The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions

Cedric Montanier[1], Victoria A. Money[2], Virginia M. R. Pires[3], James E. Flint[1], Benedita A. Pinheiro[3], Arun Goyal[3], José A. M. Prates[3], Atsushi Izumi[2], Henrik Stålbrand[4], Carl Morland[1], Alan Cartmell[1], Katarina Kolenova[4], Evangelos Topakas[1], Eleanor J. Dodson[2], David N. Bolam[1], Gideon J. Davies[2*], Carlos M. G. A. Fontes[3], Harry J. Gilbert[1*]

1 Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne, United Kingdom, 2 York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, United Kingdom, 3 CIISA - Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, Lisboa, Portugal, 4 Department of Biochemistry, Center for Chemistry and Chemical Engineering, Lund University, Lund, Sweden

Multifunctional proteins, which play a critical role in many biological processes, have typically evolved through the recruitment of different domains that have the required functional diversity. Thus the different activities displayed by these proteins are mediated by spatially distinct domains, consistent with the specific chemical requirements of each activity. Indeed, current evolutionary theory argues that the colocalization of diverse activities within an enzyme is likely to be a rare event, because it would compromise the existing activity of the protein. In contrast to this view, a potential example of multifunctional recruitment into a single protein domain is provided by *Ct*Cel5C-CE2, which contains an N-terminal module that displays cellulase activity and a C-terminal module, *Ct*CE2, which exhibits a noncatalytic cellulose-binding function but also shares sequence identity with the CE2 family of esterases. Here we show that, unlike other CE2 members, the *Ct*CE2 domain displays divergent catalytic esterase and noncatalytic carbohydrate binding functions. Intriguingly, these diverse activities are housed within the same site on the protein. Thus, a critical component of the active site of *Ct*CE2, the catalytic Ser-His dyad, in harness with inserted aromatic residues, confers noncatalytic binding to cellulose whilst the active site of the domain retains its esterase activity. *Ct*CE2 catalyses deacetylation of noncellulosic plant structural polysaccharides to deprotect these substrates for attack by other enzymes. Yet it also acts as a cellulose-binding domain, which promotes the activity of the appended cellulase on recalcitrant substrates. The CE2 family encapsulates the requirement for multiple activities by biocatalysts that attack challenging macromolecular substrates, including the grafting of a second, powerful and discrete noncatalytic binding functionality into the active site of an enzyme. This article provides a rare example of "gene sharing," where the introduction of a second functionality into the active site of an enzyme does not compromise the original activity of the biocatalyst.

Introduction

The different activities displayed by multifunctional proteins are typically domain specific. A natural enzyme system that contains large numbers of proteins with complex molecular architectures is presented by the plant cell-wall–degrading apparatus from a range of microbial species. Plant cell wall degradation, now of great environmental significance, particularly with respect to the generation of renewable and sustainable biofuels [1,2], is a challenging process that requires a large consortium of different enzyme activities. The plant cell wall consists primarily of an array of interlocking polysaccharides. While cellulose, which forms crystalline microfibrils, has a simple chemical structure consisting of β-1,4-linked glucopyranoside moieties, the matrix polysaccharides are chemically complex molecules in which the backbone polymers are protected with both sugars and organic esters [3]. Plant cell-wall–degrading systems thus feature glycoside hydrolases, which cleave the glycosidic bonds that link the sugars, and esterases that remove the diverse acylations [4,5]. In addition to their chemical complexity, plant cell walls have a physical structure that

presents a physical barrier to enzyme attack. To compensate for the accessibility problem, plant cell-wall–degrading enzymes generally contain a noncatalytic carbohydrate binding function that, by bringing the biocatalyst into prolonged and intimate contact with its substrate, increases the rate of catalysis [6]. In general these diverse catalytic and noncatalytic carbohydrate-binding activities are housed in discrete modules within the same protein [7].

Abbreviations: 4-NPAc, 4-nitrophenyl acetate; AGE, affinity gel electrophoresis; CBM, carbohydrate-binding module; ITC, isothermal titration calorimetry; rsmd, root mean square deviation

* To whom correspondence should be addressed. E-mail: davies@ysbl.york.ac.uk (GJD); H.J.Gilbert@ncl.ac.uk (HJG)

❂ These authors contributed equally to this work.

Author Summary

Proteins that display multiple activities have typically evolved through the recruitment of different domains, each of which has a specific function. Thus, in a multifunctional protein, the different activities are mediated by spatially distinct domains such that a single domain can provide the specific chemical requirements for one activity. Indeed, current evolutionary theory argues that the colocalization of diverse activities within a single-domain enzyme is likely to be a rare event, as it would compromise the existing activity of the protein when a new function evolves. Nonetheless, a potential example of multifunctional recruitment into a single protein domain is provided by an enzyme that contains a cellulase enzyme module and a discrete noncatalytic cellulose-binding module. In this article, we show that the cellulose-binding module displays esterase activity and that these diverse activities are housed within the same site on the protein. Structural analysis of the enzyme reveals that its catalytic residues also contribute to the noncatalytic cellulose-binding function. This report provides a rare example of "gene sharing," whereby the introduction of a second functionality into the active site of an enzyme does not compromise the original activity of the catalyst.

A family of carbohydrate esterases (family CE2 in the CAZy classification (http://www.cazy.org/ [7,8]) are especially intriguing with respect to the multiple functions required of plant cell-wall–degrading systems. CE2 enzymes are acetyl esterases (that are generally not appended to other catalytic modules), which are reported to be active on synthetic aryl-esters and acetylated xylan [9]. The bacterium *Clostridium thermocellum* contains a single CE2 member, designated *Ct*CE2, which is linked to the cellulase, *Ct*Cel5C, within the modular protein designated *Ct*Cel5C-CE2. The enzyme also contains a type I dockerin module that, by binding to cohesin modules in the scaffoldin protein, incorporates *Ct*Cel5C-CE2 into the multi-enzyme plant cell-wall–degrading complex known as the cellulosome [10] (Figure 1). Intriguingly, *Ct*CE2 was previously characterized as a carbohydrate-binding module (CBM) by virtue of its cellulose-binding capacity and its ability to potentiate the cellulase activity of the linked *Ct*Cel5C catalytic module [11,12]. This unusual, potentially dual,

activity of the *Ct*CE2 module prompted us to investigate the functional nuances within the CE2 family.

Here we report the biochemical properties and crystal structure of several CE2 members, both single-module CE2 enzymes and *Ct*CE2, a component of *Ct*Cel5C-CE2. The data show that CE2 enzymes are α/β-hydrolases in which *Ct*CE2 displays a unique dual function within the same region of the protein scaffold. The enzyme module displays acetyl esterase activity that is spatially coupled to a noncatalytic cellulose-binding function. The binding ability directs the appended cellulase module *Ct*Cel5C to, and facilitates its activity on, cellulose. Structural and biochemical analyses reveal that the grafting of aromatic residues into the substrate binding cleft of the enzyme, primarily in harness with His-791, a critical component of the Ser-His catalytic apparatus, also plays an important role in the cellulose-binding function. By contrast, the single-module CE2s, although exhibiting substantial structural homology and similar catalytic activities to the *Clostridium* esterase, do not bind cellulose. We demonstrate how subtle modifications in the active centres of different members of an enzyme family leads to functional divergence, which is manifested by the gain of dual function within the same environment of the protein.

Results

CE2 Is a Large Family of Diverse Esterases Displaying the α/β-Hydrolase Fold with a Serine Nucleophile

To probe the function and role of the diverse members of the CE2 esterase family, which currently contains 31 members (http://www.cazy.org/fam/CE2.html), the biochemical properties of five CE2 enzymes were assessed. In addition to *Ct*CE2, which comprises the C-terminal region of *C. thermocellum* *Ct*Cel5C-CE2 (formerly EGE; [11,12]) we also characterised three *Cellvibrio japonicus* CE2 members, *Cj*CE2A, B and C and the *Bacteroides thetaiotaomicron* esterase *Bt*CE2, identified from the genome sequence of the two bacteria [13,14]; these latter four enzymes are not appended to other enzyme modules (Figure 1). All of these CE2 enzymes act as acetyl esterases, releasing acetate from activated artificial substrates such as 4-nitrophenyl acetate (4-NPAc; see Table 1) and, to different
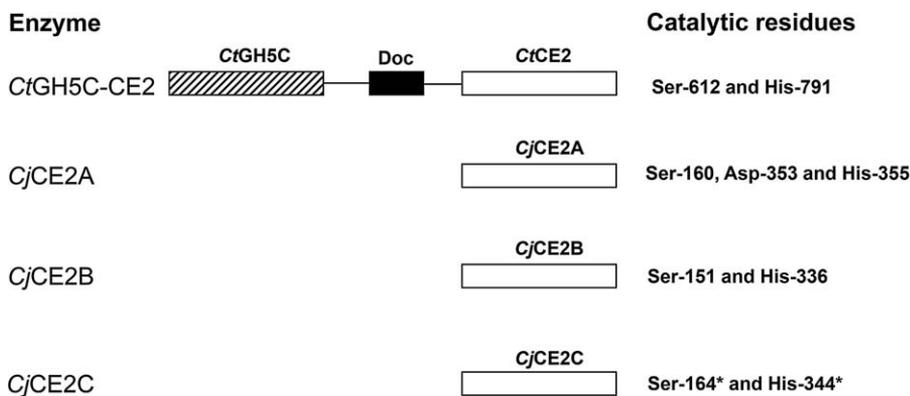
Figure 1. Schematic of the Molecular Architecture of the Enzymes Investigated in This Study

Enzyme		Catalytic residues
*Ct*GH5C-CE2	*Ct*GH5C — Doc — *Ct*CE2	Ser-612 and His-791
*Cj*CE2A	*Cj*CE2A	Ser-160, Asp-353 and His-355
*Cj*CE2B	*Cj*CE2B	Ser-151 and His-336
*Cj*CE2C	*Cj*CE2C	Ser-164* and His-344*

The mature forms of the *C. japonicus* esterases consist of a single catalytic module. The *C. thermocellum* enzyme containing a CE2 module (*Ct*Cel5C-CE2) consists of an N-terminal GH5 cellulase module (*Ct*Cel5C), a central type I dockerin module (Doc) that facilitates the integration of the enzyme into the cellulosome, and a C-terminal CE2 module (*Ct*CE2). The CE2 modules of all the esterases contains an N-terminal domain of ~160 residues that displays a jelly roll fold and a C-terminal domain that exhibits an α/β-hydrolase fold. The black lines are ~15-residue P/T linker sequences. The residues in *Cj*CE2C that carry an asterisk are predicted catalytic residues based on sequence alignments.
doi:10.1371/journal.pbio.1000071.g001

Table 1. Catalytic Activity of CE2 Enzymes against 4-Nitrophenyl Acetate (4-NPAc)

Enzyme	4-NPAc		
	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} μM^{-1})
CtCE2	7,032	165	43
D789A[a]	2,512	75	33
D789N[a]	2,976	186	16
S612A[a]	NA	NA	NA
H791A[a]	NA	NA	NA
CjCE2A	7,717	339	23
CjCE2B	49,820	262	190
CjCE2C	7,251	45	161

[a] Mutants of CtCE2.
NA no activity detected
doi:10.1371/journal.pbio.1000071.t001

extents, the acetylated plant polysaccharides xylan and glucomannan, Table 2 (note that only very small amounts of the *Bacteroides* CE2 could be produced and so only initial qualitative assays could be carried out with this enzyme). Based on their catalytic efficiencies, *CtCE2* and *CjCE2B* exhibit a significant preference for acetylated glucomannan over xylan, whereas *CjCE2A* and *CjCE2C* do not distinguish between the two polysaccharides. It should be noted that the specificity of *CtCE2* for glucomannan reflects an extremely low K_M. The esterases appear specific for acetyl groups and do not hydrolyse aryl-ferulates or aryl-coumarates (unpublished data). The only other reported analysis of the catalytic activity of CE2 enzymes are the esterases from *Neocallimastix patriciarum* [9] that display xylan esterase activity and hydro-

lyse 4-NPAc, but their activity on other substrates was not assessed. The activities reported here extend the substrates known to be deacetylated by CE2 enzymes

For three of these enzymes, we were able to obtain 3D crystal structures (Figure 2A–2C). These include the CE2 module of *CtCel5C-CE2* and two of the *Cellvibrio* enzymes, *CjCE2A* and *CjCE2B*. All three 3D structures reveal a bi-domain enzyme in which an N-terminal β-sheet "jelly roll" domain (around 130 residues) is linked to a C-terminal domain of approximately 220 residues. The C-terminal domains possess an atypical α/β-hydrolase (SGNH-hydrolase) fold [15], consisting of repeating β-α-β motifs that form a curved central five-stranded parallel β-sheet, in the strand order β2, β1, β3, β4, and β5 with strand β2 interrupted by loop insertion. The sheet packs against two α-helices (α1 and α6 on the concave side and three α-helices, α2, α4, and α5, on the convex side), all of which are antiparallel to the β-strands. There is also a small α-helix (α3 in the loop connecting β3 and α4 and a 3$_{10}$ helix between β1 and α1. Superimposition of *CtCE2* with the two *Cellvibrio* esterases reveals that both the N-terminal β-sheet domain (root mean square deviation [rmsd] of 1.5 and 1.9 over 107 and 94 Cα atoms of *CjCE2A* and *CjCE2B*, respectively) and the C-terminal catalytic domain (rsmd of 1.3 and 1.4 over 196 and 195 Cα atoms of *CjCE2A* and *CjCE2B*, respectively) display considerable structural conservation.

Structural similarity searches using secondary-structure mapping [16] indicate that the C-terminal α/β-hydrolase domain contains the esterase catalytic centre in which, to date, a Ser-His dyad is invariant across the whole CE2 landscape. *CjCE2A* is a canonical serine hydrolase, with a classical Ser-His-Asp triad, in which Ser-160 is the catalytic nucleophile, His-335 activates the serine, and Asp-333 makes a hydrogen bond with Nδ1 of His-335 thus completing the

Table 2. Activity of CE2 Esterases against Acetylated Polysaccharides

Enzyme	Birchwood Xylan			Glucomannan		
	k_{cat} (min^{-1})	K_M[a] (mM)	k_{cat}/K_M (min^{-1} mM^{-1})	k_{cat} (min^{-1})	K_M[a] (mM)	k_{cat}/K_M (min^{-1} mM^{-1})
CtCE2	12	2.7	4.4	1.1	0.019	58
D789A[b]	4.5	3.1	1.5	0.23	0.025	9.2
D789N[b]	6.5	2.9	2.2	0.71	0.031	23
S612A[b]	—	—	NA[c]	—	—	NA[c]
H791A[b]	—	—	NA	—	—	NA
Y665A[b]	15	7.7	1.9	ND[d]	ND	32
Y683A[b]	10	2.4	4.2	0.56	0.026	21
W746A[b]	13	3.3	4.1	10	0.14	71
Q780A[b]	8.5	2.2	3.9	2.0	0.046	43
W790A[b]	14	2.5	5.6	7.9	0.088	89
Y683A/W746A[b]	13	2.9	4.3	12	0.13	89
W746A/Q780A[b]	17	3.1	5.5	9.4	0.12	78
Y683A/W746/Q780A[b]	22	4.2	5.1	ND	ND	31
CjCE2A	69	1.5	46	163	4.2	39
W127A[e]	65	10	6.5	217	87	2.5
W212A[e]	55	6.2	8.9	468	58	8.1
W127A/W212A[e]	12	14	0.85	4.6	37	0.12
CjCE2B	66	4.6	14	1348	0.84	1,605
CjCE2C	119	3.2	37	1212	14	87

[a] The K_M values relates to the concentration of acetate moieties present in the polysaccharides.
[b] Mutants of CtCE2. [c] NA: No activity detected. [d] ND: The K_M and k_{cat} values could not be determined as there was a linear relationship between rate and substrate concentration up to 0.6 mM.
[e] Mutants of CjCE2A.
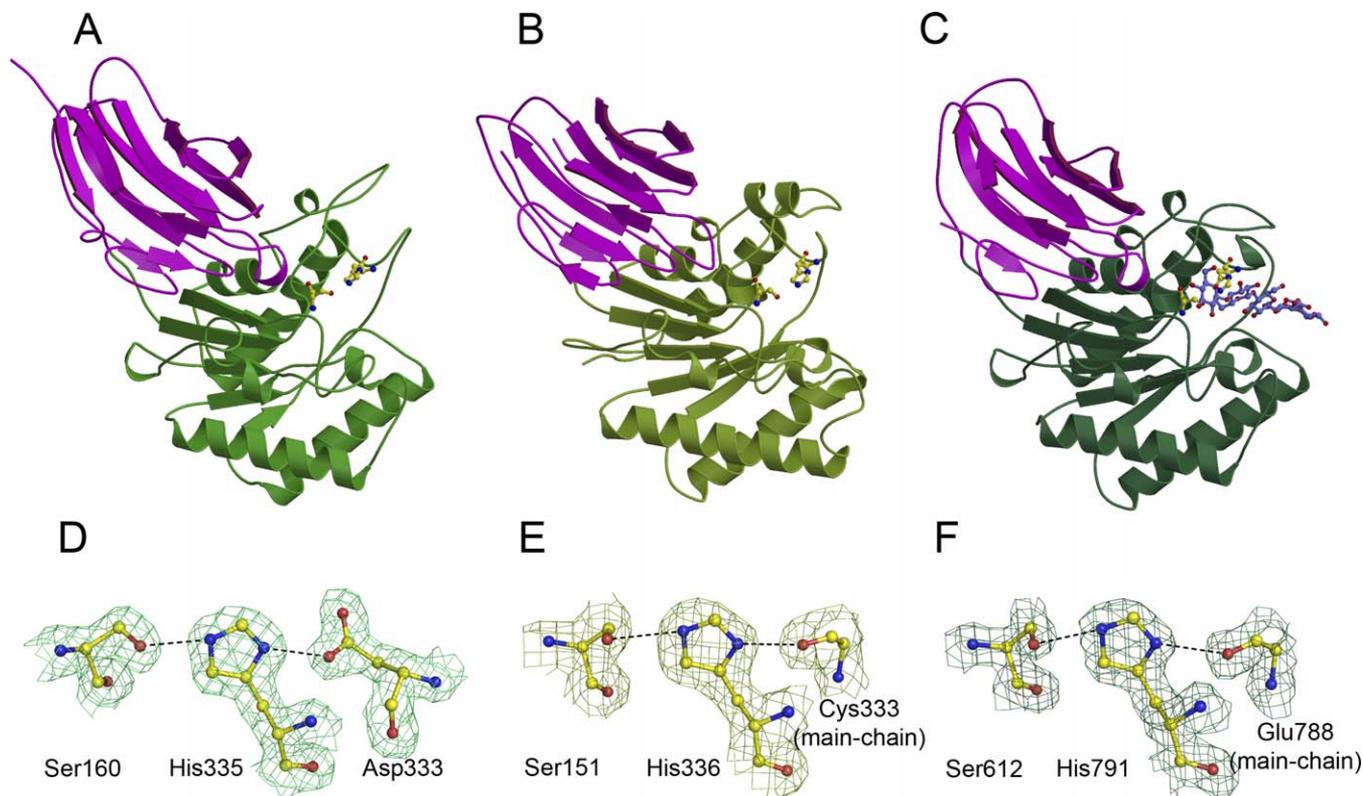doi:10.1371/journal.pbio.1000071.t002

Figure 2. 3-D Structures and Catalytic Constellation Geometries for Three CE2 Esterases

(A) 3-D structure of *Cj*CE2B with the catalytic domain in green and the β-sheet domain in magenta. The catalytic Ser and His are shown in ball-and-stick representation.
(B) *Cj*CE2A, drawn as above.
(C) *Ct*CE2 as above and with cellopentaose in blue.
(D) The catalytic Ser-His-Asp triad of *Cj*CE2A.
(E) The catalytic Ser-His dyad with the main-chain carbonyl interaction from Cys333 of *Cj*CE2B.
(F) The catalytic dyad and main-chain carbonyl of *Ct*CE2.
This figure was drawn with PyMOL (DeLano Scientific, http://pymol.sourceforge.net/).
doi:10.1371/journal.pbio.1000071.g002

catalytic triad (Figure 2D). The oxyanion hole, which stabilises the incipient tetrahedral transition-state, comprise the N of Ser-160 and Gly-205 and the Nδ2 of Asn-255. Indeed, in the pocket within the cleft of *Cj*CE2A is a formate molecule, mimicking the reactive intermediate, which makes hydrogen bonds with the residues that form the oxyanion hole.

In the case of *Ct*CE2 the putative catalytic dyad is provided by Ser-612 and His-791 consistent with the observation that the mutants S612A and H791A display no esterase activity, Tables 1 and 2. In this enzyme the oxyanion hole, similar to *Cj*CE2A, comprise the N of Ser-612 and Gly-658 and the Nδ2 of Asn-705 with formate again mimicking the tetrahedral transition state. *Ct*CE2 along with *Cj*CE2B does not possess a side-chain residue equivalent to the Asp of a classical Ser-His-Asp triad. Instead, these enzymes display a catalytic dyad with stabilisation of the histidine provided by main-chain carbonyl groups (Figure 2E and 2F). *Ct*CE2 does possess an aspartate whose carboxylate group is approximately 5–6 Å from the His-791 Nε1, which could, conceivably and with considerable conformation change, make an appropriate interaction with the imidazole ring. In order to probe this possibility, the D789A and D789N mutants were constructed but both retain significant catalytic activity against 4-NPAc and the polymeric substrates (Tables 1 and 2). Instead of the more typical

Ser-His-Asp, the "triad" geometry is completed through the Nε1 of His-791 making a hydrogen bond with the backbone carbonyl of Glu-788, in the case of *Ct*CE2 (Figure 2F), whilst the equivalent interaction in *Cj*CE2B is with the main-chain carbonyl of Cys-333 (Figure 2E). Catalytic dyads, although rare, have been observed in different forms elsewhere. Other than the CE2 enzymes reported here, IroE is an α/β-hydrolase peptidase that also features a catalytic Ser-His dyad [17]. Given that current wisdom suggests that the Asp of the triad does not act as a base [18], both IroH and now CE2 highlight the need for correct orientation of histidine, which may, apparently, equally well be achieved through an interaction with a carbonyl moiety rather than a carboxylic acid.

The unusual divergence to a dyad geometry in CE2 likely reflects the inclusion of a loop modification incorporating a tryptophan residue. In *Ct*CE2, for example, one of the key residues involved in cellulose recognition is Trp-790 in the sequence Asp-Trp-His. Inclusion of the Trp forces the adjacent aspartate into a conformation that prevents it hydrogen-bonding with His-791 (Figure 3). The dyad geometry and the topology of the active-centre binding surfaces are described below in light of the dissection of polysaccharide recognition in CE2 enzymes. Sequence analysis suggests that the dyad geometry is not unique to these two enzymes,
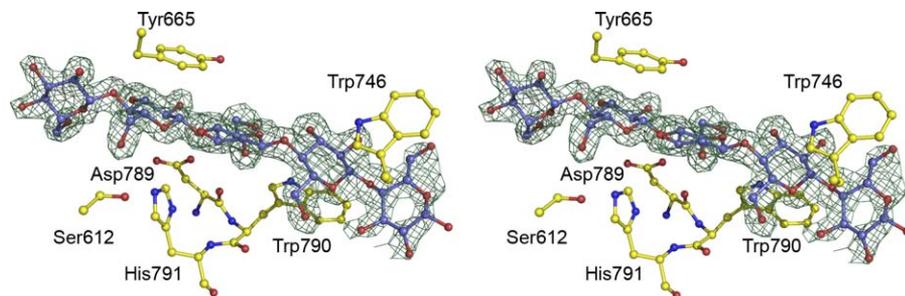
Figure 3. Binding of Cellooligosaccharides through the Esterase Active Centre of *Ct*CE2

Observed electron density (maximum-likelihood weighted $2F_{obs} - F_{calc}$ contoured at 1 σ) for cellohexaose (in which cellopentaose is ordered) bound to wild-type *Ct*CE2. Ser-612 and His-791 form the catalytic dyad, with Trp-790 causing a change in position of Asp-789 and forming the binding-platform for the second glucose. Other interactions with aromatic residues discussed in the text are shown. The figure is in divergent ("wall-eyed") stereo and was drawn with PyMOL.
doi:10.1371/journal.pbio.1000071.g003

with 14 of the 26 CE2 members lacking the putative catalytic aspartate. Furthermore, of the remaining 12 CE2 enzymes that appear to have a canonical catalytic triad, 10 have a tryptophan between the His and Asp, as observed with *Ct*CE2, suggesting that in these enzymes the aromatic ring may also place the Asp into an orientation that prevents it forming a classical triad.
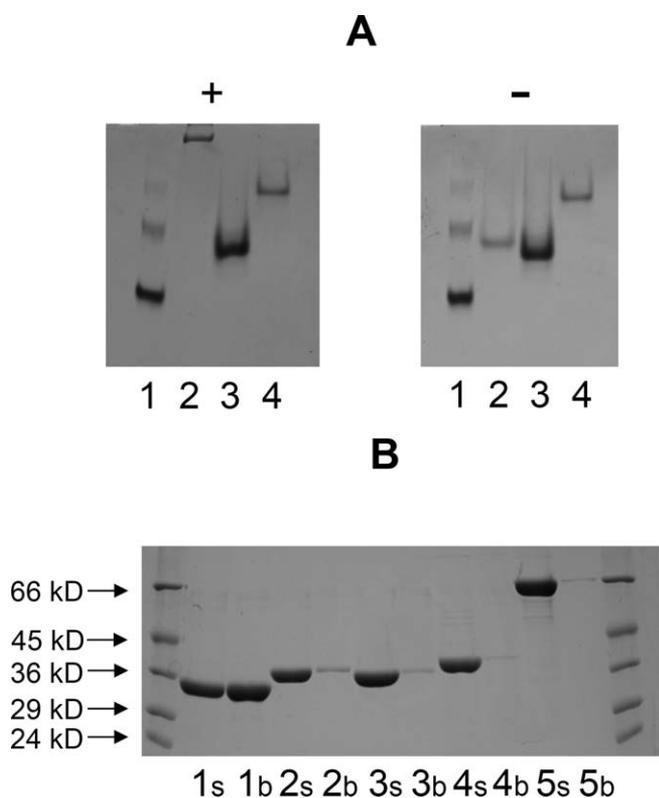

Figure 4. Binding of CE2 Esterases to Cellulose

(A) An AGE experiment in which the enzymes were subjected to nondenaturing gel electrophoresis in the absence (–) or presence (+) of 0.1 % (w/v) hydroxyethylcellulose. The lanes contained BSA (1), *Ct*CE2 (2), *Cj*CE2A (3), and *Cj*CE2C (4). *Cj*CE2B did not migrate on the nondenaturing gel. (B) Shows a pull-down experiment using insoluble cellulose. The original protein samples (s) and bound protein eluted from cellulose with 10% SDS (b) were subjected to SDS-PAGE. The lanes contained *Ct*CE2 (1), *Cj*CE2A (2), *Cj*CE2B (3), *Cj*CE2C, (4), and BSA as a noninteracting control (5). Molecular weight markers are shown.
doi:10.1371/journal.pbio.1000071.g004

The Esterase and Cellulose-Binding Functions of *Ct*CE2 Are in Close Proximity

The most intriguing CE2 member is *Ct*CE2, which is derived from the multidomain cellulase/esterase *Ct*Cel5C-CE2. Historical work had shown that *Ct*CE2 not only bound cellulose but also potentiated the activity of the cellulase catalytic module on insoluble substrates [12]: the CE2 module thus behaved as a classical CBM [19]. In this report we show that *Ct*CE2 binds to insoluble cellulose (Figure 4). Affinity gel electrophoresis (AGE) also demonstrated that the protein interacts with β-glucan (β-1,3:β-1,4 mixed linked glucan) and soluble derivatized forms of cellulose (carboxymethylcellulose and hydroxyethylcellulose; Figure 4), but displays no affinity for laminarin (β-1,3-linked glucan), α-linked glucans or the β-1,4-linked xylose polymer, xylan (unpublished data). Isothermal titration calorimetry (ITC) revealed that *Ct*CE2 binds to cellooligosaccharides with a K_D for cellohexaose of 33 µM (Figure 5; Tables 3 and 4), cellopentaose of 71 µM, cellotetraose of 333 µM and cellotriose >1 mM (Table 4), but no binding to mannohexaose or xylohexaose was detected (unpublished data). The K_D for β-glucan was found to be similar to that for cellohexaose (Table 4).

Interestingly, whilst all CE2s possess a β-sheet domain that is reminiscent of many CBMs [19], it is not this domain that interacts with cellulose in *Ct*CE2. As was revealed in the 3-D complex structure (Figure 2) and discussed in more detail below, cellooligosaccharide binding, with a stoichiometry of 1, occurs across the esterase active centre with both the Ser and His of the dyad influencing ligand recognition. The S612A mutation of *Ct*CE2 yields an approximately 8-fold increase in affinity (Figure 5; Tables 3 and 4), whilst the H791A amino acid substitution reduces binding with the K_D increasing approximately 20-fold (Tables 3 and 4). Furthermore, cellohexaose and β-glucan binding inhibit the esterase activity of the wild-type *Ct*CE2 enzyme (Table 3), which is discussed in detail below.

This capacity to bind cellulose is not a common feature within the CE2 family. Pull-down assays revealed that none of the *Cellvibrio* enzymes bound to insoluble cellulose (Figure 4). Furthermore, the *C. japonicus* CE2 enzymes and the *B. thetaiotaomicron* esterase were not inhibited by cellohexaose, mannohexaose, or xylohexaose (unpublished data) at 300 µM, while ITC also revealed no binding of the *C. japonicus* enzymes to these oligosaccharides (Table 3). While the binding of *Cj*CE2A or *Cj*CE2C to polysaccharides could not be detected
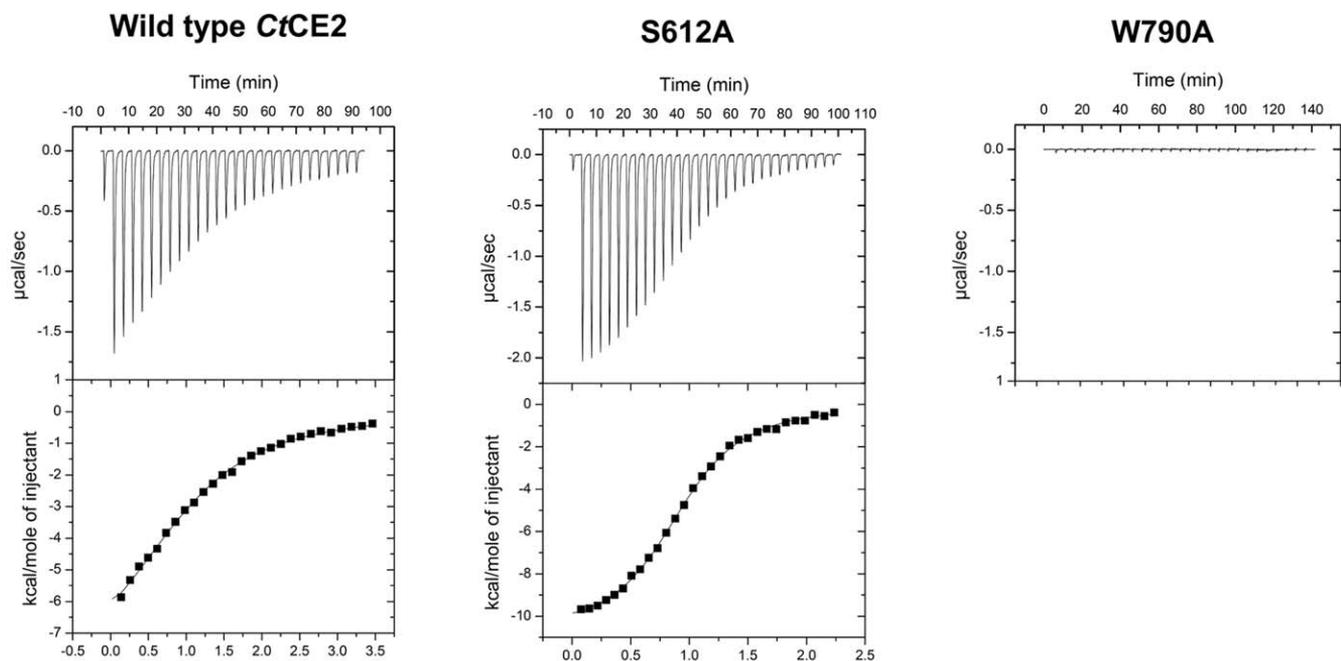
Figure 5. Examples of Isothermal Titration Calorimetry of Wild-Type and Mutants of *Ct*CE2
The proteins were titrated with cellohexaose in 50 mM sodium HEPES buffer, pH 7.0, at 25 °C. The protein concentration for each titration was 100 μM.
doi:10.1371/journal.pbio.1000071.g005

by ITC (unpublished data) or AGE (Figure 4), *Cj*CE2B interacted with, β-glucan, albeit with ~17-fold lower affinity than *Ct*CE2 (Table 3), but did not recognize carboxymethyl-cellulose (unpublished data) or hydroxyethylcellulose (Figure 4). These data reveal that the capacity of *Ct*CE2 to recognise cellulose is, possibly, a unique feature within the CE2 family, and one that prompted its study by X-ray crystallography.

The Structural Basis for Substrate Recognition in the CE2 Family

The catalytic apparatus of the CE2 esterases is located within a cleft that extends across the catalytic domain and is

Table 3. The Binding of CE2 Esterases to Cellohexaose and β-Glucan

Enzyme	Cellohexaose (K_i; μM)	Cellohexaose (K_D; μM)	β-Glucan (K_i; μM)	β-Glucan (K_D; μM)
*Ct*CE2	3.1	33	7	31
*Cj*CE2A	NI[a]	NB[b]	NI	NB
*Cj*CE2B	NI	NB	34	538
*Cj*CE2C	NI	NB	129	NB
H791A[c]	—[d]	667	—	—
S612A[c]	—	4.1	—	—
W746A[c]	—	>1,000	—	>1,000
W790A[c]	—	NB	—	NB
Y665A[c]	—	NB	—	NB

Binding was determined by ITC to yield the K_D values and inhibition of 4-NPAc hydrolysis to provide the K_i values.
[a]NI: No inhibition detected with 300 μM of ligand.
[b]NB: No binding detected with 1 mM of ligand.
[c]Mutants of *Ct*CE2.
[d]— indicates experiment not done.
doi:10.1371/journal.pbio.1000071.t003

likely to constitute the substrate binding site of these enzymes. To explore this possibility, the effect of removing the aromatic side chains (which play a pivotal role in protein-carbohydrate recognition [19]) that line the putative substrate binding site (Figure 3) on the catalytic activity of the esterases was investigated. The data, reported in Table 2, show that none of the aromatic residue mutations (Y665A, W746A, and W790A) influenced the kinetic parameters of the esterase against xylan. While the catalytic efficiencies of the *Ct*CE2 variants against glucomannan were similar to the wild-type enzyme, there was a significant increase in K_M, exemplified by the Y665A mutant ($K_M > 0.6$ mM), which was mirrored by a similar increase in k_{cat}. It should also be noted that the activity of the Tyr-665, Trp-790, and Trp-746 mutants of *Ct*CE2 against acetylated glucomannan was subject to substrate inhibition. Thus, these aromatic residues not only bind glucomannan but also guide the polysaccharide into the substrate binding cleft such that the acetyl groups are presented at the active site. These mutational studies indicate that the aromatic residues lining the substrate binding cleft of *Ct*CE2 not only contribute to the tight binding of acetylated glucomannan, reflected by the very low K_M, but also limit oligosaccharide product release after k_2, which appears to be the rate-determining step in catalysis. Mutation of the single aromatic residue in the substrate binding cleft of *Cj*CE2A, Trp-212, although not effecting k_{cat}, leads to a significant increase in K_M for both glucomannan and xylan (Table 2). These data suggest that Trp-212 contributes to substrate binding but, in contrast to *Ct*CE2, oligosaccharide product departure at the completion of k_2 is not the rate-limiting step, and thus mutation of the tryptophan does not lead to an increase in k_{cat}.

In addition to the catalytic α/β-hydrolase domain, the CE2 members also contain an all β-sheet domain, somewhat

Table 4. ITC Analysis of *Ct*CE2 Wild-Type and Mutants Binding to Cellooligosaccharides and β-Glucan

Ligand	Mutant	$K_A \times 10^4$ (M^{-1})	K_D (μM)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	TΔS (kcal mol^{-1})	n^a
Cellohexaose		3.0 ± 0.3	33	−6.1 ± 0.0	−9.1 ± 0.6	−3.0 ± 0.6	0.8 ± 0.1
Cellopentaose		1.4 ± 0.1	71	−5.6 ± 0.0	−8.5 ± 0.4	−2.9 ± 0.4	0.7 ± 0.0
Cellotetraose		0.3 ± 0.0	333	−4.8 ± 0.0	—^{b}	—	—
Cellotriose		<0.1	>1,000	—	—	—	—
β-glucan		3.2 ± 0.1	31	−6.1 ± 0.0	−7.6 ± 0.1	−1.5 ± 0.1	1.0 ± 0.0
*Ct*CE2 mutants^c	S612A	24.4 ± 1.8	4.1	−7.3 ± 0.0	−10.5 ± 0.1	−3.2 ± 0.1	0.8 ± 0.05
	H791A	0.15 ± 0.1	667	−4.3 ± 0.0	—	—	—
	Y683A	2.7 ± 0.1	37	−6.0 ± 0.2	−8.7 ± 0.4	−2.7 ± 0.6	0.9 ± 0.1

^{a}Number of binding sites on the protein.
^{b}— indicates the affinity is too low to calculate accurate thermodynamic parameters.
^{c}Values are with cellohexaose as ligand.
doi:10.1371/journal.pbio.1000071.t004

reminiscent of a CBM. In *Ct*CE2 this jelly-roll domain appears to extend the substrate/cellulose binding cleft of the catalytic domain. To address this issue Tyr-683 and Trp-127, which are located in or at the interface, of the CBM-like domains of *Ct*CE2 and *Cj*CE2A, respectively, were mutated and the activities of the two esterase variants were assessed. The data, reported in Table 2, show that the W127A *Cj*CE2A mutation caused a considerable increase in K_M but not k_{cat} for glucomannan and xylan, while the catalytic properties of the *Ct*CE2 mutant Y683A were similar to the wild-type enzyme. It would appear, therefore, that the CBM-like domain contributes to substrate recognition in *Cj*CE2A, but its role in the catalytic activity of *Ct*CE2 is less evident.

Structural Characterisation of Cellooligosaccharide Binding to *Ct*CE2

The structure of the α/β-hydrolase domain of *Ct*CE2 revealed a deep cleft. One wall of the cleft is formed by the loops derived from β2, β3, and β4, while the other face contains the extended loop that links β5 to α5. The structure of both wild-type *Ct*CE2 and the S612A mutant were determined in the presence of cellohexaose and clear electron density for five β-1,4-linked D-glucose molecules was observed in the cleft (Figure 3), consistent with the affinities revealed by ITC (Table 4). The interaction of the oligosaccharide with *Ct*CE2 is dominated by planar hydrophobic contacts between the pyranose rings of Glc-1 with Trp-746, Glc-2 and Trp-790, and Glc-4 and Tyr-665 (Figure 3). Hydrophobic interactions represent the major mechanism by which most sugars are recognised by proteins (see for example [20–22]). Although the two structurally characterised *Cellvibrio* esterases do not bind cellohexaose or cellulose, they do contain aromatic residues in the cleft that houses the active site. Thus, Tyr-665 in *Ct*CE2 is conserved in the *C. japonicus* esterases (Trp-231 in *Cj*CE2A and Tyr-206 in *Cj*CE2B), while Trp-335 in *Cj*CE2B is equivalent to Trp-790 in the *Clostridium* enzyme (Figure S1). The importance of all three hydrophobic interactions in the binding of *Ct*CE2 to cellulose was demonstrated by the observation that the mutants W746A, W790A, and Y665A displayed no, or extremely weak, affinity for cellohexaose or β-glucan (Figure 5; Table 3). Importantly, Trp-746 in *Ct*CE2 is not conserved within CE2 members (Figure S1) (only one other CE2 member appears to contain an aromatic residue at the equivalent position), including the *Cellvibrio* enzymes,

suggesting that the insertion of this residue into *Ct*CE2 contributes considerably to cellulose recognition. It should be emphasised, however, that the sequence and conformation of the loop in *Ct*CE2, which contains the critical aromatic residue Trp-746, is different in the other CE2 enzymes. Thus, cellulose recognition is not caused by the simple insertion of a tryptophan into the substrate binding site of the *Clostridium* esterase, but is also influenced by the context of the introduced aromatic residue.

In *Ct*CE2, interaction of cellohexaose with the hydrophobic platform is augmented by several polar contacts (Figure 3). Notably, the O6 of Glc-4 makes a hydrogen bond with Oγ of Ser-612 (in one of its two conformations) in the wild-type esterase and Nε2 of His-791 in the S612A mutant, while O6 of Glc-2 interacts with Oε1 of Gln-780. As discussed above, mutation of His-791 to alanine reduces affinity for cellohexaose ~20-fold but, intriguingly, the S612A mutant binds to the oligosaccharide ~8-fold more tightly than the native protein (Tables 3 and 4). In the wild type enzyme Ser-612 adopts two conformations. In one conformation the hydroxyl makes a polar contact with cellohexaose forcing the ligand away from the histidine. In the S612A mutant the hexasaccharide adopts a different conformation in which it is now able to make a, presumably more productive, hydrogen bond with the histidine. Indeed, the bond distance between Ser-612 and Glc-4 is 3.3 Å (the two conformations adopted by the serine suggests a relatively weak interaction with the glucose), while the bond distance between His-791 and Glc-4 is reduced to 2.8 Å, pointing to a stronger interaction.

It is possible that the CBM-like domain, which appears to abut onto the substrate binding cleft in *Ct*CE2, may also contribute to cellulose recognition. The observation that the Y683A mutant (Tyr-683 is the only aromatic residue in the cleft anterior to the catalytic apparatus) of the esterase retains its capacity to bind cellohexaose (Table 4) and cellulose (unpublished data), however, indicates that the recognition of this polysaccharide is mediated solely by the α/β-hydrolase domain.

The Mechanism of Cellohexaose Inhibition of *Ct*CE2 Activity

Cellohexaose inhibits the deacetylation of glucomannan by *Ct*CE2 with a K_i of 32 μM (Figure 6A), similar to the K_D of the esterase for the hexasaccharide determined by ITC (Table 4).
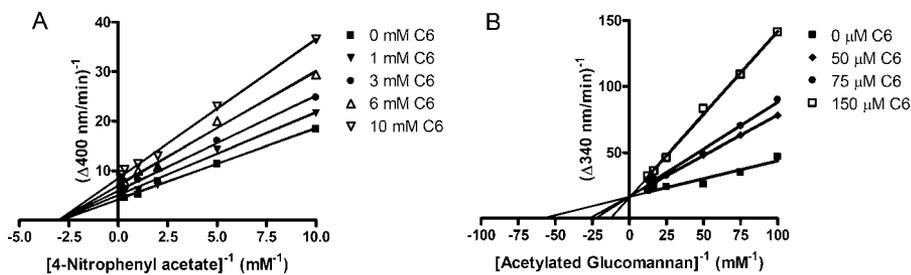
Figure 6. Inhibition of Wild-Type *Ct*CE2 by Cellohexaose
*Ct*CE2 was assayed at 37 °C using either 4-nitrophenyl acetate (A) or acetylated glucomannan (B) as the substrate in the presence of different concentrations of cellohexaose (C6). The figure displays double reciprocal plots of the data.
doi:10.1371/journal.pbio.1000071.g006

Consistent with its capacity to interact with the catalytic apparatus and with aromatic residues in the substrate binding cleft, cellohexaose is a competitive inhibitor of glucomannan, as indicated by the double reciprocal plots which intersect on the *y*-axis (Figure 6A). In contrast, cellohexaose inhibits the hydrolysis of 4-NPAc by *Ct*CE2 with a K_i of 3.1 µM, which is ~10-fold lower than the ITC-determined K_D of the enzyme for the ligand. In addition, the inhibition by cellohexaose displayed noncompetitive kinetics (Figure 6B: double reciprocal plots intersect on the *x*-axis), suggesting that the ligand inhibits the deacetylation of glucomannan and the aryl-acetate by different mechanisms.

*Ct*CE2, typical of enzymes that mediate catalysis through a covalent intermediate, displays biphasic kinetics against substrates, such as 4-NPAc, which contain good leaving groups. Pre-steady state kinetics, deploying 4-NPAc as the substrate, revealed a k_2 (pre-steady state rate of 4-nitophenolate release) value of ~21,790 s^{-1} with an amplitude that was ~80% of the predicted value, while k_3, extracted from the steady state rate, was ~4,725 s^{-1} (unpublished data). Significantly, cellohexaose inhibited enzyme deacylation (the steady state rate; k_3), with a K_i of 7 µM, but did not significantly affect enzyme acylation (the pre–steady state burst of 4-nitrophenolate release; k_2).

The data described above are consistent with the observed noncompetitive kinetics displayed by cellohexaose when 4-NPAc is the substrate, and the difference between K_i and K_D. ITC measures binding to the apoenzyme (available to the substrate at k_2) while K_i determines affinity of the oligosaccharide for the enzyme at the rate-limiting step of the reaction, k_3, when the enzyme is in complex with acetate. Indeed, the kinetics of inhibition are consistent with the observation that cellohexaose binds more tightly to the enzyme when it makes a polar contact with His-791 rather than Ser-612; when the serine is in complex with acetate (or formate in the crystal structure) it is not available to hydrogen bond with the hexasaccharide that is now able to interact with His-791. By contrast, k_2 is the rate-limiting step when glucomannan is the substrate and thus cellohexaose must now bind to the apo form of the enzyme to inhibit the reaction leading to a lower affinity (as it will make a polar contact with Ser-612 and not His-791) and competitive kinetics.

Discussion

Family CE2 esterases raise intriguing questions about the optimization of plant cell-wall–degrading systems. The most interesting feature of the divergence of function within the CE2 family is that *Ct*CE2, derived from the *Ct*Cel5C-CE2 multimodular enzyme, is uniquely able to act as a CBM, with structural and biochemical studies highlighting Trp-746, a residue observed in only one other CE2 enzyme, as a key recognition element. It is probably highly relevant that *Ct*CE2 is the only member of CE2 that is a component of a modular enzyme. Thus, *Ct*CE2 is an example of an emerging biochemical theme that protein scaffolds possess a latent potential to acquire new, orthogonal functionality [23,24].

This report is in contrast to the Ohno model of protein evolution, which proposes that mutations that provide new functionalities are introduced into a redundant copy of a duplicated gene [25,26]. A central component of the Ohno model is that the introduction of new functions into a protein compromises its original activity, hence the requirement for gene duplication subsequent to mutation. The observation that the chicken structural protein δ-crystallin and the metabolic enzyme, argininosuccinate lyase, are encoded by the same gene led to the "gene sharing" hypothesis, which states that a gene can be recruited for a novel function without significant changes to its protein coding sequence [27]. Bergthorsson et al. [28] and others argue that Ohno's model poses a dilemma, because it requires that the duplicated copy of the gene, which incurs an energetic cost, must be retained within a population prior to acquiring the mutations that confer the new activity. These concerns led to the proposal that gene products display secondary activities prior to duplication of the encoding gene, and that there is a selection for both activities subsequent to amplification (these models are reviewed in [29]). A major caveat with models based on "gene sharing," however, is that amino acid changes that introduce new functions often compromise the stability or original activity of the protein [30]. Thus, in general, the introduced functionality replaces the endogenous activity of the enzyme. Here we show that the active site of *Ct*CE2 displays dual activities; it catalyses the deacetylation of plant polysaccharides and also potentiates the activity of its appended cellulase catalytic module through its non-catalytic cellulose binding function. As such, *Ct*CE2 provides an example of gene sharing in which the new function (cellulose binding) has developed without having an obvious deleterious effect on the existing esterase activity of the enzyme. It would appear, therefore, that there are different ways to evolve new enzyme functions without compromising the original activity displayed by the protein. Which of these pathways a protein follows depends on (i) what the new

function is, (ii) whether it is advantageous to retain both activities, and (iii) whether the chemistry or steric constraints of the two divergent functions preclude the housing of these activities in the same active site.

Throughout plant cell-wall–degrading systems, noncatalytic binding to polysaccharides by enzymes that attack insoluble substrates is primarily conferred by CBMs that are linked to, but spatially distinct from, the catalytic module [19]. In general these targeting modules are joined by flexible linker sequences, although noncatalytic carbohydrate binding regions, while spatially distinct from the active site, can occasionally be distal components of the catalytic module itself [31,32]. This report shows that the presentation of an additional tryptophan residue within the substrate binding cleft of the esterase enables the enzyme to acquire cellulose-binding capacity—potentiating the activity of the appended cellulase catalytic module—whilst still retaining its original catalytic activity. Indeed, the importance of cellulose recognition by *Ct*CE2 in the function of the appended catalytic module *Ct*Cel5C is supported by the observation that mutation of Trp-790, Tyr-665, or Trp-746 in the full-length enzyme (*Ct*Cel5C-CE2) reduces the activity of the cellulase against insoluble substrates by 3- to 5-fold. Hence the CE2 family displays a spectrum of activities reflecting the grafting of new functionality upon the α/β framework. This manifests itself most powerfully in the acquisition of cellulose binding by *Ct*CE2, demonstrating how nature has exploited latent carbohydrate recognition to introduce additional "noncatalytic" polysaccharide binding features that are complementary to the activities displayed by a complex modular enzyme. The composite structure of the plant cell wall, which requires the synergistic interactions of multiple catalytic and binding functions to elicit its degradation, exerts selection pressures that lead to the generation of single proteins with multiple activities. The CE2 family, and *Ct*CE2 in particular, reveal an extreme example of this functional complexity and provides a platform for the future directed engineering of plant cell-wall–degrading enzyme systems, which is one of the key environmental goals of the twenty-first century.

Materials and Methods

Gene cloning and protein expression. The region of the *C. thermocellum* cellulase–esterase gene, *cel5C-ce2*, encoding the C-terminal CE2 esterase module, *Ct*CE2 (residues 482 to 814 of the full-length enzyme), was amplified by PCR from genomic DNA (strain ATCC 27405) using the thermostable DNA polymerase Pfu Turbo (Stratagene) and primers (Table S1) that contain NdeI and XhoI restriction sites, respectively. The DNA product was cloned into the NdeI and XhoI sites of the *Escherichia coli* expression vector pET22b (Novagen) to generate pCtCE2. *Ct*CE2 encoded by pCtCE2 contains a C-terminal His$_6$ tag. The region of the genes encoding the mature *C. japonicus* and *B. thetaiotaomicron* VPI-5482 CE2 esterases were amplified by PCR from genomic DNA, using primers listed in Table S1, and cloned into NdeI- and XhoI-restricted pET22b such that the proteins encoded by the recombinant plasmids contained a C-terminal His$_6$ tag.

Protein expression and purification. *E. coli* BL21 cells harbouring the CE2 esterase-encoding recombinant expression vectors were cultured in Luria-Bertani broth at 37 °C to mid-exponential phase (A_{600nm} ~0.6–1.0), and recombinant protein expression was induced by the addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) and incubation for a further 5 h at 37 °C. The CE2 esterases were purified by immobilized metal ion affinity chromatography (IMAC) using Talon resin (Clontech) and elution in 20 mM Tris/HCl buffer containing 300 mM NaCl and 100 mM imidazole. The eluted esterase was then dialyzed against 20 mM Tris/HCl buffer, pH 8.0 (Buffer A) and applied to a Bio-Rad Q12 column. The esterases were eluted with a 400 ml gradient of 0–500 mM NaCl in Buffer A. The

fractions containing esterase activity were concentrated using a 10 kDa MWCO Vivaspin 20 centrifugal concentrator and applied to a Superdex 200 26/60 Hiload column (Amersham) equilibrated in 10 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl. Protein was eluted at a flow rate of 1 ml/min. Both chromatography steps employed a Bio-Rad FPLC system. Purified enzymes were adjudged homogenous by SDS-PAGE. To produce seleno-methionine–containing proteins the same protocol was employed except that the enzyme was expressed in *E. coli* B834 (Novagen) using growth conditions described previously [33], and 2 mM β-mercaptoethanol was included in all buffers during purification up to the point of IMAC and 10 mM for all subsequent steps. The purified seleno-methionine enzyme eluted from the gel filtration column was exchanged into ddH$_2$O containing 10 mM DTT using Amersham PD10 column.

Mutagenesis. Site-directed mutagenesis was carried out using the PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions, using pCtCE2 or pCjCE2A as the template and primers pairs that are listed in Table S1.

Enzyme assays. Substrates used in the enzyme assays described below were purchased from Sigma except acetylated Konjac glucomannan, which was purchased from Megazyme International (Bray, Ireland); acetylated xylan, which was prepared from birchwood xylan (Sigma) using acetic anhydride following the method of Johnson et al. [34], and O-acetyl galactoglucomannan that was extracted from spruce (*Picea abies*) [35]. To determine activity against 4-NPAc, 1 ml enzyme reactions were carried out in 50 mM sodium phosphate buffer, pH 7.0, containing 1 mg/ml of BSA and substrate concentrations up to 10 mM. The reaction was initiated by the addition of an appropriate concentration of enzyme; 10 nM for wild type and 10 μM for the most inactive mutants, and the release of 4-nitrophenolate was monitored at 400 nm. To determine the rate of deacetylation of acetylated polysaccharides, the release of acetate was determined using an acetic acid detection kit (Megazyme International) following the manufacturer's recommendation except that the product was measured continuously rather than in a stopped reaction. Pre-steady-state kinetics were performed using a stopped-flow apparatus (Applied Photophysics Model SX-17MV), with the flow path thermostatically controlled at 15 °C, and a 2-mm light path to monitor the formation of the product 4-nitrophenolate at A_{400}. Equal volumes (50 μl) of solutions mixed with a dead time of 1.5 ms, contained, respectively, 10 μM *Ct*CE2 and 2 mM 4-NPAc in 50 mM sodium phosphate buffer, pH 7.0. These solutions were supplemented with 0–100 μM cellohexaose.

Carbohydrate binding studies. The capacity of the CE2 enzymes to bind to soluble saccharides was determined by AGE or by ITC. AGE was performed as described previously [33] with the polysaccharide ligand included in the polyacrylamide gel at 0.1 %. ITC [36] was carried out in 50 mM HEPES-Na buffer, pH 8.0, containing 300 mM NaCl at 25 °C. Data collected for titrations with 10 mg/ml of β-glucan were fitted with a molar concentration of 5 mM, the value at which $n = 1$. The concentrations of cellooligosaccharides in the syringe ranged from 0.7 to 5 mM and the CE2 enzymes in the reaction cell were at 80–100 μM. The determined K_A and ΔH values were used to calculate ΔS from the standard thermodynamic equation. Binding to insoluble cellulose was determined by pull-down assay using phosphoric acid swollen cellulose (PASC). Briefly, 1 mg of washed PASC in 50 mM Tris-HCl, pH 8.0, containing 300 mM NaCl (Buffer B), was mixed with 50 μg of protein in a total volume of 100 μl and incubated on ice for 30 min. This was then centrifuged for 1 min at 13,000g and the supernatant containing the unbound protein removed. The cellulose was then washed three times with 100 μl of ice-cold Buffer B and each wash discarded, before 50 μl of SDS-PAGE loading buffer (containing 10 % SDS) was added and the bound protein eluted from the cellulose by boiling for 5 min. Approximately equal volumes of starting protein and material eluted from the cellulose were then subjected to SDS-PAGE to assess binding.

Crystallisation and structure solution. *Cj*CE2A: Native and seleno-methionine crystals of *Cj*CE2A were grown by hanging drop vapour-phase diffusion in 1 M diammonium hydrogen phosphate/0.1 M sodium acetate, pH 5.0, for 1–2 d at 20 °C (30 % v/v glycerol was added as the cryoprotectant). Data from seleno-methionine derivatised *Cj*CE2A were collected on ID23-1 at a wavelength of 0.97880 Å using an ADSC Q315R CCD (charge-coupled device) detector the wavelength was optimised for the f'' component of the anomalous signal using a fluorescence scan. Data from crystals of native *Cj*CE2A were collected on ID14.2 using an ADSC Q4 CCD detector at a wavelength of 0.9330 Å. Data for all CE2 enzymes were processed with either DENZO [37] or MOSFLM from the CCP4 suite (Collaborative Computational Project Number 4 1994). The structure of *Cj*CE2A was

Table 5. Data Collection and Refinement Statistics

Data Collection and Refinement	Statistic	Subcategory	*Ct*CE2-Cellohexaose	S612A[a]-Cellohexaose	*Cj*CE2A	*Cj*CE2B
Data	**Space group**		P2$_1$	P2$_1$2$_1$2	P4$_1$2$_1$2	P4$_3$
	Cell dimensions	*a, b, c* (Å)	41.6, 60.1, 69.0	41.2, 141.2, 58.2	132.2, 132.2, 45.8	75.3, 75.3, 141.6
		α, β, γ (°)	90, 78.01, 90	90,90,90	90,90,90	90,90,90
	Resolution (Å)		67.60–1.80 (1.90–1.80)	70.53–1.90 (2.00–1.90)	46.73–1.80 (1.90–1.80)	53.23 (2.10–2.00)
	R_{merge}		0.08 (0.30)	0.10 (0.35)	0.12 (0.58)	06 (0.62)
	$<I/\sigma I>$		12.7 (3.9)	10.8 (3.9)	15.6 (3.3)	15.5 (3.0)
	Completeness (%)		99.3 (99.3)	99.8 (99.8)	100 (100)	100 (100)
	Redundancy		4.0 (4.1)	4.7 (4.8)	8.0 (8.1)	5.5 (5.4)
Refinement	**Resolution (Å)**		1.80	1.90	1.80	2.0
	No. reflections		31,011	27,641	38,157	49,944
	R_{work}/R_{free}		0.14/0.19	0.15/0.20	0.15/0.20	0.19/0.23
	No. atoms	Protein	2,569	2,574	2,725	5,141
		Ligand/ion	56	114	30	42
		Water	430	375	445	420
	***B*-factors**	Protein main chain/side chain	8/11	13/15	11/13	41/41
		Ligand/ion	25	24	30	51
		Water	25	27	28	42
	Rms deviations	Bond lengths (Å)	0.019	0.018	0.019	0.012
		Bond angles (°)	1.7	1.6	1.7	1.3
PDB Codes			2wao	2wab	2waa	2w9x

[a] The S619A mutant is a derivative of *Ct*CE2.
doi:10.1371/journal.pbio.1000071.t005

solved using the single-wavelength anomalous dispersion (SAD) method. Selenium positions were determined using SHELXD [39] and the phases were subsequently calculated using SHELXE. 5% of the data were set aside for cross validation analysis and the behaviour of R$_{free}$ was used to monitor and guide the refinement protocols. ARP/wARP [40] in conjunction with REFMAC [41] was used to automatically build the sequence into the electron density. Refinement was undertaken in REFMAC with manual correction to the model using COOT [42].

*Ct*CE2: Crystals of both wild-type and S612A *Ct*CE2 in complex with cellohexaose were grown by hanging drop vapour-phase diffusion from 20% PEG3350 and 0.2–0.3 M ammonium iodide, with protein at ~8 mg ml^{-1} and cellohexaose at approximately 1 mM. Crystals were harvested in rayon fibre loops before being bathed in cryoprotectant solution (crystallisation conditions augmented with 25% v/v glycerol) and flash frozen in liquid nitrogen. Data were collected at the ESRF from single crystals at 100 K for cellohexaose complexes of both wild-type and S612A mutant forms with data processed as previously. The structure of *Ct*CE2 was solved by molecular replacement using PHASER with the search model being the *Cj*CE2A structure. ARP/wARP was used to build the initial model and refinement was undertaken as described above.

*Cj*CE2B: Native and seleno-methionine crystals of *Cj*CE2B were grown by hanging drop vapour diffusion in 0.1 M imidazole, pH 8.0, 10 % PEG 8,000, for 3 d at 20 °C (25 % v/v glycerol was added as the cryoprotectant). Native data were collected at ESRF on ID14.2 using ADSC Q4 CCD detector at a wave length of 0.933 Å. The structure was solved using a combined approach in which molecular replacement using a combined *Ct*CE2/*Cj*CE2A model was used in PHASER [43] was used in conjunction with the Se-Met derived SAD phases for map calculation and verification of methionine positions. The final model was build using COOT and refined using REFMAC. *Cj*CE2B crystals contains two molecules in the asymmetric unit. The final models of *Cj*CE2B in chain A was ordered from 27 to 361 containing five missing regions in amino acids 41–44, 61–63, 81–84, 98–98, and 127–137, while chain B was ordered from 26 to 361 and contained one missing region between residues 99 and 107. Coordinates and

observed structure factor amplitudes have been deposited at the Worldwide Protein Data Bank (wwPDB, http://www.wwpdb.org/), and the crystal structure statistics are in Table 5.

Supporting Information

Figure S1. Alignment of CE2 Enzymes

The catalytic histidine and serine are in red, the residue equivalent to the aspartate that completes the catalytic triad is green, while the equivalent residues to the three aromatic amino acids in *Ct*CE2 that binds to cellulose are blue. The vertical arrow marks the division between the β-jelly roll fold and the α/β-hydrolase catalytic domain. The residues are labelled according to the *Ct*CE2 sequence.

Found at doi:10.1371/journal.pbio.1000071.sg001 (53 KB DOC).

Table S1. Primers Used to Mutate Residues in *Ct*CE2 or Clone Other CE2 Enzymes

Found at doi:10.1371/journal.pbio.1000071.st001 (35 KB DOC).

Acknowledgments

References

1. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, et al. (2006) The path forward for biofuels and biomaterials. Science 311: 484–489.

2. Sticklen M (2006) Plant genetic engineering to improve biomass characteristics for biofuels. Curr Opin Biotechnol 17: 315–319.

3. Brett CT, Waldren K (1996) Physiology and Biochemistry of Plant Cell

Walls. Topics in Plant Functional Biology; Black M, Charlewood B, editors. London: Chapman and Hall.

4. Tomme P, Warren RA, Gilkes NR (1995) Cellulose hydrolysis by bacteria and fungi. Adv Microb Physiol 37: 1–81.

5. Warren RA (1996) Microbial hydrolysis of polysaccharides. Annu Rev Microbiol 50: 183–212.

6. Hall J, Black GW, Ferreira LM, Millward-Sadler SJ, Ali BR, et al. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. Biochem J 309: 749–756.

7. Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In: Gilbert HJ, Davies GJ, Henrissat B, Svensson B, editors. Recent Advances in Carbohydrate Bioengineering". Cambridge: The Royal Society of Chemistry. pp. 3–12.

8. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280: 309–316.

9. Dalrymple BP, Cybinski DH, Layton I, McSweeney CS, Xue GP, et al. (1997) Three *Neocallimastix patriciarum* esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. Microbiology 143: 2605–2614.

10. Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58: 521–554.

11. Hall J, Hazlewood GP, Barker PJ, Gilbert HJ (1988) Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. Gene 69: 29–38.

12. Durrant AJ, Hall J, Hazlewood GP, Gilbert HJ (1991) The non-catalytic C-terminal region of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain. Biochem J 273: 289–293.

13. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, et al. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science 299: 2074–2076.

14. DeBoy RT, Mongodin EM, Fouts DE, Tailford LE, Khouri H, et al. (2008) Insights into Plant Cell-wall Degradation from the Genome Sequence of the Soil Bacterium *Cellvibrio japonicus*. J Bacteriol: in the press.

15. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF (2004) GDSL family of serine esterases/lipases. Prog Lipid Res 43: 534–552.

16. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D60: 2256–2268.

17. Larsen NA, Lin H, Wei R, Fischbach MA, Walsh CT (2006) Structural characterization of enterobactin hydrolase IroE. Biochemistry 45: 10184–10190.

18. Wharton CW (1997) The Serine Proteinases. In: Sinnott ML, editor. Comprehensive Biological Catalysis. London: Academic Press. pp. 345–379.

19. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382: 769–781.

20. Nagy T, Simpson P, Williamson MP, Hazlewood GP, Gilbert HJ, et al. (1998) All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands. FEBS Lett 429: 312–316.

21. Ponyi T, Szabo L, Nagy T, Orosz L, Simpson PJ, et al. (2000) Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas* xylanase A to insoluble ligands. Biochemistry 39: 985–991.

22. Flint J, Bolam DN, Nurizzo D, Taylor EJ, Williamson MP, et al. (2005) Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. J Biol Chem 280: 23718–23726.

23. O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6: R91–R105.

24. Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, et al. (2005) The 'evolvability' of promiscuous protein functions. Nat Genet 37: 73–76.

25. Ohno S (1970) Evolution by gene duplication. New York: Springer.

26. Kimura M, Ota T (1974) On some principles governing molecular evolution. Proc Natl Acad Sci U S A 71: 2848–2852.

27. Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, et al. (1988) Gene sharing by delta-crystallin and argininosuccinate lyase. Proc Natl Acad Sci U S A 85: 3479–3483.

28. Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: evolution of new genes under continuous selection. Proc Natl Acad Sci U S A 104: 17004–17009.

29. Bershtein S, Tawfik DS (2008) Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol Biol Evol 25: 2311–2318.

30. Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol 320: 85–95.

31. Robert X, Haser R, Gottschalk TE, Ratajczak F, Driguez H, et al. (2003) The structure of barley alpha-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: a pair of sugar tongs. Structure 11: 973–984.

32. Allouch J, Helbert W, Henrissat B, Czjzek M (2004) Parallel substrate binding sites in a beta-agarase suggest a novel mode of action on double-helical agarose. Structure 12: 623–632.

33. Charnock SJ, Bolam DN, Turkenburg JP, Gilbert HJ, Ferreira LM, et al. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. Biochemistry 39: 5013–5021.

34. Johnson KG, Fontana JD, MacKenzie CR (1988) Measurement of acetylxylan esterase in *Streptomyces*. Methods Enzymol 160: 551–560.

35. Lundqvist J, Teleman A, Junel L, Zacchi G, Dahlman O, et al. (2002) Isolation and characterization of galactomannan from spruce (*Picea abies*). Carbohydr Poly 48: 29–39.

36. Bolam DN, Xie H, Pell G, Hogg D, Galbraith G, et al. (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. J Biol Chem 279: 22953–22963.

37. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276: 307–326.

38. Collaborative Computational Project Number 4 (1994) Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D50: 760–763.

39. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallographica Section D-Biological Crystallography 58: 1772–1779.

40. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. Nature Structural Biology 6: 458–463.

41. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum likelihood method. Acta Cryst D 53: 240–255.

42. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D60: 2126–2132.

43. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. Journal of Applied Crystallography 40: 658–674.