JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau

Article

High Density of N- and O-Glycosylation Shields and Defines the Structural Dynamics of the Intrinsically Disordered Ectodomain of Receptor-type Protein Tyrosine Phosphatase Alpha

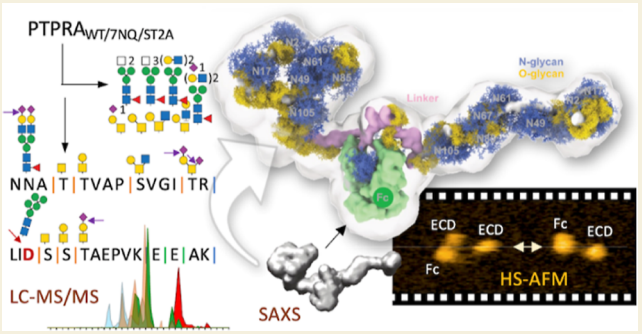
Yu-Chun Chien, Yong-Sheng Wang, Deepa Sridharan, Chu-Wei Kuo, Chih-Ta Chien, Takayuki Uchihashi, Koichi Kato, Takashi Angata, Tzu-Ching Meng, Shang-Te Danny Hsu, and Kay-Hooi Khoo*

Cite This: *JACS Au* 2023, 3, 1864–1875

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The intracellular phosphatase domain of the receptor-type protein tyrosine phosphatase alpha (PTPRA) is known to regulate various signaling pathways related to cell adhesion through c-Src kinase activation. In contrast, the functional significance of its relatively short, intrinsically disordered, and heavily glycosylated ectodomain remains unclear. Through detailed mass spectrometry analyses of a combination of protease and glycosidase digests, we now provide the first experimental evidence for its site-specific glycosylation pattern. This includes the occurrence of *O*-glycan at the N-glycosylation sequon among the more than 30 O-glycosylation sites confidently identified beside the 7 N-glycosylation sites. The closely spaced N- and O-glycans appear to have mutually limited the extent of further galactosylation and sialylation. An immature smaller form of full-length PTPRA was found to be deficient in O-glycosylation, most likely due to failure to transit the Golgi. N-glycosylation, on the other hand, is dispensable for cell surface expression and contributes less than the extensive O-glycosylation to the overall solution structure of the ectodomain. The glycosylation information is combined with the overall structural features of the ectodomain derived from small-angle X-ray scattering and high-speed atomic force microscopy monitoring to establish a dynamic structural model of the densely glycosylated PTPRA ectodomain. The observed high structural flexibility, as manifested by continuous transitioning from fully to partially extended and fold-back conformations, suggests that the receptor-type phosphatase is anchored to the membrane and kept mostly at a monomeric state through an ectodomain shaped and fully shielded by glycosylation.



KEYWORDS: *glycopeptide sequencing, glycoproteomics, mass spectrometry, SAXS, HS-AFM, receptor-type protein tyrosine phosphatase*

INTRODUCTION

The receptor-type protein tyrosine phosphatase alpha (PTPRA; encoded by *ptpra* gene) belongs to the R4 subfamily of human protein tyrosine phosphatases, which includes PTPRA and receptor-type protein tyrosine phosphatase epsilon (PTPRE).[1] Both PTPs carry a short extracellular domain and two tandem phosphatase domains in the intracellular regions. The plasma membrane-proximal D1 phosphatase domain of the ubiquitously expressed PTPRA would dephosphorylate Tyr527 of the non-receptor tyrosine kinase c-Src,[2−4] leading to conformational change-mediated kinase activation. Through the regulation of c-Src activity, PTPRA may modulate various growth factor receptor and cell adhesion signaling pathways.[5,6] For example, c-Src activation is essential for fibronectin-induced cell spreading and mobility in integrin-dependent adhesion in fibroblasts.[7] PTPRA$^{-/-}$ fibroblasts showed a decrease of c-Src, Fyn, and FAK kinase activity, resulting in a delay of fibroblast spreading on fibronectin- or vitronectin-coated surface.[6,8] Recent studies on lung fibroblast[9,10] further suggested that PTPRA is an activator of c-Src kinase to induce downstream signaling required for cell adhesion and migration.

Contrary to the relatively well-characterized functions of the D1 phosphatase domain in regulating c-Src activity, very little is known about the role of the ectodomain of PTPRA.[5] This short ectodomain of only 122 amino acids is intrinsically disordered without any defined structural features, as predicted by AlphaFold.[11] To date, PTPRA remains an orphan receptor,[12] and its ectodomain may not be responsible for ligand binding. Although it has been proposed to promote cell adhesion of

Received: March 17, 2023
Revised: June 1, 2023
Accepted: June 1, 2023
Published: June 13, 2023

ACS Publications

© 2023 The Authors. Published by American Chemical Society

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864–1875

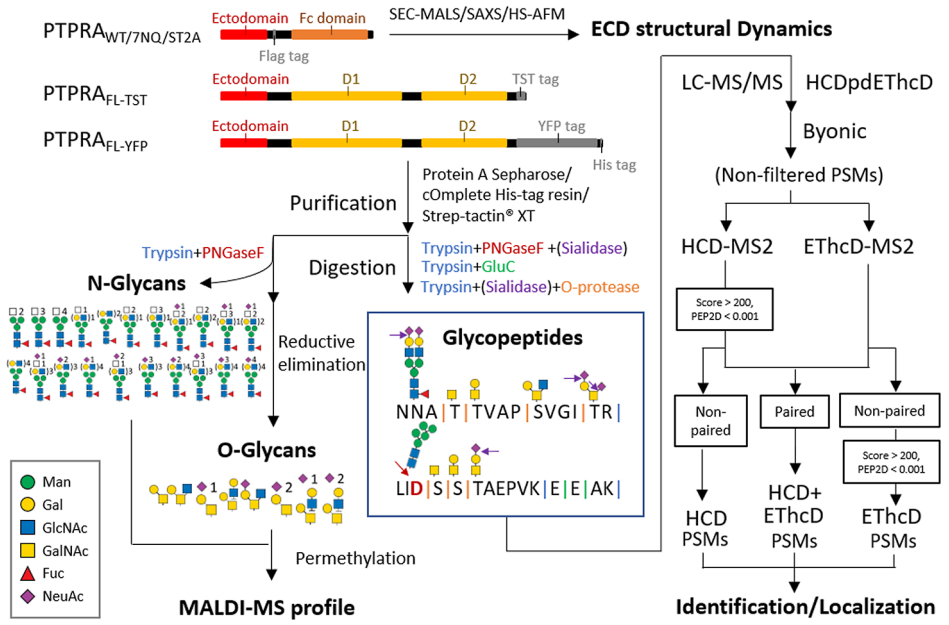


Figure 1. Workflow for glycosylation analysis of PTPRA. The recombinant PTPRA ectodomain-Fc fusion protein and full-length PTPRA were digested with various glycosidases and proteases. N- and O-glycans were released from one batch of trypsin-digested PTPRA sample and permethylated for MALDI-MS profiling. Intact glycopeptides from all different combinations of digests were analyzed by LC−MS/MS using a HCDpdEThcD data acquisition mode and processed by Byonic. The HCD-MS^2 peptide spectrum matches (PSMs) were filtered by score > 200 and PEP2D < 0.001 and then manually paired with non-filtered EThcD-MS^2 PSMs identifying glycopeptides of the same m/z, sequence, and glycan within five scans. EThcD-MS^2 PSMs that could not be paired were filtered by the same cutoff criteria. Site localization was manually verified based on the supporting c and z ions afforded by EThcD-MS^2. Glycosylation data obtained were then used to aid structural model building of the ectodomain (ECD) based on the SEC-MALS, SAXS, and HS-AFM structural data of $PTPRA_{WT}$ in comparison with non-glycosylated $PTPRA_{ST2A}$. Notation for the PTPRA constructs: WT, 7NQ, and ST2A refer to Fc fused to wild-type ECD, ECD with all 7 Asn mutated to Gln, and ECD with all Ser/Thr mutated to Ala, respectively. FL-TST and FL-YFP refer to full-length protein fused to a twin-strep-tag or a YFP. Glycan cartoons used here and in all other figures follow the recommended Symbol Nomenclature for Glycans (SNFG).[30]

fibroblast-like synoviocytes,[13] its exact function in fibronectin interaction is poorly understood. To complicate the matter, the ectodomain of PTPRA is known to be heavily glycosylated,[14] but neither the structural feature nor the biological role of the attached glycans has been definitively proven. The current Uniprot record of PTPRA notes that it is N-glycosylated at seven sites based on sequence analysis, but only one de-N-glycosylated site at Asn36 has been experimentally identified and annotated in GlyGen.[15] An examination of the ectodomain sequence shows a high occurrence of Ser and Thr, among which Ser37/Ser38/Thr39/Thr49 are recorded as O-glycosylated in the O-GalNAc Human Glycoproteome Database[16] based on the accumulated SimpleCell data.[17] In a recent mucinomics study, a mutated mucin-specific O-glycoprotease without cleavage activity was used to selectively capture glycoproteins carrying a high density of O-glycans. Among the 47 PTP family proteins known, only PTPRA was thus identified in every cancer cell line (HeLa, SKBR3, Capan2, K562, and OVCAR3) examined.[18] However, when the putative mucin-domain containing proteins identified were filtered through their mucin-domain candidacy algorithm based on predicted O-GalNAc sites by NetO-Glyc4.0,[19] cellular localization GO terms, the proportion of Thr to Ser residues, and the frequency and density of predicted O-GalNAc sites, only PTPRC and PTPRQ, but not PTPRA, qualified as mucin-domain proteins.

To gain a better insight into how glycosylation may impact the structure and functional roles of the ectodomain of PTPRA, it is imperative to address foremost whether it indeed is densely populated by O-glycans in addition to putative N-glycosylation. Identification of intact N-glycopeptides by mass spectrometry (MS)-based glycoproteomic analysis can now be handled fairly well by available data analysis software,[20] but tackling site-specific O-glycosylation is still problematic.[21−23] To localize individual O-glycans at distinct sites instead of just providing an overall glycosyl composition, good quality electron transfer dissociation (ETD) MS^2 data are required that would retain the O-glycans, which otherwise fall off easily during the normal mode of collision-induced dissociation.[24,25] When there is an additional N-glycan on the same multiply O-glycosylated peptide, it is largely an uncharted area. Moreover, the common lack of properly spaced tryptic sites within the mucin domain necessitates the use of O-glycoprotease[26,27] and other less specific cleavages that would significantly broaden the search space and exacerbate the problem of mis-assignments.

In this work, multiple LC−MS/MS analysis data sets derived from different combinations of protease and glycosidase digests of PTPRA ectodomain were analyzed in concert to build up a composite picture bottom-up. We established some ground rules governing the interpretation and correct assignment of multiple O-glycosylated peptides, particularly in confidently identifying O-glycosylation at and adjacent to the N-glycosylated sequon. We then fitted the experimentally determined site-specific glycosylation profile into the structural envelopes derived from small angle X-ray scattering (SAXS) and high-speed atomic force microscopy (HS-AFM) to build a dynamic structural model of two heavily N- and O-glycosylated PTPRA ectodomains brought together by a compact Fc domain anchor. The continuous transitioning from fully to partially extended and fold-back conformations recorded in real-time HS-AFM is provocative of how we may rethink the functional

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

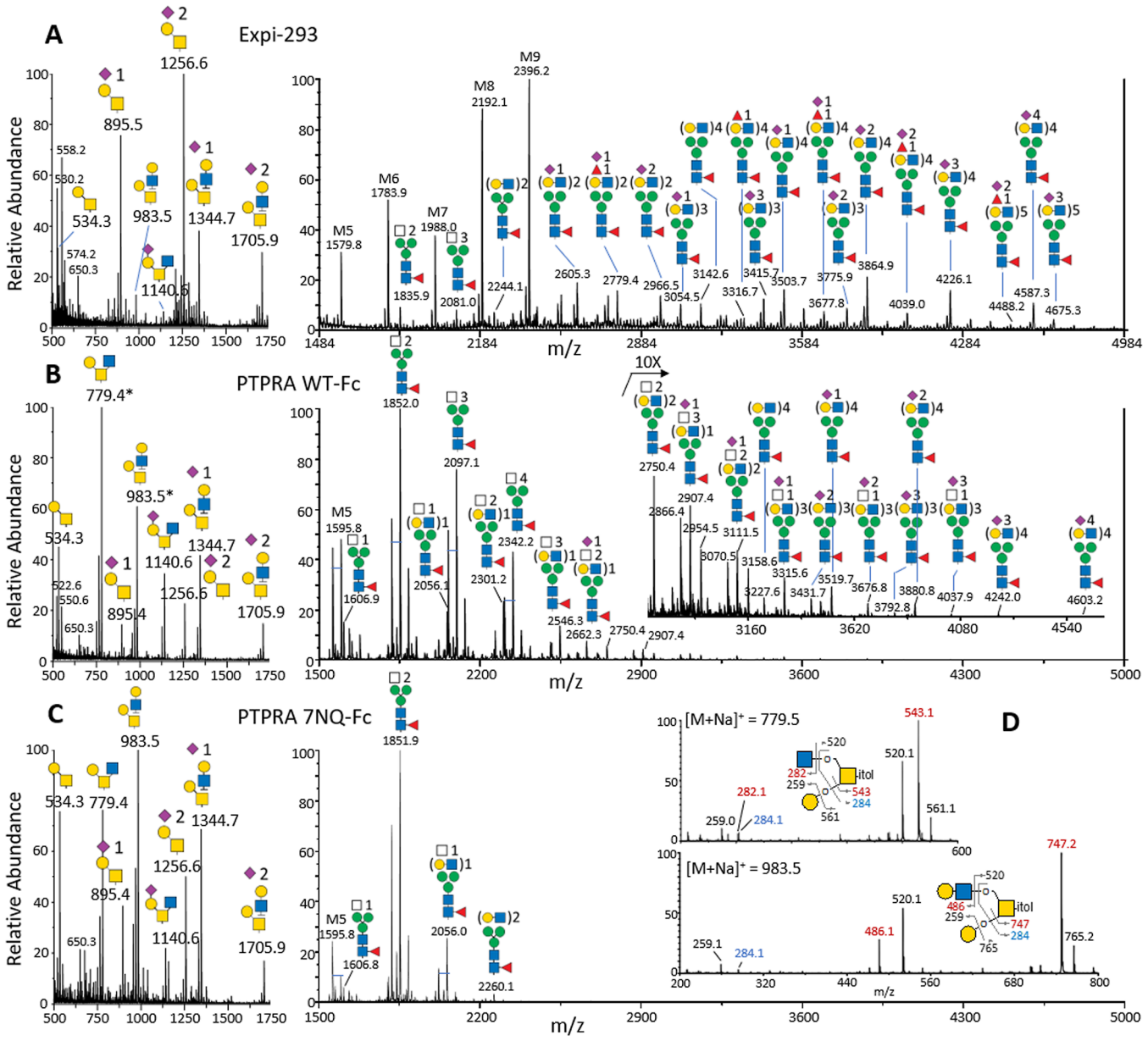


Figure 2. MALDI-MS profiles of permethylated glycans released from Expi293F cells, WT and 7NQ mutant of PTPRA-Fc expressed in Expi293F cells. The O-glycans (left panels) were obtained by reductive elimination after removing the N-glycans (right panels) by PNGase F from the tryptic digested Expi293F protein extracts (A) or isolated $PTPRA_{WT}$-Fc (B) and $PTPRA_{7NQ}$-Fc (C). The N-glycans obtained from the isolated PTPRA-Fc proteins were further reduced before permethylation, whereas those from Expi293F cells were permethylated directly for MS analysis. The comparative analyses show that other than the major G0-biantennary N-glycan (m/z 1851.9) partly contributed by the Fc domain, the N-glycans carried on recombinant $PTPRA_{WT}$ ectodomain were dominated by those with 2 (m/z 2097.1) and 3 (m/z 2342) terminal HexNAc on fucosylated trimannosyl core, accompanied by a range of further extended and sialylated complex-type structures at low abundance but with almost no high mannose type detected, which is very different from the whole cell N-glycomic profile. Its O-glycans also differ significantly from those of whole cell by a higher abundance of core 2 structures and non-galactosylated HexNAc termini, as confirmed by MS/MS analysis (D) on the two representative major structures (marked by an asterisk in B). Other detected glycans were annotated based simply on glycosyl composition assignment. Undefined terminal HexNAc not paired by a Hex on the N-glycans is represented by a non-colored square since Expi293F cells may carry terminal GalNAc-GlcNAc in addition to ± Gal-GlcNAc.

aspects of an intrinsically disordered ectodomain essentially shaped and fully shielded by glycosylation.

RESULTS

Experimental Workflow for Glycosylation Analysis of PTPRA

Detailed MS-based analyses were first performed on the ectodomain of PTPRA that was fused to Fc ($PTPRA_{WT}$) to allow it to be secreted and isolated in a larger quantity than a full-length transmembrane form ($PTPRA_{FL}$) would afford. Two alternative forms of the latter were constructed, exploring two different tags for their subsequent purification (Figure 1). The lack of tryptic sites other than suitably producing the first two N-terminal peptides necessitated additional digests by other proteases, including GluC that cleaves at Asp/Glu, and OpeRATOR, the O-glycoprotease that would cleave specifically

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

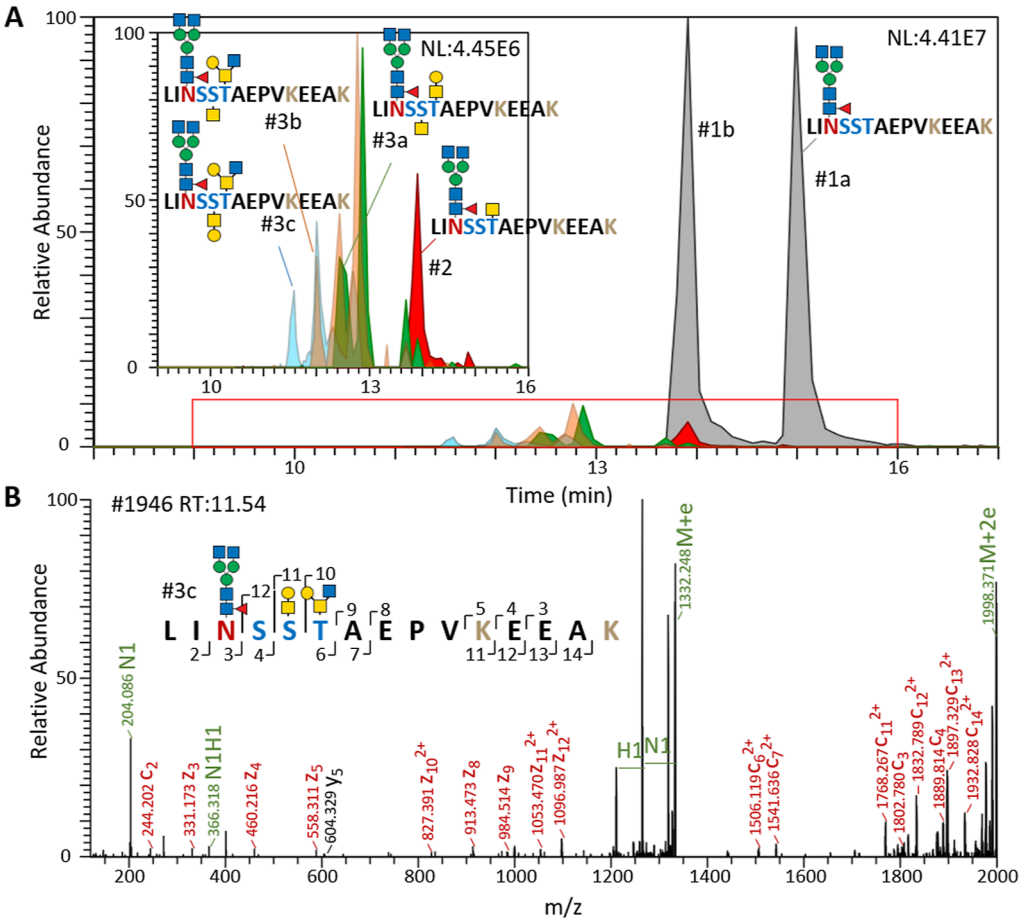


Figure 3. Identification of $PTPRA_{WT}$ glycopeptides carrying both N- and O-glycans. (A) In addition to carrying only an N-glycan of the composition $HexNAc_4Hex_3Fuc_1$ (gray, peak no. 1a), at least four other minor glycoforms (see inset for a magnified view) were identified for the same N17-containing peptide backbone, LINSSTAEPVKEEAK, derived from $PTPRA_{WT}$ digested by trypsin and OpeRATOR. Shown here are their respective extracted ion chromatograms (XICs) with overall glycosyl compositions $HexNAc_5Hex_3Fuc_1$ (red), $HexNAc_6Hex_4Fuc_1$ (green), $HexNAc_7Hex_4Fuc_1$ (orange), and $HexNAc_7Hex_5Fuc_1$ (light blue), each resolved into several peaks. Among them, only the site-specific O-glycosylation patterns of peaks annotated as no. 2, no. 3a, no. 3b, and no. 3c could be delineated by EThcD-MS^2 (see Figure S2), with the supporting EThcD-MS^2 spectrum for one of them shown here in (B).

at the N-termini of O-glycosylated Ser/Thr.[26,28,29] Removal of N-glycans by PNGase F and/or the negatively charged terminal sialic acids by sialidase was additionally undertaken to facilitate proteolytic digestion and LC−MS/MS analysis of the simplified glycopeptides. We also investigated if preventing the N-glycosylation on the ectodomain of $PTPRA_{WT}$ by mutating the Asn at the sequons to Gln ($PTPTA_{7NQ}$) may alter the extent of O-glycosylation and/or allow more efficient digestion. Higher energy collision dissociation (HCD) MS^2 scans were acquired in data-dependent mode, while those producing the glycopeptide-specific glyco-oxonium ions were further subjected to EThcD MS^2. The HCD-pd-EThcD data sets were processed by Byonic for peptide spectrum matches (PSMs), which were further filtered by stringent acceptance criteria, with the corresponding HCD and EThcD PSMs for the same precursors paired for cross validation and O-glycosylation site localization, as outlined in Figure 1. In addition, N- and O-glycans were released from the trypsin-digested proteins and profiled by MALDI-MS after permethylation to determine if the major glycans carried are consistent with those identified through intact glycopeptide analyses.

Overall Glycosylation Profile of PTPRA Ectodomain

In comparison with the glycomic profile of Expi-293 cells (Figure 2A), the N-glycans of PTPRA were strikingly dominated by a few complex type bi-, tri-, and tetra-antennary complex type N-glycans carrying non-galactosylated terminal GlcNAc (Figure 2B). High mannose $Man_5GlcNAc_2$ as well as fully extended, variably sialylated complex type N-glycans were also detected but at significantly lower abundance. A portion of the non-galactosylated biantennary N-glycans would have been contributed by that carried on the Fc domain, as revealed by the profile of $PTPRA_{7NQ}$ (Figure 2C) compared to that of $PTPRA_{WT}$ (Figure 2B). This is consistent with the well-known predominant form of Fc N-glycosylation constrained by spatial accessibility,[31−33] but itself does not account for the additional high amount of tri- and tetra-antennary structures that are authentic of $PTPRA_{WT}$. The conversion of a majority of high-mannose-type to complex-type structures and yet mostly not extended beyond the addition of first GlcNAc suggests that the processing of N-glycans is significantly impeded not at the early stage in the Golgi but later on, presumably after the trimming of mannoses, adding of antennary GlcNAc, and initiation of O-glycans.

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

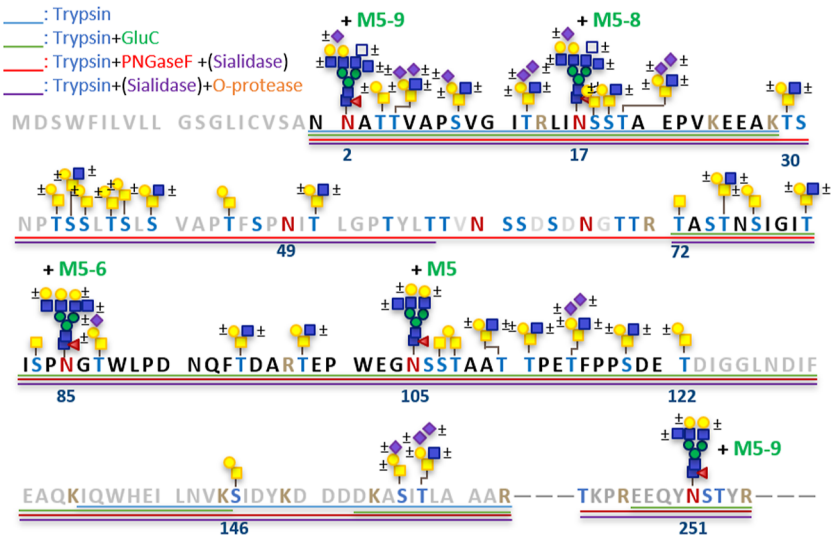


Figure 4. Identified and localized glycans on the ectodomain of $PTPRA_{WT}$. Data were generated from concerted analyses of $PTPRA_{WT}$ glycopeptides derived from a combination of PNGase F, sialidase, trypsin, GluC, and/or O-glycoprotease digests. Only those site-specific O-glycosylations that are supported by EThcD-MS^2 are annotated. The ± sign indicates that O-glycan structures with and without the extra glycosyl residue were both identified. M5-9 refers to high mannose structures $Man_{5\text{-}9}GlcNAc_2$. Additional glycosylation identified by HCD-MS^2 analysis but not further delineated nor localized to specific sites were excluded from this overall composite picture. Full listing of the identified glycopeptides is provided in Table S1.

While the overall O-glycomic profile of Expi-293F was dominated by mono- and disialylated core 1 and 2 structures (Figure 2A), the O-glycans released from $PTPRA_{WT}$ (Figure 2B) comprised a significantly higher amount of incompletely galactosylated ($HexNAc_1$-$Hex_1HexNAcitol$, m/z 779) and non-sialylated ($Hex_1HexNAc_1$-$Hex_1HexNAcitol$, m/z 983) core 2 structures. This indicates that a larger proportion of the O-glycans carried on the PTPRA were prone to core 2 branching, but their subsequent extension and sialylation were relatively incomplete. In that respect, without the potentially crowding N-glycans, the O-glycans on non-N-glycosylated $PTPRA_{7NQ}$ were found to contain a higher proportion of mono- and disialylated core 1 and 2 structures (Figure 2C), albeit not as high as detected for the Expi293 cells. It can thus be concluded that the occurrence of closely spaced N- and O-glycans is likely to have mutually exerted steric hindrance effects, leading to an obvious inhibition of the downstream glycosylation events such as further galactosylation and sialylation. Our subsequent analyses at the glycopeptide level further showed that the entire ectodomain is densely O-glycosylated, which is reflective of its initial accessibility to efficient O-glycosylation.

Identification of N- and O-Glycosylation Sites on $PTPRA_{WT}$

Initial tryptic digests of $PTPRA_{WT}$ only afforded a limited sequence coverage (24%) of the ectodomain of PTPRA, whereby a majority of confidently identified glycopeptides were contributed by cleavages at the first two N-terminal tryptic sites. Among the many contributing factors was a genuine lack of tryptic sites suitably located within the internal Ser/Thr-rich sequences, many of which were likely to be O-glycosylated. Additional use of GluC and OpeRATOR significantly increased the mapped sequence coverage to 66 and 89%, respectively. Since OpeRATOR was meant to cleave only at O-glycosylated Ser/Thr, the results provided the first evidence that many of the Ser/Thr at the N-termini of the resulting glycopeptides were O-glycosylated. However, due to the often-incomplete cleavages, the O-glycans may also be found at the internal sites and not necessarily only at the N-terminal Ser/Thr of the resulting peptides. Additional EThcD MS^2 to complement the peptide backbone identification by HCD MS^2 is therefore still required to determine the site-specific distribution.

It is clear from further manual validation of the EThcD MS^2 data that many of the identified peptides including the first two N-terminal tryptic peptides were not only N-glycosylated but also variably O-glycosylated. Based on the conserved fucosylated trimannosyl core structure, additional HexNAc and Hex could either be assigned as non-galactosylated GlcNAc or Gal-GlcNAc extension on the antennae of N-glycans, or as additional O-glycans. In the case of N49-containing glycopeptides, for example, two N-glycosylated peptides with the same peptide backbone but carrying different glycosyl compositions could be identified by HCD MS^2 and detected as two separately eluting peaks (Figure S1A, B). Although the entire $HexNAc_6Hex_{4,5}Fuc_1$ could be attributed to N-glycan at N49, it is equally likely that one or more of the Hex and HexNAc residues came from an O-glycan. This ambiguity could only be resolved if critical EThcD MS^2 cleavage ions can be detected, aided by a distinctive elution pattern. Notably, it was found that for the same N-glycosylated peptide backbone, each additional presence of O-glycan at distinct sites would contribute to an earlier elution time more significantly than having an additional Hex or/and HexNAc residue on the glycan itself, as shown in Figure 3.

In this case, the peptide LINSSTAEPVKEEAK with only an N-glycan, $HexNAc_4Hex_3Fuc_1$, was found to elute at around 15 min (peak no. 1a), while the same N-glycosylated peptide with an additional O-glycan eluted at around 14 min (peak no. 2) and that with two O-glycans eluted at around 11.5−13 min (peaks no. 3a−c, see Figure 3B and S2 for supporting EThcD MS^2 spectra). This characteristic feature is very useful in deducing the likely N- and O-glycoforms, especially when the EThcD MS^2 data, or the lack of it, failed to resolve the site localization ambiguity. The earlier eluting, non-identified peak no. 1b could thus be rationalized as the same peptide carrying an extra O-GalNAc besides a $HexNAc_3Hex_3Fuc_1$ N-glycan for it to be resolved and distinguished from the later eluting isomeric

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

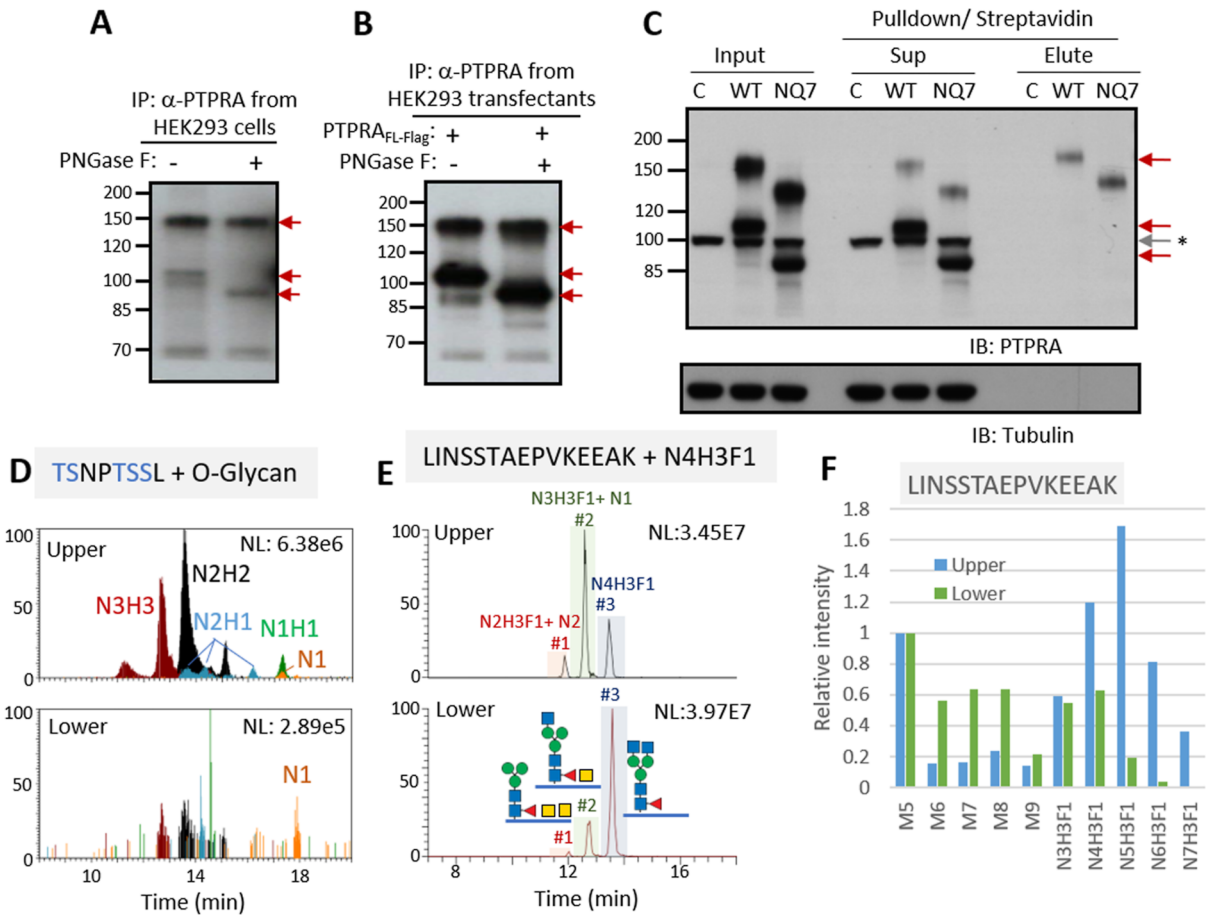


Figure 5. Glycosylation analyses of full-length PTPRA. Western blot analysis of endogenous PTPRA (A) and $PTPRA_{FL\text{-}WT}$ (B) overexpressed in HEK293 cells and immunoprecipitated by anti-PTPRA antibody revealed two gel bands for both PNGase F-treated and non-treated samples, as indicated by arrows. Only the upper band from $PTPRA_{FL\text{-}WT}$ and that from $PTPRA_{FL\text{-}7NQ}$ overexpressing HEK293 cells could be surface-labeled by sulfo-NHS-Biotin and pulled down by streptavidin (C). Non-captured and captured proteins were recovered in the supernatant (lanes 4−6) and eluted (lanes 7−9) fractions, respectively. Control cells were transfected with an empty vector. * indicates protein band from non-specific binding. The two gel bands similarly produced by $PTPRA_{FL}$ with YFP tag ($PTPRA_{FL\text{-}YFP}$) were excised and digested with trypsin, ±sialidase, and OpeRATOR for glycopeptide analysis (D−F). Samples from the upper band yielded more glycopeptides carrying additional O-glycans, as indicated by XIC plots for the peptide TSNPTSSL with a range O-glycans (D), and the peptide LINSSTAEPVKEEAK with an overall N4H3F1 composition (E). The ion intensities of the latter peptide carrying different overall glycan compositions relative to that with a single M5 N-glycan (intensity set as 1.0), as identified in samples derived from the upper and lower bands, are shown in (F). F, Fuc; H, Hex; M, Man; N, HexNAc.

glycoforms (peak no. 1a) carrying only a $HexNAc_4Hex_3Fuc_1$ N-glycan. Similarly, the peptide eluted at 13 min (peak no. 3a) could be assigned as most likely carrying two extra O-glycans, a single HexNAc and a Hex-HexNAc, in addition to the $HexNAc_4Hex_3Fuc_1$ N-glycan, despite not producing sufficiently discriminating EThcD MS^2 data (Figure S2B) to define their site-specific distribution. In total, a combination of trypsin, GluC, and OpeRATOR digestions led to the confident assignment of 5 N-glycosylated Asn sites and at least 26 O-glycosylated Ser/Thr on the ectodomain (Figure 4 and Table S1).

Additional O-Glycosylation Identified upon de-N-glycosylation and Desialylation

To avoid a biased sampling of the resulting glycopeptides due to their carrying a bulky N-glycans that adversely affected the MS ionization and detection efficiency, PNGase F was employed to remove the N-glycans from the glycopeptides, but this additional step did not contribute significantly to additional glycopeptides being positively identified. Only 2 O-glycosylation sites (Thr5 and Ser19) were additionally identified with core 2 O-glycans, when the N-glycans were first removed (Figure 4 and Table S1). However, when coupled with additional sialidase treatment ($PTPRA_{WT+NF+S}$), it allowed the two remaining N-glycosylated sites, N61 and N67, to be identified on the same de-*N*-glycosylated, desialylated peptide by virtue of detecting a deamidated Asp instead of Asn at both the N61 and N67 sequons (Figure S1C).

To prevent sialylation on the O-glycans from hindering efficient cleavages of OpeRATOR, $PTPRA_{WT}$ was also treated with sialidase before OpeRATOR digestion. This successfully led to a few additional glycopeptides being detected. It could be reasonably inferred that the O-glycans at the N-termini of newly produced glycopeptides were originally sialylated and thus not susceptible to OpeRATOR. In all, 5 additional O-glycosylation sites (Thr45, Ser83, Ser107, Thr108, and Ser119) were identified, and 6 sites (Ser41, Thr51, Thr75, Thr95, Thr112, and Ser119) were found to additionally carry truncated core 2 O-glycans ($Hex_1HexNAc_2$) (Figure 4 and Table S1), implying that sialylation may render them less amenable to detection and hence underestimated. Strikingly, core 2 O-glycans could be identified on Thr115 and Ser119 only when these residues were

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

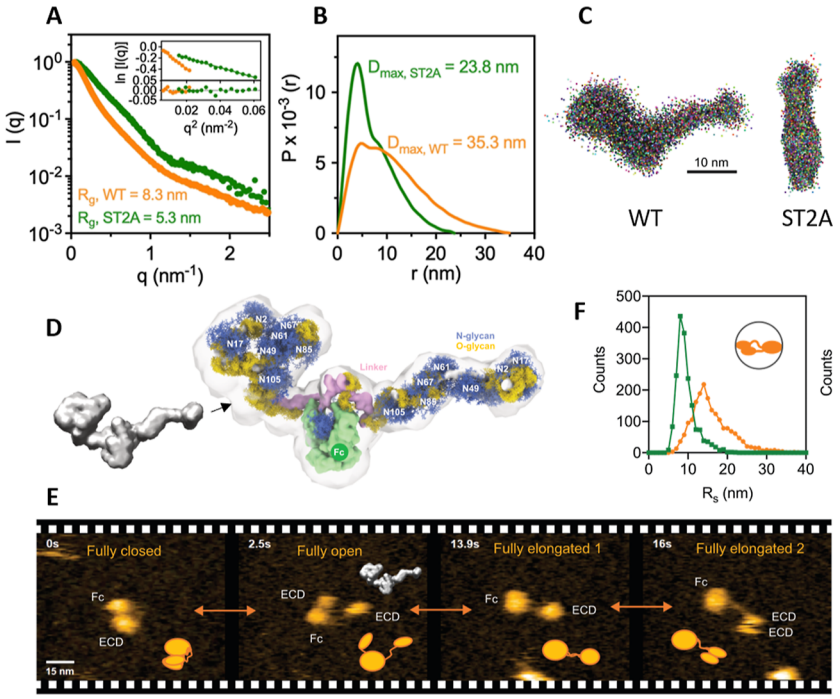


Figure 6. Conformation and structural dynamics of PTPRA ectodomain. The raw SAXS profiles (A) of $PTPRA_{WT}$ (orange) and $PTPRA_{ST2A}$ (green), with the Guinier approximation for R_g estimations and the residuals of the linear fitting results shown in the inset, pair-wise distance distribution, $P(r)$, shown in (B), and superposition of DAMMIF[39] bead models in (C). The surface representation of the most representative DAMMIF model for $PTPRA_{WT}$ (D, left) was additionally shown in semitransparent volume, low-pass filtered to 2.5 nm by ChimeraX,[40] and overlaid with its atomic model (D, right). The protein elements of Fc, linker, and ECD are low-pass filtered to a resolution of 0.8 nm by ChimeraX and colored in lime, pink and white, respectively. The ensemble structures of N- and O-glycans are shown in sticks and colored in blue and gold, respectively, with the positions of individual N-glycosylation sites indicated. Each ensemble consists of 20 structures derived from the trajectories of fully atomistic molecular dynamics simulations within the GlycoSHIELD repository.[41] Representative snapshots of HS-AFM trajectories of $PTPRA_{WT}$ at the indicated time points (E) are interpreted as adopting fully closed, open or elongated domain organizations as illustrated by the cartoon drawings, with the putative positions of Fc and ECDs indicated and the SAXS-derived molecule envelope placed next to the fully open form. The corresponding HS-AFM trajectory snapshots of $PTPRA_{ST2A}$ are shown in Figure S11E. The cumulative counts of R_s derived from five independent trajectories of $PTPRA_{WT}$ (orange) and $PTPRA_{ST2A}$ (green) are shown in (F). It defines the maximum separation between any pair of pixels within the single molecules, equivalent to the D_{max} values derived from SAXS analyses (B).

not located at the N-termini (Figure S3), suggesting that OpeRATOR failed to cleave at the core 2 O-glycosylated Ser/Thr even when non-sialylated, consistent with the specificity preference of the enzyme reported by a previous study.[29] Combining the various enzyme digestion and glycopeptide identification results, it can be concluded that the ectodomain of PTPRA was N-glycosylated at all 7 sites, accompanied by closely packed O-glycosylated sites, at least 31 of which (out of a total 43 Ser/Thr) could be confidently verified (Figure 4 and Table S1).

Interestingly, some of the identified O-glycosylated Ser actually corresponded to the Ser of the Asn-Xxx-Ser/Thr sequon, providing solid evidence for such a rarely reported occurrence[34] (Figure 3B). Within the first two tryptic N-glycopeptides, we found that the Ser or Thr proximal to the N-glycosylated Asn were O-glycosylated by a single GalNAc (Tn) or Gal-GalNAc (core 1 O-glycan), while those further away were occupied by more sialylated and/or core 2 O-glycans (Figure 4 and Table S1). This indicates that further elaboration of the O-glycans was somewhat hindered by the bulky N-glycan nearby. Other than that, there is no apparent site-specificity preference detected for any of the glycans identified on each site. To see if preventing N-glycosylation by N → Q mutation may favor the expression of more elaborated O-glycans, $PTPRA_{7NQ}$ was similarly digested and analyzed using the same analytical workflow. Indeed, many of the same peptide backbones were identified as carrying more core 1 O-glycans, with also an overall higher degree of sialylation detected (see Figure S4 for the first tryptic peptide). Collectively, the emerging picture is largely consistent with the profiles of released N- and O-glycans (Figure 2). However, it is also apparent that highly complex, extended and sialylated glycans are not well represented on the identified glycopeptides partly due to their less favorable MS response but also due to their low abundance in the first instance.

Characterization of Glycans Attached to the Full-Length PTPRA Expressed in HEK293 Cells

Previous studies have shown that recombinant full-length PTPRA produced in COS-1 cells,[14] N202.1A cells,[35] and HEK 293 cells[36] would run as two distinct gel bands on SDS-PAGE. This was recapitulated by our own Western blot analysis of both endogenous and overexpressed $PTPRA_{FL\text{-}WT}$ in HEK293T cells (Figure 5A,B). PNGaseF treatment significantly shifted the lower band at ~105 kDa to around the theoretical molecular weight (~90 kDa) for both samples but only marginally shifted the upper band at ~150 kDa. Since the non-N-glycosylated $PTPRA_{FL\text{-}7NQ}$ similarly produced two bands, with the upper band running at significantly lower position (~135 kDa) than that of $PTPRA_{FL\text{-}WT}$ (Figure 5C, left panel), it could be concluded that both upper and lower bands of $PTPRA_{FL\text{-}WT}$ were N-glycosylated but only the N-glycans from the lower band form could be removed efficiently. When the

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

cell-surface localized PTPRA was labeled by the Sulfo-NHS-Biotin and then pulled down by Streptavidin resins, it was found that only the upper band forms of both $PTPRA_{FL\text{-}WT}$ and $PTPRA_{FL\text{-}7NQ}$ were captured and therefore recovered mostly in the eluates, whereas the lower band forms were lost to the supernatant (Figure 5C). These findings indicate that preventing N-glycosylation by 7NQ mutation did not affect the cell surface expression level of $PTPRA_{FL}$ nor its segregation into two distinct forms.

To better define the glycosylation pattern contributing to the size differences between the two forms, we constructed the full-length PTPRA with the TST or YFP tag to increase the protein yield. It was found that both $PTPRA_{FL\text{-}YFP}$ and $PTPRA_{FL\text{-}TST}$ were similarly separated into two gel bands, and those of $PTPRA_{FL\text{-}YFP}$ were subjected to in-gel digestion and analyzed by the same workflow (Figure 1) applied to $PTPRA_{WT}$. While not completely devoid of O-glycans, the smaller size form yielded far less O-glycosylated peptides, most of them carrying a single non-extended O-GalNAc (Figure 5D,E, with supporting EThcD MS^2 for peaks no. 1−3 in Figure S6). Taking the second tryptic N-glycopeptide for example and normalized to the peptides N-glycosylated with Man_5 (Figure 5F), more peptides carrying complex type N-glycans and/or O-glycans were identified in the upper bands while more of those with only high mannose type N-glycans were found in the lower bands. This was further corroborated by the MS profiles of the released glycans, which also showed that more complex type relative to high-mannose-type N-glycans were carried on the larger form of PTPRA in the upper band (Figure S7). Strikingly, a full range of core 1 and 2 O-glycans was detected only in the glycan pools released from the upper but not the lower band form. The overall glycosylation profile of the cell surface-expressed, larger form of full-length PTPRA, therefore, resembled that of the $PTPRA_{WT}$-Fc constructs (Figure 2B), whereas the immature smaller size form was severely deficient in O-glycosylation.

Constructing the Fully Glycosylated $PTPRA_{WT}$ Structure

By using size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS), we were able to tease out the molecular weight contributions of the glycans to the overall molecular weight estimates of PTPRAs based on the integrated UV absorption, refractive index, and static light scattering signals[37] (Fig S8). The results showed that the glycans contributed 65.7 ± 0.4 and 3.1 ± 0.7 kDa to $PTPRA_{WT}$ and $PTPRA_{ST2A}$, respectively. The notably smaller contribution of glycosylation to $PTPRA_{ST2A}$ is consistent with the ST → A mutations that prevented all N- and O-glycosylation of the ectodomain, leaving only the Fc module to be glycosylated. Based on the released glycan profiles and full occupancy of the 7 N-glycosylation sites, it can be further estimated that the O-glycans made up about 60% of the glycan size, to be distributed over at least 31 Ser/Thr sites on the PTPRA ectodomain and three additional sites on the linker identified as O-glycosylated in this work (Figure S9). These experimentally derived structural determinants then allowed a best-fit glycosylated model of $PTPRA_{WT}$ to be built for conformational studies (Figure S10).

Flexible and Dynamic Structure of $PTPRA_{WT}$ Ectodomain by SAXS and HS-AFM

The highly glycosylated disordered structure of $PTPRA_{WT}$ is not amenable to most conventional methods for protein structural studies other than SAXS, which is well-suited to characterize the ensemble structures of highly dynamic biomolecules in solution.[38] The synchrotron-based SAXS data of Fc-fused $PTPRA_{WT}$ ectodomain showed a radius of gyration (R_g) of 8.34 ± 0.03 nm based on Guinier approximation (inset of Figure 6A). Transformation of the raw SAXS profile into the pair-wise distance distribution, $P(r)$ showed one prominent peak centered around 5 nm accompanied by a very broad peak centered around 10 nm and extended to a maximum dimension, D_{max}, of 35.3 nm (Figure 6B, orange). In contrast, the SAXS profile of the variant of Fc-fused $PTPRA_{ST2A}$ ectodomain of which all Ser and Thr within the ectodomain were replaced by Ala translated into a much small R_g value, 5.28 ± 0.04 nm. Accordingly, the $P(r)$ profile of $PTPRA_{ST2A}$ showed a much narrower distribution with a D_{max} of 23.8 nm and a peak at 5 nm that is also present in that of $PTPRA_{WT}$ (Figure 6B). The corresponding Kratky plot (Figure S11A) also indicates a higher structural flexibility of the ectodmain of $PTPRA_{WT}$ relative to that of $PTPRA_{ST2A}$. The same SAXS analysis was carried out for Fc-fused $PTPRA_{7NQ}$ ectodomain, and the corresponding R_g and D_{max} values were comparable to that of Fc-fused $PTPRA_{WT}$ ectodomain. Importantly, the overall molecular envelope of Fc-fused $PTPRA_{7NQ}$ ectodomain was very similar to that of Fc-fused $PTPRA_{WT}$ ectodomain, indicating that the extended ectodomain structure is primarily attributed to the O-glycosylation while the seven N-glycans make limited contributions to the overall solution structure of the ectodomain (Figure S11B,C).

Ab initio modeling of the molecular envelope of both $PTPRA_{WT}$ and $PTPRA_{7NQ}$ showed a broad distribution of arm-like structures with a common globular domain structure that held the two flexible arms together (Figure 6C, left and S11D). The superposition of the bead-like models suggests that the two ectodomains are loosely tethered to the compact Fc to occupy a broad conformational space in solution. In contrast, the same procedure consistently yielded a cocoon-like compact structure for $PTPRA_{ST2A}$, indicating that the removal of both the N- and O-glycans would lead to the two ectodomains forming a compact structure with the Fc domain (Figure 6C, right). We next chose the most representative ab initio model of $PTPRA_{WT}$ out of 100 permissible solutions to best fit the molecular envelope accounting for the volume exclusion conferred by $PTPRA_{WT}$ (Figure 6D, left). An ensemble of structures taken from a trajectory of fully atomistic molecular dynamics simulation for each selected N- or O-glycan (as defined in Figure S10) was generated and attached to the respective amino acid side chain using the previously described protocol, known as GlycoSHIELD.[41] The resulting atomic model clearly shows that the N- and O-glycans constitute a major part of the extended arms of $PTPRA_{WT}$ (Figure 6D, right).

To monitor for dynamic conformational changes in fully hydrated states at a single molecule level on a sub-second timescale, HS-AFM was additionally performed at a comparable resolution to that of SAXS, with the advantage that the HS-AFM images directly reveal the molecular structures without the need for additional data processing and inference.[42] Several trajectories of $PTPRA_{WT}$ and $PTPRA_{ST2A}$ were successfully collected based on which several important attributes, including R_s (Figure 6F), $R_{g,d}$, and $R_{s,main}$ (Figure S11B,D) could be derived. It is reassuring that the SAXS-derived molecular envelops of both $PTPRA_{WT}$ and $PTPRA_{ST2A}$ could find equivalent snapshots from the HS-AFM trajectories with matching sizes and shapes (Figures 6E and S11E), thereby confirming the validity of the atomic model based on SAXS and MS.

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

DISCUSSION

It is apparent that the range of N- and O-glycans carried on the PTPRA-Fc differ significantly from the overall glycomic profile of the host cells. A more extensive trimming into Man_5 from Man_9, followed by efficient initiation of antennary GlcNAc and O-GalNAc, is consistent with its rather disordered conformation not imposing any accessibility issue for the required glycosyltransferases and glycosidases. However, most of the N-glycans were then not further galactosylated, and the O-glycans were less elaborated. This argues that further downstream processing was impeded, which can best be rationalized by the crowding effect imposed by the densely spaced O-GalNAc glycans. Such a PTPRA ectodomain-specific glycosylation pattern applies to both the secreted Fc-constructs and full-length transmembrane forms (Figures 2B and S7A). It is nonetheless possible that if and when under certain pathophysiological conditions, any of the O-GalNAc initiating enzymes (GALNTs) are downregulated, or downstream processing glycosyltransferases are significantly upregulated, a higher proportion of multi-sialylated, multi-antennary complex type N-glycans, or fully sialylated core 2 O-glycans would be formed. A recent work has shown that overexpression of the branching enzyme mgat5 could produce more six-arm branching on the N-glycans carried, leading to altered interactions with underlying ECM.[43] In another study,[44] it was found that a small set of targets involved in cell−cell adhesion could be specifically glycosylated by increased expression of GALNT6 in colon cancer. Among the 20 human GALNTs, GALNT1 and T2 are ubiquitously expressed, whereas other GALNTs have more restricted and specific expression profiles.[45] Since each of the GALNTs has subtle substrate specificity[46] with respect to the acceptor peptide sequence and requirement for prior addition of GalNAc at nearby site(s), a different permutation of expressed GALNTs at different levels would likely contribute to a different combination of sites being occupied and the extent of further elongation beyond the first GalNAc. HEK293 is known to additionally express GALNT7 and GALNT11 at significant levels and several other GALNTs at varying lower levels, with an overall pattern that differs from those of various carcinomas.[47] However, since a majority of sites are likely to be O-glycosylated by the ubiquitous GALNT1 and 2, the ectodomain of PTPRA would most likely remain heavily O-glycosylated irrespective of its origin.

It is intriguing that during the biosynthesis, the O-glycosylation process of full-length PTPRA should lead to two distinct size populations. We found that only the fully O-glycosylated glycoforms are expressed on the cell surface, a process that does not require N-glycosylation and can be recapitulated by the 7NQ form. This O-glycosylation-dependent size segregation became less apparent when the ectodomain was conjugated to Fc and secreted, suggesting that it was at least partly contributed by the ER/Golgi-membrane retention and/or cytoplasmic domains. The unexpected identification of the signal peptide of PTPRA from the digests of the lower but not the upper band supports the idea that the former comprised mostly immature forms[14] resulting from being impeded in their release from ER and transit through Golgi, hence affecting their chances of being fully O-glycosylated. Previous work on the Gp160 of HIV similarly showed that precursors retaining the signal peptide were associated with a prolonged folding process and low release or failed exit from the ER.[48,49] The disordered stretch of 130 odd amino acid residues, with or without the addition of N-glycans on seven sites, may conceivably get entangled in the ER and dragged along by its bulky cytoplasmic domains instead of timely cleavage of signal peptide and release into Golgi. Although the mature forms did carry more complex type N-glycans, it is more a consequence of translocation through Golgi rather than a requirement per se since the fully O-glycosylated but non-N-glycosylated 7NQ form could reach the plasma membrane for stable cell surface expression. On the other hand, it has been reported before that deficiency in O-GalNAc glycosylation of transmembrane MUC1 would enhance endocytic trafficking and degradation,[50] thus reducing its overall surface expression. This may also contribute to the smaller size PTPRA population containing more of the glycoforms with very low levels or no O-glycans and hence not stably expressed on the cell surface.

Irrespective of the underlying mechanism, our findings firmly establish that the surface-expressed, mature form of functional PTPRA is one that is densely N- and O-glycosylated. Interestingly, our structural studies conducted in a soluble environment to maintain the native protein structure and subsequent molecular modeling show that the glycosylation apparently shields and dynamically shapes the otherwise disordered ectodomain. Unlike the well-known "bottle brush" conformation adopted by a typical mucin domain,[51] we found that the $PTPRA_{WT}$ ectodomain is more flexible and dynamic rather than persistently linear. Compared to the tandem repeat sequence of MUC1 (HGVTSAPDTRPAPGSTAPPA) that contains 5 Ser/Thr and 5 Pro per 20 residues, there are 43 Ser/Thr and 12 Pro among the first 122 residues of the PTPRA ectodomain. The density of Ser/Thr is, therefore, slightly higher but with much less occurrence of Pro, which may partly contribute to the structural difference observed. The overall size, as defined by the radius of gyration (R_g) and the maximum dimension (D_{max}), as well as the degree of flexibility, are also significantly larger than the compact and more rigid $PTPRA_{ST2A}$. With two ectodomains tethered together by the Fc, the two arms of $PTPRA_{WT}$ would cycle through various fully open, close, and elongated conformations under HS-AFM (Figure 6E), while the most representative model from SAXS data favors a fully open Y shape (Figure 6D).

This overall structural view implies that the ectodomain of PTPRA is like a floppy tentacle protruding from the plasma membrane and may be flexible enough to keep it mostly dissociated from one another. The presence of heavily negatively charged glycans favors monomerization of PTPRA on the cell surface, in line with a previously proposed model that suggests the monomer rather than the dimer of PTPRA being the active form.[52,53] It has been reported that the catalytic D1 domain of PTPRA can be auto-inhibited by the plasma membrane-distal D2 domain, thus keeping PTPRA in low activity at the resting state,[54] not causing unwarranted tyrosine dephosphorylation of its substrates. The densely glycosylated ectodomain provides an excellent means for PTPRA to float and move freely on the cell surface, where receptor and non-receptor tyrosine kinases are located nearby. As soon as a stimulus activates receptor tyrosine kinases, the Tyr789 residue located in the C-terminal tail of PTPRA is rapidly phosphorylated, leading to the recruitment of an intracellular adaptor with an SH2 domain. Complex formation between this SH2 domain-containing adaptor and the Tyr789 phosphorylated PTPRA would relieve the auto-inhibitory conformation, allowing the open conformation of D1 domain to dephosphorylate its substrates efficiently. Our findings suggest that the heavily glycosylated ectodomain may

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

augment or facilitate this surveillance and regulatory functions of PTPRA by fastening it to the plasma membrane and thereby keeping it in close proximity with myriad receptors.

MATERIALS AND METHODS

All materials and methods used including cloning, expression and purification of various PTPRA constructs, their digestion into glycopeptides, the release of glycans, sialidase treatment, MALDI and LC−MS/MS analyses, biophysical characterization, and SAXS, HS-AFM, and molecular modeling of intact proteins are described in full detail in the Supporting Information.

MS/MS Data Analysis

The raw file was processed by Byonic v4.0.12. The parameters were set as below: specific cleavage at Arg/Lys for trypsin, Asp/Glu for GluC, and Ser/Thr for O-protease; allowing up to two missed cleavages for trypsin and/or GluC digestion and five for O-protease; peptide ion tolerance at 5 ppm and fragment ion tolerance at 10 ppm; fixed cysteine carbamidomethylation at Cys, variable deamidation at Asn or Gln, and variable oxidation at Met. For samples treated with PNGase F, deamidation was set as a variable modification at Asn. The MS/MS data sets were searched against the default glycan database(s) ("N-glycan 182 human no multiple fucose" for N-glycan and "O-glycan 6 most common" for O-glycan).

The non-filtered Byonic search results at protein FDR 1% were output as Excel files for further manual data mining. First, the HCD-MS^2 peptide spectrum matches (PSMs) were filtered by score > 200 and PEP2D < 0.001, as the cutoff criteria for them to be considered as positive HCD-MS^2 PSMs. These were paired with their corresponding EThcD-MS^2 PSMs triggered within five scan numbers for the same precursor. The rest of EThcD-MS^2 PSMs without accompanying positive HCD-MS2 PSMs were then similarly filtered by score > 200 and PEP2D < 0.001. All positive MS^2 PSMs were manually checked, with at least three matching fragment ions to be considered a true identification, and the relevant site-determining EThcD-MS^2 ions to support O-glycosylation site localization. The glycopeptides identified by HCD-MS^2 and EThcD-MS^2 PSMs within five scan numbers were defined as the same glycopeptides if sharing the same precursor mass and peptide sequence. If the peptide sequence was different, the EThcD-MS^2 PSMs would be considered as identifying independent glycopeptides, which would therefore not be paired with the HCD-MS^2 PSM and need to pass the cutoff criteria applied. Any different glycosyl assignment for the same peptide sequence was defined as the same glycopeptides with different glycoforms.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.3c00124.

Detailed descriptions of materials and methods and Figure S1−S12 (PDF)

Table S1. Glycopeptide analysis of $PTPRA_{WT}$ (XLSX)

AUTHOR INFORMATION

Corresponding Author

Kay-Hooi Khoo − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan; Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Okazaki 444-8787, Japan;* orcid.org/0000-0003-2906-406X; Email: kkhoo@gate.sinica.edu.tw

Authors

Yu-Chun Chien − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan;* orcid.org/0000-0002-2485-2932

Yong-Sheng Wang − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan;* orcid.org/0000-0002-4868-4507

Deepa Sridharan − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan;* orcid.org/0000-0003-3517-8389

Chu-Wei Kuo − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan;* orcid.org/0000-0002-4388-5748

Chih-Ta Chien − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan*

Takayuki Uchihashi − *Department of Physics, Nagoya University, Nagoya 464-8602, Japan; Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Okazaki 444-8787, Japan;* orcid.org/0000-0002-0263-5312

Koichi Kato − *Exploratory Research Center on Life and Living Systems and Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki 444-8787, Japan; Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya 467-8603, Japan;* orcid.org/0000-0001-7187-9612

Takashi Angata − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan;* orcid.org/0000-0002-4171-702X

Tzu-Ching Meng − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan;* orcid.org/0000-0001-7693-2021

Shang-Te Danny Hsu − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan; International Institute for Sustainability with Knotted Chiral Meta Matter, Hiroshima University, Higashihiroshima 739-8527, Japan;* orcid.org/0000-0002-7231-0185

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.3c00124

Author Contributions

CRediT: **Yu-Chun Chien** data curation, formal analysis, investigation, methodology, visualization, writing-original draft, writing-review & editing; **Yong-Sheng Wang** data curation, formal analysis, investigation, methodology, validation, visualization; **Deepa Sridharan** data curation, formal analysis, investigation, methodology, validation, visualization; **Chu-Wei Kuo** data curation, formal analysis, investigation, methodology; **Chih-Ta Chien** investigation, methodology; **Takayuki Uchihashi** data curation, formal analysis, investigation, validation, visualization; **Koichi Kato** funding acquisition, resources; **Takashi Angata** funding acquisition, project administration, resources, supervision; **Tzu-Ching Meng** funding acquisition, project administration, resources, supervision, writing-review & editing; **Shang-Te Danny Hsu** conceptualization, funding acquisition, project administration, resources, supervision, writing-review & editing; **Kay-Hooi Khoo** conceptualization, funding acquisition, project administration, resources, super-

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875

vision, visualization, writing-original draft, writing-review & editing.

Notes

The authors declare no competing financial interest.

The .raw files and annotated spectra by Byonic (.byrslt and .byspec2 byonic files) have been deposited in MassIVE (MSV000091364). SAXS data of $PTPRA_{WT}$, $PTPRA_{7NQ}$, and $PTPRA_{ST2A}$ have been deposited in SASBDB (https://www.sasbdb.org).

■ ACKNOWLEDGMENTS

This work was supported by Academia Sinica (AS) intramural grant, an AS Grand Challenge seed grant (AS-GC-108-05 to T.A., T.C.M., and K.H.K. and AS-IR-111-01 to K.H.K.), and the Joint Research of the Exploratory Research Center on Life and Living Systems (ExCELLS) (ExCELLS program no. 20-403 to S.T.D.H., no. 22EXC304, 21-302, and 20-321 to K.H.K., and no. 18-101 to T.U.). We thank the Academia Sinica Common Mass Spectrometry Facilities for Proteomics and Protein Modification Analysis (AS-CFII-108-107) and the Academia Sinica Biophysics Core Facility (AS-CFII108-111) for data collection, both of which are funded by the Academia Sinica Core Facility and Innovative Instrument Project grants. We thank the mammalian cell culture facility of the Institute of Biological Chemistry (IBC), AS, for supporting the protein production and Dr. Meng-Chiao Ho (IBC, AS) for the PTPRA (*Homo sapiens*) construct in pDONR221 (HsCD00043172, DNASU). We also thank Mateusz Sikora at the Max Planck Institute for Biophysics who shared the GlycoSHIELD trajectories of the N- and O-glycans used in the SAXS-based modeling, which was implemented in-house with the support from Yu-Xi Tsai and Hao-Ting Chang in S.T.D.H.'s laboratory.

■ REFERENCES

(1) Tonks, N. K. Protein tyrosine phosphatases–from housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **2013**, *280*, 346−378.

(2) Su, J.; Muranjan, M.; Sap, J. Receptor protein tyrosine phosphatase alpha activates Src-family kinases and controls integrin-mediated responses in fibroblasts. *Curr. Biol.* **1999**, *9*, 505−511.

(3) Ponniah, S.; Wang, D. Z.; Lim, K. L.; Pallen, C. J. Targeted disruption of the tyrosine phosphatase PTPalpha leads to constitutive downregulation of the kinases Src and Fyn. *Curr. Biol.* **1999**, *9*, 535−538.

(4) Zheng, X. M.; Resnick, R. J.; Shalloway, D. A phosphotyrosine displacement mechanism for activation of Src by PTPalpha. *EMBO J.* **2000**, *19*, 964−978.

(5) Pallen, C. J. Protein tyrosine phosphatase alpha (PTPalpha): a Src family kinase activator and mediator of multiple biological effects. *Curr. Top. Med. Chem.* **2003**, *3*, 821−835.

(6) Zeng, L.; Si, X.; Yu, W. P.; Le, H. T.; Ng, K. P.; Teng, R. M.; Ryan, K.; Wang, D. Z.; Ponniah, S.; Pallen, C. J. PTP alpha regulates integrin-stimulated FAK autophosphorylation and cytoskeletal rearrangement in cell spreading and migration. *J. Cell Biol.* **2003**, *160*, 137−146.

(7) Kaplan, K. B.; Swedlow, J. R.; Morgan, D. O.; Varmus, H. E. c-Src enhances the spreading of src-/- fibroblasts on fibronectin by a kinase-independent mechanism. *Genes Dev.* **1995**, *9*, 1505−1517.

(8) von Wichert, G.; Jiang, G.; Kostic, A.; De Vos, K.; Sap, J.; Sheetz, M. P. RPTP-alpha acts as a transducer of mechanical force on alphav/beta3-integrin-cytoskeleton linkages. *J. Cell Biol.* **2003**, *161*, 143−153.

(9) Aschner, Y.; Khalifah, A. P.; Briones, N.; Yamashita, C.; Dolgonos, L.; Young, S. K.; Campbell, M. N.; Riches, D. W.; Redente, E. F.; Janssen, W. J.; et al. Protein tyrosine phosphatase alpha mediates profibrotic signaling in lung fibroblasts through TGF-beta responsiveness. *Am. J. Pathol.* **2014**, *184*, 1489−1502.

(10) Aschner, Y.; Nelson, M.; Brenner, M.; Roybal, H.; Beke, K.; Meador, C.; Foster, D.; Correll, K. A.; Reynolds, P. R.; Anderson, K.; et al. Protein tyrosine phosphatase-alpha amplifies transforming growth factor-beta-dependent profibrotic signaling in lung fibroblasts. *Am. J. Physiol.: Lung Cell. Mol. Physiol.* **2020**, *319*, L294−L311.

(11) AlphaFold Protein Structure Database. 2022, https://alphafold.ebi.ac.uk/entry/P18433 (accessed March 1, 2023).

(12) Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. Protein tyrosine phosphatases in the human genome. *Cell* **2004**, *117*, 699−711.

(13) Stanford, S. M.; Svensson, M. N.; Sacchetti, C.; Pilo, C. A.; Wu, D. J.; Kiosses, W. B.; Hellvard, A.; Bergum, B.; Muench, G. R.; Elly, C.; et al. Receptor Protein Tyrosine Phosphatase alpha-Mediated Enhancement of Rheumatoid Synovial Fibroblast Signaling and Promotion of Arthritis in Mice. *Arthritis Rheumatol.* **2016**, *68*, 359−369.

(14) Daum, G.; Regenass, S.; Sap, J.; Schlessinger, J.; Fischer, E. H. Multiple forms of the human tyrosine phosphatase RPTP alpha. Isozymes and differences in glycosylation. *J. Biol. Chem.* **1994**, *269*, 10524−10528.

(15) Glygen. 2019, https://www.glygen.org/protein/P18433 (accessed March 1, 2023).

(16) O-GalNAc Human Glycoproteome Database. 2018, http://www.glycoproteomics.somee.com/Data/GalNAc_SC.aspx (accessed March 1, 2023).

(17) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Schjoldager, K. T.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; Marcos-Silva, L.; et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **2013**, *32*, 1478−1488.

(18) Malaker, S. A.; Riley, N. M.; Shon, D. J.; Pedram, K.; Krishnan, V.; Dorigo, O.; Bertozzi, C. R. Revealing the human mucinome. *Nat. Commun.* **2022**, *13*, 3542.

(19) NetOGlyc—4.0. 2013, https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0 (accessed March 1, 2023).

(20) Polasky, D. A.; Nesvizhskii, A. I. Recent advances in computational algorithms and software for large-scale glycoproteomics. *Curr. Opin. Chem. Biol.* **2023**, *72*, 102238.

(21) Darula, Z.; Medzihradszky, K. F. Analysis of Mammalian O-Glycopeptides-We Have Made a Good Start, but There is a Long Way to Go. *Mol. Cell. Proteomics* **2018**, *17*, 2−17.

(22) Khoo, K. H. Advances toward mapping the full extent of protein site-specific O-GalNAc glycosylation that better reflects underlying glycomic complexity. *Curr. Opin. Struct. Biol.* **2019**, *56*, 146−154.

(23) Caval, T.; de Haan, N.; Konstantinidi, A.; Vakhrushev, S. Y. Quantitative characterization of O-GalNAc glycosylation. *Curr. Opin. Struct. Biol.* **2021**, *68*, 135−141.

(24) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int. J. Proteomics* **2012**, *2012*, 560391−560397.

(25) Yin, X.; Bern, M.; Xing, Q.; Ho, J.; Viner, R.; Mayr, M. Glycoproteomic analysis of the secretome of human endothelial cells. *Mol. Cell. Proteomics* **2013**, *12*, 956−978.

(26) Yang, W.; Ao, M.; Hu, Y.; Li, Q. K.; Zhang, H. Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). *Mol. Syst. Biol.* **2018**, *14*, No. e8486.

(27) Riley, N. M.; Malaker, S. A.; Bertozzi, C. R. Electron-Based Dissociation Is Needed for O-Glycopeptides Derived from OpeRATOR Proteolysis. *Anal. Chem.* **2020**, *92*, 14878−14884.

(28) Yang, S.; Onigman, P.; Wu, W. W.; Sjogren, J.; Nyhlen, H.; Shen, R. F.; Cipollo, J. Deciphering Protein O-Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment. *Anal. Chem.* **2018**, *90*, 8261−8269.

(29) Trastoy, B.; Naegeli, A.; Anso, I.; Sjogren, J.; Guerin, M. E. Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from Akkermansia muciniphila. *Nat. Commun.* **2020**, *11*, 4844.

(30) Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lutteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; et al. Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **2019**, *29*, 620−624.
(31) Huhn, C.; Selman, M. H.; Ruhaak, L. R.; Deelder, A. M.; Wuhrer, M. IgG glycosylation analysis. *Proteomics* **2009**, *9*, 882−913.
(32) Plomp, R.; Bondt, A.; de Haan, N.; Rombouts, Y.; Wuhrer, M. Recent Advances in Clinical Glycoproteomics of Immunoglobulins (Igs). *Mol. Cell. Proteomics* **2016**, *15*, 2217−2228.
(33) Pongracz, T.; Vidarsson, G.; Wuhrer, M. Antibody glycosylation in COVID-19. *Glycoconjugate J.* **2022**, *39*, 335−344.
(34) Tian, W.; Li, D.; Zhang, N.; Bai, G.; Yuan, K.; Xiao, H.; Gao, F.; Chen, Y.; Wong, C. C. L.; Gao, G. F. O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. *Cell Res.* **2021**, *31*, 1123−1125.
(35) Ardini, E.; Agresti, R.; Tagliabue, E.; Greco, M.; Aiello, P.; Yang, L. T.; Menard, S.; Sap, J. Expression of protein tyrosine phosphatase alpha (RPTPalpha) in human breast cancer correlates with low tumor grade, and inhibits tumor cell growth in vitro and in vivo. *Oncogene* **2000**, *19*, 4979−4987.
(36) Gil-Henn, H.; Volohonsky, G.; Elson, A. Regulation of protein-tyrosine phosphatases alpha and epsilon by calpain-mediated proteolytic cleavage. *J. Biol. Chem.* **2001**, *276*, 31772−31779.
(37) Yang, T. J.; Chang, Y. C.; Ko, T. P.; Draczkowski, P.; Chien, Y. C.; Chang, Y. C.; Wu, K. P.; Khoo, K. H.; Chang, H. W.; Hsu, S. T. D. Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 1438−1446.
(38) Kikhney, A. G.; Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570−2577.
(39) Franke, D.; Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009**, *42*, 342−346.
(40) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70−82.
(41) Gecht, M.; von Bülow, S.; Penet, C.; Hummer, G.; Hanus, C.; Sikora, M. GlycoSHIELD: a versatile pipeline to assess glycan impact on protein structures. 2022, bioRxiv, 2021.2008.2004.455134. bioRxiv:10.1101/2021.08.04.455134. February 15 (accessed March 1, 2023).
(42) Ando, T.; Uchihashi, T.; Kodera, N. High-speed AFM and applications to biomolecular systems. *Annu. Rev. Biophys.* **2013**, *42*, 393−414.
(43) Xiao, J.; Gao, Y.; Yang, F.; Wang, C.; Xu, Y.; Chang, R.; Zha, X.; Wang, L. beta1,6 GlcNAc branches-modified protein tyrosine phosphatase alpha enhances its stability and promotes focal adhesion formation in MCF-7 cells. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 1455−1461.
(44) Lavrsen, K.; Dabelsteen, S.; Vakhrushev, S. Y.; Levann, A. M. R.; Haue, A. D.; Dylander, A.; Mandel, U.; Hansen, L.; Frodin, M.; Bennett, E. P.; et al. De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. *J. Biol. Chem.* **2018**, *293*, 1298−1314.
(45) Kato, K.; Hansen, L.; Clausen, H. Polypeptide N-acetylgalactosaminyltransferase-Associated Phenotypes in Mammals. *Molecules* **2021**, *26*, 5504.
(46) Wandall, H. H.; Irazoqui, F.; Tarp, M. A.; Bennett, E. P.; Mandel, U.; Takeuchi, H.; Kato, K.; Irimura, T.; Suryanarayanan, G.; Hollingsworth, M. A.; et al. The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology* **2007**, *17*, 374−387.
(47) Narimatsu, Y.; Joshi, H. J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y. H.; Schjoldager, K. T.; Steentoft, C.; et al. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cell* **2019**, *75*, 394−407.e5.
(48) Li, Y.; Bergeron, J. J.; Luo, L.; Ou, W. J.; Thomas, D. Y.; Kang, C. Y. Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9606−9611.
(49) Land, A.; Zonneveld, D.; Braakman, I. Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J.* **2003**, *17*, 1058−1067.
(50) Altschuler, Y.; Kinlough, C. L.; Poland, P. A.; Bruns, J. B.; Apodaca, G.; Weisz, O. A.; Hughey, R. P. Clathrin-mediated endocytosis of MUC1 is modulated by its glycosylation state. *Mol. Biol. Cell* **2000**, *11*, 819−831.
(51) Brockhausen, I.; Wandall, H. H.; Hagen, K. G. T.; Stanley, P. O-GalNAc Glycans. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., et al., Eds.; Cold Spring Harbor Laboratory Press, 2022; pp 117−128.
(52) Jiang, G.; den Hertog, J.; Su, J.; Noel, J.; Sap, J.; Hunter, T. Dimerization inhibits the activity of receptor-like protein-tyrosine phosphatase-alpha. *Nature* **1999**, *401*, 606−610.
(53) Jiang, G.; den Hertog, J.; Hunter, T. Receptor-like protein tyrosine phosphatase alpha homodimerizes on the cell surface. *Mol. Cell. Biol.* **2000**, *20*, 5917−5929.
(54) Wen, Y.; Yang, S.; Wakabayashi, K.; Svensson, M. N. D.; Stanford, S. M.; Santelli, E.; Bottini, N. RPTPalpha phosphatase activity is allosterically regulated by the membrane-distal catalytic domain. *J. Biol. Chem.* **2020**, *295*, 4923−4936.

https://doi.org/10.1021/jacsau.3c00124
JACS Au 2023, 3, 1864−1875