Video Article

# Characterization of Glycoproteins with the Immunoglobulin Fold by X-Ray Crystallography and Biophysical Techniques

June Ereño-Orbea*[1], Taylor Sicard*[1,2], Hong Cui[1], Indira Akula[1], Jean-Philippe Julien[1,2,3]

[1]Program in Molecular Medicine, The Hospital for Sick Children Research Institute

[2]Department of Biochemistry, University of Toronto

[3]Department of Immunology, University of Toronto

*These authors contributed equally

Correspondence to: Jean-Philippe Julien at jean-philippe.julien@sickkids.ca

URL: https://www.jove.com/video/57750
DOI: doi:10.3791/57750

Keywords: Biochemistry, Issue 137, Glycoproteins, N-linked glycosylation, crystallization, biolayer interferometry, isothermal titration calorimetry, crystal soaking, fragment antigen binding.

Date Published: 7/5/2018

Citation: Ereño-Orbea, J., Sicard, T., Cui, H., Akula, I., Julien, J.P. Characterization of Glycoproteins with the Immunoglobulin Fold by X-Ray Crystallography and Biophysical Techniques. *J. Vis. Exp.* (137), e57750, doi:10.3791/57750 (2018).

## Abstract

Glycoproteins on the surface of cells play critical roles in cellular function, including signalling, adhesion and transport. On leukocytes, several of these glycoproteins possess immunoglobulin (Ig) folds and are central to immune recognition and regulation. Here, we present a platform for the design, expression and biophysical characterization of the extracellular domain of human B cell receptor CD22. We propose that these approaches are broadly applicable to the characterization of mammalian glycoprotein ectodomains containing Ig domains. Two suspension human embryonic kidney (HEK) cell lines, HEK293F and HEK293S, are used to express glycoproteins harbouring complex and high-mannose glycans, respectively. These recombinant glycoproteins with different glycoforms allow investigating the effect of glycan size and composition on ligand binding. We discuss protocols for studying the kinetics and thermodynamics of glycoprotein binding to biologically relevant ligands and therapeutic antibody candidates. Recombinant glycoproteins produced in HEK293S cells are amenable to crystallization due to glycan homogeneity, reduced flexibility and susceptibility to endoglycosidase H treatment. We present methods for soaking glycoprotein crystals with heavy atoms and small molecules for phase determination and analysis of ligand binding, respectively. The experimental protocols discussed here hold promise for the characterization of mammalian glycoproteins to give insight into their function and investigate the mechanism of action of therapeutics.

## Video Link

The video component of this article can be found at https://www.jove.com/video/57750/

## Introduction

Surface proteins play critical roles in cellular function. Through their extracellular domains, these membrane proteins can modulate cell-cell interactions, adhesion, transport and signalling[1,2]. The extracellular localization of these proteins makes them attractive targets for the development of therapeutics for the treatment of a wide range of diseases, including cancer and autoimmune diseases[3,4,5,6,7]. One of the most common folds of human membrane protein ectodomains is the immunoglobulin-like (Ig) fold, which is formed by seven or more β-strands arranged into two β-sheets[8,9]. Typically, Ig-containing glycoproteins are multi-domain structures with Ig domains sequentially arranged on the extracellular portion of the membrane protein[10]. Post-translational modifications of these cell-surface proteins, particularly N- and O-linked glycosylation, have been shown to play essential roles in their regulation, folding, secretion and function[11]. To improve our understanding of their function and to better design therapeutics that can target them, techniques are required that allow for their detailed molecular characterization. Here, we present a combination of techniques that allow for the biophysical (biolayer interferometry (BLI) and isothermal titration calorimetry (ITC)) and structural (X-ray crystallography) characterization of the extracellular domain of Ig-containing membrane glycoproteins, alone and in complex with their biologically relevant ligands and therapeutic molecules (**Figure 1**).

N-linked glycosylation is one of the most common post-translation modifications on mammalian proteins, and occurs during protein maturation within the endoplasmic reticulum and Golgi[12,13]. Cell lines, such as human embryonic kidney (HEK) 293 cells, have been developed for the recombinant expression of large quantities of glycosylated mammalian proteins[14,15]. This cell line has been developed in a suspension format, which allows for the ease of scaling up protein production to larger quantities in comparison to adherent cell lines. Here, we utilize two HEK293 cell lines: HEK293F and HEK293 Gnt I$^{-/-}$ (HEK293S), which differ by the absence of N-acetylglucosaminyl transferase I (Gnt I) in the latter. In turn, production of complex glycans (as seen in HEK293F) is not possible and instead high mannose-type glycans (predominantly $Man_5GlcNAc_2$) reside at N-linked glycan sites[18,19,20]. Using these two cell lines in parallel allows studying the effect of glycan size and complexity on biological function and therapeutic targeting. Indeed, glycoproteins produced in HEK293F cells will have larger, more complex glycans compared to the same glycoprotein produced in HEK293S cells. Glycoproteins produced in HEK293S cells are more amenable to crystallization, because of the reduced chemical and conformational heterogeneity of their N-linked glycans. To further improve crystallizability, glycoproteins produced in

HEK293S (but not HEK293F) cells can be treated with the enzyme endoglycosidase H (Endo H), which results in the cleavage of high mannose glycans such that only a single N-acetylglucosamine (GlcNAc) moiety remains at each N-linked glycosylation site[21,22]. Other methods can also be used to limit N-glycan processing within the cells, such as the addition of glycosyltransferase inhibitors during glycoprotein expression, including kifunensine[23]. Alternative approaches involve the expression of native glycoproteins (in HEK293F cells) followed by enzymatic deglycosylation using peptide N-glycosidase F (PNGaseF). However, deglycosylation with PNGaseF has been shown to be less effective under native conditions and increases aggregation in some proteins; in cases when the protein remains soluble after treatment, it acquires negative charges on its surface due to the deamidation of the asparagine residue to aspartic acid[24], which might be detrimental for its crystallization. Predicted N-glycosylation sites can also be mutated, most often to alanine or glutamine residues, to prevent N-linked glycosylation at these sites and to generate glycoprotein samples of high homogeneity. Alternatively, glycoproteins can be produced in other eukaryotic cell cultures, including yeast, insect, and plant systems, or other mammalian cell lines such as Chinese hamster ovarian (CHO) cells[16,17].

Many mammalian expression vectors, including pHLsec, allow for the secretion of recombinant glycoprotein ectodomains into the cell medium[25]. Secretion of glycoproteins from HEK293 cells allows for rapid and easy purification without the need for cell lysis. Addition of purification tags (*e.g.,* His-tag, Strep-tag, Flag-tag, Myc-tag, HA-tag) to the N or C terminus of the target glycoprotein allows purification by a single-step affinity chromatography. Subsequently, size exclusion chromatography can be used to yield a monodisperse sample for biophysical and structural characterization.

A highly pure and homogeneous glycoprotein sample under appropriate conditions can result in well-diffracting crystals. Once a full X-ray diffraction dataset has been obtained from such crystals, initial phases need to be determined to calculate the electron density of the glycoprotein. Thanks to an ever increasing number of structures in the Protein Data Bank (PDB), the most commonly used method for phasing has by far become molecular replacement (MR), which uses a related protein structure to obtain initial phases[26]. However, when MR fails to solve the phase problem, as has occasionally been the case for multi-Ig domain glycoproteins[27,28,29], alternative methods are required. In this article, we detail a method to soak crystals with heavy atoms (HA) for phasing, which was required for solving the structure of the CD22 ectodomain[28]. Identifying the right HA for phasing is an iterative process that depends on HA reactivity, available atoms in the glycoprotein in a given crystal lattice, and the crystallization solution[30,31]. Alternatively, natural sulfur atoms in cysteine and methionine residues can be used for phasing if present at a high enough ratio to other atoms in the glycoprotein, and if X-ray diffraction data can be collected with high enough redundancy[32,33].

The biological function of membrane glycoproteins is often mediated by protein-protein interactions or protein-ligand interactions, such as with carbohydrates. When the ligand is small enough to diffuse from the solution to the glycoprotein binding site in the crystal lattice, soaking experiments can be successful to obtain a glycoprotein-ligand co-crystal structure to better understand ligand recognition.

The protocols presented here are also relevant for understanding the interactions of surface glycoproteins with synthetic therapeutic ligands[34,35] and antibody therapeutics[36,37]. When combined with structural information, binding kinetics and thermodynamics can be powerful to understand and improve their mechanisms of action. One technique that allows for the kinetic analysis of therapeutic antibodies binding to a glycoprotein is BLI[38,39]. BLI uses biosensors with an immobilized ligand to measure the association and dissociation kinetics with a binding partner, ultimately determining an equilibrium dissociation constant ($K_D$). BLI is an attractive approach because small amounts of glycoproteins are required (<100 µg), experiment time is fast (~10-15 min per run), and it can be automated. ITC is also useful for studying affinities between glycoproteins and binding partners[40,41,42,43]. While ITC is more time and reagent intensive, valuable information can be obtained regarding the thermodynamics of the interaction (ΔG, ΔH, ΔS, and stoichiometry). ITC is also very useful for studying weak interactions that are often associated with the transient binding of surface glycoproteins to ligands. Furthermore, these techniques can be used in conjunction to evaluate the binding of various constructs and assess the effect of different N-linked glycoforms obtained from expressing the glycoprotein in different cell lines. Performing BLI and ITC with glycoproteins produced in HEK293F, HEK293S and treated with Endo H can provide an in-depth view of the role of glycans in biological activity and therapeutic engagement.

We successfully applied these protocols to characterize the extracellular domain (ECD) of human CD22[28], a glycoprotein member of the sialic acid-binding Ig-like lectins (Siglecs) family that is essential for maintaining B-cell homeostasis[44]. We performed in-depth construct design to facilitate crystallization and phased the X-ray dataset by HA soaking with Hg. We also soaked CD22 crystals with its ligand sialic acid (α2-6 sialyllactose) to obtain a structure of the immune receptor-ligand complex and thus provided the blueprints for the structure-guided design of glycan mimetics[45,46]. In addition, we generated the fragment antigen binding (Fab) of the anti-CD22 therapeutic antibody epratuzumab - a therapeutic candidate currently in phase III clinical trials for non-Hodgkin's lymphoma[47]- to determine its binding affinity by BLI and ITC to differentially glycosylated CD22 ECD constructs. These studies revealed a critical role for N-linked glycosylation in epratuzumab engagement, with potential implications for CD22 recognition on dysfunctional B cells.

## Protocol

# 1. Construct Design for Glycoprotein ECD

1. Evaluate the amino acid sequence of human CD22 (Uniprot) using the InterPro and Phyre2 servers to identify predicted domain elements and boundaries located within the protein[48,49].
2. Clone the sequence of human CD22, lacking the signal peptide, transmembrane and cytosolic domains (residues 20-687, hereafter CD22 extracellular domain, CD22 ECD) into the pHLsec mammalian expression vector[25] using restriction enzymes AgeI and KpnI (**Figure 2A**)[50]. NOTE: The pHLsec vector is optimized for the overexpression of soluble, secreted proteins in mammalian cells[25]. This vector contains a secretion signal to allow for the extracellular secretion of soluble glycoproteins. pHLsec contains a C-terminal $(His)_{6x}$ tag to facilitate affinity purification from cell supernatants using immobilized metal affinity chromatography methods.
3. Clone truncated constructs of CD22 ECD with sequential deletions of the C-terminal Ig domains: domains 1-6 (residues 20-687), domains 1-5 (residues 20-592), domains 1-4 (residues 20-504), and domains 1-3 (residues 20-330) (**Figures 2B and 2C**)[50].
4. Evaluate the primary sequence of CD22 ECD using the NetNGlyc server to identify predicted N-linked glycosylation sites present in the construct[51].

5. Using site-directed mutagenesis, by standard protocols[52] or by overlapping PCR[53], mutate each predicted N-linked glycosylation site (Asn to Gln and/or Asn to Ala) to create constructs of CD22 ECD that contain either a single or several N-linked glycosylation mutations.
6. After sequence verification of cloned constructs, transform into competent *E. coli* DH5α cells[54] and maxi-prep the DNA (as per manufacturer's instructions) to prepare for transfection.

## 2. HEK293F and HEK293S Cell Establishment

NOTE: All manipulation of HEK293F or HEK293S cells with necessary reagents and equipment must be performed in a biosafety level 2 facility in a suitable biosafety cabinet. The external surface of all items must be sterilized with a 70% ethanol solution or equivalent reagent.

1. Obtain HEK293F and HEK293S suspension cells (see **Table of Materials**) and store at -80 °C until ready for use.
2. Warm media (see **Table of Materials**) for 1 h in a 37 °C water bath. Transfer 24 mL of warmed media to a 125 mL baffled cell culture flask with a vented cap.
3. Obtain 1 mL cell aliquot from -80 °C and transfer to ice.
4. Incubate cells in a 37 °C water bath for approximately 1 min, to partially thaw the cells. Transfer 1 mL of cells from the vial to the 125 mL baffled cell culture flask containing the media.
5. Close the cell culture flask with the vented cap and place flask in a shaker set to 37 °C, 130 rpm, 70% humidity, and 8% $CO_2$.

## 3. HEK293 Cell Maintenance

NOTE: Cell density and viability of cells must be checked approximately 24 h after thawing. This step ensures that cells are recovering following inoculation; initial viability should be >80%.

1. Carefully remove 10 µL of cells from the 125 mL flask containing the fresh suspension cells and transfer it into a sterile 1.5 mL microtube. Close the flask and return it to the incubator.
2. Pipette 10 µL of Trypan blue solution into the 1.5 mL microtube containing the cells, mix thoroughly and transfer 10 µL to the chamber of the counting slide.
3. Put the counting slide into an automatic cell counter and obtain values for cell density (in units of cells $mL^{-1}$) and cell viability (in percentage).
4. Calculate the volume of cells that would be required to inoculate a fresh 200 mL culture at a final density of ~0.8 x $10^6$ cells $mL^{-1}$ using the following equations:

$$Volume\ of\ cells\ (mL) = \frac{(0.8\ x\ 10^6\ cells\ mL^{-1})(200\ mL)}{density\ of\ live\ cells\ in\ maintenance\ flask\ (cells\ mL^{-1})} \quad (1)$$

$$Volume\ of\ media\ (mL) = 200\ mL - volume\ of\ cells\ (mL) \quad (2)$$

NOTE: It may take ~ 5 d to obtain a suitable cell density for inoculation into a 200 mL culture.

5. Once the cell density is sufficient for inoculation of a 200 mL culture, warm-up media for 1 h in a 37 °C water bath and transfer the warmed media into the biosafety cabinet.
6. Using a serological pipette, carefully transfer the required volume of media (as calculated in equation 2) into a 500 mL baffled cell culture flask with a vented cap.
7. Using a serological pipette, transfer the required volume of suspension cells (calculated in equation 1) into the 500 mL baffled cell culture flask containing the media.
8. Cap the new 200 mL maintenance stock and return it to the incubator. Grow cells to a density of approximately 3 x $10^6$ cells $mL^{-1}$. Passage cells at a density of 0.8 x $10^6$ cells $mL^{-1}$ every 2-3 d to maintain a stable culture of cells (as described in section 3.4-3.7). Do not allow the cells to exceed a density of ~4 x $10^6$ cells $mL^{-1}$.

## 4. Transfection of HEK293 Cells for Glycoprotein Expression

1. Calculate the volume of cells and media that is required for a 200 mL culture for transfection at 0.8 x $10^6$ cells $mL^{-1}$ (using equations 1 and 2 from section 3.4).
   NOTE: The number of 200 mL transfections that can be performed depends on the cell density of the maintenance stock.
2. Transfer the required volume of media and cells for transfection into a new 500 mL cell culture flask with a vented cap and return the cell stock to the incubator.
3. Incubate cells for 1 h prior to transfection to allow cells to acclimatize following splitting.
4. Transfer 50 µg of DNA into a sterile 50 mL conical tube and dilute with 5 mL of media. Vacuum filter the diluted DNA using a 0.22 µm filtration system into another sterile tube.
5. Mix diluted, filtered DNA in a 1:1 mass:volume ratio with transfection reagent. Gently swirl the DNA:transfection reagent solution to mix and incubate the solution at room temperature for 10 min.
6. Add DNA:transfection reagent solution directly to the cells. Incubate transfected cells at 37 °C, 130 rpm, 70% humidity, and 8% $CO_2$ in a shaker for 5-7 d.

## 5. Optimization of Cell Transfection Conditions

NOTE: To optimize cell transfection conditions for maximum glycoprotein yield, transfect cells at a variety of initial cell densities and assess protein yield over time (**Figure 3A**). Transfect cells as described in section 4, at initial cell densities ranging from 0.5 x $10^6$ to 2 x $10^6$ cells $mL^{-1}$ [55]. Trial transfections can be scaled down to 25 mL total volume (in 125 mL baffled cell culture flask) with 6 µg of DNA to save space and reagents. The amount of DNA can also be optimized[55].

1. Every day post-transfection (days 1-7), transfer a 500 µL aliquot from the cell culture into a sterile 1.5 mL microtube (in the biosafety cabinet).

2. Spin aliquoted cells at 12,000 x g for 5 min in a microcentrifuge immediately after collection. Transfer supernatant to a new 1.5 mL microtube and store at 4 °C until all samples are obtained.
3. Quantitate secreted glycoprotein by densitometry
    1. Once all samples are obtained, aliquot 20 µL of each sample into a new 1.5 mL microtube and mix with 6 µL of non-reducing 4x Laemmli sample buffer.
    2. Boil samples for 5 min at 95 °C in a thermo-block. Spin samples for 1 min at 12,000 x g in a microcentrifuge.
    3. Load 20 µL of each sample per well in a 10-well 4-15% gradient SDS-PAGE gel. Include one lane for protein size markers. Run gel at 250 V for 20 min in a Tris/Glycine/SDS buffer.
    4. Following run completion, transfer gel to Coomassie stain (see **Table of Materials**) for 20 min. De-stain gel in $ddH_2O$ for 20 min. Image gel.
    5. Perform densitometry with ImageJ, following standard protocols[57,58].
    6. Compile and plot data with 'days post-transfection' on the X-axis and 'densitometry values' on the Y-axis (**Figure 3A**).
    Note: Alternatively, if protein expression is insufficient for visualization by SDS-PAGE, techniques such as Western blotting may be used[56].

4. Quantitate secreted glycoprotein by BLI
    1. Using Ni-NTA biosensors, quantitate the amount of secreted glycoprotein using BLI[59].
    2. Compile and plot data with 'days post-transfection' on the X-axis and 'protein concentration (µg/mL)' on the Y-axis (**Figure 3A**).

## 6. Purification of Soluble Glycoprotein from HEK293 Supernatant

1. Harvest cells by centrifugation at 6,371 x g for 20 min at 4 °C. Retain supernatant containing secreted CD22 ECD and filter using a 0.22 µm filter.
2. Load supernatant at 4 mL $min^{-1}$ on a pre-equilibrated (20 mM Tris pH 9.0, 150 mM NaCl, 5 mM imidazole) Ni-NTA column (5 mL volume) using a benchtop chromatography system.
NOTE: Other affinity-based purification techniques may be used, based on the affinity tags included in construct design in Section 1.
3. Following supernatant loading, wash the affinity column with 3-4 column volumes (CV) of wash buffer (20 mM Tris pH 9.0, 150 mM NaCl, 5 mM imidazole).
4. Elute purified glycoprotein from the column using a 4-100% gradient (4 CVs) of elution buffer (20 mM Tris pH 9.0, 150 mM NaCl, 500 mM imidazole) while collecting fractions (**Figure 3B**).
5. Pool fractions containing the eluted peak in an centrifugal filtration device with a 10 kDa nominal molecular weight limit (NMWL) and concentrate by centrifuging at 4,000 x g at 4 °C for 15 min or until the sample reaches a volume of 500 µL.
6. Inject concentrated glycoprotein into a 500 µL sample loop and load at 0.5 mL $min^{-1}$ onto a pre-equilibrated (20 mM Tris, pH 9.0, 150 mM NaCl) high-performance size exclusion column (approximately 24 mL volume) on a fast protein liquid chromatography (FPLC) system at 4 °C while collecting fractions (**Figure 3C**).
7. Run SDS-PAGE gel of eluted fractions to identify fractions containing the glycoprotein, and pool corresponding fractions. The SDS-PAGE gel can be run as described in Section 5[60].

## 7. Deglycosylation of Purified Glycoprotein

1. Measure the concentration of purified protein following size exclusion chromatography by using absorbance at 280 nm divided by the extinction coefficient (*e.g.,* 1.418 $M^{-1}$ $cm^{-1}$ for CD22 ECD).
NOTE: The theoretical extinction coefficient of proteins of interest can be calculated using servers such as ExPASy ProtParam[61].
2. Incubate purified protein with Endo H for 1 h at 37 °C, at a ratio of 1 mg of purified protein to 10 µL of commercial enzyme in 1X Endo H buffer (as per manufacturer's instructions).
NOTE: Endo H cleaves high mannose glycans produced in HEK293S leaving a single GlcNAc moiety at each glycosylation site[21]. Endo H does not cleave glycans on proteins produced in HEK293F cells[22], however other enzymes can be used for this purpose (*e.g.,*PNGaseF[24]).
3. Concentrate deglycosylated ECD to 500 µL and run gel filtration chromatography on a high-performance size exclusion column (approximately 24 mL volume) at 0.5 mL $min^{-1}$ on an FPLC to remove Endo H and separate any resulting aggregates.
4. Store the deglycosylated protein at 4 °C until use in downstream experiments.

## 8. Crystallization of Glycoproteins

NOTE: Perform crystallization trials using commercially available screens and set up sitting drop experiments using a crystallization robot.

1. Concentrate pure, deglycosylated ECD to 10 mg $mL^{-1}$ using a centrifugal filtration device with 10 kDa NMWL at 4,000 x g (4 °C) until the desired concentration is obtained.
2. Determine the protein concentration by using absorbance at 280 nm and dividing by the extinction coefficient.
3. Centrifuge sample at 12,000 x g for 5 min at 4 °C prior to crystallization trials to remove unwanted dust or other contaminants from sample.
4. Fill the reservoir wells of 96-well sitting drop crystallization plates with 80 µL of crystallization solution from a commercial crystallization screen.
NOTE: We use sparse-matrix commercial screens that have been designed based on the most successful crystallization conditions with respect to structures deposited in the PDB.
5. Using a crystallization robot, dispense drops into the well of the crystallization plate with a total drop volume of 200 nL at a ratio of purified protein:crystallization solution of 1:1.
6. Once the entire plate has been dispensed, seal the plate with tape and place into a plate imager for inspection by visible and ultraviolet light.

7. Inspect crystallization plates immediately following setup, and in the following weeks, using both visible and ultraviolet light to identify conditions that give initial glycoprotein crystal hits.
8. Further optimize crystals obtained from initial crystallization hits using fine screens based on the condition of the crystal hit or random matrix micro-seeding methods[62,63,64,65].
9. Cryo-protect any crystals lacking sufficient cryo-protectant within the crystallization condition by soaking the crystal in mother liquor solution supplemented with 20% (v/v) glycerol solution (or equivalent cryo-protectant, such as ethylene glycol or polyethylene glycol 400).
10. Mount crystals in cryoloops and flash freeze them in liquid nitrogen prior to data collection on a home source diffractometer or using synchrotron radiation.

## 9. Phasing Using Heavy Atom Derivatization

NOTE: Before any manipulation of HA compounds, safety aspects must be considered. HA compounds used in protein crystallography are selected for their strong affinity to biological molecules and pose risks to human health from prolonged exposure. Take appropriate safety steps for HA compounds as mentioned in their Material Safety Data Sheets.

1. To test different HA compounds, concentrations, and incubation times, reproduce well-diffracting crystals obtained in Section 8 in a 24-well crystallization plate using the hanging-drop vapour diffusion method[66].
2. Decide which HA will be used for crystal derivatization. Servers (*e.g.,* the Heavy-atom Database System[67]) can assist with HA compound selection, ensuring they are appropriate for the protein and crystallization condition.
   NOTE: HA screens are also commercially available for easy screening of most effective HA compounds for phasing. A set of "magic seven" HA compounds have been previously described to have high probability of success for HA derivatization[68].
3. Set up workstation for HA soaking (**Figure 4A**). Using a cryoloop, rapidly transfer crystals to a 0.2 µL drop on a 22 mm cover slip containing HA solution diluted in the crystallization condition such that the final concentration of HA ranges from 1-20 mM. Seal the drop and incubate for different lengths of time (**Figure 4B**). A good starting point is 5, 10, 60, and 90 min and overnight.
4. Visually inspect crystals with a light microscope to identify possible cracks or changes in color, which can indicate adverse effects to the glycoprotein crystal or crystal derivatization.
5. Mount crystals in cryoloops and back-soak crystals for 30 s in three consecutive 0.2 µL drops containing mother liquor solution supplemented with 20% (v/v) glycerol (or alternate cryo-protectant)[69]. Back-soaking the crystals removes HA compound that was non-specifically bound and reduces partial occupancy caused by weak HA binding. Flash freeze crystals in liquid nitrogen (**Figure 4B**).
6. For data collection, processing, structure solution, and refinement, use previously described protocols[26,70,71,72].

## 10. Soaking Glycoprotein Crystals with its Ligand

1. Reproduce well-diffracting crystals obtained in Section 8 in a 24-well crystallization plate using the hanging-drop vapor diffusion method.
2. Prepare a stock solution of 50 mM ligand in 20 mM Tris, pH 9.0, 150 mM NaCl.
   NOTE: The concentration of the ligand should be prepared according to the affinity to its glycoprotein. If the affinity is unknown, it might be required to use a method such as ITC (Section 12.2) to determine the affinity prior to beginning soaking experiments. Ensure that the ligand is soluble at the desired concentration in the required buffer.
3. Add varying concentrations of the ligand to the drop containing ECD crystals and seal the drop for incubation at time lengths ranging between 5 min to 5 d.
4. Visually track crystals with a light microscope to identify possible changes in morphology.
5. Mount crystals in cryoloops and cryo-protect them in mother liquor solution supplemented with 20% (v/v) glycerol (or other cryo-protectant such as ethylene glycol or low molecular weight polyethylene glycol 400)[69].
6. For data collection, processing, structure solution and refinement, use previously described protocols[73,74,75].

## 11. Production of Fragment Antigen Binding (Fab)

1. Subclone genes that correspond to Fab heavy chain (HC) and light chain (LC) sequences of anti-ECD antibodies, *e.g.,*epratuzumab.
   NOTE: Alternatively, IgG can be cleaved by the enzyme papain to generate Fab fragments[76].
2. Transfect cells as described in Section 4, with the following modifications:
   1. Use a total mass of DNA for transfection of Fab fragments of 90 µg per 200 mL of culture.
   2. Transfect HC and LC plasmids in a 2:1 ratio to reduce the amount of LC dimer formation.
3. After 7 d of incubation, harvest cells, retain supernatant and filter with a 0.22 µm vacuum-driven filtration device.
4. Equilibrate anti-LC (kappa or lambda) affinity columns in PBS buffer using a benchtop chromatography system.
   NOTE: If LC dimer formation is an issue during purification, Protein G affinity chromatography can be used as an alternative to kappa/lambda LC affinity purification.
5. Load supernatant on affinity column at 4 mL $min^{-1}$. Following sample loading, wash column with 3-4 CVs of PBS.
6. Elute protein from column using an isocratic elution with 100 mM glycine, pH 2.2, immediately neutralizing the eluted fractions with 10% (v/v) 1 M Tris, pH 9.0 in each fraction.
   NOTE: The eluted Fab can be further purified by ion exchange chromatography and/or size exclusion chromatography using a FPLC at 4 °C.

## 12. Characterization of Fab and Small Molecule Binding to the Glycoprotein

1. Biolayer interferometry
   1. Prepare 50 mL of 1x kinetics buffer (1x PBS, 0.002% (v/v) Tween-20, 0.01% (w/v) BSA).
   2. Hydrate six Ni-NTA biosensors in 200 µL of 1x kinetics buffer for 10 min in a pre-wetting plate.

3. Dilute His-tagged ECD in 1 mL of 1x kinetics buffer at a final concentration of 25 ng $\mu L^{-1}$. Pipette serial dilutions of purified Fab into 200 µL of 1x kinetics buffer, with a high concentration of 500 nM, and subsequent serial dilutions of 250 nM, 125 nM, and 62.5 nM.
4. Aliquot reagents into black flat bottom polypropylene 96-well microplate as shown in **Figure 5A**, where each well contains 200 µL of the indicated solution.
5. Collect data using the Kinetics Assays in the Data Acquisition software, as described previously[38,39,77] (**Figure 5A**).
    1. Briefly, transfer the biosensors into wells containing 1x kinetics buffer to baseline for 60 s before loading 25 ng $\mu L^{-1}$ of glycoprotein for 240 s (or until a threshold of 1.0 nm is reached) at 1,000 rpm.
    2. After a second baseline of 60 s in 1x kinetics buffer, transfer the biosensors into wells containing the serial dilution of Fab. The 180 s association phase is subsequently followed by a 180 s dissociation step in 1x kinetics buffer.
    Note: Biosensors can be reused if the above protocol is followed by a regeneration step, which consists of three cycles of washing the biosensors in stripping buffer (PBS with 500 mM imidazole) for 5 s followed by 5 s in 1x kinetics buffer for neutralization. Biosensors can be reused up to ~10-20 times during the same day, or until poor data quality is observed.
6. Analyze the data using the analysis software (**Figure 5A**):
    1. Under Tab 1, import and select data.
    2. Under Tab 2, Step 1: Data Selection, select 'Sensor Selection' and highlight reference wells (rows E and F, **Figure 5A**), right click and set to reference well. Under Step 2: Subtraction, select 'Reference Wells'. Under Step 3: Align Y-axis, select 'Baseline' from time range of 0.1 to 59.8 s. Under Step 4: Inter-step Correction, select 'Align to dissociation.' Under Step 5: Process, select 'Savitzky-Golay filtering' and press Process Data.
    3. Under Tab 3, select 'Association and Dissociation' under Step to Analyze with a 1:1 model. Select 'Global fitting' and 'Group by Color'. Right click curves, select 'Change color', set all curves to the color of your choosing. Select 'Fit Curves'. If data are well fitted, a report can be exported by selecting 'Save Report'.
7. Repeat experiment with glycoprotein produced in HEK293F and HEK293S cells (Section 5) and following Endo H treatment (Section 7) to assess the effect, if any, of different glycoforms on Fab recognition. Furthermore, repeat the experiment with ECD truncations to provide insight into the domain(s) bound by the Fab.

2. Isothermal titration calorimetry of Fab-glycoprotein interaction
Note: ITC experiments described here are performed using an automated ITC instrument. Experiments are carried out in a 1 mL round bottom 96-well block.
    1. Dialyze the ECD and Fab in a single 4 L beaker of 20 mM Tris, pH 8.0, 150 mM NaCl at 4 °C overnight with a stir bar.
    2. Concentrate dialyzed ECD and Fab to 5 µM and 50 µM, respectively, using a centrifugal filter with 10 kDa NMWL, ensuring to wash the concentrator membranes three times with 5 mL of dialysis buffer at 4,000 x g for 5 min at 4 °C prior to use.
    NOTE: Any mismatch in buffer between samples in the cell and syringe can cause undesired heat to be released during the ITC experiment and result in data of poor quality.
    3. For experiment 1: add 400 µL of ECD to A1 to be loaded in the cell, and 120 µL of the Fab to well A2 to be loaded in the syringe. Well A3 is left empty to return the mixed sample following experiment completion. Each subsequent experiment can be added to the plate in the same order (*i.e.,* experiment 2: cell - A4, syringe - A5, empty well - A6; **Figure 5B**).
    Note: Include buffer into buffer controls (to confirm the instrument is behaving well) at the beginning and end of each run, as well as ligand (in syringe) into buffer (in cell) controls to calculate the heat of dilution for the sample in the syringe. This calculated heat of dilution should then be subtracted from raw experimental data during data analysis (**Figure 5B**).
    4. Run a total of 16 injections with a volume of 2.5 µL for each injection. The injection duration is 5 s, with 180 s spacing between injections. Set the cell temperature to 25 °C, with a stirring speed of 750 rpm and a filter period of 5 s.
    NOTE: Based on the affinity and thermodynamics of the ECD:Fab interaction, it may be necessary to change the sample concentration, number of injections or cell temperature.
    5. Analyze the data with the analysis software, as described previously[40,41,43] (**Figure 5B**).
    6. Repeat the experiment at least in duplicates, calculating mean $K_D$ values and standard errors. Repeat experiment with ECD of different glycoforms (Sections 5 and 6) to assess the effect, if any, of glycoforms on the thermodynamics of the Fab:glycoprotein interaction.

3. For isothermal titration calorimetry of ligand-glycoprotein interactions, set up the ITC experiment as described in Section 12.2, with the following changes:
    1. Dialyze ECD in 4 L of dialysis buffer overnight. Dissolve the ligand using dialysis buffer following completion of dialysis.
    2. Perform ITC experiments at significantly higher concentrations to be able to detect low-affinity interactions. For the ECD and ligand interaction, perform ITC experiments at concentrations of 100 µM of ECD in the cell and 1 mM of ligand in the syringe.

## Representative Results

Several constructs of CD22 ECD were successfully cloned into the pHLsec expression vector and overexpressed in mammalian HEK293F and HEK293S cell lines (**Figure 2 and 3A**). All constructs were purified to size homogeneity by size exclusion chromatography and yielded a highly pure sample for crystallization studies (**Figure 3B and 3C**). The CD22 construct that led to well-diffracting crystals was the d1-d3 truncation (residues 20-330), with five of the six predicted N-linked glycosylation sites mutated from Asn to Ala (N67A, N112A, N135A, N164A and N231A), produced in HEK293S cells, such that only the glycosylation site at position N101 was retained (this construct is named $CD22_{20\text{-}330,\ 5A}$). Crystals were obtained in several conditions of the MCSG-1 sparse-matrix screen, but the best crystals were from a condition containing 30% (w/v) polyethylene glycol 4000, 0.2 M lithium chloride and 0.1 M Tris, pH 8.5. These native crystals diffracted to 2.1 Å resolution; using known structures of Ig domains of related Siglec proteins did not yield any solutions in MR searches.

To acquire phasing information, we soaked native crystals with a panel of HA compounds that included Hg, Pt, Os, Ta, and Br at concentrations ranging from 1-20 mM of HA compound for an incubation time from 5 min to 1 d (**Figure 4**). We monitored crystals for changes in morphology, and found that crystals soaked with HA compound at 20 mM resulted in the rapid cracking and dissolving of the crystal. We froze a total of 63 crystals that retained their shape following set incubation times that were soaked with tantalum bromide cluster, platinum chloride, mercuric acetate and mercuric chloride. Crystals soaked with 7 mM of mercuric chloride for 30 min showed anomalous signal on a fluorescence scan at the Canadian Light Source (CLS) 08-BM beamline (Saskatoon, Canada) and allowed for multi-wavelength anomalous dispersion X-ray data collection on a single crystal. These datasets allowed us to solve the mercury substructure of $CD22_{20\text{-}330,\ 5A}$, which revealed a single mercury atom bound to a free cysteine at position C308 and ultimately allowed us to build the structure of $CD22_{20\text{-}330,\ 5A}$ into the phased electron density map using AutoBuild[78].

Once the unliganded structure was solved, we were interested in solving the structure of CD22 bound to its ligand, α2-6 siallylactose. We first calculated the affinity of CD22 towards α2-6 sialyllactose using ITC to characterize the binding thermodynamics of the interaction. We observed an affinity of ~280 μM and used this information to identify an initial concentration (~ 100 x $K_D$) of ligand to use for soaking of our native $CD22_{20\text{-}330,5A}$ crystals. We soaked the $CD22_{20\text{-}330,\ 5A}$ crystals with 25 mM siallylactose for 5 min, 2 h, 14 h, 40 h and 5 d and monitored for changes in crystal morphology. A total of ~75 crystals were frozen from various time points and sent to the CLS synchrotron beamline 08-ID (Saskatoon, Canada) for remote data collection. A total of six X-ray datasets were collected from well-diffracting crystals. The structure from each X-ray dataset was solved by MR using the unliganded $CD22_{20\text{-}330,\ 5A}$ structure as an initial search model. The resulting electron density for all datasets was then inspected for positive density in the Fo-Fc map that would correspond to bound α2-6 sialyllactose within the binding site of CD22. Remarkably, all datasets collected, even those from crystals soaked after only 5 min of incubation time, contained positive density corresponding to the ligand in the binding site. Overall structures of the unliganded and liganded CD22 were highly similar with minimal conformational changes, which might explain the success of soaking experiments with α2-6 sialyllactose.

We next characterized the antigenic surface of CD22 recognized by therapeutic antibody epratuzumab in BLI and ITC experiments (**Figure 5**). Kinetics and thermodynamics profiles of epratuzumab Fab binding to CD22 constructs with different glycoforms revealed an increasing affinity to CD22 with reduced N-linked glycan size, with up to a 14-fold improvement in affinity for smaller glycans (327 nM vs 24 nM in BLI; 188 nM vs 58 nM in ITC). The CD22 N-linked glycan restricting access of the antibody to its epitope was identified by BLI using single-point mutants of CD22 and by solving the epratuzumab Fab-CD22 d1-d3 co-crystal structure[28].

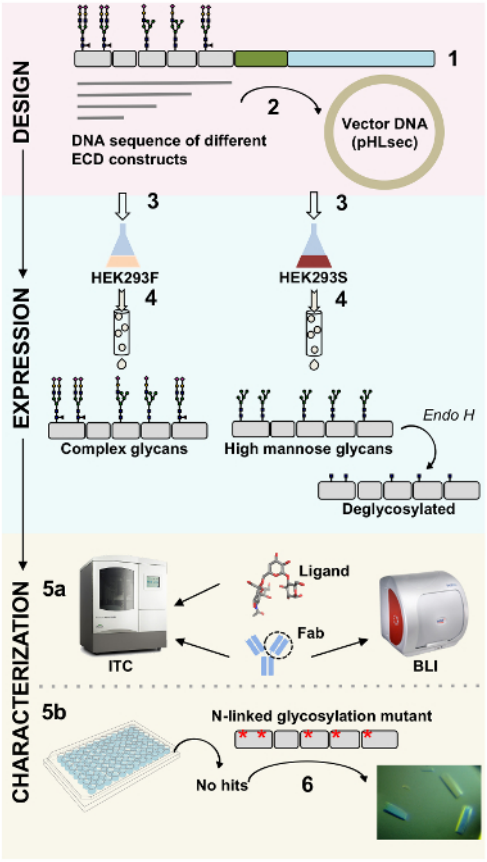


**Figure 1. Overview of glycoprotein characterization from construct design to biophysical and structural characterization.** (1) Primary sequence analysis of representative glycoprotein. In grey, the extracellular domain (ECD); in green, the transmembrane (TM) segment; and in blue, the cytosolic domain of the glycoprotein. Predicted N-linked glycans are labeled. (2) Cloning of ECD constructs. (3) Expression of ECD constructs in mammalian cells. (4) Glycoprotein purification. While proteins expressed in HEK293F will contain complex glycans, proteins expressed in HEK293S will have high mannose glycans. Enzymatic treatment of glycoproteins produced in HEK293S cells with Endo H results in glycoproteins with only a GlcNAc moiety at N-linked glycosylation sites. (5a) Glycoproteins are tested for their binding to antibodies by biolayer interferometry (BLI) and isothermal titration calorimetry (ITC). Affinity to small ligands can also be measured by ITC. (5b) Crystallization trials of glycoproteins with homogeneous N-linked glycans, such as those expressed in HEK293S and deglycosylated with Endo H. (6) In some cases, mutation of N-linked glycosylation sites is necessary to obtain crystals. Please click here to view a larger version of this figure.

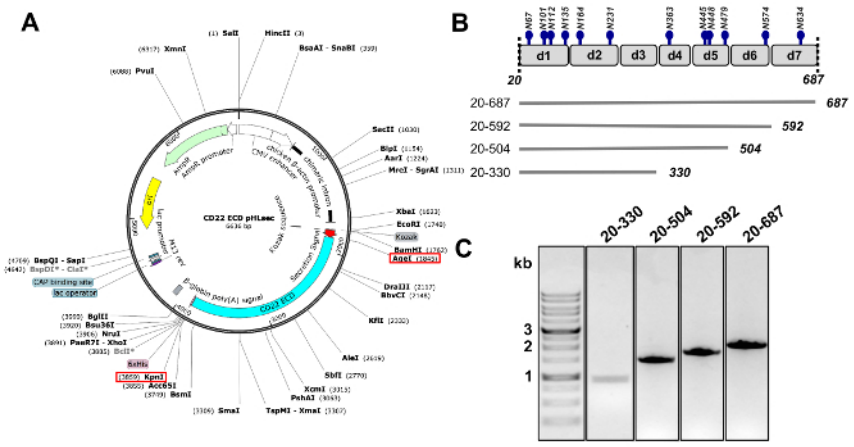


**Figure 2. Design of CD22 ectodomain DNA constructs for expression in mammalian cells. A)** Representation of the pHLsec plasmid used for transient transfection of CD22 ECD constructs. AgeI and KpnI sites used for cloning are indicated with red boxes. **B)** The CD22 ECD contains seven Ig domains (d1-d7) and 12 predicted N-linked glycosylation sites (in blue). Four constructs were designed from the CD22 ECD. **C)** 1% agarose gel showing PCR amplicons of CD22 ECD constructs for cloning into the pHLsec mammalian expression vector. First lane contains 1 kb DNA marker. Please click here to view a larger version of this figure.

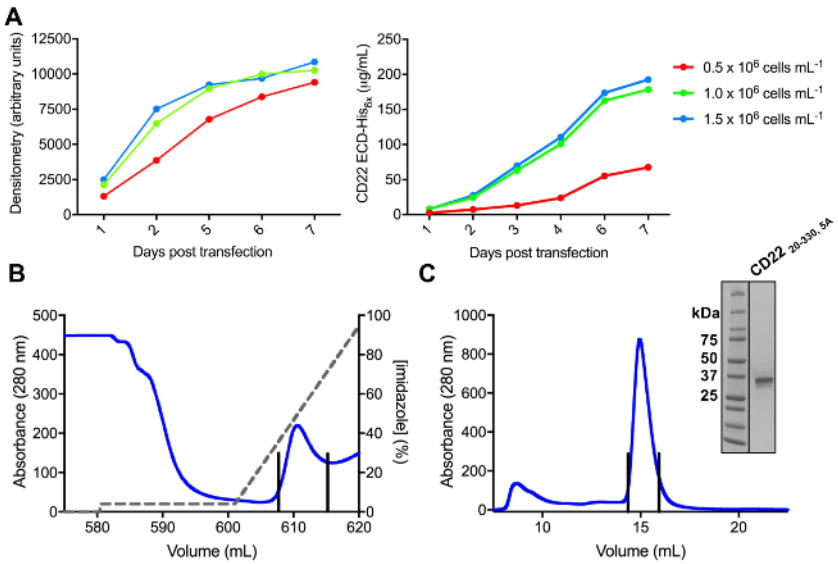


**Figure 3. Expression and purification of glycoproteins. A)** Effect of cell density on expression yields. Glycoprotein expression in small-scale 25 mL culture of HEK293F suspension cells transfected using three different starting densities of cells (0.5 x $10^6$ cells $mL^{-1}$, 1.0 x $10^6$ cells $mL^{-1}$, and 1.5 x $10^6$ cells $mL^{-1}$). Quantification performed by densitometry from SDS-PAGE in left panel and by quantitative BLI in right panel. Values are representative of one glycoprotein preparation. **B)** Chromatogram of the first purification step for construct $CD22_{20\text{-}330,\ 5A}$ from 600 mL of supernatant using a Ni-NTA affinity column. The glycoprotein was eluted using a gradient of imidazole (grey line), where 100% corresponds to the elution buffer, which contains 500 mM imidazole. Pooled fractions are depicted with vertical lines. **C)** Size-exclusion chromatogram for construct $CD22_{20\text{-}330,5A}$ using a high-performance gel filtration column. Pooled fractions from the elution peak are depicted with vertical lines. Inset: Coomassie-stained SDS-PAGE gel showing the purity of the glycoprotein. Please click here to view a larger version of this figure.

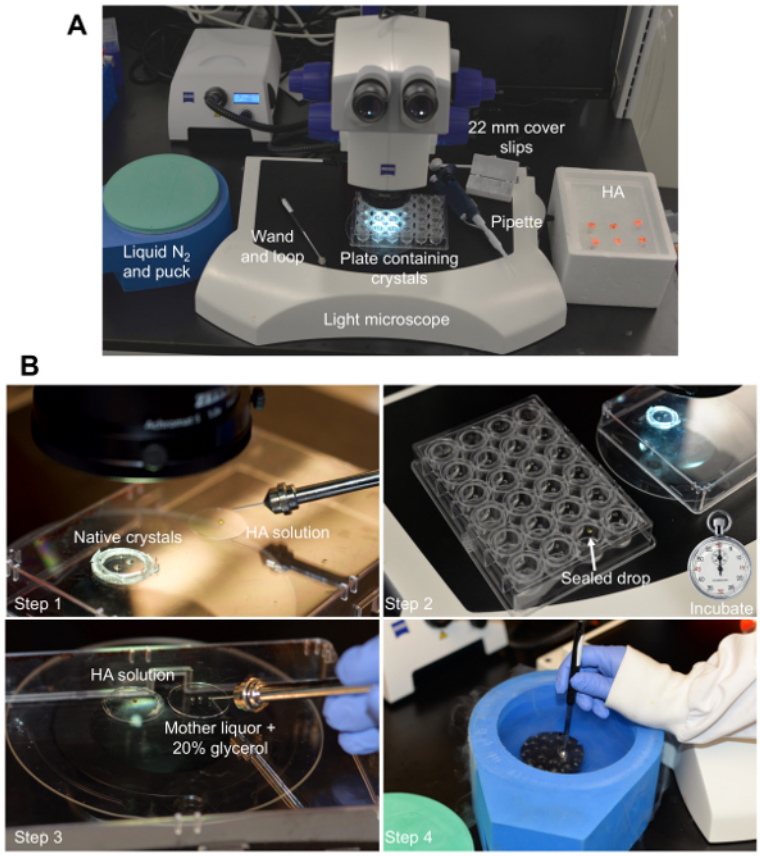


**Figure 4. Crystal soaking with heavy atoms.** A) Sample workstation for soaking native crystals with HA compounds. All required tools are labeled. B) Steps followed to soak crystals of construct $CD22_{20\text{-}330,\ 5A}$ with HA compounds. *Step 1*, Open well containing crystals, and transfer crystals using a loop to a 0.2 µL drop on a cover slip containing HA solution diluted in the crystallization condition such that the final concentration of HA ranges from 1-10 mM. *Step 2*, seal the drop in the crystallization plate and incubate crystals with the HA compound for different periods of time. *Step 3*, mount the soaked crystal in the loop and back-soak for 30 s in three consecutive 0.2 µl drops containing the mother liquor solution supplemented with 20% (v/v) glycerol dispensed on a cover slip. *Step 4*, flash freeze the crystal mounted on a loop with liquid nitrogen and place it in a puck for shipment to the synchrotron beamline. Please click here to view a larger version of this figure.

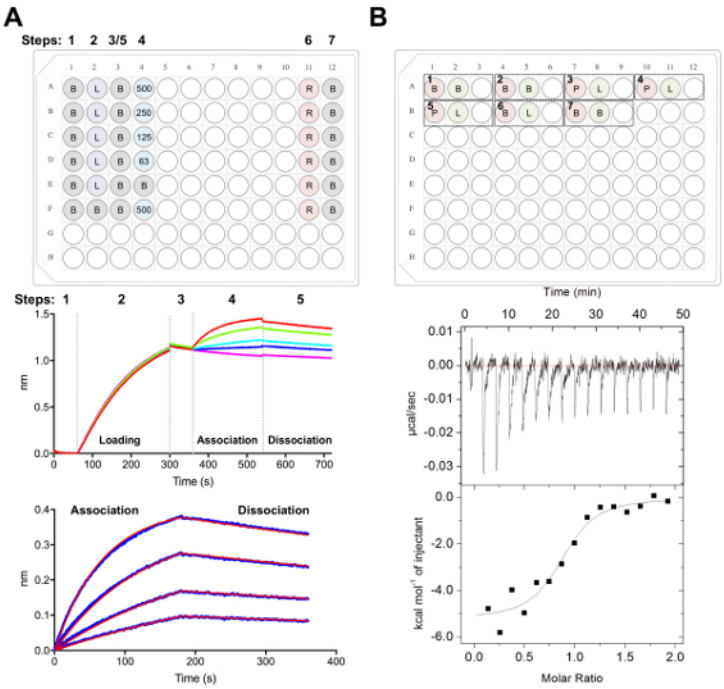


**Figure 5. Biolayer interferometry and isothermal titration calorimetry measurements. A)** Representative BLI experiment. *Top panel:* example of plate setup for a kinetics experiment, where the following are labeled: 1x kinetics buffer (B), $His_{6x}$-tagged glycoprotein loading (L), representative Fab concentrations (500, 250, 125, 62.5 nM), PBS + 500 mM regeneration buffer (R), and 1x kinetics neutralization buffer (B). Each well contains 200 µL of solution. Step number for the kinetics experiment is indicated at the top of the plate. *Middle panel:* Representative raw data of BLI experiment performed using Ni-NTA biosensors and the plate described in the top panel. Step numbers correspond to baseline (1), $His_{6x}$ glycoprotein loading (2), baseline (3), association in serial dilution of Fab (4) and dissociation (5). Regeneration steps are not represented (Steps 6-7). *Bottom panel:* Representative analyzed data showing raw association and dissociation (blue line) with the corresponding 1:1 fit (red line). **B)** *Top panel:* Representative plate setup for a single ITC run on an automated ITC instrument with seven experiments in a 96-well round bottom block. Each experiment is comprised of three wells. The first well (red) corresponds to sample for the cell (400 µL), the second well (green) corresponds to sample for the syringe (120 µL). The third well is left empty, and the mixed samples will be returned to this well following experiment completion. Experiments 1, 2, and 7 are buffer into buffer controls. Experiments 3-5 represent triplicate experiments with glycoprotein (P) in the cell and Fab or ligand (L) in the syringe. Experiment 6 represents a ligand heat of dilution control and should be subtracted from experiments 3-5 during data analysis. *Bottom panel:* Representative raw (top) and processed (bottom) ITC data showing Fab (epratuzumab) binding to CD22 ECD produced in HEK293F cells. Please click here to view a larger version of this figure.

## Discussion

Membrane-anchored glycoproteins are critical for cell function and are attractive therapeutic targets. Here, we present a protocol for the structural and biophysical characterization of the ECD of membrane glycoproteins, both alone and in complex with small molecule ligands and Fab fragments. We have successfully used this protocol to determine the crystal structure of the three N-terminal-most Ig domains of the extracellular portion of human CD22[28], a critical co-receptor on B cells involved in keeping humoral immunity in check[79]. We have also characterized the binding site of CD22 with its natural ligand α2-6 sialyllactose, and defined the mode of recognition of a therapeutic antibody towards human CD22. These results provide insights into the structure-function relationship of a key member of the Siglecs family that has restricted expression on B cells, and a molecular roadmap towards the development of new CD22 targeted small molecule and antibody-based therapeutics. While this protocol was used successfully for an Ig-containing B cell receptor, we propose that our approach can be applied for the structural and biophysical characterization of any membrane glycoprotein with a distinct domain organization. In such cases, construct design and combinatorial N-linked glycan mutations (either to Gln or Ala) can be evaluated to find a construct suitable for crystal growth and high-resolution diffraction.

Obtaining a homogeneous and pure glycoprotein sample is of critical importance for crystal growth and X-ray diffraction, as well as for downstream biophysical characterization. N-linked glycans present on glycoproteins are inherently heterogeneous, and can cause conformational and chemical heterogeneity within the glycoprotein that can deter crystal formation. To reduce this micro-heterogeneity, strategies that introduce point mutations to remove Asn residues predicted to harbor N-linked glycans, or using mutant cell lines (such as HEK293S) followed by treatment with endoglycosidases (such as EndoH) can considerably improve crystallization success[15,21,22]. In this protocol, we discuss the purification of soluble glycoproteins and Fabs that are secreted into the cell supernatant. Glycoprotein secretion provides a relatively simple route towards purity, without the need for cell lysis or the addition of harsh chemicals or detergents. The cell supernatant, obtained following cell harvesting is then run directly over a column that has affinity for the protein of interest (*e.g.,* Ni-NTA for His-tagged glycoproteins, or LC affinity for Fab fragments). However, depending on the column of use and the conditions of the cell supernatant (*e.g.,* pH), the binding capability of the protein of interest to the column may be affected. If this is the case, it may be necessary to concentrate and buffer exchange the cell supernatant to improve binding to the column. Furthermore, it is strongly recommended that quality control steps during purification be employed to help assess protein purity. Running an SDS-PAGE gel or Western blot of all samples (before, during and after purification steps)

can yield insights into whether the proposed purification scheme is suitable for the protein of interest. If contaminating bands are visible on SDS-PAGE, or if several species are obtained during purification (*e.g.,* several peaks on size exclusion), additional purification steps should be considered, *e.g.,* ion exchange chromatography, to gain in purity and increase chances of downstream crystallization[80].

For macromolecular crystallization, it is often critical to obtain high yields of the protein of interest to allow for the screening of a large number of potential crystallization conditions at high protein concentrations to find suitable crystal hits. Generally, the HEK293 cell lines discussed here (HEK293F and HEK293S) are robust expression systems, and can be easily scaled up to produce more sample as necessary. However, it is possible that the protein of interest may not express sufficiently within these cell lines. In these cases, other cell lines, such as Expi293 cells[81,82], have been found to show superior levels of protein expression and should be considered as an alternative.

If well-ordered, diffracting crystals are not obtained following testing of several constructs of the protein of interest despite high purity, it may be necessary to expand crystallization techniques to promote crystal formation. It has been shown that Fab fragments of antibodies and nanobodies can be excellent crystallization enhancers, and promote well-ordered crystal packing[83,84,85]. These fragments can be expressed and purified to homogeneity, and used in a complex with the protein of interest to promote crystallization. Importantly, Fab fragments produced as described in Section 10 can have a tendency to form non-functional LC dimers[86]. These dimers are contaminants and should be removed during purification. In our experience, LC dimers often have a different retention volume on size exclusion, or elute as a distinct peak on ion exchange chromatography, and thus can be removed from the Fab purification - however this is not always the case. If these techniques are insufficient to remove LC dimers from the Fab purification, additional purification methods, such as Protein G affinity purification, can be employed to improve purity.

Alternative to co-complexation with Fab fragments, well-documented techniques such as random matrix microseeding can improve chances of obtaining well-ordered crystals[63,70]. This method involves the addition of small amounts of crushed, suboptimal crystals into the crystallization condition, providing a crystal nucleate to promote crystal growth. This can be performed using crystals of the protein of interest, or ones with similar domain architecture and tertiary structure. Furthermore, random matrix microseeding can be performed in attempts to crystallize the protein alone, or in complex with a Fab fragment or small molecule of interest. Recent advances in cryo-electron microscopy also make this technique an attractive alternative to X-ray crystallography for obtaining high-resolution structural information for molecules with appropriate features[87,88,89,90,91].

When phasing of X-ray diffraction datasets fails by MR, HA soaking may be needed to solve the phase problem by anomalous dispersion or isomorphous replacement. Inspection of the amino acid sequence of the protein can provide clues about the strategy for HA derivatization, including optimum pH for binding. In particular, unpaired cysteines within the protein can specifically bind HA compounds that contain mercury. Soaking native crystals with HA compounds is an iterative process to determine the identity of the optimal HA compound, its concentration, and the required incubation time. If initial soaking attempts do not yield well-diffracting crystals containing a HA suitable for phasing, it may be necessary to introduce amino acid substitutions to improve the probability of HA binding and improve anomalous signal. Examples include mutations to include a free cysteine residue to efficiently bind Hg, Au, Pt or Pb. Expression of proteins for anomalous phasing in a seleno-methionine supplemented media in *E. coli* is extensively used for anomalous phasing, however an equivalent system that reliably incorporates seleno-methionine is not readily available for mammalian cells in suspension[92,93], and is an area of future development.

Once the unliganded structure of the glycoprotein of interest is obtained, soaking the crystals with small molecule ligands can be performed to obtain a structure of the immune receptor-ligand complex. These data provide a blueprint for the rational design of more specific and high-affinity ligands that can be used as small-molecule therapeutics as well as provides high resolution insights into the biological function of the glycoprotein. When attempting to soak glycoprotein crystals with small molecule ligands of interest, inspection of the unliganded crystal structure can indicate whether soaking should be possible. If close crystal-packing contacts are found around the ligand-binding site or around regions expected to undergo conformational changes upon ligand binding, soaking will likely be problematic. In this case, other methods such as co-crystallization of the protein-ligand complex should be performed.

## Disclosures

The authors declare no competing interests.

## Acknowledgements

X-ray diffraction experiments described in this paper were performed using beamlines 08-ID and 08-BM at the Canadian Light Source, which is supported by the Canada Foundation for Innovation, Natural Sciences and Engineering Research Council of Canada, the University of Saskatchewan, the Government of Saskatchewan, Western Economic Diversification Canada, the National Research Council Canada, and the Canadian Institutes of Health Research. We would like to acknowledge the Structural & Biophysical Core Facility, The Hospital for Sick Children, for access to the ITC and BLI instruments. J.E.O. was supported by Banting Postdoctoral Fellowship BPF-144483 from the Canadian Institutes of Health Research. T.S. is a recipient of a Canada Graduate Scholarship Master's Award and a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research. This work was supported by operating grant PJT-148811 (J.-P.J.) from the Canadian Institutes of Health Research. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program (J.-P.J.).

## References

1. Sachs, J.N., Engelman, D.M. Introduction to the membrane protein reviews: The interplay of structure, dynamics, and environment in membrane protein function. *Annu Rev Biochem.* **75** (1), 707-712 (2006).
2. Cournia, Z. *et al.* Membrane protein structure, function, and dynamics: A perspective from experiments and theory. *J Membr Biol.* **248** (4), 611-40 (2015).

3. Macauley, M.S. *et al.* Antigenic liposomes displaying CD22 ligands induce antigen-specific B cell apoptosis. *J Clin Invest.* **123** (7), 3074-3083 (2013).
4. Hyde, C.A.C. *et al.* Targeting extracellular domains D4 and D7 of vascular endothelial growth factor receptor 2 reveals allosteric receptor regulatory sites. *Mol Cell Biol.* **32** (19), 3802-3813 (2012).
5. Tai, W., Mahato, R., Cheng, K. The role of HER2 in cancer therapy and targeted drug delivery. *J Control Release.* **146** (3), 264-275 (2010).
6. Zarei, O., Benvenuti, S., Ustun-Alkan, F., Hamzeh-Mivehroud, M., Dastmalchi, S. Strategies of targeting the extracellular domain of RON tyrosine kinase receptor for cancer therapy and drug delivery. *J Cancer Res Clin Oncol.* **142** (12), 2429-2446 (2016).
7. Rosman, Z., Shoenfeld, Y., Zandman-Goddard, G. Biologic therapy for autoimmune diseases: an update. *BMC Med.* **11** (1), 88 (2013).
8. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature.* **409** (6822), 860-921 (2001).
9. Barclay, A.N. Membrane proteins with immunoglobulin-like domains - A master superfamily of interaction molecules. *Semin Immunol.* **15** (4), 215-223 (2003).
10. Barclay, A.N. Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition. *Proc Natl Acad Sci.* **96** (26), 14672-14674 (1999).
11. Aebi, M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta - Mol Cell Res.* **1833** (11), 2430-2437 (2013).
12. Ohtsubo, K., Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell.* **126** (5), 855-867 (2006).
13. Lodish, H., Berk, A., Zipursky, S., Al, E. Glycosylation in the ER and Golgi complex. *Mol Cell Biol.* (4), Section 17.7 (2000).
14. Thomas, P., Smart, T.G. HEK293 cell line: A vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods.* **51** (3), 187-200 (2005).
15. Lee, J.E., Fusco, M.L., Ollmann Saphire, E. An efficient platform for screening expression and crystallization of glycoproteins produced in human cells. *Nat Protoc.* **4** (4), 592-604 (2009).
16. Betenbaugh, M.J., Tomiya, N., Narang, S., Hsu, J.T.A., Lee, Y.C. Biosynthesis of human-type N-glycans in heterologous systems. *Curr Opin Struct Biol.* **14** (5), 601-606 (2004).
17. Yang, Z. *et al.* Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat Biotechnol.* **33** (8), 842-844 (2015).
18. Bláha, J., Kalousková, B., Skořepa, O., Pažický, S., Novák, P., Vaněk, O. High-level expression and purification of soluble form of human natural killer cell receptor NKR-P1 in HEK293S GnTI-cells. *Protein Expr Purif.* **140**, 36-43 (2017).
19. Bláha, J., Pachl, P., Novák, P., Vaněk, O. Expression and purification of soluble and stable ectodomain of natural killer cell receptor LLT1 through high-density transfection of suspension adapted HEK293S GnTI- cells. *Protein Expr Purif.* **109**, 7-13 (2015).
20. Chaudhary, S., Pak, J.E., Gruswitz, F., Sharma, V., Stroud, R.M. Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. *Nat Protoc.* **7** (3), 453-466 (2012).
21. Chang, V.T. *et al.* Glycoprotein structural genomics: Solving the glycosylation problem. *Structure.* **15** (3), 267-273 (2007).
22. Davis, S.J., Crispin, M. Solutions to the glycosylation problem for low- and high-throughput structural glycoproteomics. *Funct Struct Proteomics Glycoproteins.* 127-158 (2011).
23. Elbein, A.D., Tropea, J.E., Mitchell, M., Kaushal, G.P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J Biol Chem.* **265** (26), 15599-15605 (1990).
24. Zheng, K., Bantog, C., Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs.* **3** (6), 568-576 (2011).
25. Aricescu, A.R., Lu, W., Jones, E.Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr Sect D Biol Crystallogr.* **62** (10), 1243-1250 (2006).
26. Adams, P.D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods.* **55** (1), 94-106 (2011).
27. May, A.P., Robinson, R.C., Vinson, M., Crocker, P.R., Jones, E.Y. Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol Cell.* **1** (5), 719-728 (1998).
28. Ereño-Orbea, J. *et al.* Molecular basis of human CD22 function and therapeutic targeting. *Nat Commun.* **8** (1), 764 (2017).
29. Yu, X.-L. *et al.* Crystal structure of HAb18G/CD147: implications for immunoglobulin superfamily homophilic adhesion. *J Biol Chem.* **283** (26), 18056-18065 (2008).
30. Garman, E., Murray, J.W. Heavy-atom derivatization. *Acta Crystallogr - Sect D Biol Crystallogr.* **59** (11), 1903-1913 (2003).
31. Agniswamy, J., Joyce, M.G., Hammer, C.H., Sun, P.D. Towards a rational approach for heavy-atom derivative screening in protein crystallography. *Acta Crystallogr Sect D Biol Crystallogr.* **64** (4), 354-367 (2008).
32. Rose, J.P., Wang, B.C., Weiss, M.S. Native SAD is maturing. *IUCrJ.* **2** (20), 431-440 (2015).
33. Olieric, V. *et al.* Data-collection strategy for challenging native SAD phasing. *Acta Crystallogr Sect D Struct Biol.* **72** (3), 421-429 (2016).
34. Rillahan, C.D. *et al.* Disubstituted sialic acid ligands targeting Siglecs CD33 and CD22 associated with myeloid leukaemias and B cell lymphomas. *Chem Sci.* **5** (6), 2398-2406 (2014).
35. Mesch, S. *et al.* From a library of MAG antagonists to nanomolar CD22 ligands. *ChemMedChem.* **7** (1), 134-143 (2012).
36. Chiu, M.L., Gilliland, G.L. Engineering antibody therapeutics. *Curr Opin Struct Biol.* **38**, 163-73 (2016).
37. Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V., Kayser, V. The state-of-play and future of antibody therapeutics. *Adv Drug Deliv Rev.* **122** (2016), 2-19 (2017).
38. Yang, D., Singh, A., Wu, H., Kroe-Barrett, R. Determination of high-affinity antibody-antigen binding kinetics using four biosensor platforms. *J Vis Exp.* (122), e55659 (2017).
39. Kamat, V., Rafique, A. Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Anal Biochem.* **536**, 16-31 (2017).
40. Brautigam, C.A., Zhao, H., Vargas, C., Keller, S., Schuck, P. Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat Protoc.* **11** (5), 882-894 (2016).
41. Duff, Jr., M.R., Grubbs, J., Howell, E.E. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp.* (55), e2796 (2011).
42. Livingstone, J.R. Antibody characterization by isothermal titration calorimetry. *Nature.* **384** (6608), 491-492 (1996).
43. Freyer, M.W., Lewis, E.A. Isothermal titration calorimetry: Experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.* **84**, 79-113 (2008).
44. Macauley, M.S., Crocker, P.R., Paulson, J.C. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol.* **14** (10), 653-666 (2014).
45. Zaccai, N.R. *et al.* Structure-guided design of sialic acid-based Siglec inhibitors and crystallographic analysis in complex with sialoadhesin. *Structure.* **11** (5), 557-567 (2003).

46. Pantophlet, R. *et al.* Bacterially derived synthetic mimetics of mammalian oligomannose prime antibody responses that neutralize HIV infectivity. *Nat Commun.* **8** (1), 1601 (2017).
47. Leonard, J.P. *et al.* Epratuzumab, a humanized anti-CD22 antibody, in aggressive non-Hodgkin's lymphoma: phase I/II clinical trial results. *Clin Cancer Res.* **10** (16), 5327-5334 (2004).
48. Finn, R.D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190-D199 (2017).
49. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* **10** (6), 845-858 (2015).
50. Lessard, J.C. Molecular cloning. *Methods Enzymol.* **529**, 85-98 (2013).
51. Gupta, R., Jung, E., Brunak, S. *NetNGlyc: Prediction of N-glycosylation sites in human proteins.* (2004).
52. Liu, H., Naismith, J.H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
53. Heckman, K.L., Pease, L.R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc.* **2** (4), 924-932 (2007).
54. Froger, A., Hall, J.E. Transformation of plasmid DNA into E. coli using the heat shock method. *J Vis Exp.* (6), e253 (2007).
55. Akula, I., Julien, J.-P. *Optimization of glycoprotein expression by transient transfection in HEK293 F/S suspension cells.* at <https://www.polyplus-transfection.com/wp-content/uploads/2015/09/FectoPRO-Technical-Note-031716.pdf> (2015).
56. Taylor, S.C., Berkelman, T., Yadav, G., Hammond, M. A defined methodology for reliable quantification of western blot data. *Mol Biotechnol.* **55** (3), 217-226 (2013).
57. Tan, H.Y., Ng, T.W. Accurate step wedge calibration for densitometry of electrophoresis gels. *Opt Commun.* **281** (10), 3013-3017 (2008).
58. Gassmann, M., Grenacher, B., Rohde, B., Vogel, J. Quantifying Western blots: pitfalls of densitometry. *Electrophoresis.* **30** (11), 1845-1855 (2009).
59. Jonnalgadda, K., Markley, L., Estes, S., Prajapati, S., Takkar, R., Kumaraswamy, S. *Rapid, reliable quantitation of Fc-fusion protein in cell culture supernatants.* at <https://www.fortebio.com/documents/ForteBio_App_Note_13.pdf> (2018).
60. JoVE Science Education Database. Basic methods in cellular and molecular biology: Separating protein with SDS-PAGE. *J Vis Exp.* (2018).
61. Wilkins, M.R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* **112**, 531-552 (1999).
62. Till, M. *et al.* Improving the success rate of protein crystallization by random microseed matrix screening. *J Vis Exp.* (78), e50548 (2013).
63. Obmolova, G., Malia, T.J., Teplyakov, A., Sweet, R., Gilliland, G.L. Promoting crystallization of antibody-antigen complexes via microseed matrix screening. *Acta Crystallogr Sect D Biol Crystallogr.* **66** (8), 927-933 (2010).
64. Arcy, A., Bergfors, T., Cowan-Jacob, S.W., Marsh, M. Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr Sect F, Struct Biol Commun.* **70** (9), 1117-1126 (2014).
65. Luft, J.R. *et al.* Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein Sci.* **16** (4), 715-722 (2007).
66. Dessau, M.A., Modis, Y. Protein crystallization for X-ray crystallography. *J Vis Exp.* (47), e2285 (2011).
67. Sugahara, M., Asada, Y., Ayama, H., Ukawa, H., Taka, H., Kunishima, N. Heavy-atom Database System: A tool for the preparation of heavy-atom derivatives of protein crystals based on amino-acid sequence and crystallization conditions. *Acta Crystallogr D Biol Crystallogr.* **61** (9), 1302-1305 (2005).
68. Boggon, T.J., Shapiro, L. Screening for phasing atoms in protein crystallography. *Structure.* **8** (7), R143-R149 (2000).
69. Vera, L., Stura, E.A. Strategies for protein cryocrystallography. *Cryst Growth Des.* **14** (2), 427-435 (2014).
70. Pichlo, C., Montada, A.A., Schacherl, M., Baumann, U. Production, crystallization and structure determination of C. difficile PPEP-1 via microseeding and Zinc-SAD. *J Vis Exp.* (118), e55022 (2016).
71. Leslie, A.G.W. *et al.* Automation of the collection and processing of X-ray diffraction data - a generic approach. *Acta Crystallogr Sect D Biol Crystallogr.* **58** (11), 1924-1928 (2002).
72. Pike, A.C.W., Garman, E.F., Krojer, T., Von Delft, F., Carpenter, E.P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr Sect D Struct Biol.* **72** (3), 303-318 (2016).
73. Cooper, D.R., Porebski, P.J., Chruszcz, M., Minor, W. X-ray crystallography: Assessment and validation of protein-small molecule complexes for drug discovery. *Expert Opin Drug Discov.* **6** (8), 771-782 (2011).
74. Hassell, A.M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr Sect D Biol Crystallogr.* **63** (1), 72-79 (2006).
75. Müller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallogr Sect D Struct Biol.* **73** (2), 79-92 (2017).
76. Zhao, Y. *et al.* Two routes for production and purification of Fab fragments in biopharmaceutical discovery research: Papain digestion of mAb and transient expression in mammalian cells. *Protein Expr Purif.* **67** (2), 182-189 (2009).
77. Shah, N.B., Duncan, T.M. Bio-layer interferometry for measuring kinetics of protein-protein interactions and allosteric ligand effects. *J Vis Exp.* (84), e51383 (2014).
78. Terwilliger, T.C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr Sect D Biol Crystallogr.* **64** (1), 61-69 (2008).
79. Walker, J.A., Smith, K.G.C. CD22: An inhibitory enigma. *Immunology.* **123** (3), 314-325 (2008).
80. Gräslund, S. *et al.* Protein production and purification. *Nat Methods.* **5** (2), 135-146 (2008).
81. Jain, N.K. *et al.* A high density CHO-S transient transfection system: Comparison of ExpiCHO and Expi293. *Protein Expr Purif.* **134**, 38-46 (2017).
82. Fang, X.T., Sehlin, D., Lannfelt, L., Syvänen, S., Hultqvist, G. Efficient and inexpensive transient expression of multispecific multivalent antibodies in Expi293 cells. *Biol Proced Online.* **19** (1), 11 (2017).
83. Löw, C., *et al.* Nanobody mediated crystallization of an archeal mechanosensitive channel. *PLoS One.* **8** (10), e77984 (2013).
84. Hunte, C., Michel, H. Crystallization of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol.* **12** (4), 503-508 (2002).
85. Ereño-Orbea, J., Sicard, T., Cui, H., Carson, J., Hermans, P., Julien, J.-P. Structural basis of enhanced crystallizability induced by a molecular chaperone for antibody antigen-binding fragments. *J Mol Biol.* **430** (3), 322-336 (2018).
86. Spooner, J. *et al.* Evaluation of strategies to control Fab light chain dimer during mammalian expression and purification: A universal one-step process for purification of correctly assembled Fab. *Biotechnol Bioeng.* **112** (7), 1472-1477 (2015).
87. Elmlund, D., Le, S.N., Elmlund, H. High-resolution cryo-EM: The nuts and bolts. *Curr Opin Struct Biol.* **46**, 1-6 (2017).
88. Merk, A. *et al.* Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell.* **165** (7), 1698-1707 (2016).
89. Bai, X. chen, McMullan, G., Scheres, S.H.W. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci.* **40** (1), 49-57 (2015).
90. Wlodawer, A., Li, M., Dauter, Z. High-resolution cryo-EM maps and models: A crystallographer's perspective. *Structure.* **25** (10), 1589-1597 (2017).

91. Bartesaghi, A. *et al.* 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science.* **348** (6239), 1147-1151 (2015).
92. Hendrickson, W. a, Horton, J.R., LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J.* **9** (5), 1665-1672 (1990).
93. Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallogr Sect D Biol Crystallogr.* **66** (4), 352-357 (2010).