Precision genome editing: A small revolution for glycobiology

Catharina Steentoft, Eric P Bennett, Katrine T-B G Schjoldager, Sergey Y Vakhrushev, Hans H Wandall, and Henrik Clausen

Glycobiology vol. 24 no. 8 pp. 663–680, 2014
doi:10.1093/glycob/cwu046
Advance Access publication on May 26, 2014

Introduction

A major challenge for the field of glycobiology remains to translate the world of the glycome, with its diverse and heterogeneous glycan structures and complex biosynthetic pathways, into more predictable and defined networks of glycogenes, glycoenzymes, glycostructures and ultimately into defined molecular mechanisms for biological functions. The general biosynthetic pathways and structure-based functional interactions of glycans are currently fairly well understood. However, further levels of detailed insight, especially into regulation and function of site-specific glycosylation of proteins, are required to enable meaningful integrative analysis of glycoproteomes together with other Omics (genome, transcriptome, proteome, phosphoproteome, etc.) for “polyomic”-wide characterization and discovery programs. The key question of what roles, if any, a particular sugar structure at a particular site in a given protein, may play is generally not possible to address without extensive and laborious experimentation. The major reasons for this are clearly technical difficulties with characterization of glycan structures at specific sites on a proteome-wide level, as well as the complexity and heterogeneity of glycan structures at individual glycosylation sites. The inherent obstacle of site-specific glycan characterization is rapidly becoming addressable by improved mass spectrometry (MS) fragmentation strategies, particularly utilizing electron transfer dissociation (ETD) technology to enable sequencing of the protein backbone with minimal loss or fragmentation of the glycan entity (Syka et al. 2004; Hogan et al. 2005). The second obstacle of glycan heterogeneity has now become at least partly addressable through nuclease-based precision genome editing tools that enable fast generation of cell systems with simplified homogeneous glycosylation (Steentoft et al. 2011) with and without specific glycogenes (Schjoldager et al. 2012).

The large number of glycoengineered rodent models produced with systemic or organ-specific knockout or knockin of glycogenes has clearly provided important insights into the biology of glycans and distinct glycogenes through introduction of phenotypic characteristics often mirroring human diseases. Arguably, however, these models have provided less insight into specific molecular mechanisms of functions and discovery of druggable targets and pathways (Lowe and Marth 2003). In this respect, lower organisms such as Caenorhabditis elegans, Drosophila melanogaster, Danio rerio, and Xenopus laevis have more often provided model systems where distinct structure–function relationships and molecular mechanisms for glycans have been dissected (Haltiwanger and Lowe 2004; Lehle et al. 2006; Ten Hagen et al. 2009). This is presumably due to the methods available in these organisms for efficient genetic engineering and in particular efficient genome-wide and high-throughput nucleases.

Keywords: GALNT / glycoengineering / isogenic cell system / SimpleCell
knockdown and knockout strategies including RNA silencing, morpholinos, and transposons, but also because the glycome and glycogenome of these organisms are often simpler. Great advances have been made through the early work with mutated mammalian cell lines enabling dissection of individual glycosylation pathways as well determination of functional roles of glycosylation (reviewed in Maeda et al. 2006; Patnaik and Stanley 2006; Zhang et al. 2006). Nevertheless, it is still apparent that there has long been a pressing need for even simpler biological systems in which designed, fast, efficient, and stable genome-wide editing is possible. The nuclease-based editing technologies enable genome editing of cells at a level comparable to what has been available for bacteria and yeast for decades, and it is unquestionable that fast and precise gene editing have revolutionized microbiology including yeast biology and biotechnology.

One may ask why RNA silencing strategies have generally failed to advance the glycobiology field. RNA silencing strategies are fast, inexpensive and high-throughput methods allowing for genome-wide screening studies (reviewed by (Dorsett and Tuschi 2004)). However, RNAi-based silencing generally only provides reduction in transcript and protein levels, especially in mammalian cells, and this may not be sufficient for most glycogenes, as these encode highly efficient enzymes that generally do not appear to be rate-limiting in the glycosylation process. Thus even 80–90% reduction in enzyme protein may not lead to clear and discernible effects on glycosylation. Moreover, these silencing strategies often produce off-target effects as well as cellular toxicity (reviewed in Aagaard and Rossi 2007; Jackson and Linsley 2010) that are especially difficult to evaluate and qualify for glycogenes because of our limited understanding of the site-specific regulation and function of protein glycosylation.

Nuclease-based gene editing is now opening the field for development of isogenic cell models that enable detailed dissection of the non-redundant functions of individual glycosyltransferases and discovery of important biological functions (Schjoldager and Clausen 2012). More generally, the genome editing tools discussed in this review enable complete, rapid and stable knockout (and other editing possibilities) of individual or multiple glycogenes in cell lines as well as whole animals, and applications of these new tools are poised to change the glycobiology field.

Nuclease-based genome editing tools

Targeted genome editing in mammalian cells has long been hampered by the lack of efficient targeting strategies. Homologs recombination (HR) enables targeted gene integration as well as disruption. However, most mammalian cells, except some embryonic stem cells, have very low frequency of HR making biallelic (or in some cancer cell lines tri- or tetraallelic) disruption very time-consuming and laborious, often requiring screening of thousands of clones as well as extensive drug selection (Yamane-Ohnuki et al. 2004). The discovery of site-specific endonucleases that introduce a double-stranded break (DSB) and thereby stimulate HR at a specific locus has opened a new era in genome engineering (Rouet et al. 1994; Bibikova et al. 2003). This field is likely one of the most rapidly advancing areas in cell biology today. New applications and improvements to the technologies are appearing weekly and a review can merely present a snapshot of the current state of the technology (Carroll 2014). The precision genome editing strategies will clearly have tremendous impact on many areas of cell biology and medicine as, e.g., recognized by the selection for Method of the Year 2011 by Nature Methods (Editorial 2011). Today four main platforms for precision genome editing exist: the protein guided homing endonucleases; zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN) and the RNA-guided clustered regularly interspaced short palindromic repeat (CRISPR)-associated (CRISPR/Cas) nuclease system (Figure 1). All four platforms are based on modified, often chimeric, nucleases optimized to generate site-specific DSB or nicks at high activity and specificity thereby opening up for a large variety of gene editing possibilities including gene disruption, deletion, insertion, correction, mutation and tagging (Figure 2). Modifying or replacing the nuclease domain with domains having a different functionality furthermore enable control of gene expression (reviewed in Mussolino and Cathomen 2012; Mali, Esvelt, et al. 2013). A number of excellent reviews on ongoing technological developments are available (Silva et al. 2011; Stoddard 2011; Joung and Sander 2012; Mussolino and Cathomen 2012; Mali, Esvelt, et al. 2013; Ran, Hsu, Wright et al. 2013; Segal and Meckler 2013; Carroll 2014; Sander and Joung 2014) and in the following we briefly present the principles of the four editing techniques, and discuss their current and future applications with a particular focus on their use in the glycobiology field.

Homing endonucleases

Homing endonucleases were the first nuclease-based precision genome editing tools to be discovered and were used for, among other things, the initial finding of DSB stimulating HR (Rouet et al. 1994). Homing endonucleases are “selfish” mobile genetic elements divided into five families, of which members of the LAGLIDADG family, often referred to as meganucleases, are the nucleases most frequently used for gene targeting (reviewed in Silva et al. 2011; Stoddard 2011; Hafiz and Hausner 2012). Meganucleases have been exploited for gene targeting for almost two decades. While early studies relied on the endogenous targeting sequence, a lot of effort is continuously being put into understanding the structure, and thereby DNA recognition mechanism, of meganucleases in order to create engineered nucleases targeting a desired sequence (Edgell 2014; Takeuchi et al. 2014) (Figure 1). Meganucleases generally have a very high specificity, but the design challenges limit their application and until now only a relative small number of human genes has been targeted by meganucleases (Segal and Meckler 2013); there is to the authors’ knowledge only one example of a meganuclease targeting a glycogene, FUT8 (Daboussi et al. 2012).

ZFNs

ZFNs are unnatural, chimeric proteins consisting of a zinc finger (ZF) DNA-binding domain and a non-specific DNA nuclease, this most often being the catalytic domain of the FokI restriction enzyme connected by a short linker (Figure 1) (Kim et al. 1996; Smith et al. 1999, 2000). The DNA-binding domain of a ZFN is made up of 3–6 ZFs, each recognizing a DNA triplet, and by combining multiple ZF domains with known binding specificities one can rather easily assemble an
array of ZFs with high binding specificity for different gene targets. The active form of FokI is a dimer, and ZFNs are therefore introduced in pairs with the target site consisting of two binding sequences between 9 and 18 bp in length, depending on the number of ZF domains, and a 5–7 bp spacer region in between (Händel et al. 2009; Shimizu et al. 2009; Wilson et al. 2013). Currently, three major assembly platforms exist for ZFN generation. A proprietary method developed by SangamoBioSciences is custom designed, validated and commercially available from Sigma-Aldrich under the brand name CompoZr®. A simpler (and less costly) strategy is modular assembly, in which individual ZF modules do not behave independent of each other and some context-dependent effects from neighboring modules have been observed, often resulting in low success rates (Ramirez et al. 2008). Alternatively, ZF arrays can be assembled by the publicly available oligomerized pool engineering (OPEN) method (Maeder et al. 2008, 2009) or the more recently developed and somewhat simpler context-dependent assembly approach (Sander et al. 2011), strategies that both take context-dependent effects into account. ZFNs have been applied to target an impressive number of genes in a broad selection of species; however, despite continuously improved
the DNA strands are modified by replacing the nuclease with a nickase HR repair can be stimulated. In NHEJ, the most frequent repair mechanism for DSB in mammalian cells; however, the nuclease with a nickase HR repair can be stimulated. In NHEJ, the DNA strands are modified to facilitate direct end-joining of the broken strands. Since this is a non-template-driven repair mechanism, NHEJ is error-prone, often introducing small insertions or deletions and thereby frame shift mutations. NHEJ-mediated gene knockout is the most frequent application of precision genome editing to date (Segal and Meckler 2013). By applying two nuclease platforms (Wilson et al. 2013), the design and selection of ZFNs with high cutting efficiency and specificity remains a challenge.

**TALENs**

Transcription activator-like effectors (TALE) were originally discovered in plant pathogenic bacteria (genus *Xanthomonas*) that inject TALEs into plant cells, where they bind specifically to plant DNA promoter sequences and activate host genes required for pathogen infection. TALEs consist of an N-terminal domain that controls translocation, a central DNA-binding repeat domain, and a C-terminal transcription activator domain (Figure 1). The individual repeated domains in TALEs recognize the nucleotides in a 1:1 ratio. Each repeated domain consists of 34 amino acid (aa), where the aa located at positions 12 and 13, the repeat-variable di-residues (RVDs), determine the DNA-binding specificity (Boch et al. 2009; Moscou and Bogdanove 2009). TALEs have an impressive target versatility, with the only restriction being that the base 5′ of the first repeat is a thymine; however, studies now suggest that even this limitation can be circumvented (Lamb et al. 2013; Tsuji et al. 2013). Transcription activator-like effector nucleases (TALENs) are similar to ZFNs in using the FokI nuclease to make DSB and various TALEN scaffolds consist of segments of the N- and C-terminal domain and usually between 15.5 and 19.5 RVDs, with the last module only containing 20 aa therefore designated as a half repeat (Christian et al. 2010; Miller et al. 2011; Mussolino et al. 2011; Bedell et al. 2012). The optimal length of the spacer between the two binding sites is highly dependent on the scaffold used, but it is longer than that for ZFNs, often ranging from 10 to 30 bp (Bogdanove and Voytas 2011). One major challenge when constructing a TALEN is the need to clone and assemble multiple almost identical repeats; however, several platforms are now available that enable rapid TALEN array assembly (Cermak et al. 2011; Briggs et al. 2012; Reyon et al. 2012; Schmid-Burgk et al. 2013). Custom designed and validated TALENs are currently also offered commercially through several companies (Life Technologies, Cellectis).

**CRISPR/Cas**

The most recent addition to precision genome editing tools is the RNA-guided CRISPR/Cas system, an adaptive natural defence mechanism present in bacteria and archaea, acting against invading genetic elements. The type II CRISPR/Cas mechanism is a three-component system where short CRISPR RNA (crRNA) in complex with the trans-activating crRNA (tracrRNA) and the Cas9 nuclease hybridize to a 20-bp target sequence resulting in cleavage and thus disruption of the foreign DNA (reviewed in Makarova et al. 2011). Jinek et al. (2012) demonstrated that the Cas9 enzyme, together with a dual-tracrRNA:crRNA RNA chimera, named single-guided RNA (sgRNA), are able to generate site-specific DSBs in vitro (Figure 1). Not long after CRISPR/Cas-based genome editing was accomplished in mammalian cells (Cho et al. 2013; Cong et al. 2013; Jinek et al. 2013; Mali, Yang, et al. 2013) by introducing a codon optimized Cas9 nuclease together with the RNA components resulting in targeting efficiencies similar or even higher than those reported for ZFNs and TALENs. Compared with ZFNs and TALENs, no protein engineering is required and only a single sgRNA needs to be designed for each new target, although design and test of at least two sgRNA per target locus is recommended to ensure selection of one with adequate activity (Ran, Hsu, Wright, et al. 2013). Another advantage of the CRISPR/Cas system is that it can be used for multiplexing by simultaneous delivery of multiple sgRNAs targeting different loci (Cong et al. 2013; Wang, Yang, et al. 2013), which opens up the possibility for genome-wide screenings (Shalem et al. 2013; Wang, Wei, et al. 2013; Koike-Yusa et al. 2014). The only target site restriction of the CRISPR/Cas system is the presence of a proto-spacer-adjacent motif (PAM) located immediately downstream of the 20 bp target. The Cas9 derived from *Streptococcus pyogenes* requires a 5′-NGG PAM
that is on average present every 8–12 bp (Cong et al. 2013; Hsu et al. 2013), thus not presenting a substantial constraint. To add further flexibility to the system, Cas9 proteins with different PAM requirements can be derived from alternative species and applied for genome editing (Esvelt et al. 2013; Hou et al. 2013). A number of attempts have been made to further improve the CRISPR/Cas system focusing on increasing activity and in particular specificity in order to decrease off-target effects, as will be discussed in detail in a later section. In summary, the activity and specificity of CRISPR/Cas-based gene targeting seem to be highly dependent on the selected Cas9, the PAM, the sgRNA architecture, the sgRNA recognition sequence, the cell to be targeted, as well as the Cas9 and sgRNA expression concentration and duration (Fu et al. 2013; Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013; Sander and Joung 2014).

**Efficiency and specificity of targeting**

An optimal genome editing tool should have a very high activity and at the same time exhibit absolute gene target specificity. A very large proportion of currently available publications in the field focus on development of improved methodological strategies, reflecting that although the techniques provide an enormous leap forward, there are still technical challenges. Here, the main methodological challenges associated with nuclease-based genome editing in mammalian cell lines are addressed including: (i) delivery and expression, (ii) enrichment, (iii) screening and (iv) potential off-target effects.

**Efficient delivery and expression**

Nuclease delivery and expression is of great importance for activity as well as specificity, as high level but short-term expression can result in increased activity and at the same time lower risk for off-target effects (Pattanayak et al. 2011; Hsu et al. 2013). The optimal choice of delivery system is highly cell-type dependent. The nuclease and sgRNA are most commonly delivered as plasmid DNA or in vitro transcribed mRNA by electroporation, lipid-based transfection or alternatively by different viral vectors. ZFNs and TALENs can furthermore be delivered as purified protein either by exploiting the cell penetrating ability of the ZFN protein (Gaj et al. 2012), uptake mediated by linkage to transferrin (Chen et al. 2013) or cell penetrating peptides (Liu et al. 2014). Protein delivery, although more laborious, minimizes off-target effects as the protein is degraded after <4 h compared with 16–72 h expression with plasmid DNA delivery (Gaj et al. 2012; Chen et al. 2013). Once delivered, nuclease efficiency can be increased by culturing the cells at transient hypothermic conditions (30°C) (Doyon et al. 2010; Gaj et al. 2012). Alternatively, addition of proteasome inhibitors (Ramakrishna et al. 2013), cytostatic drugs (Rahman et al. 2012) or histone deacetylases (Pelascini et al. 2013) has been shown to improve ZFN activity.

**Enrichment for targeted cells**

Nuclease-driven genome-targeting activities ranging from one to almost 100% have been reported. As many mammalian cell lines have low transfection efficiency, nuclease expression or HR frequency, nuclease activity is commonly found in the low range. Furthermore, if complete knockout targeting of two or even multiple alleles is desired, as is often the case in cancer cells, the success rate will be even lower, thereby requiring screening of hundreds of clones. A number of strategies have been developed to enrich and select for targeted cells (Table I). Enrichment strategies include addition of drug selection cassettes that can later be removed by the Cre/lox system (Hockemeyer et al. 2009) or the piggyBac transposon (Yusa et al. 2011; Sun and Zhao 2013); however, these strategies often favor monoallelic insertions and suffer from low insertion frequencies compared with NHEJ-mediated targeting. Other strategies involve co-expression of reporter constructs followed by fluorescence-activated cell sorting (FACS) (Kim et al. 2011; Ramakrishna et al. 2014) or magnetic bead isolation (Kim et al. 2013), both having the disadvantage that a new reporter construct needs to be made for each new target site. Elegant enrichment systems have been created for CRISPR/Cas9, ZFNs ((Duda et al. 2014) now available through Sigma–Aldrich) and TALENs (Feng et al. 2014) with nuclease expression plasmids also encoding fluorescent proteins, allowing for FACS enrichment. Finally, in some cases the target gene may allow for a specific phenotypic enrichment strategy, as has been reported in studies targeting glycogenes using lectins either to kill or deselect untargeted cells (Malphettes et al. 2010; Wong et al. 2010; Hauschild et al. 2011).

**Screening for targeted cells**

Although conceptually simple, screening and selection of targeted cells represent technically challenging problems with respect to two key procedures: (i) identifying targeted cells in a mixture of clones and (ii) distinguishing the number and type of allelic modifications in each targeted clone. Our preferred screening methods are illustrated in Figure 3. The most commonly applied method to screen the nuclease transfected cell pool is to use a hybridization and endonuclease-based assay, where a PCR product is amplified around the target sequence, denatured and hybridized to allow for potential heteroduplex formation between mutated and wild-type strands. The mismatched nucleotides are recognized and cleaved by various endonucleases (T7EI (nuclease I), CEL I, Surveyor), and mutation efficiency can be estimated by gel analysis comparing digested and undigested products (Guschin et al. 2010) (Figure 3A). Isolated single cell colonies can be screened by DNA sequencing either by sub-cloning and Sanger sequencing or direct deep-sequencing; the first being unsuited for high-throughput analysis, and the second rather costly, making this unachievable for the hundreds of clones sometimes needed to obtain a clone with multi-allelic targeting events. An alternative indel screening procedure is based on DNA fragment analysis using target-specific fluorophore-tagged primers combined with capillary electrophoresis detection (Foley et al. 2009; Pratt et al. 2012). In order to increase the flexibility of the fragment analysis approach and thus make it applicable to high-throughput indel screening, we devised a single tube, tri-primer amplification protocol by which fluorescent amplicons are easily generated using a generic fluorescently labeled primer and can thereby demonstrate amplicon discrimination power down to 1 bp differences (Zhang et al. unpublished).
Many studies have relied on in silico off-target predictions, based on sequence similarity to the target site, followed by PCR amplification and endonuclease assay and/or sequencing of the most likely off-target sites (Perez et al. 2008). These methods are obviously biased, do not take context dependent effects into account and are often not sensitive enough to detect low-frequency off-target events. An unbiased approach using integrase-defective lentiviral vectors trapped within DSBs through NHEJ (Gabriel et al. 2011) enabled identification of ZFN-mediated cleavage events, thereby demonstrating that off-targets sites have >66.8% sequence identity to the target site and that mismatch-tolerance is position dependent. The authors additionally reported that while the sequences of the in vivo cleaved sites matched biochemical studies, ranking the target sites with respect to frequency could not be achieved predictively in silico. A similar integrase-defective lentiviral approach was taken for one TALEN pair targeting COL7A1, a gene-therapy candidate, identifying three off-target sites although these were located far from (>10 kb) a coding region (Osborn et al. 2013). Other studies have used whole-exome sequencing, which is now becoming more affordable, to exclude off-target events in coding regions (Ding et al. 2013; Ousterout et al. 2013; Cho et al. 2014), although the high

(Figure 3B). Finally, for some target genes functional screening assays can be conducted as reported for FUT8 knockout cells lacking LCA lectin staining (Wong et al. 2010), COSMC knockout cells expressing truncated Tn O-glycans detectable with either lectins or antibodies (Steentoft et al. 2011; Steentoft, Bennett, et al. 2013) (Figure 3C), and glutamine synthetase knockout cells that can be identified by growing duplicate plates in media lacking glutamine (Fan et al. 2012). Functional screening assays have the advantage of enabling direct identification of multi-allelic targeted cells even in a mixed population.

### Off-target effects

A general concern for nuclease-based genome targeting is off-target effects that may affect interpretation of gene targeting studies and more importantly pose potential safety issues for therapeutic applications. While the activity of a particular editing tool can be studied with a range of methods as listed in Table I, off-target effects are difficult to address in a non-biased fashion. Many studies have relied on in silico off-target predictions, based

### Table I. Methods to enrich and screen for targeted clones

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Pros/cons</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clonal enrichment</td>
<td>+ Effective biallelic enrichment</td>
<td>Duda et al. (2014), Feng et al. (2014)</td>
</tr>
<tr>
<td>Fluorescent nuclease</td>
<td>+ Moving nuclease into colored construct, FACS sorting</td>
<td></td>
</tr>
<tr>
<td>Drug selection</td>
<td>+ Effective for HR events</td>
<td>Hockemeyer et al. (2009), Yusa et al. (2011), Sun and Zhao (2013)</td>
</tr>
<tr>
<td>Surrogate reporter, FACS sorting</td>
<td>+ Effective biallelic enrichment</td>
<td>Kim et al. (2011), Ramakrishna et al. (2014)</td>
</tr>
<tr>
<td>− Individual reporter construct required for each new target site, FACS sorting required</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic bead isolation</td>
<td>+ Effective biallelic enrichment, no FACS sorting</td>
<td>Kim et al. (2013)</td>
</tr>
<tr>
<td>− Individual reporter construct required for each target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transposon co-selection</td>
<td>+ Effective mono- and biallelic enrichment</td>
<td>Carlson et al. (2012)</td>
</tr>
<tr>
<td>− Integration of selection-transposon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface phenotype enrichment</td>
<td>+ Powerful enrichment for biallelic targeting</td>
<td>Malphettes et al. (2010), Wong et al. (2010), Hauschild et al. (2011)</td>
</tr>
<tr>
<td>− Requires known surface phenotype upon targeting and antibody/lectin towards the phenotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening</td>
<td>+ Easy, works on pools as well as single clones</td>
<td>Malphettes et al. (2010), Steentoft et al. (2011)</td>
</tr>
<tr>
<td>− Requires known surface phenotype upon targeting and antibody/lectin towards the phenotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenotypic screen</td>
<td>+ Easy, works on pools as well as single clones</td>
<td>Miller et al. (2007), Santiago et al. (2008), Guschin et al. (2010)</td>
</tr>
<tr>
<td>− Requires access to sequenator</td>
<td>Perez et al. (2008)</td>
<td></td>
</tr>
<tr>
<td>Heteroduplex DNA digestion</td>
<td>+ Easy, ideal for detecting mutations in pools of cells</td>
<td>Miller et al. (2007), Santiago et al. (2008)</td>
</tr>
<tr>
<td>− Insensitive and low resolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragments analysis</td>
<td>+ Easy, works on pools and single clones, semi-quantitative</td>
<td>Foley et al. (2009), Pratt et al. (2012), Zhang et al. (unpublished)</td>
</tr>
<tr>
<td>− Requires access to sequenator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep sequencing</td>
<td>+ Identifies exact sequence</td>
<td>Perez et al. (2008)</td>
</tr>
<tr>
<td>− Expensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-cloning and Sanger sequencing</td>
<td>+ Identifies exact sequence, relatively cheap</td>
<td>Miller et al. (2007), Santiago et al. (2008)</td>
</tr>
<tr>
<td>− Difficult for high throughput</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss of restriction site</td>
<td>+ Easy, fast and cheap screening</td>
<td>Cost et al. (2010), Sakuma et al. (2013)</td>
</tr>
<tr>
<td>− Requires restriction site in the target sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertion of restriction site</td>
<td>+ Easy, fast and cheap screening</td>
<td>Chen et al. (2011)</td>
</tr>
<tr>
<td>− Insertion frequency for HR &lt; NHEJ</td>
<td>Moehle et al. (2007)</td>
<td></td>
</tr>
<tr>
<td>Insertion of fluorescent marker</td>
<td>+ Enables FACS enrichment</td>
<td></td>
</tr>
<tr>
<td>− Cannot distinguish mono- vs. biallelic, insertion frequency for HR &lt; NHEJ</td>
<td>Hisano et al. (2013)</td>
<td></td>
</tr>
<tr>
<td>lacZ blue white screening</td>
<td>+ Allow for direct identification of out-of-frame indels</td>
<td></td>
</tr>
<tr>
<td>− Unsuited for high throughput</td>
<td>Dahlem et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>High-resolution melt curve analysis</td>
<td>+ Simple, works on pools as well as single clones</td>
<td></td>
</tr>
<tr>
<td>− Does not distinguish between indels and SNVs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

C Steentoft et al.
false-negative rate associated with whole-exome sequencing makes it difficult to rule out missed off-target events (Cho et al. 2014). Comprehensive in vitro studies, testing nuclease activity on large libraries containing $10^{11}$–$10^{12}$ DNA sequences, followed by validation of the potential off-target sequences in cell cultures have been presented for ZFNs (Pattanayak et al. 2011), TALENs (Guilinger et al. 2014) and CRISPR/Cas (Pattanayak et al. 2013). Thirty-four potential off-target sequences were identified for one ZFN pair targeting CCR5 and nine of them had actual deletions or insertions characteristic for NHEJ, although the frequency was lower (<10%) than for the on-target site (43%) (Pattanayak et al. 2011). Compared with this 45 potential and 9 actual off-target sites were identified for TALENs-targeting CCR5 with a cleavage frequency of 0.03–2.3% efficiency (Guilinger et al. 2014). However, directly comparing off-target effects of the different editing tools can be difficult as the selected nuclease design, including target sequence length, spacer length, expression construct, etc., has a large influence on the activity. This being said, it is easier to place target sites in unique sequences for TALENs compared with ZFNs because target site design and selection is less restricted for TALENs. One study (Mussolino et al. 2011) comparing one ZFN and different TALENs targeting CCR5 showed that while similar on-target activities were observed, TALENs had lower cytotoxicity and off-target effect in the highly homologous CCR2.

Several studies addressing CRISPR/Cas off-targets (Cradick et al. 2013; Fu et al. 2013; Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013) suggest that on- and off-target rates are variable and difficult to predict and while impressive on-target activity can be achieved, off-targets with similar frequency are in some cases identified (reviewed in Sander and Joung 2014).

Considerable effort has been put into methods and designs that can decrease off-target activity. One major improvement for specificity was the development of a modified FokI domain that only functions upon heterodimerization, thus preventing off-targets due to homodimerization (Miller et al. 2007). Other attempts include varying the number of ZF modules (Shimizu et al. 2011; Yan et al. 2013) and optimizing the length of the linker between the ZF domain and the FokI domain (Händel et al. 2009; Shimizu et al. 2009). For TALENs it has been shown that mismatches in the 5’ end of the target are more disruptive than mismatches in the 3’ end, information, which should be taken into account when designing TALENs (Meckler et al. 2013). The expression level of the editing tool has high impact on off-target events and a recent study determined the dose dependent activity for ZFN off-targeting (Duda et al. 2014) and prolonged exposure to moderate levels has been shown only to result in marginal off-target effects (Wang, Wei, et al. 2013). Alternatively, destabilized ZFNs and small molecules have been used to regulate ZFN protein levels (Pruett-Miller et al. 2009) and the FokI domain has been replaced with a sequence-specific nuclease, e.g., a meganuclease, in order to improve specificity (Kleinstiver et al. 2012; Schierling et al. 2012; Boissel et al. 2013). The competition from the more frequent NHEJ, on- and off-target, is a challenge when the goal is to obtain HR-directed insertions. One possible solution is to replace the nuclease domain with a nickase domain, thereby creating only a single-stranded break or a “nick”. This has been shown to significantly reduce the NHEJ frequency at on- and off-target sites, although it also affects the HR efficiency negatively. Nickases have been developed for meganucleaseas (Metzger et al. 2011), ZFNs (Kim et al. 2012; Ramirez et al. 2012) and TALENs (Beurdeley et al. 2013), as well as for CRISPR/Cas9 (Jinek et al. 2012; Cong et al. 2013; Mali, Yang, et al. 2013). Additional strategies to improve CRISPR/Cas9 specificity includes a dual nickase method (Mali, Aach, et al. 2013; Ran, Hsu, Lin, et al. 2013; Cho et al. 2014) and modified sgRNA architecture (Pattanayak et al. 2013; Cho et al. 2014; Fu et al. 2014).
Precision genome engineering in glycobiology

The nuclease-based genome editing tools have potential to revolutionize the glycobiology field and be a game changer in molecular dissection of specific biological functions of glycosylation. The technology allows us to edit the genome of model organisms faster and in a much broader selection of species as exemplified by the generation of pigs with knockout of the α1,3galactosyltransferase (GGT A1) to eliminate the major xenotransplantation barrier (Hauschild et al. 2011; Li et al. 2012). We can edit endogenous genes of virtually any cell line to carry disease-causing mutations that have been identified in human congenital disorders of glycosylation (CDGs). This will enable development of cell systems as well as animal models with defined disease-causing genotypes such that phenotypes caused by not only complete null genotypes but also more subtle hypomorphic genotypes, can be analyzed. Investigators outside the glycobiology field can use targeted gene editing to probe the role of glycosylation genes with traditional cell biology methods, and dissect specific structure–function relationships without necessarily performing actual analysis of carbohydrate structures. With increasing knowledge of the enzymatic functions of individual glycosyltransferases in the biosynthesis of the diverse glycome (Figure 4), we will enter a phase where the glycobiology field and broaden interest and involvement to all fields of cell biology and medicine. The possibilities seem endless, and a vision of breaking the major technical barrier for the field of glycobiology seems achievable.

Precedence for the value and impact of mutant cell lines with defined glycosylation capacities for wider use even outside the glycobiology field may be found in early work with Chinese Hamster Ovary (CHO) mutants. One highly successful approach to generate CHO glycosylation mutants have been to select for mutated cells resistant to various lectins as reported as early as in 1973 (Wright 1973). The first biochemical characterizations of CHO lectin mutants identified deficiencies in the Mgtal gene encoding the GNT-1 enzyme that controls the first step in generating complex type N-glycosylation (Gottlieb et al. 1975; Stanley et al. 1975). Further pioneering work on development of panels of CHO mutants (reviewed in Maeda et al. 2006; Patnaik and Stanley 2006; Zhang et al. 2006) with different capacities for glycosylation of proteins and lipids has been used to define glycosylation pathways, to identify novel glycosyltransferase genes, to dissect biological functions of glycans as well as individual glycan sites and to engineer and analyse glycosylation of recombinantly expressed glycoproteins. These efforts were laborious, time-consuming and dependent on substantial discovery efforts to define the genetic basis underlying each selected change in glycosylation. A classical example of “non-targeted” glycoengineering of CHO cells is the CHO lDlD line selected for loss of surface expression of the low-density lipoprotein receptor (LDLR), which “fortuitously” carried a defect in the C4-Glc/GlcNAc epimerase that results...
in loss of synthesis of UDP-Gal/GalNAc and capacity for O-glycosylation (Kingsley and Krieger 1984; Kingsley et al. 1986). This cell system enabled the demonstration that stable surface expression of LDLR requires O-glycosylation in the stem region to prevent proteolytic release of the ectodomain. The CHO ldlD cells have been widely used to evaluate the function of GalNAc-type O-glycosylation, because the deficiency is partly reversible upon exogenous addition of Gal and GalNAc (Kozarsky et al. 1988; Kato et al. 2006). The past history with these CHO mutants thus clearly demonstrates how access to model systems with defined types of glycosylation can be widely used also outside the glycobiology field. Precise genome editing is still in its infancy, and in the following section current applications reported are discussed and future perspectives for the glycobiology field are provided.

Whole organisms
Phenotypic information from transgenic mice is largely based on complete null mutants, while a majority of the discovered CDGs are caused by hypomorphic missense mutations (reviewed in Freeze 2013), and there is generally poor overlap in understanding between phenotypes and biochemistry (Lowe and Marth 2003). Precise gene editing now enables reproduction of mono- and biallelic mutations found in disease in both cell systems and animals of choice (Soldner et al. 2011; Wefers et al. 2013). The technology eliminates the need for embryonic stem cells as the nuclease can be delivered directly to one cell embryos (Geurts et al. 2009; Carbery et al. 2010), or to somatic cells followed by somatic cell nuclear transfer (SCNT) (Hauschil et al. 2011). Thus in principle, targeting can be performed in any mammalian species and genetic background. Introduced modifications are efficiently transmitted through the germ line, increasing the speed of animal model generation significantly as extensive backcrossing is avoided. However, to the authors’ knowledge, there are currently no reports on animal models for genes involved in CDGs.

One of the areas where ZFNs have been used for glycoengineering is xenotransplantation. With growing demand for organs for transplantation, pigs have long been considered as alternative donors. The α1,3Gal glycan epitope (Galili epitope) found on porcine glycoconjugates has been identified as the major xenotransplantation barrier and the cause of hyper-acute rejection. This epitope is synthesized by the α1,3galactosyltransferase encoded by the GGTAT gene, which is inactivated in man (for review, see Galili 2001). It has previously been shown that knockout of GGTAT by conventional strategies improves the success rate when transplanting organs from pigs to baboons (Kuwaki et al. 2005; Yamada et al. 2005), although other xenotransplantation antigens have emerged (Chen et al. 2005). Generation of biallelic knockout pigs by conventional strategies was laborious and required repeated transfection and selection steps (Phelps et al. 2003), but these efforts have now been reproduced using the ZFN technology combined with SCNT creating live offspring with biallelic GGTAT disruption (Hauschil et al. 2011; Li et al. 2012). While the application of nuclease-based genome editing in whole animals is still in its early stages, it is clear that these technologies now offer new opportunities to introduce, e.g. disease-causing hypomorphic alleles, use more comparable animal models, and include rescue controls.

Cell lines: isogenic cell systems
The ability to generate complete, stable and stackable glyco-gene knockout cell lines and produce isogenic cell pair models (parental cell line and subclones of which only one or few defined genes have been modified) will meet many of the major obstacles faced with molecular dissection of specific structure–function relationships in glycobiology: (i) complete knockout of a glyco-gene is often necessary to produce substantial changes in the glyco-phenotype; (ii) stable knockout ensures that sufficient cell material can be produced for analysis of glycans; (iii) many glycogenes exist as homologous families with partially overlapping functions and/or function in sequential order, and stacking multiple editing events may be necessary; and (iv) access to isogenic cell systems enables detailed comparative polyO/mics analyses in a relatively homogenous cell system.

The first example of ZFN-mediated knockout of a glyco-gene was reported by Zou et al. (2009) using the “OPEN” method to construct a ZFN pair targeting PIGA required for N-acetylgalcosaminyl-phosphatidylinositol synthesis, the first intermediate in GPI-anchor biosynthesis (Figure 4). PIGA is mutated in patients with paroxysmal nocturnal hemoglobinuria, and the authors targeted PIGA in hESC and iPSC cells creating valuable cell models for the disease.

A particularly impressive example of use of TALENs was presented in a study using the haploid cell line HAP1 to identify genes involved in α-dystroglycan modifications. Haploid cells, mutagenized by retroviral gene trapping, were exposed to a replication-competent vesicular stomatitis virus expressing the receptor glycoprotein of Lassa virus. The Lassa virus receptor glycoprotein targets a complex carbohydrate structure on α-dystroglycan and cells selected for resistance to Lassa virus could be studied. The authors were thereby able to obtain the same information in one experimental setup, that has been reported in studies over the last fifteen years on identification of glycosylation related genes and dystroglycanopathies, as well as to identify a couple of novel candidate genes. To confirm the identified candidate genes and their involvement in α-dystroglycan glycosylation, all genes were individually knocked out in the haploid cell system using TALEN gene editing (Jae et al. 2013). While these authors did not use precise gene editing for the actual screening and hence were restricted to the particular haploid cell line, very recent studies demonstrate how the CRISPR/Cas technology can be used for whole-genome screening using lentiviral delivery of sgRNA libraries (Shalem et al. 2013; Wang, Wei, et al. 2013; Koike-Yusa et al. 2014). The precision genome editing technologies, thus, bring an entirely new tool to glycobiology where it is possible to approach biological functions of carbohydrates by gene editing screens rather than traditional structural analysis of glycoconjugates.

Cell lines - The SimpleCell strategy
A critical obstacle to progress in the field of glycobiology is the complex and heterogeneous nature of glycan structures found on proteins (Figure 4). It is thus currently extremely difficult to both characterize all glycan structures expressed and assign all their glycosites in glycoproteins on a proteome-wide basis, primarily because the current MS-based analytical strategies impose: (i) a
need for enrichment strategies to avoid glycopeptide precursor signal suppression from abundant unglycosylated peptides and (ii) a need for glycan simplification with respect to glycan heterogeneity as well as individual glycan size and complexity. The latter is particularly troublesome due to the often insufficient peptide fragmentation observed for glycopeptides under the conditions of collision-induced dissociation (Hogan et al. 2005). The difficulties are multiplied for spectral assignment with peptides bearing extensive and complex glycosylation. An additional practical problem concerns high-throughput data processing, particularly where extensive glycan microheterogeneity is present: the rapid expansion of the search space when each possible glycan structure represents a unique modification to be accounted for in mass calculations. These obstacles have been partially solved for N-glycoproteomics by use of multiple lectin enrichment strategies of N-glycoproteins and enzymatic release of N-glycans that simultaneously enable labeling of the Asn attachment site (Zielinska et al. 2010). While this protocol enables proteome-wide analysis of N-glycosylation sites, information about glycan structures at each N-glycosylation site is lost. In contrast to this, there are no enzymes that can broadly release O-glycans and moreover, release of O-glycans from proteins cannot be performed with concomitant labeling of the aa attachment site (O’Neill 1996). However, Burlingame and colleagues (Vosseller et al. 2006; Chalkley et al. 2009) previously demonstrated efficient lectin enrichment and sequencing of O-GlcNAc glycopeptides in studies to characterize the O-GlcNAc modification catalyzed by the cytosolic O-GlcNAc-transferase, OGT (Haltiwanger et al. 1992). The authors took advantage of the simple and homogenous nature of the O-GlcNAc modification with no elongation to develop a lectin weak-affinity chromatography (LWAC) strategy for isolation of O-GlcNAc glycopeptides. Total proteolytic digests of murine brain postsynaptic density preparations were separated on long (>3 m) wheat germ agglutinin in lectin columns and GlcNAc eluted glycopeptides were subjected to LC–MS–MS analysis (Vosseller et al. 2006; Chalkley et al. 2009). Darula et al. (Darula and Medzihrdzsky 2009; Darula et al. 2010) further demonstrated that a similar enrichment strategy could be applied to O-GalNAc glycopeptides with use of exo-glycosidase treatment prior to lectin enrichment.

Inspired by these studies, we designed the SimpleCell strategy, which entails genetically engineering cells to produce a particular glycoproteome with homogenous glycan structures (Figure 5). The strategy enables global enrichment of glycoproteins/glycopeptides from cell lysates and secretomes by specific lectin chromatography, and rapid and sensitive identification of glycosylation sites by LC–MS–MS using higher-energy collision-induced dissociation and ETD fragmentation modes, greatly facilitated by the simple homogenous glycan structure. The SimpleCell strategy thereby sacrifices structural glycan information in order to obtain information on where the glycans are located.

The first SimpleCells were developed for analysis of the O-GalNAc-type glycoproteome using ZFN targeting to abrogate the normal elongation pathway of this type of O-glycosylation. This was achieved by knockout of the private chaperone, COSMC, for the core1 enzyme, C1GalT1, which is essential for O-GalNAc elongation in most cell types. Another predominant O-GalNAc elongation pathway, primarily expressed in the gastrointestinal tract, is core3 formed by β3GnT6; however, this gene is down-regulated in cancer and generally not expressed in cancer cell lines (Iwai et al. 2005). Lack of Cosmc therefore results in biosynthesis of the truncated cancer-associated O-glycan structures Tn (GalNAcα1-O-Ser/Thr) (Ju and Cummings 2002) and/or STn (NeuAco2-6GalNAcα1-O-Ser/Thr) depending on how efficient sialylation of Tn is in the targeted cells (Steentoft, Vakhrušev, et al. 2013). Since the sialic acid can be efficiently removed by sialidases, this strategy produces a homogenous O-glycoproteome with the simple GalNAc monosaccharide attached to Ser or Thr that can serve as a ligand for lectin enrichment as well as a tag for identification of glycosites. The SimpleCell strategy was first tested on total cell lysates...
from three human SimpleCell lines (K652, COLO-205 and Capan-1) (Steentoft et al. 2011; Steentoft, Bennett, et al. 2013; Vakhrushev et al. 2013). To validate whether blocking of O-GalNAc elongation changes site occupancy, we analyzed the glycoproteome of wild-type K562 cells, capturing glycopeptides carrying T-antigen (Galβ1-3GalNAcβ1-O-Ser/Thr) through LWAC enrichment. Comparing our finding with the SimpleCell data set a substantial overlap with the SimpleCell glycoproteome was observed suggesting that COSMC knockout does not alter site occupancy. More recently, we expanded the GalNAc SimpleCell strategy to a panel of 12 human cancer cell lines derived from different organs and identified in total >600 O-glycoproteins and almost 3000 O-glycosites (http://glycomics.ku.dk/o-glycan_database/) (Steentoft, Vakhrushev, et al. 2013).

We have additionally expanded the SimpleCell strategy to mining the O-Man glycoproteome (Vester-Christensen et al. 2013), which had been particularly elusive for decades. O-Man SimpleCells were engineered by ZFN targeting the POMGNT1 gene encoding the β2GlcNAc-transferase (PomGNT1) controlling the major elongation pathway of the initial O-Man glycan. Thus, in POMGNT1 knockout cells the O-Man glycosylation complexity is reduced to the initial mannose residue (Figure 5). An alternative elongation pathway by the β1,4GlcNAc-transferase (POMGNT2/GTDC2) has been identified, although currently only at one specific O-Man site in α-DG (Yoshida-Moriguchi et al. 2013). Adapting the LWAC workflow established for GalNAc-type SimpleCells using a ConA lectin column, we identified a large number of O-Man sites including evolutionarily conserved O-Man sites on 37 members of the cadherin superfamily (Vester-Christensen et al. 2013). Cadherins are the major receptors for cell–cell interactions, and the O-Man glycosites are located on beta-strands of the EC domains (2–5) involved in cis-interactions. In agreement with this finding, O-Man glycosylation was recently found to be essential for E-cadherin-mediated cell adhesion during blastocyst formation (Lommel et al. 2013).

Cell lines: deciphering isoform-specific functions

There are several examples in glyobiology where multiple iso-enzymes catalyze the same reaction with some redundancy, but at the same time slightly different specificity, making it difficult to decipher the role of each individual isoenzyme. As an example GalNAc-type O-glycosylation is one of the most complex regulated form of protein glycosylation with up to 20 distinct GalNAc-T isoforms controlling sites of O-glycosylation, and for decades considerable efforts have been devoted to identifying unique non-redundant functions for these isoforms (reviewed in Bennett et al. 2012). Most of these studies have used in vitro enzyme assays to characterize the substrate specificities of individual enzymes. Analysis of in vivo functions has also been performed with a number of murine knockout models, and in one example two O-glycoproteins serving as specific substrates for the GalNAc-T1 isoform were identified (Miwa et al. 2010). However, in general this approach has not resulted in deeper insight into the function of the GalNAc-T family. Several studies have used RNAi-based strategies to probe in vivo functions of GalNAc-Ts, and these, while valuable, rather exemplify some of the inherent problems associated with silencing of genes encoding glycosyltransferases. These include the silencing of GalNAc-T3 and T6, which have been shown to have near identical enzymatic characteristics and constitute the Ic subfamily (Bennett et al. 2012). Expression of both GalNAc-T3 and T6 has been suggested to promote cancer development, and silencing of one or both genes has been shown to inhibit oncogenic features (Park et al. 2010, 2011; Freire-de-lima et al. 2011; Taniuchi et al. 2011), although exact mechanisms bearing on site-specific glycosylation have been difficult to determine. One fascinating although puzzling finding was that silencing of GalNAc-T6 can dramatically alter the glycosylation and stability of the heavily glycosylated oncoprotein MUC1 (Park et al. 2010), opening up a number of additional questions, such as why this is not compensated for by GalNAc-T3. Nevertheless, knockdown of individual GalNAc-Ts has presented a number of intriguing phenotypes, but caution is needed in interpretation of these findings. Nuclease-based editing now offers additional tools to explore and validate past findings.

We have started exploring non-redundant functions of GalNAc-Ts using comparative analysis of isogenic SimpleCells with and without a single isoenzyme. In a first proof of concept, we used HepG2 SimpleCells to probe the specificity of GalNAc-T2 in a liver cell line and investigate a putative role for GalNAc-T2 in HDL and triglyceride metabolism (Schjoldager et al. 2010, 2011). The HepG2 SimpleCell model enables analysis of the O-glycoproteome and identification of O-glycosites as discussed above. With further knockout of GALNT2 the O-glycoproteomes of isogenic cells can be compared and non-redundant functions of this isoform determined. In parallel, we developed HepG2 isogenic cells with knockout of only GALNT2, which retain wild-type glycan elongation capacity; candidates selected from the differential O-glycoproteome could be validated and the role of elongated normal glycans studied as well. Using this approach, we identified and validated a number of O-glycoproteins exclusively glycosylated by GalNAc-T2, including ApoC-III and ANGPTL3, which both have roles in lipid metabolism. We could furthermore validate specific effects observed in GALNT2 knockout by reintroducing GALNT2 in the AAf/SI safe harbor locus using ZFN targeting (Schjoldager et al. 2012). In preliminary studies, isogenic HepG2 cells with knockout of GALNT1 have been analyzed and a different subset of unique O-glycoproteins controlled by this isoform identified (Schjoldager et al. in preparation).

One major finding in the “SimpleCell” O-glycoproteome strategy was the discovery of O-GalNAc glycosylation sites in the linker regions between LDLR class A repeats in several LDLR-related receptors (Steentoft, Vakhrushev, et al. 2013). We have now systematically characterized recombinant expressed LDLR, and confirmed evolutionarily conserved O-glycosylation in linker regions between the EGF-like class A repeats. In vitro studies together with ZFN knockout have furthermore established GalNAc-T11 as the main enzyme responsible for this modification (Pedersen et al. 2014) and we have now in preliminary studies shown that this modification affects LDL binding (unpublished).

The use of precise gene editing to develop isogenic cell lines with defined repertoire of GalNAc-Ts isoforms clearly has the potential to uncover their biological functions and the purpose for the many members of this gene family. Recently, a similar approach was presented for sphingolipids (Yamaji and Hanada 2014) using TALENs to knockout key genes including
B4GalT5, thereby confirming previous studies showing that B4GalT5 and not B4GalT6 is the main LacCer synthase (Figure 4).

**Cell lines: Custom engineering the glycosylation capacity of CHO cells for bioprocessing**

The true workhorse for recombinant production of biologics is the mammalian CHO cell line. CHO cells have been used safely for the production of glycoprotein therapeutics over the last 25 years, being the host cells of choice due to available gene amplification systems for high titer and specific productivity, as well as capacity for posttranslational modifications especially glycosylation. CHO cells generally produce recombinant therapeutic glycoproteins with rather simple glycoforms that are fully compatible and non-immunogenic in man. The genome and transcriptome of CHO-K1 were recently reported (Xu et al. 2011), as well as a comparative analysis of multiple CHO production lines (Lewis et al. 2013), and it is evident that the compatible glycosylation profile of CHO is maintained at the regulatory level rather than by gene inactivation or loss. Thus, the typical immunogenic glycan structures expressed in non-human mammals, including NeuGc sialic acid and α1,3Gal capping of oligosaccharide chains, are generally not produced (Jenkins et al. 1996; Xu et al. 2011). CHO cell N-glycans are primarily of bi- to tetraantennary complex structures with a variable degree (up to 26) of poly-N-acetyllactosamine and α2,3NeuAc capping, while O-glycans (GalNAc-type) are core1 structures with α2,3NeuAc capping and a variable degree of α2,6NeuAc to the innermost GalNAc residue (Do and Cummings 1993; North et al. 2010). The availability of the genome and transcriptome of CHO in combination with precise gene editing technologies now open possibilities for extensive engineering of the glycosylation capacities of the CHO cell factory. Past decades of work using random plasmid transfection strategies to enhance the glycosylation capacity of CHO (reviewed in Rita Costa et al. 2010) have largely failed. This is primarily due to a lack of scalability and consistency of glycosylation as well as due to the requirement for constant antibiotic selection for transgenes that is incompatible with use for human therapeutics. However, precise gene editing has a great potential to pave the way for a new era where CHO can be engineered to produce more homogenous glycoforms, more complex glycosylation and custom designed glycosylation for new therapeutic options.

The first reported example of ZFN glycogene editing in CHO was the knockout of the α1,6-fucosyltransferase, \textit{FUT8} (Malphettes et al. 2010). Therapeutic monoclonal antibodies carrying the core α6Fuc residue have substantially reduced antibody-dependent cellular cytotoxicity effects as first demonstrated for IgG1 produced in a CHO lectin-resistant mutant (Lec13) (Shields et al. 2002). \textit{FUT8} was previously knocked out in CHO cells using sequential HR, an impressive achievement requiring screening of >100,000 clones (Yamane-Ohnuki et al. 2004). ZFN-mediated biallelic knockout of \textit{FUT8} (Malphettes et al. 2010) was achieved in 3 weeks without selection. Alternatively, a CHO cell line deficient in fucose has been accomplished by ZFN targeting of the GDP-fucose transporter (\textit{Slc35c1}) (Zhang et al. 2012).

Another example of glycoengineering using ZFN targeting of \textit{Mgat1} was recently reported in CHO (Sealover et al. 2013). \textit{Mgat1} is necessary for the synthesis of complex N-glycans through the addition of GlcNAc to the Man5GlcNAc2 (Man5) structure (Figure 4). For a number of glycoprotein products complex type N-glycans capped by sialic acids are desired as these glycans improve protein half-life and prevent immunogenicity (Hossler et al. 2009). However, high-mannose N-glycans may be desirable for targeting N-glycoprotein therapeutics to mannose receptors for immune-stimulation (Betting et al. 2009) or for use to produce N-glycoproteins for crystallization, where the N-glycans are more homogenous and can be trimmed down by glycosidase treatment (reviewed in Nettleship et al. 2010). \textit{De novo} targeting of \textit{Mgat1} by ZFNs enabled selection of clones with production and growth rates comparable or even slightly better than the wild-type CHO (Sealover et al. 2013), demonstrating the value of this technology for engineering host cells for bioprocessing. The original CHO Lec1 cell lines are also deficient in \textit{Mgat1} (Chen and Stanley 2003), and could serve the same purpose in production of N-glycoproteins with high-mannose structures suitable for crystallization studies.

Glycoengineering CHO lines for production of therapeutic glycoproteins pose a more complex task. A number of issues need to be addressed in the development of new production cell lines including the level of plasticity of the CHO cell factory for deletion and/or introduction of glycogenes. Maintenance of cell viability and growth properties as well as production and secretion levels of recombinant glycoproteins are obviously parameters that need to be selected for throughout glycoengineering efforts of CHO cells. Another essential parameter is consistency of engineered glycosylation during bioprocess scale up and high-yield production.

**Perspectives**

The precision genome editing tools presented here have already found wide applications in many fields of cell biology and applications in glycobiology are now emerging. There are technical challenges with each of the editing tools, but many of these are being overcome at a rapid pace. Especially, the latest CRISPR/Cas9 technology is developing quickly, and the first impressive reports of whole-genome knockout targeting have recently appeared (Shalem et al. 2013; Wang, Wei, et al. 2013; Koike-Yusa et al. 2014). Validated targeting constructs for all human and rodent genes are being made available through several companies. Furthermore, whole-genome lentiviral CRISPR/Cas-guide RNA libraries used in recent studies are also being made available for academic use (Shalem et al. 2013; Koike-Yusa et al. 2014). ZFNs are still the most frequently published method; however, it was the first technology to be presented and a substantial number of the publications are focused on method development and not biological applications. Publications on TALEN and CRISPR/Cas are definitely catching up. Especially, CRISPR/Cas has seen a burst in number of publications since the first studies were presented, possibly reflecting the low cost and ease of design for this tool.

The gene editing tools represent a revolution for the glycobiology field. With the ease and efficiency of precise gene editing tools...
editing techniques, it is now possible to dissect glycosylation pathways and biological functions of complex carbohydrates by a genetic approach. Most mammalian glycosyltransferase genes appear to have been identified and the enzymatic functions at least partly characterized, but we are still not able to translate gene expression directly to glycosylation capacity. This is in part due to our poor understanding of the specific functions of individual isoenzymes in the many homologous glycosyltransferase gene families, such as for example the β4Gal-Ts (Amado et al. 1999) and β3GlcNAc-Ts (Narimatsu 2006) controlling (poly)-N-acetyllactosamine biosynthesis on different glycoconjugates and oligosaccharide structures, or the many polypeptide GalNAc-Ts controlling sites of O-glycosylation of proteins (Bennett et al. 2012). Clearly, a number of other factors, besides expression of glycosyltransferases, influence the final glycosylation outcomes, including potential regulation of glycosyltransferases, availability of substrates, competition for substrates and organisation (or rather disorganization) of the secretory pathway. However, with the presented gene editing technology, it is now possible to use knockout screening strategies to define the in vivo function of all glycogenes and their roles in glycosylation. Thus, a major obstacle for in silico prediction of protein and lipid glycosylation in a cell may be overcome in the future, and this would greatly facilitate access to, and wider understanding of, complex carbohydrates beyond the glycobiology community.

Another important outcome of these editing tools is the development of isogenic cell systems that enable integrative polyOmic studies. Isogenic cells with and without any chosen glycogene offer a unique platform to study in vivo functions in a homogenous system without limitations in the amounts of material for analysis. The option to link individual glycogenes with glycosylation changes as well as variations in, for example, transcriptomes, proteomes, glycoproteomes or phosphoproteomes in simple isogenic cell systems have the potential to greatly broaden discovery efforts of biological functions of complex carbohydrates. Generation of isogenic cell models with deleted or inserted glycogenes will provide a tremendous resource for future glycobiology research. We have demonstrated how cell lines with simplified glycosylation can be applied to characterize two O-glycoproteomes, O-GalNAc (Steentoft et al. 2011) as well as O-Man (Vester-Christensen et al. 2013), and to probe isofrom-specific functions of individual members of the large GalNAc-T family (Schjoldager et al. 2012). We have now expanded the SimpleCell library to include proteoglycan SimpleCells by targeting B4GALT7 (Almeida et al. 1999) generating cells that only are able to add the initial xylose residue (Steentoft et al. unpublished). We have also expanded the SimpleCell approach to include an organotypic culture system using the spontaneously immortalized non-transformed human keratinocyte HaCaT cell line, and found that simplified O-GalNAc glycosylation (COSMC knockout) affects epithelial tissue architecture and induces oncogenic features including enhanced proliferative and invasive properties (Radhakrishnan et al. unpublished).

Moreover, the ability to precisely correct or introduce disease causing mutations will provide a powerful tool for the discovery and molecular dissection of CDGs associated with hypomorphic mutations. Introduction of mutations identified by, e.g. whole-exome sequencing studies and subsequent validation of mutation effects in a simple cell system should be invaluable. On top of this, by avoiding extensive backcrossing and slow selection, transgenic animal models can be generated much faster than previously possible.

Finally, precision genome editing clearly has an enormous potential for the generation of production cells lines customized for individual product requirements. Early efforts with ZFN engineering of CHO lines has been discussed, but given the new abilities to eliminate incompatible glycosylation in any cell one may envision development of other even better production cell lines as well. Recombinant biologics will likely be the first therapeutic applications of the editing tools, but corrective gene-therapy of glycogenes is on the horizon. ZFNs have already been in clinical use with introduction of deletions in the CCR5 co-receptor for HIV (Maier et al. 2013), and pre-clinical studies are in progress on lysosomal storage diseases, although no data have been published yet.

Funding
This work was supported by Kirsten og Freddy Johansen Fonden; A.P. Møller og Hustru Chastine Mc-Kinney Møllers Fond til Almene Formaal; The Novo Nordisk Foundation, Excellence Programme for Interdisciplinary Research University of Copenhagen; The Danish Research Councils; and the Danish National Research Foundation (DNRF107).

Acknowledgements
A version of this review constitutes a part of the PhD thesis by C Steentoft. The authors express their gratitude to the entire glycobiology team in Copenhagen for their helpful discussion and input.

Conflict of interest statement
None declared.

Abbreviations
Aa, amino acid; Cas, CRISPR associated; CDG, congenital disorder of glycosylation; CHO, Chinese Hamster Ovary; CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, CRISPR RNA; DSB, double-stranded break; ETD, electron transfer dissociation; FACS, fluorescence-activated cell sorting; GalNAc, N-acetylgalactosamine; HR, homologs recombination; indels, deletions or insertions; LDLR, low-density lipoprotein receptor; LIAC, lectin weak-affinity chromatography; MS, mass spectrometry; NHEJ, non-homologous end joining; OPEN, oligomerized pool engineering; PAM, protospacer-adjacent motif; RVD, repeat-variable di-residue; SCNT, somatic cell nuclear transfer; sgRNA, single-guide RNA; SNV, single nucleotide variant; ssODN, single-stranded oligodeoxynucleotides; TALEN, transcription activator-like effector nucleases; TALE, transcription activator-like effector; tracrRNA, trans-activating crRNA; VVA, Vicia villosa agglutinin; ZFN, zinc finger nuclease.
References


