A new insect cell line engineered to produce recombinant glycoproteins with cleavable *N*-glycans

Received for publication, August 13, 2021, and in revised form, November 16, 2021 Published, Papers in Press, November 26, 2021,
https://doi.org/10.1016/j.jbc.2021.101454

Hideaki Mabashi-Asazuma[1] **and Donald L. Jarvis**[1,2,*]

From the [1]*Department of Molecular Biology, University of Wyoming, Laramie, Wyoming, USA;* [2]*GlycoBac, LLC, Laramie, Wyoming, USA*

Edited by Gerald Hart

Glycoproteins are difficult to crystallize because they have heterogeneous glycans composed of multiple monosaccharides with considerable rotational freedom about their *O*-glycosidic linkages. Crystallographers studying *N*-glycoproteins often circumvent this problem by using β1,2-*N*-acetylglucosaminyltransferase I (MGAT1)–deficient mammalian cell lines, which produce recombinant glycoproteins with immature *N*-glycans. These glycans support protein folding and quality control but can be removed using endo-β-*N*-acetylglucosaminidase H (Endo H). Many crystallographers also use the baculovirus-insect cell system (BICS) to produce recombinant proteins for their work but have no access to an MGAT1-deficient insect cell line to facilitate glycoprotein crystallization in this system. Thus, we used BICS-specific CRISPR–Cas9 vectors to edit the *Mgat1* gene of a rhabdovirus-negative *Spodoptera frugiperda* cell line (Sf-RVN) and isolated a subclone with multiple *Mgat1* deletions, which we named Sf-RVN^{Lec1}. We found that Sf-RVN and Sf-RVN^{Lec1} cells had identical growth properties and served equally well as hosts for baculovirus-mediated recombinant glycoprotein production. *N*-glycan profiling showed that a total endogenous glycoprotein fraction isolated from Sf-RVN^{Lec1} cells had only immature and high mannose-type *N*-glycans. Finally, *N*-glycan profiling and endoglycosidase analyses showed that the vast majority of the *N*-glycans on three recombinant glycoproteins produced by Sf-RVN^{Lec1} cells were Endo H-cleavable $Man_5GlcNAc_2$ structures. Thus, this study yielded a new insect cell line for the BICS that can be used to produce recombinant glycoproteins with Endo H-cleavable *N*-glycans. This will enable researchers to combine the high productivity of the BICS with the ability to deglycosylate recombinant glycoproteins, which will facilitate efforts to determine glycoprotein structures by X-ray crystallography.

N-glycoproteins are a subclass of proteins with one or more oligosaccharide side chains, or glycans, covalently linked through amide bonds to asparagine residues in the polypeptide backbone. *N*-glycans can influence basic glycoprotein properties, including their folding, degradation, physical stability, and biological functions, as well as physiological behaviors including their half-lives in the human circulatory system (1). However, *N*-glycans usually have an adverse practical impact on efforts to determine glycoprotein structures because they can make it difficult or impossible to use X-ray crystallography for this purpose (2–4). The main reason is that "an *N*-glycan" linked to any given site on the polypeptide is actually comprised of a structurally heterogeneous population of multiple monosaccharide assemblies. In addition, each monosaccharide in those assemblies has significant mobility about its *O*-glycosidic linkage. Both the heterogeneity and mobility of these glycans contribute to the difficulties associated with efforts to produce high-quality glycoprotein crystals for downstream analysis.

Crystallographers can circumvent this problem by producing recombinant glycoproteins in mutant mammalian cell lines lacking functional β1,2-*N*-acetylglucosaminyltransferase I (MGAT1; (5)). MGAT1 adds a terminal *N*-acetylglucosamine residue to the lower branch mannose of $Man_5GlcNAc_2$, which produces the first endo-β-*N*-acetylglucosaminidase H (Endo H)-resistant intermediate in the mammalian (Fig. 1*A*) and insect (Fig. 1*B*) cell *N*-glycan processing pathways (6, 7). Thus, MGAT1-deficient mammalian cell lines such as CHO *Lec1* (8) and human embryonic kidney 293S GnT1$^-$ (9) produce immature and relatively minimally processed *N*-glycan precursors that can be removed from the polypeptide backbone using Endo H (Fig. 1*A*). Endo H cleaves the O-glycosidic linkage in the chitobiose core of all high-mannose precursors in the *N*-glycan processing pathway from $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ (Fig. 1, *A* and *B*). This releases the bulk of the immature *N*-glycan structure from the polypeptide backbone, leaving behind just a single *N*-linked *N*-acetylglucosamine residue. This monosaccharide is homogeneous and contributes far less rotational mobility than the intact *N*-glycan, which is why glycoproteins are much easier to crystallize after being deglycosylated by Endo H. This is why many crystallographers focused on determining that glycoprotein structures use MGAT1-deficient mammalian cell lines to produce recombinant glycoproteins, then deglycosylate the products with Endo H, as this approach greatly facilitates their efforts to produce high-quality crystals for X-ray crystallography.

Many crystallographers also use the baculovirus-insect cell system (BICS) to produce recombinant proteins for their

* For correspondence: Donald L. Jarvis, DLJarvis@uwyo.edu.
Present address for Hideaki Mabashi-Asazuma: Department of Nutrition and Food Science, Graduate School of Humanities and Sciences, Ochanomizu University, 2-1-1 Otsuka, Bunkyo, Tokyo 112-8610, Japan.

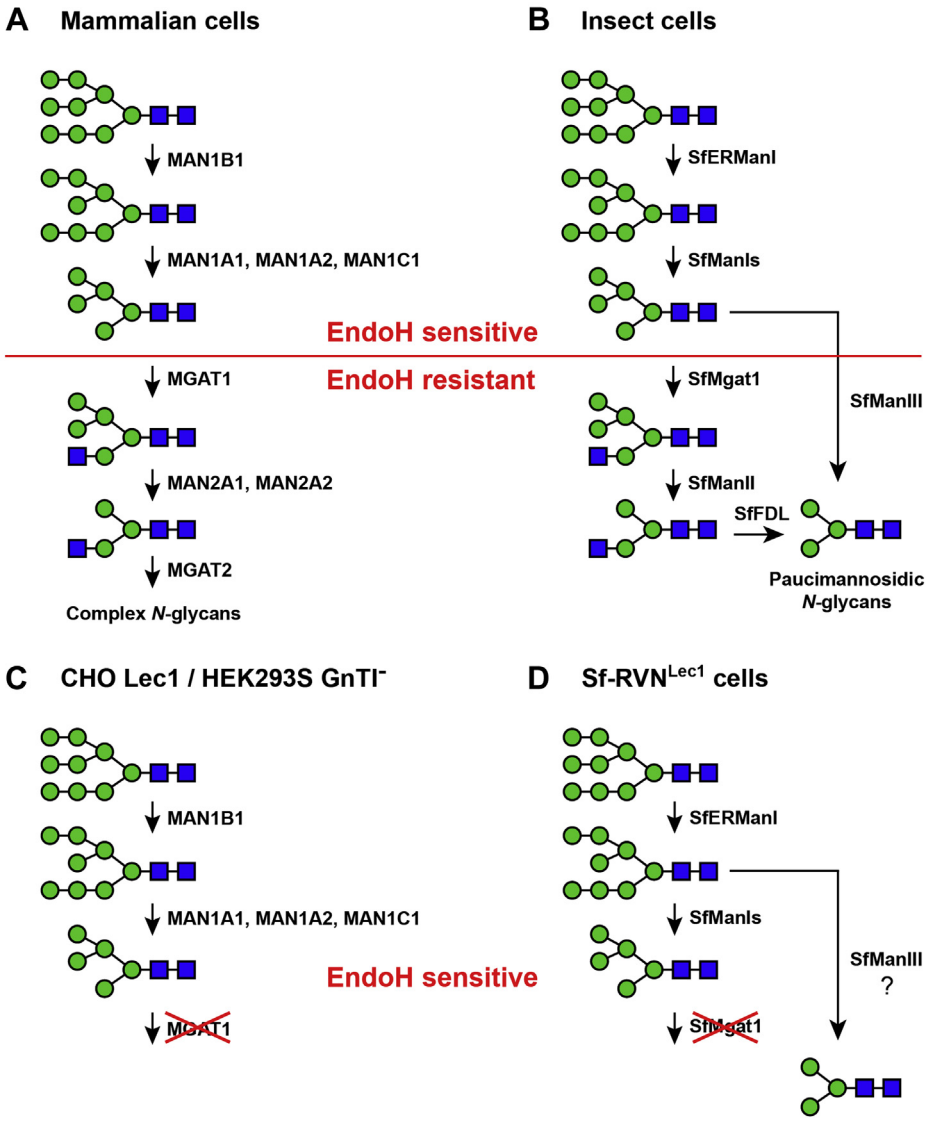

Figure 1. *N*-glycan processing pathways in mammalian and insect cells. *A* and *B*, initial steps in the *N*-glycan processing pathways of mammalian or insect cells, respectively, with a *horizontal line* delineating Endo H-sensitive and Endo H-resistant structures. *C* and *D*, *N*-glycan processing pathways of MGAT1-deficient mammalian (CHO Lec1 and HEK293S GnTI$^-$) and insect (SfRVN^{Lec1}) cells, respectively. All *N*-glycans shown in panels *C* and *D* are Endo H sensitive. For simplicity, core fucosylation is not shown. MGAT1, β1,2-*N*-acetylglucosaminyltransferase I; Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

work because this system can support high yields of at least some proteins and multiprotein complexes (10). Moreover, recombinant glycoproteins produced by insect cell lines used as hosts in the BICS have smaller and less heterogeneous *N*-glycans than their native mammalian cell counterparts (11). In some cases, the reduced heterogeneity and mobility of these smaller, simpler, and insect-type *N*-glycans enables isolation of high-quality crystals without prior deglycosylation (12, 13). However, this is not always true. In addition, because insect cells encode and express MGAT1 (14), they, too, can produce recombinant glycoproteins with Endo H-resistant *N*-glycans (Fig. 1*B*; (15)). Unfortunately, crystallographers using the BICS for recombinant glycoprotein production cannot utilize the Endo H deglycosylation approach to circumvent this problem because there is no MGAT1-deficient insect cell line available for use as an alternative host in this system. Thus, the overall purpose of this study was to create and characterize an MGAT1-deficient insect cell line that can be used as an alternative host in the BICS for the production of recombinant glycoproteins with Endo H-cleavable *N*-glycans.

Our basic approach was to use a CRISPR–Cas9 vector specifically developed for use in the BICS (16) to edit the *mgat1* gene of an insect cell line derived from the fall armyworm, *Spodoptera frugiperda* (Sf). The parental cell line we used for this purpose was rhabdovirus-negative Sf (Sf-RVN) cell line, a derivative of Sf9 (17), which has been widely used as a host in the BICS since this system was first described in the literature nearly 40 years ago (18). As indicated by its name, Sf-RVN cells have no detectable Sf-rhabdovirus, and these cells also have no other detectable adventitious viral contaminants (19, 20).

In short, we used the CRISPR–Cas9 approach to edit the Sf-RVN *mgat1* (*Sfmgat1*) gene and isolated a clonal derivative that can only produce recombinant *N*-glycoproteins with immature and Endo H-sensitive *N*-glycans. We designated this new cell line Sf-RVN^{Lec1}, in recognition of the *mgat1*-deficient CHO cell lines isolated by Stanley *et al.* (8) nearly 50 years ago. We expect that this new insect cell line will be useful as an alternative host to enhance the utility of the BICS as a tool for glycoprotein crystallography and structural biology.

Results and discussion

Identifying an effective CRISPR–Cas9 targeting sequence for the Sf-Mgat1 gene

The first step in our effort to knock out *Sfmgat1* was to identify an effective targeting sequence, or single-guide RNA (sgRNA) "spacer," for insertion into the Sf-specific CRISPR–Cas9 vector, pIE1-Cas9-SfU6-sgRNA-Puro (16). This was important because our overall goal was to create a clonal *Sfmgat1*-deficient Sf-RVN cell derivative, and an effective targeting sequence would drive efficient and specific genome editing, which would produce a higher proportion of Sf-RVN cells with *Sfmgat1* indels. In practical terms, this would enhance the probability that we could isolate single-cell clones with disruptive *Sfmgat1* indels from the polyclonal CRISPR–Cas9 edited Sf-RVN cell population. We did not expect this to be trivial, as most of the cells in this population would likely have at least one copy of the *Sfmgat1* gene with no indels or in-frame indels that failed to knock out MGAT1 function.

Inspection of the *Sfmgat1* gene sequence revealed multiple PAM sequences that allowed us to design eight distinct targeting sequences, designated SfMgat1t1 to SfMgat1t8, which were distributed over five different exons (Fig. S1 and Table S1). Eight *Sfmgat1*-specific CRISPR–Cas9 vectors encoding each of these targeting sequences were produced by inserting synthetic oligonucleotides into pIE1-Cas9-SfU6-sgRNA-Puro, as described in the Experimental procedures section. Each of these vectors encoded sgRNAs with a different targeting sequence under the transcriptional control of the endogenous Sf cell U6-3 promoter (16). These vectors were used to transfect individual Sf-RVN cultures, and polyclonal transformed cell populations were selected by growth in medium supplemented with puromycin. Finally, the gene-editing efficiencies obtained using each CRISPR–Cas9 vector were assessed, as described in the Experimental procedures section.

Typically, gene-editing efficiencies can be quickly and easily assessed by performing CEL-I nuclease assays and/or by directly sequencing relevant genomic PCR fragments and using an algorithm to deconvolute the raw sequencing data. However, in this case, we were unable to use either of these standard approaches to analyze all eight targeted sequences because the complete *Sfmgat1* sequence was not available at the time. Therefore, we invoked an alternative approach, which circumvented the need for the full *Sfmgat1* sequence and provided some information, albeit preliminary and indirect, on the phenotypic impact of each targeting sequence.

This approach involved infecting samples of the parental or polyclonal CRISPR–Cas9 transformed Sf-RVN cell derivatives with a baculovirus vector encoding human erythropoietin (hEPO). When this human glycoprotein is produced in Sf-RVN cells, it acquires three *N*-glycans with relatively small paucimannosidic structures consisting mostly of $Man_{3}GlcNAc_{2}$±Fuc (Fig. 1*B*; (19)). In contrast, recombinant hEPO produced in an MGAT1-deficient Sf-RVN cell derivative would be expected to acquire larger and high-mannose *N*-glycan structures consisting of $Man_{5-9}GlcNAc_{2}$. This would increase the total mass of the hEPO produced in MGAT1-deficient cells by an average of ∼1000 Da and reduce its electrophoretic mobility upon analysis by SDS-PAGE. Therefore, we hypothesized that we could qualitatively assess the genome-editing efficiencies obtained using CRISPR–Cas9 vectors encoding the different targeting sequences by comparing the electrophoretic mobilities of recombinant hEPOs produced by the parental and CRISPR–Cas9 transformed Sf-RVN cell derivatives. The most effective targeting sequences would drive the most efficient CRISPR–Cas9 editing, which would produce Sf-RVN cell derivatives most likely to contain disruptive *Sfmgat1* indels, and those derivatives would be the most likely to produce recombinant hEPO with a reduced electrophoretic mobility.

We were able to use this screen to successfully identify promising polyclonal Sf-RVN cell derivatives transformed with the CRISPR–Cas9 vectors encoding targeting sequences SfMgat1t3, SfMgat1t4, SfMgat1t6, SfMgat1t7, or SfMgat1t8, as each produced recombinant hEPO with a reduced electrophoretic mobility relative to the parental control (Fig. 2). The derivative transformed using the SfMgat1t4 vector produced hEPO with the slowest or among the slowest electrophoretic mobilities observed in this assay. This suggested that these cells produced very little or no functional MGAT1 and indicated SfMgat1t4 was likely the most or among the most effective targeting sequences.

The identification of SfMgat1t4 as an effective targeting sequence was not too surprising, considering it was designed to target a region in the *Sfmgat1* gene encoding several highly conserved amino acids. These included a proline residue conserved in all MGAT1 orthologs (P138 in *Cricetulus griseus*), which was shown to be essential for MGAT1 function by chemical mutagenesis and lectin selection of CHO-DUKX cells, as this untargeted approach produced a mutant with a P138L substitution and Lec1 phenotype (21). Molecular

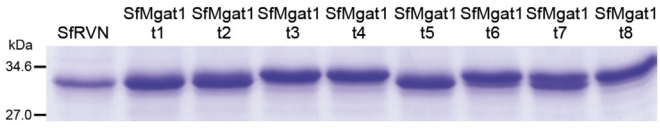

Figure 2. Screening polyclonal edited/transformed cell populations using an electrophoretic mobility shift assay. Sf-RVN cells were transfected with eight distinct CRISPR–Cas9 vectors (SfMgat1t1–SfMgat1t8), then selected for puromycin resistance to produce eight distinct CRISPR–Cas9 edited and transformed polyclonal cell populations. Samples of each were infected with a baculovirus vector encoding hEPO, and then the secreted products were purified and analyzed by SDS-PAGE with Coomassie blue staining, as described in the Experimental procedures section. hEPO, human erythropoietin; Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

modeling indicated that P138 connects two β-strands proximal to the UDP-*N*-acetylglucosamine donor substrate–binding pocket, which is probably why this proline residue is required for catalysis (21). Other conserved amino acids immediately downstream of P138 most likely contribute to the structure of the donor substrate–binding pocket, as well (22). Thus, SfMgat1t4 was likely to be a particularly effective targeting sequence, as even in-frame indels in this region could disrupt MGAT1 function. In addition, SfMgat1t4 was directed against a sequenced region of the *Sfmgat1* gene, which would enable us to PCR amplify this region of *Sfmgat1* in CRISPR–Cas9 transformed derivatives for further analysis. Based on the markedly reduced electrophoretic mobility of the hEPO produced by these cells, the presence of key and highly conserved amino acids in the targeted region, and the practical value of having the sequence of this region of the *Sfmgat1* gene in hand, we proceeded to isolate single-cell clones from the polyclonal population of Sf-RVN cells transformed with the CRISPR–Cas9 vector encoding the SfMgat1t4 targeting sequence, henceforth designated "SfMgat1t4 cells."

Isolating and genotyping SfMgat1t4 cell clones

Using a previously described end-point dilution method (23), we isolated 11 single-cell clones from the polyclonal SfMgat1t4 cell population and then expanded and genotyped each one. Our initial genotyping approach involved PCR amplifying the targeted *Sfmgat1* sequence from each clone, directly sequencing the PCR products using the Sanger method (24), and deconvoluting the raw data using the TIDE (Tracking of Indels by DEcomposition) algorithm (25), as described in the Experimental procedures section.

The first conclusion derived from the results of the TIDE analysis was that all 11 SfMgat1t4 clones had at least three distinct indels (Fig. 3). Considering we genotyped clonally derived Sf-RVN cell derivatives, this might be confusing. However, the presence of multiple indels can be easily explained by the fact that lepidopteran insect cell lines have polyploid genomes. Established insect cell lines have numerous fragmented "microchromosomes" (26–28), which have been visualized in Sf9 cell nuclei as collections of propridium iodine–stained foci averaging ∼350 per cell (29). These polyploid karyotypes have been attributed to the fact that lepidopteran insect cell chromosomes have no centromeres (30). The presence of multiple indels in a single genetic target has been observed in previous studies that included genotyping of CRISPR–Cas9 edited subclones of several different insect cell lines (16, 31–33).

Another conclusion from the results of the TIDE analysis was one of the 11 clones isolated from the polyclonal SfMgat1t4 cell population, which we designated SfMgat1t4#4, had no detectable wildtype sequences or in-frame mutations, but had three frame-shift mutations within the targeted sequence (Fig. 3). This was exciting because the detection of frame shift mutations in the apparent absence of wildtype sequences or in-frame deletions in this polyploid cell line was consistent with the possibility that it might not have a functional copy of the *Sfmgat1* gene. Unfortunately, while SfMgat1t4#4 had the most favorable genotype, it did not grow well after being scaled up to the 50 ml shake flask level (data not shown). Thus, we shifted our attention to other SfMgat1t4 clones, even though they all had one or more in-frame deletions (Fig. 3).

The potential significance of a CRISPR–Cas9–edited Sf-RVN derivative with multiple frame-shift mutations and no detectable wildtype sequences or in-frame deletions in the targeted site was duly noted previously. However, a clone with no detectable wildtype copies and one or more copies of *Sfmgat1* with an in-frame deletion in the targeted area could, nevertheless, be a loss of function mutant. The retention or loss of MGAT1 function would depend upon the specific amino acid(s) eliminated by the in-frame deletion in the nucleotide sequence. As noted previously, the SfMgat1t4 targeting sequence was directed against a region of the *Sfmgat1* gene that included several conserved amino acid residues required for catalysis (21, 22). Thus, it seemed likely that even an in-frame deletion in this region might be intolerable and could knock out MGAT1 function.

Of the original 11 clones isolated and subjected to genotyping in this study, only SfMgat1t4#1 grew well after being expanded to the 50 ml shake flask level. However, the TIDE results showed that the targeted *Sfmgat1* sequence in this clone was heavily edited. A total of ∼41% of the deconvoluted sequences included out-of-frame deletions, and another ∼17% included a 9 bp in-frame deletion in the targeted region (Fig. 3). To validate these results, we used a different approach to genotype SfMgat1t4#1, which involved PCR amplifying the targeted *Sfmgat1* fragment, cloning the amplification products, and sequencing the inserts in 30 independent clones. This analysis confirmed the presence of the −2, −7, and −9 deletions detected by TIDE analysis and revealed the additional presence of −3, −5, and −28 deletions in the targeted *Sfmgat1* sequence (Fig. 4*A*). The most likely reason the PCR, cloning, and sequencing approach did not reveal the −1, −4, and −8 deletions is that the TIDE analysis is inaccurate if the sequence is comprised of a highly complex mixture, which was true in this case. Overall, our results indicated that SfMgat1t4#1 cells included four frame shift (−2, −5, −7, and −28) mutations and not just one, but at least two potentially problematic in-frame (−3 and −9) mutations in the targeted region of *Sfmgat1*. However, closer inspection of the −3 deletion indicated it could not be tolerated because it eliminates P138, which is required for MGAT1 catalysis (Fig. 4). The impact of the −9 deletion was not as clear, as it starts immediately downstream of the codon for P138 and does not eliminate this conserved proline residue (Fig. 4). However, it eliminates a highly conserved IIV motif located in the β-sheet adjacent to the UDP-*N*-acetylglucosamine binding site (21, 22). Thus, it was likely this in-frame deletion would be intolerable, as well.

Finally, our discussion of the SfMgat1t4#1 genotype also should address an issue associated with the approach used to isolate this cell line. Sf-RVN cells were transfected using a CRISPR–Cas9 vector encoding the SfMgat1t4 targeting sequence, a constitutively expressible Cas9 gene and a

A

Cell line	Frame shift deletions	In-frame deletions
SfMgat1t4#1	-1, -2, -4, -7, -8 bp	-9 bp
SfMgat1t4#2	-4, -5 bp	-3, -6 bp
SfMgat1t4#3	-2, -4, -5, -7, -8 bp	-3, -6, -9 bp
SfMgat1t4#4	-2, -4, -7 bp	
SfMgat1t4#5	-2, -5, -7, -8 bp	-6, -9 bp
SfMgat1t4#6	-1, -4, -5, -10 bp	-6 bp
SfMgat1t4#7	-1, -5, -7, -8 bp	-3, -6, -9 bp
SfMgat1t4#8	-5, -10 bp	-6, -9 bp
SfMgat1t4#9	-2, -4 bp	-6 bp
SfMgat1t4#10	-4, -7, -8, -10 bp	-6 bp
SfMgat1t4#11	-4, -7 bp	-6, -9 bp

B

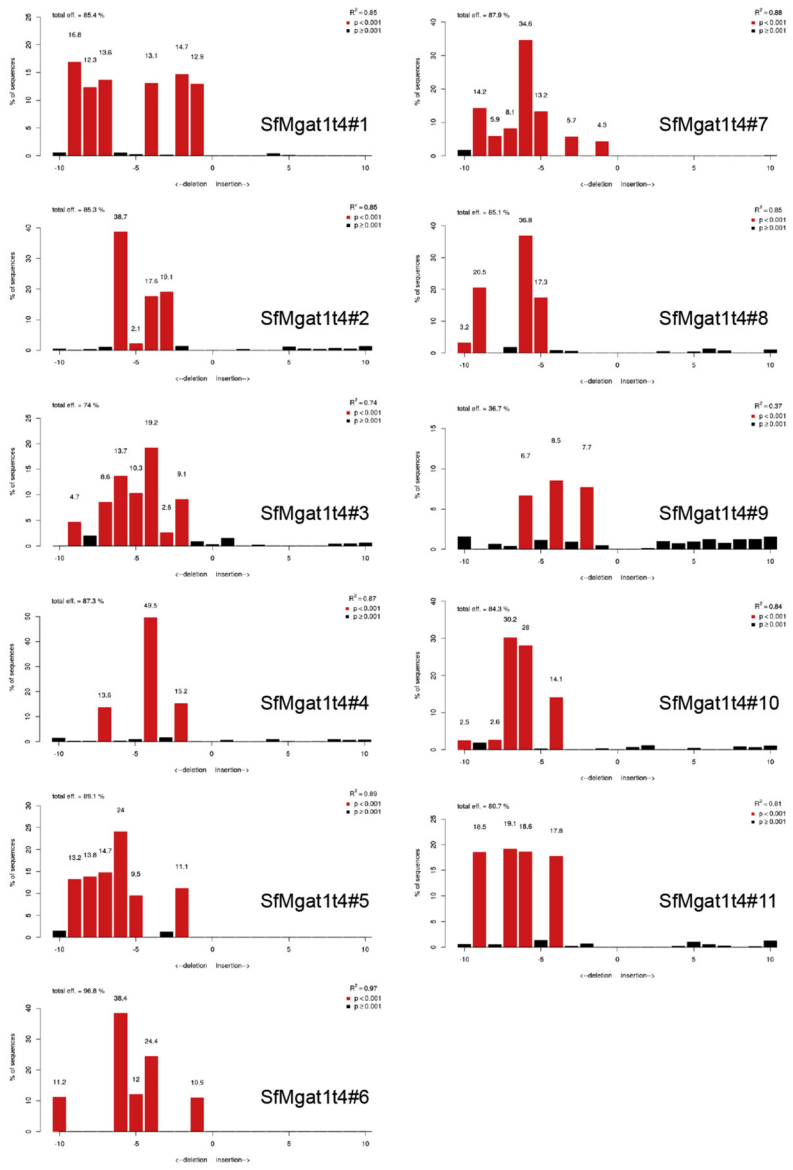

Figure 3. Initial genotyping of SfMgat1t4 cell clones. Eleven single-cell clones were isolated from the polyclonal and transformed SfMgat1t4 population, and the targeted region of the *Sfmgat1* gene in each clone was PCR amplified. The amplification products were then purified, directly sequenced, and the TIDE program was used to deconvolute the sequencing data, as described in the Experimental procedures section. *A*, summary of major indels observed and *B*, frequencies of major indels identified in each SfMgat1t4 clone. *Sfmgat1*, Sf-RVN mgat1; TIDE, Tracking of Indels by DEcomposition.

puromycin resistance marker, and then a puromycin-resistant subpopulation was selected. Thus, SfMgat1t4#1 is a genetically transformed Sf-RVN subclone designed to constitutively express the CRISPR–Cas9 machinery. This means that the cellular gene can be continuously edited as the subclone is passaged. The advantage of this approach is that any residual wildtype or repaired copies of *Sfmgat1* would be susceptible to this additional editing and potential knockout. The disadvantage is that continuous editing might also have off-target effects, which could compromise key features of the cell line, including its growth and quality as a host for baculovirus vectors, including its susceptibility to infection and its productivity, among others. The expectation that the *Sfmgat1* sequence will continue to evolve is supported by observations from our previous efforts to edit Sf cell genomes using these vectors, which showed that the proportion of indels in targeted sequences increased with increasing passage number (16). We have not yet observed any obvious or significant deleterious effects potentially induced by off-target editing with increasing passage number in any of these cell lines. Nevertheless, it is possible that this problem could be encountered with higher passages of SfMgat1t4#1 cells and, therefore, we note here, that it will be prudent to use low-passage cultures for recombinant glycoprotein production.

Based on the high efficiency of gene editing observed in a region of the *Sfmgat1* gene that encodes highly conserved amino acids in close proximity to the donor substrate–binding site as well as the significantly reduced electrophoretic

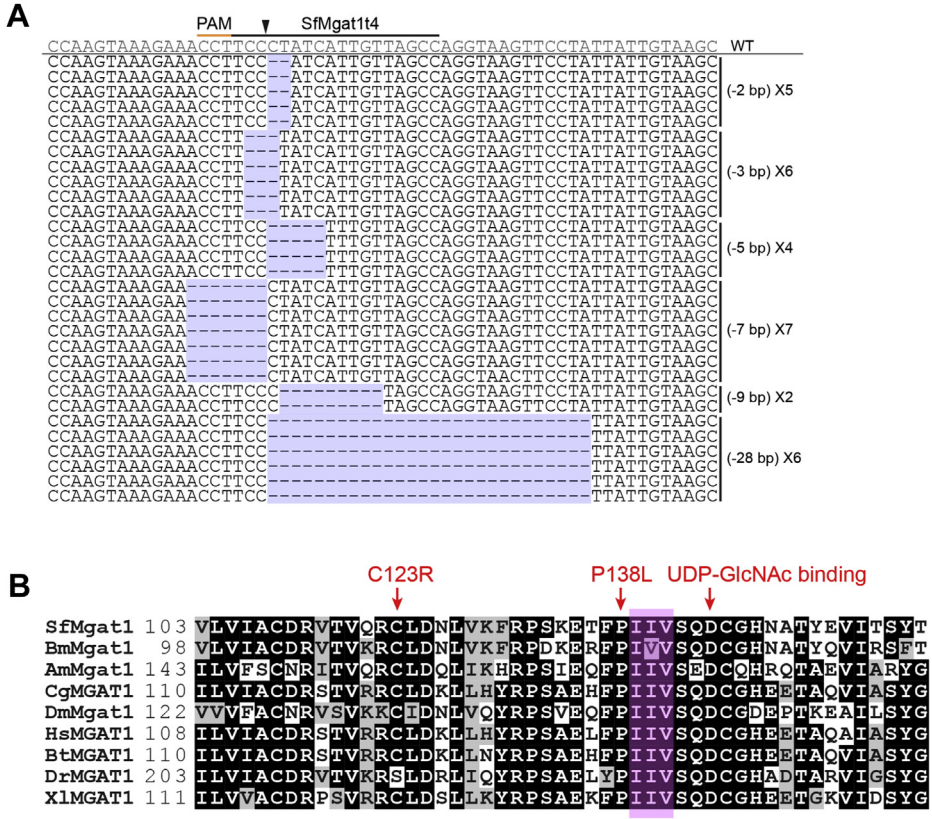

Figure 4. Extended analysis of Sf-RVN^{Lec1} cell genotype. The targeted region of the *Sfmgat1* gene in SfMgat1t4#1 (*i.e.*, Sf-RVN^{Lec1} cells) was PCR amplified, and the amplification product was purified and inserted into pGEM-T Easy. Thirty independent *Escherichia coli* transformants were randomly picked, and each insert was sequenced, as described in the Experimental procedures section. *A*, sequences of all 30 inserts showing the type and estimating the frequency of indels in Sf-RVN^{Lec1} cells. *B*, amino acid sequences and sequence conservation in the targeted region of MGAT1 from nine different species. Sf-RVN^{Lec1} cells have a 9 bp in-frame deletion, which eliminates the conserved IIV motif highlighted in *purple* and a 3 bp in-frame deletion, which eliminates the key proline residue labeled P138L. MGAT1, β1,2-*N*-acetylglucosaminyltransferase I; sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line; *Sfmgat1*, Sf-RVN *mgat1*.

mobility of the hEPO produced by the polyclonal transformed cells used as the source of this clone, we tentatively concluded that SfMgat1t4#1 cells most likely produce no functional MGAT1 and have the desired Lec1 phenotype. Thus, we renamed this new cell line Sf-RVN^{Lec1} and proceeded to characterize it in more detail.

Sf-RVN^{Lec1} growth properties and recombinant protein productivity

Our initial characterization of Sf-RVN^{Lec1} focused on its growth properties, including its growth rate and maximal cell density in suspension culture, in comparison to the parental Sf-RVN cell line (Fig. 5). We used replicate 50 ml shake flask cultures of Sf-RVN or Sf-RVN^{Lec1} cells in ESF-921 serum-free growth medium to measure these parameters, as described in the Experimental procedures section. The results showed that the growth rates and cell densities of the two cell lines were nearly identical up to 5 days after seeding. However, the Sf-RVN^{Lec1} culture achieved its maximal cell density of ∼5 × 10^6 cells/ml after 4 to 5 days of growth, whereas the Sf-RVN culture continued to grow and reached a slightly higher maximal cell density of ∼6.5 × 10^6 cells/ml after 6 days of growth. Overall, these results indicate that CRISPR–Cas9 editing had no significant adverse impact on the growth properties of Sf-RVN^{Lec1} cells, as compared with the parental line, at least not up to passage 15, which was the passage level

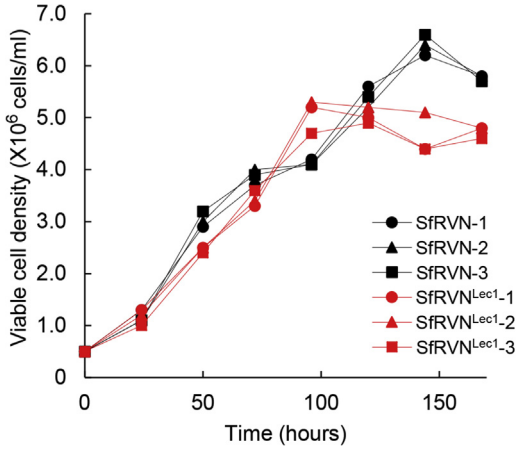

Figure 5. Growth properties of Sf-RVN^{Lec1} cells. Three replicate 50 ml shake flask cultures were seeded at a density of 5 × 10^5 cells/ml with either Sf-RVN cells (*black symbols*) or Sf-RVN^{Lec1} cells (*red symbols*). Samples were taken from each flask at 24 h intervals each day for 1 week, and then viable cell densities were measured, as described in the Experimental procedures section. The densities measured in each individual shake flask were plotted independently, with each flask represented by a different symbol. Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

of the Sf-RVN^{Lec1} culture used for this experiment. It is relevant to note that insect cell lines used as hosts in the BICS are routinely maintained and/or used for recombinant protein production at densities within the range of ~0.5 to 5.0 × 10^6 cells per ml, which can be accommodated by the growth properties of Sf-RVN^{Lec1} cells observed in this experiment.

We also examined the recombinant glycoprotein productivity of Sf-RVN^{Lec1} cells, in comparison to the parental Sf-RVN cells using hEPO as the model. Each cell line was infected with a recombinant baculovirus encoding a HIS-tagged version of hEPO, and the secreted hEPO products were purified by immobilized metal affinity chromatography (IMAC), as described in the Experimental procedures section. Samples were taken from various fractions isolated during the course of the purification scheme and analyzed by SDS-PAGE with Coomassie blue staining. The results showed that both cell lines produced approximately equal amounts of secreted hEPO (Fig. S2, starting material), and approximately equal yields of purified hEPO were obtained by IMAC (Fig. S2, elutions 1 and 2). We also found that both cell lines produced approximately equal amounts of human IgG1 Fc and mouse IgG2a Fc, which were the two other model glycoproteins used in this study (Figs. S4 and S5; data not shown). Thus, while it is fair to say we did not comprehensively analyze Sf-RVN^{Lec1} productivity using a broader selection of model glycoproteins, the results of this limited analysis showed that Sf-RVN^{Lec1} cells can support about the same levels of recombinant hEPO production as the parental cell line.

Sf-RVN^{Lec1} N-*glycosylation profile*

The definitive characteristic of the new insect cell line we hoped to isolate in this project would be its ability to produce only high mannose-type, Endo H-cleavable *N*-glycans, by analogy to mammalian *mgat1* knockout cell lines, such as CHO *Lec1* (8) and human embryonic kidney 293S GnT1$^-$ (9). Loss of MGAT1 function precludes *N*-glycan processing beyond the $Man_5GlcNAc_2$ step and results in the accumulation of this Endo H-sensitive intermediate (Fig. 1*C*). Interestingly, however, knocking out *Sfmgat1* would not necessarily be expected to produce this same phenotype because Sf cells have a novel class II processing α-mannosidase, which is called Sf-mannosidase III (SfManIII; (34, 35)). This unusual Golgi-localized enzyme can convert $Man_5GlcNAc_2$ directly to $Man_3GlcNAc_2$ with no requirement for the prior addition of a terminal *N*-acetylglucosamine residue by MGAT1 (Fig. 1*D*). Therefore, an *mgat1* knock out in Sf cells might be expected to produce an accumulation of $Man_3GlcNAc_2$, rather than $Man_5GlcNAc_2$, because of the action of SfManIII.

To directly examine the phenotypic impact of the *Sfmgat1* knock out, we produced total protein fractions from Sf-RVN and Sf-RVN^{Lec1} cells, enzymatically released and purified total *N*-glycans, and profiled their structures by MALDI-TOF–MS, as described in the Experimental procedures section. The results showed that the vast majority of the *N*-glycans isolated from the parental Sf-RVN cell glycoproteins were fucosylated and nonfucosylated paucimannosidic structures ($Man_3GlcNAc_2$±Fuc), as expected (Fig. 6*A*). In contrast, the major *N*-glycan isolated from Sf-RVN^{Lec1} cell glycoproteins was the high mannose structure ($Man_5GlcNAc_2$; Fig. 6*B*), which normally serves as the acceptor substrate for MGAT1. These results clearly demonstrated that the *Sfmgat1* gene was effectively knocked out in Sf-RVN^{Lec1} cells. However, it was surprising to find no detectable $Man_3GlcNAc_2$ in the Sf-RVN^{Lec1} *N*-glycosylation profile, considering Sf cells encode and express SfManIII, which could convert $Man_5GlcNAc_2$ to this product (34, 35). Perhaps, this can be explained by an unexpected and novel sub-Golgi compartmentalization of SfManIII. Alternatively, it is possible that the SfManIII gene was functionally inactivated by off-target effects associated with our constitutive CRISPR–Cas9 editing approach, as discussed previously.

We did not explore these possibilities because we considered that efforts to determine the reason(s) for the absence of SfManIII processing were beyond the scope of this study. In addition, we were happy to observe no evidence of SfManIII processing because this would have produced glycoproteins with mostly $Man_3GlcNAc_2$ *N*-glycans, which are Endo H resistant (15). Thus, SfManIII would have thwarted our effort to isolate an insect cell line designed to produce recombinant glycoproteins with Endo H-cleavable *N*-glycans. Instead, we found that Sf-RVN^{Lec1} cells have a Lec1 phenotype analogous to the Lec1 phenotype in mammalian cells, which means these insect cells can theoretically be used to produce recombinant glycoproteins with Endo H-cleavable *N*-glycans, just like their mammalian counterparts.

N-*glycosylation profile of recombinant glycoproteins produced by Sf-RVN^{Lec1} cells*

Subsequently, we extended the results described in the preceding section by assessing the ability of Sf-RVN^{Lec1} cells to produce recombinant glycoproteins with *N*-glycan structures expected to be Endo H cleavable. First, we produced the HIS-tagged form of recombinant hEPO in both Sf-RVN and Sf-RVN^{Lec1} cells and purified the secreted hEPO products by IMAC, as described in the Experimental procedures section. Analysis of the starting material, flow through, wash, and eluted fractions by SDS-PAGE with Coomassie blue staining indicated that the IMAC purification method was efficacious (Fig. S2). We then enzymatically released the *N*-glycans from samples of each purified hEPO preparation, purified the products, and profiled their structures by MALDI-TOF–MS, as described in the Experimental procedures section. We should note that we chose to simplify *N*-glycan profiling in this initial experiment by producing hEPO with AcRMD2-p6.9hEPO, which is a glycoengineered baculovirus vector designed to block glycan fucosylation (36). The *N*-glycan profiling results showed that the major *N*-glycan on the hEPO produced by Sf-RVN cells was the paucimannosidic structure $Man_3GlcNAc_2$, and there was a minor *N*-glycan with the hybrid structure $GlcNAcMan_3GlcNAc_2$ (Fig. 7*A*). In contrast, the major *N*-glycan on the hEPO produced by Sf-RVN^{Lec1} cells was the high-mannose structure $Man_5GlcNAc_2$, with a minor

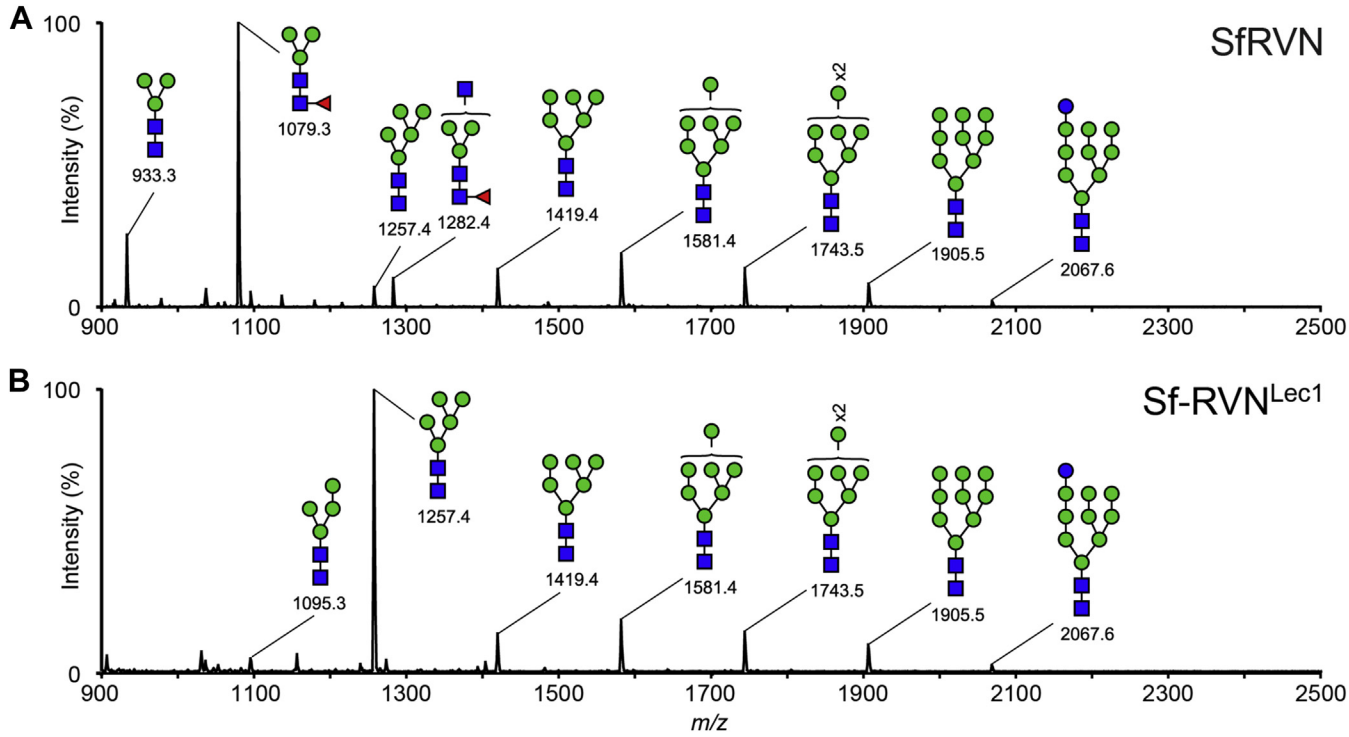

Figure 6. *N*-glycosylation profiles of endogenous insect cellular glycoproteins. Total glycoproteins were isolated from uninfected *A*, Sf-RVN or *B*, Sf-RVN^{Lec1} cells, and then total *N*-glycans were enzymatically released, purified, and analyzed by MALDI-TOF–MS, as described in the Experimental procedures section. All molecular ions were detected as $[M + Na]^+$, peaks with signal-to-noise ratios of >3 were assigned and annotated with the standard *cartoon symbolic* representations, and those were then presented as percentages of total annotated ion intensities. Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

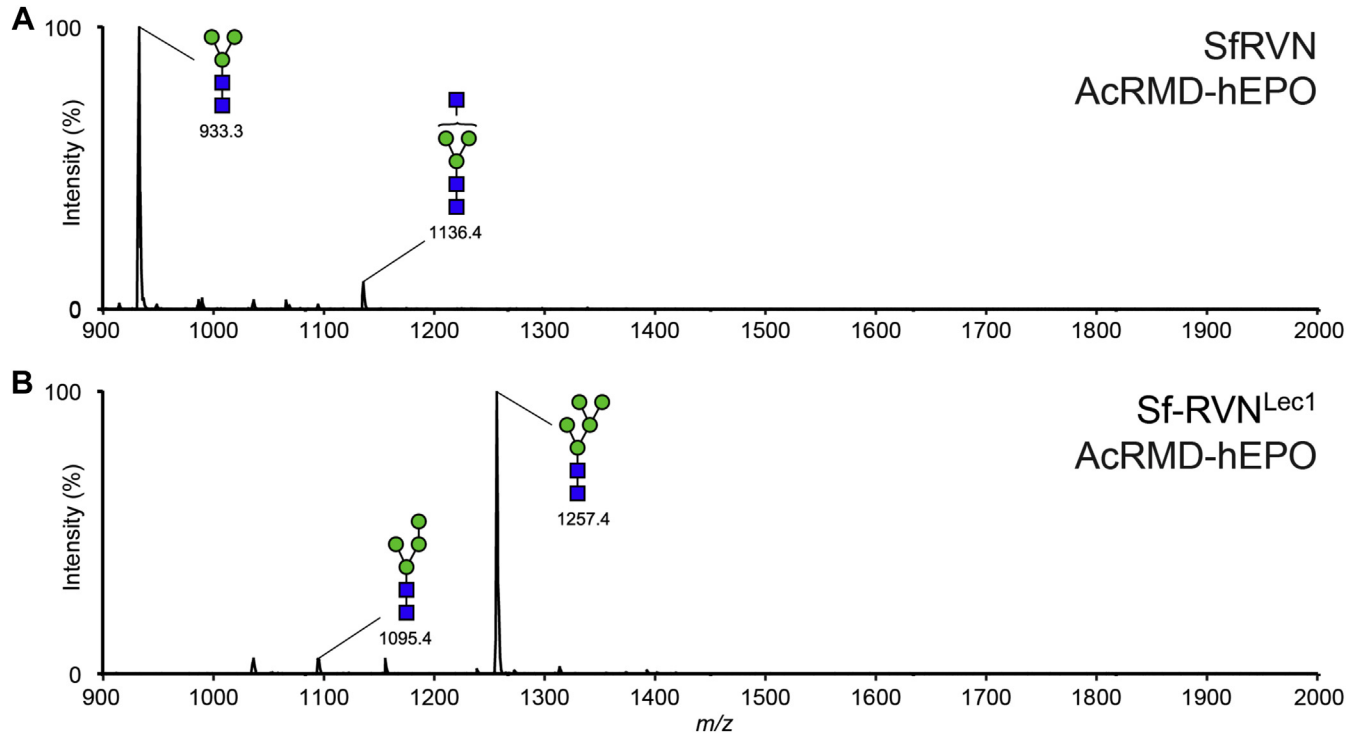

Figure 7. *N*-glycosylation profiles of recombinant hEPO. Recombinant forms of nonfucosylated hEPO were isolated from the cell-free media of *A*, Sf-RVN or *B*, Sf-RVN^{Lec1} cells infected with a glycoengineered baculovirus vector (AcRMD-hEPO) designed to block glycan fucosylation, and then total *N*-glycans were enzymatically released, purified, and analyzed by MALDI-TOF–MS, as described in the Experimental procedures section. hEPO, human erythropoietin; Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

subpopulation of $Man_4GlcNAc_2$ and no detectable $Man_3GlcNAc_2$ (Fig. 7*B*). These results clearly showed that Sf-RVN^{Lec1} cells can be used as an alternative host to produce at least one recombinant glycoprotein with high-mannose–type *N*-glycans in the BICS.

To validate the theoretical prediction that the high-mannose *N*-glycans on the recombinant hEPO expressed in Sf-RVN^{Lec1} cells would be Endo H cleavable, we treated samples of the purified hEPO preparations from both insect cell lines with PNGase-F or Endo Hf, as described in the Experimental procedures section. PNGase-F can cleave all *N*-glycans produced by Sf cells, as these cells produce no detectable core α3-fucosylated structures (36, 37), and Endo Hf is simply a commercially available recombinant form of Endo H. The SDS-PAGE and Coomassie blue staining results showed that Endo Hf treatment had no impact on the electrophoretic mobility of the hEPO from Sf-RVN cells, indicating it had Endo H-resistant *N*-glycans (Fig. 8*A*). In contrast, Endo Hf treatment increased the electrophoretic mobility of the hEPO from Sf-RVN^{Lec1} cells, indicating it had Endo H-cleavable *N*-glycans, as expected (Fig. 8*A*). The control PNGase-F treatments increased the electrophoretic mobility of the hEPO preparations from both cell lines (Fig. 8, *A* and *B*). Finally, a more sensitive Western blotting assay produced the same results, indicating the hEPO from Sf-RVN^{Lec1} cells had no detectable Endo H-resistant *N*-glycans (Fig. 8*B*). Thus, these results verified that Sf-RVN^{Lec1} can produce at least one recombinant glycoprotein with Endo H-cleavable *N*-glycans.

As noted previously, the results shown in Figure 8 were obtained using a HIS-tagged form of hEPO produced using a glycoengineered baculovirus vector designed to block glycan fucosylation in an effort to simplify the *N*-glycosylation profiles. In retrospect, we were concerned that blocking fucosylation could have had some additional and unexpected impact on the hEPO glycosylation patterns. Therefore, we repeated these experiments with a conventional and non-glycoengineered baculovirus vector to produce the HIS-tagged form of hEPO in Sf-RVN and Sf-RVN^{Lec1} cells. The results showed that the major *N*-glycan on hEPO produced by Sf-RVN cells was a fucosylated paucimannosidic structure (Fig. S3*A*), and the major *N*-glycan on the hEPO produced by Sf-RVN^{Lec1} cells was $Man_5GlcNAc_2$, which was mostly non-fucosylated, but accompanied by a small fucosylated subpopulation (Fig. S3*B*). We also validated the results of our initial endoglycosidase digestions using samples of the purified hEPO preparations produced using the nonglycoengineered baculovirus vector. The results showed that the *N*-glycans on the hEPO isolated from Sf-RVN^{Lec1} cells were Endo H cleavable, as expected (Fig. S3*C*). Thus, our initial use of the glycoengineered baculovirus vector to block hEPO fucosylation had no relevant impact on our results.

Finally, we extended the analytical results obtained using recombinant hEPO as a model glycoprotein by examining the *N*-glycosylation profiles and Endo H cleavability of human IgG1-Fc and mouse IgG2a-Fc produced in Sf-RVN^{Lec1} cells. In short, both these additional glycoproteins had *N*-glycosylation profiles and Endo H susceptibilities analogous to those observed with hEPO. That is, their major *N*-glycans had paucimannosidic structures ($Man_3GlcNAc_2$±Fuc) when produced in Sf-RVN cells (Fig. S4, *A* and *C*) and $Man_5GlcNAc_2$ when produced in Sf-RVN^{Lec1} cells (Fig. S4, *B* and *D*). Accordingly, the *N*-glycans on both these proteins also were Endo H cleavable when they were produced in Sf-RVN^{Lec1} cells (Fig. S5).

Conclusions

The major purpose of this study was to isolate an insect cell line with an *SfMgat1* knock out and Lec1 phenotype to provide crystallographers with a new tool to facilitate their ability to

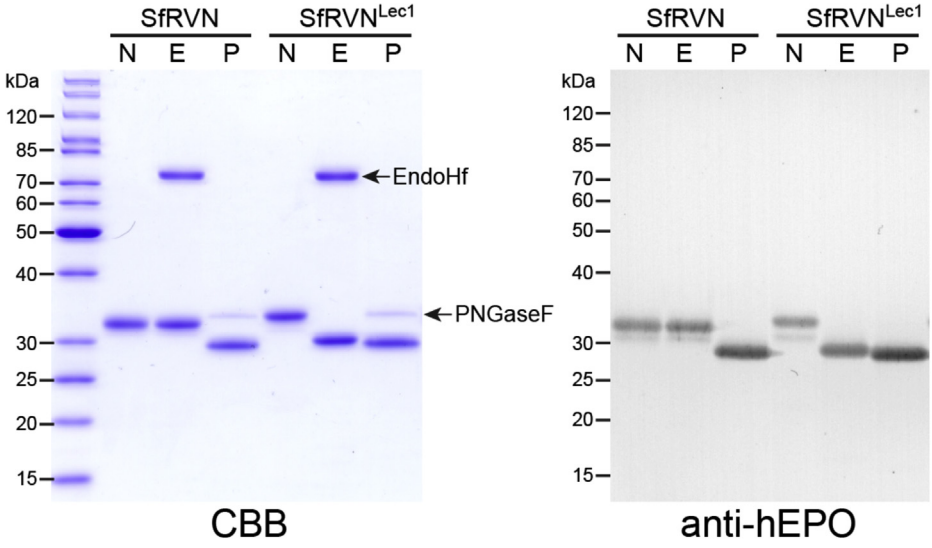

Figure 8. Endo H susceptibility of recombinant hEPO. Recombinant forms of nonfucosylated hEPO were isolated from the cell-free media of Sf-RVN or Sf-RVN^{Lec1} cells infected with AcRMD-hEPO, and then samples were treated with buffer alone (N), Endo Hf (E), or PNGase F (P) and analyzed by SDS-PAGE with Coomassie blue staining or Western blotting, as described in the Experimental procedures section. hEPO, human erythropoietin; Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line.

produce high-quality glycoprotein crystals for structural biology. While several mammalian cell lines with the Lec1 phenotype have been described (9, 21, 38–42) and used to facilitate recombinant glycoprotein crystallization, there has been no insect cell line that can be used for this purpose. The new Sf-RVN^{Lec1} cell line isolated in this study has multiple indels in the *SfMgat1* gene, grows and appears to produce recombinant glycoproteins as well as the parental Sf-RVN line, but it can only produce immature and high-mannose–type *N*-glycans that can be removed using Endo H. We expect that this new cell line will be a useful alternative host for researchers using the BICS platform to produce recombinant glycoproteins for structural biological work driven by X-ray crystallography.

Experimental procedures

Insect cells and insect cell culture

The parental insect cell line used in this study was Sf-RVN, which is an Sf-rhabdovirus–negative derivative of Sf9 cells (19). Polyclonal and monoclonal CRISPR–Cas9–edited derivatives were isolated as described later. Adherent insect cell cultures were grown at 28 °C in tight-capped T-flasks in TNM–FH medium supplemented with 10% fetal bovine serum and 0.1% pluronic F68. Suspension cultures were grown at 28 °C in loose-capped shake flasks rotated at 125 rpm in serum-free ESF 921 medium.

Plasmid constructions

Eight distinct derivatives of pIE1-Cas9-SfU6-sgRNA-Puro, a BICS-specific CRISPR–Cas9 vector (16), were constructed for this study. Each individual derivative had a different *Sfmgat1*-specific targeting sequence encoded by synthetic DNA primer pairs that were annealed, phosphorylated, and subcloned into the SapI site of this vector, as described previously (16). The sequences of the primers used to produce the eight different *Sfmgat1*-specific targeting sequences are given in Table S1.

CRISPR–Cas9 editing of SfRVN cells

Eight individual Sf-RVN cell cultures were transfected with one of the eight CRISPR–Cas9 constructs described previously using a calcium phosphate transfection method, as described previously (23). The transfected cells were selected in growth medium supplemented with puromycin for 7 days to produce eight distinct polyclonal transformed cell populations designated SfMgat1t1 to SfMgat1t8, which were passaged two times prior to single-cell cloning by end-point dilution.

Screening polyclonal transformed cell populations by electrophoretic mobility shift assays

Samples of each of the polyclonal transformed Sf-RVN cell populations obtained after editing with the eight distinct CRISPR–Cas9 vectors were infected with a previously described recombinant baculovirus vector encoding an *N*-terminally HIS-tagged version of hEPO (16). The cell-free media from each culture were harvested at 48 h postinfection, and then the secreted hEPO products were purified by IMAC, and samples were analyzed by SDS-PAGE with Coomassie blue staining, as described previously (16).

Isolating single-cell clones from SfMgat1t4

We subsequently used an end-point dilution method to isolate single-cell clones from the polyclonal SfMgat1t4 cell population, as described previously (23). Briefly, after diluting the cells to a level expected to produce one cell per well, on average, samples were seeded into 96-well plates, and wells containing single cells were identified by visual inspection under a phase contrast microscope. These wells were marked, the cells were allowed to grow in the 96-well plates for ∼2 weeks, and then the expanded single-cell clones were sequentially transferred to a 24-well plate, 6-well plate, 25 cm^2 culture flasks, 75 cm^2 culture flasks, and finally, the 75 cm^2 flasks were split 1:3 and the cells from those three confluent 75 cm^2 flasks were combined and used to seed 50 ml shake flasks.

Genotyping SfMgat1t4 cell clones by PCR and TIDE

Our initial genotypic analysis of the single-cell clones isolated from the SfMgat1t4 transformed polyclonal cell population involved PCR amplifying and sequencing the targeted region of the *Sfmgat1* gene (Fig. S1) in the parental Sf-RVN or SfMgat1t4 cell clones, and then using the TIDE program to deconvolute the results. Briefly, genomic DNA preparations were isolated from the parental cell line and samples from the first passage of all 11 SfMgat1t4 cell clones using the DirectPCR lysis reagent (Viagen Biotech). The approximate locations of the primers used for PCR and the size of the amplification product are shown in Fig. S1. The PCRs were performed using 2 μl of genomic DNA lysate as the template, Q5 DNA polymerase (New England Biolabs), and 0.2 mM dNTPs. The PCRs included 30 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 10 s, and extension of 72 °C for 30 s. The amplification products were purified on Omega Bio-Tek minispin columns, sequenced using the Sanger method (24) (GENEWIZ, Inc), and the results were analyzed using the TIDE algorithm (http://shinyapps.datacurators.nl/tide/) to predict the genome-edited ratios obtained in each SfMgat1t4 cell clone, as described previously (16, 25).

Genotyping SfMgat1t4#1 by PCR and cloning and sequencing amplimers

Additional genotyping of clone SfMgat1t4#1 was performed by PCR amplifying the targeted region of the *Sfmgat1* gene, cloning the resulting amplification product, and sequencing the inserts in 30 independent clones. Briefly, the targeted region of the genomic DNA isolated from passage 1 SfMgat1t4#1 cells was amplified as described in the preceding paragraph. The amplification product was then purified, as described previously, and cloned into the TA site of pGEM-T Easy (Promega). Finally, the inserts in 30 randomly picked independent clones were sequenced (GENEWIZ), and indels were identified in comparison to wildtype. Upon completion of

this genotyping analysis, we changed the name of this cell line from SfMgat1t4#1 to Sf-RVN^{Lec1}.

Growth curves

Sf-RVN and Sf-RVN^{Lec1} cells were seeded at 0.5 M cells/ml in ESF 921 in a 50 ml shake flask, and then, viable cell densities were measured at 24 h intervals for 1 week using a Countess automated cell counter (Thermo Fisher Scientific).

Recombinant hEPO expression and purification

An *N*-terminally HIS-tagged form of hEPO was produced in Sf-RVN and Sf-RVN^{Lec1} cells and purified by IMAC, as described previously (16). In this study, we actually used two different baculovirus vectors to produce hEPO in these two cell lines. One was Acp6.9-hEPO, which is a conventional vector encoding hEPO under the control of the baculovirus p6.9 promoter, and the other was AcRMD2p6.9-hEPO, which is the same vector in addition glycoengineered to block glycan fucosylation (36). The cell-free culture media from all hEPO production runs were collected at 48 h postinfection for subsequent IMAC, and fractions were examined for hEPO by SDS-PAGE with either Coomassie blue staining or Western blotting using polyclonal goat anti-hEPO (U-CyTech) as the primary antibody and alkaline phosphatase–conjugated polyclonal mouse antigoat IgG (MilliporeSigma) as the secondary antibody, as described previously (36).

Recombinant hIgG1-Fc and mIgG2a-Fc expression and purification

N-terminally HIS-tagged hIgG1-Fc and mIgG2a-Fc were produced in Sf-RVN and Sf-RVN^{Lec1} cells and purified by IMAC, as described previously (43). Briefly, the cells were infected with Acp6.9-hIgG1-Fc or Acp6.9-mIgG2a-Fc, which are conventional baculovirus vectors encoding the Fc fragments of either human IgG1 or mouse IgG2a under the control of the viral p6.9 promoter. Again, the cell-free culture media were collected at 48 h postinfection for subsequent IMAC, and fractions were assayed for human IgG1 or mouse IgG2a by SDS-PAGE with either Coomassie blue staining or Western blotting using alkaline phosphatase–conjugated polyclonal goat antimouse IgG or antihuman IgG Fc (MilliporeSigma), as described previously (43).

Endoglycosidase analyses

Samples of various purified glycoproteins were treated with Endo Hf (New England Biolabs) or PNGase F (New England Biolabs) at 37 °C overnight. The spent reactions were analyzed by SDS-PAGE on a 12% polyacrylamide gel, and then either stained with Coomassie blue or analyzed by Western blotting, as described previously.

MALDI-TOF MS analyses

MALDI-TOF MS analysis was performed as described previously (16). Briefly, 100 μg aliquots of a total cell protein fraction or purified glycoprotein were digested with PNGase F, then the enzymatically released *N*-glycans were purified and analyzed using an Applied Biosystems SCIEX TOF/TOF 5800 with 500 shots accumulated in reflectron positive-ion mode. The T2D raw data files produced by this instrument were converted to ASCII files, which were then uploaded to GlycoWorkbench (44) for data analysis. Molecular ions were detected as $[M + Na]^+$, and the structures of nondeisotoped peaks with signal-to-noise ratios of >3 were identified using the profiler tool and a customized *N*-glycan database in GlycoWorkBench with an accuracy of 0.2 Da. These peaks were annotated using the standard cartoon glycan symbolic representations, and annotated spectra were exported to graphical reports and finally opened in Adobe Illustrator to create our final figures, in which each peak was plotted as the percentage of total annotated ion intensities in the spectrum. Additional details of the conditions used for MALDI-TOF–MS analysis in this study are included with the raw data in the report deposited on GlycoPOST (see Data availability section).

Data availability

All data are included within the article except the raw MALDI-TOF–MS data, which were deposited to GlycoPOST (45) as part of a full project report (accession number: GPST000223).

Supporting information—This article contains supporting information.

Acknowledgments—We thank Dr Christoph Geisler for scientific advice over the course of this work and for the major contributions to the design, preparation, and submission of an Small Business Innovation Research phase I proposal that provided partial financial support. We thank Dr Steven Hallam, University of Wyoming, for organizing the raw MALDI data and technical information needed to upload this project to GlycoPOST. Finally, we thank Dr Asif Shajahan, University of Georgia Complex Carbohydrate Research Center, for assistance uploading this project to GlycoPOST. This work was supported by Award numbers GM102982 and GM136071 from the National Institutes of Health, Institute of General Medical Sciences. Acquisition of the Applied Biosystems SCIEX TOF/TOF 5800 was provided by Major Research Instrumentation Award CHE-1429615 from the National Science Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Institute of General Medical Sciences.

Author contributions—H. M.-A. and D. L. J. conceptualization; H. M.-A. methodology; H. M.-A. supervision; H. M.-A. and D. L. J. formal analysis; H. M.-A. and D. L. J. writing–original draft; H. M.-A. and D. L. J. writing–review & editing; D. L. J. funding acquisition; D. L. J. project administration.

Conflict of interest—As the President of GlycoBac, LLC, D. L. J. declares a potential conflict of interest. GlycoBac has trademarked and licensed the Sf-RVN cell line, which was mentioned in this article, as a commercial product. In addition, in the future, GlycoBac might choose to sell and/or license the Sf-RVN^{Lec1} subclone created in this study as a commercial product. H. M.-A. is a former consultant for GlycoBac but has moved on to another position, and,

since he has no financial interest in GlycoBac, H. M.-A. declares no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: BICS, baculovirus-insect cell system; Endo H, endo-β-*N*-acetylglucosaminidase H; hEPO, human erythropoietin; IMAC, immobilized metal affinity chromatography; MGAT1, β1,2-*N*-acetylglucosaminyltransferase I; Sf, *Spodoptera frugiperda*; SfManIII, Sf-mannosidase III; *Sfmgat1*, Sf-RVN mgat1; Sf-RVN, rhabdovirus-negative *Spodoptera frugiperda* cell line; sgRNA, single-guide RNA; TIDE, Tracking of Indels by DEcomposition.

References

1. Varki, A. (2017) Biological roles of glycans. *Glycobiology* **27**, 3–49
2. Aricescu, A. R., and Owens, R. J. (2013) Expression of recombinant glycoproteins in mammalian cells: Towards an integrative approach to structural biology. *Curr. Opin. Struct. Biol.* **23**, 345–356
3. Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J., and Davis, S. J. (2007) Glycoprotein structural genomics: Solving the glycosylation problem. *Structure* **15**, 267–273
4. Kamiya, Y., Satoh, T., and Kato, K. (2014) Recent advances in glycoprotein production for structural biology: Toward tailored design of glycoforms. *Curr. Opin. Struct. Biol.* **26**, 44–53
5. Columbus, L. (2015) Post-expression strategies for structural investigations of membrane proteins. *Curr. Opin. Struct. Biol.* **32**, 131–138
6. Sarkar, M., Hull, E., Nishikawa, Y., Simpson, R. J., Moritz, R. L., Dunn, R., and Schachter, H. (1991) Molecular cloning and expression of cDNA encoding the enzyme that controls conversion of high-mannose to hybrid and complex N-glycans: UDP-N-acetylglucosamine:alpha-3-D-mannoside ß-1,2-N-acetylglucosaminyltransferase I. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 234–238
7. Stanley, P. (2014) Mannosyl (alpha-1,3-)- glycoprotein beta-1,2-N-acetylglucosaminyltransferase (MGAT1). In: Taniguchi, N., Honke, K., Fukuda, M., Narimatsu, H., Yamaguchi, Y., Angata, T., eds. *Handbook of Glycosyltransferases and Related Genes*, Springer Japan, Tokyo: 183–194
8. Chen, W., and Stanley, P. (2003) Five Lec1 CHO cell mutants have distinct Mgat1 gene mutations that encode truncated N-acetylglucosaminyltransferase I. *Glycobiology* **13**, 43–50
9. Reeves, P. J., Callewaert, N., Contreras, R., and Khorana, H. G. (2002) Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13419–13424
10. Almo, S. C., Garforth, S. J., Hillerich, B. S., Love, J. D., Seidel, R. D., and Burley, S. K. (2013) Protein production from the structural genomics perspective: Achievements and future needs. *Curr. Opin. Struct. Biol.* **23**, 335–344
11. Geisler, C., Aumiller, J. J., and Jarvis, D. L. (2008) A fused lobes gene encodes the processing beta-*N*-acetylglucosaminidase in Sf9 cells. *J. Biol. Chem.* **283**, 11330–11339
12. Lu, D., Futterer, K., Korolev, S., Zheng, X., Tan, K., Waksman, G., and Sadler, J. E. (1999) Crystal structure of enteropeptidase light chain complexed with an analog of the trypsinogen activation peptide. *J. Mol. Biol.* **292**, 361–373
13. Matsuura, H., Kirschner, A. N., Longnecker, R., and Jardetzky, T. S. (2010) Crystal structure of the Epstein-Barr virus (EBV) glycoprotein H/glycoprotein L (gH/gL) complex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22641–22646
14. Altmann, F., Kornfeld, G., Dalik, T., Staudacher, E., and Glossl, J. (1993) Processing of asparagine-linked oligosaccharides in insect cells. *N*-acetylglucosaminyltransferase I and II activities in cultured lepidopteran cells. *Glycobiology* **3**, 619–625
15. Jarvis, D. L., and Summers, M. D. (1989) Glycosylation and secretion of human tissue plasminogen activator in recombinant baculovirus-infected insect cells. *Mol. Cell. Biol.* **9**, 214–223
16. Mabashi-Asazuma, H., and Jarvis, D. L. (2017) CRISPR-Cas9 vectors for genome editing and host engineering in the baculovirus-insect cell system. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9068–9073
17. Summers, M. D., and Smith, G. E. (1987) *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experiment Station, College Station, Texas
18. Smith, G. E., Summers, M. D., and Fraser, M. J. (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* **3**, 2156–2165
19. Maghodia, A. B., Geisler, C., and Jarvis, D. L. (2016) Characterization of an Sf-rhabdovirus-negative *Spodoptera frugiperda* cell line as an alternative host for recombinant protein production in the baculovirus-insect cell system. *Protein Expr. Purif.* **122**, 45–55
20. Geisler, C. (2018) A new approach for detecting adventitious viruses shows Sf-rhabdovirus-negative Sf-RVN cells are suitable for safe biologicals production. *BMC Biotechnol.* **18**, 8
21. Zhong, X., Cooley, C., Seth, N., Juo, Z. S., Presman, E., Resendes, N., Kumar, R., Allen, M., Mosyak, L., Stahl, M., Somers, W., and Kriz, R. (2012) Engineering novel Lec1 glycosylation mutants in CHO-DUKX cells: Molecular insights and effector modulation of N-acetylglucosaminyltransferase I. *Biotechnol. Bioeng.* **109**, 1723–1734
22. Gordon, R. D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C. A., Vizitiu, D., Withers, S. G., and Rini, J. M. (2006) X-ray crystal structures of rabbit *N*-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *J. Mol. Biol.* **360**, 67–79
23. Harrison, R. L., and Jarvis, D. L. (2016) Transforming lepidopteran insect cells for continuous recombinant protein expression. *Methods Mol. Biol.* **1350**, 329–348
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
25. Brinkman, E. K., Chen, T., Amendola, M., and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168
26. Jarman-Smith, R. F., Armstrong, S. J., Mannix, C. J., and Al-Rubeai, M. (2002) Chromosome instability in Spodoptera frugiperda Sf-9 cell line. *Biotechnol. Prog.* **18**, 623–628
27. Vaughn, J. L., Goodwin, R. H., Thompkins, G. J., and McCawley, P. (1977) The establishment of two insect cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro* **13**, 213–217
28. Ennis, T. J., and Sohi, S. S. (1976) Chromosomal characterisation of five lepidopteran cell lines of Malacosoma disstria (Lasiocampidae) and Christoneura fumiferana (Tortricidae). *Can. J. Genet. Cytol.* **18**, 471–477
29. Gerbal, M., Fournier, P., Barry, P., Mariller, M., Odier, F., Devauchelle, G., and Duonor-Cerutti, M. (2000) Adaptation of an insect cell line of Spodoptera frugiperda to grow at 37 degrees C: Characterization of an endodiploid clone. *In Vitro Cell. Dev. Biol. Anim.* **36**, 117–124
30. Drinnenberg, I. A., deYoung, D., Henikoff, S., and Malik, H. S. (2014) Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* **3**, e03676
31. Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H., and Jarvis, D. L. (2015) Modifying an insect cell *N*-glycan processing pathway using CRISPR-Cas technology. *ACS Chem. Biol.* **10**, 2199–2208
32. Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K. A., Weng, Z., and Zamore, P. D. (2018) The genome of the Hi5 germ cell line from Trichoplusia ni, an agricultural pest and novel model for small RNA biology. *Elife* **7**, e31628
33. Izumi, N., Shoji, K., Suzuki, Y., Katsuma, S., and Tomari, Y. (2020) Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature* **578**, 311–316
34. Jarvis, D. L., Bohlmeyer, D. A., Liao, Y. F., Lomax, K. K., Merkle, R. K., Weinkauf, C., and Moremen, K. W. (1997) Isolation and characterization of a class II alpha-mannosidase cDNA from lepidopteran insect cells. *Glycobiology* **7**, 113–127

35. Kawar, Z., Karaveg, K., Moremen, K. W., and Jarvis, D. L. (2001) Insect cells encode a class II alpha-mannosidase with unique properties. *J. Biol. Chem.* **276**, 16335–16340
36. Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H., and Jarvis, D. L. (2014) A novel baculovirus vector for the production of nonfucosylated recombinant glycoproteins in insect cells. *Glycobiology* **24**, 325–340
37. Staudacher, E., Kubelka, V., and Marz, L. (1992) Distinct N-glycan fucosylation potentials of three lepidopteran cell lines. *Eur. J. Biochem.* **207**, 987–993
38. Puthalakath, H., Burke, J., and Gleeson, P. A. (1996) Glycosylation defect in Lec1 Chinese hamster ovary mutant is due to a point mutation in N-acetylglucosaminyltransferase I gene. *J. Biol. Chem.* **271**, 27818–27822
39. Chen, W., Ünligil, U. M., Rini, J. M., and Stanley, P. (2001) Independent Lec1A CHO glycosylation mutants arise from point mutations in *N*-acetylglucosaminyltransferase I that reduce affinity for both substrates. Molecular consequences based on the crystal structure of GlcNAc-TI. *Biochemistry* **40**, 8765–8772
40. Sealover, N. R., Davis, A. M., Brooks, J. K., George, H. J., Kayser, K. J., and Lin, N. (2013) Engineering Chinese hamster ovary (CHO) cells for producing recombinant proteins with simple glycoforms by zinc-finger nuclease (ZFN)-mediated gene knockout of mannosyl (alpha-1,3-)-glycoprotein beta-1,2-*N*-acetylglucosaminyltransferase (Mgat1). *J. Biotechnol.* **167**, 24–32
41. Lee, J. S., Kallehauge, T. B., Pedersen, L. E., and Kildegaard, H. F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.* **5**, 8572
42. Stolfa, G., Mondal, N., Zhu, Y., Yu, X., Buffone, A., Jr., and Neelamegham, S. (2016) Using CRISPR-Cas9 to quantify the contributions of O-glycans, N-glycans and glycosphingolipids to human leukocyte-endothelium adhesion. *Sci. Rep.* **6**, 30392
43. Geisler, C., and Jarvis, D. L. (2012) Innovative use of a bacterial enzyme involved in sialic acid degradation to initiate sialic acid biosynthesis in glycoengineered insect cells. *Metab. Eng.* **14**, 642–652
44. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008) GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659
45. Watanabe, Y., Aoki-Kinoshita, K. F., Ishihama, Y., and Okuda, S. (2021) GlycoPOST realizes FAIR principles for glycomics mass spectrometry data. *Nucleic Acids Res.* **49**, D1523–D1528