Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR

Mohammed Kaplan[1], Abhishek Cukkemane[1,4], Gydo C P van Zundert[1], Siddarth Narasimhan[1], Mark Daniëls[1], Deni Mance[1], Gabriel Waksman[2], Alexandre M J J Bonvin[1], Rémi Fronzes[3], Gert E Folkers[1] & Marc Baldus[1]

Studying biomolecules at atomic resolution in their native environment is the ultimate aim of structural biology. We investigated the bacterial type IV secretion system core complex (T4SScc) by cellular dynamic nuclear polarization–based solid-state nuclear magnetic resonance spectroscopy to validate a structural model previously generated by combining *in vitro* and *in silico* data. Our results indicate that T4SScc is well folded in the cellular setting, revealing protein regions that had been elusive when studied *in vitro*.

Super-high-resolution light microscopy[1] and electron tomography[2] have provided unprecedented insights into cellular organization at the nanometer scale. Obtaining information at the atomic level, however, typically requires modeling using *in vitro* data provided by X-ray crystallography, nuclear magnetic resonance (NMR) or electron microscopy. In-cell solution-state NMR can track structure provided that molecular units tumble rapidly[3,4]. Solid-state NMR spectroscopy (ssNMR) has been used to study small proteins embedded in native membranes[5–8]; we have previously studied[9,10] a 150-amino-acid membrane protein using cellular preparations. However, cellular ssNMR experiments on larger molecules or complexes pose additional challenges ranging from sample preparation to data interpretation. Here, we show that the combination of dedicated labeling schemes, cellular ssNMR and dynamic nuclear polarization (DNP)[11] allows us to examine the validity of the structural model of a membrane-associated complex directly in a cellular setting (**Supplementary Fig. 1a**).

We studied the type IV secretion system core complex (T4SScc), a 1-MDa protein machine consisting of 14 copies of 3 proteins (VirB7, VirB9 and VirB10, **Supplementary Fig. 1b**)[12]. It is part of a larger machine (T4SS) that spans the periplasm and is embedded in both the inner and outer membranes of Gram-negative bacteria. The complex transports various substrates, including plasmid DNA, during bacterial conjugation or carries effector proteins into eukaryotic cells[13]. Structural information about T4SScc has been gathered on purified complexes using electron microscopy[12,14], and atomic structures have been obtained[15] for the outer layer (OL) composed of VirB7 and the C-terminal regions of VirB9 and VirB10 (**Supplementary Fig. 1b**). Recent work suggests that the core complex also largely maintains its structure in T4SS constructs comprising T4SScc components and the inner membrane components VirB3, VirB4, VirB5, VirB6 and VirB8, connected to the inner membrane complex via the inner layer (IL) of T4SScc and a flexible region called the stalk[16]. From cysteine scanning studies[17] it became evident that the IL of T4SScc contains N-terminal protein segments of VirB10 that insert into the inner membrane, which is located at a unique position to regulate substrate transfer across the cell envelope. However, atomic information on the IL has remained elusive. In fact, a comparison of the available electron microscopy data suggests that the N-terminal region of VirB10 is compact or unstructured in purified variants of the T4SScc and must adopt a more extended conformation reaching into the inner membrane in larger T4SS complexes and/or when embedded in the cellular envelope[16].

To obtain structural information about T4SScc in its cellular setting, we coexpressed all three subunits in wild-type *Escherichia coli* BL21 (DE3) (WT, **Supplementary Fig. 2a**) as well as in cells deficient in OmpA and OmpF (ref. 9) (BL21 double mutant (dm), **Supplementary Fig. 2a**) following earlier procedures[12]. In both cases, cell envelope fractions contained T4SScc components, and electron microscopy studies after further purification revealed intact protein complexes (**Supplementary Fig. 2b**). Furthermore, previous work[15] as well as our experimental data presented below strongly suggested that the complex is correctly folded in the cellular envelope. For cellular ssNMR studies, we prepared T4SScc either uniformly labeled with ^{13}C and^{15}N or selectively labeled with [^{13}C,^{15}N](Gly,Ser,Leu,Val) or with ([^{13}C,^{15}N]Thr, [^{15}N]Val) in the cell envelope of BL21dm cells (referred henceforth to as U-, GSLV- and TV-labeled T4SScc, respectively). Specific amino acid labeling was confirmed by two-dimensional (2D) ssNMR (**Supplementary Fig. 3**). For spectral evaluation, we used a hybrid approach (**Supplementary Fig. 1a**) with a T4SScc structural model combining previous electron microscopy results and X-ray data available for the OL[14,15]. In particular, the electron microscopy data provided global information about the OL and its putative

[1]NMR Spectroscopy, Utrecht University, Utrecht, the Netherlands. [2]Institute of Structural and Molecular Biology, University College London and Birkbeck, London, UK. [3]Groupe à 5 ans, Biologie structurale de la secretion bacterienne, Unité mixte de recherche Centre National de la Recherche Scientifique–Institut Pasteur 3528, Institut Pasteur, Paris, France. [4]Current address: Microbiology Department, Tuljaram Chaturchand College, Baramati, India. Correspondence should be addressed to G.E.F. (gert@nmr.chem.uu.nl) or M.B. (m.baldus@uu.nl).

RECEIVED 25 AUGUST 2014; ACCEPTED 25 MARCH 2015; PUBLISHED ONLINE 18 MAY 2015; DOI:10.1038/NMETH.3406

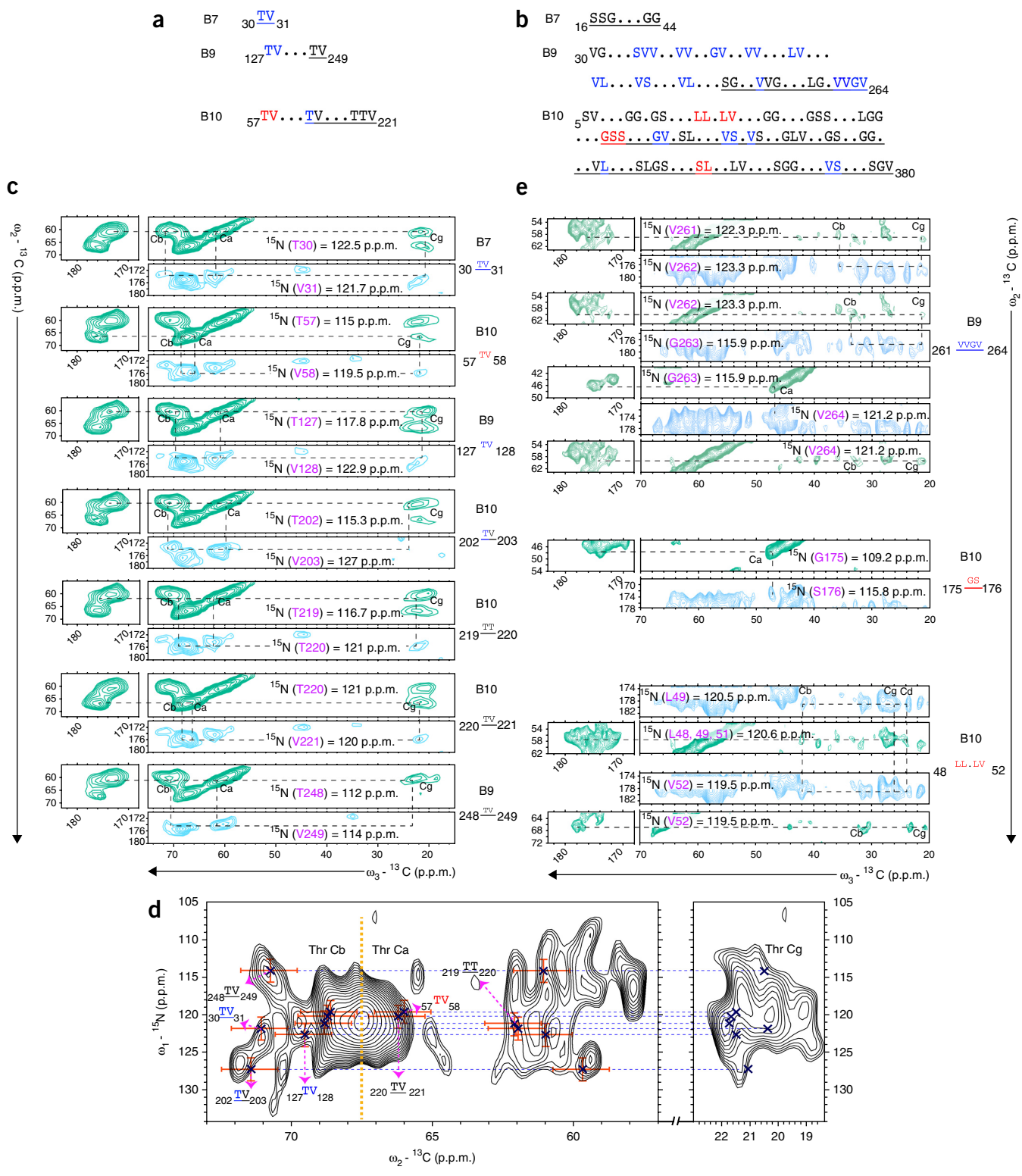

Figure 1 | Analysis of DNP-based two and three-dimensional ssNMR data sets of tailored-labeled cell-embedded T4SScc. (**a**,**b**) Predicted sequential correlations in TV-T4SScc and GSLV-T4SScc, respectively, with letters indicating correlations in random coil (black), β-strand (blue) and α-helix (red). Underlined segments reflect crystallized T4SScc subdomains. Single dots represent 1–4 residues, double dots 5–8 residues and triple dots more than 8 residues. (**c**) Two-dimensional (^{13}C,^{13}C) planes taken out of 3D NCACX (green) and 3D NCOCX (light blue) data sets of TV-T4SScc acquired at 400 MHz DNP. All seven expected correlations were identified in the spectra. (**d**) 2D NCOCX of TV-T4SScc recorded at 800 MHz DNP conditions. Crosses represent the seven correlations expected in this labeling scheme. Additional correlations can be explained by scrambling to glycine. Orange error bars for carbon are 2 p.p.m. and for nitrogen 3 p.p.m., respectively. For clarity, the dashed vertical line separates Cb (resonating above 67.5 p.p.m.) and Ca (resonating below 67.5 p.p.m.) ^{13}C resonance frequencies. Data shown in **e** were recorded under conditions similar to those in **c** on GSLV-labeled T4SScc (see **Supplementary Tables 1** and **2**). In all spectra shown in **c**–**e**, dashed lines highlight carbon backbone as well as side chain correlations.

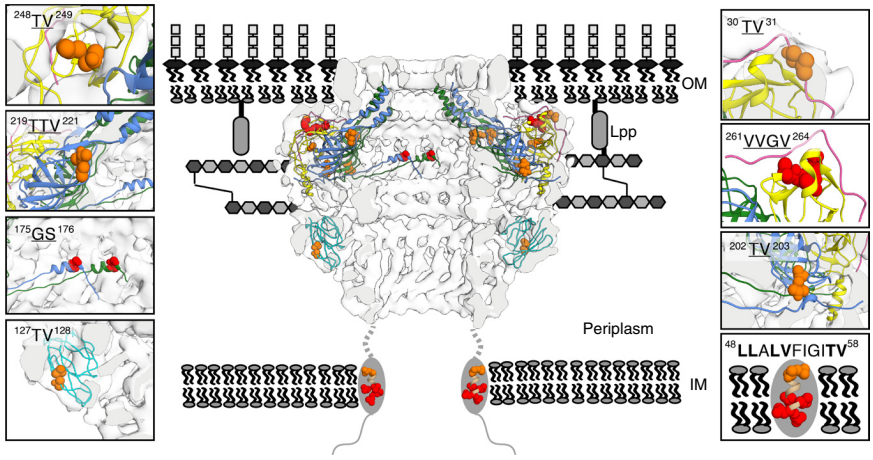

Figure 2 | Summary of residue-specific ssNMR probes and their location in reference to the cellular envelope including the T4SScc electron microscopy map with the outer membrane complex fitted inside (PDB 3JQ0). The N terminus of VirB9 is also docked into the map. Identified residues in our spectra are shown as orange balls for the TV-T4SScc and red balls for the GSLV-T4SScc. Boxes represent zoom-ins of the identified correlations. Note that in the case of the N-terminal helix of VirB10, residues given in bold were tentatively assigned. See ref. 16 for further information on the cellular location of the T4SScc subcomponents.

location relative to the outer membrane. In addition, the X-ray data were used to predict the secondary structure elements for the OL. Finally, we employed[14] structure prediction servers (**Supplementary Figs. 1b** and **4**) for T4SScc regions for which secondary structural information was lacking. Following earlier work[9,10], an SDS-PAGE analysis (**Supplementary Fig. 5**) and a comparison of ssNMR data of cell envelopes without and with protein expression (**Supplementary Fig. 6**) confirmed that T4SScc and Braun's lipoprotein (Lpp) are dominant. Subsequently, we adapted the NMR software package FANDAS[18] to predict the spectra of T4SScc, Lpp and the other dominating cellular envelope compounds[9]—lipopolysaccharides, peptidoglycans and lipids (phosphatidylethanolamine)—that all contribute to the ssNMR spectrum of uniformly ^{13}C,^{15}N-labeled cellular envelope preparations (**Supplementary Fig. 7**).

We exploited the well-established correlation between NMR resonance frequencies and protein secondary structure (see, for example, ref. 19) to obtain structural information about the backbone fold of cell-embedded T4SScc. Our optimized selective amino acid labeling was designed to maximize spectral dispersion under low-temperature DNP conditions and, at the same time, generate unambiguous structural reporters throughout VirB7, VirB9 and VirB10. For example, TV-T4SScc should give rise to seven sequential correlations in β-strand (blue), α-helical (red) and random-coil (rc) regions (black) of the entire 1-MDa complex (**Fig. 1a**). Five of these correlations stem from the previously crystallized OL segments (**Fig. 1a**), and two sequential correlations originate from the structurally elusive regions in VirB9 and VirB10. For GSLV-T4SScc, we expected 55 sequential correlations, with 31 from the OL (**Fig. 1b**) and 20 originating from N-terminal VirB9 and VirB10 (**Fig. 1b**). Lpp sequential correlations do not contribute to our spectral analysis presented in **Figure 1** (**Supplementary Fig. 8**).

We prepared ssNMR samples amenable to DNP studies at 400 and 800 MHz using AMUPol[20] (**Supplementary Fig. 9**). We observed a DNP enhancement factor of ~60, enabling 3D intra-residue (NCACX) and inter-residue (NCOCX) experiments at 400 MHz (**Fig. 1c**; see Online Methods and **Supplementary Fig. 9**). At 800 MHz, the DNP enhancement factor of ~15 allowed for 2D NCACX and NCOCX (**Fig. 1d**) experiments for cross-validation. For TV-T4SScc, a combined evaluation of these data sets confirmed the presence of resonances of α, β and rc conformations (**Fig. 1c**,**d** and **Supplementary Fig. 10**). Using FANDAS and empirically determined standard deviations[21] for protein secondary structure predictions, we identified sequential strips connected via a unique ^{15}N chemical shift along ω_1 in (ω_2,ω_3) intra-residue 2D CA-CX and CO-CX planes in our 3D data sets (**Fig. 1c**), respectively. The chemical shifts were also consistent with a 2D NCOCX experiment at 800 MHz (**Fig. 1d**) and led to tentative assignments for all sequential pairs (**Supplementary Table 1**). The increased spectral resolution (see, for example, **Supplementary Fig. 11** for a comparison of 400- and 800-MHz DNP spectra) revealed several resolved correlations in the Cβ region and allowed us to assess the spectral line width. For correlations such as ^{202}ThrVal^{203} (B10) in the OL or ^{30}ThrVal^{31} located in B7, we determined ^{13}C and ^{15}N resonance line widths of ~1.5 p.p.m. and 3 p.p.m., respectively. Two correlations, relating to α-helical elements predicted for ^{57}ThrVal^{58} and ^{220}ThrVal^{221} in VirB10, exhibited larger line width, especially in the ^{15}N dimension, consistent with variable dynamics within the T4SScc.

Next, we investigated the validity of our structural model by comparing structural predictions to 3D data on GSLV-T4SScc (**Fig. 1e**) as in **Figure 1c**. Again we could identify ω_2-ω_3 planes leading to sequential correlations in the ^{261}ValValGlyVal^{264} VirB9 stretch of the OL forming a β-strand in crystals, in agreement with the observed chemical shifts. NMR data and predictions were also consistent with the detection of the only glycine-serine pair occurring in an α-helix in (^{175}GlySer^{176}) in VirB10. We expected the sequence (^{40}KAFVILMALLALVFIGITV^{58}) in the N-terminal region of VirB10 to adopt an α-helical conformation (**Fig. 1b**). Indeed, we could identify in our 3D data correlations consistent with leucine-leucine (only found in the N-terminal B10) or leucine-valine contacts (**Fig. 1e** and **Supplementary Table 2**), in line with the identification of the unique α-helical ^{57}ThrVal^{58} pair for N-terminal VirB10 (**Fig. 1c**,**d**). ssNMR experiments at higher temperatures provided evidence against a sizable increase in protein motion for the N-terminal half of VirB10 (**Supplementary Fig. 12**) and thus suggested that this N-terminal VirB10 stretch is membrane associated.

Taken together, our results strongly suggest that the 3D fold of the T4SScc OL seen in *in vitro* crystals is maintained in the cellular preparation (**Fig. 2**). At the same time, these findings underlined the validity of our chemical-shift analysis as a means to probe protein structure in a cellular environment. Moreover, our observed correlations in the IL part of T4SScc were consistent with a β-strand arrangement for ^{127}ThrVal^{128} in VirB9, providing the first experimental atomic evidence that the N-terminal

region of VirB9 adopts a distinct conformation at the periplasmic surface of the cell-embedded core complex. Also, we could identify a series of sequential correlations (^{48}LLALVFIGITV^{58}) of the IL suggesting that VirB10 inserts into the inner membrane as an α-helix, allowing the core complex to span the entire cell envelope (**Fig. 2**). Our results also indicate that VirB9 and VirB10 are involved in forming a stable scaffold that allows the assembly of the entire complex in the cellular envelope. At the same time, differences in ssNMR line width seen in our study suggest structural flexibility that may be needed for accommodation of the pilus and substrate secretion[22].

Existing approaches rely on the molecular modeling of *in vitro* data into low-resolution cellular structures. The combination of such approaches with cellular ssNMR data such as shown here allows for direct experimental validation of atomic models in a cellular environment. In the case of T4SScc, such information may help to engineer systems as *in vivo* gene-delivery vectors, with applications in DNA vaccination and therapeutic gene therapy in human cells[23]. Furthermore, our approach may be applicable to studying molecular complexes at the atomic level in their native environment, including those occurring during gene transcription, protein folding or cellular signaling.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: PDB 3JQO, 1EQ7 and 2YPW.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank E. Koers and J. van der Zwan for helpful discussions and technical support. We are indebted to P. Tordo and his group for providing AMUPol. This work was funded in part by the Netherlands Organization for Scientific Research (NWO, grants 700.26.121 and 700.10.443 to M.B.)

AUTHOR CONTRIBUTIONS

G.E.F. and M.B. designed the research. M.K., M.D., A.C. and S.N. produced samples, and M.K. and D.M. performed ssNMR experiments. M.K., G.C.P.v.Z. and A.M.J.J.B. docked atomic models into the electron microscopy density map. R.F. provided clones and conducted electron microscopy studies. G.W. and M.B. wrote the paper and all authors edited it.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Huang, B., Babcock, H. & Zhuang, X. *Cell* **143**, 1047–1058 (2010).
2. Leis, A., Rockel, B., Andrees, L. & Baumeister, W. *Trends Biochem. Sci.* **34**, 60–70 (2009).
3. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. *Methods Enzymol.* **394**, 17–41 (2005).
4. Banci, L., Barbieri, L., Luchinat, E. & Secci, E. *Chem. Biol.* **20**, 747–752 (2013).
5. Etzkorn, M. *et al. Angew. Chem. Int. Edn Engl.* **46**, 459–462 (2007).
6. Fu, R. *et al. J. Am. Chem. Soc.* **133**, 12370–12373 (2011).
7. Jacso, T. *et al. Angew. Chem.* **124**, 447–450 (2012).
8. Miao, Y. *et al. Angew. Chem. Int. Edn Engl.* **51**, 8383–8386 (2012).
9. Renault, M. *et al. Proc. Natl. Acad. Sci. USA* **109**, 4863–4868 (2012).
10. Renault, M. *et al. Angew. Chem. Int. Edn Engl.* **51**, 2998–3001 (2012).
11. Ni, Q.Z. *et al. Acc. Chem. Res.* **46**, 1933–1941 (2013).
12. Fronzes, R. *et al. Science* **323**, 266–268 (2009).
13. Low, H.H. *et al. Nature* **508**, 550–553 (2014).
14. Rivera-Calzada, A. *et al. EMBO J.* **32**, 1195–1204 (2013).
15. Chandran, V. *et al. Nature* **462**, 1011–1015 (2009).
16. Trokter, M., Felisberto-rodrigues, C., Christie, P.J. & Waksman, G. *Curr. Opin. Struct. Biol.* **27**, 16–23 (2014).
17. Jakubowski, S.J. *et al. Mol. Microbiol.* **71**, 779–794 (2009).
18. Gradmann, S. *et al. J. Biomol. NMR* **54**, 377–387 (2012).
19. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. *J. Biomol. NMR* **44**, 213–223 (2009).
20. Sauvée, C. *et al. Angew. Chem. Int. Edn. Engl.* **52**, 10858–10861 (2013).
21. Seidel, K., Etzkorn, M., Schneider, R., Ader, C. & Baldus, M. *Solid State Nucl. Magn. Reson.* **35**, 235–242 (2009).
22. Waksman, G. & Orlova, E.V. *Curr. Opin. Microbiol.* **17**, 24–31 (2014).
23. Llosa, M., Schröder, G. & Dehio, C. *Trends Microbiol.* **20**, 355–359 (2012).

ONLINE METHODS

Sample preparation. BL21 WT and BL21 double mutant[9] cells were transformed with pKM101 following earlier work[12] and plated on LB agar plates enriched with glucose (4 g/L) and containing chloramphenicol (35 mg/L). Next day, a fresh colony of BL21 WT or BL21dm cells was inoculated in a 2 ml LB culture and grown at 37 °C until it reached OD_{600} ~1. Thereafter, the cells were centrifuged and transferred into 25 ml of unlabeled M9 medium. At OD_{600} ~1, cells were centrifuged again and transferred into 250 ml of unlabeled M9 medium. Upon reaching OD_{600} ~1, the cells were centrifuged for 10 min at 2,000*g* and transferred into 250 ml of ^{13}C,^{15}N-labeled M9 medium in the case of uniformly labeled samples and then induced with tetracycline (200 µg/L, see ref. 12). After induction, samples were incubated at 20 °C over night. 5–6 h after induction, another 200 µg/L of tetracycline and 35 mg/L chloramphenicol were added. In case of GSLV-T4SScc samples, 200 mg/L of each labeled amino acid was added to unlabeled M9 medium at OD_{600} ~1, 20 min after induction with tetracycline. In TV-T4SScc sample, in addition to the labeled amino acids, all other amino acids were added (as unlabeled at 200 mg/L) to the medium after induction to decrease scrambling. Cell lysis and cell envelope samples were prepared as described earlier[9]. The T4SScc was further purified as in ref. 12 and visualized on a Tecnai T12 120 kV BioTwin electron microscope after negative staining with 2% uranyl acetate.

For DNP measurements, the cell envelope samples were washed twice with a buffer containing 20 mM AMUPol in 32.5% D_2O, 12.5% H_2O, 10% 100 mM HEPES (pH = 7) and 45% glycerol d8 for GSLV-T4SScc and 22.5% glycerol d8 for TV-T4SScc. 50 µl of this buffer was used in each washing step. Samples were centrifuged after washing (100,000*g*, 25 min).

Solid-state NMR & DNP experiments. For NMR measurements, a standard-bore 700 MHz as well as wide-bore 800 MHz/527 GHz DNP and 400 MHz/263 GHz DNP systems (Bruker Biospin) were used. We filled cellular envelope preparations containing approximately 1 mg of ^{13}C,^{15}N-labeled T4SScc into standard 3.2 mm rotors. For all DNP measurements, samples were cooled down to 100 K in a 3.2 mm sapphire rotor. The DNP enhancement was measured by overlaying HC CP/MAS spectra recorded with and without microwave irradiation. Two- and three-dimensional NC correlation spectra were recorded using SPECIFIC-CP[24] ^{15}N-^{13}C transfers. Homonuclear (^{13}C,^{13}C) transfers were established using PARIS[25] or spin-diffusion[26] blocks. ^{1}H decoupling using SPINAL64 (ref. 27) was employed during evolution and detection periods.

Additional experimental parameters. Note that in all experiments, heteronuclear SPECIFIC transfer[24] was established between t_1 and t_2 and homonuclear (^{13}C,^{13}C) mixing was established between t_2 and t_3. ^{1}H decoupling using SPINAL64 was employed at 87 kHz during evolution and detection periods.

Additional experimental parameters are given below.

The NCACX in **Figure 1c** (400 MHz DNP) was acquired using 14 points in t_2 and 10 points in t_1 with a spectral width of 5,027 Hz and 2,269 Hz in t_2 and t_1 respectively. The spectrum was processed using a squared sine function 3 in t_3 and t_2 and sine bell 2 in t_1, with zero filling 256 in t_3 and 64 in t_2 and t_1, with 4 linear prediction coefficients in t_1 and t_2.

The NCOCX in **Figure 1c** (400 MHz DNP) was acquired using 10 points in t_1 and t_2 with spectral width of 5,027 Hz and 2,269 Hz in t_2 and t_1 respectively. The spectrum was processed using a squared sine function 3 in t_3 and t_2 and sine bell 2 in t_1, with zero filling 256 in t_3 and 64 in t_2 and t_1, with 4 linear prediction coefficients in t_1 and t_2.

The NCOCX in **Figure 1d** (800 MHz DNP) was acquired using 16 t_1 points with a spectral width of 4,055 Hz in t_1. The spectrum was processed using squared sine 3.5 function in both t_1 and t_2 with 2k and 1k zero filling points in t_2 and t_1 respectively.

The NCACX in **Figure 1e** (400 MHz DNP) was acquired using 14 points in t_1 and t_2 with spectral width of 5,027 Hz and 2,269 Hz in t_2 and t_1 respectively. The spectrum was processed using a squared sine function 3 in t_3 and t_2 and a sine bell function 2 in t_1, with zero filling 256 in t_3 and 64 in t_2 and t_1, with 8 linear prediction coefficients in t_1 and t_2.

The NCOCX in **Figure 1e** (400 MHz DNP) was acquired using 15 points in t_1 and t_2 with spectral width of 5,027 Hz and 2,269 Hz in t_2 and t_1 respectively. The spectrum was processed using a sine bell function 3 in t_3 with 2k zero filling. In t_2 and t_3 the spectrum was processed with sine bell function 2 and 64 zero filling in both dimensions, with 4 linear prediction coefficients. Further experimental parameters relating to **Figure 1** and **Supplementary Figures 3**, **6**, **7** and **11** are given in **Supplementary Tables 3**–**8**.

Modeling and structural analysis. NMR signals were predicted using crystal structures of the outer membrane complex crystal structure (PDB 3JQO) and modeled inner membrane complex. The secondary structure of the N termini of B9 and B10 were predicted using Jpred3 (ref. 28) and PSIPRED[29]. Signals stemming from Braun's lipoprotein (Lpp) were predicted using the Lpp crystal structure (PDB 1EQ7[30]). Chemical shift predictions for the crystal part of T4SScc was made using ShiftX[31] and NMR correlations were derived by FANDAS[18]. Analysis of the NMR/DNP spectra was performed using Sparky (https://www.cgl.ucsf.edu/home/sparky/).

The atomic model of the outer membrane complex of T4SScc (PDB 3JQO) was fitted into the electron microscopy (EM) density (EMD-2232)[14] as a rigid body using UCSF Chimera[32]. The predicted transmembrane helix in the N-terminal part of VirB10 was modeled by imposing a helical conformation on the amino acid sequence and then manually placed into the membrane. For the non-crystalline N-terminal part of VirB9, the already docked atomic model (PDB 2YPW)[14] in the EM density (EMD = 2,232) was used. Atomic models of the T4SS core complex in the EM density and of the N-terminal transmembrane helix of B10 are available as **Supplementary Data**.

24. Baldus, M., Petkova, A.T., Herzfeld, J. & Griffin, R.G. *Mol. Phys.* **95**, 1197–1207 (1998).
25. Weingarth, M., Demco, D.E., Bodenhausen, G. & Tekely, P. *Chem. Phys. Lett.* **469**, 342–348 (2009).
26. Bloembergen, N. *Physica* **15**, 386–426 (1949).
27. Fung, B.M., Khitrin, A.K. & Ermolaev, K. *J. Magn. Reson.* **142**, 97–101 (2000).
28. Cole, C., Barber, J.D. & Barton, G.J. *Nucleic Acids Res.* **36**, W197–W201 (2008).
29. Buchan, D.W., Minneci, F., Nugent, T.C.O., Bryson, K. & Jones, D.T. *Nucleic Acids Res.* **41**, W349–W357 (2013).
30. Shu, W., Liu, J., Ji, H. & Lu, M. *J. Mol. Biol.* **299**, 1101–1112 (2000).
31. Neal, S., Nip, A.M., Zhang, H. & Wishart, D.S. *J. Biomol. NMR* **26**, 215–240 (2003).
32. Pettersen, E.F. *et al. J. Comput. Chem.* **25**, 1605–1612 (2004).