

**Automatische sequentielle Zuordnung von mehrdimensionalen  
Protein-NMR-Spektren sowie molekulardynamisch gestützte  
stereospezifische Zuordnung von Seitenkettenamidgruppen in  
Modellpeptiden**



DISSERTATION ZUR ERLANGUNG DES  
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN  
DER UNIVERSITÄT REGENSBURG

vorgelegt von  
**Tobias Harsch**  
aus Wernberg

im Jahr 2015

Das Promotionsgesuch wurde eingereicht am:

14. Dezember 2015

Die Arbeit wurde angeleitet von:

Prof. Dr. rer. nat. Dr. med. Hans Robert Kalbitzer

Prüfungsausschuss:

Vorsitzender: Prof. Dr. rer. nat. Christoph Oberprieler

Erstgutachter: Prof. Dr. rer. nat. Dr. med. Hans Robert Kalbitzer

Zweitgutachter: Prof. Dr. rer. nat. Wolfram Gronwald

Drittprüfer: Prof. Dr. rer. nat. Werner Kremer

Unterschrift:

---

Dipl.-Phys. Tobias Harsch

# Inhalt

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in den Modellpeptiden Gly-Gly-Asn-Ala-NH <sub>2</sub> und Gly-Gly-Gln-Ala-NH <sub>2</sub>	2
1.2	Automatische sequentielle Zuordnung (SIBASA)	4
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>8</b>
2.1	Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin	8
2.1.1	Allgemeine Herangehensweise	8
2.1.2	Simulation der NOESY-Spektren	9
2.1.3	Bestimmung der chemische Austauschraten	12
2.1.4	Vorhersage von chemischen Verschiebungen aus den MD-Trajektorien	13
2.1.5	Entnahme der Dynamik Parameter aus der MD-Simulation	13
2.2	Automatische sequentielle Zuordnung (SIBASA)	17
2.2.1	Vorbereitung	17
2.2.2	Vorbereitung der experimentelle Spektren	19
2.2.3	Simulation der NOESY-Spektren	20
2.2.4	Bestimmung der Volumenskalierungsfaktoren	22
2.2.5	Bestimmung der zusätzlichen Linienverarbeitungen	25
2.2.6	Suchraum der chemischen Verschiebungen (Slots)	27
2.2.7	Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen	30
2.2.8	Pseudoenergien	34
2.2.9	Behandlung von Kernen mit nicht sichtbaren Resonanzlinien	49
2.2.10	Threshold Accepting	51
2.2.11	MOVES	53
2.2.12	Verifikation der Zuordnungen	64
2.2.13	Festlegen von Zuordnungen während der Laufzeit	65
2.2.14	Dynamische Anpassung der Volumenskalierungsfaktoren	71
2.2.15	Absolute Bewertung von Zuordnungen	72
<b>3</b>	<b>Material und Methoden</b>	<b>74</b>
3.1	Die Modellpeptide Gly-Gly-Asn-Ala-NH <sub>2</sub> und Gly-Gly-Gln-Ala-NH <sub>2</sub>	74
3.1.1	Probe	74
3.1.2	NMR Messungen	74

3.1.3	Molekulardynamiksimulationen.....	74
3.1.4	Rückrechnungen.....	75
3.2	Die Punktmutante HPr S. aureus (H15A).....	76
3.2.1	Probe.....	76
3.2.2	NMR Messungen.....	77
3.2.3	Modellierung und Molekulardynamiksimulationen.....	78
3.2.4	Simulation der NOESY-Spektren.....	79
3.2.5	Automatische Zuordnung mit SIBASA.....	80
3.3	Plasmodium falciparum Thioredoxin.....	81
3.3.1	Probe.....	81
3.3.2	NMR Messungen.....	82
3.3.3	Modellierung und Molekulardynamiksimulationen.....	83
3.3.4	Simulation der NOESY-Spektren.....	83
3.3.5	Automatische Zuordnung mit SIBASA.....	84
3.4	Ras(T35S)-GppNHp.....	85
3.4.1	Probe.....	86
3.4.2	NMR Messungen.....	87
3.4.3	Molekulardynamiksimulationen.....	88
3.4.4	Simulation der NOESY-Spektren.....	88
3.4.5	Automatische Zuordnung mit SIBASA.....	89
<b>4</b>	<b>Ergebnisse.....</b>	<b>90</b>
4.1	Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in Modellpeptiden.....	90
4.1.1	Die stereospezifische Zuordnung von Gly-Gly-Asn-Ala-NH <sub>2</sub> .....	91
4.1.2	Die stereospezifische Zuordnung von Gly-Gly-Gln-Ala-NH <sub>2</sub> .....	94
4.1.3	Erzeugung der Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der Amidprotonen.....	96
4.2	Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen.....	97
4.3	Automatische sequenzielle Zuordnung (SIBASA).....	102
4.3.1	Erzeugung der Häufigkeitsverteilungen der Pseudoenergien für die AURMOL Datenbank.....	102
4.3.2	Automatische Zuordnung der Punktmutante HPr S. aureus (H15A).....	113

4.3.3	Automatische Zuordnung des reduzierten Plasmodium falciparum Thioredoxin.....	132
4.3.4	Automatische Zuordnung von Ras(T35S)-GppNHp.....	151
<b>5</b>	<b>Diskussion.....</b>	<b>171</b>
5.1	Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in Modellpeptiden.....	171
5.1.1	Zuordnungen.....	171
5.1.2	Stereospezifische Zuordnung.....	171
5.1.3	Bestimmung der chemischen Austauschraten.....	173
5.1.4	Aus den 3D-Strukturen vorhergesagte Dichteverteilungen der chemische Verschiebungen.....	174
5.2	Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen.....	175
5.3	Automatische sequentielle Zuordnung (SIBASA).....	179
5.3.1	Ablauf der automatischen Zuordnung.....	181
5.3.2	Vorbereitung der Röntgenstrukturen.....	182
5.3.3	Automatische Bestimmung der Volumenskalierungsfaktoren und der zusätzlichen Linienverbreiterungen.....	184
5.3.4	Schätzung der Häufigkeitsverteilung der Pseudoenergien.....	187
5.3.5	Verlauf von Pseudoenergien und dynamischer Größen während der automatischen Zuordnung.....	189
5.3.6	Festlegen von Zuordnung zur Laufzeit mithilfe mehrerer Instanzen.....	190
5.3.7	Diskussion der Ergebnisse der automatischen sequentiellen Zuordnung..	191
5.3.8	Falsch bestätigte Resonanzen von Seitenketten.....	194
5.3.9	Automatische Zuordnung von HSQC-Spektren.....	195
5.3.10	Automatische Zuordnung der HSQC-Spektren von Punktmutanten.....	203
5.3.11	Ausblick.....	204
<b>6</b>	<b>Zusammenfassung.....</b>	<b>206</b>
<b>7</b>	<b>Literaturverzeichnis.....</b>	<b>208</b>
<b>8</b>	<b>Danksagung.....</b>	<b>218</b>

# 1 Einleitung

Die NMR (Nuklear **M**agnetic **R**esonance) Spektroskopie ist ein mächtiges Werkzeug um die geordnete 3D-Struktur, also die räumliche Anordnung der Atome eines Proteinen zu analysieren. Mit mehrdimensionalen NOESY-Spektren (**N**uclear **O**verhauser **E**nhancement **S**pectroscopy) (Kumar et al. 1980) können die Abstände zwischen zwei Atomkernen im Protein aus den Volumen des zugehörigen NOESY-Kreuzsignals bestimmt werden. Umgekehrt ist es möglich mit Programmen wie AUREMOL RELAX (Ried et al. 2004) die Intensitäten und Linienformen aller NOESY-Signale aus einer vorhandenen 3D-Struktur des Proteins zurückzurechnen. Für ein vollständiges simuliertes 2-D- oder 3-D-NOESY-Spektrum sind dann nur noch die chemischen Verschiebungen der Kernresonanzen notwendig.

Die Rückrechnungen der NOESY-Signale aus den 3D-Strukturen dienen in dieser Arbeit als Basis des sogenannten Top-Down Ansatzes, bei dem die zurückgerechneten Einzelsignale so angeordnet werden, dass die resultierenden simulierten NOESY-Spektren bestmöglich mit den gemessenen NOESY-Spektren übereinstimmen. Als Ergebnis können die chemischen Verschiebung aller Resonanzen des Proteins nach der Anordnung der simulierten Signale abgelesen werden.

Da zwei oder mehrerer Protonen, die kovalent an das selbe schwere Atom gebunden sind, nicht weit voneinander entfernt sind, haben sie der Regel ähnliche NOESY-Signalmuster, aber unterschiedliche chemische Verschiebungen. Sie sind bei der Aufklärung der vollständigen Liste der chemischen Verschiebungen nur schwer voneinander zu unterscheiden. Im ersten Teil dieser Arbeit, der stereospezifischen Zuordnung der Seitenkettenamidgruppen von Asparagin und Glutamin, wird mithilfe des Top-Down-Ansatzes die stereospezifische Zuordnung der Resonanzen der NH<sub>2</sub>-Protonen der Seitenketten von Asparagin und Glutamin in den Modellpeptiden Gly-Gly-Asn-Ala-NH<sub>2</sub> Gly-Gly-Gln-Ala-NH<sub>2</sub> händisch bestimmt.

Im zweiten Teil wird das neue AUREMOL Modul SIBASA, die Weiterentwicklung von ASSIGN (Brunner 2006) vorgestellt. SIBASA verwendet den Top-Down-Ansatz um die vollständige Zuordnung von Proteinen automatisch zu bestimmen. Als Eingabe dient lediglich die Röntgenstruktur des Proteins. SIBASA ist zwar darauf spezialisiert worden HSQC-Spektren (**H**eteronuclear **S**ingle **Q**uantum **C**oherence) (Bodenhausen und Ruben 1980) von Proteinen aus dem 2-D-NOESY-Spektrum und dem 3D <sup>15</sup>N-NOESY-HSQC-Spektrum zuzuordnen, ordnet aber im Prozess alle in den Eingangsspektren sichtbaren Resonanzen zu.

Im Folgenden wird noch eine Einführung in die beiden Problemstellung gegeben. In Kap. 2 werden die theoretischen Grundlagen für die beiden Teile dieser Arbeit erläutert. In Kap. 4.1 werden die Ergebnisse der stereospezifischen Zuordnung der Seitenkettenamide von Asn und Gln, die mit Hilfe des Top-Down-Ansatzes und der Molekulardynamik Simulationen der Tetrapeptide Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Asn-Ala-NH<sub>2</sub> gewonnen wurden, beschrieben. In Kap. 4.2 werden die Ergebnisse der verbesserten Vorhersagen von chemische

Verschiebungen aus 3D-Strukturen von Proteinen, die für die automatische Zuordnung mit SIBASA sehr wichtig sind, gezeigt. Im Kap. 4.3 werden schließlich die Ergebnisse der automatischen Zuordnung mit SIBASA und insbesondere die Ergebnisse der automatischen Zuordnung der HSQC-Spektren von drei Proteinen unterschiedlicher Größe dargelegt. In Kap. 5 werden die Ergebnisse dieser Arbeit diskutiert und in Kap. 6 noch einmal zusammengefasst.

## 1.1 Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in den Modellpeptiden Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Gln-Ala-NH<sub>2</sub>

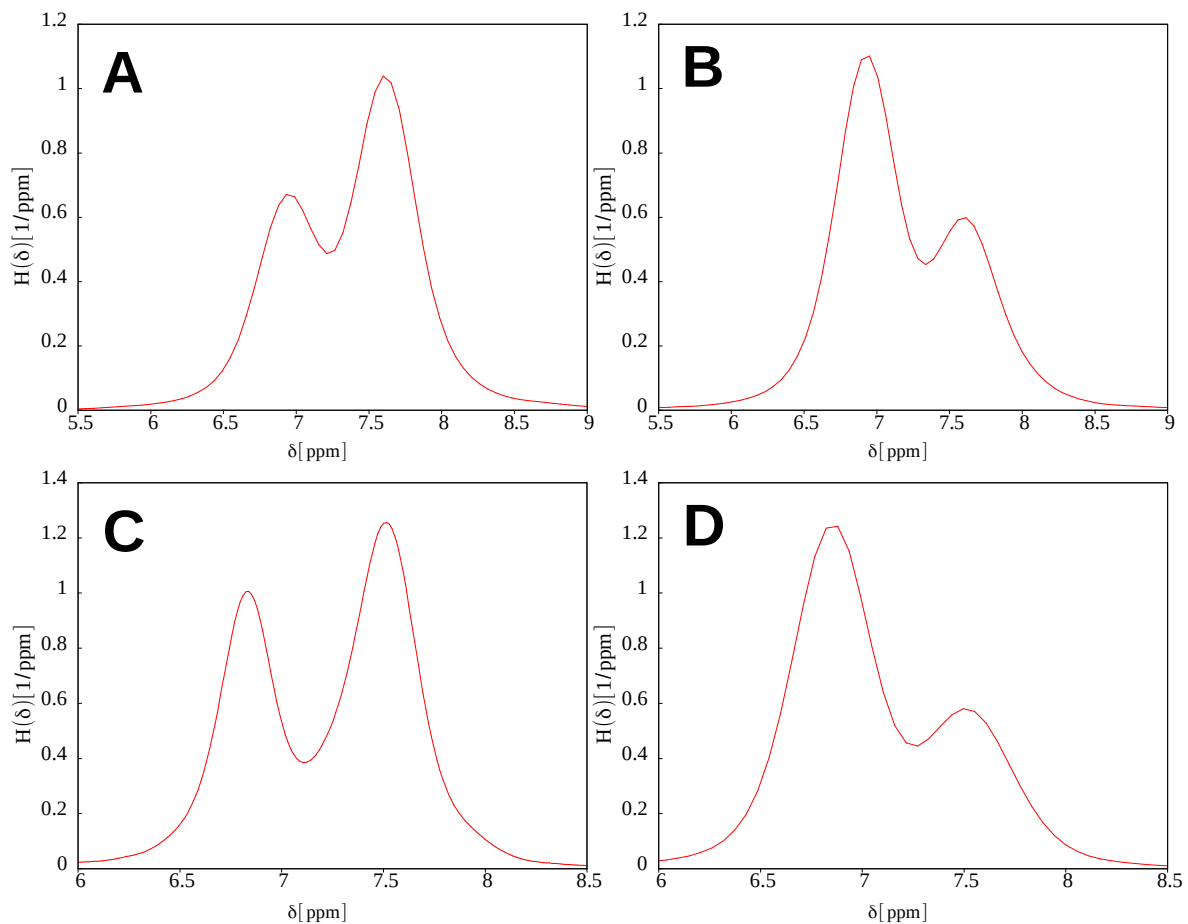


Abbildung 1: **Häufigkeitsverteilung  $H(\delta)$  der chemischen Verschiebungen  $\delta$  der Amidprotonen der Asparagin- und Glutaminreste.** Die Daten wurden aus der BMRB Datenbank entnommen und mit einem Kerndichteschätzer geglättet. **A** Asn  $H^{\delta 21}$ , **B** Asn  $H^{\delta 22}$ , **C** Gln  $H^{\delta 21}$ , **D** Gln  $H^{\delta 22}$ . Maxima für Asn  $H^{\delta 21}$  (6,93 ppm, 7,60 ppm), Asn  $H^{\delta 22}$  (6,95 ppm, 7,61 ppm), Gln  $H^{\delta 21}$  (6,84 ppm, 7,52 ppm), Gln  $H^{\delta 22}$  (6,88 ppm, 7,50 ppm).

Bei der Zuordnung der chemischen Verschiebungen eines Proteins sind

Zuordnungsdatenbanken bereits gelöster Biomoleküle hilfreich. Diese Datenbanken enthalten neben den Intervallen in denen Resonanzen von Kernen häufig vorkommen auch die Wahrscheinlichkeit  $P(\delta | \text{Aminosäure} \cap \text{Atom})$ , dass die Resonanz eines bestimmter Kern einer Aminosäure die chemische Verschiebung  $\delta$  einnimmt. Die Zuverlässigkeit der hiervon abgeleiteten Voraussagen hängt hauptsächlich von der Güte der zugrundeliegenden Daten ab. Für biologische Anwendungen ist die **Biological Magnetic Resonance Bank** (BMRB) (Ulrich et al. 2008) die am weitesten verbreitete Datenbank für chemische Verschiebungen und enthält unter anderem auch die stereospezifischen Zuordnung der Seitenkettenamide von Asparagin und Glutamin. Diese werden bei der Zuordnung von Biomolekülen experimentell über ihr NOE-Muster im homonuklearen (Wüthrich 1986) oder über das J-Kopplungsmuster im heteronuklearen Spektrum (McIntosh et al. 1997) zugeordnet. Abb. 1 zeigt die Häufigkeitsverteilungen der chemischen Verschiebungen der Seitenkettenamide von Asparagin und Glutamin aus der BMRB. Die vier Verteilungen zeigen jeweils zwei getrennte Maxima, die im Falle der beiden Verteilungen von Asparagin 0,67 ppm und im Falle der beiden Verteilungen von Glutamin 0,68 ppm voneinander entfernt liegen. Zwei Maxima in diesen Häufigkeitsverteilungen sind hier nicht nachvollziehbar, denn chemische Verschiebungen setzten sich in der Regel aus zwei Beiträgen zusammen, nämlich der intrinsischen chemischen Verschiebung des Kerns und dem Einfluss der Umgebung auf diesen. Der erste Beitrag, die sogenannte Random-Coil-Verschiebung, kann aus NMR-Messung der Tetrapeptide Gly-Gly-Asn-Ala und Gly-Gly-Gln-Ala gewonnen werden. Die gemessenen Resonanzen der Seitenkettenamide von  $H^{\delta 21}$  und  $H^{\delta 22}$  im Asn sind 0,68 ppm, die Resonanzen von Gln  $H^{\epsilon 21}$  und  $H^{\epsilon 22}$  sind 0,71 ppm voneinander entfernt (Bundi und Wüthrich 1979). Der zweite Anteil ist von Fall zu Fall verschieden, ist aber statistisch um 0 ppm verteilt und erzeugt somit in der Regel eine unimodale Verteilung um den Random-Coil-Wert. Es kann also angenommen werden, dass ein großer Teil der stereospezifisch Zuordnungen falsch in der BMRB abgelegt wurden. Der wahrscheinlichste Schluss den man aus den Verteilungen in Abb. 1 ziehen kann ist, dass Asn  $H^{\delta 21}$  gegenüber  $H^{\delta 22}$  um ca. 0,67 ppm in Richtung Tieffeld und Gln  $H^{\epsilon 21}$  gegenüber dem  $H^{\epsilon 22}$  um 0,68 ppm ebenfalls in Richtung Tieffeld verschoben ist. Eine Möglichkeit die Fehler in der BMRB zu korrigieren wäre die chemischen Verschiebungen von Asn  $H^{\delta 21}$  und  $H^{\delta 22}$  bzw. von Gln  $H^{\epsilon 21}$  und  $H^{\epsilon 22}$  in der Datenbank zu vertauschen, falls gilt:

- $\delta(\text{Asn } H^{\delta 21}) < \delta(\text{Asn } H^{\delta 22})$
- $\delta(\text{Gln } H^{\epsilon 21}) < \delta(\text{Gln } H^{\epsilon 22})$

Eine Voraussetzung dafür ist, dass die Reihenfolge der stereospezifischen Zuordnungen in den Random-Coil-Peptiden bekannt ist. Konkret haben wir NOESY-Spektren von Tetrapeptiden und deren Rückrechnungen verwendet um die Seitenkettenamide von Asn und Gln zuzuordnen. Dabei wurden für die Rückrechnungen der vollständige Relaxationsmatrixformalismus und die Ergebnisse von Molekuldynamik-Simulationen verwendet.



## 1.2 Automatische sequentielle Zuordnung (SIBASA)

Die Bestimmung von 3D-Strukturen von Protein aus NMR-Daten ist normalerweise ein sehr zeitaufwendiger Prozess. Es ist daher erstrebenswert die Strukturaufklärung zu automatisieren. Prinzipiell ist es möglich die Struktur aus der Primärsequenz mittels Molekulardynamik Simulationen (MD) zu bestimmen, was aber große Computer Cluster und sehr viel Rechenzeit benötigen würde. Da diese Voraussetzungen selten erfüllt sind, aber 3D-Strukturen in Forschung und Industrie oft benötigt werden, sucht man nach automatischen Lösungen, die auch auf handelsüblichen Computern durchführbar sind.

Für die Strukturaufklärung mit NMR ist die vollständige Liste der chemischen Verschiebungen des Proteins zwar nicht zwingend erforderlich, sie verbessert aber die Qualität der resultierenden 3D-Struktur deutlich.

Üblicherweise arbeiten Programme, die automatisch Zuordnungen finden mit dem Bottom-up-Ansatz, werten also hauptsächlich die Primärsequenz des Proteins und die experimentellen Signallisten aus. Die meisten entwickelten Algorithmen beschränken sich auf die Auswertung von Triple-Resonanz-Spektren, die das gesamte Rückgrat des Proteins über  $^1\text{J}$ - und  $^2\text{J}$ -Kopplungen miteinander verbinden. Mit diesen Spektren sind die automatische Methoden in der Lage die Resonanzen der Hauptketten und oft auch die Resonanzen der  $\text{C}^\beta$ -Kerne zuzuordnen. Dafür verwenden die meisten Programme Heuristiken, den Simulated Annealing Algorithmus, Datenbanken, Brute-Force- oder Monte-Carlo-Methoden (Güntert 2009). Vertreter dieser Programme sind z.B. MATCH (Volk et al. 2008), AUTOASSIGN (Zimmerman et al. 1997; Moseley und Montelione 1999; Moseley et al. 2001; Moseley, Hunter N B et al. 2004) und MARS (Jung und Zweckstetter 2004b, 2004a; Narayanan et al. 2010). Weitere Ansätze können in Andrec und Levy (2002), Atreya et al. (2000) (2002), Chen et al. (2005), Coggins (2003), Friedrichs et al. (1994), Güntert et al. (2000), Hare und Prestegard (1994), Kamisetty et al. (2006), Kjær et al. (1994), Leutner et al. (1998), Li und Sanctuary (1997a), Lin et al. (2005), Lukin et al. (1997), Masse und Keller (2005), Olson und Markley (1994), Vitek et al. (2005), Wang et al. (2005), Wu et al. (2006) und Xu et al. (2002) (2006) nachgelesen werden.

Sollen ebenfalls die Seitenketten automatisch zugeordnet werden, so wird meistens auf mehrdimensional  $^1\text{H}$ - $^1\text{H}$ -COSY- (**C**orrelated **S**pectroscop**F**) (Aue et al. 1976) oder  $^1\text{H}$ - $^1\text{H}$ -TOCSY-Spektren (**T**otal **C**orrelation **S**pectroscopy) (Braunschweiler und Ernst 1983) zurückgegriffen. In diesen Spektren, sind die Protonen des Proteins über die  $^2\text{J}$ - und  $^3\text{J}$ -Kopplungen miteinander verbunden, was dazu führt, dass nur intraresiduale Kreuzsignale von Protonen einer Aminosäure untereinander sichtbar sind. Ziel vieler dieser Algorithmen zur Bestimmung der chemischen Verschiebungen der Seitenketten ist die Signale der Aminosäuren in diesen Spektren zu sogenannten intraresidualen Spinsystemen zusammenzufassen und diese mit Hilfe von Triple-Resonanz-Spektren auf die Primärsequenz abzubilden (Li und Sanctuary 1997b; Eghbalnia et al. 2005). Durch Mustererkennung können

in der Regel die möglichen Aminosäuren der Spinsysteme eingeschränkt und die Kreuzsignale der Aminosäuren in den Spektren voneinander getrennt werden. In CAPRI (Choy et al. 1997) erfolgt die Mustererkennung über neuronale Netze und wird durch die Vorhersage der Sekundärstruktur des Proteins aus der Primärsequenz (Rost und Sander 1993) unterstützt. CAMRA (Gronwald et al. 1998) verwendet die chemischen Verschiebungen von homologen Proteinen als zusätzliche Informationen, um die Spinsysteme aus TOCSY-Spektren in die Primärsequenz einzuordnen. MONTE (Hitchens et al. 2003) verwendet Simulated Annealing, Triple-Resonanz-Spektren und 4D Amid-Amid-NOESY-Spektren um die Resonanzen von deuterierten Proteinen zuzuordnen. SIDELINK (Masse et al. 2006) verwendet die chemischen Verschiebungen der Hauptketten ( $H^N$ ,  $C^\alpha$ ,  $H^\alpha$ ) und mehrere 3-D-NOESY-Spektren für die Zuordnung. In ST2NMR (Pristovsek et al. 2002) und in Xu et al. (1993) (1994) werden statt Triple-Resonanz-Spektren 2-D- und 3-D-NOESY-Spektren verwendet, um die mit den TOCSY-Spektren erzeugten intraresidualen Spinsysteme auf die Primärsequenz abzubilden.

GARANT (Bartels et al. 1996; Bartels et al. 1997; Güntert 2009) verfolgt einen allgemein gültigen Ansatz. Als Eingabe dient mindestens die Primärsequenz des Proteins und eine Kreuzsignal-Liste eines NMR-Spektrums. Zusätzlich können weitere Kreuzsignallisten von weiteren NMR-Experimenten, 3D-Strukturen, bekannte und geschätzte chemische Verschiebungen übergeben werden. Aus der Primärsequenz und der Topologie Datenbank des Magnetisierungstransfers der verwendeten NMR-Experimente werden die zu erwartenden Signale bestimmt. Mit Hilfe eines genetischen Algorithmus und einer anschließenden lokalen Optimierung wird bestmöglich versucht die erwarteten auf die sichtbaren Kreuzsignale abzubilden. Als Ergebnis erhält man, soweit sie in den Eingabespektren untersucht wurden, eine vollständige Liste der chemischen Verschiebungen des Proteins. Die zu erwartenden NOESY-Kreuzsignale können, sollte keine 3D-Struktur vorhanden sein, durch Wahrscheinlichkeiten des Auftretens eines NOE-Kontaktes, die aus Datenbanken stammen, ersetzt werden. GARANT ist Teil des FLYA-Algorithmus (López-Méndez und Güntert 2006) zur automatischen Bestimmung von 3D-Strukturen mittels NMR.

Das in diesem Kapitel beschriebene SIBASA (**S**imulated **B**ased **S**equential **A**ssignment) bestimmt die vollständige Liste der chemischen Verschiebungen von Proteinen und ist im speziellen auf die Zuordnung von HSQC-Spektren (**H**eteronuclear **S**ingle **Q**uantum **C**oherence) (Bodenhausen und Ruben 1980) optimiert worden.

Die Zuordnung von HSQC-Spektren von Proteinen ist unter anderem für das Wirkstoffdesign mittels NMR wichtig, wo solche Spektren mit und ohne ein zu untersuchendes kleines Molekül, dem sogenannten Liganden, aufgenommen werden. Durch Vergleich der beiden Spektren können Signale identifiziert werden, die unter dem Einfluss des Liganden die Position verändern. Diese Signale sind in der Regel Zeichen einer lokalen Interaktion des Liganden mit dem Protein an der Stelle in der 3D-Struktur, die für das Signal im HSQC-Spektrum verantwortlich ist. Eine vollständige Zuordnung des Spektrums ist notwendig, um jedem HSQC-Signal einen Ort in der Struktur zuzuweisen zu können. Mithilfe von Docking Simulationen, die mit Programmen wie HADDOCK (Dominguez et al. 2003; de

Vries, Sjoerd J et al. 2007) und AUTODOCK (Park et al. 2006) durchgeführt werden können, ist es dann möglich durch Angabe der Interaktionsstellen die Auswirkung des Liganden auf die 3D-Struktur des Proteins zu simulieren und damit die Wirkungsweise des Stoffes zu beurteilen. Ein weiteres Einsatzgebiet ist die Untersuchung der Druckabhängigkeit der chemischen Verschiebungen von Proteinen. Um die Druckabhängigkeit der chemischen Verschiebungen herleiten und durch Hochdruck begünstigte konformationelle Änderungen in der Struktur erkennen zu können werden Druckreihen von HSQC-Spektren aufgenommen. Generell sind Problemstellungen häufig anzutreffen, bei denen eine Röntgenstruktur zu einem zu untersuchenden Proteinen vorhanden ist und ein HSQC-Spektrum zugeordnet werden soll. In vielen Fällen ist neben der Röntgenstruktur auch schon eine ungefähre Teilzuordnung des HSQC-Spektrums bekannt, die als zusätzliche Information für die automatische Zuordnung dienen kann. Bei Titrationsreihen und Druckreihen von HSQC-Spektren ist meistens die Zuordnung des Basisspektrums bekannt. Soll das HSQC-Spektrum einer Punktmutante zugeordnet werden, so steht oft bereits das zugeordnete HSQC-Spektrum des Wildtyps zur Verfügung. Vor allem für solche Probleme lohnt sich die automatische Zuordnung, da sie mit den bekannten chemischen Verschiebungen schneller arbeitet und zuverlässigere Ergebnisse erzeugt.

Idealerweise könnte man die chemischen Verschiebungen aus der 3D-Struktur direkt berechnen. Ansätze dafür liefern Softwarepakete wie SHIFTS (Osapay und Case 1991; Case 1995; Facelli und Dios 1999; Xu und Case 2001, 2002), SHIFTX (Neal 2003) und dessen Nachfolger SHIFTX2 (Han et al. 2011). Diese Programme gehen von den chemischen Verschiebungen ungefalteter Peptide aus, den sogenannten Random-Coil-Verschiebungen, die z.B. in Schwarzingler et al. (2000) bestimmt worden sind. Die Abhängigkeit der Random-Coil-Verschiebungen von der Primärsequenz wurde in Schwarzingler et al. (2001) untersucht und die Abhängigkeiten von Temperatur und pH wurde in Kjaergaard et al. (2011) diskutiert. Zusätzlich werden von den Vorhersageprogrammen der Einfluss der 3D-Struktur auf die chemische Verschiebung berechnet und zum Random-Coil-Wert hinzugezählt. SHIFTX und SHIFTX2 bestimmen diesen Einfluss mithilfe von Datenbanken, die aus den chemischen Verschiebungen und 3D-Strukturen bereits gelöster Proteinen erstellt worden sind. SHIFTX2 bezieht zusätzlich, falls vorhanden, die Liste der chemischen Verschiebungen homologer Proteine mit ein. Obwohl sich die Genauigkeit der Vorhersagen stetig verbessert, sind die Ergebnisse noch lange nicht gut genug um NMR-Spektren verlässlich zuzuordnen zu können.

Der erster Schritt für die Strukturaufklärung und die Zuordnung von HSQC-Spektren ist normalerweise die Auswertung von Triple-Resonanz-Spektren (Cavanagh 2007), die entweder manuell oder über eine der oben genannten Automatisierungsmethoden zugeordnet werden müssen. Da aber das dafür benötigte  $^{13}\text{C}$  isotopenmarkierte Protein relativ teuer ist, konzentriert sich der Ansatz von SIBASA auf 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren, deren Projektionen sowohl die zuzuordnenden HSQC-Spektren, als auch den linken Teil eines 2-D-NOESY (Nuclear Overhauser Effect Spectroscopy) Spektrums beinhalten. Das 2-D-NOESY-Spektrum, das ebenfalls als Eingabespektrum für SIBASA dient, hat in der Regel nicht nur ein

hervorragendes Signal-zu-Rausch-Verhältnis, sondern liefert ein eindeutiges Netzwerk von Wechselwirkungen zwischen nahezu allen Protonen des Proteins, das außerdem mit einer vorhanden Röntgenstruktur nachvollzogen werden kann.

Das hier beschriebene automatische Zuordnungsmodul SIBASA basiert auf dem ASSIGN-Algorithmus (Brunner 2006), das für die vorliegende Arbeit weiterentwickelt und angepasst wurde. ASSIGN basiert auf dem Top-down-Ansatz, d.h. Ausgangspunkt ist neben der Primärsequenz auch die nicht notwendigerweise perfekte 3D-Struktur des Proteins. Das AUREMOL Modul RELAX (Görler und Kalbitzer 1997; Ried et al. 2004) rechnet aus diesen Daten das 2-D-NOESY-Spektrum des Proteins zurück. RELAX basiert auf dem vollständigen Relaxationsmatrixformalismus und kann die Signalvolumen, Linienbreiten, Linienformen und Multiplettstrukturen der NMR-Signale berechnen. ASSIGN fand die richtige vollständige Zuordnung, indem es eine Anfangsliste von chemischen Verschiebungen, die auch zufällig gewählt werden konnte, so optimierte, dass die Abweichung des zurückgerechneten und des experimentelles 2-D-NOESY-Spektrums minimal wurde. Diese Prozedur konnte auch iterativ vorgenommen werden. Dabei wurden die NOE-Signale des 2-D-NOESY-Spektrums mit der optimierten Liste der chemischen Verschiebungen zugeordnet und anschließend eine neue genauere 3D-Struktur bestimmt. Die resultierende Struktur diente als Eingabe für die Rückrechnung der nächsten Iteration von ASSIGN. ASSIGN war bereits in der Lage mithilfe des AUREMOL Moduls für die Homologie-Modellierung PERMOL (Brunner 2006) aus der Struktur und der Zuordnung von HPr *S. aureus* (wt) die Struktur und die nahezu vollständige Zuordnung der Mutante H15A zu ermitteln.

SIBASA wurde an vielen Stellen angepasst und erweitert, die Geschwindigkeit erhöht und die Verlässlichkeit entscheidend verbessert. SIBASA ist nun in der Lage neben dem 2-D-NOESY-Spektrum gleichzeitig ein 3D <sup>15</sup>N-NOESY-HSQC-Spektrum zuzuordnen. SIBASA enthält zudem erstmals auch ein Bewertungssystem, das es erlaubt die automatisch ermittelten chemischen Verschiebungen und HSQC-Zuordnungen zu beurteilen.

## 2 Theoretische Grundlagen

### 2.1 Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin

#### 2.1.1 Allgemeine Herangehensweise

Im Idealfall können Protonen in starren Peptide mit gut bestimmter 3D-Struktur mithilfe ihres NOE-Musters stereospezifisch zugeordnet werden. Bei den Amidprotonen des Asn betrachtet man im einfachsten Fall die NOE-Kontakte der Asn  $H^{\delta 2}$ - mit den  $H^{\beta}$ -Protonen, beim Gln die Kontakte der Gln  $H^{\epsilon 2}$ - mit den  $H^{\gamma}$ -Protonen. Dem  $H^{\delta 21}$  des Asn bzw. dem  $H^{\epsilon 21}$  des Gln wird dabei das Stärkste der betrachteten NOESY-Kreuzsignale zugeordnet (McIntosh et al. 1997). Eine allgemeinere Herangehensweise ist dagegen der Vergleich des zurückgerechneten mit dem gemessenen NOE-Muster. Für Rückrechnung muss eine stereospezifische Zuordnung der chemischen Verschiebungen vorgegeben sein. Durch das Permutieren der chemischen Verschiebungen der nicht eindeutigen Zuordnungen ändert sich auch das Gesamtmuster im zurückgerechneten Spektrum. Man kann davon ausgehen, dass man die richtige stereospezifische Zuordnungen vorgegeben hat, wenn die Übereinstimmung zwischen experimentellen und zurückgerechneten Spektrum am höchsten ist. Random-Coil-Peptide besitzen im Gegensatz zu gefalteten Proteinen keine eindeutige 3D-Struktur, aber es können dennoch akkurate Spektren simuliert werden, falls über alle vorkommenden Konformationen in der NMR-Probe gemittelt wird. Ein solches Ensemble erhält man durch Simulation der Peptide in Lösung über einen längeren Zeitraum. Das Peptid durchläuft im Laufe dieser Simulation alle Zustände, die auch im Experiment für die beobachtbaren NOEs verantwortlich sind. Aus der resultierenden Trajektorie erhält man neben dem Ensemble des Peptids in Lösung auch dessen kinetische Parameter unter Experimentbedingungen. Für genaue Rückrechnungen fehlen nur noch die chemische Austauschraten, die aber wegen der langen Austauschzeiten nicht von der Molekulardynamik erfasst werden können. Zwei Prozesse sollten berücksichtigt und als zusätzliche Parameter in die Simulation der Spektren gesteckt werden, nämlich der Austausch der Amidprotonen mit dem Lösungsmittel und der  $180^\circ$ -Flip der  $NH_2$ -Gruppen der Seitenketten um die partielle  $\pi C=N$  Doppelbindung. Für Asn bzw. Gln sind das die Flips um die Bindungen  $C^{\gamma}-N^{\delta 2}$  bzw.  $C^{\delta}-N^{\epsilon 2}$ . Während der Austausch von Amidprotonen mit dem Lösungsmittel von der MD nicht unterstützt wird und deshalb die Raten selbst bestimmt werden müssen, sind Flip-Austausche prinzipiell auch in Molekulardynamik-Trajektorien beobachtbar und würden in die bestimmten kinetischen Parameter mit einfließen. Da aber solche Austausche sind in der Regel nicht häufig genug vorkommen (ungefähr alle 5 s was einer Austauschrate von 0,2 Hz entspricht (JURANIĆ et al. 1998) ) um in kurzen MD-Simulationszeit erfasst zu werden, müssen sie separat bestimmt

werden. Eine Möglichkeit diese Raten zu bestimmen wird im weiteren Verlauf des Kapitels noch erläutert.

### 2.1.2 Simulation der NOESY-Spektren

Die bisherigen Versionen von AUREMOL RELAX konnten den chemischen Austausch und dessen Effekt auf NOE Kontakte nicht behandeln. Die neueste Version wurde deshalb um die Funktion erweitert die Kinetikmatrix  $\mathbf{K}$  (Jeener et al. 1979), die die chemischen Austauschraten für alle Protonenpaare enthält, für die Rückrechnung zu berücksichtigen. Die chemische Formel eines Zweizustandsaustausch ist in Gl. 1 gegeben.



Die Kinetikmatrix  $\mathbf{K}$  für einen solchen Austausch sieht folgendermaßen aus:

$$\mathbf{K} = \begin{pmatrix} -k_{AB} & k_{BA} \\ k_{AB} & -k_{BA} \end{pmatrix} \quad (2)$$

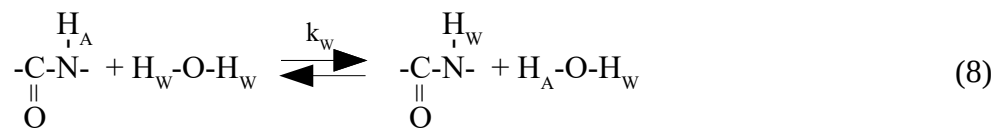
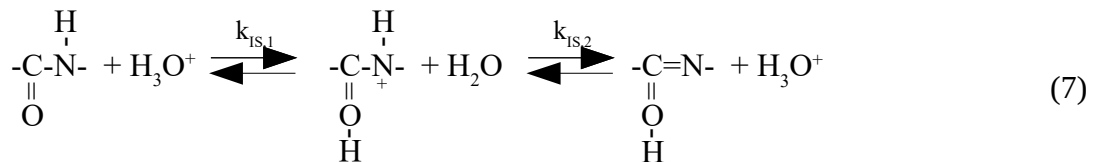
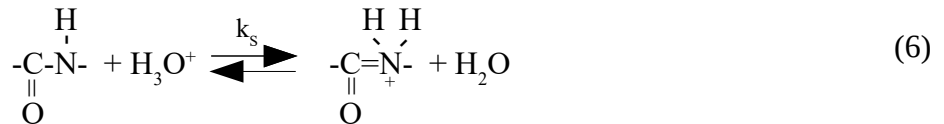
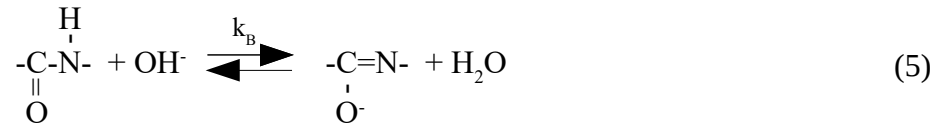
Dabei sind A und B die beiden Zustände und  $k_{AB}$  und  $k_{BA}$  die Austauschraten in beide Richtungen. Die Dynamikmatrix  $\mathbf{D}$  ist die Differenz der Relaxationsmatrix und der Kinetikmatrix:

$$\mathbf{D} = \mathbf{R} - \mathbf{K} \quad (3)$$

Für einen einfachen Austausch zwischen zwei Zuständen sieht sie folgendermaßen aus:

$$\mathbf{D} = \begin{pmatrix} \rho_A + k_{AB} & \sigma_{BA} - k_{BA} \\ \sigma_{AB} - k_{AB} & \rho_B + k_{BA} \end{pmatrix} \quad (4)$$

RELAX kann mit zwei Arten des chemischen Austausches umgehen, nämlich den Austausch von Amidprotonen mit dem Lösungsmittel und einem Positionsaustausch zweier Protonen im Molekül. Der erstgenannte Austausch kann auf vier verschiedene Arten stattfinden, nämlich basekatalytisch, säurekatalytisch, über eine imidische Säure und direkt mit dem Lösungsmittel. Die Reaktionsgleichungen für den Austausch eines Hauptkettenamids mit einem Wasserproton sehen vereinfacht so aus (Eriksson et al. 1995):



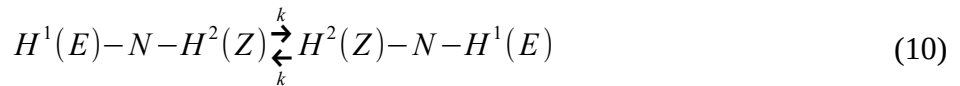
Der direkte Austausch (Gl. 8) spielt in der Regel eine geringe Rolle.

Wie bei Jeener et al. (1979) beschrieben ist, stellen diese Vorgänge chemische Austauschreaktionen, bei denen die Partner stark unterschiedliche Konzentrationen haben, dar. Mit den Annahmen, dass der Anteil des gelösten Biomoleküls  $x_A \ll 1$  und der Anteil des Lösungsmittels  $x_B \approx 1$  ist und mit  $k_{AB} = x_B k$  bzw.  $k_{BA} = x_A k$ , sind die Signalvolumen des Biomoleküls weder vom Diagonalelement des Lösungsmittels noch von den Nicht-Diagonalelementen des Biomoleküls mit dem Lösungsmittel abhängig. Das gleiche gilt auch für die Nicht-Diagonalelemente der dipolaren Wechselwirkung  $\sigma_{AB} = x_B \sigma$  und  $\sigma_{BA} = x_A \sigma$  und dessen Diagonalelement  $\rho_B$ . Außerdem rechnet AUREMOL RELAX keine Signale für das Lösungsmittel zurück, da diese in der Regel nicht benötigt werden, die entsprechenden Strukturdaten in der Regel nicht vorhanden sind und die zurückgerechneten Signale wegen Verfahren wie Wasserunterdrückung nicht mit dem Experiment vergleichbar wären. Deshalb werden in der Relaxationsmatrix alle Elemente vernachlässigt, die das Lösungsmittel direkt oder die Interaktion mit ihm beschreiben. Zur Berechnung des chemischen Austausches des Proteins benötigt man nur noch das Diagonalelement des betroffenen Protons, das den Abfluss von Magnetisierung an die Umgebung beschreibt. Es berechnet sich folgendermaßen:

$$K_{ii} = \rho_i + k \quad (9)$$

Alle anderen Elemente der Dynamikmatrix bleiben vom Austausch unbeeinflusst. Die Austauschraten  $k$  können interaktiv vom Benutzer angegeben werden. Eine Möglichkeit diese Raten zu bestimmen wird in Bai et al. (1993) beschrieben.

Der Positionsaustausch zweier Kerne im Molekül findet z.B. beim Flip einer  $\text{NH}_2$ -Gruppe statt:



Beide Protonen werden von AUREMOL RELAX berücksichtigt. Deshalb müssen vier Elemente in der Dynamikmatrix angepasst werden:

$$\begin{aligned} D_{ii} &= \rho_i + k \\ D_{jj} &= \rho_j + k \\ D_{ij} &= \sigma_{ij} - k \\ D_{ji} &= \sigma_{ji} - k \end{aligned} \quad (11)$$

Die Elemente  $\rho_i$  und der  $\sigma_{ij}$  der Relaxationsmatrix hängen von der Gesamttrotation und internen Beweglichkeit des Moleküls ab. Eine vollständige Beschreibung beider Vorgänge für  $\text{NH}_2$ -Gruppen kann in JURANIĆ et al. (1998) nachgelesen werden. Die beiden Vorgänge können in AUREMOL RELAX zusammen mit der dipolaren Relaxation gleichzeitig behandelt werden, wobei die Kinetikmatrizen  $\mathbf{K}$  der beiden Austauschreaktionen nur addiert werden müssen.

Die Dynamik des Systems wird oft mit dem modellfreien Ansatz (Lipari und Szabo 1982) beschrieben. Aus ihm ergibt sich für die Spektrale Dichte zweier Protonen  $i$  und  $j$ :

$$J_{ij}(\omega) = \frac{2}{5} \left[ \frac{S_{ij}^2 \tau_{ij,rot}}{1 + (\omega \tau_{ij,rot})^2} + \frac{(1 - S_{ij}^2) \tau_{ij}}{1 + (\omega \tau_{ij})^2} \right] \quad (12)$$

Hierbei ist  $S_{ij}^2$  der Ordnungsparameter,  $\tau_{ij,rot}$  die Korrelationszeit der Gesamttrotation im Molekül für isotrope Moleküle und  $\tau_{ij,intern}$  die Korrelationszeit der internen Beweglichkeit. Der Ordnungsparameter  $S_{ij}^2$  quantifiziert anschaulich den Anteil der starren Gesamttrotation des Moleküls an der translationsfreien Gesamtbewegung des Verbindungsvektors zwischen den beiden Protonen  $i$  und  $j$ , d.h. je größer  $S_{ij}^2$  desto ist, desto starrer ist die Verbindung zwischen den Protonen  $i$  und  $j$ .  $\tau_{ij}$  wird folgendermaßen berechnet:

$$\tau_{ij}^{-1} = \tau_{ij,rot}^{-1} + \tau_{ij,intern}^{-1} \quad (13)$$

Clore et al. (Clore et al. 1990) haben gezeigt, dass eine dritte Beweglichkeit eingeführt



werden muss, um eine bessere Beschreibung von Relaxationsprozessen im Peptid zu erhalten. Die interne Beweglichkeit wird in eine langsame interne Beweglichkeit mit der Korrelationszeit  $\tau_{ij,slow}$  und eine schnelle Beweglichkeit mit Korrelationszeit  $\tau_{ij,fast}$  und dem Ordnungsparameter  $S_{ij,fast}^2$  aufgeteilt. Falls  $\tau_{ij,fast}$  viel kleiner als  $\tau_{ij,slow}$  ist, ergibt sich die zugehörige spektrale Dichte des Kernpaares  $ij$  zu:

$$J_{ij}(\omega) = \frac{2}{5} \left[ \frac{S_{ij}^2 \tau_{ij,rot}}{1 + (\omega \tau_{ij,rot})^2} + \frac{(S_{ij,fast}^2 - S_{ij}^2) \tau_{ij}}{1 + (\omega \tau_{ij})^2} \right] \quad (14)$$

Dabei gilt diesmal:

$$\tau_{ij}^{-1} = \tau_{ij,rot}^{-1} + \tau_{ij,slow}^{-1} \quad (15)$$

### 2.1.3 Bestimmung der chemische Austauschraten

Da NOE-Intensitäten und Linienbreiten auch von Austauschprozessen abhängen, dürfen deren Beiträge in der Rückrechnung nicht vernachlässigt werden. Der chemische Austausch der Haupt- und Seitenkettenprotonen mit dem Lösungsmittel und der Flip-Austausch der  $NH_2$ -Gruppen der Seitenketten müssen für die Rückrechnungen der beiden Tetrapeptide berücksichtigt werden. Erster kann vernachlässigt werden, da die Austauschraten bei pH 3 sehr klein sind und deshalb die Linienformen nicht beeinflussen (Bai et al. 1993). Diese Annahme wurde durch den Vergleich der ohne chemischen Austausch simulierten 1-D-NOESY mit den entsprechenden experimentellen 1-D Linienformen bestätigt. Die Linienbreiten der Hauptkettenamide stimmen in beiden Fällen überein.

Die Kreuzsignale der beiden Protonen in den  $NH_2$ -Gruppen zueinander haben, im Gegensatz zur Rückrechnung des 2-D-NOESY-Spektrums ohne Austausch, positives Signalvolumen. Dies kann nur mit Flipprozessen erklärt werden, was auch in JURANIĆ et al. (1998) bestätigt wurde. Die Magnetisierung des Lösungsmittel spielt generell keine Rolle, da sein Magnetisierungstransfer zum Peptid sowohl über chemischen Austausch als auch dipolare Kopplung vernachlässigbar ist (siehe Kap. 2.1.2). Damit die Kreuzsignale in den Simulationen der Spektren ebenfalls positive Signalvolumen erhalten, muss die Austauschrate  $k$  zu den beiden Diagonalelementen der Amidprotonen  $R_{ii}$ ,  $R_{jj}$  in der Relaxationsmatrix  $\mathbf{R}$  hinzuaddiert und von den beiden Nicht-Diagonalelementen  $R_{ij}$ ,  $R_{ji}$  abgezogen werden. Falsche Relaxationsraten beeinflussen nicht nur die entsprechenden Signale, sondern über Spindiffusion auch alle anderen Kreuzsignalvolumen, insbesondere da die Mischzeit mit 0,3 s für unsere Experimente sehr hoch gewählt wurde.

Die Austauschraten wurden solange variiert, bis das Verhältnis der Volumen zweier Referenzsignale im Experiment und in der Simulation gleich groß waren. Für Asn wurden die beiden Kreuzsignale  $H^{\beta 3}/H^{\delta 21}$  und  $H^{\beta 3}/H^{\delta 22}$  für die Fliprate  $k(C^\gamma-N^{\delta 2})$  als Referenz gewählt, für

Gln waren es die beide Signale  $H^\gamma/H^{\epsilon 21}$  und  $H^\gamma/H^{\epsilon 22}$  für die Rate  $k(C^\delta-N^{\epsilon 2})$  und für die beiden Ala C-Termini waren es die Signale  $H^{21}/H^\alpha$  und  $H^{22}/H^\alpha$ .

### 2.1.4 Vorhersage von chemischen Verschiebungen aus den MD-Trajektorien

Die Ergebnisse der stereospezifischen Zuordnung der  $NH_2$ -Gruppen wurde durch die Programme SHIFTS 4.3 (Osapay und Case 1991; Case 1995; Facelli und Dios 1999; Xu und Case 2001, 2002) und SHIFTX (Neal 2003) bestätigt. Beides sind Programme, die die vollständige Zuordnung eines Biomoleküls nur mit dessen 3D-Strukturen vorhersagen können. Sie addieren hierfür die intrinsischen (Random-Coil) chemische Verschiebung zu einem von der 3D-Struktur abhängigen Beitrag. Während die Random-Coil-Verschiebung für jede Atom-Aminosäure-Kombination gleich ist, hängt der zweite Beitrag von der chemischen Umgebung der jeweiligen Kerns ab und wird aus der PDB Datei berechnet. Die vorhergesagte chemische Verschiebung ergibt sich folgendermaßen:

$$\delta_{predicted} = \delta_{random-coil} + \delta_{structural\ correction} \quad (16)$$

Mit beide Programmen wurde für jede der 1.000 Strukturen der Ensembles von Gly-Gly-Asn-Ala- $NH_2$  und Gly-Gly-Gln-Ala- $NH_2$  eine Zuordnungsliste erstellt. Die 1.000 Einzelvorhersagen einer chemischen Verschiebung wurden mit einem Kerndichteschätzer mit Gauß-Kernel, dessen optimaler Bandbreite  $\left(\frac{4}{3N}\right)^{\frac{1}{5}}\sigma$  (Berwin A. TURLACH) ( N Anzahl der Datenpunkte,  $\sigma$  Standardabweichung der Daten) zu einer Häufigkeitsverteilung zusammengefasst.

### 2.1.5 Entnahme der Dynamik Parameter aus der MD-Simulation

Da MD-Trajektorien zu jedem Zeitpunkt der Simulation sowohl die Position als auch den Impuls für alle Atome im System aufgezeichnet haben, ist es möglich sowohl Ordnungsparameter als auch Korrelationszeiten aus ihnen zu berechnen (Bora und Prabhakar 2009; Villa und Stock 2006). GROMACS stellt zu diesem Zweck eigene Werkzeuge zur Verfügung, deren Anwendung für diesen Fall im folgendem genauer beschrieben wird. Die ersten 5 ns der Trajektorie, die das System ins Gleichgewicht gebracht haben, wurden mit *trjconv* abgeschnitten. Von der verbleibenden Trajektorie wurde anschließend die Translation der Peptide entfernt. Die Dynamikparameter können durch Fits der Korrelationsfunktionen  $C_{ij}(t)$  gewonnen werden. Im allgemeinen gilt:

$$C_{ij}(t) = \langle P_2[\hat{\mu}_{ij}(0) \cdot \hat{\mu}_{ij}(t)] \rangle \quad (17)$$

Dabei ist  $\hat{\mu}_{ij}(t)$  der zeitabhängige Einheitsvektor entlang der Verbindungslinie der Kerne  $i$  und  $j$  und  $P_2$  das Legendre-Polynom der 2-ten Ordnung:

$$P_2(x) = \frac{1}{2}(3x^2 - 1) \quad (18)$$

Im erweiterten modellfreien Ansatz geht man von der Unabhängigkeit der starren Gesamttrotation des Moleküls und der internen Beweglichkeit aus. Daher wird  $C_{ij}(t)$  zu:

$$C_{ij}(t) = C_{ij,rot}(t) C_{ij,intern}(t) \quad (19)$$

wobei

$$C_{ij,rot}(t) = \frac{1}{5} \exp(-t/\tau_{ij,rot}) \quad (20)$$

und

$$C_{ij,intern}(t) = S_{ij}^2 + (S_{ij,fast}^2 - S_{ij}^2) \exp(-t/\tau_{ij,slow}) \quad (21)$$

ist, falls  $\tau_{ij,fast} \ll \tau_{ij,slow}$  (Bora und Prabhakar 2009). Die Unabhängigkeit voneinander wird im erweiterten modellfreien Ansatz auch für die langsame und schnelle interne Beweglichkeit vorausgesetzt. Durch Fouriertransformation der Korrelationsfunktion ergibt sich die Spektrale Dichte  $J_{ij}(\omega)$ , die in Gl. 14 analytisch beschrieben wurde.

Um  $C_{ij,intern}(t)$  für alle Kernpaare zu erhalten, wurde mit *trjconv* eine Trajektorie mit konstantem Gesamtdrehimpuls erzeugt. Das Peptid wurde zu jedem Zeitpunkt an die zeitlich gemittelte Struktur, die mit *g\_rmsf* erzeugt wurde, gefittet um die Gesamttrotation zu entfernen. Die numerische Umsetzung von Gl. 21 wurde mit *g\_rotacf* und folgenden Optionen erzeugt:

- -P 2 für Legendre-Polynome der 2-ten Ordnung
- -noaver damit für jedes Paar  $ij$  eine Korrelationsfunktion und keine gemittelte Korrelationsfunktion für alle Kernpaare erzeugt wird
- -d um anzugeben, dass die Korrelationsfunktionen für Paare aus jeweils 2 Kernen gebildet werden sollen

Außerdem benötigt *g\_rotacf* eine Indexdatei, die die Kernindexpaare  $ij$  mit GROMACS Atomindices für alle zu erstellenden  $C_{ij}(t)$  auflistet. Zur Automatisierung dieses Prozesses wurde ein neuer Dialog für die Erzeugung von Metafiles in AUREMOL implementiert, der

unter anderem solche Index-Dateien erstellen kann. Mit einer ebenfalls neuen AUREMOL Funktion wurde die erste ns aller  $C_{ij, \text{intern}}(t)$  der Peptide an Gl. 21 gefittet und die resultierenden Werte  $S_{ij}^2$ ,  $S_{ij, \text{fast}}^2$  und  $\tau_{ij, \text{slow}}$  im AUREMOL Metafile als Matrizen abgelegt. AUREMOL verwendet hierfür die Fitroutine von ALGLIB (Sergey Bochkonov). AUREMOL RELAX liest für seine Rückrechnung das erzeugte Metafile ein und kann somit auf die bestimmten Dynamikparameter zurückgreifen.

Zur Berechnung der Relaxationsmatrix  $\mathbf{R}$  fehlt nur noch die Korrelationszeiten der starren Gesamtrotation  $C_{ij, \text{rot}}(t)$ :

$$C_{ij, \text{rot}}(t) = \frac{C_{ij}(t)}{C_{ij, \text{intern}}(t)} \quad (22)$$

Die Korrelationsfunktion der gesamten Bewegung  $C_{ij}(t)$  werden ebenfalls mit *g\_rotacf*, mit den selben Parametern und mit der selben Indexdatei aus der translationsfreien Trajektorie erzeugt. Die mit Gl. 22 berechneten  $C_{ij, \text{rot}}(t)$  werden mit AUREMOL an Gl. 20 gefittet und die Werte von  $\tau_{ij, \text{rot}}$  im Metafile gespeichert. Eine  $\tau_{ij, \text{rot}}$ -Matrix hat den Vorteil, dass sie im Gegensatz zu einem einzelnen Wert von  $\tau_{\text{rot}}$  die Anisotropie des Moleküls mit berücksichtigt. Ein einzelner Wert kann die starre Rotation nur für sphärische Moleküle gut beschreiben. Neben der  $\tau_{ij, \text{rot}}$ -Matrix bietet AUREMOL RELAX auch noch eine weitere Anisotropiekorrektur für nicht kugelförmige Moleküle an (Görler und Kalbitzer 1997).

Da die in der Molekulardynamik-Simulation verwendeten Lösungsmittelmodelle nicht die gleiche Viskositäten wie die reale Lösungsmittel haben (Mark und Nilsson 2001), müssen die rotatorischen Korrelationszeiten  $\tau_{ij, \text{rot}}$  korrigiert werden. Mithilfe des GROMACS Tool *g\_msd* wurde dafür zuerst der Selbstdiffusionskoeffizient  $D_T$  des H<sub>2</sub>O/D<sub>2</sub>O-Lösungsmittelmodell bestimmt:

$$D_{T, \text{sim}} = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle |r(t) - r(0)|^2 \rangle \quad (23)$$

Er wurde über alle Sauerstoffatome in der Lösung gemittelt. Für das SPC/E/SPCHW-Mischmodell ergibt sich  $D_{T, \text{sim}}$  zu  $2,0496 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$ . Der korrespondierende experimentelle Wert kann durch Interpolation aus den in Longworth (1960) aufgelisteten experimentellen Werte des Selbstdiffusionskoeffizienten für H<sub>2</sub>O/D<sub>2</sub>O-Gemische bestimmt werden. Aus den Werten bei 278, 298 und 318 K ergibt sich  $D_{T, \text{exp}}$  bei 293 K zu  $1,9640 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$ .

Der Zusammenhang von  $\tau_{\text{rot}}$  eines gelösten Moleküls mit dem Selbstdiffusionskoeffizienten des Lösungsmittels wird durch die Debye-Stokes-Einstein-Gleichung beschrieben:

$$\tau_{\text{rot}} = \frac{V \eta}{k_B T} \quad (24)$$

Der Selbstdiffusionskoeffizient  $D_T$  ergibt sich dann folgendermaßen:

$$D_T = \frac{k_B T}{6\pi\eta r_H} \quad (25)$$

wobei  $\eta$  die Viskosität des Lösungsmittels,  $r_H$  dessen hydrodynamischer Radius und  $V$  das Volumen des Moleküls ist. Kombiniert man Gl. 24 und Gl. 25, so ergibt sich für  $\tau_{rot}$ :

$$\tau_{rot} = \frac{V}{6\pi r_H D_T} \quad (26)$$

Aus der folgenden Relation kann ein Korrekturfaktor  $C$  für  $\tau_{rot}$  bestimmt werden:

$$\frac{\tau_{rot,exp}}{\tau_{rot,sim}} = \frac{D_{T,sim}}{D_{T,exp}} \quad (27)$$

$$C = \frac{D_{T,sim}}{D_{T,exp}} \cdot \tau_{rot,sim} = 1,04 \tau_{rot,sim}$$

Diese Korrektur ist im Gegensatz zu der Korrektur, die für das weit verbreiteten Tip3p-Wassermodell (Jorgensen et al. 1983) benötigt wird, ziemlich gering. Das Tip3p-Modell benötigt einen Korrekturfaktor von ca. 2 (Mark und Nilsson 2001; Ferner et al. 2008) und ist somit nicht geeignet die Dynamik eines Peptidsystems zu beschreiben.

## 2.2 Automatische sequentielle Zuordnung (SIBASA)

### 2.2.1 Vorbereitung

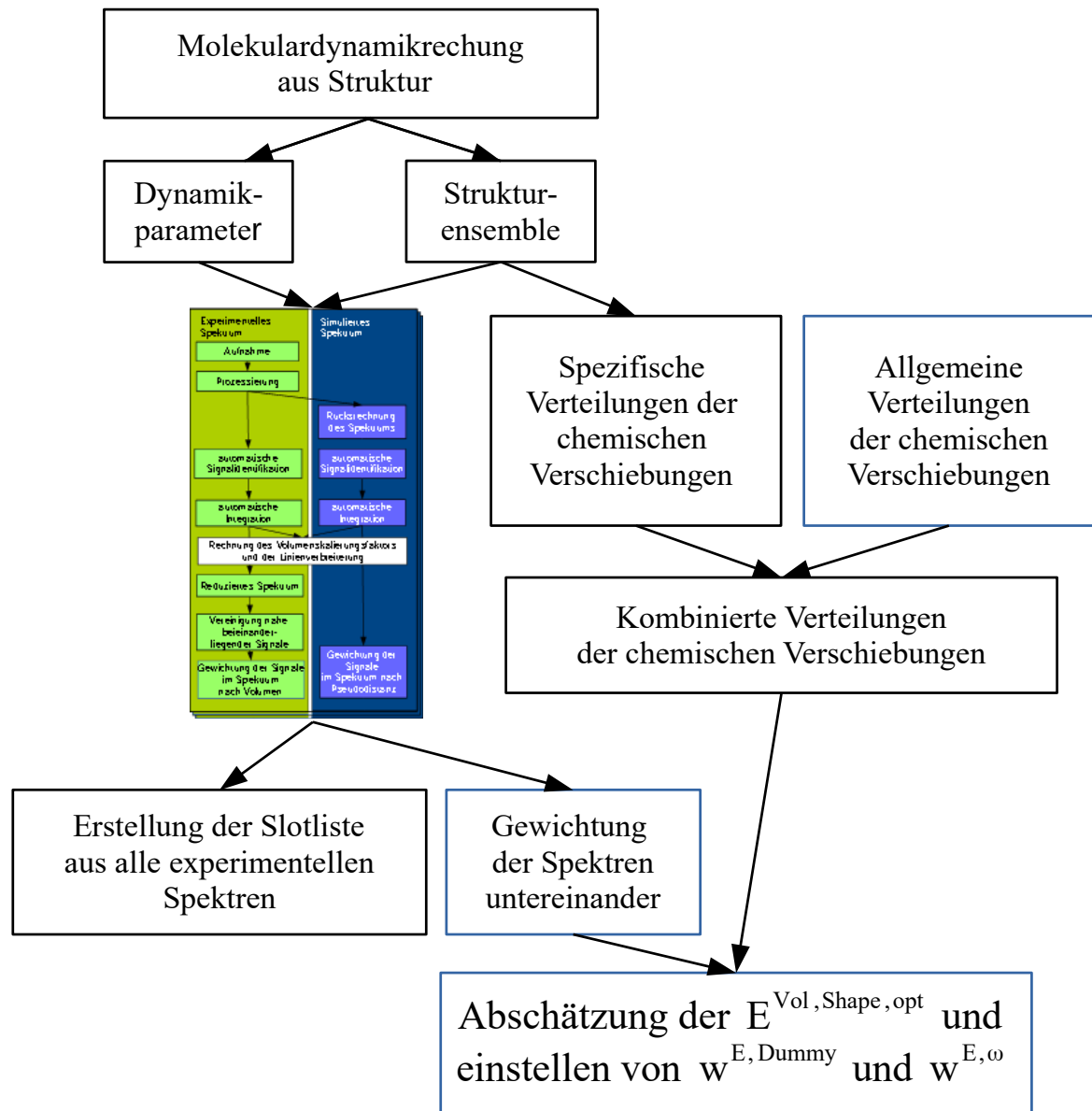


Abbildung 2: Nötige Vorbereitungen und Vorverarbeitungen für die automatische Zuordnung. Abb. 3 zeigt die nötigen Schritte für die Verarbeitung der experimentellen Daten und die Details der Rückrechnung (**Grün, Blau**) vergrößert.  $E^{\text{Vol, Shape, opt}}$  steht für den geschätzten Wert der Pseudoenergie des Volumens und der Linienform der richtigen Zuordnung (Gl. 65),  $w^{\text{E, Dummy}}$  für den Gewichtungsfaktor der Pseudoenergie des Dummyslots (Kap. 2.2.9) und  $w^{\text{E, } \omega}$  für den Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung (Gl. 63 und 64).

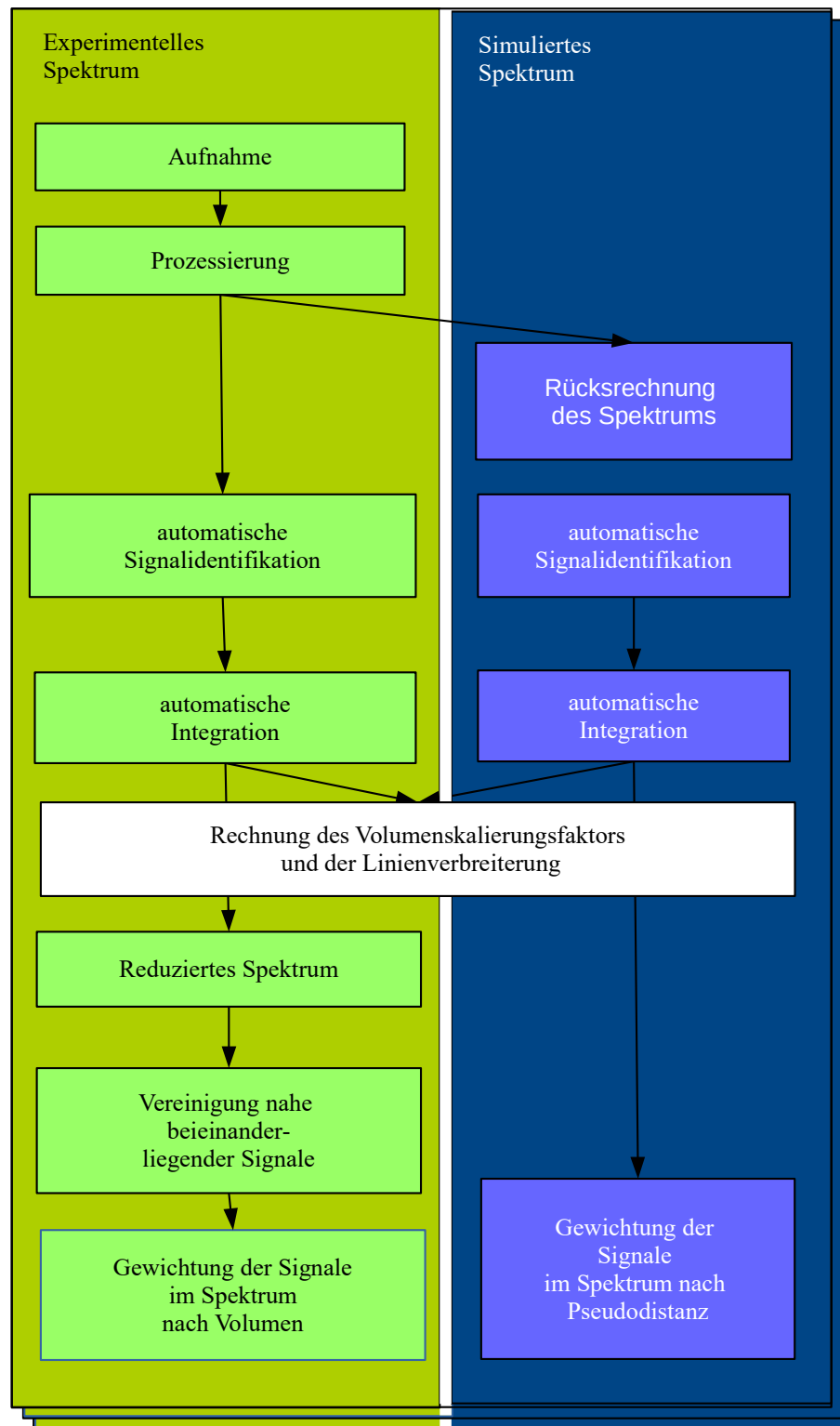


Abbildung 3: Schematische Darstellung der benötigten Schritte zur Vorbereitung eines Spektrums für die automatische Zuordnung. (Grün) Verarbeitung eines experimentellen NMR-Spektrums. (Blau) Rückrechnung des entsprechenden NMR-Spektrums.

SIBASA kann als Eingabe ein Strukturensemble, das mit dem dem Molekulardynamik (MD) Softwarepaket GROMACS (Berendsen et al. 1995; Hess et al. 2008) aus einer Röntgenstruktur erzeugt wurde, verarbeiten. Wie in Kap. 2.1.2 und in der Veröffentlichung von Harsch et al. (2013) verwendet SIBASA eine neue Version von AUREMOL RELAX, die diese großen Strukturensembles einlesen und Ordnungsparameter und Korrelationszeiten aus der MD-Trajektorie verwenden kann. Außerdem werden für SIBASA Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der zu bestimmenden Resonanzen im Vorfeld aus der MD-Trajektorie geschätzt. Die Vorbereitungen und Vorverarbeitungen der Eingangsdaten für den SIBASA Algorithmus sind schematisch in Abb. 2 gezeigt.

## 2.2.2 Vorbereitung der experimentelle Spektren

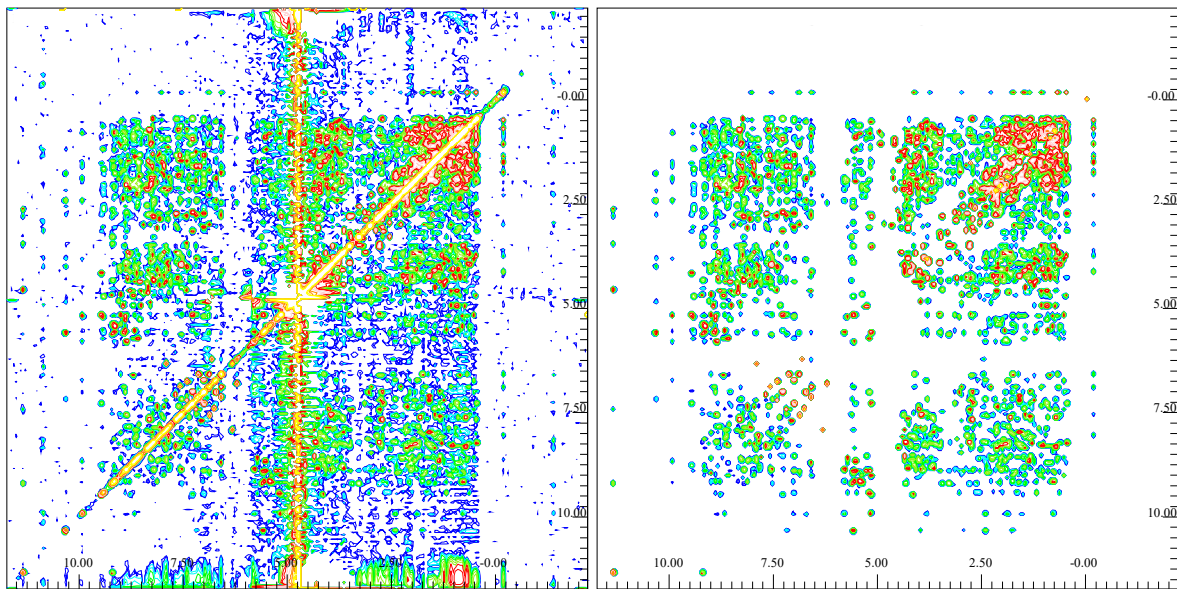


Abbildung 4: **Experimentelles und reduziertes 2-D-NOESY-Spektrum der HPr *S. aureus* Mutante H15A.** Das reduzierte Spektrum ist nahezu rausch- und artefaktfrei. Wegen dem starken Überlapp wurde die Diagonale ebenfalls entfernt.

Zusätzlich zum homonuklearen 2-D-NOESY-Spektrum, das ASSIGN als einziges Spektrum verwendet konnte, ist SIBASA für die gleichzeitige Auswertung eines zusätzlichen 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums ausgelegt. Damit das gesamte NOE-Kopplungsmuster des gemessenen Proteins möglichst vollständig zugänglich ist, ist es notwendig, die beiden Eingabespektren auf möglichst gutes Signal-zu-Rausch-Verhältnis zu prozessieren. Dabei ist für die automatische Zuordnung wichtig, dass richtige Signale von Artefakten und Rauschmaxima unterscheidbar sind. Als Filterfunktion stehen dem Benutzer zusätzlich zum Exponentialfilter (EM), den schon ASSIGN auswerten konnte, jetzt auch die aus Topspin



(Bruker Biospin, Karlsruhe) bekannte Lorenz-Gauß-Transformation (GM) und der Cosinus Filter (QSINE mit SSB 1) zur Verfügung. Im Anschluss an die Prozessierung der Spektren müssen die Signale identifiziert werden, was automatisch mit dem verbesserten BAYESSCHEN PEAKPICKING Modul von AUREMOL (Antz et al. 1995) durchgeführt werden kann. In der anschließend Korrektur der automatisch generierten Signalliste sollte der Benutzer die irrtümlich erkannten Rauschsignale und Artefakte, die z.B. vom Lösungsmittel herrühren, entfernen und nicht erkannte Signale hinzufügen. Anschließend müssen die Signalvolumen mit der AUREMOL Integration (Geyer et al. 1995) bestimmt werden. Mit ihnen kann SIBASA die Skalierungsfaktoren zwischen den experimentellen und den entsprechenden zurückgerechneten Spektren berechnen (Kap. 2.2.4). Außerdem bestimmt die Integrationsroutine die für die Berechnung der lokalen Pseudoenergien der Optimierung wichtigen Signalgrundflächen (Kap. 2.2.8) und damit den spektralen Bereich, der für die automatische Zuordnung wichtig ist. Die Verbesserte Bestimmung der Grundflächen ist Teil der neuen AUREMOL Integration (Donaubauer, Veröffentlichung in Vorbereitung), die mit Signalüberlapp besser als die Methode aus ASSIGN umgehen kann. Mithilfe der bestimmten Grundflächen werden auch die von ASSIGN bekannten reduzierten Spektren (Abb. 4) erzeugt. Reduzierte Spektren vereinen die Informationen intelligenter Algorithmen, wie des bereits erwähnten Bayesschen Peakpicking oder der Integration mit Expertenwissen und maximiert damit die für ASSIGN und SIBASA zugänglichen experimentellen Informationen. Reduzierte Spektren kommen dem rausch- und artefaktfreien zurückgerechneten Spektren am nächsten. mit Hilfe des reduzierten Spektrums können außerdem Artefakte, die anormale Grundflächen erzeugen, besser erkannt und somit die Güte des Peakpickings noch einmal verbessert werden.

### 2.2.3 Simulation der NOESY-Spektren

Der Top-down-Ansatz gewinnt für die automatische Zuordnung den Großteil der Mehrinformationen gegenüber dem Bottom-up-Ansatz durch die Auswertung von zurückgerechneten Spektren. In SIBASA wird, eine neue Version von AUREMOL RELAX verwendet. Im Gegensatz zu dem in ASSIGN verwendeten Version von RELAX, wurde die neue Version erweitert und verbessert und kann jetzt noch genauere Rückrechnungen der NOESY-Signale erstellen. Der Benutzer von ASSIGN und SIBASA muss nur folgende Parameter für die Erstellung der Rückrechnungen angeben:

- die Sequenz des Proteins
- die 3D-Struktur oder das 3D-Strukturensemble des Proteins
- die Mischzeiten für NOESY-Spektren
- die Repitionszeiten für NOESY-Spektren

Aus den Eingangsspektren und der MD-Simulation werden die restlichen Parameter

bestimmt:

- die Spektrumtypen
- die Frequenzen der verwendeten NMR-Spektrometer
- die initiale vollständige Zuordnung
- Spektrale Breiten und Offset der Spektren
- Größen der prozessierten Spektren in digitalen Datenpunkten (SI)
- FID-Filter und deren Parameter
- Dynamikparameter wie Korrelationszeiten und Ordnungsparameter
- zusätzlich Linienverbreiterung für jede Dimension und jedes Spektrum, die z.B. durch Feldinhomogenität bedingt sind (siehe Kap. 2.2.5)

Beim Top-down-Ansatz werden die einzelnen simulierten Linienformen mit einer Liste von chemischen Verschiebungen zu NMR-Spektren angeordnet, die anschließend mit den zugehörigen reduzierten Spektren verglichen werden. Aus dem Vergleich kann die Qualität der verwendet Zuordnung ermittelt werden. Gegenüber dem Bottom-up-Ansatz erhält man dadurch einen großen Informationsgewinn:

- Es werden mithilfe des vollständigen Relaxationsmatrixformalismus, der auch die Spindiffusion berücksichtigt, NOESY-Signalvolumen berechnet.
- Die Linienformen werden unter Berücksichtigung der  $^2\text{J}$ -Kopplungsmuster bzw.  $^3\text{J}$ -Kopplungsmuster, deren Kopplungskonstanten aus der AURMEMOL Datenbank bzw. der Auswertung der Karplus-Beziehung (Karplus 1963) und der Diederwinkel der Eingangsstrukturen stammen, bestimmt.
- Die Linienbreiten werden ebenfalls aus den Eingabestrukturen bestimmt und fließen in die Linienformen mit ein.
- Sich überlappende Signale können simuliert und mit den experimentellen Daten verglichen werden.

Die für SIBASA im Rahmen dieser Arbeit weiterentwickelte Version von RELAX bietet folgende Neuerungen:

- Die Rückrechnung von NMR-Spektren, bei denen der Magnetisierungstransfer über skalare Kopplung erfolgt, wie z.B. HSQC-, COSY- und Tripleresonanz-Spektren.
- Die Rückrechnung von 1-D-NOESY-Spektren.
- Ein einfacher Interpreter von Pulsprogrammen um weitere N-dimensionale, heteronukleare Spektren hinzuzufügen zu können.
- Die Berücksichtigung von chemischen Austauschraten sowohl bei der Berechnung der Signalvolumen als auch der Linienbreiten (siehe Kap. 2.1.2).

- Unterstützung großer Strukturensamples, die z.B. aus MD-Trajektorien entnommen werden können.
- Auswertung der Ergebnisse von MD-Simulation zur Bestimmung von Ordnungsparametern und Korrelationszeiten (Kap. 2.1.5 und Bora und Prabhakar 2009; Villa und Stock 2006).
- Das Clore-Szabo-Modell (Clore et al. 1990) wurde implementiert.
- Neue  $^3\text{J}$ -Kopplungsparameter (Schmidt 2007) zur Berechnung der Karplus-Beziehung wurden der AUREMOL Datenbank hinzugefügt.
- Nicht auflösbare J-Kopplungen, die in RELAX wegen ihrer hohen Anzahl und dem hohen Bedarf an Rechenzeit und Arbeitsspeicher ignoriert wurden, werden bei der Berechnung der Linienbreite berücksichtigt. Es gilt:  $LW_{\text{neu}} = LW + \sum_i J_i$ , wobei  $J_i < \frac{LW}{2}$  für alle  $J_i$  gelten muss.
- Die Fouriertransformation, der von Topspin bekannten Lorenz-Gauß-Transformation (GM) und der Cosinusfilters (QSINE mit SSB=1), wurden analytisch mit Mathematica (Wolfram Research, Inc., Champaign, Illinois) berechnet und die Ergebnisse in RELAX implementiert. Mit diesen Analytischen Beziehungen können die Auswirkung der Filter auf die zurückgerechneten Signale in der Frequenzdomäne berechnen werden. Endliche Aufnahmezeit  $t_{\text{aq}}$ , die die Signale normalerweise weiter verbreitern, wurden ebenfalls mit berücksichtigt.
- Separate Filter sind jetzt für alle Dimensionen wählbar.
- Fehlenden chemischen Verschiebungen können mit an die Temperatur (Kjaergaard et al. 2011) und an die Sequenz (Schwarzinger et al. 2000) angepassten Random-Coil-Verschiebungen (Schwarzinger et al. 2001) oder mit chemischen Verschiebungen, die statistisch aus Normalverteilungen ermittelt wurden, ersetzt werden. Die benötigten Mittelwerte und Standardabweichung sind für alle Kerne der essentiellen Aminosäuren in der BMRB (**B**iological **M**agnetic **R**esonance **B**ank) (Ulrich et al. 2008) abrufbar.

## 2.2.4 Bestimmung der Volumenskalierungsfaktoren

Für den absoluten Vergleich von experimentellen und zurückgerechneten Signalvolumen in SIBASA wird der Skalierungsfaktor  $\alpha$  zwischen den beiden Spektren benötigt. Dieser wird im Normalfall wie z.B. bei der NMR-Strukturbestimmung mithilfe eines oder mehrere zugeordneter Referenzsignale berechnet. Zu den NOESY-Referenzsignalen müssen sowohl die experimentellen Volumen als auch die zugehörigen Abstände der Protonen in der 3D-Struktur bekannt sein. Wurde ein NOESY-Spektrum mit kleine Mischzeiten aufgenommen, kann folgende Beziehung ausgenutzt werden:

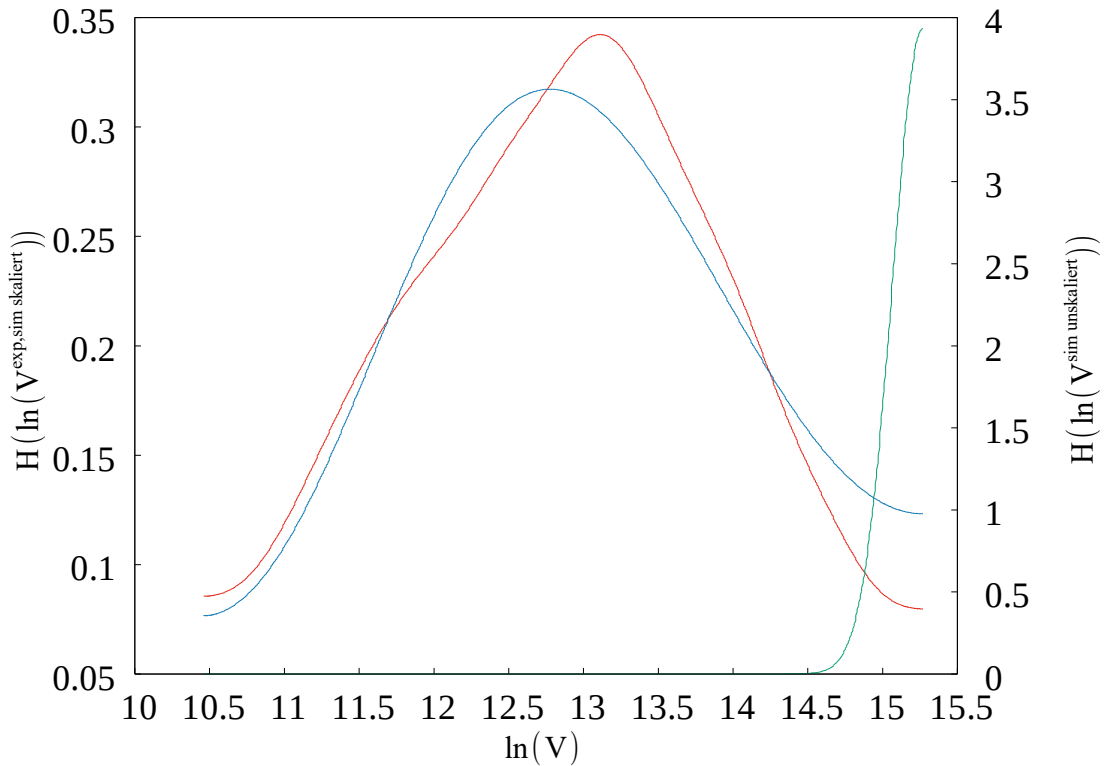


Abbildung 5: **Häufigkeitsverteilungen der logarithmierten Signalvolumen im 2-D-NOESY-Spektrum der HPr *S. aureus* Mutante H15A.** (Rot) Häufigkeitsverteilung  $H(\ln(V^{\text{exp}}))$  der experimentelles Volumen (linke y- Achse), (Grün) Häufigkeitsverteilung  $H(\ln(V^{\text{sim, unskaliert}}))$  der unskalierten simulierten Volumen (rechte y- Achse), (Blau) Häufigkeitsverteilung  $H(\ln(V^{\text{sim, skaliert}}))$  der skalierten simulierten Volumen (linke y- Achse).

$$V_{NOE} \propto d^{-6} \quad (28)$$

Da experimentelle und simulierte Signalvolumen  $j$  von der selben Distanz herrühren, gilt:

$$V_j^{\text{exp}} = \alpha V_j^{\text{sim}} \quad (29)$$

Stehen mehrere zugeordnete Referenzsignale zur Verfügung, so ist es besser den Skalierungsfaktor  $\alpha$  folgendermaßen zu mitteln (Trenner 2006):

$$\bar{\alpha} = \frac{\sum_{j=1}^N V_j^{\text{exp}} V_j^{\text{sim}}}{\sum_{j=1}^N (V_j^{\text{sim}})^2} \quad (30)$$

Eine noch bessere Methode kommt bei AUREMOL KNOWNOE (Gronwald et al. 2002) und AUREMOL REFINE (Trenner 2006) zum Einsatz. Die AUREMOL Module passen die Elemente der Referenzsignale in der Relaxationsmatrix solange iterativ an, bis alle von

RELAX berechneten Referenzvolumen bis auf den Skalierungsfaktor  $\alpha$  mit den experimentellen Referenzvolumen übereinstimmen. Diese Methode kann auch für NOESY-Spektren mit größeren Mischzeiten eingesetzt werden, falls eine 3D-Struktur vorhanden ist.

Wenn, wie im Falle der automatische Zuordnung, keine Resonanzzuordnungen vorher bekannt sind, können keine experimentellen Signale zugeordnet werden und keine der beschriebenen Methoden kann einen Skalierungsfaktor bestimmen. SIBASA greift deshalb auf einen statistischen Ansatz zurück. Es wird davon ausgegangen, dass die Häufigkeitsverteilung der richtig skalierten Signalvolumen im experimentellen und zurückgerechneten Spektrum gleich sind. Für die Bildung der Häufigkeitsverteilungen der Volumen der zurückgerechneten Spektren wird jedes mit einer zufälligen Anfangszuordnung zurückgerechnete Spektrum wie das zugehörige experimentelle Spektrum vorbereitet, d.h. die Intensitätsmaxima werden automatisch ermittelt und integriert (blauer Bereich in Abb. 3). Mithilfe eines Kerndichteschätzers und eines Gauß-Kernel mit optimaler Bandbreite werden Häufigkeitsverteilungen der logarithmierten Volumen der experimentellen und den skalierten simulierten Spektren erzeugt. Um Störungen durch schwache Rauschsignale und starke Artefakten zu vermeiden, werden jeweils die 2,5 % kleinsten und die 2,5 % größten Datenpunkte ausgeschlossen.

Das Signal-zu-Rausch-Verhältnis des simulierten Spektrum ist, Im Gegensatz zum Signal-zu-Rausch-Verhältnis des experimentellen Spektrums, ideal, wodurch schwache Signale, die im Experiment im thermisches Rauschen untergehen oder unter Artefakten verschwinden würden, sichtbar werden. Deshalb werden bei der Erzeugung der Häufigkeitsverteilungen der simulierten Volumen solange die schwächsten Datenpunkte verworfen, bis die Anzahl experimentellen und simulierten Volumen in den Häufigkeitsverteilungen gleich sind. Die Logarithmierung der Volumen glättet die Verteilungen und sorgt außerdem dafür, dass verbleibende Ausreißer in der Statistik nicht übergewichtet werden. Die Verteilungen mit Kerndichteschätzer werden folgendermaßen berechnet:

$$f(V) = \sum_{j=1}^N \frac{1}{\sqrt{2\pi N b}} \exp\left(-\frac{1}{2b^2}(\ln V - \ln V_j)^2\right) \quad (31)$$

$b$  ist die optimale Bandbreite für den Gauß-Kernel (Berwin A. TURLACH):

$$b = \left(\frac{4}{3N}\right)^{\frac{1}{5}} \cdot \sigma \quad (32)$$

$N$  ist die Anzahl der verbliebenen Signale und  $\sigma$  die Standardabweichung der logarithmierten Volumen  $\ln(V_j)$ . Der hier verwendete Kerndichteschätzer ist im Prinzip die Faltung der Datenpunkte mit einer Normalverteilung mit Standardabweichung  $b$ . Die Häufigkeitsverteilungen der logarithmierten skalierten zurückgerechneten Volumen  $\ln(\alpha V_j)$  werden mit verschiedenen Skalierungsfaktoren  $\alpha$  erzeugt. Mit dem  $\chi^2$ -Kriterium (Gl. 33) wird

diejenige Verteilung, die am besten zur Verteilung der experimentellen Volumen passt, ermittelt.

$$\chi^2 = \int \left( f^{\text{exp}}(V) - f_{\alpha}^{\text{sim}}(V) \right)^2 dV \quad (33)$$

In Abb. 5 sind die Häufigkeitsverteilungen der experimentellen, der unskalierten simulierten und der optimal skalierten simulierten logarithmierten Volumen  $\ln(V_j)$  des 2-D-NOESY-Spektrums des HPr *S. aureus* (H15A) abgebildet. Man kann eine gute Übereinstimmung zwischen den Verteilung der experimentellen und der optimal skalierten simulierten Volumen erkennen.

Während der automatischen Zuordnung werden, wie in Kap. 2.2.14 noch beschrieben wird, Skalierungsfaktoren aus bereits sicheren Zuordnungen mit Gl. 30 noch zur Laufzeit korrigiert. Diese Skalierungsfaktoren gehen danach direkt in die Bestimmung der Pseudoenergie der Optimierung ein und können das Ergebnis der automatischen Zuordnung weiter verbessern.

### 2.2.5 Bestimmung der zusätzlichen Linienverbreiterungen

Die Rückrechnung der Linienbreiten der NMR-Signale geht von einem konstantem äußeren Magnetfeld aus und verlässt sich auf die angegebene 3D-Strukturen und Dynamikparameter. In der Realität ist es oft so, dass die experimentelle Linien wegen einem Magnetfeldgradient im NMR-Spektrometer oder wegen der geringen Auflösung in einer indirekten Dimension breiter als in der Theorie sind. Der Wert der sogenannten Feldinhomogenität kann in der Praxis nur für die gut aufgelöste direkte Dimension bestimmt werden. Dafür misst man die Linienbreite des nicht durch J-Kopplungen beeinflussten DSS-Referenzsignals, das in der Regel eine Breite von ca. 1 Hz hat. Die Feldinhomogenität ist die um 1 Hz verminderte in einem hoch aufgelösten 1-D-Spektrum gemessene Linienbreite dieses Signals und kann der Rückrechnung als zusätzliche Linienverbreiterung mitgegeben werden. Die Linienverbreiterung durch geringe Auflösungen wird in der neuen Version von AUREMOL RELAX dagegen direkt berücksichtigt (siehe Kap. 2.2.3).

Eine Hauptursache für falsch berechnete Linienbreiten ist, unabhängig von der Feldinhomogenität, der für die Rückrechnung nur genäherte Wert der rotatorischen Korrelationszeit  $\tau_{\text{rot}}$ . Die Linienbreite hängt größtenteils von den Abständen eines Kerns von den anderen Kernen in der 3D-Struktur und von den spektralen Dichten ab (Cavanagh 2007) (Gl. 34).

$$LW_k \propto \sum_{l=1, l \neq k}^{N_{\text{kerne}}} \frac{1}{d_{kl}^6} \left( 4 J_{kl}(0) + J_{kl}(\omega_k - \omega_l) + 3 J_{kl}(\omega_k) + 6 J_{kl}(\omega_k + \omega_l) \right) \quad (34)$$

Falls die Korrelationszeit der schnellen internen Beweglichkeit viel kleiner als die der

langsamen internen Beweglichkeit ist ( $\tau_{kl,fast} \ll \tau_{kl,slow}$ ), ergibt sich die spektrale Dichte  $J_{kl}(\omega)$  zwischen zwei Kernen im Makromolekül im Clore-Szabo-Modell (Clore et al. 1990) (Kap. 2.1.2) zu:

$$J_{kl}(\omega) = \frac{2}{5} \left[ \frac{S_{kl}^2 \tau_{kl,rot}}{1 + (\omega \tau_{kl,rot})^2} + \frac{(S_{kl,fast}^2 - S_{kl}^2) \tau_{kl,slow}}{1 + (\omega \tau_{kl,slow})^2} \right] \quad (35)$$

Ist die Korrelationszeit der langsamen internen Beweglichkeit viel kleiner als die rotatorische Korrelationszeit ( $\tau_{kl,slow} \ll \tau_{kl,rot}$ ) und gleichzeitig  $\omega \tau_{kl,rot} \gg 1$ , was für interne Bewegungen in Proteinen in den allermeisten Fällen gilt, so kann  $J_{kl}(\omega)$  für alle  $\omega \neq 0$  in der spektralen Dichte aus Gl. 34 vernachlässigt werden und für alle  $\omega = 0$  ist  $J_{ij}(\omega)$  proportional zur rotatorischen Korrelationszeit  $\tau_{ij,rot}$ . Für annähernd sphärische Moleküle nimmt  $\tau_{ij,rot}$  außerdem den konstanten Wert  $\tau_{rot}$  ein. Die rotatorische Korrelationszeit für solche Moleküle wird mit der Stockes-Einstein-Gleichung auf Basis der Viskosität des Lösungsmittels und der Größe des Moleküls bestimmt (Görler und Kalbitzer 1997). Für die Rückrechnungen in SIBASA sollte aber die in Kap. 2.1.5 vorgestellte Methode zur Bestimmung der gesamten Matrix der rotatorischen Korrelationszeiten  $\tau_{ij,rot}$ , die auch auch nicht kugelförmige Moleküle behandeln kann, verwendet werden.

Da sowohl die eingesetzten Kraftfelder und Lösungsmittelmodelle nur Näherungen an die Realität sind, wurde für SIBASA zusätzlich ein Algorithmus implementiert, der die Linienverbreiterung  $LB_{ij}$ , welche der Unterschied zwischen zurückgerechneten und experimentellen Linienbreiten ist, für alle Spektren  $i$  und Dimensionen  $n$  bestimmt (Gl. 36).

$$LW_{ijn}^{exp} = LW_{ijn}^{sim} + LB_{ij} \quad (36)$$

Die zusätzlichen Linienverbreiterungen  $LB_{ij}$  werden während der Vorbereitung von SIBASA berechnet und an die Rückrechnung übergeben. Ein zusätzliches manuelles Messen der Feldinhomogenität ist nicht mehr nötig.

Wie beim Volumenskalierungsfaktor (Kap. 2.2.4) werden Verteilungen der experimentellen und der logarithmierten verbreiterten simulierten Linienbreiten mit einem Kerndichteschätzer und einem Gauß-Kernel erstellt. Um die Vergleichbarkeit gewährleisten zu können, wurden experimentelle und simulierte Linienbreiten mit der gleichen Methode bestimmt, d.h. die simulierten Linienbreiten sind nicht die von RELAX berechneten Linienbreiten, sondern wurden ebenfalls mithilfe der AUREMOL Methode aus dem zurückgerechneten Spektrum bestimmt. Es gelten außerdem die selben Ausschlussregeln und Filter, die schon bei der Bestimmung der Volumenskalierungsfaktoren zum Einsatz kamen (Kap. 2.2.4). Die Verteilungen  $f_{in}(LW)$  der Linienbreiten im Spektrum  $i$  und der Dimension  $n$  ergeben sich wie folgt:

$$f_{\text{in}}(LW) = \sum_{j=1}^{N_i^{\text{Signale}}} \frac{1}{\sqrt{2\pi N} b_{\text{in}}} \exp\left(-\frac{1}{2b_{\text{in}}^2} (\ln LW - \ln LW_{\text{inj}})^2\right) \quad (37)$$

$b_{\text{in}}$  ist wieder die optimale Bandbreite des Gauß-Kernels,  $N_i^{\text{Signale}}$  die Anzahl der Datenpunkte und  $LW_{\text{inj}}$  der  $j$ -te Datenpunkt im Spektrum  $i$  und der Dimension  $n$ . Bei den Verteilungen der Linienbreiten der simulierten Spektren werden statt der Linienbreiten  $LW_{\text{inj}}$  die um  $LB_{\text{in}}$  verbreiterte Werte  $LW_{\text{inj}} + LB_{\text{in}}$  verwendet. Durch die Logarithmierung der Linienbreiten werden, wie schon bei den Häufigkeitsverteilungen der Volumina, die Effekte von Ausreißern gemindert und die Verteilungen geglättet. Die Verbreiterung  $LB_{\text{in}}$  wird wieder von derjenigen Verteilung der verbreiterten simulierten Linienbreiten, für die das kleinste  $\chi^2$  mit der entsprechenden experimentellen Verteilung bestimmt wurde, übernommen.

## 2.2.6 Suchraum der chemischen Verschiebungen (Slots)

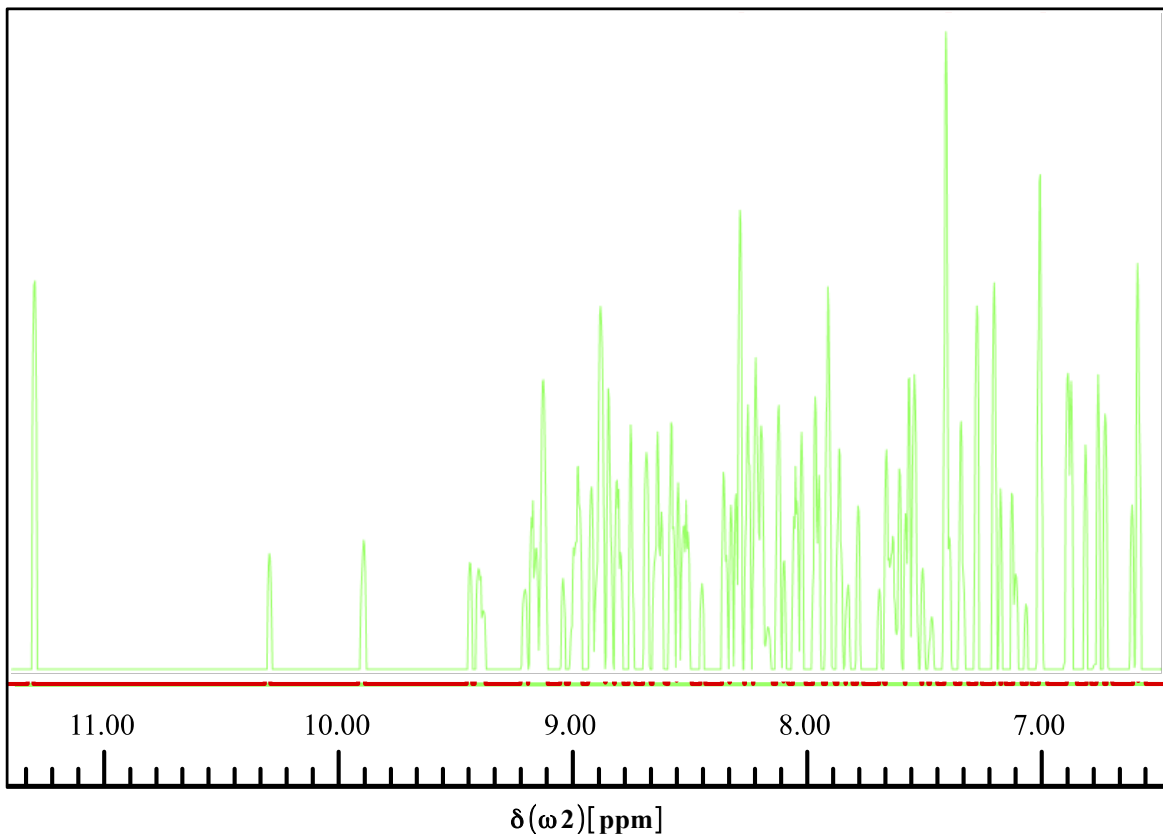


Abbildung 6: Teil der Summenprojektion des auf eine Segmentationshöhe von 0,5 reduzierten 2-D-NOESY-Spektrums von HPr *S. aureus* (H15A) auf die direkte Dimension ( $\omega_2$ ). SIBASA erzeugt für die roten Bereiche keine Slots, weil keine NOESY-Kreuzsignale an diesen Position vorhanden sind.



SIBASA testet im Laufe der automatischen Zuordnung viele verschiedenen chemische Verschiebungen für die Resonanz jedes Kerns, indem es die entsprechenden zurückgerechneten NMR-Signale an die jeweiligen Stellen in den simulierten Spektren verschiebt. Da aber nicht jede chemische Verschiebung im spektralen Bereich für jeden Kern erlaubt oder sinnvoll ist, werden die zugänglichen Bereiche abhängig vom Kern, eingeschränkt. Die sogenannten Slots definieren die möglichen chemischen Verschiebungen und enthalten einen digitalen Datenpunktvektor mit Dimension  $d = \sum_{i=1}^{N_{\text{Spektren}}} n_i$ . In ihm ist die Position einer möglichen Resonanzzuordnung im digitalen Datenpunktraaster für alle Eingangsspektren und deren Dimensionen angegeben. Wenn die Zuordnung in einer Dimension nicht möglich ist, enthält der Vektor den Eintrag -1. Jeder in den Eingangsspektren angeregte Atomtyp, also in der Regel jede Kernart mit unterschiedlichen NMR-Anregungsbereich erhält eine eigene Slotliste, also eine Liste mit erlaubten Bereichen der chemischen Verschiebung.

SIBASA sieht drei Möglichkeiten für die Erstellung der Slotlisten vor:

1. Es werden alle chemischen Verschiebungen erlaubt, die in mindestens einer Dimension eines Spektrums ein Kreuzsignalmaximum schneiden. Diese Methode kam bereits ASSIGN, für die Erstellung seiner  $^1\text{H}$ -Slotliste aus dem 2-D-NOESY-Spektrum, zum Einsatz.
2. Jeder digitale Datenpunkt wird erlaubt.
3. Es werden nur digitale Datenpunkte erlaubt, die mindestens ein experimentelles Signal im Bereich seiner Linienbreite auf halber Höhe schneiden (Abb. 6).

Die 1. Methode eignet sich vor allem für Spektren ohne großen Signalüberlapp wie dreidimensionale oder noch höherdimensionale Experimente. In ihnen sind die Positionen der tatsächlichen Signalmaxima ausreichen um die richtige Lösung zu finden. Die 2. Methode nimmt keine Einschränkung des Lösungsraums vor. Sie ist aber auch die langsamste Methode und kann ohne Einschränkung durch die 3. Methode ersetzt werden. Die 3. Methode wird in SIBASA vor allem bei 2-D-NOESY-Spektren verwendet. In diesen Spektren kann es wegen Signalüberlapp vorkommen, dass das Zentrum eines Signals nicht auf einem lokalem Maximum liegt. Bei Multipletts mit gerader Anzahl von Intensitätsmaxima liegt das gesuchte Signalzentrum sogar auf einem Minimum. Durch geeignete Filterung des NMR FIDs (**F**ree **I**nduction **D**ecay) mit genügend hoher Linienverbreiterung können zwar Multipletts zu einem Signal vereinigt werden oder mit AUREMOL GRAVITY PEAKPICKING (Donaubauer, Veröffentlichung in Vorbereitung) das wirkliche Signalzentrum trotz Überlappungen bestimmt werden, doch würden mit dieser Methode die möglichen chemischen Verschiebungen schon während der Prozessierung eingeschränkt werden. Abb. 6 veranschaulicht den Möglichkeitsraum der Zuordnungen in einem 2-D-NOESY-Spektrum. Es ist ein Teil der Summenprojektion der direkten Dimension eines auf Segmentationshöhe von 0,5 reduzierten 2-D-NOESY-Spektrums abgebildet. Für jeden digitalen Datenpunkt, an dem diese Projektion eine von 0 verschiedene Intensität hat, wird mit der 3. Methode ein Slot erstellt.

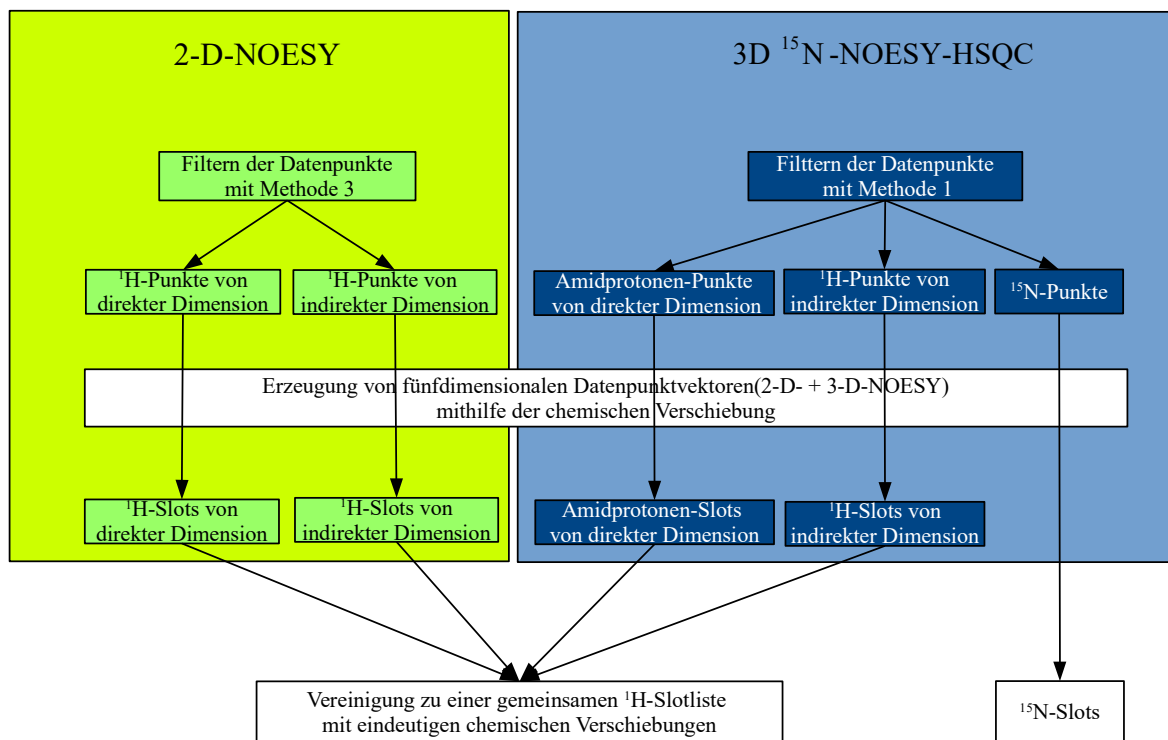


Abbildung 7: **Schematische Darstellung der benötigten Schritte zur Erzeugung der von SIBASA benötigten Slotlisten.** Für die automatische Zuordnung eines 2-D-NOESY- und eines 3D <sup>15</sup>N-NOESY-HSQC-Spektrums wird eine gemeinsame Slotliste für die Protonen <sup>1</sup>H und einer separaten Liste für Stickstoffe <sup>15</sup>N erstellt.

Folgende Schritte sind nötig um die Slotlisten für beliebig viele homo- und heteronukleare Spektren zu erstellen:

1. Sammeln aller aufgenommenen Kernarten, wie z.B. <sup>1</sup>H, <sup>15</sup>N aus den Eingabespektren. Jede Kernart erhält eine eigene Liste von Slots.
2. Sammeln der in Frage kommenden Datenpunkte für alle Kernarten in allen Spektren und Dimensionen mit der geeigneten Methode. Es können unterschiedliche Methoden für verschiedenen Spektren verwendet werden.
3. Berechnung der chemischen Verschiebung in ppm und Erzeugung der digitalen Datenpunktvektoren für alle berechneten chemischen Verschiebungen. Ist eine chemische Verschiebung nicht Teil des spektralen Bereichs einer Dimension oder wird die Kernart in dieser Dimension nicht gemessen, so erhält sie die Koordinate -1.
4. Entfernung von doppelt vorhandenen Slots.
5. Hinzufügen des Dummyslots, für Kerne, die keine sichtbaren Kreuzsignale haben. Er hat überall die Koordinate -1.

In Abb. 7 wird schematisch gezeigt, wie die <sup>1</sup>H- und <sup>15</sup>N-Slotlisten für ein 2-D-NOESY- und ein 3D <sup>15</sup>N-NOESY-HSQC-Spektrum erzeugt werden.

## 2.2.7 Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen

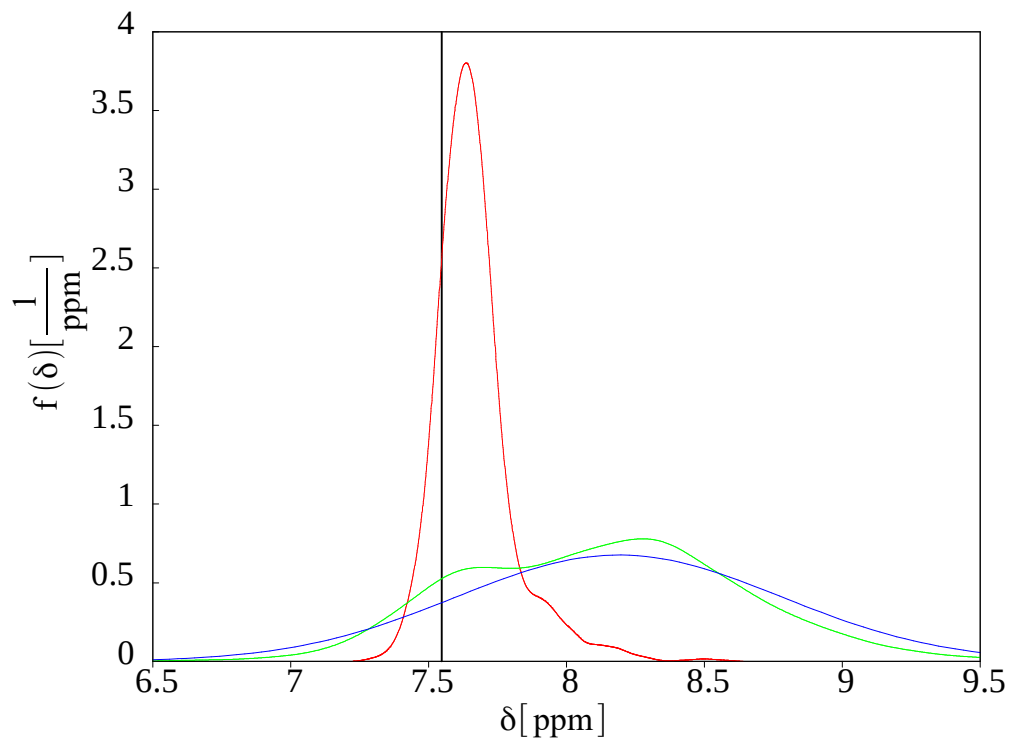


Abbildung 8: **Wahrscheinlichkeitsdichteverteilung der chemischen Verschiebung von Ala59  $H^N$  im HPr *S. aureus* (H15A).** (Rot) Dichteverteilung  $f(\delta)$  der von SHIFTX2 berechneten chemischen Verschiebung aus dem MD-Ensemble, (Grün) Dichteverteilung der von SHIFTS aus der Nh3D-Datenbank berechneten chemischen Verschiebungen, (Blau) Normalverteilung der chemischen Verschiebung aus der BMRB und (Schwarz) korrekte chemische Verschiebung von ALA59  $H^N$  bei 7,548 ppm.

Um die Effizienz der automatischen Zuordnung zu erhöhen, ist es nötig den Suchbereich möglichst effizient einzuschränken, was durch die Analyse der experimentellen Daten und das Erstellen von Slotlisten im vorherigen Kap. 2.2.6 bereits teilweise erreicht wurde. Da dies aber noch nicht genügt um den Optimierungsalgorithmus in angemessener Zeit zum Ziel zu führen, muss der verbleibende Suchraum weiter sinnvoll eingeschränkt und dazu auch noch gewichtet werden. Dafür eignen sich Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen, die allgemein für alle Aminosäuren und Kernarten oder speziell für alle Kerne im Protein aus seinem 3D-Strukturensemble berechnet werden können.

Bei der händischen Zuordnung werden in der Regel Tabellen benutzt, die die minimalen und maximalen chemischen Verschiebung aller Kerne und Aminosäuren auflisten (Cavanagh 2007). Falls zusätzlich die Sekundärstruktur des Proteins bekannt ist, können diese Bereiche noch besser eingegrenzt werden (Wishart et al. 1991). Automatische Zuordnungsalgorithmen

wie z.B. ASSIGN und SIBASA müssen sehr viele Konfigurationen von chemische Verschiebungen testen und bewerten. Sie profitieren ebenfalls von Vorwissen, wie der minimale und der maximale chemischen Verschiebung eines Kerns, aber auch von den Häufigkeiten des Auftretens innerhalb der Grenzen. Wahrscheinlichkeitsdichteverteilungen können den Suchraum gewichten, geben also dem Algorithmus vor, wie er seine Rechenzeit effektiv verteilen kann. Außerdem können gute Dichteverteilungen der chemischen Verschiebungen bei der automatischen Zuordnung die Rolle von Potentialen, die z.B. für die schnelle Konvergenz eines Gradientenverfahrens zur Minimierung verantwortlich sind, übernehmen. Es gibt verschiedene Wege die Wahrscheinlichkeitsdichteverteilungen für SIBASA zu erzeugen.

## Normalverteilungen der chemischen Verschiebungen aus der Biological Magnetic Resonance Bank

Die erste Methode erzeugt Wahrscheinlichkeitsdichteverteilungen in Form von Normalverteilungen, die von der 3D-Struktur des Proteins unabhängig sind. Auf der Webseite der BMRB (Ulrich et al. 2008) kann man sich für die alle Kernarten ( $Kern = H^N, N H^{\alpha}, H^{\beta}, \dots$ ) der essentiellen Aminosäuren AA eine aktuelle Tabelle der durchschnittlichen chemischen Verschiebungen  $\bar{\delta}_{AA, Kern}$  und der Standardabweichungen  $\sigma_{AA, Kern}$  abzurufen. Mit diesen Werten, ist es möglich normierte Wahrscheinlichkeitsdichteverteilungen nach folgenden Muster zu berechnen:

$$f_{AA, Kern}(\delta) = \frac{1}{\sqrt{(2\pi)\sigma_{AA, Kern}}} \exp\left(-\frac{1}{2}\left(\frac{\delta - \bar{\delta}_{AA, Kern}}{\sigma_{AA, Kern}}\right)^2\right) \quad (38)$$

Diese Methode ist wegen der Vollständigkeit der Tabelle immer möglich, selbst wenn mit den anderen Methoden keine Verteilungen erzeugt werden können. Da die Tabelle nicht auf die zuzuordnenden Proteine angepasst ist und mit ihr nur unimodale Gaußverteilungen erzeugt werden können und die Datenbasis der Tabelle außerdem nicht unbiased ist, sollte die Methode nur zur Vervollständigung benutzt werden.

Eine Ausnahme bildet die Optimierung von chemischen Verschiebungen. Sind bereits fast richtige Zuordnungen von z.B einer Punktmutante bekannt, so kann eine schnelle Konvergenz des Algorithmus mit Normalverteilungen, die diese chemischen Verschiebungen als Mittelwerte und ein kleines  $\sigma$  haben, gewährleistet werden.

ASSIGN hat ebenfalls Normalverteilungen mit den von SHIFTS aus einer Eingabestruktur geschätzten Mittelwerten  $\delta_{ResNr, Kern}^{SHIFTS}$  und den Standardabweichungen aus der BMRB verwendet.

## Bestimmung der allgemeinen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen für alle Aminosäuren und Kernarten mit SHIFTS und einer unbiased Strukturdatenbank

Bessere Dichteverteilungen können mithilfe von SHIFTS und der unbiased Nh3D-Datenbank (Thiruv et al. 2005) erzeugt werden (Baskaran 2010). Die Wahrscheinlichkeitsdichteverteilungen  $f_{AA,Kern}(\delta)$  werden folgendermaßen berechnet:

$$f_{AA,Kern}(\delta) = \frac{1}{\sqrt{(2\pi)b_{AA,Kern}} N_{AA,Kern}} \sum_{i=1}^{N_{AA,Kern}} \exp\left(-\frac{1}{2} \left(\frac{\delta - \delta_i^{AA,Kern}}{b_{AA,Kern}}\right)^2\right) \quad (39)$$

Hierbei ist  $\delta_i^{AA,Kern}$  die  $i$ -te aus der Nh3D-Datenbank mit SHIFTS geschätzte chemische Verschiebung der Aminosäure  $AA$  und der Kernart  $Kern$ ,  $b_{AA,Kern}$  die zugehörige optimale Bandbreite des Gauß-Kernels und  $N_{AA,Kern}$  die gesamte Anzahl der geschätzten chemischen Verschiebungen dieses Kerntyps. Da die Verteilungsform nicht mehr auf eine Normalverteilung beschränkt ist, kann der Möglichkeitsraum besser gewichtet werden. Eine chemische Verschiebung kommt in Proteinen oft als ihr  $\alpha$ -Helix-,  $\beta$ -Faltblatt- oder als ihr Random-Coil-Wert vor. Diese drei Unterverteilungen haben im allgemeinen drei verschiedene Mittelwerte und drei verschiedene Standardabweichungen. Die Normalverteilung aus der BMRB dagegen hat nur einen Mittelwert, der zudem nicht unbiased ist. Die Breite der Verteilung ist zudem höher als die Breite der Unterverteilungen. Ein Beispiel dafür ist in Abb. 8 abgebildet. Die mit der Nh3D-Datenbank und SHIFTS bestimmte Verteilung von Ala H<sup>N</sup> (**grün**) hat zwei Maxima, die Normalverteilung aus der BMRB (**blau**) nur eins. Das zweite Maximum der grünen Kurve, das nicht mit dem Maximum der Normalverteilung übereinstimmt, liegt in diesem Fall ungefähr an der Position der richtigen chemischen Verschiebung (**schwarz**).

## Vorhersage der speziellen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen aller Kerne eines Proteins aus dem Molekulardynamik-Ensemble mit SHIFTX2

Für die dritte Möglichkeit zur Erzeugung der Dichteverteilungen für SIBASA wird ein MD-Ensemble des zuzuordnenden Proteins ausgewertet. Es werden Wahrscheinlichkeitsdichteverteilungen für jeden Kern mit SHIFTX2 (Han et al. 2011), das im Gegensatz zu SHIFTS und SHIFTX chemische Verschiebungen besser vorhersagen kann und außerdem Temperatur und pH berücksichtigt, geschätzt. Die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen  $f_{ResNr,Kern}(\delta)$  ergeben sich wie folgt:

$$f_{ResNr, Kern}(\delta) = \frac{1}{\sqrt{(2\pi)} b_{ResNr, Kern} N^{Ensemble}} \sum_{i=1}^{N^{Ensemble}} \exp\left(-\frac{1}{2} \left(\frac{\delta - \delta_i^{ResNr, Kern}}{b_{ResNr, Kern}}\right)^2\right) \quad (40)$$

Die Datenbasis der Verteilungen  $f_{ResNr, Kern}(\delta)$  des Kerns *Kern* in der Aminosäure mit der Position *ResNr* in der Primärsequenz besteht aus den mit SHIFTX2 geschätzten chemischen Verschiebungen  $\delta_i^{ResNr, Kern}$  der  $N^{Ensemble}$  Strukturen.  $b_{ResNr, Kern}$  ist wieder die optimale Bandbreite des Gauß-Kernels des Kerndichteschätzers. In Abb. 8 wurde die rote Verteilung mit SHIFTX2 aus dem MD-Ensemble von HPr *S. aureus* (H15A) erzeugt. Sie schränkt den Möglichkeitsraum deutlich stärker als die allgemeinen Dichteverteilungen ein.

## Kombination der allgemeinen und speziellen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen

Die speziellen Wahrscheinlichkeitsdichten liefern zwar die besten Vorhersagen, doch es kommt häufig vor, dass SHIFTX2 mit seinen Berechnungen falsch liegt. In diesen Fällen wäre die richtige chemische Verschiebung für SIBASA schwer zu bestimmen. Daher ist es empfehlenswert spezielle und allgemeine Wahrscheinlichkeitsverteilungen miteinander zu kombinieren. Die Wahrscheinlichkeit  $f_k(\delta)$  des Kerns *k* die chemische Verschiebung  $\delta$  einzunehmen wird mit dem Gesetz der totalen Wahrscheinlichkeit folgendermaßen bestimmt:

$$f_k(\delta) = f_{ResNr, Kern}(\delta) P_{Vorhersage}^{korrekt} + f_{AA, Kern}(\delta) P_{Vorhersage}^{falsch} \quad (41)$$

Dabei ist  $P_{Vorhersage}^{korrekt}$  die Wahrscheinlichkeit, dass die eine chemische Verschiebung von SHIFTX2 richtig vorhergesagt werden kann. Für die automatische Zuordnung ist eine Vorhersage richtig, falls der speziellen Wahrscheinlichkeitsdichte von  $f_{ResNr, Kern}(\delta_k^{korrekt})$  (Gl. 40) für die richtige chemische Verschiebung  $\delta_k^{korrekt}$  einen ausreichend hohen Wert hat. *ResNr* ist dabei die Position der Aminosäure in der Primärsequenz und *Kern* das betrachtete Atom in dieser Aminosäure. Falls die Vorhersage richtig ist kann SIBASA die Zuordnung in der Regel einfacher bestimmen.  $P_{Vorhersage}^{falsch}$  ist die reziproke Wahrscheinlichkeit  $1 - P_{Vorhersage}^{korrekt}$ .  $P_{Vorhersage}^{korrekt}$  kann mithilfe bereits gelöster Proteine und folgender Formel abgeschätzt werden:

$$P_{Vorhersage}^{korrekt} = \frac{N_{Vorhergesagt}^{korrekt}}{N_{Kerne}} \text{ mit } P_{ResNr, Kern}(\delta_{ResNr, Kern}^{korrekt}) > 0,01 \max(P_{ResNr, Kern}(\delta)) \quad (42)$$

Im Falle von HPr *S. aureus* (H15A) ist diese Wahrscheinlichkeit 25 %.

Der Wertebereich der Verteilungen  $f_k(\delta)$  wird in der Vorbereitung von links und rechts

eingeschränkt, so dass für  $\delta_k^{\min}$  und  $\delta_k^{\max}$  folgende Bedingungen gelten:

$$\begin{aligned} f_k(\delta_k^{\min}) &> 0,01 \max(f_k(\delta)) \\ f_k(\delta_k^{\max}) &> 0,01 \max(f_k(\delta)) \end{aligned} \quad (43)$$

Dies entspricht einer Segmentierung der Verteilung auf das 0,01 Level und schränkt den Suchraum auf die wesentlichen spektralen Bereiche ein.

## Normierung der Wahrscheinlichkeitsdichteverteilungen und Anpassung an die experimentellen Spektren

Anschließend werden die Dichteverteilungen an den Stellen, die wegen des Fehlens von Slots (Kap. 2.2.6) nicht zugänglich sind, auf 0 gesetzt. Diese Bereiche können zwar nicht ausgewählt werden und enthalten zudem keine Signale, können aber unter Umständen auf die Bewertung einer Zuordnung Einfluss nehmen. SIBASA minimiert unter anderem die Pseudoenergie  $\sum_{k=1}^{N_{\text{Kerne}}} \ln(A_k P_k(\delta_k))$  (Gl. 59 in Kap. 2.2.8). Die Maxima der Dichteverteilungen der chemischen Verschiebungen der Kerne  $k$   $f_k(\delta)$  werden mit den Normierungsfaktoren  $A_k$  auf 1 gesetzt, damit der Kern im bestmöglichen Fall keinen Beitrag zur Pseudoenergie leistet. Falls aber das Maximum einer Verteilung keinem Slot zuordenbar ist, ist dieser minimale Beitrag zur Pseudoenergie nicht erreichbar. Dies würde den dynamischen Bereich des Beitrags der Kerne  $k$  zur Pseudoenergie beschneiden und ihn außerdem gegenüber den anderen Zuordnungen untergewichten.

Mit dem Herausschneiden der Bereiche ohne Slots aus den Verteilungen und der anschließenden Normierung wird sichergestellt, dass die maximale Wahrscheinlichkeit und damit der minimale Beitrag zur Pseudoenergie zumindest theoretisch für jede Zuordnung erreicht werden kann.

### 2.2.8 Pseudoenergien

Damit Algorithmen wie ASSIGN und SIBASA die vollständige Zuordnung finden können, benötigen sie eine skalare Größe, die die Güte einer Konfiguration von chemischen Verschiebungen global bestimmen kann. Die gesamte Pseudoenergie der Optimierung  $E^{\text{total}}$  muss so konstruiert werden, dass sie für die richtige Zuordnung minimal wird. Der erste Teil der gesamten Pseudoenergie  $E^{\text{total}}$  wird beim Top-down-Ansatz aus lokalen Pseudoenergien, die aus ausgewählten spektralen Bereichen der Eingangsspektren stammen und sowohl Signalvolumen als auch Linienformen bewerten, gebildet. Zusätzlich wurde schon in ASSIGN diese Pseudoenergie der Signalübereinstimmung zwischen experimentellen und simulierten

Spektrum  $E^{\text{Vol,Shape}}$  mit der Pseudoenergie  $E^\omega$ , die die momentanen chemischen Verschiebungen mit den Wahrscheinlichkeitsdichteverteilung aus Kap. 2.2.7 bewertet, kombiniert. In SIBASA wurde die Gesamtpseudoenergie für die Auswertung von mehreren Spektren angepasst und an vielen Stellen verbessert. Dabei wurde der neue Term  $E^{\text{Dummy}}$  hinzugefügt, der dafür sorgt dass Kerne, deren Signale in den NOESY-Spektren nicht sichtbar sind, besser als solche identifiziert werden können.

Um zu gewährleisten, dass die verschiedenen Terme in der Gesamtpseudoenergie der Optimierung nicht zu stark aber auch nicht zu schwach berücksichtigt werden, muss die Pseudoenergie der chemischen Verschiebung mit dem Faktor  $w^{E,\omega}$  und die Pseudoenergie des Dummy Slots mit dem Faktor  $w^{E,\text{Dummy}}$  gewichtet werden. Die gesamte Pseudoenergie in SIBASA ergibt sich dann zu:

$$E^{\text{total}} = E^{\text{Vol,Shape}} + w^{E,\omega} E^\omega + w^{E,\text{Dummy}} E^{\text{Dummy}} \quad (44)$$

Zusätzlich wurden für SIBASA die Pseudoenergien  $E_k^{\text{Vol,Shape,sim}}$ , die in der Lage sind die Güte von einzelnen Zuordnungen  $k$  zu bestimmen, entwickelt.

## Lokale Pseudoenergien und Wahrscheinlichkeiten

ASSIGN hat für jedes gefundene experimentelle Signal eine lokale Pseudoenergie des Volumen- und Linienformvergleichs berechnet. Wurden die experimentellen Signale vollständig und sorgfältig ausgewählt, so bildeten deren Grundflächen zusammen den für die automatische Zuordnung wichtigen Bereich der experimentellen Spektren. Dieser Bereich ist im Idealfall vollständig mit der Rückrechnung erklärbar (Kap. 2.2.2). Die lokalen Pseudoenergien werden durch die Verrechnung des Vektors  $\vec{s}_{ij}^{\text{exp}}$ , der die Intensitäten der Grundfläche des experimentellen Signals  $j$  im Eingangsspektrum  $i$  (Grundfläche  $ij$ ) enthält mit dem Vektor  $\vec{s}_{ij}^{\text{sim}}$ , der die Intensitäten der Rückrechnung im selben spektralen Bereich enthält, bestimmt. Die Intensitäten in den Vektoren  $\vec{s}_{ij}^{\text{sim}}$  hängen von der momentanen Zuordnung ab und verändern sich während der Optimierung laufend.

### Grundflächen der lokalen Pseudoenergien

In SIBASA neu hinzugekommen ist die Möglichkeit nahe beieinanderliegender Grundflächen in experimentellen Spektren zusammenzufassen. Falls ein Signal durch das automatische Peakpicking mehrmals ausgewählt worden ist, hilft das Zusammenfassen dabei, dass dieses Signal in der Pseudoenergie der Optimierung nur einmal berücksichtigt wird. Nicht optimal prozessierte Spektren können z.B. periodisch auftretende lokale Maxima



entlang von Signalausläufern haben, die auch mehrfach in die Optimierung mit eingehen würden. Selbst bei einer perfekten Signalliste würde jedes Maximum eines auflösbaren Multipletts einzeln einbezogen werden. Dadurch würden Signale mit komplexe Aufspaltungsmuster stärker in die automatische Zuordnung einfließen als Singlets. Experimentelle Signale werden in SIBASA zusammengefasst, falls ihr Abstand in keiner Dimension größer als 15 Hz ist. Das neue Zentrum des zusammengefassten Signals ist das gemittelte Zentrum der beteiligten Signale und das Volumen die Summe der Einzelvolumen. Die Grundflächen der vereinigten Signale werden so gewählt, dass sie alle beteiligten Signalfächen enthalten sind (Abb. 9). Ein zusammengefasstes Signal wächst so lange bis keine weiteren Signale mehr im Abstand von 15 Hz zu finden sind.

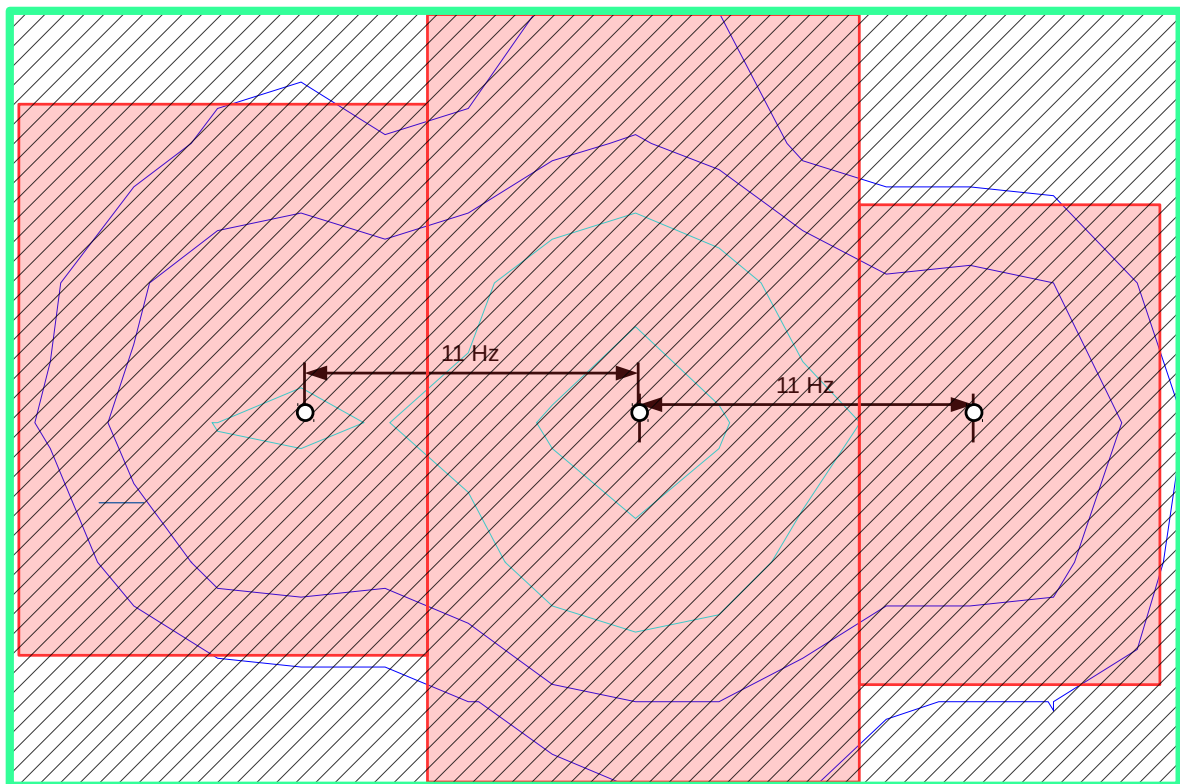


Abbildung 9: **Beispiel für das Zusammenfassen der Grundflächen eines Multiplett Signals.** Die Einzelsignale sind weniger als 15 Hz voneinander entfernt. Die drei roten Grundflächen werden von SIBASA zu einer Grundfläche (grüner Rahmen) zusammengefügt.

### Lokale Pseudoenergie des Volumenvergleichs

SIBASA übernimmt die von ASSIGN bekannten lokalen Pseudoenergien für Volumen- und Linienform, die aber auf den mit der neuen AUREMOL Integration bestimmten Signalfächen berechnet werden. Die Pseudoenergie des Volumens  $E_{ij}^{\text{Vol}}$  der Grundfläche  $ij$  wird folgendermaßen berechnet:

$$E_{ij}^{Vol} = \begin{cases} 1 - \frac{|d_{ij}^{exp} - d_{ij}^{sim}|}{d_{ij}^{exp}} & , \text{wenn } |d_{ij}^{exp} - d_{ij}^{sim}| < d_{ij}^{exp} \\ 0 & , \text{sonst} \end{cases} \quad (45)$$

Dabei werden die Abstände folgendermaßen aus den Intensitätsvektoren bestimmt:

$$d_{ij}^{exp} = (|\vec{s}_{ij}^{exp}|)^{-\frac{1}{6}} = (\sum_k \vec{s}_{ijk}^{exp})^{-\frac{1}{6}} \quad (46)$$

$$d_{ij}^{sim} = (|\vec{s}_{ij}^{sim}|)^{-\frac{1}{6}} = (\sum_k \vec{s}_{ijk}^{sim})^{-\frac{1}{6}}$$

Die Pseudoenergie  $E_{ij}^{Vol}$  vergleicht die aus den experimentellen und simulierten Signalvolumen auf der Grundfläche des experimentellen Signals  $j$  im Spektrum  $i$  berechneten Atomdistanzen. Für die Pseudoenergie des Volumens wird deshalb der Volumenskalierungsfaktor aus Kap. 2.2.4 benötigt.

### Lokale Pseudoenergie des Linienformvergleichs

Die Pseudoenergie der Linienform  $E_{ij}^{Shape}$  auf der Grundfläche  $ij$  wird dagegen folgendermaßen aus den Intensitätsvektoren  $\vec{s}_{ij}$  bestimmt:

$$E_{ij}^{Shape} = \cos(\angle(\vec{s}_{ij}^{exp}, \vec{s}_{ij}^{sim})) = \frac{\vec{s}_{ij}^{exp} \cdot \vec{s}_{ij}^{sim}}{|\vec{s}_{ij}^{exp}| \cdot |\vec{s}_{ij}^{sim}|} \quad (47)$$

Die Pseudoenergie  $E_{ij}^{Shape}$  bewertet die Ähnlichkeiten der experimentellen und simulierten Linienformens auf der Grundfläche des experimentellen Signals  $j$  im Spektrum  $i$  ohne dabei das Volumen mit einzubeziehen. Es gehen vor allem Linienbreiten, Multiplettaufspaltung und der Signalüberlapp auf der Grundfläche mit ein.

### Lokale Wahrscheinlichkeiten der richtigen vollständigen Zuordnung

Beide Pseudoenergien sind auf den Top-down-Ansatz zugeschnitten, denn sie vergleichen den gesamten Inhalt der spektralen Bereiche zwischen dem Experiment und seiner Rückrechnung und nicht nur, wie es bei dem Bottom-up Ansatz üblich ist, die Topologie, Signalpositionen.

ASSIGN und SIBASA versuchen die wahrscheinlichste vollständige richtige Zuordnung zu finden. Dafür ist es notwendig die Gesamtwahrscheinlichkeit der richtigen Zuordnung zu maximieren. Zuerst werden lokale Wahrscheinlichkeit aus den berechneten lokalen Pseudoenergien des Volumens und der Linienform mit dem Satz von Bayes bestimmt. Die Gesamtwahrscheinlichkeit ist das Produkt dieser voneinander unabhängigen lokalen Wahrscheinlichkeiten. Die Wahrscheinlichkeit  $P(\text{korrekt}|E_{ij}^{\text{Vol}})$ , dass die vollständige Zuordnung auf der Grundfläche  $ij$  mit der Pseudoenergien des Volumens  $E_{ij}^{\text{Vol}}$  richtig ist, wird folgendermaßen berechnet:

$$P(\text{korrekt}|E_{ij}^{\text{Vol}}) = \frac{P(E_{ij}^{\text{Vol}}|\text{korrekt})}{P(E_{ij}^{\text{Vol}}|\text{korrekt}) + P(E_{ij}^{\text{Vol}}|\text{falsch})} \quad (48)$$

Analog dazu berechnet sich die Wahrscheinlichkeit  $P(\text{korrekt}|E_{ij}^{\text{Shape}})$  für die Pseudoenergien der Linienform  $E_{ij}^{\text{Shape}}$  folgendermaßen:

$$P(\text{korrekt}|E_{ij}^{\text{Shape}}) = \frac{P(E_{ij}^{\text{Shape}}|\text{korrekt})}{P(E_{ij}^{\text{Shape}}|\text{korrekt}) + P(E_{ij}^{\text{Shape}}|\text{falsch})} \quad (49)$$

Die beiden Wahrscheinlichkeiten lassen sich unter der Annahme, dass die Pseudoenergien des Volumens und der Linienform auf einer Grundfläche  $ij$  voneinander unabhängig sind, zur Wahrscheinlichkeit der richtigen Zuordnung  $P_{ij}^{\text{Vol, Shape}}$  auf der Grundfläche  $ij$  kombinieren:

$$\begin{aligned} P_{ij}^{\text{Vol, Shape}} &= P(\text{korrekt}|E_{ij}^{\text{Vol}} \cap E_{ij}^{\text{Shape}}) = \\ &= \frac{P(E_{ij}^{\text{Vol}} \cap E_{ij}^{\text{Shape}}|\text{korrekt}) P(\text{korrekt})}{P(E_{ij}^{\text{Vol}} \cap E_{ij}^{\text{Shape}}|\text{korrekt}) P(\text{korrekt}) + P(E_{ij}^{\text{Vol}} \cap E_{ij}^{\text{Shape}}|\text{falsch}) P(\text{falsch})} = \\ &= \frac{P(E_{ij}^{\text{Vol}}|\text{korrekt}) P(E_{ij}^{\text{Shape}}|\text{korrekt})}{P(E_{ij}^{\text{Vol}}|\text{korrekt}) P(E_{ij}^{\text{Shape}}|\text{korrekt}) + P(E_{ij}^{\text{Vol}}|\text{falsch}) P(E_{ij}^{\text{Shape}}|\text{falsch})} \end{aligned} \quad (50)$$

Da die globale Wahrscheinlichkeit für die richtige Zuordnung  $P(\text{korrekt})$  nicht angegeben werden kann, wird folgendes vorausgesetzt:

$$P(\text{korrekt}) = P(\text{falsch}) = 0,5 \quad (51)$$

ASSIGN hat zwar ebenfalls die lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $P_{ij}^{\text{Vol, Shape}}$  bestimmt, aber die dafür benötigten Häufigkeitsverteilungen der beiden Pseudoenergien im Falle einer korrekten vollständigen Zuordnung  $H_i^{\text{korrekt}}(E^{\text{Vol}})$  und  $H_i^{\text{korrekt}}(E^{\text{Shape}})$  fehlten bislang für alle Spektren. SIBASA bezieht diese nun aus der AUREMOL Datenbank, in der sie für verschiedene Spektrentypen abgelegt worden sind. Da

sich die Häufigkeitsverteilungen der beiden Pseudoenergien bei gleicher Prozessierung und gleichen Spektrenarten bei unterschiedlichen Proteinen nicht stark voneinander unterscheiden und der Satz von Bayes gegen leichte Veränderungen tolerant ist, können sie aus den Datensätzen bereits gelöster Proteinen erzeugt werden und für alle Zuordnungsprobleme mit den gleichen Spektren verwendet werden. Zur Erstellung der Verteilungen wird ein Kerndichteschätzer mit Gauß-Kernel und optimaler Bandbreite  $b$  (Berwin A. TURLACH), statt dem von ASSIGN verwendeten Häufigkeitsverteilungen mit 10 Klassen verwendet. Die Verteilungen werden dadurch geglättet, was das Übertragen auf andere Zuordnungsprobleme weiter vereinfacht. Für jedes Spektrum  $i$ , wie z.B. 2-D-NOESY- oder das 3D  $^{15}\text{N}$ -NOES-HSQC-Spektrum werden sie mit folgenden Gleichungen erzeugt:

$$H_i^{\text{korrekt}}(E^{\text{Vol}}) = \frac{1}{\sqrt{2\pi} b_{E_i^{\text{Vol}}}} \sum_j \exp\left(-\frac{1}{2} \left(\frac{E - E_{ij}^{\text{Vol}}}{b_{E_i^{\text{Vol}}}}\right)^2\right) \quad (52)$$

$$H_i^{\text{korrekt}}(E^{\text{Shape}}) = \frac{1}{\sqrt{2\pi} b_{E_i^{\text{Shape}}}} \sum_{j=1}^{N_i^{\text{Signale}}} \exp\left(-\frac{1}{2} \left(\frac{E - E_{ij}^{\text{Shape}}}{b_{E_i^{\text{Shape}}}}\right)^2\right) \quad (53)$$

Da sowohl die Pseudoenergie des Volumens als auch die der Linienform nur Werte zwischen 0 und 1 haben können, aber der Kerndichteschätzer Ausläufer über diese Grenzen hinaus erzeugt, werden diese Ausläufer an den jeweiligen Außengrenzen achsensymmetrisch zurück gespiegelt und zu den Verteilungen addiert.

Die Verteilungen für eine falsche Zuordnung  $H_i^{\text{falsch}}(E^{\text{Vol}})$  und  $H_i^{\text{falsch}}(E^{\text{Shape}})$  wurden mit der gleiche Methode und dem gleichen Datensatz, aber mit einer zufälligen vollständigen Zuordnung erzeugt und ebenfalls in der AUREMOL Datenbank gespeichert.

### **Zusammenfassung der (lokalen) Wahrscheinlichkeiten zu globalen Pseudoenergien**

In ASSIGN und SIBASA wird immer wenn möglich ausgenutzt, dass die gesuchte und zu maximierende Gesamtwahrscheinlichkeit der richtigen Zuordnung, das Produkt seiner voneinander unabhängigen Einzelwahrscheinlichkeiten ist. Der Logarithmus der Gesamtwahrscheinlichkeit ist damit die Summe der logarithmierten Einzelwahrscheinlichkeiten und kann als Pseudoenergie, die minimiert werden soll, dienen. Ist die Pseudoenergie minimal, so ist die Zuordnung mit der maximalen Gesamtwahrscheinlichkeit für die richtige Zuordnung gefunden.

In ASSIGN gab es nur ein 2-D-NOESY-Spektrum zu beachten, dessen lokalen Bayessche Wahrscheinlichkeiten der Grundflächen  $j$  zu der globalen Gesamtpseudoenergie des Volumens

und der Linienform  $E^{\text{VolShape}}$  zusammengefasst wurden:

$$E^{\text{Vol, Shape}} = \ln \left| \prod_{j=0}^{N^{\text{Signale}}} P(\text{korrekt} | E_j^{\text{Vol}} \cap E_j^{\text{Shape}}) \right| = \sum_{j=0}^{N^{\text{Signale}}} |\ln P_j^{\text{Vol, Shape}}| \quad (54)$$

Dabei war  $N^{\text{Signale}}$  die Anzahl der NOESY-Signale und  $E_j^{\text{Vol}}$  bzw.  $E_j^{\text{Shape}}$  die Pseudoenergie des Volumens bzw. der Linienform auf der Grundfläche des experimentellen Signals  $j$ . Um numerische Stabilität zu gewährleisten, wurden Wahrscheinlichkeiten, die kleiner als 0,05 waren, auf 0,05 gesetzt.

### Gewichtung der lokalen Wahrscheinlichkeiten in der globalen Pseudoenergie

In ASSIGN sind alle lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $P_j^{\text{Vol, Shape}}$ , unabhängig von der Intensität des zugrundeliegenden Signals, gleich stark in die globale Gesamtpseudoenergie des Volumens und der Linienform  $E^{\text{Vol, Shape}}$  eingegangen. In SIBASA dagegen gehen die Pseudoenergien von stärkeren Signalen, die für die Zuordnung von chemischen Verschiebungen in NOESY-Spektren wichtiger sind als die von schwächeren Signalen, auch stärker in die globale Gesamtpseudoenergie  $E_i^{\text{Vol, Shape}}$  des Spektrums  $i$  ein. Die logarithmierten lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $|\ln P_{ij}^{\text{Vol, Shape}}|$  werden dafür mit ihren experimentellen Signalvolumen  $V_{ij}^{\text{exp}}$  in der gesamten Pseudoenergie des Signalvergleichs  $E_i^{\text{Vol, Shape}}$  gewichtet. Die Gewichtung minimiert außerdem den Einfluss von thermische Rauschsignalen und Artefakten mit geringer Intensität auf die automatische Zuordnung. Es ist außerdem wahrscheinlicher, dass schwache Signale nicht von der Rückrechnung einer nicht idealen 3D-Struktur erklärt werden können, da sich ihre Wechselwirkungsdistanzen nahe am Cut-Off Abstand der Simulation befinden. Starke Signale dagegen, sind oft selbst in den simulierten Spektren des ausgestreckten Strang sichtbar und erhalten auch bei der manuellen Zuordnung die meiste Aufmerksamkeit. Die Gewichtung  $w_{ij}^{\text{Signal, exp}}$  des Signals  $j$  im Spektrum  $i$  wird mithilfe seines Volumens  $V_{ij}^{\text{exp}}$  mit folgenden Formeln berechnet:

$$w_{ij}^{\text{Signal, exp}} = 9 \frac{(V_{ij}^{\text{exp}})^{-1/6} - (V_{i, \text{min}}^{\text{exp}})^{-1/6}}{(V_{i, \text{max}}^{\text{exp}})^{-1/6} - (V_{i, \text{min}}^{\text{exp}})^{-1/6}} + 1 \quad \text{für das 2D-NOESY Spektrum} \quad (55)$$

$$w_{ij}^{\text{Signal, exp}} = 4,5 \frac{(V_{ij}^{\text{exp}})^{-1/6} - (V_{i, \text{min}}^{\text{exp}})^{-1/6}}{(V_{i, \text{max}}^{\text{exp}})^{-1/6} - (V_{i, \text{min}}^{\text{exp}})^{-1/6}} + 0,5 \quad \text{für alle anderen Spektren}$$

Es hat sich bewährt das minimale mögliche Gewicht einer experimentellen Grundfläche  $w_{ij}^{\text{Signal, exp}}$  im führende 2-D-NOESY-Spektrum auf 1,0 und das maximale Gewicht auf 10,0 zu

setzen. Es wird, wie schon bei der Berechnung der Pseudoenergie des Volumens  $E_{ij}^{\text{Vol}}$  (Gl. 45) die Distanz aus dem Volumen berechnet. Zwischen den Grenzen besteht ein linearer Zusammenhang zwischen Gewicht  $w_{ij}^{\text{Signal,exp}}$  und der aus den Volumen berechneten Distanzen  $(V^{\text{exp}})^{-1/6}$ . In allen weiteren Spektren werden die Grenzen der Gewichte auf 0,5 und 5, also jeweils genau auf die Hälfte gesetzt. Im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum trägt der reduzierte dynamische Bereich der Gewichtungen der Tatsache Rechnung, dass dort in der Regel das Signal-zu-Rausch-Verhältnis, die Auflösung in den indirekten Dimensionen und chemische Austauschereffekte die Genauigkeiten der gemessenen Signalvolumen verschlechtern. Die gesamte Pseudoenergie des Volumens und der Linienform  $E_i^{\text{Vol,Shape}}$  des Spektrums  $i$ , wird mit den lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $P_{ij}^{\text{Vol,Shape}}$  folgendermaßen berechnet:

$$E_i^{\text{Vol,Shape}} = \sum_{j=1}^{N_i^{\text{Signale}}} w_{ij}^{\text{Signal,exp}} |\ln P_{ij}^{\text{Vol,Shape}}| \quad (56)$$

### Gewichtung der Pseudoenergien der Spektren untereinander

Im Gegensatz zu ASSIGN kann SIBASA mehrere Spektren, deren Pseudoenergien gemeinsam optimiert werden, für die automatische Zuordnung verwenden. Da z.B. ein 2-D-NOESY-Spektrum in der Regel mehr Signale und damit auch mehr aufsummiertes Gewicht  $\sum_{j=1}^{N_{2\text{-D-NOESY}}^{\text{Signale}}} w_{2\text{-D-NOESY},j}^{\text{Signal,exp}}$  als das zugehörige 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum hat, muss der Beitrag der globalen Pseudoenergie des Volumens und der Linienform  $E_{3\text{-D-NOESY}}^{\text{Vol,Shape}}$  des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums höher gewichtet werden, um in der Optimierung richtig berücksichtigt zu werden. Mit der Wichtung wird garantiert, dass kein Spektrum während der automatischen Zuordnung vernachlässigt wird.

Bei allen Gewichtungsfaktoren, die für SIBASA eingeführt wurden, um Pseudoenergien gegeneinander zu gewichten, wird ausgenutzt, dass wegen der numerischen Stabilität eine Wahrscheinlichkeiten minimal 0,05 sein kann. Dadurch kann die maximal mögliche Pseudoenergie mit der Beziehung  $\sum_{j=1}^{\text{Wahrscheinlichkeiten}} \ln(0,05)$  bestimmt werden. Ein globaler Gewichtungsfaktor wird als Referenz auf 1 gesetzt und die anderen Gewichtungsfaktoren so gewählt, dass die maximale mit dem Faktor  $w_i$  gewichtete Pseudoenergie  $w_i \sum_j \ln(0,05)$  genauso hoch wie die Pseudoenergie der Referenz ist, bzw. den gewünschten Anteil der Pseudoenergie der Referenz hat. Im Falle der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  wird der Gewichtungsfaktor des führenden 2-D-NOESY-Spektrums  $w_1^{\text{Spektrum, Vol, Shape}}$  als Referenz auf 1 gesetzt und die Gewichtungsfaktoren der anderen Spektren  $w_i^{\text{Spektrum, Vol, Shape}}$  so gewählt, dass ihre höchstmögliche gewichtete Volumen- und Linienformpseudoenergie die Hälfte der maximalen Pseudoenergie des 2-D-NOESY-Spektrums beträgt.

$$w_i^{\text{Spektrum, Vol, Shape}} = 0,5 \frac{\sum_{j=1}^{N_i^{\text{Signale}}} w_j^{\text{Signal, exp}}}{\sum_{j=1}^{N_i^{\text{Signale}}} w_{ij}^{\text{Signal, exp}}} \quad (57)$$

Das führende 2-D-NOESY-Spektrum erhält das doppelte Gewicht, da es mit seinem großen und vollständigen Korrelationsnetz der dipolaren Wechselwirkungen die Hauptinformation für die automatische Zuordnung zur Verfügung stellt. Damit ergibt sich die globale Gesamtpseudoenergie des Volumens und der Linienform aller Spektren  $E^{\text{Vol, Shape}}$  zu:

$$E^{\text{Vol, Shape}} = \sum_{i=1}^{N^{\text{Spektren}}} w_i^{\text{Spektrum, Vol, Shape}} \sum_{j=1}^{N_i^{\text{Signale}}} w_{ij}^{\text{Signal, exp}} |\ln P_{ij}^{\text{Vol, Shape}}| \quad (58)$$

## Pseudoenergie der chemischen Verschiebung

ASSIGN hat neben der Pseudoenergie der Linienformen und des Volumens  $E^{\text{Vol, Shape}}$  die Pseudoenergie der chemischen Verschiebung  $E^\omega$  eingeführt. Sie ist der Logarithmus des Produkts der unabhängigen Einzelwahrscheinlichkeiten der Zuordnung, also wieder der Logarithmus der Gesamtwahrscheinlichkeit (Gl. 59).

$$E^\omega = |\ln (\prod_{k=1}^{N^{\text{Kerne}}} P_k(\delta_k))| = \sum_{k=1}^{N^{\text{Kerne}}} |\ln P_k(\delta_k)| \quad (59)$$

Die Wahrscheinlichkeiten  $P_k(\delta_k)$  der chemischen Verschiebungen der Kerne  $k$  wurden in ASSIGN folgendermaßen bestimmt:

$$P_k(\delta_k) = \exp\left(-\frac{1}{2} \left(\frac{\delta_k - \delta_k^{\text{SHIFTS}}}{\sigma_{AA, Kern}^{\text{BMRB}}}\right)^2\right) \quad (60)$$

$\delta_k^{\text{SHIFTS}}$  war die aus der Eingabestruktur mit SHIFTS vorhergesagte chemische Verschiebung der Resonanz des Kerns  $k$  und  $\sigma_{AA, Kern}^{\text{BMRB}}$  die aus der BMRB stammende Standardabweichung der des Kerns *Kern* in der Aminosäure *AA* (siehe Kap. 2.2.7). Auch hier wurde, wie zuvor bei den lokalen Wahrscheinlichkeiten  $P_{ij}^{\text{Vol, Shape}}$  der richtigen Zuordnung (Gl. 50), wegen der numerischen Stabilität, die minimale Wahrscheinlichkeit  $P_k(\delta_k)$  auf 0,05 festgelegt. Die gesamte zu optimierende Pseudoenergie  $E^{\text{total}}$  ergab sich in ASSIGN für ein 2-D-NOESY-Spektrum zu:

$$E^{\text{total}} = E^{\text{Vol, Shape}} + w^{E, \omega} E^\omega \quad (61)$$

$w^{E, \omega}$  war dabei ein vom Benutzer vorgegebener Gewichtungsfaktor, der dafür sorgte, dass

die Pseudoenergie der chemischen Verschiebungen in der Optimierung gegenüber der Volumen- und Linienformpseudoenergie richtig gewichtet war.

Die Pseudoenergie der chemischen Verschiebung  $E^\omega$  wird auch in SIBASA als Pseudoenergie verwendet, wobei aber die verbesserten Wahrscheinlichkeitsdichteverteilungen der chemischen  $f_k(\delta)$  (Gl. 41) aus Kapitel 2.2.7 verwendet werden. Die Pseudoenergie der chemischen Verschiebung  $E^\omega$  wird folgendermaßen bestimmt:

$$E^\omega = \sum_{k=1}^{N^{Kerne}} |\ln(A_k f_k(\delta_k))| \quad (62)$$

Der Normierungsfaktor  $A_k$  normiert die Beiträge  $A_k f_k(\delta_k)$  zur Pseudoenergie auf 1 und sorgt dafür, dass der Logarithmus des optimalen Beitrags, unabhängig vom Kern  $k$  gleich 0 ist.

### **Dynamischer Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung**

In SIBASA wird der benötigte Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung in der Gesamtpseudoenergie  $w^{E,\omega}$  problemspezifisch berechnet und während der Optimierung dynamisch angepasst. Da zu Beginn die Liste der chemische Verschiebungen zufällig initialisiert wird und damit größtenteils falsch ist, sind auch die lokalen Wahrscheinlichkeiten  $P_{ij}^{Vol,Shape}$  der richtigen Zuordnung auf den Grundflächen  $ij$ , wegen ihrer Abhängigkeiten von zwei oder mehreren chemischen Verschiebungen noch wenig nützlich für die automatische Zuordnung. Deshalb ist es sinnvoll zu Beginn die Pseudoenergie der chemischen Verschiebungen relativ hoch zu gewichten, da ihre Beiträge nur von einer einzigen chemischen Verschiebung abhängig sind. Sie sind damit in der Anfangsphase besser in der Lage SIBASA in die richtige Richtung zu bringen. Im späteren Verlauf würde aber ein Algorithmus, der zu stark auf die Wahrscheinlichkeitsverteilungen der chemischen Verschiebungen setzt, die richtige vollständige Zuordnung nicht finden, da die geschätzten Wahrscheinlichkeitsdichteverteilungen zu einem nicht unerheblichen Teil falsch sind. Für den späteren Verlauf der automatische Zuordnung hat es sich als sinnvoll erwiesen, langsam auf die Pseudoenergie des Signalvergleichs  $E^{Vol,Shape}$ , die auf experimentellen NMR-Daten basiert, als Hauptkriterium der Optimierung zu wechseln. Dies wird erreicht indem der Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}(t)$  während der Optimierung fortlaufend angepasst wird. Zu Beginn wird er zunächst folgendermaßen initialisiert:

$$w^{E,\omega}(t=0) = A \frac{E^{Vol,Shape,max}}{N^{Kerne} \ln(0,05)} \quad (63)$$



$N^{Kern} \ln(0,05)$  ist dabei der maximal mögliche Wert der Pseudoenergie der chemischen Verschiebungen  $E^\omega$ . Der maximale Beitrag der Pseudoenergie der chemischen Verschiebung wird über den Gewichtungsfaktor  $w^{E,\omega}$  an den maximalen mögliche Wert der Volumen- und Linienformpseudoenergie  $E^{Vol,Shape,max}$  aller Spektren angepasst.  $A$  ist eine benutzerdefinierte Stärke, die in der Regel auf Werte zwischen 1 und 3 eingestellt werden sollte. Wird sie auf 1 eingestellt, so sind die maximalen Beiträge der Pseudoenergie  $E^{Vol,Shape}$  und der Pseudoenergie der chemischen Verschiebungen  $E^\omega$  zur gesamten Pseudoenergie zu Beginn gleich groß. Für diese Arbeit hat sich eine erhöhte Stärke  $A$  von 1,5 bewährt, da die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen sich gegenüber von ASSIGN verbessert haben.

Die Gewichtung  $w^{E,\omega}(t)$  wird während der Optimierung regelmäßig folgendermaßen aktualisiert:

$$w^{E,\omega}(t) = A \frac{(E^{Vol,Shape}(t) - E^{Vol,Shape,opt})}{N^{Kerne} \ln(0,05)} \quad (64)$$

Hierbei ist  $E^{Vol,Shape,opt}$  der im vornherein geschätzte Wert der Volumen- und Linienformpseudoenergie der korrekten vollständigen Zuordnung, also die Pseudoenergie des Endergebnisses. Ist die richtigen Zuordnung erreicht, so wird auch der Gewichtungsfaktor  $w^{E,\omega}(t)$  minimal. Da aber im Realfall, wegen Artefakten, Rausch und Ungenauigkeiten in den 3D-Strukturen und Rückrechnungen, die kombinierte Volumen- und Linienformpseudoenergie die vollständige Zuordnung nicht alleine finden kann, sorgt SIBASA dafür dass der Gewichtungsfaktor  $w^{E,\omega}(t)$  nie unter 10% seines Anfangswert aus Gl. 63 fällt.

### Abschätzung der Pseudoenergie des Signalvergleichs der richtigen Lösung

Die Pseudoenergie des Signalvergleichs der richtigen Zuordnung  $E^{Vol,Shape,opt}$  kann während der Vorbereitung der automatischen Zuordnung folgendermaßen abgeschätzt werden:

$$E^{Vol,Shape,opt} = \sum_{i=1}^{N^{Spektren}} w_i^{E,Vol,Shape} \bar{E}_i^{Vol,Shape,korrekt} \sum_{j=1}^{N_i^{Signale}} w_{ij}^{Signal,exp} \quad (65)$$

$w_i^{E,Vol,Shape}$  ist der Gewichtungsfaktor der Pseudoenergie des Signalvergleichs für das Spektrum  $i$  und  $w_{ij}^{Signal,exp}$  der Gewichtungsfaktor des Signals  $j$  im Spektrum  $i$  für die selbe Pseudoenergie.  $\bar{E}_i^{Vol,Shape,korrekt}$  ist die durchschnittliche Pseudoenergie des Volumen- und Linienformvergleichs  $E_i^{Vol,Shape}$  eines beliebigen Signals im vollständig richtig zugeordneten Spektrum  $i$ . Sie wäre im Idealfall 0, muss aber für den Realfall mit Hilfe der Häufigkeitsverteilungen der richtigen Zuordnung der beiden Pseudoenergien des Volumens

und der Linienform (  $H_i^{\text{korrekt}}(E^{\text{Vol}})$  (Gl. 52) bzw.  $H_i^{\text{korrekt}}(E^{\text{Shape}})$  (Gl. 53) ) mit folgender Gleichung für alle Eingabespektren  $i$  ermittelt werden:

$$\bar{E}_i^{\text{Vol, Shape, korrekt}} = \frac{1}{Z} \int \frac{N_i^{\text{Signale}}(E)}{N_i^{\text{Signale}}} \bar{w}_i^{\text{Signal, exp}}(E) |\ln P_i^{\text{Vol, Shape, korrekt}}(E)| dE \quad (66)$$

Dies ist im Prinzip die Formel zur Bestimmung des Mittelwertes der gewichteten Pseudoenergie der experimentellen Signale aus den Häufigkeitsverteilungen der Pseudoenergie  $E$ . Dabei ist:

- $P_i^{\text{Vol, Shape, korrekt}}(E)$  ist die mit mit Gl. 50 gemittelte lokale Wahrscheinlichkeit eines Signals mit den Pseudoenergien  $E = E^{\text{Vol}} = E^{\text{Shape}}$  im Spektrum  $i$  für die richtige vollständige Zuordnung.
- $\frac{N_i^{\text{Signale}}(E)}{N_i^{\text{Signale}}}$  ist die relative Häufigkeit eines experimentellen Signals mit den Pseudoenergien  $E = E^{\text{Vol}} = E^{\text{Shape}}$  im Spektrum  $i$ . Sie wird aus den beiden Verteilungen der Pseudoenergien der richtigen Zuordnung folgendermaßen gemittelt:  

$$\frac{N_i^{\text{Signale}}(E)}{N_i^{\text{Signale}}} = \frac{H_i^{\text{korrekt}}(E^{\text{Vol}}) + H_i^{\text{korrekt}}(E^{\text{Shape}})}{2}$$
- $\bar{w}_i(E)$  ist das durchschnittliche Gewicht eines experimentellen Signals mit  $E = E^{\text{Vol}} = E^{\text{Shape}}$  im Spektrum  $i$  des gelösten Proteins, mit dessen Datensatz auch die Verteilungen  $H_i^{\text{korrekt}}(E^{\text{Vol}})$  und  $H_i^{\text{korrekt}}(E^{\text{Shape}})$  erzeugt wurden.
- $\bar{w}_i^{\text{Signal, exp}}(E) |\ln P_i^{\text{Vol, Shape}}(E)|$  ist damit die durchschnittliche lokale Pseudoenergie des Signalvergleichs, die ein Signal zur Pseudoenergie  $E_i^{\text{Vol, Shape}}$  des richtig zugeordneten Spektrums  $i$  beiträgt.
- $Z = \int \frac{N_i^{\text{Signale}}(E)}{N_i^{\text{Signale}}} \bar{w}_i^{\text{Signal, exp}}(E) dE$  ist der Normierungsfaktor. Der durchschnittliche Gewichtungsfaktor eines experimentellen Signals mit den Pseudoenergien  $E = E^{\text{Vol}} = E^{\text{Shape}}$  im Spektrum  $i$   $\bar{w}_i^{\text{Signal, exp}}(E)$  ist eigentlich ein Teil der zu mittelnden Größe und sollte deswegen kein Teil der Normierung sein. Er wird trotzdem mit eingerechnet, da alle Signalgewichte  $w_{ij}^{\text{Signal, exp}}$  der Eingabespektren bekannt sind und somit nicht mit abgeschätzt werden müssen.

Die Annahme  $E = E^{\text{Vol}} = E^{\text{Shape}}$  in den Integralen ist möglich, da die Pseudoenergie des Volumens  $E_{ij}^{\text{Vol}}$  und der Linienform  $E_{ij}^{\text{Shape}}$  auf der Grundfläche des experimentellen Signals  $j$  im Spektrum  $i$  in der richtigen vollständigen Zuordnung stark miteinander korreliert sind.

## Bewertung von Zuordnungen

Eine der größten Neuerungen von SIBASA ist die Möglichkeit sich die Ergebnisse der automatischen Zuordnung bewerten zu lassen. Dazu wurde eine von der Pseudoenergie der experimentellen Signale  $E^{\text{Vol,Shape}}$  (Kap. 2.2.8) unabhängige neue Größe definiert. Die Pseudoenergie  $E_k^{\text{Vol,Shape,sim}}$  wird für jede Zuordnung eines Kern  $k$  berechnet und basiert auch auf dem Volumen- und Linienformvergleichen zwischen dem variablen zurückgerechneten und dem experimentellen Spektrum. Die Pseudoenergien des Volumens  $E_{ij}^{\text{Vol,Sim}}$  werden mit Gl. 45 und die der Linienformen  $E_{ij}^{\text{Shape,Sim}}$  mit Gl. 47 auf den Grundflächen der simulierten Signale  $j$  im Spektrum  $i$  berechnet. Diese Grundflächen sind um die aktuellen Koordinaten des simulierten Kreuzsignals  $ij$  zentriert und schließen die gesamte zurückgerechnete Linienform ein. Die Größe der Flächen werden wie bei den experimentellen Signalen (Kap. 2.2.2), durch eine Segmentationshöhe vorgegeben.  $E_{ij}^{\text{Vol,Sim}}$  und  $E_{ij}^{\text{Shape,Sim}}$  sind gegen Ungenauigkeiten in der Eingangsstruktur sehr empfindlich, da schwache Kreuzsignale bis zum Cut-Off Abstand, die eventuell kein experimentelles Äquivalent haben, mit berücksichtigt werden. Auch Artefakte, wie z.B. das Wassersignal im 2-D-NOESY-Spektrum, können die lokalen Pseudoenergien der simulierten Signale verfälschen, da die experimentellen Signale, die unter dem Artefakt liegen, für eine Auswertung nicht zugänglich sind. Eine Vorauswahl der simulierten Signale, deren Volumen- und Linienformpseudoenergien in die Berechnung der Pseudoenergien  $E_k^{\text{Vol,Shape,sim}}$  eingehen sollen, ist deshalb notwendig. Für die Bewertung der Sichtbarkeit eines simulierten Signal  $j$  im Spektrum  $i$  im Experiment, wird sein Pseudoabstand  $d_{kl}^{\text{pseudo}}$  berechnet:

$$d_{kl}^{\text{pseudo}} = v_{kl}^{\text{red}} \left( \sum_m \sum_n r_{mn}^{-6} \right)^{-1/6} \quad (67)$$

Hier findet eine  $r^{-6}$ -Aufsummierung der Abstände der Kerne von gleichwertigen Kreuzsignalen magnetisch gleichwertiger Kernen statt.  $k$  und  $l$  sind die Indices der von SIBASA zuzuordnenden Kerne des Signals  $j$  im Spektrum  $i$ .  $m$  und  $n$  sind die Indices der magnetisch gleichwertigen Atome der Atomgruppen  $k$  und  $l$ . Der Pseudoabstand  $d_{kl}^{\text{pseudo}}$  berücksichtigt mit dem Parameter  $v_{kl}^{\text{red}}$  Ordnungsparameter, Korrelationszeiten und andere Faktoren, die das Signalvolumen verkleinern können. Die Parameter  $v_{kl}^{\text{red}}$  stammen von AUREMOL RELAX, die diese für jede Rückrechnung mit bestimmt.

Damit die Grundfläche eines simulierten Kreuzsignals für die Berechnung der Volumen- und Linienformpseudoenergie der Zuordnung  $k$   $E_k^{\text{Vol,Shape,sim}}$  mit einbezogen wird, müssen folgende Kriterien erfüllt sein:

- Der Pseudoabstand  $d_{kl}^{\text{pseudo}}$  der beiden Atomgruppen  $k$  und  $l$  im Kreuzsignal muss kleiner als 0,4 nm sein. Dieser Threshold beseitigt bei dem gängigen Cut-Off Abstand

der Rückrechnung von 0,5 nm und normalen Signal-zu-Rausch-Verhältnis den Einfluss der meisten Fehler der 3D-Strukturen.

- Simulierte Signale, die im 2-D-NOESY-Spektrum unter dem Wassersignal zu finden wären, d.h. in der direkten Dimension zwischen 4,3 und 5,3 ppm bzw. in der indirekten Dimension zwischen 4,6 und 5,0 ppm liegen würden, werden ausgeschlossen.
- Für das 2-D-NOESY-Spektrum werden simulierte Diagonalsignale und Kreuzsignale, die nahe an die Diagonale gesetzt wurden, d.h. Signale für die  $|\delta(\omega_1) - \delta(\omega_2)|$  kleiner als 0,5 ppm ist, wegen dem starken Überlapp in diesem spektralen Bereich ausgeschlossen.
- Nicht sichtbare simulierte Kreuzsignale, also Signale bei denen mindestens ein beteiligter Kern im Dummyslot liegt, werden ebenfalls nicht berücksichtigt.

Es werden, wie schon für die lokalen Pseudoenergien der experimentellen Signale, mit dem Satz von Bayes die lokalen Wahrscheinlichkeiten für die richtigen Zuordnung  $P_{ij}^{\text{Vol,Shape,sim}}$  der Grundflächen der simulierten Signale  $j$  im Spektrum  $i$  aus den Volumen- und den Linienformpseudoenergien ( $E_{ij}^{\text{Vol,sim}}$  bzw.  $E_{ij}^{\text{Shape,sim}}$ ) gebildet. Diese Wahrscheinlichkeiten unterschieden sich von ihren experimentellen Äquivalenten nur in den spektralen Bereichen, in denen sie ausgewertet werden. Die Zentren dieser Bereiche sind mit Kreuzsignalzuordnungen verbunden und verändern während der Optimierung ihre Positionen. Die Pseudoenergien auf der Grundfläche bewerten, wie gut ein simuliertes Signal zu einem bestimmten Zeitpunkt in der Optimierung, der auch am Ende sein kann, zugeordnet wurde. Die für die Berechnung mit Gl. 50 benötigten Verteilungen der Volumen- und Linienformpseudoenergien der simulierten Signale der richtigen Lösung  $H_i^{\text{korrekt}}(E^{\text{Vol,sim}})$  und  $H_i^{\text{korrekt}}(E^{\text{Shape,sim}})$  werden mit allen simulierten Signalen eines gelösten Proteins, die die obigen Bedingungen erfüllen, gebildet. Die Verteilungen der Pseudoenergien der falschen Zuordnung  $H_i^{\text{falsch}}(E^{\text{Vol,sim}})$  und  $H_i^{\text{falsch}}(E^{\text{Shape,sim}})$  werden ebenfalls mit der gefilterten Signalliste, aber mit einer zufälligen Zuordnung erzeugt. Die Verteilungen der Pseudoenergien der simulierten Grundflächen wurden wieder in der AUREMOL Datenbank gespeichert. Analog zur Berechnung der gesamten globalen Pseudoenergie der experimentellen Spektren  $E^{\text{Vol,Shape}}$  (Gl. 58), wird für die Bewertung der Zuordnung eines Kerns  $k$  die logarithmierte Gesamtwahrscheinlichkeit, die wieder als Summe der logarithmierten Einzelwahrscheinlichkeiten der beteiligten simulierten Kreuzsignale geschrieben werden kann, verwendet. Um die für die Zuordnung wichtigen starken Signale stärker zu berücksichtigen, werden die logarithmierten Einzelwahrscheinlichkeiten der simulierten Signale  $|\ln(P_{ij}^{\text{Vol,Shape,sim}})|$ , wie die experimentellen Äquivalente, mit einem Faktor  $w_{ij}^{\text{Signal,sim}}$  (Gl. 68) gewichtet.

$$w_{ij}^{\text{Signal,sim}} = \frac{d^{\text{pseudo, max}}}{d_{kl}^{\text{pseudo}}} \quad (68)$$

Hier ist  $d^{\text{pseudo,max}}$  die maximale Pseudodistanz  $d_{kl}^{\text{pseudo}}$  (Gl. 67), die ein simuliertes Signal  $j$  im Spektrum haben darf, um für die Volumen- und Linienformpseudoenergie einer Zuordnung  $k$  berücksichtigt zu werden. Um zu verhindern, dass ein Spektrum  $i$  über- oder untergewichtet wird, werden, die simulierten Signale, wie bei der gesamten globalen experimentellen Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$ , nach dem Herkunftsspektrum gewichtet. Der Gewichtungsfaktor des simulierten Spektrums  $i$   $w_i^{\text{Spektrum,sim}}$  in der Volumen- und Linienformpseudoenergie einer Zuordnung wird mit Gl. 69, gebildet.

$$w_i^{\text{Spektrum,sim}} = 1 \quad \text{für } i=1$$

$$w_i^{\text{Spektrum,sim}} = 0,5 \frac{\sum_{j=1}^{N_i^{\text{Signale}}} w_{1j}^{\text{Signal,sim}}}{\sum_{j=1}^{N_i^{\text{Signale}}} w_{ij}^{\text{Signal,sim}}} \quad \text{sonst} \quad (69)$$

Für die Berechnung der Gewichtungsfaktoren der Spektren werden die Signale nur mit der 1. Regel, also falls sie kleines Signalvolumen haben, ausgeschlossen. Die normierte gewichtete Pseudoenergie  $E_k^{\text{Vol,Shape,sim}}$  für die Zuordnung des Kerns  $k$  ergibt sich dann zu:

$$E_k^{\text{Vol,Shape,sim}} = 1 - \frac{1}{Z_k} \sum_{ij} w_i^{\text{Spektrum,sim}} w_{ij}^{\text{Signal,sim}} |\ln(P_{ij}^{\text{Vol,Shape,sim}})| \quad (70)$$

Die Summe läuft nur über diejenigen Signale  $j$  in den Spektren  $i$ , an denen der Kern  $k$  beteiligt ist. Der minimale Wert der lokalen Wahrscheinlichkeiten für die richtige Zuordnung der simulierten Signale  $P_{ij}^{\text{Vol,Shape,sim}}$  wird, wegen des Erhalts der numerischen Stabilität des Logarithmus, wieder auf 0,05 gesetzt. Dadurch ist es auch wieder möglich die Pseudoenergie der Zuordnung des Kerns  $k$   $E_k^{\text{Vol,Shape,sim}}$  mit dem Normierungsfaktor  $Z_k$  auf Wertebereich zwischen 0 und 1 zu normieren. Der Normierungsfaktor wird folgendermaßen berechnet:

$$Z_k = \sum_{ij} w_i^{\text{Spektrum,sim}} w_{ij}^{\text{Signal,sim}} |\ln(0,05)| \quad (71)$$

Falls der Kern  $k$  keine sichtbaren NMR-Signale hat, kann seine Volumen- und Linienformpseudoenergie  $E_k^{\text{Vol,Shape,sim}}$  nicht ausgewertet werden. Sie wird in diesen Fällen auf 1 gesetzt (Begründung siehe Kap. 2.2.9). Da die lokale Wahrscheinlichkeit für die richtige Zuordnung der simulierten Grundfläche  $ij$   $P_{ij}^{\text{Vol,Shape,sim}}$ , wie schon das experimentelle Pendant  $P_{ij}^{\text{Vol,Shape}}$ , von minimal zwei chemischen Verschiebungen abhängig ist, kann die Pseudoenergie  $E_k^{\text{Vol,Shape,sim}}$  der Zuordnung  $k$  die Qualität nur akkurat messen, falls die chemischen Verschiebungen der Resonanzen der Kreuzsignalpartner größtenteils bekannt sind. Einen Teil der Unsicherheit die dadurch entsteht, kann aber durch den Einsatz des Satzes von Bayes und durch mehrere parallel laufende Instanzen von SIBASA abgefangen werden (siehe Kap. 2.2.15).

### 2.2.9 Behandlung von Kernen mit nicht sichtbaren Resonanzlinien

Durch Prozesse, wie chemischen Austausch kann es vorkommen, dass Kerne keine Signale im NMR-Spektrum erzeugen. Vor allem davon betroffen sind die für die Zuordnung von HSQC-Spektren wichtigen Rückgrat- und Seitenkettenamidprotonen. Um die Möglichkeit von nicht sichtbaren Resonanzlinien für die automatische Zuordnung berücksichtigen zu können, hat ASSIGN einen eigenen Slot eingeführt. Kerne, die in diesen Slot verschoben worden sind, haben keine Kreuzsignale auf den zurückgerechneten Spektren. Der sogenannte Dummslot konnte, wie alle anderen Slots auch, während der automatischen Zuordnung ausgewählt werden. ASSIGN vertraute darauf, dass Resonanzen von Kernen, die keine sichtbaren Kreuzsignale in den NMR-Spektren haben und nicht im Dummslot verschoben worden sind, in der richtigen Lösung stören würden, d.h. den Wert der gesamten Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  vergrößern würden. Für die Zuordnung von 2-D-NOESY-Spektren hat diese Annahme nur für simulierte Datensätze zugegriffen, den alle bisher eingeführten Pseudoenergien haben selbst Kerne mit sichtbaren Kreuzsignalen, die falsch zugeordnet wurden, kaum bestraft. Experimentelle NOESY-Kreuzsignale, die durch mehr als ein simuliertes Signal erklärt wurden, trugen selten negativ zur Gesamtpseudoenergie bei. Die Bewertung der Kerne im Dummslot bezüglich der Pseudoenergie der Optimierung musste deshalb für SIBASA verändert werden.

Zuerst wurde die Möglichkeit des Dummslots auf Resonanzlinien von Amidprotonen und an sie gebundene Stickstoffe eingeschränkt, da fast nur bei diesen Kernarten NMR-Kreuzsignale nicht sichtbar waren. Außerdem wurde eine Lücke, die ASSIGN bei der Berechnung der Pseudoenergie der chemischen Verschiebungen  $E^{\omega}$  gelassen hatte, geschlossen. Kerne, die in den Dummslot verschoben worden sind, erhalten in SIBASA nun die Wahrscheinlichkeit  $P_k(\text{Dummslot})$  von 1, was einen Beitrag von 0 zur Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  entspricht (Gl. 62).

Zu Beginn der Optimierung führt diese Maßnahme dazu, dass viele Amidgruppen keine Kreuzsignale auf den zurückgerechneten Spektren haben. Da der Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}$  (Gl. 64) im späteren Verlauf der automatischen Zuordnung stark verringert wird, reicht diese Maßnahme alleine nicht aus um die Amidgruppen ohne sichtbaren Kreuzsignale in den Spektren zu identifizieren. In SIBASA wurde deswegen, neben den Pseudoenergien des Signalvergleichs  $E^{\text{Vol,Shape}}$  (Gl. 58) und der chemischen Verschiebungen  $E^{\omega}$  (Gl. 62) eine dritte Pseudoenergie, die Pseudoenergie der nicht sichtbaren Resonanzlinien  $E^{\text{Dummy}}$  eingeführt. In ASSIGN, in dem diese Pseudoenergie noch nicht implementiert hatte, wurden während der automatischen Zuordnung die nicht eigentlich nicht sichtbaren simulierten Kreuzsignale so angeordnet, dass Rauschsignale, Artefakte im experimentellen Spektrum erklärt werden konnten. Auch die Möglichkeit durch Überlagerung von sichtbaren und nicht sichtbaren Kreuzsignalen die gesamte Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  zu verbessern, wurde immer wahrgenommen. Dabei wurde in Kauf

genommen, dass viele simulierte Signale über leerem spektralem Bereich lagen, also unerklärt blieben.

Die Volumen- und Linienformpseudoenergie der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  (Gl. 70) des Kerns  $k$  wird auf 1 gesetzt, falls er im Dummyslot ist. Die Pseudoenergie ist im Wertbereich zwischen 0 und 1 definiert, wobei ein Wert von 1 bedeutet, dass alle simulierten Signale des Kerns  $k$  perfekt durch die experimentellen NMR-Spektren erklärt werden können und ein Wert von 0 bedeutet, dass alle zurückgerechneten Signale des Kerns über leerem spektralem Bereich liegen. Die Zuordnungspseudoenergie  $E_k^{\text{Vol,Shape,sim}}$  erfüllt damit die Kriterien einer Pseudowahrscheinlichkeit, die wie anderen Wahrscheinlichkeiten in SIBASA zu globalen Pseudoenergien zusammengefasst werden können. Die Pseudoenergie des Dummyslots  $E^{\text{Dummy}}$  in SIBASA ist die Summe der Logarithmen der Pseudowahrscheinlichkeiten der Amid- und Stickstoffkerne:

$$E^{\text{Dummy}} = \sum_{k=0}^{N^{\text{Amide}} + N^{\text{Stickstoffe}}} |\ln(E_k^{\text{Vol,Shape,sim}})| \quad (72)$$

Die Summe läuft über alle Kerne von Amidgruppen, da nur sie in SIBASA in den Dummyslot verschoben werden können. Zur Erhaltung der numerischer Stabilität wird das Minimum von  $E_k^{\text{Vol,Shape,sim}}$  für diese Berechnung wieder auf 0,05 gesetzt.

Analog zur Pseudoenergie der chemischen Verschiebungen  $E^\omega$  muss die Pseudoenergie der Dummyslots  $E^{\text{Dummy}}$  gegen die anderen Pseudoenergien in der gesamten Energie der Optimierung  $E^{\text{total}}$  gewichtet werden. Der initiale Gewichtungsfaktor  $w^{E,\text{Dummy}}(t=0)$ , wird analog zum Gewichtungsfaktor der Pseudoenergie chemischen Verschiebung  $w^{E,\omega}(t=0)$  (Gl. 63), folgendermaßen gebildet:

$$w^{E,\text{Dummy}}(t=0) = A \frac{E^{\text{Vol,Shape,max}}}{(N^{\text{Amide}} + N^{\text{Stickstoffe}}) \ln(0,05)} \quad (73)$$

Hierbei ist  $A$  wieder die benutzerdefinierte Stärke und hat den selben Wert wie in Gl. 63 und Gl. 64. Da nur Amidprotonen und Stickstoffe im Dummyslot sein können, ist die maximal mögliche Pseudoenergie  $E^{\text{Dummy,max}}$  durch  $(N^{\text{Amide}} + N^{\text{Stickstoffe}}) \ln(0,05)$  gegeben. Der Gewichtungsfaktor wird außerdem während der automatischen Zuordnung regelmäßig an den momentanen Wert der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  angepasst:

$$w^{E,\text{Dummy}}(t) = A \frac{E^{\text{Vol,Shape}}(t) - E^{\text{Vol,Shape,opt}}}{(N^{\text{Amide}} + N^{\text{Stickstoffe}}) \ln(0,05)} \quad (74)$$

2.2.10 Threshold Accepting

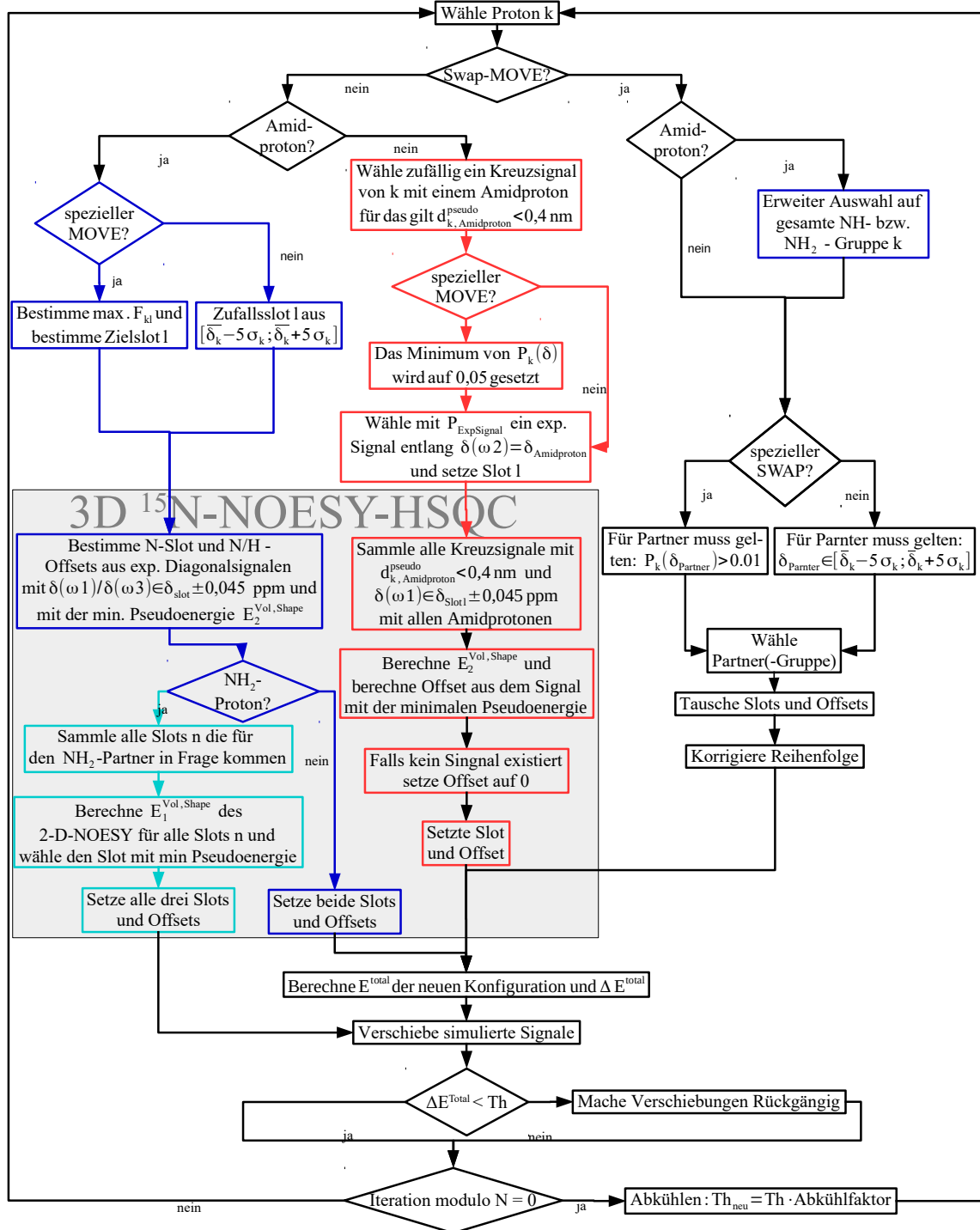


Abbildung 10: Schematische Darstellung der Optimierung von  $E^{total}$  mittels Threshold-Accepting während der automatischen Zuordnung in SIBASA.



Der eigentliche Optimierungsalgorithmus hat sich in SIBASA, gegenüber ASSIGN nicht verändert. Die vollständige Zuordnung wird durch Optimierung der gesamten Pseudoenergie  $E^{\text{total}}$  mit dem Threshold Accepting Algorithmus (Dueck und Scheuer 1990) erlangt. Threshold-Accepting ist eine Variante des Simulated Annealing Algorithmus (Kirkpatrick et al. 1983), der häufig für die physikalische Optimierung eingesetzt wird. Beide Verfahren können das Optimum einer Pseudoenergie in einem hochdimensionalen Konfigurationsraum finden.

Die Algorithmen gehen von einem beliebigen Anfangszustand, der auch zufällig sein kann, aus. Es werden wiederholt ein oder mehrere ausgewählte Parameter mehr oder weniger zufällig verändert und nach jeder Veränderung wird der Unterschied in der problembezogenen Pseudoenergie  $\Delta E^{\text{total}}$  gemessen. Die Veränderungen der Konfiguration werden übernommen falls gilt:

$$\Delta E^{\text{total}} = E^{\text{total}}(t+1) - E^{\text{total}}(t) < Th \quad (75)$$

Der Threshold  $Th$  ist das Äquivalent der Temperatur in einem physikalischen Prozess und sorgt bei Optimierungen dafür, dass der gesamte Konfigurationsraum abgesucht werden kann. Nach jeweils einer bestimmten Anzahl  $N$  von Veränderungen und Überprüfungen nach Gl. 75 wird der Threshold  $Th$  nach folgender Regel reduziert:

$$Th' = c Th \text{ mit } 0 < c < 1 \quad (76)$$

Dies nennt man logarithmisches Abkühlen. Die Abkühlrate  $c$ , die Anzahl  $N$  der versuchten Veränderungen pro Kühschritt und der Threshold  $Th$  sind problemspezifisch und müssen vorher abgeschätzt werden.  $Th$  wird solange reduziert bis sich keine Verbesserungen von  $E^{\text{total}}$  mehr einstellen. Werden keine Veränderungen mehr angenommen, so hat Threshold-Accepting-Algorithmus im Idealfall die Konfiguration mit minimalen Pseudoenergie  $E^{\text{total}}$  gefunden.

ASSIGN kannte zur Veränderung der chemischen Verschiebungen zwei Möglichkeiten:

- Der MOVE: Einem zufällig ausgewähltem Proton wird ein neue chemische Verschiebung zugeordnet. Der neue Wert der chemischen Verschiebung wurde aus der entsprechenden Normalverteilung mit SHIFTS Mittelwert und BMRB Standardabweichung (Gl. 60) und einer Zufallszahl berechnet.
- Der SWAP MOVE: Die chemischen Verschiebungen zweier zufällig ausgewählter Protonen wurden vertauscht.

In Abb. 10 wird der Ablauf Optimierung in SIBASA mit allen Anpassungen und Verbesserungen gezeigt. Der Algorithmus basiert grundsätzlich auf Threshold Accepting, aber die MOVES und SWAP MOVES wurden speziell auf die Zuordnung eines HSQC-Spektrums mithilfe eines 2-D-NOESY- und eines 3D 15N-NOESY-HSQC-Spektrums optimiert.

### 2.2.11 MOVES

Die von ASSIGN bekannten MOVES und SWAP MOVES für ein einziges homonukleares 2-D-NOESY-Spektrum wurden für SIBASA auf den Einsatz mit beliebig vielen N-dimensionalen Spektren erweitert und für die Zuordnung von HSQC-Spektren aus einem 2-D-NOESY und einem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum angepasst. Die MOVES unterscheiden nun zwischen Amidprotonen und Nichtamidprotonen. Für die beiden Arten von Protonen gibt es in SIBASA eine allgemeine und eine spezialisierte Methode für die Veränderung der chemischen Verschiebung während der automatischen Zuordnung. Die speziellen MOVES sind an das Problem angepasst und erhöht die Geschwindigkeit und Zuverlässigkeit des Algorithmus. Da sie aber stark von der PDB Struktur, der Genauigkeit der Rückrechnung und der Verlässlichkeit der Vorhersagen der chemischen Verschiebungen abhängig sind, finden sie die chemischen Verschiebungen von Ausreißern meistens nicht. Die allgemeinen MOVES dagegen gewichten den Möglichkeitsraum nicht und schränken ihn auch weit weniger ein. Die Entscheidung ob die allgemeine oder der spezielle Methode verwendet wird, wird über den Kontrollparameter  $\kappa$  kontrolliert. Er gibt an wie hoch der Anteil der speziellen MOVES zu einem Zeitpunkt in automatischen Zuordnung sein soll. Zu Beginn der automatischen Zuordnung hat es sich bewährt ihn so einzustellen, dass in 90 % der Fälle die spezielle Methode, die versucht möglichst viel Vorwissen aus den Eingabespektren, der Topologie des Proteins und den statistischen Daten einzubeziehen, verwendet wird. Der Kontrollparameter  $\kappa$  wird, wie der Threshold  $\text{Th}$ , nach jeweils  $N$  Schritten logarithmisch verkleinert (Abb. 10). Ein hoher Anteil von speziellen MOVES sorgt dafür, dass die Mehrheit der Resonanzen, die mit vorhanden Vorwissen einfach zuzuordnen sind, auch schnell zugeordnet werden. Im späteren Verlauf der automatischen Zuordnung konzentriert sich SIBASA durch die Erhöhung des Anteils der allgemeinen MOVES, die einen größeren spektralen Bereich abdecken, auf Ausreißer, die sich im Netzwerk der bereits bestimmten Zuordnungen auch besser einfügen lassen.

### Auswahl der Zielslots für Amidprotonen

Ein Proton das kovalent an ein Stickstoff gebunden ist, nimmt im zweiten für die automatische Zuordnung von HSQC-Signalen benötigten Spektrum, dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum eine besondere Rolle ein und kann deshalb von SIBASA besser behandelt werden als ein Proton das nicht skalar an ein Stickstoff gebunden ist. Falls ein Amidproton nicht in den Dummslot verschoben werden soll, wird eine der beiden Methoden zur Bestimmung des Zielslots verwendet:

1. Die allgemeine Methode: Der Zielslot wird rein zufällig aus dem chemischen Verschiebungsbereich  $[\bar{\delta}_{\text{ResNr, Kern}} - a\sigma_{\text{AA, Kern}}^{\text{BMRB}}; \bar{\delta}_{\text{ResNr, Kern}} + a\sigma_{\text{AA, Kern}}^{\text{BMRB}}]$  ausgewählt, wobei

$\bar{\delta}_{ResNr, Kern}$  ( $ResNr$  ist die Position der Aminosäure in der Primärsequenz und  $Kern$  das betrachte Amidproton wie z.B.  $H^N$ ,  $H^{\delta 21}$ , ...) die mittlere chemische Verschiebung des ausgewählten Amidprotons (Kap. 2.2.7),  $\sigma_{AA, Kern}^{BMRB}$  ( $AA$  ist die Aminosäure wie z.B. Ala, ...) die Standardabweichung aus der BMRB und  $a$  ein vorgegebener Multiplikator ist. Um alle Ausreißer berücksichtigen zu können, sollte der Multiplikator  $a$  mindestens auf 5 eingestellt werden.

- Die spezielle Methode: Es werden Pseudokräfte, die auf das zu verschiebende Amidproton wirken, von alle Slots im chemischen Verschiebungsbereich  $[\bar{\delta}_{ResNr, Kern} - a\sigma_{AA, Kern}^{BMRB}; \bar{\delta}_{ResNr, Kern} + a\sigma_{AA, Kern}^{BMRB}]$  ( $a$ ,  $\sigma_{AA, Kern}^{BMRB}$ ,  $ResNr$ ,  $AA$  und  $Kern$  sind wie in der allgemeinen Methode definiert) bestimmt. Der Slot der die größte Pseudokraft auf das Amidproton hat, wird als Zielslot ausgewählt. Falls sich dabei der Slot nicht verändern würde, wird der Slot mit der zweitstärksten Kraft verwendet.

Die Entscheidung, ob die allgemeine oder die spezielle Methode verwendet wird, wird statistisch mithilfe des Kontrollparameters  $\kappa$  (Kap. 2.2.11) getroffen.

### Filterung der möglichen Slots für Amidprotonen

Vor der Optimierung der chemischen Verschiebungen in den zurückgerechneten NOESY-Spektren filtert SIBASA für jedes Amidproton die Slots, die die zurückgerechneten Kreuzsignale des Amidprotons auf der indirekten  $^1H$ -Dimension des 2-D-NOESY-Spektrums nicht erklären können, heraus. Für alle simulierten Kreuzsignale des betrachteten Amidprotons mit dem Kern  $k$  im 2-D-NOESY-Spektrums, die einen Pseudoabstand  $d_{Amidproton, k}^{Pseudo} < 0,4$  nm (Gl. 67 in Kap. ) haben, werden mögliche experimentelle Signalzuordnung auf allen Slots gesucht. Sollte kein experimentelles Signal im 2-D-NOESY-Spektrum die Strecke zwischen den Koordinaten  $(\bar{\delta}_k - a\sigma_k; \delta_{Slot})$  und  $(\bar{\delta}_k + a\sigma_k; \delta_{Slot})$  ( $\bar{\delta}_k$  ist die durchschnittliche und  $\sigma_k$  die Standardabweichung des chemischen Verschiebung des Kerns  $k$  und  $\delta_{Slot}$  die chemische Verschiebung des betrachteten Slots  $Slot$ ) auf mindestens halber Höhe schneiden, so wird dieser Slot als Ziel für den speziellen MOVE des Amidprotons ausgeschlossen. Das aufsummierte experimentelle Volumen der NOESY-Kreuzsignale auf diesen Strecken muss außerdem mindestens 80 % des skalierten simulierten Volumens des Kreuzsignals betragen. Zurückgerechnete Kreuzsignale, deren Strecken sich ganz oder teilweise überlagern werden zu Cluster zusammengefasst und ihr summiertes Volumen wird, statt dem skalierten Volumens eines Kreuzsignals, als Kriterium verwendet (Abb. 11).

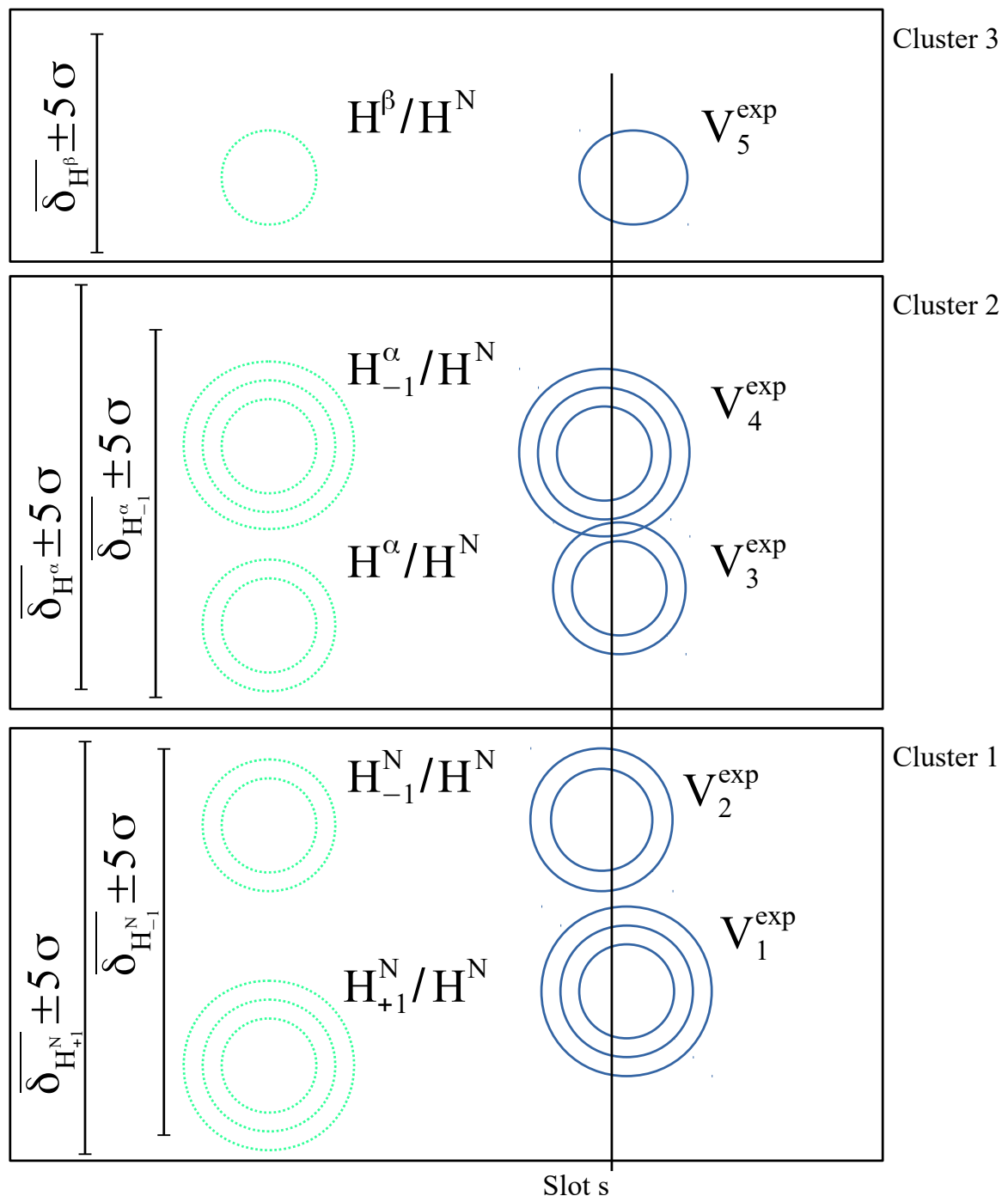


Abbildung 11: Schematische Darstellung der experimentellen 2-D-NOESY-Kreuzsignale entlang eines für den speziellen MOVE des Amidprotons  $H^N$  erlaubten Slots  $s$ . Für jedes simulierte Signal des Amidprotons  $H^N$  mit dem zugrundeliegenden Pseudoabstand  $d^{\text{pseudo}} < 0,4 \text{ nm}$  (grüne Signale) muss im spektralen Bereich  $\bar{\delta} \pm 5\sigma$  ( $\bar{\delta}$  ist die mittlere und  $\sigma$  die Standardabweichung der chemischen Verschiebung der Kreuzsignalpartner von  $H^N$ ) mindestens ein experimentelles Signal (blau) existieren. In den Clustern (schwarze Rechtecke) muss das experimentelle Volumen ( $V_1^{\text{exp}} + V_2^{\text{exp}}$ ,  $V_3^{\text{exp}} + V_4^{\text{exp}}$ ,  $V_5^{\text{exp}}$ ) höher als 80 % des gesamten simulierten Volumen ( $V_{H_{+1}^N/H^N} + V_{H_{-1}^N/H^N}$ ,  $V_{H^\alpha/H^N} + V_{H_{-1}^\alpha/H^N}$ ,  $V_{H^\beta/H^N}$ ) sein.

## Pseudokräfte

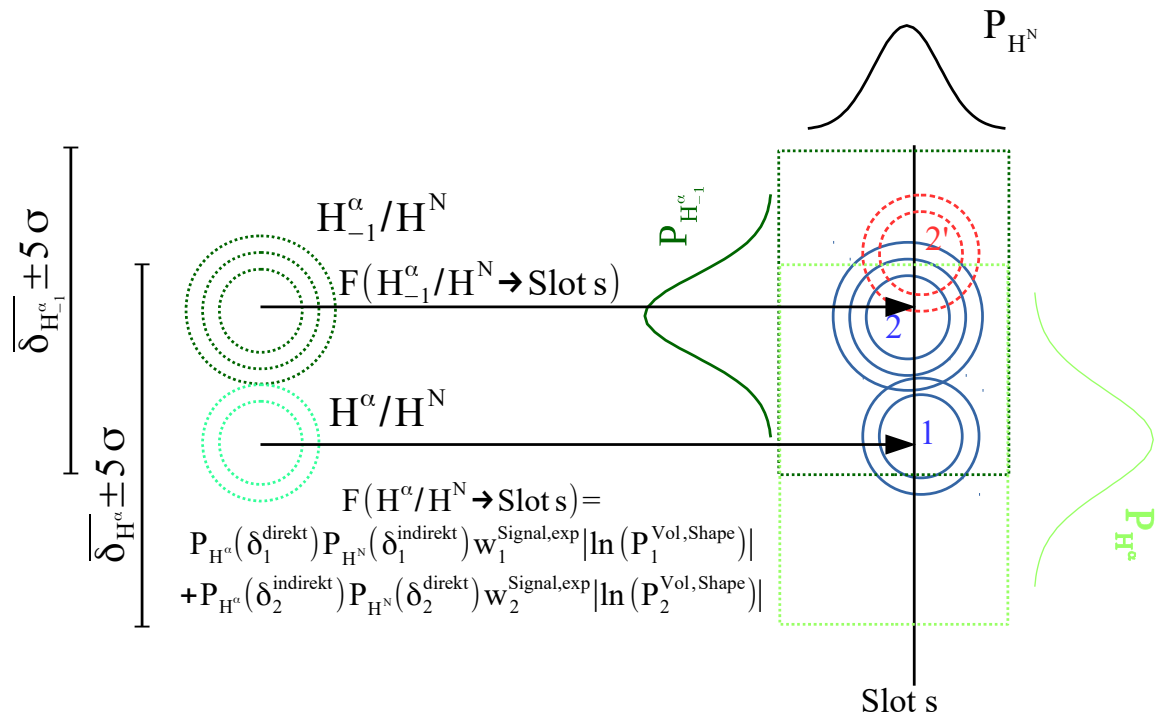


Abbildung 12: **Beispiel für die Berechnung der Pseudokraft die vom Slot s auf die Resonanz des Amidprotons  $H^N$  ausgeht.** Die experimentellen Signale 1 und 2 ziehen  $H^N$  an, wobei die Kraft die von Signal 2 ausgeht durch das dort bereits liegende simulierte Signal 2' teilweise abgeschirmt wird. Die Kräfte die von den Signalen 1 und 2 auf die starken simulierten Kreuzsignale  $H^\alpha/H^N$  bzw.  $H_{-1}^\alpha/H^N$  ausgehen, werden mit den Produkt der Wahrscheinlichkeiten der chemischen Verschiebungen des Zielorts ( $P_{H^N} P_{H^\alpha}$  bzw.  $P_{H^N} P_{H_{-1}^\alpha}$ ) gewichtet.

Für die Bestimmung der Ziele des speziellen MOVES von Amidprotonen werden die Pseudokräfte von allen nach der Filterung verbleibenden Slots auf das Amidproton  $k$  bestimmt. In die Berechnung der Pseudokraft eines Zielslots  $s$  auf das Amidproton  $k$   $F^{\text{Pseudo}}(\text{Slot } s \rightarrow \text{Amidproton } k)$  fließt die Wahrscheinlichkeiten, dass ein simuliertes Kreuzsignal einem experimentellen Kreuzsignal richtig zugeordnet werden kann und die momentane Besetzung des Zielslots  $s$  mit simulierten Signalen ein. Die Wahrscheinlichkeiten können in Abhängigkeit von den chemischen Verschiebungen des experimentellen Signals und den beteiligten Kernen des simulierten Kreuzsignals folgendermaßen berechnet werden:

$$P(\text{sim Signal} \rightarrow \text{exp Signal}) = \prod_n A_{k(n)} f_{k(n)}(\delta_n) \quad (77)$$

$\delta_n$  ist die chemische Verschiebung des experimentellen Signals in der Dimension  $n$  des 2-D-NOESY-Spektrums,  $k(n)$  der Index des Kerns des simulierten Signals in der Dimension  $n$  und

$A_{k(n)}f_{k(n)}(\delta_n)$  die Wahrscheinlichkeit der chemischen Verschiebung  $\delta_n$  des Kerns  $k(n)$  (Kap. 2.2.7). Der Wert von  $P(\text{sim Signal} \rightarrow \text{exp Signal})$  darf, wegen der numerischen Stabilität, wieder minimal 0,05 betragen. Die Pseudokraft  $F^{\text{Pseudo}}$  (Slot  $s \rightarrow$  Amidproton  $k$ ) des Slots  $s$  auf das Amidproton  $k$  berechnet sich folgendermaßen (Abb. 12):

$$F^{\text{Pseudo}}(\text{Slot } s \rightarrow \text{Amidproton } k) = \sum_{ij} P(i \rightarrow j) w_{1j} |\ln(P_{1j}^{\text{Vol, Shape}})| \quad (78)$$

$i$  ist der Index der simulierten Kreuzsignale des Amidprotons  $k$  in der direkten Dimension des 2-D-NOESY-Spektrums, die einem Pseudoabstand  $d^{\text{pseudo}} < 0,4$  nm haben.  $j$  ist der Index der experimentellen Kreuzsignale, die in der direkten Dimension den Slot  $s$  auf mindestens halber Höhe schneiden.  $P(i \rightarrow j)$  ist die Zuordnungswahrscheinlichkeit (Gl. 77). Die Abhängigkeit von der gewichten logarithmierten lokalen Wahrscheinlichkeit der richtigen Zuordnung des experimentellen Signals  $j$  im 2-D-NOESY-Spektrum  $w_{2\text{-D-NOESY},j} |\ln(P_{2\text{-D-NOESY},j}^{\text{Vol, Shape}})|$  (Gl. 54) sorgt dafür, dass nicht zugeordnete experimentelle Signale eine größere Kraftwirkung haben als bereits besetzte.

## Auswahl der Slots für H $^\alpha$ - und Seitenkettenprotonen

Protonen, die nicht kovalent an ein Stickstoff gebunden sind, haben fast immer sichtbare NOESY-Kreuzsignale und werden von SIBASA deshalb nie in den Dummyslot verschoben. Die Resonanz eines H $^\alpha$ - oder Seitenkettenproton  $k$  wird in SIBASA immer so verschoben, dass ein zufällig gewähltes starkes simuliertes Kreuzsignal des Protons mit einem Amidproton ( $d_{k, \text{Amidproton}}^{\text{pseudo}} < 0,4$  nm) auf ein experimentelles 2-D-NOESY-Kreuzsignal verschoben wird.

Als Ziele kommen die experimentellen Kreuzsignale in Frage, die die den momentanen Slot des ausgewählten Amidprotons *Amidproton* auf mindestens halber Höhe entlang der indirekten Dimension schneiden. Die Wahrscheinlichkeit  $P_{\text{ExpSignal}}^{\text{Auswahl}}$  für das experimentellen Kreuzsignal *ExpSignal* ausgewählt zu werden, wird folgendermaßen bestimmt:

$$P_{\text{ExpSignal}}^{\text{Auswahl}} = \frac{1}{Z} P_k(\delta^{\text{indirekt}}) w_{\text{ExpSignal}}^{E, \text{Signal}} |\ln(P_{\text{ExpSignal}}^{\text{Vol, Shape}})| \quad (79)$$

$\delta^{\text{indirekt}}$  ist die chemische Verschiebung des experimentellen Signals entlang der indirekten  $^1\text{H}$ -Dimension,  $P_k(\delta)$  die Wahrscheinlichkeit der chemischen Verschiebung  $\delta$  für den Kern  $k$  (Kap. 2.2.7),  $w_{\text{ExpSignal}}^{E, \text{Signal}} |\ln(P_{\text{ExpSignal}}^{\text{Vol, Shape}})|$  die gewichte logarithmierte lokale Wahrscheinlichkeit der richtigen Zuordnung des Kreuzsignals *ExpSignal* und  $Z$  der Normierungsfaktor, der folgendermaßen berechnet wird:

$$Z = \sum_{\text{ExpSignal}} A_k P_k(\delta_{\text{ExpSignal}}^{\text{indirekt}}) w_{\text{ExpSignal}}^{E, \text{Signal}} |\ln(P_{\text{ExpSignal}}^{\text{Vol, Shape}})| \quad (80)$$



Der Kontrollparameter  $\kappa$  kontrolliert auch in diesem Fall den Anteil der speziellen und der allgemeinen MOVES. Beim allgemeinen MOVE wird die minimal mögliche Wahrscheinlichkeit der chemischen Verschiebung  $P_k(\delta^{\text{indirekt}})$  auf 0,05 gesetzt, im speziellen Fall kann sie dagegen bis auf 0 sinken und so die Auswahl von unwahrscheinlichen Zuordnungen verhindern. Da die chemische Verschiebungen immer aus der indirekten  $^1\text{H}$ -Dimension des 2-D-NOESY-Spektrums stammen und diese in der Regel schlechter als die direkte Dimension aufgelöst ist, existieren in der Regel mehrere Slots, die das Maximum des ausgewählten experimentellen Signal *ExpSignal* in der indirekten Dimension schneiden. Um den optimalen Slot für den MOVE zu finden, wird für jeden dieser Zielslots die gesamte Pseudoenergie des Signalvergleichs auf dem 2-D-NOESY-Spektrum  $E_{2\text{-D-NOESY}}^{\text{Vol, Shape}}$  (Gl. 56) berechnet und der Slot, der die niedrigste Pseudoenergie erzeugt hat, als Ziel übernommen. Würde sich der momentane Slot des  $\text{H}^\alpha$ - oder Seitenkettenprotons durch diesen MOVE nicht verändern, so wird die gesamte Auswahlprozedur wiederholt.

Sollte kein starkes simuliertes Kreuzsignal des  $\text{H}^\alpha$ - oder Seitenkettenprotons  $k$  mit einem Amidproton existieren, was z.B. bei manchen äußeren Seitenkettenprotonen vorkommen kann, so wird der normale MOVE aus ASSIGN (Kap. 2.2.10) mit den neuen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen im speziellen Fall und die Methode, mit der Amidprotonen im allgemeinen Fall in SIBASA verschoben werden, als allgemeiner MOVE verwendet.

## Anpassung an das 3-D-NOESY-Spektrum

Bisher wurden nur die MOVES der Resonanzen von Protonen im [ $^1\text{H}$ - $^1\text{H}$ ] 2-D-NOESY-Spektrum behandelt. Die chemischen Verschiebungen der Stickstoffe müssen aber mit dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum bestimmt werden. Das 3-D-NOESY-Spektrum liefert außerdem zusätzliche Informationen über die Zuordnung der Resonanzen der Protonen, da weniger Signalüberlapp in diesen Spektren vorhanden ist. Für die Zuordnung einer Resonanz in mehreren Spektren müssen zusätzliche Parameter eingeführt und bestimmt werden. Die sogenannten Offsets der chemischen Verschiebungen sind in SIBASA die Differenz der chemischen Verschiebung einer Resonanz des Kerns  $k$  zwischen einer Dimension in einem Spektrum und der direkten  $^1\text{H}$ -Dimension des 2-D-NOESY-Spektrums.



### Bestimmung der Zielslots von Stickstoffatomen und der Offsets der Resonanzen NH-Gruppen

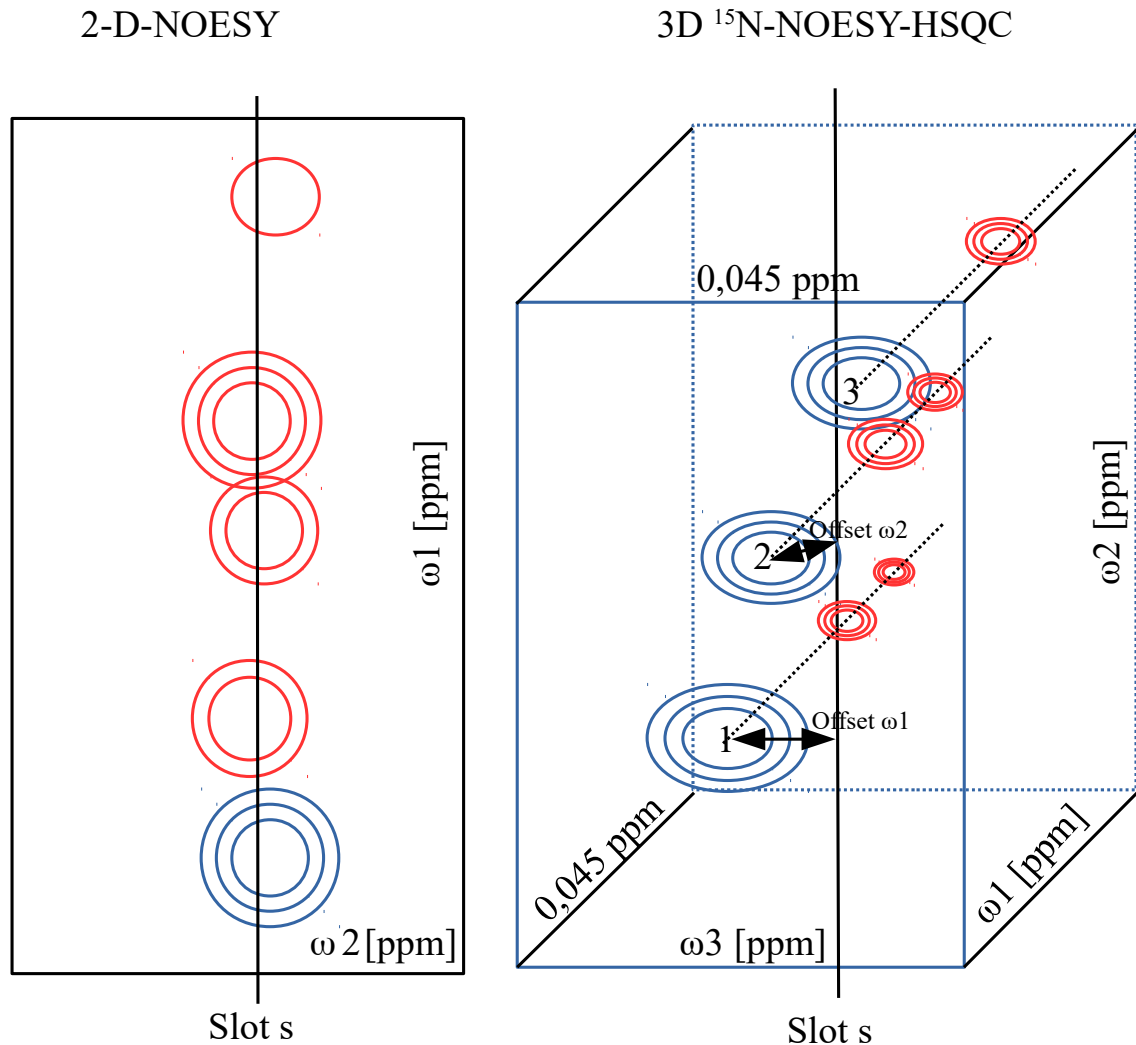


Abbildung 14: Schematische Darstellung der Bestimmung der Zielslots der Stickstoffatome und der Offsets der Amidprotonen mit Hilfe des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Nach der Auswahl des Slots  $s$  für das Amidproton werden für alle  $\omega_1\omega_3$ -Diagonalsignale des 3-D-Spektrums, deren Abstand zum Slot  $s$  in keiner  $^1\text{H}$ -Dimension größer als 0,045 ppm ist (Signale 1, 2 und 3), die Slots der  $^{15}\text{N}$ -Dimension und die Offsets der beiden  $^1\text{H}$ -Dimensionen bestimmt. Für alle in Frage kommenden Slot/Offset-Kombinationen wird die Pseudoenergie des Signalvergleichs  $E_{3\text{-D NOESY}}^{\text{Vol, Shape}}$  auf dem 3-D-NOESY-Spektrum berechnet. Die Pseudoenergie berücksichtigt auch die Kreuzsignale (rot). Der  $\omega_1$ - und  $\omega_3$ -Offsets, sowie der Zielslot für das gebundene Stickstoff werden vom Slot/Offset-Kombination mit der besten Pseudoenergie übernommen.

Da SIBASA auf die Zuordnung von HSQC-Spektren optimiert wurde, verschiebt es die Slots von NH- oder  $\text{NH}_2$ -Gruppen zusammen auf experimentelle HSQC-Signale. Nachdem der Zielslot  $s$  eines Amidprotons festgelegt wurde, wird auf Basis dieses Slots  $s$  und des 3D

$^{15}\text{N}$ -NOESY-HSQC-Spektrums, das in der  $\omega_2\omega_3$ -Summenprojektion ein 2D  $^{15}\text{N}$ -HSQC-Spektrum ist, sowohl der Zielslot des kovalent gebunden Stickstoffs als auch der des zweiten Amidprotons einer  $\text{NH}_2$ -Gruppe bestimmt. Während der Vorbereitung wurden alle experimentellen  $^1\text{H}$ - $^1\text{H}$  Diagonalsignale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum, also die Signale mit einer maximalen chemischen Verschiebungsdifferenz  $|\delta(\omega_1) - \delta(\omega_3)|$  von 0,12 ppm, gesammelt. Diese Signale sind die stärksten Repräsentanten der 2D  $^{15}\text{N}$ -HSQC-Signale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum und sind für alle zuzuordnenden HSQC-Signale vorhanden. Die gesammelten Diagonalsignale geben den Möglichkeitsraum für die Verschiebung Amidgruppen vor. Wegen dem starken Überlapp auf der  $\omega_1\omega_3$ -Diagonalen muss der Möglichkeitsraum durch zusätzliche Zielsignale erweitert werden, und zwar um 3D  $^{15}\text{N}$ -NOESY-HSQC-Kreuzsignale, die nicht von den bereits gesammelten Zielsignale erklärt werden.

Für die Fortsetzung eines MOVES eines Hauptkettenamidprotons mit dem Zielslot  $s$  im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum kommen nur Zielsignale in Frage, für die die chemische Verschiebungsdifferenz  $|\delta(\text{Slot } s) - \delta_{\text{Zielsignal}}(\omega_3)|$  kleiner als 0,045 ppm ist. Die Toleranz von 0,045 ppm, welche gleichzeitig der maximale Offset ist, berücksichtigt die potentiellen Unterschiede der chemischen Verschiebungen der Resonanzen eines Amidprotons in verschiedenen NMR-Experimenten unter gleichen Experimentbedingungen. Der maximale Wert von 0,045 ppm ist eine Abschätzung des maximal zu erwartenden Offset in der  $\omega_3$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Für alle verbliebenen Ziele, die im Bereich des maximalen Offsets von 0,045 ppm sind, werden die Zielslots des gebundene Stickstoffs und die Offsets des Amidprotons bestimmt. Für diese Slot/Offset-Kombinationen wird jeweils die Pseudoenergie des Signalvergleichs  $E_{3\text{-D-NOESY}}^{\text{Vol, Shape}}$  auf dem gesamten 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum berechnet. Der Slot des Stickstoffs und die Offsets des Amidprotons in der  $\omega_1$ - und der  $\omega_3$ -Dimension werden von dem Zielsignal, das das die beste Pseudoenergie erzeugen würde, übernommen. Gibt es keine passenden Ziele im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum zum initialen Slot  $s$  des Amidprotons, so wird die gesamte Auswahlprozedur wiederholt.

### **Bestimmung der Zielslots von Stickstoffatomen und der Offsets von $\text{NH}_2$ -Gruppen**

Für  $\text{NH}_2$ -Gruppen wird analog zu der Liste Zielsignale der Hauptkettenamidgruppen, eine separate Liste aus dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum erzeugt. Es werden zuerst Paare von Kreuzsignalen, die dem Muster der gegenseitigen Kreuzsignale von  $\text{NH}_2$ -Gruppen untereinander im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum entsprechen, gesucht. Zwei Signale  $i$  und  $j$  müssen in diesem Spektrum dafür folgende Bedingungen erfüllen:

- Die Signale  $i$  und  $j$  haben die gleiche chemische Verschiebung in der Stickstoffdimension, also  $\delta_i(\omega_2) = \delta_j(\omega_2)$ .

- Aus statistischen Daten ist bekannt, dass die maximale chemische Verschiebungsdifferenz der beiden NH<sub>2</sub>-Protonen nicht größer als 2,5 ppm ist. Deshalb muss für jedes Signal gelten:  $|\delta(\omega 1) - \delta(\omega 3)| < 2,5 \text{ ppm}$ .
- Da die beiden an der  $\omega 1 \omega 3$ -Diagonale gespiegelten Kreuzsignale zweier Protonen gesucht werden, muss für die Differenzen der chemischen Verschiebungen gelten:  $|\delta_i(\omega 1) - \delta_j(\omega 3)| < 0,12 \text{ ppm}$  und  $|\delta_i(\omega 3) - \delta_j(\omega 1)| < 0,12 \text{ ppm}$ .

Wie zuvor bei den Hauptketten, wird auch bei den NH<sub>2</sub>-Gruppen die maximale Toleranz der chemischen Verschiebungsdifferenz für die Aufnahme in die Liste der Zielsignale auf 0,12 ppm festgelegt. Eine weitere mögliche Einschränkung, die nur starke 3-D-NOESY-Kreuzsignale berücksichtigen würde, hat sich nicht in allen Fällen bewährt. Zwar haben die gegenseitigen Kreuzsignale von NH<sub>2</sub>-Protonen, wegen ihrem kleinen Abstand ungefähr 0,18 nm in der 3D-Struktur, sehr hohe NOE-Intensitäten, die aber in vielen Fällen wegen chemischen Austausch stark abgeschwächt auftreten können. Diese Kreuzsignale würden deshalb dem Auswahlprozess entgehen.

Die verbleibenden Ziele liefern, wie schon bei den Hauptkettenamidgruppen, die Offsets der chemischen Verschiebung des ersten der beiden NH<sub>2</sub>-Protonen im 3D <sup>15</sup>N-NOESY-HSQC-Spektrum und den Slot des zugehörigen Seitenkettenstickstoffs. Die möglichen Slots des zweiten Amidprotons der NH<sub>2</sub>-Gruppe wird mit dem bereits ausgewählten Zielslots und Offsets des ersten Protons und den oben genannten Regeln bestimmt. Für jedes verbleibende Ziel wird die Pseudoenergie des Volumen- und Linienformvergleichs  $E_{2\text{-D-NOESY}}^{\text{Vol, Shape}}$  auf führenden 2-D-NOESY-Spektrums bestimmt, da auf diesem Spektrum das Kreuzsignalmuster von NH<sub>2</sub>-Protonen, wegen dem besseren Signal-zu-Rausch Verhältnis, stärker ausgeprägt ist.

### Bestimmung der Offsets von H<sup>α</sup>- und Seitenkettenprotonen

Der Zielslot  $s$  für die chemische Verschiebung eines Nichtamidprotons  $k$  wurde so ausgewählt, dass eines der simulierten Kreuzsignale des Nichtamidprotons  $k$  auf ein experimentelles 2-D-NOESY-Kreuzsignal gelegt wird. Für den Offset in der indirekten <sup>1</sup>H-Dimension des 3D <sup>15</sup>N-NOESY-HSQC-Spektrum werden alle starken simulierten Kreuzsignale ( $d_{k, \text{Amid}}^{\text{Pseudo}} < 0,4 \text{ nm}$ ) des Nichtamidprotons  $k$  mit Amidprotonen im 3D <sup>15</sup>N-NOESY-HSQC-Spektrum gesammelt. Um als Zielsignal in Frage zu kommen, muss für deren chemische Verschiebungsdifferenzen zu einem beliebigen experimentellen 3D <sup>15</sup>N-NOESY-HSQC-Kreuzsignal folgendes gelten:

- $|\delta^{\text{exp}}(\omega 1) - \delta(\text{Slot } s)| < 0,12 \text{ ppm}$
- $|\delta^{\text{exp}}(\omega 2) - \delta^{\text{sim}}(\omega 2)| < 0,12 \text{ ppm}$
- $|\delta^{\text{exp}}(\omega 3) - \delta^{\text{sim}}(\omega 3)| < 0,12 \text{ ppm}$

Hierbei sind  $\delta^{\text{exp}}(\omega)$  die chemischen Verschiebungen eines experimentellen Kreuzsignals in der entsprechenden Dimension des Spektrums,  $\delta^{\text{sim}}(\omega)$  die die chemischen Verschiebungen des simulierten Kreuzsignals und  $\delta(\text{Slot } s)$  die chemische Verschiebung des vorher gewählten Zielslots  $s$ . Die maximale Toleranz von 0,12 ppm bestimmt sowohl den maximalen Abstand des experimentellen Signal zur chemischen Verschiebung des Zielslots  $s$  in der indirekten  $^1\text{H}$ -Dimension als auch die maximal erlaubten Abstände der momentanen chemischen Verschiebungen des Amidprotons und des Stickstoffs zu den  $\omega_3\omega_2$ -Verschiebungen des experimentellen Signals.

Für jedes nach dieser Filterung verbleibende experimentelle Zielsignal wird die Pseudoenergie des Signalvergleichs  $E_{3\text{-D-NOESY}}^{\text{Vol, Shape}}$  auf dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum bestimmt. Der Offset des Nichtamidprotons  $k$  wird aus der chemischen Verschiebung  $\delta^{\text{exp}}(\omega_1)$  des Zielsignals, das die geringste Pseudoenergie erzeugt hat, berechnet. Existieren unter den oben genannten Bedingungen keine experimentellen Signale, so wird der MOVE trotzdem durchgeführt, aber der Offset auf 0 ppm gesetzt.

## Der SWAP MOVE

Der SWAP MOVE aus ASSIGN (Kap. 2.2.10) wurde an für die Zuordnung von HSQC-Spektren angepasst. In SIBASA müssen die chemischen Verschiebungen und deren Offsets im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum gleichzeitig getauscht werden. In ASSIGN wurden noch zwei Protonen zufällig für einen Positionstausch ausgewählt, wobei die beiden neuen Positionen der beiden Protonen  $k$  jeweils im Bereich  $[\delta_k^{\text{SHIFTS}} - 3\sigma_{\text{AA, Kern}}^{\text{BMRB}}; \delta_k^{\text{SHIFTS}} + 3\sigma_{\text{AA, Kern}}^{\text{BMRB}}]$  liegen mussten. Hierbei ist die chemische Verschiebung  $\delta_k^{\text{SHIFTS}}$  die von SHIFTS aus der 3D-Struktur geschätzte chemische Verschiebung des Protons  $k$  und  $\sigma_{\text{AA, Kern}}^{\text{BMRB}}$  die Standardabweichung der chemischen Verschiebung der Aminosäure  $\text{AA}$  und der Kernart  $\text{Kern}$  aus der BMRB (Kap. 2.2.7).

In SIBASA wird mithilfe des Kontrollparameters  $\kappa$ , der Anteil von problemspezifischen und allgemeinen SWAP MOVES gesteuert (Kap. 2.2.11). Beim allgemeinen SWAP MOVE müssen die chemischen Verschiebungen aller Zielslots im erweiterten Bereich  $[\bar{\delta}_k - a\sigma_{\text{AA, Kern}}^{\text{BMRB}}; \bar{\delta}_k + a\sigma_{\text{AA, Kern}}^{\text{BMRB}}]$  der chemischen Verschiebung liegen, beim speziellen SWAP MOVE im Bereich der mit dem Segmentationsniveau von 0,01 abgeschnittenen Wahrscheinlichkeitsverteilung  $P_k(\delta)$  aus Kapitel 2.2.7. Dabei ist  $\bar{\delta}_k$  der Mittelwert der von SHIFTX2 aus dem MD-Ensemble geschätzten chemischen Verschiebungen und der Faktor  $a$  ein Multiplikator, der die Breite des Intervalls vorgibt. Es hat sich bewährt  $a$  wird, wie schon beim allgemeinen MOVE für Amidprotonen, auf 5 zu setzen.

Unter Beachtung dieser Regeln werden nur Hauptkettenamidgruppen mit Hauptkettenamidgruppen,  $\text{NH}_2$ -Gruppen mit  $\text{NH}_2$ -Gruppen und Nichtamidprotonen mit

anderen Nichtamidprotonen getauscht.

Zusätzlich kann der SWAP MOVE falls nötig die Reihenfolge der chemischen Verschiebungen der Resonanzen wieder herstellen. Durch falsche Offsets kann es vorkommen, dass die Reihenfolge zweier chemischen Verschiebungen im 2-D- und 3-D-Spektrum vertauscht ist. Seien  $o_i$  und  $o_j$  die Offsets der Vertauschungspartner  $i$  und  $j$  und  $\delta_i$  und  $\delta_j$  ihre chemischen Verschiebungen im 2-D-NOESY-Spektrum. Falls also  $\delta_i < \delta_j$  und  $\delta_i + o_i > \delta_j + o_j$  ist oder der umgekehrte Fall zutrifft, so werden die Slots weiterhin aber die Offsets nicht getauscht.

### 2.2.12 Verifikation der Zuordnungen

Beim Threshold Accepting handelt sich um einen Monte-Carlo Algorithmus (Müller-Gronbach et al. 2012), dessen Ergebnisse von den initialen Konfigurationen und von den Zufallszahlen abhängen können. Aus einem Optimierungslauf alleine ist in der Regel schwer festzustellen, ob das absolute Minimum der Pseudoenergie und damit im Idealfall die richtige Lösung gefunden wurde. Auch in SIBASA können sich Teile der Ergebnisse in verschiedenen Läufen unterscheiden. Eine Verifikation ist notwendig, um die verlässlichen Teile der Lösung zu identifizieren.

Eine Möglichkeit der Verifikation bei einem Monte-Carlo-Algorithmus ist die Auswertung mehrere Läufe, die mit gleichen Eingabedaten aber mit unterschiedlichen Startkonfigurationen und Zufallszahlen, gestartet wurde. Da das Muster der NOESY-Kreuzsignale eines Proteins einzigartig ist und durch die Rückrechnung im wesentlichen sicher reproduziert werden kann, sind Zuordnungen von Resonanzen, die in allen SIBASA Instanzen gleich sind, auch mit hoher Wahrscheinlichkeit richtig.

Neben der Verifikation der einzelnen Zuordnungen wird durch das unabhängige parallele Optimieren in verschiedenen Instanzen in der Regel auch eine Lösung mit niedrigerer Gesamtpseudoenergie  $E^{\text{total}}$  (Gl. 44) als für einen einzigen Lauf gefunden. Um ein gleich gutes Ergebnis mit einem Lauf zu erzielen, müsste die Rechenzeit dieser Instanz sehr stark erhöht werden. Diese Tatsache wird z.B. auch bei der NMR-Strukturrechnung berücksichtigt, wo sehr viele Strukturen mit gleicher Topologie und mit Randbedingungen aber unterschiedlichen Zufallszahlen bestimmt werden, um das korrekt gefaltete Protein mit der niedrigsten Gesamtenergie zu erhalten (Cavanagh 2007). Außerdem lässt sich durch den Einsatz mehrerer Optimierungsinstanzen der kaum parallelisierbare SIBASA-Algorithmus auf mehrere CPU-Kerne verteilen, um die Gesamtzeit der automatischen Zuordnung zu verkürzen.

## Verifikation durch Signalvergleich auf mehreren Instanzen

Eine weitere Größe, die die Richtigkeit einer Zuordnung  $k$  messen kann, ist die Pseudoenergie der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  (Gl. 70). Sie basiert auf dem Vergleich des Volumens und der Linienform über den Grundflächen der simulierten Kreuzsignale des Kerns  $k$ . Diese normierte Pseudowahrscheinlichkeit ist 1, falls alle starken simulierten Kreuzsignale des Kerns  $k$  einem experimentellen Signalen zugeordnet worden sind. Wegen der Eindeutigkeit des Korrelationsnetzes der NOESY-Signale, kommt dies fast nur bei der richtigen Zuordnungen vor. Die Pseudoenergie der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  wird kleiner je mehr simulierte Signale schlecht oder keinen experimentellen Kreuzsignal zugeordnet worden sind.

Es bietet sich an für die Verifikation der Zuordnungen die Pseudoenergie der Zuordnung mit der Analyse der unabhängigen Instanzen der automatischen Zuordnung zu kombinieren. Dafür werden für alle unabhängigen Zuordnungsinstanzen  $j$ , die für die Resonanz des Kerns  $k$  die gleiche chemische Verschiebung bestimmt haben, die Pseudoenergien der Zuordnung  $E_{k,j}^{\text{Vol,Shape,sim}}$  summiert. Durch Normierung ergibt sich die Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(m)$  des Kerns  $k$  mit der chemischen Verschiebung  $\delta_{k,m}$  des Clusters  $m$  zu:

$$E_k^{\text{Cluster,Vol,Shape,sim}}(m) = \sum_{j=1}^{N^{\text{Cluster}}(m)} E_{k,j}^{\text{Vol,Shape,sim}} / N^{\text{Instanzen}} \quad (81)$$

Hierbei ist  $j$  der Index einer der  $N^{\text{Cluster}}(m)$  Instanzen die die chemischen Verschiebung  $\delta_{k,m}$  für den Kern  $k$  gefunden haben. Die Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(m)$  hat wieder einen Wertebereich zwischen 0 und 1. Durch Auswertungen bereits gelöster Proteine kann außerdem ein Zusammenhang zwischen der Verifikationspseudoenergie eines Kerns  $k$   $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  nach der automatischen Zuordnung und der Wahrscheinlichkeit der richtigen Zuordnung  $P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  hergestellt werden.

### 2.2.13 Festlegen von Zuordnungen während der Laufzeit

Da die automatische Zuordnung mithilfe des Top-down-Ansatzes ein sehr rechenzeitintensive Aufgabe darstellt, ist die minimale Laufzeit, die man in einen SIBASA Lauf investieren muss, relativ hoch. Es ist deshalb von Vorteil, Synchronisationen zwischen den Instanzen während der Laufzeit durchzuführen. In einer Synchronisationsphase werden bereits gefundene richtige Zuordnungen für den weiteren Verlauf fixiert.

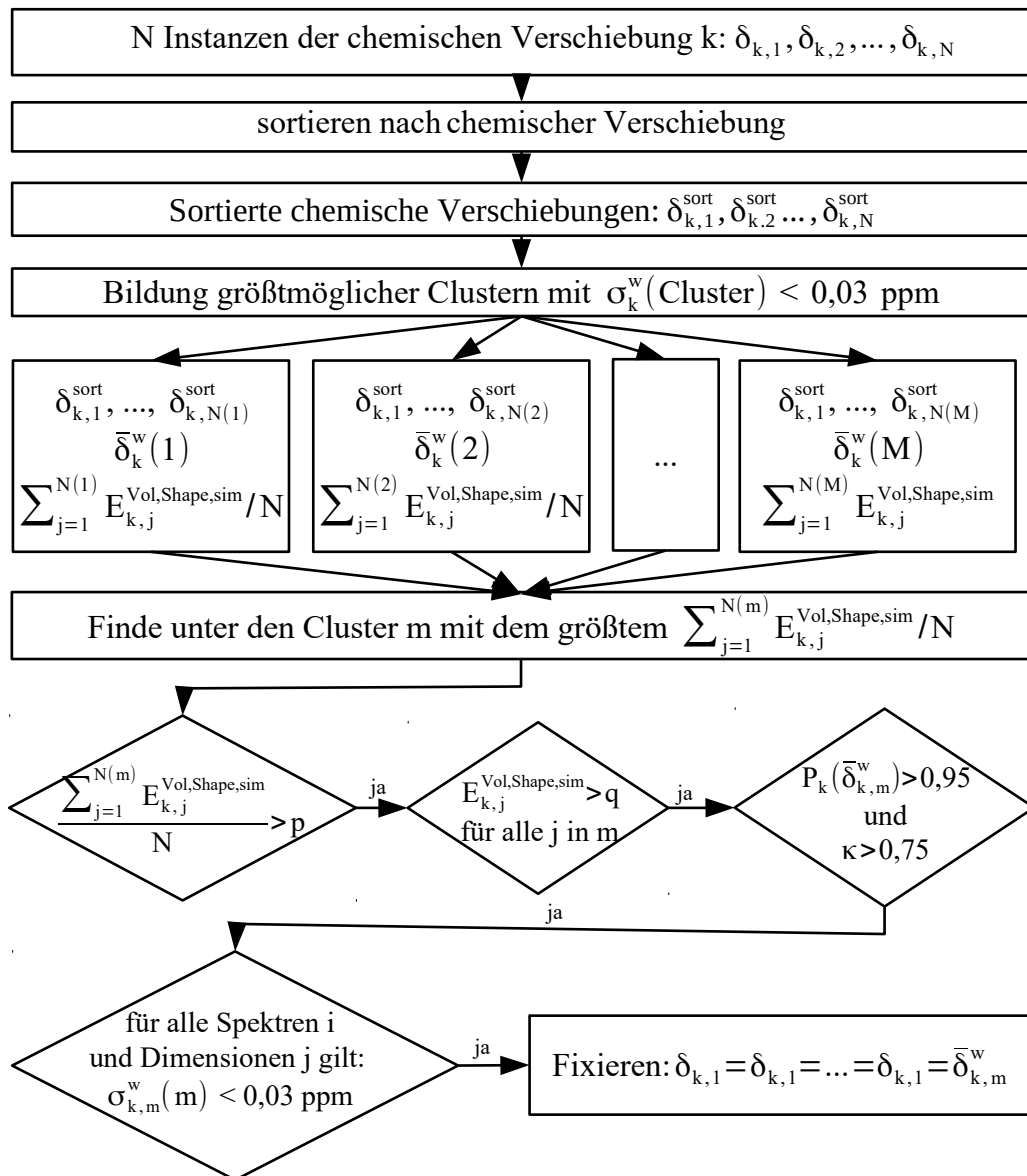


Abbildung 15: Schematische Darstellung der Fixierung der chemischen Verschiebung der Resonanz des Kerns  $k$  während einer Synchronisationsphase. Hierbei ist  $\delta_{k,j}$  die chemische Verschiebung der Resonanz des Kerns  $k$  in der Instanz  $j$ ,  $\delta_{k,j}^{\text{sort}}$  die chemische Verschiebung der nach der Größe der chemischen Verschiebungen sortierten Instanz  $j$ ,  $\bar{\delta}_k^w(m)$  ist der gewichtete Mittelwert und  $\sigma_k^w(m)$  die gewichtete Standardabweichung der chemischen Verschiebung im Cluster  $m$ ,  $E_{k,j}^{\text{Vol,Shape,sim}}$  die Pseudoenergie der Zuordnung in der sortierten Instanz  $j$ ,  $P_k(\delta)$  die Wahrscheinlichkeitsverteilung der chemischen Verschiebung aus Kap. 2.2.7,  $\kappa$  der Kontrollparameter der MOVES aus Kap. 2.2.11,  $p$  der untere Grenzwert der normierten summierten Pseudoenergie der Zuordnung und  $q$  der untere Grenzwert der Pseudoenergie der Zuordnung für eine einzelne Instanz.

Die Schwierigkeit die Resonanzzuordnung eines Kerns im Protein mit automatischen Methoden zu finden, hängt von vielen Faktoren wie z.B. Signalüberlapp in den Spektren, der Anzahl der dipolaren Kopplungspartner oder von der Güte der Voraussagen der chemischen

Verschiebungen (Kap. 2.2.7) ab. Viele Resonanzen werden beim Threshold Accepting Algorithmus (Kap. 2.2.10) schon im früheren Verlauf des Abkühlens richtig zugeordnet und es sollte, um Rechenzeit zu sparen, nicht mehr versucht werden diese zu verändern. Der dynamischen Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}$  (Gl. 64) und der dynamischen Gewichtungsfaktor der Pseudoenergie des Dummyslots  $w^{E,Dummy}$  (Gl. 74) versuchen diesen Effekt zu unterstützen. Mit hohen Startwerten dieser Gewichtungsfaktoren wird versucht, so viele richtige Zuordnungen wie möglich bereits in der Anfangsphase der Optimierung zu bestimmen. Besonders gut können die chemischen Verschiebungen der Resonanzen von Amidprotonen vorzeitig festgelegt werden, da diese sowohl im 2-D-NOESY- als auch im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum sehr viele Kreuzsignale und damit ein sehr eindeutiges Kopplungsmuster haben. Natürlich muss der minimale Grenzwert des verwendeten Qualitätsmerkmal für die Fixierung so gewählt werden, dass die Richtigkeit der während der Laufzeit festgelegten Zuordnungen gewährleistet werden kann.

Es hat sich bewährt 16 unabhängige automatische Zuordnungsläufe mit den selben Parametern und Eingabedaten, die sowohl auf mehreren Prozessoren als auch auf mehreren PCs verteilt werden können, zu starten. Nach einer vorher festgelegten Anzahl von MOVES und SWAP MOEVES, die in Abhängigkeit von der Größe des Proteins auf einen Wert zwischen 250.000 und 500.000 gesetzt werden sollte, wird regelmäßig eine Synchronisation zwischen den 16 Instanzen durchgeführt. Der Threshold  $Th$  des Threshold-Accepting-Algorithmus (Gl. 76) sollte zwischen zwei Synchronisationen fünf mal gesenkt werden.

## Erzeugung einer vereinigten Zuordnungsliste der Instanzen und Bestimmung der Qualität der Zuordnungen

Jede der  $N^{\text{Instanzen}}$  Instanzen  $j$  von SIBASA erzeugt eine eigene vollständige Zuordnungsliste, in der die chemische Verschiebungen  $\delta_{k,j}$  der Resonanzen der Kerne  $k$  mit der Pseudoenergie der Zuordnung  $E_{k,j}^{\text{Vol,Shape,sim}}$  (Gl. 70) bewertet worden sind. Die Erzeugung der vereinigten Liste der chemischen Verschiebungen aller Instanzen und die Berechnung der Qualitätsmerkmale für die Zuordnungen aller Resonanzen läuft folgendermaßen ab:

Zuerst sollen die vollständigen Zuordnungslisten aller Instanzen zu einer einzigen Liste vereinigt werden. Dazu werden die chemischen Verschiebungen  $\delta_{k,j}$  des Kerns  $k$  der Instanzen  $j$  nach Größe sortiert, um möglichst große zusammenhängende Cluster der chemischen Verschiebungen  $\delta_{k,j}$  zu bilden. Die maximale gewichtete Standardabweichung der chemischen Verschiebung im Cluster  $m$  darf dabei maximal  $\epsilon$  betragen (Gl. 82).

$$\sigma_k^w(m) = \sum_{j=1}^{N^{\text{Cluster}}(m)} ((w_{k,j}^{\text{Zuordnung, Instanz}}(m) \delta_{k,j})^2 - \bar{\delta}_k^w(m)) < \epsilon \quad (82)$$



Die  $N^{\text{Cluster}}(m)$  Instanzen, deren chemische Verschiebungen  $\delta_{k,j}$  zusammen die Bedingung aus Gl. 82 erfüllen, bilden den Cluster  $m$ .  $\bar{\delta}_k^w(m)$  ist der zugehörige gewichtete Mittelwert der chemischen Verschiebung der Resonanz des Kerns  $k$  im Cluster  $m$  und  $\varepsilon$  der maximal erlaubte Fehler, der auf 0,03 ppm gesetzt werden sollte.  $w_{k,j}^{\text{Zuordnung, Instanz}}(m)$  ist der Gewichtungsfaktor der chemischen Verschiebung  $\delta_{k,j}$  im gewichteten Mittelwert und der gewichteten Standardabweichung im Cluster  $m$  (Gl. 83).

$$w_{k,j}^{\text{Zuordnung, Instanz}}(m) = \frac{E_{k,j}^{\text{Vol,Shape,sim}}}{\sum_{j=1}^{N(m)} E_{k,j}^{\text{Vol,Shape,sim}}} \quad (83)$$

Er sorgt dafür, dass Zuordnungen mit höherer Pseudoenergie der Zuordnung, also verlässlichere Zuordnungen, höher gewichtet werden. Die gewichtete mittlere chemische Verschiebung  $\bar{\delta}_k^w(\text{best})$  des Clusters mit der höchsten Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  (Gl. 81) wird in die vereinigte Zuordnungsliste eingetragen.

### Bestimmung und Bewertung der Zuordnungen von Amidgruppen

Amidprotonen und deren kovalent gebundene Stickstoffe nehmen in SIBASA auch bei der Verifikation und Bewertung eine besondere Rolle ein. Die chemischen Verschiebungen der Instanzen werden für die Bildung der Cluster anders vorsortiert und anders gewichtet. Die chemischen Verschiebungen der Instanzen  $\delta_{k,j}$  werden für die Erzeugung der Cluster zusammen mit dem kovalent gebundenen Partner nach der kombinierten chemischen Verschiebung (Schumann et al. 2007) sortiert, um zu gewährleisten, dass die nahe beieinanderliegenden Zuordnungen von HSQC-Signale in den Instanzen auch in einem Cluster zusammengefasst werden. Die kombinierte chemische Verschiebung einer Amidgruppe  $k$  in der Instanz  $j$  ist folgendermaßen definiert:

$$\delta_{H_k N_{k,j}} = \frac{\delta_{H_{k,j}}}{\sigma_{AA,H}^{\text{BMRB}}} + \frac{\delta_{N_{k,j}}}{\sigma_{AA,N}^{\text{BMRB}}} \quad (84)$$

Hierbei sind  $\delta_{H_{k,j}}$  und  $\delta_{N_{k,j}}$  die chemischen Verschiebungen des Amidprotons und des Stickstoffs der Amidgruppe  $k$  in der Instanz  $j$ .  $\sigma_{AA,H}^{\text{BMRB}}$  und  $\sigma_{AA,N}^{\text{BMRB}}$  sind die aus der BMRB entnommen Standardabweichungen der chemischen Verschiebungen des Amidprotons und des Stickstoffs der entsprechenden Aminosäure  $AA$ . Bei der Bewertung der Zuordnung einer Amidgruppen ersetzt  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(m)$  des besten Clusters  $m$  die normierte summierte Pseudoenergie einer einzelnen Zuordnung (Gl. 81) als Verifikationspseudoenergie:

$$E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(m) = \sum_{j=1}^{N^{\text{Cluster}}(m)} E_{H^k, j}^{\text{Vol, Shape, sim}} E_{N^k, j}^{\text{Vol, Shape, sim}} / N^{\text{Instanzen}} \quad (85)$$

Hierbei ist  $N^{\text{Cluster}}(m)$  die Anzahl der Instanzen und  $j$  der zugehörige Index im Cluster  $m$ .  $E_{H^k, j}^{\text{Vol, Shape, sim}}$  und  $E_{N^k, j}^{\text{Vol, Shape, sim}}$  sind die Pseudoenergien der Zuordnung des Amidprotons und des Stickstoffs in der Amidgruppe  $k$  in der Instanz  $j$  und  $N^{\text{Instanzen}}$  ist die Anzahl der Instanzen von SIBASA. Durch die Verwendung des Produkt der beiden Pseudowahrscheinlichkeiten der Zuordnung wird die chemische Verschiebung einer Resonanz einer NH-Gruppe nur gut bewertet, falls die gesamte Gruppe und damit das HSQC-Signal sicher richtig zugeordnet worden ist.

Der Gewichtungsfaktor der Pseudoenergie der Zuordnung  $w_{k, j}^{\text{Zuordnung, Instanz}}(m)$  ist für die Amidgruppe  $k$  in der Instanz  $j$  folgendermaßen definiert:

$$w_{k, j}^{\text{Zuordnung, Instanz}}(m) = \frac{E_{H^k, j}^{\text{Vol, Shape, sim}} E_{N^k, j}^{\text{Vol, Shape, sim}}}{\sum_{j=1}^{N(m)} E_{H^k, j}^{\text{Vol, Shape, sim}} E_{N^k, j}^{\text{Vol, Shape, sim}}} \quad (86)$$

Hier ersetzt ebenfalls das Produkt der beiden Pseudoenergien der Zuordnung die einzelne Pseudoenergie der Zuordnung in Gl. 83. Der Gewichtungsfaktor wird für die Bildung der beiden gewichteten Mittelwerte der chemischen Verschiebungen  $\bar{\delta}_{H^k}^w(m)$  und  $\bar{\delta}_{N^k}^w(m)$ , sowie für die Bildung der beiden gewichteten Standardabweichung der chemischen Verschiebungen  $\sigma_{H^k}^w(m)$  und  $\sigma_{N^k}^w(m)$  der Amidgruppe  $k$  im Cluster  $m$  benötigt.

## Filtern von Zuordnungen nach ihrer Qualität

Damit der Zuordnung eines Kerns  $k$  vertraut werden kann, müssen folgende Voraussetzungen erfüllt sein:

- Die Verifikationspseudoenergie des besten Clusters muss höher als der Threshold  $p$  sein:

$$E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}) > p \text{ bzw. } E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}) > p_{\text{Amidgruppe}} \quad (87)$$

$p$  ist ein festgelegter Grenzwert zwischen 0 und 1, der die minimale gewünschte Sicherheit beschreibt.

- Die Pseudoenergie der Zuordnung jeder Instanz  $j$  im besten Clusters muss größer als der Threshold  $q$  sein:

$$\begin{aligned}
 E_{k,j}^{\text{Vol,Shape,sim}} > q \text{ bzw.} \\
 E_{\text{Amidgruppe } k,j}^{\text{Vol,Shape,sim}} > q_{\text{Amidgruppe}}
 \end{aligned}
 \tag{88}$$

$q$  ist ebenfalls ein Grenzwert zwischen 0 und 1.

- Kerne, die insgesamt fünf oder weniger starke Kreuzsignale mit  $d_{kl}^{\text{pseudo}} < 0,4 \text{ nm}$  ( $k, l$  sind die Indices der Kerne die an einem NOESY-Kreuzsignal beteiligt sind haben, sind zu wenig bestimmt um deren Zuordnungen vertrauen zu können.
- Falls die Wahrscheinlichkeit der chemischen Verschiebung  $P_k(\bar{\delta}_k^w)$  (Kap. 2.2.7) der mittleren gewichteten chemischen Verschiebung  $\bar{\delta}_k^w$  größer als 0,95 und der Kontrollparameter  $\kappa$ , der den Anteil der problemspezifischen speziellen MOVES steuert (Kap. 2.2.11), noch kleiner als 0,75 ist, wird der Zuordnung der Resonanz des Kerns  $k$  (noch) nicht während der automatischen Zuordnung vertraut. Der Einfluss der geschätzten Wahrscheinlichkeitsverteilungen  $P_k(\delta)$  auf die gefundene chemische Verschiebung wäre noch zu hoch.
- Die gewichtete Standardabweichung der chemischen Verschiebung des besten Clusters  $\sigma_k^w(\text{best})$  darf in keiner Dimension eines Spektrums den Fehler  $\varepsilon$  übersteigen (Gl. 82). Insbesondere müssen die Offsets der chemischen Verschiebung im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum (Kap. 2.2.11) berücksichtigt werden.
- Der Zuordnung von Kerne, die von SIBASA in den Dummslot verschoben worden sind, wird nicht vertraut, weil diese Zuordnungen nicht verlässlich überprüft werden können.

Eine Zuordnung, die alle diese Bedingungen erfüllt, kann schon während der Laufzeit festgelegt werden. Abb. 15 zeigt schematisch den Prüfvorgang, der zur Fixierung der chemischen Verschiebung einer Resonanz während der Laufzeit führt.

## Nicht stereospezifische Festlegung von Zuordnungen

Eine zweite Iteration, des Prüfvorgangs, legt die chemischen Verschiebungen von magnetisch gleichwertigen Kernen wie z.B.  $\text{CH}_2$ -,  $\text{C}(\text{CH}_3)_2$ - und  $\text{NH}_2$ -Gruppen fest. Bei solchen Gruppen kommt es oft vor, dass die Positionen der Resonanzen der Mitglieder in verschiedenen Instanzen vertauscht wurden. In diesen Fällen würde auch die händische Zuordnung die stereospezifische Lösung nicht immer finden, was aber für die meisten Anwendungen nicht bedenklich ist. Der nicht stereospezifische Prüfvorgang wird als zweite Iteration nach der priorisierten stereospezifische Fixierung durchgeführt.

Die Grundmenge der chemischen Verschiebungen  $\delta_{k,j}^{\text{stereospezifisch}}$  der Resonanzen des Kerns  $k$  für die Bildung der Cluster  $m$  aus den Instanzen  $j$  besteht in der zweiten Iteration aus allen

chemischen Verschiebungen der Resonanzen der Gruppe. Jeder Cluster  $m$ , der aus der sortierten erweiterten Grundmenge gebildet wird, darf aber maximal eine chemische Verschiebung aus jeder Instanz  $j$  enthalten. Der Rest des Prüfvorgangs für die Fixierung wird analog zu der stereospezifischen Festlegung durchgeführt. Bei der nicht stereospezifischen Fixierung von Zuordnungen wird die Resonanz des Gruppenmitglieds, das die meisten chemischen Verschiebung zum besten Cluster beiträgt, festgelegt.

Die chemischen Verschiebungen der beiden Resonanzen einer  $\text{NH}_2$ -Gruppe, werden wieder nach der kombinierten chemischen Verschiebung mit dem Seitenkettenstickstoff sortiert, mit dem Faktor  $w_{k,j}^{\text{Zuordnung, Instanz}}$  aus Gl. 86 gewichtet und mit der Verifikationspseudoenergie von Aminogruppen  $E_{\text{Aminogruppe } k}^{\text{Cluster, Vol, Shape, sim}}(m)$  (Gl. 85) bewertet.

Die in der ersten Iteration Zuordnungen stereospezifisch fixierten Zuordnungen, erhalten den von der BMRB vorgesehen Chemical Shift Ambiguity Code 1. Die in der zweiten Iteration nicht stereospezifische festgelegten Zuordnungen erhalten den Ambiguity Code 2.

### 2.2.14 Dynamische Anpassung der Volumenskalierungsfaktoren

Die statistisch bestimmten initialen Volumenskalierungsfaktoren aus Kapitel 2.2.4 können mit Gl. 30 und den festgelegten Zuordnungen aus Kapitel 2.2.13 noch zur Laufzeit nachjustiert werden. Dies geschieht nach jeder Synchronisation und sorgt dafür, dass die Pseudoenergien danach genauer bestimmt werden können. Um Fehler bei der Bestimmung der Volumenskalierungsfaktoren zu vermeiden, verwendet SIBASA nur eindeutig zugeordnete Kreuzsignale für die Berechnung. Die Position eines simulierten Signals darf ausschließlich von fixierten Zuordnungen festgelegt werden und muss innerhalb der Linienbreite eines einzigen experimentellen Signals liegen. Außerdem werden nur Kreuzsignalzuordnungen, bei denen genau ein simuliertes Signal genau einem experimentellen Signal zugeordnet ist, für die Berechnung der Volumenskalierungsfaktoren verwendet.

Die Zuordnungen der Kreuzsignale werden vor jeder Synchronisation gelöscht und anschließend mit den fixierten Zuordnungen neu erstellt. Da Artefakte, Überlapp, Rauschen das Volumen experimenteller Kreuzsignale und falsche Eingabestrukturen das Volumen der simulierte Kreuzsignale verfälschen können, werden nur starke simulierten NOESY-Signale, also Signale, die einen maximalen Pseudoabstand  $d^{\text{pseudo}}$  (Gl. 67) von 0,4 nm haben, berücksichtigt. Um eine gute Mittlung zu gewährleisten, wird der Skalierungsfaktor eines Spektrum nur angepasst, falls mindestens 40 Signalzuordnungen die obigen Bedingungen erfüllen.

### 2.2.15 Absolute Bewertung von Zuordnungen

Am Ende der automatischen Zuordnung werden die Ergebnisse aller Instanzen mit der in Kapitel 2.2.13 beschriebenen Methoden zu einer Zuordnungsliste zusammengeführt. Im Gegensatz zu der Vereinigung während der Synchronisationsphasen, werden nur Resonanzen, die im Dummyslot liegen, ausgeschlossen. Es werden keine Kerne mit wenigen starken Kreuzsignalen und keine Resonanzen, deren gewichtete Standardabweichung  $\sigma_k^w(\text{best})$  der chemischen Verschiebung unter Einfluss der Offsets (Gl. 82) den maximalen Fehler  $\varepsilon$  in einer oder mehreren Dimension überschreiten, ausgeschlossen. Diese Regeln sind zur Laufzeit sinnvoll, um falsch fixierte Zuordnungen zu verhindern, verlieren aber am Ende ihren Wert, da jede Zuordnung an der optimalen Position liegt.

Die verbleibenden chemischen Verschiebungen können mit den Gl. 87 bzw. Gl. 88 und einem geeigneten minimalen Grenzwert  $p$  der Verifikationspseudoenergie bzw. einem geeigneten Grenzwert  $q$  der Pseudoenergie der Zuordnung gefiltert werden. Die Verifikationspseudoenergie und die Pseudoenergie der Zuordnung sind lediglich relative Merkmale für die Qualität einer Zuordnung und sind deshalb für den Einsatz in der Praxis nicht geeignet. Die Wahrscheinlichkeit  $P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  einer richtigen Zuordnung mit der Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  dagegen ist ein absolutes Qualitätsmerkmal. Mit diesen Wahrscheinlichkeiten kann der Anwender die vereinigte Zuordnungsliste nach einem maximal tolerierten Anteil von falschen Zuordnungen, der auch 0 % sein kann, filtern.

Für die Beurteilung des Endergebnisses hat sich gezeigt, dass das Kriterium aus Gl. 88 und der zugehörige Threshold  $q$  vernachlässigbar sind. Die Sicherheitswahrscheinlichkeit  $P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  kann mit dem Satz von Bayes und den Datensätzen von bereits gelösten Proteinen berechnet werden:

$$P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})) = \frac{P(E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}) | \text{korrekt})}{P(E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}) | \text{korrekt}) + P(E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}) | \text{falsch})} \quad (89)$$

Kandidaten für die benötigten Häufigkeitsverteilungen werden mit SIBASA und dem vollständigen Datensatz von gelösten Proteinen erzeugt und in der AUREMOL Datenbank abgelegt. Die Verteilungen werden folgendermaßen erstellt:

Es werden 16 Instanzen von SIBASA mit dem 2-D-NOESY- und dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum gestartet. Während jeder Synchronisation, die nach jeweils 250.000 MOVES stattfinden, werden keine Zuordnungen fixiert, aber alle Wertepaare  $(\bar{\delta}_k^w(\text{best}); E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der gewichteten mittleren chemischen Verschiebungen und der Verifikationspseudoenergien der besten Cluster der Kerne  $k$  gespeichert. Mithilfe der

bekannten korrekten Liste der chemischen Verschiebungen der Resonanzen des Proteins werden diese Wertepaare in richtige und falsche Zuordnungen aufgeteilt und die Häufigkeitsverteilung  $H^{\text{korrekt}}(E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der richtigen Zuordnungen bzw.  $H^{\text{falsch}}(E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der falschen Zuordnungen in Abhängigkeit von der Verifikationspseudoenergie gebildet. Es kommt dafür wieder ein Kerndichteschätzer mit Gauß Kernel und optimaler Bandbreite zum Einsatz. Da die beiden Verteilungen zu jedem Zeitpunkt im automatischen Zuordnungsalgorithmus gelten müssen, können die Wertepaare aller Synchronisationsphasen gleich behandelt werden.

AUREMOL bietet für HSQC-Spektren auch die Möglichkeit der visuellen Verifikation an. Über ein Softwaretool können die gefunden Zuordnungen mit einer beliebigen Sicherheitswahrscheinlichkeit gefiltert werden und die verbliebenen Zuordnungen der HSQC-Signale direkt auf dem Spektrum betrachtet werden.

## 3 Material und Methoden

### 3.1 Die Modellpeptide Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Gln-Ala-NH<sub>2</sub>

#### 3.1.1 Probe

Die beiden Proben enthielten jeweils 5 mM Gly-Gly-Asn-Ala-NH<sub>2</sub> bzw. Gly-Gly-Gln-Ala-NH<sub>2</sub> welches in 90 % H<sub>2</sub>O/10 % D<sub>2</sub>O gelöst war. Der pH-Wert wurde durch Zugabe von DCl und NaOD auf 3 eingestellt. Als Referenz wurde 0,1 mM DSS hinzugegeben.

#### 3.1.2 NMR Messungen

Die Spektren wurden auf einem Bruker Avance-800 NMR mit einer Cryoprobe bei einer <sup>1</sup>H Resonanzfrequenz von 800,2 MHz aufgezeichnet. Die zweidimensionalen (2-D) NOESY (Nuclear Overhauser Enhancement Spectroscopy) Spektren (Kumar et al. 1980) wurden mit einer Mischzeit von 0,3 s und mit einer Repititionszeit von 3 s bei einer Temperatur von 293 K aufgenommen. Die Größe der Zeitdomäne betrug 512 x 16.384 komplexe Datenpunkte, die spektrale Breite in beiden Dimensionen war jeweils 10,0135 ppm. Die FIDs wurden mit LB = 5 Hz exponentiell gefiltert und mit TopSpin (Bruker BioSpin, Karlsruhe) Fourier transformiert. Mit den entsprechenden AUREMOL (Wolfram Gronwald und Hans Robert Kalbitzer 2004) Routinen wurde die Grundlinie automatisch korrigiert und das Wassersignal unterdrückt (Malloni et al. 2010). Die Größe der Signalvolumen wurde mit der AUREMOL-Integrationsroutine (Neidig et al. 1995) bestimmt. Die 1D-Multiplettaufspaltungen von Asn H<sup>β2,3</sup> in Gly-Gly-Asn-Ala-NH<sub>2</sub> wurden mit NMRSIM (Bruker BioSpin, Karlsruhe) simuliert.

#### 3.1.3 Molekulardynamiksimulationen

Die Molekulardynamik-Trajektorien von Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Gln-Ala-NH<sub>2</sub> wurden mit dem Softwarepaket GROMACS in der Version 4.5.1 (Hess et al. 2008; Berendsen et al. 1995) erzeugt. Es wurde das Amber94 Kraftfeld (Cornell et al. 1995), das SPC/E (Berendsen, H. J. C. et al. 1987) Wassermodell für den Wasseranteil (90 %) und das SPCHW Modell (Grigera 2001) für den Schweren Wasseranteil (10 %) verwendet. Der Einfluss von HDO im Lösungsmittel wurde vernachlässigt. Die Startstrukturen für die Molekulardynamik-

Rechnungen der beiden Tetrapeptide wurden mit Pymol modelliert und zusammen mit den Lösungsmittelmolekülen in kubische  $7,6 \times 6,7 \times 6,8 \text{ nm}^3$  Boxen gepackt. Zum Gly-Gly-Asn-Ala-NH<sub>2</sub> wurden 10.151 H<sub>2</sub>O und 1.128 D<sub>2</sub>O Moleküle hinzugefügt, beim Gly-Gly-Gln-Ala-NH<sub>2</sub> waren es 9.452 H<sub>2</sub>O und 1.050 D<sub>2</sub>O Moleküle. Die Temperatur wurde auf 293 K und der Druck auf 0,1 MPa eingestellt.

Wie von Garcia et al. (Day und García 2008) vorgeschlagen, kontrollierte ein Nosé-Hoover Thermostat die Temperatur und eine Parinello-Rahman Barostat den Druck. Die Kopplungszeiten von Temperatur und Druck wurden auf 5 ps gesetzt. Die Tetrapeptide, die H<sub>2</sub>O Moleküle und die D<sub>2</sub>O Moleküle erhielten jeweils ihre eigenen Wärmebäder und Druckklassen, in denen die Temperatur und der Druck unabhängig von den andern Klassen geregelt wurde. Die Particle-Mesh-Ewald Methode (PME) (Darden et al. 1993) berechnete während der Simulation die elektrostatischen Wechselwirkungen mit einem Gitterabstand von 0,12 nm. Van-der-Wals-Energien wurden bis zu einem Abstand von 1 nm berücksichtigt. Der Zeitschritt der Simulation wurde auf 1 fs festgelegt und die Non-Bonded-Liste alle 10 Schritte aktualisiert. Insgesamt wurden zwei 30 ns lange Trajektorien berechnet. Die ersten 5 ns dienten der Überführung des Systems in das Gleichgewicht, während die letzten 25 ns für die Simulation der NMR-Spektren verwendet wurden. Jeweils 1.000 PDB Strukturen wurden im zeitlichen Abstand von 25 ps aus den zweiten Teilen der Trajektorien entnommen, mit dem Validation Server der RCSB Protein Data Bank (Berman 2000) auf stereospezifisch korrekte Atomnomenklatur überprüft. Sie dienten als Ensemble für die Rückrechnung der 1-D- und 2-D-NOESY-Spektren. Die Rückrechnungen wurden mit AUREMOL RELAX (Ried et al. 2004) durchgeführt. AUREMOL RELAX verwertet neben den Atomabständen aus den Eingangstrukturen auch Dynamikparameter, die aus den letzten 25 ns der Simulation berechnet werden konnten. Abstände und Dynamikparameter wurden mit einer weiterentwickelten Version des Lipari-Szabo-Modells zu Relaxationsraten verrechnet. Außerdem wurde AUREMOL RELAX erweitert um chemische Austauschprozesse, die nicht von Molekulardynamik-Rechnungen abgedeckt werden, berücksichtigen zu können.

Der Selbstdiffusionskoeffizient des Lösungsmittels wurde aus einer weiteren 10 ns lange Trajektorie mit gleichen Parametern, aber ohne Peptid berechnet.

### 3.1.4 Rückrechnungen

Für die Rückrechnung von NMR-Spektren benötigte AUREMOL RELAX die Peptid-Sequenz, die chemischen Verschiebungen, mindestens eine PDB-Struktur und für jeden NOE-Kontakt einen Ordnungsparameter, eine rotatorische und eine interne Korrelationszeit. Das AURMEOL-Metafile enthielt die Dynamikparameter und die chemischen Verschiebungen. Es wurde konnte mit einer grafischen Oberfläche aus der Zuordnungsliste, die mithilfe eines TOCSY-Spektrums (**T**otal **C**orrelation **S**pectroscopy) (Braunschweiler und Ernst 1983) erzeugt wurde und den Ausgabedateien des GROMACS Tools *g\_rotacf* erstellt werden. Mit diesen Daten wurden 2-D-NOESY-Spektren mit 0,3 s Mischzeit und 3 s Repititionszeit und 1-



D-NOESY-Spektren mit 0,1 s Mischzeit und 2 s Repititionszeit simuliert.

## 3.2 Die Punktmutante HPr S. aureus (H15A)

Für die Erzeugung der Häufigkeitsverteilungen der Pseudoenergien des Volumens und der Linienform (Gl. 52 und Gl. 53) und die Erzeugung der Häufigkeitsverteilungen der Verifikationspseudoenergien (Kap. 2.2.15) der richtigen und falschen Zuordnungen wurden die NOESY-Spektren, die 3D-Strukturen und die vollständige Zuordnung der Punktmutante H15A des Histidin-Containing Phosphorcarrier (HPr) von *Staphylococcus aureus* verwendet. Mit den bestimmten Häufigkeitsverteilungen konnten die Bayesschen Wahrscheinlichkeiten der lokalen richtigen Zuordnung  $P_{ij}^{\text{Vol,Shape}}$  der experimentellen Signale  $ij$  und die Wahrscheinlichkeit der richtigen Zuordnung  $P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}})$  der Resonanzen der Kerne  $k$  bestimmt werden.

Für HPr S. aureus (H15A) standen sowohl eine vollständige Zuordnung, ein NMR Strukturensamble, ein 2-D-NOESY- und ein 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum zur Verfügung (Munte et al., Veröffentlichung in Vorbereitung). Damit eignet es sich dieser Datensatz gut, um die für SIBASA benötigten Häufigkeitsverteilungen, die anschließend in der AURMEOL Datenbank abgelegt werden, zu erstellen.

Für den ersten Testfall von SIBASA wurde die vollständige Zuordnung des HPr S. aureus (H15A) aus den beiden Spektren und der Röntgenstruktur von HPr S. faecalis (1PTF) (Jia et al. 1994) bestimmt.

HPr ist ein wichtiger Teil des Phosphotransferasesystems (PTS) von Bakterien. Es transferiert Phosphorylgruppen, die an das  $\text{N}^{\delta 1}$ -Atom des aktiven Zentrums von His15 gebunden sind, von Enzym I zu Enzym II. Die Gruppe wird von HPr an ein Cystein im Enzym II übertragen, das wiederum dann ein Kohlenhydrat durch die Zellmembran transportiert (Kalbitzer und Hengstenberg 1993; Postma et al. 1993). HPr S. aureus (H15A) hat 88 Aminosäuren und ein Molekulargewicht von 9,43 kDA.

### 3.2.1 Probe

Die Probe des 2-D-NOESY-Spektrums enthielt 3 mM der unmarkierten Punktmutante des Proteins HPr S. aureus (H15A), welche in 90 %  $\text{H}_2\text{O}$  und 10 %  $\text{D}_2\text{O}$  gelöst war. Der pH-Wert wurde durch Zugabe von NaOD, DCl und 0,5 mM Ethylendiamintetraessigsäure (EDTA) auf 7,0 eingestellt. Es wurden außerdem noch 1 mM Natriumazid ( $\text{NaN}_3$ ) und als Referenz 0,06 mM DSS hinzugefügt.

Die Probe des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums enthielt 1 mM des  $^{15}\text{N}$ - und  $^{13}\text{C}$ -

markierten Proteins HPr *S. aureus* (H15A). Das Protein war in 92 % H<sub>2</sub>O, 8 % D<sub>2</sub>O und 0,1 mM Dioxan gelöst. Als Puffer wurde 10 mM [D-11]TRIS/HCl Puffer verwendet. Der pH wurde auf 7 eingestellt. Zusätzlich wurde 0,5 mM NaN<sub>3</sub>, 0,5 mM EDTA und 0,05 mM DSS als Referenz hinzugegeben.

### 3.2.2 NMR Messungen

Die 2-D-NOESY-Spektren (Kumar et al. 1980) wurden mit einer Mischzeit von 0,1 s und mit einer Repetitionszeit von 1,54 s bei einer Temperatur von 303 K auf einem Bruker Avance-800 NMR mit einer <sup>1</sup>H-Resonanzfrequenz von 800,2 MHz aufgezeichnet. Die Größe der Zeitdomäne betrug 1.024 x 4.096 komplexe Datenpunkte, die spektrale Breite in beiden Dimensionen war 13,9486 ppm. Es wurden jeweils 36 Scans aufgenommen. Mit TopSpin (Bruker BioSpin, Karlsruhe) wurden die FIDs mit LB = -6 Hz und GB = 0,1 in der indirekten und mit LB = -4 Hz und GB = 0,12 in der direkten Dimension mit der Lorentz-Gauß-Transformation gefiltert. Das Spektrum wurde anschließend auf 1.024 x 4.096 reelle Datenpunkte Fourier transformiert. Die Signale wurden mit der entsprechenden AUREMOL (Wolfram Gronwald und Hans Robert Kalbitzer 2004) Routine gefunden. Die Signalliste wurde nachträglich manuell korrigiert, wobei die Diagonalsignale entfernt wurden. Insgesamt wurden im 2-D-NOESY-Spektrum 3.601 Signale gefunden. Die Größe der Signalvolumen wurde mit der verbesserten AUREMOL-Integrationsroutine (Donaubauer, Veröffentlichung in Vorbereitung) auf dem Segmentationsniveau von 0,1 bestimmt.

Das 3D <sup>15</sup>N-NOESY-HSQC-Spektrum wurde mit einem 600 MHz Bruker Spektrometer, einer Mischzeit von 0,12 s und einer Repetitionszeit von 1,12 s bei einer Temperatur von 303 K und einer <sup>1</sup>H-Resonanzfrequenz von 600,2 MHz aufgenommen, wobei <sup>13</sup>C entkoppelt war. Es wurden 128 x 64 x 2.048 komplexe Datenpunkte mit jeweils 24 Scans aufgenommen. Die spektrale Breite war 13,9813 x 32 x 13,9813 ppm. Der FID wurde in der indirekten <sup>1</sup>H-Dimension mit dem Cosinus Filter (QSINE mit SSB=2), die Stickstoff-Dimension und die direkte Protonen-Dimension wurden beide mit Lorentz-Gauß-Transformation gefiltert (LB=-8Hz und GB=0,1 bzw. LB=-6Hz und GB=0,12). Auf die Stickstoff-Dimension wurde eine Lineare Vorwärts Vorhersage (Lpfc) mit 32 Koeffizienten angewendet. Nach der TopSpin Fourier Transformation auf 256 x 128 x 2.048 reelle Datenpunkte, wurde die direkte Dimension mit einer Streifen-Transformation auf die endgültige Größe von 256 x 128 x 896 Datenpunkten gebracht. Die Signalliste wurde zuerst automatisch mit AUREMOL erzeugt und anschließend manuell verbessert. Im 3D <sup>15</sup>N-NOESY/HSQC-Spektrum wurden insgesamt 1.080 Signale ausgewählt. Die Signalvolumen wurden wieder mit der AUREMOL Integrationsroutine auf dem Segmentationsniveau von 0,1 bestimmt.

### 3.2.3 Modellierung und Molekulardynamiksimulationen

Im ersten Test von SIBASA wurde die vollständige Zuordnung des Protein HPr *S. aureus* (H15A) bestimmt. Ausgangsstruktur war die 0,16 nm aufgelöste Röntgenstruktur von HPr *S. faecalis* (PDB ID: 1PTF) (Jia et al. 1994). Die von SIBASA benötigte Struktur von HPr *S. aureus* (H15A) wurde mit AUREMOL PERMOL und CNS modelliert (Brunner 2006). Die Sequenzidentität zwischen den beiden Strukturen betrug dabei 63,6 %. Für PERMOL kamen die Parameter aus (Cano et al.; 2009) zum Einsatz. Es wurden alle Abstände zwischen H<sup>α</sup>-Atomen, die mindesten 2 Aminosäuren in der Primärsequenz auseinanderliegen und einen Abstand zwischen 0,18 und 1,5 nm haben und alle anderen interresidualen <sup>1</sup>H-<sup>1</sup>H-Abstände ohne H<sup>α</sup>-Beteiligung im Bereich von 0,18 bis 0,6 nm als Restraints verwendet. Zusätzlich wurden Restraints für alle Diederwinkel und alle Wasserstoffbrücken erzeugt. Insgesamt hat PERMOL 5.874 Abstands-, 353 Diederwinkel- und 62 Wasserstoffbrücken-Restraints definiert. Mit diesen Restraints wurden 1.000 Strukturen mit CNS 1.2 (Brunger 2007; Brünger et al. 1998) berechnet. Die Struktur mit der besten Energie diente als Ausgangspunkt einer freien Molekulardynamiksimulation mit GROMACS 4.6.5 (Hess et al. 2008; Berendsen et al. 1995). Die MD-Simulation und wurden mithilfe der Anleitung von Kerrigan durchgeführt. Es wurde das Amber ff99SB-ILDN Kraftfeld (Lindorff-Larsen et al. 2010), das SPC/E-Wassermodell (Berendsen, H. J. C. et al. 1987) für den Wasseranteil (90 %) und das SPCHW-Modell (Grigera 2001) für den schweren Wasseranteil (10 %) verwendet. Der Einfluss von HDO im Lösungsmittel wurde vernachlässigt. Die Temperatur wurde auf 303 K und der Druck auf 0,1 MPa eingestellt. Das Protein wurde in eine automatisch generierte Dodekaeder Box, die einen minimalen Abstand von 1,2 nm zum zentrierten Protein hat, gesteckt. Es wurden 5.044 H<sub>2</sub>O- und 560 D<sub>2</sub>O-Moleküle hinzugegeben. Zusätzlich wurde die Gesamtladung des Systems mit 0,15 mM NaCl, d.h. mit 23 Na<sup>+</sup>- und 17 Cl<sup>-</sup>-Ionen ausgeglichen. Ein Velocity-Rescale Thermostat (Bussi et al. 2007) kontrollierte die Temperatur und ein Parinello-Rahman Barostat den Druck. Die Kopplungszeiten der Temperatur wurde auf 0,1 ps, die des Drucks auf 2 ps gesetzt. Das Protein und das Lösungsmittel erhielten jeweils ihre eigenen Wärmebäder und Druckklassen, in denen Temperatur und der Druck unabhängig voneinander geregelt werden konnten. Die elektrostatischen Wechselwirkungen während der Simulation wurden mit der Particle-Mesh-Ewald Methode (PME) (Darden et al. 1993) mit einem Gitterabstand von 0,16 nm berechnet. Van-der-Waals- und Coulomb-Energien wurden bis zu einem Abstand von 1 nm berücksichtigt. Der Zeitschritt der Simulation wurde auf 2 fs festgelegt und die Non-Bonded-Liste alle 5 Schritte aktualisiert. Die besten homologe Struktur von HPr *S. aureus* (H15A) wurde zuerst im Vakuum anschließend in Lösung auf die Maximalkraft von 500 kJmol<sup>-1</sup>nm<sup>-1</sup> minimiert. Anschließend folgten zwei MD-Simulationen von jeweils 100 ps um das Gesamtsystem ins Gleichgewicht zu bringen. Zuerst wurde das minimierte Gesamtsystem des Proteins in Lösung als kanonisches (NVT-) Ensemble und anschließend unter Berücksichtigung des Drucks (NPT-Ensemble) simuliert. Aus dem Gleichgewichtssystem wurde anschließend eine 100 ns lange Trajektorie erzeugt. Für die Rückrechnung der NOESY-

Spektren wurden 200 Strukturen von HPr *S. aureus* (H15A) im Abstand von jeweils 0,5 ns entnommen. 1.000 äquidistante Strukturen im Abstand von 0,1 ns wurden für die Erstellung der speziellen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen verwendet (Kap. 2.2.7). Die rotatorischen Korrelationszeiten  $\tau_{ij,rot}$ , die Ordnungsparameter  $S_{ij}$ , die Ordnungsparameter der schnellen Beweglichkeit  $S_{ij}$  und die Korrelationszeiten der langsamen internen Beweglichkeit  $\tau_{ij,slow}$  wurden für alle NOE-Kontakte, die im Durchschnitt weniger als 0,5 nm voneinander entfernt lagen, mit der Methode aus Kap. 2.1.5 bestimmt.

### 3.2.4 Simulation der NOESY-Spektren

Die Rückrechnungen der 2-D-NOESY-Spektren und die der 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren von HPr *S. aureus* (H15A) wurden mit AUREMOL RELAX in der neuesten Version erstellt. Mischzeiten, Repetitionszeiten, Spektrometerfrequenz, Spektrale Breiten und Offsets, Anzahl der Datenpunkte, Filter und deren Parameter wurden aus den experimentellen Spektren direkt übernommen (siehe Kap. 3.2.2). Der maximale Abstand für NOE-Kontakte wurde für beide Rückrechnungen auf 0,5 nm begrenzt. Das Segmentationsniveau der simulierten Linienformen wurde auf 10 % des Signalmaximums gesetzt. Die Signalmaxima in den simulierten Spektren wurden automatisch mit AUREMOL gefunden und integriert, wobei die Diagonalsignale der 2-D-NOESY-Spektren vorher entfernt wurden.

Für die Erzeugung der Häufigkeitsverteilungen der Pseudoenergien, die in der AURMOL Datenbank gespeichert wurden die simulierten NOESY-Spektren aus den besten 20 NMR-Strukturen mit der vollständigen korrekten Zuordnungsliste zurückgerechnet. Für die Rückrechnungen wurden außerdem die AUREMOL RELAX Standardparameter für die Dynamik verwendet, d.h. es kamen die Bewegungsmodelle FAST\_JUMP, RING und LIPARI\_1 aus (Görler und Kalbitzer; 1997) zum Einsatz, wobei der Ordnungsparameter für LIPARI\_1 Hauptketten-Hauptketten Kontakte auf 0,85, für Hauptketten-Seitenketten auf 0,8 und für Seitenketten-Seitenketten Kontakte auf 0,6 gesetzt wurden (Axel T. Brünger 1992). Die rotatorische Korrelationszeit, die mit der Stokes-Einstein-Beziehung berechnet wurde, war 3,54 ns. Es wurden 10.737 2-D-NOESY-Signale und 2.132 3D  $^{15}\text{N}$ -NOESY-HSQC Signale simuliert. Die zusätzlichen Linienverbreiterungen und die Volumenskalierungsfaktoren wurden von SIBASA statistisch bestimmt (Kap. 2.2.5 und 2.2.4).

Für die Rückrechnungen des Testfalls wurde das 200 Strukturen umfassende MD-Ensemble und eine zufällige mit den Wahrscheinlichkeitsdichten der chemischen Verschiebungen aus BMRB-Daten generierte Zuordnungsliste verwendet. In den Rückrechnungen wurden folgende Kerne, die selten sichtbare NOESY-Kreuzsignale erzeugen, ignoriert:

- die an Sauerstoffe gebundenen Protonen Asp  $\text{H}^{\delta 2}$ , Glu  $\text{H}^{\epsilon 2}$ , Ser  $\text{H}^{\gamma}$ , Thr  $\text{H}^{\gamma 1}$ , Tyr  $\text{H}^{\eta}$
- das an Schwefel gebundene Proton Cys  $\text{H}^{\gamma}$
- die beiden selten sichtbaren Protonen  $\text{H}^{\delta 1}$  und  $\text{H}^{\epsilon 2}$  des Histidins

- die Protonen der beiden NH<sub>2</sub>-Gruppen des Arginins
- die Protonen der NH<sub>3</sub>-Gruppe des Lysins

Die Dynamik ging über das erweiterte Lipari-Szabo-Modell (siehe Kap. 2.1.2 und Clore et al. 1990) und die Korrelationszeiten und Ordnungsparameter, die aus der MD-Trajektorien berechnet wurden, in die Rückrechnungen ein. Es wurden 11.010 2-D-NOESY- und 2.139 3D <sup>15</sup>N-NOESY-HSQC-Linienformen simuliert. Linienverbreiterungen und Volumenskalierungsfaktoren wurden wieder statistisch bestimmt.

### 3.2.5 Automatische Zuordnung mit SIBASA

Die automatische Zuordnung mittels Threshold-Accepting wurde mit einem Anfangs-Threshold von 40 gestartet. Nach jeweils 50.000 Iterationen wurde der Threshold logarithmisch mit einem Faktor von 0,985 abgekühlt (Kap. 2.2.10).

Durch Zusammenfassung von experimentellen Signalen (Kap. 2.2.8) wurden die 3.601 2-D-NOESY-Signale auf 3.290, die 1.080 3D <sup>15</sup>N-NOESY-HSQC-Signale auf 1.075 Signale reduziert.

In die Slotliste gingen alle digitalen Datenpunkte, des 2-D-NOESY-Spektrums, die mindestens ein experimentelles Signal auf mindesten halber Höhe geschnitten haben, ein. Vom 3D <sup>15</sup>N-NOESY-HSQC-Spektrum gingen dagegen nur die Punkte, die ein Signalmaximum geschnitten haben, in die Slotlisten ein (Kap. 2.2.6). Insgesamt entstanden so eine <sup>1</sup>H-Slotliste mit 2.126 Slots und eine <sup>15</sup>N-Slotliste mit 75 Slots.

Es wurden die kombinierten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der 586 zuzuordnenden Kerne verwendet. Die mit SHIFTX2, dem 1.000 Strukturen große MD-Ensemble und einem Kerndichteschätzer vorhergesagten Verteilungen gingen mit 75 % und die aus der Nh3D-Datenbank und SHIFTS berechneten allgemeinen Dichteverteilungen gingen dabei mit 25 % in die kombinierte Wahrscheinlichkeitsdichten ein (Kap. 2.2.7).

Die Benutzerdefinierte Stärke  $A$  der Pseudoenergie der chemischen Verschiebung  $E^\circ$  wurde auf 1,5 (Gl. 63) gesetzt, der Anfangswert des Kontrollparameters  $\kappa$  für spezielle MOVES und allgemeine MOVES wurde auf 0,9 eingestellt. Der Kontrollparameter  $\kappa$  wurde bei jeder Abkühlung mit dem Faktor 0,995 multipliziert. Der chemische Verschiebungsbereich für allgemeine MOVES (Kap. 2.2.11) betrug für jeden Kern  $k$   $\bar{\delta}_k \pm 5 \sigma_{AA, Kern}^{BMRB}$ .

Insgesamt wurden 16 Instanzen mit unterschiedlichen zufälligen Zuordnungen und Random Seeds gestartet. Nach jeweils fünfmaligen Abkühlen (250.000 Iterationen) wurden alle Instanzen synchronisiert. Im Testfall wurden Zuordnungen fixiert (Kap. 2.2.13) und die Volumenskalierungsfaktoren angepasst (Kap. 2.2.14). Die gewichtete Standardabweichung  $\sigma_k^w(\text{best})$  der chemischen Verschiebungen des besten Clusters eines Kerns  $k$  durfte dabei

maximal 0,03 ppm sein. Die minimalen Grenzwerte  $p$  (Gl. 87) bzw.  $q$  (Gl. 88) der Verifikationspseudoenergien  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  bzw. der Pseudoenergien der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  einer Instanz im besten Cluster wurden für die Fixierung auf 0,85 gesetzt. Die minimalen Grenzwerte  $p_{\text{Amidgruppe}}$  bzw.  $q_{\text{Amidgruppe}}$  der Verifikationspseudoenergien von Amidgruppen  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  bzw. des Produkts der Pseudoenergien der Zuordnung  $E_{H^k}^{\text{Vol,Shape,sim}} E_{N^k}^{\text{Vol,Shape,sim}}$  einer Instanz im besten Cluster wurden für die Fixierung auf 0,73 gesetzt.

Für die Erzeugung der Häufigkeitsverteilungen der Verifikationspseudoenergien der richtigen  $H(E_k^{\text{Cluster,Vol,Shape,sim}}|\text{korrekt})$  und der falschen Zuordnungen  $H(E_k^{\text{Cluster,Vol,Shape,sim}}|\text{falsch})$  wurden 20 Synchronisationen (1.000.000 Iterationen) mit den Rückrechnungen aus den NMR-Daten durchgeführt.

Nach der 51. Synchronisation (12.750.000 Iterationen) war die automatische Zuordnung von HPr *S. aureus* aus der Röntgenstruktur von HPr *S. faecalis* abgeschlossen, weil die minimale Pseudoenergie der Optimierung erreicht worden ist und keine weiteren chemischen Verschiebungen zur Laufzeit fixiert wurden.

### 3.3 Plasmodium falciparum Thioredoxin

**Plasmodium falciparum Thioredoxin (PfTrx)** ist Teil des effektiven regulierenden Redoxsystems im Malaria Parasiten *Plasmodium falciparum*. Als enzymatische Oxidoreduktase wechselt es ständig zwischen dem hier untersuchten reduzierten Zustand und dem oxidierten Zustand. Mithilfe von PfTrx kontrolliert der Malaria Parasit den Redox-Zustand innerhalb seiner Wirtszelle, was für sein Überleben und seine Reproduktion notwendig ist (Munte et al. 2009). Der wichtigste Unterschied zwischen der reduzierten und der oxidierten Form ist die Disulfidbrücke zwischen den beiden Cysteinen Cys30 und Cys 33 im aktiven Zentrum, die sich in der oxidierten Form einstellt. PfTrx hat 104 Aminosäuren und inklusive des hier nicht betrachteten His-Tags ein Molekulargewicht von 13 kDa.

Für das reduzierte PfTrx standen die NMR-Strukturen (Munte et. al., Veröffentlichung in Vorbereitung) (PDB ID: 2MMN), die vollständige Zuordnung, das 2-D-NOESY- und das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum zur Verfügung (Munte et al. 2009).

#### 3.3.1 Probe

Die Probe des 2-D-NOESY-Spektrums enthielt 1 mM des unmarkierten reduzierten PfTrx, welches in 92 %  $\text{H}_2\text{O}$  und 8 %  $\text{D}_2\text{O}$  gelöst war. Der pH-Wert wurde durch Zugabe von 10 mM

Kaliumphosphat Puffer auf 7,0 eingestellt. Es wurden außerdem noch 1 mM Natriumazid ( $\text{NaN}_3$ ) und als Referenz 0,1 mM DSS hinzugefügt (Munte et al. 2009).

Die Probe des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums enthielt 0,1 mM der  $^{15}\text{N}$  markierten Proteins PfTrx und wurde ansonsten mit dem gleichen Puffer auf pH 7,0 eingestellt. Die Lösung enthielt ebenfalls 92 %  $\text{H}_2\text{O}$  und 8 %  $\text{D}_2\text{O}$ .

### 3.3.2 NMR Messungen

Das 2-D-NOESY-Spektrum wurden mit einer Mischzeit von 0,1 s und mit 1,78 s Repetitionszeit bei einer Temperatur von 293 K auf einem Bruker Avance-800 NMR bei einer  $^1\text{H}$ -Resonanzfrequenz von 800,2 MHz aufgezeichnet. Die Größe der Zeitdomäne betrug  $1.024 \times 4.096$  komplexe Datenpunkte, die spektrale Breite in beiden Dimensionen war 13,9486 ppm. Die Anzahl der Scans war 32. Die FIDs wurden mit  $\text{LB} = -8$  Hz und  $\text{GB} = 0,1$  in der indirekten und mit  $\text{LB} = -6$  Hz und  $\text{GB} = 0,12$  in der direkten  $^1\text{H}$ -Dimension mit einer Lorentz-Gauß-Transformation gefiltert. Das Spektrum wurde anschließend mit TopSpin (Bruker BioSpin, Karlsruhe) auf die Größe von  $1.024 \times 4.096$  reelle Datenpunkten Fourier transformiert. Die Signalliste wurden mit der automatischen AUREMOL (Wolfram Gronwald und Hans Robert Kalbitzer 2004) Routine erzeugt und anschließend manuell korrigiert, wobei auch die Diagonalsignale entfernt wurden. Insgesamt wurden im 2-D-NOESY-Spektrum 5.372 Signale ausgewählt. Die Größe der Signalvolumen wurde mit der verbesserten AUREMOL-Integrationsroutine (Donaubauer, Veröffentlichung in Vorbereitung) auf dem Segmentationsniveau von 0,1 bestimmt.

Das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum wurde mit einem 800 MHz Bruker Spektrometer, einer Mischzeit von 0,08 s und einer Repetitionszeit von 1,1 s bei einer Temperatur von 293 K und einer  $^1\text{H}$ -Resonanzfrequenz von 800,2 MHz aufgenommen. Die Zeitdomäne hatte  $128 \times 64 \times 2.048$  komplexe Datenpunkte, die spektralen Breiten war  $13,9486 \times 35 \times 13,9813$  ppm. Es wurden immer 32 Scans aufgenommen. Die FIDs wurde in der indirekten  $^1\text{H}$ -Dimension mit dem Cosinus Filter (QSINE mit  $\text{SSB} = 2$ ), die Stickstoff-Dimension bzw. die direkte Protonen-Dimension wurde mit  $\text{LB} = -8$  Hz und  $\text{GM} = 0,1$  bzw.  $\text{LB} = -6$  Hz und  $\text{GB} = 0,12$  mit der Lorentz-Gauß-Transformation gefiltert. Auf die Stickstoff-Dimension wurde eine Lineare Vorwärts Vorhersage (Lpfc) mit 64 Koeffizienten angewendet. Nach der Fourier Transformation mit TopSpin auf  $256 \times 128 \times 2.048$  reelle Datenpunkte, wurde die direkte  $^1\text{H}$ -Dimension mit einer Streifen-Transformation auf die endgültige Größe von  $256 \times 128 \times 736$  Datenpunkten gebracht. Die Signalliste wurde automatisch mit AUREMOL erzeugt und die Liste anschließend manuell verbessert. Es wurden insgesamt 950 3D  $^{15}\text{N}$ -NOESY-HSQC Signale ausgewählt. Anschließend wurden die Signalvolumen wieder mit der AUREMOL Integrationsroutine auf dem Segmentationsniveau von 0,1 bestimmt.

### 3.3.3 Modellierung und Molekulardynamiksimulationen

Als Basis der MD-Rechnungen diente die 0,237 nm aufgelöste Röntgenstruktur des reduzierten Plasmodium Thioredoxin aus dem Komplex mit Thioredoxin Reductase (Fritz-Wolf et al. 2013) (PDB ID: 4J56). Die extrahierte Struktur hatte eine Sequenzidentität von 98,1 %. In der anschließenden Homologie-Modellierung mit AUREMOL PERMOL und den Parametern aus Kap. 3.2.3 (Cano et al. 2009), wurde die Röntgenstruktur an die gegebene Primärsequenz von PfTrx angepasst. Des bedeutete im speziellen den Austausch von Ala9 mit Ser9 und den Austausch von Ser33 mit Cys33 im aktiven Zentrum. PERMOL erzeugte 8.509 Abstands-, 453 Diederwinkel- und 65 Wasserstoffbrückenbindungsrestraints. Mit CNS 1.2 (Brunger 2007; Brünger et al. 1998) wurden 1.000 Strukturen erzeugt. Die Struktur mit der geringsten Energie wurde als Basis der freien MD mit GROMACS 4.6.5 (Hess et al. 2008; Berendsen et al. 1995) verwendet. Die MD-Simulationen wurden mit den in Kap. 3.2.3 beschriebenen Parametern gestartet. Es musste nur der D<sub>2</sub>O-Gehalt auf 8 % und die Temperatur von 293 K geändert werden. Insgesamt bestand das Lösungsmittel aus 4.389 H<sub>2</sub>O-Molekülen, 488 D<sub>2</sub>O-Molekülen, 20 Na<sup>+</sup>- und 15 Cl<sup>-</sup>-Atome.

Aus 1.000 äquidistante Strukturen wurden die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen mit SHIFTX2 erzeugt. 200 äquidistante Strukturen wurden für die Rückrechnungen der NOESY-Spektren mit AUEREMOL RELAX verwendet. Die Korrelationszeiten und Ordnungsparameter für das erweiterte Lipari-Szabo-Modell stammten wieder aus der gesamten Trajektorie.

### 3.3.4 Simulation der NOESY-Spektren

Die Rückrechnungen der 2-D-NOESY-Spektren und die der 3D <sup>15</sup>N-NOESY-HSQC-Spektren wurden mit AUREMOL RELAX mit der vorgegebenen Sequenz von Plasmodium falciparum Thioredoxin erstellt. Es wurden Rückrechnungen auf Basis der NMR-Daten, also der korrekten Zuordnung und den besten 10 NMR-Strukturen erstellt. Außerdem wurden Rückrechnungen auf Basis des MD-Ensemble und einer zufälligen Zuordnung erstellt. Es wurden die gleichen Methoden und Parameter wie in Kap. 3.2.4 verwendet, wobei die Parameter soweit wie möglich an die jeweiligen experimentellen Spektren (Kap. 3.3.2) angepasst wurden. Für die Rückrechnung aus den NMR-Daten wurde eine rotatorische Korrelationszeit  $\tau_{\text{rot}}$  von 5,69 ns bestimmt. Es wurden insgesamt 13.369 2-D-NOESY- und 2.521 3D <sup>15</sup>N-NOESY-HSQC-Signale simuliert. Aus dem MD-Ensemble wurden insgesamt 13.258 2-D-NOESY- und 2.471 3D <sup>15</sup>N-NOESY-HSQC-Signale zurückgerechnet.



### 3.3.5 Automatische Zuordnung mit SIBASA

Die automatische Zuordnung mittels Threshold-Accepting wurde mit einem gegenüber der automatischen Zuordnung von HPr *S. aureus* (H15A) erhöhten Anfangs-Threshold  $Th$  von 60 gestartet, da die Pseudoenergie der Optimierung  $E^{\text{total}}$  mit steigender Komplexität des Zuordnungsproblems ebenfalls steigt. Alle anderen Parameter wurden wie bei der Zuordnung des HPr *S. aureus* (H15A) (Kap. 3.2.5) eingestellt.

Durch Zusammenfassung von experimentellen Signalen (Kap. 2.2.8) wurden die 5.372 2-D-NOESY-Signale auf 5.012, die 950 3D  $^{15}\text{N}$ -NOESY-HSQC-Signale auf 949 reduziert.

Auch die Slotlisten wurden auf die gleiche Weise wie in Kap. 3.2.5 erstellt. Insgesamt entstand so eine  $^1\text{H}$ -Slotliste mit 2.527 Slots und eine  $^{15}\text{N}$ -Slotliste mit 69 Slots.

Die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der 715 zuzuordnenden Kerne wurden wieder mit SHIFTX2 und dem 1.000 Strukturen umfassenden MD-Ensemble erzeugt und mit den allgemeinen Dichteverteilungen, die mit SHIFTS und der Nh3D-Datenbank erzeugt worden sind, kombiniert (Kap. 2.2.7).

Für die Erzeugung der Häufigkeitsverteilungen der Verifikationspseudoenergien, die in die AUREMOL Datenbank gespeichert wurden und für die Berechnung der Wahrscheinlichkeit einer richtigen Zuordnung mithilfe des Satzes von Bayes verwendet werden, wurden insgesamt 20 Synchronisationen (1.000.000 Iterationen) mit den Rückrechnungen der NMR-Strukturen durchgeführt.

Nach der 47. Synchronisation (1.175.000 Iterationen) wurde für den Testfall die minimale Pseudoenergie  $E^{\text{total}}$  erreicht und es wurden keine weiteren chemischen Verschiebungen zur Laufzeit mehr fixiert.

### 3.4 Ras(T35S)-GppNHp

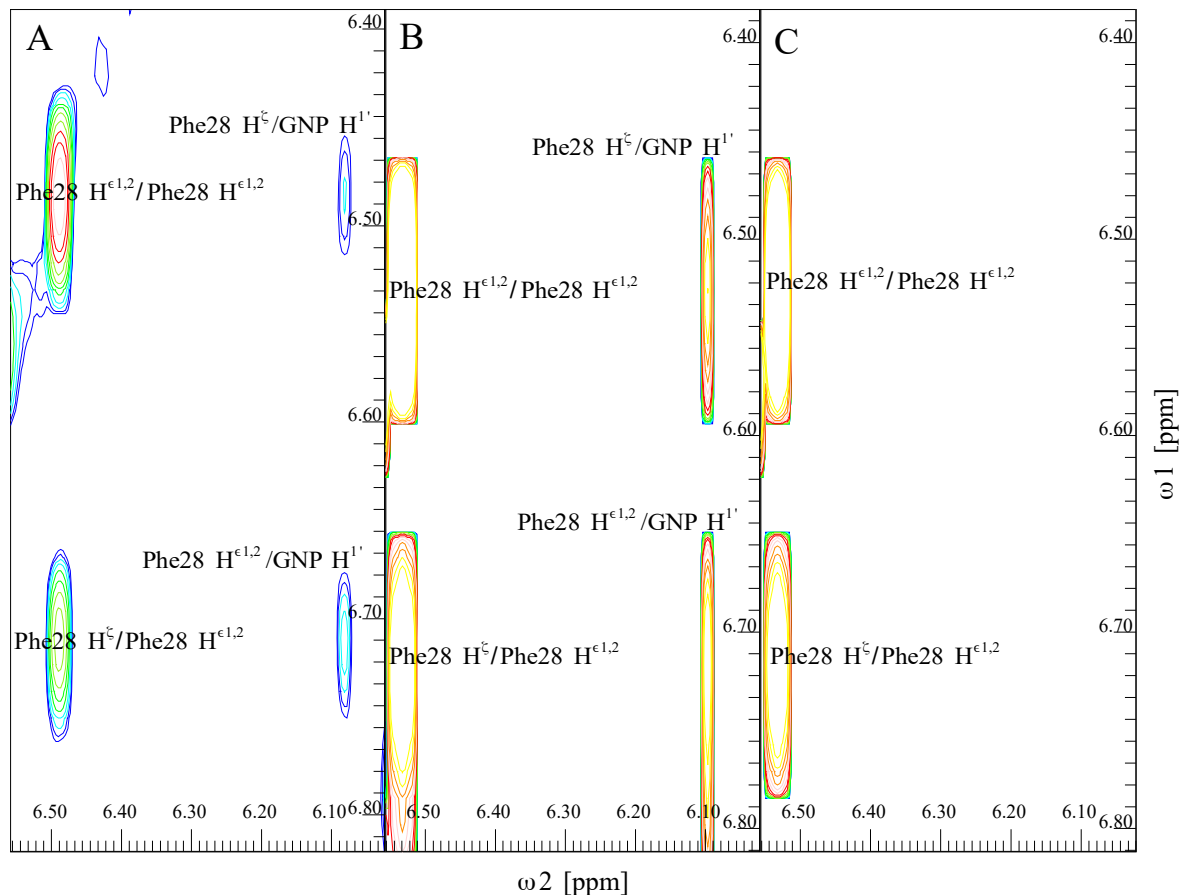


Abbildung 16: **Teil des 2-D-NOESY-Spektrums und der entsprechenden Rückrechnungen aus zwei Röntgenstrukturen von Ras(T35S)-GppNHp.** **A** Experimentelles Spektrum. **B** Rückrechnung aus der Röntgenstruktur von Ras(T35S)-GppNHp Form 2 (PDB ID: 3KKM). **C** Rückrechnung aus der Röntgenstruktur von Ras(T35S)-GppNHp Form 1 (PDB ID: 3KKN). Die Kontakte Phe28 H<sup>ε1,2</sup>/GppNHp H<sup>1'</sup> und Phe28 H<sup>ζ</sup>/GppNHp H<sup>1'</sup> des Proteins zum Nukleotid sind nur im Experiment und in der Rückrechnung von Form 2 zu sehen.

Das Ras-Protein ist ein molekularer Schalter, der zwischen dem GTP-gebundenem aktiven Zustand und dem GDP-gebundenem inaktiven Zustand, wechselt. Zur Weiterleitung eines Signals, wechselwirkt Ras im aktiven Zustand mit Effektoren, wie z.B. Raf-Kinase oder RaGDS. Dadurch werden Zellproliferation, Zelldifferenzierung oder der programmierte Zelltod eingeleitet. Der Wechsel zwischen der aktiven und der inaktiven Form von Ras wird durch GEFs (Guaninnukleotid-Austauschfaktoren) und GAPs (GTPase-aktivierende Proteine) bestimmt (Wittinghofer und Waldmann 2000). In der Krebsforschung spielt das Ras-Protein eine wichtige Rolle, da bestimmte Mutanten des Proteins keine intrinsischen und auch keine GAP-beschleunigten GTPase-Aktivität mehr haben. Fehlt diese, so kann es zu unkontrollierten Zellproliferation, was Tumorbildung begünstigen kann, kommen (Friday und

Adjei 2005). Mittels  $^{31}\text{P}$ -NMR-Spektroskopie konnte gezeigt werden, dass Ras im Komplex mit dem GTP-Analogen GppNHp (GppNHp = Guanosin-5'-( $\beta,\gamma$ -imino)triphosphat) in mindestens zwei konformationellen Zuständen vorliegt, die miteinander im Gleichgewicht stehen (Geyer et al. 1996). Sogenannte *Partial Loss-of-Function-Mutanten*, wie das hier betrachtete Ras(T35S)-GppNHp können nur einen Teil der möglichen Effektoren binden und es konnte gezeigt werden, dass diese Mutante vorwiegend im Zustand 1 vorliegt (Spörner et al. 2001).

Für die automatische Zuordnung lagen zwei Röntgenstrukturen von Ras(T35S)-GppNHp (Shima et al. 2010) vor, wovon die Struktur in Form 2 (PDB ID: 3KKM) nicht vollständig definiert ist. Ihr fehlt die Helix zwischen den Aminosäuren 61 und 71. Die Struktur in Form 1 (PDB ID: 3KKN) ist zwar vollständig, doch sie scheint nicht die in Lösung hauptsächlich vorkommende Form zu sein. In Abb. 16 ist ein Teil des 2-D-NOESY-Spektrums von Ras(T35S)-GppNHp und die entsprechenden Teile der Rückrechnungen der beiden Röntgenstrukturen von Form 1 und von Form 2 gezeigt. Die beiden Kontakte Phe28  $\text{H}^{\epsilon_{1,2}}$ /GppNHp  $\text{H}^1$  und Phe28  $\text{H}^{\delta}$ /GppNHp  $\text{H}^1$  zwischen dem Ras(T35S) und seinem Nukleotid sind sowohl im Experiment als auch in der Rückrechnung der Form 2 deutlich zu sehen (Abstände in der 3D-Struktur von Form 2: 0,412 nm und 0,512 nm bzw. 0,298 nm), fehlen aber in der Rückrechnung von Form 1 (Abstände in der 3D-Struktur von Form 1: 2,62 nm und 2,45 nm bzw. 2,68 nm). Auch die RMSD-Werte der beiden Röntgenstrukturen zu der gemittelten NMR-Struktur von Araki et al. (2011) bestätigen die Annahme, dass Form 2 näher an der Struktur in Lösung liegt als Form 1 (Abb. 58 und Tabelle 12).

Das zuzuordnende H-Ras(T35S) hat 166 Aminosäuren und ein Molekulargewicht von 18,48 kDa.

Es standen ein 2-D-NOESY-Spektrum von Ras(T35S)-GppNHp (Spörner et al.) und ein 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum (Puthenpurackal Narayanan et al.) zur Verfügung. Zur Überprüfung der Ergebnisse der automatischen Zuordnung wurde die chemischen Verschiebungen der Resonanzen der Amidprotonen und Stickstoffe händisch an das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum angepasst (Puthenpurackal Narayanan et al.), die chemischen Verschiebungen der Resonanzen der  $\text{H}^{\alpha}$ - und Seitenkettenprotonen wurden automatisch mit SIBASA optimiert.

### 3.4.1 Probe

Die Probe des 2-D-NOESY-Spektrums enthielt 0,7 mM des unmarkierten Proteins H-Ras(T35S)-GppNHp (Residuen 1-166) welches in 90 %  $\text{H}_2\text{O}$  und 10 %  $\text{D}_2\text{O}$  gelöst war. Durch Zugabe von 15 mM perdeutertem TRIS/HCl wurde der pH auf 7,4 eingestellt. Außerdem wurden 10 mM  $\text{MgCl}_2$ , 2 mM DTE und als Referenz 0,04 mM DSS hinzugegeben.

Die Probe des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums enthielt 1,2 mM des  $^{15}\text{N}$ -markierten

Proteins H-Ras(T35S)-GppNHp (Residuen 1-166), welches in 90 % H<sub>2</sub>O und 10 % D<sub>2</sub>O gelöst war. Durch Zugabe von 40 mM perdeutertem TRIS/HCl wurde der pH auf 7,4 eingestellt. Außerdem wurden 10 mM MgCl<sub>2</sub>, 2 mM DTE und als Referenz 0,04 mM DSS hinzugegeben.

### 3.4.2 NMR Messungen

Das 2-D-NOESY-Spektrum wurden mit einer Mischzeit von 0,15 s und mit 2,37 s Repetitionszeit bei einer Temperatur von 303 K auf einem Bruker Avance-800 NMR bei einer <sup>1</sup>H-Resonanzfrequenz von 800,2 MHz aufgezeichnet. Die Größe der Zeitdomäne betrug 1.024 x 8.192 komplexe Datenpunkte, die spektrale Breite in beiden Dimensionen war 13,9473 ppm. Die Anzahl der Scans war 32. Die FIDs wurden mit LB = -18 Hz und GB=0,1 in beiden Dimensionen mit einer Lorentz-Gauß-Transformation gefiltert und mit TopSpin (Bruker BioSpin, Karlsruhe) auf die Größe von 1.024 x 4.096 reelle Datenpunkte Fourier transformiert. Die Signalliste wurden automatisch mit AUREMOL (Wolfram Gronwald und Hans Robert Kalbitzer 2004) erzeugt und anschließend manuell korrigiert. Die Diagonalsignale wurden wieder entfernt. Im 2-D-NOESY-Spektrum wurden insgesamt 5.843 Signale ausgewählt und mit der verbesserten AUREMOL Integrationsroutine (Donaubauer, Veröffentlichung in Vorbereitung) auf dem Segmentationsniveau von 0,1 integriert.

Das 3D <sup>15</sup>N-NOESY-HSQC-Spektrum wurde mit einem 800 MHz Bruker Spektrometer, einer Mischzeit von 0,15 s und einer Repetitionszeit von 1,1 s bei einer Temperatur von 303 K und einer <sup>1</sup>H-Resonanzfrequenz von 800,2 MHz aufgenommen. Die Zeitdomäne enthielt 256 x 64 x 2.048 komplexe Datenpunkte, die spektrale Breite war 13,9473 x 40 x 13,9873 ppm. Es wurden immer 24 Scans aufgenommen. Der FID wurde in der indirekten <sup>1</sup>H-Dimension mit dem Cosinus Filter (QSINE mit SSB=2), in der <sup>15</sup>N- und der direkten <sup>1</sup>H-Dimension wurde er mit der Lorentz-Gauß-Transformation gefiltert (LB = -8 Hz und GM = 0,1 bzw. LB = -6 Hz und GB= 0,12). Auf die Stickstoff-Dimension wurde eine Lineare Vorwärts Vorhersage (Lpfc) mit 32 Koeffizienten angewendet. Nach der Fourier Transformation mit TopSpin auf 256 x 128 x 2.048 reelle Datenpunkte, wurde die Spektrum mir einer Streifen-Transformation auf die endgültige Größe von 256 x 128 x 544 Datenpunkten gebracht. Die Signalliste wurde automatisch mit AUREMOL erzeugt und die Liste anschließend manuell verbessert. Es wurden 1.745 3D <sup>15</sup>N NOESY-HSQC Signale ausgewählt. Die Signalvolumen wurden wieder mit der AUREMOL Integrationsroutine bestimmt, wobei das Segmentationsniveau wegen dem schlechteren Signal-zu-Rausch-Verhältnis auf 0,2 erhöht werden musste.

### 3.4.3 Molekulardynamiksimulationen

Die Energieminimierungen und MD-Simulationen wurden mit den gleichen Parametern wie in Kap. 3.2.3 durchgeführt. Als Basis der Molekulardynamiksimulation diente die 0,209 nm aufgelöste Röntgenstruktur von Ras(T35S)-GppNHp in Form 1 (Shima et al. 2010) (PDB ID: 3KKN). Die Dynamik des Proteins wurde, ohne zusätzliche Restraints, mit dem  $Mg^{2+}$ -Ion und dem Nukleotid GppNHp simuliert. Damit GROMACS das Nukleotid erkennen konnte, musste zuerst eine separate Topologie Datei erstellt werden. GppNHp wurde dafür in eine separate PDB-Datei eingefügt und mit USCF CHIMERA (Pettersen et al. 2004) in das mol2 Format konvertiert. Anschließend wurde die von benötigte GROMACS itp-Datei mit ACPYPE (Sousa da Silva, Alan W und Vranken 2012), welches das Softwarepaket ANTECHAMBER (Wang et al. 2006) verwendet, erstellt. Dem im Vakuum minimierten Komplex wurden als Lösungsmittel 7.103  $H_2O$ -Moleküle, 789  $D_2O$ -Moleküle, 30  $Na^+$ - und 24  $Cl^-$ -Atome hinzugefügt. Die Temperatur der MD-Simulation betrug 303 K.

Die 1.000 äquidistante Strukturen für die Erzeugung des Wahrscheinlichkeitsdichteverteilungen mit SHIFTX2, die 200 äquidistante Strukturen für die Rückrechnungen mit AUEREMOL RELAX, sowie Korrelationszeiten und Ordnungsparameter für das erweiterte Lipari-Szabo-Modell wurden anschließend wieder aus der Trajektorie extrahiert.

### 3.4.4 Simulation der NOESY-Spektren

Die Rückrechnungen der 2-D-NOESY- und die 3D  $^{15}N$ -NOESY-HSQC-Spektren wurden mit AUEREMOL RELAX und der vorgegebenen Sequenz von Ras(T35S)-GppNHp erstellt. Das Magnesium-Ion und das Nukleotid mussten vorher der AUEREMOL Topologie Datenbank hinzugefügt werden. Es wurden Rückrechnungen für beide NOESY-Spektren aus dem 200 Strukturen umfassenden MD-Ensemble, den Dynamikparametern und einer zufälligen Zuordnung erstellt. Die Rückrechnungen wurden wie in Kap. 3.2.4 erzeugt, wobei die Parameter soweit wie möglich von den jeweiligen experimentellen Spektren (Kap. 3.4.2) übernommen wurden. Aus dem MD-Ensemble wurden insgesamt 21.319 2-D-NOESY- und 4.325 3D  $^{15}N$ -NOESY-HSQC-Signale zurückgerechnet.

Für die Anpassung der Zuordnungen der  $H^\alpha$ - und Seitenkettenprotonen von Araki et al. (2011) auf die vorhanden experimentellen Spektren, wurden Rückrechnungen mit der veröffentlichten Zuordnung und den NMR-Strukturen (PDB ID: 2LCF) erstellt. Die Dynamikparameter blieben wie schon in Kap. 3.2.4 und Kap. 3.3.4 auf den Standardeinstellungen von AUEREMOL RELAX. Die rotatorische Korrelationszeit wurde automatisch auf 6,78 ns bestimmt. Insgesamt wurden 21.773 2-D-NOESY- und 4.441 3D  $^{15}N$ -NOESY-HSQC Signale erstellt.

### 3.4.5 Automatische Zuordnung mit SIBASA

Die automatische Zuordnung mittels Threshold-Accepting wurde mit einem Anfangs-Threshold von 60 gestartet. Die Anzahl der Iteration pro Kühschritt musste auf 100.000 verdoppelt werden. Dies schlug sich auch in Anzahl der Iterationen zwischen zwei Synchronisationsschritten nieder, die ebenfalls auf 500.000 verdoppelt wurde. Alle anderen Parameter wurden wie bei der Zuordnung des HPr *S. aureus* (H15A) (Kap. 3.2.5) eingestellt.

Durch Zusammenfassung von experimentellen Signalen (Kap. 2.2.8) wurden die 5.843 2-D-NOESY-Signale auf 5.634 reduziert. Von den 1.745 3D <sup>15</sup>N-NOESY-HSQC-Signale wurden dagegen keine zusammengefasst.

Die Slotlisten wurden auf die gleiche Weise wie in Kap. 3.2.5 erstellt. Insgesamt entstand so eine <sup>1</sup>H-Slotliste mit 2.419 Slots und eine <sup>15</sup>N-Slotliste mit 72 Slots.

Die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der 1.161 zuzuordnenden Kerne wurden wieder mit SHIFTX2 und dem 1.000 Strukturen umfassenden MD-Ensemble erzeugt und anschließend mit den allgemeinen Verteilungen kombiniert (Kap. 2.2.7).

Nach der 72. Synchronisation (7.200.000 Iterationen) hat sich im Testfall die Pseudoenergie nicht mehr stark verändert und es wurden keine weiteren chemischen Verschiebungen mehr zur Laufzeit fixiert.

### Anpassung der vorhandenen Zuordnung der H<sup>α</sup>- und Seitenkettenprotonen auf die experimentellen Spektren

SIBASA wurde für die Optimierung der chemischen Verschiebung der Resonanzen der H<sup>α</sup>- und Seitenkettenprotonen mit nur einer Instanz, einem Anfangs-Threshold von 40, einer Abkühlrate von 0,985 und mit 15.000 Iterationen zwischen zwei Abkühlungen gestartet.

Da die gleichen experimentellen Spektren wie im Testfall zum Einsatz kamen, sind auch die Ergebnisse der Signalzusammenfassung und die Slotlisten die gleichen.

SIBASA verwendet für die Optimierung von chemischen Verschiebungen Normalverteilungen als Wahrscheinlichkeitsdichtefunktionen (Kap. 2.2.7). Die Mittelwerte stammen dabei aus der zu optimierenden Liste der chemischen Verschiebungen und die Standardabweichungen wurden auf 0,02 ppm gesetzt. Außerdem durfte SIBASA die chemischen Verschiebungen der Resonanzen  $k$  nur im Bereich von  $\delta_k^{\text{Literatur}} \pm 3\sigma_k$  suchen.

## 4 Ergebnisse

### 4.1 Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in Modellpeptiden

Aminosäure	Atom	$\delta$ [ppm]	Aminosäure	Atom	$\delta$ [ppm]
Gly1	H <sup>N</sup>	8,062	Gly1	H <sup>N</sup>	8,069
	H <sup><math>\alpha</math>2</sup>	3,894		H <sup><math>\alpha</math>2</sup>	3,897
	H <sup><math>\alpha</math>3</sup>	3,894		Gly2	H <sup><math>\alpha</math>3</sup>
Gly2	H <sup>N</sup>	8,613	H <sup>N</sup>		8,610
	H <sup><math>\alpha</math>2</sup>	4,022	H <sup><math>\alpha</math>2</sup>		4,025
	H <sup><math>\alpha</math>3</sup>	4,022	H <sup><math>\alpha</math>3</sup>	4,025	
Asn3	H <sup>N</sup>	8,541	Gln3	H <sup>N</sup>	8,485
	H <sup><math>\alpha</math></sup>	4,720		H <sup><math>\alpha</math></sup>	4,351
	H <sup><math>\beta</math>2</sup>	2,838		H <sup><math>\beta</math>2</sup>	2,120
	H <sup><math>\beta</math>3</sup>	2,749		H <sup><math>\beta</math>3</sup>	1,992
	H <sup><math>\delta</math>21</sup>	7,645		H <sup><math>\gamma</math>2</sup>	2,381
	H <sup><math>\delta</math>22</sup>	6,961		H <sup><math>\gamma</math>3</sup>	2,381
			H <sup><math>\epsilon</math>21</sup>	7,593	
			H <sup><math>\epsilon</math>22</sup>	6,898	
Ala4	H <sup>N</sup>	8,433	Ala4	H <sup>N</sup>	8,493
	H <sup><math>\alpha</math></sup>	4,291		H <sup><math>\alpha</math></sup>	4,290
	H <sup><math>\beta</math></sup>	1,403		H <sup><math>\beta</math></sup>	1,401
NH <sub>2</sub>	H <sup>2(E)</sup>	7,600	NH <sub>2</sub>	H <sup>2(E)</sup>	7,608
	H <sup>3(Z)</sup>	7,081		H <sup>3(Z)</sup>	7,070

Tabelle 1: **Vollständige stereospezifische Zuordnung der Tetrapeptide Gly-Gly-X-Ala-NH<sub>2</sub>.** 5mM Gly-Gly-X-Ala-NH<sub>2</sub> und 0,1 mM DSS wurden in 90 % H<sub>2</sub>O/10 % D<sub>2</sub>O gelöst. Die Temperatur betrug 293 K, der pH 3. Die stereospezifische Zuordnungen stammt aus Harsch et al. (2013).

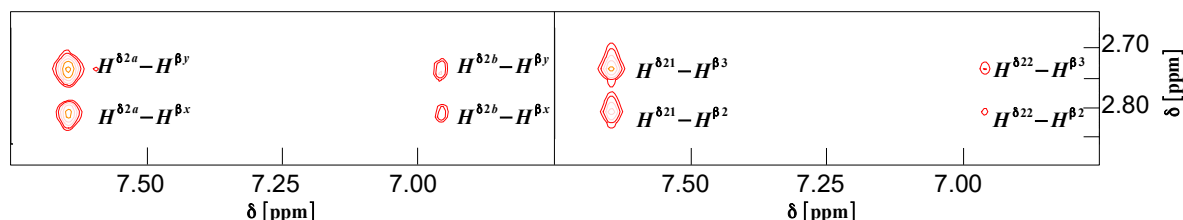
4.1.1 Die stereospezifische Zuordnung von Gly-Gly-Asn-Ala-NH<sub>2</sub>

Abbildung 17: **Ausschnitte des experimentellen und des zugehörigen simulierten 2-D-NOESY-Spektrums von Gly-Gly-Asn-Ala-NH<sub>2</sub>.** Das 800 MHz Spektrum von Gly-Gly-Asn-Ala-NH<sub>2</sub> (links) wurde in 90 % H<sub>2</sub>O/10 % D<sub>2</sub>O bei pH 3 und 293 K mit einer Mischzeit von 0,3 s gemessen. Als Referenz wurden 0,1 mM DSS hinzugegeben. Die Gesamtzeit der Messung betrug 15 h 22 min. Vor der Fouriertransformation wurde es mit 5 Hz Linienverbreiterung exponentiell gefiltert. Beim simulierten Spektrum (rechts) wurden dagegen 7,7 Hz verwendet um die Feldinhomogenität zu berücksichtigen.

Aminosäure	Kreuzsignal A	Kreuzsignal B	$V_A^{\text{exp}}/V_B^{\text{exp}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$
				a=1, b=2 x=2, y=3	a=2, b=1 x=2, y=3	a=1, b=2 x=3, y=2	a=2, b=1 x=3, y=2
Asn3	H <sup>δ2a</sup> –H <sup>βx</sup>	H <sup>δ2b</sup> –H <sup>βx</sup>	3,1	<b>3,4</b>	0,3	3,8	0,3
	H <sup>δ2a</sup> –H <sup>βy</sup>	H <sup>δ2b</sup> –H <sup>βy</sup>	3,8	<b>3,8</b>	0,3	3,4	0,3
	H <sup>δ2a</sup> –H <sup>βy</sup>	H <sup>δ2a</sup> –H <sup>βx</sup>	1,5	<b>1,4</b>	1,2	0,7	0,8
	H <sup>δ2b</sup> –H <sup>βy</sup>	H <sup>δ2b</sup> –H <sup>βx</sup>	1,3	<b>1,2</b>	1,4	0,8	0,7
	H <sup>N</sup> –H <sup>βy</sup>	H <sup>N</sup> –H <sup>βx</sup>	1,9	<b>1,5</b>	1,5	0,7	0,7
	H <sup>α</sup> –H <sup>βx</sup>	H <sup>α</sup> –H <sup>βy</sup>	1,2	<b>2,1</b>	2,1	0,5	0,5
Ala4	H <sup>α</sup> –H <sup>a</sup>	H <sup>α</sup> –H <sup>b</sup>	4,2	<b>4,1</b>	0,2	4,1	0,2

Tabelle 2: **Experimentelle und zurückgerechnete 2-D-NOESY-Kreuzsignalintensitätsverhältnisse von Gly-Gly-Asn-Ala-NH<sub>2</sub>.** Die Experimentbedingungen sind die gleichen wie in in Abb. 17. In der Tabelle werden die gemessenen Volumenverhältnisse zwischen den Kreuzsignalen A und B  $V_A^{\text{exp}}/V_B^{\text{exp}}$  mit den simulierten Verhältnissen verglichen. In den Spalten sind die Volumenverhältnisse aller möglichen stereospezifischen Zuordnungen aufgelistet. Die chemischen Verschiebungen von H<sup>δ2a,b</sup> und H<sup>βx,y</sup> können aus Abb. 17 abgelesen werden. Die Zuordnung, die das experimentelle Ergebnis am besten erklärt ist fett gedruckt.



Aminosäure	Experiment		Theoretisch	
Asn3	${}^3J_{H\alpha H\beta x}$ [Hz]	$5,9 \pm 0,5$	${}^3J_{H\alpha H\beta 2}$ [Hz]	3,7
	${}^3J_{H\alpha H\beta y}$ [Hz]	$7,8 \pm 0,5$	${}^3J_{H\alpha H\beta 3}$ [Hz]	8,1
			$LW_{H\beta 2}$ [Hz]	0,44
			$LW_{H\beta 3}$ [Hz]	0,46

**Tabelle 3: Gemessene und theoretische Werte der  ${}^3J_{H\alpha H\beta}$  Kopplungskonstanten und berechnete Linienbreiten der  $H^\beta$ -Protonen von Asn im Tetrapeptid Gly-Gly-Asn-Ala-NH<sub>2</sub>. Die experimentellen  ${}^3J$ -Kopplungskonstanten wurden mit AUREMOL aus den NMR-Spektren bestimmt, die theoretischen Werte wurden mit der Karplus Gleichung aus der MD-Trajektorie berechnet.**

Die chemischen Verschiebungen von Gly-Gly-Asn-Ala-NH<sub>2</sub> wurden mit einem 2-D-TOCSY-, einem 2-D-NOESY- und einem 1-D-NOESY-Spektrum zugeordnet (Tabelle 1). Abb. 17 zeigt den für die stereospezifische Zuordnung relevanten Bereich der  $H^{\delta 2}$ - $H^\beta$ -Kreuzsignale im experimentellen und im zurückgerechneten 2-D-NOESY-Spektrum.

Die in Abb. 17 verwendete stereospezifische Zuordnung der  $H^{\delta 2}$ -Protonen (Tabelle 1) erklärt die beobachteten 2-D-NOESY-Kreuzsignalintensitäten am besten.

In Tabelle 2 sind alle für die stereospezifische Zuordnung von Gly-Gly-Asn-Ala-NH<sub>2</sub> wichtigen experimentellen und zurückgerechneten Volumenverhältnisse aufgelistet. Die Tabelle zeigt, dass das Experiment und die Simulation nur konsistent sind, falls  $H^{\delta 2 1}$  die Position a, also die größere und  $H^{\delta 2 2}$  die Position b, also die kleiner chemische Verschiebung zugeordnet bekommt. Es geht ebenfalls hervor, dass  $H^{\beta 2}$  an der x-Position und  $H^{\beta 3}$  an der y-Position liegen muss. Um dies zu bestätigen wurde die 1-D-NOESY-Linienform der Asn3  $H^{\beta 3}$ -Signale im Gly-Gly-Asn-Ala-NH<sub>2</sub> unter Berücksichtigung der starken skalaren Kopplung mit NMRSIM simuliert. NMRSIM benötigte dafür die J-Kopplungskonstanten  ${}^3J_{H\alpha H\beta 2}$  und  ${}^3J_{H\alpha H\beta 3}$ , die mit der generalisierten Karplus-Gleichung (Karplus 1963) und den Parametern von (Schmidt 2007) aus den 3D-Strukturen von Gly-Gly-Asn-Ala-NH<sub>2</sub> berechnet und über das ganze Ensemble gemittelt wurden. Die so ermittelten  ${}^3J_{H\alpha H\beta 2,3}$ -Kopplungsraten von 3,7 und 8,1 Hz (Tabelle 3) wurden zusammen mit der gemessenen  ${}^2J_{H\beta 2 H\beta 3}$ -Kopplungskonstante von 15,6 Hz, den beiden chemischen Verschiebungen von  $H^{\beta x,y}$  und den von RELAX ermittelten Linienbreiten an NMRSIM übergeben um die Linienformen der Asn  $H^\beta$ -Signale zu simulieren (Abb. 18). Ein Vergleich mit dem experimentellen 1-D-NOESY-Spektrum zeigt, dass nur die Zuordnung aus Tabelle 1 das gemessene J-Kopplungsmuster erklärt. Die C-Terminalen NH<sub>2</sub>-Protonen von Ala4 konnten mithilfe ihrer Kreuzsignale zum Ala  $H^a$  stereospezifisch zugeordnet werden.

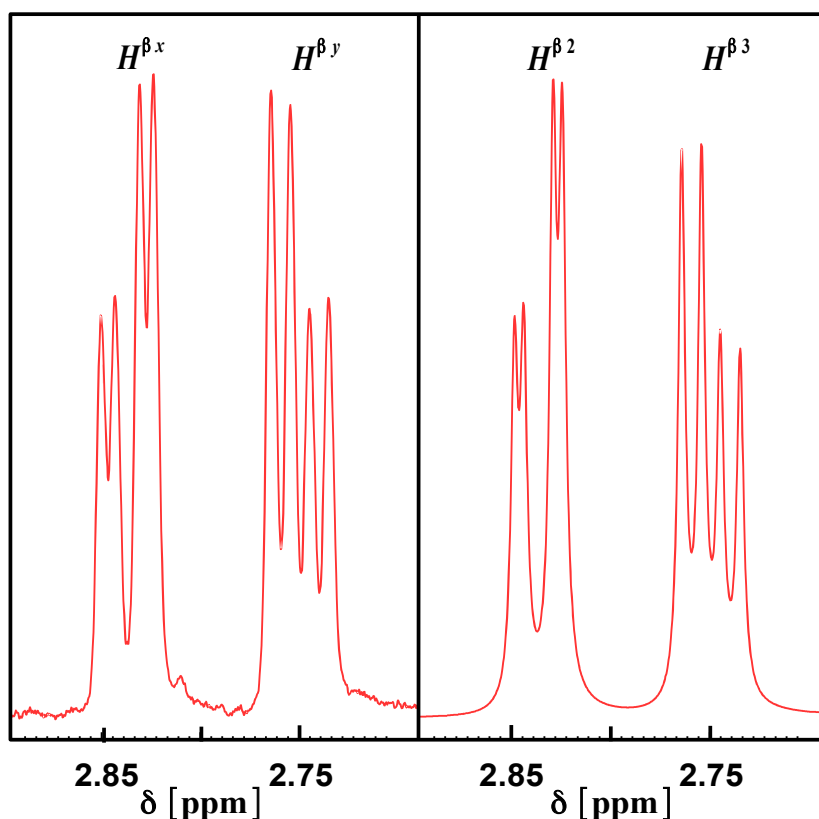


Abbildung 18: **Multiplett Aufspaltung der experimentellen und simulierten der 1-D-Resonanzlinien von  $H^{\beta 2}$  und  $H^{\beta 3}$ .** Das Spektrum von 5 mM Gly-Gly-Asn-Ala-NH<sub>2</sub> in 90 % H<sub>2</sub>O/10 % D<sub>2</sub>O Lösung wurde bei einer <sup>1</sup>H Frequenz von 800,2 MHz, pH 3, 293 K und 0,1 s Mischzeit aufgenommen. Als Referenz wurde 0,1 mM DSS hinzugegeben. Die Messung dauerte insgesamt 34 s. Das experimentelle Spektrum (links) wurde vor der Fouriertransformation exponentiell mit 0,3 Hz Linienverbreiterung gefiltert. Das simulierte Spektrum (rechts) wurde mit den theoretischen J-Kopplungskonstanten und Linienbreiten aus Tabelle 3, den chemischen Verschiebungen aus Tabelle 1 und mit 3,3 Hz Linienverbreiterung simuliert. Die zusätzliche Linienverbreiterung von 3 Hz gegenüber dem Experiment berücksichtigt die Feldinhomogenität des NMR-Spektrometers.

Das Volumenverhältnis der Signale  $H^{\alpha}/H^a$  zu  $H^{\alpha}/H^b$  stimmt bestmöglich überein, falls  $H^2(E)$  die Position a, also die größere und  $H^3(Z)$  die Position b, also die kleiner chemische Verschiebung bekommt.

Die Austauschraten der beiden Flips wurden durch Anpassung der beiden Volumenverhältnissen  $H^{\delta 22}/H^{\beta 3}$  zu  $H^{\delta 21}/H^{\beta 3}$  bzw.  $H^{21}(E)/H^{\alpha}$  zu  $H^{22}(Z)/H^{\alpha}$  an die experimentellen Volumenverhältnisse bestimmt. Die Austauschrate der Asn Seitenkette wurde mit 1,3 Hz und die der C-Terminale Amidgruppe zu 0,3 Hz bestimmt.

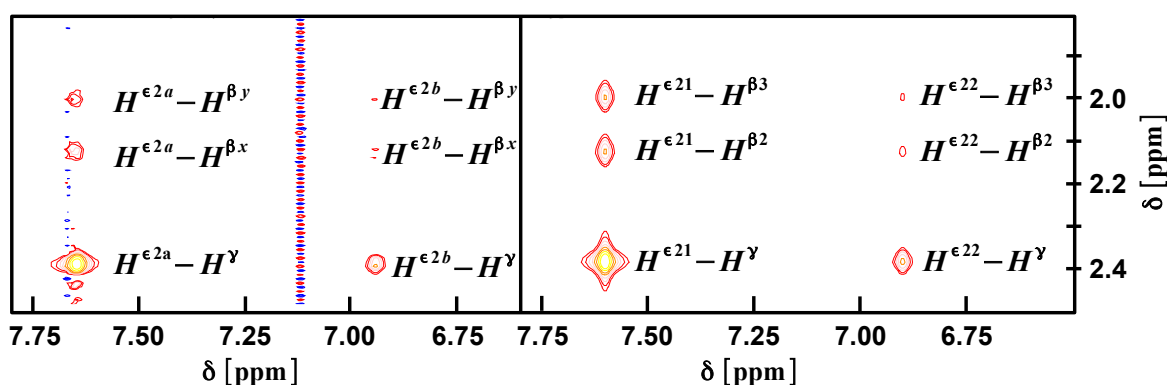
4.1.2 Die stereospezifische Zuordnung von Gly-Gly-Gln-Ala-NH<sub>2</sub>

Abbildung 19: Ausschnitte des experimentellen und des zugehörigen simulierten 2-D-NOESY-Spektrums von Gly-Gly-Gln-Ala-NH<sub>2</sub>. Das 800,2 MHz Spektrum von Gly-Gly-Gln-Ala-NH<sub>2</sub> (links) wurde in 90 % H<sub>2</sub>O/ 10 % D<sub>2</sub>O bei pH 3 und 293 K mit einer Mischzeit von 0,3 s gemessen. Als Referenz wurden 0,1 mM DSS hinzugegeben. Die Gesamtzeit der Messung betrug 15h 23min. Vor der Fouriertransformation wurde es mit 5 Hz Linienverbreiterung in beiden Dimensionen exponentiell gefiltert. Beim simulierten Spektrum (rechts) wurden dagegen 15 Hz in beide Richtungen verwendet, um die Feldinhomogenität und die Signalverbreiterungen durch die endliche Aufnahmezeit zu berücksichtigen.

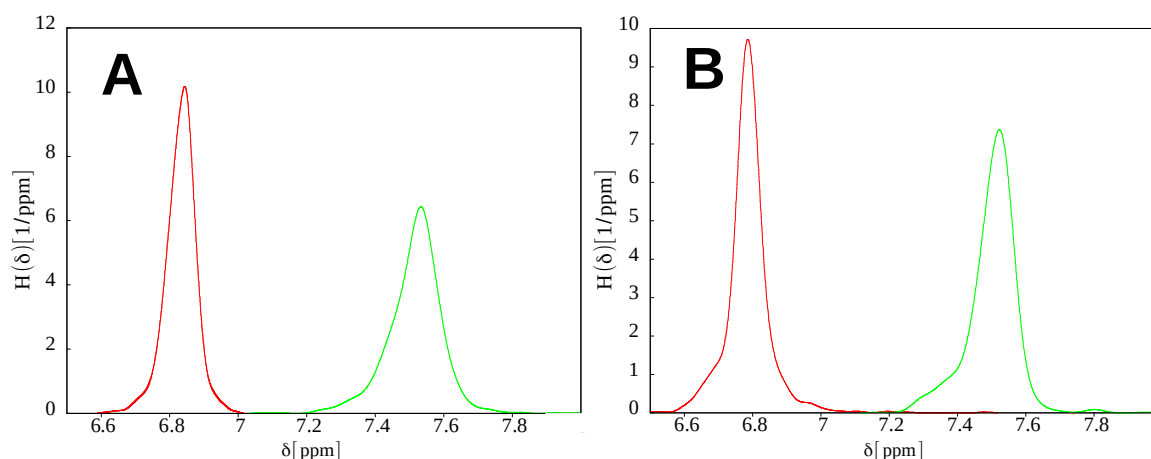
Die Zuordnung des Tetrapeptids Gly-Gly-Gln-Ala-NH<sub>2</sub> wurde ebenfalls mit Hilfe des 2-D-NOESY und des 2-D-TOCSY-Spektrums erstellt (Tabelle 1). In Abb. 19 ist der für die stereospezifische Zuordnung wichtige Abschnitt des experimentellen und des zurückgerechneten 2-D-NOESY-Spektrums abgebildet. Die chemischen Verschiebungen der beiden H<sup>γ</sup>-Protonenresonanzen sind in Gly-Gly-Gln-Ala-NH<sub>2</sub> nahezu identisch, d.h. ihre Linien können in den aufgenommenen NMR-Spektren nicht voneinander getrennt werden. Eine direkte Zuordnung des H<sup>ε21</sup> ist wegen des großen Volumenunterschiedes zwischen den Kreuzsignalen H<sup>ε21</sup>/H<sup>γ</sup> und H<sup>ε22</sup>/H<sup>γ</sup> möglich. Aus Tabelle 4 geht hervor, dass nur die stereospezifische Zuordnung der H<sup>ε2</sup>-Protonen aus Tabelle 1 die hohe Intensität des H<sup>ε2a</sup>/H<sup>γ</sup>-Kreuzsignals im experimentellen NOESY-Spektrum erklären kann. Außerdem geht auch die stereospezifische Zuordnung der H<sup>β</sup>-Protonen aus Tabelle 4 hervor. Die beste Übereinstimmung zwischen experimentellen und simulierten Volumenverhältnissen wird erreicht, wenn H<sup>β2</sup> der x-Position und H<sup>β3</sup> der y-Position (Abb. 19) zugeordnet wird.

Aminosäure	Kreuzsignal A	Kreuzsignal B	$V_A^{\text{exp}}/V_B^{\text{exp}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$	$V_A^{\text{sim}}/V_B^{\text{sim}}$
				a=1, b=2 x=2, y=3	a=2, b=1 x=2, y=3	a=1, b=2 x=3, y=2	a=2, b=1 x=3, y=2
Gln3	$H^{\epsilon 2a} - H^{\beta x}$	$H^{\epsilon 2b} - H^{\beta x}$	3,1	<b>4,2</b>	0,2	4,7	0,3
	$H^{\epsilon 2a} - H^{\beta y}$	$H^{\epsilon 2b} - H^{\beta y}$	5,2	<b>4,7</b>	0,2	4,2	0,2
	$H^{\epsilon 2a} - H^{\gamma x, y}$	$H^{\epsilon 2b} - H^{\gamma x, y}$	5,4	<b>5,5</b>	0,2	5,5	0,2
	$H^{\epsilon 2a} - H^{\beta y}$	$H^{\epsilon 2a} - H^{\beta x}$	1,5	<b>1,0</b>	0,9	1,0	1,1
	$H^{\epsilon 2b} - H^{\beta y}$	$H^{\epsilon 2b} - H^{\beta x}$	0,9	<b>0,9</b>	1,0	1,1	1,0
	$H^{\alpha} - H^{\beta x}$	$H^{\alpha} - H^{\beta y}$	1,2	<b>0,7</b>	0,7	1,4	1,4
	$H^N - H^{\beta y}$	$H^N - H^{\beta x}$	0,8	<b>0,4</b>	0,4	2,7	2,7
Ala4	$H^{\alpha} - H^a$	$H^{\alpha} - H^b$	8,7	<b>6,5</b>	0,2	6,5	0,2

Tabelle 4: *Experimentelle und zurückgerechnete 2-D-NOESY-Kreuzsignalintensitätsverhältnisse von Gly-Gly-Gln-Ala-NH<sub>2</sub>. Die Experimentbedingungen sind die gleichen wie in in Abb. 17. In der Tabelle werden die gemessenen Volumenverhältnisse zwischen den Kreuzsignalen A und B  $V_A^{\text{exp}}/V_B^{\text{exp}}$  mit den simulierten Verhältnissen verglichen. In den Spalten sind die Volumenverhältnisse aller möglichen stereospezifischen Zuordnungen aufgelistet. Die chemischen Verschiebungen von  $H^{\epsilon 2a, b}$ ,  $H^{\gamma x, y}$  und  $H^{\beta x, y}$  können aus Abb. 19 abgelesen werden. Die Zuordnung, die das experimentelle Ergebnis am besten erklärt ist fett gedruckt.*

Wie beim Asn können auch beim Gly-Gly-Gln-Ala-NH<sub>2</sub> die Protonen der C-terminale NH<sub>2</sub>-Gruppe stereospezifisch mithilfe der Ala4 H<sup>α</sup>-Kreuzsignale zugeordnet werden. Aus Tabelle 4 ist ersichtlich, dass H<sup>1</sup>(E) die Position a und H<sup>2</sup>(Z) die Postion b einnehmen muss. Auch für Gly-Gly-Gln-Ala-NH<sub>2</sub> mussten die Flipaustauschraten der Amidgruppen bestimmt werden. Die Fliprate der Gln Seitenkettenamidgruppe ergibt sich zu 0,4 Hz, die Fliprate der C-terminalen NH<sub>2</sub>-Gruppe liegt bei 0 Hz. Als Referenzvolumen wurden die beiden H<sup>ε2</sup>/H<sup>γ</sup>-Signale, bzw. die H<sup>1,2</sup>/H<sup>α</sup>-Signale des Ala4 verwendet.

### 4.1.3 Erzeugung der Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der Amidprotonen



**Abbildung 20:** Aus den Molekulardynamik Rechnungen geschätzte Häufigkeitsverteilungen der chemischen Verschiebungen der Seitenkettenamide. Die Verteilungen wurden mit den Ergebnissen der Vorhersagen von SHIFTS aus den 1.000 Strukturen des MD-Ensembles erstellt und mithilfe eines Kerndichteschätzers geglättet. Die Random-Coil-Werte  $\delta_{\text{random-coil}}$  von SHIFTS wurden mit den stereospezifischen korrigierten Werten ersetzt. Die 1.000 Strukturen waren dabei äquidistant über die gesamte Trajektorie verteilt. Die Atompositionen wurden stereospezifisch korrigiert.

**A** Gly-Gly-Asn-Ala-NH<sub>2</sub>: die Maxima für H<sup>δ21</sup> (grün) und H<sup>δ22</sup> (rot) befinden sich bei 7,53 und 6,84 ppm. Die Standardabweichungen der beiden Verteilungen betragen 0,08 bzw. 0,05 ppm.

**B** Gly-Gly-Gln-Ala-NH<sub>2</sub>: die Maxima für H<sup>ε21</sup> (grün) und H<sup>ε22</sup> (rot) befinden sich bei 7,52 und 6,79 ppm. Die Standardabweichungen der beiden Verteilungen betragen 0,10 bzw. 0,07 ppm.

Mit speziellen Softwarepaketen, wie z.B. SHIFTS (Osapay und Case 1991) ist es möglich die vollständige Zuordnungsliste direkt aus den 3D-Strukturen von Proteinen abzuschätzen. Die Random-Coil-Verschiebungen die SHIFTS für seine Berechnungen verwendet, sind 7,03 ppm für Asn H<sup>δ21</sup>, 7,74 ppm für Asn H<sup>δ22</sup>, 6,97 ppm für Gln H<sup>ε21</sup> und 7,69 ppm für Gln H<sup>ε22</sup>. Sowohl beim Gln als auch beim Asn widerspricht die Reihenfolge der Random-Coil-Werte der NH<sub>2</sub>-Protonen der Seitenketten unseren Ergebnissen. In Abb. 20 sind die stereospezifisch korrigierten Häufigkeitsverteilungen der chemischen Verschiebungen der Amidprotonen von Asn und Gln abgebildet. Die Verteilungen wurden mit SHIFTS aus den 1.000 Strukturen des Gly-Gly-Asn-Ala-NH<sub>2</sub> bzw. Gly-Gly-Gln-Ala-NH<sub>2</sub> Ensembles (Kap. 3.1.3) erzeugt, wobei die Random-Coil-Werte von Asn H<sup>δ21</sup> und Asn H<sup>δ22</sup> bzw. Random-Coil-Werte von Gln H<sup>ε21</sup> und Gln H<sup>ε22</sup> mit den chemischen Verschiebungen dieser Arbeit ersetzt wurden.

## 4.2 Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen

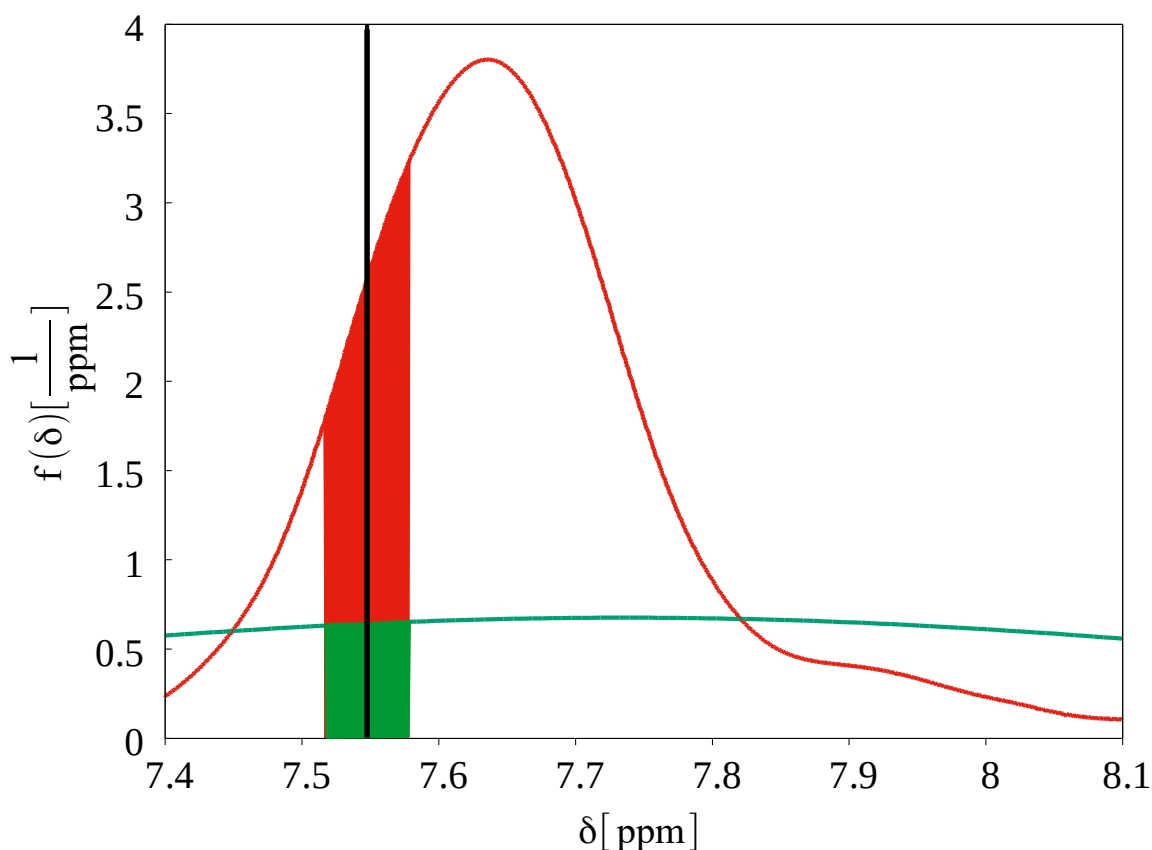


Abbildung 21: Auf zwei verschiedene Arten vorhergesagte Wahrscheinlichkeitsdichtefunktion  $f_{Ala59 H^N}(\delta)$  der chemischen Verschiebung  $\delta$  von Ala59  $H^N$  im Protein HPr *S. aureus* (H15A). In **rot** ist die mit SHIFTS2 und einem Kerndichteschätzer aus dem MD-Ensemble erzeugt Wahrscheinlichkeitsdichte eingezeichnet. In **grün** ist die Normalverteilung  $N(\delta_{Ala59 H^N}^{SHIFTS}, \sigma_{Ala H^N}^{BMRB})$  der chemischen Verschiebungen, deren Mittelwert  $\delta_{Ala59 H^N}^{SHIFTS}$  mit SHIFTS aus einer 3D-Struktur geschätzt wurde und deren Standardabweichung die Standardabweichung aller in der BMRB abgelegten chemischen Verschiebungen von Ala  $H^N$   $\sigma_{Ala H^N}^{BMRB}$  ist. Die **schwarze** Linie markiert die richtige chemische Verschiebung  $\delta_{Ala59 H^N}^{Referenz}$  der Resonanz von Ala59  $H^N$ . Die 0,06 ppm breite Fläche um den richtigen Wert der chemischen Verschiebung, die von den Graphen begrenzt wird (**grün** und **rot + grün**), ist die Wahrscheinlichkeit der richtigen Vorhersage  $P_{Ala59 H^N}^{Vorhersage(korrekt)}$  der entsprechenden geschätzten Häufigkeitsverteilung von Ala59  $H^N$  (Gl. 90).

Vorhergesagte Wahrscheinlichkeitsdichtefunktionen der chemischen Verschiebungen  $f_k(\delta)$  der Resonanzen der Kerne  $k$  eines Proteins sind wichtige Informationen für die automatische Zuordnung (Kap. 2.27). Für ASSIGN (Brunner 2006), den Vorgänger von SIBASA, wurde

angenommen, dass die chemischen Verschiebungen normalverteilt wären. Die Mittelwerte der chemischen Verschiebung  $\delta_k^{\text{SHIFTS}}$  der Normalverteilungen  $N(\delta_k^{\text{SHIFTS}}, \sigma_{\text{AA,Kern}}^{\text{BMRB}})$  wurden mit SHIFTS (Osapay und Case 1991) aus einer einzigen Struktur geschätzt und die Standardabweichung  $\sigma_{\text{AA,Kern}}^{\text{BMRB}}$  (AA = Aminosäure wie z.B. Ala, Kern = H<sup>N</sup>, N, H<sup>α</sup>, ...) stammte aus der BMRB (Ulrich et al. 2008).

SIBASA verwendet dagegen Kombinationen von allgemeinen und speziellen vorhergesagten Wahrscheinlichkeitsdichtefunktionen. Die allgemeinen Wahrscheinlichkeitsdichten  $f_{\text{AA,Kern}}(\delta)$  wurden aus der energieminierten unbiased *Nh3D*-Strukturdatenbank (Thiruv et al. 2005) mit SHIFTS geschätzt (Gl. 39), die speziellen Wahrscheinlichkeitsdichten  $f_k(\delta)$  der chemischen Verschiebungen der Resonanzen der Kerne  $k$  wurden mit SHIFTX2 (Han et al. 2011) und einem MD-Ensemble des zuzuordnenden Proteins vorhergesagt (Gl. 40). Beide Wahrscheinlichkeitsdichten wurden mit einem Kerndichtschätzer mit Gauß-Kernel gebildet und mit Gl. 41 miteinander kombiniert.

Um den Nutzen der vorhergesagten Wahrscheinlichkeitsdichten der chemischen Verschiebungen für die automatische Zuordnung messen zu können, müssen die Wahrscheinlichkeit einer richtigen Vorhersage der chemischen Verschiebung des Kerns  $k$  bestimmt werden.

$$P_k^{\text{Vorhersage}}(\text{korrekt}) = \int_{\delta_k^{\text{Referenz}} - 0,03 \text{ ppm}}^{\delta_k^{\text{Referenz}} + 0,03 \text{ ppm}} f_k(\delta) d\delta \quad (90)$$

$P_k^{\text{Vorhersage}}(\text{korrekt})$  ist die Wahrscheinlichkeit, dass die chemische Verschiebung  $\delta_k^{\text{Referenz}}$  des Kerns  $k$  innerhalb eines Fehlers von 0,03 ppm, was ungefähr einer Linienbreite entspricht, mithilfe einer vorhergesagten Wahrscheinlichkeitsdichte  $f_k(\delta)$  richtig gefunden wird. Sie ist ein Maß für den Anteil der MOVES in SIBASA, die die Zuordnung des Kerns  $k$  innerhalb der Toleranz von 0,03 ppm richtig finden. Abb. 21 zeigt auf zwei Weisen vorhergesagte Wahrscheinlichkeitsdichteverteilungen  $f_k(\delta)$  der chemischen Verschiebung von Ala59 H<sup>N</sup> im Protein HPr *S. aureus* (H15A). In **grün** ist die Dichte  $f_k(\delta)$  die ASSIGN und in **rot** die Dichte, die SIBASA vorhergesagt hat, eingezeichnet. Die Flächen um die richtige chemische Verschiebung  $\delta_{\text{Ala59 H}^{\text{N}}}^{\text{Referenz}}$  der Resonanz (**schwarz**) stellen die Wahrscheinlichkeiten der richtigen Vorhersage dar. Man kann erkennen, dass die Fläche unter der vorhergesagten Wahrscheinlichkeitsdichtekurve von SIBASA (**rot+grün**) größer ist als unter der von ASSIGN (**grün**). Die beiden Dichten sind normiert, was bedeutet, dass die Wahrscheinlichkeit der richtigen Vorhersage der chemischen Verschiebung  $P_k^{\text{Vorhersage}}(\text{korrekt})$  maximal 1 werden kann.

Die automatische Zuordnung eines Proteins kann durch eine große Anzahl von gut vorhergesagten chemischen Verschiebungen, also durch einen hohen Anteil von großen Vorhersagewahrscheinlichkeiten der richtigen chemischen Verschiebung  $P_k^{\text{Vorhersage}}(\text{korrekt})$  stark beschleunigt werden.

	HPr S. aureus (H15A)			Reduziertes PfTrx			Ras(T35S)-GppNHp		
	SHIFTS (1 Struktur)	SHIFTX2 (1 Struktur)	SHIFTX2 (MD- Ensemble)	SHIFTS (1 Struktur)	SHIFTX2 (1 Struktur)	SHIFTX2 (MD- Ensemble)	SHIFTS (1 Struktur)	SHIFTX2 (1 Struktur)	SHIFTX2 (MD- Ensemble)
$\bar{P}_{H^N}^{\text{Vorhersage}}(\text{korrekt})$ [%]	3,2	3,6 (+13 %)	4,9 (+ 53 %)	2,8	3,3 (+18 %)	4,6 (+64 %)	3,6	3,9 (+ 8 %)	3,8 (+6 %)
$\bar{P}_{^{15}\text{N}}^{\text{Vorhersage}}(\text{korrekt})$ [%]	6,8	7,7 (+13 %)	12 (+76 %)	6,0	7,6 (+27 %)	11 (+83 %)	7,3	8,8 (+21 %)	11,5 (+58 %)
$\bar{P}_{H^\alpha}^{\text{Vorhersage}}(\text{korrekt})$ [%]	5,3	5,7 (+ 8 %)	11 (+108 %)	4,8	5,2 (+8 %)	9,2 (+92 %)	5,1	5,4 (+6 %)	8,0 (+57 %)
$\bar{P}_{SC}^{\text{Vorhersage}}(\text{korrekt})$ [%]	7,4	7,9 (+8 %)	16 (+116 %)	6,9	7,5 (+9 %)	12,0 (+74 %)	6,8	7,4 (+9 %)	12,3 (+81 %)
$\bar{P}_{^1\text{H}}^{\text{Vorhersage}}(\text{korrekt})$ [%]	6,3	6,7 (+6 %)	13 (+106 %)	5,9	6,4 (+8 %)	10 (+69 %)	6,0	6,5 (+ 9 %)	10,2 (+70 %)

**Tabelle 5: Durchschnittliche Wahrscheinlichkeiten der Vorhersage der richtigen chemischen Verschiebung der Resonanzen von verschiedenen Kerntypen  $\bar{P}^{\text{Vorhersage}}(\text{korrekt})$ .** Die Mittelwerte wurden über alle Hauptkettenamidprotonen  $H^N$ , alle Hauptkettenstickstoffe  $^{15}\text{N}$ , alle  $H^\alpha$ -Protonen, alle Seitenkettenprotonen SC und über alle Protonen  $^1\text{H}$  gebildet und für drei Proteine ausgewertet. Die zugrundeliegenden vorhergesagten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebung  $f_k(\delta)$  der Kerne  $k$  wurden auf drei verschiedene Arten bestimmt (siehe Text). In Klammern sind die relativen Verbesserungen zu der alten Methode aus ASSIGN (jeweils 1. Spalte bzw. 1. Methode im Text) angegeben.

Tabelle 5 listet die mittleren Wahrscheinlichkeiten der richtigen Vorhersage der chemischen Verschiebung  $\bar{P}^{\text{Vorhersage}}(\text{korrekt})$  für die Resonanzen von  $H^N$ -,  $H^\alpha$ -, Stickstoff- und Seitenkettenatomen in den drei Testproteine HPr S. aureus (H15A), PfTrx und Ras(T35S)-GppNHp auf. Die Wahrscheinlichkeitsdichten  $f_k(\delta)$  wurden für die Tabelle auf drei verschiedene Arten vorhergesagt:

1. Mit Normalverteilungen  $N(\delta_k^{\text{SHIFTS}}, \sigma_{AA, \text{Kern}}^{\text{BMRB}})$  ( $k$  Index des Kerns im Protein,  $AA$  = Aminosäure wie z.B. Alanin,  $\text{Kern}$  = Kernart wie  $H^N$ ,  $N$ ,  $H^\alpha$ , ...), wobei die chemischen Verschiebungen  $\delta_{\text{ResNr}, \text{Kern}}^{\text{SHIFTS}}$  mit SHIFTS aus der initialen Struktur geschätzt wurden und die Standardabweichungen  $\sigma_{AA, \text{Kern}}^{\text{BMRB}}$  aus der BMRB (Ulrich et al. 2008) stammen. Dies ist die Methode, die ASSIGN verwendet hat.
2. Mit Normalverteilungen  $N(\delta_{\text{ResNr}, \text{Kern}}^{\text{SHIFTX2}}, \sigma_{AA, \text{Kern}}^{\text{BMRB}})$ , wobei diesmal die chemischen Verschiebungen  $\delta_{\text{ResNr}, \text{Kern}}^{\text{SHIFTX2}}$  mit dem neueren SHIFTX2 aus der initialen 3D-Struktur bestimmt worden sind.
3. Mit den den Wahrscheinlichkeitsdichtefunktionen  $f_k(\delta)$  aus SIBASA (siehe oben und Kap. 2.2.7).

Für alle drei Testproteine und alle Mittlungen konnten, mit der Ausnahme der  $H^N$ -Protonen des Ras(T35S)-GppNHp, durchschnittlich höhere Wahrscheinlichkeiten der richtigen



Vorhersage  $\bar{P}^{\text{Vorhersage}}(\text{korrekt})$ , sowohl mit dem neue Programm SHIFTX2 als auch mit den kombinierten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen aus SIBASA erreicht werden. Gegenüber den Wahrscheinlichkeitsdichten der chemischen Verschiebung von ASSIGN konnten mit denen aus SIBASA die Wahrscheinlichkeit der richtigen Vorhersage um bis zu 116 % gesteigert werden. Vor allem verbesserten sich die Vorhersagen der chemischen Verschiebungen der Resonanzen der Seitenkettenprotonen und Stickstoffe in SIBASA, aber auch die Genauigkeit der Vorhersagen der chemischen Verschiebungen von  $H^N$ -Protonenresonanzen konnten um bis zu 64 % gesteigert werden.

In Abb. 22 sind die Häufigkeitsverteilungen  $H(P_{i_H}^{\text{Vorhersage}}(\text{korrekt}))$  der Wahrscheinlichkeiten der richtigen Vorhersage  $P_k^{\text{Vorhersage}}(\text{korrekt})$  der Protonen  $^1H$  für die drei betrachteten Proteine und die drei Methoden der Vorhersage der Wahrscheinlichkeitsdichten der chemischen Verschiebung  $f_k(\delta)$  aufgetragen. Es wurden nur die Protonen betrachtet, da die Slots der Stickstoffe in SIBASA durch das zugehörige Amidproton, das 3D  $^{15}N$ -NOESY-HSQC-Spektrum und durch die Pseudoendenergie des Signalvergleichs ohne die Vorhersagen der chemischen Verschiebungen der Resonanzen der Stickstoffe bestimmt werden (Kap. 2.2.11). Durch den Austausch des Vorhersageprogramms SHIFTS (Methode 1) (**rot**) mit dem moderneren SHIFTX2 (Methode 2) (**grün**) wurde der Anteil von kleinen Wahrscheinlichkeiten der richtigen Vorhersage für alle drei Proteine reduziert. Durch den Einsatz der kombinierten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebung von SIBASA (Methode 3) (**blau**) werden die Graphen der Häufigkeitsverteilungen  $H(P_k^{\text{Vorhersage}}(\text{korrekt}))$  breiter. In ihnen sind die Werte von  $H(P_k^{\text{Vorhersage}}(\text{korrekt}) = 0)$  höher als für die anderen beiden Methoden, was bedeutet, dass mehr Zuordnungen von Protonenresonanzen nicht vorhergesagt werden konnten. Der Anteil der für die automatische Zuordnung sehr gut vorhergesagbaren Zuordnungen von Resonanzen  $H(P_k^{\text{Vorhersage}}(\text{korrekt}) > 15 \%)$  konnte dagegen mit Methode 3 gesteigert werden.

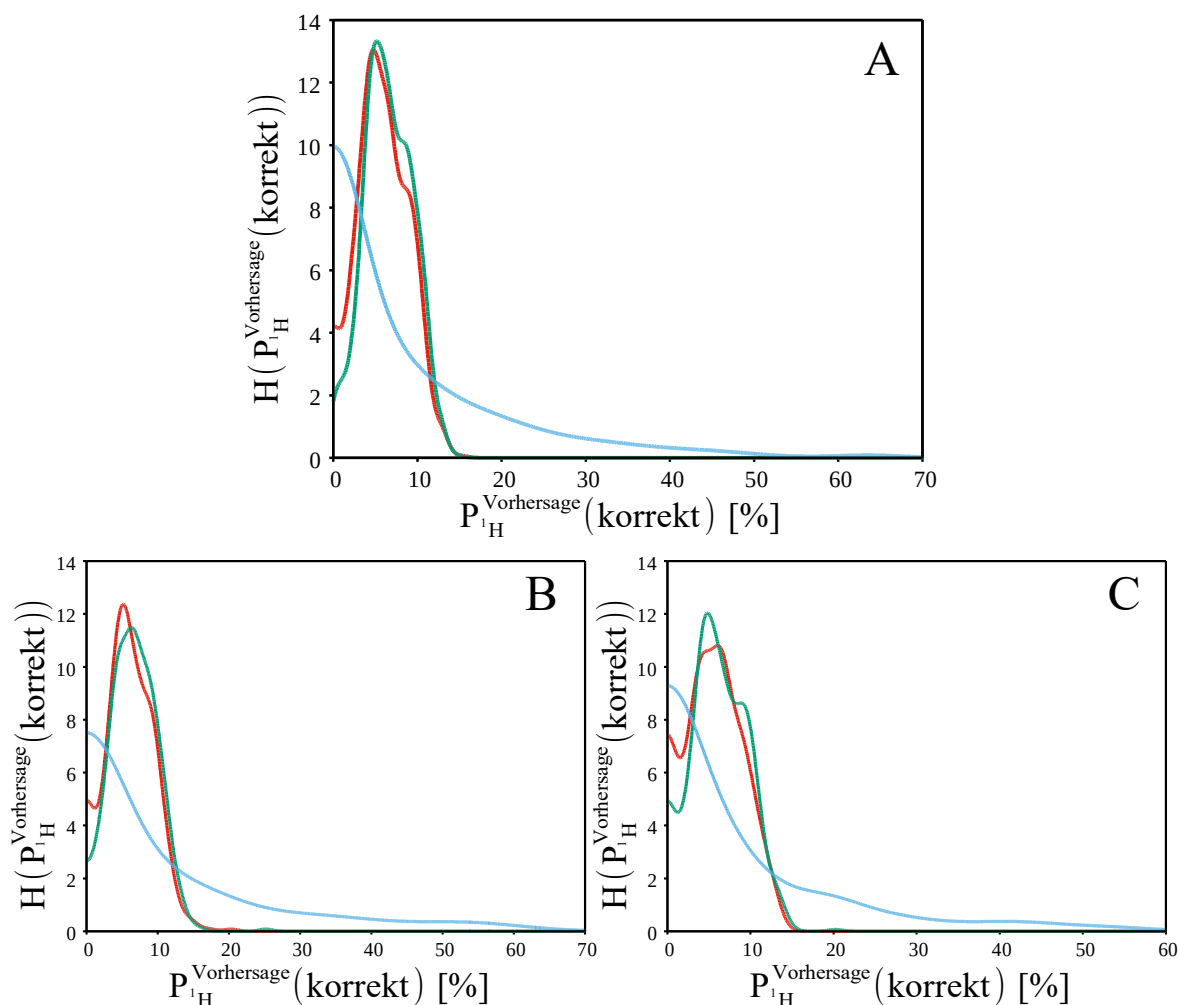


Abbildung 22: Häufigkeitsverteilungen  $H(P_{1H}^{\text{Vorhersage(korrekt)}})$  der Wahrscheinlichkeiten der richtigen Vorhersage der chemischen Verschiebungen der Resonanzen von Protonen  $^1\text{H}$   $P_{1H}^{\text{Vorhersage(korrekt)}}$ . Die den Wahrscheinlichkeiten der richtigen Vorhersage zugrundeliegenden Wahrscheinlichkeitsdichteverteilungen  $f_k(\delta)$  der chemischen Verschiebungen der Resonanzen der Protonen  $^1\text{H}$  wurden auf drei Arten vorhergesagt (siehe Text). Die **roten** Graphen tragen die Häufigkeit der Wahrscheinlichkeiten einer richtigen Vorhersage, die mit den alten Wahrscheinlichkeitsdichten der chemischen Verschiebungen aus ASSIGN (Methode 1) bestimmt worden sind, auf. Für die **grünen** Graphen wurde das Zentrum der Normalverteilung mit SHIFTX2 statt mit SHIFTS bestimmt (Methode 2). Für den **blauen** wurden die neuen kombinierten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen von SIBASA verwendet (Methode 3 und Kap. 2.2.7).

*A* Ras(T35S)-GppNHp, *B* HPr *S. aureus* (H15A) und *C* reduziertes *Plasmodium falciparum* Thioredoxin.

## 4.3 Automatische sequenzielle Zuordnung (SIBASA)

### 4.3.1 Erzeugung der Häufigkeitsverteilungen der Pseudoenergien für die AURMOL Datenbank

Für die Berechnung der lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $P_{ij}^{\text{Vol, Shape}}$  bzw.  $P_{ij}^{\text{Vol, Shape, sim}}$  in SIBASA werden die Häufigkeitsverteilungen der Pseudoenergien des Volumens und die Häufigkeitsverteilungen der Pseudoenergien der Linienform für die vollständige richtige und eine vollständige falsche Zuordnung benötigt (Kap. 2.2.8). Hierfür wurden die Rückrechnungen der NOESY-Spektren aus den 20 besten NMR-Strukturen der Punktmutante des HPr *S. aureus* (H15A) verwendet (Kap. 3.2.4).

Die Pseudoenergien des Volumens ( $E^{\text{Vol}}$  bzw.  $E^{\text{Vol, sim}}$ ) und der Linienform ( $E^{\text{Shape}}$  bzw.  $E^{\text{Shape, sim}}$ ) wurden auf den Grundflächen der experimentellen bzw. der simulierten Signale des 2-D-NOESY- und des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums berechnet. Für Häufigkeitsverteilungen der Pseudoenergien der richtigen Zuordnung  $H_i^{\text{korrekt}}(E^{\text{Vol}})$ ,  $H_i^{\text{korrekt}}(E^{\text{Shape}})$ ,  $H_i^{\text{korrekt}}(E^{\text{Vol, sim}})$  und  $H_i^{\text{korrekt}}(E^{\text{Shape, sim}})$  wurden die Pseudoenergien mit der vollständigen Zuordnung berechnet, für die Verteilungen der falschen Zuordnung  $H_i^{\text{falsch}}(E^{\text{Vol}})$ ,  $H_i^{\text{falsch}}(E^{\text{Shape}})$ ,  $H_i^{\text{korrekt}}(E^{\text{Vol, sim}})$  und  $H_i^{\text{falsch}}(E^{\text{Shape, sim}})$  wurden die Pseudoenergien mit einer zufälligen Zuordnung berechnet. Es wurde jeweils ein Kerndichteschätzer mit Gauß-Kernel verwendet, um aus den gesammelten Pseudoenergien Häufigkeitsverteilungen zu erzeugen (Gl. 52 und Gl. 53). Der Index  $i$  der Häufigkeitsverteilungen unterscheidet hierbei zwischen dem 2-D-NOESY- und 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Da sowohl die 20 besten NMR-Strukturen als auch die Liste der chemischen Verschiebungen händisch aus den beiden experimentellen NOESY-Spektren bestimmt worden sind, eignet sich der Datensatz zur Erzeugung der für die Berechnung der lokalen Wahrscheinlichkeiten der richtigen Zuordnung  $P_{ij}^{\text{Vol, Shape}}$  benötigten Häufigkeitsverteilungen.

Mit dem diesem Datensatz und einem Datensatz, der aus den NMR-Strukturen und der vollständigen Zuordnung des Proteins PfTrx (Kap. 3.3.4) bestand, wurden die Häufigkeitsverteilungen der Verifikationspseudoenergie der richtigen  $H^{\text{korrekt}}(E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  und der falschen Zuordnung  $H^{\text{falsch}}(E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  bestimmt (Kap. 2.2.15). Mit dem Satz von Bayes und den Häufigkeitsverteilungen, die in der AUREMOL Datenbank abgelegt worden sind, ist es möglich die Wahrscheinlichkeit der richtigen Zuordnung  $P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}})$  für die automatisch bestimmte chemische Verschiebung der Resonanz eines Kerns  $k$  zu bestimmen.

## Automatisch bestimmte Volumenskalierungsfaktoren

	2-D-NOESY	3D <sup>15</sup> N-NOESY-HSQC
$\alpha_i^{\text{statistisch}}$	0,00032 (-41 %)	0,092 (-60 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	480 (+23 %)	135 (-34 %)
Anzahl der Signale	3432	1030
$\bar{\alpha}_i^{\text{ref}}$	0,00044 (-19 %)	0,23
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	398 (+ 2 %)	204
Anzahl der Signale	2040	453
$\alpha_i^{\text{Ring}}$	0,00054	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	389	
Anzahl der Signale	69	

**Tabelle 6: Mit verschiedenen Methoden bestimmte Volumenskalierungsfaktoren und deren Auswirkung auf die Pseudoenergie des Volumens der beiden verwendeten NOESY-Spektren von HPr S. aureus (H15A).**  $\alpha_i^{\text{statistisch}}$  ist der statistisch bestimmte Skalierungsfaktor und  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  die über alle experimentelle Kreuzsignale  $j$  des Spektrums  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ) aufsummierte Pseudoenergie des Volumens. In Klammern sind die relativen Unterschiede zu den Werten des aus Ringkontakten bestimmten Skalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$  des 2-D-NOESY-Spektrums bzw. zu den Werten des aus allen eindeutig zuordenbaren Signalen bestimmten Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  des 3D <sup>15</sup>N-NOESY-HSQC-Spektrums angegeben. Die Rückrechnungen der NOESY-Spektren wurden mit dem NMR-Ensemble und der vollständigen richtigen Zuordnung erstellt. Die zusätzlichen Linienverbreiterungen wurden statistisch bestimmt.

Als Teil der Vorbereitungen der automatisch Zuordnung wurden die Volumenskalierungsfaktoren  $\alpha_i^{\text{statistisch}}$  mit der statistischen Methode aus Kap. 2.2.4 für alle Eingabespektren  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ) bestimmt. In Abb. 5 wurden bereits die Häufigkeitsverteilungen der logarithmierten experimentellen (**rot**), simulierten (**grün**) und skalierten simulierten 2-D-NOESY-Signalvolumen (**blau**) des Proteins HPr S. aureus (H15A) abgebildet. Abb. 23 zeigt die Verteilungen der logarithmierten Volumen des 3D <sup>15</sup>N-NOESY-HSQC-Spektrums. In beiden Abbildungen ist eine gute Übereinstimmung zwischen den experimentellen und den skalierten simulierten Volumenverteilungen zu erkennen. Mit der vorhanden Liste der chemischen Verschiebung können den experimentellen Signalen zurückgerechneten Signale zugeordnet werden, was die Bestimmung der Referenzskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$  mithilfe von Gl. 30 möglich macht.

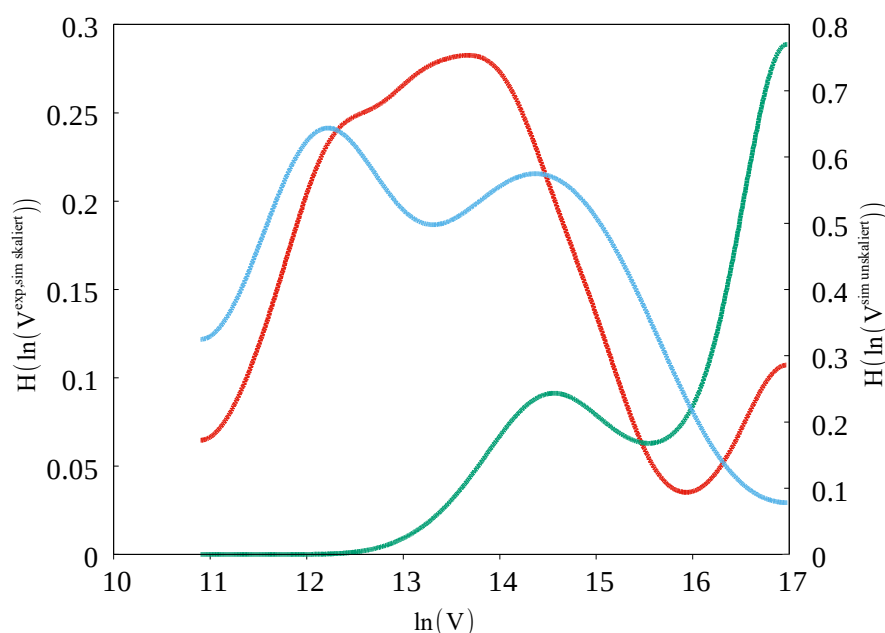


Abbildung 23: Häufigkeitsverteilungen der logarithmierten Signalvolumen  $H(\ln(V))$  im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum von HPr *S. aureus* (H15A). Die Verteilung der experimentellen Volumina  $V^{\text{exp}}$  sind in rot (y-Achse), die Verteilung der simulierten Volumina  $V^{\text{sim unskaliert}}$  in grün (y-Achse) und die Verteilungen der skalierten simulierten Volumina  $V^{\text{sim skaliert}}$  in blau (y2-Achse) eingezeichnet. Für die Rückrechnungen wurde das NMR-Ensemble und die vollständige richtige Zuordnung verwendet. Die zusätzlich Linienverbreiterung wurde statistisch bestimmt.

Tabelle 6 listet die statistisch bestimmten Skalierungsfaktoren  $\alpha_i^{\text{statistisch}}$  und die Referenzskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$  für beide NOESY-Spektren auf. Zusätzlich wurde der Volumenskalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$  nur mit den intraresidualen Ringkontakten des Phe und Tyr bestimmt. Er basiert ausschließlich auf NOE-Kontakten mit festem Abstand und ist deshalb gegen falsche bestimmte 3D-Strukturen relativ robust. Diese Methode ist für Strukturaufklärung relevant, da für kurze Mischzeiten von unter 0,1 s die  $d^{-6}$ -Abhängigkeit des Volumens vom Abstand (Gl. 28) noch ausreichend gut erfüllt ist. Außerdem sind in der Tabelle die Werte der summierten Pseudoenergie des Volumens  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  der beiden Eingabespektren  $i$  aufgelistet, wobei die Rückrechnungen mit den unterschiedlichen Methoden skaliert wurden. Je verlässlicher die Methode der Skalierung war, desto besser war die summierte Pseudoenergie der Volumens für das 2-D-NOESY-Spektrum. Im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum ist der statistische berechnete Skalierungsfaktor  $\alpha_{3\text{-D-NOESY}}^{\text{statistisch}}$  dem Referenzskalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  überlegen.

## Automatisch bestimmte zusätzliche Linienverbreiterungen

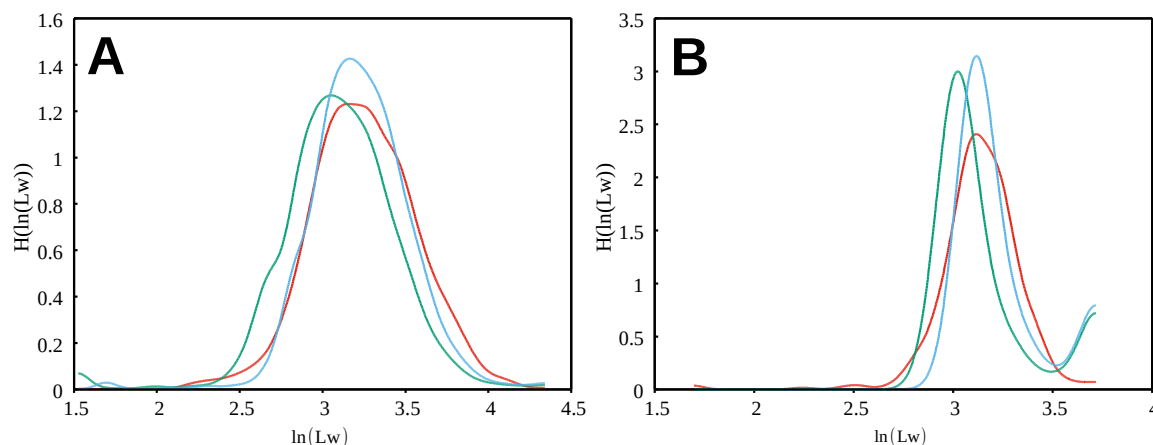


Abbildung 24: **Häufigkeitsverteilungen  $H(\ln(Lw))$  der logarithmierten Linienbreiten  $\ln(Lw)$  der direkten  $^1\text{H}$ -Dimensionen der beiden NOESY-Spektren von HPr *S. aureus* (H15A).** Die Verteilungen der experimentellen Linienbreiten sind in **rot**, die der simulierten in **grün** und die der verbreiteteren simulierten Linienbreiten in **blau** eingezeichnet. Die Rückrechnungen der NOESY-Spektren wurden mit dem NMR-Ensemble und der vollständigen richtigen Referenzzuordnung erstellt. **A** 2-D-NOESY-Spektrums und **B** 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums

Die zusätzliche Linienverbreiterung wurde statistisch mit der Methode aus Kap. 2.2.5 für beide Eingangsspektren berechnet. In Tabelle 7 sind die bestimmten Werte der Linienverbreiterung und die Pseudoenergien des Volumens  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  und die Pseudoenergien Linienform  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$  der Eingangsspektren  $i$  für die verbreiterten und nicht verbreiterten Rückrechnungen angegeben. Nur für die jeweils am besten aufgelösten Dimension der Spektren konnte eine Linienverbreiterung bestimmt werden. Werden diese als zusätzliche Signalverbreiterung an RELAX übergeben, so ergeben sich bessere Werte für die Pseudoenergien der richtig zugeordneten Spektren. Abb. 24 zeigt die Häufigkeitsverteilungen der logarithmierten experimentellen (**rot**), der logarithmierten simulierten (**grün**) und der logarithmierten verbreiterten simulierten Linienbreiten (**blau**) der direkten  $^1\text{H}$ -Dimensionen der beiden Eingangsspektren. Wie schon bei den Signalvolumen, konnte die Übereinstimmungen der Verteilungen der experimentellen und simulierten Linienbreiten durch die statistische Korrektur verbessert werden.

	2-D-NOESY	3D <sup>15</sup> N-NOESY-HSQC
Verbreiterung $\omega 1$ [Hz]	0	0
Verbreiterung $\omega 2$ [Hz]	2,1	0
Verbreiterung $\omega 3$ [Hz]		0,1
$\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Shape}}$ ohne Verbreiterung	544	267
$\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Shape}}$ mit Verbreiterung	533 (-2 %)	267 (0 %)
$\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Vol}}$ ohne Verbreiterung	396	204
$\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Vol}}$ mit Verbreiterung	389 (-2 %)	204 (0 %)

**Tabelle 7: Mit SIBASA statistisch ermittelte Werte der zusätzlichen Linienverbreiterungen und deren Auswirkungen auf die Pseudoenergien des Volumens und auf die Pseudoenergie der Linienform für die beiden NOESY-Spektren.** Hierbei sind  $\omega 1$ ,  $\omega 2$  und  $\omega 3$  die Dimensionen der Spektren,  $\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Vol}}$  bzw.  $\sum_{j=1}^{N_i^{\text{Singale}}} E_{ij}^{\text{Shape}}$  die über alle experimentelle Signale  $j$  summierten Pseudoenergien des Volumens bzw. der Linienform im Spektrum  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ). In Klammern sind die relative Verbesserung der Pseudoenergien zu den nicht verbreiterten Rückrechnungen angeben. Die Rückrechnungen der NOESY-Spektren wurden vom NMR-Ensemble mit der vollständigen richtigen Zuordnung erstellt.

## Häufigkeitsverteilungen der lokalen Pseudoenergien

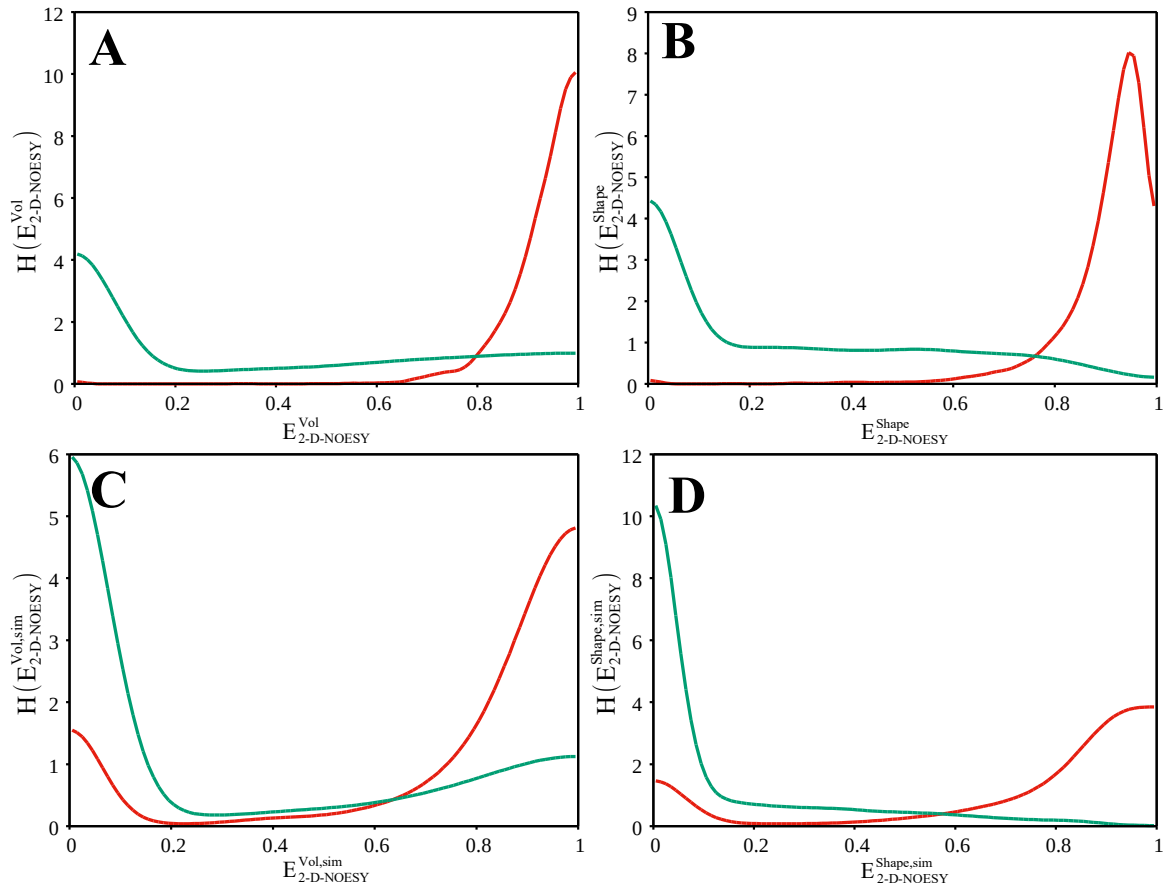


Abbildung 25: **Häufigkeitsverteilungen  $H(E)$  der lokalen Pseudoenergien, die auf den experimentellen und simulierten Signalen des 2-D-NOESY-Spektrums von HPr *S. aureus* berechnet und in die AUREMOL Datenbank aufgenommen wurden.** In rot sind die Verteilungen der Pseudoenergien der korrekten Zuordnung und in grün die Verteilungen der Pseudoenergien einer zufälligen Zuordnung aufgetragen. Die Rückrechnung wurden mit dem NMR-Ensemble, der vollständigen richtigen Zuordnung erstellt und mit dem aus Ringkontakten bestimmten Volumenskalierungsfaktor  $\alpha_{2-D-NOESY}^{Ring}$  skaliert. Die zusätzliche Linienverbreiterung wurde statistisch bestimmt. **A** Verteilungen  $H(E_{2-D-NOESY}^{Vol})$  der Pseudoenergie des Volumens  $E_{2-D-NOESY}^{Vol}$  auf den Grundflächen der experimentellen Signale, **B** Verteilungen  $H(E_{2-D-NOESY}^{Shape})$  Pseudoenergie der Linienform  $E_{2-D-NOESY}^{Shape}$  auf den Grundflächen der experimentellen Signale, **C** Verteilungen  $H(E_{2-D-NOESY}^{Vol,sim})$  Pseudoenergie des Volumens  $E_{2-D-NOESY}^{Vol,sim}$  auf den Grundflächen der simulierten Signale und **D** Verteilungen Pseudoenergie  $H(E_{2-D-NOESY}^{Shape,sim})$  der Linienform  $E_{2-D-NOESY}^{Shape,sim}$  auf den Grundflächen der simulierten Signale. Nicht zuordenbare experimentelle Signale wurden für die Häufigkeitsverteilungen in **A** und **B** weggelassen. Alle acht Verteilungen wurden in die AUREMOL Datenbank aufgenommen.



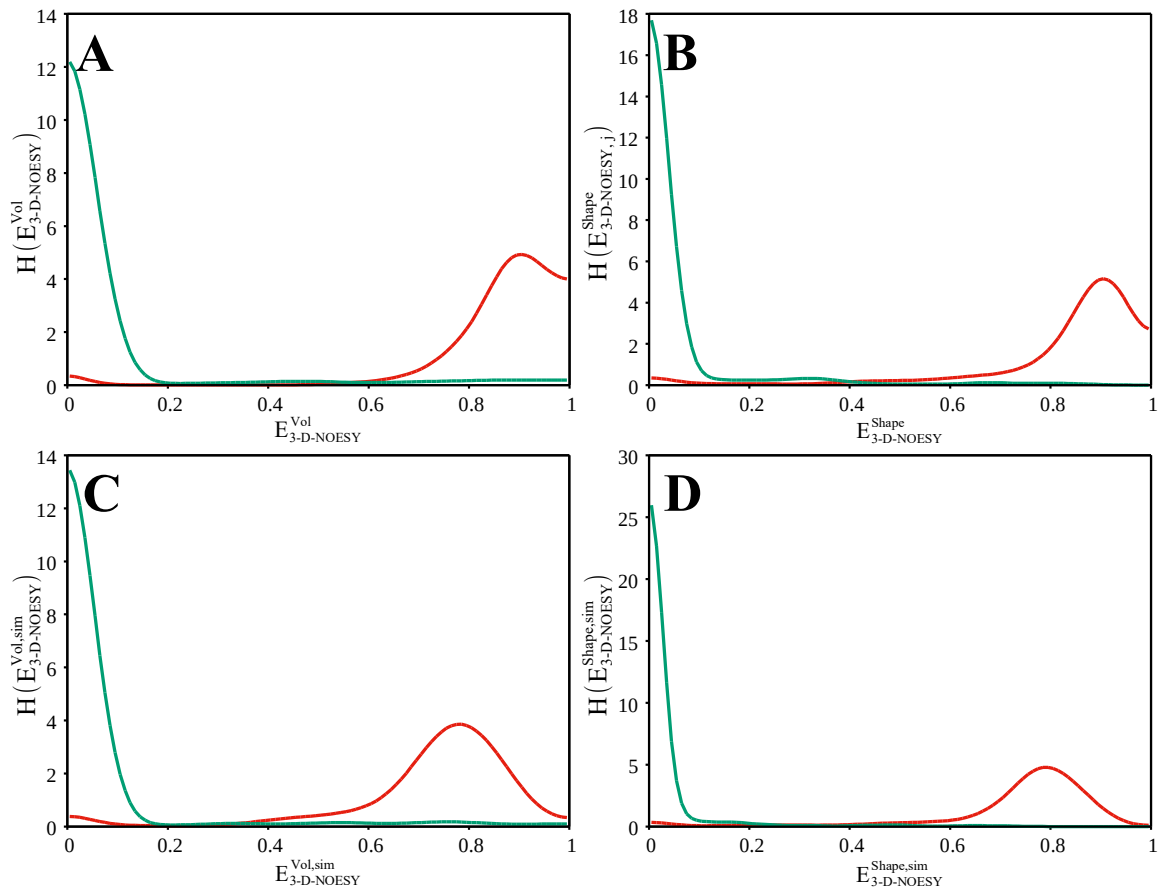


Abbildung 26: **Häufigkeitsverteilungen  $H(E)$  der lokalen Pseudoenergien, die auf den experimentellen und simulierten Signalen des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums von HPr *S. aureus* berechnet und in die AUREMOL Datenbank aufgenommen wurden.** Die Beschreibung der Abbildung und Bezeichnungen sind die gleichen wie in Abb. 25 mit dem Unterschied, dass die lokalen Pseudoenergien auf dem 3-D-NOESY-Spektrum berechnet wurden. Die Rückrechnung wurden mit dem NMR-Ensemble, der vollständigen richtigen Zuordnung erstellt und mit dem Volumenskalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  skaliert.

Um die lokale Wahrscheinlichkeit der richtigen Zuordnung  $P_{ij}^{\text{Vol, Shape}}$  auf einer Grundfläche  $ij$  mit dem Satz von Bayes bestimmen zu können, benötigt SIBASA die Häufigkeitsverteilungen der Pseudoenergien des Volumens  $E^{\text{Vol}}$  und die Häufigkeitsverteilungen der Pseudoenergien der Linienform  $E^{\text{Shape}}$  von der richtigen und einer falschen Zuordnung (Kap. 2.2.8). Diese Verteilungen wurden auf Basis des gelösten Proteins HPr *S. aureus* (H15A), zu dem alle Daten vorhanden waren, gebildet und als Referenz für SIBASA in die AUREMOL Datenbank gespeichert. Die Rückrechnungen 2-D-NOESY-Spektrums bzw. des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums wurden mit dem Volumenskalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$ , der nur mit Hilfe von Ringkontakten berechnet wurde bzw. mit Referenzskalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{ref}}$  skaliert. Die zusätzlichen Linienverbreiterungen wurden automatisch bestimmt (Tabelle 6 und 7). Damit der

dynamischen Bereich der lokalen Pseudoenergie zwischen nicht belegten und perfekt erklärten experimentellen Signalen für den Optimierungsalgorithmus hoch genug ist, war es notwendig nur experimentelle Signale, die mit der Rückrechnung und der Liste der richtigen chemischen Verschiebung zugeordnet werden konnten, für die Häufigkeitsverteilungen der Pseudoenergien der experimentellen Signale zu berücksichtigen. Dadurch wurde gewährleistet, dass nicht erklärbare experimentelle Kreuzsignale während der automatischen Zuordnung stark genug bestraft werden. Die **grünen** Kurven in Abb. 25 zeigen die Verteilungen  $H_{2-D-NOESY}^{\text{korrekt}}(E^{\text{Vol}})$  (**A**),  $H_{2-D-NOESY}^{\text{korrekt}}(E^{\text{Shape}})$  (**F**),  $H_{2-D-NOESY}^{\text{korrekt}}(E^{\text{Vol,sim}})$  (**C**) und  $H_{2-D-NOESY}^{\text{korrekt}}(E^{\text{Shape,sim}})$  (**D**) des 2-D-NOESY-Spektrums, die **roten** Kurven zeigen die entsprechenden Verteilungen für eine zufällige vollständige Zuordnung. Abb. 26 zeigt die nochmal die gleichen Verteilungen für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Abb. 27 zeigt schließlich für 2-D-NOESY-Spektren und das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren die Graphen der lokalen Wahrscheinlichkeiten  $P_i^{\text{Vol,Shape}}(E^{\text{Vol}}, E^{\text{Shape}})$  bzw.  $P_i^{\text{Vol,Shape,sim}}(E^{\text{Vol,sim}}, E^{\text{Shape,sim}})$  in Abhängigkeit von den beiden Pseudoenergien einer experimentellen Grundfläche bzw. einer simulierten Grundfläche. SIBASA verwendet bei der automatische Zuordnung, unabhängig vom Protein, diese abgebildeten Zusammenhänge für die Berechnung der lokalen Wahrscheinlichkeiten aus den gemessenen Pseudoenergien.

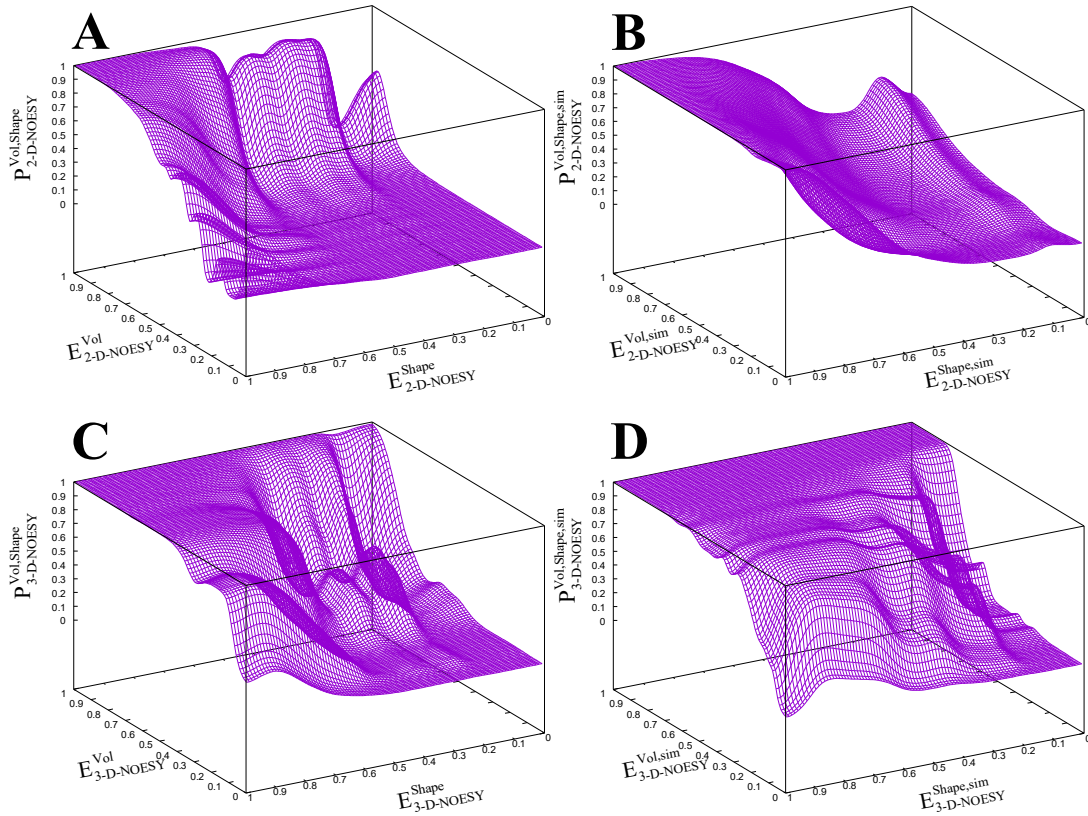


Abbildung 27: Graphen der lokalen Wahrscheinlichkeiten  $P^{Vol,Shape}$  bzw.  $P^{Vol,Shape,sim}$  in Abhängigkeit von der gemessenen Kombination der Pseudoenergien  $E^{Vol}/E^{Shape}$ . **A** Die lokale Wahrscheinlichkeit  $P^{Vol,Shape}(E_{2-D-NOESY}^{Vol}, E_{2-D-NOESY}^{Shape})$  einer experimentellen Grundfläche im 2-D-NOESY-Spektrum, **B** die lokale Wahrscheinlichkeit  $P^{Vol,Shape,sim}(E_{2-D-NOESY}^{Vol,sim}, E_{2-D-NOESY}^{Shape,sim})$  einer simulierten Grundfläche im 2-D-NOESY-Spektrum, **C** die lokale Wahrscheinlichkeit  $P^{Vol,Shape}(E_{3-D-NOESY}^{Vol}, E_{3-D-NOESY}^{Shape})$  einer experimentellen Grundfläche im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum **D** die lokale Wahrscheinlichkeit  $P^{Vol,Shape,sim}(E_{3-D-NOESY}^{Vol,sim}, E_{3-D-NOESY}^{Shape,sim})$  einer simulierten Grundfläche im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum.

Die Rückrechnungen der NOESY-Spektren wurden mit dem NMR-Ensemble, der vollständigen richtigen Zuordnung und den Volumenskalierungsfaktoren  $\alpha_{2-D-NOESY}^{Ring}$  (2-D-NOESY) bzw.  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  (3D  $^{15}\text{N}$ -NOESY-HSQC) erstellt. Die zusätzlichen Linienverbreiterungen wurden statistisch bestimmt. Nicht zuordenbare experimentelle Signale wurden für die Bestimmung der Wahrscheinlichkeitsfunktionen in **A** und **C** ignoriert.

## Häufigkeitsverteilungen der Verifikationspseudoenergien

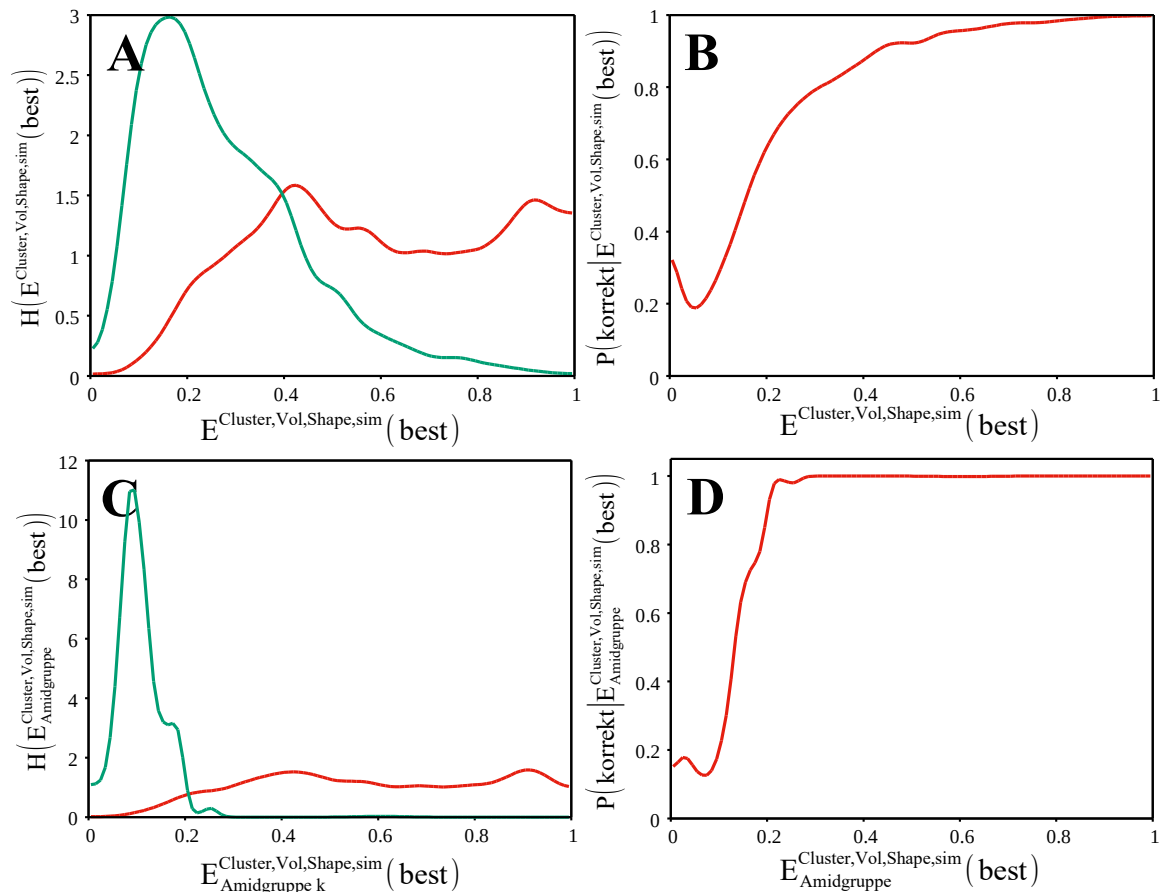


Abbildung 28: **Häufigkeitsverteilungen der Verifikationspseudoenergien und Bayesche Wahrscheinlichkeiten der richtigen Zuordnung.** **A** Häufigkeitsverteilungen  $H(E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der Verifikationspseudoenergie  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der richtigen (**rot**) und falschen Zuordnungen (**grün**), **B** Abhängigkeit der Wahrscheinlichkeit einer richtigen Zuordnung  $P(\text{korrekt}|E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  von  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$ , **C** Häufigkeitsverteilungen  $H(E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der Verifikationspseudoenergien  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der richtigen (**rot**) und falschen Zuordnungen (**grün**) von Mitgliedern von Amidgruppen, **D** Wahrscheinlichkeit der richtigen Zuordnung eines Mitglieds einer Amidgruppe  $P(\text{korrekt}|E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  in Abhängigkeit von seiner Verifikationspseudoenergie. Die Rückrechnungen von HPr *S. aureus* (H15A) und dem reduzierten PfTrx wurden mit den NMR-Ensembles und den richtigen vollständigen Zuordnungen erzeugt. Volumenskalierungsfaktoren und zusätzliche Linienverbreiterungen wurden statistisch bestimmt. Während der Synchronisationsphasen der jeweils 16 SIBASA Instanzen sind keine Zuordnungen fixiert worden.

Um mit dem Satz von Bayes die Wahrscheinlichkeit für eine korrekte Zuordnung  $k$  bestimmen zu können (Kap. 2.2.15), benötigt man die Häufigkeitsverteilungen der Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  (Gl. 70) sowohl für richtige, als auch für falsche Zuordnungen. Diese Verteilungen wurden folgendermaßen erzeugt und in die

AUREMOL Datenbank gespeichert.

Für die automatische Zuordnung der beiden Proteine HPr S. aureus (H15A) und PfTrx, zu denen die vollständige Zuordnung und ein NMR-Strukturensemble vorhanden waren, wurden jeweils 16 SIBASA Instanzen mit den Parameter aus Kap. 3.2.5 und 3.3.5 gestartet. Als Basis der Rückrechnungen dienten jeweils die NMR-Daten. Die Volumenskalierungsfaktoren und die zusätzlichen Linienverbreiterungen wurden statistisch bestimmt (Tabelle 7 für HPr S. aureus (H15A)). Nach jeweils 250.000 MOVES wurde zwar keine Synchronisation, wie sie Kap. 2.2.13 beschrieben sind, durchgeführt, aber alle Tupel  $(\bar{\delta}_k^w(\text{best}); E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der gewichten mittleren chemischen Verschiebungen und Verifikationspseudoenergien gespeichert. Es wurden insgesamt 20 Synchronisationen für beide Proteine durchgeführt. Nachdem die gesammelten Tupel mit Hilfe der bekannten vollständigen korrekten Zuordnungslisten in Richtig und Falsch aufgeteilt worden sind, wurden mithilfe eines Kerndichteschätzers und eines Gauß-Kernel (Kap. 2.2.15) die Häufigkeitsverteilungen  $H^{\text{korrekt}}(E^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  und  $H^{\text{falsch}}(E^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der Verifikationspseudoenergien der richtigen und falschen Zuordnungen (Abb. 28 A) erzeugt. Alle Tupel der beiden Proteine gingen, unabhängig vom Zeitpunkt der Synchronisation, gleich stark in die Verteilungen ein. Auf die gleiche Weise wurden die Verteilungen  $H^{\text{korrekt}}(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  und  $H^{\text{falsch}}(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (Abb. 28 C) der Verifikationspseudoenergien der Amidgruppen erstellt. Die A-priori-Wahrscheinlichkeiten  $P(\text{korrekt})$  bzw.  $P(\text{falsch}) = 1 - P(\text{korrekt})$ , wurden mit der Anzahl der richtigen  $N^{\text{richtig}}$  und falschen Tupel  $N^{\text{falsch}}$  folgendermaßen bestimmt:

$$P(\text{korrekt}) = \frac{N^{\text{korrekt}}}{N^{\text{korrekt}} + N^{\text{falsch}}} \quad (91)$$

Die Abhängigkeit der Wahrscheinlichkeit einer richtigen Zuordnung von der Verifikationspseudoenergie  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (Abb. 28 D) für Mitglieder von Amidgruppen und die Abhängigkeit der Wahrscheinlichkeit einer richtigen Zuordnung  $P(\text{korrekt} | E^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (Abb. 28 C) von der Verifikationspseudoenergie wurden mit dem Satz von Bayes berechnet (Gl. 89) und in der AUREMOL Datenbank abgelegt.

Für Synchronisation in SIBASA zur Laufzeit ist es wichtig, dass möglichst wenig falschen Zuordnungen vorzeitig fixiert werden. Aus Abb. 28 A und Abb. 28 C (jeweils **grüne** Kurve) ist zu erkennen, dass dies ab einem Wert von 0,85 für die Verifikationspseudoenergie  $E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  und ab einem Wert von 0,26 für Verifikationspseudoenergie von Mitglieder von Amidgruppen  $E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  nicht mehr geschieht. Dies entspricht einer richtigen Zuordnungswahrscheinlichkeit von 94 % (Abb. 28 A) bzw. 100 % (Abb. 28 C). Zwar ist bereits ein Grenzwert  $p_{\text{Amidgruppe}}$  (Gl. 87) von 0,26 für die Verifikationspseudoenergie von Amidgruppen ausreichend um keine falschen chemischen Verschiebungen zu fixieren, aber es hat sich gezeigt, dass dieser Wert mit zunehmender

molekularer Masse des zuzuordnenden Proteins, vor allem während der frühen Synchronisationsphasen, höher gewählt werden muss. Für HPr *S. aureus* (H15A) (9,43 kDa) war noch ein Grenzwert  $p_{\text{Amidgruppe}}$  von 0,17 ausreichend, für PfTrx (11,75 kDa) dagegen war schon ein Wert von 0,26 notwendig. Für den Grenzwert  $p$  (Gl. 87) der Verifikationspseudoenergie einer beliebigen Zuordnung  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  reicht für beide Proteingrößen ein Wert von 0,85 aus. Der Grenzwert  $p_{\text{Amidgruppe}}$  wird deshalb für die Synchronisationen zur Laufzeit auf  $0,85^2 \approx 0,73$  gesetzt.

Die Grenzwerte  $q$  und  $q_{\text{Amidgruppe}}$  (Gl. 88) der Pseudoenergien der Zuordnung eines  $H^\alpha$ - oder Seitenkettenprotons bzw. der Pseudoenergien der Zuordnung eines Mitglieds einer Amidgruppe werden ebenfalls auf 0,85 bzw. 0,73 für alle Instanzen des besten Cluster gesetzt.

### 4.3.2 Automatische Zuordnung der Punktmutante HPr *S. aureus* (H15A)

#### Vorbereitung der Röntgenstruktur

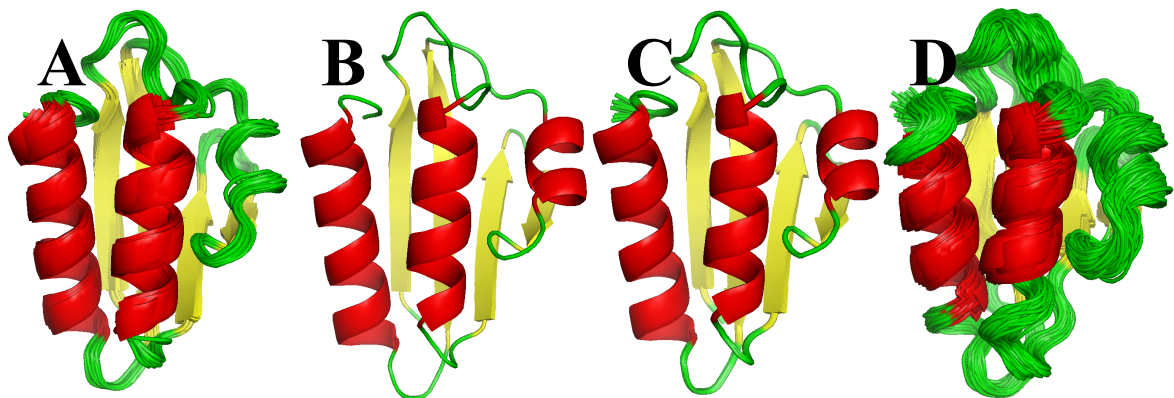


Abbildung 29: **Abbildung der betrachteten 3D-Strukturen von HPr *Staphylococcus*.** **A** Die besten 20 NMR-Strukturen von HPr *S. aureus* (H15A) (Munte, Veröffentlichung in Vorbereitung), **B** Röntgenstruktur von HPr *S. faecalis* (PDB ID: 1PTF), **C** Die besten 20 Modelle von HPr *S. aureus* (H15A) aus der Röntgenstruktur von HPr *S. faecalis*, **D** MD-Ensemble vom besten Modell (200 Strukturen). Der mittlere quadratischen Abstand (RMSD) der schweren Hauptkettenatome der gemittelten Strukturen zum gemittelten NMR-Ensemble beträgt für **B** 0,084 nm, für **C** 0,079 nm und für **D** 0,058 nm. Die Modelle wurden mit PyMOL (DeLano Scientific, LLC) erstellt.

	Top20 NMR-Ensemble von HPr. S. aureus (H15A)	Röntgenstruktur HPr S. faecalis (PDB ID: 1PTF)	Top10 Modelle von HPr. S aureus (H15A) aus der Röntgenstruktur von HPr S. faecalis	MD-Ensemble von HPr S. aureus H15A (200 Strukturen)
RMSD [nm]	0	0,084	0,079	0,058
$\sum_{ij} E_{ij}^{Vol}$	612		700 (+16 %)	626 (+2 %)
$\sum_{ij} E_{ij}^{Shape}$	825		910 (+12 %)	822 (-0,4 %)

**Tabelle 8: Summe der lokalen Pseudoenergien der Rückrechnungen und RMSD-Werte der 3D-Strukturen von HPr Staphylococcus.** Für die RMSD-Werte wurden die schweren Hauptkettenatome der gemittelten 3D-Strukturen und die schweren Hauptkettenatome des gemittelten NMR-Ensembles berücksichtigt. Die Volumen- und Linienformpseudoenergien  $E_{ij}^{Vol}$  und  $E_{ij}^{Shape}$  der Signale  $j$  im Spektrum  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ) wurden mit der vollständigen richtigen Zuordnung von HPr S. aureus (H15A) berechnet. In Klammern sind die relativen Unterschied der Pseudoenergien zu den Pseudoenergien der Rückrechnungen der NMR-Strukturen aufgelistet. Es wurde der Volumenskalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{Ring}$  für das 2-D-NOESY- und  $\bar{\alpha}_{3\text{-D-NOESY}}^{Ref}$  für das 3D  $^{15}\text{N-NOESY-HSQC}$ -Spektrum verwendet.

Als Basis der automatischen Zuordnung der Punktmutante HPr S. aureus (H15A) diente die Röntgenstruktur von HPr S. faecalis (Abb. 29 B), aus der mit AUREMOL PERMOL ein Modell der Punktmutante erzeugt wurde (Abb. 29 C). Das beste Modell diente als Startstruktur einer Molekulardynamik Simulation, aus der ein 200 Strukturen umfassendes Ensemble (Abb. 29 D), sowie Ordnungsparameter und Korrelationszeiten für die Rückrechnungen der beiden Eingabespektren hervorgingen. Tabelle 8 listet die RMSD Werte der Röntgenstruktur, des Modells und des MD-Ensembles zum NMR-Ensemble (Abb. 29 A) auf. Die RMSD-Werte verkleinerten sich mit jeder Weiterverarbeitung der Röntgenstruktur. Außerdem wurden die aufsummierten Werte der lokalen Pseudoenergien des Volumens  $E^{Vol}$  und der Pseudoenergien der Linienform  $E^{Shape}$  auf den Grundflächen der experimentellen Signale (Gl. 45 und Gl. 47 in Kap. 2.2.8) mit der richtigen vollständigen Zuordnungen berechnet. Als Volumenskalierungsfaktoren wurden die optimalen Werte  $\alpha_{2\text{-D-NOESY}}^{Ring}$  für das 2-D-NOESY- und  $\bar{\alpha}_{3\text{-D-NOESY}}^{ref}$  für das 3D  $^{15}\text{N-NOESY-HSQC}$ -Spektrum verwendet. Die Wert der Pseudoenergien für die Rückrechnungen der MD-Simulation waren besser als die Pseudoenergien für die Rückrechnung des Modells. Sie sind nur knapp höher als die Pseudoenergien der Rückrechnung des NMR-Ensembles.

## Automatische Bestimmung der Volumenskalerungsfaktoren und der zusätzlichen Linienverbreiterungen

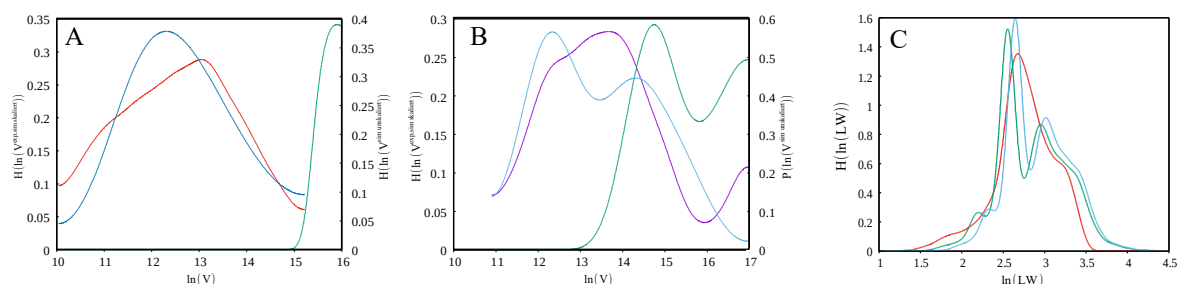


Abbildung 30: Häufigkeitsverteilung der experimentellen und simulierten logarithmierten Volumen  $H(\ln(V))$  und Linienbreiten  $H(\ln(Lw))$  im HPr *S. aureus* (H15A). Die Häufigkeitsverteilungen der experimentellen Werte sind in rot, die der simulierten Werte in grün und die der statistisch optimierten simulierten Werte in blau eingezeichnet. **A** Häufigkeitsverteilungen der Volumina im 2-D-NOESY-Spektrum, **B** Häufigkeitsverteilungen der Volumina im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum und **C** Häufigkeitsverteilungen der Linienbreiten der direkten  $^1\text{H}$ -Dimension des 2-D-NOESY-Spektrums. Die logarithmierten simulierten Volumina in **A** und **B** sind gegen die y2-Achse aufgetragen. Die aus dem MD-Ensemble zurückgerechneten Spektren hatten zufällige Startzuordnungen und wurden für die Bestimmung der Volumenskalerungsfaktoren mit der statistischen Methode verbreitert.

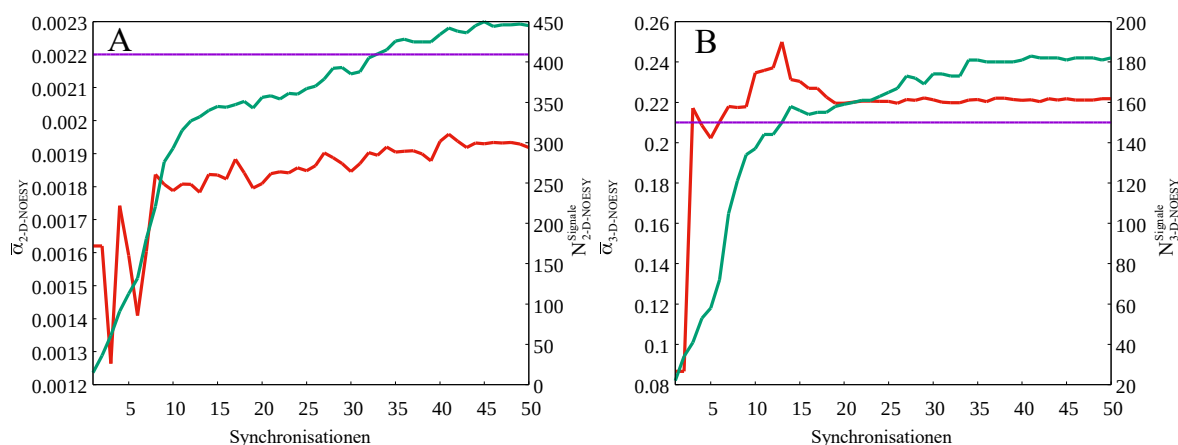


Abbildung 31: Entwicklung der Volumenskalerungsfaktoren  $\bar{\alpha}_i$  (rot) während der automatischen Zuordnung von HPr *S. aureus* (H15A). In grün (y2-Achse) ist die Anzahl der eindeutig zuordenbaren Signale  $N_i^{\text{Signale}}$  aufgetragen. In lila sind die Volumenskalerungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$ , die mit den eindeutig zuordenbaren Kreuzsignalen der vollständigen richtigen Zuordnung bestimmbar waren, eingezeichnet. **A** 2-D-NOESY-Spektrum, **B** 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum.

Für die Rückrechnungen aus dem MD-Ensemble mussten die statistischen Skalierungsfaktoren  $\alpha^{\text{statistisch}}$  (Kap. 2.2.4) und die zusätzlichen Linienverbreiterungen der Spektren (Kap. 2.2.5) bestimmt werden. Es konnten nur für die direkte  $^1\text{H}$ -Dimension des 2-



D-NOESY-Spektrums eine Linienverbreiterung gefunden werden. Wird diese auf die Rückrechnung angewendet, so verbessert sich die Summen der lokalen Pseudoenergien  $E^{\text{Vol}}$  und  $E^{\text{Shape}}$  für die richtige Zuordnung (Tabelle 9).

Abb. 30 C zeigt die Verteilung der logarithmierten Linienbreiten in dieser Dimension. Die Verteilung der verbreiterten simulierten Signale (**blau**) stimmt besser, als die der nicht angepassten Signale (**grün**), mit der Verteilung der experimentellen Linienbreiten (**rot**) überein.

Die statistischen Volumenskalierungsfaktoren  $\alpha_i^{\text{statistisch}}$  der Spektren  $i$  wurden, im Gegensatz zu den Referenzskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$  zu klein bestimmt, was im 2-D-NOESY-Spektrum in einer schlechteren Volumenpseudoenergie der richtigen Zuordnung resultiert (Tabelle 9). Die Pseudoenergie der Volumens der richtigen Zuordnung des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum profitiert dagegen von der Anpassung der simulierten Volumen durch einen statistisch bestimmten Volumenskalierungsfaktor. Abb. 30 A und B zeigen für beide Eingabespektren die Verteilungen der logarithmierten Volumen der experimentellen Signale (**rot**), der simulierten Signale (**grün**) und der angepassten simulierten Signale (**blau**). Die Verteilungen der angepassten simulierten Volumen stimmen gut mit den Verteilungen der experimentellen Volumen überein.

Der statistisch bestimmte Skalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{statistisch}}$  des 2-D-NOESY-Spektrums wurde in der 3. Synchronisationsphase durch einen aus den eindeutig zuordenbaren Signalen gemittelte Skalierungsfaktor ersetzt (Kap. 2.2.14). Nach 7 Synchronisationen hat sich der Skalierungsfaktor (Abb. 31 A **rot**) stabilisiert und tendiert immer besser zum Referenzskalierungsfaktor  $\bar{\alpha}_{2\text{-D-NOESY}}^{\text{Ref}}$  (Tabelle 9 und Abb. 31 A **lila**). Am Ende unterscheidet er sich um 13,6 % von der Referenz. Im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum wurde nach 20 Synchronisationen der richtige Skalierungsfaktor gefunden (Abb. 31 B). Hier betrug der Unterschied zur Referenz am Ende 5,6 %.

	2-D-NOESY	3D <sup>15</sup> N-NOESY-HSQC
$\alpha_i^{\text{statistisch}}$	0,0016 (-51,5 %)	0,087 (-59 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	480 (+15 %)	132 (- 33 %)
Anzahl der Signale	3431	1029
$\bar{\alpha}_i^{\text{Ref}}$	0,0022 (- 33,3 %)	0,21
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	401 (-3,6 %)	196
Anzahl der Signale	2114	479
$\alpha_i^{\text{Ring}}$	0,0033	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	416	
Anzahl der Signale	79	
Verbreiterung $\omega 1$ [Hz]	0	0
Verbreiterung $\omega 2$ [Hz]	1,2	0
Verbreiterung $\omega 3$ [Hz]		0
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ ohne Verbreiterung	430	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ mit Verbreiterung	400 (-7 %)	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ ohne Verbreiterung	557	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ mit Verbreiterung	544 (-2,3 %)	

**Tabelle 9: Auf verschiedene Arten bestimmte Volumenskalierungsfaktoren und statistisch bestimmte Linienverbreiterungen und ihre Auswirkung auf die lokalen Pseudoenergien der richtigen Zuordnung von HPr S. aureus (H15A).** Es sind die statistisch bestimmten ( $\alpha_i^{\text{statistisch}}$ ), die mit allen eindeutig zuordenbaren Signalen bestimmten ( $\bar{\alpha}_i^{\text{Ref}}$ ) und nur mit eindeutigen Ringkontakten bestimmten Skalierungsfaktoren ( $\alpha_i^{\text{Ring}}$ ) angegeben. Außerdem sind die Pseudoenergie des Volumens und die Anzahl der experimentellen Signale, die bei der Bildung des Skalierungsfaktor beteiligt waren, aufgelistet.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  bzw.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$  sind die über alle experimentelle Signale j summierten Pseudoenergien des Volumens bzw. der Linienform im Spektrum i (i = 2-D-NOESY bzw. 3D <sup>15</sup>N-NOESY-HSQC).  $\omega 1$ ,  $\omega 2$  und  $\omega 3$  sind die Dimensionen der Spektren, in denen die Signale der Rückrechnung verbreitert wurden. In Klammern sind die relativen Unterschiede der Skalierungsfaktoren und Pseudoenergien zu der Referenz bzw. der nicht verbreiterten Rückrechnung angegeben.

## Häufigkeitsverteilungen der Pseudoenergien der richtigen Zuordnung

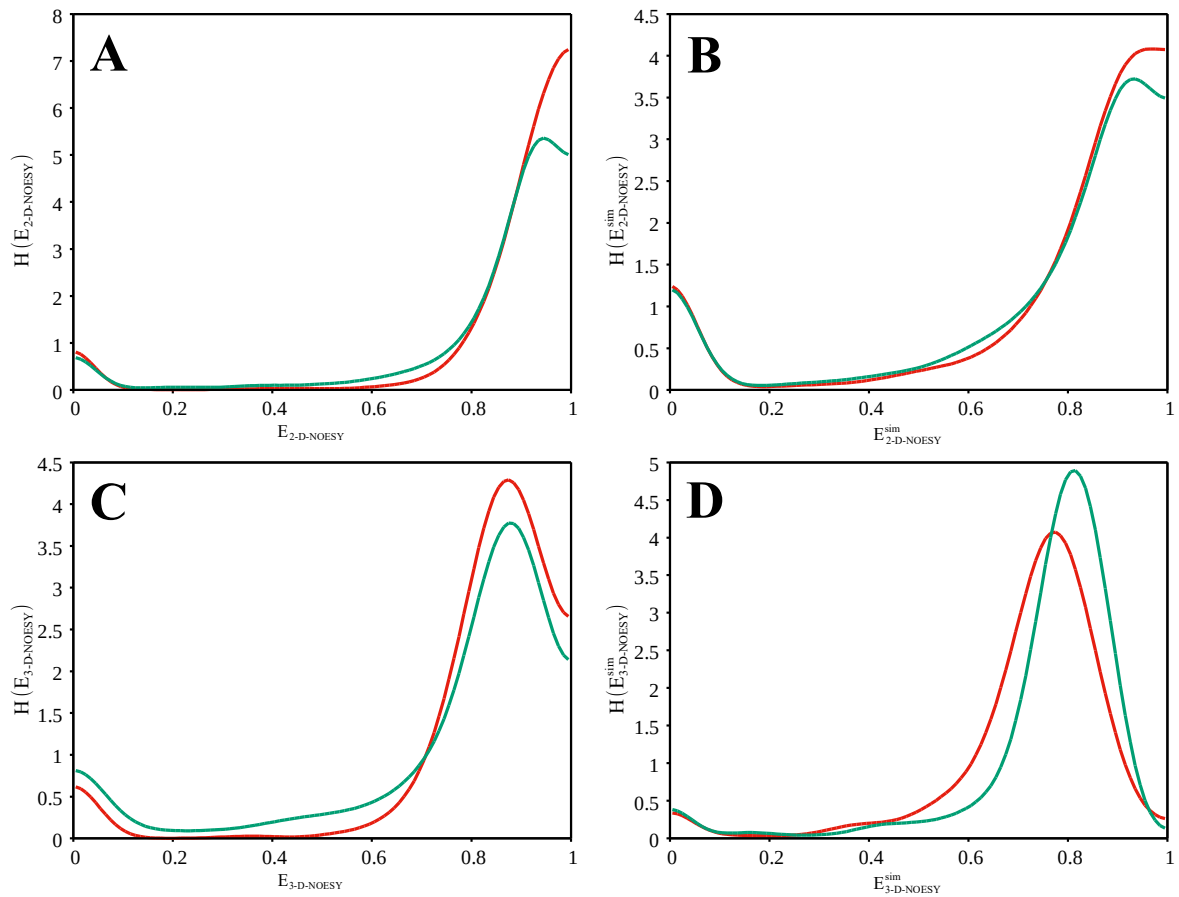


Abbildung 32: Häufigkeitsverteilungen  $H(E)$  der lokalen Pseudoenergien des Volumens  $E^{Vol}$  (rot) und der Linienform  $E^{Shape}$  (grün) der richtigen vollständigen Zuordnung von HPr *S. aureus* (H15A). **A** Auf den Grundflächen der experimentellen Signale im 2-D-NOESY-Spektrum, **B** auf den Grundflächen der simulierten Signale im 2-D-NOESY-Spektrum, **C** auf den Grundflächen der experimentellen Signale im 3D  $^{15}N$ -NOESY-HSQC-Spektrum, **D** auf den Grundflächen der simulierten Signale im 3D  $^{15}N$ -NOESY-HSQC-Spektrum. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und mit einer zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalisierungsfaktor  $\alpha_{2-D-NOESY}^{Ring}$ , für das 3D  $^{15}N$ -NOESY-HSQC-Spektrum der Skalisierungsfaktor  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.

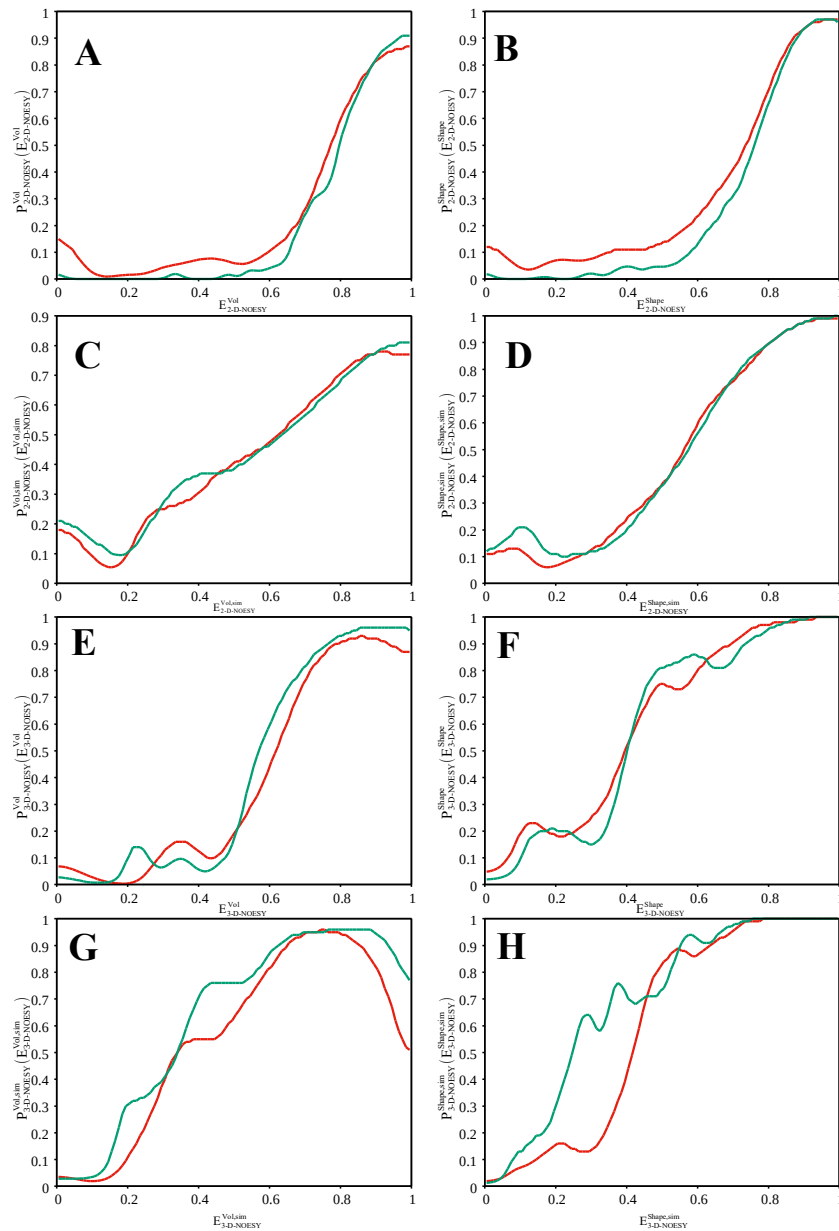


Abbildung 33: **Abhängigkeit der Bayesschen Wahrscheinlichkeit der lokalen richtigen Zuordnung  $P_i(E)$  von den lokalen Pseudoenergien  $E_i$  (rot) für die NOESY-Spektren des HPr *S. aureus* (H15A).** In grün sind die zugehörigen Wahrscheinlichkeitsfunktionen, die SIBASA während der automatischen Zuordnung verwendet (Kap. 4.1.3), abgebildet. **A**  $P_{2-D-NOESY}^{Vol}(E_{2-D-NOESY}^{Vol})$ , **B**  $P_{2-D-NOESY}^{Shape}(E_{2-D-NOESY}^{Shape})$ , **C**  $P_{2-D-NOESY}^{Vol,sim}(E_{2-D-NOESY}^{Vol,sim})$  und **D**  $P_{2-D-NOESY}^{Shape,sim}(E_{2-D-NOESY}^{Shape,sim})$  des 2-D-NOESY-Spektrums. **E**  $P_{3-D-NOESY}^{Vol}(E_{3-D-NOESY}^{Vol})$ , **F**  $P_{3-D-NOESY}^{Shape}(E_{3-D-NOESY}^{Shape})$ , **G**  $P_{3-D-NOESY}^{Vol,sim}(E_{3-D-NOESY}^{Vol,sim})$  und **H**  $P_{3-D-NOESY}^{Shape,sim}(E_{3-D-NOESY}^{Shape,sim})$  des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und einer mit zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalierungsfaktor  $\alpha_{2-D-NOESY}^{Ring}$ , für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum der Skalierungsfaktor  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.

In Abb. 32 sind die Häufigkeitsverteilungen der lokalen Pseudoenergien des Volumens  $E^{\text{Vol}}$  und der Linienform  $E^{\text{Shape}}$  der richtigen Zuordnung für die beiden NOESY-Spektren aufgetragen. Wie erwartet, sind hohe Werte der Pseudoenergien in der vollständigen richtigen Zuordnung viel häufiger vertreten als kleine oder mittlere Werte. In Abb. 33 sind die mit dem Satz von Bayes berechneten Wahrscheinlichkeiten in Abhängigkeit von den lokalen Pseudoenergien aufgetragen (**rot**). Eine lokale Wahrscheinlichkeit  $P_{ij}^{\text{Vol,Shape}}$  auf der Grundfläche  $ij$  (Gl. 50) wird immer aus der Kombination der Volumen- und der Linienformpseudoenergie gebildet. Man kann aus der Abbildung erkennen, dass alle lokalen Wahrscheinlichkeiten der richtigen Zuordnung mit den korrespondierenden Pseudoenergien steigen. In **grün** sind die entsprechende Bayesschen Wahrscheinlichkeitskurven der AUREMOL Datenbank (Kap. 4.1.3), die für die Berechnung der lokalen Wahrscheinlichkeiten bei unbekannter Zuordnung verwendet werden, abgebildet. Für die automatische Zuordnung von HPr *S. aureus* (H15A) aus der Röntgenstruktur von HPr *S. faecalis* stimmen die Datenbankverteilungen mit den realen Verteilungen, vor allem in den relevanten Bereichen hoher lokaler Pseudoenergien der richtige Zuordnung (Abb. 32), sehr gut überein. Für niedrige Pseudoenergien sind die realen Wahrscheinlichkeiten höher als die Wahrscheinlichkeiten die SIBASA verwendet, da alle experimentellen Signale in die zugrundeliegenden Häufigkeitsverteilungen eingegangen sind. Für die Häufigkeitsverteilungen der lokalen Pseudoenergien der AUREMOL Datenbank war eine vorherige Filterung notwendig um den dynamischen Bereich der von SIBASA berechneten lokalen Wahrscheinlichkeiten zu vergrößern.

## Verlauf von Pseudoenergien und dynamischer Größen während der automatischen Zuordnung

Es wurden 16 unabhängige SIBASA Instanzen mit verschiedenen Startkonfigurationen und Zufallszahlen gestartet. Jede dieser Instanzen optimierte die Pseudoenergie  $E^{\text{total}}$  (Gl. 44) um die richtige vollständige Zuordnung zu finden. Abb. 34 zeigt den Verlauf des wichtigsten Teils von  $E^{\text{total}}$ , nämlich die gesamte Pseudoenergie des Volumen- und Linienformvergleichs  $E^{\text{Vol,Shape}}$  (Gl. 58) für zwei Instanzen. Die Instanz mit dem **roten** Verlauf optimiert die Pseudoenergie deutlich besser, als die Instanz mit dem **grünen** Verlauf, was sich auch in der Anzahl der richtig gefunden Zuordnungen (Abb. 34 **B**) niederschlägt. Für die Auswertung wurde eine Zuordnung als richtig angesehen, wenn sich die chemische Verschiebung ihrer Resonanz sich maximal 0,03 ppm, was ungefähr einer Linienbreite eines NOESY-Signals entspricht, vom richtigen Wert unterschieden hat.

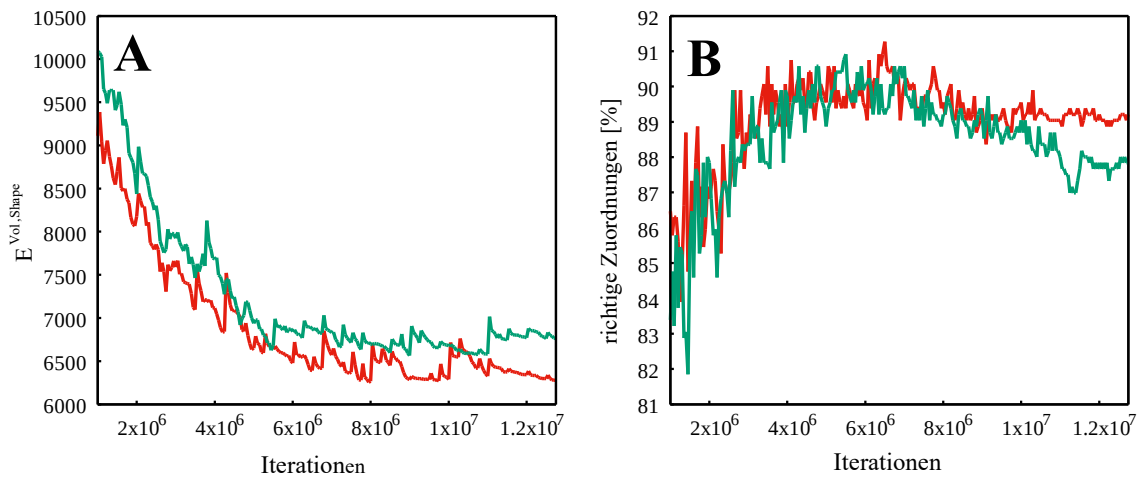


Abbildung 34: Zeitlicher Verlauf der gesamten Pseudoenergie des Signalvergleichs und des Anteil der richtig identifizierten Zuordnungen bei der automatischen Zuordnung von HPr *S. aureus* (H15A). In rot ist die Instanz der Optimierung mit der besten und in grün die Instanz mit der schlechtesten finalen Pseudoenergie eingezeichnet. **A** Verlauf der Pseudoenergie des Volumen- und Linienformvergleichs auf den Grundflächen der experimentellen Signale  $E^{\text{Vol,Shape}}$ , **B** Verlauf der gefundenen richtigen Zuordnungen im Fehlerbereich  $\delta^{\text{Ref}} \pm 0,03 \text{ ppm}$ .

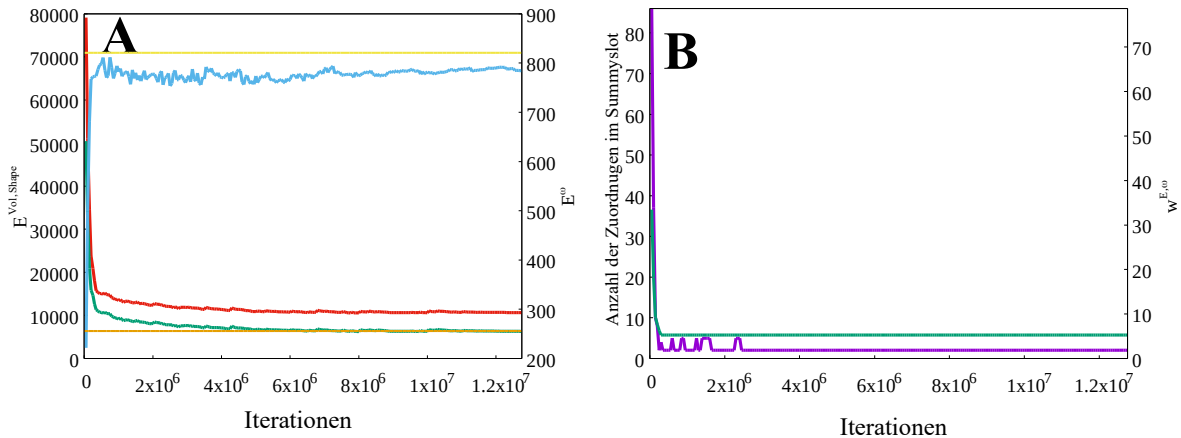


Abbildung 35: Zeitliche Verläufe wichtiger Optimierungsgrößen einer Instanz der automatischen Zuordnung von HPr *S. aureus* (H15A). **A** Verlauf der gesamten Pseudoenergie der Optimierung  $E^{\text{total}}$  (rot, y-Achse), Verlauf der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  (grün, y-Achse) und Verlauf der Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  (blau, y2-Achse). In orange (y-Achse) ist der geschätzte optimale Wert der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape,opt}}$  und in gelb (y2-Achse) die Pseudoenergie der chemischen Verschiebung der richtigen vollständigen Zuordnung eingetragen.

**B** Anzahl der Mitglieder von Amidgruppen im Dummyslot (lila, y-Achse) und Verlauf des dynamischen Gewichtungsfaktors der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}$  (grün, y2-Achse), dessen Trend auch auf den Gewichtungsfaktor der Pseudoenergie des Dummyslots  $w^{E,\text{Dummy}}$  übertragen werden kann.

Der Vergleich der besten und schlechtesten Instanz zeigt, dass sich wie bei der Strukturrechnung, das starten mehrere unabhängiger Instanzen mit unterschiedlichen

Startkonfigurationen und Zufallszahlen lohnt um das Optimum der Pseudoenergie und damit die richtige vollständige Zuordnung zu finden. Insgesamt fanden die 16 Instanzen jeweils zwischen 87,7 % und 90,1 % der Zuordnungen richtig. In Abb. 35 sind der Verlauf wichtiger Pseudoenergien und Gewichtungsfaktoren, der besten Instanz, also der Instanz die die Pseudoenergie der Optimierung  $E^{\text{total}}$  (Abb. 34 **A rot**) am besten minimiert hat, abgebildet. Die Abschätzung der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape,opt}}$  (Abb. 35 **A orange**) (Gl. 65), die für die Berechnung der dynamischen Gewichtungsfaktoren benötigt wird, wird von  $E^{\text{Vol,Shape}}$  am Ende der Minimierung (Abb. 35 **A grün**) gut erreicht. Die Abweichung lag bei knapp über 1 % ( $E^{\text{Vol,Shape,opt}} = 6404$  bzw.  $E^{\text{Vol,Shape}}(\text{Ende}) = 6392$ ).

Die Pseudoenergie der chemischen Verschiebung der richtigen Zuordnung  $E^{\omega}$  (Abb. 35 **A gelb**) wird in der automatischen Zuordnung durch die dynamische Gewichtung mit dem Faktor  $w^{E,\omega}(t)$  (Gl. 64) fast erreicht. Die Abweichung am Ende beträgt 5 % ( $E^{\omega}(\text{Referenz}) = 824$  bzw.  $E^{\omega}(\text{Ende}) = 785$ ). Durch die schnelle Minimierung der Pseudoenergie  $E^{\text{total}}$  in der Anfangsphase (Abb. 35 **A rot**) konvergieren der dynamische Skalierungsfaktor der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}(t)$  (Abb. 35 **B grün**) und der dynamische Skalierungsfaktor der Pseudoenergie des Dummyslots  $w^{E,\text{Dummy}}$  (Gl. 74) schnell gegen ihre minimalen Grenzwerte. Die Anzahl der Amidprotonen und Stickstoffe im Dummyslot (Abb. 35 **B lila**), die keine sichtbaren Signale im finalen zurückgerechneten Spektrum haben, geht auf zwei zurück. Nur die beiden außerhalb des spektralen Bereichs der Eingabespektren liegenden Zuordnungen von Arg17 H<sup>e</sup> und Arg17 N<sup>e</sup> sind auf diesem Slot geblieben.

## Festlegung von Zuordnungen während der Laufzeit

Nach jeweils fünfmaligen Abkühlen des Thresholds  $Th$  (Kap. 2.2.10) ist SIBASA in eine Synchronisationsphase, in der die chemischen Verschiebungen der Resonanzen  $\delta_{k,j}$  und die Pseudoenergien der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  aller parallel gestarteten Instanzen  $j$  ausgewertet wurden (Kap. 2.2.13), gegangen. Es wurden die sogenannten besten Cluster der Kerne  $k$  gebildet, in denen die gewichtete Standardabweichung der chemischen Verschiebung  $\sigma_k^w(\text{best})$  (Gl. 82) nicht größer als 0,03 ppm sein durfte.

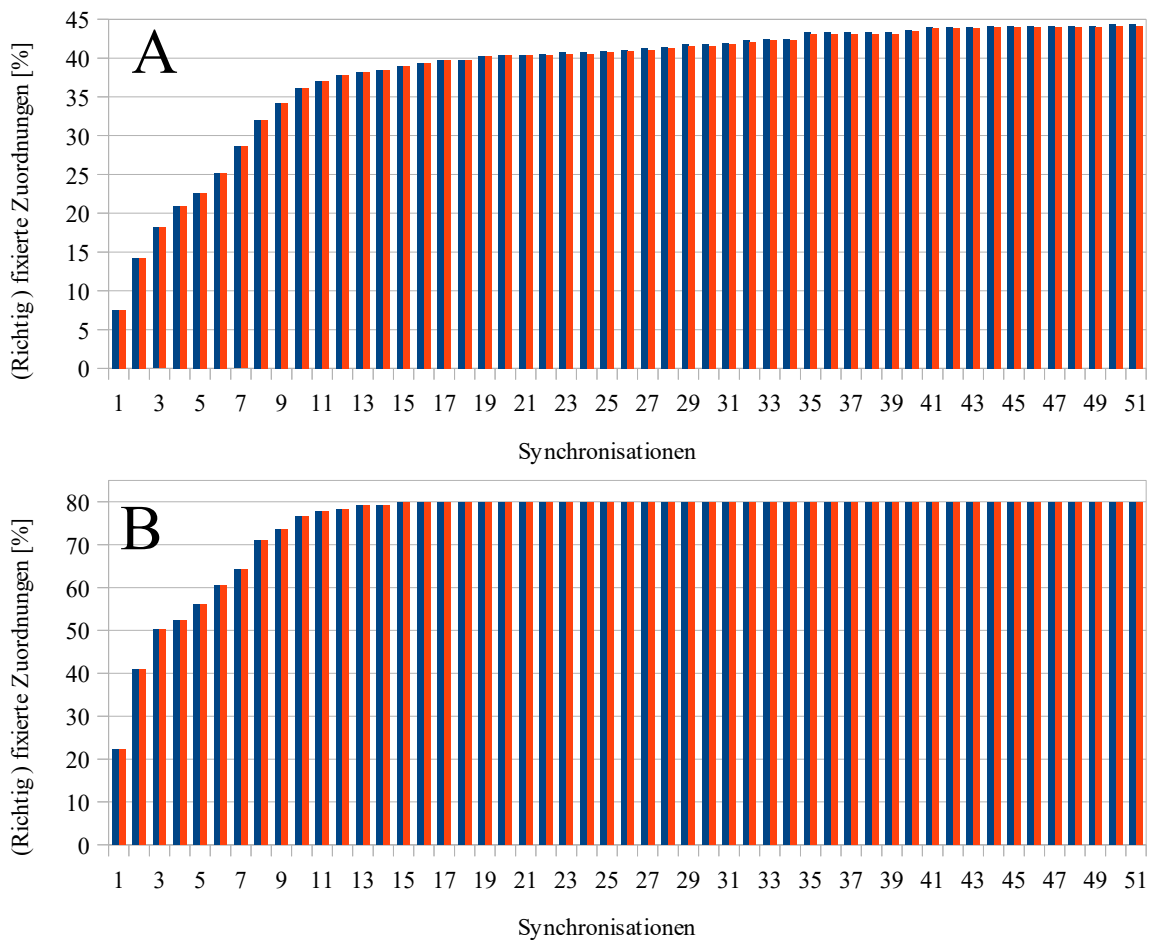


Abbildung 36: Anteil der zur Laufzeit fixierte Zuordnungen (blau) und Anteil der zur Laufzeit richtig fixierten Zuordnungen (rot) für die automatische Zuordnung von HPr *S. aureus* (H15A). **A** Alle Zuordnungen von Resonanzen, **B** Nur Zuordnungen von Resonanzen von Amidprotonen und Stickstoffen.

Die erste Bedingung für eine Fixierung war, dass der Wert Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  (Gl. 81) des Kerns  $k$  größer als sein Grenzwert  $p = 0,85$  bzw. die Verifikationspseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  (Gl. 85) für ein Mitglied einer Amidgruppe größer als der Grenzwert  $p_{\text{Amidgruppe}} = 0,73$  sein musste (Gl. 87 und Gl. 88). Außerdem muss für alle Instanzen  $j$  des besten Clusters der Wert der Pseudoenergie der Zuordnung  $E_{k,j}^{\text{Vol,Shape,sim}}$  (Gl. 70) größer als der Grenzwert  $q = 0,85$  bzw. für Mitglieder von Amidgruppen die Pseudoenergie  $E_{H^k,j}^{\text{Vol,Shape,sim}}(m) E_{N^k,j}^{\text{Vol,Shape,sim}}$  größer als der Grenzwert  $q_{\text{Amidgruppe}} = 0,73$  sein. Die Grenzwerte  $p$ ,  $q$ ,  $p_{\text{Amidgruppe}}$ ,  $q$  und  $q_{\text{Amidgruppe}}$  wurden durch Auswertung von Datensätzen bereits gelöster Proteine (Kap. 4.1.3) festgelegt.

Die **blauen** Balken in Abb. 36 zeigen Anteil der zur Laufzeit fixierten Resonanzen (**A**) bzw. den Anteil der zur Laufzeit fixierten Resonanzen von Amidgruppen (**B**). Die **rote** Balken in der Abbildung zeigen den Anteil der richtig fixierten Zuordnungen. Zur Laufzeit wurden



knapp 45 % der Resonanzen von HPr S. aureus (H15A) fixiert. Von Ihnen wurde nur eine einzige falsch zugeordnet. Bei den Mitgliedern von Amidgruppen wurden knapp 80 % der Resonanzen festgelegt. Von Ihnen wurde keine falsch zugeordnet. Man kann erkennen, dass die Anzahl der festgelegten Resonanzen am Anfang schnell ansteigt und am Ende sättigt. Bei den Amidgruppen geschieht dies schneller als bei den H<sup>α</sup>- und Seitenkettenprotonen.

Durch das Fixieren von Zuordnungen wird nicht nur in erheblichen Maß die Rechenzeit auf die noch zu bestimmenden chemischen Verschiebungen von Resonanzen umverteilt, sondern auch die Skalierungsfaktoren  $\bar{\alpha}_i$  bestimmt (Kap. 2.2.14). In Abb. 31 sind für die beiden Eingabespektren die Anzahl der eindeutig zuordenbaren Kreuzsignale (**grün**) in Abhängigkeit vom Zeitpunkt in der automatischen Zuordnung angegeben. Für die Bestimmung der Volumenskalierungsfaktoren standen am Ende 452 2-D-NOESY-Signale (insgesamt 12,6 %) und 182 3D <sup>15</sup>N-NOESY-HSQC-Signale (insgesamt 16,8 %) zur Verfügung.

## Auswertung und Verifikation der vereinigten Zuordnungsliste

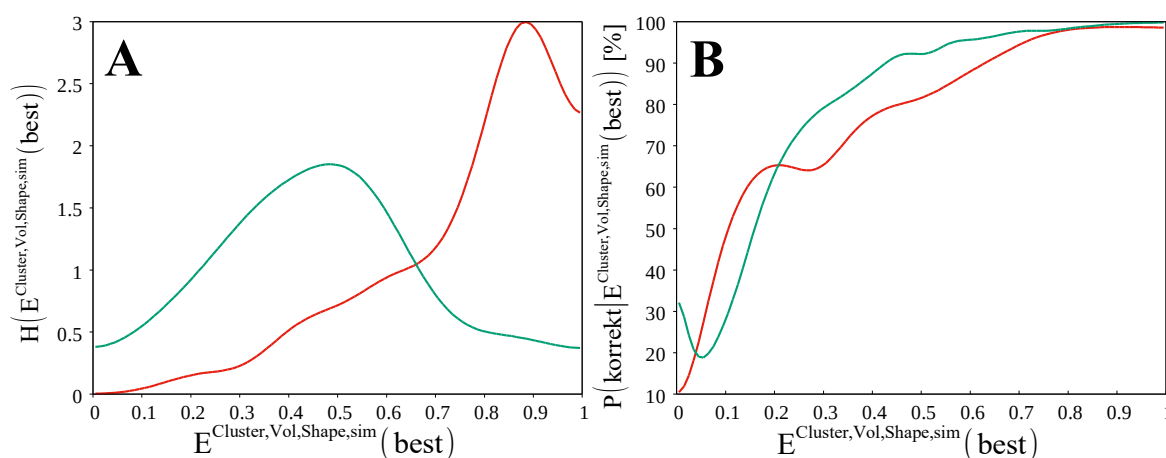


Abbildung 37: **Häufigkeitsverteilungen der Verifikationspseudoenergie  $E^{Cluster, Vol, Shape, sim}(best)$  und Wahrscheinlichkeit der richtigen Zuordnung nach der automatischen Zuordnung von HPr S. aureus (H15A).** **A** Häufigkeitsverteilung  $H(E^{Cluster, Vol, Shape, sim}(best))$  der Verifikationspseudoenergien der richtigen (**rot**) und der falschen Zuordnungen (**grün**). **B** Bayessche Wahrscheinlichkeit einer richtigen Zuordnung  $P(korrekt|E^{Cluster, Vol, Shape, sim}(best))$  (**rot**) in Abhängigkeit von der Verifikationspseudoenergie und zugehörige Wahrscheinlichkeitskurve, die mit Hilfe von bereits gelösten Proteinen erzeugt und in die AUREMOL Datenbank gespeichert wurde (**grün**) (Abb. 28).

Das Ergebnis von SIBASA wurde nach der 51. Synchronisation ausgewertet. Die Pseudoenergie des Signalvergleichs  $E^{Vol, Shape}$  hat zu diesem Zeitpunkt ihr Minimum erreicht und es wurden keine weiteren Zuordnungen mehr zur Laufzeit fixiert. Nach dieser letzten Synchronisation waren in der vereinigten Zuordnungsliste (Kap. 2.2.12), die mit einer

gewichteten Standardabweichung  $\sigma_k^w(\text{best})$  von 0,03 ppm erstellt wurde, insgesamt 539 der 586 chemischen Verschiebungen von Resonanzen im Fehlerbereich von 0,03 ppm richtig erkannt worden (91,3 %). Alle Resonanzen der 193 Amidprotonen und Stickstoffen sind richtig zugeordnet worden (100 %). In ASSIGN, dass nur ein 2-D-NOESY-Spektrum und eine NMR-Struktur als Eingabe verwendet hatte, sind beim gleichen Protein nur 31 % aller Resonanzen richtig zugeordnet worden.

### Häufigkeitsverteilung der Verifikationspseudoenergien und Wahrscheinlichkeit der richtigen Zuordnung

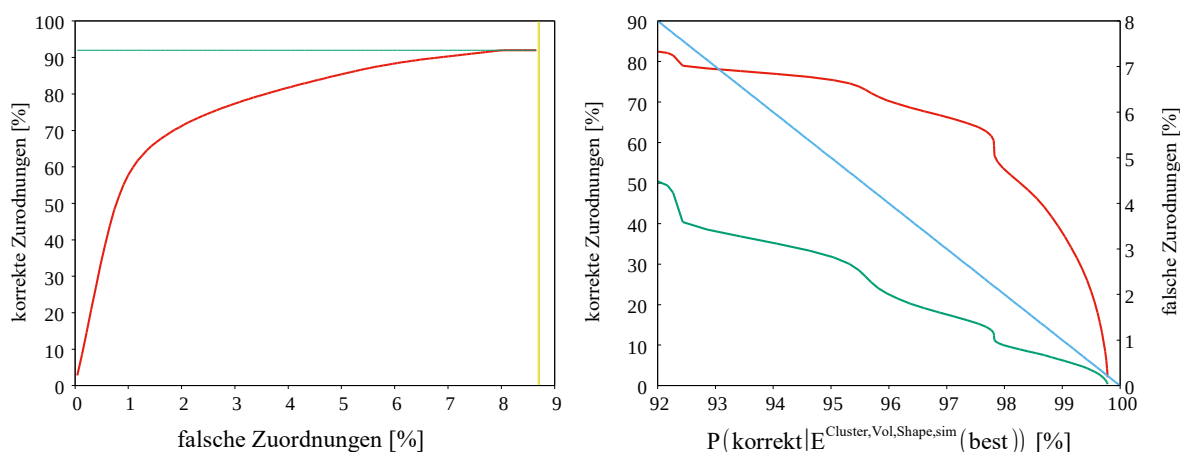


Abbildung 38: **Auswirkung eines Grenzwertes der Verifikationspseudoenergie  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  auf die Klassifizierung der Resonanzzuordnungen in der vereinten Liste der chemischen Verschiebung von HPr *S. aureus* (H15A).** **A** Modifizierte Receiver-Operator-Characteristics (ROC) Kurve der Verifikationspseudoenergie Es sind die möglichen richtigen Zuordnungen gegen die dafür in Kauf zu nehmenden falschen Zuordnungen aufgetragen (**rot**). Insgesamt sind 91,3 % der Resonanzen richtig (**grün**) und 8,7 % falsch zugeordnet worden (**gelb**).

**B** Gefundene richtige (**rot**, y-Achse) und falsche Zuordnungen (**grün**, y2-Achse) in Abhängigkeit von der Sicherheitswahrscheinlichkeit  $P(\text{korrekt}|E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$ . Der vom Benutzer tolerierten Anteil von falschen Zuordnungen ist in **blau** (Winkelhalbierende, y2-Achse) eingetragen.

In Abb. 37 **A** sind die Häufigkeitsverteilungen der Verifikationspseudoenergien  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  nach der 51. Synchronisation für falsche (**grün**) und für richtige (**rot**) Zuordnungen angegeben. Wendet man den Satz von Bayes an (Gl. 89), so erhält man die Wahrscheinlichkeit der richtigen Zuordnung in Abhängigkeit von der Verifikationspseudoenergie. In Abb. 37 **B** ist diese in **rot** zusammen mit dem Graph der Wahrscheinlichkeit einer richtigen Zuordnung, die in Kap. 4.1.3 für die Auswertung der Ergebnisse der automatischen Zuordnung von bereits gelösten Proteinen aus NMR-Daten erstellt worden ist (**grün**) (Abb. 28), aufgetragen. Wie schon die lokalen Wahrscheinlichkeiten der richtigen Zuordnung (Abb. 32), zeigen auch die mit unterschiedlichen Datensätzen und

dem Satz von Bayes gebildeten Wahrscheinlichkeitskurven, vor allem im wichtigen Bereich hoher Pseudoenergien, sehr ähnliches Verhalten. Noch deutlicher wird die Übertragbarkeit der Wahrscheinlichkeit-Verifikationspseudoenergie-Abhängigkeit zwischen den beiden Datensätzen in Abb. 38 **B** gezeigt. Die gefunden richtigen Zuordnungen (**rot**, y-Achse) und die gefunden falschen Zuordnungen (**grün**, y2-Achse) sind gegen den vom Benutzer, auf Basis der in der AUREMOL Datenbank gespeicherten Wahrscheinlichkeitsgraphen (Abb. 28) wählbaren Wahrscheinlichkeitsthreshold aufgetragen. Der Anteil der gefunden falschen Zuordnungen liegt immer unter der Winkelhalbierende **blauen** Linie (y2-Achse), die den vom Benutzer in Kauf genommenen Anteil der falschen Zuordnungen beschreibt.

Für die in Kap. 4.1.3 festgelegten minimalen Grenzwerte der Verifikationspseudoenergien  $p = 0,85$  bzw.  $p_{\text{Amidgruppe}} = 0,73$  ist die reale Wahrscheinlichkeit einer falschen Zuordnung für  $\text{H}^\alpha$ - und Seitenkettenprotonen 1,4 % und 0 % für Amidgruppen.

### Eignung der Verifikationspseudoenergie als Separator zwischen falschen und richtigen Zuordnungen

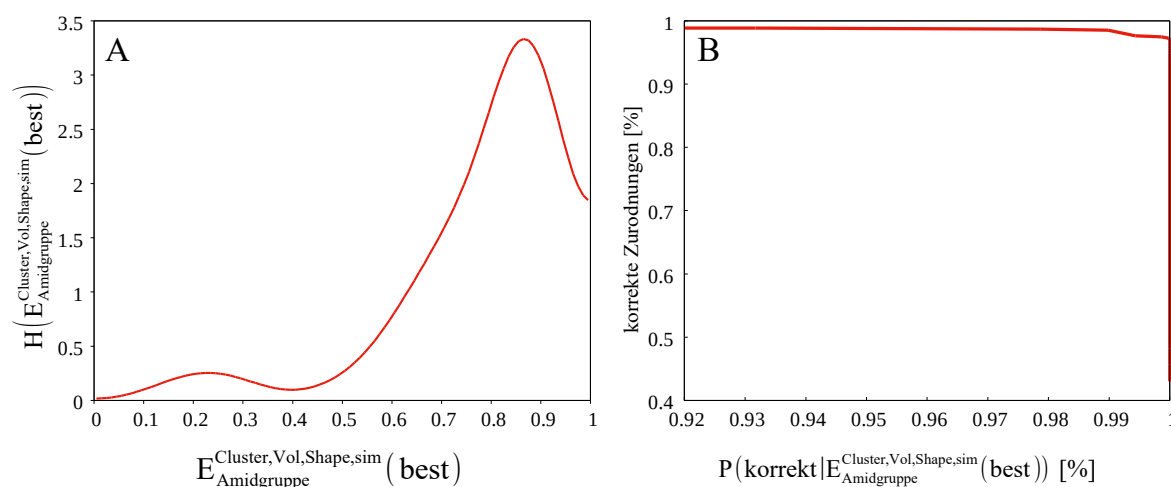


Abbildung 39: **Häufigkeitsverteilung der Verifikationspseudoenergie  $E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  und Auswirkung eines Wahrscheinlichkeitsthresholds auf den Anteil der verbleibenden richtigen Zuordnungen von Amidprotonen und Stickstoffen nach der automatischen Zuordnung von HPr S. aureus (H15A).** In der vereinigten Zuordnungsliste waren alle chemischen Verschiebungen der Resonanzen von Amidprotonen und Stickstoffen richtig zugeordnet. **A** Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der Verifikationspseudoenergien der Mitglieder von Amidgruppen. **B** Gefundene richtige Zuordnungen von Amidprotonen- und Stickstoffresonanzen in Abhängigkeit von der vom Benutzer gewählten Sicherheitswahrscheinlichkeit  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$ .

In Abb. 38 **A** ist eine modifizierte Version der **Receiver-Operator-Characteristics** Kurve (ROC) eingezeichnet. Die unmodifizierte ROC-Kurve trägt den Anteil der richtig bestätigten Zuordnungen (Sensitivität) gegen den Anteil der falsch bestätigten Zuordnungen auf. Die

Qualität der Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  als Entscheidungskriterium kann mithilfe der Fläche unter der unmodifizierten ROC-Kurve quantifiziert werden. Sie beträgt für alle automatisch von SIBASA gefunden Zuordnungen 83 % vom Maximalwert. Da keine Amidgruppen falsch zugeordnet wurden, beträgt die Fläche hier definitionsgemäß 100 %. Die modifizierte ROC-Kurve berücksichtigt nicht nur die Verifikationspseudoenergie, sondern auch die Qualität des automatischen Zuordnungsalgorithmus. Sie trägt den Anteil aller richtig gefunden Zuordnungen gegen den Anteil der in Kauf zu nehmenden falschen Zuordnungen auf. Die **grüne** und die **gelbe** Linie geben die jeweiligen maximal möglichen Anteile an. Nimmt man nach der automatischen Zuordnung von HPr S. aureus (H15A) z.B. 5 % falsche Zuordnungen in Kauf (signifikantes Niveau), so erhält man 75 % aller Zuordnungen richtig und 2,8 % aller Zuordnungen falsch.

Da nach der letzten Synchronisation für HPr S. aureus (H15A) keine falschen Zuordnungen von Amidgruppen mehr gibt, kann weder der Satz von Bayes verwendet, noch ROC-Kurven gebildet werden. Abb. 39 A zeigt die Häufigkeitsverteilung der Verifikationspseudoenergien der Amidgruppen  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$ . Es zeigt sich wieder, dass richtige Zuordnungen meistens hohe Verifikationspseudoenergien haben. Abb. 39 B trägt den Anteil der richtig gefunden Zuordnungen von Amidgruppen gegen das vom Benutzer gewählte Wahrscheinlichkeitsniveau auf. Will der Benutzer nur absolut sichere Zuordnungen bestätigen, so wählt er das 99,8 % Sicherheits-Niveau (weniger als eine falsche Zuordnung). Mit diesem Niveau wurden nur acht richtige Zuordnungen von Seitenkettenamiden und Seitenkettenstickstoffen (4,1 %) herausgefiltert.

### Abschätzung der Wahrscheinlichkeiten der richtigen Zuordnung von HSQC-Signalen

Eine weitere Methode die Wahrscheinlichkeiten der richtigen Zuordnung von Amidprotonen und Stickstoffen  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  zu bestimmen wurde bereits in Donaubauer (Veröffentlichung in Vorbereitung) im Rahmen des automatischen PEAKPICKING verwendet. Aus den Häufigkeitsverteilungen der Verifikationspseudoenergien  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der falschen Zuordnungen, die für die Berechnung der Wahrscheinlichkeiten der richtigen Zuordnung aus den NMR-Datensätzen bereits gelöster Proteinen erstellt worden ist, ist bekannt, dass kaum falsche Zuordnungen von Mitgliedern von Amidgruppen Verifikationspseudoenergien zwischen 0,73 und 1 haben (Abb. 28 **grün**).  $p_{\text{Amidgruppe}} = 0,73$  ist deshalb auch der minimale Grenzwert für die Festlegung von chemischen Verschiebungen von Amidprotonen- und Stickstoffresonanzen während der automatischen Zuordnung.

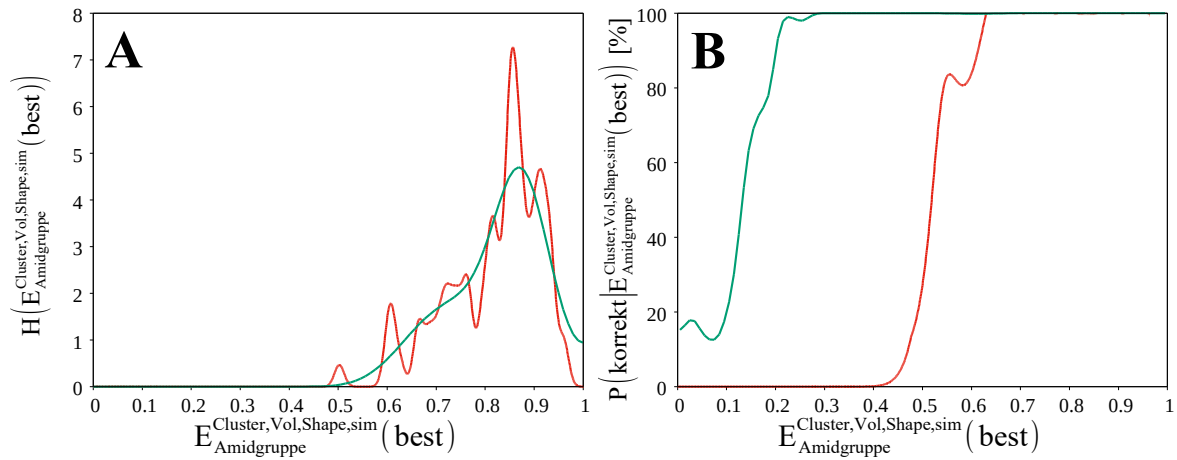


Abbildung 40: **Schätzung der Häufigkeitsverteilung der Verifikationspseudoenergien der richtig zugeordneten Resonanzen von Hauptkettenamidgruppen und resultierende Wahrscheinlichkeit der richtigen Zuordnung beim HPr *S. aureus* (H15A).** **A** Wenig geglättete (rot) und mit  $H^{\text{korrekt}}(x)$  (Gl. 92) gefittete Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (grün) der Verifikationspseudoenergien der Zuordnungen von Hauptkettenamidgruppen. **B** Mithilfe des Fits geschätzte Bayessche Wahrscheinlichkeit der richtigen Zuordnung eines Amidprotons oder Stickstoffs  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  in Abhängigkeit von der Verifikationspseudoenergie (rot). In grün ist der Wahrscheinlichkeitsgraph, der mit Hilfe von Datensätzen bereits gelöster Proteine bestimmt wurde, eingezeichnet.

Man kann also annehmen, dass der Bereich hoher Pseudoenergien in der Häufigkeitsverteilung der Verifikationspseudoenergien der Zuordnungen von Amidprotonen und Stickstoffen  $H(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  ausschließlich von richtigen Zuordnungen herrührt. Wenn man zusätzlich davon ausgeht, dass die Häufigkeitsverteilungen der Verifikationspseudoenergien der richtigen Zuordnungen von Amidprotonen und Stickstoffen  $H^{\text{korrekt}}(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  mit einer Superposition  $H^{\text{korrekt}}(x)$  einer Logarithmischen Normalverteilung und einer Normalverteilung (Gl. 92) gefittet werden können, müssen nur noch die Grenzen des Fits in der Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  festgelegt werden, um die Verteilung  $H^{\text{korrekt}}(E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  im gesamten Bereich der Verifikationspseudoenergien zwischen 0 und 1 zu erhalten.

$$H^{\text{korrekt}}(x) = \frac{A}{\sqrt{2\pi}\sigma_1 x} \exp\left(-\frac{1}{2}\left(\frac{\ln(x) - \mu_1}{\sigma_1}\right)^2\right) + \frac{F - A}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x - \mu_2}{\sigma_2}\right)^2\right) \quad (92)$$

$H^{\text{korrekt}}(x)$  ist auf den ebenfalls unbekanntem Anteil der richtigen Zuordnungen  $F$  normiert. Es muss also gelten, dass  $F \leq 1$  und  $A \in [0; F]$  sein muss.

Die wenig geglättete Häufigkeitsverteilung der Verifikationspseudoenergien der in der Regel besser zuordenbaren Hauptkettenamidgruppen nach der automatische Zuordnung von HPr *S.*

aureus (H15A)  $H(E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (Abb. 40 **A rot**) fällt bei einem Wert von 0,57 auf nahezu 0 ab. Dieser Wert wird als linke Grenze des Fits der verbreiterten Verteilung  $H(E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  an  $H^{\text{korrekt}}(x)$  (Gl. 92) verwendet. In **grün** ist das Ergebnis des Fits eingetragen.

Die Wahrscheinlichkeit  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  kann dann folgendermaßen bestimmt werden:

$$P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best})) = \frac{H^{\text{korrekt}}(E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))}{H(E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))} \quad (93)$$

Abb. 40 **B** zeigt den mit dem Fit bestimmten Graph der Wahrscheinlichkeit für richtige Zuordnungen (**rot**) in Abhängigkeit von der Verifikationspseudoenergie. Der minimale Grenzwert  $p_{\text{Amidgruppe}}$  für die Verifikationspseudoenergie von Amidgruppen für sichere Zuordnungen kann mit diesem Graph auf 0,63 bestimmt werden. Der Wahrscheinlichkeitsgraph, der mithilfe der Häufigkeitsverteilungen von bereits gelösten Proteinen erstellt wurde (**grün**), würde dagegen einen Grenzwert von 0,28 vorhersagen. Mit diesem Grenzwert würden 6 Hauptkettenamidgruppen und 5 Seitenkettenamidgruppen mehr auf dem 100 %-Sicherheitsniveau bestätigt werden als mit dem Grenzwert, der mithilfe des Fits bestimmt wurde.

## Zuordnung des HSQC-Spektrums von HPr *S. aureus* (H15A)

In Abb. 41 sind die Wahrscheinlichkeiten der richtigen Zuordnungen der HSQC-Signale in Abhängigkeit von der Position der Aminosäuren in der Primärsequenz (**grün**) abgebildet. Die Wahrscheinlichkeiten der Zuordnungen wurden mit der Wahrscheinlichkeitsfunktion, die mithilfe des Fits bestimmt wurden, berechnet (Abb. 40 **B rot**). In **gelb** ist der lokale RMSD (RMSD der betrachteten und der beiden benachbarten Aminosäuren) der schweren Hauptkettenatome des gemittelten MD-Ensembles zum gemittelten NMR-Ensemble abgebildet. 93 % der Hauptkettenzuordnungen (Abb. 40 **A**) können mit 100 % Wahrscheinlichkeit bestätigt werden. Es sind keine Abhängigkeiten der Wahrscheinlichkeiten einer richtigen Zuordnung zum insgesamt sehr geringen lokalen RMSD beobachtbar. Nur 2 der 14 Seitenkettenzuordnungen (Abb. 40 **B**) können mit 100 % Wahrscheinlichkeit bestätigt werden. Ihre Verifikationspseudoenergien sind durchweg geringer, als die der Hauptketten. Schließlich zeigt Abb. 42 **A** die mit SIBASA zugeordnete Summenprojektion des 3D  $^{15}\text{N}$ -NOESY/HSQC-Spektrums auf die  $\omega 2\omega 3$ -Ebene, die im Prinzip ein HSQC Spektrum ist.

Im Vergleich zur Sicherheitswahrscheinlichkeit von 0 % fehlen 100 %-Niveau 6 Hauptketten- und 12 Seitenkettensignale.

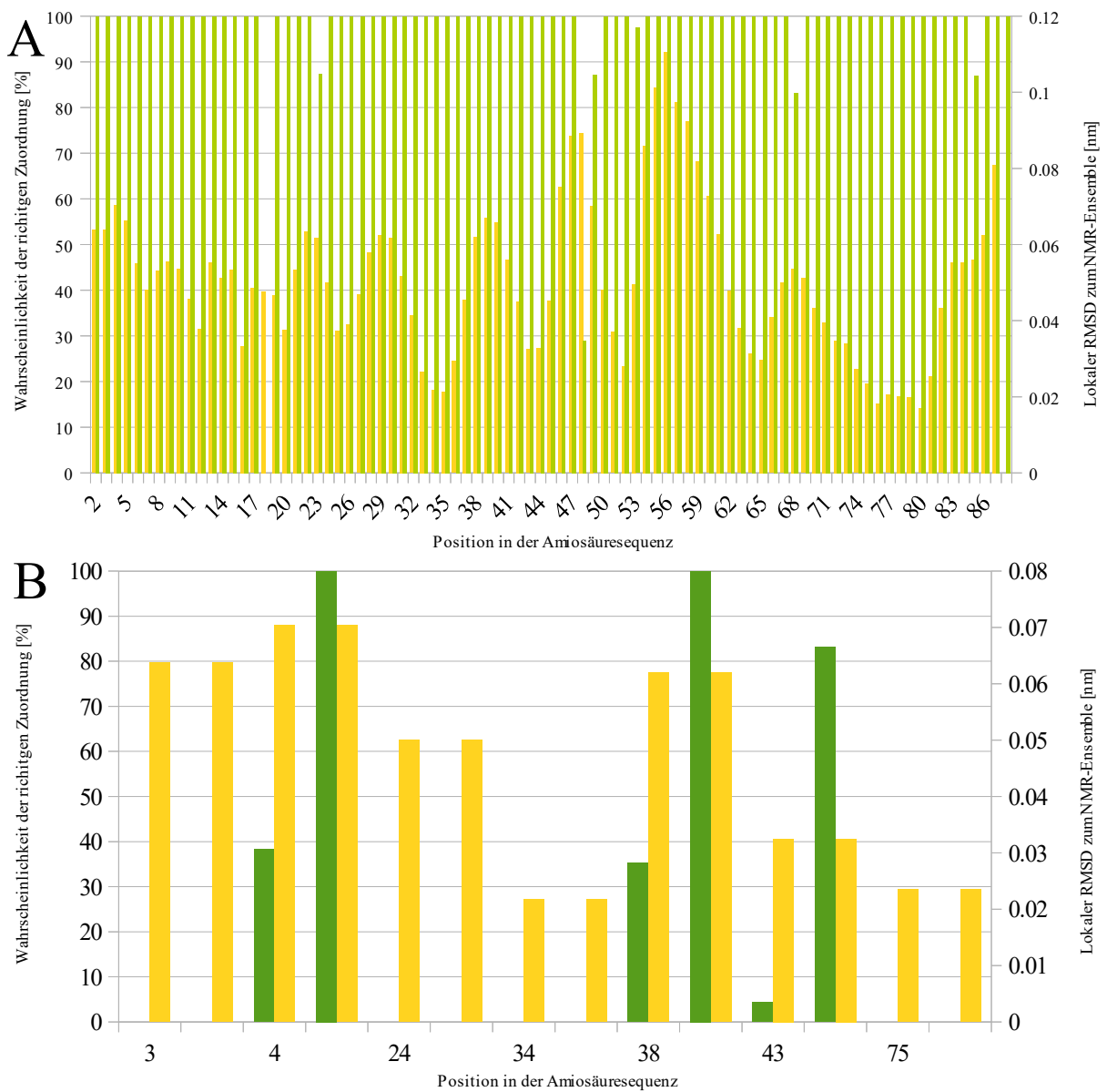


Abbildung 41: **Wahrscheinlichkeiten der richtigen Zuordnung von HSQC-Signale von HPr S. aureus (H15A) in Abhängigkeit von der Position der Aminosäure in der Primärsequenz.** Die **grünen** Balken zeigen die Wahrscheinlichkeiten von richtig gefundene Zuordnungen an. In **gelb** ist der lokale RMSD (RMSD der Aminosäure und seiner beiden nächsten Nachbarn) der schweren Hauptkettenatome des gemittelten MD-Ensembles zum gemittelten NMR-Ensemble aufgetragen.

**A** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Hauptketten.

**B** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Seitenketten von Asn und Gln. Die Wahrscheinlichkeit des HSQC-Signals von Asn  $N^{\delta 2}/Asn H^{\delta 21}$  ist immer links von der Wahrscheinlichkeit des Signals von Asn  $N^{\delta 2}/Asn H^{\delta 22}$  und die Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 2}/Gln H^{\epsilon 21}$  ist immer links von der Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 21}/Gln H^{\epsilon 22}$  aufgetragen.

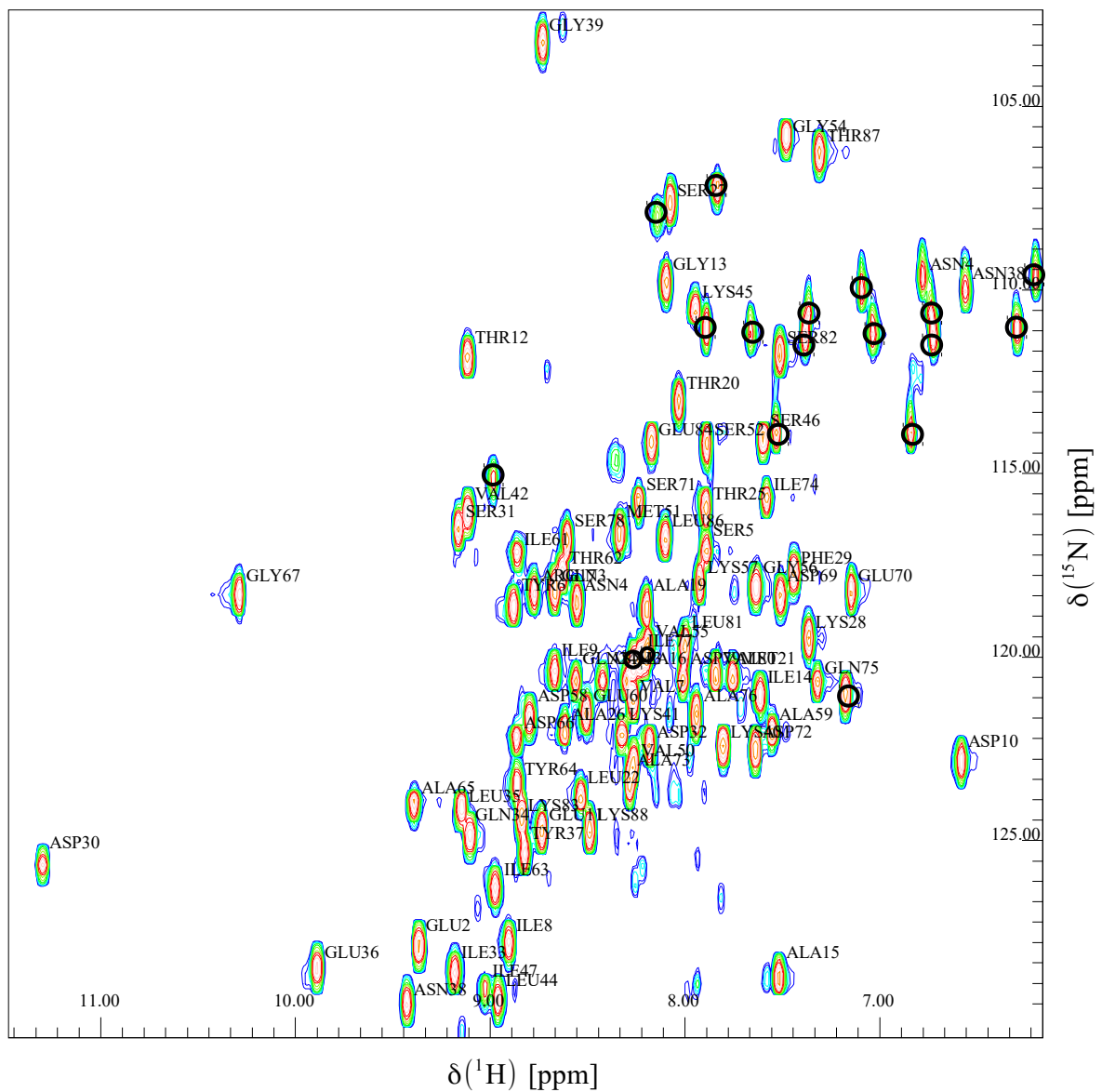


Abbildung 42: **Automatisch mit dem MD-Ensemble zugeordnetes HSQC-Spektrum von HPr *S. aureus* (H15A).** Die vom Benutzer wählbare Sicherheitswahrscheinlichkeit beträgt hier 99,8 %. In **schwarz** sind die fehlenden Zuordnungen eingekreist. Bei den Hauptketten fehlen 6 der 86 und bei den Seitenketten 12 der 14 Signale.



### 4.3.3 Automatische Zuordnung des reduzierten Plasmodium falciparum Thioredoxin

#### Vorbereitung der Röntgenstruktur

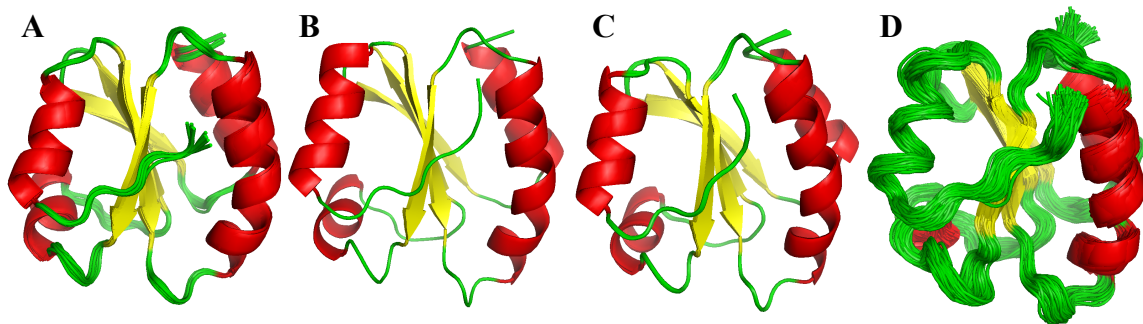


Abbildung 43: **Abbildung der betrachteten 3D-Strukturen des reduzierten PfTrx.** **A** Die besten 20 NMR-Strukturen des reduzierten PfTrx (Munte, Veröffentlichung in Vorbereitung), **B** Röntgenstruktur von PfTrx aus Komplex mit Thioredoxin Reductase (PDB ID: 4J56), **C** Die besten 20 Modelle des reduzierten PfTrx aus der Röntgenstruktur, **D** MD-Ensemble vom besten Modell (200 Strukturen). Der mittlere quadratischen Abstand (RMSD) der schweren Hauptkettenatome der gemittelten Strukturen zum gemittelten NMR-Ensemble beträgt für **B** 0,125 nm, für **C** 0,170 nm und für **D** 0,133 nm. Die Modelle wurden mit PyMOL (DeLano Scientific, LLC) erstellt.

	NMR-Ensemble	Röntgenstruktur	Top10 Modelle von PfTrx aus der Röntgenstruktur	MD-Ensemble (200 Strukturen)
RMSD [nm]	0	0,125	0,170	0,133
$\sum_{ij} E_{ij}^{Vol}$	1039		1096 (+5 %)	977 (-6 %)
$\sum_{ij} E_{ij}^{Shape}$	1348		1471 (+10 %)	1430 (+6 %)

Tabelle 10: **Summe der lokalen Pseudoenergien der Rückrechnungen und RMSD-Werte der 3D-Strukturen von PfTrx.** Für die RMSD-Werte wurden die schweren Hauptkettenatome der gemittelten 3D-Strukturen und die schweren Hauptkettenatome des gemittelten NMR-Ensembles berücksichtigt. Die Volumen- und Linienformpseudoenergien  $E_{ij}^{Vol}$  und  $E_{ij}^{Shape}$  der Signale  $j$  im Spektrum  $i$  ( $i = 2$ -D-NOESY bzw.  $3D$   $^{15}N$ -NOESY-HSQC) wurden mit der vollständigen richtigen Zuordnung von PfTrx berechnet. In Klammern sind die relativen Unterschiede der Pseudoenergien zu den Pseudoenergien der Rückrechnungen der NMR-Strukturen aufgelistet. Es wurde der Volumenskalingfaktor  $\alpha_{2-D-NOESY}^{Ring}$  für das 2-D-NOESY- und  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  für das  $3D$   $^{15}N$ -NOESY-HSQC-Spektrum verwendet.

Als Basis der automatischen Zuordnung des reduzierten Plasmodium falciparum Thioredoxin (PfTrx) wurde eine Struktur aus der Röntgenstruktur (Abb. 43 A) des Komplexes

von PfTrx mit PfTrx Reductase entnommen (Fritz-Wolf et al. 2013) (PDB ID: 4J56). Die Struktur hat im Gegensatz zur betrachteten Primärsequenz von PfTrx zwei Punktmutationen, so dass eine Homologie Modellierung mit AUREMOL PERMOL nötig war. Die 10 besten Modelle sind in Abb. 43 C abgebildet. Aus der besten der 1.000 homologen Strukturen wurde anschließend eine MD-Trajektorie erzeugt. Aus 200 äquidistanten Strukturen dieser Trajektorie (Abb. 43 D) wurden die NOESY-Spektren für SIBASA zurückgerechnet und aus 1.000 äquidistanten Strukturen wurden die speziellen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen mit SHIFTX2 (Han et al. 2011) erzeugt (Kap. 2.2.7).

Tabelle 10 listet die RMSD-Werte der schweren Hauptkettenatome der Strukturensamples zum gemittelten NMR-Ensemble (Munte et. al., Veröffentlichung in Vorbereitung) (PDB ID: 2MMN) und die Pseudoenergien des Volumens  $E^{\text{Vol}}$  und der Linienform  $E^{\text{Shape}}$  der jeweiligen Rückrechnungen mit der richtigen vollständigen Zuordnungen auf. Der RMSD der schweren Hauptkettenatome der gemittelten Strukturensamples zum NMR-Strukturensemble steigt mit jeder Weiterverarbeitung der Röntgenstruktur leicht an, bleibt aber auf niedrigem Niveau. Durch die MD-Simulation wurden die Pseudoenergie des Volumens um 11 % und die der Linienform um 3 % gegenüber dem Modell verkleinert. Die Pseudoenergie des Volumens war für die Rückrechnungen aus der MD-Trajektorie mit der vollständigen richtigen Zuordnung sogar 6 % kleiner als für die Rückrechnungen des NMR-Ensembles.

## Automatische Bestimmung der Volumenskalierungsfaktoren und Linienverbreiterungen

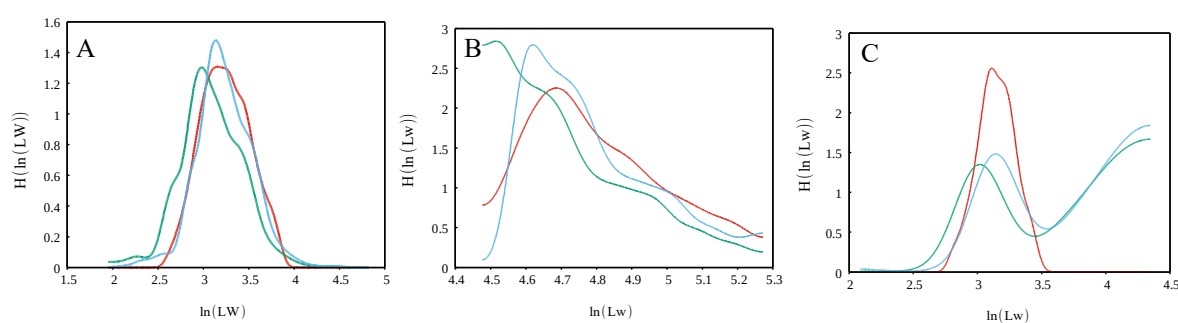


Abbildung 44: **Häufigkeitsverteilung der experimentellen und simulierten logarithmierten Linienbreiten  $H(\ln(Lw))$  im PfTrx.** Die Häufigkeitsverteilungen der experimentellen Linienbreiten sind in **rot**, die der simulierten Linienbreiten in **grün** und die der statistisch optimierten simulierten Linienbreiten in **blau** eingezeichnet. **A** Häufigkeitsverteilungen der Linienbreiten in der  $\omega_2$ -Dimension des 2-D-NOESY-Spektrums, **B** Häufigkeitsverteilung der Linienbreiten in der  $\omega_1$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums und **C** Häufigkeitsverteilung der Linienbreiten in der  $\omega_3$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die aus dem MD-Ensemble zurückgerechneten Spektren hatten zufällige Startzuordnungen.

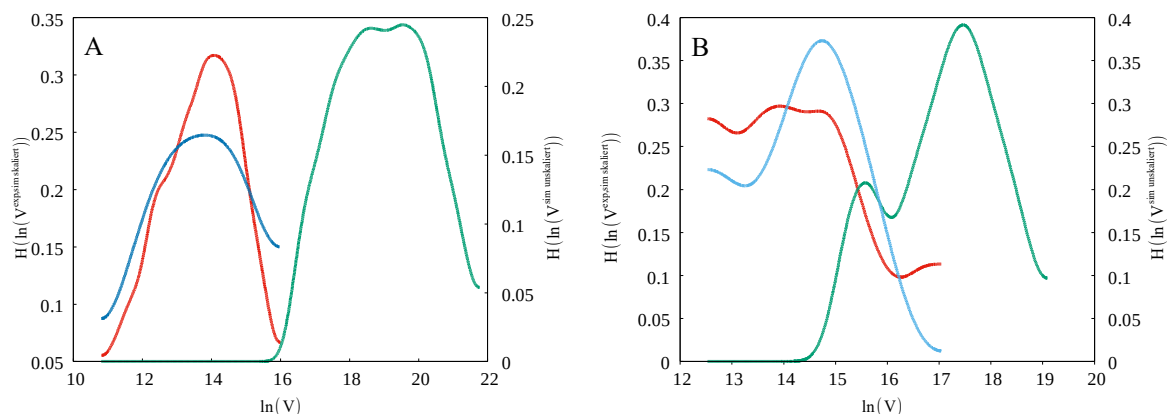


Abbildung 45: **Häufigkeitsverteilung der experimentellen und simulierten logarithmierten Volumina  $H(\ln(V))$  im PfTrx.** Die Häufigkeitsverteilungen der experimentellen Volumina sind in **rot**, die der simulierten Volumina in **grün** und die der statistisch skalierten simulierten Volumina in **blau** eingezeichnet. **A** Häufigkeitsverteilungen der Volumina im 2-D-NOESY-Spektrum, **B** Häufigkeitsverteilungen der Volumina im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Die logarithmierten simulierten Volumina sind gegen die y2-Achse aufgetragen. Die aus dem MD-Ensemble zurückgerechneten Spektren hatten zufällige Startzuordnungen und wurden mit der statistischen Methode verbreitert.

Während der Vorbereitung der automatischen Zuordnung wurden die Rückrechnungen der NOESY-Spektren wieder bestmöglich an die experimentellen Spektren angepasst. Zuerst wurde die zusätzlichen Linienverbreiterungen mit der statistischen Methode aus Kap. 2.2.5 bestimmt. Abb. 44 zeigt die Graphen der Häufigkeitsverteilungen der logarithmierten experimentellen (**rot**), logarithmierten simulierten (**grün**) und der verbreiterten logarithmierten simulierten Linienbreiten (**blau**) für die  $\omega_2$ -Dimension des 2-D-NOESY- und die  $\omega_1$ - und  $\omega_3$ -Dimensionen des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die Häufigkeitsverteilungen der verbreiterten Linien stimmen besser mit den Häufigkeitsverteilungen der experimentellen Linienbreiten überein als die Verteilungen der nicht verbreiterten Linien. Tabelle 11 listet neben den automatisch bestimmten Werten der Linienverbreiterungen auch ihren Effekt auf die Pseudoenergien des Volumens  $E_i^{\text{Vol}}$  und der Linienform  $E_i^{\text{Shape}}$  der vollständigen richtigen Zuordnung der Eingabespektren  $i$  auf. Werden die zusätzliche Linienverbreiterungen auf die Rückrechnungen angewendet, so verbessert sich die Pseudoenergie der Linienform der Referenz des 2-D-NOESY-Spektrums um über 5 %.

Anschließend wurden die Volumenskalierungsfaktoren, die vor allem für die Bestimmung der lokalen Pseudoenergien des Volumens  $E_{ij}^{\text{Vol}}$  (Gl. 45 wichtig sind, zwischen den experimentellen NOESY-Spektren und den Rückrechnungen statistisch ermittelt (Kap. 2.2.4). In Abb. 45 sind die Häufigkeitsverteilungen der logarithmierten Volumina aufgetragen. Die statistisch bestimmten Volumenskalierungsfaktoren sorgen dafür, dass sich die Häufigkeitsverteilungen der experimentellen und skalierten simulierten Volumina für beide NOESY-Spektren bestmöglich überlappen.

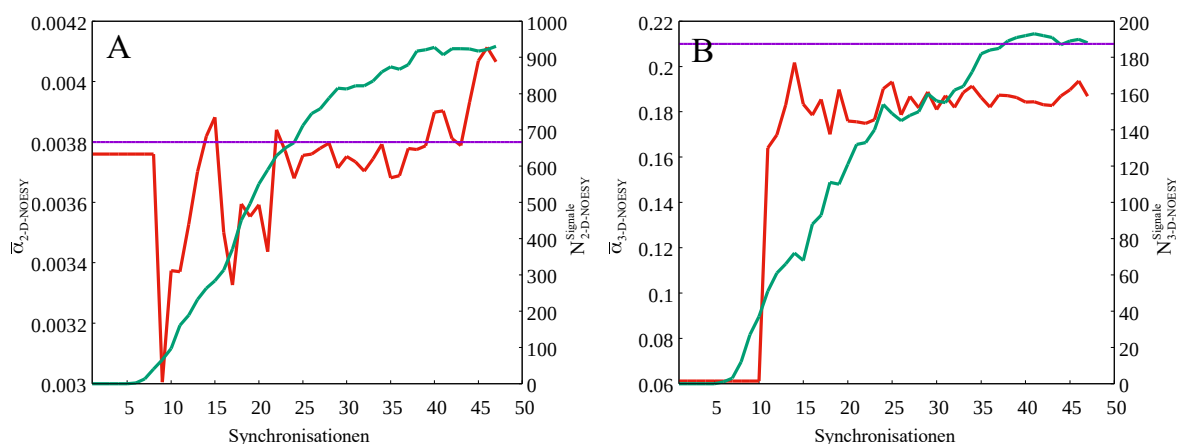


Abbildung 46: **Entwicklung der Volumenskalierungsfaktoren  $\bar{\alpha}_i$  (rot) während der automatischen Zuordnung von PfTrx.** In grün (y2-Achse) ist die Anzahl der eindeutig zuordenbaren Signale  $N_i^{\text{Signale}}$  aufgetragen. In lila sind die Volumenskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$ , die mit den eindeutig zuordenbaren Kreuzsignalen der vollständigen richtigen Zuordnung bestimmbar waren, eingezeichnet. **A** 2-D-NOESY-Spektrum, **B** 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum.

In Tabelle 11 sind die statistischen Volumenskalierungsfaktoren  $\alpha_i^{\text{statistisch}}$  und die mit der richtigen vollständigen Zuordnung bestimmbaren Volumenskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$  aufgelistet. Der statistisch gefundene Skalierungsfaktor des 2-D-NOESY-Spektrums  $\alpha_{2\text{-D-NOESY}}^{\text{statistisch}}$  war genauso groß wie der mit allen eindeutig zuordenbaren Signalen bestimmte Skalierungsfaktor  $\bar{\alpha}_{2\text{-D-NOESY}}^{\text{Ref}}$ . Der relative Unterschied zum optimalen Skalierungsfaktor, der nur mit den NOESY-Signalen der Phe- und Tyr-Ringen bestimmt worden ist, beträgt 5 %. Die beste Pseudoenergie des Volumens des 2-D-NOESY-Spektrums konnte mit dem Skalierungsfaktor aus Ringkontakten  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$  erreicht werden. Diese Pseudoenergie wird mit dem statistischen Skalierungsfaktor aber nur um 3 % verfehlt. Beim 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum des PfTrx beträgt der relative Unterschied des statistisch bestimmten Volumenskalierungsfaktors zum Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  68,6 %. Die Pseudoenergie des Volumens  $E_{3\text{-D-NOESY}}^{\text{Vol}}$  der richtigen Zuordnung war für den statistischen Skalierungsfaktor um 33 % höher als für den Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$ .

In Abb. 46 ist in rot die Abhängigkeit der Volumenskalierungsfaktoren vom Zeitpunkt in der automatischen Zuordnung für die beiden NOESY-Spektren aufgetragen. Der bereits gut bestimmte Skalierungsfaktor des 2-D-NOESY-Spektrums (Abb. 46 A) variiert nur geringfügig um seine ideale Position  $\bar{\alpha}_{2\text{-D-NOESY}}^{\text{Ref}}$  (lila). Der Volumenskalierungsfaktor des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums (Abb. 46 B) nähert sich dem realen Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  bis auf 10 % an.

	2-D-NOESY	3D <sup>15</sup> N-NOESY-HSQC
$\alpha_i^{\text{statistisch}}$	0,0038 (-5 %)	0,066 (-68,6 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	846 (+3 %)	205 (+33 %)
Anzahl der Signale	5114	905
$\bar{\alpha}_i^{\text{Ref}}$	0,0038 (-5 %)	0,21
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	845 (+3 %)	154
Anzahl der Signale	2416	472
$\alpha_i^{\text{Ring}}$	0,0040	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	821	
Anzahl der Signale	86	
Verbreiterung $\omega_1$ [Hz]	0	5,7
Verbreiterung $\omega_2$ [Hz]	3,3	0
Verbreiterung $\omega_3$ [Hz]		2,6
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ ohne Verbreiterung	822	155
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ mit Verbreiterung	821 (0 %)	153 (-1 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ ohne Verbreiterung	1151	279
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ mit Verbreiterung	1088 (-5,4 %)	282 (+1 %)

**Tabelle 11: Auf verschiedene Arten bestimmte Volumenskalierungsfaktoren und statistisch bestimmte Linienverbreiterungen und ihre Auswirkung auf die lokalen Pseudoenergien der richtigen Zuordnung von PfTrx.** Es sind die statistisch bestimmten ( $\alpha_i^{\text{statistisch}}$ ), die mit allen eindeutig zuordenbaren Signalen bestimmten ( $\bar{\alpha}_i^{\text{Ref}}$ ) und nur mit eindeutigen Ringkontakten bestimmten Skalierungsfaktoren ( $\alpha_i^{\text{Ring}}$ ) angegeben. Außerdem sind die Pseudoenergie des Volumens und die Anzahl der experimentellen Signale, die bei der Bildung des Skalierungsfaktor beteiligt waren, aufgelistet.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  bzw.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$  sind die über alle experimentelle Signale  $j$  summierten Pseudoenergien des Volumens bzw. der Linienform im Spektrum  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ).  $\omega_1$ ,  $\omega_2$  und  $\omega_3$  sind die Dimensionen der Spektren, in denen die Signale der Rückrechnung verbreitert wurden. In Klammern sind die relativen Unterschiede der Skalierungsfaktoren und Pseudoenergien zu der Referenz bzw. der nicht verbreiterten Rückrechnung angegeben.

## Häufigkeitsverteilungen der Pseudoenergien der richtigen Zuordnung

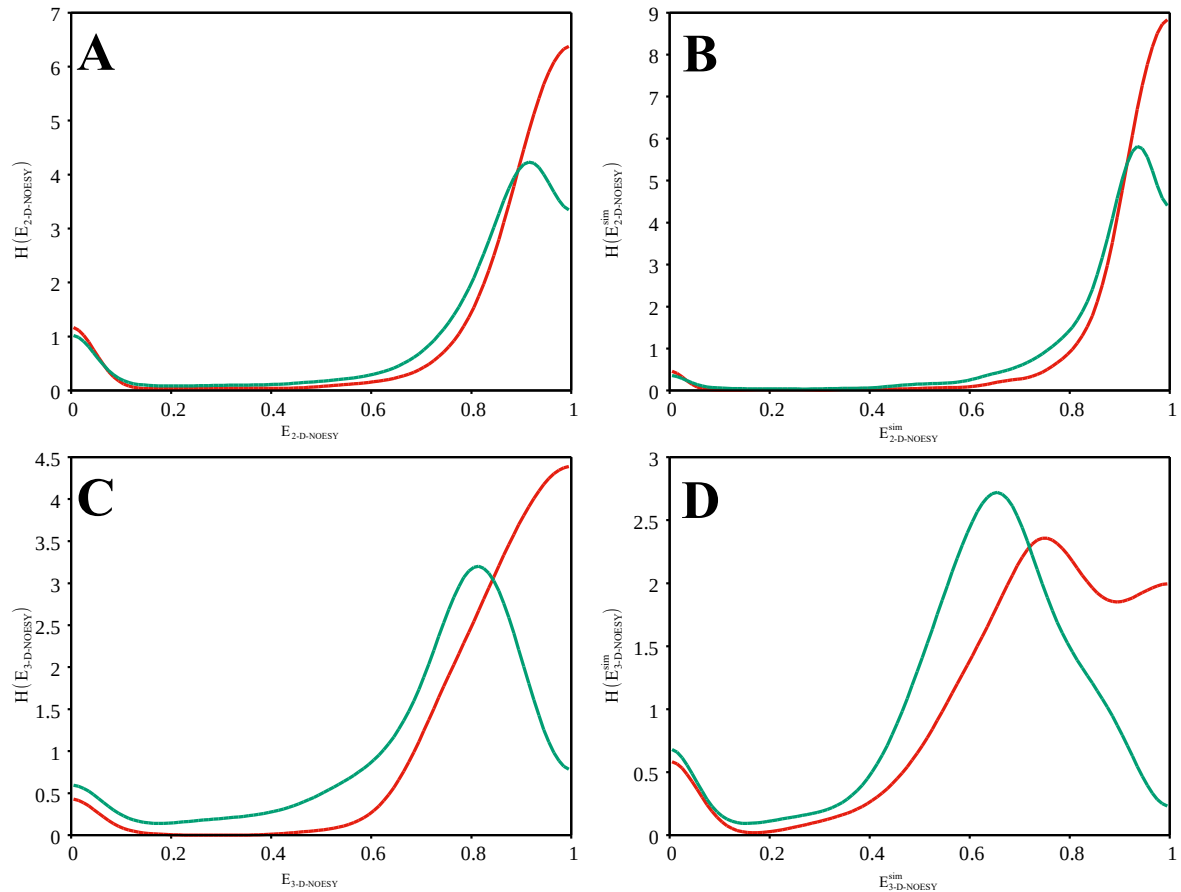
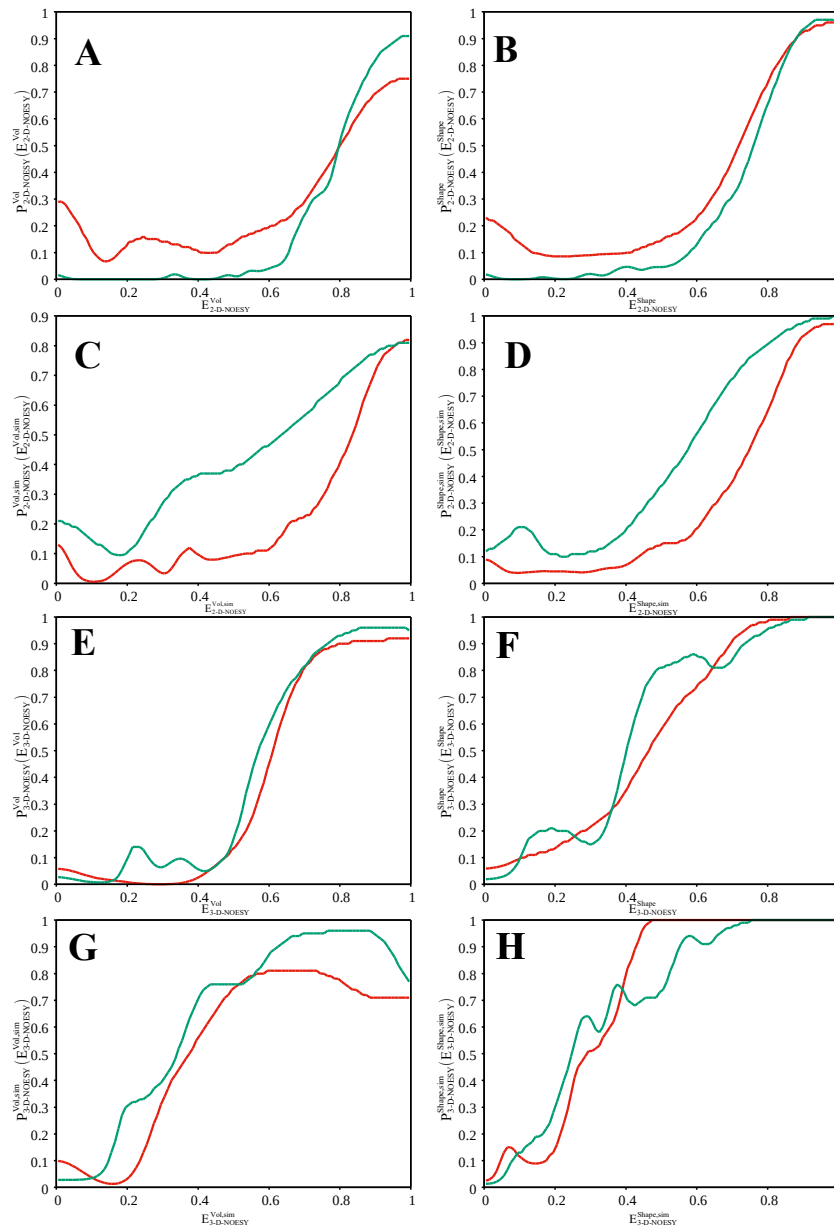


Abbildung 47: Häufigkeitsverteilungen  $H(E)$  der lokalen Pseudoenergien des Volumens  $E^{Vol}$  (rot) und der Linienform  $E^{Shape}$  (grün) der richtigen vollständigen Zuordnung von PfTrx. **A** Auf den Grundflächen der experimentellen Signale im 2-D-NOESY-Spektrum, **B** auf den Grundflächen der simulierten Signale im 2-D-NOESY-Spektrum, **C** auf den Grundflächen der experimentellen Signale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum, **D** auf den Grundflächen der simulierten Signale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und mit einer zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{Ring}$ , für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum der Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{Ref}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.



**Abbildung 48: Abhängigkeit der Bayesschen Wahrscheinlichkeit der lokalen richtigen Zuordnung  $P_i(E)$  von den lokalen Pseudoenergien  $E_i$  (rot) für die NOESY-Spektren des PfTrx. In grün sind die zugehörigen Wahrscheinlichkeitsfunktionen, die SIBASA während der automatischen Zuordnung verwendet (Kap. 4.1.3), abgebildet. A  $P_{2-D-NOESY}^{Vol}(E_{2-D-NOESY}^{Vol})$ , B  $P_{2-D-NOESY}^{Shape}(E_{2-D-NOESY}^{Shape})$ , C  $P_{2-D-NOESY}^{Vol,sim}(E_{2-D-NOESY}^{Vol,sim})$  und D  $P_{2-D-NOESY}^{Shape,sim}(E_{2-D-NOESY}^{Shape,sim})$  des 2-D-NOESY-Spektrums. E  $P_{3-D-NOESY}^{Vol}(E_{3-D-NOESY}^{Vol})$ , F  $P_{3-D-NOESY}^{Shape}(E_{3-D-NOESY}^{Shape})$ , G  $P_{3-D-NOESY}^{Vol,sim}(E_{3-D-NOESY}^{Vol,sim})$  und H  $P_{3-D-NOESY}^{Shape,sim}(E_{3-D-NOESY}^{Shape,sim})$  des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und mit einer zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalierungsfaktor  $\alpha_{2-D-NOESY}^{Ring}$ , für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum der Skalierungsfaktor  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.**

Mit Hilfe der Häufigkeitsverteilungen der lokalen Pseudoenergien der richtigen vollständigen Zuordnung wurde für PfTrx überprüft, ob die Optimierung zum richtigen Ergebnis geführt hat und ob die Häufigkeitsverteilungen der Pseudoenergien, wie beim HPr S. aureus (H15A), für die Berechnung der lokalen Wahrscheinlichkeiten  $P_{ij}^{\text{Vol,Shape}}$  durch die entsprechenden Häufigkeitsverteilungen eines bereits gelösten Proteins ersetzt werden können. Abb. 47 zeigt die Häufigkeitsverteilungen der Pseudoenergien des Volumens  $E_i^{\text{Vol}}$  (**rot**) und der Linienform  $E_i^{\text{Shape}}$  (**grün**) der richtigen Zuordnung auf den Grundflächen der experimentellen und simulierten Signale in den Spektren  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ). Als Basis der Rückrechnungen der NOESY-Spektren diente das MD-Ensemble. Für das 2-D-NOESY-Spektrum wurde der Skalierungsfaktor aus den Ringkontakten  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$  und für das  $3\text{D }^{15}\text{N-NOESY-HSQC}$ -Spektrum der Skalierungsfaktor aus allen eindeutig zuordenbaren Signalen  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  verwendet. Die zusätzlichen Linienverbreiterungen wurden mit SIBASA automatisch bestimmt. Die Verteilungen zeigen wieder, dass sich die gesuchte richtige vollständige Zuordnung durch hohe Pseudoenergien auszeichnet.

In Abb. 48 sind die Bayesschen Wahrscheinlichkeiten der lokalen richtigen Zuordnung  $P_i^{\text{Vol,Shape}}(E_i)$  in Abhängigkeit von den Pseudoenergien für PfTrx (**rot**) eingezeichnet. In **grün** sind die Abhängigkeiten der Wahrscheinlichkeiten von den Pseudoenergien, die für die Berechnung der lokalen Wahrscheinlichkeiten bei unbekannter Zuordnung verwendet werden (Kap. 4.3.1), aufgetragen. Man erkennt, dass die Graphen der Wahrscheinlichkeit für beide Datensätze sehr ähnlich verlaufen. Im Bereich niedriger experimenteller Pseudoenergien geht die **grüne** Kurve tiefer, da bei der Erzeugung der zugrundeliegenden Häufigkeitsverteilungen experimentellen Signale ohne Zuordnung nicht mit berücksichtigt wurden. Der Unterschied zwischen dem realen Zusammenhang der Wahrscheinlichkeit von den Pseudoenergien und den Wahrscheinlichkeitsgraphen, die SIBASA für die Berechnung der lokalen Bayesschen Wahrscheinlichkeiten verwendet, kann vor allem beim 2-D-NOESY-Spektrum beobachtet werden (Abb. 48 **A** und **B**). Die Wahrscheinlichkeiten von SIBASA gehen für niedrige Pseudoenergien auf 0, die realen Wahrscheinlichkeiten dagegen bleiben auf höherem Niveau.



## Verlauf von Pseudoenergien und dynamischer Größen während der automatischen Zuordnung

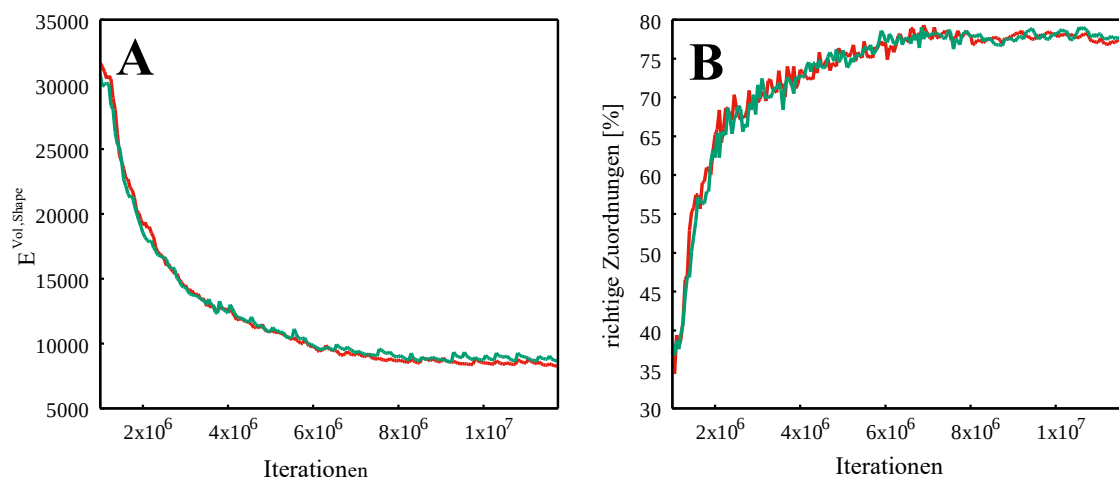


Abbildung 49: **Zeitlicher Verlauf der gesamten Pseudoenergie des Signalvergleichs und des Anteil der richtig identifizierten Zuordnungen bei der automatischen Zuordnung von PfTrx.** In rot ist die Instanz der Optimierung mit der besten und in grün die Instanz mit der schlechtesten finalen Pseudoenergie eingezeichnet. **A** Verlauf der Pseudoenergie des Volumen- und Linienformvergleichs auf den Grundflächen der experimentellen Signale  $E^{\text{Vol,Shape}}$ , **B** Verlauf der gefundenen richtigen Zuordnungen im Fehlerbereich  $\delta^{\text{ref}} \pm 0,03 \text{ ppm}$ .

Nach 47 Synchronisationen (11.750.000 Iterationen) war die automatische Zuordnung von PfTrx abgeschlossen. SIBASA hatte die minimale Pseudoenergie der Optimierung  $E^{\text{total}}$  (Gl. 44) gefunden und es wurden keine weiteren Resonanzzuordnungen mehr zur Laufzeit fixiert. Im folgenden wird der Verlauf der wichtigsten Pseudoenergien und andere dynamischer Parameter während der Optimierung ausgewertet.

In Abb. 49 **A** sind die Verläufe der Pseudoenergie des Volumen- und Linienformvergleichs  $E^{\text{Vol,Shape}}$  auf den Grundflächen der experimentellen Signale für die beiden Instanzen, die die Pseudoenergie am besten (**rot**) und am schlechtesten (**grün**) optimiert haben, aufgetragen. Bei der automatischen Zuordnung von PfTrx ist kaum ein Unterschied zwischen den Verläufen der Pseudoenergie der beiden Instanzen zu sehen. Auch beim Anteil der im Fehlerbereich von  $\pm 0,03 \text{ ppm}$  richtig gefunden chemischen Verschiebungen (Abb. 49 **B**) sind keine großen Unterschieden zwischen den beiden Instanzen sichtbar (77,34 % bzw. 77,76 % richtige Zuordnungen am Ende).

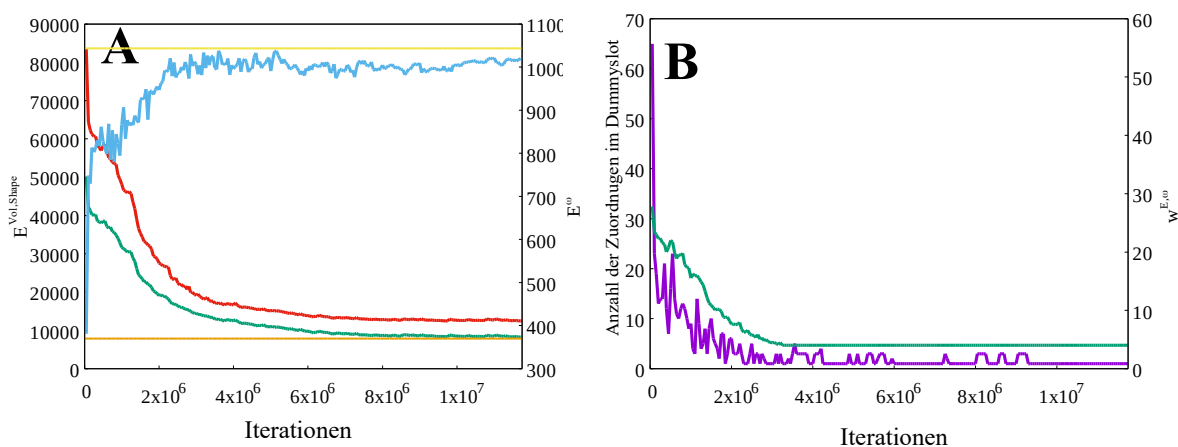


Abbildung 50: **Zeitliche Verläufe wichtiger Optimierungsgrößen einer Instanz der automatischen Zuordnung von PfTrx.** **A** Verlauf der gesamten Pseudoenergie der Optimierung  $E^{\text{total}}$  (rot, y-Achse), Verlauf der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  (grün, y-Achse) und Verlauf der Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  (blau, y2-Achse). In orange (y-Achse) ist der geschätzte optimale Wert der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape,opt}}$  und in gelb (y2-Achse) die Pseudoenergie der chemischen Verschiebung der richtigen vollständigen Zuordnung eingetragen. **B** Anzahl der Mitglieder von Amidgruppen im Dummyslot (lila, y-Achse) und Verlauf des dynamischen Gewichtungsfaktors der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}$  (grün, y2-Achse), dessen Trend auch auf den Gewichtungsfaktor der Pseudoenergie des Dummyslots  $w^{E,\text{Dummy}}$  übertragen werden kann.

In Abb. 50 **A** ist in **rot** die Pseudoenergie der Optimierung  $E^{\text{total}}$  und in **grün** die Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  einer Instanz der automatischen Zuordnung aufgetragen. Die Pseudoenergie des Signalvergleichs konvergiert bis auf 5 % ( $E^{\text{Vol,Shape}}(\text{Ende}) = 8242$ ,  $E^{\text{Vol,Shape,opt}} = 7867$ ) gegen den geschätzten Wert der vollständigen richtigen Zuordnung  $E^{\text{Vol,Shape,opt}}$  (**gelb**) (Gl. 65). Die Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  (**blau**) (Gl. 62) verfehlt den Wert der richtigen Zuordnung (**gelb**) nur um auf 2 % ( $E^{\omega}(\text{Referenz}) = 1043$ ,  $E^{\omega}(\text{Ende}) = 1018$ ). Der dynamische Skalierungsfaktor der Pseudoenergie der chemischen Verschiebung in der Gesamtpseudoenergie  $w^{E,\omega}(t)$  (Abb. 50 **B grün**) (Gl. 64) fällt dabei im Laufe der automatischen Zuordnung auf den erlaubten Minimalwert ab. Die Anzahl der Kerne, deren Resonanzen von SIBASA als nicht sichtbar identifiziert wurden, ist in **lila** aufgetragen. Nur die Zuordnung der Resonanz von Arg36 N<sup>ε</sup> ist am Ende im Dummyslot verblieben. Die Amidgruppe von Gly30, die eigentlich dem Dummyslot zugeordnet werden sollte, war am Ende dort nicht zu finden.

## Festlegung von Zuordnungen während der Laufzeit

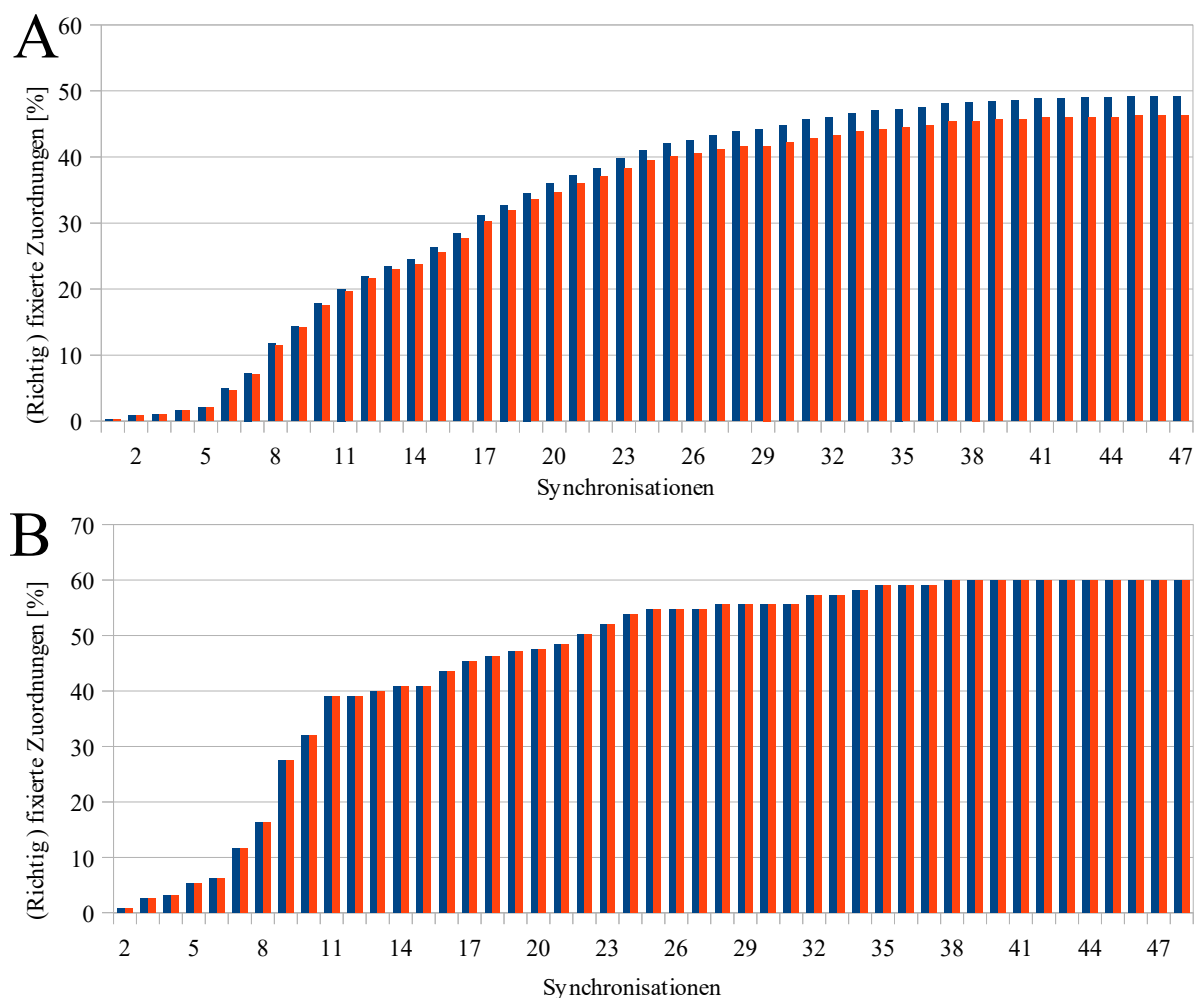


Abbildung 51: Anteil der zur Laufzeit fixierte Zuordnungen (blau) und Anteil der zur Laufzeit richtig fixierten Zuordnungen (rot) für die automatische Zuordnung von Pflrx. **A** Alle Zuordnungen von Resonanzen, **B** Nur Zuordnungen Resonanzen von Amidprotonen und Stickstoffen.

Während der Synchronisationsphasen wurden die besten Cluster der chemischen Verschiebungen der Resonanzen der Kerne  $k$  mit der maximalen gewichteten Standardabweichung der chemischen Verschiebung  $\sigma_k^w(\text{best})$  von 0,03 ppm erstellt (Kap. 2.2.13). Die gewichtete mittlere chemische Verschiebung wurde in allen Instanzen für den restlichen Verlauf der automatischen Zuordnung festgelegt, falls die Verifikationspseudoenergie des Kerns  $k$  im besten Cluster  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  größer als 0,85 oder für Amidprotonen und Stickstoffen die Verifikationspseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  größer als 0,73 war (Kap. 4.3.1). Abb. 51 A zeigt den Anteil der fixierten Resonanzen (**blau**) in Abhängigkeit vom Zeitpunkt in der automatischen Zuordnung. In **rot** ist immer der Anteil der richtig fixierten Resonanzen angegeben. Der Anteil der

fixierten Zuordnungen steigt wieder zu Beginn stark an und sättigt bei einem Anteil von ungefähr 45 %. Insgesamt wurden 353 der 715 Resonanzen von PfTrx zur Laufzeit fixiert (49,3 %). Im Gegensatz zur automatischen Zuordnung von HPr *S. aureus* (H15A), für das nur eine Resonanz zur Laufzeit falsch fixiert wurde, waren es beim PfTrx bereits 21 Zuordnungen. Das entspricht einem Anteil von 3 %.

Zuordnungen von Amidprotonen und Stickstoffen wurden dagegen zur Laufzeit nicht falsch festgelegt (Abb. 51 B). Der Anteil der fixierten Zuordnung von Amidgruppen steigt zu Beginn schneller an als der Anteil der restlichen fixierten Zuordnungen und ist, wie bei der automatischen Zuordnung von HPr *S. aureus* (H15A), mit 59,7 % am Ende auch höher.

## Auswertung und Verifikation der vereinigten Zuordnungsliste

Nach der 47. Synchronisation der automatischen Zuordnung des reduzierten PfTrx wurde die vereinigte Zuordnungsliste mit der maximalen gewichteten Standardabweichung  $\sigma_k^w(\text{Cluster})$  von 0,03 ppm (Kap. 2.2.12) aus den 16 Instanzen erstellt. In ihr sind 586 der 715 (81,9 %) der zu bestimmenden chemischen Verschiebungen von Resonanzen richtig zugeordnet. Bei den Amidgruppen mit sichtbaren NOESY-Kreuzsignalen waren es 212 der 225 Resonanzen (95,1 %).

### Häufigkeitsverteilung der Verifikationspseudoenergie und Wahrscheinlichkeit der richtigen Zuordnung

Mithilfe der Verifikationspseudoenergien  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der Kerne  $k$  (Gl. 81) bzw.  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der Amidgruppen  $k$  (Gl. 85) konnten die richtigen von den falschen Zuordnungen in der vereinigten Zuordnungsliste getrennt werden. Abb. 52 zeigt die Häufigkeitsverteilungen der Verifikationspseudoenergien und die resultierenden Bayesschen Wahrscheinlichkeiten  $P(\text{korrekt} | E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  bzw.  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$ . In A ist zu erkennen, dass falsche Zuordnung nach der automatischen Zuordnung von PfTrx niedrigere Verifikationspseudoenergien als richtige Zuordnungen haben. Im Vergleich zum Ergebnis der automatischen Zuordnung von HPr *S. aureus* (H15A), sind mehr falsche Zuordnungen mit hoher Pseudoenergie vertreten. Dies wirkt sich auf den Verlauf der Wahrscheinlichkeit der richtigen Zuordnung aus (Abb. 52 B), die nie auf 100 % ansteigt, sondern bei hohen Pseudoenergien nur gegen 95 % konvergiert.

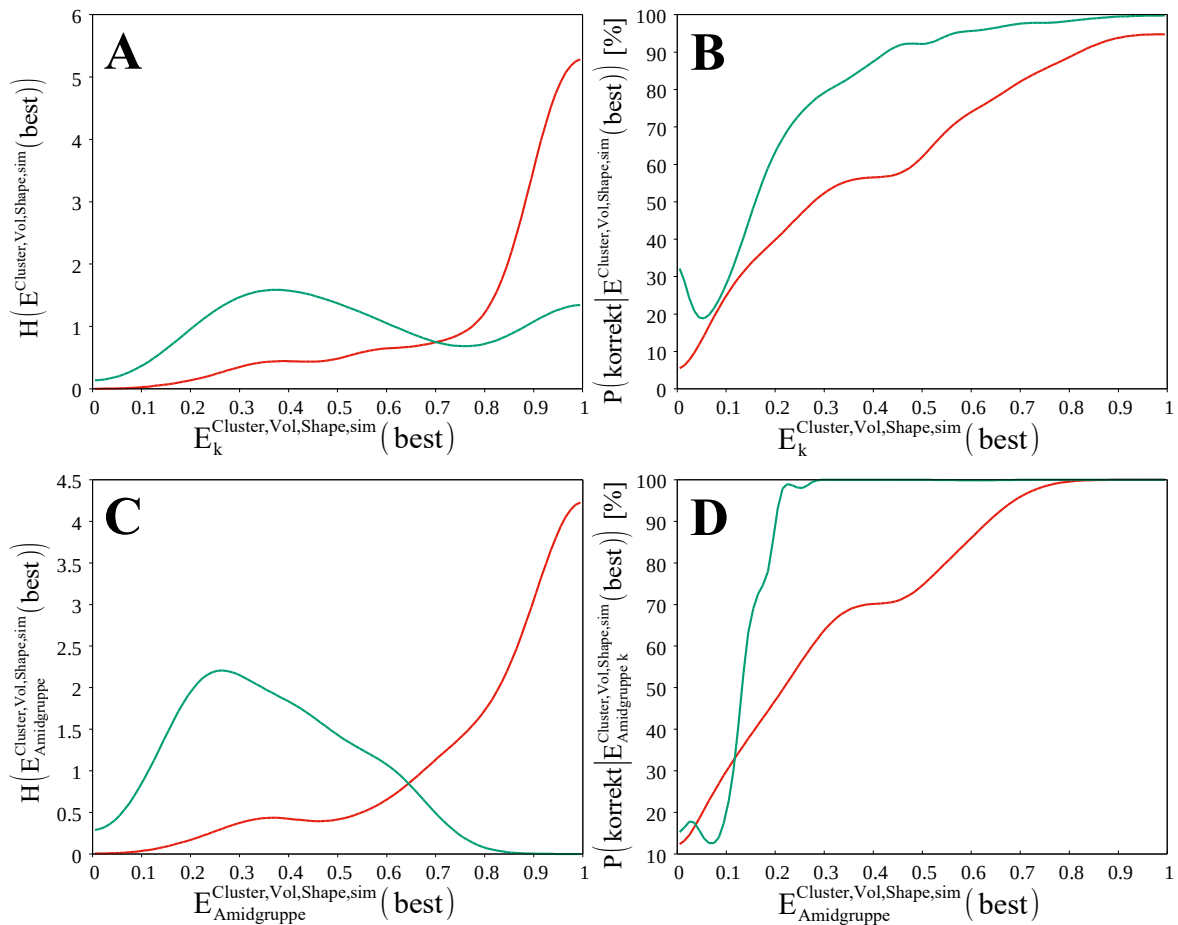


Abbildung 52: **Häufigkeitsverteilungen der Verifikationspseudoenergien  $E^{Cluster,Vol,Shape,sim}(best)$  und Wahrscheinlichkeiten der richtigen Zuordnung nach der automatischen Zuordnung von PfTrx.**

**A** Häufigkeitsverteilung  $H(E^{Cluster,Vol,Shape,sim}(best))$  der Verifikationspseudoenergien der richtigen (**rot**) und der falschen Zuordnungen (**grün**), **B** Bayessche Wahrscheinlichkeit einer richtigen Zuordnung  $P(korrekt|E^{Cluster,Vol,Shape,sim}(best))$  einer Resonanz (**rot**) in Abhängigkeit von der Verifikationspseudoenergie und zugehörige Wahrscheinlichkeitskurve, die mit Hilfe von bereits gelösten Proteinen erzeugt und in die AUREMOL Datenbank gespeichert wurde (**grün**) (Abb. 28).

**C** Häufigkeitsverteilung  $H(E_{Amidgruppe}^{Cluster,Vol,Shape,sim}(best))$  der richtigen (**rot**) und der falschen Zuordnungen (**grün**) von Amidgruppen, **D** Bayessche Wahrscheinlichkeit der richtigen Zuordnung  $P(korrekt|E_{Amidgruppe}^{Cluster,Vol,Shape,sim}(best))$  einer Amidgruppe (**rot**) in Abhängigkeit von der Verifikationspseudoenergie und zugehörige Wahrscheinlichkeitskurve aus der AUREMOL Datenbank (**grün**) (Abb. 28).

Im Gegensatz zum HPr *S. aureus* (H15A) sind in der vereinigten Zuordnungsliste falsche Resonanzzuordnungen von Amidgruppen vorhanden. Die Verteilungen in Abb. 52 C zeigen, dass die richtigen Zuordnungen (**rot**) größtenteils hohe Pseudoenergien haben. Die wenigen falschen Zuordnungen (**grün**) sind aber fast gleichmäßig über den Bereich zwischen 0,1 und 0,7 verteilt. Dies führt dazu, dass die Wahrscheinlichkeit einer richtige Zuordnung einer Amidgruppe (Abb. 52 D) fast linear im Bereich von 0 bis 0,7 auf 100 % ansteigt. Die Wahrscheinlichkeit der richtigen Zuordnungen konnten beim PfTrx weder für die

Amidgruppen noch für die H<sup>α</sup>- und Seitenkettenprotonen mithilfe des Satzes von Bayes und den Häufigkeitsverteilungen der Verifikationspseudoenergien eines Referenzdatensatzes (Abb. 52 **B** und **D** grün) gut genug genähert werden. Die Bayesschen Wahrscheinlichkeiten, die mit dem Referenzdatensatz berechnet werden können, sind höher als in der Realität, wodurch der Anteil der falschen Zuordnungen höher wird als eigentlich beabsichtigt.

Für die in Kap. 4.3.1 festgelegten minimalen Grenzwert  $p = 0,85$  bzw.  $p_{\text{Amidgruppe}} = 0,73$  der Verifikationspseudoenergien ist die reale Wahrscheinlichkeit einer falschen Zuordnung einer Resonanz eines Nichtamidprotons 8,5 % und 0 % für die Resonanzen von Amidgruppen.

### Die Güte der Verifikationspseudoenergie als Separator zwischen falschen und richtigen Zuordnungen

Um die Eignung der Verifikationspseudoenergien als Entscheidungskriterium zwischen falschen und richtige gefunden chemischen Verschiebungen zu untersuchen, ist in Abb. 53 **A** die modifizierte ROC-Kurve für die Zuordnungen aller Kerne und in Abb. 53 **C** die modifizierte ROC-Kurve für die Zuordnungen von Amidgruppen abgebildet. Der Verlauf der Kurve in Abb. 53 **A** ähnelt dem Verlauf der entsprechenden ROC-Kurve von HPr *S. aureus* (H15A). Die Fläche unter der Kurve, welche die Güte der Verifikationspseudoenergie als Separator zwischen falschen und richtig gefunden chemischen Verschiebungen misst, ist mit 80,4 % nur um 3 % kleiner. Man kann erkennen, dass es für H<sup>α</sup>- und Seitenkettenprotonen keinen minimalen Grenzwert  $p$  der Verifikationspseudoenergie, für den nur noch richtige Zuordnungen bestätigt werden, gibt. In Abb. 53 **C** ist die modifizierte ROC-Kurve der Verifikationspseudoenergie der Zuordnungen von Amidgruppen aufgetragen. Die Fläche unter der Kurve beträgt 93,2 %. Hier existiert ein Grenzwert  $p_{\text{Amidgruppe}}$  der Verifikationspseudoenergie, für den keine falsche Zuordnung in Kauf genommen werden müssen und gleichzeitig 61 % der chemischen Verschiebungen von Resonanzen von Amidgruppen richtig identifiziert werden.

Die Abhängigkeit der Wahrscheinlichkeit  $P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der richtigen Zuordnung von der Verifikationspseudoenergie wurde mit den Satz von Bayes und den Datensätzen von bereits gelösten Proteinen bestimmt und soll als absolutes Qualitätskriterium für eine Zuordnung dienen (Kap. 2.2.15). In Abb. 53 **B** ist der Anteil der richtig (**rot**) und falsch identifizierten (**grün**) chemischen Verschiebung von Resonanzen gegen die Wahrscheinlichkeit  $P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  aufgetragen. Im Idealfall würde der Anteil der falschen Zuordnungen, die als richtig identifiziert werden  $1 - P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  betragen (**blaue** Winkelhalbierende), ist aber in diesem Fall durchwegs höher. Maximal werden 2,7 % (insgesamt 5 %) mehr falsche chemische Verschiebungen in die gefilterte Zuordnungsliste aufgenommen als beabsichtigt, wenn der Benutzer eine Sicherheitswahrscheinlichkeit von 97,7 % wählt.

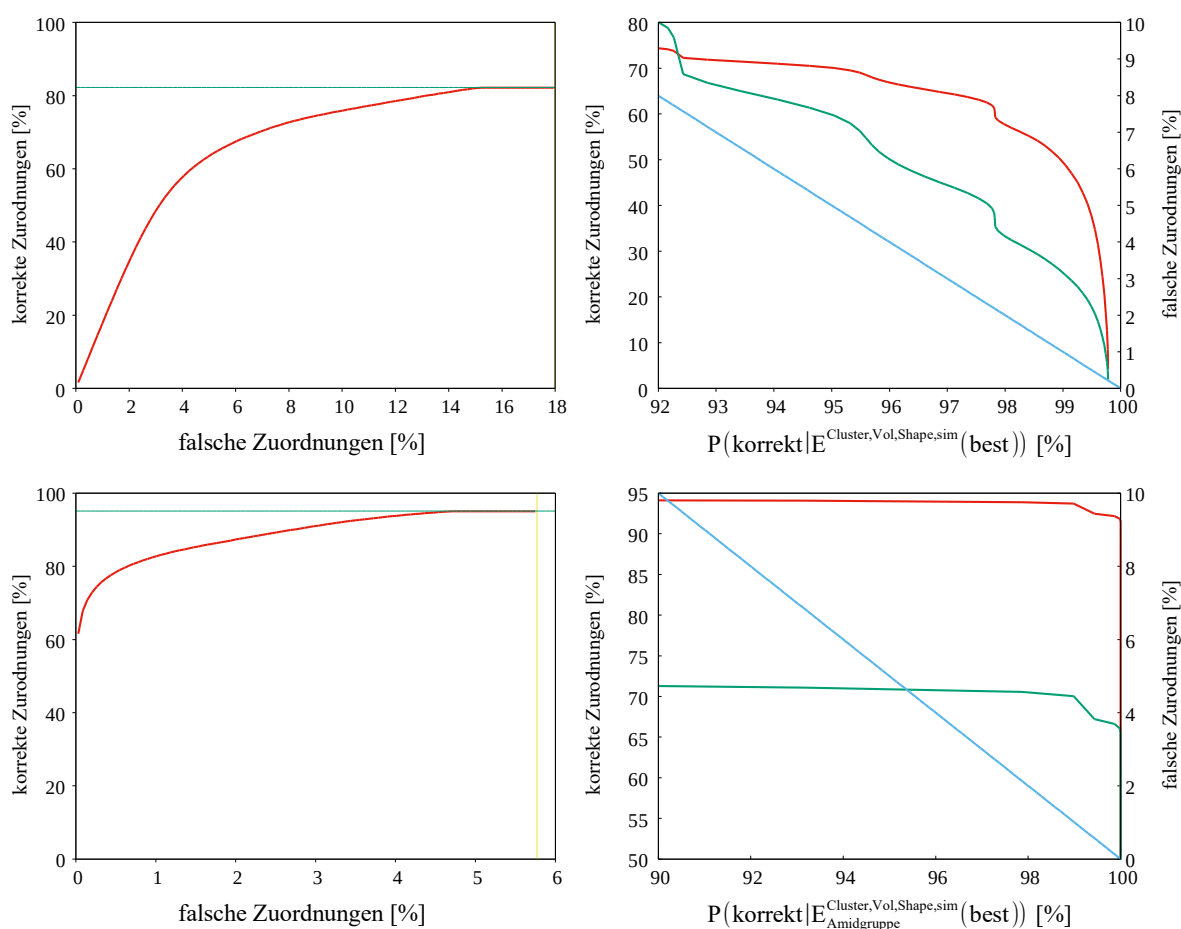


Abbildung 53: **Auswirkung eines Grenzwertes der Verifikationspseudoenergie  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  auf die Klassifizierung der Resonanzzuordnungen in der vereinten Liste der chemischen Verschiebung von PfTrx.** **A** Modifizierte ROC-Kurve der Verifikationspseudoenergie. Es sind die möglichen richtigen Zuordnungen gegen die dafür in Kauf zu nehmenden falschen Zuordnungen aufgetragen (**rot**). Insgesamt sind 82,3 % der im  $3D^{15}\text{N-NOESY-HSQC}$ -Spektrum sichtbaren Resonanzen richtig (**grün**) und 17,9 % aller Resonanzen falsch zugeordnet worden (**gelb**). **B** Gefundene richtige (**rot**, y-Achse) und falsche Zuordnungen (**grün**, y2-Achse) in Abhängigkeit von der von Sicherheitswahrscheinlichkeit  $P(\text{korrekt} | E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$ . Der vom Benutzer tolerierten Anteil von falschen Zuordnungen ist in **blau** (Winkelhalbierende, y2-Achse) eingetragen. **C** Modifizierte ROC-Kurve der Verifikationspseudoenergien von Amidgruppen  $E^{\text{Cluster,Vol,Shape,sim}}_{\text{Amidgruppe}}(\text{best})$ . Insgesamt sind 95,5 % der im  $3D^{15}\text{N-NOESY-HSQC}$ -Spektrum sichtbaren Resonanzen von Amidgruppen richtig (**grün**) und 5,3 % aller Resonanzen von Amidgruppen falsch (**gelb**) zugeordnet worden.

Beim signifikanten Niveau (95 % Sicherheitswahrscheinlichkeit) sind insgesamt 7,5 % der gefilterten Zuordnungen falsch, 70 % richtig und 22,5 % bleiben undefiniert. In Abb. 53 **D** ist die gleiche Abhängigkeiten für Mitglieder von Amidgruppen aufgetragen. Hier kann der Anteil der falschen Zuordnungen in der gefilterten Zuordnungsliste bis zu einer Sicherheitswahrscheinlichkeit von 95,5 % eingehalten werden. Der maximale Fehler liegt bei einer Sicherheitswahrscheinlichkeit von 99,9 % bei der 3,6 % mehr falschen Zuordnungen als

beabsichtigt (insgesamt 3,7 % falsche Zuordnungen), übernommen werden.

### Abschätzung der Wahrscheinlichkeiten der richtigen Zuordnung von HSQC-Signalen

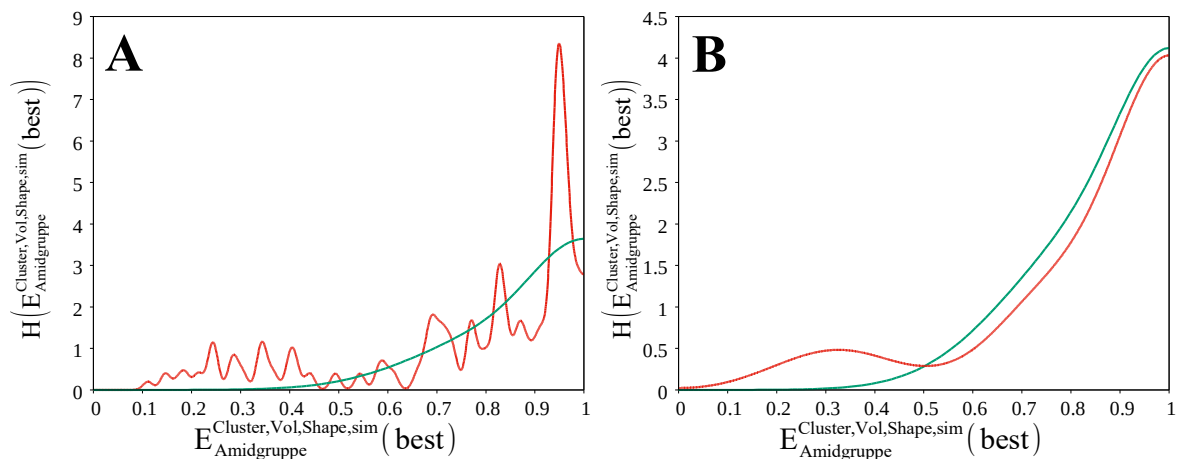


Abbildung 54: **Schätzung der Häufigkeitsverteilung der Verifikationspseudoenergien der richtig zugeordneten Resonanzen von Hauptkettenamidgruppen beim PfTrx.** **A** Wenig geglättete (rot) und mit  $H^{\text{korrekt}}(x)$  (Gl. 92) gefittete Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  (grün) der Verifikationspseudoenergien der Zuordnungen von Hauptkettenamidgruppen. **B** Vergleich zwischen der gefitteten (rot) und der realen (grün) Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der Verifikationspseudoenergien der richtigen Zuordnungen von Hauptkettenamidgruppen.

Da der Graph der Bayesschen Wahrscheinlichkeiten  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  einer richtigen Zuordnung, der mit Häufigkeitsverteilungen der Verifikationspseudoenergien bereits gelöster Proteine gebildet worden ist (Abb. 28 D), für hohe Verifikationspseudoenergien kaum mehr Abstufungen kennt, sind diese Wahrscheinlichkeiten als absolute Qualitätsmerkmal für die Zuordnung von Amidgruppen nicht geeignet. Um eine bessere Abhängigkeit der Wahrscheinlichkeiten der richtigen Zuordnung von Amidprotonen und Stickstoffen von den Verifikationspseudoenergien zu erhalten wird, wie in Kap. 4.3.2 für HPr *S. aureus* (H15A) beschrieben, die Häufigkeitsverteilung der hohen Verifikationspseudoenergien  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der Hauptkettenamidgruppen mit Gl. 92 gefittet. Dieser Fit dient wieder als Näherung der Häufigkeitsverteilung der Verifikationspseudoenergie der richtigen Zuordnung. Die linke Grenze des Fits wird wieder durch Analyse der wenig verbreiterten Häufigkeitsverteilung der Hauptkettenamidgruppen (Abb. 54 A rot) auf 0,64 gesetzt, da die Häufigkeitsverteilung  $H(E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  an diesem Punkt zum ersten mal auf nahezu 0 fällt. In Abb. 54 A und B ist jeweils in grün die gefittete Verteilung von der richtigen Zuordnungen  $H^{\text{korrekt}}(x)$  aufgetragen. Abb. 12 B vergleicht den Fit von Gl. 92 mit der realen Verteilung der richtigen Zuordnungen.



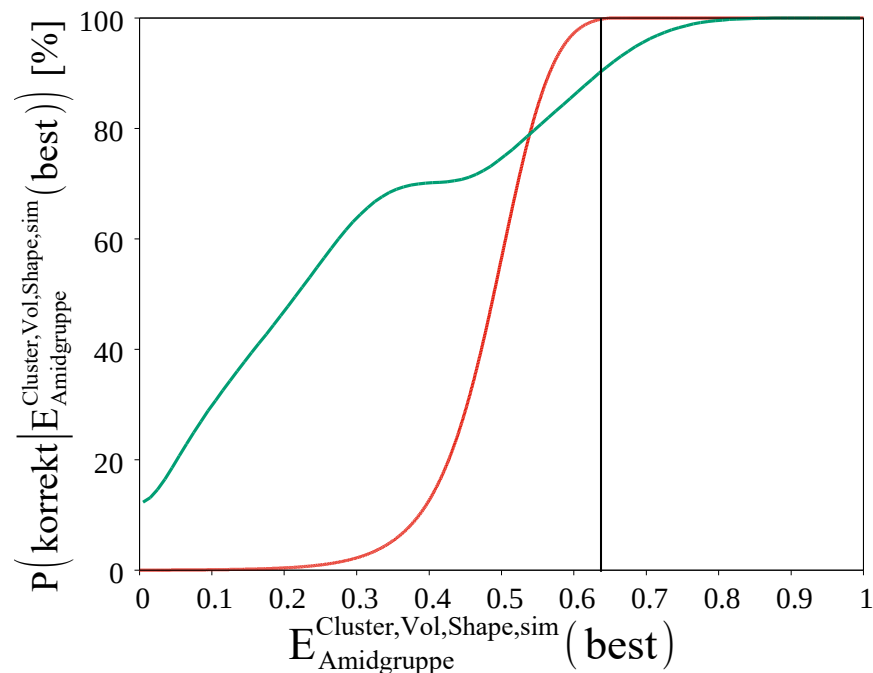
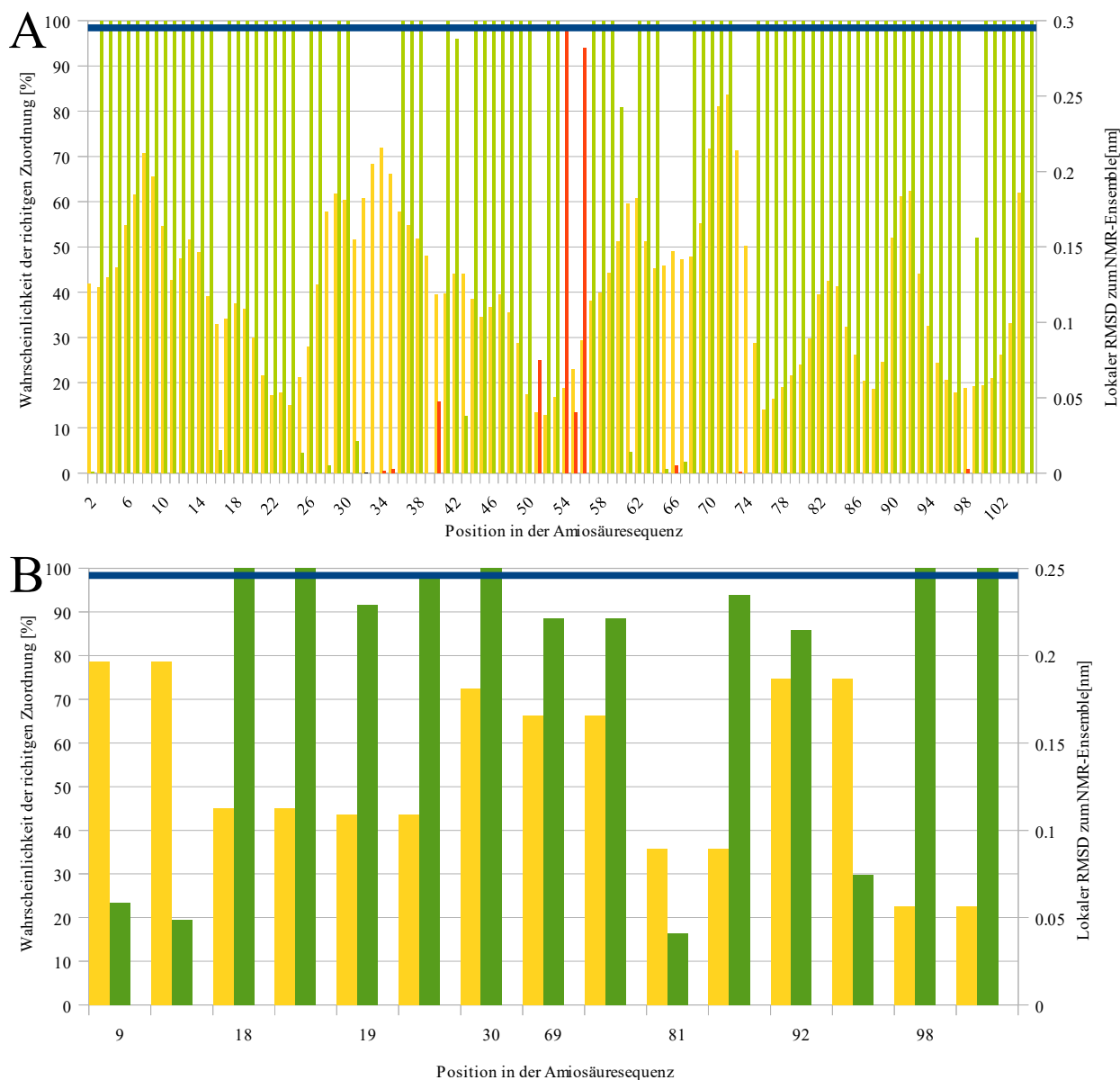


Abbildung 55: Mithilfe des Fits von Abb. 54 geschätzte Bayessche Wahrscheinlichkeit der richtigen Zuordnung eines Amidprotons oder Stickstoffs  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  in Abhängigkeit von der Verifikationspseudoenergie (rot). In grün ist der Wahrscheinlichkeitsgraph, der mit Hilfe von Datensätzen bereits gelöster Proteine bestimmt wurde, eingezeichnet. In schwarz ist der bestimmte Grenzwert der Verifikationspseudoenergie  $p_{\text{Amidgruppe}} = 0,64$  für sichere Zuordnungen eingetragen.

Man kann erkennen, dass die Näherung im Bereich der hohen Verifikationspseudoenergien gut mit der realen Kurve (rot) übereinstimmt. In Abb. 55 ist die mit dem Fit geschätzte (rot) und die reale (grün und Abb. 52 D) Abhängigkeit der Wahrscheinlichkeit  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  einer richtigen Zuordnung eines Amidprotons oder Stickstoffs von der Verifikationspseudoenergie aufgetragen. Der Grenzwert der Verifikationspseudoenergie  $p_{\text{Amidgruppe}}$ , der benötigt wird um alle falschen Zuordnungen von HSQC-Signalen aus der vereinigten Zuordnungsliste zu entfernen, kann wieder bei der Wahrscheinlichkeit von 99,8 % abgelesen werden (schwarze Linie). Er beträgt 0,639 und liegt über der höchsten Verifikationspseudoenergie einer falschen Zuordnung (0,61).

## Zuordnung des HSQC-Spektrums von PfTrx



**Abbildung 56: Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale von PfTrx in Abhängigkeit von der Position der Aminosäure in der Primärsequenz.** Die grünen Balken zeigen die Wahrscheinlichkeiten von richtig gefundenen, die roten die Wahrscheinlichkeiten von falsch gefundene Zuordnungen an. In gelb ist der lokale RMSD (RMSD der Aminosäure und seiner beiden nächsten Nachbarn) der schweren Hauptkettenatome des gemittelten MD-Ensembles zum gemittelten NMR-Ensemble aufgetragen. Die dunkelblaue Linie gibt den minimalen Grenzwert der Wahrscheinlichkeit, für den keine falschen Zuordnungen mehr bestätigt werden, an.

**A** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Hauptketten.

**B** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Seitenketten von Asn und Gln und Trp30. Die Wahrscheinlichkeit des HSQC-Signals von Asn  $N^{\delta 2}/Asn H^{\delta 21}$  ist immer links von der Wahrscheinlichkeit des Signals von Asn  $N^{\delta 2}/Asn H^{\delta 22}$  und die Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 2}/Gln H^{\epsilon 21}$  ist immer links von der Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 21}/Gln H^{\epsilon 22}$  aufgetragen.

In Abb. 56 A sind die Wahrscheinlichkeiten der richtige Zuordnung der HSQC-Signale der Hauptkette von PfTrx gegen die Primärsequenz aufgetragen. Die Wahrscheinlichkeiten wurden mithilfe der geschätzten Wahrscheinlichkeitsfunktion aus Abb. 55 (**rot**) berechnet. In **gelb** ist der lokale RMSD (RMSD der Aminosäure und der beiden nächsten Nachbarn) der schweren Hauptkettenatome vom gemittelten MD-Ensemble zum gemittelten NMR-Ensemble aufgetragen.

Eine der 11 falsch zugeordneten Amidgruppen hat keine sichtbaren Kreuzsignale (Gly32). Bei 7 der 10 verbleibenden Amidgruppen sind jeweils nur die Resonanzen der Stickstoffe falsch zugeordnet worden. Nur bei den HSQC-Signalen von Cys34 und Val56 sind die chemische Verschiebungen beider Resonanzen falsch gefunden worden. Die beiden falsch zugeordneten HSQC-Signale von Ile54 und Val56 haben Wahrscheinlichkeiten von nahezu 100 % und sind kaum noch von den richtigen gefundenen Zuordnungen zu unterscheiden. Die Zuordnungen der Aminosäure 31-35 im aktiven Zentrum haben nur kleine Wahrscheinlichkeiten. Dies betrifft die beiden Cysteine Cys31 und Cys34, die im oxidierten Zustand eine Disulfidbrücke bilden und die Amidgruppe von Gly32, die keine sichtbaren Kreuzsignale in den NOESY-Spektren hat. Die 3 Proline scheinen die Wahrscheinlichkeiten der richtigen Zuordnung in ihrer Umgebung negativ zu beeinflussen. In Abb. 56 B sind die Wahrscheinlichkeiten für die richtigen Zuordnungen der Seitenkettenamidgruppen von Asn, Gln und Trp gezeigt. Es wurden keine Signale falsch zugeordnet, aber nur 5 der 16 Amidgruppen haben Wahrscheinlichkeiten, die höher als 99,8 % sind. Eine Abhängigkeit der Wahrscheinlichkeit einer richtigen Zuordnung vom lokalen RMSD, der maximal 0,25 nm beträgt, kann bei der automatischen Zuordnung von PfTrx nicht beobachtet werden. Im aktiven Zentrum steigt der lokale RMSD zwar an, bleibt aber insgesamt auf einem niedrigen Niveau.

Abb. 57 zeigt die Summenprojektion des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums auf die  $\omega_2\omega_3$ -Ebene. Die Signale wurden mit einer Sicherheitswahrscheinlichkeit von 99,8 % zugeordnet. Insgesamt konnten 83 der 115 HSQC-Signale (72 %) (78 der 100 Hauptkettensignale) sicher zugeordnet werden. Auf dem 99,8 % -Niveau wurde keine falsche Zuordnung bestätigt.

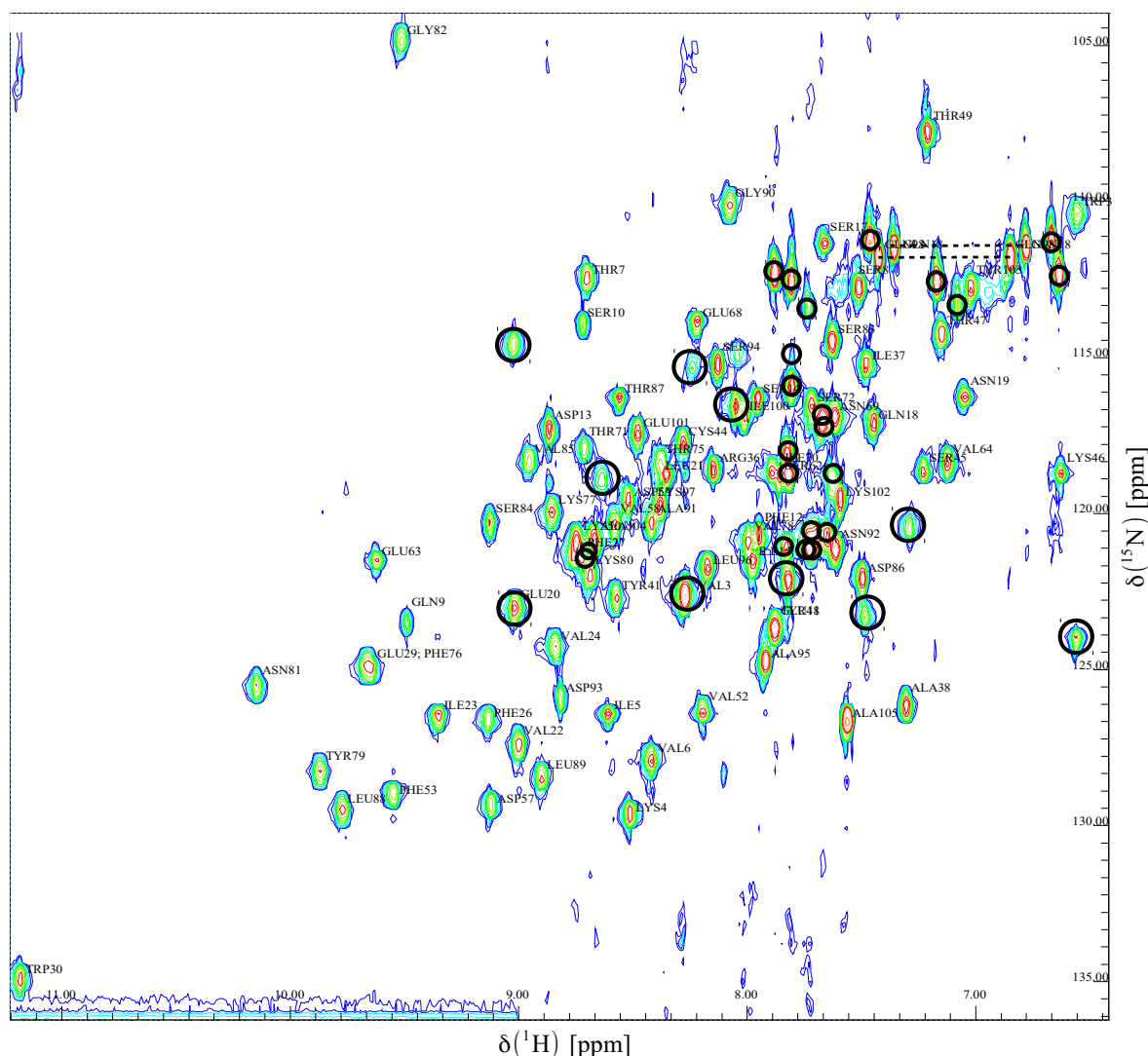


Abbildung 57: **Automatisch mit dem MD-Ensemble zugeordnetes HSQC-Spektrum von PfTrx.** Die vom Benutzer wählbare Sicherheitswahrscheinlichkeit beträgt 99,8 %. Es wurden keine Signale falsche zugeordnet. Die Positionen an denen Zuordnungen fehlen, sind **schwarz** umrandet. Insgesamt konnten 83 der 115 Signale und 78 der 100 Hauptkettensignale sicher gefunden werden.

#### 4.3.4 Automatische Zuordnung von Ras(T35S)-GppNHp

Im Gegensatz zu HPr *S. aureus* (H15A) und dem reduzierten PfTrx stand für Ras(T35S)-GppNHp keine vollständige Zuordnung auf Basis der verwendeten NOESY-Spektren zur Verfügung. Eine vollständige Zuordnung, die bei einem anderen pH (pH 6,8 statt 7,4) und anderer Temperatur (298 K statt 303 K) und mit anderen Spektren erstellt wurde, wurde von Araki et al. (2011) veröffentlicht. Da aber eine Referenzzuordnung, die nicht selbst aus den Eingabespektren stammt, systematische Fehler in der Auswertung erzeugen würde, kann sie nicht ohne Anpassung übernommen werden.

Die chemischen Verschiebungen der Resonanzen der Amidgruppen, die stark von pH und

Temperatur abhängen, wurden auf Basis der  $\omega_2\omega_3$ -Projektion des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums angepasst (Puthenpurackal Narayanan et al.), wobei die chemischen Verschiebungen der Resonanzen der 343 Amidprotonen und Stickstoffen bestimmt wurden und außerdem festgestellt wurde, dass die Resonanzen von 12 Hauptkettenamidgruppen nur über ein Kreuzsignal mit dem Lösungsmittel im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum sichtbar waren.

Der Rest der Liste der chemischen Verschiebungen von Araki et al. (2011) wurde mit den aus dem NMR-Ensemble (PDB ID: 2CLF) zurückgerechneten NOESY-Spektren und SIBASA an die experimentellen NOESY-Spektren angepasst (Kap. 3.4.5).

## Auswahl und Vorbereitung der Röntgenstruktur

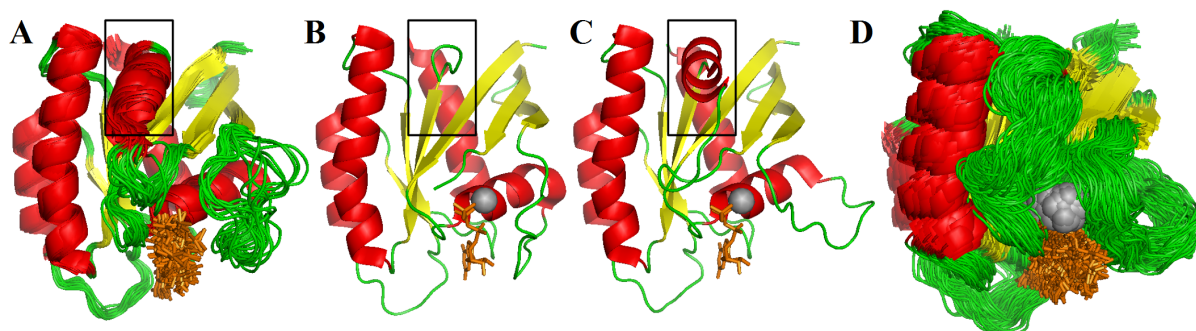


Abbildung 58: **Abbildung der betrachteten 3D-Strukturen von Ras(T35S)-GppNHp.** **A** NMR-Strukturenensemble (PDB ID: 2LCF), **B** Röntgenstruktur in Form 2 (PDB ID: 3KKM), **C** Röntgenstruktur in Form 1 (PDB ID: 3KKN), **D** 200 Strukturen umfassendes MD-Ensemble auf Basis der Röntgenstruktur Form 1. Die fehlende Helix in der Röntgenstruktur von Form 2 ist **schwarz** umrandet. Der RMSD-Wert der schweren Hauptkettenatome der gemittelten Strukturen zum gemittelten NMR-Ensemble beträgt für **B** 0,14 nm, für **C** 0,22 nm und für **D** 0,21 nm. Die Modelle wurden mit PyMOL (DeLano Scientific, LLC) erstellt.

Als Basis der MD-Simulation wurde die Röntgenstruktur von Ras(T35S)-GppNHp Form 1 (PDB ID: 3KKN) (Abb. 58 C) verwendet, weil sie im Gegensatz zu der Röntgenstruktur von Form 2 (Abb. 58 B) vollständig war (beide Strukturen stammen aus der Veröffentlichung von Shima et al. 2010). Die Struktur in Form 2 liegt zwar näher am NMR-Ensemble (Abb. 58 A) (Araki et al. 2011) (RMSD: 0,13 nm statt 0,22 nm) und kann auch experimentell Kontakte mit dem Nukleotid, die Form 1 nicht hat, erklären (Abb. 17), aber es kann vom Benutzer nicht erwartet werden, dass er die ideale Röntgenstruktur kennt oder sich die optimale Startstruktur selbst erzeugt. In diesem Fall wäre die optimale Struktur die Kombination aus der Röntgenstruktur Ras(T35S)-GppNHp Form 2 und der Helix aus Form 1.

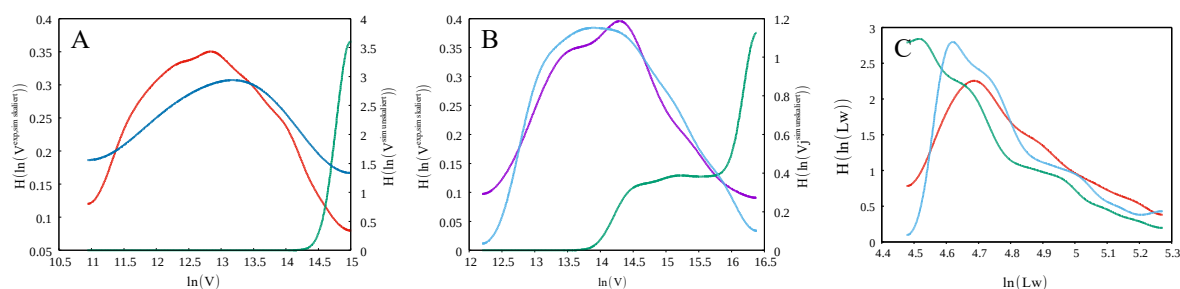
	NMR-Ensemble	Röntgenstruktur Form 2	Röntgenstruktur Form 1	MD-Ensemble (200 Strukturen)
RMSD [nm]	0	0,136	0,22	0,207
$\sum_{ij} E_{ij}^{Vol}$	1121		1260 (+12,4 %)	1109 (-1,1 %)
$\sum_{ij} E_{ij}^{Shape}$	1772		1868 (+5,4 %)	1749 (-1,3 %)

**Tabelle 12: Summe der lokalen Pseudoenergien der Rückrechnungen und RMSD-Werte der 3D-Strukturen von Ras(T35S)-GppNHp.** Für die RMSD-Werte wurden die schweren Hauptkettenatome der gemittelten 3D-Strukturen und die schweren Hauptkettenatome des gemittelten NMR-Ensembles berücksichtigt. Die Volumen- und Linienformpseudoenergien  $E_{ij}^{Vol}$  und  $E_{ij}^{Shape}$  der Signale  $j$  im Spektrum  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D }^{15}\text{N-NOESY-HSQC}$ ) wurden mit der vollständigen richtigen Zuordnung von Ras(T35S)-GppNHp berechnet. In Klammern sind die relativen Unterschiede der Pseudoenergien zu den Pseudoenergien der Rückrechnungen der NMR-Strukturen aufgelistet. Es wurde der Volumenskalingfaktor  $\alpha_{2\text{-D-NOESY}}^{Ring}$  für das 2-D-NOESY- und  $\bar{\alpha}_{3\text{-D-NOESY}}^{Ref}$  für das 3D  $^{15}\text{N-NOESY-HSQC}$ -Spektrum verwendet.

Die MD-Simulation (Abb. 58 **D**) kann den RMSD der schweren Hauptkettenatome zum NMR-Ensemble nur leicht verbessert. Tabelle 12 listet neben den RMSD-Werten auch die Summen der lokalen Pseudoenergien des Volumens  $E^{Vol}$  und der Linienform  $E^{Shape}$  (Kap. 2.2.8) für die entsprechenden Rückrechnungen und die richtige vollständige Zuordnung auf. Als Volumenskalingfaktoren wurden  $\alpha_{2\text{-D-NOESY}}^{Ring}$  für das 2-D-NOESY- und  $\bar{\alpha}_{3\text{-D-NOESY}}^{Ref}$  für das 3D  $^{15}\text{N-NOESY-HSQC}$ -Spektrum verwendet. Die Röntgenstruktur von Ras(T35S)-GppNHp Form 1 erzeugt wie erwartet die höchsten Pseudoenergien, mit den Rückrechnungen des MD-Ensemble wurden dagegen sogar bessere Werte als für die Rückrechnung des NMR-Ensembles berechnet. Durch die MD-Simulation verbessert sich die Ausgangslage von SIBASA gegenüber der Röntgenstruktur Form 1 um 12 % bei der Pseudoenergie des Volumens und um 6 % bei der Pseudoenergie der Linienform.

## Automatische Bestimmung der Volumenskalingfaktoren und Linienverbreiterungen

Die mit dem MD-Ensemble, den MD-Dynamikparametern und einer zufälligen Zuordnung erstellten Rückrechnungen wurden mithilfe der statistischen Methoden (Kap. 2.2.4 und Kap. 2.2.5) bestmöglich an die experimentellen Spektren angepasst.



**Abbildung 59: Häufigkeitsverteilung der experimentellen und simulierten logarithmierten Volumen  $H(\ln(V))$  und Linienbreiten  $H(\ln(Lw))$  im Ras(T35S)-GppNHp.** Die Häufigkeitsverteilungen der experimentellen Werte sind in rot, die der simulierten Werte in grün und die der statistisch optimierten simulierten Werte in blau eingezeichnet. **A** Häufigkeitsverteilungen der Volumen im 2-D-NOESY-Spektrum, **B** Häufigkeitsverteilungen der Volumen im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum und **C** Häufigkeitsverteilungen der Linienbreiten der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die logarithmierten simulierten Volumen in **A** und **B** sind gegen die y2-Achse aufgetragen. Die aus dem MD-Ensemble zurückgerechneten Spektren hatten zufällige Startzuordnungen und wurden für die Bestimmung der Volumenskalierungsfaktoren mit der statistischen Methode verbreitert.

SIBASA konnte nur für die indirekte  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums ( $\omega_1$ -Dimension) eine Linienverbreiterung von 7,3 Hz bestimmen. In Abb. 59 **C** sind die Häufigkeitsverteilungen der logarithmierten experimentellen (rot), logarithmierten simulierten (grün) und logarithmierten verbreiterten simulierten Linienbreiten (blau) dieser Dimension aufgetragen. Durch die Verbreiterung stimmen die Häufigkeitsverteilungen der simulierten und experimentellen Linienbreiten besser überein. Tabelle 13 listet die Veränderung der Summe der lokalen Pseudoenergien  $E_{3\text{-D-NOESY}}^{\text{Vol}}$  des Volumens und  $E_{3\text{-D-NOESY}}^{\text{Shape}}$  der Linienform des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums auf. Durch die Anpassung der Linienbreiten konnten in diesem Fall die Pseudoenergien der richtigen vollständigen Zuordnung nicht verbessert werden.

Anschließend wurden die Volumenskalierungsfaktoren statistisch ermittelt. Mit ihnen konnten die Verteilungen der logarithmierten skalierten simulierten Signalvolumen (Abb. 59 **A** und **B** blau) in beiden NOESY-Spektren bestmöglich an die Verteilungen der logarithmierten experimentellen Signalvolumen (rot) angepasst werden. In Tabelle 13 sind die statistisch und die mit der vollständigen richtigen Zuordnung bestimmten Skalierungsfaktoren zusammen mit den zugehörigen Pseudoenergien des Volumens  $E_i^{\text{Vol}}$  der Eingabespektren  $i$  ( $i = 2\text{-D-NOESY}$  bzw.  $3\text{D } ^{15}\text{N-NOESY-HSQC}$ ) aufgelistet. Im Falle des 2-D-NOESY-Spektrums war der statistisch ermittelte Volumenskalierungsfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{statistisch}}$  zu klein. Mit ihm war Pseudoenergie des Volumens  $E_{2\text{-D-NOESY}}^{\text{Vol}}$  der vollständigen richtigen Zuordnung auf dem 2-D-NOESY-Spektrum um 7 % höher als mit dem Referenzskalierungsfaktor  $\bar{\alpha}_{2\text{-D-NOESY}}^{\text{Ref}}$ .

	2-D-NOESY	3D <sup>15</sup> N-NOESY-HSQC
$\alpha_i^{\text{statistisch}}$	0,0058 (- 6,5 %)	0,25 (-66,2 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	805 (+2,9 %)	270 (-19,6 %)
Anzahl der Signale	5563	1661
$\bar{\alpha}_i^{\text{Ref}}$	0,01 (+61,3 %)	0,74
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	752 (-3,8 %)	336
Anzahl der Signale	2814	783
$\alpha_i^{\text{Ring}}$	0,0062	
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$	782	
Anzahl der Signale	144	
Verbreiterung $\omega_1$ [Hz]	0	7,3
Verbreiterung $\omega_2$ [Hz]	0	0
Verbreiterung $\omega_3$ [Hz]		0
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ ohne Verbreiterung		327
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$ mit Verbreiterung		336 (+2,7 %)
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ ohne Verbreiterung		442
$\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$ mit Verbreiterung		447 (+ 1,1 %)

**Tabelle 13: Auf verschiedene Arten bestimmte Volumenskalierungsfaktoren und statistisch bestimmte Linienverbreiterungen und ihre Auswirkung auf die lokalen Pseudoenergien der richtigen Zuordnung von Ras(T35S)-GppNHp.** Es sind die statistisch bestimmten ( $\alpha_i^{\text{statistisch}}$ ), die mit allen eindeutig zuordenbaren Signalen bestimmten ( $\bar{\alpha}_i^{\text{Ref}}$ ) und nur mit eindeutigen Ringkontakten bestimmten Skalierungsfaktoren ( $\alpha_i^{\text{Ring}}$ ) angegeben. Außerdem sind die Pseudoenergie des Volumens und die Anzahl der experimentellen Signale, die bei der Bildung des Skalierungsfaktor beteiligt waren, aufgelistet.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Vol}}$  bzw.  $\sum_{j=1}^{N_i^{\text{Signale}}} E_{ij}^{\text{Shape}}$  sind die über alle experimentelle Signale j summierten Pseudoenergien des Volumens bzw. der Linienform im Spektrum i (i = 2-D-NOESY bzw. 3D <sup>15</sup>N-NOESY-HSQC).  $\omega_1$ ,  $\omega_2$  und  $\omega_3$  sind die Dimensionen der Spektren, in denen die Signale der Rückrechnung verbreitert wurden. In Klammern sind die relativen Unterschiede der Skalierungsfaktoren und Pseudoenergien zu der Referenz bzw. der nicht verbreiterten Rückrechnung angegeben.

Im 3D <sup>15</sup>N-NOESY-HSQC-Spektrum ist die Pseudoenergie des Volumens  $E_{3\text{-D-NOESY}}^{\text{Vol}}$  mit dem zu klein bestimmten statistischen Skalierungsfaktor  $\alpha_{3\text{-D-NOESY}}^{\text{statistisch}}$  um 20 % besser als mit dem Referenzskalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$ .



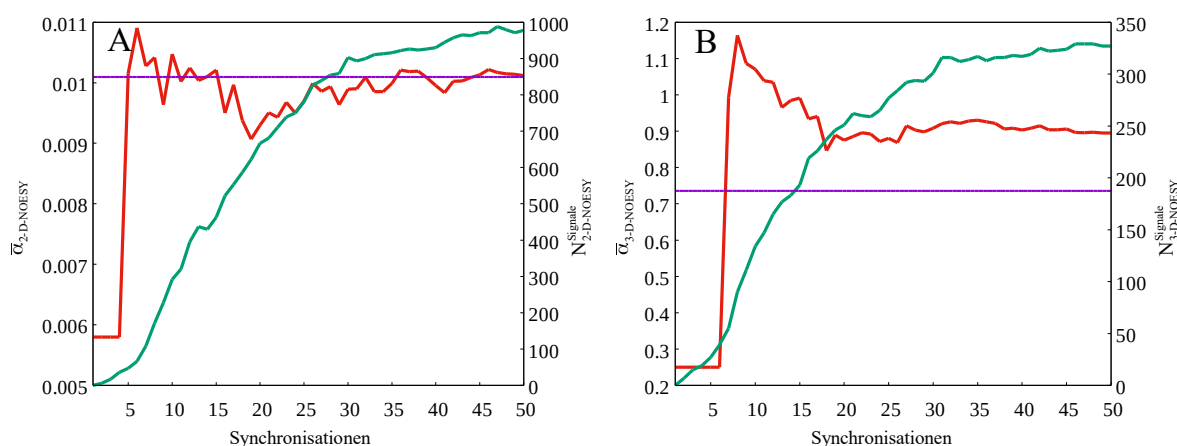


Abbildung 60: **Entwicklung der Volumenskalierungsfaktoren  $\bar{\alpha}_i$  (rot) während der automatischen Zuordnung von Ras(T35S)-GppNHp.** In grün (y2-Achse) ist die Anzahl der eindeutig zuordenbaren Signale  $N_i^{Signale}$  aufgetragen. In lila sind die Volumenskalierungsfaktoren  $\bar{\alpha}_i^{Ref}$ , die mit den eindeutig zuordenbaren Kreuzsignalen bestimmbar waren, eingezeichnet. **A** 2-D-NOESY-Spektrum, **B** 3D  $^{15}N$ -NOESY-HSQC-Spektrum.

In Abb. 60 ist die zeitliche Anpassung der verwendeten Volumenskalierungsfaktoren während der automatischen Zuordnung aufgetragen (Kap. 2.2.14). Durch das Fixieren der chemischen Verschiebungen von Resonanzen nähert sich der Skalierungsfaktor des 2-D-NOESY-Spektrums (**A rot**) zur Laufzeit dem Referenzskalierungsfaktor  $\bar{\alpha}_{2-D,NOESY}^{Ref}$  (**A lila**) an und erreicht diesen am Ende fast perfekt. Das gleiche gilt für den Skalierungsfaktor des 3D  $^{15}N$ -NOESY-HSQC-Spektrums, dessen Fehler am Ende der automatischen Zuordnung auf 19 % reduziert werden konnte.

## Häufigkeitsverteilungen der Pseudoenergien der richtigen Zuordnung

Mit der richtigen vollständigen Zuordnung können die Häufigkeitsverteilungen der lokalen Pseudoenergien  $E^{Vol}$  und  $E^{Shape}$  des erwarteten Ergebnisses der automatischen Zuordnung von Ras(T35S)-GppNHp bestimmt werden. Abb. 61 zeigt in **rot** die Häufigkeitsverteilungen der lokalen Pseudoenergien des Volumens  $E^{Vol}$  und in **grün** die Häufigkeitsverteilungen der lokalen Pseudoenergien des Linienformvergleichs  $E^{Shape}$  der beiden NOESY-Spektren auf den experimentellen und auf den simulierten Grundflächen  $ij$ . Die vollständige richtige Zuordnung zeichnet sich auch bei der automatischen Zuordnung von Ras(T35S)-GppNHp durch eine sehr großen Anteil von hohen Pseudoenergien aus.

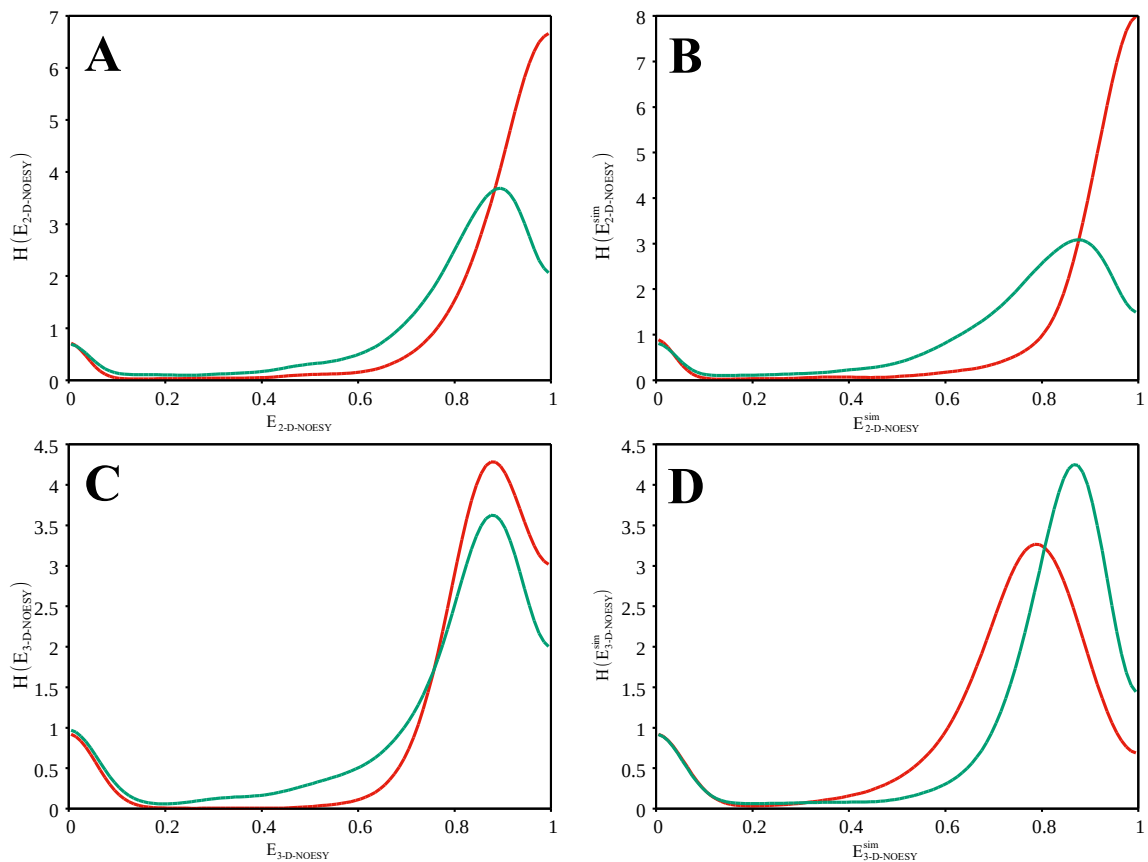
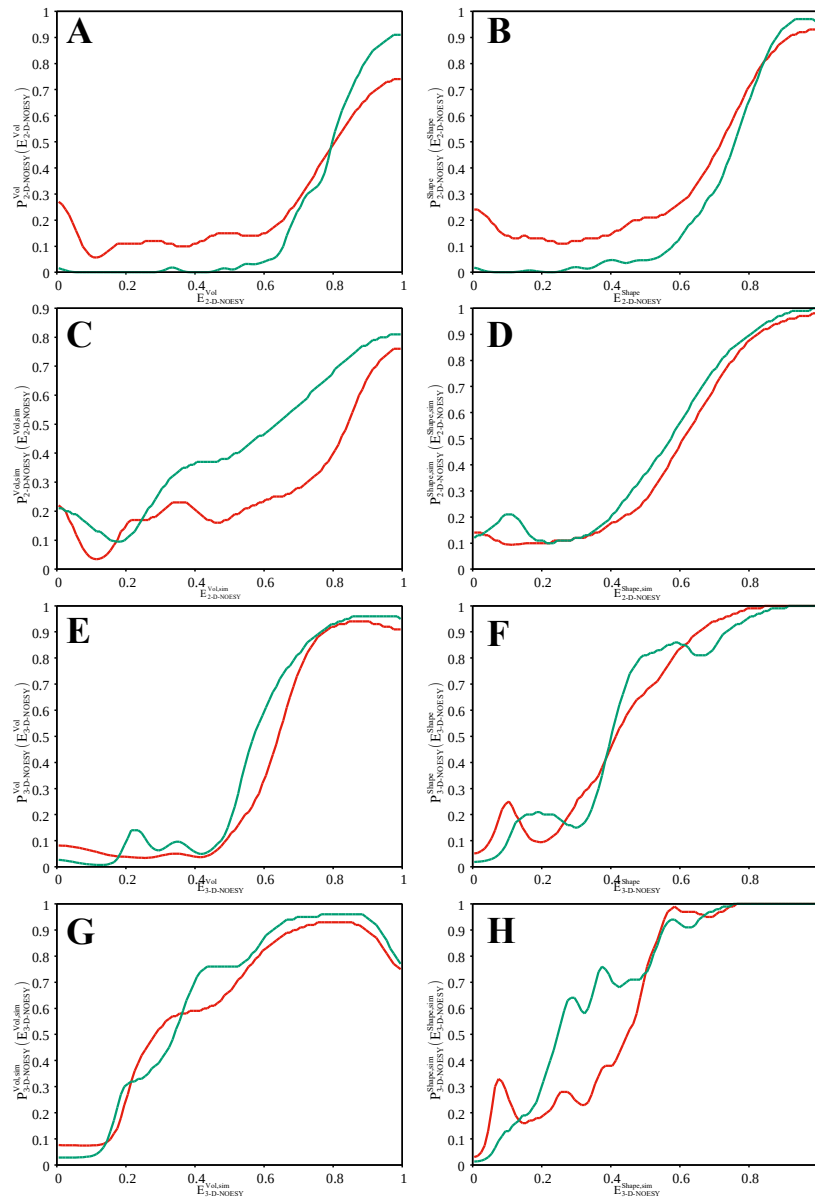


Abbildung 61: **Häufigkeitsverteilungen  $H(E)$  der lokalen Pseudoenergien des Volumens  $E^{\text{Vol}}$  (rot) und der Linienform  $E^{\text{Shape}}$  (grün) der richtigen vollständigen Zuordnung beim Ras(T35S)-GppNHp.** **A** Auf den Grundflächen der experimentellen Signale im 2-D-NOESY-Spektrum, **B** auf den Grundflächen der simulierten Signale im 2-D-NOESY-Spektrum, **C** auf den Grundflächen der experimentellen Signale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum, **D** auf den Grundflächen der simulierten Signale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und mit einer zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalingfaktor  $\alpha_{2\text{-D-NOESY}}^{\text{Ring}}$ , für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum der Skalierungsfaktor  $\bar{\alpha}_{3\text{-D-NOESY}}^{\text{Ref}}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.

Abb. 62 zeigt die Graphen der acht Bayesschen Wahrscheinlichkeiten einer lokalen richtigen Zuordnung in Abhängigkeit von den beiden Pseudoenergien  $E_{ij}^{\text{Vol}}$  und  $E_{ij}^{\text{Shape}}$  in beiden NOESY-Spektren auf jeweils experimentellen und den simulierten Grundflächen. In **grün** sind die Abhängigkeiten der Wahrscheinlichkeiten von der Pseudoenergie, die SIBASA während der automatischen Zuordnung verwendet und mithilfe des Datensatzes eines bereits gelösten Proteins erstellt wurden (Kap. 4.3.1), eingezeichnet. Die Wahrscheinlichkeitskurven (Abb. 62 **rot**) stimmen gut mit der Referenz (**grün**) überein, obwohl Ras(T35S)-GppNHp doppelt so groß wie das Protein, mit dem die Referenzverteilungen erzeugt worden sind, ist. Die größten Unterschiede sind bei den Wahrscheinlichkeiten der richtigen Zuordnung der experimentellen Signale im Bereich niedriger Pseudoenergien zu sehen. Für die Bildung der Referenzverteilungen der Pseudoenergien wurden nämlich nicht zuordenbaren

experimentellen Signale, die in der Regel Pseudoenergien von 0 haben, ausgeschlossen (Kap. 4.3.1).



**Abbildung 62: Abhängigkeit der Bayesschen Wahrscheinlichkeit der lokalen richtigen Zuordnung  $P_i(E)$  von den lokalen Pseudoenergien  $E_i$  (rot) für die NOESY-Spektren des Ras(T35S)-GppNHp. In grün sind die zugehörigen Wahrscheinlichkeitsfunktionen, die SIBASA während der automatischen Zuordnung verwendet (Kap. 4.1.3), abgebildet. **A**  $P_{2-D-NOESY}^{Vol}(E_{2-D-NOESY}^{Vol})$ , **B**  $P_{2-D-NOESY}^{Shape}(E_{2-D-NOESY}^{Shape})$ , **C**  $P_{2-D-NOESY}^{Vol,sim}(E_{2-D-NOESY}^{Vol,sim})$  und **D**  $P_{2-D-NOESY}^{Shape,sim}(E_{2-D-NOESY}^{Shape,sim})$  des 2-D-NOESY-Spektrums. **E**  $P_{3-D-NOESY}^{Vol}(E_{3-D-NOESY}^{Vol})$ , **F**  $P_{3-D-NOESY}^{Shape}(E_{3-D-NOESY}^{Shape})$ , **G**  $P_{3-D-NOESY}^{Vol,sim}(E_{3-D-NOESY}^{Vol,sim})$  und **H**  $P_{3-D-NOESY}^{Shape,sim}(E_{3-D-NOESY}^{Shape,sim})$  des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums. Die zurückgerechneten Spektren wurden aus dem MD-Ensemble und mit einer zufälligen Zuordnung erzeugt. Für das 2-D-NOESY-Spektrum wurde der Volumenskalierungsfaktor  $\alpha_{2-D-NOESY}^{Ring}$ , für das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum der Skalierungsfaktor  $\bar{\alpha}_{3-D-NOESY}^{Ref}$  verwendet. Die zusätzlichen Linienverbreiterung wurden von SIBASA statistisch bestimmt.**

## Verlauf von Pseudoenergien und dynamischer Größen während der automatischen Zuordnung

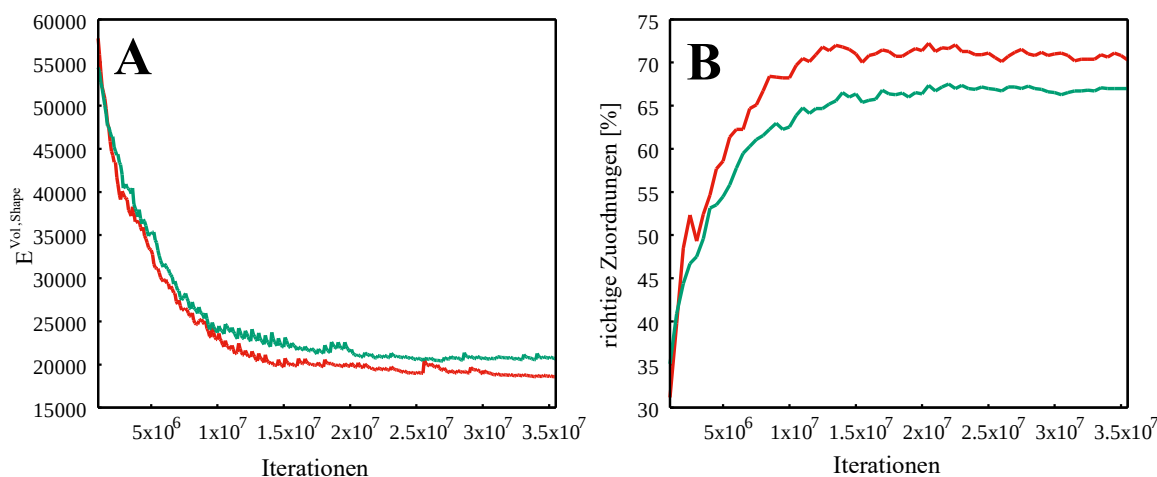


Abbildung 63: **Zeitlicher Verlauf der gesamten Pseudoenergie des Signalvergleichs und des Anteil der richtig identifizierten Zuordnungen bei der automatischen Zuordnung von Ras(T35S)-GppNHp.** In rot ist die Instanz der Optimierung mit der besten und in grün die Instanz mit der schlechtesten finalen Pseudoenergie eingezeichnet. **A** Verlauf der Pseudoenergie des Volumen- und Linienformvergleichs auf den Grundflächen der experimentellen Signale  $E^{\text{Vol,Shape}}$ , **B** Verlauf der gefundenen richtigen Zuordnungen im Fehlerbereich  $\delta^{\text{Ref}} \pm 0,03$  ppm.

Die automatische Zuordnung wird durch die Minimierung der Pseudoenergie des Volumen- und Linienformvergleichs auf den Grundflächen der experimentellen Signale  $E^{\text{Vol,Shape}}$  zur richtigen Zuordnung geführt. Abb. 63 **A** zeigt den zeitlichen Verlauf der Pseudoenergie für die am schnellsten (**rot**) und am langsamsten minimierende (**grün**) SIBASA Instanz. Man kann erkennen, dass auch im Falle des Ras(T35S)-GppNHp mithilfe mehrerer Instanzen die minimale Pseudoenergie besser erreicht werden kann, als mit einer einzigen Instanz. Der Unterschied zwischen den beiden finalen Pseudoenergien beträgt 9 % in Abb. 63 **A**. Dies wirkt sich auch wieder auf die Anzahl der richtig gefundenen Zuordnungen (Abb. 63 **B**) in diesen beiden Instanzen aus. Die beste Instanz identifiziert 70,3 %, die schlechteste Instanz 67 % der chemischen Verschiebungen von Resonanzen im Fehlerbereich von  $\delta^{\text{Ref}} \pm 0,03$  ppm richtig. In Abb. 64 sind zeitliche Verläufe von Optimierungsgrößen der besten Instanz von SIBASA aufgetragen. Die Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  (Abb. 64 **A rot**) kann den geschätzten Wert der richtigen Zuordnung  $E^{\text{Vol,Shape,opt}}$  (Abb. 64 **A orange**) (Gl. 64) nicht erreichen (+40 %), doch nähert sie sich nahe genug an, dass der dynamische Gewichtungsfaktor der Pseudoenergie der chemischen Verschiebung  $w^{\text{E},\omega}(t)$  (Abb. 64 **B lila**) (Gl. 64) und der Gewichtungsfaktor der Pseudoenergie des Dummyslots  $w^{\text{E,Dummy}}(t)$  (Gl. 74) am Ende der automatischen Zuordnung nahe an ihrem minimalen Wert lagen.

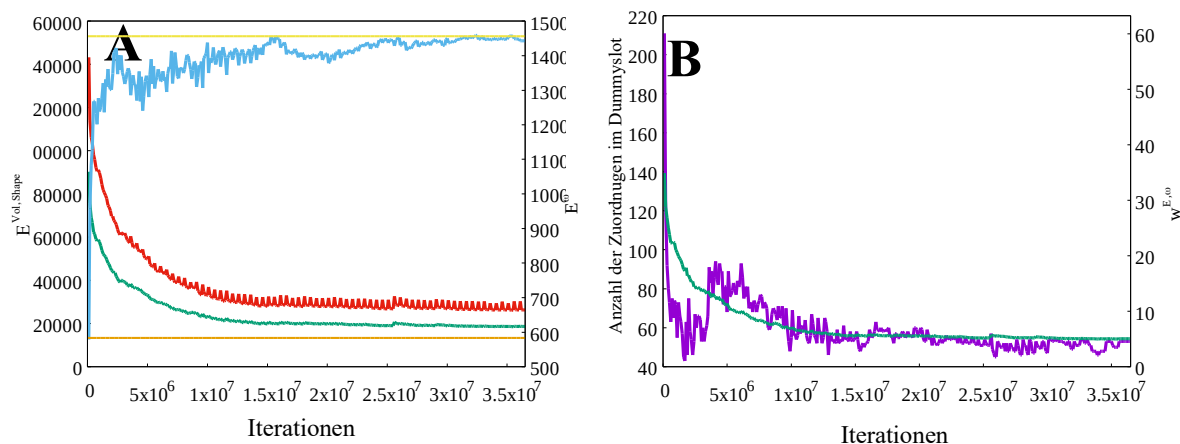


Abbildung 64: **Zeitliche Verläufe wichtiger Optimierungsgrößen einer Instanz der automatischen Zuordnung von Ras(T35S)-GppNHp.** **A** Verlauf der gesamten Pseudoenergie der Optimierung  $E^{\text{total}}$  (rot, y-Achse), Verlauf der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape}}$  (grün, y-Achse) und Verlauf der Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  (blau, y2-Achse). In orange (y-Achse) ist der geschätzte optimale Wert der Pseudoenergie des Signalvergleichs  $E^{\text{Vol,Shape,opt}}$  und in gelb (y2-Achse) die Pseudoenergie der chemischen Verschiebung der richtigen vollständigen Zuordnung eingetragen. **B** Anzahl der Mitglieder von Amidgruppen im Dummieslot (lila, y-Achse) und Verlauf des dynamischen Gewichtungsfaktors der Pseudoenergie der chemischen Verschiebung  $w^{E,\omega}$  (grün, y2-Achse), dessen Trend auch auf den Gewichtungsfaktor der Pseudoenergie des Dummieslots  $w^{E,\text{Dummy}}$  übertragen werden kann.

Durch den dynamischen Gewichtungsfaktor  $w^{E,\omega}(t)$  nähert sich die Pseudoenergie der chemischen Verschiebung  $E^{\omega}$  (blau) (Gl. 62) bis auf 1 % an die Pseudoenergie der chemischen Verschiebung der vollständigen richtigen Zuordnung an (gelb). Die Anzahl der Amidprotonen und Stickstoffe im Dummieslot ist am Ende mit 53 höher als die zu erwartende Anzahl von 36 (12 im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum nicht sichtbare Resonanzen Amidgruppen und 11 Resonanzen von Arg  $\text{N}^{\epsilon}$ , die außerhalb der spektralen Bereiche liegen).

## Festlegung von Zuordnungen während der Laufzeit

Es wurden wieder 16 Instanzen von SIBASA mit unterschiedlichen Konfigurationen und Zufallszahlen gestartet. Nach jeweils fünfmaligen Abkühlen (500.000 Iterationen) wurden diese synchronisiert. Während der Synchronisationsphasen wurden die vereinigten Zuordnungslisten mit der maximale gewichten Standardabweichung eines Clusters eines Kerns  $k$  von  $\sigma_k^w = 0,03\text{ppm}$  gebildet und versucht chemische Verschiebungen von Resonanzen zu fixieren (Kap. 2.2.13). Abb. 65 A zeigt den Anteil der fixierten (blau) und korrekt fixierten Zuordnungen (rot) in Abhängigkeit vom Zeitpunkt in der Optimierung an. Am Ende sind 40,3 % der Zuordnungen zur Laufzeit festgelegt worden. Der Großteil davon konnte bereits relativ früh festgelegt werden, wodurch im weiteren Verlauf der automatischen Zuordnung mehr Rechenzeit auf die schwieriger zuzuordnen Kerne verteilt werden konnte.

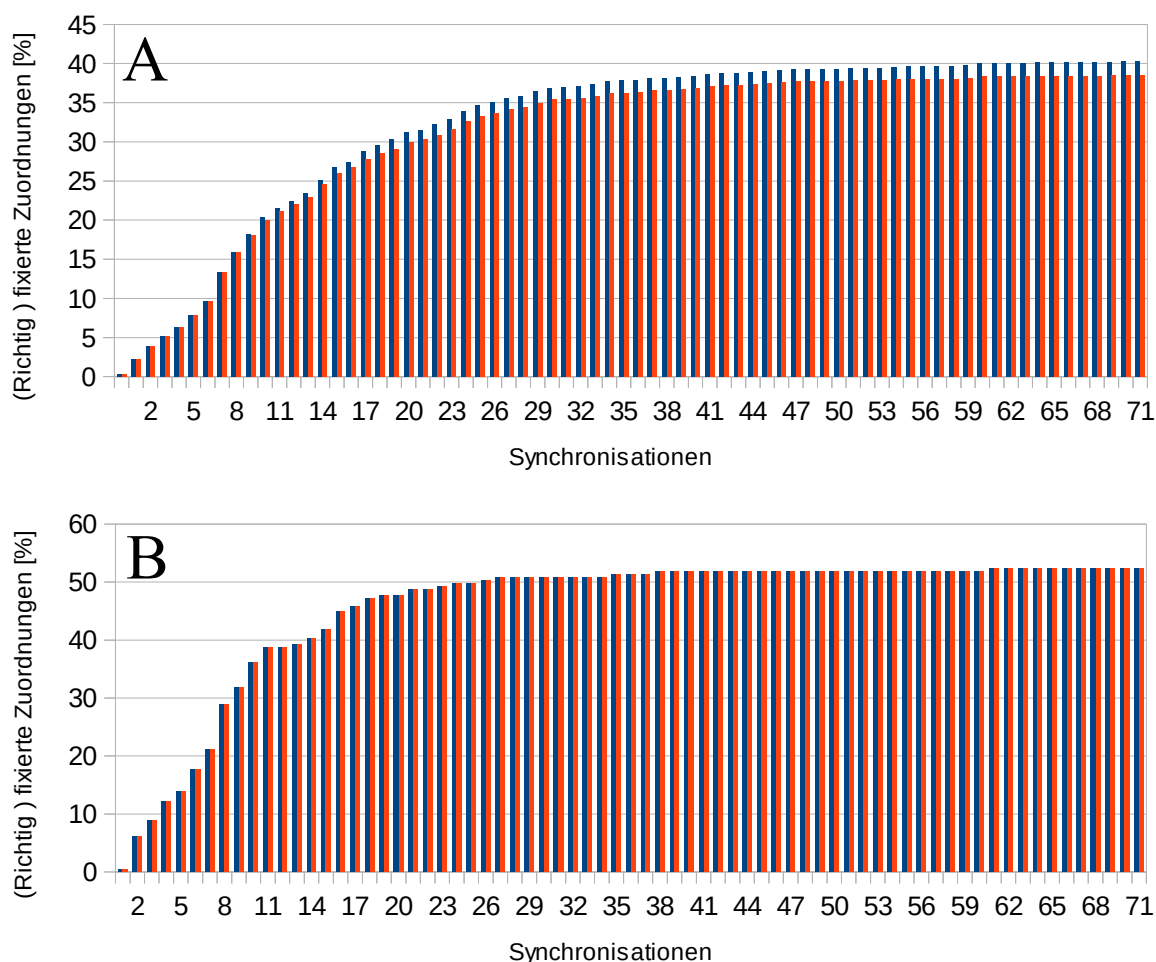


Abbildung 65: Anteil der zur Laufzeit fixierte Zuordnungen (blau) und Anteil der zur Laufzeit richtig fixierten Zuordnungen (rot) für die automatische Zuordnung von Ras(T35S)-GppNHp. **A** Alle Zuordnungen von Resonanzen, **B** Nur Zuordnungen Resonanzen von Amidprotonen und Stickstoffen.

Insgesamt sind ein Großteil der zur Laufzeit fixierten Zuordnungen auch richtig festgelegt worden. Lediglich 20 der 468 Zuordnungen (4,3 %) wurden im Fehlerbereich von  $\delta^{\text{Ref}} \pm 0,03$  ppm falsch festgelegt. Würde der maximale Fehler für die chemischen Verschiebungen der Resonanzen von H <sup>$\alpha$</sup> - und Seitenkettenprotonen auf 0,06 ppm gesetzt, welches dem Bereich entspricht, indem die chemischen Verschiebungen der Resonanzen der Referenz optimiert worden sind (Kap. 3.4.5), so wären nur 12 fixierte chemische von Resonanzen Verschiebungen falsch zugeordnet worden (2,6 %).

Abb. 65 **B** zeigt den Anteil der fixierten (**blau**) und korrekt fixierten (**rot**) Zuordnungen von Amidgruppen an. Auch bei den Amidgruppen steigt der Anteil der fixierten Resonanzen zu Beginn stark an und konvergiert gegen 54 %. Wieder sind alle fixierten Zuordnungen von Amidgruppen auch richtig.

Mit den zur Laufzeit fixierten Resonanzzuordnungen konnten experimentelle NOESY-

Kreuzsignale eindeutig zugeordnet werden, wodurch es möglich war die Volumenskalierungsfaktoren  $\bar{\alpha}_i$  (Kap. 2.2.14) zur Laufzeit der automatischen Zuordnung zu bestimmen. Abb. 60 zeigt in **grün** die eindeutig zuordenbaren Signale für beide NOESY-Spektren. Im 2-D-NOESY-Spektrum konnten am Ende 931 (16 % aller experimentellen Signale) und im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum konnten 311 (18 % aller experimentellen Signale) Signale eindeutig zugeordnet werden.

## Auswertung und Verifikation der vereinigten Zuordnungsliste

Nach 72 Synchronisationen der automatische Zuordnung von Ras(T35S)-GppNHp hat sich die Pseudoenergie der Optimierung  $E^{\text{total}}$  nicht mehr entscheiden verbessert und es wurden keine weiteren Zuordnungen mehr zur Laufzeit fixiert. Die finale vereinigte Zuordnungsliste wurde anschließend wieder mit einer maximalen gewichteten Standardabweichung  $\sigma_k^w(\text{Cluster})$  von 0,03 ppm erstellt (Kap. 2.2.13). In ihr sind insgesamt 850 der 1095 (77,6 %) chemischen Verschiebungen von Resonanzen im Fehlerbereich von  $\delta^{\text{Ref}} \pm 0,03$  ppm richtig zugeordnet worden. Von den 345 Resonanzen von Amidprotonen und Stickstoffen, die sichtbare Kreuzsignale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum hatten, sind 317 (91,8 %) richtig zugeordnet worden. Von den chemischen Verschiebungen der Resonanzen der Hauptkettenamidgruppen wurden 282 der 300 richtig zugeordnet (94 %). Von den 12 Amidgruppen, die keine sichtbaren NOESY-Kreuzsignale hatten, konnte SIBASA nur Ser35 auf den dafür vorgesehenen Dummymslot in allen 16 Instanzen setzen können. Die Zuordnung einer Amidgruppe, die keine sichtbare Kreuzsignale hatte, hat von SIBASA maximal eine Verifikationspseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  von 0,189 (Glu62) erhalten, weshalb diese Zuordnungen leicht herausgefiltert werden konnten (Abb. 66 C).

### Häufigkeitsverteilung der Verifikationspseudoenergie und Wahrscheinlichkeit der richtigen Zuordnung

Abb. 66 A zeigt die Häufigkeitsverteilungen der Verifikationspseudoenergien  $E_k^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  (Gl. 81) der richtigen (**rot**) und falschen Zuordnungen (**grün**). Abb. 66 C zeigt die gleichen Verteilungen für die Verifikationspseudoenergien  $E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  der Amidgruppen (Gl. 85). In beiden Abbildungen ist zu erkennen, dass nach der automatischen Zuordnungen die richtig identifizierten chemischen Verschiebungen im Ras(T35S)-GppNHp hauptsächlich hohe Verifikationspseudoenergien und falsche Zuordnungen vor allem niedrige Verifikationspseudoenergien haben.

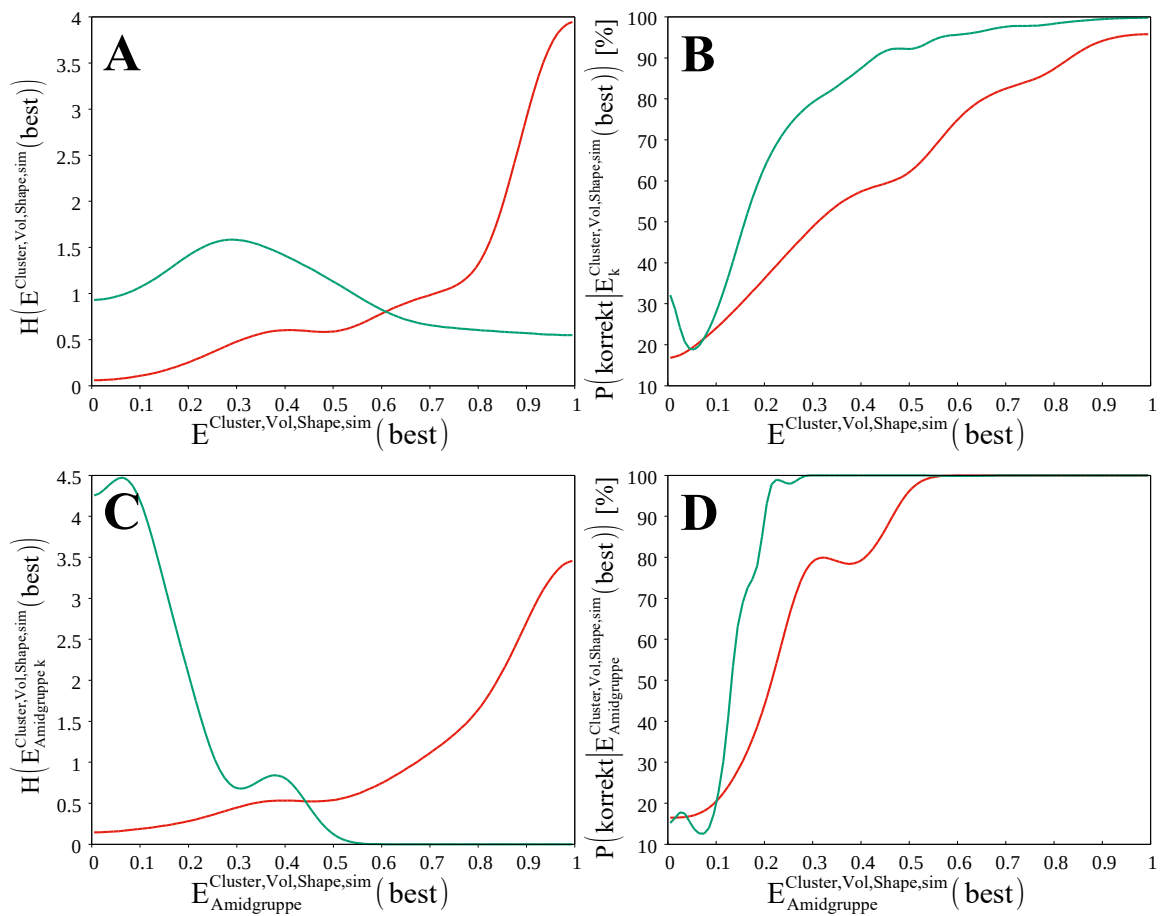


Abbildung 66: **Häufigkeitsverteilungen der Verifikationspseudoenergien  $E^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  und Wahrscheinlichkeiten der richtigen Zuordnung nach der automatischen Zuordnung von Ras(T35S)-GppNHp.** **A** Häufigkeitsverteilung  $H(E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der Verifikationspseudoenergien der richtigen (**rot**) und der falschen Zuordnungen (**grün**), **B** Bayessche Wahrscheinlichkeit einer richtigen Zuordnung  $P(\text{korrekt}|E^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  einer Resonanz (**rot**) in Abhängigkeit von der Verifikationspseudoenergie und zugehörige Wahrscheinlichkeitskurve, die mit Hilfe von bereits gelösten Proteinen erzeugt und in die AUREMOL Datenbank gespeichert wurde (**grün**) (Abb. 28). **C** Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der richtigen (**rot**) und der falschen Zuordnungen (**grün**) von Amidgruppen, **D** Bayessche Wahrscheinlichkeit der richtigen Zuordnung  $P(\text{korrekt}|E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  einer Amidgruppe (**rot**) in Abhängigkeit von der Verifikationspseudoenergie und zugehörige Wahrscheinlichkeitskurve aus der AUREMOL Datenbank (**grün**) (Abb. 28).

Dies macht es möglich die falschen von den richtigen chemische Verschiebungen im Ergebnis der automatischen Zuordnung von Ras(T35S)-GppNHp zu unterscheiden. Die zugehörigen Bayesschen Wahrscheinlichkeiten einer richtigen Zuordnung (Abb. 66 **B** und **D**) zeigen einen nahezu linearen Zusammenhang mit den Verifikationspseudoenergien. Die höhere Komplexität der automatischen Zuordnung von Ras(T35S)-GppNHp sorgt aber dafür, dass diese Wahrscheinlichkeitskurven im Gegensatz zu den Kurven, die mit Hilfe von bereits gelösten Proteinen erstellt worden sind und SIBASA als Grundlage für die Bewertung von



Zuordnung dienen sollen (**grün**) (Abb. 28), weniger steil verlaufen. Für die automatischen Zuordnung von Ras(T35S)-GppNHp sind deshalb höhere Pseudoenergien für sichere Zuordnungen nötig als für die anderen beiden Testfällen. Für die in Kap. 4.3.1 festgelegten minimalen Grenzwerte  $p = 0,85$  bzw.  $p_{\text{Amidgruppe}} = 0,73$  der Verifikationspseudoenergien ist die reale Wahrscheinlichkeit einer falschen Zuordnung 9 % für die Resonanzen von H $^{\alpha}$ - und Seitenkettenprotonen und 0 % für die Resonanzen von Amidgruppen.

### Die Güte der Verifikationspseudoenergie als Separator zwischen falschen und richtigen Zuordnungen

Die Güte der Verifikationspseudoenergie als Separator zwischen falschen und richtigen Zuordnungen wurde mit der Fläche unter den entsprechenden ROC-Kurven gemessen. Am Ende der automatischen Zuordnung von Ras(T35S)-GppNHp betrug sie für alle Resonanzen zusammen 82,1 % und nur für die Resonanzen von Amidgruppen betrug sie 96,7 %. Die Flächen waren damit ungefähr genauso groß wie die entsprechenden Flächen beim halb so großen HPr *S. aureus* (H15A) (83 % bzw. 100 %) (Kap. 4.3.2). Die modifizierten ROC-Kurven, die auch die Qualität des Zuordnungsalgorithmus mit einbeziehen, sind in Abb. 67 **A** und **C** zu sehen. Hat man 5 % falsche Zuordnungen in Kauf genommen, so betrug der Anteil der bestätigten richtigen Zuordnungen 73,4 %. 11,6 % der Zuordnungen blieben dabei undefiniert. Erlaubt man keine falsche Zuordnung von Amidgruppen, so konnten immer noch 78,3 % der Zuordnungen von Amidprotonen und Stickstoffe, die sichtbare NOESY-Kreuzsignale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum haben, als richtig betätigt werden.

Abb. 67 **B** und **D** zeigen wieder, wie gut sich der Anteil der richtigen bzw. falschen Resonanzzuordnungen im Ras(T35S)-GppNHp mit der Sicherheitswahrscheinlichkeit-Verifikationspseudoenergie-Abhängigkeit, die mithilfe des Satzes von Bayes und den Datensätzen bereits gelöster Proteine bestimmt worden sind (Kap. 2.2.15), bestimmen lässt. Beim Ras(T35S)-GppNHp übersteigt der Anteil der falschen Zuordnungen (Abb. 67 **B**, **grün**) den mit der Sicherheitswahrscheinlichkeit tolerierten Anteil (Abb. 67 **B**, **lila** Winkelhalbierende). Im schlechtesten Fall erhält der Benutzer 2 % mehr falsche Zuordnungen als er erlaubt hat. Er würde insgesamt 4,5 % falsche Zuordnungen bei einer gewählten Sicherheitswahrscheinlichkeit von 97,5 % erhalten. Würde er die Sicherheitswahrscheinlichkeit auf 95 % (signifikantes Niveau) setzen, so würde seine gefilterte Zuordnungsliste 60 % richtige und 6,8 % falsche Resonanzzuordnungen enthalten. 33,2 % der Zuordnungen würden dabei undefiniert bleiben.

Für die Zuordnungen von Amidgruppen konnte die Sicherheitswahrscheinlichkeit bis 96,6 % eingehalten werden. Darüber muss der Benutzer mit maximal 1,8 % (bei 99 % Sicherheitswahrscheinlichkeit) mehr falschen Zuordnungen als beabsichtigt rechnen.

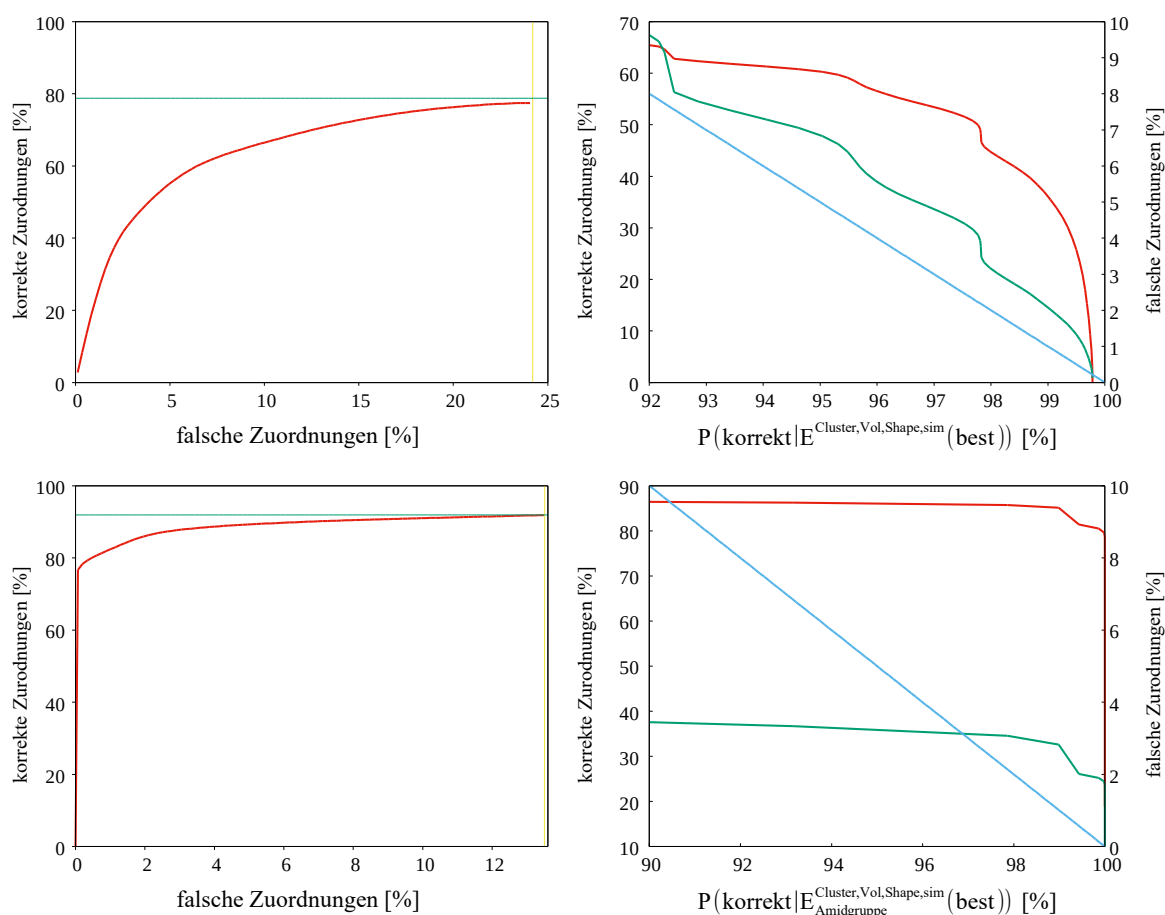


Abbildung 67: **Auswirkung eines Grenzwertes der Verifikationspseudoenergie  $E^{\text{Cluster, Vol, Shape, sim}}(\text{best})$  auf die Klassifizierung der Resonanzzuordnungen in der vereinten Liste der chemischen Verschiebung von Ras(T35S)-GppNHp.** **A** Modifizierte ROC-Kurve der Verifikationspseudoenergie. Es sind die möglichen richtigen Zuordnungen gegen die dafür in Kauf zu nehmenden falschen Zuordnungen aufgetragen (**rot**). Insgesamt sind 77,4 % der im  $3D^{15}\text{N-NOESY-HSQC}$ -Spektrum sichtbaren Resonanzen richtig (**grün**) und 24 % aller Resonanzen falsch zugeordnet worden (**gelb**). **B** Gefundene richtige (**rot**, y-Achse) und falsche Zuordnungen (**grün**, y2-Achse) in Abhängigkeit von der von Sicherheitswahrscheinlichkeit  $P(\text{korrekt} | E^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$ . Der vom Benutzer tolerierten Anteil von falschen Zuordnungen ist in **blau** (Winkelhalbierende, y2-Achse) eingetragen. **C** Modifizierte ROC-Kurve der Verifikationspseudoenergien von Amidgruppen  $E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best})$ . Insgesamt sind 91,9 % der im  $3D^{15}\text{N-NOESY-HSQC}$ -Spektrum sichtbaren Resonanzen von Amidgruppen richtig (**grün**) und 13,5 % aller Resonanzen von Amidgruppen falsch (**gelb**) zugeordnet worden. **D** Gefundene richtige (**rot**, y-Achse) und falsche Zuordnungen von Amidgruppen (**grün**, y2-Achse) in Abhängigkeit von der Sicherheitswahrscheinlichkeit  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$ . Der vom Benutzer tolerierten Anteil ist in **blau** (Winkelhalbierende, y2-Achse) eingetragen.

### Abschätzung der Wahrscheinlichkeiten der richtigen Zuordnung von HSQC-Signalen

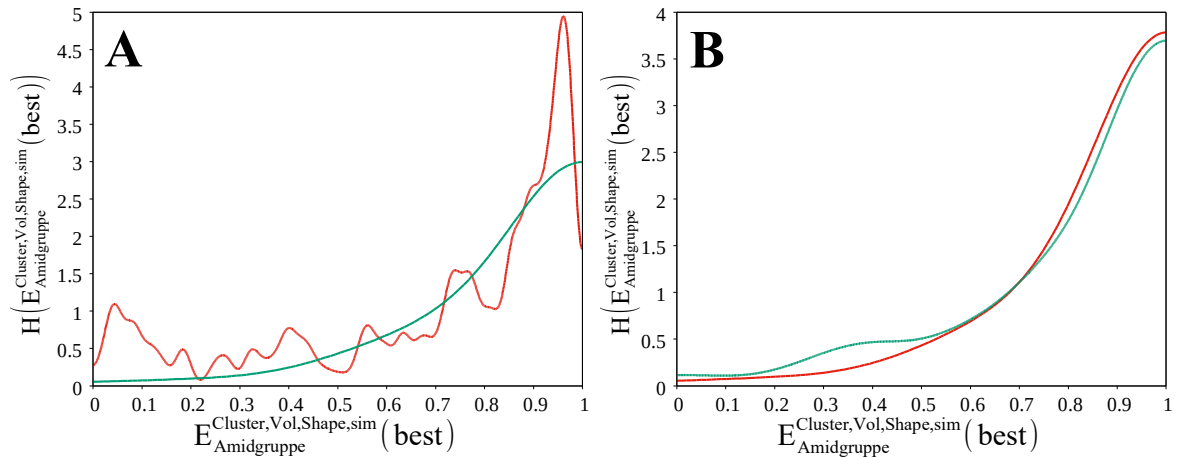


Abbildung 68: **Schätzung der Häufigkeitsverteilung der Verifikationspseudoenergien der richtig zugeordneten Resonanzen von Hauptkettenamidgruppen beim Ras(T35S)-GppNHp.** **A** Wenig geglättete (**rot**) und mit  $H^{\text{korrekt}}(x)$  (Gl. 92) gefittete Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  (**grün**) der Verifikationspseudoenergien der Zuordnungen von Hauptkettenamidgruppen. **B** Vergleich zwischen der geglätteten (**rot**) und der realen (**grün**) (Abb. 66 C) Häufigkeitsverteilung  $H(E_{\text{Amidgruppe}}^{\text{Cluster,Vol,Shape,sim}}(\text{best}))$  der Verifikationspseudoenergien der richtigen Zuordnungen von Hauptkettenamidgruppen.

Um den Grenzwert  $p_{\text{Amidgruppe}}$  für sichere HSQC-Signaluordnungen zu bestimmen, wird wieder die wenig geglättete Häufigkeitsverteilung der Verifikationspseudoenergien  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  der Hauptkettenamidgruppen des Ras(T35S)-GppNHp nach der automatischen Zuordnung betrachtet (Abb. 68 A **rot**). Bei einer Verifikationspseudoenergie von 0,53 ist ein auffallendes Minimum, das wie schon in den anderen beiden Testfällen als untere Grenze für den Fit von  $H^{\text{korrekt}}(x)$  (Gl. 92) an die stärker geglättete Häufigkeitsverteilung der Verifikationspseudoenergie diente.

Abb. 68 B zeigt das Ergebnis des Fits (**grün**) und zum Vergleich auch die reale Häufigkeitsverteilung der Verifikationspseudoenergien der richtigen Zuordnungen (**rot**) (Abb. 66 C **rot**). Man erkennt das der Verlauf der Verteilung gut abgeschätzt werden kann. Im kritischen Bereich um die Verifikationspseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best}) = 0,4$ , wo die wirkliche Grenze zwischen teilweise falschen und nur richtigen Zuordnungen liegt, unterschätzt die mit dem Fit bestimmte Häufigkeitsverteilung (**rot**) die reale Häufigkeitsverteilung (**grün**).

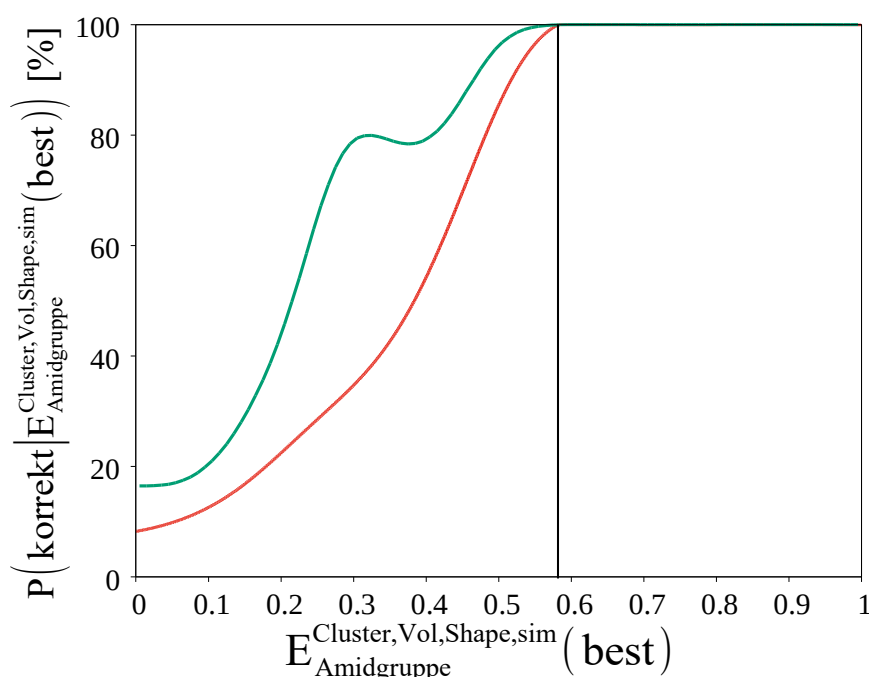


Abbildung 69: Mithilfe des Fits von Abb. 68 geschätzte Bayessche Wahrscheinlichkeit der richtigen Zuordnung eines Amidprotons oder Stickstoffs  $P(\text{korrekt} | E_{\text{Amidgruppe}}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  in Abhängigkeit von der Verifikationspseudoenergie (rot). In grün ist der reale Wahrscheinlichkeitsgraph eingezeichnet. In schwarz ist der bestimmte minimale Grenzwert der Verifikationspseudoenergie  $p_{\text{Amidgruppe}} = 0,58$  für sichere Zuordnungen eingetragen.

Abb. 69 zeigt die mit dem Fit der Häufigkeitsverteilung und Gl. 93 bestimmte Wahrscheinlichkeitskurve  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  (rot) und zum Vergleich die reale Wahrscheinlichkeitskurve (grün), die mit der richtigen vollständigen Zuordnung und dem Satz von Bayes berechnet werden konnte (Abb. 66 D rot). Der Verlauf der Wahrscheinlichkeit und auch der minimale Grenzwert der Verifikationspseudoenergie für absolut sichere Zuordnungen ( $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best})) = 100 \%$ ) konnten gut bestimmt werden.

## Zuordnung des HSQC-Spektrums von Ras(T35S)-GppNHp

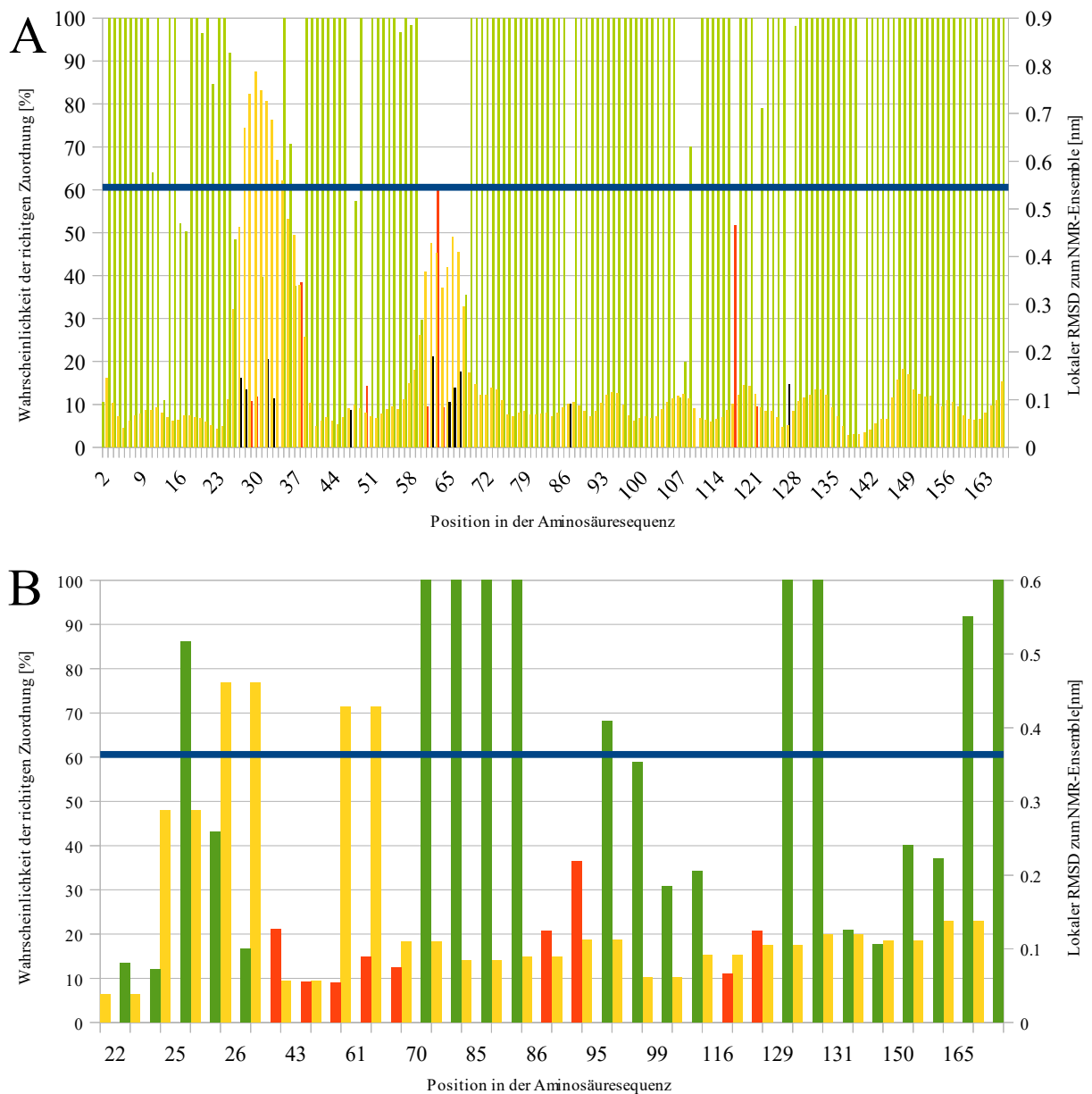
Abb. 70 A zeigt die Wahrscheinlichkeiten der richtigen Zuordnung  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der HSQC-Signale der Hauptketten in Abhängigkeit von der Position in der Aminosäuresequenz. Sie wurden mithilfe der geschätzten Wahrscheinlichkeitsfunktion aus Abb. 69 (rot) bestimmt. Die richtig zugeordneten HSQC-Signale (grüne Balken) haben größtenteils hohe Wahrscheinlichkeiten. Amidgruppen, die im

3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum keine sichtbaren Kreuzsignale haben (**schwarze** Balken) und falsche Zuordnungen von HSQC-Signale (**rote** Balken) haben in der Regel kleine Wahrscheinlichkeiten. Es ist zu erkennen, dass falsche Zuordnungen, Zuordnungen mit geringer Wahrscheinlichkeit und Hauptkettenamidgruppen, die keine Kreuzsignale im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum haben in der Primärsequenz von Ras(T35S)-GppNHp oft nebeneinander liegen.

Zusätzlich ist in **gelb** der lokale RMSD (RMSD der schweren Hauptkettenatome einer Aminosäure und seiner beiden nächsten Nachbaraminosäuren) des gemittelten MD-Ensembles zum gemittelten NMR-Ensemble (PDB ID: 2LCF) angegeben. Falsche Zuordnungen von Amidgruppen und Amidgruppen ohne sichtbare NOESY-Kreuzsignale liegen häufig im Bereich, in denen das MD-Ensemble stark vom NMR-Ensemble abweicht. Die Wahrscheinlichkeiten der richtigen Zuordnung sind in diesen Bereichen größtenteils klein, wodurch diese falschen Zuordnungen leicht herausgefiltert werden konnten.

In Abb. 70 **B** ist das gleiche Balkendiagramm für die  $\text{NH}_2$ -Gruppen von Asn und Gln abgebildet. Die Resonanzen Seitenkettenamidgruppen waren zwar alle sichtbar, aber sie wurden sehr oft falsch zugeordnet. Hohe Wahrscheinlichkeiten der richtigen Zuordnung kamen bei den Seitenketten weniger häufig als bei den Hauptketten vor. Eine Abhängigkeit der Wahrscheinlichkeit zum lokalen RMSD war für die Seitenketten nicht zu erkennen.

Um für Ras(T35S)-GppNHp nur richtig zugeordnete HSQC-Signale zu erhalten muss man den minimalen Grenzwert der Wahrscheinlichkeit  $P(\text{korrekt} | E_{\text{Amidgruppe } k}^{\text{Cluster, Vol, Shape, sim}}(\text{best}))$  der Amidgruppen  $k$  auf 100 % setzen. Abb. 71 zeigt die mit der Sicherheitswahrscheinlichkeit von 100 % zugeordnete Summenprojektion des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums auf die  $\omega_2\omega_3$ -Ebene. SIBASA konnte insgesamt 123 der 180 HSQC-Signale (68,3 %) richtig zuordnen. Von den Hauptketten waren das insgesamt 116 der 150 (77,3 %) Signale, von den Seitenketten waren es 7 der 30 Signale.



**Abbildung 70: Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale von Ras(T35S)-GppNHp in Abhängigkeit von der Position der Aminosäure in der Primärsequenz.** Die grünen Balken zeigen die Wahrscheinlichkeiten von richtig gefundenen, die roten die Wahrscheinlichkeiten von falsch gefundene Zuordnungen und die schwarzen Balken die Wahrscheinlichkeit von Zuordnungen von Amidgruppen, die keine sichtbaren Kreuzsignale in den NOESY-Spektren haben, an. In gelb ist der lokale RMSD (RMSD der Aminosäure und seiner beiden nächsten Nachbarn) der schweren Hauptkettenatome des gemittelten MD-Ensembles zum gemittelten NMR-Ensemble aufgetragen. Die dunkelblaue Linie gibt den minimalen Grenzwert der Wahrscheinlichkeit, für den keine falschen Zuordnungen mehr bestätigt werden, an.

**A** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Hauptketten.

**B** Wahrscheinlichkeiten der richtigen Zuordnung der HSQC-Signale der Seitenketten von Asn und Gln. Die Wahrscheinlichkeit des HSQC-Signals von Asn  $N^{\delta 2}/Asn H^{\delta 21}$  ist immer links von der des Signals von Asn  $N^{\delta 2}/Asn H^{\delta 22}$  und die Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 2}/Gln H^{\epsilon 21}$  ist immer links von der Wahrscheinlichkeit des Signals von Gln  $N^{\epsilon 21}/Gln H^{\epsilon 22}$  aufgetragen.

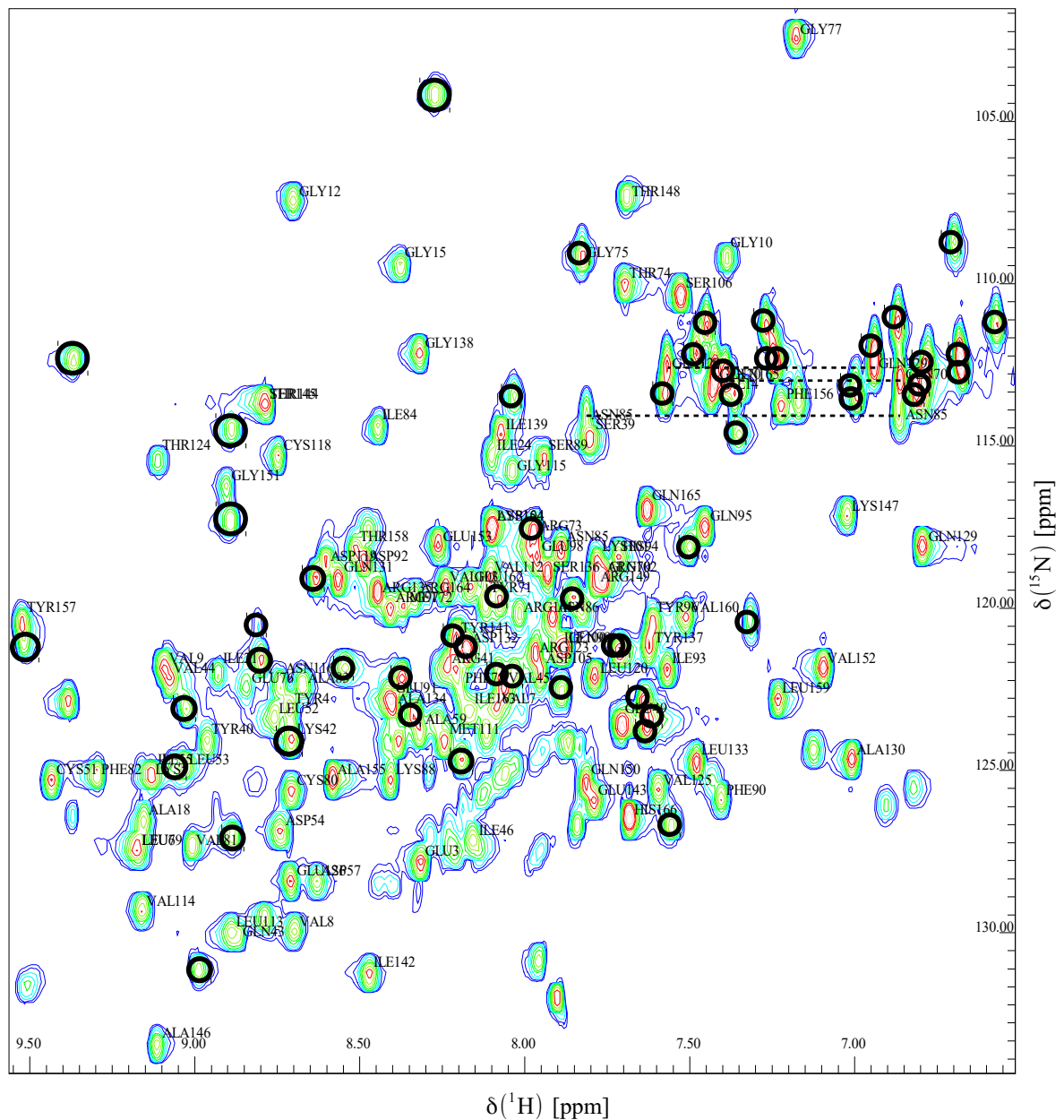


Abbildung 71: *Automatisch mit dem MD-Ensemble zugeordnetes HSQC-Spektrum von Ras(T35S)-GppNHp.* Die vom Benutzer wählbare Sicherheitswahrscheinlichkeit beträgt 100 %. Es wurden keine Signale falsch zugeordnet. Die Positionen an denen Zuordnungen fehlen, sind **schwarz** umrandet. Insgesamt konnten 123 der 180 Signale und 116 der 150 Hauptkettensignale sicher gefunden werden.

## 5 Diskussion

### 5.1 Stereospezifische Zuordnung der Seitenkettenamide von Asparagin und Glutamin in Modellpeptiden

#### 5.1.1 Zuordnungen

Die stereospezifische Zuordnung der Seitenkettenamide des Asn und des Gln sollten mithilfe von MD-Rechnungen und der Simulation eines 2-D-NOESY-Spektrums bestimmt werden.

Bei gefalteten Proteinen erhält man die chemische Verschiebung des Asn  $H^{\delta 21}$  vom stärksten der vier Asn  $H^{\delta 2}/Asn H^{\beta}$ -, die chemische Verschiebung des Gln  $H^{\epsilon 21}$  aus dem stärksten der vier Gln  $H^{\epsilon 2}/Gln H^{\gamma}$ -Kreuzsignale des experimentellen 2-D-NOESY-Spektrums. Diese Annahme ist berechtigt, da der Abstand der  $H^{\beta}$ -Protonen zum  $H^{\delta 21}$  im Asn für alle Diederwinkel  $\chi_2$  kleiner ist als zum  $H^{\delta 22}$  bzw. der Abstand der  $H^{\gamma}$ -Protonen zum  $H^{\epsilon 21}$  im Gln für alle Diederwinkel  $\chi_3$  kleiner ist als zum  $H^{\epsilon 22}$  (McIntosh et al. 1997). Diese Herangehensweise ist aber nicht mehr praktikabel, wenn die vier betrachteten Kreuzsignale nahezu gleiche Intensitäten haben. Erschwerend kam in den Spektren der beiden Tetrapeptide Gly-Gly-Asn-Ala-NH<sub>2</sub> und GlyGly-Gln-Ala-NH<sub>2</sub> hinzu, dass die Korrelationszeiten  $\tau_{rot}$  in unserem Fall ungefähr so groß wie die reziproke Spektrometer Frequenz  $1/\omega$  waren. Dies hat dafür gesorgt, dass homonuklearen NOEs im 2-D-NOESY-Spektrum nahezu verschwanden. Trotzdem war es dank eines Cryoprobekopfes, der die Sensitivität des Spektrometers erhöht, möglich bei 800 MHz NOESY-Spektren mit ausreichendem Signal-zu-Rausch-Verhältnis aufzunehmen.

Die Resonanzen von Asn  $H^{\delta 21}$  und Asn  $H^{\delta 22}$  waren 0,68 ppm, die der beiden Asn  $H^{\beta}$ -Protonen waren 0,09 ppm voneinander entfernt. Beim Gln waren die chemischen Verschiebungen der beiden  $H^{\epsilon 2}$ -Resonanzen 0,7 ppm, die der  $H^{\beta}$ -Resonanzen waren 0,13 ppm voneinander entfernt. Diese Abstände entsprechen ungefähr den Werten, die in Bundi und Wüthrich (1979) angegeben sind ( $|\delta(Asn H^{\delta 21}) - \delta(Asn H^{\delta 22})| = 0,68$  ppm,  $|\delta(Asn H^{\beta 2}) - \delta(Asn H^{\beta 3})| = 0,08$  ppm,  $|\delta(Gln H^{\epsilon 21}) - \delta(Gln H^{\epsilon 22})| = 0,72$  ppm,  $|\delta(Gln H^{\beta 2}) - \delta(Gln H^{\beta 3})| = 0,12$  ppm).

#### 5.1.2 Stereospezifische Zuordnung

Durch den Vergleich der integrierten experimentellen 2-D-NOESY-Signalvolumen mit den aus dem MD-Ensemble zurückgerechneten Volumen der beiden Tetrapeptide war es möglich die stereospezifische Zuordnungen der NH<sub>2</sub>-Gruppen der Seitenketten und der C-Termini zu bestimmen. Um den Einfluss der Skalierung zwischen simulierten und experimentellen



Spektren bei der Auswertung vernachlässigen zu können, wurden Volumenverhältnisse zweier Kreuzsignale für verschiedene Möglichkeiten der stereospezifischen Zuordnung im Experiment und in der Rückrechnung miteinander verglichen. Da sowohl die experimentellen wie auch die simulierten Volumen fehlerbehaftet waren, konnten auch die experimentellen und simulierten Verhältnisse nicht exakt gleich sein. Um trotzdem eine sichere Zuordnung gewährleisten zu können, wurden mehrere Verhältnisse für eine Zuordnung ausgewertet. Es wurden immer dann Möglichkeiten ausgeschlossen, falls sich die simulierten und die experimentellen Volumenverhältnisse stark widersprechen würden.

Aus den Volumen der vier experimentellen Asn-Kreuzsignalvolumen wurden vier der sechs möglichen Verhältnisse zwischen den nicht stereospezifisch zugeordneten  $H^{\delta_{2a,b}}$ - und den ebenfalls nicht eindeutig zuordenbaren  $H^{\beta_{x,y}}$ -Signalen ( $H^{\beta_{x,y}}/H^{\delta_{2a,b}}$ ) gebildet. In jedem Verhältnis musste eine Gruppe entweder im Zähler oder im Nenner vertreten sein. Die vier Verhältnisse wurden für alle vier stereospezifischen Kombinationen der  $H^{\delta^2}$ - und  $H^{\beta}$ -Protonen bestimmt. Nur eine stereospezifische Zuordnung konnte die vier Volumenverhältnisse im Experiment und Simulation in Einklang bringen (Tabelle 2) und zwar die, für die galt:

$\delta(H^{\delta_{21}}) > \delta(H^{\delta_{22}})$  ( $a = 1$  und  $b = 2$ ) und  $\delta(H^{\beta_2}) > \delta(H^{\beta_3})$  ( $x = 2$  und  $y = 3$ ) (Abb. 17). Alle anderen Möglichkeiten die simulieren Volumenverhältnisse zu bilden, hätten zu keiner Übereinstimmung geführt. Wäre nämlich die Reihenfolge der chemischen Verschiebungen der beiden  $H^{\delta^2}$ -Protonen vertauscht ( $a = 2$  und  $b = 1$ ), dann wären die ersten beiden Volumenverhältnisse in Tabelle 2 nicht erklärbar gewesen. Wäre die Reihenfolge der  $H^{\beta}$ -Protonen vertauscht ( $x=3$  und  $y=2$ ), dann wären unter der Bedingung  $\delta(H^{\delta_{21}}) > \delta(H^{\delta_{22}})$  die anderen vier Volumenverhältnisse nicht erklärbar gewesen. Die Reihenfolge der Asn  $H^{\delta^2}$ -Resonanzen, die mit den Volumenverhältnissen gefunden wurde, stimmt mit der Mehrzahl der Zuordnungen der BMRB überein (siehe Kap. 1.1).

Das stärkste Kreuzsignal konnte dem Asn  $H^{\delta_{21}}/Asn H^{\beta_3}$  zugeordnet werden, was sich mit der Aussage in McIntosh et al. (1997) deckt. Die stereospezifische Zuordnung der Asn  $H^{\beta}$ -Protonen, die sich aus den vier Volumenverhältnissen ergeben hat und mit zwei weiteren Verhältnissen bestätigt werden konnte, deckt sich ebenfalls mit der Reihenfolge durchschnittlichen chemischen Verschiebungen der BMRB ( $\bar{\delta}(Asn H^{\beta_2}) = 2,81 \text{ ppm} > \bar{\delta}(Asn H^{\beta_3}) = 2,75 \text{ ppm}$ . Stand: 18.06.2015).

Beim Gln wurden zusätzlich die Verhältnisse zwischen den Signalen  $H^{\epsilon_{2x,y}}/H^{\gamma_{2,3}}$  ausgewertet. Da die Linien der beiden  $H^{\gamma_{2,3}}$ -Protonen im 2-D-NOESY-Spektrum von Gly-Gly-Gln-Ala-NH<sub>2</sub> nicht unterscheidbar waren und sie in der 3D-Struktur näher als die  $H^{\beta}$ -Protonen an den  $H^{\delta^2}$ -Protonen lagen, hatten die entsprechenden Kreuzsignale ein höheres Signal-zu-Rausch-Verhältnis und waren deshalb für die stereospezifische Zuordnung aussagekräftiger. Die fünf Volumenverhältnisse konnten nur für  $\delta(Gln H^{\epsilon_{21}}) > \delta(Gln H^{\epsilon_{22}})$  und  $\delta(Gln H^{\beta_2}) > \delta(Gln H^{\beta_3})$  in Einklang gebracht werden. Die Verhältnisse der intraresidualen  $H^{\beta_{x,y}}/H^{\alpha}$ -Kreuzsignale bzw. der  $H^{\beta_{x,y}}/H^N$ -Kreuzsignale bestätigen die gefundene stereospezifische Zuordnung der  $H^{\beta}$ -Protonen.

Eine vertauschte  $H^{\epsilon 2}$ -Zuordnung ( $a = 2$  und  $b = 1$ ) hätte die ersten drei Verhältnisse in Tabelle 4 nicht erklären können. Die stereospezifische Zuordnung der  $H^{\beta}$ -Protonen ( $x = 2$  und  $y = 3$ ) hat alle Volumenverhältnisse bis auf das Verhältnis mit den  $H^{\alpha}$ -Protonen besser als die alternative Zuordnung erklärt. Da aber die an diesem Volumenverhältnis beteiligten Kreuzsignale nur durch einen einzigen Diederwinkel, nämlich  $\chi_1$ , der außerdem noch frei drehbar ist, voneinander getrennt waren, waren auch die topologischen Einschränkungen der  $H^{\alpha}$ - $H^{\beta 2,3}$ -Abstände nicht besonders hoch. Damit war das Verhältnis im allgemeinen nicht sehr gut definiert. Die stereospezifische Zuordnung der  $H^{\epsilon 2}$ -Protonen deckt sich mit der in McIntosh et al. (1997), da das zugeordnete Signal  $H^{\epsilon 21}/H^{\gamma 2,3}$  im Experiment stärker war als das Signal  $H^{\epsilon 22}/H^{\gamma 2,3}$ . Die gefundene Zuordnung stimmt ebenfalls wieder mit der Mehrheit der Einträge in der BMRB überein (Kap. 1.1). Die bestimmte stereospezifische Reihenfolge der  $H^{\beta}$ -Protonen und die der durchschnittlichen chemischen Verschiebungen der BMRB sind ebenfalls gleich ( $\bar{\delta}(\text{Gln } H^{\beta 2}) = 2,05 \text{ ppm} > \bar{\delta}(\text{Gln } H^{\beta 3}) = 2,00 \text{ ppm}$ . Stand: 18.06.2015). Durch Auswertung von Volumenverhältnissen konnten auch die C-Terminalen  $\text{NH}_2$ -Gruppen der beiden Tetrapeptide eindeutig und sicher zugeordnet werden (Tabelle 2 und 4).

Ein visueller Vergleich der 1-D-Spektren von Asn  $H^{\beta 2}$  und Asn  $H^{\beta 3}$  bestätigte das Ergebnis, das aus den Volumenverhältnissen gewonnen wurde. Die simulierten 1-D-Multiplettmuster der Asn  $H^{\beta}$ -Signale (Abb. 18) sind für beide Protonen unterscheidbar. Der zurückgerechnete spektrale Bereich stimmt nur mit der richtigen Reihenfolge der  $H^{\beta}$ -Protonen mit dem experimentellen Spektrum überein. Die für die Simulation benötigten  ${}^3J_{H^{\beta 2,3}H^{\alpha}}$ -Kopplungskonstanten wurden mit der Karplus-Gleichung aus dem MD-Ensemble berechnet (3,7 Hz und 8,1 Hz) und stimmen gut mit den Werten von Bundi und Wüthrich (1979) (5,8 Hz und 8,3 Hz) überein. Dies gilt auch für die gemessene  ${}^2J_{H^{\beta 2}H^{\beta 3}}$ -Kopplungskonstante, die mit 15,6 Hz sehr nahe an der veröffentlichten Kopplungskonstante von 16,0 Hz lag. Zusammen mit den Volumenverhältnissen konnte die stereospezifische Zuordnung der Asn  $H^{\beta}$ -Protonen in Gly-Gly-Asn-Ala- $\text{NH}_2$  zweifelsfrei bestimmt werden.

Im Prinzip sollte es möglich sein, wie beim Asn, die  $H^{\beta}$ -Protonen des Gln mithilfe des J-Kopplungsmusters zuzuordnen. Die aus der MD-Trajektorie und der Karplus-Gleichung berechneten Kopplungskonstanten ergaben sich für  ${}^3J_{H^{\alpha}H^{\beta 2}}$ ,  ${}^3J_{H^{\alpha}H^{\beta 3}}$ ,  ${}^3J_{H^{\beta 2}H^{\gamma}}$  und  ${}^3J_{H^{\beta 3}H^{\gamma}}$  zu 5,3 Hz, 5,0 Hz, 9,8 Hz und 9,8 Hz. Diese Werte würden zwar die experimentellen Linienformen erklären, aber eine stereospezifische Zuordnung der  $H^{\beta}$ -Protonen war wegen der Komplexität des Multiplettmusters nicht möglich.

### 5.1.3 Bestimmung der chemischen Austauschraten

Wie bereits erwähnt, mussten für eine möglichst genaue Vorhersage der Kreuzsignalvolumen der Flippaustausch der  $\text{NH}_2$ -Gruppen berücksichtigt werden. Bei einer  ${}^1\text{H}$ -Frequenz von 800 MHz ist für Gly-Gly-Asn-Ala- $\text{NH}_2$  und Gly-Gly-Gln-Ala- $\text{NH}_2$   $\omega \tau_{\text{rot}} \approx 1$ , d.h. die Mehrzahl der NOESY-Kreuzsignale hatten in den 2-D-NOESY-Spektren

negatives Signalvolumen. Ausnahmen davon waren die Kreuzsignale Asn  $H^{\delta 21}$ /Asn  $H^{\delta 22}$ , Gln  $H^{\epsilon 21}$ /Gln  $H^{\epsilon 22}$ ,  $H^{21}(E)/H^{22}(Z)$  und deren an der Diagonale im 2-D-NOESY-Spektrum gespiegelten Signale. Das positive Volumen ist ein Folge des Flippaustauschs der  $NH_2$ -Gruppen und wurde bereits in JURANIĆ et al. (1998) für Gln in  $H_2O/D_2O$ -Gemischen beobachtet. Die Rate des chemischen Austausches wurde dort mit 0,2 Hz bei einer Temperatur von 275 K bestimmt. Die Austauschraten konnten auch mit AUREMOL RELAX bestimmt werden. Die von RELAX bestimmte Rate von 0,4 Hz für den Flip um die partielle Doppelbindung  $C^{\gamma}/N^{\epsilon 2}$  des Gln3 im Gly-Gly-Gln-Ala- $NH_2$  lag in der gleichen Größenordnung und war wegen der höheren Temperatur (293 K) wie erwartet auch größer. Die von AUREMOL RELAX berechnete Rate  $\sigma_{H^{\epsilon 21}H^{\epsilon 22}}$  der dipolaren Kopplung, die dem chemischen Austausch im 2-D-NOESY-Spektrum von Gly-Gly-Gln-Ala- $NH_2$  entgegenwirkt, lag mit 0,23 Hz bei 293 K ebenfalls nahe an der veröffentlichten Rate von 0,2 Hz. Die veröffentlichten Raten des chemischen Austausches bzw. der dipolaren Kopplung der  $H^{\epsilon 2}$ -Protonen mit dem Lösungsmittel sind mit 0,014 Hz bzw. 0,022 Hz viel kleiner als die entsprechenden Werte der Diagonalelemente der Relaxationsmatrix  $R_{1,H^{\epsilon 21}} = 0,26$  Hz bzw.  $R_{1,H^{\epsilon 22}} = 0,18$  Hz. Diese Raten konnten deshalb für die Rückrechnungen vernachlässigt werden. Die Raten der Asparagin Seitenketten und der C-Terminalen  $NH_2$ -Gruppen des Asparagins und Glutamins lagen mit 1,3 Hz, 0,3 Hz und ungefähr 0 Hz ebenfalls im erwarteten Bereich des chemischen Austauschs.

#### 5.1.4 Aus den 3D-Strukturen vorhergesagte Dichteverteilungen der chemische Verschiebungen

Die beiden häufig verwendeten Programme für die Abschätzung von chemischen Verschiebungen aus Protein-3D-Strukturen SHIFTX (Neal 2003) und SHIFTS (Osapay und Case 1991) vertauschen in ihren Vorhersagen die chemischen Verschiebungen der Asn  $H^{\delta 2}$ -Protonen. Sie ordnen beide dem  $H^{\delta 21}$  eine kleinere chemische Verschiebung als dem  $H^{\delta 22}$  zu. Die Reihenfolge der chemischen Verschiebungen der beiden Gln  $H^{\epsilon 2}$ -Protonen wird von SHIFTX richtig und von SHIFTS falsch vorhergesagt. Die falschen Reihenfolgen der Zuordnungen sind auf fehlerhafte Datenbanken in den Programmen zurückzuführen. Die Annahme wurde im Falle des SHIFTS von der Ausgabedatei bestätigt. In dieser Datei findet man neben den vorhergesagten chemischen Verschiebungen auch die einzelnen Beiträge, aus denen sich die vorhergesagte chemische Verschiebung zusammensetzt. Es war dort zu erkennen, dass die Werte der Random-Coil chemischen Verschiebungen der Protonen der beide Seitenketten  $NH_2$ -Gruppen vertauscht sind.

Bei der Erstellung der Datenbank von SHIFTX wurde vorausgesetzt, dass die Resonanz des Asn  $H^{\delta 21}$  immer die kleinere und die des Asn  $H^{\delta 22}$  immer die größere chemische Verschiebung hat. Beim Gln wurde für die Datenbank hingegen die richtige Reihenfolge, nämlich dass die Resonanz des  $H^{\epsilon 21}$  die größere und die des  $H^{\epsilon 22}$  die kleinere chemische Verschiebung hat, vorausgesetzt.

Man kann in Abb. 20 erkennen, dass die Verteilung des Asn  $H^{\delta 21}$  breiter als die des Asn  $H^{\delta 22}$

ist. Dies lässt vermuten, dass Asn H<sup>δ21</sup> stärker mit seiner Umgebung wechselwirkt als Asn H<sup>δ22</sup> mit seiner. Die Ursache dafür könnten schwache spezifische Wechselwirkungen sein. Dies würde auch die inhomogene Form der Verteilungen des Asn H<sup>δ21</sup> erklären. Das gleiche Verhalten konnte auch für die Verteilungen der chemischen Verschiebungen der Gln H<sup>ε2</sup>-Protonen beobachtet werden.

## 5.2 Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen

Die vorhergesagten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der Resonanzen  $f_k(\delta)$  der Kerne  $k$  spielen eine sehr große Rolle bei der automatischen Zuordnung von Proteinen mit SIBASA. Die Zielslots der speziellen MOVES oder SWAP MOVES werden während der Optimierung der Zuordnungsliste mithilfe dieser Wahrscheinlichkeitsdichtefunktionen ausgewählt (Kap. 2.2.7). Außerdem tragen sie über die Pseudoenergie der chemischen Verschiebung  $E^\circ$  (Gl. 62), vor allem zu Beginn der Optimierung, einen großen Anteil zur gesamten Pseudoenergie der Optimierung  $E^{\text{total}}$  (Gl. 44) bei. Mit einer sehr gut vorhergesagten Wahrscheinlichkeitsdichten der chemischen Verschiebungen ist es im besten Fall möglich, die chemische Verschiebungen dieser Resonanzen unabhängig von den Zuordnungen anderer Resonanzen richtig zuzuordnen. Für die automatische Zuordnung ist dies besonders wichtig, da die Pseudoenergien, die auf experimentellen Daten basieren, mindestens eine richtige Zuordnungen benötigen um eine andere zu finden. Die lokale Pseudoenergie des Signalvergleichs auf den experimentellen Signalen  $j$  im Spektrum  $i$   $E_{ij}^{\text{Vol, Shape}}$  (Kap. 2.2.8) kann nur richtig zugeordnete experimentelle Kreuzsignale, also eine Kombination von mindestens zwei richtig gefundenen chemischen Verschiebungen, in der Optimierung belohnen. Je mehr Resonanzen von Kreuzsignalpartner eines Kerns  $k$  bereits richtig zugeordnet wurden, desto stärker belohnt die Pseudoenergie des Signalvergleichs  $E_{ij}^{\text{Vol, Shape}}$  die an die richtige Position verschobene Resonanz des Kerns  $k$ . Besser vorhergesagte Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen erhöhen den Anteil der unabhängig von den experimentellen NOESY-Kreuzsignalen richtig zuordenbaren Resonanzen, was der automatischen Zuordnung die Suche nach anderen Resonanzzuordnungen über experimentelle NOESY-Kreuzsignale erheblich erleichtert.

Die vorhergesagten Wahrscheinlichkeitsdichten in ASSIGN, dem Vorgänger von SIBASA, waren Normalverteilungen, deren Mittelwerte aus einer Struktur mit SHIFTS (Osapay und Case 1991) geschätzt wurden und deren Standardabweichung aus der BMRB (Ulrich et al. 2008) stammten ( $f_k(\delta) = N(\delta_k^{\text{SHIFTS}}, \sigma_{AA, \text{Kern}}^{\text{BMRB}})$ ) ( $AA$  = Aminosäure wie z.B. Alanin,  $\text{Kern}$  = Kernart wie H<sup>N</sup>, N, H<sup>α</sup>, ...). SIBASA dagegen kombiniert spezielle, von dem 3D-Strukturensemble des Proteins abhängige, mit allgemeinen von der Aminosäure und der Kernart abhängigen Wahrscheinlichkeitsdichten der chemischen Verschiebungen. In die

speziellen Wahrscheinlichkeitsdichten gehen die mit SHIFTX2 (Han et al. 2011) vorhergesagten chemischen Verschiebungen aller Einzelstruktur des MD-Ensembles ein, wodurch die konformationellen Zustände des Proteins in Lösung berücksichtigt werden können. Dies ist für die automatische Zuordnung von Vorteil, da eine einzige 3D-Struktur die Struktur des Proteins in Lösung unter Experimentbedingungen nicht repräsentieren kann. Im Fall von Röntgenstrukturen kann man sich z.B. nicht sicher sein, dass der kristallisierte Zustand auch der in Lösung hauptsächlich vorkommende Zustand ist. In den Tabellen 8, 10 und 12 ist für drei Proteine gezeigt worden, dass ein MD-Ensemble näher an den experimentellen NMR-Daten und den NMR-Strukturen liegt, als die Röntgenstruktur.

Der durchschnittliche Fehler der Vorhersage  $\epsilon_{\text{H}}^{\text{Vorhersagefehler}}$  für die Zuordnungen der Protonen  $^1\text{H}$  eines Proteins ist folgendermaßen definiert:

$$\epsilon_{\text{H}}^{\text{Vorhersagefehler}} = \frac{1}{N^{\text{H}}} \sum_{k=1}^{N^{\text{H}}} |\delta_k^{\text{vorhergesagt}} - \delta_k^{\text{Referenz}}| \quad (94)$$

Hierbei ist  $N^{\text{H}}$  die Anzahl der Protonen im Protein und  $\delta_k^{\text{vorhergesagt}}$  bzw.  $\delta_k^{\text{Referenz}}$  sind die vorgesagten bzw. richtigen chemische Verschiebungen der Resonanzen der Protonen  $k$ . Der durchschnittliche Fehler der Vorhersage  $\epsilon_{\text{H}}^{\text{Vorhersagefehler}}$  sank in SIBASA (im Gegensatz zum Vorgänger ASSIGN) im Fall von HPr S. aureus (H15A) von 0,27 ppm um 26 % auf 0,20 ppm, beim PfTrx von 0,31 ppm um 26 % auf 0,227 ppm und beim Ras(T35S)-GppNHp von 0,25 ppm um 12,4 % auf 0,22 ppm.

Mit Hilfe aller aus der MD-Trajektorie geschätzten Werte und eines Kerndichteschätzers konnten die Breiten der vorhergesagten Wahrscheinlichkeitsdichteverteilungen  $f_k(\delta)$  der chemischen Verschiebung wesentlich verkleinert werden. Die Breiten der Normalverteilungen in ASSIGN waren die Standardabweichung  $\sigma_{\text{AA,Kern}}^{\text{BMRB}}$  aus der BMRB, die aus allen dort abgelegten chemischen Verschiebung der gleichen Kernart *Kern* und der gleichen Aminosäure *AA* bestimmt worden sind. Die durchschnittliche Standardabweichung der vorhergesagten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen aller Protonen  $^1\text{H}$  sank beim HPr S. aureus (H15A) im Gegensatz zu ASSIGN von 0,375 ppm um 55 % auf 0,17 ppm, beim PfTrx sank die durchschnittliche Breite um 51 % von 0,376 ppm auf 0,186 ppm und beim Ras(T35S)-GppNHp um 53 % von 0,382 ppm auf 0,179 ppm. Außerdem sind die Dichteverteilungen durch den Einsatz eines Kerndichteschätzers in SIBASA nicht mehr auf unimodale Kurven beschränkt, wodurch Wahrscheinlichkeitsdichten von chemischen Verschiebungen, die mehrere Maxima haben, genauer vorhergesagt werden können.

Für die allgemeinen nur von Aminosäure *AA* und Kernart *Kern* abhängigen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen in SIBASA  $f_{\text{AA,Kern}}(\delta)$  wurden alle 3D-Strukturen der unbiased N<sub>h</sub>3D-Röntgenstrukturdatenbank (Thiruv et al. 2005) mit GROMACS (Hess et al. 2008; Berendsen et al. 1995) in Lösung energieminiert. Anschließend wurden mit SHIFTS (Osapay und Case 1991), einem Kerndichteschätzer und

einem Gauß-Kernel die Wahrscheinlichkeitsdichteverteilungen  $f_{AA,Kern}(\delta)$  für alle Aminosäure-Kernpaare gebildet (Baskaran 2010) (Kap. 2.2.7). Durch den Einsatz des Kerndichteschätzers sind auch die allgemeinen Verteilungen  $f_{AA,Kern}(\delta)$  nicht mehr auf Normalverteilungen beschränkt. Der chemischen Verschiebungsbereich, in dem Resonanzen von Kernen, vorkommen, ist häufig abhängig von der Sekundärstruktur ( $\alpha$ -Helix,  $\beta$ -Faltblatt, nicht strukturierten Bereich). Die allgemeinen Wahrscheinlichkeitsdichteverteilungen sollten deshalb mehr als ein Maximum haben können, falls die Resonanz in Abhängigkeit von der Sekundärstruktur in verschiedenen trennbaren spektralen Bereichen vorkommen kann. Zudem ist die Nh3D-Datenbank im Gegensatz zur BMRB eine unbiased Strukturdatenbank, wodurch keine in der Natur vorkommende Strukturart übergewichtet wird. In der BMRB ist dies nicht der Fall, da alle abgelegten chemischen Verschiebungen gleich stark zum Mittelwert  $\bar{\delta}_{AA,Kern}^{BMRB}$  und zur Standardabweichung  $\sigma_{AA,Kern}^{BMRB}$  beitragen. Durch Kombination der allgemeinen und speziellen Verteilungen konnten viele Ausreißer, deren Resonanzzuordnung mit SHIFTX2 und dem MD-Ensemble nicht vorhergesagt werden konnten, trotzdem für die automatische Zuordnung abgedeckt werden.

Die Wahrscheinlichkeit der richtigen Vorhersage der chemischen Verschiebung der Resonanz des Kerns  $k$   $P_k^{Vorhersage}(\text{korrekt})$  (Gl. 90) ist ein gutes Instrument um die Güte der mit verschiedenen Methoden vorhergesagten Wahrscheinlichkeitsdichten  $f_k(\delta)$  der chemischen Verschiebungen zu messen. Je höher diese Wahrscheinlichkeit ist, desto schneller kann SIBASA die richtige Zuordnung dieser Resonanz finden und desto besser kann SIBASA richtige Zuordnung über die Pseudoenergie der chemischen Verschiebung  $E^\circ$  bewerten. Ist die Wahrscheinlichkeit der richtigen Vorhersage der chemischen Verschiebung für eine Resonanz 0, so kann sie nur mit dem allgemeinen MOVE oder SWAP MOVE in SIBASA richtig zugeordnet werden (Kap. 2.2.11). Eine Wahrscheinlichkeit von 1, entspräche einer vorher festgelegten richtigen chemischen Verschiebung, die die automatische Zuordnung eines Proteins im erheblichen Maß vereinfachen kann.

Die durchschnittliche Wahrscheinlichkeit der richtigen Vorhersage von Protonen  $^1H$   $\bar{P}_H^{Vorhersage}(\text{korrekt})$  hat sich, im Vergleich zu ASSIGN, um 106 % beim Protein HPr S. aureus (H15A) verbessert. Für die größeren Proteine PfTrx und Ras(T35S)-GppNHp betrug die Verbesserung immer noch um die 70 %. Für die schwer vorherzusagenden chemischen Verschiebungen der Resonanzen der  $H^N$ -Protonen verbesserte sich durchschnittliche Wahrscheinlichkeit der richtigen Vorhersage  $\bar{P}_{H^N}^{Vorhersage}(\text{korrekt})$  beim PfTrx um 64 % und beim Ras(T35S)-GppNHp um 6 %.

Während des speziellen MOVES in SIBASA hängt die Pseudokraft eines experimentelles 2-D-NOESY-Kreuzsignals auf ein  $H^N$ -Proton auch von den Wahrscheinlichkeitsdichtefunktionen der Kreuzsignalpartner des  $H^N$ -Protons ab (Gl. 78). Da die durchschnittliche Wahrscheinlichkeit der richtigen Vorhersage von Seitenketten- bzw.  $H^\alpha$ -Protonen im größten Testprotein Ras(T35S)-GppNHp mit der Methode aus SIBASA um 81 % bzw. 57 % gesteigert werden konnte, profitiert auch die automatische Zuordnung der  $H^N$ -

Protonenresonanzen direkt von der neuen Methode. Die relativ guten Vorhersagen der chemischen Verschiebungen der Resonanzen von Seitenkettenprotonen aus der Röntgenstruktur des Proteins sprechen deshalb auch für die automatische Zuordnung von HSQC-Spektren über das 2-D-NOESY-Spektrum. Die Wahrscheinlichkeiten der richtigen Vorhersage sind selbst beim großen Ras(T35S)-GppNHp mit 12,3 % mehr als dreimal so hoch wie für die Resonanzen von  $H^N$ -Protonen.

In Abb. 22 wurde gezeigt, dass die Wahrscheinlichkeitsdichteverteilungen von SIBASA, die mit MD-Strukturensembles und der Nh3D-Strukturdatenbank erstellt worden sind, die Form der Häufigkeitsverteilungen der Vorhersagewahrscheinlichkeiten verändern. Die Häufigkeitsverteilungen der Wahrscheinlichkeiten einer richtigen Vorhersage  $H(P_{iH}^{\text{Vorhersage}}(\text{korrekt}))$  in SIBASA sind breiter als die Häufigkeitsverteilungen der Wahrscheinlichkeitsdichten der chemischen Verschiebungen  $f_k(\delta)$ , die nur eine 3D-Struktur berücksichtigen. Der Anteil von hohen Wahrscheinlichkeiten einer richtigen Vorhersage, konnte in SIBASA vergrößert werden. Diese Resonanzen können von SIBASA sehr schnell zugeordnet werden. Da sie eine hohe Pseudoenergie der chemischen Verschiebung  $E^\omega$  erhalten, verlassen sie die richtige Position auch nicht mehr und können die Zuordnung ihrer Kreuzsignalpartner über die lokalen Pseudoenergien des Signalvergleichs  $E_{ij}^{\text{Vol, Shape}}$  unterstützen.

In Zukunft kann die automatische Zuordnung in SIBASA mit Hilfe von verbesserten Vorhersageprogrammen und genaueren MD-Ensembles ohne Veränderung des Zuordnungsalgorithmus und ohne Zunahme zusätzlicher experimenteller Informationen entscheidend verbessert werden. Wegen der Abhängigkeit der experimentellen Pseudoenergien der richtigen Zuordnung eines Kerns  $k$  von den chemischen Verschiebung der Resonanzen aller seiner Kreuzsignalpartner, würde SIBASA sogar mehr als linear von der Erhöhung Wahrscheinlichkeiten der richtigen Zuordnungen  $P_k^{\text{Vorhersage}}(\text{korrekt})$  profitieren. Eine einfache Möglichkeit die Vorhersagen der chemischen Verschiebungen zu verbessern ist die Verwendung des Programms SHIFTY (Han et al. 2011), das Teil von SHIFTX2 ist. SHIFTY korrigiert die von SHIFTX2 vorhergesagten Werte der chemischen Verschiebungen mithilfe von veröffentlichten Zuordnungslisten von homologen Proteinen. SHIFTY kann die Qualität der Vorhersagen mit dieser Methode entscheidend verbessern.

Wenn sich die chemischen Verschiebungen eines automatisch zuzuordnenden Proteins sich, wie z.B. bei einer Punktmutante, (lokal) nicht stark von den chemischen Verschiebungen eines bereits gelösten homologen Proteins unterscheiden, können Normalverteilungen  $f_k(\delta) = N(\delta_k^{\text{homolog}}, \sigma_k)$  mit den Mittelwerten  $\delta_k^{\text{homolog}}$  des Kerns  $k$  des homologen Proteins und kleinen Standardabweichungen  $\sigma_k$  das Zuordnungsproblem erheblich vereinfachen (Kap. 2.2.7 und 5.3.10).

### 5.3 Automatische sequentielle Zuordnung (SIBASA)

Das AUREMOL Modul für die sequentielle automatische Zuordnung von HSQC-Spektren und Seitenketten SIBASA wurde für drei Proteine unterschiedlicher Größe getestet. Im Gegensatz zu seinem Vorgänger ASSIGN wurden SIBASA in vielen Punkten verbessert.

Es kann momentan neben dem 2-D-NOESY-Spektrum auch ein 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum verarbeiten und wird in Zukunft in der Lage sein weiterer NMR-Spektren auszuwerten. Die Rückrechnungen der NOESY-Spektren, die Grundlage des Top-Down-Ansatzes von SIBASA und ASSIGN sind, wurden entscheidend verbessert (Kap. 2.2.3).

Mit SIBASA war es für drei Proteine unterschiedliche Größe möglich, das HSQC-Spektrum vollständig richtig teilzuzuordnen. Jede gefundene Zuordnung eines HSQC-Signals erhält von SIBASA eine Verifikationspseudoenergie, die auf die Wahrscheinlichkeit der richtigen Zuordnung abgebildet werden kann. Vertraut der Benutzer nur Wahrscheinlichkeiten von 100 %, so bleiben keine falschen Zuordnungen zurück. Die Wahrscheinlichkeiten der richtigen Zuordnung von HSQC-Signalen waren beim größten Protein Ras(T35S)-GppNHp sehr zuverlässig. Beim kleinsten der Testproteine HPr *S. aureus* (H15A) wurden die chemischen Verschiebungen aller Amidgruppen richtig gefunden, aber nicht alle bestätigt. Hier war die Verifikation zu restriktiv, was aber besser ist als falsche Zuordnungen zu bestätigen. Beim reduzierten Plasmodium falciparum Thioredoxin wurden keine falschen HSQC-Zuordnungen bestätigt, doch war hier der Abstand zwischen der maximalen Verifikationspseudoenergie einer falschen Zuordnung eines HSQC-Signals und der minimalen Verifikationspseudoenergie für 100 % sichere Zuordnungen sehr klein. Durch einen weiteren Schritt in der Verifikation kann aber auch diese mögliche Fehlerquelle einfach behoben werden und somit gewährleistet werden, dass keine falschen zugeordneten HSQC-Signale von SIBASA bestätigt werden.

Keine andere Methode der automatischen Zuordnung kann HSQC-Spektren ohne eine  $^{13}\text{C}$ -markierte Probe des Proteins sicher auch nur teilweise zuordnen. Die meisten anderen Lösungen benötigen hierfür mindestens ein Triple-Resonanz-Spektrum (Kap. 1.2).

Mit ST2NMR (Pristovsek et al. 2002) ist es möglich mithilfe von intraresidualen Spinsystemen aus TOCSY-Spektren, einem 2-D-NOESY-Spektrum, einem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum und einer 3D-Struktur die chemischen Verschiebungen der Amidgruppen zu bestimmen. Bei der automatischen Zuordnung des oxidierten *Paracoccus Denitrificans* Cytochrome  $c_{552}$  (112 Aminosäuren) konnten 100 % der TOCSY-Spinsysteme richtig zugeordnet werden. SIBASA konnte beim ungefähr gleich großen PfTrx (104 Aminosäuren) 89 der 100 Amidgruppen von Hauptketten mit sichtbaren NOESY-Kreuzsignalen ohne das von ST2NMR benötigte TOCSY-Spektrum richtig zuordnen und 78 davon als sicher richtig bestätigen. Für das  $\alpha_2\text{u}$ -Gloulin von Mäusen (162 Aminosäuren) war, laut der Veröffentlichung, die Signalliste des 2-D-NOESY-Spektrums für ST2NMR nicht verwendbar, da das Protein zu groß für die Strukturbestimmung mit diesem Spektrum ist. Die Vermutung liegt nahe, dass eine unvollständige Signalliste, die man wegen Signalüberlapps aus dem 2-D-



NOESY-Spektrum erhalten würde, die automatische Zuordnung stören würden. Nur mit dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum konnte ST2NMR nur 110 der 162 Spinsystemen richtig zuordnen. Beim Ras(T35S)-GppNHp (166 Aminosäuren), das von der Größe her vergleichbar ist, hat SIBASA lediglich 9 der 150 Zuordnungen von Amidgruppen mit sichtbaren Kreuzsignalen falsch zugeordnet.

ST2NMR und viele andere Programme brauchen, im Gegensatz zu SIBASA, TOCSY-Spinsysteme um die Resonanzen von Seitenkettenprotonen zuzuordnen. Für größer Proteine ist das benötigte 2-D-TOCSY-Spektrum nicht geeignet, da während der langen Mischzeiten, die benötigt werden um Spinsystem vollständig aufzuklären, die dipolare Kopplung die skalare Kopplung überwiegen kann (Cavanagh 2007). Das 2-D-TOCSY-Spektrum kann durch ein 3D HCCH-TOCSY-Spektrum, das die stärkere Kopplung über die Kohlenstoffe nutzt, ersetzt werden. Dafür wird aber eine  $^{13}\text{C}$ -markierte Probe in  $\text{D}_2\text{O}$  und weitere Messzeit benötigt. Durch das Lösungsmittel ändern sich auch die chemischen Verschiebungen der gemessenen Resonanzen im 3D HCCH-TOCSY Spektrum, was die Synchronisation der Spinsysteme mit dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum erschwert. Mit dem 3D  $^{13}\text{C}$ -NOESY-HSQC-Spektrum war es mit ST2NMR möglich alle TOCSY-Spinsysteme richtig zuzuordnen.

Nur GARANT (Bartels et al. 1996; Bartels et al. 1997; Güntert 2009) (Kap. 1.2) könnte neben SIBASA durch seinen allgemeinen Aufbau, eine automatische Zuordnung von HSQC-Spektren nur mit dem 2-D-NOESY- und dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum erreichen. GARANT ist, wie in den Veröffentlichungen erwähnt, sehr stark von den Signallisten, die durch Peakpicking aus den Eingabespektren erzeugt werden müssen, abhängig. In Bartels et al. (1997) wurde auch darauf hingewiesen, dass dieser Nachteil durch die direkte Interaktion des automatischen Zuordnungsalgorithmus mit den experimentellen Spektren behoben werden kann. In SIBASA wird eben dies durch den Vergleich der spektralen Bereiche der beobachteten Signale zwischen Rückrechnung und experimentellen NMR-Spektren erreicht. Dadurch kann SIBASA als einzige Lösung der automatische Zuordnung Überlagerungen von Kreuzsignalen berücksichtigen.

Mit der NMR-Struktur, dem CBCA(CO)NH-, dem 3D  $^{13}\text{C}$ - und dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum konnte GARANT 92,7 % der HSQC-Zuordnungen des Cyclophilin A (165 Aminosäuren) richtig bestimmen. SIBASA dagegen konnte für das ungefähr gleich große Ras(T35S)-GppNHp (166 Aminosäuren) 141 der 150 (94 %) Hauptkettenamidgruppen mit sichtbaren NOESY-Kreuzsignalen nur mit dem 2-D-NOESY- und dem 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum richtig zugeordnet. SIBASA hat dafür nur eines statt drei 3D-Spektren und nur eine einfach  $^{15}\text{N}$ -markierte Probe in  $\text{H}_2\text{O}$  benötigt. Es scheint so, dass auch GARANT, wie schon ST2NMR, die gut aufgelösten 2-D-NOESY-Spektren mit sehr gutem Signal-zu-Rausch-Verhältnis für größere Proteine nicht verwenden kann.

Außerdem bieten weder GARANT noch ST2NMR die Möglichkeit an die Qualität der automatischen Zuordnungen zu verifizieren. In SIBASA kann die Qualität einer Zuordnung durch die Auswertung mehrerer unabhängiger Instanzen und durch den Vergleich der Linienformen der simulierten Signale mit dem experimentellen Spektrum bestimmt werden

(Kap. 2.2.12).

SIBASA ist der einzige Algorithmus, der HSQC-Spektren nur mit einer 3D-Struktur einer einfach markierten  $^{15}\text{N}$ -Probe und einem einzigen zeitaufwendigem 3D-Spektrum sicher (teil)zuordnen kann. Durch zukünftige Verbesserungen bei den Rückrechnung, bessere Vorhersagen der chemischen Verschiebung und genauere Verifikation wird es möglich sein nahezu alle Amidgruppen, die gut sichtbaren NOESY-Kreuzsignale haben, sicher zuzuordnen.

### 5.3.1 Ablauf der automatischen Zuordnung

Die reduzierten experimentellen Datensätze konnten durch die Segmentation der neuen AUREMOL Integration (Donaubauer, Veröffentlichung in Vorbereitung) besser auf die wesentlichen Bereiche reduziert werden (Kap. 2.22). Die möglichen chemischen Verschiebungen, die für alle Kerne im Protein während der automatischen Zuordnung getestet werden, wurden durch die neuen Slotlisten (Kap. 2.2.6) und durch die neuen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen (Kap. 2.2.7) sinnvoll eingeschränkt und gewichtet.

Die Grundflächen der experimentellen Kreuzsignale der NOESY-Spektren, auf denen die Pseudoenergie des Volumen- und Linienformvergleichs  $E^{\text{Vol,Shape}}$  berechnet werden, werden nach ihrer Relevanz gewichtet und falls nötig zusammengefasst um die Übergewichtung von Signalen zu vermeiden. Mit SIBASA wurde ein neuer Beitrag zur Gesamtenergie der Optimierung  $E^{\text{total}}$  eingeführt. Die Pseudoenergie des Dummyslots  $E^{\text{Dummy}}$  (Kap. 2.2.9) verbessert die Erkennung von Amidprotonen und Stickstoffe, die keine sichtbaren Kreuzsignale in den NOESY-Spektren haben. Alle Beiträge der Gesamtpseudoenergie der Optimierung  $E^{\text{total}}$ , nämlich die Pseudoenergien des Signalvergleichs  $E_i^{\text{Vol,Shape}}$  der beiden Eingangsspektren  $i$  (Gl. 56), die Pseudoenergie der chemischen Verschiebung  $E^\circ$  (Gl. 62) und die neu eingeführte Pseudoenergie des Dummyslots  $E^{\text{Dummy}}$  (Gl. 72) werden sinnvoll gegeneinander gewichtet. Die beiden letzten Pseudoenergien erhielten einen dynamischen Gewichtungsfaktor, der während der automatischen Zuordnung den Fokus von statistischen auf experimentelle Daten verschiebt (Kap. 2.2.8 und 2.2.9).

Nach jeder versuchten Veränderung einer chemischen Verschiebung während der Optimierung, werden die zurückgerechneten Spektren mit der aktuellen veränderten Zuordnungsliste neu aufgebaut und mit der Pseudoenergie der Optimierung  $E^{\text{total}}$  bewertet. Sowohl Neuaufbau als auch die Berechnung der Pseudoenergie benötigt Rechenzeit. Deshalb wurden die Veränderungen der chemischen Verschiebungen auf die Zuordnung eines HSQC-Spektrums angepasst und optimiert. Es wurden zwei Arten von sogenannte MOVES eingeführt, nämlich solche die Vorwissen stark berücksichtigten und allgemeine, die Ausreißer besser erkennen können (Kap. 2.2.11). Für jedes Amidprotonen werden in der Vorbereitung die wahrscheinlichsten Zielslots für den speziellen MOVE bestimmt. Die Pseudokraft eines

Zielslots  $s$  auf das Amidproton  $k$   $F^{\text{Pseudo}}$  (Slot  $s \rightarrow$  Amidproton  $k$ ) (Gl. 78) versucht unbesetzte experimentelle Signale zu besetzen und berücksichtigt auch die Wahrscheinlichkeitsdichteverteilungen der Kreuzsignalpartner des Amidprotons. Die zugehörige chemische Verschiebung der Resonanz des kovalent gebunden Stickstoffs wird anschließend auf den Wert desjenigen experimentellen Diagonalsignals des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum verschoben, das zum einen auf der vorher festgelegte chemischen Verschiebung des Amidprotons liegt und zum anderen die beste Pseudoenergie erzeugt. Die speziellen und allgemeinen MOVES verschieben die Resonanzen von  $\text{H}^\alpha$ - und Seitenkettenprotonen immer auf experimentelle Kreuzsignale mit Amidprotonen im 2-D-NOESY-Spektrum (Abb. 13).

Um das Minimum der Pseudoenergie der Optimierung zu finden, unterstützt SIBASA, im Gegensatz zu ASSIGN, den Einsatz von mehreren Instanzen mit unterschiedlichen Startzuordnungen und Zufallszahlen. Neben der besseren Ausnutzung der Rechenleistung von Mehrkernprozessoren und einer besseren Abtastung des Konfigurationsraums, werden die Instanzen zur Laufzeit regelmäßig synchronisiert, indem bereits richtig zugeordnete Resonanzen für die weitere Optimierung in allen Instanzen festgelegt werden. Durch diese Synchronisationen werden richtige Informationen zwischen den Instanzen ausgetauscht, die Rechenzeit auf noch zu optimierende chemische Verschiebungen konzentriert und außerdem die Volumenskalierungsfaktoren zwischen den experimentellen und zurückgerechneten Spektren noch zur Laufzeit verbessert (Kap. 2.2.14).

Der Einsatz mehrerer Instanzen zusammen mit der Pseudoenergie des Volumen- und Linienformvergleichs  $E_{ij}^{\text{Vol,Shape,sim}}$  auf den simulierten Grundflächen  $ij$  erlaubt die Bestimmung der Verifikationspseudoenergien  $E_k^{\text{ClusterVol,Shape,sim}}(\text{best})$  der Kerne  $k$  (Gl. 81) bzw.  $E_{\text{Amidgruppe } k}^{\text{ClusterVol,Shape,sim}}(\text{best})$  der Amidgruppen  $k$  (Gl. 85). Sie sind von der Pseudoenergie der Optimierung größtenteils unabhängig und werden auf den Grundflächen der zurückgerechneten Signale berechnet. Die Verifikationspseudoenergie gibt an, wie gut die Kreuzsignale eines Kerns  $k$  von der aktuellen vollständigen Zuordnung und den experimentellen Daten erklärt werden. Für die Bildung der Verifikationspseudoenergie werden die Pseudoenergien der Zuordnung  $E_k^{\text{Vol,Shape,sim}}$  (Gl. 70) aller Instanzen, in denen die Resonanz des Kern  $k$  die selbe chemische Verschiebung einnimmt, summiert. Die chemische Verschiebung, die dabei die höchste Summe erreicht, wird in die vereinigte vollständige Zuordnungsliste übernommen und die Summe Pseudoenergien als Verifikationskriterium übernommen. Zuordnungen, die von mehr unabhängig Instanzen gefunden wurden, werden damit insgesamt als sicherer eingestuft.

### 5.3.2 Vorbereitung der Röntgenstrukturen

Für den Top-Down-Ansatz ist es wichtig, dass die mit der richtigen vollständigen

Zuordnung zurückgerechneten Spektren die entsprechenden experimentellen Spektren bestmöglich abbilden. Nur exakte Rückrechnungen garantieren, dass die Minimierung der Pseudoenergie der Optimierung  $E^{\text{total}}$  zu der richtigen vollständigen Zuordnung führt. Röntgenstrukturen, die der automatischen Zuordnung als Basis dienen, liegen, im Gegensatz zum Protein in Lösung, nur in einer einzigen Zustand, der nicht notwendiger Weise der gleiche wie in Lösung ist, vor. Eine Aufgabe der MD-Simulation ist es die Dynamik des Proteins, also die Veränderungen der Struktur mit der Zeit im Lösungsmittel unter den Versuchsbedingung zu simulieren und die Ergebnisse der Rückrechnung zugänglich zu machen. Eine weitere Aufgabe der MD-Simulation ist durch gute Kraftfelder, gute Modelle und lange Simulationszeiten die Abweichung der Röntgenstruktur zum Zustand des Proteins in Lösung zu minimieren.

Für letztere Aufgabe sind die für diese Arbeit verwendete Simulationszeit von 100 ns zu wenig. Dies zeigt der Vergleich der RMSD-Werte der gemittelten MD-Ensembles, der gemittelten Modelle und der Röntgenstrukturen zu den zugehörigen gemittelten NMR-Ensembles (Tabelle 8, 10 und 12). Die Röntgenstruktur von HPr *S. faecalis* hat trotz der relativ geringen Sequenzidentität von 63,3 % einen RMSD von nur 0,084 nm zum NMR-Strukturensemble von HPr *S. aureus* (H15A). Durch Homologie Modellierung und MD-Simulation konnte dieser auf 0,058 nm verkleinert werden. Beim PfTrx wurde der RMSD durch Weiterverarbeitung der Röntgenstruktur von 0,125 nm auf 0,133 nm erhöht. Diese RMSD-Werte liegen aber noch unter den Auflösungen der verwendeten Röntgenstrukturen. Die zugehörigen Verläufe der lokalen RMSD Werte (Abb. 41 **gelb** und Abb. 56 **gelb**) zeigen ebenfalls keine Ausreißer nach oben an. Es kann als davon ausgegangen werden, dass die Konformation im Kristall und in der Lösung beim HPr *S. aureus* (H15A) und dem reduzierten PfTrx jeweils gleich sind.

Beim Ras(T35S)-GppNHp ist der RMSD der verwendeten Röntgenstruktur mit 0,22 nm höher und konnte mit der MD-Simulation auf 0,207 nm verkleinert werden. Außerdem hat der lokale RMSD der aktiven Zentren große Ausreißer nach oben (Abb. 70 **gelb**). Es standen zwei Formen der Röntgenstrukturen von Ras(T35S)-GppNHp zur Auswahl. Im Gegensatz zu Form 1 (PDB ID: 3KKK) hat Form 2 (PDB ID: 3KKM) vier Wasserstoffbrücken zum Nukleotid GppNHp (Shima et al. 2010). In der Rückrechnung des 2-D-NOESY-Spektrums von Form 2 konnten außerdem wichtige NOE-Kontakte zum Nukleotid, die in der Rückrechnung von Form 1 nicht sichtbar waren, beobachtet werden (Abb. 16). Form 2 hat einen kleineren RMSD zum NMR-Ensemble (Abb. 58), ist aber unvollständig und damit nicht als Eingabestruktur für SIBASA geeignet. Man kann davon ausgehen das Form 2 als Eingabestruktur für SIBASA besser geeignet gewesen wäre und das die 100 ns MD-Simulation von Form 1 die Unterschiede zwischen den beiden Formen nicht beheben konnte.

Durch das Verwenden des Ensembles und der Dynamikparameter einer MD-Simulation kann die Rückrechnung das Protein im Lösungsmittel unter Versuchsbedingung besser abbilden. Die gesamte Pseudoenergie des Volumens  $E^{\text{Vol}}$  und die der Linienform  $E^{\text{Shape}}$  der richtigen Zuordnung haben sich mit den Rückrechnungen der MD-Trajektorie für alle drei

Proteine verbessert (Tabelle 8, 10 und 12). Beim Ras(T35S)-GppNHp waren die Pseudoenergien der richtigen Zuordnung für die aus dem MD-Ensemble zurückgerechneten Spektren im Mittel sogar kleiner als für die aus dem NMR-Ensemble zurückgerechneten Spektren.

MD-Ensembles bilden auch die Basis für die Bestimmung der Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der zuzuordnenden Kerne (Kap. 2.2.7). Durch Fortschritte in den Kraftfeldern und Modellen, sowie durch Einsatz von mehr Rechenzeit konnte die Qualität der Vorhersagen der chemischen Verschiebungen verbessert werden.

Es ist vorstellbar, dass Fortschritte auf dem Gebiet der Molekulardynamik es möglich machen die richtige vollständige Zuordnung mit SIBASA mit noch weiter von der Lösungsstruktur entfernten 3D-Strukturen zu bestimmen.

### 5.3.3 Automatische Bestimmung der Volumenskalierungsfaktoren und der zusätzlichen Linienverbreiterungen

Für die Bestimmung der Pseudoenergien des Volumens  $E_{ij}^{\text{Vol}}$  auf den Grundflächen der experimentellen Signale  $ij$  (Gl. 45) ist der Volumenskalierungsfaktor  $\alpha_i$  zwischen den Volumen der simulierten und experimentellen Signalen im Spektrum  $i$  unverzichtbar. Die Pseudoenergie des Volumens differenziert besser zwischen falschen und richtigen Zuordnungen von Kreuzsignalen als die Pseudoenergie der Linienform (Abb. 32, 47 und 61 jeweils A und C). Die Volumenskalierungsfaktoren gehen zudem über Signalüberlapp auch in die Berechnung der Pseudoenergie der Linienform  $E_{ij}^{\text{Shape}}$  (Gl. 47) auf der Grundfläche des Signals  $j$  im Spektrum  $i$  ein.

Die statistischen Methode aus Kap. 2.2.4 zur Bestimmung der Skalierungsfaktoren kommt ohne Zuordnungsliste aus. Sie schafft es durch Glättung und durch Filterung die Häufigkeitsverteilungen der skalierten simulierten und die Häufigkeitsverteilungen der experimentellen Volumen für die beiden NOESY-Spektren bestmöglich aneinander zu fitten (Abb. 5, 23, 30, 45 und 59). Die Abweichungen von bis zu 70 % zu den Volumenskalierungsfaktoren, die mit der richtigen vollständigen Zuordnung berechnet worden sind, sind z.B. auf die Unterschieden zwischen dem MD-Ensemble und der 3D-Struktur in Lösung zurück zu führen. Solche Unterschiede können manchmal dafür verantwortlich sein, dass kein eindeutiger Skalierungsfaktor zwischen experimentellen und simulierten Signalvolumen gefunden werden kann. Lokale Fehler im Eingangsstrukturensemble fallen aber in der Regel weniger stark ins Gewicht, da die Häufigkeitsverteilungen der Volumen aus fast alle sichtbaren Signale des Proteins erzeugt werden.

Eine stärkere Fehlerquelle bei der Bestimmung der Volumenskalierungsfaktoren über die Häufigkeitsverteilungen sind die zufälligen initialen chemischen Verschiebungen der zurückgerechneten Spektren in Verbindung mit einer schlechten Auflösung der experimentellen Spektren. Die initialen Zuordnungslisten werden zufällig aus den Normalverteilungen  $N(\bar{\delta}_{AA, Kern}^{BMRB}, \sigma_{AA, Kern}^{BMRB})$  ( $AA$  ist die Aminosäure, wie z.B. ALA,  $Kern$  ist die Kernart wie  $H^N$ ,  $N$ ,  $H^\alpha, \dots$ ) gebildet, wobei die Mittelwerte  $\bar{\delta}_{AA, Kern}^{BMRB}$  und die Standardabweichungen  $\sigma_{AA, Kern}^{BMRB}$  aus der BMRB stammen. Die statistisch bestimmten chemischen Verschiebungen des Proteins liegen in der Regel näher beieinander als die Resonanzen der Kerne in der Realität, was zum Verschmelzen von Signalen in den zurückgerechneten Spektren führt. Das durchschnittliche Volumen in den Häufigkeitsverteilungen der simulierten Volumen wird dadurch erhöht und die resultierenden statistisch bestimmten Volumenskalierungsfaktoren  $\alpha_i^{statistisch}$  verkleinert. Geringe Auflösungen wie sie z.B. oft auf der indirekten  $^1H$ -Dimension des 3D  $^{15}N$ -NOESY-HSQC-Spektrums zu finden sind, können diesen Effekt verstärken und die Fehler der Volumenskalierungsfaktoren  $\alpha_{3-D-NOESY}^{statistisch}$  von 3D  $^{15}N$ -NOESY-HSQC-Spektren auf bis zu 70 % erhöhen. Bei 2-D-NOESY-Spektren waren die Fehler der statistisch bestimmten Volumenskalierungsfaktoren dagegen im Bereich von 5 % - 50 %.

Für die gut aufgelöste 2-D-NOESY-Spektren erzeugen Volumenskalierungsfaktoren, die aus allen eindeutig zuordenbaren Signalen oder nur mit Ringkontakten berechnet wurden, geringere Pseudoenergien des Volumens  $E^{Vol}$  mit der richtigen Zuordnung als die statistisch bestimmten Volumenskalierungsfaktoren  $\alpha_{2-D-NOESY}^{statistisch}$ . Bei den schlechter aufgelösten 3D  $^{15}N$ -NOESY-HSQC-Spektren erzeugen hingegen die statistisch bestimmten Volumenskalierungsfaktor  $\alpha_{3-D-NOESY}^{statistisch}$  geringere Pseudoenergien des Volumens  $E^{Vol}$  (Tabelle 9, 11 und 13). Die Ursache ist folgende:

Bei den Rückrechnungen werden viele schwache Signale, die im experimentellen Spektrum nicht sichtbar sind, erzeugt. Diese Signale überlappen, wegen der nahe beieinander liegenden chemischen Verschiebungen in der initialen Liste der chemischen Verschiebung, oft mit stärkeren zurückgerechneten Signalen und erhöhen damit ihr Volumen. Ein kleinerer Volumenskalierungsfaktor bringt die skalierten simulierten Volumen wieder näher an die experimentellen Volumen heran und verbessert somit die Pseudoenergien des Volumens. Im gut aufgelösten 2-D-NOESY-Spektrum ist zwischen den experimentellen Signalen meistens genug leerer Raum, in dem die schwachen simulierten Signale Platz finden, ohne die Pseudoenergie des Volumens durch Überlapp zu beeinflussen. Eine möglich Abhilfe könnte die Verkleinerung des Cut-Off-Abstands der Rückrechnungen der 3D  $^{15}N$ -NOESY-HSQC-Spektren bringen. Dies würde aber dazu führen, dass experimentelle Signale während der automatischen Zuordnung nicht mehr erklärt werden können.

Die initialen Skalierungsfaktoren werden während der automatischen Zuordnung mithilfe der zur Laufzeit festgelegten chemischen Verschiebungen und Gleichung 30, dynamisch angepasst. Abb. 31, 46 und 60 zeigen die Entwicklung der Volumenskalierungsfaktoren für

die drei Testproteine während der automatischen Zuordnung. Mit der Anzahl der während der Synchronisationsphasen festgelegten chemischen Verschiebungen steigt auch die Anzahl der eindeutig zuordenbaren Signale. Die zur Laufzeit korrigierten Volumenskalierungsfaktoren nähern sich den mit der vollständigen Zuordnung bestimmten Volumenskalierungsfaktoren  $\bar{\alpha}_i^{\text{Ref}}$  schnell an und verbleiben bis zum Ende bei diesen Werten. Der maximale Fehler der Volumenskalierungsfaktoren kann somit von maximal 58 % auf maximal 15,6 % für das 2-D-NOESY-Spektrum und von maximal 69 % auf maximal 20 % beim 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum verringert werden. Da die Korrekturen der Volumenskalierungsfaktoren schon relativ früh erfolgen und fast alle bis dahin festgelegten chemischen Verschiebungen richtig sind, können alle weiteren Zuordnungen mithilfe der richtigen Skalierungsfaktoren bestimmt und bestätigt werden.

Auch die zusätzliche Linienverbreiterungen wurden mit der statistischen Methode bestimmt (Kap. 2.2.5). Die Häufigkeitsverteilungen der verbreiterten Linienbreiten stimmen besser als die Häufigkeitsverteilungen der nicht verbreiterten Linienbreiten mit den Häufigkeitsverteilungen experimentellen Linienbreiten überein (Abb. 30, 44 und 49). Wie die statistisch bestimmten Volumenskalierungsfaktoren hängen auch die statistisch bestimmten Linienverbreiterungen von der initialen Liste der chemischen Verschiebungen der Rückrechnung ab.

Der stärkere Signalüberlapp durch die nahe beieinander liegenden chemischen Verschiebungen sorgt dafür, dass Signale in den Spektren verschmelzen und höhere simulierte Linienbreiten in die Häufigkeitsverteilungen eingehen. Die bestimmten Korrekturfaktoren für die Linienbreiten der Rückrechnungen sind somit oft zu klein oder sogar negativ. Um Fehler zu vermeiden, wurden negative Korrekturen nicht berücksichtigt. Ein Vergleich der mit der richtigen (Tabelle 7) und der zufälligen Zuordnungsliste (Tabelle 9) bestimmten zusätzlichen Linienverbreiterungen im HPr *S. aureus* (H15A) betätigt diese Annahme. Die mit der Rückrechnung der vollständigen richtigen Zuordnung bestimmten zusätzliche Verbreiterung in der  $\omega_2$ -Dimension des 2-D-NOESY-Spektrums war fast doppelt so groß (2,1 Hz) wie die mit der zufälligen Liste der chemischen Verschiebungen bestimmte Verbreiterung (1,2 Hz).

In 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren war die Auflösung der indirekten  $^1\text{H}$ -Dimension ( $\omega_1$ -Dimension) acht mal bzw. beim Ras(T35S)-GppNHp vier mal größer als die Auflösung der indirekten  $^1\text{H}$ -Dimension des 2-D-NOESY-Spektrums. Wegen dem großen Überlapp, der durch die verwendeten FID-Filter verstärkt wurde, konnten nur kleine oder gar keine zusätzlichen Verbreiterungen bestimmt werden. Die Korrekturen der Linienbreiten verkleinerten die Pseudoenergien der richtigen Zuordnung nur wenig oder vergrößerten sie sogar.

Die für den Top-Down-Ansatz wichtigen Volumenskalierungsfaktoren und zusätzlichen Linienverbreiterungen mussten in ASSIGN noch manuell eingegeben werden, was zumindest Teile der richtigen Zuordnung erfordert hat. In SIBASA können die Volumenskalierungsfaktoren ohne bereits bekannte Resonanzzuordnungen mit der statistische Methode bestimmt werden. Durch die zusätzliche Anpassungen zur Laufzeit können die

Volumenskalierungsfaktoren sehr gut genähert werden. Die Verbreiterungen durch Feldinhomogenität können in der gut aufgelösten direkten  $^1\text{H}$ -Dimension der 2-D-NOESY-Spektrums ebenfalls gut bestimmt werden. Da in der neuesten Version von AUREMOL RELAX die Filterfunktionen auf der Frequenzdomäne (Kap. 2.2.3) genauer zurückgerechnet werden können, sind die Korrekturen nicht mehr so wichtig, wie sie noch in ASSIGN waren.

### 5.3.4 Schätzung der Häufigkeitsverteilung der Pseudoenergien

Für die Berechnung der lokalen Wahrscheinlichkeiten einer richtigen Zuordnung  $P_{ij}^{\text{Vol,Shape}}$  auf den Grundflächen der experimentellen Signale  $j$  im Spektrum  $i$  (Gl. 50) bzw. der lokalen Wahrscheinlichkeiten  $P_{ij}^{\text{Vol,Shape,sim}}$  auf den Grundflächen der simulierten Signale  $j$  im Spektrum  $i$  mit dem Satz von Bayes, müssen die Häufigkeitsverteilungen der Pseudoenergien des Volumens ( $E_{ij}^{\text{Vol}}$  bzw.  $E_{ij}^{\text{Vol,sim}}$ ) und die Häufigkeitsverteilungen der Pseudoenergien der Linienform ( $E_{ij}^{\text{Shape}}$  bzw.  $E_{ij}^{\text{Shape,sim}}$ ) für die vollständige richtige und für eine falsche Zuordnungen bekannt sein (Kap. 2.2.8). Für SIBASA wurden die Häufigkeitsverteilungen mithilfe der Spektren, der bekannten vollständigen Zuordnung und den Rückrechnungen der NOESY-Spektren aus dem NMR-Strukturensemble von HPr *S. aureus* (H15A) erzeugt. Die resultierenden Häufigkeitsverteilungen  $H_i^{\text{Vol,Shape}}(E_{ij}^{\text{Vol}}, E_{ij}^{\text{Shape}})$  und  $H_i^{\text{Vol,Shape,sim}}(E_{ij}^{\text{Vol,sim}}, E_{ij}^{\text{Shape,sim}})$  wurden in die AUREMOL Datenbank aufgenommen. Mit ihnen ist es möglich die lokale Wahrscheinlichkeit einer richtigen vollständigen Zuordnung des Signals  $j$  für die Spektrentypen  $i$  ( $i$  steht für 2-D-NOESY- bzw. 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum) in Abhängigkeit von der Pseudoenergie des Volumens und der Pseudoenergie der Linienform zu berechnen. Die Spektren sollten idealerweise für die automatische Zuordnung immer mit den gleichen Parametern aufgenommen und prozessiert werden. In Abb. 25 und 26 sind die 16 Häufigkeitsverteilungen für die Pseudoenergie des Volumens und die Pseudoenergie der Linienform, die für das 2-D-NOESY-Spektrum und das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum jeweils für die richtige und eine zufällige Zuordnung erzeugt worden sind, gezeigt. Die empfohlene Prozessierung der beiden NOESY-Spektren wird in Kap. 3.2.2 beschrieben. Es kann aber, wie z.B. bei der Lorentz-Gauß-Transformation des 2-D-NOESY-Spektrum von Ras (T35S)-GppNHp, nötig sein die Parameter zu verändern, um das Signal-zu-Rausch-Verhältnis zu erhöhen. Die Häufigkeitsverteilungen der Pseudoenergien, die mit der Rückrechnungen des NMR-Strukturensembles und der richtigen vollständigen Zuordnung bestimmt wurden, haben meistens zwei Maxima. Eines liegt bei der maximal möglichen Pseudoenergie 1,0 und repräsentiert die richtig zugeordneten NOESY-Kreuzsignale. Ein viel kleineres liegt bei sehr kleinen Werten der Pseudoenergie. Die Verteilungen der Pseudoenergien der zufälligen Zuordnung haben nur ein einziges Maximum bei sehr kleinen Pseudoenergien. Dieses Maximum repräsentiert experimentelle bzw. simulierte Signale denen kein simuliertes bzw. experimentelles Signal zugeordnet werden konnte. Um den dynamischen Bereich der lokalen Wahrscheinlichkeiten zwischen falschen und richtigen



Kreuzsignalzuordnung zu erhöhen, mussten die mit der vollständigen richtigen Liste der chemischen Verschiebungen nicht zuordenbaren experimentellen Signale herausgefiltert werden.

Die Höhe und die Position des Maxima der richtig zugeordneten Signale hängt vom Spektrumtyp ab. Wegen kleinerer Auflösungen und schlechteren Signal-zu-Rausch-Verhältnissen sind die Maxima der richtig zugeordneten Kreuzsignale in den Verteilungen der Pseudoenergien der 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren breiter und in Richtung kleinere Pseudoenergie verschoben.

Die Häufigkeitsverteilungen der Pseudoenergie des Volumens der richtigen vollständigen Zuordnung sind für die automatische Zuordnung meistens aussagekräftiger, als die Häufigkeitsverteilungen der Pseudoenergien der Linienform. Die Pseudoenergie des Volumens  $E_{ij}^{\text{Vol}}$  ist der relative Unterschied zwischen den aus experimentellen und simulierten Signalvolumen bestimmten Abständen (Gl. 45). Sie ist wegen der  $V^{-1/6}$  Abhängigkeit weniger empfindlich gegenüber Abweichungen als die Pseudoenergie der Linienform  $E_{ij}^{\text{Shape}}$ . Bei den Verteilungen der Pseudoenergien der 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektren verringert sich, wegen der ungenaueren Volumenskalierungsfaktoren (Kap. 5.3.3), dieser Vorteil.

Im Vergleich zu den entsprechenden Verteilung auf den experimentellen Grundflächen sind die Maxima der Häufigkeitsverteilungen der Pseudoenergien auf den simulierten Grundflächen der richtig zugeordneten Signale in der Regel niedriger und weiter in Richtung niedriger Pseudoenergie verschoben. Die Häufigkeitsverteilungen sind stark von der Linienform der zurückgerechneten Signale abhängig. Sind die simulierten Signale breiter als die entsprechenden experimentellen Signale, so können die Pseudoenergien nicht den Bestwert erreichen. Die Höhe des Maxima bei niedrigen Pseudoenergie in den Verteilungen sind hauptsächlich von der Güte des 3D-Strukturensembles der Rückrechnung und dem maximalen erlaubten Pseudoabstand  $d_{ij}^{\text{pseudo}}$  für ein simulierten NOESY-Kreuzsignal abhängig (Kap. 2.2.8).

Die gleichen Beobachtungen können auch für die Häufigkeitsverteilungen der drei Testfälle gemacht werden. In Abb. 33, 48 und 62 sind die lokalen Wahrscheinlichkeiten einer richtigen Zuordnung  $P(\text{korrekt} | E)$  in Abhängigkeit von der Pseudoenergie für die drei Testfälle aufgetragen. Zum Vergleich sind die aus den vorher diskutierten Häufigkeitsverteilungen bestimmten Wahrscheinlichkeitsfunktion aufgetragen. Für die Berechnung der lokalen Wahrscheinlichkeiten  $P_i^{\text{Vol,Shape}}(E_{ij}^{\text{Vol}}, E_{ij}^{\text{Shape}})$  sind immer zwei in den Abbildungen nebeneinander liegende Wahrscheinlichkeitsfunktionen verantwortlich. Alle acht Wahrscheinlichkeitsfunktionen (**rot**) stimmen in allen drei Testfällen sehr gut mit den Kurven der NMR-Daten von HPr *S. aureus* (H15A) (**grün**) überein. Die Wahrscheinlichkeitsfunktionen auf den experimentellen Grundflächen, die für die Berechnung der lokalen Wahrscheinlichkeiten bei unbekannter Zuordnung verwendet werden (jeweils **A**, **B**, **E**, **F**), gehen dabei bei niedrigen Pseudoenergien tiefer als die Kurven der Testfälle, weil

sie den Unterschied zwischen richtigen und falschen Zuordnungen bezüglich Wahrscheinlichkeit maximieren wollen.

Die Wahrscheinlichkeitskurven der Pseudoenergien auf den simulierten Grundflächen (**C**, **D**, **G**, **H**) der drei Testfälle haben etwas größere Abweichungen, aber auch hier ist die grundsätzliche Form die selbe. Die vier Kurven gehen in die Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  mit ein. Bessere Häufigkeitsverteilungen können in Zukunft dafür sorgen, dass die Verifikationspseudoenergien von richtigen und falschen Zuordnungen weiter auseinander liegen.

Durch das Einbinden der optimierten Häufigkeitsverteilungen der Pseudoenergien eines bekannten Proteins in der AUREMOL Datenbank wurde eine weitere Lücke von ASSIGN geschlossen. In ASSIGN musste für die Berechnung der lokalen Bayesschen Wahrscheinlichkeiten  $P^{\text{Vol,Shape}}(E_{ij}^{\text{Vol}}, E_{ij}^{\text{Shape}})$  zumindest ein Teil der chemischen Verschiebungen bereits vorhanden sein.

### 5.3.5 Verlauf von Pseudoenergien und dynamischer Größen während der automatischen Zuordnung

Während der automatischen Zuordnung mit SIBASA wird in mehreren unabhängigen Instanzen die Pseudoenergie der Optimierung  $E^{\text{total}}$  minimiert. In Abb. 34 **A**, 49 **A** und 63 **A** sind für die drei Testproteine die zeitliche Verläufe des wichtigsten Teils von  $E^{\text{total}}$ , nämlich die Pseudoenergie des Signalvergleichs auf den Grundflächen der experimentellen Signale für die Instanz mit der niedrigsten (**rot**) und der höchsten (**grün**) finalen Pseudoenergie  $E^{\text{total}}$  gezeigt. Es besteht dabei ein direkter Zusammenhang zwischen der Pseudoenergie der Signalvergleichs und den Anteil der richtig gefunden Zuordnungen (jeweils **B**). Dies bestätigt die Wahl der verwendeten Pseudoenergien für das Zuordnungsproblem mit dem Top-Down-Ansatz auf Basis von NOESY-Spektren.

Die finale Pseudoenergie des Signalvergleichs der automatischen Zuordnung kann schon vor der eigentlichen Optimierung gut mit Gl. 65 geschätzt werden (Abb. 35 **A**, 50 **A** und 64 **A** jeweils **gelb**). Wurde der Wert von  $E^{\text{Vol,Shape,opt}}$  von SIBASA (fast) erreicht oder unterschritten kann angenommen werden, dass die automatische Zuordnung erfolgreich war. Falls sie nicht erreicht wurde, kann davon ausgegangen werden, dass SIBASA nicht erfolgreich war. Ursachen für Fehler müssen dann bei den Eingangsdaten, wie z.B. dem Signal-zu-Rausch-Verhältnis der prozessierten NOESY-Spektren, dem Peakpicking, der 3D-Struktur der Rückrechnung oder bei den Optimierungsparametern wie Anfangsthreshold oder Anzahl der Iterationen pro Kühlschritt (Kap. 2.2.10) gesucht werden. Da die finale Pseudoenergie des Signalvergleichs gut vorhergesagt werden kann, ist es auch möglich den Fortschritt der automatischen Zuordnung zu jedem Zeitpunkt anzugeben. Außerdem können die Gewichtungsfaktoren der Pseudoenergien der chemischen Verschiebung  $E^\circ$  (Gl. 62) und des

Dummyslots  $E^{\text{Dummy}}$  (Gl. 72) dynamisch angepasst werden. In die Pseudoenergie der chemischen Verschiebung  $E^{\circ}$  gehen die vorhergesagte Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen ein.  $E^{\circ}$  wird für die richtige vollständige Zuordnung nicht minimal. Ihr Beitrag zur gesamten Pseudoenergie der Optimierung wird deshalb mit dem Gewichtungsfaktor  $w^{E^{\circ}}$  (Gl. 64) während der Optimierung verringert (Abb. 35 **B**, 50 **B** und 64 **B** jeweils **grün**), was der Pseudoenergie der chemischen Verschiebung  $E^{\circ}$  in allen drei Testfällen erlaubt (**grün**) gegen den Wert der richtigen vollständigen Zuordnung (**gelb**) zu konvergieren.

Die Pseudoenergie  $E^{\text{Dummy}}$  wurde eingeführt, um Amidprotonen und Stickstoffe, die keine sichtbaren Kreuzsignale in den NOESY-Spektren haben, während der automatischen Zuordnung in den Dummyslot, zu verschieben (Kap. 2.2.9). Der Gewichtungsfaktor  $w^{E^{\text{Dummy}}}$  der Pseudoenergie  $E^{\text{Dummy}}$  wird während der automatischen Zuordnung dynamisch verkleinert, wodurch verhindert wird das sichtbare Resonanzen von Amidprotonen und Stickstoffen irrtümlich als unsichtbar eingestuft werden. Die Identifikation der nicht sichtbaren Resonanzen von Amidgruppen erfolgt in SIBASA aber in der Regel über die Verifikationspseudoenergie. Dennoch sind viele sichtbare und schwer zuordenbare Resonanzen von Amidgruppen in vielen Instanzen der automatischen Zuordnung von Ras(T35S)-GppNHp auch am Ende im Dummyslot gewesen. Dies hat dafür gesorgt, dass die Zuordnungen dieser Resonanzen mithilfe der Verifikationspseudoenergie besser als unsicher eingestuft werden konnten.

### 5.3.6 Festlegen von Zuordnung zur Laufzeit mithilfe mehrerer Instanzen

Nach jeweils fünf Abkühlritten wurde für die automatische Zuordnung der drei Testproteine das Synchronisationsverfahren aus Kap. 2.2.13 verwendet um Zuordnungen festzulegen. Durch die Synchronisationen wurden richtig gefundene Zuordnungen zwischen den Instanzen ausgetauscht und dafür gesorgt, dass keine weitere Rechenzeit an bereits optimal platzierte Resonanzen mehr verschwendet werden musste.

Bei der automatischen Zuordnung mit dem Top-Down-Ansatz gibt es Resonanzen deren chemische Verschiebungen leicht, welche deren chemische Verschiebungen schwieriger und welche deren chemische Verschiebungen nicht gefunden werden können. In welche Kategorie die Zuordnung einer Resonanz fällt hängt von der 3D-Struktur des Proteins, der Qualität der Spektren und anderen Faktoren wie Signalüberlapp oder chemischen Austausch ab. Durch das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum sind die chemischen Verschiebungen von Amidprotonen viel besser definiert und erheblich leichter zuzuordnen als die chemischen Verschiebungen der Resonanzen von  $\text{H}^{\alpha}$ - und Seitenkettenprotonen.

Die chemischen Verschiebungen der Resonanzen von  $\text{H}^{\alpha}$ - und Seitenkettenprotonen wurden

zur Laufzeit fixiert, falls ihre Verifikationspseudoenergie  $E_k^{\text{ClusterVol,Shape,sim}}(\text{best})$  höher als 0,85 war (Kap. 4.3.1). Aus den Häufigkeitsverteilungen der Verifikationspseudoenergie der falschen Zuordnungen des Endergebnisses (Abb. 37 A, 52 A und 66 A) und den ROC-Kurven (Abb. 38 A, 53 A und 67 A) der drei Proteine geht hervor, dass es keinen Grenzwert  $p < 1$  der Verifikationspseudoenergie für den nur noch richtige Zuordnungen fixiert werden, gibt.

Die  $H^\alpha$ - und Seitenkettenprotonen haben im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum nur wenige und im 2-D-NOESY-Spektrum viele sich überlappende Kreuzsignale, weswegen deren Zuordnung schwieriger als die Zuordnung der Resonanzen von Amidgruppen ist. Ein Großteil der falsch fixierten chemischen Verschiebungen von Seitenketten sind für viele Anwendungen wie Strukturrechnung und Strukturverfeinerung trotzdem verwendbar, da die falsch festgelegten Resonanzen fast ausschließlich auf Positionen von Resonanzen anderen Seitenkettenprotonen der selben Aminosäure gelegt wurden. Da diese Kerne in der 3D-Struktur nahe beieinander liegen, liegt der Fehler eines auf Basis der falschen Zuordnung bestimmten Abstandsrestraints im Toleranzbereich. Die vertauschten Kerne haben außerdem fast die gleichen Kreuzsignale und stören so den weiteren Verlauf der Optimierung nur wenig. Im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum können vertauschte Seitenkettenprotonen, wegen der schlechten Auflösung entlang der indirekten  $^1\text{H}$ -Dimension, sowieso häufig nicht voneinander unterschieden werden.

Die Festlegung von chemischen Verschiebungen von Amidgruppen zur Laufzeit verlief mit dem Grenzwert  $p_{\text{Amidgruppe}}$  der Verifikationspseudoenergie von 0,73 dagegen fehlerfrei.

### 5.3.7 Diskussion der Ergebnisse der automatischen sequentiellen Zuordnung

#### Die vereinigte Zuordnungsliste

Als letzter Schritt in der automatischen Zuordnung wurde die vereinigte Zuordnungsliste (Kap. 2.2.13) aus den 16 Optimierungs-Instanzen mit der maximalen gewichteten Standardabweichung  $\sigma^w$  von 0,03 ppm gebildet und jede einzelne Zuordnung bewertet. Der Anteil der richtigen Zuordnungen in der vereinigten Zuordnungsliste war höher als in jeder einzelnen SIBASA Instanz. Beim HPr *S. aureus* (H15A) stieg der Anteil von 90,1 % auf 91,8 %, beim PfTrx von 77,3 % auf 81,7 % und beim Ras(T35S)-GppNHp von 70,3 % auf 73,9 %.

Laut Güntert (2003) sind 90 % aller chemischen Verschiebungen nötig, um eine 3D-Struktur eines Proteins mit automatischen Methoden zu bestimmen. Nur die vereinigte Zuordnungsliste von HPr *S. aureus* (H15A) würde dieses Kriterium erfüllen, falls es möglich gewesen wäre alle falschen chemischen Verschiebungen aus dieser Liste zu entfernen. Für die beiden anderen Proteine ist der Anteil der richtig gefundenen chemischen Verschiebungen zu

klein, was aber nicht ausschließt, dass eine NMR-Struktur mittels Strukturverfeinerung aus der Röntgenstruktur bestimmt werden könnte. Der Anteil der für die Strukturverfeinerung verwendbaren chemischen Verschiebungen wird außerdem durch falschen Zuordnungen von Resonanzen von Seitenkettenprotonen, die mit Resonanzen der selben Seitenkette vertauscht worden sind, erhöht.

Wie zu erwarten war, nehmen mit der Größe der Proteine bei gleichen Eingabespektren die Anzahl der richtigen gefundenen Zuordnungen ab. Das Testprotein Ras(T35S)-GppNHp hat fast doppelt so viele zuzuordnende Resonanzen und fast doppelt so viele NOESY-Kreuzsignale wie das Testprotein HPr S. aureus (H15A). Außerdem sind die Signale des Ras(T35S)-GppNHp breiter als die Signale des HPr S. aureus (H15A). Dank der guten Definitionen der Amidgruppen in der automatischen Zuordnung durch das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum und der sehr guten Vorhersagen der chemischen Verschiebungen der Seitenketten- und  $\text{H}^\alpha$ -Protonen ist der Anteil der richtig gefundenen chemischen Verschiebung von Ras(T35S)-GppNHp nur 20 % kleiner als der von HPr S. aureus (H15A).

Beim Haupteinsatzgebiet von SIBASA, der automatischen Zuordnung von HSQC-Spektren, waren in der vereinigen Zuordnungsliste von HPr S. aureus (H15A) 100 % der Resonanzen der Amidprotonen und Stickstoffe, mit sichtbaren NOESY-Kreuzsignalen, richtig zugeordnet. Beim PfTrx waren es 95,1 % und beim Ras(T35S)-GppNHp 91,8 %. Die Verdoppelung der Komplexität des Problems hat den Anteil der richtigen gefundenen Zuordnung nur um 8,2 % reduziert.

## Die Verifikationspseudoenergie

SIBASA besitzt mit der Verifikationspseudoenergie  $E_k^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  ein Kriterium, um die Qualität einer Zuordnung zu bestimmen. Sie wurde für Amidgruppen und die restlichen Protonen unterschiedlich definiert und wurde für jede Zuordnung separat ausgewertet.

Die Häufigkeitsverteilungen der Verifikationspseudoenergien der falsch zugeordneten Resonanzen der  $\text{H}^\alpha$ - und Seitenkettenprotonen (Abb. 37 A, 52 A und 66 A jeweils **grün**) haben ein Maximum bei mittleren Verifikationspseudoenergien und fallen im gesamten Bereich nie auf 0 ab. Die Häufigkeitsverteilungen der Verifikationspseudoenergien richtig gefundenen Zuordnungen (**rot**) dagegen haben ein hohes Maximum bei der maximal möglichen Verifikationspseudoenergie von 1. Falls man alle falsche Zuordnungen von  $\text{H}^\alpha$ - und Seitenkettenprotonen im Endergebnis der automatischen Zuordnung mit der Verifikationspseudoenergie herausfiltern hätte wollen, wären auch keine richtigen Zuordnungen mehr übrig geblieben. Der Anteil der richtig gefundenen und bestätigten Zuordnungen war aber sehr hoch, wenn man wenige falsche Zuordnungen erlaubt hat. Die zugehörigen ROC-Kurven der Verifikationspseudoenergien der  $\text{H}^\alpha$ - und Seitenkettenprotonen

(Abb. 38 A, 53 A und 67 A) haben dieses Ergebnis bestätigt. Die Flächen unter den ROC-Kurven, die die Qualität eines Trennkriteriums messen, waren in allen drei Fällen über 80 % und auch unabhängig der Größe des automatisch zugeordneten Proteins.

Durch die Auswertung des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums in der automatischen Zuordnung waren die Positionen der Amidprotonen- und Stickstoffresonanzen besser festgelegt als die Positionen der restlichen Resonanzen. Die Verifikationspseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Cluster, Vols, Shape, sim}}(\text{best})$  einer Amidgruppe, die das Produkt der Verifikationspseudoenergie des Amidprotons und des Stickstoffs ist (Gl. 85), kann falsche und richtige HSQC-Signalzuordnungen in allen drei Testfällen sehr gut voneinander trennen. Die Häufigkeitsverteilungen der Verifikationspseudoenergien der richtigen und falsch gefundenen Zuordnungen in der vereinigten Zuordnungsliste (Abb. 52 C und 66 C jeweils **grün**) sehen zwar den ähnlich aus wie die entsprechenden Häufigkeitsverteilungen der Verifikationspseudoenergien der  $\text{H}^{\alpha}$ - und Seitenkettenprotonen. Sie fallen bei höheren Verifikationspseudoenergien aber auf 0 ab. Dies ist Voraussetzung für die automatische fehlerfrei Teilzuordnung eines HSQC-Spektrums. Das Ergebnis wird durch die modifizierten ROC-Kurven (Abb. 53 C und 67 C) bestätigt. Die modifizierte ROC-Kurve von PfTrx schneidet die y-Achse bei über 60 % richtig gefundenen Zuordnungen, die von Ras(T35S)-GppNHp bei über 80 %. Beim HPr *S. aureus* (H15A) wurden alle Resonanzen von Amidgruppen richtig zugeordnet. Die Fläche unter der ROC-Kurve war im schlechtesten Fall 93,2 % (PfTrx), was ein Indiz dafür ist, dass die Verifikationspseudoenergie richtige von falsche gefundene HSQC-Zuordnungen sehr gut voneinander trennen kann.

## Wahrscheinlichkeiten der richtigen Zuordnung

Da die Verifikationspseudoenergie nur eine relative und keine absolute Angabe über die Güte der Zuordnung machen kann, berechnet SIBASA die Wahrscheinlichkeit einer richtigen Zuordnung mit Hilfe des Satzes von Bayes (Kap. 2.2.15). Die dafür benötigten allgemein gültigen Häufigkeitsverteilungen der Verifikationspseudoenergien der richtigen und falsch gefundene Zuordnungen in SIBASA wurden mithilfe der vollständigen Listen der chemischen Verschiebungen, der NOESY-Spektren und der NMR-Strukturen von HPr *S. aureus* (H15A) und PfTrx erstellt.

Die resultierende Abhängigkeit der Wahrscheinlichkeit der richtigen Zuordnung von der Verifikationspseudoenergie, konnte nur für einen der drei Testfälle, nämlich für das Protein HPr *S. aureus* (H15A) (Abb. 38 B und 39 B), den realen Zusammenhang gut genug abschätzen. Für die beiden anderen Testfälle waren die mit dem Zusammenhang geschätzten Wahrscheinlichkeiten sowohl für die Zuordnungen der Resonanzen der  $\text{H}^{\alpha}$ - und Seitenkettenprotonen (Abb. 53 B und 67 B), als auch für die Resonanzen der Amidgruppen (Abb. 53 D und 67 D) zu hoch, was dazu führte dass falschen Zuordnungen eine Sicherheitswahrscheinlichkeit von 100 % erhalten haben. Die Ursache für die falsch

bestimmten Wahrscheinlichkeiten war, dass die Häufigkeitsverteilungen der Verifikationspseudoenergien, die als Berechnungsgrundlage dienten, von einem Zuordnungsproblem, das weniger komplex als die Zuordnungsprobleme von Ras(T35S)-GppNHp und PfTrx war, stammten.

Um absolut sichere Teilzuordnungen von HSQC-Spektren zu erhalten, mussten die Häufigkeitsverteilungen der Verifikationspseudoenergien für die Berechnung Wahrscheinlichkeit der richtigen Zuordnung mit dem Satz von Bayes anders berechnet werden. Unter der Annahme, dass Zuordnungen von Amidgruppen mit hohen Verifikationspseudoenergien ( $E_{\text{Amidgruppe } k}^{\text{Cluster, Vols, Shape, sim}}(\text{best}) > 0,7$ ) richtig sind und der Annahme, dass die Häufigkeitsverteilung der Verifikationspseudoenergien der richtigen Zuordnungen von Amidgruppen sich durch eine Superposition einer Normalverteilung und einer Lognormalverteilung (Gl. 92) darstellen lässt, konnten die benötigten Häufigkeitsverteilungen der Verifikationspseudoenergien für den Satz von Bayes aus den Ergebnissen der automatischen Zuordnung gefittet werden. Diese Methode berechnet zu niedrige Wahrscheinlichkeiten, hat aber erkannt, dass für die Amidgruppen von PfTrx ein höherer minimaler Grenzwert  $p_{\text{Amidgruppe}}$  der Verifikationspseudoenergie für sichere Zuordnungen als für die Amidgruppen von Ras(T35S)-GppNHp nötig war, obwohl von der Schwierigkeit des Zuordnungsproblems ausgehend eher der umgekehrte Fall zu erwarten gewesen wäre. Beim HPr *S. aureus* (H15A) hat der bestimmte Grenzwert  $p_{\text{Amidgruppe}}$  die meisten richtig zugeordneten Resonanzen der Seitenkettenamidgruppen, die in der Häufigkeitsverteilung der Verifikationspseudoenergien (Abb. 39 A) getrennt auf einem zweiten Maximum lagen, als falsch eingestuft.

### 5.3.8 Falsch bestätigte Resonanzen von Seitenketten

Beim PfTrx wurden mit dem in Kap. 4.3.1 festgelegten minimalen Grenzwert  $p$  der Verifikationspseudoenergie 25 der 306 chemische Verschiebungen von Resonanzen von  $H^{\alpha}$ - und Seitenkettenprotonen falsch zugeordnet. Die minimale Wahrscheinlichkeit einer richtigen Zuordnung  $P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}} = 0,85)$ , die mithilfe des in Abb. 28 B abgebildeten Graphen bestimmt wurde, betrug 99 %. 19 der 25 falsch zugeordneten Resonanzen wurden von SIBASA auf die Position einer Resonanz der selben Aminosäure gesetzt. Die Zuordnung zweier weiterer Resonanzen wurde von magnetisch gleichwertige Kerne gestört. Der Überlapp der Resonanzen von Ile15  $H^{\gamma 12}$  und  $H^{\gamma 13}$ , bzw. der Resonanzen von Lys67  $H^{\delta 2}$  und  $H^{\delta 3}$  hat dafür gesorgt, dass einer der beiden Resonanzen jeweils richtig, die andere dagegen falsch zugeordnet wurde. Außerdem wurden beim PfTrx die chemischen Verschiebungen der Resonanzen der im Mittel 0,33 nm voneinander entfernten Protonen Phe12  $H^{\zeta}$  und Phe26  $H^{\zeta}$  miteinander vertauscht.

Beim Ras(T35S)-GppNHp waren 12 der 311 auf dem Niveau

$P(\text{korrekt} | E_k^{\text{Cluster, Vol, Shape, sim}} = 0,85)$  fixierten chemischen Verschiebungen falsch. Fünf davon wurden auf die Position der Resonanzen von Kernen der selben Aminosäure gelegt. Bei dreien war der Überlapp der Signale zweier gleichwertiger Kerne für die falsche Zuordnung verantwortlich.

Beim HPr *S. aureus* (H15A) waren 2 der 339 festgelegten chemischen Verschiebungen falsch. Die Signale von Asn42  $H^{\alpha}$  lagen auf der Resonanzlinie des Lösungsmittels und die Resonanz von Ile33  $H^{\beta}$  hatten am Ende der automatischen Zuordnung die Position der Resonanz von Ile33  $H^{\gamma13}$  eingenommen.

### 5.3.9 Automatische Zuordnung von HSQC-Spektren

Beim PfTrx sind keine Seitenketten  $NH_2$ -Gruppen falsch und nur die Resonanzen der Amidprotonen von Val56 und Cys34 falsch zugeordnet worden. Es sind ebenfalls neun weitere chemischen Verschiebungen von Hauptkettenstickstoffen falsch zugeordnet worden.

SIBASA hat insgesamt 18 HSQC-Signale von Ras(T35S)-GppNHp falsch zugeordnet, wobei neun davon Signale von Seitenketten waren.

Im Allgemeinen wurde ein falsch zugeordnetes HSQC-Signal immer an die Position des HSQC-Signals einer anderen Amidgruppe gesetzt, was fast immer zur Folge hatte, dass an der Zielposition das Signal der richtigen Amidgruppe verdrängt wurde und dann ebenfalls falsch zugeordnet wurde. Dies führt zu Vertauschung der Zuordnungen von zwei Amidgruppen bzw. zu falschen Permutation mehrere Zuordnungen. Es hat sich bewährt mehrere Instanzen der automatischen Zuordnung zu verwenden, da die richtige Permutation der Zuordnung in den meisten Fällen am häufigsten vorkommt. Wenn ein HSQC-Signal von SIBASA falsch zugeordnet wurde, dann waren meistens auch die Permutationen mit dieser Zuordnung in allen Instanzen verschieden, wodurch sie leicht über die Verifikationspseudoenergie herausgefiltert werden konnte.

Die Ursachen für falsche automatisch gefundene Zuordnungen HSQC-Signalen waren folgende:

1. Die zurückgerechneten Kreuzsignale der Amidgruppen konnten zusammen mit den richtigen chemischen Verschiebungen nur teilweise durch die experimentellen NOESY-Spektren erklärt werden. Dies äußerte sich durch eine niedrige Volumen- und Linienformpseudoenergien der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol, Shape, sim}}(\text{Referenz})$  (Gl. 70). Ursachen dafür waren falsche (lokale) 3D-Strukturen, beim Peakpicking vergessene oder fehlende experimentelle Signale, starke Verbreiterung von Linien durch chemischen Austausch oder ein schlechtes Signal-zu-Rausch-Verhältnis. In diesen Fällen gibt es für die automatische Zuordnung keine eindeutig beste Lösung die zurückgerechneten Kreuzsignale der Amidgruppen auf die NOESY-Spektren zu legen. Die Verifikationspseudoenergien der Zuordnungen solcher Amidgruppen sind



meistens klein.

2. Amidgruppen, die nur wenige oder nur schwache Kreuzsignale in der indirekten  $^1\text{H}$ -Dimension des  $^{15}\text{N}$ -NOESY-HSQC-Spektrums hatten, waren schwerer zuzuordnen. Die Signale entlang dieser Dimension sind besonders wichtig, da sie in der Summenprojektion das HSQC-Signal ergeben. Im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum ist es möglich, dass Kreuzsignale wegen chemischen Austausch oder schlechterem Signal-zu-Rausch-Verhältnis, nicht sichtbar sind. Wegen der geringeren digitalen Auflösung der indirekten Dimension können außerdem Signale von Ausläufern, die oft von richtigen Signalen schwer zu unterscheiden sind, hinzu kommen. Resonanzen von Amidgruppen, die wegen dieser Probleme falsch zugeordnet wurden, haben oft gleichzeitig eine geringen Pseudoenergie der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$ . Besonders stark ist der spektrale Bereich der Seitenkettenamidgruppen davon betroffen. Die vielen starken Diagonal- und gegenseitigen  $\text{NH}_2$ -Kreuzsignale auf kleinem spektralem Bereich erzeugen Ausläufer, die ähnliche Intensitäten wie die intraresidualen  $\text{H}^\beta$ - oder  $\text{H}^\gamma$ -Kreuzsignale haben. Durch die Lage der  $\text{NH}_2$ -Gruppen am äußeren Ende der Seitenkette kann es zusätzlich vorkommen, dass die Seitenkettenamidprotonen kaum starke interresiduale NOEs haben. Aus diesem Grund ist es möglich, dass sie normale Wert für die Pseudoenergien der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$  haben, aber im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums nicht richtig zugeordnet werden können. Die Verifikationspseudoenergie am Ende der automatischen Zuordnung ist für solche Amidgruppen in der Regel ebenfalls sehr klein.
3. Es gibt insgesamt mehrere gleichwertige Möglichkeiten das HSQC-Signal einer Amidgruppe zuzuordnen. In mehreren Instanzen von SIABSA ist die Zuordnung solcher Amidgruppen in der Regel verschieden, was dazu führt, dass die Verifikationspseudoenergie einer einzigen dieser entarteten Zuordnungsmöglichkeiten verkleinert wird. Die Identifikation einer falschen Zuordnung wird schwierig, falls zwei Möglichkeiten ungefähr gleich große Verifikationspseudoenergien haben und die Summe der beiden Verifikationspseudoenergien ungefähr 1 ergibt. Dies führt dazu, dass die richtige Zuordnung nicht in die vereinigte Zuordnungsliste aufgenommen wird und die falschen Zuordnungen eine zu hohe Verifikationspseudoenergie hat, um sie einfach herausfiltern zu können. Ursachen für diese Entartung können zu viele oder fehlende experimentelle Signale oder Signalüberlapp im HSQC-Spektrum oder auf der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums sein.

In den Tabellen 14 und 15 sind die falsch zugeordneten Amidgruppen und die Amidgruppe(n), die in mit der richtigen Zuordnung auf der Position im HSQC-Spektrum zu finden sind, von PfTrx und Ras(T35S)-GppNHp aufgelistet. Die falschen Zuordnungen in den Tabellen sind nach den drei möglichen Fehlerquellen getrennt. In Spalte 3 ist die Pseudoenergie der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$  angegeben. Falls diese viel kleiner als der durchschnittliche Wert ist, kann davon ausgegangen werden, dass die

simulierten Kreuzsignale der Amidgruppe nicht mit den experimentellen NOESY-Spektren erklärt werden können (siehe Grund 1). Daneben in Spalte 4 steht die Pseudoenergie der falschen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{falsch})$ , die im Unterschied zur Pseudoenergie der richtigen Zuordnung aus einer Instanz des besten Clusters mit der falschen Zuordnung entnommen wurde. Ist  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{falsch})$  größer als  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$ , so erklärt die falsche Zuordnung die experimentellen Daten besser als die richtige Zuordnung. In den Spalten 5 und 6 sind die Verifikationspseudoenergien des falschen besten Clusters  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  und  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{Referenz})$  des Clusters der richtigen Zuordnung aufgelistet. Kleine Verifikationspseudoenergien sind in diesen Fällen für die Trennung von falschen und richtigen Zuordnungen erwünscht. Mittlere und hohe Verifikationspseudoenergien  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  von falschen Zuordnungen könnten in Zukunft durch tiefere Analyse aller Instanzen der automatischen Zuordnung in der Glaubwürdigkeit abgewertet werden (siehe Grund 3).

In den letzten drei Spalten ist der Beitrag der falschen Zuordnung auf die Pseudoenergie des Volumen- und Linienformvergleichs  $E^{\text{Vol,Shape}}$  der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums angegeben. Alle Signale entlang einer Linie in dieser Dimension ergeben in der Summenprojektion ein HSQC-Signal. Zur lokalen Bewertung der Zuordnung einer Amidgruppe  $k$  anhand eines einzigen gemessenen HSQC-Signals wird die experimentelle Pseudoenergie des HSQC-Signals  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  eingeführt (Gl. 95).

$$E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC}) = \sum_{j=1}^{\text{Kreuzsignale entlang von } \omega 1} \left( \left| \ln(0,05) \right| - \left| \ln(w_{2j}^{\text{Signal,exp}} P_{2j}^{\text{Vol,Shape}}) \right| \right) \quad (95)$$

Je höher die Pseudoenergie des HSQC-Signals einer Amidgruppe ist, desto mehr der experimentelle 3D  $^{15}\text{N}$ -NOESY-HSQC-Kreuzsignale erklären die momentane Zuordnung dieser Amidgruppe. Sehr hohe Werte deuten auf den Überlapp mehrerer HSQC-Signale an der Position der betrachteten Zuordnung im experimentelle HSQC-Spektrum hin. Niedrige Werte bedeuten, dass das zurückgerechnete und das experimentelle Spektrum entlang der indirekten  $^1\text{H}$ -Dimension der betrachteten Zuordnung der Amidgruppe nicht zusammenpassen oder, dass nur wenige schwache experimentelle NOESY-Kreuzsignale entlang der Linie vorhanden sind. Ein Beispiel für kleine Werte sind die Amidgruppen von Ras(T35S)-GppNHp, die keine sichtbaren experimentellen NOESY-Kreuzsignale, sondern nur ein Diagonalsignal und ein Kreuzsignal mit dem Lösungsmittel im 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum haben. Mit der richtigen Zuordnung haben diese Amidgruppen eine Pseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  von 0, da das Diagonalsignal und das Kreuzsignal mit dem Lösungsmittel nicht zur Pseudoenergie beitragen. In der 7. Spalte sind die Pseudoenergien der HSQC-Signale  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  der Amidgruppen der Spalten 1 und 2 mit der richtigen Zuordnung eingetragen. In Spalte 8 sind die Werte der Pseudoenergie der HSQC-Signale  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  der beiden Amidgruppen mit der falschen von SIBASA

gefunden Zuordnung gegeben. In Spalte 9 ist die Differenz, und damit der Effekt der falschen Zuordnung auf die Pseudoenergie der Optimierung angegeben. Negative Werte bedeuten, dass die richtige Zuordnung die beiden HSQC-Signale besser erklärt und positive Werte, dass die falsche Zuordnung die experimentellen Daten besser erklärt. Werte Nahe bei 0 bedeuten, dass beide Konfigurationen energetisch ungefähr gleichwertig sind und die Populationen der falschen und richtigen Lösung in den Instanzen der automatischen Zuordnung ähnlich hoch sein können.

## Diskussion der falsch gefunden Zuordnungen von Amidgruppen im PfTrx

In Tabelle 14 sind die falschen HSQC-Zuordnungen des PfTrx aufgelistet. Die Rückrechnung von Cys34 im aktiven Zentrum konnte mit den richtigen chemischen Verschiebungen nicht gut erklärt werden ( $E_{\text{Cys34}}^{\text{Vol,Shape,sim}}(\text{Referenz}) = 0,51$ ). Cys34 hatte außerdem entlang der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums nur wenige schwache Kreuzsignale ( $E_{\text{Cys34}}^{\text{Vol,Shape}}(\text{HSQC}) = 10$ ). Es nahm in den 16 Instanzen verschiedene Positionen ein, aber niemals die Richtige ( $E_{\text{Cys34}}^{\text{Cluster,Vol,Shape,sim}}(\text{Referenz}) = 0$ ).

Gln98 hatte ebenfalls nur wenige schwache NOESY-Kreuzsignale entlang der  $^1\text{H}$ -Dimensiones 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums ( $E_{\text{Gln98}}^{\text{Vol,Shape}}(\text{HSQC}) = 13$ ) und konnte ohne große Veränderung der Pseudoenergie der Optimierung von SIBASA an die Position des benachbarten Leu99 verschoben werden. Mit Lys35, Met51, Glu66 und Leu99 bildeten Cys34 und Gln98 einen Cluster, in dem die Zuordnungen untereinander vertauscht waren. Bis auf Cys34 hatten alle fünf HSQC-Signale des Clusters ungefähr die gleiche richtige chemische Verschiebung des Amidprotons von SIBASA zugeordnet bekommen. Die Rückrechnungen von Lys35, Met51, Glu66 und Leu99 hätten auch mit der richtigen Zuordnung die experimentellen NOESY-Spektren gut erklären können. Sie hatten auf der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums genügend Signale, doch ist der Unterschied in der Pseudoenergie der Optimierung zwischen falscher und richtiger Lösung klein genug, dass der Anteil der falschen Lösung in den 16 Instanzen von SIBASA höher war als der Anteil der richtigen Lösung. Die Tatsache, dass die Verifikationspseudoenergien des falschen besten Clusters  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$  und des richtigen zweitbesten Clusters  $E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{Referenz})$  beide hoch waren, kann aber in Zukunft zusammen mit einer tieferen Analyse der einzelnen Instanzen ausgenutzt werden, um solche falsche Zuordnungen am Ende der automatischen Zuordnung abzuwerten. Auch Phe40, Ile54, Lys55 und Val56 sind in den 16 Instanzen sowohl falsch als auch richtig zugeordnet worden, da der Unterschied in der Pseudoenergie der Optimierung der beiden Lösungen nicht groß genug war.

Falsche Zuordnung	Ziel	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{falsch})$	$E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{best})$	$E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,sim}}(\text{Referenz})$	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$	$\Delta$
Cys34	Gln98	<b>0,51</b>	0,63	0,21	0	<b>10/13</b>	16/11	+4
Gln98	Leu99	1,00	0,81	0,24	0,05	<b>13/19</b>	14/16	-2
Lys35	Gln98	0,89	0,64	<b>0,25</b>	<b>0,11</b>	17/13	16/16	+2
Phe40	<b>Ile16/Thr65</b>	0,99	0,93	<b>0,41</b>	<b>0,24</b>	29/23	27/28	+3
Met51	Glu66	0,97	0,84	<b>0,44</b>	<b>0,36</b>	27/18	14/25	-6
Ile54	<b>Lys55/Glu20</b>	0,85	0,64	<b>0,61</b>	<b>0,17</b>	21/24	26/22	+3
Lys55( <b>Glu20</b> )	Ile54	0,97	0,85	<b>0,40</b>	<b>0,30</b>	24/21	24/26	+5
Val56	Asp25	0,90	0,94	<b>0,58</b>	<b>0,36</b>	29/26	29/26	0
Glu66	Gln98	0,94	0,89	<b>0,28</b>	<b>0,17</b>	18/13	16/17	+2
Met73( <b>Ala104</b> )	Lys80	0,85	0,72	0,15	0	30/26	29/30	+3

Tabelle 14: *Erklärung der falschen automatisch mit SIBASA gefundenen Zuordnungen von HSQC-Signalen im PfTrx.* In der 2. Spalte ist die richtige Zuordnung(en) des Signals angegeben. In Klammern sind andere Amidgruppen, die sich mit der falsch zugeordneten Amidgruppe im experimentellen Spektrum überlappen, angegeben. In den Spalten 3 und 4 sind die Volumen- und Linienformpseudoenergien der Zuordnung der Amidgruppe  $k$  der richtigen und der falschen Zuordnung aufgelistet. In den Spalten 5 und 6 sind die Verifikationspseudoenergien der 16 Instanzen des besten falschen Clusters und des Clusters mit der richtigen Lösung eingetragen. In den Spalte 7 und 8 sind für die vollständige richtige Zuordnung und für eine Instanz des besten falschen Clusters die experimentellen Pseudoenergien des HSQC-Signals entlang der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  (Gl. 95) für die Amidgruppen in den Spalten 1 und 2 angegeben. In Spalte 9 steht die Differenz  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})(\text{best}) - E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{Referenz})$  der beiden Pseudoenergien. In fetter Schrift sind jeweils die Größen oder überlappenden Zuordnungen, die zur falschen automatischen Zuordnung in SIBASA geführt haben, markiert.

Der Mittelwert von  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$  der von SIBASA richtig gefundenen Zuordnungen von HSQC-Signalen war 0,89 und die Standardabweichung 0,15. Der Mittelwert von  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  der vollständigen richtigen Zuordnung war 20 und die Standardabweichung war 7.

Die HSQC-Signale von Met73 und Ala104 haben sich überlappt und konnten die Rückrechnungen der NOESY-Signale nicht optimal erklären ( $E_{\text{Met73}}^{\text{Vol,Shape,sim}}(\text{Referenz}) = 0,85$ ). Die richtige Zuordnung von Met73 konnten von keiner Instanz der automatischen Zuordnung gefunden werden, aber auch die falsche Lösung wurde nur mit einer niedrigen Verifikationspseudoenergie bestätigt. Der Überlapp zweier HSQC-Signale spielte auch bei der falschen automatischen Zuordnung von Ile54 und Lys55 eine Rolle, da Glu20 und Lys55 die gleiche Position im HSQC-Spektrum hatten. Außerdem konnte SIBASA für die Zuordnung der Amidgruppe von Phe40 ein energetisch günstiges Ziel im gemeinsamen HSQC-Signal von Ile16/Thr65 finden.

Falsche Zuordnung	Ziel	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}$ (Referenz)	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}$ (falsch)	$E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,s}}$ (best)	$E_{\text{Amidgruppe } k}^{\text{Cluster,Vol,Shape,s}}$ (Ref)	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}$ (HSQC)	$E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}$ (HSQC)	$\Delta$
Val29	Gly13	<b>0,49</b>	0,40	0,06	0,03	18/9	9/20	+2
Asp38	Asp108	<b>0,32</b>	0,65	0,27	0,11	<b>12/28</b>	26/12	-2
Thr50	Glu63	<b>0,35</b>	0,29	0,12	0	14/12	12/0	-14
Glu63	<b>Gln61/Tyr64/Asp132</b>	<b>0,57</b>	0,93	0,41	0	<b>12/41</b>	42/11	0
Asp30	Glu37	0,78	0,52	0,09	0,04	<b>9/30</b>	24/9	-6
Lys117	Asn116 H <sup>δ22</sup>	0,96	0,60	0,39	0	<b>9/17</b>	20/9	+3
Ala121	Ile36	0,68	0,64	0,04	0	<b>9/24</b>	19/6	-8
Asn26 H <sup>δ21</sup>	Gln70 H <sup>e21</sup>	<b>0,21</b>	0,41	0,19	0,02	<b>6/8</b>	10/6	+2
Gln43 H <sup>e21</sup>	Gln150 H <sup>e21</sup>	0,83	0,47	0,03	0	<b>9/9</b>	9/12	+3
Gln43 H <sup>e22</sup>	Gln150 H <sup>e22</sup>	0,88	0,55	0,03	0	<b>12/9</b>	12/12	+3
Gln61 H <sup>e21</sup>	Gln150 H <sup>e21</sup>	<b>0,46</b>	0,57	0,1	0	<b>6/9</b>	9/9	+3
Gln61 H <sup>e22</sup>	Gln150 H <sup>e22</sup>	<b>0,31</b>	0,71	0,1	0	<b>6/9</b>	12/5	+2
Asn86 H <sup>δ21</sup>	Gln131 H <sup>e21</sup>	<b>0,11</b>	0,64	0,31	0	<b>3/9</b>	9/0	-3
Asn86 H <sup>δ22</sup>	Gln131 H <sup>e22</sup>	<b>0,15</b>	0,56	0,19	0	<b>3/8</b>	9/0	-2
Asn116 H <sup>δ22</sup>	Lys117	0,69	0,64	0,19	0	17/9	9/19	+2
Asn116 H <sup>δ21</sup>	<b>Thr144/Ser145</b>	0,74	0,61	0,08	0	17/50	9/54	<b>-4</b>
Gln61 (Tyr64/Asp132)	Gln22	0,73	0,69	0,04	0	41/30	24/37	-10
Tyr64 (Asp132/Gln61)	Val152	0,72	0,51	0,04	0	42/36	36/42	0

Tabelle 15: *Erklärung der falschen automatisch mit SIBASA gefundenen Zuordnungen von HSQC-Signalen im Ras(T35S)-GppNHp. Die Beschreibung ist die gleiche wie in Tabelle 14. Der Mittelwert von  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}$  (Referenz) der von SIBASA richtig gefundenen Zuordnungen von HSQC-Signalen war 0,84 und die Standardabweichung 0,17. Der Mittelwert von  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}$  (HSQC) der vollständigen richtigen Zuordnung war 26 und die Standardabweichung 11.*

Durch Abwertung der Zuordnungen von Amidgruppen, die in den Instanzen der automatischen Zuordnung nur in einem einzigen falschen und dem richtigen Zustand vorliegen, könnte der minimale benötigte Grenzwert  $p_{\text{Amidgruppe}}$  der Verifikationspseudoenergie für sichere Zuordnungen von 0,61 auf 0,24 abgesenkt werden. Statt 72,2 % könnten 89,6 % der Resonanzen der Amidgruppen sicher richtig zugeordnet werden. Die einzigen verbleibenden richtig gefundenen Zuordnungen, die nicht bestätigt werden könnten, wären die am N-Terminus liegende und über NOESY-Signale nur schwer zu identifizierenden Amidgruppe von Gln2 und die Amidgruppe von Thr65, deren HSQC-Signal sich mit dem der Amidgruppe von Ile6 überlappt und außerdem durch die falsche Zuordnung der Amidgruppe von Phe40 im Laufe der Optimierung gestört wurde.

## Diskussion der falsch gefundenen Zuordnungen von Amidgruppen im Ras(T35S)-GppNHp

Die Auflösung des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum von Ras(T35S)-GppNHp war mit 256 Datenpunkte mehr als doppelt so hoch, wie beim PfTrx. Falsch zugeordnete Resonanzen von Stickstoffen, bei denen die Resonanz des zugehörigen Amidprotons richtig zugeordnet wurden, traten nicht mehr auf. In Tabelle 15 werden die falschen Zuordnungen von HSQC-Signalen im Ras(T35S)-GppNHp diskutiert.

Die durchschnittliche Pseudoenergie einer richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$  war für eine Amidgruppe, die mit SIBASA richtig gefunden wurde, 0,84 und die zugehörige Standardabweichung war 0,17. Die Pseudoenergie  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$  der Amidgruppen von Val29, Asp38, Thr50, Glu63 und der Seitenketten von Asn26, Gln61 und Asn86 unterschritten diesen mittleren Wert deutlich, was bedeutet, dass die Rückrechnungen der entsprechenden NOESY-Kreuzsignale durch die experimentellen Spektren nicht gut genug erklärt werden konnten. Bis auf die Hauptkettenamidgruppe von Thr50 und die Seitenkettenamidgruppe von Asn86 lagen diese Amidgruppen in einem Bereich, in dem sich das von SIBASA verwendete MD-Ensemble stark vom NMR-Strukturensemble unterschieden hat (Abb. 70 A). Die falsche Lösung verbessert für die meisten Amidgruppen (alle bis auf Val29 und Thr50) die Pseudoenergie der Zuordnung ( $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz}) < E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{falsch})$ ). Auch Ala121 und Asn116 haben relativ niedrige Pseudoenergien der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape,sim}}(\text{Referenz})$ , was aber nicht die einzige Ursache für diese falsch Zuordnungen war. Von den falsch Zuordnungen, die auf Ursache 1 zurückzuführen sind, sind nur die von Asp38 und Glu63 kritisch, da sie mit einer hohen Verifikationspseudoenergie bestätigt wurden.

Die Amidgruppen von Asp30, Lys117, Ala121 sowie fast alle falsch zugeordneten Seitenkettenamidgruppen (alle bis auf Asn116) hatten wenige Kreuzsignale entlang der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums von Ras(T35S)-GppNHp. Die Werte der experimentellen Pseudoenergie des HSQC-Signals der richtigen Zuordnung  $E_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC})$  (Gl. 95) sind weit unter dem durchschnittlichen Wert für erfolgreich zugeordnete Amidgruppen ( $\bar{E}_{\text{Amidgruppe } k}^{\text{Vol,Shape}}(\text{HSQC}) = 26$  und  $\sigma = 11$ ). Asp30 liegt in einem Bereich der 3D-Struktur, in dem sich das MD-Ensemble stark vom NMR-Ensemble unterscheidet. Es hat eine hohe Pseudoenergie der richtigen Zuordnung  $E_{\text{Asp30}}^{\text{Vol,Shape,sim}}(\text{Referenz})$ , weil die Signale von Val109 und Ala122 in unmittelbarer Nachbarschaft liegen und seine zurückgerechneten Kreuzsignale mit erklären. Lys117 liegt im spektralen Bereich der  $\text{NH}_2$ -Gruppen, in dem Artefakte häufig sind. Ein vollständiges Peakpicking war in diesem Bereich nicht möglich, was sich durch kleine experimentelle

Pseudoenergie der HSQC-Zuordnung  $E_{\text{Lys117}}^{\text{Vol,Shape}}(\text{HSQC})$  feststellen lässt. Die Zuordnung von Lys117 ist wegen ihrer hohen Verifikationspseudoenergie die einzige kritische Zuordnung, die wegen Grund 2 falsch zugeordnet wurde.

Durch die falsche Zuordnung einer Amidgruppe von Ras(T35S)-GppNHp kann die Pseudoenergie der Optimierung  $E^{\text{total}}$  fast immer verbessert werden. Die Summe der beiden von der falschen Zuordnung einer Amidgruppe betroffenen lokalen Pseudoenergie des Volumen- und Linienformvergleichs  $E_k^{\text{Vol,Shape}}(\text{HSQC})$  entlang indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums ist für die falsche Zuordnung immer kleiner oder kaum größer als für die richtige Zuordnung. Für das HSQC-Signal der Seitenkettenamidgruppe Asn116  $\text{H}^{\delta 22}$  ist dies ein Grund für die falsche Zuordnung. Außerdem liegt das HSQC-Signal von Lys117, im Gegensatz zu dem von Asn116  $\text{H}^{\delta 22}$ , im spektralen Bereich der Seitenkettenamidgruppen von Ras(T35S)-GppNHp. Dies hat zur Vertauschung der beiden Zuordnungen der HSQC-Signale geführt.

Das zugehörige HSQC-Signal von Asn116  $\text{H}^{\delta 21}$  wurde an die Position des gemeinsamen Signals von Thr144/Ser145 gelegt. Die experimentelle NOESY-Kreuzsignale von Thr144/Ser145 konnten wegen dem Überlapp auch die Rückrechnungen der Signale von Asn116  $\text{H}^{\delta 21}$  ausreichend gut erklären. Der starke Überlapp im Bereich der HSQC-Signale von Gln61/Tyr64/Asp32 führte dazu, dass Gln61 und Tyr64 falsch zugeordnet wurden, wobei Thr50 ebenfalls an der Störung beteiligt war.

Das größte Optimierungspotential in SIBASA für die automatische Zuordnung des HSQC-Spektrums Ras(T35S)-GppNHp bietet wieder die Verifikation. Um die 15 falschen HSQC-Signale herauszufiltern, mussten auch 42 richtige zugeordnete Signale verworfen werden (Abb. 71). Hätten die falschen Zuordnungen der Hauptketten von Asp38, Glu63 und Lys117 geringere Verifikationspseudoenergien erhalten, so könnte der Grenzwert  $p_{\text{Amidgruppe}}$  erniedrigt werden, und 23 zusätzliche Hauptkettensignale bestätigt werden. Die vollständig richtige Teilzuordnung des HSQC-Spektrums würden 91,3 % statt 77,3 % der Hauptkettensignale enthalten. Nur noch vier von SIBASA richtige gefundene Zuordnungen von Hauptkettensignale hätten zu kleine Verifikationspseudoenergien um als sicher bestätigt zu werden (Thr2, Gly13, Asp107, Asp108). Thr2 am N-Terminus des Proteins ist nicht stark über NOE-Kontakte definiert, die automatische Zuordnung der Amidgruppe von Gly13 wurde von der falschen Zuordnung von Val29 und die automatische Zuordnung von der Amidgruppe von Asp108 von der falschen Zuordnung von Asp38 gestört. Asp107 hat nur eine kleine Pseudoenergie des HSQC-Signals (Gl. 197) auf der indirekten  $^1\text{H}$ -Dimension des  $^{15}\text{N}$ -NOESY-HSQC-Spektrums und ist deshalb nur schwer zuzuordnen (siehe Grund 2). Von den falschen Zuordnungen der Seitenketten hat nur die Seitenkettenamidgruppe von Asn86  $\text{H}^{\delta 22}$  eine hohe Verifikationspseudoenergie erhalten.

### 5.3.10 Automatische Zuordnung der HSQC-Spektren von Punktmutanten

Eine weitere Anwendung für SIBASA ist die Zuordnung von HSQC-Spektren von Punktmutanten. SIBASA kann die vorhanden chemischen Verschiebungen und die 3D-Struktur des Wildtyps dafür als Eingabe verwenden. Der Vorgänger ASSIGN (Brunner 2006) war so schon in der Lage 96,7 % der chemischen Verschiebungen von HPr *S. aureus* (H15A) mit der 3D-Struktur und der Liste der chemischen Verschiebungen von HPr *S. aureus* (wt) zu bestimmen.

Während die Struktur des Wildtyps nach der Homologie Modellierung und MD-Simulation direkt als Eingabe für die Rückrechnung von SIBASA dient, werden die bekannten chemischen Verschiebungen  $\delta_k^{\text{wt}}$  des Wildtyps in Form von Normalverteilungen  $N(\delta_k^{\text{wt}}, \sigma_k)$  übergeben, wo sie die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebung bei der Auswahl der Zielslots (Kap. 2.2.11) und der Berechnung der Pseudoenergie der chemischen Verschiebung  $E^\omega$  übernehmen (Gl 62). Wie in Bartels et al. (1997) vorgeschlagen, sollten die Standardabweichung  $\sigma_k$  für H $^\alpha$ - und Seitenkettenprotonen 0,05 ppm, für Amidprotonen 0,15 ppm und für Stickstoffe 0,42 ppm betragen.

Die chemischen Verschiebungen des Wildtyps schränken den Möglichkeitsraum der automatischen Zuordnung ein und können so die automatische Zuordnung erheblich beschleunigen. Da aber die minimale Pseudoenergie der Optimierung  $E^{\text{total}}$ , die die beste vollständige Zuordnung definiert fast nur vom Signalvergleich und damit der Rückrechnung der 3D-Struktur abhängt, kann SIBASA trotz der zusätzlichen Informationen keine experimentell schlecht definierten Zuordnungen finden, und schon gar nicht als sicher bestätigen. Die chemischen Verschiebungen des Wildtyps können aber dabei helfen, die richtige Zuordnung aus mehreren bezüglich der Pseudoenergie der Optimierung gleichwertigen Konfigurationen auszuwählen oder die Verifikationspseudoenergie einer falschen Zuordnung verkleinern.

Sollten genügend chemische Verschiebungen des Wildtyps vorhanden sein, so sollte es selbst für größere Punktmutanten möglich sein das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum durch ein  $^1\text{H}$ - $^{15}\text{N}$  2-D-HSQC-Spektrum zu ersetzen. Die automatische Zuordnung der Protonenresonanzen kann höchstwahrscheinlich mit dem 2-D-NOESY-Spektrum, der 3D-Struktur des Wildtyps und den engen Wahrscheinlichkeitsdichtefunktionen der chemischen Verschiebungen durchgeführt werden. Unterstützt wird die Zuordnung der Protonen durch die experimentellen 2-D-HSQC-Signale, mit deren Hilfe die Resonanzen der Amidprotonen von den Resonanzen der Seitenkettenprotonen im 2-D-NOESY-Spektrum getrennt werden können.

Für die Zuordnung des HSQC-Spektrums muss für jedes Amidproton das passende HSQC-Signal gefunden werden. Dies kann schwierig werden, falls mehr als ein Signal von der chemischen Verschiebung eines Amidprotons in der  $^1\text{H}$ -Dimension des HSQC-Spektrums



geschnitten wird. Falls dies der Fall ist müssen die Uneindeutigkeiten mit den chemischen Verschiebungen der Stickstoffresonanzen des Wildtyps oder mithilfe von vorhergesagten Wahrscheinlichkeitsdichten der chemischen Verschiebungen aufgelöst werden. Falls die Daten dafür nicht ausreichen, ist eine eindeutige Zuordnung der beteiligten HSQC-Signale nicht möglich.

Das größte Problem bei dieser Art der automatischen Zuordnung des HSQC-Spektrums sind sich überlappende HSQC-Signale. Wenn die Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der beteiligten Stickstoffe sich ebenfalls überlappen ist es, wenn überhaupt, nur über die Reihenfolge der Resonanzen der chemischen Verschiebungen der Stickstoffe im Wildtyp möglich, die HSQC-Signale richtig zuzuordnen.

### 5.3.11 Ausblick

Mit SIBASA wurde ein AUREMOL Modul zur zuverlässigen automatischen Zuordnung von Protein-HSQC-Spektren aus deren Röntgenstrukturen entwickelt. Durch optimierte Pseudoenergien und MOVES, sowie den Einsatz von mehreren Instanzen ist es möglich HSQC-Spektren sicher teilzuzuordnen. Außerdem wurde der Grundstein gelegt, um in Zukunft zusätzliche Spektren, wie z.B. Triple-Resonanz-Spektren, 3D  $^{13}\text{C}$ -NOESY-HSQC-Spektrum oder  $^{13}\text{C}$ -HSQC-Spektren mit natürlicher Häufigkeit, auswerten zu können.

Mit Hilfe von Triple-Resonanz-Spektren wie HNCA-, HN(CO)CA-, CBCA(CO)NH-, HNCACB-Spektren, die bereits in AUREMOL RELAX implementiert sind, könnten vertauschte Zuordnungen von HSQC-Signalen vermieden, die vollständige Zuordnung des Rückgrats automatisiert und mit Hilfe der chemischen Verschiebungen der  $\text{C}^\alpha$ - und  $\text{C}^\beta$ -Kerne auch die automatische Zuordnung von Seitenkettenprotonen verbessert werden.

Intraresiduale vertauschte Seitenkettenatome könnten durch die Verwendung eines  $^{13}\text{C}$ -HSQC-Spektrum mit natürlicher Häufigkeit und den zugehörigen geschätzten Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen der Kohlenstoffe korrigiert werden. Es ist außerdem davon auszugehen, dass durch die Anpassung des SIBASA-Algorithmus an ein zusätzliches 3D  $^{13}\text{C}$ -NOESY-HSQC-Spektrum die Zuordnung der Seitenkettenprotonen das gleiche hohe Niveau wie die Zuordnung der Amidprotonen erreicht.

Je mehr Informationen SIBASA zur Verfügung hat, desto verschiedener können Eingabestruktur und Lösungsstruktur sein. Die Auswertung von TOCSY- bzw. 3D HCCH-TOCSY Spektren kann in Zukunft helfen interresiduale von intraresidualen NOESY-Kreuzsignalen zu trennen. Die Abhängigkeit des automatischen Zuordnung von der 3D-Struktur würde dadurch verringert werden. Um die Abhängigkeit von der Eingabestruktur weiter zu verkleinern kann in Zukunft weiteres Datenbankwissen als Pseudoenergie oder für die Bewertung einzelner Terme der Pseudoenergie verwendet werden. Einerseits profitiert

SIBASA sehr stark von den Vorhersagen der chemischen Verschiebungen, andererseits können Diederwinkeldatenbanken und Abstandsdatenbanken, wie sie z.B. in Nasser (2006) beschrieben sind, helfen die wichtigen zurückgerechneten Signale zu bestimmen und zu gewichten. In diese Datenbanken könnten zusätzlich tiefere von der Primärsequenz abhängige Korrelation eingefügt werden. Es sollte sogar möglich sein, eine vorläufige Struktur, die mit vorhanden experimentellen Daten und Datenbankwissen erzeugt wurde, iterativ mit den automatisch von SIBASA bestimmten chemischen Verschiebungen zu verbessern und so die Lösungsstruktur und die vollständige Zuordnung zu erhalten.

Mit genügend Rechenzeit wäre es möglich von der Lösungsstruktur abweichende 3D-Strukturen mit MD-Simulationen in einen Zustand zu bringen, der für SIBASA nahe genug an der Lösungsstruktur liegt. Durch Weiterentwicklung der MD-Simulationen, könnte es in Zukunft möglich sein chemischen Austausch für die Rückrechnung und die Auswirkungen des Lösungsmittels auf die chemische Verschiebung von Amidprotonen durch Analyse des lokalen elektrischen und magnetischen Feldes vorherzusagen.

Ein großes Potential zur Verbesserung liegt in der Anpassung der Verifikationspseudoenergie. Ziel dieser Anpassung sollte es sein, die falschen Zuordnungen mit hoher Verifikationspseudoenergie abzuwerten, um den Grenzwert für das 100 %-Sicherheitsniveau von HSQC-Signalen absenken zu können. Folgende Verbesserungen könnten hierbei helfen:

- Für die Verifikation der Zuordnung von HSQC-Signalen könnten zusätzlich zu den lokalen Pseudoenergie der Grundflächen der simulierten Signale auch die experimentellen Signale entlang der indirekten  $^1\text{H}$ -Dimension des 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrums mit berücksichtigt werden.
- Es könnte Datenbankwissen von Abstandsdatenbanken und chemischen Verschiebungsdatenbanken in die Verifikation einfließen.
- Eine tiefere Analyse der einzelnen Instanzen könnte Vertauschungen und falsche Permutationen von Zuordnungen aufdecken.
- Durch die Berücksichtigung des zweitbesten Clusters in der Verifikationspseudoenergie könnten Zuordnungen, die zwischen zwei Zuständen schwanken, abgewertet werden.

Für die Zuordnung von Seitenkettenprotonen kann die Einführung einer weiteren Pseudoenergie, die das Zusammenlegen der Resonanzen von magnetisch gleichwertiger Kerne wie z.B.  $\text{CH}_2$ -Gruppen belohnt, dabei helfen überlappende Zuordnung besser zu finden.

## 6 Zusammenfassung

Im ersten Teil dieser Arbeit wurde die bisher unbekannte stereospezifische Zuordnung der Seitenkettenamidgruppen in den Random-Coil-Modellpeptiden Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Gln-Ala-NH<sub>2</sub> ermittelt. Um eine korrekte stereospezifische Zuordnung von NMR-Resonanzsignalen zu gewährleisten wird meistens auf Listen bereits bekannter chemischer Verschiebungen zurückgegriffen. Die für diesen Zweck am häufigsten verwendeten Quellen sind die chemischen Verschiebungen von Zuordnungsdatenbanken von bereits ausgewerteten Biomolekülen. Die bekannteste dieser Datenbanken ist die **Biological Magnetic Resonance Data Bank (BMRB)**. Unterzieht man aber die in der BMRB gespeicherten chemischen Verschiebungen einer genaueren Untersuchung, so kann man Inkonsistenzen bei den stereospezifischen Zuordnungen finden. Es wurde außerdem festgestellt, dass die beiden Programme SHIFTS und SHIFTX, die chemischen Verschiebungen aus den 3D-Strukturen von Peptiden und Proteinen vorhersagen können, wegen Fehlern in ihren jeweiligen Datenbanken diese chemischen Verschiebungen ebenfalls stereospezifisch falsch zuordnen können.

Für die stereospezifische Zuordnung wurden **Nuclear Overhauser Enhancement Spectroscopy (NOESY)** Spektren der Random-Coil-Peptide von Gly-Gly-Asn-Ala-NH<sub>2</sub> und Gly-Gly-Gln-Ala-NH<sub>2</sub> bei 800 MHz aufgenommen und mit AUREMOL RELAX, das den vollständigen Relaxationsmatrixformalismus verwendet, entsprechende Spektren aus **Molekulardynamik (MD)** Rechnungen der beiden Tetrapeptide simuliert. Ein Vergleich der experimentellen und simulierten Signalvolumina erbrachte eine eindeutige stereospezifische Zuordnung der Random-Coil-Verschiebungen der Seitenkettenamid- und H<sup>β</sup>-Protonen der beiden Aminosäuren.

Die vorgestellte Methode hat das Potential in zukünftigen Arbeiten auf einen großen Teil aller Aminosäuren übertragen zu werden, um eine vollständige stereospezifische Random-Coil-Verschiebungsdatenbank zu erzeugen.

Im zweiten Teil der Arbeit wurde SIBASA, das neue AUREMOL Modul zur automatischen Zuordnung von HSQC-Spektren vorgestellt. Mit ihm ist es nicht nur möglich die chemischen Verschiebungen der Resonanzen von Amidprotonen und Stickstoffen, sondern auch die chemischen Verschiebungen der Resonanzen von H<sup>α</sup>- und Seitenkettenprotonen automatisch zu bestimmen. SIBASA basiert auf dem Top-Down-Ansatz und bestimmt die vollständige Zuordnung eines Proteins, indem es die optimale Übereinstimmung zwischen experimentellen und mit variablen chemischen Verschiebungen zurückgerechnete NOESY-Spektren findet. Auf diese Weise können auch Spektren mit starkem Signalüberlapp, wie 2-D-NOESY-Spektren von großen Proteinen, die ein sehr gutes Signal-zu-Rausch-Verhältnis haben, als primäre experimentelle Informationsquelle der automatische Zuordnung verwendet werden. SIBASA benötigt für die vollständige Zuordnung der Resonanzen aller Protonen <sup>1</sup>H und Stickstoffe <sup>15</sup>N eine ungefähre 3D-Struktur des Proteins, ein [<sup>1</sup>H-<sup>1</sup>H] 2-D-NOESY-Spektrum

und nur ein einziges zeitaufwendiges 3-D-NMR-Spektrum, nämlich das 3D  $^{15}\text{N}$ -NOESY-HSQC-Spektrum. Es muss auch keine teure  $^{13}\text{C}$ -markierte Probe des Proteins vorhanden sein. Die Rückrechnungen der NOESY-Spektren werden mit AUREMOL RELAX, das den vollständigen Relaxationsmatrixformalismus verwendet, erstellt. RELAX wertet zudem die Ergebnisse von Molekular-dynamik (MD) Rechnungen aus, um Informationen über den Zustands des Proteins und dessen Beweglichkeit in Lösung berücksichtigen zu können. Mithilfe der Programme SHIFTS und SHIFTX2 kann SIBASA die individuellen Wahrscheinlichkeitsdichteverteilungen der chemischen Verschiebungen aller Protonen und Stickstoffe des Proteins aus der MD-Trajektorie des betrachteten Proteins vorhersagen, was zur Verbesserung der Zuverlässigkeit und Geschwindigkeit der automatischen Zuordnung führt.

Die optimale Übereinstimmung der experimentellen und der aus den simulierten Signalen zusammengesetzten NOESY-Spektren wird durch den Threshold-Accepting-Algorithmus, der in mehreren unabhängigen Instanzen mit unterschiedlichen Startzuordnung und Zufallszahlen ausgeführt wird, bestimmt. Mehrere Instanzen helfen SIBASA die wahrscheinlichste vollständige Zuordnung zu finden und sind Voraussetzung für die Verifikation der automatisch gefundenen Resonanzzuordnungen. Nach der automatischen Zuordnung weist SIBASA jeder gefundene Zuordnung eine Wahrscheinlichkeit der korrekten Zuordnung zu.

Die automatische Zuordnung wurde mit den NOESY-Spektren und den Röntgenstrukturen der Proteine HPr *S. aureus* (H15A) (88 Aminosäuren), Plasmodium falciparum Thioredoxin (PfTrx) (104 Aminosäuren) und Ras(T35S)-GppNHp (166 Aminosäuren) getestet. SIBASA konnte 91,3 % der Resonanzen von HPr *S. aureus* (H15A), 81,9 % der Resonanzen von PfTrx und 77,6 % der Resonanzen von Ras(T35S)-GppNHp richtig zuordnen. Eine Verifikation auf dem signifikanten Niveau ermöglicht es, einen großen Teil der falschen Zuordnungen von den richtigen zu trennen. Insgesamt erhielten 77,8 % der automatisch gefundenen Resonanzzuordnungen von HPr *S. aureus* (H15A), 77,5 % der gefundenen Resonanzzuordnungen von PfTrx und 66,8 % der Resonanzzuordnungen von Ras(T35S)-GppNHp von SIBASA eine Wahrscheinlichkeit von mindestens 95 %. Von diesen Resonanzen sind beim HPr *S. aureus* (H15A) nur 3,5 % beim PfTrx nur 9,7 % und beim Ras(T35S)-GppNHp nur 10,2 % falsch zugeordnet worden.

Es wurde anhand der drei Proteine gezeigt, dass SIBASA in der Lage ist HSQC-Spektren sicher teilzuordnen. Das HSQC-Spektrum von HPr *S. aureus* (H15A) konnte von SIBASA vollständig richtig zugeordnet werden. Beim PfTrx waren 90 % und beim Ras(T35S)-GppNHp 88 % der automatisch gefundenen HSQC-Zuordnungen richtig. Vertraut man nur Zuordnungen von HSQC-Signalen, die von SIBASA bestätigt wurden, so konnten 82 % der Signale von HPr *S. aureus* (H15A), 72 % der Signale von PfTrx und 68 % der HSQC-Signale des Ras(T35S)-GppNHp richtig zugeordnet werden. In keinem Fall wurde von SIBASA eine falsche Zuordnung eines HSQC-Signals bestätigt.

Das vorgestellte Modul ermöglicht es für die Wirkstoffentwicklung wichtige [ $^1\text{H}$ ,  $^{15}\text{N}$ ]-HSQC-Spektren automatisch zuzuordnen, ohne auf die umständliche Markierung der Proteine mit dem Isotop  $^{13}\text{C}$  zurückgreifen zu müssen.

## 7 Literaturverzeichnis

- Andrec, Michael; Levy, Ronald M. (2002): Protein sequential resonance assignments by combinatorial enumeration using  $^{13}\text{C}^\alpha$  chemical shifts and their (i, i - 1) sequential connectivities. In: *Journal of Biomolecular NMR* 23 (4), S. 263–270. DOI: 10.1023/A:1020236105735.
- Antz, C.; Neidig, K. P.; Kalbitzer, H. R. (1995): A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis. In: *Journal of Biomolecular NMR* 5 (3), S. 287–296. DOI: 10.1007/BF00211755.
- Araki, Mitsugu; Shima, Fumi; Yoshikawa, Yoko; Muraoka, Shin; Ijiri, Yuichi; Nagahara, Yuka et al. (2011): Solution structure of the state 1 conformer of GTP-bound H-Ras protein and distinct dynamic properties between the state 1 and state 2 conformers. In: *The Journal of biological chemistry* 286 (45), S. 39644–39653. DOI: 10.1074/jbc.M111.227074.
- Atreya, H. S.; Chary, K. V.R.; Govil, Girjesh (2002): Automated NMR assignments of proteins for high throughput structure determination: TATAPRO II. In: *CURRENT SCIENCE* 83 (11), S. 1372.
- Atreya, H. S.; Sahu, S. C.; Chary, K.V.R.; Govil, Girjesh (2000): A tracked approach for automated NMR assignments in proteins (TATAPRO). In: *Journal of Biomolecular NMR* 17 (2), S. 125–136. DOI: 10.1023/A:1008315111278.
- Aue, W. P.; Bartholdi, E.; Ernst, R. R. (1976): Two-dimensional spectroscopy. Application to nuclear magnetic resonance. In: *J. Chem. Phys.* 64 (5), S. 2229–2246. DOI: 10.1063/1.432450.
- Axel T. Brünger (1992): X-PLOR MANUAL. A System for X-ray Crystallography and NMR. Version 3.1. In: *New Haven, Yale University Press*.
- Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. (1993): Primary structure effects on peptide group hydrogen exchange. In: *Proteins* 17 (1), S. 75–86. DOI: 10.1002/prot.340170110.
- Bartels, C.; Billeter, M.; Güntert, P.; Wüthrich, K. (1996): Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. In: *Journal of Biomolecular NMR* 7 (3), S. 207–213. DOI: 10.1007/BF00202037.
- Bartels, Christian; Güntert, Peter; Billeter, Martin; Wüthrich, Kurt (1997): GARANT-a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. In: *J. Comput. Chem.* 18 (1), S. 139–149. DOI: 10.1002/(SICI)1096-987X(19970115)18:1<139::AID-JCC13>3.0.CO;2-H.
- Baskaran, Kumaran (2010): Chemical shift optimization and ensemble averaging in protein NMR spectroscopy.
- Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. (1995): GROMACS: A message-passing parallel molecular dynamics implementation. In: *Computer Physics Communications* 91 (1-3), S. 43–56. DOI: 10.1016/0010-4655(95)00042-E.
- Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. (1987): The missing term in effective pair potentials. In: *The Journal of Physical Chemistry* 91 (24), S. 6269–6271. DOI: 10.1021/j100308a038.

- Berman, H. M. (2000): The Protein Data Bank. In: *Nucleic Acids Research* 28 (1), S. 235–242. DOI: 10.1093/nar/28.1.235.
- Berwin A. TURLACH: Bandwidth selection in kernel density estimation: a review. Humboldt Universitaet Berlin (9307). Online verfügbar unter <http://ideas.repec.org/p/wop/humbse/9307.html>.
- Bodenhausen, Geoffrey; Ruben, David J. (1980): Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. In: *Chemical Physics Letters* 69 (1), S. 185–189. DOI: 10.1016/0009-2614(80)80041-8.
- Bora, Ram Prasad; Prabhakar, Rajeev (2009): Translational, rotational and internal dynamics of amyloid beta-peptides (Abeta40 and Abeta42) from molecular dynamics simulations. In: *J. Chem. Phys.* 131 (15), S. 155103. DOI: 10.1063/1.3249609.
- Braunschweiler, L.; Ernst, R.R (1983): Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. In: *Journal of Magnetic Resonance (1969)* 53 (3), S. 521–528. DOI: 10.1016/0022-2364(83)90226-3.
- Brunger, Axel T. (2007): Version 1.2 of the Crystallography and NMR system. In: *Nature protocols* 2 (11), S. 2728–2733. DOI: 10.1038/nprot.2007.406.
- Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W. et al. (1998): Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. In: *Acta Crystallogr D Biol Crystallogr* 54 (5), S. 905–921. DOI: 10.1107/S0907444498003254.
- Brunner, Konrad (2006): Modellierung, Strukturverbesserung und sequentielle Zuordnung als vollautomatische Module für die automatisierte Proteinstrukturbestimmung im Softwareprojekt AUREMOL.
- Bundi, Arno; Wüthrich, Kurt (1979): <sup>1</sup>H-nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. In: *Biopolymers* 18 (2), S. 285–297. DOI: 10.1002/bip.1979.360180206.
- Bussi, Giovanni; Donadio, Davide; Parrinello, Michele (2007): Canonical sampling through velocity rescaling. In: *J. Chem. Phys.* 126 (1), S. 014101. DOI: 10.1063/1.2408420.
- Cano, Carolina; Brunner, Konrad; Baskaran, Kumaran; Elsner, Ralph; Munte, Claudia E.; Kalbitzer, Hans Robert (2009): Protein structure calculation with data imputation: the use of substitute restraints. In: *Journal of Biomolecular NMR* 45 (4), S. 397–411. DOI: 10.1007/s10858-009-9379-y.
- Case, David A. (1995): Calibration of ring-current effects in proteins and nucleic acids. In: *Journal of Biomolecular NMR* 6 (4). DOI: 10.1007/BF00197633.
- Cavanagh, John (2007): Protein NMR spectroscopy. Principles and practice. 2nd ed. Amsterdam, Boston: Academic Press.
- Chen, Zhi-Zhong; Lin, Guohui; Rizzi, Romeo; Wen, Jianjun; Xu, Dong; Xu, Ying; Jiang, Tao (2005): More reliable protein NMR peak assignment via improved 2-interval scheduling. In: *Journal of computational biology : a journal of computational molecular cell biology* 12 (2), S. 129–146. DOI: 10.1089/cmb.2005.12.129.
- Choy, W. Y.; Sanctuary, B. C.; Zhu, G. (1997): Using Neural Network Predicted Secondary Structure Information in Automatic Protein NMR Assignment. In: *J. Chem. Inf. Model.* 37 (6),

S. 1086–1094. DOI: 10.1021/ci970012c.

Clore, G. Marius; Szabo, Attila; Bax, Ad; Kay, Lewis E.; Driscoll, Paul C.; Gronenborn, Angela M. (1990): Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. In: *Journal of the American Chemical Society* 112 (12), S. 4989–4991. DOI: 10.1021/ja00168a070.

Coggins, Brian E. (2003): PACES: protein sequential assignment by computer-assisted exhaustive search. In: *Journal of Biomolecular NMR* 26 (2), S. 93–111. DOI: 10.1023/A:1023589029301.

Cornell, Wendy D.; Cieplak, Piotr; Bayly, Christopher I.; Gould, Ian R.; Merz, Kenneth M.; Ferguson, David M. et al. (1995): A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. In: *Journal of the American Chemical Society* 117 (19), S. 5179–5197. DOI: 10.1021/ja00124a002.

Darden, Tom; York, Darrin; Pedersen, Lee (1993): Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. In: *J. Chem. Phys.* 98 (12), S. 10089. DOI: 10.1063/1.464397.

Day, Ryan; García, Angel E. (2008): Water penetration in the low and high pressure native states of ubiquitin. In: *Proteins: Structure, Function, and Bioinformatics* 70 (4), S. 1175–1184. DOI: 10.1002/prot.21562.

de Vries, Sjoerd J; van Dijk, Aalt D J; Krzeminski, Mickaël; van Dijk, Mark; Thureau, Aurelien; Hsu, Victor et al. (2007): HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. In: *Proteins* 69 (4), S. 726–733. DOI: 10.1002/prot.21723.

Dominguez, Cyril; Boelens, Rolf; Bonvin, Alexandre M J J (2003): HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. In: *Journal of the American Chemical Society* 125 (7), S. 1731–1737. DOI: 10.1021/ja026939x.

Dueck, Gunter; Scheuer, Tobias (1990): Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing. In: *Journal of Computational Physics* 90 (1), S. 161–175. DOI: 10.1016/0021-9991(90)90201-B.

Eghbalnia, Hamid R.; Bahrami, Arash; Wang, Liya; Assadi, Amir; Markley, John L. (2005): Probabilistic Identification of Spin Systems and their Assignments including Coil-Helix Inference as Output (PISTACHIO). In: *Journal of Biomolecular NMR* 32 (3), S. 219–233. DOI: 10.1007/s10858-005-7944-6.

Eriksson, M. A.; Härd, T.; Nilsson, L. (1995): On the pH dependence of amide proton exchange rates in proteins. In: *Biophysical Journal* 69 (2), S. 329–339. DOI: 10.1016/S0006-3495(95)79905-2.

Facelli, Julio C.; Dios, Angel C. de (Hg.) (1999): Modeling NMR Chemical Shifts. Washington, DC: American Chemical Society (ACS Symposium Series).

Ferner, Jan; Villa, Alessandra; Duchardt, Elke; Widjajakusuma, Elisabeth; Wöhnert, Jens; Stock, Gerhard; Schwalbe, Harald (2008): NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. In: *Nucleic Acids Research* 36 (6), S. 1928–1940. DOI: 10.1093/nar/gkm1183.

Friday, Bret B.; Adjei, Alex A. (2005): K-ras as a target for cancer therapy. In: *Biochimica et*

- biophysica acta* 1756 (2), S. 127–144. DOI: 10.1016/j.bbcan.2005.08.001.
- Friedrichs, Mark S.; Mueller, Luciano; Wittekind, Michael (1994): An automated procedure for the assignment of protein  $^1\text{HN}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ ?,  $^1\text{H}$ ?,  $^{13}\text{C}$ ? and  $^1\text{H}$ ? resonances. In: *Journal of Biomolecular NMR* 4 (5), S. 703–726. DOI: 10.1007/BF00404279.
- Fritz-Wolf, Karin; Jortzik, Esther; Stumpf, Michaela; Preuss, Janina; Iozef, Rimma; Rahlfs, Stefan; Becker, Katja (2013): Crystal structure of the Plasmodium falciparum thioredoxin reductase-thioredoxin complex. In: *Journal of Molecular Biology* 425 (18), S. 3446–3460. DOI: 10.1016/j.jmb.2013.06.037.
- Geyer, M.; Neidig, K. P.; Kalbitzer, H. R. (1995): Automated Peak Integration in Multidimensional NMR Spectra by an Optimized Iterative Segmentation Procedure. In: *Journal of Magnetic Resonance, Series B* 109 (1), S. 31–38. DOI: 10.1006/jmrb.1995.1143.
- Geyer, M.; Schweins, T.; Herrmann, C.; Prisner, T.; Wittinghofer, A.; Kalbitzer, H. R. (1996): Conformational transitions in p21ras and in its complexes with the effector protein Raf-RBD and the GTPase activating protein GAP. In: *Biochemistry* 35 (32), S. 10308–10320. DOI: 10.1021/bi952858k.
- Görler, A.; Kalbitzer, H. R. (1997): Relax, a flexible program for the back calculation of NOESY spectra based on complete-relaxation-matrix formalism. In: *J. Magn. Reson.* 124 (1), S. 177–188. DOI: 10.1006/jmre.1996.1033.
- Grigera, J. Raul (2001): An effective pair potential for heavy water. In: *J. Chem. Phys.* 114 (18), S. 8064. DOI: 10.1063/1.1359183.
- Gronwald, Wolfram; Moussa, Sherif; Elsner, Ralph; Jung, Astrid; Ganslmeier, Bernhard; Trenner, Jochen et al. (2002): Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). In: *Journal of Biomolecular NMR* 23 (4), S. 271–287. DOI: 10.1023/A:1020279503261.
- Gronwald, Wolfram; Willard, Leigh; Jellard, Timothy; Boyko, Robert F.; Rajarathnam, Krishna; Wishart, David S. et al. (1998): CAMRA: chemical shift based computer aided protein NMR assignments. In: *Journal of Biomolecular NMR* 12 (3), S. 395–405. DOI: 10.1023/A:1008321629308.
- Güntert, Peter (2003): Automated NMR protein structure calculation. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 43 (3-4), S. 105–125. DOI: 10.1016/S0079-6565(03)00021-9.
- Güntert, Peter (2009): Automated structure determination from NMR spectra. In: *European biophysics journal : EBJ* 38 (2), S. 129–143. DOI: 10.1007/s00249-008-0367-z.
- Güntert, Peter; Salzmann, Michael; Braun, Daniel; Wüthrich, Kurt (2000): Sequence-specific NMR assignment of proteins by global fragment mapping with the program Mapper. In: *Journal of Biomolecular NMR* 18 (2), S. 129–137. DOI: 10.1023/A:1008318805889.
- Han, Beomsoo; Liu, Yifeng; Ginzinger, Simon W.; Wishart, David S. (2011): SHIFTX2: significantly improved protein chemical shift prediction. In: *Journal of Biomolecular NMR* 50 (1), S. 43–57. DOI: 10.1007/s10858-011-9478-4.
- Hare, Brian J.; Prestegard, James H. (1994): Application of neural networks to automated assignment of NMR spectra of proteins. In: *Journal of Biomolecular NMR* 4 (1). DOI: 10.1007/BF00178334.



- Harsch, T.; Dasch, C.; Donaubaue, H.; Baskaran, K.; Kremer, W.; Kalbitzer, H. R. (2013): Stereospecific Assignment of the Asparagine and Glutamine Side Chain Amide Protons in Random-Coil Peptides by Combination of Molecular Dynamic Simulations with Relaxation Matrix Calculations. In: *Appl Magn Reson* 44 (1-2), S. 319–331. DOI: 10.1007/s00723-012-0416-8.
- Hess, Berk; Kutzner, Carsten; van der Spoel, David; Lindahl, Erik (2008): GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. In: *J. Chem. Theory Comput.* 4 (3), S. 435–447. DOI: 10.1021/ct700301q.
- Hitchens, T. Kevin; Lukin, Jonathan A.; Zhan, Yiping; McCallum, Scott A.; Rule, Gordon S. (2003): MONTE: an automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins. In: *Journal of Biomolecular NMR* 25 (1), S. 1–9. DOI: 10.1023/A:1021975923026.
- Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. (1979): Investigation of exchange processes by two-dimensional NMR spectroscopy. In: *J. Chem. Phys.* 71 (11), S. 4546. DOI: 10.1063/1.438208.
- Jia, Zongchao; Vandonselaar, Margaret; Hengstenberg, Wolfgang; Wilson Quail, J.; Delbaere, Louis T.J. (1994): The 1.6 Å structure of histidine-containing phosphotransfer protein HPr from *Streptococcus faecalis*. In: *Journal of Molecular Biology* 236 (5), S. 1341–1355. DOI: 10.1016/0022-2836(94)90062-0.
- Jorgensen, William L.; Chandrasekhar, Jayaraman; Madura, Jeffrey D.; Impey, Roger W.; Klein, Michael L. (1983): Comparison of simple potential functions for simulating liquid water. In: *J. Chem. Phys.* 79 (2), S. 926. DOI: 10.1063/1.445869.
- Jung, Young-Sang; Zweckstetter, Markus (2004a): Backbone assignment of proteins with known structure using residual dipolar couplings. In: *Journal of Biomolecular NMR* 30 (1), S. 25–35. DOI: 10.1023/B:JNMR.0000042955.14647.77.
- Jung, Young-Sang; Zweckstetter, Markus (2004b): Mars -- robust automatic backbone assignment of proteins. In: *Journal of Biomolecular NMR* 30 (1), S. 11–23. DOI: 10.1023/B:JNMR.0000042954.99056.ad.
- JURANIĆ, NENAD; MACURA, SLOBODAN; ZOLNAI, ZSOLT (1998): Two-dimensional exchange NMR spectroscopy in fractionally deuterated molecules: separation of exchange and cross-relaxation processes by proton dilution. In: *Molecular Physics* 95 (5), S. 833–840. DOI: 10.1080/002689798166459.
- Kalbitzer, Hans Robert; Hengstenberg, Wolfgang (1993): The solution structure of the histidine-containing protein (HPr) from *Staphylococcus aureus* as determined by two-dimensional <sup>1</sup>H-NMR spectroscopy. In: *Eur J Biochem* 216 (1), S. 205–214. DOI: 10.1111/j.1432-1033.1993.tb18134.x.
- Kamisetty, Hetunandan; Bailey-Kellogg, Chris; Pandurangan, Gopal (2006): An efficient randomized algorithm for contact-based NMR backbone resonance assignment. In: *Bioinformatics (Oxford, England)* 22 (2), S. 172–180. DOI: 10.1093/bioinformatics/bti786.
- Karplus, Martin. (1963): Vicinal Proton Coupling in Nuclear Magnetic Resonance. In: *Journal of the American Chemical Society* 85 (18), S. 2870–2871. DOI: 10.1021/ja00901a059.

- Kerrigan, John E.: GROMACS Tutorial for Solvation Study of Spider Toxin Peptide. Gromacs 4.6 Tutorial. Online verfügbar unter [http://cinjweb.umdj.edu/~kerrigje/pdf\\_files/fwspidr\\_tutor.pdf](http://cinjweb.umdj.edu/~kerrigje/pdf_files/fwspidr_tutor.pdf), zuletzt geprüft am 30.06.2015.
- Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. (1983): Optimization by simulated annealing. In: *Science* 220 (4598), S. 671–680. DOI: 10.1126/science.220.4598.671.
- Kjær, Mogens; Andersen, Kim Vilbour; Poulsen, Flemming M. (1994): [8] Automated and semiautomated analysis of homo- and heteronuclear multidimensional nuclear magnetic resonance spectra of proteins: The program pronto. In: *Nuclear Magnetic Resonance, Part C*, Bd. 239: Elsevier (Methods in Enzymology), S. 288–307.
- Kjaergaard, Magnus; Brander, Søren; Poulsen, Flemming M. (2011): Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. In: *Journal of Biomolecular NMR* 49 (2), S. 139–149. DOI: 10.1007/s10858-011-9472-x.
- Kumar, Anil; Ernst, R. R.; Wüthrich, K. (1980): A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. In: *Biochemical and Biophysical Research Communications* 95 (1), S. 1–6. DOI: 10.1016/0006-291X(80)90695-6.
- Leutner, Michael; Gschwind, Ruth M.; Liermann, Jens; Schwarz, Christian; Gemmecker, Gerd; Kessler, Horst (1998): Automated backbone assignment of labeled proteins using the threshold accepting algorithm. In: *Journal of Biomolecular NMR* 11 (1), S. 31–43. DOI: 10.1023/A:1008298226961.
- Li, K.-B.; Sanctuary, B. C. (1997a): Automated Resonance Assignment of Proteins Using Heteronuclear 3D NMR. 1. Backbone Spin Systems Extraction and Creation of Polypeptides. In: *J. Chem. Inf. Model.* 37 (2), S. 359–366. DOI: 10.1021/ci960045c.
- Li, K.-B.; Sanctuary, B. C. (1997b): Automated Resonance Assignment of Proteins Using Heteronuclear 3D NMR. 2. Side Chain and Sequence-Specific Assignment. In: *J. Chem. Inf. Model.* 37 (3), S. 467–477. DOI: 10.1021/ci960372k.
- Lin, Hsin-Nan; Wu, Kun-Pin; Chang, Jia-Ming; Sung, Ting-Yi; Hsu, Wen-Lian (2005): GANA--a genetic algorithm for NMR backbone resonance assignment. In: *Nucleic Acids Research* 33 (14), S. 4593–4601. DOI: 10.1093/nar/gki768.
- Lindorff-Larsen, Kresten; Piana, Stefano; Palmo, Kim; Maragakis, Paul; Klepeis, John L.; Dror, Ron O.; Shaw, David E. (2010): Improved side-chain torsion potentials for the Amber ff99SB protein force field. In: *Proteins* 78 (8), S. 1950–1958. DOI: 10.1002/prot.22711.
- Lipari, Giovanni; Szabo, Attila (1982): Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. In: *Journal of the American Chemical Society* 104 (17), S. 4546–4559. DOI: 10.1021/ja00381a009.
- Longworth, L. G. (1960): THE MUTUAL DIFFUSION OF LIGHT AND HEAVY WATER. In: *The Journal of Physical Chemistry* 64 (12), S. 1914–1917. DOI: 10.1021/j100841a027.
- López-Méndez, Blanca; Güntert, Peter (2006): Automated protein structure determination from NMR spectra. In: *Journal of the American Chemical Society* 128 (40), S. 13112–13122. DOI: 10.1021/ja061136l.

- Lukin, Jonathan A.; Gove, Andrew P.; Talukdar, Sarosh N.; Ho, Chien (1997): Automated probabilistic method for assigning backbone resonances of (C-13, N-15)-labeled proteins. In: *Journal of Biomolecular NMR* 9 (2), S. 151–166. DOI: 10.1023/A:1018602220061.
- Malloni, Wilhelm M.; Sanctis, Silvia de; Tomé, Ana M.; Lang, Elmar W.; Munte, Claudia E.; Neidig, Klaus Peter; Kalbitzer, Hans Robert (2010): Automated solvent artifact removal and base plane correction of multidimensional NMR protein spectra by AUREMOL-SSA. In: *Journal of Biomolecular NMR* 47 (2), S. 101–111. DOI: 10.1007/s10858-010-9414-z.
- Mark, Pekka; Nilsson, Lennart (2001): Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. In: *J. Phys. Chem. A* 105 (43), S. 9954–9960. DOI: 10.1021/jp003020w.
- Masse, James E.; Keller, Rochus (2005): AutoLink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. In: *J. Magn. Reson.* 174 (1), S. 133–151. DOI: 10.1016/j.jmr.2005.01.017.
- Masse, James E.; Keller, Rochus; Pervushin, Konstantin (2006): SideLink: automated side-chain assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. In: *J. Magn. Reson.* 181 (1), S. 45–67. DOI: 10.1016/j.jmr.2006.03.012.
- McIntosh, Lawrence P.; Brun, Emmanuel; Kay, Lewis E. (1997): Stereospecific assignment of the NH<sub>2</sub> resonances from the primary amides of asparagine and glutamine side chains in isotopically labeled proteins. In: *Journal of Biomolecular NMR* 9 (3), S. 306–312. DOI: 10.1023/A:1018635110491.
- Moseley, Hunter N. B.; Montelione, Gaetano T. (1999): Automated analysis of NMR assignments and structures for proteins. In: *Current Opinion in Structural Biology* 9 (5), S. 635–642. DOI: 10.1016/S0959-440X(99)00019-6.
- Moseley, Hunter N.B.; Monleon, Daniel; Montelione, Gaetano T. (2001): Automatic Determination of Protein Backbone Resonance Assignments from Triple Resonance Nuclear Magnetic Resonance Data. In: *Nuclear Magnetic Resonance of Biological Macromolecules - Part B*, Bd. 339: Elsevier (Methods in Enzymology), S. 91–108.
- Moseley, Hunter N B; Sahota, Gurmukh; Montelione, Gaetano T. (2004): Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. In: *Journal of Biomolecular NMR* 28 (4), S. 341–355. DOI: 10.1023/B:JNMR.0000015420.44364.06.
- Müller-Gronbach, Thomas; Novak, Erich; Ritter, Klaus (2012): Monte Carlo-Algorithmen. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Munte, Claudia Elisabeth; Becker, Katja; Schirmer, Rolf Heiner; Kalbitzer, Hans Robert (2009): NMR assignments of oxidised thioredoxin from *Plasmodium falciparum*. In: *Biomolecular NMR assignments* 3 (2), S. 159–161. DOI: 10.1007/s12104-009-9163-7.
- Narayanan, Rhagavendran L.; Dürr, Ulrich H N; Bibow, Stefan; Biernat, Jacek; Mandelkow, Eckhard; Zweckstetter, Markus (2010): Automatic assignment of the intrinsically disordered protein Tau with 441-residues. In: *Journal of the American Chemical Society* 132 (34), S. 11906–11907. DOI: 10.1021/ja105657f.
- Nasser, Adel (2006): Optimierung der Zuordnung mehrdeutiger NOESY-NMR-Signale unter Anwendung einer Datenbank nichtredundanter Proteinstrukturen.

- Neal, Stephen (2003): Rapid and accurate calculation of protein  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts. In: *Journal of Biomolecular NMR* 26 (3), S. 215–240. DOI: 10.1023/A:1023812930288.
- Neidig, K. P.; Geyer, M.; Görler, A.; Antz, C.; Saffrich, R.; Beneicke, W.; Kalbitzer, H. R. (1995): AURELIA, a program for computer-aided analysis of multidimensional NMR spectra. In: *Journal of Biomolecular NMR* 6 (3), S. 255–270. DOI: 10.1007/BF00197807.
- Olson, JohnB.; Markley, JohnL. (1994): Evaluation of an algorithm for the automated sequential assignment of protein backbone resonances: A demonstration of the connectivity tracing assignment tools (CONTRAST) software package. In: *Journal of Biomolecular NMR* 4 (3). DOI: 10.1007/BF00179348.
- Osapay, Klara; Case, David A. (1991): A new analysis of proton chemical shifts in proteins. In: *Journal of the American Chemical Society* 113 (25), S. 9436–9444. DOI: 10.1021/ja00025a002.
- Park, Hwangseo; Lee, Jinuk; Lee, Sangyoub (2006): Critical assessment of the automated AutoDock as a new docking tool for virtual screening. In: *Proteins* 65 (3), S. 549–554. DOI: 10.1002/prot.21183.
- Pettersen, Eric F.; Goddard, Thomas D.; Huang, Conrad C.; Couch, Gregory S.; Greenblatt, Daniel M.; Meng, Elaine C.; Ferrin, Thomas E. (2004): UCSF Chimera--a visualization system for exploratory research and analysis. In: *J. Comput. Chem.* 25 (13), S. 1605–1612. DOI: 10.1002/jcc.20084.
- Postma, P. W.; Lengeler, J. W.; Jacobson, G. R. (1993): Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. In: *Microbiological Reviews* 57 (3), S. 543–594. Online verfügbar unter <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC372926/>.
- Pristovsek, Primoz; Rüterjans, Heinz; Jerala, Roman (2002): Semiautomatic sequence-specific assignment of proteins based on the tertiary structure--the program st2nmr. In: *J. Comput. Chem.* 23 (3), S. 335–340. DOI: 10.1002/jcc.10011.
- Ried, Andreas; Gronwald, Wolfram; Trenner, Jochen M.; Brunner, Konrad; Neidig, Klaus-Peter; Kalbitzer, Hans Robert (2004): Improved simulation of NOESY spectra by RELAX-JT2 including effects of J-coupling, transverse relaxation and chemical shift anisotropy. In: *Journal of Biomolecular NMR* 30 (2), S. 121–131. DOI: 10.1023/B:JNMR.0000048945.88968.af.
- Rost, B.; Sander, C. (1993): Prediction of protein secondary structure at better than 70% accuracy. In: *Journal of Molecular Biology* 232 (2), S. 584–599. DOI: 10.1006/jmbi.1993.1413.
- Schmidt, Jürgen (2007): Asymmetric Karplus curves for the protein side-chain  $^3\text{J}$  couplings. In: *Journal of Biomolecular NMR* 37 (4), S. 287–301. DOI: 10.1007/s10858-006-9140-8.
- Schumann, Frank H.; Riepl, Hubert; Maurer, Till; Gronwald, Wolfram; Neidig, Klaus-Peter; Kalbitzer, Hans Robert (2007): Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. In: *Journal of Biomolecular NMR* 39 (4), S. 275–289. DOI: 10.1007/s10858-007-9197-z.
- Schwarzinger, Stephan; Kroon, Gerard J.A.; Foss, Ted R.; Wright, Peter E.; Dyson, H. Jane (2000): Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift

- data in NMRView. In: *Journal of Biomolecular NMR* 18 (1), S. 43–48. DOI: 10.1023/A:1008386816521.
- Schwarzinger, Stephan; Kroon, Gerard J. A.; Foss, Ted R.; Chung, John; Wright, Peter E.; Dyson, H. Jane (2001): Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. In: *Journal of the American Chemical Society* 123 (13), S. 2970–2978. DOI: 10.1021/ja003760i.
- Sergey Bochkano: ALGLIB. Version. Online verfügbar unter [www.alglib.net](http://www.alglib.net).
- Shima, Fumi; Ijiri, Yuichi; Muraoka, Shin; Liao, Jingling; Ye, Min; Araki, Mitsugu et al. (2010): Structural basis for conformational dynamics of GTP-bound Ras protein. In: *The Journal of biological chemistry* 285 (29), S. 22696–22705. DOI: 10.1074/jbc.M110.125161.
- Sousa da Silva, Alan W; Vranken, Wim F. (2012): ACPYPE - AnteChamber PYthon Parser interfacE. In: *BMC research notes* 5, S. 367. DOI: 10.1186/1756-0500-5-367.
- Spoerner, M.; Herrmann, C.; Vetter, I. R.; Kalbitzer, H. R.; Wittinghofer, A. (2001): Dynamic properties of the Ras switch I region and its importance for binding to effectors. In: *Proceedings of the National Academy of Sciences of the United States of America* 98 (9), S. 4944–4949. DOI: 10.1073/pnas.081441398.
- Thiruv, B.; Quon, G.; Saldanha, S. A.; Steipe, B. (2005): Nh3D: a reference dataset of non-homologous protein structures. In: *BMC structural biology* 5, S. 12. DOI: 10.1186/1472-6807-5-12.
- Trenner, Jochen Markus (2006): Accurate proton-proton distance calculation and error estimation from NMR data for automated protein structure determination in AUREMOL.
- Ulrich, Eldon L.; Akutsu, Hideo; Doreleijers, Jurgen F.; Harano, Yoko; Ioannidis, Yannis E.; Lin, Jundong et al. (2008): BioMagResBank. In: *Nucleic Acids Research* 36 (suppl 1), S. D402-D408. DOI: 10.1093/nar/gkm957.
- Villa, Alessandra; Stock, Gerhard (2006): What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study. In: *J. Chem. Theory Comput.* 2 (5), S. 1228–1236. DOI: 10.1021/ct600160z.
- Vitek, Olga; Bailey-Kellogg, Chris; Craig, Bruce; Kuliniewicz, Paul; Vitek, Jan (2005): Reconsidering complete search algorithms for protein backbone NMR assignment. In: *Bioinformatics (Oxford, England)* 21 Suppl 2, S. ii230-6. DOI: 10.1093/bioinformatics/bti1138.
- Volk, Jochen; Herrmann, Torsten; Wüthrich, Kurt (2008): Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. In: *Journal of Biomolecular NMR* 41 (3), S. 127–138. DOI: 10.1007/s10858-008-9243-5.
- Wang, Jianyong; Wang, Tianzhi; Zuideweg, Erik R P; Crippen, Gordon M. (2005): CASA: an efficient automated assignment of protein mainchain NMR data using an ordered tree search algorithm. In: *Journal of Biomolecular NMR* 33 (4), S. 261–279. DOI: 10.1007/s10858-005-4079-8.
- Wang, Junmei; Wang, Wei; Kollman, Peter A.; Case, David A. (2006): Automatic atom type and bond type perception in molecular mechanical calculations. In: *Journal of Molecular Graphics and Modelling* 25 (2), S. 247–260. DOI: 10.1016/j.jmgm.2005.12.005.
- Wishart, D. S.; Sykes, B. D.; Richards, F. M. (1991): Relationship between nuclear magnetic

- resonance chemical shift and protein secondary structure. In: *Journal of Molecular Biology* 222 (2), S. 311–333. DOI: 10.1016/0022-2836(91)90214-Q.
- Wittinghofer, Alfred; Waldmann, Herbert (2000): Ras – ein molekularer Schalter bei der Tumorentstehung. In: *Angew. Chem.* 112 (23), S. 4360–4383. DOI: 10.1002/1521-3757(20001201)112:23<4360::AID-ANGE4360>3.0.CO;2-J.
- Wolfram Gronwald; Hans Robert Kalbitzer (2004): Automated structure determination of proteins by NMR spectroscopy. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 44 (1–2), S. 33–96. DOI: 10.1016/j.pnmrs.2003.12.002.
- Wu, Kun-Pin; Chang, Jia-Ming; Chen, Jun-Bo; Chang, Chi-Fon; Wu, Wen-Jin; Huang, Tai-Huang et al. (2006): RIBRA--an error-tolerant algorithm for the NMR backbone assignment problem. In: *Journal of computational biology : a journal of computational molecular cell biology* 13 (2), S. 229–244. DOI: 10.1089/cmb.2006.13.229.
- Wüthrich, Kurt (1986): NMR of proteins and nucleic acids. New York: Wiley (The George Fisher Baker non-resident lectureship in chemistry at Cornell University).
- Xu, J.; Straus, S. K.; Sanctuary, B. C.; Trimble, L. (1994): Use of Fuzzy Mathematics for Complete Automated Assignment of Peptide <sup>1</sup>H <sup>2</sup>D NMR Spectra. In: *Journal of Magnetic Resonance, Series B* 103 (1), S. 53–58. DOI: 10.1006/jmrb.1994.1006.
- Xu, Jun; Straus, Suzana K.; Sanctuary, B. C.; Trimble, Laird (1993): Automation of protein <sup>2</sup>D proton NMR assignment by means of fuzzy mathematics and graph theory. In: *J. Chem. Inf. Model.* 33 (5), S. 668–682. DOI: 10.1021/ci00015a004.
- Xu, Xiao-Ping; Case, David A. (2001): Automated prediction of <sup>15</sup>N, <sup>13</sup>C<sub>α</sub>, <sup>13</sup>C<sub>β</sub> and <sup>13</sup>C' chemical shifts in proteins using a density functional database. In: *Journal of Biomolecular NMR* 21 (4), S. 321–333. DOI: 10.1023/A:1013324104681.
- Xu, Xiao-Ping; Case, David A. (2002): Probing multiple effects on <sup>15</sup>N, <sup>13</sup>C alpha, <sup>13</sup>C beta, and <sup>13</sup>C' chemical shifts in peptides using density functional theory. In: *Biopolymers* 65 (6), S. 408–423. DOI: 10.1002/bip.10276.
- Xu, Ying; Xu, Dong; Kim, Dongsup; Olman, Victor; Razumovskaya, Jane; Jiang, Tao (2002): Automated Assignment of Backbone NMR Peaks Using Constrained Bipartite Matching. In: *Computing in Science & Engineering* 4 (1), S. 50–62. DOI: 10.1109/5992.976437.
- Xu, Yizhuang; Wang, Xiaoxia; Yang, Jun; Vaynberg, Julia; Qin, Jun (2006): PASA--a program for automated protein NMR backbone signal assignment by pattern-filtering approach. In: *Journal of Biomolecular NMR* 34 (1), S. 41–56. DOI: 10.1007/s10858-005-5358-0.
- Zimmerman, Diane E.; Kulikowski, Casimir A.; Huang, Yuanpeng; Feng, Wenqing; Tashiro, Mitsuru; Shimotakahara, Sakurako et al. (1997): Automated analysis of protein NMR assignments using methods from artificial intelligence. In: *Journal of Molecular Biology* 269 (4), S. 592–610. DOI: 10.1006/jmbi.1997.1052.

## 8 Danksagung

An dieser Stelle möchte ich allen danken, die mich unterstützt und somit zum Gelingen dieser Arbeit beigetragen haben:

An erster Stelle bedanke ich mich bei meinem Betreuer, Herrn Prof. Dr. Dr. Hans Robert Kalbitzer, für das interessante Thema, die Betreuung und die Freiräume, die er mir bei der Anfertigung der Arbeit gelassen hat.

Ich bedanke mich bei Claudia Munte, Michael Spörner und Sunilkumar Puthenpurackal Narayanan für die mir zur Verfügung gestellten NMR-Spektren und Zuordnungslisten.

Ferner danke ich meinen Kollegen im AUREMOL Team Harald Donaubauer, Bärbel Kieninger, Nina Danilova, Willhelm Malloni, Silvia de Sanctis, Kumaran Baskaran, Carolina Cano und Konrad Brunner die mir bei Fragen stets zur Seite standen und ein hervorragendes Arbeitsklima geschaffen haben.

Ich danke allen Mitgliedern des Lehrstuhl für Biophysik für für Biophysik für die freundliche und inspirierende Atmosphäre in unserem Team.

Ich bedanke mich auch bei meinen Kommilitonen die mich durch das Physik Studium und durch die Promotion begleitet und mit fachbereichsübergreifenden Ratschlägen zu dieser Arbeit beigetragen haben.

Schließlich bedanke ich mich bei meinen Eltern die mir das Physik Studium und die Promotion erst ermöglicht haben und die ganze Zeit hinter mir standen.