Calcium and domain interactions contribute to the thermostability of domains of the multimodular cellobiohydrolase, CbhA, a subunit of the *Clostridium thermocellum* cellulosome

Irina A. KATAEVA^{*1}, Vladimir N. UVERSKY^{†‡} and Lars G. LJUNGDAHL^{*}

^{*}Department of Biochemistry and Molecular Biology, and Center for Biological Resources Recovery, University of Georgia, Athens, GA 30602-7229, U.S.A., [†]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, U.S.A., and [‡]Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region 142292, Russia

Each of three internal domains of multi-modular cellobiohydrolase CbhA from *Clostridium thermocellum*, X1₁, X1₂ (previously designated as fibronectin type 3-like modules, Fn3₁ and Fn3₂) and family 3 carbohydrate-binding module (CBM3) binds 1 mol of Ca²⁺. Structures and thermal stabilities of X1₁, X1₂, CBM3, X1₁X1₂, and X1₁X1₂-CBM3 containing Ca²⁺ (holo-proteins) and without Ca²⁺ (apo-proteins) have been studied using CD spectroscopy. All domains are β -proteins with irregular far-UV CD spectra due to the aromatic side chain contributions. The positive signal at 294 nm in the near-UV CD spectrum of X1₁ lacking a tryptophan residue might be attributed to the presence of aromatic clusters. Thermal denaturation of all proteins is reversible and results in the total loss of tertiary structure and preservation of significant amount of ordered secondary structure. Removal of Ca²⁺ destabilizes polypeptides in a different way and to a different extent. It decreases the melting temperature (T_m) (by 20 °C) and co-operativity of thermal transition of X1₁, increases the number of transitions and lowers the co-operativity of

unfolding of CBM3, and slightly decreases T_m s (2.4–4.2 °C) of X1₂, X1₁X1₂, and X1₁X1₂-CBM3. Transitions of X1₁X1₂ and X1₁X1₂-CBM3 follow a two-state model regardless of the presence of Ca²⁺. X1₁ is strongly stabilized in the apo-X1₁X1₂ and apo-X1₁X1₂-CBM3 as they display T_m s similar to those of individual and combined holo-modules. Observed CD spectra of X1₁X1₂ and X1₁X1₂-CBM3 differ from those calculated as the simple weighted sum of individual modules. These differences are more prominent in spectra of apo-proteins. The results indicate the presence of inter-domain interactions in CbhA. Holo-modules, i.e. containing Ca²⁺, behave essentially independently, but in the absence of Ca²⁺ domain interactions are more important for the conformation of the polypeptides.

Key words: domain-structured polypeptide, glycoside hydrolase, multimodular protein, stabilization of protein fold, temperature-induced denaturation, thermodynamics.

INTRODUCTION

Polypeptide chains beyond 50–100 amino acid residues tend to form domains [1]. 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’ [2]. As domains have an intrinsic capacity to form their native fold spontaneously and autonomously, appearance of multi-modular proteins seems to be a great achievement of evolution, as long polypeptide chains start to fold in several places, independently minimizing misfolding due to wrong intramolecular interactions [1].

Microbial extracellular glycoside hydrolases, which are active against insoluble plant cell wall biopolymers, are exclusively domain-structured polypeptides comprising a catalytic domain with different ‘accessory’ modules attached to the N- or C-terminus [3]. Catalytic modules and carbohydrate-binding modules (CBMs) have been classified into several families on the basis of amino-acid sequence similarity [4]. It has been noted that domains in glycoside hydrolases are not randomly combined. CBMs of a particular family have a tendency to be associated with a catalytic domain of a particular family, and to be located in a specific place of the polypeptide [5]; Ig-like domains often, but not necessarily, precede catalytic domains belonging to family 9 glycoside hydrolases [3]; fibronectin type 3-like domains (Fn3)

are often multiple and located between catalytic and CBMs [6]. These observations suggest the existence of specific interactions between domains in glycoside hydrolases similar to that of some other modular proteins [7–9].

The importance of correct domain interactions in glycoside hydrolases has been demonstrated. Truncation or transposition of some ancillary domains resulted in changes in activity towards insoluble substrates [10–12], in the mode of action [13], in the shift of optimal temperature to a lower range, and decreased thermostability [14,15] of catalytic modules. Although there have been some random observations, there is no deep understanding of how domains in bacterial glycoside hydrolases interact and affect overall properties of the enzymes.

Calcium is involved in a diversity of biological processes. One major role of Ca²⁺ is to stabilize native folds of proteins. For this reason Ca²⁺ is a constituent of many thermostable proteins [16]. Cellobiohydrolase CbhA is a component of the cellulolytic/hemicellulolytic complex termed the ‘cellulosome’ of the anaerobic thermophilic bacterium *Clostridium thermocellum*. CbhA is a typical representative of thermostable multi-modular Ca²⁺-containing enzymes. It is composed of an N-terminal family 4 CBM (CBM4), an Ig-like module, a family 9 glycoside hydrolase catalytic domain (GH9), X1₁ and X1₂ modules (formerly called Fn3₁ and Fn3₂), a family 3 CBM (CBM3) and a

Abbreviations used: CBM, carbohydrate-binding module; GH, glycoside hydrolase module; X1₁ and X1₂, modules in CbhA previously designated as fibronectin type 3-like modules; $[\theta]_{\text{mrv}}$, mean residue ellipticity; T_m , mid-point transition temperature.

¹To whom correspondence should be addressed (e-mail kataeva@arches.uga.edu).

Table 1 Oligonucleotides used for cloning

Amino acid residue numbering is according to that of CbhA [18]. Restriction sites are underlined.

Primer	Sequence	Restriction site	Location	Direction
1. X1 ₁ F	5'-CTAGCTAGCACAGACAGTGAACCGATAAG-3'	<i>NheI</i>	2887–2907	Forward
2. X1 ₂ R	5'-TTTTCCTTTTTCGCGCCGCTACAAATATTTTACGGAATC-3'	<i>NotI</i>	3415–3435	Reverse
3. X1 ₁ F(DKV)	5'-CTAGCTAGCGATAAGGTAAGTATTGATTCGCCT-3'	<i>NheI</i>	2902–2923	Forward
4. X1 ₂ R	5'-TTTTCCTTTTTCGCGCCGCTTTCACAATTACCTTAACAAG-3'	<i>NotI</i>	3155–3174	Reverse
5. X1 ₂ F	5'-CTAGCTAGCCCTACTGTAAACTTACT-3'	<i>NheI</i>	3277–3315	Forward
6. CBM3F	5'-CTAGCTAGCGATGTAAAGTACAGTATTG-3'	<i>NheI</i>	3448–3468	Forward
7. CBM3R	5'-TTTTCCTTTTTCGCGCCGACCGCCCGGCGCGTCCCCA-3'	<i>NotI</i>	3862–3882	Reverse

dockerin domain [17]. In a previous publication we demonstrated that X1₁X1₂ modified the surface of cellulose and promoted its hydrolysis by the catalytic domain [18]. It was also found that each X1₁, X1₂ and CBM3 domain contained 1 mol of Ca²⁺ [18]. Despite some contradictory data, the most probable role of Ca²⁺ in CBMs is to stabilize the native protein structure [19].

In the present paper we present a thermal denaturation study of native (holo-) and Ca²⁺-free (apo-) individual domains X1₁, X1₂ and CBM3, and domain combinations X1₁X1₂ and X1₁X1₂-CBM3 of CbhA by CD spectroscopy. The results indicated that both Ca²⁺ binding and domain interactions contributed to the structure and stability of the domains in modular CbhA.

MATERIALS AND METHODS

Bacterial strains, culture conditions and plasmids

C. thermocellum JW20 was used as a source of genomic DNA. The bacterium was grown anaerobically under a N₂ atmosphere at 60 °C in a pre-reduced medium with 1% (w/v) cellobiose [20]. *Escherichia coli* BL21(DE3)pLys (Stratagene Cloning Systems, La Jolla, CA, U.S.A.) was used as cloning host for the T7 RNA polymerase expression vector pET-21b(+) (Novagen, Madison, WI, U.S.A.). It was grown in Luria–Bertani medium supplemented with ampicillin (100 µg/ml).

Primer design, PCR and cloning

Flanking primers containing restriction sites were designed according to the DNA sequence of *cbhA* (GenBank® accession number X80993) (Table 1) and synthesized with an Applied Biosystems DNA synthesizer. DNA fragments were amplified by PCR using the primers in combination with purified genomic DNA [20] as a template. PCRs were done on a 480 Thermal Cycler (Perkin–Elmer, Norwalk, CT, U.S.A.). The reactions were carried out with *Taq* polymerase (New England Biolabs, Beverly, MA, U.S.A.). The annealing temperature was 54 °C and the extension time depended on the length of the fragment. PCR products were separated by 1% agarose gel electrophoresis and extracted from the gel using the GeneClean II Kit (Bio 101, La Jolla, CA, U.S.A.). The extracted DNA fragments were digested with restriction enzymes and ligated into the pET-21b(+) vector linearized with the same enzymes. The ligation products were used to transform BL21(DE3)pLys competent cells. Each construct was verified by both restriction analysis and DNA sequencing.

Protein purification

All polypeptides were His₆ tagged at the C-terminus. They were purified from BL21(DE3)pLys cultures harbouring the pET-

21b(+) vector with the DNA fragment of interest. Harvesting was carried out 5 h after induction with 1 mM isopropyl β-D-thiogalactopyranoside. All purification steps were done at 4 °C, except for the FPLC column chromatography, which was run at 23 °C. After collection, the cells were washed with 20 mM sodium phosphate, pH 7.5, containing 0.5 M NaCl (buffer A) and disintegrated with a French Press. Cell debris was removed by centrifugation. The clear supernatant was applied to a Ni²⁺-nitrilotriacetate agarose (Qiagen, Valencia, CA, U.S.A.) column equilibrated with buffer A. The column was washed with 20 mM sodium phosphate/0.5 M NaCl, pH 6.0. Proteins were eluted by a gradient of 0.0–0.5 M imidazole, pH 6.0. Fractions containing proteins were combined, concentrated by precipitation with (NH₄)₂SO₄, and dialysed against 20 mM Tris/HCl buffer, pH 7.5, containing 0.1 M NaCl. Dialysed proteins were further purified by gel-filtration on a TSK 3000SW column (TosoHaas, Montgomeryville, PA, U.S.A.). Eluted samples were concentrated using Centricon 3 concentrators (Amicon, Beverly, MA, U.S.A.) and stored at 4 °C. All proteins were purified to near homogeneity as ascertained by SDS/PAGE.

Metal content

The metal contents of the proteins were determined by inductively coupled plasma MS (EPA Method 610) using VG PlasmaQuad 3 ICP-MS. Samples (1 ml) containing proteins at concentrations of 50–200 µM were quantitatively analysed for elements of interest using multi-element calibration solutions. A single calibration set was prepared and run at the beginning of the analysis procedure. Prior analysis the proteins were intensively dialysed against 1 litre of 20 mM Tris/HCl buffer, pH 7.5, used as a blank. Each data point was a mean of three replicates.

Protein determination

During purification, protein levels were determined with the Coomassie Protein Assay Reagent (Pierce, Rockford, IL, U.S.A.). Concentrations of purified proteins were determined on the basis of A₂₇₈ values. Each protein was extensively dialysed against 20 mM sodium phosphate buffer, pH 6.0, and then the solution was centrifuged at 20 000 g for 60 min. The absorbance was read using the dialysis buffer as a blank. Molar absorption coefficients calculated from tryptophan and tyrosine residue contents were (M⁻¹·cm⁻¹): 34 480 for X1₁X1₂-CBM3; 14 650 for X1₁X1₂; 3 840 for X1₁; 10 810 for X1₂; and 19 830 for CBM3.

Preparation of Ca²⁺-free polypeptides

Calcium was removed from the proteins dissolved in 20 mM sodium phosphate buffer, pH 6.0, by overnight incubation at

4 °C in the presence of 5 % (w/v) Chelex-100 resin (Bio-Rad Laboratories, Hercules, CA, U.S.A.). The resin was removed by centrifugation and the polypeptides were dialysed against three changes of Ca²⁺-free 20 mM sodium phosphate buffer, pH 6.0. To remove traces of multivalent cations the buffer was stirred overnight with 5 % (w/v) Chelex-100. All glassware, plasticware and quartz cells were soaked overnight in 4 M HCl and then thoroughly rinsed with deionized water.

CD spectra

CD measurements were carried out using a Jasco J-710 spectropolarimeter with a jacketed quartz cell with a 1.0-mm (far-UV region) or 10-mm (near-UV region) pathlength. The cell temperature was controlled to within ± 0.1 °C by circulating (15 l/min) a mixture of water/antifreeze (1:1) via a Neslab R-111 water bath (NESLAB Instruments, Portsmouth, NH, U.S.A.) through a cell jacket. The results were expressed as mean residue ellipticity $[\theta]_{\text{mrw}}$. The spectra obtained were averages of five scans. The spectra were smoothed using an internal algorithm in the Jasco software package, J-710 for Windows. Protein samples were in 20 mM sodium phosphate buffer (pH 6.0) at a concentration of 10 μ M for the far-UV CD and at an A_{278} of 0.5 for the near-UV CD.

Temperature-induced denaturation

Thermal denaturation of different constructs was monitored by CD in both near- (190–250 nm) and far-UV (250–320) regions. For the analysis of thermostability, the temperature was increased from 25 to 100 °C with a step size of 10 or 5 °C, and with the CD spectra measured at each temperature after incubation for 30 min. At the end of each thermal denaturation experiment, the sample was cooled for 10 min to 23 °C and spectra were recorded to determine the extent of refolding.

Thermodynamic analysis

Denaturation curves were analysed using the two-state model. The difference in free energy (Gibbs energy) between the native and the denatured states, ΔG^{eff} , was calculated by the equation:

$$\Delta G^{\text{eff}} = -RT \ln K^{\text{eff}} \quad (1)$$

where R is the gas constant (1.987 cal/deg/mol), T is the absolute temperature and K^{eff} is the effective equilibrium constant. The corresponding thermodynamic parameters, T_m (the midpoint of the thermal denaturation curve), ΔH_m (the enthalpy change at T_m) and ΔS_m (the entropy change at T_m), were determined from ΔG^{eff} versus T plots. In the midpoint of a thermal denaturation curve

$$\Delta G = 0 = \Delta H_m - T_m \Delta S_m \quad (2)$$

This gives ΔS_m as the slope of the ΔG versus T plot at T_m , and $\Delta H_m = T_m \Delta S_m$ [22]. The difference in the free energy change associated with the Ca²⁺ binding, $\Delta(\Delta G)$, was calculated as

$$\Delta(\Delta G) = \Delta(T_m) \Delta S_m = \Delta(T_m) (\Delta H_m / T_m) \quad (3)$$

where $\Delta(T_m)$ is the difference between the ΔT_m values measured for the given protein in the presence and the absence of Ca²⁺, whereas ΔS_m and ΔH_m are values for the Ca²⁺-free forms of the proteins [21,22].

Analysis of spectroscopic data in a form of parametric dependencies

The 'phase diagram' method analysis of spectroscopic data is extremely sensitive for the detection of intermediate states [23–26]. Although this method was developed for the analysis of fluorescence data, it can be used with any spectroscopic technique. The essence of this method is to build up the diagram of I_{λ_1} versus I_{λ_2} , where I_{λ_1} and I_{λ_2} are the spectral intensity values measured at wavelengths λ_1 and λ_2 under different experimental conditions for a protein undergoing structural transformations. As spectral intensity is the extensive parameter, it will describe any two-component system by a simple relationship:

$$I(\lambda) = \alpha_1 I_1(\lambda) + \alpha_2 I_2(\lambda) \quad (4)$$

$$\alpha_1 + \alpha_2 = 1,$$

where $I_1(\lambda)$ and $I_2(\lambda)$ are the intensities of the first and second components, whereas α_1 and α_2 are their relative concentrations. Substituting for α_1 ($\alpha_1 = 1 - \alpha_2$), eqn (4) can be rewritten as:

$$I(\lambda) = (1 - \alpha_2) I_1(\lambda) + \alpha_2 I_2(\lambda) = I_1(\lambda) + \alpha_2 [I_2(\lambda) - I_1(\lambda)] \quad (5)$$

α_2 may be determined from the intensity measurements at two different wavelengths, λ_1 and λ_2 . In fact:

$$I(\lambda_1) = I_1(\lambda_1) + \alpha_2 [I_2(\lambda_1) - I_1(\lambda_1)] \quad (6)$$

$$I(\lambda_2) = I_1(\lambda_2) + \alpha_2 [I_2(\lambda_2) - I_1(\lambda_2)] \quad (7)$$

and

$$\alpha_2 = [I(\lambda_2) - I_1(\lambda_2)] / [I_2(\lambda_2) - I_1(\lambda_2)] \quad (8)$$

This allows determination of the relationship between $I(\lambda_1)$ and $I(\lambda_2)$ by substitution of α_2 in equations (9) from (11):

$$I(\lambda_1) = I_1(\lambda_1) + \frac{I(\lambda_2) - I_1(\lambda_2)}{I_2(\lambda_2) - I_1(\lambda_2)} [I_2(\lambda_1) - I_1(\lambda_1)]$$

$$= I_1(\lambda_1) - \frac{I_2(\lambda_1) - I_1(\lambda_1)}{I_2(\lambda_2) - I_1(\lambda_2)} I_1(\lambda_2) + \frac{I_2(\lambda_1) - I_1(\lambda_1)}{I_2(\lambda_2) - I_1(\lambda_2)} I(\lambda_2) \quad (9)$$

or:

$$I(\lambda_1) = a + b I(\lambda_2) \quad (10)$$

where

$$a = I_1(\lambda_1) - \frac{I_2(\lambda_1) - I_1(\lambda_1)}{I_2(\lambda_2) - I_1(\lambda_2)} I_1(\lambda_2) \quad \text{and} \quad b = \frac{I_2(\lambda_1) - I_1(\lambda_1)}{I_2(\lambda_2) - I_1(\lambda_2)} \quad (11)$$

In application to protein unfolding, the relationship in eqn (10) predicts that $I(\lambda_1) = f[I(\lambda_2)]$ will be linear if changes in protein environment lead to an all-or-nothing transition between two different conformations. Alternatively, the non-linearity of this function reflects the sequential character of structural transformations. Moreover, each linear portion of the $I(\lambda_1) = f[I(\lambda_2)]$ dependence will describe an individual all-or-nothing transition.

Figure 1 Near-UV CD spectra of individual domains (A) X1₁, (B) X1₂ and (C) CBM3, and domain combinations (D) X1₁X1₂ and (E) X1₁X1₂-CBM3

Spectra were recorded at 25 °C (continuous lines) or 100 °C (broken lines) in the absence (lines with circles) or presence of Ca²⁺ (lines only).

RESULTS

Metal content

Individual domains X1₁, X1₂ and CBM3 bind 1 mol of Ca²⁺ each [18]. Domain combinations X1₁X1₂ and X1₁X1₂-CBM3 bind 2 (1.95 ± 0.2) and 3 (2.03 ± 0.25) mol of Ca²⁺ respectively. Calcium content in all constructs treated with Chelex-100 was found to be less than 0.02 mol per mol of protein.

Structural properties of CbhA domains

CD spectra in the aromatic region

Figure 1 represents near-UV CD spectra of holo- and apo-forms of individual domains X1₁, X1₂ and CBM3, and domain combinations X1₁X1₂ and X1₁X1₂-CBM3 recorded at 25 °C and 100 °C. The individual domains show differences in the shape

and intensity of the near-UV CD signals correlating with the different contents of aromatic amino acid residues. Analysis of the amino-acid sequence of CbhA revealed that X1₁ domain (residues 822–912) contains three tyrosine, two phenylalanine and no tryptophan residues; X1₂ domain (residues 913–999) has four tyrosine, three phenylalanine and one tryptophan residues; and CBM3 (residues 1004–1148) includes eleven tyrosine, nine phenylalanine and one tryptophan residues.

Figure 1 shows that not only tryptophan-containing X1₂ and CBM3, but also the tryptophan-free X1₁ have well-pronounced peaks in the vicinity of 294 nm. As there is no tryptophan residue in X1₁, the presence of a longer wavelength peak in the near-UV CD spectrum suggests the existence of specific clusters of aromatic residues.

It should be noted that the spectral shape of all constructs studied (except X1₁) is almost independent of Ca²⁺ presence (Figure 1). This reflects the fact that Ca²⁺ binding does not affect tertiary structures of X1₂ and CBM3, but causes some changes

Figure 2 Comparison of near-UV CD spectra

Comparison of spectra recorded at 25 °C of X_1X_{12} in the presence (A) and absence (C) of Ca^{2+} , and X_1X_{12} -CBM3 in the presence (B) and absence (D) of Ca^{2+} to the spectra calculated as simple weighted sums based on spectra recorded for the individual domains: $(\text{X}_1 + \text{X}_{12})/2$ (A and C, $\cdots$) and $(\text{X}_1 + \text{X}_{12} + \text{CBM3})/3$ (B and D, $\cdots$) or $(\text{X}_1\text{X}_{12} + \text{CBM3})/2$ (B and D, $---$). Experimental spectra of the domain combinations X_1X_{12} and X_1X_{12} -CBM3 are shown by continuous lines.

in the structure of X_{11} . These changes appear most likely in the aromatic cluster of the X_{11} domain, as the intensity of the longer wavelength signal is increased for apo- X_{11} . As Ca^{2+} does not have any visible effect on the CD spectra of the combined domains, it might be suggested that within those domain combinations X_{11} is stabilized by the interaction with X_{12} . The near-UV CD signals of all polypeptides measured at high temperatures are almost completely absent regardless of the presence of Ca^{2+} , reflecting the temperature-induced loss of specific tertiary structure, i.e. heat-induced denaturation.

To reveal possible interactions between the domains and their effect on the conformation of CbhA, we compared the near-UV CD spectra of domain combinations X_1X_{12} and X_1X_{12} -CBM3 with the spectra calculated as simple weighted sums of $\text{X}_1 + \text{X}_{12}$ (' X_1X_{12} ' construct), and $\text{X}_1 + \text{X}_{12} + \text{CBM3}$ or $\text{X}_1\text{X}_{12} + \text{CBM3}$ (' X_1X_{12} -CBM3' construct) spectra (Figure 2). If the domains were absolutely independent, then spectra of X_1X_{12} and X_1X_{12} -CBM3 would be equal to the spectra calculated as weighted sums of the spectra of individual domains. However, if the domain interactions induce changes in the tertiary structure, then some difference between the observed and calculated spectra must be seen. Figure 2 shows that inter-domain interactions play an important role in the conformation of X_1X_{12} -CBM3 either in the presence (Figure 2B) or in the absence (Figure 2D) of Ca^{2+} . As for the X_1X_{12} construct, a considerable difference between measured and calculated spectra was observed in the absence of Ca^{2+} (Figure 2C). Upon Ca^{2+} binding this difference was essentially eliminated (Figure 2A). Thus, Ca^{2+} is important for the interactions between the X_{11} and X_{12} domains of CbhA.

CD spectra in the peptide region

The far-UV CD spectra of individual X_{11} , X_{12} and CBM3 domains, and domain combinations X_1X_{12} and X_1X_{12} -CBM3 are given in Figure 3. The spectral shapes of all constructs recorded at 25 °C indicate mainly β -structure, although they are rather unusual. This is due to the contribution of aromatic side groups, the effect of which is more pronounced in β -proteins [27–31]. There are no structural data on proteins displaying sequence similarity to X_{11} and X_{12} , but structures of all known CBMs comprise mainly β -sheet [5,14,19]. Ca^{2+} causes different effects on the far-UV spectra of the constructs. Loss of Ca^{2+} does not affect the CD spectra of X_{12} , CBM3, and X_1X_{12} -CBM3 recorded at 25 °C; it slightly changes spectrum of X_1X_{12} , and it has a dramatic effect on the spectrum of X_{11} (Figure 3).

At elevated temperatures Ca^{2+} considerably stabilizes these domains against thermal unfolding. Interestingly, denaturation of CbhA constructs leads to the 'normalization' of their far-UV CD spectra. Figure 3 shows that at 100 °C spectra of apo- and holo-CBM3 resemble typical spectra of β -proteins; spectra of holo- X_{11} , X_{12} and X_1X_{12} -CBM3 look like spectra of $\alpha + \beta$ proteins; spectra of apo- X_{11} and apo- X_{12} , and holo- and apo- X_1X_{12} are typical of essentially disordered proteins.

Conformational stabilities of CbhA domains

Equilibrium thermal denaturation and 'phase-diagram' analysis

Complete or almost complete reversibility of denaturation of all constructs allowed the application of thermodynamic analysis. It

Figure 3 Far-UV CD spectra of individual domains X1, (A), X2 (B) and CBM3 (C), and domain combinations X1X2 (D) and X1X2-CBM3 (E)

Spectra were recorded at 25 °C (continuous lines) or 100 °C (broken lines) in the absence (lines with circles) or presence of Ca^{2+} (lines only).

is very important to remember that far-UV CD spectra of the constructs measured at 25 °C are significantly affected by the contribution of aromatic side groups (Figure 3). Moreover, thermal denaturation, while accompanied by complete disruption of the tertiary structure (Figure 1), does not lead to the unfolding of the proteins, as all of them are characterized by significant amounts of ordered secondary structure, even at 100 °C (Figure 3). Thus temperature-denatured constructs are characterized by the absence of rigid tertiary structure and the presence of essential secondary structure, i.e. they resemble molten globules [32]. This means that unambiguous thermodynamic analysis can be performed only for the near-UV CD data.

To elucidate the mechanism of temperature-induced structural transformations we used the method of 'phase diagrams' (see the Materials and methods section). Figure 4 shows phase diagrams for the temperature-induced denaturation of holo-constructs and

apo-constructs. Analysis of these diagrams makes two important points: mechanisms of denaturation are different for different constructs; and Ca^{2+} affects denaturation of different domains in different manners. The phase diagram of holo-X1 consists of two linear parts. Removal of Ca^{2+} dramatically changes the shape of the phase diagram. However, the plot still has two linear parts indicating the existence of at least two independent transitions linking three different conformations regardless of Ca^{2+} binding (Figure 4A). Figure 4(B) shows that the denaturation of both holo- and apo-X2 most likely represents an all-or-nothing transition. Similar behaviour is characteristic for the denaturation of X1X2 (Figure 4D). The denaturation of CBM3 is Ca^{2+} sensitive (Figure 4C). Thus apo-CBM3 denatures according to the three-state model (the phase diagram has two linear parts), whereas the denaturation of holo-CBM3 is described as an all-or-nothing transition. Finally, Figure 4(E) shows that removal of Ca^{2+} has

Figure 4 Phase diagrams based on $[\theta]_{\lambda,1}$ versus $[\theta]_{\lambda,2}$ (see text for details), for the heat-induced denaturation of different domains of CbhA, based on the temperature-induced changes in the near-UV CD spectra of individual domains X1 (A), X2 (B) and CBM3 (C), and domain combinations X1X2 (D) and X1X2-CBM3 (E)

Data for holo- and apo-proteins are given in closed circles and continuous lines, and open circles and broken lines, respectively.

no effect on the mechanism of X1X2-CBM3 melting, with the apo- and holo-forms being denatured according to the two-state scheme.

Equilibrium thermal denaturation monitored by changes in the near-UV CD spectra

Figure 5 presents denaturation curves monitored by changes in the near-UV CD spectra of holo- and apo-proteins. All constructs are relatively thermostable with transition temperatures (T_m) lying in the range of 78–83 °C (Table 2). Results presented in Figure 5(A) are consistent with the conclusion that Ca^{2+} significantly contributes to the thermal stability of X1: T_m of apo-X1 is 20 °C less than T_m of holo-X1 (approx. 62 °C and 82 °C respectively). This corresponds to a 4.2 kcal/mol decrease in ΔG

value, reflecting dramatic destabilization of protein associated with the release of Ca^{2+} (Table 2). In contrast, stabilities of X2, CBM3, X1X2 and X1X2-CBM3 are not affected much by Ca^{2+} , as T_m values of the corresponding apo-forms decrease by only 2.5, 0.0, 4.3 and 1.8 °C respectively (with corresponding changes in ΔG of 0.59, 0.00, 1.02 and 0.49 kcal/mol). These changes in T_m values are significantly smaller than determined for typical Ca^{2+} -binding proteins [33].

Melting of holo-CBM3 is characterized by T_m of approx. 80–82 °C (Table 2), with a sharp transition occurring within 6–7 °C. The denaturation transition of apo-CBM3 has a comparable T_m value, is broader and occurs within an interval of 15 °C (Figure 5C). This suggests that Ca^{2+} influences the co-operativity of thermal denaturation of the domain. Similar results are obtained for X1X2 (Figure 5D). Thermal denaturation of X2 is mostly unaffected by Ca^{2+} as both T_m values, and co-operativities of the

Figure 5 Temperature-induced denaturation of different domains of CbhA monitored by changes in the near-UV CD spectra for individual domains X1₁ (A), X1₂ (B), CBM3 (C) and domain combinations X1₁X1₂ (D) and X1₁X1₂-CBM3 (E)

Transition curves of holo- and apo-proteins are given in continuous lines with circles and broken lines with triangles respectively. For X1₁ transition curves represent changes at 276 nm (closed symbols) and 295 nm (open symbols); for X1₂ at 291 nm; for CBM3 at 277 nm (open symbols) and 295 nm (closed symbols); for X1₁X1₂ at 290 nm; for X1₁X1₂-CBM3 at 276 nm.

denaturation of the holo- and apo-polypeptides, are very similar (Figure 5B). In the absence of Ca²⁺, X1₁ is significantly stabilized by the interaction with X1₂, as *T_m* of apo-X1₁X1₂ is 20 °C higher than that of apo-X1₁ (Table 2). Thus X1₁ by itself is stabilized by Ca²⁺, whereas in X1₁X1₂ the stabilization of the domain occurs via the inter-domain interactions. In agreement with this proposal, Figure 4(D) shows that the melting of X1₁X1₂ appears to be a simple all-or-nothing process.

It is possible that the inter-domain surface in X1₁X1₂ has a hydrophobic character, as the solubility of X1₁ and X1₂ is limited to 10–12 mg/ml, whereas X1₁X1₂ is soluble at 70 mg/ml. This situation is similar to that described for a protein S composed of two domains with an extensive hydrophobic interface [34].

Some interesting information on the nature of thermal transitions of different CbhA modules can be extracted from the comparison of melting data presented in a traditional form (Figure 5) and in the form of parametric dependencies (Figure 4).

As shown in Figure 4, different domains denature in a different manner. For example, melting of X1₂ and X1₁X1₂ represents a simple all-or-nothing process (Figures 4B and 4D respectively), while thermal denaturation of X1₁ represents at least two sequential transitions (Figure 4A). Figure 5(A) shows that in the case of apo-X1₁, increasing temperature first results in ellipticity changes in the vicinity of 276 nm, followed by an ellipticity decrease at longer wavelengths (in the vicinity of 295 nm, see also Table 2). This indicates that melting of the environment of individual aromatic residues happens at lower temperatures than the melting of aromatic cluster(s). Ca²⁺ removal leads to dramatic destabilization of the environment of individual aromatic residues (by approx. 20 °C), whereas the stability of the aromatic cluster is less affected (*T_m* decreased by 1.5 °C). As for the individual CBM3 domain, removal of Ca²⁺ changes its mechanism of thermal denaturation from a two-state to a three-state model, without considerable change in the stability of this domain (Figure 4C and Table 2). Finally, regardless of the presence of

Table 2 Thermodynamic parameters of melting of individual domains X1₁, X1₂ and CBM3, and their combinations X1₁X1₂, and X1₁X1₂-CBM3 derived from the analysis of the temperature-induced changes in their near-UV CD spectra

Protein, form	ΔS_m^* (cal/mol/deg)	$\Delta H_m^\dagger$ (kcal/mol)	$T_m^\ddagger$ (°C)	$\Delta(T_m)^\S$ (°C)	$\Delta(\Delta G_m)^\parallel$ (kcal/mol)
X1 ₁ , apo					
[θ] ₂₇₆	208 ± 15	69.8 ± 5.2	61.8 ± 0.9		
[θ] ₂₉₅	177 ± 16	61.2 ± 5.7	72.3 ± 0.9		
X1 ₁ , holo					
[θ] ₂₇₆	132 ± 9	46.8 ± 3.2	82.0 ± 0.9	+ 20.2 ± 0.9	+ 4.20 ± 0.09
[θ] ₂₉₅	189 ± 28	65.5 ± 9.5	73.8 ± 0.9	+ 1.5 ± 0.9	+ 0.27 ± 0.12
X1 ₂ , apo					
[θ] ₂₉₀	236 ± 14	82.3 ± 4.5	75.8 ± 0.9		
X1 ₂ , holo					
[θ] ₂₉₀	234 ± 12	82.5 ± 4.2	78.3 ± 0.8	+ 2.5 ± 0.9	+ 0.59 ± 0.08
CBM3, apo					
[θ] ₂₇₇	222 ± 16	78.2 ± 5.5	79.4 ± 0.9		
[θ] ₂₉₅	139 ± 5	49.5 ± 1.5	82.4 ± 1.2		
CBM3, holo					
[θ] ₂₇₇	156 ± 9	55.0 ± 3.5	79.4 ± 1.0	0.0 ± 1.0	0.00 ± 0.10
[θ] ₂₉₅	187 ± 14	66.3 ± 5.1	81.5 ± 1.0	-0.9 ± 1.2	-0.12 ± 0.15
X1 ₁ X1 ₂ , apo					
[θ] ₂₉₀	238 ± 13	83.7 ± 3.5	78.7 ± 0.7		
X1 ₁ X1 ₂ , holo					
[θ] ₂₉₀	291 ± 14	103.6 ± 4.2	83.0 ± 0.8	+ 4.2 ± 0.8	+ 1.02 ± 0.08
X1 ₁ X1 ₂ -CBM3, apo					
[θ] ₂₇₇	206 ± 11	72.7 ± 2.5	79.2 ± 0.6		
X1 ₁ X1 ₂ -CBM3, holo					
[θ] ₂₇₇	293 ± 11	103.6 ± 2.5	81.6 ± 0.6	+ 2.4 ± 0.6	+ 0.49 ± 0.06

* Slope of ΔG versus T plot at T_m .† $\Delta H_m = T_m(K) \times \Delta S_m$.

‡ Midpoint of thermal denaturation curve.

§ Difference between the T_m values measured for the given protein in the holo- and apo-form.|| $\Delta(\Delta G) = \Delta(T_m) \times \Delta S_m = \Delta(T_m) \times (\Delta H_m/T_m)$, where ΔS_m and ΔH_m are values for the apo-forms of proteins.

Ca²⁺, X1₁X1₂-CBM3 is denatured according to a two-state model (Figures 4E and 5E, and Table 2). In other words, all three domains in this construct melt simultaneously, once again underlining the stabilizing effect of inter-domain interactions.

DISCUSSION

CbhA of *C. thermocellum* comprises seven modules and binds perhaps as many as nine Ca²⁺ ions. X1₁, X1₂ and CBM3 contain one Ca²⁺ each. Two structural Ca²⁺ ions have been found in highly conserved dockerin domains [35] and are likely to be present in the dockerin domain of CbhA. In a previous study [36], we reported the presence of 1 mol of Ca²⁺ in the N-terminal CBM4 from the highly similar cellobiohydrolase CelK [36]. As this domain displays over 80% identity with the corresponding module from CbhA, we assume that the latter domain also binds Ca²⁺. Recently, we found three Ca²⁺ ions in an Ig-GH9 combination from CbhA (I. A. Kataeva, unpublished work). All three ions are likely situated in the catalytic module GH9, as was found in the crystal structure of Ig-GH9 combination from endoglucanase CelD [37]. In CelD Ca²⁺ enhances thermal stability of the enzyme [38].

In the present study, the structures and stabilities of individual domains and domain combinations of CbhA have been compared. In the presence of Ca²⁺, all the constructs show high co-operativity of thermal unfolding and high intrinsic stability with mid-point transition temperatures of about 80 °C. These T_m values exceed the optimal growth temperature of *C. thermocellum* by 20 °C. Removal of Ca²⁺ destabilizes polypeptides in a different way, and to a different extent: it decreases the melting temperature of X1₁ by 20 °C; increases the number of transitions and decreases the co-operativity of unfolding of CBM3 by 7 °C; and slightly (2–4 °C) decreases T_m s of the other constructs. Despite the fact that individual X1₁, X1₂ and CBM3 are independently folded units, as indicated by the reversible unfolding, their interactions in domain combinations X1₁X1₂ and X1₁X1₂-CBM3 have been demonstrated. Thus X1₁ is significantly stabilized in the apo-X1₁X1₂ and apo-X1₁X1₂-CBM3 by interaction with X1₂. Thermal denaturations of holo- and apo-variants of X1₁X1₂ and X1₁X1₂-CBM3, comprising two and three different domains respectively, follow two-step mechanisms with no intermediate detected. This so-called ‘coupling phenomenon’ in folding/unfolding transitions when a modular protein is denatured by a simpler model than expected based on its multi-domain nature was described, in particular, for γ II-crystallin from calf eye lens [39] and xylanase

XynA from the extreme thermophilic bacterium *Thermotoga maritima* [9]. Absence of long flexible linkers rich in proline residues and hydroxylic amino acids (threonine and serine residues) often found between modules in glycoside hydrolases, suggests tighter contacts between the domains in CbhA and probably contributes to the observed one-or-nothing transitions of two- and three-domain composed constructs.

X1₁ and X1₂ display 26–48 % identity with the modules found mostly in chitinases from mesophilic bacteria [18]. The presence of Ca²⁺ has not been reported for other modules displaying similarity to X1₁ and X1₂, but the aspartic acid residue involved in binding of Ca²⁺ in X1₂ is conserved in these modules [18]. Thus we assume that they bind Ca²⁺. Ca²⁺ probably plays a leading role in the stabilization of the domains from mesophilic organisms. In X1₁ and X1₂, however, its role might be decreased due to the acquisition of other stabilizing increments during natural evolution of coding sequences.

X1₁ is strongly stabilized by Ca²⁺, whereas apo-X1₂ is almost as stable as holo-X1₂. Comparison of the amino acid sequences of X1₁ and X1₂, which display 33 % identity, shows that they do not have any cysteine residues, and consequently no disulphides, and contain an almost identical number of hydrophobic, polar and charged amino acid residues. One of the factors, which may influence the lower thermostability of apo-X1₁, is a decreased content of aromatic amino acid residues. Aromatic interactions are known to be one of factors responsible for the thermal stability of proteins [40,41]. A survey of the protein database shows that proteins from thermophiles usually have more aromatic pairs and aromatic clusters or additional aromatic residues in cluster than mesophilic proteins [41]. It is also possible that hydrophobic surface exposed to the solvent is larger in X1₁ than in X1₂.

Microbial glycoside hydrolases constitute a large group of extracellular modular-composed proteins. Some of them have complex structures with multiple modules [3]. Ca²⁺ is a usual component of these enzymes from both mesophilic and thermophilic micro-organisms. The remarkable feature of glycoside hydrolases is their activity against insoluble carbohydrates. Correspondingly, to reach their substrate the enzymes must be secreted outside the cytosolic compartment and act under extreme conditions. To overcome this environmental factor, proteins require molecular adaptation at different levels of their structure. Besides, domains performing particular biological functions and combined in one polypeptide chain must act in concert with elements of regulation. Binding of Ca²⁺ and formation of interacting surfaces between domains probably not only play a stabilizing role, but also are involved in the co-ordination of different modules. Finally, inter-domain interactions stabilizing domains in the absence of Ca²⁺ might be of importance during protein biosynthesis.

We thank Leon S. Dure (Department of Biochemistry and Molecular Biology, University of Georgia, U.S.A.) for interpretation of CD spectra, critical notes and suggestions. Support from the U.S. Department of Energy is gratefully acknowledged (grant number DE-FG02-93ER20127).

REFERENCES

- Janin, J. and Wodak, S. J. (1983) Structural domains in protein and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol.* **42**, 21–78
- Jaenicke, R. (1999) Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155–241
- Bayer, E. A., Shoham, Y. and Lamed, R. (2000) Cellulose-decomposing bacteria and their enzyme systems. In *The Prokaryotes, an Evolving Electronic Resource for the Microbiological Community* (Dvorkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. and Stackebrandt, E., eds), 3rd edn, pp. 1–41. Springer Verlag, New York
- Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-Active Enzymes server at <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>
- Tomme, P., Warren, R. A. J., Miller, Jr, R. C., Kilburn, D. G. and Gilkes, N. R. (1995) Cellulose-binding domains: classification and properties. In *Enzymatic Degradation of Insoluble Carbohydrates* (Saddler, J. N. and Penner, M. P., eds), pp. 142–163. American Chemical Society, Washington DC
- Little, E., Bork, P. and Doolittle, R. F. (1994) Tracing and spread of fibronectin type III domain in bacterial glycohydrolases. *J. Mol. Evolution* **39**, 631–643
- Wenk, M. and Jaenicke, R. (1998) Kinetic stabilization of a modular protein by domains interactions. *FEBS Letters* **438**, 127–130
- Wenk, M. and Jaenicke, R. (1999) Calorimetric analysis of the Ca²⁺-binding β -crystallin homolog protein S from *Myxococcus xanthus*: intrinsic stability and mutual stabilization of domains. *J. Mol. Biol.* **293**, 117–124
- Wassenberg, D., Schurig, H., Liebl, W. and Jaenicke, R. (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
- Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, J. P. and Harry, J. (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331**, 775–781
- Kataeva, I. A., Li, X.-L., Chen, H. and Ljungdahl, L. G. (1999) CelK – a new cellobiohydrolase from *Clostridium thermocellum* cellulosome: role of N-terminal cellulose-binding domain. In *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmura, K., Sakka, K., Karita, S., Hayashi, K., Kobayashi, Y. and T. Kimura eds), pp. 454–460. UniPublishers Co., Tokyo, Japan
- Shin, E.-S., Yang, M.-J., Jung, K.-H., Kwon, E.-J., Jung, J. S., Park, S. K., Kim, J., Yun, H. D. and Kim, H. (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
- Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J. Bacteriol.* **180**, 1709–1714
- Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M. G. A., Dias, F. M. V., Ferreira, L. M. A. and Bolam, D. N. (2001) *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry* **40**, 9167–9176
- Kataeva, I. A., Blum, D. L., Li, X.-L. and Ljungdahl, L. G. (2001) Do domain interactions of glycosyl hydrolases from *Clostridium thermocellum* cellulosome contribute to protein thermostability? *Protein Eng.* **14**, 167–172
- Notenboom, V., Boraston, A. B., Kilburn, D. G. and Rose, D. R. (2001) Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound form. *Biochemistry* **40**, 6248–6256
- Zverlov, V. V., Velikodvorskaya, G. V., Schwarz, W. H., Bronnenmeier, K., Kellermann, J. and Staudenbauer, W. L. (1998) Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *J. Bacteriol.* **180**, 3091–3099
- Kataeva, I. A., Seidel, R. D., III, Shah, A., West, L. T., Li, X.-L. and Ljungdahl, L. G. (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Appl. Environ. Microbiol.* **68**, 4292–4300
- Johnson, P. E., Creagh, A. L., Brun, E., Joe, K., Tomme, P., Haynes, C. A. and McIntosh, L. P. (1998) Calcium binding by the N-terminal cellulose-binding domain. *Biochemistry* **37**, 12772–12781
- Kataeva, I. A., Li, X.-L., Chen, H., Choi, S.-K. and Ljungdahl, L. G. (1999) Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum*: evidence of gene duplication and recombination. *J. Bacteriol.* **181**, 5288–5295
- Pace, C. N. and Scholtz, J. M. (1997) Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 291–321. Oxford University Press, Oxford
- Becktel, W. J. and Schellman, J. A. (1987) Protein stability curves. *Biopolymers* **26**, 1859–1877
- Burstein, E. A. (1976) Intrinsic Protein Fluorescence: Origin and Applications. In series: *Biophysics 7*, VINITI, Moscow
- Permyakov, E. A., Yarmolenko, V. V., Emelyanenko, V. I., Burstein, E. A., Closset, J. and Gerday, C. (1980) Fluorescence studies of the calcium binding to whiting (*Gadus merlangus*) parvalbumin. *Eur. J. Biochem.* **109**, 307–315
- Kuznetsova, I. M., Stepanenko, O. V., Turoverov, K. K., Zhu, L., Zhou, J. M., Fink, A. L. and Uversky, V. N. (2002) Unraveling multistate unfolding of rabbit muscle creatine kinase. *Biochim. Biophys. Acta* **1596**, 138–155

- 26 Bushmarina, N. A., Kuznetsova, I. M., Biktashev, A. G., Turoverov, K. K. and Uversky, V. N. (2001) Partially folded conformations in the folding pathway of bovine carbonic anhydrase II: a fluorescence spectroscopic analysis. *ChemBiochem* **2**, 813–821
- 27 Bolotina, I. A. and Lugauskas, V. Yu. (1985) Determination of the secondary structure of proteins from circular dichroism spectra. IV. Contribution of aromatic amino acid residues into circular dichroism spectra of proteins in the peptide region. *Mol. Biol. (Moscow)* **19**, 1409–1421
- 28 Manning, M. C. and Woody, R. W. (1989) Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry* **28**, 8609–8613
- 29 Chakrabarty, A., Kortemme, T., Padmanabhan, S. and Baldwin, R. L. (1993) Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**, 5560–5565
- 30 Uversky, V. N. and Ptitsyn, O. B. (1996) Further evidence on the equilibrium 'pre-molten globule state': four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255**, 215–228
- 31 Abramov, V. M., Vasiliev, A. M., Vasilenko, R. N., Kulikova, N. L., Kosarev, I. V., Khlebnikov, V. S., Ishchenko, A. T., MacIntyre, S., Khurana, R., Gillespie, J. R. et al. (2001) Structural and functional similarity between *Yersinia pestis* capsular protein Caf1 and human interleukin-1 β . *Biochemistry* **40**, 6076–6084
- 32 Ptitsyn O. B. (1995) Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83–229
- 33 Permyakov, E. A. (1993) Calcium-Binding Proteins. pp.1–190, Nauka, Moscow
- 34 Mayr, E. M., Jaenicke, R. and Glockshuber, R. (1997) The domains in γ -crystallin: identical fold – different stabilities. *J. Mol. Biol.* **269**, 260–269
- 35 Lytle, B. L., Volkman, B. F., Westler, W. H. and Wu, J. H. D. (2000) Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. *Arch. Biochem. Biophys.* **379**, 237–244
- 36 Kataeva, I. A., Seidel, III, R. D., Li, X.-I. and Ljungdahl, L. G. (2001) Properties and mutation analysis of the CelK cellulose-binding domain from the *Clostridium thermocellum* cellulosome. *J. Bacteriol.* **183**, 5288–5295
- 37 Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeysens, M., Béguin, P., and Aubert, J.-P. (1992) Crystal structure of a thermostable bacterial cellulose-degrading enzyme. *Nature (London)* **357**, 89–91
- 38 Chauvaux, S., Souchon, S., Alzari, P. M., Chariot, P. and Beguin, P. (1995) Structural and functional analysis of the metal-binding sites of *Clostridium thermocellum* endoglucanase CelD. *J. Biol. Chem.* **270**, 9757–9762
- 39 Rudolph, R., Siebendritt, R., Nessler, G., Sharma, A. K. and Jaenicke, R. (1990) Folding of an all- β protein: independent domain folding in γ -II-crystallin from calf eye lens. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4625–4629
- 40 Burley, S. K. and Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science (Washington, D.C.)* **229**, 23–28
- 41 Kannan, N. and Vishveshwara, S. (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **13**, 753–761

Received 17 October 2002/30 January 2003; accepted 5 February 2003

Published as BJ Immediate Publication 5 February 2003, DOI 10.1042/BJ20021621