doi:10.1016/j.jmb.2010.05.027 J. Mol. Biol. (2010) **400**, 724–742

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding

Margaret A. Johnson[1]†, Amarnath Chatterjee[1]†, Benjamin W. Neuman[2,3] and Kurt Wüthrich[1,4,5]*

[1] *Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*

[2] *Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA 92037, USA*

[3] *School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK*

[4] *Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA*

[5] *The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*

Received 19 February 2010; received in revised form 12 May 2010; accepted 13 May 2010
Available online 21 May 2010

Edited by M. F. Summers

Nonstructural protein 3 of the severe acute respiratory syndrome (SARS) coronavirus includes a "SARS-unique domain" (SUD) consisting of three globular domains separated by short linker peptide segments. This work reports NMR structure determinations of the C-terminal domain (SUD-C) and a two-domain construct (SUD-MC) containing the middle domain (SUD-M) and the C-terminal domain, and NMR data on the conformational states of the N-terminal domain (SUD-N) and the SUD-NM two-domain construct. Both SUD-N and SUD-NM are monomeric and globular in solution; in SUD-NM, there is high mobility in the two-residue interdomain linking sequence, with no preferred relative orientation of the two domains. SUD-C adopts a frataxin like fold and has structural similarity to DNA-binding domains of DNA-modifying enzymes. The structures of both SUD-M (previously determined) and SUD-C (from the present study) are maintained in SUD-MC, where the two domains are flexibly linked. Gel-shift experiments showed that both SUD-C and SUD-MC bind to single-stranded RNA and recognize purine bases more strongly than pyrimidine bases, whereby SUD-MC binds to a more restricted set of purine-containing RNA sequences than SUD-M. NMR chemical shift perturbation experiments with observations of ^{15}N-labeled proteins further resulted in delineation of RNA binding sites (i.e., in SUD-M, a positively charged surface area with a pronounced cavity, and in SUD-C, several residues of an anti-parallel β-sheet). Overall, the present data provide evidence for molecular mechanisms involving the concerted actions of SUD-M and SUD-C, which result in specific RNA binding that might be unique to the SUD and, thus, to the SARS coronavirus.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: severe acute respiratory syndrome (SARS); nonstructural protein 3 (nsp3); RNA-binding proteins; macrodomains; frataxins

Introduction

Severe acute respiratory syndrome (SARS) is an atypical pneumonia with flu-like symptoms that can lead to respiratory failure.[1,2] SARS is caused by a coronavirus (CoV), SARS-CoV, which first became evident in 2003. CoVs are enveloped viruses with single-stranded positive-sense 30-kb RNA genomes.[3,4] The SARS-CoV and 'SARS-like' viruses that have since been found in bats were phylogenet-

**Corresponding author.* Department of Molecular Biology, MB-44, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. E-mail address: wuthrich@scripps.edu.

† M.A.J. and A.C. contributed equally to the present work.

Abbreviations used: SARS, severe acute respiratory syndrome; SUD, SARS-unique domain; CoV, coronavirus; nsp, nonstructural protein; NOESY, nuclear Overhauser enhancement spectroscopy; NOE, nuclear Overhauser enhancement; 2D, two-dimensional; HSQC, heteronuclear single quantum coherence; EMSA, electrophoretic mobility shift assay; ssRNA, single-stranded RNA; PDB, Protein Data Bank; TRS, transcription regulatory sequence; GDAP2, ganglioside-induced differentiation-associated protein 2; TEV, tobacco etch virus; 4D, four-dimensional; APSY, automated projection spectroscopy; 5D, five-dimensional; TXI, triple-resonance inverse; 3D, three-dimensional.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.

ically classified as CoV subgroup 2b.[5] They are only distantly related to group 1 and group 2a human CoVs, which cause the common cold and other respiratory illnesses.[3,6] Nonetheless, many genomic features are shared between SARS-CoV and other CoVs, particularly in two-thirds of the genome encoding the nonstructural proteins (nsp) which are needed for genome replication and RNA processing and are thought to function in membrane-associated replicase complexes.[4,7] The structural and accessory proteins encoded by the remainder of the genome vary greatly between different CoVs.

Nsps are initially expressed as two polyproteins, pp1a and pp1ab, which are then cleaved by the action of a main protease (nsp5) and of one or two papain-like proteases (PLpro) found in nsp3 to form mature individual polypeptides.[4,8,9] As one of the products of polyprotein cleavage, nsp3 is a large multidomain polypeptide that is found in all CoVs, with several of the domains being conserved among CoVs.[4,9–11] These include one or two PLpro domains, an 'X' domain that has been shown to form a macrodomain fold and to function as an ADP-ribose-1″-phosphatase and in poly-ADP-ribose binding,[12,13] a 'Y' domain of as yet unknown function, and an N-terminal acidic domain that has been shown to be an RNA-binding protein with a ubiquitin-like fold.[14] SARS-CoV was found to further contain a polypeptide segment in nsp3 that was not found in any other CoVs known at the time, had no apparent sequence homology with any other known protein, and was therefore termed the 'SARS-unique domain' (SUD).[15] Here, we present work that completes the structural coverage of the SUD in solution and provides new insights into its mode of RNA binding.

The SUD was initially annotated as a continuous polypeptide segment of 357 amino acid residues located in sequence positions 366–722 of the SARS-CoV nsp3.[15] In the meantime, atomic resolution structural studies revealed that the SUD actually contains three distinct globular domains, SUD-N (N-terminal region of SUD), SUD-M (middle region of SUD), and SUD-C (C-terminal region of SUD), with residues 387–524, 527–651, and 655–720, respectively, which are connected by short linker peptide segments. NMR structure determination showed that the 'middle domain' SUD-M forms a macrodomain fold, and biochemical experiments and NMR chemical shift mapping resulted in the identification of a putative RNA binding site on the protein surface.[16] The NMR structure of SUD-M was then used in a molecular replacement approach to solve a crystal structure of SUD-NM, which contains the two domains SUD-N and SUD-M in a construct of residues 389–652.[17] The crystal structure was found to be a dimer of this two-domain construct, with SUD-N also forming a macrodomain fold. The linker peptide between SUD-N and SUD-M was not observed in the protein crystals. With biochemical experiments, it was further shown that SUD-NM forms complexes with G-quadruplexes.[17]

This work describes new NMR structure determinations of SUD-C and of a two-domain construct SUD-MC with residues 527–720. Combined with the previous structure determination of SUD-M,[15] these results are used for detailed comparisons of the SUD-M and SUD-C domains in isolated form and in the two-domain construct, and for investigations of the nature of the link between the two domains. We further present NMR data on SUD-N and SUD-NM in constructs comprising residues 387–524 and 387–651, respectively, which supplement the aforementioned crystal structure of SUD-NM with a description of the behavior of this protein in solution. Specifically, biochemical data and NMR experiments define the oligomerization state of SUD-NM and the interactions between the N- and M-domains in SUD-NM. The structure determinations are supplemented with investigations of the RNA binding properties of SUD-M, SUD-C, and SUD-MC based on biochemical data and NMR chemical shift perturbation experiments.

Results

Solution structure of SUD-C

The backbone assignment of SUD-C was essentially complete, with the only unassigned atoms being ^{15}N and ^{1}H^{N} of Ser655, ^{15}N of Pro700, and ^{13}C′ of Ser699. Table 1 summarizes the statistics of the structure calculation, which indicate a high-quality structure determination.

SUD-C adopts a fold consisting of seven β-strands arranged in an anti-parallel β-sheet, and two α-helices located at the N-terminus and C-terminus of the sequence, which are packed against the same side of the β-sheet (Fig. 1a and b). Helix α1 (residues 655–666) is followed by a small anti-parallel β-hairpin of strands β1 and β2 (residues 668–669 and 672–673), a short extended region, and another short strand, β3, of residues 678–679. There are two longer strands, β4 and β5, of residues 682–688 and 691–695, followed by a seven-residue loop, another hairpin of strands β6 and β7 (residues 703–705 and 708–710), and helix α2 (712–719). The seven extended strands form a twisted anti-parallel β-sheet (Fig. 1b), with the topology shown in Fig. 1c, where the two short strands β2 and β3 are both paired in anti-parallel fashion with β4. The fold is classified in the SCOP database[19] as 'N-terminal domain of CyaY-like.'

Solution structure of SUD-MC

The backbone assignment was essentially complete; the only unassigned atoms were ^{13}C$^{\alpha}$ and ^{1}H$^{\alpha}$ of Gly -4; ^{15}N, ^{1}H^{N}, ^{13}C$^{\alpha}$, and ^{1}H$^{\alpha}$ of Ser -3; ^{15}N and ^{1}H^{N} of His -2 (these residues result from the vector-derived N-terminal expression tag); ^{15}N of all proline residues, and ^{13}C′ of the residues preceding prolines. These assignments were used as input for the analysis of the nuclear Overhauser enhancement

Table 1. Input for the structure calculations of the proteins SUD-C and SUD-MC and the statistics of the ensembles of 20 energy-minimized CYANA conformers used to represent the NMR structures

Quantity[a]	SUD-M[b]	SUD-C[b]	SUD-C[c]
NOE upper distance limits	2288	1399	2336
Intraresidual	555	339	362
Short range	586	346	601
Medium range	466	291	530
Long range	681	423	843
Restraints/residue	19	20	35
Long-range restraints/residue	6	6	13
Dihedral angle constraints	673	347	329
Residual target function value (Å^2)	2.12±0.33	1.20±0.24	6.27±0.18
Residual NOE violations			
Number >0.1 Å	27±6	14±7	13±3
Maximum (Å)	0.21±0.17	0.13±0.01	0.13±0.01
Residual dihedral angle violations			
Number >2.5°	0±1	1±1	4±0
Maximum (°)	2.19±0.85	2.06±0.89	28.88±0.31
Amber energies (kcal/mol)			
Total	−4847.45±76.87	−2474.23±83.15	−2530.17±49.79
Van der Waals	−441.68±17.24	−201.07±9.39	−198.23±9.28
Electrostatic	−5379.82±75.29	−2811.92±73.72	−2865.62±46.61
RMSD from ideal geometry			
Bond lengths (Å)	0.0075±0.0002	0.0078±0.0005	0.0078±0.0003
Bond angles (°)	1.834±0.046	2.047±0.139	2.021±0.059
RMSD to the mean coordinates (Å)[d]			
BB	0.55±0.08 (527–648)	0.54±0.08 (655–720)	0.31±0.04 (655–720)
HA	0.95±0.08 (527–648)	0.91±0.06 (655–720)	0.69±0.06 (655–720)
Ramachandran plot statistics (%)[e]			
Most favored regions	81.8	79.7	84.1
Additionally allowed regions	16.9	15.8	10.9
Generously allowed regions	1.0	3.9	5.0
Disallowed regions	0.4	0.6	0.0

[a] The top eight entries describe the input from NMR experiments. The other entries refer to a bundle of 20 CYANA conformers after energy minimization with OPALp. The ranges indicate standard deviations.

[b] SUD-M and SUD-C within the SUD-MC construct. Structure calculations were performed for the intact construct of SUD-MC and also for the two individual domains of residues 527–648 (SUD-M) and residues 655–720 (SUD-C) using the input measured with SUD-MC, since no medium-range or long-range distance constraints between the two domains were observed. In the table, we only list the statistics for calculations with the individual domains in SUD-MC (PDB codes 2KQV and 2KQW for SUD-M and SUD-C, respectively), since these coincide very closely with the result of a calculation for the intact SUD-MC, as presented by Fig. 2a and b, with backbone RMSDs of 0.23 and 0.21 Å for SUD-M and SUD-C, respectively.

[c] Isolated SUD-C (PDB code 2KAF).

[d] BB indicates the backbone atoms N, C$^\alpha$, and C′; HA stands for "all-heavy atoms." The numbers in parentheses indicate the residues for which the RMSD was calculated.

[e] As determined by PROCHECK.[18]

spectroscopy (NOESY) spectra with UNIO-ATNOS/ASCAN[20,21] and UNIO-ATNOS/CANDID,[22] which yielded amino acid side-chain assignments and input for the structure calculation with CYANA[23] (for details, see Materials and Methods). The resulting structure (Fig. 2a and b) was based on 3750 nuclear Overhauser enhancement (NOE) restraints (916 intraresidual, 954 short range, 767 medium range, and 1113 long range). While there were a small number of NOEs between domain hydrogen atoms and hydrogens in the linker peptide or in the short, chain-terminal tails, no NOEs that would connect hydrogen atoms in the two different domains were observed. In Fig. 2a and b, SUD-M and SUD-C are superimposed independently, and it is apparent that there is a large manifold of possible orientations for the domain that was not used for the superposition, showing that the two domains do not adopt a unique orientation relative to each other. In view of this result, we repeated the last cycle of the structure calculation separately for each domain, using NOE restraints within the two individual globular domains (residues 527–648 and 654–720) that had been measured in intact SUD-MC. The results of these two structure calculations (Table 1) show that the two domains are individually well defined in SUD-MC.

To further characterize the dynamics of the SUD-MC two-domain construct, we collected a steady-state ^{15}N{^{1}H} NOE experiment, which is sensitive to the rapid motion of ^{15}N–^{1}H moieties on the picosecond-to-nanosecond timescale (Fig. 2c). The linker residues 651–653 that connect the two domains exhibit ^{15}N{^{1}H} NOE values ranging from 0.2 to 0.4, indicating significant segmental mobility, whereas the mobility of the residues within both globular domains is essentially limited to the overall molecular tumbling of the protein, with ^{15}N{^{1}H} NOE values of 0.7–0.8. Overall, these data support that the range of interdomain orientations observed in Fig. 2a and b is a realistic indication of the SUD-MC conformation in solution.

Additional evidence for a loose, flexible linkage of the two domains in SUD-MC comes from a com-

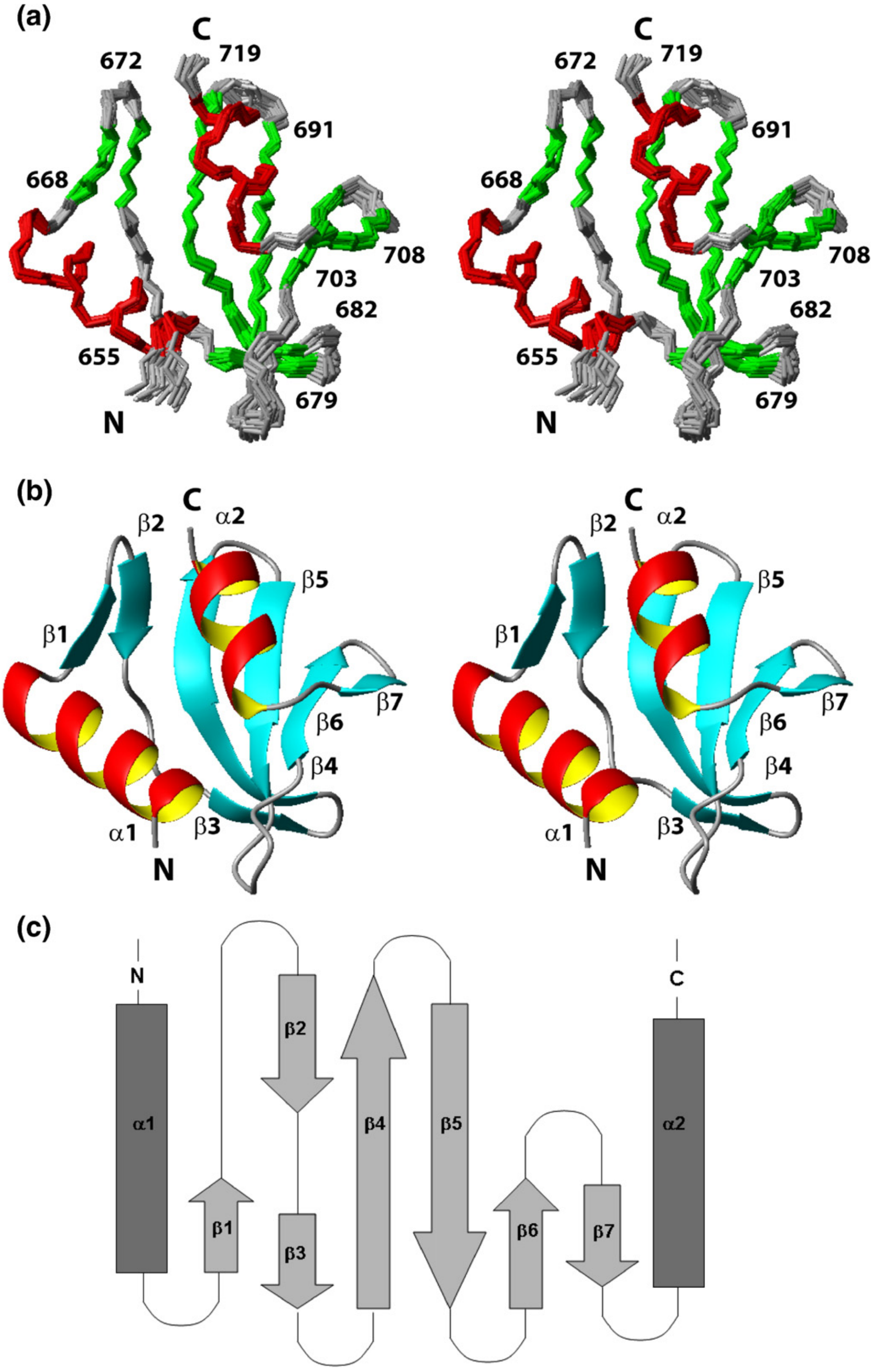

Fig. 1. NMR solution structure of SUD-C. (a) Bundle of 20 energy-minimized NMR conformers representing the solution structure of SUD-C, superimposed for minimal RMSD of the N, C^{α}, and C′ atoms of residues 655–720. Residues in α-helices are shown in red, those in β-strands are shown in green, and regions without regular secondary structure are shown in gray. Selected sequence positions are indicated by numerals. (b) Ribbon presentation of the conformer with the lowest RMSD to the mean coordinates of the ensemble shown in (a). Regular secondary structures are identified. (c) Topology diagram of SUD-C, where α-helices are represented by rectangles and β-strands are represented by arrows. Dark gray, the plane closest to the viewer on which the α-helices lie; light gray, the plane farthest from the viewer on which the β-strands lie.

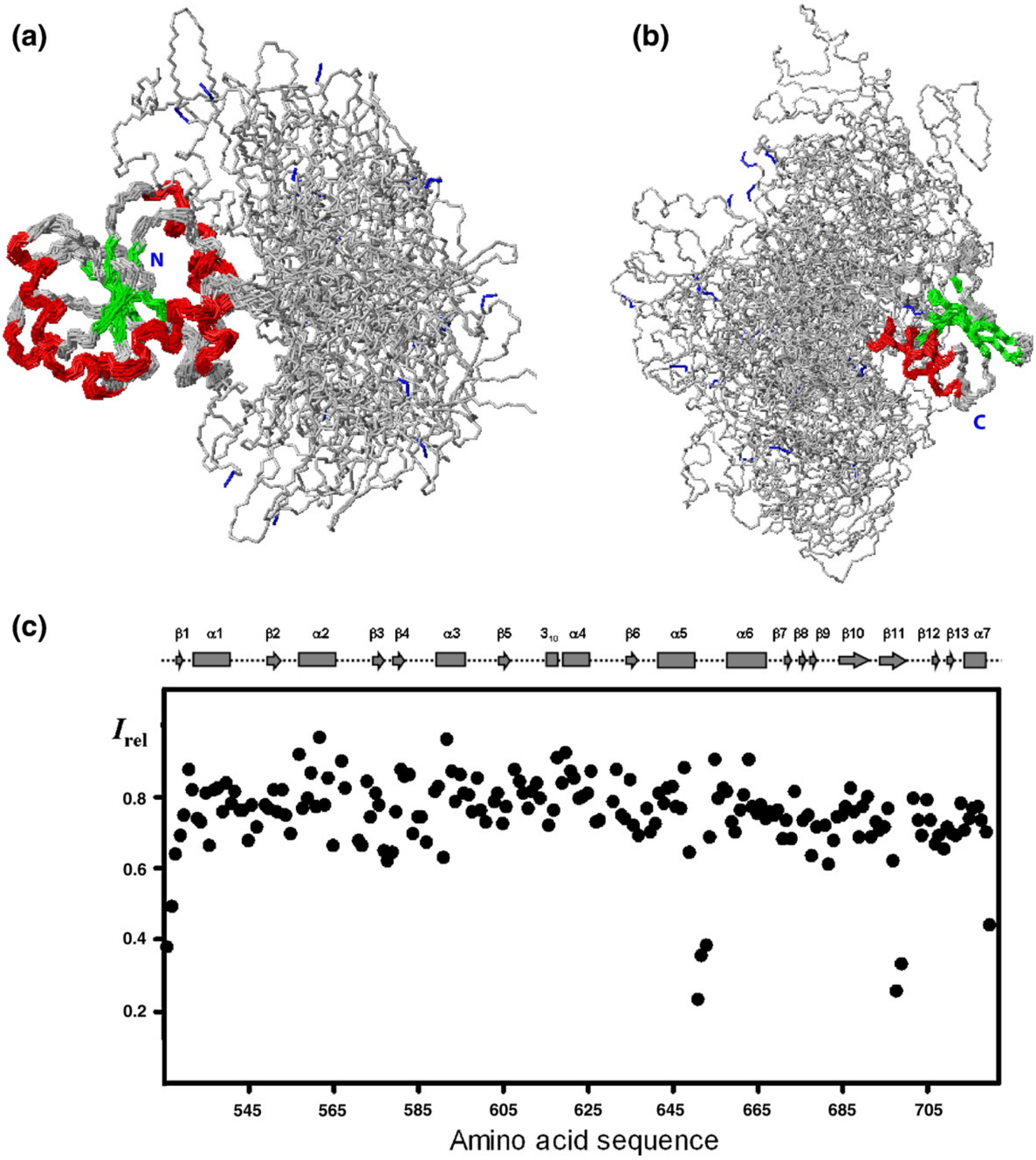

Fig. 2. NMR solution structure of SUD-MC. (a) Bundle of 20 energy-minimized NMR conformers superimposed for minimal RMSD of the backbone N, C^{α}, and C′ atoms of SUD-M (residues 527–648). In SUD-M, residues in α-helices are shown in red, those in β-strands are shown in green, and regions without regular secondary structure are shown in gray. SUD-C is shown in gray. The N-terminus of SUD-M is labeled, and the backbone of the C-terminal residue of SUD-C is shown in blue. (b) The same bundle of conformers as in (a) superimposed for minimal RMSD of the backbone N, C^{α}, and C′ atoms of SUD-C (residues 655–720). In SUD-C, residues in α-helices are shown in red, those in β-strands are shown in green, and regions without regular secondary structure are shown in gray. SUD-M is shown in gray. The C-terminus of SUD-C is labeled, and the backbone of the N-terminal residue of SUD-M is shown in blue. (c) $^{15}N\{^{1}H\}$ NOE values (I_{rel}) plotted *versus* the sequence of SUD-MC. The sequence positions of regular secondary structures are indicated at the top of the panel.

parison with structures determined from data collected in solutions of separately expressed SUD-M and SUD-C. The root-mean-square deviation (RMSD) value between the mean coordinates of the backbone N, C^{α}, and C′ atoms of the SUD-C domain in SUD-MC and in the isolated form is 0.86 Å, and the corresponding value for the SUD-M domain in SUD-MC and in the previously reported structure of the isolated form is 0.97 Å. The structure superpositions in Fig. 3a and b, together with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift data presented in Fig. 3c and d, further show that there are also no outstanding local differences between the SUD-M and the SUD-C polypeptide backbone folds determined either from data collected with individual domains or from data collected with SUD-MC.

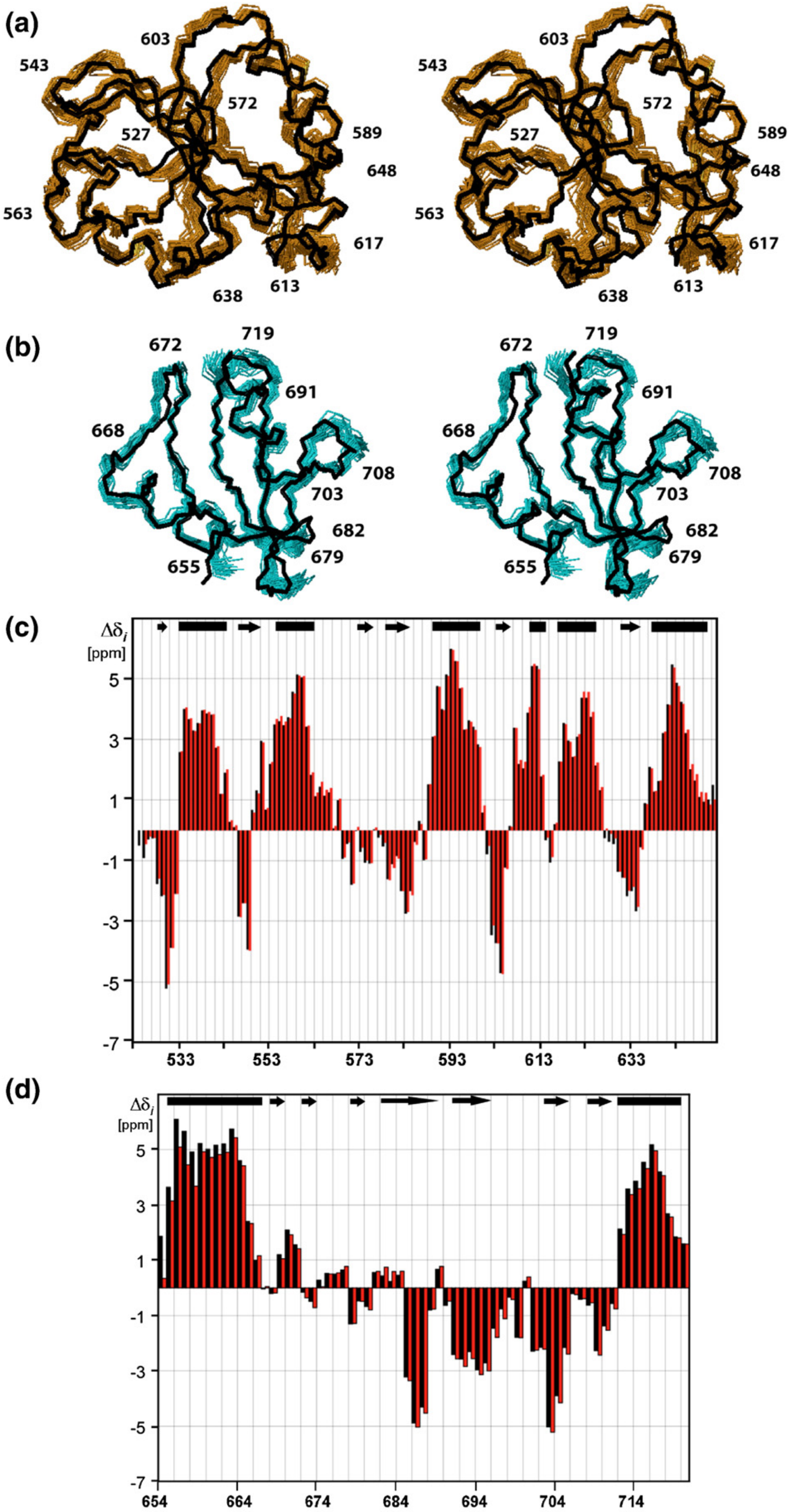

Fig. 3 (*legend on next page*)

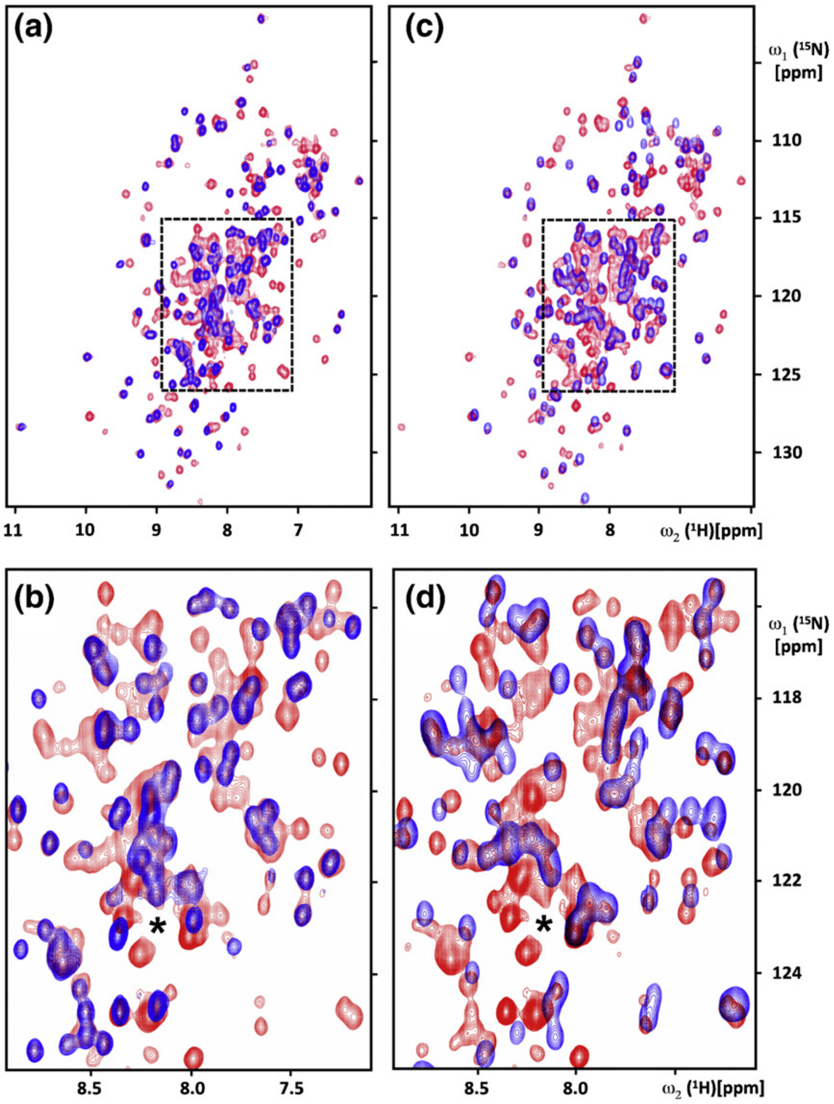

Fig. 4. Comparison of the NMR correlation spectra of SUD-NM, SUD-N, and SUD-M. (a) Overlay of the 2D ^{15}N,^{1}H HSQC spectra of SUD-NM (red) and SUD-N (blue). (b) Expanded presentation of the spectral region indicated by the box in (a). (c) Overlay of the 2D ^{15}N,^{1}H HSQC spectra of SUD-NM (red) and SUD-M (blue). (d) Expanded presentation of the spectral region indicated by the box in (c). In (b) and (d), the asterisk indicates a peak that was, by exclusion, tentatively assigned to the interdomain linker peptide segment between SUD-N and SUD-M in SUD-NM (see the text).

Two-dimensional ^{15}N,^{1}H heteronuclear single quantum coherence spectrum of SUD-N and solution oligomeric states of SUD-N and SUD-NM

The isolated SUD-N formed by the polypeptide of nsp3 residues 387–524 is monomeric in solution, as judged by the two-dimensional (2D) ^{15}N,^{1}H heteronuclear single quantum coherence (HSQC) spectrum (Fig. 4a and b, blue peaks), which contains 130 of the 133 expected backbone ^{15}N–^{1}H correlation peaks and shows line shapes that are typical of a small globular protein. This conclusion is supported by size-exclusion chromatography (Fig. 5a), where SUD-N elutes at an apparent molecular mass of 19 kDa, which is close to the actual molecular mass of 15.2 kDa and would be inconsistent with a 30.4-kDa dimer.

Fig. 3. (a) Stereo view of the bundle of 20 energy-minimized NMR conformers calculated from data collected with the isolated SUD-M (brown),[16] superimposed for minimal RMSD of the N, C$^{\alpha}$, and C′ atoms of residues 527–648. This bundle has been superimposed with the NMR structure of SUD-M calculated from data collected with SUD-MC, which is represented by the conformer that has the minimal RMSD to the mean coordinates of the bundle of 20 conformers in Fig. 2a (black). (b) Stereo view of the bundle of 20 energy-minimized NMR conformers calculated from data collected with the isolated SUD-C (cyan), superimposed for minimal RMSD of the N, C$^{\alpha}$, and C′ atoms of residues 655–720. This bundle has been superimposed with the NMR structure of SUD-C calculated from data collected with SUD-MC, which is represented by a single conformer as described in (a) (black). (c and d) Chemical shift deviations from random-coil values in SUD-M and SUD-C, respectively. Values of $\Delta\delta(^{13}C^{\alpha})$ and $\Delta\delta(^{13}C^{\beta})$ were determined with the program UNIO by subtracting random-coil chemical shifts from experimentally observed chemical shifts. The $\Delta\delta_i$ value for each residue is an average value over three consecutive residues $i-1$, i, and $i+1$, given by $\Delta\delta_i=(\Delta\delta(^{13}C^{\alpha})_{i-1}+\Delta\delta(^{13}C^{\alpha})_i+\Delta\delta(^{13}C^{\alpha})_{i+1}-\Delta\delta(^{13}C^{\beta})_{i-1}-\Delta\delta(^{13}C^{\beta})_i-\Delta\delta(^{13}C^{\beta})_{i+1})/3$.[24] Residues in helices typically have positive $\Delta\delta_i$ values, while those in β-strands have negative values. The $\Delta\delta_i$ values for the isolated domains are plotted in black, where the data for the isolated SUD-M were taken from Chatterjee *et al.*,[16] and those in the intact SUD-MC construct are shown in red. Above the plots, the locations of regular secondary structures, as determined by PROCHECK, are shown by rectangles (helices) and arrows (β-strands).

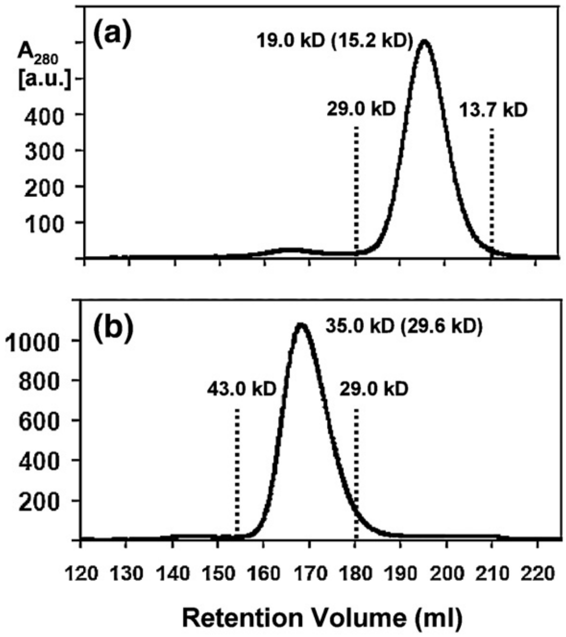

Fig. 5. Size-exclusion chromatograms from a Superdex 75 26/60 column. (a) SUD-N. (b) SUD-NM. The broken lines indicate the elution volumes of protein standards (13.7 kDa, ribonuclease A; 29.0 kDa, carbonic anhydrase; 43.0 kDa, ovalbumin). The numbers near the top of the elution peaks indicate the apparent molecular masses calculated from the observed elution volumes, and the numbers in parentheses indicate the actual molecular masses of the proteins. Protein elution was monitored by the absorbance at 280 nm (A_{280}).

SUD-NM was reported to form a dimer in the single crystals used for X-ray structure determination, with a disulfide bond joining the SUD-N and SUD-M domains within each subunit, in addition to the backbone link between these domains.[17] In contrast, we found that SUD-NM is monomeric in solution. The monomeric state is clearly apparent from the NMR data, as the 2D ^{15}N,^{1}H HSQC spectrum (Fig. 4, red peaks) would be inconsistent with a molecular size of 60 kDa. Independently, size-exclusion chromatography showed that the protein eluted at an apparent molecular mass of 35 kDa, which fits closely with its monomeric size (Fig. 5b). Based on the similarity of the correlation peaks (Fig. 4), we conclude that the globular structures of SUD-N and SUD-M are conserved in the SUD-NM construct. Furthermore, SUD-N and SUD-M are flexibly connected with each other, as evidenced by the close similarity of the peak positions and peak shapes of the isolated domains and the SUD-NM construct (Fig. 4). Since the resonances of SUD-N have been identified as a group, but not individually assigned, the peaks of the ^{15}N–^{1}H moieties of residues 525–526, which form the interdomain linker, had to be assigned by exclusion. There is one ^{15}N–^{1}H correlation peak (identified in Fig. 4b and d) that has no counterpart in the spectra of the individual domains and was therefore tentatively assigned to the linker peptide segment. This is one of about 20 peaks in SUD-NM that exhibit ^{15}N{^{1}H} NOE values smaller than 0.6 (Fig. S1), thus indicating that there is increased mobility in parts of SUD-N and in the linker peptide when compared to the bulk of either of the two globular domains.

RNA binding to SUD-C and SUD-MC

We collected new data on the RNA affinity of SUD-MC, SUD-M, and SUD-C, and then included earlier data on SUD-M[16] in comparative studies. Initially, gel-shift experiments [electrophoretic mobility shift assay (EMSA)] were used to screen a wider selection of RNA ligands, and evidence for binding was then followed up with chemical shift perturbation NMR experiments.

EMSA experiments showed weak binding of SUD-MC to A_{10}, U_{10}, “TRS(+)” (5′-CUAAACGAAC-3′), “TRS(−)” (5′-GUUCGUUUAG-3′), and “GAUA” (5′-CCCGAUACCC-3′) (Fig. 6a and b). Binding was observable at RNA/protein ratios of 1.8:1 or higher, except that binding to C_{10} could not be determined in the experiment of Fig. 6a due to low staining efficiency. There was no observable binding of SUD-C to the same set of RNA ligands (data not shown).

Both SUD-M and SUD-MC bind to a 20-base RNA containing only purine bases, $(GGGA)_5$. The SUD-MC/$(GGGA)_5$ complex is seen as a discrete band in EMSA (Fig. 6b and e). SUD-M does not enter the polyacrylamide gel due to its basic p*I* (calculated p*I*=9.0); therefore, complex formation is inferred by the decrease in free RNA at a 10:1 protein/RNA ratio (Fig. 6c). SUD-MC has a p*I* value that is close to neutral (calculated p*I*=6.7) and does not enter the gel, but its $(GGGA)_5$ complex is sufficiently stable and has enough negative charge to enter the gel and to be observed as a discrete band (Fig. 6b and e). In contrast, the acidic SUD-C (calculated p*I*=5.0) does enter the gel, as shown by the protein staining in Fig. 6h, but we do not observe significant RNA staining at the corresponding position (Fig. 6g) and thus conclude that SUD-C does not bind to $(GGGA)_5$ under the assay conditions used. From Fig. 6e and g, it is also seen that, in contrast to SUD-M (Fig. 6c and previous work[16]), SUD-MC and SUD-C do not bind to $(ACUG)_5$ under these assay conditions. Finally, both SUD-M and SUD-MC bind to a mixture of random 20-base RNA sequences, while there was no evidence of binding to a mixture of random 20-base DNA sequences (Fig. 6c, e, and g).

Binding to purine-containing RNAs was further investigated with six 10-base oligonucleotides containing variable combinations of G and A (Fig. 7). SUD-M was found to bind to all six of these sequences and also to the octamer GGGAGGGA (Fig. 7a, c, and e), showing that neither the number of consecutive guanosines nor their positions in the sequence had significant effects on binding. In contrast, SUD-MC bound only to GGGAGGGAGG (Fig. 7d and f). The different affinities of SUD-MC for GGGAGGGA (low affinity), GGGAGGGAGG (high affinity), and

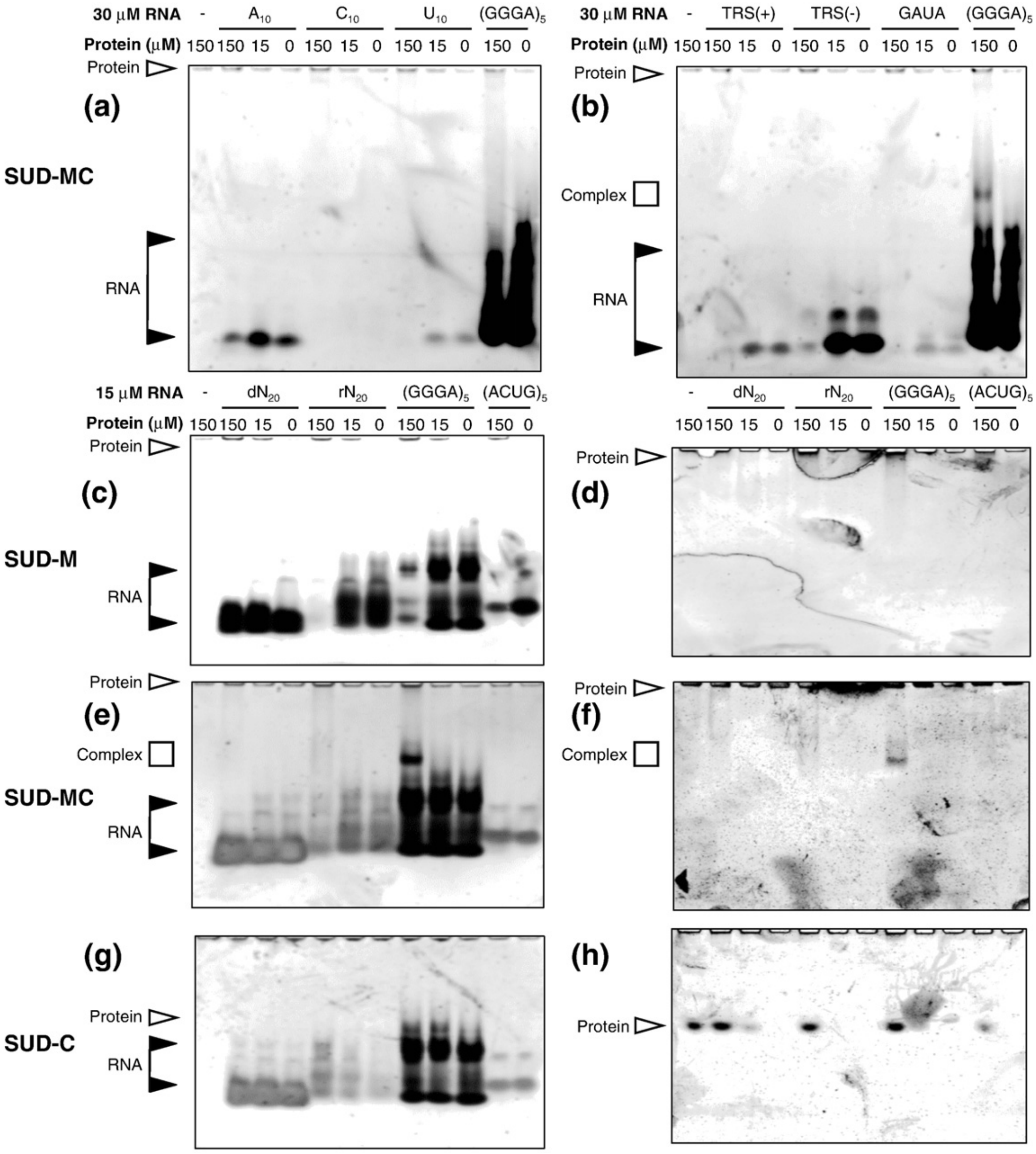
30 μM RNA - A10 C10 U10 (GGGA)5
Protein (μM) 150 150 15 0 150 15 0 150 15 0 150 0
Protein
(a)
SUD-MC
RNA
30 μM RNA - TRS(+) TRS(-) GAUA (GGGA)5
Protein (μM) 150 150 15 0 150 15 0 150 15 0 150 0
Protein
(b)
Complex
RNA
15 μM RNA - dN20 rN20 (GGGA)5 (ACUG)5
Protein (μM) 150 150 15 0 150 15 0 150 15 0 150 0
Protein
(c)
SUD-M
RNA
- dN20 rN20 (GGGA)5 (ACUG)5
150 150 15 0 150 15 0 150 15 0 150 0
Protein
(d)
Protein
(e)
Complex
SUD-MC
RNA
Protein
(f)
Complex
(g)
Protein
SUD-C
RNA
(h)
Protein

Fig. 6. Gel-shift (EMSA) assays probing the interactions of the proteins with ssRNA. (a) SUD-MC with A_{10}, C_{10}, U_{10}, and $(GGGA)_5$. (b) SUD-MC with TRS(+), TRS(−), GAUA, and $(GGGA)_5$ (see the text for the notation used). The protein concentrations are indicated above the gels. RNA (30 μM) is present in all, except for the leftmost lane. (c and d) SUD-M with mixtures of random DNA 20-mers, mixtures of random RNA 20-mers, and the RNA 20-mers $(GGGA)_5$ and $(ACUG)_5$. The same gel is stained for nucleic acid in (c) and for protein in (d). The protein concentrations are indicated above the gels, and the same concentrations apply to (e)–(h). RNA (15 μM) is present in all, except for the leftmost lane. (e and f) SUD-MC with the same nucleic acids as in (c) and (d). (g and h) Same as (e) and (f) for the protein SUD-C. In all panels, RNA bands are indicated by filled triangles, the position of the protein is indicated by open triangles, and the RNA/protein complexes are indicated by open squares. The analysis of these data (see the text) considered that SUD-M does not enter the polyacrylamide gels because of its basic p*I* (calculated p*I*=9.0) even when in complex with RNA. Binding to SUD-M was therefore inferred by the decrease in free RNA in the gel. SUD-MC has a p*I* value that is close to neutral (calculated p*I*=6.7) and does not enter the gel on its own, but the $(GGGA)_5$ complex is sufficiently stable and has enough negative charge to enter the gel and to be observed as a discrete band. SUD-C has a calculated p*I* value of 5.0 and enters the gel also in the absence of RNA (h). The appearance of multiple bands on the native gels for some of the G-rich RNAs is discussed in the text.

$(GGGA)_5$ (low affinity) observed in buffer containing potassium chloride (Fig. 7f) further suggests that the presence of 3′-terminal G residues is an important determinant of SUD-MC binding. Two oligonucleotides, GGGAGGGAGG and $(GGGA)_5$, were tested both in saline buffer (Fig. 6a–f) and in KCl buffer (Fig. 7e and f), whereby the binding of both SUD-M and SUD-MC to $(GGGA)_5$ was found to be weaker in KCl buffer. Binding of G-rich sequences to SUD-M and SUD-MC was also inferred by smearing of the bands

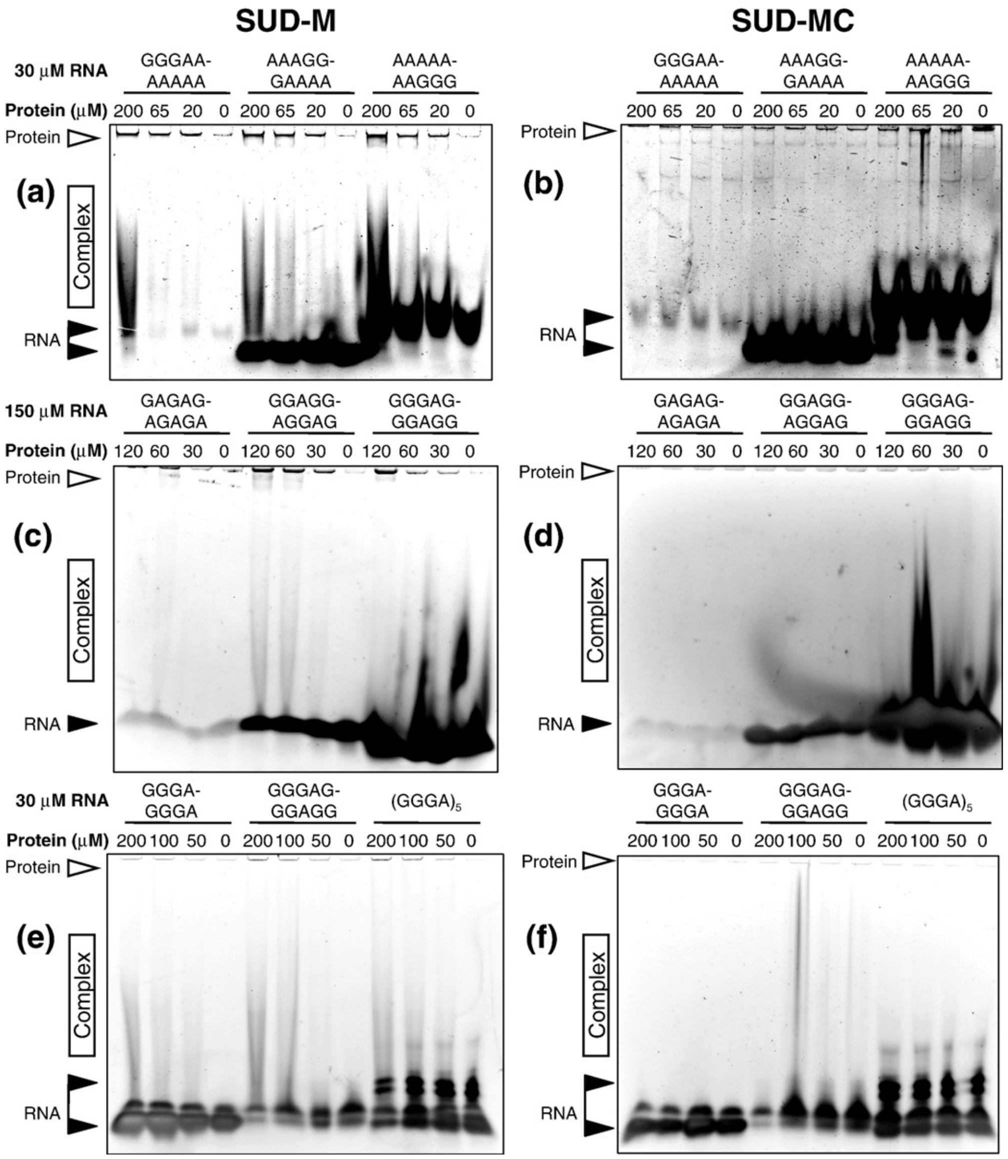

Fig. 7. Gel-shift (EMSA) assays probing the interactions of SUD-M (left) and SUD-MC with 10-base RNAs containing different patterns of G and A. Above the gels, RNA sequences and protein concentrations are indicated. RNA (30 μM) is present in all lanes of (a), (b), (e), and (f), and there is 150 μM RNA in all lanes of (c) and (d). The protein–nucleic acid mixtures in (e) and (f) were incubated in KCl buffer (for details, see Materials and Methods). In all panels, RNA bands are indicated by filled triangles, the position of the protein is indicated by open triangles, and the RNA/protein complexes are indicated by open rectangles labeled 'Complex.' The appearance of multiple bands on the native gels for some of the G-rich RNAs is discussed in the text, as is the smearing of some of the bands.

(Fig. 7), indicating that there was a range of RNA electrophoretic mobilities rather than a single complex with a unique mobility; we concluded that this probably manifests reversible dissociation of the complexes during electrophoresis.

NMR spectra of SUD-MC in the presence and in the absence of the single-stranded RNA (ssRNA) A_{10} (Fig. 8a and b) revealed highly specific changes in a small number of peaks, whereas the positions of the other peaks were unchanged. The magnitudes of the chemical shift changes ($\Delta\delta$) for each residue are plotted in Fig. 8c. The molecular surface area of SUD-M that is affected by A_{10} binding to SUD-MC was found to localize in a positively charged surface cavity (Fig. 9b). In contrast, the molecular surface of SUD-C that is affected by A_{10} binding to SUD-MC (Fig. 8a–c) shows an excess of negative charge (Fig. 9c and d).

NMR chemical shift perturbation measurements revealed that the isolated SUD-C also binds to A_{10} and, additionally, to G_{10} (Fig. 8d). Both homooligonucleotides caused similar patterns of chemical shift perturbations, as described in detail in the Discussion. Much smaller perturbations were induced by U_{10} (Fig. 8d). SUD-C thus shows similarity of RNA binding to SUD-M in the sense that stronger binding is observed for RNA sequences containing purine bases. No binding was apparent in EMSAs at RNA/protein ratios between 0.5:1 and 2.25:1 (data not shown), which is likely due to the fact that RNA binding to the

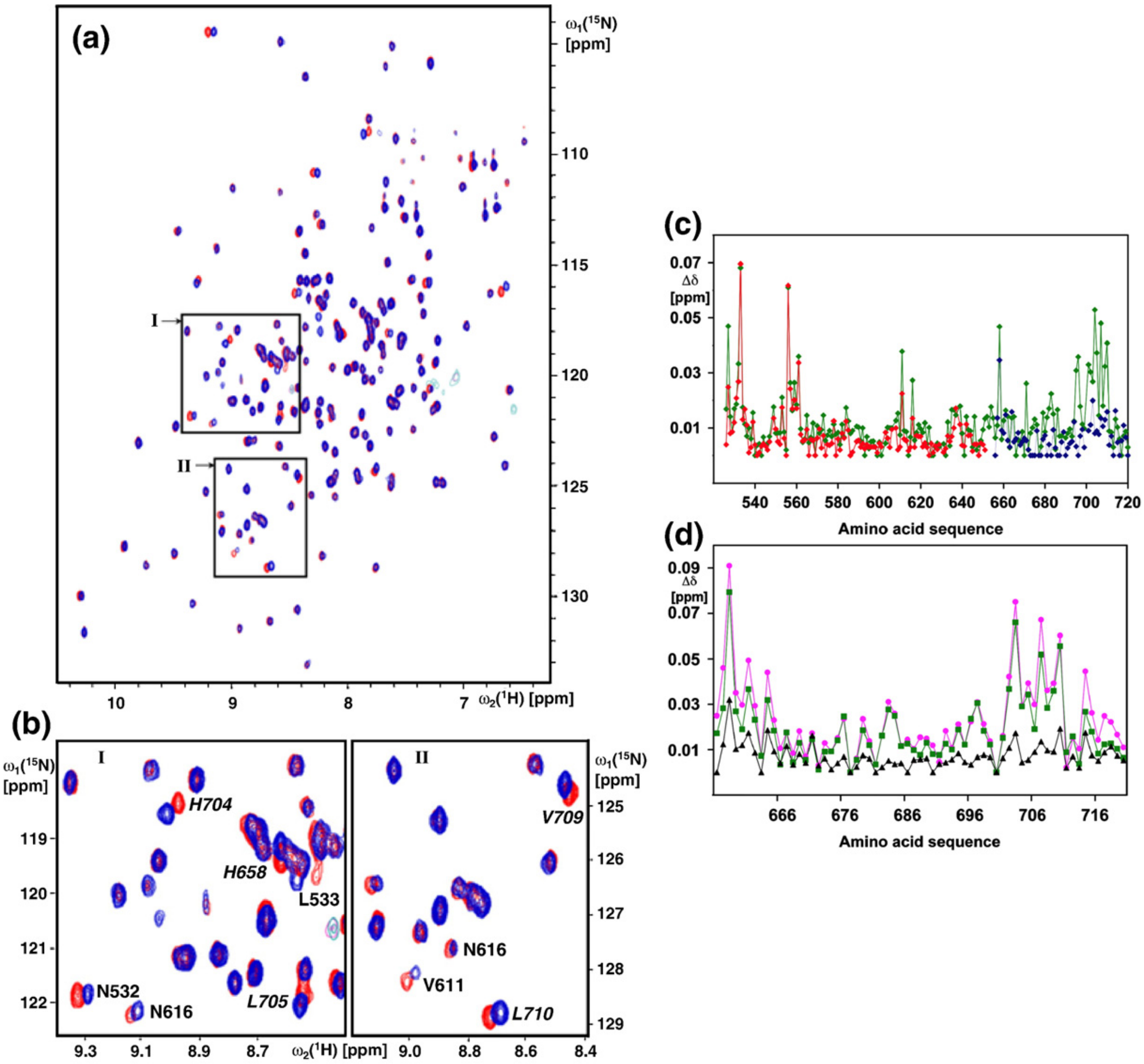

Fig. 8. (a) Overlay of the 2D ^{15}N,^{1}H HSQC spectra of SUD-MC in the presence (blue) and in the absence (red) of A_{10} at an RNA/protein ratio of 1:1. (b) Expansions of regions I and II in the spectra of (a). Residues with chemical shift changes $\Delta\delta \geq 0.03$ ppm are labeled, where the labeling of those belonging to SUD-C is in italics. (c) Plot of chemical shift changes $\Delta\delta$ induced by A_{10} binding to SUD-MC, SUD-M, and SUD-C. Data are plotted *versus* the amino acid sequence of SUD-MC, using the following color code: SUD-MC, green; SUD-M, red; SUD-C, blue. The RNA/protein ratio was 1:1. For the amide group of each residue, $\Delta\delta$ was calculated as $[\Delta\delta(^{1}\mathrm{H})^2 + (\Delta\delta(^{15}\mathrm{N})/5)^2]^{1/2}$. (d) Plot of chemical shift perturbations $\Delta\delta$ induced by RNA binding to the isolated SUD-C. The RNAs used were A_{10} (magenta), G_{10} (green), and U_{10} (black), and the RNA/protein ratio was 5:1.

isolated SUD-C is weaker than RNA binding to SUD-MC; a 5:1 excess of A_{10} over SUD-C was required to induce chemical shift perturbations comparable to those observed for SUD-MC at a 1:1 ratio.

Discussion

SUD-C completes the structural coverage of nsp3(1–1203)

With the SUD-C structure determination, there is now continuous structural coverage for the nsp3 polypeptide segment of residues 1–1203, which comprises 63% of the protein. This includes high-resolution structures of all globular domains and characterization of flexibly disordered polypeptide segments linking these domains (Fig. 10). The data presented in Figs. 2 and 4 provide additional information on the plasticity in solution of two two-domain constructs (see the following section, which extends previous NMR characterizations of flexibly extended interdomain regions of nsp3 (Fig. 10) in constructs where they were linked to a single globular domain.

With the exception of nsp3e, which is an RNA-binding domain and putative nucleic acid chaperone,[11,25] all nsp3 domains in Fig. 10 adopt known protein folds, although some of the domains have very low levels of sequence identity to other proteins. This is also the case for the new domain structure presented in this work, SUD-C, which has a fold that is named after

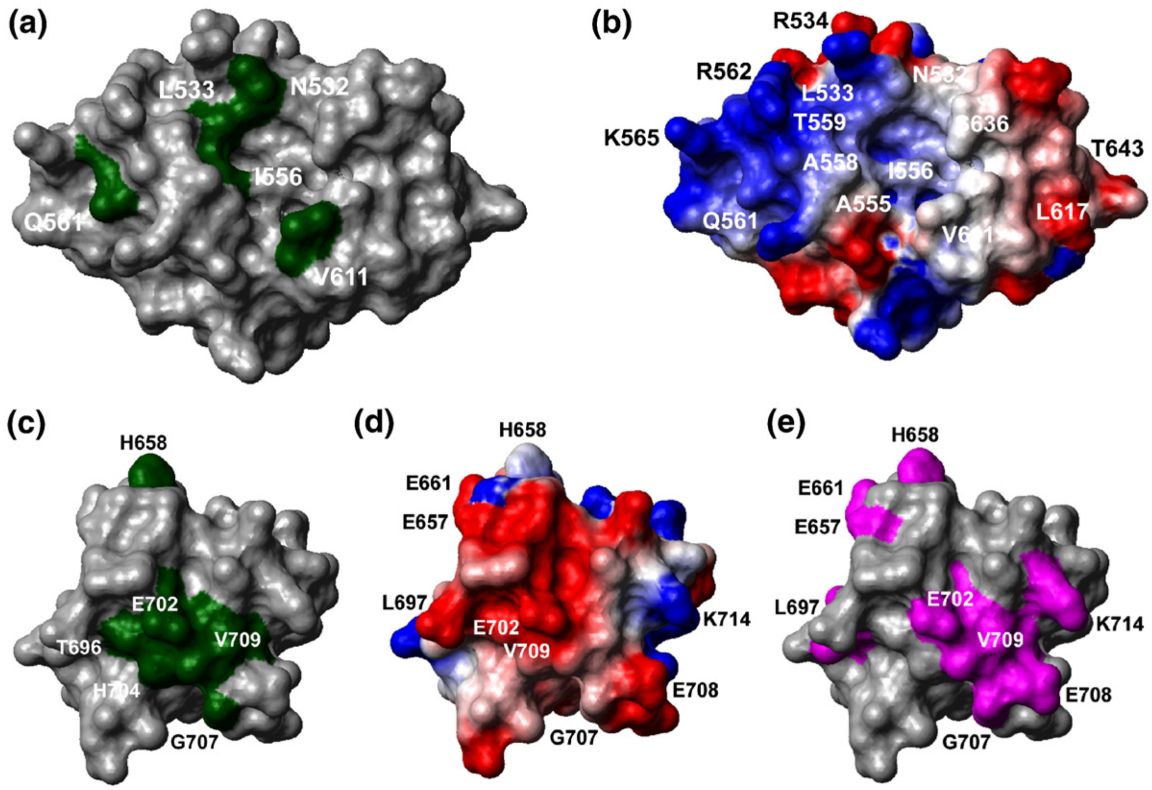

Fig. 9. Visualization of the RNA-binding protein surface areas implicated by the NMR chemical shift perturbation experiments of Fig. 8. (a) Perturbations induced on the surface of SUD-M by the addition of 1 Eq of A_{10} to SUD-MC. The residues with $\Delta\delta \geq 0.03$ (data plotted in green in Fig. 8c) are shown in green and labeled. (b) Electrostatic potential surface of SUD-M. The molecule is shown in the same orientation as in (a). Red, negatively charged areas; blue, positively charged areas; white, neutral areas. Selected surface residues are labeled. (c) Perturbations induced on the surface of SUD-C by the addition of 1 Eq of A_{10} to SUD-MC. The residues with $\Delta\delta \geq 0.03$ (data plotted in green in Fig. 8c) are shown in green and labeled. (d) Electrostatic potential surface of SUD-C. The molecule is shown in the same orientation as in (c), with the same color scheme as in (b). (e) Perturbations induced on the surface of SUD-C by the addition of a 5-fold excess of A_{10} to SUD-C. The residues with $\Delta\delta \geq 0.03$ (data plotted in magenta in Fig. 8d) are shown in magenta and labeled. In (c) to (e), the orientation of SUD-C is related to that of Fig. 1 by a rotation of approximately 90° about the horizontal axis in the projection plane.

the N-terminal domain of CyaY, a member of the frataxin family of metal-binding proteins involved in iron homeostasis.[26,27] Despite the structural similarity, there is only a 10% sequence identity between SUD-C and CyaY, including that only one of the eight residues believed to be involved in iron binding by CyaY is conserved in SUD-C. This makes it unlikely that SUD-C performs the same function. A search of the

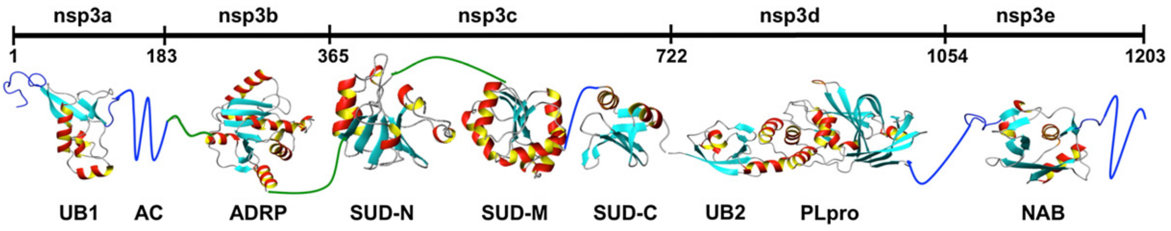

Fig. 10. Globular domains and nonglobular linker peptide segments formed by residues 1–1203 of nsp3. The globular domains are shown as ribbon presentations, flexibly disordered linker segments characterized by NMR spectroscopy are shown as blue lines, and disordered segments implicated by the absence of X-ray diffractions in crystallographic studies are represented by green lines. The black line at the top indicates the initial domain annotation based on bioinformatics and phylogenetic analyses.[15] Abbreviations for the common names of the globular domains are shown below the structures: UB1, first ubiquitin-like domain; AC, acidic domain; ADRP, ADP-ribose-1″-phosphatase; SUD-N, N-terminal region of SUD; SUD-M, middle region of SUD; SUD-C, C-terminal region of SUD; UB2, second ubiquitin-like domain; PLpro, papain-like protease; NAB, nucleic-acid-binding domain.

Protein Data Bank (PDB) with the program Dali[28,29] identified several additional structural homologues with Z-scores of ≤4.4 (Table S1). These include anthrax edema factor,[30] which has adenylate cyclase activity, and AcsD, which is an adenylating enzyme of which SUD-C resembles the N-terminal dimerization domain.[31] Similar to the comparison with CyaY, SUD-C lacks any apparent conservation of the active-site residues of these proteins; therefore, the observation that two different adenylate-binding enzymes have SUD-C-like folds does not appear to provide a new lead toward the as yet unknown SUD-C function. Among the structural homologues of SUD-C (Table S1), there are also small protein domains that perform regulatory functions for DNA-modifying enzymes. Examples are the N-terminal domain of DNA primase, which is a zinc-binding domain that is responsible for DNA binding and also regulates the activity of the catalytic domain of the enzyme,[32] and double-wing motif DNA-binding proteins such as YjbR and MotA.[33,34] Although the key DNA-binding residues are in no case conserved in SUD-C, it is interesting that other proteins with SUD-C-like folds bind nucleic acids, especially when considering that nucleic acids were observed to copurify with recombinant SUD-C expressed in *Escherichia coli* (see Materials and Methods).

In summary, data bank searches for structural homologues of SUD-C show that metal binding, adenylate binding, and nucleic acid binding may be key functions associated with its fold. Due to lack of active-site conservation, however, the present homology search alone did not yield straightforward leads to possible SUD-C functions, although its results are of considerable interest in the context of the experiments in Figs. 6–8 (see the text below).

Solution behavior of the SUD inferred from investigations of SUD-NM and SUD-MC

Previously reported NMR solution and X-ray crystal structures[16,17] and the data reported in this work show that the SUD forms three globular domains. The NMR structure determination of SUD-MC in solution further reveals that the short linker peptide segment of residues 652–654 between SUD-M and SUD-C allows nearly unrestricted freedom of movement of the two domains relative to each other (Fig. 2), with no evidence for a preferred interdomain orientation detected by the methods used. Similarly, solution studies of the construct SUD-NM, for which crystal structure determination has recently been presented,[17] indicate extensive freedom of motion of SUD-N and SUD-M relative to each other. SUD-NM is monomeric in solution (Fig. 5), contrasting with the observation of a dimeric form in the crystal structure.[17] The observed flexible linkage of the two domains further indicates that an interdomain disulfide bond that has been reported to link SUD-N and SUD-M in the crystal structure is not present in the reducing solvent milieu used for the NMR studies. Since the eukaryotic cell cytoplasm, where the SUD is expected to be located,[10,25] also presents a reducing milieu,[35] it would appear that the monomeric sulfhydryl form observed in solution, with flexibly linked SUD-N and SUD-M domains, might be closely related to the physiologically relevant form of the protein. There are no significant chemical shift perturbations that would indicate tight transient or static contacts between the two domains, and the SUD-NM ^{15}N–^{1}H correlation NMR signals have linewidths narrower than one expects for a compact globular domain with a molecular mass of 29.6 kDa. In view of these observations, the following discussion on nucleic acid binding and on the possible physiological roles of the SUD will be based on an SUD molecular model containing three flexibly linked globular domains.

RNA binding properties of the SUD

Previous NMR studies[16] had shown that SUD-M binds A_{10} with higher affinity than U_{10}, and EMSAs further revealed weak binding to A_{15}, $(ACUG)_5$, and the reverse complement of the SARS-CoV transcription regulatory sequence (TRS), TRS(−). No affinity was observable by EMSA either for TRS(+) or for an RNA sequence 'GAUA,' which had previously been found to be recognized by the N-terminal domain of nsp3, nsp3a.[14] The binding studies of Figs. 6–8 now add the information that SUD-C also binds preferentially to purine-containing RNAs and further provide a basis for a comparison of nucleic acid binding by SUD-M and SUD-C in isolated form and in SUD-MC.

As a first step toward gaining insight into the mode of nucleic acid binding by SUD-C, we evaluated the locations of residues that experience chemical shift perturbations upon addition of purine-containing ssRNAs (Fig. 8), in context with electrostatic potential distribution on the molecular surface (Fig. 9d). This shows that the binding surface of SUD-C does not have the characteristics of a typical RNA binding site, since besides two positively charged residues, of which the histidine in position 658 has the largest perturbation (Fig. 8d), there are also several negatively charged amino acid side chains (Fig. 9e). In SUD-MC, the Δδ values of SUD-C residues at a 1:1 A_{10}/protein ratio (Fig. 8c) are significantly increased relative to those of the isolated SUD-C. The presence of SUD-M in SUD-MC thus stabilizes SUD-C–RNA interactions, indicating that the coordinated binding of the linked domains might reduce the repulsion of RNA due to the negative surface charge of SUD-C (Fig. 9d). A putative binding site on SUD-C in SUD-MC is indicated by the fact that the largest Δδ values were seen for H658, H695, T696, E702, G707, and V709 (Fig. 9c). Except for H658 in helix α1, these residues are all in strands β5, β6, and β7 of the β-sheet (Fig. 1), or in loops joining these strands, and the set of residues with large Δδ values is somewhat smaller but nonetheless clearly similar to that seen for RNA binding to the isolated SUD-C (Fig. 9e).

Compared to SUD-MC, SUD-M binds to a larger range of purine-containing RNA sequences, including that it binds to $(ACUG)_5$ and interacts more strongly with a mixture of random RNA 20-mers than does SUD-MC. It thus appears that coordinated action of the two domains in SUD-MC results in greater specificity of RNA binding by SUD-MC than by SUD-M, and may thus lead to binding interactions that are unique to SUD-MC. For example, SUD-MC binds to SARS-CoV TRS(+) (5′-CUAAACGAAC-3′) and to its reverse complement in the anti-genome, TRS(−) (5′-GUUCGUUUAG-3′) (Fig. 6b). NMR chemical shift perturbations (data not shown) were larger in the presence of TRS(+) than in the presence of TRS(−), and localized to the same site of SUD-M in SUD-MC that was previously identified for A_{10} binding (Fig. 9a). However, in contrast to A_{10} binding (Fig. 8a–c), there was minimal contact of SUD-C with TRS. Since the isolated SUD-M did not bind either to TRS(+) or to TRS(−), this shows that the SUD-C in SUD-MC modifies binding without being directly involved in the ligand interaction. Since TRS(+) and TRS(−) contain six and four purine residues, respectively, the aforementioned larger shift perturbations by TRS(+) again seem to reflect the preference of SUD for purine binding. Poly-A is found at the 3′-end of the viral RNA genome, and these interactions could thus indicate a role for SUD-MC in viral genome or subgenomic RNA transcription.

We further used NMR spectroscopy to explore the binding of SUD-M and SUD-MC to other purine RNAs, which yielded results that could only tentatively be interpreted due to line broadening and/or precipitation. Addition of $(GGA)_4$ resulted in broadening of SUD-M resonance linewidths in SUD-MC with increasing RNA/protein ratio. Thus, at a $(GGA)_4$/protein ratio of 0.5:1, most peaks of SUD-M were broadened beyond detection at the contour levels shown in Fig. S2, whereas those of SUD-C were nearly unaffected. For GGGAGGGAGG binding with SUD-MC, a similar pattern of line broadening was observed at a 1:1 RNA/protein ratio, and both $(GGA)_4$ and GGGAGGGAGG caused precipitation when added to SUD-M. Eventually, chemical shift perturbations in SUD-MC and SUD-M were measurable only for $(GGA)_4$ at a 0.2:1 RNA/protein ratio, where the binding of $(GGA)_4$ involves the same surface residues (Fig. S2c and d) as previously observed for A_{10} binding (Fig. 9a). In all of these experiments, SUD-C NMR signals showed neither sizeable chemical shift perturbations nor line broadening (Fig. S2a–c and e). We speculate that this indicates the formation of high-molecular-mass aggregates characterized by binding of SUD-M to RNA, with the NMR signals of the flexibly linked SUD-C only minimally affected by this process. These aggregates would involve multiple SUD-MC proteins binding via their SUD-M to the same RNA. Similar to conclusions from the data in Figs. 6–8, in this interpretation, SUD-C would again appear to regulate the specificity of SUD-MC for RNA binding without direct contacts with the ligand, which may be physiologically important in directing the protein only to certain viral or host cell RNA sequences. Alternatively, chemical exchange between free SUD-MC and a 1:1 complex with RNA in the intermediate rate regime on the chemical shift timescale could also lead to line broadening; however, in this situation, one would expect the line broadening to be limited to binding site residues rather than affecting the entire SUD-M.

Comparison with RNA binding by other macrodomains

ADRP-type macrodomains have been shown to have poly-ADP-ribose binding activity, with so far unknown binding mechanisms.[13] Human macrodomains have also been found to have different specificities for monomeric and poly-ADP-ribose and for RNA. For example, the human macrodomain ganglioside-induced differentiation-associated protein 2 (GDAP2), which is implicated in neuronal differentiation and expressed during brain development, does not bind to poly-ADP-ribose or hydrolyze ADPR-1″-phosphate, but does recognize poly-A RNA.[36] Since we and others have shown that neither SUD-M[16] nor SUD-NM[17] binds to monomeric ADP-ribose, the substrate specificity of SUD appears to be similar to that of GDAP2. No sequence similarity between GDAP2 and SUD-M was detected by BLAST, but a pairwise alignment with FFAS[37] identified about a 15% sequence identity.

While cellular and viral macrodomains occur in combination with a variety of C-terminal domains, SUD-M appears to be a first example of a macrodomain occurring together with a C-terminal domain that has similar RNA binding specificity and regulates the RNA binding behavior of the macrodomain. Since the fifth globular domain of nsp3, nsp3e, which is separated from SUD-MC in the nsp3 sequence by a papain-like protease (Fig. 10), also binds to G-containing and A-containing RNAs,[25] there is also the possibility of a longer-range concerted action with SUD-MC in RNA recognition. Physiologically, this increased specificity is likely to be important in directing nsp3 activity to the desired RNA sequences; binding to a broader range of RNA sequences, as observed for the isolated SUD-M, might affect host cell viability and thus reduce the ability of the virus to replicate.

Implications for possible functions of the SUD

Several lines of evidence[11,14,16,17,25,38] have shown links between nsp3 and RNA, and here we characterize SUD-MC as a purine-binding protein. SUD-M was previously shown to bind to poly-A RNA, which might relate to a possible function in initiating negative strand synthesis by binding to poly-A tails of the viral genome.[16] It has also been proposed that due to its affinity for G-quadruplexes, SUD-NM could bind to G-quadruplex-containing host cell mRNAs encoding proteins involved in cellular signaling pathways, which could lead to disruption of the host cell environment and could favor viral replication.[17] When we tested $(GGGA)_5$ binding to

SUD-MC in KCl buffer, which favors the formation of G-quadruplexes, the binding was much weaker than that in saline buffer (Fig. 7), which would argue against a G-quadruplex being the actual molecular structure recognized by SUD-MC. We note, however, that several of the G-rich RNA sequences displayed multiple bands on native gels (Figs. 6 and 7), which may be due to the presence of different secondary structures, suggesting that variable types of folded RNA structure could indeed be recognized or induced by the interaction with the protein. In the following, we investigate the occurrence of alternative oligopurine sequences in SARS-CoV genomes that could function as additional SUD binding sites.

In the SARS-CoV Tor2 strain, there are three G_6 stretches and three G_5 stretches, but none of these is conserved in all SARS-CoV strains. Of 25 G_4 sequences (on both (+) and (−) strands), eight are conserved in all known SARS-CoV strains; seven of these are in protein-coding regions, and one is in the 3′-untranslated region. Of these eight conserved G_4 sequences, all but one are unique to SARS-CoV, and if these sequences were indeed physiologically relevant substrates for SUD-MC, this would be consistent with the finding that SUD is unique to SARS-CoV. An additional search for the GGGAGGGAGG sequence in the SARS-CoV Tor2 genome did not result in an exact hit, but we found three nucleotide segments that differed only by one base: nt 1461–1470 on the (+) strand in the nsp2 coding region have the sequence 5′-GGGAGGUAGG-3′; nt 12723–12732 on the (+) strand in the nsp12 coding region have the sequence 5′-GGGAGGUAGG-3′; and nt 25383–25392 on the (−) strand have the sequence 5′-GGGAGUGAGG-3′. The two sequences on the (+) strand are highly conserved, with nt 1461–1470 being found in all known SARS-CoV strains and with nt 12723–12732 showing only one point mutation. In contrast, the sequence on the (−) strand is not highly conserved among SARS strains. These three sequences, along with the previously discussed poly-A, TRS(+), and TRS(−) sequences, are all of potential interest as physiological substrates for SUD.

Within nsp3 and possibly in concerted action with other nsp3 domains (see the text above), the SUD may have multiple roles, perhaps interacting with viral RNA to initiate transcription and interfering with mRNAs of the host cell. There is a precedent for a remarkable variety of biological functions carried out by single viral proteins, where nsp1, which is a key virulence factor of the influenza A virus,[39] provides a particularly striking illustration.

Materials and Methods

Preparation of SUD-N, SUD-NM, SUD-MC, and SUD-C

The DNA sequence encoding nsp3 residues 1–723 obtained as a synthetic gene from DNA 2.0 (Menlo Park, CA) was codon optimized for expression in *E. coli*. All four constructs (Fig. S3) were cloned from this starting sequence into pET-28b. SUD-N and SUD-NM consisting of nsp3 residues 387–524 and 387–651, respectively, were cloned with a tobacco etch virus (TEV) protease cleavage site, and SUD-MC consisting of residues 527–720 was cloned with a thrombin cleavage site after an N-terminal 6× His tag. For SUD-C, the highest expression levels were obtained for a construct with a long linker to the 6× His tag, containing both a thrombin and a TEV protease cleavage site.

The four constructs were expressed in *E. coli* strains BL21(DE3)-RIL (SUD-N and SUD-NM), Rosetta(DE3) (SUD-MC), and BL21(DE3) (SUD-C). To produce uniformly ^{15}N-labeled or ^{13}C,^{15}N-labeled protein, we grew cultures in M9 minimal medium containing 1 g/L $^{15}NH_4Cl$ as the sole nitrogen source and 4 g/L of either unlabeled glucose or [$^{13}C_6$]D-glucose as the sole carbon source. For SUD-N, SUD-NM, and SUD-MC, the cell cultures were grown at 37 °C with shaking to an optical density at 600 nm of 0.8. The temperature was then lowered to 18 °C, expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside, and the cultures were grown for a further 18 h. For SUD-C, the cultures were grown for 28 h after induction with 1 mM isopropyl-β-D-thiogalactopyranoside at 18 °C.

The proteins were extracted from frozen cell pellets by sonication in lysis buffer (Table 2). After centrifugation to remove cell debris, the supernatants were applied to a 5-ml HisTrap Crude column (GE) equilibrated with buffer A

Table 2. Buffers used in the purification of the four proteins

	Lysis[a]	Buffer A	Cleavage	Final NMR sample
SUD-N	25 mM NaP_i (pH 8.0), 200 mM NaCl, 2 mM DTT, 1% TX-100	25 mM NaP_i (pH 8.0), 200 mM NaCl, 2 mM DTT, 20 mM imidazole	25 mM NaP_i (pH 8.0), 200 mM NaCl, 20 mM imidazole, 2 mM DTT	25 mM NaP_i (pH 6.6), 150 mM NaCl, 2 mM DTT, 2 mM NaN_3
SUD-NM	25 mM NaP_i (pH 6.6), 200 mM NaCl, 2 mM DTT, 1% TX-100	25 mM NaP_i (pH 6.6), 200 mM NaCl, 2 mM DTT, 20 mM imidazole	25 mM NaP_i (pH 6.6), 200 mM NaCl, 20 mM imidazole, 2 mM DTT	25 mM NaP_i (pH 6.4), 150 mM NaCl, 4 mM DTT, 2 mM NaN_3
SUD-MC	25 mM NaP_i (pH 8.0), 200 mM NaCl, 2 mM DTT, 1% TX-100	25 mM NaP_i (pH 8.0), 200 mM NaCl, 2 mM DTT, 20 mM imidazole	50 mM Tris–HCl (pH 8.0), 200 mM NaCl, 5 mM $CaCl_2$	25 mM NaP_i (pH 6.8), 150 mM NaCl, 2 mM NaN_3
SUD-C	20 mM NaP_i (pH 7.3), 500 mM NaCl, 10 mM imidazole	20 mM NaP_i (pH 7.3), 500 mM NaCl, 10 mM imidazole	20 mM NaP_i (pH 7.0), 50 mM NaCl	20 mM NaP_i (pH 6.5), 20 mM NaCl, 3 mM NaN_3

[a] All lysis buffers also contained Complete EDTA-Free protease inhibitors (Roche).

(Table 2). The bound proteins were eluted with a gradient to 500 mM imidazole, concentrated, and diluted into cleavage buffer. The 6× His tag was removed with agarose-linked thrombin (Thrombin CleanCleave kit; Sigma) or TEV protease using the following different conditions and buffers (Table 2) for the individual proteins.

For SUD-N, tag cleavage was performed overnight at room temperature after addition of 1 ml of 0.69 mg/ml TEV protease solution per 10 ml of protein solution. The solution was applied to a HisTrap column equilibrated with buffer A (Table 2). After elution in the flow-through, the protein was chromatographed on a size-exclusion column (Superdex 75 26/60; GE) equilibrated with 25 mM sodium phosphate buffer (NaP_i; pH 6.6) containing 150 mM NaCl and 2 mM DTT. The yield was about 10 mg of pure protein from 500 ml of culture. The final product contained an N-terminal GHM tripeptide segment from the expression tag. For preparation of the NMR sample, the solution was concentrated to 500 μl using Amicon UltraFree centrifugal devices with a 3-kDa molecular mass cutoff (Millipore), and 50 μl of D_2O and NaN_3 to a concentration of 2 mM were added. The protein concentration was adjusted to 0.8 mM, since higher concentrations led to precipitation.

For SUD-NM, the same procedure was used as for SUD-N, except that the final NMR sample contained 4 mM DTT, and the molecular mass cutoff of the centrifugal device was 10 kDa. The yield was about 20 mg of pure protein from 1 L of culture, and the protein concentration in the NMR sample was adjusted to 1.0 mM.

For SUD-MC, tag cleavage was performed overnight at room temperature with thrombin. For size-exclusion chromatography, we used 25 mM NaP_i (pH 6.8) with 150 mM NaCl. The N-terminal residual tag-related sequence was GSHM. The yield was about 15 mg of pure protein from 1 L of culture, and the protein concentration in the NMR sample was adjusted to 1.0 mM. For SUD-C, the proteins eluting from the first HisTrap column were concentrated 35-fold, then diluted 10-fold with cleavage buffer (Table 2). Four milliliters of TEV protease solution were added, and the reaction was incubated at 37 °C for 2 days. The resulting solution was chromatographed on a HisTrap column, and the protein eluted in the flow-through. SUD-C was then applied to a HiTrap Q FF column in 20 mM NaP_i (pH 7.0) and eluted with a 0–1 M NaCl gradient. The protein did not bind to this column, but was isolated from a nucleic acid fraction with which it had previously copurified. The yield was about 3 mg of purified protein from 1 L of culture, which contained an N-terminal G from the expression tag. A 330-μl NMR sample with 1.2 mM SUD-C in NMR buffer (Table 2) containing 7.5% D_2O was obtained.

NMR structure determination of SUD-C and SUD-MC

The backbone assignment of SUD-C was carried out based on four-dimensional (4D) automated projection spectroscopy (APSY) HACANH, five-dimensional (5D) APSY-CBCACONH, and 5D APSY-HACACONH experiments,[40] which were recorded in a total measurement time of 16.5 h on a Bruker Avance 600 spectrometer with a triple-resonance inverse (TXI) *z*-gradient probe. APSY data were analyzed with the software GAPRO, and GAPRO peak lists were used as input for automated backbone assignment with the program MATCH.[41] The peak lists produced by MATCH were checked and completed interactively. A three-dimensional (3D) ^{15}N-resolved ^{1}H,^{1}H NOESY spectrum and two 3D ^{13}C-resolved ^{1}H,^{1}H NOESY spectra with the carrier frequency centered on the aliphatic and aromatic carbon regions, respectively, were recorded with a mixing time of 150 ms on a Bruker Avance 800 spectrometer with a TXI *z*-gradient probe. Automated side-chain assignment was performed with the program ASCAN,[20] using backbone chemical shifts and the three NOESY data sets as input. During the input preparation, the backbone chemical shifts in the MATCH output were adjusted for optimal fit with the NOESY spectra, with adjustments of 0.01 ppm for $^{1}H^{N}$, 0.02 ppm for all other ^{1}H, 0.09 ppm for ^{15}N, 0.1 ppm for Gly $^{13}C^{\alpha}$, and 0.3 ppm for all other $^{13}C^{\alpha}$ positions and all $^{13}C^{\beta}$ positions.

The chemical shift lists from the ASCAN resonance assignments were used for an initial structure calculation with the stand-alone program suite ATNOS/CANDID 2.2 and the torsion angle molecular dynamics program CYANA 3.0.[21–23] The ASCAN chemical shift lists were then interactively corrected and extended for the final structure calculation. Backbone φ and ψ dihedral angle constraints derived from the $^{13}C^{\alpha}$ chemical shifts were used as supplementary data in the input for the structure calculation.[42,43] The 20 conformers with the lowest residual CYANA target function values obtained from the seventh ATNOS/CANDID/CYANA cycle were energy minimized in a water shell with the program OPALp,[44,45] using the AMBER force field.[46] The program MOLMOL[47] was used to analyze the ensemble of 20 energy-minimized conformers. The stereochemical quality of the models was analyzed using the PDB validation server‡.

SUD-MC structure determination was performed with the same strategy and the same NMR equipment, using the following experimental data: 4D APSY-HACANH, 4D APSY-HNCOCA, 5D APSY-CBCACONH, and 5D APSY-HACACONH experiments[40] recorded in a total of 120 h. A 3D ^{15}N-resolved ^{1}H,^{1}H NOESY spectrum and two 3D ^{13}C-resolved ^{1}H,^{1}H NOESY spectra were recorded with a mixing time of 60 ms. The adjustments of backbone chemical shifts to fit the NOESY data sets were 0.016 ppm for $^{1}H^{N}$, 0.01 ppm for all other ^{1}H, 0.12 ppm for ^{15}N, and 0.31 ppm for all $^{13}C^{\alpha}$ and $^{13}C^{\beta}$. The software versions employed were UNIO 1.0.4 (which includes MATCH, ASCAN, ATNOS, and CANDID) and CYANA 3.0. In the seventh ATNOS/CANDID/CYANA cycle, 40 conformers were generated and subjected to energy minimization in a water shell with OPALp,[44,45] and the 20 best energy-minimized conformers were selected to represent the solution structure.

NMR experiments with SUD-N and SUD-NM

One-dimensional ^{1}H NMR and 2D ^{15}N,^{1}H HSQC spectra of the uniformly ^{15}N-labeled SUD-N (0.8 mM) and SUD-NM (1.0 mM) were recorded on a Bruker DRX 700 spectrometer with a 1.7-mm TXI *z*-gradient probe. ^{15}N{^{1}H} NOE experiments were measured using experiments based on transverse relaxation optimized spectroscopy,[48,49] with a saturation period of 3.0 s and a total interscan delay of 5.0 s. The peaks in the SUD-NM 2D ^{15}N,^{1}H HSQC spectrum arising from SUD-M were identified by comparison with the previously obtained assignments for the single-domain construct SUD-M.[48,49]

The residual peaks in the 2D ^{15}N,^{1}H HSQC spectrum of SUD-NM were assigned as a group to SUD-N. One peak

‡ http://deposit.pdb.org/validate

was tentatively assigned to the interdomain linker based on the lack of any apparent corresponding peak in either the SUD-N spectrum or the SUD-M spectrum.

NMR studies of RNA binding

The interactions of SUD-C and SUD-MC with RNA were evaluated by comparing the 2D ^{15}N,^{1}H HSQC spectra in the presence and in the absence of ssRNA. For the experiments with SUD-C, ssRNA solutions in water were added to an empty tube and frozen, and water was removed by speed vacuum. Ten microliters of 0.68 mM uniformly ^{15}N-labeled protein were added, and the RNA was gently resuspended. The ^{15}N,^{1}H HSQC spectra were recorded on a Bruker DRX 700 spectrometer with a 1-mm TXI *z*-gradient probe. The combined changes in ^{1}H and ^{15}N chemical shifts in the presence versus the absence of the RNA were calculated as $\Delta\delta = [\Delta\delta(^{1}\mathrm{H})^2 + (\Delta\delta(^{15}\mathrm{N})/5)^2]^{1/2}$.

For RNA binding experiments with SUD-M and SUD-MC, a similar procedure was applied, except that for most experiments, the ssRNA solutions in buffer were added directly to the protein solution. Exceptions are the experiments with a 5:1 $(\mathrm{GGA})_4$/SUD-MC ratio and with equimolar mixtures of both proteins and $(\mathrm{GGGA})_5$ or GGGAGGGAGG, where the procedure described for SUD-C was used. Uniformly ^{15}N-labeled SUD-MC at 0.38 mM concentration and uniformly ^{15}N,^{13}C-labeled SUD-M at 0.5 mM concentration, both in 25 mM phosphate buffer (pH 6.8) containing 150 mM NaCl, were used for these experiments. For these proteins, 40-μl samples were measured on a Bruker DRX 700 spectrometer with a 1.7-mm TXI *z*-gradient probe.

Electrophoretic mobility shift assays

Purified SUD-M, SUD-MC, or SUD-C was incubated with ssRNA or single-stranded DNA oligomers either in "saline buffer" containing 150 mM NaCl, 7% glycerol, 4 mM $\mathrm{MgCl_2}$, and 50 mM $\mathrm{NaP_i}$ (pH 6.5), or in "KCl buffer" containing 100 mM KCl, 50 mM NaCl, and 50 mM $\mathrm{NaP_i}$ (pH 6.5). The protein–nucleic acid mixtures were incubated at 37 °C for 60–90 min and analyzed by native electrophoresis on precast 6% acrylamide DNA retardation gels (Invitrogen). Nucleic acid was detected by SYBR-Gold (Invitrogen) staining and photographed using a UV light source equipped with a digital camera. SYBR-Gold was rinsed out, and the protein was subsequently detected by SYPRO-Ruby staining (Invitrogen).

Data bank depositions

The chemical shifts of SUD-C and SUD-MC were deposited in the BioMagResBank§ under accession numbers 16008 and 16613, respectively. The atomic coordinates of the three ensembles of 20 conformers used to represent the SUD-C structure and the structures of the individual domains SUD-M and SUD-C in SUD-MC were deposited in the PDB|| with codes 2KAF, 2KQV, and 2KQW, respectively.

Acknowledgements

This study was supported by National Institute of Allergy and Infectious Diseases/National Institutes of Health contract HHSN266200400058C ("Functional and Structural Proteomics of the SARS-CoV"; Principal Investigator, Dr. P. Kuhn) and the Joint Center for Structural Genomics through National Institutes of Health/National Institute of General Medical Sciences grant U54-GM074898 (Principal Investigator, Dr. I. A. Wilson). Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at The Scripps Research Institute. We thank Dr. Biswaranjan Mohanty and Dr. Bill Pedrini for help with the recording of APSY experiments and MATCH calculations, Dr. Torsten Herrmann for assistance with the use of the various softwares, Dr. Jeremiah S. Joseph for providing the nsp3 construct used for the cloning of the constructs studied in this work, and Dr. Pedro Serrano for helpful discussions.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.027

References

1. Peiris, J. S., Guan, Y. & Yuen, K. Y. (2004). Severe acute respiratory syndrome. *Nat. Med.* **10**, S88–S97.
2. Poon, L. L., Guan, Y., Nicholls, J. M., Yuen, K. Y. & Peiris, J. S. (2004). The aetiology, origins, and diagnosis of severe acute respiratory syndrome. *Lancet Infect. Dis.* **4**, 663–671.
3. Perlman, S. & Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450.
4. Masters, P. S. (2006). The molecular biology of coronaviruses. *Adv. Virus Res.* **66**, 193–292.
5. Gorbalenya, A. E., Snijder, E. J. & Spaan, W. J. (2004). Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.* **78**, 7863–7866.
6. Navas-Martin, S. & Weiss, S. R. (2003). SARS: lessons learned from other coronaviruses. *Viral Immunol.* **16**, 461–474.
7. Weiss, S. R. & Navas-Martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **69**, 635–664.
8. Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M. R. (2004). Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.* **78**, 9977–9986.
9. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37.
10. Oostra, M., Hagemeijer, M. C., van Gent, M., Bekker, C. P., te Lintelo, E. G., Rottier, P. J. & de Haan, C. A. (2008). Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* **82**, 12392–12405.

§ http://www.bmrb.wisc.edu
|| http://www.rcsb.org/pdb

11. Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A. *et al.* (2008). Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* **82**, 5279–5294.
12. Saikatendu, K. S., Joseph, J. S., 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.
13. Egloff, M. P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A. *et al.* (2006). Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *J. Virol.* **80**, 8493–8502.
14. Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S. *et al.* (2007). Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 12049–12060.
15. Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. *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.
16. Chatterjee, A., Johnson, M. A., Serrano, P., Pedrini, B., Joseph, J. S., Neuman, B. W. *et al.* (2009). Nuclear magnetic resonance structure shows that the severe acute respiratory syndrome coronavirus-unique domain contains a macrodomain fold. *J. Virol.* **83**, 1823–1836.
17. Tan, J., Vonrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y. *et al.* (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* **5**, e1000428.
18. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
19. Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G. & Chothia, C. (2000). SCOP: a Structural Classification of Proteins database. *Nucleic Acids Res.* **28**, 257–259.
20. Fiorito, F., Herrmann, T., Damberger, F. F. & Wüthrich, K. (2008). Automated amino acid side-chain NMR assignment of proteins using ^{13}C- and ^{15}N-resolved 3D [^{1}H,^{1}H]-NOESY. *J. Biomol. NMR*, **42**, 23–33.
21. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
22. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). 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.
23. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
24. Metzler, W. J., Constantine, K. L., Friedrichs, M. S., Bell, A. J., Ernst, E. G., Lavoie, T. B. & 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.
25. Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J. *et al.* (2009). Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* **83**, 12998–13008.
26. Cho, S. J., Lee, M. G., Yang, J. K., Lee, J. Y., Song, H. K. & Suh, S. W. (2000). Crystal structure of *Escherichia coli* CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family. *Proc. Natl Acad. Sci. USA*, **97**, 8932–8937.
27. Nair, M., Adinolfi, S., Pastore, C., Kelly, G., Temussi, P. & Pastore, A. (2004). Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites. *Structure*, **12**, 2037–2048.
28. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
29. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
30. Drum, C. L., Yan, S. Z., Bard, J., Shen, Y. Q., Lu, D., Soelaiman, S. *et al.* (2002). Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature*, **415**, 396–402.
31. Schmelz, S., Kadi, N., McMahon, S. A., Song, L., Oves-Costales, D., Oke, M. *et al.* (2009). AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat. Chem. Biol.* **5**, 174–182.
32. Pan, H. & Wigley, D. B. (2000). Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase. *Structure*, **8**, 231–239.
33. Singarapu, K. K., Liu, G., Xiao, R., Bertonati, C., Honig, B., Montelione, G. T. & Szyperski, T. (2007). NMR structure of protein yjbR from *Escherichia coli* reveals 'double-wing' DNA binding motif. *Proteins*, **67**, 501–504.
34. Li, N., Sickmier, E. A., Zhang, R., Joachimiak, A. & White, S. W. (2002). The MotA transcription factor from bacteriophage T4 contains a novel DNA-binding domain: the 'double wing' motif. *Mol. Microbiol.* **43**, 1079–1088.
35. Schäfer, F. Q. & Buettner, G. R. (2001). Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radic. Biol. Med.* **30**, 1191–1212.
36. Neuvonen, M. & Ahola, T. (2009). Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *J. Mol. Biol.* **385**, 212–225.
37. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. (2005). FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res.* **33**, W284–W288.
38. Tan, J., Kusov, Y., Mutschall, D., Tech, S., Nagarajan, K., Hilgenfeld, R. & Schmidt, C. L. (2007). The "SARS-unique domain" (SUD) of SARS coronavirus is an oligo (G)-binding protein. *Biochem. Biophys. Res. Commun.* **364**, 877–882.
39. Hale, B. G., Randall, R. E., Ortin, J. & Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. *J. Gen. Virol.* **89**, 2359–2376.
40. Hiller, S., Fiorito, F., Wüthrich, K. & Wider, G. (2005). Automated projection spectroscopy (APSY). *Proc. Natl Acad. Sci. USA*, **102**, 10876–10881.
41. Volk, J., Herrmann, T. & Wüthrich, K. (2008). Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J. Biomol. NMR*, **41**, 127–138.
42. 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.

43. Luginbühl, P., Szyperski, T. & Wüthrich, K. (1995). Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination. *J. Magn. Reson. Ser. B*, **109**, 229–233.
44. Koradi, R., Billeter, M. & Güntert, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147.
45. Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136–146.
46. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M. *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
47. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
48. Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. (2000). Protein dynamics measurements by TROSY-based NMR experiments. *J. Magn. Reson.* **143**, 423–426.
49. Renner, C., Schleicher, M., Moroder, L. & Holak, T. A. (2002). Practical aspects of the 2D ^{15}N–$\{^{1}H\}$-NOE experiment. *J. Biomol. NMR*, **23**, 23–33.