den Vorangehenden dienen die Überlappatome in der Bausteinbibliothek: die ersten und letzten drei Atome jedes Bibliothekseintrags (für Aminosäurereste in der Standardbibliothek die Atome C, O und N; vgl. S. 68ff.). Die Atomkoordinaten \dot{r}''_{α} des angedockten Bausteins ergeben sich durch die Transformation

$$\dot{r}''_{\alpha} = \dot{p}_1 + R(\dot{p}_3 - \dot{p}_1, \dot{p}_2 - \dot{p}_1)^{-1}(\dot{r}'_{\alpha} - \dot{r}'_1), \quad [60]$$

wobei \dot{p}_1 , \dot{p}_2 und \dot{p}_3 die Positionen der hinteren Überlappatome des vorangehenden Rests und \dot{r}'_{α} die Koordinaten des anzudockenden Rests in Standardorientierung gemäß Gl. [58] bezeichnen. Die Transformation

$$\dot{r}'''_{\alpha} = \dot{r}_a + R(\dot{e}_a, \phi_a - \phi_a^0)(\dot{r}''_{\alpha} - \dot{r}_a) \quad (\alpha \in M_a) \quad [61]$$

stellt die gewünschten Diederwinkelwerte ein. In Gl. [61] bezeichnet \dot{r}_a den Ortsvektor des Anfangspunktes, \dot{e}_a einen Einheitsvektor entlang der drehbaren Bindung a , $R(\dot{e}, \phi)$ die Rotation¹ mit der Achse \dot{e} und dem Drehwinkel ϕ , ϕ_a den gewünschten Diederwinkelwert, ϕ_a^0 den Diederwinkelwert der in der Bi-

¹ Für einen beliebigen Vektor \dot{x} gilt: $R(\dot{e}, \phi)\dot{x} = (\dot{e} \cdot \dot{x})\dot{e} + (\dot{e} \wedge \dot{x})\sin\phi + (\dot{x} - (\dot{e} \cdot \dot{x})\dot{e})\cos\phi$.

bibliothek gespeicherten Konformation und M_a die Menge der Atome die bei einer Drehung um die Bindung a bewegt werden. Gl. [61] muß für alle m drehbaren Bindungen des angedockten Rests nacheinander in absteigender Reihenfolge $m, m-1, \dots, 1$ angewandt werden.

Die Berechnung der Atomkoordinaten nach Gl. [60/61] ist ineffizient, weil ein Atom im allgemeinen mehreren Koordinatentransformationen unterworfen wird. Das Programm DIANA verwendet Gl. [60/61] daher nur, um am Anfang, nach dem Lesen der Aminosäuresequenz, die Koordinaten der Konformation mit $\phi_1 = \dots = \phi_n = 0$ zu berechnen, die als Ausgangspunkt für den während der Minimierung verwendeten, schnellen Algorithmus dient, der zuerst das »Gerüst« der drehbaren Bindungen aufbaut und anschließend die Atome mit einer einzigen Koordinatentransformation positioniert. Dazu werden, ausgehend von der Konformation $\phi_1 = \dots = \phi_n = 0$ mit Atomkoordinaten \dot{r}_α , die starren Einheiten analog zu Gl. [58] einzeln in Standardorientierung gebracht:

$$\dot{r}_\alpha = R(\dot{r}_a^{(1)} - \dot{r}_a^{(2)}, \dot{r}_a^{(3)} - \dot{r}_a^{(2)})(\dot{r}_\alpha - \dot{r}_a^{(2)}) \quad . \quad [62]$$

$\dot{r}_a^{(i)}$ bezeichnet den Ortsvektor des i -ten Atoms in der Definition¹ des Diederwinkels a . Die Transformation [62] wird auf alle Atome in der starren Einheit a angewandt. Zusätzlich wird für jede starre Einheit a eine Rotationsmatrix R_a und ein Translationsvektor \dot{t}_a berechnet — die zwei Größen, die das Gerüst des Baums bilden:

$$R_a = R(\dot{r}_a^{(1)} - \dot{r}_a^{(2)}, \dot{r}_a^{(3)} - \dot{r}_a^{(2)}) R(\dot{r}_{p(a)}^{(1)} - \dot{r}_{p(a)}^{(2)}, \dot{r}_{p(a)}^{(3)} - \dot{r}_{p(a)}^{(2)})^{-1} \quad , \quad [63]$$

$$\dot{t}_a = R(\dot{r}_a^{(1)} - \dot{r}_a^{(2)}, \dot{r}_a^{(3)} - \dot{r}_a^{(2)})(\dot{r}_a^{(2)} - \dot{r}_{p(a)}^{(2)}) \quad . \quad [64]$$

Die Größen \dot{r}_α , R_a und \dot{t}_a brauchen nur einmal — nach dem Einlesen der Aminosäuresequenz — berechnet zu werden. Für jede neue Konformation (ϕ_1, \dots, ϕ_n) baut dann

$$R_a = R(\dot{\epsilon}_z \phi_a) R_a R_{p(a)}^{-1} \quad (\dot{\epsilon}_z = (0, 0, 1)) \quad , \quad [65]$$

$$\dot{r}_a = \dot{r}_{p(a)} + R_{p(a)}^{-1} \dot{t}_a \quad [66]$$

¹ Ein Diederwinkel a wird durch vier Atome mit den Koordinaten $\dot{r}_a^{(1)}$, $\dot{r}_a^{(2)}$, $\dot{r}_a^{(3)}$ und $\dot{r}_a^{(4)}$ definiert (vgl. S. 66), wobei die drehbare Bindung das zweite mit dem dritten Atom in der Diederwinkeldefinition verbindet. Das zweite Atom in der Diederwinkeldefinition ist der Anfangspunkt $\dot{r}_a = \dot{r}_a^{(2)}$ der drehbaren Bindung; für den Einheitsvektor entlang ihrer Achse gilt $\dot{\epsilon}_a = (\dot{r}_a^{(3)} - \dot{r}_a^{(2)}) / |\dot{r}_a^{(3)} - \dot{r}_a^{(2)}|$. Der Wert $\phi_a \in [-\pi, \pi]$ des Diederwinkels a ist mit den Koordinaten $\dot{r}_a^{(1)}$, $\dot{r}_a^{(2)}$, $\dot{r}_a^{(3)}$ und $\dot{r}_a^{(4)}$ durch $\cos \phi_a = (\dot{\epsilon}_1 \cdot \dot{\epsilon}_2) / |\dot{\epsilon}_1| |\dot{\epsilon}_2|$ verknüpft, wobei $\dot{\epsilon}_1 = \dot{\epsilon}_a \wedge (\dot{r}_a^{(1)} - \dot{r}_a^{(2)})$ und $\dot{\epsilon}_2 = \dot{\epsilon}_a \wedge (\dot{r}_a^{(4)} - \dot{r}_a^{(3)})$ sind und ϕ_a das gleiche Vorzeichen wie $\dot{\epsilon}_a \cdot (\dot{\epsilon}_1 \wedge \dot{\epsilon}_2)$ erhält.

rekursiv das Gerüst der drehbaren Bindungen und

$$\dot{\gamma}'_{\alpha} = \dot{\gamma}'_a + R_a \dot{\gamma}'_{\alpha} \quad [67]$$

die Koordinaten $\dot{\gamma}'_{\alpha}$ der Atome auf.

Minimierung der Zielfunktion

Die Zielfunktion wird mit der bekannten Methode der konjugierten Gradienten minimiert (Powell, 1977). Der dafür notwendige Gradient der Zielfunktion kann effizient berechnet werden, weil die Zielfunktion [49] in eine Summe von Funktionen einzelner interatomarer Distanzen und Diederwinkel zerfällt (Noguti & Gō, 1983; Abe *et al.*, 1984). Die partielle Ableitung der Zielfunktion [49] nach einem Diederwinkel a' ist gegeben durch

$$\frac{\partial T}{\partial \phi_{a'}} = -\hat{e}_{a'} \cdot \dot{\gamma}'_{a'} - (\hat{e}_{a'} \wedge \dot{\gamma}'_{a'}) \cdot \dot{\mathbf{g}}_{a'} + 2w_w \sum_{a \in I_w} \left(1 - \left(\frac{\Delta_a}{\Gamma_a}\right)^2\right) \Delta_a \delta_{aa'} \quad [68]$$

mit

$$\begin{aligned} \dot{\gamma}'_{a'} &= \sum_{c=o, u, v} w_c \sum_{\substack{\alpha\beta \in I_c \\ \alpha \in M_{a'}}} \frac{\Theta_c(d_{\alpha\beta}^2 - c_{\alpha\beta}^2)}{c_{\alpha\beta}^2} (\dot{\gamma}'_{\alpha} \wedge \dot{\gamma}'_{\beta}), \\ \dot{\mathbf{g}}_{a'} &= \sum_{c=o, u, v} w_c \sum_{\substack{\alpha\beta \in I_c \\ \alpha \in M_{a'}}} \frac{\Theta_c(d_{\alpha\beta}^2 - c_{\alpha\beta}^2)}{c_{\alpha\beta}^2} (\dot{\gamma}'_{\alpha} - \dot{\gamma}'_{\beta}). \end{aligned} \quad [69]$$

$\dot{\gamma}'_{\alpha}$ und $\dot{\gamma}'_{\beta}$ bezeichnen die Ortsvektoren der Atome α und β , $\hat{e}_{a'}$ einen Einheitsvektor entlang der Bindung a' , $\dot{\gamma}'_{a'}$ den Ortsvektor ihres Anfangspunktes und $M_{a'}$ die Menge aller Atome, deren Position von einer Änderung des Diederwinkels a' betroffen wird, wenn der N-Terminus des Proteins festgehalten wird.

In jedem Minimierungsschritt wird so lange minimiert, bis eine der Abbruchbedingungen erfüllt ist: der Betrag des Gradienten sinkt unter einen vorgegeben Wert, die Zielfunktionswerte nehmen in einer vorgegebenen Anzahl von Iterationen um weniger als 1% ab, die maximal erlaubte Anzahl von Iterationen wird erreicht oder die Minimierung kann aus numerischen Gründen nicht weitergehen. Die Aktualisierung der Liste möglicher van der Waals Wechselwirkungen kann Unstetigkeiten der Zielfunktion zur Folge haben, die aus numerischen Gründen die Minimierung vorzeitig abbrechen lassen. Daher wird die konjugierte Gradienten-Minimierung bei Problemen wegen Sprüngen der Zielfunktion automatisch neu gestartet.

Optimierung der Strukturrechnung mit DIANA

Die Auswertung der Zielfunktion und ihres Gradienten ist im Programm DIANA dadurch sehr effizient, da die Berechnung in mehrere Schritte zerlegt wird, die nur dann ausgeführt werden, wenn es nötig ist. Am Anfang einer Strukturrechnung wird der unveränderliche Teil der Liste möglicher van der Waals Wechselwirkungen aufgebaut (vgl. S. 112), irrelevante Einschränkungen werden gestrichen (vgl. S. 109) und Einschränkungen mit diastereotopen Gruppen werden behandelt (vgl. S. 110). Zu Beginn jedes Minimierungsschritts werden die Teilmengen $I_c \subseteq I_c$ ($c = o, u, v$; vgl. Gl. [50]) der während diesem Minimierungsschritt gebrauchten Einschränkungen entsprechend der gewählten Minimierungsweiten L_c bestimmt, wobei der veränderliche Teil der Liste möglicher van der Waals Wechselwirkungen dem Aktualisierungsverfahren unterworfen bleibt (vgl. S. 113). Der Minimierungsalgorithmus ist so aufgebaut, daß in jeder Iteration die Zielfunktion und ihr Gradient an der gleichen Stelle ausgewertet werden müssen; die Berechnung der kartesischen Atomkoordinaten aus den Diederwinkelwerten und der gemeinsamen Terme der Gl. [49] und [68/69] erfolgt daher nur einmal pro Iteration. Die aufwendigsten Schritte in der Auswertung der Zielfunktion und des Gradienten sind die Berechnung der kartesischen Atomkoordinaten aus den Diederwinkelwerten und der interatomaren Distanzen für die Atompaaire in den Mengen I_o , I_u und vor allem I_v .

Das Programm DIANA ist für Vektorrechner, insbesondere Cray Y/MP Hochleistungsrechner, optimiert, läuft aber, da es soweit wie möglich in Standard FORTRAN-77 (Kießling & Lowes, 1987) geschrieben ist, auf einer Vielzahl von anderen Rechnern. Die folgende Tabelle gibt für einige Rechner die Rechenzeiten¹ an, die eine vollständige, realistische Strukturrechnung mit dem Protein BPTI benötigt. Ausgehend von Zufallsstartstrukturen, wurden mit Hilfe der REDAC-Strategie zur Konvergenzbeschleunigung (Güntert & Wüthrich, 1991) 50 BPTI-Konformere berechnet (für Einzelheiten vgl. Kapitel *Konvergenzbeschleunigung*, S. 134ff.). Pro Konformer wurden die Zielfunktion und ihr Gradient im Durchschnitt rund 12'250 mal, insgesamt also über 600'000 mal, ausgewertet. Auf allen aufgeführten Rechnern wurde nur ein Prozessor eingesetzt. Zusätzlich zur gesamten Rechenzeit gibt die Tabelle den Anteil der aufwendigen Programmteile an der Rechenzeit an: »Koordinaten« bezeichnet die Berechnung der Atomkoordinaten bei gegebenen Diederwinkelwerten (vgl. S. 114ff.), »Verletzungen« das Feststellen der verletzten Einschränkungen und die Berechnung von Termen, die der Zielfunktion [49] und dem Gradienten [68] gemeinsam sind, »Gradienten« die eigentliche Berechnung des Gradienten [68]

¹ Die in der Tabelle angegebenen Rechenzeiten sind niedriger als die auf S. 136, weil eine neuere, besser optimierte Version des Programms DIANA zum Einsatz kam.

und »Paarliste« die Aktualisierung der Liste möglicher van der Waals Einschränkungen (vgl. S. 113).

Rechenzeiten für eine vollständige Strukturrechnung von 50 BPTI-Konformeren

Rechner	Rechenzeit [Std.]	Anteil an der Rechenzeit für die Berechnung der			
		Koordinaten	Verletzungen	Gradienten	Paarliste
NEC SX-3	0,31	31%	31%	22%	10%
Cray YMP	0,34	27%	34%	22%	11%
VAX 9000-420	2,2	19%	38%	14%	20%
SGI Iris 4D/35	3,8	26%	38%	15%	16%
Convex C220	4,3	30%	39%	11%	16%
Sun 4/690	5,5	27%	40%	14%	12%
VAX 6000-520	6,0	33%	30%	13%	18%

Interaktive Eingaben für DIANA

Das Programm DIANA ist grundsätzlich ein interaktives Programm das aber (für längere Rechnungen) auch nicht-interaktiv verwendbar ist. Dieser Abschnitt führt alle Fragen auf, die das Programm während eines Laufs *möglicherweise* stellen wird. *Nicht alle dieser Fragen werden in jedem Lauf gestellt*, weil eine Frage aufgrund vorheriger Eingaben keinen Sinn mehr haben kann. Zu fast allen Fragen schlägt DIANA eine Standardantwort vor, die verwendet wird, wenn der Benutzer mit einer Leerzeile antwortet. Bei Fragen nach Dateinamen schlägt DIANA eine Namensergänzung vor, die Verwendung findet, wenn der Benutzer nicht explizit eine angibt. Ein senkrechter Strich »|« am Ende des Dateinamens verhindert das Anhängen der Standardnamensergänzung. Die Antwort zu einer Frage muß immer auf einer Zeile stehen, und numerische Eingaben müssen den FORTRAN-77-Regeln für listengesteuerte Eingabe (Kießling & Lowes, 1987) gehorchen. Wenn bei numerischen Eingaben nicht alle verlangten Größen vom Benutzer angegeben werden, verwendet DIANA für die Restlichen die Standardwerte. Um Dateien mit Antworten zu Fragen des Programms DIANA, wie sie im nicht-interaktiven Betrieb Verwendung finden, besser lesbar zu machen, können Kommentare eingefügt werden, die sich vom Kommentarzeichen »#« bis zum Ende der Zeile erstrecken.

Die folgende Liste zeigt die Fragen des Programms DIANA in der Reihenfolge, in der sie während des interaktiven Dialogs auftreten (Güntert, 1992):

Output result file (res, default: diana.res):

fragt nach dem Namen der Ergebnisdatei (vgl. S. 109), die immer geschrieben wird.

Input library file (lib, default: Standardbausteinbibliothek):

fragt nach dem Namen der Bausteinbibliothek (vgl. S. 66). Der Name der *Standardbausteinbibliothek* wird mit dem Parameter **DEFLIB** in der Datei **precis.incl** festgelegt.

Input sequence file (seq):

fragt nach dem Namen der Sequenzdatei (vgl. S. 73), der immer anzugeben ist.

Input stereospecific assignments file (ssa, default: none):

fragt nach dem Namen der Datei mit stereospezifischen Zuordnungen (vgl. S. 76). Distanzeinschränkungen, die Atompaaire betreffen, die in dieser Datei auftreten, werden von der Behandlung gemäß Gl. [54–57] nicht erfaßt, d. h. es wird angenommen, daß korrekte stereospezifische Zuordnungen für sie vorliegen. Wird die Frage mit einer Leerzeile beantwortet, werden überall stereospezifische Zuordnungen angenommen; wenn sie mit »*« beantwortet wird, nimmt das Programm an, daß keinerlei stereospezifische Zuordnungen vorhanden sind.

Input upper distance limits file (upl, default: none):

fragt nach dem Namen der Datei mit oberen Distanzschranken (vgl. S. 77).

Input lower distance limits file (lol, default: none):

fragt nach dem Namen der Datei mit unteren Distanzschranken (vgl. S. 77).

Input angle constraints file (aco, default: none):

fragt nach dem Namen der Datei mit Diederwinkeleinschränkungen (vgl. S. 78).

Input minimization parameters file (min, default: none):

fragt nach dem Namen der Datei mit Minimierungsparametern (vgl. S. 107). Falls die Frage mit einer Leerzeile beantwortet wird, führt das Programm keine Minimierung durch und stellt die nächste Frage. Ein Stern »*« als Antwort bedeutet, daß Standardminimierungsparameter (vgl. S. 124) für die Minimierung verwendet werden; das Programm fährt mit der übernächsten Frage fort. Im allgemeinen sind die Standardminimierungsparameter zu empfehlen. Eine Datei mit Minimierungsparametern enthält für jeden Minimierungsschritt eine Zeile, die der Reihe nach die folgenden Größen angibt (vgl. S. 107): die Minimierungsweiten L_o , L_u und L_v für obere, untere und van der Waals Distanzschranken; die Gewichtungsfaktoren w_o , w_u , w_v und w_w für obere, untere und van der Waals Distanzschranken sowie Diederwinkeleinschränkungen; die maximale Anzahl Iterationen (Zielfunktionsauswertungen); die minimale Änderung eines Diederwinkels, damit die Liste möglicher van der Waals Kontakte erneuert wird und die maximale Anzahl Iterationen zwischen zwei Aktualisierungen dieser Liste. Diese Größen werden wie interaktive Eingaben gemäß den FORTRAN-77 Regeln für listengesteuerte Eingabe (Kießling & Lowes, 1987) gelesen. Für Größen, die bei einem Minimierungsschritt nicht angegeben werden, verwendet das Programm die Werte vom vorhergehenden Schritt. Eine Datei mit Minimierungsparametern kann zum Beispiel so aussehen:

#	Weiten	Gewichtungsfaktoren	Iterationen	van der Waals
#				Aktualisierung
0	0 0 0	1.0 1.0 0.2 5.0	100	10.0 100
1	1 1 1	1.0 1.0 0.2 5.0	100	10.0 100
2	2 2 2	1.0 1.0 0.2 5.0	100	10.0 100
3	3 3 3	1.0 1.0 0.2 5.0	200	10.0 100
4	4 4 4	1.0 1.0 0.2 5.0	100	10.0 100
5	5 5 5	1.0 1.0 0.2 5.0	100	10.0 100
10	10 10 10	1.0 1.0 0.2 5.0	400	10.0 100
10	10 10 10	1.0 1.0 0.6 5.0	400	10.0 100
10	10 10 10	1.0 1.0 0.2 5.0	400	10.0 100

Entsprechend den FORTRAN-77 Regeln für listengesteuerte Eingabe ist die folgende, gekürzte Version äquivalent zum obigen Beispiel:

#	Weiten	Gewichtungsfaktoren	Iterationen	van der Waals
#				Aktualisierung
	3*0	2*1.0 0.2 5.0	100	10.0 100

3*1	4*	100/
3*2	4*	100/
3*3	4*	200/
3*4	4*	100/
3*5	4*	100/
3*10	4*	400/
3*10	2*1.0 0.6 5.0	400/
3*10	2*1.0 2.0 5.0	400/

Der Gewichtungsfaktor w_v für van der Waals Einschränkungen kann auch mit einem negativen Vorzeichen versehen werden: in diesem Fall werden die van der Waals Abstoßung der Wasserstoffatome nicht berücksichtigt, anderenfalls ($w_w > 0$) werden alle Atome (außer Pseudoatomen) in die van der Waals Abstoßung miteinbezogen.

Maximal residue index differences and weights (default: 3* L_{\max} 1.0 1.0 2.0 5.0):

fragt nach den Minimierungsweiten L_o , L_u und L_v für obere, untere und van der Waals Distanzschranken und den Gewichtungsfaktoren w_o , w_u , w_v und w_w für obere, untere und van der Waals Distanzschranken sowie Diederwinkelschranken für die die Zielfunktion [49] ausgewertet werden soll. Standardmäßig werden alle vorhandenen Einschränkungen verwendet (L_{\max} bezeichnet die größte Minimierungsweite, d. h. die Anzahl Aminosäurereste). Diese Frage erscheint nur dann, wenn keine Minimierung sondern lediglich eine einzige Auswertung der Zielfunktion gemacht werden soll, d. h. wenn die Antwort auf die vorangehende Frage eine Leerzeile war.

Minimization level range, iteration limits (default: 0 L_{\max} 150 400 50 100):

fragt nach einigen Parametern für die Minimierung und erscheint nur, wenn Standardminimierungsparameter (vgl. S. 124) verwendet werden, d. h. wenn die Frage nach einer Datei mit Minimierungsparametern mit einem Stern »*« beantwortet wird. Die fünf Größen haben folgende Bedeutung: die kleinste und größte Minimierungsweite (L_{\max} bezeichnet die größte mögliche Minimierungsweite, d. h. die Anzahl Aminosäurereste); die maximale Anzahl Iterationen in den Minimierungsschritten, in denen die van der Waals Abstoßung nur für Nicht-Wasserstoffatome berechnet wird; die maximale Anzahl Iterationen in den letzten drei Minimierungsschritten, in denen die van der Waals Abstoßung für alle Atome berechnet wird; die Anzahl Iterationen, während denen die Zielfunktion um mindestens 1% abnehmen muß, damit das Abbruchkriterium »flat« in den Minimierungsschritten, in denen die van der Waals Abstoßung nur für Nicht-Wasserstoffatome berechnet wird, nicht eintritt und die analoge Größe für die letzten drei Minimierungsschritte.

Input structure file or number of structures and seed (ang, default: 1 3771):

fragt nach dem Namen der Dateien mit Startkonformeren *oder* nach der Anzahl der vom Programm zu generierenden Zufallsstartkonformeren und einem Startwert für den Zufallszahlengenerator. Startkonformere können entweder aus Koordinaten- oder Diederwinkeldateien eingelesen (vgl. S. 74/75) oder vom Programm intern gewählt werden. Beim Einlesen von Startkonformeren bestimmt das Programm selbständig, ob eine Koordinaten- oder Diederwinkeldatei vorliegt. Von einer Koordinatendatei wird angenommen, daß sie genau ein Konformer enthält; Diederwinkeldateien können mehrere Konformere enthalten, die durch genau eine Kommentarzeile voneinander getrennt sind. Eine andere Möglichkeit, mehrere Startkonformere anzugeben, ist die Eingabeumleitung: An Stelle einer Koordinaten- oder Diederwinkeldatei kann das Eingabeumleitungszeichen »<<« angegeben werden, gefolgt vom Namen einer Datei, die ihrerseits pro Zeile den Namen einer Koordinaten- oder Diederwinkeldatei enthält. Wird die Frage mit zwei ganzen Zahlen beantwortet, so gibt die erste an, wieviele Startkonformere vom Programm intern aufzubauen sind, und die zweite setzt den Startwert des Zufallszahlengenerators, der dafür verwendet wird. Die Diederwinkel der Zufallsstartkonformere werden als gleichverteilte unabhängige Zufallsvariablen im Intervall $[-180^\circ, +180^\circ]$ angenommen. Der Zufallszahlengenerator ist maschinenunabhängig (Press *et al.*, 1986). Es ist möglich, die Anzahl Startkonformere auf 0 zu setzen, in diesem Fall berechnet das Programm keine Strukturen, sondern behandelt lediglich die Distanzeinschränkungen gemäß Gl. [52–57].

Output cutoffs: f, upper/lower/vdW/angle, fraction (default: 100 3*0.2 5 100):

fragt nach Schwellenwerten für die Ausgabe von Verletzungen der Einschränkungen in der Ergebnis- bzw. Übersichtsdatei. Die erste Schwelle betrifft den Endwert der Zielfunktion: In der Übersichtsdatei werden Verletzungen, Wasserstoffbrücken und Abweichungen der Atompositionen nur für Konformere angezeigt, deren Zielfunktionswert unterhalb dieser Schwelle liegt. Die nächsten vier Schwellenwerte beziehen sich auf obere, untere und van der Waals Distanzschranken sowie Diederwinkeleinschränkungen: nur Verletzungen, die größer als die entsprechende Schwelle sind, werden aufgelistet. Ebenso beziehen sich in der Übersichtsdatei angegebene Anzahlen von Verletzungen nur auf solche, die die entsprechende Schwelle übersteigen. Um alle — auch kleine — Verletzungen zu sehen, müssen die Schwellenwerte auf 0 gesetzt werden. Die letzte Schwelle gibt an, welchen prozentualen Anteil am gesamten Wert der Zielfunktion eine einzelne Verletzung erreichen muß, um in der Ergebnisdatei angezeigt zu werden. Ein Wert $\geq 100\%$ unterdrückt die Anzeige von Verletzungen in der Ergebnisdatei vollständig und verkürzt sie damit erheblich. Diese Schwelle hat keinen Einfluß auf die Übersichtsdatei.

Output modified upper distance limits file (upl, default: none):

fragt nach dem Namen einer Ausgabedatei, in die die oberen Distanzschranken nach der Behandlung gemäß Gl. [52–57], d. h. nachdem irrelevante Einschränkungen eliminiert und solche zu nicht stereospezifisch zugeordneten diastereotopen Substituenten behandelt worden sind, geschrieben werden.

Output modified lower distance limits file (lol, default: none):

fragt nach dem Namen einer Ausgabedatei, in die die unteren Distanzschranken nach der Behandlung gemäß Gl. [52–57], d. h. nachdem irrelevante Einschränkungen eliminiert und solche zu nicht stereospezifisch zugeordneten diastereotopen Substituenten behandelt worden sind, geschrieben werden.

Target function cutoffs to eliminate bad structures (default: none):

fragt nach Zielfunktionsschwellenwerten zum Eliminieren »schlechter« Strukturen. Die Minimierung einer Struktur wird abgebrochen, wenn die Zielfunktionswert am Ende eines Minimierungsschritts die für diesen Schritt angegebene Schwelle übersteigt. Dies steht im Gegensatz zu den Schwellenwerten der vorherigen Frage, die nur die Anzeige von Verletzungen usw. regeln, aber die Berechnung und Ausgabe von Strukturen nicht beeinflussen. Ein oder mehrere Schwellenwerte können in einer der folgenden Formen angegeben werden:

Minimierungsweite=Schwellenwert

Minimierungsschritt==Schwellenwert

Minimierungsweite..Minimierungsweite=Schwellenwert

Minimierungsschritt..Minimierungsschritt==Schwellenwert

Minimierungsweite bezeichnet das Maximum $L = \max(L_o, L_u, L_w)$ der einzelnen Minimierungsweiten. Um zum Beispiel alle Strukturen zu eliminieren, deren Zielfunktion bei den Minimierungsweiten $L = 5-10$ den Wert 2 \AA^2 und bei allen anderen Weiten den Wert 10 \AA^2 übersteigt, lautet die Antwort auf diese Frage:

0..5=2.0 0..100=10.0

Dabei wird benützt, daß die niedrigste Schwelle gilt, wenn für einen Minimierungsschritt mehrere Werte angegeben werden.

Output overview file (ovw, default: diana.ovw):

fragt nach dem Namen der Übersichtsdatei. Diese Datei wird immer geschrieben und heißt, falls kein Name angegeben wird, **diana.ovw**. Während der Berechnung eines Bündels von Konformeren wird die Übersichtsdatei nach jedem fertiggestellten Konformer auf den neusten Stand gebracht, damit der Benutzer den Ablauf der Strukturrechnung verfolgen kann.

Options: vdW violations, H-bonds, RMSDs, connectivities (default: ynnn)?

fragt nach Optionen für die Ausgabe und entscheidet damit, ob Verletzungen von van der Waals Distanzschranken, Wasserstoffbrücken, Abweichungen der Atompositionen (McLachlan, 1979) in der Übersichtsdatei und kovalente Bindungen in Koordinatendateien angezeigt werden. Die Antwort besteht aus vier Buchstaben — »y« bedeutet, daß die entsprechende Ausgabe erfolgt, »n«, daß sie unterdrückt wird. Es ist zu beachten, daß Abweichungen der Atompositionen nur für eine begrenzte Anzahl von Konformeren berechnet werden können¹, weil dazu ihre Diederwinkelwerte im Arbeitsspeicher aufbewahrt werden müssen.

Maximal H-bond length [Å] and angle [deg] (default: 2.4 35.0):

erscheint nur, wenn die Option zum Auflisten von Wasserstoffbrücken gesetzt ist, und fragt nach der maximalen Wasserstoff-Akzeptor-Distanz in Å und dem maximalen Wasserstoff-Donor-Akzeptor-Winkel in Grad. »Donor« ist das Atom mit dem der Wasserstoff kovalent gebunden ist. Im Fall einer Wasserstoffbrücke zwischen einem Amidproton und einem Sauerstoffatom ist der Stickstoff der Donor, der Sauerstoff der Akzeptor.

Residue ranges and segment length for RMSDs (default: M..N 3):

erscheint nur, wenn die Option zum Auflisten von Abweichungen der Atompositionen gesetzt ist, und fragt nach der Teilmenge aller Aminosäurereste für die Berechnung globaler sowie der Segmentlänge (in Anzahl Aminosäurereste) für die Berechnung lokaler paarweiser Abweichungen der Atompositionen. Ein oder mehrere Bereiche der Form

Aminosäurenummer..Aminosäurenummer

legen die genannte Teilmenge für globale Überlagerung fest. Standardmäßig werden alle Reste eingeschlossen (*M* bezeichnet die erste, *N* die letzte Aminosäurenummer) und lokale Überlagerungen werden für Segmente von drei Resten gebildet.

Output angle files (ang, default: none):

fragt nach dem Namen der Ausgabedateien für mit den Diederwinkelwerten der berechneten Konformere (vgl. S. 75). In den gegebenen Dateinamen fügt das Programm vor der Namensergänzung (standardmäßig »**ang**«) für jedes Konformer eine fortlaufende Nummer »001«, »002«, ... ein. Auf Wunsch können auch Konformere am Ende eines beliebigen Zwischenschritts ausgegeben werden, indem anstatt des bloßen Dateinamens

Minimierungsschritt=Dateiname

angegeben wird. In diesem Fall wird die Frage so lange wiederholt, bis sie mit einem bloßen Dateinamen oder einer Leerzeile beantwortet wird.

Output coordinate files (cor, default: none):

fragt nach dem Namen der Ausgabedateien für mit den kartesischen Koordinaten der berechneten Konformere (vgl. S. 74). In den gegebenen Dateinamen fügt das Programm vor der Namensergänzung (standardmäßig »**cor**«) für jedes Konformer eine fortlaufende Nummer »001«, »002«, ... ein. Auf Wunsch können auch Konformere am Ende eines beliebigen Zwischenschritts ausgegeben werden, indem anstatt des bloßen Dateinamens

Minimierungsschritt=Dateiname

angegeben wird. In diesem Fall wird die Frage so lange wiederholt, bis sie mit einem bloßen Dateinamen oder einer Leerzeile beantwortet wird.

Output angle constraint file (aco, default: none):

fragt nach dem Namen der Ausgabedatei mit Diederwinkeleinschränkungen (vgl. S. 78), die zur Konvergenzbeschleunigung mit Hilfe der REDAC-Strategie (Güntert & Wüthrich, 1991; vgl. S. 132ff.) verwendet werden. Die ausgegebenen enthalten die eingegebenen Diederwinkeleinschränkungen, sind aber im allgemeinen enger.

¹ Entsprechend dem Parameter **MAXRMS** in der Datei **lisvio.f**.

Angle constraint cutoffs: f/res., no. of struct., width (default: 0.2 10 270):

erscheint nur, wenn redundante Diederwinkeleinschränkungen gewünscht werden, und fragt nach dem maximalen Zielfunktionswert pro Aminosäurerest, der für das Erzeugen redundanter Diederwinkeleinschränkungen berücksichtigt werden soll, der minimalen Anzahl von Konformeren, auf die sich eine redundante Diederwinkeleinschränkung gründen muß, und nach der maximalen Breite redundanter Diederwinkeleinschränkungen (in Grad). Für Einzelheiten der REDAC-Strategie vgl. S. 132ff.

Standardminimierungsparameter für DIANA

Der Benutzer hat eine weitgehende Kontrolle über die Details der einzelnen Minimierungsschritte, wenn er sie in einer Datei mit Minimierungsparametern (vgl. S. 120) eingibt. Häufig ist dies jedoch unnötig kompliziert, weil das Programm DIANA bis auf sechs vom Benutzer explizit anzugebende Minimierungsparameter (vgl. S. 121), Standardminimierungsparameter annimmt, die in fast allen Fällen gute Ergebnisse liefern.

Die ersten beiden der sechs vom Benutzer anzugebenden Parameter bezeichnen die Minimierungsweiten L_0 und L_{\max} im ersten bzw. letzten Minimierungsschritt. Dazwischen wird die Zielfunktion bei jeder (oder für $L = L_o = L_u = L_v > 10$ jeder zweiten) Weite L_i minimiert, wenn für die Anzahl berücksichtigter oberer und unterer Distanzschranken im Minimierungsschritt i , N_i , entweder $N_i \geq N_{i-1} + 5$ oder gleichzeitig $N_i \geq N_{i-1} + 2$ und $N_{i+1} > N_{i-1} + 5$ gilt oder wenn $L_i \geq L_{i-1} + 10$ ist.

Während der gesamten Rechnung sind die Gewichtungsfaktoren für obere und untere Distanzschranken $w_o = w_u = 1$ und für Diederwinkeleinschränkungen $w_w = 5 \text{ \AA}^2$. Für alle Minimierungsweiten außer der letzten erhalten von der Waals Distanzschranken ein Gewicht $w_v = 0.2$, die von der Waals Abstoßung der Wasserstoffatome wird vernachlässigt und die gleiche vom Benutzer spezifizierte maximale Anzahl von Iterationen wird erlaubt. Für die größte Minimierungsweite L_{\max} führt das Programm drei Minimierungsschritte mit ansteigenden Gewichtungsfaktoren $w_v = 0.2, 0.6$ und 2.0 für von der Waals Distanzschranken durch, die sich von den vorangehenden dadurch unterscheiden, daß die von der Waals Abstoßung der Wasserstoffatome berücksichtigt und eine andere ebenfalls vom Benutzer spezifizierte — im allgemeinen größere — maximale Anzahl von Iterationen erlaubt wird.

Während der ganzen Rechnung wird die Liste möglicher von der Waals Kontakte auf den neusten Stand gebracht, wenn ein Diederwinkel seit der letzten Aktualisierung um mehr als 10° geändert hat, spätestens aber nach 100 Iterationen.

Um die Rechnung etwas abzukürzen, gelangen für die Minimierung zwei zusätzliche Abbruchkriterien zur Anwendung, die nicht gebraucht werden, wenn eine Datei mit Minimierungsparametern eingelesen wird: Das Abbruchkriterium »flat« stoppt die Minimierung, wenn der Wert der Zielfunktion in einer vorgegebenen Anzahl von Iterationen um weniger 1% sinkt. Wenn am Anfang eines Minimierungsschritts der Wert der Zielfunktion um weniger als 10% und der Betrag des Gradienten um weniger als 50% gegenüber dem Ende des vorangegangenen Minimierungsschritts zunimmt, wird der Minimierungsschritt übergangen. Dieses Abbruchkriterium heißt »skip« und wird für alle Minimierungsschritte mit Ausnahme des ersten und letzten angewandt.

Beispiel einer Strukturrechnung mit DIANA

Als Beispiel für eine Strukturrechnung mit dem Programm DIANA zeigt dieser Abschnitt die Eingabe- und Ausgabedaten von DIANA für die Berechnung von 10 Konformeren eines 11 Aminosäureresten langen Polypeptids — der zweiten Helix der *Antennapedia(C39S)* Homeodomäne von *Drosophila melanogaster* — mit experimentellen Daten (Güntert *et al.*, 1991b). Die DIANA-Rechnungen starten von Zufallsstartstrukturen und laufen auf einem Sun-4 Rech-

ner mit dem Betriebssystem UNIX (oder, nach sehr geringen Änderungen, auf einem anderen Rechner) ab. Dieses einfache Beispiel dient dazu, die Ein- und Ausgabedaten des Programms DIANA zu erläutern; für Strukturrechnungen mit Proteinen ist es sehr empfehlenswert, die Konvergenz mit Hilfe redundanter Diederwinkeleinschränkungen zu beschleunigen (vgl. S. 132ff.).

Die Strukturrechnung wurde mit Hilfe des folgenden UNIX-Skripts durchgeführt, das die Antworten auf die Fragen, die das Programm stellt, enthält (vgl. S. 119–124). Kommentare geben an, auf welche Frage sich die Antwort bezieht. Gewisse Antwortzeilen sind bis auf den Kommentar leer; in diesen Fällen verwendet das Programm Standardwerte.

```
diana << EOF
example.res          # Ergebnisdatei (Ausgabe)
../diana.lib        # Bausteinbibliothek (Eingabe)
example.seq         # Aminosäuresequenz (Eingabe)
example.ssa        # Stereospezifische Zuordnungen (Eingabe)
example.upl        # obere Distanzschranken (Eingabe)
example.lol        # untere Distanzschranken (Eingabe)
example.aco        # Diederwinkeleinschränkungen (Eingabe)
*                 # Minimierungsparameter (* = Standardwerte)
                 # Standardminimierungsparameter
10 15191           # Anzahl Konformere und Zufallsstartwert
10.0 3*0.1 1.0 0.0 # Schwellen für Ausgabe von Verletzungen usw.
example_mod.upl    # Modifizierte obere Schranken (Ausgabe)
8==0.3           # Schwellen zum Streichen schlechter Strukturen
example.ovw       # Übersichtsdatei (Ausgabe)
yyyn            # Ausgabeoptionen
2.4 35.0         # Kriterien für H-Brücken: Länge und Winkel
30..38 3         # Bereiche für globale und lokale Überlagerung
example          # Diederwinkel (Ausgabe)
                 # kartesische Koordinaten (Ausgabe)
example_stat.aco  # Diederwinkeleinschränkungen (Ausgabe)
0.1 5 180.0      # Parameter für Diederwinkeleinschränkungen
EOF
```

Die Ergebnisdatei, die während der Strukturrechnung geschrieben wird, beginnt mit einer Liste der Eingabedateinamen, die von einer Liste allgemeiner Parameter gefolgt wird:

Result file of the program DIANA (UNIX version 2.1, single precision)

```
Input files:
Library          : ../diana.lib
Sequence         : example.seq
Stereospecific assignments : example.ssa
Upper distance limits : example.upl
Lower distance limits : example.lol
Angle constraints : example.aco

Number of residues      : 11
Number of atoms        : 240
Number of dihedral angles : 67 (57 rotatable)
Number of stereospec. assignmts: 8 (30 %)
Number of upper distance limits: 146
Number of lower distance limits: 14
Number of angle constraints : 26
Number of structures   : 10
Number of minimization steps : 8
Maximal number of iterations : 1950
Random number generator seed : 15191
Cutoff for upper limits : 0.10 A
Cutoff for lower limits : 0.10 A
Cutoff for van der Waals limits: 0.10 A
Cutoff for angle constraints : 1.00 deg
Cutoff for target fn. fraction : 0.00 %
```

In der Liste gibt »Number of dihedral angles« die gesamte Anzahl Diederwinkel an; auch solche, die bei einem Wert fixiert sind (in Proteinen normalerweise die ω -Winkel der Peptidgruppen; vgl. S. 73), werden mitgezählt. Die Anzahl drehbarer Bindungen, d. i. die Anzahl echter Freiheitsgrade, steht in Klammern dahinter. Der prozentuale Anteil stereospezifischer Zuordnungen bezieht sich auf die gesamte Anzahl prochiraler Zentren im Molekül, die durch die letzte Kolonne der Atomdeklarationen in der Bausteinbibliothek (vgl. S. 66ff.) festgelegt wird — er beträgt immer 100%, wenn die Frage nach der Datei mit stereospezifischen Zuordnungen mit einer Leerzeile beantwortet wird, weil dann durchgehend stereospezifische Zuordnungen angenommen werden (vgl. S. 120). Der letzte Parameter der obigen Liste ist der Schwellenwert, den der prozentuale Beitrag einer Verletzung zum Zielfunktionswert überschreiten muß, damit die Verletzung in der Ergebnisdatei aufgelistet wird.

Als nächster Teil der Ergebnisdatei folgt eine Liste der Änderungen gemäß Gl. [54–57] von Distanzschranken, die nicht stereospezifisch zugeordnete diastereotop Substituenten betreffen:

Modifications for floating stereospecific assignments:										
Atom(s) A		Atom(s) B		Input constraint(s)				modified to		
				A1-B1	A1-B2	A2-B1	A2-B2	Ai-Bj	QA-QB	
Upper	HN	ARG+	29 - HG2/3	ARG+	29	4.00	4.80	4.80	4.33	
Upper	HA	ARG+	29 - HG2/3	ARG+	29		3.60		3.87	
Upper	HN	ARG+	30 - HB2/3	ARG+	30	3.40	2.80	3.40	2.99	
Upper	HA	ARG+	30 - HB2/3	ARG+	30	2.70	3.00	3.00	2.72	
Upper	HB2/3	ARG+	30 - HN	ARG+	31		4.00			
Upper	HN	ARG+	31 - HG2/3	ARG+	31	4.40	4.80	4.80	4.52	
Upper	HA	ARG+	31 - HB2/3	ARG+	31		2.80			
Upper	HA	ARG+	31 - HG2/3	ARG+	31		3.00		3.60	
Upper	HE	ARG+	31 - HG12/3	ILE	35		4.80		5.68	
Upper	HB3	ARG+	32 - HD2/3	ARG+	32	3.00	3.90	3.90	3.37	
Upper	HN	ILE	33 - HG12/3	ILE	33	3.80	4.00	4.00	3.80	
Upper	HA	ILE	33 - HG12/3	ILE	33	3.50	2.80	3.50	3.05	
Upper	HG12/3	ILE	33 - HN	GLU-	34	4.80			5.68	
Upper	HA	ILE	35 - HG12/3	ILE	35		3.10		3.62	
Upper	HG12/3	ILE	35 - QD2	LEU	39	3.40			4.28	

Jede Zeile in dieser Liste behandelt ein Paar von Distanzschranken, wenn ein prochirales Zentrum ohne stereospezifische Zuordnung betroffen ist, oder vier Distanzschranken, wenn diese zwei prochirale Zentren ohne stereospezifische Zuordnung verbinden, wobei in der Eingabe natürlich nicht beide bzw. alle vier Distanzschranken vorhanden sein müssen. Eingabedistanzschranken von einem Atom zum ersten bzw. zweiten diastereotopen Substituenten stehen in den Kolonnen »A1–B1« bzw. »A1–B2«, solche zwischen den zweiten diastereotopen Substituenten zweier prochiraler Zentren in der Kolonne »A2–B2« usw. Die Eingabedistanzschranken in den vier Kolonnen »A1–B1«, »A1–B2«, »A2–B1« und »A2–B2« ersetzt das Programm durch die Distanzschranken in den Kolonnen »Ai–Bj« und »QA–QB«, die unter Vertauschungen von diastereotopen Partnern invariant sind: »Ai–Bj« bezeichnet identische Distanzschranken zu allen (zwei bzw. vier) diastereotopen Substituenten, »QA–QB« solche zu Pseudoatomen, die im Zentrum der diastereotopen Substituenten liegen. Manchmal werden keine modifizierten Distanzschranken angegeben (in der obigen Liste z. B. für HB2/3—HN von Arg 30), weil sie als irrelevant (vgl. S. 109) erkannt wurden.

Zusätzlich gibt die Ergebnisdatei alle irrelevanten oder zu restriktiven Distanzeinschränkungen an:

Meaningless distance constraints:									
						limit	diagnosis		
Upper	HD2	ARG+	31 - QH1	ARG+	31	4.90	no restriction		
Upper	HD2	ARG+	31 - QH2	ARG+	31	5.80	no restriction		
Upper	HD3	ARG+	31 - QH1	ARG+	31	5.20	no restriction		
Upper	HD2	ARG+	32 - QH1	ARG+	32	5.80	no restriction		
Upper	HD2	ARG+	32 - QH2	ARG+	32	5.80	no restriction		
Upper	HD3	ARG+	32 - QH1	ARG+	32	5.80	no restriction		

Number of modified upper limits: 145

Distanzeinschränkungen können aus einem der folgenden Gründe vom Programm als irrelevant erkannt und gestrichen werden (vgl. S. 109):

- »fixed distance« steht bei Distanzschranken, die konformationsunabhängige interatomare Distanzen betreffen;
- »no restriction« bezeichnet Distanzschranken, die von keiner möglichen Konformation verletzt werden können;
- »duplicate constraint« steht bei mehrmals vorkommenden Distanzschranken;
- »weaker than steric limit« bezeichnet untere Distanzschranken, die kleiner als die vom Programm gesetzten van der Waals Distanzschranken, sind.

»Number of modified upper limits« gibt die Anzahl oberer Distanzschranken nach der Behandlung fehlender stereospezifischer Zuordnungen und dem Streichen irrelevanter Einschränkungen an (nicht die Anzahl geänderter Distanzschranken). Diese Zahl kann je nach Anzahl stereospezifischer Zuordnungen größer oder kleiner als die Anzahl eingelesener Distanzschranken sein. Im vorliegenden Beispiel treten keine Änderungen unterer Distanzschranken auf.

Der nächste Teil der Ergebnisdatei zeigt die Namen der Ausgabedateien und die Minimierungsparameter an. »%%« in den Namen von Ausgabedateien mit Diederwinkelwerten oder kartesischen Koordinaten wird für die erste Struktur durch »001«, für die zweite durch »002« usw. ersetzt. In diesem Beispiel wurden Standardminimierungsparameter (vgl. S. 124) verwendet; der Aufbau der Minimierungsparameterliste ist analog zum dem einer Datei mit Minimierungsparametern (vgl. S. 120). Die Buchstaben »H« und »A« nach dem Gewichtungsfaktor w_v für van der Waals Einschränkungen geben an, ob die van der Waals Abstoßung der Nicht-Wasserstoff- (»H«) oder aller Atome (»A«) berücksichtigt wird. Die mit »funct. cutoff« überschriebene Kolonne enthält Zielfunktionsschwellenwerte zum Eliminieren »schlechter« Strukturen (vgl. S. 122).

```
Output files:
Modified upper distance limits : example_mod.upl
Overview                       : example.ovw
Dihedral angles                 : example%%.ang
Angle constraints                : example_stat.aco
```

```
Minimization parameters (standard parameters):
  level |i-j|      weighting factors iter. vdw-update  funct. residue
  upl lol vdw  upper lower   vdw angle limit angle iter  cutoff range(s)
  1   0   0   0 1.0E+0 1.0E+0 2.0E-1H 5.0E+0   150 10.0 100 1.0E+10 29..39
  2   1   1   1 1.0E+0 1.0E+0 2.0E-1H 5.0E+0   150 10.0 100 1.0E+10 29..39
  3   2   2   2 1.0E+0 1.0E+0 2.0E-1H 5.0E+0   150 10.0 100 1.0E+10 29..39
  4   3   3   3 1.0E+0 1.0E+0 2.0E-1H 5.0E+0   150 10.0 100 1.0E+10 29..39
  5   4   4   4 1.0E+0 1.0E+0 2.0E-1H 5.0E+0   150 10.0 100 1.0E+10 29..39
  6  11  11  11 1.0E+0 1.0E+0 2.0E-1A 5.0E+0   400 10.0 100 1.0E+10 29..39
  7  11  11  11 1.0E+0 1.0E+0 6.0E-1A 5.0E+0   400 10.0 100 1.0E+10 29..39
  8  11  11  11 1.0E+0 1.0E+0 2.0E+0A 5.0E+0   400 10.0 100 3.0E-01 29..39
```

Nun folgt der Hauptteil der Ergebnisdatei, in dem für jedes berechnete Konformer der Startwert des Zufallszahlengenerators, verschiedene Einzelheiten über den Verlauf der Minimierung bei jedem Minimierungsschritt sowie eine Liste »wichtiger« Verletzungen von Einschränkungen angegeben wird. Eine Einschränkung wird angezeigt, wenn ihre Größe und ihr relativer Beitrag zum Zielfunktionswert die entsprechenden Schwellen übersteigen. Zur Platzersparnis ausgelassen sind im folgenden die Daten der Strukturen 3–9. Am Ende der Ergebnisdatei stehen Angabe über die gesamte verbrauchte Rechenzeit und über den Anteil verschiedener zeitaufwendiger Programmteile an ihr:

```
Structure 1: (random number generator seed: 15191)
```

```
Minimization:
  lev  upper  lower      vdw  angle target funct. |grad| #up  #f stop
```

		# act	# act	# act	# act	begin	end	end							
1	0	69	9	0	0	180	10	26	0	158.45	0.10	2.1E-3	0	63	gradtl
2	1	96	13	0	0	339	22	26	0	5.56	0.79	2.9E-3	21	120	gradtl
3	2	101	15	0	0	376	26	26	0	5.50	0.39	1.2E-2	10	79	flat
4	3	120	19	0	0	434	31	26	0	312.82	0.16	4.6E-2	13	150	maxit
5	4	145	19	14	2	486	34	26	1	10.68	0.19	0.12	8	150	maxit
6	11	145	20	14	3	1560	33	26	1	0.42	0.12	3.4E-2	12	400	maxit
7	11	145	23	14	2	1553	33	26	1	0.25	0.23	2.3E-2	2	143	flat
8	11	145	22	14	3	1549	32	26	1	0.59	0.39	3.5E-2	9	400	maxit

Structure not accepted.

Structure 2: (random number generator seed: 97312)

Minimization:

lev	upper	lower		vdw	angle	target	funct.	grad	#up	#f	stop				
	# act	# act	#	# act	# act	begin	end	end							
1	0	69	8	0	0	180	8	26	2	148.93	5.8E-2	5.3E-3	0	150	maxit
2	1	96	13	0	0	337	27	26	0	2.74	6.8E-2	8.6E-3	10	122	flat
3	2	101	16	0	0	362	28	26	2	1.89	7.8E-2	2.1E-3	5	108	gradtl
4	3	120	14	0	0	431	27	26	1	152.37	7.5E-2	6.1E-2	15	150	maxit
5	4	145	17	14	6	473	28	26	3	3.39	0.14	0.16	6	150	maxit
6	11	145	22	14	4	1499	25	26	4	0.34	5.3E-2	6.1E-2	9	400	maxit
7	11	145	23	14	4	1494	22	26	4	0.11	8.3E-2	1.7E-2	3	222	flat
8	11	145	22	14	3	1490	21	26	2	0.17	0.15	1.4E-2	3	125	flat

Constraint violations:

						limit(s)	value	viol.	fraction		
Upper	HA	HIS	37	-	HD2	HIS	37	4.30	4.49	0.19	25.32 %
VdW	CG2	ILE	33	-	CD1	ILE	33	2.80	2.66	0.14	23.52 %
Non-listed constraint violations:									51.16 %		

(Hier stünden Ergebnisse für die Strukturen 3–9.)

Structure 10: (random number generator seed: 158120)

Minimization:

lev	upper	lower		vdw	angle	target	funct.	grad	#up	#f	stop				
	# act	# act	#	# act	# act	begin	end	end							
1	0	69	9	0	0	180	9	26	1	168.01	6.2E-2	2.9E-3	0	102	gradtl
2	1	96	7	0	0	337	24	26	0	5.70	4.8E-2	2.5E-3	24	135	gradtl
3	2	101	9	0	0	357	23	26	1	0.84	5.2E-2	2.8E-3	3	74	gradtl
4	3	120	19	0	0	421	24	26	1	39.73	7.9E-2	7.7E-3	6	136	flat
5	4	145	18	14	5	470	33	26	1	53.80	0.27	0.24	9	150	maxit
6	11	145	23	14	4	1506	37	26	3	0.48	0.15	6.9E-2	8	389	flat
7	11	145	22	14	5	1500	33	26	3	0.24	0.21	0.10	3	204	flat
8	11	145	23	14	5	1492	32	26	3	0.35	0.31	0.22	4	210	flat

Structure not accepted.

CPU times:

for the whole program	:	2.32 min
to generate Cartesian coord.	:	0.62 min (26.6 %)
to determine violations	:	1.00 min (43.3 %)
to calculate the gradient	:	0.31 min (13.5 %)
to update the vdW contact list	:	0.14 min (6.1 %)
for the rest	:	0.24 min (10.4 %)

Die detaillierten Angaben über die Minimierung bestehen aus einer Zeile für jeden Minimierungsschritt, wo die erste Kolonne den (fortlaufenden) Index des Minimierungsschritts und die zweite die größte Minimierungsweite $L = \max(L_\sigma, L_w, L_v)$ angibt. Es folgen für alle vier Arten von Einschränkungen — obere, untere und van der Waals Distanzschranken sowie Diederwinkeleinschränkungen — je zwei Kolonnen: die mit »#« gekennzeichnete gibt die gesamte Anzahl, die mit »act« gekennzeichnete die Anzahl aktiver Einschränkungen an, die am Ende des Minimierungsschritts berücksichtigt werden.¹ Danach stehen der Zielfunktionswert am Anfang sowie der Zielfunktionswert und der Betrag des Gradienten am Ende des Minimierungsschritts, die Anzahl Aktualisierungen der Liste möglicher van der Waals Wechselwirkungen (vgl. S.

113), die Anzahl Auswertungen der Zielfunktion und ihres Gradienten und ein Code für das Abbruchkriterium der Minimierung. Es können folgende Codes auftreten:

gradtl	Das Betragsquadrat des Gradienten ist kleiner als der Parameter GSQTOL in der Datei cgmin.f .
maxit	Die maximale Anzahl Iterationen wurde erreicht.
linmin	Die maximal erlaubte Anzahl Iterationen in der eindimensionalen Minimierung wurde überschritten.
nostep	Die Schrittweite während der eindimensionalen Minimierung wurde zu klein.
uphill	Die Richtung eines konjugierten Gradientenschritts zeigt aufwärts.
const	Mehrere aufeinanderfolgende konjugierte Gradientenschritt vermochten den Wert der Zielfunktion nicht zu verringern (vgl. Parameter MAXCON in der Datei cgmin.f).
flat	Der Wert der Zielfunktion sank in einer gegebenen Anzahl Iterationen um weniger als 1%. Dieses Abbruchkriterium tritt nur bei der Verwendung von Standardminimierungsparametern (vgl. S. 124) auf.
skip	Die Minimierung wurde gar nicht erst gestartet, weil am Anfang des Minimierungsschritts der Wert der Zielfunktion um weniger als 10% und der Betrag des Gradienten um weniger als 50% gegenüber dem Ende des vorangegangenen Minimierungsschritts zugenommen hat.

Auf Rechnern mit 64 Bit Genauigkeit für Fließkommazahlen sind »gradtl«, »maxit«, »flat« und »skip« normale Abbruchkriterien, die anderen sollten nur selten auftreten. Bei 32 Bit Genauigkeit treten wegen (harmlosen) numerischen Problemen ab und zu auch andere Abbruchkriterien auf.

Neben der Ergebnisdatei schreibt das Programm DIANA auch eine Übersichtsdatei, die mit jeder fertiggestellten Struktur auf den neusten Stand gebracht wird und einen Überblick über die Zielfunktionswerte, Verletzungen, Wasserstoffbrücken und paarweisen Abweichungen der Atompositionen enthält. Der erste Teil der Übersichtsdatei besteht aus einer nach Zielfunktionswerten geordneten Liste, die neben den Zielfunktionswerten die den entsprechenden Schwellenwert übersteigende Anzahl, die Summe und das Maximum der Verletzungen von oberen, unteren und van der Waals Distanzschranken sowie Diederwinkeleinschränkungen. Jedem Konformer entspricht in dieser Liste eine Zeile, deren erste Kolonne den Rang und deren zweite Kolonne den Index des Konformers angibt. Die Struktur mit dem niedrigsten Zielfunktionswert hat *Rang* 1; die erste berechnete Struktur — mit der Dateinamensergänzung ».001« — den *Index* 1 usw. Die Tabelle enthält nur 6 Konformere, weil die restlichen 4 aufgrund des Kriteriums, daß der Zielfunktionswert akzeptabler Konformere am Ende des achten (und letzten) Minimierungsschritts kleiner als $0,3 \text{ \AA}^2$ sein muß, eliminiert wurden. Die Rechenzeit und die durchschnittliche Anzahl Iterationen pro Struktur entsprechen der Summe dieser Größen über alle Konformere dividiert durch die Anzahl akzeptabler Konformere. Am Ende der Tabelle stehen Mittelwert, Standardabweichung, Minimum und Maximum der entsprechenden Größen:

Overview:

```

Number of accepted structures :      6 (10 structures started)
Residue range for upper limits :      11
                    lower limits :      11
                    van der Waals:      11
Cutoff for upper limits      :      0.10 A
                    lower limits      :      0.10 A
                    van der Waals      :      0.10 A
                    angle constraints  :      1.00 deg
CPU time                      :      2.32 min
CPU time per structure        :      0.39 min

```

¹ *Aktive* Einschränkungen sind solche, die einen nicht-verschwindenden Beitrag zur Zielfunktion liefern.

Average number of iterations : 2453

struct	target function	upper limits			lower limits			van der Waals			torsion angles			
		#	sum	max	#	sum	max	#	sum	max	#	sum	max	
1	4	7.25E-02	1	0.4	0.13	0	0.1	0.07	1	0.3	0.12	0	0.1	0.1
2	8	7.52E-02	1	0.6	0.11	0	0.1	0.08	1	0.4	0.11	0	0.5	0.5
3	5	8.29E-02	1	0.6	0.12	0	0.1	0.07	1	0.4	0.12	0	0.3	0.3
4	3	0.11	1	0.5	0.12	0	0.1	0.08	2	0.5	0.13	0	0.1	0.1
5	2	0.15	1	0.8	0.19	0	0.1	0.03	1	0.5	0.14	0	0.2	0.2
6	6	0.25	3	1.2	0.20	1	0.2	0.13	2	0.8	0.14	1	1.4	1.4
Average		0.12	1	0.7	0.15	0	0.1	0.08	1	0.5	0.13	0	0.4	0.4
+/-		6.32E-02	1	0.3	0.04	0	0.0	0.03	0	0.1	0.01	0	0.4	0.4
Minimum		7.25E-02	1	0.4	0.11	0	0.1	0.03	1	0.3	0.11	0	0.1	0.1
Maximum		0.25	3	1.2	0.20	1	0.2	0.13	2	0.8	0.14	1	1.4	1.4

Der zweite Teil der Übersichtsdatei besteht aus einer Liste aller Verletzungen von Einschränkungen, die in mindestens einem Konformer, dessen Zielfunktion unter dem Zielfunktionswellenwert liegt, den entsprechenden Schwellenwert überschreiten. Im vorliegenden Beispiel ist der Zielfunktionswellenwert so groß gewählt, daß Verletzungen aus allen sechs akzeptierten Konformern in die Übersicht eingehen. Nach dem Maximalwert der Verletzung folgt für jedes Konformer eine Kolonne, in der ein »*«, wenn dieses Konformer die maximale Verletzung zeigt, ein »+«, wenn die Verletzung in diesem Konformer größer als der zugehörige Schwellenwert ist, oder sonst nichts steht. Die Konformere sind entsprechend ihrem Rang — d. h. nach ansteigenden Zielfunktionswerten — geordnet. Konsistente Verletzungen sind solche, die in allen Konformern den Schwellenwert überschreiten. Gleich wie Verletzungen der Einschränkungen werden auch Wasserstoffbrücken angezeigt:

Constraint violation and hydrogen bond overview (structures ordered):

Cutoff for target function	:	1.00E+01
Number of structures included	:	6
Number of violated constraints	:	8
Number of consistent violations	:	0
Maximal hydrogen bond length	:	2.40 A
Maximal hydrogen bond angle	:	35.00 deg
Number of hydrogen bonds	:	9
Number of consistent H-bonds	:	3
		max 1 5
Upper HA	GLU-	34 - HB3 GLU- 34 0.14 ++++ *
Upper QG1	ILE	35 - QD2 LEU 39 0.20 *
Upper HA	HIS	37 - HD2 HIS 37 0.19 *
Upper HD2	HIS	37 - HA ALA 38 0.10 *
Lower O	ILE	33 - N HIS 37 0.13 *
VdW CG2	ILE	33 - CD1 ILE 33 0.14 +*+
VdW CG	GLU-	34 - C GLU- 34 0.13 ++++ *
Angle CHI1	LEU	39 1.37 *
Hbond HH22	ARG+	31 - OE1 GLU- 34 1 +
Hbond HN	ILE	33 - O ARG+ 29 6 ++++++
Hbond HN	GLU-	34 - O ARG+ 30 6 ++++++
Hbond HN	ILE	35 - O ARG+ 31 6 ++++++
Hbond HN	ALA	36 - O ILE 33 5 ++++ +
Hbond HN	HIS	37 - O ILE 33 5 +++++
Hbond HN	HIS	37 - O GLU- 34 1 +
Hbond HN	ALA	38 - O GLU- 34 3 ++ +
Hbond HN	LEU	39 - O ILE 35 5 ++++ +

Der letzte Teil der Übersichtsdatei enthält paarweise Abweichungen der Atompositionen: global für alle Reste im angegebenen Bereich (vgl. S. 123), wobei sich in der Matrix Einträge oberhalb der Diagonalen auf die Rückgratátome (in Proteinen N, C^α und C') und unterhalb auf alle »Schwerátome« (in der Standardbausteinbibliothek alle Átome außer Wasserstoff- und Pseudátomen; vgl. »ATOMTYPES« auf S. 67) beziehen, lokal für Segmente mit der angegebenen Länge (»local RMSD/segment«; vgl. S. 123) und Verschiebungen der Atompositionen (»global-RMSD/residue«; *Displacements* bei Billeter *et al.*, 1989 und Billeter, 1992).

Pairwise RMSDs (structures ordered):

Number of backbone atoms : 27
 Number of heavy atoms : 78
 Residues considered : 30..38
 Local RMSD segment length : 3 residues
 Mean global backbone RMSD : 0.18 +/- 0.09 A (0.06..0.38 A)
 Mean global heavy atom RMSD: 1.43 +/- 0.28 A (0.79..1.88 A)

	1	2	3	4	5	6
1		0.09	0.10	0.06	0.26	0.22
2	0.79		0.07	0.06	0.20	0.26
3	1.29	1.12		0.08	0.20	0.25
4	1.33	1.36	1.38		0.24	0.23
5	1.87	1.88	1.77	1.28		0.38
6	1.42	1.46	1.69	1.49	1.34	

Residue	local RMSD/segment		global RMSD/residue	
	backbone	heavy atoms	backbone	heavy atoms
29	0.00	0.00	0.28	2.07
30	0.07	1.56	0.13	2.34
31	0.04	1.40	0.09	1.01
32	0.05	1.05	0.11	1.68
33	0.06	1.17	0.12	0.64
34	0.07	0.92	0.16	1.26
35	0.10	0.92	0.14	0.71
36	0.10	0.96	0.16	0.25
37	0.14	0.97	0.32	1.87
38	0.19	0.94	0.22	0.28
39	0.00	0.00	0.41	0.55

Konvergenzbeschleunigung in Strukturrechnungen mit dem Programm DIANA

Die Grundidee des Verfahrens der variablen Zielfunktion (Braun & Gö, 1985) ist die schrittweise Anpassung einer Zufallsstartkonformation an in Kernresonanzexperimenten gewonnene geometrische Einschränkungen, beginnend mit Einschränkungen innerhalb der einzelnen Aminosäurereste und sich nach und nach ausdehnend auf Einschränkungen zwischen entlang der Kette weiter und weiter entfernten Resten, bis alle Einschränkungen berücksichtigt sind. Das Verfahren der variablen Zielfunktion ist in den Programmen DISMAN (Braun & Gö, 1985), DADAS (Kohda *et al.*, 1988) und DIANA (Güntert *et al.*, 1991a) verwirklicht und diente zur Strukturberechnung vieler Proteine (z. B. Wagner *et al.*, 1987; Arseniev *et al.*, 1988; Kline *et al.*, 1989; Qian *et al.*, 1989; Güntert *et al.*, 1991b; Ikura *et al.*, 1991), so daß sein Verhalten in der Praxis gut beurteilt werden kann. Vorteile der Methode sind, daß sie konzeptionell einfach ist und im Diederwinkelraum, d. h. mit fester Standardgeometrie, arbeitet. Ein wesentlicher Nachteil war bislang, daß für alle außer die einfachsten Faltungen bloß ein kleiner Teil der Zufallsstartstrukturen zu Lösungen mit genügend kleinen Restverletzungen der Einschränkungen konvergiert: ein typisches Problem lokaler Minima (Li & Scheraga, 1987). Wegen der niedrigen Ausbeute von annehmbaren Konformeren mußten die Rechnungen von einer hohen Anzahl von Zufallsstartstrukturen ausgehen, um eine relativ kleine Gruppe guter Lösungen zu finden. In manchen Fällen zwang dies gar zu Kompromissen zwischen den Anforderungen an die Qualität der Strukturen und der verfügbaren Rechenzeit (Kline *et al.*, 1989; Widmer *et al.*, 1989). Die Einführung des hochoptimierten Programms DIANA, das die Zeit zur Berechnung eines einzelnen Konformers stark verminderte, erlaubte zwar die Behandlung von α -Proteinen mit vertretbarem Aufwand, aber für β -Proteine blieb die Situation oft unbefriedigend (Vendrell *et al.*, 1991; Sodano *et al.*, 1991).

Mit dem Gebrauch redundanter Diederwinkeleinschränkungen, wie er in diesem Kapitel eingeführt wird (Güntert & Wüthrich, 1991), beschleunigt sich die Konvergenz des Verfahrens der variablen Zielfunktion wesentlich, und die Ausbeute konvergierender Rechnungen wird auch für β -Proteine hoch.

Redundante Diederwinkelschränkungen (REDAC)

Die neue Strategie für die Verwendung des Programms DIANA mit redundanten Diederwinkelschränkungen (REDAC, »*redundant dihedral angle constraints*«, Güntert & Wüthrich, 1991) ist im folgenden Flußdiagramm dargestellt. Sie ist eine Erweiterung des ursprünglichen, »direkten« Verfahrens der variablen Zielfunktion, wie es von Braun & Gö (1985) vorgeschlagen wurde, und Verbesserungen mit dem neuen Verfahren werden relativ zum direkten Verfahren gewertet.

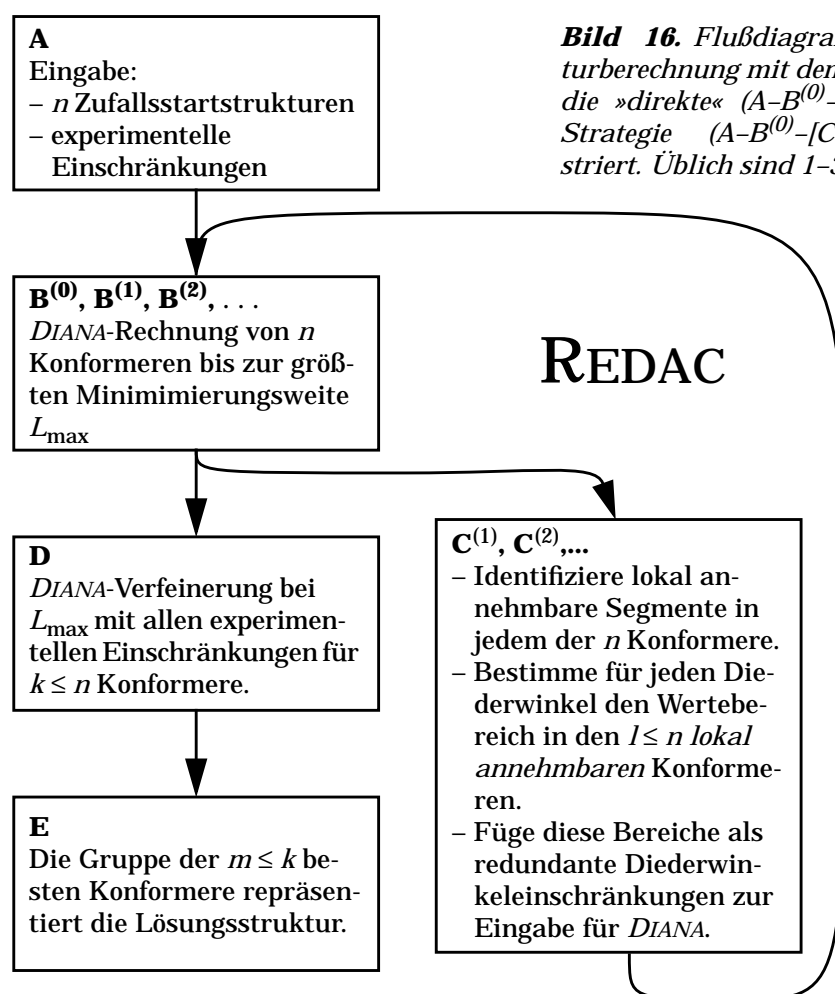


Bild 16. Flußdiagramm einer Proteinstrukturberechnung mit dem Programm DIANA, das die »direkte« ($A-B^{(0)}-D-E$) und die REDAC-Strategie ($A-B^{(0)}-[C^{(1)}-B^{(1)}-...]-D-E$) illustriert. Üblich sind 1-3 REDAC-Zyklen.

Das direkte Verfahren beginnt mit n Zufallsstartstrukturen und experimentellen Einschränkungen nach einer ersten Analyse der lokalen Einschränkungen mit dem Programm HABAS (Güntert *et al.*, 1989; Schritt **A** im Flußdiagramm). Die n Konformere werden einer DIANA-Minimierung der variablen Zielfunktion unterworfen ($B^{(0)}$). Die Erfahrung hat gezeigt, daß sich die

Zielfunktion bereits gut konvergierter Lösungen noch weiter verringern läßt, indem man die Minimierung mit *allen* Einschränkungen fortsetzt und dabei das Gewicht der van der Waals Einschränkungen verändert. Dieser Verfeinerung wird eine begrenzte Zahl k ($m \leq k \leq n$) von Konformeren unterzogen (D). Schließlich repräsentiert die Gruppe der m Konformere mit den niedrigsten Werten der Zielfunktion die Lösungsstruktur des Proteins (E). Im allgemeinen wird n so angepaßt, daß sich $m = 10\text{--}20$ annehmbare Konformere ergeben.

Um redundante Diederwinkeleinschränkungen zu verwenden, wird die direkte Strategie durch einen oder mehrere (im allgemeinen 1–3) REDAC-Zyklen ($C^{(j)}\text{--}B^{(j)}$) erweitert, die eine Rückkopplung von Information aus dem gesamten Satz von Konformeren bewirken, die bis zur größten Minimierungsweite L_{\max} (vgl. Gl. [50]) in $B^{(i-1)}$ gerechnet wurden. Im Schritt $C^{(j)}$ betrachtet man die Konformation eines gegebenen Aminosäurerests als *lokal annehmbar*, wenn Verletzungen von Einschränkungen, die Atome oder Diederwinkel des Rests betreffen, eine vorgegebene Schwelle, typischerweise $0,4 \text{ \AA}$, nicht überschreiten und die gleiche Bedingung auch für die beiden benachbarten Reste gilt. Redundante Diederwinkeleinschränkungen werden für all jene Reste erzeugt, die in mindestens einer vorgegebenen Zahl von Konformeren lokal annehmbar sind, typischerweise 10 bei $n = 50$ Zufallsstartstrukturen, wobei die beiden Extremwerte eines Diederwinkels in der Gruppe lokal annehmbarer Konformere die Schranken des erlaubten Bereichs bilden. Falls dieser Bereich nicht zu breit, typischerweise schmaler als 270° ist, wird er als redundante Diederwinkeleinschränkung der Eingabe für die nächste DIANA-Minimierung im Schritt $B^{(j)}$ beigefügt. Das Programm DIANA erzeugt redundante Diederwinkeleinschränkungen automatisch.

Ausmaß der Konvergenzbeschleunigung

Um die Konvergenzbeschleunigung dank REDAC im Vergleich zur direkten Methode zu untersuchen, wurden, ausgehend von den Kernresonanzdatensätzen für die *Antennapedia(C39S)* Homeodomäne (Güntert *et al.*, 1991b), den basischen pankreatischen Trypsininhibitor (BPTI; Berndt *et al.*, 1992) und die Aktivierungsdomäne von Procarboxypeptidase B (Vendrell *et al.*, 1991) mit DIANA Strukturen berechnet. Die Homeodomäne ist ein typisch α -helikales Protein, BPTI enthält sowohl α als auch β Sekundärstruktur, und die Aktivierungsdomäne besitzt eine kompliziertere Faltung, bestehend aus einem viersträngigen β -Faltblatt und zwei α -Helices sowie drei Schleifen, deren Konformation die Kernresonanzdaten nur ungenügend festlegen (Vendrell *et al.*, 1990).

Für jedes Protein wurde eine Strukturrechnung mit der REDAC Strategie, ausgehend von $n = 50$ Zufallsstartstrukturen durchgeführt und die Gruppe der $m = 20$ Konformere mit den tiefsten Zielfunktionswerten für die weitere Analyse ausgewählt. Die DIANA-Minimierung in den Schritten $B^{(j)}$ arbeitete mit den Standardwerten für die Minimierungsparameter, d. h. höchstens 150 Auswertungen der Zielfunktion und einem Gewicht $w_v = 0.2$ für van der Waals Einschränkungen in allen Minimierungsschritten außer dem letzten bei L_{\max} , wo dreimal eine Grenze von höchstens 400 Zielfunktionsauswertungen und van der Waals Gewichte w_v von 0,2, 0,6 und 2,0 galten. Die übrigen Gewichte, $w_u = w_l = 1$ und $w_a = 5 \text{ \AA}^2$, blieben während der ganzen Rechnung unverändert. Schritt D wurde für alle $k = n = 50$ Konformere durchgeführt, wobei Grenzen von dreimal höchstens 1000 Zielfunktionsauswertungen und wiederum van der Waals Gewichte w_v von 0,2, 0,6 und 2,0 galten. In den Schritten $C^{(j)}$ wurden redundante Diederwinkelschränkungen von höchstens 270° Breite erzeugt, falls der Zielfunktionsanteil pro Rest in mindestens 10 Konformeren kleiner als $0,4 \text{ \AA}^2$ war (vgl. S. 134). Für die Homeodomäne und BPTI

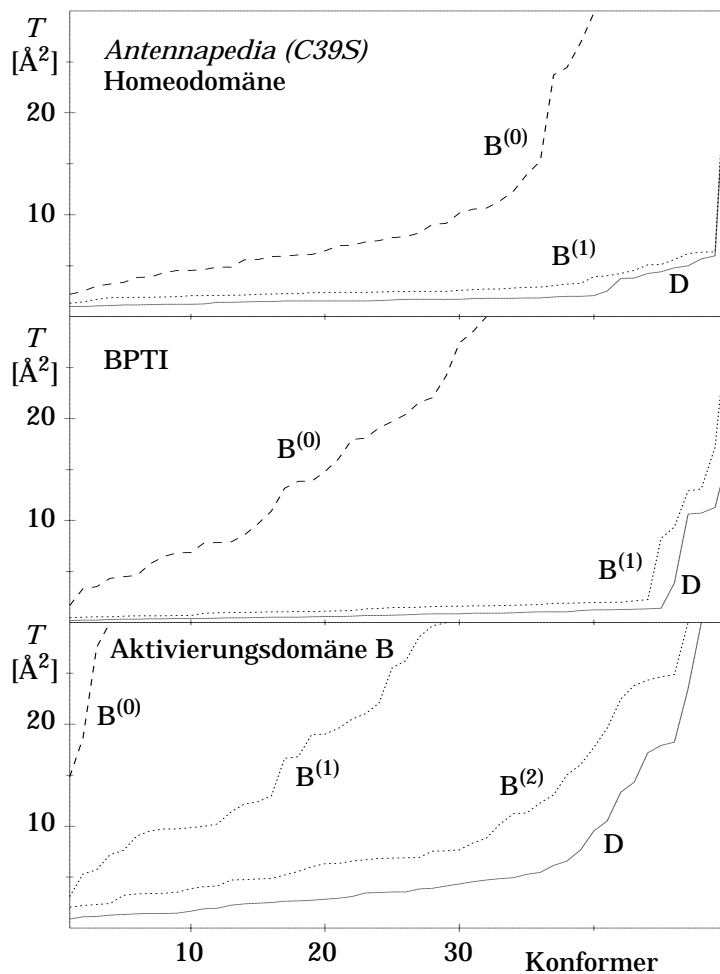


Bild 17. Verteilung der DIANA Zielfunktionswerte T von 50 mit REDAC berechneten Konformeren für die Antennapedia(C39S) Homeodomäne, BPTI und die Aktivierungsdomäne B. Entlang der Horizontalen sind die 50 Konformere nach ansteigenden Zielfunktionswerten geordnet. Die Rechnung bestand für die Homeodomäne und BPTI aus den Schritten $A-B^{(0)}-C^{(1)}-B^{(1)}-D$ und für die Aktivierungsdomäne aus $A-B^{(0)}-C^{(1)}-B^{(1)}-C^{(2)}-B^{(2)}-D$ (vgl. Bild 16). Aufgetragen sind die Ergebnisse am Ende der Schritte $B^{(j)}$ und D.

genügte jeweils ein REDAC-Zyklus, um mehr als 40 Konformere mit Zielfunktionswerten unter 2,1 bzw. 1,3 Å² zu erhalten. Die Aktivierungsdomäne erforderte zwei REDAC-Zyklen für eine Gruppe von 20 Konformeren mit Zielfunktionswerten unter 2,9 Å².

Für die Homeodomäne und BPTI wurden auch Strukturrechnungen gemäß der direkten Strategie mit dem Ziel durchgeführt, eine Gruppe von 20 Konformeren gleicher Qualität, d. h. mit gleichen Zielfunktionswerten, wie mit der REDAC-Strategie zu erzeugen. Für die Homeodomäne benötigte dies $n = 400$, für BPTI gar $n = 2000$ Zufallsstartstrukturen. Um einen fairen Vergleich zu ermöglichen, wurde die maximal erlaubte Anzahl von Zielfunktionsauswertungen im Schritt B⁽⁰⁾ gegenüber der Rechnung mit REDAC verdoppelt und nur die $k = 50$ Konformere mit den niedrigsten Zielfunktionswerten am Ende des Schritts B⁽⁰⁾ wurden in Schritt D verfeinert. Für die Aktivierungsdomäne lieferte das direkte Verfahren aus 400 Zufallsstartstrukturen nur ein akzeptables Konformer, es wären also rund 8000 Zufallsstartstrukturen nötig, um ein mit der REDAC-Strategie vergleichbares Ergebnis zu erzielen. Die folgende Tabelle zeigt, daß die Verwendung redundanter Diederwinkelsinschränkungen die Rechenzeit für Strukturrechnungen mit den drei Proteinen 5,7, 29 bzw. etwa 100 mal zu verringern vermag:

Effizienz von DIANA Rechnungen mit direkter und REDAC-Strategie

	<i>Antennapedia</i> Homeodomäne		BPTI		Aktivierungsdomäne B	
	direkt	REDAC	direkt	REDAC	direkt	REDAC
Nötige Anzahl Startstrukturen n	400	50	2000	50	≈8000	50
Rechenzeit [h] (auf Cray Y/MP)	3,8	0,66	17,7	0,61	≈140	1,48

Die Konvergenzbeschleunigung ist für β -Proteine am ausgeprägtesten, also dort wo die direkte Strategie die geringste Ausbeute konvergierter Strukturen ergibt.

Die erzielte Konvergenzbeschleunigung mit REDAC ist natürlich nur dann aussagekräftig, wenn die DIANA-Rechnungen mit und ohne REDAC äquivalente Gruppen von Strukturen liefern, die sich sowohl in den Restverletzungen der Einschränkungen als auch in der Verteilung der Konformere im erlaubten Konformationsraum nicht merklich unterscheiden. Dies ist tatsächlich der

Fall, wie die folgende Tabelle mit der Analyse der jeweils 20 besten Homeodomänen- und BPTI-Konformere zeigt; für die zwei Proteine zeigen sich zwischen den beiden Strategien kaum Unterschiede in den Werten der Zielfunktion, der verschiedenen Maße für Restverletzungen von Einschränkungen und der globalen, paarweisen Abweichungen der Atompositionen (McLachlan, 1979) im Bereich der gut definierten Aminosäurereste 7–59 für die Homeodomäne (Güntert *et al.*, 1991b) und 3–55 für BPTI:

Analyse der Strukturen aus DIANA-Rechnungen mit und ohne REDAC

	<i>Antennapedia</i> Homeodomäne		BPTI	
	direkt	REDAC	direkt	REDAC
Zielfunktionswert [\AA^2]	1,31±0,31	1,29±0,19	0,40±0,12	0,49±0,11
Restverletzungen von Distanzschranken:				
Anzahl > 0,2 \AA	3±2	3±2	0	0
Maximum [\AA]	0,28±0,06	0,28±0,06	0,18±0,02	0,19±0,03
Summe [\AA]	9,8±0,9	9,8±0,8	3,1±0,6	3,1±0,6
Restverletzungen von Diederwinkelschranken:				
Anzahl > 5°	0	1±1	0	0
Maximum [°]	4,5±1,5	5,0±1,4	1,9±1,0	1,8±1,0
Summe [°]	25,8±4,8	24,7±4,7	5,9±2,3	5,7±2,0
Abweichungen der Atompositionen:				
verglichene Reste	7–59	7–59	3–55	3–55
Rückgratatome [\AA]	0,76±0,14	0,80±0,16	0,67±0,12	0,67±0,13
Schweratome [\AA]	1,70±0,13	1,76±0,16	1,49±0,12	1,48±0,14

Paarweise Abweichung der Atompositionen zwischen den mit direkter und REDAC-Strategie berechneten Strukturen

Atome	Homeodomäne (Reste 7–59)	BPTI (Reste 3–55)
Rückgratatome [\AA]	0,77 ± 0,15	0,67 ± 0,14
Schweratome [\AA]	1,74 ± 0,15	1,47 ± 0,16

Darüberhinaus sind die Abweichungen der Atompositionen im paarweisen Vergleich von Konformeren aus *verschiedenen* Bündeln praktisch gleich den entsprechenden Werten *innerhalb* der beiden Bündel. Auch die Atomverschiebungen der einzelnen Aminosäurereste nach globaler Überlagerung (Billeter *et al.*, 1989) weichen kaum voneinander ab:

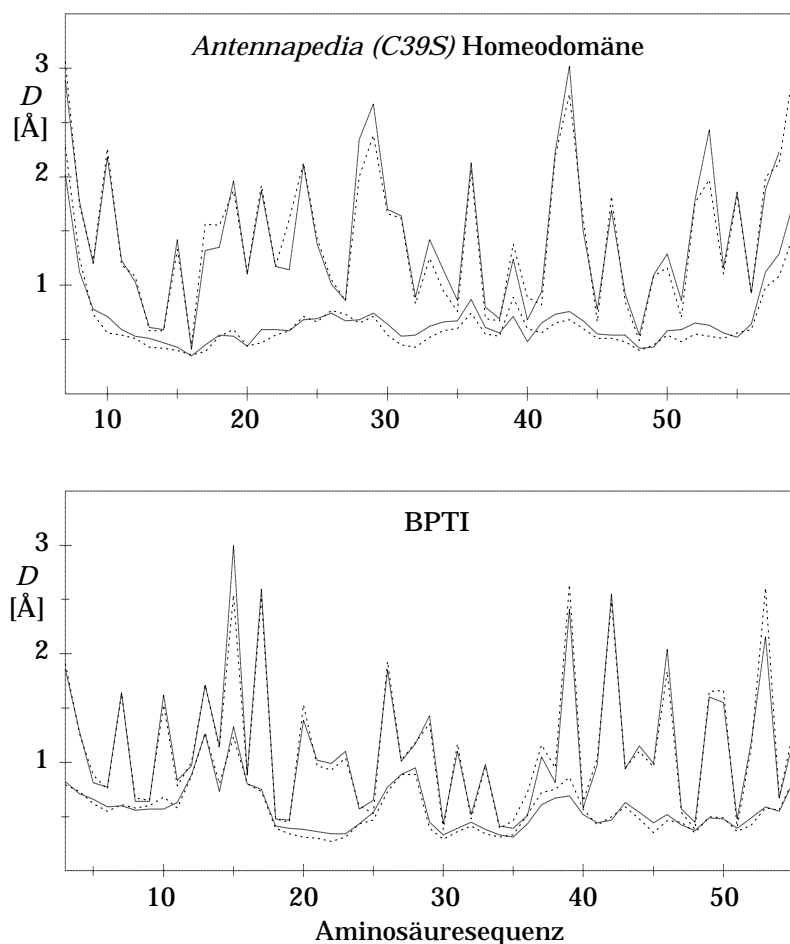


Bild 18. Mittlere Atomverschiebungen D der Rückgratatome N, C^α und C' (untere Kurven) und aller Schweratome (obere Kurven) der einzelnen Aminosäurereste nach globaler Überlagerung der 20 besten DIANA Konformere, die mit der direkten Strategie (gepunktete Kurven) und mit REDAC (ausgezogene Kurven) berechnet wurden. Die globale Überlagerung wurde für den Bereich der gut definierten Reste, d. h. für die Reste 7–59 der Homeodomäne und 3–55 von BPTI gemacht, und auch nur für diese Reste sind die Atomverschiebungen aufgetragen.

Die mit beiden Strategien erzeugten Bündel von Strukturen besitzen also die gleiche Qualität in bezug auf Restverletzungen der Einschränkungen und die gleiche Ausdehnung und Lage im Konformationsraum.

Diese Daten führen zum wichtigen Schluß, daß die Gruppen von Strukturen, welche mit der direkten und der REDAC-Strategie berechnet wurden, sich nicht signifikant voneinander unterscheiden, und daß somit die Verwendung redundanter Diederwinkelschränkungen ein zulässiges und erfolgreiches Mittel zur Konvergenzbeschleunigung ist, das zum Standardverfahren für Proteinstrukturberechnungen mit der Methode der variablen Zielfunktion werden sollte (Güntert & Wüthrich, 1991).

Grundlage der Konvergenzbeschleunigung

Die empirisch gefundene höhere Ausbeute konvergenter Strukturen bei der Verwendung redundanter Diederwinkeleinschränkungen kann so begründet werden: In verschiedenen Bereichen einer Proteinstruktur, insbesondere in β -Strängen, wird die lokale Konformation nicht allein durch lokale Einschränkungen festgelegt, sondern auch durch Einschränkungen zwischen bezüglich der Sequenz weiter auseinanderliegenden Aminosäureresten, zum Beispiel durch Einschränkungen zwischen den einzelnen Strängen eines β -Faltblatts. Umgekehrt bedeutet dies, daß die lokalen Einschränkungen lokale Konformationen zulassen, die bei kleinen Minimierungsweiten am Anfang einer DIANA-Rechnung auftreten können, aber nicht mehr erlaubt sind, sobald im weiteren Verlauf der Rechnung größere Minimierungsweiten gelten. Offensichtlich bilden falsche lokale Konformationen, die mit den lokalen Einschränkungen verträglich sind, potentielle lokale Minima, welche nur dann von vornherein vermieden werden können, wenn die Information der weitgespannten Einschränkungen bereits bei engen Minimierungsweiten am Anfang der DIANA-Rechnung nutzbar wäre. Der Gebrauch redundanter Diederwinkeleinschränkungen erreicht dieses Ziel: Information aus dem gesamten Datensatz wird in — per Definition lokale — Diederwinkeleinschränkungen übersetzt, die vom Programm effizient verwertet werden können. Wie wichtig der Einbezug weitgespannter Einschränkungen ist, zeigt der nur begrenzte Erfolg früherer Versuche zur Verwendung redundanter Diederwinkeleinschränkungen, die aus Einschränkungen mit einer Weite von weniger als fünf Resten gewonnen wurden (Kline *et al.*, 1988; Widmer *et al.*, 1989; Billeter *et al.*, 1990a). Auch die höhere Konvergenzrate der direkten Strategie bei α - im Vergleich zu β -Proteinen erklärt sich damit, daß die Konformation einer α -Helix besonders gut durch lokale Einschränkungen festgelegt ist (Wüthrich *et al.*, 1984; Wüthrich, 1986).

Der Erfolg der REDAC-Strategie beruht auf einer Rückkopplung nutzbringender Information über die Konformation, die aus den lokal annehmbaren Teilen *aller* zuvor berechneten Konformere stammt, womit auch lokal annehmbare Abschnitte von »schlechten« Konformeren mit hohem Zielfunktionswert noch gebraucht werden können, während diese Information beim direkten Verfahren ungenutzt bleibt.

Die Anzahl lokal annehmbarer Konformationen der einzelnen Aminosäurereste für das Erzeugen redundanter Diederwinkeleinschränkungen (Bild 19) ist im Mittel bei der Homeodomäne am höchsten, obwohl die Endwerte der Zielfunktion bei BPTI niedriger sind. Bei BPTI und der Aktivierungsdomäne ergeben sich in den β -Strängen besonders wenige lokal annehmbare Konformationen, währendem Schleifen, Helices und die oft schlecht bestimmten un-

regelmäßigen Abschnitte der Kette im allgemeinen zu einer höheren Zahl lokal annehmbarer Konformationen führen — eine weitere Bestätigung der oben gegebenen Erklärung der Konvergenzbeschleunigung.

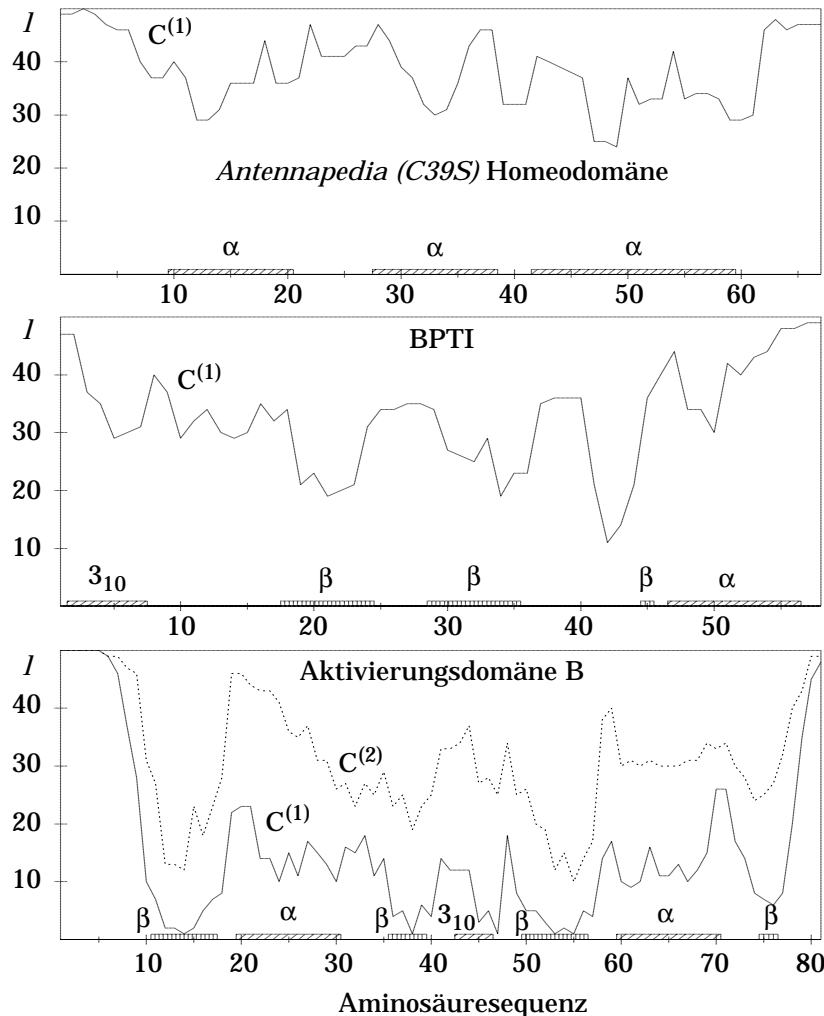


Bild 19. Anzahl lokal annehmbarer Konformationen l der einzelnen Aminosäurereste (im Sinne der Definition von S. 134) im Schritt $C^{(i)}$ der DIANA-Rechnungen mit der REDAC-Strategie (vgl. Bild 16). Die Lage der Helices (α , 3_{10}) und β -Stränge ist mit schraffierten Balken wiedergegeben.

Gebrauch der REDAC-Strategie mit dem Programm DIANA

Um die Konvergenzbeschleunigung mit Hilfe redundanter Diederwinkelsinschränkungen im praktischen Gebrauch mit dem Programm DIANA zu erzielen, sind mehrere Läufe des Programms DIANA nötig. So entspricht zum Beispiel in einer Strukturrechnung mit einem REDAC-Zyklus der erste Lauf den Schritten A, $B^{(0)}$ und $C^{(1)}$ in Bild 16; der zweite dem Schritt $B^{(1)}$ und der letzte dem Schritt D. Der Berechnung der obigen BPTI-Strukturen diente das folgende UNIX-Skript:

```

diana << EOF
bptila.res          # Ergebnisdatei
                    # Bausteinbibliothek
  ../bpti.seq       # Aminosäuresequenz
                    # Stereospezifische Zuordnungen
  ../bpti.upl       # obere Distanzschranken
  ../bpti.lol       # untere Distanzschranken
  ../bpti.aco       # Diederwinkeleinschränkungen
  *                 # Minimierungsparameter (* = Standardwerte)
                    # Standardminimierungsparameter
50 21263            # Anzahl Konformere und Zufallsstartwert
2.0 3*0.2 5.0 100  # Schwellen für Ausgabe von Verletzungen usw.
                    # Schwellen zum Streichen schlechter Strukturen
bptila.ovw         # Übersichtsdatei
ynnn               # Ausgabeoptionen
bptila.ang         # Diederwinkel
                    # kartesische Koordinaten
bptila.aco         # redundante Diederwinkeleinschränkungen
0.4 10 270.0      # Parameter für Diederwinkeleinschränkungen
EOF
cat bptila???.ang >bptila.ang
rm bptila???.ang
# =====
diana << EOF
bptilb.res         # Ergebnisdatei
                    # Bausteinbibliothek
  ../bpti.seq       # Aminosäuresequenz
                    # Stereospezifische Zuordnungen
  ../bpti.upl       # obere Distanzschranken
  ../bpti.lol       # untere Distanzschranken
bptila.aco         # Diederwinkeleinschränkungen
  *                 # Minimierungsparameter (* = Standardwerte)
                    # Standardminimierungsparameter
bptila.ang         # Diederwinkel der Ausgangsstrukturen
2.0 3*0.2 5.0 100  # Schwellen für Ausgabe von Verletzungen usw.
                    # Schwellen zum Streichen schlechter Strukturen
bptilb.ovw        # Übersichtsdatei
ynnn               # Ausgabeoptionen
bptilb.ang        # Diederwinkel
                    # kartesische Koordinaten
                    # redundante Diederwinkeleinschränkungen
EOF
cat bptilb???.ang >bptilb.ang
rm bptilb???.ang
# =====
diana << EOF
bptilc.res         # Ergebnisdatei
                    # Bausteinbibliothek
  ../bpti.seq       # Aminosäuresequenz
                    # Stereospezifische Zuordnungen
  ../bpti.upl       # obere Distanzschranken
  ../bpti.lol       # untere Distanzschranken
  ../bpti.aco       # Diederwinkeleinschränkungen
  *                 # Minimierungsparameter (* = Standardwerte)
                    # Standardminimierungsparameter
2*58 0 1000 2*100  # Diederwinkel der Ausgangsstrukturen
bptilb.ang         # Schwellen für Ausgabe von Verletzungen usw.
2.0 3*0.2 5.0 100  # Schwellen zum Streichen schlechter Strukturen
bptilc.ovw        # Übersichtsdatei
ynnn               # Ausgabeoptionen
bptilc.ang        # Diederwinkel
                    # kartesische Koordinaten
                    # Diederwinkeleinschränkungen
EOF
cat bptilc???.ang >bptilc.ang
rm bptilc???.ang

```

Die Namen der Ausgabedateien des ersten Laufs beginnen mit »bptila«, die des zweiten mit »bptilb« und die des dritten mit »bptilc«. Um die Anzahl der Ausgabedateien zu verrin-

gern, werden nach jedem Lauf die Diederwinkel aller 50 Konformere mit Hilfe der UNIX-Befehle *cat* und *rm* in einer Datei zusammengefaßt.

Eine Ausgabedatei mit redundanten Diederwinkelschränkungen, wie sie im Schritt C^(j) der REDAC-Strategie vom Programm DIANA geschrieben wird, enthält zusätzlich zu den Diederwinkelschränkungen auch Angaben über die Verteilung der Diederwinkelwerte in den lokal akzeptablen Segmenten, die zur Information des Benutzers dienen und beim Einlesen der Datei am Anfang des nächsten Laufs B^(j) als Kommentare übergangen werden. Der Teil der Datei »bpt11a.aco«, der die Aminosäurereste Leu 6 und Glu 7 betrifft, sieht folgendermaßen aus:

```
# Angle constraint output from DIANA

# Parameters:
# Maximal target function per residue: 0.400
# Minimal number of structures : 10
# Maximal width of allowed interval : 270.0 deg
# * ... there was an input angle constraint
# ** ... input and output angle constraint are the same

# Residue Angle Allowed range Struct Angle distribution [% struct. * 10]
# -180 -90 0 90 170
# | | | | |
# 6 LEU PHI -130.0 -75.0 #* 30 . 114320 . . .
# 6 LEU CHI1 -70.0 -20.0 # 30 . . 01242 . .
# 6 LEU CHI2 90.0 140.0 # 30 . . . 04410
# 6 LEU PSI -40.0 50.0 #* 30 . . 110120121 .
# 7 GLU- PHI -120.0 -20.0 #* 31 . 0221120020 . .
# 7 GLU- CHI1 -325.0 -25.0 #** 31 . 000020 0 10 . 0122 0
# 7 GLU- CHI2 -250.0 -50.0 # 31 110 00110.1 0 . . 1000011
# 7 GLU- PSI 120.0 170.0 #* 31 . . . . 13421
```

Am Anfang stehen die Kriterien für lokal akzeptable Segmente. Diese Angaben entsprechen der Antwort zur letzten Frage im interaktiven Dialog des Programms DIANA (vgl. S. 124). Es folgen die redundanten Diederwinkelschränkungen im üblichen Format (vgl. S. 78). Hinter dem Kommentarzeichen »#« steht ein Stern »*«, wenn für den betreffenden Diederwinkel bereits in der Eingabe eine Einschränkung vorlag, die aufgrund der Diederwinkelwerte in den lokal akzeptablen Segmenten restitiver gefaßt werden konnte, zwei Sterne »**« stehen, wenn die Ein- und Ausgabe die gleiche Diederwinkelschränkung enthalten. Alle anderen Diederwinkelschränkungen sind neu, d. h. ihnen entspricht keine Einschränkung in der Eingabe. Die mit »Struct« überschriebene Kolonne gibt die Anzahl Konformere an, in denen der betreffende Aminosäurerest in einem lokal akzeptablen Segment lag. Der Rest der Zeile gibt die Verteilung der Diederwinkelwerte in den lokal akzeptablen Segmenten wieder, wobei für jedes 10°-Intervall die Ziffer einem Zehntel des prozentualen Anteils der lokal akzeptablen Konformere, deren Diederwinkelwert im 10°-Intervall liegt, entspricht. In Intervallen, für die keine Ziffer steht, wurden keine erlaubten Diederwinkelwerte gefunden. Falls die Kriterien zur Erzeugung einer redundanten Diederwinkelschränkung nicht erfüllt sind, wird die entsprechende Einschränkung aus der Eingabe kopiert.

Anwendungen

Dieses Kapitel zeigt die Anwendung der Rechenverfahren, die in den vorangehenden Kapiteln eingeführt wurden, für die Lösungsstrukturbestimmung von zwei Proteinen — der *Antennapedia (C39S)* Homeodomäne (Güntert *et al.*, 1991b) und von Hirudin(1–51) (Szyperski *et al.*, 1992b, c). Außerdem werden einige interessante Aspekte der Lösungsstrukturen diskutiert. Die ebenfalls im Rahmen dieser Dissertation durchgeführten Lösungsstrukturbestimmungen des basischen pankreatischen Trypsininhibitors (Berndt *et al.*, 1992) und des Dendrotoxins K (Berndt *et al.*, 1993) werden in den genannten Veröffentlichungen behandelt.

Analyse und Vergleich von Proteinstrukturen

Dreidimensionale Proteinstrukturen werden sowohl interaktiv mit Hilfe von Molekülgrafikprogrammen wie CONFOR (Billeter, 1985; Billeter *et al.*, 1985; Xia, 1992), MIDAS (Ferrin *et al.*, 1988) oder XAM (Xia, 1992), als auch quantitativ mit Hilfe mathematischer Größen wie Abweichungen der Atompositionen (McLachlan, 1979), Atomverschiebungen (Billeter *et al.*, 1989; Billeter, 1992) und Verteilungen der Diederwinkelwerte untersucht.

Zur Berechnung der *gemittelten Struktur* eines Strukturbündels werden die kartesischen Koordinaten der einzelnen Konformere nach bestmöglicher Überlagerung (im Sinne der untenstehenden Gl. [70]) der Rückgratatome N, C α und C' der »wohlbestimmten« Aminosäurereste¹ gemittelt.

Die *Abweichung der Atompositionen*, A , zweier Konformationen $\{\dot{r}_1, \dots, \dot{r}_n\}$ und $\{\dot{p}_1, \dots, \dot{p}_n\}$, deren Schwerpunkte im Ursprung liegen, ist durch

$$A = \min_{R \in SO(3)} \sqrt{\frac{1}{n} \sum_{i=1}^n |\dot{r}_i - R\dot{p}_i|^2} \quad [70]$$

¹ Die Reste 7–59 für die *Antennapedia(C39S)* Homeodomäne; 3–30 und 37–48 für Hirudin(1–51).

gegeben. R bezeichnet eine Rotationsmatrix und $SO(3)$ die Drehgruppe. A entspricht der Wurzel aus dem quadratischen Mittel der Abstände entsprechender Atome für die bestmögliche Überlagerung der beiden Konformationen. Zur Bestimmung der optimalen Rotation R wird die 3×3 Matrix B berechnet (McLachlan, 1979):

$$B_{kl} = \frac{1}{n} \sum_{i=1}^n p_{ik} r_{il}. \quad [71]$$

p_{ik} bezeichnet die k -Komponente des Ortsvektors \vec{p}_i des Atoms i ; Entsprechendes gilt für r_{il} . Die Singulärwertzerlegung (Press *et al.*, 1986) $B = UWV^T$ stellt B als Produkt einer Matrix U mit orthonormierten Spaltenvektoren, einer Diagonalmatrix $W = \text{diag}(w_1, w_2, w_3)$ mit $w_1 \geq w_2 \geq w_3 \geq 0$ und dem Transponierten einer orthogonalen Matrix V dar. Die gesuchte optimale Rotationsmatrix R ist dann (McLachlan, 1979)

$$R = U \text{diag}(1, 1, \text{sgn det } B) V^T, \quad [72]$$

wobei $\text{sgn det } B$ das Vorzeichen der Determinante der Matrix B bezeichnet.¹

Atomverschiebungen D (*»displacements«*; Billeter *et al.*, 1989; Billeter, 1992) sind eine Verallgemeinerung von Abweichungen der Atompositionen, indem eine Menge $M_{\bar{U}}$ von Atomen optimal überlagert und anschließend die Abweichung der Atompositionen für eine andere Menge M_D von Atomen ausgewertet wird. So umfaßt zum Beispiel für die Berechnung der Atomverschiebung der Seitenkette des Aminosäurerests i nach globaler Überlagerung die Menge $M_{\bar{U}}$ die Rückgratatome N, C $^\alpha$ und C' der Reste mit wohlbestimmter Konformation und die Menge M_D die Nicht-Wasserstoffatome der Seitenkette des Rests i .

Antennapedia(C39S) Homeodomäne

Das Interesse an der Strukturbestimmung einer Mutante der *Antennapedia* Homeodomäne rührt von einer grundlegenden Frage der Entwicklungsbiologie her: Wie entsteht aufgrund des Erbguts in einer Eizelle die komplexe Struktur des Embryos? In der Fruchtfliege *Drosophila melanogaster* wurden drei Klassen von Genen gefunden, die den Körperbauplan bestimmen (Gehring, 1987). Gene aller drei Klassen enthalten ein hochkonserviertes kodierendes Fragment von etwa 180 Basenpaaren, die *Homeobox* (McGinnis *et al.*, 1983a, b; Scott & Weiner, 1984). Hochgradig homologe Homeoboxen wur-

¹ Dies stellt sicher, daß R immer eine eigentliche Drehung (mit $\text{det } R = 1$) und keine Drehspiegelung ist (mit $\text{det } R = -1$) ist.

den in vielen anderen Organismen — von Fadenwürmern bis zum Menschen — gefunden (Scott *et al.*, 1989). Sie kodieren für *Homeodomänen*, die für die DNS-Bindung des gesamten Proteins verantwortlich sind, das in einem Gen, welches eine Homeobox enthält, kodiert ist. So besitzt ein rekombinantes Polypeptid, das die 68 C-terminalen Aminosäurereste des *Antennapedia* (*Antp*) Proteins umfaßt, die Fähigkeit, spezifisch an die DNS zu binden (Müller *et al.*, 1988). In einem Projekt mit dem Ziel, den Mechanismus der Homeodomänen-DNS Wechselwirkung aufzuklären, wurde zunächst die dreidimensionale Struktur dieser rekombinanten *Antp* Homeodomäne mit Hilfe der magnetischen Kernspinresonanz in Lösung bestimmt (Otting *et al.*, 1988; Qian, 1992; Qian *et al.*, 1989; Billeter *et al.*, 1990a). Ein einzelner Cysteinrest in der Wildtyp-*Antp* Homeodomäne erfordert, daß den Proben für die Kernresonanzexperimente ein Reduktionsmittel zugesetzt werden muß. Aus diesem Grund wurde für alle folgenden Untersuchungen Cys 39 durch einen Serinrest ersetzt. Die Strukturbestimmung der *Antp(C39S)* Homeodomäne und ihr Vergleich mit der Wildtyp-*Antp* Homeodomäne bilden daher die Grundlage für die Strukturbestimmung eines *Antp(C39S)*-DNS Komplexes (Otting *et al.*, 1990; Qian, 1992; Qian *et al.*, 1993).

Die dreidimensionale Struktur der *Antp(C39S)* Homeodomäne in Lösung ist die erste Proteinstruktur, die mit Hilfe des Programms DIANA und seiner Unterstützungsprogramme bestimmt wurde (Güntert *et al.*, 1991b). Bis auf die später eingeführte Verwendung redundanter Diederwinkeleinschränkungen zur Konvergenzbeschleunigung während der DIANA-Rechnungen (vgl. S. 132ff.) und dem heute im allgemeinen nicht mehr praktizierten Einbezug von Distanzeinschränkungen für Wasserstoffbrücken (Williamson *et al.*, 1985) hat sich am Ablauf der Strukturrechnung (vgl. Bild 4) seither nichts Grundlegendes mehr verändert.

Die Strukturrechnung beruhte auf einem Datensatz von 857 relevanten (vgl. S. 109) oberen Distanzschranken aus Kern-Overhausereffekten, 50 oberen und 50 unteren Distanzschranken (Williamson *et al.*, 1985) für die 25 experimentell bestimmten Wasserstoffbrückenbindungen und 171 Diederwinkeleinschränkungen, die das Programm HABAS aufgrund von 151 experimentell gemessenen skalaren Spin-Spin-Kopplungen und lokaler Distanzschranken bestimmte. Die Programme HABAS und GLOMSA lieferten stereospezifische Zuordnungen für 35 der 120 diastereotopen Protonenpaare und alle diastereotopen Methylgruppen, sowie individuelle Zuordnungen für die Protonen von 10 der 22 NH₂-Gruppen von Asn, Gln und Arg (Güntert *et al.*, 1991b).

Weil die Konvergenzbeschleunigung mittels redundanter Diederwinkeleinschränkungen noch nicht bekannt war, mußte die Strukturrechnung mit dem Programm DIANA von 250 Startkonformeren mit zufällig gewählten Werten

für die Diederwinkel ausgehen, um ein Bündel von 20 konvergierten Konformeren zu erhalten. Auf einem Cray X-MP Hochleistungsrechner benötigte die gesamte Strukturrechnung 8,4 Std. Rechenzeit. Die 20 Konformere mit den niedrigsten Zielfunktionswerten wurden anschließend mit einer modifizierten Version des Programms AMBER (Weiner *et al.*, 1986), die Pseudoenergieterme für Distanz- und Diederwinkeleinschränkungen beinhaltet (Billeter *et al.*, 1990b), energieminiert. Die Pseudoenergieterme wurden so kalibriert, daß eine Verletzung von 0,2 Å einer Distanzeinschränkung oder 5° einer Diederwinkeleinschränkung einer Energie von $k_B T/2$ bei Raumtemperatur entspricht.

Einen Überblick über die Ergebnisse der Strukturrechnung gibt die folgende Tabelle (der »Kern« besteht aus den Rückgratatomen N, C^α, C' der Aminosäurereste 7–59 und den Schweratomen der Reste 13, 16, 26, 34, 45, 37, 38, 40, 45, 48 und 49; Billeter *et al.*, 1990a):

Analyse der 20 Antp(C39S) Konformere vor und nach der Energieminimierung

Größe	<i>vor der Energie- minimierung</i>	<i>nach der Energie- minimierung</i>
DIANA Zielfunktionswert [Å ²]	1,67 ± 0,32	(undefiniert)
AMBER Energie [kcal/Mol]	40 ± 65	-829 ± 61
Mittlere Restverletzungen oberer Distanzschranken:		
Anzahl > 0,2 Å	2,7 ± 1,3	1,9 ± 1,1
Summe [Å]	5,7 ± 0,8	11,1 ± 1,0
Maximum [Å]	0,34 ± 0,10	0,21 ± 0,01
von Diederwinkelschranken:		
Anzahl > 5°	0,8 ± 0,8	3,6 ± 1,5
Summe [°]	28,0 ± 8,1	67,9 ± 11,8
Maximum [°]	5,2 ± 1,4	6,4 ± 1,0
Mittlere Abweichungen von der gemittelten Struktur [Å]:		
Rückgratatom N, C ^α , C' 0–67	3,69 ± 1,06	3,68 ± 1,09
Rückgratatom N, C ^α , C' 7–59	0,53 ± 0,17	0,55 ± 0,18
Kern	0,56 ± 0,16	0,57 ± 0,17
Schweratome 7–59	1,22 ± 0,31	1,23 ± 0,32

Die 20 besten DIANA-Konformere erfüllen die Distanz- und Diederwinkeleinschränkungen fast perfekt und weisen eine relativ tiefe AMBER-Energie

auf — ein Zeichen dafür, daß sterisch ungünstige Überlappungen von Atomen bereits vom Programm DIANA erfolgreich beseitigt wurden. Zu jedem der 20 DIANA-Konformere fand das Programm AMBER in unmittelbarer Nähe eine Konformation mit tiefer Energie und nur leicht erhöhter Summe der Verletzungen. Die mittlere Abweichung der Atompositionen zwischen den DIANA-Konformeren und ihren entsprechenden energieminierten Konformeren ist mit 0,24 Å für die Rückgrat-atome N, C^α, C' aller Aminosäurereste viel geringer als innerhalb der Strukturbündel. Weiter zeigt die obige Tabelle, daß die Energieminimierung (mit experimentellen Distanz- und Diederwinkeleinschränkungen) weder die Lage noch die Ausdehnung des Strukturbündels im Konformationsraum ändert und daß die Seitenkettenkonformationen der Reste, die dem Lösungsmittel kaum zugänglich sind (Billeter *et al.*, 1990a) gleich präzise bestimmt sind wie der wohlbestimmte Teil des Rückgrats (Reste 7–59).

Die *Antp(C39S)* Homeodomäne ist ein α -helikales Protein (Güntert *et al.*, 1991b); die vier Helices I–IV bestehen aus den Resten 10–21, 28–38, 42–52 bzw. 53–59. Die Helix II, die Schleife 39–41 und die »Erkennungshelix« III bilden ein Helix-Schleife-Helix Motiv, das eine spezifische Bindung des Proteins an DNS ermöglicht, wobei, wie die spätere Strukturbestimmung eines *Antp(C39S)*-DNS Komplexes gezeigt hat (Otting *et al.*, 1990; Qian *et al.*, 1993), die Helix III in die große Furche der DNS zu liegen kommt. Helix IV schließt mit einem »Knick« direkt an Helix III an. Die Reste 0–6 am N- und 60–67 am C-Terminus sind in Lösung ungeordnet.



Bild 20. Stereobild des Rückgrats der 20 energieminierten DIANA-Konformere der *Antp(C39S)* Homeodomäne. Die Lage einiger Aminosäurereste ist im Einbuchstabenkode angegeben (Güntert *et al.*, 1992b).

Ein Vergleich mit der Wildtyp-*Antp* Homeodomäne (Qian *et al.*, 1989; Billeter *et al.*, 1990a) zeigt signifikante Unterschiede zwischen den beiden Lösungsstrukturen nur im Bereich von Helix IV (Güntert *et al.*, 1991b). In der folgenden Tabelle äußert sich dies darin, daß die Reste 7–52 sich viel besser

überlagern lassen als die Reste 7–59, welche die Helix IV einschließen.

Abweichungen der Atompositionen zwischen den gemittelten Lösungsstrukturen der Wildtyp-Antp und der Antp(C39S) Homeodomäne

Menge der Atome	Abweichung der Atompositionen
Rückgratátome N, C ^α , C' der Reste 0–67	2,64 Å
Rückgratátome N, C ^α , C' der Reste 7–59	1,01 Å
Kern (vgl. S. 146)	1,00 Å
Schwerátome der Reste 7–59	1,47 Å
Rückgratátome N, C ^α , C' der Reste 7–52	0,52 Å

Die Veröffentlichung von Güntert *et al.* (1991b) zeigt Bilder der dreidimensionalen Struktur der *Antp(C39S)* Homeodomäne und vertieft den Vergleich mit der Wildtyp-*Antp* Homeodomäne.

Die geringen Abweichungen der beiden Proteinstrukturen voneinander sind angesichts der konservativen Mutation eines Aminosäurerests an der Proteinoberfläche (bei der nur ein einziges Atom ausgetauscht wird) nicht überraschend. Vom methodischen Standpunkt der Strukturbestimmung aus gesehen, ist die gute Übereinstimmung der beiden Strukturen, die unabhängig voneinander und mit verschiedenen Programmen berechnet wurden¹, sehr befriedigend. Der in der Struktur des Wildtyps stärker ausgeprägte Knick zwischen den Helices III und IV dürfte keine direkte Folge der Mutation Cys39→Ser sein, sondern daher rühren, daß im Fall der *Antp(C39S)* Homeodomäne die ¹⁵N-Markierung des Proteins einen gegenüber dem Wildtyp qualitativ verbesserten Datensatz lieferte (Güntert *et al.*, 1991b). Vermutlich kommt daher der schwache Knick in der Struktur der Mutante den Verhältnissen in Lösung näher.

¹ Für die Wildtyp-*Antp* Homeodomäne wurde das Programm DISMAN (Braun & Gö, 1985) verwendet.

Hirudin(1-51)

Hirudin ist ein kleines Protein aus 65 Aminosäureresten, das in den Speicheldrüsen des Blutegels *Hirudo Medicinalis* vorkommt (Markwardt, 1970). Frühere Kernresonanzuntersuchungen (Folkers *et al.*, 1989; Haruyama & Wüthrich, 1989) haben gezeigt, daß Hirudin eine kompakte N-terminale Domäne, die die Aminosäurereste 1–48 umfaßt, und ein flexibles, nicht-strukturiertes carboxyterminales Polypeptidsegment 59–65 enthält. Hirudin ist der effektivste und gleichzeitig spezifischste bekannte Inhibitor des Enzyms Thrombin, einer Serinprotease, die eine zentrale Rolle bei der Blutgerinnung spielt (Johnson *et al.*, 1989), und hat deshalb interessante medizinische Anwendungen (Lent, 1986; Märki & Wallis, 1990).

Hirudin bildet einen sehr stabilen 1:1 Komplex mit Thrombin, der durch eine Dissoziationskonstante von $2 \cdot 10^{-14} M$ charakterisiert ist (Stone & Hofsteenge, 1986). Kürzlich wurden die Röntgenstrukturen von Thrombin (Bode *et al.*, 1989) und zweier Hirudin-Thrombin-Komplexe (Grütter *et al.*, 1990; Rydel *et al.*, 1990, 1991) gelöst. Dabei hat sich gezeigt, daß die aminoternale Domäne von Hirudin nahe dem aktiven Zentrum des Thrombins bindet, während das in wässriger Lösung nicht strukturierte C-terminale Segment im Komplex strukturiert ist und sich an die Exobindungsstelle des Fibrinogens anlagert. Weiter erlaubte der Vergleich der Röntgenstrukturen des Komplexes und des freien Thrombins die Untersuchung struktureller Änderungen des Thrombins

bei Bindung von Hirudin (Rydel *et al.*, 1991).

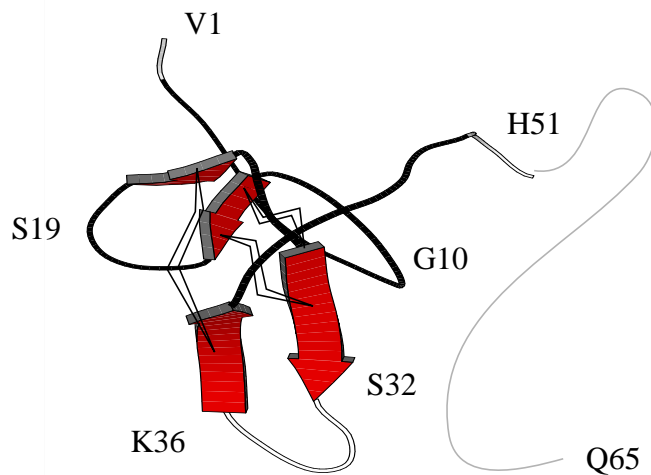


Bild 21. Schematische Darstellung von Hirudin(1-65), erzeugt mit dem Programm MOLSCRIPT (Kraulis, 1991), die für die Aminosäurereste 1-51 das Hirudin(1-51) Konformer mit dem niedrigsten DIANA-Zielfunktionswert zeigt und für die Reste 52-65 eine zufällige Konformation andeutet, da das Segment 49-65 in Lösung strukturell ungeordnet vorliegt. Die Pfeile stehen für die Stränge der β -Faltblätter, schwarze Linien markieren strukturell gut definierte Bereiche, während schlecht definierte Bereiche (Reste 1-2, 32-36 und 49-51) in weiß und die drei Disulfidbrücken 6-14, 16-28 und 22-39 durch »Blitze« angegeben sind.

Für eine vollständige Beschreibung der Protein-Protein Wechselwirkungen war eine qualitativ hochwertige Strukturbestimmung von Hirudin notwendig. Zwar wurden Lösungsstrukturen für natürliches Hirudin (Cloure *et al.*, 1987), das ein sulfatiertes Tyrosin in Position 63 enthält, rekombinantes Desulfatohirudin (Folkers *et al.*, 1989; Haruyama & Wüthrich, 1989) und die Mutante Lys47→Glu des rekombinanten Hirudins (Folkers *et al.*, 1989) bestimmt. Eine weitere Verfeinerung der Lösungsstruktur für die gesamte Polypeptidkette wurde jedoch durch Überlappungen von Signalen des flexiblen carboxyterminalen Polypeptidsegmentes mit solchen der aminoterminalen Domäne erschwert. Um eine hochaufgelöste NMR-Struktur der aminoterminalen Domäne zu erhalten, wurde deshalb Hirudin(1-51) (Dennis *et al.*, 1990) eingesetzt (Szyperski, 1992; Szyperski *et al.*, 1992b), das mit Hilfe einer CNBr-Spaltung aus der Mutante Asn52→Met des rekombinanten Hirudins erhalten wurde.

Sammlung von Konformationseinschränkungen und Strukturrechnung

Die Einzelheiten der Kernresonanzexperimente und der sequentiellen Resonanzzuordnung von Hirudin(1-51) werden von Szyperski *et al.* (1992b) beschrieben. Distanzeinschränkungen wurden aus zwei NOESY-Spektren erhalten, die mit einer Mischzeit von 50 ms in D₂O bzw. H₂O aufgenommen wurden (Szyperski *et al.*, 1992b). Insgesamt wurden 1070 Signale in den beiden

NOESY-Spektren zugeordnet, von denen 580 relevante Distanzeinschränkungen ergaben (Bild 22). Die hohe Anzahl langreichweitiger Distanzeinschränkungen ist für ein β -Faltblattprotein zu erwarten.

Die skalaren Spin-Spin-Kopplungskonstanten ${}^3J_{NH\alpha}$ wurden aus dem in H_2O aufgenommenen NOESY-Spektrum durch inverse Fourier-Transformation von In-Phase-Multipletts (Szyperski *et al.*, 1992a; vgl. S. 58ff.) erhalten. Analog wurden die ${}^3J_{\alpha\beta}$ -Kopplungskonstanten von Thr, Ile und Val aus dem NOESY-Spektrum in D_2O bestimmt. Die skalaren Kopplungskonstanten ${}^3J_{\alpha\beta}$ der übrigen Aminosäurereste wurden in einem E.COSY-Spektrum (Griesinger *et al.*, 1985) gemessen. In allen Fällen, in denen die Werte der ${}^3J_{\alpha\beta}$ -Kopplungskonstanten die Belegung mehrerer gestaffelter Rotamere implizieren (Nagayama & Wüthrich, 1981), wurden diese ${}^3J_{\alpha\beta}$ Kopplungskonstanten nicht zur Ableitung von Diederwinkeleinschränkungen genutzt und die betroffenen Distanzeinschränkungen zu den β -Protonen auf die entsprechenden Pseudoatome bezogen (Wüthrich *et al.*, 1983).

Die Programme HABAS und GLOMSA lieferten für Hirudin(1–51) stereospezifische Zuordnungen für 23 der 48 nicht-entarteten diastereotopen Protonenpaare und für 3 Isopropylgruppen sowie individuelle Zuordnungen für die Protonen von 4 NH_2 -Gruppen von Asn oder Gln.

Der endgültige Datensatz bestand aus 580 Distanzeinschränkungen aus Kern-Overhauserereffekten, 109 durch HABAS erzeugten Diederwinkeleinschränkungen (39 für ϕ , 39 für ψ und 31 für χ^1) und 9 unteren und 9 oberen Distanzschranken (Williamson *et al.*, 1985) für die drei Disulfidbrücken 6–14, 16–28 und 22–39. Es wurden keine Distanzeinschränkungen für Wasserstoff-

brückenbindungen eingebracht.

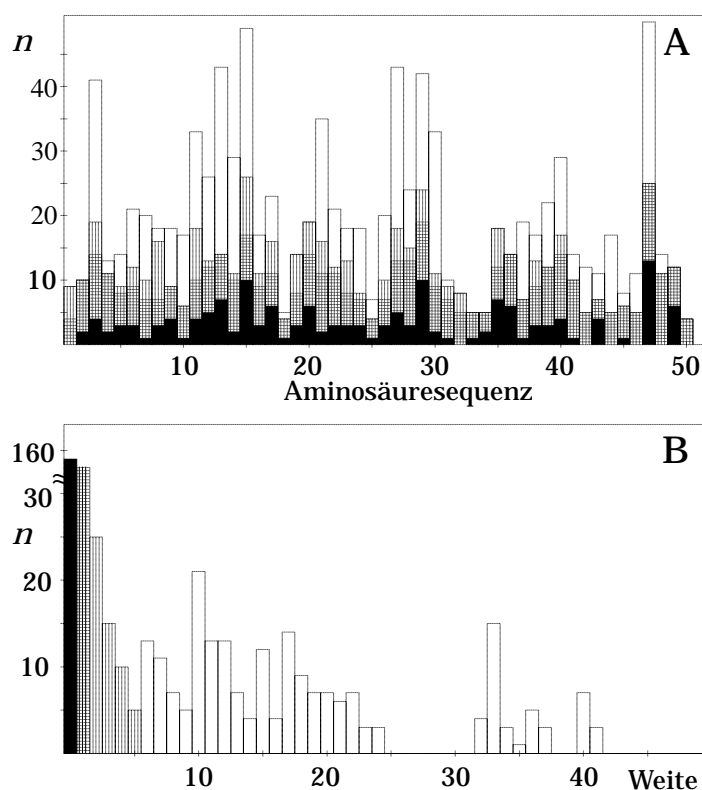


Bild 22. Grafische Darstellung der Anzahl relevanter oberer Distanz-schranken, n , für die Strukturrechnung von Hirudin(1-51):

(A) Aufgetragen gegen die Aminosäuresequenz.

(B) Aufgetragen gegen die Weite, d. h. den Abstand der beteiligten Reste entlang der Aminosäuresequenz (Die Ordinate ist oberhalb von 30 unterbrochen).

Distanzschranken der Weite 0 (innerhalb eines Rests) sind schwarz, 1 (zwischen sequentiell benachbarten Resten) kariert, 2-5 längsschraffiert und für alle größeren Weiten weiß dargestellt.

Für die Strukturrechnung mit dem Programm DIANA wurde die REDAC-Strategie zur Konvergenzbeschleunigung genau so verwendet, wie auf S. 132ff. für die Strukturrechnung der Aktivierungsdomäne B beschrieben. Lediglich die maximale Anzahl von Auswertungen der Zielfunktion wurde leicht reduziert und der maximale Wert der Zielfunktion pro Aminosäurerest im zweiten REDAC-Schritt, $C^{(2)}$ (vgl. Bild 16 auf S. 133), wurde auf $0,1 \text{ \AA}^2$ gesetzt. 49 von 50 Strukturrechnungen konvergierten mit einem Endwert der Zielfunktion von weniger als $0,9 \text{ \AA}^2$. Die vollständige DIANA-Rechnung für 50 Startkonformere benötigte 27,4 Minuten Rechenzeit auf einem Cray Y-MP Hochleistungsrechner. Die besten 20 DIANA-Konformere besaßen sehr niedrige verbleibende Verletzungen der Distanz- und Diederwinkelschranken. Anschließend wurden diese 20 Konformere einer eingeschränkten Energieminimierung mit einer modifizierten Version des Programms AMBER unterworfen, wie auf S. 145 für die *Antennapedia(C39S)* Homeodomäne beschrieben. Aus der folgenden Tabelle ergeben sich die gleichen Schlußfolgerungen wie für die *Anten-*

napedia(C39S) Homeodomäne (vgl. S. 146):

Analyse der 20 *Hirudin(1-51)* Konformere vor und nach der Energieminimierung

Größe	vor der Energie- minimierung	nach der Energie- minimierung
DIANA Zielfunktionswert [\AA^2]	$0,12 \pm 0,03$	(undefiniert)
AMBER Energie [kcal/Mol]	-197 ± 45	-654 ± 40
Mittlere Restverletzungen oberer Distanzschranken:		
Anzahl $> 0,2 \text{ \AA}$	0	0
Summe [\AA]	$1,0 \pm 0,2$	$2,7 \pm 0,2$
Maximum [\AA]	$0,09 \pm 0,02$	$0,09 \pm 0,01$
von Diederwinkelschranken:		
Anzahl $> 5^\circ$	0	0
Summe [$^\circ$]	$3,2 \pm 0,7$	$17,5 \pm 3,4$
Maximum [$^\circ$]	$0,9 \pm 0,1$	$2,5 \pm 0,4$

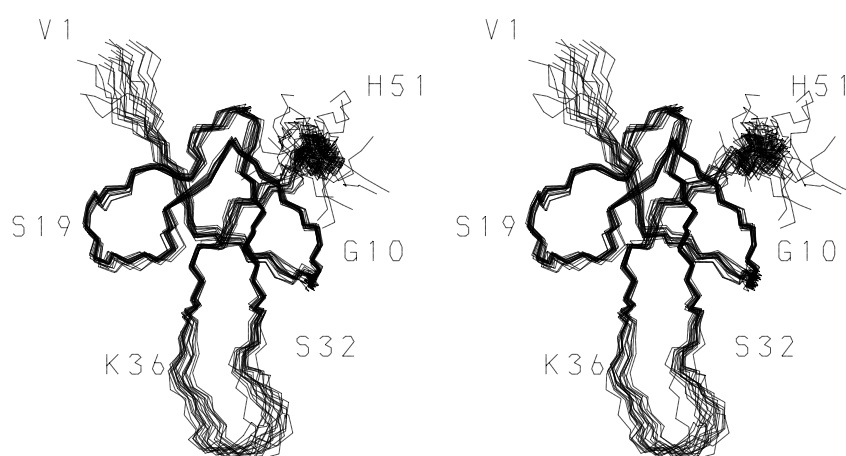


Bild 23. Stereobild des Rückgrats der 20 energieminimierten DIANA-Konformere von *Hirudin(1-51)*. Die Lage einiger Aminosäurereste ist im Einbuchstabenkode angegeben (Szyperski et al., 1992b).

Die dreidimensionale Struktur von *Hirudin(1-51)*

Bild 23 zeigt das Rückgrat der 20 energieminierten DIANA-Konformere von Hirudin(1–51) nach Überlagerung der Rückgratatompositionen N, C^α und C' der Aminosäurereste 3–30 und 37–48.

Die niedrigen Abweichungen der Atompositionen für verschiedene Teile des Proteins sind charakteristisch für eine qualitativ hochwertige Lösungsstruktur:

Mittlere Abweichungen der Atompositionen der 20 energieminierten Hirudin (1–51) Konformere von der gemittelten Struktur

Menge der Atome	Abweichung der Atompositionen
Rückgratatompositionen (N, C ^α , C') der Reste 1–51	0,90 ± 0,19 Å
Alle Schweratome der Reste 1–51	1,40 ± 0,19 Å
Rückgratatompositionen der Reste 3–30 und 37–48	0,38 ± 0,07 Å
Rückgrat 3–30, 37–48 und Schweratome von 19 »gut bestimmten« Seitenketten (Reste 4–7, 9, 12, 14–16, 19, 20, 22, 26, 28, 29, 39, 40 und 46)	0,48 ± 0,07 Å
Alle Schweratome der Reste 3–30 und 37–48	0,77 ± 0,07 Å

Die Atomverschiebungen nach globaler Überlagerung, D_g , die lokalen Abweichungen der Atompositionen A_l des Rückgrats von Hirudin(1–51) in Bild 24A und B sowie die Bereiche der Diederwinkel ϕ und ψ des Rückgrats in Bild 25 erlauben eine quantitative Beurteilung der Präzision der Beschreibung der lokalen Konformation entlang der Polypeptidkette, wie sie mit Hilfe der Kernresonanzdaten möglich ist.

Das aminoternale Tripeptidsegment ist lokal recht gut bestimmt — die Rückgratdiederwinkel liegen im Bereich, der für β -Faltblätter typisch ist. Es folgt ein gut definiertes Hexapeptidsegment mit nicht regulärer Sekundärstruktur. Eine Schleife des Typs II (Chou & Fasman, 1977) ist dann über einen Knick (Reste 12 und 13) mit einem Strang des ersten β -Faltblattes verbunden. Die beiden Stränge dieses β -Faltblattes sind über eine Schleife des Typs II', die von den Aminosäureresten 17–20 gebildet wird, miteinander verbunden. Eine weitere Schleife vom Typ II, bestehend aus den Aminosäureresten 23–26, verbindet die beiden β -Faltblätter miteinander. Die lokale Konformation der drei Schleifen 8–11, 17–20 und 23–26 ist durch die Kernresonanzdaten gut bestimmt — unter anderem deshalb, weil für die α -Protonen aller Glycine in diesen Schleifen stereospezifische Zuordnungen erhalten wurden. Im Vergleich

dazu ist die Konformation der Schleife 31–36, die die beiden Stränge des zweiten β -Faltblattes verbindet, deutlich weniger gut bestimmt. Das Polypeptid

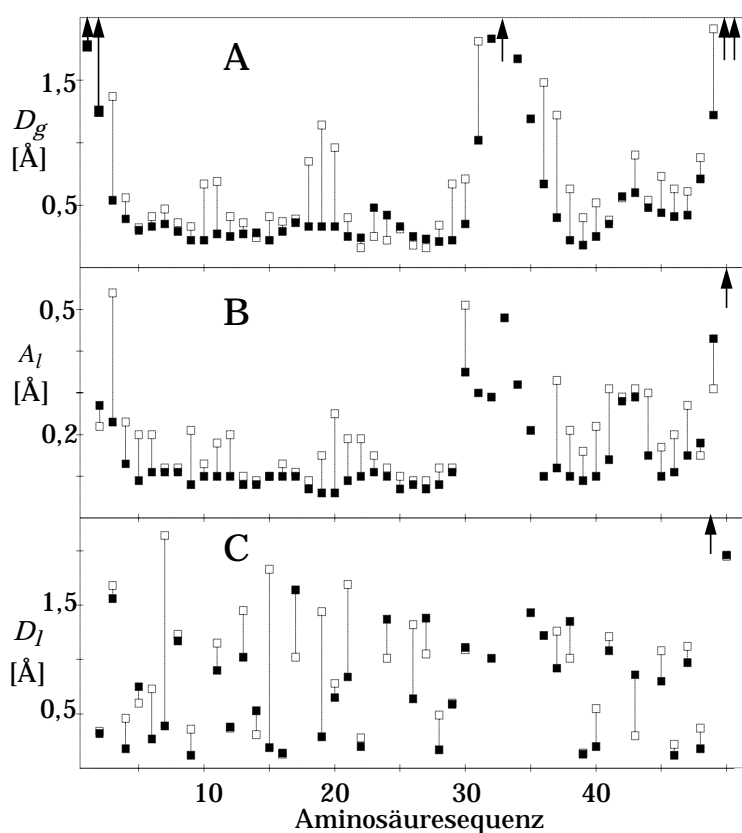


Bild 24. Die folgenden Größen sind gegen die Aminosäuresequenz von Hirudin(1–51) aufgetragen, wobei schwarze Quadrate das Mittel für die 20 energieminierten DIANA-Konformere relativ zur gemittelten Lösungsstruktur; offene Quadrate die Werte für Hirudin im Hirudin-Thrombin-Komplex relativ zur gemittelten Lösungsstruktur angeben: (A) Atomverschiebungen D_g des Rückgrats nach globaler Überlagerung des Rückgrats der Reste 3–30 und 37–48. (B) Lokale Abweichungen der Atompositionen, A_l , für die Rückgratatome N, C^α und C' von Tripeptiden. (C) Atomverschiebungen D_l der Seitenketten nach lokaler Überlagerung des Rückgrats von Tripeptiden. Pfeile stehen für Werte außerhalb der dargestellten Bereiche.

segment 41–45, das auf das zweite β -Faltblatt folgt, zeigt leicht erhöhte Atomverschiebungen, während die Aminosäurereste 46–48, die eine kurze Polyprolinhelix II (Sasisekharan, 1959; Yonath & Traub, 1969) bilden, durch die Kernresonanzdaten gut bestimmt sind. Im Einklang mit früheren Untersuchungen des intakten Hirudins (Folkers *et al.*, 1987; Haruyama & Wüthrich, 1989), erweisen sich die carboxyterminalen Reste 49–51 als nahezu völlig unstrukturiert. Etwas willkürlich lassen sich 20 »gut definierte« Aminosäurereste mit lokalen Atomverschiebungen der Seitenkettenatome D_l unter 0,75 Å festlegen (4–7, 9, 12, 14–16, 19, 20, 22, 26, 28, 29, 39, 40 und 46). Für sämtliche β -Methylenprotonen und Isopropyl-Methylgruppen dieser Aminosäurereste konnten stereospezifische Resonanzzuordnungen erhalten werden. Die Konformation der Seitenketten, Wasserstoffbrücken sowie Vergleiche mit früheren Lösungsstrukturen des intakten Hirudins(1–65) werden von Szyperski *et al.* (1992b) ausführlich behandelt und zeigen, daß die dreidimensionalen Strukturen von Hirudin(1–51) und dem entsprechenden Segment im intakten Hirudin sind weitgehend identisch sind. Dies ermöglicht den Vergleich von Hirudin(1–

51) mit dem entsprechenden Polypeptidfragment im Hirudin-Thrombin-Komplex und somit eine detaillierte Untersuchung der Protein-Protein Wechselwirkungen im Komplex (Szyperski *et al.*, 1992c).

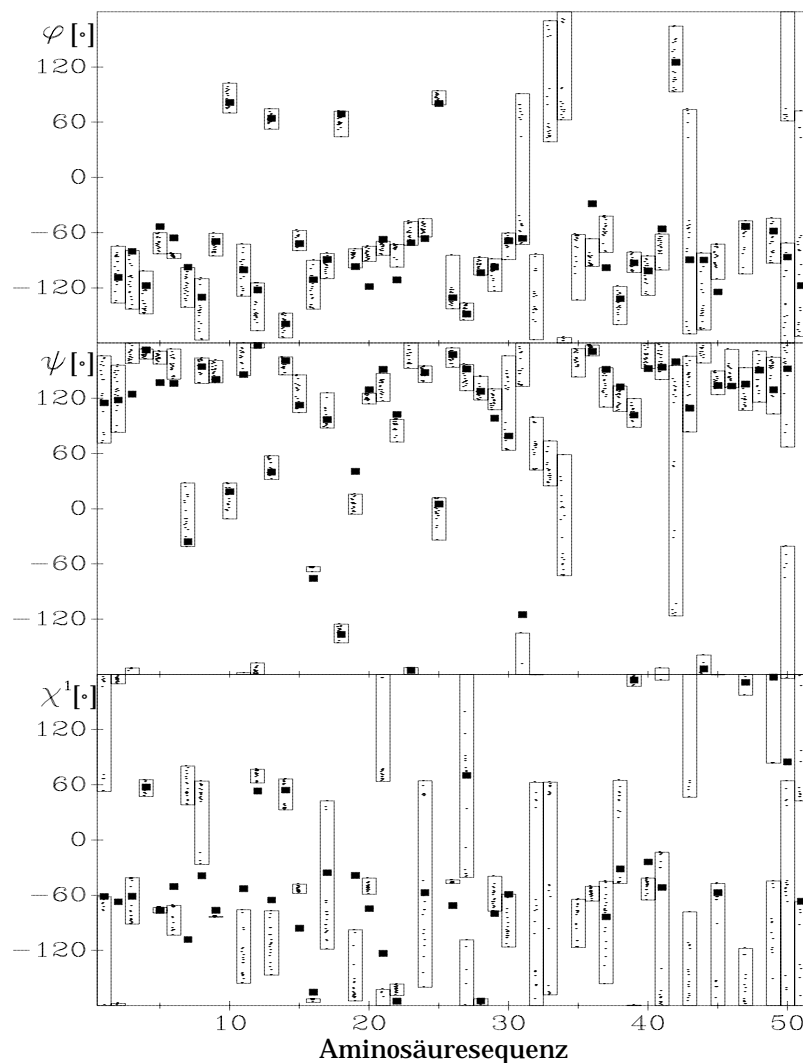


Bild 25. Werte der Die-
derwinkel ϕ , ψ und χ^1
von Hirudin(1-51) in der
Kristallstruktur des
Hirudin-Thrombin-
Komplexes (schwarze
Quadrate) und für die 20
energieminimierten DIA-
NA-Konformere von
Hirudin(1-51) (Punkte,
die von offenen Balken
umschlossen sind).

Vergleich der NMR-Struktur von Hirudin(1-51) mit dem entsprechenden Poly- peptidfragment im Hirudin-Thrombin-Komplex

Der niedrige Abweichung der Atompositionen von 0,61 Å für die Rückgratato-
me N, C $^\alpha$ und C' der Aminosäurereste 3–30 und 37–48 von Hirudin(1–51) in Lö-
sung und im Hirudin-Thrombin-Komplex bedeutet, daß sich die Konformation
des Rückgrats dieser Aminosäurereste bei der Bindung an Thrombin nur we-

nig ändert:

Abweichungen der Atompositionen zwischen der gemittelten Lösungsstrukturen von Hirudin(1-51) und dem entsprechenden Polypeptidfragment im Hirudin-Thrombin-Komplex

Menge der Atome	Abweichung der Atompositionen
Rückgratatome der Reste 3-30 und 37-48	0,61 Å
Rückgrat 3-30, 37-48 und Schweratome von 19 »gut bestimmten« Seitenketten (Reste 4-7, 9, 12, 14-16, 19, 20, 22, 26, 28, 29, 39, 40 und 46)	0,84 Å
Alle Schweratome der Reste 3-30 und 37-48	0,91 Å

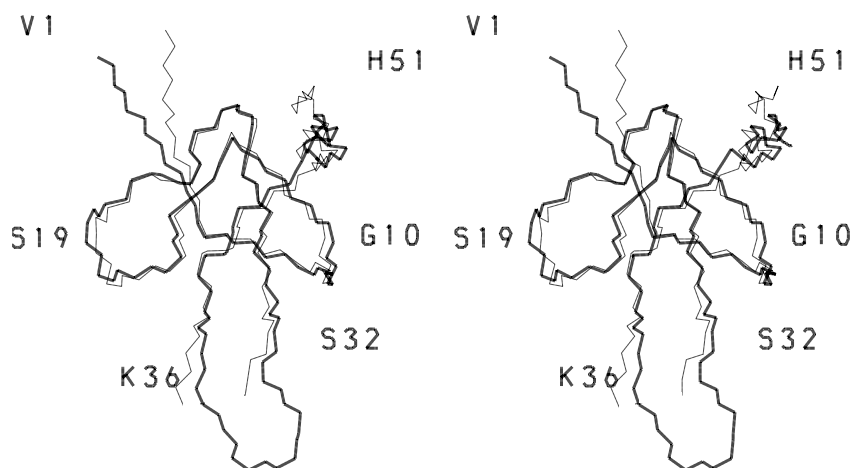


Bild 26. Stereodarstellung des Rückgrats der mittleren NMR-Struktur von Hirudin(1-51) (fett) und des entsprechenden Segments von rekombinantem Hirudin in der Röntgenstruktur des Hirudin-Thrombin-Komplexes von Rydel *et al.* (1991) (dünn). In der Kristallstruktur fehlen die Koordinaten der Aminosäurereste 32-35.

Hinzu kommt, daß die Aminosäurereste 32-35, deren Konformation die Kernresonanzdaten nicht gut bestimmen, in der Kristallstruktur des Komplexes keine Elektronendichte aufweisen (Rydel *et al.*, 1990). Die globalen Atomverschiebungen, D_g , der Rückgratatome zeigen, daß beide Strukturen für die Aminosäurereste 4-9, 12-17, 22-28 und 38-48 fast identische Rückgratkonformationen besitzen, während für die Reste 1-3, 10-11, 17-21, 29-31 und 36-37 Unterschiede auftreten (Bild 24A). Der Vergleich der Rückgratdiederwinkel ϕ und ψ (Bild 25) zeigt, daß die minimale Abweichung einer der 20 energieminierten DIANA-Konformere von der Röntgenstruktur nur für die ϕ -Winkel von Asn 20 und Lys 36, sowie die ψ -Winkel von Tyr 3 und Ser 19 größer als 20° ist.

Auch die Konformationen vieler Seitenketten ist in Lösung und im Komplex sehr ähnlich (Szyperski *et al.*, 1992b). Die Seitenketten von Thr 7, Leu 15, Ser 19, Val 21 und Asn 26 sind in beiden Strukturen gut bestimmt und der χ^1 -Winkel der Röntgenstruktur liegt klar außerhalb des Bereichs, der von den 20 DIA-NA-Konformeren abgedeckt wird (Bild 25). Die Überlagerung der gemittelten Lösungsstruktur von Hirudin(1–51) mit dem entsprechenden Polypeptidfragment im Hirudin-Thrombin-Komplex in Bild 26 zeigt die große Ähnlichkeit der beiden Strukturen. Erkennbar sind jedoch auch die bereits oben angesprochenen Unterschiede der Segmente 1–3 und 17–21 (Szyperski *et al.*, 1992c).

Die strukturellen Unterschiede lassen sich durch Protein-Protein Kontakte zwischen Hirudin und Thrombin sowie zwischen Hirudin und Hirudin-Thrombin-Komplexen in anderen asymmetrischen Einheiten im Kristall erklären oder sind eine Folge der unterschiedlichen Aminosäuresequenz der beiden Hirudinvarianten¹. Für eine ausführliche Diskussion siehe Szyperski *et al.*, 1992c).

¹ Die Hirudinvariante 2 in der Kristallstruktur unterscheidet sich von der Variante 1 in Lösung im Bereich der Reste 1–51 durch die 5 Aminosäuresubstitutionen Val 1 → Ile, Val 2 → Thr, Asp 33 → Asn, Glu 35 → Lys und Lys 36 → Gly.

Literaturverzeichnis

- Abe, H., Braun, W., Noguti, T. & Gō, N. (1984). Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles in proteins. General recurrent equations. *Computers & Chemistry*, **8**, 239–247.
- Abragam, A. (1961). *Principles of Nuclear Magnetism*, Clarendon Press, Oxford.
- Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*, Clarendon Press, Oxford.
- Anil-Kumar, Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Comm.* **95**, 1–6.
- Anil-Kumar, Wagner, G., Ernst, R. R. & Wüthrich, K. (1981). Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: implications for studies of protein conformation. *J. Amer. Chem. Soc.* **103**, 3654–3658.
- Antuch, W., Berndt, K. D., Chávez, M. A., Delfin, J. & Wüthrich, K. (1993). The NMR solution structure of a Kunitz-type proteinase inhibitor from the sea anemone *Sichodactyla helianthus*. *Eur. J. Biochem.* **212**, 675–684.
- Arseniev, A., Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R. & Wüthrich, K. (1988). Three-dimensional structure of rabbit liver [Cd₇]-metallothionein-2a in aqueous solution determined by nuclear magnetic resonance. *J. Mol. Biol.* **201**, 637–657.
- Barkhuijsen, H., De Beer, R. & van Ormondt, D. (1987). Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *J. Magn. Reson.* **73**, 553–557.
- Bax, A., Ikura, M., Kay, L. E. & Zhu, G. (1991). Removal of F_1 baseline distortion and optimization of folding in multidimensional NMR spectra. *J. Magn. Reson.* **91**, 174–178.
- Berndt, K. D., Güntert, P., Orbons, L. P. M. & Wüthrich, K. (1992). Determination of a high-quality NMR solution structure of the bovine pancreatic trypsin inhibitor (BPTI) and comparison with three crystal structures. *J. Mol. Biol.* **227**, 757–775.
- Berndt, K. D., Güntert, P. & Wüthrich, K. (1993). The NMR solution structure of dendrotoxin K from the venom of *Dendroaspis polylepis polylepis*. *J. Mol. Biol.*, im Druck.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
- Billeter, M. (1985). *Strukturbestimmung von Polypeptiden und kleinen Proteinen auf der Basis von ¹H-NMR Daten mittels Computergraphik und Optimierung*, Diss. ETH Nr. 7723, Zürich.
- Billeter, M. (1992). Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals. *Q. Rev. Biophys.* **25**, 325–377.
- Billeter, M., Braun, W. & Wüthrich, K. (1982). Sequential resonance assignments in protein ¹H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J. Mol. Biol.* **155**, 321–346.

- Billeter, M., Engeli, M. & Wüthrich, K. (1985). Interactive program for investigation of protein structures based on ^1H NMR experiments. *J. Mol. Graph.* **3**, 79–83, 97–98.
- Billeter, M., Havel, T. F. & Wüthrich, K. (1987). The ellipsoid algorithm as a method for the determination of polypeptide conformations from experimental distance constraints and energy minimization. *J. Comp. Chem.* **8**, 132–141.
- Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). Comparison of the high-resolution structures of the α -amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. Mol. Biol.* **206**, 677–687.
- Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990a). Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ^1H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **214**, 183–197.
- Billeter, M., Schaumann, T., Braun, W. & Wüthrich, K. (1990b). Restrained energy refinement with two different algorithms and force fields of the structure of the α -amylase inhibitor tendamistat determined by NMR in solution. *Biopolymers*, **29**, 695–706.
- Billeter, M., Neri, D., Otting, G. & Wüthrich, K. (1992a). Precise vicinal coupling constants $^3J_{\text{HN}\alpha}$ in proteins from nonlinear fits of J -modulated [^{15}N , ^1H]-COSY experiments. *J. Biomol. NMR*, **2**, 257–274.
- Bloch, F. (1946). Nuclear induction. *Phys. Rev.* **70**, 460–474.
- Blumenthal, L. M. (1970). *Theory and Applications of Distance Geometry*, Chelsea, New York.
- Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 Å crystal structure of human α -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467–3475.
- Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–188.
- Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1989). Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer. *J. Magn. Reson.* **82**, 290–308.
- Braun, W. (1987). Distance geometry and related methods for protein structure determination from NMR data. *Q. Rev. Biophys.* **19**, 115–157.
- Braun, W. & Gö, N. (1985). Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J. Mol. Biol.* **186**, 611–626.
- Braun, W., Bösch, C., Brown, L. R., Gö, N. & Wüthrich, K. (1981). Combined use of proton-proton overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. application to micelle-bound glucagon. *Biochim. Biophys. Acta*, **667**, 377–396.
- Braun, W., Wider, G., Lee, K. H. & Wüthrich, K. (1983). Conformation of glucagon in a lipid-water interphase by ^1H nuclear magnetic resonance. *J. Mol. Biol.* **169**, 921–948.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.
- Brooks, C. L., Karplus, M. & Pettitt, B. M. (1988). *Proteins: A theoretical perspective of dynamics, structure, and thermodynamics*, Wiley, New York.
- Brown, D. E., Campbell, T. W. & Moore, R. N. (1989). Automated phase correction of FT NMR spectra by baseline optimization. *J. Magn. Reson.* **85**, 15–23.
- Brown, L. R. & Wüthrich, K. (1992). Nuclear magnetic resonance solution structure of the α -neurotoxin from the black mamba (*Dendroaspis polylepis polylepis*). *J. Mol. Biol.* **227**, 1118–1135.
- Brown, L. R., Mronga, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Wüthrich, K. (1993). The NMR solution structure of the pheromone Er-10 from the ciliated protozoan *Euplotes raikovi*. *J. Mol. Biol.*, im Druck.

- Bruker Analytische Meßtechnik GmbH (1991). *UXNMR, Messen und Verarbeiten von NMR Daten*, Rheinstetten, Deutschland.
- Brünger, A. T. (1992). *X-PLOR version 3.0, user manual*, Yale University, New Haven.
- Brünger, A. T., Clore, G. M., Gronenborn, A. M. & Karplus, M. (1986). Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3801–3805.
- Cavanagh, J. & Rance, M. (1990). Sensitivity improvement in isotropic mixing (TOCSY) sequences. *J. Magn. Reson.* **88**, 72–85.
- Cieslar, C., Clore, G. M. & Gronenborn, A. M. (1988). Automatic phasing of pure phase absorption two-dimensional NMR spectra. *J. Magn. Reson.* **79**, 154–157.
- Clore, G. M. & Gronenborn, A. M. (1991). Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. NMR Spectrosc.* **23**, 43–92.
- Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1985). Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*. Combined use of ^1H nuclear magnetic resonance and restrained molecular dynamics. *J. Mol. Biol.* **186**, 435–455.
- Clore, G. M., Sukumaran, D. K., Nilges, M., Zarbock, J. & Gronenborn, A. M. (1987). The conformation of hirudin in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **6**, 529–539.
- Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. (1991) Four-dimensional $^{13}\text{C}/^{13}\text{C}$ -edited nuclear Overhauser enhancement spectroscopy of a protein in solution: Application to Interleukin- 1β . *Biochemistry*, **30**, 12–18.
- Chou, P. Y. & Fasman, G. D. (1977). β -turns in proteins. *J. Mol. Biol.* **115**, 135–175.
- Crippen, G. M. (1977). A novel approach to the calculation of conformation: distance geometry. *J. Comp. Phys.* **24**, 96–107.
- Deisenhofer, J. & Steigemann, W. (1975). Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta. Crystallogr. B*, **31**, 238–250.
- DeMarco, A. & Wüthrich, K. (1976) Digital filtering with a sinusoidal window function: An alternative technique for resolution enhancement in FT NMR. *J. Magn. Reson.* **24**, 201–204.
- DeMarco, A., Llinás, M. & Wüthrich, K. (1978) Analysis of the ^1H -NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers*, **17**, 617–636.
- Denk, W., Baumann, R. & Wagner, G. (1986). Quantitative evaluation of cross-peak intensities by projection of two-dimensional NOE spectra on a linear space spanned by a set of reference resonance lines. *J. Magn. Reson.* **67**, 386–390.
- Dennis, S., Wallace, A., Hofsteenge, J. & Stone, S. R. (1990). Use of fragments of hirudin to investigate thrombin-hirudin interaction. *Eur. J. Biochem.* **188**, 61–66.
- Dickerson, R. E. & Geis, I. (1971). *Struktur und Funktion der Proteine*, Verlag Chemie, Weinheim.
- Dietrich, W., Rüdell, C. H. & Neumann, M. (1991). Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. *J. Magn. Reson.* **91**, 1–11.
- Dobson, C. M., Olejniczak, E. T., Poulson, F. M. & Ratcliffe, R. G. (1982). Time development of proton nuclear Overhauser effects in proteins. *J. Magn. Reson.* **48**, 97–110.
- Driscoll, P. C., Gronenborn, A. M., Beress, L. & Clore, G. M. (1989). Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry*, **28**, 2188–2198.
- Dubs, A., Wagner, G. & Wüthrich, K. (1979). Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. *Biochim. Biophys. Acta*, **577**, 177–194.

- Eccles, C. (1990). *EASY manual. December 90*, Institut für Molekularbiologie und Biophysik, ETH Zürich.
- Eccles, C., Güntert, P., Billeter, M. & Wüthrich, K. (1991). Efficient analysis of protein 2D NMR spectra using the software package EASY. *J. Biomol. NMR*, **1**, 111–130.
- Ernst, R. R. (1969). Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy. *J. Magn. Reson.* **1**, 7–26.
- Ernst, R. R. (1992). Kernresonanz-Fourier-Transformations-Spektroskopie. *Angew. Chem.* **104**, 817–836.
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987). *The principles of nuclear magnetic resonance in one and two dimensions*, Clarendon, Oxford.
- Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13–27.
- Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Kohler, S., & Gronenborn, A. M. (1989). Solution structure of recombinant hirudin and the Lys47-Glu mutant: a nuclear magnetic resonance a hybrid geometry-dynamical simulated annealing study. *Biochemistry*, **28**, 2601–2617.
- Freeman, R. (1988). *A handbook of nuclear magnetic resonance*, Longman, Harlow, Vereinigtes Königreich.
- Freiherr von Freyberg-Eisenberg, B. (1993). *Studium der Mobilität von Polypeptid- und Proteinstrukturen mittels Monte-Carlo-Verfahren und unter Berücksichtigung von Solvations-effekten*, Diss. ETH Nr. 10031, Zürich.
- Friedrichs, M. S., Metzler, W. J. & Mueller, L. (1991). Removal of diagonal peaks in two-dimensional NMR spectra by means of digital filtering. *J. Magn. Reson.* **95**, 178–183.
- Gehring, W. J. (1987). Homeo boxes in the study of development. *Science*, **236**, 1245–1252.
- Gladden, L. F. & Elliott, S. R. (1986). A numerical phasing technique for application to one-dimensional NMR spectra. *J. Magn. Reson.* **68**, 383–388.
- Glusker, J. P. & Trueblood, K. N. (1985). *Crystal Structure Analysis*, Oxford University Press, New York.
- Gordon, S. L. & Wüthrich, K. (1978). Transient proton-proton Overhauser effects in horse ferrocycytochrome C. *J. Amer. Chem. Soc.* **100**, 7094–7096.
- Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1985). Two-dimensional correlation of connected NMR transitions. *J. Amer. Chem. Soc.* **107**, 6394–6396.
- Grütter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. & Stone, S. R. (1990). Crystal structure of the thrombin-hirudin complex: a novel mode of serine protease inhibition. *EMBO J.* **9**, 2361–2365.
- Güntert, P. (1987). *Modellstudien zur Ermittlung von Proteinstrukturen mittels NMR und Distanzgeometrierechnungen: Der Einfluß stereospezifischer Zuordnungen*, Diplomarbeit, Abteilung für Mathematik und Physik, ETH Zürich.
- Güntert, P. (1992). *DIANA program version 2.1. User's manual and instructions*, Institut für Molekularbiologie und Biophysik, ETH Zürich.
- Güntert, P. & Wüthrich K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR*, **1**, 446–456.
- Güntert, P. & Wüthrich, K. (1992). A new procedure for high-quality baseline correction of two- and higher-dimensional NMR spectra. *J. Magn. Reson.* **96**, 403–407.
- Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). Automated stereospecific ¹H NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Amer. Chem. Soc.* **111**, 3997–4004.
- Güntert, P., Braun, W. & Wüthrich, K. (1991a). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.

- Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich K. (1991b). Structure determination of the *Antp(C39→S)* homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 531–540.
- Güntert, P., Dötsch, V., Wider, G. & Wüthrich K. (1992). Processing of multi-dimensional NMR data with the new software PROSA. *J. Biomol. NMR*, **2**, 619–629.
- Güntert, P., Schaefer, N., Otting, G. & Wüthrich K. (1993). POMA, a complete *Mathematica* implementation of the NMR product operator formalism. *J. Magn. Reson. A*, **101**, 103–105.
- Hare Research, Inc. (1992). *FELIX version 2.1*, Bothell, Washington, USA.
- Haruyama, H. & Wüthrich, K. (1989). Conformation of recombinant desulfatohirudin in aqueous solution determined by nuclear magnetic resonance. *Biochemistry*, **28**, 4301–4312.
- *Havel, T. F. (1991). An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog. Biophys. mol. Biol.* **56**, 43–78.
- *Havel, T. F. & Wüthrich, K. (1984). A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^1H - ^1H proximities in solution. *Bull. Math. Biol.* **46**, 673–698.
- *Havel, T. F. & Wüthrich, K. (1985). An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J. Mol. Biol.* **182**, 281–294.
- Hendrickson, W. A. & Wüthrich, K. (ed.) (1991). *Macromolecular structures 1991. Atomic Structures of biological macromolecules reported during 1990*, Current Biology, London.
- Hendrickson, W. A. & Wüthrich, K. (ed.) (1992). *Macromolecular structures 1992. Atomic Structures of biological macromolecules reported during 1991*, Current Biology, London.
- Henrichs, P. M., Hewitt, J. M. & Young, R. H. (1986). Baseline roll and phase distortion of NMR spectra. An algorithm for the generation of reliable powder spectra of solids. *J. Magn. Reson.* **69**, 460–466.
- Heuer, A. (1991). A new algorithm for automatic phase correction by symmetrizing lines. *J. Magn. Reson.* **91**, 241–253.
- Hockney, R. W. & Eastwood, J. W. (1981). *Computer Simulations Using Particles*, McGraw-Hill, New York.
- Hoffman, R. E., Delaglio, F. & Levy, G. C. (1992). Phase correction of two-dimensional NMR spectra using DISPA. *J. Magn. Reson.* **98**, 231–237.
- Hoult, D. I., Chen, C.-N., Eden, H. & Eden, M. (1983). Elimination of baseline artifacts in spectra and their integral. *J. Magn. Reson.* **51**, 110–117.
- Hyberts, S. G., Märki, W. & Wagner, G. (1987). Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. *Eur. J. Biochem.* **164**, 625–635.
- Ikura, T., Gō, N. & Inagaki, F. (1991). Refined structure of melittin bound to perdeuterated dodecylphosphocholine micelles as studied by 2D-NMR and distance geometry calculation. *Proteins*, **9**, 81–89.
- IUPAC-IUB Commission on Biochemical Nomenclature (1970). Abbreviations and symbols for the description of polypeptide chains. *J. Mol. Biol.* **52**, 1–17.
- Johnson, P. H., Sze, P., Winant, R., Payne, P. W. & Lazar, J. B. (1989). Biochemistry and genetic engineering of hirudin. *Sem. Thromb. Hemost.* **15**, 302–315.
- Kalk, A. & Berendsen, H. J. C. (1976). Proton magnetic relaxation and spin diffusion in proteins. *J. Magn. Reson.* **24**, 343–366.
- Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. Lac repressor headpiece. *J. Mol. Biol.* **182**, 179–182.
- Karplus, M. (1963). Vicinal proton coupling in nuclear magnetic resonance. *J. Amer. Chem. Soc.* **85**, 2870–2871.

- Keepers, J. W. & James, T. L. (1984). A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J. Magn. Reson.* **57**, 404–426.
- Keller, R. M. & Wüthrich, K. (1980). Structural study of the heme crevice in cytochrome b based on individual assignments of the ^1H -NMR lines of the heme group and selected amino acid residues. *Biochim. Biophys. Acta*, **621**, 204–217.
- Kendrew, J. C. (1963). Myoglobin and the structure of proteins. Crystallographic analysis and data processing techniques reveal the molecular architecture. *Science* **139**, 1259–1266.
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of myoglobin. *Nature* **185**, 422–427.
- Kessler, H., Gehrke, M. & Griesinger, C. (1988) Zweidimensionale NMR-Spektroskopie, Grundlagen und Übersicht über die Experimente. *Angew. Chem.* **100**, 507–554.
- Kießling, I. & Lowes, M. (1987). *Programmierung mit FORTRAN-77*, Teubner, Stuttgart.
- Kline, A. D., Braun, W. & Wüthrich, K. (1988). Determination of the complete three-dimensional structure of the α -amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **204**, 675–724.
- Kohda, D., Gō, N., Hayashi, K. & Inagaki, F. (1988). Tertiary structure of mouse epidermal growth factor determined by two-dimensional ^1H NMR. *J. Biochem.* **103**, 741–743.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
- Kumaresan, R. & Tufts, D. W. (1982). Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP-30**, 833–840.
- Kuntz, I. D., Crippen, G. M., Kollman, P. A. & Kimmelman, D. (1976). Calculation of protein tertiary structure. *J. Mol. Biol.* **106**, 983–994.
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, **245**, 635–637.
- Lent, C. (1986). New medical and scientific uses of the leech. *Nature*, **323**, 494.
- Li, Z. & Scheraga, H. A. (1987). Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci., U.S.A.* **84**, 6611–6615.
- Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance in macromolecules. 1. Theory and range of validity. *J. Amer. Chem. Soc.* **104**, 4546–4559.
- Luginbühl, P. (1991). *The influence of internal motion in proteins on nuclear magnetic relaxation*. Diplomarbeit, Abteilung für Mathematik und Physik, ETH Zürich.
- Macura, S. & Ernst, R. R. (1980). Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Mol. Phys.* **41**, 95–117.
- Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H - ^1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Comm.* **113**, 967–974.
- Marion, D. & Bax, A. (1989). Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time domain data. *J. Magn. Reson.* **83**, 205–211.
- Marion, D., Ikura, K. & Bax, A. (1989). Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J. Magn. Reson.* **84**, 425–430.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.
- Märki, W. E. & Wallis, R. B. (1990). The anticoagulant and antithrombotic properties of hirudins. *Thromb. Haemostasis*, **64**, 344–348.
- Marple, S. L. (1987). *Digital Spectral Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- Markwardt, F. (1970). Hirudin as an inhibitor of thrombin. *Meth. Enzymology*, **19**, 924–932.

- Marquardt, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. B*, **39**, 480–490.
- McCammon, J. A. & Harvey, S. C. (1987). *Dynamics of proteins and nucleic acids*, Cambridge University Press, Cambridge.
- McConnell, J. (1987). *The theory of nuclear magnetic relaxation in liquids*, Cambridge University Press, Cambridge.
- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. (1984a). A conserved DNA sequence in homeotic genes of the *Drosophila antennapedia* and *bithorax* complexes. *Nature*, **308**, 428–433.
- McGinnis, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. (1984b). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell*, **38**, 675–680.
- McLachlan, A. D. (1979). Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79.
- Mertz, J. E., Güntert, P., Wüthrich, K. & Braun, W. (1991). Complete relaxation matrix refinement of NMR structures of proteins using analytically calculated dihedral angle derivatives of NOE intensities. *J. Biomol. NMR.* **1**, 257–269.
- Messiah, A. (1979). *Quantenmechanik*, de Gruyter, Berlin.
- Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975). Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **79**, 2361–2381.
- Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988). Isolation and sequence-specific DNA binding of the *Antennapedia* homeodomain. *EMBO J.* **7**, 4299–4304.
- Müller, N., Ernst, R. R. & Wüthrich, K. (1986). Multiple-quantum-filtered two-dimensional correlated NMR spectroscopy of proteins. *J. Amer. Chem. Soc.* **108**, 6482–6492.
- Nagayama, K. & Wüthrich, K. (1981). Structural interpretation of vicinal proton-proton coupling constants $^3J_{\alpha\beta}$ in the basic pancreatic trypsin inhibitor measured by two-dimensional J-resolved NMR spectroscopy. *Eur. J. Biochem.* **115**, 653–657.
- Neff, B. L., Ackerman, J. L. & Waugh, J. S. (1977). Fully automatic software correction of Fourier transform NMR spectra. *J. Magn. Reson.* **25**, 335–340.
- Nelson, S. J. & Brown, T. R. (1989). The accuracy of quantification from 1D NMR spectra using the PIQABLE algorithm. *J. Magn. Reson.* **84**, 95–109.
- Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983). Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. *J. Phys. Chem.* **87**, 1883–1887.
- Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labelling. *Biochemistry*, **28**, 7510–7516.
- Neri, D., Otting, G. & Wüthrich, K. (1990). New nuclear magnetic resonance experiment for measurements of the vicinal coupling constants $^3J_{\text{HN}\alpha}$ in proteins. *J. Amer. Chem. Soc.* **112**, 3663–3665.
- Neri, D., Billeter, M. & Wüthrich, K. (1992). Determination of the NMR solution structure of the DNA-binding domain(1–69) of the 434 repressor and comparison with the X-ray crystal structure. *J. Mol. Biol.* **223**, 743–767.
- Noggle, J. H. & Schirmer, R. E. (1971). *The nuclear Overhauser effect*, Academic Press, New York.
- Noguti, T. & Gō, N. (1983). A method for rapid calculation of a second derivative matrix of conformational energy for large molecules. *J. Phys. Soc. Jpn.* **52**, 3685–3690.

- O'Connell, J. F., Bougis, P. E. & Wüthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of cardiotoxin CTX IIB from *Naja mossambica mossambica*. *Eur. J. Biochem.* **213**, 891–900.
- Olejniczak, E. T. & Eaton, H. L. (1990). Extrapolation of time-domain data with linear prediction increases resolution and sensitivity. *J. Magn. Reson.* **87**, 628–632
- Otting, G., Widmer, H., Wagner, G. & Wüthrich, K. (1986). Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. *J. Magn. Reson.* **66**, 187–193.
- Otting, G., Qian, Y. Q., Müller, M., Gehring, W. J. & Wüthrich, K. (1988). Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J.* **7**, 4305–4309.
- Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.
- Pardi, A., Billeter, M. & Wüthrich, K. (1984). Calibration of the angular dependence of the amide proton- C^α proton coupling constants, $^3J_{HN\alpha}$, in a globular protein. Use of $^3J_{HN\alpha}$ for identification of helical secondary structure. *J. Mol. Biol.* **180**, 741–751.
- Pearson, G. A. (1977). A general baseline-recognition and baseline-flattening algorithm. *J. Magn. Reson.* **27**, 265–272.
- Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming* **12**, 241–254.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. (1986). *Numerical Recipes. The art of scientific computing*, Cambridge University Press, Cambridge.
- Qian, Y. Q. (1992). *Determination of the NMR solution structures of the Antennapedia homeodomain and an Antennapedia homeodomain-DNA complex*, Diss. ETH Nr. 9972, Zürich.
- Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.
- Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W. & Wüthrich, K. (1993). NMR spectroscopy of a DNA complex with the uniformly ^{13}C -labeled *Antennapedia* homeodomain and structure determination of the DNA-bound homeodomain. *J. Mol. Biol.*, im Druck.
- Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wüthrich, K. (1983). Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Comm.* **117**, 479–485.
- Rydel, T. J., Ravichandran, K. G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C. & Fenton, J. W. II (1990). The structure of a complex of recombinant hirudin and human α -thrombin. *Science*, **249**, 277–280.
- Rydel, T. J., Tulinsky, A., Bode, W. & Huber, R. (1991). Refined structure of the hirudin-thrombin complex. *J. Mol. Biol.* **221**, 583–601.
- Sasisekharan, V. (1959). Structure of poly-L-proline II. *Acta Crystallogr.* **12**, 897–903.
- Schaumann, T. (1988). *Strukturverfeinerung von Polypeptiden und Proteinen mit Newton-Raphson-Energieminimierung*, Diss. ETH Nr. 8469, Zürich.
- Schaumann, T., Braun, W. & Wüthrich K. (1990). The program FANTOM for energy refinement of polypeptides and proteins using a Newton-Raphson minimizer in torsion angle space. *Biopolymers*, **29**, 679–694.
- Schott, O. (1992). *Bestimmung der NMR Struktur von 434-Repressor(1–63)* in Lösung, Diplomarbeit, Abteilung für Mathematik und Physik, ETH Zürich.
- Schulz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, Springer, New York.
- Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R. & Wüthrich, K. (1988). Conformation of [Cd₇]-metallothionein-2 from rat liver in aqueous solution determined by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **203**, 251–268.

- Scott, M. P. & Weiner, A. J. (1984). Structural relationships among genes that control development: Sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc. Nat. Acad. Sci. USA* **81**, 4115–4119.
- Scott, M. P., Tamkun, J. W. & Hartzell, G. W. III (1989). The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**, 25–48.
- Senn, H., Werner, B., Messerle, B. A., Weber, C., Traber, R. & Wüthrich, K. (1989). Stereospecific assignment of the methyl ^1H NMR lines of valine and leucine in polypeptides by non-random ^{13}C labelling. *FEBS Lett.* **249**, 113–118.
- Shriver, J. W. (1991). NMR product-operator calculations in *Mathematica*. *J. Magn. Reson.* **94**, 612–616.
- Singh, U. C., Weiner, P. K., Caldwell, J. W. & Kollman, P. A. (1986). Amber 3.0, University of California at San Francisco.
- Sodano, P., Xia, T., Bushweller, J. H., Björnberg, O., Holmgren, A., Billeter, M. & Wüthrich, K. (1991). Sequence-specific ^1H assignments and determination of the three-dimensional structure of reduced *Escherichia coli* glutaredoxin. *J. Mol. Biol.* **221**, 1311–1324.
- Solomon, I. (1955). Relaxation processes in a system of two spins. *Phys. Rev.* **99**, 559–565.
- Sørensen, O. W., Eich, G. M., Levitt, M. H., Bodenhausen, G. & Ernst, R. R. (1983). Product operator formalism for the description of NMR pulse experiments. *Progr. NMR Spectrosc.* **16**, 163–192.
- Stephenson, D. S. (1988). Linear prediction and maximum entropy methods in NMR spectroscopy. *Prog. NMR Spectrosc.* **20**, 515–626.
- Stejskal, E. O. & Schaefer, J. (1974). Comparisons of quadrature and single-phase Fourier transform NMR. *J. Magn. Reson.* **14**, 160–169.
- Stone, S. R. & Hofsteenge, J. (1986). Kinetics of the inhibition of thrombin by hirudin. *Biochemistry*, **25**, 4622–4624.
- Stryer, L. (1985). *Biochemie*, Vieweg, Braunschweig.
- Szyperski, T. (1993). *Charakterisierung der strukturellen Eigenschaften von Hirudin mit NMR in Lösung sowie Entwicklung und Anwendung neuer NMR-Methoden zur Untersuchung von Proteinen in Lösung*, Diss. ETH Nr. 9981, Zürich.
- Szyperski, T., Güntert, P., Otting, G. & Wüthrich, K. (1992a). Determination of scalar coupling constants by inverse Fourier transformation of in-phase multiplets. *J. Magn. Reson.* **99**, 552–560.
- Szyperski, T., Güntert, P., Stone, S. R. & Wüthrich, K. (1992b). The NMR solution structure of hirudin(1–51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J. Mol. Biol.* **228**, 1193–1205.
- Szyperski, T., Güntert, P., Stone, S. R., Tulinsky, A., Bode, W., Huber, R. & Wüthrich, K. (1992c). Impact of protein-protein contacts on the conformation of thrombin-bound hirudin studied by comparison with the NMR solution structure of hirudin(1–51). *J. Mol. Biol.* **228**, 1206–1211.
- Szyperski, T., Neri, D., Leiting, B., Otting, G. & Wüthrich, K. (1992d). Support of ^1H NMR assignments in proteins by biosynthetically directed fractional ^{13}C -labeling. *J. Biomol. NMR*, **2**, 323–334.
- Szyperski, T., Luginbühl, P., Otting, G., Güntert, P. & Wüthrich, K. (1993a). Protein dynamics studied by rotating frame ^{15}N spin relaxation times. *J. Biomol. NMR* **3**, 151–164.
- Szyperski, T., Wider, G., Bushweller, J. H. & Wüthrich, K. (1993b). 3D ^{13}C - ^{15}N heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C - ^{15}N -double-labeled proteins. *J. Biomol. NMR*, **3**, 127–132.
- Torda, A. E., Scheek, R. M. & van Gunsteren, W. (1989). Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Lett.* **157**, 289–294.

- Torda, A. E., Brunne, R. M., Huber, T., Kessler, H. & van Gunsteren, W. (1993). Structure refinement using time-averaged J -coupling constant constraints. *J. Biomol. NMR*, **3**, 55–66.
- Torda, A. E., Scheek, R. M. & van Gunsteren, W. (1990). Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. Mol. Biol.* **214**, 223–235.
- TRIPOS Associates, Inc. (1992). *SYBYL/NMR TRIAD version 5.5.1*, St. Louis, Missouri, USA.
- van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J. & Postma, J. P. M. (1983). Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. *Proc. Natl. Acad. Sci., U.S.A.* **80**, 4315–4319.
- Vendrell, J., Wider, G., Avilés, F. X. & Wüthrich, K. (1990). Sequence-specific ^1H NMR assignments and determination of the secondary structure for the activation domain isolated from pancreatic procarboxypeptidase B. *Biochemistry*, **29**, 7515–7522.
- Vendrell, J., Billeter, M., Wider, G., Avilés, F. X. & Wüthrich, K. (1991). The NMR structure of the activation domain isolated from porcine procarboxypeptidase B. *EMBO J.* **10**, 11–15.
- Verlet, L. (1967). Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103.
- Wagner, G. & Wüthrich, K. (1979). Truncated driven nuclear overhauser effect (TOE). A new technique for studies of selective ^1H - ^1H Overhauser effects in the presence of spin diffusion. *J. Magn. Reson.* **33**, 675–680.
- Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gō, N. & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using the algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
- Wako, H. & Gō, N. (1987). Algorithm for rapid calculation of hessian of conformational energy function of proteins by supercomputer. *J. Comp. Chem.* **8**, 625–635.
- Weber, P. L., Morrison, R. & Hare, D. (1988). Determining stereo-specific ^1H nuclear magnetic resonance assignments from distance geometry calculations. *J. Mol. Biol.* **204**, 483–487.
- Weiner, P. K., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.
- Wider, G. (1990). Elimination of baseline artifacts in NMR spectra by oversampling. *J. Magn. Reson.* **89**, 406–409.
- Wider, G., Lee, K. H. & Wüthrich, K. (1982). Sequential resonance assignments in protein ^1H nuclear magnetic resonance spectra. Glucagon bound to perdeuterated dodecylphosphocholine micelles. *J. Mol. Biol.* **155**, 367–388.
- Widmer & Wüthrich, K. (1986). Simulation of two-dimensional NMR experiments using numerical density matrix calculations. *J. Magn. Reson.* **70**, 270–279.
- Widmer, H., Billeter, M. & Wüthrich, K. (1989). Three-dimensional structure of the neurotoxin ATX Ia from *anemonia sulcata* in aqueous solution determined by nuclear magnetic resonance spectroscopy. *Proteins*, **6**, 357–371.
- Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIa from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315.
- Wolfram, S. (1988). *Mathematica. A system for doing mathematics by computer*, Addison-Wesley, Redwood City, California.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
- Wüthrich, K. (1988). Proteinstrukturermittlung in Lösung mittels kernmagnetischer Resonanzspektroskopie. *Phys. Bl.* **44**, 103–109.
- Wüthrich, K. (1989a). The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Acc. Chem. Res.* **22**, 36–44.
- Wüthrich, K. (1989b). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, **243**, 45–50.

- Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.* **155**, 311–319.
- Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.
- Wüthrich, K., Billeter, M. & Braun, W. (1984). Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J. Mol. Biol.* **180**, 715–740.
- Wüthrich, K., Güntert, P. & Berndt, K. D. (1993). Computer-supported NMR structure determination of proteins in solution illustrated with studies of protein proteinase inhibitors. In *Proceedings of the workshop on innovations on proteases and their inhibitors: fundamental and applied aspects* (Ed. F. X. Avilés), de Gruyter, Berlin, im Druck.
- Xia, T. H. (1992). *Software for determination and visual display of NMR structures of proteins: The distance geometry program DGPLAY and the computer graphics programs CONFOR and XAM*, Diss. ETH Nr. 9831, Zürich.
- Xia, T. H., Bushweller, J. H., Sodano, P., Billeter, M., Björnberg, O., Holmgren, A. & Wüthrich, K. (1992). NMR structure of oxidized *Escherichia coli* glutaredoxin: Comparison with reduced *E. coli* glutaredoxin and functionally related proteins. *Protein Science*, **1**, 310–321.
- Yip, P. & Case, D. A. (1989). A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra. *J. Magn. Reson.* **83**, 643–648.
- Yonath, A. & Traub, W. (1969). Polymers of tripeptides as collagen models. IV. Structure of poly(L-prolyl-glycyl-L-proline). *J. Mol. Biol.* **43**, 461–477.
- Zhu, G. & Bax, A. (1990). Improved linear prediction for truncated signals of known phase. *J. Magn. Reson.* **90**, 405–410.
- Zuiderweg, E. R. P., Nettesheim, D. G., Mollison, K. W. & Carter, G. W. (1989). Tertiary structure of human complement C5a in solution from nuclear magnetic resonance data. *Biochemistry*, **28**, 172–185.
- Zuiderweg, E. R. P., Petros, A. M., Fesik, S. W. & Olejniczak, E. T. (1991). Four-dimensional [^{13}C , ^1H , ^{13}C , ^1H] HMQC-NOE-HMQC NMR spectroscopy: resolving tertiary NOE distance constraints in the spectra of larger proteins. *J. Amer. Chem. Soc.* **113**, 370–372.
-

Abkürzungen

BPTI	<u>B</u> asischer <u>p</u> ankreatischer <u>T</u> rypsin <u>i</u> nhibitor
ADB	<u>A</u> ktivierungs <u>d</u> omäne von Procarboxypeptidase <u>B</u>
<i>Antp</i>	<i>Antennapedia</i>
CALIBA	» <u>c</u> alibration«; Programm zur Umwandlung von NOE Intensitäten in Distanzeinschränkungen
COFIMA	» <u>c</u> oordinate <u>f</u> ile <u>m</u> anipulation«; Programm zur Umwandlung und Bearbeitung von Koordinaten- und Einschränkungdateien.
$d_{ab}(i,j)$	Distanz zwischen dem Atom <i>a</i> im Aminosäurerest <i>i</i> und dem Atom <i>b</i> im Rest <i>j</i>
DIANA	<u>D</u> istanzgeometrie <u>a</u> lgorithmus für <u>N</u> MR <u>A</u> nwendungen
DNS	Desoxyribonukleinsäure
EASY	» <u>E</u> TH <u>a</u> utomated <u>s</u> pectroscopy«; Programm zur Interpretation mehrdimensionaler Kernresonanzspektren
GLOMSA	» <u>g</u> lobal <u>m</u> ethod for obtaining <u>s</u> tereospecific <u>a</u> ssignments«; Programm zur Bestimmung von stereospezifischen Zuordnungen
HABAS	Programm zur Bestimmung stereospezifischer Zuordnungen für β -Protonen und von Diederwinkeleinschränkungen in Proteinen
${}^3J_{\text{HN}\alpha}$	Spin-Spin Kopplungskonstante zwischen dem Amid- und dem α -Proton eines Aminosäurerests
${}^3J_{\alpha\beta}$	Spin-Spin Kopplungskonstante zwischen dem α - und einem β -Proton eines Aminosäurerests
NMR	<u>n</u> uclear <u>m</u> agnetic <u>r</u> esonance; magnetische Kernspinresonanz
NOE	<u>n</u> uclear <u>O</u> verhauser <u>e</u> ffect; Kern-Overhauserereffekt
NOESY	<i>two-dimensional nuclear Overhauser effect spectroscopy</i> ; Zweidimensionale Kern-Overhauserereffektspektroskopie
POMA	<u>P</u> roduktoperatorformalismus in <u>M</u> athematica
REDAC	» <u>u</u> se of <u>r</u> edundant <u>d</u> ihedral <u>a</u> nge <u>c</u> onstraints«; Verwendung redundanter Diederwinkeleinschränkungen

Lebenslauf

Am 29. März 1964 wurde ich als Sohn von Werner Güntert und Alice Güntert, geb. Heusi, in Schaffhausen geboren. Ich bin Bürger von Schleithem (Kanton Schaffhausen) und deutscher Staatsangehöriger. Zwischen 1970 und 1974 besuchte ich die Grund- und Hauptschule in Büsingen (Deutschland), anschließend die Real- und Kantonsschule in Schaffhausen, die ich 1982 mit der Matura, Typus C, beendete. Danach begann ich das Studium der Physik an der ETH Zürich, wo ich 1987 mit dem Diplom abschloß. Von 1988 bis 1993 arbeitete ich als Doktorand und Assistent unter der Leitung von Prof. Kurt Wüthrich am Institut für Molekularbiologie und Biophysik der ETH Zürich an der vorliegenden Abhandlung.

Zürich, im März 1993 Peter Güntert