Supplementary Data

A Highly Efficient Tree Structure for the Biosynthesis of Heparan Sulfate Accounts for the Commonly Observed Disaccharides and Suggests a Mechanism for Domain Synthesis.

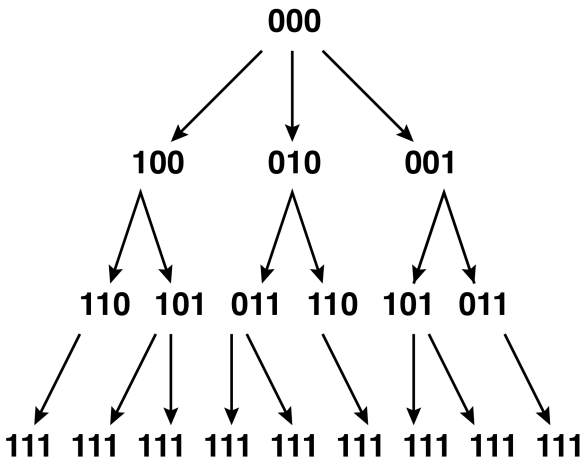
Timothy R. Rudd[1,2] and Edwin A. Yates[2,*]

[1]Istituto di Ricerche Chimiche e Biochimiche "G. Ronzoni", Via Giuseppe Colombo, 81 Milano 20133 ITALIA

[2]Institute of Integrative Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB UK

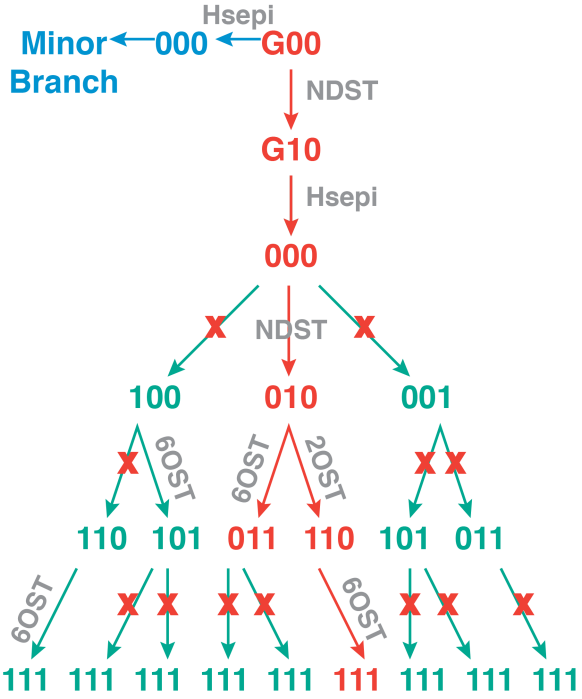
*To whom correspondence should be addressed, eayates@liv.ac.uk

Starting with the enzyme activities in any order and on any substrate, constraints corresponding to known experimental observations of the system were progressively imposed and the results of these models these are shown in **Supplementary Data Figures 1-4**.



Supplementary Figure 1. Hypothetical biosynthesis possibilities from IdoA-GlcNAc (000) if no constraints of enzyme specificity or order are imposed.

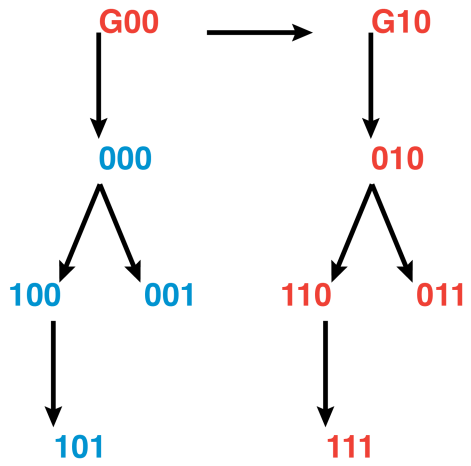
These schemes, in combination with ^{13}C NMR data can be used as an indicator of the conformational space that can be visited during a number of hypothetical biosynthetic routes although detailed geometry cannot be recovered directly. Before application of the known constraints, an unconstrained model was generated. In this, all modifications were permitted, leading to the situation depicted (for a part of the system only; that starting from 000) by the *graph* in **Supplementary Data Figure 1.** This represents a hypothetical system in which numerous biosynthetic pathways are feasible for each structure. A modification of this basic system, moving closer to that thought to prevail (assuming there is just one) in biology, is then generated by applying constraints to the order of enzyme application and restricting the substrates to those known (**Supplementary Data Figures 2 and 3**). A key point is that it is not possible to generate IdoA2S-GlcNAc (100) from IdoA-GlcNAc (000). Instead, 000 must originate from G00 by action of the Hsepi enzyme, which has been observed. In the quest to relate the substitution pattern of HS (or its close analogue, heparin) to its conformational properties and hence its interactions with proteins, culminating in biological activity, key relationships between substitution pattern and molecular geometry are sought. Molecular details are currently obtained largely using NMR spectra and, although information-rich, these remain difficult to interpret because of the complex, charged nature of HS and the challenges associated with modelling them, although significant progress is being made [1-3].



Supplementary Figure 2. Formation of the major branch (red) following the application of contraints of the order of enzyme action (enzyme shown in grey) and substrate. Note that disallowed steps are shown by a magenta cross on the arrow, which not only prevents that step but also subsequent steps that, otherwise, would be allowed e.g. conversion of 110 to 111. In effect, this limits the number of routes that are available to make any given structure. Note also that production of the IdoA2S series (originating from 100) is forbidden because 000 cannot be converted to 100. This must occur *via* direct conversion of G00 to 000 (blue) by epimrase,

which is known, and this gives rise to the minor branch.

In the course of pursuing these aims, we have collected NMR data for a number of systematically modified heparin derivatives, containing the basic units of the heparan sulfate and heparin polysaccharides. These have been analysed previously in terms of the effects on chemical shifts, as a function simply of substitution pattern, for the purposes of characterisation and to reveal the likely presence of hydrogen bonding [4], examined in terms of structural changes, especially at glycosidic bonds through $^{1}J_{CH}$ coupling constants [5] and, more recently, following statistical analysis of chemical shift values [6], employed as the starting point for modelling redundancy between substitution pattern and variation in the glycosidic linkages [7]. This last study revealed that the glycosidic linkages were differentially influenced by the substitution of the basic disaccharide repeating unit, but that there was limited variability and high redundancy in these patterns.

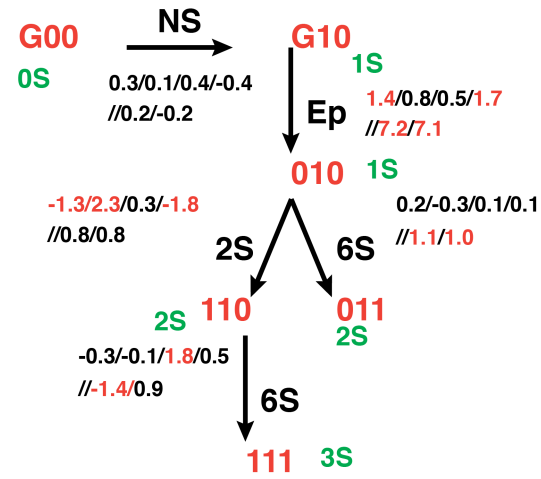


Supplementary Figure 3. The resulting pathway showing major (red) and minor (blue) pathways and applying the same principle in the minor branch. Further elaboration also allows the series of compounds containing GlcA and GlcA2S to be produced. The final pathway is shown in the paper (**Figure 2**).

Variation in ^{13}C chemical shifts at I-3 and I-5 showed essentially two groups, suggesting two broad conformations in these model polysaccharides. Those structures with O-sulfate groups at I-2 and N-acetyl at A-2 were clearly distinct from all other combinations of O-sulfation, N-sulfation and N-acetylation [4] and also had distinct chemical shifts at I-4. Here, we extend this line of thought to examine the variation in glycosidic linkage geometry and variation in iduronate ring flexibility, which is known to depend on substitution pattern [8] that can be explored during sequential structural modifications during biosynthesis following a number of potential routes.

^{13}C δ values of 6 principal structures in HS/heparin

	UA-4	UA-1	A-4	A-1	UA-3	UA-5
G00	79.1	105.2	81.1	99.5	78.8	79.1
G10	78.8	105.1	80.7	99.9	78.7	79.3
010	77.4	104.3	80.2	98.2	71.5	72.2
110	78.7	102.0	80.5	100.0	70.7	71.4
011	77.2	104.6	80.1	98.1	70.4	71.2
111	79.0	102.1	78.8	99.5	72.1	72.3



Changes in δ - shifts > 1 ppm in red.
Quoted in order: UA4/UA1/A4/A1 //UA3/UA5

Supplementary Figure 4. ^{13}C NMR chemical shift changes (ppm) at the glycosidic linkage positions (UA-4, UA-1, A-1 and A-4) and positions 3 and 5 (UA-3 and UA-5) of uronic acid residues for the structures in the major branch of the proposed biosynthetic route. The number of sulfates is shown alongside each structure in green.

The extent and location of structural variation throughout the repeating units of the polysaccharide can be monitored through changes in ^{13}C chemical shifts, even if at present they remain difficult to interpret in detailed geometric terms. Each disaccharide repeating unit can be viewed as having four degrees of freedom at the glycosidic linkages, namely: O_{g1}...A-4, A-1...O_{g2}, O_{g2}...I-4 and I-1...O_{g2}. The geometry of these are influenced in distinct ways and in a range of patterns by the substitution pattern in the disaccharide, but not all combinations of variation are possible. Four degrees of freedom implies 2^4 (=16) possible combinations of the *presence* of effects (that is, before considering the extent of such effects) at these positions but, for a disaccharide comprising IdoA and GlcN, there are only 8 possible substitution patterns if O-sulfation, N-acetyl and N-sulfate substitutions are considered, or 12, if free amino groups on glucosamine are also included. Consequently, even with this simple argument, it is clear that not all combinations of (the existence of)

variation at the four positions can be sampled in the basic disaccharide unit. In fact, there is evidence of considerable degeneracy even within these limited combinations and some combinations of variation at the linkages are not attainable in model polysaccharides [8]. Extensive variation in $^{1}J_{CH}$ coupling constants, particularly those involved in the glycosidic linkage positions and in the iduronate residues, has also been observed in heparin derivatives [5], often at sites remote from the position of substituents such as O- or N-sulfates and this suggests a substantial degree of geometric change at both linkage positions and in the uronate residues. The extent of variability at the four positions involved in formation of the glycosidic linkages, A-4, A-1, I-4 and I-1, but also in iduronate (I-3 and I-5, indicative of conformational change) can be monitored through ^{13}C NMR chemical shift changes and its development through the biosynthetic pathway, connecting biosynthesis with a measure of geometric variation, can be studied.

References

[1]. J. Am. Chem. Soc., (2008) 130, 12435-12444
[2]. J. Am. Chem. Soc., (2010) 130, 13132-13134
[3]. Glycobiol., (2009) 19, 1185-1196
[4]. Carbohydr.Res., (1996) 294, 15-27
[5]. Carbohydr. Res., (2000) 329, 239-247
[6]. Glycobiol., (2009) 19, 52-67
[7]. Mol. Biosys. (2010) 6, 902-908
[8]. Carbohydr.Res., (1987) 165, c1-c5