Re:GEN Open
Volume 4, Number 1, 2024
DOI: 10.1089/regen.2023.0027

Re:GEN OPEN
Mary Ann Liebert, Inc. publishers

ORIGINAL RESEARCH

Predicting Potential miRNA Targets for Antibody Glycosylation in CHO Cells Using a Simplified Evaluation Workflow

Patrick Schlossbauer*, Florian Klingler, and Friedemann Hesse

Abstract

Introduction: In the past decade, microRNAs (miRNAs), which are small, noncoding RNAs of ~22 nucleotides length, have gained attention as novel engineering tools for biopharmaceutical cell line development. As miRNAs provide the ability to fine-tune the regulation of their targets, they offer attractive options for the development of differentially glycosylated monoclonal antibodies (mAbs) in production hosts like Chinese hamster ovary (CHO) cells. However, as one miRNA can potentially interact with several hundred gene transcripts, targeted miRNA mediated glycosylation regulation is complex to implement. This circumstance raises the need for computational aid in the prediction of miRNA targets.

Method: In our study, we present a workflow using the target prediction tool RNA22 in combination with a comprehensive dataset of sequences for CHO cells, followed by curation of prediction results.

Results: We created a knowledge-based database consisting of biochemically relevant genes for mAb *N*-glycosylation, to rationally process relevant results. Comparison to experimental data of target regulation unraveled the potential of our method to correctly predict 55 of 69 (79%) regulations caused by 16 different miRNAs known to affect mAb glycosylation.

Conclusion: This work could potentially serve as a starting point for the development of new bioinformatics-assisted workflows to select miRNAs serving the user s exact needs.

Keywords: CHO; microRNA; target prediction; glycosylation; engineering

Pages: 21–32

Introduction

To this date, mammalian expression hosts, especially Chinese hamster ovary (CHO) cells, are the most popular production host for biopharmaceutical production of recombinant proteins.[1,2] In CHO cell line development, the traditional gene knock-out or overexpression approaches can often just switch on or off a special phenotype or characteristic, but lack the ability to modulate regulatory effects.[3–5] Therefore, the demand for new engineering approaches filling this gap rises. In the past decade, microRNAs (miRNAs) have drawn increased attention, as they provide the ability to fine-tune the regulation of their targets.[3,4,6–9] miRNAs are short, single-stranded molecules, typically 20–25 base pairs long, which exhibit imperfect sequence complementarity to their respective messenger RNA (mRNA) target.[10–13] As a consequence, one miRNA can interact with up to hundreds of target genes rendering miRNA-mediated regulation complex, as several pathways can be simultaneously regulated by a single miRNA directly or indirectly.[14–17] Therefore, identifying the targets of miRNAs is a crucial step in understanding the biological functions of miRNAs.[18–20] However, experimental approaches for the identification of miRNA targets can be time-consuming, expensive, and have limited specificity. This circumstance raises the necessity for

Institute for Applied Biotechnology, University of Applied Sciences Biberach, Hubertus-Liebrecht-Str. 35, Biberach 88400, Germany

*Address correspondence to: *Patrick Schlossbauer, Institute for Applied Biotechnology, University of Applied Sciences Biberach, Patrick Schlossbauer, Hubertus-Liebrecht-Str. 35, 88400 Biberach, Germany*, Email: Schlossbauer@hochschule-bc.de

computational aid in the prediction of miRNA targets to narrow down potential effects of a miRNA of interest or to choose a miRNA based on the desired mode of action.[21] Bioinformatic approaches incorporate the use of algorithms and databases to analyze the sequence and structural features of miRNAs and their target mRNAs.[22–24] There are several bioinformatic tools and databases[25] available for predicting miRNA targets, such as TargetScan,[26] miRDB,[27–29] PicTar,[30] PITA, RNAhybrid,[31,32] DIANA,[33,34] or RNA22.[35] These tools use different algorithms to predict the binding sites of miRNAs on target mRNAs, based on sequence complementarity, thermodynamic stability, and site conservation and summarized context scores of the aforementioned parameters.[17,23,36–42] Bioinformatics-supported experimental design has several advantages compared to simple trial-and-error experimental validation. It is faster, cheaper, and more scalable than unguided large-scale experimental methods, and it can provide a systematic analysis of miRNA targets.[39] Moreover, bioinformatics-based prediction can identify potential miRNA targets that may be missed by experimental approaches. As it is not possible to experimentally test all potential targets of an miRNA, which would be all existing transcripts of an organism, bioinformatic target prediction is a prerequisite to rationally select a smaller subset of candidates for a validation.

However, it is important to note that bioinformatics-based miRNA target prediction is far from being perfect and can produce large numbers of false positive or false negative results depending on the algorithm and the parameter settings applied for the computation.[43] Furthermore, the method is also very much dependent on the databases used.[37,40] Therefore, it is essential to validate the highly plausible predicted targets using experimental methods such as miRNA transfection followed by luciferase assays, Western blotting, or quantitative reverse transcription polymerase chain reaction (qPCR).[23,41]

In our study, we present a workflow using the existing bioinformatic tool RNA22[35] in combination with a comprehensive dataset of sequences to perform target prediction of selected miRNAs. Manual curation of prediction results and comparison with experimental data of target regulation unraveled the potential and limitations of target prediction. The sequences of miRNAs with in-house-identified target mRNAs leading to a phenotypic change in antibody *N*-glycosylation from a previously conducted comprehensive miRNA screen were used as validation data for target prediction.[44] The targets of these screened miRNAs were predicted on the basis of a dataset containing all mRNA transcripts of the CHO cell line using different parameter settings including binding quality and thermodynamic stability. Followed by semi-automated filtering for targets highly relevant for glycosylation, the results were condensed and ranked by probability. These predictions were compared to the results of experiments where CHO cells were transiently transfected with the respective miRNAs and examined for the resulting target regulation via expression analysis. This work could serve as a starting point for the development of fine-tuned target prediction tools to select miRNAs that exactly provide specific metabolic functions.

Material and Methods

MiRNA target prediction

All target predictions in this work were conducted using the locally executable version of the published RNA22 algorithm (Version v2, released 2015) as this tool allows the user to submit batch requests of large custom databases.[35] Predictions of miRNAs were automatically executed against the National Center for Biotechnology Information (NCBI) publicly available transcriptome database for the organism *Cricetulus griseus* (Chinese hamster, RefSeq: GCF_000223135.1) containing ~46,000 mRNA transcripts and their transcript variants at the date of experiments,[45] as this database resembles the closest relative to the CHO cell line. The following target prediction parameters were tested: seed composition, minimum number of paired-up bases in the heteroduplex, and maximum folding energy for heteroduplexes.

Library of glycosylation relevant genes in CHO cells

A database containing the transcripts of 152 genes was established, which were classified as highly relevant for *N*-glycosylation. Relevant genes were collected from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the organism Chinese hamster (*C. griseus:* cge00510, cge00520, cge00051, cge00511, and cge00052).[46] Sequence data for the transcripts of these genes were obtained from the NCBI transcript database described earlier. The collected genes are mainly relevant for the sequential build-up or degradation of the *N*-glycan tree of the protein. Furthermore, sugar-nucleotide transporters[47] and sugar conversion enzymes are also present in the collection. The complete list of the selected glycosylation relevant genes can be found in Supplementary Table 1.

Filtering of relevant hits

Relevant hits were collected by comparing the complete output of the target prediction step to the glycosylation

database. This filtering step was performed in a semiautomated way using Microsoft Excel. The follow-up manual curation of the resulting condensed dataset included the following steps. Hits were ranked for binding probability based on their predicted thermodynamic stability and furthermore classified by their binding location on the mRNA transcript (5' untranslated region (UTR), coding sequence (CDS), and 3'UTR). Furthermore, a binding to all possible transcript variants of one gene was mandatory for a hit to be further processed.

MiRNA mimics and sequences

The experimental setup and the implementation of murine miRNA transfection into antibody-producing CHO cells are already described elsewhere.[44,48] In short, experimental validation was performed by transient transfection of miRNA mimics that were known to have an influence on glycosylation into a mAb-producing CHO-K1 cell line, followed by expression analysis of relevant genes using qPCR. The results of these experiments were used as a model dataset to test and validate our predictions. Sequences for the tested murine miRNAs were obtained from the manufacturer and were based on the miRBase release 21.[49] MiRNA sequences used in this study are listed in Supplementary Table 2.

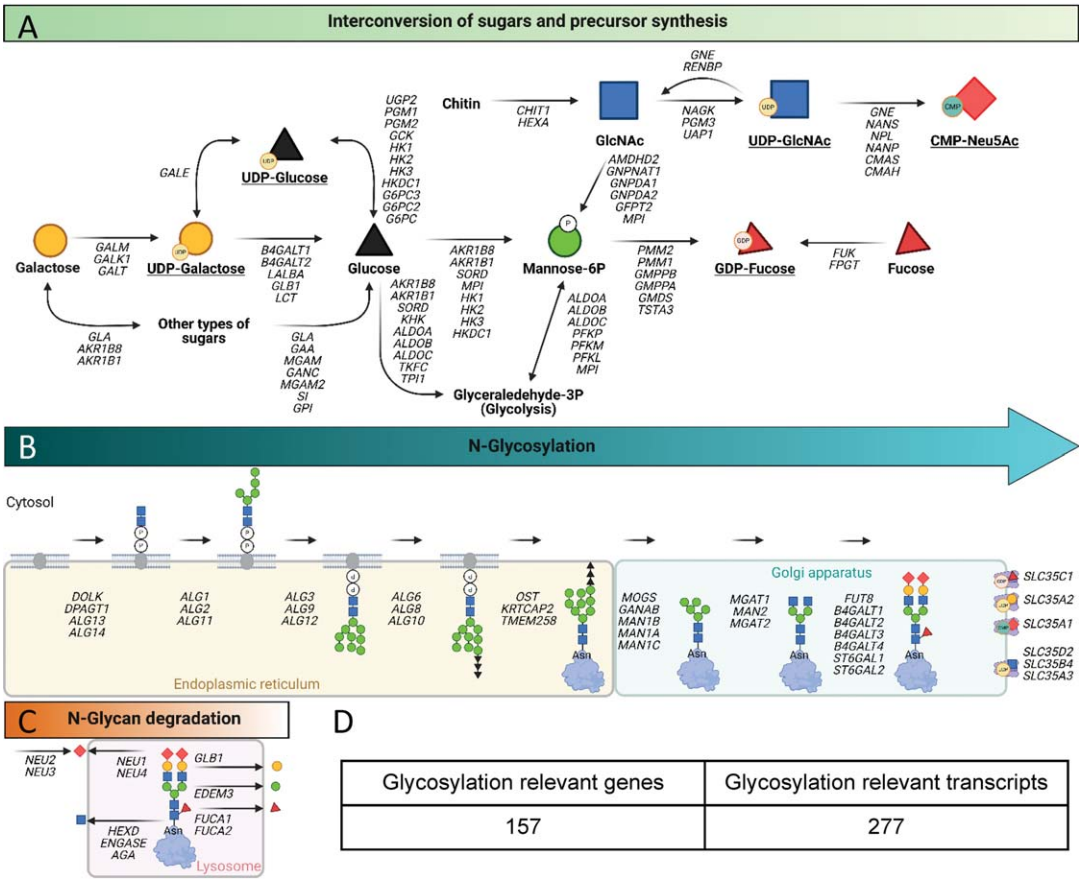
Results

Generation of a CHO glycosylation database

In this study, mAb glycosylation was used as phenotypic model readout caused by miRNA-mediated regulation. To narrow down the list of potentially interesting prediction hits, we generated a database consisting of highly relevant genes for protein *N*-glycosylation in CHO cells. This list consists of three main parts, which are the interconversion of sugar precursors, the following *N*-glycosylation of a mAb, and the degradation of the glycan tree. The first part of the list describes the interconversion of free galactose, glucose, mannose, fucose, *N*-acetylglucosamine (GlcNAc), and *N*-acetylneuraminic acid (Neu5Ac) into each other and toward the nucleotide sugars uridine diphosphate (UDP)-galactose, UDP-glucose, guanosine diphosphate (GDP)-fucose, UDP-GlcNAc, and cytidine-5′-monophosphate (CMP)-Neu5Ac (Fig. 1A). These nucleotide sugars reflect the active precursors, which are later attached to the glycan tree for *N*-glycosylation. The *N*-glycosylation database sums up all reactions, which lead from a membrane-bound dolichol to a complex glycan tree attached to a mAb (Fig. 1B). The assembly of the core structure consisting of mannose residues attached to dolichol takes place on the membrane of the endoplasmic reticulum, first directed into the cytosol, later on into the endoplasmic reticulum. After the transfer of the mannosylated glycan tree to the mAb in the endoplasmic reticulum, initial trimming of galactose and the mannose tree takes place. Subsequently, the addition of GlcNAc, galactose, Neu5Ac, and fucose takes place in the Golgi apparatus by their respective sugar transferases.[50,51] The nucleotide sugars are thus transported into the Golgi apparatus via solute carrier (SLC) transporters. As the cell also has the ability to degrade the glycan tree, the carbohydrate cleaving enzymes were also added to the database (Fig. 1C). They are located in the lysosome, the cytosol, or the extracellular space and cleave their specific carbohydrate moiety off the glycan tree to release the free sugar into the cytosol.[52] Overall, our database of highly relevant targets for *N*-glycosylation is composed of 157 genes with 277 transcript variants (Fig. 1D). This database represents our current knowledge of the biochemical pathways that lead to *N*-glycosylation and was used as a filter to rationally reduce the hits from the target prediction of a miRNA against the whole transcriptome database of CHO-K1, which typically yields thousands of predicted hits.

Parameter testing for initial hit generation

miRNA target prediction typically yields hundreds of results, mainly false positives.[43] The amount of generated data is highly dependent on the applied parameter settings, which are used for the prediction. Therefore, we initially aimed to evaluate the influence of the three main parameters *seed region*, *number of paired-up bases in the heteroduplex*, and *folding energy* on the raw number of generated hits. Therefore, we ran test predictions of mmu-miR-3096b-5p, a well-characterized miRNA from our previous study to regulate fucosylation, against 100 randomly chosen transcripts from our database for *N*-glycosylation and evaluated the number of raw hits. The RNA22 algorithm that was used in this study offered adjustability for all three mentioned parameters. The parameter *seed region* describes the composition of the seed region, which represents the initial recognition site of a miRNA and its respective mRNA target. The algorithm offers the settings 8mer, 7mer, and 6mer without mismatches (8mer0mm, 7mer0mm, and 6mer0mm), 8mer or 7mer with up to one mismatch (8mer1mm and 7mer1mm) or 8mer with up to two mismatches (8mer2mm). The resulting number of raw hits was highly dependent on these settings as we got only 16 hits with the setting 8mer0mm; however, as one or two additional mismatches were allowed in the same seed size, the number of hits increased to 49 and 83 for mmu-miR-3096b-5p predicted against the 100 randomly chosen transcripts, respectively. The same



Glycosylation relevant genes	Glycosylation relevant transcripts
157	277

FIG. 1. Establishment of a database of biochemically relevant genes for the *N*-glycosylation of a monoclonal antibody (mAb) in Chinese hamster ovary (CHO) cells. Created with BioRender.com. (A) The first set of genes describes the interconversion of the sugars galactose, glucose, mannose, fucose, *N*-acetylglucosamine (GlcNAc), and *N*-acetylneuraminic acid (Neu5Ac) into each other and their respective sugar nucleotides inside of the cell. (B) These genes describe the sequential build-up of the glycan tree on a mAb in the cytosol, endoplasmic reticulum, and Golgi apparatus. Furthermore, the respective transporters for the nucleotide sugars into the Golgi apparatus are also included. (C) These genes describe the degradation of the glycan tree on a mAb. (D) The overall numbers of genes and their resulting summarized transcript variants used for knowledge-based filtering of target prediction hits in this work.

tendency could be observed for the 7mer0mm compared to the 7mer1mm setting (Fig. 2A). The setting 8mer2mm was chosen as fixed setting for the optimization of the next parameters, to initially keep a high amount of information—potentially containing all targets that are currently known to lead to *N*-glycosylation—for the later implemented postprocessing based on biochemical knowledge. Next, the *number of paired-up bases* was tested, describing the number of bases in the miRNA–mRNA duplex, which must be paired. Here, a stringent decrease in the number of obtained hits could be observed with an increasing number of pairing bases. Most hits were obtained with 10 pairing bases resulting in 117 hits, while 19 paired bases resulted in only a single hit (Fig. 2B). As we wanted to carry on high numbers of initial hits before postprocessing, 10 paired-up bases were chosen as a fixed parameter for the last optimization step using the parameter *folding energy*. The *folding energy* describes the strength of the duplex binding, with lower values representing stronger binding. Here, similar to the previous parameter, a stringent correlation between the number of predicted hits and a lower folding energy (stronger binding) could be observed. The highest folding energy (−5 kcal/mol) yielded 302 potential hits, while the lowest folding energy that yielded a result was −29 kcal/mol with a single hit. Lower folding energies than −29 kcal/mol predicted no hits (Fig. 2C). The folding energy of −5 kcal/mol was chosen as the final setting to not lose potential hits with a weak duplex binding. Overall, the

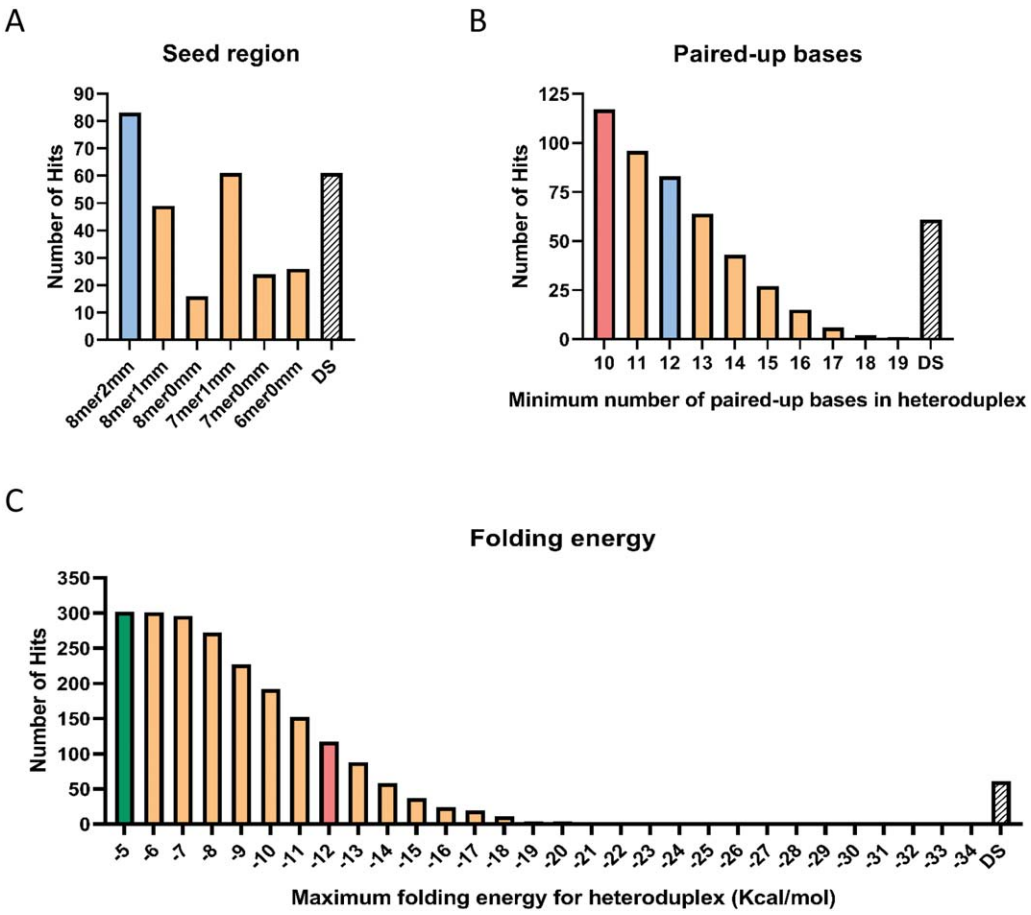


FIG. 2. Initial testing of the target prediction parameters seed region, paired-up bases, and folding energy on the number of resulting raw hits. The target prediction of the model mmu-miR-3096b-5p was conducted against 100 randomly chosen transcripts from a database with genes relevant for antibody *N*-glycosylation in Chinese hamster ovary (CHO) cells. For all predictions, the locally executable version of the RNA22 algorithm was used. (A) The target prediction parameter seed region describes the composition of the initial recognition site of a miRNA on an mRNA transcript. Permitted settings were as follows: 8mer, 7mer, and 6mer without mismatches (8mer0mm, 7mer0mm, and 6mer0mm), 8mer or 7mer, with up to one mismatch (8mer1mm, 7mer1mm) or 8mer with up to two mismatches (8mer2mm). (B) The total number of bases that pair in a miRNA–mRNA duplex is described by the parameter paired-up bases and is a quality attribute for strong binding. The tested settings were 10–19 paired bases. (C) The folding energy parameter is related to the two aforementioned parameters, as it describes the sum of the folding energy of all pairing bases in a miRNA–mRNA duplex. Its range is from −5 kcal/mol to −4 kcal/mol, where a lower binding energy represents stronger binding. The final settings for the first parameter are marked blue, the final settings from the second parameter are in red, and the final setting for all further experiments in green. DS = default setting of the algorithm (seed: 7mer1mm; 12 paired bases; −12 kcal/mol folding energy).

final setting for all further experiments was set to 8mer2mm, 10 paired-up bases, and a folding energy of −5 kcal/mol, to initially generate as many results as possible, which could then be narrowed down by further filtering steps. The same prediction was also conducted using the default settings (DSs) of the RNA22 algorithm, which were as follows: seed: 7mer1mm; 12 paired bases; −12 kcal/mol folding energy. Applying these more stringent parameter settings, the prediction only yielded 61 raw hits (Fig. 2A–C). Subsequently, the results for our less stringent parametrization were compared to the results obtained from default parametrization.

Filtering and processing of hits

The experimental validation of thousands of predicted hits is impossible and would yield high numbers of false

positive predictions depending on the algorithm used. Therefore, we used the above described database of highly glycosylation relevant genes as a filter, to narrow down the results based on biochemical knowledge of glycosylation to reasonable numbers of target genes, which could afterward be experimentally assayed. We selected 16 validated miRNAs from our previous studies to shed light on the filtering and processing of miRNA target predictions and their accuracy. Nine miRNAs showed in the initial experiments decreased fucosylation of a mAb,[44] five miRNAs led to decreased galactosylation, and two miRNAs increased the galactosylation on the mAb after transient transfection into CHO cells.[48] A target prediction of these 16 miRNAs was conducted against the complete CHO-K1 transcript collection using RNA22, yielding an initial set of raw data hits, which was processed stepwise (Fig. 3A). In a first step, if multiple transcript variants of one gene were targeted by a miRNA, these hits were combined, resulting in a reduced dataset (Dataset 1). This was further processed by filtering for glycosylation relevant genes by applying the collection of genes described above leading to Dataset 2. As the phenotypic change (change in fucosylation or galactosylation) was already known from our previous work with transiently transfected miRNAs in CHO cells, potential targets could be narrowed down to target genes highly relevant for the respective result, leading to Dataset 3. This last dataset was compared to the experimental data, which was obtained in our previous work via qPCR from CHO cells transfected with the respective miRNAs to either verify or contradict the predictions (Fig. 3B). The whole process of filtering reduced the number of hits from several thousands to only a few hits. In the case of the 16 tested miRNAs, raw data ranged between 1,610 hits for mmu-let7c-1-3p and 65,375 hits for mmu-miR-432, which, after filtering, could be narrowed down to 4–13 hits with a very high potential to be relevant for the respective glycosylation phenotype, depending on the miRNA (Fig. 3C). The combination of transcript variants reduced the raw data on average by 40.3% (Raw data &cenveo_unknown_entity_wingdings_F0E0; Dataset 1), the second filtering for glycosylation relevant genes showed the strongest reduction by 99.17% on average (Dataset 1 &cenveo_unknown_entity_wingdings_F0E0; Dataset 2), and the last reduction decreased the potential hits by another 79.06% on average (Dataset 2 &cenveo_unknown_entity_wingdings_F0E0; Dataset 3) (Fig. 3D). This filtering step enabled the rational testing and validation of predicted targets in a reasonable time via qPCR.

Validation of predicted miRNA targets

To evaluate the results of our previously conducted miRNA target prediction of 16 miRNAs with known phenotypic influences on the glycosylation of a mAb, we compared the generated Dataset 3 to qPCR data obtained in our previous studies,[44,48] applying default parametrization or our earlier described less stringent parametrization. Post-processing for both approaches was the identical. For many miRNAs, the cross-section of predicted and validated data showed overlapping results for the less stringent approach, and most of the experimentally observed regulations could be verified by a prediction (Fig. 4A). The 16 described miRNAs showed 69 confirmed (via qPCR) regulations in the assayed set of targets. Applying the DSs, only 33 of 69 (47%) observed regulations could be explained by a prediction (Fig. 4B). This number could be strongly increased by our less stringent target prediction, where 55 regulations could be explained (79%) (Fig. 4C). The 55 correctly predicted targets, however, represented 61% of total predictions from Dataset 3 for the less stringent settings, resulting in 39% false positive predictions (Fig. 4E). These rates are however similar for the DSs, where 32% of the predictions were false positive (Fig. 4D). These results demonstrated that by applying less stringent parameter settings, an additional 22 (32%) experimentally observed miRNA targets could be explained compared to DSs using our described workflow. This strong increase in prediction capacity was accompanied by only 7% increase in false positive results.

Discussion

In recent years, miRNAs have gained increased interest as engineering tools, as they allow for a more subtle regulation compared to standard approaches like knock-out or overexpression of genes.[3–5] They have already been successfully used in many engineering approaches in production cells, targeting titer, quality attributes or biomass-related parameters such as apoptosis.[3,6,8,53–66] As CHO cells are currently one of the most frequently used production hosts for therapeutic proteins and new engineering techniques apart from classical gene overexpression or knock-out are developed for biosimilar production or fine-tuning of phenotypic effects, the application of miRNAs in this host seems even more attractive.[2,9] However, the versatility of one miRNA to potentially regulate hundreds of genes due to its imperfect target binding also represents one of its major drawbacks when using them in a directed manner.[10,13,67–69] To understand the underlying mechanism behind an effect of a miRNA on the cell is crucial for the targeted development of miRNAs as a toolbox to fine-tune cellular behavior. Therefore, prediction of said targets is a key element to unravel the full potential of miRNAs as regulators in CHO production cells.[18,21,24,40] Our study provides

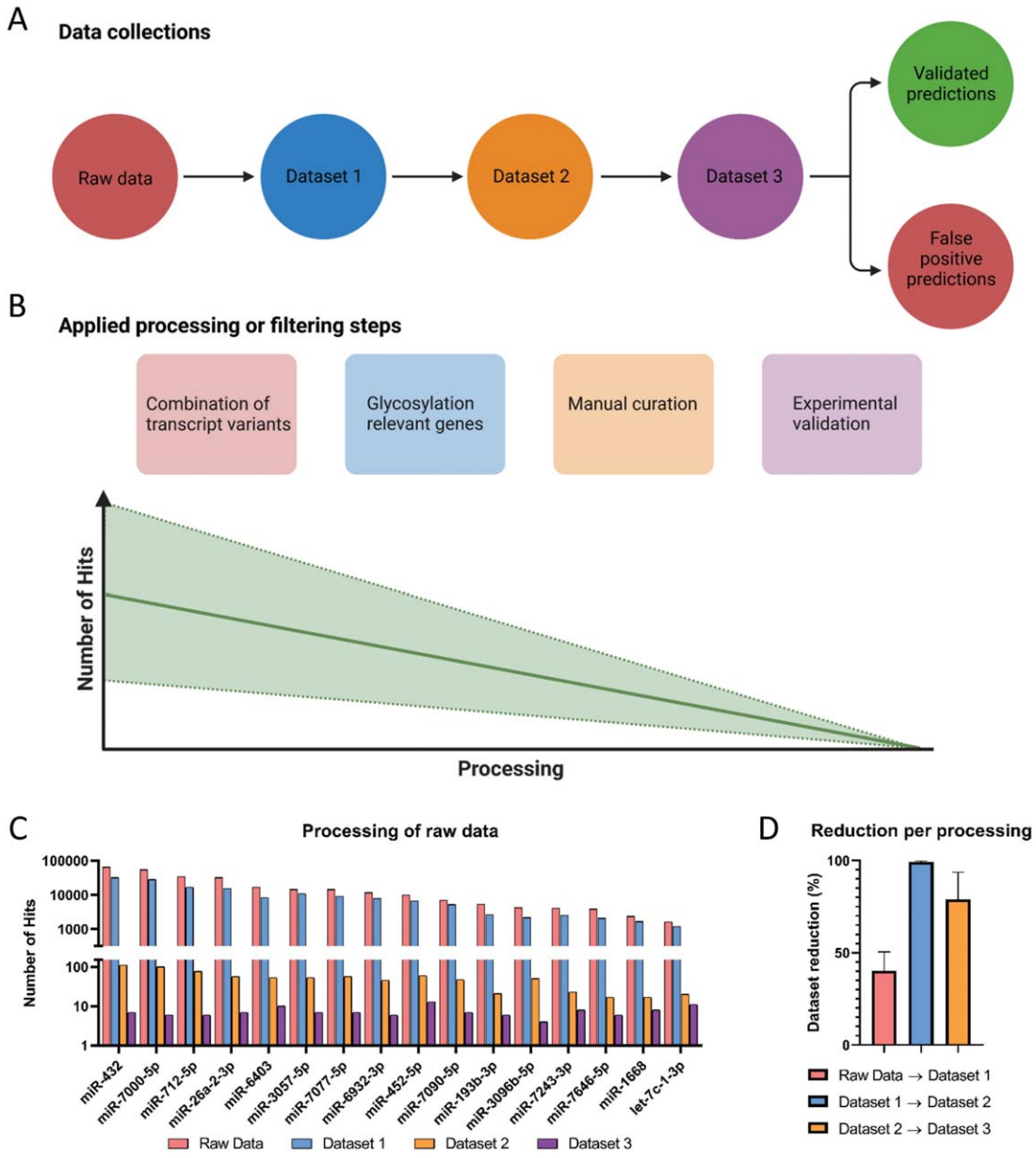


FIG. 3. Overview of the postprocessing, which was applied to target prediction hits in this study. (A) Based on the raw data from the prediction of a miRNA against the used transcriptome, three databases were processed sequentially, and the resulting selection of hits was compared to experimental data from transient miRNA transfections. (B) Schematic description of the reduction step, which is applied to each dataset stepwise to generate the following dataset. Created with BioRender.com. (C) 16 miRNAs known to regulate glycosylation in Chinese hamster ovary (CHO) cells were used as models in this study. Raw data and the number of hits for each of the above-mentioned datasets, after processing of raw data are shown. (D) Reduction rate of each processing step. Data are given as mean ± SD of all 16 tested miRNAs.

further insights into the impact of different target prediction parameters on the number of generated raw data and presents a simple workflow to handle large numbers of prediction results using a knowledge-based dataset of biochemically highly interesting genes combined with an experimentally observed phenotype. Furthermore, we finally compared the processed prediction data to experimental results obtained in previous studies.[44,48]

We decided to use the RNA22 algorithm,[35] as we wanted to use transcript data from CHO cells for the

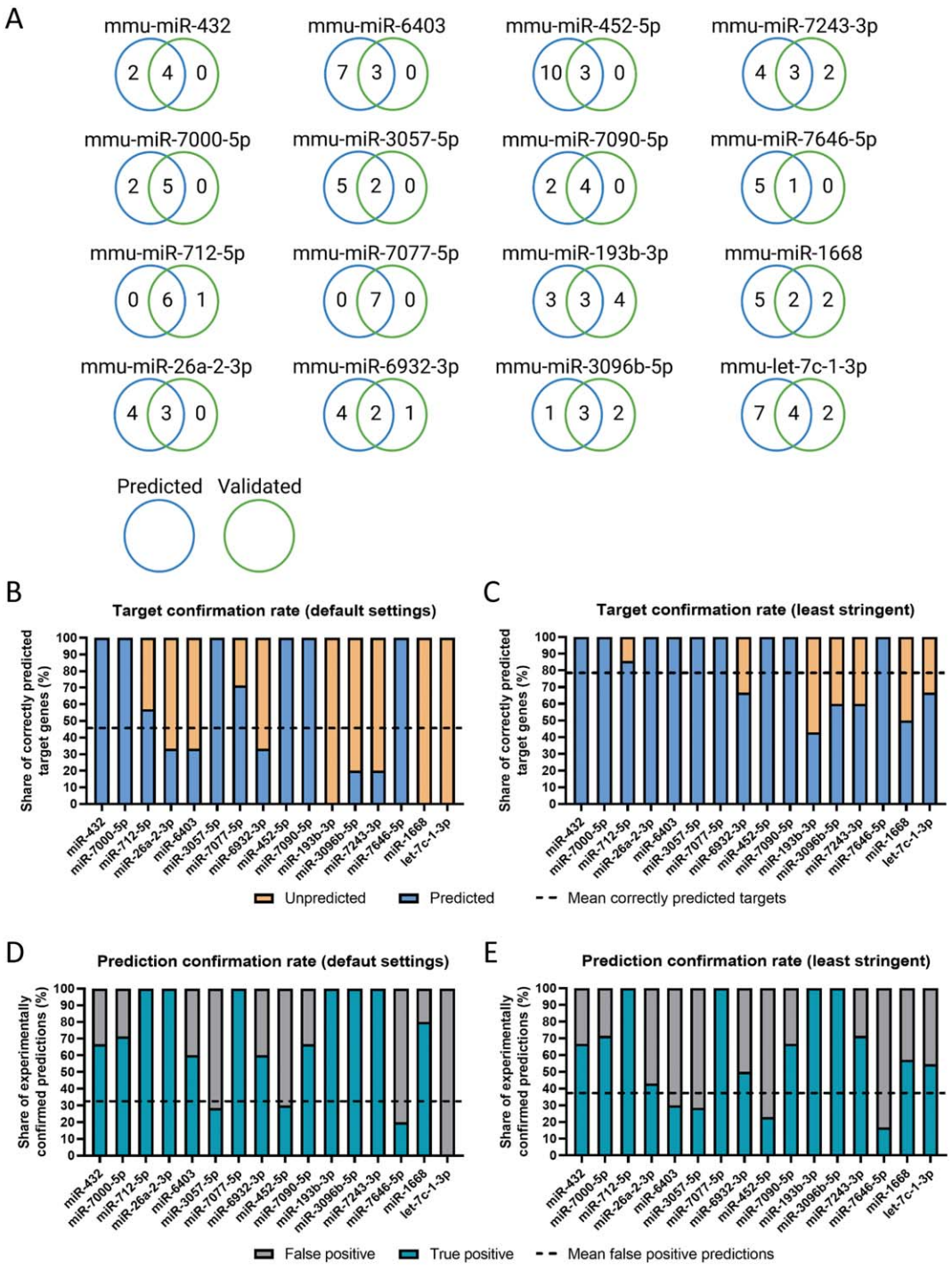


FIG. 4. Comparison of final target prediction hits after postprocessing for highly relevant targets to experimental data. (A) Overlap of target predictions using loose parameter stringency after the final selection step (blue) and confirmed regulations by the respective miRNA (green). The regulation of 16 model miRNAs was assayed in our previous work, using transcription analysis after transient transfection of antibody producing Chinese hamster ovary (CHO) cells with the respective miRNA. (B) For each miRNA, the target confirmation rate is shown for the default prediction settings and the (C) less stringent parameter settings. (D) For each miRNA, the prediction confirmation rate is shown for the default prediction settings and the (E) less stringent parameter settings. False positive describes the hits, which are predicted for the respective miRNA, but not confirmed experimentally.

predictions, which was not possible with other algorithms like Target Scan or miRDB. These algorithms typically only offer predictions for human, mouse, rat, or similar model organisms. Furthermore, RNA22 offered the advantage to perform target prediction against complete transcripts, not only 3'UTRs. In this study, we established a database containing the spectrum of known reactions, which are necessary for mAb *N*-glycosylation, glycan degradation, and precursor synthesis, as we used mAb glycosylation as a model readout. To our knowledge, this database represents a complete collection of the known enzymes in CHO for this subset of metabolic reactions. However, for the observed and assayed phenotypes—decrease in fucosylation, decrease in galactosylation, or increase in galactosylation—several enzymes of this collection were of special interest. For fucosylation, fucokinase (FUK), fucose-1-phosphate guanylyltransferase (FPGT), GDP-mannose 4,6-dehydratase (GMDS), GDP-L-fucose synthetase (TSTA3), GDP-L-fucose transporter (SLC35C1), fucosyltransferase 8 (FUT8), and fucosidase 1 (FUCA1) were identified as the most important enzymes.[44] The mentioned enzymes mark the key synthesis pathway for core-fucosylation on a mAb produced in CHO cells. A regulation of the genes FUK and FPGT, or GMDS and TSTA3 could block synthesis of the nucleotide sugar GDP-L-fucose, from either free fucose or mannose, and therefore lower fucosylation by depleting the precursor. Furthermore, as FUT8 is the only known fucosyltransferase in CHO to perform an α-1,6 linkage of fucose to the core of the glycan tree, this enzyme is a bottleneck for fucosylation, and a frequent target for miRNA regulation leading to the strongest afucosylation capacities, as shown previously. The same is true for the only validated fucose SLC35C1, where a downregulation would lead to lowered availability of GDP-L-fucose in the Golgi. A strong miRNA-mediated downregulation of these genes resulted in drastically decreased mAb fucosylation in CHO cells.[44] For galactosylation, galactose mutarotase (GALM), galactokinase (GALK1), galactose-1-phosphate uridylyltransferase (GALT), phosphoglucomutase-1 (PGM1), phosphoglucomutase-2 (PGM2), UDP-glucose pyrophosphorylase 2 (UGP2), UDP-galactose-4-epimerase (GALE), UDP-galactose transporter (SLC35A2), beta-1,4-galactosyltransferase 1–4 (B4GALT1, B4GALT2, B4GALT3, and B4GALT4), and beta-galactosidase-1 (GLB1) were considered critical.[48] Similar to fucosylation, a regulation of the enzymes that synthesize the precursor nucleotide sugar UPD-galactose (GALM, GALK1, GALT, PGM1, PGM2, UGP2, and GALE) potentially leads to a lack of precursors for terminal galactosylation. The galactosyltransferases B4GALT1-4 are similar to FUT8 the only known galactosyltransferases to catalyze a specific linkage in CHO cells, and a strong miRNA-mediated regulation of these targets could translate into a decreased linkage of UPD-galactose to the glycan and therefore a lower rate of galactosylation, as experimentally validated before.[48] Regulation of the transporter SLC35A2 could lower UDP-galactose levels in the Golgi, which is a prerequisite for efficient galactosylation. Estimation of parameters in a test prediction of an miRNA against 100 random transcripts revealed a great importance of the quality of the seed region, as the lowest amount of hits was detected, when allowing no mismatches in this region. With each additionally allowed mismatch, the number of hits increased by over 50%. When allowing no mismatches, shortening of the seed region from 8 to 7 or 6 nucleotides only added few of more results. This was due to the moveable position of the mismatch in the seed sequence, which allows more versatility in base-pairing. The importance of seed-pairing quality and parametrization was already described as critical quality factor elsewhere.[26,27,70] However, seed-pairing is not the only critical nucleotide-based parameter for miRNA–mRNA recognition.[71] Regarding the paired-up bases in the whole duplex, a linear decrease was expected and observed. These bases determine the overall stability of the duplex and the specificity of the miRNA toward its target.[70,71] Folding energy is somewhat dependent on paired-up bases, as each base pair lowers the folding energy of the duplex, thereby improving its thermodynamic stability.[19] For folding energy, a sigmoidal course of hits was observed with a plateau for very high and very low folding energies. Almost all predicted hits fulfilled the loose criteria for high folding energies between −5 kcal/mol and −7 kcal/mol as well as very few but strong duplexes, which reached very low folding energies between −19 kcal/mol and −29 kcal/mol. As we did not want to restrict our results by strict parametrization and thereby potentially lose true predictions, which simply do not exhibit a certain threshold, we decided to conduct all further experiments with the lowest possible parameter stringency. As the phenotypic outcome of the respective miRNA regulation for the 16 tested miRNAs was already known, we decided to rather generate a broad dataset with as much information as possible, which is afterward curated by a knowledge-based approach. As latter results showed, the more stringent default parameter settings restricted the prediction too much. When the prediction results of the 16 miRNAs against the whole transcriptome of CHO were first filtered by the glycosylation relevant genes and then manually by the above-mentioned key enzymes for the respective phenotype, the amount of results could be reduced from several thousands to ∼10 hits. This number was reasonable to be assayed via qPCR.

The greatest reduction in this processing took place, when the key enzymes from the glycosylation database were identified based on experimental knowledge from previous studies. This highlights the importance of experimental validation of the prediction results as already described elsewhere.[19,34,72,73] With the knowledge-based reduction of hits, based on previous results,[44,48] we could predict 79% of actual regulations. We could successfully demonstrate an increase in positive hits by 32% compared to default parametrization by applying low parametrization stringency. This result hints at a too stringent parametrization by the DSs and highlights the importance of variable parameter settings. Accompanied with the higher accuracy, the number of false positive hits increased only by 7%, as rational-based postprocessing of hits by our glycosylation database allowed for filtering. These numbers are highly dependent on the quality of the data processing and the availability of experimental data, as earlier studies have shown that target prediction in general can produce high numbers of false-positive results.[43] Also, we did not observe a pattern in incorrectly predicted targets in our study. Our data provide a first hint on how to improve the quality of prediction data but could be extended by an analysis of target site accessibility in mRNA,[22] by applying machine learning from a training data set for the optimization of the parameters[74] or by a combination of several target prediction algorithms.[75] Furthermore, the filtering workflow could be improved by a ranking for multiple bindings of a miRNA on the same transcript in spacial neighborhood, which can enhance target repression or lead to competitive binding of miRNAs.

In conclusion, our study demonstrates the influence of several target prediction parameters on the number of predicted targets and offers a workflow to filter the high number of data based on biochemically relevant targets. We could successfully evaluate results from our previous work using glycosylation as model readout and clearly demonstrate the importance of database quality and manual curation to pass on the right amount of data. In the end, we were able to explain 79% of actually occurring gene regulations with a prediction. This result is similar compared to 83% of confirmed binding sites that could be predicted by the developers of RNA22 in their work.[35] However, we highlight that our test dataset included 69 regulations, whereas the previous work only included 35 regulations. When investigating the number of predictions from our final dataset, which could be actually confirmed, our workflow produced 39% false predictions, which again highlights the issue of miRNA target prediction to produce false-positive results, which are estimated at 20–30% for most of the well-known prediction algorithms.[24] However, as previously hinted, dose sensitivity could play a role in such cases, as an allegedly false-positive miRNA prediction might actually be true when tested with a different miRNA concentration.[43] Our results show the potential and remaining challenge of target prediction-assisted experimental validation, which is of increasing interest for the implementation of miRNAs as engineering tools in bioprocessing relevant CHO cell lines. We want to highlight the relevance of biochemical knowledge, to interpret the high amount of data resulting from target prediction and to select highly plausible hit candidates for further experimental validation. This background knowledge also requires experiments in advance; however, many labs have generated valuable results over the past decades, which could be reevaluated and implemented in such databases, as recently demonstrated for glycosylation.[76] Efforts in future optimization strategies could therefore be made toward the generation of biochemical databases to accompany target prediction processing instead of changing the target prediction itself. Additionally, it must be mentioned that the addition of data from transient transfections is a valuable source and allows screening of large numbers of molecules. However, future models would benefit from data generated by stable miRNA overexpression, as this will be the goal of this technology, and discrepancies might arise between both modes of molecule delivery.

Contributions: F.H. supervised the project. P.S. and F.K. planned and conducted experiments. P.S. analyzed and evaluated data. P.S. wrote the manuscript. All authors read, approved the final manuscript, and provided critical feedback.

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgments

The authors would like to thank Rentschler Biopharma SE, represented by Dr. Dietmar Andreas Lang, for providing the CHO-K1 model cell line used in our previous work and for valuable discussion and support. Further thanks to Bruker Daltonik GmbH for providing analytical expertise. Special thanks to the Federal Ministry of Education and Research (BMBF), Germany, for funding of this study.

Grant Support: The Federal Ministry of Education and Research (BMBF), Germany, funded this study (Grand number: IP2018/722674).

References

1. Kim JY, Kim Y-G, Lee GM. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Appl Microbiol Biotechnol*. 2012; 93: 917–930. DOI: 10.1007/s00253-011-3758-5.

2. Walsh G, Walsh E. Biopharmaceutical benchmarks 2022. *Nat Biotechnol*. 2022; 40: 1722–1760. DOI: 10.1038/s41587-022-01582-x.
3. Fischer S, Handrick R, Otte K. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnol Adv*. 2015; 33: 1878–1896. DOI: 10.1016/j.biotechadv.2015.10.015.
4. Edwards E, Livanos M, Krueger A, et al. Strategies to control therapeutic antibody glycosylation during bioprocessing: Synthesis and separation. *Biotechnol Bioeng*. 2022; 119: 1343–1358. DOI: 10.1002/bit.28066.
5. Yamane-Ohnuki N, Kinoshita S, Inoue-Urakubo M, et al. Establishment of FUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol Bioeng*. 2004; 87: 614–622. DOI: 10.1002/bit.20151.
6. Fischer S, Marquart KF, Pieper LA, et al. miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. *Biotechnol Bioeng*. 2017; 114: 1495–1510. DOI: 10.1002/bit.26280.
7. Wang P, Zhou Y, Richards AM. Effective tools for RNA-derived therapeutics: SiRNA interference or miRNA mimicry. *Theranostics*. 2021; 11: 8771–8796. DOI: 10.7150/thno.62642.
8. Loh WP, Loo B, Zhou L, et al. Overexpression of microRNAs enhances recombinant protein production in Chinese hamster ovary cells. *Biotechnol J*. 2014; 9: 1140–1151. DOI: 10.1002/biot.201400050.
9. Hackl M, Borth N, Grillari J. miRNAs–pathway engineering of CHO cell factories that avoids translational burdening. *Trends Biotechnol*. 2012; 30: 405–406. DOI: 10.1016/j.tibtech.2012.05.002.
10. Bartel DP. MicroRNAs: Target recognition and regulatory functions. *Cell*. 2009; 136: 215–233. DOI: 10.1016/j.cell.2009.01.002.
11. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet*. 2010; 11: 597–610. DOI: 10.1038/nrg2843.
12. Gurtan AM, Sharp PA. The role of miRNAs in regulating gene expression networks. *J Mol Biol*. 2013; 425: 3582–3600. DOI: 10.1016/j.jmb.2013.03.007.
13. Brennecke J, Stark A, Russell RB, et al. Principles of microRNA-target recognition. *PLoS Biol*. 2005; 3: E85. DOI: 10.1371/journal.pbio.0030085.
14. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120: 15–20. DOI: 10.1016/j.cell.2004.12.035.
15. John B, Enright AJ, Aravin A, et al. Human microRNA targets. *PLoS Biol*. 2004; 2: E363. DOI: 10.1371/journal.pbio.0020363.
16. Grimson A, Farh KK-H, Johnston WK, et al. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol Cell*. 2007; 27: 91–105. DOI: 10.1016/j.molcel.2007.06.017.
17. Friedman RC, Farh KK-H, Burge CB, et al. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009; 19: 92–105. DOI: 10.1101/gr.082701.108.
18. Hammell M. Computational methods to identify miRNA targets. *Semin Cell Dev Biol*. 2010; 21: 738–744. DOI: 10.1016/j.semcdb.2010.01.004.
19. Kuhn DE, Martin MM, Feldman DS, et al. Experimental validation of miRNA targets. *Methods*. 2008; 44: 47–54. DOI: 10.1016/j.ymeth.2007.09.005.
20. Mazière P, Enright AJ. Prediction of microRNA targets. *Drug Discov Today*. 2007; 12: 452–458. DOI: 10.1016/j.drudis.2007.04.002.
21. Watanabe Y, Tomita M, Kanai A. Computational methods for microRNA target prediction. In *Methods in Enzymology: MicroRNA Methods*. Vol. 427. Cambridge, MA: Academic Press, 2007: 65–86.
22. Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007; 39: 1278–1284. DOI: 10.1038/ng2135.
23. Riolo G, Cantara S, Marzocchi C, et al. miRNA targets: From prediction tools to experimental validation. *Methods Protoc*. 2020; 4(1): 1. DOI: 10.3390/mps4010001.
24. Min H, Yoon S. Got target? Computational methods for microRNA target prediction and their extension. *Exp Mol Med*. 2010; 42: 233–244. DOI: 10.3858/emm.2010.42.4.032.
25. Kozomara A, Griffiths-Jones S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucl Acids Res*. 2014; 42: D68–D73. DOI: 10.1093/nar/gkt1181.
26. Agarwal V, Bell GW, Nam J-W, et al. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015; 4. DOI: 10.7554/eLife.05005.
27. Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol*. 2019; 20: 18. DOI: 10.1186/s13059-019-1629-z.
28. Chen Y, Wang X. miRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020; 48: D127–D131. DOI: 10.1093/nar/gkz757.
29. Wang X. miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA*. 2008; 14: 1012–1017. DOI: 10.1261/rna.965408.
30. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005; 37: 495–500. DOI: 10.1038/ng1536.
31. Rehmsmeier M, Steffen P, Höchsmann M, et al. Fast and effective prediction of microRNA/target duplexes. *RNA*. 2004; 10: 1507–1517. DOI: 10.1261/rna.5248604.
32. Krüger J, Rehmsmeier M. RNAhybrid: MicroRNA target prediction easy, fast and flexible. *Nucleic Acids Res*. 2006; 34: W451–W454. DOI: 10.1093/nar/gkl243.
33. Maragkakis M, Alexiou P, Papadopoulos GL, et al. Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinform*. 2009; 10: 295. DOI: 10.1186/1471-2105-10-295.
34. Kiriakidou M, Nelson PT, Kouranov A, et al. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev*. 2004; 18: 1165–1178. DOI: 10.1101/gad.1184704.
35. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006; 126: 1203–1217. DOI: 10.1016/j.cell.2006.07.031.
36. McGeary SE, Lin KS, Shi CY, et al. The biochemical basis of microRNA targeting efficacy. *Science*. 2019; 366. DOI: 10.1126/science.aav1741.
37. Yue D, Liu H, Huang Y. Survey of computational algorithms for microRNA target prediction. *Curr Genomics*. 2009; 10: 478–492. DOI: 10.2174/138920209789208219.
38. Rajewsky N. microRNA target predictions in animals. *Nat Genet*. 2006; 38: S8–S13. DOI: 10.1038/ng1798.
39. Saito T, Saetrom P. MicroRNAs—Targeting and target prediction. *New Biotechnol*. 2010; 27: 243–249. DOI: 10.1016/j.nbt.2010.02.016.
40. Hamzeiy H, Allmer J, Yousef M. Computational methods for microRNA target prediction. *Methods Mol Biol*. 2014; 1107: 207–221. DOI: 10.1007/978-1-62703-748-8_12.
41. Reyes-Herrera PH, Ficarra E. One decade of development and evolution of microRNA target prediction algorithms. *Genomics Proteomics Bioinform*. 2012; 10: 254–263. DOI: 10.1016/j.gpb.2012.10.001.
42. Peterson SM, Thompson JA, Ufkin ML, et al. Common features of microRNA target prediction tools. *Front Genet*. 2014; 5: 23. DOI: 10.3389/fgene.2014.00023.
43. Pinzón N, Li B, Martinez L, et al. MicroRNA target prediction programs predict many false positives. *Genome Res*. 2017; 27: 234–245. DOI: 10.1101/gr.205146.116.
44. Klingler F, Naumann L, Schlossbauer P, et al. A novel system for glycosylation engineering by natural and artificial miRNAs. *Metab Eng*. 2023; 77: 53–63. DOI: 10.1016/j.ymben.2023.03.004.
45. National Center for Biotechnology Information: 2022. Available at https://www.ncbi.nlm.nih.gov/. Accessed April 13, 2022.
46. Kegg Genome: *Cricetulus griseus* (Chinese hamster). 2023. Available at https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=cge. Accessed February 22, 2023.
47. Maszczak-Seneczko D, Wiktor M, Skurska E, et al. Delivery of nucleotide sugars to the mammalian golgi: A very well (un)explained story. *Int J Mol Sci*. 2022; 23: 8648. DOI: 10.3390/ijms23158648.
48. Klingler F, Schlossbauer P, Naumann L, et al. Developing microRNAs as engineering tools to modulate monoclonal antibody galactosylation. *Biotechnol Bioeng*. 2023; 121: 1355–1365. DOI: 10.1002/bit.28616.
49. miRBase: 2023. Available at https://www.mirbase.org/ftp.shtml. Accessed February 22, 2023.
50. Fisher P, Thomas-Oates J, Wood AJ, et al. The N-glycosylation processing potential of the mammalian golgi apparatus. *Front Cell Dev Biol*. 2019; 7(157). DOI: 10.3389/fcell.2019.00157.
51. Schoberer J, Shin Y-J, Vavra U, et al. Analysis of protein glycosylation in the ER. *Methods Mol Biol*. 2018; 1691: 205–222. DOI: 10.1007/978-1-4939-7389-7_16.
52. Varki A, Cummings RD, Esko JD, et al. *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2015.
53. Müller D, Katinger H, Grillari J. MicroRNAs as targets for engineering of CHO cell factories. *Trends Biotechnol*. 2008; 26: 359–365. DOI: 10.1016/j.tibtech.2008.03.010.
54. Clarke C, Henry M, Doolan P, et al. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in

controlling CHO cell growth rate. *BMC Genomics*. 2012; 13: 656. DOI: 10.1186/1471-2164-13-656.
55. Jadhav V, Hackl M, Bort JAH, et al. A screening method to assess biological effects of microRNA overexpression in Chinese hamster ovary cells. *Biotechnol Bioeng*. 2012; 109: 1376–1385. DOI: 10.1002/bit.24490.
56. Jadhav V, Hackl M, Klanert G, et al. Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *J Biotechnol*. 2014; 175: 38–44. DOI: 10.1016/j.jbiotec.2014.01.032.
57. Fischer S, Buck T, Wagner A, et al. A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnol J*. 2014; 9: 1279–1292. DOI: 10.1002/biot.201400306.
58. Strotbek M, Florin L, Koenitzer J, et al. Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells. *Metab Eng*. 2013; 20: 157–166. DOI: 10.1016/j.ymben.2013.10.005.
59. Fischer S, Mathias S, Schaz S, et al. Enhanced protein production by microRNA-30 family in CHO cells is mediated by the modulation of the ubiquitin pathway. *J Biotechnol*. 2015; 212: 32–43. DOI: 10.1016/j.jbiotec.2015.08.002.
60. Leroux A-C, Bartels E, Winter L, et al. Transferability of miRNA-technology to bioprocessing: Influence of cultivation mode and media. *Biotechnol Prog*. 2021; 37: E3107. DOI: 10.1002/btpr.3107.
61. Stiefel F, Fischer S, Sczyrba A, et al. miRNA profiling of high, low and non-producing CHO cells during biphasic fed-batch cultivation reveals process relevant targets for host cell engineering. *J Biotechnol*. 2016; 225: 31–43. DOI: 10.1016/j.jbiotec.2016.03.028.
62. Loh WP, Yang Y, Lam KP. miR-92a enhances recombinant protein productivity in CHO cells by increasing intracellular cholesterol levels. *Biotechnol J*. 2017; 12. DOI: 10.1002/biot.201600488.
63. Liu H-N, Dong W-H, Lin Y, et al. The effect of microRNA on the production of recombinant protein in CHO cells and its mechanism. *Front Bioeng Biotechnol*. 2022; 10(832065). DOI: 10.3389/fbioe.2022.832065.
64. Mori K, Kuni-Kamochi R, Yamane-Ohnuki N, et al. Engineering Chinese hamster ovary cells to maximize effector function of produced antibodies using FUT8 siRNA. *Biotechnol Bioeng*. 2004; 88: 901–908. DOI: 10.1002/bit.20326.
65. Raymond C, Robotham A, Spearman M, et al. Production of α2,6-sialylated IgG1 in CHO cells. *MAbs*. 2015; 7: 571–583. DOI: 10.1080/19420862.2015.1029215.
66. Weis BL, Guth N, Fischer S, et al. Stable miRNA overexpression in human CAP cells: Engineering alternative production systems for advanced manufacturing of biologics using miR-136 and miR-3074. *Biotechnol Bioeng*. 2018; 115: 2027–2038. DOI: 10.1002/bit.26715.
67. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nat Rev Genet*. 2008; 9: 102–114. DOI: 10.1038/nrg2290.
68. Bartel DP, Chen C-Z. Micromanagers of gene expression: The potentially widespread influence of metazoan microRNAs. *Nat Rev Genet*. 2004; 5: 396–400. DOI: 10.1038/nrg1328.
69. Chi SW, Hannon GJ, Darnell RB. An alternative mode of microRNA target recognition. *Nat Struct Mol Biol*. 2012; 19: 321–327. DOI: 10.1038/nsmb.2230.
70. Seok H, Ham J, Jang E-S, et al. MicroRNA target recognition: Insights from transcriptome-wide non-canonical interactions. *Mol Cells*. 2016; 39: 375–381. DOI: 10.14348/molcells.2016.0013.
71. Chipman LB, Pasquinelli AE. miRNA targeting: Growing beyond the seed. *Trends Genet*. 2019; 35: 215–222. DOI: 10.1016/j.tig.2018.12.005.
72. Ritchie W, Rasko JEJ, Flamant S. MicroRNA target prediction and validation. In *MicroRNA Cancer Regulation*. Dordrecht: Springer, 2013: 39–53.
73. Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res*. 2006; 34: 1646–1652. DOI: 10.1093/nar/gkl068.
74. Liu H, Yue D, Chen Y, et al. Improving performance of mammalian microRNA target prediction. *BMC Bioinform*. 2010; 11: 476. DOI: 10.1186/1471-2105-11-476.
75. Oliveira AC, Bovolenta LA, Nachtigall PG, et al. Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Front Genet*. 2017; 8(59). DOI: 10.3389/fgene.2017.00059.
76. Kotidis P, Donini R, Arnsdorf J, et al. CHOGlycoNET: Comprehensive glycosylation reaction network for CHO cells. *Metab Eng*. 2023; 76: 87–96. DOI: 10.1016/j.ymben.2022.12.009.

Abbreviations

B4GALT1-4	beta-1,4-galactosyltransferase 1–4
CDS	coding sequence
CHO	Chinese hamster ovary
CMP	cytidine-5′-monophosphate
FPGT	fucose-1-phosphate guanylyltransferase
FUCA1	fucosidase 1
FUK	fucokinase
FUT8	fucosyltransferase 8
GALE	UDP-galactose-4-epimerase
GALK1	galactokinase
GALM	galactose mutarotase
GALT	galactose-1-phosphate uridylyltransferase
GDP	guanosine diphosphate
GLB1	beta-galactosidase-1
GlcNAc	*N*-Acetylglucosamine
GMDS	GDP-mannose 4,6-dehydratase
KEGG	Kyoto Encyclopedia of Genes and Genomes
mAb/mAbs	monoclonal antibody/antibodies
miRNA	microRNA
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
Neu5Ac	*N*-Acetylneuraminic acid
PGM1	phosphoglucomutase-1
PGM2	phosphoglucomutase-2
qPCR	quantitative reverse transcription polymerase chain reaction
TSTA3	GDP-l-fucose synthetase
SLC35A2	UDP-galactose transporter
SLC35C1	GDP-l-fucose transporter
UDP	uridine diphosphate
UGP2	UDP-glucose pyrophosphorylase 2
UTR	untranslated region