Toward a Robust Computational Screening Strategy for Identifying Glycosaminoglycan Sequences that Display High Specificity for Target Proteins

*Nehru Viji Sankaranarayanan and Umesh R. Desai**

Department of Medicinal Chemistry and Institute for Structural Biology and Drug Discovery
Virginia Commonwealth University, Richmond, VA 23219

*Address for correspondence: Dr. Umesh R. Desai, 800 E. Leigh Street, Suite 212, Richmond, VA 23219.
Ph (804) 828-7328; Fax (804) 827-3664; urdesai@vcu.edu

SUPPLEMENTARY MATERIAL

Table S1. H/HS hexasaccharide sequences used for algorithm optimization procedure.

(G_{DEF}, G_{EFG} and G_{FGH} hexasaccharide sequences contain DEF, EFG and FGH structures, respectively, of the DEFGH pentasaccharide; G_{USU} sequences contain fewer sulfate group in comparison to G_{DEF}, while G_{SSS} sequences contain more sulfate groups.)

	Hexasaccharide sequences	No. of OSO_3^-		Hexasaccharide sequences	No. of OSO_3^-
The G_{DEF} group					
1	ua2A-YbC6A-ZbB-Yb236A-uc2A-YbH36A	8	33	ucA-YbCA-ZbB-Yb236A-uc2A-Yb26A	6
2	uaA-YbC6A-ZbB-Yb236A-uaA-Yb23A	5	34	Zb2B-Yb2A-uaA-Yb236A-uc2A-Yb26A	8
3	uaA-YbC6A-ZbB-Yb236A-ZbB-YbCA	4	35	ZbB-Yb23A-ucA-Yb236A-uc2A-Yb26A	8
4	uc2A-YbC6A-ZbB-Yb236A-ucA-YbC6A	6	36	ZbB-Yb2A-ZbB-Yb236A-uc2A-Yb26A	7
5	ucA-YbC6A-ZbB-Yb236A-uc2A-Yb26A	7	37	ua2A-Yb23A-uaA-Yb236A-uc2A-Yb26A	9
6	ZbB-YbC6A-ZbB-Yb236A-ua2A-YbCA	5	38	ua2A-Yb2A-ucA-Yb236A-uc2A-Yb26A	8
7	ua2A-Yb26A-ZbB-Yb236A-ua2A-Yb23A	9	39	ua2A-YbCA-ZbB-Yb236A-uc2A-Yb26A	7
8	ua2A-Yb26A-ZbB-Yb236A-Zb2B-Yb23A	9	40	uaA-Yb236A-uaA-Yb236A-uc2A-Yb26A	9
9	uaA-Yb26A-ZbB-Yb236A-uc2A-YbC6A	7	41	uaA-Yb26A-ucA-Yb236A-uc2A-Yb26A	8
10	uc2A-Yb26A-ZbB-Yb236A-ua2A-YbH3A	8	42	uaA-YbC6A-ZbB-Yb236A-uc2A-Yb26A	7
11	uc2A-Yb26A-ZbB-Yb236A-ZbB-Yb2A	7	43	uc2A-Yb26A-uaA-Yb236A-uc2A-Yb26A	9
12	ucA-Yb26A-ZbB-Yb236A-ucA-Yb26A	7	44	uc2A-YbC6A-ucA-Yb236A-uc2A-Yb26A	8
13	Zb2B-Yb26A-ZbB-Yb236A-uaA-YbCA	6	45	uc2A-YbH36A-ZbB-Yb236A-uc2A-Yb26A	9
14	ZbB-Yb26A-ZbB-Yb236A-ua2A-Yb2A	7	**The G_{USU} group**		
15	ZbB-Yb26A-ZbB-Yb236A-Zb2B-Yb2A	7	46	ZbB-YbCA-ZbB-Yb2A-uc2A-YbH36A	4
The G_{EFG} group			47	ZbB-YbHA-ZbB-YbHA-ZbB-YbHA	0
16	ucA-Yb23A-ZbB-Yb236A-uc2A-YbH3A	7	48	ZbB-YbHA-ZbB-YbHA-ZbB-YbC6A	1
17	ucA-YbC6A-ZbB-Yb236A-uc2A-Yb2A	6	49	ZbB-YbHA-ZbB-YbCA-ZbB-YbHA	0
18	Zb2B-Yb23A-ZbB-Yb236A-uc2A-YbH36A	9	50	ZbB-YbHA-ZbB-YbC6A-ZbB-YbHA	0
19	ZbB-Yb236A-ZbB-Yb236A-uc2A-Yb26A	9	51	ZbB-YbHA-ZbB-Yb2A-ZbB-YbHA	1
20	ZbB-Yb26A-ZbB-Yb236A-uc2A-YbCA	6	52	ZbB-YbCA-ZbB-YbHA-ZbB-YbHA	0
21	ZbB-YbCA-ZbB-Yb236A-uc2A-Yb23A	6	53	ZbB-YbCA-ZbB-YbCA-ZbB-Yb236A	3
22	ua2A-Yb23A-ZbB-Yb236A-uc2A-Yb23A	9	54	ZbB-YbCA-ZbB-YbC6A-Zb2B-Yb26A	4
23	ua2A-Yb2A-ZbB-Yb236A-uc2A-YbC6A	7	55	ZbB-YbCA-ZbB-YbC6A-ZbB-Yb26A	3
24	ua2A-YbCA-ZbB-Yb236A-uc2A-YbH3A	6	**The G_{SSS} group**		
25	uaA-Yb236A-ZbB-Yb236A-uc2A-Yb2A	8	56	uc2A-YbH36A-uc2A-YbC6A-ZbB-Yb236A	8
26	uaA-Yb26A-ZbB-Yb236A-uc2A-YbH36A	8	57	ZbB-Yb236A-ZbB-Yb236A-ua2A-Yb23A	9
27	uaA-YbCA-ZbB-Yb236A-uc2A-Yb26A	6	58	ZbB-Yb236A-Zb2B-Yb2A-uaA-YbCA	5
28	uc2A-Yb26A-ZbB-Yb236A-uc2A-YbCA	7	59	ZbB-Yb236A-Zb2B-Yb26A-ZbB-Yb236A	9
29	uc2A-YbCA-ZbB-Yb236A-uc2A-Yb23A	7	60	ZbB-Yb236A-ucA-Yb236A-ZbB-Yb236A	9
30	uc2A-YbH3A-ZbB-Yb236A-uc2A-YbC6A	7	61	ZbB-Yb236A-ucA-Yb236A-uc2A-YbH36A	9
The G_{FGH} group			62	ZbB-Yb236A-uc2A-YbH36A-ua2A-YbH36A	9
31	ucA-Yb23A-uaA-Yb236A-uc2A-Yb26A	8	63	Zb2B-Yb26A-uc2A-Yb26A-Zb2B-Yb26A	9
32	ucA-Yb2A-ucA-Yb236A-uc2A-Yb26A	7	64	Zb2B-Yb23A-uc2A-Yb26A-Zb2B-Yb23A	9
			65	ucA-YbC6A-ZbB-Yb2A-uaA-Yb236A	5

Table S2. H/HS tetrasaccharide sequences from two libraries of 1,444 sequences each that satisfied the dual filter CVLS strategy.

#	Tetrasaccharide Sequence[a]	GoldScore[b]	# of H-bonds[c]
UA$_{NRE}$ Library			
1	uc2A-Yb236A-ZbB-Yb23A	110.49	11
2	ua2A-Yb26A-Zb2B-Yb23A	110.32	9
3	uaA-Yb26A-Zb2B-Yb23A	110.10	9
4	ucA-Yb26A-Zb2B-Yb23A	107.53	10
5	ua2A-YbC6A-ZbB-Yb236A	107.14	10
6	uc2A-Yb26A-Zb2B-Yb23A	106.26	11
7	uc2A-YbC6A-ZbB-Yb23A	103.92	12
8	ua2A-YbC6A-Zb2B-Yb23A	103.76	10
9	ucA-YbC6A-Zb2B-Yb23A	103.60	10
10	ucA-YbC6A-ZbB-Yb23A	103.15	11
11	ua2A-YbH36A-ZbB-Yb236A	101.79	8
12	uc2A-YbH36A-ucA-Yb26A	101.03	13
13	ZbB-Yb236A-ZbB-Yb236A	100.96	10
14	ZbB-Yb26A-ZbB-Yb236A	100.16	11
GlcN$_{NRE}$ Library			
1	Yb26A-ZbB-Yb23A-ZbB	116.96	12
2	Yb26A-ZbB-Yb236A-ZbB	110.41	11
3	YbC6A-ZbB-Yb23A-Zb2B	110.40	11
4	YbC6A-ZbB-Yb236A-ucA	109.82	14
5	YbC6A-ZbB-Yb236A-uc2A	108.51	11
6	YbC6A-ZbB-Yb236A-ZbB	107.56	10
7	Yb26A-ZbB-Yb23A-ucA	106.62	16
8	Yb2A-ZbB-Yb23A-uc2A	105.39	10
9	Yb26A-ZbB-YbH3A-uc2A	104.26	10
10	Yb2A-Zb2B-Yb236A-uc2A	103.75	9
11	Yb23A-ZbB-Yb23A-uc2A	103.41	10
12	Yb23A-ZbB-Yb23A-ucA	103.01	12
13	Yb23A-Zb2B-Yb236A-uc2A	102.62	12
14	Yb26A-Zb2B-Yb236A-ZbB	99.28	11
15	YbC6A-ZbB-Yb23A-uc2A	98.38	11
16	YbCA-ZbB-Yb23A-uc2A	97.68	9

[a] See definitions of residue labels and substitution in Table 1. [b] Refers to modified GoldScore, as defined in the Methods section. [c] Number of hydrogen bonds calculated using LIGPLOT.

Table S3. H/HS disaccharide sequences from two libraries of 38 sequences each that satisfied the CVLS strategy.

#	Disaccharide Sequence[a]	Gold Score[b]	# of H-bonds[c]
$\mathbf{GlcN_{NRE}}$ Library			
	None		
$\mathbf{UA_{NRE}}$ Library			
1	Zb2B-Yb23A	71.18	8
2	ZbB-Yb236A	67.22	7
3	ZbB-Yb23A	68.58	8

[a] See definitions of residue labels and substitution in Table 1. [b] Refers to modified GoldScore, as defined in the Methods section. [c] Number of hydrogen bonds calculated using LIGPLOT.

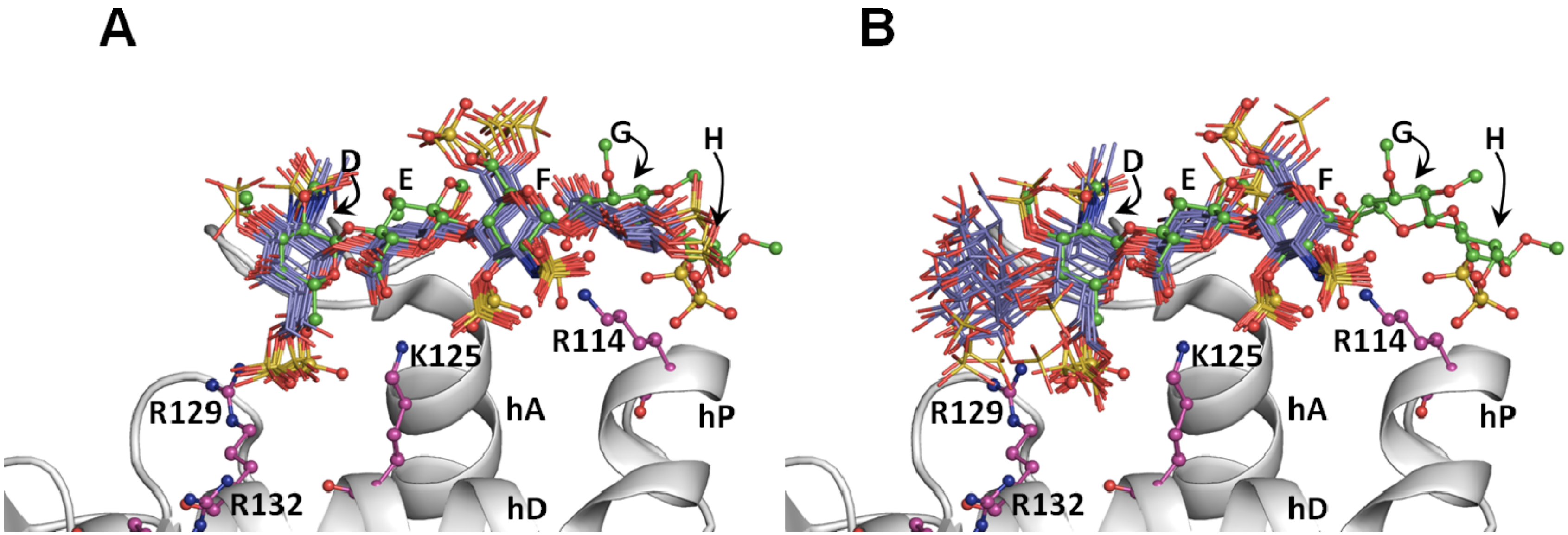

Figure S1. CVLS predicted tetrasaccharide sequences from the GlcN_{NRE} (A) and UA_{NRE} libraries (B) containing 1,444 sequences each. Shown are overlays of the docked poses of tetrasaccharide sequences that bind AT with 'high specificity' by satisfying the dual filter strategy. A) shows 16 sequences (blue sticks) from the GlcN_{NRE} library and B) shows 14 sequences (blue sticks) from UA_{NRE} tetrasaccharide library. Helices A (hA), D (hD) and P (hP) of antithrombin are shown in ribbon form and residues Arg132, Arg129, Lys125 and Arg114 are shown in ball and stick display. The crystal structure of DEFGH in green ball and sticks display is shown to highlight correspondence with the CVLS predicted poses.

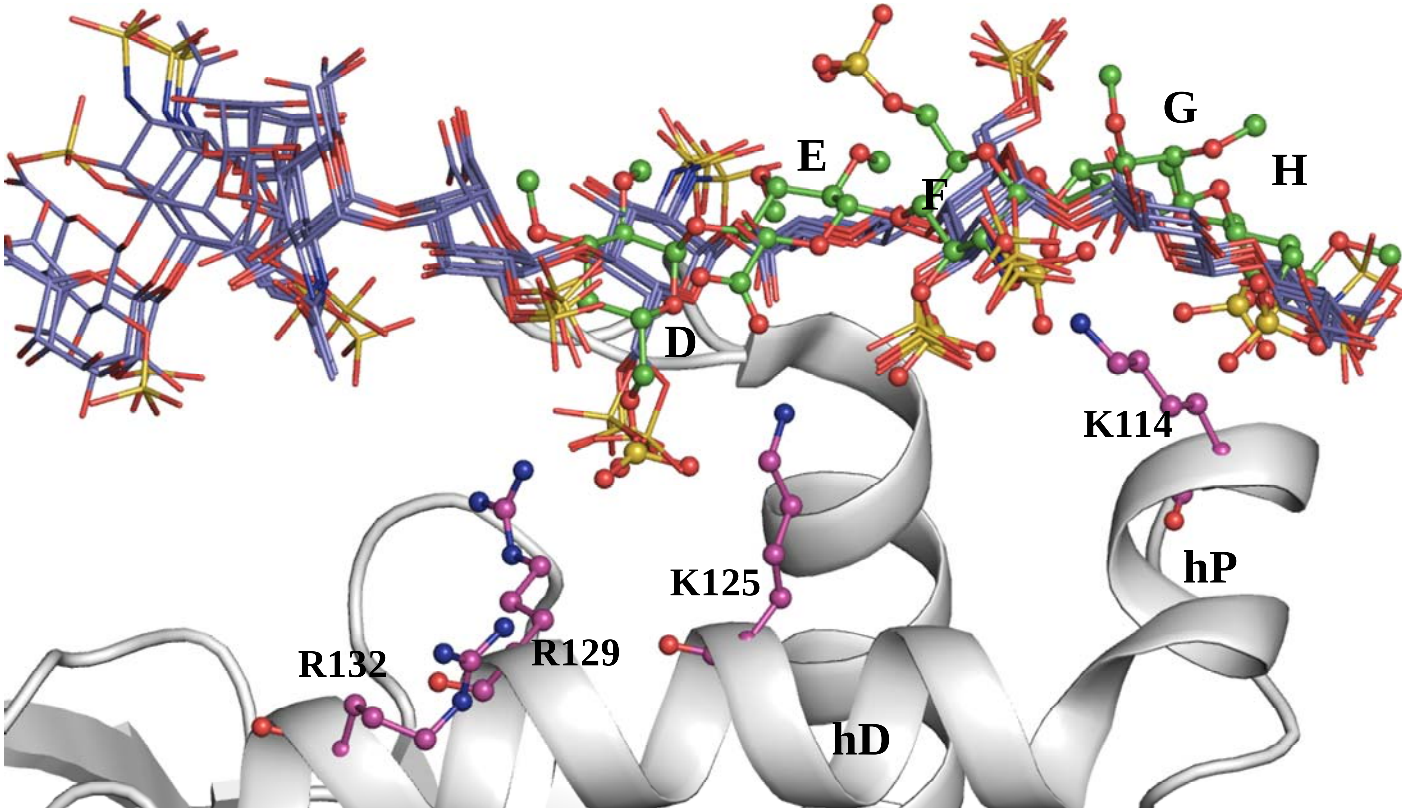

Figure S2. CVLS predicted octasaccharide sequences from the UA_{NRE}. Six sequences satisfied the CVLS dual filter criteria. Helices A (hA), D (hD) and P (hP) of antithrombin are shown in ribbon form and residues Arg132, Arg129, Lys125 and Arg114 are shown in ball and stick display. The crystal structure of DEFGH in green ball and sticks display is shown to highlight correspondence with the CVLS predicted poses.