Doctoral Thesis

NMR methods for studies of folded and unfolded forms of globular proteins

Author(s):
Hiller, Sebastian

Publication Date:
2006

Permanent Link:
https://doi.org/10.3929/ethz-a-005275871

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
NMR methods for studies of folded
and unfolded forms of globular proteins

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Natural Sciences

presented by
Sebastian Hiller
Dipl. Natw. ETH Zurich
born 14.12.1976
citizen of Germany

accepted on the recommendation of

Prof. Dr. Kurt Wüthrich, examiner
Prof. Dr. Beat H. Meier, co-examiner
Prof. Dr. Gerhard Wider, co-examiner

2006
Acknowledgments

I would like to thank Prof. Dr. Kurt Wüthrich for the great opportunity to work in his laboratory. His continuous support, his interest in my work and the excellent infrastructure he provided were essential for the success of this work. The numerous discussions and seminars with him contributed substantially to my education.

I also want to thank Prof. Dr. Gerhard Wider, who spent a lot of time teaching me the skills necessary to operate an NMR spectrometer and supervising all projects in this thesis. I am particularly grateful for his collaboration and his strong support in the APSY project.

My thanks go to Prof. Dr. Beat H. Meier for accepting to examine this work as a co-referee.

I want to thank the Swiss science foundation and Prof. Dr. Rüdiger Wehner for the permanent intellectual and financial support during my undergraduate and doctoral studies.

I had the pleasure to share the lab with Alvar Gossert. I cordially thank him for his friendship, for the uncounted fruitful and unfruitful discussions and for the many other funny things we did.

Dr. César Fernández is kindly acknowledged for his great scientific spirit and the guidance in the initial phase of this work.

I want to thank Dr. Francesco Fiorito for the good collaboration that lead to the APSY technique.

Many thanks also go to Lukas Imbach and Christian Wasmer, who have contributed to this work with their diploma theses.

Isabelle Allen, Ruedi Baumann, Johanna Blatter and Monika Marti are kindly acknowledged for their daily efforts to maintain the technical infrastructure.

I would like to thank all members of the Wüthrich groups in Zürich and La Jolla, especially Pascal Bettendorf, Dr. Sophie Bonjour, Dr. Fred Damberger, Lars Dreier, Dr. Touraj Etezady-Esfarjani, Michael Geraldt, Dr. Torsten Herrmann, Dr. Christian Hilty and Dr. Reto Horst for a good working atmosphere, all kinds of help and many valuable discussions.

I deeply thank my parents Barbara and Manfred for their constant support and my brother Niffe for many joyful hours. This thesis is dedicated to my family.

Thanks to all friends, far and close, not mentioned by names here, for a lot of fun during the last years.

I am particularly glad about and grateful for the cover page designed by Didi Fromherz, and the inspiring discussions associated with it.

Charlotte, the major sunshine of my life, has given me a lot of love during the last seven years. Thank you.
# Table of Contents

**Introduction** ............................................................................................................................ 10

1. **Automated projection spectroscopy and its applications** ................................................. 11

   Automated projection spectroscopy (APSY) ....................................................................... 11
   The sensitivity of APSY experiments .................................................................................. 24
   A library of APSY experiments for protein backbone NMR assignments ...................... 30
   Automated backbone resonance assignment of proteins...................................................... 37
   Automated backbone assignment of folded proteins using strategy I–A......................... 45
   Automated backbone assignment of a folded protein using strategy IIb–B ...................... 60
   Automated backbone assignment of a denatured protein................................................... 67

2. **Characterization of the denatured form of the membrane protein OmpX** ............. 81

   Introduction to studies of proteins under denaturing conditions ...................................... 81
   Identification of nonrandom structure in urea-unfolded OmpX ......................................... 84
   Independent formation of hydrophobic clusters ............................................................... 91
   Binding of hydrophobic clusters to micelles ..................................................................... 98
   Structural characterization of denatured OmpX in complex with DHPC micelles .......... 104
   A rapid and sensitive method to detect hydrophobic clusters .......................................... 111
   Materials and Methods ...................................................................................................... 117

**Conclusion** ............................................................................................................................. 121

**Appendix** ............................................................................................................................... 122

   Nomenclature .................................................................................................................... 122
   Calculation of envelope functions for time domain cross-sections .................................. 124
   A library of transfer elements for APSY backbone experiments ..................................... 125

**References** ............................................................................................................................. 138

**Curriculum Vitae** .................................................................................................................. 153

**Publications** .......................................................................................................................... 154
Summary

Nuclear magnetic resonance (NMR) spectroscopy is an important tool in contemporary biological research. For efficient studies of biological macromolecules, in particular in high-throughput setups, robust automated experimental procedures are required. Despite several automation attempts, substantial human interactions are still needed for the acquisition and analysis of NMR spectra. Automated projection spectroscopy (APSY), which is presented in the first part of this thesis, is a new approach for the combined acquisition and analysis of $N$-dimensional ($N \geq 3$) NMR experiments. The typical result of APSY is a peak list that contains only the $N$-dimensional resonances, with efficient suppression of artifacts and noise peaks. High-dimensional peak lists have a large information content, but they are not accessible with conventional experiments due to the prohibitively long sampling times. In APSY, the high-dimensional spectrum is not measured, but only a set of low-dimensional projection spectra, which can be recorded in a short time. The new algorithm GAPRO extracts the $N$-dimensional chemical shift information from the projection spectra and circumvents any manual interaction. The projection angles and the number of projections can be chosen freely with APSY, which is an important prerequisite to optimize the experimental sensitivity and the quality of the results.

APSY provides precise, artifact-free and complete peak lists in a fully automated way and is therefore an ideal acquisition technique for automated sequence-specific backbone resonance assignments of folded and unfolded proteins. With unfolded proteins, APSY benefits from the long spin relaxation times typically encountered in these systems, enabling experiments with up to seven dimensions. The high dimensionality in turn greatly reduces the resonance overlap typically caused by the small chemical shift dispersion of unfolded proteins. Using APSY and the assignment algorithm GARANT, all detectable backbone resonances of the urea-denatured form of the 148-residue outer membrane protein X (OmpX) from *E. coli* were correctly and fully automatically assigned. Only 20 h of spectrometer time and a few minutes of calculation time on a standard PC were required, which is an order of magnitude smaller than the conventional, interactive approach.

Sequence-specific backbone resonance assignments are a prerequisite for the NMR studies of denatured OmpX, which are reported in the second part of this thesis. Studies of unfolded proteins have contributed substantial knowledge to the understanding of protein folding mechanisms in the past. Whereas much research focuses on denatured soluble proteins and their folding transition states, only little is known about the *in vitro* folding mechanisms of integral membrane proteins. The present study shows that urea-denatured OmpX is globally
unfolded, but adopts two local non-random conformations in well separated segments of the polypeptide chain. The two clusters were structurally characterized and appear to be stabilized by hydrophobic interactions of the side chains. NMR measurements comparing the denatured form of wild-type OmpX and single-amino acid variants of OmpX show that the two hydrophobic clusters can be disrupted independently, implying that cluster formation is encoded entirely in the local amino acid sequence. A comparison of one of the hydrophobic clusters with a peptide analog supports this conclusion. OmpX can be refolded efficiently in vitrō from its urea-denatured state into micelles of the detergent DHPC. NMR experiments showed that urea-denatured OmpX binds DHPC micelles specifically via its two hydrophobic clusters, but that other segments of the denatured polypeptide chain do not interact with the micelles. Structural models of the complex of denatured OmpX and DHPC micelles indicate that the two hydrophobic clusters undergo a structural change upon binding to DHPC micelles. Overall, these experiments lead to a comprehensive understanding of denatured OmpX and its micelle-bound form, which are possible conformations on the in vitro folding pathway of this integral membrane protein.

The APSY techniques which enabled the automated sequence-specific backbone resonance assignments of denatured OmpX can be extended to folded proteins. In folded proteins, spin relaxation times are shorter than in unfolded proteins, and thus experimental sensitivity represents the limiting factor for APSY in these systems. Based on theoretical considerations on sensitivity, a library of APSY pulse sequence elements for three different classes of automated backbone assignment strategies was created. Two examples of these strategies were applied to folded proteins.

In the first example, the new 6D APSY-seq-HNCOCANH NMR experiment was used to assign the backbone resonances of two folded proteins consisting of 63 and 115 amino acid residues, respectively. In both proteins, all NMR-observable nuclei were correctly assigned with the assignment algorithm GARANT. In the second example, a more sensitive assignment strategy consisting of three APSY experiments was applied to the protein 434-repressor(1–63). GARANT assigned 98% of the protein backbone atoms correctly. These examples show that automated sequence-specific backbone resonance assignment of folded proteins with APSY and a suitable assignment algorithm is possible. Based on this experience, it is likely that standard setups for the routine use of the APSY assignment strategies can be established.
Zusammenfassung

Kernspinresonanz ist ein wichtiges Werkzeug der modernen biologischen Forschung. Für effiziente Studien von biologischen Makromolekülen, besonders in Massenanwendungen, werden robuste automatische Methoden benötigt. Trotz verschiedener Versuche zur Automatisierung ist aber nach wie vor ein beträchtlicher Aufwand an menschlicher Interaktion für die Aufnahme und Auswertung von NMR Spektren nötig. Automatische Projektionsspektroskopie (APSY), eine Technik, die im ersten Teil dieser Arbeit vorgestellt wird, ist ein neuer Ansatz zur kombinierten Aufnahme und Auswertung N-dimensionaler \( (N \geq 3) \) NMR-Experimente. Das typische Ergebnis von APSY ist eine Peakliste, die nur die N-dimensionalen Resonanzen enthält, mit effizienter Unterdrückung spektraler Artefakte und statistischen Rauschens. Hochdimensionale Peaklisten haben einen hohen Informationsgehalt, sind aber aufgrund der langen Aufnahmezeiten mit konventionellen Methoden nicht zugänglich. Mit APSY wird nicht das hochdimensionale Spektrum, sondern nur ein Satz niedrigdimensionaler Projektionsspektren in vergleichsweise kurzer Zeit aufgenommen. Der neue Algorithmus GAPRO ermittelt die \( N \)-dimensionalen Resonanzen direkt aus den Projektionen und umgeht dabei jegliche menschliche Interaktion. Die Projektionswinkel und die Zahl der Projektionen können mit APSY frei gewählt werden, was eine wichtige Vorraussetzung für die Optimierung der experimentellen Empfindlichkeit und der Ergebnisqualität ist.


Sequenzspezifische Resonanzzuordnungen sind eine Grundvoraussetzung für NMR-Studien mit denaturiertem OmpX, die im zweiten Teil dieser Arbeit gezeigt werden. In den letzten Jahrzehnten haben Forschungen über ungefaltete Proteine beträchtlich zum Verständnis von Proteinfaltungsmechanismen beigetragen. Während die meisten dieser Studien mit entfalteten


Abbreviations

1D (2D, 3D, 4D, 5D, 6D, 7D)  one- (two-, three-, four-, five-, six-, seven-) dimensional
APSY      automated projection spectroscopy
C₈POE     n-Octylpolyoxyethylene
cmc       critical micelle concentration
DC₇PC     1,2-Diheptanoyl-sn-glycero-3-phosphocholine
DHPC      1,2-Dihexanoyl-sn-glycero-3-phosphocholine
DM        n-Dodecyl-β-D-maltoside
DQ        double-quantum
DSS       2,2-Dimethyl-2-silapentane-5-sulfonic acid
FT        Fourier transform
GAPRO     geometric analysis of projections
GFT       G-matrix Fourier transform
HSQC      heteronuclear single quantum coherence
INEPT     insensitive nuclei enhanced by polarization transfer
NMR       nuclear magnetic resonance
NOE       nuclear Overhauser effect
NOESY     NOE spectroscopy
OG        n-Octyl-β-D-glucoside
OmpA      outer membrane protein A from Escherichia coli
OmpX      outer membrane protein X from Escherichia coli
PC        personal computer
ppb       parts per billion
ppm       parts per million
PR        projection-reconstruction
RMSD      root mean square deviation
TM        Thermotoga maritima
TOCSY     total correlation spectroscopy
ZQ        zero-quantum
## APSY parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>dimensionality of the APSY experiment</td>
</tr>
<tr>
<td>$S$</td>
<td>support of a candidate point</td>
</tr>
<tr>
<td>$S_{\text{min},1}$</td>
<td>minimal support 1</td>
</tr>
<tr>
<td>$S_{\text{min},2}$</td>
<td>minimal support 2</td>
</tr>
<tr>
<td>$k$</td>
<td>number of GAPRO iterations</td>
</tr>
<tr>
<td>$w$</td>
<td>iterations for coordinate convergence</td>
</tr>
<tr>
<td>$\Delta \nu_{\text{min}}$</td>
<td>intersection tolerance in the direct dimension</td>
</tr>
<tr>
<td>$r_{\text{min}}$</td>
<td>intersection tolerance in the indirect dimension</td>
</tr>
<tr>
<td>$R_{\text{min}}$</td>
<td>signal-to-noise ratio for peak picking</td>
</tr>
</tbody>
</table>
Introduction

Modern Nuclear Magnetic Resonance spectroscopy is a key technique for biophysical studies of proteins and other biological macromolecules in solution (Wüthrich, 1986; 2003). Information on structure and dynamics which may not be accessible by other methods can be obtained at atomic resolution with NMR (Bax and Grzesiek, 1993; Wüthrich, 1994; Kay and Gardner, 1997). In particular for disordered systems, such as denatured or intrinsically unstructured proteins, which can not be transferred into a crystalline state, NMR provides the only source of atomic resolution data (Dyson and Wright, 2002). These systems are of high interest for fundamental questions of structural biology: The principles behind protein folding mechanisms can be elucidated from studies of denatured proteins and their transitions towards collapsed forms (Anfinsen, 1973; Itzhaki et al., 1995; Mayor et al., 2003) and unstructured proteins play important roles in cellular processes and aggregation diseases (Plaxco and Gross, 1997; Dobson, 1999; Wright and Dyson, 1999; Dunker et al., 2001).

One of the drawbacks in modern bio-NMR are the long instrument times needed for the conventional Fourier transform schemes (Ernst et al., 1987). The amount of time needed to acquire data can be larger than would be necessary from the sensitivity requirements. This situation has been termed the sampling limit and it occurs quite often for small and medium sized proteins with modern high-field spectrometers or cryogenic probes (Styles et al., 1984; Kim and Szyperski, 2003; Freeman and Kupce, 2004). The need of extensive human work to analyze the spectral data is a further bottleneck (Herrmann et al., 2002; Szyperski et al., 2002; Malmodin et al., 2003). For each new system, the sequence-specific resonance assignment is a mandatory piece of work, without which no further information can be acquired (Wüthrich, 1986). It is thus highly desirable to automate this process, also in order to eliminate human bias and thus to make results more comparable (Baran et al., 2004).

The first part of this thesis introduces APSY, a technique for the combined automated acquisition and analysis of projection spectra. Compared to conventional techniques, APSY substantially reduces the time requirements for the determination of resonance frequencies in multidimensional correlation spectra. The possibilities of APSY for automated backbone assignment of folded and unfolded proteins are analyzed and the feasibility is demonstrated. The second part of this thesis describes the biophysical characterization of the denatured form of the integral membrane protein OmpX. A series of high-resolution NMR and fluorescence experiments with wild type OmpX, single amino acid variants and a fragment of OmpX lead to a comprehensive picture of urea-denatured OmpX and its interactions with detergent micelles.
1. Automated projection spectroscopy and its applications

Automated projection spectroscopy (APSY)

This chapter is reproduced with minimal adaptations from:

Introduction

In NMR studies of biological macromolecules in solution (Wüthrich, 1986; Bax and Grzesiek, 1993; Kay and Gardner, 1997; Wüthrich, 2003), multidimensional NMR data are commonly acquired by sampling the time domain in all dimensions equidistantly at a resolution adjusted to the populated spectral regions (Ernst et al., 1987). With recent advances in sensitivity, due to high field strengths and/or cryogenic detection devices, the time required to explore the time domain in the conventional way typically exceeds by far the time needed for sensitivity considerations, so that the desired resolution in the indirect dimensions determines the duration of the experiment. In this situation of the “sampling limit”, which is common in 3- and higher-dimensional experiments with small and medium-size proteins (Szyperski et al., 2002), the desired chemical shift information has been collected using “unconventional” experimental schemes, such as non-uniform sampling of the time domain (Orekhov et al., 2003; Rovnyak et al., 2004) or combination of two or more indirect dimensions (Szyperski et al., 1993; Freeman and Kupce, 2004).

The concept of combining indirect dimensions has lead to reduced-dimensionality experiments (Szyperski et al., 1993) and G-Matrix Fourier transform (GFT-) NMR (Kim and Szyperski, 2003; Kozminski and Zhukov, 2003). In GFT-NMR, several evolution periods of a multidimensional NMR experiment are combined, the data are processed using a G-matrix, and the resulting set of spectra is analyzed jointly to identify the peaks that arise from the same spin system and to calculate their resonance frequencies (Kim and Szyperski, 2003). In another approach, projection-reconstruction (PR-) NMR (Kupce and Freeman, 2003a; b; 2004a; b), the projection–cross-section theorem (Bracewell, 1956; Nagayama et al., 1978) is combined with reconstruction methods from imaging techniques (Mersereau and Oppenheim, 1974; Lauterbur, 1973). In particular, a scheme for quadrature detection along tilted planes in the time domain allows the direct recording of orthogonal projections of any multidimensional experiment at arbitrary projection angles (Kupce and Freeman, 2004b).
PR-NMR, the full multidimensional spectrum is then reconstructed from the projections of
the multidimensional spectral data. (Kupče and Freeman, 2003a; b; 2004a; b).
The analysis of complex NMR spectra typically involves intensive human interaction, and
automation of NMR spectroscopy with macromolecules is still in development. Thereby the
distinction of real peaks from random noise and spectral artifacts, as well as peak overlap
represent major challenges (Koradi et al., 1998; Herrmann et al., 2002; Baran et al., 2004). On
grounds of principle, automated analysis benefits from higher-dimensionality of the spectra
(Moseley et al., 2004; Xia et al., 2004), since the peaks are then more widely separated, and
hence peak overlap is substantially reduced (Fig. 1.1).

Figure 1.1. Distribution of frequency differences, $d$, up to 800 Hz between pairs of nearest cross peaks in
multidimensional NMR spectra of the protein TM1290 for the following triple resonance experiments: (a) 3D HNCO, (b) 4D HNCOCA, (c) 5D HACACONH. $E$ is the number of times a given value of $d$ is expected. The calculations were based on the published chemical shift assignment of TM1290 (BMRB entry 5560) and assumed a field strength of 750 MHz.

In the present work we combine technologies to record projections of high-dimensional NMR
experiments described by Kupče and Freeman (2004b) and automated peak-picking using a
scheme of Herrmann et al. (2002) with a new algorithm, GAPRO (geometric analysis of
projections). Based on geometrical considerations, GAPRO identifies peaks in the projections
that arise from the same resonance in the $N$-dimensional frequency space, and subsequently
calculates the resonance frequencies in the $N$-dimensional spectrum without ever considering
the high-dimensional data set itself. This automated analysis of projected spectra, APSY
(automated projection spectroscopy), yields a peak list of the original multidimensional
experiment without any human interaction. In the following sections, the foundations of
APSY are introduced and characteristic properties of APSY are discussed. Two examples of
APSY are a 4D HNCOCA experiment (Grzesiek and Bax, 1992a; Yang and Kay, 1999) with
the 63-residue protein 434-repressor(1–63) (Neri et al., 1992b), and a 5D HACACONH
experiment (Kim and Szyperski, 2003) with the 115-residue protein TM1290 (Etezady-
Esfarjani et al., 2004).
Theoretical Background

Recording of projection spectra. The projection–cross-section theorem by Bracewell (Bracewell, 1956), which was introduced into NMR by Nagayama et al. (Nagayama et al., 1978), states that an \( m \)-dimensional (\( m < N \)) cross section, \( c_m(t) \), through \( N \)-dimensional time domain data is related by an \( m \)-dimensional Fourier transformation, \( \mathcal{F}_t \), and its inverse, \( \mathcal{F}_\omega \), to an \( m \)-dimensional orthogonal projection of the \( N \)-dimensional NMR spectrum, \( P_m(\omega) \), in the frequency domain. Thereby, \( P_m(\omega) \) and \( c_m(t) \) are oriented by the same angles with regard to their corresponding coordinate systems (Fig. 1.2). On this basis, Kupčé and Freeman proposed to record projections \( P_m(\omega) \) by sampling the corresponding time domain data, \( c_m(t) \), along a straight line (dashed line in Fig. 1.2). Quadrature detection is obtained from corresponding positive and negative projection angles for the subsequent hypercomplex Fourier transformation (Kupce and Freeman, 2004b).

![Figure 1.2. Illustration of the projection–cross-section theorem (Bracewell, 1956; Mersereau and Oppenheim, 1974; Nagayama et al., 1978) for a 2D subspace of an N-dimensional frequency space with two indirect dimensions x and y. A one-dimensional signal \( c_1^{xy}(t) \) on a straight line in the time domain (\( t_x, t_y \)) (left) is related to a one-dimensional orthogonal projection \( P_1^{xy}(\omega) \) of the spectrum (\( \omega_x, \omega_y \)) in the frequency domain (right) by a one-dimensional Fourier transformation, \( \mathcal{F}_t \), and the inverse transformation, \( \mathcal{F}_\omega \). The projection angle \( \alpha \) describing the slope of \( c_1^{xy}(t) \) defines also the slope of \( P_1^{xy}(\omega) \). The cross peak \( Q \) appears at the position \( Q \) in the projection \( f \). Further indicated are the spectral widths in the two dimensions of the frequency domain, \( SW_x \) and \( SW_y \).](image)

Projections of cross peaks. We describe here 2D projections, \( P_2(\omega) \), of an \( N \)-dimensional spectrum (\( N > 2 \)). \( P_2(\omega) \) represents spectral data in a 2D plane, which is spanned by an indirect dimension with unit vector \( \hat{p}_1 \), and the direct dimension, with \( \hat{p}_2 \). The indirect dimension is a 1D projection of the \( N-1 \) indirect dimensions. The same projection angles that parametrize the 2D time domain cross section, \( c_2(t) \), with respect to the \( N \) time domain axes, define the position of \( P_2(\omega) \) with respect to the \( N \) frequency axes (Table 1.1).
Table 1.1. Coordinates of the unit vector $\tilde{p}_1$ used in this work.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$N = 5$</th>
<th>$N = 4$</th>
<th>$N = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>$\sin(\gamma)$</td>
<td>$\sin(\beta)$</td>
<td>$\sin(\alpha)$</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>$\sin(\beta) \cdot \cos(\gamma)$</td>
<td>$\sin(\alpha) \cdot \cos(\beta)$</td>
<td>$\cos(\alpha)$</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>$\sin(\alpha) \cdot \cos(\beta) \cdot \cos(\gamma)$</td>
<td>$\cos(\alpha) \cdot \cos(\beta)$</td>
<td>-</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>$\cos(\alpha) \cdot \cos(\beta) \cdot \cos(\gamma)$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

For example, with $N = 5$, the two unit vectors $\tilde{p}_1$ and $\tilde{p}_2$ are expressed in the coordinate system of the $N$-dimensional frequency space by

$$
\tilde{p}_1^{N=5} = \begin{pmatrix}
\sin(\gamma) \\
\sin(\beta) \cdot \cos(\gamma) \\
\sin(\alpha) \cdot \cos(\beta) \cdot \cos(\gamma) \\
\cos(\alpha) \cdot \cos(\beta) \cdot \cos(\gamma) \\
0
\end{pmatrix}, \quad \tilde{p}_2^{N=5} = \begin{pmatrix}
0 \\
0 \\
0 \\
1
\end{pmatrix}.
$$

(1.1)

In the coordinate system of a 2D projection spectrum, $P_2(\omega)$, a projected cross peak $Q'_j$ has the position vector $\tilde{Q}'_j = [v'_{j,1}, v'_{j,2}]$, with $v'_{j,1}$ and $v'_{j,2}$ being the chemical shifts along the projected indirect dimension and the direct dimension, respectively. If the origins of both the $N$-dimensional coordinate system and the 2D coordinate system are in all dimensions in the center of the spectral range, the position vector $\tilde{Q}'_j$ in the $N$-dimensional frequency space is

$$
\tilde{Q}'_j = v'_{j,1} \cdot \tilde{p}_1 + v'_{j,2} \cdot \tilde{p}_2.
$$

(1.2)

The $N$-dimensional cross peak $Q'$ is located in an ($N$-2)-dimensional subspace $L$, which is orthogonal to the projection plane at the point $Q'_j$ (see Fig. 1.2). In APSY, a set of $j$ projections is recorded. From each $N$-dimensional cross peak $Q'$, a projected peak $Q'_j (j = 1, \ldots, j)$ appears in each of the $j$ projections. The set of projected peaks that arise from the same $N$-dimensional peak $Q'$, $\{Q'_1, \ldots, Q'_j\}$, is defined as a peak subgroup of $Q'$. Once the subgroup is known and contains a sufficiently large number of elements, the coordinates of $Q'$ can be calculated from the intersections of the subspaces from all subgroup elements. For 2D projections, at least $N-1$ elements are needed.
**Figure 1.3.** Flowchart of APSY. Square boxes indicate processes, and boxes with rounded corners denote intermediate or final results. The steps underlayed in grey are repeated $k$ times, and thus generate $k$ lists of peak subgroups.

**The APSY procedure.** The procedure is outlined in Figure 1.3 and illustrated in Figure 1.4. The operator selects the desired $N$-dimensional NMR experiment, the dimension of the projection spectra and $j$ sets of projection angles ($j \geq N-1$), and records the projection spectra. These are automatically peak picked using ATNOS (Herrmann et al., 2002), which yields $j$ peak lists. These lists contain the peaks $Q_{gf}$, where $g$ is an arbitrary numeration of the peaks in each projection $f (f = 1, ..., j)$. GAPRO arbitrarily selects $N-1$ from among these peak lists, and generates the subspaces $L_{gf}$ that are associated with the peaks $Q_{gf}$ in these projections (Fig. 1.4(a)). The intersections of the subspaces $L_{gf}$ in the $N$-dimensional space are candidates for the positions of $N$-dimensional cross peaks (open circles in Fig. 1.4(a)), where the number of candidate points typically exceeds the number of peaks in the spectrum. For each of these candidate points, the support, $S$, is then calculated as the number of subspaces from all $j$ projections that intersect at this point. Thereby at most one subspace from each projection is considered (Fig. 1.4(b); see text below), so that $N-1 \leq S \leq j$. The peaks $Q'_f$ associated with the subspaces that contribute to the support of a given candidate point form a “peak subgroup”. The subgroups are ranked for high $S$-values, and the top-ranked subgroup is selected (in case of degeneracy, one of the top-ranked subgroups is arbitrarily selected). The subspaces originating from the peaks $Q'_f$ in this subgroup are removed, and new $S$-values for the residual candidate points are calculated from the remaining subspaces (Fig. 1.4(c)). This procedure is repeated until the value of $S$ for all remaining candidate points falls below a user-defined threshold, $S_{\min,1}$, at which point a list of peak subgroups is generated.
Figure 1.4. Illustration of the algorithm GAPRO for \( N = 3, j = 5, k = 1 \) and \( S_{\text{min},1} = 3 \) (see Fig. 1.3). The two indirect dimensions are in the paper plane, and the acquisition dimension \( \omega_3 \) is orthogonal to it. The central part of each panel shows a 2D \((\omega_1, \omega_2)\)-cross section at the frequency \( \omega_y \) through the 3D spectrum. This is surrounded by five 1D cross sections at \( \omega_y \) through the five experimental 2D projections of the 3D spectrum with projection angles \( \alpha = 0^\circ, 90^\circ, 45^\circ, -25^\circ, \) and \(-60^\circ\). (a) The blue dots mark the result for the cross sections of the automatic peak picking of the 2D projections (Fig. 1.3). The algorithm then arbitrarily selected \( N-1 = 2 \) of the \( j \) projections for the first round of spectral analysis (Fig. 1.3), with \( \alpha = 0^\circ \) and \( \alpha = -60^\circ \). The intersections of the subspaces corresponding to the peaks in these two projections (green lines) identify 8 candidate points in the 3D spectrum (open circles). (b) Using the subspaces from all 5 projections, the support \( S \) (number of intersecting subspaces, see text) is calculated for each candidate point. Yellow and red dots indicate \( S = 2 \) and \( S = 5 \), respectively. (c) One of the three candidate points with the highest support (\( S = 5 \)) is arbitrarily selected. All peaks in the projections that contribute to the selected candidate point are identified as a peak subgroup (grey dots in the projections labeled with number 1). The subspaces from this subgroup are removed (grey dashed lines) and the support \( S \) of remaining candidate points is recalculated (there remains one point with \( S = 5 \), and another one with \( S = 4 \) is shown in orange). (d) After two more rounds of the procedure indicated in c), two additional subgroups are identified and labeled with numbers 2 and 3, respectively. From the three subgroups, the positions of three peaks in the 3D spectrum are calculated (black dots).

The subgroup identification is repeated with \( k \) different, randomly chosen starting combinations of \( N-1 \) projections (the user-defined parameter \( k \) is a small fraction of the total number of possible combinations of \( N-1 \) projections), and \( k \) peak subgroup lists are thus obtained (grey box in Fig. 1.3). These lists are merged into a single list, which is again subjected to the same type of ranking procedure, so that all subgroups with \( S < S_{\text{min},2} \) are eliminated. From the resulting “final” list of subgroups, the peak positions in the \( N \)-dimensional space are calculated (Fig. 1.4(d)). In the following, the computational techniques used for individual steps in Figure 1.3 are described.

Intersection of subspaces. To simplify the mathematical treatment, we describe a \((N-l)\)-dimensional subspace \( L \) \((1 < l < N)\) by a point \( Q^L \) in this subspace and a set of orthonormal vectors,
\{\vec{p}_i^{L},...,\vec{p}_r^{L}\}, orthogonal to L. To intersect, for example, four 3D subspaces in 5D frequency space, two of the 3D subspaces can intersect to a 2D subspace, which can then intersect with one of the remaining 3D subspaces to a 1D subspace, which can intersect with the fourth 3D subspace to a point.

L and M are two subspaces of dimensionality \((N-l)\) and \((N-m)\), with \(1 < l < N\) and \(1 < m < N\).

\(L\) is described by \(\{\vec{p}_i^{L},...,\vec{p}_r^{L}\}\) and the point \(Q^L\) by \(\{\vec{p}_1^{L},...,\vec{p}_m^{L}\}\) and \(Q^M\). Both \(L\) and \(M\) are orthogonal to the direct dimension, and therefore both \(\{\vec{p}_i^{L},...,\vec{p}_r^{L}\}\) and \(\{\vec{p}_1^{M},...,\vec{p}_m^{M}\}\) include the unit vector of the direct dimension. If Equations (1.3) and (1.4) are satisfied, the subspaces \(L\) and \(M\) intersect in a subspace \(K\) of dimensionality \((N-k)\), with \(k = l + m - 1\):

\[
\left| Q^L(N) - Q^M(N) \right| \leq \Delta \nu_{\text{min}} \tag{1.3}
\]

\[
\dim \{\vec{p}_1^{L},...,\vec{p}_r^{L},\vec{p}_1^{M},...,\vec{p}_r^{M}\} = l + m - 1 \tag{1.4}
\]

\(Q^L(N)\) and \(Q^M(N)\) are the \(N\)th coordinates of \(Q^L\) and \(Q^M\), respectively, \(\Delta \nu_{\text{min}}\) is a user-defined intersection tolerance in the direct dimension, and \(\dim\) stands for “dimension of”. Eq. (1.4) implies that \(\{\vec{p}_1^{L},...,\vec{p}_r^{L}\}\) and \(\{\vec{p}_1^{M},...,\vec{p}_m^{M}\}\) share only the direct dimension. The subspace \(K\) is then described by the orthonormal basis \(\{\vec{p}_1^{K},...,\vec{p}_k^{K}\}\), and by a point \(Q^K\) with its coordinates 1 to \((N-1)\) given by the \(l+m\) scalar products of Eq. (1.5):

\[
\vec{p}_z^{L} \cdot (Q^KQ^L) = 0 \quad z = 1,...,l
\]

\[
\vec{p}_z^{M} \cdot (Q^KQ^M) = 0 \quad z = 1,...,m \tag{1.5}
\]

The \(N\)th coordinate of \(Q^K\) is the arithmetic average of the \(N\)th coordinates of \(Q^L\) and \(Q^M\).

**Distance between a point and a subspace.** The distance \(r\) between a point \(Q\) and a \((N-l)\)-dimensional subspace \(L\), as described by a point \(Q^L\) and an orthonormal set of vectors orthogonal to \(L\), \(\{\vec{p}_i^{L},...,\vec{p}_r^{L}\}\), is given by

\[
r = \sqrt{\sum_{z=1}^{l} (\vec{p}_z^{L} \cdot (QQ^L))^2} \tag{1.6}
\]
Peak positions in the N-dimensional space. The fact that the peak positions are generally overdetermined by the experimental data is used to refine the peak coordinates. From each subgroup, \(N-1\) elements are arbitrarily chosen and their associated subspaces intersected to yield the position of the \(N\)-dimensional peak (Fig. 1.4). This procedure is repeated \(w\) times, where \(w\) is a user-defined parameter. Because of the limited precision of the individual chemical shift measurements, this results in \(w\) slightly different peak positions, which are then averaged in each dimension to obtain the final positioning of the \(N\)-dimensional peak \(Q^f\).

Materials and Methods

Sample preparation. \([U^{-13}C, ^{15}N]\)-labeled 434-repressor(1–63) was produced following published procedures (Neri et al., 1989; 1992b). For the NMR measurements, a 0.9 mM sample was prepared in 20 mM sodium phosphate buffer at pH 6.5. \([U^{-13}C, ^{15}N]\)-labeled TM1290 was produced as described (Etezady-Esfarjani et al., 2003). A NMR sample with 3.2 mM protein concentration in 20 mM phosphate buffer at pH 6.0 was prepared. Both NMR samples contained 95% / 5% H\(_2\)O / D\(_2\)O and 0.1% NaN\(_3\).

NMR spectroscopy. A 4D APSY-HNCOCA experiment (Fig. 1.5) was recorded with the 434-repressor(1–63) at 30°C on a Bruker DRX 750 MHz spectrometer equipped with a z-gradient triple resonance probe. The GAPRO algorithm was applied with the parameters \(S_{\text{min},1} = S_{\text{min},2} = 3\) (Fig. 1.3), \(k = 100\) (Fig. 1.3), \(w = 400\) (Fig. 1.3), \(\Delta \nu_{\text{min}} = 7.5\) Hz (Eq. (1.3)), a signal-to-noise threshold for ATNOS peak picking of \(R_{\text{min}} = 4.0\) (Herrmann et al., 2002), and a minimal distance for subspace intersection of \(r_{\text{min}} = 100\) Hz. The calculation time for GAPRO on a standard PC running Linux with a 2.8 GHz Pentium 4 processor was approximately 10 min.

A 5D APSY-HACACONH experiment (Kim and Szyperski, 2003) was recorded with the protein TM1290 at 40°C on a Bruker DRX 500 MHz spectrometer equipped with a z-gradient triple resonance cryogenic probehead. The GAPRO algorithm was applied with the following parameters: \(S_{\text{min},1} = S_{\text{min},2} = 4, k = 100, w = 800, \Delta \nu_{\text{min}} = 5\) Hz, \(R_{\text{min}} = 8.0, r_{\text{min}} = 50\) Hz. The GAPRO calculation time was approximately 30 min. Further details on the NMR experimental schemes used are described in the Supplementary Material of Hiller et al. (2005).
Figure 1.5. Pulse sequence used for 4D APSY-HNCOCA. Radio-frequency pulses were applied at 118.0 ppm for $^{15}$N, at 174.0 ppm for $^{13}$C', at 56.0 ppm for $^{13}$C$^{\alpha}$, and at 4.7 ppm for $^1$H. The carrier frequency on the carbon channel was switched between the $^{13}$C' and $^{13}$C$^{\alpha}$ carrier positions where indicated by the two vertical arrows. Narrow and wide bars represent 90º and 180º pulses, respectively. Pulses marked with an upper case letter were applied as shaped pulses; A: Gaussian shape, duration 100$\mu$s; B: Gaussian shape, 120$\mu$s; C: I-burp (Geen and Freeman, 1991), 200$\mu$s; D: Gaussian shape, 80$\mu$s; E: RE-burp (Geen and Freeman, 1991), 350$\mu$s (durations depend on the spectrometer frequency, here 750 MHz). All other pulses were rectangular pulses applied with high power. The last six $^1$H pulses represent a 3-9-19 WATERGATE element (Sklenar et al., 1993). Grey $^{13}$C$^{\alpha}$-pulses were applied to compensate for off-resonance effects of selective pulses (McCoy and Mueller, 1992). Decoupling using DIPSI-2 (Shaka et al., 1988) on $^1$H and WALTZ-16 (Shaka et al., 1983) on $^{15}$N is indicated by rectangles. $t_i$ is the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped pulsed magnetic field gradients applied along the z-axis, with the following durations and strengths: $G_1$, 700 $\mu$s, 13 G/cm; $G_2$, 1000 $\mu$s, 35 G/cm; $G_3$, 1000 $\mu$s, 35 G/cm; $G_4$, 800 $\mu$s, 16 G/cm; $G_5$, 800 $\mu$s, 13 G/cm; $G_6$, 800 $\mu$s, 18 G/cm; $G_7$, 1000 $\mu$s, 28 G/cm; $G_8$, 1000 $\mu$s, 28 G/cm. Phase cycling: $\phi_1 = \{y, -y\}$ $\phi_2 = \{x, x, -x, -x\}$, $\phi_3 = \{4x, 4-x\}$, $\phi_4 = \{x, -x, -x, x, -x, x, x, -x\}$, all other pulses $= x$. The following delays were used: $\tau = 2.7$ ms, $\delta = 13.75$ ms, $\eta = 4.5$ ms, $\lambda = 200$ $\mu$s. The indirect evolution times $t_1 - t_3$ were incremented according to the projections angles $\alpha$ and $\beta$ (see text and Table 1.1). Quadrature detection for the indirect dimensions was achieved using the hypercomplex Fourier transformation method for projections (Kupce and Freeman, 2004b), with the phases $\psi_{1-3}$.

**Peak picking.** Automated peak picking of 2D projection spectra was performed with a peak picking routine derived from ATNOS (Herrmann et al., 2002), which recognizes all local maxima in the spectrum. In the 4D APSY-HNCOCA experiment, peaks from glutamine and asparagine side chains were eliminated by removing all pairs of peaks with the following parameters: proton chemical shifts in the range of 6.5–8.2 ppm, with chemical shift differences in the indirect dimension < 10 Hz, and chemical shift differences in the direct dimension between 400 and 700 Hz.

**Results**

APSY-NMR spectroscopy is illustrated here with a 4D APSY-HNCOCA and a 5D APSY-HACACONH experiment. The 4D APSY-HNCOCA experiment was recorded with the 6.9 kDa protein 434-repressor(1–63). In a total spectrometer time of 4 h, $j = 27$ 2D projections
were recorded with the following projection angles (Table 1.1): \((\alpha, \beta) = (0°, 0°), (0°, 90°), (90°, 0°), (±30°, 0°), (±60°, 0°), (0°, ±30°), (0°, ±60°), (90°, ±30°), (90°, ±60°), (±30°, ±30°), (±60°, ±30°), (±45°, ±60°).\) The projection spectra were peak picked with ATNOS (Herrmann et al., 2002) to generate the input for GAPRO. The 4D peak list that resulted after about 15 minutes of GAPRO computation time contained 59 peaks, which is to be compared with a total of 60 peaks expected from the chemical structure of the molecule.

Although on average 18 ± 9 noise artifacts were picked in each projection, the final 4D peak list generated by the GAPRO algorithm contained 59 cross peaks and not a single artifact. Only the peak that would correlate the residues of the N-terminal dipeptide was missing. It had a signal intensity below the noise level in all projections. The precision of the chemical shifts in the final APSY peak list has been estimated to be 1 Hz in the direct dimension and 8 Hz in each of the three indirect dimensions.

![Figure 1.6](image-url)

**Figure 1.6.** 28 2D projections a1–a28 of the 5D APSY-HACACONH experiment measured with the protein TM1290 on a 500 MHz spectrometer equipped with a z-gradient triple resonance cryogenic probehead. The dimension \(\omega_1–4\) is the projection of the four indirect dimensions \(\omega_1(\alpha^H), \omega_2(\alpha^C'), \omega_3(\alpha^{13C}),\) and \(\omega_4(15N)\) obtained with the projection angles \(\alpha, \beta\) and \(\gamma\) (Table 1.1). The scales are centered on the carrier frequencies of 118.0 ppm for \(^{15}N\), 173.0 ppm for \(^{13}C'\), 54.0 ppm for \(^{13}C\), and 4.7 ppm for \(^1H\). The projection angles and the spectral widths in the projected indirect dimension are given in the supplementary material of Hiller et al. (2005). All projections are plotted with identical contour parameters.

The 5D APSY-HACACONH experiment was recorded with the 12.4 kDa protein TM1290. In this experiment, 28 2D projections were recorded in 11 h (Fig. 1.6). The final 5D peak list produced from the 5D APSY-HACACONH experiment with the protein TM1290 contained all the peaks that were expected from the chemical structure of the molecule and the
previously published NMR assignments (Etezady-Esfarjani et al., 2003), and there were no artifacts contained in the final peak list.

Discussion

In this chapter we presented the foundations of automated projection spectroscopy (APSY), and introduced the new algorithm GAPRO for automated spectral analysis. We then implemented APSY for high-dimensional heteronuclear correlation NMR experiments with proteins. In two applications without any human intervention after the initial set-up of the experiments, we obtained complete peak lists with high precision chemical shifts for 4D and 5D triple resonance spectra. For the future, we anticipate that APSY will be the first step, after protein preparation, in a fully automated process of protein structure determination by NMR.

In addition to providing automated peak picking and computation of the corresponding chemical shift lists, as described in this paper, APSY is expected to support automated sequential resonance assignment. For these envisaged goals, APSY has the promise of being a valid alternative to related NMR techniques that have recently been introduced for similar purposes. Thus, when compared to PR-NMR (Kupce and Freeman, 2003a; b; 2004a; b), APSY has the advantage of relying exclusively on the analysis of experimental low-dimensional projection spectra, with no need to ever reconstruct the parent high-dimensional spectrum. When compared to GFT-NMR (Kim and Szyperski, 2003), APSY differs in that there are no restrictions on the selection of the number of projections or the combinations of projection angles. The strongest asset of APSY, however, is that the new algorithm GAPRO enables fully automated analysis of the experimental projection spectra. As a primary result, complete peak picking and computation of high precision chemical shift lists are obtained without any bias that could result from human intervention.

APSY and protein size. APSY has so far been applied to the 6.9 kDa protein 434-repressor(1–63) and the 12.4 kDa protein TM1290. To obtain an estimate of possible limitations for APSY applications with larger proteins due to spectral overlap, we analyzed the peak separations in 4- and 5-dimensional triple resonance spectra of a sample of 54 proteins with sizes from $n = 50$ to 300 residues, which were simulated from the BMRB (Seavey et al., 1991) chemical shift deposits. Considered were the average of the distances from each peak to its nearest-by neighbor, $d_{av}$, and the distance between the two most closely spaced peaks in the entire spectrum, $d_{min}$. Figure 1.7 represents the data for the two experiments 4D HNCOCA and 5D HACACONH. There is no obvious correlation between $n$ and either $d_{av}$ or $d_{min}$, indicating that close approach of peaks is distributed statistically and depends on particular properties of the
protein, irrespective of its size. In Figure 1.7, the statistical probability to encounter pairs of peaks that could not be resolved by APSY is less than 1% for protein sizes up to at least 300 residues, which is representative for 4D and 5D triple resonance data sets that contain one peak per residue. Foreseeably, sensitivity of signal detection will therefore be a more stringent limitation for APSY applications than spectral crowding. From our experience to-date, projection spectra with a signal-to-noise ratio of approximately 3:1 are required for efficient use of APSY with automated peak picking.

For the few expected closely spaced pairs of peaks, APSY is in a good position to resolve potential difficulties, since it is not required that a given N-dimensional resonance is found in all projection spectra. Peaks with overlap in one or several projections will usually be resolved in many other projections (Fig. 1.4). Similar to GFT-NMR (Kim and Szyperski, 2003), APSY is also well prepared to deal with inaccurate peak positions from automated peak picking, which may arise from peak overlap. Since the final N-dimensional APSY peak list is computed as the average of a large number of measurements (Fig. 1.3), inaccurate peak positions in some of the projections have only a small influence on the overall precision.

**APSY and spectral artifacts.** The positions of the peaks belonging to a peak subgroup (Fig. 1.3) are correlated in all the projection spectra, whereas, in contrast, the positions of random noise are uncorrelated. This different behavior of the two types of peaks, which cannot readily be distinguished in the initial automatic peak picking, efficiently discriminates against artifacts. Artifacts are therefore unlikely to pass through the ranking filters (Fig. 1.3) and thus to appear in the final peak list. For the two APSY applications presented in this study, the
final peak lists contained no artifacts, although the initial peak picking routine was applied with low signal-to-noise thresholds, $R_{\text{min}}$, and included a large number of spurious signals in the initial peak lists. This result supports that the presently used ranking criteria (Fig. 1.3) were well chosen.

Outlook to future implementations of APSY. The present set-up of an APSY experiment (Fig. 1.3) requires the operator to define seven parameters, in addition to the selection of the $N$-dimensional experiment, the number and dimension of the projections, and the projection angles. These are the signal-to-noise threshold, $R_{\text{min}}$, for ATNOS peak picking, the thresholds for the two ranking filters, $S_{\text{min},1}$ and $S_{\text{min},2}$, the number of subgroup list calculations, $k$, the number of final peak coordinates calculations, $w$, the minimal distance for intersections, $r_{\text{min}}$, and the intersection tolerance in the direct dimension, $\Delta v_{\text{min}}$. For routine use of APSY, optimized sets of these parameters will emerge, and some of the free variables might be replaced by novel convergence criteria. Further refinements will also apply to selecting optimal numbers of and dimensions for the projections, as well as to identifying optimal combinations of projection angles for given experimental situations. This may include consideration of the different relaxation properties of the nuclei involved, leading to the use of projection angles with short evolution times for the most rapidly relaxing spins. Further technical improvements may lead, for example, to higher spectral resolution by linear prediction, or to improved spectrometer time usage by concurrent data acquisition and data analysis, with feedback to prevent unneeded data accumulation once the final peak list converges.

The sensitivity of APSY experiments

A prerequisite for optimal results of NMR experiments is the selection of the most sensitive pulse schemes for a given purpose and the optimal choice of acquisition, processing and analysis parameters. These parameters and the best pulse sequences are however not yet known for APSY NMR applications. The present chapter derives equations for the sensitivity of projection spectra and discusses how these results can be used to optimize the performance of APSY NMR. The theory of sensitivity of NMR spectra in general is well established and the present chapter follows the notation used in Ernst et al. (1987), where applicable.

For the purpose of the analysis, APSY NMR experiments are classified according to Figure 1.8. For the first step in the classification, the nucleus of detection, $A$, is used. For each such nucleus, usually a large number of different APSY NMR experiments are imaginable. The second step in the classification is done according to the experiment, $m$, and the third according to the projection angles. For an $N$-dimensional experiment, the projection angles are summarized in the $(N-2)$-dimensional vector $\vec{\phi} = (\alpha, \beta, \gamma, \ldots)$. Without loss of generality, the present discussion is limited to 2D projections. The extension of the result to 3D or higher-dimensional projections is straightforward.

![Figure 1.8](image_url)

**Figure 1.8.** Classification of APSY NMR experiments. The set of all APSY NMR experiments (outermost box) is divided into subsets according to the nucleus of acquisition $A$, which in turn are subdivided into different experiments $m$. The experimental input for each experiment $m$ is a set of projections with variable projection angles $\vec{\phi}$ (innermost boxes).
The orientation of 2D cross-sections of an \(N\)-dimensional experiment \(m\) with respect to the time domain axes is described by two \(N\)-dimensional Cartesian vectors \(\vec{p}_1(\vec{\phi})\) and \(\vec{p}_2\) for the indirect and the direct dimension, respectively (Eq. 1.1, Table 1.1). The \(N\)-dimensional time domain, \((t_1, t_2, \ldots, t_N)\), is sampled at discrete points on the hyperplane \(t \cdot \vec{p}_1(\vec{\phi}) + t_N \cdot \vec{p}_2\), yielding a complex time domain signal \(s_m(\vec{\phi}, t, t_N)\), where \(t\) is a time parameter along \(\vec{p}_1(\vec{\phi})\) in the \((N-1)\)-dimensional time domain subspace of indirect dimensions. All cross-sections start at the time domain origin and the signal amplitude at the time origin of experiment \(m\), \(s_m(0)\), is thus independent of \(\vec{\phi}\) (Fig. 1.9).

![Figure 1.9](image)

**Figure 1.9.** Schematic representation of data sampling and processing for projections of two indirect dimensions \(x\) and \(y\). (a) The time domain is sampled along cross-sections starting at the origin. The green and the red dots are sample points for projections along the dimensions \(t_x\) and \(t_y\), respectively. The black dots are sample points for two projections with angles \(\alpha_1\) and \(\alpha_2\). (b) Signal envelopes \(s'(t)\) along the time domain cross-sections. The relaxation rate of the signal depends on the projection angle (see text) (c) After Fourier transformation the NMR spectrum \(I(\omega)\) is obtained. The intensity at peak maximum depends on the relaxation rates of the signal in (b) and on the processing parameters.

With the currently available technology, it is not possible to record projections combining the acquisition dimension and indirect dimensions. The time domain signal of a single resonance can thus be factorized into expressions for the direct and the indirect dimensions. For a 2D projection of an APSY-NMR experiment \(m\) with projection angles \(\vec{\phi}\), it can thus be described by

\[
s_m(\vec{\phi}, t, t_N) = s_m(0) \cdot s^x_m(\vec{\phi}, t) \cdot \exp[i\Omega(\vec{\phi}) t] \cdot s^t_m(t_N) \cdot \exp[i\Omega_{\text{d}} t_N]
\]

(1.7)

where the term \(\exp[i\Omega_{\text{d}} t_N]\) describes the oscillation in the direct dimension at a frequency \(\Omega_{\text{d}}\) and \(\exp[i\Omega(\vec{\phi}) t]\) is the oscillation at a frequency \(\Omega(\vec{\phi})\) in the indirect dimension (This dependency of the frequencies on the angles is exploited by the algorithm GAPRO, which determines the resonance frequencies of the individual dimensions from the Fourier
transformed signal). The functions $s_n^e(t_N)$ and $s_m^e(\phi, t)$ are the signal envelopes in the direct and indirect dimension, respectively. The envelope function in the direct dimension is usually exponentially dependent on the transverse relaxation rate of the acquisition nucleus $A$. The envelope function in the indirect dimension is a projection-angle–dependent composition of the relaxation along the dimension axes. It is shown in the appendix that $s_m^e(\phi, t)$ is monoexponential, if the relaxation is monoexponential in all indirect dimensions. The signal envelopes have a strong impact on the experimental sensitivity of a projection (Figure 1.9(c)).

After acquisition, the complex time domain signal $s_m^e(\phi, t, t_N)$ is multiplied with a window function $h_m(\phi, t, t_N)$ that can be chosen differently for each experiment $m$ and for the angles $\phi$. Subsequent Fourier transformation with respect to $t$ and $t_N$ yields a resonance peak signal with frequencies $\Omega(\phi)$ and $\Omega_A$ in the 2D projection spectrum (Ernst et al., 1987). Its intensity at peak maximum, $S_m^e(\phi)$, is given by

$$S_m^e(\phi) = s_m^e(0) \cdot n(\phi) \cdot \frac{M(\phi)}{M_A} \cdot \frac{t_{\text{max}}(\phi)}{t_{\text{max,A}}} \int_0^{t_{\text{max}}(\phi)} \int_0^{t_{\text{max,A}}} dt_N s_m^e(\phi, t) \cdot s^e_A(t_N) \cdot h_m(\phi, t, t_N)$$ (1.8)

where $M(\phi)$ and $M_A$ are the number of complex points in the indirect and direct dimension, respectively, $t_{\text{max}}(\phi)$ and $t_{\text{max,A}}$ are the maximal evolution times in the indirect and direct dimension, respectively and $n(\phi)$ is the number of scans. $M(\phi)$, $n(\phi)$ and $t_{\text{max}}(\phi)$ can be chosen differently for each $\phi$.

As the signal function is a product of the direct and the indirect dimension, it is convenient to use a window function of the form $h_m(\phi, t, t_N) = h_m(\phi, t) \cdot h_A(t_N)$, which is a product of two window functions $h_m(\phi, t)$ and $h_A(t_N)$ for the indirect and direct dimension, respectively. This allows a separation of the acquisition and processing parameters for the direct dimension, which can then be optimized independently of the projection angles. The expressions $\overline{h^2}(\phi)$ and $\overline{h^2_A}$ describe the weighting of thermal noise by the window functions (Table 1.2, (Ernst et al., 1987)).

$$\overline{h^2}(\phi) = \frac{1}{t_{\text{max}}(\phi)} \int_0^{t_{\text{max}}(\phi)} dt \ h(\phi, t)^2$$ (1.9)

$$\overline{h^2_A} = \frac{1}{t_{\text{max,A}}} \int_0^{t_{\text{max,A}}} dt_N \ h_A(t_N)^2$$ (1.10)
Table 1.2. The thermal noise weighting factor $\bar{h}^2$ for common window functions.

<table>
<thead>
<tr>
<th>Window function</th>
<th>$h(t)$</th>
<th>$\bar{h}^2 = \frac{1}{t_{\text{max}}} \int_0^{t_{\text{max}}} dt \ h(t)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosine</td>
<td>$\cos \left( \frac{\pi}{2} \cdot \frac{t}{t_{\text{max}}} \right)$</td>
<td>0.5</td>
</tr>
<tr>
<td>Cosine square</td>
<td>$\cos^2 \left( \frac{\pi}{2} \cdot \frac{t}{t_{\text{max}}} \right)$</td>
<td>0.375</td>
</tr>
<tr>
<td>Sine shifted by phase $\phi$</td>
<td>$\sin \left( \phi - (\phi - \pi) \cdot \frac{t}{t_{\text{max}}} \right)$</td>
<td>$0.5 + \sin 2\phi / 4(\pi - \phi)$</td>
</tr>
<tr>
<td>Sine square shifted by phase $\phi$</td>
<td>$\sin^2 \left( \phi - (\phi - \pi) \cdot \frac{t}{t_{\text{max}}} \right)$</td>
<td>$0.375 + \sin 2\phi / 4(\pi - \phi) - \sin 4\phi / 32(\pi - \phi)$</td>
</tr>
<tr>
<td>Hamming window</td>
<td>$0.54 + 0.46 \cdot \cos \left( \pi \cdot \frac{t}{t_{\text{max}}} \right)$</td>
<td>0.3974</td>
</tr>
<tr>
<td>Exponential with line broadening factor $L$</td>
<td>$\exp \left( - \pi \cdot \frac{L}{t} \right)$</td>
<td>$\frac{[1 - \exp \left( -2 \cdot \pi \cdot \frac{L}{t_{\text{max}}} \right)]}{(2\piLt_{\text{max}})}$</td>
</tr>
<tr>
<td>Sine shifted by $75^\circ$</td>
<td>$\sin \left( \left(5 + 7 \cdot \frac{t}{t_{\text{max}}} \right) \frac{\pi}{12} \right)$</td>
<td>0.5682</td>
</tr>
<tr>
<td>Sine square shifted by $75^\circ$</td>
<td>$\sin^2 \left( \left(5 + 7 \cdot \frac{t}{t_{\text{max}}} \right) \frac{\pi}{12} \right)$</td>
<td>0.4580</td>
</tr>
</tbody>
</table>

The sensitivity of an NMR spectrum is defined by the ratio of signal amplitude at peak maximum and the root mean square amplitude of thermal noise. It is thus commonly called signal-to-noise ratio. The root mean square noise amplitude $\sigma(\bar{\phi})$ for the 2D projection spectra is given by

$$
\sigma(\bar{\phi}) = \sqrt{2^n n(\bar{\phi}) M(\bar{\phi}) \bar{h}^2(\bar{\phi}) \bar{h}^2 M_A \rho \frac{1}{t_{\text{max},A}}} \tag{1.11}
$$

where $\rho$ is the square root of the frequency-independent power spectral density (Ernst et al., 1987). $q$ is the number of angles in the vector $\bar{\phi}$ that are not $0^\circ$ or $90^\circ$ and $2^q$ is the resulting number of subspectra that need to be combined to obtain the pure sine and cosine modulated terms for the Fourier transformation (Brutscher et al., 1995; Kupce and Freeman, 2004b). $n(\bar{\phi})$ is the number of scans, with which each of the subspectra was recorded.

An expression for the sensitivity of a projection, $[S / \sigma]^m(\bar{\phi})$, is obtained from a division of the signal (Eq. 1.8) by the noise (Eq. 1.11). The resulting equation can be written as a product of three terms, $K_A$, $s_m(0)$ and $f_m(\bar{\phi})$, where each term corresponds to one classification level of the experiments (Figure 1.8):

$$
[S / \sigma]^m(\bar{\phi}) = K_A \cdot s_m(0) \cdot f_m(\bar{\phi}) \tag{1.12}
$$

The three factors in Equation 1.12 are discussed in the following.
1.) $K_A$ contains all factors depending on the nucleus of detection, $A$.

$$K_A = \frac{1}{\rho \sqrt{h_{\text{max},A}}} \int_0^{t_{\text{max},A}} dt_N \, s^\rho_A(t_N) h_A(t_N)$$ (1.13)

The sensitivity of the probe head and the field strength is included in $\rho$. For each nucleus of detection $A$ on a given NMR instrument, the optimal acquisition time $t_{\text{max},A}$ and the optimal window function $h_A(t_N)$ can be found to maximize $K_A$.

2.) $s_m(0)$ is the signal intensity of experiment $m$ at the time domain origin. Knowledge of $s_m(0)$ allows a comparison of the relative sensitivity of different APSY experiments and thus the selection of the most sensitive experiments. Values for $s_m(0)$ can be obtained either experimentally, e.g. from a comparison of the 1D spectra that correspond to the time domain origin of the different experiments. Alternatively, model calculations can be done to derive numerical estimates for $s_m(0)$.

3.) $f_m(\phi)$ describes the dependence of the sensitivity of an experiment $m$ on the projection angles and the acquisition and processing parameters.

$$f_m(\phi) = \frac{1}{(\sqrt{2})^d} \frac{\sqrt{h(\phi) \cdot M(\phi)}}{h(\phi)} \cdot \frac{1}{t_{\text{max}(\phi)}} \int_{t_{\text{max}(\phi)}}^{t_{\text{max}(\phi)}} dt \, s^\phi_m(\phi, t) \cdot h(\phi, t)$$ (1.14)

Equation 1.14 allows the optimal choice of acquisition and processing parameters. The main factor influencing $f_m(\phi)$ comes from the signal envelope in the indirect dimension, $s^\phi_m(\phi, t)$. It is shown in the appendix that monoexponential relaxation in all indirect dimensions leads to a monoexponential envelope $s^\phi_m(\phi, t) = \exp\left(-R^*_2(\phi)t\right)$ with an effective relaxation rate

$$R^*_2(\phi) = \bar{\rho}_1(\phi) \cdot \bar{R}_{2,m}$$ (1.15)
where $\bar{R}_{2,m}$ is an $N$-dimensional vector containing the transverse relaxation rates for each dimension. In the analysis with GAPRO, the set of projections is analyzed with equal weight for each projection and it is therefore desirable to have similar sensitivities in all projections. In this case, the selection of the most sensitive experiment among a group of experiments that share the same detection nucleus $A$ reduces to a direct comparison of the $s_m(0)$ values. If the projection-angle–dependency of $s_m^*(\tilde{\phi}, t)$ is known, equation 1.14 can be used to produce similar sensitivities for all projections by variation of the user-defined parameters $n(\tilde{\phi})$, $M(\tilde{\phi})$, $h_m(\tilde{\phi}, t)$ and $t_{\text{max}}(\tilde{\phi})$. This is particularly important for experiments, which involve nuclei with strongly different relaxation properties. The sensitivity decrease by $\sqrt{2}^q$ due to the combination of subspectra, can be corrected conveniently with $n(\tilde{\phi})$.

Summarized, the theory developed in this chapter allows the calculation of the sensitivity of individual projections used for APSY experiments and thus the selection of the most sensitive experiments out of a group of experiments and the sensitivity management for a set of projections from the same experiments. Besides these options, a further possibility to optimize APSY experiments is to record some projections, which do not require the full length of the magnetization pathway using shorter pulse sequences. The relative sensitivity of the resulting spectra can also be quantified and adjusted with the present theory.
A library of APSY experiments for protein backbone NMR assignments

The method APSY provides a general tool for the analysis of sets of projection spectra derived from all kinds of multidimensional NMR experiments. A promising application seems to be sequence-specific backbone resonance assignment of proteins, as in triple-resonance correlation experiments for the protein backbone a discrete number of peaks is expected and the signal intensities are only of minor interest. The present chapter aims at the creation of a systematic library of APSY experiments for the protein backbone atoms HN, N, CO, CA, HA and the side chain atoms CB and HB (Figure 1.10).

Figure 1.10. Definition of the notation used in the library of APSY protein backbone experiments. A dipeptide fragment from a protein backbone is shown. The atoms are represented with grey boxes and the relative amino acid position are indicated. Important $J$-coupling constants are indicated by arrows between the nuclei. The Greek letters denote the time periods used in the library elements with evolution due to the corresponding $J$-couplings. For transfers between heavy atoms with different relaxation properties, two Greek letters are given. In the corresponding elements, the letter is used, which is shown here next to the grey box corresponding to the nucleus with a product operator component in the x/y-plane.

Triple resonance experiments for sequence-specific resonance assignment of the protein backbone are numerous and have been developed for many years (Bax and Grzesiek, 1993; Kay, 1995; Cavanagh et al., 1996; Wider, 1998; Sattler et al., 1999). In the majority of the experimental approaches, the nucleus of detection is the amide proton. Following this experience, the present analysis focuses on experiments with signal detection on HN. Alternatives are detection on the protons HA or HB, or on the carbon atoms CO, CA or CB. The proximity of the water resonance makes detection on HA and to less extent also detection on HB problematic. Detection on HB is further hampered by homonuclear $J$-couplings to other aliphatic protons and by the frequent occurrence of non-degenerate chemical shifts for
the HB atoms in CH₂ groups of proteins (Wüthrich, 1986). Detection on ¹³C has a reduced sensitivity compared to ¹H due to the lower gyromagnetic ratio, however, the resulting shorter pulse sequences could possibly compensate for this deficiency. Several interesting experiments with carbon-detection have recently been developed (Oh et al., 1988; Eletsky et al., 2003; Bermel et al., 2006) and the discussion in this chapter can easily be extended to such experiments. Detection on ¹⁵N seems not a promising alternative due to its small gyromagnetic ratio.

For a systematic analysis of the NMR experiments, a library of pulse program elements was created using a consistent notation. The elements of the library are time periods such as INEPT transfer steps (Morris and Freeman, 1979) or evolution periods (Aue et al., 1976). 28 different transfer elements (numbers 1–28) for the backbone atoms and 10 evolution periods (numbers E1–E10) were analyzed and are listed in the Appendix. As an example, transfer element 1 is shown in Figure 1.11. Element 1 is the INEPT step converting amide proton polarization into longitudinal two-spin order of amide proton and nitrogen. In the product operator formalism the two operators connected by element 1 are given by HNᵀ_z and -2HNᵀ_zNᵀ_z, respectively (Sørensen et al., 1983). Normalization constants and signs of the operators have been omitted here for clarity of the notation. The normalization constants of product operators are 2⁽ⁿ⁻¹⁾, where n is the number of operators in the product. As all currently investigated APSY NMR experiments start from single nuclei polarization and are detected as in-phase coherence on single nuclei, omitting of the normalization constants for the product operators does not lead to wrong results for calculations of complete experiments.

**Figure 1.11.** Example for a transfer element from the library (Appendix). Element 1, the INEPT step between amide proton and nitrogen is shown. Indicated are (a) possible coherence transfers with this element, (b) the relevant J-coupling and relaxation constants and the transfer amplitude using the simplifications discussed in the text, (c) practical aspects for the adjustment of delays and (d) the pulse scheme for the element using standard pulse sequence notation.
Each transfer element is associated with a transfer function, which gives the fraction of the magnetization or coherence that is transferred by the element into the target magnetization or coherence. The transfer function can thus assume values between 0 and 1. Values for the transfer functions of the different elements can be determined experimentally for a given protein. An alternative are model calculations using analytical expressions (Figure 1.11(b)). The calculation of such transfer functions \textit{ab initio} is difficult and thus several simplifications have been made to yield the analytical functions given in the library: (i) All 90° and 180° pulses are ideal pulses with unlimited bandwidth for the nuclei of interest but without off-resonance effects on other nuclei. (ii) Magnetization losses occur only by transverse relaxation and from incompletely evolved $J$-couplings. (iii) The relaxation rates of product operators is written as the sum of the relaxation rates of the individual nuclei and the relaxation of longitudinal operators is neglected, \textit{e.g.} $R_2(\text{HN},N_z) = R_2(\text{HN}) + R_2(N)$ and $R_2(\text{HN},N_y) = R_2(\text{HN})$. (iv) Short delays that may be needed experimentally due to hardware requirements are set to 0. These assumptions are a good approximation to first order (Sattler et al., 1999). It would be desirable to have expressions closer to the experimental reality, but this is not the scope of the present work.

The 28 transfer elements of the library connect various product operators of backbone atoms following the scheme of Figure 1.12. For example, element 1 from Figure 1.11 and the operator $\text{HN}'_z$ are located at the lower end in the middle of Figure 1.12. The 10 evolution elements $E_1$–$E_{10}$ do not change the product operator and are therefore not appearing in Figure 1.12. Whereas most elements of the library are familiar to the experienced NMR spectroscopist, the recording of CB in APSY experiments shall be briefly discussed here. In conventional correlation experiments for the CB chemical shift, such as the 3D HNCACB or 3D CBCA(CO)NH experiment (Grzesiek and Bax, 1992b; Wittekind and Mueller, 1993), the initial intensity is usually distributed on two peaks, which correlate the nuclei $\text{HN}–\text{N}–\text{CA}$ and $\text{HN}–\text{N}–\text{CB}$, respectively. This is an undesirable situation for APSY, even more, if the two peaks are of opposite signs, because the superposition of positive and negative signs in projections leads to signal deletions and thus hinders the analysis with GAPRO. It seems more convenient to design the APSY experiments in a way to record a single $\text{HN}–\text{N}–\text{CA}–\text{CB}$ peak encoding the CA and CB chemical shift in two different dimensions. Compared to the conventional situation, the total number of peaks is reduced by about 50% and because of the higher dimensionality, the peak density in the $N$-dimensional space is further reduced (Figure
The price to pay are longer delay times and thus less magnetization transfer in the transfer elements 9b and 14b, compared to their classical analogs 9 and 14.

Figure 1.12. Overview of possible magnetization pathways for the protein backbone including CB and HB. The terms are described using the product operator formalism without normalization constants, as discussed in the text. The definitions in Figure 1.10 have been used. Numbers denote the transfer element number as given in the Appendix. Greek letters denote the relevant time periods for the transfer function. Circles denote elements, where the magnetization is split into two pathways.

The creation of NMR experiments from the element library is straightforward. In a first step, the elements of the desired magnetization transfer pathway have to be added sequentially. In a second step, the evolution periods have to be introduced. In APSY experiments, evolution periods can be introduced for each nucleus that is covered along the magnetization pathway without reducing the signal intensity at the time domain origin, \( s_m(0) \). Two common schemes, the constant-time and the semi-constant time element are usually used for this purpose (Bax et al., 1979; Grzesiek et al., 1993a). As the time increments for the indirect dimensions depend
on the projection angles, it is usually necessary to use evolution elements that switch between
the two modes. Figure 1.13 shows how the evolution periods are introduced into the INEPT
steps from the library.

For a complete pulse sequence, the following features have to be accounted for: Pulsed field
gradients acting in z-filters can be introduced between the elements; decoupling pulses and
broad-band decoupling schemes should be adjusted; phase cycling schemes for cancellation
of artifacts should be added, and finally, the water handling and suppression has to be
included in the pulse scheme. As an example for such a pulse sequence, the 4D APSY-
HNOCA experiment consisting of elements 1, 2, 16, E2, 16, 3 and 4 is shown in Figure
1.14.

In this way, a large number of experiments can be built up from the library elements. Table
1.3 lists a selection of the most promising APSY backbone experiments. It is not surprising
that the conventional counterparts of most of these experiments are well known since many
years. The signal intensity at time domain origin, \( s_m(0) \), of an experiment \( m \) consisting of \( k \)
transfer elements is given by

\[
s_m(0) = P(X) \cdot \prod_{i=1}^{k} T_m(i)
\]

(1.16)
where $T_m(i)$ denotes the transfer function of element $i$ and $P(X)$ is the steady state polarization of the nucleus X, on which the experiment starts. Using the elementary composition of the experiments and the analytical transfer functions of the library, transfer amplitudes for NMR experiments can thus be calculated. For example, the transfer amplitude $s_{\text{HINCOCA}}(0)$ is the product of the transfer functions of elements $1, 2, (16)^2, 3$ and $4$ and $P(HN)$.

$$s_{\text{HINCOCA}}(0) = P(HN) \cdot \sin^2(\pi J_{HN,N} \xi) \cdot \exp(-2\xi R_2(HN)) \cdot \sin^2(\pi J_{N,CO} \tau) \cdot \exp(-2\tau R_2(N)) \cdot \sin^2(\pi J_{CA,CO} \theta) \cdot \exp(-2\theta R_2(CO))$$

Figure 1.14. Composition of the 4D APSY-HNCOCA experiment from elements 1, 2, 16, E2, 16, 3 and 4 (shaded segments). In addition to these elements from the library, phase cycling, pulsed field gradients, decoupling schemes and water suppression techniques were added to yield the complete experiment. Evolution schemes according to Figure 1.13 have been introduced in element 2 and the first appearing element 16. The resulting pulse sequence is identical to the Figure 1.5, where experimental details can be found. Product operators are given below the gradients before and after each transfer element.
Table 1.3. Composition of selected APSY NMR experiments from the library elements. Evolution periods that do not contribute to the transfer function have been omitted.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Elements</th>
<th>Experiment</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>4D APSY-COHNC</td>
<td>1.19.22.4</td>
<td>4D APSY-iHNCA</td>
<td>1.17.15.15.22.4</td>
</tr>
<tr>
<td>5D APSY-COHNCACO</td>
<td>1.19.15.15.22.4</td>
<td>or</td>
<td>1.17.12.12.22.4</td>
</tr>
<tr>
<td>or</td>
<td>1.19.12.12.22.4</td>
<td>or</td>
<td>1.17.15.15.18.4</td>
</tr>
<tr>
<td>5D APSY-COHNCABA</td>
<td>1.19.21.21.22.4</td>
<td>or</td>
<td>1.17.12.12.18.4</td>
</tr>
<tr>
<td>5D APSY-COHNCACB</td>
<td>1.19.14.14.22.4</td>
<td>3D APSY-HNCA</td>
<td>1.5.6.4</td>
</tr>
<tr>
<td>or</td>
<td>1.19.14b.14b.22.4</td>
<td>4D APSY-HNCA</td>
<td>1.5.15.15.6.4</td>
</tr>
<tr>
<td>4D APSY-HNCOC</td>
<td>1.2.16.16.3.4</td>
<td>or</td>
<td>1.5.12.12.6.4</td>
</tr>
<tr>
<td>5D APSY-HNCOCAB</td>
<td>1.2.16.21.21.16.3.4</td>
<td>4D APSY-HNCACB</td>
<td>1.5.14.14.6.4</td>
</tr>
<tr>
<td>or</td>
<td>1.2.16.14b.14b.16.3.4</td>
<td>or</td>
<td>1.5.14b.14b.6.4</td>
</tr>
<tr>
<td>3D APSY-iHNCA</td>
<td>1.17.22.4</td>
<td>4D APSY-HACANH</td>
<td>7.11.6.4</td>
</tr>
<tr>
<td>or</td>
<td>1.17.18.4</td>
<td>5D APSY-HACACONH</td>
<td>7.23.10.3.4</td>
</tr>
<tr>
<td>4D APSY-iHNCAB</td>
<td>1.17.14.14.22.4</td>
<td>5D APSY-CBCACONH</td>
<td>7.8.9b.10.3.4</td>
</tr>
<tr>
<td>or</td>
<td>1.17.14b.14b.22.4</td>
<td>4D APSY-CBCANH</td>
<td>7.8.13b.6.4</td>
</tr>
<tr>
<td>or</td>
<td>1.17.14.14.18.4</td>
<td>6D APSY-seq-HNCOCANH</td>
<td>1.2.16.20.6.4</td>
</tr>
<tr>
<td>or</td>
<td>1.17.14b.14b.18.4</td>
<td>3D APSY-HNCO</td>
<td>1.2.3.4</td>
</tr>
</tbody>
</table>

Summarized, the library of elements established in this chapter allows a systematic coverage and comparison of APSY backbone NMR experiments. With the transfer functions, estimates for the sensitivities of different experiments can be obtained, allowing the selection of the most promising approaches for a given problem. The library can be extended modularly and each transfer element can be optimized separately or be replaced by an improved version.
Automated backbone resonance assignment of proteins

The experience gained so far with APSY indicates that high-dimensional, virtually artifact-free and highly precise peak lists can be obtained from the automated analysis of projection spectra with GAPRO. It seems therefore to be a promising approach to use APSY in combination with a suitable assignment algorithm to achieve fully automated sequence specific resonance assignment of proteins (Fig. 1.15). Furthermore, as the runtime parameters of APSY can be chosen based on general considerations, it seems likely that standard parameter sets for a routine use can be found. Such setups could then form the first stage of protocols that lead to fully automated structure determination.

**Figure 1.15.** Flowchart of the general strategy employed in this work for automated sequence-specific backbone assignment with APSY. The assignment algorithm can be any automated method that performs sequence-specific resonance assignment using as sole input the amino acid sequence and a set of experimental peak lists, such as GARANT (Bartels et al., 1997).

It is the scope of the present chapter to identify sets of APSY NMR experiments for automated backbone assignment according to Figure 1.15 using the library of experiments derived in the previous chapter. Following the analysis in the first chapter, it can be assumed that if sufficient projections are recorded for a given experiment, the result from APSY is a precise and artifact-free peak list including all peaks appearing in the projection spectra and therefore, the experimental sensitivity of the single projections is the critical factor determining the size limit of the method. In the present chapter, we identify groups of experiments, called “strategies”, which allow an assignment of the backbone resonances and perform model calculations of experimental sensitivity to select the most promising strategies among them.

The current work uses APSY as the method of choice to obtain the \( N \)-dimensional peak lists from the projection spectra. Several alternative methods have recently been developed (Kim and Szyperski, 2003; Moseley et al., 2004; Eghbalnia et al., 2005; Luan et al., 2005; Malmodin and Billeter, 2005a; b). It is interesting to see that many different approaches can be chosen to determine the peak list of an \( N \)-dimensional NMR experiment from low-dimensional projection spectra. It is currently unclear, how these methods compare relative to
each other. Important factors for the applicability of these methods in routine setups are the overall sensitivity, the universality, the user-friendliness, the reproducibility, the requirements on computational time and the robustness versus spectral artifacts and user-errors such as misadjusted parameters. The present analysis of the most sensitive assignment strategies is transferable from APSY to other methods that use projection spectroscopy in combination with an assignment algorithm following the scheme of Figure 1.15.

For the process of sequence-specific resonance assignment of a protein backbone, two distinct steps are commonly recognized to be important (Güntert et al., 2000; Moseley et al., 2001; Baran et al., 2004). These are firstly the building of sequential fragments and secondly the mapping of these fragments onto the sequence. Even though a good assignment algorithm addresses these two points simultaneously, the data for the two steps may be acquired with different experiments. For the building of fragments, the overlap of one or more nuclei between the experiments is needed. Mapping of fragments onto the sequence is typically done based on chemical shift database information (Seavey et al., 1991). The CA chemical shift is important for the building of fragments and allows the unambiguous identification of Glycine residues. On the other hand, the CB chemical shift may not be required for the building of sequential connectivities, its knowledge increases the quality of sequence specific assignments strongly and an assignment of small fragments can be expected (Richarz and Wüthrich, 1978; Oh et al., 1988; Grzesiek and Bax, 1993; Güntert et al., 2000).

As introduced in the previous chapter, the present analysis focuses on experiments with signal acquisition on the amide proton. Therefore the two chemical shifts of the amide moiety, HN and N, are contained in each experiment and form one anchor to match peaks from different experiments for the building of fragments. For establishing the sequential connectivities between two of these amide anchors several possibilities exist and we classify the experimental strategies into 3 groups, I–III (Figure 1.16).

Group I includes all strategies in which a single multidimensional peak directly correlates two sequential amide anchors. Group II contains strategies, where the amide anchors are connected by two multidimensional peaks that overlap on two other atoms. These peaks can come from two different, but also from the same experiment. Group III consists of strategies, where the amide anchors are connected by two multidimensional peaks that overlap on one other atom. In principle, strategies similar to group II with overlap of three or four atoms are also imaginable, but such approaches seem not to be promising strategies due to the need for branched, insensitive experiments, and are therefore not further considered here.
In group II, three different atom combinations for the two overlapping atoms are possible. One of the two atoms is always the CA atom, and the second atom can be either CB, HA or CO. The corresponding strategies are denoted IIa, IIb and IIc. Using pairs of atoms without CA in strategy II is not a practical approach, since in the required magnetization transfer pathways, the CA atom would anyway be included. In strategies III, the only reasonable possibility for the overlap at a single atom is CA. Experiments overlapping on HA or CB atoms would include CA in the magnetization transfer pathway and the resulting strategy would be equivalent to IIa or IIb. Overlap on CO is in principle an alternative, but it would require an experiment correlating HN, N, CA, and CO. This experiment is very insensitive and the resulting strategy could be extended to the more robust strategy IIc without changing the sensitivity limiting experiment.

A second classification of the strategies depending in the recorded chemical shift information is done. Strategies, which determine the chemical shifts of the backbone atoms HN, N, CA and CO, but not the CB, are denoted “A” and strategies which additionally include chemical shift information on CB are denoted “B”. As introduced above, peak fragments are mapped onto the sequence based on comparisons of the measured chemical shifts with expected values from chemical shift data bases. The polypeptide chain may be divided by proline residues and dynamic regions into several segments, for which continuous peak fragments are obtained. In strategies of type A, the mapping of these fragments onto the sequence relies primarily from the glycine residues, which are the only residues that can be identified unambiguously based on their CA chemical shift. Depending on the number and sizes of the fragments and the

**Figure 1.16.** Classification of strategies for backbone assignment of proteins with multidimensional NMR experiments based on the amide moiety. Each colored shape stands for a multidimensional experiment that correlates the nuclei contained in the shape. The yellow areas indicate, for which nuclei the experiments overlap. The numbers on top define three main strategies I–III, including 3 variants, a–c, of strategy II. The capital letters B and A in the two rows indicate, whether the CB chemical shift is determined in the strategy or not, respectively.
relative positions of glycine residues, this assignment strategy may not suffice to assign the protein. Strategies of type B are superior in such situations, as the CB chemical shift information allows the recognition of further amino acids. Alanine residues can be unambiguously identified and serine and threonine residues can be distinguished from all other residues. Even though methods of type A can work out in many cases, in particular, if all expected resonances of the protein can be observed, we recommend the use of type B strategies in routine standard setups.

The robustness of each strategy is determined by the expected probability of the assignment algorithm to find the global assignment solution based on the input, and thus to a major part by the number of combinatorial assignment possibilities. Assuming the same experimental precision in all experiments, strategy I with one peak per amide group will be a more robust strategy than strategies II and III with two peaks per residue. Further, strategy II with two overlapping atoms is more robust than strategy III with one overlapping atom. A more precise quantification of the relative robustnesses of the three approaches can be achieved either based on theoretical considerations or with simulation runs for a certain assignment algorithm, but this is beyond the scope of the present work. We presently conclude that the robustness of the strategies decreases from I to III and assume that the three strategies IIa–c are of approximately equal robustness.

<table>
<thead>
<tr>
<th>Coupling</th>
<th>Value [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^1J_{HN,N})</td>
<td>93</td>
</tr>
<tr>
<td>(^1J_{N,CA})</td>
<td>10 (7–11)</td>
</tr>
<tr>
<td>(^2J_{N,CA})</td>
<td>7 (4–9)</td>
</tr>
<tr>
<td>(^1J_{N,CO})</td>
<td>14</td>
</tr>
<tr>
<td>(^1J_{CO,CA})</td>
<td>53</td>
</tr>
<tr>
<td>(^1J_{HA,CA})</td>
<td>140</td>
</tr>
<tr>
<td>(^1J_{HB,CB})</td>
<td>132</td>
</tr>
<tr>
<td>(^1J_{CA,CB})</td>
<td>37</td>
</tr>
</tbody>
</table>

In the next step, we identify groups of experiments that fulfill the different strategies. All transfer elements in the library use the \(J\)-couplings between the backbone atoms to transfer magnetization. The smallest \(J\)-couplings in the protein backbone are the two \(J_{N,CA}\)-couplings (Table 1.4) and therefore the bottleneck of all assignment strategies is the correlation of the
amide moiety with the sequential CA using the $^{1}J_{N,CA}$-coupling. The even weaker $^{2}J_{N,CA}$-coupling is not the limiting factor, as the sequential correlation can be obtained with experiments that use the strong and selective $J$-coupling pathway via the carbonyl.

However, the existence of the $^{2}J_{N,CA}$-coupling provides a complication for the intraresidual experiment, as $^{1}J_{N,CA}$ and the $^{2}J_{N,CA}$ are of similar size and it is not possible to experimentally discriminate between the two. Four basic experiments have been developed for the correlation of the sequential CA with the amide group (Figure 1.17). The HNCA and the COHNCA experiment result in two peaks per residue and the iHNCA and the DQ-HNCA experiment yield one peak per amide moiety. In the DQ-HNCA experiment, the resonance frequency in the CA dimension is the sum of the resonance frequencies of the sequential and the intraresidual CA nuclei. The four experiments can be extended by additional transfer elements to include more atoms, for example by addition of transfer elements 14b, E9 and 14b, the experiments HNCACB / COHNCA / iHNCACB / DQ-HNCACB result. The use of DQ-experiments is an interesting option for the resonance assignment, which however was not investigated further in the present work, because the available assignment routine GARANT can not use the resulting peak lists as input.

Another source of experimental alternatives is the possibility to implement some magnetization transfer pathways in either the “out-and-back” or the “out-and-stay”-variant. The corresponding pairs of such experiments are given in Table 1.5.

Based on these considerations, sets of experiments for the strategies I–III can be defined. Table 1.6 lists sets of experiments for each strategy using the classical variant for the sequential experiment (“HNCA”-variant, Fig. 1.17(a)) and the out-and-back variants of the experiments from Table 1.5. Starting from Table 1.5, all possible combinations of experiments
for the different strategies can be obtained by exchanging the sequential experiment with one of the versions b–c from Figure 1.17, or by using the alternatives from Table 1.6.

Table 1.5. Composition of backbone NMR experiments, for which an “out-and-back” and an “out-and-stay” version is possible (Bax and Grzesiek, 1993; Wider, 1998).

<table>
<thead>
<tr>
<th>out-and-back elements</th>
<th>out-and-stay elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNCAHA 1.5.21.E10.21.6.4</td>
<td>HACANH 7.11.6.4</td>
</tr>
<tr>
<td>HNCOCABA 1.2.16.21.E10.21.16.3.1</td>
<td>HACACONH 7.23.10.3.4</td>
</tr>
<tr>
<td>HNCOCACB 1.2.16.14.E9.14.16.3.1</td>
<td>CBCACONH 7.8.9.10.3.4</td>
</tr>
</tbody>
</table>

For a decision on the optimal strategy, sensitivity calculations are required. As discussed in the previous chapters, sensitivity comparisons of experiments that are detected on the same nucleus reduce to a calculation of the signal intensity at time domain origin, $s_m(0)$, which can be calculated from the transfer functions of the elements from the Appendix. Using the simplifying assumptions discussed in the previous chapter, these calculations were done using the $J$-coupling constants (Table 1.4) and the transverse relaxation rates $R_2$ for the different backbone nuclei (Table 1.7). Table 1.8 shows the result for all experiments for three different protein sizes.

Table 1.6. Experiments for the different assignment strategies I–III for automated backbone assignment with APSY (Fig. 1.16). The table lists only the regular variant for the sequential experiment (Fig. 1.17) and the out-and-back versions of the experiments (Table 1.5).

<table>
<thead>
<tr>
<th>Strategy</th>
<th>NMR Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I–A</td>
<td>6D APSY-seq-HNCO</td>
</tr>
<tr>
<td>I–B</td>
<td>6D APSY-seq-HNCOCA + 5D APSY-HNCO</td>
</tr>
<tr>
<td>IIa</td>
<td>4D APSY-HNCACB + 5D APSY-HNCO</td>
</tr>
<tr>
<td>IIb–A</td>
<td>4D APSY-HNCAHA + 5D APSY-HNCO</td>
</tr>
<tr>
<td>IIb–B</td>
<td>4D APSY-HNCAHA + 5D APSY-HNCOCA + 5D APSY-HNCO</td>
</tr>
<tr>
<td>IIc–A</td>
<td>4D APSY-HNCO + 4D APSY-HNCOCA</td>
</tr>
<tr>
<td>IIc–B</td>
<td>4D APSY-HNCOCA + 4D APSY-HNCOCA + 5D APSY-HNCO</td>
</tr>
<tr>
<td>III–A</td>
<td>3D APSY-HNCA + 4D APSY-HNCO</td>
</tr>
<tr>
<td>III–B</td>
<td>3D APSY-HNCA + 4D APSY-HNCOCA + 5D APSY-HNCO</td>
</tr>
</tbody>
</table>

42
Table 1.7. Relaxation rates $R_2$ (in $s^{-1}$) in proteins with rotation correlation times $\tau_c$ of 3 ns, 6 ns, and 10 ns, which roughly relate to masses of 6, 12 and 20 kDa under normal conditions. Bold values are from the literature, other values were extrapolated thereof, assuming a linear dependence of $R_2$ on $\tau_c$.

<table>
<thead>
<tr>
<th></th>
<th>$\tau_c = 3$ ns</th>
<th>$\tau_c = 6$ ns</th>
<th>$\tau_c = 10$ ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1H^a$ / HN</td>
<td>15</td>
<td>30</td>
<td>47$^d$, 51$^e$</td>
</tr>
<tr>
<td>$^1H^b$ / HA</td>
<td>25</td>
<td>50</td>
<td>82$^d$, 83$^e$</td>
</tr>
<tr>
<td>$^1H^b$ / HB</td>
<td>33</td>
<td>66</td>
<td>110$^f$</td>
</tr>
<tr>
<td>$^{13}C^a$ / CA</td>
<td>13</td>
<td>26</td>
<td>44$^d$, 41$^e$</td>
</tr>
<tr>
<td>$^{12}C^b$ / CB</td>
<td>22</td>
<td>44</td>
<td>74$^e$</td>
</tr>
<tr>
<td>$^{13}C^c$ / CO (500 MHz)</td>
<td>4.5$^a$, 5.0$^b$</td>
<td>10$^b$</td>
<td>17$^b$</td>
</tr>
<tr>
<td>$^{13}C^e$ / CO (900 MHz)</td>
<td>11$^b$</td>
<td>22$^b$</td>
<td>37$^b$</td>
</tr>
<tr>
<td>$^{15}N$ / N</td>
<td>3</td>
<td>6</td>
<td>10$^d$, 14$^e$</td>
</tr>
</tbody>
</table>

$^a$ (Allard and Härd, 1997)
$^b$ The equation $R_2(CO) / s^{-1} = \tau_c / 3 \cdot (B / T)^2 / 150$ was used.
$^c$ (Grzesiek, 2002)
$^d$ (Cavanagh et al., 1996)

Table 1.8. Relative sensitivities $s_{ex}(0)$ of the NMR experiments shown in Table 1.3 for proteins with different rotational correlation times $\tau_c$. For experiments with an intraresidual and a sequential peak (Fig. 1.17), the value refers to the more intense intraresidual peak, which is the peak of interest for the assignment (Fig. 1.16). The values were calculated using the analytical transfer functions from the element library and the values from Table 1.4 and 1.7. The steady state polarization of single protons was assumed to be independent of the chemical position, $P(HN) = P(HA) = P(HB)$ (Eq. 1.16).

<table>
<thead>
<tr>
<th>Rotational correlation time $\tau_c$</th>
<th>3 ns</th>
<th>6 ns</th>
<th>10 ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference: $[^{15}N,^1H]$-HSQC</td>
<td>85</td>
<td>73</td>
<td>62</td>
</tr>
<tr>
<td>3D-APSY-HNCO</td>
<td>69</td>
<td>49</td>
<td>32</td>
</tr>
<tr>
<td>4D-APSY-HNCOCA</td>
<td>63</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>3D-APSY-HNCA</td>
<td>29</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>3D-APSY-iHNCAC</td>
<td>48</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>5D-APSY-HNCOCAHA</td>
<td>49</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>4D-APSY-COHNCA</td>
<td>29</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>5D-APSY-HACACONH</td>
<td>34</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>5D-APSY-HNCOCA</td>
<td>45</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>4D-APSY-HNCA</td>
<td>23</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>4D-APSY-iHNCACH</td>
<td>37</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>5D-APSY-COHNCA</td>
<td>22</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>4D-APSY-HACANH</td>
<td>22</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>4D-APSY-HNCACO</td>
<td>21</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>4D-APSY-HNCA</td>
<td>21</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>4D-APSY-iHNCA</td>
<td>34</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>4D-APSY-iHNCAO</td>
<td>34</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>5D-APSY-COHNCAO</td>
<td>20</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>5D-APSY-COHNCA</td>
<td>20</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>5D-APSY-CBACONH</td>
<td>32–14</td>
<td>16–7</td>
<td>8–3</td>
</tr>
<tr>
<td>6D-APSY-seq-HNCOCAH</td>
<td>20</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>4D-APSY-CBCANH</td>
<td>10–5</td>
<td>6–3</td>
<td>3–1</td>
</tr>
</tbody>
</table>

$^a$ For these experiments, the intensity depends on the side chain type. The range gives the minimal and maximal value occurring for the 20 common amino acids (see also Appendix).
These calculations are a good guideline for the selection of the best experimental strategies in a given experimental situation. The limitations of the assumptions behind these model calculations, leading to large errors, have been discussed previously, so that relative sensitivity differences up to a factor of 2 are not significant.

For a set of experiments, the one with the smallest sensitivity is the limiting factor. As anticipated, the assignment strategies have increased sensitivity from I to III, in particular for larger proteins. For small proteins with a rotational correlation time of up to 3 ns, the sensitivities for all three classes of strategies are comparable (Table 1.6). For a protein of 10 ns, however, strong differences between the sensitivities exist and strategy I can not be expected to work. Among the strategies of type II, the most sensitive experiment seems to be the HNCAHA experiment with a relative sensitivity of 8, however, several other experiments with a value of 5 follow. As the assumptions behind these values are rather crude (see previous chapter) the best strategy should be determined experimentally in this case.

Based on these considerations, the following three chapters show applications of APSY for automated sequence-specific backbone resonance assignment. The first case shows automated backbone assignment of folded proteins using strategy I–A. In the second case, a folded protein was assigned using strategy IIb–B. The third case shows automated assignment of a denatured protein using strategies I–A and I–B. Together, the three following chapters demonstrate that automated backbone assignment with APSY and a suitable assignment algorithm is possible.
Automated backbone assignment of folded proteins using strategy I–A

This chapter is reproduced with minimal adaptations from:

Introduction
Projection NMR spectroscopy provides access to high-dimensional NMR experiments which would not, in practice, be available using conventional NMR, due to prohibitively long measurement times even at low digital resolution (Brutscher et al., 1995; Freeman and Kupce, 2004; Kim and Szyperski, 2004). In higher-dimensional experiments, peak overlap is reduced, and the information content increases, enabling more efficient and reliable data analysis (Szyperski et al., 2002). In the context of the present work it is of particular interest that higher dimensionality enables longer-range correlations to be established along polypeptide chains. Automated projection spectroscopy (APSY) data sets analyzed with the algorithm GAPRO (Hiller et al., 2005) enables the generation of accurate and virtually artifact-free peak lists of high-dimensional NMR experiments from low-dimensional projection spectra, without the need of reconstruction of the high-dimensional spectrum. Here, such high-quality peak lists have been used to achieve fully automated resonance assignment of proteins, by combining an appropriate APSY-NMR experiment with a suitable automatic assignment algorithm.

The 6D APSY-seq-HNCOCANH experiment connects two sequentially neighboring amide moieties in polypeptide chains via the carbonyl and Cα atoms, with the back transfer from Cα to the originating amide moiety suppressed in this sequential (seq) experiment (Figure 1.18). Thus, each peak from the 6D peak list generated by GAPRO correlates the resonance frequencies of two sequentially adjacent amide moieties. Clearly, this information will be adequate for the sequential assignment of the backbone resonances, provided that there is sufficient dispersion of the amide group chemical shifts. In this paper, we used the 6D peak lists thus obtained as input for the program GARANT (Bartels et al., 1997) in order to obtain NMR assignments for the two proteins 434-repressor(1–63) (Neri et al., 1992b) and TM1290 (Etezady-Esfarjani et al., 2004), in an approach which does not involve any further human interaction after starting the APSY-NMR experiments.
Materials and methods

NMR measurements. The 6D APSY-seq-HNCOCANH experiment with a 0.9 mM solution of 434-repressor(1–63) was recorded at 30°C on a Bruker DRX 750 MHz spectrometer equipped with a triple-resonance probehead with a z-gradient coil. The spectral widths were 3000 Hz, 1600 Hz, 1900 Hz and 5700 Hz in the $^1\text{H}$, $^{15}\text{N}$, $^{13}\text{C}'$ and $^{13}\text{C}\alpha$ dimensions, respectively. The interscan delay was 1.0 s, and 32 transients were accumulated for each increment of the combined evolution time. 1024 complex points were recorded in the acquisition dimension, with a sweep width of 11.0 ppm. In the indirect dimension, 64 complex points were measured. The total spectrometer time used for the recording of 25 2D projections (Table 1.9) was 40 h.

The corresponding experiment with a 3.0 mM solution of TM1290 was recorded at 35°C on a Bruker DRX 500 MHz spectrometer equipped with a triple-resonance cryogenic probehead with a z-gradient coil. The spectral widths were 2000 Hz, 1650 Hz, 1500 Hz, and 3800 Hz in the $^1\text{H}$, $^{15}\text{N}$, $^{13}\text{C}'$ and $^{13}\text{C}\alpha$ dimensions, respectively. The interscan delay was 1.0 s, and 8 transients were accumulated. 1024 complex points were recorded in the acquisition dimension, with a sweep width of 12.0 ppm. In the indirect dimension, 128 complex points were measured. The total spectrometer time used for the recording of 25 2D projections (Table 1.9) was 19.6 h.

![Figure 1.18](image)

**Figure 1.18.** Magnetization transfer in the 6D APSY-seq-HNCOCANH experiment. The pathway represented by the dashed arrows a to e (Equation (1.21)) leads to a single peak in the 6-dimensional frequency space. The dotted arrow indicates an undesired magnetization back transfer, which is suppressed by the experimental scheme shown in Figure 1.19 (see text).

Recording of projection spectra. For a 2D projection of the 6D APSY-seq-HNCOCANH experiment with the four projection angles $\alpha$, $\beta$, $\gamma$ and $\delta$, the time domain was sampled along a straight line defined by the unit vector $\vec{p}_1$,
\[ \tilde{p}_i = \begin{pmatrix} \sin \delta \\ \sin \gamma \cdot \cos \delta \\ \sin \beta \cdot \cos \gamma \cdot \cos \delta \\ \sin \alpha \cdot \cos \beta \cdot \cos \gamma \cdot \cos \delta \\ \cos \alpha \cdot \cos \beta \cdot \cos \gamma \cdot \cos \delta \\ 0 \end{pmatrix}. \] (1.16)

The spectral width, \( SW \), of each 2D projection was calculated from

\[ SW = \sum_{i=1}^{5} p'_i \cdot SW_i, \] (1.17)

where \( p'_i \) represents the coordinates of the vector \( \tilde{p}_i \) (see Equation (1.16)), and \( SW_i \) are the spectral widths of the five indirect dimensions. The dwell time for the recording of discrete data points, \( \Delta \), was calculated as

\[ \Delta = \frac{1}{SW}, \] (1.18)

and the resulting increments for the evolution times \( t_i, \Delta_i, \) in the five indirect dimensions are given by

\[ \Delta_i = p'_i \cdot \Delta. \] (1.19)

Quadrature detection was achieved using hypercomplex Fourier transformation on pure sine and cosine terms of corresponding positive and negative projection angles, which were obtained using the trigonometric addition theorem (Brutscher et al., 1995; Freeman and Kupce, 2004). In the 6D-seq-HNCOCANH experiment (Figure 2), the five evolution time periods \( t_i \) (\( i = 1, \ldots, 5 \)) were performed as semi-constant time or as constant time periods, depending on the value of the increment \( \Delta_i \), and on the number of complex points, \( n \), recorded in the indirect dimension. The initial values for \( t^a_i, t^b_i \) and \( t^c_i \) are given in the caption to Figure 1.19. For \( \Delta_i / 2 > t^c_i / n \), semi-constant time evolution periods were applied, with \( t^c_i \) being decremented by \( \Delta_i^c = t^c_i / n \), and \( t^a_i \) and \( t^b_i \) incremented by \( \Delta_i^a = \Delta_i / 2 \) and
\[ \Delta_j^b = \Delta_j^c - \Delta_j^e, \] respectively. In all other situations, constant time evolution periods were used, with \( t_i^e \) decremented by \( \Delta_i^e = \Delta_j^e / 2 \), \( t_i^a \) incremented by \( \Delta_i^a = \Delta_j^a / 2 \), and \( t_i^b \) maintaining its initial value.

### Table 1.9. Values of the projection angles \( \alpha, \beta, \gamma \) and \( \delta \) (see text), and of the spectral widths (\( SW \)) in the dimension \( \omega_{1.5} \) used here for the recording of 25 2D projections of the 6D APSY-seq-HNOCANH experiment. The resulting linear combination of frequencies are given in column LC.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \delta )</th>
<th>( SW [\text{Hz}] )</th>
<th>LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>1650</td>
<td>( \omega_5 )</td>
</tr>
<tr>
<td>90°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>3800</td>
<td>( \omega_4 )</td>
</tr>
<tr>
<td>0°</td>
<td>90°</td>
<td>0°</td>
<td>0°</td>
<td>1500</td>
<td>( \omega_5 )</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>90°</td>
<td>0°</td>
<td>1650</td>
<td>( \omega_2 )</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>90°</td>
<td>2000</td>
<td>( \omega_1 )</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>90°</td>
<td>±30°</td>
<td>2429</td>
<td>( \omega_2 \cos(30°) \pm \omega_1 \sin(30°) )</td>
</tr>
<tr>
<td>0°</td>
<td>90°</td>
<td>0°</td>
<td>±60°</td>
<td>2482</td>
<td>( \omega_3 \cos(60°) \pm \omega_1 \sin(60°) )</td>
</tr>
<tr>
<td>90°</td>
<td>0°</td>
<td>0°</td>
<td>±30°</td>
<td>4290</td>
<td>( \omega_4 \cos(30°) \pm \omega_1 \sin(30°) )</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>±60°</td>
<td>2557</td>
<td>( \omega_5 \cos(60°) \pm \omega_1 \sin(60°) )</td>
</tr>
<tr>
<td>0°</td>
<td>90°</td>
<td>±30°</td>
<td>0°</td>
<td>2124</td>
<td>( \omega_3 \cos(30°) \pm \omega_2 \sin(30°) )</td>
</tr>
<tr>
<td>90°</td>
<td>0°</td>
<td>±60°</td>
<td>0°</td>
<td>3229</td>
<td>( \omega_4 \cos(60°) \pm \omega_2 \sin(60°) )</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>±30°</td>
<td>0°</td>
<td>2254</td>
<td>( \omega_5 \cos(30°) \pm \omega_2 \sin(30°) )</td>
</tr>
<tr>
<td>90°</td>
<td>±60°</td>
<td>0°</td>
<td>0°</td>
<td>3199</td>
<td>( \omega_4 \cos(60°) \pm \omega_3 \sin(60°) )</td>
</tr>
<tr>
<td>±30°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>3329</td>
<td>( \omega_5 \cos(30°) \pm \omega_4 \sin(30°) )</td>
</tr>
</tbody>
</table>

\(^a\) The dimension \( \omega_{1.5} \) consists of projections of the five indirect dimensions \( \omega_3 (^1H) \), \( \omega_2 (^{15}N) \), \( \omega_3 (^{13}C) \), \( \omega_3 (^{15}C) \) and \( \omega_4 (^{15}N) \) (see column LC and the text).

**Processing of the 2D projection spectra.** All data processing steps were fully automated. Zero-order phase correction for the direct dimension was determined with PROSA (Güntert et al., 1992). The use of constant-time or semi-constant time evolution periods and proper phase settings in the pulse program (Figure 1.19) provides absorptive in-phase signals for the indirect dimension, without need for phase correction. In both dimensions, the FID was multiplied with a 75°-shifted sine bell prior to Fourier transformation (De Marco and Wüthrich, 1976), and zero-filled to the next power of two complex points. The baseline was corrected using the IFLAT method (Bartels et al., 1995a) in the direct dimension, and polynomials in the indirect dimension.
**APSY analysis.** The 2D projection spectra were automatically peak picked with the peak picking routine of ATNOS (Herrmann et al., 2002), with $R_{\text{min}} = 4.0$ for 434-repressor(1–63) and $R_{\text{min}} = 5.0$ for TM1290. The resulting peak lists of the projections were analyzed with the algorithm GAPRO (Hiller et al., 2005), which identifies all projected peaks arising from the same 6D peak, and then calculates the final 6D peak list. For 434-repressor(1–63), GAPRO was applied with the parameters $j = 25$, $S_{\text{min},1} = S_{\text{min},2} = 5$, $k = 50$, $w = 200$, $r_{\text{min}} = 50$ Hz, $\Delta\nu_{\text{min}} = 7.5$ Hz. The calculation time on a standard LINUX PC with a 2.8 GHz Pentium 4 processor was approximately 10 min. For TM1290, GAPRO was applied with the parameters $j = 25$, $S_{\text{min},1} = S_{\text{min},2} = 5$, $k = 100$, $w = 400$, $r_{\text{min}} = 50$ Hz, $\Delta\nu_{\text{min}} = 5.0$ Hz. The calculation time was approximately 30 min.

**Protocol and parameters used for automated assignment with GARANT.** The precise and artifact-free 6D peak lists obtained for the two proteins from the 6D-APSY-seq-HNCOCANH experiment were used as input for the assignment algorithm GARANT (Bartels et al., 1997), which had been modified by increasing the maximally allowed dimension from 4 to 6. GARANT was applied to the 6D peak lists using a standard annealing protocol (Bartels et al., 1997). For each peak list, 30 resonance assignment calculations were performed, using as input the same 6D-APSY-seq-HNCOCANH peak list but different, randomly chosen starting conditions. The 30 resulting sequence-specific resonance assignments were merged into one list. An assignment for a given atom was accepted if the same result, within 0.02 ppm and 0.4 ppm for protons and heavy atoms, respectively, was obtained in at least 50% of the calculations. Otherwise, the backbone atom was considered to remain unassigned.

**Results**

*The 6D APSY-seq-HNCOCANH NMR experiment.* This experiment connects sequentially adjacent amide moieties. Previous experimental schemes providing corresponding information include 3D and 4D HN(COCA)NH (Bax and Grzesiek, 1993), which also connect sequential amide moieties via $C^\alpha$ and the carbonyl carbon. With APSY, the full potential of this magnetization transfer pathway is exploited in a six-dimensional experiment (Figure 1.18).

The existing 3D and 4D versions give rise to an intraresidual and a sequential peak (Grzesiek et al., 1993b; Matsuo et al., 1996a). Since the intraresidual peak resulting from the “back transfer” magnetization pathway would not contribute any new information, it is efficiently
suppressed in the 6D APSY-seq-HNOCANH experiment, and only one peak for each pair of sequentially adjacent amide moieties appears in the 6D frequency space. In the coherence transfer scheme of Equation (1.20), the transfer steps of Figure 1.18 are represented by Cartesian product operators (Sørensen et al., 1983), where chemical shift evolution and transfer efficiencies are not represented, and only relevant magnetization components are retained.

\[
\begin{align*}
H_z^{N_i,j} & \rightarrow H_z^{N_i,j} & \rightarrow N_z^{i-1} C_z^{i-1} & \rightarrow N_z^{i} C_z^{i} & \rightarrow N_z^{i} C_z^{i,1} \\
\frac{J_{NH} + J_{13C}^{i}}{a} & \rightarrow N_z^{i} C_z^{i,1} & \rightarrow \frac{J_{NC}^{i} + J_{NH} + J_{13C}}{b} & \rightarrow \frac{J_{13C}^{i} + J_{NC}^{i}}{d} & \rightarrow \frac{J_{13C}^{i} + J_{NC}^{i}}{e}
\end{align*}
\]

Above the arrows which connect subsequent states, we indicate the active \( J \) coupling constants and the evolution periods \( t_i \), with the active nucleus in parentheses. Below the arrows, the letters a to e relate the magnetization transfer steps to the corresponding steps in Figure 1. The characters s, t, u, v and w at the end of the individual product operator expressions indicate time points in the experimental scheme shown in Figure 1.19. The suppression of the back transfer pathway (dotted line in Figure 1.18) is implemented between the time points u and v, where the one-bond scalar coupling \( J_{NC}^{i} \) as well as the two-bond coupling \( J_{NC}^{i} \) are active, leading to the terms \( C_z^{i-1} N_z^{i-1} \) and \( C_z^{i-1} N_z^{i} \) at time v. The term \( C_z^{i-1} N_z^{i-1} \) is associated with the sequential pathway and eventually leads to \( H_{x/y}^{N,j-1} \), which is the desired signal measured during the acquisition period. The term \( C_z^{i-1} N_z^{i} \) leads to back transfer and ends up as undesired magnetization on the amide proton from which the experiment was started, \( H_{x/y}^{N,j} \) (Panchal et al., 2001). The transfer efficiencies for the terms \( C_z^{i-1} N_z^{i-1} \) and \( C_z^{i-1} N_z^{i} \), respectively, are proportional to the expressions (1.21) and (1.22):

\[
\begin{align*}
\sin(\pi \cdot T \cdot 1 J_{NC}^{i}) \cdot \sin(\pi \cdot T \cdot 2 J_{NC}^{i}) & \quad (1.21) \\
\cos(\pi \cdot T \cdot 1 J_{NC}^{i}) \cdot \cos(\pi \cdot T \cdot 2 J_{NC}^{i}) & \quad (1.22)
\end{align*}
\]
Figure 1.19. 6D-APSY-seq-HNCOCANH experimental scheme, which correlates pairs of sequentially adjoining amide protons (Figure 1.18). The alternative out-and-back magnetization transfer pathway, which ends at the starting amide moiety, is efficiently suppressed (see text). Radio-frequency pulses were applied at 118.0 ppm for $^{13}N$, 173.0 ppm for carbonyl carbons ($^{13}C'$), and 54.0 ppm for $^{13}C\alpha$. At the outset of the pulse sequence ($^1H$ on the line $^1H$), the proton carrier frequency was set in the amide proton region at 8.2 ppm, and at the time point "H2O" the carrier was set to 4.7 ppm. Narrow and wide rectangular bars represent non-selective 90° and 180° pulses, respectively. Thin sine bells are shaped 90° pulses, and fat sine bells are shaped 180° pulses. Individual shaped pulses are identified with the capital letters A to F. The actual shapes depend on the purpose of the pulses, and the durations depend on the shape and the spectrometer frequency (here: 500 MHz): A, Gaussian shape, 150 µs; B, Gaussian, 180 µs; C, I-burp (Geen and Freeman, 1991), 300 µs; D, Gaussian, 120 µs; E, E-burp (Geen and Freeman, 1991), 400 µs; F, RE-burp (Geen and Freeman, 1991), 525 µs. The last six hard pulses on the $^1H$ line represent a 3-9-19 WATERGATE element (Sklenar et al., 1993). The four pulses marked with an asterisk are continuously centered with respect to the time periods $t^a + t^b$, $t^a + t^c$, and $t^a$, respectively. Grey pulses were applied for compensation of off-resonance effects (Bloch and Siegert, 1940; McCoy and Mueller, 1992). Decoupling using DIPSI-2 on $^1H$ (Shaka et al., 1988) and WALTZ-16 on $^{15}N$ (Shaka et al., 1983) is indicated by white rectangles. The triangle with $t_\phi$ represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped, pulsed magnetic field gradients applied along the z-axis, with the following durations and strengths: $G_1$, 700 µs, 13 G/cm; $G_2$, 1000 µs, 35 G/cm; $G_3$, 800 µs, 20 G/cm; $G_4$, 800 µs, 32 G/cm; $G_5$, 800 µs, 13 G/cm; $G_6$, 1000 µs, 35 G/cm; $G_7$, 800 µs, 18 G/cm. The following phase cycling was used: $\psi_1 = \{y, -y\}$, $\phi_2 = \{x, y, -x, -y\}$, $\phi_3 = \{x, -x, -x, x\}$, $\psi_4 = y$, $\psi_5 = x$, $\psi_6 = y$, $\psi_7 = y$, $\psi_8 = x$, $\psi_9 = x$, and all pulses without indication of a phase above the pulse symbol were applied with phase x. The time points s, t, u, v, and w are discussed in the text. The following initial delays were used: $t_1 = t_{1s} = 2.7 ms$, $t_2 = t_{2s} = 350 ms$, $t_3 = t_{3s} = 4.75 ms$, $t_4 = t_{4s} = 4.75 ms$, $t_5 = t_{5s} = 25.0 ms$, $t_6 = t_{6s} = 14.0 ms$, and $t_7 = t_{7s} = 0 ms$. The delay $\tau = 2.7 ms$ was invariant during the experiment, and the delay $\delta$ was continuously adjusted to $\delta = (t_{6s} + t_{7s} - t_{5s})/2$. In the five indirect evolution periods, constant-time or semi-constant-time periods were applied (see text). Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem to obtain pure cosine and sine terms for a subsequent hypercomplex Fourier transformation (Brutscher et al., 1995; Kupce and Freeman, 2004a). The phase pulses $\psi_1$, $\psi_2$, $\psi_3$, and $\psi_4$ were used for this purpose for the dimensions $\omega_3$ to $\omega_8$, respectively, and only the pulse phases of the evolution periods which are part of the given projection (Table 1.9) are phase cycled. For consecutive FIDs, $\psi_1$, $\psi_2$, $\psi_3$, and $\psi_4$ were simultaneously incremented in 90° steps, and $\psi_5$ was decremented in 90° steps.

The values of $1/(2J_{NC})$ and $1/(2J_{NC})$ are similar, and adjusting the time period $T = t_{1s} + t_{3s} + t_{4s}$ between u and v (Figure 2) to the average of these two values results in strong attenuation of the back transfer peak (Equations (1.20)–(1.22)). During the period $T$, the magnetization evolves also due to the scalar coupling between $\alpha$- and $\beta$-carbons, and values
of $T$ close to multiples of 27 ms are therefore preferred for optimal sensitivity. Overall, $T = 50$ ms was found to be a good compromise. Considering the variations of $J_{\text{NC}}$ and $J_{\text{NC}}$ along a polypeptide chain, the back transfer of magnetization (dotted line in Figure 1.8) can then reach at most 10% of the intensity of the corresponding desired peak (Brutscher, 2002). In the 6D APSY-seq-HNCOCANH experiment the undesired resonance was very well suppressed with $T = 50$ ms, retaining high sensitivity for the observation of the desired sequential signal (Figure 1.20).

**Figure 1.20.** Suppression of the $^{13}\text{C}_{\alpha} \rightarrow ^{15}\text{N}$ back transfer (Figure 1.18) in the 6D-APSY-seq-HNCOCANH spectrum of $[^{13}\text{C},^{15}\text{N}]-$labeled 434-repressor(1–63) (protein concentration = 0.9 mM, 20 mM sodium phosphate at pH 6.5, $T = 15^\circ$ C). Shown are 1D cross sections through the projection spectra recorded with the projection angles ($\alpha$, $\beta$, $\gamma$, $\delta$) = (0°, 90°, 0°, 0°), which were taken at the Arg 5 $^{15}\text{N}$-frequency of 125.06 ppm and cover the $^{1}\text{H}^\text{N}$ chemical shift range containing the intraresidual correlation to Arg 5 and the sequential correlation to Ser 4. The time period $T = t_4a + t_4b + t_4c$ (Figure 1.19) was set to 50 ms (top) and 27 ms (bottom), resulting in experiments with (top) and without (bottom) suppression of back transfer.

**APSY-NMR with the proteins 434-repressor(1–63) and TM1290.** In the context of this paper, 6D APSY-seq-HNCOCANH experiments (Figure 2) were recorded with the two proteins TM1290 and 434-repressor(1–63). For each protein, 25 projections were measured, using the projection angles given in Table 1.9. 20 projection angles were chosen such that all possible pairs of evolution dimensions are combined in two projections. In addition, 5 “direct” projections along a single evolution dimension were measured. These projections fully exploit the six-dimensionality of the experiment, even though none of the selected individual projection includes chemical shift information of more than 3 nuclei (Table 1.9). Combinations of more than two evolution dimensions could readily be recorded and analyzed with the present setup. The price to pay would be that these projection experiments have reduced sensitivity, since the sensitivity is proportional to $(1/\sqrt{2})^{n-1}$, where $n$ is the number of combined evolution dimensions.

52
The projection spectra were automatically peak picked, and the 2D peak lists were used as input for the GAPRO algorithm in order to compute the 6D peak positions (Hiller et al., 2005). Since peak picking uses the individual projections, the sensitivity exploited in this step is in principle that of the individual 2D projections. However, GAPRO makes further use of the fact that the NMR peak positions in different projections are correlated, so that noise peaks can be efficiently eliminated. The signal-to-noise ratio can therefore be as low as 3–5 without generating artifacts in the final results from the peak picking (S/N ratio = peak maximum divided by the standard deviation of the noise).

For the two protein samples presented here, sensitivity was not a limiting factor. In both applications, a complete 6D APSY peak list was obtained for the residues with observable $^{15}$N-$^1$H signals (see below). Since the 6D APSY-seq-HNCOCANH scheme correlates two sequential amide protons, the N-terminal residue, the Pro residues, and possibly other residues without observable $^1$H$^N$ signal are not contained in this peak list.

For 434-repressor(1–63), the peak list contained 56 out of 57 peaks that would be expected from the chemical structure of the molecule, and there were no artifacts. The connection between residues 2 and 3 was missing, probably because local dynamics lead to a reduction of the signal below the noise level in all projection spectra.

For TM1290, all but three of the peaks expected from the previous, interactively determined assignment (Etezady-Esfarjani et al., 2003) were contained in the final 6D peak list. TM1290 contains dynamically disordered regions, which makes the sequential $^1$H$^N$–$^1$H$^N$ connectivities in the regions 46–52 and 72–73 NMR-unobservable (Etezady-Esfarjani et al., 2003). Three additional peaks were missing in the GAPRO peak list, i.e., the connections 19–20, 20–21 and 71–72, because the signal intensities were below the noise level in the 6D APSY-seq-HNCOCANH data set but not in the conventional triple resonance experiments (Etezady-Esfarjani et al., 2003). The final GAPRO peak list thus contained 98 sequential peaks, and there were no artifacts.

In Figure 1.21, the 6D peak list obtained from 25 2D projections (Table 1.9) of the 6D APSY-seq-HNCOCANH experiment with TM1290 is projected onto two experimental 2D projections, visualizing the excellent agreement between the calculated and measured peak positions. The precision of the chemical shifts can be directly measured in the 6D APSY-seq-HNCOCANH data set, since the resonance of each amide moiety is part of two different 6D peaks. In one of these two peaks, the amide proton chemical shift is measured in the direct dimension ($\omega_6$) during acquisition, and in the second peak the same $^1$H$^N$ proton shift contributes to the indirect dimension ($\omega_1$); the corresponding $^{15}$N chemical shift is also...
measured twice, in $\omega_5$ and $\omega_2$. There are 93 amide moieties that contribute to two peaks in the 6D peak list of TM1290, and from these the chemical shift differences $\omega_1(^{1}H) - \omega_6(^{1}H)$ and $\omega_2(^{15}N) - \omega_5(^{15}N)$ were calculated. The resulting standard deviation for the proton chemical shifts was 0.0014 ppm (0.72 Hz), with a maximal deviation of 0.005 ppm (2.5 Hz), and the standard deviation for the nitrogen chemical shifts was 0.0137 ppm (0.69 Hz), with a maximal deviation of 0.048 ppm (2.4 Hz).

**Figure 1.21.** Visualization of part of the result of the fully automated sequence-specific backbone resonance assignment of the $[^{13}C,^{15}N]$-labeled protein TM1290 (protein concentration 3.0 mM, 20 mM phosphate buffer at pH 6.0, T = 35ºC). (a) and (b) are regions of two orthogonal projection spectra from a 6D APSY-seq-HNCACONH experiment with projection angles ($\alpha, \beta, \gamma, \delta$) = (0º, 0º, 0º, 0º) and (90º, 0º, 0º, 0º), respectively. The 6D APSY peak list is projected onto the spectra by black dots. The sequence-specific resonance assignment, as found by GARANT, is written next to each peak, using the one-letter amino acid code and the sequence position.

**Backbone resonance assignment of 434-repressor(1–63).** The software GARANT had previously been shown to yield reliable automated sequential NMR assignments, provided that an input of precise, artifact-free peak lists was available (Bartels et al., 1997). We therefore chose GARANT to perform sequential assignments based on the presently obtained 6D APSY peak lists.

For the 434-repressor(1–63), 30 GARANT calculations all converged to the correct sequence-specific resonance assignment for all the resonances contained in the 6D peak list. This assignment procedure based on the 6D APSY-seq-HNCACANH experiment yielded three fragments of length 39, 3 and 17 residues, which could unambiguously be fitted to the sequence of 434-repressor(1–63) in the positions 3–41, 43–45, and 47–63, respectively. These
fragments are separated by the two prolines in positions 42 and 46. For the residues 1 and 2, the \(^{15}\text{N}^{1}\text{H}\) NMR signals were not observed. A plot corresponding to Figure 1.22 therefore contains a continuous thick line at the level \(N = 30\) for the polypeptide segments 3–41, 43–45, and 47–63 (not shown).

![Figure 1.22](image)

**Figure 1.22.** Automated sequence-specific resonance assignment for the protein TM1290 obtained with GARANT (Bartels et al., 1997), using as input the peak lists from a 6D APSY-seq-HNCOCANH experiment calculated with the program GAPRO. GARANT was run 30 times with the same input list of 6D peaks but with different, randomly chosen starting conditions. For each residue the most frequent \(^{1}\text{H}\) assignment was counted and plotted as a square along the vertical \(N\)-axis (thick black lines correspond to a sequence of squares). If \(N \geq 15\) (dashed horizontal line), the sequence-specific assignment was accepted. All other \(^{1}\text{H}\) were considered to remain unassigned. The amide protons of residues 1, 6, 20, 46–52 and 72–73, which include the prolines 6 and 51, are not contained in the 6D APSY-seq-HNCOCANH resonances. With the acceptance criterion used here, all these amide protons remained unassigned.

**Backbone resonance assignment for TM1290.** 30 GARANT calculations made with the 6D APSY-peak list of TM1290 did not all converge to the correct solution, but for all the resonances observed in the 6D experiment the correct sequence-specific assignment was found in at least half of the calculations. Erroneous assignments of residues for which no NMR signals had been observed were obtained only in a small number of the calculations (Figure 1.22). This result can be rationalized based on previous work with TM1290, which had shown that the proline residue in positions 6, and other residues with NMR-unobservable amide moieties divide the TM1290 sequence into sequentially connected segments of residues 2–5, 7–45, 53–71 and 74–115 (Etezady-Esfarjani et al., 2003). As mentioned above, three signals expected from the results of this earlier work could not be detected by the 6D APSY experiment, so that overall the 6D correlation NMR signals with the residues 1, 6, 20, 46–52 and 72–73 were absent.
Discussion

The results obtained for the 434-repressor(1–63) and for TM1290 (Figures 1.21 and 1.22) document that 6D APSY-seq-HNCOCANH data sets can provide the information needed for obtaining de novo backbone resonance assignments for small and medium-size proteins. Here we want to further investigate the robustness of the procedure with respect to deterioration of the 6D data sets, and to replacement of 6D APSY-NMR with lower-dimensional APSY-NMR experiments.

Model calculations on the impact of type and quality of APSY-NMR data on the resonance assignments. Peak lists from 6D APSY-seq-HNCOCANH have outstandingly high information content, with six chemical shifts per peak, high accuracy of the chemical shifts, and no or very few artifacts. The influence of variable quality of the input on the automated assignment procedure was evaluated by computationally generating lower-dimensional peak lists, which in their combinations had, in principle, the same information content as the 6D peak list. We thus generated peak lists for combinations of the two 4D spectra HNCOCA and HNCA, and the four 3D spectra HNCA, HN(CO)CA, HN(CA)CO and HNCO from the 6D APSY-NMR peak list of TM1290. These peak lists contain overlapping peaks that could not be resolved in the 3D or 4D spectra, but are resolved in the 6D APSY-HNCOCANH spectrum. The two 4D and the four 3D peak lists, respectively, were used as input for 30 GARANT calculations (Figure 1.23, (c) and (d)). Although these computationally generated peak lists had the same chemical shift precision as the experimental 6D peak list, 4 and 15 of the previously assigned residues, respectively, were not assigned by GARANT when using the 4D and 3D data sets, and the 4D data sets led to two erroneous assignments.

The influence of the accuracy of the chemical shifts, and of the presence of artifacts in the input peak lists on the automated assignment procedure was tested by randomly varying the peak positions in the 6D APSY-NMR peak list of TM1290, and by adding artifacts to it. In one of the “deteriorated” 6D peak lists, the peak positions in the dimensions $\omega_6(\overset{1}{H})$ and $\omega_5(\overset{15}{N})$ were randomly changed within ranges of 0.02 ppm and 0.4 ppm, respectively (Figure 1.23(e)), ensuring that all connectivities could be found within the tolerance range used by GARANT. In another modification of the 6D peak list, 10 artifacts were added at random positions within the experimental spectral ranges (Figure 1.23(f)). Finally, chemical shift variation of individual peaks and the addition of artifacts were combined in a single list (Figure 1.23(g)).
Figure 1.23. Extent of the automated backbone assignment of the protein TM1290 achieved with GARANT when using input peak lists of different quality. The height of the diagonally dashed areas along the vertical axis indicates the number of $^1$H$^N$-$^{15}$N moieties for which both $^1$H$^N$ and $^{15}$N were correctly assigned in the final chemical shift list; gray areas indicate the number of amide group resonances that remained unassigned; vertically dashed areas represent the number of NMR-unobservable amide groups (see text); black areas indicate the number of erroneous assignments, which all occur for residues for which no NMR signals were observed; the white area at the top indicates the residues for which no $^1$H$^N$ signals are expected, i.e., the N-terminal residue and the two prolines. The interpretation of this data can be based on the following properties of TM1290, which are visualized in (a): the protein contains 115 residues with 112 backbone amide groups that should normally be observable by NMR. Of these, 8 were not detected in conventional 2D and 3D NMR experiments, so that sequence-specific assignments were obtained for 104 amide groups (Etezady-Esfarjani et al., 2003). (b) to (g) represent the extent of assignments obtained with GARANT when using the following six different input data sets: (b): final 6D peak list from the 6D APSY-seq-HNCOCANH measurement (same as Figure 1.22); (c) Two 4D peak lists computed from the original 6D peak list for the experiments 4D APSY-HNCOCA and 4D APSY-HNCACO; (d) Four 3D peak lists computed from the original 6D peak list for the experiments 3D APSY-HNCA, 3D APSY-HN(CO)CA, 3D APSY-HNCO and 3D APSY-HN(CA)CO; (e): Ten 6D peak lists generated by randomly varying the peak positions in the experimental 6D peak list (b) along $\omega_6$($^1$H) and $\omega_5$($^{15}$N) within 0.02 and 0.4 ppm, respectively; (f) Ten 6D peak lists generated by adding 10 artifacts at random positions to the experimental 6D peak list (b); (g) Combination of (e) and (f). For each data set, the GARANT analysis was repeated 30 times with different, randomly chosen starting conditions (Bartels et al., 1997), and the results were analyzed as described in Material and methods. For each of the columns (e) to (g), the average of the results obtained with the 10 deteriorated peak lists is plotted in the figure (see text).

Each type of deteriorated list was generated 10 times, with different, randomly chosen chemical shift changes and/or additions of artifacts at different random positions. For each of the resulting 30 peak lists, 30 GARANT calculations were performed and analyzed as described above for the experimentally obtained 6D APSY-seq-HNCOCANH peak list. The results for the 10 lists with random chemical shift changes, the 10 lists with artifact peaks, and the 10 lists containing both manipulations, respectively, were then averaged. Figure 1.23, e–g, shows the average number of correctly assigned residues (diagonally dashed areas), the average number of unassigned residues for which an assignment is expected (gray areas), the
average number of “correctly unassigned residues”, i.e., residues for which the $^{15}$N–$^1$H NMR signals could not be detected (vertically dashed areas), and the average number of erroneously assigned residues (black areas). Overall, these model calculations revealed an encouraging robustness of the APSY-NMR-based assignment, since throughout at least 85% of the NMR-observable residues were assigned, with a correctness rate above 97% (Figure 1.23). Nonetheless, deterioration of the high-quality peak list obtained from the 6D APSY-NMR experiment did lead to fewer residues being assigned, and to introducing some wrong assignments, which however always occur for residues for which no NMR signals were observed. Similar effects resulted from the replacement of 6D APSY-seq-HNCOCANH with, in principle, equivalent combinations of lower-dimensional APSY-NMR experiments, where the combination of four 3D data sets resulted in the lowest yield of correct assignments. The combination of two 4D data sets, in contrast, gave nearly the same extent of correct assignments as the 6D data set, albeit with introduction of erroneous assignments for two NMR-unobservable residues, which we presently attribute to an intrinsic weakness in the GARANT assignment protocol (see below).

**Outlook.** The use of a single 6D APSY experiment to perform the backbone resonance assignment avoids calibration problems between different spectra, the high dimensionality eliminates resonance overlaps, and the high information content of a 6D peak reduces the number of possible assignment combinations and thus makes the procedure of sequential backbone assignments efficient and robust. Based on the large dispersion of the $^{15}$N chemical shifts in polypeptide chains (Braun et al., 1994; Wüthrich, 1986), the 6D-APSY-seq-HNCOCANH experiment should in principle be widely applicable.

For practical applications, it will be of interest to further optimize the sensitivity of APSY-NMR experiments. Obviously, minor but not negligible improvements may result from adding water flip-back pulses (Grzesiek and Bax, 1993) and sensitivity enhancement schemes (Palmer et al., 1991; Kay et al., 1992). Furthermore, part of the projections could be measured with individually optimized, shortened pulse sequences, which would improve the sensitivity for these projections (Kupče and Freeman, 2003). For example, direct projection of the $\omega_5^{(15)}$N-dimension could be replaced by a standard $[^{15}\text{N},^1\text{H}]-\text{HSQC}$ experiment. In this context, the outcome of the model studies with 3D and 4D APSY-NMR data sets (Figure 1.23, c and d) is also highly encouraging. For these intrinsically more sensitive APSY experiments, which are based on combining HNCACO and HNCOCA elements, there is a clear promise for use with larger proteins. Current work in our laboratory is focused on
generating combinations of lower-dimensional APSY-NMR experiments suitable for use with larger proteins.

The software GARANT used here to derive automated backbone resonance assignments had been developed to work with peak lists obtained from “conventional” heteronuclear correlation and triple resonance NMR spectra. It allows multiple alternative assignments in situations of peak overlap, and it has been laid out to deal with the presence of artifacts. High-dimensional APSY-NMR data sets typically contain no or at most few peak overlaps and artifacts. Substituting GARANT with a new assignment algorithm that is tailored specifically for the needs to analyze APSY-NMR data for automated resonance assignment, promises to further enhance the potentialities of APSY-NMR-based automated resonance assignment for proteins.

In conclusion, APSY-NMR combined with a suitable assignment algorithm promises to enable fully automated sequence-specific backbone resonance assignments for small and medium-sized proteins. Eliminating the extensive human interactions typically needed so far for obtaining NMR assignments will further add to making high-resolution NMR with proteins an attractive tool in structural biology and structural genomics.

Automated backbone assignment of a folded protein using strategy IIb–B

In the previous chapter, it was demonstrated that strategy I–A, the single 6D APSY-seq-HNCOCANH experiment in combination with the algorithm GARANT, can yield fully automated backbone assignment of small proteins. The method showed to be reliable even in a case where larger stretches of a protein were not observable. The present chapter describes the application of strategy IIb–B for the automated sequence-specific backbone resonance assignment of 434-repressor(1–63) with the set of the three experiments 4D APSY-HACANH, 5D APSY-HACACONH and the 5D APSY-CBCACONH. The establishment of a type II strategy is a step towards the extension of the size limit of automated backbone assignment with APSY, as the size limit of strategy IIb–B is higher than the size limit of strategy I–A (Table 1.8).

With the present work, we introduce a new peak picking routine for 2D spectra that replaces the ATNOS peak picking routine used so far (Herrmann et al., 2002). The new peak picking algorithm consists of three steps: (i) The determination of spectral noise, (ii) the identification of all local maxima with a signal-to-noise ratio above the user-defined threshold $R_{\text{min}}$, (iii) the interpolation of peak coordinates (Fig. 1.24).

![Figure 1.24](figure.png)  
Figure 1.24. The GAPRO peak picker. (a) For the determination of the thermal noise RMSD, the spectrum is divided into 64 submatrices. (b) Scheme for the interpolation of the position of the peak maximum. Three data points are indicated as open rectangles, with the height denoting the intensity and the width the digital resolution of the spectrum, $\Delta$. Starting from the central data point as a local maximum, the peak position (indicated by the arrow) is interpolated such that the light and the dark grey area have the same size. (c) Result of GAPRO peak picking on two representative backbone amide peaks from the $(0^\circ,0^\circ,0^\circ)$ projection of the 5D APSY-HACACONH experiment with 434-repressor(1–63). The peak positions before application of the interpolation scheme (left) and after the interpolation (right) are indicated by black crosses. The size of a digital point is indicated by the grey square in the lower left corner.

In the first step, the determination of spectral noise, the data matrix representing the 2D NMR spectrum is divided into $8 \times 8 = 64$ submatrices of equal size (Fig. 1.24(a)). For each of these
64 submatrices, the root mean square deviation of the spectral intensity is calculated. Using the assumption that at least one of the 64 submatrices is devoid of peaks and $t_1$-noise, the smallest of the 64 intensity RMSD values is a measure for the spectral noise. This assumption was found to be fulfilled in several hundreds of different projections of all kinds of APSY experiments with amide proton detection. In the second step, the identification of local maxima, all data points in the intensity matrix with intensities larger than the eight neighboring data points are identified. For pairs of points with equal intensity, only one of the two points is taken. A point is kept, if the signal intensity at this point divided by the spectral noise is larger or equal than the user defined threshold, $R_{\text{min}}$. In the third step, coordinates for the peak maximum are calculated using the interpolation scheme shown in Figure 1.24b in each of the two dimensions. The only line shape model implied is that the resonance line is symmetric, which usually applies for all correctly phased NMR peaks. The effect of the interpolation scheme on the peak positions is shown in Figure 1.24(c).

Pulse sequences for the three experiments necessary for strategy IIb–B were created using the transfer elements in the library according to Table 1.3. Exemplary for the three pulse sequences, the 5D APSY-HACACONH pulse sequence is shown here with experimental details (Fig. 1.25). The 5D ASPY-HACACONH uses the transfer element 23 for the conversion of longitudinal two-spin order $HA_zCA_z$ to $CA_zCO_z$. The transfer function for this element depends on the spin system topology and is given by $\sin(\pi \cdot J_{HA,CA} \cdot \delta')$ and $2 \cdot \sin(2 \cdot \pi \cdot J_{HA,CA} \cdot \delta')$ for CH groups (non-glycine residues) and CH$_2$ groups (glycine residues), respectively. In many conventional applications, a delay time $\delta'$ of 2.3 ms is used (Burum and Ernst, 1980; Kay et al., 1993; Matsuo et al., 1996b), resulting in similar values for the two transfer functions. However, the peaks for glycine and non-glycine-residues are of opposite sign because of the $^1$H 180° pulse in the first INEPT step of the pulse sequence, element 7. The simultaneous occurrence of positive and negative peaks can be unfavorable in projection spectra because of signal cancellations depending on the projection angles. This situation can be avoided by using a delay time of $\delta' = 4.75$ ms, leading to similar signal intensity and to peaks of equal sign for CH and CH$_2$ groups. The use of this longer transfer time does not prolong the minimal length of the experiment. The same situation occurs for the 4D APSY-HACANH experiment and transfer element 11.

The three APSY experiments for strategy IIb–B were recorded with a 0.9 mM sample of $^{13}$C, $^{15}$N-labeled 434-repressor(1–63) at 15°C. Important experimental and analysis parameters and features of the resulting peak lists are given in Table 1.10.
Figure 1.25. Pulse sequence used for the 5D APSY-HACACONH experiment consisting of elements 7, 23, 10, 3 and 4 from the Appendix. Radio-frequency pulses were applied at 4.7 ppm for $^1$H, at 118.0 ppm for $^{15}$N, at 173.0 ppm for $^{13}$C and at 54 ppm for $^{13}$C'. Narrow and wide bars represent 90° and 180° pulses, respectively. Pulses marked “A” were applied as Gaussian shapes, with 120 µs duration on a spectrometer with a $^1$H frequency of 500 MHz. All other pulses on $^{13}$C' and $^{13}$C had rectangular shape, with a duration of $\sqrt{3/(\Delta \omega (C',C)\cdot 2)}$ and $\sqrt{15/(\Delta \omega (C',C)\cdot 4)}$ for 90° and 180° pulses, respectively. Pulses on $^1$H and $^{15}$N were applied with rectangular shape and high power. The last six $^1$H pulses represent a 3-9-19 WATERGATE element (Sklenar et al., 1993). Grey pulses on $^{13}$C' and $^{13}$C were applied to compensate for off-resonance effects of selective pulses (McCoy and Mueller, 1992). Decoupling using DIPSI-2 (Shaka et al., 1988) on $^1$H and WALTZ-16 (Shaka et al., 1983) on $^{15}$N is indicated by rectangles. $t_5$ represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped pulsed magnetic field gradients applied along the z-axis with the following durations and strengths: $G_1$, 800 µs, 18 G/cm; $G_2$, 800 µs, 26 G/cm; $G_3$, 800 µs, 13 G/cm; $G_4$, 800 µs, 23 G/cm; $G_5$, 800 µs, 26 G/cm; $G_6$, 800 µs, 23 G/cm. Phase cycling: $\phi_1 = \{x, x, -x, -x\}$ $\phi_2 = \{x, -x\}$, $\phi_3 = \{x, -x, -x, x\}$, $\psi_1 = y$, all other pulses = x. The delays were used as given in the Appendix and the evolution elements were used according to Figure 1.13. Quadrature detection for the indirect dimension was achieved using the hypercomplex Fourier transformation method for projections (Kupce and Freeman, 2004b) with the angles $\psi_1$, $\psi_2$, $\psi_3$ and $\psi_4$ ($\psi_1$, $\psi_2$ and $\psi_3$ were incremented and $\psi_4$ was decremented in 90° steps).

For a proper evaluation of the resulting peak lists, it is necessary to consider the particular features of glycine residues in the three experiments. For all non-glycine side chains, the 5D APSY-CBCACONH experiment correlates the CB and CA chemical shifts in the first and the second of the five dimensions, respectively. Despite the lack of a CB atom, glycine residues give rise to a 5D peak in this experiment, which contains the CA chemical shift in the first as well as in the second dimension. In the 4D ASPY-HACANH and the 5D APSY-HACACONH experiments, each glycine residue with non-degenerate HA chemical shifts gives rise to a pair of peaks, which are degenerate in all except the HA dimension. The analysis of the projection spectra by GAPRO with the present parameters can usually detect only one of these two peaks. The reason for this is that all projections with projection angles that do not resolve the HA dimension can contribute at most one subspace for the two peaks. During the identification of peak subgroups with GAPRO (Figure 1.4), the peak subgroup of the glycine peak with the higher support $S$ is identified first and all supporting subspaces are
removed from the analysis. This step likely decreases the support of the second glycine peak below the minimal support, $S_{\text{min}}$, and this peak will thus not appear in the final peak list. If desired, this problem can be encountered by recording sufficient projections that resolve the HA dimension. Alternatively, an improved version of GAPRO could be imagined, which is tailored to recognize such a situation and resolve both peaks in these experiments. However, as the absence of one of the glycine peaks does not pose a major hindrance for automated backbone resonance assignment, no further efforts to improve the situation were made here.

A further complicating situation occurs in the 4D APSY-HACANH experiment, where each amide moiety is included in a pair consisting of sequential and an intraresidual peak (Fig. 1.17), with degenerate frequencies in 1 of the 3 indirect dimensions. This pair of peaks is resolved in all projections except the one corresponding to the 2D $[^{15}\text{N},^{1}\text{H}]$-HSQC. As the $^{1}J_{\text{N,CA}}$-coupling is in most cases larger than the $^{2}J_{\text{N,CA}}$-coupling, the intraresidual peak is usually stronger than the sequential peak. The $^{2}J_{\text{N,CA}}$-coupling can be as small as 4 Hz (Table 1.4) and thus in many cases the intensity of the sequential peak is below the noise level in the projection spectra. In the present experiment, 60 of the expected 60 intraresidual peaks were found and 48 of the expected 60 sequential peaks. As the sequential peaks are obtained from the 5D APSY-HACACONH experiment, this result is more than sufficient for the backbone resonance assignment.

### Table 1.10. Summary of important experimental and analysis parameters and the resulting peak list for the three APSY experiments with a 0.9 mM sample of 434-repressor(1–63) at 15°C. The important peaks for the assignment strategy are indicated in bold

<table>
<thead>
<tr>
<th>Experiment</th>
<th>4D APSY-HACANH</th>
<th>5D APSY-HACACONH</th>
<th>5D APSY-CBCACONH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projections recorded</td>
<td>23</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>Time per projection</td>
<td>40 min</td>
<td>20 min</td>
<td>10 min</td>
</tr>
<tr>
<td>Total time</td>
<td>15 h 20 min</td>
<td>4 h</td>
<td>4 h 40 min</td>
</tr>
<tr>
<td>Spectrometer $^{1}\text{H}$ frequency</td>
<td>750 MHz</td>
<td>750 MHz</td>
<td>500 MHz</td>
</tr>
<tr>
<td>$S_{\text{min}}$</td>
<td>14</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Seq. peaks expected $^{a}$</td>
<td>60 / 4</td>
<td>60 / 4</td>
<td>60</td>
</tr>
<tr>
<td>Seq. peaks found $^{a}$</td>
<td>48 / 0</td>
<td>59 / 0</td>
<td>59</td>
</tr>
<tr>
<td>Intrares. peaks expected $^{a}$</td>
<td>60 / 4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Intrares. peaks found $^{a}$</td>
<td>60 / 0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Artifacts</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$^{a}$ In the two experiments involving HA atoms, for all residues, one peak is counted before the dash. For Glycine residues with degenerate HA chemical shift, a second peak is counted behind the dash.
With these considerations, 60 relevant peaks are expected in each of the three experiments. All these 60 peaks were contained in the 4D APSY-HACANH peak list and from the other two experiments, the 5D APSY-HACACONH and the 5D APSY-CBCACONH experiment, 59 peaks were obtained (Table 1.10, Fig. 1.26). The missing peak in the 5D APSY-CBCACONH peak list correlates CB, CA and CO of Arg 10 with the amide moiety of Ile 11. An inspection of the projection spectra showed that this resonance did not appear in the projection spectra and was thus not contained in the final result. The missing peak in the 5D APSY-HACACONH experiment was the peak correlating HA, CA and CO of Arg 43 with the amide moiety of Phe 44. However, the absence of this peak should not cause difficulties for the assignment procedure, as the sequential connectivity information between residues 43 and 44 is contained in the CBCACONH peak and the building of fragments with the corresponding HACANH peak is possible based on the CA chemical shift.

Figure 1.26. Result of the experiments 4D APSY-HACANH, 5D APSY-HACACONH and 5D APSY-CBCACONH for 434-repressor(1–63). Each peak is evaluated below the residue, whose amide moiety it contains. Open circles denote peaks contained in the final peak lists. Filled circles denote pairs of Glycine peaks, of which one peak was found (see text). Crosses denote peaks which were not contained in the final peak list due to intensities below the noise level in the projections. A star denotes Gly 53, which is the only Glycine in this protein with degenerate HA chemical shifts. In the HACANH experiment, 80% of the sequential peaks were contained in addition to the intraresidual peaks. No other peaks were contained in the peak lists.

Besides the peaks mentioned so far, the 5D APSY-CBCACONH peak list contained 7 peaks from glutamine side chains. Based on the chemical shifts, these peaks are unambiguously distinguishable from backbone peaks and they were removed by a filter routine in GAPRO. For the seven side chain peaks the chemical shift values were in each dimension within 1 ppm from the BMRB values for the glutamine sidechains: 29.2 ppm for CB, 33.7 ppm for CG, 179.7 ppm for CD, 111.89 ppm for NE and 7.03 / 7.21 ppm for HE1/2 atoms (Seavey et al., 1991). No other peaks were contained in the three peak lists. Regarding the completeness and the quality of the three peak lists, with the absence of any artifacts, the sequence-specific resonance assignment of the protein backbone should be straightforward. The three peak lists
and the amino acid sequence were given as input to the program GARANT (Bartels et al., 1997). 10 independent assignment runs from random starting conditions were performed using the same standard protocol as in the previous chapter. Also the acceptance criterion of 50% for an assignment was chosen as in the previous chapter. With this procedure, all atoms of the protein backbone except the last 2 residues and except three Glycine HA atoms were correctly assigned (Figures 1.27 and 1.28).

![Figure 1.27](image)

**Figure 1.27.** Result of the automated sequence-specific backbone resonance assignment with GARANT using the three peak lists from Figure 1.26 as input. The heights of the black bars show the number of times \( n \), in which the correct assignment for the amide proton was found in 10 independent runs. The white dashed line denotes the cutoff \( n = 5 \) for the acceptance of an assignment.

In the final result, two out of the five Glycine residues, Gly 37 and Gly 53, were correctly assigned and the other three, Gly 14, 25 and 62 were incorrectly assigned at the HA positions. For Gly 14, only one HA of the two HA chemical shifts was contained in the input peak lists from the 4D APSY-HACANH and the 5D APSY-HACACONH experiment and thus only one HA atom was correctly assigned by GARANT. For Gly 25, the peaks from the 5D APSY-HACACONH and the 4D APSY-HACANH peak list had two different HA shifts, so that correct assignment of the HA shifts in this residue would have been in principle possible. GARANT however, assigned an averaged chemical shift to both HA atoms. Further, for the assignment of the C-terminal residues GARANT used some of the peaks twice, resulting in identical chemical shift assignments for HN and N atoms in Gly 62 and Thr 63 for CA and HA in Asn 61 and Gly 62. This is a particularly unrealistic solution, as the Gly CA atom was assigned with the value 50.6 ppm, which is not in the expected range from the BMRB data.
base. Thr 63 was not assigned by GARANT despite the correct peaks were in the input data set.

Figure 1.28. Result of the automated sequence-specific resonance assignment of the backbone atoms HN, N, CO, CA, HA and CB with GARANT using the three peak lists from Figure 1.26 as input. The result of the assignment is indicated by the color. Green color denotes correct assignment, pink color denotes wrong assignment for glycine HA and red color denotes wrong assignments (see also text). Grey color denotes atoms, where no chemical shift information can be obtained. The column HN, corresponds to Figure 1.27, where all residues except Thr 63 were correctly assigned.

Summarized, the findings in this chapter show that automated sequence-specific backbone assignment of proteins with strategy IIb–B is possible. Despite good input data from the APSY experiments, the assignment algorithm GARANT did not assign all atoms of the protein backbone correctly. The algorithm GARANT was developed to deal with peak lists from conventional spectra, which can contain substantial amounts of noise artifacts, strong signal overlap and inaccurate chemical shifts (Bartels et al., 1997). The quality of APSY peak lists is different in all these three points and it seems therefore likely that a different assignment algorithm, perhaps one that is tailored for the requirements of APSY, could perform better than GARANT. Still, in the present approach 59 of the 60 amide moieties, corresponding to 98%, were correctly assigned. The results obtained are very encouraging and it seems likely that the chosen strategy can be extended to larger proteins.
Automated backbone assignment of a denatured protein

Sequence-specific resonance assignment of denatured proteins using conventional 3D NMR techniques is a cumbersome and time consuming process, as the small chemical shift dispersion encountered in such systems requires the recording and analysis of a large number of spectra (Wüthrich, 1994). The application of APSY for this task seems very promising, as APSY benefits from the long relaxation times, which enable 4D–6D or even higher-dimensional experiments. The high dimensionality, accuracy and quality of the resulting APSY peak lists may further present a basis for fully automated sequence-specific backbone resonance assignments of such systems.

It is the aim of the present chapter to explore the potential of APSY NMR for the fully automated sequence-specific backbone resonance assignment of a denatured protein. For this purpose, the 148-residue outer membrane protein X (OmpX) from *E. coli* has been chosen as a model system. The urea-denatured state of this protein is well studied and reference assignments are known (Tafer et al., 2004). It is a further aim to establish a standard parameter set for the repetition of the experiments on proteins of similar size. For this purpose, the influence of the GAPRO parameters on the result has been analyzed.

As a result of the small chemical shift dispersion of denatured proteins, the frequency spaces of multidimensional NMR experiments are densely packed with resonance peaks and only high-dimensional experiments are devoid of peak overlap (Fig. 1.29). In the 5D HNOCACAB frequency space of denatured OmpX at 750 MHz, two pairs of peaks are less than 20 Hz separated. Spaces of lower dimensionality will be even more crowded and not all peaks can usually be expected to be resolved. In the frequency space of the 6D seq-HNOCANH, however, the closest pair of peaks is more than 60 Hz apart. The main reason for these larger peak separations in the 6D experiment compared to the 5D experiment is the inclusion of a second nitrogen atom, which has a large dispersion. The 7D seq-HNCO(CA)CBCANH does not substantially spread out the closest peaks compared to the 6D experiment, because the CB chemical shifts are often identical for all amino acids of the same kind. The closest pairs of peaks in the 6D experiment consist of residues with equal side chains and these pairs of peaks are not further separated by inclusion of the CB chemical shift in an additional dimension. Similar plots can be expected for other denatured proteins of the same size, because the chemical shift values in denatured proteins depend only on the local amino acid sequence and the composition of these plots is therefore statistical.
These considerations show that the use of high-dimensional spectra is desirable for denatured proteins and we thus chose the backbone assignment strategies I–A and I–B for the assignment. The 6D APSY-seq-HNOCANH was recorded with a 3 mM sample of \([U^{-13}\text{C},^{15}\text{N}]-\text{labeled OmpX at 15°C on a Bruker DRX 750 MHz spectrometer equipped with a regular probe. Using a phase cycle of 4 scans per increment, a single projection of the 6D experiment with 192 complex points in the indirect dimension was recorded in 33 minutes.}

In the 6D APSY-seq-HNOCANH experiment, each 6D peak correlates two sequential amide moieties with the CA and the CO between them (Fig. 1.18). The intraresidual transfer is suppressed in this experiment, such that one peak per sequential pair of amide moieties can be expected. A 148-residue protein would thus lead to 147 peaks and under regard of the 4 proline residues and the N-terminus, 138 peaks are expected from OmpX. From the expression of OmpX in inclusion bodies usually two protein species with different N-terminus are obtained (Tafer et al., 2004). Both species have identical chemical shifts for the major part of the polypeptide chain, but different chemical shifts for nuclei in residues 1–5, resulting in 142 expected peaks for OmpX. Further, the 6D peaks connecting residues 99/100 and 72/73 were found to be of reduced intensity compared to the other peaks. The reasons for the weak signal intensity of the peak 99/100 is linebroadening of the nuclei of His 100, probably due to chemical exchange of the His side chain, which has a pKₐ value close to the pH value of 6.5. The weak intensity of the peak 72/73 is likely due to the hydrophobic cluster of residues 73–
82. We classify the 140 remaining peaks as “expected peaks” in the further analyses of the 6D APSY-seq-HNCOCANH experiment. Besides these 140 peaks, 14 further peaks of smaller intensity were observed in the spectra, including the two peaks of residues 99/100 and 72/73 (Fig. 1.30). The origin of these further peaks is unclear. The sample was purified by ion exchange chromatography and as judged on a polyacrylamide gel, the purity was > 95%. As these peaks are regular 6D peaks, with chemical shift values in all six indirect dimensions, they are not artifacts but additional correct peaks. All other peaks that appear in APSY peak lists are counted as artifacts.

![Figure 1.30](image)

**Figure 1.30.** Cross section through the (0°, 0°, 0°, 0°)-projection of the 6D APSY-seq-HNCOCANH experiment of denatured OmpX at a 1^5^N frequency of 115.7 ppm. The large peak at a proton frequency of 2720 Hz is T92, the small peak at 2650 Hz is of unknown origin (see text). The intensity ratio of the two peaks is 6.4.

105 2D projection spectra were acquired of the 6D APSY-seq-HNCOCANH experiment. The projections were peak-picked with the GAPRO peak picker and a signal-to-noise ratio of $R_{\text{min}} = 4.0$. An analysis of these 105 projections with GAPRO and the parameters $\Delta\nu_{\text{min}} = 3.0$, $r_{\text{min}} = 8.0$ and $S_{\text{min},1/2} = 25$ resulted in a peak list that contained the 140 expected, 8 additional peaks, but no artifacts. The large number of $j = 105$ projections was chosen to provide a large data set for the further analyses and it is not the minimal number of projections that is needed to obtain this result. The dependence of the result on $j$, the number of input spectra, was determined by a reduction of this data set with simultaneous adaptation of $S_{\text{min}}$, and with all other parameters left constant (Fig. 1.31). In this experiment, the minimal number of projections yielding a complete and artifact-free peak list was found to be 24, with a simultaneous minimal support $S_{\text{min}} = 12$. For a repetition of the experiment on another denatured protein of similar size, this number of input projections can however not be recommended, as the set of 24 projections was created *a posteriori* from a reduction of the larger set of 105 projections. We recommend using a set of 40 input projections with the projection angles given in Table
1.11 for a repetition of the experiment. The spectrometer time required for this set is 20 h with the present acquisition parameters.

![Figure 1.31](image)

**Figure 1.31.** Dependence of the number of peaks on the number of projection spectra, \( j \), and the minimal support, \( S_{\text{min}} \), in the 6D APSY-seq-HNCOCANH experiment with denatured OmpX. Blue are the expected OmpX backbone peaks (see text).

| \( j \) | \( S_{\text{min}/2} \) |
|-------|-----------------
| 68    | 18              |
| 64    | 16              |
| 60    | 16              |
| 56    | 16              |
| 52    | 16              |
| 48    | 16              |
| 44    | 16              |
| 40    | 16              |
| 36    | 16              |
| 32    | 16              |
| 28    | 16              |
| 24    | 16              |
| 20    | 16              |
| 16    | 16              |
| 12    | 16              |
| 18    | 12              |
| 10    | 12              |
| 8     | 12              |
| 6     | 12              |

**Table 1.11.** Recommended projection angles for a set of \( j = 40 \) projections for the 6D APSY-seq-HNCOCANH experiment. The projections are sorted for increasing importance, with the last 24 projections after the bold line being the minimal set, with which the optimal result was still obtained.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \delta )</th>
<th>number</th>
</tr>
</thead>
<tbody>
<tr>
<td>90°</td>
<td>± 45°</td>
<td>0°</td>
<td>0°</td>
<td>2</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>90°</td>
<td>± 45°</td>
<td>2</td>
</tr>
<tr>
<td>± 45°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>2</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>90°</td>
<td>1</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>± 45°</td>
<td>0°</td>
<td>2</td>
</tr>
<tr>
<td>90°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>1</td>
</tr>
<tr>
<td>0°</td>
<td>90°</td>
<td>0°</td>
<td>0°</td>
<td>1</td>
</tr>
<tr>
<td>0°</td>
<td>± 45°</td>
<td>0°</td>
<td>± 45°</td>
<td>4</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>0°</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0°</td>
<td>± 45°</td>
<td>± 45°</td>
<td>4</td>
</tr>
<tr>
<td>90°</td>
<td>± 45°</td>
<td>± 45°</td>
<td>0°</td>
<td>4</td>
</tr>
<tr>
<td>± 50°</td>
<td>0°</td>
<td>± 40°</td>
<td>± 25°</td>
<td>8</td>
</tr>
<tr>
<td>90°</td>
<td>± 50°</td>
<td>± 40°</td>
<td>± 25°</td>
<td>8</td>
</tr>
</tbody>
</table>

In the following, we want to identify a robust parameter set and thus investigate the influence of the GAPRO analysis parameters on the resulting peak list. As a starting point, we refer to the set of 40 projections with the projection angles of Table 1.11, which was analyzed with the parameters \( \Delta \nu_{\text{min}} = 3.0, r_{\text{min}} = 8.0 \) and \( S_{\text{min}} = 15 \).
Starting from this basic parameter set, the intersection tolerances in the direct dimension, $\Delta \nu_{\text{min}}$, was varied between 0 and 7 Hz and the intersection tolerance in the indirect dimension, $r_{\text{min}}$, between 0 and 10 Hz (Fig. 1.32 (a) and (b)). The dependence of the result on the two parameters is similar. For all values larger than the “critical” values of $\Delta \nu_{\text{min}} \approx 2.0$ Hz and $r_{\text{min}} \approx 3.0$ Hz the resulting peak list contains all expected peaks. These critical values relate to the spectral resolution of the experiments in the direct and indirect dimension, respectively. Below these critical values, subspaces from corresponding peaks in the different projection do no longer intersect within the tolerances, the minimal support is not reached for some peaks and the number of resulting peaks is below the expected number. The chosen values of $\Delta \nu_{\text{min}} = 3.0$ and $r_{\text{min}} = 8.0$ in the standard set are above the critical values and thus represent stable parameters.

![Figure 1.32. Dependence of the number of 6D APSY-seq-HNCOCANH peaks of denatured OmpX on (a) the intersection tolerance of subspaces in the direct dimension, $\Delta \nu_{\text{min}}$ (b) the intersection tolerance of subspaces in the indirect dimension, $r_{\text{min}}$ (c) the minimal support $S_{\text{min,1/2}}$. Blue are the expected OmpX backbone peaks and red are artifacts (see text). In (c), the data set of $j = 40$ projections is denoted with lighter colors and the data set with $j = 105$ projections with darker colors.

During the analysis with GAPRO, the subspaces originating from each peak are intersected and only peaks that are supported by a minimal support $S_{\text{min,1/2}}$ are contained in the result. The dependence of the resulting peaks in the result on $S_{\text{min,1/2}}$ is shown in Figure 1.32(c) for the set of 105 and the set of 40 input spectra. Artifacts, i.e. the generation of peaks from random noise occurs only with $S_{\text{min,1/2}}$ values below 11 and 9, respectively. As the support is calculated using all input spectra, this value is larger in the case of 105 input spectra than for
40 input spectra. Peak list results with all expected peaks were obtained for $S_{\text{min},1/2}$ values in the range of 9–21 for 40 spectra and in the range of 11–55 for 105 spectra. Thus, for the analysis of the 40 input spectra, a standard value of $S_{\text{min},1/2} = 15$ is recommended. The fact that wide ranges of $S_{\text{min},1/2}$ values exist, in which the expected result is continuously obtained demonstrates the robustness of APSY. It is likely that with these parameters robust results can be expected from a repetition of the experiment with other proteins. For all three parameters, the result of APSY is stable with regard to variations. We have thus identified a parameter set that is likely to yield results with similar quality for other denatured proteins of the same or smaller size.

In the 6D APSY-seq-HNCOCANH experiment, all chemical shifts of amide moieties that are not at the end of a continuous fragment, appear in two different peaks and can thus be used to determine the precision of chemical shift measurement (Fig. 1.33). In the present experiment, the differences $\omega_1-\omega_6$ and $\omega_2-\omega_5$ with 40 input spectra was $0.0 \pm 0.65$ Hz for $^1\text{H}$ and $-0.22 \pm 0.63$ Hz for $^{15}\text{N}$. With 105 input spectra, the precision is increased due to the statistical averaging, as each peak position is on average calculated from more input data. The resulting values are $0.0 \pm 0.30$ Hz for $^1\text{H}$ and for $-0.19 \pm 0.33$ Hz for $^{15}\text{N}$.

![Figure 1.33. Precision of chemical shift measurements in the 6D APSY-seq-HNCOCANH experiment. The histograms show the abundance $n$ of chemical shift differences between the two independent measurements of amide proton and amide nitrogen chemical shifts in two sequential peaks. Blue color corresponds to the data set with 40 input spectra and red color to the set with 105 spectra. In each plot, two points of the blue data set are outside the region shown: For $\omega_1-\omega_6(\text{H})$, these are -2.4 Hz and 3.0 Hz. For $\omega_2-\omega_5(\text{N})$, 2.3 Hz and -2.7 Hz.](image)

Considering the high precision, completeness and artifact-freeness of the 6D APSY-seq-HNCOCANH peak list, the sequence-specific backbone resonance assignment of denatured OmpX using this peak list as the only input can be expected to be a straightforward task for an assignment algorithm. We used the available algorithm GARANT with a standard assignment protocol for this purpose (Bartels et al., 1997). The only input besides the protein sequence...
was the 6D APSY-seq-HNCOCANH peak list, which was used directly from the GAPRO output. The resulting sequence-specific backbone resonance assignment was complete and correct for the whole protein (Fig. 1.34). As mentioned above, the input peak list contains two consistent fragment sets of peaks for the N-terminal 5 residues, branching at residue 6. Out of the two possibilities, GARANT used the set with better matching at the amide group of residue 6 for the assignment and the other set of peaks was not regarded. The presence of 8 additional correct peaks (Fig. 1.30) did not have an impact on the assignment result, as these peaks remained unassigned in the GARANT output.

The assignment strategy using the 6D APSY-seq-HNCOCANH as the only experiment corresponds to strategy I–A (Table 1.16). In a total experiment time of 20 h and in a calculation time of about 10 minutes for GAPRO and 20 minutes for each run of GARANT, all detectable backbone atoms HN, N, CA, CO were correctly assigned for the protein.

OmpX provides a rather optimal case for assignment strategy I–A. The four prolines are evenly distributed on the sequence, resulting on five long sequential fragments of 33, 31, 25, 19 and 30 peaks, respectively. Due to their different lengths and the occurrence of glycine residues at distinct positions in these five fragments, their position on the sequence can unambiguously be determined based on the HN, N, CA and CO chemical shifts. This may however not be possible for all denatured proteins and in such cases, the knowledge of CB chemical shifts would improve the robustness of the method. One possibility to record the resulting strategy I–B (Table 1.16) is the use of the 5D APSY-HNCOCACB experiment (Fig. 1.35) in addition to the 6D APSY-seq-HNCOCANH experiment. 60 2D projections of this experiment were recorded on a 750 MHz spectrometer with 16 minutes per projection, resulting in a total experiment time of 15 hours.
Radio-frequency pulses were applied at 118.0 ppm for $^{15}\text{N}$, at 173.0 ppm for $^{13}\text{C}'$ and, unless otherwise stated, at 42.0 ppm for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$. The $^1\text{H}$ carrier frequency was set at 4.7 ppm. Black and white symbols represent 90°- and 180°-pulses, respectively. Unlabeled rectangular bars stand for rectangular pulses applied at maximum power. Pulses marked with capital letters have special length and shapes, depending on their purpose. Here, all pulse lengths are given for a $^1\text{H}$ frequency of 750 MHz: $^{13}\text{C}'$-pulses: A, 180°, rectangular shape, 38.26 µs; B, 90°, rectangular shape, 42.77 µs; C, 180°, I-Burp (Geen and Freeman, 1991), 220 µs.

$^{13}\text{C}^\alpha$-pulses: D, 180°, I-Burp (applied at 51.0 ppm), 220 µs; E, 90°, Gaussian Cascade Q5 (Emsley and Bodenhausen, 1990), 280 µs; F, 180°, Gaussian Cascade Q3 (Emsley and Bodenhausen, 1990), 185 µs; G, 90°, Gaussian Cascade Q5 time reversed, 280 µs. The six last pulses on the $^1\text{H}$ line represent a 3-9-19 Watergate pulse train (Piotto et al., 1992). Decoupling using DIPSI-2 (Shaka et al., 1988) on $^1\text{H}$ and WALTZ-16 (Shaka et al., 1983) on $^{15}\text{N}$ is indicated by white rectangles. The triangle with $t_5$ represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped pulsed magnetic field gradients along the z-axis, with the following durations and strengths: G₁, 600 µs, 13 G/cm; G₂, 1000 µs, 37 G/cm; G₃, 1000 µs, 32 G/cm; G₄, 800 µs, 29 G/cm; G₅, 600 µs, 19 G/cm; G₆, 600 µs, 27 G/cm; G₇, 800 µs, 27 G/cm; G₈, 1000 µs, 37 G/cm; G₉, 1000 µs, 32 G/cm; G₁₀, 800 µs, 16 G/cm. Pulse phases other than x are indicated above the pulses. Phase cycling: $\phi_1 = \phi_5 = \psi_5 = \{x, -x\}$. The initial delays were $t_1 = t_5 = 14.0$ ms, $t_2 = t_4 = 4.7$ ms, and $t_3 = t_5 = t_6 = 0$ ms. Further delays were $\tau_1 = 2.7$ ms, $\tau_2 = 6.8$ ms, $\lambda_1 = 4.7$ ms and $\zeta_1 = 14.0$ ms. Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem (Kupce and Freeman, 2004b; Brutscher et al., 1995) with the phases $\psi_1, \psi_2, \psi_3, \psi_4$ and $\psi_5$ for $t_1, t_2, t_3$ and $t_5$, respectively, which were simultaneously incremented in 90°-steps for consecutive FIDs. Time evolution was performed in switched semi-constant time/constant-time fashion (Figure 1.13) for $t_1$ and $t_2$, in direct evolution fashion for $t_3$ and in constant-time fashion for $t_5$.

In the 5D APSY-HNCOCACB experiment, one resonance peak is expected per detectable amide moiety of the protein, correlating this amide moiety with the sequential CO, CA and CB atoms of the preceding residue, in dimensions 2, 3 and 4, respectively. For glycine residues, which lack the CB atom, the CA chemical shift is appearing twice, in dimensions 3 and 4. A particular feature of the pulse program relates to the glycine residues. At the onset of the first INEPT step transferring magnetization from CA to CB (Fig. 1.35), the proton decoupling is interrupted for a time of 6.8 ms. Without this interruption, the signs of glycine and non-glycine peaks would be opposite, leading to signal cancellations for some projection angles. During the 6.8 ms time period without decoupling, evolution due to $^{1}J_{HA,CA}$ couplings invert the signals of alpha-CH₂ groups relative to alpha-CH groups, leading to peaks of equal sign for all residues. In principle, the existence of positive and negative peaks could have advantages in some experimental situation for the distinction of two groups of peaks. In the
In the present situation, however, the glycine and non-glycine peaks can be unambiguously distinguished on the basis of their chemical shifts. Further, the introduction of the decoupling interruption does not prolong the pulse sequence and thus the use of equal sign for all peaks is the method of choice.

In the final peak list of the 5D APSY-HNCOCACB experiment, a correlation peak for each amide moiety was contained except for the amides of residues 29, 32, 73 and 99 with the side chains types Tyr, Glu, Arg and Tyr, respectively. The reason for the absence of these peaks is the high peak density in the 5D frequency space (Fig. 1.29). The four missing resonances are close in spectral space to the peaks involving the amide groups 72, 33, 30 and 72, and the same side chain types, respectively, and were not resolved with the present experimental parameters. It would probably be possible to resolve these peaks with more projections, possibly at a higher resolution. However, as the 5D APSY-HNCOCACB experiment does not yield connectivity information, but only contributes the CB chemical shifts, the lack of four peaks should hardly make any differences for the automated assignment and thus no further efforts to resolve these peaks were undertaken here.

During the assignment process using the 6D APSY-seq-HNCOCANH and the 5D APSY-HNCOCACB peak list as input, corresponding peaks from the two experiments have to be matched based on the chemical shifts of the four atoms HN, N, CO and CA. Thus, the precision of chemical shift measurement in these experiments is important. Comparison of the CO, CA, N and H chemical shift values in the 5D with the 6D experiment resulted in precisions of 1.3 Hz for HN, 1 Hz for N, 0.8 Hz for CO and 1.7 Hz for CA, respectively. These are again remarkably low values and the sequence specific resonance assignment with this input data can again be expected to be straightforward for the assignment algorithm GARANT. With the 5D and the 6D APSY peak lists as input, the backbone was assigned with the 6D peaks identically as previously and all available 5D peaks were added at the correct position (Figure 1.36). Thus, with the combined usage of the 5D APSY-HNCOCACB and the 6D APSY-seq-HNCOCANH experiment, the backbone atoms HN, N, CA, CO and 97% of the CB atoms were assigned in 35 hours of experiment time. In the case of OmpX, the 6D experiment alone is already sufficient to provide the complete sequence-specific backbone resonance assignment and the addition of the 5D experiment simply adds the CB chemical shift. However, depending on the amino acid sequence, some proteins could contain two fragments between prolines of equal length that can be only assigned unambiguously, if further information is available. In such cases, the recording of the 5D APSY-HNCOCACB experiment will be necessary.
Figure 1.36. Result of automated sequence-specific backbone resonance assignment of the denatured membrane protein OmpX with GARANT using the 6D APSY-seq-HNCOANH and the 5D APSY-HNCOACB peak lists as input. Above the amino acid sequence of OmpX, each correctly assigned 6D seq-HNCOANH peak is represented by a blue symbol, connecting two residues. Grey symbols show connections involving proline residues, for which no peak is expected. Green dots show the correctly assigned 5D peaks at the position of the residue, whose CB atom is contained. The four red dots are unassigned CB atoms due to missing 5D peaks in the 5D APSY-HNCOACB peak list.

An alternative implementation of strategy I–B is the 7D APSY-seq-HNCO(CA)CBCANH (Figure 1.38). In this experiment, two sequential amide groups are correlated with a CO, CA and CB atom. As in the 6D APSY-seq-HNCOANH experiment, the intraresidual peak is well suppressed. An inspection of the most sensitive projections of a single nucleus showed that the intraresidual peaks were suppressed below the noise level in our experiments.

Table 1.12. Sweep widths of the six indirect dimensions in the 7D APSY-seq-HNCO(CA)CBCANH experiment.

<table>
<thead>
<tr>
<th>dimension</th>
<th>nucleus</th>
<th>sweep width [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>$^1$HN</td>
<td>800</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>$^{15}$N</td>
<td>1850</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>$^{13}$CO</td>
<td>1100</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>$^{13}$CB</td>
<td>11000</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>$^{12}$CA</td>
<td>8000</td>
</tr>
<tr>
<td>$\omega_6$</td>
<td>$^{15}$N</td>
<td>1850</td>
</tr>
</tbody>
</table>

The sensitivity of the 7D APSY-seq-HNCO(CA)CBCANH experiment was found to be about 80% of the sensitivity of the 6D seq-HNCOANH experiment. The experiment was recorded with 98 projections in 45 minutes per projection in a total experiment time of 3 days. 192 complex points were recorded for each projection in the indirect dimension. The presence of the CB in one of the indirect dimension leads to a particular situation in this experiment compared to the 6D seq-HNCOANH. The sweep width of CB with 11000 Hz is 10 times larger than the sweep width of CO, which is 1100 Hz. Similar ratios exist between the CB and the N or HN dimension (Table 1.12). In combination with the strong chemical shift
degeneracies in denatured proteins, where CB atoms of the same side chain kind have often identical chemical shifts, the choice of reasonable projection angles has to be carefully considered.

![Diagram](image)

**Figure 1.37.** Pulse sequence used for the 7D APSY-HNCO(CA)CBCANH with urea-denatured OmpX. Radio-frequency pulses were applied at 118.0 ppm for $^{15}\text{N}$, at 173.0 ppm for $^{13}\text{C}'$, and, unless otherwise stated, at 42.0 ppm for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$. At the start of each transient, the $^1\text{H}$ carrier frequency was set at 8.24 ppm, indicated by “$\text{H}^\text{cm}$”, and at the time point “H₂O” the carrier was changed to 4.7 ppm. Black and white symbols represent 90°- and 180°-pulses, respectively. Unlabeled rectangular bars stand for rectangular pulses applied at maximum power. Pulses marked with capital letters have special lengths and shapes, depending on their purpose. Here, all pulse lengths are given for a $^1\text{H}$ frequency of 750 MHz: $^{13}\text{C}'$-pulses: A, 180°, rectangular shape, 38.26 µs; B, 90°, rectangular shape, 42.77 µs; C, 180°, I-Burp (Geen and Freeman, 1991), 220 µs. $^{13}\text{C}^\alpha$-pulses: D, 180°, I-Burp (applied at 51.0 ppm), 220 µs; E, 90°, Gaussian Cascade Q5 (Emsley and Bodenhausen, 1990), 280 µs; F, 180°, Gaussian Cascade Q3 (Emsley and Bodenhausen, 1990), 185 µs; H, 180°, rectangular shape, 38.26 µs. The $^{15}\text{N}$-pulses labelled with an asterisk are centred with respect to $t_3^a + t_3^b$ and to $t_2^c$, respectively. The $^{13}\text{C}'$-pulses labelled with an asterisk are centred with respect to $\rho + t/2$ and to $\rho - t/2$, respectively. The last six pulses on the $^1\text{H}$ line represent a 3-9-19 Watergate pulse train (Piotto et al., 1992). Decoupling using DIPSI-2 (Shaka et al., 1988) on $^1\text{H}$ and WALTZ-16 (Shaka et al., 1983) on $^{15}\text{N}$ is indicated by white rectangles. The triangle with $t_1$ represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped, pulsed magnetic field gradients along the z-axis with the following durations and strengths: $G_1$, 600 µs, 13 G/cm; $G_2$, 1000 µs, 37 G/cm; $G_3$, 800 µs, 16 G/cm; $G_4$, 800 µs, 34 G/cm; $G_5$, 600 µs, 19 G/cm; $G_6$, 600 µs, 27 G/cm; $G_7$, 800 µs, 13 G/cm; $G_8$, 1000 µs, 37 G/cm; $G_9$, 800 µs, 16 G/cm. Pulse phases other than x are indicated above the pulses. Pulse phase cycling: $\phi_1 = \psi_2 = \phi_3 = \{x, -x\}$, $\psi_4 = \{x, x, -x, -x\}$, $\psi_6 = y$. The initial delays were $t_1^4 = 2.7$ ms, $t_2^4 = t_2^5 = 14.0$ ms, $t_3^4 = t_3^5 = 4.7$ ms, $t_4^4 = t_4^5 = 14.0$ ms and $t_1^6 = t_1^7 = t_2^6 = t_2^7 = t_3^6 = t_3^7 = t_4 = t_5 = t_6^5 = 0$ ms. Further delays were $\tau = 2.7$ ms, $\zeta = 14.0$ ms, $\eta = 6.8$ ms, $\lambda = 4.7$ ms, $\rho = 20.75$ ms. Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem (Kupe and Freeman, 2004b; Brutscher et al., 1995) with the phases $\psi_1$, $\psi_2$, $\psi_3$, $\psi_4$, $\psi_6$, $\psi_7$, $\psi_8$, $\psi_9$, and $\psi_{10}$ for $t_1$, $t_2$, $t_3$, $t_4$, $t_5$ and $t_6$, respectively, which were simultaneously incremented in 90°-steps for consecutive FIDs. Time evolution was performed in switched semi-time/constant-time fashion (Figure 1.13) for $t_1$, $t_2$, $t_3$ and $t_6$, in direct evolution fashion for $t_4$ and in constant-time fashion for $t_5$.

The situation is illustrated in Figure 1.38. For a pair of peaks with degenerate chemical shifts in the CB dimension, but non-degenerate shifts in the CO dimension, a projection with an intuitively chosen intermediate projection angles of e.g. 42° will be similar not yield much new information for the analysis with GAPRO compared to the projection at 0°,
corresponding to the direct CB projection. A better choice for the projection angle is the angle $\phi^*$, which projects the two sweep widths to the same size (Fig. 1.39).

We have termed $\phi^*$ the “sweep-width matching projection angle(s)” or short, the “matching angle(s)”. For two dimensions i and j, the matching angle is calculated by

$$\tan \alpha^* = \frac{SW_i}{SW_j} \quad (1.23)$$

where $SW_i$ and $SW_j$ are the sweep widths of the two dimension (Fig. 1.39). The matching angle for the addition of a third dimension k with sweep width $SW_k$ is given by.
\[
\tan \beta^* = \sin \alpha^* \frac{SW_j}{SW_k}
\] (1.24)

Similar expressions can be found for more dimensions. Depending on the ratios between the sweep widths, \(\phi^*\) can assume values close to the 0° or 90°. In the case of CB and CO with 11000 Hz and 1100 Hz, respectively, the matching angle is 84° (Fig. 1.38). Using these considerations, all 98 projection spectra of the 7D APSY-seq-HNCO(CA)CBCANH experiment were recorded using sweep-width matching projection angles for all dimensions.

**Figure 1.39.** Calculation of the sweep-width matching projection angle \(\alpha^*\) for two indirect dimensions \(i\) and \(j\) with sweep widths \(SW_i\) and \(SW_j\), respectively, defining the spectrum (blue rectangle). The projection angle \(\alpha^*\) projects the two sweep widths to the same size (red line).

**Figure 1.40.** Result of automated sequence-specific backbone resonance assignment of the denatured membrane protein OmpX with GARANT using the 7D APSY-seq-HNCO(CA)CBCANH peak lists as input. Above the amino acid sequence of OmpX, each correctly assigned 7D seq-HNCO(CA)CBCANH peak is represented by a blue symbol. Grey symbols show connections involving proline residues, for which no peak is expected. The three red dots are expected 7D peaks, which were not contained in the 7D APSY-seq-HNCO(CA)CBCANH peak list.
The projection spectra from the 7D seq-HNCO(CA)CBCANH experiment was analyzed with the same parameters as the 6D experiment. The resulting peak list contained all expected 7D peaks except three, which were absent due to reduced signal intensity. The absent peaks were those connecting residues 72/73 and 99/100, which were already weak in the 6D experiment and additionally the peak connecting residues 100/101. As this latter peak also involves the chemically exchanging His 100, its absence is well rationalizable. Further, four additionally correct peaks, but no artifacts were contained in the final result. Using the 7D peak list as input for GARANT, a sequence-specific backbone resonance assignment of denatured OmpX was obtained, where all residues covered by the detected peaks were correctly assigned (Fig. 1.40).

In summary, the results obtained in this chapter show that APSY can be used beneficially for the sequence-specific backbone resonance assignment of denatured proteins. In very short experiment times of 1–2 days, complete or nearly complete peak lists can be obtained that provide the ideal input for an assignment algorithm. It seems likely that the same experiments can be repeated with other denatured proteins of similar size. The technique should also be applicable to intrinsically unstructured or otherwise denatured proteins. Compared to conventional techniques, the savings in instrument times and human work are substantial and the described approach thus opens the door to systematic high-throughput studies on larger sets of denatured proteins.
2. Characterization of the denatured form of the membrane protein OmpX

Introduction to studies of proteins under denaturing conditions

Folding into a unique three-dimensional structure is a requirement for the biological activity of proteins. For several decades it is known that the structure is encoded in the amino acid sequence of proteins (Anfinsen, 1973), but still the understanding of protein folding remains one of the central challenges in structural biology. Detailed insight into any protein folding mechanism will require characterization of all important states along the folding pathway, including the “denatured state”, which may possibly include local residual non-random structure and restricted internal dynamics. In this context, studies under denaturing conditions may reveal non-random conformations which are not sufficiently accumulated under solution conditions that favor folding, although they may appear early in the folding pathway. Residual non-random structure and local intramolecular interactions found under strongly denaturing solution conditions may thus be indicative of nucleation sites for the folding process under physiological conditions (Radford et al., 1992; Shortle, 1993; Wüthrich, 1994; Dyson and Wright, 1998; 2001). For membrane-associated proteins, structural characterization of non-native states is further of interest because of their potential role in the transport of proteins across the membranes and in cellular processes such as signal transduction. Added interest in such studies comes from the demonstration that numerous proteins are intrinsically unstructured in their biologically functional states, and only become structured upon binding to their biological target (Plaxco and Gross, 1997; Dunker et al., 2001; Wright and Dyson, 1999; Dyson and Wright, 2005).

In the denatured state, a polypeptide chain generally loses its tertiary and secondary structure, however, for many proteins in denaturing conditions, non-random behavior of short segments of the polypeptide chain has been reported, which can be rationalized with the assumption that hydrophobic forces lead to local clustering of amino acid side chains. (Neri et al., 1992a; Wüthrich, 1994; Logan et al., 1994; Lumb and Kim, 1994; Frank et al., 1995; Schwalbe et al., 1997; Fong et al., 1998; Mok et al., 2000; Demarest et al., 2001). For some proteins, hydrophobic clusters sample native-like secondary structure and may thus be potential nucleation sites for the folding process (Garvey et al., 1989; Neri et al., 1992a; Hodsdon and
For example, tertiary contacts between multiple hydrophobic clusters in denatured lysozyme were inferred from recent experimental studies (Klein-Seetharaman et al., 2002), \textit{i.e.}, a single amino acid exchange W62G in one of six hydrophobic clusters was found to affect the transverse $^{15}$N NMR relaxation rates in all six clusters. The possibility of such long-range interaction between multiple hydrophobic clusters could be an important new facet of protein folding (Baldwin, 2002). It is therefore of interest to further investigate whether such long-range interactions are a general feature of denatured proteins.

The use of NMR spectroscopy for the characterization of denatured proteins is especially useful as information at atomic resolution can be obtained (Wüthrich, 1986; 1994; Dyson and Wright, 2002). For the identification of hydrophobic clusters, usually a number of NMR parameters are recorded and analyzed (Dyson and Wright, 2002): chemical shifts, coupling constants, temperature coefficients, amide proton exchange rates, spin relaxation data, and nuclear Overhauser effects (NOEs). In principle, each parameter could detect the presence of hydrophobic clusters, but as all of these parameters are population averaged over the ensemble of structured and unstructured conformers, their interpretation is not always straightforward (Wüthrich, 1994). Usually, most of the mentioned parameters must be analyzed jointly to determine the existence and the properties of significantly populated hydrophobic clusters in proteins dissolved under denaturing conditions (Dyson and Wright, 2002).

The folding mechanisms of integral membrane proteins are biophysical processes of high complexity both \textit{in vivo} and \textit{in vitro} (Kleinschmidt, 2003; Booth et al., 2001; Popot and Engelman, 1990; White and von Heijne, 2004; Engelman et al., 2003; Tamm et al., 2004). The topology of the folded state of integral membrane proteins is adapted to an environment of alternating hydrophobic and hydrophilic phases and so their folding and insertion mechanism must deal simultaneously with both phases (White and Wimley, 1999). For many $\beta$-barrel membrane proteins efficient refolding was achieved from denatured states into micelles, bilayer vesicles or other model membranes (Surrey and Jähnig, 1992; Buchanan, 1999; Booth et al., 2001; Tamm et al., 2004). Typical \textit{in vitro} refolding procedures start from denaturing aqueous solutions, \textit{e.g.} solutions containing high molar concentrations of the chemical denaturants urea or guanidine hydrochloride. Subsequently, the micelle or bilayer vesicle concentration is raised, often with a simultaneous dilution of the sample (Buchanan, 1999). Refolding is usually exothermic and occurs spontaneously; however, the refolding rates and yields depend on several parameters, such as the temperature, pH and the type of
micelle or bilayer. For instance, the eight-stranded β-barrel of OmpA from *E. coli* refolds into micelles within minutes and into lipid bilayers within hours (Kleinschmidt and Tamm, 1996; Kleinschmidt et al., 1999; Tamm et al., 2004).

The present work attempts a characterization of the urea-denatured form of the outer-membrane protein OmpX from *E. coli*. The protein consists of 148 amino acid residues, forming an eight stranded β-barrel protein in its native form (Fernández et al., 2004; Vogt and Schulz, 1999). The unfolded state of OmpX in 8 M urea at pH 6.5 was characterized using high-resolution NMR data. It was found to contain two regions with non-random conformations, which comprise the residues 73–82 and 137–145, respectively. Structural models for the two clusters were calculated. The possibility of a long-range interaction between the two clusters was investigated using variants of OmpX with single amino acid substitutions, and a peptide analog of OmpX(70–87).

Further, the ternary system 8M urea–OmpX–DHPC micelles was characterized under equilibrium conditions in aqueous solution using high-resolution NMR and fluorescence spectroscopy. It was found that urea-denatured OmpX binds to micelles specifically with the two hydrophobic clusters characterized in 8 M urea. Binding constants and standard transfer free energies were determined, and structural models of the hydrophobic clusters in OmpX–micelle complexes were calculated. As binding of urea-denatured OmpX to DHPC micelles occurs in the early stage of the standard refolding protocol, the data presented here are a basis for further studies on the folding mechanism OmpX and other β-barrel membrane proteins.

Finally, it is suggested to use the specific binding interaction with detergent micelles as a general tool for an efficient identification of hydrophobic clusters in urea-denatured proteins. The binding to micelles causes chemical shift changes which permit an unambiguous identification of the residues involved in the interactions with the detergent micelles. The use of this approach is demonstrated with the denatured forms of the proteins OmpX, 434-repressor(1–63) (Neri et al., 1992a) and with two OmpX variants which each lack one of the two hydrophobic clusters present in the wild type protein.
Identification of nonrandom structure in urea-unfolded OmpX

Based on sequence-specific resonance assignments for the complete backbone and most of the side chains, a set of three heteronuclear resolved 3D NOESY spectra was analyzed to identify NOEs between pairs of protons. The 3D $^{15}$N-resolved [$^{1}$H, $^{1}$H]-NOESY spectrum of OmpX in 8 M urea showed 783 unambiguously assigned NOEs, among which 95% were intraresidual, sequential and medium-range $d(i,i+2)$ NOEs. This coincides with model considerations which predict that sequential $d_{\text{NN}}$, $d_{\alpha\text{N}}$ and $d_{\beta\text{N}}$, and medium-range $d_{\text{NN}}(i,i+2)$, $d_{\alpha\text{N}}(i,i+2)$ and $d_{\beta\text{N}}(i,i+2)$ NOEs should have detectable intensities for all residues in “random coil” polypeptide chains (Wüthrich, 1986; Smith et al., 1996). In contrast to this homogeneous distribution of sequential and $d(i,i+2)$ NOEs throughout the sequence, the presence of longer-range NOEs is limited to two short segments of the OmpX sequence, which comprise the residues 73–82 (I) and 137–145 (II) (Figure 2.1). This augmented occurrence of longer-range NOEs indicates that the segments I and II adopt sizeable populations of non-random local conformations in 8 M urea solution. Besides the NOESY information, two further experimental observations support this conclusion: Residues with large $^{1}$H$^\alpha$ and $^{1}$H$^\text{N}$ chemical shift differences from the random coil values and residues with outstanding temperature coefficients of the amide proton chemical shifts (Ohnishi and Urry, 1969; Baxter and Williamson, 1997) are mostly located in the segments I and II of the polypeptide (Figure 2.1). For the two polypeptide segments I and II, structure calculations with the program DYANA (Güntert et al., 1997) were performed (Figures 2.2 and 2.3). For the segment I, a total of 31 $d(i,i+3)$, $d(i,i+4)$ and $d(i,i+5)$ NOE cross peaks were observed in 3D $^{15}$N- and $^{13}$C-resolved [$^{1}$H, $^{1}$H]-NOESY spectra. Figure 2.2 shows that with this input the polypeptide backbone of residues 73–81 is well defined, but all amino acid side chains are conformationally disordered. This is reflected by RMSD values of 1.06 ± 0.28 Å for the backbone atoms, and 1.79 ± 0.37 Å for all heavy atoms. The residual DYANA target function value and the residual constraint violations were very small, indicating that the input data represent a self-consistent set, but do not represent tight constraints on the conformation. The hydrophilic side chains of Asn74, Asp75 and Ser78 point towards the solvent. The side chains of Ile73, Trp76 and Ile79 are located close to each other as evidenced by NOEs between them, and by the observation that the $^{1}$H resonances of Ile79 are shifted upfield by ring current effects from Trp76 (Figure 2.1).
Figure 2.1. Amino acid sequence of OmpX with a survey of the sequential and medium-range NOE connectivities, and supplementary data collected for identification of residual structure in 8 M urea. The medium-range connectivities are shown by lines starting and ending at the positions of the two residues related by the NOEs; “sc” stands for “any side chain proton”. The data were obtained from a 3D 15N-resolved-[1H,1H]-NOESY spectrum (mixing time = 100 ms) and two 3D 13C-resolved-[1H,1H]-NOESY spectra (mixing time = 120 ms) recorded with the carrier frequency in the aliphatic and the aromatic region, respectively. In the row “Δδ”, a rectangle identifies the residues, for which a significant 1H or 1H chemical shift deviation relative to the corresponding random coil values was observed. In the row “Δδ/ΔT”, residues with amide proton temperature coefficients smaller than 4 ppb/K are identified with a triangle (Tafer et al., 2004). The locations of the eight β-strands in folded OmpX in DHPC micelles are indicated at the bottom with “β1” to “β8”. Two segments with non-random structure are underlined and labeled I and II.

There are also NOEs from the Trp76 side chain to other protons in the aromatic chemical shift region. These other aromatic proton shifts most likely belong to Tyr80 but could not be unambiguously assigned and were thus not considered for the structure calculations. The side chains of Trp76 and Tyr80 are therefore not well defined in the structure, although there might be a hydrophobic interaction between them. The polypeptide backbone is well defined in a 3_10-helical conformation. In 17 of the 20 conformers, at least one of the two backbone hydrogen bonds Trp76 NH–O Ile73 and Ala77 NH–O Asn74 is present, and in 9 of the
conformers both of them are formed. The formation of these backbone hydrogen bonds is also supported by the observation of low temperature coefficients for the residues 76 and 81 (Baxter and Williamson, 1997). The picture emerges that the hydrophobic Ile73–Trp76 interactions as well as for technical reasons unidentified Trp76–Tyr80 interactions drive the backbone to adopt its helical conformation.

Figure 2.2. NMR structure of the polypeptide segment I (Figure 2.1) in urea-unfolded OmpX calculated from the NOE upper distance constraints with |i-j| ≥ 3. (a) Bundle of the 20 best energy-minimized conformers superimposed for minimal RMSD of the backbone atoms of residues 73–82, showing the polypeptide backbone. (b) Ball-and-stick presentation of the conformer with the lowest residual DYANA target function value in the same orientation as (a). The experimental upper-limit NOE distance constraints used as input for the structure calculations are shown as yellow lines. (c) Stereo view of the all-heavy-atom presentation of the bundle in (a), where the backbone is drawn with a thicker line than the side chains. The figures have been prepared with the program MOLMOL (Koradi et al., 1996).

For the segment II, a total of 29 $d(i,i+3)$, $d(i,i+4)$ and $d(i,i+5)$ NOE cross peaks were observed. Figure 2.3 shows that the polypeptide backbone of residues 137–145 and the orientation of the side chains of Trp140, Ile141, Ala142, and Val144 are well defined, whereas the side chains of Val137 and Thr139 are disordered. The RMSD values are 0.50 ± 0.08 Å for the backbone atoms, and 0.75 ± 0.07 Å for all heavy atoms. Close packing contacts are observed on both sides of the indole ring of Trp140, with Val137 covering the indole ring on one side, and the backbone atoms of Gly143, Val144, and Gly145 covering the other side. The methyl groups of Ile141, Ala142, and Val144 are all clustered against each
other. The upfield proton chemical shifts for residues 141, 142, 144, and 145 seem to be a consequence of the packing against the indole ring of Trp140.

**Figure 2.3.** NMR structure of the polypeptide segment II in urea-unfolded OmpX. Same presentation as in Figure 2.2. The superposition of the 20 best conformers is for best fit of the residues 137–145.

A major issue to be treated is the extent to which the local structures of Figures 2.2 and 2.3 are populated in OmpX solutions in 8 M urea. A starting point for this discussion is the observation of a single set of NMR lines for the OmpX polypeptide chain, which would be compatible either with the presence of only a single conformation or with rapid interconversion, on the chemical shift timescale, of a manifold of different conformers (Wüthrich, 1994).

For the calculations of the structures in Figures 2.2 and 2.3 exclusively NOE distance constraints were used as input, and assumed that all observable long-range NOEs originate from the same OmpX molecules (see below). The observed NOE intensities are an average over the contributions from all rapidly interconverting random and non-random OmpX conformations (Wüthrich, 1994). In fully randomized unfolded proteins, intense sequential $d_{NN}$ and $d_{\alpha N}$ NOEs, as well as weaker medium-range $d_{NN}(i,i+2)$ and $d_{\alpha N}(i,i+2)$ NOEs would therefore be expected (Wüthrich, 1986; Smith et al., 1996). Because of the $1/r_{ij}^6$ dependence of the NOE, where $r_{ij}$ is the distance between the two protons related by the NOE, $i$ and $j$, non-random conformers containing short distances between discrete pairs of hydrogen atoms
can make a dominant contribution to the NOEs between these proton pairs, even if they represent only a small fraction of the structural ensemble. With the assumption that the contributions from fully randomized molecules to longer-range NOEs, with $|i-j| \geq 3$, are negligibly small, the input for the structure calculations in Figures 2.2 and 2.3 included only these longer-range NOEs. Since for both non-random polypeptide segments all observed upper-limit NOE distance constraints with $|i-j| \geq 3$ were simultaneously satisfied by a single, quite well-defined structure (Figures 2.2 and 2.3), it can be concluded that all other conformations present in 8 M urea did indeed not significantly contribute to the NOEs used as input for the structure calculations. This then also implies that no other non-random conformations with similar short $^1$H–$^1$H distances are significantly populated in these solutions.

Two different experimental observations enable to obtain an estimate of the population of the non-random conformations. Firstly, the contribution from a given molecular species to the observed average $^1$H chemical shifts is weighted simply by its population in solution. The different averaging of NOE-intensities and chemical shifts is clearly manifested if the chemical shift deviations from the random coil values (Figure 2.1) are compared with the predictions from ring current calculations for the aliphatic protons in the structures of Figures 2.2 and 2.3. With the implicit assumption of 100% population of these structures, these calculations predict larger deviations from the random-coil shifts than those observed, leading to estimates for the populations of the local structures I and II of 30% and 25%, respectively.

Secondly, a statistical analysis of the relative intensities of short-range and longer-range NOE cross peaks opens another, independent way for estimating populations in the ensemble of conformers represented by the NMR spectrum. The simple fact that the $d(i,i+4)$ and $d(i,i+5)$ NOEs have readily detectable intensity shows that the population of the non-random structures must be not less than about 10% (Wüthrich, 1994). A more precise value for this lower bound can be obtained by accounting for the relative intensities of all observed NOEs in the polypeptide segments with and without residual non-random structure. In this novel approach it is assumed that the observed intensities of the $d(i,i+3)$, $d(i,i+4)$ and $d(i,i+5)$ NOEs arise entirely from the non-random conformation, whereas the observed intensity of the intraresidual, sequential and $d(i,i+2)$ NOEs arises from the entire population of protein molecules. Since the intensities of the intraresidual, sequential and medium-range $d(i,i+2)$ NOEs in the structured segments can be calculated from the corresponding $^1$H–$^1$H distances in the NMR-structures (Figures 2.2 and 2.3), their relative populations can be assessed from the observed signal intensities for these NOEs. Pairs of protons with scalar coupling were
excluded from this analysis, to exclude possible errors that could arise from contributions to the NOE intensities from ZQ transitions (Macura et al., 1982a, b). The $d(i,i+3)$, $d(i,i+4)$ and $d(i,i+5)$ NOEs were used for the calibration of the NOE cross peak intensity-to-$^1$H-$^1$H distance ratio. Using the statistical F-test, lower bounds for the populations of the non-random conformations I and II were thus obtained to be $\geq 30\%$ and $\geq 20\%$, respectively, which is in close agreement with the corresponding data from the comparison of NOE and chemical shift averaging.

The assumption of cooperative folding of the local non-random structures needs some detailed discussion. For non-globular polypeptide chains with lowly populated local non-random conformations, one cannot a priori distinguish between the following two limiting situations, or combinations thereof: (i) The individual NOE-observable long-range $^1$H-$^1$H contacts are statistically distributed among the entire population of polypeptide chains. (ii) There is co-operative folding of local non-random structure, and all the NOE-observable long-range $^1$H-$^1$H contacts are located in the same polypeptide chains and fulfilled simultaneously.

As mentioned at the outset of this discussion, the structural interpretation of the NMR data presented in Figures 2.2 and 2.3 is arbitrarily based on the situation (ii). The results obtained support that this working hypothesis is reasonable for the treatment of OmpX in 8 M urea, since within the precision of the experimental measurements there are no inconsistencies between the different NMR data. In particular, the ensemble of all NOE distance measurements and of all chemical shift data, which follow different rules for ensemble averaging, are compatible with the unique bundles of conformers in the Figures 2.2 and 2.3. Additional support for the assumption of co-operative folding in the OmpX polypeptide chain in 8 M urea appears to come from considerations on the life-time of the local non-random structures. The long-range NOE distance constraints used as input for the structure calculations derive from the observation of negative NOEs, i.e., from $^1$H-$^1$H NOEs related to effective correlation times longer than one nanosecond, which would then also be a lower limit on the lifetime of the local structures defined by the NOEs. It seems unlikely that such long-lived local structures would prevail in the absence of co-operative effects including a sizeable number of simultaneous close interatomic contacts (Figures 2.2 and 2.3). In contrast, since the local non-random structures I and II have been calculated independently, no statements can be made on whether there is a cooperative effect, with the two structures predominantly present either only in the same molecules or only in different molecules, or rather a statistical distribution of the two local structures among all OmpX molecules. This question is investigated further in the following chapter.
Comparison of the structure of denatured OmpX (Figures 2.2 and 2.3) with the structure of the folded protein in DHPC micelles (Fernández et al., 2001a; b; 2004) and in mixed crystals with n-octylpolyoxyethylene (C₈POE) (Vogt and Schulz, 1999) shows that the residual structures in 8 M urea do not coincide with the spatial arrangement of the corresponding peptide segments in the folded protein. The residues 73–77 in the structure I (Figure 2.2) correspond to the turn connecting the β-strands β₄ and β₅, and all other residues in the structures I and II are in β-strands in the folded protein (Figure 2.1). In the urea-unfolded OmpX, the non-random structure II appears to be stabilized by hydrophobic interactions in a small cluster of amino acid side chains formed around Trp 140 (Figures 2.2 and 2.3). These hydrophobic contacts would sterically not be possible in the folded OmpX, but hydrophobic clustering around Trp residues has been reported previously in studies of unfolded proteins and short peptides (Neri et al., 1992a; Schwalbe et al., 1997; Klein-Seetharaman et al., 2002).
Independent formation of hydrophobic clusters

OmpX in 8 M urea with its two hydrophobic clusters provides an ideal model system for probing long-range interactions between such clusters. A polypeptide chain with two separated hydrophobic clusters can adopt four possible states, $\alpha$, $\beta$, $\gamma$, and $\delta$, in which none, the first, the second, and both clusters are formed, respectively (Figure 2.4(a)). The experimentally observed populations of the two hydrophobic clusters are the starting point of the analysis. For simplicity, the same population $q = 25\%$ is assumed here for each of the two clusters of denatured OmpX. This assumption fits the data well within experimental error. Equations for a situation with differently populated clusters or more than two hydrophobic clusters can be derived analogously. Knowledge of the observed populations $q$ is however not sufficient to determine the populations $p_i$ of the four states ($i = \alpha$, $\beta$, $\gamma$, $\delta$), since $q = p_\beta + p_\delta$ and $q = p_\gamma + p_\delta$. The populations further depend on a possible interaction between the clusters.

Three scenarios are imaginable for the interaction between the clusters: (i) A stabilizing interaction leading to molecules with predominantly two clusters in the same molecules; (ii) A destabilizing interaction leading to predominantly one cluster per molecule; (iii) the

Figure 2.4. Schematic population analysis for the possible states of a polypeptide with hydrophobic clusters: (a) polypeptide chain with two clusters representing denatured wt OmpX and (b) polypeptide chain with one cluster representing denatured OmpX W140A. The denatured polypeptide chain is indicated by a black line. Black and grey circles symbolize cluster I and II, respectively, a cross denotes the disrupted cluster II in OmpX W140A. On the lines (i), (ii), (iii) the relative populations are given below each state for three scenarios: (i) completely cooperative interaction between the clusters, (ii) completely anti-cooperative interaction between the clusters, and (iii) non-interacting clusters. The populations are based on an observed population $q$ for each cluster in the wild type (The relations given are valid for $0 \leq q \leq 0.5$).
absence of an interaction between the clusters resulting in a statistical distribution of the hydrophobic clusters among the OmpX molecules. Figure 2.4(a) shows the resulting populations of the four states in the limiting cases of (i) complete cooperativity and (ii) complete anti-cooperativity and (iii) for no interaction. Scenarios with intermediate interactions would result in averaged population numbers between these cases, with weighting factors depending on the strength of the interaction.

If one of the clusters is disrupted, so that it can no longer be formed by the polypeptide chain, two states $\alpha'$ and $\beta'$ remain (Figure 2.4(b)). The resulting populations $p_\alpha$ and $p_\beta$ can be calculated using the equilibrium condition for the formation of the remaining cluster $p_\alpha / p_\beta = p_{\alpha'} / p_{\beta'}$. The population of the remaining hydrophobic cluster, $q' = p_{\beta'}$, is depending on the scenario: (i) A stabilizing interaction between the intact clusters leads to $q' < q$, in the limiting case of complete cooperativity to $q' = 0$. (ii) A destabilizing interaction leads to $q' > q$, in the limiting case of complete anti-cooperativity to $q' = q / (1-q)$. (iii) For non-interacting clusters, the population is unchanged $q' = q$.

Measurement of the population of one cluster after disruption of the other cluster thus allows a distinction between the scenarios (i)–(iii) and therefore to detect and characterize a possible interaction between the hydrophobic clusters. Based on these considerations, the following experiment was designed. Two single amino acid variants of OmpX were identified in each of which one of the hydrophobic clusters is disrupted. In these variants the population of the other, unmodified cluster was measured.

For the determination of populations in an ensemble of rapidly interconverting states, the chemical shift is a suitable and sensitive physical property. The chemical shift $\delta$ for a given nucleus of the ensemble is given by $\delta = \sum p_i \delta_i$, where $p_i$ is the population of state $i$, $\delta_i$ is the chemical shift of state $i$ and the summation covers the ensemble with $\sum p_i = 1$. The random coil can be considered to be a single state representing the rapid equilibrium between numerous random conformations of the polypeptide chain (Wüthrich, 1994; Smith et al., 1996; Mohana-Borges et al., 2004). The $^1$H and $^{15}$N chemical shift values in the random coil state, $\delta_{rc}$, are in first order close to a standard value for the amino acid type and depend in second order on the neighboring sequence. Correction factors for the sequence effects have been determined for up to 2 amino acids in each direction of the polypeptide (Braun et al., 1994; Wishart et al., 1995a; Schwarzinger et al., 2001). The non-random structures of the hydrophobic cluster are considered a second state with chemical shifts $\delta_{hc}$ that can be strongly different from the random coil values (Wüthrich, 1994). The resulting chemical shift $\delta$ of a
nucleus in a hydrophobic cluster with population $q$ is thus given by $\delta = \delta_{rc}(1-q) + \delta_{hc}q$. For the chemical shift difference $\Delta\delta$ between the wild type with cluster population $q$ and the variant protein with cluster population $q'$ follows $\Delta\delta = \Delta\delta_{hc} (q' - q)$, where $\Delta\delta_{hc} = \delta_{hc} - \delta_{rc}$. Rewriting yields

$$q' = q + \Delta\delta / \Delta\delta_{hc} \tag{2.1}$$

Equation 2.1 is used below to calculate the population differences.

Figure 2.5. 2D $^{15}$N,1H]-HSQC NMR fingerprints showing the influence of single amino acid replacements on the hydrophobic clusters in OmpX denatured in 8 M urea. (a) wild type OmpX, (b) OmpX[W76A], (c) OmpX[W140A]. Green and orange circles mark the resonances of amide moieties involved in the two hydrophobic clusters I and II, respectively, which are observed in the wild type protein. The resonance assignments are indicated in (a) using the one-letter amino-acid code and the sequence number. Grey circles in (b) and (c) point out chemical shift changes of the cluster residues due to the amino acid exchange: corresponding circles are connected by green (b) and orange (c) lines unless they touch each other. The insets show the Trp indole resonances with the assignments given in (b) and (c).

Single amino acid variants of OmpX were created by mutagenesis to generate candidates for OmpX variants devoid of clusters. NMR spectra of wild type OmpX and the two variants
OmpX[W76A] and OmpX[W140A] are shown in Figure 2.5. Most resonances are observed to be at the same position in all three spectra, however, in the variant OmpX[W76A] (Figure 2.5(b)) the resonances of all residues belonging to cluster I are shifted away from their positions in wild type OmpX (Figure 2.5(a)). The resonances of all residues involved in cluster II are similarly shifted in the variant OmpX[W140A] (Figure 2.5(c)). The sequence effects of amino acid exchanges on the chemical shift are expected to comprise the two neighboring residues in each direction of the unstructured polypeptide (Schwarzinger et al., 2001). The observation that a single amino acid exchange in denatured OmpX affects 8 or 9 residues thus indicates that tertiary contacts between side chains have been altered. The chemical shifts in the variants are closer to random coil values than in the wild type, which is particularly obvious for the downfield shifts for residues I79, G81, I141 and A142. An inspection of 3D $^{15}\text{N}$-resolved $[^{1}\text{H},^{1}\text{H}]$-NOESY spectra showed that the medium- and longer-range NOEs, which were observed between protons in the clusters of wild type OmpX, are no longer present for the residues in the “cluster segments” where Trp was replaced by Ala. Thus, in the amino acid variants OmpX[W76A] and OmpX[W140A], clusters I and II, respectively, are not formed and the corresponding polypeptide segments are in a random coil state. The amino acid variants OmpX[W76A] and OmpX[W140A] are thus suitable systems for the characterization of long-range forces of the type discussed in Figure 2.4.

The chemical shift differences between wild-type and variant OmpX were also analyzed by calculating combined deviations $\Delta\delta(\text{HN})$ of the $^{15}\text{N}$ and the $^{1}\text{H}$ resonances for each individual amide moiety, using the equation $\Delta\delta(\text{HN}) = \sqrt{\Delta\delta^{2}(^{1}\text{H}) + (0.2\cdot\Delta\delta^{2}(^{15}\text{N}))}$, where $\Delta\delta(^{1}\text{H})$ and $\Delta\delta(^{15}\text{N})$ are the chemical shift differences in ppm for the amide proton and the amide nitrogen, respectively, between the spectrum of wild type OmpX and the spectrum of the species where amino acids were exchanged (Grzesiek et al., 1996; Pellecchia et al., 1999) (Figure 2.6). Before the calculation of $\Delta\delta(\text{HN})$, the chemical shift values of the four neighboring residues were corrected for the sequence effects of the amino acid exchange (Schwarzinger et al., 2001).

The standard deviation of the $\Delta\delta(\text{HN})$ chemical shift measurement is 0.004 ppm, as determined from the values of residues 1–60 and 90–130, which are not involved in hydrophobic clusters. We therefore consider chemical shift changes $\Delta\delta(\text{HN}) > 0.01$ ppm, which corresponds to 2.5 standard deviations, to be significant. With this definition, the variant OmpX[W76A] shows significant chemical shift changes for residues 71–85, and the variant OmpX[W140A] shows significant changes for residues 134–148 when compared to
wild-type OmpX. These changes coincide closely with the segments 73–82 and 137–145 that were previously identified to form the two clusters. The small changes around His 100 probably arise from small pH differences between the samples, which may have affected the protonation state of the imidazole ring (The nominal pH value of 6.5 in this experiment is close to the random coil pKₐ value of histidine).

![Figure 2.6. Chemical shift difference ∆δ(HN) relative to wild type OmpX, corrected for sequence effects (Schwarzinger et al., 2001) (a) OmpX[W76A], (b) OmpX[W140A]. The data for residues 73–82 (cluster I) and 137–145 (cluster II) are colored green and orange, respectively.](image)

In both OmpX variants, no significant chemical shift changes are observed for residues in the cluster for which the sequence was not varied. Upper limits for the population difference between wild type OmpX and the variants can be calculated with Equation 2.1, if values of ∆δₘ are known. Considering that each amino acid exchange destroys the corresponding cluster, ∆δₘ can be obtained from the observed chemical shift changes for the residues involved in a cluster that is destroyed, as these chemical shift changes correspond to a fraction q of ∆δₘ. For a conservative estimate of ∆δₘ, we use a value of 0.1 ppm or larger for the chemical shift change caused by the amino acid exchange, which is true for 7 and 4 residues in cluster I and II, respectively. With q = 25%, for these residues follows ∆δₘ ≥ 0.1 ppm / q = 0.4 ppm. Using the standard deviation of 0.004 ppm for the chemical shift measurement of ∆δ, Equation 2.1 yields q' = q ± 0.01 for each of the two clusters. Thus, any scenario (i) or (ii)
with an interaction between the two hydrophobic clusters that would change the populations by more than 1% compared to a statistical distribution can be clearly ruled out. The most likely case for the mutual impact of the clusters on each other is therefore scenario (iii) with non-interacting clusters, resulting in a statistical distribution of the two hydrophobic clusters among the molecules of an ensemble. With a population level of $q = 25 \pm 5\%$ for each cluster, in $6 \pm 2\%$ of the molecules, both clusters I and II are formed, in $38 \pm 6\%$ one of the two clusters is present and in the remaining $56 \pm 6\%$ of the molecules, the whole polypeptide chain is in a random coil conformation.

The experiments with the amino acid variants imply that the formation of hydrophobic clusters in denatured OmpX is encoded entirely in the local amino acid sequence. This conclusion is additionally supported by NMR experiments with the polypeptide H–AYRINDWASIYGVGVGY–OH, which corresponds to residues 70–87 of wild type OmpX. NMR measurements of an unlabeled sample of this polypeptide were done under identical solution conditions as for the full length protein. Based on homonuclear 2D $[^1H,^1H]$-NOESY and 2D $[^1H,^1H]$-TOCSY NMR experiments, complete sequence-specific proton resonance assignment was obtained. The proton chemical shifts for the peptide were found to be virtually identical to those in the full length protein. This is illustrated for residues N74 and I79 with strips from a 3D $^{15}N$-resolved $[^1H,^1H]$-TOCSY spectrum of the full length protein and the corresponding amide proton strips from the 2D $[^1H,^1H]$-TOCSY of the peptide (Figure 2.7(a)).

The chemical shift differences between the peptide and the full length protein of all 49 protons in the residues 73–82 follow a statistical Gaussian distribution about 0, with a standard deviation of 0.016 ppm. The differences between wild type protein and peptide are thus much smaller than the deviations of these chemical shifts from random coil values, which are for most protons in residues 73–82 larger than 0.1 ppm, reaching 0.4 ppm for some resonances (Tafer et al., 2004). This is particularly obvious for the side chain resonance frequencies of I79 (Figure 2.7(b)). Because of the strong dependence of the chemical shift on structural and population changes, the observation of identical chemical shifts for all protons indicates that cluster I is formed in the peptide OmpX(70–87) in the same way as in the full length protein. The absence of most parts of the polypeptide chain, including the second cluster, does not have an impact on structure and population of cluster I.
Figure 2.7. Chemical shifts of sidechains in denatured OmpX and in the heptadecamer peptide H–OmpX(70–87)–OH. (a) ($\omega_3(1^H)$, $\omega_1(1^H)$) strips for residues N74 and I79 from a 3D $^{15}$N-resolved [$^{1}H,^{1}H$]-TOCSY experiment of OmpX (left strips in the strip pairs) and corresponding ($\omega_2(1^H)$, $\omega_1(1^H)$) strips from a 2D [$^{1}H,^{1}H$]-TOCSY spectrum of the peptide OmpX(70–87) (right strips). (b) Comparison of [$\omega_1(1^H)$, $\omega_3(1^H)$] strips for two Ile residues from a 3D $^{15}$N-resolved-$[^{1}H,^{1}H$]-TOCSY spectrum of denatured OmpX with the Ile random coil values, $I_{rc}$. I40 is in an unstructured region of urea-denatured OmpX, and I79 is part of cluster I.

In summary, the data from two different experimental approaches presented in this chapter show that formation of the hydrophobic clusters is encoded entirely in the local amino acid sequence. We thus have here a clearcut illustration of a denatured protein with multiple hydrophobic clusters that are formed independently.
Binding of hydrophobic clusters to micelles

A standard in vitro refolding protocol for OmpX and other β-barrel membrane proteins is the rapid dilution of urea-denatured OmpX with solutions of DHPC or other detergents. As a step towards the elucidation of initial steps in membrane protein folding, the present chapter aims at a characterization of the ternary system OmpX–8 M urea–DHPC. First, the relevant micelle parameters of DHPC in 8 M urea were determined, i.e. the critical micelle concentration, cmc, and the average micelle size, expressed as the aggregation number \( n \) (Tanford, 1973; Helenius and Simons, 1975; Hauser, 2000). Literature values are only available for DHPC in H\(_2\)O, but not in 8 M urea solutions. The values can be determined from the translational diffusion constant, \( D_s \), which is a population-weighted average of the diffusion constants of DHPC monomers and micelles (Figure 2.8).

With NMR-based diffusion measurements, the cmc of DHPC in 8 M urea was found to be 45 ± 3 mM, which is about 2.8 times higher than the cmc in water (Table 2.1). An up to 4-fold increase of the cmc is a common feature of many ionic and non-ionic detergents in 8 M urea, due to the increased solubility of the hydrophobic moiety of the monomers (Caponetti et al., 1992; Abuin et al., 1997; Schick, 1964; Tanford, 1973).

![Figure 2.8](image.png)

**Figure 2.8.** Diffusion constant, \( D_s \), of DHPC in 8M aqueous urea solution measured with BPP-LED NMR diffusion experiments (Chou et al., 2004). The circles are experimental data points, the line is the result of a least-squares fit of the data. The transition at 45(±3) mM is indicative of micelle formation. The diffusion constant of DHPC micelles in 8 M urea at infinite dilution was determined to be \( D_0 = 62 \times 10^{-12} \text{ m}^2 \text{ s}^{-1} \) in this experiment.

The value for the diffusion constant of DHPC micelles in 8 M urea at infinite dilution was determined to be \( D_0 = 62 \times 10^{-12} \text{ m}^2 \text{ s}^{-1} \) from the measured diffusion constants above the cmc after correction for the monomer contributions (Chou et al., 2004). For spherically shaped micelles, an estimate of the micelle size is obtained with the Stokes-Einstein equation,
$D_0 = \frac{kT}{6\pi \eta R_H}$, \hspace{1cm} (2.2)

where $\eta$ is the viscosity, $R_H$ the particle radius, $k$ the Boltzmann constant, $T$ the temperature and $D_0$ the diffusion constant. The assumption of a spherical shape for the DHPC micelles seems reasonable, as even for oblate ellipsoids with an aspect ratio of 2, a shape that is expected for bicelles rather than micelles, the error in $D_0$ is only about 4% (Cantor and Schimmel, 1980; Vold and Prosser, 1996; Chou et al., 2004). With the viscosity of 1.66 of 8 M urea relative to H$_2$O the radius of DHPC micelles becomes $R_H(8\text{ M urea}) = 1.05\ R_H(H_2O)$. The micelle is of similar size under the two solution conditions and thus $n(8\text{ M urea}) = (1.05)^3\ n(H_2O) = 31$. For a conservative estimation of the error, the three reported values of $n$ for DHPC micelles in H$_2$O of 19, 27 and 35 were used (Table 2.1), resulting in $n = 31 \pm 9$ for DHPC in 8 M urea.

<table>
<thead>
<tr>
<th>Solution condition</th>
<th>H$_2$O$^a$</th>
<th>8M urea$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmc [mM]</td>
<td>15 ± 1 (Hauser, 2000)</td>
<td>45 ± 3</td>
</tr>
<tr>
<td></td>
<td>14 ± 1 (Chou et al., 2004)</td>
<td></td>
</tr>
<tr>
<td>aggregation number $n$</td>
<td>19 (Tausk et al., 1974)</td>
<td>31 ± 9$^c$</td>
</tr>
<tr>
<td></td>
<td>27 (Chou et al., 2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35 (Lin et al., 1986)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1. Micelle parameters of DHPC

$^a$ pure H$_2$O or H$_2$O with low salt concentrations, 25°C

$^b$ aqueous solution of 8 M urea, 20 mM sodium phosphate, pH 6.5, 15°C; this study

$^c$ error based on the uncertainty of the reference measurements in H$_2$O

With these micelle parameters at hand, the ternary system OmpX–8 M urea–DHPC was studied using high-resolution NMR experiments. A first experiment compared the resonance frequencies of the amide moieties in denatured OmpX in the presence and absence of 200 mM DHPC. All amide resonances markedly shifted in the presence of 200 mM DHPC correspond to residues in the two hydrophobic clusters, which comprise residues 73–82 and 137–145, whereas resonances from all other amide moieties show only small changes (Figure 2.9). This observation indicates a specific interaction between the two hydrophobic clusters and the detergent DHPC.

The concentration of 200 mM in this experiment is far above the cmc of DHPC in 8 M urea, and OmpX could interact with DHPC monomers, with preformed DHPC micelles or with
both species. This question was addressed by stepwise addition of DHPC to urea-denatured OmpX. Figure 2.10 shows a superposition of four 2D \[^{15}\text{N},^{1}\text{H}]-\text{HSQC} \) spectra of denatured OmpX with different DHPC concentrations. In this spectral region, the three amide resonances of residues 122 and 148, which are both located in an unstructured region, are unaffected by DHPC. The resonance of residue 142, which is located in the hydrophobic cluster II, shifts substantially.

Figure 2.9. Chemical shift differences \( \Delta\delta(\text{HN}) \) measured for OmpX denatured in 8 M urea with 200 mM DHPC and without DHPC. The chemical shift difference for the amide moiety, \( \Delta\delta(\text{HN}) \), was calculated as

\[
\Delta\delta(\text{HN}) = \sqrt{\left(\Delta\delta(\text{HN})\right)^2 + \left(0.2 \cdot \Delta\delta(\text{HN})\right)^2}
\]

The data for the two hydrophobic clusters of OmpX (residues 73–82 and 137–145) are indicated by open bars.

Figure 2.10. Displacement of \(^{15}\text{N}−^{1}\text{H} \) resonances of OmpX denatured in 8 M urea at different concentrations of DHPC; a superposition is shown of identical spectral regions of four 2D \[^{15}\text{N},^{1}\text{H}]-\text{HSQC} \) spectra measured with DHPC concentrations of 0, 70, 110 and 200 mM, respectively. The spectral region shown contains the resonances of the backbone NH moieties of residues A122, F148 (both not part of a hydrophobic cluster) and A142 (involved in hydrophobic cluster II).
The chemical shift differences of the amide moieties, $\Delta \delta$(HN), measured as a function of the DHPC concentration for all residues involved in the two hydrophobic clusters are shown in Figure 2.11(a) and 2.11(b). All $\Delta \delta$(HN) values for DHPC concentrations up to 35 mM are zero within experimental precision, except for Tyr80, which seems to have some interactions with DHPC monomers. At DHPC concentrations around 35 mM, the $\Delta \delta$(HN) values suddenly increase for all residues belonging to the hydrophobic clusters.

These data indicate an interaction of the entire hydrophobic clusters with DHPC micelles. The value of about 35 mM for the cmc of mixed DHPC–OmpX micelles is somewhat lower than the cmc for pure DHPC micelles (Table 2.1). This is plausible, considering the high molar concentrations of OmpX molecules compared to DHPC micelles, and the fact that each cluster contains several hydrophobic side chains. For the analysis of the titration data, the bi-molecular reaction
was chosen as kinetic model, where C stands for hydrophobic cluster, Mi for micelle and C · Mi denotes their complex. In this simple model, the DHPC micelle parameters are assumed to be unchanged upon binding to one of the hydrophobic clusters. For a given resonance, the combined chemical shift difference, ∆δ(HN), in dependence of the micelle concentration is then given by

\[ \Delta\delta(HN) = \Delta\delta_{\text{max}} \frac{[\text{Mi}]}{K_D + [\text{Mi}]} \]  

(2.4)

where \( \Delta\delta_{\text{max}} \) is the value of the chemical shift difference for a particular resonance in the cluster without micelles and with limiting high concentration of micelles, and \( K_D = [C] [\text{Mi}] / [C\cdot\text{Mi}] \) is the dissociation constant defined for the reaction (2.3). In a detergent solution above the cmc, the micelle concentration as a function of the detergent monomers is given by

\[ [\text{Mi}] = ([\text{Mo}] - \text{cmc}) / n, \]  

where Mo stands for monomers and \( n \) for the aggregation number. Equation (2.4) can thus be rewritten as

\[ \Delta\delta(HN) = \Delta\delta_{\text{max}} \frac{[\text{Mo}] - \text{cmc}}{K_D n + [\text{Mo}] - \text{cmc}} \]  

(2.5)

Equation (2.5) was used to obtain nonlinear fits of the titration data for DHPC with the free parameters cmc, \( \Delta\delta_{\text{max}} \) and \( K_D n \) for each residue, using least-squares minimization. For each residue, Eq. (2.5) yields a good fit of the experimental data (Figure 2.11(c)), but the values for cmc and \( K_D n \) resulting from the fits varied slightly for individual residues. The average cmc values obtained for residues 74–84 (cluster I) and 137–145 (cluster II) were 32.0 ± 5.4 mM, and 38.6 ± 2.4 mM, respectively. To reduce the number of free parameters in a second round of fitting, the cmc values were fixed to 32.0 mM and 38.6 mM for all residues of cluster I and II, respectively. This second nonlinear fitting resulted in an average \( K_D n \) of 115 ± 40 mM for the first cluster and 157 ± 23 mM for the second cluster.

With the assumption that the micelle size does not substantially change upon binding to OmpX, the dissociation constant \( K_D \) was evaluated with the aggregation number \( n = 31 \pm 9 \) for DHPC micelles in 8 M urea. This results in \( K_D \) values of 3.7 ± 1.7 mM and 5.1 ± 1.6 mM.
for the hydrophobic clusters I and II in urea-denatured OmpX, respectively. From the
dissociation constants, standard free energies for the transfer of a hydrophobic cluster from
the aqueous phase to the micellar lipid phase were calculated according to

\[
\Delta G^0 = -RT \ln\left([H_2O]/K_D\right)
\]  

(2.6)

where \( T \) is the temperature, \( R \) the molar gas constant, and \([H_2O]\) the molar concentration of
water (White and Wimley, 1999; Seelig, 2004). The resulting standard transfer free energies
from water to DHPC micelles are \( \Delta G^0 = -5.5 \pm 0.4 \text{ kcal/mol} \), and \( \Delta G^0 = -5.3 \pm 0.3 \text{ kcal/mol} \)
for cluster I and cluster II, respectively.

Additional data for micelle binding of denatured OmpX was obtained with fluorescence
spectroscopy experiments, using that each of the hydrophobic clusters contains one
tryptophan residue. Figure 2.12 shows the fluorescence spectrum of OmpX in 8 M urea
with and without 200 mM DHPC. The fluorescence intensity is slightly increased in the presence
of DHPC, and the emission maximum undergoes a blue shift of 5 nm, from 352 to 347 nm. A
blue shift is typical for a change to a more hydrophobic environment (Lakowicz, 1983), and is
thus consistent with insertion of the tryptophan side chains into the micelle. The steady-state
fluorescence anisotropy of the tryptophans is increased in the presence of DHPC micelles
(Figure 2.12(b)), which indicates a higher rotational correlation time and thus the transition of
the tryptophan into a more rigid position or into a larger particle (Lakowicz, 1983). The
fluorescence steady-state anisotropy data thus confirms the observation that the hydrophobic
clusters bind to DHPC micelles, which increases the effective size.
Structural characterization of denatured OmpX in complex with DHPC micelles

For proteins in denaturing solutions, the chemical shift values are population-weighted averages over the conformational ensemble. The observed chemical shift differences of nuclei in the hydrophobic clusters in the presence of micelles relative to 8 M urea solution therefore indicate differently populated polypeptide chain conformations under the two conditions (Wüthrich, 1994). A simple possibility behind these observations is a shifted equilibrium of the structures already present in 8 M urea and the random coil state (Tafer et al., 2004). However, in the presence of a lipid phase an intriguing alternative is the population of one or several new non-random conformations that are not significantly populated in the aqueous urea phase. To clarify this point, the present chapter attempts a structural characterization of the hydrophobic clusters in complex with DHPC micelles, using high-resolution NMR experiments.

Figure 2.13. $[\omega_1(\text{H}),\omega_3(\text{H})]$-strips for residues 72–76 from a 3D $^{15}$N-resolved-$[^1\text{H},^1\text{H}]$-NOESY spectrum (mixing time = 120 ms). The strips were taken at the backbone amino $^{15}$N chemical shifts of the residues indicated at the top of each strip, and centered about the corresponding $^1\text{H}$ chemical shifts. The horizontal dashed arrow indicates a long-range NOE between $^1\text{H}$ of W76 and $^1\text{H}$ of R72. Also indicated are the intraresidual $^1\text{H}$–$^1\text{H}$ cross peaks (marked by $^1\text{H}$) for all residues except I73, where the cross peak is outside the spectral range shown.
All side chain and backbone resonances of urea-denatured OmpX in 200 mM DHPC were assigned using standard triple-resonance experiments. For the determination of structural constraints, three high-resolution 3D \([^1H,^1H]\)-NOESY spectra were recorded, resolved for the heteronuclei \(^{15}\text{N}, ^{13}\text{C}\) (aliphatic) and \(^{13}\text{C}\) (aromatic), respectively. Figure 2.13 shows strips from the 3D \(^{15}\text{N}\)-resolved-\([^1H,^1H]\)-NOESY with an example of a \(d(i,i+4)\) NOE between \(\text{H}^\alpha\) of Arg72 and \(\text{H}^N\) of Trp76. From all three NOESY spectra, only NOEs between residues separated by at least two residues along the polypeptide chain were considered for the structure calculations. Other short-range NOEs may contain contributions from both random-coil and non-random conformations (Wüthrich, 1986; Smith et al., 1996; Tafer et al., 2004). The distribution of the two experimentally observed sets of NOEs for the two hydrophobic clusters bound to DHPC micelles is shown in Table 2.2, together with the NOEs of the hydrophobic clusters without micelles (Tafer et al., 2004).

<table>
<thead>
<tr>
<th>NOE Range</th>
<th>Cluster I‡</th>
<th>Cluster I bound to micelles</th>
<th>Cluster II‡</th>
<th>Cluster II bound to micelles</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d(i,i+3))</td>
<td>19</td>
<td>13</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>(d(i,i+4))</td>
<td>7</td>
<td>13</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>(d(i,i+5))</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>(d(i,i+6))</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>31</td>
<td>37</td>
<td>29</td>
<td>26</td>
</tr>
</tbody>
</table>

‡ same data as in Figure 2.1 (Tafer et al., 2004).

For each cluster, the number of NOEs is similar under the two conditions, but the individual NOEs are different. In cluster I, 5 of the 31 / 37 NOEs were observed in both environments, and for cluster II, this applies for 3 of the 26 / 29 total NOEs. This nearly complete change of the NOE patterns indicates that both hydrophobic clusters undergo structural changes when binding to micelles. For the calculation of structural models of the two clusters bound to micelles, the NOEs were transformed into upper limit distance constraints and structure calculations were performed with the program DYANA (see Materials and Methods). For both clusters, the resulting 20 conformations with the lowest target function form a well defined bundle (Figure 2.14). All experimental constraints are satisfied simultaneously in each of the corresponding 20 structures. In cluster I, the polypeptide backbone adopts a helical conformation, with the hydrophobic side chains exposed to the outside. The helical conformation is stabilized by a hydrogen bond between \(\text{H}^N\) of Val82 and the carbonyl O of Ile
79, which occurred in 14 of the 20 conformers. The backbone RMSD for residues 73–82 is 0.51 ± 0.19 Å. In cluster II, the backbone also adopts a helical conformation, which is, however, more extended than in cluster I, and the hydrophobic side chains are also exposed to the exterior. Two hydrogen bonds, Ala 142-H\textsuperscript{N} to Thr 139-CO and Gly 143-H\textsuperscript{N} to Trp 140-CO, appear in more than two thirds of the conformers. The RMSD for the backbone atoms of residues 137–145 is 0.40 ± 0.13 Å. The exposure of hydrophobic side chains to the outside and the compact conformation of the backbone are plausible conformations for hydrophobic clusters embedded in DHPC micelles.

![Figure 2.14. Results of a structure calculation for the experimentally determined constraints for cluster 73–82 (I) and cluster 137–145 (II) of OmpX denatured in 8M urea and 200 mM DHPC. (a) Bundle of 20 energy-minimized DYANA conformers superimposed for best fit to the mean coordinates. (b) same representation as in (a), including side chain heavy atoms. (c) Conformer with the lowest target function value in ball and stick representation; the backbone is shown in dark red, with the carbonyl oxygens in light red and the side chains in white (only heavy atoms shown).](image)

Further, intermolecular NOEs between protons in the hydrophobic clusters and protons of DHPC were observed. In total, 28 of 38 intermolecular NOEs involve protons in the hydrophobic tail of DHPC. In both hydrophobic clusters, the protons involved in the NOEs are located on the outside of the calculated structures, which is what one would expect from insertion of the clusters into micelles (Figure 2.15).
A comparison of the structures of the hydrophobic clusters in 8 M urea in the presence and absence of DHPC micelles is shown in Figure 2.16. For cluster I, both models present a compact fold. The difference between the two structures consists of a different sense of handedness of the backbone. For cluster II, the structures in the two environments are clearly different. In denatured OmpX in the absence of micelles, the tryptophan ring is buried in the interior of the cluster and shielded from the outside by other hydrophobic side chains, whereas in the more extended structure in the complex with DHPC micelles, the hydrophobic side chains are directed to the exterior.

![Chemical structure of DHPC and intermolecular NOEs](image)

**Figure 2.15.** Analysis of intermolecular NOEs between protons in DHPC and in the hydrophobic clusters. (a) Chemical structure of DHPC indicating the number (red) of NOEs to the different classes of DHPC protons. (b) and (c) Intermolecular NOEs mapped onto the structure of cluster I and cluster II, respectively; the protons showing NOEs are highlighted in red.

As discussed above for the interpretation of the experimental data of the residual structure of denatured OmpX in 8 M urea, it is not possible to distinguish a priori between two kinds of scenarios that could give rise to the observed set of NOEs. The signals could arise either from a single conformer with well defined structure, or from several lowly-populated conformers, or even from random close contacts between the side chains. As in the case of OmpX in 8 M urea, the experimental data is interpreted here in terms of a single conformer for each cluster. This is supported by two observations: (i) The calculated conformers fulfill the corresponding sets of medium- and longer-range NOEs simultaneously. (ii) The long-range NOE distance
constraints used as input for the structure calculations derive from the observation of negative NOEs, i.e., from $^1\text{H}$$^1\text{H}$ NOEs related to effective correlation times longer than one nanosecond, which would then also be a lower limit on the lifetime of the local structures defined by the NOEs. It seems unlikely that such long-lived local structures would prevail in the absence of co-operative effects including a sizeable number of simultaneous close interatomic contacts (Otting et al., 1991; Tafer et al., 2004).

The observation of two different self-consistent sets of constraints for the same polypeptide segment in two different environments raises the general question, whether or not any randomly generated set of upper limit distance constraints would yield a well-defined bundle of conformers in structure calculations. This issue was addressed with structure calculations using simulated sets of structural constraints. Sets with different numbers of $^1\text{H}$$^1\text{H}$ constraints, $n_s$, of 20, 25, 30, 35 and 40 were randomly generated, with a range distribution of 56% $d(i,i+3)$-constraints, 28% $d(i,i+4)$-constraints, 10% $d(i,i+5)$-constraints, and 6% $d(i,i+6)$-constraints. This distribution corresponds to the experimentally observed NOE data sets (Table 2.2). Following the calculation procedure from the experimental NOE data, the upper distance limit was set to 5.5 Å for all constraints. For each cluster and each value of $n_s$, 100 independent structure calculations from randomly generated constraint sets were performed,
each resulting in 500 conformers, from which the 20 conformers with the lowest residual DYANA target function values were analyzed with regard to the number of violated constraints, $n_v$, and the backbone RMSD (Figure 2.17).

![Figure 2.17](image)

**Figure 2.17.** Analysis of structure calculations for the amino acid sequences corresponding to the two hydrophobic clusters with simulated data. Each data point is the average of 100 independent simulations, the vertical bars represent the standard deviation. (a) and (b) Number of violated constraints $n_v$ in dependence of the number of used constraints $n_s$ for clusters I and II, respectively. A constraint was considered violated, if the upper distance limit of 5.5 Å between the pair of protons was violated on at least one of the 20 conformers. The dashed lines are the results of linear regressions. (c) and (d) RMSD of the bundle of 20 conformers with the lowest DYANA target function for clusters I and II, respectively.

A constraint was considered violated, if the upper distance limit of 5.5 Å between the corresponding protons was violated in one or more of the 20 DYANA conformers, which is the same criterion that was used for the experimental constraint sets. With this criterion $n_v = 0$ is obtained for the structure determinations with the experimental sets of constraints. The number of violated constraints, $n_v$, as a function of the number of used constraints $n_s$ is described by the equations $n_v = (n_s - 14.6)$ and $n_v = 0.8 \cdot (n_s - 13)$ for cluster I and II, respectively (dashed lines in Figure 2.17). The backbone RMSD for all analyzed artificial structures is similar and lies between 0.5 and 0.7 Å. It seems unlikely that the experimentally observed sets of about 30 NOEs, which do not lead to constraint violations, arise from random processes. The above made assumption that the self-consistent sets of experimental constraints arise from single, non-random conformer for each cluster is thus additionally supported.
Discussion

Could the interaction between micelles and hydrophobic clusters, as observed with the membrane protein OmpX be of relevance for membrane protein folding mechanisms? The interaction of micelles with the denatured soluble protein 434-repressor(1–63) shows that the affinity of hydrophobic clusters to micelles is a fundamental physicochemical property of the clusters regardless of their biological context. It seems thus likely that in any biochemical situation, in which micelles and hydrophobic clusters are present, the observed binding takes place. For example, such a situation occurs in refolding processes of β-barrel membrane proteins. Based on time-resolved fluorescence data, a model for the in vitro folding and insertion into bilayers for the membrane protein OmpA was proposed (Kleinschmidt and Tamm, 1996; Tamm et al., 2004). In a first stage, denatured OmpA binds to the bilayer via a water soluble collapsed intermediate, IM, and subsequently inserts and folds into the bilayer with a multistep mechanism. A similar mechanism was suggested for the refolding into preformed detergent micelles (Kleinschmidt et al., 1999). The kinetic studies on OmpA are not directly comparable with the current OmpX studies under equilibrium conditions in 8 M urea. However, it is tempting to speculate that the bilayer or micelle bound state IM shows similar characteristics as the system observed here and that possible hydrophobic clusters in OmpA are involved in the binding process. The different affinity to the hydrophobic phase presents a striking difference between polypeptide segments that form hydrophobic clusters under denaturing conditions and other parts of the polypeptide. Other scenarios can also be imagined, e.g. the binding of hydrophobic clusters to micelles constitute a misfolding pathway resulting in aggregation for some proteins. In such a situation, refolding conditions that inhibit the binding of hydrophobic clusters to micelles would increase folding yields. The data presented in this work provide a platform for further studies that may give more insight into a possible involvement of hydrophobic clusters in in vitro membrane protein folding mechanisms.
A rapid and sensitive method to detect hydrophobic clusters

The present chapter investigates the possibilities of exploiting the previously described specific binding between hydrophobic clusters of urea-denatured OmpX and DHPC micelles as a general method to detect hydrophobic clusters. To this end, we repeated the binding experiments with other polypeptides than OmpX.

The urea-denatured form of the protein 434-repressor(1–63) is known to have a hydrophobic cluster comprising residues 41–60, which was identified based on the analysis of NOEs and chemical shift deviations (Neri et al., 1992a; c). When DHPC is added to a solution of 434-repressor(1–63) in 7 M urea, several amide cross peaks in 2D $^{15}$N-$^1$H-HSQC spectra change their resonance position. The spectral region shown in Figure 2.18 contains amide resonances of five residues, of which only one, S50, is involved in the hydrophobic cluster. The chemical shifts of residues 25, 26, 27 and 39 stay at the same position in the two conditions, whereas residue 50 shows a large chemical shift difference between 0 and 200 mM DHPC.

![Figure 2.18](image)

**Figure 2.18.** Displacement of $^{15}$N-$^1$H resonances in of 2D $^{15}$N-$^1$H-HSQC spectra of 434-repressor(1–63) in 7 M urea without (left) DHPC and with 200 mM DHPC (right). The spectral region shown contains the resonances of the backbone NH moieties of residues G25, T26, T27 and T39 which are all not part of a hydrophobic cluster, and of S50, which is involved in a hydrophobic cluster. The intensity of the S50 resonance is also decreased in the presence of the micelles.

These chemical shifts result from changes of conformation and population distribution that are caused by the binding of the hydrophobic clusters to DHPC micelles. Micelle binding also increases the effective rotational correlation time for the amide moiety of S50, resulting in a decreased signal intensity for this resonance (Figure 2.18). The $^1$H, $^{15}$N and $^{13}$Cα chemical...
shift differences of the urea-denatured proteins OmpX and 434-repressor(1–63) between solutions with 200 mM DHPC and without DHPC are shown in Figure 2.19.

The chemical shift change caused by the addition of DHPC is mostly upfield for $^1$H$^N$ and $^{15}$N, and mostly downfield for $^{13}$C$^\alpha$. Resonances from residues outside of the hydrophobic clusters show much smaller chemical shift differences than residues involved in clusters. The discrimination between residues inside and outside the hydrophobic clusters is stronger for $^{15}$N and $^{13}$C$^\alpha$ than for $^1$H$^N$. The size of the chemical shift changes for $^1$H$^N$ and $^{15}$N caused by binding to DHPC micelles is larger for 434-repressor(1–63) than for OmpX by a factor of about 2. The reasons behind this observation remain unclear, a possible explanation could be that the hydrophobic clusters in denatured 434-repressors(1–63) are higher populated than the clusters in denatured OmpX (Neri et al., 1992a; Wüthrich, 1994; Tafer et al., 2004). This
could apply for the clusters in the denatured polypeptide chains alone or also in their complexes with DHPC micelles.

For the detection of hydrophobic clusters in denatured proteins, it is convenient to use the chemical shift differences for amide moieties, \( \Delta \delta(HN) = \sqrt{\left(\Delta \delta(1^N)\right)^2 + (0.2 \cdot \Delta \delta(1^N))^2} \), between solutions with 0 and 200 mM DHPC, as shown in Figure 2.20. Compared to the data from the individual nuclei in Figure 2.19, the combined chemical shift differences of individual amide groups, \( \Delta \delta(HN) \), permit a clearer identification of the hydrophobic clusters. In principle, the available \( ^{13}\text{C}^\alpha \) data could be included into the analysis to further increase the difference between the clusters and other regions (Grzesiek et al., 1996), but when using \( \Delta \delta(HN) \), \( ^{13}\text{C} \)-labeling is not required for the DHPC titration experiments.

\[ \Delta \delta(HN) = \sqrt{\left(\Delta \delta(1^N)\right)^2 + (0.2 \cdot \Delta \delta(1^N))^2} \]

If one assumes that polypeptide segments form a hydrophobic cluster whenever \( \Delta \delta(HN) > 0.05 \) ppm for at least 3 consecutive residues, then one finds for OmpX, that there are two hydrophobic clusters involving residues 75–84 and 137–145. These ranges match well with the earlier analysis based on NOEs, which found two clusters for the segments 73–82 and 137–145. For 434-repressor(1–63), the “triplet” \( \Delta \delta(HN) \) values indicate one large hydrophobic cluster comprising residues 43–61. Previously, residues 54–60 have been identified as a hydrophobic cluster based on NOEs (Neri et al., 1992a), and residues 41–53 were identified as belonging to a hydrophobic cluster based on chemical shift data (Neri et al., 1992c). Thus, for 434-repressor(1–63) the proposed approach again identifies the hydrophobic cluster region correctly.

Wild type OmpX contains two tryptophan residues, Trp76 and Trp140, one in each hydrophobic cluster. As shown in a previous chapter, a single-amino acid exchange of one
tryptophan to alanine leads to the destruction of the corresponding cluster but does not affect the other cluster, so that the variant OmpX[W76A] lacks the cluster 73–82, and the variant OmpX[W140A] lacks the cluster 137–145. The sensitivity of the method for identifying hydrophobic clusters using $\Delta \delta$(HN) values was further explored with these two OmpX variants. As shown in Figure 2.21, the destruction of one hydrophobic cluster by the corresponding amino acid variation prevents its micelle binding and the concomitant effect on the $\Delta \delta$(HN) values.

![Figure 2.21. Combined chemical shift differences of the amide moiety in 2D $^{15}$N-$^1$H-HSQC spectra for urea-denatured OmpX variants. (a) Data for OmpX[W76A] (b) Data for OmpX[W140A]. The results for the residues involved in the hydrophobic clusters of OmpX (residues 73–82 and 137–145) are indicated by open bars.](image)

On the basis of the OmpX data alone, two alternative explanations for these findings are imaginable. The binding specificity of DHPC micelles could be either for hydrophobic clusters or for certain amino acid types, in particular for tryptophan residues. However, specificity of the binding for tryptophan residues would be in contrast to the data from 434-repressor(1–63), where micelle binding occurs for residues 43–60, up to 15 residues away from the only tryptophan in this protein, Trp 58 (Fig. 2.22). Specificity to a further amino acid type would also fail to explain the micelle binding of this long segment, as all other amino acids can be found in segments that do not interact with micelles. An example is phenylalanine 44, which would be a likely candidate due to its aromaticity. In the sequence of OmpX, 7 phenylalanine residues are contained and none of them binds to DHPC micelles (Fig. 2.22). Thus, the most likely reason for micelle binding to the polypeptide segment 43–51 of denatured 434-repressor(1–63) is the presence of a hydrophobic cluster formed by the augmentedly occurring hydrophobic residues other than tryptophan. This demonstrates that DHPC micelle binding is sensitive to the presence or absence of hydrophobic clusters, and not to similarity of closely related amino acid sequences. There is a strong correlation between the
presence of tryptophan residues in a polypeptide segment and the occurrence of clusters in the same region (Neri et al., 1992a; Dyson et al., 1992; Ragona et al., 1997; Crowhurst et al., 2002; Tafer et al., 2004; Wirmer et al., 2004), and given that micelles bind to hydrophobic clusters in general, the same correlation can also be expected between the presence of tryptophan residues and micelle binding to the corresponding polypeptide segment.

Besides these implications for the detection of hydrophobic clusters, the data shown in Figures 2.20 and 2.21 yields further support to the conclusion obtained previously concerning the absence of a long-range interaction between the two hydrophobic clusters in denatured OmpX. By an analysis of chemical shift values in the single amino acid variants of OmpX and from NMR measurements with a peptide analog of cluster I, it was found that the two hydrophobic clusters are forming independent of each other and do not exhibit a long-range interaction. As the chemical shift changes caused by micelle binding are identical in wild type OmpX and OmpX[W140A] for the first cluster, and in wild type OmpX and OmpX[W76A] for the second cluster (Fig. 2.20 and 2.21), it can be concluded that the features of each cluster to bind micelles do not depend on the second cluster. On the one hand, this is an additional piece of support for the absence of any cluster–cluster interactions in the absence of micelles. On the other hand, this observation implies that a single DHPC micelle binds at most one hydrophobic cluster. OmpX–micelle complexes, where both hydrophobic clusters are bound at the same time to a single DHPC micelle, are not significantly populated in the ensemble.

Generalizing the results obtained with wild type OmpX, the two OmpX variants and 434-repressor(1–63), it would appear that hydrophobic clusters in unfolded proteins can be detected by the measurement of combined chemical shift differences, $\Delta\delta$(HN), in solutions
containing 0 mM and 200 mM DHPC, leading to identification of polypeptide segments that show large Δδ(HN)-values. This analysis requires resonance assignments in 8 M urea with and without addition of DHPC. If assignments in the absence of DHPC are available, they can usually be carried over to the sample with 200 mM DHPC by a stepwise titration with DHPC, since typically only a few resonances move significantly. These assignments can then be checked by high-resolution 3D $^{15}$N-resolved [$^1$H,$^1$H]-NOESY spectra (Talluri and Wagner, 1996) measured with and without DHPC, and a comparison of the peak patterns in the individual NOESY strips of corresponding amide resonances. For OmpX both approaches were tested, and both resulted in unambiguous identical assignments. In exceptional cases, $^{13}$C, $^{15}$N-isotope labeling may be necessary for the assignment of resonances after adding DHPC.

To expand the range of applications, the interaction between micelles and hydrophobic clusters was characterized in the case of wild type OmpX under different solution conditions. Starting from the reference conditions in 8 M urea at pH 6.5 and 15°C, the addition of 200 mM DHPC resulted in similar Δδ(HN)-plots in the pH range from 4.5 to 8.0 at 15°C and in the temperature range from 5°C–40°C at pH 6.5. Further, the following detergents were found to bind the two hydrophobic clusters at the reference conditions in a similar way as 200 mM DHPC: 200 mM 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DC$_7$PC), 270 mM n-Octyl-β-D-glucoside (OG) and 300 mM n-dodecyl-β-D-maltoside (DM). With DM and OG, however, the Δδ(HN) values of the first cluster were less pronounced compared to the random coil values and it is therefore recommended to use DHPC or DC$_7$PC, which both resulted in virtually identical Δδ(HN)-plots. Since the presence of denaturants increases the critical micelle concentration of most detergents, it is important to use detergent concentrations high above the aqueous cmc (Tanford, 1973). For example, the cmc of DHPC is 15 mM in H$_2$O compared to 45 mM in 8 M urea.

In conclusion, the chemical shift changes caused by binding of hydrophobic clusters to detergent micelles provide a sensitive and simple method for the detection of hydrophobic clusters in urea-denatured proteins. The two hydrophobic clusters in urea-denatured OmpX are not distinguishable from other regions of the polypeptide by measurement of relaxation data or scalar coupling constants. Only the complete, time-consuming analysis of 3D NOESY spectra and chemical shift data had previously allowed the identification of the two clusters (Tafer et al., 2004). The chemical shift changes caused by micelle binding provide a sensitive and efficient alternative to conventional techniques.
Materials and Methods

Production of OmpX and NMR Sample Preparation. OmpX was expressed in *E. coli* as reported (Fernández et al., 2001a). After extraction from the inclusion bodies the protein was further purified by anion-exchange chromatography. A HiTrap Q sepharose HP column with a volume of 5 ml used on an Äkta Prime chromatography system (Amersham Pharmacia Biotech) was equilibrated with buffer A (8 M urea, 20 mM Tris-HCl, pH 8.5). The protein was loaded onto the column, washed with buffer A, and eluted with a 60 ml gradient of 0–100% of buffer B (8 M urea, 20 mM Tris-HCl, 1 M NaCl, pH 8.5) at a flow rate of 1 ml/min. The fractions containing OmpX were transferred to a Centricron ultrafiltration device with 3 kDa molecular mass cutoff (Millipore), and the buffer was exchanged against freshly prepared NMR buffer (8 M urea, 20 mM phosphate, 0.1 mM NaN₃, 5% D₂O, 95% H₂O, pH 6.5) by successive dilution/concentration steps. Buffers were freshly prepared for each experiment with ultrapure urea from USB (Cleveland, Ohio). The NMR sample for the identification of the hydrophobic clusters was [U-¹³C,¹⁵N]-labeled with a protein concentration of about 3 mM. For the NMR studies in complexes with micelles, a [U-¹³C,¹⁵N]-labeled sample with a protein concentration of 1 mM was used. For all other experiments, [U-¹⁵N]-labeled samples of 0.5 mM concentration were used. [U-¹⁵N,¹³C]-labeled 434-repressor(1–63) was expressed and purified as described (Neri et al., 1989; Neri et al., 1992b). For the NMR studies, 434-repressor(1–63) was dissolved in a buffer at pH 4.8 containing 7 M urea, 20 mM phosphate, 0.1 mM NaN₃, 5% D₂O, 95% H₂O, which are the reference conditions of previous studies for the identification of hydrophobic clusters (Neri et al., 1992a).

Calculation of Three-dimensional Structures. An upper limit distance constraint of 5.5 Å was applied for all medium-range and long-range NOEs (Wüthrich, 1986; Neri et al., 1992a). With these constraints as input, structure calculations were performed independently for the two polypeptide segments 73 – 82 and 137 – 145, using the program DYANA (Güntert et al., 1997). To limit chain-end effects, two extra residues were included at the N- and C-termini of these polypeptide segments. DYANA calculations started from 500 conformers with random torsion angle values, and 8000 torsion angle dynamics steps were performed using the standard simulated annealing procedure. Finally, the 20 conformers with lowest final target function values were subjected to energy minimization with the program OPALp (Luginbühl et al., 1996; Koradi et al., 2000) and used to represent the NMR solution structures.

Calculations with simulated sets of NOEs. For these calculations, sets of constraints were randomly generated that consisted to 56% of constraints between residue *i* and residue *i*+3,
28% between \(i\) and \(i+4\), 10% between \(i\) and \(i+5\), 6% between \(i\) and \(i+6\). This distribution corresponds to the experimentally observed NOE data (Table 2.1). NOEs between backbone and side chain protons within these boundary conditions were randomly chosen and had equal probabilities to be created. The calculations were carried out with the same protocol as for the experimental constraints, except that the energy minimization step was omitted. For all sets of constraints, the RMSD for the backbone and its standard deviation and the number of violated constraints for the bundle of 20 DYANA conformers were calculated.

**Amino acid exchange.** Site-directed mutagenesis was done using a QuickChange mutagenesis kit (Stratagene, La Jolla, California, USA). The single amino acid variants OmpX[W76A] and OmpX[W140A] were obtained with a single reaction from a pET3b plasmid containing the OmpX wild type gene. Before further use, the correct DNA sequence of the variant genes was confirmed.

**Peptide studies.** A peptide with the sequence H–AYRINDWASIYGVGVGY–OH, corresponding to OmpX(70–87), was synthesized by Bio-Synthesis (Lewisville, Texas, USA) with a purity > 95% as confirmed by HPLC. For the NMR studies, lyophilized peptide was dissolved in freshly prepared NMR buffer (8 M urea, 20 mM phosphate, 0.1 mM Na\(\text{N}\), 5% D\(\text{2}\)O, 95 H\(\text{2}\)O\%, pH 6.5) to a concentration of 1 mM and used without further purification. A peptide with the sequence H–SVDVGTWIAGVGYRF–OH corresponding to OmpX(134–148), was not soluble at concentrations higher than 0.01 mM in the same buffer.

**NMR spectroscopy.** All experiments were carried out at 15°C. The experimental parameters for the experiments that lead to sequence-specific resonance assignments and structures for denatured wild-type OmpX can be found in Tafer et al. (2004). All NMR experiments required for the sequence-specific backbone and side chain resonance assignments and the structure determination of denatured wild-type OmpX in complex with DHPC micelles were measured at 750 MHz. The following experiments were recorded: 2D \([^{15}\text{N},^{1}\text{H}]\)-HSQC (Bodenhausen and Ruben, 1980), 2D \(ct-[^{13}\text{C},^{1}\text{H}]\)-HSQC (Vuister and Bax, 1992), 3D \(ct\)-HNCA (Yamazaki et al., 1994), 3D CBCA(CO)NH (Grzesiek and Bax, 1992b), 3D H(CCO)NH-TOCSY (Grzesiek et al., 1993a), 3D (H)C(CO)NH-TOCSY (Grzesiek et al., 1993a), 3D \(^{15}\text{N}\)-resolved-[\(^{1}\text{H},^{1}\text{H}\)]-NOESY (Talluri and Wagner, 1996), two 3D \(^{13}\text{C}\)-resolved-[\(^{1}\text{H},^{1}\text{H}\)]-NOESYs for aliphatic and aromatic protons, respectively (Muhandiram et al., 1993). 2D \([^{15}\text{N},^{1}\text{H}]\)-HSQC spectra (Bodenhausen and Ruben, 1980) with a WATERGATE sequence for water suppression (Sklenar et al., 1993; Wider, 1998) were recorded on a Bruker DRX 750 spectrometer with an interscan delay of 1s. 2k complex points were recorded with an acquisition time of 194 ms, and prior to Fourier transformation the FID was multiplied with a
75°-shifted sine bell and zero-filled to 4k complex points. In the $^{15}$N dimension, 256 complex points were measured, with a maximal evolution time of 112 ms, and the data was multiplied with a 75°-shifted sine bell and zero-filled to 512 complex points before Fourier transformation. The baseline was corrected using the IFLAT method (Bartels et al., 1995a) in the $\omega_2(^1H)$-dimension, and polynomials of 5th order in the $\omega_1(^{15}N)$-dimension.

With the peptide OmpX(70–87), a 2D [$^1H$,$^1H$]-TOCSY experiment (Braunschweiler and Ernst, 1983) and a 2D [$^1H$,$^1H$]-NOESY experiment (Kumar et al., 1980) with a mixing time of $\tau_m = 120$ ms were recorded. For both experiments, the following parameters were used: The interscan delay was 1s; 4k complex points were recorded with an acquisition time of 388 ms, and prior to Fourier transformation the FID was multiplied with a 75°-shifted sine bell and zero-filled to 8k complex points. In the $\omega_1(^1H)$-dimension, 512 complex points were measured, with a maximal evolution time of 195 ms, and the data was multiplied with a 75°-shifted sine bell and zero-filled to 1024 complex points before Fourier transformation. The baseline was corrected using the IFLAT method (Bartels et al., 1995a) in the $\omega_2(^1H)$-dimension and polynomials of 5th order in the $\omega_1(^1H)$-dimension.

**Data analysis.** The proton chemical shifts were referenced to internal DSS (Markley et al., 1998), and those for nitrogen-15 and carbon-13 were indirectly referenced (Wishart et al., 1995b; Wider, 1998). All spectra were processed with the software PROSA (Güntert et al., 1992) and analyzed using the software XEASY (Bartels et al., 1995b) and CARA (www.nmr.ch). All analyses were based on sequence-specific resonance assignments for wild-type OmpX obtained with standard triple resonance experiments (Tafer et al., 2004). Sequence-specific resonance assignments for the amide moieties of OmpX[W76A] and OmpX[W140A] were derived from the wild-type assignments with 3D $^{15}$N-resolved [$^1H$,$^1H$]-NOESY experiments (Talluri and Wagner, 1996), $\tau_m = 120$ ms. Unambiguous sequence-specific resonance assignments could be obtained, since the chemical shifts were found to be unchanged for the vast majority of the residues. The titrations from 0 mM to 200 mM DHPC were done in steps of 20–40 mM, and the sequence-specific resonance assignment of the amide moieties was carried over during the stepwise titration from one spectrum to the next one. The assignment was independently verified with a 3D $\alpha$-HNCA spectrum (Yamazaki et al., 1994) for 434-repressor(1–63), and with a 3D $^{15}$N-resolved [$^1H$,$^1H$]-NOESY spectrum (Talluri and Wagner, 1996) in the case of OmpX. Alternatively for OmpX, the assignment of the spectrum at 200 mM DHPC was obtained directly from the assignment at 0 mM DHPC and analysis of the 3D $^{15}$N-resolved [$^1H$,$^1H$]-NOESY spectrum.
The overall chemical shift difference of the amide moiety, $\Delta \delta(HN)$, was defined as $\Delta \delta(HN) = \sqrt{(\Delta \delta(^1H_N))^2 + (0.2 \cdot \Delta \delta(^15N))^2}$, where $\Delta \delta(^1H_N)$ and $\Delta \delta(^15N)$ are the chemical shift differences for $^1H_N$ and $^15N$, respectively, between an urea-denatured protein at 0 mM DHPC and at 200 mM DHPC ($\Delta \delta = \delta(200 \text{ mM DHPC}) - \delta(0 \text{ mM DHPC})$). The weighting factor of 0.2 reflects the naturally occurring dispersion of chemical shift values in denatured proteins and is in line with studies on folded proteins (Pellecchia et al., 1999; Grzesiek et al., 1996).

**Fluorescence measurements.** Steady state fluorescence measurements were made at 15°C on a PTI Quantamaster QM-7/2003 spectrofluorometer in a T-format configuration. The excitation wavelength was 290 nm, the emission spectrum was measured between 300 nm and 400 nm with a scan speed of 1 nm/s. All slits were opened to a wavelength range of 2 nm. The signal was averaged over 24 scans. The baseline was corrected with spectra of the corresponding buffers without protein. Anisotropy was calculated according to $r = \frac{I_{\text{parallel}} - I_{\text{perpendicular}}}{I_{\text{parallel}} + 2I_{\text{perpendicular}}}$ and a G-factor correction was applied (Lakowicz, 1983). The sample conditions for fluorescence measurements were identical to those of the NMR experiments, except that no D$_2$O and NaN$_3$ were used, and that the protein concentration was 10 µM.

**Diffusion measurements.** Measurements of the diffusion constants of DHPC in 8M urea were done using 1D BPP-LED experiments with varying field gradient strength (Chou et al., 2004). The data was fitted to the equation $\ln \left[ \frac{I(f)}{I(f_0)} \right] = - (\gamma \cdot \delta \cdot G_{\text{max}}) \cdot (f^2 - f_0^2) \cdot (\Delta + 2/3 \cdot \delta + 3/2 \cdot \tau) \cdot D_{\text{s}}$, where $I$ is the intensity (integral) of the NMR signal, $f$ is the fractional gradient strength, $f_0$ is the fractional gradient strength of the reference spectrum (0.02), $\gamma$ is the gyromagnetic ratio of $^1H$, $G_{\text{max}}$ is the maximal gradient strength at $f = 1.0$, $\Delta$ is the diffusion time, $\delta$ is the gradient length, $\tau$ is gradient recovery delay and $D_{\text{s}}$ is the diffusion constant. The gradient strengths were calibrated on the residual $^1H$ signal in a 99.8% D$_2$O sample, using the published value of $1.902 \times 10^{-9}$ m$^2$ s$^{-1}$ for the self-diffusion coefficient of HDO at 25°C (Mills, 1973; Chou et al., 2004).

**Titrations with detergents.** For the DHPC titrations of OmpX and its variants, samples with protein concentrations of 0.5 mM were used. For 434-repressor(1–63), a 1 mM sample was used. For both proteins, 50 mM DHPC stock solutions in the corresponding NMR buffers were used for titrations up to 30 mM DHPC, for further steps, solid DHPC was added to the sample. The detergents DHPC, DC$_7$PC, OG and DM were purchased from Avanti Polar Lipids (Alabaster, AL) and used without further purification. Detergent concentrations were monitored with 1D $^1H$-NMR experiments.
Conclusion

Automated projection spectroscopy is a method with many interesting application possibilities in biomolecular NMR. Automated sequence-specific backbone assignment is one of them, with the goal of establishing standard setups, including pulse sequences and all necessary parameters, to reliably provide backbone assignments for regular small and medium-sized proteins. The theory of sensitivity and the classification of strategies presented in this thesis are an important prerequisite to this goal and the presented applications show that automated assignment of protein backbone resonances with APSY in combination with a suitable algorithm is a realistic endeavor. The next steps would be to extend the size limit of the method and to gain more experience with a series of typical proteins. A current estimate for the size limit is 15 kDa for routine use and 20 kDa in preferable cases. Currently, the main obstacles on this way seem to be the purity of samples and the existence of multiple protein species, such as caused by conformational or chemical exchange. Some effort will have to be put into matching the quality of the GAPRO output to the requirements of the assignment algorithm. Based on the experience gained in this work, the algorithm GARANT should probably be replaced by another algorithm tailored to the high quality of the APSY peak lists.

At the time when this thesis was started, the backbone of denatured OmpX was assigned using conventional techniques, requiring more than a week of spectrometer time and a substantial amount of human work. With the APSY-based approach, the assignment can be done in 1–2 days and without any human interaction.

The characterization of denatured OmpX, described in the second part of the thesis, have brought insights into the properties of this denatured membrane protein. An important result is the clear demonstration of the independent formation of the two hydrophobic clusters with the implication that the properties of the denatured polypeptide are encoded entirely in the local amino acid sequence. It would be very interesting to generalize this finding to other unstructured polypeptides. Another interesting question is whether the cooperative effects of hydrophobic cluster formation comprise the whole cluster. It might be possible to reduce a hydrophobic cluster in size without destroying it by exchanges of amino acids at the edge of the cluster-forming polypeptide segments. The identified interactions between the clusters and DHPC micelles are an intriguing feature of hydrophobic clusters with possible relevance for membrane protein folding. Overall, the understanding of the denatured state of the membrane protein OmpX achieved in this work forms an ideal platform for further folding studies.
Appendix

Nomenclature

HN\textsuperscript{i} amide proton of residue \textit{i}
HN\textsuperscript{i−1} amide proton of residue \textit{i−1}
HN\textsuperscript{i+1} amide proton of residue \textit{i+1}

HA\textsuperscript{i} alpha-proton of residue \textit{i}
HA\textsuperscript{i−1} alpha-proton of residue \textit{i−1}
HA\textsuperscript{i+1} alpha-proton of residue \textit{i+1}

HB\textsuperscript{i} beta-proton of residue \textit{i}
HB\textsuperscript{i−1} beta-proton of residue \textit{i−1}
HB\textsuperscript{i+1} beta-proton of residue \textit{i+1}

CO\textsuperscript{i} carbonyl of residue \textit{i}
CO\textsuperscript{i−1} carbonyl of residue \textit{i−1}
CO\textsuperscript{i+1} carbonyl of residue \textit{i+1}

CA\textsuperscript{i} alpha-carbon of residue \textit{i}
CA\textsuperscript{i−1} alpha-carbon of residue \textit{i−1}
CA\textsuperscript{i+1} alpha-carbon of residue \textit{i+1}

CB\textsuperscript{i} beta-carbon of residue \textit{i}
CB\textsuperscript{i−1} beta-carbon of residue \textit{i−1}
CB\textsuperscript{i+1} beta-carbon of residue \textit{i+1}

N\textsuperscript{i} amide nitrogen of residue \textit{i}
N\textsuperscript{i−1} amide nitrogen of residue \textit{i−1}
N\textsuperscript{i+1} amide nitrogen of residue \textit{i+1}

N\textsuperscript{i}CA\textsuperscript{i−1}CO\textsuperscript{i−1} example for a product operator

\textsuperscript{1}J_{HN,N} 1-bond scalar coupling between amide proton and amide nitrogen
\textsuperscript{1}J_{HA,CA} 1-bond scalar coupling between alpha-proton and alpha-carbon
\textsuperscript{1}J_{HB,CB} 1-bond scalar coupling between beta-proton and beta-carbon
\textsuperscript{1}J_{N,CA} 1-bond scalar coupling between amide nitrogen and alpha-carbon
\textsuperscript{2}J_{N,CA} 2-bond scalar coupling between amide nitrogen and alpha-carbon
\textsuperscript{1}J_{N,CO} 1-bond scalar coupling between amide nitrogen and carbonyl
\textsuperscript{1}J_{CA,CB} 1-bond scalar coupling between alpha-carbon and beta-carbon

R\textsubscript{2}(HN) transverse relaxation rate of amide proton
R\textsubscript{2}(HA) transverse relaxation rate of alpha-proton
R\textsubscript{2}(HB) transverse relaxation rate of beta-proton
R\textsubscript{2}(CO) transverse relaxation rate of carbonyl
R\textsubscript{2}(CA) transverse relaxation rate of alpha-carbon
R\textsubscript{2}(CB) transverse relaxation rate of beta-carbon
R\textsubscript{2}(N) transverse relaxation rate of amide nitrogen
A  nucleus of detection
m  an APSY / projection-NMR experiment
\vec{\phi}  vector of projection angles \alpha, \beta, \gamma, ...
\vec{p}_1(\vec{\phi})  projection vector 1
\vec{p}_2  projection vector 2
t  time parameter for the indirect dimension
t_N  time parameter for the direct dimension
s_m(\vec{\phi},t,t_N)  complex 2D time domain signal of experiment \text{m} at angles \vec{\phi}
s_m(0)  signal amplitude of experiment \text{m} at \text{t} = 0 \text{ and } t_N = 0.
s_m^i(\vec{\phi},t)  envelope function of the indirect dimension of experiment \text{m}, recorded with angles \vec{\phi}
i  complex number \sqrt{-1}
\Omega(\vec{\phi})  resonance frequency with angles \vec{\phi}
s_A^i(t_N)  signal envelope function in the direct dimension for nucleus A
\Omega_A  resonance frequency of nucleus A
h_m(\vec{\phi},t,t_N)  2D window function
S_m(\vec{\phi})  Peak height at maximum in experiment \text{m} at angles \vec{\phi}
t_{\text{max}}(\vec{\phi})  maximal evolution time in the indirect dimension, chosen by the operator for angles \vec{\phi}
t_{\text{max,A}}  maximal evolution time in the direct dimension for nucleus A
M(\vec{\phi})  number of complex data points in the indirect dimension, chosen by the operator for angles \vec{\phi}
M_A  number of complex data points in the direct dimension
n(\vec{\phi})  number of scans, chosen by the operator for angles \vec{\phi}
h_m(\vec{\phi},t)  window / weighting function for the indirect dimension, chosen by the operator for angles \vec{\phi}
h_A(t_N)  window / weighting function for the direct direction
\bar{h}_m^i(\vec{\phi})  thermal noise weighting factor at angle \vec{\phi}
\bar{h}_A^i  thermal noise weighting factor for nucleus A
\sigma(\vec{\phi})  noise RMSD of the projection at angles \vec{\phi}
q  number of projection angles that are not 0° or 90°.
\rho  square root of the frequency-independent power spectral density
f_m(\vec{\phi})  angle-dependent modulation of the sensitivity in experiment \text{m}
K_A  sensitivity factor of nucleus A
\vec{R}_{2,m}  transverse relaxation vector of experiment \text{m}
R_{2,m}(\vec{\phi})  effective transverse relaxation rate of experiment \text{m} at angles \vec{\phi}
T(\text{m})_i  transfer function of element \text{i} of experiment \text{m}
P(\text{X})  steady-state polarization of nucleus \text{X}
Calculation of envelope functions for time domain cross-sections

In multidimensional NMR experiments, the relaxation along the individual dimension axes is often a monoexponential decay with the transverse relaxation rate \( R_2(X) \) of the nucleus or coherence \( X \) evolving in the particular dimension. A constant time evolution period without relaxation is described with \( R_2(X) = 0 \). For each experiment \( m \) with monoexponential relaxation in all dimensions, an \( N \)-dimensional vector \( \vec{R}_{2,m} \) can be defined, that contains the relaxation rates for the \( N \) dimensions.

The present paragraph aims at the calculation of the envelope function for a time domain cross-section at angle \( \phi \). The situation is illustrated in Figure A.1 for two indirect dimensions \( x \) and \( y \) with transverse relaxation rates relaxation rates \( R_{2,x} \) and \( R_{2,y} \). The signal envelopes along the axes \( t_x \) and \( t_y \) is given by \( \exp(-R_{2,x} t_x) \) and \( \exp(-R_{2,y} t_y) \), respectively. A projection with projection angle \( \tilde{\phi} = \alpha \) has the envelope \( s^*_m(\tilde{\phi}, t) = \exp\{-R_{2,x} t_x\}\exp\{-R_{2,y} t_y\} = \exp\{-\sin\alpha R_{2,x} t\}\exp\{-\cos\alpha R_{2,y} t\} \). This is a monoexponential decay \( s^*_m(\tilde{\phi}, t) = \exp\{- R^*_{2,m}(\tilde{\phi}) t\} \) along the time parameter \( t \) with an effective relaxation rate \( R^*_{2,m}(\tilde{\phi}) = \sin \alpha R_{2,x} + \cos \alpha R_{2,y} \).

Generalized to \( N \) dimensions, \( R^*_{2,m}(\tilde{\phi}) \) is given by

\[
R^*_{2,m}(\tilde{\phi}) = \vec{p}_1(\tilde{\phi}) \cdot \vec{R}_{2,m}
\]

(A.1)

Figure A.1. 3D-representation of the NMR signal envelope \( s^\circ \) (green surface) for two indirect dimensions \( x \) and \( y \). The pink and the red line are exponential relaxation decays along \( t_x \) and \( t_y \), respectively. The yellow line is the signal envelope for a time domain cross-section at angle \( \alpha \), parameterized with \( t \).
A library of transfer elements for APSY backbone experiments
The following library of transfer elements uses standard notations. It seems not feasible to identify for each element the original references, where it was used first. Representative references are the first INEPT step (Morris and Freeman, 1979), a popular textbook (Cavanagh et al., 1996) and three common reviews (Bax and Grzesiek, 1993; Kay, 1995; Sattler et al., 1999). For some of the transfer functions, the amino acid side chain topology is important (Fig. A.2).

**Figure A.2.** Spin systems for the 20 common amino acids. In addition to the nuclei of interest (Fig. 1.10), the gamma-carbon CG is of relevance for all experiments on CB.

**Element 1.**

\[
\begin{align*}
HN_z^{i'} & \rightarrow HN_z^{i'}N_z^{i'} \\
HN_z^{i'}N_z^{i'} & \rightarrow HN_z^{i'}
\end{align*}
\]

relevant constants: \(J_{HN,N}^{i}, R_2(HN)\)

transfer amplitude: \(\sin(\pi J_{HN,N}^{i} \xi) \exp(-\xi R_2(HN))\)

maximum at \(\xi = \tan^{-1}(\pi J_{HN,N}^{i} / R_2(HN)) / \pi J_{HN,N}^{i}\)

value without relaxation \(\xi = 5.4\) ms

typical value \(\xi = 5.4\) ms

**Element 2.**

\[
\begin{align*}
HN_z^{i'}N_z^{i'} & \rightarrow N_z^{i'}CO_z^{i-1}
\end{align*}
\]

relevant constants: \(J_{HN,N}^{i}, J_{N,CO}^{i}, R_2(N)\)
transfer amplitude: $\sin(\pi \, J_{N,CO} \tau) \exp(-\tau \, R_2(N))$

maximum at $\tau = \tan^{-1}(\pi \, J_{N,CO} / R_2(N)) / \pi \, J_{N,CO}$

value without relaxation $\tau = 36$ ms

typical value $\tau = 27$ ms

$\xi' = 5.4$ ms fixed

Element 3.

$N_z^i CO_z^{i-1} \rightarrow HN_z^i N_z^i$

relevant constants: $J_{HN,N}, J_{N,CO}, R_2(N)$

transfer amplitude: $\sin(\pi \, J_{N,CO} \tau) \exp(-\tau \, R_2(N))$

maximum at $\tau = \tan^{-1}(\pi \, J_{N,CO} / R_2(N)) / \pi \, J_{N,CO}$

value without relaxation $\tau = 36$ ms

typical value $\tau = 27$ ms

$\xi' = 5.4$ ms fixed

Element 4.

$HN_z^i N_z^i \rightarrow HN_z^i$

relevant constants: $J_{HN,N}, R_2(HN)$

transfer amplitude: $\sin(\pi \, J_{HN,N} \xi) \exp(-\xi \, R_2(HN))$

maximum at $\xi = \tan^{-1}(\pi \, J_{HN,N} / R_2(HN)) / \pi \, J_{HN,N}$

value without relaxation $\xi = 5.4$ ms

typical value $\xi = 5.4$ ms

Element 4b.

$HN_z^i N_z^i \rightarrow HN_z^i$

relevant constants: $J_{HN,N}, R_2(HN)$

transfer amplitude: $\sin(\pi \, J_{HN,N} \xi) \exp(-2\xi \, R_2(HN)) / \sqrt{2}$

maximum at $\xi = \tan^{-1}(\pi \, J_{HN,N} / 2 \, R_2(HN)) / \pi \, J_{HN,N}$

value without relaxation $\xi = 5.4$ ms

typical value $\xi = 5.4$ ms

can be optimized with Gradient selection.

preferable versus element 4., if $R_2(HN) < (15 \text{ ms})^{-1}$
Element 5.

(i) \( {\text{HN}}_z^i {\text{N}}_z^j \rightarrow N_z^i {\text{CA}}_z^{i-1} \)
(ii) \( {\text{HN}}_z^i {\text{N}}_z^j \rightarrow N_z^i {\text{CA}}_z^i \)

relevant constants: \( ^1J_{\text{HN,N}}, \, ^1J_{\text{N,CA}}, \, ^2J_{\text{N,CA}}, \, R_2(\text{N}) \)
transfer amplitudes:
(i): \( \exp(-\chi/R_2(\text{N})) \cos(\pi \, ^1J_{\text{N,CA}} \, \chi) \) \sin(\pi \, ^2J_{\text{N,CA}} \, \chi) \)
(ii): \( \exp(-\chi/R_2(\text{N})) \sin(\pi \, ^1J_{\text{N,CA}} \, \chi) \) \cos(\pi \, ^2J_{\text{N,CA}} \, \chi) \)

value without relaxation \( \chi = 25 \text{ ms} \)
typical value \( \chi = 22 \text{ ms} \)
\( \xi' = 5.4 \text{ ms fixed} \)

Element 6.

(i) \( N_z^i {\text{CA}}_z^{i-1} \rightarrow {\text{HN}}_z^i {\text{N}}_z^j \)
\( \rightarrow {\text{HN}}_z^i {\text{N}}_z^j {\text{CA}}_z^{i-1} {\text{CA}}_z^i \)
(ii) \( N_z^i {\text{CA}}_z^i \rightarrow {\text{HN}}_z^i {\text{N}}_z^j \)
\( \rightarrow {\text{HN}}_z^i {\text{N}}_z^j {\text{CA}}_z^{i-1} {\text{CA}}_z^i \)

relevant constants: \( ^1J_{\text{HN,N}}, \, ^1J_{\text{N,CA}}, \, ^2J_{\text{N,CA}}, \, R_2(\text{N}) \)
transfer amplitudes:
(i): \( \exp(-\chi/R_2(\text{N})) \sin(\pi \, ^1J_{\text{N,CA}} \, \chi) \cos(\pi \, ^2J_{\text{N,CA}} \, \chi) \)
(ii): \( \exp(-\chi/R_2(\text{N})) \cos(\pi \, ^1J_{\text{N,CA}} \, \chi) \sin(\pi \, ^2J_{\text{N,CA}} \, \chi) \)

value without relaxation \( \chi = 25 \text{ ms} \)
typical value \( \chi = 22 \text{ ms} \)
\( \xi' = 5.4 \text{ ms fixed} \)

Element 7.

(i) \( \text{HA}^i_z \rightarrow \text{HA}^i_z \, \text{CA}^i_z \)
(ii) \( \text{HB}^i_z \rightarrow \text{HB}^i_z \, \text{CB}^i_z \)

relevant constants: \( ^1J_{\text{HB,CB}}, \, ^1J_{\text{HA,CA}}, \, R_2(\text{HA}), \, R_2(\text{HB}) \)
transfer amplitudes:
(i): \( \sin(\pi \, ^1J_{\text{HA,CA}} \, \delta) \exp(-\delta \, R_2(\text{HA})) \)
(ii): \( \sin(\pi \, ^1J_{\text{HB,CB}} \, \delta) \exp(-\delta \, R_2(\text{HB})) \)

maximum at \( \delta = \tan^{-1}(\pi \, ^1J_{\text{HA,CA}} / R_2(\text{HA})) / \pi \, ^1J_{\text{HA,CA}} / \tan^{-1}(\pi \, ^1J_{\text{HB,CB}} / R_2(\text{HB})) / \pi \, ^1J_{\text{HB,CB}} \)
value without relaxation \( \delta = 3.6 \text{ ms for HA, } \delta = 3.8 \text{ ms for HB} \)
typical value \( \delta = 3.6 \text{ ms} \)
Element 8.

(i) $\text{HA}_z^i \text{CA}_x^i \rightarrow \text{CA}_x^i$
(ii) $\text{HB}_z^i \text{CB}_z^i \rightarrow \text{CB}_z^i \text{CA}_y^i$

relevant constants: $^1J_{\text{CA,CB}}$, $^1J_{\text{HB,CA}}$, $^1J_{\text{HA,CA}}$, $R_2(\text{CA})$, $R_2(\text{CB})$

transfer amplitudes:

(i):
- $G$: $0.90 \exp(-\kappa R_2(\text{CA}))$
- others: $0.85 \exp(-\kappa R_2(\text{CA})) \cos(\pi ^1J_{\text{CA,CB}} \kappa)$

(ii):
- $G$: $0$
- S, C, D, N, W, Y, F, H: $0.94 \exp(-\kappa R_2(\text{CB})) \sin(\pi ^1J_{\text{CA,CB}} \kappa)$
- A: $0.82 \exp(-\kappa R_2(\text{CB})) \sin(\pi ^1J_{\text{CA,CA}} \kappa)$
- V, I: $0.81 \exp(-\kappa R_2(\text{CB})) \sin(\pi ^1J_{\text{CA,CA}} \kappa) \cos^2(\pi ^1J_{\text{CB,CG}} \kappa)$
- L, T, K, M, R, Q, E, P: $0.94 \exp(-\kappa R_2(\text{CB})) \sin(\pi ^1J_{\text{CA,CA}} \kappa) \cos(\pi ^1J_{\text{CB,CG}} \kappa)$

$\delta' = 2.3$ ms leads to (Burum and Ernst 1980):

- alpha-CH: $\sin(\pi ^1J_{\text{HA,CA}} \delta') = 0.85$
- alpha-CH$_2$: $2 \sin(\pi ^1J_{\text{HA,CA}} \delta') \cos(\pi ^1J_{\text{HA,CA}} \delta') = 0.90$
- beta-CH: $\sin(\pi ^1J_{\text{HB,CA}} \delta') = 0.81$
- beta-CH$_2$: $2 \sin(\pi ^1J_{\text{HB,CA}} \delta') \cos(\pi ^1J_{\text{HB,CA}} \delta') = 0.94$
- beta-CH$_3$: $3 \sin(\pi ^1J_{\text{HB,CA}} \delta') \cos^2(\pi ^1J_{\text{HB,CA}} \delta') = 0.82$

typical value $\kappa = 6.6$ ms

Element 9.

(i) $\text{CA}_x^i \rightarrow \text{CA}_x^i \text{CO}_z^i$
(ii) $\text{CB}_x^i \text{CA}_y^i \rightarrow \text{CA}_x^i \text{CO}_z^i$

relevant constants: $^1J_{\text{CA,CO}}$, $^1J_{\text{CA,CA}}$, $R_2(\text{CA})$

transfer amplitudes:

(i):
- $G$: $\exp(-\lambda R_2(\text{CA})) \sin(\pi ^1J_{\text{CA,CO}} \lambda)$
- others: $\exp(-\lambda R_2(\text{CA})) \sin(\pi ^1J_{\text{CA,CO}} \lambda) \cos(\pi ^1J_{\text{CA,CA}} \lambda)$

(ii):
- $G$: $0$
- others: $\exp(-\lambda R_2(\text{CA})) \sin(\pi ^1J_{\text{CA,CO}} \lambda) \sin(\pi ^1J_{\text{CA,CA}} \lambda)$

typical value $\lambda = 7.4$ ms
Element 9b.

\[ \text{CB}_z \text{CA}_y^i \rightarrow \text{CA}_y^i \text{CO}_x^j \]

relevant constants: \( {}^1J_{\text{CA},\text{CO}}, {}^1J_{\text{CA},\text{CB}}, R_{2}(\text{CA}) \)

transfer amplitude: \( \exp(-\kappa R_{2}(\text{CA})) \sin(\pi \, {}^1J_{\text{CA},\text{CB}} \, \kappa) \)

maximum at \( \kappa = \tan^{-1}(\pi \, {}^1J_{\text{CA},\text{CB}} / R_{2}(\text{CA})) / \pi \, {}^1J_{\text{CA},\text{CB}} \)

value without relaxation \( \kappa = 13.5 \text{ ms} \)

typical value \( \kappa = 13.5 \text{ ms} \)

\( \lambda' = 9.5 \text{ ms fixed} \)

Element 10.

(i) \( \text{CA}_z^i \text{CO}_x^i \rightarrow \text{N}_z^i \text{CO}_x^i \)

(ii) \( \text{N}_z^i \text{CO}_x^i \rightarrow \text{CA}_z^i \text{CO}_x^i \)

relevant constants: \( {}^1J_{\text{CA},\text{CO}}, {}^1J_{\text{N},\text{CO}}, R_{2}(\text{CO}) \)

transfer amplitude: \( \sin(\pi \, {}^1J_{\text{N},\text{CO}} \, \eta) \exp(-\eta R_{2}(\text{CO})) \)

maximum at \( \eta = \tan^{-1}(\pi \, {}^1J_{\text{N},\text{CO}} / R_{2}(\text{CO})) / \pi \, {}^1J_{\text{N},\text{CO}} \)

value without relaxation \( \eta = 36 \text{ ms} \)

typical value \( \eta = 23 \text{ ms} \)

\( \theta' = 9.5 \text{ ms fixed} \)

Element 11.

(i) \( \text{CA}_z^i \text{HA}_x^i \rightarrow \text{N}_z^i \text{CA}_x^i \)

(ii) \( \text{CA}_z^i \text{HA}_x^i \rightarrow \text{N}_z^i \text{CA}_x^i \)

relevant constants: \( {}^1J_{\text{HA},\text{CA}}, {}^2J_{\text{N},\text{CA}}, {}^1J_{\text{N},\text{CA}}, {}^1J_{\text{CA},\text{CB}}, R_{2}(\text{CA}) \)

transfer amplitudes:

(i):

\( G: \quad 0.86 \exp(-\rho R_{2}(\text{CA})) \cos(\pi \, {}^2J_{\text{N},\text{CA}} \, \rho) \sin(\pi \, {}^1J_{\text{N},\text{CA}} \, \rho) \)

others: \( 0.86 \exp(-\rho R_{2}(\text{CA})) \cos(\pi \, {}^2J_{\text{N},\text{CA}} \, \rho) \sin(\pi \, {}^1J_{\text{N},\text{CA}} \, \rho) \cos(\pi \, {}^1J_{\text{CA},\text{CB}} \, \rho) \)

(ii):

\( G: \quad 0.86 \exp(-\rho R_{2}(\text{CA})) \sin(\pi \, {}^2J_{\text{N},\text{CA}} \, \rho) \cos(\pi \, {}^1J_{\text{N},\text{CA}} \, \rho) \cos(\pi \, {}^1J_{\text{CA},\text{CB}} \, \rho) \)

others: \( 0.86 \exp(-\rho R_{2}(\text{CA})) \sin(\pi \, {}^2J_{\text{N},\text{CA}} \, \rho) \cos(\pi \, {}^1J_{\text{N},\text{CA}} \, \rho) \cos(\pi \, {}^1J_{\text{CA},\text{CB}} \, \rho) \)

value without relaxation \( \rho = 26 \text{ ms} \)

typical value \( \rho = 23 \text{ ms} \)

\( \delta' = 4.75 \text{ ms} \)
Element 12.

\[ \begin{align*}
N'_{y}CA_{z}^{-1} & \rightarrow N'_{y}CA_{z}^{-1}CO'_{z}^{-1}CB_{y}^{-1} \\
N'_{y}CA_{z}^{-1} & \rightarrow N'_{y}CA_{z}^{-1}CO'_{z}CB_{y}^{-1} \\
N'_{y}CA_{z}^{-1}CO'_{z}^{-1} & \rightarrow N'_{y}CA_{z}^{-1}CB_{y}^{-1} \text{(non-Gly)} \\
& \rightarrow N'_{y}CA_{z}^{-1} \text{ (Gly)} \\
N'_{y}CA_{z}^{-1}CB_{y}^{-1} & \rightarrow N'_{y}CA_{z}^{-1}CO'_{z}^{-1} \text{(non-Gly)} \\
N'_{y}CA_{z}^{-1} & \rightarrow N'_{y}CA_{z}^{-1}CO'_{z}^{-1} \text{(Gly)} \\
\end{align*} \]

relevant constants: \( ^{1}J_{CA,CO} \), \( ^{1}J_{CA,CB} \), \( R_{2}(CA) \)

transfer amplitude: \( \exp(-\kappa R_{2}(CA)) \sin(\pi \cdot ^{1}J_{CA,CB} \kappa) \)

maximum at \( \kappa = \tan^{-1}(\pi \cdot ^{1}J_{CA,CB} / R_{2}(CA)) / \pi \cdot ^{1}J_{CA,CB} \)

value without relaxation \( \kappa = 13.5 \text{ ms} \)

typical value \( \kappa = 13.5 \text{ ms} \)

\( \lambda' = 9.5 \text{ ms} \) fixed

Element 13.

(i) \( CB_{z}^{-1}CA_{y}^{-1} \rightarrow N'_{y}CA_{z}^{-1} \)

\[ \rightarrow N'_{y}CA_{z}^{-1} \]

(ii) \( CA_{x}^{-1} \rightarrow N'_{y}CA_{z}^{-1} \)

\[ \rightarrow N'_{y}CA_{z}^{-1} \]

relevant constants: \( ^{1}J_{CA,CB} \), \( ^{2}J_{N,CA} \), \( R_{2}(CA) \)

transfer amplitudes:

(i): \( \exp(-\rho R_{2}(CA)) \sin(\pi \cdot ^{2}J_{N,CA} \rho) \sin(\pi \cdot ^{1}J_{CA,CB} \rho) \cos(\pi \cdot ^{1}J_{N,CA} \rho) \)

(ii): \( \exp(-\rho R_{2}(CA)) \sin(\pi \cdot ^{2}J_{N,CA} \rho) \cos(\pi \cdot ^{1}J_{CA,CB} \rho) \cos(\pi \cdot ^{1}J_{N,CA} \rho) \)


typical value \( \rho = 22 \text{ ms} \)

Element 13b.

Pulse sequence as element 13.

\[ \begin{align*}
CB_{z}^{-1}CA_{y}^{-1} & \rightarrow N'_{y}CA_{z}^{-1} \\
& \rightarrow N'_{y}CA_{z}^{-1} \\
\end{align*} \]

relevant constants: \( ^{1}J_{CA,CB} \), \( ^{2}J_{N,CA} \), \( R_{2}(CA) \)

transfer amplitude: \( \exp(-\rho R_{2}(CA)) \sin(\pi \cdot ^{2}J_{N,CA} \rho) \cos(\pi \cdot ^{1}J_{N,CA} \rho) \)


typical value \( \rho = 13.5 \text{ ms or } 40 \text{ ms} \)
Element 14.

(i) \( N_z^i C_A^{-1} \rightarrow N_z^i C_A^{-1} \)
(ii) \( N_z^i C_A^{-1} \rightarrow N_z^i C_A^{-1} C_B^{-1} \)
(iii) \( N_z^i C_A^{-1} \rightarrow N_z^i C_A^{-1} \)
(iv) \( N_z^i C_A^{-1} C_B^{-1} \rightarrow N_z^i C_A^{-1} \)

relevant constants: \( J_{CA,CB}, R_2(CA) \)
transfer amplitudes:
(i), (iii): \( \exp(-\kappa R_2(CA)) \sin(\pi J_{CA,CB} \kappa) \)
(ii), (iv): \( \exp(-\kappa R_2(CA)) \cos(\pi J_{CA,CB} \kappa) \)

value without relaxation \( \kappa = 6.8 \text{ ms} \)
typical value \( \kappa = 6.8 \text{ ms} \)

Element 14b.

Pulse sequence as element 14.
\( N_z^i C_A^{-1} \rightarrow N_z^i C_A^{-1} C_B^{-1} \)

relevant constants: \( J_{CA,CB}, R_2(CA) \)
transfer amplitude: \( \exp(-\kappa R_2(CA)) \sin(\pi J_{CA,CB} \kappa) \)

value without relaxation \( \kappa = 13.5 \text{ ms} \)
typical value \( \kappa = 10 \text{ ms} \)

Element 15.

\( N_z^i C_A^{-1} \rightarrow N_z^i C_A^{-1} C_O^{-1} \)
\( N_z^i C_A^{-1} C_O^{-1} \rightarrow N_z^i C_A^{-1} \)

relevant constants: \( J_{CA,CO}, J_{CA,CB}, R_2(CA) \)
transfer amplitude: \( \exp(-\lambda R_2(CA)) \sin(\pi J_{CA,CO} \lambda) \cos(\pi J_{CA,CB} \lambda) \)
typical value \( \lambda = 6.8 \text{ ms} \)

Element 16.

\( N_z^i C_O^{-1} \rightarrow N_z^i C_A^{-1} C_O^{-1} \)
\( N_z^i C_A^{-1} C_O^{-1} \rightarrow N_z^i C_O^{-1} \)

relevant constants: \( J_{CA,CO}, R_2(CO) \)
transfer amplitude: \( \exp(-\theta R_2(CO)) \sin(\pi J_{CA,CO} \theta) \)
value without relaxation $\theta = 9.5$ ms

**Typical value** $\theta = 8$ ms

**Element 17.**

$\text{HN}^i_2\text{N}_z^i \rightarrow \text{N}^i_2\text{CA}^i_2\text{CO}^i_z^{-1}$

relevant constants: $^1J_{\text{N,CA}}, ^2J_{\text{N,CA}}, R_2(\text{CO}), R_2(\text{N})$

transfer amplitude: $\exp(-\chi/R_2(\text{N})) \exp(-\theta R_2(\text{CO}))$

$\sin(\pi ^2J_{\text{N,CA}} \chi) \sin(\pi ^1J_{\text{N,CA}} \chi)$

$\Delta' = 5.4$ ms fixed

$\theta' = 9.5$ ms fixed

**Typical value** $\chi = 43.5$ ms

**Element 18.**

$\text{N}^i_2\text{CA}^i_2\text{CO}^i_z^{-1} \rightarrow \text{HN}^i_2\text{N}_z^i$

relevant constants: $^1J_{\text{N,CA}}, ^2J_{\text{N,CA}}, R_2(\text{CO}), R_2(\text{N})$

transfer amplitude: $\exp(-\chi/R_2(\text{N})) \exp(-\theta R_2(\text{CO}))$

$\sin(\pi ^2J_{\text{N,CA}} \chi) \sin(\pi ^1J_{\text{N,CA}} \chi)$

$\xi' = 5.4$ ms fixed

$\theta' = 9.5$ ms fixed

**Typical value** $\chi = 43.5$ ms

**Element 19.**

(i) $\text{HN}^i_2\text{N}_z^i \rightarrow \text{N}^i_2\text{CA}^i_2\text{CO}^i_z^{-1}$

(ii) $\text{HN}^i_2\text{N}_z^i \rightarrow \text{N}^i_2\text{CA}^i_2\text{CO}^i_z^{-1}$

relevant constants: $^1J_{\text{N,CA}}, ^2J_{\text{N,CA}}, ^1J_{\text{N,CO}}, R_2(\text{N})$

transfer amplitudes:

(i): $\exp(-\chi/R_2(\text{N})) \sin(\pi ^2J_{\text{N,CA}} \chi) \sin(\pi ^1J_{\text{N,CO}} \chi) \cos(\pi ^1J_{\text{N,CA}} \chi)$

(ii): $\exp(-\chi/R_2(\text{N})) \sin(\pi ^1J_{\text{N,CA}} \chi) \sin(\pi ^1J_{\text{N,CO}} \chi) \cos(\pi ^2J_{\text{N,CA}} \chi)$

**Typical value** $\chi = 28$ ms

$\xi' = 2.7$ ms fixed
Element 20.

(i) \( N_z^i C_{A_z}^{i-1} CO_z^{i-1} \rightarrow N_z^i C_{A_z}^{i-1} CA_z^{i-1} \)
(ii) \( N_z^i C_{A_z}^{i-1} CO_z^{i-1} \rightarrow N_z^i C_{A_z}^{i-1} CA_z^{i-1} \)

relevant constants: \( J_{HA,CA}, J_{CA,CB}, R_2(CA) \)
transfer amplitudes:
(i): \( \exp(-\rho R_2(CA)) \sin(\pi J_{N,CA} \rho) \sin(\pi J_{N,CA} \rho) \cos(\pi J_{CA,CB} \rho) \)
(ii): \( \exp(-\rho R_2(CA)) \cos(\pi J_{N,CA} \rho) \cos(\pi J_{N,CA} \rho) \cos(\pi J_{CA,CB} \rho) \)

typical value \( \rho = 28 \text{ ms} \)
\( \theta' = 9.5 \text{ ms fixed} \)

Element 21.

\( N_z^i C_{A_z}^{i-1} CO_z^{i-1} \rightarrow N_z^i C_{A_z}^{i-1} CO_z^{i-1} HA_z^{i-1} \)
\( N_z^i C_{A_z}^{i-1} CO_z^{i-1} HA_z^{i-1} \rightarrow N_z^i C_{A_z}^{i-1} CO_z^{i-1} \)

relevant constants: \( J_{HA,CA}, J_{CA,CB}, R_2(CA) \)
transfer amplitudes:
CH: \( \exp(-\delta R_2(CA)) \sin(\pi J_{HA,CA} \delta) \cos(\pi J_{CA,CB} \delta) \)
CH2: \( \exp(-\delta R_2(CA)) \sin(\pi J_{HA,CA} \delta) \cos(\pi J_{CA,CB} \delta) \cos(\pi J_{HA,CA} \delta) \)

typical value \( \delta = 2.3 \text{ ms} \)

Element 22.

(i) \( N_z^i C_{A_z}^{i-1} CO_z^{i-1} \rightarrow HN_z^i N_z^i \)
(ii) \( N_z^i C_{A_z}^{i-1} CO_z^{i-1} \rightarrow HN_z^i N_z^i \)

relevant constants: \( J_{N,CA}, J_{HN,N}, J_{N,CA}, J_{N,CO}, R_2(N) \)
transfer amplitudes:
(i): \( \exp(-\chi R_2(N)) \sin(\pi J_{N,CA} \chi) \sin(\pi J_{N,CO} \chi) \cos(\pi J_{N,CA} \chi) \)
(ii): \( \exp(-\chi R_2(N)) \sin(\pi J_{N,CA} \chi) \sin(\pi J_{N,CO} \chi) \cos(\pi J_{N,CA} \chi) \)

typical value \( \chi = 28 \text{ ms} \)
\( \xi' = 5.4 \text{ ms fixed} \)

Element 23.

\( HA_z^{i-1} C_{A_z}^{i-1} \rightarrow CA_z^{i-1} CO_z^{i-1} \)

relevant constants: \( J_{HA,CA}, J_{CA,CB}, R_2(CA) \)
transfer amplitude:
\( \exp(-\lambda R_2(CA)) \sin(\pi J_{CA,CO} \lambda) \cos(\pi J_{CA,CB} \lambda) \)
for small proteins, $\lambda = 26$ ms and $\delta' = 4.75$ ms
otherwise $\lambda = 6.8$ ms, and an additional $^1$H-180° pulse is included to allow $\delta' = 4.75$ ms.

Element 24.
(i) $\text{CA}^{i-1}\text{CO}^{-1} \rightarrow N_z^i\text{CA}^{-1}$
(ii) $\text{CA}^{i-1}\text{CO}^{-1} \rightarrow N_z^{i-1}\text{CA}^{-1}$

relevant constants: $^1J_{N,CA}, ^2J_{N,CA}, ^1J_{CA,CO}, R_2(\text{CA})$
transfer amplitudes:
(i) : $\exp(-\rho/R_2(\text{CA})) \sin(\pi J_{N,CA} \rho) \cos(\pi J_{CA,CO} \rho) \cos(\pi J_{N,CA} \rho)$
(ii): $\exp(-\rho/R_2(\text{CA})) \cos(\pi J_{N,CA} \rho) \sin(\pi J_{CA,CO} \rho) \cos(\pi J_{N,CA} \rho)$

typical value $\rho = 26$ ms
$\theta' = 9.5$ ms fixed

Element 25.
(i) $N_z^i\text{CA}^{i-1} \rightarrow N_z^i\text{CA}^{-1}$
(ii) $N_z^{i-1}\text{CA}^{-1} \rightarrow N_z^{i-1}\text{CA}^{i-1}$

relevant constants: $^1J_{N,CA}, ^2J_{N,CA}, ^1J_{CA,CB}, R_2(\text{CA})$
transfer amplitudes:
(i) : $\exp(-\rho/R_2(\text{CA})) \cos(\pi J_{N,CA} \rho) \cos(\pi J_{CA,CB} \rho)$
(ii): $\exp(-\rho/R_2(\text{CA})) \sin(\pi J_{N,CA} \rho) \sin(\pi J_{CA,CB} \rho) \cos(\pi J_{CA,CB} \rho)$

typical value $\rho = 50$ ms

Element 26.
$\text{HN}^{i}_z\text{N}^i_z \rightarrow \text{N}^i_z\text{CA}^{-1}\text{CO}^{-1} \text{CA}^{-1}$

relevant constants: $^2J_{N,CA}, ^1J_{N,CA}, ^1J_{N,CO}, R_2(\text{N})$
transfer amplitude: $\exp(-\chi/R_2(\text{N})) \cos(\pi J_{N,CA} \chi) \cos(\pi J_{N,CA} \chi)$

$\xi' = 4.7$ ms fixed
$\tau' = 36$ ms fixed

typical value $\chi = 50$ ms
Element 27.

\[ ^{1}N_2 ^{1}C_A ^{1}C_O x^{-1} C_A x^{-1} \rightarrow ^{1}H N_2 ^{1}N_x \]

relevant constants: \(^{2}J_{N,C_A}, ^{1}J_{N,C_A}, ^{1}J_{N,C_O}, R_2(N)\)
transfer amplitude: \(\exp(-\chi/R_2(N))\cos(\pi^{2}J_{N,C_A} \chi)\cos(\pi^{1}J_{N,C_A} \chi)\)

typical value \(\chi = 50 \text{ ms}\)
\(\xi' = 4.7 \text{ ms fixed}\)
\(\tau' = 36 \text{ ms fixed}\)

Element 28.

(i) \(^{1}N_2 ^{1}C_A x^{-1} \rightarrow ^{1}N_2 ^{1}C_A x^{-1} ^{1}C_B y^{-1}\)
(ii) \(^{1}N_2 ^{1}C_A x^{-1} ^{1}C_B y^{-1} \rightarrow ^{1}N_2 ^{1}C_A x^{-1}\)

relevant constants: \(^{1}J_{C_A,C_B}, R_2(CA)\)
transfer amplitude: \(\exp(-\chi/R_2(CA))\cos(\pi^{1}J_{C_A,C_B} \kappa)\)

typical value \(\kappa = 27 \text{ ms}\)

Element E1.

Evolution of CO

relevant constants: \(R_2(CO)\)
transfer amplitude: \(\exp(-t R_2(CO))\)

Element E2.

Evolution of CA or CA–CB DQ

relevant constants: \(^{1}J_{C_A,C_B}, R_2(CA)\)
transfer amplitude: \(\exp(-t R_2(CA)) \cos(\pi^{1}J_{C_A,C_B} t)\)

Element E3.

CT-Evolution of CA

relevant constants: \(^{1}J_{C_A,C_B}, R_2(CA)\)
transfer amplitude: \(\exp(-T R_2(CA)) \cos(\pi^{1}J_{C_A,C_B} T)\)

\(T = 26.6 \text{ ms}\)
**Element E4.**

HMQC CO → CA

relevant constants: $J_{CA,CB}, J_{CA, CO}, R_2(CO), R_2(CA)$

transfer amplitude: $\exp(-(20 + t) R_2(CO)) \exp(-t R_2(CA)) \sin^2(\pi J_{CA,CO} t) \cos(\pi J_{CA,CB} t)$

value without relaxation $\theta = 9.5$ ms

typical value $\theta = 7.0$ ms

**Element E5.**

HMQC CA → CO

relevant constants: $J_{CA,CB}, J_{CA, CO}, R_2(CO), R_2(CA)$

transfer amplitude: $\exp(-(2\lambda + T) R_2(CA)) \exp(-T R_2(CO)) \sin^2(\pi J_{CA,CO} \lambda)$

constant time $T$, so that $2\lambda + T = 1/1 J_{CA,CB} = 26.6$ ms

value without relaxation $\lambda = 9.5$ ms

typical value $\lambda = 7.0$ ms

**Element E6.**

Evolution of CA or CA–CB DQ

relevant constants: $J_{CA,CB}, R_2(CA)$

transfer amplitude: $\exp(-t R_2(CA)) \cos(\pi J_{CA,CB} t)$

**Element E7.**

HMQC N → CO, CA

relevant constants: $J_{CA,CB}, J_{CA,CO}, J_{N,CA}, J_{N,CA}, R_2(CO), R_2(N), R_2(CA)$

transfer amplitude:

$\exp(-t_1 R_2(CO)) \exp(-t_2 R_2(CA)) \exp(-(2\tau + t_1 + t_2)/ R_2(N)) \sin^2(\pi J_{CA,CO} \tau) \sin^2(\pi J_{N,CA} \tau) \cos^2(\pi J_{N,CA} \tau) \cos(\pi J_{CA,CB} t_2)$

typical value $\tau = 28.0$ ms
Element E8.
Evolution of N
relevant constants: $R_2(N)$
transfer amplitude: $\exp(-t \, R_2(N))$

Element E9.
Evolution of CB
relevant constants: $R_2(CB)$
transfer amplitude: $\exp(-t \, R_2(CB))$
side chain dependent

Element E10.
Evolution of HA
relevant constants: $R_2(HA)$
transfer amplitude: $\exp(-t \, R_2(HA))$
References


Curriculum Vitae

Personal:
Name: Sebastian Hiller
Date of birth: December 14, 1976
Place of birth: Pforzheim, Germany
Nationality: German
Marital status: Single

Education:

- Since 1998: Member of the Swiss science foundation / Schweizerische Studienstiftung.

- Teaching assistances in geometry, statistics and physical chemistry.

- Erasmus exchange studies at Cambridge University, UK, for 2 trimesters in 2000/2001.

- Diploma thesis with Prof. Dr. Kurt Wüthrich. Title: “Application of the spin relaxation agent Gd(DOTA) in NMR spectroscopy of large proteins in aqueous solution.”

Publications


