

Schnelle Bestimmung hochaufgelöster Proteinstrukturen mit der FastNMR-Methode**

Jegannath Korukottu, Monika Bayrhuber, Pierre Montaville, Vinesh Vijayan, Young-Sang Jung, Stefan Becker und Markus Zweckstetter*

Die NMR-Spektroskopie ermöglicht die Bestimmung hochaufgelöster Strukturen von Biomolekülen unter annähernd physiologischen Bedingungen. In den letzten 20 Jahren wurden erhebliche Fortschritte in der bioanalytischen NMR-Spektroskopie erzielt,^[1] durch die moderne Genforschung, die eine Vielzahl sequenzierter Genome hervorgebracht hat, werden jedoch Hochdurchsatzverfahren zur Bestimmung der Tertiärstruktur von Proteinen immer wichtiger.^[2,3] Die Datenaufnahme kann erheblich beschleunigt werden, wenn moderne NMR-Spektrometer mit Methoden kombiniert werden, die chemische Verschiebungen in mehrdimensionalen NMR-Experimenten messen.^[4] Die zeitlichen Engpässe bei einer schnellen NMR-basierten Proteinstrukturbestimmung betreffen daher die Datenanalyse und insbesondere die sequenzspezifische Zuordnung chemischer Verschiebungen zu den Atomen der Seitenketten. Hier stellen wir eine als FastNMR (fast structure determination by NMR) bezeichnete Methode vor, die aus nichtzugeordneten NMR-Daten die automatisierte Bestimmung hochaufgelöster Strukturen von eindomänigen Proteinen ermöglicht. Zur Erprobung der Methode haben wir de novo die Struktur des aus 65 Aminosäuren bestehenden Neurotoxins Konkunitzin-S2 der Kegelschnecke bestimmt.

Trotz beachtlicher Fortschritte ist die Qualität der Vorhersagen von 3D-Proteinstrukturen auf der Basis der Aminosäuresequenz in vielen Fällen nicht ausreichend.^[5] Eine Verbesserung kann erzielt werden, indem man eine beschränkte Zahl leicht zugänglicher NMR-Daten zu den Strukturvorhersagen hinzuzieht.^[6–8] Um hochaufgelöste Proteinstrukturen zu erhalten, benötigt man experimentelle Informationen über Abstände zwischen Atomen. Diese können mit geringem experimentellem Aufwand aus NOE-Spektren gewonnen werden, wobei die sequenzspezifische Zuordnung und die Strukturrechnungen iterativ durchgeführt

werden.^[9–13] Um zu einer korrekten Struktur zu gelangen, ist jedoch eine praktisch vollständige und fehlerfreie Zuordnung der chemischen Verschiebungen notwendig.^[14] Daneben kann es gelingen, 3D-Proteinstrukturen ohne jegliche Zuordnung chemischer Verschiebungen aus den in NOESY-Spektren enthaltenen Abstandsinformationen zu bestimmen. Dies setzt jedoch ausgezeichnete, eindeutig zugeordnete NOE-Signallisten voraus.^[15,16] Die FastNMR-Methode unterscheidet sich von diesen Ansätzen darin, dass sie von nicht zugeordneten chemischen Verschiebungen, NOEs und dipolaren Restkopplungen (RDCs) ausgeht und dabei falsche Strukturen durch Kreuzvalidierung ausschließt, auf experimentelle Daten anwendbar ist, nur eine geringe Zahl an NMR-Spektren erfordert und hochaufgelöste Strukturen ($< 1 \text{ \AA}$) liefert.

Die FastNMR-Methode folgt dem in der manuellen Strukturbestimmung bewährten Ansatz. Dieser umfasst die Verwendung von Dreifachresonanzexperimenten zur sequenziellen Zuordnung des Proteinrückgrats, die iterative NOE-Zuordnung und Strukturrechnung sowie die Strukturverfeinerung anhand von RDCs (siehe Hintergrundinformationen). Der entscheidende Fortschritt der FastNMR-Methode ist die gleichzeitige Signalzuordnung und Strukturbestimmung des Proteinrückgrats noch vor der NOE-Analyse. Grundlage der Methode ist die iterative Anwendung des Zuordnungsprogramms MARS und des Strukturvorhersageprogramms Rosetta.^[6,8] In einer zweiten Stufe gelangt man von der Struktur des Proteinrückgrats zur vollständigen 3D-Struktur einschließlich der Seitenketten. Hierzu werden die Seitenketten an das Proteinrückgrat angefügt und die chemischen Verschiebungen der Protonen und Kohlenstoffatome auf der Basis der 20 niederenergetischsten RosettaNMR-Strukturen mithilfe empirischer Formeln und künstlicher neuronaler Netzwerke vorhergesagt.^[17,18] Dann werden die experimentellen chemischen Verschiebungen mit den vorhergesagten Werten verglichen und entsprechend der geringsten Abweichung zugeordnet. Anschließend werden die Struktur des Proteinrückgrats und die zugeordneten chemischen Verschiebungen des Proteinrückgrats und der Seitenketten für eine automatisierte NOE-Zuordnung mit dem Programm Cyana verwendet.^[9] Cyana wird jedoch nicht nur für die NOE-Zuordnung verwendet. Alle Abstandsrandbedingungen, die chemische Verschiebungen betreffen, die nicht eindeutig über den Vergleich mit vorhergesagten Werten zugeordnet werden konnten, werden als mehrdeutige NOEs behandelt. Dadurch wird ein Teil der Zuordnung der chemischen Verschiebungen innerhalb der automatisierten NOE-Analyse vorgenommen. Mithilfe der NOE-basierten 3D-Struktur wird die Qualität der vorhergesagten chemischen Verschiebungen verbessert, und ein zweiter Durchgang

[*] J. Korukottu, M. Bayrhuber, Dr. P. Montaville, V. Vijayan, Dr. Y.-S. Jung, Dr. S. Becker, Dr. M. Zweckstetter
Abteilung für NMR-basierte Strukturbiologie
Max-Planck-Institut für Biophysikalische Chemie
Am Fassberg 11, 37077 Göttingen (Deutschland)
Fax: (+49) 551-201-2202
E-Mail: mzwecks@gwdg.de

[**] Wir danken Christian Griesinger für wertvolle Diskussionen, Karin Giller und Kamila Sabagh für technische Hilfe und Baldomero M. Olivera für die cDNA-Klone von Konk-S1 und Konk-S2. Diese Arbeit wurde durch die Max-Planck-Gesellschaft unterstützt. M.Z. erhält ein Emmy Noether-Stipendium der DFG (ZW 71/1-5).

Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://www.angewandte.de> zu finden oder können beim Autor angefordert werden.

automatisierter NOE-Zuordnung wird gestartet. Schließlich werden alle experimentellen Daten zusammengeführt, und die hochaufgelöste Proteinstruktur wird berechnet (siehe Hintergrundinformationen). FastNMR arbeitet automatisiert: Sie nimmt als Eingabe Listen von nichtzugeordneten NMR-Daten auf und gibt eine hochaufgelöste 3D-Struktur aus.

Die FastNMR-Methode wurde an den Proteinen Konkunitzin-S1 (Konk-S1; 60 Aminosäuren) und Ubiquitin (76 Aminosäuren) erprobt. Für beide Proteine sind sowohl die Zuordnungen der chemischen Verschiebungen als auch die 3D-Strukturen bekannt.^[22,27] Zudem wurde die hochaufgelöste Struktur des aus 65 Aminosäuren bestehenden Toxins Konkunitzin-S2 (72 % Sequenzidentität mit Konk-S1) mit FastNMR bestimmt. Von Konk-S2 gab es bisher weder NMR-Daten noch eine 3D-Struktur. Abbildung 1 und Tabelle 1 belegen, dass wir von allen drei Proteinen aus nichtzugeordneten NMR-Daten hochaufgelöste 3D-Strukturen berechnen konnten. Die Streubreite im Ensemble der 20 niederenergetischsten Strukturen betrug für alle drei Proteine weniger als 0.7 Å für das Proteinerückgrat und weniger als 1.4 Å für alle schweren Atome. Die manuell und automatisiert bestimmten Strukturen haben ähnliche Energien. Die mit FastNMR berechneten Strukturen von Konk-S1 (Abbildung 1c) und Ubiquitin (Abbildung 1b) weichen um 0.4 Å bzw. 0.6 Å von den bereits beschriebenen Strukturen ab.^[22,27] Je Protein benötigten die FastNMR-Rechnungen weniger als 24 Stunden.

Wie oben erwähnt werden bei der FastNMR-Methode die chemischen Verschiebungen der Seitenketten automatisiert zugeordnet, durch Vergleich mit Werten, die aus der Struktur des Proteinerückgrats vorhergesagt wurden. Kontrollen belegen, dass die mittlere quadratische Abweichung (rmsd) zwischen vorhergesagten und experimentellen chemischen Verschiebungen 0.19 ppm für Protonen und 1.1 ppm für Kohlenstoffatome beträgt (sichtbar in HCONH- und CCONH-TOCSY-Spektren; siehe Hintergrundinformationen).^[17,18] Daher werden nur Zuordnungen berücksichtigt, für die die Abweichung zwischen vorhergesagten und gemessenen chemischen Verschiebungen weniger als 0.3 ppm für Protonen und weniger als 1.3 ppm für Kohlenstoffatome beträgt. Wenn außerdem zwei gemessene ¹H-chemische Verschiebungen, die entsprechend den HCONH- und CCONH-TOCSY-Spektren zum selben Rest gehören, um weniger als 0.3 ppm voneinander abweichen, werden alle NOEs zu diesen chemischen Verschiebungen in der automatisierten NOE-Zuordnung als mehrdeutig behandelt. Diesem Ansatz folgend konnten alle experimentellen che-

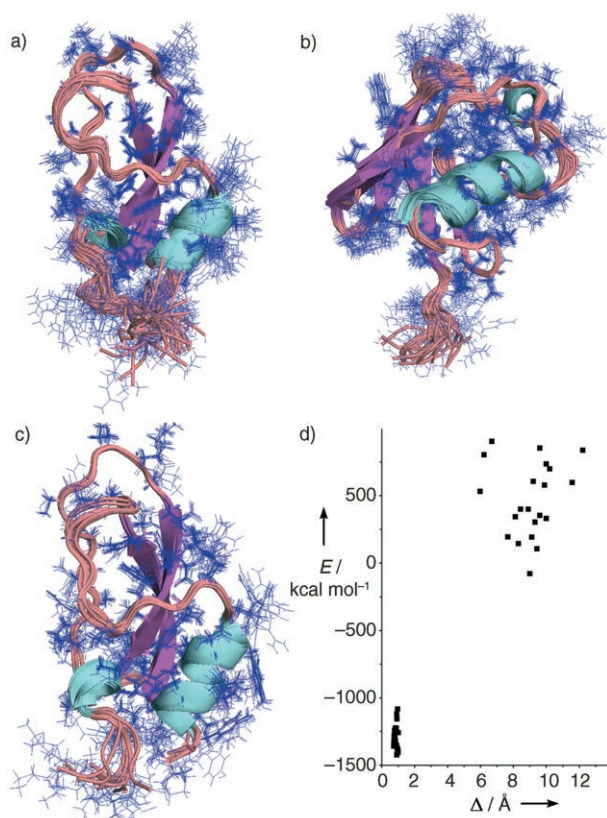


Abbildung 1. Mit FastNMR berechnete 3D-Strukturen von a) Konk-S2, b) Ubiquitin und c) Konk-S1. d) Energien der mit FastNMR berechneten Teststrukturen von Konk-S1, aufgetragen gegen die Abweichung Δ von der nativen Struktur (siehe Hintergrundinformationen).

Tabelle 1: Statistische Daten der untersuchten Proteine.^[a]

	Konk-S1		Ubiquitin		Konk-S2
	2CA7	FastNMR	1D3Z ^[b]	FastNMR	FastNMR
Zahl der NOEs ^[c]	551	464	1744 ^[d]	635	570
große Reichweite	113	72	731	119	160
mittlere Reichweite	73	57	291	106	79
kurze Reichweite	365	335	722	410	331
Zahl der Torsionswinkel	126	126	98	127	167
Verletzungen > 5°	3 ± 1	3 ± 1	0	1 ± 1	2 ± 1
Zahl der RDCs	190	190	200	200	138
RDC-Arten ^[e]	1,2,3,4	1,2,3,4	1,2,3	1,2,3	1,2,3
Energie [kcal mol ⁻¹]	-1267.2	-1387.9	-2767.9	-2247.6	-1025.0
Ramachandran-Statistik [%]					
bevorzugt	88.2	87.5	95.0	95.5	84.5
nicht erlaubt	2.0	1.0	0.0	0.0	1.9
Genauigkeit [Å] ^[f]					
Proteinerückgrat	0.6	0.7	0.3	0.4	0.6
alle schweren Atome	1.2	1.4	0.9	1.1	1.2

[a] Statistik für Ensembles aus 20 Strukturen. [b] Daten beziehen sich auf eine Struktur, die auf Basis der experimentellen Daten von 1D3Z neu berechnet wurde. [c] Bei keiner der Strukturen gab es Verletzungen der Abstandsrandbedingungen von mehr als 0.5 Å. [d] Nur 58 % der Weitbereichs-NOEs sind nicht-redundant. [e] 1, 2, 3, 4 steht für die RDCs ¹D_{N-H}, ¹D_{C-N}, ¹D_{Cα-C}, ¹D_{Cα-Hα}. [f] Definiert als die mittlere quadratische Abweichung (rmsd) zwischen den 20 FastNMR-Strukturen und den gemittelten Koordinaten der Reste 2–72 (Ubiquitin), 3–60 (Konk-S1) und 5–60 (Konk-S2).

mischen Verschiebungen der Kohlenstoffatome in Conk-S1, Conk-S2 und Ubiquitin eindeutig zugeordnet werden. Chemische Verschiebungen der Protonen sind oft entartet, sodass etwa 10 % der gemessenen chemischen Verschiebungen der Seitenkettenprotonen nicht eindeutig zugeordnet werden konnten (siehe Hintergrundinformationen).

In früheren Arbeiten wurde erkannt,^[14] dass für eine erfolgreiche automatisierte NOE-Zuordnung mehr als 90 % der ¹H-chemischen Verschiebungen zugeordnet sein müssen. In der aktuellen Version nutzt FastNMR jedoch nur 3D-CCONH- und HCCONH-TOCSY-Experimente, und nur ca. 60 % aller Protonen konnten vor Beginn der NOE-Analyse zugeordnet werden (Hintergrundinformationen). Für alle Protonen, für die keine experimentellen chemischen Verschiebungen verfügbar sind, nutzt FastNMR vorhergesagte Werte in der automatisierten NOE-Zuordnung. Da die vorhergesagten chemischen Verschiebungen nicht sehr genau sind, wird das Datenfenster, das für den Vergleich der NOEs mit den ¹H-chemischen Verschiebungen verwendet wird, von 0.05 ppm auf 0.3 ppm erhöht. Darüber hinaus werden NOE-Abstandsrandbedingungen für Protonen mit vorhergesagten chemischen Verschiebungen in der abschließenden Strukturverfeinerung nur verwendet, wenn das gleiche Proton (vorhergesagte chemische Verschiebung) mindestens zwei NOE-Signalen zugeordnet ist und die experimentellen chemischen Verschiebungen der beiden NOE-Signale um weniger als 0.1 ppm voneinander abweichen. In Verbindung mit der Struktur des Proteinrückgrats, die bereits vor Beginn der NOE-Analyse bestimmt wird, ermöglicht dies die Bestimmung hochaufgelöster Proteinstrukturen.

Zur Erprobung der FastNMR-Methode an den drei genannten Proteinen wurden alle hinlänglich bekannten Probleme wie Signalüberlappungen, fehlende Signale des Proteinrückgrats, Rauschsignale und Mehrfachkonformationen miteinbezogen. NOE-Signale mit mehrfachen Zuordnungsmöglichkeiten wurden durch die Verwendung von mehrdeutigen Abstandsrandbedingungen ohne Einschränkungen berücksichtigt. Darüber hinaus wurde der Einfluss reduzierter Datenqualität und falscher Zuordnungen des Proteinrückgrats untersucht (Hintergrundinformationen). Trotz dieser Schwierigkeiten konnten hochaufgelöste Strukturen berechnet werden, insbesondere auch für das bisher nicht charakterisierte Neurotoxin Conk-S2. Diese Ergebnisse belegen, dass die FastNMR-Methode sehr robust ist.

Um sicherzustellen, dass FastNMR keine falschen Strukturen berechnet, wird eine Kreuzvalidierung genutzt. Zur Signalzuordnung und Strukturbestimmung des Proteinrückgrats werden ausschließlich RDCs und chemische Verschiebungen verwendet, wohingegen keine RDCs in die automatisierte NOE-Zuordnung einfließen. Für den Fall einer falschen Struktur des Proteinrückgrats ist es dann äußerst unwahrscheinlich, dass eine ausreichende Zahl an NOEs im Verlauf der automatisierten strukturbasierten NOE-Analyse zugeordnet werden kann. Selbst wenn eine große Zahl an NOEs zugeordnet wurde, wird die resultierende NOE-basierte Struktur voraussichtlich von der zuvor bestimmten Struktur des Proteinrückgrats abweichen und den RDCs widersprechen. In diesem Fall ist in der abschließenden Stufe der FastNMR-Rechnung – wenn alle experimentellen NMR-

Daten kombiniert werden – keine Konvergenz zu einer niedrigerenergetischen Struktur möglich. Dies wird aus Abbildung 1d sowie zusätzlichen Stabilitätsuntersuchungen deutlich: Nur bei „richtigen“ Strukturen erzielt FastNMR niedrige Energien. Zudem müssen mit FastNMR berechnete Strukturen folgende Kriterien erfüllen: 1) Am Ende jeder FastNMR-Stufe müssen die berechneten Strukturen zu einer einzigen Konformation konvergiert sein. 2) Strukturelle Änderungen, die im Verlauf der FastNMR-Stufen von der anfänglichen Struktur des Proteinrückgrats bis zur hochaufgelösten Struktur auftreten, dürfen nur weniger als 3.5 Å betragen. 3) Mehr als 85 % aller Signale des Proteinrückgrats müssen vor Beginn der automatisierten NOE-Analyse zugeordnet worden sein. 4) FastNMR-Strukturen müssen die in der NMR-Spektroskopie üblichen Standardqualitätsmerkmale erfüllen, z.B. eine möglichst geringe Zahl an Verletzungen der experimentellen Randbedingungen (Tabelle 1).

In der aktuellen Implementierung kann FastNMR nur auf Proteine angewendet werden, die aus einer einzigen Domäne bestehen. Dies liegt vor allem daran, dass die chemischen Verschiebungen der Seitenketten ausschließlich aus CCONH- und HCCONH-Experimenten gewonnen werden. Die Leistungsfähigkeit dieser Experimente nimmt jedoch mit zunehmendem Molekulargewicht der Proteine ab, und zudem sind die chemischen Verschiebungen der aromatischen Gruppen nicht verfügbar. Eine größere Zahl chemischer Verschiebungen ist mit 3D-HCCH-COSY- und 3D-HCCH-TOCSY-Spektren zugänglich.^[19] Darüber hinaus können chemische Verschiebungen aromatischer Gruppen aus zweidimensionalen ($(H_\beta)C_\beta(C_\gamma C_\delta)H_\delta$ - und $(H_\beta)C_\beta(C_\gamma C_\delta C_\epsilon)H_\epsilon$ -Spektren gewonnen werden.^[20] Die Implementierung solcher Experimente in die FastNMR-Methode ist Gegenstand laufender Arbeiten.

Zusammenfassend haben wir eine als FastNMR bezeichnete Methode entwickelt, mit der hochaufgelöste Strukturen von Proteinen mittlerer Größe aus nichtzugeordneten chemischen Verschiebungen, RDCs und NOE-Signallisten innerhalb von 24 Stunden bestimmt werden können. Die Methode wurde verwendet, um de novo die Struktur des Neurotoxins Conk-S2 zu bestimmen. Sie läuft automatisiert, verhindert falsche Strukturen durch Kreuzvalidierung, greift direkt auf experimentelle Daten zu, erfordert nur eine beschränkte Zahl an NMR-Spektren und erzeugt hochaufgelöste Strukturen. Eine manuelle Zuordnung chemischer Verschiebungen oder Korrelationen zwischen Aminosäuren sind nicht erforderlich.

Experimentelles

Ubiquitin und Conk-S1 wurden nach Literaturangaben rekombinant hergestellt.^[21,22] Die Herstellung und Strukturdetails von Conk-S2 werden an anderer Stelle veröffentlicht. Die NMR-Spektren wurden je nach Verfügbarkeit an 600-, 700- oder 900-MHz-Spektrometern (Bruker) gemessen. Spektren und Bedingungen finden sich in den Hintergrundinformationen. Zur Referenzierung der Spektren, Signaldetektion und Signalgruppierung wurde das Programm Sparky verwendet. Torsionswinkel χ_1 wurden aus 2D-¹⁵N-¹³C γ - und ¹³C-¹³C γ -Spin-Echo-Differenzexperimenten gewonnen,^[23] NOE-Signalvolumen aus 3D-¹⁵N- und ¹³C-editierten NOESY-Spektren.^[24,25] RDCs wurden mit verschachtelten 3D-TROSY-HNCO- und 3D-CBCA-

(CO)NH-Spektren gemessen.^[26] Die FastNMR-Rechnungen wurden auf einem aus zehn 3.06-GHz-Linux-PCs bestehenden Verbund ausgeführt. Für den Großteil der Rechnungen wurde nur eine CPU verwendet. Die Struktur von Conk-S2 wurde in der Protein Data Bank hinterlegt (PDB 2J6D). Software zur Ausführung der FastNMR-Methode ist von den Autoren erhältlich.

Eingegangen am 8. August 2006

Online veröffentlicht am 5. Januar 2007

Stichwörter: Genomik · NMR-Spektroskopie · Proteinstrukturen · Strukturaufklärung

- [1] K. Wüthrich, *Angew. Chem.* **2003**, *115*, 3462; *Angew. Chem. Int. Ed.* **2003**, *42*, 3340.
- [2] A. Abbott, *Nature* **2005**, *435*, 547.
- [3] G. T. Montelione, D. Y. Zheng, Y. P. J. Huang, K. C. Gunsalus, T. Szyperski, *Nat. Struct. Biol.* **2000**, *7*, 982.
- [4] G. H. Liu, Y. Shen, H. S. Atreya, D. Parish, Y. Shao, D. K. Sukumaran, R. Xiao, A. Yee, A. Lemak, A. Bhattacharya, T. A. Acton, C. H. Arrowsmith, G. T. Montelione, T. Szyperski, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10487.
- [5] P. Bradley, K. M. Misura, D. Baker, *Science* **2005**, *309*, 1868.
- [6] Y. S. Jung, M. Sharma, M. Zweckstetter, *Angew. Chem.* **2004**, *116*, 3561; *Angew. Chem. Int. Ed.* **2004**, *43*, 3479.
- [7] J. Meiler, D. Baker, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15404.
- [8] C. A. Rohl, D. Baker, *J. Am. Chem. Soc.* **2002**, *124*, 2723.
- [9] P. Guntert, *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *43*, 105.
- [10] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, *Bioinformatics* **2003**, *19*, 315.
- [11] J. Kuszewski, C. D. Schwieters, D. S. Garrett, R. A. Byrd, N. Tjandra, G. M. Clore, *J. Am. Chem. Soc.* **2004**, *126*, 6258.
- [12] W. Gronwald, S. Moussa, R. Elsner, A. Jung, B. Ganslmeier, J. Trenner, W. Kremer, K. P. Neidig, H. R. Kalbitzer, *J. Biomol. NMR* **2002**, *23*, 271.
- [13] Y. J. Huang, R. Tejero, R. Powers, G. T. Montelione, *Proteins Struct. Funct. Genet.* **2006**, *62*, 587.
- [14] J. Jee, P. Guntert, *J. Struct. Funct. Genomics* **2003**, *4*, 179.
- [15] A. Grishaev, M. Llinas, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10941.
- [16] A. Grishaev, C. A. Steren, B. Wu, A. Pineda-Lucena, C. Arrowsmith, M. Llinas, *Proteins Struct. Funct. Genet.* **2005**, *61*, 36.
- [17] J. Meiler, *J. Biomol. NMR* **2003**, *26*, 25.
- [18] K. Osapay, D. A. Case, *J. Biomol. NMR* **1994**, *4*, 215.
- [19] L. E. Kay, G. Y. Xu, A. U. Singer, D. R. Muhandiram, J. D. Forman-Kay, *J. Magn. Reson. Ser. B* **1993**, *101*, 333.
- [20] T. Yamazaki, J. D. Forman-Kay, L. E. Kay, *J. Am. Chem. Soc.* **1993**, *115*, 11054.
- [21] G. A. Lazar, J. R. Desjarlais, T. M. Handel, *Protein Sci.* **1997**, *6*, 1167.
- [22] M. Bayrhuber, R. Graf, M. Ferber, M. Zweckstetter, J. Imperial, J. E. Garrett, B. M. Olivera, H. Terlau, S. Becker, *Protein Expression Purif.* **2006**, *47*, 640.
- [23] A. Bax, G. W. Vuister, S. Grzesiek, F. Delaglio, A. C. Wang, R. Tschudin, G. Zhu, *Methods Enzymol.* **1994**, *239*, 79.
- [24] S. Talluri, G. Wagner, *J. Magn. Reson. Ser. B* **1996**, *112*, 200.
- [25] Y. Xia, A. Yee, C. H. Arrowsmith, X. Gao, *J. Biomol. NMR* **2003**, *27*, 193.
- [26] V. Vijayan, M. Zweckstetter, *J. Magn. Reson.* **2005**, *174*, 245.
- [27] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 6836.