Document S2. Data and summary statistics for protein expression, purification, crystallization, and structure determination.

Protein Expression and Purification

The following genes were amplified from their respective genomic DNA sources and cloned using ligation-independent cloning (LIC) into expression vector BG1861, which provides a non-cleavable His_6 tag prior to the ORF (Aslanidis and de Jong 1990):

- the 172-residue soluble domain from residues 51 to 222 of the *Bartonella grahamii* VirB8 gene (UniProt accession code C6AER9, SSGCID identifier BagrA.18388.a)
- the 170-residue soluble domain from residues 65 to 234 of the *Bartonella quintana* VirB8 gene (UniProt accession code Q6FYW3, SSGCID identifier BaquA.18388.b)
- the 172-residue soluble domain from residues 51 to 222 of the *Bartonella tribocorum* VirB8 gene (UniProt accession code A9IWN6, SSGCID identifier BatrA.18388.b)
- the 170-residue soluble domain from residues 63 to 232 of the *Bartonella birtlesii* TrwG gene (UniProt accession code D0AAZ5, SSGCID identifier BabiA.18388.a)
- the 172-residue soluble domain from residues 63 to 233 of the *Bartonella grahamii* TrwG gene (UniProt accession code C6AAT5, SSGCID identifier BagrA.18388.b)
- the 172-residue soluble domain from residues 61 to 232 of the *Rickettsia typhi* RvhB8-II gene (UniProt accession code Q68X84, SSGCID identifier RityA.18390.a)

The sequence of the entire N-terminal tag is MAHHHHHHM, followed by the recombinant protein product.

After transformation into chemically competent *Escherichia coli* BL21 (DE3) Rosetta cells, starter cultures of LB broth for each VirB8 or TrwG construct were grown for ~18 hours at 37 °C. Protein was expressed in a LEX bioreactor in the presence of ampicillin (50 µg ml^{-1}) in 2 liters of sterilized ZYP-5052 auto-induction media (Studier 2005) inoculated with the overnight starter culture. After 24 hours at 25 °C the temperature was reduced to 15 °C for another 60 hours. The sample was centrifuged at 4000 x g for 20 minutes at 4 °C. Cell paste was flash frozen in liquid nitrogen and stored at -80 °C.

The frozen cells were thawed and resuspended in lysis buffer {20 mM HEPES pH 7.4, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 0.5% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 10 mM $MgCl_2$, 3 mM [beta]-mercaptoethanol, 1.3 µg ml-1 protease-inhibitor cocktail, 0.05 mg ml-1 lysozyme} at 4 °C. Lysis was achieved by sonication on ice with Branson Digital 450D Sonifier (70% amplitude with alternating cycles of 5 s pulse on and 10 s

pulse off), followed by incubation with Benzonase (20 µL of 25 unit/µL) at 37 °C for 40 minutes. Insoluble proteins and other cellular components were removed by centrifugation with a Sorvall RC5 at 10 000 rev min^{-1} for 60 min at 4°C in an F14S rotor (Thermo Fisher) and the supernatant was syringe-filtered through a 0.45 µm cellulose acetate filter (Corning Life Sciences, Lowell, Massachusetts, USA). The soluble fraction was loaded onto a Ni-NTA His-Trap FF 5 ml column (GE Biosciences, Piscataway, New Jersey, USA). The column was washed with binding buffer [20 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 1 mM tris(2-carboxyethyl)phosphine (TCEP)] and eluted with 500 mM imidazole in the same buffer. The collected protein was concentrated and further resolved by size-exclusion chromatography (SEC) using a Hiload 26/60 Superdex 75 prep grade column (GE Biosciences) equilibrated in SEC buffer (20 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol, 1 mM TCEP) attached to an ÄKTA FPLC system (GE Biosciences). Peak fractions were collected and pooled based on purity-profile assessment by SDS-PAGE. Concentrated pure protein was flash-frozen in liquid nitrogen and stored at -80°C. The final protein concentrations (*B. grahamii* VirB8, 6.35 mg ml^{-1}; *B. quintana* VirB8, 41.9 mg ml^{-1}; *B. tribocorum* VirB8, 3.84 mg ml^{-1}; *R. typhi* RvhB8-II, 11.2 mg ml^{-1}; *B. birtlesii* TrwG, 25.7 mg ml^{-1}; *B. grahamii* TrwG, 39.0 mg ml^{-1}) were determined by UV spectrophotometry at 280 nm using a Nanodrop spectrophotometer and were calculated using extinction coefficients (*B. grahamii* VirB8, 24410; *B. quintana* VirB8, 25900; *B. tribocorum* VirB8, 24410; *R. typhi* RvhB8-II, 20400; *B. birtlesii* TrwG, 21890; *B. grahamii* TrwG, 23380) obtained with ProtParam (Gasteiger et al. 2005). Final purity (>95%) was assayed by SDS-PAGE electrophoresis.

anSEC Data

To understand the oligomeric state in solution, we analyzed our VirB8, TrwG, and RvhB8-II protein samples using analytical size exclusion chromatography (anSEC) (**fig 1**). In comparison with sizing standards, most of these protein samples gave a single peak with a retention time in between that expected for a monomer or a dimer. This is reflected in the large bubbles, whereas samples that exhibited more than one peak are displayed with small medium and large bubbles to represent the size of different populations. Although some protein samples appeared closer to a monomer and other samples closer to a dimer, there did not appear to be a correlation with our ability to determine a crystal structure or the oligomeric state in the crystal. For example, *B. quintana* VirB8 (BaquA.18388.b.B2.PS01732) appeared closest to a dimer by anSEC but was the only protein to appear as a monomer in the crystal structure. In comparison, *B. birtlesii* TrwG (BabiA.18388.a.B2.PS01693) appeared much closer to a

monomer in anSEC than a dimer, but was a dimer in the crystal structure. Unsurprisingly, protein samples that gave a single peak by anSEC appeared more likely to result in a crystal structure, although *B. tribocorum* VirB8 (BatrA.18388.b.B2.PS0799) had the least clean anSEC trace yet resulted in a crystal structure. Taken together, it is difficult to draw significant conclusions on the oligomeric state in solution from this data.

MS Data on RvhB8-II

The analyses of RvhB8-II potential protein cleavage by MALDI imaging mass spectrometry (MALDI-IMS) and liquid chromatography-electrospray ionization-tandem mass spectrometry (LC/ESI-MS/MS) were conducted in the Medicinal Chemistry Mass Spectrometry Center at the University of Washington (Seattle, WA, USA). The MALDI data was acquired on the Bruker Autoflex II MALDI (Bruker Daltonics, Billerica, MA, USA) in linear mode using a 2:1 mixture of alpha-cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) in acetonitrile as matrix. Equal volumes (1.3uL) of the matrix solution and sample (protein) solution were mixed on the MALDI target plate and allowed to dry before placing the plate into the MALDI. LC/ESI-MS/MS was performed using a Shimadzu QP2010GCMS system (Shimadzu Scientific Instruments, Columbia, MD) with a dual wavelength detector, interfaced to a Waters Micromass Quattro Micro API (Waters Corporation, Milford, MA, USA). The UV wavelengths monitored were both 254 nm. The expected molecular weight of the *R. typhi* RvhB8-II soluble domain with N-terminal His6 tag is approximately 21 kDa. If cleavage of the protein around residues 206-211 had occurred before crystallization, peaks would be expected at approximately 2.7 and 18 kDa, but for both the native (MALDI, see **fig. 2*A***) and denatured (ESI, see **fig. 2*B*,*C***) protein samples MS data show the 21 kDa peak.

Crystallization

R. typhi RvhB8-II crystallized at 11.2 mg/mL against the JCSG+ screen condition H9, which contains 0.2 M lithium sulfate, 0.1 M BisTris pH 5.5, 25% w/v PEG 3350. *B. grahamii* VirB8 crystallized at 6.35 mg/mL against the MCSG1 screen condition E6, which contains 0.2 M potassium sulfate and 20% w/v PEG 3350. *B. quintana* VirB8 crystallized at 21 mg/mL against the JCSG+ screen condition, which contains 0.1 M sodium citrate pH 5.0 and 20% w/v PEG 6000. *B. tribocorum* VirB8 crystallized at 3.84 mg/mL against JCSG+ screen condition F10, which contains 1.1 M sodium malonate, 0.1 M Hepes pH 7.0, and 0.5% w/v Jeffamine ED2001. *B. birtlessi* TrwG crystallized at 25.7 mg/mL against the JCSG+ screen condition A3, which contains 0.2 M citrate and 25% w/v PEG 3350. *B. grahamii* TrwG crystallized at 39 mg/mL

against the PACT screen condition B11, which contains 0.2 M $CaCl_2$, 0.1 M MES pH 6.0, 20% w/v PG 6000. An equal volume of protein (0.4 µL) and precipitant was set up at 16 °C against reservoir (80 µL) in sitting drop vapor diffusion format. Crystals appeared within a week and exhibited a brick-shaped morphology, often with irregular edges.

Data Collection and Processing

Crystals were cryo-protected with reservoir supplemented with 15% w/v ethylene glycol (*R. typhi* RvhB8-II, *B. grahamii* VirB8), w/v 20% ethylene glycol (*B. birtlesii* TrwG, *B. quintana* VirB8, *B. tribocorum* VirB8) or 20% w/v glycerol (*B. grahamii* TrwG) and flash frozen in liquid nitrogen. All data sets were collected at 100 K under a stream of liquid nitrogen (**table 1**). Data were reduced with XDS/XSCALE (Kabsch 2010).

Structure Solution and Refinement

The apo structures (**table 2**) were solved by molecular replacement using the following search models in Phaser (McCoy et al. 2007) from the CCP4 suite (Collaborative Computational Project 1994): *R. typhi* RvhB8-II (PDB ID 4O3V) used PDB ID 4JF8 (*B. birtlesii* TrwG), *B. grahamii* VirB8 (PDB ID 4KZ1) used PDB ID 4JF8 (*B. birtlesii* TrwG), *B. quintana* VirB8 (PDB ID 4LSO) used PDB ID 4KZ1 (*B. grahamii* VirB8), *B. tribocourm* VirB8 (PDB ID 4MEI) used PDB ID 4KZ1 (*B. grahamii* VirB8), *B. birtlesii* TrwG (PDB ID 4JF8) used PDB ID 2BHM (*Brucella suis* VirB8), and *B. grahamii* TrwG (PDB ID 4NHF) used PDB ID 4JF8 (*B. birtlesii* TrwG). The structures were re-built using automated building ARP/wARP (Langer et al. 2008). The final models were obtained after numerous rounds of refinement in REFMAC (Murshudov et al. 1997) and manual re-building in COOT (Emsley and Cowtan 2004). Structures were assessed for correctness and validated using Molprobity (Davis et al. 2007).

REFERENCES

Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. [Internet] 18:6069–6074. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=332407&tool=pmcentrez&rendertype=abstract

Collaborative Computational Project N 4. 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. [Internet] 50:760–763. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15299374

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, et al. 2007. MolProbity: all-atom contacts and structure validation for proteins and

nucleic acids. Nucleic Acids Res. [Internet] 35:W375–W383. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1933162&tool=pmcentrez&rendertype=abstract

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. [Internet] 60:2126–2132. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15572765

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Humana Press. p. 571–607.

Kabsch W. 2010. XDS. Acta Crystallogr. D. Biol. Crystallogr. [Internet] 66:125–132. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815665&tool=pmcentrez&rendertype=abstract

Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. [Internet] 3:1171–1179. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2582149&tool=pmcentrez&rendertype=abstract

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. J. Appl. Crystallogr. [Internet] 40:658–674. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2483472&tool=pmcentrez&rendertype=abstract

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. [Internet] 53:240–255. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15299926

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. [Internet] 41:207–234. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15915565

Fig. 1 (Document S2)

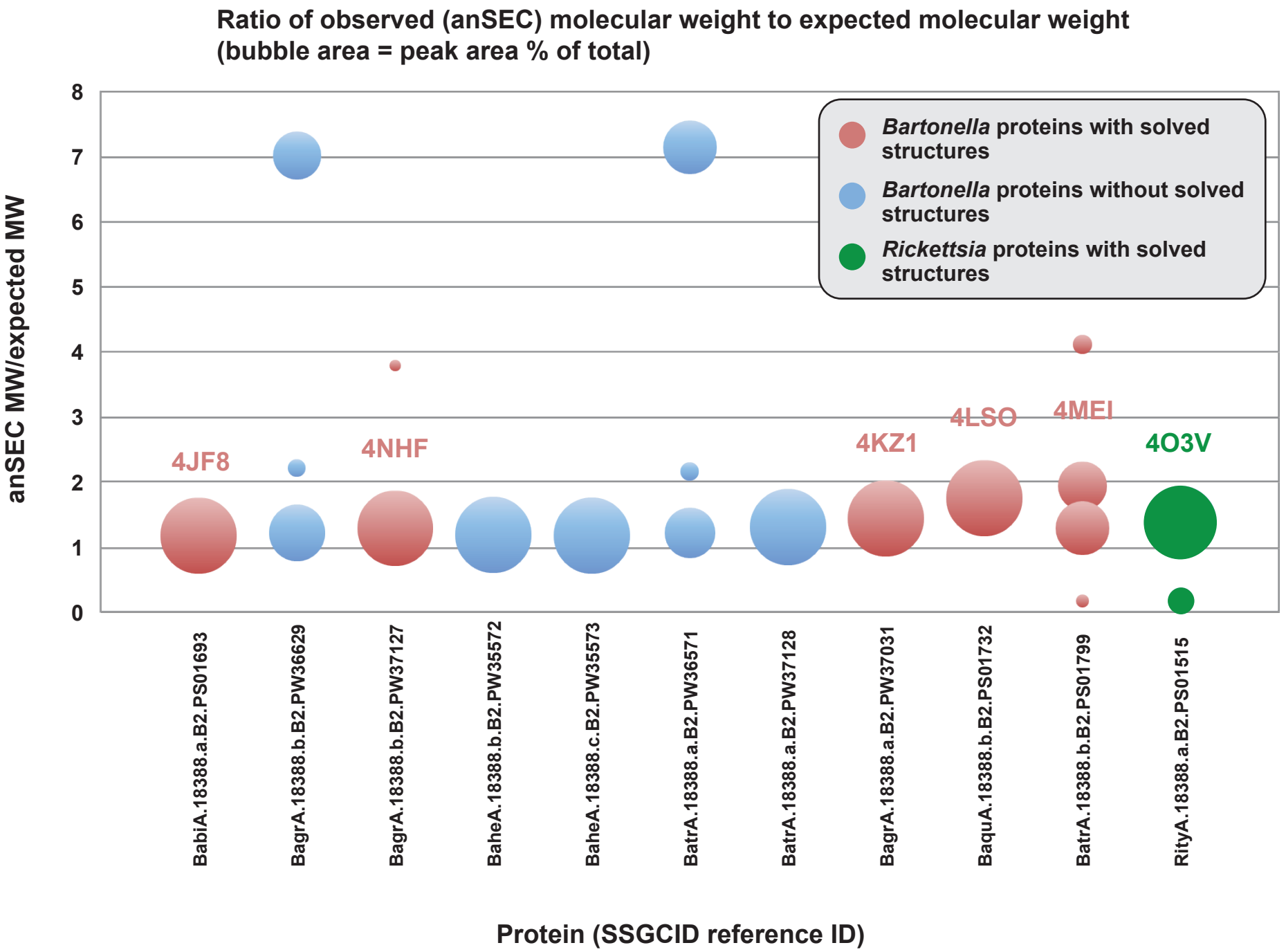

Fig. 2 (Document S2)

A

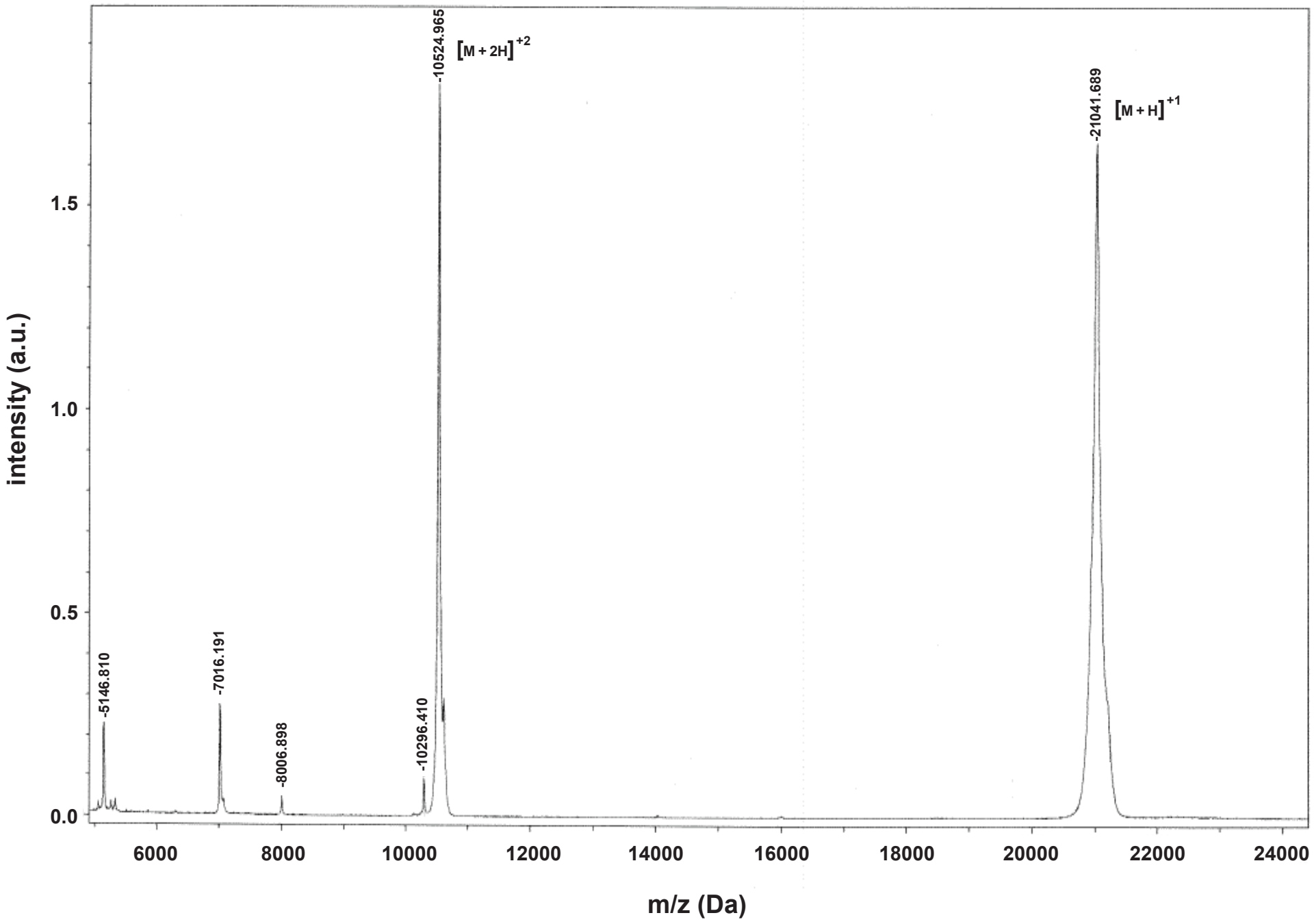

Fig. 2 (Document S2)

B

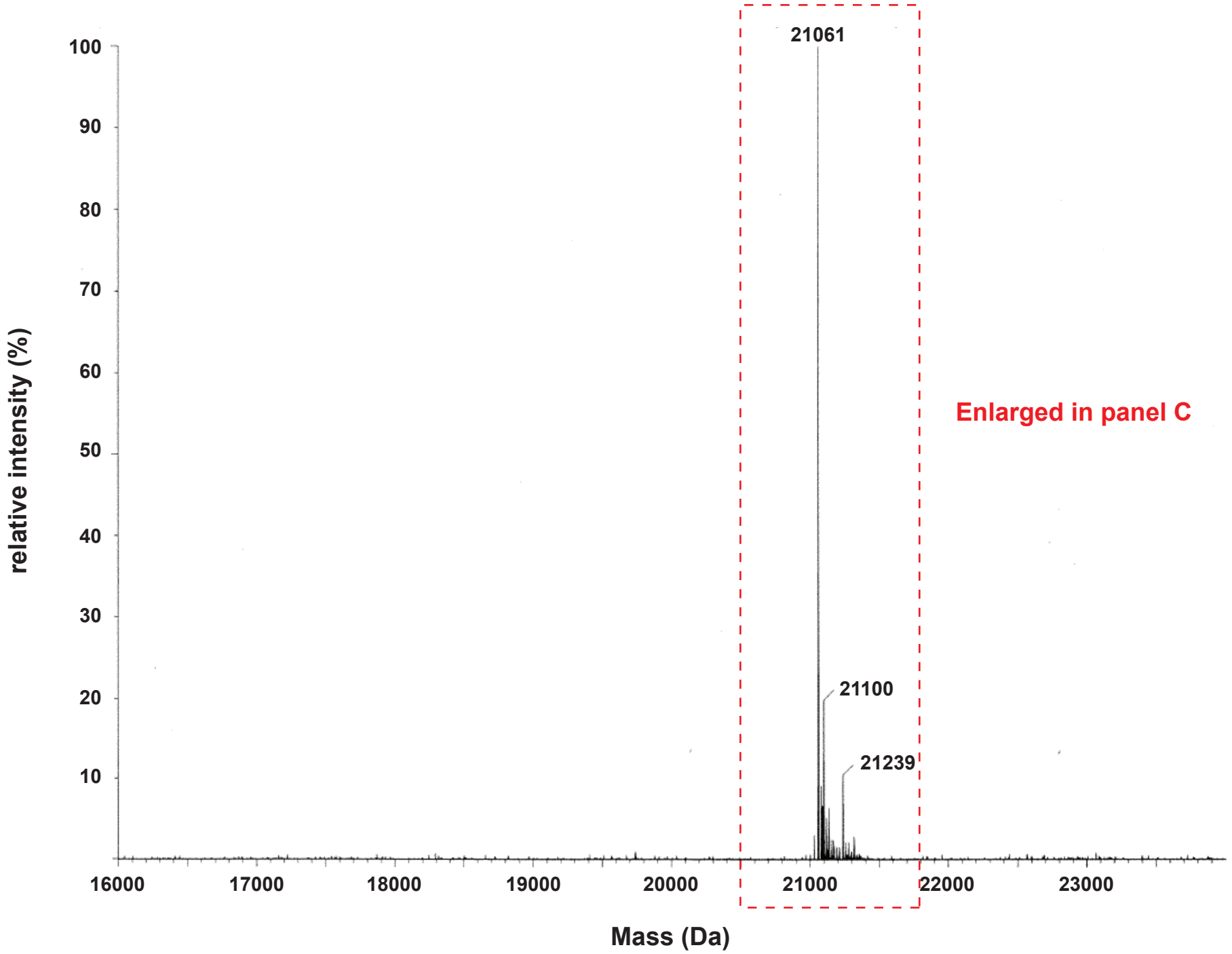

Fig. 2 (Document S2)

C

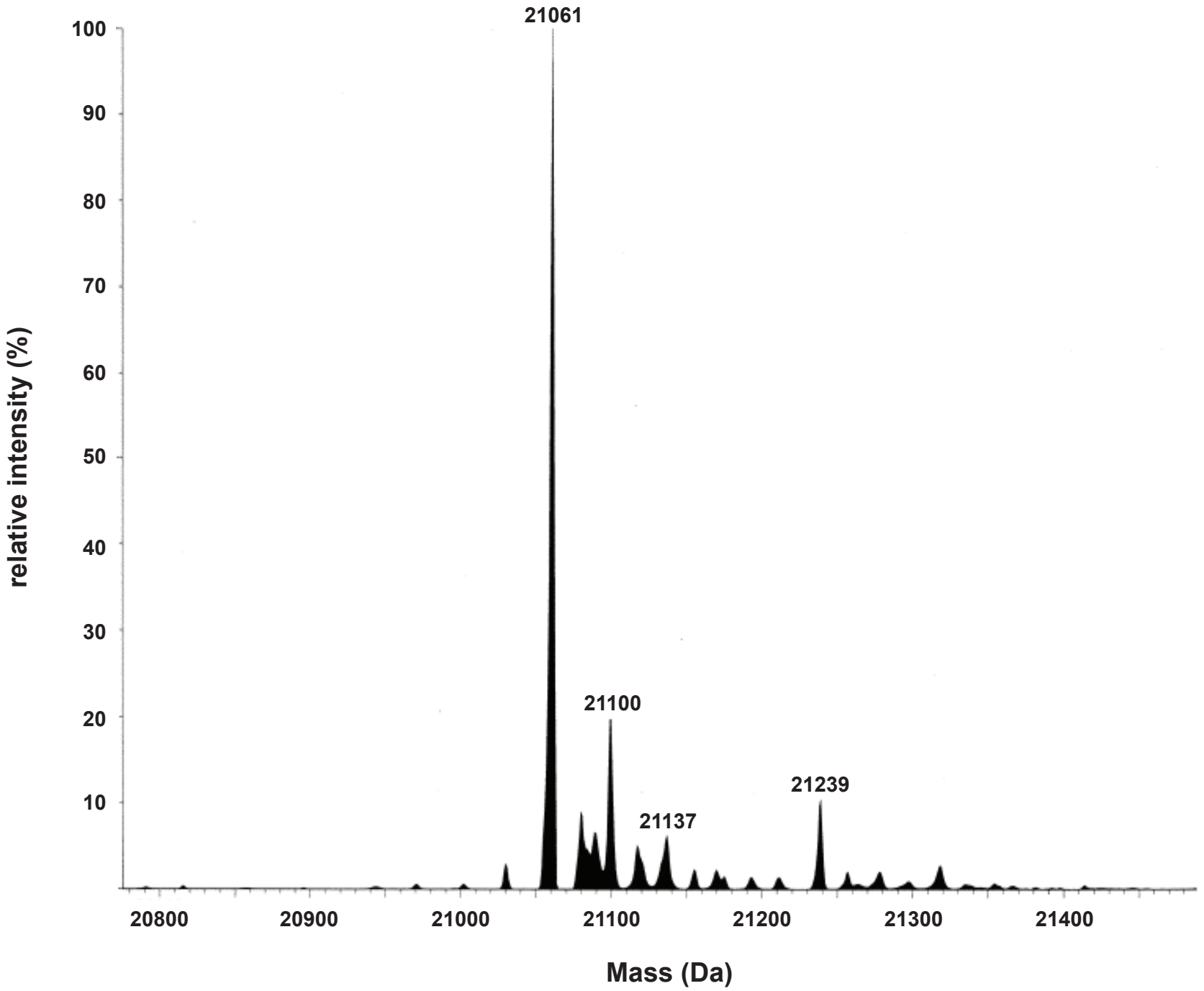

Table 1 (Document S2)

Table 1 (Document S2). Data collection and processing [a].

Species	***B. quintana***	***B. grahamii***	***B. tribocorum***
Strain	**str. Toulouse**	**str. as4aup**	**str. CIP 105476**
Molecule	**VirB8**	**VirB8**	**VirB8**
Wavelength (Å)	0.97872	0.97740	0.97856
Crystal-detector distance (mm)	160	300	370
Rotation range per image (°)	1.0	1.0	1.0
Total rotation range (°)	90	90	115
Exposure time per image (s)	1.5	3	3
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
a, b, c (Å)	60.73, 60.73, 87.86	60.62, 60.62, 124.57	61.54. 61.54, 126.47
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Mosaicity (°)	0.6	0.5	0.3
Resolution range (Å)	50-1.7 (1.74-1.70)	50-2.55 (2.62-2.55)	50-2.85 (2.92-2.85)
Total No. of reflections	132,357 (9,760)	53,029 (4,024)	52,362 (4,091)
No. of unique reflections	18,752 (1,364)	8,063 (579)	6,166 (450)
Completeness (%)	99.8 (100)	99.6 (99.7)	99.7 (100)
Multiplicity	7.1 (7.2)	6.6 (6.9)	8.5 (9.1)
⟨ I/σ(I)⟩	34.3 (4.72)	43.0 (3.8)	23.0 (3.9)
R-factor	0.034 (0.429)	0.030 (0.506)	0.077 (0.589)
CC 1/2	100 (92.8)	100 (92.5)	99.9 (91.0)

Species	***B. birtlesii***	***B. grahamii* str.**	***R. typhi***
Strain	**str. LL-WM9**	**str. as4aup**	**str. Wilmington**
Molecule	**TrwG**	**TrwG**	**RvhB8-II**
Wavelength (Å)	1.12709	0.97872	1.00003
Crystal-detector distance (mm)	150	180	275
Rotation range per image (°)	1.0	1.0	1.0
Total rotation range (°)	150	150	120
Exposure time per image (s)	2.5 at 10% att	1	2.5s
Space group	$P3_121$	$P2_12_12_1$	$P4_12_12$
a, b, c (Å)	72.15, 72.15, 47.44	62.95, 87.65, 191.83	117.68, 117.68, 83.82
α, β, γ (°)	90, 120, 90	90, 90, 90	90, 90, 90
Mosaicity (°)	0.6	0.2	0.3
Resolution range (Å)	50-1.35 (1.39-1.35)	50-2.0 (2.05-2.00)	50-1.95 (2.00-1.95)
Total No. of reflections	261,694 (10,349)	446,829 (32,774)	415,915 (29,249)
No. of unique reflections	31,529 (2,286)	71,950 (5,262)	43,393 (3,147)
Completeness (%)	99.7 (98.6)	99.1 (100)	100 (100)
Multiplicity	8.3 (4.5)	6.2 (6.2)	9.6 (9.3)
⟨ I/σ(I)⟩	34.6 (2.90)	15.4 (3.61)	24.3 (4.15)
R-factor	0.038 (0.497)	0.080 (0.513)	0.065 (0.477)
CC 1/2	100 (84.4)	99.9 (90.9)	99.9 (95.2)

[a] Values for the outer shell are given in parentheses.

Table 2 (Document S2)

Table 2 (Document S2). Structure solution and refinement [a].

Species	***B. quintana***	***B. grahamii* str.**	***B. tribocorum***
Strain	**str. Toulouse**	**str. as4aup**	**str. CIP 105476**
Molecule	**VirB8**	**VirB8**	**VirB8**
Resolution range (Å)	50-1.7 (1.74-1.70)	50-2.55 (2.62-2.55)	50-2.85 (2.92-2.85)
Completeness (%)	99.8 (100)	99.6 (99.8)	99.8 (100)
No. of reflections, total	17,738 (1,272)	8,063 (548)	6,131 (417)
No. of reflections, test set	960 (85)	369 (31)	286 (31)
Final Rcryst	0.176 (0.203)	0.205 (0.337)	0.200 (0.271)
Final Rfree	0.207 (0.235)	0.249 (0.399)	0.247 (0.294)
No. of non-H atoms			
Protein	1,120	1,076	1,092
Water	162	20	14
R.m.s. deviations			
Bonds (Å)	0.011	0.010	0.012
Angles (°)	1.407	1.346	1.477
Average B factors (Å^2)			
Protein	24.8	66.7	59.0
Water	37.3	55.2	50.6
Ramachandran plot			
Most favored (%)	100	96.2	96.3
Allowed (%)	100	100	100
PDB identifier	4LSO	4KZ1	4MEI

Species	***B. birtlesii***	***B. grahamii* str.**	***R. typhi***
Strain	**str. LL-WM9**	**str. as4aup**	**str. Wilmington**
Molecule	**TrwG**	**TrwG**	**RvhB8-II**
Resolution range (Å)	50-1.35 (1.39-1.35)	50-2.0 (2.04-2.00)	50-1.95 (2.00-1.95)
Completeness (%)	99.6 (98.1)	99.1 (99.9)	100 (100)
No. of reflections, total	29,908 (2,161)	71,949 (5,006)	43,389 (2,991)
No. of reflections, test set	1,588 (107)	3,620 (249)	2,182 (150)
Final Rcryst	0.155 (0.213)	0.198 (0.243)	0.167 (0.199)
Final Rfree	0.175 (0.226)	0.240 (0.276)	0.189 (0.240)
No. of non-H atoms			
Protein	1,134	6,640	2,226
Water	229	612	318
R.m.s. deviations			
Bonds (Å)	0.009	0.013	0.013
Angles (°)	1.304	1.495	1.464
Average B factors (Å^2)			
Protein	13.4	31.7	31.5
Water	25.5	34.8	41.8
Ramachandran plot			
Most favored (%)	98.7	97.9	99.3
Allowed (%)	100	99.8	100
PDB identifier	4JF8	4NHF	4O3V

[a] Values for the outer shell are given in parentheses.