

# A highly conserved ABC transporter mediates cello-oligosaccharide uptake in the extremely thermophilic, lignocellulolytic bacterium *Anaerocellum bescii* (f. *Caldicellulosiruptor bescii*)

Hansen Tjo,<sup>1</sup> Virginia Jiang,<sup>1</sup> Anherutowa Calvo,<sup>1</sup> Jerelle A. Joseph,<sup>1,2,3,4,5</sup> Jonathan M. Conway<sup>1,2,3,6,7</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 17.

**ABSTRACT** Cellulose deconstruction and utilization are foundational to renewable biofuel and biochemical production. *Anaerocellum bescii* (formerly *Caldicellulosiruptor bescii*) is an extremely thermophilic cellulolytic bacterium, notable for its multi-domain cellulases and hemicellulases that efficiently degrade lignocellulosic biomass. However, the mechanisms by which *A. bescii* transports cello-oligosaccharides released during cellulose degradation into the cell for catabolism remain unclear. Among its 23 ATP-binding cassette (ABC) sugar transporters, we identified a conserved ABC transporter locus (*athe\_0595-0598*) encoding two extracellular binding proteins: Athe\_0597 and Athe\_0598. Biophysical analyses using differential scanning calorimetry and isothermal titration calorimetry revealed that Athe\_0597 binds cello-oligosaccharides of varying lengths (G2-G5), while Athe\_0598 is specific to cellobiose (G2). Ligand docking simulations supported these findings and shed light on the subsite configuration of these substrate-binding proteins (SBPs). To assess its physiological importance, we genetically deleted this transporter locus in *A. bescii* strain HTAB187, which grew poorly on cellobiose and did not grow on cellulose. Comparison of growth with a *msmK* deletion strain that cannot consume oligosaccharides showed that HTAB187 retains growth on non-cello-oligosaccharides and monosaccharides. Taken together, these results integrate biophysical characterization, structural modeling, and genetic perturbation to elucidate how *A. bescii* transports cello-oligosaccharides released from cellulose, providing mechanistic insight relevant to consolidated bioprocessing applications.

**IMPORTANCE** *Anaerocellum bescii* is the most thermophilic lignocellulolytic bacterium known and holds potential for bioprocessing lignocellulosic biomass into renewable fuels. Its diverse ATP-binding cassette (ABC) sugar transporters make it a valuable model for studying thermophilic sugar uptake. Here, we identify a single ABC transporter with two substrate-binding proteins (Athe\_0597 and Athe\_0598) responsible for cello-oligosaccharide uptake. Genetic deletion of this transporter locus impaired growth on cellobiose and eliminated growth on cellulose. This is the first genetic manipulation in *A. bescii* to modulate transport of a specific sugar. We also characterize the substrate specificity of the extracytoplasmic binding proteins associated with the locus. One binds various cellodextrins (G2-G5), while the other specifically binds cellobiose (G2). Molecular modeling depicts how each oligosaccharide is docked within the binding pocket of these proteins. Understanding the mechanism of cello-oligosaccharide uptake by *A. bescii* expands opportunities for its metabolic engineering and furthers our understanding of its carbohydrate utilization systems.

**Editor** Nicole R. Buan, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

Address correspondence to Jonathan M. Conway, [jmconway@princeton.edu](mailto:jmconway@princeton.edu).

The authors declare no conflict of interest.

See the funding table on p. 17.

**Received** 26 June 2025

**Accepted** 24 November 2025

**Published** 18 December 2025

Copyright © 2025 Tjo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

**KEYWORDS** thermophile, biotechnology, cellulose, cello-oligosaccharide, ABC sugar transporter, substrate-binding protein, *Caldicellulosiruptor*, *Anaerocellum bescii*, genetics

Bio-based fuels and chemicals derived from plant biomass are an attractive and renewable replacement for petroleum-based supply chains (1, 2). However, for decades, technological progress has been stalled by the physical and chemical recalcitrance of lignocellulose, impeding deconstruction and conversion at the large scales needed to compete with comparatively cheaper petroleum (3, 4). Lignocellulolytic thermophiles, with their native ability to overcome lignocellulose recalcitrance, have therefore garnered significant interest, as reflected in recent advances in their genetic and metabolic engineering (5–7). Fully realizing their potential will require a more comprehensive understanding of their physiology and of their native ability to utilize complex carbohydrate substrates (8, 9).

*Anaerocellum bescii* is the most thermophilic, lignocellulolytic bacterium known, with an optimal growth temperature of 75°C (10–12). It possesses vital attributes for consolidated bioprocessing of lignocellulosic feedstocks: efficiently degrading cellulose and hemicellulose into oligosaccharides while simultaneously metabolizing pentose and hexose sugars (13). Multiple genetic tools, including methods for chromosomal modification, an antibiotic-based selection marker, and a xylose-inducible promoter, have been developed over the past decade (14, 15). These tools have enabled metabolic engineering in *A. bescii* to produce a range of valuable products, including ethanol, acetone, and 2,3-butanediol (2,3-BDO), at elevated yields (6, 16–18).

*A. bescii* naturally secretes a suite of multi-domain carbohydrate-active enzymes (CAZymes) for plant biomass deconstruction (19). Its unique glucan degradation locus (GDL) encodes six multi-domain glycoside hydrolases (GHs) with catalytic activity on a broad range of plant biomass substrates, including cellulose, xylans, and mannans (20, 21). Among these, CelA (Athe\_1867), CelC (Athe\_1857), and CelE (Athe\_1865) show particularly strong synergistic activity on cellulose (21). CelA is among the most effective bacterial cellulases known and one of the most abundant proteins in the *A. bescii* secretome (22, 23). It contains an N-terminal GH9 endoglucanase domain, followed by three carbohydrate-binding module family 3 (CBM3) domains for binding cellulose, and a C-terminal GH48 domain exhibiting exo- $\beta$ -1,4-glucanase activity. Synergistic action between the GH9 and GH48 enables CelA to excavate pits in cellulose microfibrils, exposing interior fibrils, with the GH9 generating new chain ends for cleavage by the GH48, primarily into cellobiose (22–24). Inter-enzyme synergy of CelA with other GDL CAZymes, particularly the GH10-GH48 domains in CelC and GH9-GH5 domains in CelE, further enhances cellulose degradation (21–23). Beyond the GDL, the *A. bescii* secretome also includes auxiliary catalytic and non-catalytic proteins that aid in the deconstruction and uptake of plant biomass polysaccharides (21, 25–27). But despite considerable advances in our biochemical understanding of *A. bescii* plant biomass deconstruction, comparatively little is known about how this bacterium imports the released oligosaccharides into the cell for metabolism (8).

To facilitate carbohydrate uptake, *A. bescii* utilizes ATP-binding cassette (ABC) sugar transporters that hydrolyze two molecules of ATP for the translocation of one sugar molecule into the cytoplasm (13, 28, 29). Twenty-three such ABC transporters have been annotated in the bacterium (13, 30). In addition, the genome encodes a phosphotransferase system (PTS) transporter, which is thought to be fructose specific (30). ABC transporters comprise a hydrophobic transmembrane protein domain, a cytoplasmic ATPase domain, and an extracytoplasmic substrate-binding domain (28, 29). This substrate-binding domain, also referred to as a substrate-binding protein (SBP), is responsible for binding extracellular sugars and delivering them to the transmembrane domain and, consequently, is indicative of the transporter's carbohydrate specificity (8, 31, 32). There is also genetic evidence to suggest that MsmK, a promiscuous ATPase encoded by *athe\_1803*, is responsible for powering all oligosaccharide ABC sugar transporters in *A. bescii* (13). Though a few transcriptomic studies have attempted to

predict the specificity of the ABC sugar transporters in *A. bescii*, only those for the maltodextrin system have been experimentally validated thus far (13, 30, 33). No studies have examined how *A. bescii* uptakes oligosaccharides derived from cellulose, a major growth substrate (25).

In this study, we report the first biochemical investigation of cello-oligosaccharide transport in *A. bescii*. We show that two SBPs in *A. bescii*—Athe\_0597 and Athe\_0598—are specific to cello-oligosaccharides. We demonstrate that these proteins bind cello-oligosaccharides at characteristically high binding affinities, with *in silico* models that structurally elucidate how these sugars are bound. We also genetically deleted both the substrate-binding and transmembrane domains comprising the cello-oligosaccharide ABC transporter locus (*athe\_0595–0598*) in *A. bescii*, which resulted in reduced growth on the disaccharide cellobiose and complete loss of growth on cellulose, demonstrating the critical role of this locus in cellulose utilization. These insights reveal essential features of sugar transport in lignocellulolytic *A. bescii* and open doors to engineering strains with novel modes of plant biomass utilization.

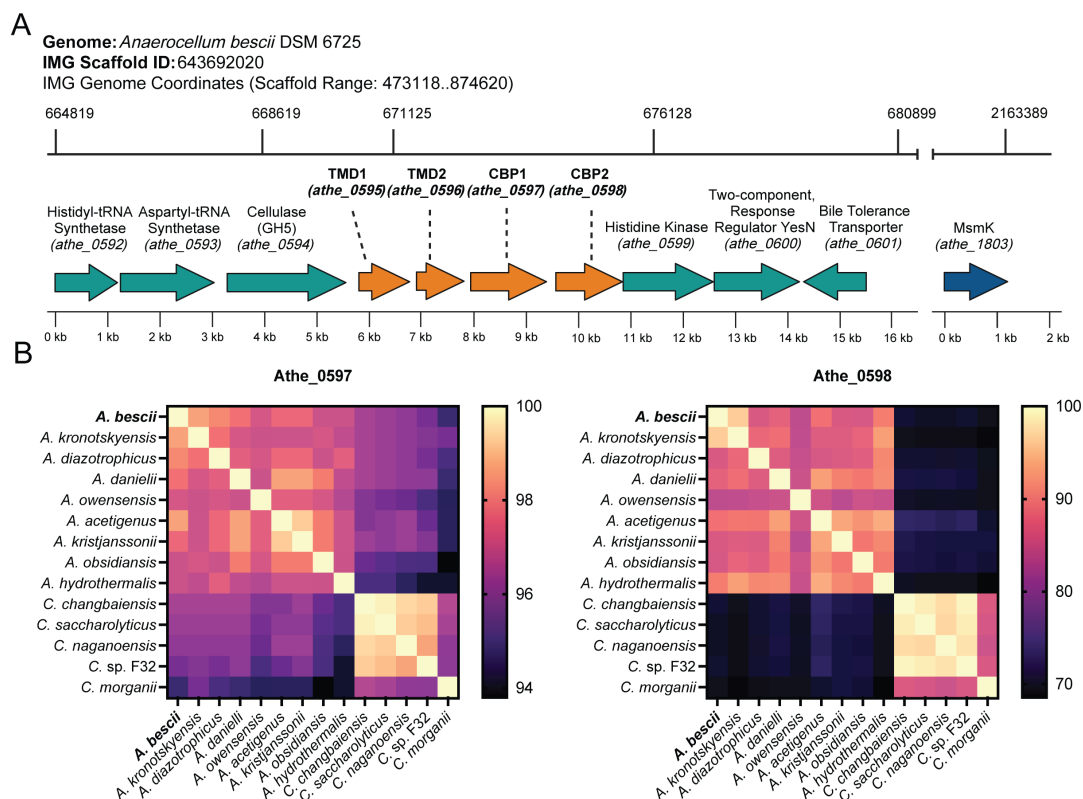
## RESULTS

### Identification of the cello-oligosaccharide ABC transport system in *A. bescii*

*A. bescii* contains a gene cluster that encodes an ABC transporter (*athe\_0595–0598*) predicted to mediate cello-oligosaccharide transport (Fig. 1A) based on transcriptomic work by VanFossen et al. (2009) in *Caldicellulosiruptor saccharolyticus* and Rodionov et al. (2021) in *A. bescii* (13, 30). Rodionov et al. (2021) further predicted that *athe\_0597* and *athe\_0598* encode for ABC SBPs with specificity toward oligosaccharides (13, 28). Notably, their expression was upregulated when *A. bescii* was grown on cellulose (13). ABC transporters with multiple SBPs have been documented in other bacteria (34, 35). However, to date, both transporter substrate predictions from Vanfossen et al. (2009) and Rodionov et al. (2021) have not been experimentally validated (13, 30).

The genomic context of the *athe\_0595–0598* locus suggests that this ABC transporter plays a role in cello-oligosaccharide uptake. Both genes are co-localized with *athe\_0594*, which encodes a cell surface localized GH5-CBM28 endoglucanase with activity on  $\beta$ -glucans from barley, lichenan, and cellulose (20, 36, 37). While Athe\_0594 potentially generates cello-oligosaccharide products from endo activity on cellulose, it is thought to play a more minor role in cellulose hydrolysis compared to the GDL enzymes, particularly Cella (20, 21). The *athe\_0594–0598* locus was found to be upregulated over 10-fold in *A. bescii* strains grown on cellulose and cellobiose growth substrates (13). Because no ATPase gene is encoded nearby and deletion of the promiscuous *msmK* ATPase (*athe\_1803*) disrupts growth on oligosaccharides, including cello-oligosaccharides, this indicates that the Athe\_0595–0598 cello-oligosaccharide ABC transporter is powered by MsmK (13).

The *athe\_0595–0598* ABC transporter gene cluster appears widely conserved across the *Anaerocellum* and *Caldicellulosiruptor* genera, though perhaps to a lesser extent in the latter (the *Anaerocellum* genus was recently split from the *Caldicellulosiruptor* genus) (Fig. 1B) (12). All *Anaerocellum* and *Caldicellulosiruptor* species possess genes encoding the heterodimeric transmembrane domain, as well as genes encoding for the two separate SBPs. The SBP Athe\_0597 appears much more highly conserved across both genera. Its *C. morgani* homolog, being the most distinct in amino acid sequence, still retains a relatively high amino acid sequence identity of 94.8% compared to Athe\_0597. Conversely, Athe\_0598 homologs exhibit higher amino acid sequence variability, particularly among *Caldicellulosiruptor* species as in the cases of *C. saccharolyticus*, *C. changbaiensis*, *C. naganoensis*, *C. sp. F32* and *C. morgani*. Yet even within the *Anaerocellum* genus, *A. owensensis* has the lowest amino acid sequence identity at 83.6%.

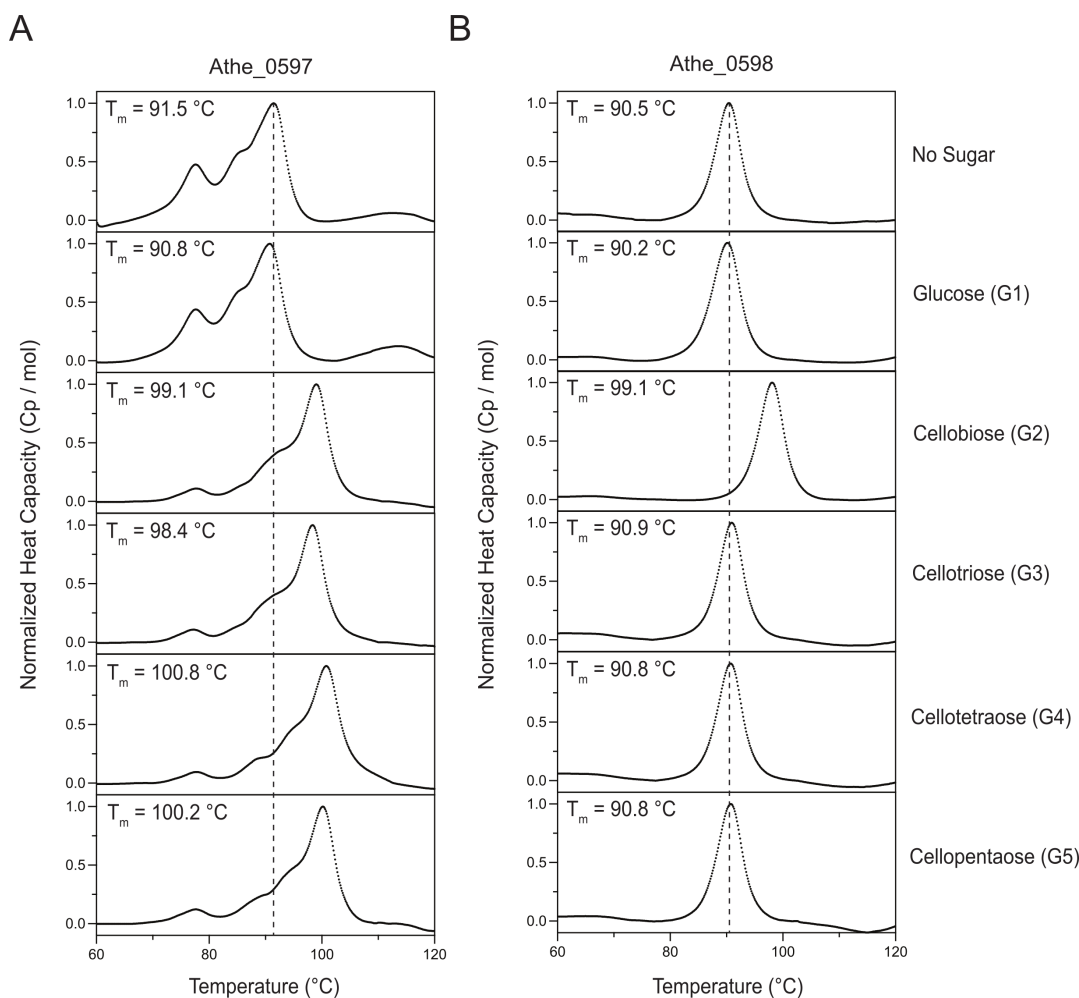


**FIG 1** (A) Genomic organization and coordinates of the *Anaerocellum besicii* cello-oligosaccharide ATP-binding cassette (ABC) transporter and its neighboring genes. The promiscuous ATPase MsmK, encoded by *athe\_1803*, is not a neighboring gene but is also shown due to its involvement in ABC transport of cello-oligosaccharides. TMD = transmembrane domain; CBP = cello-oligosaccharide-binding protein. (B) BLAST analysis of protein sequence conservation of cello-oligosaccharide binding proteins Athe\_0597 and Athe\_0598 across the *Anaerocellum* and *Caldicellulosiruptor* genus.

## Biophysical determination of cello-oligosaccharide specificity

Differential scanning calorimetry (DSC) is useful for screening the carbohydrate specificity of highly thermostable SBPs (33, 38). DSC can reach temperatures as high as 130°C, necessary for denaturing *A. besicii* SBPs that typically possess native melting temperatures above 92°C, the limit of traditional circular dichroism instruments (33). In DSC, a constant rate of heating raises the heat capacity of the protein until it fully denatures. The temperature corresponding to the maximum heat capacity of the protein, implying complete denaturation, is the native melting temperature  $T_m$  of the SBP in its *apo* state. When bound to its cognate substrates, however, these SBPs acquire a closed conformation with higher thermal stability—a process referred to as the “Venus flytrap” mechanism (8, 31). The difference in melting temperatures across the *apo* and *holo* states, defined as  $|\Delta T_m| = T_{m,holo} - T_{m,apo}$ , denotes the sugar specificity of a given SBP. Sugars with higher affinity will yield greater  $\Delta T_m$  values (33).

DSC melting curves showed that Athe\_0597 has a strong affinity for cello-oligosaccharide substrates (G2 – G5). In its *apo* state, Athe\_0597 exhibited a melting curve with a lower peak at  $T_{m,1} = 77.6^\circ\text{C}$  and a higher peak at  $T_{m,2} = 91.5^\circ\text{C}$  (Fig. 2). The higher peak at  $T_{m,2} = 91.5^\circ\text{C}$  is the presumptive melting temperature, but it is possible that in the absence of a bound ligand, a subunit begins denaturing at  $T_{m,1} = 77.6^\circ\text{C}$  (Table 1). The melting curve of Athe\_0597 resolves as a single peak in the presence of cognate sugar ligands: cellobiose (G2), cellotriose (G3), cellotetraose (G4), and cellopentaose (G5) with increased melting temperatures of 98.4°C to 100.8°C (Table 1). Curiously, melting temperature shifts are relatively similar for all cello-oligosaccharides that lead to substrate binding ( $\Delta T_m = 6.89 - 9.30^\circ\text{C}$ ). The rank ordering of melting temperature increases also does not neatly correlate with oligosaccharide size. This contrasts the



**FIG 2** Normalized differential scanning calorimetry (DSC) screens of (A) Athe\_0597 and (B) Athe\_0598, mixed with glucose and cello-oligosaccharides of different lengths: cellobiose (G2), cellotriose (G3), cellotetraose (G4), and cellopentaose (G5). Apo melting temperatures for Athe\_0597 and Athe\_0598 are denoted by a vertical dashed line, respectively. Athe\_0597 is shown to bind all tested cello-oligosaccharides from cellobiose to cellopentaose, whereas Athe\_0598 only binds cellobiose. All DSC screens were performed at a temperature range of 60–130°C, in 50 mM HEPES and 300 mM NaCl, pH 7.0 buffer.

larger differences in  $\Delta T_m$  values observed for the binding of *A. besicii* maltodextrin-binding protein Athe\_2574 with maltodextrins of various lengths, from maltose ( $\Delta T_m = 5.03^\circ\text{C}$ ) to maltoheptaose ( $\Delta T_m > 12^\circ\text{C}$ ) (Table 1) (33). Athe\_0597 also showed no changes in its melting profile relative to its apo state in the presence of glucose, demonstrating its lack of affinity for the monosaccharide.

Athe\_0598 showed uniform melting even in its unliganded apo state (Fig. 2). Binding to cellobiose (G2) did not appreciably alter the melting curve profile compared to the apo form. Athe\_0598 only exhibited affinity toward cellobiose as a substrate. Larger cello-oligosaccharide substrates from cellotriose to cellopentaose, as well as the monosaccharide glucose, did not result in any  $T_m$  shifts, suggesting that none of these ligands were able to bind Athe\_0598 (Fig. 2).

Next, ITC experiments were performed to probe the binding interactions between the two SBPs and their cognate cello-oligosaccharide ligands. Figure 3 illustrates the binding isotherms and isothermal titration calorimetry (ITC) curves for all sugar–protein pairs that were found to bind according to the DSC, with key measurements including dissociation constants  $K_d$  and stoichiometry ( $n$ ) summarized in Table 1. ITC data reinforced the results from the DSC, showing that Athe\_0597 indeed binds cellobiose, cellotriose, cellotetraose, and cellopentaose with micromolar dissociation constants typical of high-affinity

**TABLE 1** Biophysical and thermodynamic parameters of binding between substrate-binding proteins (SBPs) and celloextrins of various lengths as determined by DSC and ITC<sup>a</sup>

Protein	Sugar	$T_m$ (°C)	$\Delta T_m$ (°C)	$n$	$K_d$ ( $\mu\text{M}$ )	$K_a$ ( $\times 10^5 \text{ M}^{-1}$ )	$\Delta H$ (kcal/mol)	$T \times \Delta S$ (kcal/mol)	$\Delta G$ (kcal/mol)
Athe_0597	Glucose	90.79	-0.70	-	-	-	-	-	-
	CelloG2	99.06	7.57	0.95±0.02	0.228±0.10	43.86	3.33±0.13	12.4	-9.07
	CelloG3	98.38	6.89	1.13±0.01	0.942±0.12	10.62	3.94±0.08	12.2	-8.22
	CelloG4	100.79	9.30	0.95±0.03	1.97±0.52	5.08	6.04±0.32	13.8	-7.79
	CelloG5	100.21	8.72	1.02±0.08	1.18±0.12	8.47	4.61±0.08	12.7	-8.09
Athe_0598	Glucose	90.17	-0.36	-	-	-	-	-	-
	CelloG2	99.05	8.53	2.13±0.24	2.84±0.36	3.52	27.2±10.9	34.7	-7.57
	CelloG3	90.91	0.39	-	-	-	-	-	-
	CelloG4	90.81	0.28	-	-	-	-	-	-
	CelloG5	90.79	0.26	-	-	-	-	-	-

<sup>a</sup>SBP-sugar combinations denoted with a “-” were not tested via the ITC as DSC measurements indicated either weak or absent binding. All measurements were performed at 25°C.

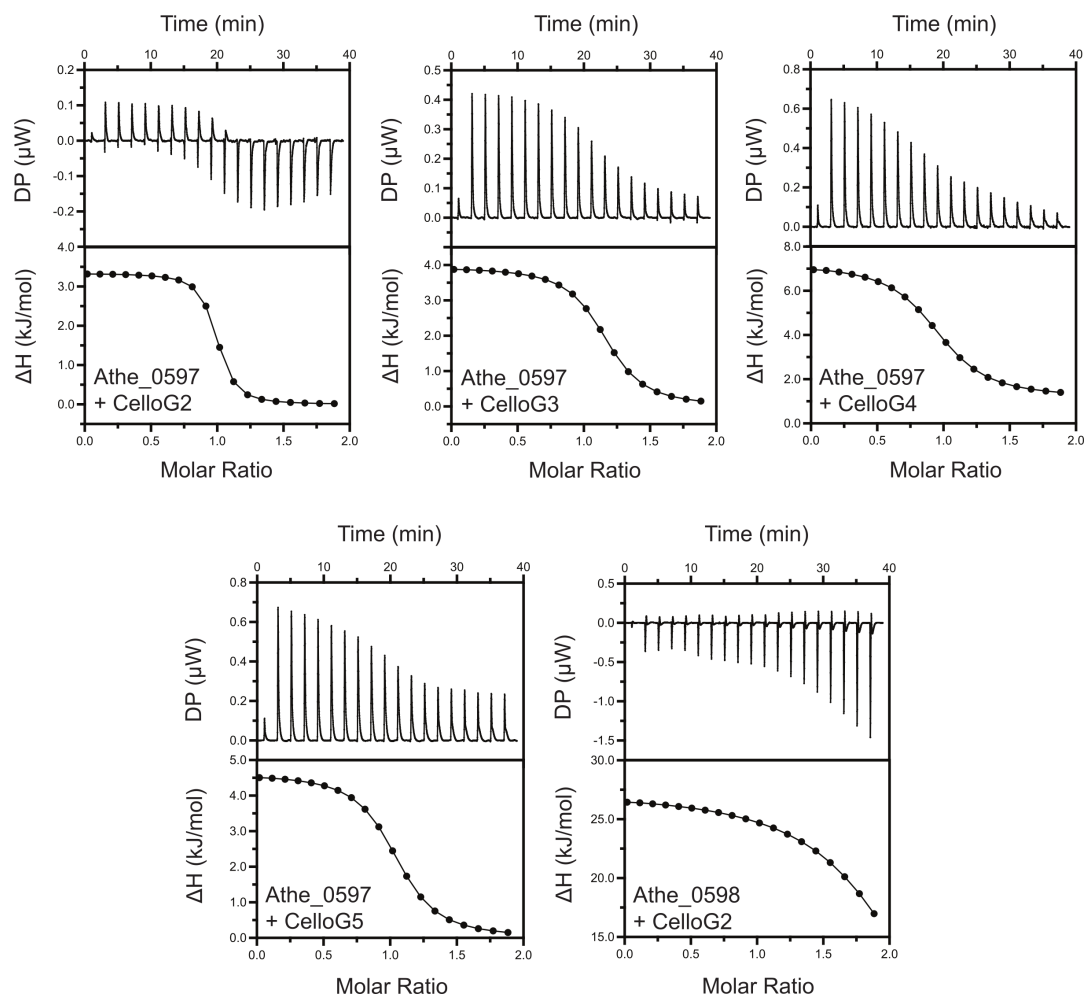
substrates in ABC sugar transporter systems (39, 40). Closer inspection of the association constant  $K_a$  values reveals that Athe\_0597 has the highest affinity for cellobiose (Table 1). The  $K_a$  for cellobiose is one order of magnitude higher than for all other cello-oligosaccharides, including cellotriose after taking into account the observed error (Table 1). The endothermic nature of binding between Athe\_0597 and its cello-oligosaccharides is consistent with other known cello-oligosaccharide-binding proteins (39, 40). The binding of both Athe\_0597 and Athe\_0598 to cello-oligosaccharide substrates also appears to be consistently entropically driven, with  $T\Delta S$  values significantly exceeding  $\Delta H$  values in all measured SBP and cello-oligosaccharide combinations (Table 1).

### Docking simulations elucidate structural context to cello-oligosaccharide binding

Computational models of protein-carbohydrate interactions showed that Athe\_0597 can bind cellobiose, cellotriose, cellotetraose, and cellopentaose, while Athe\_0598 primarily favors binding to cellobiose (Fig. 4C; Table S1). These conclusions largely match results obtained through experiments. The modeled Athe\_0598 incorrectly retains some affinity for cellotriose, as the closest-match homologs with crystal structures used as templates for the model have affinity for larger oligosaccharides (41). Simulations revealed that each SBP possesses a ligand-binding pocket composed of subsites, where glucosyl monomers can be docked (Fig. 4). Favorable binding free energies are driven by entropic contributions, such as the displacement of water molecules from the binding pocket, and by the formation of specific enthalpic interactions between each glucosyl monomer and the protein at each subsite.

The rank ordering for substrate-binding affinity, as measured by ligand docking for Athe\_0597, aligns with the rank ordering indicated by DSC, where cellotetraose is the most favored substrate (Fig. 4C). The non-reducing end of the cellotetraose is bound by subsite 1, with hydrogen bonds between R217 and R323. An internal glucosyl monomer 2 occupies subsite 2 and forms a hydrogen bond with Y80, while a third monomer occupies subsite 3. The reducing end of cellotetraose is bound by subsite 4 through hydrogen bonds between K286, E301, and S305. As shown through ITC, entropic contributions primarily drive substrate affinity (Table 1). Each subsite contains hydrophobic amino acids: W222 and G327 in subsite 1; A100, F98, and W393 in subsite 2; I44, M394, and W77 in subsite 3; and Y211 in subsite 4 (Fig. 4D). When cellobiose is docked in the binding pocket, the disaccharide is stabilized by interactions and residues similar to those that make up subsites 3 and 4 (Fig. 4E).

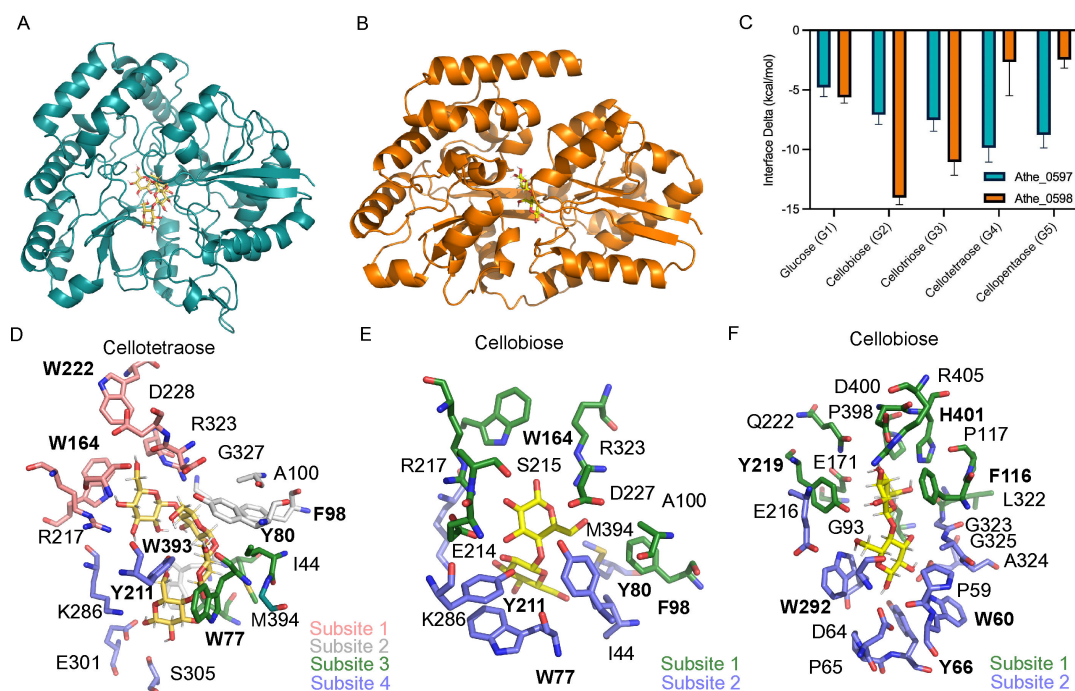
In the Athe\_0598 case, there were fewer available subsites for forming enthalpic interactions and fewer hydrophobic residues in the binding pocket contributing to desolvation entropy. Athe\_0598 had the most favorable binding interface energy with



**FIG 3** Representative isothermal titration calorimetry (ITC) screens of all SBP and cello-oligosaccharide pairings that were successful from DSC (Fig. 2). These combinations include Athe\_0597 with cellobiose (G2), cellotriose (G3), cellotetraose (G4), and cellopentaose (G5), as well as Athe\_0598 with cellobiose (G2). Both raw isothermal titration curves and integrated binding isotherms are shown for each SBP and cello-oligosaccharide pairing.

the disaccharide cellobiose, as only two subsites are available (Fig. 4F). The non-reducing end of cellobiose is bound by subsite 1, where D400, H401, and E171 form hydrogen bonds and I169, L322, P117, P398, and F116 form hydrophobic contacts. The reducing end occupies subsite 2, where E167, Y219, and D64 form hydrogen bonds and W60, Y66, and W292 form hydrophobic contacts (Fig. 4F). Moreover, the selective binding of cellobiose to Athe\_0598 may be due to the limited availability of subsites that could form possible enthalpic interactions, or the lack of available hydrophobic patches for entropic desolvation.

Broadly, our computational analysis suggests a possible mechanism for the entropically driven binding of cello-oligosaccharides to Athe\_0597 and Athe\_0598. This binding mode is driven by the entropically favorable liberation of solvent molecules when the docked sugar interacts with the hydrophobic aromatic residues located in the hinge region of the proteins. Entropically driven binding is consistent with previous studies showing that protein interactions often include a significant hydrophobic component (42). In line with our findings, our docking models—and those from other studies—indicate that aromatic residues form a hydrophobic cleft around the sugar ring faces.



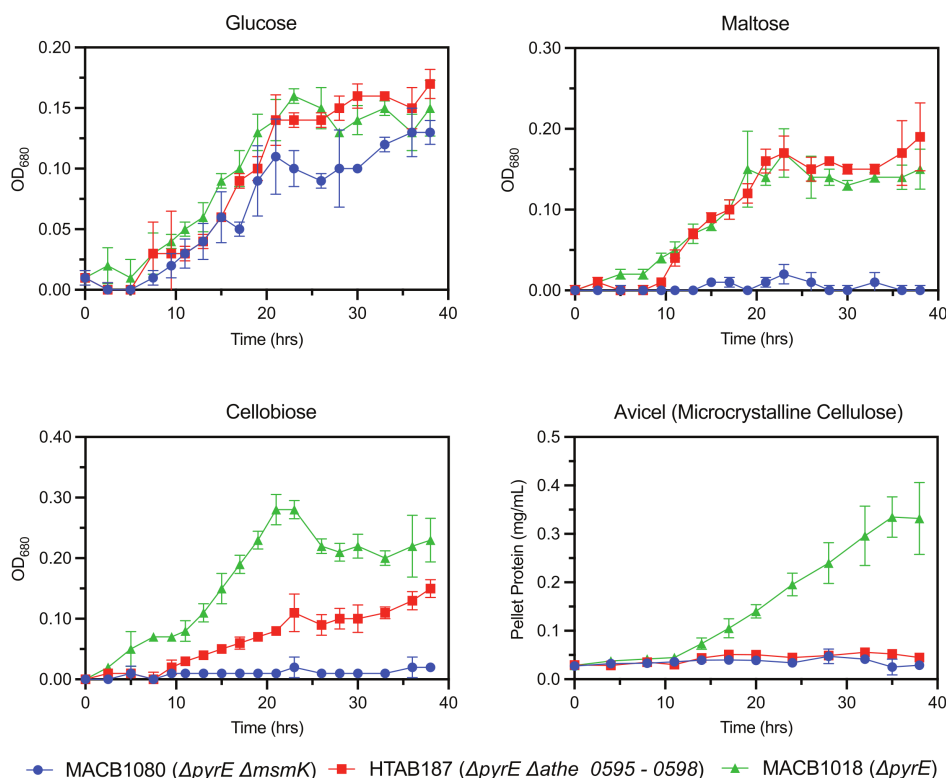
**FIG 4** (A) Computational model of *Athe\_0597* as visualized on PyMol. (B) Computational model of *Athe\_0598* as visualized on PyMol. (C) Interface energy deltas, as a proxy for free energies of binding, are shown for each simulated combination of cello-oligosaccharide and SBP. The rank-order affinity of *Athe\_0597* and *Athe\_0598* with their respective cello-oligosaccharide substrates is aligned with experimental results. (D) Binding of the highest affinity ligand to *Athe\_0597*, cellotetraose (G4), is mediated by multiple residues across four subsites in its binding pocket. (E) Only subsites 3 and 4 are used to coordinate binding from *Athe\_0597* to its cognate disaccharide ligand cellobiose (G2). (F) Multiple residues in the binding pocket of *Athe\_0598* mediate the two subsites that coordinate its primary cognate ligand, cellobiose (G2). Residues with aromatic rings are written in bold.

### Genetic deletion of the cello-oligosaccharide ABC transporter locus (*athe\_0595–0598*) disrupts growth on cellobiose and cellulose

To probe its function in carbohydrate assimilation, the cello-oligosaccharide ABC transporter locus (*athe\_0595–0598*) was deleted *in vivo* via homologous recombination from the uracil biosynthesis-deficient parent strain MACB1018 ( $\Delta$ *pyrE*) (15). The cello-oligosaccharide transporter deletion strain HTAB187 ( $\Delta$ *pyrE*  $\Delta$ *athe\_0595–0598*) was generated using maltose as the carbon source following transformations.

We conducted growth curves using strain HTAB187 on individual carbon sources, including glucose, maltose, cellobiose, and Avicel (microcrystalline cellulose) (Fig. 5). We also monitored growth curves for the parent strain MACB1018 ( $\Delta$ *pyrE*) and the MsmK deletion strain MACB1080 ( $\Delta$ *pyrE*  $\Delta$ *msmK*). For the latter, we sought to investigate whether a single cello-oligosaccharide transporter deletion results in growth disruption similar to that observed with the inactivation of all oligosaccharide transporters in *A. bescii* (13). To ensure complete inactivation of the transporter, we deleted the entire *athe\_0595–0598* locus rather than individual components, as has been done in *Clostridium thermocellum* transporter studies (39).

Unsurprisingly, all three strains demonstrated similar growth behavior on glucose: HTAB187 grows similarly to MACB1018, while MACB1080 exhibits a slightly lower growth rate compared to MACB1018 during the exponential phase (Fig. 5), consistent with prior investigations (13). HTAB187 also exhibited no deviations in growth behavior compared to MACB1018 on maltose, whereas growth on this disaccharide is impaired for MACB1080 due to inactivation of its maltodextrin transporters through loss of ATPase MsmK. On cellobiose, however, all three strains displayed distinct growth behavior. Growth on cellobiose was clearly impaired for HTAB187 compared to MACB1018, although the defect was not severe enough to completely prevent growth. Surprisingly, MACB1080 showed no growth whatsoever on cellobiose in contrast with literature



**FIG 5** Growth curves of axenic, engineered *A. besicii* strains MACB1018 (green triangles), MACB1080 (blue circles), and HTAB187 (red squares) on carbohydrate substrates glucose, maltose, cellobiose, and Avicel (microcrystalline cellulose). All growth curves were performed in biological triplicate. For growth curves on glucose, maltose, and cellobiose, each data point represents the mean optical density at 680 nm (OD<sub>680</sub>) value. The growth curve on Avicel was based on quantification of pellet protein as previously described (43). Error bars denote the standard deviation.

results that suggest minimal growth on the disaccharide (13). And despite some growth on cellobiose, HTAB187 appears to be entirely incapable of growth on cellulose. MACB1080 also showed no growth on cellulose, consistent with previous reports (13). Visual inspection of the turbidity of cultures from all three strains on cellulose suggested that only MACB1018 was thriving (Fig. S1). Broadly, deletion of the *athe\_0595–0598* cello-oligosaccharide transporter locus results in (i) significant, but partial, impairment of growth on cellobiose, and (ii) elimination of growth on cellulose, without apparent defects in the uptake of other glucose-based substrates, such as glucose or maltose. Although it is possible that cellobiose can be extracellularly hydrolyzed into glucose, which HTAB187 can uptake, it is more likely that low-affinity transport through another ABC transporter explains the limited growth on cellobiose. If the former was the case, we would have expected partial MACB1080 growth on cellobiose given that it can still grow on glucose. Overall, these results indicate that *Athe\_0595–0598* is the major cello-oligosaccharide transporter in *A. besicii*.

## DISCUSSION

Here, using biophysics, structural modeling, and genetic knockouts, we show that *Athe\_0595–0598* is the principal ABC transporter that enables *A. besicii* to utilize cello-oligosaccharides. Through DSC, we screened the substrate specificity of SBPs *Athe\_0597* and *Athe\_0598*. We further characterized and quantified their substrate-binding thermodynamics using ITC, showing that both SBPs bind to cello-oligosaccharides with dissociation constants in the  $\mu\text{M}$  range, characteristic of high-affinity ABC sugar transporters (33, 40, 44). Using sequence and structural alignment to known thermophilic homologs, as well as molecular modeling, we further elucidated how

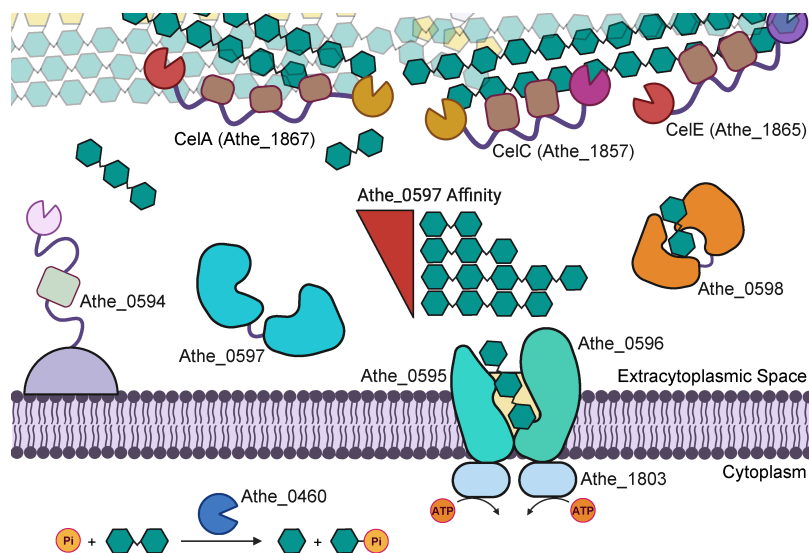
these carbohydrate ligands are docked within their respective SBP binding pockets. Finally, we complemented *in vitro* and *in silico* characterization with knockout of the *athe\_0595-0598* locus *in vivo* and subsequently studied changes to growth behavior on multiple carbohydrate sources. We showed that deletion of this transporter locus impaired growth on cellobiose and eliminated growth on cellulose, illustrating its critical role in cello-oligosaccharide and cellulose utilization.

Both ITC and DSC corroborated that Athe\_0597 and Athe\_0598 bind cello-oligosaccharide substrates with  $K_d$  values in the  $\mu\text{M}$  range (Table 1). While Athe\_0597 binds to cello-oligosaccharides of various lengths, from cellobiose to cellopentaose with similar  $T_m$  shifts, Athe\_0598 only binds to cellobiose. Although ITC and DSC largely agree that cello-oligosaccharides of lengths G2–G5 induce high-affinity binding with Athe\_0597, differences in thermal stabilization and  $K_d$  values are not as tightly correlated. We found a decreasing rank order between  $T_m$  shift and oligosaccharide length for Athe\_0597 (G4 > G5 > G2 > G3), whereas the rank order for decreasing binding affinity (decreasing  $K_d$ ) follows the trend of G2 > G3 > G5 > G4 (Table 1). In the context of the maltodextrin transport system, which includes two orthogonal SBPs, Athe\_0598 could be considered analogous to Athe\_2310 as the SBP with higher specificity toward shorter oligosaccharides (33). However, unlike the maltodextrin system where one SBP does not display a significant affinity toward the disaccharide maltose, both Athe\_0597 and Athe\_0598 bind cellobiose as a cognate ligand. Previously, Yokoyama et al. (45) proposed Athe\_0597 as a secreted SBP meant for plant cell wall adhesion (45). Our biophysical measurements demonstrated substrate specificities consistent with those reported in their study. However, our genetic and biochemical evidence indicates that Athe\_0597 is not simply a secreted adhesin, but rather the major substrate-binding component of the Athe\_0595-0598 ABC transporter system mediating cello-oligosaccharide uptake.

The Athe\_0597 and Athe\_0598 SBPs are highly conserved in the genera *Anaerocellum* and *Caldicellulosiruptor*, indicating their likely importance (Fig. 1B). Among them, Athe\_0597 is more broadly conserved across *Caldicellulosiruptor* species compared to Athe\_0598. Athe\_0597 is also multi-functional, binding an array of cello-oligosaccharides while binding cellobiose at association constants comparable to those of Athe\_0598. This suggests that Athe\_0597 plays the more important role in cello-oligosaccharide assimilation. Although all *Anaerocellum* and *Caldicellulosiruptor* species are hemicellulolytic, and most are cellulolytic, some *Caldicellulosiruptor* species, such as *C. kristjanssonii*, lack a GDL and are non-cellulolytic (19). Even in non-cellulolytic species, retaining the cello-oligosaccharide transporter locus may confer a metabolic advantage by enabling the scavenging of cello-oligosaccharides generated by neighboring cellulolytic species in their native hot springs.

Previous investigations have demonstrated that purified GDL enzyme cocktails, or even the concentrated native secretome itself, are not as effective as a live *A. besicii* culture in solubilizing cellulose, underscoring the importance of simultaneous uptake of cello-oligosaccharides to remove end-product inhibition and maximize cellulose deconstruction (21). Based on these results, we propose a model describing cellulose deconstruction by the coordinated action of GDL CAZymes and the transport of released cello-oligosaccharides by the Athe\_0595–0598 transporter (Fig. 6). As Athe\_0597 binds cello-oligosaccharides of lengths G2–G5 yet displays the highest affinity for cellobiose (G2), and Athe\_0598 binds strictly to cellobiose, cellobiose generated by exoglucanases (e.g., GH48 domains in CelA and CelC) likely predominates among imported sugars over larger cello-oligosaccharides produced by endoglucanase activity (e.g., GH9, GH5).

Transporter selectivity carries important bioenergetic consequences. Multiple studies have suggested that the bioenergetic benefit of transporting higher-order oligosaccharide substrates is vital for offsetting the significant metabolic costs of expressing a large array of extracellular CAZymes for cellulose deconstruction (46, 47). The cost of translocating one sugar molecule across the membrane remains fixed at two ATP molecules regardless of sugar size. An affinity for transporting larger sugars implies a more efficient use of resources (46). Cellobiose is a predominant product of cellulase activity and serves



**FIG 6** A proposed mechanism of cellulose deconstruction and transport in *Anaerocellum besicii*. In the extracellular space, secreted GDL CAZymes CelA, CelC, and CelE work in concert to deconstruct cellulose. Surface layer associated Athe\_0594 also potentially supports cellulose deconstruction through endoglucanase activity that generates cello-oligosaccharides. The released cello-oligosaccharides are first captured by ABC clodextrin-binding proteins Athe\_0597 (with affinity to cello-oligosaccharides of various lengths, ranked in order of decreasing  $K_d$  as determined by ITC in Table 1) and Athe\_0598 (with affinity only to cellobiose), and then delivered to the Athe\_0595–0596 permease. Cellodextrin translocation across the membrane is coupled to ATP hydrolysis by the MsmK (Athe\_1803) nucleotide-binding domain. Within the cytoplasm, cellobiose phosphorylases, such as those encoded by Athe\_0460, phosphorylate cellobiose into glucose and glucose-1-phosphate.

as a ligand for both Athe\_0597 and Athe\_0598, underscoring its central role in substrate uptake strategies of *A. besicii*. Additionally, *A. besicii* carries intracellular cellobiose phosphorylases, encoded by *athe\_0459* and *athe\_0460*, which cleave cellobiose into glucose and glucose-1-phosphate, thereby enabling an ATP-saving strategy observed in other cellulolytic anaerobes such as *Acetivibrio thermocellus* (basonym *Clostridium thermocellum*) and *Ruminiclostridium cellulolyticum* (13, 46, 48, 49).

We examined how this transporter's function contributes to growth on cellulose *in vivo*. Both *A. besicii* strains HTAB187 ( $\Delta pyrE \Delta athe_{0595-0598}$ ) and MACB1080 ( $\Delta pyrE \Delta msmK$ ) failed to grow on microcrystalline cellulose, suggesting that the promiscuous *msmK* (Athe\_1803) is the ATPase associated with the *athe\_0595-0598* cello-oligosaccharide transporter locus. This result is also consistent with a prior MACB1080 growth study performed on microcrystalline cellulose (13). Interestingly, while this previous study (13) showed that MACB1080 can experience modest growth on cellobiose, we were unable to reproduce this result. We suspect potential variance in inoculation method, or perhaps fructose carryover from the inoculum, could account for the observed growth of MACB1080 in cellobiose-containing media (13, 39).

HTAB187 demonstrated modest growth on cellobiose compared to the MACB1018 parent strain. Because HTAB187 growth on cellobiose reached final OD<sub>680</sub> levels comparable to growth on glucose, we speculate two possible causes for this phenomenon. First, it is possible that extracellular  $\beta$ -glucosidase activity could hydrolyze cellobiose into glucose, the latter of which strain HTAB187 can import. However, this seems unlikely, as we did not observe growth of MACB1080 under these conditions, which can also grow on glucose. Additionally, no extracellular  $\beta$ -glucosidases dedicated to hydrolyzing cellodextrins to glucose have been identified in the *A. besicii* secretome, and heterologous expression of an extracellular  $\beta$ -glucosidase in *A. besicii* only modestly affected growth on cellulose (50). Alternatively, a more plausible explanation is that *A.*

*bescii* possesses other MsmK-dependent oligosaccharide transporters with some affinity for cellobiose enabling partial growth on the disaccharide. However, because these redundant transporters may not possess as high a binding affinity for cellobiose, or because they are not as highly expressed, the growth rate of HTAB187 on cellobiose remains noticeably impaired.

The lignocellulolytic bacterium *Ruminiclostridium cellulolyticum* also employs an ABC transporter for cellodextrin uptake (51). Within its seven-gene cellulose utilization associated (*cua*) gene cluster, the SBP CuaD, which is specific only to cellobiose, appears to contribute to regulation of transporter expression, as mutation of *cuaD* impaired growth on cellobiose (51). By analogy, it remains possible that in the *athe\_0595–0598* locus, Athe\_0598—which only recognizes cellobiose—could similarly regulate carbohydrate utilization in *A. bescii*. However, unlike CuaD, which exhibited significantly higher affinity for cellobiose than the transport-associated SBP CuaA, Athe\_0598 showed lower affinity for cellobiose compared to Athe\_0597 (Table 1). Cellodextrin transport also plays a regulatory role in *Clostridium thermocellum*, whereby cellobiose uptake directly induces cellulosome expression (39). In the *C. thermocellum* system, initial biophysical studies suggested that four distinct SBPs were capable of binding cello-oligosaccharides, but subsequent genetic analyses demonstrated that only a single transporter is critical for uptake, with the others potentially serving regulatory or accessory functions (39, 40). In contrast, transcriptomic and proteomic analyses suggest that expression of the GDL CAZymes in *A. bescii* is consistent across diverse growth conditions (13, 25, 39). Decoupling the root cause of growth disruption on cellulose in *A. bescii*—potentially by observing simultaneous cellulose deconstruction alongside growth on non-cellodextrin sugars in a panel of additional mutants—is the most logical next step for disentangling transport and regulation in *A. bescii*.

In sum, this study identifies Athe\_0595–0598 as the primary ABC transporter responsible for cello-oligosaccharide uptake in *A. bescii*. Our results reveal that Athe\_0597 and Athe\_0598 exhibit distinct binding preferences, with Athe\_0597 recognizing a range of cello-oligosaccharides and Athe\_0598 specifically binding cellobiose. Knockout of the *athe\_0595–0598* ABC transporter locus in *A. bescii* strain HTAB187 abolished cellulose utilization and severely impaired growth on cellobiose, underscoring its essential role in assimilation of the cello-oligosaccharide substrates derived from lignocellulose. Without its ability to grow on cellulose, strain HTAB187 could conceivably be co-cultured with other cellulose-utilizing thermophiles, such as *Clostridium thermocellum*, for consolidated bioprocessing of plant biomass (52). Our findings illuminate core features of carbohydrate utilization and bacterial physiology in *A. bescii*, as well as in related *Anaerocellum* and *Caldicellulosiruptor* species, and establish a framework for future metabolic engineering efforts that involve manipulation of sugar transport.

## MATERIALS AND METHODS

### Bacteria and growth conditions

*E. coli* strains, including NEB 5- $\alpha$  (New England Biolabs) and BL21 (DE3) pRosetta2 (EMD Millipore), were plated routinely on Luria-Bertani medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with 1.5% agar. All plates were supplemented with 50  $\mu$ g/mL kanamycin for selection. All *E. coli* strains were cultured routinely in Luria-Bertani broth containing an additional 24 g/L yeast extract. All plates and cultures for growing *E. coli* BL21 (DE3) pRosetta2 strains were supplemented with 33  $\mu$ g/mL chloramphenicol antibiotic. *A. bescii* strains DSMZ 6725, MACB1018, and MACB1080 were provided by the labs of Dr. Robert Kelly (North Carolina State University) and Dr. Michael W. W. Adams (University of Georgia) and cultured on complex media (C516) containing 5 g/L maltose, 0.5 g/L yeast extract, and 40  $\mu$ M uracil as described previously (15, 53). Late exponential phase *A. bescii* cells were harvested and pelleted by centrifugation at 5,000  $\times$  *g* for 20 min. Genomic DNA was extracted from pelleted cells using a Quick-DNA Miniprep kit



quantity and purity were selected, pooled, concentrated, and buffer-exchanged using a 10 kDa MWCO PES filter 20 mL Spin-X Concentrator (Corning). The final protein storage and characterization buffer is comprised of 50 mM HEPES, 300 mM NaCl, pH 7.0. A bicinchoninic acid (BCA) assay (Thermo Fisher Scientific) was used to quantify pHT007 concentrated to 50 mg/mL and pHT017 concentrated to 10 mg/mL. Both proteins were run on SDS-PAGE to verify protein identity by weight and sample purity (Fig. S2).

### Isothermal titration calorimetry (ITC)

Methods were performed as previously described (33). Briefly, measurements were performed at 25°C using a MicroCal PEAQ Isothermal Titration Calorimeter (Malvern Panalytical) hosted by the Princeton University Biophysics Core Facility. Standard titration experiments were performed using a protein sample volume of 280  $\mu$ L at a concentration of 50  $\mu$ M and a sugar ligand volume of 40  $\mu$ L at a concentration of 500  $\mu$ M. A cell stir speed of 750 rpm was used throughout the entire titration experiment. The protein in the cell was initially injected with a priming aliquot of 0.2  $\mu$ L sugar, followed by 19 injections of 2.0  $\mu$ L each, with 2-minute intervals between injections. A single-site binding model was used to determine integrated heat effects via non-linear regression (Microcal PEAQ-ITC Analysis). Binding dissociation constant and key thermodynamic parameters were determined from the fitted isotherms using the Gibbs free energy form  $\Delta G = -RT\ln(K_d)$ . The dissociation constant  $K_d$  was calculated based on the binding isotherm's slope at the equivalence point, and its inverse was taken as the association rate constant  $K_a$ . 280  $\mu$ L of Milli-Q water was used in the reference cell.

### Differential scanning calorimetry (DSC)

Methods were performed as previously described (33). Briefly, the melting curve profiles of hexa-histidine-tagged Athe\_0597 and Athe\_0598 were determined via DSC using the MicroCal PEAQ-DSC (Malvern Panalytical) hosted at the Princeton University Biophysics Core Facility. Athe\_0597 and Athe\_0598 were screened against a palette of cello-oligosaccharide substrates: cellobiose (G2), cellotriose (G3), cellotetraose (G4), and cellopentaose (G5). Glucose (G1) was also tested as a negative control. Each protein–ligand mixture contained final concentrations of 2 mg/mL protein and 5 mM sugar, with the same protein storage buffer (50 mM HEPES, 300 mM NaCl, pH 7.0) used as the reference. All scan rates were set at 4°C/min, reaching a maximum temperature of 130°C. Raw data consisted of the heat capacity ( $C_p$ ) plotted as a function of temperature, followed by normalization to each respective run's maximum heat capacity value,  $C_{p,max}$ .  $C_{p,max}$  is defined as  $C_{p,max} = C_p(T = T_m)$ , where  $T_m$  is the melting temperature of the protein mixture. Each protein–carbohydrate mixture was ascribed a  $\Delta T_m$ , defined as  $\Delta T_m = T_{m,holo} - T_{m,apo}$ , where  $T_{m,holo}$  is the melting temperature of a given protein–sugar mixture and  $T_{m,apo}$  is the melting temperature of the protein in the absence of sugar.

### Construction of *A. besicii* $\Delta$ athe\_0595–0598 knockout strain

Starting with the MACB1018 background parent strain ( $\Delta$ pyrE), the *athe\_0595-0598* gene cluster was chromosomally deleted via homologous recombination using previously described methods (15) (Fig. S3). The plasmid pKHT017 contained a  $\sim$  1 kb flanking region upstream of the *athe\_0595* gene and a  $\sim$  1 kb flanking region downstream of the *athe\_0598* gene. Both flanking regions are adjacent to one another in the knockout vector (Fig. S2). Flanking regions were PCR-amplified from *A. besicii* DSMZ 6725 genomic DNA using Phusion polymerase (New England Biolabs) and inserted into a pGL103 backbone, as previously described (15). The knockout vector was transformed into *E. coli* NEB 5- $\alpha$  via heat shock at 42°C, followed by recovery in SOC outgrowth medium (New England Biolabs) and then growth on plates and in cultures supplemented with 50  $\mu$ g/mL kanamycin and 50  $\mu$ g/mL apramycin selection pressures for plasmid maintenance. pKHT017 was prepared and purified using the Zymo Research Plasmid Miniprep – Classic kit (Zymo Research) and sequenced via Plasmidsaurus' whole-plasmid

sequencing service, as described above. Sequence confirmed pKHT017 transformants were grown in 250 mL Luria-Bertani medium supplemented with additional 24 g/L yeast extract, as well as 50 µg/mL kanamycin and 33 µg/mL apramycin antibiotics, in 1 L shake flasks at 37°C. Incubator shake speed was set at 250 rpm. Cells were pelleted and processed via a ZymoPURE II Plasmid Purification Maxiprep kit. The pKHT017 plasmid was then methylated *in vitro* with M.CbeI methyltransferase, as previously described (15, 56). Successful methylation was verified by protection against HaeIII restriction endonuclease activity.

Competent MACB1018 cells were prepared in low-osmolarity-defined (LOD) media containing 5 g/L cellobiose and 40 µM uracil, as described previously (15, 53). LOD media and all media hereafter described were degassed and made anaerobic for 15 min using a vacuum pump, with a gaseous headspace comprising 80% (vol/vol) nitrogen and 20% (vol/vol) carbon dioxide gas. Competent cells were electroporated using a Bio-Rad Gene Pulser with 2 µg of methylated pKHT017 plasmid at 2.2 kV, resistance of 400 Ω, and capacitance of 25 µF. One milliliter of recovery media, containing modified DSM 516 media supplemented with 5 g/L of yeast extract and preheated to 70°C, was immediately used to resuspend electroshocked cells (15). After both 60 min and 120 min post-electroporation, 5 mL of electroporated cells in recovery medium were transferred to 50 mL (selective) DSM 516 medium supplemented with 50 µg/mL kanamycin. Cloudy media after 48–72 h of growth at 70°C suggested successful transformations; 1 mL samples of such cultures were passaged into 10 mL fresh selective DSM 516 media containing 50 µg/mL kanamycin. Passaged transformants were then plated in solid, selective modified DSM 516 medium with 1.5% agar, grown for 48–72 h at 70°C under 95% (vol/vol) nitrogen and 5% (vol/vol) hydrogen gas, and single colonies were selected and screened via PCR to confirm knockout vector chromosomal integration adjacent to the *athe\_0595-0598* locus (13). A PCR-verified first crossover colony is then plated on solid, modified DSM 516 medium containing 8 mM 5-fluoroorotic acid (5-FOA) and 40 µM uracil. The 5-FOA counter-selection step also marks a permanent change in growth substrate from 5 g/L cellobiose to 5 g/L maltose. Single colonies were picked and screened for complete deletion of the *athe\_0595-0598* locus following counter-selection, using colony PCR with primers directly outside the 1 kb upstream and downstream flanking regions present in the pKHT017 vector.

### A. *bescii* growth curves on select carbohydrate substrates

Freezer stocks of MACB1018 ( $\Delta$ *pyrE*), MACB1080 ( $\Delta$ *pyrE*  $\Delta$ *msmK*), and HTAB187 ( $\Delta$ *pyrE*  $\Delta$ *athe\_0595-0598*) were each thawed and used to inoculate 50 mL of anaerobic, non-selective modified DSM 516 medium containing 40 µM uracil and 5 g/L of fructose in 100 mL serum bottles for overnight growth. Strains were passaged once more into anaerobic, non-selective modified DSM 516 medium containing 40 µM uracil and 5 g/L of fructose. Upon reaching late exponential phase at approximately 16 h of growth at 70°C, cultures were cooled to room temperature. Ten milliliters of cells from each strain was extracted and centrifuged at 14,000 rpm for 2 min. The supernatant was primarily decanted by pouring out the liquid, with residual supernatant removed via pipetting. The pellet was then washed and resuspended in 1 mL of non-selective modified DSM 516 sugar-free medium, and then injected into 10 mL of the same sugar-free medium contained within a 18 × 125 mm anaerobic tube sealed with 20 mm butyl rubber stoppers (Duran Wheaton Kimble), resulting in a sugar-free bacterial culture with final OD<sub>680</sub> ~0.10.

One milliliter of the sugar-free bacterial culture was used to inoculate 50 mL of non-selective modified DSM 516 medium containing 40 µM uracil and a single carbohydrate source at 5 g/L, sealed by 20 mm butyl rubber stoppers (Duran Wheaton Kimble). All inoculated cultures reported an initial OD<sub>680</sub> ~ 0.00 and were grown at 70°C, without shaking, for 40 h in biological triplicate, as previously described (13). The OD<sub>680</sub> of growth cultures was measured at intervals of roughly 2.5–6.0 h using the cuvette setting of a UV-Vis Nanodrop Spectrophotometer, with 1× DSM 516 salt solution as the blank.

Cell protein content, used as a proxy for cell density under Avicel growth conditions, was quantified using the Bradford method with bovine serum albumin (BSA) protein standards (Pierce Thermo Scientific) in a 96-well plate format, as previously described (13, 39, 43).

## Bioinformatics

Through PDB sequence similarity search, glucan-binding protein orthologs with solved crystal structures were identified using BLAST, with Athe\_0597 and Athe\_0598 as the query protein sequences (57). For Athe\_0597, structures were all from *Streptococcus pneumoniae* (PDB: 5SUO, 5SWA, 5SWB) (41). For Athe\_0598, structures included those from *Streptococcus pneumoniae* (PDB: 5SWA), as well as *Caldanaerobius polysaccharolyticus* (PDB: 4G68), *Paenibacillus sp. str. FPU-7* (PDB: 7EHP), and *Thermotoga maritima* (PDB: 2O7I) (41, 58–60). Amino acid sequences of all homologs used in this study are summarized in Table S3. PROMALS3D and ESPript 3.0 visualization software were used to identify similarities in secondary and tertiary protein structures from structure-based sequence alignment of each SBP and their orthologs (61, 62).

## Ligand docking models

ColabFold v1.5.5 supplied with AlphaFold 2 parameters was used to generate the initial models. The best-match homolog was used as the template, with other homologs providing the MSA (63, 64). The “ref2015” score function under the Rosetta *relax* protocol was used to minimize AlphaFold-generated structures. Docking of carbohydrate ligands onto protein binding pockets was performed using the ROSIE server (65, 66). Initial posing of the carbohydrate ligand within the protein binding pocket was based on homology to the corresponding crystal structure. For Athe\_0597, glucose was positioned in the binding cleft from crystal structure 5SWI; cellobiose was positioned in the two subsites of 5SWA; cellotriose was positioned in three subsites of 5SWB; cellotetraose was positioned in four subsites of 5SWB; and cellopentaose was built by extending the non-reducing end of the bound cellotetraose with another glucosyl monomer (41). For Athe\_0598, glucose was positioned by homology to 5SWI; cellobiose was positioned in the sites aligned with 2O7I; cellotriose was initially positioned by homology to 4G68; cellotriose was positioned in subsites from 7C68; and cellopentaose was positioned using the aligned subsites in 7C6R (41, 58–60). Two hundred conformers of each carbohydrate substrate were generated using BCL (67). Each protein–substrate combination yielded 200 docked structures, and interface energies of binding were quantified from the 10 lowest energy structures. KDEEP, a 3D convolutional neural network method for assessing interface energy, was used to further evaluate the ligand-bound pose (68). All protein structures were visualized using PyMOL (69).

## ACKNOWLEDGMENTS

This work was supported by a Roberto Rocca Graduate Fellowship from Techint Group to H.T.; by the High Meadows Environmental Institute at Princeton University through the generous support of the William Clay Ford, Jr. '79 and Lisa Vanderzee Ford '82 Graduate Fellowship Fund to H.T.; by a National Science Foundation Graduate Research Fellowship DGE-2039656 to V.J.; by a Gordon Wu Fellowship to V.J.; by startup funds from the Department of Chemical and Biological Engineering and the Omenn-Darling Bioengineering Institute at Princeton University to J.A.J.; and by seed funds from the School of Engineering and Applied Science and startup funds from the Department of Chemical and Biological Engineering at Princeton University to J.M.C. All authors acknowledge financial support from Princeton University's Omenn-Darling Bioengineering Institute.

We acknowledge Venu Gopal Vandavasi, Director of the Princeton Biophysics Core, for instrument training and advice on experimental design. We also acknowledge Tom Silhavy, Ned Wingreen, and A. James Link for helpful discussions.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey, USA

<sup>2</sup>Omenn-Darling Bioengineering Institute, Princeton University, Princeton, New Jersey, USA

<sup>3</sup>Andlinger Center for Energy and the Environment, Princeton University, Princeton, New Jersey, USA

<sup>4</sup>Department of Chemistry, Princeton University, Princeton, New Jersey, USA

<sup>5</sup>Princeton Institute for Computational Science and Engineering, Princeton University, Princeton, New Jersey, USA

<sup>6</sup>Molecular Biology Department, Princeton University, Princeton, New Jersey, USA

<sup>7</sup>High Meadows Environmental Institute, Princeton University, Princeton, New Jersey, USA

## AUTHOR ORCID*s*

Hansen Tjo  <http://orcid.org/0009-0008-5994-8281>

Virginia Jiang  <http://orcid.org/0000-0002-7380-1306>

Jerelle A. Joseph  <http://orcid.org/0000-0003-4525-180X>

Jonathan M. Conway  <http://orcid.org/0000-0002-2715-2149>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">Techint Group</a>	Roberto Rocca Graduate Fellowship	Hansen Tjo
<a href="#">High Meadows Environmental Institute, Princeton University</a>	Graduate Fellowship	Hansen Tjo
<a href="#">National Science Foundation Graduate Research Fellowship Program</a>	DGE-2039656	Virginia Jiang
<a href="#">Princeton University</a>		Jerelle A. Joseph
<a href="#">Princeton University</a>		Jonathan M. Conway

## AUTHOR CONTRIBUTIONS

Hansen Tjo, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft | Virginia Jiang, Data curation, Investigation, Methodology, Visualization, Writing – review and editing | Anherutowa Calvo, Investigation, Writing – review and editing | Jerelle A. Joseph, Funding acquisition, Resources, Supervision, Writing – review and editing | Jonathan M. Conway, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplemental material (AEM01284-25-S0001.pdf).** Figures S1 to S3; Tables S1 to S3.

## REFERENCES

- Zuiderveen EAR, Kuipers KJJ, Caldeira C, Hanssen SV, van der Hulst MK, de Jonge MMJ, Vlysidis A, van Zelm R, Sala S, Huijbregts MAJ. 2023. The potential of emerging bio-based products to reduce environmental impacts. *Nat Commun* 14:8521. <https://doi.org/10.1038/s41467-023-43797-9>
- El-Chichakli B, von Braun J, Lang C, Barben D, Philp J. 2016. Policy: five cornerstones of a global bioeconomy. *Nature* 535:221–223. <https://doi.org/10.1038/535221a>
- Gilna P, Lynd LR, Mohnen D, Davis MF, Davison BH. 2017. Progress in understanding and overcoming biomass recalcitrance: a bioenergy

- science center (BESC) perspective. *Biotechnol Biofuels* 10:285. <https://doi.org/10.1186/s13068-017-0971-1>
4. Bing RG, Sulis DB, Wang JP, Adams MWW, Kelly RM. 2021. Thermophilic microbial deconstruction and conversion of natural and transgenic lignocellulose. *Environ Microbiol Rep* 13:272–293. <https://doi.org/10.1111/1758-2229.12943>
  5. Crosby JR, Laemthong T, Lewis AM, Straub CT, Adams MW, Kelly RM. 2019. Extreme thermophiles as emerging metabolic engineering platforms. *Curr Opin Biotechnol* 59:55–64. <https://doi.org/10.1016/j.copbio.2019.02.006>
  6. Straub CT, Bing RG, Otten JK, Keller LM, Zeldes BM, Adams MWW, Kelly RM. 2020. Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. *Biotechnol Bioeng* 117:3799–3808. <https://doi.org/10.1002/bit.27529>
  7. Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM. 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr Opin Biotechnol* 19:210–217. <https://doi.org/10.1016/j.copbio.2008.04.007>
  8. Tjo H, Conway JM. 2024. Sugar transport in thermophiles: bridging lignocellulose deconstruction and bioconversion. *J Ind Microbiol Biotechnol* 51:kuae020. <https://doi.org/10.1093/jimb/kuae020>
  9. de Vos WM, Kengen SW, Voorhorst WG, van der Oost J. 1998. Sugar utilization and its control in hyperthermophiles. *Extremophiles* 2:201–205. <https://doi.org/10.1007/s007920050061>
  10. Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM. 2020. The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in “Caldi world”. *Extremophiles* 24:1–15. <https://doi.org/10.1007/s00792-019-01116-5>
  11. Yang S-J, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MWW. 2010. Classification of “*Anaerocellum thermophilum*” strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *Int J Syst Evol Microbiol* 60:2011–2015. <https://doi.org/10.1099/ijs.0.017731-0>
  12. Bing RG, Willard DJ, Crosby JR, Adams MWW, Kelly RM. 2023. Whither the genus *Caldicellulosiruptor* and the order Thermoanaerobacterales: phylogeny, taxonomy, ecology, and phenotype. *Front Microbiol* 14:1212538. <https://doi.org/10.3389/fmicb.2023.1212538>
  13. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM, Klingeman DM, Poole FL 2nd, Zhang Y, Kelly RM, Adams MWW. 2021. Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *mSystems* 6:e0134520. <https://doi.org/10.1128/mSystems.01345-20>
  14. Williams-Rhaesa AM, Awuku NK, Lipscomb GL, Poole FL, Rubinstein GM, Conway JM, Kelly RM, Adams MWW. 2018. Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium *Caldicellulosiruptor bescii*. *Extremophiles* 22:629–638. <https://doi.org/10.1007/s00792-018-1023-x>
  15. Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MWW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. *Appl Environ Microbiol* 82:4421–4428. <https://doi.org/10.1128/AEM.00570-16>
  16. Chung D, Cha M, Guss AM, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc Natl Acad Sci USA* 111:8931–8936. <https://doi.org/10.1073/pnas.1402210111>
  17. Williams-Rhaesa AM, Rubinstein GM, Scott IM, Lipscomb GL, Poole FL, Kelly RM, Adams MWW. 2018. Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. *Metab Eng Commun* 7:e00073. <https://doi.org/10.1016/j.mec.2018.e00073>
  18. Tanwee TNN, Lipscomb GL, Vaillonis JL, Zhang K, Bing RG, O’Quinn HC, Poole FL, Zhang Y, Kelly RM, Adams MWW. 2024. Metabolic engineering of *Caldicellulosiruptor bescii* for 2,3-butanediol production from unpretreated lignocellulosic biomass and metabolic strategies for improving yields and titers. *Appl Environ Microbiol* 90:e0195123. <https://doi.org/10.1128/aem.01951-23>
  19. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL 2nd, Adams MWW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. *J Bacteriol* 194:4015–4028. <https://doi.org/10.1128/JB.00266-12>
  20. Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, Giannone RJ, Hettich RL, Williams-Rhaesa AM, Lipscomb GL, Adams MWW, Kelly RM. 2017. Functional analysis of the glucan degradation locus in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. *Appl Environ Microbiol* 83:e01828–17. <https://doi.org/10.1128/AEM.01828-17>
  21. Conway JM, Crosby JR, McKinley BS, Seals NL, Adams MWW, Kelly RM. 2018. Parsing *in vivo* and *in vitro* contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. *Biotech & Bioengineering* 115:2426–2440. <https://doi.org/10.1002/bit.26773>
  22. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang S-J, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ. 2013. Revealing nature’s cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 342:1513–1516. <https://doi.org/10.1126/science.1244273>
  23. Brunecky R, Donohoe BS, Yarbrough JM, Mittal A, Scott BR, Ding H, Taylor II LE, Russell JF, Chung D, Westpheling J, Teter SA, Himmel ME, Bomble YJ. 2017. The multi domain *Caldicellulosiruptor bescii* CelA cellulase excels at the hydrolysis of crystalline cellulose. *Sci Rep* 7:9622. <https://doi.org/10.1038/s41598-017-08985-w>
  24. Zverlov V, Mahr S, Riedel K, Bronnenmeier K. 1998. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile “*Anaerocellum thermophilum*” with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology (Reading)* 144 (Pt 2):457–465. <https://doi.org/10.1099/00221287-144-2-457>
  25. Poudel S, Giannone RJ, Basen M, Nookaew I, Poole FL 2nd, Kelly RM, Adams MWW, Hettich RL. 2018. The diversity and specificity of the extracellular proteome in the cellulolytic bacterium *Caldicellulosiruptor bescii* is driven by the nature of the cellulosic growth substrate. *Biotechnol Biofuels* 11:80. <https://doi.org/10.1186/s13068-018-1076-1>
  26. Blumer-Schuette SE, Alahuhta M, Conway JM, Lee LL, Zurawski JV, Giannone RJ, Hettich RL, Lunin VV, Himmel ME, Kelly RM. 2015. Discrete and structurally unique proteins (täpirins) mediate attachment of extremely thermophilic *Caldicellulosiruptor* species to cellulose. *J Biol Chem* 290:10645–10656. <https://doi.org/10.1074/jbc.M115.641480>
  27. Crosby JR, Laemthong T, Bing RG, Zhang K, Tanwee TNN, Lipscomb GL, Rodionov DA, Zhang Y, Adams MWW, Kelly RM. 2022. Biochemical and regulatory analyses of xylanolytic regulons in *Caldicellulosiruptor bescii* reveal genus-wide features of hemicellulose utilization. *Appl Environ Microbiol* 88:e01302–22. <https://doi.org/10.1128/aem.01302-22>
  28. Davidson AL, Dassa E, Orelle C, Chen J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* 72:317–364. <https://doi.org/10.1128/MMBR.00031-07>
  29. Schneider E. 2001. ABC transporters catalyzing carbohydrate uptake. *Res Microbiol* 152:303–310. [https://doi.org/10.1016/s0923-2508\(01\)01201-3](https://doi.org/10.1016/s0923-2508(01)01201-3)
  30. Vanfossen AL, Verhaart MRA, Kengen SMW, Kelly RM. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. *Appl Environ Microbiol* 75:7718–7724. <https://doi.org/10.1128/AEM.01959-09>
  31. de Boer M, Gouridis G, Vietrov R, Begg SL, Schuurman-Wolters GK, Husada F, Eleftheriadis N, Poolman B, McDevitt CA, Cordes T. 2019. Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. Edited by R. Aldrich, B. Chanda, B. Chanda, and M. P. Goldschen-Ohm. *eLife* 8:e44652. <https://doi.org/10.7554/eLife.44652>
  32. Bosdriesz E, Magnúsdóttir S, Bruggeman FJ, Teusink B, Molenaar D. 2015. Binding proteins enhance specific uptake rate by increasing the substrate–transporter encounter rate. *FEBS J* 282:2394–2407. <https://doi.org/10.1111/febs.13289>
  33. Tjo H, Jiang V, Joseph JA, Conway JM. 2025. Maltodextrin transport in the extremely thermophilic, lignocellulose degrading bacterium *Anaerocellum bescii* (f. *Caldicellulosiruptor bescii*). *J Bacteriol* 207:e0040124. <https://doi.org/10.1128/jb.00401-24>
  34. Thomas G. 2010. Homes for the orphans: utilization of multiple substrate-binding proteins by ABC transporters. *Mol Microbiol* 75:6–9. <https://doi.org/10.1111/j.1365-2958.2009.06961.x>
  35. Chen C, Malek AA, Wargo MJ, Hogan DA, Beattie GA. 2010. The ATP-binding cassette transporter Cbc (choline/betaine/carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds. *Mol Microbiol* 75:29–45. <https://doi.org/10.1111/j.1365-2958.2009.06962.x>

36. Laemthong T, Bing RG, Crosby JR, Adams MWW, Kelly RM. 2022. Engineering *Caldicellulosiruptor bescii* with surface layer homology domain-linked glycoside hydrolases improves plant biomass solubilization. *Appl Environ Microbiol* 88:e01274–22. <https://doi.org/10.1128/aem.01274-22>
37. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM. 2016. Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. *J Biol Chem* 291:6732–6747. <https://doi.org/10.1074/jbc.M115.707810>
38. Gill P, Moghadam TT, Ranjbar B. 2010. Differential scanning calorimetry techniques: applications in biology and nanoscience. *J Biomol Tech* 21:167–193.
39. Yan F, Dong S, Liu Y-J, Yao X, Chen C, Xiao Y, Bayer EA, Shoham Y, You C, Cui Q, Feng Y. 2022. Deciphering cellodextrin and glucose uptake in *Clostridium thermocellum*. *mBio* 13:e0147622. <https://doi.org/10.1128/mBio.01476-22>
40. Nataf Y, Yaron S, Stahl F, Lamed R, Bayer EA, Schepers T-H, Sonenshein AL, Shoham Y. 2009. Cellodextrin and laminaribiose ABC transporters in *Clostridium thermocellum*. *J Bacteriol* 191:203–209. <https://doi.org/10.1128/JB.01190-08>
41. Robb M, Hobbs JK, Woodiga SA, Shapiro-Ward S, Suits MDL, McGregor N, Brumer H, Yesilkaya H, King SJ, Boraston AB. 2017. Molecular characterization of N-glycan degradation and transport in *Streptococcus pneumoniae* and its contribution to virulence. *PLoS Pathog* 13:e1006090. <https://doi.org/10.1371/journal.ppat.1006090>
42. Oldham ML, Chen S, Chen J. 2013. Structural basis for substrate specificity in the *Escherichia coli* maltose transport system. *Proc Natl Acad Sci USA* 110:18132–18137. <https://doi.org/10.1073/pnas.1311407110>
43. Hong W, Zhang J, Feng Y, Mohr G, Lambowitz AM, Cui G-Z, Liu Y-J, Cui Q. 2014. The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol Biofuels* 7:80. <https://doi.org/10.1186/1754-6834-7-80>
44. Shukla S, Bafna K, Gullett C, Myles DAA, Agarwal PK, Cuneo MJ. 2018. Differential substrate recognition by maltose binding proteins influenced by structure and dynamics. *Biochemistry* 57:5864–5876. <https://doi.org/10.1021/acs.biochem.8b00783>
45. Yokoyama H, Yamashita T, Morioka R, Ohmori H. 2014. Extracellular secretion of noncatalytic plant cell wall-binding proteins by the cellulolytic thermophile *Caldicellulosiruptor bescii*. *J Bacteriol* 196:3784–3792. <https://doi.org/10.1128/JB.01897-14>
46. Zhang YHP, Lynd LR. 2005. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc Natl Acad Sci USA* 102:7321–7325. <https://doi.org/10.1073/pnas.0408734102>
47. Yi-Heng ZP, Lynd LR. 2005. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. *J Bacteriol* 187:99–106. <https://doi.org/10.1128/JB.187.1.99-106.2005>
48. Liu N, Fosses A, Kampik C, Parsieglä G, Denis Y, Vita N, Fierobe H-P, Perret S. 2019. *In vitro* and *in vivo* exploration of the cellobiose and cellodextrin phosphorylases panel in *Ruminiclostridium cellulolyticum*: implication for cellulose catabolism. *Biotechnol Biofuels* 12:208. <https://doi.org/10.1186/s13068-019-1549-x>
49. Bai S, Yang L, Wang H, Yang C, Hou X, Gao J, Zhang Z. 2022. Cellobiose phosphorylase from *Caldicellulosiruptor bescii* catalyzes reversible phosphorolysis via different kinetic mechanisms. *Sci Rep* 12:3978. <https://doi.org/10.1038/s41598-022-08036-z>
50. Kim SK, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2017. Heterologous expression of a  $\beta$ -D-glucosidase in *Caldicellulosiruptor bescii* has a surprisingly modest effect on the activity of the exoprotoeme and growth on crystalline cellulose. *J Ind Microbiol Biotechnol* 44:1643–1651. <https://doi.org/10.1007/s10295-017-1982-4>
51. Fosses A, Maté M, Franche N, Liu N, Denis Y, Borne R, de Philip P, Fierobe H-P, Perret S. 2017. A seven-gene cluster in *Ruminiclostridium cellulolyticum* is essential for signalization, uptake and catabolism of the degradation products of cellulose hydrolysis. *Biotechnol Biofuels* 10:250. <https://doi.org/10.1186/s13068-017-0933-7>
52. Tjo H, Blundin K, Conway JM. 2025. Designing thermophilic, synthetic microbial communities for consolidated bioprocessing. *BioDesign Research* 7:100024. <https://doi.org/10.1016/j.bidere.2025.100024>
53. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J. 2013. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. *J Ind Microbiol Biotechnol* 40:41–49. <https://doi.org/10.1007/s10295-012-1202-1>
54. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>
55. Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234. <https://doi.org/10.1016/j.pep.2005.01.016>
56. Chung D, Farkas J, Huddleston JR, Olivar E, Westpheling J. 2012. Methylation by a unique  $\alpha$ -class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. *PLoS One* 7:e43844. <https://doi.org/10.1371/journal.pone.0043844>
57. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
58. Han Y, Agarwal V, Dodd D, Kim J, Bae B, Mackie RI, Nair SK, Cann IKO. 2012. Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *J Biol Chem* 287:34946–34960. <https://doi.org/10.1074/jbc.M112.391532>
59. Cuneo MJ, Beese LS, Hellinga HW. 2009. Structural analysis of semi-specific oligosaccharide recognition by a cellulose-binding protein of *Thermotoga maritima* reveals adaptations for functional diversification of the oligopeptide periplasmic binding protein fold. *J Biol Chem* 284:33217–33223. <https://doi.org/10.1074/jbc.M109.041624>
60. Itoh T, Yaguchi M, Nakaichi A, Yoda M, Hibi T, Kimoto H. 2021. Structural characterization of two solute-binding proteins for N,N'-diacetylchitobiose/N,N',N''-triacetylchitotriose of the gram-positive bacterium, *Paenibacillus* sp. str. FPU-7. *J Struct Biol* X 5:100049. <https://doi.org/10.1016/j.jysbx.2021.100049>
61. Pei J, Grishin NV. 2014. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information, p 263–271. In Russell DJ (ed), *Multiple sequence alignment methods* [Internet]. Vol. Available from. Humana Press, Totowa, NJ.
62. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–4. <https://doi.org/10.1093/nar/gku316>
63. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
64. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* 19:679–682. <https://doi.org/10.1038/s41592-022-01488-1>
65. DeLuca S, Khar K, Meiler J. 2015. Fully flexible docking of medium sized ligand libraries with rosettaligand. *PLoS One* 10:e0132508. <https://doi.org/10.1371/journal.pone.0132508>
66. Lyskov S, Chou F-C, Conchúir SÓ, Der BS, Drew K, Kuroda D, Xu J, Weitzner BD, Renfrew PD, Sripakdeevong P, Borgo B, Havranek JJ, Kuhlman B, Kortemme T, Bonneau R, Gray JJ, Das R. 2013. Serverification of the molecular modeling applications: the rosetta online server that includes everyone (ROSIE). *PLoS One* 8:e63906. <https://doi.org/10.1371/journal.pone.0063906>
67. Kothiwale S, Mendenhall JL, Meiler J. 2015. BCL::Conf: small molecule conformational sampling using a knowledge based rotamer library. *J Cheminform* 7:47. <https://doi.org/10.1186/s13321-015-0095-1>
68. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. 2018.  $K_{DEEP}$ : protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J Chem Inf Model* 58:287–296. <https://doi.org/10.1021/acs.jcim.7b00650>
69. PyMOL. 2025. The PyMOL Molecular Graphics System. Version 3.0 Schrödinger, LLC.