WURCS2.0 undefined LINs input 2018-10-06

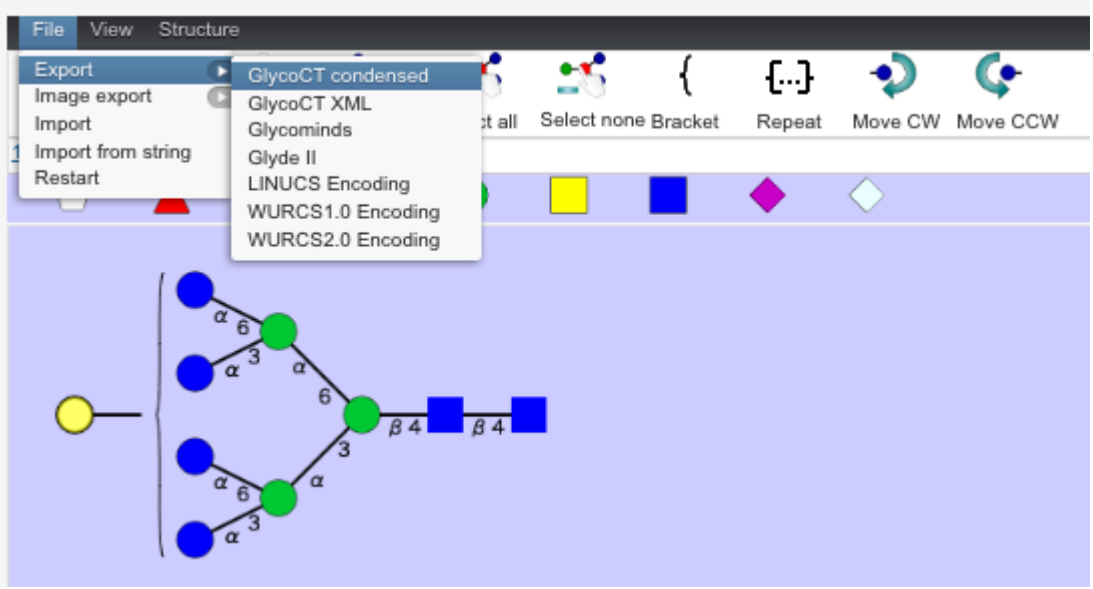
2018-10-06

Issaku Yamada

Generation of WURCS2.0 with undefined LINs

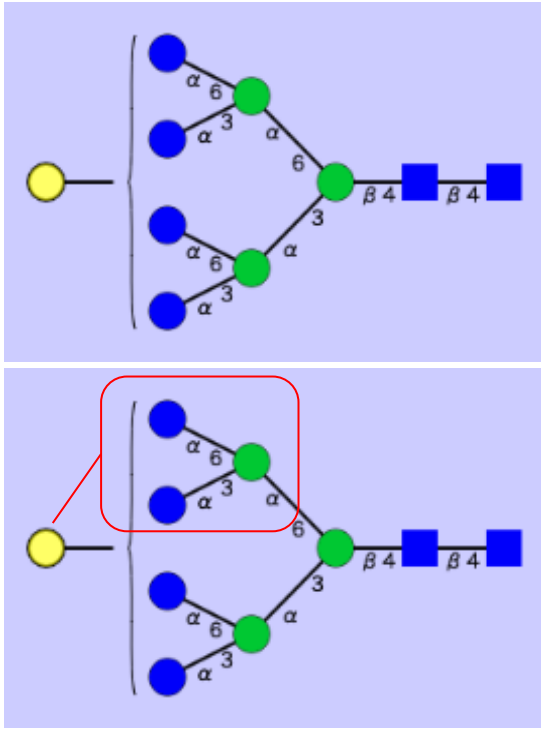
Using GlycoCT

1. https://glytoucan.org/Structures/graphical



2. download xxxxxxxxxxxx.glycoct_condensed file
3. open xxxxxxxxxxxx.glycoct_condensed file

RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dglc-HEX-1:5 8b:a-dglc-HEX-1:5 9b:a-dman-HEX-1:5 10b:a-dglc-HEX-1:5 11b:a-dglc-HEX-1:5 LIN	1. Exported GlycoCT condensed 2. Fixing unidentified linkage information for SubtreeLinkageID.

1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:6o(3+1)7d 7:6o(6+1)8d 8:5o(6+1)9d 9:9o(3+1)10d 10:9o(6+1)11d UND UND1:100.0:100.0 ParentIDs:1|3|5|6|7|8|9|10|11 SubtreeLinkageID1:u(-1+1)u RES 12b:x-dgal-HEX-1:5	
RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dglc-HEX-1:5 8b:a-dglc-HEX-1:5 9b:a-dman-HEX-1:5 10b:a-dglc-HEX-1:5 11b:a-dglc-HEX-1:5 LIN 1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:6o(3+1)7d 7:6o(6+1)8d 8:5o(6+1)9d 9:9o(3+1)10d 10:9o(6+1)11d UND UND1:100.0:100.0 ParentIDs:1|3|5|6|7|8|9|10|11 SubtreeLinkageID1:o(-1+1)d RES 12b:x-dgal-HEX-1:5	1. Chenage u(-1+1)u in UND. SubtreeLinkageID[1-9]+ 2. Delete unlinked RES number(s) (ex. 6,7,8) 
RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dglc-HEX-1:5 8b:a-dglc-HEX-1:5 9b:a-dman-HEX-1:5 10b:a-dglc-HEX-1:5	WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-a?|b?|c?|g?|h?|i?}

11b:a-dglc-HEX-1:5 LIN 1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:6o(3+1)7d 7:6o(6+1)8d 8:5o(6+1)9d 9:9o(3+1)10d 10:9o(6+1)11d UND UND1:100.0:100.0 ParentIDs:9|10|11 SubtreeLinkageID1:o(-1+1)d RES 12b:x-dgal-HEX-1:5	

4. Convert GlycoCT to WURCS http://www.wurcs-wg.org/tool/converter/glycocttowurcs/input

inut GlycoCT condenced	output WURCS
RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dglc-HEX-1:5 8b:a-dglc-HEX-1:5 9b:a-dman-HEX-1:5 10b:a-dglc-HEX-1:5 11b:a-dglc-HEX-1:5 LIN 1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:6o(3+1)7d 7:6o(6+1)8d 8:5o(6+1)9d 9:9o(3+1)10d 10:9o(6+1)11d UND UND1:100.0:100.0 ParentIDs:9|10|11 SubtreeLinkageID1:o(-1+1)d RES 12b:x-dgal-HEX-1:5	WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

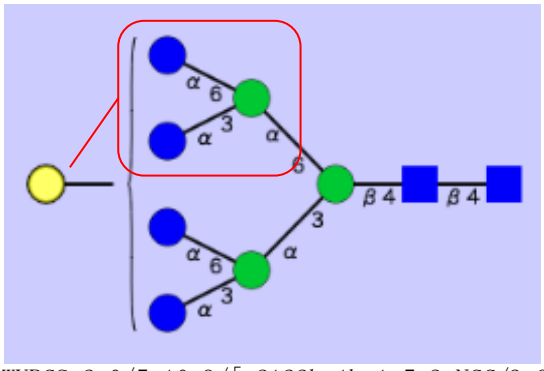
5. Check WURCS String http://www.wurcs-wg.org/tool/WURCSchecker.php

input	output
WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2	The character strings at WURCSinput and WURCS output are equal.

112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}	WURCSinput WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?} WURCSoutput WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

6. WURCSoutput is standard WURCS2.0
   WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

Using WURCS

1. check RES index of WURCS
2. 

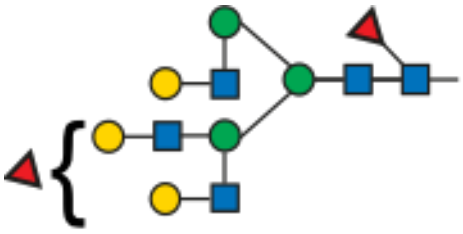
3. WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-a?|b?|c?|d?|e?|f?|g?|h?|i?}
4. Delete undefined LINs contains unlinked RES index(s) (ex. a,b,c,d,e,f)
5. WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}
6. Check WURCS String http://www.wurcs-wg.org/tool/WURCSchecker.php
7.

input	output
WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}	The character strings at WURCSinput and WURCS output are equal. WURCSinput WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?} WURCSoutput WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

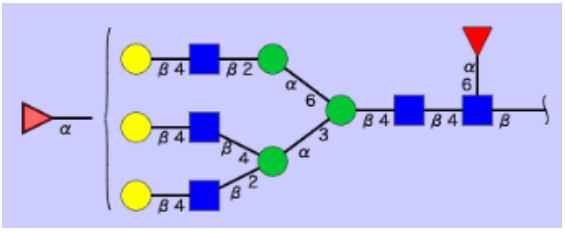
8. WURCSoutput is standard WURCS2.0
   WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

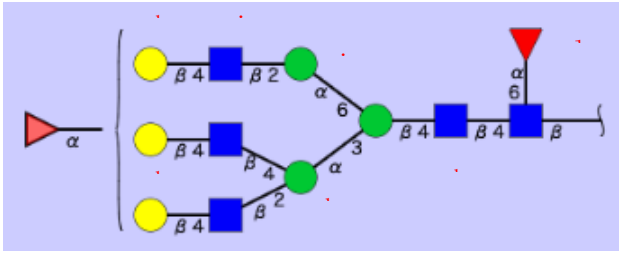
GlycoCTtoWURCS vs WURCS modify metod

GlycoCTtoWURCS	WURCS modify
WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}	WURCS=2.0/5,10,9/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5][a2122h-1a_1-5][a2112h-1x_1-5]/1-1-2-3-4-4-3-4-4-5/a4-b1_b4-c1_c3-d1_c6-g1_d3-e1_d6-f1_g3-h1_g6-i1_j1-g?|h?|i?}

Example of 

beta linkage represent dashed line in oxford style. anomeric information in above structure is unclear for me, it assigned general anomeric form in some linkages.

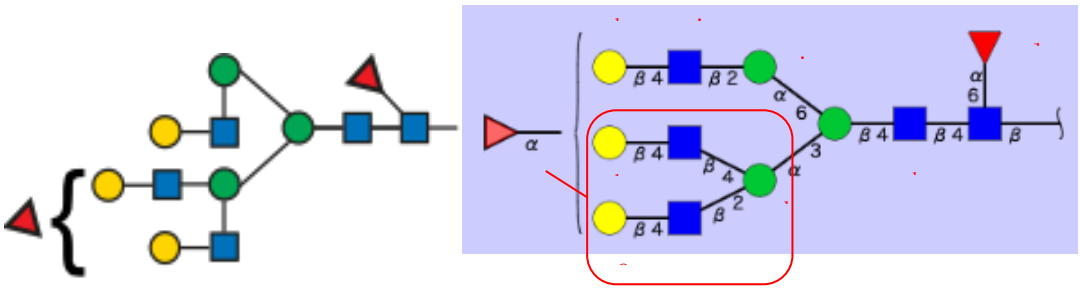




GlycoCT	WURCS
RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:b-dglc-HEX-1:5 8s:n-acetyl 9b:b-dgal-HEX-1:5 10b:b-dglc-HEX-1:5 11s:n-acetyl 12b:b-dgal-HEX-1:5 13b:a-dman-HEX-1:5 14b:b-dglc-HEX-1:5 15s:n-acetyl 16b:b-dgal-HEX-1:5 17b:a-lgal-HEX-1:5|6:d LIN 1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:6o(2+1)7d 7:7d(2+1)8n 8:7o(4+1)9d 9:6o(4+1)10d	WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-a?|b?|c?|d?|e?|f?|g?|h?|i?|j?|k?|l?}

10:10d(2+1)11n 11:10o(4+1)12d 12:5o(6+1)13d 13:13o(2+1)14d 14:14d(2+1)15n 15:14o(4+1)16d 16:1o(6+1)17d UND UND1:100.0:100.0 ParentIDs:1|3|5|6|7|9|10|12|13|14|16|17 SubtreeLinkageID1:o(-1+1)d RES 18b:a-lgal-HEX-1:5|6:d	

WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-a?|b?|c?|d?|e?|f?|g?|h?|i?|j?|k?|l?}



Delete undefined LINs contains unlinked RES index(s) (ex. a,b,c,i,j,k,i)

WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-d?|e?|f?|g?|h?}

Check WURCS String http://www.wurcs-wg.org/tool/WURCSchecker.php

input	output
WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-d?|e?|f?|g?|h?}	The character strings at WURCSinput and WURCS output are equal. WURCSinput WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-d?|e?|f?|g?|h?} WURCSoutput WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-d?|e?|f?|g?|h?}

WURCSoutput is standard WURCS2.0

WURCS=2.0/5,13,12/[a2122h-1b_1-5_2*NCC/3=0][a1122h-1b_1-5][a1122h-1a_1-5][a2112h-1b_1-5][a1221m-1a_1-5]/1-1-2-3-1-4-1-4-3-1-4-5-5/a4-b1_a6-l1_b4-c1_c3-d1_c6-i1_d2-e1_d4-g1_e4-f1_g4-h1_i2-j1_j4-k1_m1-d?|e?|f?|g?|h?}