Mechanism of pH-dependent N-Acetylgalactosamine Binding by a Functional Mimic of the Hepatocyte Asialoglycoprotein Receptor*

Hadar Feinberg‡, Dawn Torgersen§, Kurt Drickamer§, and William I. Weis‡‡

From the ‡Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305 and the §Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom

Received for publication, June 25, 2000, and in revised form, August 4, 2000
Published, JBC Papers in Press, August 7, 2000, DOI 10.1074/jbc.M005557200

Efficient release of ligands from the Ca\(^{2+}\)-dependent carbohydrate-recognition domain (CRD) of the hepatic asialoglycoprotein receptor at endosomal pH requires a small set of conserved amino acids that includes a critical histidine residue. When these residues are incorporated at corresponding positions in an homologous galactose-binding derivative of serum mannos-binding protein, the pH dependence of ligand binding becomes more like that of the receptor. The modified CRD displays 40-fold preferential binding to N-acetylgalactosamine compared with galactose, making it a good functional mimic of the asialoglycoprotein receptor. In the crystal structure of the modified CRD bound to N-acetylgalactosamine, the histidine (His\(^{202}\)) contacts the 2-acetamido methyl group and also participates in a hydrogen bond with the sugar amide group. These interactions appear to produce the preference for N-acetylgalactosamine over galactose and are also likely to influence the pH\(_m\) of His\(^{202}\). Protonation of His\(^{202}\) would disrupt its interaction with an asparagine that serves as a ligand for Ca\(^{2+}\) and sugar. The structure of the modified CRD without sugar displays several different conformations that may represent structures of intermediates in the release of Ca\(^{2+}\) and sugar ligands caused by protonation of His\(^{202}\).

The hepatocyte asialoglycoprotein receptor mediates uptake of glycoproteins bearing oligosaccharides terminating in galactose or N-acetylgalactosamine (GalNAc) residues. The receptor contains two homologous polypeptide chains: the major form is designated rat hepatic lectin 1 (RHL-1), and the minor form is designated RHL-2/3 (2). Each of these component polypeptides consists of a short N-terminal cytoplasmic domain, a single transmembrane \(\alpha\)-helix, an \(\alpha\)-helical stalk, and a C-terminal carbohydrate-recognition domain (CRD). The stalk mediates trimerization of the receptor, probably by formation of a parallel coiled-coil of \(\alpha\)-helices (3). The CRD is a member of the C-type (Ca\(^{2+}\)-dependent) lectin superfamily (4). Structural analysis of the homologous C-type CRDs of rat mannos-binding proteins (MBP-A and MBP-C) (5–7) has shown that two Ca\(^{2+}\) stabilize an extensive series of loops at one end of the domain. Sugar ligands interact directly with Ca\(^{2+}\) at site 2, which is designated the principal Ca\(^{2+}\) site. Vicinal hydroxyl groups on the pyranose ring of the sugar form both direct coordination bonds with this Ca\(^{2+}\) and hydrogen bonds with amino acid side chains that also serve as Ca\(^{2+}\) coordination ligands.

Structure-based sequence alignments between RHL-1 and MBP-A and site-directed mutagenesis have led to the identification of several residues that provide selective affinity for galactose in C-type lectins (8, 9). Substitution of Glu and Asn for Gln and Asp at positions 185 and 187 of MBP-A and also the presence of a tryptophan residue at position 189 confer upon MBP-A galactose affinity that is comparable with that of RHL-1 (9). Discrimination against mannose is provided by a glycine-rich sequence immediately following Trp\(^{189}\), which forms a compact, well ordered loop that holds Trp\(^{189}\) in position for interaction with the B face of galactose while sterically excluding mannose (10). This galactose-binding mutant of MBP-A is designated QPDWG. RHL-1 displays a 60-fold preference for GalNAc over Gal, whereas QPDWG does not discriminate between these two sugars. Introduction of a histidine residue found in RHL-1 into the equivalent position, 202, of QPDWG confers approximately 9-fold selectivity for GalNAc over Gal, thereby providing a partial mimic of RHL-1. The crystal structure of this mutant complexed with GalNAc (11) revealed that the distal edge of the His\(^{202}\) imidazole ring directly contacts the methyl group of the acetamido moiety.

RHL-1 and many other C-type lectins exploit the Ca\(^{2+}\)-dependent sugar binding to release their endocytosed ligands in endosomes. Acidification of endosomes reduces the affinity for Ca\(^{2+}\) and hence sugar. The ligand and receptor are sorted from one another, with the receptor returning to the cell surface for another round of ligand binding, whereas the ligand is usually delivered to lysosomes for degradation. For different C-type lectins, the pH at which Ca\(^{2+}\) affinity is sufficiently weakened to result in loss of sugar ligands is tuned so that separation of ligand and receptor occurs in an appropriate endosome. For RHL-1 in isolation, the pH at which half-maximal sugar binding occurs (pH\(_B\)) at 1 mM Ca\(^{2+}\) is 7.1 (at 5 mM Ca\(^{2+}\) the pH dependence of ligand binding becomes more like that of the receptor. The modified CRD displays 40-fold preferential binding to N-acetylgalactosamine compared with galactose, making it a good functional mimic of the asialoglycoprotein receptor. In the crystal structure of the modified CRD bound to N-acetylgalactosamine, the histidine (His\(^{202}\)) contacts the 2-acetamido methyl group and also participates in a hydrogen bond with the sugar amide group. These interactions appear to produce the preference for N-acetylgalactosamine over galactose and are also likely to influence the pH\(_m\) of His\(^{202}\). Protonation of His\(^{202}\) would disrupt its interaction with an asparagine that serves as a ligand for Ca\(^{2+}\) and sugar. The structure of the modified CRD without sugar displays several different conformations that may represent structures of intermediates in the release of Ca\(^{2+}\) and sugar ligands caused by protonation of His\(^{202}\).

The hepatocyte asialoglycoprotein receptor mediates uptake of glycoproteins bearing oligosaccharides terminating in galactose or N-acetylgalactosamine (GalNAc) residues. The receptor contains two homologous polypeptide chains: the major form is designated rat hepatic lectin 1 (RHL-1), and the minor form is designated RHL-2/3 (2). Each of these component polypeptides consists of a short N-terminal cytoplasmic domain, a single transmembrane \(\alpha\)-helix, an \(\alpha\)-helical stalk, and a C-terminal carbohydrate-recognition domain (CRD). The stalk mediates trimerization of the receptor, probably by formation of a parallel coiled-coil of \(\alpha\)-helices (3). The CRD is a member of the C-type (Ca\(^{2+}\)-dependent) lectin superfamily (4). Structural analysis of the homologous C-type CRDs of rat mannos-binding proteins (MBP-A and MBP-C) (5–7) has shown that two Ca\(^{2+}\) stabilize an extensive series of loops at one end of the domain. Sugar ligands interact directly with Ca\(^{2+}\) at site 2, which is designated the principal Ca\(^{2+}\) site. Vicinal hydroxyl groups on the pyranose ring of the sugar form both direct coordination bonds with this Ca\(^{2+}\) and hydrogen bonds with amino acid side chains that also serve as Ca\(^{2+}\) coordination ligands.

Structure-based sequence alignments between RHL-1 and MBP-A and site-directed mutagenesis have led to the identification of several residues that provide selective affinity for galactose in C-type lectins (8, 9). Substitution of Glu and Asn for Gln and Asp at positions 185 and 187 of MBP-A and also the presence of a tryptophan residue at position 189 confer upon MBP-A galactose affinity that is comparable with that of RHL-1 (9). Discrimination against mannose is provided by a glycine-rich sequence immediately following Trp\(^{189}\), which forms a compact, well ordered loop that holds Trp\(^{189}\) in position for interaction with the B face of galactose while sterically excluding mannose (10). This galactose-binding mutant of MBP-A is designated QPDWG. RHL-1 displays a 60-fold preference for GalNAc over Gal, whereas QPDWG does not discriminate between these two sugars. Introduction of a histidine residue found in RHL-1 into the equivalent position, 202, of QPDWG confers approximately 9-fold selectivity for GalNAc over Gal, thereby providing a partial mimic of RHL-1. The crystal structure of this mutant complexed with GalNAc (11) revealed that the distal edge of the His\(^{202}\) imidazole ring directly contacts the methyl group of the acetamido moiety.

RHL-1 and many other C-type lectins exploit the Ca\(^{2+}\)-dependent sugar binding to release their endocytosed ligands in endosomes. Acidification of endosomes reduces the affinity for Ca\(^{2+}\) and hence sugar. The ligand and receptor are sorted from one another, with the receptor returning to the cell surface for another round of ligand binding, whereas the ligand is usually delivered to lysosomes for degradation. For different C-type lectins, the pH at which Ca\(^{2+}\) affinity is sufficiently weakened to result in loss of sugar ligands is tuned so that separation of ligand and receptor occurs in an appropriate endosome. For RHL-1 in isolation, the pH at which half-maximal sugar binding occurs (pH\(_B\)) at 1 mM Ca\(^{2+}\) is 7.1 (at 5 mM Ca\(^{2+}\)}

A. K. Powell and K. Drickamer, unpublished observations.
Ca$^{2+}$ pH$_{eq}$ in 6.3) (12), which corresponds to the separation of this receptor from its sugar ligands in mildly acidic early endosomes. In contrast, MBP-A, which is not an endocytic receptor, shifts to a state with weaker Ca$^{2+}$ affinity only when the pH is reduced to below 5.0. These results indicate that different pH sensitivity for Ca$^{2+}$ binding can be encoded in the common C-type CRD structural framework.

Mutagenesis of the RHL-1 CRD has led to the identification of several residues that are necessary for release of ligands at endosomal pH. Key residues include His$_{256}$, Asp$_{266}$, and Arg$_{270}$. Introduction of these RHL-1 residues at corresponding positions in MBP-A results in shifts in the pH dependence of ligand binding. Introduction of His$_{202}$ to generate the GalNAc-selective mutant of QPDWG, QPDWG-H, changes pH$_{eq}$ from 5.0 to 5.7 at 1 mM Ca$^{2+}$ and from 3.5 to 5.3 at 4 mM Ca$^{2+}$ (12). Inspection of the structure of QPDWG-H (11) reveals that His$_{202}$ acts as a hydrogen bond acceptor from the amide group of Asn$_{210}$, which is a Ca$^{2+}$ and sugar ligand (see Fig. 1a).

Introduction of aspartic acid at position 212 of QPDWG-H to generate the mutant QPDWG-HD, results in a 8-fold reduction of Ca$^{2+}$ and sugar ligand (see Fig. 1a), although in the absence of His$_{202}$ this substitution has no effect (12). These results suggest that Asp$_{212}$ alters the pK$_a$ of His$_{202}$. Replacement of a loop at positions 216–218 in QPDWG-HD with the corresponding RHL-1 sequence Arg-Pro-Tyr produces the mutant QPDWG-HDPRPY, which shows a further 5-fold reduction Ca$^{2+}$ affinity and a raised pH$_{eq}$ (6.4 at 4 mM Ca$^{2+}$) (12). Although this mutant binds Ca$^{2+}$ approximately 15 times more weakly than RHL-1 at pH 7.8, its properties make it the best available mimic of RHL-1.

Biochemical and structural data on QPDWG-HDPRPY presented here provide a structural basis for the strong GalNAc selectivity of RHL-1. A high resolution crystal structure of native QPDWG-HDPRPY reveals conformational differences that likely represent intermediates leading to the loss of Ca$^{2+}$ at endosomal pH. An intricate network of interactions appears to be responsible for adjusting the pH at which Ca$^{2+}$ is released from the structure.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—Following expression in the previously described bacterial system (12), the arginine residue that was introduced into QPDWG-HDPRPY proved to be accessible to clostridial digestion. Thus, it was not possible to use clostridipain to remove the N-terminal extension sequence following the approach used in previous crystallization studies (13). To obviate this problem, the expression vector was modified by replacing appropriate restriction fragments with synthetic oligonucleotides to encode a cleavage site for factor Xa. In the final vector, the sequence encoding the bacterial ompA signal sequence (14) is followed by the sequence GCTGAAATTTTTCCAGCTGGATAAAAATCGAGGGTAGA, which is fused to codon 73 of the MBP-A cDNA (15). The N-terminal sequence of the secreted protein, which was confirmed by automated Edman degradation, is Ala-Glu-Leu-Ile-Pro-Ser-Leu-Asp-Lys-Ile-Glu-Val-Lys-Leu-Ala. Following cleavage with Factor Xa (New England Biolabs), the resulting protein thus has exactly the same N terminus as the trimeric fragments of wild type MBP-A and other mutants previously studied by crystallography (10, 11, 13, 16).

Cultures (6 l) of Escherichia coli strain JA221 containing the expression plasmid were grown in Luria-Bertani broth containing 50 μg/ml ampicillin at 25 °C to an A$_{600}$ of 0.8. Following induction with 40 μM isopropyl-β-D-thiogalactoside and addition of CaCl$_2$ to a final concentration of 150 μM, the culture was incubated at 18 h at 25 °C. Cells were harvested by centrifugation for 15 min at 4000 × g, suspended in 500 ml of loading buffer (1.25 mM NaCl, 25 mM Tris-HCl, pH 7.8, and 2.5 mM CaCl$_2$) and sonicated with the 4-mm probe of a Branson model 250 sonifier for a total of 10 min in 2-min bursts interspersed with cooling on ice. Debris was removed by centrifugation for 15 min at 27,000 × g and by further centrifugation for 60 min at 100,000 × g.

**Crystallization and Structure Determination**—Lyophilized protein was dissolved in 10 mM NaCl, 10 mM CaCl$_2$ and neutralized with 50 mM NaOH to give a final pH of ~7.5. The protein was diluted to a final concentration of 15–25 mg/ml with 10 mM NaCl, 10 mM CaCl$_2$. Crystals of QPDWG-HDPRPY were grown at 20 °C by microbatch hanging drop vapor diffusion, by mixing 1–2 μl of protein solution with 1–2 μl of reservoir solution containing 13.5% polyethylene glycol 8000, 100 mM Tris-HCl pH 8.0, 10 mM NaCl, and 20 mM CaCl$_2$ (solution A). Crystals appeared within 3–4 days and grew to full size in 7–14 days. Prior to data collection, the crystals were adapted in a stepwise fashion to solution A plus 0, 5, 7.5, 10, 15, and 20 mg/ml 2-methyl-2-pentanediol and flash cooled by plunging into liquid nitrogen. The complex with GalNAc was prepared by including 200 mM GalNAc (Sigma) in the soaking solutions.

Diffraction data from native and GalNAc-soaked crystals were measured to maximum Bragg spacings of 1.9 Å on a MAR345 Image plate detector at the Stanford Synchrotron Radiation Laboratory beam line 9-1. A total of 168 images were measured from a single native crystal (size ~ 0.2 × 0.2 × 0.2 mm$^3$), with oscillation ranges of 1.2, 1.0, or 0.6° depending on the portion of reciprocal space being measured. Exposure times were 10 s for the 1.2 and 1.0° sweeps and 5 s for the 0.6° sweep. For the GalNAc-soaked crystal (size 250 × 250 × 60 μm$^3$), 96 1.2° oscillation sweeps were measured at 60 s/image. The same images were then remeasured with a 10 s/image exposure time to measure reflections that were saturated in the high dose pass. Intensities were integrated, scaled and merged with DENZO and SCALEPACK (18) (Table I).

Crystals of QPDWG-HDPRPY are nearly isomorphous with those of the QPDWG-H mutant and permitted structure solution by rigid body refinement of the QPDWG-H model (11) against the QPDWG-HDPRPY data set. The temperature factors from the QPDWG-H model were retained, and all water molecules were omitted from the model. All refinement procedures employed the maximum likelihood amplitude target in CNS (19). An overall anisotropic temperature factor and a bulk solvent correction were applied throughout. The trimer was refined as a single rigid body against data from 30–2.5 Å (R$_{free}$ = 0.408), and then the protomers were refined as individual rigid bodies extending the resolution range to 30–2.2 Å (R$_{free}$ = 0.393). After several rounds of positional and isotropic temperature factor refinement, the side chains of the mutated residues at positions 212 and 216–218 could be seen clearly in the maps and were added to the model using the program O (20). The electron density maps clearly indicated that the glycine-rich loop comprising residues 188–197 in protomer B has a different conformation from that of the search model. These residues were removed from the model, and after several rounds of

| TABLE I | Data collection statistics for QPDWG-HDPRPY |
|----------|------------------------------------------|
| **The crystals belong to the space group C2:** |
| **Native** | **With GalNAc** |
| Unit cell parameters* | | |
| a (Å) | 80.34 | 81.49 |
| b (Å) | 84.83 | 84.55 |
| c (Å) | 97.14 | 98.54 |
| β (°) | 107.88 | 105.43 |
| Resolution range (Å) | 30.0–1.95 | 30.0–1.95 |
| Measured reflections | 120,159 | 173,378 |
| Unique reflections | 42,884 | 43,653 |
| Completeness (%)b | 94.7 (93.0) | 92.7 (92.7) |
| R$_{sym}$ (%)f | 5.3 (15.7) | 5.7 (32.8) |

* From postrefinement in SCALEPACK.

Numbers in parentheses represent the values in the range 2.02–1.95 Å.

R$_{sym} = \Sigma_{h}^{m} I(h) / (1/h) - \{1/h\} \Sigma_{h}^{m} I(h)$, where I(h) = observed intensity, and (1/h) = mean intensity obtained from multiple measurements.
positioned and isotropic temperature factor refinement alternating with manual model adjustment (initially using data from 30–2.2 Å and then extending the range to 30–1.95 Å), this loop region could be built unambiguously. Water molecules were assigned to peaks in $F_o - F_c$ electron density maps at a contour greater than 3.5 $\sigma$ that were within hydrogen bonding distance from a potential partner. The sugar complex was refined by similar strategy, starting from the refined model for the QPDWG-HDRPY data set. As found in other studies of this crystal form (10, 11), the positioning of GalNAc was unambiguous for protomer A and C and less well defined for protomer B. Only the $\beta$ anomer of GalNAc was modeled. In addition to the three GalNAc residues found in the principal Ca$^{2+}$I-binding site (one for each protomer), another GalNAc molecule was found between two protomers, near the end of the neck and the first $\beta$-strand (residues 100–107) of protomer C and the first $\alpha$-helix of protomer A. This binding site probably has no biological relevance and is observed because of the high concentrations of sugar used in the soaking experiment. The final native model contains residues 73–226 of each protomer, 412 water molecules, 9 Ca$^{2+}$ ions, and 2 Cl$^{-}$. The GalNAc complex consists of residues 73–226 of each protomer, 171 water molecules, 9 Ca$^{2+}$ ions, 3 Cl$^{-}$, and 4 molecules of GalNAc.

### RESULTS

**GalNAc Selectivity of MBP-A/HHL-1 Chimeras**—Previous studies have revealed that the imidazole ring of His$^{202}$ plays a critical role in preferential binding of GalNAc compared with Gal, as well as serving as part of the pH-sensing mechanism in the asialoglycoprotein receptor (11, 12). For this reason, it was of interest to examine the sugar binding properties of other asialoglycoprotein receptor mimics created in the context of the CRD from MBP to assess the possible influence of other residues in this region (Table II). Addition of Asp$^{212}$ (QPDWG-HD) increases selectivity for GalNAc over Gal from 9- to 14-fold. Remarkably, addition of the Arg-Pro-Tyr (RPy) sequence to QPDWG-HD generates further selectivity for GalNAc, with the resulting QPDWG-HDRPY mutant showing a 40-fold preference for GalNAc over Gal. Therefore, introduction of just four residues beyond the QPDWG-HI mutant generates a chimeric protein that displays nearly the full GalNAc selectivity of HHL-1, as well as having a pH dependence of sugar binding close to that of the hepatic asialoglycoprotein receptor (12).

**Structural Correlates of GalNAc Selectivity**—To examine the structural basis for preferential binding of GalNAc, the structure of QPDWG-HDRPY complexed with GalNAc was determined at 1.95 Å resolution (Tables I and III). The crystal form used in these studies contains a trimer in the asymmetric unit, providing three independent views of the molecule for each structure. As found in other studies of this crystal form (10, 11), differences in overall temperature factors lead to copy A of the trimer being best defined, copy C being slightly less well ordered, and copy B being the least well defined. Nonetheless, in all three copies the mode of binding appears to be identical.

The complex with GalNAc can be best understood with reference to the structures of the QPDWG-HI mutant studied previously. The principal Ca$^{2+}$ and sugar binding site is identical to that observed in QPDWG-HI, and GalNAc is bound in this site as before (Figs. 1, a and b, and 2a). The network of Ca$^{2+}$ coordination and hydrogen bonds that forms the ternary protein/Ca$^{2+}$/sugar complex is retained. Trp$^{133}$ and the glycine-rich loop adopt identical positions in the complex as those in QPDWG-HI, with the 5 and 6 positions of GalNAc stacking over the Trp ring. Thus, the interactions that have previously been observed to define galactose binding in C-type lectins (10, 11) are observed in this complex.

There are several novel features of the interaction of GalNAc with QPDWG-HDRPY arising from additional contacts between the acetamido moiety and the residues that have been introduced into the MBP framework. First, there is a water-mediated hydrogen bond between the NH group of the sugar and Asp$^{212}$ (Fig. 1, a and b). Asp$^{212}$ and the water molecule are positioned by an elaborate network of hydrogen bonds involving the newly introduced Arg$^{216}$ and Tyr$^{218}$, the Ca$^{2+}$ and sugar ligand Glu$^{198}$, and another water molecule (Fig. 1, a and b). The geometry of these interactions appears to be specifically optimized for interaction with the acetamido moiety. Second, the methyl group of GalNAc forms van der Waals’ contacts with the distal edge of the His$^{202}$ ring as described previously (11). However, the imidazole ring is rotated ($\chi$2) by approximately 60° with respect to its position in the QPDWG-H structure (Fig. 3). The repositioning of the His$^{202}$ ring appears to be due to its interaction with the aromatic ring of Tyr$^{218}$. The two rings are roughly parallel, but are laterally displaced with respect to one another, and are also slightly canted with their distal portions closer to one another (Fig. 1c). This arrangement places the partial positive charge of the His$^{202}$ side chain near the $\pi$-electron system of the Tyr$^{218}$ ring (Fig. 1b), an electrostatically favorable interaction commonly found in proteins (21).

The crystal structure of the major subunit of the human hepatocyte asialoglycoprotein receptor (human hepatic lectin-1 (HHL-1)) in its uncomplexed form was recently solved (22). The structures of HHL-1 and QPDWG-HDRPY are virtually identical in the principal Ca$^{2+}$/binding site and the glycine-rich loop (Fig. 1d). Of the residues that provide GalNAc specificity, His$^{202}$, Asp$^{212}$, and Arg$^{216}$, as well as the water molecules positioned by these residues, are also identical to their HHL-1 counterparts. There are, however, two significant differences between HHL-1 and QPDWG-HDRPY, the ring of Tyr$^{218}$ is rotated approximately 20° with respect to Tyr$^{272}$ of HHL-1, and Ser$^{154}$, although in the same general vicinity as Asn$^{206}$ of HHL-1, is positioned differently (Fig. 1d). The latter difference can be attributed to an insertion of one residue in this loop of HHL-1 relative to MBP-A, which moves the backbone slightly. The difference in the tyrosine ring position appears to be due to several effects. In HHL-1, Asn$^{206}$ C$\beta$ is in van der Waals’ contact with Tyr$^{272}$. Also, in QPDWG-HDRPY, the side chain of Arg$^{118}$ would clash with the tyrosine if it were in the conformation seen in HHL-1. This region is more open in HHL-1 because of the presence of glycine at the position equivalent to Arg$^{118}$ of QPDWG-HDRPY. These differences result in a relative shift in the tyrosine ring, so that Tyr$^{272}$ of HHL-1 does not

### Table II

| Protein       | $K_{\text{GalNAc}}$/$K_{\text{Gal}}$ | $R_{\text{cryst}}$ | $R_{\text{free}}$ | $R_{\text{cryst}}$ | $R_{\text{free}}$ |
|---------------|-------------------------------------|--------------------|--------------------|--------------------|--------------------|
| QPDWG         | 0.6 ± 0.1                           | 19.3               | 23.7               | 22.8               | 26.9               |
| QPDWG-H       | 9.2 ± 0.6                           |                    |                    |                    |                    |
| QPDWG-HD      | 14 ± 2                              |                    |                    |                    |                    |
| QPDWG-HDRPY   | 41 ± 4                              |                    |                    |                    |                    |

### Table III

| Data set | Native | With GalNAc |
|----------|--------|-------------|
| Resolution range (Å) | 30.0–1.95 | 30.0–1.95 |
| $R_{\text{cryst}}$ | 18.5/20.1/23.0 | 31.4/43.8/36.0 |
| $R_{\text{free}}$ | 21.9/24.2/26.5 | 35.4/46.9/39.7 |
| Water molecules | 30.2 | 38.9 |
| All atoms | 23.1 | 39.0 |

### Notes

- $R_{\text{cryst}} = \sum |F_o| - |F_c|/\sum |F_o|$, where $|F_o|$ = observed structure factor amplitude and $|F_c|$ = calculated structure factor amplitude for the working and test sets.

### References

1. Previous studies have revealed that the imidazole ring of His$^{202}$ plays a critical role in preferential binding of GalNAc compared with Gal, as well as serving as part of the pH-sensing mechanism in the asialoglycoprotein receptor (11, 12).

2. Remarkably, addition of the Arg-Pro-Tyr (RPy) sequence to QPDWG-HD generates further selectivity for GalNAc, with the resulting QPDWG-HDRPY mutant showing a 40-fold preference for GalNAc over Gal.

3. The repositioning of the His$^{202}$ ring appears to be due to its interaction with the aromatic ring of Tyr$^{218}$. The two rings are roughly parallel, but are laterally displaced with respect to one another, and are also slightly canted with their distal portions closer to one another.

4. This arrangement places the partial positive charge of the His$^{202}$ side chain near the $\pi$-electron system of the Tyr$^{218}$ ring, an electrostatically favorable interaction commonly found in proteins.

5. The crystal structure of the major subunit of the human hepatocyte asialoglycoprotein receptor (human hepatic lectin-1 (HHL-1)) in its uncomplexed form was recently solved (22).
make a direct cation-π interaction with His256, unlike the Tyr218-His202 interaction in QPDWG-HDRPY. Despite these differences, Fig. 1d demonstrates that the mode of GalNAc binding observed in QPDWG-HDRPY provides a good structural mimic of the binding site in the hepatocyte asialoglycoprotein receptors.
A modest enhancement of GalNAc selectivity is achieved by addition of Asp$^{212}$ to QPDWG-H, whereas further addition of the RPY loop greatly enhances selective GalNAc binding. These data suggest that the precise positioning of Asp$^{212}$ through its interactions with Arg$^{216}$ and Tyr$^{218}$ is important in creating an energetically significant water-mediated hydrogen bond with...
the N-acetyl amide group. The structure of the QPDWG-HD mutant shows that Asp 212 interacts with His 218 found in the parent MBP-A sequence, changing its position relative to that described here, so this mutant does not allow dissection of the structural role of Asp 212 from the positioning effect of the RPY loop. The positioning of Tyr 218 by the network of interactions shown in Fig. 1b appears to orient His 218 optimally in the binding site. The strong preference of the mutant studied here for GalNAc suggests that the complete network of interactions is essential in creating a subsite for the N-acetyl group.

Structural Transitions in Ca 2+ Binding—The structure of QPDWG-HDRPY was determined in its sugar-free form in the presence of Ca 2+. Unlike the case of the GalNAc complex, the three independent copies of the molecule in the asymmetric unit show significant differences in structure, with each copy in the native structure displaying a different conformation in the principal Ca 2+- and sugar-binding site. In copy A, the conformation of the protein is the same as that in the GalNAc complex (Fig. 2a and b). Two water molecules form the same Ca 2+ coordination and hydrogen bonds as the 3- and 4-OH groups of GalNAc. The replacement of vicinal sugar hydroxyl groups by water has been observed previously in wild type and mutant MBPs (7, 10, 16). Thus, copy A represents a conformation fully competent to bind to sugar ligands.

In copy C, the conformation is similar to copy A, except that in 50% of the molecules Asn 210 adopts a different χ1 rotamer and is swung out of the Ca 2+ site (Fig. 2c). A water molecule replaces the amide group of Asn 210 and forms hydrogen bonds with His 210. In this conformation, the Asn side chain would have to rotate back to bind Ca 2+ and sugar. Moreover, in this conformation the Asn 210 side chain amide cannot form a hydrogen bond with His 210 (Fig. 1, a and b). In addition to the alternate conformation of Asn 210, Ser 154, which forms a hydrogen bond with the other side of the His 210 imidazole ring, adopts two conformations. In the ligand-bound conformation, the singly protonated His 210 gives an unambiguous arrangement of hydrogen bond donors and acceptors (Fig. 1b); His 210 is a hydrogen bond acceptor from Asn 210 and is a donor to the OOH group of Ser 154. In the alternate Asn 150 conformation, however, the location of the proton on His 210 is not defined uniquely.

Copy B shows the most dramatic conformational differences with respect to the GalNAc-bound structure. Ca 2+ prefers a pentagonal bipyramidal coordination state, with five oxygen atoms arranged in a pentagonal plane and two or three ligands at apices 90° away from the plane. In wild type and mutant MBP sugar complexes studied to date (5–7, 10, 11, 16), side chain oxygen atoms of Gln 185, Asp 187, Ghu 198, and Asn 210, as well as the main chain carbonyl oxygen of Asp 211, form the pentagonal equatorial plane; a side chain oxygen atom of Asp 211 forms one apex, and the 3- and 4-OH groups of the sugar bisect the other apex to form an eight-coordinated Ca 2+ (QP-DWG-HDRPY numbering) (Fig. 2a). This eight-coordinate arrangement is found in copy A of the native QPDWG-HDRPY (Fig. 2b). In copy B, however, Asn 210 is rotated out of the site as observed in copy C, and a water molecule occupies a nearby coordination position (Fig. 2d). The position formerly occupied by the amide group of Asn 210 is replaced by a water molecule, which forms hydrogen bonds with His 210, Asp 212, and another Ca 2+-coordinating water molecule. The γ angle of Glu 185 changes by approximately 75° (Fig. 2e). The γ angle of Glu 185 also shows small differences relative to the sugar-bound structure (Fig. 2e). These changes produce an altered Ca 2+ coordination geometry, such that the pentagonal equatorial plane of the eight-coordinated, ligand bound state (Fig. 2f) is rotated approximately 90° with respect to the sugar-binding conformation. Side chain oxygen atoms of Gln 185, Ghu 198, and Asp 211, along with two water molecules, form the equatorial pentagonal plane (Fig. 2g). The main chain carbonyl oxygen of Asp 211 and a side chain oxygen of Asp 210 form the two apices, giving a seven-coordinate state. Thus, the coordination number and geometry is altered in copy B, with oxygen atoms from Asp 187 and Asn 210 now, respectively, apical and equatorial, reversing their roles in the sugar-bound structure (Fig. 2, f and g).

In addition to the local changes in the principal Ca 2+ site, the glycine-rich loop is dramatically rearranged in copy B. The loop still adopts a well structured conformation, but it is displaced away from the protein (Fig. 4a and b). As part of this change, Trp 189 changes position and is considerably more solvent exposed than when it is packed against the loop in the sugar-binding conformation. The loop rearrangement cannot be ascribed to a simple hinge motion, because many of the backbone Ramachandran angles differ between the two structures. The distal end of the loop forms contacts with neighboring molecules in the lattice, which may explain its order, but in any case the structure indicates that the loop is dynamic.

The rearrangement of the structure in copy B is reminiscent of changes in the equivalent region of mannose-binding protein C (MBP-C) upon loss of Ca 2+ (23). In the crystal structure of apo-MBP-C, a conserved cis-proline in the Cu 2+ site (position 194; equivalent to 186 in MBP-A) is observed to isomerize to the trans-form in some copies of the molecule. The equivalent Pro 186 of QPDWG-HDRPY retains the cis-configuration, and in this sense is most similar to copy B of the apo-MBP-C structure (Fig. 4c). Comparison of these structures (Fig. 4d) shows that the region equivalent to the glycine-rich loop moves in a similar direction. Thus, copy B may represent a structural rearrangement related to loss of Ca 2+ from the CRD.

DISCUSSION

GalNAc Selectivity—The 40-fold preference of QPDWG-HDRPY for GalNAc over Gal is nearly as great as the 60-fold preference shown by RHL-1. The QPDWG-HDRPY structure presented here reveals a specific subsite for the 2-acetamido group that is formed by an intricate network of interactions among His 202, Asp 212, Asn 210, and Tyr 218, which are highly conserved in C-type lectins that display preferential binding to GalNAc. Comparison with the unliganded structure of the major subunit of the human hepatocyte asialoglycoprotein receptor (22) reveals that the present structure is virtually identical in the sugar-binding site. These observations indicate that the structure of QPDWG-HDRPY in complex with GalNAc is likely

3 B. Zagrovic, H. Feinberg, A. Kolatkar, K. Drickamer, W. Weis, unpublished data.

4 A. Kolatkar, K. Ng, S. Park-Snyder, K. Drickamer, W. Weis, manuscript in preparation.
to represent faithfully the mode of interaction in these proteins. Unlike the hepatocyte asialoglycoprotein receptor, the macrophage galactose receptor (MGR) shows no preference for GalNAc over Gal. All of the RHL-1 residues introduced into QPDWG to give RHL-1-like GalNAc selectivity, i.e., His202, Asp212, Arg216, Pro217, and Tyr218 (QPDWG numbering), are found in the MGR. However, it has been shown that position 208 of RHL-1 (230 in the MGR) influences the GalNAc selectivity of this receptor. In RHL-1, this position is asparagine. Replacement of this asparagine with the valine residue present in MGR or isoleucine almost abrogates GalNAc selectivity (24). The equivalent residue in MBP-A is Ser154. Replacement of Ser154 of QPDWG-H with Val reduces GalNAc selectivity from 9- to 3-fold (11). The structure of QPDWG-H containing Ser154→Val showed that the β-branched methyl group sterically forces the imidazole ring of His202 to rotate 25°, but the van der Waals contacts with the acetamido methyl group are retained (Fig. 3) (11). The relationship of the altered His202 position to the reduced GalNAc affinity is unclear from this structure. It should be noted that the rotation of the His202 ring in the presence of Val154 relative to its position in QPDWG-H is opposite to the direction that would be needed to superimpose it on that in QPDWG-HDRPY (Fig. 3).

In QPDWG-HDRPY, rotation of the His202 imidazole ring is restricted because of its interaction with Tyr218. The presence of Val230 in the MGR may therefore be more disruptive to the GalNAc binding site than observed in the Ser154→Val mutant of QPDWG-H (11). In particular, the steric clash between valine and the imidazole ring cannot be relieved without forcing the ring of Tyr218 to move. This in turn would disrupt the intricate network of hydrogen bonds in the acetamido subsite. It must be noted that the natural galactose receptors contain an extra residue in the loop that contains Ser154 in QPDWG-HDRPY, and the position of these residues is not strictly equivalent (Fig. 1d). Ultimately, structural analysis will be required to determine the effects of β-branched amino acids at this position.

**Structural Basis of Altered pH and Ca2+ Affinity**—The three conformations of the uncomplexed QPDWG-HDRPY observed in the present crystals may represent discrete structural states associated with Ca2+ binding and release by this protein. The crystals were grown at pH 8.0 at a nominal Ca2+ concentration of 15 mM. The effective Ca2+ affinity, $K_{Ca}$, obtained by assaying sugar ligand binding at pH 7.8 as a function of Ca2+ is 7.2 mM (12). Thus, the Ca2+ concentration is not much over the $K_{Ca}$, and it is plausible that three conformations represent intermediates between Ca2+-bound and -free forms found in solution. The fact that the three independent copies in the GalNAc-soaked crystals are identical reflects the fact that sugar and Ca2+ binding are linked equilibria, so that the effective Ca2+ affinity is higher in the presence of GalNAc. The change in Ca2+ coordination geometry and the fact that the movement of the glycine-rich loop in copy B is similar to changes in this region in Ca2+-free MBP-C (23) suggests that this conformation represents a structural state on the pathway to Ca2+ release.

Based on the available structural data, the following pathway of transitions from Ca2+-bound to Ca2+-free forms of C-type CRDs can be proposed. Copy A of the uncomplexed QPDWG-HDRPY is identical to the GalNAc-bound structure, with the exception of the two water molecules that substitute for the 3- and 4-OH groups of the sugar. This structure therefore represents the fully bound Ca2+-site structure in a conformation competent to bind to sugar ligands. Copy C is identical to A, except that Asn210 is rotated out of the site. The amide group of the Asn210 side chain serves as a hydrogen bond donor.
to His\textsuperscript{202}. At acid pH, His\textsuperscript{202} will be protonated, making it unable to accept the Asn\textsuperscript{210} hydrogen bond. This electrostatic clash can be relieved by the Asn\textsuperscript{210} side chain rotation observed in copy C, and this conformation would represent the first step in Ca\textsuperscript{2+} release. This would be followed by the side chain rotations of Gln\textsuperscript{185} and Gln\textsuperscript{393}, the resultant change in coordination geometry and coordination number from 8 to 7 (Fig. 2, f and g), and the movement of the glycine-rich loop observed in copy B (Fig. 4). These changes may be sufficient to release sugar ligands from the protein. It is not known whether the Ca\textsuperscript{2+} is actually released, but if it is this pathway would ultimately produce a state like that observed in apo-MBP-C. The changes may be accompanied by loss of the auxiliary Ca\textsuperscript{2+}. Previous work on the MBPs has shown that the Ca\textsuperscript{2+}-free state corresponds to an ensemble of conformations. These include states in which Pro\textsuperscript{186} of MBP-A (194 in MBP-C) has isomerized to the trans-configuration. Kinetic measurements of this transition have shown that 80% of the Ca\textsuperscript{2+}-free molecules are in the trans-form (25), and the slow halftime (several minutes) of cis-trans-proline isomerization slows attainment of the cis-conformation required to bind Ca\textsuperscript{2+}. It was proposed that this may serve as a kinetic trap to ensure that receptor and ligand can be sorted from one another (25).

In a recent structure of the fourth CRD from the macrophage mannose receptor, another C-type lectin, the principal Ca\textsuperscript{2+} site is occupied, but the auxiliary site is vacant (26). In this structure, the principal site is rearranged in a manner inconsistent with the geometry of mannose binding in the MBPs. Based on these observations, it was proposed that this CRD may use a different mechanism to release sugar ligands, in which loss of the auxiliary Ca\textsuperscript{2+} leads to changes in the principal site, rendering the latter unable to bind sugar. Although the macrophage mannose receptor and RHL-1 may differ in how they sense acid pH, these two distinct kinds of C-type CRDs may have in common the use of pH-dependent changes outside of the immediate coordination shell of the principal Ca\textsuperscript{2+} site to alter their ability to interact with sugar ligands.

**pH Dependence of Ca\textsuperscript{2+} Binding**—Carbohydrate binding in the C-type lectins is strictly dependent on the precise 8-fold coordination found in the principal Ca\textsuperscript{2+} site (4) (Fig. 2, a, b, and f). This dependence couples sugar binding with pH. As the pH drops, a higher fraction of molecules become protonated, resulting in a reduced effective Ca\textsuperscript{2+} affinity. MBPs, which are not endo-rectifiers, show a sharp pH dependence of ligand binding, in which binding to multivalent glycoprotein ligands is lost as the pH is lowered between 5.4 and 4.8. The pKa\textsubscript{H} defined as the pH at which half-maximal sugar binding occurs at a given Ca\textsuperscript{2+} concentration, is 5.0 for the QPDWG mutant of MBP-A at 1 mM Ca\textsuperscript{2+} (12). This effect is likely due to direct protonation of one or more Ca\textsuperscript{2+} ligands. In contrast, endorectifiers such as RHL-1 must release their ligands in more mildly acidic conditions and show a more gentle dependence of ligand binding on pH. Mutagenesis of RHL-1 (12) has implicated His\textsuperscript{256} as an important determinant of pKa\textsubscript{H}. Introduction of His\textsuperscript{202} into RHL-1 increases pKa\textsubscript{H}, Introduction of Asp\textsuperscript{252} further raises pKa\textsubscript{H}, but Asp\textsuperscript{212} has no effect unless His\textsuperscript{202} is present (12). These data imply that His\textsuperscript{202} is largely responsible for the pH sensitivity of ligand binding.

Although pKa\textsubscript{H} is dependent on the Ca\textsuperscript{2+} concentration and is therefore not an intrinsic property of the protein, it is likely to be correlated with the pKa of His\textsuperscript{202}. At 1 mM Ca\textsuperscript{2+}, pKa\textsubscript{H} of QPDWG-H is 5.7. The solution pKa of free histidine is approximately 6.5. The structures of GalNAc-bound and copy A of native QPDWG-HDRPY show that His\textsuperscript{202} accepts a hydrogen bond from Asn\textsuperscript{210}. Protonation of His\textsuperscript{202} will disrupt this interaction, so this arrangement would be expected to lower the effective pKa of His\textsuperscript{202}. The addition of Asp\textsuperscript{212} to QPDWG-H further raises the pKa to 6.1. The present structures show that Asp\textsuperscript{212} is approximately 3.9 Å from the NE\textsubscript{2} imidazole nitrogen of His\textsuperscript{202}. This distance is too long for a hydrogen bond interaction but would be expected to provide a significant perturbation of the electrostatic environment of His\textsuperscript{202}. In particular, the presence of negative charge would be predicted to favor protonation of His\textsuperscript{202} and raise the pKa.

Addition of the RPY loop to QPDWG-HD further raises the pKa\textsubscript{H}. Changes in pH dependence produced by introduction of these residues are likely due to a combination of effects. Arg\textsuperscript{216} interacts directly with Asp\textsuperscript{212}, which would be predicted to lessen the effect of the latter on the pKa of His\textsuperscript{202} by neutralizing the negative charge. It would be expected that the presence of Arg\textsuperscript{216} in the vicinity of His\textsuperscript{202} would lower the pKa of the histidine. This notion is supported by the observation that mutation of Arg\textsuperscript{270} to leucine in RHL-1 raises the pKa\textsubscript{H} (12). On the other hand, there are two well ordered water molecules held in position by this arginine and two acidic residues (Fig. 1, a and b). The network of interactions (Fig. 1, a and b) serves to position Tyr\textsuperscript{218}, so that its delocalized π-electron system forms an electrostatically favorable interaction with His\textsuperscript{202}. Formation of a hydrogen bond by the phenolic hydroxyl proton with Asp\textsuperscript{212} will enhance the partial negative charge of the aromatic system. Protonation of His\textsuperscript{202} would be expected to enhance this interaction, as has been observed in other systems (27). As noted above, however, the position of Tyr\textsuperscript{270} in HHL-1 differs slightly from that of Tyr\textsuperscript{218} in QPDWG-HDRPY (Fig. 1d), so that a direct cation-π interaction with His\textsuperscript{256} does not exist in HHL-1. Nonetheless, the proximity of the delocalized π system of the tyrosine may provide a small electrostatic contribution that would tend to raise the pKa of His\textsuperscript{202}. Finally, as discussed above, Val\textsuperscript{300} of the MGR, equivalent to Ser\textsuperscript{154} of QPDWG-HDRPY, would be expected to disrupt this network, which may account for the fact that the pKa of 6.3 (1 mM Ca\textsuperscript{2+}) found for the MGR\textsubscript{A} is lower than the pH\textsubscript{B} of RHL-1. Collectively, these observations indicate that the interactions of the Asp\textsuperscript{212}, Arg\textsuperscript{216}, Tyr\textsuperscript{218}, and water must influence the electrostatic environment of His\textsuperscript{202} to alter its pKa. More detailed dissection of the contributions of these residues awaits direct measurement of the His\textsuperscript{202} pKa and detailed electrostatic calculations.

Acknowledgments—We are grateful to Markus Meier and Peter Burkhard for providing the HHL-1 coordinates prior to general release. We thank Andy May and Katy Spink for data collection at Stanford Synchrotron Radiation Laboratory and for comments on the manuscript.

REFERENCES

1. Spiess, M. (1990) Biochemistry 29, 10009–10018
2. Halberg, D. F., Wagner, R. E., Farrell, D. C., Hildreth, J., IV, Quesenberry, M. S., Leob, J. A., Holland, E. C., and Drickamer, K. (1987) J. Biol. Chem. 262, 9628–9638
3. Beavil, A. J., Edmeades, R. L., Gould, H. J., and Sutton, B. J. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 753–757
4. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) Immunochem. Rev. 163, 19–34
5. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991) Science 254, 1608–1615
6. Weis, W. I., Drickamer, K., and Hendrickson, W. A. (1992) Nature 360, 127–134
7. Ng, K. K.-S., Drickamer, K., and Weis, W. I. (1996) J. Biol. Chem. 271, 663–674
8. Drickamer, K. (1992) Nature 360, 183–186
9. Iobst, S. T., and Drickamer, K. (1994) J. Biol. Chem. 269, 15512–15519
10. Kolatkar, A. R., and Weis, W. I. (1996) J. Biol. Chem. 271, 6679–6685
11. Kolatkar, A. R., Leung, A. K., Isecke, E. B., Brossmer, R., Drickamer, K., and Weis, W. I. (1998) J. Biol. Chem. 273, 19502–19508
12. Wragg, S., and Drickamer, K. (1999) J. Biol. Chem. 274, 35490–35406

5 H. Dobbyn and S. Wragg, unpublished observations.
pH-dependent GalNAc Binding in C-type Lectins

13. Weis, W. I., and Drickamer, K. (1994) *Structure* **2**, 1227–1240
14. Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y., and Inouye, M. (1984) *EMBO J.* **3**, 2437–2442
15. Drickamer, K., Dordal, M. S., and Reynolds, L. (1986) *J. Biol. Chem.* **261**, 6878–6886
16. Ng, K. K.-S., and Weis, W. I. (1997) *Biochemistry* **36**, 979–988
17. Weis, W. I., Crichlow, G. V., Murthy, H. M. K., Hendrickson, W. A., and Drickamer, K. (1991) *J. Biol. Chem.* **266**, 20678–20686
18. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
19. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
20. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
21. Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9459–9464
22. Meier, M., Bider, M. D., Malashkevich, V. N., Spiess, M., and Burkhard, P. (2000) *J. Mol. Biol.* **300**, 857–865
23. Ng, K. K.-S., Park-Snyder, S., and Weis, W. I. (1998) *Biochemistry* **37**, 17965–17976
24. Iobst, S. T., and Drickamer, K. (1996) *J. Biol. Chem.* **271**, 6686–6693
25. Ng, K. K.-S., and Weis, W. I. (1998) *Biochemistry* **37**, 17977–17989
26. Feinberg, H., Park-Snyder, S., Kelatkar, A. R., Heise, C. T., Taylor, M. E., and Weis, W. I. (2000) *J. Biol. Chem.* **275**, 21539–21548
27. Loewenthal, R., Sancho, J., and Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 759–770