A carbohydrate binding module as a diversity-carrying scaffold

L.Cicortas Gunnarsson[1], E.Nordberg Karlsson[2], A.-S.Albrekt[1], M.Andersson[3], O.Holst[2] and M.Ohlin[1,4]

[1]Department of Immunotechnology, Lund University, PO Box 7031, S-220 07 Lund, [2]Department of Biotechnology, Lund University, PO Box 124, S-221 00 Lund and [3]Alligator Bioscience, Scheelevägen 19 A, S-223 70 Lund, Sweden

[4]To whom correspondence should be addressed.
E-mail: mats.ohlin@immun.lth.se

The growing field of biotechnology is in constant need of binding proteins with novel properties. Not just binding specificities and affinities but also structural stability and productivity are important characteristics for the purpose of large-scale applications. In order to find such molecules, libraries are created by diversifying naturally occurring binding proteins, which in those cases serve as scaffolds. In this study, we investigated the use of a thermostable carbohydrate binding module, CBM4-2, from a xylanase found in *Rhodothermus marinus*, as a diversity-carrying scaffold. A combinatorial library was created by introducing restricted variation at 12 positions in the carbohydrate binding site of the CBM4-2. Despite the small size of the library (1.6×10^6 clones), variants specific towards different carbohydrate polymers (birchwood xylan, Avicel and ivory nut mannan) as well as a glycoprotein (human IgG4) were successfully selected for, using the phage display method. Investigated clones showed a high productivity (on average 69 mg of purified protein/l shake flask culture) when produced in *Escherichia coli* and they were all stable molecules displaying a high melting transition temperature (75.7 ± 5.3°C). All our results demonstrate that the CBM4-2 molecule is a suitable scaffold for creating variants useful in different biotechnological applications.

Keywords: binding specificity/carbohydrate binding module/molecular library/phage selection/scaffold

Introduction

Nature has created a great diversity among binding proteins, in terms of both affinity and specificity. This has been evolution's way of solving the environmental challenges, such as substrate accessibility at the different pH and temperature scales faced by these molecules. Recently, synthetic approaches have been used to establish such diversity, most often by creating large random mutant peptide or protein libraries and then selecting for the desired binding specificities (Nygren and Uhlén, 1997). A feature that all these molecules have in common is a stable and well defined structure, a scaffold, on which the diversification process can be built (Skerra, 2000a). The biggest and most successful efforts so far have been put on diversifying antibodies (Söderlind *et al*., 2001), a large group of naturally diverse binders, but also other types of binding proteins, such

as lipocalins (Skerra, 2000b) and a single protein A domain (Nord *et al*., 1997) have been used in this kind of molecular engineering.

Except for the well known group of antibodies, there are also other types of proteins displaying natural binding diversity, such as, for example, the carbohydrate binding modules (CBMs), which are non-catalytic modules connected to a variety of glycoside hydrolysing or modifying enzymes. The diversity of CBMs is demonstrated by their different ligand specificities and, based on the topology of the binding site, CBMs have been assigned into three major types: A, B and C (Boraston *et al*., 1999). Type A modules, with a flat binding surface, bind to insoluble crystalline glucans (Linder *et al*., 1996; Simpson and Barras, 1999). Type B modules, displaying a binding cleft, have affinity for free single carbohydrate chains (Simpson *et al*., 2002). Type C modules, which possess a solvent-exposed binding slot, have the ability to bind mono- and disaccharides (Boraston *et al*., 2001). The driving force for binding is determined by enthalpic and entropic forces, and their respective contribution varies depending on the specific carbohydrate–CBM interaction (Gilbert *et al*., 2002). The residues involved in binding are consequently also varying, and thus far aromatic residues are the most established mediators of carbohydrate–protein interactions (Boraston *et al*., 1999). Altering the binding ability of a single CBM is a relatively unexplored area. The few reports that have been published have in principle utilized the rare examples, occurring in nature, of closely related CBMs with clear differences in binding specificity. In this way, a type A and a type B module, both classified into CBM family 2 have been specificity swapped solely by a single point mutation in the binding site (Simpson *et al*., 2000). Such investigations, however, demand high similarities, in order to restrict the putative target residues to be modified.

In this work, we have chosen to explore the potential for diversification of a CBM by constructing a combinatorial library followed by a selection procedure using different target substrates, an approach that only rarely (Smith *et al*., 1998; Lehtiö *et al*., 2000) has been employed for CBMs. For this purpose we used CBM4-2 of the *Rhodothermus marinus* xylanase Xyn10A (Nordberg Karlsson *et al*., 1997), a protein module with a molecular weight of 18 kDa, as a template. It is a well characterized module, with a known β jelly-roll structure (Simpson *et al*., 2002), and defined substrate affinity and specificity (Abou-Hachem *et al*., 2000). This module displays the type B topology of the binding site (Figure 1) and it is also thermostable (Abou-Hachem *et al*., 2002), being a suitable scaffold for evolution of new binding properties.

Materials and methods

Strains and vectors

Escherichia coli strain Top10F′ (Invitrogen, Paisley, UK) was used as host for construction of the library and for phage

L.Cicortas Gunnarsson *et al.*

display work. The phagemid used was a variant of pFab5c.His (Engberg *et al*., 1996) carrying the part of gene III encoding only the final C-terminal domain of M13 protein 3. Genes coding for selected clones were inserted in the expression vector pET-22b(+) (Novagen, Madison, WI, USA) and then transformed into a cloning host, *E.coli* strain XL1-Blue (Stratagene, La Jolla, CA, USA). Production of recombinant protein was achieved using *E.coli* strain BL21(DE3) (Novagen).

CBM library construction

The gene encoding the CBM4-2 from Xyn10A (GenBank accession No. AY347870) was initially amplified from a pre-existing construct cloned into the pET-25b(+) vector (Abou-Hachem *et al*., 2000) using the following primers: 5′-GGCCGATTCATTAATGCAGCTGGCCCAGCCGGCCAT-GGCCCTTGTCGCCAACATCAACGGT-3′ and 5′-GCCTT-TAGCGTCAGACTGTAGTGCGGCCGCGTCCACAATGG-CCAGGCCATC-3′. The product was cloned in the phagemid vector pFab5c.His and sequenced to confirm the correct gene sequence. Restricted diversity was created at 12 residues (W69 → F, H, L, W, Y; E72, Q111, E112, E118 → D, E, H, Q; F76, F110, H117, H146, Y149 → F, H, L, Y; R115 → K, R; R142 → R, S) in the binding site of CBM4-2 (Figure 1) [amino acid numbering according to Simpson *et al*. (2002)]. Mutations were introduced into the CBM gene by an initial amplification of four gene fragments (fragments 1a and 1b are two variants of the same fragment introducing different mutations in position 69) using gene- and vector-specific primers (Table I) (30 cycles, 94°C for 1 min, 50 or 55°C for 1 min, 72°C for 2 min). The

Fig. 1. The structure of CBM4-2 (PDB 1k45) showing the targeted residues for mutation in the binding cleft in two slightly different projections. The modified residues known to undergo large chemical shift (dark grey) upon titration with xylo-oligosaccharide (Simpson *et al*., 2002) are spelled out in (**A**). Other modified residues (light grey) are spelled out in (**B**).

fragments were assembled by an overlap-extension polymerase chain reaction (PCR) (eight cycles, 94°C for 1 min, 52°C for 1 min, 72°C for 2 min) and the resulting PCR product was first amplified using the vector-specific primers (Table I) (25 cycles, 94°C for 1 min, 55°C for 1 min, 72°C for 2 min) before it was digested by *Sfi*I and *Not*I and inserted back into the phagemid vector, digested with the same restriction enzymes. The size of the library after transformation into competent *E.coli* Top10F′ was calculated to be 1.6×10^6 clones.

Phage stocks

An exponentially growing culture of cells from the constructed library or of a selected clone [2×YT, 1% (w/v) glucose, 100 µg/ml ampicillin, 10 µg/ml tetracycline at OD_{600} ~0.5] was infected with helper phages VCSM13 (multiplicity of infection: ~20) (Stratagene) for 30 min at 37°C. The medium was changed (2×YT, 100 µg/ml ampicillin, 10 µg/ml tetracycline, 50 µg/ml kanamycin and 0.25 mM isopropyl-β-D-thiogalacto-side) and the cultures were grown at 30°C overnight. After centrifugation, supernatants were filtered through 0.45 µm filters and phage particles were concentrated by polyethylene glycol/NaCl precipitation if they were to be used for further selection attempts. The precipitated phages were dissolved in phosphate-buffered saline (PBS) containing 0.1% bovine serum albumin (BSA) (Sigma-Aldrich Inc., St Louis, MO, USA) and stored at 4°C.

Phages, prepared as described above but displaying a single-chain antibody fragment (scFv) recognizing fluorescein (kindly provided by Helena Persson, Department of Immunotechnology, Lund University) were used as a negative control in this study.

Selections

Phage selections were performed either on insoluble carbohydrate substrates [the birchwood xylan (Sigma-Aldrich) part that remained insoluble after two overnight washes, Avicel (~50% amorphous and ~50% crystalline cellulose) (Merck, Darmstadt, Germany) and ivory nut mannan (Megazyme, Bray, Ireland)] or on Dynabeads coated with a human monoclonal IgG4 (Pan Mouse IgG; Dynal A/S, Oslo, Norway). Prior to selection the substrate (~50 mg of a carbohydrate or 2×10^7 Dynabeads) was washed twice with 500 µl of PBS. Phage stocks, 450–500 µl containing ~10^{11} phage particles, displaying different variants of the CBM molecule were added to 450–500 µl of a substrate suspension in PBS containing 2% BSA and 0.1% (v/v) Tween-20. The mixture was incubated for 1 h on an end-to-end rotator at room temperature or, in the case of selections at elevated
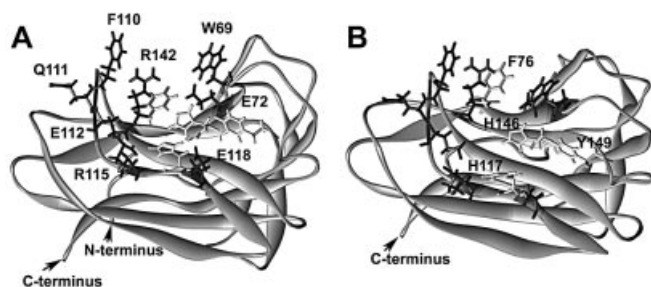
Table I. Oligonucleotide primers used for library construction

Name	Sequence
Fragment 1-5′	5′-TTTCACACAGGAAACAGCTATG[a]
Fragment 2-5′	5′-GAACGTACGTCCGGGCGTGA
Fragment 3-5′	5′-GCAGATCACGACCGGAGTGGC
Fragment 4-5′	5′-GCGGCCAACGTCGGCAATAC
Fragment 1a-3′	5′-CACGCCCGGACGTACGTTCACCGGRWGGCCGTCGCNTSGATGTCRWGGGGGTTGTTGCCCACCCCGT[b]
Fragment 1b-3′	5′-CACGCCCGGACGTACGTTCACCGGRWGGCCGTCGCNTSGATGTCCCAGGGGTTGTTGCCCACCCCGT[b]
Fragment 2-3′	5′-GCCACTCGGTCGTGATCTGCTGNTSRWRCAGYYTCCCGTANTSNTSRWRCGACTGGTTCCCCACCG[b]
Fragment 3-3′	5′-ATTGCCGACGTTGGCCGCAWRGCCAAARWRGATCGGCGCNCTAATGACCGTCTCCTGATCA[b]
Fragment 4-3′	5′-GCCTTTAGCGTCAGACTGTAG[a]

[a]Vector-specific primer.
[b]R = A or G; S = C or G; W = A or T; N = A, C, G or T.

214

temperature, in a water bath with repeated manual vortexing. The substrate was subsequently washed four times with PBS containing 1% BSA and 0.05% (v/v) Tween-20 and twice with PBS. Bound phages were eluted using 100 μl of trypsin (0.5 mg/ml) (Invitrogen) (Johansen *et al.*, 1995) by incubation for 30 min at room temperature with shaking. After removing the substrate the activity of trypsin was neutralized by the addition of 100 μl of aprotinin (0.1 mg/ml) (Roche Diagnostic Corporation, IN, USA). A few microlitres of the elution solution were saved for titration experiments while the rest was used to transfect exponentially growing *E.coli* Top10F′ (~10 ml of OD_{600} ~0.5) for preparation of new phage stocks, as described above. Both input and output phages were titrated on exponentially growing *E.coli* Top10F′. The selection process was repeated three to four times before clones were picked for analysis.

Enzyme-linked immunosorbent assay (ELISA)

Ninety-six-well ELISA plates were coated with 1 μg/ml human monoclonal IgG4 (Dynal) in PBS at 4°C overnight. After washing the plates with washing buffer [154 mM NaCl and 0.05% (v/v) Tween-20 in PBS], the phages were added diluted in a blocking buffer [PBS containing 1% BSA and 0.05% (v/v) Tween-20] and the plates were incubated for 2 h at room temperature. Washing of the plates was followed by addition of rat anti-M13 protein 8 diluted in blocking buffer for 1 h at room temperature. Plates were washed and horseradish peroxidase (HRP)-conjugated rabbit-anti-mouse immunoglobulins that cross-react with rat IgG (DAKO A/S, Glostrup, Denmark) diluted in blocking buffer were added to the wells and incubated for 1 h at room temperature. For detection, substrate (0.67 mg *o*-phenylenediamine/ml 35 mM citrate and 67 mM phosphate buffer pH 5 and 0.012% H_2O_2) was added. The reaction was stopped by addition of H_2SO_4 to a final concentration of 0.6 M and the absorbance was recorded at 490 nm.

In phage-ELISA experiments performed to confirm the display of functional CBMs, plates were coated with 2 μg/ml rabbit anti-CBM4-2 (immunoglobulin fraction from serum drawn from a rabbit immunized with purified CBM4-2). After incubating with the phages as above, HRP-conjugated anti-M13 antibody (Amersham Pharmacia Biotech Inc., Piscataway, NJ, USA) diluted in blocking buffer was added to the wells and incubated for 1 h at room temperature. The detection method followed the one described above.

To confirm binding of soluble variants of IgG4 specific CBM, purified proteins (1 μg/ml in PBS) were coated onto plates at room temperature overnight and their binding to human IgG4 was assessed by ELISA. After washing the plates, human IgG4, 1 μg/ml diluted in blocking buffer, was added and the plates were incubated for 1 h at room temperature. The bound human IgG4 molecules were detected by adding HRP-conjugated goat anti-human IgG antibody (Zymed, South San Francisco, CA, USA) for 1 h at room temperature and using the detection method described above.

Protein production and purification

Genes encoding CBM variants were cloned in between the *Nde*I and *Xho*I sites found in the pET-22b(+) vector. Production of CBM carrying a C-terminal His_6 tag in BL21(DE3) and purification of the soluble protein using metal-ion-affinity chromatography have been described earlier (Abou-Hachem *et al.*, 2000). The only differences in this work were that 20 mM sodium phosphate buffer (pH 7.4)/20 mM imidazole/0.75 M NaCl was used as binding buffer and the chromatography columns were packed with Ni-NTA (Qiagen, Hilden, Germany). The concentration of each purified protein was determined spectrophotometrically at A_{280nm} using extinction coefficients individually calculated for each CBM variant.

Affinity electrophoresis (AE)

The AE method was used in the Bio-Rad (Hercules, CA, USA) mini-gel apparatus as earlier described for the wild-type CBM4-2 (Abou-Hachem *et al.*, 2000). Purified CBM variants (3 μg per gel) were separated at room temperature and 90 V on two different native gels, with or without 0.02% (w/v) oat spelt xylan (Sigma-Aldrich). A Kaleidoscope prestained standard (Bio-Rad) was used as a negative, non-interacting, control and the proteins were detected by staining with Coomassie Brilliant Blue.

Phylogenetic tree

A consensus tree was computed using the phylogenetic packages MEGA (Kumar *et al.*, 2001) and PHYLIP 3.2 (Felsenstein, 1993) on a data set consisting of 165 amino acids from 37 different clones. Three different phylogenetic algorithms were used to investigate the robustness of the tree: neighbour-joining (NJ) (Felsenstein, 1978), minimum evolution (ME) (Edwards, 1996) and maximum parsimony (MP) (Fitch, 1971). For MP, 500 bootstrap replicas (Felsenstein, 1985) were performed, while for NJ and ME, 5000 were done. The alignment contained no gaps or otherwise ambiguously aligned regions. The tree was produced using TreeView (Page, 1996).

Differential scanning calorimetry (DSC)

DSC measurements were performed on a MicroCal differential scanning calorimeter (MicroCal Inc., Northampton, MA, USA) with a cell volume of 0.5072 ml. All samples were degassed for 15 min at room temperature prior to scanning. Baseline scans were collected with buffer in both the reference and sample cells and later subtracted from sample scans. Protein samples (0.5 mg/ml) in 20 mM sodium phosphate buffer pH 7.5 were scanned in the temperature range 25–100°C at a rate of 1°C/min.

Circular dichroism (CD)

CD spectroscopy was carried out at 20°C on a J-720 spectropolarimeter (Jasco Inc., Easton, MD, USA) in a 2 mm cuvette with a protein concentration of ~0.1 mg/ml in 20 mM sodium phosphate buffer (pH 7.5). Spectra were scanned six times from 250 to 205 nm at a scan speed of 10 nm/min, a time constant of 8 s, a bandwidth of 1 nm, resolution of 1 nm and sensitivity of 20 mdeg. The scans were combined to produce a final spectrum that is presented as mean residue molar ellipticity, calculated by using the mean residue weight of each CBM.

Results

Construction of the combinatorial library

The module CBM4-2 from the *R.marinus* xylanase Xyn10 A is a type B CBM displaying a cleft capable of enthalpically driven binding to different xylans and β-glucans. It is also relatively unusual in its capacity to bind both to xylan and to amorphous cellulose, although the affinity for the latter is much lower (Abou-Hachem *et al.*, 2000). This raised the possibility that additional changes in the binding site could further diversify the binding specificity of CBM4-2. High productivity in *E.coli*

(Abou-Hachem *et al.*, 2000) and high thermostability (Abou-Hachem *et al.*, 2002) were other properties that made this CBM a suitable candidate for use in many applications and another reason why molecular evolution of this module was considered. After demonstrating that it could be displayed on phages and that the displayed protein could be bound to an anti-CBM polyclonal antibody (Figure 2) we decided to explore the potential of this module as a diversity-carrying scaffold for selection of new binding specificities.

A combinatorial library was constructed by limited substitutions of 12 residues around the binding site of CBM4-2 (Figure 1), identified by NMR to undergo large chemical shifts upon titration with xylo-oligosaccharide (W69, E72, F110, Q111, E112, R115, E118, R142) (Simpson *et al.*, 2002), or located very close to residues showing such shifts (F76, H117, H146 and Y149). The substitutions introduced in this library were mainly restricted to related residues, in order not to destabilize the structure to a large extent and to maintain similar but modified interactions, introducing the possibility to improve the fit for other target substrates. Such modifications tended to preserve the aromatic nature of residues 69 and 110 believed to be important for the binding. It also largely preserved the hydrophilic and charged character of other parts of the surface as such interactions fit with the enthalpic changes observed upon substrate binding to type B modules (Pell *et al.*, 2003).

Degenerate primers were used to introduce the designed mutations, creating a library with a theoretical maximum size of 5.6×10^6 variants. The library obtained after transformations contained 1.6×10^6 clones, of which 86% had an insert. Sequencing of 11 randomly picked clones showed that six of

11 were in-frame and six of 11 had mutations not intended by the primer design but probably introduced by the polymerase. Despite the small number of clones sequenced, the diversity of the library was confirmed by identifying 42 of 45 mutated residue variants.

Selections

Initially we observed that not only the wild-type CBM but also the created library was successfully displayed on phages (Figure 2), as fusion proteins together with phage coat protein 3. This allowed us to use it in the well established selection procedure of phage display (Smith, 1985) for selection of functional binders.

In order to find CBM variants with new binding specificities, the phage-displayed library was subjected to several series of selections, performed on different substrates. Three carbohydrate polymers, birchwood xylan, Avicel (~50% amorphous and ~50% crystalline cellulose) and ivory nut mannan, as well as a glycoprotein, human IgG4, were used for this purpose. Three to four selection rounds per substrate resulted in most cases in higher selection ratios (output phages:input phages) (Table II), a fact that often indicates an enrichment of target-specific binders. This increase in selection ratio was most apparent with xylan and human IgG4, as background binding on these substrates was much less pronounced.

Binding specificity of selected clones

Having demonstrated an enrichment of some variants from the library during the selections with different targets, it became important to determine whether these variants had any significant difference in binding specificity. We therefore determined the ability of selected clones, displayed on phages, to bind to the different insoluble carbohydrate substrates or human IgG4, which in turn was bound onto either a polystyrene surface or on paramagnetic beads. Figure 3A and B illustrates that CBM variants selected on a specific carbohydrate polymer have enhanced recognition of that substrate as compared with another carbohydrate structure. Similarly, clones selected on IgG4 specifically recognized this target and had lost their ability to bind insoluble xylan (Figure 3C and D).

The binding studies using phage-displayed CBMs were later confirmed, when some of the clones were produced in soluble form (see below) and assessed for their binding properties (Figure 4). Affinity electrophoresis using xylan as a substrate (Figure 4A) indicated that clones selected on Avicel or mannan had lost some or all of their affinity to xylan while clones selected on xylan had high but diverse affinity for this substrate. Clones selected on IgG4 proved again to be specific for the selection target (Figure 4C) and to have completely lost their affinity for xylan (Figure 4A). We thus conclude that the approach used to create diversity in the CBM molecule allowed for different specificities to be established.
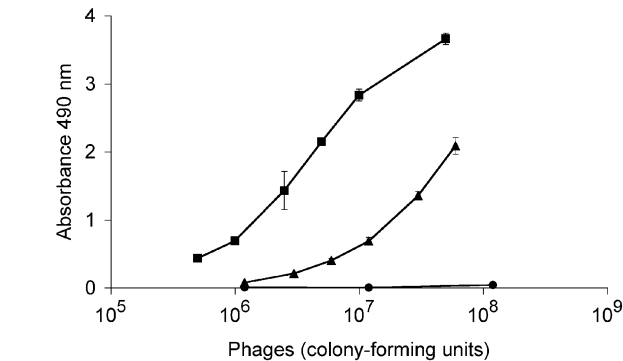

Fig. 2. Display of functional CBMs, both wild-type (squares) and library variants (triangles), on phages, as confirmed by phage-ELISA. Plates were coated with rabbit anti-CBM4-2 antibody. Phage particles displaying a scFv molecule (circles) were used as a negative control. The error bars represent the standard deviation of experiments performed in quadruplicates.

Table II. Selection ratios (output phages:input phages) achieved using different substrates at different temperatures

Selection round	Selections at room temperature			Human IgG4	Selections at 60°C[a]		
	Xylan	Avicel	Mannan		Xylan	Avicel	Mannan
1	1:2 000 000	1:43 000	1:4700	1:3 000 000	1:220 000	1:79 000	1:23 000
2	1:100 000	1:27 000	1:4400	1:20 000	1:310 000	1:300 000	1:16 000
3	1:100	1:26 000	1:100	1:100	1:220 000	1:150 000	1:7000
4		1:400			1:70 000	1:100 000	1:30

[a]The fourth selection in this series was performed at room temperature.
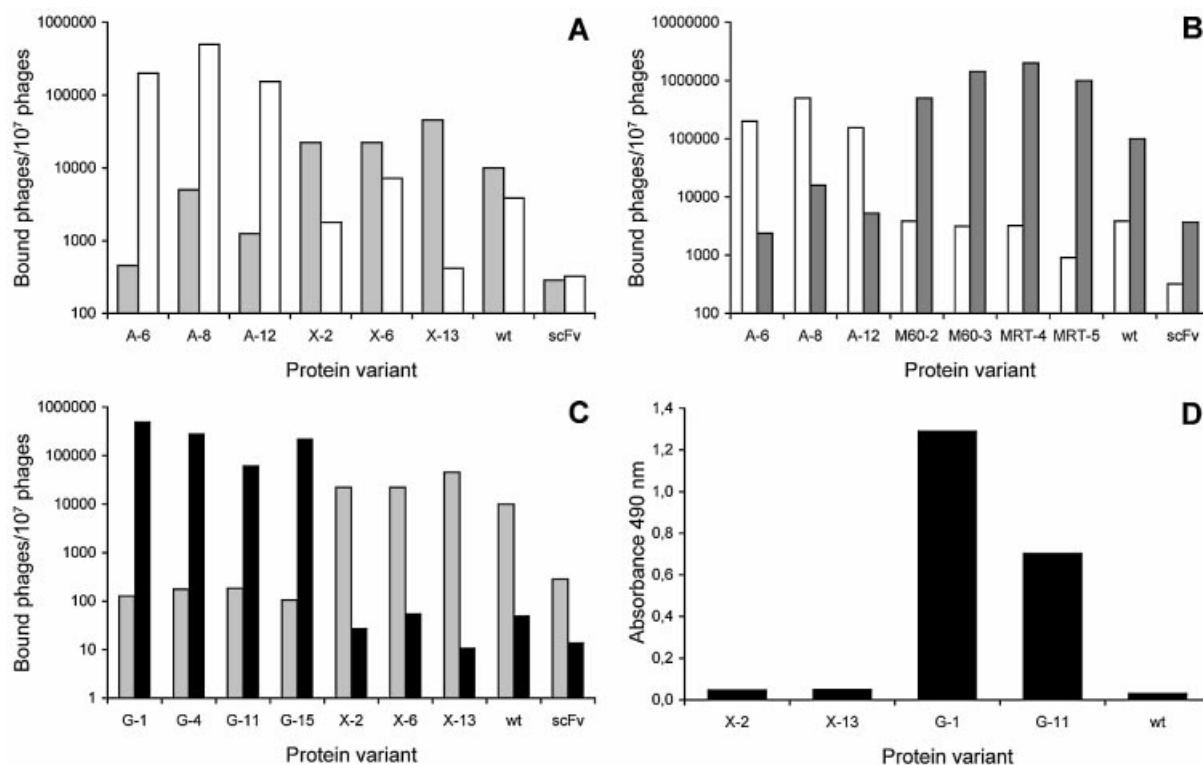
216

Fig. 3. Binding studies on different substrates, birchwood xylan (light grey), Avicel (white), ivory nut mannan (dark grey) and human IgG4 (black), using selected clones and wild-type (wt) CBM4-2 displayed on phages. Experiments included phages displaying a scFv recognizing fluorescein as a negative control. (**A**) Selection on xylan and Avicel selected for variants (X- for xylan and A- for Avicel) with preferred recognition of the substrate used for selection over the other substrate. (**B**) Selection on ivory nut mannan selected for variants with enhanced binding to mannan as opposed to Avicel. This was true irrespective of whether the clones were selected at room temperature (MRT-) or initially at 60°C (M60-). (**C**) Variants of CBM selected on human IgG4 displayed on phages could not bind to insoluble xylan but to monoclonal human IgG4 immobilized on Dynabead particles. The reverse was true for variants selected on xylan and displayed on phages. Similarly, CBMs selected on human IgG4 molecule but not the other CBMs could bind to the monoclonal antibody (human IgG4) immobilized onto microtitre plates (**D**).

Sequence studies

In order to further understand how the evolution process selected for specific variants against the different targets, clones (~15–20) from the final round of selection on each substrate were sequenced. Although some of them were identical, several sequence types of CBMs were identified from each selection (Figure 5) (complete gene sequences are available at GenBank, accession Nos AY534546–AY534578). Except for the introduced mutations, some other spontaneous ones occurred as well in the sequences of selected clones, just as they were found among randomly picked clones (see above). A phylogenetic tree calculated from the amino acid sequences of the selected clones demonstrated that clones selected on one target often grouped together to form one or a few neighbouring groups in the tree (Figure 6).

The sequence studies also revealed that among the 12 target residues there were positions where only one specific amino acid of all possible mutant alternatives had been selected for. These positions and their location in the binding site varied between the clones selected on different substrates (Figure 7), thus indicating that the binding specificity in each case could be coupled to certain key residues in the binding site.

Production and purification of single clones

An important feature of a diversity-carrying scaffold is that selected variants can be produced in large quantities, especially if these are to be used in large-scale applications. Previous work had shown that wild-type CBM is efficiently produced in *E.coli* (Abou-Hachem *et al.*, 2000), but it was not known what effect evolution and selection may have had on the productivity of the selected CBMs. In order to investigate this property, a total of 14 clones of CBM4-2 chosen from all four substrate selections plus the wild-type were therefore produced and on average 67 mg of a mutated CBM/l shake flask culture (range 38–111 mg/l, wild-type 40 mg/l) were obtained after purification. Six of these production schemes were reproduced, this time the average being 70 mg of a mutated CBM/l shake flask culture (range 58–84 mg/l, wild-type 72 mg/l) and showing not surprisingly, due to the lack of optimization, the presence of batch variation for some of the CBMs, including the wild-type. Nevertheless, despite the variation between different production batches it was shown that the diversification and selection approach had no tendency to select for variants with poor productivity properties from this library.

Stability and structure of selected CBMs

Apart from high productivity, high stability of the scaffold is another advantageous property when its variants are to be used in many different applications. The wild-type CBM is a thermostable molecule (Abou-Hachem *et al.*, 2002) and in order to preferentially select for variants with retained thermostability, selections were also performed on the carbohydrate substrates at 60°C, a temperature at which phage particles remain stable. After three selection rounds no
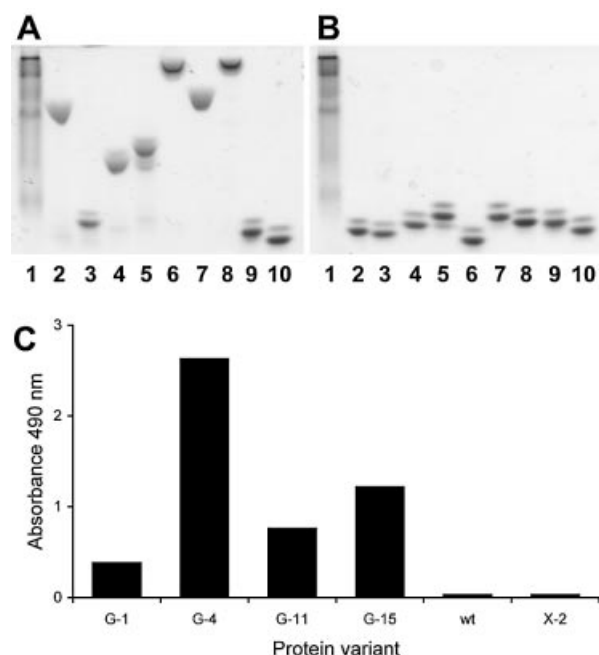
Fig. 4. Binding studies using soluble variants of CBM. Non-denaturing affinity electrophoresis in the presence (**A**) or in the absence of oat spelt xylan (**B**). Proteins: lane 1, Kaleidoscope prestained standard; lanes 2 and 3, mannan-specific clones (MRT-4 and MRT-5, respectively); lanes 4 and 5, Avicel-specific clones (A-6 and A-12, respectively); lane 6, wild-type CBM4-2; lanes 7 and 8, xylan specific clones (X-2 and X-6, respectively); and lanes 9 and 10, human IgG4 specific clones (G-1 and G-4, respectively). (**C**) The binding of human IgG4 molecules to microtitre plates coated with different CBM variants selected on IgG4 (G-1, G-4, G-11 and G-15) but not to wild-type CBM4-2 or X-2 selected on xylan.

remarkable differences were observed in this study, meaning that the mutations introduced in the selected variants of CBM allowed them to bind to different substrates and still retain the same overall structure. In summary, these investigations demonstrated that CBM4-2 is a stable scaffold with a capacity to harbour diversity suitable for molecular evolution of the binding properties.

Discussion

Protein scaffolds have a great potential for identification of molecules with particular properties for use in specific biotechnological and biomedical applications. Many such scaffolds have been explored, with antibodies being the most notable but not the only example. For many applications involving extreme conditions, such as high temperatures, conventional scaffolds do not fulfil the necessary requirements. We have now demonstrated that the CBM CBM4-2 from the thermostable *R.marinus* xylanase Xyn10A is a suitable carrier of diversity in the search for specific binders with a high-stability character.

The minimum requirement on a protein, in order to be suitable as a diversity-carrying scaffold in molecular libraries is its ability to be displayed properly in a chosen display system. The fact that the CBM is native to *R.marinus*, a prokaryotic organism phylogenetically placed among the Gram-negative bacteria, suggests that it should be appropriate for display systems relying on *E.coli*. This was indeed also shown to be the case, the wild-type CBM being functionally displayed on filamentous phages, allowing us to use the generally known phage display system (Smith, 1985) as a selection method. Also, the combinatorial library seemed to be well expressed on phages. The assay signals were weaker in this case compared with those received from phages displaying only the wild-type CBM, probably due to a combination of different factors. First, the fact that some of the clones were out of frame reduces the overall display level. Nevertheless, some of this diversity may be accessible as certain display systems even tend to positively select for out-of-frame deletions (Jacobsson and Frykberg, 1995), possibly as a consequence of toxicity of the fusion protein. However, none of the clones selected from this library showed any sequence changes of that kind, demonstrating the usefulness of CBM as a display scaffold in *E.coli* and showing that it is well tolerated by this environment. Another possible reason for the weaker signals of the library as opposed to the wild-type sequence may be that some of the displayed clones were not well recognized by the polyclonal antibody used in the assay. This loss of recognition could be due to modifications of the epitopes displayed by the CBMs, producing an artificially low level of detectable display in this assay system. Thirdly, it cannot be ruled out that not all sequence variants were efficiently folded in the unselected library. The fact that a wide range of sequence substitutions was found in all modified positions but one suggests though that much of the introduced diversity is compatible with a proper protein fold. The one residue rarely found to be diversified among the selected clones was R142 (only three of 33 selected clones had a serine in position 142 compared with seven of 11 for unselected clones). Either this residue is not able to harbour diversity or the rather large difference between arginine and serine prevented efficient incorporation of the latter in functional proteins. This agrees with studies of diversity in the Z-domain derived from protein A that similarly

enrichment effect was seen in any of the cases, and this is why a fourth selection round was performed, this time at room temperature (Table II). Although no substantial increase in selection ratio was seen with xylan and Avicel, selection on mannan clearly demonstrated an enrichment of specific binders also after initial selection at 60°C. This suggests that the previous rounds of selections on this substrate had enriched for a set of binders that at the lower temperature could be more efficiently identified.

Irrespective of the selection temperature used, high thermostability may still be an attractive feature of selected clones. In order to assess this matter, the thermal transition temperature (T_m) of 14 clones, selected on four substrates and purified, was determined by DSC. Although the T_m for these clones was lower than that of the wild-type CBM ($T_m = 89.3°C$), these were stable modules with a melting temperature of 75.7 ± 5.3°C (mean ± standard deviation). No major difference in T_m was found between clones selected on mannan only at room temperature or clones first selected at 60°C and then at room temperature (all had a T_m between 77 and 80.4°C). Thus, no selection pressure was necessary for finding thermostable binders, indicating that the structure of the CBM is a stable scaffold, as all the variants created in this library expressed high thermostability.

The stability of a module is related to its physical characteristics, such as the structure that the protein adopts. CD spectra were gathered (Figure 8) for the selected clones in order to register any structural changes compared with the known β-sheet structure of the wild-type CBM4-2. No

```
              67                78    108              120    140            150      other
wt    N P W D I E A T A F P V   Q S F Q E Y G R L H E Q Q   V I R A P I H F G Y A   substitutions
X-2   . . F N . Q . . . L . .   . . L D . . . . . . H . .   . . . . . . . . . . .   W91R
X-3   . . . . . Q . . . Y . .   . . Y . . . . . . L Q . .   . . . . . . L . . . .
X-5   . . . . . Q . . . Y . .   . . Y E . . . . . . H . .   . . . . . . F . . . .   I6T
X-6   . . . . . Q . . . L . .   . . Y L . . . . . . H . .   . . . . . . . . . H .
X-13  . S . . . . . . . . . .   . . . D Q . . . . . Q . .   . . . . . . . . . . .
X-14  . . . . . Q . . . . . .   . . . H . . . . . . Q . .   . . . . . . . . . H .   V101A
X-15  . . . . . Q . . . Y . .   . . Y E . . . . . . L Q . .  . . . . . . L . . . .
X-18  . . . . . H . . . L . .   . . Y . . . . . . . Q . .   . . . . . . . . . . .
X-19  . . . . . Q . . . L . .   . . Y E . . . . . . H . .   . . . . . . F . . . .
X-20  . . . . . Q . . . Y . .   . . Y D . . . . . . H . .   . . . . . . . . . . F   L30P
A-3   . . . . . Q . . . H . .   . . Y H . . . K . F Q . .   . . S . . . . . . Q .
A-5   . . . . . Q . . . H . .   . . . D . . . . . Y Q . .   . . S . . . . . . F .
A-6   . . . . . Q . . . L . .   . . Y . D . . . . L Q . .   . . . . . . . . . . L
A-8   . . . . . Q . . . H . .   . . . H D . . K . Y Q . .   . . . . . . . . . . F
A-9   . . . . . Q . . . L . .   . . Y . D . . K . F . . .   . . . . . . . . . . .
A-12  . . . . . Q . . . L . .   . . Y H D . . K . Y H R .   . . . . . . . . . . F
A-15  . . . . . Q . . . Y . .   . . Y H D . . G . Y Q . .   . . . . . . . . . . F
A-19  . . . . . Q . . . L . .   . . Y E D . . K . F H . .   . . . . . . . . . . F   R81C
MRT-3 . . Y . . H . . . L . .   . . . E . . . . . . Q . .   . . . . . . . . . . .
MRT-4 . . . . . H . . . L . .   . . Y D . . . . . . H . .   . . . . . . . . . . .
MRT-5 . . . . . H . . . Y . .   . . Y E . . . . . . Y . .   . . S . . . . . . . .
MRT-18 . . . . . H . . . L . .   . . H H . . . . . . . . .   . . . . . . . . . . .
M60-2 . . . . . H . . . L . .   . . Y . . . . . . . Q . .   . . . . . . . . . . .
M60-3 . . . . . H . . . L . .   . . Y D . . . . . . . . .   . . . . . . . . . . .
M60-7 . . . . . H . . . L . .   . . Y . . . . . . . Q . .   . . . . . . . . . . .   V165G, D166G
M60-9 . . . . . H . . . L . .   . . H . . . . . . . Q . .   . . . . . . . . Y . . .
M60-11 . . . . . H . . . L . .  . . . . . . . . . . . . .   . . . . . . . . . . F   Y88H
M60-12 . . . . . H . . . L . .  P . . . . . . . . . . . .   . . . . . . . . . . F
G-1   . . Y . . D . . . L . .   . . L E H . . K . F H . .   . . . . . . . Y . . F   E138G
G-3   . . L . . D . . . L . .   . . L H Q . . K . Y D . .   . . . . . . . Y . . .   E138G
G-4   . . F . . D . . . L . .   . . L E H . . . . F . . .   . . . . . . . Y . . .   E138A
G-11  . . . . . D . . . L . .   . . L H H . . . . F D . .   . . . . . . . Y . . F   D136G
G-15  . . L . . D . . . L . .   R . L D H . . . . L H . .   . . . . . . . Y . . .
```

Fig. 5. Amino acid sequence alignment of intentionally mutated regions of the primary structure of CBM-variants selected on xylan (X-), Avicel (A-), mannan (MRT- and M60-) and human IgG4 (G-). The amino acid numbering in this figure is according to Simpson *et al.* (2002).
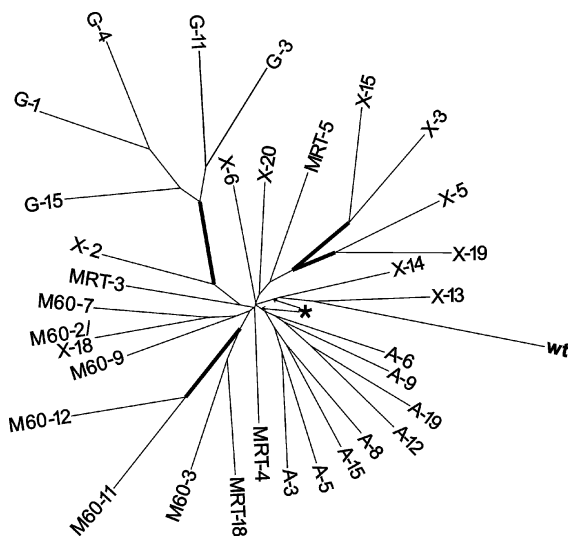

Fig. 6. Clonal relationship as a majority rule consensus tree calculated using parsimony. Branches obtaining more than 50% support in bootstrap analysis in minimum evolution (ME) and neighbour-joining (NJ) only are indicated by asterisks in the tree and branches supported to more than 50% in all analyses by thick branches. For identification, the clones were given letters from the name of the selection substrate like X for xylan, A for Avicel, M for mannan and G for human IgG4 plus a number. The names of clones selected on mannan were given an extra RT for those selected only at room temperature and 60 for those selected three times at 60°C and once at room temperature. The wild-type sequence is denoted wt.
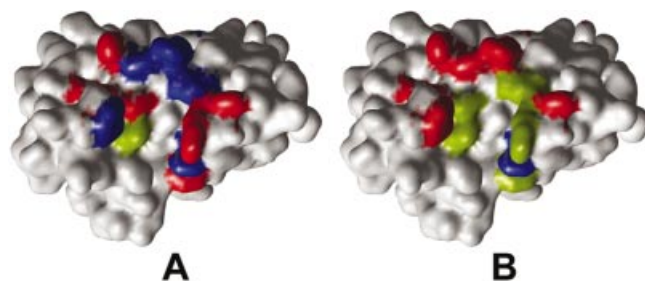

Fig. 7. Diversity in clones selected on birchwood xylan (*n* = 10) (**A**) or human monoclonal IgG4 (*n* = 5) (**B**) shown in relation to the wild-type sequence of CBM4-2. Diversity is shown on the known structure of CBM4-2 (PDB 1k45) and illustrates modifications preferentially selected in the 12 positions allowed to vary by library design. Differences are colour-coded: blue, conserved residues ($\geqslant$4/5 identical to wt CBM4-2); green, conserved ($\geqslant$4/5) but mutated related to wild-type CBM4-2; red, no single particular residue was highly selected for (<4/5 of independent, selected clones carried a specific residue).

suggested that some types of diversity in certain residues were not well tolerated by the scaffold (Wahlberg *et al.*, 2003). Future library designs would need to take this factor into account and possibly introduce another type of diversity in this position. Furthermore, it cannot be excluded that specific combinations of residues are not well tolerated, but this problem is associated with all kinds of large library design and
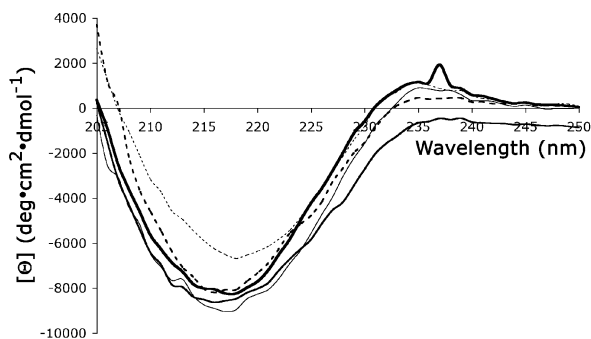
Fig. 8. The wild-type CBM and clones selected from the library have the same overall structure. CD spectra of wild-type CBM (very thick solid line), A-6 selected on Avicel (thick solid line), MRT-5 selected on mannan (thin solid line), G-11 selected on human IgG4 (thick broken line) and X-2 selected on xylan (thin broken line) have a significant absorption minimum at 217 nm which is typical for a β-sheet structure.

is not specifically related to the use of CBM4-2 as a diversity-carrying scaffold.

The fact that some members of a protein library can be poorly folded does not necessarily have to be a problem. It has been shown that such molecules can still act as specific binders and they can also be specifically selected by phage display (Wahlberg *et al.*, 2003). Such proteins were, however, not selected from this library based on the CBM scaffold. CD spectra and high thermal transition temperatures of the selected clones confirmed a stable β-sheet structure just like that of the wild-type module (Simpson *et al.*, 2002).

Many molecules used as scaffolds have one or a few disulphide bridges in the protein structure (Skerra, 2000a,b), which often relates to their stability. In many display systems, the formation of such disulphide bonds is problematic as the protein is synthesized in a reducing environment where such stabilizing bonds cannot form. Such scaffolds may be difficult to use for display. This is, however, not a problem in our case, as the wild-type CBM4-2 is highly thermostable, although it does not possess any disulphide bridges. In this study, we have demonstrated that an active molecule can be formed both in a reducing (intracellular) and in an oxidizing environment (export into periplasmic space). This fact is due to the lack of disulphide bonds and it strongly argues in favour of the CBM's adaptability to various display and selection conditions.

In addition to stability, productivity and yield of purification are important parameters (Nygren and Uhlén, 1997) in order to find a selected molecule that is useful in large-scale applications. The most common molecules used for diversification are antibodies that sometimes have a poor productivity, in particular in prokaryotic systems where disulphide bonds form poorly. In contrast, both the wild-type CBM, and even more importantly, 14 selected variants of this protein have all proved to be easily and reliably produced in *E.coli* and purified by affinity chromatography, yielding in several cases over 80 mg of purified protein/l in a simple production system like shake flask culture. This observation and the fact that very high yields of the wild-type protein can be achieved in fed-batch cultures using 3 l fermenters (data not shown) suggest that mutant proteins could also be produced with very good yield, in particular if the production system is optimized further.

Apart from being dynamically stable and easily produced, scaffolds are defined as molecules able to switch biochemical function when changes are introduced at the binding site. In this study, a small combinatorial library, of a size of 1.6×10^6 clones, was created by allowing restricted variation at 12 positions in the binding site of CBM4-2 and different binding specificities have successfully been selected for by using not only the carbohydrate ligands birchwood xylan, Avicel, ivory nut mannan, but also the glycoprotein human IgG4. Theoretical considerations (Perelson and Oster, 1979) as well as practical experience (Hoogenboom, 2002) suggest that a small library, such as the one assessed here, will only give rise to low-affinity binders. Despite the small size of the library we could clearly identify variants occupying different parts of the sequence space. The characteristic evolution patterns chosen by the target-specific clones were identified and revealed some of the residues important for the binding to the different substrates. For example, the IgG4 selection resulted in the clones carrying a particular residue in 6/12 intentionally mutated positions spread over the xylan binding site (Figure 7B). These included the replacement of one aromatic residue (F110L) and extensive diversification of another aromatic residue (W69), both believed to be important for the xylan binding. These clones selected on human IgG4 had lost the ability to bind the original substrate, xylan, confirming that available interactions, their position or the distance between them in the binding site of IgG4-specific variants were entirely different as compared with the xylan-specific clones. This is despite the fact that the overall fold appeared to be conserved in all molecules. The monoclonal IgG4 selecting for these more divergent binders is to a large extent of a type (i.e. protein versus carbohydrate) different from xylan and the carbohydrates it carries are also likely to be diverse. Furthermore, carbohydrates found in the constant part of IgG are partly hidden in between protein modules, making them difficult to access from the outside. Altogether, this suggests that selected variants may not necessarily recognize carbohydrates but the protein itself. Studies are currently under way to clarify this point (L.Cicortas Gunnarsson, E.Nordberg Karlsson, M.Andersson, O.Holst and M.Ohlin, unpublished data). It is not unreasonable to believe that a molecular library, like the one based on CBM4-2, may contain variants able to recognize very different types of targets. This has already been shown in other cases, as exemplified by an antibody fragment library based on a single scaffold that has been used to select for binders specifically recognizing either proteins, peptides, carbohydrates or small organic molecules (Söderlind *et al.*, 2000), and it is thus not a unique feature of the CBM4-2 library. In fact, given the high selective power of phage display, such variants can be retrieved even if they are only very rarely found in the library.

In summary, we have shown that CBM4-2 can serve as a good scaffold in molecular engineering for evolving new binding specificities. The CBM has a stable β-sheet structure that remained largely intact despite the mutations that were introduced in the binding site of the molecule. The high productivity and the high yield of purification are two other properties that support the large potential of this molecule as a diversity-carrying scaffold and demonstrate its usefulness in downstream biotechnological applications.

References

Abou-Hachem,M., Nordberg Karlsson,E., Bartonek-Roxå,E., Raghothama,S., Simpson,P.J., Gilbert,H.J., Williamson,M.P. and Holst,O. (2000) *Biochem. J.*, **345**, 53–60.

Abou-Hachem,M. *et al.* (2002) *Biochemistry*, **41**, 5720–5729.

Boraston,A.B., McLean,B.W., Kormos,J.M., Alam,M., Gilkes,N.R., Haynes,C.A., Tomme,P., Kilburn,D.G. and Warren,R.A.J. (1999) In Gilbert,H.J., Davies,G.J., Svensson,B. and Henrissat,B. (eds), *Recent Advances in Carbohydrate Engineering.* Royal Society of Chemistry, Cambridge, pp. 202–211.

Boraston,A.B., Creagh,A.L., Alam,M.M., Kormos,J.M., Tomme,P., Haynes,C.A., Warren,R.A. and Kilburn,D.G. (2001) *Biochemistry*, **40**, 6240–6247.

Edwards,A.W.F. (1996) *Syst. Biol.*, **45**, 79–91.

Engberg,J. andersen,P.S., Nielsen,L.K., Dziegiel,M., Johansen,L.K. and Albrechtsen,B. (1996) *Mol. Biotechnol.*, **6**, 287–310.

Felsenstein,J. (1978) *Syst. Zool.*, **27**, 401–410.

Felsenstein,J. (1985) *Evolution*, **39**, 783–791.

Felsenstein,J. (1993) *Cladistics*, **5**, 164–166.

Fitch,W.M. (1971) *Syst. Zool.*, **20**, 406–416.

Gilbert,H.J. *et al.* (2002) In Teeri,T.T., Svensson,B., Gilbert,H.J. and Feizi,T. (eds), *Carbohydrate Bioengineering Interdisciplinary Approaches.* Royal Society of Chemistry, Cambridge, pp. 89–98.

Hoogenboom,H.R. (2002) *Methods Mol. Biol.*, **178**, 1–37.

Jacobsson,K. and Frykberg,L. (1995) *Biotechniques*, **18**, 878–885.

Johansen,L.K., Albrechtsen,B., Andersen,H.W. and Engberg,J. (1995) *Protein Eng.*, **8**, 1063–1067.

Kumar,S., Tamura,K., Jakobsen,I.B. and Nei,M. (2001) *Bioinformatics*, **17**, 1244–1245.

Lehtiö,J., Teeri,T.T. and Nygren,P.-Å. (2000) *Proteins*, **41**, 316–322.

Linder,M., Salovuori,I., Ruohonen,L. and Teeri,T.T. (1996) *J. Biol. Chem.*, **271**, 21268–21272.

Nord,K., Gunneriusson,E., Ringdahl,J., Ståhl,S., Uhlén,M. and Nygren,P.-Å. (1997) *Nat. Biotechnol.*, **15**, 772–777.

Nordberg Karlsson,E., Bartonek-Roxå,E. and Holst,O. (1997) *Biochim. Biophys. Acta*, **1353**, 118–124.

Nygren,P.-Å. and Uhlén,M. (1997) *Curr. Opin. Struct. Biol.*, **7**, 463–469.

Page,R.D. (1996) *Comput. Appl. Biosci.*, **12**, 357–358.

Pell,G., Williamson,M.P., Walters,C., Du,H., Gilbert,H.J. and Bolam,D.N. (2003) *Biochemistry*, **42**, 9316–9323.

Perelson,A.S. and Oster,G.F. (1979) *J. Theor. Biol.*, **81**, 645–670.

Simpson,H.D. and Barras,F. (1999) *J. Bacteriol.*, **181**, 4611–4616.

Simpson,P.J., Xie,H., Bolam,D.N., Gilbert,H.J. and Williamson,M.P. (2000) *J. Biol. Chem.*, **275**, 41137–41142.

Simpson,P.J., Jamieson,S.J., Abou-Hachem,M., Karlsson,E.N., Gilbert,H.J., Holst,O. and Williamson,M.P. (2002) *Biochemistry*, **41**, 5712–5719.

Skerra,A. (2000a) *J. Mol. Recognit.*, **13**, 167–87.

Skerra,A. (2000b) *Biochim. Biophys. Acta*, **1482**, 337–350.

Smith,G.P. (1985) *Science*, **228**, 1315–1317.

Smith,G.P., Patel,S.U., Windass,J.D., Thornton,J.M., Winter,G. and Griffiths,A.D. (1998) *J. Mol. Biol.*, **277**, 317–332.

Söderlind,E. *et al.* (2000) *Nat. Biotechnol.*, **18**, 852–856.

Söderlind,E., Carlsson,R., Borrebaeck,C.A. and Ohlin,M. (2001) *Comb. Chem. High Throughput Screen.*, **4**, 409–416.

Wahlberg,E., Lendel,C., Helgstrand,M., Allard,P., Dincbas-Renqvist,V., Hedqvist,A., Berglund,H., Nygren,P.A. and Hard,T. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 3185–3190.