

# **A brief and Informationally Rich Naming System for Oligosaccharide**

## **Motifs of Heteroxylans Found in Plant Cell Walls**

*Régis Fauré,<sup>A,B,C</sup> Christophe M. Courtin,<sup>D</sup> Jan A. Delcour,<sup>D</sup> Claire Dumon,<sup>A,B,C</sup> Craig B. Faulds,<sup>E</sup> Geoffrey B. Fincher,<sup>F</sup> Sébastien Fort,<sup>G</sup> Stephen C. Fry,<sup>H</sup> Sami Halila,<sup>G</sup> Mirjam A. Kabel,<sup>I,†</sup> Laurice Pouvreau,<sup>I</sup> Bernard Quemener,<sup>J</sup> Alain Rivet,<sup>G</sup> Luc Saulnier,<sup>J</sup> Henk A. Schols,<sup>I</sup> Hugues Driguez,<sup>G,K</sup> and Michael J. O'Donohue<sup>A,B,C,K</sup>*

<sup>A</sup>Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France.

<sup>B</sup>INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France.

<sup>C</sup>CNRS, UMR5504, F-31400 Toulouse, France.

<sup>D</sup>Laboratory of Food Chemistry and Biochemistry, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20 bus 2463, B-3001, Leuven, Belgium.

<sup>E</sup>Institute of Food Research, Sustainability of the Food Chain Exploitation Platform, Norwich Research Park, Norwich NR4 7UA, UK.

<sup>F</sup>Australian Centre for Plant Functional Genomics, School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia.

<sup>G</sup>Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS),<sup>‡</sup> B.P. 53, F-38041, Grenoble cedex 9, France.

<sup>H</sup>The Edinburgh Cell Wall Group, Institute of Molecular Plant Sciences, School of Biological Sciences, The University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JH, UK.

<sup>I</sup>Wageningen University, Laboratory of Food Chemistry, Bomenweg 2, 6703HD Wageningen, The Netherlands

<sup>J</sup>UR1268 Biopolymères Interactions Assemblages, INRA, F-44300 Nantes, France.

<sup>K</sup>Corresponding authors. Email: [hugues.driguez@cermav.cnrs.fr](mailto:hugues.driguez@cermav.cnrs.fr); [michael.odonohue@insa-toulouse.fr](mailto:michael.odonohue@insa-toulouse.fr)

This manuscript is dedicated to Robert V. Stick, Professor of The University of Western Australia (WA, Australia), for his contribution in Glycochemistry.

**Running title:** Short names for heteroxylans

**Footnotes to title page**

<sup>†</sup>Present address: Royal Nedalco, PO Box 6, 4600 AA Bergen op Zoom, The Netherlands.

<sup>‡</sup>Affiliated with Joseph Fourier University and member of the Institut de Chimie Moléculaire de Grenoble FR-CNRS 2607.

## **Abstract**

The one-letter code system proposed here is a simple method to accurately describe structurally diverse oligosaccharides derived from heteroxylans. Substitutions or ‘molecular decoration(s)’ of main-chain D-xylosyl moieties are designated by unique letters. Hence, an oligosaccharide is described by a series of single letters, beginning with the non-reducing D-xylosyl unit. Superscripted numbers are used to indicate the linkage position(s) of main-chain substitution(s) and, where necessary, superscripted lowercase letter(s) indicate the nature of non-glycosidic groups (*e.g.* methyl, acetyl or phenolic derivative moieties) that can be present on the substituents. Although relatively simple and practical to use, this abbreviated system lends itself to the naming of a large number of different combinations of structural building blocks and substituents. In its present state, this system is therefore adequate to name and differentiate all currently known complex oligosaccharides derived from heteroxylans and is sufficiently flexible to accommodate new structures as they become available.

**Keywords:** heteroxylans, arabinoxylans, oligosaccharides, AXOS

Difficulties associated with the naming of complex oligosaccharide motifs that are derived from heteroxylans often results in the use of erroneous or highly abbreviated, informationally-poor terms in the text of articles. Here, we propose a new convenient method for generating useful short names based on a one-letter code that should help glyco-chemists and biochemists to more easily convey both compositional and structural information concerning heteroxylan motifs.

Heteroxylans, especially arabinoxylans (AXs), are chemically and structurally complex polysaccharides that form up to 35% of the dry weight of hardwoods and cereals.<sup>[1]</sup> Although the main-chains of heteroxylans are almost exclusively consist of D-xylopyranosyl residues linked through  $\beta$ -(1 $\rightarrow$ 4) glycosidic linkages, they also carry a wide diversity of main-chain substituents. The most frequently reported type of main-chain substituents are L-arabinofuranosyl residues that are linked to the main-chain xylosyl units through *O*-2 and/or *O*-3 bonds. Similarly, other substituting groups that can be linked through *O*-2 and/or *O*-3 bonds are D-xylosyl and D- and L-galactopyranosyl residues, uronic acids (notably D-glucuronic acid and its 4-methylated derivative), and acetyl moieties. In AXs, further chemical complexity is introduced through the substitution of the substituents themselves. Typical examples of this are phenolic acid derivatives that are linked to L-arabinofuranosyl residues. The two major types of phenolic substituents, (*E*)-4-hydroxy-3-methoxycinnamoyl (*trans*-feruloyl) and (*E*)-4-hydroxy-cinnamoyl (*trans*-coumaroyl), form ester bonds with the *O*-5 primary alcohol group of L-arabinofuranosyl units.

Because of the high degree of chemical diversity of heteroxylans and heteroxylan-derived oligosaccharides, the naming of these compounds is a difficult task for most researchers. Indeed, the frequent use in the literature of erroneous or approximate chemical names provides ample proof for this statement. The only established rules that do exist are those laid down by IUPAC-IUBMB.<sup>[2]</sup> These rules generate complex names that are unfamiliar to many researchers and are often quite difficult to understand. Thus, clearly a simpler, more useful, comprehensive naming system would be welcome.

Fifteen years ago, Fry *et al.* (1993)<sup>[3]</sup> proposed a simplified nomenclature for xyloglucan-derived oligosaccharides and more recently an easy-to-use structural code system has been developed for glycosaminoglycans.<sup>[4]</sup> The system proposed by Fry *et al.* (1993)<sup>[3]</sup> attributes a unique letter to each main-chain segment according to the substituent that it bears. With this simple system it has been possible to name a variety

of molecules, including a recently prepared library of xylogluco-oligosaccharides.<sup>[5,6]</sup> Inspired by these initiatives, we propose a similar strategy to name oligosaccharides derived from heteroxylyans. Likewise, we believe that the use of a single letter code system makes it possible to create simpler abbreviations that will improve clarity and accuracy, while faithfully reflecting the vast diversity of heteroxylyan structures, notably those of arabinoxylyan oligosaccharides (AXOS).

The key to our proposed short name system is summarized in Tables 1 and 2. Similar to the system described by Fry *et al.* (1993),<sup>[3]</sup> we have adopted a single letter code, where uppercase letters are unique identifiers of main-chain substituents. To form a name, these are arranged in a consecutive manner starting from the non-reducing terminal residue of the main-chain. The letter 'X' is attributed to unsubstituted or terminal  $\beta$ -(1 $\rightarrow$ 4)-linked, D-xylosyl residues. Other letters describe glycosyl units that decorate the xylan main-chain. The majority of these letters have been used to describe single glycosyl units (*e.g.* 'A' for L-arabinosyl), but oligosaccharide motifs that contain an arabinosyl unit at the reducing end of the sidechain have also been attributed letters that follow on from 'A' in alphabetical order (*i.e.* arabinobiosyl is 'B' and so forth). So far, we have introduced 10 uppercase letters. A second level of information is provided by one (or several) superscript number(s) that indicate(s) the position(s) of the substitution(s). Punctuated by 'commas', these are written in a directional manner beginning with the non-reducing terminal glycosyl unit of the side-chain. The last number in the sequence indicates the position on the main-chain D-xylosyl unit that is decorated. Accordingly, A<sup>2</sup> describes a L-arabinosyl residue that substitutes a main-chain D-xylosyl unit at its O-2 position, the presence of this latter being considered implicit. With regard to other punctuations, the 'plus' symbol is used when a main-chain D-xylosyl residue is double substituted. For example, A<sup>2+3</sup> is used to indicate the presence of two side-chain L-arabinosyl groups that are linked to the same main-chain D-xylosyl residue at O-2 and O-3 respectively. Building on this simple code system, further side-chain complexity is handled through the use of superscripted letters, which can take two forms. First, to account for the fact that glycosyl units that form a side-chain decoration can themselves be substituted, superscript lowercase letters are used to designate non-glycosidic substituents. These are preceded by a number that indicates the position of the hydroxyl group engaged in the substitution. For example, A<sup>5f2</sup> describes a main-chain D-xylosyl unit that is substituted at O-2 by an L-arabinosyl residue that is itself substituted at O-5 by a *trans*-feruloyl moiety. Second, to describe more complex, relatively uncommon side-chains that cannot be defined by a single letter, a composite letter is formed from single letters that are organized in the following manner. As before, the reducing terminal residue (or a defined oligosaccharide motif) is

represented by its assigned letter, and then further side-chain glycosyl moieties are indicated by their letter, which is written in superscript and associated with a superscript number. Therefore,  $D^{M3,L2,2,5f3}$  describes a main-chain D-xylosyl unit that is substituted at *O*-3 by a  $\alpha$ -D-Galp-(1 $\rightarrow$ 3)- $\alpha$ -L-Galp-(1 $\rightarrow$ 2)- $\beta$ -D-Xylp-(1 $\rightarrow$ 2)-[5-*O*-Feruloyl]- $\alpha$ -L-Araf motif, where D designates the  $\beta$ -D-Xylp-(1 $\rightarrow$ ?) - $\alpha$ -L-Araf motif and L and M designate the  $\alpha$ -L-Galp and  $\alpha$ -D-Galp units respectively. Overall, by combining the single letter codes with the superscript numbers, uppercase and lowercase letters and punctuation, a large number of permutations can be generated to describe variously substituted  $\beta$ -D-xylosyl residues (and the reducing terminal D-xylose unit) in the xylan backbone of heteroxylans. Finally, it is noteworthy that the structural code system can be applied to describe simple motifs even when the oligosaccharide structures are only partially determined (*i.e.* the positions of the linkages are not defined). In this case, uncertainty associated with linkage position can be indicated by the use of one or more superscript ‘0’ characters.

To provide a more complete guide to how the system works, the short names of seven representative AX-derived structures are shown in Table 3 and Fig. 1. The benefits of the new system can be clearly illustrated by considering some of these structures. Firstly, in the past the feruloylated trisaccharide 5-*O*-feruloyl- $\alpha$ -L-Araf-(1 $\rightarrow$ 3)- $\beta$ -D-Xylp-(1 $\rightarrow$ 4)-D-Xylp has been variously named in the literature as Fe5Ara $\alpha$ 3Xyl $\beta$ 4Xyl,<sup>[2]</sup> *O*-[5-*O*-(*trans*-feruloyl)- $\alpha$ -L-Araf]-(1 $\rightarrow$ 3)-*O*- $\beta$ -D-Xylp-(1 $\rightarrow$ 4)-D-Xylp or simply FAXX,<sup>[7]</sup> terms that are either highly complex or uninformative. Using the new system, the same trisaccharide can be designated as A<sup>5f3</sup>X, which is a very simple but totally unambiguous name. Secondly, related arabinoxylan-octasaccharides such as Xyl $\beta$ 4[Ara $\alpha$ 3]Xyl $\beta$ 4[(Ara $\alpha$ 2)(Ara $\alpha$ 3)]Xyl $\beta$ 4Xyl $\beta$ 4Xyl, Xyl $\beta$ 4[(Ara $\alpha$ 2)(Ara $\alpha$ 3)]Xyl $\beta$ 4[Ara $\alpha$ 3]Xyl $\beta$ 4Xyl $\beta$ 4Xyl, and Xyl $\beta$ 4[Ara $\alpha$ 3]Xyl $\beta$ 4Xyl $\beta$ 4[Ara $\alpha$ 3]Xyl $\beta$ 4Xyl $\beta$ 4Xyl (shown in Fig. 1) can be attributed the simple, visually unambiguous designations XA<sup>3</sup>A<sup>2+3</sup>XX, XA<sup>2+3</sup>A<sup>3</sup>XX, and XA<sup>3</sup>XA<sup>3</sup>XX, respectively. Therefore, through the use of this new one-letter code system, regions of heteroxylans can be easily named in a uniform way that is accessible to both specialists and non-specialists working in a wide variety of scientific fields.

Although the abbreviations described above will generate quite simple names for almost all known structures, two noteworthy exceptions require extra punctuation. Asymmetric double substitutions of a D-xylosyl residue by glycosyl moieties can be dealt with using ‘curly brackets’ that are used to associate two code letters with a single main-chain D-xylosyl unit (Fig. 2), while multiple glycosidic substitutions of a side-chain decoration (*i.e.* branch-point within a side-chain) could be represented using the ‘slash’ symbol.

In conclusion, we propose a system that we believe is sufficiently simple to enable its use by non-specialists and robust enough to accommodate the introduction of new, complex structures. For these reasons, we hope that this system can be widely adopted by the scientific community and that it will have significant impact in reducing ambiguities linked to poor use or understanding of current nomenclature. It is clear that the system must be sufficiently flexible to accommodate new, hypothetical structures that might be reported in the future, and the authors envisage that the list of code uppercase letters, superscript numerals and lowercase letters could be easily extended, if and as required. However, in order to maintain the universal nature of the short name system, it is vital that colleagues wishing to modify it first consult with H. Driguez, M. O'Donohue and their co-authors before publishing any new codes. To further facilitate the use of the structural code system and its future administration, the creation of an interactive internet-based interface is underway (<http://www.cermav.cnrs.fr/AXonym>). This interface will provide up-to-date code tables, a brief tutorial to describe how the system works, and an automatic abbreviated name generator.

## References

- [1] L. Saulnier, C. Marot, E. Chanliaud, J.-F. Thibault, *Carbohydr. Polym.* **1995**, *26*, 279.
- [2] A. D. McNaught, *Pure Appl. Chem.* **1996**, *68*, 1919.
- [3] S. C. Fry, W. S. York, P. Albersheim, A. Darvill, T. Hayashi, J. P. Joseleau, Y. Kato, E. P. Lorences, G. A. Maclachlan, M. McNeil, A. J. Mort, G. J. S. Reid, H. U. Seitz, R. R. Selvendran, A. G. J. Voragen, A. R. White, *Physiol. Plant.* **1993**, *89*, 1.
- [4] R. Lawrence, H. Lu, R. D. Rosenberg, J. D. Esko, L. Zhang, *Nat. Methods* **2008**, *5*, 291.
- [5] R. Fauré, D. Cavalier, K. Keegstra, S. Cottaz, H. Driguez, *Eur. J. Org. Chem.* **2007**, 4313.
- [6] R. Fauré, M. Saura-Valls, H. Brumer III, A. Planas, S. Cottaz, H. Driguez, *J. Org. Chem.* **2006**, *71*, 5151.
- [7] L. Saulnier, J. Vigouroux, J.-F. Thibault, *Carbohydr. Res.* **1995**, *272*, 241.
- [8] J.-P. Utille, I. Jeacomine, *Carbohydr. Res.* **2007**, *342*, 2649.
- [9] H. Gruppen, R. A. Hoffmann, F. J. M. Kormelink, A. G. J. Voragen, J. P. Kamerling, J. F. G. Vliegthart, *Carbohydr. Res.* **1992**, *233*, 45.
- [10] G. Wende, S. C. Fry, *Phytochemistry* **1997**, *45*, 1123.
- [11] J. J. Ordaz-Ortiz, F. Guillon, O. Tranquet, G. Dervilly-Pinel, V. Tran, L. Saulnier, *Carbohydr. Polym.* **2004**, *57*, 425.

- [12] B. Quemener, J. J. Ordaz-Ortiz, L. Saulnier, *Carbohydr. Res.* **2006**, *341*, 1834.
- [13] A. A. Shatalov, D. V. Evtuguin, C. Pascoal Neto, *Carbohydr. Res.* **1999**, *320*, 93.
- [14] V. M. F. Goncalves, D. V. Evtuguin, M. R. M. Domingues, *Carbohydr. Res.* **2008**, *343*, 256.



CODE LETTERS		
Uppercase letter	Structures represented	Mnemonic
X	$\beta$ -D-Xylp	Xylosyl of the backbone
Y	$\beta$ -D-Xylp-(1 $\rightarrow$ ?)-	xYlosyl
A	$\alpha$ -L-Araf-(1 $\rightarrow$ ?)-	Arabinosyl
B	$\alpha$ -L-Araf-(1 $\rightarrow$ ?)- $\alpha$ -L-Araf-(1 $\rightarrow$ ?)-	arabinoBiosyl
C	$\alpha$ -L-Araf-(1 $\rightarrow$ ?)- $\alpha$ -L-Araf-(1 $\rightarrow$ ?)- $\alpha$ -L-Araf-(1 $\rightarrow$ ?)-	follows B
D	$\beta$ -D-Xylp-(1 $\rightarrow$ ?)- $\alpha$ -L-Araf-(1 $\rightarrow$ ?)-	follows C
L	$\alpha$ -L-Galp-(1 $\rightarrow$ ?)-	gaLactosyl
M	$\alpha$ -D-Galp-(1 $\rightarrow$ ?)-	follows L
N	$\beta$ -D-Galp-(1 $\rightarrow$ ?)-	follows M
U	$\alpha$ -D-GlcpA-(1 $\rightarrow$ ?)-	glUcuronosyl

**Table 1. The elementary letters of the short name system.**

Uppercase letters are the basic components of the code. ‘X’ represents unsubstituted main-chain  $\beta$ -D-xylopyranosyl units (including reducing terminal D-xylose). Other letters represent various decorations of the main-chain  $\beta$ -D-xylopyranosyl residues. It is expected that the letters ‘F’, ‘G’, ‘P’, ‘R’, and ‘V’ could be used in the future to represent fucosyl, glucosyl, mannosyl, rhamnosyl, and galacturonyl residues, respectively.

SUPERSCRIPT AUXILIARY COMPONENTS		
Lowercase letter	Structures represented	Mnemonic
a	<i>O</i> -Acetyl-	acetyl
f	<i>O</i> -Feruloyl-	feruloyl
c	<i>O-p</i> -Coumaroyl-	coumaroyl
m	<i>O</i> -Methyl-	methyl
Numerals	Linkage	
2	(1→2)	
3	(1→3)	
4	(1→4)	
5	(1→5)	
0	incertitude associated with the linkage position	
Punctuation	Definition	
,	separation of the successive residue linkages engaged inside a single side-chain	
/	multiple glycosidic substitutions of a single glycosyl residue within a side-chain	
+	double substitutions of a single main-chain D-xylosyl residue	
{	complex asymmetric double substitutions of a single main-chain D-xylosyl residue by glycosidic side-chains	

**Table 2. The auxiliary components of the short name system.**

The proposed nomenclature is completed by three other types of constituents: superscript lowercase letters that furnish additional chemical information relative to non-glycosidic substituents, superscript Arabic numerals that provide information about the position of substitutions, and punctuation. The use of uppercase letters in combination with superscripted uppercase letters provides a way to describe more complex side-chain motifs (*ibid.* Table 1).

<b>Structures of the xylosyl backbone</b>	<b>Name</b>
D-Xylp	X
3- <i>O</i> -Acetyl-β-D-Xylp	X <sup>3a</sup>
2,3-Di- <i>O</i> -acetyl-β-D-Xylp	X <sup>2a+3a</sup>
<b>Side-chain with a xylosyl residue at the reducing terminal position</b>	
β-D-Xylp-(1→2)-	Y <sup>2</sup>
<b>Side-chains with an arabinosyl residue at the reducing terminal position</b>	
α-L-Araf-(1→2)-	A <sup>2</sup>
α-L-Araf-(1→2)-[α-L-Araf-(1→3)]-	A <sup>2+3</sup>
5- <i>O</i> -Feruloyl-α-L-Araf-(1→3)-	A <sup>5f3</sup>
5- <i>O</i> - <i>p</i> -Coumaryl-α-L-Araf-(1→2)-	A <sup>5c2</sup>
5- <i>O</i> -Feruloyl-α-L-Araf-(1→3)-[α-L-Araf-(1→2)]-	A <sup>2+5f3</sup>
β-D-Galp-(1→5)-α-L-Araf-(1→2)-	A <sup>N5,2</sup>
α-L-Araf-(1→5)-α-L-Araf-(1→3)-	B <sup>5,3</sup> *
β-D-Xylp-(1→3)-α-L-Araf-(1→3)-α-L-Araf-(1→2)-	B <sup>Y3,3,2</sup> *
α-L-Araf-(1→2)-α-L-Araf-(1→3)-α-L-Araf-(1→2)-	C <sup>2,3,2</sup> *
β-D-Xylp-(1→3)-α-L-Araf-(1→2)-	D <sup>3,2</sup> *
β-D-Xylp-(1→2)-[5- <i>O</i> -Feruloyl]-α-L-Araf-(1→3)-	D <sup>2,5f3</sup> *
α-L-Galp-(1→4)-β-D-Xylp-(1→2)-α-L-Araf-(1→3)-	D <sup>L4,2,3</sup> *
α-L-Galp-(1→4)-β-D-Xylp-(1→2)-[5- <i>O</i> -Feruloyl]-α-L-Araf-(1→2)-	D <sup>L4,2,5f2</sup> *
<b>Side-chains with a glucuronyl residue at the reducing terminal position</b>	
α-D-GlcpA-(1→2)-	U <sup>2</sup>
α-D-GlcpA-(1→2)-3- <i>O</i> -acetyl-	U <sup>2+3a</sup>
4- <i>O</i> -Methyl-α-D-GlcpA-(1→2)-	U <sup>4m2</sup>
4- <i>O</i> -Methyl-α-D-GlcpA-(1→2)-3- <i>O</i> -acetyl-	U <sup>4m2+3a</sup>
α-D-Galp-(1→2)-4- <i>O</i> -Methyl-α-D-GlcpA-(1→2)-	U <sup>M2,4m2</sup>

\* The availability of letters that describe monosaccharidic side-chains and others that represent common disaccharidic side-chains means that it is theoretically possible to provide alternative names for some structures. However, to avoid ambiguity, side-chain motifs must be decomposed into the largest defined motifs, beginning with the reducing terminal unit of the side-chain.

**Table 3. Simplified names for a variety of heteroxyylan-derived oligosaccharide motifs.**

Structures	New abbreviations
$\begin{array}{c} \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 2 \\ \alpha\text{-L-Araf} \\ \uparrow 1 \end{array}$	A <sup>2</sup> X
$\begin{array}{c} 5\text{-O-Feruloyl-}\alpha\text{-L-Araf} \\ \downarrow 1 \\ \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 3 \end{array}$	A <sup>5f3</sup> X
$\begin{array}{c} \alpha\text{-L-Araf} \quad \alpha\text{-L-Araf} \\ \downarrow 1 \quad \downarrow 1 \\ \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 3 \quad \uparrow 3 \\ \alpha\text{-L-Araf} \\ \uparrow 2 \\ \uparrow 1 \end{array}$	XA <sup>3</sup> A <sup>2+3</sup> XX
$\begin{array}{c} \alpha\text{-L-Araf} \quad \alpha\text{-L-Araf} \\ \downarrow 1 \quad \downarrow 1 \\ \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 3 \quad \uparrow 3 \\ \alpha\text{-L-Araf} \\ \uparrow 2 \\ \uparrow 1 \end{array}$	XA <sup>2+3</sup> A <sup>3</sup> XX
$\begin{array}{c} \alpha\text{-L-Araf} \quad \alpha\text{-L-Araf} \\ \downarrow 1 \quad \downarrow 1 \\ \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 3 \quad \uparrow 3 \end{array}$	XA <sup>3</sup> XA <sup>3</sup> XX
$\begin{array}{c} \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 2 \quad \uparrow 2 \\ 4\text{-O-Methyl-}\alpha\text{-D-GlcpA} \quad 4\text{-O-Methyl-}\alpha\text{-D-GlcpA} \\ \uparrow 1 \quad \uparrow 1 \\ \alpha\text{-D-Galp} \\ \uparrow 2 \end{array}$	XU <sup>4m2</sup> XU <sup>M2,4m2</sup> X
$\begin{array}{c} O\text{-Acetyl} \quad O\text{-Acetyl} \quad O\text{-Acetyl} \\ \downarrow 1 \quad \downarrow 1 \quad \downarrow 1 \\ \beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-}\beta\text{-D-Xylp}-(1\rightarrow 4)\text{-D-Xylp} \\ \uparrow 3 \quad \uparrow 3 \quad \uparrow 3 \\ \alpha\text{-D-GlcpA} \quad O\text{-Acetyl} \quad O\text{-Acetyl} \\ \uparrow 2 \quad \uparrow 1 \quad \uparrow 1 \end{array}$	U <sup>2+3a</sup> XX <sup>3a</sup> X <sup>2a</sup> X <sup>2a+3a</sup>

**Fig. 1. Demonstration of the new abbreviated structural code system for AXOs generated by chemical or enzymatic treatment of arabinoxylans.**<sup>[8-14]</sup>

Structure	New abbreviation
<p> <math>\alpha\text{-L-Araf}</math>  <math>\downarrow</math>  <math>1</math>  <math>\downarrow</math>  <math>3</math>  <math>\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}</math>  <math>\uparrow</math>  <math>2</math>  <math>\uparrow</math>  <math>1</math>  <math>\alpha\text{-L-Araf}</math>  <math>\uparrow</math>  <math>2</math>  <math>\uparrow</math>  <math>1</math>  <math>\beta\text{-D-Xylp}</math> </p> <p> <math>\beta\text{-D-Xylp}</math>  <math>\downarrow</math>  <math>1</math>  <math>\downarrow</math>  <math>3</math>  <math>\beta\text{-D-Xylp}</math>  <math>\downarrow</math>  <math>1</math>  <math>\downarrow</math>  <math>3</math>  <math>\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}-(1 \rightarrow 4)-\beta\text{-D-Xylp}</math>  <math>\uparrow</math>  <math>2</math>  <math>\uparrow</math>  <math>1</math>  <math>5\text{-O-Feruloyl-}\alpha\text{-L-Araf}</math>  <math>\uparrow</math>  <math>3</math>  <math>\uparrow</math>  <math>1</math>  <math>\beta\text{-D-Xylp}</math> </p> <p> <math>\beta\text{-D-Xylp}</math>  <math>\downarrow</math>  <math>1</math>  <math>\downarrow</math>  <math>3</math>  <math>\alpha\text{-L-Araf}</math>  <math>\downarrow</math>  <math>1</math>  <math>\downarrow</math>  <math>3</math>  <math>\beta\text{-D-Xylp}</math>  <math>\uparrow</math>  <math>2</math>  <math>\uparrow</math>  <math>1</math>  <math>\alpha\text{-L-Araf}</math> </p>	$\{\text{D}^{2,2}\text{A}^3\} \text{X} \{\text{D}^{3,5\text{f}2}\text{Y}^3\} \text{X} \{\text{A}^2\text{D}^{3,3}\}$

**Fig. 2. The use of ‘curly brackets’ symbol to represent exceptional side-chain complexity.**

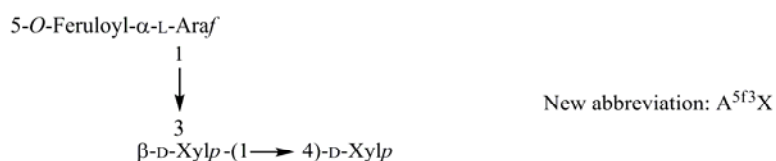
Curly brackets are used to indicate the asymmetric double substitution of a single main-chain D-xylosyl residue.<sup>[1]</sup>

## Table of Contents entry

### A brief and informationally-rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls

Régis Fauré, Christophe M. Courtin, Jan A. Delcour, Claire Dumon, Craig B. Faulds, Geoffrey B. Fincher, Sébastien Fort, Stephen C. Fry, Sami Halila, Mirjam A. Kabel, Laurice Pouvreau, Bernard Quemener, Alain Rivet, Luc Saulnier, Henk A. Schols, Hugues Driguez, and Michael J. O'Donohue

This article describes a convenient and information-rich method for naming heteroxylans. IUPAC rules do not provide simple nomenclature for heteroxylans, but current short names are informationally-poor. The naming system described provides a single letter-based system that should radically improve the published descriptions of heteroxylan structures, while remaining accessible to most researchers.



Nomenclature from IUPAC carbohydrate rules: *O*-[5-*O*-(*trans*-feruloyl)- $\alpha$ -L-Araf ]-(1,3)-*O*- $\beta$ -D-Xylp-(1,4)-D-Xylp  
Short form from IUPAC carbohydrate rules: Fe5Ara $\alpha$ 3Xyl $\beta$ 4Xyl  
Old abbreviation: FAXX