

Chapter 9

Contribution of Domain Interactions and Calcium Binding to the Stability of Carbohydrate-Active Enzymes

LARS G. LJUNGAHL, IRINA A. KATAEVA, AND VLADIMIR N. UVERSKY

Production of ethanol from renewable plant biomass and its use as an alternative fuel or as additive to traditional fuels will partially contribute to the solution of the “global warming” problem. Biomass consists mainly of cellulose, hemicelluloses, and lignin. The conventional method to release the sugars in cellulose and hemicelluloses involves acidic hydrolysis. Recent research is focused on more “eco-friendly” enzymatic processes involving carbohydrate active enzymes that attack plant biomass efficiently, leading to high yields of sugars released not only from cellulose and hemicellulose, but also from pectin and other carbohydrate polymers present in plant cell walls. The sugars are then fermented to ethanol. In the industrial environment, strains of the yeast *Saccharomyces cerevisiae* have been used. Unfortunately, *S. cerevisiae* ferments only glucose, the product of cellulose hydrolysis. Thus, xylose and other sugars, derived from hemicelluloses and constituting a significant part of the biomass, are not fermented. Breakthroughs in fermentation technology in the past decade, i.e., the discovery and constructions of new yeast and bacterial strains with the ability to ferment the full spectrum of available sugars into ethanol, have led to the possible commercialization of biomass conversion technology (Kerr, 2004).

Industrial enzymatic hydrolysis of plant biomass is still more expensive than chemical hydrolysis. A search for cellulases and other carbohydrate-active enzymes with high specific activities and stabilities with regard to pH and temperature is under way. In addition, methods to recycle the enzymes are being investigated and if successful, will make industrial processes more productive and cost-effective. In this regard, the stability of the carbohydrate-active enzymes for use in bio-

reactors is one of the most important properties to be considered for scaling up the process.

Most carbohydrate-active enzymes are modular proteins, usually with calcium presented in some of the domains. The final fold, enzymatic properties, and stability of these enzymes are significantly affected by specific interactions between domains. Calcium plays a structural role and mediates proper intradomain interactions. In this chapter, we focus on the contributions that domain interactions and calcium have on the properties of carbohydrate-active enzymes.

PROTEIN DOMAINS: AN OVERVIEW

Many proteins in nature are constructed from a relatively small number of domains or modules. One domain definition reads: “Domains are topological entities which, at the atomic level, exhibit more pronounced interactions within the structural unit than with other parts of the polypeptide chain” (Janin and Wodak, 1983). Domains have an intrinsic capacity to form their native fold spontaneously and autonomously, often mediate specific biological functions, and combine to form larger multidomain proteins with segregated functions. A specific domain can be found in different proteins, and several different domains can be found within a given protein. Proteins thus can be viewed as being built of a finite set of domains, which are joined together in diverse combinations. Thus, domains may be considered basal units of the structure, function, and evolution of proteins.

Domains range in size from 25 to 500 amino acid residues. The robustness of the characteristic size of a domain suggests a simple underlying physical principle

Lars G. Ljungdahl • Department of Biochemistry and Molecular Biology, A216 Fred Davison Life Sciences Complex, University of Georgia, Athens, GA 30602. Irina A. Kataeva • Department of Biochemistry and Molecular Biology, B222A Fred Davison Life Sciences Complex, University of Georgia, Athens, GA 30602. Vladimir N. Uversky • Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 635 Narnhill Dr. MS 4021, Indianapolis, IN 46202-3763.

that is determined by only a few parameters (Shen et al., 2005). The optimal size of a globular protein domain, estimated by using a sphere-packing model, ranges from 117 to 213 residues, with an average of 165 residues. The model takes into account four parameters: (i) hydrophilic amino acid content, (ii) protein size, (iii) surface polarity, and (iv) eccentricity of a protein. These parameters are in good agreement with numerous experimental data (Shen et al., 2005). There are two points of view on folding of multidomain proteins: (i) simultaneous domain folding when a long polypeptide chain starts to fold in several places independently (Jaenicke, 1999), and/or (ii) sequential or cotranslational domain folding when domains fold one by one starting with an N-terminal domain (Frydman et al., 1999; Maity et al., 2005; Rumbley et al., 2001). At any rate, the modularity of proteins is a great achievement of evolution as it minimizes protein misfolding due to wrong intradomain interactions and provides biological systems with a convenient way of presenting binding sites on a stable protein scaffold, in the correct position of function; it also allows regulation by modular rearrangement (Campbell, 2003; Jaenicke, 1999).

Knowledge about modular proteins is increasing rapidly due to good databases and more-systematic approaches to protein expression and structure resolution. Several classification systems have attempted to combine domains sharing a similar sequence/fold. Some of the commonly used databases are InterPro, InterDom, ProDom, BLOCKS, Pfam, PRINTS, SBASE, SMART, TIGRFAM, and EVEREST. Studies show that the main task of structural biology is to move from determination of structures of individual modules to the task of assessing the ways by which these modules interact and bind their various ligands. Since modules occur in many proteins from various genomes, information about their structure and function can have an impact in a wide range of fields.

MODULARITY OF CARBOHYDRATE-ACTIVE ENZYMES AND DOMAIN CLASSIFICATION

Plant biomass is composed mainly of insoluble carbohydrates. Enzymes involved in the degradation of these compounds display modular architecture. In the simplest case, an enzyme is composed of only the catalytic module. However, in most cases, catalytic domains are surrounded by different noncatalytic ancillary modules, the total number of which within one protein can be eight and even more (Ahsan et al., 1996; Kataeva et al., 1999; Zverlov et al., 1998). The most common modules flanking catalytic domains from N or C termini are carbohydrate-binding mod-

ules (CBMs) to bring catalytic centers in close contact with the substrate. Other known domains are immunoglobulin (Ig)-like modules, fibronectin type 3 (Fn3)-like modules now designated X-modules, a second catalytic domain of similar or different activity, more than one CBM of similar or different binding affinity, and dockerin domains. The last named are found in catalytically active subunits of exocellular protein complexes called cellulosomes, which are produced by many anaerobic bacteria and fungi (Bayer et al., 2004; Doi and Kosugi, 2004; Ljungdahl et al., 2007). The function of the dockerins is to bind the catalytic subunits to special domains called cohesins of the special scaffolding proteins of the cellulosomes. Dockerin and cohesin domains are discussed in chapter 8 and also in reviews by Bayer et al. (2000, 2004). The biological function of some domains remains unknown (Bayer et al., 2000, 2004; Doi and Kosugi, 2004; Kataeva and Ljungdahl, 2003; Zverlov and Schwarz, 2007).

Several thousand sequences of carbohydrate-active enzymes have been described. For convenience, a classification system of catalytic domains of these enzymes based on amino acid sequence similarity was constructed in the early nineties (Henrissat, 1991). The classification system of catalytic and CBMs is available at the Carbohydrate-Active Enzymes (CAZy) server (<http://afmb.cnrs-mrs.fr/~cazy/CAZY>). The database includes glycoside hydrolases (GH) (108 families), glycosyltransferases (87 families), polysaccharide lyases (18 families), carbohydrate esterases (14 families), and CBMs (48 families). The family classification of catalytic modules and CBMs is expected to (i) aid in their identification, (ii) predict catalytic activity or binding specificity, (iii) aid in identifying functional residues, (iv) reveal evolutionary relationships, and (v) be predictive of polypeptide folds. The database is updated periodically and includes the GenBank/GenPept, Swiss-Prot, and PDB/3D entries. Information regarding catalytic domains, when they are known, includes the nature of residues participating in catalysis, molecular mechanism of the reaction, and availability of three-dimensional structure. Representatives of a particular family have highly conserved residues involved in catalysis, hydrolyze substrates by the same mechanism, and possess a similar fold allowing a homology modeling of known sequences (Henrissat and Davies, 2000). With a growing number of new sequences and three-dimensional protein structures, it has been noted that sometimes a new sequence shares similarity to sequences of different families, and vice versa, i.e., sequence-unrelated proteins display similar fold and mechanistic properties (Dominguez et al., 1995; Henrissat and Romeu, 1995). Such observations have led to the enzymes of some families being grouped into "clans."

DOMAIN INTERACTIONS IN CARBOHYDRATE-ACTIVE ENZYMES

Little is known about domain communications in carbohydrate-active enzymes, although the importance of such interactions has been demonstrated in various mammalian, fungal, bacterial, and viral proteins (Berr et al., 2000; Clout et al., 2000; Jaenicke, 1999; Smith et al., 2007; Wassenberg et al., 1997, 1999; Wenk and Jaenicke, 1998, 1999). It has been noted that modules in carbohydrate-active enzymes are not randomly combined. They often are located in a particular place within a protein and are associated with a particular type of neighboring module(s). Thus, CBMs of a particular family have a tendency to be associated with catalytic domains of a particular family and to be located in a specific place on the polypeptide chain (Tomme et al., 1995, 1998). Ig-like domains often, but not always, precede catalytic domains belonging to family GH9 (Bayer et al., 2000; Zverlov and Schwarz, 2007). Fn3-like domains, also designated X1, are common in microbial chitinases, often are duplicated or triplicated, and are located between catalytic modules and CBMs (Little et al., 1994; Shen et al., 1995). Surface layer homologous domains are often in duplicate and triplicate and located at C termini (Ohmiya et al., 1997). Dockerin domains are usually located at the C termini of polypeptide chains (Bayer et al., 2000, 2004). It has been demonstrated that deletion of a CBM from a catalytic domain drastically decreases catalytic activity against insoluble substrates (Bolam et al., 1998; Kataeva et al., 1999, 2001a, 2001b; Srisodsuk et al., 1993) and, in some cases, the thermostability of the residual protein (Araki et al., 2006; Fontes et al., 1995; Hayashi et al., 1997; Kataeva et al., 1999, 2005). Fusion of catalytic and CBM domains of different origins decreased the thermostability of the chimeric polypeptide. Attachment of an N-terminal CBM to the C terminus of the catalytic domain led to a decrease in binding affinity, enzymatic capacity, and thermostability (Kataeva et al., 2001a, 2001b). Transposition of the domains within the polypeptide chain resulted in a drop of thermostability (Shin et al., 2002). These observations imply the existence of specific interactions between domains in the carbohydrate-active enzymes similar to those found in some other modular proteins.

ROLE OF LINKERS LOCATED BETWEEN FUNCTIONAL DOMAINS

Many carbohydrate-active enzymes contain linkers of different length, which are rich in hydroxylic amino acid residues (threonine, serine, and proline), two of which, Ser and Pro, are disorder-promoting residues

(Dunker et al., 2001). Linker sequences are often highly O glycosylated (Gilkes et al., 1991). It has been shown that a deletion of a 23-residue linker separating the catalytic module from the CBM in CenC from *Cellulomonas fimi* changed the relative orientation of these domains and affected both catalytic activity and desorption from cellulose, although it did not affect the overall fold of each domain (Shen et al., 1991). Similarly, the deletion of part of the linker located between the catalytic module and CBM in CBHI from *Trichoderma reesei* had a serious effect on binding and activity against crystalline cellulose (Srisodsuk et al., 1993). These findings imply that linker sequences serve to set proper interactions between domains. As the presence of linkers negatively affects crystallization of modular proteins, only a limited number of structures of domain combinations have been resolved so far. As a result, although structural and functional properties of many individual domains are well studied, very little is known about the role and structural aspects of the linker peptide.

More information about conformational versatility of linkers in carbohydrate-active proteins was obtained by exploring a small-angle X-ray scattering technique (Receveur et al., 2002; von Ossowski et al., 2005). Cel45 from *Humicola insolens* is composed of catalytic domain GH45 and a CBM1 separated by a 36-residue glycosylated linker. Combination of light-scattering data with a known crystal structure of the GH45 catalytic domain and a modeled structure of CBM1 (based on a 45% identity to Cel7A CBM1) helped to evaluate hydrodynamic dimensions of the enzyme. The average size of a protein is estimated by measuring its radius of gyration (R_g). The R_g of Cel45 was much higher than what would be expected for a spherical protein with a similar number of residues. Cel45 had an elongated shape in solution despite the fact that its large catalytic domain is a globular polypeptide and that a small-size CBM1 did not greatly affect R_g . The data showed that the dimension of the linker is very large with respect to its mass. As R_g reflects a mixture of all possible conformations the protein can adopt in solution, one might conclude that the linker has some internal flexibility.

Further progress in understanding the role of linkers in carbohydrate-active enzymes was made by studying hydrodynamic properties of the chimeric fungal cellulase Cel6AB (von Ossowski et al., 2005). Wild-type Cel6A and Cel6B are both domain proteins. Cel6A contains an N-terminal CBM followed by a 52-residue linker and a C-terminal catalytic module (CBM-L₅₂-GH6_A). Cel6B is composed of an N-terminal catalytic module and a C-terminal CBM separated by a 36-residue linker (GH6_B-L₃₆-CBM). In wild-type proteins linkers and CBDs are small and they are difficult to

distinguish clearly. To further evaluate the linkers and CBMs, the CBMs were removed and the catalytic modules with natural linkers were fused together so that the two linkers formed a combined linker of 88 residues in total length located between the catalytic modules (GH_{6B}-L₃₆-L₅₂-GH_{6A}). The extended conformation of the linker within the chimeric construct was confirmed by measuring the R_g value and the distance distribution function $P(r)$. The shape of the experimental distance distribution profile $P(r)$ indicated that there is a distribution of conformations with various distances between the GH_{6A} and GH_{6B} domains. Interestingly, molecular modeling of the linker length in combination with the small-angle X-ray scattering data implied that the linker is not just an extended polypeptide chain; it adopts a much more compact conformation, which is the most stable. It is, however, with a relatively low energy cost able to unwind, forming a stretched-out longer peptide. This suggests that the linker may function as a "molecular spring" between the two functional domains. With regard to cellulose hydrolysis, the linker would thus give the catalytic site the ability to reach out and hydrolyze new glycosidic bonds while the CBM is still bound to the substrate surface (Receveur et al., 2002; von Ossowski et al., 2005).

DOMAINS IN THE CLOSTRIDIUM THERMOCELLUM CELLULOSOME

Clostridium thermocellum, a thermophilic, anaerobic bacterium, produces an exocellular multiprotein complex cellulosome highly active against plant cell wall carbohydrates (Bayer et al., 2000). Besides the cellulosome, the bacterium synthesizes several noncellulosomal enzymes with different hydrolytic activities (Zverlov and Schwarz, 2007). The assembly of the cellulosome occurs by specific interactions between cohesin domains of scaffoldin protein CipA and dockerin domains (DD) of the enzymes (Bayer et al., 2000; Béguin and Lemaire, 1996). Analysis of the whole genome of *C. thermocellum* revealed over 70 putative open reading frames bearing DD and several other components without dockerins (Zverlov and Schwarz, 2007). Below we focus on the interactions between domains in one of the largest cellulosomal catalytic components, cellobiohydrolase A (CbhA) (Schubot et al., 2004; Zverlov et al., 1998). It has recently been shown to also possess endoglucanase activity (McGrath, 2007). This enzyme is a thermostable, multimodular, calcium-containing protein. It has a molecular mass of 135 kDa and is composed of seven modules: an N-terminal CBM4, an Ig-like domain, a GH9 domain, two Fn3-like domains (X1

and X1₂), a CBM3, and a DD (Fig. 1). All modules, except the Ig-like domain, bind calcium. One major role of calcium is to stabilize native folds of polypeptides. For this reason, calcium is a constituent of many thermostable proteins (Medved et al., 1995; Notenboom et al., 2001; Wenk and Jaenicke, 1998, 1999). CbhA is the only *C. thermocellum* cellulosomal enzyme whose domain interactions and role of calcium have been studied in detail by different techniques including genetic manipulations, crystallography, circular dichroism (CD) spectroscopy, and differential scanning calorimetry (DSC).

INTERACTIONS BETWEEN THREE INTERNAL CbhA MODULES PROBED BY CD SPECTROSCOPY

Three domains of CbhA, X1₁, X1₂, and CBM3, each bind 1 mol of calcium (Fig. 1). Expressed either individually or in combinations X1₁X1₂ and X1₁X1₂-CBM3, all variants, when saturated with calcium, are designated holoproteins. When calcium is reversibly removed by incubating the proteins with Chelex-100, they are designated apoproteins. Thermal unfolding of all variants was totally reversible (Kataeva et al., 2003). The recorded near-UV CD spectra of domain combinations were compared with the spectra calculated as simple weighted sums of X1₁ + X1₂ (X1₁X1₂ construct) and X1₁ + X1₂ + CBM3 or X1₁X1₂ + CBM3 (X1₁X1₂-CBD3 construct) (Fig. 2). In the case of total domain independency, the recorded spectra of X1₁X1₂ and X1₁X1₂-CBD3 were expected to be equal to the spectra calculated as weighted sums of the recorded spectra of individual domains. When domain interactions induced changes in the tertiary structure, differences between the observed and calculated spectra were expected. Figure 2 demonstrates that domain interactions play an important role in the conformation of X1₁X1₂-CBM3 either in the presence (Fig. 2B) or in the absence (Fig. 2D) of Ca²⁺. As for the X1₁X1₂ construct, a considerable difference between recorded and calculated spectra was observed in the absence of Ca²⁺ (Fig. 2C), whereas upon calcium binding, this difference was considerably reduced or eliminated (Fig. 2A). Thus, (i) the three domains interact, and (ii) Ca²⁺ is involved as mediator of these interdomain interactions (Kataeva et al., 2003).

To gain insight into the mechanisms of thermal denaturation of individual and linked domains, phase diagram (PD) analyses were performed with spectroscopic data. Such an analysis is extremely sensitive for the detection of intermediate states (Kuznetsova et al., 2002). The essence of this method is to build up the diagram of *L* versus *T* where *L* and *T* are the

tral intensity values measured at wavelength λ_1 and λ_2 , under different experimental conditions for a protein undergoing structural transformations. The relationship $I_{\lambda_1} = f[I_{\lambda_2}]$ is linear if changes in protein environment lead to an all-or-nothing transition between two different conformations. Alternatively, nonlinearity of this function reflects the sequential character of structural transformations; and each linear portion of the $I_{\lambda_1} = f[I_{\lambda_2}]$ dependence describes an individual all-or-nothing transition. Phase diagrams of individual domains and domain combinations are given in Fig. 3. Figure 3A shows that the PD of holo- $X1_1$ consists of two linear parts. Removal of Ca^{2+} dramatically changes the shape of the PD, although the plot still has two linear parts indicating the existence of two independent transitions. Figure 3B shows that the denaturation of holo- and apo- $X1_2$ most likely represents an all-or-nothing transition. The denaturation of CBM3 is Ca^{2+} sensitive (Fig. 4C). Thus, apo-CBM3 denatures according to the three-state model (the PD has two linear parts), whereas holo-CBM3 denatures as an all-or-nothing transition. Finally, panels D and E of Fig. 3 show that removal of calcium has no effect on the mechanisms of $X1_1X1_2$ and $X1_1X1_2$ -CBM3 melting, respectively, with the apo- and

holoforms being denatured by the two-state scheme. Thermodynamic analysis proved the assumption that the stability of individual $X1_1$ is Ca^{2+} dependent. Removal of calcium resulted in a 4.2 kcal/mol decrease in ΔG value and a 20°C decrease of T_m of the domain (82°C versus 61.8°C, respectively). In contrast, the individual $X1_2$ domain and the $X1_1X1_2$ domain combination are relatively calcium independent with T_m s of 75.8 and 78.3°C ($X1_2$ apo and holo) and 78.7 and 83.0°C ($X1_1X1_2$ apo and holo), respectively. This leads to the conclusion that the individual $X1_1$ is stabilized by calcium whereas in the $X1_1X1_2$ and $X1_1X1_2$ -CBM3 constructs, the stabilization is by interaction of $X1_1$ with $X1_2$. Finally, in the presence of calcium, the stabilities of the domains are relatively independent, while in the absence of Ca^{2+} , domain interactions play a stabilizing role (Kataeva et al., 2003).

DOMAIN COUPLING IN CbhA PROBED BY DSC

Comparison of thermal denaturation of truncated variants of CbhA of increasing length as detected by DSC yields a good illustration of domain coupling (Fig.

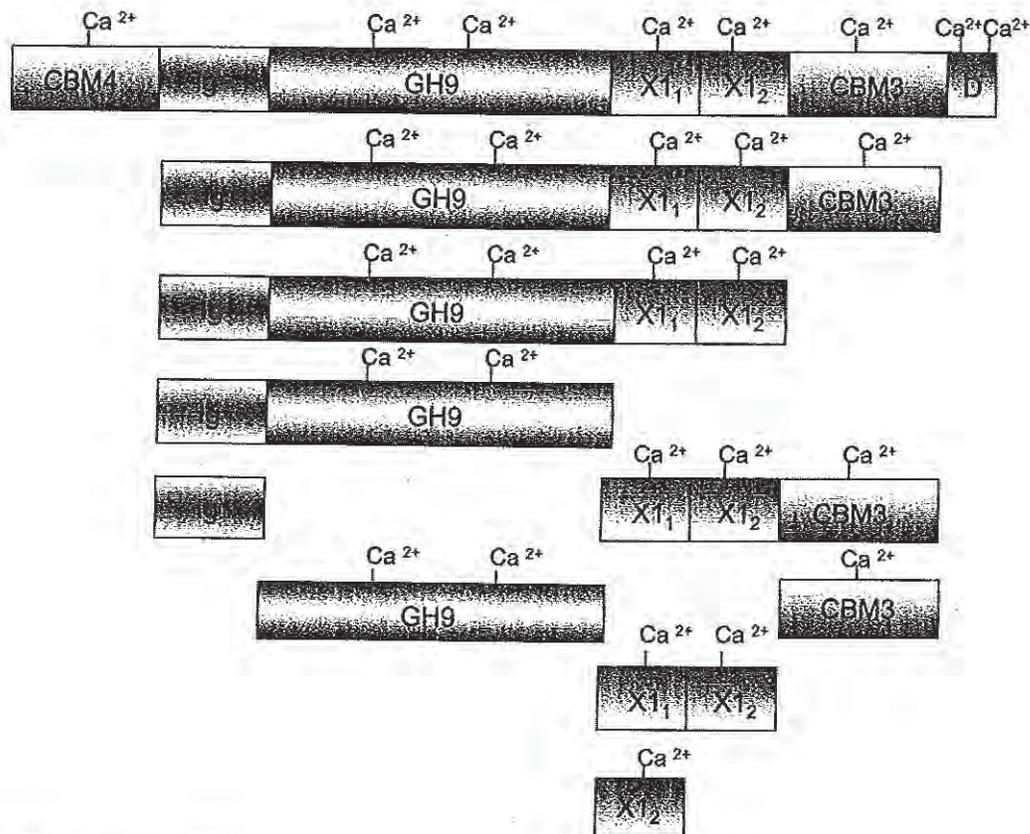


Figure 1. Domain structure of CbhA and its truncated variants. Abbreviations: CBD4 and CBD3, carbohydrate-binding domains of family 4 and 3, respectively; Ig, immunoglobulin-like domain; GH9, catalytic domain of family 9 glycoside hydrolases; $X1_1$ and $X1_2$, X domains of family 1; DD, duplicated dockerin domain. The content of calcium is also shown. (From Kataeva et al., 2005.)

4) (Kataeva et al., 2005). As all domains except the Ig-like domain bind calcium (Fig. 1) (Béguin and Alzari, 1998; Henrissat and Romeu, 1995), the thermal denaturation was studied (i) with 20 mM sodium phosphate buffer, pH 6.0; (ii) with 2 mM calcium; and also (iii) by heating with 2 mM EDTA. Individual domains possessed relatively independent folding since they unfolded as cooperative units (Fig. 4A, C, E, F). As expected, the Ig-like domain was insensitive to Ca^{2+} and EDTA (Fig. 4E). The effects of calcium and EDTA on thermal denaturation of the other proteins can be summarized as follows: (i) in buffer (i.e., only endogenous calcium present), all proteins except CBM3 and $\text{X1}_1\text{X1}_2$ -CBM3 denatured less cooperatively with lower mid-point temperature (T_d) and higher number of cooperative units (r) than in the presence of Ca^{2+} ; (ii) calcium increased T_d and the cooperativeness of unfolding; (iii) EDTA decreased T_d and, in most cases, r values. In other words, Ca^{2+} significantly stabilizes the proteins. As in buffer, all constructs contain Ca^{2+} , and the stabilizing effect of external calcium occurs due to the shift of the dissociation equilibrium towards associ-

ation of Ca^{2+} and proteins. The higher cooperativeness of thermal unfolding in the presence of EDTA in comparison to that in buffer is a result of lower heterogeneity of the protein population. Thermal unfolding in buffer assumes coexistence of protein molecules (i) with calcium bound to all binding sites, (ii) with partially lost calcium, and (iii) without calcium. In the presence of EDTA, only apoprotein molecules are present.

Measured calorimetric denaturation enthalpy ($\Delta_d H_{\text{cal}}$) of domain combinations was higher than the sum of $\Delta_d H_{\text{cal}}$ s of individual domains (Table 1), indicating the importance of domain communications. The number of cooperative units in the system r , calculated as $\Delta_d H_{\text{cal}}/\Delta_d H_{\text{v.H}}$ is often used as a measure of cooperativeness of thermal transitions (Kozhevnikov et al., 2001). In an ideal case, the polypeptide is one cooperative unit, with an r value of 1. In Table 1 the r values deviate from 1. For individual domains (X1_2 , Ig-like, and GH9), this means that although they preserved a relatively independent fold, the cooperativeness of unfolding is low. Upon gradual association of domains (X1_1 , X1_1 - X1_2 , and $\text{X1}_1\text{X1}_2$ -CBM3), the r value de-

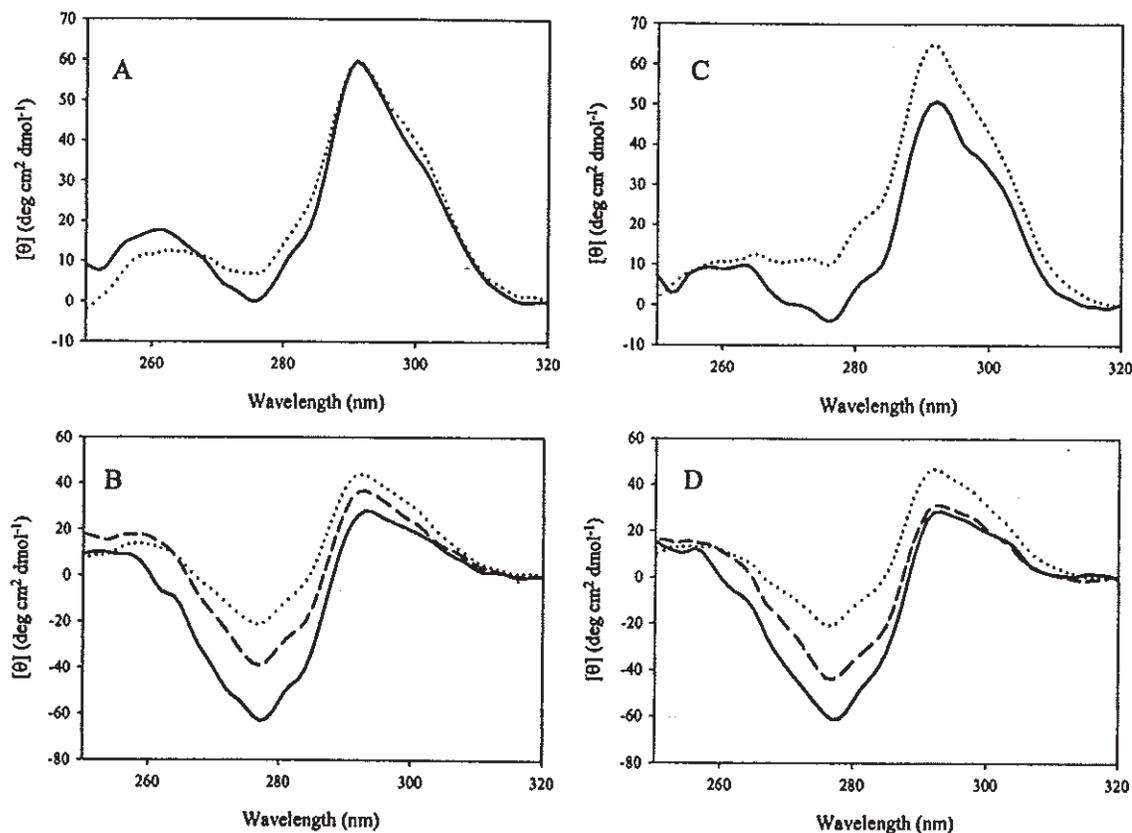


Figure 2. Comparison of near-UV CD spectra recorded at 25°C of $\text{Fn3}_{1,2}$ in the presence (A) and absence (C) of calcium and of $\text{Fn3}_{1,2}$ -CBM3 in the presence (B) and absence (D) of calcium to the spectra calculated as simple weighted sums based on spectra recorded for the individual domains: $(\text{Fn3}_1 + \text{Fn3}_2)/2$ (A and C, dotted lines) and $(\text{Fn3}_1 + \text{Fn3}_2 + \text{CBM3})/3$ (B and D, dashed lines) or $(\text{Fn3}_{1,2} + \text{CBM3})/2$ (B and D, dotted lines). Experimental spectra of the domain combinations $\text{Fn3}_{1,2}$ and $\text{Fn3}_{1,2}$ -CBM3 are shown with solid lines. (From Kataeva et al., 2003.)

creases and reaches 1 in the three-domain construct which unfolds as one cooperative unit. The large GH9 domain is the least-stable construct, with a highly asymmetric thermogram (Fig. 4; Table 1), r value of 2.8, and the lowest T_d (68.3°C). Addition of the Ig-like domain decreased the r value in buffer to 1.83. Addition of other domains to GH9 led to an increase of cooperativeness of thermal unfolding. Binding calcium further increased the cooperativity of thermal denaturation.

Thus, in a four- and a five-domain construct, the r values in the presence of calcium are 1.12 and 1.86, respectively. In other words, these multidomain polypeptides thermally unfold by a more simple mechanism than expected from their domain architecture. This phenomenon, first observed in some proteins from thermophiles and called "domain coupling" (Wassenberg et al., 1999), is clear evidence of domain interactions.

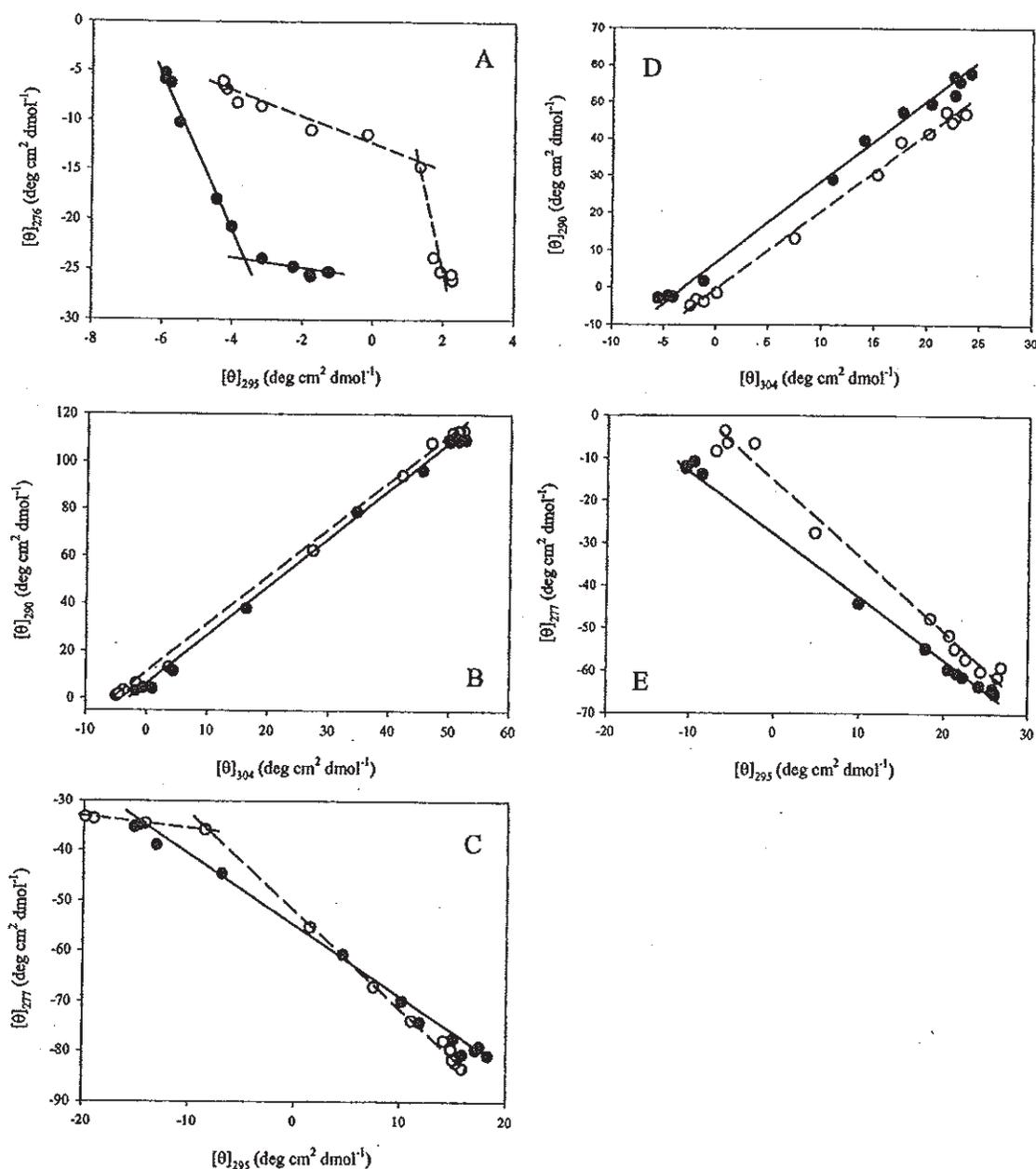


Figure 3. Phase diagrams based on $[\theta]_{\lambda_1}$ versus $[\theta]_{\lambda_2}$ (see the text for details) characterize heat-induced denaturation of different domains of CbhA, based on the temperature-induced changes in the near-UV CD spectra of individual domains Fn3₁ (A), Fn3₂ (B), and CBM3 (C) and of domain combinations Fn3_{1,2} (D) and Fn3_{1,2}-CBM3 (E). Data for holo- and apo-proteins are given with closed symbols and solid lines, and open symbols and dashed lines, respectively. (From Kataeva et al., 2003.)

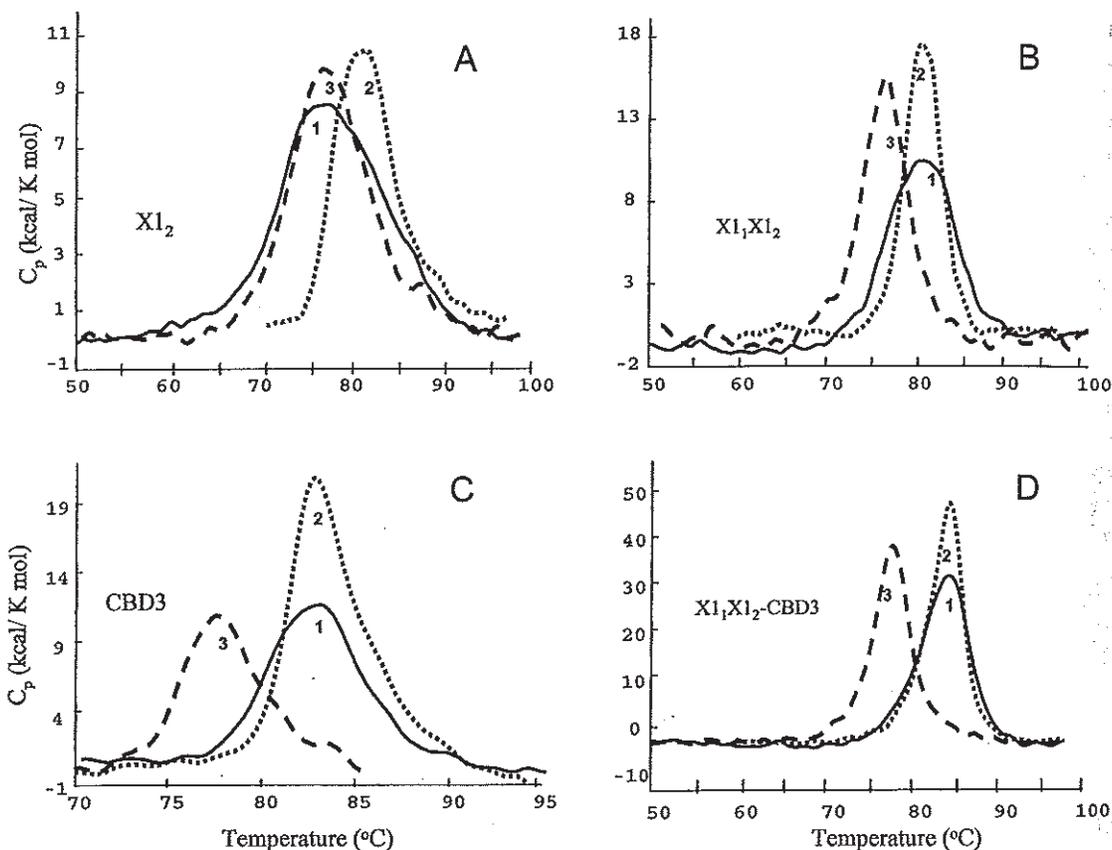


Figure 4. Denaturation peaks obtained for different constructs of CbhA in 20 mM sodium-phosphate buffer, pH 6.0 (A through D) and in the presence of 2 mM Ca^{2+} as well (E through H) or in the presence of 2 mM EDTA (I). (From Kataeva et al., 2005.)

SPECIFICITY OF DOMAIN INTERACTIONS: Ig-GH9 DOMAIN PAIR

The *C. thermocellum* genome encodes several cellulosomal subunits (proteins containing DD) with catalytic domains belonging to family GH9. Based on different arrangements around GH9, these enzymes have been divided into two groups: (i) a GH9 domain is N terminal and followed by other domains; (ii) a GH9 domain is located internally and is preceded by an Ig-like domain or CBM-Ig-like domain pair (Zverlov and Schwarz, 2007). The biological role of the Ig-like domain in plant cell wall active enzymes remains unknown. The enzymes originally lacking the Ig-like domain are enzymatically active (Mandelman et al., 2003; Parsiegla et al., 2002; Sakon et al., 1997). Surprisingly, the deletion of the Ig-like domain from the GH9s of Cel9A and CbhA, both belonging to group 2, resulted in a complete loss of activity (Béguin and Alzari, 1998; Juy et al., 1992; Kataeva et al., 2004). In both enzymes, the characteristic linker sequence between the Ig-like domain and the GH9 is missing. This leads to the possibility of a tight interaction between the two domains, and it has been assumed that the inactivation of the GH9 domain upon deletion

of the Ig-like domain occurs as a result of breakage of multiple bonding stabilizing the fold of GH9. Indeed, in individually expressed, both domains were unstable units of low cooperativity (Fig. 3) (Kataeva et al., 2004). Analysis of the crystal structures of the domain pair Cel9A and CbhA revealed an extensive domain interface composed of over 40 amino acid residues from both domains involved in numerous hydrophobic and hydrophilic interactions (Kataeva et al., 2003, 2004; Schubert et al., 2004). Comparison of the interfaces between the two domains revealed that 3 of 10 hydrogen bonding pairs were conserved in both proteins, suggesting that they play an important role in the maintenance of the stable domain interaction and affect the overall fold of the combined Ig-like and GH9 domains. Thr230 of the Ig-like domain and Gly221 of the GH9 domain formed one of these pairs. This interaction helped to stabilize an otherwise flexible loop, which in turn interacts with another loop that is part of the catalytic domain. Asp262 and Asp264 of the Ig-like domain form two other conserved hydrogen bonds with Gly221 and Tyr676, respectively, of the GH9 domain. To alter the H-bonding network between the domains and evaluate the importance of domain interactions in the domain pair, these residues

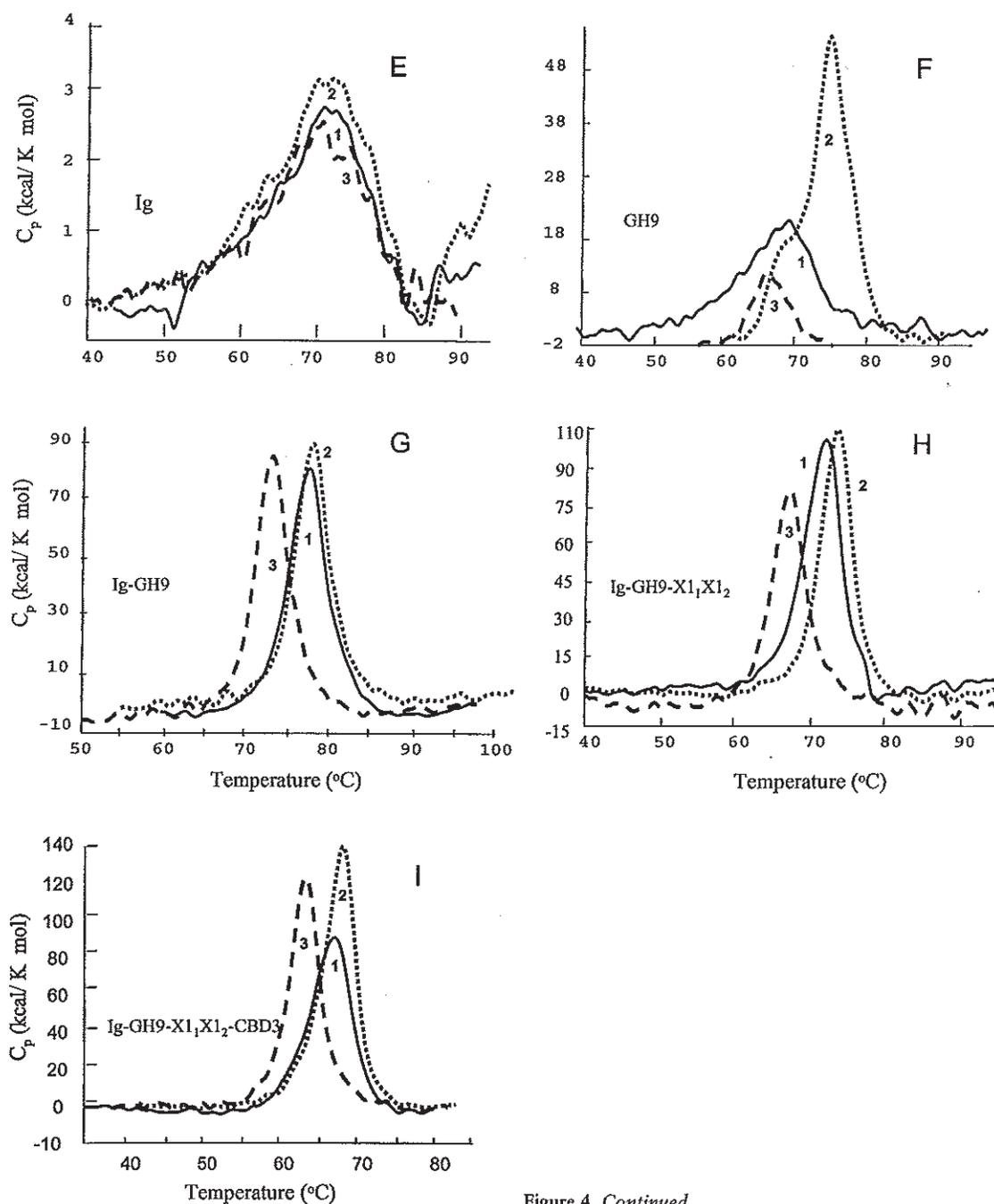


Figure 4. Continued

were replaced with alanine residues, giving a double mutant, T230A/D262A, and a single mutant, D264A. Figure 5 shows thermograms of the Ig-GH9 domain pair and its T230A/D262A and D264A mutants. Both mutants significantly differed from the original Ig-GH9 by their denaturation temperatures (T_d) and calorimetric enthalpies ($\Delta_d H_{cal}$). The T_d values of Ig-GH9, T264A, and T230A/D262A were 77.9, 71.9, and 69.2°C, respectively; and the $\Delta_d H_{cal}$ values were 530.5, 469.8, and 296.4 kcal/mol⁻¹, respectively. Enzymatic properties of

the mutants were comparable with those of the Ig-GH9. The values of K_M and k_{cat} for Ig-GH9, D264A, and T230A/D262A were 2.0 ± 0.15 , 2.3 ± 2.25 , and 2.0 ± 1.73 mM and 18.4 ± 2.11 , 16.7 ± 2.42 , and 19.0 ± 2.35 s⁻¹, respectively. Correspondingly, the efficiencies of catalysis were similar. In summary, mutation of two and even one residue from the domain interface significantly destabilized the combined Ig-GH9 protein structure. The data imply that the domain interactions are very specific and cannot be transferred in artificial fusion

Table 1. Denaturation parameters of CbhA individual domains and their combinations^e

Protein	Mol mass (kDa)	Solvent	T_d^a (°C)	$\Delta_d H_{cal}^b$ (kcal mol ⁻¹)	$\Sigma \Delta_d H_{cal}^c$ (kcal mol ⁻¹)	r^d ($\Delta_d H_{cal} / \Delta_d H^{HH}$)
X1 ₂	9.4	Buffer	77.6	110.9		1.72
		CaCl ₂	81.3	131.8		1.26
		EDTA	76.1	106.9		1.09
X1 ₁ X1 ₂	19.6	Buffer	78.7	126.9		1.31
		CaCl ₂	80.5	131.2		1.03
		EDTA	75.2	127.2		0.93
CBD3	16.4	Buffer	82.7	85.5		0.59
		CaCl ₂	82.8	116.2		0.71
		EDTA	77.2	101.6		0.74
X1 ₁ X1 ₂ -CBD3	36.5	Buffer	83.7	252.0	212.4	0.99
		CaCl ₂	84.5	259.5	249.4	1.04
		EDTA	77.6	261.8	228.8	0.84
Ig	11.7	Buffer	73.6	66.7		1.41
		CaCl ₂	72.5	75.9		1.37
		EDTA	73.8	65.7		1.29
GH9	56.9	Buffer	68.3	279.8		2.80
		CaCl ₂	74.2	353.9		1.71
		EDTA	64.2	135.4		0.91
Ig-GH9	68.5	Buffer	77.9	565.6	346.5	1.83
		CaCl ₂	78.4	596.2	429.8	1.76
		EDTA	73.1	526.1	201.1	2.11
Ig-GH9-X1 ₁ X1 ₂	88.4	Buffer	77.7	593.5	473.4	2.28
		CaCl ₂	79.0	666.9	561.0	1.12
		EDTA	75.7	625.5	328.3	2.81
Ig-GH9-X1 ₁ X1 ₂ -CBD3	105.3	Buffer	82.3	797.3	558.9	2.14
		CaCl ₂	83.1	835.0	679.2	1.86
		EDTA	77.4	734.5	429.9	2.90

^aDenaturation temperature.^bCalorimetric denaturation enthalpy.^cSum of calorimetric denaturation enthalpies of individual domains.^dNumber of cooperative units.^eFrom Kataeva et al., 2005.

constructs (Kataeva et al., 2001a, 2001b). Finally, the fact that the addition of each next domain to the C terminus of the existing construct stabilized its structure indicates that at least CbhA is folding sequentially and each preexisting construct acts as a chaperone assisting correct folding of each C-terminal fused domain.

BIOLOGICAL FUNCTIONS OF INTERNAL MODULES

Although it is clear that the Ig domain stabilizes the GH9 structure of CbhA, where they are tightly bound to each other, and appears to retain the configuration of the active site of the GH9, allowing it to be enzymatically active (Kataeva et al., 2004; Schubot et al., 2004), not much is known about the biological activity of Ig-like modules in other enzymes.

CBMs are very important modules in carbohydrate-active enzymes. They bind to the insoluble substrates and bring catalytic domains in close contact with the substrate to hydrolyze it. The CBMs differ by binding specificity (Tomme et al., 1995). In particular, some of them bind amorphous insoluble substrates like cellulose or xylan (Brun et al., 2000; Johnson et al., 1996). Some of the CBMs binding amorphous insoluble carbohydrates also bind some soluble substrates such as soluble xylans, lichenan, laminarin, beta-glucan, mannans, or carboxymethyl cellulose (Arai et al., 2003; Kataeva et al., 2001a, 2001b). Other CBMs have the ability to bind highly crystalline recalcitrant substrates (Tomme et al., 1996). It has been shown that the binding specificity of CBMs depends on the topology of the binding surface. In those CBMs binding amorphous soluble and insoluble substrates, the binding surface is a cleft accommodating a single carbohydrate chain (Brun et al.,

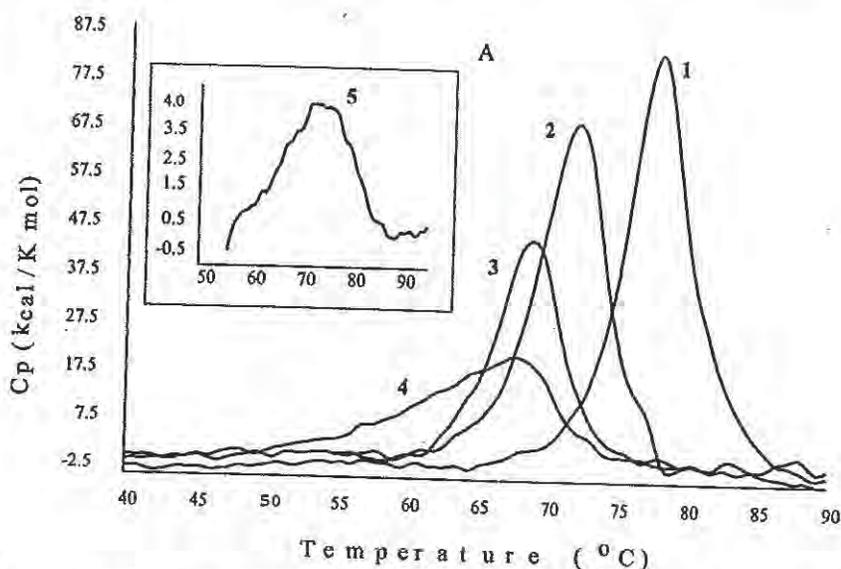


Figure 5. DSC thermograms of Ig-GH9 module pair (1), its D264A (2) and T230/D262A (3) mutants, and individual GH9 module (4) and Ig-like module (5). The protein concentrations and scan rate were 6 mg/ml and 60°C/h, respectively. All thermal transitions were completely irreversible, so that second scans of the proteins were used as baselines. (From Kataeva et al., 2004.)

2000; Johnson et al., 1996). The CBMs binding to the crystalline cellulose with carbohydrate chains connected by numerous hydrogen bonds have flat binding surfaces fitting very well the surface of the crystalline carbohydrate (Tomme et al., 1996). In some cases, the carbohydrate-active enzyme might contain more than one CBM with different binding specificities, the presence of which increases the hydrolytic potential of the enzyme.

Another interesting group of domains classified based on sequence similarity into family "X1" (B. Henrissat, personal communication) are less common. They are present in some bacterial chitinases and a few cellulases (Kataeva et al., 2001a, 2001b). These domains have a beta-sandwich fold typical of representatives of the Ig superfamily. This superfamily incorporates proteins involved in different binding functions. The duplicated version of the X1 domains, X1₁ and X1₂, is present in CbhA (Fig. 1). It has been demonstrated that the X1₁X1₂ domain pair of CbhA is involved in a nonhydrolytic loosening of the cellulose surface, thus enhancing catalytic activity of CbhA against insoluble cellulose. This effect is even more pronounced when the X1₁X1₂ domain pair is attached to the CBM3 in the three-domain construct X1₁X1₂-CBM3 (Kataeva et al., 2002). It is still unclear if this function is general for all X1 domains.

CONCLUSIONS

This chapter describes modular architectures of carbohydrate-active enzymes and analyzes the role of interdomain interactions in the structure, stability, and functionality of these interesting and important pro-

teins. Although the majority of these proteins are complex machines, in which catalytic domains are surrounded by different noncatalytic ancillary modules, these modules are not randomly combined. They are located in a particular place within a protein often being associated with a particular type of neighboring modules. Modules vary in length and possess diverse functions. They are involved in complex communications and interactions. Entire proteins are greatly stabilized via a series of specific interdomain interactions. Structures and communications between modules in large cellulosomal catalytic enzymes like CbhA from *C. thermocellum* are modulated by calcium binding. The complex nature of interdomain interactions and cross talking has been confirmed by structural and conformational analyses. Even linkers between domains are crucial for the functionality of carbohydrate-active enzyme in that they serve as "molecular springs" allowing catalytic sites to reach and hydrolyze new glycosidic bonds, while the CBM is still bound to the substrate surface. As a general conclusion one may say that domain interactions in modular carbohydrate hydrolytic enzymes enhance the activity of the catalytic domains of these enzymes.

REFERENCES

- Ahsan, M. M., T. Kimura, S. Karita, K. Sakka, and K. Ohmiya. 1996. Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *J. Bacteriol.* 178:5732-5740.
- Arai, T., R. Araki, A. Tanaka, S. Karita, T. Kimura, K. Sakka, and K. Ohmiya. 2003. Characterization of a cellulose containing a family 30 carbohydrate-binding module (CBM) derived from

- Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. *J. Bacteriol.* 185:504–512.
- Araki, R., S. Karita, A. Tanaka, T. Kimura, and K. Sakka. 2006. Effect of family 22 carbohydrate-binding module on the thermostability of Xyn10B catalytic module from *Clostridium sterco-rarium*. *Biosci. Biotechnol. Biochem.* 70:3039–3041.
- Bayer, E. A., J.-P. Belaich, Y. Shoham, and R. Lamed. 2004. The cel-lulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* 58:521–554.
- Bayer, E. A., Y. Shoham, and R. Lamed. 2000. Cellulose-decompos-ing bacteria and their enzyme systems, p. 1–41. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes, an Evolving Electronic Resource for the Micro-biological Community*, 3rd ed. Springer Verlag, New York, NY.
- Béguin, P., and P. Alzari. 1998. The cellulosome of *Clostridium ther-mocellum*. *Biochem. Soc. Trans.* 26:178–185.
- Béguin, P., and M. Lemaire. 1996. The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Crit. Rev. Biochem. Mol. Biol.* 31:201–236.
- Berr, K., D. Wassenberg, H. Lilie, J. Behlke, and R. Jaenicke. 2000. e-Crystallin from duck eye lens: comparison of its quaternary struc-ture and stability with other lactate dehydrogenases and complex formation with α -crystallin. *Eur. J. Biochem.* 267:5413–5420.
- Bolam, D. N., A. Ciruela, S. McQueen-Mason, P. Simpson, M. P. Williamson, J. E. Rixon, A. Boraston, J. P. Hazlewood, and J. Harry. 1998. *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* 331:775–781.
- Brun, E., P. E. Johnson, A. L. Creagh, P. Tomme, P. Webster, C. A. Haynes, and I. P. McIntosh. 2000. Structure and binding specificity of the second N-terminal cellulose-binding domain from *Cellulo-monas fimi* endoglucanase C. *Biochemistry* 39:2445–2450.
- Campbell, I. D. 2003. Modular proteins at the cell surface. *Biochem. Soc. Trans.* 31:1107–1114.
- Clout, N. J., A. Basak, K. Wieligmann, O. A. Bateman, R. Jaenicke, and C. Slingsby. 2000. The N-terminal domain of betaB2-crystallin resembles the putative ancestral homodimer. *J. Mol. Biol.* 304: 253–257.
- Doi, R. H., and A. Kosugi. 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev.* 2:541–551.
- Dominguez, R., H. Souchon, S. Spinelli, Z. Dauter, K. S. Wilson, S. Chauvaux, and P. Béguin. 1995. A common protein fold and similar active site in two distinct families of beta-glycanases. *Nat. Struct. Biol.* 2:569–576.
- Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59.
- Fontes, C. M., G. P. Hazlewood, E. Morag, J. Hall, B. H. Hirst, and H. J. Gilbert. 1995. Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacte-ria. *Biochem. J.* 307:151–158.
- Frydman, J., H. Erdjument-Bromage, P. Tempst, and F. U. Hartl. 1999. Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly luciferase. *Nat. Struct. Biol.* 6:697–705.
- Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, and R. A. J. Warren. 1991. Domains in microbial beta-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Rev.* 55: 303–315.
- Hayashi, H., K. I. Takagi, M. Fukumura, T. Kimura, S. Karita, K. Sakka, and K. Ohmiya. 1997. Sequence of xynC and properties of XynC, a major component of the *Clostridium thermocellum* cel-lulosome. *J. Bacteriol.* 179:4246–4253.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280:309–316.
- Henrissat, B., and G. J. Davies. 2000. Glycoside hydrolases and gly-cosyltransferases. Families, modules, and implications for ge-nomics. *Plant Physiol.* 124:1515–1519.
- Henrissat, B., and A. Romeu. 1995. Families, superfamilies and sub-families of glycosyl hydrolases. *Biochem. J.* 311:350–351.
- Jaenicke, R. 1999. Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* 71:155–241.
- Janin, J., and S. J. Wodak. 1983. Structural domains in protein and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol.* 42:21–78.
- Johnson, P. E., M. D. Joshi, P. Tomme, D. G. Kilburn, and I. P. McIntosh. 1996. Structure of the N-terminal cellulose-binding do-main of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* 35:14383–14394.
- Juy, M., A. G. Amit, P. M. Alzari, R. J. Poljak, M. Claeysens, P. Béguin, and J.-P. Aubert. 1992. Three-dimensional structure of a thermostable bacterial cellulase. *Nature* 357:89–91.
- Kataeva, I. A., D. L. Blum, X.-L. Li, and L. G. Ljungdahl. 2001a. Do domain interactions of glycosyl hydrolases from *Clostridium ther-mocellum* cellulosome contribute to protein thermostability? *Protein Eng.* 14:167–172.
- Kataeva, I. A., J. M. Brewer, V. N. Uversky, and L. G. Ljungdahl. 2005. Domain coupling in a multimodular cellobiohydrolase CbhA from *Clostridium thermocellum*. *FEBS Lett.* 579:4367–4373.
- Kataeva, I. A., X.-L. Li, H. Chen, and L. G. Ljungdahl. 1999. CelK—a new cellobiohydrolase from *Clostridium thermocellum* cellulosome: role of N-terminal cellulose-binding domain, p. 454–460. In K. Ohmiya, K. Sakka, S. Karita, M. Hayashi, Y. Kobayashi, and T. Kimura (ed.), *Genetics, Biochemistry and Ecology of Cellulose Degradation*. UniPublishers Co., Tokyo, Japan.
- Kataeva, I. A., and L. G. Ljungdahl. 2003. The *Clostridium thermo-cellum* cellulosome: a multi-protein complex of domain-composed components, p. 651–666. In V. Uversky (ed.), *Protein Structures. Kaleidoscope of Structural Properties and Functions*. Research Signpost, Kerala, India.
- Kataeva, I. A., R. D. Siedel III, X.-L. Li, and L. G. Ljungdahl. 2001b. Properties and mutation analysis of the CelK cellulose-binding do-main from the *Clostridium thermocellum* cellulosome. *J. Bacteriol.* 183:1552–1559.
- Kataeva, I. A., R. D. Seidel III, A. Shah, L. T. West, X.-L. Li, and L. G. Ljungdahl. 2002. The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hy-drolysis of cellulose by modifying its substrate. *Appl. Environ. Microbiol.* 68:4292–4300.
- Kataeva, I. A., V. N. Uversky, J. M. Brewer, F. Schubot, J. Rose, B.-C. Wang, and L. G. Ljungdahl. 2004. Interactions between immunoglobulin-like and catalytic modules in *Clostridium thermo-cellum* cellobiohydrolase CbhA. *Protein Eng. Des. Sel.* 17:759–769.
- Kataeva, I. A., V. N. Uversky, and L. G. Ljungdahl. 2003. Calcium and domain interactions contribute to the thermostability of do-mains of the multimodular cellobiohydrolase, CbhA, a subunit of the *Clostridium thermocellum* cellulosome. *Biochem. J.* 372:151–161.
- Kerr, E. 2004. Broadened applicability of use for industrial enzymes. *Genet. Eng. News* 24:1–5.
- Kozhevnikov, G. O., A. N. Danilenko, E. E. Braudo, and K. D. Schwenke. 2001. Comparative studies of thermodynamic char-acteristics of pea legumin and legumin-T thermal transition. *Int. J. Biol. Macromol.* 29:225–236.
- Kuznetsova, I. M., O. V. Stepanenko, K. K. Zhu, J. M. Zhou, A. L. Fink, and V. N. Uversky. 2002. Unraveling multistate unfolding of rabbit muscle creatine kinase. *Biochim. Biophys. Acta* 1596:138–155.

- Little, E., P. Bork, and R. F. Doolittle. 1994. Tracing and spread of fibronectin type III domain in bacterial glycohydrolases. *J. Mol. Evol.* 39:631-643.
- Ljungdahl, L. G., H. J. M. Op den Camp, H. J. Gilbert, H. R. Harangi, P. J. M. Steenbakkers, and X.-L. Li. 2007. Cellulosomes of anaerobic fungi, p. 271-303. In V. Uversky and I. A. Kataeva (ed.), *Cellulosome*. Nova Science Publishers, Inc, New York, NY.
- Maity, H., M. Maity, M. M. G. Krishna, L. Mayne, and S. W. Englander. 2005. Protein folding; the step-wise assembly of foldon units. *Proc. Natl. Acad. Sci. USA* 102:4741-4746.
- Mandelman, D., A. Belaich, J.-P. Belaich, N. Aghajari, H. Driguez, and R. Haser. 2003. X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J. Bacteriol.* 185:4127-4135.
- McGrath, C. 2007. Mechanistic and functional characterization of glycosyl hydrolases involved in biomass degradation: *Thermobifida fusca* LAM81A and CHI18A and *Clostridium thermocellum* CbhA. Thesis. Cornell University, Ithaca, NY.
- Medved, L. V., C. L. Orthner, H. Lubon, T. K. Lee, W. N. Drohan, and K. C. Ingham. 1995. Thermal stability and domain-domain interactions in natural and recombinant protein C. *J. Biol. Chem.* 270:13652-13659.
- Notenboom, V., A. B. Boraston, D. G. Kilburn, and D. R. Rose. 2001. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* 40:6248-6256.
- Ohmiya, K., K. Sakka, S. Karita, and T. Kimura. 1997. Structure of cellulases and their applications. *Biotechnol. Genet. Eng. Rev.* 14: 365-414.
- Parsiegla, G., A. Belaich, J.-P. Belaich, and R. Haser. 2002. Crystal structure of the cellulase Cel9M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases. *Biochemistry* 41:11134-11142.
- Receveur, V., M. Czjzek, M. Schulein, P. Panine, and B. Henrissat. 2002. Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. *J. Biol. Chem.* 273:40888-40892.
- Rumbley, J., L. Hoang, L. Mayne, and S. W. Englander. 2001. An amino acid code for protein folding. *Proc. Natl. Acad. Sci. USA* 98:105-112.
- Sakon, J., D. Irwin, D. B. Wilson, and P. A. Karplus. 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.* 4:810-818.
- Schubot, F. D., I. A. Kataeva, J. Chang, A. K. Shah, L. G. Ljungdahl, J. P. Rose, and B.-C. Wang. 2004. Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. *Biochemistry* 41:1163-1170.
- Shen, H., N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren. 1995. Cellobiohydrolase B, a second exo-cellobiohydrolase from cellulolytic bacterium *Cellulomonas fimi*. *Biochem. J.* 311:67-74.
- Shen, H., M. Schmuck, I. Pilz, N. R. Gilkes, D. G. Kilburn, R. C. Miller, and R. A. J. Warren. 1991. Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity. *J. Biol. Chem.* 266:11335-11340.
- Shen, M.-Y., E. P. Davis, and A. Sali. 2005. The optimal size of a globular protein domain: a simple sphere-packing model. *Chem. Phys. Lett.* 405:224-228.
- Shin, E.-S., M.-J. Yang, K.-H. Jung, E.-J. Kwon, J. S. Jung, S. Park, J. Kim, H. D. Yun, and H. Kim. 2002. Influence of the transposition of the thermostabilizing domain of *Clostridium thermocellum* xylanase (XynX) on xylan binding and thermostabilization. *Appl. Environ. Microbiol.* 68:3496-3501.
- Smith, M. A., O. A. Bateman, R. Jaenicke, and C. Slingsby. 2007. Mutation of interfaces in domain-swapped human β -B2-crystallin. *Protein Sci.* 16:615-625.
- Srisodsuk, M., T. Reinikainen, M. Penttila, and T. T. Teeri. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *J. Biol. Chem.* 268:20756-20761.
- Tomme, P., A. Boraston, B. McLean, J. Kormos, A. L. Creagh, K. Sturch, N. R. Gilkes, C. Haynes, R. A. J. Warren, and D. G. Kilburn. 1998. Characterization and affinity applications of cellulose-binding domains. *J. Chromatogr.* 715:283-296.
- Tomme, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, and T. A. Steitz. 1996. Crystal structure of a bacterial family III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* 15:5739-5751.
- Tomme, P., R. A. J. Warren, R. C. Miller, D. G. Kilburn, and N. R. Gilkes. 1995. Cellulose-binding domains: classification and properties, p. 142-163. In J. M. Saddler and M. H. Penner (ed.), *Enzymatic Degradation of Insoluble Polysaccharides*. American Chemical Society, Washington, DC.
- von Ossowski, I., J. T. Eaton, M. Czjzek, S. J. Perkins, T. P. Frandsen, P. M. Schulein, P. Panine, B. Henrissat, and V. Receveur-Brechot. 2005. Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulose. *Biochem. J.* 88: 2823-2832.
- Wassenberg, D., H. Schurig, W. Liebl, and R. Jaenicke. 1997. Xylanase XynA from the hyperthermophilic bacterium *Thermotoga maritima*: structure and stability of the recombinant enzyme and its isolated cellulose-binding domain. *Protein Sci.* 6:1718-1726.
- Wassenberg, D., C. Welker, and R. Jaenicke. 1999. Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* 289:187-193.
- Wenk, M., and R. Jaenicke. 1998. Kinetic stabilization of a modular protein by domain interactions. *FEBS Lett.* 438:127-130.
- Wenk, M., and R. Jaenicke. 1999. Calorimetric analysis of the Ca²⁺-binding $\beta\gamma$ -crystallin homolog protein S from *Myxococcus xanthus*: intrinsic stability and mutual stabilization of domains. *J. Mol. Biol.* 293:117-124.
- Zverlov, V. V., and W. H. Schwarz. 2007. The *C. thermocellum* cellulosome: novel components and insights from the genomic sequence, p. 119-151. In I. A. Kataeva (ed.), *Cellulosome*. Nova Science Publishers, Inc., Hauppauge, NY.
- Zverlov, V. V., G. V. Velikodvorskaya, W. H. Schwarz, K. Bronnenmeier, J. Kellermann, and W. L. Staudenbauer. 1998. Multi-domain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *J. Bacteriol.* 180:3091-3099.