Exploring the Glycans of Euglena Gracilis

Ellis C. O'Neill,[1]* Sakonwan Kuhaudomlarp,[2] Martin Rejzek,[2] Jonatan U. Fangel,[3]

Kathirvel Alagesan,[4,5] Daniel Kolarich,[4,5] William G. T. Willats,[6] Robert A. Field[2]*

[1] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK.

[2] Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

[3] Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg 1871, Denmark

[4] Institute for Glycomics, Gold Coast Campus, Griffith University, Southport, QLD, 4222, Australia

[5] Department of Biomolecular Sciences, Max Planck Institute of Colloids and Interfaces, 14424 Potsdam, Germany

[6] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

* Correspondence: *Email: ellis.oneill@plants.ox.ac.uk. Tel: +44 (0)1865 275130; Email: rob.field@jic.ac.uk. Tel.: +44 (0) 1603 450720

Abstract: Euglena gracilis is an alga of great biotechnological interest and extensive metabolic capacity, able to make high levels of bioactive compounds, such as polyunsaturated fatty acids, vitamins and β-glucan. Previous work has shown that Euglena expresses a wide range of carbohydrate-active enzymes, suggesting an unexpectedly high capacity for the synthesis of complex carbohydrates for a single celled organism. Here we present an analysis of some of the carbohydrates synthesised by Euglena gracilis. Analysis of the sugar nucleotide pool showed that there are the substrates necessary for synthesis of complex polysaccharides, including the unusual sugar galactofuranose. Lectin- and antibody-based profiling of whole cells and extracted carbohydrates revealed a complex galactan, xylan and aminosugar based surface. Protein N-glycan profiling, however, indicated that just simple high mannose-type glycans are present and that they are partially modified with putative aminoethylphosphonate moieties. Together, these data indicate that Euglena possesses a complex glycan surface, unrelated to plant cell walls, while its' protein glycosylation is simple. Taken together, these findings suggest that Euglena gracilis may lend itself to production of pharmaceutical glycoproteins.

Keywords: algae; Euglena; biotechnology; carbohydrates; N-glycan; sugar nucleotide

1. Introduction

Euglena are a group of fast growing, mainly freshwater algae, that are distantly related to other algae. They have long been studied for both their fundamental biology and for their production of a range of high value products, including vitamins, amino acids and polyunsaturated fatty acids.[1] As they are easy to grow and have a large metabolic capacity, Euglena have been proposed as a platform for the production of proteins and high value metabolites of biotechnological interest [2] and for the synthesis of complex natural products.[3]

The genome sequence of Euglena is not yet available, but *de novo* transcriptome sequencing of the most heavily studied species, *Euglena gracilis*, has unveiled the huge metabolic capacity of these organisms. In fact, Euglena possesses more genes than are evident in the human genome.[4] Of particular note is the large number of carbohydrate-active enzymes (CAZymes), comparable in number to multicellular animals, and much higher than other single-celled algae. However, predicting the specificity of CAZymes is notoriously difficult and it is not currently possible to predict the structure of the carbohydrates made from enzyme sequences alone.

Euglena is not reported to have a carbohydrate-based cell wall, but glucose, galactose, mannose, fucose, xylose, and rhamnose have been detected in cell surface extracts by paper chromatography.[5] It has also been reported that Euglena cells can become encased in a carbohydrate sheath, forming cysts.[6] Euglena has a full complement of genes for the biosynthesis of glycosylphosphatidylinositol

49 (GPI) membrane anchors which may be involved in anchoring these glycans to internal membranes
50 or the cell surface.[4] *N*-Acetylglucosamine-1-phosphate transferase activity has been detected in
51 membrane preparations of Euglena cells [7] and a complex xylose-containing material has been found
52 associated with the flagella.[8] A wide range of glycosyltransferases are present in the Euglena
53 transcriptome, supporting the capacity for synthesis of complex glycans, although the exact structure
54 of such glycans remains to be elucidated.

55   Protein glycosylation appears to be fairly typical in Euglena, synthesising the same *N*-glycan
56 precursor oligosaccharide as animals and fungi (Glc$_3$Man$_9$GlcNAc$_2$-Asn),[9] with sequences for all of
57 the enzymes required for the synthesis present in the transcriptome.[4] There are three sequences for
58 the oligosaccharyltransferases that couple this pre-formed oligosaccharide to proteins, in common
59 with *Trypanosomes* where the enzymes have different substrate specificities.[10] In Euglena, flagella-
60 associated glycoproteins are affected by tunicamycin, an inhibitor of protein glycosylation.[11]
61 However, the exact structure of the glycans and the identity of glycosylated proteins remain elusive.
62 There are also three members of GT41 family of protein glycosyltransferases present in the
63 transcriptome,[12] more than in animals, which have one protein *O*-GlcNAc transferase [13] or plants,
64 which have both an *O*-GlcNAc transferase and an *O*-fucosyltransferase in this family.[14]

65   The range of tantalising previous work and the rather high number of glycosyltransferases led
66 us to speculate that Euglena has great capacity for the synthesis of complex glycans and we were
67 therefore minded to investigate the carbohydrates synthesised by these organisms.
68

69 **2. Materials and Methods**

70 *Growth of Euglena*

71   *Euglena gracilis* var. *saccharophila* Klebs (strain 1224/7a) was obtained from the Culture Collection
72 of Algae and Protozoa (CCAP) (http://www.ccap.ac.uk/) and treated with antibiotics.[12] For
73 heterotrophic cultures, cells were grown in the dark in EG:JM media containing additional glucose
74 (15 g/L) at 30 °C. For photoheterotrophic cultures, cells were grown in the same media at 22 °C on a
75 14:10 light cycle with a light intensity of 100 µmol.m^{-2}.s^{-1}. For phototrophic cultures cell were grown
76 in JM media in the same growth conditions. All cultures were shaken at 150 rpm.

77 *Lectin labelling*

78   Heterotrophic cells were collected at mid log phase, resuspended in the recommended lectin
79 buffers and incubated with the appropriate lectins (Vector labs). For aggregation, cells were then
80 visualised after 1 hour. For the fluorescently labelled samples, cells were washed with fresh buffer
81 and then visualised using a Leica DM 6000 equipped with a DFC420 camera. 450-490 nm and 515-560
82 nm excitation were used for visualising the fluorescent lectins.

83 *Sugar nucleotide profiling*

84   The methods used were essentially as reported by Rejzek et al.[33] and Wagstaff et al.[34] In
85 brief, mid-log phase (OD$_{600}$ = 1.1 in about 6 days) cultures were harvested (3 biological replicates).
86 UDP-α-D-GlcNAcA was added to the cell pellet as internal standard (1.46 nmol / g wet pellet). Cell
87 lysis was performed with cold (-20 °C) 70% ethanol (20 mL). The cell debris was removed by
88 centrifugation (28 928 g, 20 min, 4 °C) and the supernatant was concentrated to dryness. Lipophilic
89 components were removed by partitioning the sample between water and butan-1-ol.[35] Solid phase
90 extraction (SPE) on a graphitised carbon column (EnviCarb, Supelco, 250 mg, 3 mL) was performed
91 essentially as described by Rabina and co-workers.[18] LC-MS/MS profiling of sugar nucleotides was
92 performed on a Xevo TQ-S tandem quadrupole mass spectrometer (Waters) operated in multiple
93 reaction monitoring (MRM) mode coupled to an Acquity UPLC. MRM transitions for sugar
94 nucleotide standards in negative ESI mode were generated using IntelliStart software (Table S2).
95 Samples (10 µM) were introduced at 10 µL/min combined with a flow from the HPLC pump typical
96 of an LC run. MassLynx software (Waters) was used to collect, to analyse and to process data.

97   Separation of sugar nucleotides was achieved on a surface-conditioned PGC column (Hypercarb,
98   Thermo Scientific, dimensions 1 x 100 mm, particle size 5 μm) equipped with a column guard
99   (Hypercarb, 5 μm, 1 x 10 mm).[19] Analytes were eluted using mobile phase A: formic acid 0.3%
100 brought to pH 9.0 with ammonia and mobile phase B: acetonitrile using the following multistep
101 gradient at a flow rate 80 μL/min: 0 min: 2% B; 20 min: 15% B; 26 min: 50% B; 27 min: 90% B; 30 min:
102 90% B; 31 min: 2% B; 50 min: 2% B. Available sugar nucleotide standards (10 μM) were injected (5
103 μL) to determine retention time (Table S2). UDP-α-D-Glc, UDP-α-D-GlcNAc, UDP-α-D-GlcA, UDP-
104 α-D-Gal, TDP-α-D-Glc, ADP-α-D-Glc, ADP-D-Rib, GDP-α-D-Man and GDP-β-L-Fuc were obtained
105 commercially from Sigma Aldrich (Haverhill, UK). UDP-α-D-GlcNAcA,[36] TDP-β-L-Rha,[34] UDP-
106 β-L-Rha,[34] UDP-α-D-Xyl[37] and UDP-β-L-Ara*p*[38] were prepared as previously described.
107 Although between runs there were significant differences in absolute retention times of standards,
108 relative retentions were reasonably reproducible (Table S2). Using a serial dilution of UDP-α-D-Glc
109 limit of detection was determined to be 10 fmol on column. Where in doubt, co-injection of sample
110 with appropriate standard sugar nucleotide was used for positive identification.

111 *Immuno-carbohydrate microarray profiling*

112 The samples were snap frozen in liquid nitrogen and ground with steel balls in a tissue lyser
113 (Qiagen) for 1 min at 30 s^{-1} frequency. AIR was isolated by washing the homogenate with 70%
114 ethanol followed by drying. Sequential extraction was carried out as described in Moller et. al.[24]
115 Briefly, 300 μL of CDTA was added to 10.0 mg of material, and the microtubes were shaken at RT for
116 2 h. The samples were centrifuged at 4,000 *g* for 10 min, and the supernatant was removed and
117 pipetted to a new microtube. NaOH at 4 M + 0.1% (w/v) NaBH$_4$ was added to the sediment and
118 shaken for 2 h. The samples were centrifuged at 4,000 *g* for 10 min, and the supernatant was removed
119 and pipetted to new microtubes. The printing on the nitrocellulose membrane was done on an
120 ArrayJet Sprint printer in two replicates and four dilutions, probed, and quantified as described
121 by Moller et al.[39]

122 *N- and O-Glycan analysis*

123 Sample preparation

124 Euglena cells were grown in EG:JM+Glc media in the dark and the flagella isolated by cooling
125 to 4°C for 1 hour prior to centrifugation at 800 g to remove the cells and at 20,000 g to collect the
126 flagella. About 5 μg of the obtained flagellar protein purified from *Euglena gracilis* was dot-blotted on
127 to the PVDF membrane. *N*- and *O*-glycans were released from the dot-blotted protein as described
128 by Jensen et al.[25] Briefly, The *N*-linked glycans were released by incubation with PNGase F (3 U)
129 overnight at 37°C. The released *N*-linked glycans were isolated, dried and reduced with 20 μL of 1 M
130 NaBH$_4$ in 50 mM KOH at 50°C for 3 h. The reduction was quenched with 1 μL glacial acetic acid and
131 *N*-linked glycans were purified using cation exchange columns made from 30 μL AG50W-X8 cation-
132 exchange resin (BioRad, Hercules, CA) packed on μC18 ZipTips. The residual borate was removed
133 by addition of methanol (200 μL) and dried under vacuum. *O*-linked glycans were then released from
134 the N-glycan liberated sample by reductive β-elimination performed overnight in 20 μL of 0.5 M
135 NaBH$_4$ dissolved in 50 mM KOH at 50°C. The reduction was quenched with 1 μL glacial acetic acid
136 and *O*-linked glycans were desalted using the cation exchange columns as described above for the *N*-
137 glycans. The purified glycans were then further purified by PGC-SPE and vacuum dried prior PGC
138 LC ESI MS/MS.

139 Mass spectrometric analysis

140 PGC-LC was performed on an Ultimate 3000 UHPLC system (Dionex, Part of Thermo Fisher,
141 Germany) online coupled to an amaZon speed ETD ion trap mass spectrometer (Bruker Daltonics,
142 Bremen, Germany). Dried samples (*N*- and *O*-glycans) were dissolved in 6 μL water and 5 μL were
143 used for separation by PGC-LC (Hypercarb, 30 x 0.32 mm, 5 μm particle size (pre-column), 100 mm
144 x 180 μm, 3 μm particle size (separation column), Thermo Hypersil, Runcorn UK). The sample was

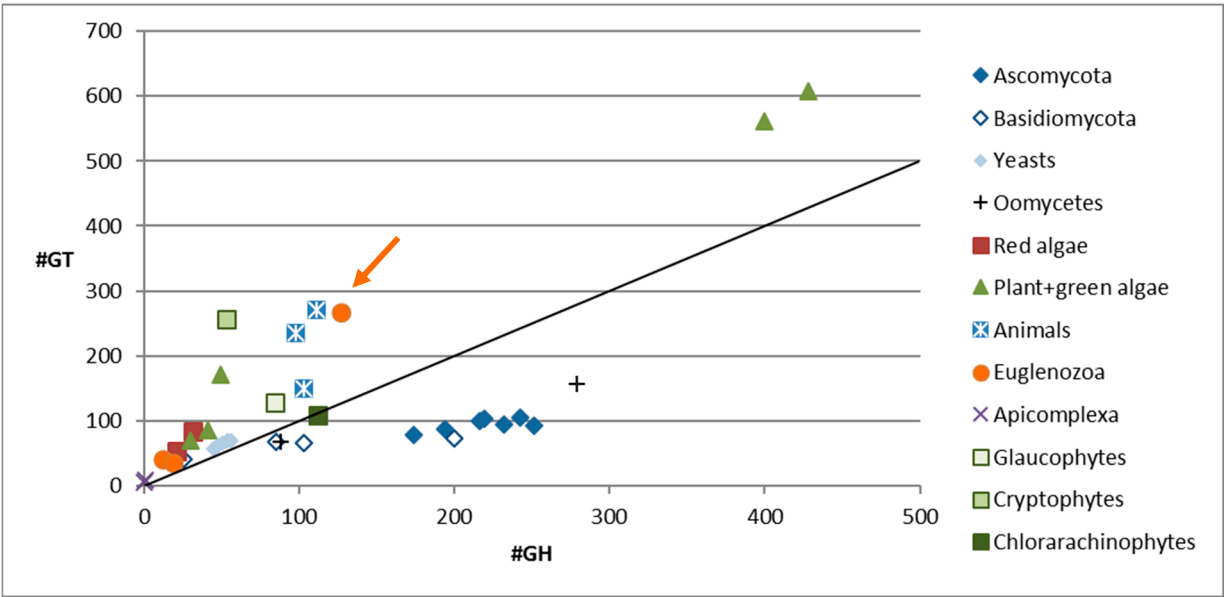
145 loaded on the pre-column at a flow rate of 6 µL/min in buffer A (10 mM ammonium bicarbonate) for
146 5 min before a gradient of 5.8% to 30.3% buffer B (60% ACN in 10 mM ammonium bicarbonate) was
147 applied over 54 min at a flow-rate of 1 µL/min. The column oven temperature was set to 35°C. The
148 LC system was connected to the mass spectrometer using the nano-flow ESI sprayer (Bruker
149 Daltonics, Bremen, Germany). The MS settings were as follows: capillary outlet 1100 V. Negative ion
150 mode MS spectra were obtained within a mass range of 380 – 1800 m/z. Smart parameter setting (SPS)
151 was set to 900 m/z. For MS/MS precursor selection the three most intense ions, including singly
152 charged ones, above the absolute intensity of 34.000 and 20% relative intensity threshold were
153 isolated with a width of 1.0 m/z. The ICC was set to 40.000 with a maximum accumulation time of
154 200 ms. CID was performed using helium as collision gas. The fragmentation amplitude was set to
155 100% using in addition SmartFragTM Enhanced for amplitude ramping (30 - 120%). Fragmentation
156 time was set to 32 ms. The instrument was controlled using Hystar 3.2 software. MS and MS/MS data
157 were processed using DataAnalysis 4.1. The glycan peaks were semi-quantified using the extracted
158 ion chromatogram peak areas.

3. Results

159

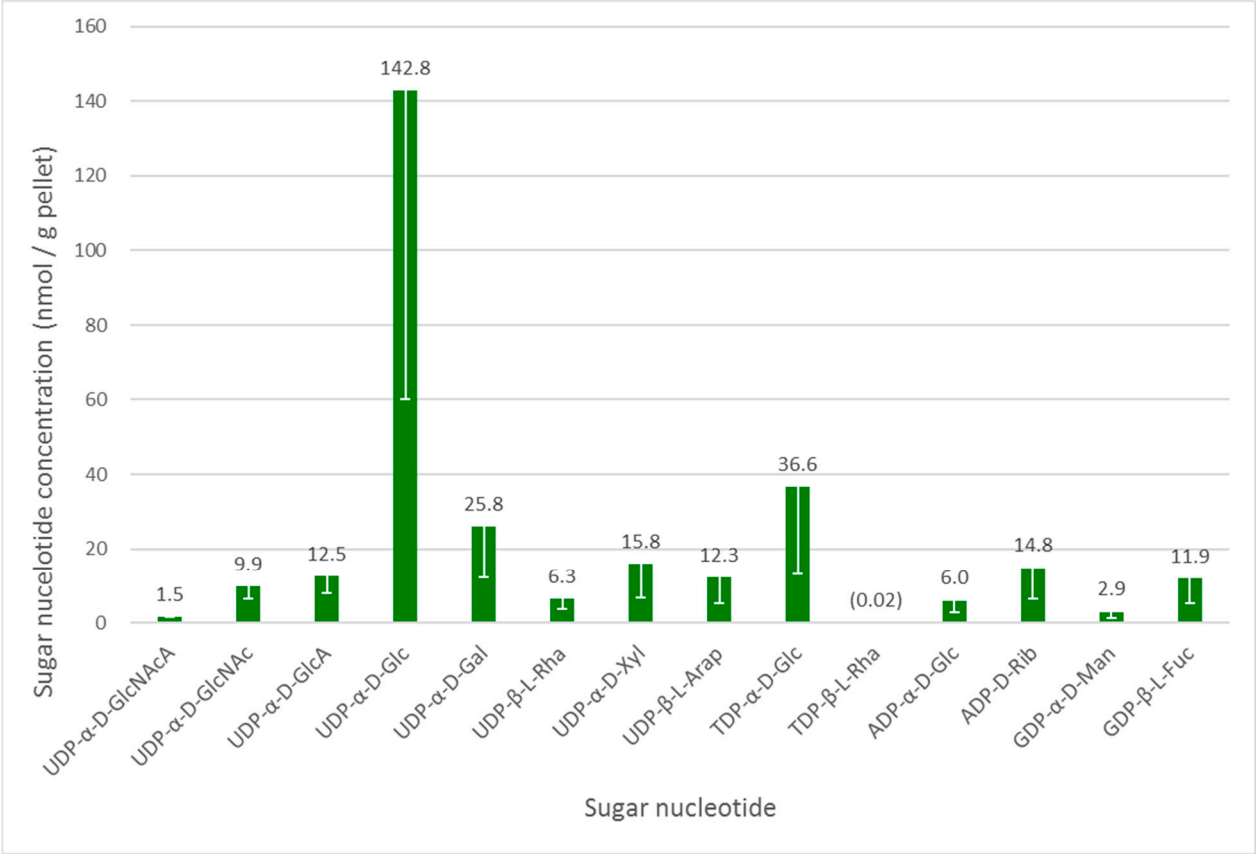
160 *3.1 Carbohydrate active enzymes in Euglena*

161 Inspection of the transcriptome of Euglena showed that there are many more CAZymes encoded
162 [4] than in the other sequenced Euglenozoa, such as the human pathogens *Trypanosoma brucei* and
163 *Leishmania braziliensis* (see Figure 1). There are also more CAZymes than in most other sequenced
164 algae, but fewer than in the land plants, which require complex cell walls to support their growth.
165 One major change is the loss of enzymes involved in α-glucan biosynthesis and an increase in those
166 involved in synthesis of β-glucans.[4] This reflects the use of paramylon, an insoluble β-1,3-
167 glucan,[15] as the storage polysaccharide, as opposed to the α-1,4/6-linked glucan used as the energy
168 store in plants, animals and bacteria (starch or glycogen, respectively). Most β-glucans found in
169 nature are either branched or mixed linkage, such as the 1,3 and 1,6 linkages in fungal cell walls and
170 the storage polysaccharides of sea weeds, or 1,3 and 1,4 linkages found in cereals, such as oat.[16] In
171 contrast, the pure 1,3 linkage of the Euglena paramylon makes it an ideal model system to study the
172 synthesis of this type of carbohydrate.

173

174 **Figure 1: The number of carbohydrate active enzymes annotated in the genomes of selected
175 organisms**. Most organisms have more glycosyltransferases then glycoside hydrolases, except
176 amongst the saprophytic fungi and oomycetes. *Euglena gracilis* is indicated with an arrow. See Table
177 S1 for a breakdown of the CAZyme families encoded in the Euglena transcriptome.

178   *3.2 Sugar nucleotide profiling in Euglena*

179   As in other organisms, the glycosylation machinery in Euglena depends on the availability of
180 activated sugar building blocks, namely the sugar nucleotides.[17] Euglena cell cultures grown under
181 phototrophic conditions were analysed to determine the range of these substrates available. To
182 ensure reproducibility, mid-log phase cultures were harvested and an internal standard was added
183 at this stage to enable quantification. UDP-α-D-GlcNAcA was selected as internal standard, based on
184 the fact that it was absent in the Euglena extracts and that it did not co-elute (see Table S2) with any
185 significant analytes. The intracellular sugar nucleotides were extracted by cell lysis with cold 70%
186 ethanol, followed by removal of lipophilic components and solid phase extraction on a graphitised
187 non-porous carbon column, as described previously.[18] Liquid chromatography separation of sugar
188 nucleotides was achieved on a surface-conditioned porous graphitic carbon (PGC) column using a
189 gradient of acetonitrile against 0.3% (v/v) formic acid adjusted to pH 9.0 with ammonia.[19] Sugar
190 nucleotides were detected using negative mode electrospray ionisation tandem mass spectrometry
191 (ESI-MS/MS). The identity of all nucleotide diphosphate sugar (NDP-sugar) species was confirmed
192 by comparison with authentic standards (see Figure 2).

193

194   **Figure 2: Intracellular sugar nucleotide profile of Euglena gracilis grown under phototrophic**
195   **conditions.** An internal standard UDP-α-D-GlcNAcA (1.46 nmol / g pellet) was added to the samples.
196   The data are mean of three biological replicates, error bars (negative value only) indicate standard
197   error. The identity of all NDP-sugar species was confirmed using authentic standards.

198

199   UDP-α-D-Glc is clearly the most abundant sugar nucleotide in these cells, consistent with its role
200 in the biosynthesis of the major storage polysaccharide paramylon and its' central role as a
201 biosynthetic precursor to other sugar nucleotides.[17] There is a much smaller amount of ADP-α-D-
202 Glc, which is used in the biosynthesis of α-glucans, such as starch and glycogen, in other organisms.
203 These glycans are absent from Euglena, but this sugar nucleotide may be used in the biosynthesis of
204 trehalose.[20]

205 As mentioned above, Euglena is known to make *N*-glycans in a similar manner to other
206 Eukaryotes. The biosynthesis of the conserved Man5GcNAc2 core requires the transfer of GlcNAc,
207 Man and Glc from UDP-α-D-GlcNAc, GDP-α-D-Man and UDP-α-D-Glc respectively, all of which can
208 be detected in the cell extracts. These glycans can be further elaborated by the addition of further
209 GlcNAc and Gal residues, with the latter coming from the UDP-α-D-Gal, which is present in quite
210 high concentrations in these cells. However, these modifications were not detected in the present
211 protein-bound glycan analysis (*vide infra*).
212 The surface of Euglena is known to contain glucose, galactose, mannose, fucose, xylose, and
213 rhamnose [5] and the corresponding sugar nucleotides are present for all of these. Indeed, both UDP-
214 β-L-Rha and trace amounts of TDP-β-L-Rha were detected, confirmed against authentic standards by
215 LCMS. Although TDP-α-D-Glc is a precursor of the latter nucleotide, the detected high TDP-α-D-Glc
216 concentrations indicate it may be used for the synthesis of alternative carbohydrates, or it may be
217 used interchangeably with UDP-α-D-Glc. The presence of UDP-β-L-Ara*p* and UDP-α-D-GlcA at
218 similar concentrations to GDP-β-L-Fuc and UDP-α-D-Xyl indicate there may also be capacity to
219 synthesise significant quantities of complex glycans in Euglena.
220 The other major use for sugar nucleotides is in the biosynthesis of the antioxidant ascorbate, for
221 which there are three different pathways in Euglena.[4] UDP-α-D-Glc, UDP-α-D-GlcA and GDP-α-D-
222 Man could all be detected, but UDP-α-D-GalA and GDP-β-L-Gal were not found, despite being
223 compared to authentic standards. The other notable sugar nucleotide detected was ADP-D-Rib, which
224 may be involved ADP ribosylation of proteins [21] or as part of the NAD recycling pathway.[22]
225

3.3 Glycan analysis of Euglena

3.3.1 Surface glycans

228 In order to investigate the carbohydrates present on the surface of Euglena, lectins were used as
229 probes. Initially a panel of unlabelled lectins were incubated with Euglena and cell aggregation was
230 monitored by eye and then by light microscopy (see Figure 3a). It was clear from this that the *Ricinus*
231 *communis* agglutinin (RCA) and an antibody raised against horseradish peroxidase (HRPAb) show
232 reactivity. These agents recognise galactose or *N*-acetylgalactosamine (GalNAc) and fucose or xylose,
233 respectively. The absence of aggregation with Concanavalin A (ConA), *Sambucus nigra* agglutinin
234 (SNA) or Wheat Germ Agglutinin (WGA) indicates that there is insufficient glucose- or mannose-,
235 sialic acid- or GlcNAc-bearing oligosaccharide accessible on the surface to cross-link the cells. The
236 small amount of crosslinking from *Ulex europaeus* agglutinin (UEA) indicates there may be some
237 fucose displayed on the surface of Euglena cells.
238 In order to further probe Euglena cell surface glycans, fluorescently labelled lectins were
239 employed, in the presence or absence of competing ligands (see Figure 3B). The strong labelling with
240 RCA and *Erythrina cristagalli* lectin (ECL) indicated that there are likely to be either GalNAc or Gal
241 and Gal-β-1,4-GlcNAc moieties on the cell surface. Consistent with these observations, weak labelling
242 with WGA indicated that there may be some GlcNAc present. ConA labelling was only apparent in
243 cells that were disrupted and lysed, with the lectin probably labelling the paramylon storage
244 carbohydrate (ConA binds β-glucans, but with a 28 fold lower affinity than for α-glucans [23]).
245 Additionally, some cells in the last stages of mitosis appeared to have some labelling adjacent to the
246 point of cell separation or at the base of the flagella (see Figure 3C), indicating there may be some α-
247 Glc or α-Man residues accessible at the fission point.

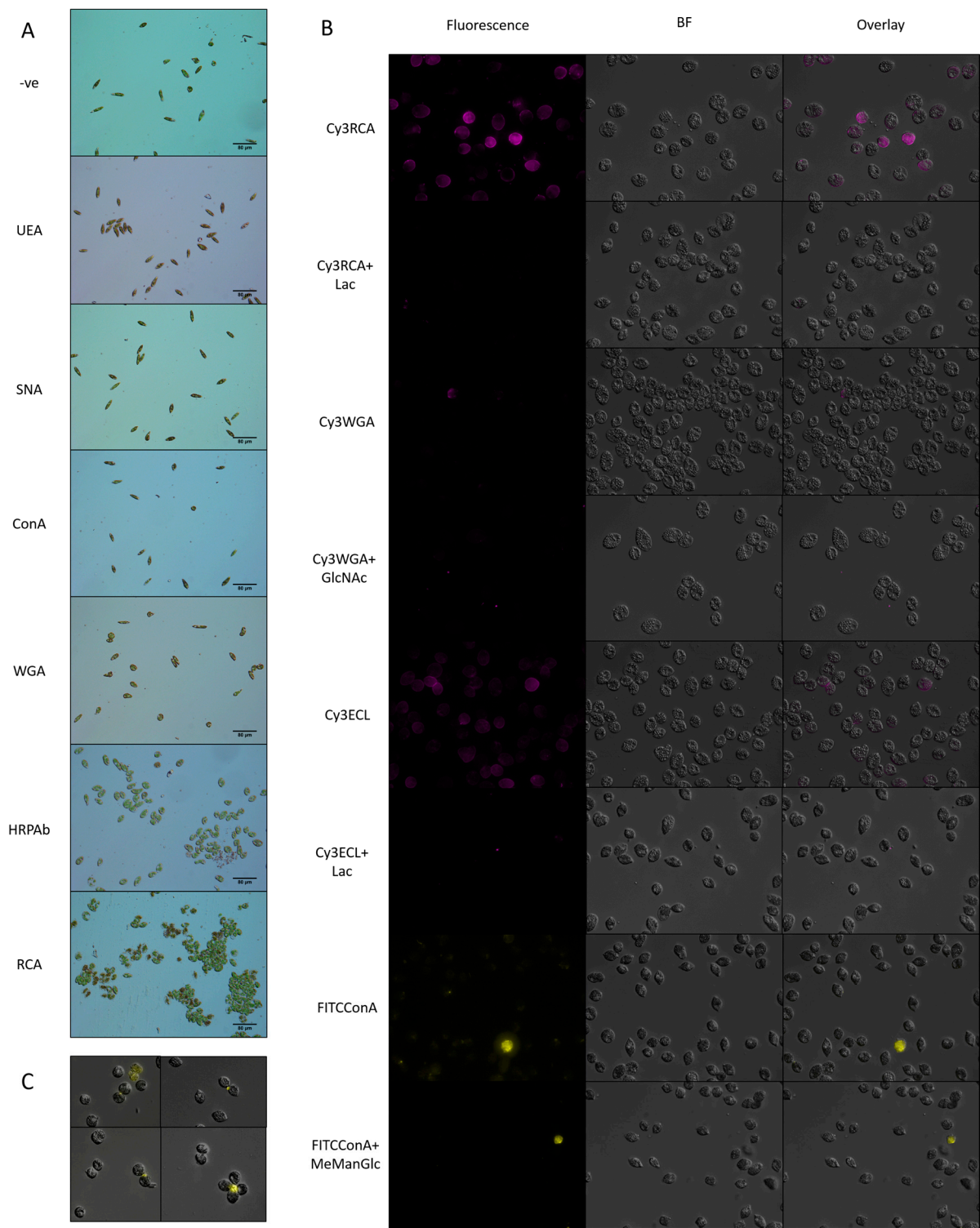
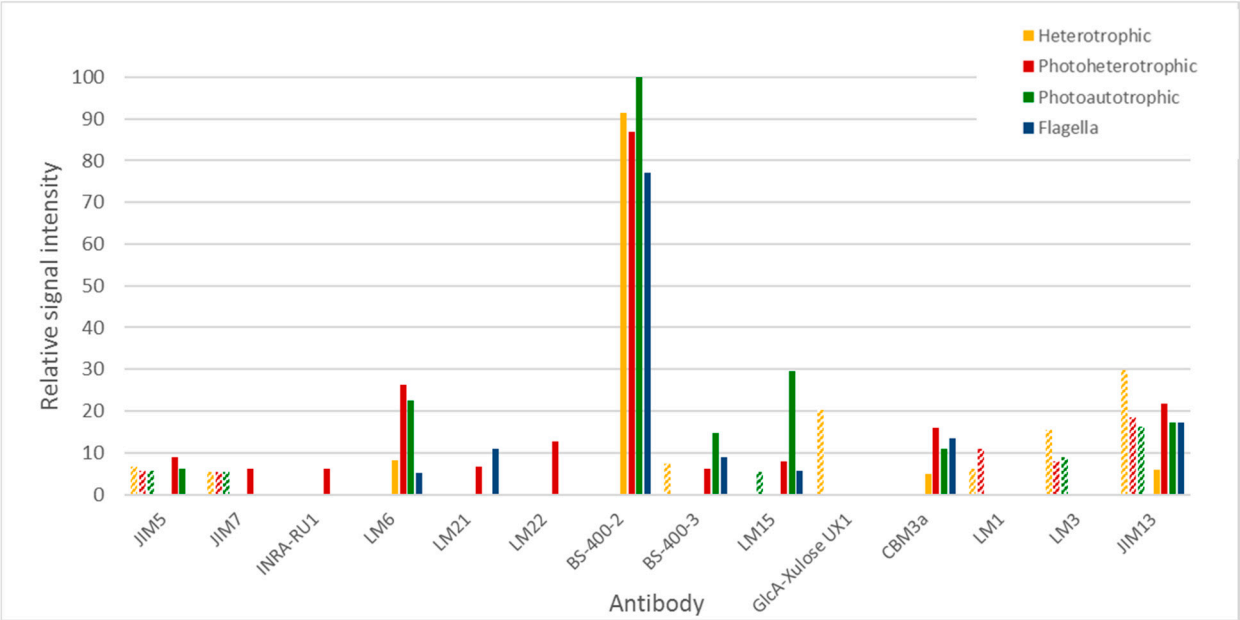



Figure 3: Labelling of Euglena cells with lectins. A. Cells of Euglena were mixed with a variety of lectins and allowed to precipitate. They were then imaged using bright field microscopy. B. Cells of Euglena were labelled with various fluorescently labelled lectins and imaged using bright field microscopy (BF) and fluorescence microscopy at the appropriate wavelength. C. Representative images of dividing Euglena cells labelled with FITC-ConA.

3.3.2 Immunocarbohydrate microarray profiling

Immunocarbohydrate microarray profiling is a semi-quantitative high-throughput method whereby cell wall polysaccharide components are extracted and then printed as microarrays onto membranes and probed with panels of monoclonal antibodies and/or carbohydrate-binding modules

258 (CBMs).[24] This provides information about the relative abundance of polysaccharide epitopes and
259 is useful for obtaining a comparative overview of glycans across sample sets. Euglena cells grown
260 under three different conditions (phototrophic, mixotrophic and heterotrophic) were analysed, as
261 well as flagella isolated from heterotrophically grown cells. Two extraction solvents were used
262 sequentially, first 1,2-diaminocyclohexanetetraacetic acid (CDTA) and then NaOH. As these
263 extraction methods are optimised for their use on land plant cell wall materials the data requires
264 careful interpretation (see Figure 4).
265   From the milder CDTA treatment, it is apparent that no carbohydrates were released from the
266 flagella material. CDTA treatment of the cells released material contained glycan epitopes recognised
267 by anti-homogalacturonan, anti-extensin and anti-AGP antibodies (mAbs JIM5 and JIM7, LM1 and
268 LM3, and JIM13 respectively). Heterotrophic cells also released material with a relatively high
269 abundance of epitopes recognised by anti-glucuronoxylan and anti-mixed linkage glucan mAbs
270 (mAb GlcA-Xylose UX1 and BS-400-3 respectively), and mixotrophic cells released material
271 containing the xyloglucan epitope recognised by mAb LM15. (see Figure 4). The material released by
272 NaOH had a strikingly different profile to that released by CDTA.   The most notable signal was that
273 obtained from the anti-β-1,3-glucan mAb BS-400-2 and this binding was most likely due to the
274 presence of the major storage polysaccharide, paramylon. In addition, NaOH also released material
275 containing epitopes recognised by cellulose binding CBM3a, and the anti-arabinan and anti-
276 arabinogalactan mAbs LM6 and JIM13 respectively. It is also worth noting that distinctly different
277 profiles were obtained for photoheterotrophic and photoautotrophic cells. For example, in NaOH
278 released material, photoheterotrophic cells contained more mannan (recognised by mAbs LM21 and
279 LM22), rhamnogalacturonan (recognised by mAbs INRA-RU1) and methylesterfied
280 homogalacturonan (recognised by mAb JIM7) than photoautotrophic cells. In contrast,
281 photoautotrophic cells contained relatively higher levels of xyloglucan (recognised by mAb LM15)
282 and mixed linkage glucan (recognised by mAb BS-400-3). Taken together, these results suggest that
283 the major carbohydrate in Euglena is paramylon, but other more structurally complex polymers
284 related to xylan, mannan, arabinan and arabinogalactan are also present.

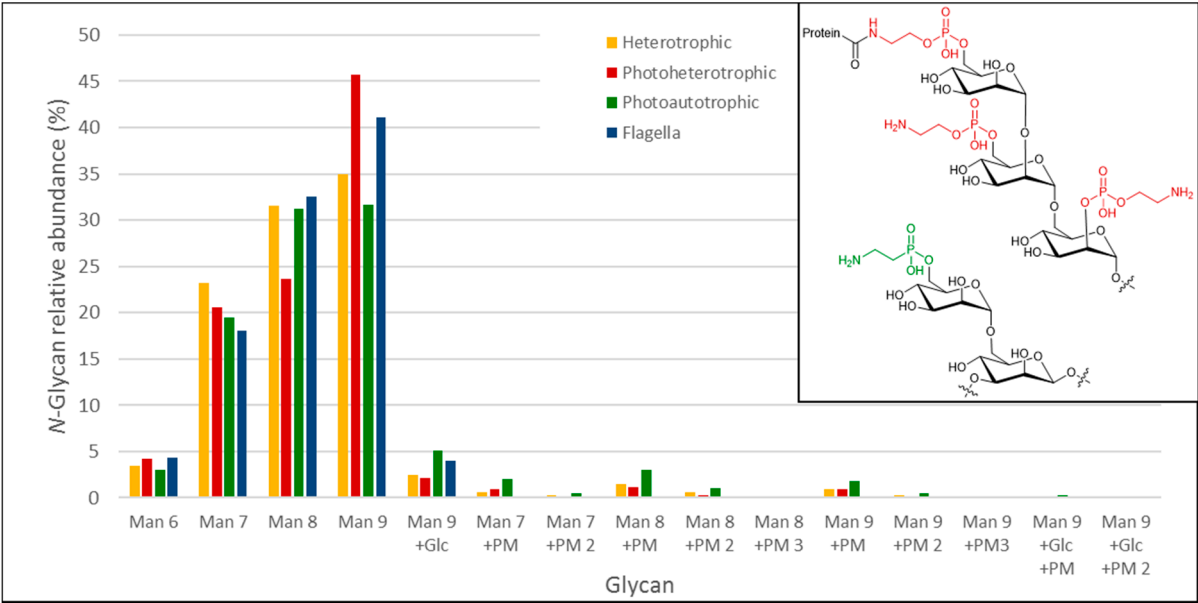
285

286 **Figure 4: Immunocarbohydrate microarray profiling of Euglena samples.** Only those antibodies that
287   had some reactivity to the Euglena samples are presented and a full list of the antibodies used and
288   their specificity are presented in Table S3. Hashed bars represent CDTA released material and solid
289   bars represent NaOH released material. Signal is normalised to the strongest signal.

290

291   3.3.3 Analysis of Euglena protein bound glycans

292   We used a porous graphitized carbon (PGC)-LC ESI MS/MS approach to identify the *N*- and *O*-
293   glycans from Euglena photoheterotrophic, heterotrophic and photoautotrophic cells and isolated
294   flagella. Proteins were dot-blotted onto PVDF membranes to enzymatically release *N*-glycans by
295   PNGase F, followed by a reductive β-elimination step to release any *O*-glycans present.[25] After
296   identification, relative quantitative data was obtained by integrating the area under the curve (AUC)
297   obtained for the respective Extracted Ion Chromatogram (EIC's) for each individual identified glycan
298   compound. The *N*-glycan analysis of all of the Euglena samples revealed the exclusive presence of
299   oligomannose-type *N*-glycans (Man9 > Man8 > Man7 > Man6 > Man9Glc) with no additional
300   carbohydrate moieties (Figure 5 and Table 1). However, the glycomics analyses of whole protein
301   extracts obtained from photoheterotrophic, heterotrophic and photoautotrophic *Euglena* showed that
302   a small fraction of oligomannose-type *N*-glycans also exhibited an additional modification of 107 Da
303   (Figure 5). Based on the detected mass differences, this 107 Da addition could reflect the attachment
304   of a putative 2-aminoethylphosphonate to the non-reducing end of the oligomannose *N*-glycans. 2-
305   Aminoethylphosphonate has previously been found to be attached to *N*-glycans from insects,[26] and
306   derivatives of this compound were also described to be present on marine invertebrates *N*-
307   glycans.[27] *N*-glycans carrying a putative aminoethylphosphonate were significantly elevated in
308   photoautotrophic (almost 10% of released glycans) and less pronounced in heterotrophic *Euglena*
309   (~4%). Putative aminoethylphosphonate carrying *N*-glycans could only just be detected in the flagella
310   samples, but at levels below the level of quantitation and these signals could also not be verified by
311   tandem MS data. We did not find any indication for complex type *N*-glycans as they have been
312   described for other algae, such as *Chlamydomonas reinhardtii*, where *N*-glycans were identified
313   carrying both core and antenna xylose residues and some methylation.[28] Beta-elimination after
314   PNGase F treatment did not result in any signals consistent with any *O*- or *N*-glycan-specific
315   signatures. This leads us to conclude that *Euglena gracilis* does not possess any PNGase F resistant,
316   complex type glycans, previously described in other Euglenozoa,[29,30] at least not in levels
317   detectable by the overall highly sensitive and selective approach. Overall this comprehensive *N*-
318   glycan profiling of Euglena shows that it produces predominantly high mannose type glycans with
319   a small proportion containing a modification consistent with the mass of an aminoethylphosphonate.

320

321

322   Figure 5: *N*-glycans found on Euglena (glyco)protein extracts and on purified flagella. For each
323   measurement, the obtained values are normalised to the sum of all detected *N*-glycans. PM - glycan
324   modification. The inset shows a fragment of the structure of a human GPI anchor core (upper) [31]
325   and a fragment of a *Locusta migratoria N*-glycan (lower).[26] The aminoethylphosphate and
326   aminoethylphosphonate are highlighted in red and green respectively.

327

Table 1: *N*-glycans identified in Euglena by PGC-LC-ESI-MS/MS						
Observed m/z	Charge state	Calculated mass (Da)	Theoretical mass (Da)	Delta mass (Da)	Glycan identified	
698.24	2	1398.48	1398.5	0.02	(Hex)3 + (Man)3(GlcNAc)2	Man 6
779.25	2	1560.50	1560.55	0.05	(Hex)4 + (Man)3(GlcNAc)2	Man 7
860.26	2	1722.52	1722.60	0.08	(Hex)5 + (Man)3(GlcNAc)2	Man 8
941.37	2	1884.74	1884.66	-0.08	(Hex)6 + (Man)3(GlcNAc)2	Man 9
1022.32	2	2046.64	2046.71	0.07	(Hex)7 + (Man)3(GlcNAc)2	Man 9 +Glc
832.75	2	1667.50	1667.55	0.05	(107)(Hex)4 + (Man)3(GlcNAc)2	Man 7 +PM
886.25	2	1774.50	1774.55	0.05	(107)2(Hex)4 + (Man)3(GlcNAc)2	Man 7 +PM 2
913.76	2	1829.52	1829.60	0.08	(107)(Hex)5 + (Man)3(GlcNAc)2	Man 8 +PM
967.26	2	1936.52	1936.60	0.08	(107)2(Hex)5 + (Man)3(GlcNAc)2	Man 8 +PM 2
1020.76	2	2043.52	2043.60	0.08	(107)3(Hex)5 + (Man)3(GlcNAc)2	Man 8 +PM 3
994.87	2	1991.74	1991.66	-0.08	(107)(Hex)6 + (Man)3(GlcNAc)2	Man 9 +PM
1048.37	2	2098.74	2098.66	-0.08	(107)2(Hex)6 + (Man)3(GlcNAc)2	Man 9 +PM 2
1101.87	2	2205.74	2205.66	-0.08	(107)3(Hex)6 + (Man)3(GlcNAc)2	Man 9 +PM 3
1075.82	2	2153.64	2153.71	0.07	(107)(Hex)7 + (Man)3(GlcNAc)2	Man 9 +Glc +PM
1129.32	2	2260.64	2260.71	0.07	(107)2(Hex)7 + (Man)3(GlcNAc)2	Man 9 +Glc +PM 2

Hex - Hexose; HexNAc - *N*-Acetyl Hexosamine; Man - Mannose; GlcNAc - *N*-acetyl Glucosamine; PM - glycan modification (107 Da) putatively assigned as 2-aminoethylphosphonate, previously described for *Locusta migratoria* apolipophorin III protein.[26]

328   **4. Conclusions**

329   *Euglena gracilis* expresses a wide range of glycosyltransferases and also produces the precursor
330   sugar nucleotides necessary for the synthesis of some very complex glycans. Our
331   immunocarbohydrate microarray profiling indicates that the surface of Euglena has some xylan- and
332   arabinan-type material, bearing GalNAc and GlcNAc moieties. However, these monosaccharides
333   appear not to be used to form complex type *N*-glycans, as only oligomannose type *N*-glycans were
334   detected in this study. There is evidence for the addition of an unusual aminoethylphosphonate to a
335   small proportion of these *N*-glycans; aminoethylphosphate-bearing mannose residues are more
336   typically recognised as a core component of human GPI anchor structures. Because Euglena
337   apparently does not produce classic complex-type *N*-glycans, it offers a platform to be used for the
338   heterologous expression and manufacture of medicinal proteins that will exclusively carry

339 oligomannose-type *N*-glycans. Targeted engineering of selected glycosylation pathways, such as
340 sialylation, has been successfully achieved in plants.[32] As most important nucleotide precursors are
341 naturally available in Euglena, a similar strategy could transform it into an attractive recombinant
342 glycoprotein production platform. This will require more detailed work to determine if the presence
343 of such complex type *N*-glycans affects the growth and fermentation potential of Euglena. Overall,
344 these data indicate that *Euglena gracilis* has potential as an alga platform for the manufacture of
345 pharmaceutical glycoproteins.

346 **Supplementary Materials:** The following are available online at www.mdpi.com/link, Table S1:
347 Number of Euglena transcripts encoding proteins in each of the CAZyme families, Table S2:
348 Standards of sugar nucleotides detected in the phototrophic culture of Euglena gracilis: relative
349 retention times and MRM transitions, Table S3: Specificity of the antibodies used in the
350 immunocarbohydrate microarray profiling

351 **Acknowledgments:** This work was supported by the UK BBSRC Institute Strategic Programme
352 on Understanding and Exploiting Metabolism (MET) [BB/J004561/1]; the John Innes Foundation; the
353 Royal Thai Government Scholarship programme. Ellis O'Neill is supported by a Violette and Samuel
354 Glasstone Independent Research Fellowship. DK acknowledges support by the European Union
355 (Seventh Framework Programme ''Glycoproteomics'' project, grant number PCIG09-GA-2011-
356 293847). We thank the Beilstein-Institut for supporting KA with a PhD scholarship and the Max
357 Planck Society for financial support. The authors would like to thank Lionel Hill for assistance with
358 mass spectrometry of sugar nucleotides, and Ben Wagstaff, Ed Hems and Nick Smirnoff for sugar
359 nucleotide standards.

360 **Author Contributions:** ECO and RAF conceived and designed the experiments. Sugar
361 nucleotide profiling was performed by MR. Surface labelling was performed by SK and ECO.
362 Microarray polymer profiling was performed by JUF and WGTW. N-Glycan analysis was performed
363 by KA and DK. ECO wrote the paper with input from all authors.

364 **Conflicts of Interest:** The authors declare no conflict of interest. The funding sponsors had no
365 role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of
366 the manuscript, and in the decision to publish the results.

References

368 1.   1.   Schwartzbach, S.; Shigeoka, S. Euglena: Biochemistry, cell and molecular biology. Springer: 2017.
369 2.   2.   Krajčovič, J.; Matej, V.; Schwartzbach, S.D. Euglenoid flagellates: A multifaceted biotechnology
370   platform. Journal of Biotechnology 2015, 202, 135-145.
371 3.   3.   O'Neill, E.C.; Saalbach, G.; Field, R.A. Gene discovery for synthetic biology: exploring the novel
372   natural product biosynthetic capacity of eukaryotic microalgae. Methods in Enzymology 2016, 576, 99-120.
373 4.   4.   O'Neill, E.C.; Trick, M.; Hill, L.; Rejzek, M.; Dusi, R.G.; Hamilton, C.J.; Zimba, P.V.; Henrissat, B.; Field,
374   R.A. The transcriptome of Euglena gracilis reveals unexpected metabolic capabilities for carbohydrate and
375   natural product biochemistry. Molecular Biosystems 2015, 11, 2808-2820.
376 5.   5.   Barras, D.R.; Stone, B.A. Chemical composition of pellicle of Euglena gracilis var bacillaris.
377   Biochemical Journal 1965, 97, 14-15.
378 6.   6.   Jahn, T.L. Manual of phycology: an introduction to the algae and their biology. Chronica Botanica Co.:
379   1951.
380 7.   7.   Ivanova, I.M.; Nepogodiev, S.A.; Saalbach, G.; O'Neill, E.C.; Urbaniak, M.D.; Ferguson, M.A.J.;
381   Gurcha, S.S.; Besra, G.S.; Field, R.A. Fluorescent mannosides serve as acceptor substrates for
382   glycosyltransferase and sugar-1-phosphate transferase activities in Euglena gracilis membranes.
383   Carbohydrate Research 2017, 438, 26-38.
384 8.   8.   Bouck, G.B.; Rogalski, A.; Valaitis, A. Surface organization and composition of Euglena. 2. Flagellar
385   mastigonemes. Journal of Cell Biology 1978, 77, 805-826.
386 9.   9.   de la Canal, L.; Parodi, A.J. Glycosylation of proteins in the protozoan Euglena gracilis. Comparative
387   Biochemistry and Physiology Part B: Comparative Biochemistry 1985, 81, 803-805.
388 10.   10.   Izquierdo, L.; Schulz, B.L.; Rodrigues, J.A.; Güther, M.L.S.; Procter, J.B.; Barton, G.J.; Aebi, M.;
389   Ferguson, M.A.J. Distinct donor and acceptor specificities of Trypanosoma brucei
390   oligosaccharyltransferases. The EMBO Journal 2009, 28, 2650-2661.

391　11.　11.　Geetha-Habib, M.; Bouck, G.B. Synthesis and mobilization of flagellar glycoproteins during
392　　　regeneration in Euglena. The Journal of Cell Biology 1982, 93, 432-441.

393　12.　12.　O'Neill, E.C.; Trick, M.; Henrissat, B.; Field, R.A. Euglena in time: Evolution, control of central
394　　　metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry.
395　　　Perspectives in Science 2015, 6, 84-93.

396　13.　13.　Olszewski, N.E.; West, C.M.; Sassi, S.O.; Hartweck, L.M. O-GlcNAc protein modification in plants:
397　　　Evolution and function. Biochimica Et Biophysica Acta 2010, 1800, 49-56.

398　14.　14.　Zentella, R.; Sui, N.; Barnhill, B.; Hsieh, W.-P.; Hu, J.; Shabanowitz, J.; Boyce, M.; Olszewski, N.E.;
399　　　Zhou, P.; Hunt, D.F., et al. The Arabidopsis O-fucosyltransferase SPINDLY activates nuclear growth
400　　　repressor DELLA. Nature Chemical Biology 2017, 13, 479-485.

401　15.　15.　Rodríguez-Zavala, J.S.; Ortiz-Cruz, M.A.; Mendoza-Hernández, G.; Moreno-Sánchez, R. Increased
402　　　synthesis of α-tocopherol, paramylon and tyrosine by Euglena gracilis under conditions of high biomass
403　　　production. Journal of Applied Microbiology 2010, 109, 2160-2172.

404　16.　16.　Stone, B.A. Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides.
405　　　Academic Press: San Diego, 2009.

406　17.　17.　Freeze, H.H.; Hart, G.W.; Schnaar, R.L. Glycosylation precursors. In Essentials of Glycobiology, 3rd
407　　　ed.; Varki, A.; Cummings, R.D.; Esko, J.D.; Stanley, P.; Hart, G.W.; Aebi, M.; Darvill, A.G.; Kinoshita, T.;
408　　　Packer, N.H.; Prestegard, J.H., et al., Eds. Cold Spring Harbor Laboratory Press

409　18.　The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), 2017.

410　19.　18.　Räbinä, J.; Mäki, M.; Savilahti, E.M.; Järvinen, N.; Penttilä, L.; Renkonen, R. Analysis of nucleotide
411　　　sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid
412　　　chromatography. Glycoconjugate Journal 2001, 18, 799-805.

413　20.　19.　Pabst, M.; Grass, J.; Fischl, R.; Léonard, R.; Jin, C.; Hinterkörner, G.; Borth, N.; Altmann, F. Nucleotide
414　　　and nucleotide sugar analysis by liquid chromatography-electrospray ionization-mass spectrometry on
415　　　surface-conditioned porous graphitic carbon. Analytical Chemistry 2010, 82, 9782-9788.

416　21.　20.　Qu, Q.; Lee, S.J.; Boos, W. TreT, a novel trehalose glycosyltransferring synthase of the
417　　　hyperthermophilic archaeon Thermococcus litoralis. Journal of Biological Chemistry 2004, 279, 47890-
418　　　47897.

419　22.　21.　Liu, C.; Yu, X. ADP-Ribosyltransferases and Poly ADP-Ribosylation. Current protein & peptide
420　　　science 2015, 16, 491-501.

421　23.　22.　Noctor, G.; Queval, G.; Gakière, B. NAD(P) synthesis and pyridine nucleotide cycling in plants and
422　　　their potential importance in stress conditions*. Journal of Experimental Botany 2006, 57, 1603-1620.

423　24.　23.　So, L.L.; Goldstein, I.J. Protein-carbohydrate interaction XX. On the number of combining sites of
424　　　concanavalin A, the phytohemagglutinin of the jack bean. Biochimica et Biophysica Acta (BBA) - General
425　　　Subjects 1968, 165, 398-404.

426　25.　24.　Moller, I.; Sorensen, I.; Bernal, A.J.; Blaukopf, C.; Lee, K.; Obro, J.; Pettolino, F.; Roberts, A.; Mikkelsen,
427　　　J.D.; Knox, J.P., et al. High-throughput mapping of cell-wall polymers within and between plants using
428　　　novel microarrays. Plant Journal 2007, 50, 1118-1128.

429　26.　25.　Jensen, P.H.; Karlsson, N.G.; Kolarich, D.; Packer, N.H. Structural analysis of N- and O-glycans
430　　　released from glycoproteins. Nature Protocols 2012, 7, 1299-1310.

431　27.　26.　Hard, K.; Van Doorn, J.M.; Thomas-Oates, J.E.; Kamerling, J.P.; Van der Horst, D.J. Structure of the
432　　　Asn-linked oligosaccharides of apolipophorin III from the insect Locusta migratoria. Carbohydrate-linked
433　　　2-aminoethylphosphonate as a constituent of a glycoprotein. Biochemistry 1993, 32, 766-775.

434　28.　27.　Eckmair, B.; Jin, C.; Abed-Navandi, D.; Paschinger, K. Multistep fractionation and mass dpectrometry
435　　　reveal zwitterionic and anionic modifications of the N- and O-glycans of a marine snail. Molecular &
436　　　Cellular Proteomics 2016, 15, 573-597.

437　29.　28.　Mathieu-Rivet, E.; Scholz, M.; Arias, C.; Dardelle, F.; Schulze, S.; Le Mauff, F.; Teo, G.; Hochmal, A.K.;
438　　　Blanco-Rivero, A.; Loutelier-Bourhis, C., et al. Exploring the N-glycosylation pathway in Chlamydomonas
439　　　reinhardtii unravels novel complex structures. Molecular & Cellular Proteomics 2013, 12, 3160-3183.

440　30.　29.　Zamze, S.E.; Ashford, D.A.; Wooten, E.W.; Rademacher, T.W.; Dwek, R.A. Structural characterization
441　　　of the asparagine-linked oligosaccharides from Trypanosoma brucei type II and type III variant surface
442　　　glycoproteins. Journal of Biological Chemistry 1991, 266, 20244-20261.

443 31. 30. Jones, D.C.; Mehlert, A.; Guther, M.L.; Ferguson, M.A. Deletion of the glucosidase II gene in
444   Trypanosoma brucei reveals novel N-glycosylation mechanisms in the biosynthesis of variant surface
445   glycoprotein. Journal of Biological Chemistry 2005, 280, 35929-35942.

446 32. 31. Deeg, M.A.; Humphrey, D.R.; Yang, S.H.; Ferguson, T.R.; Reinhold, V.N.; Rosenberry, T.L. Glycan
447   components in the glycoinositol phospholipid anchor of human erythrocyte acetylcholinesterase. Novel
448   fragments produced by trifluoroacetic acid. Journal of Biological Chemistry 1992, 267, 18573-18580.

449 33. 32. Kallolimath, S.; Castilho, A.; Strasser, R.; Grunwald-Gruber, C.; Altmann, F.; Strubl, S.; Galuska, C.E.;
450   Zlatina, K.; Galuska, S.P.; Werner, S., et al. Engineering of complex protein sialylation in plants.
451   Proceedings of the National Academy of Sciences 2016, 113, 9498-9503.

452 34. 33. Rejzek, M.; Hill, L.; Hems, E.S.; Kuhaudomlarp, S.; Wagstaff, B.A.; Field, R.A. Chapter Seven -
453   Profiling of sugar nucleotides. Methods in Enzymology 2017, Volume 597, 209-238.

454 35. 34. Wagstaff, B.A.; Rejzek, M.; Kuhaudomlarp, S.; Hill, L.; Mascia, I.; Nepogodiev, S.A.; Field, R.A. NDP-
455   β-L-rhamnose biosynthesis across the algal taxonomic groups; an evolutionary perspective.

456 36. 35. Turnock, D.C.; Ferguson, M.A.J. Sugar nucleotide pools of Trypanosoma brucei, Trypanosoma cruzi,
457   and Leishmania major. Eukaryotic Cell 2007, 6, 1450-1463.

458 37. 36. Rejzek, M.; Mukhopadhyay, B.; Wenzel, C.Q.; Lam, J.S.; Field, R.A. Direct oxidation of sugar
459   nucleotides to the corresponding uronic acids: TEMPO and platinum-based procedures. Carbohydrate
460   Research 2007, 342, 460-466.

461 38. 37. Errey, J.C.; Mann, M.C.; Fairhurst, S.A.; Hill, L.; McNeil, M.R.; Naismith, J.H.; Percy, J.M.; Whitfield,
462   C.; Field, R.A. Sugar nucleotide recognition by Klebsiella pneumoniae UDP-D-galactopyranose mutase:
463   Fluorinated substrates, kinetics and equilibria. Organic & Biomolecular Chemistry 2009, 7, 1009-1016.

464 39. 38. Errey, J.C.; Mukhopadhyay, B.; Kartha, K.P.R.; Field, R.A. Flexible enzymatic and chemo-enzymatic
465   approaches to a broad range of uridine-diphospho-sugars. Chemical Communications 2004, 2706-2707.

466 40. 39. Pedersen, H.L.; Fangel, J.U.; McCleary, B.; Ruzanski, C.; Rydahl, M.G.; Ralet, M.-C.; Farkas, V.; von
467   Schantz, L.; Marcos, S.E.; Andersen, M.C.F., et al. Versatile high-resolution oligosaccharide microarrays for
468   plant glycobiology and cell wall research. Journal of Biological Chemistry 2012, 287, 39429-39438.

469