Carbohydrate Research 518 (2022) 108574

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Binding of synthetic carbohydrate receptors to enveloped virus glycans: Insights from molecular dynamics simulations

Beicer Tapia [a,b], Genrietta Yagudayeva [a], M. Fernando Bravo [a], Khushabu Thakur [d], Adam B. Braunschweig [a,b,c,d,**], Mateusz Marianski [a,b,c,d,*]

[a] Department of Chemistry and Biochemistry, Hunter College, The City University of New York, 695 Park Ave, New York, NY, 10065, USA
[b] The PhD Program in Biochemistry, The Graduate Center of The City University of New York, 365 5th Ave, New York, NY, 10016, USA
[c] The PhD Program in Chemistry, The Graduate Center of The City University of New York, 365 5th Ave, New York, NY, 10016, USA
[d] Advanced Science Research Center of The City University of New York, 85 Nicolas Terrace, New York, NY, 10031, USA

ARTICLE INFO

Keywords:
Antiviral agents
Glycans
Viruses
Molecular dynamics simulations
Supramolecular chemistry

ABSTRACT

Can envelope glycans be targeted to stop viral pandemics? Here we address this question by using molecular dynamics simulations to study the binding between 10 synthetic carbohydrate receptors (SCRs) and the 33 *N*-glycans most commonly found on the surfaces of enveloped viruses, including Zika virus and SARS-CoV-2. Based on association quotients derived from these simulations, we classified the SCRs as weak binders, promiscuous binders, or selective binders. The SCRs almost exclusively associate at the $Man_3GlcNAc_2$ core, which is common to all *N*-glycans, but the binding affinity between the SCR·glycan pair depends on the noncovalent interactions between the heterocycle rings and the glycan antennae. Systematic variations in the glycan and SCR structures reveal relationships that could guide the design of SCRs to attain affinity and selectivity towards a chosen envelope glycan target. With these results, envelope glycans, which are currently considered "undruggable", could become viable targets for new therapeutic strategies.

1. Introduction

Enveloped viruses (EnV) – viruses surrounded by a glycosylated lipoprotein bilayer envelope [1] – include coronaviruses, retroviruses, flaviviruses, bunyaviruses, alphaviruses, togaviruses, filoviruses, and others [2]. Collectively, they are responsible for many recent health crises, including the HIV epidemic [3], the Zika virus (ZIKV) outbreak in 2016 [4], and the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic [5]. Common strategies for antiviral drug development include entry inhibition, RNA-dependent RNA polymerase inhibitors, protease inhibitors, and others [6], but none of these have resulted in FDA-approved treatments for most flaviviridae or coronaviruses, including SARS-CoV-2 or ZIKV, and only two have Emergency Use Authorization by FDA [7]. Instead treatments focus on repurposing already approved drugs, monoclonal antibodies, or palliative care [8,9]. Given the limitations of current drug design strategies to mitigate the intensity of the SARS-CoV-2 pandemic, or to protect from present or future EnV threats, there is no alternative than to consider new antiviral drug design strategies.

EnV glycoproteins, occurring as membrane-anchored peplomers or "spikes", bind receptors on the surface of the host cell, and this step precedes virus entry and viral replication [10–13]. Further, these glycoproteins are densely decorated with *N*-glycans and *O*-glycans which constitute up to 25% of their molecular weight [14]. The *N*-glycans, which share the $Man_3GlcNAc_2$ core sequence, are involved in a range of essential processes, including viral protein folding, evasion of the host's immune system, and facilitating attachment to the host receptors [15]. For instance, selective binding of HIV-1 and Ebola viruses to the host cell is facilitated by an oligomannose *N*-glycan mediating interactions with the DC-SIGN receptor [15], and the EnV glycans of SARS-CoV-2 are known to stabilize the open state of the spike protein' [16–18]. Thus, synthetic molecules that selectively bind EnV glycans and disrupt these processes could act as broad spectrum antivirals (BSAs) - agents that target a wide-range of viruses and which could be deployed immediately to mitigate the threat of the infection [13]. The challenge with this strategy, however, is that glycans are considered "undruggable targets",

* Corresponding author. Department of Chemistry and Biochemistry, Hunter College, The City University of New York, 695 Park Ave, New York, NY, 10065, USA.
** Corresponding author. Advanced Science Research Center of The City University of New York, 85 Nicolas Terrace, New York, NY, 10031, USA.
E-mail address: mmarians@hunter.cuny.edu (M. Marianski).

https://doi.org/10.1016/j.carres.2022.108574
Received 3 December 2021; Received in revised form 15 April 2022; Accepted 22 April 2022
Available online 8 May 2022

meaning they have a known role in disease progression, but no widely adopted therapeutic strategies exploit this information [19]. Although natural lectins and antibodies recognize specific glycans, their toxicity [20] has limited their use as therapeutic or imaging agents [21]. There is only one glycan-targeting antibody that has received FDA approval – a drug of last resort for high-risk childhood neuroblastoma [22]– and there are no approved antiviral treatments whose mechanism involves binding EnV glycans.

To challenge this *status quo* and redefine glycans as feasible targets for therapeutics and chemosensors, several research groups [23–25] have developed synthetic carbohydrate receptors (SCRs) – small molecules that bind carbohydrates through noncovalent interactions. Inspired by the structure and binding thermodynamics of natural glycan binding proteins, we have recently reported a series of flexible, tetrapodal SCRs based upon a biaryl core, that are selective for non-glucosides [13,26–30], and some of these tetrapodal SCRs display nanomolar inhibition against live ZIKV, with data suggesting that the mode of action is the prevention of viral attachment/binding/entry [13,28]. Despite these encouraging results, without a molecular-level understanding of the binding of the glycans by SCRs, their structures cannot be rationally redesigned to achieve strong and selective binding towards EnV glycans. Fine-tuning the specificity of the SCRs towards selected EnV glycans would (1) improve the inhibition of the viral docking to the host receptor, thus reducing the chance of the infection, and (2) decrease the likelihood of off-site toxicity caused by binding to the host *N*-glycans. In this study, we report detailed molecular dynamics (MD) simulations analysis of the binding between 10 tetrapodal SCRs and 33 *N*-glycans common to surfaces of ZIKV [31], MERS-CoV, SARS-CoV, and SARS-CoV-2 [32–34]. These simulations capture how small modifications in *N*-glycan or SCR structures alter the binding geometries and association strengths, and provide a molecular-level rationale for the differences in the observed antiviral activity of the SCRs, which can guide the design of the next generation of EnV glycan binding-receptors.

2. Results and discussion

The SCRs and EnV glycans investigated in this work are shown in Fig. 1. We selected three SCRs (**SCR001**, **SCR005** and **SCR007**) that have displayed potent antiviral activity against live ZIKV [28] and seven recently synthesized SCRs (**SCR007**, **SCR018**, **SCR019**, **SCR020**, **SCR021**, **SCR022**, **SCR023**) that have shown high selectivity towards biologically relevant monosaccharides [29]. All tetrapodal SCRs share the same chemical design: a biaryl core decorated with four linkers terminated with a heterocycle (Fig. 1A) [26]. The investigated SCRs consist of four pairs of regioisomers, which differ by the point-of-attachment of the heterocyclic ring, and have pyrrole (**SCR001** and **SCR017**), indole (**SCR018** and **SCR019**), pyridine (**SCR020** and **SCR021**) or phenol (**SCR022** and **SCR023**) groups attached to the amine-containing linker. The other two receptors have 2-furan (**SCR005**) or 2-thiophene (**SCR007**) rings attached to the imine-containing linker.

As for the binding targets, we selected 33 *N*-glycans common to the surfaces of flavi- and coronaviruses [31–34]. These *N*-glycans possess conserved structural features that were used to sort them into three groups (Fig. 1B and C): (1) oligomannose *N*-glycans bearing 3 or 5–9 mannose units (abbreviated as **MX**, where **X** = 3 or 5–9); (2) hybrid *N*-glycans with either LacNAc (**G1**) or Neu5Acα(2 → 6)LacNAc (**S1**) on the α 1–3 arm and a Man_3 (**M3**) or Man_5 (**M5**) group on the α1–6 arm; (3) complex *N*-glycans which were sorted into four subgroups based on the type of the carbohydrate at the antennae termini: GlcNAc-terminated (**NX**, where **X** = 1–4), Gal-terminated (**GX**, where **X** = 2–4), GalNAc-terminated (**L2**), and Neu5Ac-terminated (**SX**, where **X** = 1–4). Their structures can include an optional bisecting GlcNAc (**-B**) or fucosyl at the reducing (**F**) and an additional fucosyl on the non-reducing (**F-3/6**) GlcNAc. The attachment of a third antenna to either the α 1–3 or the α 1–6 arm is indicated by the **-3** or **-6** suffixes. Terminal Neu5Ac residues (**S1**, **S2**, **S3-3**, **S3-6** and **S4**) are always attached to the preceding Gal with the $\alpha(2 \rightarrow 6)$ glycosidic bond and were neutralized with K^+ cations.

MD simulations were performed and analyzed using the GROMACS-

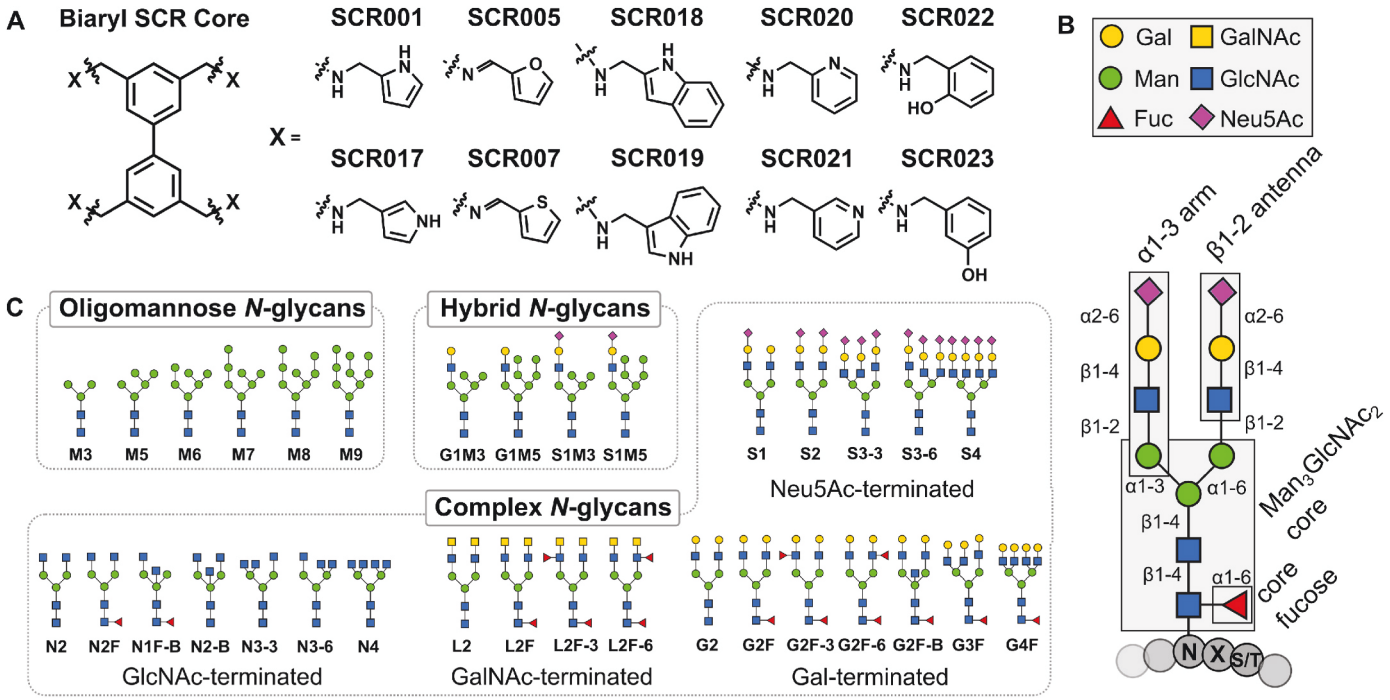

Fig. 1. **(A)** 10 Tetrapodal SCRs investigated in this work. **(B)** A biantenary complex *N*-glycan showing configurations of glycosidic bonds and labelling scheme. The monosaccharides are coded according to the SNFG notation [35]. **(C)** 33 studied *N*-glycans common to EnV [32–34], divided into three major types. The complex *N*-glycan are divided into four subgroups, depending on the type of a carbohydrate at the non-reducing end.

2020.1 MD engine [36]. The CHARMM36 force field was selected for *N*-glycans because of its parametrization for carbohydrates [37] and carbohydrate-protein interactions [38]. The SCRs were parameterized using CHARMM36-compatible CGenFF [39]. For the simulation of the SCR·glycan pairs (1 μs each), we monitored the distance between their center-of-masses. As an example, Fig. 2A compares the binding of **SCR007** (2-thiophene) to **N3-3**, which contains an additional GlcNAcβ (1 → 4) on the α1–3 arm, and to **N2**, which lacks the GlcNAc residue. The histograms of the distances reveals that **SCR007** binds preferably to the tri-antennary **N3-3**, and this differential binding can be understood by inspection of the most abundant **SCR007·N3-3** clusters present in the molecular trajectory. An overlay of 10 conformers that belong to the cluster is shown in Fig. 2B. The structure of the complex shows that the additional GlcNAcβ(1 → 4) in the glycan α1–3 arm engages in C–H···π interactions with one of the thiophene heterocycles, thus stabilizing the **SCR007·N3-3** complex more than the **SCR007·N2** complex, which does not have this additional stabilization. We defined the association quotient (Q_a) as the ratio between a population of bound and unbound states, and use it as a metric to compare the binding of the 330 flexible host-guest complexes (Table S1 in the SI). The Q_a can be converted to the binding constant K_a via the equation $K_a = Q_a(Q_a + 1)/C_o$ (see SI for details), where $C_o = 7.69$ mM, which is equivalent to one molecule per 216 nm^3. With this equation, Q_as of 1, 5, and 10 are equal to K_as of $2.6 \cdot 10^2$, $3.9 \cdot 10^3$, and $1.4 \cdot 10^4$ M^{-1}, which are in the same range as the K_as determined experimentally between monosaccharides and SCRs [29]. To validate that the Q_as are reproducible, we repeated the simulations of 8 SCRs and three *N*-glycans (**G1M3**, **G2F** and **S2**) three times to calculate the standard deviation (Table S2). Out of 24 SCR·*N*-glycan pairs, 17 pairs yielded a standard deviation below 0.33, six pairs yielded between 0.33 an 0.66, and one pair (**SCR019·S2**) had a standard deviation of 0.98. Thus, the Q_a is able to identify the receptors that bind a specific glycan, while avoiding computationally-demanding derivations of fully converged binding constants [40].

All plots showing the center-of-mass distance are presented in Figs. S2–S34 in the SI and the populations of the most abundant clusters are shown in Tables S3–S8. The overview of Q_as of six SCRs (**SCR005**, **SCR007**, **SCR018**, **SCR019**, **SCR022**, **SCR023**) that have either strong or selective binding to 33 *N*-glycans is shown in Fig. 3 and illustrative examples of host·guest complexes are discussed in detail. An overview of three other SCRs (**SCR001**, **SCR017**, **SCR020**), which show weak or no biding towards, *N*-glycans and one SCR (**SCR021**), which is selective to two glycans, are shown in Fig. S35 in the SI.

2.1. Oligomannose N-glycans

Only three SCRs, **SCR018**, **SCR019**, and **SCR022**, bind the oligomannose *N*-glycans with $Q_a > 1$, and their selectivity depends on the mannose content of the glycan. Of these, **SCR022** possesses the largest Q_a (1.59) towards **M9**, but the affinity of other oligomannose *N*-glycans towards the receptor decreases with decreasing mannose content in the α1–6 arm, to Q_a of 0.70 (**M8**) and 0.71 (**M7**). Further change in the number of mannose residues in the α1–3 arm in **M6** and **M5** increases the Q_as to 1.28 and 1.28, respectively. The representative structures of the most populated complexes, shown in Fig. S36A, can be used to rationalize this trend. The receptor initially associates at the $\alpha(1 \rightarrow 6)$ glycosidic bond in the Man_5 in the α1–6 arm. This allows the phenol heterocycles to form interactions with the adjacent mannosyl residues. Reducing the mannosyl content in this arm results in a weaker association. In the complexes with **M6** and **M5** glycans, the receptor moves towards the central mannosyl moiety in the $Man_3GlcNAc_2$ core, which allows the heterocycles to form C–H···π and H-bonding interactions with the α1–3 and α1–6 arms, resulting in a similar magnitude of the association.

The other two receptors that bind oligomannose *N*-glycans selectively are indole-bearing **SCR018** and **SCR019**. The first, which has a linker attached to the C2 of the indole heterocycle, is selective towards **M8** ($Q_a = 1.12$), while association with other oligomannose glycans show $Q_a < 1$. Its regioisomer **SCR019**, which has the linkers attached to C3 of the indole hetorocycle, is selective towards **M7** ($Q_a = 1.41$). The most abundant clusters (Figs. 4A and S36B) reveal that in both complexes the biaryl core of the receptor aligns along the $\alpha(1 \rightarrow 6)$ glycosidic bond connecting the two mannoses in the $Man_3GlcNAc_2$ core. In this position, the receptor associates at the more hydrophobic α-face of the central mannosyl group, opposite to the hydrophilic axial O2. Then, four heterocycle groups of the receptor engage in C–H···π interactions with two mannosyl groups in the α1-6 arm, one mannosyl in the α1–3 arm, and the GlcNAc moiety at the reducing end of the *N*-glycan. In the **SCR019·M7** complex, these interactions with the α1–6 arm appear stronger than in **SCR018·M8**. The additional mannose residues on the α1–6 arm in **M9** sterically hinder the alignment of the heterocycle rings, whereas reducing the number of mannose in the α1–3 arm diminishes the interactions of the heterocycle with this arm. Finally, we looked at the binding of the receptor with the model **M3** glycan, which consists solely of the $Man_3GlcNAc_2$ core. It shows a trend similar to other oligomannose *N*-glycans, and only **SCR022** has $Q_a > 1$, and other receptors do not bind to this model glycan.

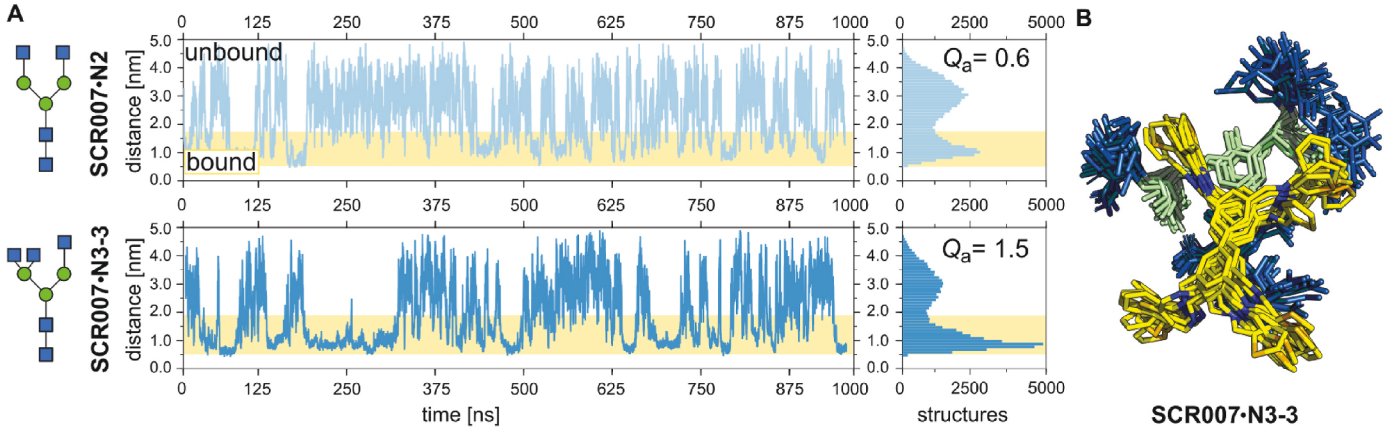

Fig. 2. **(A)** MD simulations of binding between **SCR007** or **N2** (top) and **N3-3** (bottom), with histograms illustrating the time spent in bound and unbound states. The ratio of the number of bound and unbound structures defines the association quotient, Q_a, shown in the histogram. **(B)** Overlay of 10 conformers selected from the most abundant cluster of the **N3-3** glycan and **SCR007** (yellow) complex, showing its fluctuation around the equilibrium structure.

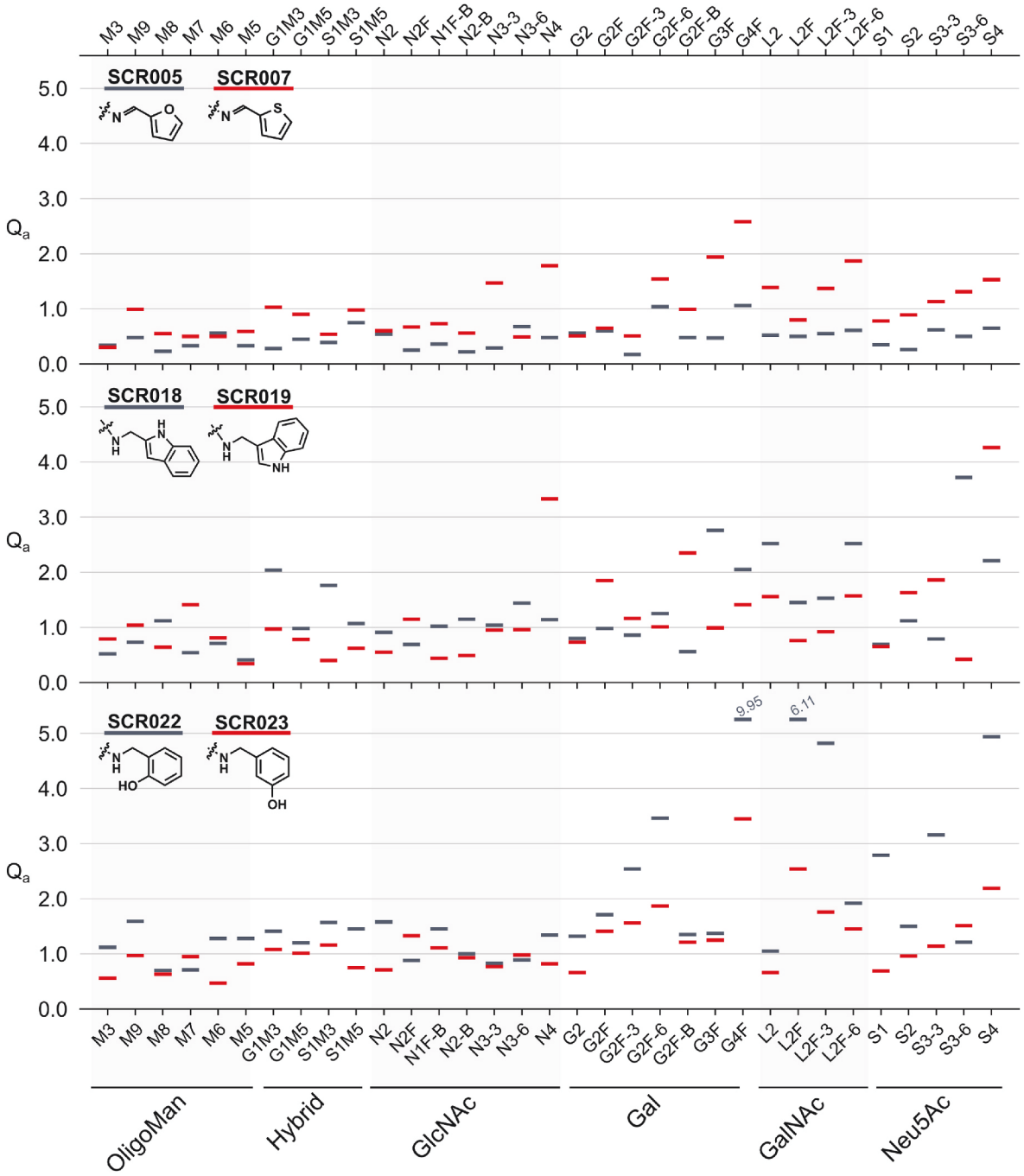

Fig. 3. Association quotients (Q_as) of six SCRs binding to the library of 33 *N*-glycans. The SCRs are sorted in regioisomeric pairs, as well as a pair of **SCR005** and **SCR007**, to help visualize the differences in Q_as between receptors. Each SCR pair is color coded. Two Q_as of **SCR022** are above the limit of 5.0 and have their numerical value included above the bar. The *N*-glycans are sorted into groups based upon their composition and the type of a carbohydrate at the non-reducing end. The data for the remaining four SCRs is shown in Fig. S35.

2.2. Hybrid N-glycans

The four hybrid *N*-glycans can be divided into two sets based upon the composition of the α1–3 arm (**G1**-and **S1**-) and the mannose content in the α1–6 arm (-**M3** and -**M5**). **SCR018** is selective towards the two glycans that contain three mannose residues on the α1–6 arm, as its supramolecular complex with the **G1M3** and **S1M3** glycans yields Q_as of 2.04 and 1.76, respectively. When the number of mannoses in the α1–6 arm increases to five units, the receptor's association becomes weaker, and the Q_as decrease to 0.98 (**G1M5**) and 1.07 (**S1M5**). The structures of the complexes reveal that the stronger binding to hybrid glycans with a Man_3 group is driven by the interactions of the SCR with both α1–3 and α1–6 arms (Figs. 4B and S37A), as the receptor inserts into the central position above the central mannosyl in the $Man_3GlcNAc_2$ core. Increasing the mannose content on the α1–6 arm blocks this binding conformations, and the receptor instead associates on the α1–6 arm, leading to a less stable complex (Fig. S37A). The other two receptors which bind strongly to hybrid *N*-glycans are **SCR022** and **SCR023**, but unlike **SCR018** these receptors are promiscuous binders, meaning they will bind all four hybrid *N*-glycans with similar affinity. The representative structures of these complexes reveal that the receptors associate at the $Man_3GlcNAc_2$ core (Fig. S37B) which allows their linkers and

heterocycles to engage in H-bond and C–H···π interactions with the antennae.

2.3. Complex N-glycans

GlcNAc-terminated: The smallest complex *N*-glycan, **N2**, has $Q_a < 1$ with all SCRs except for **SCR022**, which has Q_a of 1.58. Fucosylation of the reducing GlcNAc (**N2F**) reduces its Q_a to 0.88 but increases the affinity of other three receptors, **SCR019** (1.15), **SCR021** (1.14), and **SCR023** (1.33), all of which have the heterocyclic ring attached to the linker at the 3-position. The simulations reveal that the binding, which in all cases occurs at the $Man_3GlcNAc_2$ core, is driven by hydrophobic C–H···π interactions between the SCR biaryl core and the β-face of the non-terminal GlcNAc residue and is boosted by analogous interactions between the 3-heterocyclic rings and GlcNAcs in two antennae. The adjacent Fuc residue further expands the hydrophobic surface of the glycan, allowing for the formation of additional C–H···π interactions with the SCRs containing hydrophobic 3-heterocycles (Figs. 4C and S38A). Changing the attachment of GlcNAc from the α1–6 arm (**N2F**) to the bisecting position (**N1F-B**) weakens the binding of these three receptors, a result of losing favorable interactions between one of the receptor arms and the GlcNAc residue in the flexible antenna (Fig. S38B), as GlcNAc in the bisecting position is less available for forming C–H···π interactions. Comparison of **N2** and **N2-B** *N*-glycans shows a similar trend, as the addition of the bisecting GlcNAc weakens the overall binding by removing the H-bond that forms with the O4-hydroxy groups of the central mannosyl residue in the $Man_3GlcNAc_2$ core, which restrains the flexibility of both antennae.

Increasing the number of GlcNAc-terminated antennae in **N3-3**, **N3-6**, and **N4** glycans has the largest effect on the selectivity of the **SCR007** and **SCR021** receptors. An additional GlcNAc residue attached to the mannose in the α1–3 arm increases the **SCR007** binding affinity to **N3-3** and **N4** glycans from 0.60 (**N2**) to 1.47 and 1.78, respectively, whereas addition of the GlcNAc residue to the α1–6 arm does not affect the receptor's association. Figs. 4D and S38B reveal that the extra GlcNAc in the α1–3 arm provides some stabilizing C–H···π interactions with the 2-thiophene ring of **SCR007**, yet the interactions of the heterocyclic rings with the GlcNAc and Man in the α1–6 arm contribute more to the stability of the complex. An additional GlcNAc in the α1–6 arm disrupts these interactions, which renders the **SCR007** complex with **N3-6** less stable ($Q_a = 0.49$). Another receptor, **SCR021**, is specific towards the **N3-6** glycan, as its association quotient of 2.36 is the largest value observed for this receptor. Inspection of the most abundant cluster (Fig. S38C) shows that the receptor inserts between the two GlcNAcs in the α1–6 arm, such that each of phenyl rings forms C–H···π interactions with the β-face of the glycan ring. Then, this position is further stabilized by other contacts between the 3-pyridine heterocycles and same GlcNAcs. This binding mode is not observed for any other receptor or glycan, and might explain why **SCR021**·**N3-6** is the only stable complex formed by **SCR021**. The tetra-antennary **N4** glycan forms the most stable complexes with **SCR007** and **SCR019** (Q_a of 3.33, Fig. S38C). **SCR007**, instead of binding facially at the $Man_3GlcNAc_2$ core, inserts its biaryl core between the core and the GlcNAc in the α1–6 arm. The sandwiched complex enables C–H···π interactions between the α-face of the GlcNAc and the biaryl core, and the *N*-acetyl group and the biaryl core on the opposite side. **SCR019**, on the other hand, associates at the α-face of the central mannosyl between the two antennae, and this central position allows the four indole heterocycles to engage in C–H···π interactions with the four GlcNAc groups in both arms.

Gal-terminated: Extending the **N2** glycan's antennae by two Gal residues (**G2**) does not alter the SCRs' binding affinities significantly, and **SCR022** remains the only receptor that displays $Q_a > 1$ with **G2**. While the core-fucosylation in the **G2F** glycan nominally increases the association to all SCRs, it substantially increases the binding with **SCR019** from $Q_a = 0.73$ for **G2** to $Q_a = 1.85$ for **G2F**. In a complex similar to that formed between the receptors and **N2F**, the biaryl core of **SCR019** aligns parallel to the β-face of the GlcNAc and α-face of the central mannosyl of the $Man_3GlcNAc_2$ core, which allows three out of four arms to engage in C–H···π interactions with two other GlcNAc residues, one in each antenna, and the fucose residue adjacent to the reducing end (S39A). Similarly, this synergy of multiple SCR-glycan contacts increases Q_a from 1.32 (**SCR022**·**G2**) to 1.71 for the **SCR022**·**G2F** complex (Fig. 4E) and from 0.66 to 1.41 for **SCR023**·**G2F**. Further fucosylation in either of the antennae in **G2F-3** and **GF2-6** glycans affects their binding with **SCR019** and **SCR022** in an opposite manner. On the one hand, the additional fucose attached to the GlcNAc in α1–3 or α1–6 destabilizes the **SCR019**·glycan complexes by disrupting the interactions of the heterocycles with either of the antenna, yielding Q_as of 1.16 and 1.01 for **G2F-3** and **G2F-6**, respectively. On the other hand, the fucose stabilizes both **SCR022**· glycan complexes by enhancing H-bonding interactions with the 2-phenol arms (Q_as of 2.54 for **G2F-3** and 3.46 for **G2F-6**, Fig. S39B). **SCR023** experiences similar but weaker enhancement. An addition of a bisecting GlcNAc residue in the **G2F-B** glycan decreases the Q_a of the glycan complex with **SCR022** to 1.35. The lower Q_a highlights the importance of the H-bond with the O4 atom of the central mannosyl residue for achieving the strong bonding. The same bisecting GlcNAc influences the binding of **SCR018** and **SCR019** in opposite ways: by decreasing Q_a of **SCR018** from 0.98 (**G2F**) to 0.56 (**G2F-B**) but increasing Q_a of **SCR019** from 1.85 (**G2F**) to 2.35 (**G2F-B**).

Addition of the third LacNAc antenna to the α1–3 arm of **G2F** to form **G3F** either weakens or leaves the association unchanged for all SCRs except for **SCR007** and **SCR018**, whose Q_as increase, respectively, from 0.65 (**G2F**) to 1.94 (**G3F**) and from 0.98 (**G2F**) to 2.76 (**G3F**). In both complexes with **G3F**, the biaryl core of **SCR017** or **SCR018** assumes the same central position as **G2F**, and the three receptor heterocycles form C–H···π interactions with core fucose and the GlcNAc in the β1–2 antennae, but these contacts are more frequent in tri-antennary **G3F** (Figs. 4F and S39C). Subsequent addition of the fourth LacNAc antenna to the α1–6 arm in **G4F** compounds with the effect of the previous antenna. Although the association of **SCR018**·**G4F** decreases to 2.05, the extra antenna increases the association with **SCR007**, **SCR023** and **SCR022**, complexes which display Q_as of 2.58, 3.45 and 9.95, respectively. The **SCR007** receptor forms a similar complex to that already observed for GlcNAc-terminated glycans, and the addition of the antennae to the α1–3 arm increases its Q_a, which compounds with the addition of the fourth moiety. Finally, the large Q_as of the phenol-based receptors is somewhat unexpected, as their Q_as with tri-antennary (**N3-3**, **N3-6**, and **G3F**) and tetra-antennary (**N4**) *N*-glycans are all smaller than 1.5. However, the inspection of the molecular structures of the most populated clusters of **SCR022**·**G4F** and **SCR023**·**G4F** reveals H-bond interactions between the phenol and the amine linkers and the two GlcNAc residues at the reducing end (Fig. S39D). This binding conformation is further stabilized by the bending of the β1–2 antenna in the α1–6 arm and forming C–H···π interactions with another phenol group.

GalNAc-terminated: Next, we investigated how changing the terminal Gal to GalNAc, which results in additional *N*-acetyl groups in the antennae, affects the binding of the SCRs and complex *N*-glycans. SCRs have a larger affinity to the smallest LacdiNAc-bearing *N*-glycan, **L2**, than to analogous Gal-terminated **G2**. We observed the largest Q_a of 2.52 and 1.56 with the two indole-bearing SCRs, **SCR018** and **SCR019**, as well as a Q_a of 1.39 for binding with **SCR007** (Fig. S40A). The core fucosylation in **L2F** decreases the association of the indole-bearing SCRs receptors to 1.45 and 0.76, respectively, and **SCR007** to 0.80, but it increases the association of the phenol-bearing receptors **SCR022** and **SCR023** to 6.11 and 2.54, respectively. The representative structures of the most abundant cluster of the complex with **SCR022** reveals that the receptor, as observed in previous complexes, aligns with the biaryl core parallel to the $Man_3GlcNAc_2$ core (Figs. 4G and S40B). One of the phenol rings engages in a C–H···π interaction with the core fucose and also with the GlcNAc in the α1–6 arms. Additional fucoses in either of the antennae decreases the nominal Q_as of both of the phenol SCRs

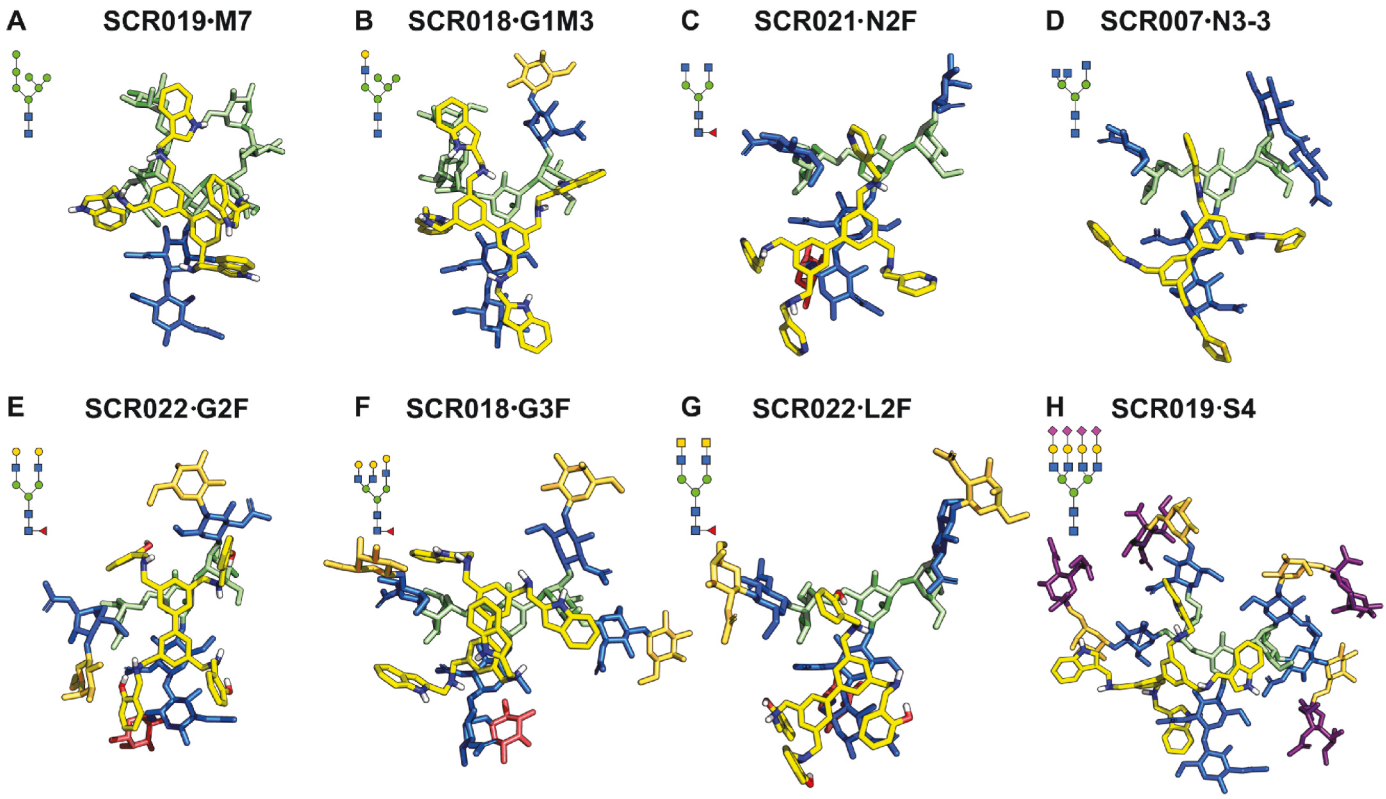

Fig. 4. Representative structures of eight SCR·*N*-glycan complexes discussed in the text. The color-coding of the glycans follows the SNFG notation, shown in the cartoon, and hydrogens were omitted for clarity. SCRs are shown in yellow, with nitrogen atoms highlighted in blue, oxygen atoms in red, sulphur atoms in orange, and polar hydrogen atoms in white. Note that the SCRs tend to bind at the $Man_3GlcNAc_2$ core of the *N*-glycan, facing the α-face of the central mannosyl residue. To show the binding, the *N*-glycans were rotated 180^o with respect to the orientation shown in the SNFG representation.

(Fig. S40C). However, the association of **SCR022·L2F-3** remains strong with a Q_a of 4.82, whereas shifting the fucose to the α1–6 arm (**L2F-6**) decreases the Q_a to 1.92. This agrees with the observed binding mode for the **L2F** glycan, as extra fucosylation in the α1–6 antenna disrupts the C–H$\cdots\pi$ interactions with the heterocycle. Adding the fucose to the α1–6 arms brings the Q_a of **SCR007** and **SCR019** to similar values as for **L2**.

Neu5Ac-terminated: The last set of glycans feature bi- (**S2**), tri- (**S3-3** and **S3-6**), and tetra- (**S4**) antennary glycans terminated with Neu5Ac residues attached with the $\alpha(2 \rightarrow 6)$ linkages, and a **G2** glycan featuring only one Neu5Ac on the α1–3 arm (**S1**). Only **SCR022** is a strong binder towards the **S1** glycan, with a Q_a of 2.79, whereas all other Q_as are smaller than 1. Addition of the second Neu5Ac in the α1-6 antennae in **S2** weakens the binding of **SCR022** to 1.49, but simultaneously increases the binding of **SCR019** from 0.65 to 1.63. The overall binding to these negatively charged bi-antennary glycans remains weak, but increasing the number of antennae to three results in large changes in the Q_as of several receptors. First, the binding of **SCR019** appears to be selective towards attachment of the antenna: additional antenna in the α1–3 arm (**S3-3**) increases its binding to 1.86, but adding it to the α1–6 arm (**S3-6**) lowers significantly the receptor affinity towards the sialylated glycan ($Q_a = 0.42$). Second, **SCR018** experiences the opposite effect, as it binds to **S3-3** weaker (0.79), but its binding to **S3-6** is strongly enhanced with Q_a of 3.72. The inspection of the structures shows that the two regio-siomeric receptors associate such that they form interactions with the new antenna in the respective arm (Fig. S41A). Both receptors then are strong binders to the tetra-antennary **S4** glycan, with Q_as of 2.21 and 4.26, respectively, for **SCR018** and **SCR019** (Figs. 4H and S41B), as the **S4** glycan features the additional antenna available for the interactions with either of the receptors.

3. Conclusions

In conclusion, a systematic MD study of 330 SCR·glycan pairs revealed that the SCRs are selective binders towards *N*-glycans, and their specificity depends on the type of the linker, heterocyclic groups, and their point-of-attachment, consistent with our previous conclusions [29, 30]. The investigated SCRs can be divided, based upon their Q_as towards the library of 33 *N*-glycans, into three groups: weak binders, promiscuous binders and selective binders. The first group consists of weakly binding SCRs that have Q_as with all *N*-glycans smaller than 1. The members of this group are **SCR001**, **SCR005**, **SCR017**, and **SCR020**. The second group consists of promiscuous SCRs, which have $Q_a > 1$ towards more than half of the investigated *N*-glycans. **SCR018**, **SCR022** and **SCR023**, which have $Q_a > 1$ towards 19, 27 and 17 different *N*-glycans, respectively, are members of this group. **SCR022** binds almost all investigated glycans, and the other two receptors show binding preferences between the types of *N*-glycans. For instance, **SCR023** does not bind oligomannose *N*-glycans, but binds almost all Gal-, GalNAc-, and Neu5Ac-terminated complex *N*-glycans. **SCR018** has more nuanced binding preferences. Although it would bind 19 different glycans with the $Q_a > 1$, it has $Q_a > 2$ towards seven specific *N*-glycans, which enables the receptor to differentiate between minute structural details, such as the number of mannosyls in the α1-6 arm of the hybrid *N*-glycans or the position of the antenna in the Neu5Ac-terminated complex *N*-glycans. However, because these receptors have significant Q_as towards more than half of the *N*-glycans, they would probably display high off-site toxicity as a consequence of binding to non-EnV glycans. The third group is composed of three selective binders, **SCR007**, **SCR019** and **SCR021**, which have large Q_as towards few specific *N*-glycans. **SCR021** has only two Q_as> 1 which means it would bind selectively to the **N2F** and **N3-6** glycans. **SCR007** associates with $Q_a > 1$ with 12 *N*-glycans, which share some common structural features such as additional

antenna in the α1–3 arm. We can conclude that this receptor would bind to complex *N*-glycans, which have more than three Gal-terminated antennae. Finally, **SCR019**, which has $Q_a > 1$ for complexes with 14 different *N*-glycans, is strongly selective for tetra-antennary glycans **N4** (3.33), and **S4** (4.26) and tri-antennary glycans with the additional antennae in the α1-3 arm.

The analysis of the structures of the most abundant complexes observed in the MD simulations revealed several key features of SCR·glycan binding. First, we observe that the receptors associate almost exclusively at the $Man_3GlcNAc_2$ core, which is common to all EnV glycans (Fig. 4). This association is driven primarily by the C–H···π interactions between the SCR biaryl core and the hydrophilic surfaces of the pyranose rings. Second, the strength of the association depends on the noncovalent interactions (H-bonding and C–H···π) between the heterocycle rings and H-bonding linker, and the decoration of the *N*-glycan that involves the composition and connectivity of the antennae, optional fucosylation and/or bisecting GlcNAc. The weak binding with the **M3** *N*-glycan, which constitutes the $Man_3GlcNAc_2$ core in other *N*-glycans, confirms the importance of the interactions with the antennae. However, each of these structural features affects the binding of the receptor in a non-trivial way and complicates the rational design of the receptors towards specific targets. This shortcoming, as shown in this work, can be alleviated by the computational screening the SCRs against the library of the EnV glycans.

In summary, the MD simulations illustrate how computations can guide the experimental design to address one of the most challenging and pressing problems in medicinal chemistry: developing glycan-binding antiviral agents. These data suggest that it might be possible to dial-in affinity and selectivity of SCRs towards a particular glycan target. If the computational predictions are corroborated by the experimental studies, the proposed selective binding could reclassify glycans from "undruggable" to viable targets for antivirals and helps to open new avenues of investigation in medicine, biology, sensors, and diagnostics.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Adam Braunschweig reports was provided by US Army Research Laboratory. Adam Braunschweig reports a relationship with Dultech LLC that includes: board membership. Adam Braunschweig has patent licensed to Dultech LLC.

Acknowledgements

M.M. and A. B. B. thank the Army Research Office (W911NF2010271) for the financial support and M.M. thanks the National Institute of General Medical Sciences (SC2GM135145) for the financial support and COVID-19 High Performance Computing Consortium for support of the computational work (MCB200141). The simulations were performed and analyzed on supercomputers at Lawrence Livermore National Laboratory (LLNL) and Texas Advanced Computing Center (TACC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carres.2022.108574.

References

[1] C.J. Burrell, C.R. Howard, F.A. Murphy, Virion Structure and Composition, Academic Press, 2017, pp. 27–37, https://doi.org/10.1016/B978-0-12-375156-0.00003-5, chapter 3.
[2] C.J. Burrell, C.R. Howard, F.A. Murphy, Virus Replication, Academic Press, 2017, pp. 39–55, https://doi.org/10.1016/B978-0-12-375156-0.00004-7, chapter 4.
[3] UNAIDS, UNAIDS global AIDS update - confronting inequalities - lessons for pandemic responses from 40 years of AIDS, URL: https://www.unaids.org/en/resources/documents/2021/2021-global-aids-update, 2021. (Accessed 30 November 2021).
[4] A. Gulland, Zika virus is a global public health emergency, declares WHO, BMJ 352 (2016) i657, https://doi.org/10.1136/bmj.i657.
[5] E. Mahase, Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction, BMJ 368 (2020), https://doi.org/10.1136/bmj.m1036 m1036.
[6] J.A. Bernatchez, L.T. Tran, J. Li, Y. Luan, J.L. Siqueira-Neto, R. Li, Drugs for the treatment of Zika virus infection, J. Med. Chem. 63 (2020) 470–489, https://doi.org/10.1021/acs.jmedchem.9b00775.
[7] FDA, Emergency use authorization for drugs for COVID-19, URL: https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization#coviddrugs. (Accessed 10 April 2022).
[8] C. Shan, X. Xie, A.D.T. Barrett, M.A. Garcia-Blanco, R.B. Tesh, N. da Costa Vasconcelos Pedro Fernando, Vasilakis, S.C. Weaver, P.Y. Shi, Zika virus: diagnosis, therapeutics, and vaccine, ACS Infect. Dis. 2 (2016) 170–172, https://doi.org/10.1021/acsinfecdis.6b00030.
[9] F. Jiang, L. Deng, L. Zhang, Y. Cai, C.W. Cheung, Z. Xia, Review of the clinical characteristics of coronavirus disease 2019 (COVID-19), J. Gen. Intern. Med. 35 (2020) 1545–1549, https://doi.org/10.1007/s11606-020-05762-w.
[10] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, Z. Qian, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, Nat. Commun. 11 (2020) 1620–1632, https://doi.org/10.1038/s41467-020-15562-9.
[11] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292, https://doi.org/10.1016/j.cell.2020.02.058.
[12] O.C. Grant, D. Montgomery, K. Ito, R.J. Woods, Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition, Sci. Rep. 10 (2020) 14991, https://doi.org/10.1038/s41598-020-71748-7.
[13] M.F. Bravo, M.A. Lema, M. Marianski, A.B. Braunschweig, Flexible synthetic carbohydrate receptors as inhibitors of viral attachment, Biochem 60 (2021) 999–1018, https://doi.org/10.1021/acs.biochem.0c00732.
[14] Y. Watanabe, Z.T. Berndsen, J. Raghwani, G.E. Seabright, J.D. Allen, O.G. Pybus, J. S. McLellan, I.A. Wilson, T.A. Bowden, A.B. Ward, M. Crispin, Vulnerabilities in coronavirus glycan shields despite extensive glycosylation, Nat. Commun. 11 (2020) 2688, https://doi.org/10.1038/s41467-020-16567-0.
[15] Y. Watanabe, T.A. Bowden, I.A. Wilson, M. Crispin, Exploitation of glycosylation in enveloped virus pathobiology, Biochim. Biophys. Acta Gen. Subj. 1863 (2019) 1480–1497, https://doi.org/10.1016/J.BBAGEN.2019.05.012.
[16] L. Casalino, Z. Gaieb, J.A. Goldsmith, C.K. Hjorth, A.C. Dommer, A.M. Harbison, C. A. Fogarty, E.P. Barros, B.C. Taylor, J.S. McLellan, E. Fadda, R.E. Amaro, Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein, ACS Cent. Sci. 6 (2020) 1722–1734, https://doi.org/10.1021/acscentsci.0c01056.
[17] M. Gur, E. Taka, S.Z. Yilmaz, C. Kilinc, U. Aktas, M. Golcuk, Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states, J. Chem. Phys. 153 (2020), 075101, https://doi.org/10.1063/5.0011141.
[18] T. Sztain, S.H. Ahn, A.T. Bogetti, L. Casalino, J.A. Goldsmith, E. Seitz, R.S. McCool, F.L. Kearns, F. Acosta-Reyes, S. Maji, G. Mashayekhi, J.A. McCammon, A. Ourmazd, J. Frank, J.S. McLellan, L.T. Chong, R.E. Amaro, A glycan gate controls opening of the SARS-CoV-2 spike protein, Nat. Chem. 13 (2021) 963–968, https://doi.org/10.1038/s41557-021-00758-3.
[19] J.S. Lazo, E.R. Sharlow, Drugging undruggable molecular cancer targets, Annu. Rev. Pharmacol. Toxicol. 56 (2016) 23–40, https://doi.org/10.1146/annurev-pharmtox-010715-103440.
[20] I.M. Vasconcelos, J.T.A. Oliveira, Antinutritional properties of plant lectins, Toxicon 44 (2004) 385–403, https://doi.org/10.1016/j.toxicon.2004.05.005.
[21] F.D. Vrionis, C.J. Wikstrand, P. Fredman, J.E.M. ansson, L. Svennerholm, D. D. Bigner, Five new epitope-defined monoclonal antibodies reactive with GM2 and human glioma and medulloblastoma cell lines, Cancer Res. 49 (1989) 6645–6651.
[22] N.K.V. Cheung, I.Y. Cheung, K. Kramer, S. Modak, D. Kuk, N. Pandit-Taskar, E. Chamberlain, I. Ostrovnaya, B.H. Kushner, Key role for myeloid cells: phase II results of anti-GD2 antibody 3F8 plus granulocyte-macrophage colony-stimulating factor for chemoresistant osteomedullary neuroblastoma, Int. J. Cancer 135 (2014) 2199–2205, https://doi.org/10.1002/ijc.28851.
[23] M. Mazik, Design of lectin mimetics, Chembiochem 9 (2008) 1015–1017, https://doi.org/10.1002/cbic.200800038.
[24] A.P. Davis, Synthetic lectins, Org. Biomol. Chem. 7 (2009) 3629–3638, https://doi.org/10.1039/b909856a.
[25] O. Francesconi, C. Nativi, G. Gabrielli, I.D. Simone, S. Noppen, J. Balzarini, S. Liekens, S. Roelens, Antiviral activity of synthetic aminopyrrolic carbohydrate binding agents: targeting the glycans of viral gp120 to inhibit HIV entry, Chem. Eur J. 21 (2015) 10089–10093, https://doi.org/10.1002/chem.201501030.
[26] S. Rieth, M.R. Miner, C.M. Chang, B. Hurlocker, A.B. Braunschweig, Saccharide receptor achieves concentration dependent mannoside selectivity through two distinct cooperative binding pathways, Chem. Sci. 4 (2013) 357–367, https://doi.org/10.1039/C2SC20873C.
[27] K. Palanichamy, M.F. Bravo, M.A. Shlain, F. Schiro, Y. Naeem, M. Marianski, A. B. Braunschweig, Binding studies on a library of induced-fit synthetic carbohydrate receptors with mannoside selectivity, Chem. Eur J. 24 (2018) 13971–13982, https://doi.org/10.1002/chem.201803317.
[28] K. Palanichamy, A. Joshi, T. Mehmetoglu-Gurbuz, M.F. Bravo, M.A. Shlain, F. Schiro, Y. Naeem, H. Garg, A.B. Braunschweig, Anti-zika activity of a library of

synthetic carbohydrate receptors, J. Med. Chem. 62 (2019) 4110–4119, https://doi.org/10.1021/acs.jmedchem.9b00142.
[29] M.F. Bravo, K. Palanichamy, M.A. Shlain, F. Schiro, Y. Naeem, M. Marianski, A. B. Braunschweig, Synthesis and binding of mannose-specific synthetic carbohydrate receptors, Chem. Eur J. 26 (2020) 11782–11795, https://doi.org/10.1002/chem.202000481.
[30] K. Thakur, M.A. Shlain, M. Marianski, A.B. Braunschweig, Regiochemical effects on the carbohydrate binding and selectivity of flexible synthetic carbohydrate receptors with indole and quinoline heterocyclic groups, Eur. J. Org Chem. 2021 (2021) 5262–5274, https://doi.org/10.1002/ejoc.202100763.
[31] N.K. Routhu, S.D. Lehoux, E.A. Rouse, M.R.M. Bidokhti, L.B. Giron, A. Anzurez, S. P. Reid, M. Abdel-Mohsen, R.D. Cummings, S.N. Byrareddy, Glycosylation of Zika virus is important in host–virus interaction and pathogenic potential, Int. J. Mol. Sci. 20 (2019) 5206–5225, https://doi.org/10.3390/ijms20205206.
[32] Y. Watanabe, J.D. Allen, D. Wrapp, J.S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike, Science 369 (2020) 330–333, https://doi.org/10.1126/science.abb9983.
[33] Y. Zhang, W. Zhao, Y. Mao, Y. Chen, S. Wang, Y. Zhong, T. Su, M. Gong, D. Du, X. Lu, J. Cheng, H. Yang, Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins, Mol. Cell. Proteomics 20 (2021) 100058, https://doi.org/10.1074/MCP.RA120.002295.
[34] B.G. Cho, S. Gautam, W. Peng, Y. Huang, M. Goli, Y. Mechref, Direct comparison of N-glycans and their isomers derived from spike glycoprotein 1 of MERS-CoV, SARS-CoV-1, and SARS-CoV-2, J. Proteome Res. 20 (2021) 4357–4365, https://doi.org/10.1021/acs.jproteome.1c00323.
[35] S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lütteke, N. O'Boyle, N.H. Packer, P. Stanley, P. Toukach, A. Varki, R.J. Woods, Updates to the symbol nomenclature for glycans guidelines, Glycobiology 29 (2019) 620–624, https://doi.org/10.1093/glycob/cwz045.
[36] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1-2 (2015) 19–25, https://doi.org/10.1016/J.SOFTX.2015.06.001.
[37] O. Guvench, S.S. Mallajosyula, E.P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F.W. Jamison, A.D. MacKerell, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling, J. Chem. Theor. Comput. 7 (2011) 3162–3180, https://doi.org/10.1021/ct200328p.
[38] W.K. Lay, M.S. Miller, A.H. Elcock, Optimizing solute–solute interactions in the GLYCAM06 and CHARMM36 carbohydrate force fields using osmotic pressure measurements, J. Chem. Theor. Comput. 12 (2016) 1401–1407, https://doi.org/10.1021/acs.jctc.5b01136.
[39] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell, CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (2009) 671–690, https://doi.org/10.1002/jcc.21367.
[40] Z. Tang, C.A. Chang, Binding thermodynamics and kinetics calculations using chemical host and guest: a comprehensive picture of molecular recognition, J. Chem. Theor. Comput. 14 (2018) 303–318, https://doi.org/10.1021/acs.jctc.7b00899.