Biomol NMR Assign (2007) 1:191–194
DOI 10.1007/s12104-007-9052-x

ARTICLE

NMR assignment of the domain 513–651 from the SARS-CoV nonstructural protein nsp3

Amarnath Chatterjee · Margaret A. Johnson · Pedro Serrano · Bill Pedrini · Kurt Wüthrich

Received: 15 September 2007 / Accepted: 21 September 2007 / Published online: 30 October 2007

Abstract Sequence-specific NMR assignments of an internal domain of the protein nsp3, nsp3(513–651), which is a part of the SARS coronavirus (SARS-CoV) replicase polyprotein, have been determined, using triple-resonance NMR experiments with the uniformly [^{13}C,^{15}N]-labeled protein. The complete assignments (>99%) provide the basis for the ongoing three-dimensional structure determination.

Keywords SARS-CoV · Nonstructural protein · nsp3 · nsp3c · SARS-unique domain

Biological context

The SARS coronavirus (SARS-CoV) is the etiological agent of the severe acute respiratory syndrome (SARS), a type of pneumonia with a high mortality rate (Peiris et al. 2003). To date there is no vaccine or effective treatment for SARS, and intense ongoing work is focused on the structural and functional characterization of the proteins encoded by the SARS-CoV genome (Bartlam et al. 2005; Chen et al. 2006). The viral RNA genome is translated into polyproteins, which are long polypeptide chains that are subsequently processed by virus-encoded proteases to produce a manifold of functional proteins important for the life-cycle of the virus (Prentice et al. 2004; Schiller et al. 1998). One of these is the nonstructural protein nsp3, which is thought to be involved in RNA replication and processing, and which has been predicted to consist of multiple domains (Thiel et al. 2003). We identified a construct of residues 513–651 that has long-term stability in aqueous solution. It is part of the domain nsp3c, which extends from residues 366–722 and is apparently unique to the SARS-CoV ('SARS-unique domain'; Snijder et al. 2003). As the physiological role of nsp3 is still poorly understood, a complete characterization of each of its domains is essential for deciphering its physiological function. To establish a basis for structural and functional studies by NMR spectroscopy, we report here the sequence-specific resonance assignment of the 139-residue independently folded polypeptide nsp3(513–651).

Methods and experiments

The starting construct nsp3(451–651) was obtained from the cloning and expression pipeline of the FSPS consortium (http://www.visp.scripps.edu/SARS/default.aspx). 1D ^{1}H NMR spectroscopy showed that this 22 kD construct contained a globular domain, but the protein was chemically unstable, being subject to proteolytic degradation after purification from either the soluble or the insoluble fraction of the bacterial cell lysate by Ni^{2+} affinity and size-exclusion chromatography. This proteolytic degradation could not be suppressed by the addition of protease inhibitors, and it persisted also after additional purification steps. We then noticed that proteolysis assays consistently yielded a fragment of about 15.5 kD. Edman degradation of this fragment identified a single

A. Chatterjee · M. A. Johnson · P. Serrano · B. Pedrini · K. Wüthrich (✉)
Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., MB-44, La Jolla CA 92037, USA
e-mail: wuthrich@scripps.edu

K. Wüthrich
Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., MB-44, La Jolla CA 92037, USA

cleavage site between residues 512 and 513. We subsequently cloned and expressed nsp3(513–651) and found it to be chemically stable, while still containing the globular domain.

To prepare the protein for the NMR structure determination, the construct encoding nsp3(513–651) was expressed in the *E. coli* strain BL21(DE3) (Stratagene), using the vector pET-28b, which encodes an N-terminal 6×His tag followed by a thrombin cleavage site. Thrombin cleavage leaves a tag-related N-terminal tetrapeptide segment, GSHM. Cells were grown at 37°C, induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) at an OD_{600} of 0.7, and then grown for another 20 h at 18°C. For the protein purification, the cells were disrupted by sonication in 25 mM phosphate buffer, pH 6.5, with 250 mM NaCl, 2 mM DTT, 5% Triton X-100 and Complete protease inhibitor tablets (Roche). The solution was then centrifuged (18,000*g*, 4°C, 30 min) and the supernatant was applied to a 5 ml His-Trap HP column (Amersham) equilibrated with 25 mM phosphate buffer (pH 6.5, 250 mM NaCl, 2 mM DTT), using a flow rate of 1 ml/min. The bound proteins were eluted with a linear imidazole gradient (5–500 mM) at 8.25 mM/min. The eluted proteins were fractionated and monitored at wavelengths of 280 and 254 nm. Fractions containing the protein nsp3(513–651) were pooled and exchanged with 25 mM phosphate buffer, pH 6.5, containing 150 mM NaCl and 2 mM DTT, to remove imidazole. The histidine tag was then removed from the protein by digestion during 15 min at room temperature with 50 units of thrombin (Enzyme Research Laboratories, USA) per 10 ml of protein solution. After the cleavage, the protein was passed through a size-exclusion Superdex 75 column (Amersham). The fractions containing the protein (as determined by SDS-PAGE) were pooled, exchanged with 25 mM phosphate buffer, pH 6.5, containing 150 mM NaCl and 2 mM NaN_3, and concentrated to a volume of 500 µl using Vivaspin 15 R centrifugal devices (Vivascience) and supplemented with 10% D_2O. Isotope labeling was accomplished by growing cultures in minimal medium containing either 1 g/l of $^{15}NH_4Cl$ as the sole nitrogen source, yielding the uniformly ^{15}N-labeled protein, or 1 g/l of $^{15}NH_4Cl$ and 4 g/l of [$^{13}C_6$]-D-glucose (Cambridge Isotope Laboratories), yielding uniformly [^{13}C,^{15}N]-labeled nsp3(513–651). Growth in M9 minimal medium yielded about 15 mg of pure protein from 1 l of culture. In the 550 µl NMR samples, the protein concentration was adjusted to 1.2 mM, since higher concentrations led to precipitation.

NMR experiments were recorded at 298 K on Bruker Avance 600 and Avance 800 spectrometers equipped with TXI HCN *z*- or *xyz*-gradient probes. The sequence-specific H^N, ^{15}N, C^α and C' backbone assignments were based on the following experiments (Sattler et al. 1999): 2D [^{15}N,^1H]-HSQC, 3D HNCA, 3D HNCO, 3D HNCACB and 3D CBCA(CO)NH. The side chain assignments for the non-aromatic residues were based on 3D ^{15}N-resolved [^1H,^1H]-TOCSY (τ_m = 20 ms), 3D H(C)C(CO)NH-TOCSY, 3D (H)CC(CO)NH-TOCSY, 3D HC(C)H-TOCSY, 3D (H)CCH-COSY, 3D ^{15}N-resolved [^1H,^1H]-NOESY (τ_m = 60 ms), and 3D ^{13}C-resolved [^1H,^1H]-NOESY (τ_m = 60 ms) experiments. The assignment of the aromatic side chain resonances was based on 3D ^{13}C-resolved [^1H,^1H]-NOESY (τ_m = 60 ms) (Muhandiram et al. 1993; Wüthrich 1986) and 2D [^{13}C,^1H]-HSQC experiments. Internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as a chemical shift reference for ^1H, and ^{15}N and ^{13}C shifts were referenced indirectly using the absolute frequency ratios (Wishart et al. 1995).

Assignments and data deposition

The protein nsp3(513–651) has the following amino acid sequence:

```
-4 -1  515  520        530        540        550        560
 GSHMVLPSEAPNAKEEILGTVSWNLREMLAHAEETRKLMPICMDVRAIMATIQRKYKGIK
 1         10         20         30         40         50

  570        580        590        600        610        620
 IQEGIVDYGVRFFFYTSKEPVASIITKLNSLNEPLVTMPIGYVTHGFNLEEAARCMRSLK
 61         70         80         90         100        110

  630        640        650
 APAVVSVSSPDAVTTYNGYLTSS
 121        130        140
```

The residue numbering above the sequence refers to the full-length nsp3 protein, and the tag-related N-terminal tetrapeptide

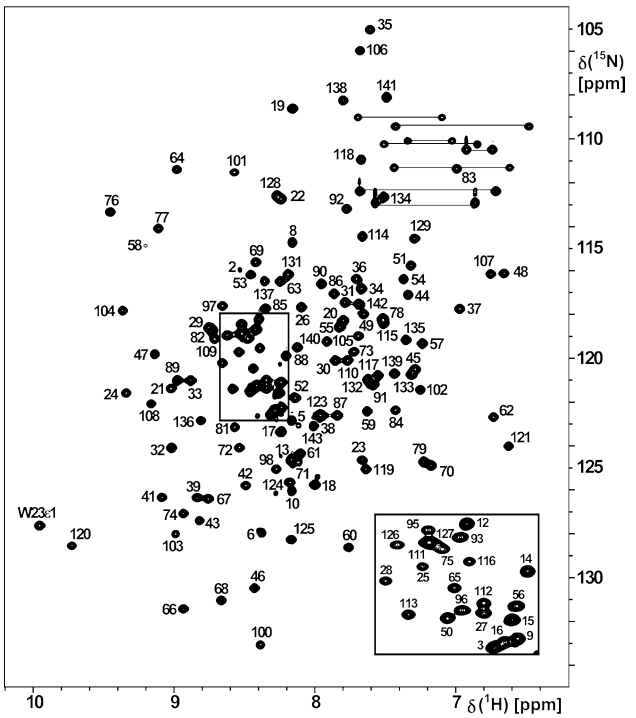

Fig. 1 2D [^{15}N,^{1}H]-HSQC spectrum of uniformly ^{13}C,^{15}N-labeled nsp3(513–651) (protein concentration 1.2 mM, 25 mM sodium phosphate buffer at pH 6.5, 150 mM NaCl, 2 mM NaN_3). The spectrum was recorded at 600 MHz ^{1}H frequency at a temperature of 25°C, with 256 increments in the ^{15}N dimension and 4 scans/increment. The side chain amide resonances of asparagine and glutamine are connected by horizontal lines. The backbone resonances are identified by continuous numbering in the construct used, where the numbers 5–143 identify the residues 513–651 of nsp3 (see text)

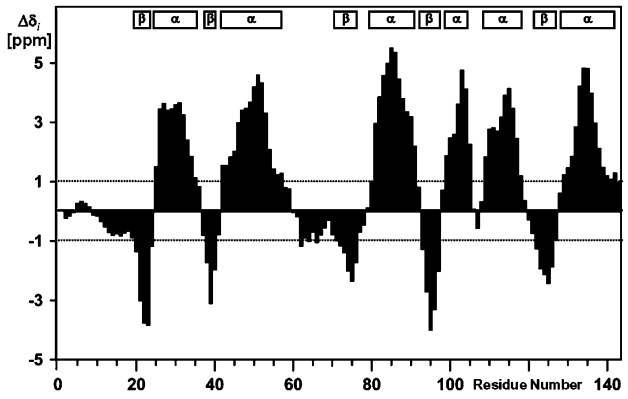

Fig. 2 Chemical shift deviations from the random coil values in the protein nsp3(513–651). Values of $\Delta\delta C^{\alpha}$ and $\Delta\delta C^{\beta}$ were determined within the program package ATNOS/CANDID (Herrmann et al. 2002a, b) by subtracting the random coil shifts from the experimentally determined chemical shifts. The $\Delta\delta_i$ value for a residue i represents a three-point average over the three consecutive residues $i-1$, i and $i+1$, calculated as $\Delta\delta_i = (\Delta\delta C^{\alpha}_{i-1} + \Delta\delta\, C^{\alpha}_{i} + \Delta\delta C^{\alpha}_{i+1} - \Delta\delta C^{\beta}_{i-1} - \Delta\delta C^{\beta}_{i} - \Delta\delta C^{\beta}_{i+1})/3$ (Metzler et al. 1993). A positive value for $\Delta\delta_i$ indicates that residue i is located in a regular helical structure, while a negative value indicates its location in a regular β-strand. Tentative positions for regular secondary structures indicated at the top of the figure, were obtained with the criterion that $|\Delta\delta_i| \geq 1$ for three or more contiguous residues

segment extends from positions –4 to –1. Below the sequence, the residues in the presently studied construct are numbered consecutively from 1 to 143, so that the segment Val 5–Ser 143 corresponds to the segment Val 513–Ser 651 of nsp3 (used in Fig. 1). High-quality NMR data for the protein nsp3(513–651) were obtained, as illustrated by the [^{15}N,^{1}H]-HSQC spectrum shown in Fig. 1. Assignments are complete, with the sole exceptions of the backbone $^{13}C'$ of Leu 514, Ala 518, Met 547, Glu 587, Glu 601, Met 606, Lys 628 and Ser 637 (these residues precede Pro residues), γCH_3 of Val 513 and ζCH of Phe 581 (could not be assigned due to spectral overlap). The side chain amide groups of all eight Asn and Gln were also assigned. The chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 15469.

Identification of regular secondary structures from ^{13}C chemical shifts

Here we use well-established empirical relations between ^{13}C chemical shifts and regular polypeptide secondary structures (Saito 1986; Pastore and Saudek 1990; Spera and Bax 1991; Wishart et al. 1994; Luginbühl et al. 1995) to identify polypeptide segments that form α-helices and β-strands. Specifically we used $\Delta\delta_i = (\Delta\delta C^{\alpha}_{i-1} + \Delta\delta C^{\alpha}_{i} + \Delta\delta C^{\alpha}_{i+1} - \Delta\delta C^{\beta}_{i-1} - \Delta\delta C^{\beta}_{i} - \Delta\delta C^{\beta}_{i+1})/3$, where $\Delta\delta C$ is the deviation of the experimentally observed ^{13}C chemical shift from the corresponding random coil value (Metzler et al. 1993). In a plot of $\Delta\delta_i$ versus the amino acid sequence (Fig. 2), a positive value for $\Delta\delta_i$ indicates that residue i is located in a regular helical structure, while a negative value indicates its location in an extended β-strand. This approach does not enable predictions on the assembly of β-strands into β-sheets.

Acknowledgments We thank Jeremiah Joseph, Vanitha Subramanian, Benjamin W. Neuman, Michael J. Buchmeier, Raymond C. Stevens and Peter Kuhn of the Consortium for Functional and Structural Proteomics of the SARS-CoV for providing us with the nsp3(451-651) and nsp3(466-651) constructs for the initial NMR screening. This study was supported by the NIAID/NIH contract #HHSN266200400058C "Functional and Structural Proteomics of the SARS-CoV" to P. Kuhn and M. J. Buchmeier, and by the Joint Center for Structural Genomics through the NIH/NIGMS grant #U54-GM074898. Additional support was obtained for M. A. J., P. S. and B. P. through fellowships from the Canadian Institutes of Health Research, the Spanish Ministry of Science and Education, and the Swiss National Science Foundation (PA00A-109047/1), respectively, and by the Skaggs Institute for Chemical Biology. Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at TSRI.

References

Bartlam M, Yang H, Rao Z (2005) Structural insights into SARS coronavirus proteins. Curr Opin Struct Biol 15:664–672

Chen S, Luo H, Chen L, Chen J, Shen J, Zhu W, Chen K, Shen X, Jiang H (2006) An overall picture of SARS coronavirus (SARS-CoV) genome-encoded major proteins: structures, functions and drug development. Curr Pharm Des 12:4539–4553

Herrmann T, Güntert P, Wüthrich K (2002a) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE-identification in NOESY spectra using the new software ATNOS. J Biomol NMR 24: 171–189

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1995) Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination. J Magn Reson 109:229–233

Metzler WJ, Constantine KL, Friedrichs MS, Bell AJ, Ernst EG, Lavoie TB, Mueller L (1993) Characterization of the three-dimensional solution structure of human profilin: ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern. Biochemistry 32:13818–13829

Muhandiram DR, Farrow NA, Xu G-Y, Smallcombe SH, Kay LE (1993) A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H_2O. J Magn Reson B 102:317–321

Pastore A, Saudek V (1990) The relationship between chemical shift and secondary structure in proteins. J Magn Reson 90:165–176

Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT, Cheng VC, Chan KH, Tsang DN, Yung RW, Ng TK, Yuen KY, SARS study group (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361:1319–1325

Prentice E, McAuliffe J, Lu X, Subbarao K, Denison MR (2004) Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. J Virol 78:9977–9986

Saito H (1986) Conformation-dependent ^{13}C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR. Magn Reson Chem 24: 835–852

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog NMR Spectrosc 34:93–158

Schiller JJ, Kanjanahaluethai A, Baker SC (1998) Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1A. Virology 242:288–302

Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 331:991–1004

Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance chemical shifts. J Am Chem Soc 113:5490–5492

Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, Weissbrich B, Snijder EJ, Rabenau H, Doerr HW, Gorbalenya AE, Ziebuhr J (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol 84: 2305–2315

Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4:171–180

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140

Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York