Quantification of Glycan in Glycoproteins via AUCAgent-enhanced Analytical Ultracentrifugation

Xiaojuan Yu[1], Zhaoxing Wang[2,3], Chengshi Zeng[3], Ruifeng Zhang[2,3], Qing Chang[2,3], Wendan Chu[2,3], Qinghua Ma[4], Ke Ma[4], Lan Wang[1]✉, Chuanfei Yu[1]✉ & Wenqi Li[2,3,5]✉

Abstract

As essential biomolecules composed of proteins and carbohydrate moieties, glycoproteins play pivotal roles in numerous biological processes. The glycosylation level plays a crucial role in determining the functionality of glycoproteins. Therefore, the precise quantification of glycan components in proteins holds significant importance for research on and development of polysaccharide-protein-conjugated vaccines. In this study, a novel glycan quantification approach was developed, leveraging analytical ultracentrifugation (AUC) technology that synergistically utilizes ultraviolet wavelength absorption and interference data to directly determine glycan mass fractions in glycoproteins. This methodology expands the analytical framework for glycoproteins while retaining the intrinsic advantages of AUC, enabling analysis in native states with demonstrated robustness. The approach was implemented in our proprietary AUC data analysis software called AUCAgent.

Keywords: Glycoprotein, Quantification, Analytical ultracentrifugation, AUCAgent

Introduction

Glycosylation modifications of proteins are ubiquitous across all domains of life. These modifications provide proteins with diverse properties that critically influence

[1]National Institutes for Food and Drug Control, State Key Laboratory of Drug Regulatory Science, NHC Key Laboratory of Research on Quality and Standardization of Biotech Products, NMPA Key Laboratory for Quality Research and Evaluation of Biological Products, Beijing, China. [2]Core Facility for Biomolecule Preparation and Characterization, Technology Center for Protein Sciences, Tsinghua University, Beijing, China. [3]Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing, China. [4]Beijing Zhifei Lvzhu Biopharmaceutical Co., Ltd.; Beijing Economic Development Zone, Beijing, China. [5]State Key Laboratory of Membrane Biology, Membrane Structure and Artificisl intelligence Biology Branch, Hangzhou, China. These authors contributed equally: Xiaojuan Yu, Zhaoxing Wang, Chengshi Zeng.

✉ e-mail: wanglan@nifdc.org.cn; yucf@nifdc.org.cn; liwenqi@mail.tsinghua.edu.cn.

their function[1–3] and are a key mechanism for regulating multiple cellular processes, including signal transduction, protein folding, localization, stability, cell-cell interactions, virus-cell recognition, and host immune responses[4,5]. Glycosylation-mediated diversity plays a central role in glycoprotein biosynthesis and biological activity involved in antigen recognition. Notably, viruses exploit certain advantages conferred by glycosylation to assemble their own envelope glycoproteins through host cell glycosylation mechanisms, thereby evading immune surveillance[5,6]. Conversely, biopharmaceutical development has leveraged the ability of glycan structures to specifically recognize various receptors on cell surfaces. Glycosylated therapeutics are increasingly emerging as a favored and valuable therapeutic strategy due to their enhanced targeting specificity and reduced adverse effects as compared to small molecule drugs[7–9].

Many protein drugs possess glycosylation modifications, and the structure of the attached glycan chains significantly affects drug efficacy. The quantity and composition of these glycan chains regulate protein folding, solubility, and intracellular transport processes[10], and result in an extremely broad molecular weight(MW) distribution range in glycoproteins[1]. The average MW of glycan chains directly determines many physicochemical properties, including density, viscosity, diffusivity, sedimentation coefficient, electrophoretic mobility, and specific heat capacity. It also influences protein oligomerization states and the actual mass concentration determined by ultraviolet (UV) absorption methods[1,9,11,12].These properties influence the biological activity and therapeutic efficacy of glycoprotein drugs. Therefore, the accurate determination of glycoprotein MW has become an indispensable parameter in drug production processes. These measurements are also crucial for comparing batch-to-batch consistency and assessing post-translational modification levels[1,11].

Currently, the determination of protein MW primarily employs techniques such as sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), size-exclusion chromatography (SEC), multi-angle light scattering (MALS), analytical ultracentrifugation (AUC), and mass spectrometry (MS). SDS-PAGE is a classic electrophoresis-based method that offers advantages of simplicity and low cost. However, SDS-PAGE accuracy is susceptible to post-translational modifications, such as glycosylation, often leading to MW estimation errors due to migration anomalies[1,13,14]. SEC estimates MW based on protein elution behavior in solution,

typically requiring calibration with standards. However, since its response signal depends on the protein's hydrodynamic radius rather than absolute mass, measurements for non-spherical structures or glycoproteins (where glycan chains significantly increase the hydrodynamic radius) often deviate from the true MW[1,15–17]. MALS can directly determine the absolute MW of proteins in solution and effectively assess sample homogeneity without requiring calibration with standards. however, it is typically used with separation techniques such as SEC or AF4[15]. As an increasingly important analytical tool, AUC is widely regarded as the gold standard technique for evaluating the molecular integrity and non-aggregated state of monoclonal antibodies. AUC can determine MW in solutions approximating natural conditions while simultaneously obtaining multiple other parameters, such as conformation and the sedimentation coefficient[18,19].

However, all the above techniques have certain limitations in determining glycoprotein MW. SDS-PAGE may exhibit MW deviations due to migration abnormalities caused by glycosylation[1,13] and SEC may overestimate actual MW due to the effect of glycan chains on the overall hydrodynamic radius[1,16,17]. Currently, MALS combined with UV detection and differential refractive index detection enables the precise MW resolution of both glycan and protein components within glycoprotein complexes, making it a highly accurate method[20]. Conventional AUC detection can utilize dual signals from UV-visible light and interferometric detectors for analysis, a technique already applied in membrane protein complex studies[21]. Consequently, dual-detector AUC analysis holds significant potential for glycoprotein complex characterization.

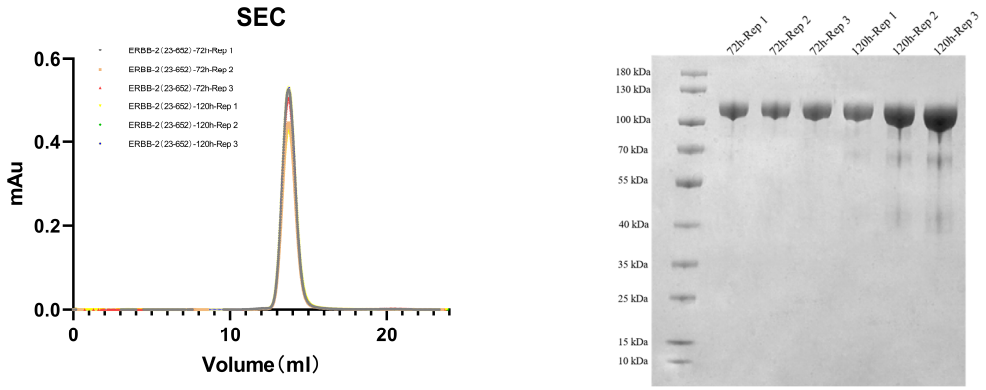
This study developed a novel absorbance-interference glycoprotein analysis method based on AUC technology. It employs a dual detection system combining UV-visible and interferometric measurements to simultaneously capture sample sedimentation signals. The truncated form of the receptor tyrosine-protein kinase erbB-2（ERBB-2） (UniProt ID: P04626) was employed as a glycosylation pattern protein to analyze the MW and proportion of glycosylated proteins. We integrated SDS-PAGE, SEC, MALS, MP, and LC-MS techniques to cross-validate the algorithm's reliability. Systematic comparisons were performed to clarify the strengths and limitations of each method in determining glycoprotein MW, demonstrating this approach's superiority in assessing glycan ratios and MW. The method was integrated into the AUC data analysis software AUCAgent. This study

further applied the AUC dual-signal detection strategy to determine the polysaccharide content in polysaccharide-conjugated vaccines. The study provides novel methodological support for analyzing MW or component ratios in biological agents such as glycosylated proteins and polysaccharide-conjugated vaccines.

Results

SEC and SDS-PAGE analyses confirm the high consistency of glycoprotein purification outcomes

SEC separates samples through porous stationary phases according to molecular hydrodynamics volume, in which large molecules are eluted first because their size prevents them from entering the smaller pores, and small molecules are eluted later due to their retention by the pores, representing an important means to studying MW and MW distribution. SEC analysis revealed that the retention volumes of ERBB-2 (23–652) proteins expressed in a 293F cell system at 72 and 120 h under the same chromatographic conditions (column model: SuperdexTM-200 Increase 10/300 GL column; mobile phase: phosphate-buffered saline (PBS)) were highly consistent, indicating that the hydrodynamic volume and MW distribution of these glycoproteins in solution were not significantly different (Fig. 1a). This phenomenon suggested that the glycoproteins of the proteins with different expression times had similar molecular size characteristics in SEC separation. However, SEC results for modified proteins, such as glycoproteins, need to be interpreted with caution, as the glycan chains may mask the MW differences of the proteins; thus these proteins need to be further analyzed in combination with techniques such as MALS.



(a) (b)

Fig. 1 (a) Expression of 72-h and 120-h receptor tyrosine-protein kinase（ERBB-2） (23–652) proteins analyzed by SuperdexTM-200 Increase 10/300 GL size-exclusion chromatography. (b) Sodium dodecyl sulfate-

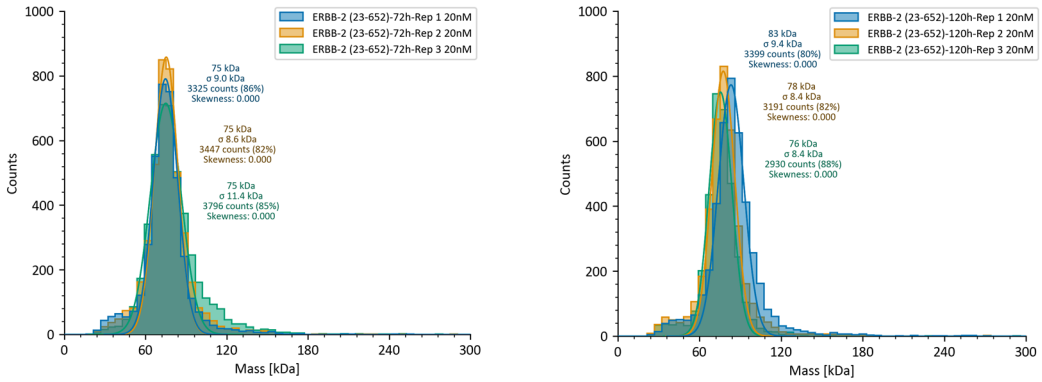
polyacrylamide gel electrophoresis (SDS-PAGE) results of 72-h and 120-h ERBB-2 (23–652) protein expression.

SDS denatures and uniformly negatively charges proteins, and PAGE separates them by MW, with smaller MW proteins migrating faster. Thus, SDS-PAGE is commonly used to estimate the MW and purity of proteins. In reducing SDS-PAGE analysis, both 72-h and 120-h ERBB-2 (23–652) protein samples showed single bands with consistent migration positions, indicating no significant difference in MWs after SDS binding in the unfolded state (Fig. 1b). Notably, all bands migrated at a larger MW position than the theoretical protein MW (71 kDa), which reflect a migration lag due to glycosylation modification. Thus, the MW of glycoproteins estimated by SDS-PAGE would be overestimated.

MP characterization of the consistency of the ERBB-2 (23–652) protein

MP technology, based on the principle of single-molecule scattering, enables the label-free “weighing” and consistency analysis of molecules in their native state by detecting the intensity of light scattered from biomolecules in solution under laser illumination; this intensity is proportional to MW[22]. We characterized the homogeneity of ERBB-2 (23–652) protein samples using MP to validate the SEC measurement results.

MP analysis results indicated a MW of approximately 76 kDa in the 72-h expression sample, with high uniformity (Fig. 2a). In contrast, the 120-h expression sample showed inconsistent results, at 83 kDa, 78 kDa, and 76 kDa, demonstrating lower consistency (Fig. 2b). Although the 72-h sample exhibited good uniformity, the MW deviated significantly from the MALS results. The consistency of the 120-h sample also fell below the SEC analysis results. This discrepancy may stem from the characteristics of the standard used for MP calibration or fitting. The standard used in this study was non-glycosylated bovine serum albumin (BSA), whereas the actual samples were glycosylated proteins. The presence of glycan chains may influence scattering behavior in MP, introducing systematic errors into calibration established using BSA. In contrast, both MALS and AUC are absolute MW determination methods that do not require calibration with standards.

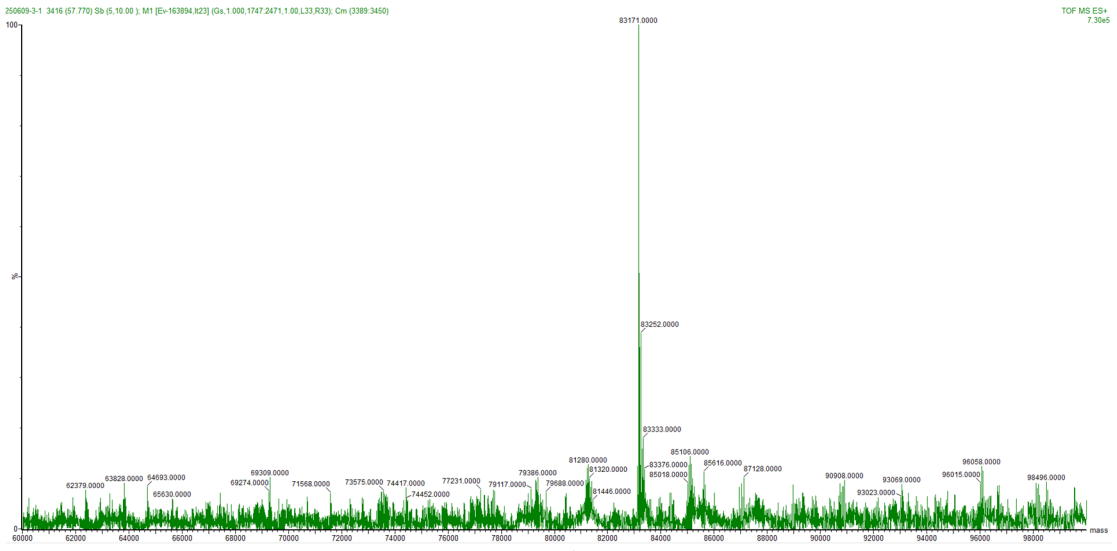


(a) (b)

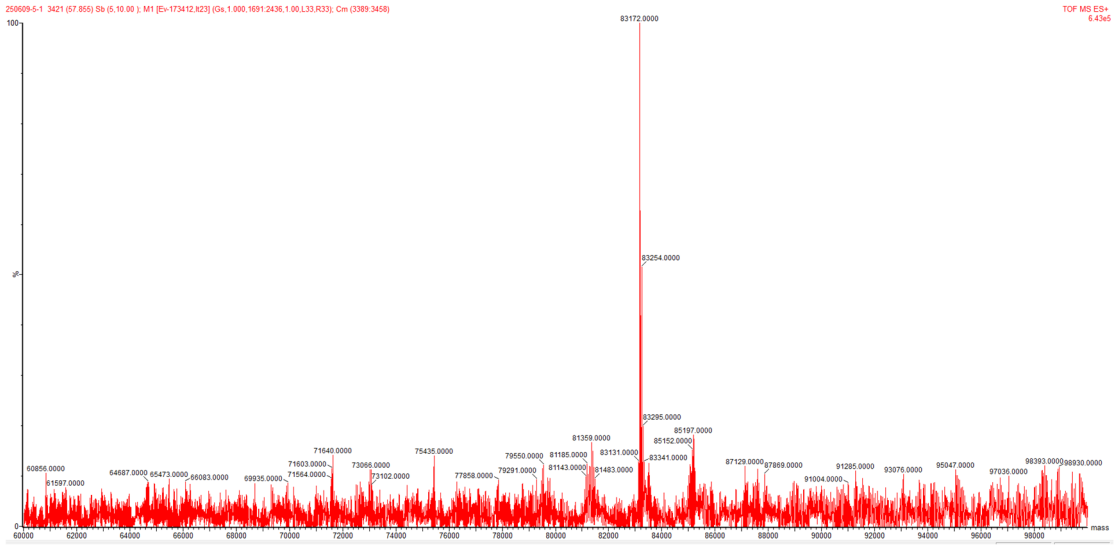
Fig. 2 Mass photometry (MP) Detection of glycoprotein MW. (a) MW measurement of ERBB-2 (23–652) protein using MP at 72 h of expression; (b) MW measurement of ERBB-2 (23–652) protein using MP at 120 h of expression.

LC-MS analysis of the average MW of the ERBB-2 (23–652) protein

LC-MS, as a mass spectrometry technique capable of high-precision MW determination, is increasingly becoming an important complementary or alternative method in protein analysis and characterization. This study performed LC-MS analysis on ERBB2-(23–652) proteins expressed at 72 and 120 h. The results showed measured MWs of 83,171 Da and 83,172 Da, respectively (Fig. 3). However, complex spectra with low signal-to-noise ratios were observed in all replicate samples, making effective deconvolution processing using software challenging. This reflects the limitations of the current algorithms when handling such highly glycosylated complex molecules. These findings are consistent with a study by Puranik et al[11]. reporting that glycosylation interference affected deconvolution in Fc fusion proteins.



（a）



（b）

Fig. 3 Liquid chromatography-mass spectroscopy (LC-MS) results. (a) LC-MS results for ERBB-2 (23−652) protein expressed at 72 h (Rep1). (b) LC-MS results for ERBB-2 (23−652) protein expressed at 120 h.

SEC-MALS enables high-precision determination of the MW of intact glycoproteins

MALS is a technique used to determine the absolute molecular mass of biomolecules. When a macromolecule in solution is irradiated by a beam of laser light, the charge in the molecule is vibrated by the electric field of the light, generating scattered light.

The intensity of the scattered light detected at a fixed angle (I(θ) scatteered) is directly proportional to the MW of the substance (M), the quadratic of the

concentration (c), and the increment of the refractive index increment (dn/dc), as shown in Equation (1):

$$I(\theta)_{scattered} \propto Mc\left(\frac{dn}{dc}\right)^2 \quad (1)$$

MALS is often used in conjunction with an SEC and a differential refractive index detector to acquire three detector signals simultaneously and characterize the MW and size information of biomolecules. For samples with known dn/dc or UV extinction coefficients, either UV or dRI data can be selected as the source of the concentration to calculate the MW.

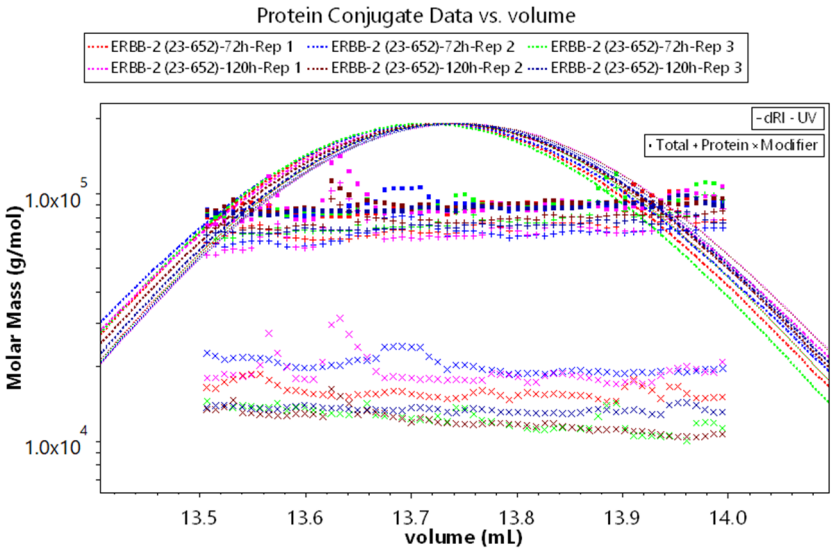


Fig. 4 Molecular weight (MW) detected by Multi-angle light scattering (MALS). The MW determination of glycoprotein complex components involves two distinct calculations: using protein ultraviolet (UV) extinction coefficients for protein moieties and polysaccharide dn/dc values for carbohydrate portions. The ■ symbols denote the total molecular mass, + denotes the protein molecular mass, and × denotes the glycan molecular mass.

Table 1 Results of molecular weight(MW) and glycan content calculated by Multi-angle light scattering (MALS)

Sample ID	Total MW	Protein MW	Gly MW	Gly %
ERBB2-(23–652)-72h-Rep1	88880(±5.081%)	73020(±5.075%)	15860(±5.107%)	17.84
ERBB2-(23–652)-72h-Rep2	89630(±4.388%)	69220(±4.385%)	20410(±4.4%)	22.77
ERBB2-(23–652)-72h-Rep3	90610(±5.82%)	78140(±5.83%)	12470(±5.579%)	13.76
ERBB2-(23–652)-120h-Rep1	91510(±6.187%)	72150(±6.153%)	19360(±6.317%)	21.16
ERBB2-(23–652)-120h-Rep2	91030(±5.73%)	78870(±5.723%)	12160(±5.778%)	13.36
ERBB2-(23–652)-120h-Rep3	87600(±1.622%)	74120(±1.625%)	13470(±1.614%)	15.38
Average				17.38
Std				3.9196

Since scattered light intensity is proportional to the quadratic of the electric field strength (E^2), the free molecules produce incoherent light with the scattered light intensity proportional to $(E1^2 + E2^2)$, whereas the bound molecular complexes produce coherent light with the scattered light intensity proportional to $(E1 + E2)^2$. Thus, free A/B molecules and bound AB complexes produce different scattering signals, so the overall parameters of the protein complexes need to be calculated by combining UV and dRI using the UV extinction coefficients and dn/dc of the two substances in the complex. The overall extinction coefficients of the complex and the calculation of dn/dc are shown in Equations (3) and (4), respectively[20]:

$$C = \frac{A}{\varepsilon_{complex} \times L} = \frac{dRI}{\left(\frac{dn}{dc}\right)_{complex}} \quad (2)$$

$$\left(\frac{dn}{dc}\right)_{complex} = \left(\frac{dn}{dc}\right)_{protein} \times X_{protein} + \left(\frac{dn}{dc}\right)_{modifier} \times \left(1 - X_{protein}\right) \quad (3)$$

$$\varepsilon_{compelx} = \varepsilon \times X_{protein} + \varepsilon \times \left(1 - X_{protein}\right) \quad (4)$$

$$Mw \propto \frac{I_{LS}}{\left(\frac{dn}{dc}\right)^2_{complex} \times c} \quad (5)$$

where $X_{protein}$ is the weight fraction of the protein.

In light of the above, MALS could be used in conjunction with SEC and an differential refractive index detector to obtain information on the total MW of ERBB-2 (23–652) protein, as well as the molecular mass of the protein and glycan. The MW distributions of the components of ERBB-2 (23–652) protein expressed for 72 h and 120 h are shown in Fig. 4. Their specific MW distributions are shown in Table 1. Some differences in the glycosylation modifications of ERBB-2 (23–652) protein under 72 h and 120 h expression conditions were observed. The polysaccharide percentage (Gly %) of the 72-h samples (Rep1-3) ranged from 13.76% to 22.77%, while the 120-h samples (Rep1-3) ranged from 13.36% to 21.16%, with the overall trend showing slightly less glycosylation in the 120-h samples, but some overlap between the two sets of data. Notably, the polysaccharide content fluctuated widely between replicate samples within the same time point (e.g., 9.01% difference between

Rep1 and Rep3 at 72 h, and 7.8% difference between Rep1 and Rep2 at 120 h), suggesting that the reproducibly of glycoprotein MALS measurements is limited. This variability may arise from glycosylation heterogeneity, where slight site modification differences may alter the light scattering signal and affect the polysaccharide content calculation, even if the degree of glycosylation modification is uniform. Alternatively, it may result from differences in instrument sensitivity in response to the dynamic structure of the glycan chain. Although by this method, SEC-MALS can measure the MW of whole glycoproteins with high accuracy and provide the MWs of both protein and glycan fractions, further validation by other technical means is recommended.

AUC-based methods offer higher stability than other methods

Five methods, including LC-MS, SEC-MALS, MP, GUSSI's, and absorbance-interference, were adopted to analyze glycoproteins. The distribution of the sedimentation coefficient of ERBB-2 (23-652) protein obtained by AUC is shown in Fig. 5.

Among all these methods, GUSSI's method demonstrated superior stability. The average glycan fraction remained stable at about 16.7% in the six groups of experimental data, with a standard deviation of 0.4382 (Fig. 6 and Table 2). The absorbance-interference method ranked second in reliability. The average glycan fraction calculated by AUCAgent was 12.73%, with a standard deviation of 0.7284 (Table 3). In comparison to the two preceding AUC-based methods, SEC-MALS exhibited relatively lower stability, with an average value of 17.38% and a standard deviation of 3.9196.

The absorbance-interference method works by comparing concentration differences between two measurement signals to directly obtain the ratio of glycans and proteins in glycoproteins. For reliable results, on the one hand, it is crucial to have precise calibration between the signals and the corresponding concentration values. Thus, two crucial optical properties, the protein extinction coefficient and the refractive index increments of both proteins and glycans in glycoproteins, must be considered when using this method. However, the accuracy of MW measurement depends on the accuracy of frictional ratio (f/f_0) fitting by $c(s)$ sedimentation

coefficient distribution analysis in SEDFIT. More accurate frictional ratio determinations yield more reliable MW measurements.

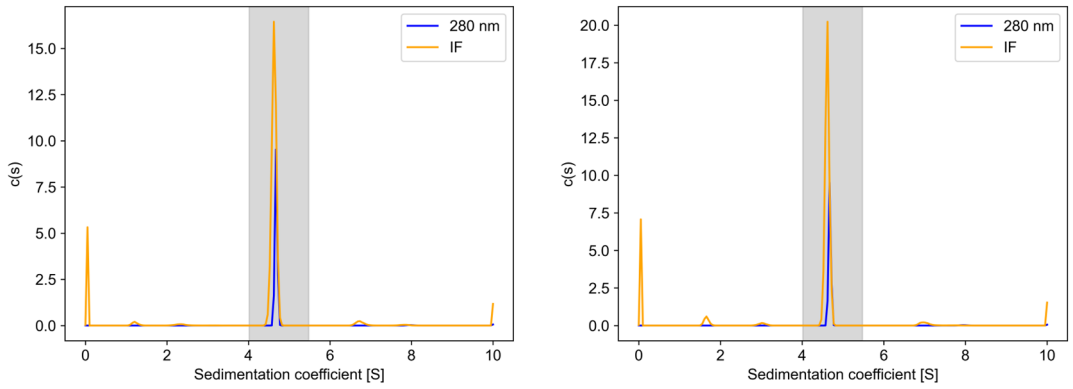
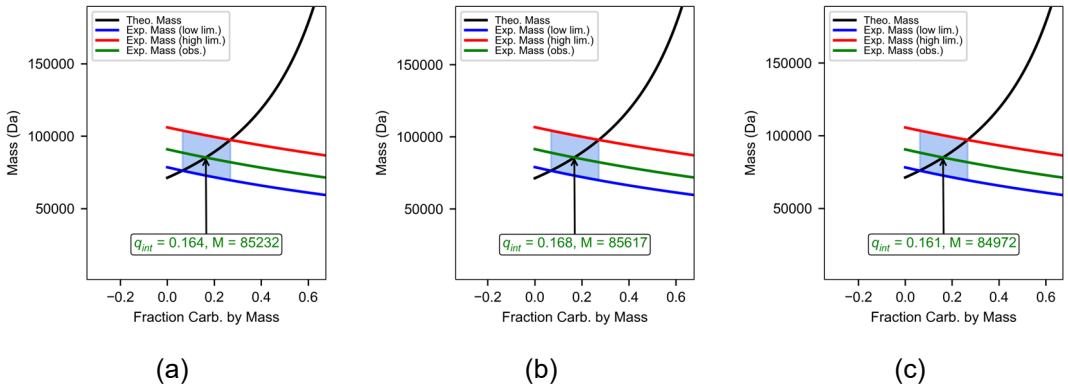


Fig. 5 c(s) distributions from SEDFIT. (a) Sedimentation distribution of ERBB-2 (23–652) protein at 72 h of expression (Rep1). (b) Sedimentation distribution of ERBB-2 (23–652) protein at 120 h of expression (Rep1). The shaped area is the glycoprotein integration region in GUSSI and AUCAgent.

Table 2 Results of MW and glycan content calculated by GUSSI

Sample ID	Total MW	Protein MW	Gly MW	Gly %
ERBB2-(23–652)-72h-Rep1	85232	71254	13978	16.4
ERBB2-(23–652)-72h-Rep2	85617	71233	14384	16.8
ERBB2-(23–652)-72h-Rep3	84972	71292	13680	16.1
ERBB2-(23–652)-120h-Rep1	86211	71210	15001	17.4
ERBB2-(23–652)-120h-Rep2	85583	71291	14292	16.7
ERBB2-(23–652)-120h-Rep3	85604	71223	14381	16.8
Average				16.7
Std				0.4382



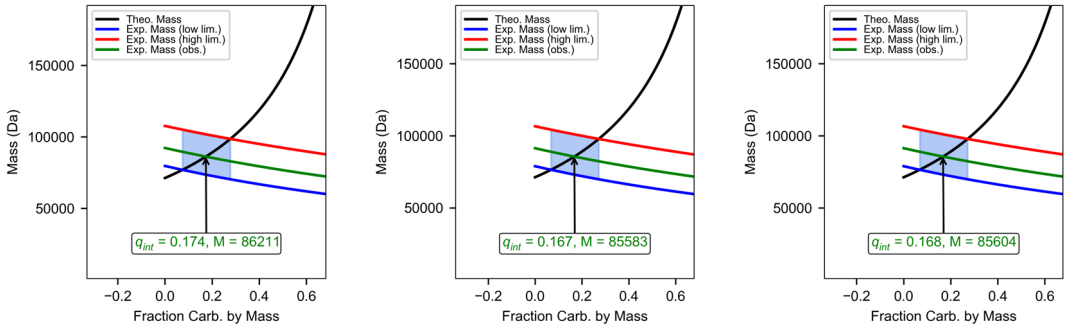


Fig. 6 Glycoprotein MW calculated by GUSSI: (a-c) ERBB-2 (23–652) protein MW measurement at 72 h of expression. (d-f) MW measurement of ERBB-2 (23–652) protein at 120 h of expression.

Table 3 Results of MW and glycan content calculations by AUCAgent

Sample ID	Total MW	Protein MW	Gly MW	Gly %
ERBB2-(23–652)-72h-Rep1	86792	76646	10146	11.69
ERBB2-(23–652)-72h-Rep2	87128	76507	10621	12.19
ERBB2-(23–652)-72h-Rep3	86173	75393	10780	12.51
ERBB2-(23–652)-120h-Rep1	87630	76159	11471	13.09
ERBB2-(23–652)-120h-Rep2	86731	75222	11509	13.27
ERBB2-(23–652)-120h-Rep3	86632	74824	11808	13.63
Average	-	-	-	12.73
Std				0.7284

Table 4 Mass spectroscopy (MS) , MALS, Mass photometry (MP) , and Analytical ultracentrifugation (AUC) results for calculating total protein MW and glycan content

Sample ID	MS MW total	MALS MW total	MP MW total	AUC-SEDFIT-280 MW total	AUC-SEDFIT-IF MW total	GUSSI MW total	MS Gly %	MALS Gly %	GUSSI Gly %
ERBB2-(23–652)-72h-Rep1	83171	88880(± 5.081%)	75000	91555	97525	85232	14.33%	17.80%	16.40%
ERBB2-(23–652)-72h-Rep2	/	89630(± 4.388%)	75000	92087	95735	85617	/	22.80%	16.80%

ERBB2-(23–652)-72h-Rep3	/	90610(± 5.82%)	75000	91189	94251	84972	/	13.80%	16.10%
ERBB2-(23–652)-120h-Rep1	83172	91510(± 6.187%)	83000	92938	95444	86211	/	21.20%	17.40%
ERBB2-(23–652)-120h-Rep2	/	91030(±5.73%)	78000	92047	93917	85583	14.33%	13.40%	16.70%
ERBB2-(23–652)-120h-Rep3	/	87600(± 1.622%)	76000	92070	94257	85604	/	15.40%	16.80%

AUC can accurately determine the polysaccharide ratio in polysaccharide-conjugated vaccines

Polysaccharide-conjugated vaccines represent a novel vaccine type created by chemically linking bacterial capsular polysaccharides to protein carriers via covalent bonds. They effectively stimulate T-cell-dependent immune responses, overcoming the limitations of conventional vaccines, such as weak immunogenicity in infants and young children and an inability to induce lasting immune memory. This significantly enhances the protective efficacy and induces long-term immunity. Currently, this type of vaccine has been successfully applied to prevent infections caused by multiple pathogens, including *Hemophilus influenzae* type b, *Streptococcus pneumoniae*, and *Neisseria meningitidis*[18,23,24]. Although different polysaccharide-conjugated vaccines vary in their specific conjugation processes and typically involve multi-step reactions, their polysaccharide ratios are critical quality control indicators. These factors directly impact the stability, safety, efficacy, and batch-to-batch consistency of the conjugates. Therefore, this study applied the aforementioned algorithm to further explore its ability to determine polysaccharide ratios in polysaccharide-conjugated vaccines (Table 5).

Both vaccines utilize tetanus toxoid as the carrier protein but exhibit significant differences in protein content. We measured sedimentation velocity in vaccine samples using AUC, simultaneously collecting absorbance values at 280 nm and interference light signals. The sedimentation distribution results are shown in Fig. 7. The polysaccharide ratios in the samples were obtained through algorithmic analysis of the data. The summary results presented in Table 6.

Table 5 Polysaccharide-conjugated vaccine information

Sample ID	Buffer	Polysaccharide ratio (%)
Polysaccharide-conjugated vaccines-1	0.85% NaCl	37.1
polysaccharide-conjugated vaccines-2	0.85% NaCl	57.6

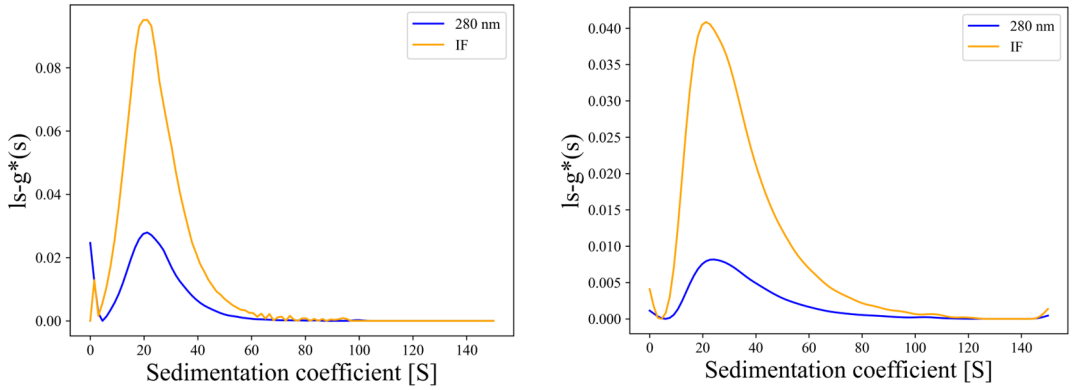


Fig. 7 ls-g*(s) distributions from AUCAgent. (a) Sedimentation distribution of Polysaccharide-conjugated vaccines-1 (Rep1). (b) Sedimentation distribution of Polysaccharide-conjugated vaccines-2 (Rep1).

Table 6 AUC results were used to calculate the polysaccharide ratio of polysaccharide-conjugated vaccines

Sample ID	Buffer	Polysaccharide ratio (%)
Polysaccharide-conjugated vaccines-1-Rep1	0.85% NaCl	40.9
Polysaccharide-conjugated vaccines-1-Rep2	0.85% NaCl	41.1
Polysaccharide-conjugated vaccines-1-Rep3	0.85% NaCl	40.01
Polysaccharide-conjugated vaccines-2-Rep1	0.85% NaCl	57.1
Polysaccharide-conjugated vaccines-2-Rep2	0.85% NaCl	59.7
Polysaccharide-conjugated vaccines-2-Rep3	0.85% NaCl	60.2

Both vaccines exhibited broad sedimentation coefficient distributions (Fig. 7), consistent with the broad MW distribution observed by Jia et al[25]. for polysaccharide-conjugate vaccines using MALS technology. The results reflect significant compositional heterogeneity within the samples in solution. Therefore, based on the difference between the 280 nm and interference light dual signals, we selected the entire sedimentation distribution region and used the absorbance-interference method

to calculate polysaccharide ratios. Each vaccine was tested in triplicate, and the results showed good consistency between replicates. The values deviated from the theoretical polysaccharide ratios by less than 5%, indicating that AUC combined with this method exhibits high accuracy and reproducibility in vaccine component distribution analysis.

Discussion

GUSSI's method balances simplicity and robustness by fixing the protein's MW to its theoretical value. However, in practice, experimentally observed apparent MW often deviates from theoretical values. Enforcing strict MW constraints could systematically skew results, particularly for glycoproteins with heterogeneous glycosylation. Another essential factor in GUSSI's method is the oligomeric state of the protein. This is difficult to estimate experimentally, even if the actual protein monomer sequence is known. The absorbance-interference method entirely bypasses the need to account for protein multimerization. It relies solely on the protein's molar extinction coefficient, the refractive index increment (dn/dc) of the protein, and the dn/dc of the glycan to accurately quantify glycan content. If the protein sequence is known, its extinction coefficient and dn/dc value can be computationally predicted. For glycans, dn/dc values show minimal variation across types (e.g., 0.142–0.150 mL/g), allowing the use of a generic value when experimental data is unavailable. Building on these parameters, this method enables the straightforward quantification of glycan content in glycoproteins, overcoming the limitations of GUSSI's method requiring knowledge of the multimerization state. Predicted or generic parameters may not precisely match sample-specific values, but deviations from true values are typically minor. Furthermore, our findings demonstrate that the combined application of the absorbance-interference method and GUSSI's method enables the estimation of glycoprotein oligomeric states.

While SEC analysis cannot provide detailed glycan composition data, it confirmed batch-to-batch consistency across the six sample sets. MALS shares a similar fundamental principle with the AUC-based absorbance-interference method, given that both methods rely on the integration of two detection modalities for molecular characterization. However, MALS exhibits significantly higher variability compared to the AUC method, which underscores the superior reproducibility and precision of AUC in glycoprotein analysis.

Furthermore, the analysis of polysaccharide-conjugated vaccines showed that the absorbance interference method could directly quantitate polysaccharide ratios. Given the common polydispersity of such vaccines, the absorbance interference strategy employed by AUCAgent not only offers superior reproducibility and precision in glycoprotein quantification compared to other methods but also demonstrates significant advantages in analyzing ratios in polysaccharide-conjugated vaccines, indicating substantial application potential.

Methods

Protein expression and purification

The receptor tyrosine-protein kinase erbB-2 (ERBB-2) (23–652) gene sequence was synthesized by Ruibiotech. The truncated protein contained seven N-glycosylation sites located at positions 68, 125, 187, 259, 530, 571, and 629 of the sequence. These were amplified by PCR after cloning into a pVAX1 vector, a 6×His-tagged fusion protein was generated, and ERBB-2 (23–652) protein was expressed in a 293F cell system. The 293F cells were cultured to a cell concentration of 2×10^6 cells/mL, and the mass ratio of plasmid to PEI was about 1:4 for transfection, and 1 mg of plasmid per liter of cells. The cells were transfected under the same conditions, and the cellular supernatants were collected at 72 and 120 h. Three sets of replicates were constructed at the two time points, with 1 L of cells for each set of replicates.

The cell supernatant was collected and then subjected to PBS (pH 7.4) buffer replacement using a 10 kDa small ultrafiltration membrane pack (Merck, P2C100C01) via an ultrafiltration concentration system (Merck Millipore Mini Pellicon, Merck, Darmstadt, Germany). The concentrated protein solution was subjected to affinity chromatography using a His affinity column (Cytiva, 17531802). Heterogeneous proteins were removed using wash buffer (PBS, 20 mM imidazole, pH 7.4), and target proteins were eluted using elution buffer (PBS, 400 mM imidazole). The eluted proteins were concentrated using a 10 kDa centrifugal filter (Merck, UFC905008) and further purified using a SuperdexTM-200 Increase 5/150 GL column (Cytiva, 28990945) equilibrated with PBS buffer. The proteins were finally stored in PBS buffer.

polysaccharide-conjugated vaccines manufacturing process

The Streptococcus pneumoniae serotypes (22F using the amine reduction method and 19A using the 1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP) method) were conjugated to the protein of diphtheria toxoid as the carrier to form covalently bound polysaccharide-protein conjugates. The purification was carried out by tangential flow ultrafiltration and molecular sieve chromatography. The glycan content was determined according to the anthrone-sulfuric acid method in the current edition of the Chinese Pharmacopoeia (Part III).

Size-exclusion chromatography

An Agilent 1260 infinity liquid chromatography system, equipped with an autosampler, a thermostatic column chamber, a UV detector, and a temperature control at 20°C, was used for chromatographic separation. A SuperdexTM-200 Increase 10/300 GL column (Cytiva, 28990944) was used with PBS as the mobile phase and a detector wavelength of 280 nm. The protein concentration of ERBB-2 (23–652) purified at 72 h and 120 h was quantified uniformly to an A280 nm value of about 0.6, 100 μL of each sample was uploaded, the run flow rate was 0.5 mL/min, and the run time was 48 minutes.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Genscript YoungPAGE™ Protein Prep Gel (Genscript, M00928) was used for the electrophoretic separation of the samples. Protein samples were mixed with protein loading buffer (SDS-PAGE sample loading buffer, 5X) at a volume ratio of 4:1 prior to sampling and denatured at 100°C for 5 min, after which gel electrophoresis was performed. Pre-stained protein marker (Thermo Scientific, 26616), containing 10 MW markers, 180, 130, 100, 70, 55, 40, 35, 25, 15, and 10 kDa protein bands, was used for protein MW estimation. Staining was performed using the eStain® L1 Protein Stainer (Genscript, Nanjing, China) after completing electrophoresis.

Liquid chromatography-mass spectroscopy

In LC-MS analysis, the analytes were separated by a 60 min gradient elution at a flow rate 0.5 μL/min with a nanoACQUITY UPLC system, which was directly interfaced with a SYNAPT-G2-Si mass spectrometer produced by the Waters company. The analytical column was a Protein BEH C4 silica capillary column (150 μm ID, 100 mm long) packed with C-4 resin (300 Å, 1.7 μm) purchased from the Waters company. Mobile phase A consisted of 0.1% formic acid aqueous solution, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid.

Multi-angle light scattering

An Agilent 1260 infinity liquid chromatography system configured with an autosampler, a thermostat column chamber, a UV detector, and temperature control at 20°C was used for MALS analysis. Chromatographic separation was achieved using a SuperdexTM-200 Increase 10/300 GL column (Cytiva, 28990944) connected to a Wyatt DAWN HELEOS laser detector (Wyatt, CA, USA) and a Wyatt Optilab T-rEX differential detector (Wyatt, CA, USA). A PBS mobile phase overnight equilibration system and a detector were used. The 72-h and 120-h purified ERBB-2 (23–652) protein concentrations were quantified to an A280 nm value of about 0.6, and 100 μL of samples were uploaded. Light scattering, UV, and differential signals were collected, and the data were analyzed using ASTRA 8.1.2 software.

Mass photometry

The solution-phase mass determination of native glycoproteins was acquired using a TwoMP (Refeyn, Ltd, Oxford, UK) mass photometer calibrated with BSA (66 kDa). Experimental data were obtained in the form of mass photometry movies acquired over 6,000 frames (60 s) using the AcquireMP software (2024R2) on precleaned, high-sensitivity microscope slides [n]. The test samples (200 nM) were mixed with 18 μL of PBS to a final concentration of 20 nM, and then loaded onto a glass slide for measurement. The final protein concentrations were determined empirically to achieve approximately 50 binding events per second.

Analytical ultracentrifugation

AUC (ProteomelabXL-A, BeckmanCoulter, USA) sedimentation velocity experiments were performed using both absorbance (280 nm) and interference (660 nm) optical detection systems. The rotor speed was maintained at 45,000 rpm throughout the sedimentation process. Real-time scanning data were acquired at 120-second intervals, with absorbance recorded at 280 nm (for protein detection) and interference signals monitored at 660 nm (for high-precision concentration profiling).

The collected data were processed using sedfit17 software with the fitting parameters set to 200 for resolution, 0 for Smin, 10 for Smax, and a confidence level of 0.68 to obtain the results of the continuous $c(s)$ distribution of the 280 nm and interferometric data for each group of samples. Both distributions were imported into GUSSI 2.1.0 software and analyzed for glycoproteins using the Glycoprotein Calcs module, selecting a sedimentation coefficient range of 4.0–5.5 S. The protein fitting parameters were set to a UV extinction coefficient of $0.9\text{L} \cdot g^{-1} \cdot \text{cm}^{-1}$, and the protein MW was set to 71,250 Da (the above parameters were predicted from https://www.expasy.org/ based on the protein amino acid sequence). The corresponding f/f_0 values were entered as distribution parameters, which were obtained by fitting with sedfit17 software to obtain the final glycoprotein mass distribution results.

In the sedimentation velocity experiments, the particle movement process in solution can be described by the Lamm equation:

$$\frac{\partial c}{\partial t} = \frac{1}{r}\frac{\partial}{\partial t}\left(D\frac{\partial c}{\partial r}r\right) - \omega^2 r^2 sc \tag{6}$$

where c, t, r, D, ω, and s represents the concentration of the solute, sedimentation time, radial distance from the axis of rotation, diffusion coefficient, angular velocity of the rotor, and sedimentation coefficient, respectively. The sedimentation coefficient is defined as

$$\text{s} = \frac{u}{\omega^2 r} \tag{7}$$

and is a measure of the sedimentation speed of macromolecules when using the centrifugation method, which is equal to the speed per unit of the centrifugal field.

The sedimentation coefficient unit is Svedberg (symbol S), which is a unit of time and is equal to 10^{-13}s. The diffusion coefficient of a molecular is related to its translational frictional coefficient, f, through the Einstein-Smoluchowski relationship:

$$D = \frac{RT}{N_0 f} = \frac{k_B}{f} \tag{8}$$

The unit of diffusion coefficient is cm^2/s, also called Fick's (symbol F), where 1 F = $10^{-7}\mathrm{cm}^2/\mathrm{s}$.

GUSSI glycoprotein calculations

GUSSI (Version 2.1.0) includes a glycoprotein analysis module that efficiently calculates glycan composition ratios. It uses $c(s)$ distributions from SEDFIT, combined with partial specific volumes, experimental temperature, the experimental solution density, viscosity, and f/f_0, to directly quantify the fraction of the total mass of the carbohydrate in glycoprotein complexes. GUSSI's theory for calculating glycoproteins is as follows. First, the assumed glycan fraction is q, and the hypothetical total glycoprotein molar mass is $M_{GP,h}$:

$$\mathrm{M_{GP,h}} = \frac{nM_{p,mono}}{1-q} \tag{9}$$

The apparent molecular mass $M_{GP,s}$ can also be calculated from the experimental data according to the Svedberg equation:

$$\mathrm{M_{GP,s}} = \frac{s_{GP}RT}{D_{GP}(1-\bar{v}_{GP}\rho)} \tag{10}$$

where s_{GP} is the sedimentation coefficient of glycoprotein. The $\bar{v}$ of glycoprotein needs to be recalculated based on the glycan fraction:

$$\bar{v}_{GP} = q\bar{v}c + (1-q)\bar{v}_P \tag{11}$$

It is easy to calculate the value of q, so that the hypothetical molar mass and the apparent molecular mass are consistent:

$$\mathrm{q} = \frac{1-[\psi(1-\bar{v}_P)]}{1+[\psi\rho(\bar{v}_P-\bar{v}_C)]} \tag{12}$$

where

$$\psi = \frac{nM_{P,mono}D_{GP}}{s_{GP}RT} \tag{13}$$

The sedimentation distributions used in GUSSI are all taken from absorbance-type data, because the frictional ratio is affected by all species, as more species can be detected in the interference optics.

AUCAgent glycoprotein calculations

Unlike GUSSI, the analysis of glycoprotein by AUCAgent is based on absorbance optics and interference optics, namely the absorption-interference difference method. In the interference optical system, the number of fringes J follows[26]:

$$J = c\frac{dn}{dc}\frac{l}{\lambda} \tag{14}$$

where c is the concentration, and dn/dc is the specific refractive increment. We used 0.185 mL/g for proteins and 0.15 mL/g for glycans. l is the optical pathlength, and λ is the wavelength of the light used. In the absorbance optical system, the measured absorbance follows the Beer-Lambert law:

$$A = \varepsilon lc \tag{15}$$

where ε is the molar extinction coefficient per cm, c is the concentration, and l is the optical pathlength determined by the centerpiece type.

The quantitative determination of glycan occupancy in glycoprotein involves three sequential calculations: total glycoprotein (c_{GP}) concentration is calculated using Equation (16):

$$c_{GP} = \frac{J\lambda}{l \cdot \frac{dn}{dc}} \tag{16}$$

protein core concentration (c_P) is calculated using Equation (15):

$$\mathrm{c_P} = \frac{A}{l \cdot \varepsilon} \tag{17}$$

and the glycan ratio is computed using c_{GP} and c_P:

$$q = \frac{c_{GP} - c_P}{c_{GP}}. \quad (18)$$

As mentioned before, $\bar{v}$ needs to be recalculated (Equation (11)) before calculating the glycoprotein total mass. The corresponding glycan proportion can be calculated using the process described above under the relevant parameters.

Acknowledgments

This work was supported by the Beijing Frontier Research Center for Biological Structure and the Technology Center for Protein Sciences, Tsinghua University; and the State Key Laboratory of Drug Regulatory Science Project (2025SKLDRS0318).

Declarations

The authors declare no conflicts of interest.

References

1. Wang, G., De Jong, R. N., Van Den Bremer, E. T. J., Parren, P. W. H. I. & Heck, A. J. R. Enhancing Accuracy in Molecular Weight Determination of Highly Heterogeneously Glycosylated Proteins by Native Tandem Mass Spectrometry. *Anal. Chem.* **89**, 4793–4797 (2017).
2. Mitra, N., Sinha, S., Ramya, T. N. C. & Surolia, A. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in Biochemical Sciences* **31**, 156–163 (2006).
3. Taylor, M. E. & Drickamer, K. *Introduction to Glycobiology*. (Oxford University Press, 2011). doi:10.1093/hesc/9780199569113.001.0001.
4. Nothaft, H. & Szymanski, C. M. Bacterial Protein N-Glycosylation: New Perspectives and Applications. *Journal of Biological Chemistry* **288**, 6912–6920 (2013).
5. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370–2376 (2001).
6. Parekh, R. B. *et al.* Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452–457 (1985).
7. Gupta, S. K. & Shukla, P. Glycosylation control technologies for recombinant therapeutic proteins. *Appl Microbiol Biotechnol* **102**, 10457–10468 (2018).

8. Su, L. *et al.* Carbohydrate-Based Macromolecular Biomaterials. *Chem. Rev.* **121**, 10950–11029 (2021).

9. Carter, P. J. Introduction to current and future protein therapeutics: A protein engineering perspective. *Experimental Cell Research* **317**, 1261–1269 (2011).

10. Dicker, M. & Strasser, R. Using glyco-engineering to produce therapeutic proteins. *Expert Opinion on Biological Therapy* **15**, 1501–1516 (2015).

11. Puranik, A., Saldanha, M., Dandekar, P. & Jain, R. A comparison between analytical approaches for molecular weight estimation of proteins with variable levels of glycosylation. *Electrophoresis* **43**, 1223–1232 (2022).

12. Folta-Stogniew, E. & Williams, K. R. Determination of Molecular Masses of Proteins in Solution: Implementation of an HPLC Size Exclusion Chromatography and Laser Light Scattering Service in a Core Laboratory. *LASER LIGHT SCATTERING*.

13. Guan, Y. *et al.* An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Sci Rep* **5**, 13370 (2015).

14. Sonboli, R., Najafi, Z., Zarezadeh, N., Yazdani, M. & Behrouz, H. Improving SDS-PAGE method for monoclonal antibodies: The advantages of Tris-Acetate over Tris-Glycine SDS-PAGE system and comparison with CE-SDS method. *Protein Expression and Purification* **182**, 105845 (2021).

15. Some, D., Amartely, H., Tsadok, A. & Lebendiker, M. Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *JoVE* 59615 (2019) doi:10.3791/59615.

16. Hong, P., Koza, S. & Bouvier, E. S. P. A REVIEW SIZE-EXCLUSION CHROMATOGRAPHY FOR THE ANALYSIS OF PROTEIN BIOTHERAPEUTICS AND THEIR AGGREGATES. *Journal of Liquid Chromatography & Related Technologies* **35**, 2923–2950 (2012).

17. Wen, J., Arakawa, T. & Philo, J. S. Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions. *Analytical Biochemistry* **240**, 155–166 (1996).

18. MacCalman, T. E., Phillips-Jones, M. K. & Harding, S. E. Glycoconjugate vaccines: some observations on carrier and production methods. *Biotechnology and Genetic Engineering Reviews* **35**, 93–125 (2019).

19. *Carbohydrate-Based Vaccines: Methods and Protocols*. vol. 1331 (Springer New York, New York, NY, 2015).

569 20. Bender, M. F., Li, Y., Ivleva, V. B., Gowetski, D. B. & Paula Lei, Q. Protein and glycan
570 molecular weight determination of highly glycosylated HIV-1 envelope trimers by HPSEC-
571 MALS. *Vaccine* **39**, 3650–3654 (2021).
572 21. Li, D., Chu, W., Sheng, X. & Li, W. Optimization of Membrane Protein TmrA Purification
573 Procedure Guided by Analytical Ultracentrifugation. *Membranes* **11**, 780 (2021).
574 22. Sonn-Segev, A. *et al.* Quantifying the heterogeneity of macromolecular machines by
575 mass photometry. *Nat Commun* **11**, 1772 (2020).
576 23. Li, Y. *et al.* Polysaccharide conjugate vaccines: From historical milestones to future
577 strategies. *Carbohydrate Polymers* **371**, 124469 (2026).
578 24. Micoli, F. *et al.* Glycoconjugate vaccines: current approaches towards faster vaccine
579 design. *Expert Review of Vaccines* **18**, 881–895 (2019).
580 25. Jia, X. *et al.* Characterization of pneumococcal conjugates in vaccine process
581 development by multi-detection hydrodynamic chromatography. *Journal of Pharmaceutical*
582 *and Biomedical Analysis* **261**, 116826 (2025).
583 26. Richards, E. Glen., Teller, D. C. & Schachman, H. K. Ultracentrifuge studies with
584 Rayleigh interference optics. II. Low-speed sedimentation equilibrium of homogeneous
585 systems. *Biochemistry* **7**, 1054–1076 (1968).
586