Computational and Structural Biotechnology Journal 18 (2020) 3925–3935

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Computational insights into O-glycosylation in a CTLA4 Fc-fusion protein linker and its impact on protein quality attributes

Yuanli Song, Yueming Qian *, Zhe Huang, Sarwat F. Khattak, Zheng Jian Li

Biologics Development, Bristol Myers Squibb Company, 38 Jackson Road, Devens, MA 01434, USA

ARTICLE INFO

Article history:
Received 7 September 2020
Received in revised form 21 November 2020
Accepted 23 November 2020
Available online 01 December 2020

Keywords:
O-glycosylation
Protein linker
Fc-fusion protein
Sialic acid
Protein aggregation

ABSTRACT

The hinge region of immunoglobulin G1 (IgG1) is used as a common linker for Fc-fusion therapeutic proteins. With the advances of high-resolution mass spectrometry and sample treatment strategies, unexpected O-linked glycosylation has been observed in the linker. However, the molecular mechanism involved in this unusual posttranslational modification is unknown. In this study, we applied site-direct mutagenesis, mass spectrometry, analytical chromatography, and computational modeling to investigate O-glycosylation processes in a clinically used CTLA4 Fc-fusion protein and its impacts on protein quality attributes. Surprisingly, O-glycans could be formed at new sites when an initial O-glycosylation site was eliminated, and continued to occur until all potential O-glycosylation sites were nulled. Site-preference of O-glycosylation initiation was attributed to the complex formation between the linker peptide and glycan transferase whereas the O-glycosylating efficiency and the linker flexibility were correlated using molecular modeling and simulations. As predicted, O-glycan-free CTLA4 Fc-fusion proteins were more homogenous for sialylation, and interestingly less prone to protein aggregation. Attenuating protein aggregation was a desirable effect, and could be related to the reduced presence of linker O-glycans that hindered inter-chain disulfide bond reformation. Findings from this study shed light on new therapeutic protein design and development.

1. Introduction

Fusion proteins are proteins with multiple functional domains connected via protein linkers, and these proteins may functionally retain multiple distinct activities derived from each individual domain. Genetically engineered recombinant fusion proteins are widely applied in basic biological research, such as protein purification [1], protein crystallization [2], drug delivery [3], and imaging [4]. These proteins have also emerged as an important category of biopharmaceuticals to treat a number of diseases [5]. However, the importance of linkers in addition to serving as connectors in fusion proteins has been overlooked compared to the novel functions these therapeutic fusion proteins offer. Improper linker structure in the fusion protein may lead to undesirable outcomes, such as low yield, protein aggregation, and reduced potency [6,7]. Elegant reviews on protein linker design and selection have been published [8,9]. However, studies have not focused on O-glycosylation in fusion protein linkers.

Protein glycosylation, including N- and O-glycosylation, is involved in numerous biological processes, such as cellular targeting, secretion, protein folding, enhancing stability and solubility, cell signaling, and hematopoiesis. N-glycosylation begins with the addition of a preformed oligosaccharide unit (consisting of 14 monosaccharide residues) followed by the sequential addition or trimming of certain sugar residues one at a time in a defined process regulated by different enzymes [10–13]. In comparison, O-glycosylation involves attachment of a single monosaccharide residue to the hydroxyl group of serine or threonine and further sequential addition of many diverse monosaccharide residues regulated by different glycosyltransferase enzymes [14,15]. Based on the monosaccharide residue linked to serine or threonine, O-glycosylation can be categorized into mucin and nonmucin types [16]. In the mucin type, the attached monosaccharide residue is N-acetylgalactosamine (GalNAc), whereas the attached residue is N-acetylglucosamine (GlcNAc), fucose, mannose, glucose, xylose, or galactose in the nonmucin type [17,18]. Compared to the well-studied N-glycosylation, O-glycosylation is difficult to study for at least three reasons: (1) lack of a consensus sequence for the glycosylation sites on the polypeptide, (2) high heterogeneity both in

* Corresponding author at: Bristol Myers Squibb Company, 38 Jackson Road, Devens, MA 01434, USA.
E-mail address: Yueming.Qian@bms.com (Y. Qian).

https://doi.org/10.1016/j.csbj.2020.11.037

the number of glycans and in the extent of their occupancy, and (3) lack of a universal enzyme to release O-glycans from proteins.

Some of our therapeutic fusion proteins comprise a linker derived from the hinge region of IgG1. Abatacept (ORENCIA®) and belatacept (NULOJIX®) are typical examples of such fusion proteins, which consist of a modified IgG1-Fc and a native or high affinity variant of the extracellular domain of cytotoxic T lymphocyte antigen-4 (CTLA4). CTLA4 is expressed as an extracellular receptor on the surface of activated T cells, and transmits a competitive inhibitory signal [19]. As immunosuppressant drugs, CTLA4-Fc fusion proteins are clinically used to treat autoimmune disorders such as rheumatoid arthritis, or protect a transplanted organ from rejection. Interestingly, the linker between CTLA4 and IgG1-Fc is highly O-glycosylated as mucin type in the fusion proteins but not in IgG1 antibodies [20,21]. The change in the engineered linker is the substitution of two cysteine residues to serine residues, which converts CPPC motif to SPPS motif and eliminates the disulfide bond pairs between two chains. The unintended introduction of O-glycosylation may increase glycan heterogeneity in the recombinant proteins. Prompted by the regulatory guidance for product quality control and consistency, we performed a mechanistic study on O-glycan initialization and investigated the impacts of O-glycosylation in fusion protein linkers on protein quality attributes, such as total sialic acid and protein aggregation levels. In this study, the following questions are addressed. Why is the linker O-glycosylated in the fusion protein but not in the hinge region of IgG1 antibody although the linker is derived from the hinge region of IgG1? What determines O-glycosylation efficiency at multiple O-glycosylation sites in the fusion protein linker? What is the relationship between O-glycosylation and protein quality attributes? Answers to these questions advance our fundamental understanding of biological processing for potential O-linked glycosylation in Fc-fusion proteins and shed light on strategies to improve protein attributes through protein engineering and process development.

2. Materials and methods

2.1. Cell lines and culture condition

HEK293-F (Invitrogen) suspension cells were grown in serum-free FreeStyle 293 expression medium (Invitrogen) and passaged in shake flasks every three days. Cells were incubated at 130 rpm on an orbital shaker platform in a 37 °C incubator with 6% CO2 in air.

2.2. Construction of fusion protein mutants

Site-directed mutagenesis was performed as previously described [22]. In brief, the mutagenesis reaction was performed using methylated CTLA4 Fc-fusion protein plasmid DNA as the template and with two overlapping primers, one of which contained the target mutation that led to the encoded protein with one amino acid deletion or substitution with alanine or glutamine. The mutagenesis products were transformed into E. coli competent cells, where unmethylated linear mutated DNA was circularized and replicated. Plasmid DNA was purified with a Maxi-prep plasmid kit (Qiagen, Mississauga, ON, Canada) and sequenced to confirm the correct mutation was present (Cogenics, Houston, TX).

2.3. Transient expression of CTLA4 Fc fusion protein and its variants in HEK-293F suspension cells

One day prior to transfection, cells were seeded at 0.6 million cells/ml and agitated on an orbital shaker platform rotating at 135 rpm at 37 °C with 6% CO2. On the day of transfection, cells were diluted to 1.0 million cells/ml and added into a 125-ml shake flask at 30 ml culture volume. Forty micrograms of plasmid DNA were diluted into OptiPro SFM (Invitrogen) to a total volume of 0.6 ml and mixed. In a separate tube, 40 μL of FreeStyle MAX Reagent (Invitrogen) was also diluted with Opti-Pro SFM to a total volume of 0.6 ml. Diluted DNA solution and diluted transfection reagent were mixed gently and incubated at room temperature for 20 min. DNA-transfection-reagent mixture was then slowly added to freshly diluted cells in 125-ml flasks. Transfected cells were incubated at 37 °C and 6% CO_2 on an orbital shaker platform rotating at 135 rpm for 7 d in batch mode.

2.4. Determination of CTLA4 Fc-fusion mutant concentration (titer) and protein purification

Cell culture samples were centrifuged at 1000 rpm for ten minutes and the supernatants were stored at −20 °C prior to the product titer assay. Fusion protein concentrations were measured with Protein-A biosensors on an Octet QK (ForteBio, Menlo Park, CA), according to the manufacturer's instructions. The harvest samples were purified with Protein-A spin columns (Sartorius), yielding a final concentration of 0.2–0.3 g/L.

2.5. Determination of binding kinetics

Binding kinetics of each mutant and wild-type to CD80-Ig was determined by immobilizing the CD80-Ig on the BIAcore sensor chip surface at densities of 200 RUs. Purified CTLA4 Fc fusion proteins in the concentration range of 10–200 nM in HBS-EP buffer (10 mM Hepes, 150 mM NaCl, 3 mM EDTA, 0.005% Tween-20) (pH 7.4) were injected over the sensor chip surface at a flow rate of 30 μL/min. Association and dissociation data were collected for 3 and 5 min, respectively, at different concentrations.

2.6. Determination of O-linked glycans

Protein-A purified samples were diluted in 100 mM Tris, 25 mM NaCl, pH 7.6 and incubated with PNGase F (1:50 enzyme:substrate) at 37℃ overnight to remove N-linked oligosaccharides. Samples were then spiked with an internal standard of insulin, loaded onto a Waters Oasis HLB cartridge, and washed with 0.1 formic acid in 5% acetonitrile followed by gradient elution with 0.1 formic acid in acetonitrile. The method for intact mass analysis was the same as described previously [23]. Briefly, processed sample was injected at 0.5 μg onto a reversed phase column (10 μm, 2.1 × 10 0 mm Poros®, Applied Biosystems, Foster City, CA) equilibrated with 20% acetonitrile containing 0.1% formic acid, followed by gradient elution from 20% to 50% acetonitrile in 25 min, at a flow rate of 0.25 ml/min. The capillary and cone voltages were 3 kV and 30 V, respectively, and scans were performed from 800 to 2500 m/z. Spectra were deconvoluted using the MaxEnt1 algorithm to determine molecular masses and O-linked glycan species. The mass difference between the detected mass and the predicted mass from protein sequence (78837 Da for wild type protein) was used to determine O-glycan levels. Theoretical masses of O-glycan units GalNAc, Gal, and Sialic Acid are 203 Da, 162 Da, and 291 Da, respectively. The most common O-glycan forms of GalNAc-Gal-SIA and GalNAc-(SIA)-Gal-SIA have expected masses of 656 and 947 Da, respectively. The differences between the measured mass and expected mass are 3 Da or less.

2.7. Determination of total sialylation

Total sialic acid content was determined by the method described previously [24]. Sialic acids (N-acetylneuraminic acid

and N-glycolylneuraminic acid) from N- and O-linked glycosylation were released by mild acid hydrolysis and then separated by reversed-phase HPLC. The sialic acid was monitored with its absorbance at 206 nm and quantitated via peak area and linear regression of a standard curve. The results were reported as a molar ratio of N-acetylneuraminic acid to protein. N-glycolylneuraminic acid was below the limit of detection in all cases (<0.1 mol per mol glycoprotein).

2.8. Size exclusion chromatography (SEC)

SEC was used to determine protein aggregation by measuring the percentage of high molecular weight (HMW) species in wild-type and mutant samples as described previously [25]. Protein-A purified samples and reference standard were directly injected into an HPLC equipped with a 7.8 × 300 mm Toso Haas Biosep TSK G3000SWXL column. The separation was isocratic using PBS at pH 7.2 at room temperature at a flow rate of 0.5 ml/minute for 30 min. The isocratic elution profile was monitored at 280 nm. Quantitation was performed by calculating relative peak area percentages of HMW, monomer and low molecular weight (LMW) species between the void and the inclusion volumes.

2.9. Molecular modeling and molecular dynamics simulation

2.9.1. Homology model of CTLA4 Fc-fusion protein and flexibility of the linker region of the fusion protein

To understand why linker region can be glycosylated in the Fc-fusion protein but not in IgG1 mAbs, homology models of CTLA4 Fc-fusion protein with SPPS motif and CPPC motif were generated Molecular dynamics (MD) simulation of homology models was conducted to evaluate flexibility of the linker regions of two homology models. Briefly, 3D structures of the fusion protein and its CPPC motif containing analog were generated using a comparative homology modeling method in Modeller [26] with CTLA4 crystal structure 3OSK ([27] and human IgG1 Fc crystal structure 5JII [28] as the 3D structure templates for the CTLA4 domain and Fc domain of the fusion protein, respectively. The Fc-fusion protein and its CPPC motif containing analogy were used for molecular dynamics (MD) simulation. Both proteins were sugar-free to simulate linker flexibility with and without inter-chain disulfide bonds prior to glycosylation. Fc fusion proteins were solvated using extended simple point charge (SPCE) water [29] in GROMACS (V5.0) in the OPLS all-atom force field [30,31]. The solvated and electro-neutralized system was energy minimized to release steric clashes and inappropriate geometry. To equilibrate solvents around the protein, a solvent equilibrium step including a 200 ps NVT (constant Number of particles, Volume, and Temperature) step and a 200 ps NPT (constant Number of particles, Pressure, and Temperature) step was conducted with a position restraining force applying on all non-hydrogen protein atoms. The all-atom MD simulations were performed for 10 ns with a time step of 2 fs, temperature coupling on and a reference temperature of 300 K, pressure coupling on and a reference pressure of 1 atm. The flexibility of linker in each Fc-fusion was assessed using the root mean square fluctuation (RMSF) of alpha carbons calculated using the MD simulation trajectory.

2.9.2. Homology model of mouse polypeptide N-acetylgalactosaminyltransferase and its complex with the linker region of the fusion protein

The homology model of the catalytic domain (resi85-resi485) of mouse polypeptide N-acetylgalactosaminyltransferase (GALNT) (UniPort Sequence ID: Q921L8) was generated using the structure of the catalytic domain (resi71-resi441) of human UDP-GalNAc: polypeptide alpha-N-acetylgalactosaminyltransferase (PDB Code: 2FFU) [32] using the same comparative homology modeling method. Sequence alignment is reported in the supplemental material. Peptides from the linker region of the fusion protein covering residues from 125 to 143 were docked to the active site of the mouse GALNT using residues S129, S130, T133, T135, S136, and S139 as the glycan acceptor in Molecular Operating Environment (MOE) from Chemical Computing Group. The top hits of each complex were selected according to the E_conf score to evaluate the conformational preference of six different receptors at S129, S130, T133, T135, S136, and S139. The representative models and size of different sugar units were displayed measured using PyMol V1.7 (Schrödinger, LLC.).

3. Results

Previous studies suggested that our CTLA4 Fc fusion protein was highly glycosylated with three N-glycosylation sites (two in the CTLA4 domain and one in Fc domain) and two or more O-glycosylation sites [20,21] in the linker region for each single chain as shown in Fig. 1 (A, B). To understand the mechanism of O-glycosylation in the linker and the impact of O-glycan in the linker region on product quality, mutations were introduced to the all six putative O-glycosylation sites (all S/T sites), which were predicted by the neural network method NetOGlyc3.1 at www.cbs.dtu.dk/services/netoglyc, in the linker region as shown in Fig. 1 (C).

3.1. Protein engineering, expression, and purification

Site-directed mutagenesis was performed via either complete deletion or substitution with alanine or glutamine. Typically, residues that protrude out of the surface of the molecule are substituted with alanine and buried residues are substituted with glutamine to avoid structural collapse. With respect to creating a functional protein, the only difference between an alanine and a serine is the substitution of hydrogen with a hydroxyl, which has minimal interference on protein structure. In this model protein, even a complete deletion of all six amino acids still resulted in secretion of a protein at comparable expression level of the wild type protein. DNA sequences of all mutants were confirmed prior to protein expression.

Both the wild type Fc-Fusion protein and its mutants were robustly expressed in Human embryonic kidney (HEK) transient expression system, and the protein titers of all mutants ranged from 0.10 to 0.15 g/L. In particular, the successful synthesis and secretion of mutants in which all six putative O-glycosylation sites were eliminated indicates that O-linked glycosylation is not essential for CTLA4 Fc-fusion protein synthesis and secretion. After harvest by centrifugation, proteins were purified using Protein-A spin columns (Sartorius) to a final concentration of 0.2–0.3 g/L. Given that the mutation sites are located on the opposite side of the target binding site, we expect that the mutation would not affect protein potency, which was confirmed by potency assay using Biacore. The dissociation constant Kd of analyzed mutants is within a range from 2.0 to 6.6 nM, which is comparable to the wild type Kd of 4.2 nM and previously published results [33].

3.2. O-Glycosylation of mutants at the linker region

The O-glycosylation amino acid consensus sequence is unknown, although neural network approaches have been developed to predict mucin-type O-linked sites [34,35]. At the time of designing this work, the only published work [20] on a similar protein indicated that two O-glycans were present in the linker region as shown in Fig. 1 (B). We mutated the linker region to address three questions in this study: (1) Are other O-glycosylation sites

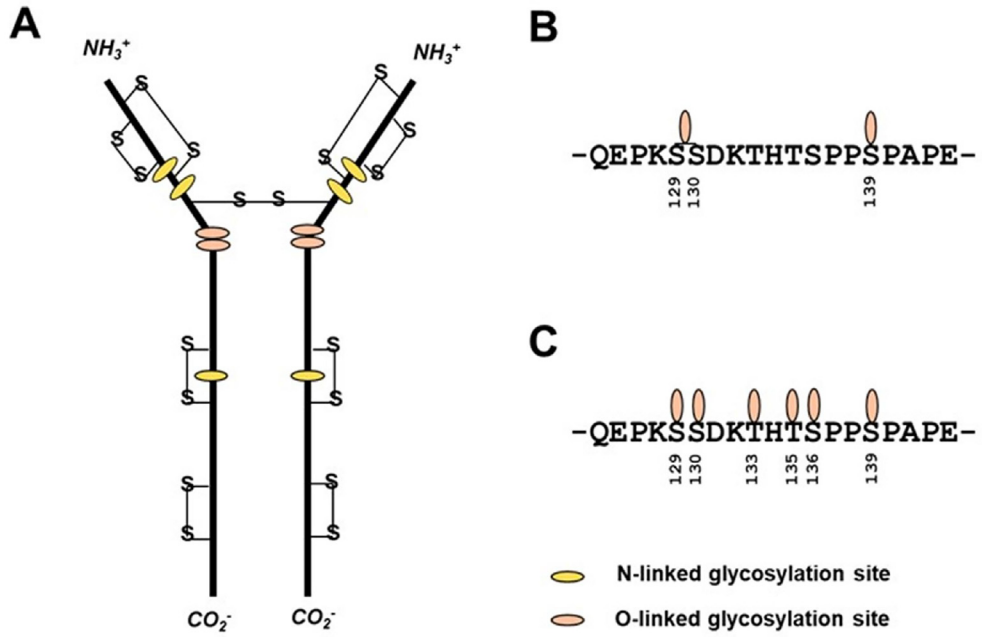

Fig. 1. Glycosylation sites in the CTLA4 Fc-Fusion protein. (A) Identified N-linked glycosylation and O-linked glycosylation sites deduced from the published data [20,21,46]. (B) O-linked glycosylation sites mapped by mass spectrometric analysis on tryptic peptide fragments [20]. One O-linked site could not be definitively assigned to S129 or S130. (C) The O-linked glycosylation sites predicted by the neural network method NetOGlyc3.1 [34,35]. The Belatacept amino acid sequence was used as input into the NetOGlyc3.1 software online at www.cbs.dtu.dk/services/netoglyc. All six putative O-linked glycosylation sites were mutated in this work.

present in addition to the two reported sites? (2) How does the local sequence affect the O-glycan pattern? (3) How can we generate an O-glycan-free construct?

3.2.1. Additional O-glycosylation site(s)

We deleted all three previously identified O-glycosylation sites (S129/S130 and S139) to generate the Δ129Δ130Δ139 construct. Surprisingly, our data showed that 83.1% of monomers possess two SIA-Gal-(SIA)-GalNAc (SIA in the parenthesis is branched sialic acid) glycans, and 16.9% of monomers possess three SIA-Gal-(SIA)-GalNAc (as shown in Fig. 2 C) glycans, which suggests that T133, T135, and S136 are also possible O-glycosylation sites. A recent publication [21] reported O-glycosylation at these sites, which is consistent with our data.

3.2.2. Local sequence impact on O-glycan pattern

A group of mutants were created at S129 with S129 being deleted or substituted with Ala or Gln to generate Δ129, S129A, and S129Q, respectively. Similarly, Δ139, S139A, and S139Q were generated. The O-glycosylation patterns of these mutants are shown in Table 1. Our data revealed that mutagenesis at the 129 or 139 sites affected O-glycosylation patterns differently. Deleting S129 or S139 led to more complicated O-glycosylation, e.g., protein with two SIA-Gal-(SIA)-GalNAc glycans and protein with three SIA-Gal-(SIA)-GalNAc glycans. Substitution with Ala or Gln at 129 or 139 simplified and mitigated O-glycosylation. Ala substitution at site 139 had a more significant impact on simplifying and mitigating O-glycosylation. When both 129 and 139 were deleted simultaneously (Δ129Δ139), more complicated O-glycans were produced.

3.2.3. O-glycan-free construct

The construct with all possible O-glycan sites mutated (Δ129Δ130A133A135A136Δ139) was expected to have no O-glycans since all possible O-glycosylation sites are either deleted or replaced with Ala. S129 and S130 were detected as O-glycosylation sites in the wild type construct in current study and others [20,21]. Interestingly, the construct possessing only S129 or S130 with other possible O-glycosylation sites replaced with Ala exhibits no O-glycan (Table 1s to last row and third to last row), which implies that O-glycosylation at S129 and S130 may not stand alone and is not a prerequisite for further O-glycosylation. Our data strongly suggest that the local sequence has a significant impact on O-glycosylation; however, no consensus sequence of O-glycosylation has been identified. Inconsistent sialylation is a major problem for manufacturing consistency. If eliminating O-glycosylation from the linker region does not negatively affect product quality, it is desirable to have the simplest O-glycosylation profile from a manufacturing perspective. Eliminating redundant glycosylation sites will lead to homogenous glycoproteins.

3.3. Impacts of mutagenesis on protein critical quality

The level and pattern of O-glycan in a protein may affect chemical, physical, and biological properties. In the development of biologics, a critical quality attribute is defined as a physical, chemical, biological, or microbiological property or characteristics that should be within an appropriate limit, range, or distribution to ensure the desired product quality (FDA Q8(R2)). The presence of sialic acid in the terminus of glycans is a typical critical quality attribute, which affects absorption and the serum half-life of therapeutic proteins.

3.3.1. Sialic acid levels in mutant proteins

The total sialic acid (TSA) level is the sum of N-glycan sialic acid residues in the CTLA4 domain and O-glycan sialic acid residues in the linker region. A trend of decreased total sialic acid is not observed with an increasing number of mutations until all O-glycosylation sites are eliminated as shown in Fig. 3 (A). Interestingly, the TSA level is constant in these mutants with no O-glycosylation (as presented by circled data). Our data suggest that eliminating O-glycosylation sites will improve product homogeneity and process consistency.

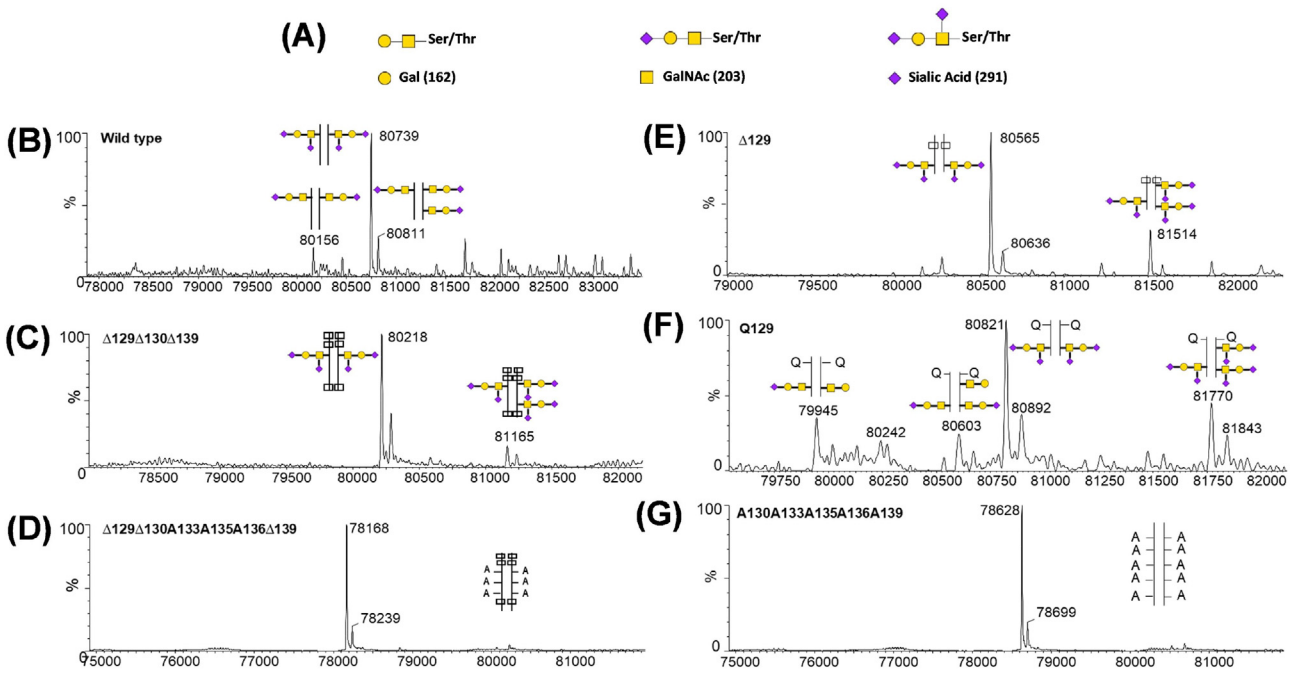

Fig. 2. Representative mass spectrometry analysis of O-glycan of Fc fusion protein and its mutants. (A) Diagram illustrating common O-glycan structures. The number in parenthesis indicates the molecular weight. Gal - galactose, GalNAc - N-acetylgalactosamine. (B) Deconvoluted mass spectra showing O-linked glycosylation in wild-type Fc-fusion protein. (C) Deconvoluted mass spectra showing O-linked glycosylation in the mutant that did not have S129, S130 and S139. (D) Deconvoluted mass spectra showing no O-linked glycosylation in the mutant that did not have S129, S130 and S139 and in which T133, T135 and S136 were substituted with alanine. (E) Deconvoluted mass spectra showing O-linked glycosylation in the mutant with S129 deletion. (F) Deconvoluted mass spectra showing O-linked glycosylation in the mutant with glutamine substitution at position 129. (G) Deconvoluted mass spectra showing no O-linked glycosylation in the mutant with S129 intact and alanine substitution at position 130, 133, 135, 136 and 139.

Table 1
Mutants generated in this study and their O-glycans determined using mass spectrometry.

Mutants	Possible O-sites	Non-O	1 GalNAc-Gal-SIA	2 GalNAc-Gal-SIA	3 GalNAc-Gal-SIA	1 GalNAc-(SIA)-Gal-SIA	2 GalNAc-(SIA)-Gal-SIA	3 GalNAc-(SIA)-Gal-SIA
WT	S129S130T133T135S136S139			75.1%	24.9%			
Δ129	S130T133T135S136S139						75.5%	24.5%
Q129	S130T133T135S136S139			19.6%			51.8%	28.6%
A129	S130T133T135S136S139		29.6%	40.8%		29.6%		
Δ139	S129S130T133T135S136						72.7%	27.3%
Q139	S129S130T133T135S136	20.7%				12.5%	66.8%	
A139	S129S130T133T135S136	43.5%				28.7%	27.9%	
Δ129Δ139	S130T133T135S136						80.4%	19.6%
Q129A139	S130T133T135S136	33.2%				18.9%	36.1%	11.8%
Δ129Δ130Δ139	T133T135S136						83.1%	16.9%
Δ129Δ130A136Δ139	T133T135						83.1%	16.9%
Δ129Δ130A133A136Δ139	T135	60.4%				39.6%		
A130A133A135A136A139	S129	100.0%						
A129A133A135A136A139	S130	100.0%						
Δ129Δ130A133A135A136Δ139	NONE	100.0%						

3.3.2. HMW levels of mutant proteins

Product HMW is also a critical quality attribute due to its potential involvement in immunogenicity. The trend of relative HMW is clearly decreased as the number of mutated O-glycosylation sites increases as shown in Fig. 3 (B). HMW levels are consistently 40% lower for most mutants relative to wild type protein. Reducing HMW level at the cell culture stage is desirable since it can reduce the burden for HMW removal in the subsequent purification steps. An explanation on why the HMW level is negatively correlated to the O-glycosylation level in this study is proposed in the discussion section.

3.4. Linker region flexibility

The linker of this fusion protein originated from the hinge region of the IgG1 monoclonal antibody. The only change was the substitution of two cysteine residues to serine residues. These two cysteine residues in the hinge region, known as CPPC motif, covalently bind to the other two cysteine residues from the CPPC motif in the other heavy chain in a monoclonal antibody. The substitution of cysteine residues using serine eliminates these disulfide bond pairs. O-glycosylation has never been reported in the hinge region of any mAbs. However, multiple O-glycosylation sites

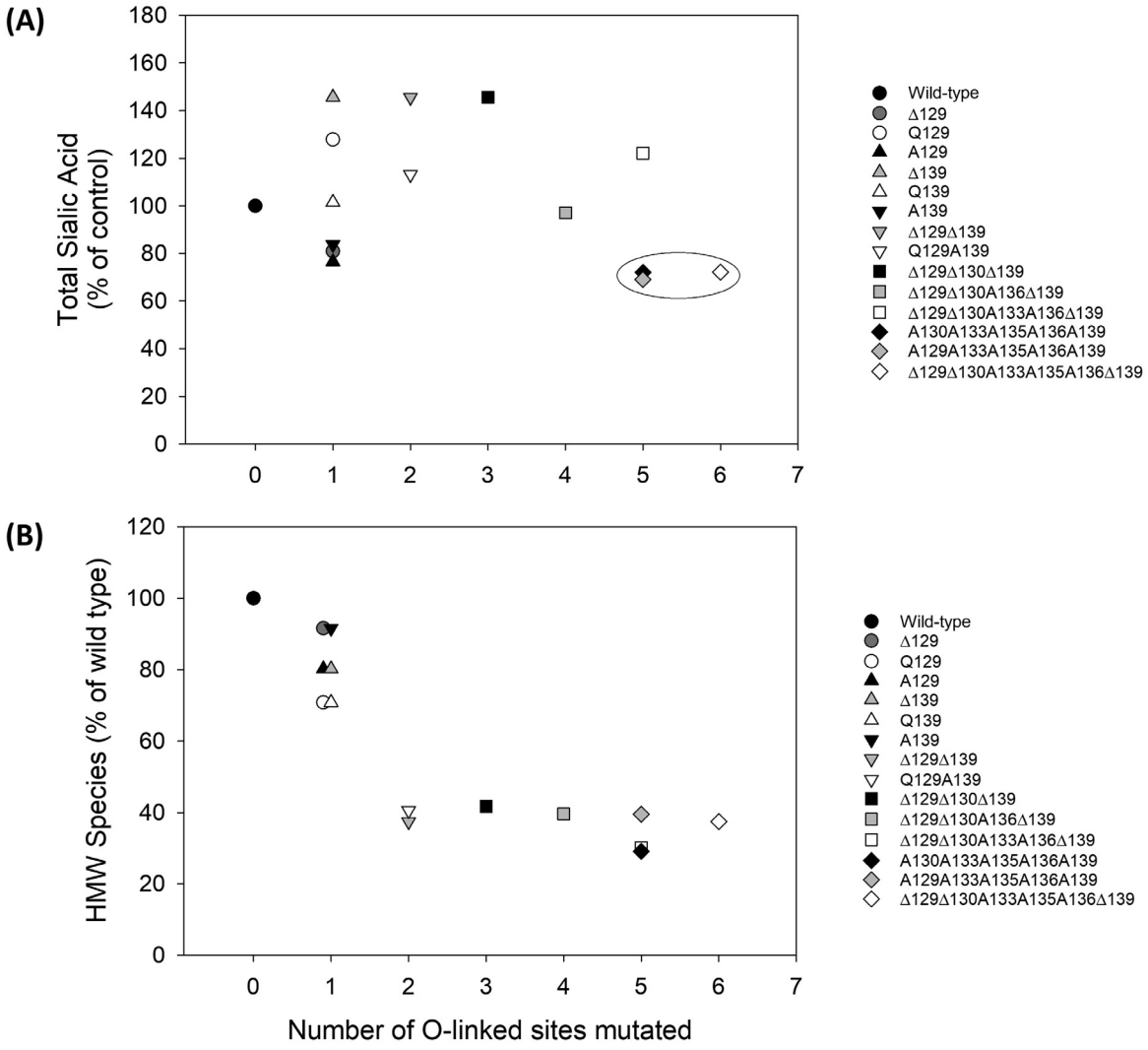

Fig. 3. Impact of mutagenesis at O-glycosylation sites in the linker region of CTLA4 Fc fusion protein on quality attributes. (A) Total sialic acid (TSA) of mutants relative to that of wild type. TSA data for mutants without O-glycosylation are circled. (B) Percentage of High Molecular Weight (HMW) species of mutants compared to wild type. HMW species is a parameter indicating protein aggregation.

were identified after the substitution in the linker region in this fusion protein. We hypothesize that the substitution leads to the high flexibility of the linker region, and the high flexibility results in high accessibility of the linker region to glycan transferase to initialize O-glycosylation. The root mean square fluctuation of alpha carbon (Cα RMSF) of the linker region was calculated to evaluate the impact of the mutation introduced to the linker (Fig. 4). The Cα RMSF of the linker region containing SPPS motif is greater than that of the linker region containing the CPPC motif. We speculate that the high flexibility may result in high accessibility to glycan transferase to initialize O-glycosylation.

3.5. Model of the linker peptide and polypeptide N-acetylgalactosaminyltransferase (GALNT) complex

To understand why eliminating previously identified O-glycosylation sites cannot lead to O-glycan-free constructs, we generated a model of the complex of the linker with the enzyme that initializes the mucin-type O-glycosylation. The catalytic domain of mouse GALNT has high sequence identity (51.2%) to that of human GALNT (PDB code: 2FFU) with key residues involved in catalysis and substrate binding highly conserved as shown in the supplemental material (Fig. S1). The homology model of CHO polypeptide N-acetylgalactosaminyltransferase was generated using the structure of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase [32] as shown in Fig. 5 (A).

To understand the site preference of the initial O-glycan in the substrate, six complexes were generated to compare the geometric complement and steric hindrance of the complexes. S/T residues of the complexes were close to the enzyme center. The conformational energy of the complexes was calculated. Interactions between the glycan acceptor residue (S/T) and the enzyme active center as well as between neighbor residues and the enzyme active site contribute to the energy score. To demonstrate this interaction, three representative complex models with the substrate (linker peptide) at different positions are shown in Fig. 5 (B-D). Steric hindrance from neighboring residues plays critical role in complex formation. S139 is flanked with Pro and small residue such as Ala. This allows the substrate to adopt an elongated and narrow conformation (as shown in Fig. 5 (E-F)) to fit the narrow groove of the enzyme active site. The alpha carbon trace of the substrate in Fig. 5 (E-F) shows the conformation of the polyproline-II helix

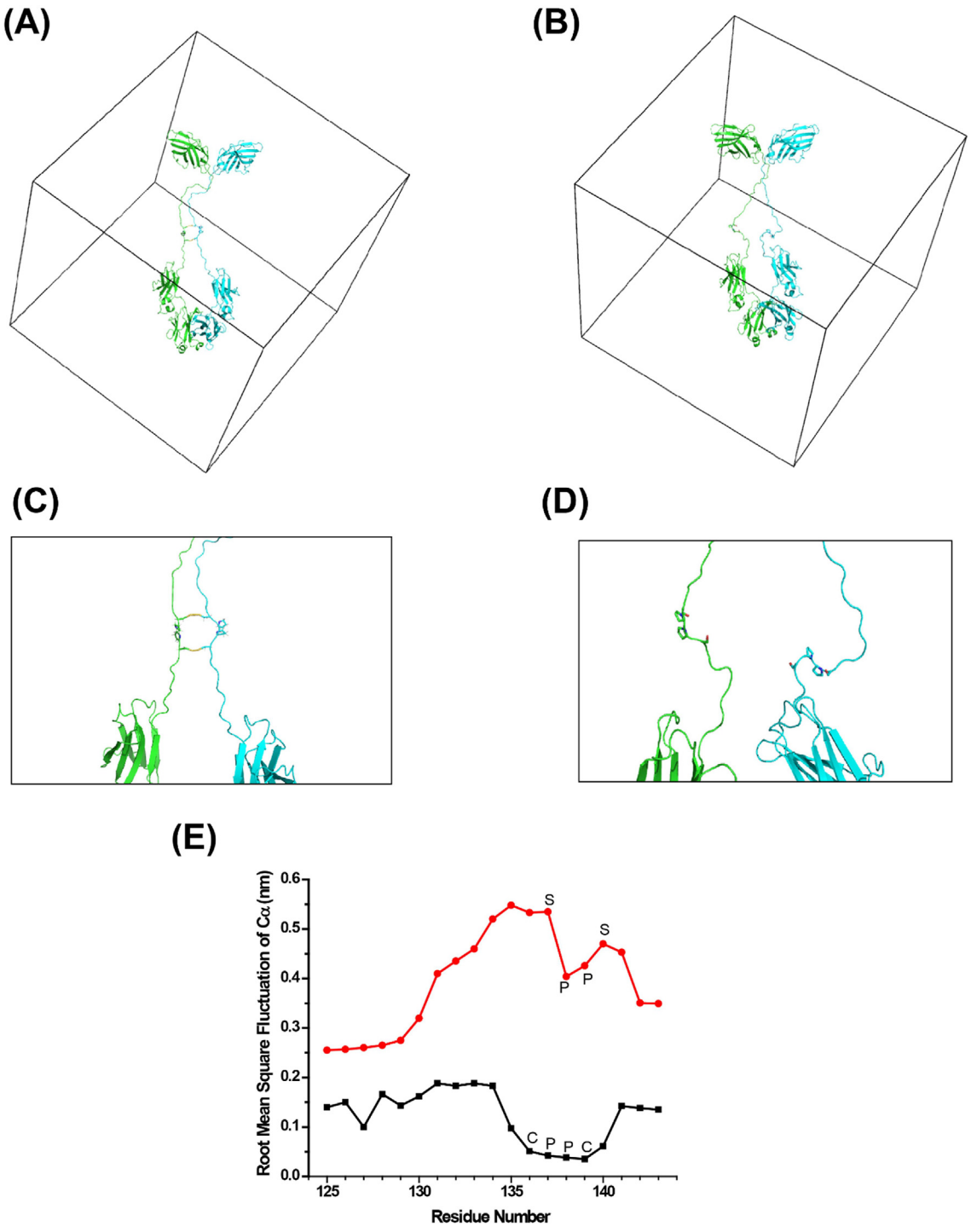

Fig. 4. The flexibility of CTLA4 Fc-fusion protein linker was calculated from trajectory of molecular dynamics (MD) simulation of structure models with and without cysteine substitutions. (A) A CTLA4 Fc-fusion protein with a CPPC motif (wild type in IgG1 antibody) in the linker region in an MD simulation box. (B) A CTLA4 Fc-fusion protein with a SPPS motif (serine replacement of cysteine) in the linker region in an MD simulation box. (C) Zoomed-in region of the linker in (A) panel. (D) Zoomed-in region of the linker in (B) panel. (E) RMSF of alpha carbon of linker regions in proteins containing the CPPC motif (black) and SPPS motif (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[36]. Large residues, such as Lys next to T133 and S129, are not energetically favorable for the formation of the polyproline-II helix and do not fit such a narrow groove.

3.6. Impact of O-Glycan size on linker conformation

The size of O-glycans with different number of sugar unit in the linker region was measured. One two-sugar unit Gal-GalNAc is approximately 11.5 Å long, one linear three-sugar unit SIA-Gal-GalNAc is 18.7 Å long, and one branched four-sugar unit SIA-Gal-(SIA)-GalNAc is 7.0 Å long and 18.2 Å wide as shown in Fig. 6 (A-C). The distance from the end of the CTLA4 domain to the start of the Fc domain is approximately 80 Å in an extended conformation of the linker region as shown in Fig. 6 (D). Multiple glycans in the linker region may cause steric hindrance for the formation of fusion protein homodimers. During the protein production stage,

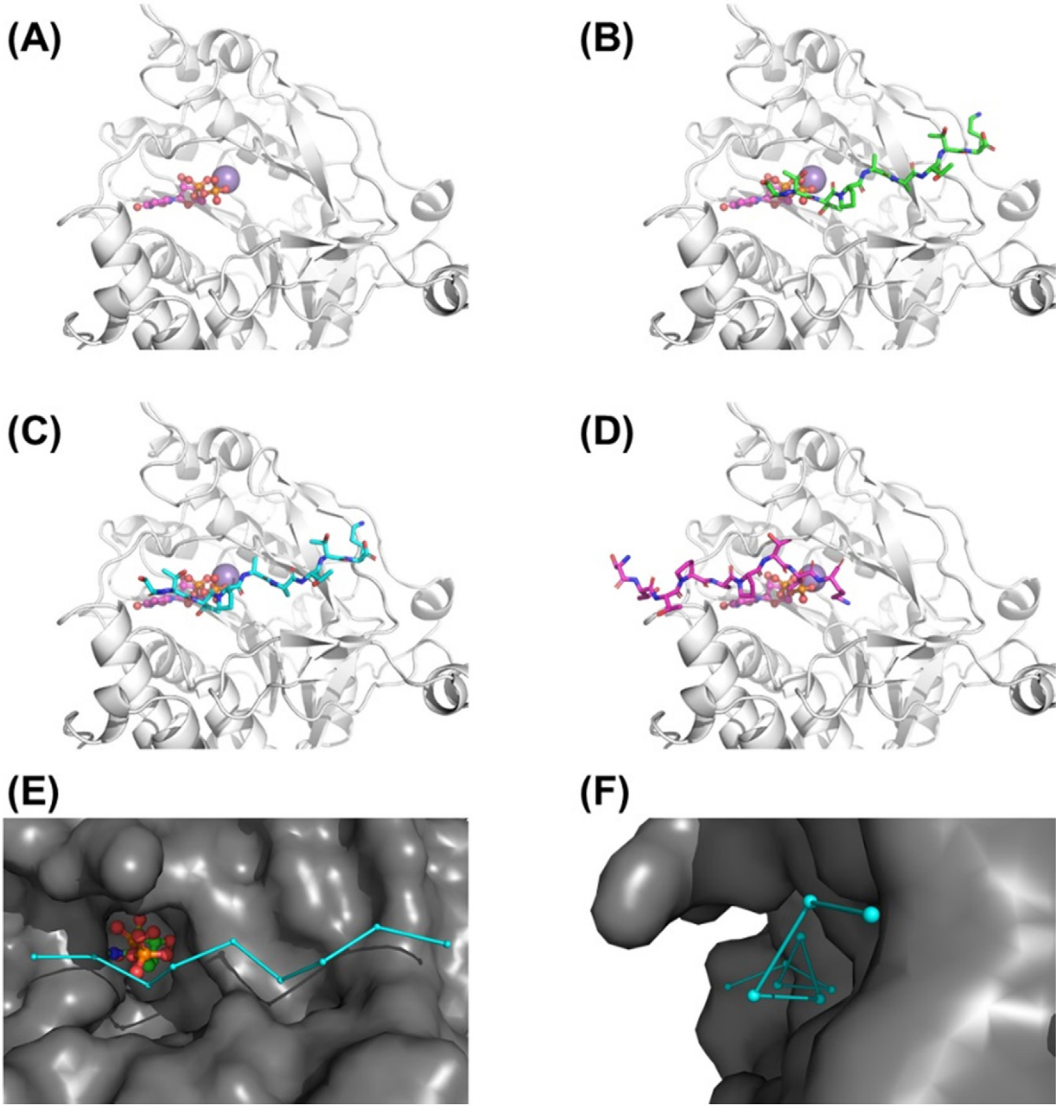

Fig. 5. Docking of linker peptide substrate into UDP containing the CHO GALNT2 homology model. (A) UDP containing the CHO GALNT2 homology model. In the enzyme active center, UDP-GalNac is shown in pink, and Mn^{2+} is noted in the purple sphere. (B) Position 1 of the linker peptide in GALNT2 with S129 at the reaction center. (C) Position 2 of the linker peptide in GALNT2 with S130 at the reaction center. (D) Position 3 of the linker peptide in GALNT2 with S139 at reaction center. (E) Side view of alpha carbon trace of the substrate in the narrow groove of GALNT2. (F) Axis view of alpha carbon trace of the substrate in the narrow groove of GALNT2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the interchain disulfide at the end of the CTLA4 domain may be reduced to free thiols. The large steric hindrance may prevent interchain disulfide bond reformation. However, the CH3 domain in the Fc still maintains its dimerization status. The free thiol from one monomer may cross-link to the free thiol from another monomer to form high molecular weight species as demonstrated in Fig. 6 (D). The existence of more species in a nondisulfide bond state may lead to an increased tendency to form aggregates.

4. Discussion

4.1. Impacts on other properties

Mutagenesis may impact both biological and physicochemical properties. Solubility is critical for the development of high concentrations of biologics. Although an exclusive test for solubility is not available, we believe that the risk of low solubility of these mutants is low for two reasons. First, these mutants have similar yields compared to wild type. Second, these mutants did not exhibit elevated HMW levels. The evidence suggests that the negative impact of these mutants on solubility is negligible. O-glycosylation has been reported to affect the half-life of recombinant proteins in the body and immunogenicity, which are not covered in this study. Further work is needed to answer these intriguing questions.

4.2. Why does O-glycosylation occur at linker regions in proteins after CPPC is changed to SPPS?

Our MD simulation data suggest that CPPC-containing linker region exhibits less flexibility. The reduced flexibility will limit the accessibility of the Ser/Thr residue in the linker region. The two pairs of disulfide bonds in the linker lock two chains together, preventing the linker from docking to peptide GalNAc-transferases. On the other hand, the linker in the construct containing SPPS is flexible and less paired, which facilitates docking the linker to peptide GalNAc-transferases initializing O-glycosylation in the linker.

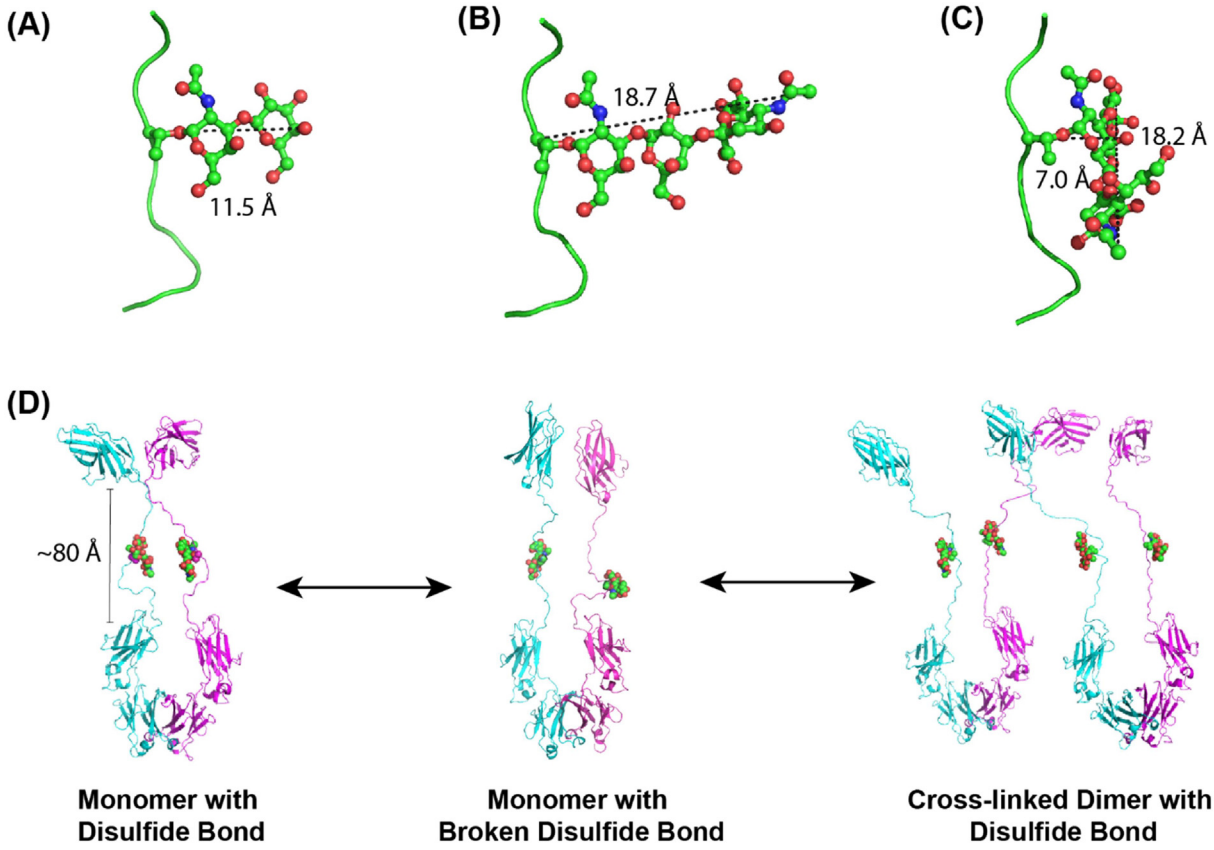

Fig. 6. (A-C) O-glycan size in the linker region. (D) Hypothetical mechanism of O-glycan in CTLA4 Fc fusion protein linker facilitating the formation of High Molecular Weight (HMW) species. HMW species is a parameter indicating protein aggregation.

4.3. Which site exhibits efficient O-glycosylation? (O-glycosylation site preference)

Mucin-type O-glycosylation occurs in the region rich in Ser, Thr, and Pro [37]. Mucin-type O-glycosylation biosynthesis involves more than twenty glycosyltransferases. The reactions are confined in the cis-Golgi and the trans-Golgi [38]. The initial step is the addition of GalNac to the peptide, which is catalyzed by polypeptide N-acetylgalactosaminyltransferases (GALNTs) [39]. The first step of O-GalNAc glycosylation is clearly regulated by the amino acid sequence of the protein substrate. However, the sequence of amino acids with a Ser/Thr that will receive a GalNAc has not been identified. The initial reaction occurs between UDP-GalNac and the peptide catalyzed by GalNAc-transferases. The GalNAc-transferase structure reveals that the donor UDP-GalNac (activated form of GalNac) is located in the center of the catalytic domain. In addition, the acceptor peptide fits in the groove of the catalytic domain with the hydroxyl group of the acceptor residue, which is either serine or threonine, facing the donor, and a metal ion, such as Mn^{2+}, joins the donor and the acceptor. The Pro residue plays a role in peptide substrate recognition by GalNAc-transferases [40]. To understand the order of O-glycosylation initialization, we conducted molecular docking experiments to determine how the linker region of the CTLA4 Fc-fusion protein fits the groove of the complex of UDP-GalNac and GalNAc-transferases. The order of O-glycosylation from our analysis on the geometric complement and steric hindrance primarily in relation to the neighboring Pro residue is S139, S136, T135, T133, S129, and S130. S139 and S129/S130 were previously reported as the O-glycosylation sites [20]. This finding seems to conflict with the O-glycosylation preference order from our modeling study. Our O-glycosylation preference order only considers the first O-glycosylation initialization in the region assuming no second O-glycosylation. O-glycosylation detected in real product involves initialization of the O-glycosylation at a second site.

Other protein domain in GLANT, namely, the lectin domain, dictates the initialization of O-glycosylation at a second site [41]. After the first GalNac is added to the substrate peptide, the lectin domain binds to the GalNac-glycopeptide, which directs and facilitates the catalytic domain to catalyze a second initial GalNAc to the acceptor glycopeptide [42]. The residue with the GalNAc bound to lectin domain is approximately 8–10 residues apart from the residue to be catalyzed in the catalytic domain. This notion explains why our previous study showed O-glycosylation at 129/130 and 139, which are 9–10 residues apart from each other. Some O-glycosylation sites are highly efficient, whereas others are less efficient. When these efficient O-glycosylation sites are not available for glycosylation or mutated, less efficient O-glycosylation sites become available and are glycosylated. O-glycosylation at different neighboring sites is orchestrated by multiple GALNTs in the secretory pathway. Early studies showed that O-glycosylation only occurs at 129/130 and 139. Pro residues in the vicinity of an O-glycosylation site play a role in creating a suitable acceptor by exposing hydroxyls to be glycosylated or by binding to the enzyme in a Pro-binding motif. Thus, we argue that 139 is the most favorable site since S139 is followed by a Pro residue. Our data show that mutants that exclusively harbor an O-glycosylation site at 129 or 130 exhibit no O-glycosylation. This finding suggests that O-glycosylation at 129 or 130 requires O-glycosylation at another site, such as 139. A similar phenomenon has been reported in a different protein, namely, fibroblast growth factor 23 [43]. Our study also showed that the mutations intro-

duced to these three sites could not eliminate O-glycosylation in the linker region. This finding suggests that other O-glycosylation sites emerged as primary O-glycosylation sites when mutations were introduced to sites 129, 130, and 139. Therefore, we conclude that it is both O-glycosylation catalytic domain and the lectin binding domain that orchestrate the initialization of the first O-glycosylation site and the second O-glycosylation site.

4.4. Why O-glycan sialic acid residues do not proportionally contribute to the total sialic acid level?

The total sialic acid level is the sum of N-glycan sialic acid residues and O-glycan sialic acid residues, and expected to decrease with increasing number of O-glycosylation eliminations. However, the data in Fig. 3 (A) does not show this correlation until all O-glycosylation sites are eliminated. This inconsistent correlation could be caused by the following two factors. First, when one O-glycosylation site is void in the linker region, O-glycosylation occurs in other potential O-glycosylation sites in the surrounding area. Second, sialylation on N-glycans is the major contributor to the total sialic acid level, and may not be directly affected by O-glycosylation at a distal region.

4.5. How do reductions in O-glycan levels lead to low HMW?

The fusion protein has an interchain disulfide bond before the linker region. The disulfide bond is subject to reduction during the cell culture stage. The reduced fusion protein remains as a monomer given the strong dimerization propensity of CH3 domains in the Fc region. The free thiol in the reduced fusion protein can cross-link to another reduced fusion protein, forming a cross-linked dimer. Cross-linking reactions produce the formation of high molecular weight species. The broken disulfide bond can reform reversibly to prevent dimer and HMW formation. However, O-glycans in the linker region in the monomer prevent disulfide bond reformation due to space hindrance as hypothesized in Fig. 6 D. As shown in Fig. 6 A-C, the O-glycan can be as long as 18.7 Å depending on the number of sugar units. The presence of more O-glycans in the linker region reduces the opportunity for disulfide bond reformation, leading to higher HMW.

5. Conclusion

Glycosylation is critical for the quality of therapeutic proteins because it determines molecular stability, pharmacodynamic responses, and pharmacokinetic profiles. Thus, therapeutic efficacy can be optimized by tuning glycosylation. However, glycosylation is affected by various factors, including protein sequence, host cell, and cell culture condition. Achieving optimized glycosylation requires harmonization from early stage protein design, cell line development to late stage cell culture process development [44,45]. This study sought to understand O-glycosylation in the linker region of a CTLA4 Fc-fusion protein exclusively from a protein perspective. We found that (1) changing or even eliminating O-glycosylation in the linker region did not affect protein production and protein *in vitro* potency, (2) linker region flexibility facilitated O-glycosylation, and (3) altering O-glycosylation in the linker region changed product quality attributes, such as sialic acid and HMW levels. These molecular insights into O-glycosylation in the linker region of Fc-fusion protein will be beneficial to protein design in early stages of biologics development. However, questions, such as whether eliminating O-glycosylation in the linker region will result in glycosylation and sialic acid consistency in manufacturing processes and what the impacts of changing O-glycosylation in the linker region are on clinical efficacy, still remain to be answered.

CRediT authorship contribution statement

Yuanli Song: Data curation, Formal analysis, Writing - original draft. **Yueming Qian:** Conceptualization, Data curation, Formal analysis, Writing - review & editing. **Zhe Huang:** Data curation, Formal analysis. **Sarwat F. Khattak:** Conceptualization, Formal analysis. **Zheng Jian Li:** Project administration.

Declaration of Competing Interest

Yueming Qian, Sarwat Khattak, and Zheng Jian Li are the inventors of the patent "methods for reducing glycoprotein aggregation" (patent number: 9926365), which covered the experimental data of the current manuscript. The authors declare no other financial interests that might be perceived to influence the interpretation of the article.

Acknowledgments

Zhe Huang and Sarwat F. Khattak worked at Bristol Myers Squibb Co. in Syracuse during early stage of this study. The authors would like to thank their colleagues in Process Analytical Sciences at Bristol Myers Squibb Co. at the Syracuse site for their analytical support. The authors also would like to thank Drs. Nels Thorsteinson and Olivier Rousseau from the Chemical Computing Group for their support and advice in protein homology modeling and protein docking.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.11.037.

References

[1] Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. Curr Protocols Protein Sci 2013;73(1).
[2] Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. Acta Crystallogr F Struct Biol Commun 2015;71(Pt 7):861–9.
[3] Boado RJ, Zhang Y, Zhang Y, Wang Y, Pardridge WM. IgG-paraoxonase-1 fusion protein for targeted drug delivery across the human blood–brain barrier. Mol Pharmaceutics 2008;5(6):1037–43.
[4] Crivat G, Taraska JW. Imaging proteins inside cells with fluorescent tags. Trends Biotechnol 2012;30(1):8–16.
[5] Schmidt SR. Fusion-proteins as biopharmaceuticals–applications and challenges. Curr Opin Drug Discov Devel 2009;12(2):284–95.
[6] Zhao HL, Yao XQ, Xue C, Wang Y, Xiong XH, Liu ZM. Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. Protein Expr Purif 2008;61(1):73–7.
[7] Amet N, Lee H-F, Shen W-C. Insertion of the designed helical linker led to increased expression of Tf-based fusion proteins. Pharm Res 2009;26(3):523–8.
[8] Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev 2013;65(10):1357–69.
[9] Yu K, Liu C, Kim B-G, Lee D-Y. Synthetic fusion protein design and applications. Biotechnol Adv 2015;33(1):155–64.
[10] Bieberich E. Synthesis, processing, and function of N-glycans in N-glycoproteins. Adv Neurobiol 2014;9:47–70.
[11] Roth J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. Chem Rev 2002;102(2):285–303.
[12] Galleguillos SN, Ruckerbauer D, Gerstl MP, Borth N, Hanscho M, Zanghellini J. What can mathematical modelling say about CHO metabolism and protein glycosylation?. Comput Struct Biotechnol J 2017;15:212–21.
[13] Sha S, Agarabi C, Brorson K, Lee D-Y, Yoon S. N-glycosylation design and control of therapeutic monoclonal antibodies. Trends Biotechnol 2016;34(10):835–46.
[14] Wopereis S, Lefeber DJ, Morava E, Wevers RA. Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. Clin Chem 2006;52(4):574–600.

[15] Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature 2011;469 (7331):564–7.
[16] Tran DT, Ten Hagen KG. Mucin-type O-glycosylation during development. J Biol Chem 2013;288(10):6921–9.
[17] Gill DJ, Clausen H, Bard F. Location, location, location: new insights into O-GalNAc protein glycosylation. Trends Cell Biol 2011;21(3):149–58.
[18] Luther KB, Haltiwanger RS. Role of unusual O-glycans in intercellular signaling. Int J Biochem Cell Biol 2009;41(5):1011–24.
[19] Linsley PS, Nadler SG, Bajorath J, Peach R, Leung HT, Rogers J, Bradshaw J, Stebbins M, Leytze G, Brady W, Malacko AR, Marquardt H, Shaw S-Y. Binding Stoichiometry of the Cytotoxic T Lymphocyte-associated Molecule-4 (CTLA-4): a disulfide-linked homodimer binds two cd86 molecules. J Biol Chem 1995;270(25):15417–24.
[20] Bongers J, Devincentis J, Fu J, Huang P, Kirkley DH, Leister K, Liu P, Ludwig R, Rumney K, Tao Li, Wu W, Russell RJ. Characterization of glycosylation sites for a recombinant IgG1 monoclonal antibody and a CTLA4-Ig fusion protein by liquid chromatography–mass spectrometry peptide mapping. J Chromatogr A 2011;1218(45):8140–9.
[21] Zhu L, Guo Q, Guo H, Liu T, Zheng Y, Gu P, Chen Xi, Wang H, Hou S, Guo Y. Versatile characterization of glycosylation modification in CTLA4-Ig fusion proteins by liquid chromatography-mass spectrometry. mAbs 2014;6 (6):1474–85.
[22] Qian Y, Chen Z, Huang X, Wang X, Xu X, Kirov S, Ludwig R, Qian N-X, Ravi K, Tao L, Borys MC, Li ZJ. Early identification of unusually clustered mutations and root causes in therapeutic antibody development. Biotechnol Bioeng 2018;115(9):2377–82.
[23] Fu J, Bongers J, Tao Li, Huang D, Ludwig R, Huang Y, Qian Y, Basch J, Goldstein J, Krishnan R, You Li, Li ZJ, Russell RJ. Characterization and identification of alanine to serine sequence variants in an IgG4 monoclonal antibody produced in mammalian cell lines. J Chromatogr B 2012;908:1–8.
[24] Jing Y, Qian Y, Li ZJ. Sialylation enhancement of CTLA4-Ig fusion protein in Chinese hamster ovary cells by dexamethasone. Biotechnol Bioeng 2010;107 (3):488–96.
[25] Qian Y, Jing Y, Li ZJ. Glucocorticoid receptor-mediated reduction of IgG-fusion protein aggregation in Chinese hamster ovary cells. Biotechnol Progress 2010;26(5):1417–23.
[26] Šali A, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol 1993;234(3):779–815.
[27] Yu C, Sonnen A-P, George R, Dessailly BH, Stagg LJ, Evans EJ, Orengo CA, Stuart DI, Ladbury JE, Ikemizu S, Gilbert RJC, Davis SJ. Rigid-body ligand recognition drives cytotoxic t-lymphocyte antigen 4 (CTLA-4) receptor triggering. J Biol Chem 2011;286(8):6685–96.
[28] Lobner E, Humm AS, Goritzer K, Mlynek G, Puchinger MG, Hasenhindl C, et al. Fcab-HER2 Interaction: a Menage a Trois. lessons from X-Ray and solution studies. Structure 2017;25(6):878–889 e875.
[29] Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. J Phys Chem 1987;91(24):6269–71.
[30] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701–18.
[31] Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118(45):11225–36.
[32] Fritz TA, Raman J, Tabak LA. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. J Biol Chem 2006;281(13):8613–9.
[33] Larsen CP, Pearson TC, Adams AB, Tso P, Shirasugi N, StrobertM E, Anderson D, Cowan S, Price K, Naemura J, Emswiler J, Greene JoAnne, Turk LA, Bajorath J, Townsend R, Hagerty D, Linsley PS, Peach RJ. Rational development of LEA29Y (belatacept), a high-affinity variant of CTLA4-Ig with potent immunosuppressive properties. Am J Transplant 2005;5(3):443–53.
[34] Julenius K, Molgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 2005;15(2):153–64.
[35] Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J 2013;32(10):1478–88.
[36] Adzhubei AA, Sternberg MJE, Makarov AA. Polyproline-II Helix in Proteins: Structure and Function. J Mol Biol 2013;425(12):2100–32.
[37] Bergstrom KSB, Xia L. Mucin-type O-glycans and their roles in intestinal homeostasis. Glycobiology 2013;23(9):1026–37.
[38] Hanisch FG. O-glycosylation of the mucin type. Biol Chem 2001;382(2):143–9.
[39] Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology 2012;22(6):736–56.
[40] Yoshida A, Suzuki M, Ikenaga H, Takeuchi M. Discovery of the shortest sequence motif for high level mucin-type O-glycosylation. J Biol Chem 1997;272(27):16884–8.
[41] Raman J, Fritz TA, Gerken TA, Jamison O, Live D, Liu M, et al. The catalytic and lectin domains of UDP-GalNAc:polypeptide alpha-N-Acetylgalactosaminyltransferase function in concert to direct glycosylation site selection. J Biol Chem 2008;283(34):22942–51.
[42] de las Rivas M, Lira-Navarrete E, Daniel EJP, Compañón I, Coelho H, Diniz A, Jiménez-Barbero J, Peregrina JM, Clausen H, Corzana F, Marcelo F, Jiménez-Osés G, Gerken TA, Hurtado-Guerrero R. The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. Nat Commun 2017;8(1).
[43] de las Rivas M, Paul Daniel EJ, Narimatsu Y, Compañón I, Kato K, Hermosilla P, Thureau A, Ceballos-Laita L, Coelho H, Bernadó P, Marcelo F, Hansen L, Maeda R, Lostao A, Corzana F, Clausen H, Gerken TA, Hurtado-Guerrero R. Molecular basis for fibroblast growth factor 23 O-glycosylation by GalNAc-T3. Nat Chem Biol 2020;16(3):351–60.
[44] Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. Glycobiology 2009;19(9):936–49.
[45] Butler M. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. Cytotechnology 2006;50(1-3):57–76.
[46] Schwartz JC, Zhang X, Fedorov AA, Nathenson SG, Almo SC. Structural basis for co-stimulation by the human CTLA-4/B7-2 complex. Nature 2001;410 (6828):604–8.