NMR assignment of the nonstructural protein nsp3(1066–1181) from SARS-CoV

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

Received: 7 May 2008/Accepted: 1 July 2008/Published online: 22 July 2008 © Springer Science+Business Media B.V. 2008

Abstract  Sequence-specific NMR assignments of the globular core comprising the residues 1066–1181 within the non-structural protein nsp3e from the SARS coronavirus have been obtained using triple-resonance NMR experiments with the uniformly [13C, 15N]-labeled protein. The backbone and side chain assignments are nearly complete, providing the basis for the ongoing NMR structure determination. A preliminary identification of regular secondary structures has been derived from the 13C chemical shifts.

Keywords  Severe acute respiratory syndrome · SARS coronavirus · Nonstructural protein · NMR structure determination

Biological context

The SARS coronavirus (SARS-CoV) is the infectious agent responsible for the severe acute respiratory syndrome (SARS) and represents one of the largest currently known RNA genomes. It is composed of at least 14 functional ORFs, which encode three classes of proteins, i.e., the structural proteins S, M, E, N, 3a, 7a and 7b, the nonstructural proteins nsp1 to nsp16, and the accessory proteins 3b, 6, 8, 9b and 14 (Snijder et al. 2003). Nsp3 is the largest of the nonstructural proteins. It is predicted to have multiple functional domains (Snijder et al. 2003), and large parts of this 1922-residue protein have already been structurally characterized. This includes the functional domains nsp3a, which is a single-stranded RNA-binding protein exhibiting a ubiquitin-like fold (Serrano et al. 2007), nsp3b, which has been described as an ADP ribose-1′-phosphatase (Saikatendu et al. 2005), and nsp3d, which contains two structural domains involved in the proteolytic processing of the polyproteins pp1a and pp1ab, and a third structural domain with a ubiquitin-like fold (Ratia et al. 2006). Further structure determinations of structural domains within the “SARS-unique” functional domain nsp3e are in advanced stages. Despite this progress with the structural and functional characterization, the overall physiological role of the nsp3 protein remains poorly understood and further work is required. Here, we report the sequence-specific NMR assignment of the core structural domain of residues 1066–1181 within the functional domain nsp3e, to establish a basis for a next step of structural and functional characterization of nsp3.

Methods and experiments

Protein preparation

A starting construct, nsp3(1066–1226), obtained from the cloning and expression pipeline of the FSPS consortium (http://visp.scripps.edu/SARS/default.aspx), was found by 1D 1H NMR spectroscopy to contain a globular domain as well as non-globular polypeptide segments. To focus on the globular domain, we cloned the truncated construct nsp3(1066–1181) into the vector pET-25b and expressed
the protein in the *E. coli* strain BL21(DE3)-RIL CodonPlus (Stratagene). Cells were grown at 37°C, induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) at an OD600 of 0.8, and then grown for another 18 h at 18°C. For the protein purification, the cells were disrupted by sonication in 50 mM phosphate buffer, pH 6.5, with 50 mM NaCl, 1% Triton X-100 and Complete protease inhibitor tablets (Roche). The solution was then centrifuged for 30 min at 18,000g and 4°C, and the supernatant was loaded with a flow rate of 1 ml/min onto a 5 ml Hitrap Q SP column (Amersham) equilibrated with 50 mM phosphate buffer at pH 6.5, containing 50 mM NaCl. The bound proteins were eluted with a linear NaCl gradient from 50 to 1,000 mM, fractionated and monitored at wavelengths of 280 and 254 nm. Fractions containing nsp3(1066–1181) were pooled and concentrated to 10 ml using Amicon ultracentrifugal filter devices with 5 kDa cutoff (Millipore). Subsequently, the protein solution was passed through a Superdex 75 size-exclusion column (Amersham). The fractions containing the protein, as determined by SDS-PAGE, were again pooled and concentrated to a volume of 550 μl using 5 kDa cutoff Amicon ultracentrifugal filter devices (Millipore).

Isotope labeling was accomplished by growing cultures in minimal medium containing either 1 g/l of 15NH4Cl as the sole nitrogen source, yielding the uniformly 15N-labeled protein, or 1 g/l of 15NH4Cl and 4 g/l of [13C6]-D-glucose (Cambridge Isotope Laboratories), yielding the uniformly [13C, 15N]-labeled protein. Growth in M9 minimal medium yielded about 10 mg of pure protein from 1 l of culture. This then provided 550 μl of a 1.4 mM NMR sample in 50 mM phosphate buffer at pH 6.5 containing 50 mM NaCl, which was supplemented with 10% D2O and 5.5 μl of a 200 mM NaN3 solution.

NMR spectroscopy

NMR experiments were recorded at 298 K on Bruker Avance 600, DRX 700 and Avance 800 spectrometers equipped with TXI HCN z- or xyz-gradient probes. The sequence-specific HN, 15N, Cα and C0 backbone assignments were carried out with the following experiments (Sattler et al. 1999): 2D [15N, 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 H(CCO)NH-TOCSY, 3D HC(C)H-TOCSY, 3D 15N-resolved [1H, 1H]-NOESY (τm = 60 ms), and 3D 13C-resolved [1H, 1H]-NOESY (τm = 60 ms) experiments. The assignment of the aromatic side chain resonances was based on 3D 13C-resolved [1H, 1H]-NOESY (τm = 60 ms) and 2D [13C, 1H]-HSQC experiments. Internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as a chemical shift reference for 1H, and the 15N and 13C shifts were referenced indirectly using the absolute frequency ratios (Wishart et al. 1995).

NMR assignments and data deposition

The protein nsp3e(1066–1181) has the following amino acid sequence:

```
MYTEQPIDLVPTQPLPNASFDNFKLTCSNTKFADDLNQMTGFTKPASRELSVTFFPDLNGDVVAIDYRHYSASFKKGAKLLHKPIVWHINQATTKTFPKNTWCLRCLWSTKPVDT
```

The numbers above the sequence represent the numbering in the full-length nsp3 protein, and below the sequence the numbering for the construct used in this study is given, which is also used in the Figs. 1 and 2. High-quality NMR data were obtained, as illustrated by the [15N, 1H]-HSQC spectrum shown in Fig. 1. Assignments are complete, except for the backbone 15N and H N of the residues Asn 21, Ser 72, and Lys 94, for which no peaks could be observed in the HSQC spectrum, the backbone 13C of the residues Gln 4, Val 9, Gln 12, Leu 14, Lys 43, Phe 54, Lys 82, Lys 98 and Lys 111, which all precede a Pro residue, the backbone C’ of Asp 20, Ala 70 and Thr 93, which precede the three residues for which the backbone amide resonances could not be observed, and CH of Phe

![Fig. 1 2D [15N, 1H]-HSQC spectrum of uniformly [13C, 15N]-labeled nsp3(1066–1181) (protein concentration 1.4 mM, 50 mM phosphate buffer at pH 6.5, 50 mM NaCl, 2 mM NaN3). The spectrum was recorded at a 1H frequency of 600 MHz, T = 25°C, with 256 increments in the 15N dimension and 4 scans/increment. Resonance assignments are indicated by the sequence numbers in the presently studied construct, which correspond to the residues 1066–1181 of nsp3 (see text). Side chain amide resonances were assigned for eight residues of asparagine and glutamine; these are connected by horizontal lines](image-url)
Deviations from random coil $^{13}$C$^a$ and $^{13}$C$^g$ chemical shifts in the protein nsp3(1066–1181) plotted versus the amino acid sequence. The $\Delta \delta^a$ and $\Delta \delta^g$ values were determined with the program package ATNOS/CANDID (Herrmann et al. 2002a, b) by subtracting the random coil $^{13}$C$^a$ and $^{13}$C$^g$ shifts from the experimentally determined chemical shifts. The $\Delta \delta_i$ value for residue $i$ represents a three-point average value over the three consecutive residues $i - 1, i$ and $i + 1$, calculated as follows: $\Delta \delta_i = 1/3 (\Delta \delta^a_{i-1} + \Delta \delta^a_i + \Delta \delta^a_{i+1} - \Delta \delta^g_{i-1} - \Delta \delta^g_i - \Delta \delta^g_{i+1})$ (Metzler et al. 1993). A positive value for $\Delta \delta_i$ indicates that the residue $i$ is located in a regular helical structure, while a negative value indicates its location in a regular $\beta$-strand. The positions of regular secondary structures indicated at the bottom of the figure were obtained with the criterion that $|\Delta \delta_i| \geq 1$ for three or more sequentially adjacent residues.

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

Regular secondary structures were identified using the well-established empirical relationships with $^{13}$C chemical shifts (Saito 1986; Pastore and Saudek 1990; Spera and Bax 1991; Wishart and Sykes 1994; Luginbühl et al. 1995). The Fig. 2 shows a plot of the deviations of the experimentally observed $^{13}$C chemical shifts from the random coil values, $\Delta \delta_i$, versus the amino acid sequence, where $\Delta \delta_i$ is defined as $1/3 (\Delta \delta^a_{i-1} + \Delta \delta^a_i + \Delta \delta^a_{i+1} - \Delta \delta^g_{i-1} - \Delta \delta^g_i - \Delta \delta^g_{i+1})$ (Metzler et al. 1993). A positive value for $\Delta \delta_i$ indicates that the residue $i$ is located in a regular helical structure, while a negative value indicates that it is located in a regular $\beta$-strand. The tentative positions of regular secondary structures, which are identified at the bottom of the figure, were obtained with the criterion that $|\Delta \delta_i| \geq 1$ for three or more sequentially adjacent residues. No information is as yet available on the assembly of the $\beta$-strands into $\beta$-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 samples of nsp3(1066–1226) for the initial NMR screening. This study was supported by the NIAID/NIH contract #HHSN2662004000058C “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 P. S., M. A. J. and B. P. through fellowships from the Spanish Ministry of Science and Education, the Canadian Institutes of Health Research, and the Swiss National Science Foundation (fellowship 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

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. doi:10.1016/S0022-2836(02)00241-3

Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24:171–189. doi:10.1023/A:1021641154332

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1995) Statistical basis for the use of $^{13}$C chemical shifts in protein structure determination. J Magn Reson 109:229–233. doi:10.1006/jmr.1995.0016

Metzler WJ, Constantine KL, Friedrichs MS, Bell AJ, Ernst EG, Lavoie TB, et al (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. doi:10.1021/bi00213a010
Pastore A, Saudek V (1990) The relationship between chemical shift and secondary structure in proteins. J Magn Reson 90:165–176
Ratia K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC, et al (2006) Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. Proc Natl Acad Sci USA 103:5717–5722. doi: 10.1073/pnas.0510851103
Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, et al (2005) Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1'-phosphate dephosphorylation by a conserved domain of nsp3. Structure 13:1665–1675. doi:10.1016/j.str.2005.07.022
Saito H (1986) Conformation-dependent 13C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state 13C NMR. Magn Reson Chem 24:835–852. doi:10.1002/mrc.1260241002
Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158. doi:10.1016/S0079-6565(98)00025-9
Serrano P, Johnson MA, Almeida MS, Horst R, Herrmann T, Joseph JS, et al (2007) NMR structure of the N-terminal domain of the nonstructural protein 3 from the SARS coronavirus. J Virol 81:12049–12060. doi: 10.1128/JVI.00969-07
Snijder EJ, Bredeenbeck PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, et al (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. doi: 10.1016/S0022-2836(03)00865-9
Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and Cα and Cβ 13C nuclear magnetic resonance chemical shifts. J Am Chem Soc 113:5490–5492. doi:10.1021/ja00014a071
Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:135–140. doi:10.1007/BF00175245
Wishart DS, Sykes BD, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140. doi:10.1007/BF00211777