

# A theoretical estimate for nucleotide sugar demand towards Chinese Hamster Ovary cellular glycosylation

Ioscani Jimenez del Val<sup>a,b,\*</sup>, Karen M Polizzi<sup>c,d</sup> and Cleo Kontoravdi<sup>b</sup>

<sup>a</sup> School of Chemical & Bioprocess Engineering, University College Dublin, Belfield campus, Dublin 4, Ireland

<sup>b</sup> Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K.

<sup>c</sup> Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K.

<sup>d</sup> Centre for Synthetic Biology and Innovation, Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

\* Author to whom correspondence should be addressed: Email: [ioscani.jimenezdelval@ucd.ie](mailto:ioscani.jimenezdelval@ucd.ie); Tel.: +353-(1)716-1923; Fax: +353-(1)716-1177

## Abstract

Glycosylation greatly influences the safety and efficacy of many of the highest-selling recombinant therapeutic proteins (rTPs). In order to define optimal cell culture feeding strategies that control rTP glycosylation, it is necessary to know how nucleotide sugars (NSs) are consumed towards host cell and rTP glycosylation. Here, we present a theoretical framework that integrates the reported glycoproteome of CHO cells, the number of N-linked and O-GalNAc glycosylation sites on individual host cell proteins (HCPs), and the carbohydrate content of CHO glycosphingolipids to estimate the demand of NSs towards CHO cell glycosylation. We have identified the most abundant N-linked and O-GalNAc CHO glycoproteins, obtained the weighted frequency of N-linked and O-GalNAc glycosites across the CHO cell proteome, and have derived stoichiometric coefficients for NS consumption towards CHO cell glycosylation. By combining the obtained stoichiometric coefficients with previously reported data for specific growth and productivity of CHO cells, we observe that the demand of NSs towards glycosylation is significant and, thus, is required to better understand the burden of glycosylation on cellular metabolism. The estimated demand of NSs towards CHO cell glycosylation can be used to rationally design feeding strategies that ensure optimal and consistent rTP glycosylation.

**Keywords:** Chinese Hamster Ovary cells, N-linked glycosylation, O-GalNAc glycosylation, Host cell proteins, Pharmaceutical bioprocessing, Biopharmaceutical quality control.

# 1 Introduction

2 Glycosylation is the most prevalent and impactful post-translational modification of  
3 recombinant therapeutic proteins (rTP)<sup>1,2</sup>. Twenty of the fifty highest selling  
4 pharmaceutical products in 2014 were glycoproteins and all twenty contain asparagine-  
5 linked (N-linked) complex carbohydrates (glycans), with two also containing  
6 serine/threonine-linked mucin type (O-GalNAc) glycans<sup>3</sup>. The relative abundance and  
7 monosaccharide composition of therapeutic protein glycoforms has been widely  
8 reported to influence the bioactivity, bioavailability, and biocompatibility of these  
9 molecules<sup>1,2</sup>. Given the close relationship between glycoform distribution and rTP safety  
10 and efficacy, understanding the fundamental aspects of protein glycosylation is of critical  
11 importance for the biopharmaceutical industry<sup>1,2,4</sup>.

12 Nucleotide sugars (NSs) are the direct link between cellular metabolism and the  
13 glycosylation process<sup>4,5</sup>. NSs are synthesised in the cytoplasm from nutrients commonly  
14 found in cell culture medium (e.g. glucose, glutamine, asparagine, and glycine) and are  
15 subsequently transported into the Golgi apparatus, where they are consumed in the  
16 glycosylation reactions. Several reports have linked decreased availability of NSs with  
17 reduced complexity of secreted glycans<sup>6-8</sup>. In addition, direct precursors for NS  
18 biosynthesis, such as N-acetylglucosamine (GlcNAc), galactose, N-acetylmannosamine  
19 (ManNAc), mannose, uridine, and cytosine, have been fed to the culture to ensure that  
20 NSs are sufficiently available to achieve adequate recombinant product glycosylation  
21 (reviewed recently in ref. 9). However, addition of these NS precursors has commonly  
22 been associated with reduced cell growth<sup>10-13</sup>; therefore, optimal NS precursor feeding  
23 strategies that can modulate recombinant product glycosylation while minimising the  
24 impact on cell growth are needed<sup>12,14</sup>. Alongside experimental work aimed at identifying

1 such optimal feeding strategies, recently developed mathematical models<sup>15-17</sup> could  
2 become a powerful tool for these purposes in the near future.

3 Crucially, the glycosylation of recombinant product (rTP), host cell proteins (HCPs), and  
4 lipids occurs simultaneously. This is substantiated by the close correlation observed  
5 between cell surface and rTP glycans reported recently<sup>12</sup>. As a consequence, a fraction of  
6 fed NS precursors does not reach their intended target, and identifying the demand of  
7 NSs required for HCP glycosylation is therefore necessary to define optimal NS precursor  
8 feeding strategies.

9 To date, two estimates for the demand of NSs towards mammalian cell HCP glycosylation  
10 have been made<sup>16,17</sup>. Both were obtained by combining the frequency of N-linked and O-  
11 GalNAc glycosylation sites reported across the entire SwissProt database<sup>18</sup> with the  
12 relative abundance of each monosaccharide present in the reported glycome of human  
13 activated B-cells<sup>16</sup> or CHO cells<sup>17</sup>. Despite providing useful initial approximations, these  
14 estimates have limitations given that they do not account for the relative abundance of  
15 each protein within the host organism's proteome or the number of glycosylation sites  
16 on each protein. In addition, these estimates have not included the consumption of NSs  
17 towards glycolipid synthesis.

18 Here, we present a theoretical framework to estimate the demand of all nucleotide sugars  
19 consumed towards CHO HCP N-linked and O-GalNAc glycosylation, as well as for  
20 glycolipid synthesis. Our strategy integrates recent CHO proteomic<sup>19</sup> and glycomic<sup>20,21</sup>  
21 data to include (i) the relative abundance of all reported glycoproteins within the CHO  
22 proteome, (ii) the frequency of N-linked and O-GalNAc glycosylation sites across the  
23 proteome, and (iii) the monosaccharide composition of all N-linked and O-GalNAc HCP

glycans. We also include the reported monosaccharide content of CHO lipids<sup>22</sup>, so that this sink for NSs is accounted for.

Our framework derives stoichiometric coefficients for the demand of all nucleotide sugars consumed for CHO HCP glycosylation and glycosphingolipid synthesis. The obtained stoichiometric coefficients are subsequently used to estimate how NS demand is partitioned between HCP, lipid, and rTP glycosylation<sup>23</sup> in industrially-relevant mAb-producing CHO cells<sup>24-26</sup>. Our results indicate that NS consumption rates towards cellular and rTP glycosylation are within the same order of magnitude, suggesting that considering both is imperative to achieve a comprehensive description of NS metabolism.

The overall framework presented herein provides an approximation of NS demand towards CHO cellular glycosylation, which can aid in developing NS precursor feeding strategies that ensure control of rTP glycosylation without compromising cell growth and product yield. In addition, use of the estimated stoichiometric coefficients in advanced mathematical models for NS metabolism and glycosylation should improve model fidelity and take us one step closer to model-based design and control of feeding strategies that ensure optimal and consistent rTP glycosylation. Finally, we report the most abundant host cell N-linked and O-GalNAc glycoproteins in CHO cells as well as their individual contribution towards overall NS consumption. Monitoring the glycosylation profiles of this reduced set of glycoproteins could be used to track and predict the effect of NS precursor feeding strategies in the future.

## Materials and Methods

In order to obtain an improved estimate for the demand of NSs towards CHO cell glycosylation, we required an estimate of (i) the relative abundance of CHO HCPs, (ii) the

weighted frequency of N-linked and O-GalNAc glycan sites per protein, (iii) the average glycoform distribution of each type, and (iv) NS demand for glycolipid synthesis. The workflow employed to obtain each of these components and the stoichiometric coefficients of NS consumption towards CHO HCP glycosylation are detailed in Figure 1.

## Step 1: CHO HCP relative quantification

The reported proteome of CHO cells<sup>19</sup> was compared with the SwissProt database<sup>27</sup> using BLAST<sup>28</sup> to obtain the closest reviewed homolog for each HCP. The sequence, length ( $Len_i$ ), and molecular weight ( $MW_i$ ) for each homolog were also retrieved from SwissProt. The obtained amino acid lengths and reported spectral counts<sup>19</sup> ( $SC_i$ ) were then used to calculate the relative abundance ( $Z_i$ ) of each CHO HCP using Eqs. 1 and 2, as described previously<sup>29,30</sup>. With the obtained values for  $Z_i$ , the weighted average amino acid length ( $\overline{Len}_{HCP}$ ) and molecular weight ( $\overline{MW}_{HCP}$ ) of CHO HCPs were calculated using Eqs. 3 and 4. The list for all reported HCPs, the obtained homologs, and the calculations for  $Z_i$ ,  $\overline{Len}_{HCP}$ , and  $\overline{MW}_{HCP}$  are presented in Supplementary table ST1. The weighted relative abundance of the amino acids present in CHO HCPs ( $x_{Aac_j}$ ) was obtained with Eq. 5, using the amino acid frequency in the sequence of each homologous protein ( $n_{Aac_j}$ ) and  $Z_i$ . The calculation of  $x_{Aac_j}$  across the whole CHO proteome is presented in Supplementary table ST2.

$$SAF_i = \frac{SC_i}{Len_i} \quad \text{Eq. 1}$$

$$NSAF_i = \frac{SAF_i}{\sum_{i=1}^{NP} SAF_i} = Z_i \quad \text{Eq. 2}$$

$$\overline{Len}_{HCP} = \sum_{i=1}^{NP} Len_i Z_i \quad \text{Eq. 3}$$

$$\overline{MW}_{HCP} = \sum_{i=1}^{NP} MW_i Z_i \quad \text{Eq. 4}$$

$$x_{Aac_j} = \frac{\sum_{i=1}^{NP} n_{Aac_{j,i}} Z_i}{\sum_{j=1}^{NAac} \sum_{i=1}^{NP} n_{Aac_{j,i}} Z_i} \quad \text{Eq. 5}$$

## 1 Step 2: N-linked glycan sites per protein

2 The vast majority (>90%) of all reported N-linked glycosites in SwissProt have not been  
3 experimentally confirmed to be occupied<sup>31</sup>. In order to further refine the estimate by  
4 considering macroheterogeneity (variations in N-glycan occupancy of each potential  
5 site), the amino acid sequences for all potential N-linked glycoproteins were analysed  
6 using the NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>)<sup>32</sup>. To obtain  
7 a conservative estimate for N-linked glycosites, only ones found to be within the upper  
8 thresholds (++ and +++) were considered to be occupied. The obtained number of  
9 occupied N-linked glycosites ( $S_{Nglyc,i}^{HCP}$ ) was multiplied by the respective protein's relative  
10 abundance ( $Z_i$ ) and then summed to obtain the weighted number of N-linked glycosites  
11 across the entire CHO proteome ( $S_{Nglyc}^{HCP}$ ), as shown in Eq. 6. All calculations performed to  
12 estimate  $S_{Nglyc}^{HCP}$ , along with the contribution of each predicted N-glycoprotein to the  
13 weighted number of N-glycosites, are presented in Supplementary table ST3.

$$S_{Nglyc}^{HCP} = \sum_{i=1}^{NP} S_{Nglyc,i}^{HCP} Z_i \quad \text{Eq. 6}$$

### 1 Step 3: O-linked glycan sites per protein

2 As of February 2016, 2,375 O-GalNAc glycosylation sites were reported in a total of  
3 182,101 eukaryotic glycoproteins in the SwissProt database<sup>27</sup>. However, a recent study  
4 by Yang and collaborators<sup>21</sup> found 1,548 O-GalNAc glycosites in the proteome of CHO cells  
5 alone, suggesting that O-GalNAc glycosylation may be underreported in SwissProt. The  
6 low frequency of reported O-GalNAc sites is possibly due to the challenging analytical  
7 methods required for mapping, identifying, and quantifying these glycan types which lack  
8 a consensus amino acid sequon, may have core structure variability, and cluster densely  
9 in Ser/Pro/Thr-rich domains<sup>33</sup>.

10 In order to obtain the frequency of these post-translational modifications, the reported  
11 O-GalNAc glycoproteome of CHO cells was used<sup>21</sup>. The work of Yang et al. relied on  
12 engineering CHO cells to produce truncated O-GalNAc glycans that simplified  
13 glycopeptide enrichment using lectin chromatography and streamlined data analysis and  
14 interpretation for mapping all O-GalNAc glycosylation sites. The glycoproteome from this  
15 study was BLASTed<sup>28</sup> against the SwissProt<sup>27</sup> database to obtain a set of confirmed  
16 homologous O-GalNAc glycoproteins. This set was then aligned with the closest  
17 homologous proteins obtained for the entire CHO proteome. Once aligned, the number of  
18 O-GalNAc glycosites per HCP ( $S_{N_{glyc},i}^{HCP}$ ) reported by Yang et al.<sup>21</sup> was multiplied by the  
19 relative abundance of each HCP ( $Z_i$ ) and summed (Eq. 7) to obtain the weighted average  
20 number of O-GalNAc glycosites across the CHO proteome ( $S_{Oglyc}^{HCP}$ ). All calculations and the  
21 contribution of individual proteins to  $S_{Oglyc}^{HCP}$  are presented in Supplementary table ST4.

$$S_{Oglyc}^{HCP} = \sum_{i=1}^{NP} S_{Oglyc,i}^{HCP} Z_i \quad \text{Eq. 7}$$



## Step 4. Weighted average monosaccharide composition of CHO HCP and mAb glycans

In order to obtain the weighted average monosaccharide composition of CHO HCP glycans ( $f_{NS_k}^{NglycHCP}$  and  $f_{NS_k}^{OglycHCP}$ ), the reported<sup>20</sup> relative abundance of each N-linked glycan species  $l$  and O-GalNAc glycan species  $m$  ( $x_{Nglyc,l}^{HCP}$  and  $x_{Oglyc,m}^{HCP}$ , respectively) presented in Figure 2 was multiplied by the frequency with which the monosaccharide species appears in each glycan ( $n_{NS_k,l/m}$ ), as shown in Eqs. 8 and 9. Each monosaccharide present on the reported HCP glycans was assumed to be sourced from the respective nucleotide sugar (NS). All N-linked HCP glycans were assumed to contain three glucose (sourced from uridine diphosphate glucose – UDP-Glc) and nine mannose (sourced from guanosine diphosphate mannose – GDP-Man) residues, given that all N-linked glycans are synthesised from the Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> precursor oligosaccharide depicted on the top left corner of Figure 2. These calculations are detailed in Supplementary table ST5 and ST6.

The weighted average monosaccharide composition of mAb glycans ( $f_{NS_k}^{Nglyc,mAb}$ ) was obtained by combining the relative abundance of mAb constant fragment (Fc) N-glycans ( $x_{Nglyc,q}^{mAb}$ ) reported for trastuzumab<sup>23</sup> (Figure 2) with the frequency of each monosaccharide residue in these Fc carbohydrates ( $n_{NS_k,q}$ ), as shown in Eq. 10. The Fc glycoform distribution of trastuzumab<sup>23</sup> was used because product glycan data is not reported in the work cited herein for the estimation of NS consumption between host cell and product glycosylation<sup>24-26</sup>. Furthermore, data for trastuzumab glycosylation was selected for the calculation of  $f_{NS_k}^{Nglyc,mAb}$  given that this product contains glycans that are representative of almost all commercially-available mAbs (complex biantennary with varying degrees of fucosylation and galactosylation<sup>1,2,4</sup>) and can be considered a ‘model’

CHO-derived product due to its isotype (IgG1 – the most common among therapeutic mAbs<sup>4</sup>) and the fact that it has been manufactured using CHO cells for nearly two decades<sup>3</sup>. Similarly to HCP glycans, all mAb Fc N-linked glycans were assumed to require three glucose and nine mannose residues for their formation. All calculations performed to obtain  $f_{NS_k}^{Nglyc,mAb}$  are presented in Supplementary table ST7.

$$f_{NS_k}^{NglycHCP} = \sum_{l=1}^{NGlycans} n_{NS_k,l} x_{Nglyc,l}^{HCP} \quad \text{Eq. 8}$$

$$f_{NS_k}^{Oglyc,HCP} = \sum_{m=1}^{OGlycans} n_{NS_k,m} x_{Oglyc,m}^{HCP} \quad \text{Eq. 9}$$

$$f_{NS_k}^{Nglyc,mAb} = \sum_{q=1}^{mAbGlycans} n_{NS_k,q} x_{Nglyc,q}^{mAb} \quad \text{Eq. 10}$$

The monosaccharide content of total CHO cell lipids relative to total protein content ( $F_{NS_k}^{GL}$ ) was obtained from the work by Briles and collaborators<sup>22</sup>. With these values, the distribution of CHO cell glycosphingolipids (Figure 2) was calculated as shown in Supplementary table ST8.

## Step 5: Stoichiometric coefficients of NSs per cell

The final step was to calculate the stoichiometric coefficients for the demand of NSs towards CHO HCP N-linked and O-GalNAc glycan synthesis ( $v_{NS_k}^{NglycHCP}$  and  $v_{NS_k}^{OglycHCP}$ , respectively), CHO cell glycolipid synthesis ( $v_{NS_k}^{GL}$ ), and mAb N-linked glycosylation ( $v_{NS_k}^{Nglyc,mAb}$ ).  $v_{NS_k}^{NglycHCP}$ ,  $v_{NS_k}^{OglycHCP}$ , and  $v_{NS_k}^{GL}$  represent the weighted average number of nanomoles of NSs consumed towards each type of glycosylation per million cells. The stoichiometric coefficient of NS demand towards HCP N-linked glycosylation was obtained by combining the dry weight of CHO cells ( $M_{cell}$ ), the protein content of CHO

cells ( $Z_{prot}$  in %w/w), the weighted average molecular weight of CHO HCPs ( $\overline{MW}_{HCP}$ ), the weighted average number of N-linked glycosites across the CHO proteome ( $S_{Nglyc,i}^{HCP}$ ), and the weighted average monosaccharide composition of CHO HCP glycans ( $f_{NS_k}^{NglycHCP}$ ), as shown in Eq. 11. The stoichiometric coefficient for NS demand towards CHO HCP O-GalNAc glycosylation was calculated in a similar fashion, as presented in Eq. 12. For these calculations, the dry weight of CHO cells reported by Carinhas et al.<sup>34</sup> ( $M_{cell} = 271 \text{ pg/cell}$ ) and a cellular protein content of  $Z_{prot} = 74.2\%$ <sup>35</sup> were used. The value for dry cell weight was selected because it is representative of those that have previously been reported for CHO cells<sup>36,37</sup>, and  $Z_{prot} = 74.2\%$  was selected given that similar values have been reported for several mammalian cell types, including CHO<sup>34-38</sup>. Although different assays have been used to quantify  $Z_{prot}$  (e.g. Bradford<sup>34</sup> and Biuret<sup>38</sup>), the values that are reported typically range between 70% and 80%. The stoichiometric coefficient for NS consumption towards lipid glycosylation ( $v_{NS_k}^{GL}$ ) was obtained by multiplying the monosaccharide content of CHO cell lipids relative to total protein content ( $F_{NS_k}^{GL}$ )<sup>22</sup> by the above mentioned values for  $M_{cell}$  and  $Z_{prot}$ , as shown in Eq. 13. The calculations to obtain  $v_{NS_k}^{NglycHCP}$ ,  $v_{NS_k}^{OglycHCP}$ , and  $v_{NS_k}^{GL}$  are shown in Supplementary table ST9.

The stoichiometric coefficients for NS demand towards mAb Fc N-linked glycosylation ( $v_{NS_k}^{Nglyc,mAb}$ ) were calculated with Eq. 14 using the values of  $f_{NS_k}^{Nglyc,mAb}$  that were obtained as described in Step 4. The mAb was assumed to contain two N-linked glycosites on its Fc ( $S_{Nglyc}^{mAb} = 2$ ). The molecular weight of the mAb was calculated using the amino acid sequence of trastuzumab and its weighted average glycan composition<sup>23</sup> to yield a value of  $MW_{mAb} = 148,545 \text{ g/mol}$ . The calculation of  $v_{NS_k}^{Nglyc,mAb}$  is presented in Supplementary table ST9.

$$v_{NS_k}^{NglycHCP} = \left( \frac{M_{cell} Z_{prot}}{MW_{HCP}} \right) S_{Nglyc}^{HCP} f_{NS_k}^{NglycHCP} \quad \text{Eq. 11}$$

$$v_{NS_k}^{OglycHCP} = \left( \frac{M_{cell} Z_{prot}}{MW_{HCP}} \right) S_{Oglyc}^{HCP} f_{NS_k}^{OglycHCP} \quad \text{Eq. 12}$$

$$v_{NS_k}^{GL} = M_{cell} Z_{prot} F_{NS_k}^{GL} \quad \text{Eq. 13}$$

$$v_{NS_k}^{Nglyc,mAb} = \left( \frac{1}{MW_{mAb}} \right) S_{Nglyc}^{mAb} f_{NS_k}^{Nglyc,mAb} \quad \text{Eq. 14}$$

## 1 Other types of cellular glycosylation

2 Although other types of glycosylation exist in mammalian cells, they have not been  
3 considered to greatly impact cellular NS demand due to their low abundance relative to  
4 the amount of HCP N-linked and O-GalNAc glycans and glycosphingolipids<sup>39,40</sup>. For  
5 example, O-mannose and O-fucose linked glycans have been indeed observed in the CHO  
6 cell glycoproteome, but at nearly undetectable levels<sup>20</sup>.

7 Similarly, dolichol pyrophosphate-linked oligosaccharides (Dol-PP-OS) have also been  
8 reported to be present in CHO cells<sup>41</sup>, but it is likely that NS demand towards this  
9 glycolipid pool is small compared to that of HCP and glycosphingolipid glycosylation.

10 Intracellular accumulation of Dol-PP-OSs has been reported to depend on the availability  
11 of dolichol phosphate (Dol-P)<sup>42</sup>, which in turn, has been identified as a small fraction (only  
12 a few percent) of total cellular lipids in rat and mouse tissues<sup>43</sup>. In addition, the majority  
13 of monosaccharide constituents of the Dol-PP-OS precursor are cleaved off during glycan  
14 processing within the Golgi apparatus and are likely recycled for *de novo* synthesis of Dol-  
15 PP-OS. Therefore, it is possible to assume that NS consumption towards Dol-PP-OS

synthesis is low compared to glycosphingolipid and HCP glycosylation and is, thus, negligible within our proposed framework.

### Proteoglycans

Proteoglycans, which contain linear carbohydrate chains constituted by tandem disaccharide motifs (glycosaminoglycans), are more commonly associated with cell adhesion in connective tissue<sup>44</sup>, and thus, may not be prevalent in CHO cells<sup>39</sup>. In order to evaluate whether this potential source of NS consumption is low compared to cellular N-linked, O-GalNAc, and glycosphingolipid glycosylation, the frequency of this glycosylation type was estimated based on the CHO proteome quantification described in Step 1 above.

Specifically, the database of closest confirmed homologs to the CHO proteome (ST1) was queried for presence of the forty-three known proteoglycans<sup>44</sup>. The weighted average frequency of glycosaminoglycan (GAG) sites per CHO HCP ( $S_{GAG}^{HCP}$ ) was obtained by multiplying the reported number of glycosaminoglycan (GAG) sites per proteoglycan ( $S_{GAG,i}^{HCP}$ ) by the relative abundance of the corresponding protein ( $Z_i$ ) and summing this product across the entire CHO proteome, as shown in Eq. 15. These calculations are shown in Supplementary table ST10.

$$S_{GAG}^{HCP} = \sum_{i=1}^{NP} S_{GAG,i}^{HCP} Z_i \quad \text{Eq. 15}$$

### Glycosylphosphatidylinositol (GPI) anchors

The glycosylphosphatidylinositol (GPI) anchor, which attaches proteins to the membrane of cells, also contains a carbohydrate component. The amount of free GPI and GPI-anchored proteins has also been reported to be low compared to N-linked, O-GalNAc, and glycosphingolipid glycosylation<sup>39,40</sup>. In order to estimate the abundance of GPI glycans in

1 CHO cells, the number of potential GPI-anchored CHO HCPs ( $S_{GPI,i}^{HCP}$ ) was obtained by  
 2 submitting the amino acid sequences of all confirmed homologous CHO HCPs (ST1) to the  
 3 PredGPI GPI-anchor prediction server (<http://gpcr.biocomp.unibo.it/predgpi/pred.htm>)  
 4 <sup>45</sup>. The obtained  $S_{GPI,i}^{HCP}$  values were then multiplied by the relative abundance of each  
 5 corresponding HCP ( $Z_i$ ) and summed (Eq. 16) to obtain the weighted average frequency  
 6 of GPI-anchored HCPs ( $S_{GPI}^{HCP}$ ) across the CHO proteome. These calculations are detailed  
 7 in Supplementary table ST10. The number of free GPI sites within CHO cells ( $S_{GPI}^{free}$ ) was  
 8 obtained by multiplying  $S_{GPI}^{HCP}$  by the ratio of free to occupied GPI sites ( $3.7 oGPI/fGPI$ )  
 9 reported for CHO cells<sup>39</sup>, as shown in Eq. 17. These calculations are also presented in  
 10 Supplementary table ST10.

$$S_{GPI}^{HCP} = \sum_{i=1}^{NP} S_{GPI,i}^{HCP} Z_i \quad \text{Eq. 16}$$

$$S_{GPI}^{free} = 3.7 S_{GPI}^{HCP} \quad \text{Eq. 17}$$

## 11 Demand of NSs towards cellular and mAb glycosylation

12 Considering that the demand for NSs is partitioned between HCP, lipid, and mAb  
 13 glycosylation, two limits can be identified: (i) when cell growth greatly exceeds mAb  
 14 productivity and (ii) when mAb productivity considerably outweighs growth. Both  
 15 scenarios could be observed in typical fed-batch culture, where during exponential  
 16 growth phase, the demand of NSs towards cellular glycosylation would predominate,  
 17 whereas during stationary phase, the demand of NSs towards cellular glycosylation  
 18 would decrease considerably and demand towards mAb glycosylation would become  
 19 significant. In order to cover both possibilities, data for the specific growth rate ( $\mu_g$ ) and

specific mAb productivity ( $q_p$ ) corresponding to a breadth of productivity to growth ratios ( $R_{q_p/\mu_g} = q_p/\mu_g$ ) reported for industrially-relevant CHO cells have been retrieved from recent publications<sup>24-26</sup>.

Specifically, two sets of values for  $\mu_g$  and  $q_p$  were extracted from each reference, the first corresponding to a low productivity to growth rate ratio (Dool-L<sup>24</sup>, Chus-L<sup>25</sup>, and Kant-L<sup>26</sup>), and the second corresponding to a high productivity to growth ratio (Dool-H<sup>24</sup>, Chus-H<sup>25</sup>, and Kant-H<sup>26</sup>) (Table 1). All but the  $\mu_g$  and  $q_p$  values obtained from Kant-H correspond to industrial CHO cell lines continuously expressing a recombinant mAb while growing exponentially. The selected  $\mu_g$  and  $q_p$  values are considered typical for industrial CHO bioprocesses ( $\mu_g$  between  $0.03\text{h}^{-1}$  and  $0.05\text{h}^{-1}$  and  $q_p$  between  $20\text{pg/cell/day}$  to  $70\text{pg/cell/day}$ )<sup>46,47</sup>. The values for Kant-H correspond to engineered CHO cells undergoing stationary growth while being cultured under mild hypothermic conditions and in the presence of sodium butyrate. The Kant-H values were selected to have an extremely high  $R_{q_p/\mu_g}$  that would be representative of the limit for scenario (ii) described above ( $q_p = 94.05\text{pg/cell/day}$  and  $\mu_g = 0.011\text{h}^{-1}$ ). We must note that a  $q_p$  value of  $100\text{pg/cell/day}$  is commonly viewed as the upper threshold for CHO cell mAb specific productivity<sup>48</sup>. In all references,  $q_p$  was determined based on extracellular measurements of mAb titre. Chusainow *et al.* and Kantardjieff *et al.* determined mAb titre using ELISA, while Doolan *et al.* used protein A HPLC for mAb quantification.

In order to obtain estimates for NS demand towards HCP ( $D_{NS_k}^{NglycHCP}$  and  $D_{NS_k}^{OglycHCP}$ ), lipid ( $D_{NS_k}^{GL}$ ), and mAb ( $D_{NS_k}^{Nglyc,mAb}$ ) glycosylation under different growth and mAb productivity conditions, the  $\mu_g$  and  $q_p$  values were simply multiplied by the stoichiometric coefficients

1 calculated in Step 5, as shown in Eqs. 18-21. These calculations are detailed in  
 2 Supplementary table ST11.

$$D_{NS_k}^{NglycHCP} = v_{NS_k}^{NglycHCP} \mu_g \quad \text{Eq. 18}$$

$$D_{NS_k}^{OglycHCP} = v_{NS_k}^{OglycHCP} \mu_g \quad \text{Eq. 19}$$

$$D_{NS_k}^{GL} = v_{NS_k}^{GL} \mu_g \quad \text{Eq. 20}$$

$$D_{NS_k}^{Nglyc, mAb} = v_{NS_k}^{Nglyc, mAb} q_p \quad \text{Eq. 21}$$

3 By combining the assumed values for  $M_{cell} = 271 \text{ pg/cell}^{34}$  and  $Z_{prot} = 74.2\%^{35}$  with  
 4 the  $\mu_g$  values obtained from literature<sup>24-26</sup>, cellular protein productivity can be estimated  
 5 (Eq. 22). The obtained HCP, mAb, and total protein synthesis rates are shown in Table 1  
 6 and the corresponding calculations are presented in supplementary table ST11.

$$q_{prot}^{HCP} = M_{cell} Z_{prot} \mu_g \quad \text{Eq. 22}$$

## 7 Cellular glycosylation calculator

8 Although literature-derived experimental data has been used to perform all calculations  
 9 reported herein, we must highlight that the supplementary file is a set of integrated  
 10 spreadsheets that is amenable to analysis and modification by users to include  
 11 experimental data for specific rTGP-producing CHO cell cultures. Users may readily  
 12 adjust the relative abundance of HCP N-linked and O-GalNAc glycans (ST5 and ST6), the  
 13 glycoform distribution of the recombinant product (ST7), the molecular weight and  
 14 number of N-linked glycosites on the product (ST9), and the specific growth rate and  
 15 productivity (ST11) to obtain stoichiometric coefficients for NS consumption towards  
 16 recombinant product and cellular glycosylation for any system. These coefficients can, in



turn, be used to estimate the partition of NS consumption towards cellular and rTGP glycosylation for user-defined CHO cell culture data and can also be used as inputs for models of CHO cell metabolism.

## Results and Discussion

### Relative abundance of CHO cell proteins

The most highly abundant individual proteins are those associated with the cytoskeleton (e.g.  $\gamma$ -actin,  $\beta$ -actin,  $\beta$ -tubulin,  $\alpha$ -tubulin, vimentin), glycolysis (glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase A), chromatin (e.g. histones H4 and H2B), ribosomes (e.g. 40S ribosomal protein S12 and S3A), and those involved in protein transcription, translation, and folding (e.g. heat shock cognate 71 kDa protein, eukaryotic translation initiation factor 5A-1, elongation factor 1 $\alpha$ -1, and peptidyl-prolyl cis-trans isomerase A). A full list of the CHO cell proteome including the relative abundance of each protein is presented in Supplementary Table ST1.

Figure 3A compares the most abundant categories (as obtained through relative quantification and gene ontology analysis) of CHO HCPs with those obtained from protein relative abundance data of a murine ovary cell line (<http://pax-db.org/>)<sup>49,50</sup>. Both datasets are in relative agreement for cytoskeletal, chromatin, and glycolysis and respiration proteins. Large deviations can be observed for HCPs involved in protein transcription, translation, and folding. Specifically, HCPs involved in protein biosynthesis (transcriptional, translational, and ribosomal proteins) are in considerably higher abundance in CHO cells than in the murine ovary cell line. These differences are likely to

be species-specific and are consistent with the phenotypic traits (rapid proliferation and high protein productivity) that characterise CHO cells.

In order to further validate the proteome-based quantification, the relative abundance of amino acids in the CHO HCPs was compared with previously reported data on murine<sup>38</sup> and CHO<sup>51</sup> HCP amino acid content (Figure 3B). In general, the trends in amino acid content are similar across all datasets. When compared with the murine hybridoma data, the proteome-derived amino acid distribution diverges only slightly, with the highest observed deviations being for glutamate (+2.0%), proline (-1.7%), and lysine (+1.6%). These differences could be attributed to species-specific variations. When compared to the CHO dataset, the proteome-derived amino acid frequency diverges more. Particular differences can be seen for glutamate (-4.1%), aspartate (-2.9%), and lysine (-1.5%). As the major deviations with respect to the CHO dataset are observed for amino acids that are products of deamination, this result could be an artefact of the analytical method. Another possibility is that the analytical method used by Selvarasu et al.<sup>51</sup> quantified total amino acid concentration (free amino acids plus those contained in protein). Because glutamate and aspartate are the key amino acids involved in the aspartate-malate shuttle, and CHO cells have high activity of this process<sup>52</sup>, a considerable amount of free glutamate and aspartate would be expected in this cell type. However, we cannot rule out that the deviation is due to the spectral counting quantification method employed.

## **Frequency of N-linked and O-GalNAc glycosylation sites in CHO**

The obtained frequencies for N-linked and O-GalNAc glycosylation sites in CHO HCPs are presented in Table 2. The frequency of O-GalNAc glycosylation is estimated to be 26.64 sites per 100 host cell proteins, while the frequency of N-linked glycosylation is lower, at

8.09 sites per 100 host cell proteins. These values are significantly lower than those derived in previous studies<sup>16,17</sup>, where N-glycosylation frequency was considered to be 170 sites per 100 host cell proteins and 158 O-GalNAc sites per 100 host cell proteins. These studies did not consider HCP relative abundance or the number of glycosylation sites per HCP. Precisely to refine these initial estimates, both of these components are included in our strategy to obtain more representative values of HCP N-linked and O-GalNAc glycosylation sites across the CHO proteome. When considering relative HCP abundance, only ~4% (mol/mol) of CHO HCPs contain N-glycosites and ~6.5% (mol/mol) contain O-GalNAc sites. Furthermore, a recent publication estimates the frequency of N-linked glycosylation sites across the *Pichia pastoris* proteome at 1 site for every 4,691 amino acid residues<sup>53</sup>. Using our estimated value for weighted average CHO HCP length ( $\overline{Len}_{HCP} = 413.6 \text{ aac/HCP}$ ) and N-glycosite frequency ( $S_{Nglyc}^{HCP} = 0.0809 \text{ Sites/HCP}$ ) a value of 1 N-glycosite for every 5,102 amino acid residues is obtained (a deviation of only 8%). Although differences between *P. pastoris* and mammalian cell lines are expected, it would seem unusual for them to span two orders of magnitude. Irani *et al.*<sup>53</sup> obtained their N-glycosite frequency value by dividing the previously reported number of N-linked glycosites in *P. pastoris* by the total number of amino acid residues in that organism's proteome. Our strategy for obtaining O-GalNAc glycosylation site frequency is analogous, given that data for the O-GalNAc glycosylation sites of CHO cells was available<sup>21</sup>. Overall, we believe that the estimates presented herein and those reported by Irani *et al.*<sup>53</sup> are more accurate than the previous ones because HCP relative abundance and glycosylation site frequency have been considered. The relative abundance of all N-linked and O-GalNAc glycoproteins, along with their weighted

contribution to glycosylation site frequency are presented in Supplementary Tables ST3 and ST4, respectively.

An interesting feature of our estimates is that the ten highest contributors (out of a total of 437 predicted N-linked glycoproteins) account for 50% of the overall N-linked glycosite frequency. Similarly, the ten highest contributors to O-GalNAc site frequency (out of 346 predicted O-GalNAc glycoproteins) account for 43% of the total O-GalNAc glycosite frequency (Figure 4). The fact that a small subset of proteins contributes so heavily to CHO HCP glycosylation could prove extremely useful for monitoring HCP glycosylation experimentally. If dynamic variations in abundance, site occupancy, and glycoform distribution of these top ten contributors is representative of the remaining host cell glycoproteins, they could be used as markers for dynamic variations in HCP glycosylation.

Nine of the eighteen most abundant glycoproteins in the CHO proteome are involved in protein folding (Figure 4). Five of these are chaperones (endoplasmic reticulum chaperone protein 1, clusterin, 78 kDa glucose-regulated protein, and serpin H1), while four are either disulphide isomerases (protein disulphide-isomerase A3 and protein disulphide isomerase) or peptide isomerases (peptidyl-prolyl cis-trans isomerase A, peptidyl-prolyl cis-trans isomerase FKB10). It would be interesting to evaluate if the carbohydrates bound to these proteins influence their activity. If so, we could envision a situation where HCP glycosylation is limited by NS availability, which, in turn, would lead to accumulation of unfolded or misfolded proteins in the endoplasmic reticulum. Furthermore, many of the enzymes involved in glycosylation are themselves glycoproteins<sup>54,55</sup>. Again, if activity of these enzymes relies on their carbohydrate composition, reduced NS availability could directly influence recombinant protein

glycosylation while simultaneously decreasing glyco-enzyme activity, thus impacting overall protein glycosylation in a feed-forward manner. Although yet to be confirmed, these mechanisms could cause cascading effects whereby reduced NS availability may negatively impact protein folding as well as HCP and recombinant product glycosylation.

The remaining highly-abundant glycoproteins are cell adhesion and extracellular matrix components (chondroitin sulfate proteoglycan 4, laminin, fibronectin, annexin), cell surface receptors (CD44 antigen, integrin  $\beta$ ), the C-C motif chemokine 2, the Lysosome-associated membrane glycoprotein 1, and the amyloid beta A4 protein. Given the role of carbohydrates in cell-cell interactions<sup>56</sup>, the high observed abundance of cell surface, extracellular matrix, and secreted glycoproteins is expected.

## **Frequency of GAG, GPI, and O-Man glycosylation sites**

### **GAG glycosites**

Eleven proteoglycans were found in the CHO proteome, including syndecan, glypican, agrin, decorin, and lumican (ST10). These proteoglycans are reported to contain heparan, chondroitin, and keratin sulphate GAGs<sup>44</sup>. The total weighted frequency of proteoglycans across the CHO proteome was found to be 0.38 sites per 100 CHO HCPs (ST10). This value is over twenty-fold lower than the one obtained for N-linked glycosites and 70-fold lower than that of O-GalNAc glycosites. These results suggest that, although present, proteoglycans are in low abundance compared to N-linked and O-GalNAc glycoproteins, and do not contribute considerably to NS consumption towards CHO cellular glycosylation, an observation that has also been made experimentally<sup>39</sup>. We therefore believe that our assumption of disregarding these HCP glycosylation types as considerable sinks for NS consumption is valid.

## 1 GPI glycosites

2 The PredGPI GPI-anchor prediction server (<http://gpcr.biocomp.unibo.it/predgpi/>)  
3 identified 48 GPI-anchoring sites across the CHO proteome (ST10). The predicted GPI-  
4 anchoring sites were found on cell surface proteins (e.g. legumain, ephrin, the renin  
5 receptor, tissue plasminogen activator), secretory pathway proteins (e.g. ER-Golgi 24 kDa  
6 SNARE, transmembrane protein 115, Syntaxin-16), as well as various enzymes and  
7 chaperones (e.g. UMP-CMP kinase, lipoprotein lipase, tissue  $\alpha$ -L-fucosidase, proteome  
8 assembly chaperone 2, peptidyl-prolyl cis-trans isomerase FKBP8). The weighted  
9 frequency of free and occupied GPI sites in CHO was found to be 0.447 *sites/100 HCP*  
10 and 1.655 *sites/100 HCP*, respectively. The total number of GPI sites  
11 (2.103 *sites/100 HCP*) is approximately four-fold lower than the frequency of HCP N-  
12 linked glycans and over 12-fold lower than HCP O-GalNAc glycosites. Given that the  
13 difference in glycosite frequency is relatively low, the demand of NSs towards GPI  
14 glycosylation was included in our estimates. The monosaccharide composition of GPI  
15 glycans was assumed to be the one reported for the Hamster prion protein GPI-anchor  
16 (Man<sub>4</sub>Neu5AcGalNAcGlcN)<sup>57</sup>. The stoichiometric coefficients of NS consumption towards  
17 GPI synthesis were calculated using the obtained GPI frequency and the monosaccharide  
18 distribution for the Hamster prion protein GPI anchor, as described for the N-linked and  
19 O-GalNAc glycans in the materials and methods section. The stoichiometric coefficients  
20 for lipid glycosylation presented in Table 2 are the sum of the stoichiometric coefficients  
21 for glycosphingolipid and GPI glycosylation. Individual values for GPI and  
22 glycosphingolipid stoichiometric coefficients are presented in Supplementary table ST9.

## 23 O-Man glycosite frequency

Cadherins and plexins have been recently reported as the major O-mannosylated glycoproteins in humans<sup>58</sup>. This study reports 133 O-Man sites across 37 cadherins and 8 O-Man sites in 6 plexins out of a total of 235 identified O-Man glycosites. A third major source of O-Man glycosites was reported to be  $\alpha$ -dystroglycan, with 13 sites. In order to evaluate whether O-mannose glycans are considerable sinks for NS consumption in CHO cells, the database for closest homologs to the CHO proteome (ST1) was queried for the presence of cadherins, plexins, and  $\alpha$ -dystroglycan. Three cadherins (protocadherin fat1, protocadherin fat 3, and cadherin 9), four plexins (A4, C1, D1, and A1), and  $\alpha$ -dystroglycan were found across the CHO proteome. The total relative abundance of the above seven proteins plus  $\alpha$ -dystroglycan was found to be 0.014% of the CHO proteome. In total, these eight proteins contain 11 confirmed O-Man sites<sup>58</sup>. When this value is multiplied by the total relative abundance of these proteins, this yields a weighted O-Man glycosite frequency of 0.154 sites per 100 HCPs. This value is 50-fold lower than the number of N-linked glycosites and nearly 200-fold lower than that of O-GalNAc glycosites. This result confirms that O-Man glycans are in considerably lower abundance than other types of HCP glycosylation, and do not significantly contribute to NS consumption.

## **Stoichiometric coefficients for NSs consumed towards CHO cell glycosylation**

The stoichiometric coefficients range over three orders of magnitude (Table 2) and are higher for monosaccharides that are abundant in the average glycoform structure (e.g. GlcNAc in N-linked glycosylation) and lower for less abundant species (e.g. Neu5Gc). The high stoichiometric coefficients for GDP-Man and UDP-Glc towards N-linked glycosylation reflect the presence of their corresponding monosaccharides in the

precursor oligosaccharide that initiates N-linked glycosylation, even though most of the mannose and all of the glucose residues are trimmed from the final structure through processing along the secretory pathway. It is worth noting that demand for these NSs may be lower considering that their monosaccharide components could be recycled after being cleaved off of the glycoprotein during the processing reactions.

Interestingly, the obtained stoichiometric coefficients for uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) and uridine diphosphate N-acetylgalactosamine (UDP-GalNAc) towards HCP glycosylation are quite similar ( $1.248 \text{ nmol}/10^6 \text{ cells}$  and  $1.160 \text{ nmol}/10^6 \text{ cells}$ , respectively) despite the notable difference in frequency of each glycosylation type. The difference in N-linked and O-GalNAc glycosite frequencies is offset by the high average GlcNAc content of HCP N-linked glycans ( $3.54 \text{ nmol}_{\text{GlcNAc}}/\text{nmol}_{\text{N-glycan}}$ ) compared with that for O-GalNAc HCP glycosylation ( $1.0 \text{ nmol}_{\text{GalNAc}}/\text{nmol}_{\text{O-glycan}}$ ).

Only two NSs, uridine diphosphate galactose (UDP-Gal) and cytosine monophosphate N-acetylneuraminic acid (CMP-Neu5Ac), are consumed for all considered forms of glycosylation (N-linked and O-GalNAc HCP, lipid glycosylation, and mAb N-linked glycosylation). The highest obtained stoichiometric coefficient for UDP-Gal consumption corresponds to O-GalNAc HCP glycosylation ( $v_{\text{UDPGal}}^{\text{OglycHCP}} = 1.16 \text{ nmol}/10^6 \text{ cells}$ ) and is due to the high frequency of identified O-GalNAc glycosylation sites across the CHO proteome<sup>21</sup> (Table 2) and the presence of galactose in all O-GalNAc glycans<sup>20</sup> (Figure 2).

The stoichiometric coefficient of UDP-Gal towards lipid glycosylation was calculated as  $v_{\text{UDP-Gal}}^{\text{GL}} = 0.623 \text{ nmol}/10^6 \text{ cells}$ , and surpasses that for HCP N-linked glycosylation ( $v_{\text{UDPGal}}^{\text{NglycHCP}} = 0.504 \text{ nmol}/10^6 \text{ cells}$ ). The higher stoichiometric coefficient for lipid



glycosylation is likely due to the presence of galactose in the most abundant glycosphingolipids reported for CHO cells<sup>22,59</sup>.

Similarly to UDP-Gal, the highest obtained stoichiometric coefficient for CMP-Neu5Ac consumption is towards O-GalNAc HCP glycosylation ( $v_{UDPGal}^{OglycHCP} = 1.443 \text{ nmol}/10^6 \text{ cells}$ ). Again, this is due to the high frequency of O-GalNAc glycans across the CHO proteome and the high Neu5Ac content (at least one residue per glycan) of O-GalNAc glycans (Figure 2). The second largest coefficient for CMP-Neu5Ac consumption is towards lipid glycosylation ( $v_{CMPNeu5Ac}^{GL} = 0.403 \text{ nmol}/10^6 \text{ cells}$ ), and is due to the presence of Neu5Ac on the G<sub>M3</sub> ganglioside, which is the most abundant glycosphingolipid in CHO cells<sup>22,59</sup>. The low stoichiometric demand of CMP-Neu5Ac towards HCP N-linked glycosylation ( $v_{CMPNeu5Ac}^{NglycHCP} = 0.034 \text{ nmol}/10^6 \text{ cells}$ ) is due to the low abundance of the corresponding monosaccharide in CHO HCP N-linked glycans (Figure 2).

## Partition of NS fluxes towards host cell and rTP glycosylation

### Total NS fluxes

Total NS fluxes towards cellular and mAb glycosylation were estimated as described in the materials and methods section and Supplementary table ST11. The highest demand of NSs towards cellular glycosylation can be observed at  $R_{q_p/\mu_g}^{Low}$ , whereas the highest demand for product glycosylation will be observed at  $R_{q_p/\mu_g}^{High}$ , thus bounding a range of metabolic demand for cellular versus product glycosylation. The values for  $q_p$  and  $\mu_g$  extracted from these publications were multiplied by the corresponding stoichiometric coefficients presented in Table 2 to obtain the percentage of NS fluxes directed towards cellular N-linked, O-GalNAc, and lipid glycosylation, as well as product N-linked

glycosylation (Figure 5). For analysis, the fluxes of UDP-GlcNAc and UDP-GalNAc towards cellular glycosylation have been summed (UDP-HexNAc), given that they share a common biosynthetic pathway where UDP-GalNAc is produced from the epimerization reaction of UDP-GlcNAc<sup>16</sup>.

Figure 5 shows that the flux of UDP-Gal and CMP-sialic acid towards glycosylation is highest at  $R_{q_p/\mu_g}^{Low}$ . This trend is possibly due to the higher total protein productivity (HCP and mAb) observed under this regime (Table 1), which ranges from  $6.93 \frac{pg_{prot}}{cell\ h}$  to  $7.62 \frac{pg_{prot}}{cell\ h}$  at  $R_{q_p/\mu_g}^{Low}$ , and is reduced to between  $3.20 \frac{pg_{prot}}{cell\ h}$  to  $6.11 \frac{pg_{prot}}{cell\ h}$  for  $R_{q_p/\mu_g}^{High}$ . In contrast, more guanosine diphosphate fucose (GDP-Fuc) is consumed for  $R_{q_p/\mu_g}^{High}$  in the Chus and Kant datasets, while UDP-HexNAc is also consumed more for  $R_{q_p/\mu_g}^{High}$  in the Kant dataset (Figure 5). Specifically, the flux of GDP-Fuc towards glycosylation ( $\frac{pmol}{10^6 cells\ h}$ ) goes from 24.5 (Chus-L) to 28.1 (Chus-H) and from 27.2 (Kant-L) to 49.4 (Kant-H) (Figure 5C-F). Similarly, the UDP-HexNAc flux increases from 157.2 (Kant-L) to 219.8 (Kant-H). For GDP-Fuc, this trend is likely due to the high percentage of protein productivity destined for mAb synthesis in the Chus-H and Kant-H datasets (48.8% and 64.1%, respectively), along with the considerably lower fucose content of HCP glycans ( $0.40 \frac{mol_{Fuc}}{mol_{HCP\ glycan}}$ ) compared to that of mAb glycans ( $0.91 \frac{mol_{Fuc}}{mol_{mAb\ glycan}}$ ). Similar effects cause the observed increase in UDP-HexNAc flux towards glycosylation in the Kant-H scenario, although it may mainly occur due to the high  $q_p$ , given that the HexNAc content ( $\frac{mol_{HexNAc}}{mol_{glycan}}$ ) of HCP and mAb glycans is relatively similar (4.54 and 3.65, respectively).

## Flux distributions towards cellular and mAb glycosylation

## **GDP-Fuc and UDP-HexNAc flux towards HCP and mAb glycosylation**

GDP-Fuc is exclusively consumed for N-linked glycosylation and its flux is distributed between host cell proteins and recombinant mAb according to demand. In each of the cases tested, the majority of GDP-Fuc flux goes towards mAb production (between 73% and 97%) due to the low abundance of fucosylated species in HCP glycans (Figure 2). The flux of UDP-HexNAc is more evenly distributed between HCP and mAb glycosylation. Except for the Dool-L dataset, more than half of the UDP-HexNAc flux goes towards mAb glycosylation (Figure 5), which is likely due to the high GlcNAc content of mAb N-glycans (Figure 2). Despite the typically higher demand towards mAb glycosylation, the flux of UDP-HexNAc towards HCP glycosylation is significant and ranges between 11.9% and 59.8%. UDP-HexNAc consumption towards HCP glycosylation is almost evenly split between N-linked and O-GalNAc glycosylation. This arises from the similar stoichiometric coefficients obtained for UDP-GlcNAc towards HCP N-linked glycosylation and UDP-GalNAc towards HCP mucin-type glycosylation.

## **UDP-Gal and CMP-Neu5Ac flux towards cellular and mAb glycosylation**

For all scenarios, HCP mucin-type glycosylation consumes the majority of CMP-Neu5Ac (between 73.8% and 75.8%), and the second highest demand for this NS is observed for lipid glycosylation (21.1% to 21.7%). Interestingly, the distribution of CMP-Neu5Ac consumption is almost independent of  $R_{q_p/\mu_g}$ , which is due to the scarcity of sialic acids in the mAb. For this reason, CMP-Neu5Ac consumption depends almost exclusively on the specific growth rate of CHO cells, and thus the distribution of its consumption remains nearly constant across different  $R_{q_p/\mu_g}$  values.

The fraction of UDP-Gal flux consumed towards mucin-type glycosylation varies between 23.9% and 46.2% and is higher under conditions where  $R_{q_p/\mu_g}$  is low (and thus, antibody synthesis rate and utilisation of UDP-Gal for mAb glycosylation is low). Under  $R_{q_p/\mu_g}^{Low}$ , the UDP-Gal flux towards glycolipid synthesis is the next highest (21.9% to 24.8%), followed closely by HCP N-linked glycosylation (17.7% to 20.1%). Due to the low mAb productivity to growth ratio, only between 8.8% and 19.6% of the consumed UDP-Gal goes towards mAb N-linked glycosylation. We must note that UDP-Gal consumption towards mAb glycosylation surpasses HCP N-linked glycosylation in the Kant-L dataset, a result that is due to the relatively high  $R_{q_p/\mu_g}^{Low}$  value observed for this engineered cell line.

Under the  $R_{q_p/\mu_g}^{High}$  scenarios, consumption of UDP-Gal towards mAb glycosylation surpasses that of any other type of glycosylation, except for the Doolan-H dataset, where UDP-Gal consumption towards HCP O-GalNAc glycosylation is higher. Across all tested conditions for  $R_{q_p/\mu_g}$ , between 47.2% and 91.2% of UDP-Gal is consumed towards HCP N-linked, O-GalNAc, and lipid glycosylation, the only exception being the Kant-H dataset where 52.8% of all consumed UDP-Gal goes towards mAb glycosylation. In this context, we must highlight that the Kant-H data correspond to engineered CHO cells (overexpressing anti-apoptotic genes) being cultured under mild hypothermic conditions and in presence of sodium butyrate<sup>26</sup>. Under these conditions, the cells characterised by Kantardjieff *et al.* yield an exceptionally high specific mAb productivity (~100pg/cell/day), which causes the observed demand in UDP-Gal towards mAb glycosylation. Thus, it is possible to consider all other tested scenarios of  $R_{q_p/\mu_g}$  as more representative of typical CHO cell culture. In these typical CHO culture scenarios, more UDP-Gal is consumed towards cellular glycosylation than to product glycosylation, which

is particularly relevant when considering NS precursor feeding strategies. Our results suggest that over half of NS precursors fed to control mAb galactosylation would be destined to cellular components and would not reach their intended target. This is consistent with previous findings where cell surface galactosylation was observed to increase more (up to four-fold) than mAb galactose content (up to two-fold) with different uridine-manganese-galactose feeding strategies<sup>12</sup>.

The changes in UDP-Gal flux distributions observed at varying  $R_{qp/\mu_g}$  are the broadest observed in this study and are most likely due to the fact that this NS is the only one that is consumed for all considered forms of glycosylation (HCP N-linked and O-GalNAc, lipid, and mAb glycosylation). Given the observed variability of UDP-Gal flux distributions, it is no surprise that cell surface galactose content has been reported to correlate well with antibody galactosylation<sup>12</sup>. Furthermore, these results indicate that the partition of UDP-Gal between cellular and product glycosylation may be one of the underlying causes of galactosylation-associated microheterogeneity of mAbs produced in CHO: if a substantial proportion of available UDP-Gal is continuously consumed for cellular glycosylation, availability of this NS may fluctuate considerably throughout cell culture, and thus, lead to variations in rTP galactose content.

When considering that excess uridine and galactose supplementation have been reported to hinder cell growth and final product titre<sup>10,12,13</sup>, our results for UDP-Gal flux distributions further highlight the importance of having accurate estimates for NS consumption towards HCP glycosylation. Indicative estimates for the consumption of UDP-Gal towards HCP glycosylation are necessary to determine optimal uridine and galactose feeding strategies that ensure adequate rTP galactosylation while minimising (or completely avoiding) deleterious effects on cell growth and product yield.

The results obtained with our proposed framework are in qualitative agreement with previous findings, but the estimates could be influenced by: *i*) variations in the relative abundance of CHO HCPs and lipids, *ii*) variations in the occupancy of N-linked, O-GalNAc, and lipid glycosylation sites, and *iii*) changes in glycosylation microheterogeneity. All three of the above are likely to vary over time, depending on cell culture conditions, NS precursor supplementation strategy, growth phase, and cellular metabolic state. Indeed (and possibly due to a combination of proteomic, macroheterogeneity, and microheterogeneity effects), qualitative variations in CHO cell surface glycosylation have been reported<sup>12,60</sup>. The overall framework presented herein could be readily applied, via the CHO cellular glycosylation calculator provided in the supplementary file, to any experimental data so that variations in the demand for NSs towards cellular and rTGP glycosylation can be represented for different cell lines grown under different culture conditions. Experimentally, these limitations can also be addressed by monitoring the relative abundance and glycosylation of the subset of CHO proteins that were found to contribute most to HCP glycosylation.

## Conclusions

The work presented herein describes a strategy to estimate the metabolic demand of NSs towards protein and lipid glycosylation in CHO cells. The estimate combines the relative abundance of individual CHO host proteins with the number of N-linked and O-GalNAc glycosylation sites present on each HCP and the reported CHO HCP glycome. The stoichiometric demand of NSs towards glycolipid (glycosphingolipid and GPI anchor) synthesis has also been included. Overall, our results show that the consumption of NSs towards HCP glycosylation is, in most cases, significant and cannot be neglected when rationally designing NS precursor feeding strategies or while developing mechanistic

mathematical models for cell growth and recombinant protein glycosylation. The obtained stoichiometric coefficients for NS consumption towards cellular glycosylation are a first approximation towards a fuller quantitative understanding of how the process of HCP, glycosphingolipid, and rTP glycosylation are integrated. In addition, we have identified a subset of CHO HCPs which, if monitored for abundance, glyco-site occupancy, and glycan microheterogeneity, could refine the obtained stoichiometric coefficients, and thus improve our quantitative understanding of cellular and rTP glycosylation. In the future, analysis of the overall burden of protein glycosylation on cellular metabolism will lead to mechanistically-defined feeding strategies for quantitative and optimal control of rTP glycosylation.

## Acknowledgments

The authors gratefully acknowledge the financial support of the Bioprocess Research Industry Club, jointly funded by the UK BBSRC, EPSRC, and industrial members (BB/I017011/1). KMP and CK thank Research Councils UK for their fellowships. CK thanks LONZA Biologics for their financial support of her fellowship.

## Author contributions

IJV collected the data and developed the spreadsheets. IJV, KMP, and CK designed the study, analysed the data, and wrote the paper. All authors discussed and commented on the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

1 **Competing financial interests:** The authors declare no competing financial interests.

UNCORRECTED PROOF



# 1 References

- 2 1 Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nat*  
3 *Rev Drug Discov* **8**, 226-234 (2009).
- 4 2 Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug*  
5 *Discov Today* **15**, 773-780, doi:DOI 10.1016/j.drudis.2010.06.009 (2010).
- 6 3 EvaluatePharma. World Preview 2015, Outlook to 2020. 69 (Evaluate Ltd., London,  
7 UK, 2015).
- 8 4 del Val, I. J., Kontoravdi, C. & Nagy, J. M. Towards the Implementation of Quality by  
9 Design to the Production of Therapeutic Monoclonal Antibodies with Desired  
10 Glycosylation Patterns. *Biotechnol Progr* **26**, 1505-1527, doi:Doi 10.1002/Btpr.470  
11 (2010).
- 12 5 Varki, A. *Essentials of glycobiology*. 2nd edn, (Cold Spring Harbor Laboratory Press,  
13 2009).
- 14 6 Nyberg, G. B., Balcarcel, R. R., Follstad, B. D., Stephanopoulos, G. & Wang, D. I. C.  
15 Metabolic effects on recombinant interferon-gamma glycosylation in continuous  
16 culture of Chinese hamster ovary cells. *Biotechnol. Bioeng.* **62**, 336-347,  
17 doi:10.1002/(sici)1097-0290(19990205)62:3<336::aid-bit10>3.0.co;2-n (1999).
- 18 7 Wong, D. C. F., Wong, K. T. K., Goh, L. T., Heng, C. K. & Yap, M. G. S. Impact of dynamic  
19 online fed-batch strategies on metabolism, productivity and N-glycosylation quality  
20 in CHO cell cultures. *Biotechnol. Bioeng.* **89**, 164-177, doi:10.1002/bit.20317  
21 (2005).
- 22 8 Kochanowski, N. *et al.* Influence of intracellular nucleotide and nucleotide sugar  
23 contents on recombinant interferon-gamma glycosylation during batch and fed-  
24 batch cultures of CHO cells. *Biotechnol. Bioeng.* **100**, 721-733,  
25 doi:10.1002/bit.21816 (2008).
- 26 9 Jedrzejewski, P., del Val, I., Polizzi, K. & Kontoravdi, C. Applying quality by design to  
27 glycoprotein therapeutics: experimental and computational efforts of process  
28 control. *Pharm Bioprocess* **1**, 51-69, doi:doi:10.4155/pbp.13.4 (2013).
- 29 10 Carvalhal, A. V., Santos, S. S., Calado, J., Haury, M. & Carrondo, M. J. Cell growth arrest  
30 by nucleotides, nucleosides and bases as a tool for improved production of  
31 recombinant proteins. *Biotechnol Prog* **19**, 69-83, doi:10.1021/bp0255917 (2003).
- 32 11 Yang, M. & Butler, M. Effects of ammonia and glucosamine on the heterogeneity of  
33 erythropoietin glycoforms. *Biotechnol Prog* **18**, 129-138, doi:10.1021/bp0101334  
34 (2002).
- 35 12 Grainger, R. K. & James, D. C. CHO cell line specific prediction and control of  
36 recombinant monoclonal antibody N-glycosylation. *Biotechnol. Bioeng.* **110**, 2970-  
37 2983, doi:10.1002/bit.24959 (2013).
- 38 13 Altamirano, C., Cairo, J. J. & Godia, F. Decoupling cell growth and product formation  
39 in Chinese hamster ovary cells through metabolic control. *Biotechnol. Bioeng.* **76**,  
40 351-360 (2001).
- 41 14 Blondeel, E. J. *et al.* Tuning a MAb glycan profile in cell culture: Supplementing N-  
42 acetylglucosamine to favour G0 glycans without compromising productivity and  
43 cell growth. *J Biotechnol* **214**, 105-112, doi:10.1016/j.jbiotec.2015.09.014 (2015).
- 44 15 Ohadi, K. *et al.* in *12th IFAC Symposium on Computer Applications in Biotechnology*  
45 (The International Federation of Automatic Control, Mumbai, India, 2014).

- 16 Jedrzejewski, P. M. *et al.* Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation. *International journal of molecular sciences* **15**, 4492-4522, doi:10.3390/ijms15034492 (2014).
- 17 Sou, S. N. *et al.* How does mild hypothermia affect monoclonal antibody glycosylation. *Biotechnol. Bioeng.* **112**, 1165-1176 (2014).
- 18 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Bba-Gen Subjects* **1473**, 4-8, doi:10.1016/S0304-4165(99)00165-8 (1999).
- 19 Baycin-Hizal, D. *et al.* Proteomic Analysis of Chinese Hamster Ovary Cells. *J Proteome Res* **11**, 5265-5276, doi:10.1021/pr300476w (2012).
- 20 North, S. J. *et al.* Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals N-Glycans of a Novel Size and Complexity. *J Biol Chem* **285**, 5759-5775, doi:10.1074/jbc.M109.068353 (2010).
- 21 Yang, Z. *et al.* The GalNAc-type O-glycoproteome of CHO cells characterized by the SimpleCell strategy. *Molecular & cellular proteomics : MCP*, doi:10.1074/mcp.M114.041541 (2014).
- 22 Briles, E. B., Li, E. & Kornfeld, S. Isolation of wheat germ agglutinin-resistant clones of Chinese hamster ovary cells deficient in membrane sialic acid and galactose. *J Biol Chem* **252**, 1107-1116 (1977).
- 23 Nakano, M. *et al.* Capillary electrophoresis-electrospray ionization mass spectrometry for rapid and sensitive N-glycan analysis of glycoproteins as 9-fluorenylmethyl derivatives. *Glycobiology* **19**, 135-143 (2009).
- 24 Doolan, P. *et al.* Transcriptomic analysis of clonal growth rate variation during CHO cell line development. *J Biotechnol* **166**, 105-113, doi:10.1016/j.jbiotec.2013.04.014 (2013).
- 25 Chusainow, J. *et al.* A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol. Bioeng.* **102**, 1182-1196, doi:10.1002/bit.22158 (2009).
- 26 Kantardjieff, A. *et al.* Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. *J Biotechnol* **145**, 143-159, doi:10.1016/j.jbiotec.2009.09.008 (2010).
- 27 UniProt Consortium. UniProt: a hub for protein information. *Nucleic acids res* **43**, D204-212, doi:10.1093/nar/gku989 (2015).
- 28 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Bio* **215**, 403-410, doi:10.1006/jmbi.1990.9999 (1990).
- 29 Zybaylov, B., Coleman, M. K., Florens, L. & Washburn, M. P. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Analytical chemistry* **77**, 6218-6224, doi:10.1021/ac050846r (2005).
- 30 Zybaylov, B. *et al.* Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **5**, 2339-2347, doi:10.1021/pr060161n (2006).
- 31 Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**, doi:10.1038/srep00090 (2011).
- 32 Gupta, R., Jung, E. & Brunak, S. *NetNGlyc 1.0 Server: Prediction of N-glycosylation sites in human proteins*, <<http://www.cbs.dtu.dk/services/NetNGlyc/>> (2004) (Date of access: 30/10/2015).

- 33 Houel, S. *et al.* N- and O-glycosylation analysis of etanercept using liquid chromatography and quadrupole time-of-flight mass spectrometry equipped with electron-transfer dissociation functionality. *Analytical chemistry* **86**, 576-584, doi:10.1021/ac402726h (2014).
- 34 Carinhas, N. *et al.* Metabolic Signatures of GS-CHO Cell Clones Associated With Butyrate Treatment and Culture Phase Transition. *Biotechnol. Bioeng.* **110**, 3244-3257, doi:10.1002/bit.24983 (2013).
- 35 Sheikh, K., Forster, J. & Nielsen, L. K. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*. *Biotechnol Prog* **21**, 112-121, doi:10.1021/Bp0498138 (2005).
- 36 Lee, M. S. & Lee, G. M. Effect of hypoosmotic pressure on cell growth and antibody production in recombinant Chinese hamster ovary cell culture. *Cytotechnology* **36**, 61-69, doi:10.1023/A:1014032701800 (2001).
- 37 Nolan, R. P. & Lee, K. Dynamic model of CHO cell metabolism. *Metab Eng* **13**, 108-124, doi:10.1016/j.ymben.2010.09.003 (2011).
- 38 Bonarius, H. P. *et al.* Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol. Bioeng.* **50**, 299-318, doi:10.1002/(SICI)1097-0290(19960505)50:3<299::AID-BIT9>3.0.CO;2-B (1996).
- 39 Norgard-Sumnicht, K., Bai, X., Esko, J. D., Varki, A. & Manzi, A. E. Exploring the outcome of genetic modifications of glycosylation in cultured cell lines by concurrent isolation of the major classes of vertebrate glycans. *Glycobiology* **10**, 691-700 (2000).
- 40 Fujitani, N. *et al.* Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. *Proc Natl Acad Sci U S A* **110**, 2105-2110, doi:10.1073/pnas.1214233110 (2013).
- 41 Liu, B. *et al.* The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. *Journal of Biotechnology* **170**, 17-27, doi:10.1016/j.jbiotec.2013.11.007 (2014).
- 42 Gao, N. & Lehrman, M. A. Coupling of the dolichol-P-P-oligosaccharide pathway to translation by perturbation-sensitive regulation of the initiating enzyme, GlcNAc-1-P transferase. *J Biol Chem* **277**, 39425-39435, doi:10.1074/jbc.M205195200 (2002).
- 43 Cantagrel, V. & Lefeber, D. J. From glycosylation disorders to dolichol biosynthesis defects: a new class of metabolic diseases. *J Inherit Metab Dis* **34**, 859-867, doi:10.1007/s10545-011-9301-0 (2011).
- 44 Iozzo, R. V. & Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol* **42**, 11-55, doi:10.1016/j.matbio.2015.02.003 (2015).
- 45 Pierleoni, A., Martelli, P. L. & Casadio, R. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9**, 392, doi:10.1186/1471-2105-9-392 (2008).
- 46 Omasa, T., Onitsuka, M. & Kim, W. D. Cell engineering and cultivation of chinese hamster ovary (CHO) cells. *Curr Pharm Biotechnol* **11**, 233-240 (2010).
- 47 Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs* **2**, 466-479 (2010).
- 48 De Jesus, M. & Wurm, F. M. Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors. *Eur J Pharm Biopharm* **78**, 184-188, doi:10.1016/j.ejpb.2011.01.005 (2011).

- 1 49 Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0  
2 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and  
3 cell-lines. *Proteomics* **15**, 3163-3168, doi:10.1002/pmic.201400441 (2015).
- 4 50 Geiger, T. *et al.* Initial quantitative proteomic map of 28 mouse tissues using the  
5 SILAC mouse. *Molecular & cellular proteomics : MCP* **12**, 1709-1722,  
6 doi:10.1074/mcp.M112.024919 (2013).
- 7 51 Selvarasu, S. *et al.* Combined in silico modeling and metabolomics analysis to  
8 characterize fed-batch CHO cell culture. *Biotechnol. Bioeng.* **109**, 1415-1429,  
9 doi:10.1002/bit.24445 (2012).
- 10 52 Zagari, F. *et al.* High expression of the aspartate-glutamate carrier Aralar1 favors  
11 lactate consumption in CHO cell culture. *Pharmaceutical Bioprocessing* **1**, 19-27,  
12 doi:doi:10.4155/pbp.13.5 (2013).
- 13 53 Irani, Z. A., Kerkhoven, E. J., Shojaosadati, S. A. & Nielsen, J. Genome-scale metabolic  
14 model of *Pichia pastoris* with native and humanized glycosylation of recombinant  
15 proteins. *Biotechnol. Bioeng.*, doi:10.1002/bit.25863 (2015).
- 16 54 Tempel, W. *et al.* Structure of mouse Golgi alpha-mannosidase IA reveals the  
17 molecular basis for substrate specificity among class 1 (family 47  
18 glycosylhydrolase) alpha1,2-mannosidases. *J Biol Chem* **279**, 29774-29786,  
19 doi:10.1074/jbc.M403065200 (2004).
- 20 55 Liu, T. *et al.* Human plasma N-glycoproteome analysis by immunoaffinity  
21 subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* **4**, 2070-  
22 2080, doi:10.1021/pr0502065 (2005).
- 23 56 Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation:  
24 diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**, 448-462,  
25 doi:10.1038/nrm3383 (2012).
- 26 57 Paulick, M. G. & Bertozzi, C. R. The glycosylphosphatidylinositol anchor: a complex  
27 membrane-anchoring structure for proteins. *Biochemistry* **47**, 6991-7000,  
28 doi:10.1021/bi8006324 (2008).
- 29 58 Vester-Christensen, M. B. *et al.* Mining the O-mannose glycoproteome reveals  
30 cadherins as major O-mannosylated glycoproteins. *Proc Natl Acad Sci U S A* **110**,  
31 21018-21023, doi:10.1073/pnas.1313446110 (2013).
- 32 59 Warnock, D. E. *et al.* Determination of plasma membrane lipid mass and  
33 composition in cultured Chinese hamster ovary cells using high gradient magnetic  
34 affinity chromatography. *J Biol Chem* **268**, 10145-10153 (1993).
- 35 60 Tateno, H. *et al.* A novel strategy for mammalian cell surface glycome profiling using  
36 lectin microarray. *Glycobiology* **17**, 1138-1146, doi:10.1093/glycob/cwm084  
37 (2007).

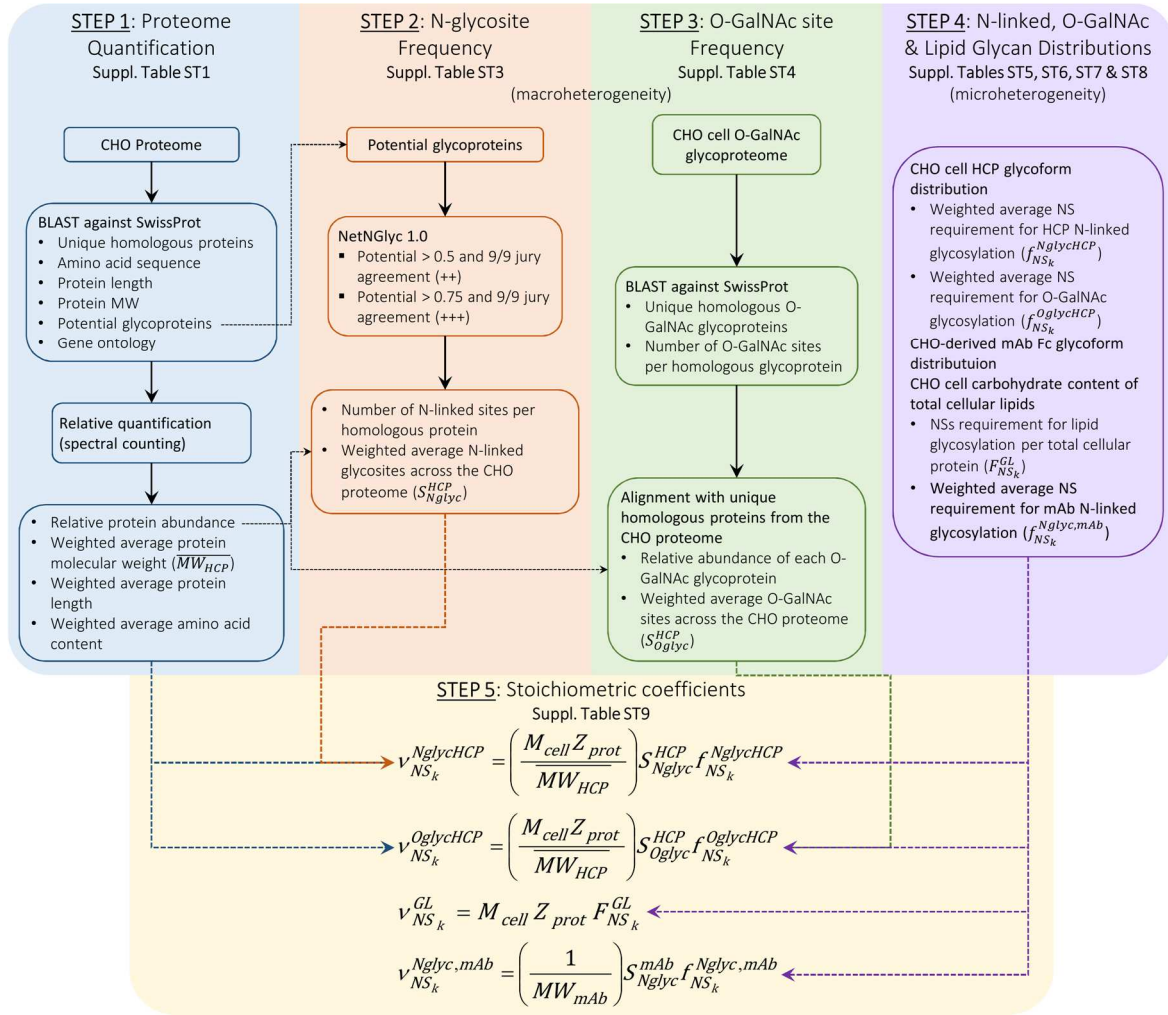
1 **Table 1. mAb productivity to growth rate ratios and total protein synthesis rates**

	Productivity to growth ratios ( $R_{q_p/\mu_g}$ )					
	Low			High		
	Doolan et al. <sup>24</sup>	Chusainow et al. <sup>25</sup>	Kantardjieff et al. <sup>26</sup>	Doolan et al. <sup>24</sup>	Chusainow et al. <sup>25</sup>	Kantardjieff et al. <sup>26</sup>
$\mu_g$ ( $h^{-1}$ )	0.031	0.03	0.025	0.011	0.011	0.011
$q_p$ ( $\frac{pg}{cell\ day}$ )	23.1	39.95	46.52	23.7	52.07	94.05
$R_{q_p/\mu_g}$ ( $\frac{pg_{produced}}{cell_{produced}}$ )	31.1	56.2	78.2	89.8	192	359.5
	Protein synthesis rates					
	Low			High		
	Doolan et al. <sup>24</sup>	Chusainow et al. <sup>25</sup>	Kantardjieff et al. <sup>26</sup>	Doolan et al. <sup>24</sup>	Chusainow et al. <sup>25</sup>	Kantardjieff et al. <sup>26</sup>
HCP ( $\frac{pg_{prot}}{cell\ h}$ )	6.23	5.95	4.99	2.21	2.27	2.19
mAb ( $\frac{pg_{prot}}{cell\ h}$ )	0.96	1.66	1.94	0.99	2.17	3.92
Total ( $\frac{pg_{prot}}{cell\ h}$ )	7.20	7.62	6.93	3.20	4.44	6.11
% to mAb	13.4%	21.9%	28.0%	30.9%	48.8%	64.1%

**Table 2. HCP glycan frequency and NS stoichiometric coefficients for CHO cell and mAb glycosylation**

HCP Glycan frequency $\left(\frac{sites}{100\ HCPs}\right)$	N-linked		O-GalNAc		
	8.09		26.64		
Nucleotide sugar stoichiometric coefficients					
	N-linked HCP $\left(\frac{nmol_i}{10^6\ cells}\right)$	O-GalNAc HCP $\left(\frac{nmol_i}{10^6\ cells}\right)$	HCP Total $\left(\frac{nmol_i}{10^6\ cells}\right)$	Lipids $\left(\frac{nmol_i}{10^6\ cells}\right)$	N-linked mAb $\left(\frac{nmol_i}{mg_{mAb}}\right)$
UDP-GlcNAc	1.248	----	1.248	----	49.14
GDP-Man	3.172	----	3.172	0.366	121.2
UDP-Glc	1.057	----	1.057	0.503	40.39
UDP-Gal	0.504	1.160	1.665	0.623	7.119
GDP-Fuc	0.140	----	0.140	----	12.23
CMP-Neu5Ac	0.034	1.409	1.443	0.403	0.155
CMP-Neu5Gc	0.007	----	0.007	----	----
UDP-GalNAc	----	1.160	1.160	0.092	----

# 1 List of Figures



2

3 **Figure 1. Workflow for the estimation of stoichiometric coefficients for nucleotide**  
4 **sugar demand towards CHO cell and mAb glycosylation**

5 Step 1: the CHO proteome<sup>19</sup> was BLASTed<sup>28</sup> against the SwissProt database<sup>27</sup> to obtain  
6 the corresponding confirmed homologous proteins. The length, amino acid sequence,  
7 molecular weight, and potential N-linked glycosylation sites for each homologous protein  
8 were extracted from the database. Using the protein length and the spectral counts  
9 available from literature<sup>19</sup>, the relative abundance of each homologous protein ( $Z_i$ ) was

obtained<sup>29,30</sup> (Eq. 2). With  $Z_i$ , the weighted average protein length ( $\overline{Len}_{HCP}$ ), molecular weight ( $\overline{MW}_{HCP}$ ), and amino acid content were calculated (Eqs. 3-5).

Step 2: the amino acid sequences of potentially N-glycosylated homologous proteins were analysed with the NetNGlyc 1.0 server<sup>32</sup> to obtain the number of occupied N-linked sites per HCP ( $S_{Nglyc,i}^{HCP}$ ). The weighted average number of N-linked glycosites across the CHO proteome ( $S_{Nglyc}^{HCP}$ ) was obtained by combining  $S_{Nglyc,i}^{HCP}$  with  $Z_i$  (Eq. 6).

Step 3: the CHO O-GalNAc glycoproteome<sup>21</sup> was BLASTed<sup>28</sup> against the SwissProt database<sup>27</sup>. The obtained list of homologous proteins was aligned with the one from the CHO proteome to obtain the weighted average number of O-GalNAc glycosites across the CHO proteome ( $S_{Oglyc}^{HCP}$ ) (Eq. 7).

Step 4: the weighted average monosaccharide composition of CHO HCP N-linked ( $f_{NS_k}^{NglycHCP}$ ) and O-GalNAc glycans ( $f_{NS_k}^{OglycHCP}$ ) was obtained from the CHO glycome<sup>20</sup> (Figure 2) using Eqs. 8-9. The average N-linked glycan monosaccharide composition for a commercial mAb ( $f_{NS_k}^{Nglyc,mAb}$ ) was also obtained from literature<sup>23</sup> (Figure 2). The demand of NSs towards glycolipid synthesis was obtained from the reported monosaccharide content of CHO cell glycosphingolipids relative to total protein content ( $F_{NS_k}^{GL}$ )<sup>22</sup>.

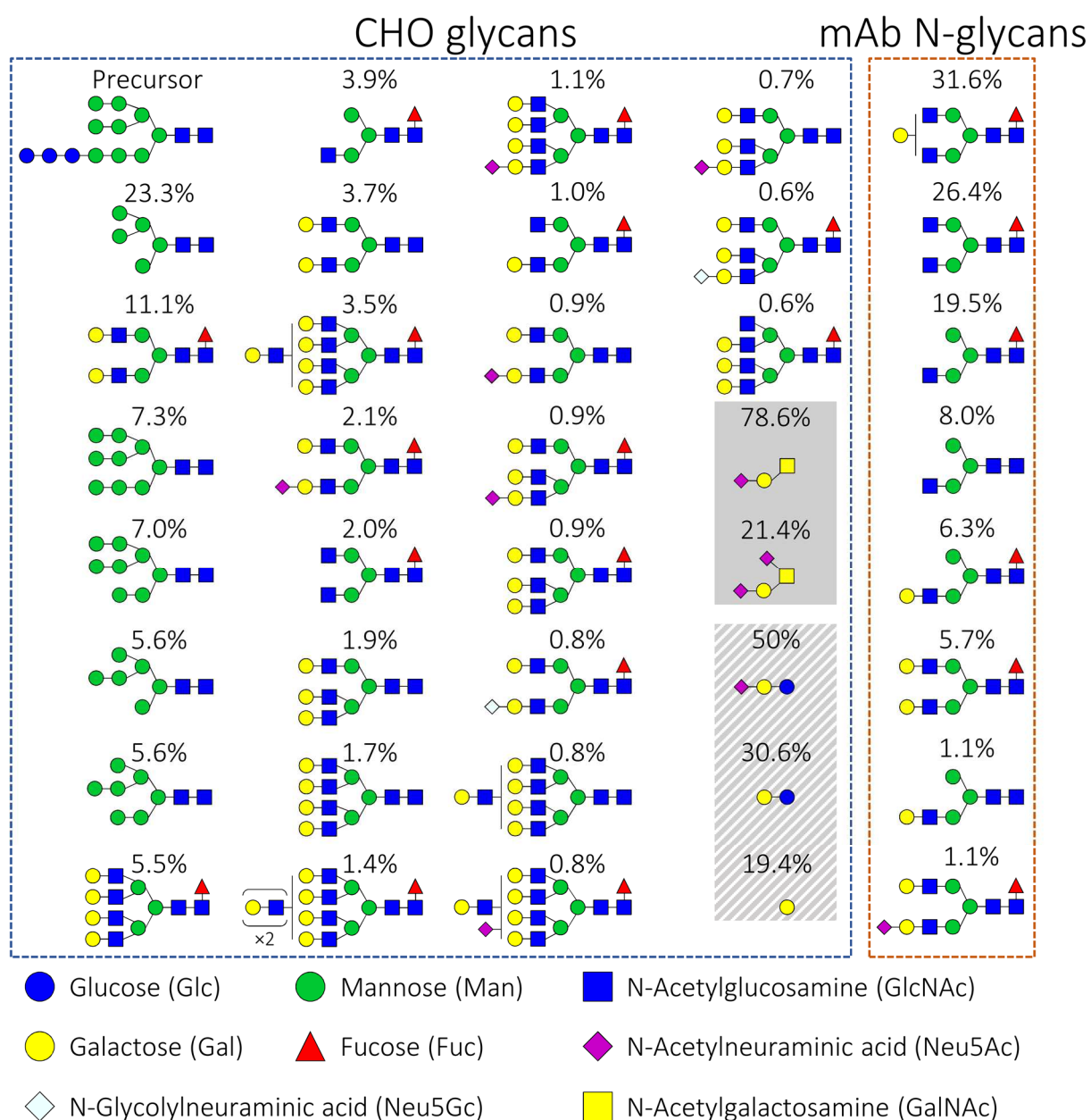
Step 5: the stoichiometric coefficients for the demand of NSs towards HCP N-linked ( $v_{NS_k}^{NglycHCP}$ ), O-GalNAc ( $v_{NS_k}^{OglycHCP}$ ), lipid ( $v_{NS_k}^{GL}$ ), and mAb N-linked ( $v_{NS_k}^{mAb,Nglyc}$ ) glycosylation were calculated using the corresponding equations included at the bottom of the figure using a dry cell weight ( $M_{cell}$ ) of 271pg/cell<sup>34</sup>, a protein content ( $Z_{prot}$ ) of



1 74.2% (w/w)<sup>35</sup>, and a mAb molecular weight ( $MW_{mAb}$ ) of 148,545g/mol. All data and  
2 calculations are presented in Supplementary Tables ST1-ST10.

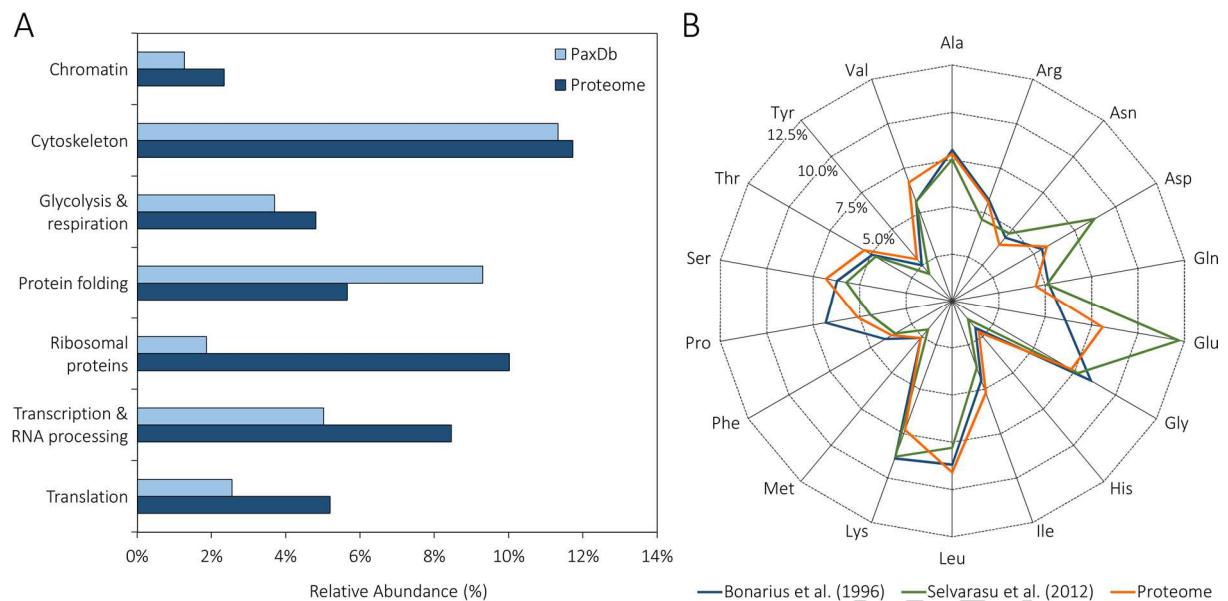
3

UNCORRECTED PROOF



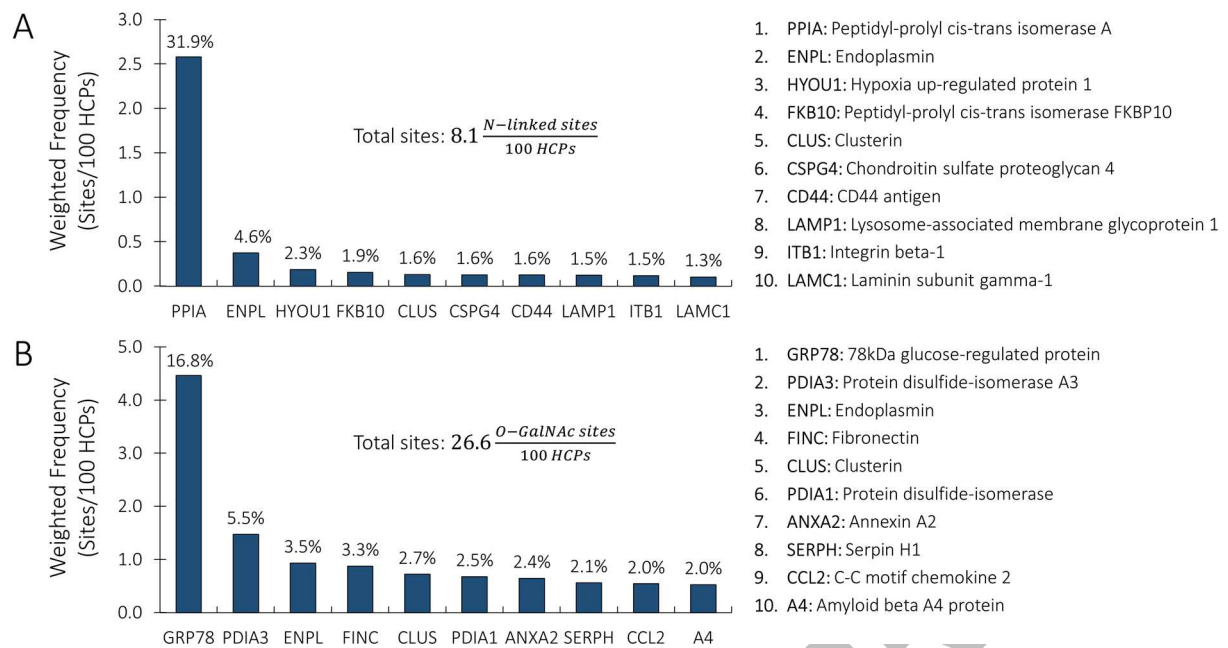
**Figure 2. Glycan distributions produced by CHO cells**

The distribution of N-linked and O-GalNAc glycans found on CHO HCPs is presented within the blue dashed box<sup>20</sup>. The O-GalNAc glycans are highlighted within the solid shaded box. The distribution of lipid-bound glycans is shown in the belted grey box<sup>22</sup>. The average composition of mAb Fc-bound N-linked glycans present on a commercial mAb (trastuzumab) produced by CHO cells<sup>23</sup> is shown inside the orange dashed box.



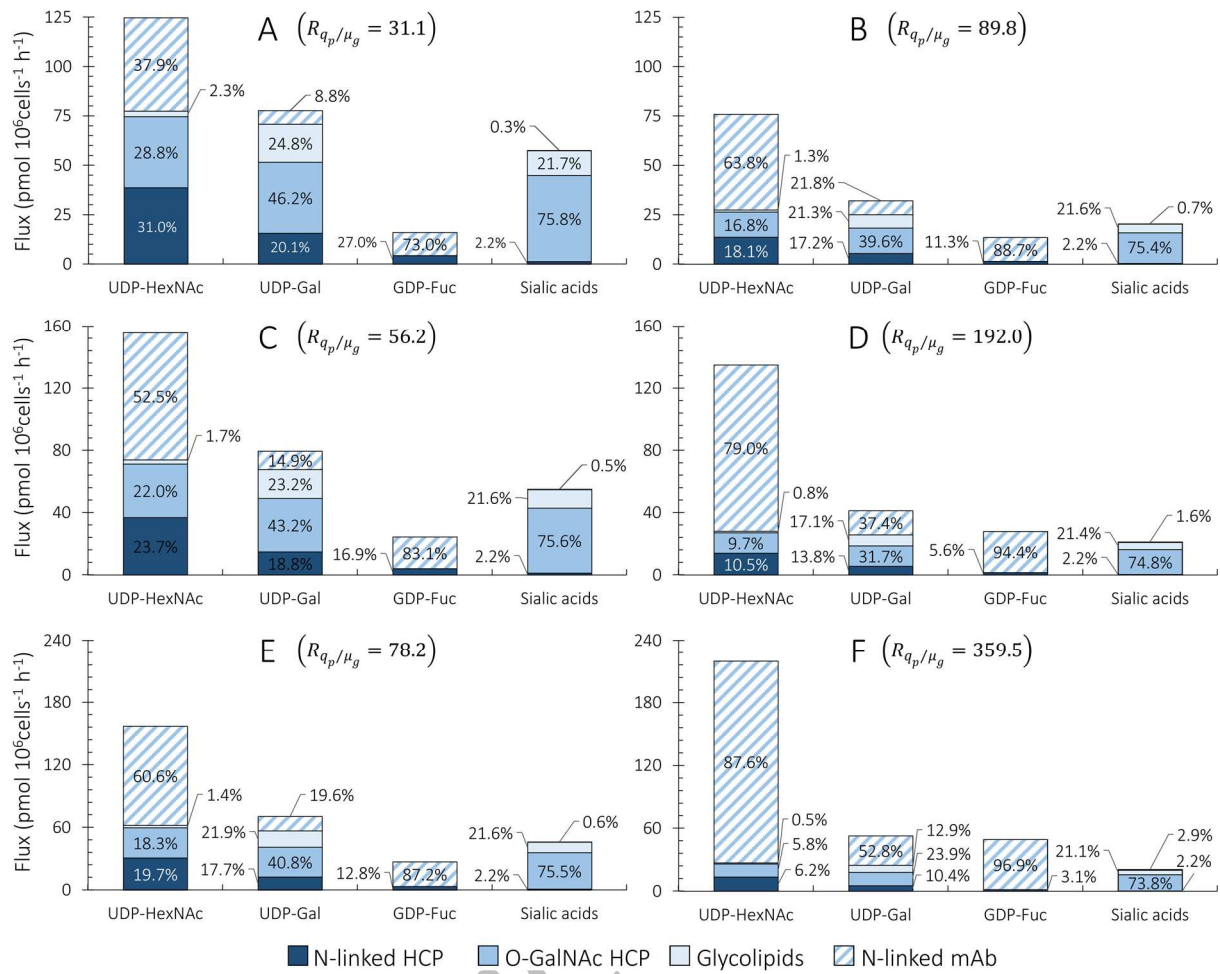
**Figure 3. Relative abundance and amino acid composition of CHO host cell proteins**

The relative abundance of the major categories for CHO HCPs are compared with those obtained from murine ovary cells<sup>49,50</sup> (A). The amino acid composition of murine<sup>38</sup> and CHO<sup>51</sup> host cell proteins used in previous studies is compared with the one estimated from the CHO proteome (B).



**Figure 4. Major N-linked and O-GalNAc glycoproteins in CHO**

The ten proteins that contribute most to total N-linked (A) and O-GalNAc (B) glycosites are shown. The labels above the bars show the percent contribution of each individual glycoprotein to the total number of N-linked and O-GalNAc glycosylation sites. The name of each glycoprotein is shown on the right.



**Figure 5. Distribution of nucleotide sugar fluxes towards cellular and mAb glycosylation**

Theoretical distributions of NS fluxes towards N-linked HCP, O-GalNAc HCP, glycosphingolipids, and mAb N-linked glycosylation at low (A, C, and E) and high (B, D, and F)  $R_{q_p/\mu_g}$  values. The distributions were estimated by multiplying the stoichiometric coefficient for each NS (Table 2), by values of  $q_p$  and  $\mu_g$  reported for industrial cell lines (Table 1)<sup>24-26</sup>. The different datasets that are presented are Dool-L (A), Dool-H (B), Chus-L (C), Chus-H (D), Kant-L (E) and Kant-H (F). The percentage of total NS flux (axis) towards N-linked HCP (dark blue), O-GalNAc HCP (intermediate blue), lipid (light blue), and N-linked mAb (belted pattern) glycosylation is presented within the figure.