Sulfoglycodendrimer Therapeutics for HIV-1 and SARS-CoV-2

Lauren Wells, Cory Vierra, Janee’ Hardman, Yanxiao Han, Dustin Dimas, Lucia Gwarada-Phillips, Rachel Blackeye, Daryl K. Eggers, Celia C. LaBranche, Petr Král*, Katherine D. McReynolds*

Supporting Information

Contents:

NMR Data Interpretation   S2
NMR Data:                 S3-S39
MST Raw Data:             S40
MD Data:                  S41-S55
References:               S54
NMR Data Interpretation: NMR data was collected on a 500 MHz Bruker Avance III spectrometer. All NMR peaks are referenced relative to NMR residual solvent peaks and all $^{13}\text{C}$ data collected in D$_2$O are referenced to an internal methanol standard. For all hexavalent products, the integrations were normalized to 1/3 of the molecule. Additionally, for all oxime containing products, Compounds 5-18e, it should be noted that while there are two primary stereoisomers of the open chain oxime product, the $E$ and the $Z$ isomers, there are also ring-closed $\alpha$- and $\beta$-anomer products possible (Figure S1). This explains the lower than anticipated sum (<1 proton for the sugar linkers and <2 (for 1/3 of the molecule) protons for the GD and SGD hexavalent molecules) for the $E$ and $Z$ isomer peaks observed between 7-8 ppm and the presence of small/fractional peaks in the CHOH regions of the $^1\text{H}$ spectra. The % open/closed can be estimated using the integrals for the $E$ and $Z$ isomers (Table S1). It should also be noted that when exposed to protic solvents, the isomeric distribution will change with time.

![Diagram](image)

Figure S1. Oxime major and minor products formed.

Table S1. Estimation of % open and closed rings on reducing end sugar oximes.

| Compound | $E + Z$ integral sum | $E$ to $Z$ Ratio | % Open Oxime ($E/Z$ isomers) | % Closed Glycoside ($\alpha/\beta$-anomers) |
|----------|----------------------|------------------|-----------------------------|------------------------------------------|
| 5        | 0.26 + 0.05 = 0.31   | 5.2:1            | 31%                         | 69%                                      |
| 6        | 0.20 + 0.07 = 0.27   | 2.9:1            | 27%                         | 73%                                      |
| 7        | 0.57 + 0.12 = 0.69   | 4.8:1            | 69%                         | 31%                                      |
| 8        | 0.45 + 0.10 = 0.55   | 4.5:1            | 55%                         | 45%                                      |
| 9        | 0.48 + 0.12 = 0.60   | 4.0:1            | 60%                         | 40%                                      |
| 10       | 0.42 + 0.11 = 0.53   | 3.8:1            | 53%                         | 47%                                      |
| 11       | 0.51 + 0.09 = 0.60   | 5.7:1            | 60%                         | 40%                                      |
| 12d      | 0.88 + 0.16 = 1.04   | 5.5:1            | 52%                         | 48%                                      |
| 13d      | 0.77 + 0.24 = 1.01   | 3.2:1            | 51%                         | 49%                                      |
| 14d      | 0.90 + 0.17 = 1.07   | 5.3:1            | 54%                         | 46%                                      |
| 15d      | 0.72 + 0.14 = 0.86   | 5.1:1            | 43%                         | 57%                                      |
| 16d      | 0.61 + 0.09 = 0.70   | 6.8:1            | 35%                         | 65%                                      |
| 17d      | 0.91 + 0.22 = 1.13   | 4.1:1            | 57%                         | 43%                                      |
| 18d      | 1.03 + 0.18 = 1.21   | 5.7:1            | 61%                         | 39%                                      |
NMR Data:

Figure S2: $^1$H NMR (500 MHz, D$_2$O) of Compound 4.
Figure S3: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 4.
Figure S4: $^1$H NMR (500 MHz, D$_2$O) of Compound 5.
Figure S5: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 5.
Figure S6: $^1$H NMR (500 MHz, D$_2$O) of Compound 6.
Figure S7: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 6.
Figure S8: $^1$H NMR (500 MHz, D$_2$O) of Compound 7.
Figure S9: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 7.
Figure S10: $^1$H NMR (500 MHz, D$_2$O) of Compound 8.
Figure S11: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 8.
Figure S12: $^1$H NMR (500 MHz, D$_2$O) of Compound 9.
Figure S13: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 9.
Figure S14: $^1$H NMR (500 MHz, D$_2$O) of Compound 10.
Figure S15: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 10.
Figure S16: $^1$H NMR (500 MHz, D$_2$O) of Compound 11.
Figure S17: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 11.
Figure S18: $^1$H NMR (500 MHz, D$_2$O) of Compound 12d.
Figure S19: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 12d.
Figure S20: $^1$H NMR (500 MHz, D$_2$O) of Compound 13d.
Figure S21: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 13d.
Figure S22: $^1$H NMR (500 MHz, D$_2$O) of Compound 14d.
Figure S23: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 14d.
Figure S24: $^1$H NMR (500 MHz, D$_2$O) of Compound 15d.
Figure S25: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 15d.
Figure S26: $^1$H NMR (500 MHz, D$_2$O) of Compound 16d.
Figure S27: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 16d.
Figure S28: $^1$H NMR (500 MHz, D$_2$O) of Compound 17d.
Figure S29: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 17d.
Figure S30: $^1$H NMR (500 MHz, D$_2$O) of Compound 18d.
Figure S31: $^{13}$C NMR (125 MHz, D$_2$O with internal methanol std.) of Compound 18d.
Figure S32: $^1$H NMR (500 MHz, D$_2$O) of Compound 12e.
Figure S33: $^1$H NMR (500 MHz, D$_2$O) of Compound 13e.
Figure S34: $^1$H NMR (500 MHz, D$_2$O) of Compound 14e.
Figure S35: $^1$H NMR (500 MHz, D$_2$O) of Compound 15e.
Figure S36: $^1$H NMR (500 MHz, D$_2$O) of Compound 16e.
Figure S37: $^1$H NMR (500 MHz, D$_2$O) of Compound 17e.
Figure S38: $^1$H NMR (500 MHz, D$_2$O) of Compound 18e.
Figure S39. Sample data from MST trial with lactose SGD. A. Initial scan of 12 capillary tubes indicating nearly equal amounts of labeled gp120 (constant fluorescence) and no protein adsorption to capillary surface (sharp symmetrical peaks). B. Thermophoresis responses for capillary tubes in panel A containing serial dilution of lactose SGD. Arrow on graph indicates order of tubes from #1-12. C. Final binding curve for lactose SGD; error bars denote error of precision for three trials. See Experimental Methods for more details.
MD Data:

Binding of SGDs to gp120:

Figures S39-S40 show the final snapshots of SGDs and GDs binding with the gp120 trimer. Initially, six SGDs/GDs were put near the V3 loops in each system. However, not all the SGDs/GDs stayed bound to gp120, since some of the SGDs/GDs left its surface. Figure S39 shows that some of the remaining SGDs hang in the grooves of the gp120 monomers, while others sit directly on the V3 loops.

![Diagram of binding of SGDs to gp120](image)

Figure S40. Binding of SGDs with the gp120 trimer. A. Cellobiose 12e (148 ns); B. Lactose 15e (189 ns); C. Maltotriose 17e (150 ns); D. Melibiose 18e (190 ns). Only SGDs within 3 Å of gp120 are shown. Proteins are shown in cartoon presentation, and SGDs are shown in atomic detail. The V3 loop is colored in purple; The atom coloring scheme: C - grey, O - red, S - yellow, N - blue, H - omitted. Water and ions are omitted for better visualization.

In comparison to SGDs (Figure S40), more GDs leave the surface of the gp120, and those that remain attached show a rather small affinity to the V3 loops. The number of GDs binding to V3 loops was quantified in Figure S42. Among the four SGDs, Cellobiose 12e shows a slightly higher number of attached compounds (4-5), while Lactose 15e, Maltotriose 17e, and Melibiose 18e show similar number of attached compounds (2-3). For dendrimers with the same sugar terminus, there
are more SGDs interacting with V3 than GD, especially, the ratio of SGD to GD (number of interacting compounds) is about 4:1 in the case of Cellobiose 12e and 2:1 in Maltotriose 17e. In general, SGDs show higher affinity to the V3 loop than GDs, because of the enhanced Coulombic interactions.

Figure S41. Binding of GDs with the gp120 trimer. A. Cellobiose (100 ns); B. Lactose (120 ns); C. Maltotriose (100 ns); D. Melibiose (130 ns). Only SGDs within 3 Å of gp120 are shown. Proteins are shown in cartoon presentation, and SGDs are shown in atomic detail. The V3 loop is colored in purple; The atom coloring scheme: C - grey, O - red, S - yellow, N - blue, H - omitted. Water and ions are omitted for better visualization.

In the following, we focus on analysis of the interaction between SGDs and gp120. The contact/interaction is defined by a cutoff distance, chosen to be 3 Å. If the amino acids were within 3 Å of SGD/GD, we considered that they were interacting with each other, and vice versa. Amino acids of the V3 loops interacting with SGD/GDs were quantified in Figures S42-S43. The interacting sugar units of the SGDs with V3 loops were also analyzed in Figure S44. Since the gp120 protein is a trimer, the analysis in Figures S42-S44 was conducted by averaging the interactions over three monomers.
Figure S42. Number of SGD/GDs contacting with the V3 loop. A. Cellobiose 12e; B. Lactose 15e; C. Maltotriose 17e; D. Melibiose 18e. Red: SGDs; green: GDs.
Analysis of interacting residues of V3 loop towards SGD/GDs:

Figure S43. The contact times of amino acids from the V3 loop (per monomer) interacting with SGDs, averaged over the last 500 frames (20 ns). A. Cellobiose 12e; B. Lactose 15e; C. Maltotriose 17e; D. Melibiose 18e.

Figure S43 shows the contact times of amino acids of V3 loop interacting with different SGDs. The contact time is defined as the number of frames out of the last 500, where the interaction or contact happens. The two most frequent amino acids in the binding of 12e/17e to V3 loop are ARG and ILE, while in the case of 15e, they are ARG and ALA; in the case of 17e, they are ARG and LYS. The frequently interacting amino acids among the four SGDs are slightly different, but they stand for the two major amino acids categories, basic and hydrophobic. Among the four SGDs, ARG is the common amino acid which contacts for more than half of the counting time (250). The long binding time of amino acids arises from the strong Coulombic interaction between the positively charged amino acids and the sulfate groups of SGDs.
Figure S44. The contact times of amino acids of the V3 loop (per monomer) interacting with GDs, averaged over the last 500 frames (20 ns). A. Cellobiose 12d; B. Lactose 15d; C. Maltotriose 17d; D. Melibiose 18d.

As negative controls, the number of interacting amino acids with GDs are far less than SGDs. The contact times of the amino acids interacting with GDs are less dense than their SGDs partners as shown in Figure S44. As the GDs are neutral, the charged interaction is impaired in the whole interaction, which could significantly weaken the binding strength of GDs to V3 loops as discussed in the SGD cases.
Analysis of SGDs interacting with V3 loops:

**Figure S45.** Contact times of sugar groups of SGDs with the V3 loops. A. Cellobiose 12e; B. Lactose 15e; C. Maltotriose 17e; D. Melibiose 18e. Naming rule: the ring close to compound core is named as ring 1, the ring far from compound core is named as ring 2. The atoms are named by the element symbol, followed by the ring name, then normal naming number for ring structure, e.g. C12 means carbon in ring 1 at position 2; C23 means carbon in ring 2 at position 3; H1O3 means hydrogen in ring 1 in hydroxyl group at position 3; H2O2 means hydrogen in ring 2 in hydroxyl group at position 2.

**Figure S45** shows the contact times of atoms in sugar units interacting with the V3 loop averaged over 36 branches in the cases of SGDs. For 12e, the total contacting times are 213 for the second glucose, and 132 for the first glucose. For 15e, there are 117 contacting times for the first sugar ring (glucose), and 113 contacting times for the second sugar ring (galactose). For 18e, there are 139 contacting times for the first sugar ring (glucose) and 84 contacting times for the second sugar ring (galactose). For 17e, there are 180 contacting times for the third glucose, 66 contacting times...
for the second glucose, and 63 contacting times for the first glucose. Note here, not all the branches bind to V3, the numbers obtained above cannot be used directly to compare among different SGDs. We further defined a term, sugar ring attaching factor, to make the data comparable among different cases. Sugar ring attaching factor equals the difference of contacting times between sugar ring 1 and 2 divided by the total contacting time of ring 1 and 2. A smaller value of sugar ring attaching factor corresponds to a higher potential for the multivalent binding. For example, the sugar ring attaching factor of 12e equals (213-132)/(213+132), which gives 0.23. Similarly, the sugar ring attaching factor is 0.02 for 15e and 0.25 for 18e. As 17e has three sugar units with the ring 1 and ring 2 presenting very similar contacting times, we consider ring 1 and ring 2 as one unit which contribute 129 contacting times. So, the sugar ring attaching factor of 17e is 0.17.

Collectively, we find 15e shows the best sugar ring attaching factor (0.02) followed by 17e (0.17), 12e (0.23), then, 18e (0.25). Only 15e with the lowest sugar ring attaching factor shows comparable contacting time for both sugar units, which indicates a more ideal multivalent interaction against V3 loop; while the rest of the SGDs favor the terminal sugar binding style which is a sign of weaker multivalency. Given that 15e and 12e with similar sulfates are different by the terminal sugars, we propose that galactose (terminal unit of 15e, with three OH groups toward one side of the ring) could provide a more polar side concentrated with more OH groups. The concentrated OH groups in galactose unit of 15e could increase the polar interactions between SGD and V3 loop.

As a positive control (Figure S46), dextran sulfate interacts with V3 loops by a relative stable number of amino acids (5) over the last 30 ns. This can be attributed to the large molecular weight of dextran sulfate, which makes it less mobile.

![Figure S46](image)

**Figure S46.** The total number of residues of the V3 loops in contact with the dextran sulfate (molecular weight 40368 g/mol).
Application of GDs in binding with RBD of SARS-CoV-2:

Figures S47-S53 show the binding modes of SGD/GD and HGD (hybrid structures with ½ of the GD molecule unsulfated and half sulfated) with the RBD of SARS-CoV-2. Table S2 summarizes the information for each kind of compound. SGDs bind to both the top and middle part of the RBD, while GDs only bind to the top region of the RBD. SGDs bind to the middle part of the RBD due to the charge interaction between the sulfate groups and the basic amino acids of the RBD. GDs only binds to the top part of RBD, due to the lack of sulfate groups.

Figure S47. Lactose SGD binding with RBD of SARS-CoV-2 (less favorable binding modes). The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H - omitted, interacting ARG - pink, LYS - grey. Water and ions are omitted for better visualization (two binding modes A and B).
Figure S48. Cellobiose SGD binding with RBD of SARS-CoV-2. The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H - omitted, interacting ARG - pink, LYS - grey. Water and ions are omitted for better visualization (four binding modes A, B, C and D).
Figure S49. Cellobiose GD binding with RBD of SARS-CoV-2. The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H – omitted. Water and ions are omitted for better visualization (two binding modes A and B).

Figure S50. Lactose GD binding with RBD of SARS-CoV-2. The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H - omitted, interacting ARG- pink, LYS- grey. Water and ions are omitted for better visualization (two binding modes A and B).
**Figure S51.** Lactose HSG and Cellobiose HGD.

**Figure S52.** Cellobiose HGD binding with RBD of SARS-CoV-2. The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H - omitted, interacting ARG- pink, LYS- grey. Water and ions are omitted for better visualization (two binding modes A and B).
Figure S53. Lactose HGD binding with RBD of SARS-CoV-2. The coloring scheme: C - cyan, O - red, S - yellow, N - blue, H - omitted, interacting ARG- pink, LYS- grey. Water and ions are omitted for better visualization (two binding modes A and B).

Table S2. Summary of GDs binding to RBD of SARS-CoV-2.

|                  | Cellobiose SGD | Lactose SGD | Cellobiose GD | Lactose GD |
|------------------|----------------|-------------|---------------|------------|
| Bind to top      | YES            | YES         | YES           | YES        |
| Bind to middle   | YES            | YES         | NO            | NO         |

The SGDs bind with RBD either to the top or to the middle, however, GDs only target the top region of RBD. The SGDs and GDs are not observed to bind the top and middle of RBD simultaneously.

Figure S54. Number of amino acids of RBD interacting with GDs. Naming rule: C - Cellobiose; L - Lactose; H - HGD; C-GD-top: Cellobiose GD binding to the top region of RBD; C-SGD-mid: Cellobiose SGD binding to the middle region of RBD; 1: the first trial; 2: the second trial.
Figure S54 shows the number of amino acids of the RBD interacting with GDs. The octavalent HGDs were not always the best GDs for interacting with residues of the RBD. To further analyze the binding affinity of HGDs to RBD, the percentage of different interactions was shown in Figure S55, and the free energy of binding was calculated as shown in Figure S56.

Figure S55. Percentage of different interactions contributing to the binding between HGDs and RBD of SARS-CoV-2. Polar: interaction with polar amino acids; hydrophobic: interaction with hydrophobic amino acids; charge: interaction with positively charged amino acids. C - Cellobiose; L - Lactose; H - hybrid.
As the sulfation level of C-H (Cellobiose hybrid glycodendrimers) and L-H (Lactose hybrid glycodendrimers) are the same during the computational design, the percentages of charged interactions are very similar in the two cases as shown in Figure S55, as well as the free energy of binding shown in Figure S56.

In summary, the best SGD (15e) in binding with gp120 also presents promising binding affinity towards SARS-CoV-2. The binding of SGDs against SARS-CoV-2 is also largely determined by the number of sulfate groups in SGDs as analyzed in the case of gp120. Although, HGDs could not provide a significantly higher binding affinity towards RBD than SGDs, they could target both the top and middle part of RBD. The design of the HGDs could be improved by increasing the sulfation level to strengthen the Columbic interaction or increasing the overall valency.