# The Jo-In protein welding system is a relevant tool to create CBM-containing plant cell wall degrading enzymes

Louise Badruna, Vincent Burlat, Pierre Roblin, Thomas Enjalbert, Guy
Lippens, Immacolata Venditto, Michael O'donohue, Cédric Montanier

1    **The Jo-In protein welding system is a relevant tool to create CBM-containing plant cell**

2    **wall degrading enzymes**

3

4    Louise Badruna[1], Vincent Burlat[2], Pierre Roblin[3], Thomas Enjalbert[1], Guy Lippens[1],

5    Immacolata Venditto[4], Michael J. O'Donohue[1], Cédric Y. Montanier[1]*

6    *Correspondence: cedric.montanier@insa-toulouse.fr; Tel. +33 (0)5 61 55 97 13

7

8    [1] TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

9    [2] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS,

10   Toulouse INP, 24 chemin de Borde Rouge, 31320 Auzeville-Tolosane, France

11   [3] Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse,

12   France

13   [4] Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne

14   NE2 4HH, UK

15

16

17     Abstract:

18         Irrespective of their biological origin, most proteins are composed of several

19     elementary domains connected by linkers. These domains are either functionally independent

20     units, or part of larger multidomain structures whose functions are defined by their spatial

21     proximity. Carbohydrate-degrading enzymes provide examples of a range of multidomain

22     structures, in which catalytic protein domains are frequently appended to one or more non-

23     catalytic carbohydrate-binding modules which specifically bind to carbohydrate motifs. While

24     the carbohydrate-binding specificity of these modules is clear, their function is not fully

25     elucidated. Herein, an original approach to tackle the study of carbohydrate-binding modules

26     using the Jo-In biomolecular welding protein pair is presented. To provide a proof of concept,

27     recombinant xylanases appended to two different carbohydrate-binding modules have been

28     created and produced. The data reveal the biochemical properties of four xylanase variants

29     and provide the basis for correlating enzyme activity to structural properties and to the nature

30     of the substrate and the ligand specificity of the appended carbohydrate-binding module. It

31     reveals that specific spatial arrangements favour activity on soluble polymeric substrates and

32     that activity on such substrates does not predict the behaviour of multimodular enzymes on

33     insoluble plant cell wall samples. The results highlight that the Jo-In protein welding system

34     is extremely useful to design multimodular enzyme systems, especially to create rigid

35     conformations that decrease the risk of intermodular interference. Further work on Jo-In will

36     target the introduction of varying degrees of flexibility, providing the means to study this

37     property and the way it may influence multimodular enzyme functions.

38

39     Keywords:

40     GH family 11 endo-1,4-β-xylanase, CBM family 2, CBM family 3, Bio Molecular Welding,

41     spatial proximity

42

**Introduction**

Domains constitute the key building blocks of proteins, conferring their structural integrity and/or functionalities. They can form independent structural and or functional units, but are often combined in multidomain organizations in which neighboring domains associate to define protein structures and/or functions [1]. The majority of proteins from all taxa are composed of several domains connected by linkers [2,3]. This quite complex organization is probably the source of functional diversity and also responsible for functional fine-tuning [4], including that of enzymes whose catalytic sites are often formed at the interface of several domains [5].

Plant cell wall-degrading enzymes provide excellent examples of multidomain proteins. This large group is characterized by a multitude of structures and functions, their diversity matching the complexity of plant cell wall (PCW) structures [6]. Lignified PCWs are macromolecular networks, composed of cellulose, hemicelluloses, pectins, proteins and lignins that interact and, in some cases, crosslink to form insoluble, three dimensional matrices [7]. The main enzymes involved in the breakdown of PCWs are glycoside hydrolases (GHs), a large and diverse group including cellulases, hemicellulases and pectinases. Most GHs display modular architecture, containing catalytic and non-catalytic domains [8,9]. Regarding the latter, carbohydrate binding modules (CBMs) are prominent.

67   When appended to a catalytic module, CBMs target specific regions in polysaccharides,

68   consequently increasing the local concentration of catalytic domains and favoring intimate

69   contacts between the substrate and the enzyme. In certain cases, CBMs also disrupt the

70   surface of tightly packed polymers, such as cellulose or starch, thus facilitating enzyme action

71   [10].

72

73       In PCW-degrading enzymes, domain organizations are numerous and quite varied. In

74   some cases, the catalytic GH domain and the CBM are combined in a single prolonged

75   domain, while in others the CBM is appended to the GH *via* a linker peptide of variable size

76   (from 4 to 158 residues), composition and structure [11]. Considering the number of known

77   GH and CBM domains and the number of possible combinations,  the organizational diversity

78   of PCW-degrading enzymes is considerable [12].

79

80       Previous studies described the creation of synthetic multimodular GH constructs [13–15]

81   designed to investigate enzymatic activity. Often, a domain fusion strategy was adopted,

82   using short linker sequences to associate different domains in recombinant proteins. While

83   this strategy is frequently employed, its success hinges on expression of the fusion protein and

84   requires the cloning and expression of each studied domain combinations. An alternative

85   route to obtain a range of different domain combinations is to express single domains

86   separately and then perform module linkage in a second *in vitro* step. Several strategies to

87   achieve this have also been described [16–21]. One of the more recent additions to the protein

88   engineer's toolbox is genetically encoded click chemistry (GECC),  based on a naturally

89   occurring phenomenon identified in bacterial pili, where certain protein subunits are linked

90   together *via* an isopeptide bond [22]. Exploiting this for protein engineering has led to the

91   development of SpyTag-SpyCatcher [23] and the Biomolecular Welding tool [24]. The latter

92    comprises two proteins, designated Jo and In (10.6 and 16.5 kDa, respectively), which

93    spontaneously form an intramolecular isopeptide bond when mixed in solution, leading to a

94    two-domain protein measuring 6 nm in length. When Jo and In are individually fused to other

95    protein domains, it is possible to create  domain combinations *in vitro*, with the Jo-In

96    intramolecular complex acting as the linker. Accordingly, Jo-In were recently used to

97    combine two different GHs, producing bifunctional enzymes [25].

98

99    In the current study, focusing on the well characterized xylanase Xyn11A from

100   *Neocallimastix patriciarum* [26] Jo-In is used to link this GH to two different, non-cognate

101   CBMs of bacterial origin either targeting the substrate of Xyn11A (xylan) or another PCW

102   polymer (cellulose), the aim being to understand how the specific properties of these CBMs

103   affect enzyme activity on simple substrates and complex PCW networks.

104

105   **Materials and methods**

106   **Gene cloning**

107   Plasmid constructs used are summarized in Supplementary Table S1. For some cloning

108   purposes, PCR (Phusion™ High-Fidelity DNA Polymerase, ThermoFischer Scientific,

109   Waltham, MA, USA) was used to amplify target sequences, generally introducing restriction

110   enzyme target sequences, and to introduce PCR amplicons into linearized plasmid vectors by

111   homologous recombination (In-Fusion® HD cloning kit, Clonetech, Mountain View, CA,

112   USA). Alternatively, target sequences were synthesized by Genscript HK limited

113   (Piscataway, NJ, USA) introducing appropriate restriction enzyme target sequences for

114   subsequent plasmid construction.

115

116   **Protein expression and purification**

117    Proteins were expressed in *Escherichia coli* strain BL21 (DE3) or Tuner (DE3) harbouring

118    the relevant plasmids and purified using standard methods described in Supplementary Table

119    S2.

120

121    **Covalent chimeric protein complexes**

122    To prepare chimeric protein complexes, purified Jo and In fusion proteins were mixed, using

123    a slightly lower concentration for the xylanase than for the CBMs (typically 2 µmoles of

124    xylanase for 3 µmoles of CBM, 8.3 and 8.8 mg/ml of protein, respectively), for 1 h at 21°C

125    and then stored overnight at 4 °C. Protein complexes were isolated from solution using a

126    XK16 Hiload 16/600 Superdex S75 prep-grade gel filtration column (GE Healthcare Life

127    Sciences, Chicago, IL, USA) connected to an Äkta Pure system. Elution was performed at 1

128    mL/min using 50 mM sodium phosphate buffer pH 7.4 supplemented with 150 mM NaCl.

129    Subsequently, NaCl was removed by dialysis and purified chimeric complexes were judged

130    homogenous by sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE).

131

132    **Enzymatic activity measurements**

133    The apparent kinetic parameters $K_{Mapp}$, $V_{max}$, $k_{cat}$, and $k_{cat}/K_{M\ app}$ of Jo-*Np*Xyn11A were

134    measured using the dinitrosalicylic acid (DNSA) assay as previously described [25], using

135    various concentrations (0.3 to 30 mg/mL) of beechwood xylan (BWX, Megazyme) in activity

136    assay buffer (50 mM sodium phosphate, 12 mM sodium citrate pH 6, supplemented with 1

137    mg/mL of bovine serum albumin (BSA, Merck KGaA, Darmstadt, Germany). DNSA assays

138    were performed to determine the specific activity (SA) of xylanase derivatives (5 nM) on

139    0.5% (w/v) wheat arabinoxylan (WAX), 0.5% (w/v) rye arabinoxylan (RAX) or 1% (w/v)

140    BWX respectively (all from Megazyme, Bray, Ireland) in 50 mM Tris-HCl pH 7.5,

141    supplemented with 1 mg/mL of BSA as previously described [27]. SA of the xylanases (100

142  nM) was also determined using 4-nitrophenyl-β-D-xylotrioside (*p*NP-X$_3$, LIBIOS, France)

143  using 5 mM of substrate in activity assay buffer as previously described [25]. SA were

144  expressed in µmoles of product formed per min per µmole of enzyme (IU/µmole) in order to

145  integrate the differences of mass concentration [25]. To investigate enzyme activity on

146  complex substrates, 5 g destarched wheat bran and 10 g wheat straw (0.5 mm) (both from

147  ARD, Pomacle, France), were each washed in 2 L of deionized water for 1 h at 4 °C and

148  recovered by filtration (0.45 µm) before drying at 50 °C for 3 d. For enzyme assays, wheat

149  bran or straw (2 % w/v) were incubated overnight in 1.9 mL of activity assay buffer before

150  the addition of enzyme (final concentration 1 µM). Reactions were conducted at 37 °C under

151  constant mixing at 1200 rpm (ThermoMixer® C, Eppendorf, Hamburg, Germany). Reaction

152  progress was monitored by regular sampling as previously described [25]. All experiments

153  were performed in triplicate, and the reported values are the means of three experiments ± SD.

154  Kinetic parameters were derived from the data using the Michaelis-Menten equation

155  embedded in SigmaPlot 11.0 (Systat Software,San Jose, CA, USA).

156

157  **Isothermal titration calorimetry (ITC)**

158  ITC measurement was carried out using a MicroCal VP-ITC titration calorimeter (Malvern

159  Panalytical, Malvern, UK) as detailed in Supporting Information. Integrated heat effects were

160  analyzed by non-linear regression using a single-site binding model (Microcal ORIGIN

161  software, version 7.0, Microcal Software), yielding values for the association constant Ka (M$^{-1}$

162  ) and the binding enthalpy ΔH (J.mol$^{-1}$). Other thermodynamic parameters were calculated

163  using the standard thermodynamic equation:

164  -RTlnKa= ΔG=ΔH-TΔS.

165

166  **Microscale Thermophoresis (MST)**

167  MST measurement was carried out using a Monolith NT115 (NanoTemper Technologies

168  GmbH, München, Germany) at 25 °C, 20% LED power and 40 % MEST power as detailed in

169  Supporting Information. Data analysis was performed with MO Affinity software 2.1

170  (Nanotemper). The Hill equation was chosen to determine a value for EC50 [30]. EC50 is the

171  half-maximal effective concentration, meaning the more affinity for the substrate is, the

172  smaller the value of the EC50 will be.

173

174  **Nuclear Magnetic Resonance (NMR)**

175  NMR experiments were performed on a Bruker Avance III HD 800 MHz spectrometer

176  equipped with a 5 mm quadruple resonance QCI-P (H/P-C/N/D) cryogenically cooled probe

177  head (Bruker, Billerica, MA, USA) as detailed in Supporting Information. Analysis of the

178  data and fitting of the chemical-shift perturbation to the standard equation for a saturation

179  isotherm was performed using SigmaPlot 11.0 (Systat Software, San Jose, CA, USA).

180

181  **Small angle X-ray scattering (SAXS)**

182  SAXS measurements were performed at Laboratoire de Génie Chimique, Toulouse, on the

183  XEUSS 2.0 bench (Xenocs, Grenoble, France) equipped with a copper internal source

184  (Genix3D) that produces a X-ray beam (8 keV and $30.10^6$ ph.s$^{-1}$) providing a size resolution

185  of approximately 500 × 500 μm. Proteins were concentrated to approximately 10 mg/mL

186  using a centrifugal filter device (Amicon® Ultra 30 or 50K, Merck KGaA). To remove

187  aggregates and obtain a monodisperse solution, samples (50 µL) were injected onto a size

188  exclusion column mounted on a HPLC coupled to the SAXS. For direct analysis, sample

189  aliquots (40 µL) were transferred from the sample holder (maintained at 18 °C using a

190  circulating water bath) to the measurement cell placed under vacuum to limit air absorption.

191  Data were collected on a 150 × 150 mm area DECTRIS detector (Pilatus 1M) at a sample-

192     detector distance of 1.216 m, thus procuring a measurement range from 0.005 to 0.5 $\text{Å}^{-1}$.

193     Each sample dataset is an average of at least 6 measurements with a data collection time of

194     1,800 s. The averaged curves obtained using direct injection and SEC-HPLC were merged to

195     obtain a composite curve devoid of an aggregation contribution at small angles and displaying

196     low noise at high angles. Finally, to obtain the absolute scattering intensity I(q) for the

197     solutes, the background buffer solution contribution was subtracted from the total SAXS

198     profile. Data integration and reduction were performed using FOXTROT software. The

199     biophysical parameters, such as gyration radius (Rg), maximal distance (Dmax) and Porod

200     volume were calculated using PRIMUS [28] from the ATSAS suite. Low resolution shapes

201     were calculated with DAMMIF and rigid body molecular modelling using the SAXS data was

202     performed using CORAL.

203

204     **Paraffin embedding, microtomy, on-section enzymatic digestion and double**

205     **immunofluorescence labelling.**

206     Sample preparation was performed essentially as previously described [29] with minor

207     modifications. Briefly, wheat straw (1 cm long) and wheat bran were fixed in acetic

208     acid/ethanol without aldehyde, infiltrated in paraplast and assembled as tissue arrays in

209     paraplast. Tissue arrays corresponded to hundreds of wheat bran fragments and at least four

210     wheat straw cross sections. Tissue array blocks were soaked for several weeks in acetic

211     acid/ethanol softening solution at 4 °C [30] and 14 µm thick serial sections were displayed on

212     silane coated slides. Individual slides were dewaxed and covered with a microincubation

213     chamber (22×40×0.2 mm deep; (200 µL), #70324-20, Electron Microscopy Science, Hatfield,

214     PA, USA) containing either 50 mM sodium-phosphate 12 mM citrate buffer pH 6 alone or

215     containing enzymes (300 nM final concentration). Slides were incubated for 24 h at 37 °C in a

216     humid atmosphere and then recombinant enzymes were digested with proteinase K (5 µg/ml

217 in 0.1 M Tris-HCl pH 8-50 mM EDTA pH 8) for 30 min at 37 °C. Double

218 immunofluorescence labelling was performed as previously described [31] using xylan-

219 specific LM11 monoclonal antibody (PlantProbes, Leeds, UK) and cellulose specific His$_6$-

220 Tagged $Ct$CBM3a recombinant protein (PlantProbes, Leeds, UK) as primary probes. These

221 were labelled with goat anti-rat IgG-Alexa Fluor 488 (ThermoFischer, Waltham, MA, USA)

222 and His$_6$-Tag monoclonal antibody (4E3D10H2/E3)-Alexa Fluor 555 (ThermoFischer,

223 Waltham, MA, USA), respectively. Slides were mounted in ProLong™ Gold Antifade

224 mounting medium (ThermoFischer, Waltham, MA, USA) and scanned with a Nanozoomer

225 2.0 RS scanner (Hamamatsu photonics, Hamamatsu City, Japan) using a 40x objective and

226 7×1 µm Z stacks. Scans were analysed with NDP view (Hamamatsu photonics). The lignin

227 autofluorescence observed for the A488 channel in all the cell walls from the

228 untreated/unlabelled sections was set to a minimum threshold to observe the LM11/A488

229 specific labelling on the other sections. No such an autofluorescence is seen for the A555

230 channel. Figures were assembled using Photo-paint (Corel draw graphics suite 2018, Corel,

231 Ottawa, Canada).

232

233 **Results and Discussion**

234

235 **Investigating the impact of alternative multidomain arrangements on biological activity**

236 In previous work [25] it was demonstrated how Jo and In can be used to conveniently link

237 proteins together [24]. Therefore, Jo and In were exploited to create a series of GH-CBM

238 chimeric proteins based on $Np$Xyn11A. The expression of His$_6$-tagged $Np$Xyn11A linked to

239 either Jo or In at its $N$-terminus [27] yielded active enzymes. Data regarding In-$Np$Xyn11A is

240 already available [27] so work was restricted to the comparison of the kinetic parameters of

241 Jo-$Np$Xyn11A with those of wild type $Np$Xyn11A (Supplementary Table S3). Values of $K_M$

242 $_{app}$ for In-*Np*Xyn11A and Jo-*Np*Xyn11A (1.8 mg.mL$^{-1}$ and 2.76 mg.mL$^{-1}$) are 2.4- and 3.7-

243 fold higher than that of the wild type enzyme (0.75 mg.mL$^{-1}$), indicating that the affinity for

244 BWX is lowered. As discussed previously [25], additions at the *N*-terminal extremity of the

245 *Np*Xyn11A probably hinder the flexible loops that connect the β-sheets and form the catalytic

246 pocket that accommodates glycone moieties. This is inferred by the fact that even subtle

247 modifications in this highly conserved region lead to significant alterations in enzyme activity

248 [32]. SAXS data (see Figure 3 below) revealed that in solution In-*Np*Xyn11A, displayed a

249 partially unfolded conformation compared to *Np*Xyn11A. These observations can possibly be

250 correlated with the relatively high value of the standard deviation of the kinetic parameters.

251 Despite changes to $K_{Mapp}$ and $k_{cat}$, the overall catalytic efficiency of the two chimeric enzymes

252 and *Np*Xyn11A were similar due to compensatory effects on $k_{cat}$ and $K_{Mapp}$.

253

254 To use the Jo-In system to assess rapidly the effect of covalent linkage of *Np*Xyn11A to

255 *Ct*CBM3a and *Cf*CBM2b-1, the same strategy was used to prepare CBM chimeras, yielding

256 the proteins Jo-CBM3a, In-CBM3a, Jo-CBM2b-1 and In-CBM2b-1. A pull down assay [33]

257 performed using insoluble cellulose confirmed the binding ability of *Ct*CBM3a and its Jo and

258 In derivatives (see supporting information for detailed protocol, Supplementary Figure S1).

259 Further investigation using Isothermal titration calorimetry (ITC) and either regenerated

260 cellulose (RC) or cellulose nanocrystals (CN) confirmed that *Ct*CBM3a showed similar $K_a$

261 values in the case of both ligands (see supporting information for detailed protocol,

262 Supplementary Figure S2). It is noteworthy that ITC data also indicate that the *Ct*CBM3a-

263 ligand interaction is enthalpy-driven, with the entropic component being unfavorable,

264 consistent with previous data related to CBMs binding to RC [34]. Having demonstrated that

265 CNs constitute a suitable ligand for *Ct*CBM3a, they were used to evaluate the binding

266 properties of Jo-CBM3a and In-CBM3a using microscale thermophoresis (MST) (see

267    supporting information for detailed protocol, **Table 1** and Supplementary Figure S3). Since

268    Jo- and In-CBM3a are optimally stable in different buffers, to compare them with *Ct*CBM3a

269    it was necessary to determine CN $EC_{50}$ values for the latter in both buffers. Comparing Jo-

270    CBM3a with *Ct*CBM3a revealed that the $EC_{50}$ values were highly similar (0.21 $g.L^{-1}$ and 0.14

271    $g.L^{-1}$, respectively). However, the $EC_{50}$ value characterizing the interaction of In-CBM3a with

272    CN was 60-fold lower than that of *Ct*CBM3a (0.003 $g.L^{-1}$ and 0.18 $g.L^{-1}$, respectively),

273    indicating a higher affinity of In-CBM3a for CN. In this respect, it is noteworthy that

274    *Ct*CBM3a binding to CN is only moderately sensitive to buffer changes, because the

275    difference between the two $EC_{50}$ values was only 1.28-fold. Therefore, accounting for buffer

276    effects, while the appendage of Jo to the *N*-terminal extremity of *Ct*CBM3a has a relatively

277    minor impact on ligand binding, the appendage of In significantly reinforces it.

278

279    To investigate the ligand binding properties of *Cf*CBM2b-1 and its Jo and In derivatives for

280    soluble oligosaccharides, NMR was used (see supporting information for detailed protocol).

281    This revealed that CBM binding to $X_6$ significantly perturbed the chemical shift of the side

282    chains $NH^{\epsilon}$ signals of two solvent exposed Trp residues (Supplementary Figure S4).

283    Monitoring these shifts provided a $K_d$ value for the *Cf*CBM2b-1/$X_6$ interaction (**Table 2**),

284    with data being in the same order of magnitude as a previously reported value [35].

285    Determination of the $K_d$ value for $X_6$ binding to In-CBM2b-1 gave a similar value, but in

286    identical assay conditions no interaction (i.e. $K_d > 10$ mM) between Jo-CBM2b-1 and $X_6$ was

287    evidenced. The reason for this difference is unknown, but clearly Jo engages in unfavorable

288    interactions with CBM2b-1 that possibly lead to the steric hindrance of one or more of the

289    CBM's ligand binding determinants [35].

290

291    In summary, the attachment of Jo or In to the *N*-terminus of *Np*Xyn11A yielded an active

292    xylanase. However, attachment of these elements to the *N*-termini of *Ct*CBM3a or *Cf*CBM2b-

293    1 yielded variable results. The attachment of In increased the binding affinity for both CBMs

294    (significantly for CBM3a) while the attachment of Jo decreased both binding affinities

295    (drastically for CBM2b-1). Nevertheless, despite this it was decided to proceed with the

296    creation of Jo-In linked multimodular chimeras, because previous work has shown that the

297    covalent association of Jo and In leads to a stable complex devoid of flexibility [24].

298    Similarly, it was postulated that any deleterious effects arising from the linkage of either Jo or

299    In to a CBM might be attenuated once the Jo-In complex is formed.

300

301    **Creation and biochemical characterization of multimodular chimeric enzymes**

302    Using the different Jo and In derivatives, four multimodular chimeric xylanases were

303    prepared and purified (**Figure 1**A-B). As shown by SDS-PAGE, the apparent molecular

304    weight (MW) of *Np*Xyn11A-[Jo-In]-CBM3a, *Np*Xyn11A-[In-Jo]-CBM3a, *Np*Xyn11A-[Jo-

305    In]-CBM2b-1 and *Np*Xyn11A-[In-Jo]-CBM2b-1 are consistent with predicted values,

306    obtained by summing the MWs of the individual modules (Figure 1B, Supplementary Table

307    S2).

308

309    Measurement of hydrolytic activity of multimodular chimeric xylanases on *p*NP-X$_3$, a

310    substrate small enough to avoid major interference from the CBM, showed that all the

311    enzymes display activities in the same order of magnitude as recombinant *Np*Xyn11A (**Table

312    3**). The activity of *Np*Xyn11A-[In-Jo]-CBM2b-1 was reduced by 36%, even though In-

313    *Np*Xyn11A displayed almost the same activity as *Np*Xyn11A. Conversely, while the

314    attachment of Jo to *Np*Xyn11A led to a 20% reduction in activity on *p*NP-X$_3$, adding In-

315    CBM2b-1 restored activity to a level almost identical to that of *Np*Xyn11A. Moreover, the

316    addition of In-CBM3a was not deleterious (Table 3). These results are consistent with current

317    knowledge that the presence of CBM appendages does not enhance the activity of GHs on

318    soluble substrates such as *p*NP-glycosides or short oligosaccharides [36].

319

320    Regarding the activities of the four chimeric multimodular xylanases on soluble polymeric

321    substrates (RAX, WAX and BWX), the enzymes fall into one of two categories. When In-

322    *Np*Xyn11A is linked to either of the Jo-CBMs the activity on all three substrates is mostly

323    lower than *Np*Xyn11A, with CBM2b-1 having the most deleterious effect of up to 66% loss

324    of activity. This result could be related to the loss of binding properties of Jo-CBM2b-1

325    (Table 2) and does not follow examples in the literature suggesting that activity enhancement

326    could be expected [37]. Nevertheless, when Jo-*Np*Xyn11A is linked to either In-CBMs,

327    activity on the different substrates is mostly increased, with the exception of *Np*Xyn11A-[Jo-

328    In]-CBM2b-1 on BWX (10% activity loss) (Table 3). Remarkably, the activity of *Np*Xyn11A-

329    [Jo-In]-CBM3a on RAX was increased by 135% compared to that of *Np*Xyn11A on the same

330    substrate (2.92 ± 0.04 mM xylose equivalent and 1.25 ± 0.16 mM xylose equivalent after 15

331    min of reaction, respectively). Clearly, activity increases related to the presence of CBM2b-1

332    can be tentatively attributed to the specific ligand binding ability of the CBM. However, the

333    significant increase correlated with the presence of the cellulose-targeting CBM3a is less

334    intuitive. Nevertheless, a recent study also revealed that the appendage of a cellulose-specific

335    CBM family 1 to xylanase *Np*Xyn11C [38] increased catalytic efficiency by 21% on BWX.

336    The underlying reasons for such activity enhancements is unclear and are often treated

337    cautiously [39]. The fold and architecture of a xylanase core was proposed to explain the

338    positive effect of a CBM targeting xylan on the catalytic activity of GH family 11 towards

339    soluble xylan [37]. In the case of the present results, accounting for the relatively large MW

340    of the xylan polymers (~ 350 kDa) [27] and the shape of the chimeric enzyme (see SAXS data

341     below, Figure 5B), it may be postulated that non-specific interactions cannot be excluded,

342     although with no evidence for this.

343

344     The enzymatic activity of *Np*Xyn11A derivatives was also evaluated using destarched wheat

345     bran (DWB), which displays a high arabinoxylan:cellulose ratio [40] and wheat straw (WS),

346     which conversely displays a low arabinoxylan:cellulose ratio [41]. After 23 h, chimeric

347     xylanases had released ~7.5-fold more reducing sugars from DWB than from WS (**Figure 2**),

348     consistent with the greater availability of arabinoxylan in the former and also the greater

349     structural and chemical complexity of WS. Moreover, the presence of either CBM clearly

350     enhanced final reducing sugar yield (by ~17% in the case of DWB), even with the presence of

351     *Ct*CBM3a that apparently reduced the initial reaction rate (Figure 2A). The impact of the

352     nature of the Jo-In linkage was also significant with *Np*Xyn11A-[Jo-In]-CBM2b-1 displaying

353     a faster initial rate than *Np*Xyn11A-[In-Jo]-CBM2b-1. On WS, the presence of *Ct*CBM3a

354     proved to be a severe handicap, because activity was ~75% lower than that of the catalytic

355     domain alone (Figure 2B). Instead, the impact of *Ct*CBM2b-1 was imperceptible, since the

356     activities of the *Ct*CBM2b-1 chimeras were nearly identical to that of the catalytic domain

357     alone. However, the discriminating nature of the Jo-In linkage was again perceptible, because

358     despite its faster initial rate, *Np*Xyn11A-[In-Jo]-CBM2b-1 generated ~15 % less reducing

359     sugars when compared to *Np*Xyn11A-[Jo-In]-CBM2b-1. These results demonstrate the

360     importance of substrate targeting by CBMs, especially in complex environments such as

361     PCWs [15]. DWB provides *Cf*CBM2b-1 with an abundant source of highly accessible ligand

362     binding sites, whereas the cellulose-specific *Ct*CBM3a probably hinders the early progression

363     of the enzyme in this matrix. In contrast, WS provides *Ct*CBM3a with abundant crystalline

364     cellulose. Thus, binding of *Ct*CBM3a chimeras to cellulose sequesters the enzyme and

365     prevents it from reaching its arabinoxylan target substrate [15,42]. It is noteworthy that

366    although MST and NMR measurements revealed that In-CBM3a and Jo-CBM2b-1 display

367    impaired ligand binding (Tables 1 and 2) and experiments using purified xylan substrates

368    suggest that the activities of *Np*Xyn11A chimeras are sensitive to the exact nature of the Jo-In

369    linkage (Table 3), these factors did not appear to be major determinants of activity on

370    complex insoluble substrates.

371

372    **SAXS and NMR analysis of the multimodular xylanases**

373    As previously described [24], the anti-parallel organization of the stable complex Jo-In and

374    the possibility to link Jo or In to the *N*- and *C*-termini of proteins of interest [25] offers the

375    ability to create chimeric proteins and modulates the relative spatial orientation of linked

376    protein domains. To examine the structures of the protein chimeras created in this work,

377    SAXS data (**Figure 3**) were recorded and biophysical parameters were extracted (Figure 3C).

378    The single domain *Np*Xyn11A generated a SAXS curve typical of a globular, folded protein

379    (Figure 3A) that fits well with the theoretical  curve calculated using CRYSOL and crystal

380    structure data (PDB id: 2C1F) (data not shown). However, addition of the In domain resulted

381    in a modified solution structure, with the Rg and Dmax values (46.2Å and 160Å respectively)

382    being considerably higher than those of *Np*Xyn11A (18.5Å and 60Å respectively), in

383    agreement with the increase of the MW. It appears that the In domain of In-*Np*Xyn11A is

384    present as a long unfolded tail, while the *Np*Xyn11A maintains its globular structure. SAXS

385    curves of Jo and In derivatives of *Cf*CBM2b-1 displayed the characteristics of unfolded

386    proteins with a constant decreased at medium angles (0.008 to 0.1Å$^{-1}$). The proton NMR

387    spectrum of the isolated *Cf*CBM2b-1 is characteristic of a well-folded protein domain, with

388    several methyl resonances below 0 ppm and well-defined Trp side chain signals (**Figure 4**).

389    The Jo-CBM2b-1 and In-CBM2b-1 constructs display similar resonances devoid of chemical

390    shifts or spectral broadening, indicating that the CBM domain maintains its 3D fold within

391  these constructs. Indeed, calculation of a difference spectrum (i.e. subtracting the spectrum of

392  the isolated *Cf*CBM2b-1 domain from that of the Jo/In derivatives) confirmed that the spectra

393  of the Jo/In -attached *Cf*CBM2b-1 are simple composites of the spectra of the constituent

394  proteins domains. This implies that when linked to *Cf*CBM2b-1, neither Jo nor In intrinsically

395  alter the structure of the CBM. However, the fact that Jo-CBM2b-1 binds less effectively than

396  *Cf*CBM2b-1 to $X_6$ (Table 2) suggests that the unstructured Jo domain obstructs access to the

397  CBM's ligand binding site. A similar conclusion possibly explains the lower activity of In-

398  *Np*Xyn11A against BWX, since the SAXS data (Figure 3) indicates that the In domain is

399  partially unstructured and might obstruct access to the catalytic site (Supplementary Table S3

400  and Table 3).

401  The second set of SAXS curves presented in Figure 3B are those of the multimodular

402  xylanases. These display a similar profile, with intensity decreasing to a plateau at small

403  angles (q <0.01 $\text{Å}^{-1}$), corresponding to the Guinier region. At higher angles (0.01<q<0.08 $\text{Å}^{-1}$

404  and 0.08<q<0.2 $\text{Å}^{-1}$) the curves decay, consistent with the power law function $I(q) = q^{-p}$, with

405  p value $\approx$ 2 and 4 respectively. This is characteristic of elongated proteins. For the

406  multimodular xylanases, *Np*Xyn11A-[In-Jo]-CBM2b-1 and *Np*Xyn11A-[Jo-In]-CBM2b-1,

407  the plots of P(r) *versus* r are very similar (**Figure 5**A) and reflect a multidomain, elongated,

408  global form (Figure 5B). In contrast, although the P(r) profiles of *Np*Xyn11A-[In-Jo]-CBM3a

409  and *Np*Xyn11A-[Jo-In]-CBM3a are also highly similar, they nevertheless differ from those of

410  *Np*Xyn11A-[In-Jo]-CBM2b-1 and *Np*Xyn11A-[Jo-In]-CBM2b-1, displaying more marked

411  oscillations that reflect the larger size of *Ct*CBM3a (17.1 kDa compared to 9.1 kDa for

412  CBM2b-1). Despite this difference, the curves are also indicative of elongated shapes

413  composed of distinct domains (Figure 5B). The comparison of experimental SAXS data

414  acquired for *Np*Xyn11A-[In-Jo]-CBM2b-1 and *Np*Xyn11A-[Jo-In]-CBM2b-1 with that of

415  model curves generated using crystallographic data revealed that these were highly similar,

416    with a goodness of fit $\chi^2$ = 1.25 and 1.57 respectively (**Figure 6**A). For each chimera,

417    modelling and superimposing the theoretical structure that displayed the best $\chi^2$ value

418    provided low-resolution hypothetical structures (Figure 6B-C). Irrespective of the Jo-In

419    configuration, the distance between *Np*Xyn11A and *Cf*CBM2b-1 is quite similar in both

420    models (118.0 ± 9.6Å for *Np*Xyn11A-[In-Jo]-CBM2b-1 and 120.6 ± 5.5 Å for *Np*Xyn11A-

421    [Jo-In]-CBM2b-1). However, due to the axial asymmetry in the Jo-In complex, the torsion

422    angle is + 66.3 ± 27.6° for *Np*Xyn11A-[In-Jo]-CBM2b-1 and -145.4 ± 21.1° for *Np*Xyn11A-

423    [Jo-In]-CBM2b-1 (Supplemnetary Figure S6). Predictive structural modelling of *Np*Xyn11A-

424    [In-Jo]-CBM3a and *Np*Xyn11A-[Jo-In]-CBM3a using CORAL was not possible because of

425    significant structural variability (i.e. $\chi^2$ > 2) that prevented fitting to the molecular envelope of

426    the SAXS data.

**Differences in targeting multimodular xylanases in wheat straw**

428    To further test the possibility of using the chimeric xylanases on raw substrate, experiments

429    were performed *in situ* on wheat bran and wheat straw PCW sections and the accessible

430    xylans and cellulose localization were monitored using immunological labelling. Following

431    treatment of the wheat bran PCW sections with multimodular enzymes (*Np*Xyn11A-[Jo-In]-

432    CBM3a was omitted from this study), LM11/A488-specific xylan labelling between the

433    pericarp and the nucellar epidermis (**Figure 7** C1-F1) and intense continuous CBM3a/A555-

434    specific labelling of the nucellar epidermis (Figure 7 C2-F2) were observed. This contrasts

435    with the untreated sections that displayed no LM11/A488 labelling and only faint

436    CBM3a/A555 labelling (Figure 7B). Although increased labelling after enzymatic treatment

437    appears counter intuitive, it is almost certainly a consequence of PCW complexity. The action

438    of the chimeric xylanases leads to the removal of xylan and concomitant exposure of hitherto

439    masked PCW components, such as cellulose and xylan, which constitute new ligands for

440    antibody or CBM binding. However, experiments performed on DWB failed to reveal any

441   major changes to labelling. Therefore, subsequent work focused on WS (**Figure 8**,

442   Supplementary Figures S7 and S8). The use of either probe on untreated sections produced

443   homogenous labelling of both PCW components (Figure 8B). However, after treatment with

444   *Np*Xyn11A (Figure 8C) and *Np*Xyn-[Jo-In]-CBM2b-1 (Figure 8D), LM11/A488 labelling

445   was repeatedly found to be diminished when observing different vascular bundles

446   (Supplementary Figure S7). This suggests that the presence of CBM2b-1 did not enhance the

447   activity of *Np*Xyn11A on WS. Additionally, compared to *Np*Xyn11A, *Np*Xyn-[Jo-In]-

448   CBM2b-1 was apparently less active on intervascular fibres. Finally, diminution of the

449   CBM3a/A555 labelling (cellulose-specific probe) was correlated with xylan hydrolysis.

450   Conversely, treatment with *Np*Xyn11A-[In-Jo]-CBM2b-1 or *Np*Xyn11A-[In-Jo]-CBM3a

451   generally did not affect LM11/A488 labelling intensity (Figure 8E-F; Supplementary Figure

452   S7), although this was strongly enhanced in patches around the phloem, the protoxylem and in

453   vascular bundle cell corners. In distal zones, intense patches of LM11/A488 labelling were

454   also observed in sclerenchyma cell corners and around the pith parenchyma intercellular

455   spaces (Supplementary Figure S8). For PCW sections treated with *Np*Xyn11A-[In-Jo]-

456   CBM2b-1, additional patches of LM11/A488 labelling were also observed in the intervascular

457   fibres, but this was not the case when *Np*Xyn11A-[In-Jo]-CBM3a was used (Figure 8E-F;

458   Supplementary Figure S7). The impact on cellulose labelling was also more evident for

459   *Np*Xyn11A-[In-Jo]-CBM2b-1 than for *Np*Xyn11A-[In-Jo]-CBM3a (Figure 8E-F;

460   Supplementary Figure S7).

461

462   Previously, both *Ct*CBM3a and *Cf*CBM2b-1 were shown to potentiate the activity of a

463   cognate xylanase on tobacco PCWs, presumably by improving its substrate targeting

464   capability [15]. However, the present results do not confirm this and may signify limited

465   usefulness for the Jo-In system when studying enzyme activity on complex matrices, even

466   though the structural environment of xylan and xylan-cellulose interactions are different in

467   tobacco PCW [43]. It can be postulated that Jo-In introduces a high degree of rigidity

468   compared to natural linkers found in PCW-degrading enzymes. While rigidity might be

469   undesirable in certain circumstances, it can be useful in others. Specifically, greater protein

470   rigidity will be useful in investigating different spatial organizations in multidomain proteins,

471   locking chimeric protein isoforms in different domain configurations.

472

473   **Conclusions**

474         A previous study on *Ct*CBM3a linked to endoglucanase CelA of *Clostridium*

475   *thermocellum* revealed that the CBM must be correctly oriented to potentiate enzymatic

476   activity, especially on insoluble substrates [44]. In this regard, the nature of the linker is

477   certainly important because its structure will strongly contribute to the spatial orientations of

478   linked protein domains. Here, the Jo-In complex has been used to link protein domains, the

479   resultant linker complexes being stable, rather rigid protein structures. The first implication of

480   this rigidity is the likelihood that there will be no physical interference between the attached

481   protein domains. In the case of *Np*Xyn11A linked to CBMs, this is certainly the case because

482   the solution structures are elongated and characterized by well separated GH and CBM

483   domains. The second implication is that the spatial orientation (torsion angle) of the two

484   linked proteins domains is locked. If this is optimal then synergy should be possible and

485   enhancement of the activity of the GH domain will be a likely outcome. Conversely, if this is

486   suboptimal, the chimera will be definitively impaired, lacking sufficient linker flexibility to

487   allow alternative solution conformers. Accordingly, it is proposed that data related to the

488   hydrolysis of insoluble substrates described herein reflects this fact, revealing that

489   *Np*Xyn11A-[Jo-In]-CBM2b-1 has a more optimal configuration for hydrolysis than

490   *Np*Xyn11A-[In-Jo]-CBM2b-1. The third implication of linker rigidity is that protein chimeras

491     are likely to encounter difficulty in penetrating structurally complex three-dimensional

492     matrices. In this work, the fact that the addition of CBMs to *Np*Xyn11A failed to potentiate

493     the hydrolysis of wheat straw possibly supports this hypothesis.

494        In summary, this work confirms the usefulness of the Jo-In system in creating

495     multidomain GHs. However, one caveat is the intrinsic rigidity of Jo-In. When designing

496     chimeric proteins, this property must be considered with respect to the intended purpose.

497

506

507 **Author contributions**

508     LB, VB, CYM conceived and designed the experiments. LB, PR, TE, GL, IV and CYM

509     performed the experiments. LB, VB, PR, GL, MJO and CYM analyzed the data. LB, MJO

510     and CYM wrote the paper.

511

512 **References**

513     [1]    Wang Y, Zhang H, Zhong H, Xue Z. Protein domain identification methods and online

514         resources. Comput Struct Biotechnol J 2021;19:1145–53.

515         https://doi.org/10.1016/j.csbj.2021.01.041.

516  [2]  Apic G, Gough J, Teichmann SA. Domain combinations in archaeal, eubacterial and

517       eukaryotic proteomes. J Mol Biol 2001;310:311–25.

518       https://doi.org/10.1006/jmbi.2001.4776.

519  [3]  Apic G, Huber W, Teichmann SA. Multi-domain protein families and domain pairs:

520       comparison with known structures and a random model of domain recombination. J

521       Struct Funct Genomics 2003;4:67–78. https://doi.org/10.1023/A:1026113408773.

522  [4]  Altschuh D, Tessier DC, Vernet T. Modulation of the enzymatic activity of papain by

523       interdomain residues remote from the active site. Protein Eng Des Sel 1994;7:769–76.

524       https://doi.org/10.1093/protein/7.6.769.

525  [5]  Valentini G, Chiarelli L, Fortin R, Speranza ML, Galizzi A, Mattevi A. The Allosteric

526       Regulation of Pyruvate Kinase. J Biol Chem 2000;275:18145–52.

527       https://doi.org/10.1074/jbc.M001870200.

528  [6]  Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The

529       carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res

530       2014;42:D490–5. https://doi.org/10.1093/nar/gkt1178.

531  [7]  Amos RA, Mohnen D. Critical review of plant cell wall matrix polysaccharide

532       glycosyltransferase activities verified by heterologous protein expression. Front Plant

533       Sci 2019;10. https://doi.org/10.3389/fpls.2019.00915.

534  [8]  Wu H, Ioannou E, Henrissat B, Montanier CY, Bozonnet S, O'Donohue MJ, et al.

535       Multimodularity of a GH10 Xylanase Found in the Termite Gut Metagenome. Appl

536       Environ Microbiol 2021;87. https://doi.org/10.1128/AEM.01714-20.

537  [9]  Henrissat B, Davies GJ. Glycoside Hydrolases and Glycosyltransferases. Families,

538       Modules, and Implications for Genomics. Plant Physiol 2000;124:1515–9.

539       https://doi.org/10.1104/pp.124.4.1515.

540  [10] Shoseyov O, Shani Z, Levy I. Carbohydrate binding modules: biochemical properties

541      and novel applications. Microbiol Mol Biol Rev 2006;70:283–95.

542      https://doi.org/10.1128/MMBR.00028-05.

543 [11] Sammond DW, Payne CM, Brunecky R, Himmel ME, Crowley MF, Beckham GT.

544      Cellulase linkers are optimizedbBased on domain type and function: insights from

545      sequence analysis, biophysical measurements, and molecular simulation. PLoS One

546      2012;7:e48615. https://doi.org/10.1371/journal.pone.0048615.

547 [12] Sidar A, Albuquerque ED, Voshol GP, Ram AFJ, Vijgenboom E, Punt PJ.

548      Carbohydrate binding modules: diversity of domain architecture in amylases and

549      cellulases from filamentous microorganisms. Front Bioeng Biotechnol 2020;8.

550      https://doi.org/10.3389/fbioe.2020.00871.

551 [13] Pasari N, Adlakha N, Gupta M, Bashir Z, Rajacharya GH, Verma G, et al. Impact of

552      Module-X2 and Carbohydrate Binding Module-3 on the catalytic activity of associated

553      glycoside hydrolases towards plant biomass. Sci Rep 2017;7:3700.

554      https://doi.org/10.1038/s41598-017-03927-y.

555 [14] Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B, et al. How nature

556      can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic

557      specificity. Proc Natl Acad Sci 2012;109:20889–94.

558      https://doi.org/10.1073/pnas.1212034109.

559 [15] Herve C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. Carbohydrate-

560      binding modules promote the enzymatic deconstruction of intact plant cell walls by

561      targeting and proximity effects. Proc Natl Acad Sci 2010;107:15293–8.

562      https://doi.org/10.1073/pnas.1005732107.

563 [16] Shah NH, Muir TW. Inteins: nature's gift to protein chemists. Chem Sci 2014;5:446–

564      61. https://doi.org/10.1039/C3SC52951G.

565 [17] Pinto F, Thornton EL, Wang B. An expanded library of orthogonal split inteins enables

566        modular multi-peptide assemblies. Nat Commun 2020;11:1529.

567        https://doi.org/10.1038/s41467-020-15272-2.

568   [18]  Zhao X, Scheffner M, Marx A. Assembly of branched ubiquitin oligomers by click

569        chemistry. Chem Commun 2019;55:13093–5. https://doi.org/10.1039/C9CC07303E.

570   [19]  You C, Piehler J. Multivalent chelators for spatially and temporally controlled protein

571        functionalization. Anal Bioanal Chem 2014;406:3345–57.

572        https://doi.org/10.1007/s00216-014-7803-y.

573   [20]  Singh S, Kluger R. Self-Assembly of a Functional Triple Protein: Hemoglobin-Avidin-

574        Hemoglobin via Biotin–Avidin Interactions. Biochemistry 2016;55:2875–82.

575        https://doi.org/10.1021/acs.biochem.6b00215.

576   [21]  Fierobe H-P, Mechaly A, Tardif C, Belaich A, Lamed R, Shoham Y, et al. Design and

577        Production of Active Cellulosome Chimeras. J Biol Chem 2001;276:21257–61.

578        https://doi.org/10.1074/jbc.M102082200.

579   [22]  Sun F, Zhang W-B. Unleashing chemical power from protein sequence space toward

580        genetically encoded "click" chemistry. Chinese Chem Lett 2017;28:2078–84.

581        https://doi.org/10.1016/j.cclet.2017.08.052.

582   [23]  Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, et al. Peptide

583        tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin.

584        Proc Natl Acad Sci 2012;109:E690–7. https://doi.org/10.1073/pnas.1115485109.

585   [24]  Bonnet J, Cartannaz J, Tourcier G, Contreras-Martel C, Kleman JP, Morlot C, et al.

586        Autocatalytic association of proteins by covalent bond formation: a Bio Molecular

587        Welding toolbox derived from a bacterial adhesin. Sci Rep 2017;7:43564.

588        https://doi.org/10.1038/srep43564.

589   [25]  Enjalbert T, De La Mare M, Roblin P, Badruna L, Vernet T, Dumon C, et al.

590        Characterisation of the effect of the spatial organisation of hemicellulases on the

591      hydrolysis of plant biomass polymer. Int J Mol Sci 2020;21:4360.

592      https://doi.org/10.3390/ijms21124360.

593   [26]   Vardakou M, Dumon C, Murray JW, Christakopoulos P, Weiner DP, Juge N, et al.

594      Understanding the structural basis for substrate and inhibitor recognition in eukaryotic

595      GH11 xylanases. J Mol Biol 2008;375:1293–305.

596      https://doi.org/10.1016/j.jmb.2007.11.007.

597   [27]   Montanier CY, Fanuel M, Rogniaux H, Ropartz D, Di Guilmi A-M, Bouchoux A.

598      Changing surface grafting density has an effect on the activity of immobilized xylanase

599      towards natural polysaccharides. Sci Rep 2019;9:5763. https://doi.org/10.1038/s41598-

600      019-42206-w.

601   [28]   Konarev P V., Volkov V V., Sokolova A V., Koch MHJ, Svergun DI. PRIMUS : a

602      Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr

603      2003;36:1277–82. https://doi.org/10.1107/S0021889803012779.

604   [29]   Badruna L, Burlat V, Montanier CY. CBMs as probes to explore plant cell wall

605      heterogeneity using immunocytochemistry, 2017, p. 181–97.

606      https://doi.org/10.1007/978-1-4939-6899-2_14.

607   [30]   Brandizzi F. Plant microtechnique and microscopy. vol. 86. 2000.

608      https://doi.org/10.1006/anbo.2000.1231.

609   [31]   Francoz E, Ranocha P, Le Ru A, Martinez Y, Fourquaux I, Jauneau A, et al. Pectin

610      demethylesterification generates platforms that anchor peroxidases to remodel plant

611      cell wall domains. Dev Cell 2019;48:261-276.e8.

612      https://doi.org/10.1016/j.devcel.2018.11.016.

613   [32]   Song L, Siguier B, Dumon C, Bozonnet S, O'Donohue MJ. Engineering better

614      biomass-degrading ability into a GH11 xylanase using a directed evolution strategy.

615      Biotechnol Biofuels 2012;5:3. https://doi.org/10.1186/1754-6834-5-3.

616    [33]    Crouch LI, Labourel A, Walton PH, Davies GJ, Gilbert HJ. The Contribution of non-

617           catalytic carbohydrate binding modules to the activity of lytic polysaccharide

618           monooxygenases. J Biol Chem 2016;291:7439–49.

619           https://doi.org/10.1074/jbc.M115.702365.

620    [34]    Boraston AB. The interaction of carbohydrate-binding modules with insoluble non-

621           crystalline cellulose is enthalpically driven. Biochem J 2005;385:479–84.

622           https://doi.org/10.1042/BJ20041473.

623    [35]    Simpson PJ, Bolam DN, Cooper A, Ciruela A, Hazlewood GP, Gilbert HJ, et al. A

624           family IIb xylan-binding domain has a similar secondary structure to a homologous

625           family IIa cellulose-binding domain but different ligand specificity. Structure

626           1999;7:853–64. https://doi.org/10.1016/S0969-2126(99)80108-7.

627    [36]    Tomme P, Tilbeurgh H, Pettersson G, Damme J, Vandekerckhove J, Knowles J, et al.

628           Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain

629           function in two cellobiohydrolases by limited proteolysis. Eur J Biochem

630           1988;170:575–81. https://doi.org/10.1111/j.1432-1033.1988.tb13736.x.

631    [37]    Kittur FS, Mangala SL, Rus'd AA, Kitaoka M, Tsujibo H, Hayashi K. Fusion of family

632           2b carbohydrate-binding module increases the catalytic activity of a xylanase from

633           Thermotoga maritima to soluble xylan. FEBS Lett 2003;549:147–51.

634           https://doi.org/10.1016/S0014-5793(03)00803-2.

635    [38]    Zhang Y, Yang H, Yu X, Kong H, Chen J, Luo H, et al. Synergistic effect of acetyl

636           xylan esterase from Talaromyces leycettanus JCM12802 and xylanase from

637           Neocallimastix patriciarum achieved by introducing carbohydrate-binding module-1.

638           AMB Express 2019;9:13. https://doi.org/10.1186/s13568-019-0740-6.

639    [39]    Hoffmam ZB, Zanphorlin LM, Cota J, Diogo JA, Almeida GB, Damásio ARL, et al.

640           Xylan-specific carbohydrate-binding module belonging to family 6 enhances the

641          catalytic performance of a GH11 endo-xylanase. N Biotechnol 2016;33:467–72.

642          https://doi.org/10.1016/j.nbt.2016.02.006.

643    [40]    Bouraoui H, Desrousseaux M-L, Ioannou E, Alvira P, Manaï M, Rémond C, et al. The

644          GH51 α-l-arabinofuranosidase from Paenibacillus sp. THS1 is multifunctional,

645          hydrolyzing main-chain and side-chain glycosidic bonds in heteroxylans. Biotechnol

646          Biofuels 2016;9:140. https://doi.org/10.1186/s13068-016-0550-x.

647    [41]    Motte J-C, Escudié R, Beaufils N, Steyer J-P, Bernet N, Delgenès J-P, et al.

648          Morphological structures of wheat straw strongly impacts its anaerobic digestion. Ind

649          Crops Prod 2014;52:695–701. https://doi.org/10.1016/j.indcrop.2013.11.038.

650    [42]    Gao D, Chundawat SPS, Sethi A, Balan V, Gnanakaran S, Dale BE. Increased enzyme

651          binding to substrate is not necessary for more efficient cellulose hydrolysis. Proc Natl

652          Acad Sci 2013;110:10922–7. https://doi.org/10.1073/pnas.1213426110.

653    [43]    Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, et al.

654          The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils

655          as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana.

656          Plant J 2014;79:492–506. https://doi.org/10.1111/tpj.12575.

657    [44]    Sajjad M, Khan MIM, Zafar R, Ahmad S, Niazi UHK, Akhtar MW. Influence of

658          positioning of carbohydrate binding module on the activity of endoglucanase CelA of

659          Clostridium thermocellum. J Biotechnol 2012;161:206–12.

660          https://doi.org/10.1016/j.jbiotec.2012.05.023.

661    [45]    Wilkins MR, Gasteiger E, Bairoch A, Sanchez J-C, Williams KL, Appel RD, et al.

662          Protein identification and analysis tools in the ExPASy Server. 2-D Proteome Anal.

663          Protoc., New Jersey: Humana Press; n.d., p. 531–52. https://doi.org/10.1385/1-59259-

664          584-7:531.

665    [46]    Fumagalli M, Ouhab D, Boisseau SM, Heux L. Versatile gas-phase reactions for

666      surface to bulk esterification of cellulose microfibrils aerogels. Biomacromolecules

667      2013;14:3246–55. https://doi.org/10.1021/bm400864z.

668    [47]   Wu H, Montanier CY, Dumon C. Quantifying CBM carbohydrate interactions using

669      microscale thermophoresis, 2017, p. 129–41. https://doi.org/10.1007/978-1-4939-6899-

670      2_10.

671    [48]   Kaufmann M, Klinger C, editors. Functional Genomics. vol. 815. New York, NY:

672      Springer New York; 2012. https://doi.org/10.1007/978-1-61779-424-7.

673    [49]   Millard P, Lippens G. Interact: Automated analysis of protein-ligand interactions by

674      1D and 2D NMR. BioRxiv 2018:215194. https://doi.org/10.1101/215194.

675

676

677

678    **Legends to Figures and Tables**

679    **Figure 1**: (A) Representation of the proteins studied in this work. The multimodular enzymes

680    are linked *via* an isopeptide bond between the Lys191 of Jo and the Asp695 of In (double fine

681    black line). Although Jo and In are fused at the *N*-termini of *Np*Xyn11A and the CBMs

682    respectively, the covalent Jo-In or In-Jo association between the GH and the CBM is shown in

683    square brackets for simplicity. (B) SDS-PAGE of the protein used in this study. Lanes: M,

684    molecular markers; 1, *Np*Xyn11A; 2, Jo-*Np*Xyn11A; 3, *Ct*CBM3a; 4, Jo-CBM3a; 5,

685    *Np*Xyn11A-[In-Jo]-CBM3a; 6, *Cf*CBM2b-1; 7, Jo-CBM2b-1; 8, In-CBM2b-1; 9, *Np*Xyn11A-

686    [In-Jo]-CBM2b-1; 10, *Np*Xyn11A-[Jo-In]-CBM2b-1. For illustration 11, *Np*Xyn11A-[Jo-In]-

687    CBM3a; 12, In-CBM3a. Original gels are provided in Supplementary Information (Figure

688    S5).

689

690    **Figure 2**. Degradation of complex substrates by *Np*Xyn11A and derivatives thereof. (A)

691    Wheat bran. (B) Wheat straw. Enzyme reactions were conducted in 50 mM sodium

692    phosphate, 12 mM sodium citrate pH6, supplemented with 1 mg/ml BSA, at 37°C. Enzyme

693    loading was at 1 µM. Substrate concentration was at 2% w/v.

694

695    **Figure 3**: (A) SAXS data recorded for *Np*Xyn11A, In-*Np*Xyn11A, In-CBM2b-1 and Jo-

696    CBM2b-1. (B) SAXS data recorded for *Np*Xyn11A-[In-Jo]-CBM2b-1, *Np*Xyn11A-[Jo-In]-

697    CBM2b-1,   *Np*Xyn11A-[In-Jo]-CBM3a   and   *Np*Xyn11A-[Jo-In]-CBM3a.   (C)   Table

698    summarizing all parameters extracted from the Guinier plot and pair-atom distribution

699    function P(r) such as gyration radius Rg and maximum internal distance Dmax. The folding

700    state deduced from the parameters and the shape of the curves are also mentioned in the table

701    for each fragment.

702

703    **Figure 4**: 1D proton spectrum of (A) isolated *Cf*CBM2b-1 (bottom, black), of In-CBMb2-1

704    (middle, red) and the difference spectrum (*Cf*CBM2b-1 – In-CBMb2-1) (top, green). (B)

705    Isolated *Cf*CBM2b-1 (bottom, black), of Jo-CBMb2-1 (middle, red) and the difference

706    spectrum (*Cf*CBM2b-1 – Jo-CBMb2-1) (top, orange).

707

708    **Figure 5**: (A) Pair distribution function calculated from SAXS data of *Np*Xyn11A-[In-Jo]-

709    CBM2b-1 and *Np*Xyn11A-[Jo-In]-CBM2b-1 (blue and red curves respectively) and from

710    SAXS data of *Np*Xyn11A-[In-Jo]-CBM3a and *Np*Xyn11A-[Jo-In]-CBM3a (green and yellow

711    curves respectively). The P(r) function is plotted as P(r)/Pmax(r) *vs* distance r in order to

712    compare the different curves by normalizing with Pmax(r). (B) Low resolution shape

713    calculated with the DAMMIF program from ATSAS suite for the four constructs. The shapes

714    are built with Pymol in mesh representation and filled with transparent spheres.

715

**Figure 6**: (A) Comparison of experimental curves generated using *Np*Xyn11A-[In-Jo]-CBM2b-1 and *Np*Xyn11A-[Jo-In]-InCBM2b-1 and the theoretical curves (dotted black line) calculated using CRYSOL and crystallographic data. (B) Models of *Np*Xyn11A-[In-Jo]-CBM2b-1 and (C) *Np*Xyn11A-[Jo-In]-InCBM2b-1 were calculated using the CORAL program. In both models, Jo-In are coloured red and green respectively. *Np*Xyn11A is in cyan and *Ct*CBM2b-1in yellow. Catalytic residues of *Np*Xyn11A and residues involved in ligand binding recognition of *Ct*CBM2b-1 are represented by red lines. The domains and the linker are modelled with Pymol using cartoon and grey sphere representations respectively.

724

**Figure 7**: Double immunofluorescence of paraffin-embedded wheat bran serial sections showing the impact of enzymatic on-section treatment on accessible xylan and cellulose immunolabelling. Serial sections of wheat bran were incubated for 24 h using the xylanase derivatives as labelled on the images and further used for double indirect immunofluorescence using LM11 (xylan specific antibody) and $His_6$-CBM3a (cellulose specific CBM) as primary probes and anti-rat IgG-Alexa 488 and anti-$His_6$-Alexa 555, respectively. The individual fluorescence channels are shown in the two first rows as labelled, and the merge of both fluorescence channels with the bright field channel is shown is the third row. Arrowheads: pericarp/nucellar epidermis interface; Bars: 300 µm.

734

**Figure 8**: Double immunofluorescence of paraffin-embedded wheat straw serial sections showing the impact of enzymatic on-section treatment on xylan and cellulose immunolabelling. Serial sections of wheat straw were incubated for 24 h using the recombinant xylanase derivatives as labelled on the images and further used for double indirect immunofluorescence using LM11 (xylan specific antibody) and $His_6$-CBM3a

740    (cellulose specific CBM) as primary probes and anti-rat IgG-Alexa 488 and anti-His$_6$-Alexa

741    555, respectively. The individual fluorescence channels are shown as labelled for a wide field

742    view (two left rows) and for a vascular bundle magnified view (two right rows). Note, that to

743    enable fair comparison between the treatment/controls, the same zones on the different serial

744    sections are displayed. Note also that an intermediate magnification is shown for this vascular

745    bundle in the ROI of Supplementary Figure S7, as well as two additional ROIs. bsf, bundle

746    sheath fibres; cp, cortical parenchyma zones; ep, epidermis; if, intervascular fibres; mx,

747    metaxylem; p, pith; ph, phloem; pp, pith parenchyma; px, protoxylem; s, sclerenchyma; vb,

748    vascular bundle. Bars: 250 µm (two left rows); 25 µm (two right rows).

749

750

751

752    **Table 1**: Binding affinity of *Ct*CBM3a and derivates against cellulose nanocrystals. Buffer 1:

753    50 mM Tris HCl pH 7.4, 150 mM NaCl, 10 mM MgCl$_2$, 0.05 % Tween 20. Buffer 2: 50 mM

754    sodium phosphate buffer, pH 7 and 0.05 % pluronic acid. See Supplementary Figure S3 for

755    chart. EC$_{50}$ is the half-maximal effective concentration, i.e. the higher the affinity for the

756    substrate, the smaller the value of the EC$_{50}$ .

757

758    **Table 2**:  Ligand affinities of *Cf*CBM2b-1, In-CBM2b-1 and Jo-CBM2b-1 for xylohexaose as

759    measured by 1D NMR by titrating the resonances of the NH$\varepsilon$ of Trp 259 and Trp 291.

760    Experiments were conducted in 50 mM sodium phosphate pH 7 at 298 °K.

761
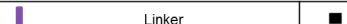
762    **Table 3**: Specific activity of *Np*Xyn11A and derivatives, as single enzymes or in complex.

763    Reactions were performed in 50 mM sodium phosphate, 12 mM sodium citrate pH 6,

764    supplemented with 1mg/mL BSA, at 37°C. Substrate concentrations were 5 mM    4-

765    nitrophenyl-β-D-xylotrioside ($p$NP-X$_3$), 0.5% w/v wheat arabinoxylan (WAX), 0.5% w/v rye

766    arabinoxylan (RAX) and 1% w/v beechwood xylan (BWX). The values are shown as means ±

767    standard deviation of replicate n = 3.

768

**A**

| | Protein name | Protein scheme | Position in Fig.2B |
|---|---|---|---|
| Single proteins | *Np*Xyn11A | N [*Np*Xyn11A] C | 1 |
| | *Ct*CBM3a | N [CBM3a] C | 3 |
| | *Ct*CBM2b-1 | N [CBM2b-1] C | 6 |
| Jo/In fusion proteins | In-*Np*Xyn11A | N [In][*Np*Xyn11A] C | - |
| | Jo-*Np*Xyn11A | N [Jo][*Np*Xyn11A] C | 2 |
| | In-CBM3a | N [In]■[CBM3a] C | 12 |
| | Jo-CBM3a | N [Jo][CBM3a] C | 4 |
| | In-CBM2b-1 | N [In]■[CBM2b-1] C | 8 |
| | Jo-CBM2b-1 | N [Jo][CBM2b-1] C | 7 |
| | His-Tag | Linker ■ | |
| Multi-modular enzymes | *Np*Xyn11A-[In-Jo]-*CBM3a* | C [CBM3a][Jo] N [In][*Np*Xyn11A] C | 5 |
| | *Np*Xyn11A-[Jo-In]-CBM3a | C [CBM3a]■[In] N [Jo][*Np*Xyn11A] C | 11 |
| | *Np*Xyn11A-[In-Jo]-CBM2b-1 | C [CBM2b-1][Jo] N [In][*Np*Xyn11A] C | 9 |
| | *Np*Xyn11A-[Jo-In]-CBM2b-1 | C [CBM2b-1]■[In] N [Jo][*Np*Xyn11A] C | 10 |

**B**
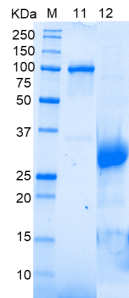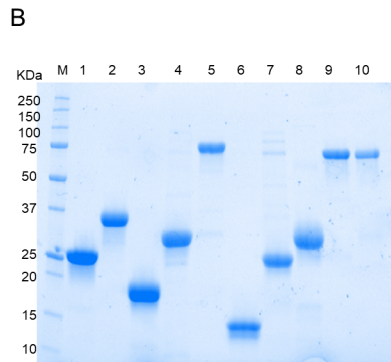
A
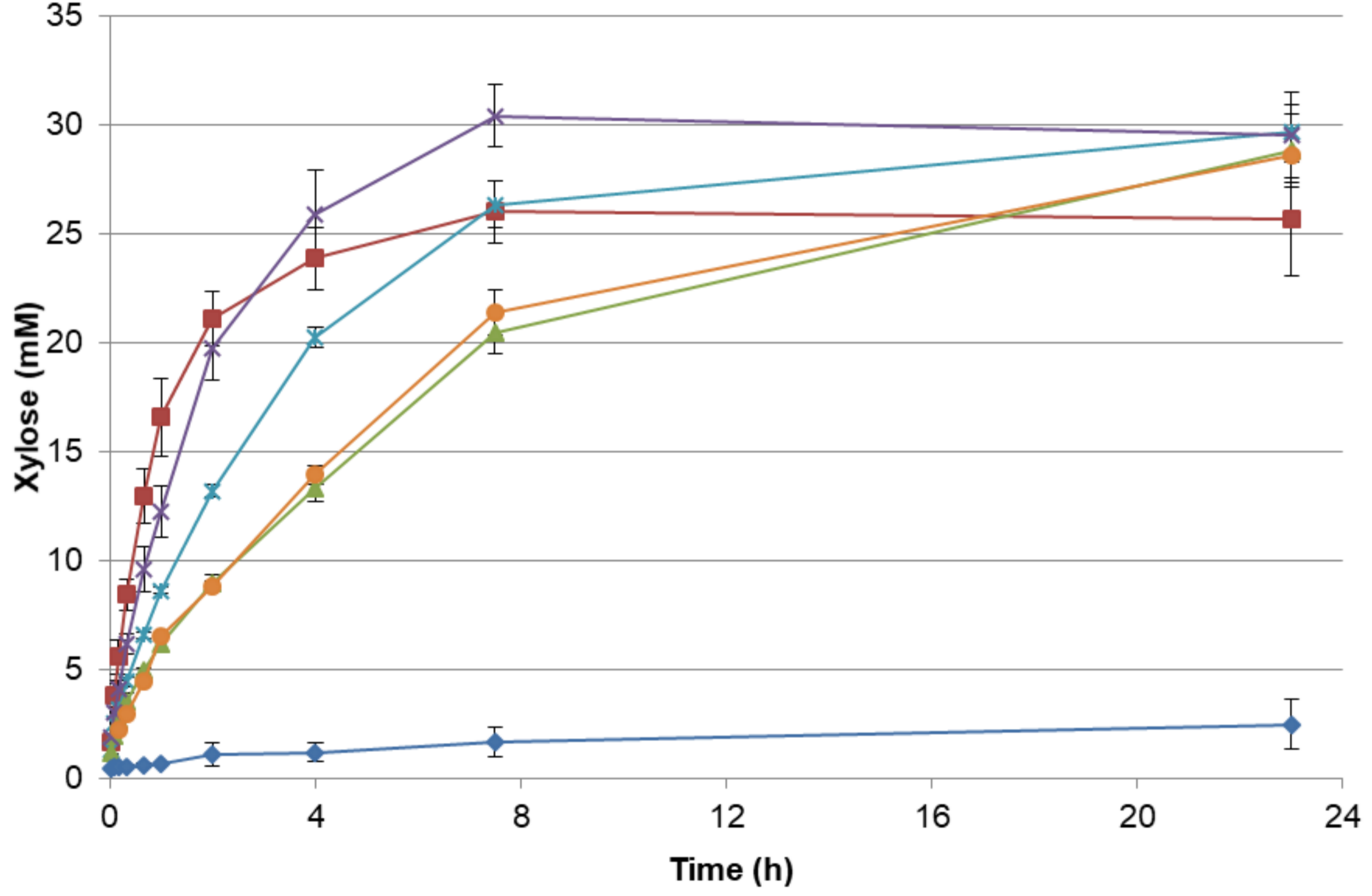
B
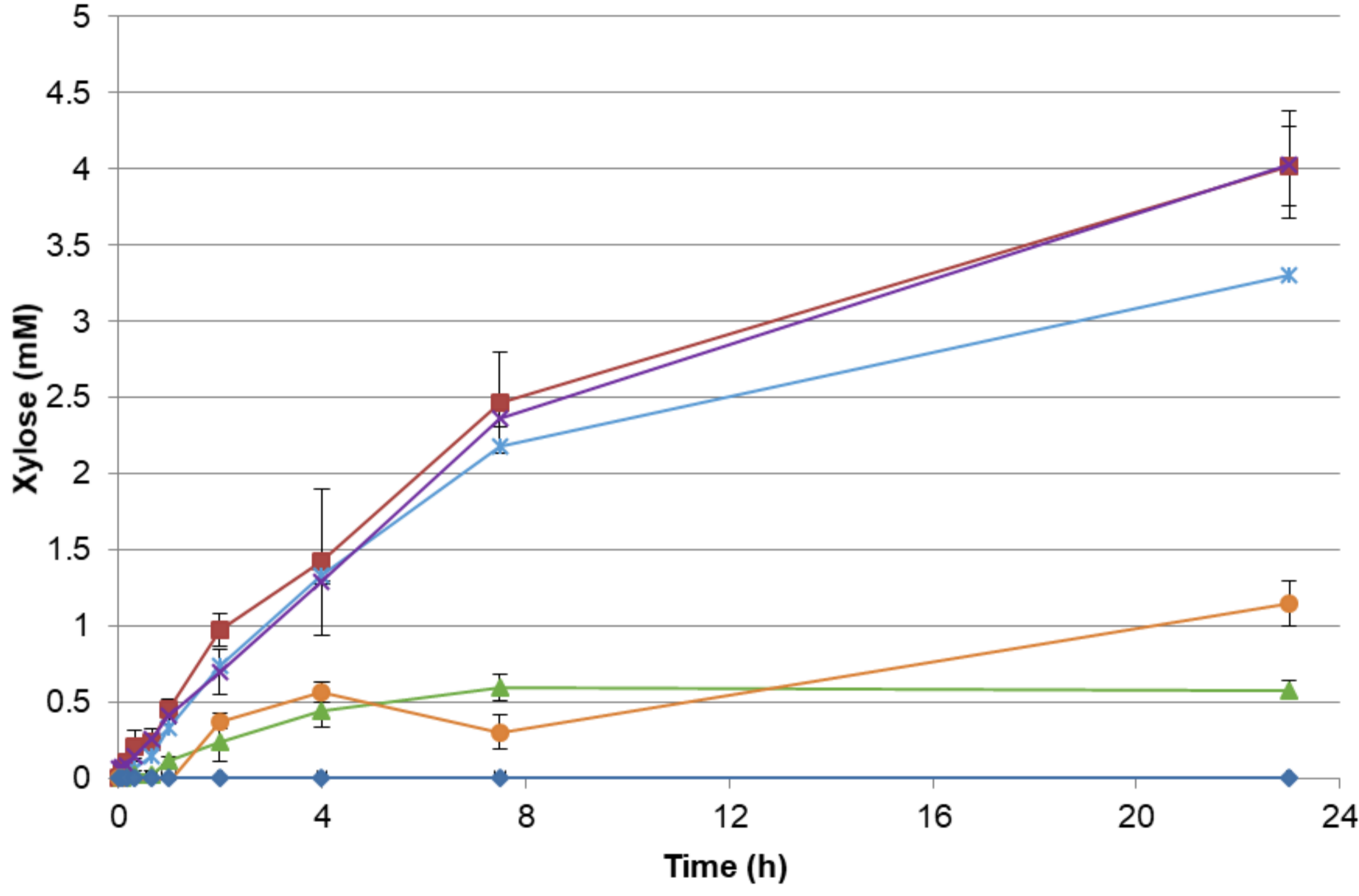
■ *Np*Xyn11A    ✳ *Np*Xyn11A-[ln-Jo]-CBM2b-1    ● *Np*Xyn11A-[Jo-ln]-CBM3a    ✳ *Np*Xyn11A-[Jo-ln]-CBM2b-1    ▲ *Np*Xyn11A-[ln-Jo]-*CBM3a*    ◆ control
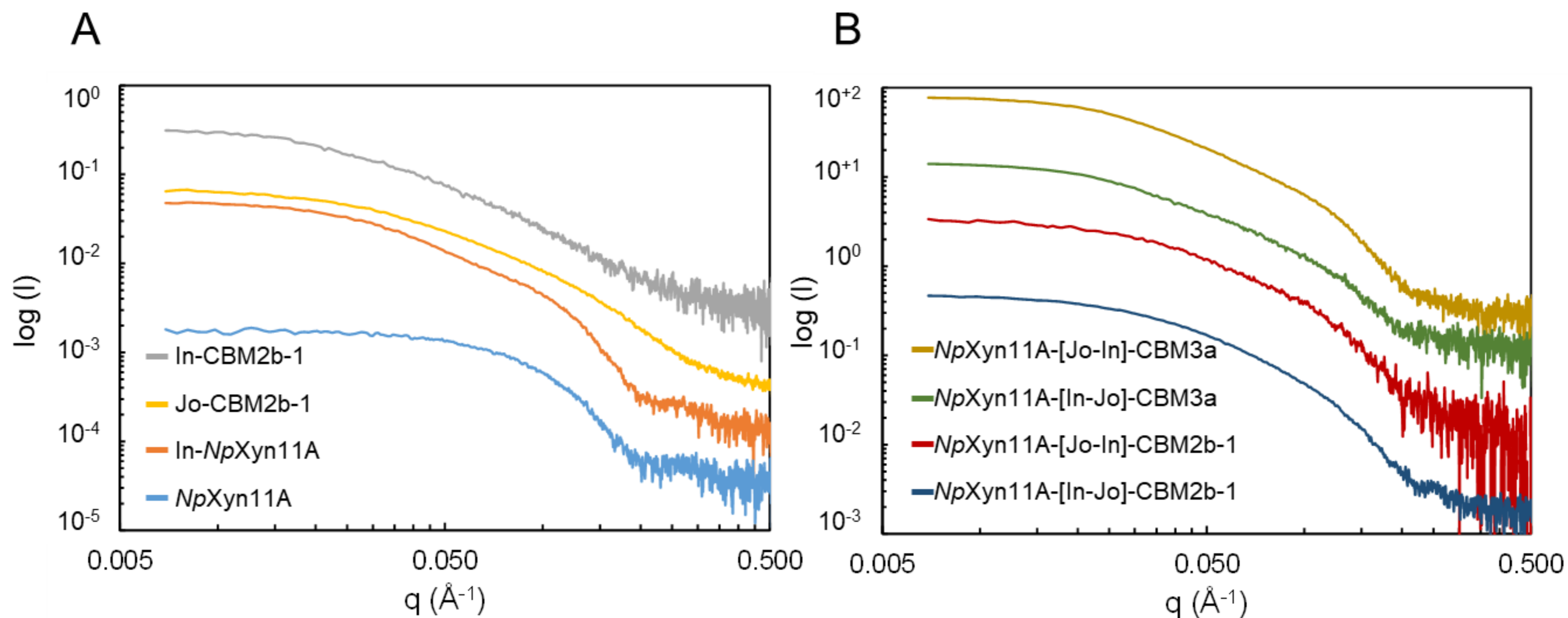
| Sample | Molecular Mass (Kda) | Gyration radius Rg (Å) | Maximal distance D (Å) | Folding |
|---|---|---|---|---|
| *Np*Xyn11A-[Jo-In]-CBM3a | 71.7 | 47.8 | 185 | Folded |
| *Np*Xyn11A-[In-Jo]-CBM3a | 71.0 | 46.5 | 180 | Folded |
| *Np*Xyn11A-[Jo-In]-CBM2b-1 | 63.4 | 42.1 | 160 | Folded |
| *Np*Xyn11A-[In-Jo]-CBM2b-1 | 62.7 | 41.8 | 150 | Folded |
| Jo-CBM2b-1 | 21.1 | 46.1 | 190 | Unfolded |
| In-CBM2b-1 | 27.0 | 58.0 | 210 | Unfolded |
| In-*Np*Xyn11A | 41.5 | 46.2 | 160 | Partially Folded |
| *Np*Xyn11A | 25.9 | 18.5 | 60 | Folded |

**A**

Trp side chain H<sub>N</sub>   H₂O   Aliphatic protons

10  9  8  7  6  5  4  3  2  1  0  ppm

Amide protons                Methyl protons

**B**

Trp side chain H<sub>N</sub>   H₂O   Aliphatic protons

10  9  8  7  6  5  4  3  2  1  0  ppm

Amide protons                Methyl protons

**A**

*Np*Xyn11A-[In-Jo]-CBM2b-1
*Np*Xyn11A-[Jo-In]-CBM2b-1

*Np*Xyn11A-[In-Jo]-CBM3a
*Np*Xyn11A-[Jo-In]-CBM3a

**B**

*Np*Xyn11A-[In-Jo]-CBM2b-1

*Np*Xyn11A-[Jo-In]-CBM2b-1

*Np*Xyn11A    In-Jo    CBM

*Np*Xyn11A-[In-Jo]-CBM3a

*Np*Xyn11A-[Jo-In]-CBM3a

200 Å

**A**

*Np*Xyn11A-[In-Jo]-CBM2b-1

Chi²=1.25

*Np*Xyn11A-[Jo-In]-CBM2b-1

Chi²=1.57

log (I)

q(Å⁻¹)

**B** *Np*Xyn11A-[In-Jo]-CBM2b-1

**C** *Np*Xyn11A-[Jo-In]-CBM2b-1

|  | LM11-A488 | CBM3a-A555 | Merge |
|---|---|---|---|
| A | A1 autofluo | A2 | A3 |
| B | B1 buffer control | B2 | B3 |
| C | C1 *Np*Xyn11A | C2 | C3 |
| D | D1 *Np*Xyn11A-[Jo-In]-CBM2b-1 | D2 | D3 |
| E | E1 *Np*Xyn11A-[In-Jo]-CBM2b-1 | E2 | E3 |
| F | F1 *Np*Xyn11A-[In-Jo]-CBM3a | F2 | F3 |

|  | LM11-A488 | CBM3a-A555 | LM11-A488 | CBM3a-A555 |
|---|---|---|---|---|
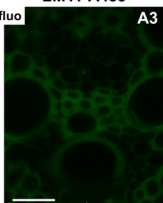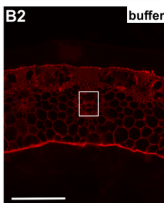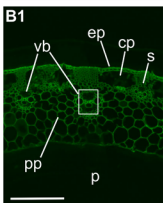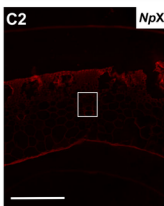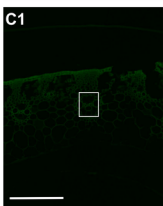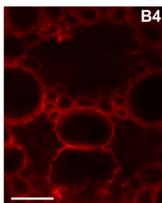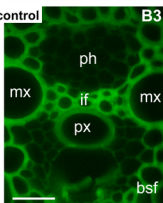
**A1**     **A2** *autofluo*     **A3**     **A4**

**B1**     **B2** *buffer control*     **B3**     **B4**

vb, ep, cp, s, pp, p (B1); mx, ph, if, px, mx, bsf (B3)

**C1**     **C2** *Np*Xyn11A     **C3**     **C4**

**D1**     **D2** *Np*Xyn11A-[Jo-In]-CBM2b-1     **D3**     **D4**
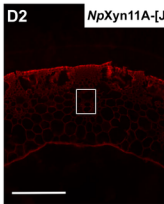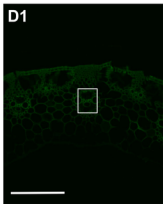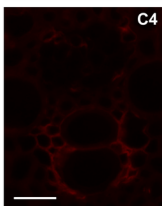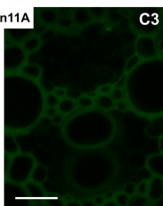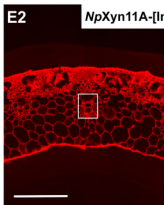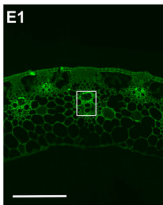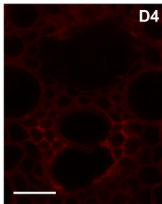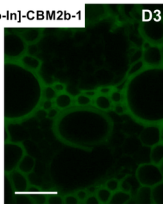
**E1**     **E2** *Np*Xyn11A-[In-Jo]-CBM2b-1     **E3**     **E4**
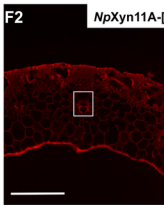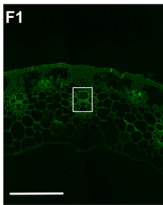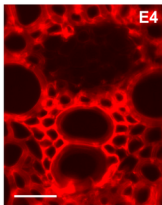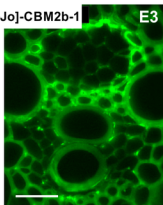
**F1**     **F2** *Np*Xyn11A-[In-Jo]-CBM3a     **F3**     **F4**
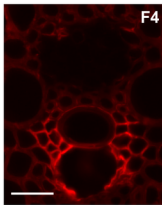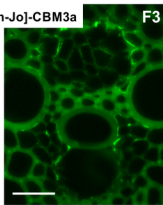
**Table 1**:

| Protein | EC$_{50}$ (g.L$^{-1}$) | |
| --- | --- | --- |
| | *Buffer 1* | *Buffer 2* |
| In-CBM3a | - | 0.003 ± 0.0001 |
| *Ct*CBM3a | 0.14 ± 0.01 | 0.18 ± 0.04 |
| Jo-CBM3a | 0.21 ± 0.12 | - |

**Table 2**:

| Protein | $K_d$ (mM) | | |
|---------|------------------------|------------------------|------|
|         | This work (Trp 259) | This work (Trp 291) | [35] |
| *Cf*CBM2b-1 | 0.84 | 1.02 | 0.29 |
| In-CBM2b-1 | 0.22 | 0.52 | - |
| Jo-CBM2b-1 | > 10 | > 10 | - |

**Table 3**:

| Single enzymes | Specific Activity | | | |
| --- | --- | --- | --- | --- |
| | $p$**NP-X$_3$** (IU/µmole) | **WAX** $10^3$ (IU/µmole) | **RAX** $10^3$ (IU/µmole) | **BWX** $10^3$ (IU/µmole) |
| *Np*Xyn11A | 155.78 ± 4.07 | 114.59 ± 2.67 | 83.44 ± 8.21 | 65.03 ± 2.19 |
| In-*Np*Xyn11A | 153.17 ± 1.68 | - | - | - |
| Jo-*Np*Xyn11A | 124.63 ± 3.64 | - | - | - |
| *Np*Xyn11A-[In-Jo]-CBM3a | 148.83 ± 14.19 | 87.29 ± 6.32 | 81.58 ± 3.03 | 53.30 ± 0.70 |
| *Np*Xyn11A-[Jo-In]-CBM3a | 115.88 ± 25.78 | 160.72 ± 6.34 | 196.17 ± 0.60 | 93.40 ± 1.34 |
| *Np*Xyn11A-[In-Jo]-CBM2b-1 | 98.87 ± 4.01 | 62.81 ± 8.28 | 28.57 ± 3.03 | 39.50 ± 0.61 |
| *Np*Xyn11A-[Jo-In]-CBM2b-1 | 153.01 ± 5.30 | 126.91 ± 9.05 | 93.86 ± 9.59 | 58.30 ± 2.29 |

① Jo-In welding system = chimeric multi-modular enzyme
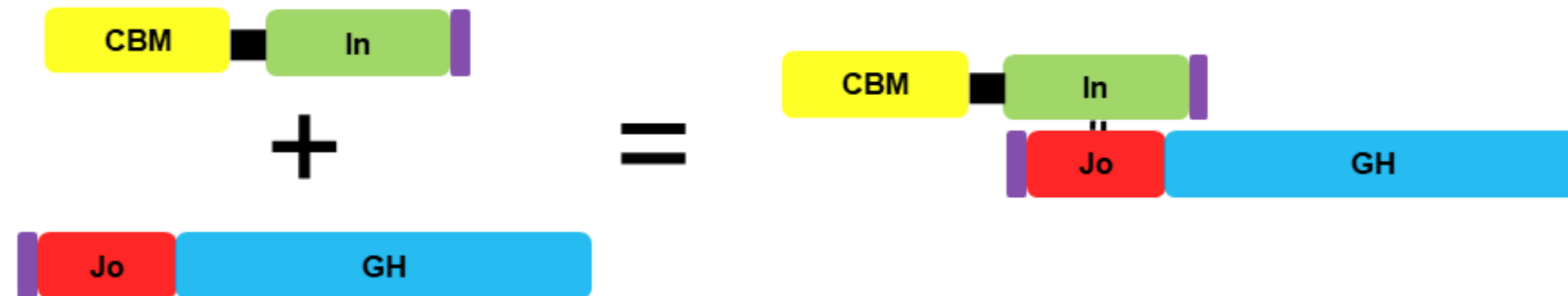
② Spatial orientation relatively locked

*Np*Xyn11A-[In-Jo]-CBM2b-1

*Np*Xyn11A-[Jo-In]-CBM2b-1
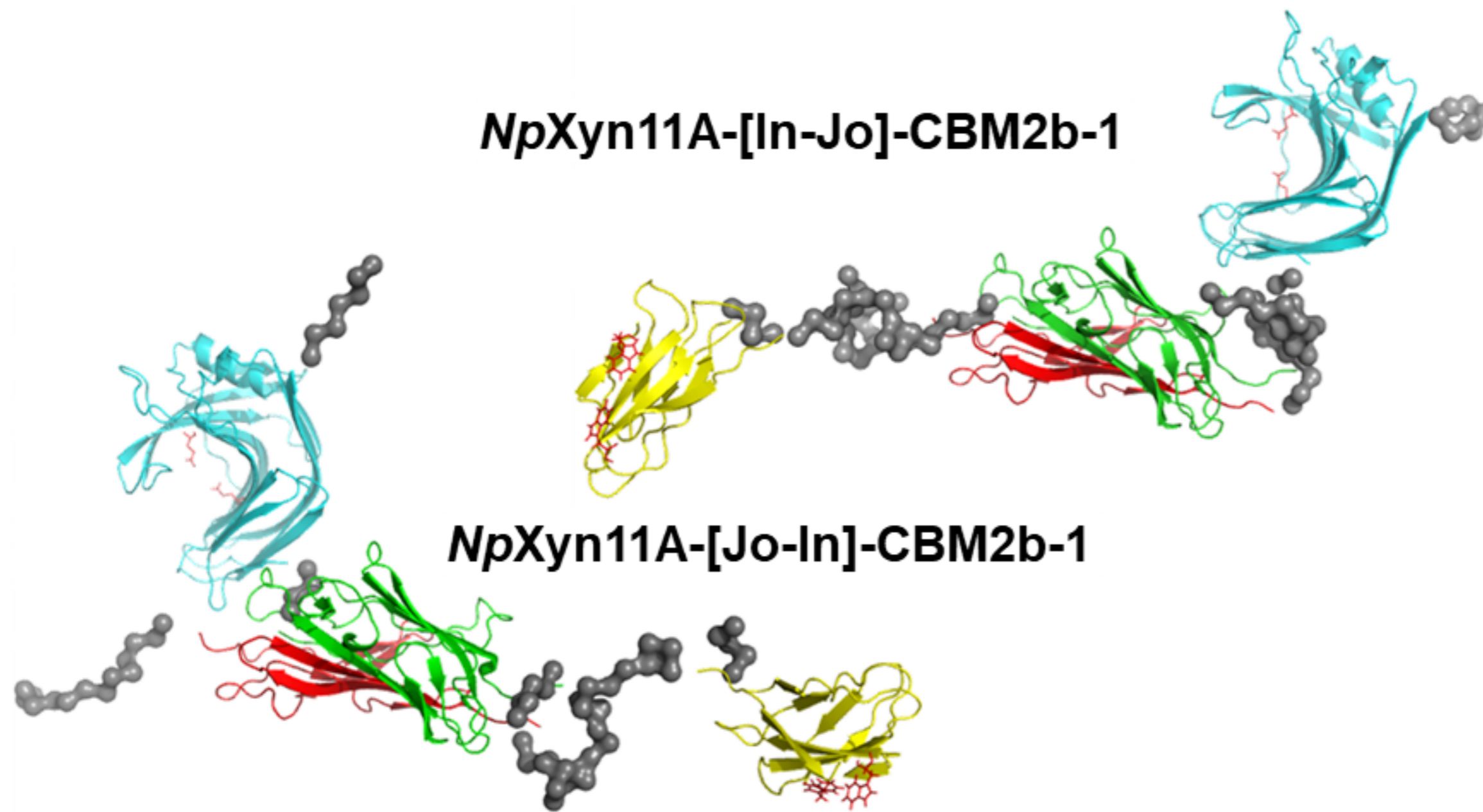
③ Optimal configuration in plant cell wall degradation

*Np*Xyn11A-[Jo-In]-CBM2b-1     D4

*Np*Xyn11A-[In-Jo]-CBM2b-1     E4