CarbBuilder: Software for Building Molecular Models of Complex Oligo- and Polysaccharide Structures.

Michelle M. Kuttel,[1] Jonas Ståhle,[2] and Göran Widmalm [2]

Correspondence to: Michelle M. Kuttel (E-mail: *mkuttel@cs.uct.ac.za*)

[1] M. M. Kuttel
Department of Computer Science, University of Cape Town, 7701, Cape Town, South Africa
[2] J. Ståhle, G. Widmalm
Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

ABSTRACT

CarbBuilder is a portable software tool for producing three-dimensional molecular models of carbohydrates from the simple text specification of a primary structure. CarbBuilder can generate a wide variety of carbohydrate structures, ranging from monosaccharides to large, branched polysaccharides. Version 2.0 of the software, described here, supports monosaccharides of both mammalian and bacterial origin and a range of substituents for derivatization of individual sugar residues. This improved version has a sophisticated building algorithm to explore the range of possible conformations for a specified carbohydrate molecule. Illustrative examples of models of complex polysaccharides produced by CarbBuilder demonstrate the capabilities of the software. CarbBuilder is freely available under the Artistic License 2.0 from https://people.cs.uct.ac.za/~mkuttel/Downloads.html.

Introduction

Carbohydrates encompass an incredibly diverse range of molecular structures; from the ubiquitous disaccharides to highly complex, multiply branched polysaccharides. The field of glycobiology is expanding as carbohydrates are increasingly recognized as key molecules in important biological processes, such as cell-cell interaction and host-pathogen recognition.[1-3] Knowledge of the three-dimensional structures of carbohydrates can greatly assist in a mechanistic understanding of these processes. As the characterization of carbohydrate molecular conformation remains a central problem in glycobiology, three-dimensional theoretical models are often useful for interpreting experimental results.[4-7] However, there remain few software tools for building models of complex, possibly branched, polysaccharides from the specification of the primary sequence. Currently available software includes two web-based tools – SWEET-II[8] and the GLYCAM carbohydrate builder[9] – and two standalone software packages – POLYS[10,11] and our CarbBuilder tool.[12] All these software packages follow the well-established 'pragmatic' approach to generating a three-dimensional (3D) model of a carbohydrate: monosaccharides are treated as stable, rigid building blocks and glycosidic linkage dihedral angles are used as the primary variable for building. An alternative approach to structure building is embodied in the fast structural prediction software (FSPS) which searches for energy minima in glycosidic conformation space with the assistance of NMR data.[13] However, FSPS is not publically available[14] and hence we could not determine the usefulness of this more

complex approach software for general, fast building of carbohydrate structures.

For the pragmatic approach to building, basic criteria for a valid 3D model of a carbohydrate is that it reflects the 2D connectivity accurately and contains no physically impossible self-intersections or steric clashes of the constituent residues. A more stringent criterion for validity would ensure that the structure does not contain conformations of the glycosidic linkages at odds with experimental or calculated conformations. However, given the scattered experimental and computational data on glycosidic linkages, this criterion is harder to enforce.

The pragmatic software tools have differing sets of monosaccharides available for building and employ different approaches to validation of the 3D models produced, as follows. The POLYS software currently supports 14 different residue building blocks in its MONOBANK database, a number which increases to 84 when variations on basic monosaccharide units resulting from α/β anomers, D/L enantiomers, furanose/pyranose forms of the sugar rings and derivatization of the residues with substituents (such as β-D-GlcNAc) are included. In POLYS, the user may supply a list of glycosidic torsion angles for each glycosidic linkage, or use the default conformational information from a database. POLYS may produce invalid structures with steric clashes, although the building procedure will give warnings if these occur. In this case, it is up to the user to establish better dihedral angles, which can be a difficult task in a constrained glycosidic linkage without prior knowledge of the preferred conformations.

The GLYCAM carbohydrate builder supports building with 29 residues, a range which is considerably extended by specification of α/β anomers, D/L enantiomers, furanose/pyranose forms of the sugar rings and derivatization of the residues with substituents. The GLYCAM carbohydrate builder has its own internal defaults for specific dihedral angles and allows the user to specify alternatives. However, the building process (which can be done either with the graphical input tool or the condensed GLYCAM notation[15]) requires multiple clicks on the web-page interface and can be problematic if a user wishes to explore the effect of a range of dihedral rotations on a large structure. It remains difficult to input a very large, branched structure or multiples of a repeating unit in GLYCAM. It is also not possible to use the GLYCAM carbohydrate builder without an internet connection, which makes it difficult to incorporate reliably into a software workflow.

The SWEET-II web tool supports building with about 50 basic residues, a range that is again extended to over 200 by specification of residue anomers, enantiomers, and substituents. SWEET-II contains a useful set of example templates and structures for input, including a text facility for input of very large, branched structures. SWEET-II has a similar building approach to GLYCAM, but valid models are ensured by optimization of the 3D structure with the MM3 force field.

We designed CarbBuilder to combine the best aspects of existing carbohydrate structure building software into a standalone, command-line tool, with a focus on simplicity and modularity. We aim for software that is robust, easy to use and portable across different architectures. Here we report a new version 2.0 of CarbBuilder, which incorporates an expanded set of monosaccharide building blocks, an extended range of substituents for generalized derivatization of sugar residues and further improvements to the building algorithm to facilitate rapid building of valid 3D models of a broader range of oligo- and polysaccharides structures than any other current software.

Features and Improvements

CarbBuilder is an command-line tool that takes as input a text string describing the primary

Table 1. Carbohydrate entities and fragments currently supported by CarbBuilder		
Anomers/ Enantiomers/ ring forms	Common name	CarbBuilder /CASPER designation
Aldose and ketose monosaccharides		
α, β / D[a] / p	abequose (3,6-dideoxy-D-galactose)	Abe
α, β / D, L / f,p	allose	All
α, β / D, L / p	altrose	Alt
α, β / D, L / p	6-deoxy altrose	Alt6d
α, β / D, L / f,p	arabinose	Ara
α, β / L[a] / p	colitose (3,6-dideoxy-L-galactose)	Col
α, β / D / f	erythose	Ery
α, β / D / f,p	fructose	Fru
α, β / L / p	fucose (6-deoxy-L-galactose)	Fuc
α, β / D, L / f,p	galactose	Gal
α, β / D, L /f,p	glucose	Glc
α / L / p	gulose	Gul
α, β / D, L / p	idose	Ido
α, β / D, L / f,p	lyxose	Lyx
α, β / D[a] / p	Kdo (3-deoxy-D-manno-oct-2-ulosonic acid)	Kdo
α, β / D, L / f,p	mannose	Man
α / D / p	sialic acid	Neu5Ac
α, β / D[a] / p	paratose (3,6-dideoxy-D-glucose)	Par
α, β/ D / p	quinovose (6-deoxy-D-glucose)	Qui
α, β/ L / p	rhamnose (6-deoxy-L-mannose)	Rha
α, β/ D, L / f,p	ribose	Rib
α, β / D[a] / p	tyvelose (3,6-dideoxy-D-mannose)	Tyv
α, β/ D, L / f,p	xylose	Xyl
Alditol monosaccharides		
D	glucitol	Glcol
D	glycerol	Gro
D	ribitol	Ribol
Derivatives		
	uronic acid	A
	acetyl	Ac
	amine	N
	N-acetyl	NAc
	N-amidino	NAm
	O-methyl	OMe
	phosphate	P
	phosphoethanolamine	PEtN
	4,6 pyruvate _(axial Me)_	46PyA[b]
	4,6 pyruvate _(equatorial Me)_	46PyE[b]

[a] when the absolute configuration is part of the trivial name, it may be omitted from the input.
[b] substituent only supported for DGal, DGlc, DMan.

structure of a carbohydrate in the simple CASPER format[12,16] and returns a three-dimensional structure in the Protein Data Bank (PDB) format.[17] For simplicity, we use the CASPER notation in all our discussions of structure below.

CarbBuilder allows for specification of multiples of a basic repeating unit to enable building of very large, branched polysaccharides. CarbBuilder produces either a single optimal 3D model, or, now, a whole family of possible structures, for any number of repeating units for a wide variety of complex, branched polysaccharides.

CarbBuilder is written in C# and hence is portable across different computer architectures.

Additional Monosaccharides and Substituents

CarbBuilder version 2.0 supports the common mammalian monosaccharides[3] and a broad range of bacterial monosaccharides, including linear alditols. Table 1 lists the monosaccharide residues supported in CarbBuilder version 2.0. This list has extended considerably since the first version of the software and is continuously updated (the "-help" command-line option will display the current list of monosaccharide residues and substituents). Taking into account the α/β anomers, the D/L enantiomers and the furanose/pyranose forms of the aldose and ketose residues, CarbBuilder currently supports over 100 basic monosaccharide components. Further, CarbBuilder allows for derivatization of existing residues with a range of substituents, also listed in Table 1. In version 2.0, CarbBuilder implements recursive parsing to allow for multiple substituents on a single residue, such as "aDGlcNAc46PyA", "aDGlcA2OMe4Ac" or "aLRha2Ac3Ac", which considerably extends the range of different monosaccharide building blocks into the thousands.

Building polysaccharide structures

CarbBuilder has a new, recursive routine for generating polysaccharide conformations. This procedure detects steric collisions or self-intersection of a chain when building (atomic 'collisions') which are reported to the user. CarbBuilder responds by backtracking and attempting linkages with alternative dihedral

values. Only after exhausting all possible valid dihedral combinations, will CarbBuilder

announce failure of the building routine.

Molecule	Command-line arguments to CarbBuilder , with primary structure in CASPER notation [branches in brackets].
	Table 2. Examples of the required input to CarbBuilder to generate a range of carbohydrates.
Chitin 100 RU	-i "->4)bDGlcNAc(1->" -r 100
Glycogen 20 RU	-i "->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)**[aDGlc(1->4)aDGlc(1->4)aDGlc(1->4***[aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->6)]**aDGlc(1->4)aDGlc(1->4)aDGlc(1->4)aDGlc(1->6)]**aDGlc(1->" -r 20
Sialylated N-glycan	-i "aDNeu5Ac(2->3)**[bDGalNAc(1->4)]**bDGal(1->4)bDGlcNAc(1->3)bDGal(1->4)bDGlcNAc(1->2)**[aDNeu5Ac(2->3)***[bDGalNAc(1->4)]***bDGal(1->4)bDGlcNAc(1->6)]**aDMan(1->6)**[aDNeu5Ac(2->3)***[bDGalNAc(1->4)]***bDGal(1->4)bDGlcNAc(1->2)***[aDNeu5Ac(2->6)bDGal(1->4)bDGlcNAc(1->4)]***aDMan(1->3)]**bDMan(1->4)bDGlcNAc(1->4)**[aLFuc(1->6)]**bDGlcNAc"
C. albicans serotype A cell wall mannan [26].	-i "aDMan(1->6)**[aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->3)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->3***[aDMan(1->6)]***aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->3)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2)aDMan(1->3***[aDMan(1->6)]***aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)aDMan(1->3)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)bDMan(1->2)aDMan(1->3)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[bDMan(1->2)bDMan(1->2)bDMan(1->2)bDMan(1->2)aDMan(1->2)aDMan(1->2)aDMan(1->2)]**aDMan(1->6)**[aDMan(1->2***[bDMan(1->2)bDMan(1->2)bDMan(1->2)aDMan1P(O->6)]***aDMan(1->2)aDMan(1->2)]**aDMan"

For the success of the building procedure, CarbBuilder relies on good initial values and alternatives for the dihedral angles in glycosidic linkages. A (1->X) glycosidic linkage is defined by the torsion angles ϕ = H'-C1'-O1'-CX and ψ = C1'-O1'-CX-HX. These definitions are analogous to ϕ and ψ in IUPAC convention. CarbBuilder has a large list of pre-computed optimal values of ϕ,ψ dihedral angle pairs for specific glycosidic linkages, determined from disaccharide potential of mean force (PMF) calculations in vacuum.[18] (These calculations are done using the new CHARMM force field for carbohydrates[19-21] and the metadynamics method in the NAMD[22] molecular dynamics program version 2.9.) For each dihedral angle

in the database, values of ϕ,ψ angle pairs for multiple minima are listed, in decreasing order of predicted probability. During the building routine, CarbBuilder attempts each of the torsion values in order, until a viable structure is generated, or all options are exhausted. If a clash occurs, CarbBuilder investigates whether a ±10 degree rotation in ψ dihedral angle will relax the strain, before moving on to the next dihedral angle pair.

The list of linkage dihedrals is regularly updated with additional linkages, but is not exhaustive (which would be a Sisyphean task, given the range of possible disaccharides). When specific values are required for building a structure, but

are not present in the database, CarbBuilder attempts to find reasonable defaults for linkages. If these are not present, generic α- or β-linkage defaults are used. For example, if the structure

"aDGlcNAc3Ac(1->4)aDGlcNAc3Ac"

is specified and the values for this linkage are not found in the database, CarbBuilder will use values from the 'core' linkage "aDGlc(1->4) aDGlc" in preference to the generic α-linkage defaults. Users may also specify values for one or more of the glycosidic linkages in a structure in a separate file, using the *–d [filename]* command-line option.

Finally, CarbBuilder can also generate *all* possible structures for a polysaccharide, given the dihedral values in the database. This is done using the using the *–all* command-line option to CarbBuilder. This option initiates the same recursive building procedure, but, instead of halting at the first successful build, CarbBuilder continues to build all non-intersecting structures. The resulting structures are output as a single PDB file containing multiple frames. However, as the number of possibilities grows exponentially with the number of residues, this option is impractical to use with very large molecules.

Application Examples

Large, branched polysaccharides

The command-line arguments to CarbBuilder required to build several representative large polysaccharides structures are listed in Table 2. The first two examples illustrate the use of the "-r" argument to specify multiple repeating units in a model. Figure 1 shows visualizations using the Visual Molecular Dynamics (VMD) software[23] of the resultant structures produced by CarbBuilder.
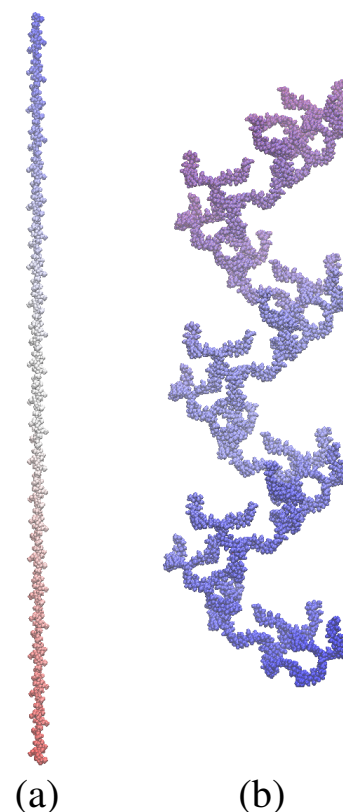

(a)   (b)

Figure 1. Examples of repeating polysaccharide structures created with CarbBuilder from the sequence specification in Table 2 and visualized with VMD[23]. (a) 100 RU of the chitin polysaccharide. (b) a highly branched glycogen-like structure containing 10503 atoms.

The linear chitin molecule contains only one type of glycosidic linkage: bDGlcNAc(1->4) bDGlcNAc. As CarbBuilder currently does not have torsion values for this specific linkage, dihedral angle values from the 'core' linkage bDGlc(1->4)bDGlc are used instead. There are four ϕ,ψ dihedral angle pairs (separated by commas) listed for this linkage, in decreasing order of preference, corresponding to minima calculated for the maltose disaccharide:

47 2.5,-25 -25,15 -175,180 0.

In the case of chitin, the first pair of ϕ,ψ dihedral angle values, 47 2.5, produces a valid 100 RU elongated structure with no atomic

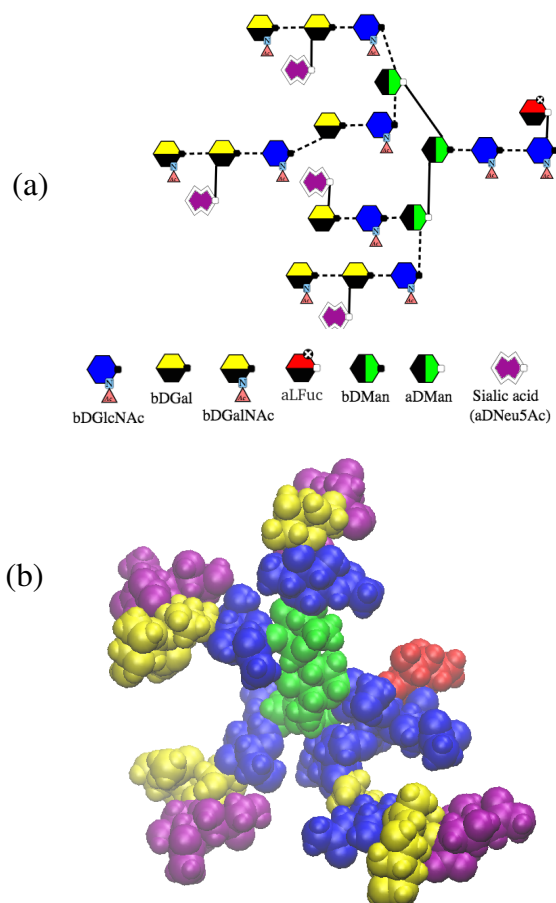collisions to be resolved, which is output by CarbBuilder (Figure 1a).

(a)

(b)

Figure 2. Typical complex N-glycan structure found on mature glycoproteins.[25] (a) 2D diagram of structure. (b) 3D model produced by CarbBuilder and visualized with VMD[23] from sequence specification in Table 2.

Glycogen contains one type of residue, aDGlc, and two types of linkage: (1->4) on the backbone and (1->6) linkages at the branches (Table 2). When building this molecule, CarbBuilder encounters steric clashes between hydroxymethyl groups on adjacent residues at the aDGlc(1->4)[aDGlc(1->6)]aDGlc branch points[24], which are only resolved on using the third pair, 25 25, of the ϕ,ψ dihedral angle values for the aDGlc(1->4)aDGlc linkage (listed

above). The resultant large model of a highly branched glycogen-like molecule (visualized in Figure 1b) thus has these alternate values of the aDGlc(1->4)aDGlc linkage at each branch point, but not elsewhere.

(a)

(b)
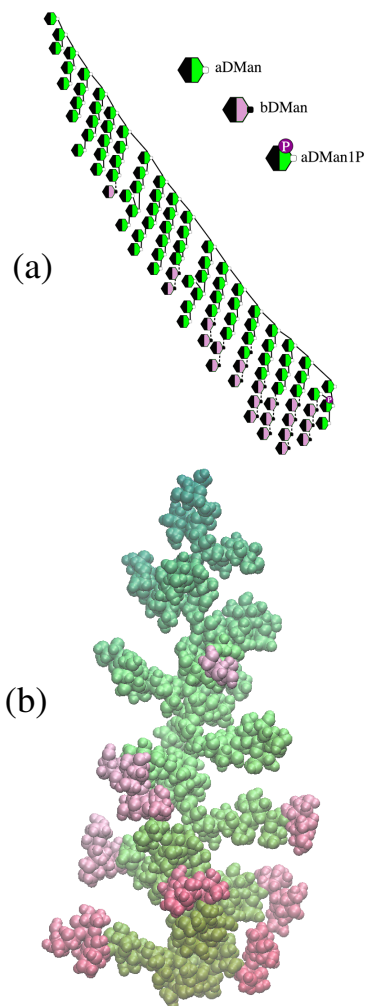
Figure 3. The complex cell wall mannan molecule of *C. albicans* serotype A[26]. (a) 2D diagram of structure from the sequence specification in Table 2. (b) The 2296 atom 3D model produced by CarbBuilder and visualized with VMD.

Our next two examples illustrate building of large, branching polysaccharides with no clear repeat unit in CarbBuilder. The N-glycan structures found on mature glycoproteins are composed of a variety of monsaccharides, including sialic acid (Table 2). Figure 2a shows a diagram of a typical complex structure,[25] with

Figure 2b visualizing the 3D structure produced by CarbBuilder. In this example, CarbBuilder contains very few pre-calculated values for the specific dihedral angles in the structure and has to resort to default values. The 3D model produced is thus valid in that it contains no steric clashes, but it may violate experimental measurements on specific linkages. Nevertheless, a structure is produced, which can be a starting point for more accurate modeling.

The cell wall mannan of *Candida albicans* serotype A is a prime example of a highly complex, highly branched polysaccharide. This structure comprises a backbone of a(1->6)-linked mannose residues and a(1->2)-linked mannose branches, with occasional a(1->3) linkages and capping β-mannose residues, as illustrated in Figure 3a.[26] As there is no repeating unit for this molecule, the entire structure was specified as input to CarbBuilder.

Figure 3b shows the resultant 2296 atom 3D model generated by CarbBuilder. For this example, we ensured that pre-computed optimal values of all the ϕ,ψ dihedral angle pairs were present in CarbBuilder for all the specific glycosidic linkages in the molecule. No steric clashes occurred on building. Thus, this structure can be considered to be a better estimate of the actual molecular structure than in the case of the N-glycan molecule. Further, as far as we could determine, this is the first 3D model of this complex polysaccharide produced.

Bacterial polysaccharides

To demonstrate CarbBuilder's ability to successfully build bacterial polysaccharide models containing variety of residues with unusual substitutions, we focus on the *Shigella flexneri* bacterium. On the basis of O-antigens, *S. flexneri* is divided into at least 19 serotypes[27], the majority of which are modifications of the same basic O-antigen by glucosylation and/or O-acetylation of the sugar residues (Table 3). These complex, branched structures with multiple residue types and substituents represent an effective demonstration of the efficacy of the CarbBuilder building routine for

Table 3. Primary sequences in CASPER notation for the O-polysaccharide repeating unit in 19 serotypes of *S. flexneri*.

Serotype	Command-line arguments to CarbBuilder , with primary structure in CASPER notation [branches in brackets].
Y	-i "->2)aLRha(1->2)aLRha(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
Yv	-i "->2)aLRha3PEtN(1->2)aLRha(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
X	-i ">2)[aDGlc(1->3)]aLRha(1->2)aLRha(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
Xv	-i "->2)[aDGlc(1->3)]aLRha(1->2)aLRha3PEtN(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
1a	-i "->2)aLRha(1->2)aLRha(1->3)aLRha(1->3)[aDGlc(1->4)]bDGlcNAc(1->" -r 6
1b	-i "->2)aLRha(1->2)aLRha(1->3)aLRha2Ac(1->3)[aDGlc(1->4)]bDGlcNAc(1->" -r 6
1d	-i "->2)[aDGlc(1->3)]aLRha(1->2)aLRha(1->3)aLRha(1->3)[aDGlc(1->4)]bDGlcNAc(1->" -r 6
2a	-i "->2)aLRha(1->2)aLRha(1->3)[aDGlc(1->4)]aLRha(1->3)bDGlcNAc(1->" -r 6
2b	-i "->2)[aDGlc(1->3)]aLRha(1->2)aLRha(1->3)[aDGlc(1->4)]aLRha(1->3)bDGlcNAc(1->" -r 6
3a	-i "->2)[aDGlc(1->3)]aLRha(1->2)aLRha(1->3)aLRha2Ac(1->3)bDGlcNAc(1->" -r 6
3b	-i "->2)aLRha(1->2)aLRha(1->3)aLRha2Ac(1->3)bDGlcNAc(1->" -r 6
4a	-i "->2)aLRha(1->2)aLRha(1->3)aLRha(1->3)[aDGlc(1->6)]bDGlcNAc(1->" -r 6
4av	-i "->2)aLRha3PEtN(1->2)aLRha(1->3)aLRha(1->3)[aDGlc(1->6)]bDGlcNAc(1->" -r 6
4b	-i "->2)aLRha(1->2)aLRha(1->3)aLRha2Ac(1->3)[aDGlc(1->6)]bDGlcNAc(1->" -r 6
5a	-i "->2)aLRha(1->2)[aDGlc(1->3)]aLRha(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
5b	-i "->2)[aDGlc(1->3)]aLRha(1->2)[aDGlc(1->3)]aLRha(1->3)aLRha(1->3)bDGlcNAc(1->" -r 6
6a	-i "->2)aLRha3Ac(1->2)aLRha(1->4)bDGalA(1->3)bDGalNAc(1->" -r 6
7a	-i "->2)aLRha(1->2)aLRha(1->3)aLRha(1->3)[aDGlc(1->2)aDGlc(1->4)]bDGlcNAc(1->" -r 6
7b	-i "->2)aLRha(1->2)aLRha(1->3)aLRha2Ac(1->3)[aDGlc(1->2)aDGlc(1->4)]bDGlcNAc(1->" -r 6

bacterial polysaccharides.

We used the CarbBuilder software to build six repeating units (6 RU) of each of the 19 *S. flexneri* O-polysaccharides. From the primary sequence inputs for each of the serotypes in CASPER format (Table 3), CarbBuilder produced the 3D models depicted in Figure 4. As some of the *S. flexneri* polysaccharides are extremely sterically crowded (e.g. serotype 7)[28], we found that pre-computed optimal values of all the ϕ,ψ dihedral angle pairs was vitally important for successful building of these branched polysaccharides; use of default torsion values resulted in unresolvable atomic clashes. It is clear from the structures shown in Figure 4 that the conformations of O-antigens within a serotype tend to have strong similarities, whereas serotypes are usually clearly distinct from each other. Compare, for example, serotype Y with serotype 2. However, in some cases, CarbBuilder predicts significant differences for the conformations of structures within a serotype. Serotype 1 and serotype 7 are predicted to have the most within-group conformational variability. However, due to the 'pragmatic' approach we have taken with CarbBuilder, the user should be wary of drawing too many conclusions from a single model conformation. Carbohydrates are known to be extremely flexible and more sophisticated modeling methods that take conformational families in solution into account are required to make accurate predictions.

Summary and Future Developments

The CarbBuilder software enables automated building of polysaccharides of arbitrary length from the text-representation of the primary structure. Version 2.0 includes additional residues and substituents and a robust building routine. The applications detailed here demonstrate CarbBuilder's ability to generate valid 3D models of complex, branched polysaccharides and bacterial polysaccharides with non-mammalian residues.

Models of carbohydrate molecules produced by CarbBuilder may be used to interpret experimental observations, or as a starting point for more complex molecular simulations.

Future developments will focus on adding an even wider range of monosaccharides and substituents, in order to cover as broad a range of as possible of polysaccharide structures, as well as automated building of glycoproteins and glycolipids.

CarbBuilder is implemented in C# and is freely available under Artistic License 2.0 from https://people.cs.uct.ac.za/~mkuttel/Downloads.html.
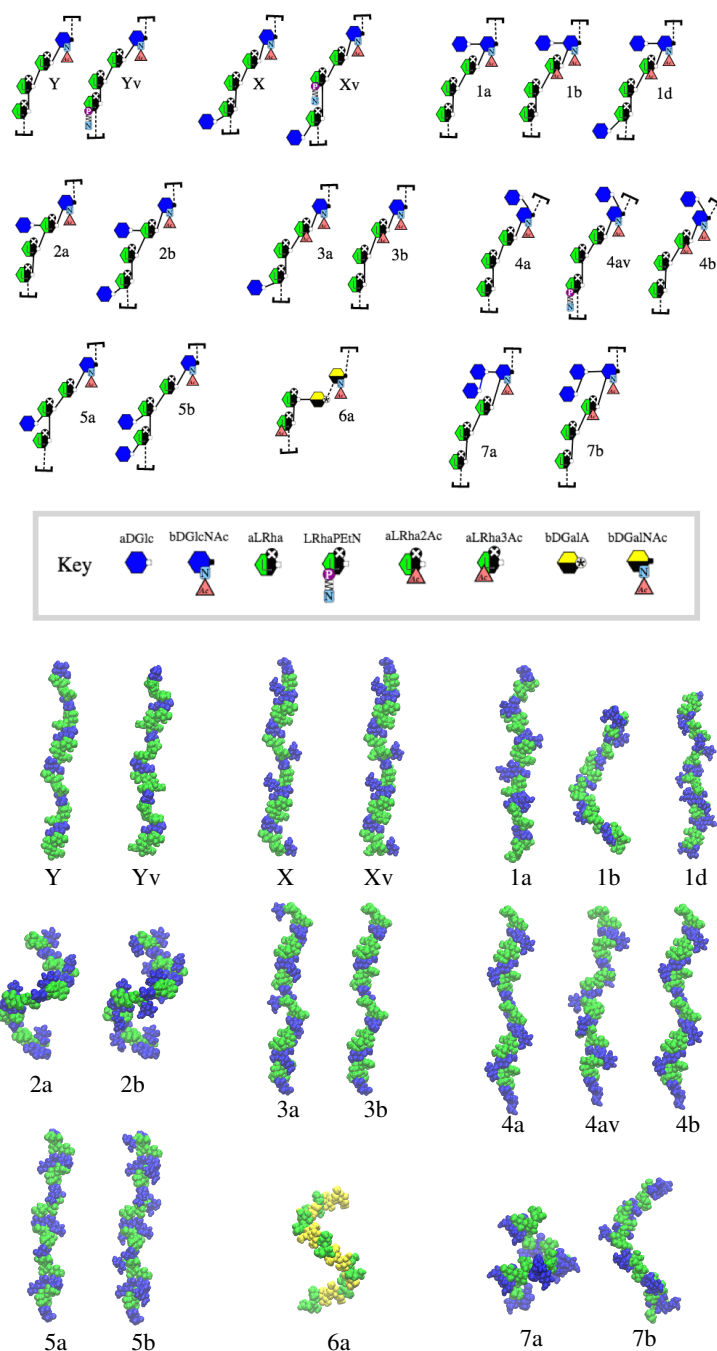
Acknowledgments

Figure 4. Graphical representations of the primary O-antigen structure of the 19 serogroups of *S. flexneri* (top) and 3D models of 6 repeating units of each of the O-antigens of *S. flexneri* (bottom) produced by CarbBuilder from the primary structure specifications in Table 3. (Molecular images are visualized with the Visual Molecular Dynamics[23] package). Residues with *gluco*-, *manno*- and *galacto*-configuration are colored blue, green and yellow, respectively (L-Rha = 6-deoxy-L-Mannose).

References and Notes

1. P. M. Rudd, T. Elliott, P. Cresswell, I. A. Wilson, R. A. Dwek, *Science* **2001,** *291*, 2370–2376.

2. G. A. Rabinovich , M. A. Toscano, *Nat. Rev. Immunol.* **2009**, *9*, 338-352

3. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler, Eds.; **2009**, Essentials of glycobiology, 2nd edn. Cold Spring Harbor Laboratory Press, Plainview, NY

4. M. L. DeMarco, R. J. Woods, *Glycobiology* **2008,** *18*, 426–440.

5. J. C. Muñoz-García, E. Chabrol, R. R. Vivès, A. Thomas, J. L. de Paz, J. Rojo, A. Imberty, F. Fieschi, P. M. Nieto, J. Angulo, *J. Am. Chem. Soc.* **2015**, *137*, 4100–4110

6. A. Sarkar, C. Fontana, A. Imberty, S. Pérez, G. Widmalm, *Biomacromolecules* **2013**, *14*, 2215–2224

7. M. Kuttel, G. E. Jackson, M. Mfata, N. Ravenscroft, *Carbohydr. Res*. **2015,** *406*, 27-33.

8. A. Bohne, E. Lang, C. W. von der Lieth, *Bioinformatics* **1999**, *15*, 767–768.

9. Woods Group (2005–2013) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. http://www.glycam.com (accessed Oct 10, 2015).

10. S. B. Englesen, P. I. Hansen, S. Pérez, *Biopolymers* **2014**, *101*, 733–743.

11. S. B. Englesen, S. Cros, W. Mackie, Pérez, *Biopolymers* **1996**, *39*, 417-433

12. M. Kuttel, Y. Mao, G. Widmalm, M. Lundborg in Proceedings of 7th IEEE International Conference on e-Science, Stockholm, Sweden, Dec 5-8, 2011, IEEE Computer Society Press, USA.

13. J. Xia, C. Margulis, *J Biomol NMR* **2008**, *42*, 241–256.

14. M. Frank, S. Schloissnig, *Cell. Mol. Life Sci.* **2010**, *67*, 2749–2772.

15. GLYCAM "condensed" oligosaccharide notation,http://glycam.org/docs/custombuilders/click-here-to-learn-condensed-notation/ (accessed Oct 10, 2015).

16. M. Lundborg, G. Widmalm, *Anal. Chem.* **2011**, *83*, 1514-1517.

17. Worldwide Protein Data Bank, File Format Documentation, http://www.wwpdb.org/documentation/file-format (accessed Oct 10, 2015).

18. M. Kuttel, M. Gordon, N. Ravenscroft, *Carbohydr. Res*. **2014** *390*, 20-27.

19. O. Guvench, S. N. Greene, G. Kamath, J.W. Brady, R.M. Venable, R. W. Pastor, A. D. MacKerell Jr., *J. Comput. Chem.* **2008**, *29*, 2543-2564.

20. O. Guvench, E. Hatcher, R.M. Venable, R. W. Pastor, A. D. MacKerell Jr., *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370.

21. S. S. Mallajosyula, O. Guvench, E. Hatcher E, A. D. MacKerell Jr., *J. Chem. Theory Comput.* **2012**, *8*, 759-776.

22. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781-1802.

23. W. Humphrey, A. Dalke, K. Schulten, *J Mol. Graph.* **1996**, 14, 33-38.

24. F. Corzana, M. S. Motawia, C. Hervé du Penhoat, F. van den Berg, A. Blennow, S. Perez S. B. Engelsen, *J. Am. Chem. Soc.*, **2004**, *126 (40)*, 13144–13155

25. P. Stanley, H. Schachter, N. Taniguchi. N-Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; **2009**. Chapter 8.

26. N. Shibata, A. Suzuki, H. Kobayashi, Y. Okawa, *Biochem. J.* **2007** *404*, 365–372

27. Q. Sun, R. Lan, J. Wang, S. Xia, Y. Wang, Y. Wang, D. Jin, B. Yu, Y. A. Knirel, J. Xu, *PLoS ONE* **2013** *8,* e70238.

28. R. A. Foster, N. I. Carlin, M. Majcher, H. Tabor, L. K. Ng, G. Widmalm, *Carbohydr. Res.* **2011**, *346*, 872-6.

GRAPHICAL ABSTRACT

AUTHORS

Michelle M. Kuttel, Jonas Ståhle, Göran Widmalm

TITLE

CarbBuilder: Software for Building Molecular Models of Complex Oligo- and Polysaccharide Structures.

TEXT

CarbBuilder is a command-line software tool for rapid, automated building of 3D models of complex polysaccharides of arbitrary length from the text-representation of the primary structure. CarbBuilder version 2.0 supports the common mammalian monosaccharides and a broad range of bacterial monosaccharides, including linear alditols, and a range of residue substituents. CarbBuilder can generate 3D models of large linear or branching polysaccharides containing hundreds of iterations of a single repeating unit.

GRAPHICAL ABSTRACT FIGURE