Convergent evolution of processivity in bacterial and fungal cellulases

Taku Uchiyama[a,1], Takayuki Uchihashi[b,c,1], Akihiko Nakamura[d], Hiroki Watanabe[c], Satoshi Kaneko[e], Masahiro Samejima[a], and Kiyohiko Igarashi (五十嵐 圭日子)[a,f,2]

[a]Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan; [b]Department of Physics and Structural Biology Research Center, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan; [c]Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Aichi 444-8787, Japan; [d]Institute for Molecular Science, National Institutes of Natural Sciences, Aichi 444-8787, Japan; [e]Department of Subtropical Biochemistry and Biotechnology, Faculty of Agriculture, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan; and [f]Protein Discovery and Engineering Team, VTT Technical Research Center of Finland Ltd., FI-02044 VTT, Espoo, Finland

Edited by Daniel J. Cosgrove, Pennsylvania State University, University Park, PA, and approved June 28, 2020 (received for review June 3, 2020)

Cellulose is the most abundant biomass on Earth, and many microorganisms depend on it as a source of energy. It consists mainly of crystalline and amorphous regions, and natural degradation of the crystalline part is highly dependent on the degree of processivity of the degrading enzymes (i.e., the extent of continuous hydrolysis without detachment from the substrate cellulose). Here, we report high-speed atomic force microscopic (HS-AFM) observations of the movement of four types of cellulases derived from the cellulolytic bacteria *Cellulomonas fimi* on various insoluble cellulose substrates. The HS-AFM images clearly demonstrated that two of them (*Cf*Cel6B and *Cf*Cel48A) slide on crystalline cellulose. The direction of processive movement of *Cf*Cel6B is from the nonreducing to the reducing end of the substrate, which is opposite that of processive cellulase Cel7A of the fungus *Trichoderma reesei* (*Tr*Cel7A), whose movement was first observed by this technique, while *Cf*Cel48A moves in the same direction as *Tr*Cel7A. When *Cf*Cel6B and *Tr*Cel7A were mixed on the same substrate, "traffic accidents" were observed, in which the two cellulases blocked each other's progress. The processivity of *Cf*Cel6B was similar to those of fungal family 7 cellulases but considerably higher than those of fungal family 6 cellulases. The results indicate that bacteria utilize family 6 cellulases as high-processivity enzymes for efficient degradation of crystalline cellulose, whereas family 7 enzymes have the same function in fungi. This is consistent with the idea of convergent evolution of processive cellulases in fungi and bacteria to achieve similar functionality using different protein foldings.

cellobiohydrolase | processivity | crystalline cellulose | biomass utilization | cellulase

Efficient production of biofuels and biobased products from biomass is one of the biggest challenges for the circular bioeconomy because of its potential to replace fossil-based fuels and products with biomass-based renewable fuels and products. In nature, biomass is broken down by the collaborative action of multiple functionally complementary enzymes that are secreted by biomass-degrading microorganisms. Cellulose, which is the most abundant biomass, is the main structural polysaccharide of plant cell walls and consists of linear polymers of β-1,4-linked D-glucose units. The chains are stabilized by intra- and intermolecular hydrogen bonds, forming an insoluble structure that is ∼70% crystalline (1). One of the most important factors influencing the efficiency of degradation of recalcitrant crystalline cellulose is the processivity of the cellobiohydrolases (CBHs) that degrade crystalline cellulose to soluble products. Processivity is defined as the ability of the cellulases to catalyze continuous degradation of a cellulose chain through successive hydrolytic catalytic reactions without dissociation from the chain. Recently, our research group succeeded in visualizing processive movements of fungal glycoside hydrolase (GH) family 7 CBHs by means of high-speed atomic force microscopy (HS-AFM). We found that *Trichoderma reesei* (*Tr*Cel7A) and *Phanerochaete chrysosporium* (*Pc*Cel7C and *Pc*Cel7D) cellulase molecules move unidirectionally from the reducing end to the nonreducing end of the chain on crystalline cellulose (2–4). Cellulose-degrading filamentous fungi also secrete other CBHs belonging to GH family 6, but their processive movement could not be visualized by HS-AFM under the same conditions, suggesting that family 6 and 7 CBHs differ in their ability to remain on the crystal surface (5). Comparison of their three-dimensional (3D) structures indicated that the catalytic domain of fungal family 6 CBHs is covered by a short loop, while that of family 7 CBHs has four loops covering the active site. These observations suggest that cellulolytic fungi secrete two types of CBHs, of which one binds strongly to the cellulose chain (like *Tr*Cel7A, *Pc*Cel7C, and *Pc*Cel7D) and shows processive degradation, while the other binds weakly to the cellulose chain (like *Tr*Cel6A), showing little or no processivity. However, less is known about bacterial cellulases, and the processivity of bacterial CBHs has never been confirmed by direct observation.

In the present study, we observed the processive movement of four bacterial cellulases, *Cf*Cel6A, *Cf*Cel6B, *Cf*Cel9A, and *Cf*Cel48A [formerly known as CenA (6), CbhA (7), CenB (8),

Significance

Some cellulases exhibit "processivity": the ability to degrade crystalline cellulose through successive hydrolytic catalytic reactions without the release of the enzyme from the substrate surface. We previously observed the movement of fungal processive cellulases by high-speed atomic force microscopy, and here, we use the same technique to directly observe the processive movement of bacterial cellobiohydrolases settling a long-standing controversy. Although fungal and bacterial processive cellulases have completely different protein folds, they have evolved to acquire processivity through the same strategy of adding subsites to extend the substrate-binding site and forming a tunnel-like active site by increasing the number of loops covering the active site. This represents an example of protein-level convergent evolution to acquire the same functions from different ancestors.

Author contributions: M.S. and K.I. designed research; T. Uchiyama, T. Uchihashi, and H.W. performed research; A.N. and S.K. contributed new reagents/analytic tools; T. Uchiyama, T. Uchihashi, and K.I. analyzed data; and T. Uchiyama and K.I. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]T. Uchiyama and T. Uchihashi contributed equally to this work.

[2]To whom correspondence may be addressed. Email: aquarius@mail.ecc.u-tokyo.ac.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011366117/-/DCSupplemental.

First published August 3, 2020.

and CbhB (9), respectively] from the cellulolytic actinobacteria *Cellulomonas fimi.* This microorganism is one of the best-known aerobic cellulolytic bacteria and expresses three endoglucanases, two exocellobiohydrolases and one xylanase in the supernatants of cultures grown on cellulose (10). *Cf*Cel6A and *Cf*Cel9A are thought to be endoglucanases (6, 8), while *Cf*Cel9A has a high amino acid sequence similarity to the possibly processive endoglucanase *Tf*Cel9A from *Thermobifida fusca* (11). *Cf*Cel6B and *Cf*Cel48A are CBHs (7, 9), and both have been predicted to have processivity. We also compare the findings for fungal and bacterial cellulases and discuss the requirements for processivity, as well as the likelihood of convergent evolution of processive cellulases in bacteria and fungi.

Results

General Characteristics of the Four *C. fimi* Cellulases. The cellulolytic activities of the four cellulases were tested on several insoluble substrates, i.e., phosphoric acid–swollen cellulose (PASC), microcrystalline cellulose (Avicel), and crystalline celluloses I_α and III_I from green algae, and the amounts of reaction products were measured by high performance liquid chromatography (HPLC). Crystalline cellulose in nature has structures called I_α and I_β (12, 13), which differ in interlayer chain stacking. When I_α and I_β are treated with supercritical ammonia, the chain stacking arrangement changes to a completely different structure, called III_I (14). Cellulose III_I forms staggered layers with intra- and interlayer hydrogen-bonding interactions, unlike cellulose I_α, in which the chains pack into flat layers. Because of this change, cellulose III_I has a larger hydrophobic surface than cellulose I_α, making it more susceptible to attack by cellulase (15). Cellulose III_I is highly susceptible to degradation by cellulase while maintaining high crystallinity (16) and therefore is suitable for observing the degradation of crystalline cellulose by cellulase by means of HS-AFM. The hydrolytic progress curves are shown in *SI Appendix*, Fig. S1 *A–D*. The major product of *Cf*Cel6A, *Cf*Cel6B, and *Cf*Cel48A was cellobiose, whereas the major products of *Cf*Cel9A were glucose and cellotriose, in addition to cellobiose (*SI Appendix*, Fig. S1 *E–H*). *Cf*Cel6A degraded PASC and Avicel more efficiently than did the other three cellulases, while *Cf*Cel9A efficiently hydrolyzed crystalline cellulose I_α. *Cf*Cel6B degraded crystalline cellulose III_I more efficiently than did the other three enzymes. *Cf*Cel48A showed the lowest degradation activity toward insoluble cellulose substrates among the four cellulases.

In the exocellulase reaction, cellobiose is produced mainly through processive hydrolysis, whereas glucose and cellotriose are the primary products released at initial hydrolysis events. Therefore, processivity can be assessed in terms of the ratio of cellobiose/(glucose + cellotriose), as discussed previously (4, 17). Processivity ratios above 10 were achieved for all four insoluble substrates by *Cf*Cel6B and *Cf*Cel48A; these ratios are within the range measured for processive GH family 7 cellulases (4, 17). In contrast, ratios under 5 and under 0.7 were observed for *Cf*Cel6A and *Cf*Cel9A, respectively. These results suggest that *Cf*Cel6B and *Cf*Cel48A are highly processive cellulases, whereas *Cf*Cel6A and *Cf*Cel9A are less processive cellulases.

The random hydrolysis catalyzed by an endocellulase results in a greater decrease of viscosity compared with the restricted action of an exocellulase, and therefore, plots of viscosity-lowering versus incubation time during the hydrolysis of carboxymethyl cellulose (CMC) are commonly used to characterize exocellulase or endocellulase activity. The viscometric activities of the four cellulases are shown in *SI Appendix*, Fig. S2. *Cf*Cel6A and *Cf*Cel9A quickly reduced CMC viscosity, indicating that they are very active endocellulases. This is consistent with an earlier observation that *Cf*Cel6A acts randomly on cellulose, leading to particle fragmentation (18).

HS-AFM Observation of the Four *C. fimi* Cellulases. The processive movements of the four cellulases was observed by means of HS-AFM (Fig. 1). A suspension of crystalline cellulose III_I (average degree of polymerization [DP] of 880 ± 500; see *SI Appendix*, Fig. S3) was dropped on a highly oriented pyrographite (HOPG) disk for HS-AFM, and the cellulases were added to the observation buffer during the imaging. As shown in Fig. 1*B* and Movie S1, *Cf*Cel6B moved unidirectionally (from right to left in this image) on the surface of the cellulose III_I. *Cf*Cel48A also moved unidirectionally on the surface of cellulose III_I, as shown in Fig. 1*D* and Movie S3, but molecules moving on the substrate were rarely observed, compared to the case of *Cf*Cel6B. This is consistent with the fact that the cellulolytic activity of *Cf*Cel6B toward crystalline cellulose III_I was 170-fold higher than that of *Cf*Cel48A (*SI Appendix*, Fig. S1*D*). *Cf*Cel6B showed faster unidirectional movement and faster detachment from the cellulose III_I surface compared with *Tr*Cel7A, as described later. Furthermore, *Cf*Cel6A and *Cf*Cel9A molecules were bound to cellulose but did not move and were easily detached from the surface (Fig. 1 *A* and *C* and Movie S2 [*Cf*Cel9A]).

Different Orientations of Movement of *Cf*Cel6B and *Cf*Cel48A or of *Cf*Cel6B and *Tr*Cel7A. During the HS-AFM observation, enzymes can be added to the observation solution, and their real-time movements on the same microfibrils can be followed before and after the addition of another enzyme. In the present work, we directly confirmed that molecules of *Cf*Cel6B and *Cf*Cel48A or of *Cf*Cel6B and *Tr*Cel7A move in opposite directions using crystalline cellulose with parallel orientation of substrate chains (Fig. 2 *A* and *B* and Movies S4 and S5). When *Cf*Cel48A alone was initially observed, the enzyme molecules moved from the right to the left, as shown in Fig. 2*A, Left*, while movement from the left to the right was observed after *Cf*Cel6B was added (Fig. 2*A, Right* and Movie S4). Fig. 2*B* and Movie S5 show *Tr*Cel7A molecules moving from the left to the right, while *Cf*Cel6B molecules move in the opposite direction. The absolute direction of movement of cellulase molecules cannot be determined from HS-AFM observation, but reported kinetic data for labeled oligosaccharides (19, 20), morphological observations of crystalline cellulose after degradation by cellulases (21, 22), and the high-resolution molecular structure determination of the cellulases (23–31) all clearly indicate that *Tr*Cel7A and *Cf*Cel48A move from the reducing end to the nonreducing end, and *Cf*Cel6B moves from the nonreducing end to the reducing end.

The HS-AFM observations revealed that *Cf*Cel6B molecules showed shorter and faster movement compared with *Tr*Cel7A molecules. Sometimes *Cf*Cel6B and *Tr*Cel7A collided head-on, and in all these cases, *Cf*Cel6B was knocked off the crystalline cellulose substrate (Movie S6 and Fig. 2 *C* and *D*). We also observed that these "traffic accidents" between *Cf*Cel6B and *Tr*Cel7A triggered a "traffic jam" of *Tr*Cel7A (Movie S6 and Fig. 2*D*). The halting of one *Tr*Cel7A molecule blocked the following *Tr*Cel7A molecules (moving from the right to the left in Movie S6), which then stacked up behind it. However, mixing *Cf*Cel6B and *Tr*Cel7A still resulted in synergistic degradation of crystalline cellulose (Fig. 2*E*), whereas a combination of enzymes from the same family, i.e., *Cf*Cel6B and *Tr*Cel6A, resulted in impaired cellulose degradation, possibly because of competition between the two enzymes (Fig. 2*F*). A competitive effect of *Tr*Cel6A and *Tf*Cel6B (formally *Thermomonospora fusca* cellulase E3, bacterial GH6) in the degradation of filter paper has been reported, in contrast to a synergistic effect of the combination of *Tr*Cel7A and *Tf*Cel6B (32).

Analysis of Moving Velocity and Processivity of *Cf*Cel6B. The movements of *Cf*Cel6B molecules were further analyzed in order to estimate their kinetic characteristics. As shown in *SI Appendix*,

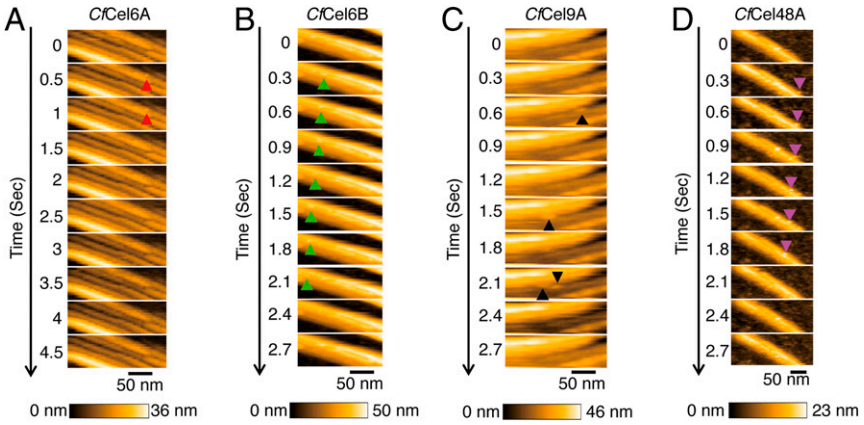

Fig. 1. HS-AFM images of the four *C. fimi* cellulases. (*A*) Real-time observation of crystalline cellulose III_I incubated with 1 μM *Cf*Cel6A. The time interval between images is 0.5 s (200 × 70 nm^2, 120 × 42 pixels). The positions of *Cf*Cel6A are marked by red arrowheads. (*B*) Real-time observation of crystalline cellulose III_I incubated with 1 μM *Cf*Cel6B. The time interval between images is 0.3 s (200 × 80 nm^2, 120 × 48 pixels). These images are taken from Movie S1. The positions of *Cf*Cel6B are marked by green arrowheads. (*C*) Real-time observation of crystalline cellulose III_I incubated with 1 μM *Cf*Cel9A. The time interval between images is 0.3 s (200 × 80 nm^2, 120 × 40 pixels). These images are taken from Movie S2. The positions of *Cf*Cel9A are marked by black arrowheads. (*D*) Real-time observation of crystalline cellulose III_I incubated with 1 μM *Cf*Cel48A. The time interval between images is 0.3 s (200 × 100 nm^2, 120 × 50 pixels). These images are taken from Movie S3. The positions of *Cf*Cel48A are marked by purple arrowheads.

Fig. S4*A*, the time courses of movement of *Cf*Cel6B showed that 24.0% of observed *Cf*Cel6B molecules (total of 500 molecules) stopped after moving on the cellulose for a while and then became detached from the substrate. *SI Appendix*, Fig. S4*B* shows a histogram of the moving velocities of *Cf*Cel6B (n = 500), which exhibited a Gaussian distribution with a mean ± SD of 12.7 ± 7.6 nm s^{-1} (Table 1). The average movement velocity of *Cf*Cel6B was twice that of *Tr*Cel7A (6.8 ± 3.5 nm s^{-1}) and was nearly equal to that of *Pc*Cel7C (14.7 ± 9.1 nm s^{-1}), which shows the fastest velocity we have ever measured (4). We made a histogram of the distribution of the duration of processive movement (*SI Appendix*, Fig. S4*C*) and estimated the half-life of processive movements of *Cf*Cel6B molecules to be 1.3 ± 0.1 s. From the velocity and half-life values, the processivity of *Cf*Cel6B was estimated to be 16.1 successive reactions, taking the length of the cellobiose repeating unit as 1.0 nm in crystalline cellulose (14).

We also analyzed the molecular movements when *Cf*Cel6B and *Tr*Cel7A were mixed and worked together and estimated the kinetic characteristics. *SI Appendix*, Fig. S4*D* shows the time courses of movement of *Cf*Cel6B (total of 625 molecules) in the presence of *Tr*Cel7A. *Cf*Cel6B molecules moved for longer in the presence of *Tr*Cel7A than in its absence, and it was rare to observe *Cf*Cel6B molecules stopping and then becoming detached (1.8% of 625 total molecules) (*SI Appendix*, Fig. S4*D*). The movement of molecules showed a Gaussian distribution with a mean ± SD of 19.4 ± 8.1 nm s^{-1}, i.e., 1.5 times faster than the movement of *Cf*Cel6B molecules alone (*SI Appendix*, Fig. S4*E* and Table 1). The half-life of processive movements of *Cf*Cel6B in the presence of *Tr*Cel7A was estimated to be 1.5 ± 0.1 s (*SI Appendix*, Fig. S4*F* and Table 1), corresponding to ∼28.8 reactions, nearly double that in the absence of *Tr*Cel7A (16.1). In contrast, the average velocity of *Tr*Cel7A molecules (n = 214) was 4.6 ± 1.8 nm s^{-1} in the presence of *Cf*Cel6B (*SI Appendix*, Fig. S4*H*), which is lower than that of *Tr*Cel7A alone (6.8 ± 3.5 nm s^{-1}) (Table 1). However, the half-life of processive movements of *Tr*Cel7A in the presence of *Cf*Cel6B was estimated to be 13.5 ± 4.6 s (*SI Appendix*, Fig. S4*I*), which is 3.9 times longer than that in the absence of *Cf*Cel6B (3.5 ± 0.2 s) and corresponds to ∼62.2 reactions, i.e., 2.6 times higher than that in the absence of *Cf*Cel6B (23.6) (Table 1). These results can well explain the synergy detected by biochemical measurement (Fig. 2*E*).

Discussion

The processivity of glycanases is a critical property for effective degradation of crystalline polysaccharides. In this study, we employed HS-AFM to directly observe the movements of four bacterial cellulases from *C. fimi* on insoluble cellulose substrates. We observed unidirectional movement of *Cf*Cel6B (GH6) and *Cf*Cel48A (GH48) on crystalline cellulose III_I. The other two cellulases, *Cf*Cel6A (GH6) and *Cf*Cel9A (GH9), were bound to cellulose III_I but did not move on the cellulose surface. To understand these results, we compared the molecular structures of the four *C. fimi* cellulases with those of four previously studied fungal cellulases: *Tr*Cel7A (33), *Tr*Cel7B (34), and *Tr*Cel6A (24), derived from *T. reesei*, and *Pc*Cel7D (35), derived from *P. chrysosporium* (Fig. 3*A*). The structures of the four *C. fimi* cellulases were predicted using the Phyre2 server (36). Phyre2 is a web-based bioinformatics server for the prediction of protein structure based on the principles of homology-based modeling. Among these cellulases, processive movements of *Tr*Cel7A, *Pc*Cel7D, *Cf*Cel6B, and *Cf*Cel48A have been confirmed by HS-AFM, but processive movements of *Tr*Cel7B, *Tr*Cel6A, *Cf*Cel6A, and *Cf*Cel9A have not been detected.

The difference in processivity between the two bacterial GH6 cellulases *Cf*Cel6A and *Cf*Cel6B is presumably due to their structural difference (Fig. 3*A*). As proposed by Varrot et al., GH family 6 cellulases can be clustered in four subfamilies based on amino acid sequence similarity (37), and *Cf*Cel6A belongs to subfamily 2, while *Cf*Cel6B belongs to subfamily 4. Among subfamily 2 cellulases, the structures of *Mt*Cel6A (amino acid identity with the catalytic domain of *Cf*Cel6A of 47%) and *Tf*Cel6A (identity of 43%) have been solved (37, 38), and these two molecules possess shortened loops or lack two loops that cover the active cleft in exotype *Tr*Cel6A (Fig. 3*A*); consequently, they exhibit open-cleft active sites that are typical of GH6 endoglucanases (37, 38). Moreover, GH6 cellulases in subfamily 2 lack the +4 tryptophan subsite that plays an essential role in crystalline cellulose degradation by the fungal family 6 cellulase *Tr*Cel6A (39). In contrast, subfamily 4 cellulase *Cf*Cel6B, which has 53% amino acid identity with the catalytic domain of *Tf*Cel6B, has the +4 subsite, and there are four loops covering the active sites, producing a tunnel-shaped active-site structure (29, 30). There is also a +6 subsite. These differences between *Cf*Cel6A and *Cf*Cel6B seem consistent with the

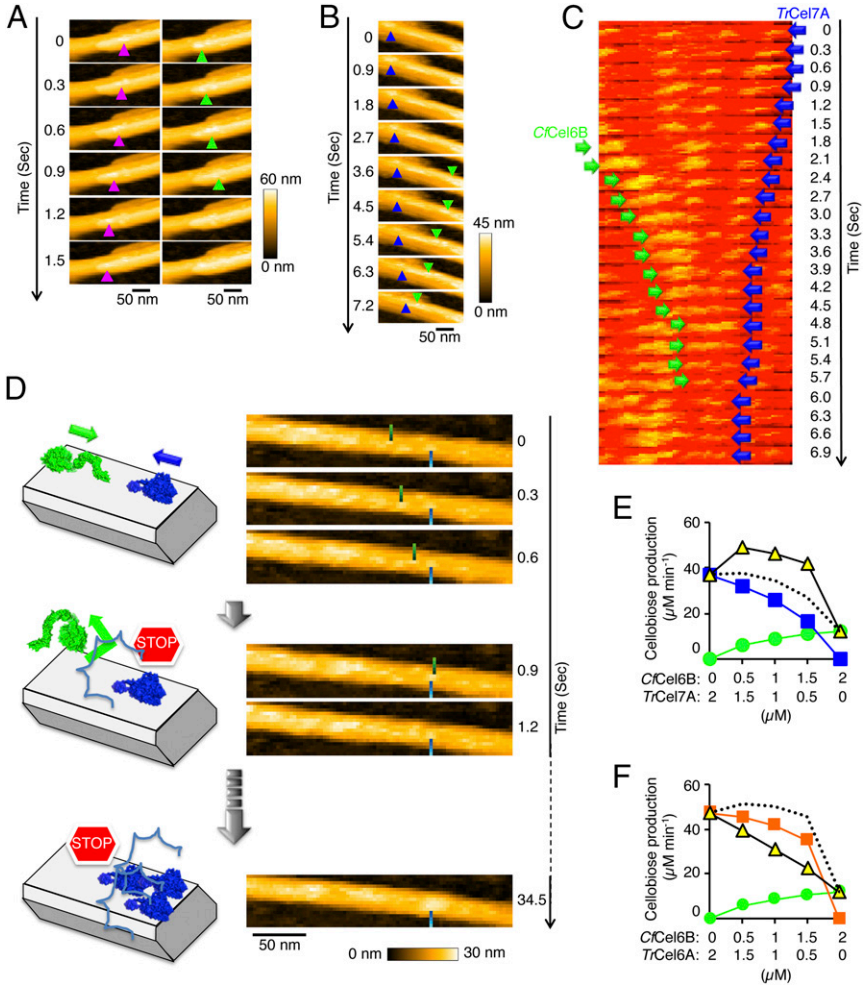

Fig. 2. Opposite directions of processive movement, leading to traffic accidents. (*A*) HS-AFM images of crystalline cellulose III_I degradation by *Cf*Cel6B and *Cf*Cel48A. The time interval between images is 0.3 s (200 × 120 nm^2, 120 × 60 pixels). These images are taken from Movie S4. Purple and green arrowheads show the positions of *Cf*Cel48A and *Cf*Cel6B, respectively. The reference processive direction was taken as that when *Cf*Cel48A alone was incubated with the substrate. (*B*) HS-AFM images of crystalline cellulose III_I degradation by *Cf*Cel6B and *Tr*Cel7A. The time interval between images is 0.3 s (200 × 80 nm^2, 120 × 48 pixels). These images are taken from Movie S5. Blue and green arrowheads show the positions of *Tr*Cel7A and *Cf*Cel6B, respectively. The reference processive direction was taken as that when *Cf*Cel6B alone was incubated with the substrate. (*C*) Kymographic analysis of *Cf*Cel6B and *Tr*Cel7A in a traffic accident on crystalline cellulose III_I. These images are taken from Movie S6. The time interval between images 0.3 s (254 × 75 nm^2, 153 × 45 pixels). (*D*) Actual HS-AFM images and schematic illustration of the traffic accident between *Cf*Cel6B and *Tr*Cel7A, followed by a traffic jam. The images are taken from Movie S6. Blue and green bars show the positions of *Tr*Cel7A and *Cf*Cel6B, respectively. (*E* and *F*) Synergy and negative synergy between *Cf*Cel6B and *Tr*Cel7A or *Cf*Cel6B and *Tr*Cel6A in cellobiose production from crystalline cellulose III_I. Green circles, blue squares, and orange squares show the rates of cellobiose production by *Cf*Cel6B, *Tr*Cel7A, and *Tr*Cel6A, respectively. The yellow triangles show the activity between the two cellulases. The dotted line indicates the sum of cellobiose production calculated from the green, blue, and orange plots. $n = 3$.

observed difference in processivity (Fig. 3*A*). Although our previous HS-AFM study of *Tr*Cel6A demonstrated that the processivity of the fungal family 6 enzyme is less than that of family 7 cellulase (3), the present work clearly demonstrates a strong processive movement of *Cf*Cel6B, suggesting that the presence of four loops that cover the active-site cleft and construct a longer tunnel than in *Tr*Cel6A could be important for enabling GH6 subfamily 4 cellulases to grip the cellulose chain well and thus to achieve increased processivity.

Our previous study of the processivity of three fungal GH7 CBHs revealed that the processivity is strongly influenced by the presence and nature of the loop regions covering the active-site cleft (4). The high processivity observable with HS-AFM seems to be associated with the presence of three or more loops covering the active cleft and a substrate-binding/product-binding subsite (Sub/Pro) ratio of 3 or more (Fig. 3*C*). Cellulases without a loop covering the active cleft, such as *Cf*Cel9A, *Cf*Cel6A, and *Tr*Cel7B, can all be classified biochemically as endotype cellulases. For two GH7 cellulases, *Pc*Cel7D and *Tr*Cel7B, which have the same Sub/Pro ratio, the presence or absence of the loop covering the active cleft is considered to lead to the presence or absence of processivity, respectively. Interestingly, there is a clear difference between these processive cellulases and GH18 chitinases, which exhibit processivity, as observed by HS-AFM, even though the chitinases have no loop covering the active cleft (40). The Sub/Pro ratio in the catalytic domain of GH18 chitinases is only about 2, indicating that the catalytic domain itself cannot show processivity, at least according to our Sub/Pro ratio vs. number of loops analysis. However, they are connected to a chitin-binding module via a short linker peptide, and the connected two-domain structure extends the subsites and affords a Sub/Pro ratio of more than 5, which is consistent with higher processivity. In fact, the higher Sub/Pro ratio of chitinase A (ChiA) than chitinase B (ChiB) is reflected in higher processivity

Table 1. Comparison of statistical analysis data of *Cf*Cel6B and GH family 7 cellulases

Cellulase	Dissociation rate constant,* s^{-1}	Half-life of movement,* s	Average velocity,† nm s^{-1}	Calculated processivity, No. of reactions
*Cf*Cel6B	0.55 ± 0.02	1.3 ± 0.1	12.7 ± 7.6	16.1
*Pc*Cel7C‡	0.51 ± 0.01	1.4 ± 0.0	14.7 ± 9.1	19.9
*Pc*Cel7D‡	0.32 ± 0.04	2.2 ± 0.1	9.4 ± 3.7	20.3
*Tr*Cel7A‡	0.20 ± 0.01	3.5 ± 0.2	6.8 ± 3.5	23.6
*Cf*Cel6B (with *Tr*Cel7A)	0.47 ± 0.03	1.5 ± 0.1	19.4 ± 8.1	28.8
*Tr*Cel7A (with *Cf*Cel6B)	0.05 ± 0.02	13.5 ± 4.6	4.6 ± 1.8	62.2

*The values are mean ± SE of the fitting with exponential decay.
†The values are mean ± SD for Gaussian distribution.
‡Data are from ref. 4.

of ChiA than ChiB. Thus, in cellulases, the processivity arises from the combination of loops covering the active cleft and the Sub/Pro ratio of the catalytic domain, whereas chitinases acquire processivity only via the Sub/Pro ratio. It should be noted that chitin is a polymer of acetylglucosamine, which has an acetoamide modification at the C2 position of glucose and thus is bulkier than cellulose, a simple glucose polymer. This could be the reason why the active cleft of chitinase is not covered with loops.

Next, we employed HS-AFM observation to see what kinds of phenomena occur at the single-molecule level to produce the synergistic effect of the combination of GH6 and GH7 cellulases. Statistical analysis revealed that in the presence of both *Cf*Cel6B and *Tr*Cel7A, the moving distances of the two cellulases on the substrate became longer, and the processivity of both enzymes was increased. This may imply that the use of two complementary types of cellulases is effective for removing obstacles on the cellulose crystal surface. This, in turn, might increase the area of the cellulose surface accessible to the degradative enzymes, as well as reducing the frequency of their detachment from the cellulose surface, thus promoting decomposition of the cellulose chain (Fig. 2*E*) and increasing the processivity (Table 1). An alternative explanation could be that once a chain has been removed from the surface by the action of one enzyme, the

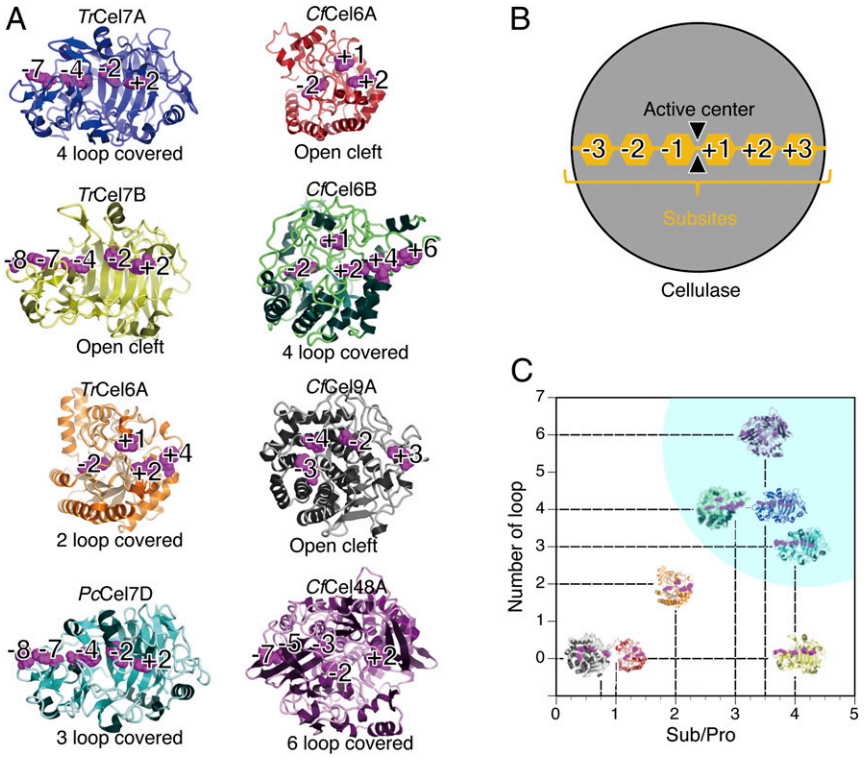

Fig. 3. (*A*) Structural comparison of the catalytic domains of the four fungal cellulases and four *C. fimi* cellulases. The fungal cellulases are *T. reesei Tr*Cel7A (PDB:1CEL), *Tr*Cel7B (PDB:1EG1), and *Tr*Cel6A (PDB:1QK2) and *P. chrysosporium Pc*Cel7D (PDB:1GPI). The catalytic domain structures of the *C. fimi* cellulases were predicted using Phyre2. Aromatic residues forming the subsites are shown in purple. (*B*) Definition of the Sub/Pro ratio. The active cleft of a cellulase consists of multiple binding sites for glucose units of cello-oligo sugars. These sites are defined as "subsites" and are designated as −3, −2, −1, active center, +1, +2, and +3 here. The subsites before the active center are defined as subsites for substrate (Sub) binding. The subsites after the active center are defined as subsites for product (Pro) binding. The Sub/Pro ratio was calculated. Cellulases which degrade substrates from the reducing end side to the nonreducing end of sugar chains are expressed as −3, −2, and −1 (Sub); active center; and +1, +2, and +3 (Pro). Cellulases that degrade substrates from the nonreducing end to the reducing end are expressed as +3, +2, and +1 (Sub); active center; and −1, −2, and −3 (Pro). (*C*) Structural classification of the eight cellulase catalytic domains. The vertical axis shows the number of loops covering the catalytic cleft. The horizontal axis shows the Sub/Pro ratio, which is the ratio of the number of binding subsites for Sub divided by the number of binding subsites for Pro. The Sub/Pro ratios are *Tr*Cel7A, 3.5; *Tr*Cel7B, 4.0; *Pc*Cel7D, 4.0; *Tr*Cel6A, 2.0; *Cf*Cel6A, 1.0; *Cf*Cel6B, 3.0; *Cf*Cel9A, 0.75; and *Cf*Cel48A, 3.5. Blue highlighting indicates the area observed by HS-AFM.

neighboring chains are less tightly held together and thus are more accessible to the other enzyme. In addition, it was observed that *Cf*Cel6B sometimes collided with *Tr*Cel7A, causing a traffic jam (Fig. 2*D* and Movie S6). *T. reesei* is thought to degrade crystalline cellulose by employing the combination of *Tr*Cel7A and *Tr*Cel6A, both of which are thought to be processive CBHs. However, we could not observe processivity of *Tr*Cel6A by means of HS-AFM, which may mean that the retaining force on the cellulose surface is weaker than that of *Tr*Cel7A or *Cf*Cel6B. This situation could have evolved to prevent the occurrence of immovable traffic jams because *Tr*Cel6A would be knocked off the substrate when a collision occurs. The combination of *Cf*Cel6B and *Tr*Cel7A did show collisions resulting in congestion, which may be the reason why the synergistic effect between *Cf*Cel6B and *Tr*Cel7A (1.4 times, Fig. 2*E*) is less than that between *Tr*Cel6A and *Tr*Cel7A [2.3 times (3)]. Similarly, *C. fimi* secretes *Cf*Cel9A, which is considered to be the major endo-processive enzyme, but no processive action was observed with HS-AFM, suggesting that the relationship between *Cf*Cel9A and *Cf*Cel6B may be similar to that between *Tr*Cel7A and *Tr*Cel6A from *T. reesei.* Although *T. reesei* and *C. fimi* are completely different types of microorganisms (eukaryote and prokaryote, respectively), they both share a combination of cellulases with high and low processivity, clearly indicating that such a combination is favorable for the deconstruction of cellulose. However, it is noteworthy that in *C. fimi*, the highly processive *Cf*Cel6B moves from the nonreducing to the reducing end, while in *T. reesei*, the highly processive *Tr*Cel7A moves in the opposite direction, from the reducing to the nonreducing end (Fig. 2*B*).

A comparison of GH6 cellulases with different processivities clearly indicated that increased processivity is favored by 1) an increase in the number of subsites available for substrate binding (41), 2) the presence of an asymmetric active site with a decreased number of subsites for the product (39), and 3) an increase in the number of loop structures to hold the substrate more tightly in the active site (29), as shown in Fig. 4. The importance of these features is supported by the results of a molecular phylogenetic analysis of GH6 cellulases (*SI Appendix*, Fig. S5). The 3D structures of two GH7 cellulases and one GH16 endo-β-1,3-1,4-glucanase are compared in Fig. 4. The two GH families are classified into clan GH-B, sharing the β-jellyroll fold (42). *Pt*Lic16A, a GH16 β-1,3-1,4-glucanase from *Paecilomyces thermophila*, has an almost symmetrical subsite (Sub/Pro ratio of 5/4) with no loop covering the active center (43). However, in GH7 endocellulase *Tr*Cel7B the subsite is asymmetric (Sub/Pro ratio of 8/2), while in *Tr*Cel7A the active center is covered with loops (23). Molecular phylogenetic analysis of GH7 cellulases and GH16 β-1,3-1,4-glucanases also supports the significance of these features (*SI Appendix*, Fig. S6). The structures of GH8 cellulase and GH48 cellulase are compared in Fig. 4. The two GH family enzymes are both classified in clan GH-M, sharing the $(\alpha/\alpha)_6$ fold. GH8 from *Clostridium thermocellum* (*Ht*Cel8A) has a symmetrical subsite (Sub/Pro ratio of 3/3) with no loop covering the active center (44), while GH48 cellulase from *Clostridium cellulolyticum* (*Rc*Cel48F) has an asymmetric subsite (Sub/Pro ratio of 7/3), and the active center is covered with loops (26). The results of molecular phylogenetic analysis of GH8 β-1,3-1,4-glucanases and cellulases and GH48 cellulases also support the relationship of higher processivity with increasing numbers of subsites in the active site and loop structures over the active site (*SI Appendix*, Fig. S7). For example, the structure of *Cf*Cel6B with four loops and a Sub/Pro ratio of 6/2 is quite similar to that of *Tr*Cel7A, which has four loops and a Sub/Pro ratio of 8/2, even though their degradation directions are opposite to each other. Bacteria do not have GH7 cellulases, but they have evolved GH6 cellulases, which have different protein foldings, into highly processive cellulases that can degrade recalcitrant crystalline cellulose.

The well-known concept of "convergent evolution," which is often applied to explain the shape of animals, such as sharks and dolphins having a long nose and streamlined body for effective movement in water even though they were originally different animals (fish and mammalian), also seems applicable to the changes in the shape of endoglucanase to generate a processive cellulase. Thus, the GH6, GH48, and GH7 processive cellulases all utilize common strategies: extending the protein "nose" in the

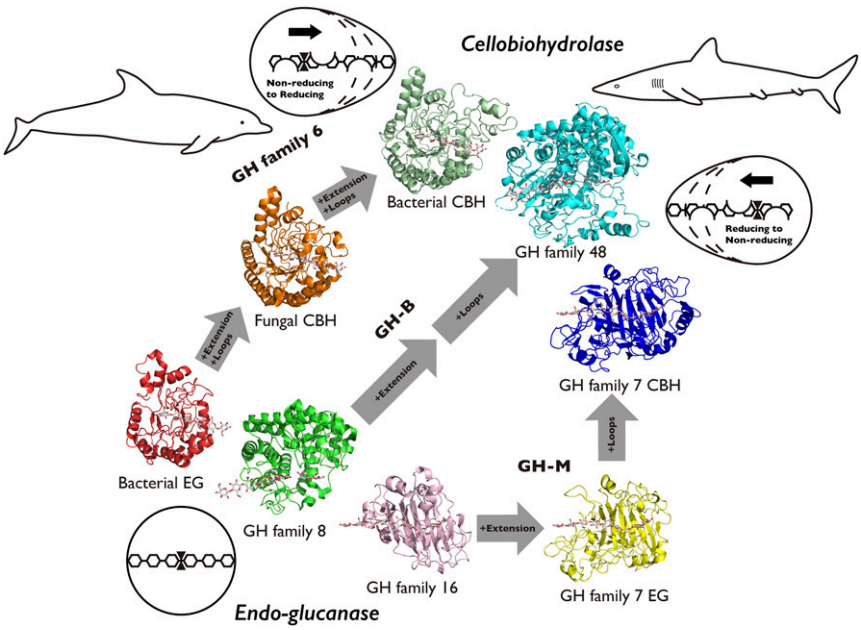

Fig. 4. Schematic representation of the putative convergent evolution of bacterial and fungal cellulases, employing different protein foldings. Modification of the symmetric architecture at positive and negative subsites in endoglucanases to generate asymmetric architecture by blocking part of the product binding region and/or by extending the substrate binding subsites increases the processivity (lower left to middle part). The presence of loops covering the active site increases the processivity (upper right). A combination of additional loops and asymmetricity results in high processivity for cellobiohydrolases.

substrate direction to form asymmetric subsite architecture and covering the subsite with loops. Degradation of crystalline cellulose is essential for both fungi and bacteria, and processive CBHs are necessary for this purpose. However, since fungi and bacteria produce glycosyl hydrolases with different folding types, they have evolved processive CBHs using different protein folds as templates. In other words, completely different evolutionary strategies of cellulolytic enzymes in bacteria and fungi have yielded functionally similar and highly processive cellulases, providing an interesting example of "protein-level convergent evolution" in cellulolytic microorganisms.

Methods

Materials. Crystalline celluloses I_α and III_I were prepared from green algae *Cladophora* spp. according to a previous report (14). PASC was prepared from Avicel (Funacel II; Funakoshi Chemical) according to a previous report (4). CMC was purchased from Hercules. Cellooligosaccharides were acquired from Seikagaku Kogyo. *Cellulomonas fimi* ATCC 484 was used as a source of *Cf*Cel6A, *Cf*Cel6B, *Cf*Cel9A, and *Cf*Cel48A genes. *Trichoderma reesei* ATCC 66589 was used as a source of the *Tr*Cel6A gene. *Escherichia coli* JM109 (Takara) was used as the subcloning host. *E. coli* Rosetta2 (DE3), Tuner (DE3) (Novagen), and *Brevibacillus choshinensis* HPD31-SP3 (Takara) were used for heterologous production of the recombinant proteins.

Cloning of Four Bacterial Cellulases from *C. fimi*. Four *C. fimi* cellulase genes were amplified by PCR. The *Cf*Cel6A gene and *Cf*Cel6B gene were cloned into the *E. coli* expression vector. The *Cf*Cel9A gene and *Cf*Cel48A gene were cloned into the *B. choshinensis* expression vector. For details, see *SI Appendix*.

Production and Purification of Four Bacterial Cellulases from *C. fimi*. Four recombinant *C. fimi* cellulases were purified (*SI Appendix*, Fig. S8) according to the following protocol, as described in detail in *SI Appendix*.

The concentrations of the four cellulases were determined using the following values of molecular absorption coefficient at 280 nm: 107,745 M^{-1} cm^{-1} for *Cf*Cel6A, 134,105 M^{-1} cm^{-1} for *Cf*Cel6B, 217,875 M^{-1} cm^{-1} for *Cf*Cel9A, and 280,325 M^{-1} cm^{-1} for *Cf*Cel48A. Calculations were performed using the ProtParam tool (45).

Enzyme Assay. Enzyme activity toward insoluble celluloses was determined in a 150 μL reaction mixture containing 100 mM sodium acetate buffer (pH 5.0), 1.0 mg mL^{-1} substrate, and 1 μM enzyme at 37 °C. The reaction was stopped by heating the sample at 98 °C for 5 min. Soluble sugars were recovered by passing the reaction mixture through a 0.22 μm Durapore polyvinylidene difluoride (PVDF) membrane filter (MultiScreen $_{HTS}$-GV; Merck Millipore) and were analyzed by HPLC (LC-2000 series; Jasco) using a Corona charged aerosol detector (Corona CAD; ESA Biosciences) and a prepacked Shodex Asahipak NH2P-50 column (4.6 × 250 mm; Showa Denko K. K.). The eluent was water with a linear gradient of CH_3CN from 60 to 40% over 12.5 min. The column was kept at 40 °C and eluted at a flow rate of 1.0 mL min^{-1}. The amounts of glucose oligomers up to cellohexaose were identified and quantitated by comparing the retention times and peak areas with those of authentic standards. Concentrations of soluble sugars in the reaction mixtures were fitted to the equations previously described (4) to calculate the production velocity of soluble sugars. The ratio of cellobiose/glucose + cellotriose production was estimated according to a previous report (17).

Viscometric Assay. Enzymatic hydrolysis of CMC was measured on an Automated Micro Viscometer (AMVn; Anton Paar) based on the rolling time of a ball through viscous liquids according to the Höppler principle. The AMVn measuring system employed a 1.6 mm diameter glass capillary and a 1.5 mm diameter steel ball (density; 7.67 g/cm^3). Calibration parameters were 0.797 mPa s for dynamic viscosity, 0.99336 g/cm^3 for buffer density, 20° for minimum angle, and 80° for maximum angle. The reaction mixture contained 20 mM sodium acetate buffer (pH 5.0), 5.0 mg mL^{-1} CMC, and 10 nM enzyme, and the enzyme reaction was performed at 37 °C.

High-Speed AFM Observation. The movements of molecules of the four bacterial cellulases and *Tr*Cel7A were observed by high-speed atomic force microscopy as described in previous reports (2–4, 40, 46). A suspension (1 mg mL^{-1}) of highly crystalline cellulose III_I was dropped on a stage of HOPG and incubated at 25 °C for 10 min. Unbound cellulose III_I was washed off with 100 mM sodium acetate buffer (pH 5.0). Then the HOPG substrate was immersed in the observation pool filled with the same buffer (volume of 76 μL), and 4 μL of 20 μM cellulase solution were added to the observation pool. We observed molecules whose movement started and ended within the observation area. More than five independent cellulose microfibrils with a length greater than 0.8 μm were observed to estimate the velocity and processivity of cellulases.

For statistical analysis of the moving velocities of individual cellulase molecules, semiautomatic tracking of each molecule was carried out with a kymograph (40). The moving velocities of cellulase molecules were analyzed by IGOR Pro-6 (WaveMetrics) and fitted to a Gaussian distribution. Histogram data were tested for normality using the Shapiro–Wilk test (Prism7, GraphPad). Values satisfied the Shapiro–Wilk test, ensuring a normal distribution ($P \geq 0.05$). Adsorption times of molecules were calculated from the number of frames in which the molecules were observed, and plots of the number of molecules in each range of adsorption time were fitted to the previously reported equation (4, 40).

Determination of the Degree of Polymerization of Cellulose III_I. The DP of cellulose III_I was estimated from the fiber lengths measured in HS-AFM scanning images (scanning area: 0.64 to 9.00 μm^2). A histogram of fiber lengths ($n = 363$) was created from the results. Histogram data were tested for normality using the Shapiro–Wilk test (Prism7, GraphPad). Values satisfied the Shapiro–Wilk test, ensuring a normal distribution ($P \geq 0.05$). The DP was calculated by taking the length of the cellobiose repeating unit in crystalline cellulose III_I to be 1.0 nm (14).

Phylogenetic Tree Analysis. The amino acid sequences of the enzymes belonging to each GH family whose function is known were collected from the Carbohydrate-Active enZymes (CAZy) database (47). The obtained amino acid sequences were subjected to BlastP analysis, and homologous sequences [conserved protein domain family (48)] representing each GH family were excised. These sequences were subjected to phylogenetic analysis. Sequence alignment and phylogenetic analysis were performed using the Phylogeny.fr web server (49). The "one-click" analysis mode was applied to each set of sequences, using MUSCLE (version 3.5) (50) for multiple sequence alignment, PhyML (version 3.0) (51) and aLRT (52) for phylogenetic analysis, and TreeDyn (53) for tree rendering. Branches with aLRT branch support values under 50% were collapsed.

Data Availability. All data obtained and analyzed in the present work are available in *SI Appendix* and Movies S1 to S6.

ACKNOWLEDGMENTS. This research was supported by Grants-in-Aid for Innovative Areas (Grant 18H05494 to K.I. and Grant 18H04512 and 19H05389 to T. Uchihashi) from the Japanese Ministry of Education, Culture, Sports and Technology and by Grants-in-Aid for Scientific Research (B: Grant 18H02252 to K.I. and Grant 18H01837 to T. Uchihashi; C: Grant 15K07383 to T. Uchiyama) from the Japan Society for the Promotion of Science. K.I. thanks the Finnish Funding Agency for Innovation for the support of the Finland Distinguished Professor Program "Advanced approaches for enzymatic biomass utilisation and modification (BioAD)."

1. S. Park, J. O. Baker, M. E. Himmel, P. A. Parilla, D. K. Johnson, Cellulose crystallinity index: Measurement techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels* **3**, 10 (2010).
2. K. Igarashi *et al.*, High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *J. Biol. Chem.* **284**, 36186–36190 (2009).
3. K. Igarashi *et al.*, Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279–1282 (2011).
4. A. Nakamura *et al.*, Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. *J. Am. Chem. Soc.* **136**, 4584–4592 (2014).
5. A. Nakamura *et al.*, Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *Trichoderma reesei* Cel6A and its domains on crystalline cellulose. *J. Biol. Chem.* **291**, 22404–22413 (2016).
6. W. K. R. Wong *et al.*, Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. *Gene* **44**, 315–324 (1986).
7. A. Meinke *et al.*, Cellobiohydrolase A (CbhA) from the cellulolytic bacterium *Cellulomonas fimi* is a β-1,4-exocellobiohydrolase analogous to *Trichoderma reesei* CBH II. *Mol. Microbiol.* **12**, 413–422 (1994).
8. J. B. Owolabi, P. Beguin, D. G. Kilburn, R. C. Miller Jr., Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. *Appl. Environ. Microbiol.* **54**, 518–523 (1988).
9. H. Shen, N. R. Gilkes, D. G. Kilburn, R. C. Miller Jr., R. A. J. Warren, Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium *Cellulomonas fimi*. *Biochem. J.* **311**, 67–74 (1995).
10. L. E. Sandercock, A. Meinke, N. R. Gilkes, D. G. Kilburn, R. A. J. Warren, Degradation of cellulases in cultures of *Cellulomonas fimi*. *FEMS Microbiol. Lett.* **143**, 7–12 (1996).

11. J. Sakon, D. Irwin, D. B. Wilson, P. A. Karplus, Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.* **4**, 810–818 (1997).
12. R. H. Atalla, D. L. Vanderhart, Native cellulose: A composite of two distinct crystalline forms. *Science* **223**, 283–285 (1984).
13. D. L. VanderHart, R. H. Atalla, Studies of microstructure in native cellulose using solid-state 13 C NMR. *Macromolecules* **17**, 1465–1472 (1984).
14. M. Wada, H. Chanzy, Y. Nishiyama, P. Langan, Cellulose III_I crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. *Macromolecules* **37**, 8548–8555 (2004).
15. K. Igarashi, M. Wada, M. Samejima, Activation of crystalline cellulose to cellulose III_I results in efficient hydrolysis by cellobiohydrolase. *FEBS J.* **274**, 1785–1792 (2007).
16. M. Wada, L. Heux, J. Sugiyama, Polymorphism of cellulose I family: Reinvestigation of cellulose IV_I. *Biomacromolecules* **5**, 1385–1391 (2004).
17. I. von Ossowski *et al*., Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with *Phanerochaete chrysosporium* Cel7D. *J. Mol. Biol.* **333**, 817–829 (2003).
18. H. Stålbrand *et al*., Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β-1,4-glucanases. *Appl. Environ. Microbiol.* **64**, 2374–2379 (1998).
19. M. Vrsanská, P. Biely, The cellobiohydrolase I from *Trichoderma reesei* QM9414: Action on cello-oligosaccharides. *Carbohydr. Res.* **227**, 19–27 (1992).
20. B. K. Barr, Y.-L. Hsieh, B. Ganem, D. B. Wilson, Identification of two functionally different classes of exocellulases. *Biochemistry* **35**, 586–592 (1996).
21. H. Chanzy, B. Henrissat, Unidirectional degradation of *Valonia* cellulose microcrystals subjected to cellulase action. *FEBS Lett.* **184**, 285–288 (1985).
22. T. Imai, C. Boisset, M. Samejima, K. Igarashi, J. Sugiyama, Unidirectional processive action of cellobiohydrolase Cel7A on *Valonia* cellulose microcrystals. *FEBS Lett.* **432**, 113–116 (1998).
23. C. Divne, J. Ståhlberg, T. T. Teeri, T. A. Jones, High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **275**, 309–325 (1998).
24. J. Y. Zou *et al*., Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure* **7**, 1035–1045 (1999).
25. A. Varrot, M. Schülein, G. J. Davies, Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry* **38**, 8884–8891 (1999).
26. G. Parsiegla *et al*., Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry* **39**, 11238–11246 (2000).
27. B. G. Guimarães, H. Souchon, B. L. Lytle, J. H. David Wu, P. M. Alzari, The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the *Clostridium thermocellum* Cellulosome. *J. Mol. Biol.* **320**, 587–596 (2002).
28. W. Ubhayasekera, I. G. Muñoz, A. Vasella, J. Ståhlberg, S. L. Mowbray, Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. *FEBS J.* **272**, 1952–1964 (2005).
29. M. Wu *et al*., Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. *J. Biol. Chem.* **288**, 33107–33117 (2013).
30. M. Sandgren *et al*., The structure of a bacterial cellobiohydrolase: The catalytic core of the *Thermobifida fusca* family GH6 cellobiohydrolase Cel6B. *J. Mol. Biol.* **425**, 622–635 (2013).
31. M. Kostylev *et al*., Cel48A from *Thermobifida fusca*: Structure and site directed mutagenesis of key residues. *Biotechnol. Bioeng.* **111**, 664–673 (2014).
32. D. C. Irwin, M. Spezio, L. P. Walker, D. B. Wilson, Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.* **42**, 1002–1013 (1993).
33. C. Divne *et al*., The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524–528 (1994).
34. G. J. Kleywegt *et al*., The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 A resolution, and a comparison with related enzymes. *J. Mol. Biol.* **272**, 383–397 (1997).
35. I. G. Muñoz *et al*., Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. *J. Mol. Biol.* **314**, 1097–1111 (2001).
36. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
37. A. Varrot *et al*., *Mycobacterium tuberculosis* strains possess functional cellulases. *J. Biol. Chem.* **280**, 20181–20184 (2005).
38. M. Spezio, D. B. Wilson, P. A. Karplus, Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* **32**, 9906–9916 (1993).
39. A. Koivula *et al*., Tryptophan 272: An essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. *FEBS Lett.* **429**, 341–346 (1998).
40. K. Igarashi *et al*., Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. *Nat. Commun.* **5**, 3975 (2014).
41. A. M. Larsson *et al*., Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion. *Biochemistry* **44**, 12915–12922 (2005).
42. B. Henrissat, A. Bairoch, Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696 (1996).
43. Y.-S. Cheng *et al*., Structural and mutagenetic analyses of a 1,3-1,4-β-glucanase from *Paecilomyces thermophila*. *Biochm. Biophys. Acta* **1844**, 366–373 (2014).
44. D. M. A. Guérin *et al*., Atomic (0.94 A) resolution structure of an inverting glycosidase in complex with substrate. *J. Mol. Biol.* **316**, 1061–1069 (2002).
45. E. Gasteiger *et al*., "Protein identification and analysis tools on the ExPASy server" in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Springer, Totowa, NJ, 2005), pp. 571–607.
46. K. Igarashi *et al*., Visualization of cellobiohydrolase I from *Trichoderma reesei* moving on crystalline cellulose using high-speed atomic force microscopy. *Methods Enzymol.* **510**, 169–182 (2012).
47. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
48. S. Lu *et al*., CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
49. A. Dereeper *et al*., Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
50. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
51. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
52. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
53. F. Chevenet, C. Brun, A. L. Bañuls, B. Jacq, R. Christen, TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).