Computational investigation of glycosylation effects on a Family 1 carbohydrate-binding module
Courtney B. Taylor, 1 M. Faiz Talib, 1 Clare McCabe, 1,2,*, Lintao Bu, 3 William S. Adney, 4 Michael E. Himmel, 4 Michael F. Crowley, 4 and Gregg T. Beckham 3,5,*
1 Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville TN 37235
2 Department of Chemistry, Vanderbilt University, Nashville, TN 37235
3 National Bioenergy Center, National Renewable Energy Laboratory, Golden CO 80401
4 Biosciences Center, National Renewable Energy Laboratory, Golden CO 80401
5 Department of Chemical Engineering, Colorado School of Mines, Golden, CO 80401
* Corresponding Authors: c.mccabe@vanderbilt.edu, Gregg.Beckham@nrel.gov

Capsule
Background: Family 1 carbohydrate-binding modules (CBMs) are often components of cellulases for binding to cellulose.

Results: Family 1 CBM binding affinity can be dramatically affected by the presence of O-glycosylation near the CBM binding face.

Conclusion: Glycosylation should be accounted for in CBM binding affinity studies.

Significance: Glycosylation can be harnessed to tune cellulase binding affinity, which is known to affect activity.

Abstract
Carbohydrate-binding modules (CBMs) are ubiquitous components of glycoside hydrolases, which degrade polysaccharides in nature. CBMs target specific polysaccharides, and CBM binding affinity to cellulose is known to be proportional to cellulase activity, such that increasing binding affinity is an important component of performance improvement. To ascertain the impact of protein and glycan engineering on CBM binding, we use molecular simulation to quantify cellulose binding of a natively-glycosylated Family 1 CBM. To validate our approach, we first examine aromatic-carbohydrate interactions on binding, and our predictions are consistent with previous experiments, showing that a tyrosine to tryptophan mutation yields a 2-fold improvement in binding affinity. We then demonstrate that enhanced binding of 3 to 6-fold over a non-glycosylated CBM is achieved by the addition of a single, native mannose or a mannose dimer, respectively, which has not been considered previously. Furthermore, we show that the addition of a single, artificial glycan on the anterior of the CBM, with the native, posterior glycans also present, can have a dramatic impact on binding affinity in our model, increasing it up to 140-fold relative to the non-glycosylated CBM. These results suggest new directions in protein engineering, in that modifying glycosylation patterns via heterologous expression, manipulation of culture conditions, or introduction of artificial glycosylation sites, can alter CBM binding affinity to carbohydrates and may thus be a general strategy to enhance cellulase performance. Generally, our results suggest that CBM binding studies should consider the effects of glycosylation on binding and function.

Introduction
Carbohydrate-binding modules (CBMs) represent a primary biological means for protein-carbohydrate recognition (1). CBMs recognize and bind to a wide range of carbohydrates and are often components of multi-modular cellulase, hemicellulase, or chitinase enzymes, which are able to deconstruct plant, fungal, or algal cell wall carbohydrates (1-5). It has been shown that CBMs serve two roles as components of multi-modular, carbohydrate-active enzymes: to target the specific structures of interest for catalytic action and to maintain proximity to a given carbohydrate surface (1,2). A third hypothesized role is disruption of the carbohydrate crystal packing , making the chains easier to decrystallize for enzymatic attack (6-8), but the mechanism of this function remains unknown and evidence to support this role is limited (1).

Engineering enhanced cellulase enzymes is currently a topic of significant worldwide interest, a situation primarily driven by research to enable commercialization of renewable biofuels from lignocellulosic biomass (6-10). Several routes exist for increasing cellulase performance, including rational and evolutionary methods to improve specific activity (8), screening for...
improved thermal tolerance (11,12), and the often overlooked strategy of increasing binding affinity to carbohydrates via CBM engineering (13-16). In the lattermost case, several groups have shown that by increasing CBM binding affinity to cellulose via single point mutations or by swapping complete CBMs from other enzymes with higher binding affinity, cellulase enzyme activity can be improved (15,16). This effect is likely due to higher enzyme concentrations on the cellulose surface, which potentially leads to a higher fraction of catalytically engaged enzymes. Higher binding affinity and thus higher activity, in turn, will lead to lower enzyme loadings and the eventual realization of more cost-effective biofuels processes (6). To examine CBM binding at the molecular level, we examine a novel strategy, namely the use of CBM glycosylation, to enhance the affinity of cellulases, which we predict will improve affinity more so than standard protein engineering strategies wherein amino acids are mutated to other residues. Specifically, using simulations we quantify the impact of natural and engineered glycosylation on the binding affinity of a model CBM, the well-characterized Family 1 CBM from the Trichoderma reesei (Hypocrea jecorina) Family 7 processive cellobiohydrolase (Cel7A). The Cel7A enzyme is shown in Figure 1 (17-19).

Family 1 CBMs are ~36 residue proteins that exhibit high sequence homology and are predominantly produced by fungi (20). As shown in Figure 2, the Cel7A CBM exhibits a planar face with three tyrosine residues that are hypothesized to form the binding face to cellulose (21,22). The Cel7A CBM, linker, and catalytic domain (CD) have all been shown to be glycosylated when expressed in the native host (23). Two residues on the CBM, threonine 1 and serine 3 (Figure 2), are natively glycosylated with at least one mannose and potentially up to three mannoses each, but as Harrison et al. discuss, the O-glycosylation assignments on the Cel7A CBM and linker are at best an average extent of glycosylation. An engineered (non-native) glycan was also chosen for our study and is shown on serine 14 in Figure 2.

Experimental studies to date have examined the role of aromatic and polar residues on the Cel7A CBM binding affinity (13-15,24), but no work to our knowledge has considered the role of natural, recombinant, or engineered O-glycosylation on the Family 1 CBM binding affinity. Boraston et al. demonstrated that the addition of high mannose N-glycans (GlcNAc2Man8 and higher) to a Family 2 CBM (CBM2a) from Cellulomonas fimii, expressed by the yeast Pichia pastoris, was detrimental to binding affinity. This result was attributed to the size of the N-glycan inhibiting the ability of the CBM to interact with the carbohydrate substrate (25). In a later study, Boraston et al. expressed a modified CBM2a with the N-glycosylation sites removed via mutations of asparagine to alanine in P. pastoris. The authors demonstrated that the O-glycosylation did not impact binding affinity in this case, because the serine and threonines are located above the cellulose surface where the glycans could neither interact with cellulose nor inhibit the CBM-cellulose interaction as the N-glycan did previously (25,26).

In many filamentous fungi and yeasts, O-glycans often contain fewer carbohydrate moieties than N-glycans (26,27). In the current work we examine natural or small artificial O-glycans, neither of which are expected to restrict access of the CBM to cellulose. Before studying the impact of O-glycans on binding affinity, to validate the simulation approach, we first study a number of amino acid mutations for which we can compare with the experimental results of Linder et al., who showed that the non-glycosylated Cel7A CBM binding affinity is improved by the mutation of residue 5 from tyrosine to tryptophan (denoted Y5W) (14). Experimentally bound and unbound concentrations of the non-glycosylated CBM variants were used to generate adsorption isotherms (13,14) from which partition coefficients can be determined and the relative binding free energy ($\Delta\Delta G$) calculated as:

$$\Delta\Delta G = -RT\ln\left(\frac{K_{\text{Mut}}}{K_{\text{WT-NG}}}\right)$$ (1)

where $K_{\text{Mut}}$ and $K_{\text{WT-NG}}$ are the partition coefficients for the mutant CBM of interest and
the wild-type CBM without glycosylation (NG = no glycosylation), respectively, $T$ is temperature, and $R$ is the gas constant. We note that the partition coefficients are defined between the free CBM in solution and bound to the hydrophobic face of cellulose (3). All of the variants studied by Linder et al., except the Y5W mutation, were found to exhibit lower binding affinity to cellulose than the wild-type, non-glycosylated Cel7A CBM with $\Delta\Delta G$ values of +0.2 to +1.7 kcal/mol. The Y5W mutation exhibited an increase in $K_{\text{Mut}}$ over $K_{\text{WT-NG}}$ by a factor of 1.6 (-0.26 kcal/mol) (13,14). In similar work, Takashima et al. examined the enzymatic activity of various Humicola grisea Cel7A enzymes with mutant CBMs (15). The mutation of Y5W produced a favorable 33% increase in binding and a 12% increase in overall enzymatic activity, similar to the results reported by Linder et al. (14,15).

Although the effects on binding of O-glycosylation on the Cel7A CBM have not been studied, both N- and O-glycosylation in cellulases are known to vary with expression host and growth conditions and have the potential ability to affect the activity and stability of cellulases (25,28-31). To understand the impact of glycosylation on CBM function and to potentially engineer enhanced cellulases by increasing CBM binding affinity through introduction of artificial glycosylation sites through amino acid mutation or altering glycosylation via heterologous expression or manipulation of growth conditions.

**Approach**

The systems studied by thermodynamic integration (TI) and long MD simulations are summarized with their respective results in Tables 1 to 3, along with the nomenclature that will be used throughout the text and in the Supplemental Information. All binding affinity comparisons, $K_{\text{Mut}}/K_{\text{WT-NG}}$, are performed as in Equation 1 by comparing to the wild-type non-glycosylated CBM. The TI simulations (32,33) are designed to measure the relative binding free energy between the CBM in solution and the CBM on the hydrophobic face of cellulose as a function of amino acid mutation or addition of glycosylation. The thermodynamic cycle thus consists of TI calculations of the CBM in solution (without cellulose) and the CBM on the hydrophobic face of cellulose. Full details of the simulations performed are provided in the Supplemental Information and are summarized in Figure S2.

**Validation simulations of amino acid mutations**

Because the Tyr-5 residue has been well characterized for the Cel7A CBM (13,14,21), we use this mutation as a validation of our computational approach. These simulations were run with no glycosylation at Thr-1 and Ser-3 for direct comparison to binding experiments where the CBMs were produced via solid-state peptide synthesis, and thus the CBMs have no glycosylation (13,14). For Tyr-5, four TI calculations were performed: Y5A, Y5W, Y5F, and F5A (Figure S2). Y5A was chosen to modify
the polarity of the Tyr-5 residue by mutation to Ala-5, which is one of the simplest non-polar amino acids, while Y5W replaces Tyr-5 with Trp-5, a larger aromatic residue with known higher binding affinity (14). Y5F and F5A (Tyr-5 to Phe-5 and Phe-5 to Ala-5, respectively) were selected as a control to ensure internal consistency, such that:

$$\Delta \Delta G(Y5A) - \Delta \Delta G(F5A) = \Delta \Delta G(Y5F) \quad (2)$$

**Glycosylation simulations** Previous work by Harrison et al. indicated the presence of at least one mannose at both threonine 1 (Thr-1) and serine 3 (Ser-3) on the Family 1 CBM expressed in a particular T. reesei strain (23). Based on the NMR structure of the CBM (22), which suggests that the sugar on Thr-1 may be too far above the cellulose surface to interact with the cellulose directly, we have not examined mutations of the native O-glycosylation on Thr-1. An engineered glycosylation site was also studied at Ser-14, which is positioned near the cellulose surface at the anterior of the CBM. The Cel7A CBM contains eight Thr and Ser residues, but only two, Ser-3 and Ser-14, are located near the surface. As previously discussed, Boraston et al. showed that O-glycans far above the surface do not impact binding affinity (26); anticipating that glycosylation in Cel7A far above the surface would have a similar result, we thus focused our study on Ser-14. This TI simulation was conducted both with and without the native glycans at Thr-1 and Ser-3 (Figure S2). The four glycan mutations studied were used to test the impact of natural (S3M1 and S3M2) and engineered (S14M1-NG, S14M1) glycosylation on binding affinity. Additionally, to test the impact of glycosylation on amino acid mutations, we repeated the Y5A and Y5W simulations with the native Thr-1 and Ser-3 mannose pattern, denoted Y5A-G and Y5W-G. This also provides a second check for internal consistency in that:

$$\Delta \Delta G(Y5W) + \Delta \Delta G(S3M1) = \Delta \Delta G(S3M1+Y5W-G) \quad (3)$$

**MD simulations to examine the impact of glycosylation on CBM stability** In addition to the TI calculations, 100 ns MD simulations of the bound wild-type CBM, the bound Y5A mutant CBM, and the bound Y5W mutant CBM were all performed without glycosylation. MD runs of bound CBMs using the glycan patterns found experimentally (23), were also conducted. Finally, simulations of the bound, native mannose at Thr-1 and dimer at Ser-3 and the bound, native mannoses at Thr-1 and Ser-3 with the engineered mannose at Ser-14 were also performed.

**Results**

**Thermodynamic integration simulations** The $\Delta \Delta G$ results from the TI calculations are provided in Tables 1 to 3. We constructed Langmuir isotherms for key mutations, as is typical with binding affinity measurements. The experimental and predicted changes in the mutant and wild-type partition coefficients, $K_{\text{Mut}}$ and $K_{\text{WT-NG}}$, respectively, are shown in Figure 3. For the simulations where S3M1 is the intermediate step between the non-glycosylated wild-type and the mutation, the total impact of the mutation is calculated by adding $\Delta \Delta G_{S3M1}$ and the $\Delta \Delta G$ of the current run, resulting in S3M1+Y5A-G, S3M1+Y5W-G, S3M1+S3M2, and S3M1+S14M1 (Table 1 to 3).

The $\Delta \Delta G$ results for the amino acid TI calculations are shown in Table 1. We find agreement between the experimental data (13-15) and our computational predictions for the two large mutations: Y5W and Y5A, which provides direct quantitative evidence for the viability of our computational approach. Congruent with the results from Linder et al. (13,14) for the non-glycosylated systems, mutation of Tyr to Ala was found here to decrease binding affinity, whereas mutation to Trp increases binding affinity. The nearly 2-fold improvement in Y5W $K_{\text{Mut}}/K_{\text{WT-NG}}$ is also the same order of magnitude as the 1.3-fold improvement measured by Takashima et al. for a similar Family 1 CBM (15). As discussed earlier, the F5A mutation was studied to validate the simulation approach, and Equation 2 holds nearly within error for the Y5A, F5A, and Y5F mutations, the details of which are documented in the Supplemental Information. Loss of the large binding surface area and a hydrogen bond in the Y5A case is clearly unfavorable for binding and is confirmed by the long MD simulations described in the Supplemental Information, where a decrease in overall interaction energy with the surface is
observed for the Y5A mutant compared to the wild-type CBM or the Y5W mutant. The relative binding free energy for the Y5F mutant is between that of the Y5A and Y5W mutants, which is expected, given that loss of the hydroxyl group on Tyr removes a hydrogen bonding site but the retention of the aromatic ring maintains the shape of the planar face of the CBM. For the Y5W mutant, the loss of the Tyr hydroxyl is offset by an increase in the surface area of the side chain, which is corroborated by the long MD simulations in that the local Trp-5-surface non-bonded interaction is primarily mediated by van der Waals interactions.

We found no differences in the Y5A and Y5A-G mutant simulations; mutation to Ala-5 is very detrimental to binding with or without the native glycosylation pattern from Harrison *et al.* (23). However, we find that the binding affinity improvement for the Y5W mutant increases 16-fold over the non-glycosylated wild-type CBM with native glycosylation (S3M1+Y5W-G) and also confirmed that Equation 3 holds near within error for Y5W, S3M1, and Y5W-G, the details of which are again documented in the Supplemental Information.

The glycosylation TI simulations were designed to understand the changes in the CBM binding affinity upon (i) the addition of a single, native mannose at Ser-3 (S3M1), (ii) the addition of a mannose dimer at Ser-3 (S3M1+S3M2), and (iii) the addition of a non-native mannose at Ser-14, with the lattermost examined both with and without the native glycosylation at Thr-1 and Ser-3 (S14M1 and S14M1-NG, respectively). For all the glycosylation simulations, we find that CBM binding from solution to cellulose improves with the addition of mannose, increasing both the potential for hydrogen bonding and the surface area for interaction. The ΔΔG results for the S3M1 and S3M2 TI simulations are given in Table 2. The S3M1 glycan TI simulation demonstrates that the addition of the single native mannose at Ser-3 located near the posterior of the CBM improves the binding affinity by 3-fold. The addition of a second mannose at Ser-3 (S3M2), from a second TI calculation, improves the binding affinity by 2-fold increase relative to the S3M1 case, resulting in a total 6-fold improvement over the non-glycosylated wild-type CBM in binding affinity for the addition of the dimer (S3M1+S3M2).

The TI results for the addition of the artificial glycan site on Ser-14, located towards the anterior of the CBM, both with and without native glycosylation, are shown in Table 3. The addition of the single glycan without the native glycans in the posterior region improves the binding affinity by 20-fold in $K_{\text{Mut}}/K_{\text{WT-NG}}$ (S14M1-NG). When the native glycosylation is present and the engineered Ser-14 mannose is added, the favorable ΔΔG increases to 3 kcal/mol, resulting in a 140-fold increase in $K_{\text{Mut}}$ over $K_{\text{WT-NG}}$ (S3M1+S14M1).

For the Ser-14 systems studied, the change in binding affinity is significantly higher than in the Ser-3 cases. To gain insights into the reasons for this increase in binding affinity, we examined thermodynamic and structural properties from the 100 ns MD simulations. First, we examine the interaction energies: Figure 4 shows the total interaction energies, comprising electrostatic and van der Waals energies, between the glycans and the cellulose surface, for the four scenarios studied: S3M1, S14M1-NG, S3M1+S3M2, and S3M1+S14M1. If we consider these interactions, the addition of a single glycan at Ser-3 (S3M1) or Ser-14 (S14M1-NG) adds an additional 4 to 7.5 kcal/mol of favorable interaction with the cellulose. Furthermore, we find that whereas the addition of the second mannose at Ser-3 increases the hydrogen bonding potential (see Supplemental Information), the total interaction energy with the cellulose surface does not change compared to the S3M1 case alone (Figure 4). In contrast, when a mannose is added at Ser-14 with the native glycans present (S14M1), the interaction energy between the Ser-3 mannose and cellulose remains constant, while the Ser-14 mannose adds an additional 3 kcal/mol of favorable interaction with the surface, increasing the total slightly over the non-glycosylated (S14M1-NG) case. In our simulations the ability of the critical CBM binding-face amino acids (13,21) to interact with the surface does not appear to be negatively impacted by the presence of the Ser-3 or Ser-14 mannoses. The glycans also form hydrogen bonds with the surface in all cases, and when the Ser-14 mannose is present, the number of hydrogen bonds possible between the Ser-3 mannose and surface
nearly doubles. While we cannot quantitatively delineate the contributions of mannose-cellulose hydrogen bonding relative to other enthalpic and entropic contributions to improvements in binding free energy, the data suggest that increased hydrogen bonding of the CBM-glycan system with the cellulose surface correlates with improved binding and stability of the CBM on the surface. Details of the mannose-cellulose hydrogen bond potential during the 100 ns simulations can be found in the Supplemental Information.

Finally, we calculated the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) for the extended MD simulations described previously, and found a slight improvement in stability of the CBM backbone over the cellulose surface for the native glycosylated versus non-glycosylated systems. The details of this analysis, along with a comparison of the interaction energy between residues of interest and the cellulose surface are provided in the Supplemental Information.

Discussion

We have used TI simulations (32,33) to examine changes in the binding affinity of a Family 1 CBM with both amino acid mutations and native and artificial O-glycans. The well-characterized T. reesei Cel7A CBM, for which glycosylation has been quantified experimentally (23) and biochemical mutation data exist (13,14), was used as a model CBM. From these biochemical data, the computational approach was validated by demonstrating that we can achieve quantitative agreement in binding affinity changes for two large amino acid mutations, as shown in Figure 3. Specifically, we have shown that the Y5W mutation for the T. reesei Cel7A CBM improves the binding affinity by 2-fold, whereas the Y5A mutation is detrimental to binding as shown experimentally (13-15). By producing results consistent with experiments for this system, we have demonstrated that TI can potentially be a useful screening tool for mutations that modify CBM binding affinity.

Following these initial validation simulations, we extended our approach to predict the change in binding affinity for both a native and an artificial O-glycan on specific regions of the Cel7A CBM. We predict that a single native O-glycan near the posterior of the CBM interacts directly with cellulose and can change the binding affinity by 3-fold, and that the addition of a second, independent, engineered glycan combined with the native glycosylation can change the binding affinity by 140-fold, which is a striking increase over an amino acid mutation alone. The results of the glycan simulations reported here are a promising demonstration of the potential for engineering improved cellulases via the introduction of non-native glycosylation, considering that only a 2-fold increase in binding affinity for the Y5W mutation relative to the wild-type non-glycosylated Cel7A CBM resulted in higher activity for the enzyme with the mutant CBM (15). Since even small glycosylation motifs (a single mannose) can impact the binding affinity, the addition of artificial glycosylation sites via site-directed mutagenesis to either N-glycan or O-glycan motifs is a potentially powerful strategy to improve the specific activity of glycoside hydrolase enzymes.

Modification of culture growth conditions and expression hosts for glycoproteins is generally known to affect the glycosylation pattern of a given secreted protein (31). This previous finding has significant implications in light of the current study for comparing the binding affinity and activities of enzymes and enzyme cocktails for biomass conversion purposes. For example, if the expression host or growth conditions vary between protein cultures, changes can arise in experimental observables (e.g., binding and/or enzyme activity) from differences in the extent of glycosylation alone, which can alter the outcome of enzyme screening or directed evolution experiments.

Lastly, many groups are constructing quantitative, mesoscale models of cellulase action to predict enzyme synergy and similar phenomena with the aim to design enhanced cellulase cocktails for biomass conversion (34-36). These models rely on accurate thermodynamic and kinetic measurements of cellulase-cellulose interactions and insights from advanced experimental techniques and simulation predictions related to the molecular-level mechanisms of cellulase action (10,37-39). Measurement or simulations of the absolute or relative binding free energies (and
partition coefficients) of CBMs to cellulose should account for native glycosylation patterns to obtain accurate measurements or predictions, respectively.

An important question from this study is how both the natural and artificial glycans affect the structure and function of the CBM. We show in the Supplemental Information that the native glycosylation pattern (23) studied here stabilizes the CBM structure slightly, as demonstrated by the changes in hydrogen bonding, RMSD, and RSMF results. In terms of the CBM function, we note that it is commonly stated in the literature that aromatic amino acids on the binding faces of CBMs, like Tyr-5 in the Cel7A CBM, primarily interact with cellulose via hydrophobic interactions (13,14,22,24). However, there are two types of functions to consider when discussing binding affinity: first, an absolute binding affinity wherein the CBM binds to cellulose from solution, and second, the function of the CBM after it is bound, which is typically thought of as translation or processivity along the surface. Here we have examined the effect of the former (CBM binding from solution) with TI calculations. It has yet to be definitively shown if the binding effect of aromatic residues in Family 1 CBMs is primarily due to enthalpic or entropic contributions, which could be demonstrated with isothermal titration calorimetry. For translation along the surface, which is likely relevant to CBM function as part of an enzyme (40), we have shown in a previous study (21) that the aromatic residues (Tyr-5 and Tyr-31) and several polar residues form hydrogen bonds approximately every 1 nm (or one cellobiose unit) on the hydrophobic surface of cellulose. These residues are likely critical for the function of Family 1 CBMs, and the importance of these hydrogen bonds is demonstrated in that either Tyr or Trp is usually preferred in these sequence positions in Family 1 CBMs over Phe (21). In the case of adding mannose to Ser-3 or Ser-14 in the Cel7A CBM, the glycans could have a similar effect on CBM translation as the existing aromatic and polar residues, because mannose presents a large, planar face accompanied by the ability to form hydrogen bonds with primary alcohol groups on cellulose that are positioned uniformly along a given polymer chain (see the Supplemental Information). Therefore it is unlikely that the addition of mannose will significantly affect CBM translation once bound to cellulose, but rather serve as additional surface area for binding. This hypothesis is validated experimentally in that a single mannose exists on the CBM already in functional enzymes at Ser-3 (23). Moreover, it is unknown for a consistent set of CBMs and whole cellulases, either experimentally or computationally, if the CBM processivity rate differs from the processivity rate (i.e., hydrolysis rate) of an engaged enzyme. Recently, Igarashi et al. used high-speed atomic force microscopy to measure the rate of the T. reesei Cel7A enzyme acting on cellulose (38), but the diffusion coefficient of Family 1 CBMs on cellulose has not been explicitly measured to our knowledge. If the processivity rate of a CBM is much faster than the combined hydrolysis and processivity rate of the whole enzyme, which it likely is, addition of glycans should not affect the ability of the CBM to translate on a biologically relevant timescale. Thus it is unlikely that the CBM will get “stuck” such that CBM translation becomes the rate-limiting step in processive hydrolysis.

Finally, we note that experimental validation of these computational results is of paramount importance. Expression of the Family 1 CBM in hosts that do not impart glycosylation or production via solid-state synthesis as was conducted previously can yield a non-glycosylated CBM (13,14). Expression and purification of the CBM from T. reesei, or other expression hosts that impart glycosylation, or via chemical synthesis procedures in which glycosylation can be chemically added, could produce a CBM with the glycosylation patterns examined here and binding isotherms measured. However, from a computational standpoint, we note that carefully conducted TI simulations and free energy calculations in general for ligand binding have been shown to yield agreement with experiment within several kcal/mol (41-46). Here we have obtained results consistent with available experimental data on amino acid mutations, which at the least, lends confidence to the qualitative nature of the predictions regarding glycosylation.

Conclusions
In summary, our results indicate that CBM glycosylation is a likely contributor to enzyme binding affinity. To our knowledge, this study is the first to apply TI calculations to test the effects of glycosylation on a cellulase enzyme, which is a computational approach that has broad applicability as many cellulases contain both N- and O-linked glycosylation. For the Cel7A CBM, the addition of glycans increases hydrogen bonding potential and hydrophobic stacking with cellulose via glycoprotein-carbohydrate interactions, stabilizing the CBM on the surface and improving binding affinity. Our results highlight the need for consideration of post-translational modifications when selecting expression hosts and growth conditions for these types of enzymes (25,28,30,31). Glycosylation, or lack thereof, could have an impact on binding that can translate into effects on enzyme mechanistic action as a whole. Moreover, the manipulation of glycan sites via recombinant expression, by varying growth conditions, or via addition of artificial glycan sites could be used as a general protein engineering strategy to tune protein-carbohydrate binding affinity for improving cellulases.

**Simulation Details**

CHARMM (47) was used to build the hybrid protein structures and initial coordinate files from the original wild-type structure. NAMD (48) was used for all equilibrations and thermodynamic integration (TI) calculations and VMD (49) was used for visualization. The CHARMM27 force field with the CMAP correction (47,50,51) was used to describe the protein, while cellulose and the O-glycosylation were modeled using the CHARMM35 carbohydrate force field (52,53). Water was modeled using the modified TIP3P force field (54,55).

The cellulose Iβ crystal structure was used to generate the cellulose slab (56). The cellulose slab thickness, the CBM positioning above the surface, and overall dimensions, were taken from Beckham et al. (21). The mannose dimer is linked α-1,2 and all O links to Ser and Thr are in the α-configuration (27,57). The solvated, bound system contained approximately 18,100 atoms. Particle mesh Ewald (58) was used to describe the long-range electrostatic interactions with a sixth order b-spline interpolation, a Gaussian distribution with a width of 0.312 Å, and a mesh size of 60 x 60 x 45. A non-bonded interaction cutoff of 10 Å was used. The SHAKE algorithm (59) was employed to fix covalent bonds to hydrogen atoms. The CBM-cellulose system (referred to as the bound system), was minimized for 2,000 steps, and then equilibrated in the NVT ensemble at 300 K using a 2 fs timestep for 2 ns, at which time the RMSD of the protein backbone had stabilized. To calculate relative binding free energy, systems without cellulose (referred to as the free system), were also prepared. The wild-type CBM and mutated CBMs structures were solvated in CHARMM with approximately 4,000 water molecules and simulations performed under the same conditions as those with the cellulose surface. The equilibrated, final coordinates of each system, bound and free, were used as the starting coordinates for the TI simulations.

In NAMD, TI was performed using the dual-topology method (33), implemented by equilibrating a single structure with a hybrid residue containing both the wild-type and mutated atoms. The electrostatic and van der Waals calculations were decoupled, reducing computational effort and eliminating instabilities arising from large energy interactions (33). The electrostatic calculations comprised 11 equidistant λ windows from 0 to 1, each equilibrated for 0.5 ns before 10 ns TI NVT runs. Two additional windows, 0.05 and 0.95 were added to the glycan electrostatic cases to improve the precision of the results. Van der Waals calculations required more windows and longer equilibrations, especially near the endpoints of λ = 0 and 1, as well as longer overall run lengths to reduce statistical error. Additional windows were selected by examining the probability histograms of $dU/d\lambda$ at each λ value (33,60). Details of the steps taken to ensure simulation convergence and the error analysis performed following the methods of Steinbrecher et al. (60) and Paliwal and Shirts (61) are provided in the Supplemental Information.

Long MD runs were also performed for selected systems. In these simulations, the system size and simulation parameters were identical to those used in the TI studies. Each system was run for 100 ns in the NVT ensemble at 300 K with a 2 fs
timestep. The non-bonded interaction energy between the residues of interest and the surface was calculated in NAMD and the error for these values over the total run was determined using block averaging (62). Hydrogen bonding potentials between residues of interest and the surface were calculated in VMD for comparison to the interaction energy values. RMSF and RMSD of the backbone were also calculated and the typical movement and rotation of the glycan sugars during the simulation determined.

Acknowledgements
We thank the DOE Office of the Biomass Program for funding. We thank Peter Ciesielski, Christina Payne, Michael Resch, and Deanne Sammond for a critical reading of the manuscript and for helpful discussions. We thank the reviewers for helpful, detailed suggestions to improve the manuscript. Computational time for this research was provided using the TACC Ranger cluster and the NICS Athena and Kraken clusters under the NSF Teragrid grant number MCB090159, and resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supplemental Information
A description of the TI methodology, free energy curves from the Y5F simulation, additional methods used in post-analysis including method validation, error analysis, Y5F frequency histograms, autocorrelation data for Y5F TI output, interaction energies between the residues and cellulose surface, hydrogen bonding potentials between the Ser-3 and Ser-14 glycans and the cellulose, and RMSF and RMSD curves for the 100 ns simulations, are provided.

REFERENCES
1. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769-781
2. Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., and Knox, J. P. (2006) J. Biol. Chem. 281, 29321-29329
3. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 484-489
4. Horn, S. J., Sikorski, P., Cederkivist, J. B., Vaaje-Kolstad, G., Sorlie, M., Synstad, B., Vriend, G., Varum, K. M., and Eijsink, V. G. H. (2006) Proc. Natl. Acad. Sci. 103, 18089-18094
5. van Aalten, D. M. F., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G. H., and Wierenga, R. K. (2000) Proc. Natl. Acad. Sci. 97, 5842-5847
6. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Science 315, 804-807
7. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) Science 311, 484-489
8. Wilson, D. B. (2009) Curr. Opin. Biotech. 20, 295-299
9. Zhang, Y. H. P., Himmel, M. E., and Mielenz, J. R. (2006) Biotechnol. Adv. 24, 452-481
10. Chundawat, S. P. S., Beckham, G.T., Himmel, M.E., and Dale, B. E. (2011) Annu. Rev. Chem. Biomol. Eng. 2, 61-6.25
11. Lantz, S. E., Goedegebuur, F., Hommes, R., Kaper, T., Kelemen, B. R., Mitchinson, C., Wallace, L., Stahlberg, J., and Larenas, E. A. (2010) Biotechnol. Biofuels 3
12. Heinzelman, P., Snow, C. D., Smith, M. A., Yu, X. L., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F. H. (2009) J. Biol. Chem. 284, 26229-26233
13. Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg, G., Stahlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995) Protein Sci. 4, 1056-1064
14. Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. T., and Pettersson, G. (1995) FEBS Lett. 372, 96-98
15. Takashima, S., Ohno, M., Hidaka, M., Nakamura, A., and Masaki, H. (2007) FEBS Lett. 581, 5891-5896
16. Kim, T. W., Chokhawala, H. A., Nadler, D., Blanch, H. W., and Clark, D. S. Biotechnol. Bioeng. 107, 601-611
17. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonieni, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1999) Science 285, 524-528
18. Divne, C., Stahlberg, J., and Jones, T. A. (1998) J. Mol. Biol. 275, 309-325
19. Zhong, L. H., Matthews, J. F., Hansen, P. I., Crowley, M. F., Cleary, J. M., Walker, F. R., and Nimlos, M. R., Brooks, C. L., Adney, W. S., Himmel, M. E., and Brady, J. W. (2009) Carbohydr. Res. 344, 1984-1992
20. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) Nucleic Acids Res. 37, D233-D238
21. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L. T., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2009) J. Phys. Chem. B 114, 1447-1453
22. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Biochemistry 28, 7241-7257
23. Harrison, M. J., Nouwens, A. S., Jardine, D. R., Zachara, N. E., Gooley, A. A., Nevalainen, H., and Packer, N. H. (1998) Eur. J. Biochem. 256, 119-127
24. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12229-12234
25. Boraston, A. B., Warren, R. A. J., and Kilburn, D. G. (2001) Biocatalysis 18, 423-430
26. Boraston, A. B., Sandercoc, L. E., Warren, R. A. J., and Kilburn, D. G. (2003) J. Mol. Microbiol. Biotechnol. 5, 29-36
27. Deshpande, N., Wilkins, M. R., Packer, N., and Nevalainen, H. (2008) Glycobiology 18, 626-637
28. Adney, W. S., Jeoh, T., Beckham, G. T., and Adney, W. S., Baker, J. O., Michener, W., Brunecky, R., and Himmel, M. E. (2009) Cellulose 16, 699-709
29. Beckham, G. T., Bomble, Y. J., Matthews, J. F., Taylor, C. B., Resch, M. G., Yarbrough, J. M., Decker, S. R., Bu, L. T., Zhao, X. C., McCabe, C., Wohler, J., Bergenstrahle, M., Brady, J. W., Adney, W. S., Himmel, M. E., and Crowley, M. F. (2010) Biophys. J. 99, 3773-3781
30. Jeoh, T., Michener, W., Himmel, M. E., Decker, S. R., and Adney, W. S. (2008) Biotechnol. Biofuels 1, 12
31. Stals, I., Sandra, K., Geysens, S., Contreras, R., Van Beeumen, J., and Claeyssens, M. (2004) Glycobiology 14, 713-724
32. Kollman, P. (1993) Chem. Rev. 93, 2395-2417
33. Pohorille, A., Jarzynski, C., & Chipot, C. (2010) J. Phys. Chem. B, 10235-10253
34. Bansal, P., Hall, M., Reall, M. J., Lee, J. H., and Bommarius, A. S. (2009) Biotechnol. Adv. 27, 833-848
35. Levine, S. E., Fox, J. M., Blanch, H. W., and Clark, D. S. Biotechnol. Bioeng. 107, 37-51
36. Zhou, W., Hao, Z. Q., Xu, Y., and Schuttler, H. B. (2009) Biotechnol. Bioeng. 104, 275-289
37. Beckham, G. T., Bomble, Y. J., Bayer, E. A., Himmel, M. E., and Crowley, M. F. (2011) Current Opinion in Biotechnology 22, 231-238
38. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) J. Biol. Chem. 284, 36186-36190
39. Zhao, X., Rignall, T. R., McCabe, C., Adney, W. S., and Himmel, M. E. (2008) Chem. Phys. Lett. 460, 284-288
40. Bu, L. T., Beckham, G. T., Crowley, M. F., Chang, C. H., Matthews, J. F., Bomble, Y. J., Adney, W. S., Himmel, M. E., and Nimlos, M. R. (2009) J. Phys. Chem. B 113, 10994-11002
41. Mobley, D. L., Chodera, J. D., and Dill, K. A. (2007) Journal of Chemical Theory and Computation 3, 1231-1235
42. Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A. (2007) J. Mol. Biol. 371, 1118-1134
43. Deng, Y., and Roux, B. (2009) J. Phys. Chem. B 113, 2234-2246
44. Wang, J., Deng, Y., and Roux, B. (2006) Biophys. J. 91, 2798-2814
Figure 1. The catalytically-active complex of the *T. reesei* Family 7 cellobiohydrolase (light purple) on cellulose (green) (17-19). The native O-glycosylation is shown in yellow and the native N-glycans are shown in dark blue. The CBM is the small protein domain on the left containing 2 native O-glycans, and the catalytic domain is the large protein domain on the right with a cellulose chain threaded into the tunnel.

Figure 2: Side view (A) and top view (B) of the Cel7A CBM and the top layer of the cellulose slab. The tyrosine residues are shown in yellow. The native O-glycans considered here are shown on Thr-1 and Ser-3. The artificial glycan examined here is shown on Ser-14.

Figure 3: Langmuir isotherms of bound CBM as a function of free CBM as measured by Linder et al. (symbols) (13,14) and predicted by the free energy simulations conducted in this work (lines). Our results indicate improvement in binding affinity over that of just amino acid mutation with the incorporation of a single or dimer O-mannose at S3 (3 to 6-fold improvement in binding affinity) and a synthetic glycan site
at Ser-14 with and without the native glycans (20 to 140-fold improvement in binding affinity, respectively).

**Figure 4:** Comparison of the total interaction energy (Total Energy = Electrostatic + van der Waals) between the glycans and cellulose surface present in the S3M1, S14M1-NG, S3M2, and S14M1 100 ns MD simulations. The Thr-1 mannose has zero interaction energy with the surface in each simulation.

**Table 1:** Relative binding free energy ($\Delta\Delta G_{TOT}$, kcal/mol) from amino acid TI calculations, including the two containing native glycosylation (Y5A-G and Y5W-G), and associated change in partition coefficient ($K_{Mut}/K_{WT-NG}$). The S3M1+Y5A-G and S3M1+Y5W-G entries are the sum of the Y5A-G and Y5W-G (data not shown) and S3M1 (see Table 2) entries, respectively. Errors were calculated as described in the Supplemental Information.

| Mutation    | Y5A   | Y5F   | Y5W   | S3M1 + Y5A-G | S3M1 + Y5W-G |
|-------------|-------|-------|-------|--------------|--------------|
| $\Delta\Delta G_{TOT}$ (kcal/mol) | 2.6 ± 0.14 | 0.32 ± 0.04 | -0.40 ± 0.14 | 1.5 ± 0.05 | -1.7 ± 0.07 |
| $K_{Mut}/K_{WT-NG}$ | 0.01 ± 0.00 | 0.59 ± 0.04 | 1.9 ± 0.45 | 0.08 ± 0.01 | 16 ± 1.8 |

**Table 2:** Relative binding free energy ($\Delta\Delta G_{TOT}$, kcal/mol) from the native glycosylation TI calculations, for a single mannose at Ser-3 (S3M1) and a dimer mannose at Ser-3 (S3M2), and associated change in partition coefficient ($K_{Mut}/K_{WT-NG}$). The partition coefficient change was not calculated for the intermediate step, S3M2. The Thr-1 mannose does not impact binding affinity directly, so it was not considered part of the thermodynamic cycle. In the schematics, the green circles represent wild-type mannoses present in the reactant and product state, and the red circles represent the mannose(s) added to Ser-3 as the product of the thermodynamic cycle for the S3M1 and S3M2 cases. The S3M1+S3M2 entry is the sum of the previous two entries. Errors were calculated as described in the Supplemental Information.

| Mutation    | S3M1  | S3M2  | S3M1 + S3M2 |
|-------------|-------|-------|-------------|
| $\Delta\Delta G_{TOT}$ (kcal/mol) | -0.75 ± 0.03 | -0.42 ± 0.03 | -1.2 ± 0.04 |
| $K_{Mut}/K_{WT-NG}$ | 3.5 ± 0.17 | -- | 6.9 ± 0.49 |
Table 3: Relative binding free energy (ΔΔG_{TOT}, kcal/mol) from the engineered glycosylation TI calculations, with a single mannose at Ser-14 (S14M1-NG) with no glycosylation and a single mannose at Ser-14 with the native glycans present (S14M1), and associated change in partition coefficient (K_{Mut}/K_{WT-NG}). The partition coefficient change was not calculated for the intermediate step, S14M1. The Thr-1 mannose does not impact binding affinity directly, so it was not part of the mutation. In the schematics, the green circles represent wild-type mannoses present in the reactant and product state, and the red circles represent the mannose mutations at Ser-14 and Ser-3 for the S14M1-NG and S14M1 cases. The S3M1+S14M1 entry is the sum of the S14M1 and S3M1 (see Table 2) entries. Errors were calculated as described in the Supplemental Information.

| Mutation          | S14M1-NG | S14M1 | S3M1 + S14M1 |
|-------------------|----------|-------|--------------|
| ΔΔG_{TOT} (kcal/mol) | -1.9 ± 0.31 | -2.3 ± 0.09 | -3.0 ± 0.10 |
| K_{Mut}/K_{WT-NG}   | 23 ± 12  | --    | 149 ± 24     |
Figure 3

Figure 4
Computational investigation of glycosylation effects on a Family 1 carbohydrate-binding module
Courtney B. Taylor, M. Faiz Talib, Clare McCabe, Lintao Bu, William S. Adney, Michael E. Himmel, Michael F. Crowley and Gregg T. Beckham

J. Biol. Chem. published online December 6, 2011

Access the most updated version of this article at doi: 10.1074/jbc.M111.270389

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/12/06/M111.270389.DC1