

1    **Supplementary Information**

2

3

4

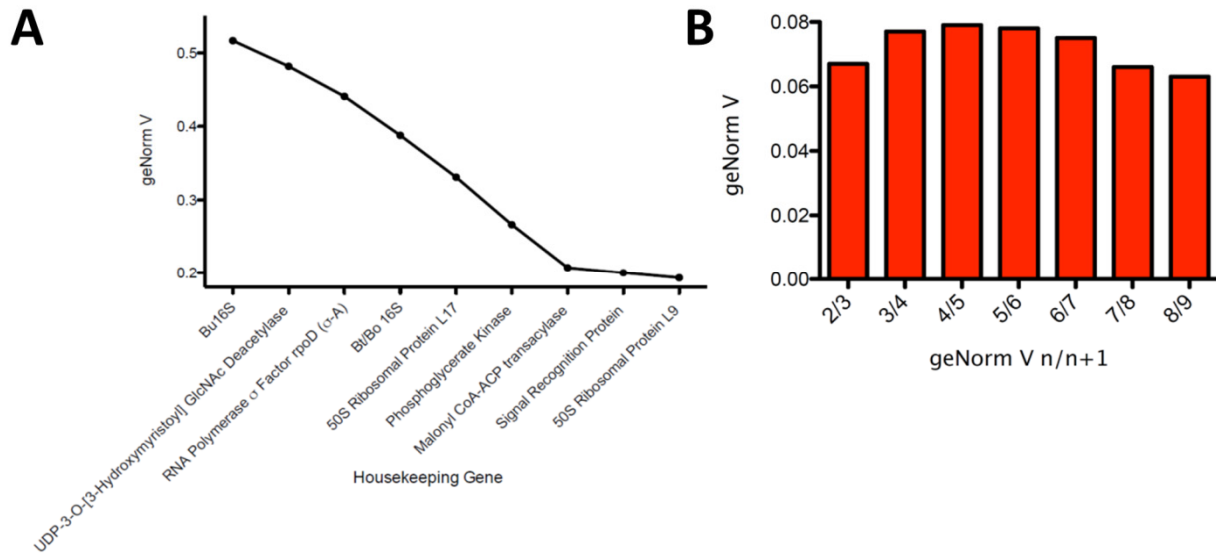
5

6

7

8

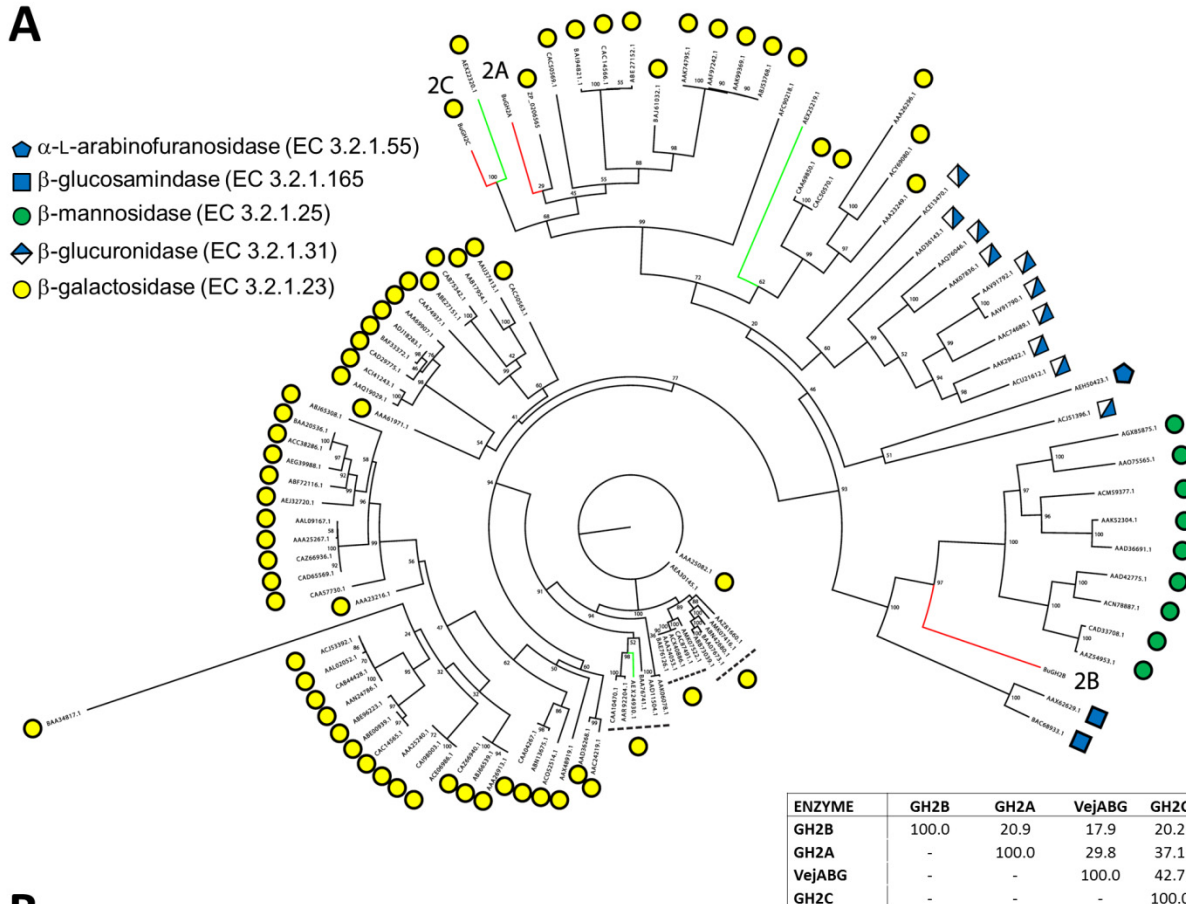
9



**Supplementary Figure 1. Putative reference gene stability evaluation results using geNorm.**

(a) The putative housekeeping genes are ranked from most unstable (left, high M value, *Bu16S*) to the most stable (right, low M value, 50S Ribosomal Protein L9). geNorm V is < 0.15 when comparing the normalization factors based on using the two or three most stable reference genes. (b) Optimal normalization factors were calculated as the geometric mean of reference targets SRP and 50S\_RP\_L9 (23). Using these two genes as the reference targets resulted in a very high reference target stability (average geNorm M  $\leq$  0.2).

A



B

GH2C --YALNNNSDENNCWKFLLSDSVCYSFVNNPSSWATVNLPHDWSVOLFESTAEGETGLKGGIGWYSKIFDTPDNFVDKRCYITVFDGV  
VejABG MHNSPRSTTLFNDWLFQAKD-----KPNTKQNSTVLPHDWSVASAFSPQYDGTGMLPGGIGWYKCFKNPLNKHYSKCIITVFDGI

GH2C YNNSEYINGRKLGRHPYGYSPFFYDISDYLNPKGOENRHSVRIDHSRYADSRWYTGSGIYRETQLIFDOKLHIEVWGTFVITPVSSESR  
VejABG YNNATININGYDIHQAYGYAPPNIETDYLKS---DNVITIHVDRRYIDSRWYTGSGIYRDIEMVLTKDVFVEITENHIKASVSSNQI

GH2C ATVNIEVRVKNDYSGPRAGEVTSYFDSKNKKVGEKLTSLFLIEA-----GKEMKINQSVESISNFSIWDVDSPSMYLAKSEILVDGNVVD  
VejABG GHIHQQLMIEAK-----TKTHYLTIVSRLLPNSDNCVATARTHSVNNREVCDELTCDDLSLWSDSPILYKLETCIYENGCVVD

GH2C TKETPHGIRSIKFDAAKNGFFLNGKNMKIKGVCLHHDASMGAAALVEDVWRRLQTLKOGGONAIRLSHNPAGDAFLCLCDPMGFIVQEEF  
VejABG KVSENIGFRSTIEFSPEQGFLLNGMPTKVRGVCLHHDGGLVGAAVPDETWIRRLSKLKOCGVNAIRIAHNPASKRLLHLCDDPMGFIVQDEF

GH2C FDEWDYPKDKRLNMDQSIDYITRGYCEYFQFAERDLKNVMLRSRNHPCIFQWSIGNEIEWTYKGCRESSTGFFSADAGGYFWNCPPPYS  
VejABG FDEWDYPKDKRLNMDQHDDEFSQCYTEFFQIFRAKIDLOTLKCHINHPSIFQWSIGNEIEWTYPRNVGATGFFSADAGGYFWNCPPPYS

GH2C TORIREEWAKCPKOTYDIGRTAKLAFTWREMDITREVTANCILPSISYETGYIDALDVAGFSYRRVMYDYAHKNYPDKPAMGTENLCQW  
VejABG PDEIKDKLKNLPQHTYDICKTAKLAFWVKAIDOTREVTANCILPSSSYHSYADALDVAGFSYRRVMYDYGHEIRENLPITIGNENLPQW

GH2C HEWKAVIERDYIPEMFIITGTVDYLGEGVGTGKREWPORAIKGLLDLAGFEKPSFHHMKSLWTDALFFIAIYSQANKSSYVEKDGRKIT-DK  
VejABG HEWKAVLERNHVSGELFIITGTGINYMGES---HGKWEVRTTDSGLLDLAGFEKPSYALFKSLWTDPEYVKVFTQADLTQLKFDEQTEVAFE

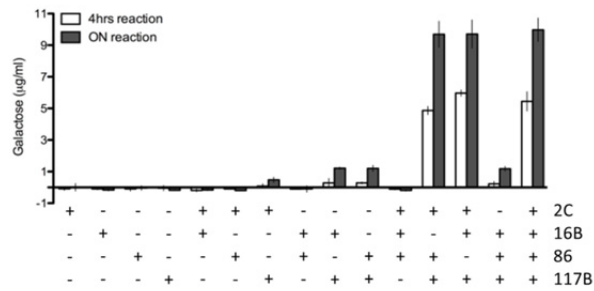
GH2C DPKKFWTORLWVVDVNSHWNYTKGERVVEHYSNCEIBLFQNGKSTIGKRRLKDFEDHIYKFSVDFKDGNTVAKGRKNGKKTTSIAIYTT  
VejABG HDENAFQKRLWVVDVNSHWNYENEQVITDEAYSNCPPQCLYINDELVTQCLEKQIDRVERWALPYRAGKISLVGRKNDVEVTRDEIVT

GH2C KE-TNSIKKLSVDKVAVDANNTDVIHVTAQLDRNGRNSWEEKIITENIGGNYRLIGVENGHLNVLNYSNIVKTYKGRALLVLQAT-D  
VejABG SGVPRKTSI-VDET--HEGSSSYRQLIVQMDKDNHPSVEEALLESVRGC-EWIGADNGSISSINAYNSPILATREGRVLAIVKSSQG

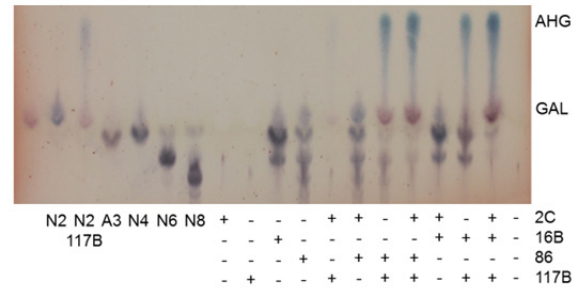
GH2C KAGILNINANSKGSISSNDIKVEVK  
VejABG QSGDIEIYSNSGVKASFSL----

**Supplementary Figure 2: Phylogeny of characterized GH2 enzymes. (a)** Sequences were extracted from the CAZy database (17), trimmed to their GH2 catalytic module using dbCAN (18), and aligned using MUSCLE (24). The phylogenetic tree was generated using RAxML (20) and FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). Ascribed activities for GH2 include  $\beta$ -glucosaminidase (EC 3.2.1.165; blue square),  $\beta$ -mannosidase (EC 3.2.1.25; green circle),  $\beta$ -glucuronidase (EC 3.2.1.31; blue/white diamond),  $\beta$ -galactosidase (EC 3.2.1.23; yellow circle), and  $\alpha$ -L-arabinofuranosidase (EC 3.2.1.55; blue pentagon). Lineage of GH2s from *Vibrio* sp. EJY3 (25) are shown with green lines; and GH2s from *Bu* NP1 with red lines. Sequence identities between the GH2s from *Bu* NP1 and the GH2 from *Vibrio* sp. EJY3, VejABG, are shown as percentages. **(b)** Alignment of GH2C and VejABG. Residues comprising the putative +1 subsite in GH2C are denoted using white triangles. The catalytic acid/base (E436) and nucleophile (E552) are indicated using black triangles.

**A**



**B**



37

38 **Supplementary Figure 3: Production of GAL and AHG by *Bu* NP1 Ag-PUL agarases. (a)**

39 Combinatorial cocktail digestions of agarose using GH86, GH16B, GH2C, and GH117B were

40 performed and incubated for 4 hours or overnight. GAL was quantified in the resulting

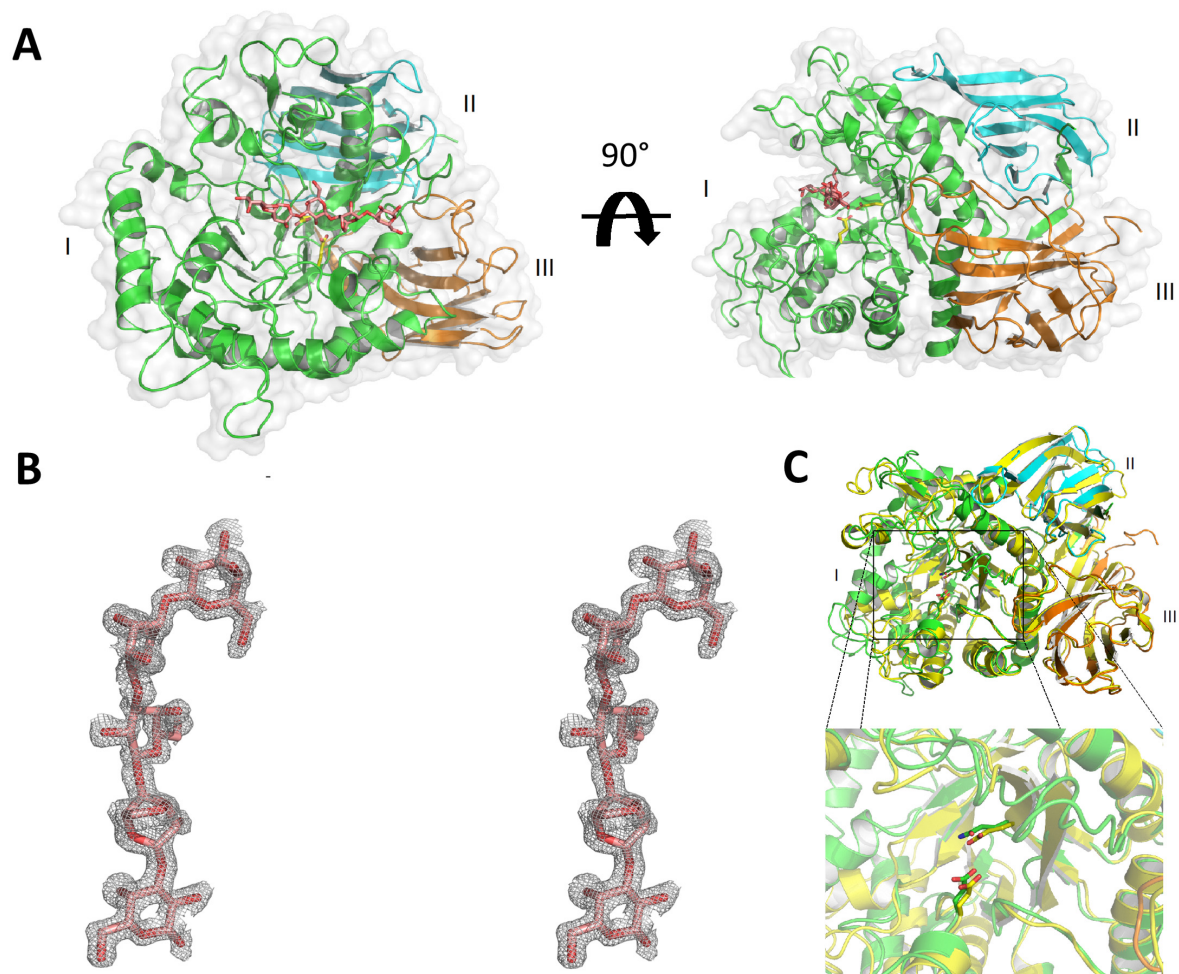
41 supernatants using a reducing sugar assay. Error (s.d.) is represented by error bars over four

42 replicates. (b) TLC analysis of the complete factorial of GH86, GH16B, GH2C, and GH117B

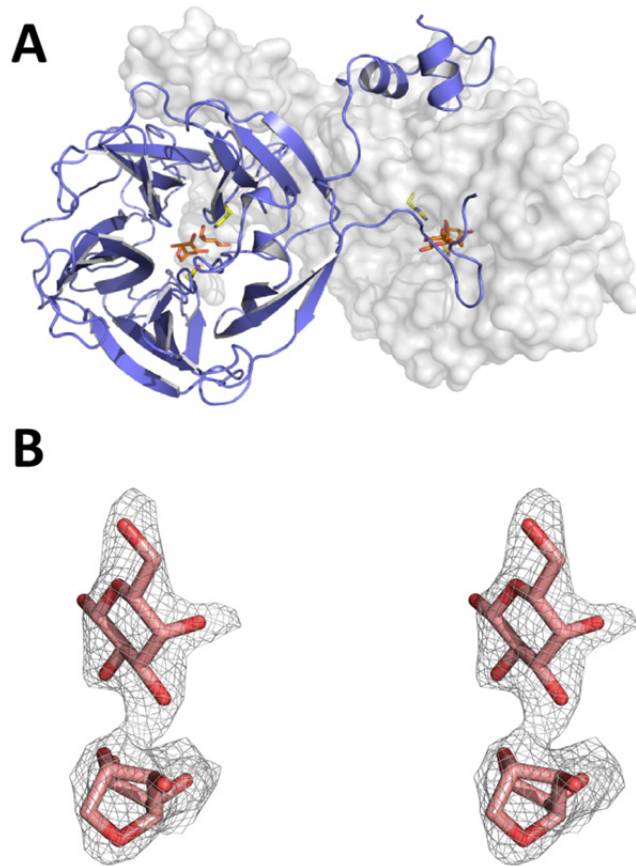
43 cocktail agarose digestions.

44

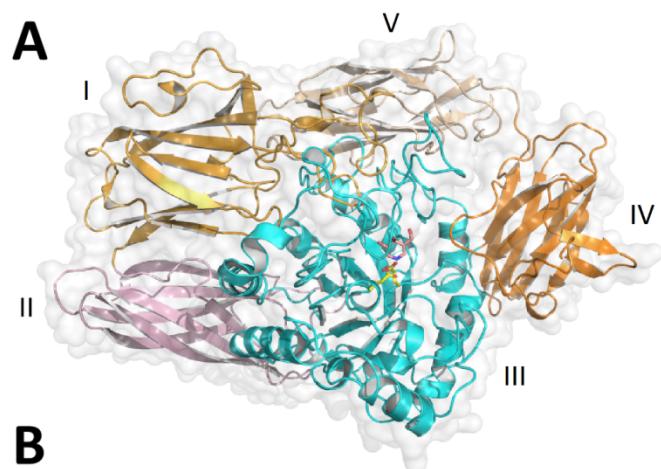
45



**Supplementary Figure 4: Structural analysis of GH86 from the *Bu* NP1 Ag-PUL.** (a) Structure of the inactive E322Q mutant of GH86 in complex with N8 (pink). Domains are indicated by Roman numerals. (b) Despite crystal soaks with excess N8, electron density allowed for the modelling of agaropentaose (A5) with a maximum likelihood<sup>σ<sub>a</sub></sup> weighted  $2F_{obs} - F_{calc}$  electron density map contoured at 0.35 electrons<sup>Å<sup>-3</sup></sup>. (c) Structural alignment of *Bu*GH86 (green, cyan, orange) with *Bp*GH86A [yellow, pdb entry: 4AW7, (12)], with structurally conserved catalytic residues shown in stick format (inset). Domains are indicated by Roman numerals.

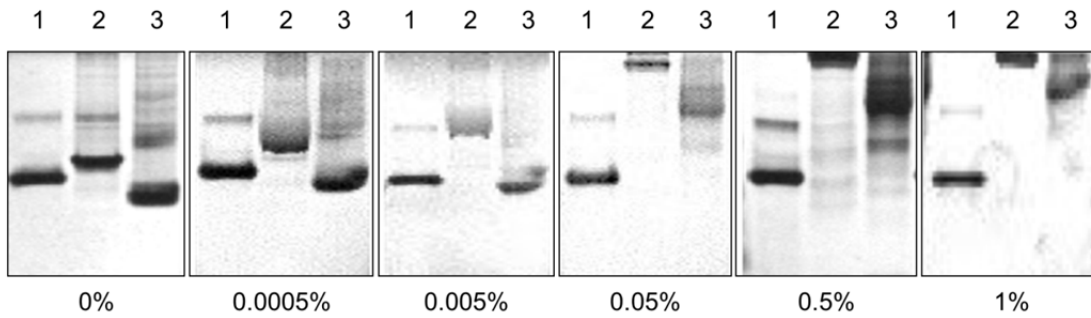


**Supplementary Figure 5: Structural features of GH117B from the *Bu* NP1 Ag-PUL. (a)** Chain A (cartoon, blue) and Chain B (surface, grey) of GH117B in complex with GAL and AHG (sticks, pink). **(b)** Electron density in the active site of GH117B was sufficient to model GAL and AHG with a maximum likelihood<sup>σ<sub>a</sub></sup> weighted  $2F_{obs}-F_{calc}$  electron density map contoured at 0.25 electrons<sup>Å<sup>3</sup></sup>.

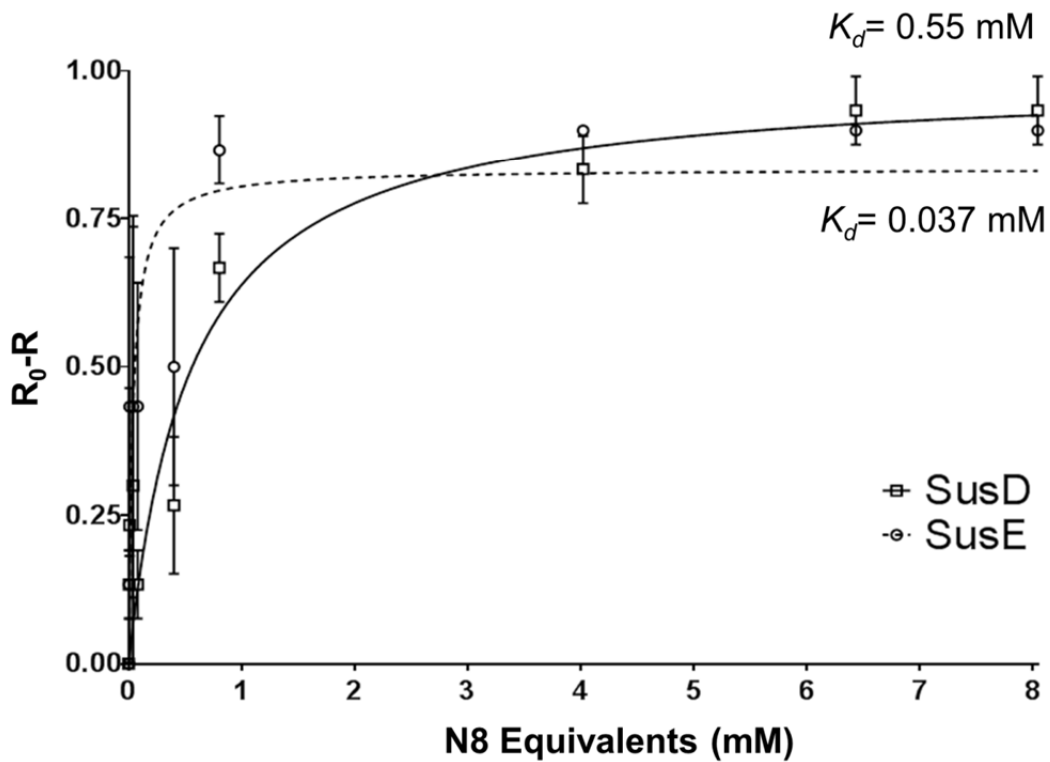


**Supplementary Figure 6: Structure of GH2C in complex with the inhibitor galactoisofagomine (GIF).** (a) Cartoon representation of the catalytic region comprising the catalytic domain (III, cyan), surrounded by four IG-like domains (I, II, IV, and V). The bound GIF molecule is shown in pink and the protein surface in transparent grey. (b) Electron density in the active site of GH2C was sufficient to model GIF with a maximum likelihood  $\sigma_a$  weighted  $2F_{obs}-F_{calc}$  electron density map contoured at 0.27 electrons  $\text{\AA}^3$ .

**A**



**B**



71

72 **Supplementary Figure 7: Affinity gel electrophoresis of recombinant SusD- (NP1\_10) and**

73 **SusE-like (NP1\_9) proteins. (a)** Quantitative AGE using stepwise gradient of agarose-containing

74 native gels, showing the relative mobilities for a BSA negative control (lane 1), SusE- (lane 2)

75 and SusD-like (lane 3) proteins. **(b)** Normal relative mobility of SusD-like and SusE-like proteins

76 plotted against N8 equivalents. Binding constants were fit to a one-site binding model and

77 calculated using GraphPad software. Error (s.d.) is represented by vertical error bars over three  
78 replicates.

79

80

81

82

83

84

85

86

87

88

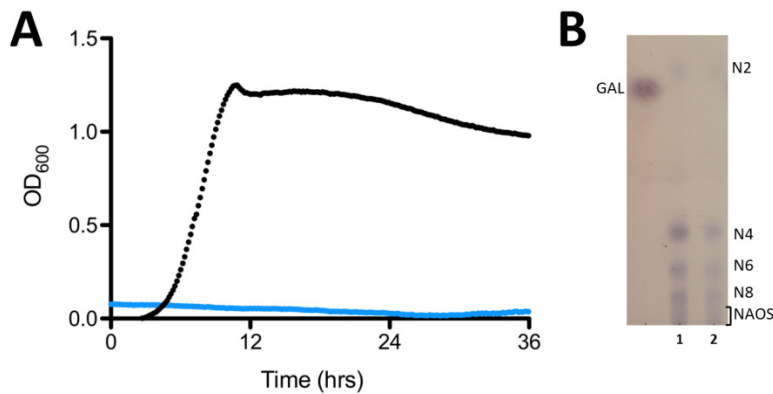
89

90

91

92

93



**Supplementary Figure 8: Growth of *Bu* ATCC 8492 on spent supernatants from growth of *Bu*NP1 on LMPA.** (a) Growth curves of agarolytic *Bu* ATCC 8492 grown on GAL (black) and the spent supernatants of *Bu* NP1 following growth on LMPA (blue). The experiment was replicated three times, with four to eight observations per replicate; error (s.d.) is represented by trace thickness. (b) TLC analysis of the spent supernatants of *Bu* NP1 (lane 1) following growth on LMPA, and *Bu* ATCC 8492 following growth on the *Bu* NP1 spent supernatants (lane 2).

**Supplementary Table 1: Annotation of the Agarolytic PUL from *B. uniformis* NP1**

| Gene          | Predicted Protein       | LipoP prediction <sup>1</sup>   | Function <sup>2</sup>               |
|---------------|-------------------------|---------------------------------|-------------------------------------|
| <i>np1_1</i>  | Cyclase                 | Cytoplasmic                     | Unknown                             |
| <i>np1_2</i>  | DUF2582                 | Cytoplasmic                     | Unknown                             |
| <i>np1_3</i>  | Unknown                 | SPI                             | Unknown                             |
| <i>np1_4</i>  | BuAs1A-like (sulfatase) | SPII (2+=G, OM-CM) <sup>2</sup> | Sulfatase                           |
| <i>np1_5</i>  | BuGH117A                | SPI                             | Agarase/Porphyrinase                |
| <i>np1_6</i>  | BuGH16A                 | SPI                             | Agarase/Porphyrinase                |
| <i>np1_7</i>  | BUGH86                  | SPI                             | β-agarase*                          |
| <i>np1_8</i>  | BuGH16B                 | SPII (2+=S, OM)                 | β-agarase*                          |
| <i>np1_9</i>  | SusE-like               | SPII (2+=Q, OM)                 | CBM-Containing, Transport           |
| <i>np1_10</i> | SusD-like               | SPII (2+=E, OM)                 | CBM-Containing, Transport           |
| <i>np1_11</i> | SusC-like               | ND                              | Outer Membrane Transport            |
| <i>np1_12</i> | BuAs1A-like (sulfatase) | Cytoplasmic                     | Sulfatase                           |
| <i>np1_13</i> | Transferase/Transposase | Cytoplasmic                     | Chromosome Integration              |
| <i>np1_14</i> | DDE_1                   | Cytoplasmic                     | Transposase                         |
| <i>np1_15</i> | Transferase/Transposase | Cytoplasmic                     | Chromosome Integration              |
| <i>np1_16</i> | DDE_3                   | Cytoplasmic                     | Transposase                         |
| <i>np1_17</i> | HTCS                    | ND                              | Transcriptional Regulation          |
| <i>np1_18</i> | Unknown                 | Cytoplasmic                     | Unknown                             |
| <i>np1_19</i> | BuAs1A-like             | SPI                             | Sulfatase                           |
| <i>np1_20</i> | GH16C                   | SPII (2+=N, OM)                 | β-Porphyrinase*                     |
| <i>np1_21</i> | GH29                    | SPII (2+=N, OM)                 | L-Galactosidase                     |
| <i>np1_22</i> | GH2A                    | Cytoplasmic                     | β-Galactosidase, Porphyrinase*      |
| <i>np1_23</i> | Unknown                 | Cytoplasmic                     | Unknown                             |
| <i>np1_24</i> | KdgK-like               | Cytoplasmic                     | KDG Kinase                          |
| <i>np1_25</i> | RhaT-like               | ND                              | Sugar Importer                      |
| <i>np1_26</i> | GalM-like               | Cytoplasmic                     | Galactose Mutarotase                |
| <i>np1_27</i> | MR-like                 | Cytoplasmic                     | Epimerase/Dehydratase/Decarboxylase |
| <i>np1_28</i> | AldA-like               | Cytoplasmic                     | Aldehyde Dehydrogenase              |
| <i>np1_29</i> | KduD                    | Cytoplasmic                     | KDG Oxidoreductase                  |
| <i>np1_30</i> | MDR-like                | Cytoplasmic                     | Alcohol Dehydrogenase               |
| <i>np1_31</i> | GH2B                    | SPI                             | β-mannosidase                       |
| <i>np1_32</i> | GH117B                  | SPII (2+=V, OM)                 | 3,6-Anhydryo-L-Galactosidase*       |
| <i>np1_33</i> | GH2C                    | SPI                             | β-Galactosidase, Agarase*           |

<sup>1</sup>(26)

<sup>2</sup>Characterized enzymes are indicated with an asterisk. See Supplementary Table 8 for more detail.

<sup>3</sup>(27)

**Supplementary Table 2: Distance identity tree of metabolic enzymes with the AHG pathway from *Vibrio* sp. VEJ3<sup>1</sup>**

| Protein           | AHGD   | NP1_24 | NP1_25 | NP1_26 | NP1_27 | NP1_28 | NP1_29 | NP1_30 |
|-------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| AHGD <sup>2</sup> | 100.0  | 18.0   | 19.0   | 17.6   | 11.7   | 64.6   | 14.8   | 13.2   |
| NP1_24            | -      | 100.0  | 12.7   | 12.6   | 9.6    | 17.9   | 8.3    | 10.3   |
| NP1_25            | -      | -      | 100.0  | 15.3   | 13.7   | 19.3   | 15.2   | 15.1   |
| NP1_26            | -      | -      | -      | 100.0  | 11.3   | 16.3   | 13.2   | 11.2   |
| NP1_27            | -      | -      | -      | -      | 100.0  | 14.4   | 22.9   | 21.5   |
| NP1_28            | -      | -      | -      | -      | -      | 100.0  | 13.3   | 13.6   |
| NP1_29            | -      | -      | -      | -      | -      | -      | 100.0  | 16.7   |
| NP1_30            | -      | -      | -      | -      | -      | -      | -      | 100.0  |
| Protein           | AHGI   | NP1_24 | NP1_25 | NP1_26 | NP1_27 | NP1_28 | NP1_29 | NP1_30 |
| AHGI <sup>3</sup> | 100.00 | 15.70  | 13.30  | 10.64  | 72.10  | 16.13  | 20.00  | 16.37  |
| NP1_24            | -      | 100.00 | 12.50  | 15.58  | 14.05  | 15.85  | 9.76   | 18.14  |
| NP1_25            | -      | -      | 100.00 | 22.86  | 14.22  | 15.69  | 12.02  | 14.65  |
| NP1_26            | -      | -      | -      | 100.00 | 10.04  | 15.16  | 5.81   | 11.49  |
| NP1_27            | -      | -      | -      | -      | 100.00 | 14.29  | 22.18  | 17.80  |
| NP1_28            | -      | -      | -      | -      | -      | 100.00 | 13.36  | 19.62  |
| NP1_29            | -      | -      | -      | -      | -      | -      | 100.00 | 12.89  |
| NP1_30            | -      | -      | -      | -      | -      | -      | -      | 100.00 |

<sup>1</sup>(28).

<sup>2</sup>AHGD (VEJY3\_RS09290) = 3,6-anhydro-L-galactose dehydrogenase.

<sup>3</sup>AHGI (VEJY3\_RS09420) = 3,6-anhydro-L-galactose isomerase.

129 **Supplementary Table 3: Primers used for RT-PCR of *np1\_1-33***  
130

| Gene          | Forward Primer (5' to 3') | Reverse Primer (5' to 3') |
|---------------|---------------------------|---------------------------|
| <i>np1_1</i>  | AGCCAATTTGGAGAGAAACGA     | CTTTCCTTTCTCCGCTATACCA    |
| <i>np1_2</i>  | CCTATCTATCGCACCTCTCC      | CACCATTTTTACCTTCGAGCA     |
| <i>np1_3</i>  | CAATACATGCTTTGCCTCCT      | CCTTGAGGTAAACAGTACGCTGA   |
| <i>np1_4</i>  | GGCATCTTGGTCTTGGTGATA     | AAGCTCTTCCGTTTTCAATCC     |
| <i>np1_5</i>  | ATTGGGAAGAACAGGGGATAG     | CGGTACTTCCGCTCATTACAG     |
| <i>np1_6</i>  | GGCAAGGAAAGGTGGTTTAAG     | TCATCCCTGTCATTACTGTTGC    |
| <i>np1_7</i>  | TGGGGATATAATCCCAAAACG     | ACCCTTGTAACCTTCACATCC     |
| <i>np1_8</i>  | CCGTTTCAGGACTATCAACCA     | TCCCAGATACCCTTCTTACCA     |
| <i>np1_9</i>  | GCTTCGGAAGATCCTGTTGA      | GGAATCATATTTTCCCCGTTG     |
| <i>np1_10</i> | AGTTGAGAAAGGTGACGGTGA     | TCCACTCTATCCCAACGTGTC     |
| <i>np1_11</i> | GATTACGACAAAACGCGGTAA     | CCACCTAAAGCACGCTGATAC     |
| <i>np1_12</i> | GGGTGTGAAAACAGACTTGA      | TACAGCGTTGAATGCCAGATA     |
| <i>np1_13</i> | TAATGGATTGTTCCGATGAGG     | CTTGACCAATGCTCCTAAAA      |
| <i>np1_14</i> | AGAATGGAAGCCAACATTTCC     | CGCTAATCTTCCAATAATCC      |
| <i>np1_15</i> | CGAGGAGTCGTCCCTAAATTC     | CGGGAAATAAGGTTGGGTATT     |
| <i>np1_16</i> | CTGTCCATAGGAACCGAAAGA     | CAGCAGGTCTTATCCATTGTC     |
| <i>np1_17</i> | GGAGACTAACCAGTCGGAGGA     | CAAGACCTTCCTTCCGTTTC      |
| <i>np1_18</i> | TTCAGTGGGGATAGTGA CTGG    | CAGCCCGTTCTTTGTAGTCAG     |
| <i>np1_19</i> | GGAAGAAAAATGGGGTGAGAA     | GACGATGATATCCCTCGACTG     |
| <i>np1_20</i> | CCTGCAAACCTGGAAAGATGAG    | GCCCATTGTCTTTTCGTCATA     |
| <i>np1_21</i> | CTCATCAACGTTTTGGCTTTC     | TTTCCGTTGAAATCACCAGTC     |
| <i>np1_22</i> | ACGTAGGACCAAAAGTTGACG     | GTTCCCTTGTTTTCTCCTTCG     |
| <i>np1_23</i> | GTAATAACTGCCGATGGCAA      | TGCAGCATTACCGTTTCCTAC     |
| <i>np1_24</i> | CGACAATGAAGAACTGGTGAC     | CCATCCTCACACAAGCCTTTA     |
| <i>np1_25</i> | ATTACCCAGCCATTTCTGAT      | ATGGACGCATAATGGACTGAG     |
| <i>np1_26</i> | GGCTGTTGTAACACAGGGAAA     | TTCTTAAACAGACCCCACGTC     |
| <i>np1_27</i> | GGGGTTCCTGGTAATGTAAC      | CTCCATAAATCCCGAAAGATTG    |
| <i>np1_28</i> | TGCTTCATTGCAATTTCTG       | GTCAAAGGGAACAGAAGGTC      |
| <i>np1_29</i> | CAGTTGGAGTTGGGAGGAAAG     | CTCGTGAAGATACAGGCGTTC     |
| <i>np1_30</i> | AAAGGATATGGTGCAAGAGG      | TTACCAGCCCATTATTAGCA      |
| <i>np1_31</i> | CCGGTGTTGAGGTACGGTTA      | ACACGAGCTCCCATGAATGTA     |
| <i>np1_32</i> | TCCCCGATTCTCCTTAACAT      | TCCGGTACGATAGGGTAATCC     |
| <i>np1_33</i> | AAGCGGATTCTTTGGGAATAC     | CCTTCCTCATAAGCCAAATCA     |

132 **Supplementary Table 4: Primers used for RT-PCR of housekeeper genes**

| Gene                                               | Forward Primer (5' to 3') | Reverse Primer (5' to 3') |
|----------------------------------------------------|---------------------------|---------------------------|
| 50S Ribosomal Protein L17                          | GTCCGGGTGGTTACACTCGT      | AGCTTCCTCAGCAGCAGGAG      |
| 50S Ribosomal Protein L9                           | AAAGAAGACATTGTAACTTGGG    | GATACGCCTTCCAGCTTAGC      |
| Malonyl CoA-ACP transacylase                       | GTTTGCGTGGCAGCCAACTA      | ATGGATTTCTGTTGCATTGA      |
| Phosphoglycerate Kinase                            | TTATTGCCGGCGGTATGACT      | GGGCAGAACTTGGTGTGGC       |
| RNA Polymerase $\sigma$ Factor rpoD ( $\sigma$ -A) | TCCGTTGAACCAGGTAGGCT      | AGCTGTTGTCCTCTCCTTCT      |
| Signal Recognition Protein                         | GCTATCGAGCAGTTGCGTGT      | CGGCAATCTCGTTCATCATC      |
| UDP-3-O-[3-Hydroxymyristoyl] GlcNAc Deacetylase    | CAAGCCGTACTTTCGTGTTT      | AGGTTTGTGGTTGATATATCC     |
| <i>Bt/Bo_16S</i>                                   | AGGTTTGTGGTTGATATATCC     | GGTAGTCCACACAGTAAACGATGAA |
| Bt16S1                                             | GGTAGTCCACACAGTAAACGATGAA | CCCGTCAATTCCTTTGAGTTTC    |

133  
134

**Supplementary Table 5: Primer sequences used to recombinant *Bu* NP1 Ag-PUL proteins (*NdeI* or *NheI* and *XhoI*) restriction sites are underlined**

| Primer         | Sequence 5' -> 3'                                             |
|----------------|---------------------------------------------------------------|
| GH2A_Foward    | TTT GGA <u>GCT AGC</u> GTA ACG GAT GGT ATC TCC TTC            |
| GH2A_Reverse   | CTT TGT <u>CTC GAG</u> TTA AAG TTC TAC ATT TAT TGT TGT C      |
| GH2B_Foward    | TTT GGA <u>GCT AGC</u> CAG CCG GGC TTC GAA TCA TTT TC         |
| GH2B_Reverse   | CTT TGT <u>CTC GAG</u> CTA TTT AGC TTT TCC AAA ACC TTT GG     |
| GH2C_Foward    | TTT GGA <u>GCT AGC</u> TAT GCG TTA AAT AAT AAT TCC G          |
| GH2C_Reverse   | CTT TGT <u>CTC GAG</u> TCA TTT AAC TTC AAC CTT TAA ATC        |
| GH16A_Foward   | TTT GGA <u>GCT AGC</u> CAA AAC CTT CCG GAG GTT CCG            |
| GH16A_Reverse  | CTT TGT <u>CTC GAG</u> TCA GTC TAC ATA CAT AGT ATT G          |
| GH16B_Foward   | TTT GGA <u>GCT AGC</u> GGA GAT AAT AAT GGA GAA GAG            |
| GH16B_Reverse  | CTT TGT <u>CTC GAG</u> TTA TTT ATC CTC TAC CGG TTT ATA AAT TC |
| GH16C_Foward   | TTT GGA <u>GCT AGC</u> AAT GAA GTT ATT ATA GAA ACA GG         |
| GH16C_Reverse  | CTT TGT <u>CTC GAG</u> TTA TAA AGT TTT CAA AAT ATA TG         |
| GH86_Foward    | TTT GGA <u>GCT AGC</u> CAG ACT CAT GTG CAA TTG AAT TTG        |
| GH86_Reverse   | CTT TGT <u>CTC GAG</u> TCA TAT ACC TCT AAT ATT ATT TG         |
| GH86_E322Q_Fwd | CCT CTT ATG ATC TCG CAA TAC AGT GCG                           |
| GH86_E322Q_Rev | C ATG CGT CTG CGC ACT ATG TTG CG                              |
| GH117A_Foward  | TTT GGA <u>GCT AGC</u> CAG AAT CCA TTC CCT TAT G              |
| GH117A_Reverse | CTT TGT <u>CTC GAG</u> TTA GAT AAT CTC CTT CTG CTT TTT AG     |
| GH117B_Foward  | TTT GGA <u>GCT AGC</u> ACA GTG TCT ACT GAT GAT AGT GC         |
| GH117B_Reverse | CTT TGT <u>CTC GAG</u> TTA TTT TCG CTT TGT TGA TTC ACC        |
| NP1_9_Foward   | TTA <u>AGC TAG CTG</u> CCA AAA GGA CTA TGA GTT TAA C          |
| NP1_9_Reverse  | TTA <u>ACT CGA GCT</u> ATG GAC GTT TTT CTA TTT TAG TC         |
| NP1_10_Foward  | TTA <u>AGC TAG CTG</u> TGA AGA TGC CTT GCT GTC AC             |
| NP1_10_Reverse | TTA <u>ACT CGA GTT</u> AAT TCT GTA ATG CTT TAT TAT TCT G      |

139 **Supplementary Table 6: Characterized enzyme activities from Ag-PUL**  
 140

| Glycoside Hydrolase | Agarase |
|---------------------|---------|
| GH2A                | -       |
| GH2B                | -       |
| GH2C                | +       |
| GH16A               | -       |
| GH16B               | +       |
| GH16C               | -       |
| GH29                | -       |
| GH86                | +       |
| GH117A              | -       |
| GH117B              | +       |

141

**Supplementary Table 7: X-ray data collection and statistics of apo structures**

|                                      | GH2C                                      | GH16B                            | GH86                                 | GH117B                                        |
|--------------------------------------|-------------------------------------------|----------------------------------|--------------------------------------|-----------------------------------------------|
| <i>Data Collection</i>               |                                           |                                  |                                      |                                               |
| Beamline                             | CLS - 08ID-1                              | CLS - 08ID-1                     | CLS - 08ID-1                         | CLS - 08ID-1                                  |
| Wavelength (Å)                       | 1.0000                                    | 1.28215                          | 0.97949                              | 0.97949                                       |
| Space Group                          | P1                                        | P4 <sub>3</sub> 2 <sub>1</sub> 2 | I2 <sub>3</sub>                      | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell Dimensions                      |                                           |                                  |                                      |                                               |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)   | 75.1, 121.4, 124.5                        | 109.7, 109.7, 239.3              | 167.7, 167.7, 167.7                  | 84.3, 100.3, 112.7                            |
| $\alpha$ , $\beta$ , $\gamma$ (°)    | 70.9, 87.5, 79.0                          | 90.0, 90.0, 90.0                 | 90.0, 90.0, 90.0                     | 90.0, 90.0, 90.0                              |
| Resolution (Å) <sup>1</sup>          | 73.73-2.50<br>(2.64-2.50)                 | 80.00-2.50<br>(2.64-2.50)        | 53.02-2.30<br>(2.42-2.30)            | 65.00-2.35<br>(2.48-2.35)                     |
| R <sub>merge</sub>                   | 0.121 (0.548)                             | 0.13 (0.549)                     | 0.109 (0.566)                        | 0.147 (0.463)                                 |
| R <sub>pim</sub>                     | 0.104 (0.475)                             | 0.076 (0.327)                    | 0.034 (0.177)                        | 0.059 (0.186)                                 |
| CC(1/2)                              | 0.935 (0.298)                             | 0.987 (0.721)                    | 0.997 (0.916)                        | 0.993 (0.886)                                 |
| <I/σI>                               | 5.7 (2.4)                                 | 7.6 (2.7)                        | 14.8 (4.7)                           | 11.6 (4.2)                                    |
| Completeness (%)                     | 97.8 (97.1)                               | 99.7 (100.0)                     | 100.0 (100.0)                        | 99.6 (98.4)                                   |
| Redundancy                           | 2.2 (2.2)                                 | 3.6 (3.6)                        | 11.2 (11.2)                          | 7.1 (7.0)                                     |
| No. of reflections                   | 303,722                                   | 181,493                          | 395,865                              | 285,291                                       |
| No. Unique                           | 137,926                                   | 50,683                           | 34,836                               | 40,335                                        |
|                                      |                                           |                                  |                                      |                                               |
| <i>Refinement</i>                    |                                           |                                  |                                      |                                               |
| Resolution (Å)                       | 2.50                                      | 2.50                             | 2.30                                 | 2.35                                          |
| R <sub>work</sub> /R <sub>free</sub> | 0.21/0.26                                 | 0.17/0.21                        | 0.17/0.21                            | 0.16/0.20                                     |
| No. of atoms                         |                                           |                                  |                                      |                                               |
| Protein                              | 6508 (A), 6494 (B),<br>6451 (C), 6530 (D) | 2428 (A), 2400 (B),<br>2383 (C)  | 5069                                 | 2978 (A), 2996 (B)                            |
| Ligand                               | 10 (SO4), 28 (EDO)                        | 3 (NA), 15 (IMD)                 | 3 (CA), 6 (GOL), 24<br>(EDO)         | 2 (MG), 44 (EDO),<br>16 (TRS)                 |
| Water                                | 755                                       | 394                              | 261                                  | 590                                           |
| <i>B</i> -factors                    |                                           |                                  |                                      |                                               |
| Protein                              | 29.9 (A), 30.3 (B),<br>35.2 (C), 28.3 (D) | 31.0 (A), 31.2 (B),<br>33.0 (C)  | 43.5                                 | 18.3 (A), 18.4 (B)                            |
| Ligand                               | 56.3 (SO4), 41.2<br>(EDO)                 | 18.0 (NA), 49.2<br>(IMD)         | 38.1 (CA), 58.5<br>(GOL), 53.2 (EDO) | 13.0 (MG), 41.6<br>(EDO), 22.2 (TRS)          |
| Water                                | 22.9                                      | 30.3                             | 42.2                                 | 21.2                                          |
| r.m.s.d                              |                                           |                                  |                                      |                                               |
| Bond lengths (Å)                     | 0.008                                     | 0.012                            | 0.010                                | 0.007                                         |
| Bond angles (°)                      | 1.195                                     | 1.604                            | 1.340                                | 1.296                                         |
| Ramachandran (%)                     |                                           |                                  |                                      |                                               |
| Preferred                            | 95.5                                      | 96.9                             | 95.7                                 | 98.1                                          |
| Allowed                              | 4.5                                       | 3.1                              | 4.1                                  | 1.6                                           |
| Disallowed                           | 0.0                                       | 0.0                              | 0.2                                  | 0.3                                           |
| PDB code                             | 5T9A                                      | 5T9X                             | 5TA1                                 | 5TA7                                          |

<sup>1</sup>Values for highest resolution shells are shown in parenthesis.

146  
147

**Supplementary Table 8: X-ray data collection and statistics of agarase complexes**

|                                      | GH2C GIF complex                          | GH86 E322Q N8 complex                                                      | GH117B N2                                     |
|--------------------------------------|-------------------------------------------|----------------------------------------------------------------------------|-----------------------------------------------|
| <i>Data Collection</i>               |                                           |                                                                            |                                               |
| Beamline                             | SSRL BL7-1                                | CLS - 08ID-1                                                               | CLS - 08ID-1                                  |
| Wavelength (Å)                       | 0.9753                                    | 0.9840                                                                     | 0.97949                                       |
| Space Group                          | P1                                        | P1                                                                         | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell Dimensions                      |                                           |                                                                            |                                               |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)   | 75.2, 121.1, 124.5                        | 60.7, 73.5, 83.2                                                           | 83.5, 104.4, 198.5                            |
| $\alpha$ , $\beta$ , $\gamma$ (°)    | 70.7, 86.8, 78.8                          | 85.9, 86.2, 71.9                                                           | 90.0, 90.0, 90.0                              |
| Resolution (Å) <sup>1</sup>          | 73.79-2.40 (2.53-2.40)                    | 48.39-1.40 (1.48-1.40)                                                     | 76.97-2.40 (2.53-2.40)                        |
| R <sub>merge</sub>                   | 0.142 (0.638)                             | 0.070 (0.324)                                                              | 0.160 (0.540)                                 |
| R <sub>pim</sub>                     | 0.085 (0.383)                             | 0.029 (0.136)                                                              | 0.059 (0.193)                                 |
| CC(1/2)                              | 0.990 (0.786)                             | 0.998 (0.952)                                                              | 0.989 (0.825)                                 |
| $\langle I/\sigma \rangle$           | 8.1 (2.0)                                 | 15.7 (5.4)                                                                 | 8.2 (3.6)                                     |
| Completeness (%)                     | 87.6 (82.5)                               | 95.7 (93.4)                                                                | 96.0 (90.0)                                   |
| Redundancy                           | 3.6 (3.4)                                 | 6.6 (6.6)                                                                  | 8.3 (8.6)                                     |
| No. of reflections                   | 504,239                                   | 1,705,176                                                                  | 543,181                                       |
| No. Unique                           | 139,311                                   | 256,936                                                                    | 65,697                                        |
|                                      |                                           |                                                                            |                                               |
| <i>Refinement</i>                    |                                           |                                                                            |                                               |
| Resolution (Å)                       | 2.40                                      | 1.40                                                                       | 2.40                                          |
| R <sub>work</sub> /R <sub>free</sub> | 0.21/0.26                                 | 0.14/0.16                                                                  | 0.21/0.25                                     |
| No. of atoms                         |                                           |                                                                            |                                               |
| Protein                              | 6533 (A), 6496 (B),<br>6554 (C), 6547 (D) | 5413 (A), 5292 (B)                                                         | 2970 (A), 2963 (B),<br>2935 (C), 2930 (D)     |
| Ligand                               | 40 (GIF), 56 (EDO)                        | 40 (SO <sub>4</sub> ), 120 (EDO), 4<br>(CA), 54 (N8),<br>54 (GOL)          | 48 (GAL), 44 (AAL),<br>4 (MG)                 |
| Water                                | 851                                       | 1312                                                                       | 447                                           |
| <i>B</i> -factors                    |                                           |                                                                            |                                               |
| Protein                              | 30.8 (A), 33.8 (B),<br>30.5 (C), 30.5 (D) | 12.3 (A), 13.5 (B)                                                         | 37.3 (A), 36.0 (B),<br>48.4 (C), 49.3 (D)     |
| Ligand                               | 29.0 (GIF), 39.7 (EDO)                    | 34.4 (SO <sub>4</sub> ), 28.8 (EDO),<br>10.4 (CA) 21.6 (N8), 29.8<br>(GOL) | 41.2 (GAL), 51.7 (AAL),<br>26.0 (MG)          |
| Water                                | 22.2                                      | 25.8                                                                       | 32.0                                          |
| r.m.s.d                              |                                           |                                                                            |                                               |
| Bond lengths (Å)                     | 0.009                                     | 0.010                                                                      | 0.008                                         |
| Bond angles (°)                      | 1.292                                     | 1.488                                                                      | 1.304                                         |
| Ramachandran (%)                     |                                           |                                                                            |                                               |
| Preferred                            | 96.3                                      | 96.7                                                                       | 96.8                                          |
| Allowed                              | 3.6                                       | 3.1                                                                        | 3.1                                           |
| Disallowed                           | 0.1                                       | 0.2                                                                        | 0.1                                           |
| PDB code                             | 5T9G                                      | 5TA0                                                                       | 5TA9                                          |

<sup>1</sup>Values for highest resolution shells are shown in parenthesis.

148  
149

## 150      **Supplementary References**

- 151      1.      Loytynoja A & Goldman N (2005) An algorithm for progressive multiple alignment of sequences  
152              with insertions. *Proceedings of the National Academy of Sciences of the United States of America*  
153              102(30):10557-10562.
- 154      2.      Pluvinae B, Hehemann JH, & Boraston AB (2013) Substrate recognition and hydrolysis by a  
155              family 50 exo-beta-agarase, Aga50D, from the marine bacterium *Saccharophagus degradans*.  
156              *The Journal of biological chemistry* 288(39):28078-28088.
- 157      3.      Rebuffet E, *et al.* (2011) Discovery and structural characterization of a novel glycosidase family  
158              of marine origin. *Environmental microbiology* 13(5):1253-1270.
- 159      4.      Gasteiger E, *et al.* (2003) ExPASy: The proteomics server for in-depth protein knowledge and  
160              analysis. *Nucleic acids research* 31(13):3784-3788.
- 161      5.      Jackson P (1993) Fluorophore-assisted carbohydrate electrophoresis: a new technology for the  
162              analysis of glycans. *Biochemical Society transactions* 21(1):121-125.
- 163      6.      Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta*  
164              *crystallographica. Section D, Biological crystallography* 55(Pt 10):1690-1695.
- 165      7.      Collaborative Computational Project N (1994) The CCP4 suite: programs for protein  
166              crystallography. *Acta crystallographica. Section D, Biological crystallography* 50(Pt 5):760-763.
- 167      8.      McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of applied crystallography*  
168              40(Pt 4):658-674.
- 169      9.      Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein  
170              chains. *Acta crystallographica. Section D, Biological crystallography* 62(Pt 9):1002-1011.
- 171      10.      Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta*  
172              *crystallographica. Section D, Biological crystallography* 60(Pt 12 Pt 1):2126-2132.
- 173      11.      Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the  
174              maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* 53(Pt  
175              3):240-255.
- 176      12.      Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal  
177              structures. *Nature* 355(6359):472-475.
- 178      13.      Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and  
179              nucleic acids. *Nucleic acids research* 35(Web Server issue):W375-383.
- 180      14.      Chen VB, *et al.* (2010) MolProbity: all-atom structure validation for macromolecular  
181              crystallography. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 1):12-21.
- 182      15.      Hehemann JH, Kelly AG, Pudlo NA, Martens EC, & Boraston AB (2012) Bacteria of the human gut  
183              microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from  
184              extrinsic microbes. *Proceedings of the National Academy of Sciences of the United States of*  
185              *America* 109(48):19786-19791.
- 186      16.      Hehemann JH, Smyth L, Yadav A, Vocadlo DJ, & Boraston AB (2012) Analysis of keystone enzyme  
187              in Agar hydrolysis provides insight into the degradation (of a polysaccharide from) red  
188              seaweeds. *The Journal of biological chemistry* 287(17):13985-13995.
- 189      17.      Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-  
190              active enzymes database (CAZy) in 2013. *Nucleic acids research* 42(Database issue):D490-495.
- 191      18.      Yin Y, *et al.* (2012) dbCAN: a web resource for automated carbohydrate-active enzyme  
192              annotation. *Nucleic acids research* 40(Web Server issue):W445-451.
- 193      19.      Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.  
194              *Nucleic acids research* 32(5):1792-1797.

- 195 20. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large  
196 phylogenies. *Bioinformatics* 30(9):1312-1313.
- 197 21. Darriba D, Taboada GL, Doallo R, & Posada D (2011) ProtTest 3: fast selection of best-fit models  
198 of protein evolution. *Bioinformatics* 27(8):1164-1165.
- 199 22. Abbott DW & Boraston AB (2012) Quantitative approaches to the analysis of carbohydrate-  
200 binding module function. *Methods in enzymology* 510:211-231.
- 201 23. Hellemans J, Mortier G, De Paepe A, Speleman F, & Vandesompele J (2007) qBase relative  
202 quantification framework and software for management and automated analysis of real-time  
203 quantitative PCR data. *Genome biology* 8(2):R19.
- 204 24. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space  
205 complexity. *BMC bioinformatics* 5:113.
- 206 25. Lee CH, *et al.* (2014) A novel agarolytic beta-galactosidase acts on agarooligosaccharides for  
207 complete hydrolysis of agarose into monomers. *Applied and environmental microbiology*  
208 80(19):5965-5973.
- 209 26. Juncker AS, *et al.* (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria.  
210 *Protein science : a publication of the Protein Society* 12(8):1652-1662.
- 211 27. Seydel A, Gounon P, & Pugsley AP (1999) Testing the '+2 rule' for lipoprotein sorting in the  
212 *Escherichia coli* cell envelope with a new genetic selection. *Molecular microbiology* 34(4):810-  
213 821.
- 214 28. Yun EJ, *et al.* (2015) The novel catabolic pathway of 3,6-anhydro-L-galactose, the main  
215 component of red macroalgae, in a marine bacterium. *Environmental microbiology* 17(5):1677-  
216 1688.