

**Proteinstrukturen****Simultane Signalzuordnung und Strukturbestimmung des Rückgrats von Proteinen mithilfe dipolarer NMR-Kopplungen\*\***

Young-Sang Jung, Mukesh Sharma und Markus Zweckstetter\*

Durch die Vielzahl sequenziertener Genome<sup>[1]</sup> werden Hochdurchsatzverfahren zur Bestimmung der Tertiärstruktur von Proteinen immer wichtiger. Zwar liegt der Schwerpunkt der so genannten „strukturellen Genomik“ bisher auf der Röntgenkristallographie, doch ist die NMR-Spektroskopie in Lösung eine vielversprechende Alternative.<sup>[2,3]</sup> In der traditionellen NMR-basierten Proteinstrukturbestimmung steht vor der eigentlichen Strukturbestimmung eine zeitaufwändige, vorwiegend manuelle sequenzielle Signalzuordnung.<sup>[3]</sup> Hier beschreiben wir ein automatisches Verfahren, bei dem, auf der Basis chemischer Verschiebungen und dipolarer Restkopplungen (residual dipolar couplings, RDCs), die sequenzielle Signalzuordnung und die Bestimmung von Proteinrückgratstrukturen in mittlerer Auflösung gleichzeitig

erfolgen. Die neue Strategie mit Namen Itas (ITerative Assignment and Structure) ist für Proteine mit bis zu ca. 125 Aminosäuren anwendbar und beschleunigt die NMRspektroskopische Proteinstrukturbestimmung erheblich.

Itas startet mit einer Liste nichtzugeordneter chemischer Verschiebungen und RDCs, die entsprechend ihrer gemeinsamen <sup>1</sup>H/<sup>15</sup>N-Verschiebungen gruppiert sind (siehe Hintergrundinformationen). Zunächst werden die Signale anhand der chemischen Verschiebungen automatisch zugeordnet. Hierzu werden Methoden verwendet, die ausgehend von früheren Verfahren weiterentwickelt wurden.<sup>[4]</sup> Information über die Tertiärstruktur ist zu diesem Zeitpunkt noch nicht verfügbar. Daher hängt die Qualität der sequenziellen Signalzuordnung stark von der Primärsequenzlänge sowie der Art und der Qualität vorhandener NMR-Spektren ab (Tabelle 1). Zugeordnete RDCs und chemische Verschiebungen werden anschließend in einer Molekülfragmentsuche verwendet, in der Fragmente aus einer nichtredundanten Datenbank von Proteinkristallstrukturen mit einer Auflösung von mindestens 2 Å extrahiert werden. Dies geschieht mithilfe eines Mehrfachsequenzabgleichs sowie anhand der Übereinstimmung zwischen experimentellen und für jedes Strukturfragment vorhergesagten RDCs und der chemischen Verschiebungen.<sup>[5,6]</sup> Anschließend wird aus diesen Fragmen-

**Tabelle 1:** Gleichzeitige Zuordnung und Strukturbestimmung von Proteinen unterschiedlicher Topologie und Primärsequenzlänge.

Protein (PDB-Schlüssel)	Länge	CS <sup>[a]</sup>	Zuordnung [%] (Fehlerzahl) <sup>[c]</sup>		RDC-Typ <sup>[d]</sup>	Zahl an RDCs	Aminosäuren <sup>[e]</sup>	Rückgrat rmsd [Å] <sup>[f]</sup>	
			Start	Ende				Itas	Voll
Interleukin-1β <sup>[g]</sup> (211B)	153	Ca/Cβ	60 (0)	98 (2)	1,2,3,4	553	2–140	2.1	1.9
Calmodulin (1J7O)	148	Ca/Cβ	49 (0)	99 (0)	1,2,3	396	5–75	2.0	2.3
Calmodulin <sup>[h]</sup> (1J7P)							82–146	1.4	2.4
Profilin <sup>[g]</sup> (1ACF)	125	Ca/Cβ	82 (0)	98 (2)	1,2,3	341	2–125	1.5	1.3
KH-Domäne (1KHM)	89	Ca	8 (0)	98 (3)	1,2,3	153	12–84	1.8	1.6
DinI (1GHH)	81	Ca	38 (0)	97 (1)	1,2,3,4	271	2–80	5.1	5.4
Ubiquitin (1UBI)	76	Ca	22 (0)	96 (2)	1,2,4	185	2–72	2.0	1.4
Z-Domäne <sup>[g]</sup> (2SPZ)	71	Ca/Cβ <sup>[b]</sup>	77 (0)	99 (2)	1,2,3	192	15–70	3.6	2.4
		Ca	15 (0)	78 (6)	1,2,3	174	15–70	5.0	2.4
Protein G <sup>[i]</sup> (1PGB)	56	Ca	2 (0)	95 (0)	1,2,3	145	2–55	0.7	1.9

[a] Chemische Verschiebungen für sequenzielle Konnektivität. Die Schwellenwerte für Konnektivität waren 0.3 ppm für C<sub>a</sub> und 0.5/0.5 ppm für C<sub>a</sub>/C<sub>β</sub>.

[b] Die Schwellenwerte waren 0.3 ppm für C<sub>a</sub> und 0.5 ppm für C<sub>β</sub>. [c] Zuordnungsanteile wurden analog zu früheren Verfahren bestimmt.<sup>[4]</sup> Start: ohne RDCs; Ende: am Ende von Itas. [d] <sup>1</sup>D<sub>NH</sub>, <sup>1</sup>D<sub>CaC</sub>, <sup>1</sup>D<sub>CaHw</sub>, <sup>1</sup>D<sub>NC</sub> sind mit 1, 2, 3 und 4 gekennzeichnet. [e] Strukturierter Bereich des Proteinrückgrats (Schweratome), der für die Berechnungen der mittleren quadratischen Abweichung (rmsd) verwendet wurde. [f] Für die Struktur niedrigster Energie relativ zur nativen Struktur. „Voll“ kennzeichnet Strukturen berechnet auf der Basis veröffentlichter Zuordnungen. [g] RDCs, simuliert durch 3D-Formvorhersage.<sup>[9]</sup> [h] RDCs, simuliert für Reste 77–81. [i] Der Ausrichtungstensor nach der ersten Zuordnung wurde durch 3D-Formvorhersage ermittelt.<sup>[9]</sup>

[\*] Y.-S. Jung, M. Sharma, Dr. M. Zweckstetter  
NMR-basierte Strukturbioologie  
Max-Planck-Institut für Biophysikalische Chemie  
Am Faßberg 11, 37077 Göttingen (Deutschland)  
Fax: (+49) 551-201-2202  
E-mail: mzwecks@gwdg.de

[\*\*] Wir danken Gaetano Montelione für die Signaltabellen der Z-Domäne und Christian Griesinger für konstruktive Diskussionen. M.Z. wird durch das Emmy Noether-Programm (ZW 71/1-3) der DFG gefördert.

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

ten mit dem Programm RosettaNmr eine Proteinstruktur aufgebaut.<sup>[6]</sup> 1000 Strukturmodelle niedriger Auflösung werden erzeugt und entsprechend ihrer Übereinstimmung mit experimentellen RDCs und energetisch günstigen Fernwechselwirkungen (Packungsdichte, hydrophobe Wechselwirkungen, Paarung von β-Strängen) bewertet.

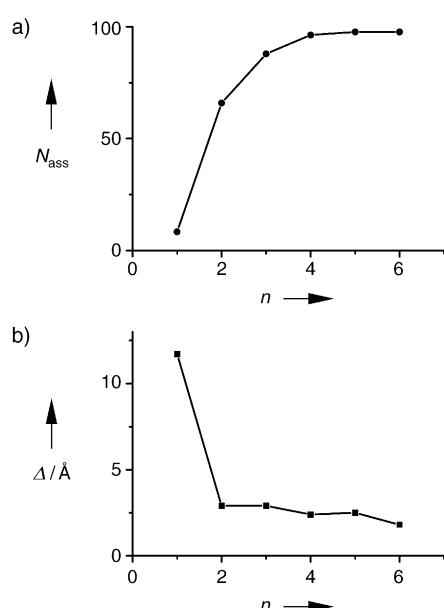
Die 20 Strukturen mit der niedrigsten Energie werden in einer zweiten Iterationsstufe in Itas für eine verbesserte Signalzuordnung verwendet: Über den Vergleich von RDCs, die aus diesen Strukturmodellen zurückgerechnet wurden, mit experimentellen Werten wird die Zahl der zugeordneten Signale erheblich vergrößert. Die Signalzuordnung beruht

somit nicht mehr nur auf chemischen Verschiebungen. Da jedoch diese Strukturmodelle von geringer Auflösung sind, ist eine Signalzuordnung für alle 20 Strukturen erforderlich, und nur Zuordnungen, die für alle 20 Strukturen identisch sind, werden weiterverwendet (siehe Hintergrundinformationen). Ein höherer Zuordnungsgrad ist gleichbedeutend mit einer höheren Zahl an chemischen Verschiebungen und RDCs, die für eine neue Molekülfragmentsuche und Strukturrechnung zur Verfügung stehen. Das iterative Verfahren wird fortgeführt, bis die Zahl der zugeordneten Signale stagniert und eine Tertiärstruktur mittlerer Auflösung erreicht ist. In der Regel sind vier bis acht Iterationsschritte erforderlich. Itas wird mit einer kurzen Energieminimierung der Strukturen (die mit der größten Zahl zugeordneter Signale berechnet wurden) abgeschlossen.

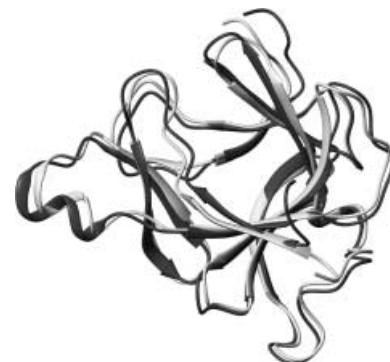
Itas wurde an acht Proteinen unterschiedlicher Topologie und mit Primärsequenzlängen im Bereich von 56 bis 153 Aminosäuren getestet (Tabelle 1). Zu Beginn der Iteration konnten auf der Basis von  $^{13}\text{C}^\alpha$ -Verschiebungen (0.3 ppm maximale Abweichung für die Etablierung von sequenzieller Konnektivität) 8 % der C-terminalen KH-Domäne des heterogenen Ribonucleoproteins zugeordnet werden. Damit standen nur insgesamt 14 RDCs für die erste Strukturrechnung zur Verfügung, und die Abweichung zwischen der mit hochauflösender NMR-Spektroskopie bestimmten Struktur und den 20 Strukturmodellen mit der niedrigsten Energie betrug im Mittel ca. 14.5 Å (Minimalwert 9.3 Å). Trotz der großen Unterschiede zur nativen Struktur erlauben diese Strukturmodelle bereits eine Verbesserung der Signalzuordnung. Der Vergleich von  $^1\text{D}_{\text{NH}}$ ,  $^1\text{D}_{\text{CaC}}$ - und  $^1\text{D}_{\text{CaHa}}$ -Kopplungen mit aus den Strukturen zurückgerechneten Werten erhöht den Zuordnungsgrad auf 66 % (Abbildung 1). Eine solche

erhebliche Verbesserung ist möglich, da Teile der mit RosettaNmr bestimmten 14.5-Å-Strukturen (besonders Sekundärstrukturelemente und teilweise deren relative Orientierung) bereits mit der nativen Struktur übereinstimmen.

Im weiteren Verlauf stieg die Zahl der Signalzuordnungen für alle getesteten Proteine wegen der kontinuierlich besser werdenden Strukturmodelle rapide an: Ausgehend von z.B. nur 2 % zugeordneter Signale für Protein G wurden nach Abschluss von Itas im Mittel mehr als 96 % aller Reste richtig zugeordnet. Folglich standen viele chemische Verschiebungen und RDCs für die abschließenden Strukturrechnungen zur Verfügung. Im Fall der KH-Domäne waren es 153 RDCs, und die Struktur mit der niedrigsten Energie weicht nur 1.8 Å von der hochauflösten NMR-Struktur ab (Tabelle 1). Bei Verwendung aller 157 gemessenen RDCs ergibt sich entsprechend der veröffentlichten Zuordnung eine Abweichung von 1.6 Å. Dies belegt, dass die drei falschen und zwei fehlenden Signalzuordnungen letztlich keinen großen Einfluss auf die Qualität der Rückgratstruktur haben. Ähnliche Ergebnisse wurden für sieben andere Proteine einschließlich des aus 153 Resten bestehenden Interleukin 1 $\beta$  erhalten (Abbildung 2 und Tabelle 1).



**Abbildung 1.** a) Anstieg des Anteils an zugeordneten Resten  $N_{ass}$  der KH-Domäne im Verlauf von Itas. b) Verringerung der mittleren quadratischen Abweichung  $\Delta$  der Itas-Struktur (Mittelwert über die fünf Strukturen niedrigster Energie) von der nativen Struktur (PDB-Schlüssel: 1KHM).  $n$  ist die Iterationsnummer.



**Abbildung 2.** Vergleich der automatisch bestimmten, „zuordnungs-freien“ Itas-Struktur (hellgrau) mit der 2.0-Å-Kristallstruktur von Interleukin-1 $\beta$  (dunkelgrau) (PDB-Schlüssel: 2I1B).

Für die KH-Domäne wurden die chemischen Verschiebungen und RDCs aus der BMRB-Datenbank entnommen.<sup>[7]</sup> Um Itas einer strengeren Prüfung zu unterziehen, wurde das Verfahren auf Signaltabellen der Z-Domäne des Staphykokkenproteins A angewendet. Die Tabellen wurden ohne nachträgliche Veränderungen über eine automatische Signalerkennungsroutine aus den NMR-Spektren gewonnen.<sup>[8]</sup> Werden nur chemische Verschiebungen der  $\text{C}^\alpha$ -Atome für die Ermittlung sequenzieller Konnektivität (Schwellenwert 0.3 ppm) sowie RDCs dreier verschiedener internuclearer Vektorarten verwendet, so erhöht sich der Grad der Signalzuordnung von 15 auf 78 %, und die resultierende Struktur weicht um 5 Å von der nativen Struktur ab. Mithilfe von RosettaNmr erhält man bei Verwendung aller 201 für die Z-Domäne beobachtbaren RDCs (dies entspricht 100 % Signalzuordnung) eine 2.4-Å-Struktur. Die schlechtere Qualität der Struktur ist in diesem Fall nicht überraschend, da mit

Abschluss von Itas eine erheblich geringere Zahl an RDCs für die Strukturrechnung zur Verfügung stand. Entscheidend ist jedoch, dass die sechs falschen Zuordnungen zu keiner Verschlechterung der Strukturqualität führten, wie eine Korrektur der Zuordnungen mit anschließender Strukturrechnung (basierend auf 51 richtigen Signalzuordnungen) zeigte. Die hohe Stabilität gegenüber einer kleinen Zahl an Fehlzuordnungen wird dadurch erreicht, dass 25 % der Molekülfragmente, mit denen in RosettaNmr die Tertiärstruktur aufgebaut wird, ohne chemische Verschiebungen und RDCs selektiert werden, d.h., die Auswahl basiert ausschließlich auf dem Mehrfachsequenzabgleich und der vorhergesagten Sekundärstruktur. Weiterhin konnte die Abweichung der Itas-Struktur von der nativen Struktur der Z-Domäne durch Verwendung von C<sup>α</sup>- und C<sup>β</sup>-Konnektivitätsinformation auf 3.6 Å reduziert werden. Zudem verringerte sich die Zahl der Fehlzuordnungen auf zwei, wobei die Zuordnungen für L58 und N65 vertauscht waren (beide Reste befinden sich in Helix 3 und haben keine C<sup>α</sup>- und C<sup>β</sup>-Verschiebungen).

Itas ermöglicht die Strukturbestimmung des Rückgrats verschiedener Proteine mit einer Auflösung von weniger als 2–3 Å. Für Proteine wie das RecA bindende Protein DinI beträgt die Abweichung jedoch 5–6 Å, auch wenn vor der Strukturrechnung eine vollständige manuelle Signalzuordnung durchgeführt wurde (Tabelle 1, mit „Voll“ gekennzeichnete mittlere quadratische Abweichungen). Um eine hochaufgelöste Struktur für diese Proteine zu erhalten, sind zusätzliche experimentelle Daten erforderlich. Falsche Strukturen werden durch Itas nicht erzeugt: Eine Itas-Struktur wird nur als verlässlich eingestuft, wenn mehr als 95 % der Signale des Proteinrückgrats zugeordnet wurden. Dadurch wird gewährleistet, dass die abschließende Strukturrechnung vorwiegend auf experimentellen RDCs beruht und statistische, von Rosetta verwendete Energieterme eine untergeordnete Rolle spielen. Sollte die Qualität und Menge der vorhandenen experimentellen Daten nicht ausreichen, werden die 20 Strukturen mit der geringsten Energie nicht die gleiche Faltung haben. Darüber hinaus kann die Abweichung von der nativen Struktur durch einen Vergleich mit experimentellen H<sup>N</sup>-H<sup>N</sup>-NOEs, die nicht in der Itas-Rechnung verwendet wurden, überprüft werden (siehe Hintergrundinformationen).

Ist eine mit Itas erhaltene Struktur von geringer Auflösung, so wird dies in bestimmten Fällen auch durch den Vergleich experimenteller RDCs (die unter Verwendung nahezu ungeladener Bizepse gemessen und durch Itas zugeordnet wurden) mit RDC-Werten, die aus der dreidimensionalen Form der Itas-Struktur vorhergesagt wurden, deutlich.<sup>[9]</sup> Obwohl bewegliche Proteinbereiche einen großen Einfluss auf die Proteinform haben können, ist für kleine, kompakte Proteine eine niedrige Korrelation R<sub>S</sub> zwischen experimentellen und formbasierten RDCs ein Zeichen für strukturelle Abweichungen. So bilden die Reste 75 bis 78 in der nativen Struktur von DinI eine kleine α-Helix, die L21 berührt. In der 5.1-Å-Itas-Struktur hingegen ist die Orientierung dieser Helix invertiert und zeigt in Richtung S11. Dies führt zu einem R<sub>S</sub>-Wert von 0.55 im Vergleich zu 0.84 im Falle der 0.7-Å-Itas-Struktur von Protein G. Hierbei muss jedoch

berücksichtigt werden, dass eine Struktur geringer Auflösung auch einen hohen R<sub>S</sub>-Wert haben kann.

Dieser Beitrag zeigt, dass ausgehend von nichtzugeordneten chemischen Verschiebungen und RDCs die Struktur des Rückgrats von Proteinen schnell und ohne manuelles Eingreifen bestimmt werden kann. In früheren Untersuchungen wurde nachgewiesen, dass die Faltung von Proteinen unter ausschließlicher Verwendung von RDCs erhalten werden kann. Hierzu musste jedoch zuerst die zeitaufwändige Signalzuordnung durchgeführt werden.<sup>[5,10]</sup> Ein erster Schritt in Richtung einer zuordnungsfreien Strukturbestimmung wurde kürzlich für das 54-Reste-Protein Rubredoxin gemacht: Ein umfangreicher Satz an RDCs aus zwei flüssigkristallinen Medien wurde als Strukturinformation und als Filter für die sequenzielle Verbindung von Proteinresten verwendet.<sup>[11]</sup> Itas geht erheblich weiter, indem es die automatische Faltungsbestimmung von Proteinen bis zu einer Länge von ca. 125 Aminosäuren mit weniger umfangreichen Datensätzen erlaubt. Darüber hinaus kann – da sequenzielle C<sup>β</sup>-Information nicht erforderlich ist – Itas auf kleine bis mittlere Proteine mit natürlichem <sup>13</sup>C-Anteil angewendet werden. Strukturmodelle mittlerer Auflösung, wie sie von Itas erzeugt werden, ermöglichen oftmals Rückschlüsse auf die biochemische Funktion.<sup>[12]</sup> Zudem können sie, wenn zusätzliche NOEs verfügbar sind, als nützliche Startpunkte für die Bestimmung einer hochauflösten 3D-Struktur dienen.

Eingegangen am 19. Dezember 2003,  
veränderte Fassung am 16. Februar 2004 [Z53588]

**Stichwörter:** Automatische Verfahren · Genomik · NMR-Spektroskopie · Proteinstrukturen

- [1] a) International Human Genome Sequencing Consortium, *Nature* **2001**, 409, 860; b) J. C. Venter et al. *Science* **2001**, 291, 1304.
- [2] a) A. Abbott, *Nature* **2000**, 408, 130; b) G. T. Montelione, D. Y. Zheng, Y. P. J. Huang, K. C. Gunsalus, T. Szyperski, *Nat. Struct. Biol.* **2000**, 7, 982.
- [3] a) K. Wüthrich, *Angew. Chem.* **2003**, 115, 3462; *Angew. Chem. Int. Ed.* **2003**, 42, 3340.
- [4] Y. S. Jung, M. Zweckstetter, *J. Biomol. NMR*, im Druck.
- [5] a) G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* **1999**, 13, 289; b) F. Delaglio, G. Kontaxis, A. Bax, *J. Am. Chem. Soc.* **2000**, 122, 2142.
- [6] a) C. A. Rohl, D. Baker, *J. Am. Chem. Soc.* **2002**, 124, 2723; b) K. T. Simons, C. Kooperberg, E. Huang, D. Baker, *J. Mol. Biol.* **1997**, 268, 209.
- [7] J. F. Doreleijers, S. Mading, D. Maziuk, K. Sojourner, L. Yin, J. Zhu, J. L. Markley, E. L. Ulrich, *J. Biomol. NMR* **2003**, 26, 139.
- [8] D. E. Zimmerman, C. A. Kulikowski, Y. P. Huang, W. Q. Feng, M. Tashiro, S. Shimotakahara, C. Y. Chien, R. Powers, G. T. Montelione, *J. Mol. Biol.* **1997**, 269, 592.
- [9] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, 122, 3791.
- [10] J. C. Hus, D. Marion, M. Blackledge, *J. Am. Chem. Soc.* **2001**, 123, 1541.
- [11] F. Tian, H. Valafar, J. H. Prestegard, *J. Am. Chem. Soc.* **2001**, 123, 11791.
- [12] a) M. J. Berardi, C. H. Sun, M. Zehr, F. Abildgaard, J. Peng, N. A. Speck, J. H. Bushweller, *Struct. Folding Des.* **1999**, 7, 1247; b) K. T. Simons, C. Strauss, D. Baker, *J. Mol. Biol.* **2001**, 306, 1191.