Biochemistry (Moscow), Vol. 67, No. 8, 2002, pp. 850-871. Translated from Biokhimiya, Vol. 67, No. 8, 2002, pp. 1026-1050.

REVIEW

The Structure and Mechanism of Action of Cellulolytic Enzymes

M. L. Rabinovich*, M. S. Melnick, and A. V. Bolobova

Bach Institute of Biochemistry, Russian Academy of Sciences, Leninsky pr. 33, Moscow, 119071 Russia; fax: (095) 954-2732; E-mail: mrabinovich@inbi.ras.ru

Received January 24, 2002
Revision received February 20, 2002

Abstract—The modern structural classification of polysaccharases comprising cellulase–hemicellulase enzyme systems is discussed. Their catalytic domains are currently grouped into 15 of more than 80 known glycosyl hydrolase families, whereas substrate binding domains fall into 13 families. The structures of catalytic and substrate binding domains, as well as linker sequences, are briefly considered. A hypothetical mechanism of concerted action of catalytic and substrate binding domains of cellobiohydrolases on the surface of highly ordered cellulose is suggested.

Key words: glycosyl hydrolase families, cellulases, xylanases, endoglucanases, cellobiohydrolases, catalytic domains, cellulose binding domains

In the past decade a clear trend of revision of the currently used glycosyl hydrolase nomenclature was revealed. Besides substrate specificity, modern classifications increasingly involve consideration of structural characteristics of the enzyme molecule. This trend has a quite certain basis. First, the central paradigm of the conventional nomenclature (i.e., absolute or narrow specificity towards the type of glycosidic bond) was seriously revised. Indeed, many examples of enzymes were described whose active sites displayed several activities, e.g., a number of cellulases–xylanases. Second, molecular structures of the enzymes of the same specificity revealed a wide diversity, being in many instances more distant from each other than from glycanases of rather different specificity. And, finally, modular chimeric enzymes were described that bear two or more catalytic domains of different specificity on C- and N-termini of the same polypeptide chain [1, 2].

According to the modern concepts [3, 4], most of the cellulolytic enzymes comprise modular multidomain proteins containing at least three separate structural elements of different functions, i.e., catalytic domain (CD), cellulose binding domain (CBD), and interdomain linker. Also, other structural elements can also be inserted into the cellulase structures, i.e., a second catalytic domain, fibronectin 3-like modules, repeated hydrophobic sequences of the so-called cohesin–dockerin type, etc. (Fig. 1). Here we shall consider only CDs, CBDs, and (very briefly) interdomain linkers, because the importance of these structural elements for cellulase action on the surface of cellulose is the most obvious.

As seen from Table 1, cellulolytic enzymes are grouped in at least 15 of more than 80 known structural families of glycosyl hydrolases. Apart from true cellulases, two families of a variety of xylanases (10/F and 11/G) were also included in the cellulase classification. Indeed, family 10/F represent xylanases of a broad specificity and this may be the reason of its formal assignment to cellulases. However, enzymes of family 11/G do not meet this criterion, because they are strictly specific towards xylan [5]. On the other hand, families 1 and 3 that include β-glucosyl hydrolases capable of splitting short-chain cellooligosaccharides, or families 16 and 17 that include 1,3(4)-β-glucanases are not considered as the members of cellulase classification, although these enzymes reveal similar specificity towards 1,4-β-glucosidic bonds and share tertiary structure of typical cellulases (particularly members of families 1, 16, and 17).

These examples demonstrate an uncertainty in the current definition of true cellulase families. If the type of attacked glycosidic bond were accepted as the major criterion, then all hydrolases capable of splitting cellooligosaccharides (EC 3.2.1.4; 21; 74; 91) should be included, whereas the enzymes strictly specific towards other types of bonds should be excluded. If a stricter criterion (decomposition of highly ordered forms of insoluble cellulose) were applied, then cellulase classification would be

Abbreviations: CD) catalytic domain; PBP) polysaccharide binding protein; CBD) cellulose binding domain; TIM) triosephosphate isomerase; CBH) cellobiohydrolase; GH) glycosyl hydrolase.

* To whom correspondence should be addressed.

0006-2979/02/6708-0850$27.00 ©2002 MAIK "Nauka/Interperiodica"

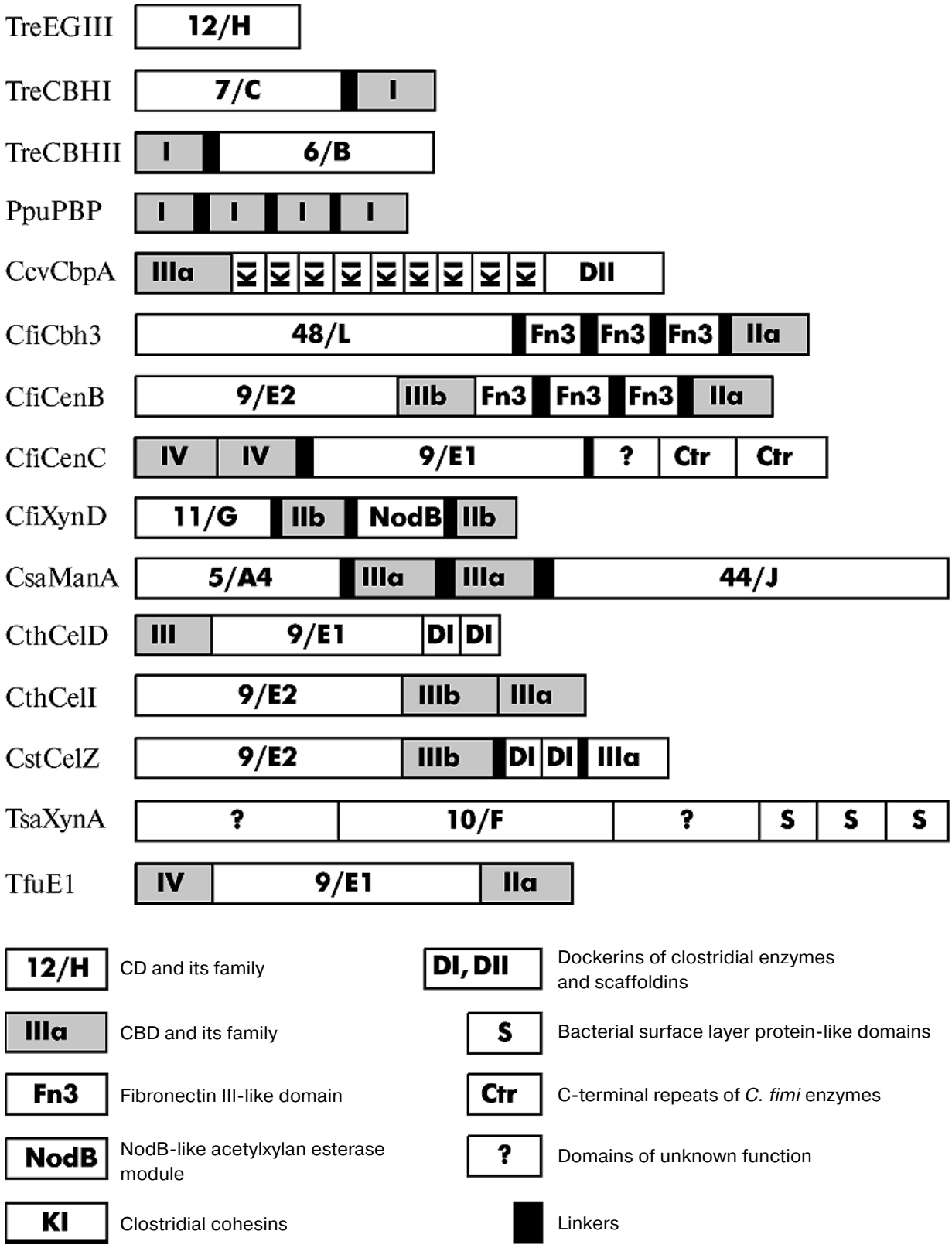

Fig. 1. Modular structures of cellulases and related enzymes (according to [4, 52]). Abbreviations: Tre, *Trichoderma reesei*; Ppu, *Porphyra purpurea*; Ccv, *Clostridium cellulovorans*; Cfi, *Cellulomonas fimi*; Csa, *Caldicellulosiruptor saccharolyticum*; Cth, *Clostridium thermocellum*; Cst, *Clostridium stercorarium*; Tsa, *Thermoanaerobacter saccharophilum*; Tfu, *Thermomonospora fusca*; PBP, polysaccharide binding protein; CbpA, cellulosome binding protein (scaffoldin A of mesophilic *Clostridium*).

Table 1. Systematic positions of the enzymes capable of splitting 1,4-β-glucosidic bonds in the general structural classification of glucosyl hydrolases (according to [4, 6-9])

Family	Type and origin of the enzyme	EC	Clan	Fold	Type of catalysis
1	2	3	4	5	6
5/A1	fungal endo-1,4-β-mannanases; endoglucanases of aerobic and anaerobic bacteria	3.2.1.78 3.2.1.4	GH-A	TIM-barrel = (α/β) 8-barrel = (β/α)8-barrel; active site includes nucleophilic glutamate at the C-terminus of β-strand 7 and general acid-base asparagine-glutamate at C-terminus of β-strand 4	retain configuration of the anomeric carbon in the reaction products
5/A2	endoglucanases of actinomycetes and aerobic and anaerobic bacteria, as well as animals (nematodes)	3.2.1.4	same	same	same
5/A3	exo-1,3-β-glucanases; endoglucanases/1,3-1,4-β-glucanases; 1,3-β-glucanases (yeast, *Clostridium*)	3.2.1.58 (cryst.) 3.2.1.4/73 3.2.1.39	same	same; insertion of an additional helical subdomain near active site	same
5/A4	endoglucanases and mannanases of actinomycetes, aerobic and anaerobic bacteria, and anaerobic fungi	3.2.1.4 3.2.1.78	same	same	same
5/A5	endoglucanases of filamentous fungi and aerobic bacteria	3.2.1.4	same	same	same
5	endo-1,6-β-glucanases ?	3.2.1.75 3.2.1.123	same	same	same
6/B	endoglucanases and cellobiohydrolases of aerobic bacteria, actinomycetes, and anaerobic and filamentous fungi	3.2.1.91 (cryst.) 3.2.1.4 (cryst.)		incomplete TIM-barrel	inverse configuration
7/C	cellobiohydrolases and endoglucanases of filamentous fungi	3.2.1.91(cryst.) 3.2.1.4	GH-B	concanavalin A-type β-sandwich	retain configuration
8/D	1,3-1,4-β-glucanases and endoglucanases of aerobic and anaerobic bacteria	3.2.1.73 3.2.1.4 (cryst.)	not defined	(α,α)6-barrel	inverse
9/E1	endoglucanases of actinomycetes and aerobic and anaerobic bacteria; cellobiohydrolases of anaerobic bacteria	3.2.1.4 (cryst.) 3.2.1.91	same	same	same
9/E2	endoglucanases of aerobic and anaerobic bacteria, plants, and insects (termites); endo/exoglucanase of actinomycetes	3.2.1.4(cryst.) 3.2.1.4/91	same	same; additional strongly fixed type III CBD	same

Table 1. (Contd.)

1	2	3	4	5	6
10/F	xylanases and xylanases/endoglucanases of actinomycetes, aerobic bacteria, anaerobic archae- and eubacteria, and aerobic and anaerobic fungi; bacterial xylanase/exoglucanase	3.2.1.4/8 3.2.1.8/91 (cryst.) 3.2.1.8 (cryst.)	GH-A	TIM-barrel	retain
11/G	strictly specific xylanases of aerobic and anaerobic fungi, aerobic and anaerobic bacteria, and actinomycetes	3.2.1.8 (cryst.)	GH-C	"rib-cage"-like β-sandwich	same
12/H	endoglucanases of aerobic fungi, actinomycetes, aerobic bacteria, and anaerobic archaebacteria	3.2.1.4 (cryst.)	same	same; two β-sheets of six and nine strands, and α-helix across	same
26/I	endoglucanases; endoglucanases/xylanases; endomannanases of aerobic and anaerobic bacteria	3.2.1.78 3.2.1.4 3.2.1.4/8	GH-A	TIM-barrel	same
44/J	endoglucanases of aerobic and anaerobic bacteria	3.2.1.4	not defined	not defined	inverse
45/K1	endoglucanases of filamentous and anaerobic fungi and aerobic bacteria	3.2.1.4	not defined	6β-barrel	same
45/K2	endoglucanase of mollusks				
48/L	endoglucanases; cellobiohydrolases; endoglucanases/cellobiohydrolases of actinomycetes and aerobic and anaerobic bacteria	3.2.1.4 3.2.1.91 3.2.1.4/91	not defined	(α,α)6-barrel	same
51	α-L-arabinofuranosidases of ascomycetes and actinomycetes; endoglucanase of *F. succinogenes*	3.2.1.55 3.2.1.4	GH-A	TIM-barrel	retain
60	*Clostridium* endoglucanases	3.2.1.4	not defined	not defined	not defined
61	endoglucanases of asco- and basidiomycetes	3.2.1.4	not defined	not defined	same

Note: EC, enzyme classification according to IUB; (cryst.), proteins with resolved crystalline structure.

restricted by only cellobiohydrolases (EC 3.2.1.91) and some processive endoglucanases (EC 3.2.1.4).

It is obvious, however, that current "intuitive" separation of "true cellulase families" has a serious basis. In this respect we suggest that either the enzymes capable of splitting 1,4-β-glucosidic bonds in the high molecular weight celluloses, or enzymes of any other specificity, or even noncatalytic proteins capable of working on cellulose surface (i.e., all proteins containing CBDs) were considered in the context of cellulase classification [10].

This broad definition includes an array of proteins, which specifically recognize either isolated cellulose molecule or cellulose surface as a whole. This includes not only cellobiohydrolases responsible for decomposition of highly ordered structure of cellulose microfibril, but also endoglucanases, xylanases, mannanases, and other enzymes, which are directly or indirectly (as several glycosyl hydrolases, acetylxylan esterases, and feruloyl esterases) involved in the hydrolysis of amorphous polysaccharide matrix and its separation from the surface of cellulose microfibril, as well as cellulosomal noncatalytic proteins of anaerobic bacteria (scaffoldins) containing CBDs. Furthermore, any other proteins sharing common structures with typical cellulase CDs or CBDs along with

a variety of structural elements that bear cellulases and related proteins (Fig. 1) are also put on the list.

Hence, substrate specificity and mode of action are no longer the sole criteria for enzyme assignment. Moreover, since structural data started to accumulate, they have become a primary classification criterion. With such an approach to cellulase classification, a diversity of other proteins is also involved in the consideration, whose function may, in fact, be rather distant from that of cellulases. The value of this classification approach is in its predictive force. When a structural family of a new protein is identified, this provides a high level of probability in the prediction of its important properties and serves as a key for further site-directed mutagenesis. In addition, such classification allows for finding unexpected relationships between such objects and phenomena that seem rather distant at first look.

The enzymes discussed here are sometimes designated in the special literature by a general name of "cellulase–hemicellulase systems". This means a complete enzyme system secreted by an organism for total decomposition of plant cell wall [11]. This is the group of enzymes responsible for polysaccharide degradation in an absolute majority of species capable of consuming plant raw [12-15]. The only exceptions are brown rot decay basidiomycetes. Their cellulose-degrading systems apparently include non-enzymatic low molecular weight factors: peptide chelators of transient metal ions forming reactive oxygen species (OH-, peroxide-, or superoxide-radicals) [16].

CLASSIFICATION OF CATALYTIC DOMAINS OF THE ENZYMES OF CELLULASE–HEMICELLULASE SYSTEMS

Classification of cellulases and hemicellulases according to the structural features of their CDs was first introduced in the end of the 1980s and beginning of the 1990s [8, 17]. It was based on hydrophobic cluster analysis (HCA) and later spread to all glycosyl hydrolases [9]. According to HCA methodology, protein amino acid sequence is represented as two-dimensional longitudinal section of a cylinder formed by a hypothetical helical folding of polypeptide chain.

Current classification of cellulase–hemicellulase enzymes is partially reproduced in the Table 1. Corresponding CDs are grouped in 15 families: 5/A, 6/B, 7/C, 8/D, 9/E, 10/F, 11/G, 12/H, 26/I, 44/J, 45/K, 48/L, 51, 60, and 61, three of which (5/A, 9/E, and 45/K) being additionally divided in subfamilies. Separated by the slash lettering designations correspond to the initial structural nomenclature of cellulases introduced in 1991. Besides cellulases, other types of glycosyl hydrolases involved in plant cell wall polysaccharide degradation are also included.

Diversity of structural types of 1,4-β-glycosyl hydrolases is particularly striking in comparison with the enzymes attacking α-glucosidic bonds [18]. The absolute majority of the latter is grouped in only three families (13, 14, and 15). It is also interesting to note that xylan degradation by evolutionarily distant organisms also involves the enzymes of only two families (10/F and 11/G, Table 1). In nature, xylanases most often accompany cellulases and their substrate is comparable to cellulose in the reproduction scale. Nevertheless, xylanases, like amylases, are built more uniformly, although xylan has more complex chemical composition than cellulose. Apparently, xylans, alike α-glucans, appeared far later in the evolution process than cellulose.

Diversity of cellulases is remarkable even if we restrict our consideration to the enzymes capable of degradation highly ordered cellulose (cellobiohydrolases). As compared with similar β-amylases and glucoamylases belonging to families 14 and 15, respectively, cellobiohydrolases are spread in five families: 6/B, 7/C, 9/E(1), 10/F, and 48/L. A tentative reason for such diversity is polymorphous structure of native cellulose, which consists of at least two crystal modifications (triclinic Iα and monoclinic Iβ). This may require a concerted action of different types of enzymes for its complete decomposition to soluble sugars. Because of this, at least two different cellobiohydrolases participate, as a rule, in cellulose hydrolysis. Fungal cellobiohydrolases belong to the families 6/B and 7/C, those of aerobic bacteria *Cellulomonas fimi* to the families 6/B, 10/F, and 48/L, and those of *Clostridium* to the families 9/E(1) and 48/L.

There is one family (7/C) known to contain only fungal cellobiohydrolases, some (e.g., 48/L) only bacterial ones, and at least one containing both fungal and bacterial enzymes (6/B). On one hand, this may be evidence of specific "fungal" and "bacterial" mechanisms of cellulose decomposition, in contrast to the universal mechanism of xylan degradation. On the other hand, this may demonstrate a possibility of gene transfer among different microbial species. A good example of the latter is given in the review [4]. Endoglucanase V of the gram-negative bacterium *Erwinia chrysanthemi* was shown to have high degree of homology with endoglucanase I of the gram-positive bacterium *Cellulomonas uda*. The content of G + C in the gene of the first enzyme is equal to its average content in the genome of the gram-negative bacterium (58%). In contrast with that, the gene of second enzyme strikingly differs in G + C content from the average value of corresponding genome of gram-positive bacterium (61 and 72%, respectively). Hence, the horizontal transfer of cellulose gene from gram-negative to gram-positive organism seems very like. Another example includes family 5 endoglucanases CelE and CelB of anaerobic rumen fungi belonging to the genera of *Orpinomyces* and *Neocallimastix*. Their intron-less genes have high degree of homology with typical genes of family 5 bacterial cellulases and, tentatively, were transferred from anaerobic bacteria to the new fungal hosts [19].

FOLDING OF CATALYTIC DOMAINS OF VARIOUS CELLULASE FAMILIES

A decade has passed since the first cellulase 3D structures (fungal cellobiohydrolase II and clostridial endoglucanase D) were published [20, 21]. Since then, dozens of new structures were studied. Most of them derive from the well known folds of immunoglobulin-like "jelly roll", lectin-like β-sandwich, (β/α)8-barrel, and (α,α)6-barrel of various sizes [22]. Below some general features of the CD folds typical for the given families will be briefly considered.

The largest family 5/A is separated in five subfamilies. It includes both typical endoglucanases and endomannanases of bacteria and fungi, and some minor endo- and exoxylanases, cellodextrinases, and endo-1,6-β-glucanases.

In this family, the CD structures of clostridial endoglucanases (families A3 [23] and A4 [24]) and of the temperature stable endoglucanase EI from the thermophilic bacterium *Acidothermus cellulolyticus* (family A1 [25]) were described. The general folding motif of these enzymes is the well-known TIM-barrel (triosephosphate isomerase fold) with the active site localized near the C-terminus (Fig. 2).

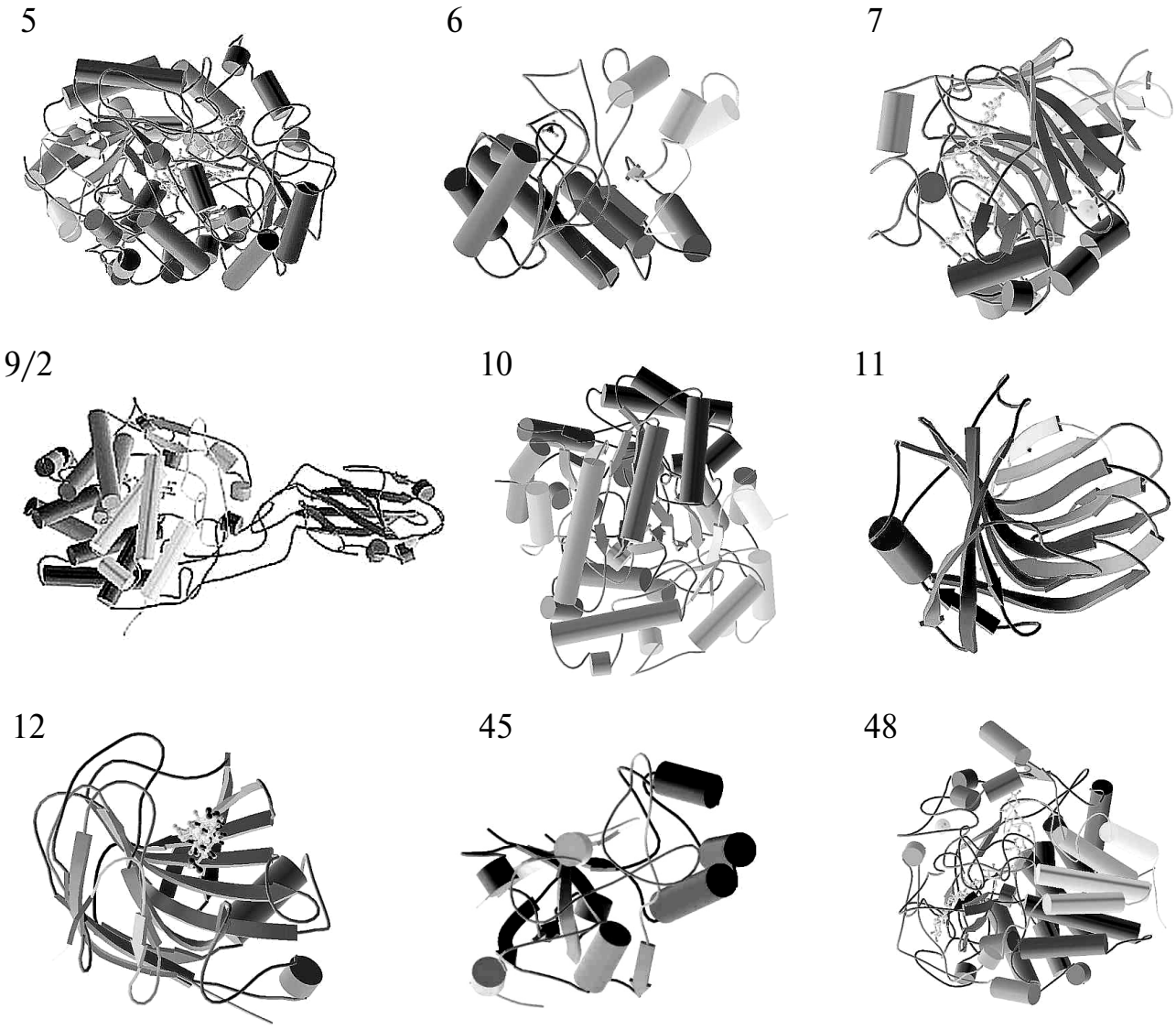

Fig. 2. Examples of catalytic domains of different glycosyl hydrolase families. Families: 5) endoglucanase EI of *Acidothermus cellulolyticus* in the complex with cellotetraose (1ECE); 6) endoglucanase E2 of *Thermomonospora fusca* (1TML); 7) cellobiohydrolase I of *Trichoderma reesei* with cellononaose model in the active site (8CEL); 9/2) endo-exocellulase E4 of *T. fusca* containing CD of family 9 with strongly fixed CBD of family III (1JS4); 10) xylanase Z of *Clostridium thermocellum* (1XYZ); 11) xylanase II of *Trichoderma harzianum* (1XND); 12) endoglucanase of *Streptomyces lividans* in the complex with 2-deoxy-2-F-cellotrioside (2NLR); 45) endoglucanase Y of *Humicola insolens* (2ENG); 48) endo/exocellulase of *C. cellulolyticum* in the complex with cellotetraose (1F9D) [18, 39] (PDB codes are given in parentheses).

Family 6/B consists of typical cellobiohydrolases and endoglucanases of fungi and aerobic bacteria. Among representatives of this family, 3D structure is studied in detail for CBH II of *Trichoderma reesei* [21]. Its fold also derives from TIM-barrel. However, the numbers of α-helices and β-strands in this barrel-like fold are reduced to five and seven ones, respectively. Because of this it can be defined as incomplete TIM-barrel. This structure was also confirmed in another family 6/B member: endoglucanase II of thermophilic actinomycete *Thermomonospora fusca* [26, 27] (Fig. 2). The structures of both enzymes differ in general by localization of the flexible loops, which define the structures of their active sites.

The 3D-structure of cellulases of the unique fungal family 7/C was studied in detail on the example of fungal CBH I [28] and later confirmed on endoglucanases I also assigned to this family [29]. It is folded in a distorted β-sandwich with its concave and convex sites formed by seven and eight anti-parallel β-strands, respectively. The β-strands are in some instances connected with short α-helices (Fig. 2). The CD of CBH I consists of 434 amino acid residues and has the size 40 × 50 × 60 Å. A similar fold is also shared by other groups of carbohydrate binding proteins: leguminous lectins and bacterial 1,3;1,4-β-glucanases, the latter being members of family 16.

Polypeptide chain of the member of family 8/D, the CD of cellulase CelC from mesophilic *Clostridium* species is folded in (α,α)6-barrel containing 365 amino acid residues [30]. This family also includes some 1,3(4)-β-glucanases.

The structure of the CDs of family 9/E (bacterial endoglucanases) was studied on the example of endoglucanase D from thermophilic *Clostridium* species [20], a member of subfamily 9/E2. It also comprises an (α,α)6-barrel, i.e., right-handed toroidal fold of six almost parallel inner α-helices surrounded with six outer α-helices going in the opposite direction. Each inner helix is connected with the proximal opposite outer helix as in helical membrane proteins. Near C-termini of inner helices three extended connective loops with insertions of short α-helices form the active site. The total length of the polypeptide chain folded in this structure is nearly 450 residues.

In the CD of endoglucanase D, the N-terminus of this fold is rigidly bound with a smaller immunoglobulin-like domain containing 90 amino acid residues. The latter consists of eight β-strands and is related to the connective domain localized between variable and constant parts of antibodies. Such a fold (so-called "jelly roll") forms a hydrophobic core fully composed of non-polar aromatic and aliphatic residues (Fig. 2). CDs of some other family 9/E members also contain similar domains, for example, CD of cellulase E4 from thermophilic actinomycete. It is suggested that these domains are a sort of CBD tightly bound with CD [31]. They are spatially oriented with respect to the position of (α,α)6-barrel CD so that the entrance in the active site extends the CBD binding site. Because of this, the cellulose molecule captured with the CBD is further directed in the active site. For this reason, the sequences of this sort of CBDs are often included in the CD structures of family 9/E2 members.

Two following families comprise a majority of known xylanases of a wide variety of species. The CD structure in the family 10/F was studied on the example of exoglucanase–xylanase of aerobic bacterium *C. fimi*, as well as xylanases of *Clostridium* and streptomycetes [32-34]. This type of CD is folded in TIM-barrel with the active site localized near C-terminus. This fold is very similar to that of CDs of family 5/A members (Fig. 2) [18]. Certain members of family 10/F split both xylan and cellulose derivatives.

The structures of smaller (near 185 amino acid residues) CDs of strictly specific xylanases assigned to family 11/G were determined for the enzymes of bacilli and fungi (*Trichoderma*) [35, 36]. Typical structure consists of three β-sheets and one α-helix. Sheets 1 and 2, each composed of five strands, for a structure of a "rib-cage", in which the second β-sheet is folded in the anti-parallel manner with respect to the first one and forms an angle with it. Both sheets are connected by means of short loops. The third sheet is composed of six strands. It comprises an extension of the second one, but is turned by 90° with respect to the latter. An α-helix lies transverse the background of the third sheet, the two final strands of the latter covering one of the sides of the helix. The enzyme active site is extended between second and third sheets along the whole width of the globule, the two strands of the third sheet forming "a lip" transverse to the active site (Fig. 2).

CD family 12/H includes endoglucanases of actinomycetes and fungi. The structures of endoglucanase produced by microscopic fungi *Aspergillus* and *Trichoderma* were resolved [37, 38]. Their folds are very similar to that of CD family 11/G members, in spite of the full absence of homology [38].

CDs of the family 26/I members are folded in already mentioned TIM-barrel with the active site localized at the C-terminus.

The structure of the CDs of endoglucanases assigned to family 44/J is not yet elucidated. These enzymes were shown to invert configuration of the anomeric center during hydrolysis.

In the following CD family of endoglucanases (45/K), the structure was resolved for endoglucanase V of thermophilic fungus *Humicola insolens* [39] (Fig. 2). Its CD has the size 40 × 40 × 25 Å and consists of 213 amino acid residues folded in seven β-strands and three short α-helices. The structural background of this fold is six-stranded β-barrel resembling "jelly roll" structure but composed of both parallel and anti-parallel β-strands. The seventh strand is attached through hydrogen bonding

to one of the strands of the main structure, but formally is not included in the β-barrel. The latter is similar with six-stranded β-barrel fold of barwin, a defense protein from barley, which is assumed to be a chitinase. In this respect, it is interesting to note that, according to the current structural classification, almost all chitinases including plant chitinases are assigned to the CD families 18 and 19.

The first enzyme of the bacterial endoglucanase and cellobiohydrolase family 48/L with resolved CD structure was CelF, the processive cellulase from mesophilic *Clostridium* species revealing the properties of both endo- and exoglucanase. Its major structural motif is also (α,α)6-barrel composed of 629 amino acid residues with the active site localized near the N-terminus [40].

Among others, CD structure was obtained for the enzymes of family 51, which includes endoglucanases and α-L-arabinosidases. These CDs are also folded in TIM-barrel like structure. In the endoglucanase families 60 and 61 the CD structures are not yet resolved.

Taken together, the diversity of structural families separated by hydrophobic cluster analysis far exceeds the limited number of typical CD folds of various cellulases and hemicellulases. Moreover, these enzymes share common folds with other glycosyl hydrolases and carbohydrate binding proteins. The most common fold among them is a full-size or incomplete TIM-barrel ((β/α)8-barrel)) shared by the enzymes of CD families 5/A, 6/B, 10/F, 26/I, and 51.

Topology of (α,α)6-barrel is typical for the enzymes of CD families 8/D, 9/E, and 48/L, although they significantly differ in size [18].

The enzymes of CD families 11/G and 12/H are folded in the "rib cage"-like structure, whereas the enzymes of fungal CD family 7/C have the structure of Con A-type β-sandwich. This topology is not typical for other cellulases but is shared by bacterial lichenases (CD family 16). The folds of CD families 7, 11, and 12 are grouped in a general 12-14-stranded β-sandwich category [18].

So-called closed 6-β-barrel of CD family 45/K belongs to a less common type of structure. However, it is also apparently shared by other carbohydrate binding proteins and, moreover, is structurally related to some immunoglobulin-like CBDs.

Common folding and localization of catalytic sites in the enzymes of different families has led to their grouping in superfamilies or clans. Five clans were separated among glucosyl hydrolases (GH), i.e., GH-A (families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, and 53), GH-B (families 7 and 16), GH-C (families 11 and 12), GH-D (families 27 and 36), and GH-E (families 33 and 34). The enzymes grouped in the clan GH-A (another name is superfamily 4/7) are characterized by the common fold of elliptic (β/α)8-barrel with the major axis lying between first and fifth β-strands. In this structure, the general acid-base and nucleophilic catalytic groups, which are responsible for double displacement (i.e., retention) of anomeric center configuration in the hydrolysis products, are localized at the termini of fourth and seventh strands, respectively (being the reason for the superfamily name) [41]. Apparently, with accumulation of data on glycosyl hydrolase structures some of their families will be grouped in a limited number of clans. However, the number of clans will certainly exceed the number of known folds, because even the enzymes sharing common fold (e.g., TIM-barrel) may differ in catalytic mechanism, e.g., retain or invert configuration of anomeric carbon center in the hydrolysis products.

CLASSIFICATION OF THE SUBSTRATE BINDING DOMAINS

CBD is the second important and the most widespread element of cellulase structure involved in cellulose transformation. In fact, CBD often plays the role of a recognition factor, which is used by enzyme-producing microorganism to address secreted polysaccharide hydrolases to the plant cell wall to be decomposed. The first evidence of their existence was obtained more than two decades ago [42-44]. Since then, the number of CBDs identified as the elements of cellulases, xylanases, mannanases, and other enzymes now exceeds 150 and constantly increases. Current structural classification of CBDs includes at least thirteen different families [45, 46].

Variable localization of CBDs (at the C- or N-terminus, or in the middle of polypeptide chain) as well as the presence of some identical or even different CBDs in one protein molecule, and, finally, the existence of non-hydrolytic proteins containing only CBD repeats suggest an independent function of these structural modules and high level of their autonomy. Therefore, their coupling with the enzyme CDs may be a recent evolutionary event.

Table 2 illustrates the CBD classification developed by the scientists of the University of British Columbia (Vancouver). It aims to overview the spread of CBDs among different species as well as the diversity of their structural forms and localization in the molecules of different proteins.

As seen from this table, bacterial CBDs significantly differ in their polypeptide chain length and form nearly a decade of families (II to X), large families II and III being additionally divided in subfamilies (a) and (b). Contrary to that, known eucaryotic CBDs are (with minor exceptions) highly homologous and are grouped in one family I. Apart from CBDs originating from the enzymes of basidiomycetous, filamentous, and anaerobic fungi, this family also includes CBDs of the polysaccharide binding protein from the alga *Porphyra purpurea*, which is, in fact, the fourfold repeat of fungal CBD. However, family I does not include any procaryotic CBDs.

Table 2. Classification of cellulose binding domains (according to [4])

Family	Organism	Enzyme	Location*	Length**	Fold	CD family	Source***
1	2	3	4	5	6	7	8
I	*Agaricus bisporus*	Cel1	C	36	wedge-like, triple-stranded antiparallel β-sheet	61	M86356
	Fusarium oxysporum	Xyn	N	36		10/F	L29380
	Neocallimastix patriciarum	XylB	C	33		10/F	S71569
	Humicola insolens	EGY	C	33		45/K	[51]
	Neurospora crassa	CBHI	C	33		7/C	X7778
	Penicillium janthinellum	CBHI	C	33			X59054
	Phanerochaete chrysosporium	CBHI	C	34			M22220
	Porphyra purpurea	PBP	×4	33		polysaccharide binding protein	U08843
	Trichoderma reesei	CBHI	C	33		7/C	P00725
		CBHII	N	36		6/B	M16190
		EGI	C	33		7/C	M15665
		EGII	N	36		5/A5	M19373
		EGY	C	36		45/K	Z33381
		Man	C	34		5/A1	L25310
	Myceliophthora thermophila	CDH	I			cellobiose dehydrogenase	[88]
IIa	*Butirivibrio fibrisolvens*	End1	C	95	ten antiparallel β-strands, nine of them forming β-barrel	5/A4	X17538
	Cellulomonas fimi	CenA	N	106		6/B	M15823
		CenB	C	103		9/E2	M64644
		CenD	C	105		5/A1	L02544
		CbhA	C	106		6/B	L25809
		CbhB	C	104		48/L	L29042
		Cex	C	106		10/F	L11080
	Clostridium cellulovorans	EngD	C	108		5/A4	M37434
	Dictiostelium discoideum	SGSP	C,I	98/106		? (is formed during spore germination)	M33861
	Microbispora bispora	CelA	C	100		6/B	P26414
	Pseudomonas fluorescens	CelA	C	100		9/E1	X12570
		CelB	N	102		45/K	X52615
		CelC	N	99		5/A1	X61299
		CelE	C	100		5/A2	X86798
		XynA	N	101		10/F	X15429
	Streptomyces lividans	CelA	N	108		5/A2	M82807
		CelB	C	106		12/H	U04629
		ChiC	N	105		18/chitinases	D12647

Table 2. (Contd.)

1	2	3	4	5	6	7	8
	Thermomonospora fusca	E2 E3 E5	C N N	96 103 103		6/C 6/C 5/A2	M73321 U18978 L01577
IIb	*Cellulomonas fimi*	XynD	×2C	90	skew β-sandwich	11/G	X76729
	Streptomyces lividans	Axe	C	86		acetylxylan esterase	[52]
	Thermomonospora fusca	XynA	C	86		11/G	U01242
IIIa	*Bacillus lautus*	CelA	C	150	nine-stranded β-sandwich binding Ca^{2+}	44/J	M76588
	Bacillus subtilis	Cel	C	133		5/A2	X67044
	Caldicellulosiruptor (Caldocellum) saccharolyticum	CelA CelB CelC ManA	×2I I ×2I ×2I	172 172 172 172		9/E2//48/L 10/F//5/A1 9/E2//5/A4 5/A4//44/J	L32742 X13602 [62] L01257
	Clostridium cellulovorans	CbpA	N	161		scaffoldin	M73817
	Clostridium stercorarium	CelZ	C	133		9/E2	X55299
	Clostridium thermocellum	Cbh3 CipA CelI	C I C	132 156 150		9/E1 scaffoldin 9/E2	X80993 L08665 L04735
	Erwinia carotovora	CelV	C	156		9/E2	X76000
IIIb	*Cellulomonas fimi*	CenB	I	138		9/E2	M64644
	Clostridium stercorarium	CelZ		144			X55299
	Clostridium thermocellum	CelI		137			L04735
IV	*Cellulomonas fimi*	CenC	×2N	148	"jelly roll" of two β-sheets each containing five (or five and six) strands	9/E1	X57858
	Clostridium cellulolyticum	CelE	N	168			M87018
	Streptomyces reticuli	Cel1		125			L04735
	Thermomonospora fusca	E1		141			L20094
	Myxococcus xantus	CelA		139		6/B	X76726
V	*Erwinia chrysanthemi*	EgZ	C	63	"ski boot", β-sheet of three antiparallel β-strands with perpendicular flexible loop	5/A2	Y00540
VI	*Bacillus polymyxa*	XynD	C	90		?	X57094
	Clostridium stercorarium	XynA	×2C	87/92		11/G	D13325

Table 2. (Contd.)

1	2	3	4	5	6	7	8
	Clostridium thermocellum	XynZ	I	92		10/F	M22624
	horseshoe crab	G-α	×2C	87		16 (coagulation factor)	D16623
	Microbispora bispora	BglA	C	85		?	L06134
VII	*Clostridium thermocellum*	CelE	I	240		5/A4	M22759
VIII	*Dictiostelium discoideum*	CelA	N	152		9/E2	M33861
IX	*Clostridium thermocellum*	XynX	×2C	174/189		10/F	M67438
	Thermoanaerobacterium sp.	XynA		174/187			M97882
	Thermotoga maritima	XynA		170/180			Z46264
X	*Cellvibrio mixtus*	XynA	C	51	two β-sheets of two and three antiparallel strands,	11/G	Z48925
	Pseudomonas fluorescens	CelA	I	55	α-helix across larger	9/A1	X12570
		CelB		55	sheet	45/K	X52615
		CelC		53		5/A1	X61299
		CelE	C	53		5/A2	X86798
		XynA		53		10/F	X15428
		XynE		55		11/G	Z48927

* Localization of CBD at C-, N-terminus, or in the middle of protein chain; ×4 or ×2 indicates the number of CBD repeats.
** Approximate length of CBD polypeptide [4].
*** Codes of amino acid sequences from SWISS-PROT, EMBL/GenBank, or PIR as given in corresponding reviews.

All CBDs of family I consist of 35 to 40 amino acid residues. Its length as well as variability of N-terminal sequence apparently depend on the localization at C- or N-terminus of polypeptide chain of fungal enzyme, because N-terminal CBDs may include some residues of signal peptide, whereas C-terminal ones may capture some residues of interdomain linker. According to 2D-NMR data obtained for CBD of *T. reesei* cellobiohydrolase I, they form a wedge-like fold with exposed hydrophilic and hydrophobic planes, which contains as a basic structure distorted β-sheet of three short anti-parallel strands. This structure is stabilized by at least two conservative disulfide bonds (9-26 and 20-36) (Fig. 3) and contains four highly conservative aromatic residues on its hydrophobic plane, at least three of which are involved in the binding with cellulose surface [47]. Interestingly, topology of fungal CBD is similar to that of potato low molecular mass carboxypeptidase inhibitor [48], although comparison of their primary sequences does not show clear homology. In this respect, it is important to note formal resemblance of fungal CBDs and protein inhibitors of proteinases in some structural aspects. This includes, at first, highly or absolutely conserved cysteines, apparently as a result of instability of secondary structure of all these small-size proteins in the absence of stabilizing disulfide bonds. The second common feature is often di- or oligomerization of both structures [49], the example of which is the tetrameric CBD isolated from red alga. Third, both CBDs and protein inhibitors may be attached to the respective CD through a flexible linker sequence. For example, procarboxypeptidase A contains its own protein inhibitor [50], whose removal during proteolytic activation of the proenzyme is very similar to the separation of CBD from cellulase CD during specific proteolysis with papain.

High level of homology of all known CBDs of fungal family I, as well as their compatibility with six fungal CD families (5/A1,5, 6/B, 7/C, 10/F, 45/K, and 61) and relatively small size suggest their high efficiency and perfect evolutionary construction.

Family II includes CBDs of various bacterial enzymes produced by aerobic and anaerobic bacteria and actinomycetes. CBDs of the largest subfamily IIa of this family are compatible with CDs of families 5/A1,2,4,

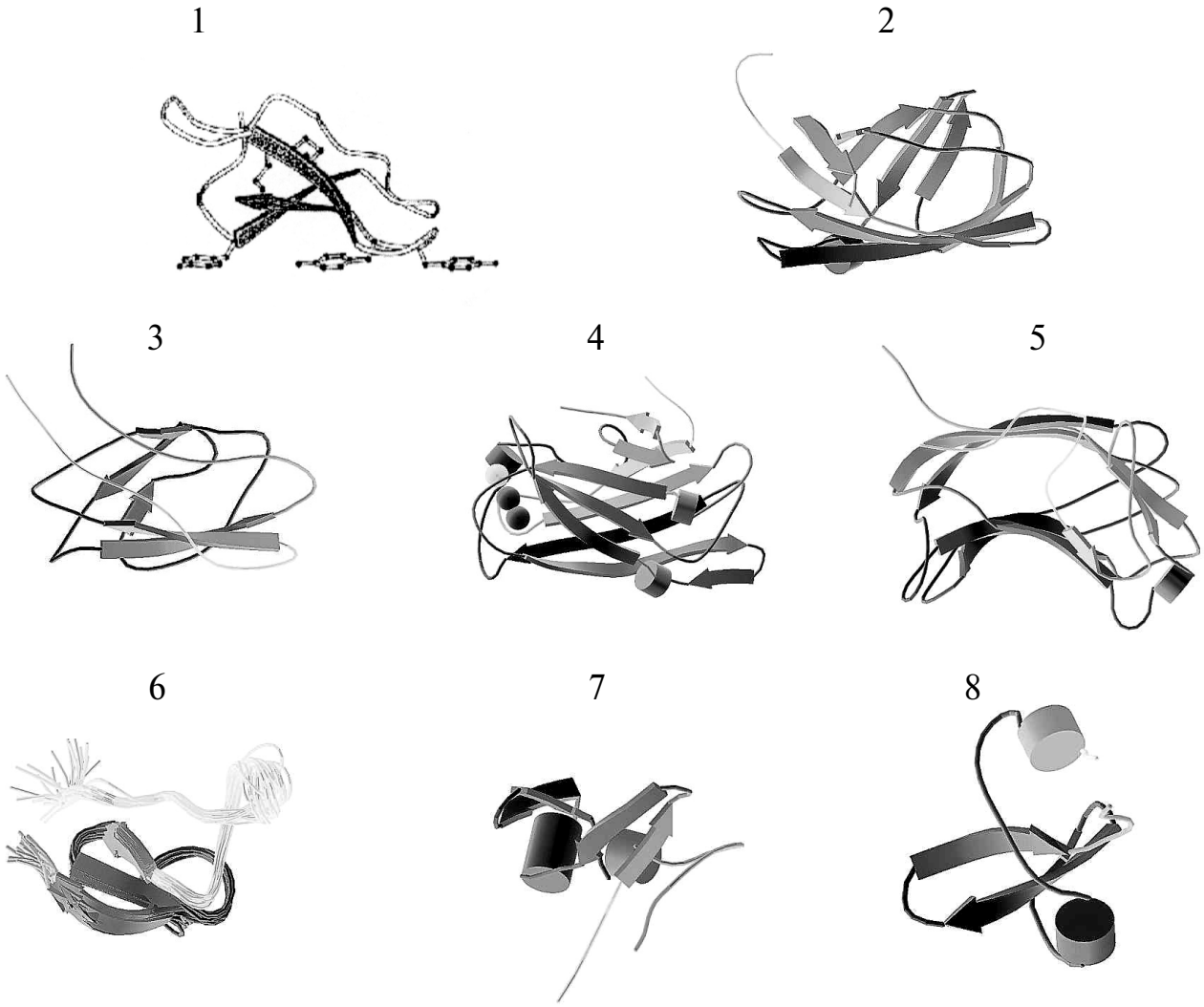

Fig. 3. Examples of noncatalytic modules of cellulases and xylanases: *1*) CBD of family I from *T. reesei* cellobiohydrolase I with aromatic residues important for cellulose binding [52]; *2*) CBD of subfamily IIa of *C. fimi* exoglucanase–xylanase Cex (1EXG); *3*) CBD of subfamily IIb of *C. fimi* family 11 xylanase (1HEH); *4*) CBD of subfamily IIIa from *C. cellulolyticum* scaffoldin with bound Ca ions (1G43); *5*) CBD_{N1} of family IV of *C. fimi* endoglucanase CenC (1ULO); *6*) CBD of family V of *E. chrysanthemi* endoglucanase Z having similarities with bacterial chitinase domains (1AIW); *7*) CBD of family X of *Pseudomonas* xylanase having similarities with oligonucleotide/oligosaccharide binding domain OB (1E8R); *8*) one of two repeats of N-terminal dockerin of endoglucanase Cel45A of anaerobic fungus *Piromyces equi* demonstrating a new folding motif of a polypeptide containing 42 residues (1E8P) (PDB codes are given in parentheses).

6/B, 9/E1,2, 10/F, 12/H, 45/K, and 48/L. CBDs of the subfamily IIb were yet found only in combination with CDs of family 11/G (strictly specific xylanases without cellulase activity). Apart from cellulases and xylanases, CBDs of family II were also found in other enzymes involved in hemicellulose decomposition, i.e., arabinosidases, acetyl esterases, as well as in chitinases. They are also found in the protein synthesized during spore germination of the myxomycete *Dictyostelium discoideum*, the organism sometimes assigned to fungi and sometimes to Protozoa Kingdom. In any event, this is one of the rare exceptions of a eucaryotic CBD that does not belong to the CBD family I.

In this respect it is important to note that CBDs of families I and II differ not only by their origin from the organisms of different complexity. They also differ in the binding specificity. According to published results, fungal CBDs can bind to cellulose but not to chitin. In contrast, some CBDs of family II were found to also bind to xylan or chitin [53]. This leads to an assumption that binding to cellulose may, in fact, be a secondary function of CBD from the spore-germinating protein of myxomycete.

The ability of some family II CBDs to bind to chitin allows for a comparison with polysaccharide binding domains of glucoamylases from Aspergilli. As shown for glucoamylase I of *A. awamori* var. *kawachi* capable of

binding to both raw starch and chitin, limited proteolysis of this full-size enzyme released small glycopeptide fragments capable of binding to either chitin or raw starch [54]. Although the structure of the glycopeptides was not elucidated, this demonstrates the presence in fungal enzymes of special domains responsible for specific binding to cell wall chitin. These domains of chitin-forming eucaryotes may not be similar to family I CBDs but may resemble chitin-binding CBDs of other families.

CBDs of family IIa consist of 95 to 108 amino acid residues folded, according to 2D-NMR data for the CBD of *C. fimi* exoglucanase–xylanase [55], in ten anti-parallel β-strands, nine of them forming an extended β-barrel (45 × 25 × 25 Å) of tendamistat type. The β-strands are extended along the longest direction of the β-barrel (Fig. 3). A cone-shaped cavity of unknown function (7 Å in the depth and 5 Å across the basis) was also identified in this structure. This fold in CBDs of family II is, as a rule, stabilized by disulfide bridge formed by two cysteine residues, one near the C-terminus, whereas the other is near the N-terminus. Subfamily IIa CBDs contain four highly conservative tryptophan residues, three of them (W14, W34, and W68) being exposed in the solution and accessible for binding with cellulose. Site-directed mutagenesis revealed the involvement of W14 and W34 or W68 in the binding to cellulose surface in various CBDs.

CBDs of subfamily IIb differ from the major subfamily by shorter length (86 to 90 amino acid residues) resulting from a number of short deletions. Deletion of eight amino acid residues near the C-terminus including highly conservative W68 is the most important one. Apparently because of this, CBDs of subfamily IIb demonstrate differences in binding specificity. Part of them binds only to xylan, whereas the other to both xylan and cellulose. In this respect, this subfamily should more correctly be designated as xylan binding domains, whose affinity for cellulose is a secondary function. Tomme et al. [4] believe that variations in specificity of family II CBDs may arise from wide variations in the distribution of the charged groups on the surface of the protein globule. However, on the first look this should have more effect on the sorption to glucuronoxylan bearing negatively charged side-chain substituents, but not on cellulose sorption.

Family III comprises a number of relatively large CBDs of aerobic and anaerobic bacteria with polypeptide chain lengths from 131 to 172 residues. It is also separated in two subfamilies. The main subfamily IIIa contains CBDs of cellulases, whose CDs are assigned to the families 5/A1,2,4, 9/E1,2, 10/F, 44/J, and 48/L. Also CBDs of scaffoldins, i.e., clostridial cellulosome integrating proteins are also assigned to the subfamily IIIa. These CBDs are most often localized at the C-termini of mature proteins. However, in the natural chimerical enzymes from bacterium *Caldicellulosiruptor (Caldocellum) saccharolyticum*, which contain two CDs of different glycosyl hydrolase families, these CBDs may be localized in the middle of polypeptide chain and even as a repeat (Table 1). Clostridial scaffoldins may also bear internal or even N-terminal CBDs, one of which was also found to be capable of chitin binding.

Subfamily IIIb CBDs differ from the representatives of the major family in the deletion of a number of highly conservative aromatic residues. N-Terminal conservative motif RY(Y/W)Y of subfamily IIIa CBDs is reduced in these CBDs to RYF. All the latter are internal and are combined only with glycosyl hydrolase family 9/E2 CDs, where they may be associated with subfamily IIa CBDs. In comparison with CBDs of subfamilies IIa and IIIa, they provide weaker cellulose binding.

Structural studies of family III CBDs were accomplished on various examples. Among them, subfamily IIIa CBD of *Clostridium thermocellum* scaffoldin is folded in nine-stranded β-sandwich containing a Ca^{2+} binding site. Conserved amino acid residues are distributed between two separate faces. In one face, these residues form a shallow groove of yet unknown function. On the other face, conserved aromatic and polar residues are aligned in a planar strip, which is supposed to provide contact of CBD with three adjacent polymeric chains on a cellulose surface. The most important binding is supposed to involve four anhydroglucose units of the central cellulose chain and the strip-constituting CBD residues W118, H57, Y67, and a salt bridge formed by D56 and R112. Additional interaction of the CBD with two adjacent cellulose chains apparently involves two groups of polar residues: N10, N16, and Q110 on one site of the strip, and S12 and S133 on the other [56].

The CBDs assigned to other CBD families are comparatively rare. CBD family IV includes CBDs of anaerobic and aerobic bacteria and of actinomycetes. Having similar size with CBDs of family III (139-168 residues), they reveal a weak homology with them. These CBDs are found only in combination with CDs of families 9/E1, 6/B, and 10/F and only from N-termini, where they may form a repeat [57]. A unique feature of some of them is their ability to bind onto amorphous (i.e., phosphoric acid-swollen) cellulose, not crystalline cellulose. According to the data of heteronuclear NMR, one of them, CBD_{N1} of CenC cellulase from *C. fimi* is folded in a "jelly roll"-like structure formed by two β-sheets each containing five β-strands. β-Strands of the one face form a concave binding site capable of fixing up to six anhydroglucose moieties. It contains a hydrophobic strip bordered with hydrophilic residues. In contrast with other CBDs, these domains do not contain typical tryptophan binding sites [58].

A unique CBD of endoglucanase Z from Gram-negative bacterium *E. chrysanthemi* forms separate CBD family V. It consists of 63 residues and is located at the C-terminus of the enzyme, whose catalytic domain belongs to the glycosyl hydrolase family 5/A2. 2D-NMR did not

reveal any structural similarity of this domain with CBDs of other families. Like fungal CBDs, it has a flat face containing one Y and two W residues supposed to be involved in cellulose binding [59].

Family VI represents CBDs of clostridial xylanases belonging to glycosyl hydrolase families 10/F and 11/G, as well as CBD of actinomycetous β-glucosidase A. In addition, a duplicated domain of 1,3-β-glucan-specific factor of coagulation G-α from hemolymph of horseshoe crab is assigned to this structural family. All these CBDs contain 85 to 92 amino acid residues and are located at either C-termini or in the middle of polypeptide chains. They may exist as a repeat. One of them was confirmed to be capable of binding to both cellulose and purified xylan.

Both family VII and family VIII consist of only one member CBD with demonstrated affinity for cellulose. CBD of clostridial endoglucanase E (glycosyl hydrolase family 5/A4) represents family VII, whereas that of cellulase A from myxomycete (glycosyl hydrolase family 9/E2) forms CBD family VIII. The former is the largest among all known CBDs and consists of 240 amino acid residues. The later has the size similar to those of CBDs from families III and IV (152 residues) and is located at the N-terminus.

CBDs of family IX were found in duplicated form at the C-termini of some heat stable xylanases of glucosyl hydrolase family 10/F produced by three species of thermophilic anaerobic bacteria. They are relatively large and contain 170 to 189 amino acid residues.

CBDs of family X represent smaller modules of 51-55 residues located at the C-termini or in the central parts of polypeptides of bacterial enzymes (mainly from pseudomonads) of glycosyl hydrolase families 5/A1,2, 9/E1, 10/F, 11/G, and 45/K.

In the recent years this CBD classification was extended and the new families XI-XIII were formed [60].

A different type of xylan binding domain was also identified in heat stable modular xylanases of some thermophilic bacteria (*Caldibacillus cellulovorans*, *Caldicellulosiruptor cellulolyticum*) and was initially designated as TSD (thermostabilizing domain), because its separation decreased temperature optimum of xylanase activity by 10°C. Later it was shown to be capable of binding to soluble xylan and hydroxyethyl cellulose. Interestingly, modular endoxylanase XynA bearing this domain at its N-terminus also bears two family IIIb CBDs at the C-terminus [61, 62].

Taken together, CBDs of different families significantly vary in their polypeptide chain length. Those of fungal CBDs are the shortest. They are three- to sevenfold shorter than those of bacterial CBDs of families II, III, IV, VI, VII, and IX. In contrast, identified bacterial CBDs have comparable size with small CDs of some glycosyl hydrolases of families 11/G, 12/H, or 45/K. Moreover, folding of β-strands of CBDs of families II, III, and IV reveals structural similarity with CD family 45/K immunoglobulin-like or "jelly roll" fold. In general, a relationship of both types of β-glycan binding structures (CBDs and CDs) can be revealed. Both of them comprise relatively small protein domains with both spatial structure and carbohydrate recognition sites defined by a limited number of amino acid residues arranged in a pack of distorted β-sheets consisting mainly of anti-parallel strands. This type of arrangement of aromatic and some other amino acid residues capable of anchoring hexameric fragment of linear 1,4-β-glycan on a short distorted anti-parallel β-sheet was first described for hen egg-white lysozyme more than three decades ago [63]. Since then, many similar folds were found among both CDs and CBDs of cellulases. Although some CD families of cellulases (e.g., 8/D, 9/E, and 48/L) reveal different structural fold of (α,α)6-barrel, they comprise larger structures containing from 350 to 600 residues. Furthermore, even such a large structure is sometimes insufficient for anchoring 1,4-β-glucan and needs an "assistance" of strongly fixed immunoglobulin-like CBD, as in the case of glycosyl hydrolase subfamily 9/E2. Therefore, distorted antiparallel β-sheet is apparently the most compact and evolutionary perfect protein basis for fixing of six or more sugar moieties of linear 1,4-β-glucans.

We have already mentioned formal structural similarities of fungal CBDs and low molecular weight protease inhibitors. It should also be noted that duplication ability typical for inhibitor structures is also widespread in both fungal and other CBD families. Many bacterial CBDs of various families are often present as a repeat in cellulase, xylanase, and mannanase molecules, etc. As in the case of proteinase inhibitors [49], this may result from some common features of β-folds, e.g., their ability to form stable interdomain bonds through exchange by homologous β-strands. Indeed, such a phenomenon was observed in highly concentrated solution of family IIa CBD by using 2D-NMR [55]. Furthermore, formal structural similarity of CBDs and protease inhibitors may reflect possible bifunctionality of both groups of protein structures. As evidence, an ability of some plant proteinase inhibitors to suppress polysaccharide hydrolase (amylase) activity can be mentioned [49].

Comparing two major structural elements of cellulases, i.e., CDs and CBDs, it is reasonable to briefly consider their specificity of interaction with 1,4-β-glucans. In contrast with CDs, which are usually strictly specific for the structure of sugar residue preceding cleavage bond, assignment of CBD to a distinct structural family does not necessarily define its binding specificity. Within the same CBD family specificity of different members for different linear 1,4-β-glycans (cellulose, chitin, xylan) may vary significantly.

Another distinguishing feature of CBDs from CDs is their far stronger affinity for insoluble polysaccharides as compared to their soluble derivatives. Such an affinity

results from major difference of CBDs from glycosyl hydrolase CDs: their polysaccharide binding sites comprise flat or slightly convex surfaces, in contrast with deep grooves or tunnels of glycosyl hydrolases. Aromatic residues Trp and Tyr playing the key role in sugar binding are exposed in CBDs outside in contrast with those of CDs. Owing to this, can simultaneously anchor several adjacent polymeric molecules of the exposed surface of polysaccharide [56], this being a principal difference from CDs. Excluding specific for amorphous cellulose CBD_{N1}, other CBDs do not usually form complexes with cellooligosaccharides in solution, which can be observed and studied by 2D-NMR, even at high concentrations of soluble ligands. This is the major reason of the phenomenon described more than to decades ago: inability of soluble cellulose derivatives to displace insoluble cellulose in the complex with the adsorbed enzyme and by this manner to release cellulases tightly bound onto the surface of crystalline substrate [42, 64]. As was mentioned above, CBDs on the surface can apparently rather with certain functional groups of sugar residues belonging to the adjacent chains than with glucose residue as a whole. This can provide sufficient binding affinity specific for a flat surface [65]. In addition, CBDs probably do not bear interfering barrier sites, which are typical for CDs and provide distortion of the native chair conformation of sugar residue preceding bond cleavage for acceleration of the catalytic event.

LINKERS

In most cases CDs and CBDs are separated in the molecules of cellulases and other enzymes by linker sequences. Usually, linkers comprise flexible disordered chains rich in proline and hydroxy amino acid residues (serine and threonine), as well as glycine and alanine (so-called PT-linkers similar with extensins of plant cell walls) (Table 3) [8]. Their length can vary from 5-6 to 100 residues, although most often is limited by the range 20 to 50 residues. In many linkers repeated motifs of four-seven residues can be identified, where some positively, or negatively charged, or hydrophobic residues are inserted within PTS-rich sequence.

Periodicity with high proline content can result in collagen-like secondary structure typical for extensins [4]. However, their crystal structure cannot be obtained because numerous threonines serve as O-glycosylation sites, while glycines provide increased flexibility. Because of this, linker sequences along with attached CBDs are usually removed by specific proteolysis before crystallization. Proteolysis occurs in the most flexible part of linker usually adjacent to one the functional domains, where G and P residues are localized. More rigid linker parts containing numerous hydroxyamino acids are supposed to be protected from proteolysis by glycosylation.

Table 3 also demonstrates similarity of the primary structures of fungal linkers and the central parts of polypeptide chains of fungal CBDs of family I (residues 15-25). In the CBDs this part includes typical linker motif GPT with absolutely conserved G16. This motif may also be found at the N-termini of some CBDs and may originate from preceding linker sequence. Central GPT-part of linkers functions as a connective loop between two conservative structural fragments forming anti-parallel β-sheet. The most striking difference of this flexible loop of CBDs from typical linker sequences is the presence of two or even three Cys residues, which are inserted either between typical linker residues or instead of Pro and covalently fix the loop on β-sheet. This leads to a conclusion that fungal CBD structure originally appeared from linker sequence by insertion of two small structural-functional blocks rich in aromatic and other residues and capable of forming short β-strands and β-turns. Then flexible linker loop was fixed onto the β-fold through insertion of conserved cysteines.

Linkers are believed to provide spatial separation of CDs from CBDs to allow their autonomous function on the surface of insoluble substrate. Rigid glycosylated part is responsible for spatial separation, whereas flexible part provides autonomous domain function and is supposed to play a role of a hinge. Reduction of linker's length has almost no effect on binding and within the certain limits only slightly decreases enzyme activity towards insoluble substrate. However, mutant fungal cellobiohydrolases I with deleted hinge and rigid parts of linker demonstrated reduced activity towards ordered cellulose while had almost the same affinity for the insoluble substrate [66]. Therefore, the role of linkers may not be restricted by the solely spatial separation of functional domains; rather, they provide a concerted action of all parts of the enzymatic molecular machine on the cellulose surface.

It is interesting to note in this respect, that extensins, the proteoglycans of primary cell walls of higher plants, whose amino acid sequence reveals similarity with linkers, play, according to one of hypotheses, a principal role in the shift of protein matrix along cellulose fibers during expansion growth of plant cells.

ON THE MECHANISM OF ENZYMATIC DEGRADATION OF HIGHLY ORDERED FORMS OF CELLULOSE

We have considered the impressive data obtained in the last 10-15 years on the structure of individual enzymes of microbial cellulase–hemicellulase systems. However, understandings of how cellulases bring about degradation of highly ordered natural cellulose is still far from complete.

Figure 4a reproduces without any changes the scheme of enzymatic hydrolysis of cellulose published as

Table 3. Linker sequences of some cellulolytic enzymes and their comparison with central parts of family I CBD sequences (according to [4, 8])

Organism	Enzyme	Sequence	Length*	Pro	Tre + Ser
1	2	3	4	5	6
Bacillus spp.	CelA	$T_2P_2SDPTP_2SDPDGEPGPDPGEPDPTP_2SDP$	33	15	7
	CelB	(**1**)$P_2SDPTP_2SDPDPGEPDPTP_2SDPGEYP$	28	13	5
		(**2**)$P_2SEPSDP_4SEPE(PDPGE)_3PDPTP_2SDPEYP$	42	20	5
	Egl1	$T_2EPVEPEPVDPGE_2TP_2$	18	6	3
Butyrivibrio spp.	End1	$(PDPTPVD)_4PDPQPVDPTP$	38	17	5
	Xyn	$PGSFTPQPTITPQ(PT)_2PSGQT$	26	7	8
Caldicellulosiruptor (Caldocellum) saccharolyticum	CelB	(**1**)$T_2S_2(PT)_4(VT)_2(PT)_5VTAT(PT)_3PVSPAT$	43	14	23
		(**2**)$PAPTMTVAPTAT(PT)_2LSPTV(TP)_2APTQTAI(PT)_2LTPN(PT)_2$	44	14	17
Cellulomonas fimi	CenA	$PT_2S(PT)_4T(PT)_7VTPQPT$	33	14	17
	CenB	(**1**)$PTGT_3DT_2P_2T_2PGTP$	17	5	9
		(**2**)$T_2DT_2GETEP_2T_2PGTP$	17	4	8
		(**3**)$T_2A_2PVDTVAPTVPGTP$	17	4	5
		(**4**)$S_2PVTFT_2LPVTSTPS$	16	3	8
	CenC	(**1**)SLT_2SATP_3	10	3	5
		(**2**)PVPTAP	6	3	1
	Cex	$(PT)_3T(PT)_3T(PT)_3S$	21	9	12
Clostridium spp.	Egl	$T_2PTS_3PVYTSPITISKT_3$	21	3	13
	CelA	$PLSDLSGQPTP_2SNPTPSLP_2$	21	8	6
	CelB	$TPSVT(PS)_2ATPSPT_2ITAP_2T$	22	7	11
	CelE	$PLVS(PT)_3LMPTPSPTVT$	20	7	8
	CelH	(**1**)$(PT)_3WTSTP_2S_3P$	16	6	9
		(**2**)$PGTYPSYSPKPSPTPRPTKP_2VTP$	24	10	7
	XynZ	(**1**)TPVPTPSPKP	10	5	3
		(**2**)TPNPSVTPTQTPIPT	15	6	6
Erwinia chrysanthemi	CelZ	$T_2DPSTDT_2MTP_2LTNRPQPT$	21	5	9
Fibrobacter succinogenes	CelC	$PVS_3DMSPTS_2DAVIDPTS_3A_2V_2DPST$	30	4	13
Microbispora bispora	CelA	$P_2TYSPSPTPST(PS)_3QSDPGS(PS)_3$	30	12	14

Table 3. (Contd.)

1	2	3	4	5	6
Xanthomonas campestris	EngA	$T_2(PT)_{11}$	24	11	13
Streptomyces sp.	CasA	(**1**)$PRT_2(PT)_2P$	9	4	4
		(**2**)$PA_2TGA(SP)_2AP_2ASPAPSADS$	22	7	6
Pseudomonas fluorescens	CelA	(**1**)$S_{11}VPVS_7I_2PS_6IQPS_6MPS_8V_2AS_5VS$	59	4	44
		(**2**)$S_4ASNINS_{12}AIVS_5V_2S_6$	37	0	28
	CelB	(**1**)$S_2APS_2VAS_7V_2S_2TPRS_5VS_3VPGTS_7$	42	3	30
		(**2**)$STS_3TPLS_6RS_2VAS_4LS_2ATS_3AS_2VS_2$	37	1	28
	XynA	(**1**)$S_3APAS_2VPS_2IAS_3PS_2VAS_2VIS_2MAS_3PVS_4VAS_2TPGS_3$	49	5	28
	XynB	(**2**)$S_6LS_4V_2S_2IRS_9$	26	0	21
		(**1**)$SAT_2S_2VAS_4TPT_2S_4AS_2VAS$	26	1	19
		(**2**)$SVS_5VQS_9A_2S$	21	0	16
Trichoderma reesei	EGI	PPPPPASSTTFSTTRRSSTTSSSPS	25	6	14
	EGII	PGATTITTSTRPPSGPTTTTRATSTSSSTPPTSS	34	6	21
	CBHI	GNPSGGNPPGGNRGTTTTRRPATTTGSSPGPT	34	6	11
	CBHII	$PGAAS_5TRAASTTSRVSPTTSRSSSATPPPGST_2RVPPVG$	44	7	22
Fragments 15-26 of family I CBDs					
Trichoderma reesei	EGI	SGcKTcTSGTTc	12(3)	0	6
	EGII	SGPTNcAPGSAcST	14(2)	2	5
	CBHI	SGPTVcASGTTc	12(2)	1	5
	CBHII	SGPTccASGSTc	12(3)	1	5
Agaricus bisporus	Cel3a	TGPTTcASGSTc	12(2)	1	6
Humicola insolens	EgY	SGcTTcVAGSTcT	13(3)	0	6
Neurospora crassa	CBHI	SGPTTcPEPYTc	12(2)	3	4

* In parentheses, the numbers of Cys residues which are not present in the linkers.

early as in 1984 [67]. This scheme was formulated on the basis of investigation of a decade of cellulase systems of various fungi [43, 68, 69]. For comparison, Fig. 4b reproduces the scheme published fifteen years later, in 1999 [70] in coauthorship with one of the best known cellulase researchers, Goran Petersson, the discoverer of fungal cellobiohydrolase II and whose studies in the 50-70s along with those of Reese, Wood, and Brown have created the basis of modern views on the structure and mechanism of action of cellulases of soft rot and white rot wood decay fungi.

As seen from Fig. 4a, already early in the 80s it had become evident that hydrolysis of crystalline cellulose requires two types of endoglucanases differing in the presence or absence of additional adsorption center later designated as CBD. It was suggested that weakly binding endoglucanases hydrolyze outer amorphous regions leaving crystalline parts intact [64, 68, 71]. These endoglucanases were found as minor components of cellulase systems of *T. reesei* and *P. chrysosporium*. Later these enzymes of *T. reesei* and *P. chrysosporium* were identified as endoglucanase III and endoglucanase 28, respectively, and designated as TrCel12A and PcCel12A in the modern cellulase nomenclature. However, the central role in the decomposition of crystalline cellulose was assigned to the tightly binding enzymes dominating in these enzyme systems and represented by endoglucanases I and II (Cel7B and Cel5A, respectively), and cellobiohydrolases I and II (Cel7A, Cel6A, respectively) of *T. reesei* with p*I* values of 4.8, 5.3, 4.2, and 5.9, respectively [72] (Fig. 4a). Later structural studies of specialists from VTT (Finland) confirmed the presence of CBDs in all these enzymes of *T. reesei* [73].

Transfer of CBD-bearing enzymes among separate particles of insoluble substrate was first detected by using simultaneous decomposition of combinations of colorless and colored cellulose substrates [44, 72, 74-76]. This reversibility of CBD adsorption was also confirmed later in an independent investigation [77].

The absence of major endoglucanase components with high affinity for cellulose in cellulase complexes of a number of Aspergilli (*A. niger*, *A. foetidus*) not capable of ordered cellulose decomposition was also confirmed later. Indeed, endoglucanases of Aspergilli were predominantly assigned to the fungal enzymes of family 12/H, which do not bear CBDs [4]. Hence, the need for a special sorption center (CBD in the current designation) in the enzymatic decomposition of the ordered cellulose is accepted now by all the well-known specialists as one of the milestones of our understanding of this phenomenon.

Tentative mechanism of action of such a sorption center was first suggested in [78]. This center was assumed to provide concentration and tight binding of cellulases within the defects of ordered cellulose structure. Furthermore, these domains were suggested to wedge in the defects thus spreading microfibril bundles of cellulose and releasing separate fibrils. Such an effect typical only for tightly binding endoglucanases and cellobiohydrolases was observed, when hydrolytic activity of the CBD-bearing enzymes was suppressed by the addition of an excess of inhibitor. The hypothesis of noncatalytic assistant effect of CBD was later considered by other authors [73, 77] and successfully confirmed by using individual bacterial CBD of family II [53]. Although similar data for isolated fungal CBDs were not obtained, this may result from the difference in experimental conditions. We observed spreading effect of CBD-containing fungal enzymes on specially pretreated (sonicated) suspension of cellulose substrates, whereas other authors did not apply sonication for this purpose.

Studies on the dynamics of accumulation of the chain end groups on the surface of microcrystalline cellulose during enzymatic decomposition [79] revealed predominantly end-wise mode of hydrolysis of polymeric molecules in the ordered regions by both exo- and endoglucanases capable of attacking both non-reducing and reducing chain end groups (Fig. 4). This suggestion was confirmed by demonstration of preferable attack of cellobiohydrolase I and cellobiohydrolase II on the opposite (reducing and non-reducing, respectively) chain end groups of cellulose molecules. Finally, evidence of some random-wise constituent in the fungal cellobiohydrolase action [80] along with the discovery of the entire family 48/L of predominantly end-wise acting bacterial cellulases of a transient type between exo- and endoenzymes [81] erases the borders between these two types of cellulases and supports all principle statements of the mechanism demonstrated by Fig. 4a.

Presently, multidomain cellulase structure should obviously be considered from the point of view of a concerted action of the enzyme modules as the elements of a molecular device providing dissociation and hydration of the chain end group tightly packed in the surface structure of elementary cellulose fibril. This task cannot be solved by the solely structural studies.

Here one of possible explanations of the unique ability of fungal cellobiohydrolase I to decompose structurally ordered cellulose is suggested. Cleavage of glycosidic bond itself is obviously not a rate-limiting step of this rather slow process. More probably, the overall process is restricted by the hydration of cellulose molecule necessary for enzyme action. The hydration process may proceed through several steps.

An initial step (binding of CBD) results in swelling of surface cellulose molecules owing to the effect of noncatalytic spread of microfibril bundles. The following stages apparently include cooperative action of the general structural elements of the enzyme, i.e., CBD, CD, and linker. Here we propose a mechanism of consecutive movement of the enzyme along the reaction pathway comprising a set of energetically equivalent enzyme–substrate intermediates: (CBD**–CD) (initial sorption) →

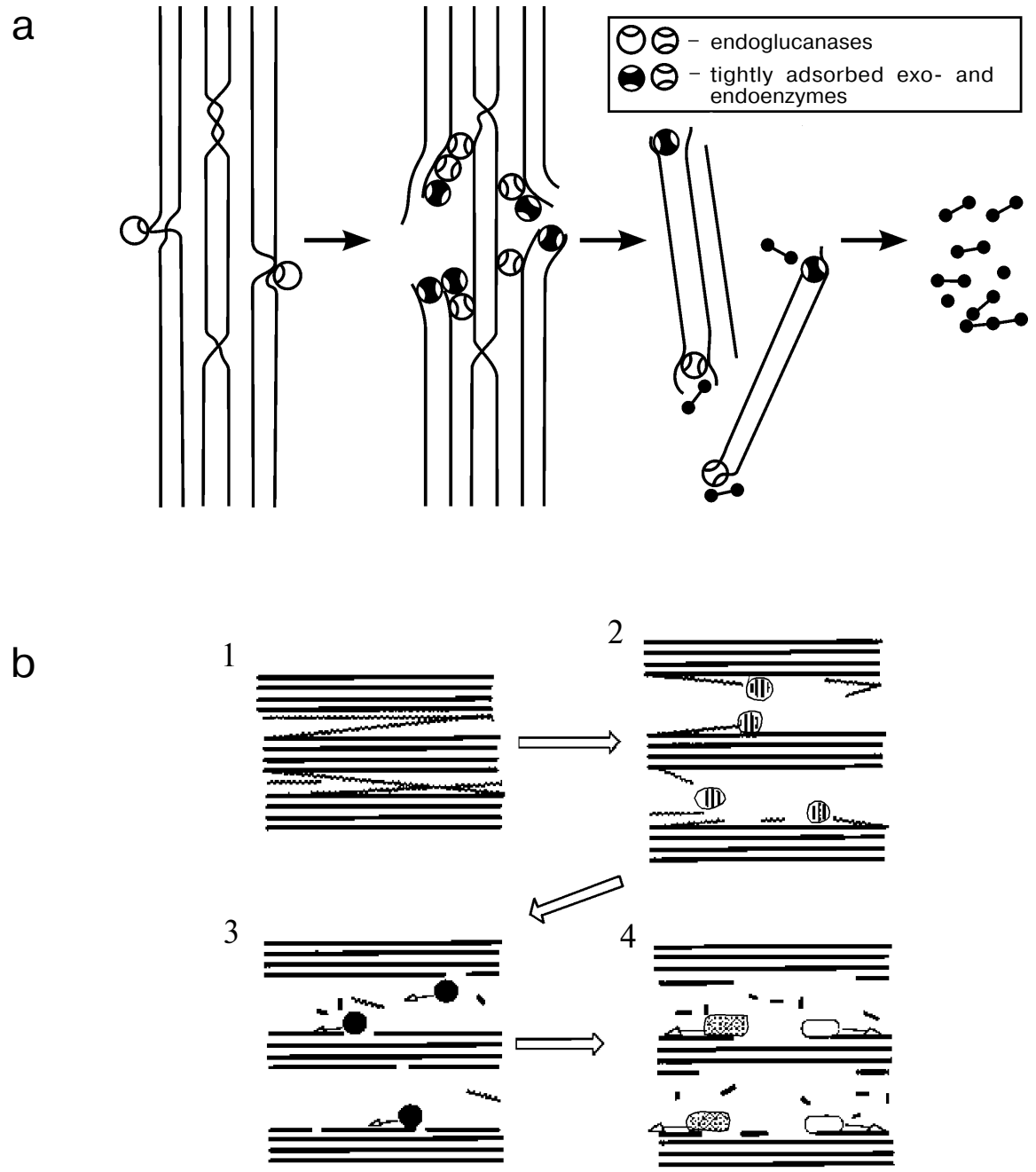

Fig. 4. a) Mechanism of ordered cellulose decomposition by fungal enzyme systems suggested in the beginning of 80s. Open circles correspond to endoglucanases of low (one segment) or high affinity for cellulose (two segments corresponding to the active site and an adsorption site, i.e., CBD in the current abbreviation). Filled circles correspond to cellobiohydrolases characterized by high affinity for cellulose, which have both active and adsorption sites. Decomposition starts from the attack of low affinity endoglucanases (minor components of *T. reesei* cellulase system) on the outer amorphous regions of bundles of elementary fibrils. This opens inner defects in the ordered structure, which are destroyed by high affinity endoglucanases. Formed separate microcrystallites are then hydrolyzed to soluble sugars by high affinity cellobiohydrolases and endoglucanases by the end-wise manner from both reducing and non-reducing chain ends (reproduced from [78]). b) Current concept of ordered cellulose decomposition by mold (*T. reesei*) and white-rot (*P. chrysosporium*) fungal enzymes (according to [70]). Minor family 12 endoglucanases of fungal cellulase systems, which do not bear CBDs (stripped circles), produce primary attack on amorphous parts of the bundles of elementary fibrils thus opening inner fibrils for CBD-bearing endoglucanases I and II of families 7 and 5 (dark circles bearing wedge-like tails). The latter form polymer chain end groups in more ordered regions. Further decomposition resulting in liberation of soluble sugars proceeds from both reducing and non-reducing chain ends owing to family 7 cellobiohydrolases I (ellipses bearing wedge-like tails and filled with points) and family 6 cellobiohydrolases II (open ellipses bearing wedge-like tails),

(CBD*–CD*) → (CBD–CD**) → (CBD*–CD*) + P (the number of asterisks at corresponding domain designates the relative degree of its involvement in cellulose binding; where P is the soluble sugar molecule released by a productive complex CD**). Formation of the tightest complex (CBD**–CD**), where both domains form all possible bonds with the substrate, is considered to be unlike for two reasons. At first, because of steric incompatibility of two modes of binding. CBD binds to the hydrophobic sites of anhydroglucose units on the planar face of cellulose microcrystal, whereas CD to the hydrated cellulose chain end group. Therefore, formation of complex (CBD**–CD**) is, in principle, possible only with a rather long and flexible interdomain linker. Second, too tight and stable complex (CBD**–CD**) would be a long-lived limiting intermediate in the catalytic turnover. Although highly extended and flexible linker would provide autonomous action of CD on larger area around anchor CBD, such a construction should be ineffective at low surface concentrations of the adsorbed enzyme because of rapid exhaustion of adjacent substrate and reversible enzyme inactivation in the tight nonproductive complex CBD**–CD*.

We speculate that cellobiohydrolase-like enzymes of very high affinity for the substrate (like fungal cellobiohydrolase I or clostridial cellobiohydrolase A [82, 83]) utilize the difference between free energies of the tightest and partially weakened binding of both domains for the hydration of the cleaved cellulose molecule. Cellobiohydrolase I can do it owing to rigid glycosylated part of the interdomain linker, which resists mechanical deformation and provides the "see-saw" mode of action, when either CBD or CD with entrapped polymeric chain end group are dissociated from cellulose surface. The chain end can initially be trapped by CD in the form of nonproductive complex CD*, where only gluconic subsites near the entrance of the active site tunnel are occupied. Simultaneously, partial dissociation of surface-bound CBD proceeds (CBD** transforms into CBD*). Entrapment of the chain end (reducing for fungal cellobiohydrolase I or non-reducing for clostridial cellobiohydrolase A) results in hydration of the adjacent segment of polymeric chain. Owing to flexibility of the linker's hinge this provides further slicing of the chain, until the later reaches the agluconic subsites in the tunnel and forms productive CD** complex. This, in turn, should release CBD from its complex or, at least, substantially weaken its binding because of local surface hydration and additional increase in elastic deformation of the rigid part of interdomain linker. Liberation of soluble sugar molecule and transformation of productive CD** complex to a new nonproductive CD* complex opens the possibility for CBD to bind again onto the surface (on certain distance from CD*). In this manner a new distorted complex (CBD*–CD*) is formed and the catalytic cycle repeats. In this mode of action CD moves along the splitting cellulose molecule as a snail shell followed by the leader CBD, which functions as "a head choosing the direction for further travel".

Such a mechanism, named "caterpillar catalysis" by the analogy with inchworm-type transportation [10], is not yet proved. However, some positions can be verified in the experiments, e.g., non-additive affinity for cellulose of CBD and CD in the native two-domain enzyme [82, 84]. An alternative mechanism of action of modular enzymes with linkers consisting of rigid and flexible parts suggests strong fixation of CBD on the surface followed by catching short hydrated cellulose chain ends by CBD. In this case CBD plays a role of a "fisher", CD is "fish hook", while linker functions as an elastic fishing rod gradually pooling the hydrated cellulose chain end group from hydrophobic environment of adjacent polymeric molecules. In this case catalytic cycle would consist of the following steps: (CBD**–CD*) → (CBD**–CD**) → (CBD**–CD*) + P and the reaction velocity would gradually decrease after initial burst since steady state determined by surface diffusion of CBD is reached. Such a mechanism of reversible surface inactivation of the enzyme could explain a phenomenon described in [85, 86] and our publications [72, 87], which have not yet received strict theoretical interpretation. Its experimental verification will include analysis of the relationship between the reaction rate and degree of enzymatic conversion of cellulose, on one hand, and the length and composition of linker for a series of enzymes with identical or similar structures of the active site, on the other.

However, in some instances, the role of CBD may, indeed, be limited by the solely transportation of the enzyme to certain structural elements of native cellulose. One of the examples is CBD-bearing noncatalytic cellulosome integrating protein (scaffoldin) of anaerobic organisms, which provides self-assembling of a cluster of hydrolases and their transportation to cellulose surface [4, 13]. Fungal cellobiose dehydrogenase (CDH) can be considered as another example. CDH is an extracellular flavohemoprotein, which couples the processes of cellulose and lignin biodegradation. It was found in both basidiomycetes and filamentous fungi [88]. Although it does not split cellulose molecules, just oxidizing their reducing chain end groups to carboxylic groups, it has very high affinity for cellulose. The CDH of one filamentous fungus contains typical fungal CBD of family I, whereas CDH of basidiomycete contains a unique cellulose binding sequence rich in aromatic residues, which has no homology with any of CBDs in the described 13 families [88].

It is obvious that the mode of action of a certain enzyme on cellulose surface is defined by a number of parameters: length and structure of its active site, the strength of fixation of the polypeptide loops forming active site, the probability of free rotation of the segments of polymeric substrate molecule within the active site,

affinities of different subsites for monomeric units of polymer, location of CBD with respect to CD and its affinity for cellulose surface, length, and degree of conformational elasticity of the interdomain linker, the presence of other domains in the enzyme structure, which can form protein–protein aggregates on cellulose surface. Because of strictly topochemical character of the overall process, structural data should be considered in the context of heterogenous kinetics and thermodynamics, taking into account evolution of the enzyme–substrate interactions during hydrolysis [72, 89].

REFERENCES

1. Henrissat, B. (1997) *Mol. Microbiol.*, **23**, 848-849.
2. Nutt, A., Sild, V., and Petterson, G. (1998) *Eur. J. Biochem.*, **259**, 200-206.
3. Kubicek, C. P. (1992) *Adv. Biochem. Eng. Biotechnol.* (Fiechter, A., ed.) Vol. 45, Springer-Verlag, Berlin-Heidelberg, pp. 1-27.
4. Tomme, P., Warren, R. A. J., and Gilkes, N. R. (1995) *Adv. Microb. Physiol.*, **37**, 1-81.
5. Bennett, N. A., Ryan, J., Biely, P., Vrsanska, M., Kremnicky, L., Macris, B. J., Kekos, D., Christakopoulos, P., Katapodis, P., Claeyssens, M., Nerinckx, W., Ntauma, P., and Bhat, M. K. (1998) *Carbohydr. Res.*, **306**, 445-455.
6. Beguin, P. (1990) *Annu. Rev. Microbiol.*, **44**, 219-265.
7. Beguin, P., and Alzari, P. M. (1998) *Biochem. Soc. Trans.*, **26**, 178-185.
8. Gilkes, N. R., Henrissat, B., and Kilburn, D. G. (1991) *Microbiol. Rev.*, **55**, 303-315.
9. Henrissat, B., and Bairoch, A. (1996) *Biochem. J.*, **316**, 722-726.
10. Rabinovich, M. L., and Melnick, M. S. (2000) *Advances in Biological Chemistry* [in Russian], Vol. 40, ONTI PNTs RAN, Pushchino, pp. 205-266.
11. Rodionova, N. A. (1989) *Advances in Science and Technology. Biotechnology* [in Russian], Vol. 19, VINITI, Moscow.
12. Sinitsyn, A. P., Gusakov A. V., and Chernoglazov, V. M. (1995) *Bioconversion of Lignocellulose Materials* [in Russian], Moscow State University Publishers, Moscow.
13. Beguin, P., and Aubert, J.-P. (1994) *FEMS Microbiol. Rev.*, **13**, 25-58.
14. Bolobova, A. V., and Kurakov, A. V. (1999) *Prikl. Biokhim. Mikrobiol.*, **35**, 402-408.
15. Kurakov, A. V., and Bolobova, A. V. (1999) *Prikl. Biokhim. Mikrobiol.*, **35**, 332-341.
16. Erikson, K.-E., Blanchette, R. A., and Ander, P. (1990) *Microbial and Enzymatic Degradation of Wood and Wood Components*, Springer Verlag, Berlin-Heidelberg, Germany.
17. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.-P. (1989) *Gene*, **81**, 83-92.
18. Himmel, M. E., Karplus, P. A., Sakon, J., Adney, W. S., Baker, J. O., and Thomas, S. R. (1997) *Appl. Biochem. Biotechnol.*, **63-65**, 315-325.
19. Chen, H., Li, X. L., Blum, D. L., and Ljungdahl, L. G. (1998) *FEMS Microbiol. Lett.*, **159**, 63-68.
20. Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P., and Aubert, J.-P. (1992) *Nature*, **357**, 89-91.
21. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990) *Science*, **249**, 380-386.
22. Tavares, G. A., Beguin, P., and Alzari, P. M. (1997) *J. Mol. Biol.*, **273**, 701-713.
23. Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D., and Derewenda, Z. S. (1994) *J. Biol. Chem.*, **269**, 20811-20819.
24. Roig, V., Fierbode, H.-P., Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Belaich, J.-P., and Haser, R. (1993) *J. Mol. Biol.*, **223**, 325-330.
25. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., and Karplus, P. A. (1996) *Biochemistry*, **35**, 10648-10660.
26. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) *Biochemistry*, **32**, 9906-9916.
27. Spezio, M., Irwin, D., Karplus, P. A., Zhang, S., Taylor, J., and Wilson, D. B. (1993) *Genetics*, *Biochemistry and Ecology of Lignocellulose Degradation* (Shimada, K., and Ohmiya, K., eds.) Uni Publishers, Tokyo, pp. 308-317.
28. Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1994) *Science*, **265**, 524-530.
29. Davies, G., Tolley, S., Wilson, K., Schulein, M., and Dodson, G. (1992) *J. Mol. Biol.*, **228**, 970-972.
30. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (1998) *EMBO J.*, **17**, 5551-5562.
31. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) *J. Bacteriol.*, **180**, 1709-1714.
32. Dominguez, R., Souchon, H., and Alzari, P. M. (1994) *Proteins: Structure, Function and Genetics*, **19**, 158-163.
33. Souchon, H., Spinelli, S., Beguin, P., and Alzari, P. M. (1994) *J. Mol. Biol.*, **255**, 1348-1356.
34. White, A., Withers, S. G., Gilkes, N. R., and Rose, D. R. (1994) *Biochemistry*, **33**, 12546-12554.
35. Arase, A., Yomo, T., Urabe, I., Hata Y., Katsube, Y., and Okada, H. (1993) *FEBS Lett.*, **316**, 123-127.
36. Torronen, A., and Rouvinen, J. (1995) *Biochemistry*, **34**, 847-855.
37. Okada, H. (1991) *Microbial Utilization of Renewable Resources* (Kinoshila, S., and Bhumiratana, A., eds.) International Center of Cooperative Research in Biotechnology, Japan, pp. 1-7.
38. Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Oarkson, K., and Bott, R. (1993) *Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P., and Reinikainen, T., eds.) Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, pp. 153-158.
39. Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, M. J., Rasmussen, G., and Schulein, M. (1993) *Nature*, **365**, 362-366.
40. Parsiegla, G., Juy, M., and Haser, R. (1998) *EMBO J.*, **17**, 5551-5562.
41. Davies, G. J. (1998) *Biochem. Soc. Trans.*, **26**, 167-173.
42. Rabinovich, M. L. (1977) *Carbohydrase Kinetics (Lysozyme and Cellulolytic Enzymes)*: Candidate's dissertation [in Russian], School of Chemistry, Moscow State University, Moscow.

43. Klyosov, A. A., and Rabinovich, M. L. (1980) *Enzyme Engineering: Future Directions* (Wingard, L. V., and Berezin, I. V., eds.) Plenum Press, New York-London, pp. 83-165.
44. Klyosov, A. A., and Rabinovich, M. L. (1989) *Tricel 89*, Technical University, Vienna, L. 11.
45. Mattinen, M. L., Kontteli, M., and Kerovuo, J. (1997) *Protein Sci.*, **6**, 294-303.
46. Jervis, E. J., Haynes, C. A., and Kilburn, D. G. (1997) *J. Biol. Chem.*, **272**, 24016-24023.
47. Kraulis, P. J., Clore, G. M., Nigles, M., Jones, T. A., Pettersson, G., Knowles, J. K. C., and Gronenborn, A. M. (1989) *Biochemistry*, **28**, 7241-7257.
48. Hass, G. M., Hermodson, M. A., Ryan, C. A., and Gentry, L. (1982) *Biochemistry*, **21**, 762-766.
49. Mosolov, V. V., and Valueva, T. A. (1993) *Plant Proteinase Inhibitors* (Kretovich, V. L., ed.) Institute of Biochemistry, Russian Academy of Sciences, Moscow.
50. McFallen, C. A., Svendsen, I., Jonassen, I., and James, M. N. G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7242-7246.
51. Schou, C., Rasmussen, G., Kalloft, M. B., Henrissat, B., and Schulein, M. (1993) *Eur. J. Biochem.*, **217**, 947-957.
52. Tomme, P., Warren, R. A. J., Miller, R. C., Jr., Kilburn, D. G., and Gilkes, N. R. (1995) *Enzymatic Degradation of Insoluble Carbohydrates* (Saddler, J., and Penner, M., eds.) ACS, Washington D. C., USA, pp. 1-12.
53. Kilburn, D. G., Assouline, Z., Din, N., Gilkes, N. R., Ong, E., Tomme, P., and Warren, R. A. J. (1993) *Trichoderma reesei Cellulases and Other Hydrolases. Foundation for Biotechnical and Industrial Fermentation* (Suominen, P., and Reinikainen, T., eds.) Helsinki, pp. 281-290.
54. Hayashida, S., and Yoshioka, H. (1980) *Agric. Biol. Chem.*, **44**, 481-487.
55. Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) *Biochemistry*, **34**, 6993-7009.
56. Bayer, E. A., Morag, E., Lamed, R., and Shoham, Y. (1998) *Curr. Opin. Struct. Biol.*, **8**, 548-557.
57. Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) *Biochem J.*, **345**, 53-60.
58. Creagh, A. L., Koska, J., Johnson, P. E., Tomme, P., Joshi, M. D., McIntosh, L. P., Kilburn, D. G., and Haynes, C. A. (1998) *Biochemistry*, **37**, 3529-3537.
59. Chapon, V., Czjzek, M., El Hassouni, M., Py, B., Juy, M., and Barras, F. (2001) *J. Mol. Biol.*, **310**, 1055-1066.
60. Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A. L., Sturch, K., Gilkes, N. R., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (1998) *J. Chromatogr. B. Biomed. Sci. Appl.*, **715**, 283-296.
61. Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2000) *Microbiology*, **146**, 2947-2955.
62. Sunna, A., Gibbs, M. D., and Bergquist, P. L. (2000) *Biochem. J.*, **346**, Pt. 3, 583-586.
63. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1965) *Nature*, **206**, 757-763.
64. Rabinovich, M. L., Nguen Van Viet, and Klyosov, A. A. (1982) *Biokhimiya*, **47**, 465-477.
65. Rabinovich, M. L. (1988) *Microbiology and Biochemistry of the Decomposition of Plant Raw Materials* (Skryabin, G. K., and Golovlev, E. L., eds.) [in Russian], Nauka, Moscow, pp. 70-108.
66. Srisodsuk, M., Reinikainen, T., and Teeri, T. T. (1993) *J. Biol. Chem.*, **268**, 20756-20761.
67. Rabinovich, M. L. (1984) *Materials of Soviet-Finland Seminar on Bioconversion of Plant Raw Materials by Microorganisms*, Institute of Biochemistry and Physiology of Microorganisms, Pushchino, pp. 31-48.
68. Rabinovich, M. L., Chernoglazov, V. M., Klyosov, A. A., Nguen Van Viet, and Berezin, I. V. (1981) *Doklady AN SSSR. Ser. Biokhim.*, **260**, 1481-1486.
69. Klyosov, A. A., and Rabinovich, M. L. (1980) *Materials II Int. Symp. on Bioconversion and Biochem. Engineering*, Scylark Printers, New Dehli, pp. 53-56.
70. Henriksson, G., Nutt, A., Henriksson, H., and Pettersson, L. G. (1999) *Eur. J. Biochem.*, **259**, 88-95.
71. Rabinovich, M. L., Chernoglazov, V. M., and Klyosov, A. A. (1983) *Biokhimiya*, **48**, 369-377.
72. Rabinovich, M. L., Melnick, M. S., Badalov, A. B., Klyosov, A. A., and Berezin, I. V. (1985) *Doklady AN SSSR. Ser. Biokhim.*, **282**, 1013-1017.
73. Knowles, J., Teerei, T. T., Lehtovaara, P., Penttila, M., and Saloheimo, M. (1988) *Biochemistry and Genetics of Cellulose Degradation* (Aubert, J.-P., and Beguin, B., eds.) Academic Press, Paris, pp. 153-169.
74. Rabinovich, M. L., Klyosov, A. A., and Berezin, I. V. (1984) *Doklady AN SSSR. Ser. Biokhim.*, **274**, 758-763.
75. Rabinovich, M. L., Novikova, T. V., Klyosov, A. A., and Berezin, I. V. (1985) *Bioorg. Khim.*, **11**, 1343-1347.
76. Rabinovich, M. L., Klyosov, A. A., and Melnick, M. S. (1986) *Analyt. Biochem.*, **156**, 489-494.
77. Linder, M., and Teerei, T. T. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12251-12255.
78. Rabinovich, M. L. (1984) *Abst. XVI FEBS Conf.*, Moscow, p. 393.
79. Rabinovich, M. L., Nguen Van Viet, and Klyosov, A. A. (1986) *Prikl. Biokhim. Mikrobiol.*, **22**, 70-79.
80. Barr, B. K., Hsieh, Y.-L., Ganem, B., and Wilson, D. B. (1996) *Biochemistry*, **35**, 586-592.
81. Reverbel-Leroy, C., Parsiegla, G., and Moreau, V. (1998) *Acta Crystallogr. D. Biol. Crystallogr.*, **54**, Pt. 1, 114-118.
82. Herner, M. L., Melnick, M. S., and Rabinovich, M. L. (1999) *Biochemistry (Moscow)*, **64**, 1012-1020.
83. Melnick, M. S., Rabinovich, M. L., and Voznyi, Ya. V. (1991) *Biokhimiya*, **56**, 1787-1797.
84. Herner, M. L., Melnick, M. S., Leontyeva, G. V., Romenskaya, I. G., and Rabinovich, M. L. (2000) *Prikl. Biokhim. Mikrobiol.*, **36**, 5-7.
85. Gusakov, A. V., Sinitsyn, A. P., and Klyosov, A. A. (1986) *Prikl. Biokhim. Mikrobiol.*, **22**, 59-69.
86. Converse, A. O., Ooshima, H., and Burns, D. S. (1990) *Appl. Biochem. Biotechnol.*, **24/25**, 67-73.
87. Obraztsova, I. N., Fedosenko, M. A., and Rabinovich, M. L. (1993) *Prikl. Biokhim. Mikrobiol.*, **29**, 684-689.
88. Henriksson, G., Johansson, G., and Pettersson, G. (2000) *J. Biotechnol.*, **78**, 93-113.
89. Yuldashev, B. T., Rakhimov, M. M., and Rabinovich, M. L. (1993) *Prikl. Biokhim. Mikrobiol.*, **29**, 58-68.