



UNIVERSITY OF LEEDS

This is a repository copy of *Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/161770/>

Version: Accepted Version

---

**Article:**

Madej, M, White, JBR, Nowakowska, Z et al. (10 more authors) (2020) Structural and functional insights into oligopeptide acquisition by the RagAB transporter from *Porphyromonas gingivalis*. *Nature Microbiology*, 5 (8). pp. 1016-1025. ISSN 2058-5276

<https://doi.org/10.1038/s41564-020-0716-y>

---

© The Author(s), under exclusive licence to Springer Nature Limited 2020. This is an author produced version of an article published in *Nature Microbiology*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

1  
2  
3  
4

## 1. Extended Data

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: <i>Smith_ED_Fig1.jpg</i>	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	<b>Unbiased peptide density in RagAB</b>	EDFig1.tiff	Stereo diagrams showing 2Fo-Fc ( <b>a</b> ) and Fo-Fc density ( <b>b</b> ) in RagAB KRAB before any modelling and refinement of the peptide. Selected segments of RagA (cyan) and RagB (green) neighbouring the peptide are shown. Map contouring parameters are $1.0\sigma$ , carve = 2 for the 2Fo-Fc and $3.0\sigma$ , carve = 2 for the Fo-Fc map. The extensive contacts of RagA with the peptide are confirmed by a PISA interface analysis <sup>17</sup> which shows that 26 RagA residues form an interface with the peptide compared to only 8 for RagB, generating interface areas of 620 and 240 Å <sup>2</sup> with RagA and RagB respectively. The PISA CSS (complexation significance score) is maximal (1.0) for peptide-RagA while it is only 0.014 for peptide-RagB. This suggests that the observed co-crystal structure represents a state where the ligand has been partially transferred from an initial, presumably low(er)-affinity binding site on RagB to a high(er)-affinity binding site in the RagAB complex, allowing co-purification.
Extended Data Fig. 2	<b>RagAB binds a wide range of oligopeptides</b>	EDFig2.tiff	<b>a-c</b> , LC-MS/MS analysis of peptides bound to RagAB W83 KRAB, showing length distribution ( <b>a</b> ), total charge ( <b>b</b> ) and pI ( <b>c</b> ). <b>d-f</b> , Analysis of peptides

			<p>bound to RagAB W83 wild-type, showing length distribution (<b>d</b>), total charge (<b>e</b>) and pI (<b>f</b>). For charge calculations, the pH was assumed to be 7.0 and contributions of any His residues were ignored. <b>g</b>, Amino acid frequency of RagAB-bound peptides (KRAB and wild-type combined; black) vs. the amino acid composition in the <i>P. gingivalis</i> proteome (gray), showing a substantial enrichment of Ala, Glu, Lys, Thr and Val. By contrast, aromatics (Phe, Trp) and bulky hydrophobics (Leu) appear to be under-represented. The peptides bound to RagAB from W83 KRAB vary in length from 7 to 29 residues, with a broad maximum of around 13 residues that fits well with the peptide density observed in the structures. Assuming equal abundance of each detected peptide, there is a slight preference for neutral to slightly acidic peptides, and the pI distribution has a bimodal shape, with maxima for acidic and slightly basic peptides. Analysis of the smaller RagAB-bound peptide set from wild-type W83 (<b>d-f</b>) yields a slightly wider size range from 5-36 residues, but overall there are no dramatic differences in the collective length, net charge and pI of the RagAB-bound peptide populations from W83 KRAB and wild-type strains.</p>
Extended Data Fig. 3	<p><b>Molecular dynamics simulations of RagAB show lid opening</b></p>	EDFig3.tiff	<p><b>a</b>, C<sub>α</sub>-rmsd values of RagB lids in apo-RagAB (red) and peptide-bound RagAB (green) with reference to the starting crystal structure in the closed conformation. The C<sub>α</sub>-rmsd values of the RagB lids in RagA<sub>2</sub>B<sub>2</sub> are shown in blue with reference to the OO EM state. Each point indicates an average of 50 ns simulation trajectory. <b>b</b>, Comparison of the RagA<sub>2</sub>B<sub>2</sub> open conformation from EM (magenta)</p>

			<p>with the snapshot of the most open simulation at 2500 ns (green). <b>c</b>, Internal surface of peptide binding cavities in closed holo-RagAB and apo-RagAB, generated with CASTp<sup>68</sup>. The bound peptide from a RagAB subunit in the crystal structure was removed <i>in silico</i> to generate a closed apo-complex, and three independent MD simulations were performed. For one of the simulations, a clear opening of the RagB lid was observed, reminiscent of recent results for a SusCD transporter and supporting the notion that ligand removal resets the transporter to favour the open state<sup>16</sup>. None of the peptide-bound complexes shows lid opening on the timescale of the simulations. While this suggests that lid opening is less favourable in the ligand-bound state, it does not contradict our observation of open, ligand-bound complexes via EM. The EM structures allowed us to compare both open states, which showed that the RagB lid in the simulation opens less wide than that in the EM structure, at least during the timescale of the simulation. We also observed a partial closing of both RagB lids during a 1000 ns simulation starting from the OO EM state (<b>a</b>, blue curves). The r.m.s.d. values of both RagB subunits decrease from ~30 Å in the EM structure (t = 0 ns) to ~15 Å, which is similar to the opening observed in one of the apo-RagAB simulations starting from the closed structure. Thus, it appears that the energy minimum for the open state in the simulations is different from that in solution, for reasons that are not clear.</p>
Extended Data Fig. 4	<b>RagB moves as a rigid body during</b>	EDFig4.tiff	<b>a</b> , Superposition of RagA subunits in the open (yellow) and closed (cyan) RagAB complexes in the

	<b>lid opening</b>		OC state, showing the rigid-body movement of RagB. Equivalent points are indicated by x,y,z (closed RagB; green) and by x*, y*, z* (open RagB; red). The arrow indicates the approximate pivot point in the N-terminus of RagB at the back of the complex. <b>b</b> , Superposition as in (a), viewed from the extracellular side. Lid opening results in displacements of up to 45 Å for main chain atoms at the front of the complex, furthest away from the RagB N-terminus. <b>c</b> , Superposition of the open and closed states of RagB, with the N-termini indicated. <b>d</b> , Extracellular view of superposed RagA, with selected loops labelled. The conformational changes upon lid movement are mostly confined to those parts of the protein that continue to interact with RagB at the back of the complex (L7-L9).
Extended Data Fig. 5	<b>Local resolution-filtered cryoEM maps and evidence for NTE density</b>	EDFig5.tiff	<b>a</b> , CC, OC and OO states of RagAB filtered and coloured by local resolution. Corresponding Fourier Shell Correlation (FSC) curves are shown (right). <b>b</b> , Unsharpened maps of RagAB displayed at low contour levels to reveal diffuse density attributed to the NTE. CC, OC and OO states are coloured purple, blue and green respectively.
Extended Data Fig. 6	<b>Analysis of RagAB mutants</b>	EDFig6.tiff	<b>a</b> , Representative growth curves (n = 3, mean ± standard error of the mean) for mutant W83 <i>ragAB</i> variants on BSA-MM. For comparison, W83 WT and $\Delta ragAB$ strains are shown as well. <b>b</b> , Representative SDS-PAGE gel (n = 2) showing OM protein expression levels following removal of inner membrane proteins by sarkosyl treatment. The RagAB <sub>mono</sub> and both RagA hinge loop mutant strains

			<p>have a similar phenotype as <math>\Delta</math>RagAB, suggesting they cannot take up oligopeptides produced by gingipains. However, the OM protein levels show that very little RagAB is present, so that no conclusions about functionality can be drawn. The RagB acidic loop mutants show intermediate growth on BSA-MM, suggesting that oligopeptide uptake is somewhat impaired. However, the OMP levels of the acidic loop mutants are substantially lower than wild type, suggesting that both acidic loop mutants are likely functional and arguing that the slow growth of <i>rag-4 P. gingivalis</i> strain ATCC 33277 on BSA-MM is not due to the absence of the acidic loop in RagB.</p>
Extended Data Fig. 7	<b>RagAB and RagB bind peptides selectively</b>	EDFig7.tiff	<p>MST titration curves with the P4-FAM peptide for (a) RagAB from ATCC 33277, (b) RagAB from W83 and (c) Omp40-41 from W83 (negative control). The other panels show MST profiles for unlabelled P21, P12 or P4 binding to His-tag labelled W83 RagAB (d-f) and His-tag labelled W83 RagB (g-i). Experiments and listed <math>K_d</math> values represent the mean of three independent experiments <math>\pm</math> SD.</p>
Extended Data Fig. 8	<b>Electron density comparison for different peptide ensembles</b>	EDFig8.tiff	<p>a, Fo-Fc electron density maps (contoured at <math>3.0 \sigma</math>) for the modelled peptide following final refinement for RagAB purified from W83 KRAB (left panel; 3.4 Å resolution), RagAB from W83 wild type co-crystallised with excess P21 peptide (middle panel; 2.6 Å), and RagAB from W83 wild type (right panel; 3.0 Å). Peptide sequence of the P21 co-crystal structure is arbitrarily modelled as QNGGANTSRGSAG, with numbering in italics according to the W83 KRAB peptide. Neighbouring residues D400-M407 of RagA are shown as cyan</p>

			stick models for orientation purposes. <b>b</b> , Simulated annealing composite 2Fo-Fc omit maps for peptides bound to RagAB complexes as in <b>(a)</b> . An annealing temperature of 500 K was used, with 5% of the models omitted. The fact that two different peptide ensembles produce similar maps as the P21 peptide, together with the inability to model the P21 sequence, suggests that the substrates are bound with register shifts and perhaps different chain directions. <b>c</b> , Stereo view of superposed 2Fo-Fc maps (made within Phenix; 1.0 $\sigma$ , carve = 2.0) for WT RagAB in the absence (blue) and presence (orange) of P21 peptide, generated with the same high resolution cutoff (3.0 Å). For orientation purposes, the P21 peptide model is shown as sticks.
Extended Data Fig. 9	<b>Schematic demonstrating the proposed mechanism of substrate capture and translocation by RagAB</b>	EDFig9.tiff	Peptide ligands to be imported by the RagAB system are predominantly generated by the action of gingipains on serum and tissue-derived proteins. <b>1</b> . A lid-open state of RagAB permits peptide binding. <b>2</b> . Contributions from both RagA and RagB to peptide binding elicits closure of the lid, forming the transport-competent state of the complex. This is signalled across the OM by perturbation of the TonB box region on the periplasmic side of the plug domain, making it accessible to TonB. <b>3</b> . According to the literature consensus, TonB-mediated disruption of the plug permits substrate translocation and a return to the open state of RagAB.
Extended Data Fig. 10	<b>Cryo-electron microscopy data support substrate-induced lid closure in RagAB</b>	EDFig10.tiff	<b>a,b</b> , 3D classes for RagAB 'as purified' <b>(a)</b> and in the presence of 50-fold excess P21 peptide <b>(b)</b> . Classes corresponding to the CC, OC and OO states are coloured purple, blue and green respectively. Junk or ambiguous classes <i>e.g.</i> where RagA barrels are

			incomplete are coloured grey. In the presence of P21 there was no clear OO state whilst the the proportion of the CC state increased, supporting the proposed mechanism of substrate capture.
--	--	--	---

5 **2. Supplementary Information:**

6

7 **A. Flat Files**

8

9

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: <i>Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.</i>
Supplementary Information	Yes	SI-final.pdf	Supplementary Figures 1-3, Supplementary Tables 1-4, Supplementary Movie legend
Reporting Summary	Yes		

10

11

12 **B. Additional Supplementary Files**

13

14

Type	Number If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: <i>Smith_</i>	Legend or Descriptive Caption Describe the contents of the file
------	--	--	--



		<i>Supplementary_Video_1.mov</i>	
Supplementary Video		RagAB-movie.mp4	<b>Supplementary Movie</b> Dynamics of the RagAB transporter. Cryo-EM map of open-closed (OC) RagAB, with cartoon models shown for RagA (blue) and RagB (yellow or grey) subunits.
Supplementary Data		RagAB-peptidomics.xlsx	<b>Supplementary Table 2</b> RagAB peptidomics and <i>in vitro</i> binding of peptides to RagAB and RagB. Both analyses contain two separate Excel spreadsheets: all information retrieved from Mascot (Mascot) and a reduced spreadsheet with summed spectra (duplicates) and one charge variant of each peptide (Spectral C.) For spectral C., additional statistics were calculated: Spectral count peptide - summed number of

			<p>spectra of particular peptide; Spectral count protein - summed number of spectra per particular protein; Spectral count sample - summed number of spectra per particular sample; Ratio peptide/protein - ratio of total number of particular peptide spectra to total number of spectra per protein; Ratio peptide/sample - ratio of total number of particular peptide spectra to total number of spectra per protein.</p>
--	--	--	--

15  
16

17 **3. Source Data**

18

19

Figure	Filename	Data description
	<p>This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: <i>Smith_SourceData_Fig1.xls</i>, or <i>Smith_</i></p>	<p>i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.</p>

	<i>Unmodified_Gels_Fig1.pdf</i>	
Source Data Fig. 1	Uncropped-Fig1a.pdf	Unprocessed SDS-PAGE gel for Fig. 1a
Source Data Fig. 4	Source_Data-Fig4.xlsx	Excel spreadsheet with source data for Fig. 4
Source Data Extended Data Fig. 2	Source_Data-EDFig2.xlsx	Excel spreadsheet with source data for ED Fig. 2
Source Data Extended Data Fig. 6	Source_Data-EDFig6a.xlsx	Excel spreadsheet with source data for ED Fig. 6a
Source Data Extended Data Fig. 6	Uncropped-EDFig6b.pdf	Unprocessed SDS-PAGE gel for ED Fig. 6b
Source Data Extended Data Fig. 7	Source_Data-EDFig7.xlsx	Excel spreadsheet with source data for ED Fig. 7

20

21

22

23

24

25 **Structural and functional insights into oligopeptide acquisition**  
26 **by the RagAB transporter from *Porphyromonas gingivalis***

27

28 Mariusz Madej<sup>1,2\*</sup>, Joshua B. R. White<sup>3\*</sup>, Zuzanna Nowakowska<sup>1</sup>, Shaun Rawson<sup>3‡</sup>, Carsten  
29 Scavenius<sup>4</sup>, Jan J. Enghild<sup>4</sup>, Grzegorz P. Bereta<sup>1</sup>, Karunakar Pothula<sup>5</sup>, Ulrich Kleinekathoefer<sup>5</sup>,  
30 Arnaud Baslé<sup>7</sup>, Neil Ranson<sup>3#</sup>, Jan Potempa<sup>1,6#</sup>, Bert van den Berg<sup>7#</sup>

31

32 <sup>1</sup>Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian  
33 University, 30-387 Krakow, Poland

34

35 <sup>2</sup>Malopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland

36

37 <sup>3</sup>Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of  
38 Leeds, Leeds, LS2 9JT, UK.

39

40 <sup>4</sup>Interdisciplinary Nanoscience Center (iNANO), and the Department of Molecular Biology, Aarhus  
41 University, Aarhus, DK-8000, Denmark

42

43 <sup>5</sup>Department of Physics and Earth Sciences, Jacobs University Bremen, Campus Ring 1, 28759  
44 Bremen, Germany

45

46 <sup>6</sup>Department of Oral Immunology and Infectious Diseases, University of Louisville School of  
47 Dentistry, Louisville, KY40202, USA

48

49 <sup>7</sup>Biosciences Institute, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH,  
50 UK.

51

52

53 ‡ Present address: The Harvard Cryo-Electron Microscopy Center for Structural Biology, Harvard  
54 Medical School, Boston, MA, USA.

55

56 \* Authors contributed equally

57

58 # Correspondence to: n.a.ranson@leeds.ac.uk

59 jan.potempa@icloud.com

60 bert.van-den-berg@ncl.ac.uk (lead contact)

61

## 62 **Abstract**

63 *Porphyromonas gingivalis*, an asaccharolytic member of the *Bacteroidetes*, is a keystone pathogen  
64 in human periodontitis that may also contribute to the development of other chronic inflammatory  
65 diseases. *P. gingivalis* utilizes protease-generated peptides derived from extracellular proteins for  
66 growth, but how those peptides enter the cell is not clear. Here we identify RagAB as the outer  
67 membrane importer for peptides. X-ray crystal structures show that the transporter forms a dimeric  
68 RagA<sub>2</sub>B<sub>2</sub> complex, with the RagB substrate-binding surface-anchored lipoprotein forming a closed  
69 lid on the RagA TonB-dependent transporter. Cryo-electron microscopy structures reveal the  
70 opening of the RagB lid and thus provide direct evidence for a "pedal bin" nutrient uptake  
71 mechanism. Together with mutagenesis, peptide binding studies and RagAB peptidomics, our  
72 work identifies RagAB as a dynamic, selective OM oligopeptide acquisition machine that is  
73 essential for the efficient utilisation of proteinaceous nutrients by *P. gingivalis*.

74

## 75 **Introduction**

76 The Gram-negative Bacteroidetes are abundant members of the human microbiota, especially in  
77 the gut. Outside the gut, Bacteroidetes often cause disease, with the best-known examples being  
78 the oral Bacteroidetes *Porphyromonas gingivalis* and *Tannerella forsythia* that are part of the "red  
79 complex" involved in periodontitis<sup>1</sup>, the most prevalent infection-driven chronic inflammation in the  
80 Western world<sup>2</sup>. Accumulating evidence suggests a link between periodontitis and other chronic  
81 inflammatory diseases, including rheumatoid arthritis, Alzheimer's disease, chronic obstructive  
82 pulmonary disease and cardiovascular disease<sup>3-7</sup>. Given this link, and its fastidious growth  
83 requirements, it is important to understand how *P. gingivalis*, thrives and causes dysbiosis of the  
84 oral microbiota, leading to inflammation and periodontal tissue destruction.

85

86 Unlike many human gut Bacteroidetes that specialise in degrading glycans, *P. gingivalis* is  
87 asaccharolytic and exclusively utilises peptides for growth<sup>8</sup>. Those peptides are generated by  
88 multiple proteases<sup>9,10</sup>, the best-known of which are the gingipains. These large and abundant  
89 surface-anchored cysteine endoproteases possess cumulative trypsin-like activity, and are  
90 essential for *P. gingivalis* virulence and growth on proteins as the sole source of carbon<sup>11</sup>.  
91 Crucially, it is not clear how gingipain-generated peptides are taken up by *P. gingivalis*, although a  
92 role for the RagAB outer membrane (OM) protein complex has been proposed<sup>12</sup>. RagAB consists

93 of a TonB dependent transporter (TBDT) RagA (PG\_0185) and a substrate-binding surface  
94 lipoprotein RagB (PG\_0186). However, a recent crystal structure of *P. gingivalis* RagB was  
95 claimed to contain bound monosaccharides<sup>13</sup>, and a role in polysaccharide utilisation was  
96 proposed based on sequence similarity with SusCD systems from gut Bacteroidetes<sup>13-15</sup>. Recent  
97 structures suggest that SusC and SusD proteins from *Bacteroides thetaiotaomicron* form stable  
98 dimeric complexes (SusC<sub>2</sub>D<sub>2</sub>), with SusD capping the extracellular face of the SusC transporter.  
99 Molecular dynamics (MD) simulations and electrophysiology studies led to the proposal that  
100 nutrient uptake likely occurs via a hinge-like opening of the SusD lid, a so-called pedal bin  
101 mechanism<sup>16</sup>. However, the molecular details of nutrient acquisition by SusCD-like complexes  
102 remain unclear, as do any differences between putative OM peptide and glycan transporters.

103

104 Here we report integrated structural and functional studies of RagAB purified from *P. gingivalis*  
105 W83 via X-ray crystallography and cryo-electron microscopy (cryo-EM). The crystal structure  
106 shows a dimeric RagA<sub>2</sub>B<sub>2</sub> complex in which RagB caps the RagA transporter, forming a large  
107 internal chamber that is occupied by co-purified bound peptides. Remarkably, and in sharp  
108 contrast to the crystal structure of RagAB (and indeed of SusCDs), cryo-EM reveals large  
109 conformational changes, with three distinct states found in a single dataset of the detergent-  
110 solubilized RagA<sub>2</sub>B<sub>2</sub> transporter. Together with RagAB peptidomics, structure-based site-directed  
111 mutagenesis, and peptide binding studies, we show that RagAB is a dynamic OM oligopeptide  
112 acquisition machine, with considerable substrate selectivity that is essential for the efficient  
113 utilisation of protein substrates by *P. gingivalis*.

114

## 115 **Results**

### 116 **Purification and X-ray crystal structure determination of RagAB from *P. gingivalis***

117 Since RagA does not express in the *E. coli* OM, we purified RagAB directly from *P. gingivalis* W83  
118 KRAB ( $\Delta kgp/\Delta rgpA/\Delta rgpB$ ). This strain lacks gingipains, reducing proteolysis of many OM  
119 proteins. RagAB is one of the most abundant OM proteins in *P. gingivalis*, and with the exception  
120 of the 230 kDa Hemagglutinin A (HagA), no co-purifying proteins are present (Fig. 1a). This  
121 indicates that the RagAB complex represents the complete transporter. Diffracting crystals were  
122 obtained by vapour diffusion, and the structure was solved by molecular replacement using data to  
123 3.4 Å resolution (Methods and Supplementary Table 1).

124

125 RagAB is dimeric (RagA<sub>2</sub>B<sub>2</sub>), with the same subunit arrangement and architecture (Fig. 1)  
126 observed in two SusCD complexes from *Bacteroides thetaiotaomicron*<sup>16</sup>. RagB caps RagA,  
127 burying an extensive surface area (~3800 Å<sup>2</sup>) and forming a large, closed internal cavity which  
128 contains clearly resolved density for an elongated, ~40 Å long molecule that is bound at the RagAB  
129 interface (Extended Data Fig. 1). The electron density strongly suggests that the bound molecule is  
130 a peptide of ~13 residues in length (Figs. 1d-g), although most side chains are truncated after the

131 C $\beta$  or C $\gamma$  positions. Exceptions are residues 9-10 of the peptide, where well-defined side chain  
132 density comes close to residues D99/D101 in RagB. The main chain of the peptide has somewhat  
133 higher average B-factors than the neighbouring RagA (~95 Å<sup>2</sup> and 73 Å<sup>2</sup> respectively), suggesting  
134 that it corresponds to an ensemble of different peptides with similar backbone conformations,  
135 hereafter referred to as "the peptide". We modelled the peptide arbitrarily as  
136 A<sup>1</sup>STTG<sup>5</sup>ANSQR<sup>10</sup>GSG<sup>13</sup> (Fig. 1e), which decreased the R<sub>free</sub> of the model by about 0.25%. D99  
137 and D101 of RagB are part of an acidic loop insertion (Asp<sup>99</sup>-Glu-Asp-Glu<sup>102</sup>) into an  $\alpha$ -helix in  
138 RagB's ligand binding site that protrudes into the RagAB binding cavity (Fig. 1g). The acidic nature  
139 of this loop suggests that *P. gingivalis* W83 RagAB may prefer to take up basic peptides.  
140 Interestingly, this acidic loop is absent in several RagB orthologs, including that from strain ATCC  
141 33277 (Supplementary Figs. 1 and 2). Several RagAB residues likely form hydrogen bonds with  
142 the backbone of the peptide (Fig. 1e), e.g. the side chain of N894 in RagA forms hydrogen bonds  
143 with residues 6 and 8 of the peptide, and may therefore be important for substrate binding.  
144 Collectively, RagAB residues interact with the backbone of 8 out of the 13 visible peptide residues,  
145 and most of those interactions are provided by RagA.

146

147 To confirm the presence of co-purified peptides in RagAB we employed a LC-MS/MS peptidomics  
148 approach. RagAB from W83 KRAB was associated with several hundred unique peptides  
149 (Supplementary Table 2) with none present in the negative control Omp40-41 sample (Methods),  
150 confirming our structure-based hypothesis that RagAB contains an ensemble of peptides. The  
151 peptides originate mostly from abundant *P. gingivalis* proteins representing all cellular  
152 compartments. As a comparison, we also analysed peptides bound to RagAB purified from the  
153 gingipain-expressing W83 wild-type strain. This complex is identical to that of the KRAB strain but,  
154 due to gingipain activity, the RagAB-associated peptides in wild-type W83 are different from those  
155 in the gingipain-deficient KRAB strain (Supplementary Table 2). Any in-depth analysis of RagAB-  
156 bound peptides is difficult because of the non-quantitative nature of MS. Assuming equal  
157 abundance of each detected peptide, there are no dramatic differences in length distribution, pI  
158 profiles and amino acid frequencies within the RagAB-bound peptides from W83 KRAB and wild-  
159 type strains (Extended Data Fig. 2).

160

### 161 **Conformational changes in RagAB**

162 The crystal structure of RagAB is very similar to previously determined SusCD structures. Both  
163 TBDT barrels are closed on the extracellular side by their RagB (or SusD) caps, even when no  
164 substrate is bound<sup>16</sup>. This suggests either that the closed state is energetically favourable, or that  
165 crystallisation selects for closed states from a wider conformational ensemble. We therefore  
166 investigated the structure of detergent-solubilized RagAB in solution using single particle cryo-EM.  
167 RagAB is unambiguously dimeric in solution. Strikingly, following initial 2D classification it was  
168 immediately apparent that multiple conformations of the RagA<sub>2</sub>B<sub>2</sub> dimer were present, and after

169 further classification steps, three distinct conformations were identified. Following 3D  
170 reconstruction and refinement, three structures were obtained at near-atomic resolution,  
171 corresponding to the three possible combinations of open and closed dimeric transporters: closed-  
172 closed (CC; 3.3 Å), open-closed (OC; 3.3 Å) and open-open (OO; 3.4 Å) (Figs. 2 and 3;  
173 Supplementary Table 3).

174

175 The CC state is essentially identical to the X-ray crystal structure (backbone r.m.s.d. values of ~0.6  
176 Å), and the bound peptide occupies the same position. In the OO state, each RagB lid has  
177 undergone a substantial conformational rearrangement that swings RagB upward, exposing the  
178 peptide binding site and plug domain within the barrel interior. The bound peptide is present in both  
179 barrels of the OO state, at the same position as in the CC state, but the density is weaker (Fig. 3),  
180 suggesting lower occupancy or greater mobility for the bound peptide in the open state. The  
181 presence of peptide in open complexes is consistent with a PISA interface analysis<sup>17</sup> that shows  
182 the observed peptide binding site is mainly formed by RagA.

183

184 The most interesting state is the open-closed transporter (OC), in which the internal symmetry of  
185 the RagA<sub>2</sub>B<sub>2</sub> complex is broken. The density for bound peptide in the open barrel of the OC RagAB  
186 is much weaker than in either barrel of the OO state (Fig. 3) even when both maps are generated  
187 without symmetry, suggesting cross-talk between RagAB units. The OC state unambiguously  
188 shows that the RagB lids in the dimeric complex can open and close separately. Fascinatingly, no  
189 "intermediate" states in the opening of the transporter are observed: essentially all of the individual  
190 RagAB pairs are either open or closed, and these are dimerised to give each of the three possible  
191 states observed (CC, OC and OO). This suggests that either the dynamics of lid opening and  
192 closing are very rapid, and/or that the energy landscape for opening and closing is very rugged,  
193 with distinct stable minima only existing for the three discrete states observed here. MD  
194 simulations performed prior to obtaining the EM structures show a similar, albeit less wide opening  
195 of the RagB lid upon removal of the peptide (Extended Data Fig. 3). Superposition of the RagA  
196 subunits of the open and closed states reveals that the N-terminal ~10 residues of the RagB lid  
197 and the lipid anchor at the back of the complex remain stationary during lid opening (Extended  
198 Data Fig. 4), acting as a hinge about which the rest of the protein moves as a rigid body  
199 (Supplementary Movie).

200

### 201 **The open-closed complex reveals changes in the RagA plug domain**

202 In the consensus model of TonB-dependent transport<sup>14</sup>, extracellular substrate binding induces a  
203 conformational change that is transmitted through the plug domain of the TBDT and results in  
204 disordering of its N-terminal Ton box. This increases accessibility of the Ton box from the  
205 periplasm, permitting interaction with TonB in the inner membrane. The disordered state of the Ton  
206 box can therefore be considered a signal, ensuring that only substrate-loaded transporters form



207 productive complexes with TonB, avoiding futile transport cycles. In crystal structures of TBDTs,  
208 Ton box conformation varies. Where visible, this region interacts with other parts of the plug  
209 domain and is assumed to be inaccessible to TonB. In other structures, electron density for the  
210 Ton box is missing, indicating increased mobility. Importantly, the correlation between ligand  
211 binding and visibility of the Ton box is poor<sup>18</sup>, likely owing to crystallisation<sup>19</sup>. In contrast, the OC  
212 state observed by cryo-EM provides an unbiased structure that relates ligand binding site  
213 occupancy to changes in the plug domain (Fig. 3). In the open side, the first visible residue is Q103  
214 at the start of the TonB box, whereas in the closed side (and in the crystal and CC EM structures),  
215 the density starts at L115. The conformation of L115-S119 is also different from that in the open  
216 complex, with shifts for L115 as large as 10 Å (Fig. 3d). Interestingly, the plug region A211-A219 is  
217 also different in the two states. Notably, the pronounced shift for R218 (~9 Å for the head group;  
218 Fig. 3e) may be important given its location at the bottom of the binding cavity. A211-A219  
219 contacts the region following the Ton box, highlighting a potential allosteric route to communicate  
220 binding site occupancy to the Ton box. This is a mechanism that would be consistent with previous  
221 AFM data suggesting that the plugs of TBDTs consist of two domains<sup>20</sup>: an N-terminal, force-labile  
222 domain that is removed by TonB to form a channel and a more stable C-terminal domain that  
223 would include R218 in RagA. The difference in dynamics of the N-terminus of RagA is most  
224 evident in unsharpened maps of the OC state contoured at low levels, which clearly show globular  
225 density connected to the plug domain only in the open state (Extended Data Fig. 5). This density  
226 corresponds to an ~80-residue N-terminal extension (NTE; Pfam 13715, Carboxypeptidase D  
227 regulatory-like domain) that is present in a relatively small subset of TBDTs<sup>21</sup> and which precedes  
228 the Ton box. In the closed state, the NTE and Ton box are not resolved in the cryo-EM maps,  
229 presumably because of increased mobility.

230

### 231 **RagAB is important for growth of *P. gingivalis* on proteins as carbon source**

232 We initially assessed whether the *P. gingivalis* strains W83 and ATCC 33277 (hereafter named  
233 ATCC) could grow on minimal medium supplemented with BSA as a sole carbon source (BSA-  
234 MM; Methods). Interestingly, while growth on rich medium is identical, robust growth on BSA-MM  
235 is observed only for the *rag-1* locus W83 strain<sup>22</sup>. By contrast, the *rag-4* strain ATCC grows slowly,  
236 and after a substantial lag phase (Figs. 4a,c). Given the high sequence similarity of the RagA and  
237 RagB proteins (Supplementary Figs. 1 and 2), the distinct growth profiles of W83 and ATCC  
238 suggest that relatively small differences in RagAB could alter substrate uptake sufficiently to  
239 dramatically affect growth on the relatively small set of peptides that can be generated from BSA.

240

241 To further investigate potential functional differences between the RagAB complexes from W83  
242 and ATCC we constructed an ATCC strain in which W83 *ragAB* was expressed from a single-copy  
243 plasmid in the ATCC  $\Delta$ *ragAB* background. We also made a strain in which the genomic copy of  
244 ATCC *ragB* was replaced by W83 *ragB*. Remarkably, replacement of either RagB or RagAB from

245 ATCC with the corresponding orthologs from W83 results in robust growth of ATCC on BSA-MM  
246 (Fig. 4c). Since the OM levels of RagAB for all strains are similar (Fig. 4d), these results have  
247 several important implications. Firstly, they confirm that RagAB is required for growth on  
248 extracellular protein-derived oligopeptides. Secondly, they suggest that different RagB lipoproteins  
249 can form functional complexes with the same RagA transporter. Finally, and perhaps most  
250 surprisingly, RagB appears to determine the substrate specificity of the complex. We speculate  
251 that TonB-dependent transporters such as SusC and RagA may have acquired SusD and RagB  
252 substrate binding lipoproteins to confer substrate specificity.

253

254 The long lag phase during growth of ATCC on 1% BSA-MM is consistent with previous data  
255 showing that this strain requires either high substrate concentrations (3% BSA) or supplementation  
256 with rich medium for growth on BSA or human serum albumin (HSA)<sup>23-25</sup>. This may seem  
257 surprising, given that HSA is an abundant component of gingival crevicular fluid (GCF), an  
258 inflammatory exudate which is dominated by abundant plasma proteins such as albumin<sup>26</sup>.  
259 However, GCF does also contain many other proteins that could be substrates for ATCC, and  
260 many other bacteria are present and could generate different substrates from those proteins. Thus,  
261 slow growth on a single substrate *in vitro* does not translate to poor fitness *in vivo*. Poor growth on  
262 BSA also does not imply that ATCC RagAB does not take up any BSA peptides, as the phenotype  
263 could be caused by inefficient uptake of a particular BSA peptide that provides one or more  
264 essential amino acids. Our Supplementary Table 2 identifies only 63 BSA-derived peptides after  
265 trypsin digest, so the substrate pool is relatively small. This notion is supported by previous data  
266 showing that while ATCC efficiently cleaves human albumin, transferrin, immunoglobulin G and  
267 collagen, the two latter substrates don't support growth<sup>24</sup>.

268

269 To better define the role of RagAB in peptide utilisation, we constructed clean deletions in *P.*  
270 *gingivalis* W83 for *ragA*, *ragB*, and *ragAB* and grew the resulting strains on BSA-MM. Compared  
271 to wild-type W83 RagAB, growth of the *ragA* and *ragAB* deletion strains exhibits long (~15-20 hrs)  
272 lag periods (Fig. 4a). Unlike the gingipain-*null* KRAB strain<sup>11</sup>, both mutant strains grow eventually,  
273 most likely due to passive, RagAB-independent uptake of small peptides produced after prolonged  
274 BSA digestion. Interestingly, the  $\Delta ragB$  strain grows better on BSA than the  $\Delta ragAB$  strain. Since  
275 the strain still produces RagA (albeit at very low levels; Fig. 4b and Extended Data Fig. 6) this  
276 demonstrates that RagB, in contrast to RagA, is not required for growth on BSA-derived  
277 oligopeptides *in vitro*, in accordance with previous work<sup>12</sup>. Collectively, the growth data suggest  
278 that RagAB mediates uptake of extracellular oligopeptides produced by gingipains, and help  
279 explain data showing that the transporter is important for *in vivo* fitness and virulence of *P.*  
280 *gingivalis*<sup>22,27</sup>.

281

282 We next constructed mutant strains for structure-function studies (see Methods), and tested growth  
283 on BSA-MM (Figs. 4a and Extended Data Fig. 6). Most mutant strains show very low OM levels of  
284 RagAB, complicating conclusions about functionality (Extended Data Fig. 6). Exceptions are the  
285 Ton box deletion ( $\Delta$ Ton) and the NTE deletion ( $\Delta$ NTE), which are expressed at reasonable levels  
286 (Fig. 4b). The  $\Delta$ Ton strain has a similar phenotype as  $\Delta$ RagAB, demonstrating that this variant is  
287 inactive and therefore that RagAB is a *bona-fide* TBDT. Strikingly, the  $\Delta$ NTE variant resembles the  
288 KRAB strain and does not grow even after a prolonged period (Fig. 4a). Since the  $\Delta$ RagAB strain  
289 (which also lacks the NTE) does grow after a lag phase, the results suggest that the NTE is  
290 important. In Proteobacteria, structures of TBDTs with a different N-terminal domain (Pfam 07660;  
291 STN) have been reported<sup>28</sup>. This domain interacts with an anti-sigma factor in the inner membrane  
292 to stimulate TBDT expression in response to the presence of their cognate substrates, exemplified  
293 by the *E. coli* FecA-FecIR system<sup>29-31</sup> in which the anti-sigma factor *fecR* and the sigma factor *fecI*  
294 are located in the same operon as *fecA*. By contrast, while *P. gingivalis* has five anti-sigma factors,  
295 none of these are located near the *ragAB* locus. Moreover, deletion of the N-terminal domain in  
296 FecA only affects FecA upregulation and has no effect on transport activity<sup>29</sup>, contrasting with the  
297 dramatic phenotype for the RagA  $\Delta$ NTE mutant. Thus, while an involvement of the RagA NTE in  
298 signalling seems likely, establishing the signalling partner(s) and mechanism remains the subject  
299 of further work.

300

### 301 **Peptide binding by RagAB is selective**

302 We next characterised peptide binding to RagAB in more detail. Since RagAB is purified with an  
303 ensemble of peptides that cannot be removed without denaturing the complex, added peptides  
304 compete with bound peptides, resulting in the measurement of apparent rather than true  
305 dissociation constants. The 21-residue peptide KATAEALKKALEEAGAEVELK (henceforth named  
306 P21; charge -1) from the C-terminus of ribosomal protein L7 was abundant in W83 KRAB RagAB  
307 (Methods), and was synthesised in addition to its 12-residue core sequence (DKATAEALKKAL,  
308 denoted P12; charge +1) that is also present in a number of similar peptides but, crucially, is not  
309 identified as a RagAB-bound peptide. We also tested an 11-residue peptide (named P4; sequence  
310 NIFKKNVGFKK) of the arginine deiminase ArcA from *Streptococcus cristatus* that was proposed to  
311 bind to RagB from ATCC<sup>32</sup>.

312

313 We used microscale thermophoresis (MST) to assess peptide binding to RagAB. Initially, we used  
314 N-terminally fluorescein-labelled P4 peptide (P4-FAM) and unlabelled RagAB to maximise the  
315 signal-to-noise, and obtained apparent dissociation constants of  $\sim 2 \mu\text{M}$  for W83 RagAB and  $\sim 0.2$   
316  $\mu\text{M}$  for ATCC RagAB (Extended Data Fig. 7). The negative control Omp40-41 did not show  
317 binding, demonstrating that these results are not due to non-specific partitioning of fluorescently-  
318 labelled peptide into detergent micelles. However, the fluorophore may affect peptide binding in an  
319 unpredictable way, and therefore we next performed MST experiments for unlabelled P21, P12

320 and P4 peptides and His-tag labelled RagAB from wild-type W83. Remarkably, we observed  
321 robust binding only for P21, with an apparent  $K_d$  ( $\sim 0.4 \mu\text{M}$ ; Extended Data Fig. 7) similar to that  
322 observed for substrate binding to classical TBDTs<sup>33-34</sup>. While the fit to a single binding site model is  
323 reasonable, models that assume two binding sites yield significantly better fits (Supplementary Fig.  
324 3). This provides an indication that the two RagA<sub>2</sub>B<sub>2</sub> binding sites may not be equivalent, perhaps  
325 due to cooperativity. The data for P12 and P4 suggest that these peptides cannot displace the  
326 bound endogenous peptides. Notably, the result for unlabelled P4 contrasts with that for P4-FAM  
327 (Discussion). We next asked whether the three peptides bind to RagB, purified from *E. coli*. W83  
328 RagB produced good-quality binding curves with P21 and P4, with similar apparent dissociation  
329 constants of  $\sim 2 \text{ mM}$ , but no binding was observed for P12 (Extended Data Fig. 7). The relatively  
330 low apparent affinities of RagB for oligopeptides are in good agreement with literature values for  
331 oligosaccharide binding to SusD proteins determined via ITC<sup>16,35,36</sup>. Collectively, the P21 data  
332 suggest that peptides bind with lower affinities to RagB (P21  $K_d$  app.  $\sim 2 \text{ mM}$ ) than to RagAB (P21  
333  $K_d$  app.  $\sim 0.4 \mu\text{M}$ ), which makes sense assuming that after initial capture by RagB, the peptide  
334 needs to be transferred to RagA. The data also show that the P4 and P12 peptides are not good  
335 substrates for W83 RagAB, confirming that the transporter has considerable substrate selectivity.

336

337 Since we can measure peptide binding to purified RagAB *in vitro* by MST, added peptides compete  
338 successfully with co-purified endogenous peptides. This, together with the robust growth observed  
339 in BSA-MM, led us to ask whether we could detect acquisition of BSA tryptic peptides by W83  
340 RagAB *in vitro* (Methods). Indeed, the sample incubated with the BSA digest revealed six bound  
341 BSA peptides in addition to endogenous peptides, suggesting that the BSA peptides only partly  
342 replace the co-purified ensemble (Supplementary Table 2). This is due to the fact that the BSA  
343 digest contains 63 different peptides, such that the concentration of BSA "binder" peptides is not  
344 high enough to replace all endogenous peptides. By contrast, a 100-fold excess of P21 completely  
345 displaced the co-purified peptides (Supplementary Table 2). The fact that only a subset of BSA  
346 peptides binds to RagAB demonstrates that the transporter is selective. We also incubated the  
347 BSA tryptic digest experiment with W83 RagB. Prior to incubation, RagB contains only one bound  
348 co-purified peptide. After incubation and post-SEC, two BSA peptides are detected, demonstrating  
349 that at least some peptides bind to RagB with sufficient affinity to survive SEC (Supplementary  
350 Table 2).

351

352 In an attempt to build a unique peptide sequence into the electron density maps, we crystallised  
353 RagAB purified from wild-type W83 in the absence and presence of a 50-fold molar excess of P21.  
354 Comparison of the maps reveals peptide density at the same site, but with some differences  
355 particularly at the N-terminus (Extended Data Fig. 8). However, the P21 sequence does not fit the  
356 density unambiguously, and we therefore speculate that P21 and other peptides are bound with  
357 register shifts and perhaps different chain directions. In all three structures the same RagAB

358 residues hydrogen bond with the backbones of the modelled substrates, providing a clear rationale  
359 how the transporter can bind many oligopeptides. Future studies will be required to define  
360 precisely how peptides bind to RagAB, perhaps by collecting anomalous X-ray data on P21-like  
361 peptides labelled at unique positions with heavy atoms.

362

## 363 **Discussion**

364 Our combined data show that RagAB is a dynamic OM oligopeptide transporter that is important  
365 for growth of *P. gingivalis* on extracellular protein substrates and possibly for peptide-mediated  
366 signalling<sup>32</sup>. We propose a transport model in which the open RagB lid binds substrates before  
367 delivering them to RagA via lid closure (Extended Data Fig. 9). This would make the RagA Ton box  
368 accessible for interaction with TonB, permitting formation of a transport channel into the  
369 periplasmic space. To test the premise that substrate binding induces lid closure we collected cryo-  
370 EM data on RagAB in the absence and presence of excess P21. Particle classification indeed  
371 shows a decrease in OO states and a clear increase in CC states in the P21 sample, in  
372 accordance with our model (Extended Data Fig. 10).

373

374 How does signalling to TonB occur, and how are unproductive interactions of TonB with "empty"  
375 transporters avoided? In the classical, smaller TBDTs such as FecA and BtuB, the ligand binding  
376 sites involve residues of the plug domain, suggesting that ligand binding is allosterically  
377 communicated to the periplasmic face of the plug. However, there are no direct interactions  
378 between the visible, well-defined part of the peptide substrates and the RagA plug. What, then,  
379 causes the observed conformational changes in the plug? One possibility is that parts of peptide  
380 substrates that are invisible (*e.g.* due to mobility) contact the plug. Given the very large solvent-  
381 excluded RagAB cavity even with the modelled 13-residue peptide ( $\sim 9800 \text{ \AA}^3$ ;  $\sim 11500 \text{ \AA}^3$  without  
382 peptide; Extended Data Fig. 3), there is enough space to accommodate the long substrates  
383 identified by the peptidomics, and these could contact the plug directly. However, the presence of  
384 substrate density in the open states of the cryo-EM structures suggests that it may not be the  
385 occupation of the binding site *per se* that is important for TonB interaction, but closure of the RagB  
386 lid (Extended Data Fig. 9). Thus, we hypothesise that certain peptides may bind to RagAB, but do  
387 not generate the closed state of the complex and signal occupancy of the binding site to TonB.  
388 This may provide an alternative explanation for the different MST results for the P4 and P4-FAM  
389 peptide titrations to RagAB (Extended Data Fig. 7). Given the nature of the MST signal, titrating  
390 unlabelled P4 to labelled RagAB is likely to give a signal only if the binding causes a  
391 conformational change (*e.g.* lid closure in the case of RagAB). In the "reverse" experiment, the  
392 readout is on the labelled peptide (P4-FAM), and the large change in mass upon binding could  
393 generate a thermophoretic signal in the absence of any conformational change. Thus, P4 may be  
394 an example of a substrate that binds non-productively to RagAB.

395

396 A fascinating question is why RagAB and related transporters are dimeric. To our knowledge, there  
397 are no other TBDTs that function as oligomers, and there is no obvious reason why dimerisation  
398 would be beneficial. There are few clues in the structures, but when RagA in, *e.g.*, the closed  
399 complexes of the CC and OC states is superposed, RagB shows a rigid-body shift of  $\sim 3\text{-}5$  Å, and  
400 *vice versa*. A similar trend is observed when the open complexes of the OC and OO states are  
401 considered. Moreover, the peptide densities in the OO state are stronger than that of the open  
402 complex in the OC state (Fig. 3). Together, this suggests that the individual RagAB complexes  
403 could exhibit some kind of cross-talk such as cooperative substrate binding. This notion is  
404 supported by the MST data for P21 binding to RagAB, suggesting the presence of two, non-  
405 equivalent ligand binding sites.

406

407

408

409

410

## 411 **Author contributions**

412 MM, JP and BvdB initiated the project. MM cultured cells, purified and crystallised proteins and  
413 performed MST binding experiments, with guidance from JP and BvdB. JBRW and SR determined  
414 cryo-electron microscopy structures, supervised by NR. ZN performed cloning and strain  
415 construction. GB carried out qPCR experiments, and CS and JJE performed the peptidomics  
416 analysis. KP performed the MD simulations, supervised by UK. BvdB purified and crystallised  
417 proteins and determined the RagAB crystal structures. AB collected crystallography data. The  
418 manuscript was written by BvdB with input from MM, JBRW, NR and JP.

419

## 420 **Acknowledgements**

421 This work was supported by a Wellcome Trust Investigator award (214222/Z/18/Z to BvdB). We  
422 would further like to thank personnel of the Diamond Light Source for beam time (Block Allocation  
423 Group numbers mx-13587 and mx-18598) and assistance with data collection. All EM was  
424 performed at the Astbury Biostructure Laboratory which was funded by the University of Leeds  
425 and the Wellcome Trust (108466/Z/15/Z). We thank Drs. Rebecca Thompson, Emma Hesketh  
426 and Dan Maskell for EM support. We would also like to thank T. Kantyka for help in designing MS  
427 experiments. This study was supported in part by National Science Centre, Poland grants UMO-  
428 2015/19/N/NZ1/00322 and UMO-2018/28/T/NZ1/00348 to MM, UMO-2016/23/N/NZ1/01513 to ZN,  
429 UMO-2018/29/N/NZ1/00992 to GPB, and UMO-2018/31/B/NZ1/03968 and NIDCR/DE 022597  
430 (NIH) to JP. LC-MS/MS analysis was supported by the Novo Nordisk Foundation (BioMS). JBRW  
431 was supported by a Wellcome Trust 4 year PhD studentship (215064/Z/18/Z).

432

433 **Ethics declarations**

434 **Competing interests**

435 The authors declare no competing interests.

436

437 **Methods**

438 **Bacterial strains and general growth conditions.** *Porphyromonas gingivalis* strains (listed in  
439 Table S4) were grown in enriched tryptic soy broth (eTSB per liter: 30 g tryptic soy broth, 5 g yeast  
440 extract; further supplemented with 5 mg hemin; 0.25 g L-cysteine and 0.5 mg menadione) or on  
441 eTSB blood agar (eTSB medium containing 1.5% [w:v] agar, further supplemented with 5%  
442 defibrinated sheep blood) at 37 °C in an anaerobic chamber (Don Whitley Scientific, UK) with an  
443 atmosphere of 90% nitrogen, 5% carbon dioxide and 5% hydrogen. *Escherichia coli* strains (listed  
444 in Supplementary Table 4), used for all plasmid manipulations, were grown in Luria–Bertani (LB)  
445 medium and on 1.5% agar LB plates. For antibiotic selection in *E. coli*, ampicillin was used at 100  
446 µg/ml. *P. gingivalis* mutants were grown in the presence of erythromycin at 5 µg/ml and/or  
447 tetracycline at 1 µg/ml.

448

449 **Growth of *P. gingivalis* in minimal medium supplemented with BSA.** An important general  
450 limitation of studies on TonB-dependent transport is that it has not yet been reconstituted *in vitro*  
451 from purified components, due to the involvement of two membranes, one of which has to be  
452 energised and contain the ExbBD-TonB complex. This limits the ability to directly link any TBDT to  
453 substrate import. Therefore, to determine the role of RagAB in the transport of peptides *in vivo*, a  
454 growth assay with minimal medium containing BSA as the sole carbon source (BSA-MM) was  
455 used. Notably, the peptides generated by extracellular *P. gingivalis* proteases such as gingipains  
456 must be intracellular to allow growth, giving a qualitative measure of peptide import. *P. gingivalis*  
457 strains were grown overnight in eTSB. The cultures were washed twice with enriched Dulbecco's  
458 Modified Eagle's Medium (eDMEM per liter: 10 g BSA; further supplemented with 5 mg hemin;  
459 0.25 g L-cysteine and 0.5 mg menadione) and finally resuspended in eDMEM. The OD600 in each  
460 case was adjusted to 0.2 and bacteria were grown for 40 hours. OD600 was measured at 5 hour  
461 intervals.

462

463 **Mutant construction.** For RagA, the following mutants were made: a Ton box deletion ( $\Delta$ Ton;  
464 residues V100-Y109 deleted), a NTE domain deletion ( $\Delta$ NTE; residues V25-K99 deleted) a  
465 putatively monomeric RagAB version made via introduction of a His6-tag between residues Q570-  
466 G571; RagAB<sub>mono</sub>), and deletions of the two loops (L7 and L8) that remain associated with RagB  
467 during lid opening as based on the cryo-EM structures ( $\Delta$ hinge1, residues Q670-G691 deleted and  
468 replaced by one Gly;  $\Delta$ hinge2, residues L731-N748 deleted and replaced by one Gly). For RagB,  
469 two variants were made, both involving the acidic loop: a deletion of this loop ( $\Delta$ AL; residues R97-

470 S104) and conversion of the acidic loop into a basic loop (RagB<sub>BL</sub>; D99/E100/D101/E102 replaced  
471 by R99/K100/R101/K102). All *P. gingivalis* mutants were prepared using wild-type W83 strain or its  
472 derivatives (unless stated otherwise) and constructed by homologous recombination<sup>37</sup>. Mutants  
473 were generated in wild type W83 rather than KRAB W83 due to the presence of multiple antibiotic  
474 cassettes in the latter strain, resulting from removal of the gingipain genes. The “swap” mutants  
475 were prepared using the ATCC33277 strain as background. For deletion strains, 1 kb regions  
476 upstream (5') and downstream (3') from the *ragA*, *ragB* and both *ragAB* genes, as well as chosen  
477 antibiotic resistance cassettes, were amplified by PCR. Obtained DNA fragments were cloned into  
478 pUC19 vector using restriction digestion method and/or Gibson method<sup>38</sup>. For the construction of  
479 master plasmids for RagA and RagB mutants the whole gene sequences were amplified with  
480 addition of antibiotic cassettes, 1kb downstream fragments and cloned into pUC19 vector. Desired  
481 mutations were introduced into master plasmids by SLIM PCR<sup>39</sup>. A similar method was used for  
482 the “swap” RagB-W83inATCC and RagB-ATCCinW83 strains. In both deletion RagB plasmids we  
483 inserted the gene of the other strain (i.e. *ragB* from W83 into  $\Delta$ RagB-ATCC). We also obtained a  
484 RagAB “swap” strains using the pTIO-1 plasmid<sup>40</sup>. The DNA sequences of *ragAB* from both strains  
485 were cloned with the addition of their promoters into pTIO plasmids and further conjugated with the  
486 opposite deletion strains using *E. coli* S-17  $\lambda$ pir (i.e. RagAB-W83-pTIO into  $\Delta$ RagAB-ATCC)<sup>41</sup>.  
487 Primers used for plasmid construction and mutagenesis are listed in Supplementary Table 4. All  
488 plasmids were analyzed by PCR and DNA sequencing. *P. gingivalis* competent cells<sup>42</sup> were  
489 electroporated with chosen plasmids and plated on TSBY with appropriate antibiotics –  
490 erythromycin (5  $\mu$ g/ml) or tetracycline (1  $\mu$ g/ml) and grown anaerobically for approximately 10  
491 days. Clones were selected and checked for correct mutations by PCR and DNA sequencing.  
492 Bacterial strains generated and used in this study are listed in Supplementary Table 4.

493

494 **W83 KRAB RagAB production and purification.** The non-His tagged RagAB complex from *P.*  
495 *gingivalis* W83 KRAB was isolated from cells grown in rich media. In brief, cells from 6 l of culture  
496 were lysed by 1 pass through a cell disrupter (0.75 kW; Constant Biosystems) at 23,000 psi,  
497 followed by ultracentrifugation at 200,00 x g for 45 minutes to sediment the total membrane  
498 fraction. The membranes were homogenised and pre-extracted with 100 ml 0.5% sarkosyl in 20  
499 mM Hepes pH 7.5 (20 min gentle stirring at room temperature) followed by ultracentrifugation  
500 (200,000 x g; 30 min) to remove inner membrane proteins<sup>43</sup>. The sarkosyl wash step was repeated  
501 once, after which the pellet (enriched in OM proteins) was extracted with 100 ml 1% LDAO (in 10  
502 mM Hepes/50 mM NaCl pH 7.5) for 1 hour by stirring at 4 °C. The extract was centrifuged for 30  
503 min at 200,000 x g to remove insoluble debris. The solubilised OM was loaded on a 6 ml  
504 Resource-Q column and eluted with a linear NaCl gradient to 0.5 M over 20 column volumes.  
505 Fractions containing RagAB were ran on analytical SEC (Superdex 200 Increase GL 10/300) in 10  
506 mM Hepes/100 mM NaCl/0.05% LDAO pH 7.5 in order to obtain RagAB of sufficient purity. Finally,



507 the protein was detergent-exchanged to C<sub>8</sub>E<sub>4</sub> using two rounds of ultrafiltration (100 kDa MWCO),  
508 concentrated to 15-20 mg/ml and flash-frozen in liquid nitrogen.

509

510 **Wild type W83 RagAB production and purification.** Homologous recombination was used to  
511 add a 8×His-tag to the C terminus of genomic *ragB* in the *P. gingivalis* W83 strain. The mutant was  
512 grown about 20 h in rich medium under anaerobic conditions. The cells from 6 l of culture were  
513 collected and processed as outlined above. The insoluble material was homogenised with 1.5%  
514 LDAO (in 20 mM Tris-HCl/300 mM NaCl pH 8.0) and the complex was purified by nickel-affinity  
515 chromatography (Chelating Sepharose; GE Healthcare) followed by gel filtration using a HiLoad  
516 16/60 Superdex 200 column in 10 mM Hepes/100 mM NaCl, 0.05% DDM pH 8.0.

517

518 **Purification of RagB from W83 and ATCC 33277 expressed in *E. coli*.** Genes encoding for the  
519 mature parts of RagB from W83 and ATCC 33277 (with His6-tags at the C-terminus) were  
520 amplified by PCR from genomic DNA extracted from *P. gingivalis*. The DNA fragments were  
521 purified and cloned into the arabinose-inducible pB22 expression vector<sup>44</sup> using NcoI/XbaI  
522 restriction enzymes. The obtained expression plasmid was transformed into *E. coli* strain BL21  
523 (DE3). Transformed *E. coli* cells were grown in LB media containing ampicillin (100 µg/ml) at 37 °C  
524 to an OD<sub>600</sub> ~ 0.6 and expression of the recombinant protein was induced with 0.1% arabinose.  
525 After ~2.5 h at 37 °C, cells were collected by centrifugation (5,000 × g; 15 min), resuspended in 20  
526 mM Tris-HCl/300 mM NaCl pH 8.0 and lysed by 1 pass through a cell disrupter (0.75 kW; Constant  
527 Biosystems) at 23,000 psi, followed by ultracentrifugation at 200,000 × g for 45 minutes. The  
528 supernatant was loaded on nickel-affinity resin (Chelating Sepharose; GE Healthcare) and after  
529 washing with 30 mM imidazole, protein was eluted with buffer containing 250 mM imidazole.  
530 Protein was further purified by gel filtration in 10 mM Hepes/100 mM NaCl pH 7.5 using a HiLoad  
531 16/60 Superdex 200 column.

532

533 **Crystallisation and structure determination of RagAB from W83 KRAB.** Sitting drop vapour  
534 diffusion crystallisation trials were set up using a Mosquito crystallisation robot (TTP Labtech)  
535 using commercially available crystallisation screens (MemGold 1 and 2; Molecular Dimensions).  
536 Initial hits were optimised manually by hanging drop vapour diffusion. Crystals were cryoprotected  
537 by transferring them for 5-10 s in mother liquor containing an additional 10% PEG400. A few  
538 crystals optimised from MemGold 2 condition C8 (18% PEG200, 0.1 M KCl, 0.1 M K-phosphate pH  
539 7.5) diffracted anisotropically to below 4 Å at the Diamond Light Source (DLS) synchrotron at  
540 Didcot, UK (space group C222<sub>1</sub>; cell dimensions ~190 x 377 x 369 Å, with four RagAB complexes  
541 in the asymmetric unit). Data were processed via Xia2<sup>45</sup> or Dials<sup>46</sup>. The structure was solved by  
542 molecular replacement with Phaser<sup>47</sup>, using data to 3.4 Å resolution. The structures of RagB (PDB  
543 5CX8) and a Sculptor-modified model of BT2264 SusC (PDB 5FQ8) were used as search models.  
544 The RagAB model was built iteratively by a combination of manually building in COOT<sup>48</sup> and the

545 AUTOBUILD routine within Phenix<sup>49</sup>, and was refined with Phenix<sup>50</sup> using TLS refinement with 1  
546 group per chain. Given that the R values are reasonable and the R-R<sub>free</sub> gaps are not excessively  
547 large we did not use NCS in the final rounds of refinement. Using NCS also did not improve the  
548 peptide density. The final R and R<sub>free</sub> factors of the RagAB structure are 20.5 and 25.5%,  
549 respectively (Supplementary Table 1). Structures of RagAB purified from wild type W83 (+/- P21)  
550 were solved via molecular replacement using Phaser, using the best-defined RagAB complex as  
551 search model. Structures were refined within Phenix as above (Supplementary Table 1) and  
552 structure validation was carried out with MolProbity<sup>51</sup>.

553

554 **Crystallisation and structure determination of wild type RagAB W83 in the absence and**  
555 **presence of P21.** Crystallisation trials were performed as outlined above with the following  
556 modifications: the protein was not detergent-exchanged after gel filtration in DDM; two other  
557 commercially available crystallisation screens were used (MemChannel and MemTrans; Molecular  
558 Dimensions); For co-crystallisation of RagAB W83 with P21 peptide, RagAB W83 at a  
559 concentration of 17 mg/ml (~0.1 mM) was incubated overnight at 4 °C with 5 mM P21 peptide. For  
560 RagAB W83 the crystals were optimised from MemTrans condition F6 (22% PEG400, 0.07 M  
561 NaCl, 0.05 M Na-citrate pH 4.5), for RagAB W83 + P21 the crystals were optimised from  
562 MemChannel condition D3 (15% PEG1000, 0.05 M Li-sulfate, 0.05 M Na-phosphate monobasic,  
563 0.08 M citrate pH 4.5). Crystals were cryoprotected by transferring them for 5-10 s in mother liquor  
564 containing additional PEG400 to generate a final concentration of ~25%.

565

566 **Microscale thermophoresis.** Initial isothermal titration calorimetry (ITC) experiments showed that  
567 P4 addition to buffer without protein generated very large heats, precluding ITC as a method to  
568 assess peptide binding. For MST, the Monolith NT.115 instrument (NanoTemper Technologies  
569 GmbH, Munich, Germany) was used to analyse the binding interactions between P4, P12 and P21  
570 peptides and RagAB from W83 as well as RagB from W83 and ATCC 33277. Proteins were  
571 labelled with Monolith His-tag Labeling Kit RED-tris-NTA 2nd Generation (NanoTemper  
572 Technologies). For all interactions the concentrations of both fluorescently labelled molecule and  
573 ligand were empirically adjusted using Binding Check mode (MO.Control software, NanoTemper  
574 Technologies). Experiments were performed in assay buffer (10 mM Hepes/100 mM NaCl pH 7.5)  
575 with addition of 0.03% DDM in the case of RagABs. For the measurements, sixteen 1:1 ligand  
576 dilutions were prepared and then mixed with one volume of labelled protein followed by loading  
577 into Monolith NT.115 Capillaries. Initial fluorescence measurements followed by thermophoresis  
578 measurement were carried out using 100% LED power and medium MST power, respectively.  
579 Data for three independently pipetted measurements were analysed (MO.Affinity Analysis  
580 software, NanoTemper Technologies), allowing for determination of apparent dissociation  
581 constants (K<sub>D</sub> app.). The data was presented using GraphPad Prism 8. The interactions between

582 unlabelled W83 and ATCC 33277 RagABs and FAM-labelled P4 peptide were determined in the  
583 same way.

584

585 **Isolation of the outer membrane fractions for quantitation of RagAB.** Outer membrane  
586 fractions from 1 l of culture were isolated using sarkosyl extraction method (see RagAB W83  
587 KRAB production and purification). After extraction in LDAO, the samples in 10 mM Hepes/50 mM  
588 NaCl pH 7.5 containing 1% LDAO were diluted 3 times and loaded on SDS-PAGE. The bands  
589 were analysed quantitatively using Image Lab 6.0.1 software (BIO-RAD). The band at ~70 kDa  
590 was used as a reference sample (loading control). This analysis allows comparison of the levels of  
591 RagAB integrated into the OM and, in combination with the growth assays, a qualitative  
592 assessment of function.

593

594 **Peptide identification by LC-MS.** Bound peptides were isolated by precipitation via addition of  
595 trichloroacetic acid to a final concentration of 30% and incubation at 4 °C for 2 hrs. Subsequently  
596 the peptide-containing supernatants were collected by centrifugation at 17,000xg. The isolated  
597 peptides were micropurified using Empore™ SPE Disks of C18 octadecyl packed in 10 µl pipette  
598 tips.

599

600 LC-MS/MS was performed using an EASY-nLC 1000 system (Thermo Scientific) connected to a  
601 QExactive+ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific). Peptides were  
602 dissolved in 0.1 % formic acid and trapped on a 2 cm ReproSil-Pur C18-AQ column (100 µm inner  
603 diameter, 3 µm resin; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The peptides were  
604 separated on a 15-cm analytical column (75 µm inner diameter) packed in-house in a pulled  
605 emitter with ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH, Ammerbuch-Entringen,  
606 Germany). Peptides were eluted using a flow rate of 250 nl/min and a 20-minute gradient from 5%  
607 to 35% phase B (0.1% formic acid and 90% acetonitrile or 0.1 % formic acid, 90 % acetonitrile and  
608 5% DMSO). The collected MS files were converted to Mascot generic format (MGF) using  
609 Proteome Discoverer (Thermo Scientific).

610

611 The data were searched against the *P. gingivalis* proteome (UniRef at uniprot.org) or the Swiss-  
612 prot database using a bovine taxonomy. Database searches were conducted on a local mascot  
613 search engine. The following settings were used: MS error tolerance of 10 ppm, MS/MS error  
614 tolerance of 0.1 Da, and either non-specific enzyme or trypsin. Based on chromatogram peak  
615 heights and the number of MS/MS spectra observed for a particular peptide in the mass  
616 spectrometry data, the 21-residue peptide KATAEALKKALEEAGAEVELK (henceforth named P21;  
617 charge -1) from ribosomal protein L7 is very abundant in W83 KRAB RagAB.

618

619 **Generation of BSA tryptic mixture.** BSA was solubilized in 100 mM ammonium bicarbonate/8M  
620 urea pH 8.0. DTT was added to a final concentration of 10 mM and the mixture was incubated for  
621 60 min at RT. Protein was alkylated by addition of iodoacetamide to a final concentration of 30 mM  
622 and incubation for 60 min at RT in the dark. Next, the concentration of DTT was adjusted to 35 mM  
623 followed by dilution of the sample to 1 M urea. For cleavage, trypsin from bovine pancreas (Sigma)  
624 was added at 1:25 (trypsin : BSA) mass ratio and the sample was incubated overnight at 37 °C.  
625 Digestion was stopped by acidifying the sample to pH < 2.5 with formic acid. Peptides were  
626 purified using Peptide Desalting Spin Columns (Thermo Scientific) and dried using Speed Vacuum  
627 Concentrator Savant SC210A (Thermo Scientific).

628

629 **Acquisition of BSA-derived peptides and P21 by RagAB and RagB *in vitro*.**

630 W83 RagAB and W83 RagB were incubated overnight at 4 °C with an ~100-fold molar excess of  
631 tryptic BSA digest (see Generation of BSA tryptic mixture) or P21 peptide. The samples were then  
632 ran on Superdex 200 Increase GL 10/300 in 10 mM Hepes/100 mM NaCl pH 7.5 with addition of  
633 0.03% DDM for RagAB W83. Fractions containing protein were pooled, acidified by addition of  
634 formic acid to a final concentration 0.1% and analysed by MS.

635

636 **qPCR.** Samples of 1 ml of bacterial cultures ( $OD_{600} = 1.0$ ) were centrifuged (5,000 x g; 5 min) at 4  
637 °C, pellets were resuspended in 1 ml Tri Reagent (Ambion), incubated at 60 °C for 20 minutes,  
638 cooled to room temperature and total RNA was isolated according to manufacturer instructions.  
639 Genomic DNA was removed from samples by digestion with DNase I (Ambion); 2 µg of RNA was  
640 incubated with 2 U of DNase for 60 minutes at 37 °C. Following digestion, RNA was purified using  
641 Tri Reagent. Reverse transcription of 50 ng of RNA was performed with High-Capacity cDNA  
642 Reverse Transcription Kit (Life Technologies), and reaction mixture was then diluted 20 times.  
643 Real-time PCR was done in 10 µl reaction volume, using KAPA SYBR FAST qPCR Master Mix  
644 (Kapa Biosystems) with 2 µl of diluted reverse transcription mixture as template. Primers used are  
645 listed in Table S4. Reaction conditions were 3 minutes at 95 °C, followed by 40 cycles of  
646 denaturation for 3 seconds at 95 °C and annealing/extension for 20 seconds at 60 °C. The reaction  
647 was carried out with a CFX96 thermal cycler (Bio-Rad), and data was analysed in Bio-Rad CFX  
648 Manager software.

649

650 **CryoEM sample preparation and data collection.** A sample of purified RagAB solubilised in a  
651 DDM-containing buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 0.03 % DDM) was prepared at 1.75  
652 mg/ml (principle dataset) or 3 mg/ml (P21 addition experiment). For the P21 addition experiment,  
653 control and P21-doped grids (with 50-fold molar excess of P21 peptide) were prepared at the same  
654 time from the same purified stock of RagAB for consistency. In all cases, a 3.5 µL aliquot was  
655 applied to holey carbon grids (Quantifoil 300 mesh, R1.2/1.3), which had been glow discharged at  
656 10 mA for 30 s before sample application. Blotting and plunge freezing were carried out using a

657 Vitrobot Mark IV (FEI) with chamber temperature set to 6 °C and 100 % relative humidity. A blot  
658 force of 6 and a blot time of 6 s were used prior to vitrification in liquid nitrogen-cooled liquid  
659 ethane.

660

661 Micrograph movies were collected on a Titan Krios microscope (Thermo Fisher) operating at 300  
662 kV with a GIF energy filter (Gatan) and K2 summit direct electron detector (Gatan) operating in  
663 counting mode. Data acquisition parameters for each data set can be found in Supplementary  
664 Table 3.

665

666 **Image processing.** Image processing was carried out using RELION (v2.1 and v3.0)<sup>52,53</sup>. Drift  
667 correction was performed using MotionCor2<sup>54</sup> and contrast transfer functions were estimated using  
668 gCTF<sup>55</sup>. Micrographs with estimated resolutions poorer than 5 Å and defocus values >4 µm were  
669 discarded using a python script<sup>56</sup>. For the principle RagAB dataset, particles were autopicked in  
670 RELION using a gaussian blob with a peak value of 0.3. Control and experimental datasets for the  
671 P21 addition experiment were autopicked using the 'general model' in crYOLO<sup>57</sup>. In both cases,  
672 particles were extracted in 216 x 216 pixel boxes and subjected to several rounds of 2D  
673 classification in RELION<sup>52</sup>. 3D starting models were generated *de novo* from the EM data by  
674 stochastic gradient descent in RELION. Processing of RagAB control and P21-doped datasets was  
675 only taken as far as 3D classification.

676

677 For the principle RagAB dataset, three conformational states representing the CC, OC and OO  
678 states were apparent in the first round of 3D classification and the corresponding particle stacks  
679 were treated independently in further processing. C2 symmetry was applied to both the CC and  
680 OO reconstructions. Post-processing was performed using soft masks and yielded reconstructions  
681 for the CC, OC and OO states with resolutions of 3.7 Å, 3.7 Å and 3.9 Å respectively, as estimated  
682 by gold standard Fourier Shell Correlations using the 0.143 criterion. The original micrograph  
683 movies were later motion corrected in RELION 3.0<sup>53</sup>. Particles contributing to the final  
684 reconstructions were re-extracted from the resulting micrographs. Following reconstruction,  
685 iterative rounds of per-particle CTF refinement, with beam tilt estimation, and Bayesian particle  
686 polishing were employed which improved the resolution of post-processed CC, OC and OO maps  
687 to 3.3 Å, 3.3 Å and 3.4 Å respectively.

688

689 **Model building into cryoEM maps.** Examination of the maps revealed that their handedness was  
690 incorrect. Maps were therefore Z-flipped in UCSF Chimera<sup>58</sup>. The RagAB W83 KRAB crystal  
691 structure was rigid-body fit to the CC density map and subjected to several iterations of manual  
692 refinement in COOT and 'real space refinement' in Phenix<sup>59</sup>. The asymmetric unit was  
693 symmetrised in Chimera after each iteration. Starting models for the OC and OO states of the  
694 complex were obtained from the CC structure by rigid-body fitting of one or both RagB subunits to

695 their cognate open density in the OC and OO maps respectively. These too were subjected to  
696 several iterations of manual refinement in COOT and 'real space refinement' in Phenix. Molprobity  
697 was used for model validation<sup>51</sup>.

698

699 **Molecular Dynamics simulations.** The bound peptide in the refined crystal structure of a RagAB  
700 "monomer" was removed *in silico* and the system was inserted into a palmitoyloleoyl-  
701 phosphatidylethanolamine (POPE) bilayer using the CHARMM-GUI Membrane Builder<sup>59</sup>. The  
702 systems were solvated using a TIP3P water box and neutralized by adding the required counter-  
703 ions. Simulations were performed using GROMACS 5.1.2<sup>60</sup> and the all-atom CHARMM36 force  
704 fields<sup>61,62</sup>. For the long-range Coulomb interactions, the particle-mesh Ewald (PME) summation  
705 method<sup>63</sup> has been employed with a short-range cutoff of 12 Å and a Fourier grid spacing of 0.12  
706 nm. In addition, the Lennard-Jones interactions were considered up to a distance of 10 Å and a  
707 switch function was used to turn off interactions smoothly at 12 Å. Achieved by semi-isotropic  
708 coupling to a Parrinello-Rahman barostat<sup>64</sup> at 1 bar with a coupling constant of 5 ps, the final  
709 unbiased simulations were performed in the isothermal-isobaric (NPT) ensemble. A Nosé-Hoover  
710 thermostat<sup>65,66</sup> was used to keep the temperature at 300 K with a coupling constant of 1 ps. A total  
711 of three simulations of 2500 ns were carried out for the apo complex with a time step of 2 fs by  
712 applying constraints on hydrogen atom bonds using the LINCS algorithm<sup>67</sup>. Similarly, the dimeric  
713 complex of RagA<sub>2</sub>B<sub>2</sub> in the OO state was simulated for 500 ns. Cavity calculations were performed  
714 using CASTp<sup>68</sup>.

715

716

717

718

## 719 **Data availability**

720 The data supporting the findings of this study are available from the corresponding authors upon  
721 reasonable request. Coordinates and structure factors that support the findings of this study have  
722 been deposited in the Protein Data Bank with accession codes 6SLI (KRAB RagAB), 6SLJ (WT  
723 RagAB) and 6SLN (WT RagAB + P21). EM structure coordinates have been deposited in the  
724 Electron Microscopy Data Bank with accession codes 6SM3 (CC), 6SMQ (OC) and 6SML (OO).  
725 The raw cryoEM movie mode micrographs for the primary dataset containing the CC, OC and OO  
726 structures, will be deposited in the EMPIAR database. Source data for Figs. 1a, 4a-d and  
727 Extended Data Figs. 2, 6a,b and 7 are included in this article and its Supplementary Information  
728 files.

729

## 730 **References**

- 731 1. Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C. & Kent, R. L. Jr. Microbial  
732 complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–44 (1998).
- 733 2. Eke, P. I. et al. Update on prevalence of periodontitis in adults in the United States:  
734 NHANES 2009 to 2012. *J. Periodontol.* **86**, 611–22 (2015).
- 735 3. Dominy, S. S. et al. *Porphyromonas gingivalis* in Alzheimer's disease brains: evidence for  
736 disease causation and treatment with small-molecule inhibitors. *Sci. Adv.* **5**, eaau3333  
737 (2019).
- 738 4. Potempa, J., Mydel, P. & Koziel, J. The case for periodontitis in the pathogenesis of  
739 rheumatoid arthritis. *Nat. Rev. Rheumatol.* **13**, 606–620 (2017).
- 740 5. Tonetti, M. S. et al. Treatment of periodontitis and endothelial function. *N. Engl. J. Med.*  
741 **356**, 911–920 (2007).
- 742 6. Usher, A. K. & Stockley, R. A. The link between chronic periodontitis and COPD: a  
743 common role for the neutrophil? *BMC Med.* **11**, 241 (2013).
- 744 7. Bui, F. Q. et al. Association between periodontal pathogens and systemic disease. *Biomed*  
745 *J.* **42**, 27–35 (2019).
- 746 8. Mayrand, D. & Holt, S. C. Biology of asaccharolytic black-pigmented *Bacteroides* species.  
747 *Microbiol. Rev.* **52**, 134–152 (1988).
- 748 9. Nemoto, T. K., Ohara-Nemoto, Y., Bezerra, G. A., Shimoyama, Y. & Kimura, S. A.  
749 *Porphyromonas gingivalis* periplasmic novel exopeptidase, acylpeptidyl oligopeptidase,  
750 releases N-acylated di- and tripeptides from oligopeptides. *J. Biol. Chem.* **291**, 5913–5925  
751 (2016).
- 752 10. Potempa, J., Banbula, A. & Travis, J. Role of bacterial proteinases in matrix destruction and  
753 modulation of host responses. *Periodontol. 2000* **24**, 153–192 (2000).
- 754 11. Grenier, D. et al. Role of gingipains in growth of *Porphyromonas gingivalis* in the presence  
755 of human serum albumin. *Infect. Immun.* **69**, 5166–5172 (2001).
- 756 12. Nagano, K. et al. Characterization of RagA and RagB in *Porphyromonas gingivalis*: study  
757 using gene-deletion mutants. *J. Med. Microbiol.* **56**, 1536–1548 (2007).
- 758 13. Goulas, T. et al. Structure of RagB, a major immunodominant outer-membrane surface  
759 receptor antigen of *Porphyromonas gingivalis*. *Mol. Oral Microbiol.* **31**, 472–485 (2016).
- 760 14. Bolam, D. N. & van den Berg, B. TonB-dependent transport by the gut microbiota: novel  
761 aspects of an old problem. *Curr. Opin. Struct. Biol.* **51**, 35–43 (2018).
- 762 15. Bolam, D. N. & Koropatkin, N. M. Glycan recognition by the Bacteroidetes Sus-like  
763 systems. *Curr. Opin. Struct. Biol.* **22**, 563–569 (2012).
- 764 16. Glenwright, A. J. et al. Structural basis for nutrient acquisition by dominant members of the  
765 human gut microbiota. *Nature* **541**, 407–411 (2017).
- 766 17. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J.*  
767 *Mol. Biol.* **372**, 774–797 (2007).

- 768 18. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters:  
769 regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43-60 (2010).
- 770 19. Freed, D. M., Horanyi, P. S., Wiener, M. C. & Cafiso, D. S. Conformational exchange in a  
771 membrane transport protein is altered in protein crystals. *Biophys. J.* **99**, 604-10 (2010).
- 772 20. Hickman, S. J., Cooper, R. E. M., Bellucci, L., Paci, E. & Brockwell, D. J. Gating of TonB-  
773 dependent transporters by substrate-specific forced remodelling. *Nat. Commun.* **8**, 14804  
774 (2017).
- 775 21. Gómez-Santos, N., Glatter, T., Koebnik, R., Świątek-Połatyńska, M. A. & Søgaard-  
776 Andersen, L. A. TonB-dependent transporter is required for secretion of protease PopC  
777 across the bacterial outer membrane. *Nat. Commun.* **10**, 1360 (2019).
- 778 22. Hall, L. M. et al. Sequence diversity and antigenic variation at the rag locus of  
779 *Porphyromonas gingivalis*. *Infect. Immun.* **73**, 4253–4262. (2005).
- 780 23. Milner, P., Batten, J. E. & Curtis M. A. Development of a simple chemically defined medium  
781 for *Porphyromonas gingivalis*: requirement for alpha-ketoglutarate. *FEMS Microbiol. Lett.*  
782 **140**, 125-30 (1996).
- 783 24. Grenier, D. et al. Role of gingipains in growth of *Porphyromonas gingivalis* in the presence  
784 of human serum albumin. *Infect. Immun.* **69**, 5166-72 (2001).
- 785 25. Nagano, K. et al. Trimeric structure of major outer membrane proteins homologous to  
786 OmpA in *Porphyromonas gingivalis*. *J. Bacteriol.* **187**, 902-11 (2005).
- 787 26. Lamster, I. B. & and Ahlo, J. K. Analysis of gingival crevicular fluid as applied to the  
788 diagnosis of oral and systemic diseases. *Ann. N. Y. Acad. Sci.* **1098**, 216–29 (2007).
- 789 27. Curtis, M. A., Hanley, S. A. & Aduse-Opoku, J. The rag locus of *Porphyromonas gingivalis*:  
790 a novel pathogenicity island. *J. Periodontal Res.* **34**, 400–405 (1999).
- 791 28. Josts, I., Veith, K. & Tidow, H. Ternary structure of the outer membrane transporter FoxA  
792 with resolved signalling domain provides insights into TonB-mediated siderophore uptake.  
793 *Elife* **8**, pii: e48528 (2019).
- 794 29. Kim, I., Stiefel, A., Plantör, S., Angerer, A. & Braun, V. Transcription induction of the ferric  
795 citrate transport genes via the N-terminus of the FecA outer membrane protein, the Ton  
796 system and the electrochemical potential of the cytoplasmic membrane. *Mol. Microbiol.* **23**,  
797 333-44 (1997).
- 798 30. Braun, V., Mahren, S. & Ogierman, M. Regulation of the FecI-type ECF sigma factor by  
799 transmembrane signalling. *Curr. Opin. Microbiol.* **6**, 173-80 (2003).
- 800 31. Koebnik R. TonB-dependent trans-envelope signalling: the exception or the rule? *Trends*  
801 *Microbiol.* **13**, 343-7 (2005).
- 802 32. Ho, M. H., Lamont, R. J. & Xie, H. Identification of *Streptococcus cristatus* peptides that  
803 repress expression of virulence genes in *Porphyromonas gingivalis*. *Sci. Rep.* **7**, 1413  
804 (2017).



- 805 33. Moynié, L. et al. The complex of ferric-enterobactin with its transporter from *Pseudomonas*  
806 *aeruginosa* suggests a two-site model. *Nat. Commun.* **10**, 3673 (2019).
- 807 34. Balhesteros, H. et al. TonB-Dependent Heme/Hemoglobin Utilization by *Caulobacter*  
808 *crescentus* HutA. *J. Bacteriol.* **199**, e00723-16 (2017).
- 809 35. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a  
810 prominent human gut symbiont is directed by the recognition of amylose helices. *Structure*  
811 **16**, 1105-15 (2008).
- 812 36. Rogowski, A., et al. Glycan complexity dictates microbial resource allocation in the large  
813 intestine. *Nat Commun.* **6**, 7481 (2015).
- 814 37. Nguyen, K. A., Travis, J. & Potempa, J. Does the importance of the C-terminal residues in  
815 the maturation of RgpB from *Porphyromonas gingivalis* reveal a novel mechanism for  
816 protein export in a subgroup of Gram-Negative bacteria? *J. Bacteriol.* **189**, 833–843 (2007)
- 817 38. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred  
818 kilobases. *Nat. Methods* **6**, 343–345 (2009).
- 819 39. Chiu, J., March, P. E., Lee, R. & Tillett, D. Site-directed, Ligase-Independent Mutagenesis  
820 (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.*  
821 **32**, e174 (2004).
- 822 40. Tagawa, J. et al. Development of a novel plasmid vector pTIO-1 adapted for  
823 electrotransformation of *Porphyromonas gingivalis*. *J. Microbiol. Methods.* **105**, 174–179  
824 (2014).
- 825 41. Belanger, M., Rodrigues, P. & Progulsk-Fox, A. Genetic manipulation of *Porphyromonas*  
826 *gingivalis*. *Curr. Protoc. Microbiol.* **5**, 13C.2.1–13C.2.24 (2007).
- 827 42. Smith, C. J. Genetic transformation of *Bacteroides* spp. using electroporation. *Methods Mol.*  
828 *Biol.* **47**, 161–169 (1995).
- 829 43. Filip, C., Fletcher, G., Wulff, J. L. & Earhart, C. F. Solubilization of the cytoplasmic  
830 membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. *J. Bacteriol.*  
831 **115**, 717–722 (1973).
- 832 44. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and  
833 high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol.* **177**,  
834 4121–4130 (1995).
- 835 45. Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol.*  
836 *Crystallogr.* **69**, 1260–73 (2013).
- 837 46. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta*  
838 *Crystallogr. D Struct. Biol.* **74**, 85–97 (2018).
- 839 47. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674  
840 (2007).
- 841 48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*  
842 *Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

- 843 49. Afonine, P. V. et al. Towards automated crystallographic structure refinement with  
844 phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
- 845 50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular  
846 structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- 847 51. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular  
848 crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
- 849 52. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure  
850 determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- 851 53. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination  
852 in RELION-3. *Elife* **7**, e42166 (2018).
- 853 54. Li, X., Zheng, S. Q., Egami, K., Agard, D. A. & Cheng, Y. Influence of electron dose rate on  
854 electron counting images recorded with the K2 camera. *J. Struct. Biol.* **184**, 251–260  
855 (2013).
- 856 55. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12  
857 (2016).
- 858 56. Thompson, R. F., Adnan, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection,  
859 pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-  
860 electron microscopy. *Nat. Protoc.* **14**, 100–118 (2019).
- 861 57. Wagner T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for  
862 cryo-EM. *Commun. Biol.* **2**, 218 (2019).
- 863 58. Pettersen, E. F. et al. USCF Chimera – A visualization system for exploratory research and  
864 analysis. *J. Comput. Chem.* **25**, 1605–1612. (2004).
- 865 59. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface  
866 for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
- 867 60. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-  
868 level parallelism from laptops to supercomputers. *SoftwareX.* **1**, 19–25 (2015).
- 869 61. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting  
870 improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles.  
871 *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- 872 62. Klauda, J. B. et al. Update of the CHARMM all-atom additive force field for lipids: validation  
873 on six lipid types. *J. Phys. Chem. B.* **114**, 7830–7843 (2010).
- 874 63. Darden, T., York, D. & Pedersen, L. Particle mesh ewald: an N log (N) method for ewald  
875 sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- 876 64. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular  
877 dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
- 878 65. Nosé, S. A. Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol.*  
879 *Phys.* **52**, 255–268 (1984).

- 880 66. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.*  
881 **31**, 1695 (1985).
- 882 67. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint  
883 solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472. (1997).
- 884 68. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface  
885 topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
- 886 69. Schrödinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.8.

887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903

## 904 **Figure legends**

905

906 **Figure 1 Crystal structure of the RagA<sub>2</sub>B<sub>2</sub> transporter suggests bound peptides.** **a**, SDS-  
907 PAGE gel showing purified RagAB from W83 KRAB before (left lane) and after boiling in SDS-  
908 PAGE sample buffer. The gel is representative of three independent purifications. **b,c**, Views of  
909 RagA<sub>2</sub>B<sub>2</sub> from the plane of the OM (**b**) and from the extracellular space (**c**). **d**, Side views of RagAB  
910 (right panel; surface representation) showing the bound peptide as a red space filling model. The  
911 plug domain of the RagA TBDT (cyan) is dark blue. **e**, 2Fo-Fc density (blue mesh; 1.0  $\sigma$ , carve =  
912 1.8) of the peptide after final refinement (left panel). The right panel shows potential hydrogen  
913 bonds (distance < 3.6 Å) between RagAB and the peptide backbone. Arbitrary peptide residue

914 numbering is indicated in italics. **f**, Extracellular views of RagAB with (left panel) and without the  
915 RagB lid (green). **g**, Side view (top panel) and close-up of the RagB acidic loop (blue) and the  
916 bound peptide. Potential hydrogen bonds are indicated by dashed lines. Structural figures were  
917 made with Pymol<sup>69</sup>.

918

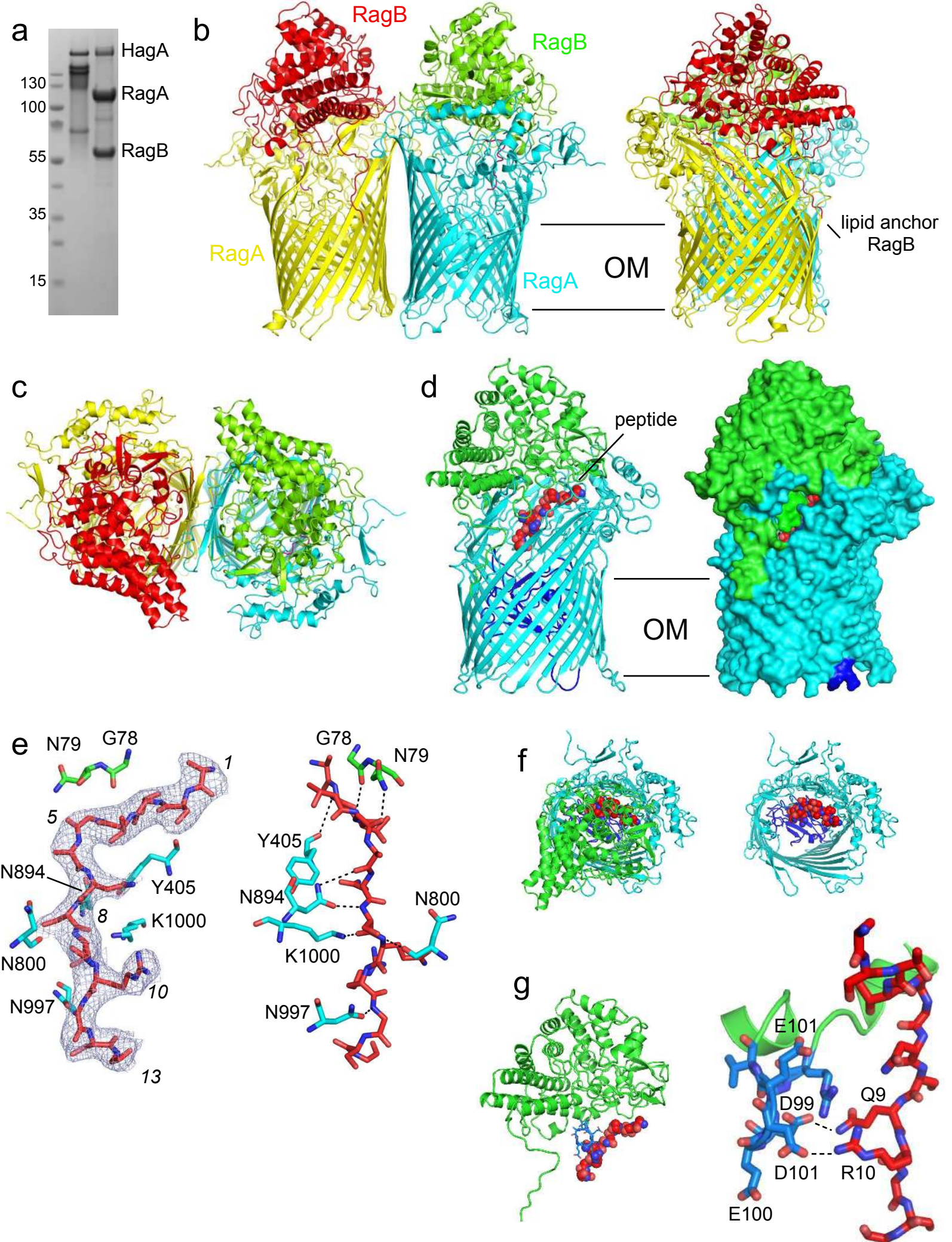
919 **Figure 2 Different conformations of the RagA<sub>2</sub>B<sub>2</sub> transporter revealed by cryo-EM. a-c**,  
920 Cartoon views for closed-closed RagA<sub>2</sub>B<sub>2</sub> (CC; **a**), open-closed (OC; **b**) and open-open (OO; **c**).  
921 Views are from the plane of the OM (left and middle panels) and from the extracellular space (right  
922 panel). Bound peptide is shown as space filling models in magenta. The plug domains of RagA are  
923 coloured dark blue.

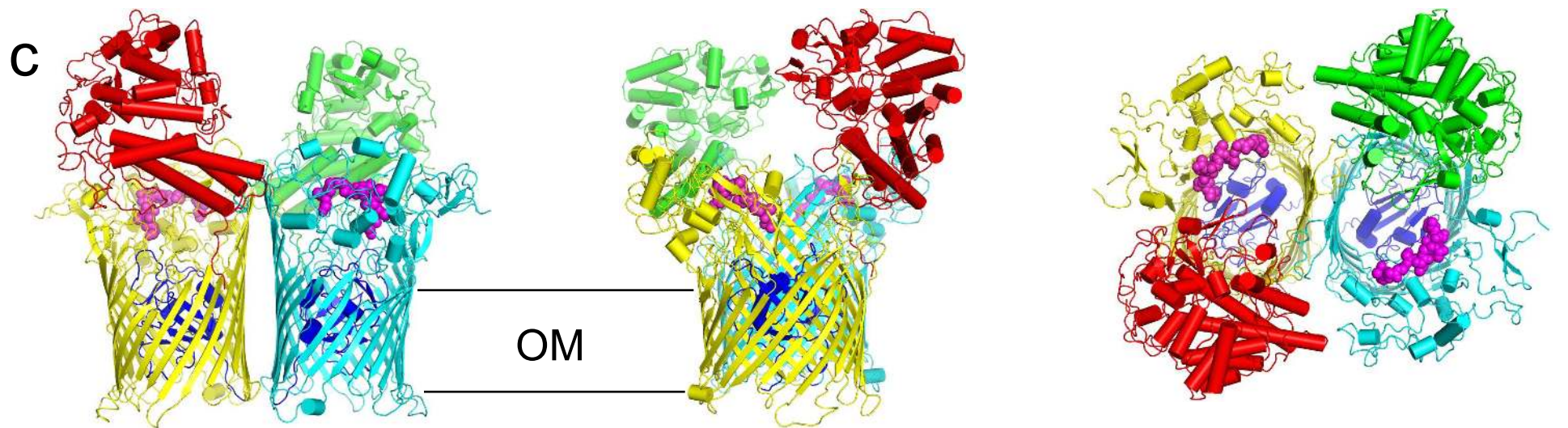
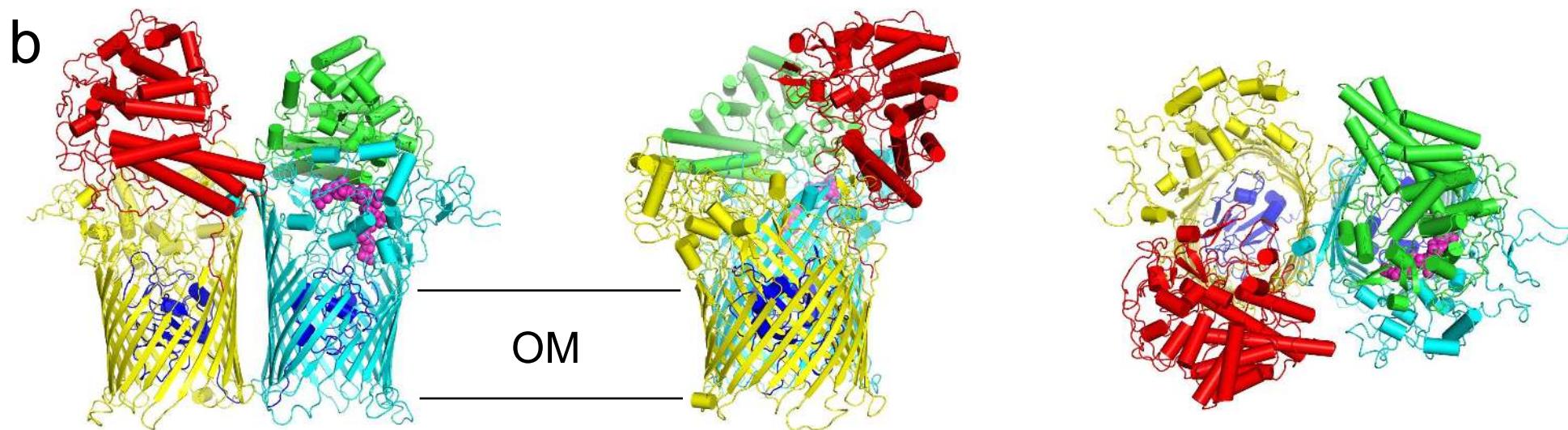
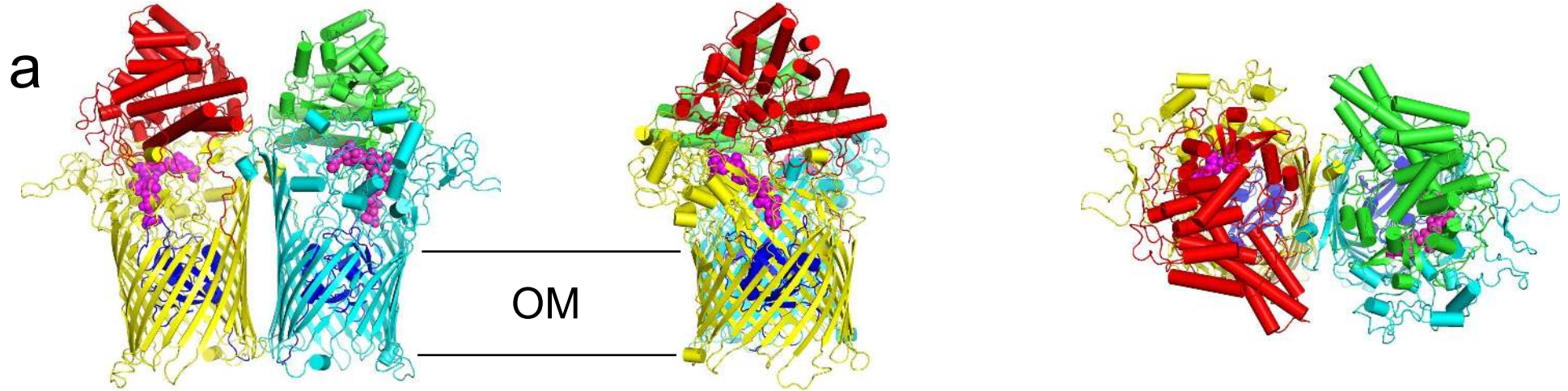
924

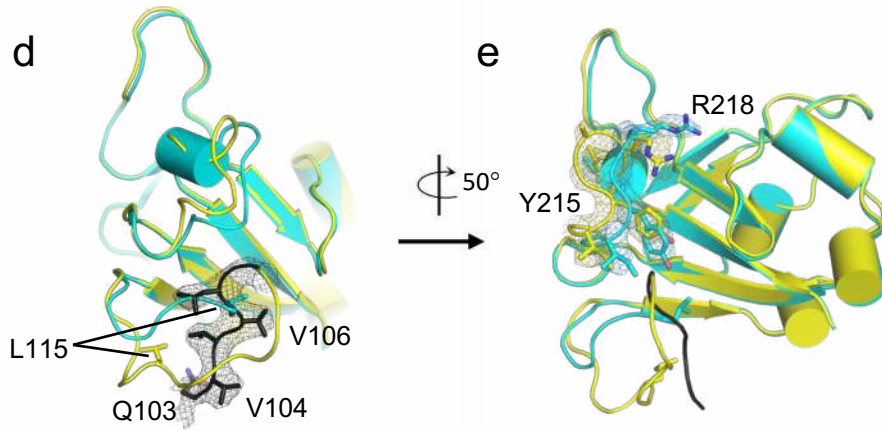
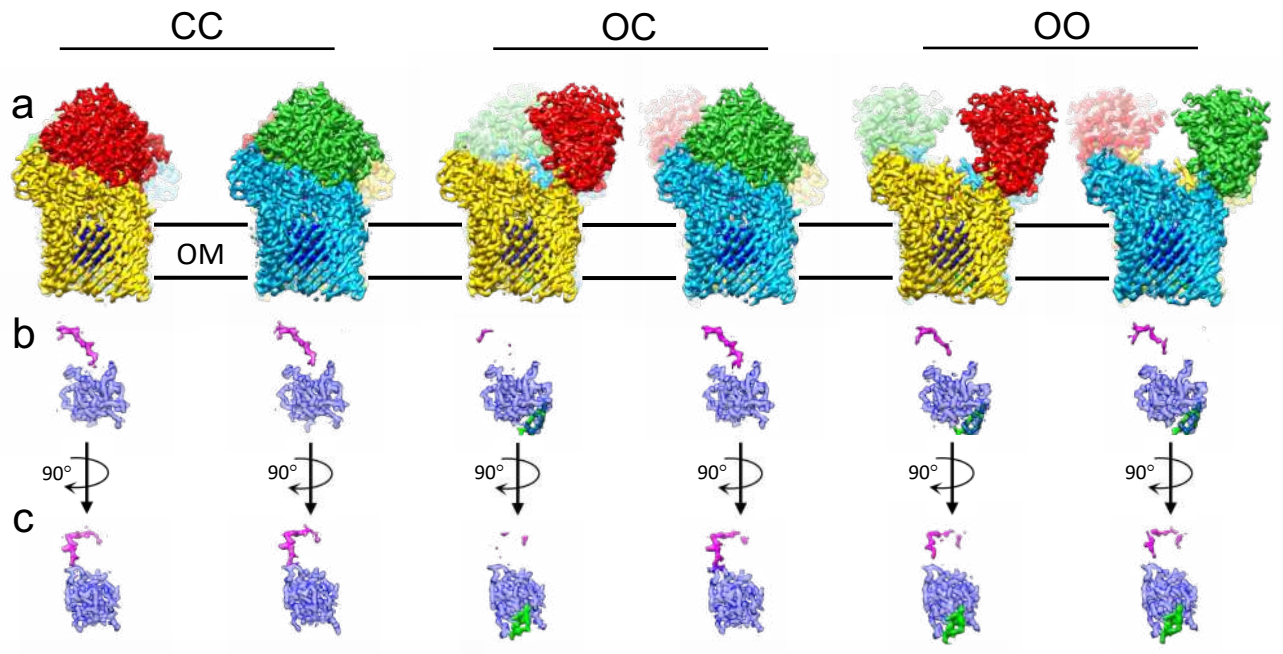
925 **Figure 3 Peptide and plug dynamics in RagA<sub>2</sub>B<sub>2</sub> revealed via cryo-EM. a**, Density maps  
926 viewed from the membrane plane of the closed-closed, open-closed and open-open RagA<sub>2</sub>B<sub>2</sub>  
927 complexes coloured as in Figure 2. **b, c**. Isolated density for the plug domain and the bound  
928 peptide ensemble viewed as in (**a**) and after a 90 degree clockwise rotation (**c**). Additional density  
929 for the plug domains of open RagAB complexes is shown in green. **d,e**, Superposition of RagA  
930 plug domains in the open (yellow) and closed (cyan) RagAB complexes of the OC state, showing  
931 different dynamics (**d**; view approximately as in **c**) and conformational changes (**e**) in going from  
932 the open to the closed state. The Ton box region Q103-G108, visible only in the open state, is  
933 coloured black with electron density shown in (**d**). In **e**, density for open and closed complexes is  
934 shown for the region I214-R218.

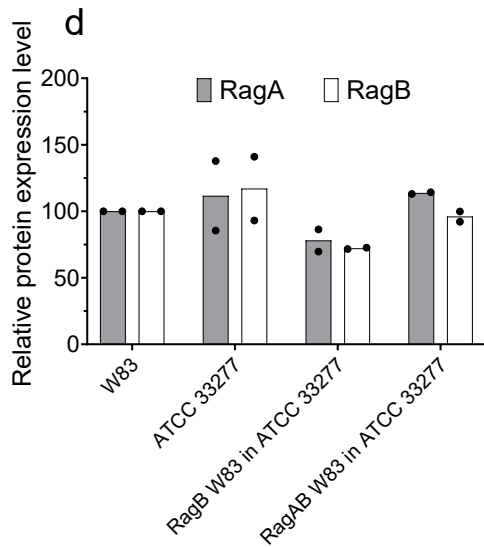
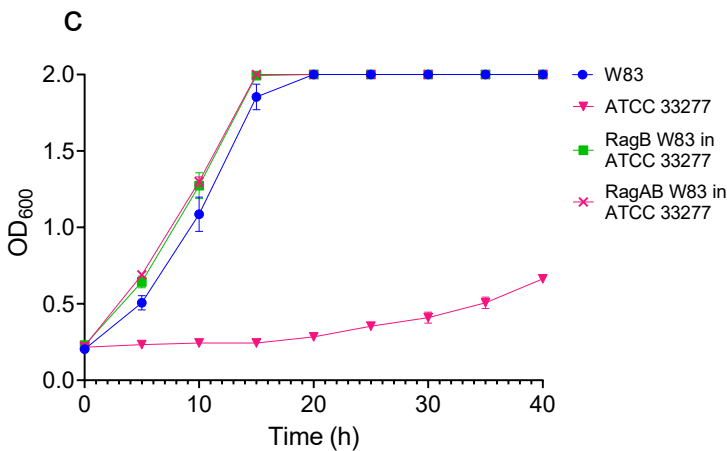
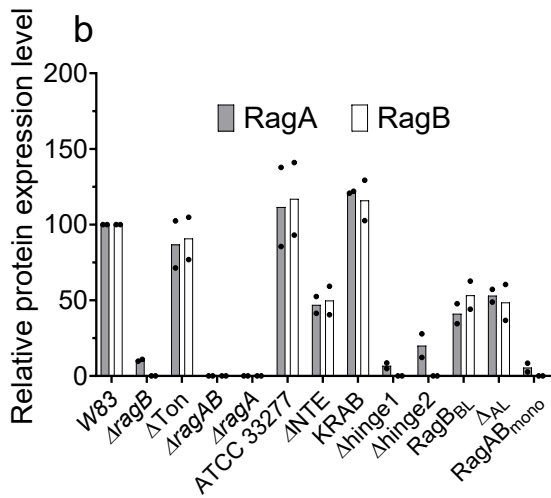
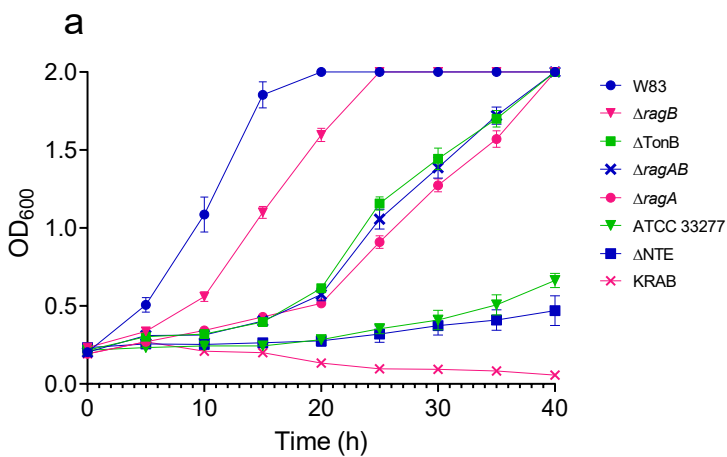
935

936 **Figure 4 RagAB is important for growth on BSA. a**, Representative growth curves (n = 3, mean  
937 ± standard error of the mean) for different *P. gingivalis* W83 and ATCC strains and mutant W83  
938 *ragAB* variants on BSA-MM. **b**, Outer membrane RagAB protein expression levels for strains and  
939 mutants determined by SDS-PAGE densitometry. **c**, Representative growth curves (n = 3, mean ±  
940 standard error of the mean) for growth on BSA-MM for wild type *P. gingivalis* W83 and ATCC and  
941 strains in which RagAB or RagB from ATCC was replaced with the corresponding ortholog from  
942 W83. **d**, Corresponding OM protein expression levels of RagAB. Graphs in **b** and **d** show the mean  
943 of two independent replicates. The dots show the individual replicates.

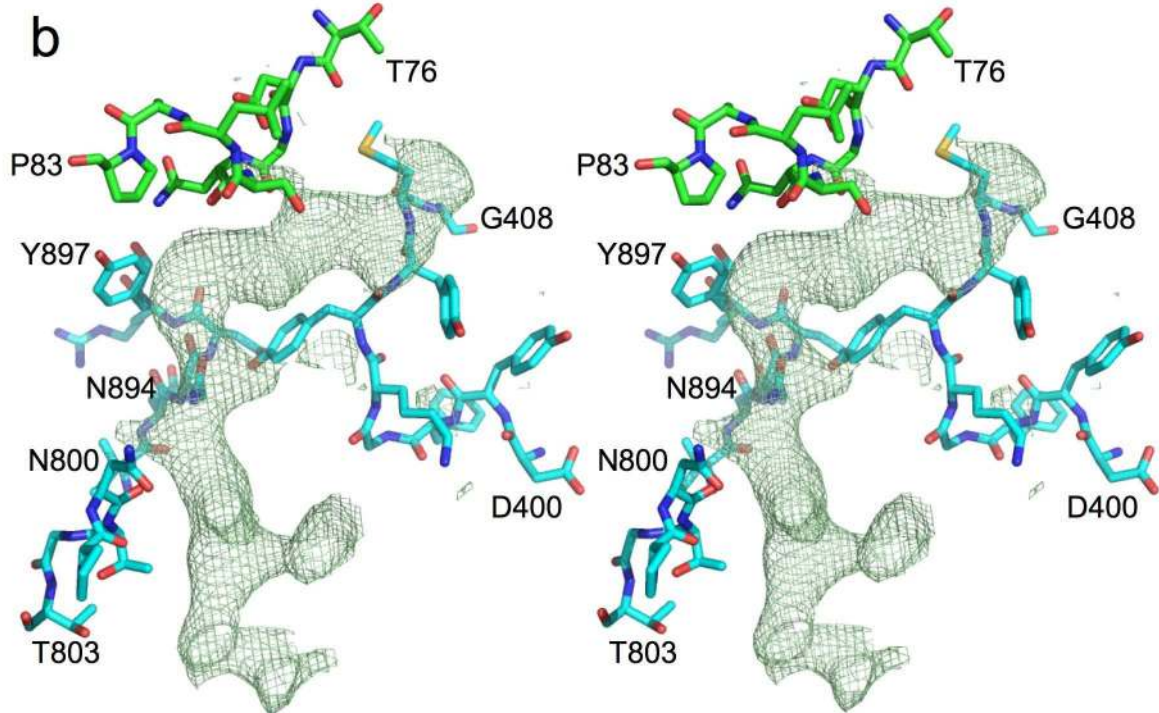
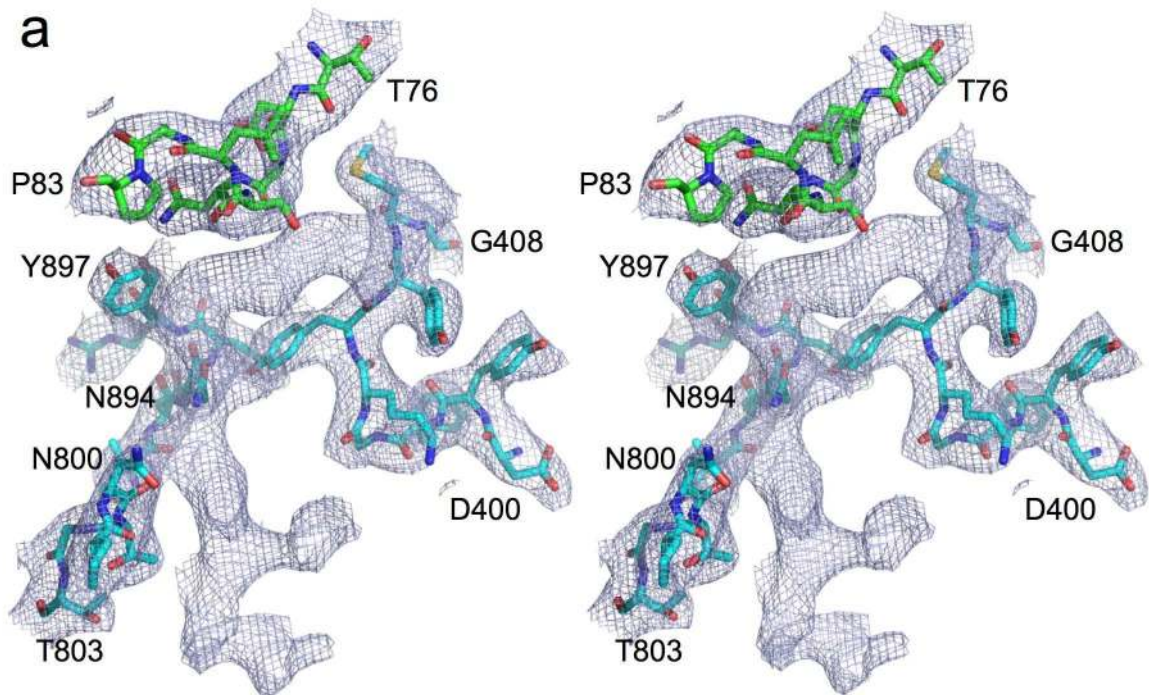


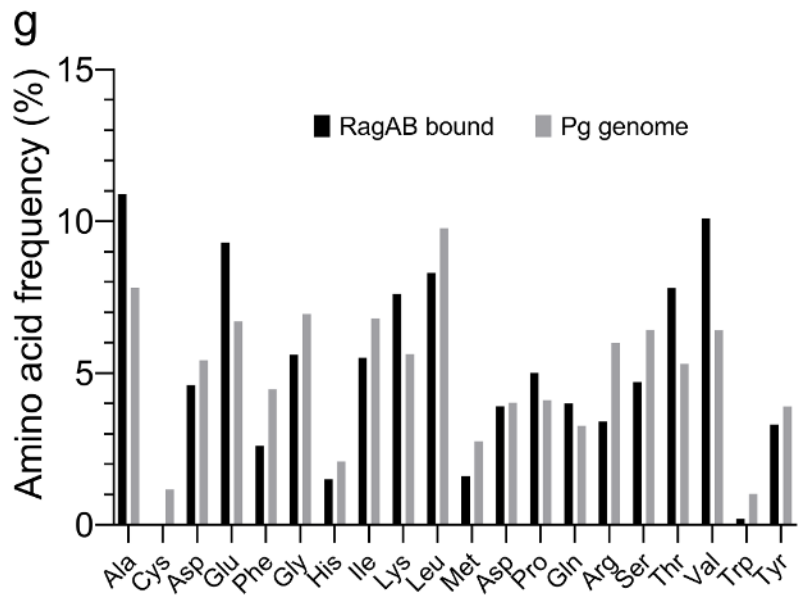
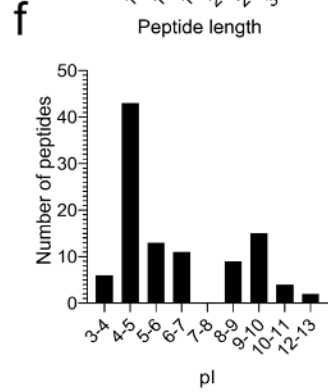
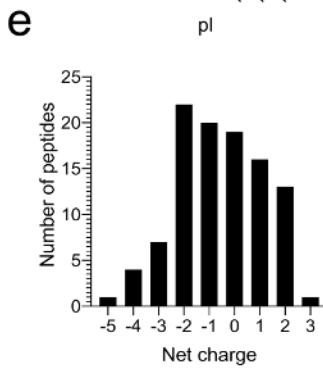
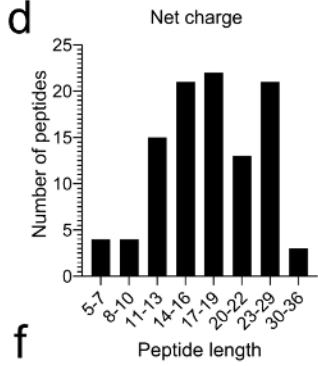
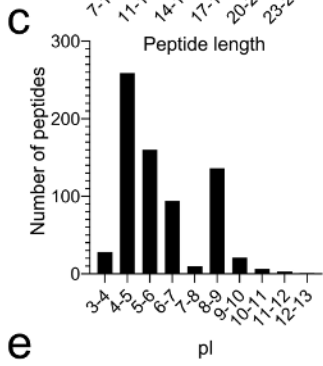
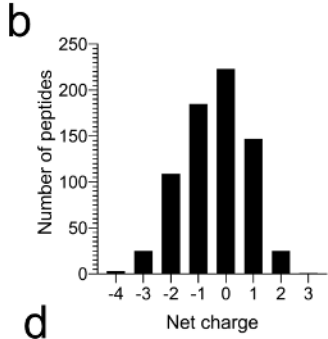
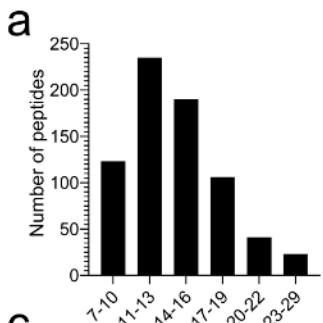


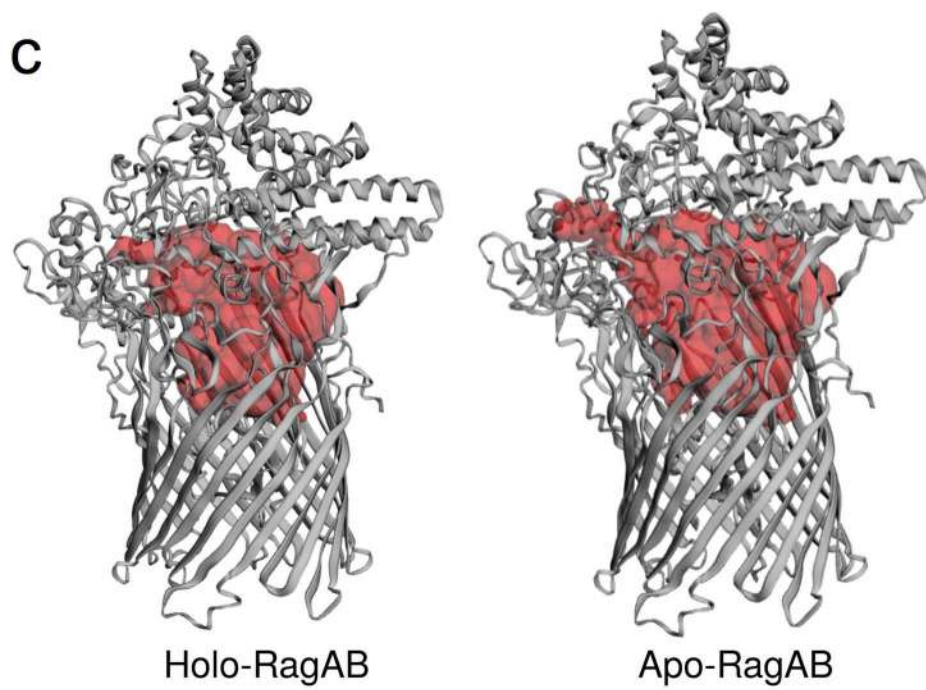
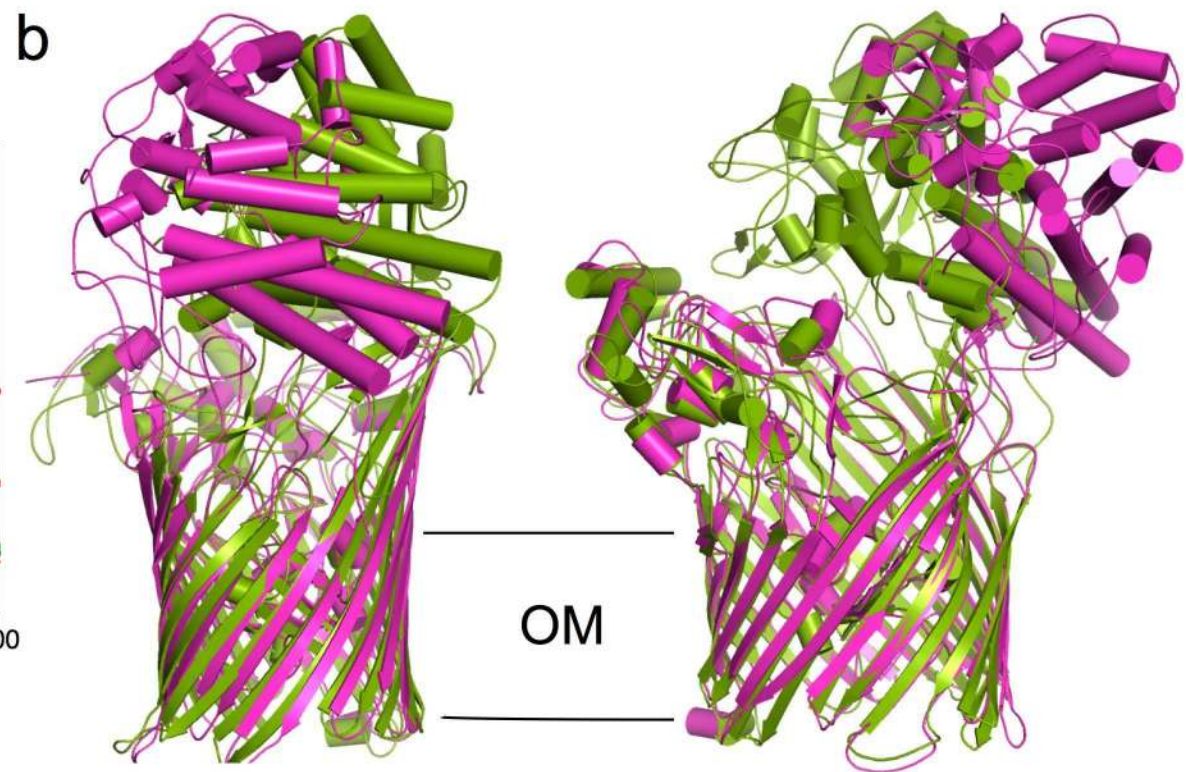
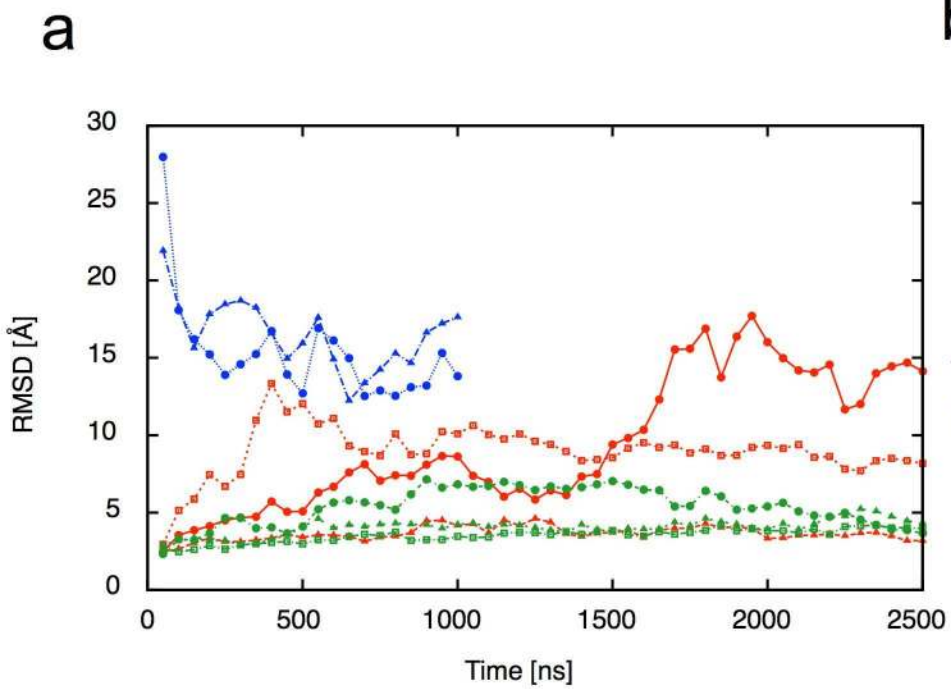


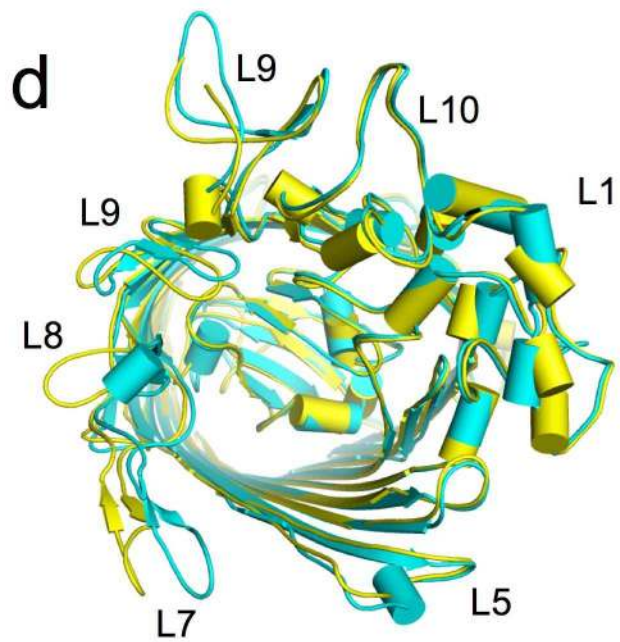
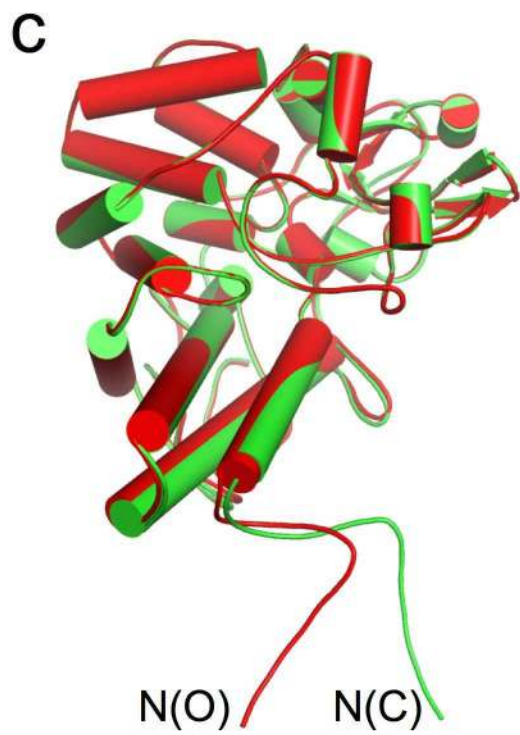
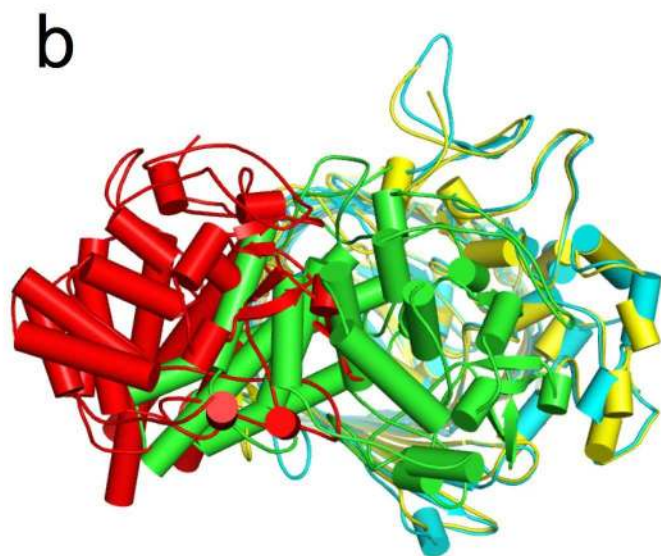
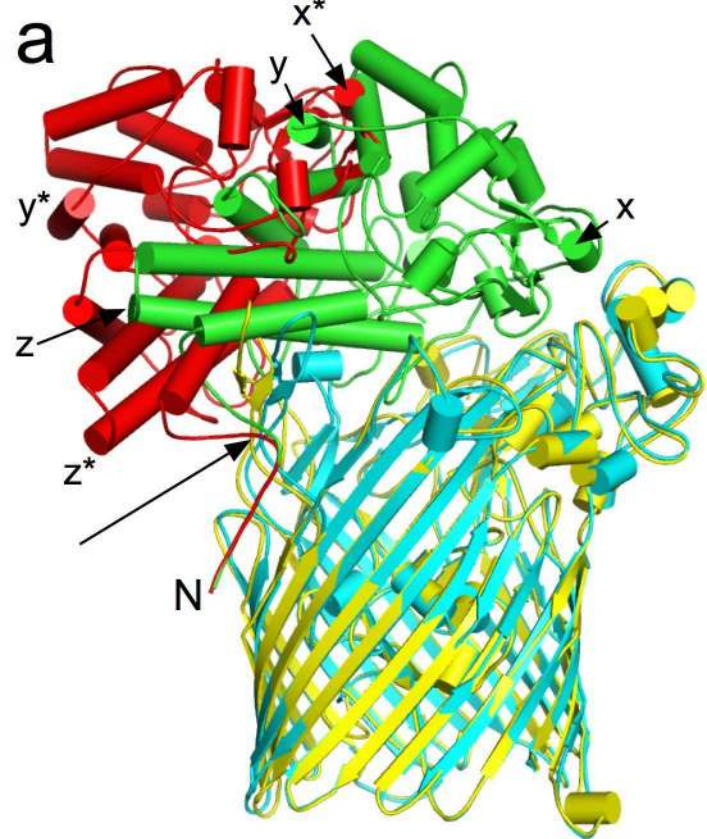


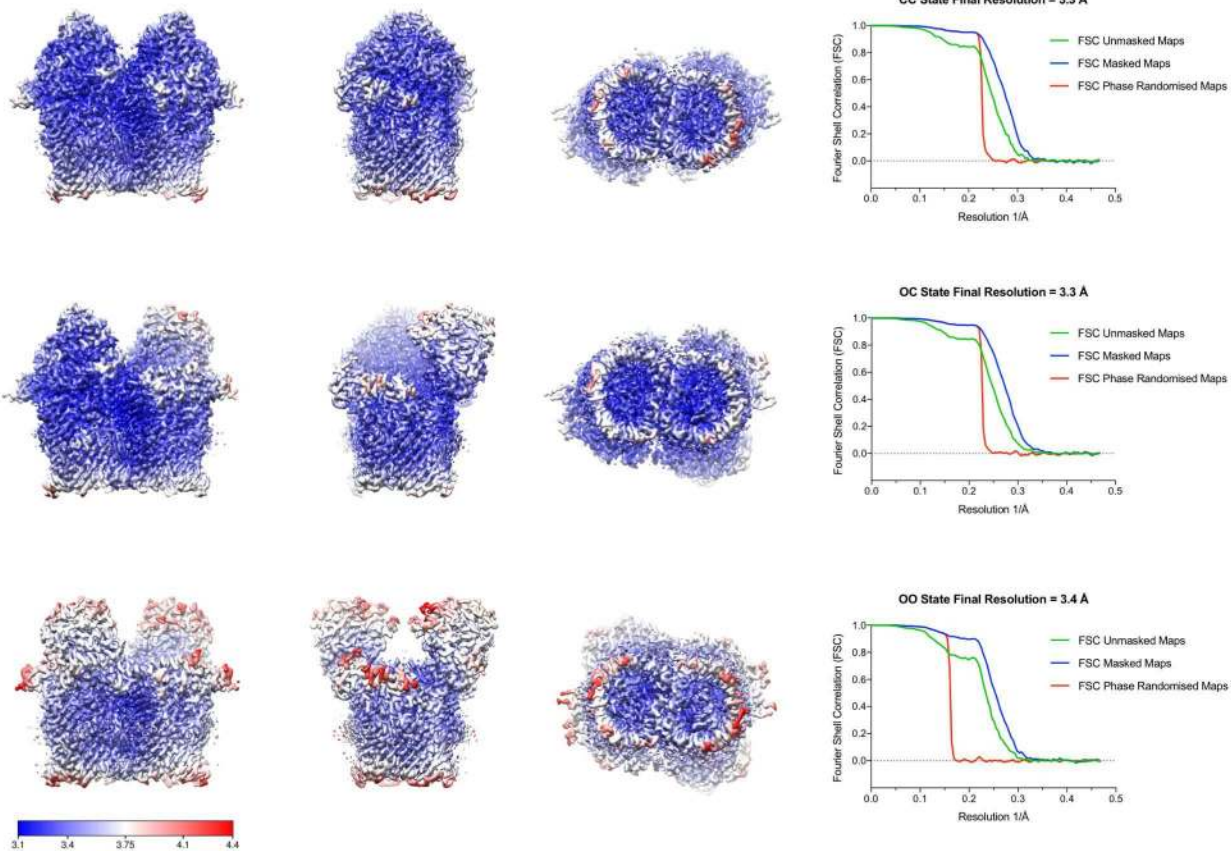
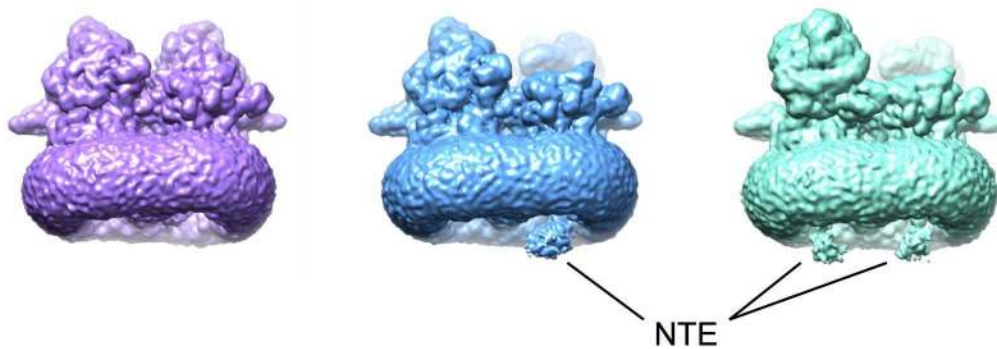


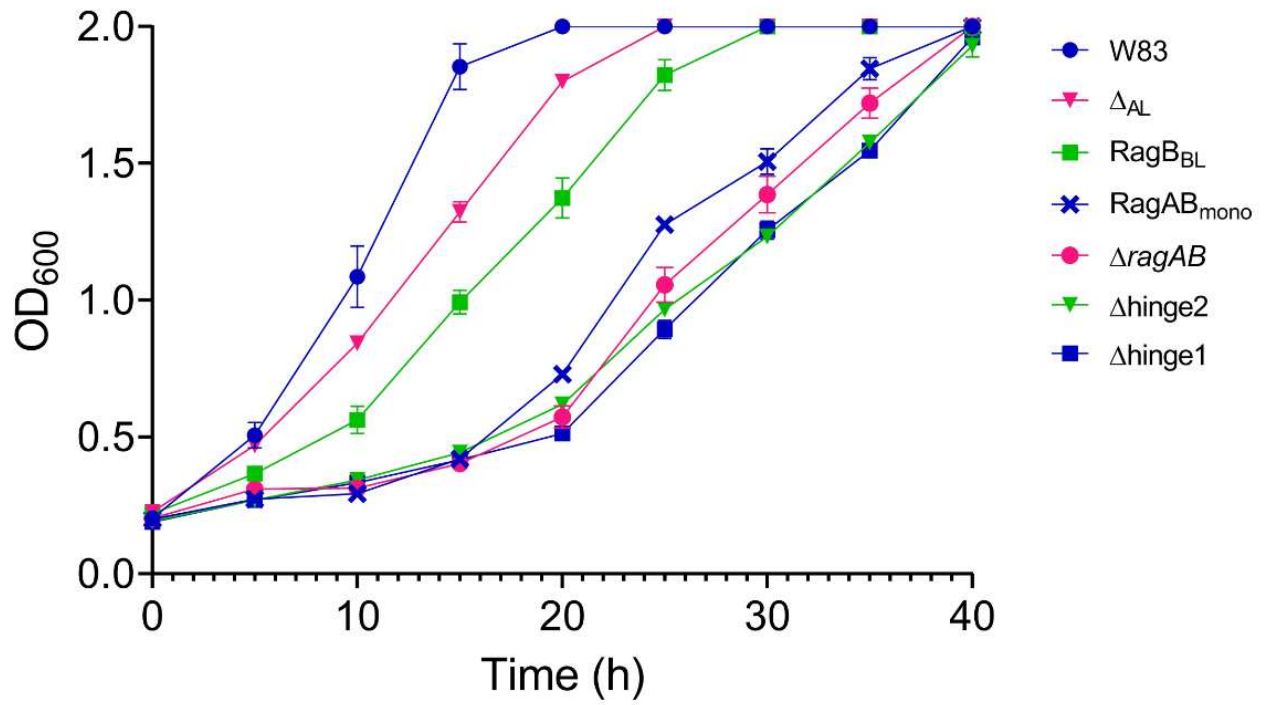
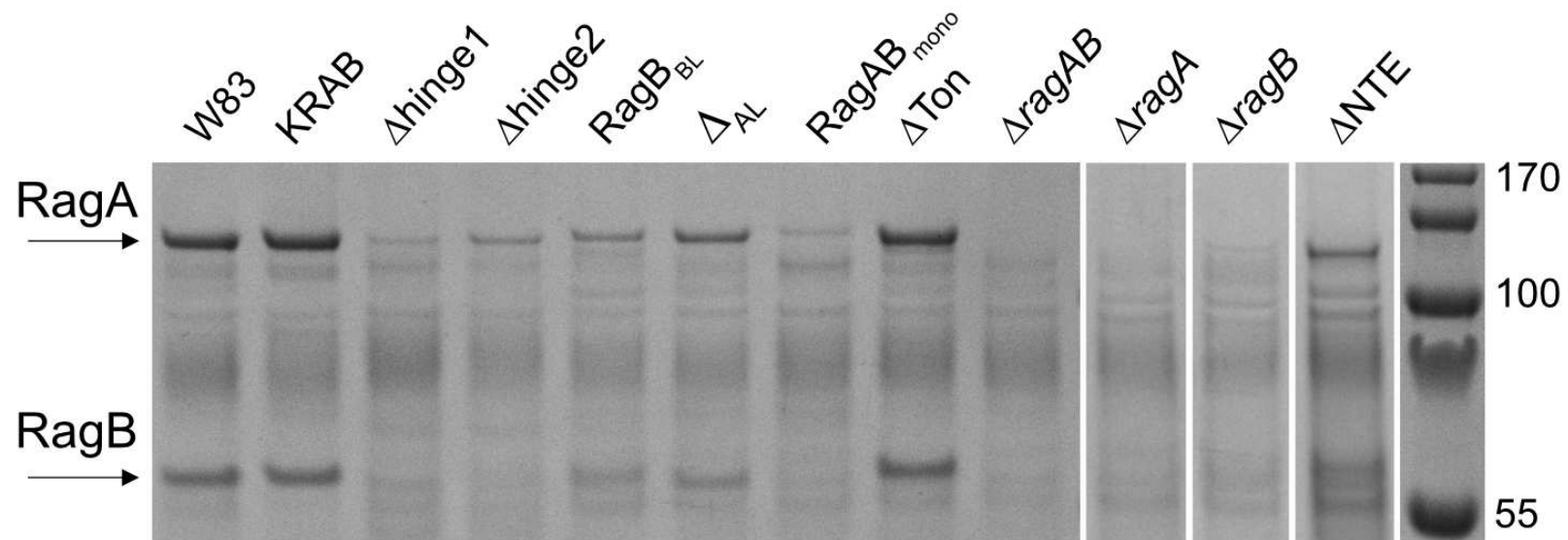


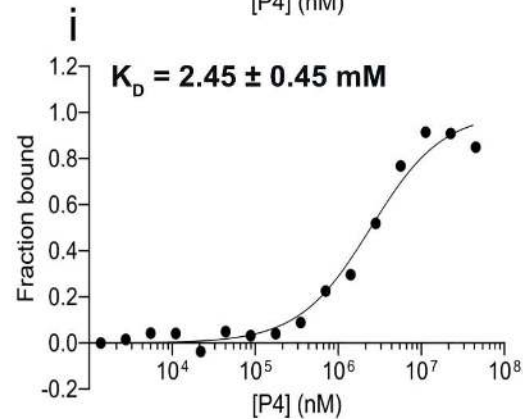
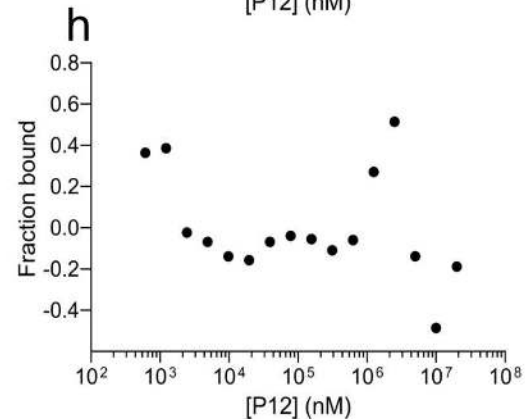
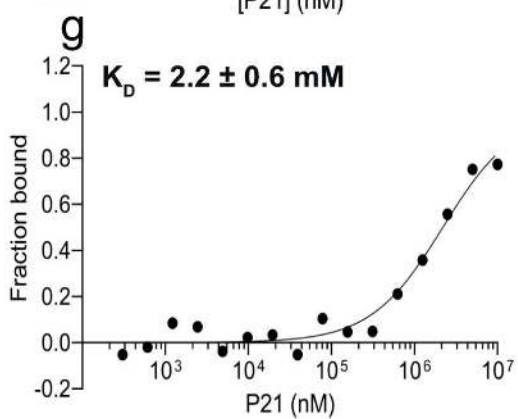
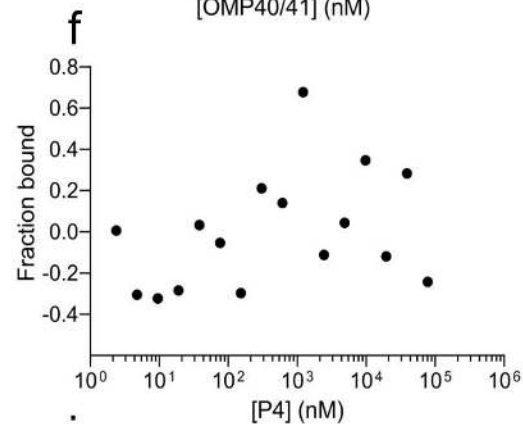
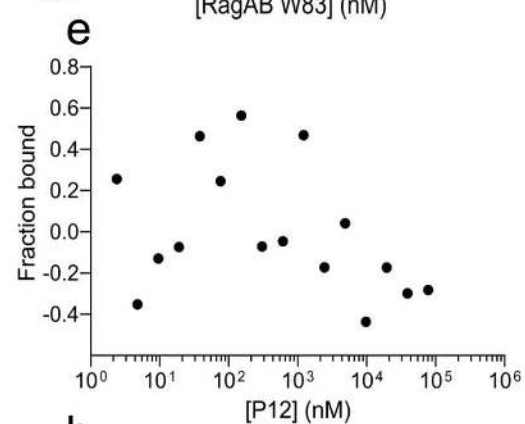
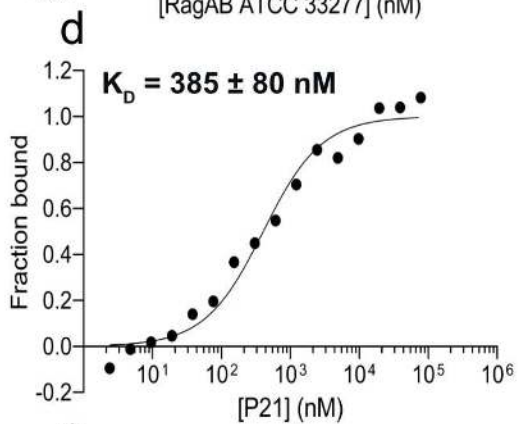
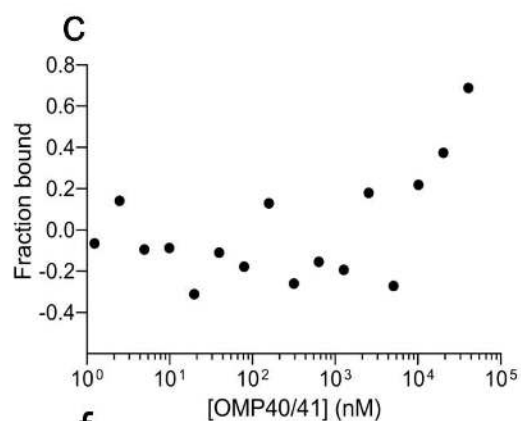
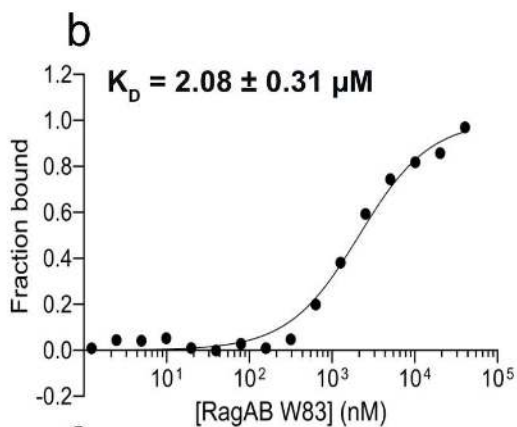
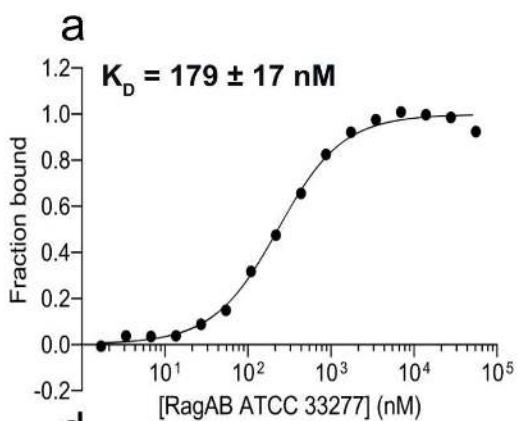


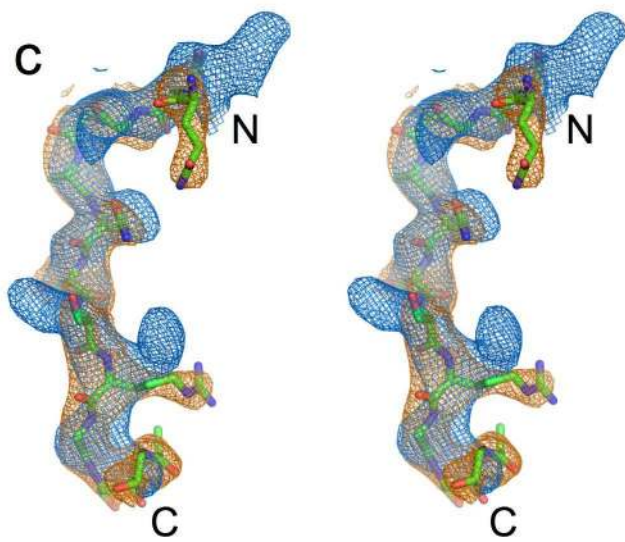
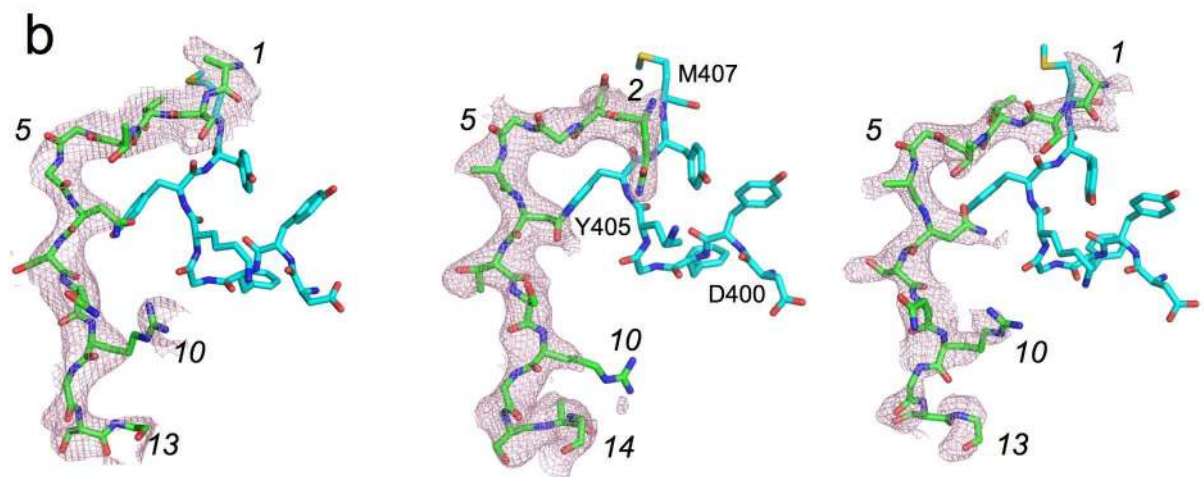
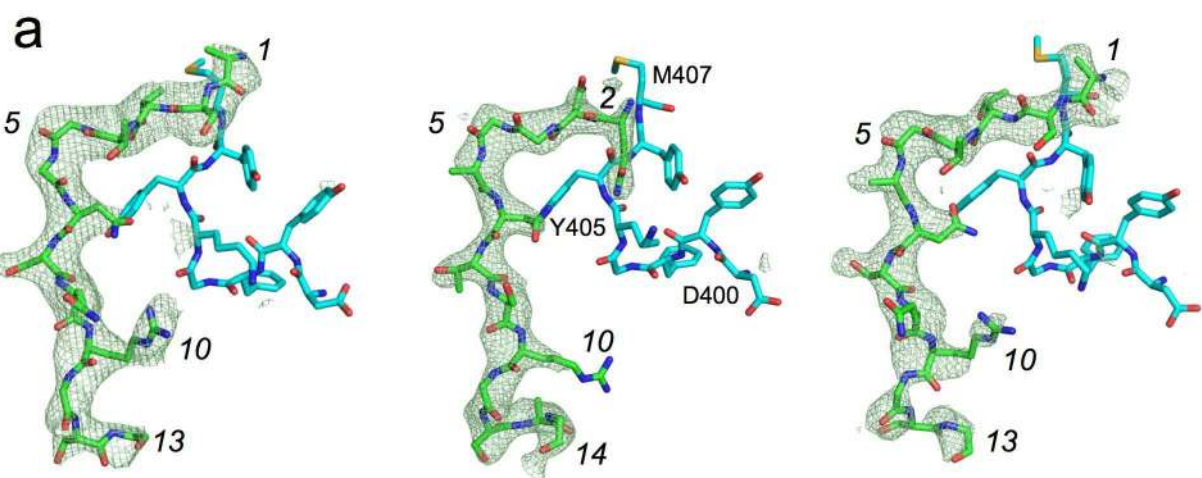




**a****b**

**a****b**







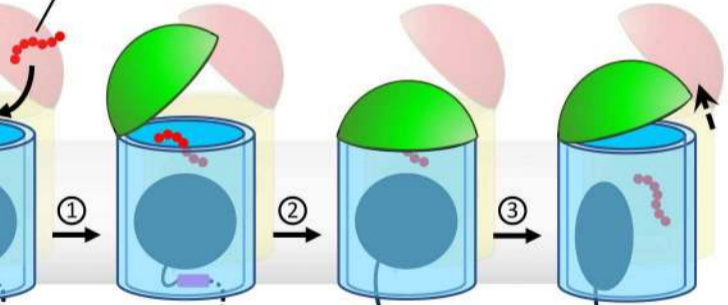
Serum and tissue-derived proteins (e.g. albumin)



Gingipain



Peptide substrate



Plug domain  
TonB box

TonB

