Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Using structure to inform carbohydrate binding module function

D Wade Abbott[1] and Alicia Lammerts van Bueren[2]

Generally, non-catalytic carbohydrate binding module (CBM) specificity has been shown to parallel the catalytic activity of the carbohydrate active enzyme (CAZyme) module it is appended to. With the rapid expansion in metagenomic sequence space for the potential discovery of new CBMs in addition to the recent emergence of several new CBM families that display diverse binding profiles and novel functions, elucidating the function of these protein modules has become a much more challenging task. This review summarizes several approaches that have been reported for using primary structure to inform CBM specificity and streamlining their biophysical characterization. In addition we discuss general trends in binding site architecture and several newly identified functions for CBMs. Streams of investigation that will facilitate the development and refinement of sequence-based prediction tools are suggested.

Addresses
[1] Lethbridge Research Centre, Agriculture and Agri-Food Canada, 5403-1st Ave, Lethbridge, AB, Canada T1J4B1
[2] Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen 9747AG, The Netherlands

Corresponding author: Abbott, D Wade (wade.abbott@agr.gc.ca)

Current Opinion in Structural Biology 2014, **28**:32–40

This review comes from a themed issue on **Carbohydrate–protein interactions and glycosylation**

Edited by **Harry Brumer** and **Harry J Gilbert**

http://dx.doi.org/10.1016/j.sbi.2014.07.004

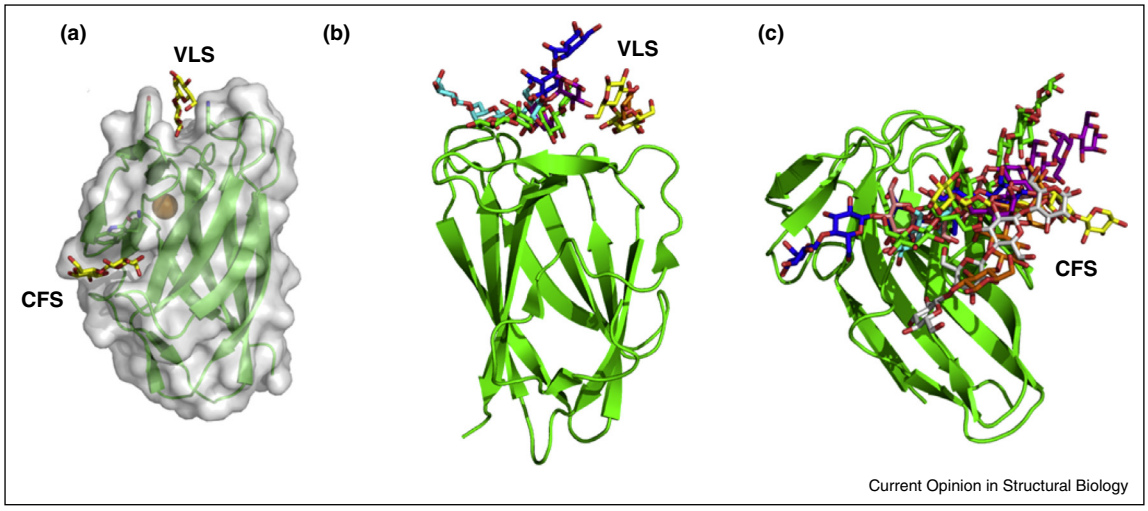
Introduction

Non-catalytic carbohydrate binding modules (CBMs) are components of carbohydrate active enzymes (CAZymes) that fold independently into discreet functional units. CBMs bind carbohydrates, as opposed to catalytically modifying carbohydrate structure, and they have been assigned several global roles in CAZyme biology, including coordinated glycan recognition, general substrate adherence, and structure–function contributions to the catalytic site (see [1•,2••,3] for reviews). Although analogous in function to lectins in that they bind carbohydrate (see Drickamer and Taylor this issue for further discussion), CBMs may be differentiated by their association with CAZymes; for the few members of sequence-based CBM families that are not components of CAZymes (e.g. [4]), they could thus be defined henceforth as 'orphan' CBMs. With the rapid expansion in metagenomic sequence space and characterized CBM families (www.cazy.org; [5]), several consistent trends in CBM structure and function have emerged. For example, the β-sandwich has proven to be an abundant CBM fold found within CAZymes from microbial and fungal organisms that colonize diverse ecosystems, including soil, fresh water, the ocean, and the intestine of animals (see Etzold and Juge this issue for further discussion on mucin binding CBMs). The plasticity of this scaffold is highlighted by the presence of two distinct surface locations for binding sites (Figure 1), which include the variable loop site (VLS) that interconnects the β-strands at one end of the β-sandwich (previously referred to as site 1 in CBM6 and site 2 in α-glucan binding CBMs)[3] and the concave face site (CFS; previously referred to as site 2 in CBM6 and site 1 in α-glucan binding CBMs). There have been a few reported examples in which both binding sites are operational within the same protein [6–10] (Figure 1a). Across the β-sandwich CBM superfamily, the VLS has proven to interact with a diverse palette of linkages and sterochemistries, including distinct carbohydrate ring conformations, epimers, anomers, and degrees of polymerization. The capacity to bind diverse chemistries and structures suggests that the variable loops provide a tunable platform to accommodate different requirements for binding site depth and shape. Alternatively, the CFS is the predominant location for pure α-glucan and β-glycan binding activity (Table 1; [2••,3]). This more restricted profile may result from the limited flexibility of functional groups (i.e. aromatics) displayed on the β-sheet surface.

One of the greatest challenges in CBM research is experimentally determining its binding specificity. This process requires the identification of appropriate modular boundaries, accumulating sufficient amounts of soluble protein, and qualitatively or quantitatively defining the

[3] Currently in the field there is discrepancy in how the two binding sites are labeled. For CBM6 and CBM35, which interact with heterogeneous carbohydrates and linkages, site 1 = VLS and site 2 = CFS [19••]. For α-glucan binding CBMs site 1 = CFS and site 2 = VLS [10]. We have proposed here the VLS and CFS nomenclature that will help to unify these two labeling systems and provide consistency for comparing the binding site architectures of diverse β-sandwich CBM families.

Figure 1



Distribution of binding sites in β-sandwich CBM6s in complex with heterogeneous β-linked carbohydrates. **(a)** Location of the VLS and CFS on the surface of the CBM6 from *Cellvibrio mixtus* (CmCBM6; 1UYY, [6]). The backbone is displayed as a green cartoon with a transparent solvent accessible surface. CmCBM6 is in complex with cellobiose (VLS) and xylobiose (CFS) displayed as yellow sticks, and a structural calcium displayed as an orange sphere. **(b)** Structural superimposition of ligands bound within the VLS of CBM6s. Green = CBM6 (1UY4, [6]); blue = CBM32 (2W1U, [38]); magenta = CBM35 (2VZQ, [14]); orange = CBM36 (1UX7, [39]); yellow = CBM60 (2XFD, [40]); cyan = CBM61 (2XOM, [41]). The backbone β-sandwich is CmCBM6 displayed as a green cartoon. Ligand chemistries are detailed in Table 1. **(c)** Structural superimposition of ligands bound within CFS of CBM6s. Blue = CBM4 (1GU3, [42]); green = CBM6 (1UYY, [6]); yellow = CBM15 (1GNY, [43]); pink = CBM16 (3OEA, [44]); gray = CBM17 (1J84, [45]); magenta = CBM27 (1OF4, [46]); orange = CBM28 (3ACH, [47]); cyan = CBM29 (1GWN, [48]). The backbone β-sandwich is CmCBM6 displayed as a green cartoon. Ligand chemistries are detailed in Table 1.

correct ligand(s) using an appropriate biophysical assay [11]. Further confounding this practice has been the emergence of CBM families that display diverse binding profiles (e.g. CBM6, CBM32, and CBM35), cryptic modular architectures (e.g. the bivalent starch binding CBM41s from *Streptococcus pneumoniae* [12]), and novel modes of function (e.g. calcium meditated oligomerization [13,14] and modulation of enzymatic specificity [15••]. Designing predictive models that use primary structure to illuminate ligand binding specificity, and refining these models using insights gleaned from structure–function investigations, are helping to develop functional hypotheses and propel innovation toward the biotechnological application of CBMs. This review will summarize the status of these predictive models, and introduce several new topics worthy of further sequence-based predictive model development.

Phylogenetic predictions

In cases where the activity of the parent enzyme is known and CBM binding specificity is invariant within a family (e.g. starch binding CBM20 and CBM21) binding function can be inferred and its characterization streamlined using carbohydrate ligands that reflect structures or substructures within the target substrate (e.g. α-maltooligosaccharides, amylose, amylopectin, glycogen, pullulan, dextran). In recent years, however, examples where enzyme and CBM specificity are uncoupled have become more common. One standout example is a cohort of uronic acid binding CBM35s that harness a calcium cofactor to bind unsaturated or saturated GalA/GlcA ligands [14]. Despite conserved binding mechanisms, these CBM35s are appended to widely divergent catalytic modules that are active on xylan (EC 3.2.1.8), chitosan (EC 3.2.1.165), and rhamnogalacturonan acetyl esters (EC 3.2.1.-). This represents a more recent and profound example to the seminal reports of family 2 CBMs appended to both cellulase (GH6) and xylanase (GH10) catalytic modules [16].

Phylogenetic comparisons of CBM sequences within defined families and across clans (i.e. which include distantly related families that share ancestral folds) have been successful for guiding the determination of CBM binding mechanisms and ligand specificity [17,18,19••,20,21]. This approach directly compares the similarity in functional residues between CBM sequences that have been extracted from the parent polypeptide. Alignments provide insights into CAZyme–CBM relationships with differential specificities or orphan CBMs, and functions that extend beyond catalytic potentiation effects [4,17,22]. In particular, this approach has utility for clustering similar specificities within families that bind a diverse portfolio of carbohydrates.