Protein Science (1992), *1*, 1293-1297. Cambridge University Press. Printed in the USA.

Specificity mapping of cellulolytic enzymes: Classification into families of structurally related proteins confirmed by biochemical analysis

MARC CLAEYSSENS[1] AND BERNARD HENRISSAT[2]

[1] Laboratorium voor Biochemie, Faculteit Wetenschappen, Rijksuniversiteit, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium
[2] Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53X, F-38041 Grenoble, France

(RECEIVED April 10, 1992; REVISED MANUSCRIPT RECEIVED May 13, 1992)

Abstract

The specificities of 15 cellulolytic enzymes have been examined using chromophoric glycosides derived from D-glucose, cellobiose, higher cellooligosaccharides, lactose, D-xylose, and β-(1,4)-xylobiose. Coinciding with a classification based on hydrophobic cluster analysis of amino acid sequences, six families each showing a characteristic specificity pattern were observed. Furthermore, in these cases where the anomeric forms of reaction products were determined, results seem to indicate conservation of intrinsic reaction mechanism (single or double displacement) within each family. On the other hand, the low molecular weight substrates do not discriminate exo- from endocellulases. This functional differentiation is speculated to originate from the presence, in exoenzymes, of a tunnel-shaped active site formed by extra loops in their structure.

Keywords: cellulases; chromophoric substrates; classification; mechanism of action; structure-function relationships

Cellulase and xylanase primary structures become more and more available and a comparative study by hydrophobic cluster analysis (HCA) of the catalytic domains has been particularly successful in detecting sequence similarities between distantly related members of this ubiquitous class of enzymes (Henrissat et al., 1989). The method, claimed to detect folding similarities between proteins with extensively diverging primary structures (Gaboriaud et al., 1987), originally revealed six cellulase-xylanase families with no interfamily similarity (Henrissat et al., 1989). With the new sequence data the number of families has recently been extended to nine (Gilkes et al., 1991b). Because the sequence identity between some members of these families can be as low as 12-15%, it is important to verify if they indeed share a similar fold and exhibit the same topology at their active site. This could become obvious by studying the catalytic properties and substrate specificities of these enzymes. We have developed a rapid and sensitive method for screening cellulase specificities using small, soluble, and well-characterized substrates (van Tilbeurgh et al., 1988). In the present paper, evidence is provided for a correlation between these specificities and the structural classification reported earlier (Henrissat et al., 1989; Béguin, 1990; Gilkes et al., 1991b).

Reprint requests to: Marc Claeyssens, Laboratorium voor Biochemie, Faculteit Wetenschappen, Rijksuniversiteit, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium.

Results and discussion

The sequences of nearly 80 cellulase and xylanase catalytic domains are known and can be classified into at least nine different families on the basis of primary structure similarities (Gilkes et al., 1991b). The present study deals only with cellulases belonging to the six families originally defined by Henrissat et al. (1989).

Chothia and Lesk (1986) have shown that the deviation of the polypeptide backbone atoms of structurally related proteins increases as a function of sequence divergence. They have also shown that sequence as well as geometry divergence is less pronounced in the hydrophobic core than in the peripheral loops. This observation seems valid for most globular proteins, hence probably also for cellulases where, within a given family of catalytic domains,

the spatial arrangement of the active site residues should be conserved (Zvelebil & Sternberg, 1988). This should be reflected by the conservation of the substrate specificity and of the molecular mechanism which, in the case of glycosyl hydrolases, determines the stereochemical outcome of the reaction (retention or inversion of configuration) (Sinnott, 1990).

Fluorophoric or chromophoric glycosides derived from the lower cellodextrins, lactose, D-xylose and β-(1,4)-xylobiose are shown in Figure 1. They are adequate tools in rapid screening of cellulase activities by HPLC analysis (van Tilbeurgh et al., 1988). Most often the degradation patterns coincide with those found for the unsubstituted ($_1H^3$-reducing end-labeled) cellooligosaccharides (Claeyssens et al., 1990b). In a few cases, however, deviations due to unspecific interactions of the (hydrophobic) chromophore are observed (Bhat et al., 1990), but the use of these artificial substrates for screening purposes remains valid as is discussed below.

Figure 2 summarizes the results obtained when the specificities of 15 cellulolytic enzymes (Table 1) were mapped in this way.

Five enzymes from family A were studied and their typical hydrolysis patterns of small substrates are shown (Fig. 2A). Their catalytic action is restricted to β-(1,4)-glucosidic linkages, and their turnover numbers are typically low ($k_{cat} < 50\ min^{-1}$) and unaffected by the substrate chain length (van Tilbeurgh et al., 1988). With the higher cellodextrins, cellotriose seems preferentially formed at the nonreducing end. Some enzymes (e.g., endoglucanase CelC from *Clostridium thermocellum*) are active against the lactoside and/or the cellobioside and this may reflect the existence of subfamilies as recently proposed (Béguin, 1990). All investigated enzymes in this family have been shown to hydrolyze glycosidic bonds with overall retention of anomeric configuration (Table 1), and it can be concluded that a double displacement mechanism is typical (Sinnott, 1990).

This and the single displacement mechanism (inversion of configuration) to describe β-glycosidase action are generally accepted (Sinnott, 1990). Controversies persist regarding certain aspects, e.g., the direction of the protonation by an enzyme electrophilic group and the possibility of substrate distortion (Franck, 1992).

For two cellulases of family B, a strict glycon specificity has been demonstrated (Fig. 2B). Both display a hydrolytic mechanism promoting inversion of configuration at the anomeric carbon (Table 1), compatible with a single displacement mechanism. The trioside is the minimum substrate, and the turnover numbers increase as a

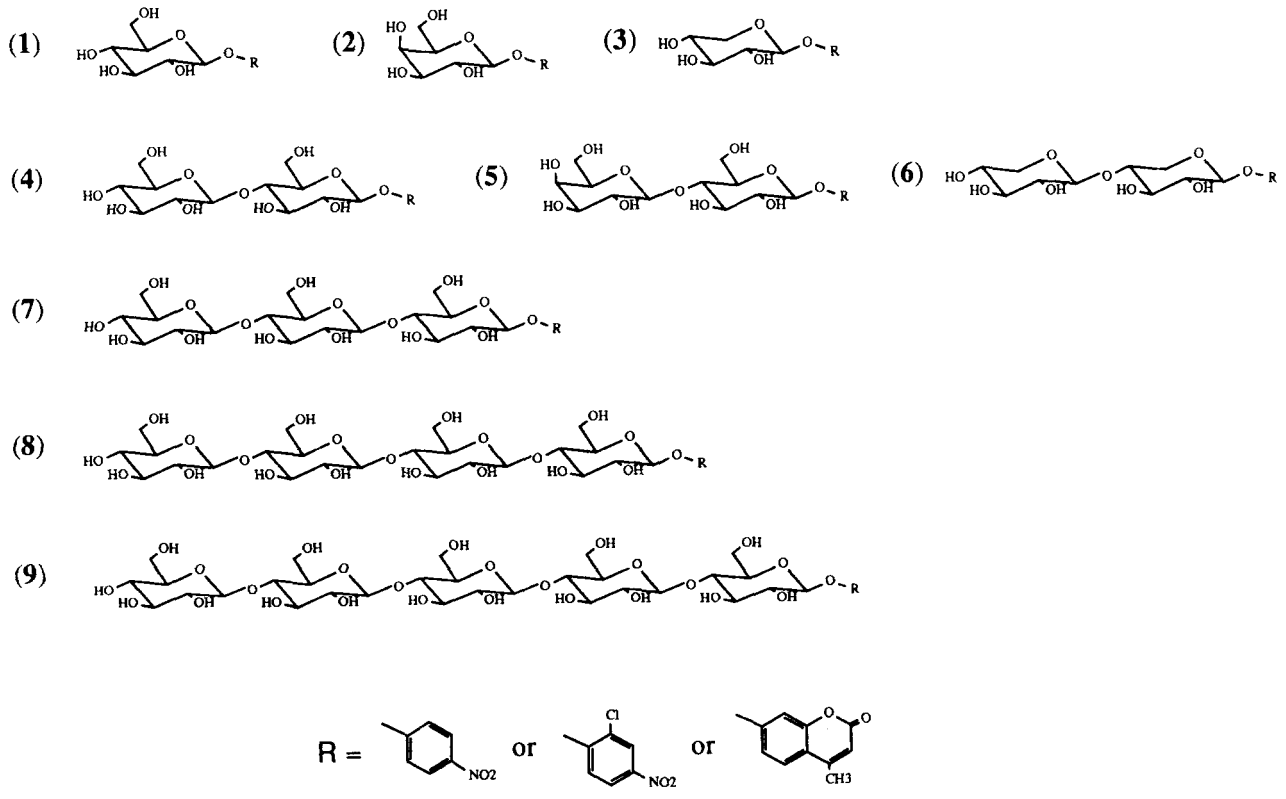

Fig. 1. Chemical structures of compounds used in the specificity study of β-glucanases. β-Glycosides of D-glucose (1), D-galactose (2), D-xylose (3), cellobiose (4), lactose (5), β(1,4)-xylobiose (6), cellotriose (7), cellotetraose (8), and cellopentaose (9). Structures of aglycons (R) as shown.

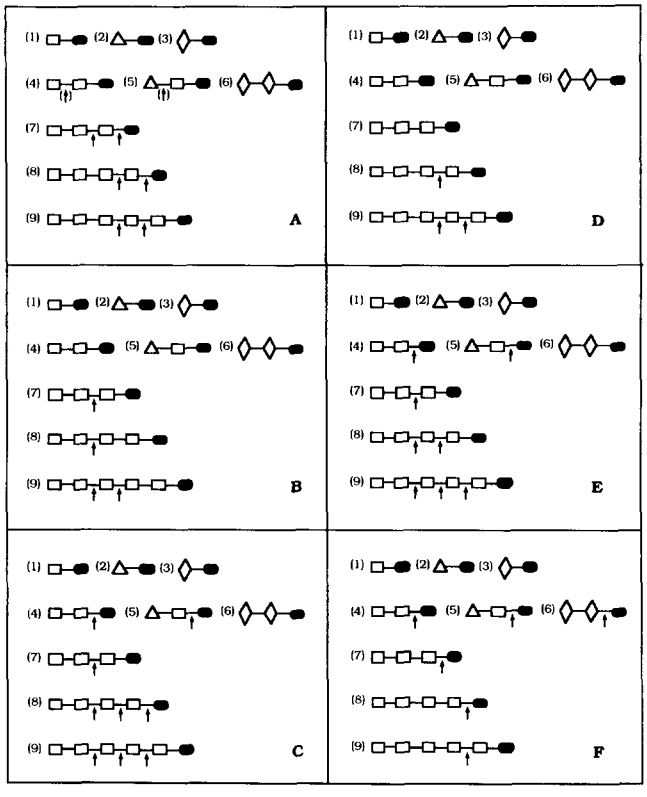

Fig. 2. Specificity mapping of β-glucanases with various chromophoric glycosides. Enzymes (families A–F) and substrates (1–9) used in this study are listed in Table 1 and Figure 1, respectively. Schematic representation of the glycosides; filled oval, chromophoric aglycon (R); open rectangle, β-1,4-linked D-glucose; open triangle, β-1,4-linked D-galactose; open diamond, β-1,4-linked D-xylose (see structures in Fig. 1). Arrows indicate observed hydrolysis sites.

function of chain length (van Tilbeurgh et al., 1988). Cellobiohydrolase II from *Trichoderma reesei* has been classified as an exoenzyme (Rouvinen et al., 1990), whereas endoglucanase A from *Cellulomonas fimi* is regarded as a typical endoglucanase (Gilkes et al., 1991a). Could discrete structural variations in active site architecture introduce important differences in specificity? Sequence alignment of the family B endoglucanases revealed several deletions in areas corresponding to extended surface loops in the parent exoglucanase (Rouvinen et al., 1990; Gilkes et al., 1991a). These loops form an integral part of the tunnel-like active site as evident in the resolved three-dimensional structure of cellobiohydrolase II core protein. Based on these observations it was suggested that, in contrast to the exoenzymes, endoglucanases lacking these loops exhibit "cleft"-like active sites such as those observed in lysozymes (Rouvinen et al., 1990). It is interesting to note that a recently crystallized endoglucanase (E2 from *Thermonospora fusca*), also belonging to family B, lacks some peptide sequences typical for *T. reesei* cellobiohydrolase II and can be differentiated from the latter enzyme by the absence of a tunnel-shaped active site and the presence of a cleft (Wilson, 1992). Thus, it seems that similarities in the sequences of the catalytic domains are paralleled by similarities in reaction mechanisms and hydrolysis patterns. On the other hand, endo vs. exo specificity could be induced by the presence, size, and properties of certain peripheral loops.

The specificities of two enzymes from *T. reesei* (cellobiohydrolase I and endoglucanase I), belonging to family C, have been extensively studied (van Tilbeurgh et al.,

Table 1. *Stereochemistry of the hydrolysis reaction catalyzed by 15 β-1,4-glycanases*

Family[a]	Enzyme	Source	Stereochemistry	References
A	CelB	*Clostridium thermocellum*	Retention	Gebler et al., 1992
	CelC	*C. thermocellum*	Retention	Gebler et al., 1992
	CelH	*C. thermocellum*	Retention	Gebler et al., 1992
	EG Z	*Erwinia chrysanthemi*	Retention	Barras et al., 1992
	EG III	*Trichoderma reesei*	Retention	Gebler et al., 1992
B	CBH II	*T. reesei*	Inversion	Knowles et al., 1988; Claeyssens et al., 1990a
	CenA	*Cellulomonas fimi*	Inversion	Withers et al., 1988
C	CBH I	*T. reesei*	Retention	Knowles et al., 1988; Claeyssens et al., 1990a
	EG I	*T. reesei*	Retention	Claeyssens et al., 1990b
	CBH I	*Phanerochaete chrysosporium*	n.d.	
D	CelA	*C. thermocellum*	n.d.	
E	CenB	*C. fimi*	Inversion	Meinke et al., 1991
	CelD	*C. thermocellum*	Inversion	Gebler et al., 1992
F	XynZ	*C. thermocellum*	Retention	Gebler et al., 1992
	Cex	*C. fimi*	Retention	Withers et al., 1988

[a] Families as defined by Henrissat et al. (1989); n.d., not determined.

1988; Claeyssens et al., 1990b). Both operate with retention of anomeric configuration (Table 1) and display similar hydrolytic patterns with small substrates (Fig. 2C), although with important differences in their kinetic parameters (van Tilbeurgh et al., 1988). Again, as in family B, both endoglucanases and cellobiohydrolases are present in family C, and the sequences of the latter are strikingly longer, especially at predicted loop locations (Henrissat et al., 1989).

Only one enzyme from family D could be examined (Fig. 2D). Its activity is restricted to higher cellodextrins, and, as in family A, cellotriose is preferentially released.

As in family C the two examined enzymes from family E show a more random degradation pattern (Fig. 2E). Both families can however be differentiated by their stereospecificity, the first operating with retention of configuration, the latter with inversion (Table 1).

The recently elucidated structure of CelD (Juy et al., 1992) shows a groovelike active site that could be typical for an endocellulase. This adds further evidence for a distinction in active site architecture for both enzyme types, one tunnel-shaped (exocellulases) and the other showing an open cleft (endocellulases). Other features could lead to a similar differentiation as further newly discovered structural details become available. It is conceivable that an endoglucanase can be turned into an exo-type enzyme by subtle structural changes such as obstructing the groove at one end.

The specificities of two enzymes from family F have been studied (Fig. 2F). They are not strictly specific for β-(1,4)-D-glucosidic bonds as they also hydrolyze β-(1,4)-D-xylosidic linkages (Grépinet et al., 1986; Gilkes et al., 1991a). In fact, the catalytic efficiency against xylobiosides was found to be 30–50-fold higher than those observed with most other substrates (Gilkes et al., 1991a), and therefore these enzymes should perhaps better be classified as xylanohydrolases. With the other substrates, the enzymes show a rather undiscriminating preference for the glycosidic bond carrying the chromophore, and turnover numbers seem unaffected by chain length (Claeyssens, unpubl.). Although this could be artifactual (induced by the chromophore) this behavior remains typical for family F. Retention of configuration in the reaction products was observed with the two enzymes studied (Table 1).

Thus, the classification of cellulases based on hydrophobic cluster analysis of amino acid sequences is corroborated by specificity mapping. This adds evidence to the presumption that each family of cellulases presents a typical fold and active site topology. To date, enzymes classified as cellobiohydrolases are found in three families (B, C, and F) always together with typical endoglucanases. The present work also demonstrates that for these families the small soluble substrates cannot be used to differentiate between endo- and exocellulases. On the other hand, the reaction mechanism (single or double displacement) is clearly conserved within each family.

Materials and methods

Endoglucanases I and III (EG I and EG III) from *T. reesei* were prepared by classical methods, whereas cellobiohydrolases I and II (CBH I and CBH II) from the same organism were purified by affinity chromatography (Tomme et al., 1988). Other enzymes (wild type or cloned) were gifts from several laboratories: CBH I from *Phanerochaete chrysosporium*, endoglucanases A, B, C, D, H, and xylanase Z (CelA, CelB, CelC, CelD, CelH, and XynZ) from *Clostridium thermocellum*; exoglucanase (Cex) and endoglucanases A and B (CenA and CenB) from *C. fimi*; and endoglucanase Z (EGZ) from *Erwinia chrysanthemi*. If necessary these were purified by the reported affinity procedure (Tomme et al., 1988). Purity was checked by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (PAGE) and isoelectric focusing–PAGE.

The substrates, either 4′-methylumbelliferyl-, 4′-nitrophenyl-, or 2′-chloro,4′-nitrophenyl-β-glycosides derived from D-glucose, cellobiose, higher cellooligosaccharides, D-xylose, and lactose were prepared by conventional methods (van Tilbeurgh et al., 1988). The β-(1,4)-xylobiose derivative was synthesized enzymatically (Claeyssens et al., 1966). Reactions were followed at 25 °C under optimal conditions of pH (5.0–7.3) and substrate concentration (0.1–5 mM) and the chromophoric products were analyzed by thin-layer chromatography (silica, MeOH/water, 80/20) or by high performance liquid chromatography (Polyolsilica 10 mm, Bio-Rad, acetonitrile/water, 75/25). Relative migration distances and retention times were determined with appropriate standards. Phenol concentrations could alternatively be determined by direct spectrophotometry (2-chloro,4-nitrophenol, 405 nm) or fluorimetry (4-methylumbelliferone, excitation 366 nm, emission >450 nm) (van Tilbeurgh et al., 1988).

Acknowledgments

We are grateful to Drs. F. Barras (Marseille), P. Béguin (Paris), N. Gilkes (Vancouver), and G. Pettersson (Uppsala) for gifts of enzyme samples. M.C. thanks the Région Rhône-Alpes for a fellowship.

References

Barras, F., Bortoli-German, I., Bauzan, M., Rouvier, J., Gey, C., Heyraud, A., & Henrissat, B. (1992). Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi*. *FEBS Lett. 300*, 145–148.

Béguin, P. (1990). Molecular biology of cellulose degradation. *Annu. Rev. Microbiol. 44*, 219–248.

Bhat, K.M., Hay, A.J., Claeyssens, M., & Wood, T.M. (1990). Study of the mode of action and site specificity of the endo-(1 → 4) β-D-glucanases of the fungus *Penicillium pinophilum* using normal, 1-H^3-labeled, reduced and chromogenic cellooligosaccharides. *Biochem. J. 266*, 371–376.

Chothia, C. & Lesk, A.M. (1986). The relation between the divergence of sequence and the structure in proteins. *EMBO J. 5*, 823–826.

Claeyssens, M., Tomme, P., Brewer, C.F., & Hehre, E.J. (1990a). Stereochemical course of hydrolysis and hydration reactions catalysed by cellobiohydrolases I and II from *Trichoderma reesei*. *FEBS Lett. 263*, 89–92.

Claeyssens, M., van Tilbeurgh, H., Kamerling, J.P., Berg, J., Vrsanska, M., & Biely, P. (1990b). Studies of the cellulolytic system of the filamentous fungus *Trichoderma reesei* QM 9414. Substrate specificity and transfer activity of endoglucanase I. *Biochem. J. 270*, 251–256.

Claeyssens, M., Van Leemputten, E., & De Bruyne, C.K. (1966). Transfer reactions catalysed by a fungal β-D-xylosidase: Enzymatic synthesis of phenyl-D-xylobiosides. *Carbohydr. Res. 3*, 32–37.

Franck, R.W. (1992). The mechanism of β-glycosidases: A reassessment of some seminal papers. *Bioorg. Chem. 20*, 77–88.

Gaboriaud, C., Bissery, V., Benchetrit, T., & Mornon, J.P. (1987). Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences. *FEBS Lett. 224*, 149–155.

Gebler, J., Gilkes, N.R., Claeyssens, M., Wilson, D.B., Béguin, P., Wakarchuk, W.W., Kilburn, D.G., Miller, R.C., Jr., Warren, R.A.J., & Withers, S.J. (1992). Stereospecific hydrolysis catalysed by related β-1,4 glucanases and β-1,4 xylanases, *J. Biol. Chem. 267*, 12559–12561.

Gilkes, N.R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H.D., Kilburn, D.G., Warren, R.A.J., & Miller, R.C. (1991a). Structural and functional relationships in two families of β-1,4-glycanases. *Eur. J. Biochem. 202*, 367–377.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., & Warren, R.A.J. (1991b). Domains in microbial β-1,4-glycanases: Sequence conservation, function and enzyme families. *Microbiol. Rev. 55*, 303–315.

Grépinet, O., Chebrou, M.C., & Béguin, P. (1986). Nucleotide sequence and deletion analysis of the xylanase gene *xynZ* of *Clostridium thermocellum*. *J. Bacteriol. 170*, 4582–4588.

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., & Mornon, J.P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene 81*, 83–95.

Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Béguin, P., & Aubert, J.-P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature 357*, 89–91.

Knowles, J.K.C., Lehtovaara, P., Murray, M., & Sinnott, M.L. (1988). Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. *J. Chem. Soc. Chem. Commun.*, 1401–1402.

Meinke, A., Braun, C., Gilkes, N.R., Kilburn, D.G., Miller, R.C., & Warren, R.A.J. (1991). Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. *J. Bacteriol. 173*, 308–314.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., & Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science 249*, 380–386.

Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev. 90*, 1171–1202.

Tomme, P., McCrae, S.I., Wood, T.M., & Claeyssens, M. (1988). Chromatographic separation of cellulolytic enzymes. *Methods Enzymol. 160*, 187–193.

van Tilbeurgh, H., Loontiens, F.G., De Bruyne, C.K., & Claeyssens, M. (1988). Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. *Methods Enzymol. 160*, 45–57.

Wilson, D.B. (1992). Biochemistry and genetics of actinomycete cellulases. *Crit. Rev. Biotechnol. 12*, 45–63.

Withers, S.G., Dombroski, D., Berven, L.A., Kilburn, D.G., Miller, R.C., Warren, R.A.J., & Gilkes, N.R. (1988). Direct ^{1}H NMR determination of the stereochemical course of hydrolysis catalyzed by glucanase components of the cellulase complex. *Biochem. Biophys. Res. Commun. 139*, 487–494.

Zvelebil, M.J.J.M. & Sternberg, M.J.E. (1988). Analysis and prediction of the location of catalytic residues in enzymes. *Protein Eng. 2*, 127–138.