Paul Wolski
April 10, 2008

Characterization of the *Thermobifida fusca* Xylanase 11A Carbohydrate-Binding Module by X-ray Crystallography

Abstract:

Xylanase 11A is an enzyme that degrades xylan. It was isolated from *Thermobifida fusca,* a thermophilic bacterium found in compost. Xyl11A also contains a family IIb carbohydrate-binding module (CBM), which has been found to be capable of binding to both cellulose and xylan, unlike some other family IIb CBMs. For this reason, this binding domain was chosen for characterization by X-ray crystallography. The gene for this binding domain was cloned into *Escherichia coli* and purified with column chromatography. Crystals were found in some crystallization screens, and diffraction of these crystals is currently pending.

Introduction:

The sugar present in cellulose provides a massive resource for energy usage, as well as for polymers. Both of these—energy, in the form of petroleum and other fossil fuels, and polymers, in the form of plastics, such as polystyrene, are currently derived from sources that are finite. The polymers, such as polystyrene, are also non-biodegradable. Additionally, it is becoming more and more evident that the burning of fossil fuels is adding to global warming.

The current problem with harnessing the energy stored in cellulose, the primary carbohydrate source on the planet, as well as the most abundant source of organic material (3), is its tough, fibrous structure. As such, the need for developing more efficient cellulases is of key importance in harnessing this energy. Cellulases are either endo- or exo- meaning they either begin their cleavage any where along the chain or at the end, respectively.

In order for effective, continued cleavage, these enzymes need to be in contact with their substrate. Since the substrate is insoluble, simply relying on thermodynamic collisions in solution would be inefficient. Therefore, carbohydrate-binding modules (CBMs) have evolved to bind the enzyme to the substrate. These domains are connected by a linker to the catalytic domain of the enzyme. Based on homology, they are classified into several different families.

There is a high level of secondary structure similarity between family IIa cellulose-binding domains and family IIb xylan-binding domains (9), including that of Xylanase 11A from the bacterium, *Thermobifida fusca*. Generally, family IIa cellulose binding domains have 3 conserved aromatic residues that all are important for binding (2,7,8). Family IIb xylan binding domains have only two of these three residues (see Figure 1), and they are perpendicular, unlike the family IIa cellulose binding domains, which have these residues in a coplanar arrangement (1). These arrangements are suitable for binding the different types of substrates—cellulose or xylan.

The family IIb CBM of the xylanase, Xyl11A, a 10-kDa globular protein from the bacterium *T. fusca,* is like other family IIb CBMs in the number and arrangement of conserved aromatic binding residues. In general it has been found that most of the family IIb xylanase binding domains bind only to xylan and not to cellulose, but the CBM IIb of *T. fusca* Xyl11A has the ability to bind to both cellulose and xylan (4). Thus, for this reason, it is especially interesting to try to determine the structure of this protein, especially to identify what allows it to bind to both cellulose and xylan. It is possible that its structure, especially at the surface near the binding site, is different from both classes of CMBIIs.

A potential application of the knowledge of the structure might be to design more effective enzymes, perhaps ones that are capable of efficiently binding both cellulose and xylan.

It also might be possible to design a highly effective binding domain and fuse it to a highly effective catalytic domain. Additionally, knowledge of this structure, compared to its performance as a binding domain, might aid it the knowledge of the design of other, potentially more effective cellulose binding domains.

Although there does exist an NMR structure of a similar domain, having the same name, from *Cellulomas fimi*, there is decent level of sequence dissimilarity between them.

The goal of this project is to determine the structure of *Thermobifida fusca* Xyl11A CBM by X-ray crystallography. We want this information to determine at what angles the key binding residues are and how and why this enzyme is able to bind to both cellulose and xylan.

Materials and Methods:

The gene for *T fusca* Xyl11A was already cloned into *Escherichia coli* (5). Specifically, the gene for this enzyme had been cloned into pYY01 + pUC18. The sequence of the CBM IIb was already known (4), and polymerase chain reaction (PCR) was used to amplify just the DNA for the binding domain, after designing primers that contain the proper sites for restriction enzymes (NOT 1 and XHO1) for later ligation into an expression plasmid. The desired PCR product of DNA for Xyl11A CBM was selected by gel electrophoresis using a 1 % agarose gel and then gel-purified. Once this was done, the desired DNA was blunt-ligated into the Promega PGEM T-Easy vector (10) and the mixture of plasmids was transformed into *E. coli* by heat shock using the Promega protocol.

Next, the bacteria were placed onto Petri dishes that contain ampicillin as well as X-gal. The desired T. easy plasmid contains a gene for ampicillin resistance and a beta galactosidase gene. The colonies that grow should have the desired plasmid. In order to detect colonies that

have the desired insert in this plasmid, the CBM for Xyl11A was cloned into the beta gal gene, so bacteria with the correct plasmid and the insert will not be blue (blue color results when the X-gal is degraded).

Several white colonies were selected, and their plasmid DNA was isolated by a miniprep procedure, and purified by gel electrophoresis and extraction of the complete plasmid, whose insert was sequenced. Upon finding no mutations in the sequence, the desired fragment was cut out of the T. easy plasmid using NOT 1 and XHO1, after growing up this strain in Luria broth + ampicillin + glucose. An *E. coli* strain containing an expression plasmid (Novagen Pet26b+), containing a T7 promoter and a kanamycin resistance gene was grown in Luria broth + kanamycin + glucose. This plasmid was then isolated by miniprep and was then restricted with NOT1 and XHO1. The cut plasmid was isolated by gel electrophoresis (on a .8 % agarose gel) and gel purification. This cut plasmid was then ligated with the fragment cut out from the T. easy vector. The ligated plasmid, after gel electrophoresis (.8% agarose) for purification, was transformed by heat shock into *E. coli* D H 5 alpha cells.

Upon growing up enough of the plasmid in the D H 5 alpha cells, the plasmid was isolated by miniprep and then transformed into *E. coli* BL21 DE3 cells for expression. These cells were then tested by miniprep and gel electrophoresis to make sure the correct plasmid had been taken up.

From SDS-PAGE analysis of a small amount of supernatant from and, separately, of lysed cells of transformed BL21 DE3, I showed that the desired protein is excreted. Thus, Xyl11A CBM was taken from the centrifuged supernatant of the fluid in which the bacteria were grown.

A 50-ml culture in LB media + 60μg/ml kanamycin + .5% glucose was initially grown overnight at 30°C with shaking. Then, 30 ml of this culture was added to 1 L of M9 media + .5% glucose + 60μg/ml kanamycin + 2 mM magnesium sulfate + .1 mM calcium chloride. This culture was also initially grown at 30°C with shaking. When the optical density at 600 nm reached 1.2, the culture was induced with IPTG to final a concentration of .8 mM and harvested 16 hours later. The cell culture was filtered through glass wool. Then, it was centrifuged in a Beckman centrifuge at 6500 rpm at 4ºC for 20 minutes. Sodium chloride was added to a concentration of 1 M, and phenyl-methyl-sulfonyl-floride was added to a concentration of .1 mM.

The protein was purified by column chromatography (6), first on a column of about 2.5 cm in diameter filled with 100 ml of packed cellulose (CF 11). The column was equilibrated with 200ml of 1 M NaCl + 50mM potassium phosphate (KPi) (pH 6.8). Then, the entire centrifuged supernatant was loaded onto this column and allowed to flow through. The protein was eluted off the column with 200 ml of ethylene glycol, while collecting 90-drop fractions in a fraction collector. The loading and eluting steps were all done in a 4ºC cold room.

The fractions were tested for presence of Xyl11A CBM by SDS PAGE. Fractions containing Xyl11A CBM, as indicated by the presence of a band at 10 kDa, were pooled and sealed in cellulose dialysis tubing with a 6-kDa molecular weight cut-off in order to remove the ethylene glycol. The sample was added to 1 L of 5mM KPi pH 6.8 in the 4ºC cold room. The buffer was replaced the next day.

After 24 hrs, the dialyzed sample, on ice, was added to 15 ml packed hydroxy apatite on a column equilibrated with 50 ml of 5mM KPi pH 6.8 at room temperature, and the flow-through was collected (on ice). Then, it was eluted with 2 separate 10-ml aliquots of 5mM KPi pH 6.8.

The flow through and elutions were tested for the presence of Xyl11A CBM by SDS PAGE stained with Comassie Blue and, separately, Western Blot using rabbit anti-Xyl11A antibodies.

The active fractions were pooled and concentrated to 7.5 mg/ml using, first, a high pressure-nitrogen filtration system to push the solvent through a 5-kDa-cutoff Amicon membrane. The volume was reduced to 10 ml (from 40 ml) this way. The remaining 10 ml was concentrated to .75 ml on a 3-kDa-cutoff Millipore centrifuge tube/membrane that was spun in a Beckman centrifuge at 6500 rpm. It took about 40 minutes to have 2 ml flow through.

The concentrated sample was taken to Dr. Ailong Ke's lab, and 48 wells of crystal screens using the Hampton Crystal Screen HR2-110 were set up. (Screens 49 and 50 were omitted). For each solution, 0.5 ml was added to each reservoir. On the siliconized cover plate for each crystal screen condition, a drop of 2 μL of protein solution and 2 μL reservoir solution was added, making one 4-μL drop.

Future experiments include diffracting crystals with X-rays's from Cornell's CHESS synchrotron facility. Also, I will soak the protein crystals with cellulose or xylan oligosaccharides and then diffract this complex. The structures will be solved by molecular replacement based on homologous domains, such as Xyl11A from *Cellulomas fimi*.

Results:

The purification method mentioned above was successful in developing a pure sample of Xyl11A CBM (figure 2). I increased the yield by eliminating a phenyl sepharose column in between the two purification steps mentioned above, as I determined by SDS PAGE that this step did not add to the purity of the sample. From a 1 L M9 culture, 5 mg of pure Xyl11A CBM were

collected, as calculated by optical density at 280 nm in a quartz cuvette, and this was concentrated to 7.5 mg/ml.

Western blot analysis of another gel like Figure 2, but not stained with Comassie blue, using rabbit anti-Xyl11A showed fluorescence for each of the Xyl11A CBM bands at 10 kDa, meaning that each of the faint bands of hydroxyl apatite flow-through and KPi elution was Xyl11A CBM.

Upon crystallizing the sample, crystals formed within 48 hrs after set up in the following two conditions (conditions 1 and 24 in the Hampton Screen):

1. .2 M Calcium chloride dehydrate, .1 M Sodium acetate trihydrate, pH 4.6 with 30% w/v of: (+/-)-2-Methyl-2,4-pentanediol
2. .2 M Calcium chloride dehydrate, .1 M Sodium acetate trihydrate, pH 4.6 with 20% w/v of: 2-Propanol.

In the first condition three crystals formed. In the 24th condition, 1 larger crystal formed. The exact sizes were not measured, but all of the crystals are suitable for X-ray diffraction.

Discussion:

In general, it seems as if this experiment is proceeding smoothly, albeit slowly. Some future projects include crystallization of the Xyl11A CBM bound to different types of cellulose and, separately, xylan oligosaccharides. Additionally, site directed mutagenesis and then either X-ray crystallography or binding assays to various carbohydrate samples will be conducted. It would be interesting to be able to change the binding properties of Xyl11A CBM to properties more like those found in other family IIb CBMs, which bind only xylan or like family IIa CBMs, which bind only cellulose. Engineering the native Xyl11A or a mutant binding domain onto a

catalytic domain of a different enzyme and studying cellulose hydrolysis rates would also be very interesting.

Acknowledgements:

A very sincere thank-you goes to Dr. Wilson for his support and mentoring. I also could not have succeeded without the mentoring and technical support from Diana Irwin, Thu Vuong, Yongchao Li, and John Dingee. I'd also like to thank the Dextra Undergraduate Endowment for helping to fund this project.

References:

1. Bolam, D.N. et. Al. "Evidence for Synergy between Family 2b Carbohydrate Binding Modules in *Cellulomonas fimi* Xylanase 11A." *Biochemistry*. (2001), 40: 2468-2477.
2. Bray, M.R., et. Al. "Probing the role of tryptophan residues in a cellulose-binding domain by chemical modificatin." *Protein Science* (1996), 5: 2311-2318.
3. "Cellulose." Encylcopaedia Britannica Online Academic Edition. http://search.eb.com/eb/article-9022028. Accessed 19 March 2008.
4. Irwin, Diana, ElyseD. Jung, David B. Wilson. Characterization and Sequence of a *Thermonospora fusca* Xylanase. *Applied and Environmental Microbiology* Vol. 60 No. 3: 763-770.
5. Ghangas, Gurdev S. You-Ji Hu, David B. Wilson. Cloning of a *Thermonospora fusca* Xylanase Gene and its Expression in *Escherichia coli* and *Streptomyces lividans*. *Journal of Bacteriology* Vol. 171 No. 6: 2963-2969.
6. Leskinen, S. et. Al. Thermostable xylanases Xyn10A, and Xyn11A, from the actinomycete *Nonomuraea flexuosa*: isolation of the genes and characterization of recombinant Xyn11A polypeptides produced in *Trichoderma reesei*. *Appl Microbiol Biotechnol*. 67: 495-505.
7. Mclean, BW., et. Al. "Analysis of Binding of the Family 2a Carbohydrate-Binding Module from *Cellulomas fimi* Xylanase 10A to Cellulose: Specificity and Identification of Functionally Important Amino Acid Residues." *Protein Engineering*. Vol. 13 No. 11: 801-809, 2000.
8. Nagy, T., et. Al. "All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands." *FEBS Letters*. 429 (1998) 312-316.
9. Simpson, Peter J. , et. Al. A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure* Vol. 7 Issue 7: 853-864.
10. "Technical Manual: pGEM®-T and pGEM®-T Easy Vector Systems." May 2007. http://www.promega.com/tbs/tm042/tm042.pdf. Accessed 25 March 25, 2008.

Figures:

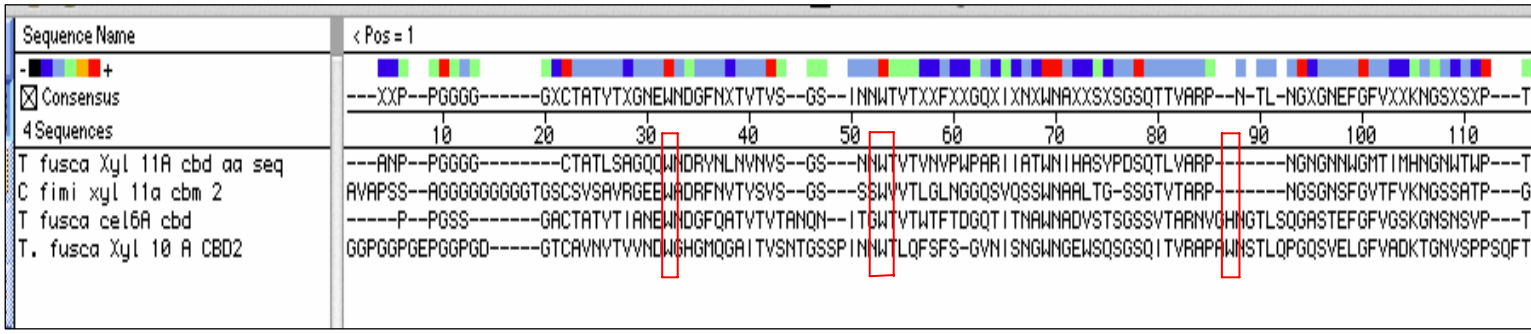

Figure 1: Alignment of two family IIa and two family IIb carbohydrate-binding modules (CBM). (The upper two—*Thermobifida fusca* Xylanase 11A CBM, the protein of interest, and *Cellulomas fimi* Xylanase 11A CBM—are from family IIb [note the only two conserved aromatic amino acids], while the lower two—*Thermobifida fusca* Cellulase 6A CBM and *Thermobifida fusca* Xylanase 10A CBM belong to family IIa, binding primarily to cellulose.) The three conserved aromatic surface residues (the third of which is absent in IIb) are boxed. (His and Trp are the third).

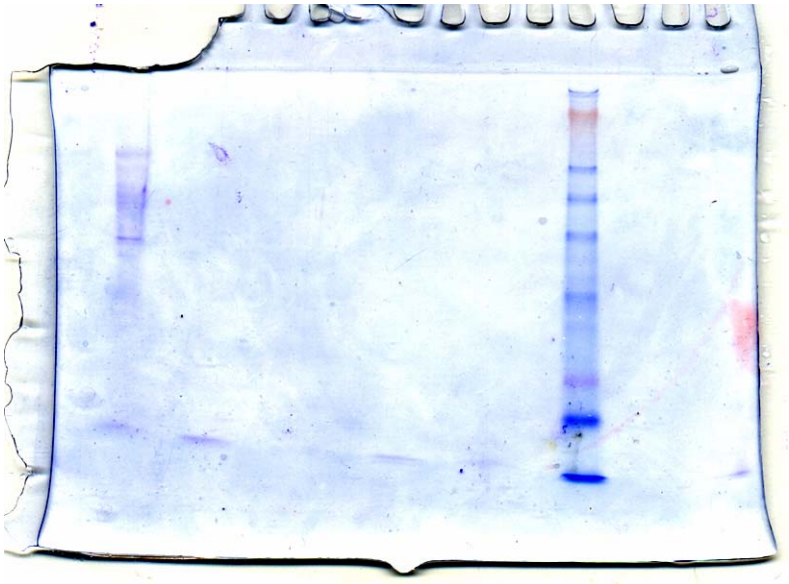

Figure 2. 12% SDS-PAGE gel stained with Comassie Blue of purified Xyl11A CBM after hydroxy apatite column. The lane in the far left is the sample before hydroxy apatite purification. The subsequent blue lines near the bottom going right are all that is left from the purification and elutions with potassium phosphate—Xyl11A CBM (10 kDa). The lane at the right is the molecular marker with the bottom band being a 5 kDa protein.