Supplementary Data

Supplementary Table 1 Systematically modified heparins used in the analysis.

	Position of Substitution		
Notation	**I2 / R_1**	**A6 / R_2**	**A2 / R_3**
$I_{2OH}\ A^{6OH}_{NH2}$	H	H	H
$I_{2OH}\ A^{6OH}_{NS}$	H	H	SO_3^-
$I_{2OH}\ A^{6S}_{NS}$	H	SO_3^-	SO_3^-
$I_{2S}\ A^{6S}_{NH2}$	SO_3^-	SO_3^-	H
$I_{2S}\ A^{6S}_{NAc}$	SO_3^-	SO_3^-	$COCH_3$
$I_{2S}\ A6^{OH}_{NS}$	SO_3^-	H	SO_3^-
$I_{2OH}\ A^{6S}_{NH2}$	H	SO_3^-	H
$I_{2OH}\ A^{6S}_{NAc}$	H	SO_3^-	$COCH_3$
$I_{2OH}\ A^{6OH}_{NAc}$	H	H	$COCH_3$
$I_{2S}\ A^{6OH}_{NH2}$	SO_3^-	H	H
$I_{2S}{}^{A6OH}_{NAc}$	SO_3^-	H	$COCH_3$
$I_{2S}\ A^{6S}_{NS}$	SO_3^-	SO_3^-	SO_3^-

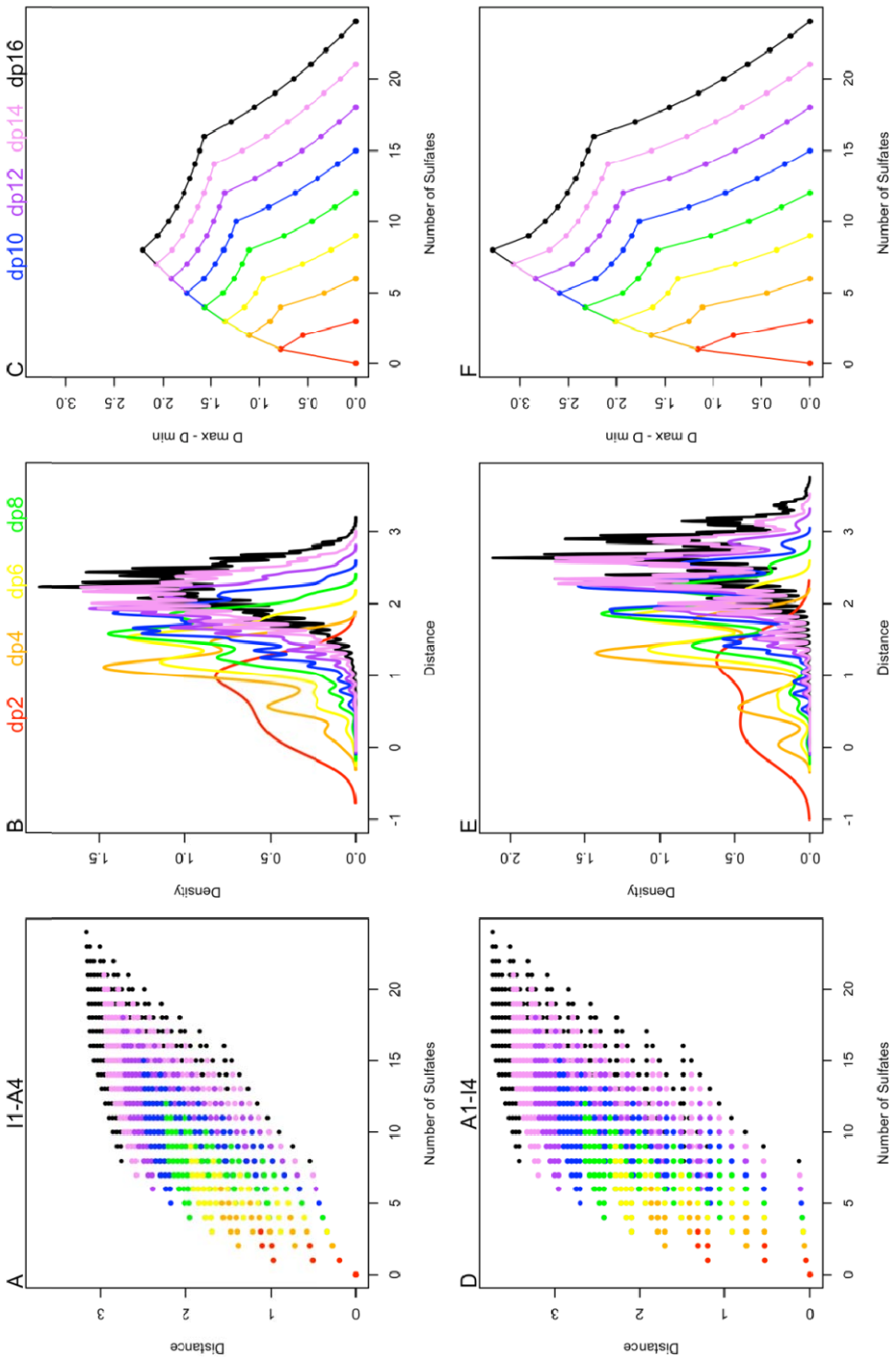
I-*N* refers to the N^{th} position of iduronate and A-*N* to the N^{th} position of glucosamine (aminosugar) residues respectively.

Supplementary Table 2 ^{13}C chemical Shift assignments (ppm) for 12 chemically modified heparins

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A1	95.8	98.2	98.1	93.7	96.6	100.0	97.9	97.1	97.1	97.2	96.8	99.5
A2	57.0	60.5	60.3	57.1	56.2	60.8	57.6	56.2	56.2	58.0	56.6	60.7
A3	71.8	72.5	72.4	70.9	73.0	72.4	73.9	72.5	72.3	73.5	72.9	72.5
A4	79.0	80.2	80.1	78.2	79.3	80.5	79.6	79.6	79.6	80.1	80.6	78.8
A5	74.3	73.5	71.5	72.0	72.3	73.8	72.7	71.8	73.7	74.2	74.2	72.0
A6	61.9	62.4	68.7	68.8	69.6	62.6	69.0	68.8	62.3	62.8	62.9	69.2
I1	104.5	104.3	104.6	101.3	102.2	102	104.8	104.6	104.3	101.9	102.3	102.1
I2	71.7	72.2	71.1	75.3	76.8	77.6	72.9	72.0	72.5	76.0	76.6	78.9
I3	71.8	71.5	70.4	65.4	67.3	70.7	72.4	71.4	72.2	67.0	67.1	72.1
I4	77.4	77.8	77.2	73.0	74.2	78.7	77.8	77.0	77.1	74.4	74.1	79.0
I5	71.8	72.2	71.2	69.8	70.8	71.4	72.8	71.9	72.6	70.1	70.6	72.3

Supplementary Table 3. The relative contribution of the first three principal components (c1 (accounting for 38.8 % of the variance in the data), c2 (22.3 %) and c3 (18.5 %) to the range of ^{13}C chemical shift values observed at the linkage positions among the 12 modified polysaccharides.

Position	**Component 1**	**Component 2**	**Component 3**
A1	0.05	0.87	0.04
A4	0.02	0.19	0.51
I1	0.98	0.02	-0.01
I4	0.52	0.84	0.02



Supplementary Figure 1. Greater diversity is evident in A1--I4 linkages then I-1--A-4 linkages. Analysis of variations in individual glycosidic linkages from combinatorial modelling of all possible sequences containing from 2 to 16 residues, A-1, I-4 and I-1, A-4 individually. **A.** (A-1, I-4), **D.** (I-1, A-4) linkage variation against number of sulfates, **B.** (A-1, I-4), **E.** (I-1, A-4) kernel density plots of the linkage variation distributions (in **A.** and **D).** and **C.** (A-1, I-4), **F.** (I-1, A-4) variation of the linkage distance at every level of sulfation for 2-16 residues.

Supplementary Table 4 Analysis of the variation in individual glycosidic linkages from combinatorial modelling of all possible oligosaccharide stretches from 2 to 16 residues, linkages A-1, I-4 and I-1, A-4 individually.

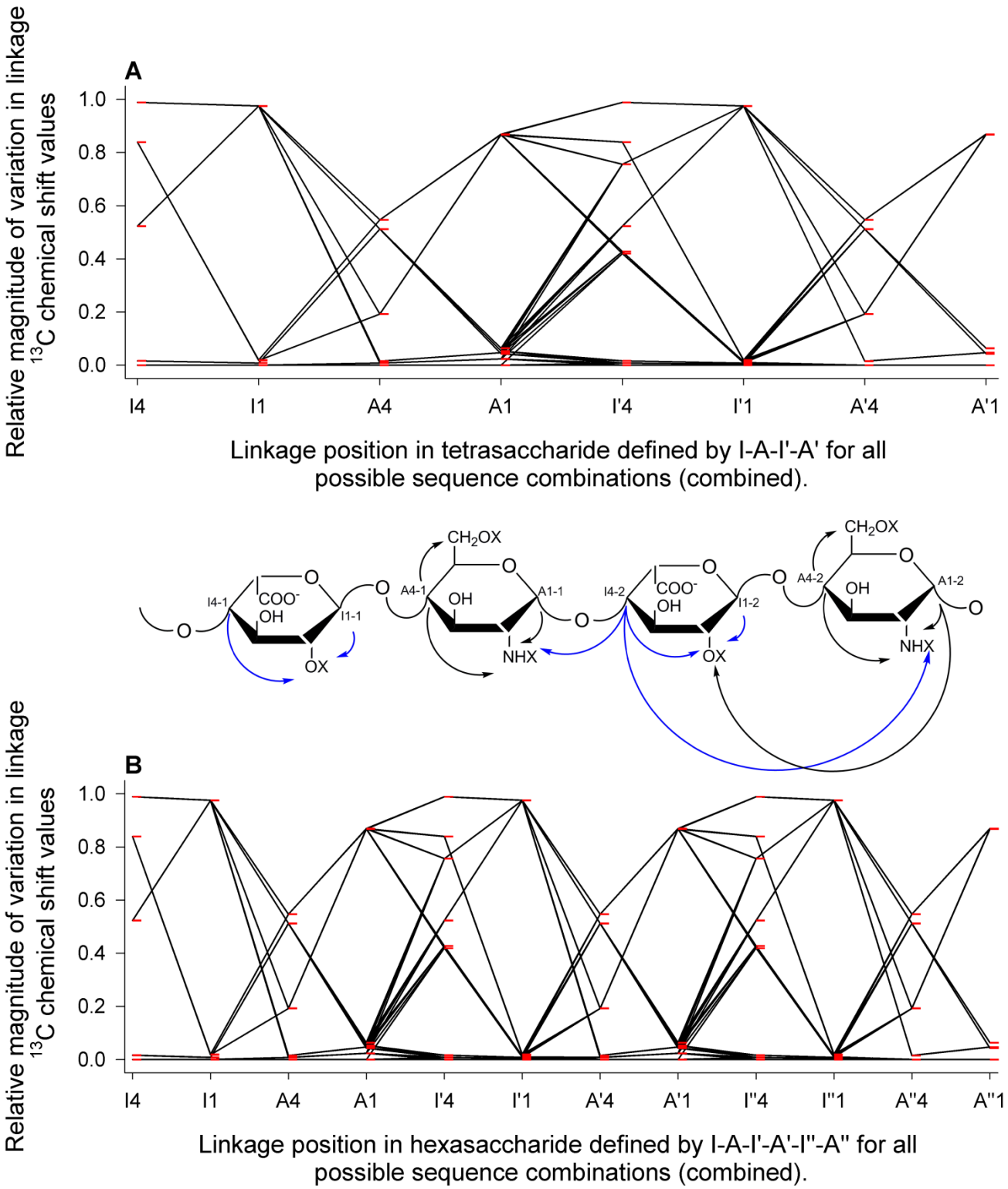
A1-I4

	Distance					*Sulfation*		
residues	**Mean**	**Skew**	**Max. D.**	**M.V.D[a]**	**N° of S that M.V.D occurs at**	**Mean**	**Skew**	**Max. n° of S**
2	0.77	-0.31	1.32	1.16	1.0	1.5	0.0	3.0
4	1.22	-0.82	1.86	1.64	2.0	3.0	0.0	6.0
6	1.54	-0.88	2.28	2.01	3.0	4.5	0.0	9.0
8	1.81	-0.80	2.63	2.32	4.0	6.0	0.0	12.0
10	2.04	-0.70	2.94	2.60	5.0	7.5	0.0	15.0
12	2.24	-0.62	3.23	2.84	6.0	9.0	0.0	18.0
14	2.43	-0.56	3.48	3.07	7.0	10.5	0.0	21.0
16	2.60	-0.51	3.72	3.28	8.0	12.0	0.0	24.0

I1-I4

	Distance					*Sulfation*		
residues	**Mean**	**Skew**	**Max. D.**	**M.V.D[a]**	**N° of S that M.V.D occurs at**	**Mean**	**Skew**	**Max. n° of S**
2	0.68	-0.42	1.12	0.78	1.0	1.5	0.0	3.0
4	1.05	-0.80	1.58	1.11	2.0	3.0	0.0	6.0
6	1.32	-0.78	1.94	1.35	3.0	4.5	0.0	9.0
8	1.54	-0.69	2.24	1.56	4.0	6.0	0.0	12.0
10	1.74	-0.60	2.50	1.75	5.0	7.5	0.0	15.0
12	1.91	-0.54	2.74	1.92	6.0	9.0	0.0	18.0
14	2.07	-0.49	2.96	2.07	7.0	10.5	0.0	21.0
16	2.21	-0.45	3.16	2.21	8.0	12.0	0.0	24.0

[a] **Maximum Variation in Distance (M.V.D),** D_{max}-D_{min} at a specific level of sulfation

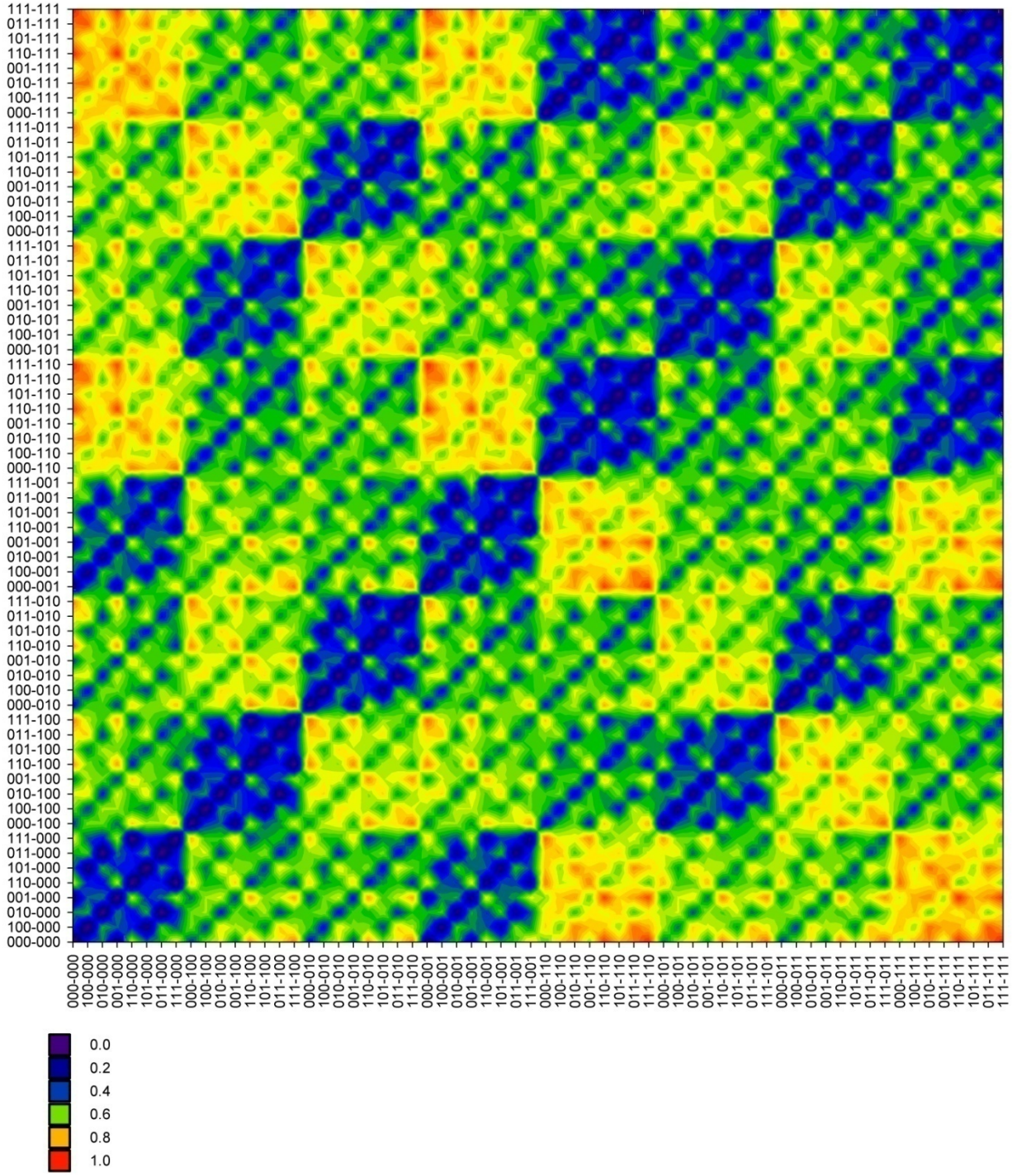


Supplementary Figure 2. Distribution of linkage variation (*r.s.s.* distances) along all possible sequence combinations for; **A.** For stretches of 4 residues defined by I-A-I'-A' (superimposed) and **B.** For stretches of 6 residues defined by I-A-I'-A'-I"-A" (superimposed). The main substitutions influencing the glycosidic linkage positions are also shown in the schematic.

Supplementary Figure 3 Variability of linkage positions in all possible tetrasaccharide stretches at each linkage position. The vertical axis is the relative magnitude of variation in linkage ^{13}C chemical shift values.

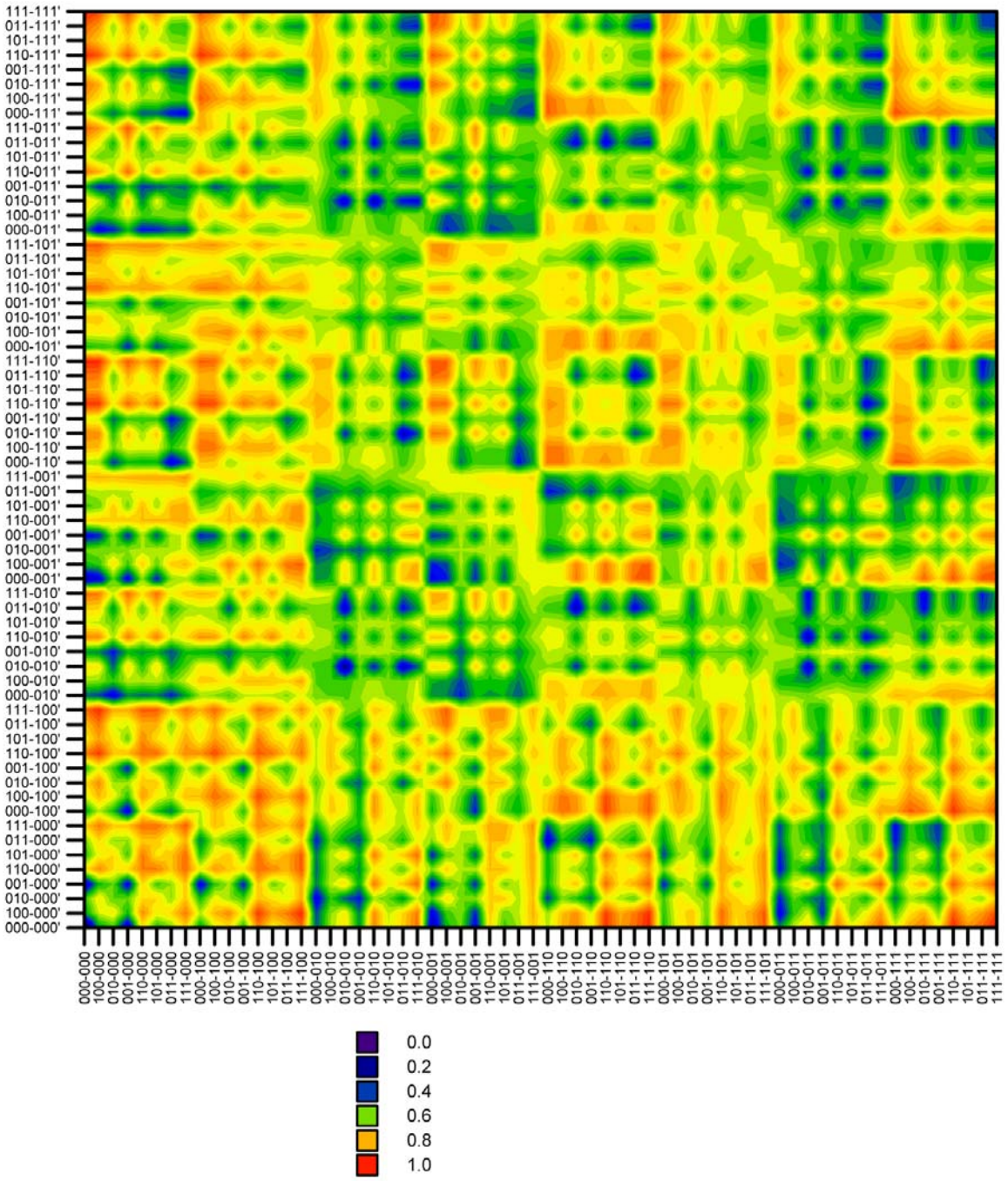
Supplementary Table 5 Breakdown of individual linkage variation for all possible sequences of 4 residues

I4-1		I1-1		A4-1		A1-1		I4-2		I1-2		A4-2		A1-2	
SEQ	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*	*SEQ*	*VAR*
000-XXX	0.0001	**000-XXX**	0.0001	**000-0XX**	0.0001	**000-0XX**	0.0001	**X00-000**	0.0001	**X00-000**	0.0001	**XXX-000**	0.0001	**XXX-000**	0.0001
001-XXX	0.0160	**001-XXX**	0.0080	**000-1XX** **100-0XX**	0.0076	**100-0XX** **000-1XX**	0.0236	**X01-000** **X00-001**	0.0081	**X01-000** **X00-001**	0.0041	**XXX-100**	0.0150	**XXX-001**	0.0430
100-XXX	0.5230	**010-XXX**	0.0180	**100-1XX**	0.0150	**001-0XX**	0.0430	**X10-001**	0.0160	**X01-001**	0.0080	**XXX-010**	0.1920	**XXX-100**	0.0470
101-XXX	0.5232	**011-XXX**	0.0197	**010-0XX**	0.1920	**100-1XX**	0.0470	**X10-000** **X00-010**	0.4196	**X10-000** **X00-010**	0.0091	**XXX-110**	0.1926	**XXX-101**	0.0637
010-XXX	0.8390	**100-XXX**	0.9750	**010-1XX** **110-0XX**	0.1923	**101-0XX** **001-1XX**	0.0534	**X11-000** **X00-011**	0.4196	**X11-000**	0.0099	**XXX-001**	0.5120	**XXX-010**	0.8670
011-XXX	0.8392	**101-XXX**	0.9750	**101-100**	0.1926	**101-1XX**	0.0637	**X01-010** **X10-001**	0.4275	**X01-010** **X10-001**	0.0130	**XXX-101**	0.5122	**XXX-011**	0.8681
110-XXX	0.9887	**110-XXX**	0.9752	**001-0XX**	0.5120	**010-0XX**	0.8670	**X11-001** **X01-011**	0.4276	**X11-001** **X01-011**	0.0138	**XXX-011**	0.5468	**XXX-110**	0.8683
111-XXX	0.9888	**111-XXX**	0.9752	**101-0XX** **001-1XX**	0.5121	**110-0XX** **010-0XX**	0.8676	**X00-100**	0.5230	**X10-010**	0.0180	**XXX-111**	0.5470	**XXX-111**	0.8693
				101-1XX	0.5122	**011-0XX**	0.8681	**X01-100** **X00-101**	0.5231	**X11-010** **X10-011**	0.0188				
				011-0XX	0.5468	**110-1XX**	0.8683	**X01-101**	0.5232	**X11-011**	0.0197				
				111-0XX **011-1XX**	0.5469	**111-0XX** **011-1XX**	0.8687	**X10-100**	0.7558	**X00-100**	0.9750				
				111-1XX	0.5470	**111-1XX**	0.8693	**X11-100** **X00-111**	0.7559	**X01-100** **X00-101**	0.9750				
								X0X-110 **X1X-101**	0.7560	**X01-101**	0.9750				
								X11-101 **X01-111**	0.7560	**X10-100** **X00-110**	0.9751				
								X10-010	0.8390	**X11-100** **X01-110** **X10-101** **X00-111**	0.9751				
								X11-010 **X10-011**	0.8391	**X10-110**	0.9752				
								X11-011	0.8392	**X10-111**	0.9752				
								X10-110	0.9887	**X11-111**	0.9752				
								X11-110 **X10-111**	0.9887						
								X11-111	0.9888						
Min Var	**000-XXX**		**000-XXX**		**000-0XX**		**000-0XX**	**X00-000**		**X00-000**		**XXX-000**		**XXX-000**	
Max Var	**11X-XXX**		**11X-XXX**		**X1X-XXX**		**X1X-XXX**	**X1X-11X**		**XXX-1XX**		**XXX-X11**		**XXX-X1X**	

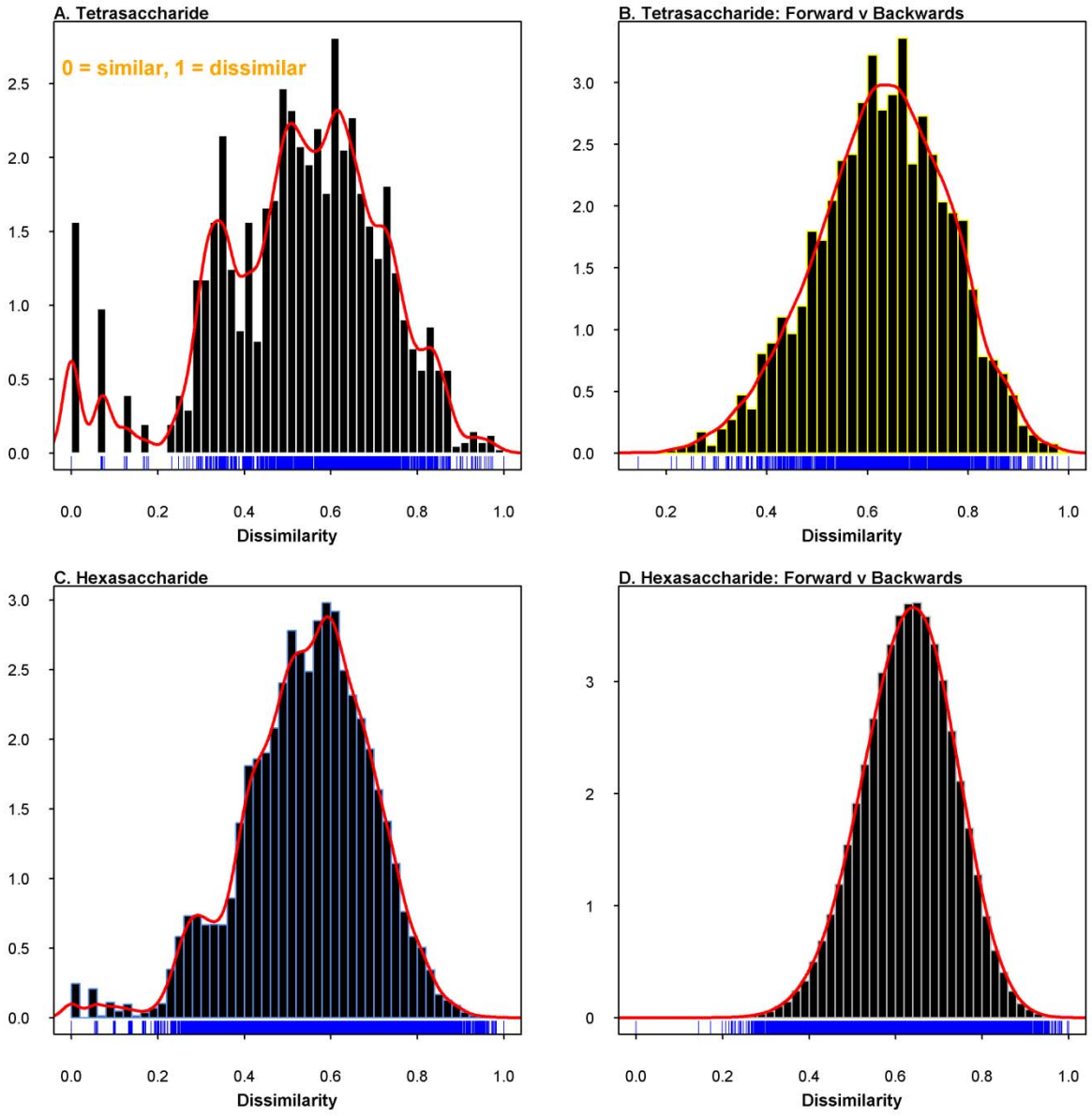


Supplementary Figure 4A Dissimilarities between all stretches of 4 residues running in the same direction. The geometric patterning is the results of the grouping of structures along the axes and systematic changes between members of the groups (diagonal lines). The data are symmetrical about the diagonal [running bottom left to top right], with zero dissimilarity on the diagonal when each sequence is compared with itself. Deep blue-purple

(representing 0.0) signifies complete similarity, while red (representing 1.0) signifies complete dissimilarity.



Supplementary Figure 4B Dissimilarities between all stretches of 4 residues running in opposite directions. The geometric patterning is the result of the grouping of structures along the axes. XXX-XXX' signifies the reverse structures, so the linkage order runs A1-4, A4-4, I1-3, I4-3, A1-2, A4-2, I1-1, I4-1 instead of I4-1, I1-1, A4-2, A1-2, I4-3, I1-3, A4-4, A1-4.



Supplementary Figure 5 Histogram of the dissimilarity values for all stretches of 4 residues and 6 residues running in the same direction and the comparison of forward and backwards measuring dissimilarity along the sequence, rather than overall variation in the linkage.

Dissimilarity Calculation

Dissimilarities were calculated using SPSS [proximity algorithm], which is a comparison of Euclidean distances from an origin between two samples, the values being scaled to the range 0 to 1, (with 0 being similar and 1 disimilar). So, for example, the dissimilarities between two dissaccahrides I4a, I1a, A4a, A1a and I4b, I1b, A4b, A1b is:

$$Da,b = \sqrt{(I4a - I4b)^2 + (I1a - I1b)^2 + (A4a - A4b)^2} + (A1a - A1b)^2$$

In general;

$$D_{ij}^2 = \sum_{v=1}^{n} \left(x_i - x_j\right)^2$$