Convergent evolution of processivity in bacterial and fungal cellulases

Taku Uchiyama1,a, Takayuki Uchihashi1,b,c,1, Akihiko Nakamura4, Hiroki Watanabe3, Satoshi Kaneo5, Masahiro Samejima2, and Kiyohiko Igarashi6 (五十嵐 圭日子)2,5,6

*Department of Biomaterial Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan; 1Department of Physics and Structural Biology Research Center, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan; 2Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Aichi 444-8787, Japan; 3Institute for Molecular Science, National Institutes of Natural Sciences, Aichi 444-8787, Japan; 4Department of Subtropical Biochemistry and Biotechnology, Faculty of Agriculture, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan; and 5Protein 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 (CfCel6B and CfCel48A) slide on crystalline cellulose. The direction of processive movement of CfCel6B is from the nonreducing to the reducing end of the substrate, which is opposite that of processive cellulase Cel7A of the fungus Trichoderma reesei (TrCel7A), whose movement was first observed by this technique, while CfCel48A moves in the same direction as TrCel7A. When CfCel6B and TrCel7A were mixed on the same substrate, “traffic accidents” were observed, in which the two cellulases blocked each other’s progress. The processivity of CfCel6B 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.

cellulohydrolase | 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 ~75% 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 (TrCel7A) and Phanerochaete chrysosporium (PcCel7C and PcCel7D) 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 TrCel7A, PcCel7C, and PcCel7D) and shows processive degradation, while the other binds weakly to the cellulose chain (like TrCel6A), 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, CfCel6A, CfCel6B, CfCel9A, and CfCel48A (formerly known as CenA (6), CbhA (7), CenB (8), CenC, and CenD) on various insoluble cellulose substrates using HS-AFM. We observed that two bacterial CBHs, CfCel6B and CfCel48A, move unidirectionally on the crystalline cellulose surface, while the other two, CfCel9A and CfCel6A, do not. These results suggest that bacterial cellulases exhibit processivity, just like fungal CBHs, but their processivity is lower than that of fungal CBHs.

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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1T. Uchiyama and T. Uchihashi contributed equally to this work.
2To whom correspondence may be addressed. Email: aquarius@email.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.

19896-19903 | PNAS | August 18, 2020 | vol. 117 | no. 33 | www.pnas.org/cgi/doi/10.1073/pnas.2011366117
and ChbB (9), respectively] from the cellulolytic actinobacteria *Cellulomonas fimi*. This microorganism is one of the best-known aerobic cellulolytic bacteria and expresses three endoglucanases, two exoglucobiohydrolases and one xylanase in the supernatants of cultures grown on cellulose (10). CjCel6A and CjCel9A are thought to be endoglucanases (6, 8), while CjCel9A has a high amino acid sequence similarity to the possibly processive endoglucanase TfCel9A from *Thermobifida fusca* (11). CjCel6B and CjCel48A 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 Iα and IIII from green algae, and the amounts of reaction products were measured by high performance liquid chromatography (HPLC). Crystalline cellulose in nature has structures called Iα and Iβ (12, 13), which differ in interlayer chain stacking. When Iα and Iβ are treated with supercritical ammonia, the chain stacking arrangement changes to a completely different structure, called IIII (14). Cellulose IIII forms staggered layers with intra- and interlayer hydrogen-bonding interactions, unlike cellulose Iα, in which the chains pack into flat layers. Because of this change, cellulose IIII has a larger hydrophobic surface than cellulose Iα, making it more susceptible to attack by cellulase (15). Cellulose IIII 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 CjCel6A, CjCel6B, and CjCel48A was cellobiose, whereas the major products of CjCel9A were glucose and cellotriose, in addition to cellulbiose (*SI Appendix*, Fig. S1 E–H). CjCel6A degraded PASC and Avicel more efficiently than did the other three cellulases, while CjCel9A efficiently hydrolyzed crystalline cellulose Iα. CjCel6B degraded crystalline cellulose IIII more efficiently than did the other three enzymes. CjCel48A 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 CjCel6B and CjCel48A; 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 CjCel6A and CjCel9A, respectively. These results suggest that CjCel6B and CjCel48A are highly processive cellulases, whereas CjCel6A and CjCel9A 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. CjCel6A and CjCel9A quickly reduced CMC viscosity, indicating that they are very active endocellulases. This is consistent with an earlier observation that CjCel6A 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 IIII (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. 1B and Movie S1, CjCel6B moved unidirectionally (from right to left in this image) on the surface of the cellulose IIII. CjCel48A also moved unidirectionally on the surface of cellulose IIII, as shown in Fig. 1D and Movie S3, but molecules moving on the substrate were rarely observed, compared to the case of CjCel6B. This is consistent with the fact that the cellulolytic activity of CjCel6B toward crystalline cellulose IIII was 170-fold higher than that of CjCel48A (*SI Appendix*, Fig. S1D). CjCel6B showed faster unidirectional movement and faster detachment from the cellulose IIII surface compared with TrCel7A, as described later. Furthermore, CjCel6A and CjCel9A molecules were bound to cellulose but did not move and were easily detached from the surface (Fig. 1 A and C and Movie S2 [CjCel9A]).

**Different Orientations of Movement of CjCel6B and CjCel48A or of CjCel6B and TrCel7A.** 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 CjCel6B and CjCel48A or of CjCel6B and TrCel7A move in opposite directions using crystalline cellulose with parallel orientation of substrate chains (Fig. 2 A and B and Movies S4 and S5). When CjCel48A alone was initially observed, the enzyme molecules moved from the right to the left, as shown in Fig. 2A, *Left*, while movement from the left to the right was observed after CjCel6B was added (Fig. 2A, *Right* and Movie S4). Fig. 2B and Movie S5 show TrCel7A molecules moving from the left to the right, while CjCel6B 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 TrCel7A and CjCel48A move from the reducing end to the nonreducing end, and CjCel6B moves from the nonreducing end to the reducing end.

The HS-AFM observations revealed that CjCel6B molecules showed shorter and faster movement compared with TrCel7A molecules. Sometimes CjCel6B and TrCel7A collided head-on, and in all these cases, CjCel6B was knocked off the crystalline cellulose substrate (Movie S6 and Fig. 2 C and D). We also observed that these “traffic accidents” between CjCel6B and TrCel7A triggered a “traffic jam” of TrCel7A (Movie S6 and Fig. 2D). The halting of one TrCel7A molecule blocked the following TrCel7A molecules (moving from the right to the left in Movie S6), which then stacked up behind it. However, mixing CjCel6B and TrCel7A still resulted in synergistic degradation of crystalline cellulose (Fig. 2E), whereas a combination of enzymes from the same family, i.e., CjCel6B and TrCel6A, resulted in impaired cellulose degradation, possibly because of competition between the two enzymes (Fig. 2F). A competitive effect of TrCel6A and CjCel6B (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 TrCel7A and CjCel6B (32).

**Analysis of Moving Velocity and Processivity of CjCel6B.** The movements of CjCel6B 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, incubated with 1 μM CfCel6A. The time interval between images is 0.5 s (200 × 70 nm², 120 × 42 pixels). The positions of CfCel6A are marked by red arrowheads. (B) Real-time observation of crystalline cellulose III, incubated with 1 μM CfCel6B. The time interval between images is 0.3 s (200 × 80 nm², 120 × 48 pixels). These images are taken from Movie S1. The positions of CfCel6B are marked by green arrowheads. (C) Real-time observation of crystalline cellulose III, incubated with 1 μM CfCel9A. The time interval between images is 0.3 s (200 × 80 nm², 120 × 40 pixels). These images are taken from Movie S2. The positions of CfCel9A are marked by black arrowheads. (D) Real-time observation of crystalline cellulose III, incubated with 1 μM CfCel48A. The time interval between images is 0.3 s (200 × 100 nm², 120 × 50 pixels). These images are taken from Movie S3. The positions of CfCel48A are marked by purple arrowheads.

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 CfCel6B (GH6) and CfCel48A (GH48) on crystalline cellulose III. The other two cellulases, CfCel6A (GH6) and CfCel9A (GH9), were bound to cellulose III 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: TrCel7A (33), TrCel7B (34), and TrCel6A (24), derived from T. reesei, and P. chrysosporium (Fig. 3). 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 TrCel7A, PcCel7D, CfCel6B, and CfCel48A have been confirmed by HS-AFM, but processive movements of TrCel7B, TrCel6A, CfCel6A, and CfCel9A have not been detected.

The difference in processivity between the two bacterial GH6 cellulases CfCel6A and CfCel6B is presumably due to their structural difference (Fig. 3A). As proposed by Varrot et al., GH family 6 cellulases can be clustered in four subfamilies based on amino acid sequence similarity (37), and CfCel6A belongs to subfamily 2, while CfCel6B belongs to subfamily 4. Among subfamily 2 cellulases, the structures of MtCel6A (amino acid identity with the catalytic domain of CfCel6A of 47%) and TfCel6A (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 TrCel6A (Fig. 3A); 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 cellulase degradation by the fungal family 6 cellulose TrCel6A (39). In contrast, subfamily 4 cellulase CfCel6B, which has 53% amino acid identity with the catalytic domain of TjCel6B, 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 CfCel6A and CfCel6B seem consistent with the...
observed difference in processivity (Fig. 3A). Although our previous HS-AFM study of TrCel6A 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 CfCel6B, suggesting that the presence of four loops that cover the active-site cleft and construct a longer tunnel than in TrCel6A 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. 3C). Cellulases without a loop covering the active cleft, such as CfCel9A, CfCel6A, and TrCel7B, can all be classified biochemically as endotype cellulases. For two GH7 cellulases, PtCel17D and TrCel7B, 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 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.
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 acetamide 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 CfCel6B and TrCel7A, 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. 2E) 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

| Cellulase                | Dissociation rate constant, s⁻¹ | Half-life of movement, s | Average velocity, nm s⁻¹ | Calculated processivity, No. of reactions |
|--------------------------|---------------------------------|--------------------------|--------------------------|------------------------------------------|
| CfCel6B                  | 0.55 ± 0.02                     | 1.3 ± 0.1                | 12.7 ± 7.6               | 16.1                                     |
| PcCel7C‡                 | 0.51 ± 0.01                     | 1.4 ± 0.0                | 14.7 ± 9.1               | 19.9                                     |
| PcCel7D‡                 | 0.32 ± 0.04                     | 2.2 ± 0.1                | 9.4 ± 3.7                | 20.3                                     |
| TrCel7A‡                 | 0.20 ± 0.01                     | 3.5 ± 0.2                | 6.8 ± 3.5                | 23.6                                     |
| CfCel6B (with TrCel7A)   | 0.47 ± 0.03                     | 1.5 ± 0.1                | 19.4 ± 8.1               | 28.8                                     |
| TrCel7A (with CfCel6B)   | 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.

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 TrCel7A (PDB:1CEL), TrCel7B (PDB:1EG1), and TrCel6A (PDB:1QK2) and P. chrysosporium PcCel7D (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 TrCel7A, 3.5; TrCel7B, 4.0; PcCel7D, 4.0; TrCel6A, 2.0; CfCel6A, 1.0; CfCel6B, 3.0; CfCel9A, 0.75; and CfCel48A, 3.5. Blue highlighting indicates the area observed by HS-AFM.
neighboring chains are more tightly held together and thus are more accessible to the other enzyme. In addition, it was observed that C/Cel6B sometimes collided with TrCel7A, causing a traffic jam (Fig. 2D and Movie S6). T. reesei is thought to degrade crystalline cellulose by employing the combination of TrCel7A and TrCel6A, both of which are thought to be processive CBHs. However, we could not observe processivity of TrCel6A by means of HS-AFM, which may mean that the retaining force on the cellulose surface is weaker than that of TrCel7A or C/Cel6B.

This situation could have evolved to prevent the occurrence of immovable traffic jams because TrCel6A would be knocked off the substrate when a collision occurs. The combination of C/Cel6B and TrCel7A did show collisions resulting in congestion, which may be the reason why the synergistic effect between C/Cel6B and TrCel7A (1.4 times, Fig. 2E) is less than that between TrCel6A and TrCel7A [2.3 times (3)]. Similarly, C. fimti secretes C/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 C/Cel9A and C/Cel6B may be similar to that between TrCel7A and TrCel6A from T. reesei. Although T. reesei and C. fimti 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. fimti, the highly processive C/Cel6B moves from the nonreducing to the reducing end, while in T. reesei, the highly processive TrCel7A moves in the opposite direction, from the reducing to the nonreducing end (Fig. 2B).

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 create a 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). P. 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 TrCel7B the subsite is asymmetric (Sub/Pro ratio of 8/2), while in TrCel7A 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 (α/α)6 fold. GH8 from Clostridium thermocellum (HtCel8A) has a symmetrical subsite (Sub/Pro ratio of 3/3) with no loop covering the active center (44), while GH48 cellulase from Clostridium cellulolyticum (RcCel48F) 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 C/Cel6B with four loops and a Sub/Pro ratio of 6/2 is quite similar to that of TrCel7A, 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 endoglucanases to generate a processive cellulase. Thus, the GH6, GH48, and GH7 processive cellulases all utilize common strategies: extending the protein “nose” in the
substrate direction to form asymmetric subite 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 1, and III were prepared from green algae Cladophora spp. according to a previous report (14). PASC was prepared from Avicel (Fancieli et al.) According to a previous report (4). CMc was purchased from Herculos. Cellulose-assay publites were acquired from Sekagaku Kogyo. Cellulomonas fimih ATCC 4848 was used as a source of CFe6A, CFe6B, CFe6H, and CFe64A genes. Trichoderma reesei ATCC 66589 was used as a source of the TrCeA gene. Escherichia coli JM109 (Takara) was used as the subcloning host. E. coli Rosetta2 (DE3), Tuner (DE3) (Novagen), and Brevibacillus choshiriensis HPD31-S3 (Takara) were used for heterologous production of the recombinant proteins.

**Cloning of Four Bacterial Cellulases from C. finli.** Four C. finli cellulase genes were amplified by PCR. The CFe6A gene and CFe66B gene were cloned into the E. coli expression vector. The CFe6A gene and CFe64A gene were cloned into the B. choshiriensis expression vector. For details, see SI Appendix.

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

**Determination of the Degree of Polymerization of Cellulose III.** The DP of cellulose III, was estimated from the fiber lengths measured in HS-AFM scanning images (scanning area: 0.64 to 9.00 μm²). 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 ≥ 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).

**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 server (version 9). The “optimal” analysis mode was applied to each set of sequences, using MUSCLE (version 3.8) (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 19H04512 and 19H05389 to T. Uchiyama) 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 19H01827 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 utilization and modification (BioAB).”

---

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 cellulase. 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 censa of Cellulomonas fimih. Gene 44, 315–324 (1986).
7. A. Meinke et al., Cellobiohydrolase A (CbhA) from the cellulolytic bacterium Cellulomonas fimih 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 fimih 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., Advancements in cellulase engineering. Adv. Biochem. Eng. 54, 67–74 (1995).
10. L. E. Sanders, A. Meinke, N. R. Gilkes, D. G. Kilburn, R. A. J. Warren, Degradation of cellulases in cultures of Cellulomonas fimih. FEBS 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. H. Souchon, B. L. Lytle, J. H. David Wu, P. M. Alzari, The crystal structure and catalytic mechanism of cellobiohydrolase Cel7A. FEBS J. 274, 1785–1792 (2007).
14. M. Wada, L. Heux, J. Sugiyama, Polymorphism of cellulose I family: Reinvestigation of cellulose IV. Cellulose 6, 331–337 (1999).
15. I. von Ossowski et al., The structure of a bacterial cellobiohydrolase: The catalytic core of cellulobiohydrolase Cel7D (CBH58) at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. 272, 383–397 (1997).
16. M. Wada, H. Chanzy, Y. Nishiyama, P. Langan, Cellulose III crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. Macromolecules 33, 8548–8555 (2004).
17. 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. Costylev 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. E. Dume et al., The three-dimensional crystal structure of the catalytic core of cellulobiohydrolase I from Trichoderma reesei. Science 265, 524–528 (1994).
34. J. Klevegård et al., The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. 272, 383–397 (1997).
35. M. Wada, L. Heux, J. Sugiyama, Polymorphism of cellulose I family: Reinvestigation of cellulose IV. Cellulose 6, 331–337 (1999).
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. M. Spezio, D. B. Wilson, P. A. Karplus, Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry 32, 9906–9916 (1993).
38. A. Kivula et al., Tryptophan 272: An essential determinant of crystalline cellulose degradation by Trichoderma reesei cellobiohydrolase Cel6A. FEBS Lett. 429, 341–346 (1998).
39. K. Igarashi et al., Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. Nat. Commun. 5, 3975 (2014).
40. A. M. Larson et al., Crystal structure of Thermobifida fusca endoglucanase Cel6A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion ring. Biochemistry 44, 12915–12922 (2005).
41. A. Dereeper et al., Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W507–W510 (2008).
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. F. M. A. Gutirrez et al., Atomic (0.94 Å) 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. Golaoncza Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D269–D275 (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. Barfuß, B. Jacq, R. Christen, Treedyn: Towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7, 439 (2006).