Insertion of Endocellulase Catalytic Domains into Thermostable Consensus Ankyrin Scaffolds: Effects on Stability and Cellulolytic Activity

Eva S. Cunha, a,b Christine L. Hatem, b Doug Barrick b
Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins University, Baltimore, Maryland, USA a; Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland, USA b

Degradation of cellulose for biofuels production holds promise in solving important environmental and economic problems. However, the low activities (and thus high enzyme-to-substrate ratios needed) of hydrolytic cellulase enzymes, which convert cellulose into simple sugars, remain a major barrier. As a potential strategy to stabilize cellulases and enhance their activities, we have embedded cellulases of extremophiles into hyperstable α-helical consensus ankyrin domain scaffolds. We found the catalytic domains CelA (CA, GH8; Clostridium thermocellum) and Cel12A (C12A, GH12; Thermotoga maritima) to be stable in the context of the ankyrin scaffold and to be active against both soluble and insoluble substrates. The ankyrin repeats in each fusion are folded, although it appears that for the C12A catalytic domain (CD; where the N and C termini are distant in the crystal structure), the two flanking ankyrin domains are independent, whereas for CA (where termini are close), the flanking ankyrin domains stabilize each other. Although the activity of CA is unchanged in the context of the ankyrin scaffold, the activity of C12A is increased between 2- and 6-fold (for regenerated amorphous cellulose and carboxymethyl cellulose substrates) at high temperatures. For C12A, activity increases with the number of flanking ankyrin repeats. These results showed ankyrin arrays to be a promising scaffold for constructing designer cellulosomes, preserving or enhancing enzymatic activity and retaining thermostability. This modular architecture will make it possible to arrange multiple cellulase domains at a precise spacing within a single polypeptide, allowing us to search for spacings that may optimize reactivity toward the repetitive cellulose lattice.

Plant biomass is a worldwide natural resource that can be attained in large quantities from waste sources (crop residues and manure) as well as from primary agricultural sources (1). Plant biomass recalcitrance results in part from the rigidity of hydrogen-bonded cellulose fibrils. Although cellulose can be degraded to fermentable sugars, such as glucose, by hydrolytic enzymes (cellulases), the rigid, insoluble nature of cellulose fibers in biomass renders this process slow and inefficient. Because of this, large quantities of cellulase enzymes are needed, significantly raising the production costs of cellulosic biofuels.

In nature, there is a wide variety of microorganisms (aerobic and anaerobic eubacteria, archeabacteria, and fungi) responsible for the degradation of cellulose. Many of these microbes produce a battery of secreted cellulases with a range of structures, activities, and stabilities. There have been considerable efforts to find enzymes that are resistant to the high temperatures and acidity conditions often used to facilitate access of enzymes to cellulosic substrates (2, 3). As a result, extremophiles are often isolated from hot environments, such as Clostridium thermocellum and Thermotoga maritima, having been the subjects of considerable research (4–8). Enzymes from these extremophiles are natural starting points for optimization of cellulase stability and activity.

Enzyme stabilization and lifetime improvement have been achieved through point substitutions and chemical modifications of single catalytic domains (CDs), either through rational, structure-based modeling or through directed evolution (9–14). Another approach to increase enzyme stability and enhance activity is insertion of protein domains into stable scaffolds (15). For example, Kim et al. inserted exoinulinase E1, an unstable glycohydrolase, into a thermophilic scaffold maltodextrin-binding protein from Pyrococcus furiosus and found that the scaffold-inserted E1 had improved kinetic stability (16). In this type of scaffolding strategy, the distance between the N and C termini of the inserted protein (in this study, cellulase catalytic domains) is an important parameter. The termini must have about the same spacing as the insertion site, to avoid strain and disruptions of structure within the inserted and scaffold domains (17). Cellulase catalytic domains belong to various glycoside hydrolase (GH) families, as reported in the CAZY database, with different folds, and thus have variable N-to-C distances (18). The CDs used in this study have their termini separated by 8.3 Å (CA CD; α/αs) and 30.1 Å (C12A CD; β-jelly) (Fig. 1B). CA is synthesized by Clostridium thermocellum as part of a macromolecular complex called the cellulosome, whereas C12A is synthesized as the catalytic domain alone (7, 19, 20).

Insertion into a scaffold protein requires appropriate selection of insertion sites to avoid disruption of structure of the inserted protein or scaffold (21, 22). Proteins with linear, repeated architectures, such as ankyrin repeats, have structurally modular architectures that should be able to accommodate insertions locally,
without affecting more distant regions of the scaffold. Ankyrin repeat proteins are roughly linear arrays of 33-residue repeating units (23, 24). Each unit consists of two short helices connected by a short, conserved turn and connects to neighboring repeats through an extended –hairpin-containing loop (25). Adjacent repeats strongly stabilize each other through highly stabilizing interfaces. The repeating unit spans about 11 Å, approximately the size of each cellobiose unit (2 units of glucose [10.4 Å]). Studies have found that ankyrin repeats are greatly stabilized by consensus sequence substitutions (26, 27). For a five-repeat consensus ankyrin protein, the thermal denaturation midpoint (Tm) is around 90°C (28, 29). Therefore, consensus ankyrin seems to be an ideal scaffold to study the effects of insertion of cellulase CDs with various N-to-C distances.

Here, we inserted the CD of CelA from *Clostridium thermocellum* (CA) and that of Cel12A from *Thermotoga maritima* (C12A) into hyperstable consensus ankyrin repeat scaffolds. We cloned, expressed, and purified the cellulases on their own and inserted them into arrays of flanking ankyrin repeats of increasing length. By comparing the stabilities and cellulolytic activities of scaffolded versus free cellulase domains, we determined the optimal configuration for this chimerization strategy and identified constructs that led to activity enhancements. We determined the thermal denaturation of these constructs as well as their cellulolytic activities against CMC (carboxymethyl cellulose) and RAC (regenerated amorphous cellulose) substrates. We found that the consensus ankyrin scaffold is able to accommodate active cellulase domain insertion but that the structural integrity of the resulting constructs depends on the proximity of the cellulase termini.

MATERIALS AND METHODS

Cloning. Cellulase catalytic domains were cloned by PCR from genomic DNA from *Clostridium thermocellum* (ATCC 27405) and *Thermotoga maritima* (ATCC 43589). To clone isolated cellulase CD sequences without flanking ankyrin repeats, we used PCR primers 5′-AATGGGTCGCTCCGACAGGTGTGCCTTT TAACACAA-3′ and 5′-GGTGGTGGTCGCTCGAGAAAGTCACTCAAAG GATTCGG-3′ for (residues A1 to L363) C A and 5′-AATGGGTCGCTCCGACAGGTGTGCCTTT TAACACAA-3′ and 5′-GGTGGTGGTCGCTCGAGAAAGTCACTCAAAG GATTCGG-3′ for C12A (residues V10 to E265). Resulting PCR products were inserted into the expression vector pET-24a at BamHI and XhoI sites by using an In-Fusion kit (Clontech).

To create an expression construct that inserted cellulase coding se-
quences for CDs into consensus ankyrin domains, we began with a single consensus repeat (R) cloned into the expression vector pET-15b, described by Aksel et al. (28), and modified the repeat sequence by using the QuikChange system (Agilent Technologies) to insert Agel and KpnI restriction sites and form R13 (Fig. 1A). Cellulase catalytic domain sequences were amplified using the PCR primers 5'-GGGCGGCGGTGGCGGCGGTGCGAGGTGCGGCTTCTTACAACAAGG-3' and 5'-CTTCCCGAATCTCTGGTAGTCGATCCGCGGCGGCTGGTACCAGAAATCGGTG-3' for the Ca CD and 5'-GGAAACCTTATACGCGTTGCGGCCGTGGTACGTAGACCAAAAAAGCGGGCTGGTGAGGTAGGAGAAGGCCGCGGGTTGATAGGGCAGAAATCGGTG-3' for the C12A CD. Resulting PCR products were cloned onto the R13 consensus ankyrin repeat by using an In-Fusion kit (Clontech). Additional ankyrin repeats were added on both the 5' and 3' sides of the CD genes by using flanking BamHI and BglII sites, along with a distal HindIII site (28, 30). All DNA sequences were confirmed by gel electrophoresis and sequencing.

Protein expression and purification. Expression vectors were transformed into Escherichia coli BL21(DE3) cells. Cells were grown at 37°C in autoinduction medium for 24 h (31). Cells were collected by centrifugation and stored at −80°C. Cells were resuspended in 300 mM NaCl and 50 mM NaH2PO4 (pH 6.5; for CA constructs) or 300 mM NaCl, 50 mM Tris-HCl (pH 8.0; for C12A constructs). Cells were lysed with an Avestin EmulsiFlex-C3 homogenizer and were then treated with DNase I at 4°C for 1 h with stirring. The lysed cells were pelleted, and cell pellets were resuspended in the same buffers with 8 mM urea. Resuspended pellets were loaded onto a Ni column (Qiagen, Valencia, CA). After washing with low-salt buffer, bound proteins were concentrated by centrifugation for 1 h at stirring. The lysed cells were pelleted, and cell pellets were resuspended in the same buffers with 8 mM urea. Resuspended pellets were loaded onto a Ni column (Qiagen, Valencia, CA). The column was washed with the same buffered urea solution and then with 100 mM NaCl and 50 mM NaH2PO4 (pH 6.5; for CA constructs) or 100 mM NaCl, 50 mM Tris-HCl (pH 8; for C12A constructs) for refolding on the column. Proteins were eluted with 250 mM imidazole in the same salt and buffer, in the smallest volume possible. As a result of the cellulyotic activity of the CDs, we were unable to remove imidazole by dialysis and were unable to concentrate proteins by using filter-based concentrators. Thus, to facilitate soluble exchange of concentrated protein samples, we used ion-exchange chromatography. Proteins eluted from nickel columns were diluted 2- to 3-fold with the corresponding buffer at 50 mM and loaded onto ion-exchange columns (Q-Sepharose for CA constructs and SP-Sepharose for C12A constructs; Qiagen, Valencia, CA). After washing with low-salt buffer, bound proteins were eluted with a sharp buffer 1 M NaCl step, with discard of the first column volume. Protein concentrations were determined based on the A280 (32).

Spectroscopic methods. Circular dichroism measurements were made with an AVIV model 400 spectrometer (Aviv Associates, Lakewood, NJ). Far-UV circular dichroism spectra were collected with a 1-nm step size and a 30-s averaging time in a 1-mm-path-length quartz cuvette. Protein concentrations ranged from 5 μM to 20 μM. Near-UV circular dichroism spectra were collected with a 0.25-nm step size and a 1-s averaging time in a 1-cm cell, and the protein concentration was 20 μM. For each construct, spectra were generated by averaging five wavelength scans. Thermal unfolding transitions were monitored at 222 nm, with a 2°C step size, within the range of 20 to 98°C, in a thermal-resistant 10-μm quartz cuvette. Each step had a temperature equilibrium time of 3 min and a 30-s signal-averaging time. The protein concentration ranged from 1 to 2 μM. Thermal denaturations at pH 4.75 were carried out in 1 M NaCl, 50 mM Na-acetate buffer.

Reversible guanidine-HCl (GdnHCl)-induced unfolding transitions were performed with 50 mM acetate buffer (pH 4.75), 1 M NaCl, at 35°C. GdnHCl titration mixtures were generated using a Hamilton 500 titrator (Reno, NV). Measurements were made in a silanized 10-μm quartz cuvette. The protein concentration used was 2 μM. Each step on the native and denatured baseline had a 5-min equilibration time and a 30-s signal-averaging time. During the transition, we found conformational relaxation times to be slowed, and therefore we used a 15-min equilibration time with a 30-s signal-averaging time. Trp and Tyr fluorescence measurements were carried out by using wavelengths of 280 nm for excitation and 350 nm for emission in an Aviv model 105 automated titrating differential/ratio spectrophluorometer (Aviv Associates, Lakewood, NJ).

Two-state and multistate analyses of equilibrium GdnHCl-induced unfolding transitions. To analyze unfolding transitions (native [N], de-natured [D], and intermediate [I]), we assumed signals resulted from a population-weighted average of each state, and the free energy of unfolding was assumed to vary linearly with denaturant concentration (33–38). The data for equilibrium unfolding were fitted by using a nonlinear least-squares analysis, and the thermodynamic parameters were estimated from the fit (see the supplemental material for a further description of the methods).

Cellulase assays. Cellulase activity was measured on two substrates: CMC (soluble, chemically modified cellulose) and RAC (prepared by acid-base-treated insoluble microcrystalline cellulose) (39). For preparation of low-viscosity CMC solution, powdered CMC (degree of polymerization [DP], 400; degree of substitution [DS], 0.65 to 0.9; Sigma-Aldrich) was dissolved to 2% (wt/vol) in 1 M NaCl–50 mM Na-acetate (pH 4.7). Cellulase activities were measured through a modified version of the dinitorosalic acid (DNS) method described by Ghose (40). All activity measurements reported here were performed at pH 4.7. We chose this pH because our constructs showed the highest activity at this pH, although the activity was similar at pH 6.0 (41–43). Moreover, this pH is close to that used in the IUPAC protocol for measuring cellulase activity (40).

We adapted the DNS method to a high-throughput 96-well format. The 30-min degradation step was carried out in a thermally resistant plate sealed with sticky aluminum foil sheets (Corning, NY). Plates were incubated in a Veriti thermocycler (Applied Biosystems, CA) at temperatures ranging from 20°C to 95°C. The lid was kept at 2°C above the well temperature to avoid condensation. The final concentration of CMC used was 1%. All assays were carried out with 4 μmol of enzyme. After the degradation step, 140 μl of DNS solution (40) was added, and the samples were subsequently incubated at 100°C for 10 min in the thermocycler, with a lid temperature of 105°C. Samples were transferred to an optically clear plate, and the absorbance was read in a GloMax Modulus microplate reader (Promega, WI) at 510 nm. (Although the product of the DNS reaction is usually detected at 540 nm, we measured product formation at 510 nm, as we found this wavelength to have an improved signal-to-noise ratio in the plate reader.) Each measurement was made in triplicate in the same plate and repeated in a minimum of two separate plates. Absorbances were corrected by using values from undigested, identically processed substrates (i.e., substrates without enzyme). We included a set of glucose standards on each plate. The scaling factor between absorbance at 510 nm and the micromoles of glucose hydrolyzed was 5.23 ± 0.05 μmol (mean ± standard error).

RESULTS

Determination of insertion sites and insertion target enzymes. To identify viable insertion sites for cellulase CDs within an ankyrin array, we used a hidden Markov model (HMM) logo from the Pfam database (Fig. 1A). HMM logos depict the degree of conservation, insertion, and deletion at each position in a multiple sequence alignment. For the ank protein family, there are 5,717 sequences in the Pfam database (version 26.0) (44). Of the 33 positions depicted in the HMM logo, positions 13 and 23 (boxed in pink) represent the most probable sites for insertion. We chose position 13 to insert cellulase CDs. To create insertion sites, we modified a gene encoding the consensus sequence R (depicted below the HMM logo in Fig. 1A) to include two restriction sites (Agel and KpnI) at position 13. This substitution introduced glycines at the site of insertion (modified consensus sequence R13 in Fig. 1A). The R hydrophobic consensus repeats are capped by N and C repeats that have four hydrophilic substitutions to maintain...
solubility (29). Since the driving force for ankyrin repeat folding has been shown to be derived from the interfaces between repeats (28), arrays of two or more repeats are needed to achieve a stable fold, since this is the minimum number of repeats needed to create an intact interface. Thus, for the constructs here, we used a minimum of five repeats (2.5 repeats on each side), leaving an intact interface on each side of the inserted domain and the possibility to form a stable interface across the inserted CD (28, 29).

The cellulase target organisms (C. thermocellum and T. maritima) were chosen because they are known to encode a number of stable cellulase enzymes, some of which can be expressed in E. coli (6, 45–48). From these organisms, we cloned and expressed several cellulases. From these initial studies (data not shown), the CDs with the highest expression levels and cellulolytic activities form a stable interface across the inserted CD (28, 29).

Effects on thermal stability of insertion of the cellulase CA catalytic domain into consensus ankyrin repeats. Cellulose deg-
The fluorescence signal is expected to report exclusively on the cellulase domain, as the consensus has no ordered Trp residues. The circular dichroism at 222 nm should report the folding of both domains (49). Transition midpoint temperatures ($T_{m}$) are reported in Table 1.

For the CA CD, there is a single transition for both FL and circular dichroism, with a midpoint of 75°C at pH 6.5. The FL native baseline is strongly sloped, suggesting that there may be a change in local environment of the Trp residues in the CA CD prior to the major thermal transition (Fig. 3A). R$_{2.5}$-CA-R$_{2.5}$ showed the same behavior as the isolated CA CD, with a single transition at 75°C and a sloped FL native baseline (Fig. 3B). The thermal denaturation of R$_{3.5}$-CA-R$_{3.5}$ by fluorescence also showed a single transition with a midpoint of 76°C, although when followed by circular dichroism, there appeared to be two separate, albeit incompletely baseline separated transitions, with midpoints of 76°C and 84°C (Fig. 3C). The multistage unfolding transitions of R$_{3.5}$-CA-R$_{3.5}$ suggest that CA and ankyrin repeat regions may be unfolding at different temperatures.

Effects upon lowering the pH on the thermal stability of scaffolded cellulase CDs. Since cellulase degradation is more effective in the range pH 4.7 to 4.8 than at neutral pH, we examined the effect of lowering the pH on the circular dichroism-monitored thermal denaturation of the CDs and consensus ankyrin constructs. The consensus ankyrin domain was destabilized at lower pHs (see Fig. S3 in the supplemental material). In contrast, thermal denaturation of the isolated CA CD was similar at pH 4.75 and pH 6.5 (Fig. 4A): a $T_{m}$ of 75°C was observed at both pH values, but the steepness was increased at low pH. Like the isolated consensus ankyrin domains, R$_{2.5}$-CA-R$_{2.5}$ and R$_{3.5}$-CA-R$_{3.5}$ have lower thermal denaturation midpoints at pH 4.75 than at pH 6.5 (Fig. 4B and C; Table 1). To examine the structural origins of the transitions observed for the R$_{2.5}$-CA-R$_{2.5}$ construct, we determined the NaCl dependence of thermal denaturation. NaCl changes the stability of the consensus ankyrin domains but has little effect on the CA CD. Increasing the NaCl concentration to 1 M (the concentration used in the present study) led to a flattening of the native baseline but had no effect on the midpoint of the transition (Fig. 4D).

The thermal denaturation transition of the C12A CD occurs at very high temperatures at both pH 8.0 and 4.7, relative to the CA CD. The C12A-ankyrin fusions were less thermally stable than the CA fusions. Unlike R$_{2.5}$-CA-R$_{2.5}$ and R$_{3.5}$-CA-R$_{3.5}$, there was considerable unfolding detected by circular dichroism for the C12A CD. Moreover, the isolated C12A CD was stable to very high temperatures. In the context of ankyrin repeats, it appeared to be slightly destabilized.

Reversible GdnHCl denaturations of consensus ankyrin and the CA CD. To obtain a more accurate thermodynamic description of the effects of cellulase fusion with ankyrin repeats and to provide quantitative analysis of stability changes in protein engineering studies of cellulases, we sought to identify conditions for reversible equilibrium unfolding. Because we can refold these consensus ankyrin cellulase scaffolds from inclusion bodies by using chemical denaturation, we surmised that strong denaturants should be able to promote reversible folding. We also sought conditions that were as close as possible to those we used to assay cellulase activity (i.e., pH 4.75 and 1 M NaCl). We found that at moderately elevated temperatures (e.g., 35°C, pH 4.75) we obtained identical folding and unfolding transitions for the catalytic domain of CelA (CA CD), indicating a reversible, pathway-independent process (see Fig. S4 in the supplemental material).
As determined by circular dichroism and fluorescence spectroscopy, the GdnHCl denaturation profile of the CA CD involves a single broad transition. However, the transition followed by fluorescence did not match that by circular dichroism, supporting the possibility that the CA CD unfolds through intermediates. Consistent with this, for both probes the GdnHCl dependence of the transition (apparent $m$ values) were much smaller than expected for a protein of this size (363 residues). To capture this complexity, we globally fitted both curves with a three-state model that assumed a linear dependence of free energy on GdnHCl concentration. (In this model, we assumed the spectroscopic signals of the intermediate to be GdnHCl independent.) This model gave a good fit to both transitions (Fig. 5 and 6) and provided quantitative estimates of unfolding midpoints ($C_m$), free energies of un-
folding in the absence of GndHCl (ΔG°_H2O), and denaturant dependencies (m values) (Table 2). This minimal three-state model provides a quantitative comparison to GndHCl unfolding for scaffolded CA constructs, although it may miss further complexities in the unfolding of this rather large protein. (The total m value for the N-to-D reaction was calculated to be 2.30 kcal mol⁻¹ M⁻¹ [m₂ + m₃], compared to a predicted m value of 12.3 kcal mol⁻¹ (Table 2).)

To help analyze the reversible GndHCl unfolding transitions of CA-ankyrin fusions, we examine the guanidine-induced unfolding transition of the consensus ankyrin NRR13RC under these conditions. NRR13RC GndHCl denaturation by circular dichroism follows a single transition that is well-fitted with a two-state model (Fig. 5). The slope of the transition was high, with an estimated m value of 7.2 kcal mol⁻¹ M⁻¹ and an estimated free energy of 8.2 kcal mol⁻¹ (Table 2). These thermodynamic parameters are consistent with the expected values for unfolding of a five-repeat consensus ankyrin protein at low pH and 35°C (28).

Reversible GndHCl denaturations of ankyrin-CA fusions.

To assess how insertion of the cellulase CA CD into the consensus ankyrin domain affects the reversible unfolding of the CA CD and ankyrin domains, we measured GndHCl-induced unfolding transitions for constructs in which CA CD was inserted into five- and seven-repeat consensus ankyrin domains (R₂₅-CA-R₂₅ and R₃₅-CA-R₃₅) by circular dichroism. R₂₅-CA-R₂₅ unfolds in a multistage transition that can be fitted with a four-state model, N ↔ I₁ ↔ I₂ ↔ D (Fig. 5). As with CA and NRR13RC, the native and denatured baselines were well defined. There were three clear transitions with calculated C_m values of 1.1, 2.3, and 3.3 M GndHCl. The first transition was sharp (m₁ value, 7.1 kcal mol⁻¹ M⁻¹), with a free energy of 8.1 kcal mol⁻¹. The parameters determined for this first transition closely matched those of NRR13RC. The second transition was broader (m₂ value, 3.35 kcal mol⁻¹ M⁻¹) and thus was less well-defined than the first transition. The estimated unfolding free energy for this second transition (I₁ ↔ I₂) was 7.5 kcal mol⁻¹. The sharpness of the third transition (m₃ value) was 4.25 kcal mol⁻¹ M⁻¹, with an estimated unfolding free energy (I₂ ↔ D) of 7.5 kcal mol⁻¹ (Table 2). The determined C_m and C_m values closely matched those determined for the CA CD, indicating that these transitions correspond to the CA CD.

GndHCl unfolding of the R₂₅-CA-R₂₅ constructs followed by fluorescence involves a single transition that can be fitted with a two-state model (Fig. 6B). The estimated m_F value was 1.9 kcal mol⁻¹ M⁻¹, with a free energy of 4.5 kcal mol⁻¹ (Table 3). The estimated fluorescence midpoint (C_mF) was close to the estimated
C<sub>m</sub> values (I<sub>1</sub> ↔ I<sub>2</sub> transition) as measured by circular dichroism, supporting the evidence for partial unfolding of the CA CD domain.

To test the effect of adding additional flanking repeats, we measured the chemical denaturation of R<sub>3.5</sub>-CA-R<sub>3.5</sub> by circular dichroism. Addition of repeats should help in interpreting the conformational changes in the unfolding transitions of the consensus cellulase fusions. The R<sub>3.5</sub>-CA-R<sub>3.5</sub> construct unfolded in two separate stages (Fig. 5). These transitions could be fitted with a three-state model (Fig. 5). As with the other constructs, both native and denatured baselines were well-defined. There were two clear transitions, with calculated C<sub>m</sub> values of 2.1 and 3.4 M GndHCl. The first transition was sharp (m<sub>1</sub> value, 6.37 kcal mol<sup>−1</sup> M<sup>−1</sup>), with an unfolding free energy of 13.47 kcal mol<sup>−1</sup>. The second transition was broader (m<sub>2</sub> value, 4.42 kcal mol<sup>−1</sup> M<sup>−1</sup>), with an unfolding free energy of 15.0 kcal mol<sup>−1</sup>. The two transitions of R<sub>3.5</sub>-CA-R<sub>3.5</sub> closely matched the two high-denaturant transitions seen for R<sub>2.5</sub>-CA-R<sub>2.5</sub>-CA-R<sub>3.5</sub>.

R<sub>3.5</sub>-CA-R<sub>3.5</sub> has a fluorescence transition that can be fitted with a two-state model (Fig. 6C). Like R<sub>2.5</sub>-CA-R<sub>2.5</sub>, the unfolding transition of R<sub>3.5</sub>-CA-R<sub>3.5</sub> was moderately sharp (m<sub>1</sub> value, 2.13 kcal mol<sup>−1</sup> M<sup>−1</sup>), with a free energy of 5.13 kcal mol<sup>−1</sup>. These parameters were similar to the fluorescence values for R<sub>2.5</sub>-CA-R<sub>2.5</sub> with calculated C<sub>m</sub> values of about 2.4 M GndHCl for both constructs (Table 3).

Effects on cellulolytic activity of insertion of cellulase catalytic domains into consensus ankyrin repeats. Degradation of cellulose was determined by the DNS method. The DNS method is a popular but somewhat cumbersome and lengthy method and requires rather large volumes and quantities of enzymes (50). To quickly measure and accurately compare the activities of multiple enzymes, with use of multiple replicates for different conditions and substrates, we developed a 96-well plate assay that employed the DNS reaction (described in Materials and Methods). In addition to its speed and high-throughput format, the use of low-volume reactions in 96-well plates minimizes enzyme consumption, shortens thermal equilibration times, and minimizes evaporative loss.

We used this assay to determine the activity of CA and C12A constructs (both for isolated CDs and for CDs in consensus ankyrin scaffolds of different lengths) against CMC (Fig. 7A and B) and RAC (Fig. 7C and D), both at 50°C and 75°C. Typically, the DNS assay is carried out at 50°C. However, we found these enzymes to have increased activities at higher temperatures, consistent with previous studies (42, 43); the 75°C measurements are presented to provide activity measurements near the various maxima. (For both substrates, activities were measured over a temperature range of 45°C to 95°C (see Fig. S5 in the supplemental material for CMC data). As expected, the activities were much higher against CMC than RAC, consistent with CMC being soluble whereas RAC is insoluble amorphous cellulose.

Under all the conditions tested, the CA CD had the same activity alone as when inserted into the consensus ankyrin scaffold. However, the C12A CD showed increased activity upon insertion into the ankyrin scaffold under all conditions tested. The largest increase in activity resulting from the C12A CD insertion into the ankyrin scaffold was found at 75°C against CMC (Fig. 7B); R<sub>2.5</sub>-C12A-R<sub>2.5</sub> and R<sub>3.5</sub>-C12A-R<sub>3.5</sub> showed a 6.2-fold increase in activity relative to the isolated C12A CD in CMC at 75°C. Against RAC, these constructs showed a 2.5-fold increase in activity (Fig. 7D).

### DISCUSSION

Cellulose degradation is one of the key problems that have to be overcome to achieve large quantities of commercially viable production of cellulosic ethanol (51). Here, we focused on increasing cellulase thermostability and activity by inserting cellulase catalytic domains into a stable scaffold. Our long-term goal is to use the ankyrin scaffold to mimic the multivalency of the cellulosome by displaying multiple cellulase CDs, allowing for a high level of control of CD identity and separation. This approach is similar to recent protein engineering methods that have combined multiple cellulase domains by using noncovalent interactions, such as minicellulosome complexes (52, 53) and oligomeric architectures termed “rosettazymes” (54). Our approach differs in that our arrays are single polypeptide chains with cellulase domains internal to the scaffolding protein.

**Structural integrity of scaffolded CDs.** We tested the effects on the structures of scaffolding of CA and C12A within consensus ankyrin domains by using far-UV circular dichroism spectroscopy. Far-UV circular dichroism spectra indicated that the consensus ankyrin repeats remained folded upon cellulase insertion at room temperature (Fig. 2; see also Fig. S1 in the supplemental material). Moreover, the α-helical structures seen for the scaffolded CA constructs were consistent with the CA CD domain retaining a folded (α-helical) structure. Though we cannot make the same conclusion for the C12A CD domain, the scaffolded C12A CD domain retained the same overall tertiary structure, as seen with near-UV circular dichroism (see Fig. S2 in the supple-
The CD structure was maintained at pH 4.7, the pH at which we measured cellulase activity. As a further indication that the CA and C12A CDs remain structured in the ankyrin scaffold, both scaffolded CDs remained enzymatically active against a range of substrates.

**Thermostability of scaffolded CDs.** By comparing the stabilities of the CD-ankyrin fusions with the stabilities of the separate scaffolds and CDs, we can infer the contributions of each domain to thermal denaturation transitions. Furthermore, we can determine whether insertion of a CD into a consensus ankyrin domain is stabilizing, destabilizing, or neutral. The CA CD $T_m$ (75°C) corresponds to the temperature of maximal cellulolytic activity, similar to previous results (55). R2.5-CA-R2.5 behaves as a single unit, with a $T_m$ matching that of the isolated CA CD, but it is midway between the $T_m$ of the full five-repeat ankyrin scaffold and the three-repeat N- and C-terminal fragments (see Fig. S3 in the supplemental material). We concluded that when the CA CD is folded, the two halves of the consensus ankyrin scaffold stabilize each other. However, when the CA CD unfolds, the ankyrin scaffold unfolds (Fig. 4B and 6A). At pH 6.5 (Fig. 4B; see also Fig. S3), it appears that the thermal unfolding of the CD significantly destabilizes the flanking ankyrin repeats. We propose that this destabilization results from the high entropic cost of loop closure that would be required to form an ankyrin repeat interface across an unfolded polypeptide of 379 residues (Fig. 6A).

For R3.5-CA-R3.5, the low-temperature unfolding transition matched the FL transition midpoint, indicating that the CD of R3.5-CA-R3.5 unfolds first (Fig. 3C). The high-temperature midpoint is likely a result of unfolding of the flanking ankyrin repeats. At pH 6.5 and 1 M NaCl, the seven-repeat consensus ankyrin domain remained folded at a temperature near 100°C; however, a three-repeat ankyrin had a melting point of 70°C (see Fig. S3B). As with R2.5-CA-R2.5, these data suggest that two sides of the ankyrin domain can remain folded and coupled when the CA CD is folded. Surprisingly, there appears to be some coupling of the N- and C-terminal ankyrin repeats, even when the CA CD is unfolded (Fig. 8).

Decreases in the pH destabilize the scaffolded CA CD but not the isolated CD (Fig. 4A, B, and C). Based on thermostability criteria alone, the ankyrin scaffold might be further improved by making it insensitive to pH by sequence substitution. However, if enzyme stability were enhanced by conformational flexibility of the adjacent scaffold, a rigid scaffold may also have detrimental effects on reactivity (56).

**FIG 7** Cellulolytic activities of the ankyrin-cellulase constructs. Cellulose degradation assays products were measured via a DNS assay in a 96-well plate format, with CMC (A and B) and RAC (C and D) as the substrates at 50°C (A and C) and 75°C (B and D). Cellulolytic activity of the CA CD was not affected by insertion into the consensus ankyrin domain. Cellulolytic activity of the C12A CD was enhanced by insertion into the consensus ankyrin domain. The scaling factor between absorbance at 510 nm and micromoles of glucose hydrolyzed is 5.23 μmol.
Variations in the NaCl concentration affect neither the thermal transition of the CA CD nor the midpoint of R2.5-CA-R2.5; however, the native baseline of R2.5-CA-R2.5 becomes sloped at lower salt concentrations. Salt stabilizes the consensus ankyrin domain (T. Aksel and D. Barrick, unpublished data) (57) but has no effect on the thermal denaturation of the isolated CA CD in the temperature range tested. Thus, the effect of NaCl on the native baseline is likely to result from prevention of fraying of the flanking ankyrin scaffold, and it may impart scaffold rigidity at higher temperatures (Fig. 4D).

Despite efforts, we were unable to follow the thermal denaturation of the C12A constructs due to aggregation at high temperatures. Nonetheless, the high thermal resistance of C12A from T. maritima is greater than for any known glycoside hydrolase family 12 (GH12) domain, natural or engineered (for a review on thermal melting points of GH12 hydrolases, see references 5, 48, 58, 12 (GH12) domain, natural or engineered (for a review on ther-

FIG 8 Thermal unfolding model for five-repeat ankyrin-cellulase constructs. (A) R2.5-CA-R2.5 thermal unfolding follows a two-state model, with both the CA CD and ankyrin repeat scaffold unfolding as a single unit. (B) For R3.5-C12A-R3.5, the consensus ankyrin repeats unfold first, in the temperature range of 50 to 75°C, and the C12A CD remains folded. The C12A CD precipitates at higher temperatures.

Unlike the CA-ankyrin fusions, the thermal denaturation of C12A fused to a consensus ankyrin scaffold did not show well-defined (low-temperature) native baselines, indicating that with C12A the consensus ankyrin segments do not remain as a single cohesive unit (Fig. 4F and G). R2.5-C12A-R2.5 and R3.5-C12A-R3.5 are expected to form one and two intact interfaces between the repeats, respectively, on each side of the cellulase. However, the major unfolding transition that precedes the thermal denaturation of the C12A CD appears to result from unfolding of two independent (i.e., noninteracting) ankyrin arrays (Fig. 6B). This is likely to result from the larger distance between the N and C termini in folded C12A (about 30Å), which is likely to be too large for the consensus repeats to remain as a single unit. The loss of the interface in the middle of the consensus ankyrin scaffold is responsible for the destabilization of this domain. Because the N- and C-terminal capping repeats are likely to be uncoupled and are of different stability, it is not surprising that we saw a broad loss of signal.

Reversible GndHCl denaturation of the CA CD. Though there have been a number of studies of chemical stability and thermostability of cellulase catalytic domains, such studies have generally involved irreversible unfolding, precluding an equilibrium thermodynamic analysis (55). By comparing unfolding and refolding transitions using GndHCl as a denaturant, we were able to find conditions under which we could measure reversible unfolding. The circular dichroism and fluorescence transitions differed, demonstrating that partly folded states are populated in the transition region (Fig. 5 and 9A; Table 2). The multistate fitting gives a quantitative basis for analysis of scaffolded CA constructs. We noted that even when summed, the combined m values for the two transitions were lower than expected for a protein of 363 residues, suggesting that there may be additional intermediates not accounted for in the three-state model (61).

Reversible GndHCl denaturation of R2.5-CA-R2.5. For R2.5-CA-R2.5, we analyzed the GndHCl-induced unfolding transitions globally by using a four-state model. The low GndHCl transition value matched the transition seen for a five-repeat consensus ankyrin scaffold, NRR13RC (Fig. 5), as did the fitted parameters (Table 2). Therefore, it is likely that in the construct R2.5-CA-R2.5, the repeats are folded, behaving as a single unit, and are not destabilized by insertion of the CA CD (Table 2). However, since a five-repeat ankyrin domain is less stable than the CA CD, the consensus ankyrin repeats unfold first. The Cm of the middle transition of R2.5-CA-R2.5 closely matched the fluorescence Cm and the Cm2 of the CA CD (Table 3), which suggested that the CA CD of R2.5-CA-R2.5 partially unfolds in this transition (Fig. 6). Importantly, the GndHCl-induced unfolding followed by fluorescence of R2.5-CA-R2.5 showed a sharper transition and a better-defined native baseline than for the CA CD alone, suggesting that the consensus ankyrin array stabilizes the CA CD at low GndHCl concentration.
concentrations. Furthermore, the $C_m$ of the high-GndHCl transition for R$_{2.5}$-CA-R$_{2.5}$ matched the $C_m$ determined for the CA CD, which suggests that the CA CD unfolds through a stable intermediate, with the fluorescence showing the transition from the native to the intermediate state and the circular dichroism reporting on both transitions (Fig. 5; Table 2). Taken together, these results indicated that consensus repeats of the construct R$_{2.5}$-CA-R$_{2.5}$ unfold at lower GndHCl concentrations, followed by multistate unfolding of the CA CD (Fig. 9B).

Reversible chemical denaturation of R$_{3.5}$-CA-R$_{3.5}$. For R$_{3.5}$-CA-R$_{3.5}$, we observe a multistate GndHCl-induced unfolding transition, similar to that seen for R$_{2.5}$-CA-R$_{2.5}$. The addition of two repeats is expected to stabilize the consensus ankyrin array, since two new repeat interfaces should be formed (28, 29). Therefore, it was not surprising that the low-GndHCl transition corresponding to a five-repeat consensus ankyrin array ($C_m$ of about 1 M) was not seen for the longer construct. Instead, for R$_{3.5}$-CA-R$_{3.5}$, the transitions seemed to match the high-denaturant transitions seen for R$_{2.5}$-CA-R$_{2.5}$ (Table 2). Furthermore, the ellipticity change for the R$_{3.5}$-CA-R$_{3.5}$ low-denaturant transition was similar to the summed ellipticity change of the two lowest denaturant transitions of R$_{2.5}$-CA-R$_{2.5}$, suggesting a loss of structure of the consensus repeats and partial unfolding of the CA CD (Fig. 5). The partial unfolding of the CA CD in R$_{3.5}$-CA-R$_{3.5}$ was supported by the finding that the fluorescence midpoint closely resembled the value for the low-GndHCl transition followed by circular dichroism (Tables 2 and 3). As with R$_{3.5}$-CA-R$_{3.5}$, the sharper transition and better-defined native baseline, compared to those for the CA CD alone, suggested that the consensus ankyrin scaffold stabilizes the CA CD of R$_{3.5}$-CA-R$_{3.5}$ at low GndHCl concentrations.

The $C_m$ value of the highest GndHCl transition of R$_{3.5}$-CA-R$_{3.5}$ was very close to the $C_m$ values for both the CA CD and R$_{2.5}$-CA-R$_{2.5}$, reflecting the complete unfolding of the CA CD (Fig. 9C). Taken together, these data suggest that both the CA CD and the consensus ankyrin scaffold remain folded when fused. The improved baseline imparted by the consensus ankyrin domain allowed us to better resolve the unfolding transitions that the CA CD undergoes upon GndHCl-induced denaturation. Furthermore, the consensus ankyrin repeats remained folded as a single cohesive unit upon insertion of the CA CD, a 363-residue globular domain.

Effects of the ankyrin scaffold on cellulolytic activity. The N and C termini of the CA CD are in close proximity and are distant from the active site. The structural integrity of the CA CD was supported by the finding that the cellulolytic activity of the CA CD against both CMC and RAC substrates remained unaltered upon removal of the CA CD undergoes upon GndHCl-induced denaturation. Further, the consensus ankyrin repeats decrease the rigidity of the CA CD, which may result in higher cellulolytic activity.

Conclusions. The close proximity of the termini of the embedded cellulase CDs seems to be critical for the structural integrity and stability enhancement of the scaffold and CDs. However, the integrity of the scaffold appears not to be necessary for higher activity. These results showed that a consensus ankyrin domain can be engineered to serve as a folded (and for C12A at high temperature, unfolded) scaffold to array folded, active cellulase enzymes. These consensus ankyrin scaffolds may be used as versatile platforms to combine cellulase domains of various reactivities. This scaffolding approach may provide insight into how cellulase CDs behave when attached to other domains found in nature, such as dockerins and cellulose-binding modules.

ACKNOWLEDGMENTS
We thank Bertrand Garcia-Moreno, Mario Amzel, Richard Cone, Juliette Lecomte, Jan Hoh, and Brian Barr for prolific discussions during the development of the work. We also thank Tural Aksel for providing the consensus ankyrin repeats and the altered PET-15B+ vector.

This work was supported by doctoral fellowships to E.S.C. from the Portuguese FLAD/Fulbright and from the Portuguese Foundation for Science and Technology (SFRH/BD/36119/2007), by a training grant from the Department of Energy (DE-FG02-04ER25626), by NIH grant GM068462 to D.B., by the Environment, Energy, Sustainability, and Health Institute at Johns Hopkins, and by generous funding from the Office of the Provost at Johns Hopkins University.

REFERENCES


