



**HAL**  
open science

## Engineered display of ganglioside-sugars on protein elicits a clonally and structurally constrained B cell response

Lachlan P Deimel, Xiaochao Xue, Aziz Khan, Lucile Moynie, Charles J Buchanan, Guoxuan Sun, Ryan McBride, Heiko Schuster, Charles Gauthier, Regis Saliba, et al.

### ► To cite this version:

Lachlan P Deimel, Xiaochao Xue, Aziz Khan, Lucile Moynie, Charles J Buchanan, et al.. Engineered display of ganglioside-sugars on protein elicits a clonally and structurally constrained B cell response. 2023. hal-04307337

**HAL Id: hal-04307337**

**<https://hal.science/hal-04307337>**

Preprint submitted on 26 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Engineered display of ganglioside-sugars on protein elicits a clonally and**  
2 **structurally constrained B cell response**

3

4 **Authors**

5 Lachlan P. Deimel<sup>1</sup>, Xiaochao Xue<sup>1,2</sup>, Aziz Khan<sup>2,3</sup>, Lucile Moynie<sup>3</sup>, Charles J.  
6 Buchanan<sup>2</sup>, Guoxuan Sun<sup>3</sup>, Ryan McBride<sup>4</sup>, Heiko Schuster<sup>2</sup>, Charles Gauthier<sup>2,5</sup>,  
7 Regis Saliba<sup>2</sup>, Karolis Leonavicus<sup>2</sup>, Leanne Minall<sup>2</sup>, Guillaume Bort<sup>2</sup>, Rebecca A.  
8 Russell<sup>1</sup>, Erdinc Sezgin<sup>6</sup>, James C. Paulson<sup>4</sup>, Daniel C. Anthony<sup>7</sup>, Andrew J.  
9 Baldwin<sup>2,3</sup>, James Naismith<sup>3</sup>, Torben Schiffner<sup>4,8</sup>, Benjamin G. Davis<sup>2,3,7\*</sup> & Quentin J.  
10 Sattentau<sup>1,9\*</sup>

11

12 <sup>1</sup> Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK

13 <sup>2</sup> Department of Chemistry, University of Oxford, Oxford, OX1 3TA, UK

14 <sup>3</sup> Rosalind Franklin Institute, Harwell Science and Innovation Campus, Oxford OX11  
15 0FA UK

16 <sup>4</sup> Scripps Research, La Jolla, CA 92037, USA

17 <sup>5</sup> Current address: Armand-Frappier Santé Biotechnologie, Institut National de la  
18 Recherche Scientifique (INRS), Laval, H7V 1B7, Canada

19 <sup>6</sup> Science for Life Laboratory, Department of Women's and Children's Health,  
20 Karolinska Institute, Solna, Sweden

21 <sup>7</sup> Department of Pharmacology, University of Oxford, Oxford, OX1 3QT, UK

22 <sup>8</sup> Institute for Drug Discovery, University Leipzig Medical School, Leipzig, 04104,  
23 Germany

24 <sup>9</sup> The Max Delbrück Centre for Molecular Medicine, Campus Berlin-Buch, 13125  
25 Berlin, Germany

26

27 \*Correspondence: [Quentin.Sattentau@path.ox.ac.uk](mailto:Quentin.Sattentau@path.ox.ac.uk) (QJS), [Ben.Davis@rfi.ac.uk](mailto:Ben.Davis@rfi.ac.uk)  
28 (BGD)

29

30 **Key words:** glycan, ganglioside, antibody, B cell, immunogenicity, vaccination, cancer

31 **Abstract:** 267

32 **Work count** (excluding methods, figure captions and references): 6,619

33 **Main figures:** 5

34 **Supplementary figures:** 14

35 **Supplementary tables:** 5

36 **Supplementary documents:** 3

37 S1, Sugar Synthesis and Log Generation; S2, Further Discussion of Structural  
38 Analyses of Bar1 Fab; S3, Supplementary Glycan Microarray Document

39

40 **Abstract [267 words]**

41 Ganglioside sugars, as Tumour-Associated Carbohydrate Antigens (TACAs), are  
42 long-proposed targets for vaccination and therapeutic antibody production, but their  
43 self-like character imparts immunorecessive characteristics that classical vaccination  
44 approaches have to date failed to overcome. One prominent TACA, the glycan  
45 component of ganglioside GM3 (GM3g), is over-expressed on diverse tumours. To  
46 probe the limits of glycan tolerance, we used protein editing methods to display GM3g  
47 in systematically varied non-native presentation modes by attachment to carrier  
48 protein lysine sidechains using diverse chemical linkers. We report here that such  
49 presentation creates glycoconjugates that are strongly immunogenic in mice and elicit  
50 robust antigen-specific IgG responses specific to GM3g. Characterisation of this  
51 response by antigen-specific B cell cloning and phylogenetic and functional analyses  
52 suggests that such display enables the engagement of a highly restricted naïve B cell  
53 class with a defined germline configuration dominated by members of the *IGHV2*  
54 subgroup. Strikingly, structural analysis reveals that glycan features appear to be  
55 recognised primarily by antibody CDRH1/2, and despite the presence of an antigen-  
56 specific Th response and B cell somatic hypermutation, we found no evidence of  
57 affinity maturation towards the antigen. Together these findings suggest a ‘reach-  
58 through’ model in which glycans, when displayed in non-self formats of sufficient  
59 distance from a conjugate backbone, may engage ‘glycan ready’ V-region motifs  
60 encoded in the germline. Structural constraints define why, despite engaging the  
61 trisaccharide, antibodies do not bind natively-presented glycans, such as when linked  
62 to lipid GM3. Our findings provide an explanation for the long-standing difficulties in  
63 raising antibodies reactive with native TACAs, and provide a possible template for  
64 rational vaccine design against this and other TACA antigens.



## 65 **Highlights**

- 66 • GM3g synthetically coupled via a longer, orthogonal (from backbone)  
67 glycoconjugate (LOG) presentation format (thioethyl-lysyl-amidine) display  
68 elicits high-titre IgG responses in mice.
- 69 • The germinal centre experience of LOG glycoconjugate-specific B cell  
70 responses is directly influenced by the protein backbone.
- 71 • Structural characterisation of the antibody response to LOGs reveals highly  
72 restricted germline-encoded glycan-engaging motifs that mediate GM3g  
73 recognition.
- 74 • Failure of antibodies to bind the native trisaccharide highlights barriers to be  
75 overcome for the rational design of anti-TACA antibodies.
- 76

## 77 **Introduction**

78 Glycosylation is a widespread enzymatic process with critical roles in modulating and  
79 controlling protein and lipid structure, function, and stability. Since mammalian glycans  
80 are endogenously added and processed, they and their native glycoconjugates have  
81 restricted immunogenicity to avoid autoreactive responses. Immunogenicity may be  
82 limited by different mechanisms, including central and peripheral immunological  
83 tolerance and intrinsic lack of antigen immunogenicity due to biophysical constraints  
84 such as antigen size, charge and accessibility. Many adaptive tolerance mechanisms  
85 are well-understood and can be partitioned into several main themes: i) B cell negative  
86 selection that eliminates self-glycan-reactive precursors during their development in  
87 the bone marrow (1); ii) antigen interactions with immunoinhibitory lectins such as  
88 CD22 and Siglec-G (2, 3), and iii) lack of T cell help to class-switch and affinity-mature  
89 B cells (4, 5). However, in contrast to self-glycans, antibodies can be readily elicited  
90 against foreign glycans, such as those from bacteria, where the same tolerance  
91 constraints do not apply. Correspondingly, classical glycoconjugate vaccines  
92 artificially display bacteria-derived polysaccharides in a format where the sugar  
93 polymer is typically presented in a non-specific, 'parallel'-mode that the immune  
94 system readily responds to (6, 7). By contrast, the immunological 'blind-spots' present  
95 for self-like glycans and glycoconjugates may render the host vulnerable to pathogen  
96 attack and other pathology such as cancers that routinely exploit glycosylation  
97 processes as an immune evasion tactic. This limits the effective targeting of certain  
98 pathology-associated glycans via vaccination.

99

100 The selective breaking of B cell tolerance to glycans may therefore have profound  
101 utility in certain settings, a highly relevant example being approaches to developing

102 neutralising antibody-based vaccines against HIV-1 (8). The HIV-1 envelope  
103 glycoprotein (Env), the only target of neutralising antibody elicitation and attack,  
104 exploits glycosylation to shield underlying sensitive peptidic epitopes (9, 10). However,  
105 a small subset of HIV-1-infected individuals develop rare B cell clones that produce  
106 potent broadly neutralising antibodies (bNAbs) which interact with glycans or glycan-  
107 protein composite epitopes on Env, and mediate broad and potent neutralisation (11–  
108 14). In general, these bNAbs have undergone significant somatic hypermutation  
109 (SHM) resulting in the progressive accommodation of Env glycans into their respective  
110 epitopes via affinity maturation (15, 16). Thus, most inferred germline revertants (iGL)  
111 of bNAbs fail to recognise natively-glycosylated Env. This creates a formidable  
112 ‘moving target’ for vaccine design as the extent of SHM generated during natural  
113 infection is difficult to recapitulate by current vaccination approaches. Other  
114 approaches may therefore prove necessary for success.

115

116 In principle, the odds of achieving a functional anti-self-like glycan response by  
117 vaccination may be improved by reducing the need for SHM. In this scenario, the  
118 target glycan would ideally be recognised directly by the germline repertoire to initiate  
119 a B cell response and so avoid the requirement for extensive SHM for initial target  
120 epitope recognition. To our knowledge, this has not yet been observed but would prove  
121 widely valuable.

122

123 Tumour-associated carbohydrate antigens (TACAs), which are also self-like, may be  
124 over-expressed or modified on tumour cells compared to their normal counterparts  
125 (17). TACAs present even greater vaccine design challenges compared with viral

126 glyco-antigens because i) several major classes of TACAs are presented on  
127 glycosphingolipids (including gangliosides) rather than proteins, removing the T helper  
128 (Th) component of the adaptive response, and ii) proximity of the carbohydrate to the  
129 membrane likely limits B cell receptor accessibility and downstream antibody  
130 engagement (18). The glycan component of the ganglioside GM3 (GM3g, 3'-O-  
131 sialyllactosyl) is of particular interest for its elevated expression in melanoma and  
132 neuroectodermal tumours (19, 20). GM3 has only very weak immunogenicity, with  
133 experimental studies showing limited success in eliciting anti-GM3 antibody responses  
134 to GM3 or GM3g on various carrier proteins (21–25). Early mouse immunisation  
135 studies with purified GM3 reported an apparent borderline IgM response (21, 26).  
136 Notably, conjugation of GM3g to carrier proteins such as Keyhole Limpet  
137 Haemocyanin (KLH) or bacterial cells (21) provided T cell help to the emerging B cell  
138 responses but in these early studies it remained unclear from the serological analyses  
139 whether the resulting antibody responses were truly SiaLac/GM3-specific or instead  
140 driven in part by artefacts of linker immunogenicity and cross-reactivity in associated  
141 assays (27).

142

143 To further probe our understanding of antibody responses against potentially useful  
144 small self-glycans in a manner that might enable vaccination approaches, we  
145 combined synthetic glycan-protein engineering with detailed B cell immunological  
146 analyses to probe germline-targeted responses leading to the elicitation of glycan-  
147 reactive antibodies. We have exploited bespoke chemical linkages of precisely  
148 modulated format and length, not found in nature, using an orthogonal/'side-on' mode  
149 from protein carrier side-chains to probe glycan tolerance mechanisms. The resulting

150 'reach through' presentation by longer, orthogonal glycoconjugates (LOGs) was  
151 designed to allow glycans to engage a subset of otherwise inaccessible naïve B cells.

152

153 We demonstrate here proof-of-principle of this concept by presenting GM3g on  
154 different carrier protein backbones. GM3g-specific IgG titres were readily elicited via  
155 a highly restricted clonotypic B cell response using distinct B cell receptor (BCR) heavy  
156 chain-mediated glycan recognition. Antibodies binding GM3g do not react with GM3  
157 itself confirming the key role that the presentation of the glycan plays. These data not  
158 only provide a rational basis for the key role of glycan presentation in the specificity of  
159 corresponding B cell clones elicited, but also represent the first evidence of a TACA-  
160 directed germline-targeting immunogen with implications for the future design of  
161 glycan reactive antibody-based vaccine approaches.

162

## 163 **Results**

### 164 *Differing presentation of GM3g modulates B cell immunogenicity*

165 GM3 presents its glycan (GM3g) (**Fig 1a**) natively at short distance (estimated at 6 Å  
166 based on native O-glycoside, three-bond O-hydroxymethyl spaced display from the  
167 head group) from its native macromolecular (lipid membrane) assembly surface.

168 We first chose to interrogate the inherent immunogenicity in mice of natively-presented  
169 GM3g in the context of intact GM3 lipid. Assembly of GM3-bearing liposomes  
170 (PG:PC:Chol:GM3 = 39:39:19:3) created an appropriate macromolecular assembly  
171 bearing multi-copy GM3g (**Fig S1a**). Following immunisation of WT BALB/c mice  
172 formulated with the TLR-4-agonist-based adjuvant Monophosphoryl-Lipid A (MPLA)  
173 (Mata-Haro et al., 2007), antisera against both the GM3-containing and GM3-free

174 control liposome displayed modest IgM reactivity with ceramide (median EPT across  
175 groups of 2,156), and similarly low anti-GM3 IgM titres (EPT = 492) (**Fig S1b–d**).  
176 These findings are consistent with non-specific IgM binding and the absence of  
177 specific GM3g-binding antibodies, reflecting the low-affinity, high-avidity nature of IgM  
178 in ELISA formats (29, 30). Unsurprisingly, in the absence of T cell help (classically  
179 provided by protein in the antigenic complex), antigen-specific IgG was not detected  
180 against either ceramide or GM3 (**Fig S1e,f**). These data confirm the profoundly limited  
181 immunogenicity of GM3g in this macromolecular format.

182

183 We next explored an alternative non-native macromolecular assembly upon which to  
184 display GM3g. Orthogonal display on macromolecular protein scaffolds has the  
185 potential to mimic membrane-like multi-copy GM3g display, yet allowing control of  
186 copy-number density, site-specific conjugation and, critically, distance from the  
187 surface in terms of longer orthogonal display. The use of precise protein-editing  
188 methods via lysine (Lys)-selective (31) ‘tag-and-modify’ methods (32) allows GM3g  
189 presentation in diverse protein scaffolds (**Fig 1b**). Synthetic, protein-compatible  
190 methods were developed that accessed three presentation modes that were  
191 systematically varied for both O- vs S- glycoside display via different amidine [–  
192 C(NH)NH–], amide [–C(O)NH–], or aminoalkyl [–(CH<sub>2</sub>)<sub>2</sub>NH–] linkers all at a similar,  
193 extended nine/ten-bond length, corresponding to ~11 Å from the peptide backbone  
194 (for synthesis, refer **Document S1**). These chemistries probed diverse non-native  
195 linkage motifs with features that modulate charge/pK<sub>a</sub>, hydrogen-bonding ability and  
196 hydrophobicity that are absent from the mammalian glycome, to create longer,  
197 orthogonal glycoconjugates (LOGs).

198

199 Initial application to the model protein antigen wild-type Hen Egg Lysozyme (wtHEL)  
200 generated corresponding biochemically homogeneous LOG products HEL-[amidine-  
201 GM3g]<sub>n</sub>, HEL-[amide-GM3g]<sub>n</sub> and HEL-[aminoalkyl-GM3g]<sub>n</sub> with full (n = 6) glycan  
202 occupancy in an efficient manner (**Fig S2a–c**) and LOG products were screened for  
203 endotoxin (**Fig S2d,e; Document S1**).

204

205 Mice were immunised with the three different HEL-[X-GM3g]<sub>6</sub> LOG antigens in MPLA  
206 adjuvant. Serum IgG titres against the glycoconjugate was assayed by ELISA against  
207 a corresponding LOG constructed from an unrelated protein carrier, gp120-[amidine-  
208 GM3g]<sub>16</sub> (**Fig 1c,d**). Antibody responses against the autologous LOG were considered  
209 a proxy for overall immunogenicity, whereas responses against the heterologous LOG  
210 indicated glycan cross-reactivity. High IgG titres were detected against autologous  
211 LOGs in antisera from all LOG-immunised mice. Antisera from HEL-[amidine-GM3g]<sub>6</sub>  
212 and HEL-[amide-GM3g]<sub>6</sub> were mutually cross-reactive with each other, reflective of  
213 only a small atomic variation (O versus NH) in display (**Fig 1d**). However, strikingly,  
214 whilst antisera from the aminoalkyl LOG was cross-reactive with both amidine- and  
215 amide-LOGs, the converse was not the case: amidine- and amide-LOG antisera were  
216 not reactive with the aminoalkyl LOG. Antibody titres against the protein carrier HEL  
217 revealed high-titre IgG in all immunisation groups, indicating that all methods used to  
218 achieve LOG glycoconjugation largely preserved native HEL epitopes. Since the  
219 amidine-based LOG provided the highest homologous anti-GM3g titres, we chose to  
220 prioritise its investigation.

221

222 *Conjugation to protein is necessary for LOGs to elicit glycoconjugate-specific IgG*

223 To evaluate whether covalent linkage to the carrier protein was required for LOG  
224 immunogenicity, we compared immunisation with either LOG HEL-[amidine-GM3g]<sub>6</sub>  
225 or instead with wtHEL that had been non-covalently mixed with stoichiometrically  
226 equivalent (n = 6) amounts of a corresponding, non-conjugated ('free') side-chain-only  
227 amidine-GM3g [-C(NH)NH-GM3g] (**Fig S2f**). To test GM3g-specific effects in  
228 particular, cross-reactive IgG titres were measured via ELISA against gp120-[amidine-  
229 GM3g]<sub>16</sub>. The HEL-based LOG antiserum had substantial titres of amidine-GM3g-  
230 reactive IgG even in the absence of exogenous adjuvant (EPT = 1,430), which  
231 increased with the addition of adjuvant. However, no LOG-specific responses were  
232 detected in the groups immunised with mixed, unconjugated HEL-plus-amidine-GM3g  
233 ( $P < 0.0001$ , Tukey's post-hoc) (**Fig S2g,h**). These data imply that LOG  
234 immunogenicity is contingent on conjugation of the glycan to a protein carrier to  
235 facilitate B cell activation and isotype switching via T cell help, anticipated from  
236 classical hapten-carrier biology (33).

237

238 *Precise LOG-editing maps the role of glycan site and stoichiometry in modulating*  
239 *immunogenicity*

240 Given the robust immunogenicity of LOGs, we next set out to better understand the  
241 molecular basis for glycan moiety immunogenicity by mapping the functional roles of  
242 both glycan site and copy number in precise structure-activity relationships.  
243 Importantly, our 'tag-and-modify' LOG construction methods (32) allowed ready LOG  
244 'editing' simply via corresponding control of 'tag' site and copy number. In this way,  
245 site-directed mutagenesis of Lys to Arg allowed codon assignment whilst leaving  
246 global protein physicochemical properties including charge essentially unchanged. We  
247 designed a set of mutant HEL constructs to control the number of Lys and



248 subsequently GM3g copy number and spacing (**Fig 1e,f; Fig S3**). This set of mutants  
249 permitted the dissection of features including moiety spacing, such as proximal versus  
250 distal GM3g glycoconjugates in HEL-[–amidine-GM3g]<sub>3p</sub> and HEL-[–amidine-  
251 GM3g]<sub>3d</sub>). Notably, predicted pI values were essentially unaltered: wtHEL was 9.32,  
252 whereas HEL-null (in which all Lys were mutated to Arg) was 9.48. In this way, full  
253 control of Lys sites and copy numbers (n = 0–6) allowed editing of the GM3g in  
254 corresponding LOGs to generate a comprehensive panel of HEL LOGs (HEL-[–  
255 amidine-GM3g]<sub>0-6</sub>). These allowed dissection of the individual contributions of HEL-[–  
256 amidine-GM3g]<sub>6</sub> in what represents, to our knowledge, an unprecedented parsing of  
257 the site-specific roles of glycan moieties in probing glycoconjugate immunogenicity.  
258 Strikingly, these revealed that not only is copy number a determining factor, but that  
259 contrary to prior avidity-centric perceptions, maximal loading does not deliver  
260 maximum titres. Indeed, optimal sugar loading with respect to anti-glycoconjugate  
261 antibody production was not proportional to the number of modifications but was found  
262 to be 2–4 (for HEL-[–amidine-GM3g]<sub>2-4</sub>) in the absence of adjuvant, with significant  
263 reductions in IgG titres for HEL-[–amidine-GM3g]<sub>5</sub> and HEL-[–amidine-GM3g]<sub>6</sub>  
264 ( $P < 0.0001$ ) (**Fig 1g**). Interestingly, the glycoconjugate spacing in the case of HEL-[–  
265 amidine-GM3g]<sub>3p</sub> and HEL-[–amidine-GM3g]<sub>3d</sub> had no obvious bearing on the final  
266 GM3g-specific IgG titres.

267

268 To understand the origins of this counterintuitive outcome, we evaluated possible  
269 mechanisms. First, we tested whether the increased GM3g-specific titres arising from  
270 HEL-[–amidine-GM3g]<sub>2-4</sub> immunisation were a consequence of Lys-to-Arg mutations  
271 changing the T cell immunogenicity of the protein backbone, possibly introducing  
272 artificial T cell epitopes that enhanced the response rather than a genuine GM3g

273 loading effect. To assess this, we immunised mice with incompletely amidine-GM3g-  
274 modified wtHEL derived from chemical modification conditions adjusted to instead  
275 yield a product where the mean glycan occupancy was lowered to ~3.7 per HEL. Mice  
276 immunised with this alternative lower copy product again showed greater GM3g-  
277 specific IgG titres compared to the high copy number LOG, HEL-[amidine-GM3g]<sub>6</sub>  
278 (**Fig S4a,b**) ( $P = 0.029$ ), implying that the differential GM3g titres were unlikely to  
279 result from protein carrier amino acid substitutions impacting T cell help. Notably, HEL  
280 is a weak T cell antigen in BALB/c mice(34), and though the high IgG titres imply that  
281 sufficient T help is generated to facilitate reliable antigen-specific isotype switching,  
282 we were unable to detect Th recall responses, including in mice that had received HEL  
283 in MPLA (**Fig S4c–g**).

284

285 To further probe the relationship between glycoconjugate occupancy and the  
286 downstream humoral response, we evaluated the anti-GM3g IgM response two weeks  
287 post-prime (**Fig S4h**). These titres reflect the early humoral response which may not  
288 necessitate Th support. Although IgM titres were lower and data more dispersed  
289 compared to IgG, the trends with respect to glycan occupancy were the same, again  
290 implying that this is likely to be a Th cell-independent effect. This GM3g occupancy  
291 phenomenon was distinct from that observed against the HEL backbone, which was  
292 found to largely be adjuvant- ( $P < 0.0001$ ) rather than sugar loading-dependent ( $P =$   
293  $0.3496$ , two-way ANOVA) effect (**Fig S4i**). Collectively, these data therefore highlight  
294 that glycan occupancy may have a substantial effect on antibody outcomes,  
295 suggesting that the titration of optimal loading can be leveraged to deliver higher titres.  
296 Interestingly, HEL-[amidine-GM3g]<sub>0</sub> in which all lysines were mutated to arginine  
297 elicited a low titre anti-GM3g response (EPT = 2,940) in formulation with MPLA (**Fig**

298 **1g**). These data, along with mass spectrometric analysis (**Fig S3b**) suggest that even  
299 partial incorporation of GM3g onto the *N*-terminal primary amine is sufficient to initiate  
300 a response against the glycoconjugate.

301

### 302 *GM3g-specific antibodies raised with multiple protein carriers*

303 Having demonstrated that HEL LOGs elicit substantial IgG titres even with relatively  
304 low glycan copy numbers, we next tested the immunogenicity of the amidine-GM3g  
305 LOG on a different protein carrier, truncated gp120. This provided an excellent  
306 additional test of the LOG method, with more potential Lys 'tag' sites and a backbone  
307 that supplies multiple Th epitopes. Notably, while the total number of lysines on the  
308 gp120 construct used was 25, after application of the same benign editing methods  
309 for LOG generation, we estimated via electrophoretic analysis and densitometry data  
310 that amidine-GM3g loading delivered a mean of approximately 16 modifications  
311 (gp120-[amidine-GM3g]<sub>16</sub>, **Fig S5a,b**). This partial lysine occupancy may be a  
312 consequence of the heavy endogenous *N*-linked glycosylation on gp120 reducing the  
313 accessibility of some lysine sidechains.

314

315 To assay longitudinal outcomes, animals were immunised with gp120 or gp120-[  
316 amidine-GM3g]<sub>16</sub> and bled periodically (**Fig S5c**). gp120-[amidine-GM3g]<sub>16</sub> rapidly  
317 induced GM3g-reactive IgG even after a single immunisation in the absence of  
318 adjuvant (IgG EPTs ~10<sup>3</sup>), which further increased after boosting (~10<sup>5</sup>–10<sup>6</sup>), unlike  
319 the unmodified gp120-only counterpart ( $P < 0.0001$ ) (**Fig S5d–f**). These titres further  
320 increased with adjuvantation, with titres approximately an order of magnitude greater  
321 at the terminal timepoint ( $P = 0.006$ ). Interestingly, gp120-[amidine-GM3g]<sub>16</sub> antisera

322 displayed dramatically less antibody reactivity against the unmodified gp120 protein  
323 backbone compared with the unmodified gp120 antiserum against the unmodified  
324 gp120 protein backbone, implying that the GM3g modifications disrupted or masked  
325 immunodominant native gp120 epitopes (**Fig S5g,h**). This is consistent with GM3g  
326 ‘plugging gaps’ between the extensive native *N*-linked glycosylation sites. Similar  
327 antibody outcomes were also observed after immunisation with a corresponding LOG  
328 based on influenza A virus H1N1-NC99-HA-trimer ( $P = 0.016$ ) (**Fig S5i-k**), H1N1-HA-  
329 [-amidine-GM3g]<sub>26</sub>. These data collectively demonstrate that GM3g-reactive antibody  
330 responses may be elicited regardless of the carrier protein. These responses were  
331 also irrespective of mouse sex and genetic background (**Fig S6**).

332

### 333 *Antigen-specific T helper responses are unaltered in LOGs*

334 Any protein alteration, including the methods we used here to generate LOGs, may  
335 also affect downstream peptide processing and antigen presentation. We therefore  
336 tested the specific impact of LOGs on T cell antigen-specific recall responses. Whole  
337 spleen suspensions from gp120-[-amidine-GM3g]<sub>16</sub>-immunised mice were stimulated  
338 *in vitro* with unmodified gp120, gp120-[-amidine-GM3g]<sub>16</sub> and HEL-[-amidine-GM3g]<sub>6</sub>  
339 for 16 h (adding Brefeldin A for the final 6 h). IFN- $\gamma^+$  CD4 T cells were quantified and  
340 contrasted between the vaccination and re-stimulatory conditions (**Fig S7a-d**).  
341 Detectable antigen-specific responses were found only in the adjuvanted groups  
342 irrespective of the GM3g-presentation status of the immunogen. Moreover, the recall  
343 response was of equal magnitude whether gp120 or gp120-[-amidine-GM3g]<sub>16</sub> were  
344 used. HEL-[-amidine-GM3g]<sub>6</sub> did not induce any recall responses, confirming the  
345 important role of the conjugated carrier in providing T cell help. Together these  
346 suggested that the presentation of GM3g with LOGs did not inhibit the capacity for

347 corresponding antigen to be processed nor for corresponding T cells to recognise  
348 anchored peptide ( $P > 0.9999$ ). We further evaluated secretion of a broader panel of  
349 cytokines in supernatant after 72 h and observed similar trends in both IL-2 and IL-4  
350 (**Fig S7e–g**). As is classical in the Th2-biased BALB/c background, IgG1 was the  
351 predominant isotype, with the TLR-4/Th1-biasing MPLA adjuvant bolstering IgG2a  
352 production (**Fig S7h,i**).

353

#### 354 *Variation of the glycan in LOGs elicits orthogonal antibody outcomes*

355 Having demonstrated that GM3 LOGs may be created in forms that are strongly  
356 immunogenic for B cell responses, we tested the extension of this phenomenon to  
357 other self-glycans. We chose the Lewis group trisaccharide Lewis-X ( $\text{Le}^{\text{X}}$ ) as another  
358 representative glycan for its similar size (trisaccharidic) and yet differing sugar content  
359 and arrangement (branched, non-linear) and charge state (neutral) (**Fig 1h**).  
360 Corresponding gp120-[–amidine- $\text{Le}^{\text{X}}$ ]<sub>n</sub> LOG was constructed in an essentially  
361 identical manner and used in formulation with MPLA in identical immunisation  
362 protocols. Antibodies were similarly raised against the  $\text{Le}^{\text{X}}$  LOG, with significantly  
363 greater titres compared with animals immunised with unmodified gp120 ( $P = 0.005$ )  
364 (**Fig 1i**). Notably, antiserum raised against either corresponding  $\text{Le}^{\text{X}}$ g or GM3g LOGs  
365 were orthogonal, strictly binding autologous glycan, implying tight glycan specificity.

366

#### 367 *B cell clonality against GM3g LOG is narrow*

368 To dissect the molecular mechanisms underpinning the surprisingly robust B cell  
369 response against the LOGs, we conducted comprehensive clonotyping using animals  
370 primed with the HEL-[–amidine-GM3g]<sub>6</sub> LOG. Antigen-specific B cells were sorted

371 from mice, sorting on pre-gated IgD<sup>-</sup> B cells according to molecular probes specific  
372 either to the glycoconjugate or the protein backbone (**Fig 2a,b; Fig S8a**). Heavy chain  
373 variable regions (V<sub>H</sub>) were recovered from one mouse and sequenced from 87 events,  
374 for which the majority (80/87) were GM3g-specific (**Fig 2c**). Clonality was defined  
375 according to the inferred heavy chain VDJ gene origins (**Fig 2d; Fig S8b**). Antigen-  
376 specific events were found in the spleen and bone marrow rather than inguinal lymph  
377 nodes, suggesting that draining follicular responses had ceased by four weeks post-  
378 administration (**Fig 2e**).

379

380 The specific gene segments present in the isolated clones (**Fig 2f–h; Fig S8c**) reveal  
381 striking homology in their IGHV utilisation. In particular, the *IGHV2* subgroup was the  
382 predominant V<sub>H</sub>-gene class used in the GM3g-specific events and was expressed in  
383 > 80% of sorted B cells. The phylogenetically-related *IGHV2-3\*01*, *IGHV2-6-5\*01* and  
384 *IGHV2-9\*02* members were the most well-represented in the GM3g-binders (**Fig 2i**).  
385 By contrast, the proportionality of V-genes utilised among HEL-binding B cells was  
386 significantly more diverse. Furthermore, D- and J-gene usage was highly diverse  
387 among these clones, implying that they tolerate broad CDRH3s and joining  
388 orientations.

389

390 V<sub>H</sub>-gene utilisation was also highly related between animals, implying a striking  
391 consistency in the use of this V<sub>H</sub>-gene-dependent clonal class in facilitating LOG  
392 binding (**Fig 2j**). This was unlike the HEL-binding clones; for these a broader, more  
393 diverse set of clonotypes was isolated, fully consistent with the larger antigenic protein  
394 surface compared with the more restricted but seemingly immunodominant glycan

395 surface in corresponding LOGs (**Fig 2k**). The corresponding odds ratio that a given V-  
396 gene would be shared with respect to the antibody binding target revealed that for all  
397 animals, there is significantly narrower V-gene utilisation against LOG than the protein  
398 backbone alone (**Fig 2l**).

399

400 Given the strikingly restricted clonotypology of the anti-[ $\alpha$ -amidine-GM3g] response in  
401 the context of the broad tolerance to diverse D<sub>H</sub> and J<sub>H</sub> genes, the LOG was  
402 hypothesised to access a high frequency of naïve B cells. To interrogate this, LOG-  
403 binding naïve B cells from murine splenocytes were detected at a strikingly high  
404 frequency of 0.025% of IgD<sup>+</sup>IgM<sup>mid-hi</sup> B cells (**Fig S8d,e**). These events were  
405 sequenced from one mouse, revealing similar enrichment of the *IGHV2* subgroup  
406 (88%) compared with the immunised mice (**Fig S8f,g**).

407

408 A representative subset of several GM3g-binding IgGs from the *IGHV2*-subgroup  
409 origin were recombinantly synthesised and supernatant screened against gp120-[ $\alpha$ -  
410 amidine-GM3g]<sub>16</sub> (**Fig 2m**) – all bound specifically, confirming functionality. The best  
411 binder amongst these antibodies, termed BAR-1 with inferred germline V<sub>H</sub>-gene  
412 *IGHV2-9\*02* (**Fig. 2m**), was purified for further analysis.

413

414 *The influence of the protein backbone on B cell outcomes does not perturb narrow*  
415 *anti-glycan clonal responses*

416 Having isolated and identified the role of the *IGHV2* subgroup in the binding of HEL  
417 LOGs, we aimed to determine the effects of the protein backbone on B cell clonal  
418 outcomes. We similarly sorted B cells from gp120-[ $\alpha$ -amidine-GM3g]<sub>16</sub>-immunised



419 mice (4-weeks post-prime). B cells that bound the gp120 backbone were not identified  
420 (**Fig 3a,b**), consistent with undetectable gp120 serum antibody binding in these  
421 animals (**Fig 3c**) and other animals primed with this LOG as an immunogen (**Fig S5**).  
422 Strikingly, the V-gene usage of antibodies raised against gp120-[–amidine-GM3g]<sub>16</sub>  
423 again revealed that the *IGHV2* subgroup dominates, representing > 90% of clones  
424 (**Fig 3d,e**), of the same clonotype as that observed in the HEL-[–amidine-GM3g]<sub>6</sub>-  
425 immunised mice.

426

427 We observed in gp120-[–amidine-GM3g]<sub>16</sub>-immunised mice that a higher proportion  
428 of B cells were members of clonal families compared with HEL-[–amidine-GM3g]<sub>6</sub>-  
429 immunised mice, with an average of 2.67-fold increase in the proportion of non-  
430 singleton B cells (**Fig 3f,g**). This may imply that the gp120 protein backbone offers  
431 greater clonal expansion, probably as a function of its improved T cell immunogenicity  
432 compared with HEL (**Fig S4, Fig S7**). We further assessed the impact of the protein  
433 backbone on clonal diversity by performing a Chao1 estimate test (35, 36). While there  
434 was a trend for lower class sampling values in gp120-[–amidine-GM3g]<sub>16</sub>-immunised  
435 mice (which implies narrow clonal diversity), this was not statistically significant (**Fig**  
436 **3h**). We also observed that at four-weeks post-prime, there were some antigen-  
437 specific B cells found in the iLN (**Fig 3i**) – this was not seen in the HEL-[–amidine-  
438 GM3g]<sub>6</sub>-immunised mice and may suggest that the different protein backbone  
439 maintains activated B cells within the secondary or tertiary lymphoid organ (S/TLO)  
440 structures, where much of the antigen persists, driving increased maintenance of the  
441 follicular response. We observed in the sequences isolated from gp120-[–amidine-  
442 GM3g]<sub>16</sub>-immunised mice that the degree of SHM undergone was compartment-  
443 specific (**Fig 3j**): the mean nucleotide mismatch of VH sequences derived from the



444 lymph node was 6.8, spleen was 1.4 and bone marrow 0.8. Moreover, the extent of  
445 SHM undergone by the clones raised against gp120-[–amidine-GM3g]<sub>16</sub> were  
446 significantly greater than that against HEL-[–amidine-GM3g]<sub>6</sub> ( $P = 0.0059$ ,  
447 Kolmogorov–Smirnov test) (**Fig 3k**). These data implicate the protein backbone in  
448 determining the maintenance of the primary germinal centre (GC) reaction conditions.

449

450 To understand the cellular underpinnings of the improved GC experience of gp120-[–  
451 amidine-GM3g]<sub>16</sub>-raised clones, we measured the induction of follicular helper T (Tfh)  
452 cells with respect to protein carrier. We demonstrated that the gp120 carrier elicits a  
453 larger Tfh population (**Fig 3l–n; Fig S9**), which is coordinate with the concept that the  
454 extent of SHM experienced by the glycoconjugate-specific B cells can be toggled by  
455 changing the T cell immunogenicity of the carrier protein.

456

#### 457 *LOGs induce minimal affinity maturation despite SHM*

458 Having shown differential SHM rates with respect to the protein carrier, we next  
459 evaluated the functional effect of SHM on antibody affinity. First, we analysed the  
460 mutation frequencies across the V<sub>H</sub> gene in an unbiased manner to identify whether  
461 there were codons that were commonly mutated across the gp120-[–amidine-  
462 GM3g]<sub>16</sub>-immunised mice (**Fig S10a**) and identified that positions in CDRH1—namely  
463 T6I and S7N—were frequently mutated across multiple animals (**Fig S10b**). To  
464 evaluate the effect of these mutations, we introduced these changes into BAR-1 and  
465 screened their binding via ELISA; the data revealed no significant differences in  
466 binding compared to the wild-type mAb (**Fig S10c**), implying a lack of affinity  
467 maturation associated with these mutations. Second, we selected the largest clonal

468 family, which had undergone significant expansion and diversification and was of an  
469 inferred *IGHV2-9\*02* origin (**Fig 3o**). These antibodies were expressed recombinantly  
470 and screened via ELISA against HEL-[ $\alpha$ -amidine-GM3g]<sub>6</sub> and the EC<sub>50</sub> values were  
471 compared against that of the iGL (**Fig 3p**). Our data showed no evidence of increased  
472 affinity against the glycoconjugate, despite substantial SHM, collectively suggesting a  
473 strongly limited capacity for B cells to further improve binding against the carbohydrate.

474

#### 475 *LOGs raise a specific anti-sugar polyclonal antibody specificity*

476 To dissect anti-glycan specificity, we screened antisera derived from gp120 and from  
477 gp120-[ $\alpha$ -amidine-GM3g]<sub>16</sub> LOG against a panel of 137 mammalian glycans (**Fig 4a**;  
478 **Table S3, Table S4**)(37, 38). This broad assessment revealed strikingly focused and  
479 specific binding against only nine glycans of >220. Indeed, cross-reactivity was seen  
480 only to very subtly altered features:  $\alpha$ -2,3  $\rightarrow$   $\alpha$ -2,6  $\rightarrow$  monohydroxylated *N*-acetyl-Neu  
481  $\rightarrow$  *N*-glycolyl-Neu or OH-2-Glc  $\rightarrow$  NHAc-2-GalNAc. To further interrogate the  
482 specificity of the polyclonal antibody response, we designed a soluble ligand  
483 competition assay for the binding of the antiserum to arrayed HEL-[amidine-GM3g]<sub>6</sub>.  
484 Consistent with the glycan panel analysis, GM3g antiserum bound essentially  
485 equivalently to its OH-2-Glc- and NHAc-2-GalNAc variants (IC<sub>50</sub> of 7.76 mM and 9.90  
486 mM, respectively) (**Fig 4b,c**). Two truncated variants further mapped GM3g specificity  
487 and saccharidic moiety dependency: the disaccharide variant SiaGal competed  
488 relatively weakly (47.8 mM), implying some role for the ‘inner’ reducing-end  
489 interactions, whereas GM3g disaccharide lacking ‘tip’ non-reducing-end Sia showed  
490 no detectable competition, implying the presence of more critical contacts made with  
491 the terminal sialic acid. These findings were rationalised by our subsequent structural  
492 analysis.

493

494 Although the data imply that the complete GM3g glycan structure is a required  
495 component of antibody binding, we also observed broad, substantial contributions  
496 from differing non-reducing aglycones (**Figure 4d, left**): enhanced binding for amide  
497 and aminoalkyl aglycones was potentiated further by the presence of an amidine. Any  
498 such potentiation was notably lost in the absence of incorrect glycan (**Fig 4d, right**),  
499 further highlighting the role of tight glycan recognition in driving affinity, despite  
500 apparent engagement both of glycan and aliphatic constituents. Thus, although these  
501 data suggest that the antibody response targets the linker-glycan motif, it is  
502 nonetheless specific to the GM3g glycan.

503

504 We next interrogated the binding of GM3g LOG-raised antibodies against native GM3g  
505 display through ELISA screening gp120-[–amidine-GM3g]<sub>16</sub> antiserum against GM3  
506 and a ceramide control. Data revealed no indication of GM3-specific binding, but rather  
507 elevated non-specific reactivity with both ceramide and GM3 in an MPLA-dependent  
508 manner, potentially a function of the adjuvant mounting non-specific antibody  
509 responses with a substantial hydrophobic element (**Fig 4e,f**). To eliminate any  
510 serological background and control for the non-specific binding observed in the MPLA-  
511 adjuvanted gp120-[–amidine-GM3g]<sub>16</sub> antiserum, GM3g LOG-reactive monoclonal  
512 antibodies of an *IGHV2* origin were purified and again screened via ELISA. No binding  
513 was detected against either ceramide or GM3 in any of the 11 clones tested (**Fig 4g**).  
514 These data imply that the antibodies raised against GM3g presented synthetically in  
515 this manner fail to elicit reactivity against native glycan presentation.

516

517 *Biophysical, biochemical, and structural properties of the dominant GM3g-engaging*  
518 *clonal class*

519 We generated and purified the Fab of the GM3g-binding mAb clone, BAR-1 and  
520 quantified binding using surface plasmon resonance (SPR) against an amidine  
521 (C(NH)NH)-GM3g-coated surface, bearing the same extended side-chain motif as  
522 used in LOGs, generating a  $K_D = 17 \pm 1 \mu\text{M}$  (**Fig 5a**). Next, we synthesised an  
523 equivalent soluble ligand, Lys–amidine-GM3g, as a representative minimal LOG motif,  
524 and a truncated variant Me–amidine-GM3g and conducted solution-phase isothermal  
525 titration calorimetry (ITC), generating respective similar  $K_D = 5.4 \pm 1.2 \mu\text{M}$  (Lys–  
526 amidine-GM3g, **Fig 5b, Fig S11a**) and  $K_D = 2.1 \pm 0.7 \mu\text{M}$  (Me–amidine-GM3g, **Fig**  
527 **S11b,c**). Notably, consistent with LOG design, rather than display entropic cost, both  
528 displayed balanced binding thermodynamics ( $\Delta S = -1 \text{ kcal/mol}$  and  $-5.7 \text{ kcal/mol}$ ,  
529 respectively). Competition ELISAs using these soluble ligands were consistent with  
530 that observed using polyclonal sera, namely, that binding could be competed out using  
531 soluble GM3g, but that Me–amidine-GM3g was more competitive (**Fig 5c**).

532

533 Dynamic structural interrogation of the BAR1•Lys–amidine-GM3g complex using  
534 universal standard transfer analysis (uSTA) protein NMR (39) (**Fig 5d,e; Fig S11c–l**)  
535 gave a  $K_D = 49 \pm 10 \mu\text{M}$  and  $k_{off} = 3.77 \pm 3 \text{ s}^{-1}$ , consistent with values obtained by  
536 complementary methods (**Fig 5a,b**). Atomic-level ‘heat maps’ of magnetization  
537 transfer in uSTA revealed a ligand pose with primary engagement of BAR1 with the  
538 glycan motif of GM3g over the Lys–amidine-linker moiety. Interestingly, analyses of  
539 the interaction of two truncated ligand variants – GM3g itself and just the tip  
540 disaccharide Neu5Ac-Gal (**Fig 5e–g**) – further identified relaxation of GM3g alone into  
541 a pose that creates even greater contact of the Gal upon removal of the LOG longer-

542 linker moiety in Lys–amidine-GM3g. This suggested topological frustration in the  
543 complex with Lys–amidine-GM3g (and by extension the LOG) that, when removed,  
544 allows a relaxation further into the binding motif.

545

546 Next, the atomic level basis of these interactions was probed through complementary  
547 methods, allowing structural analysis of BAR-1 in complex with Lys–amidine-GM3g.  
548 Crystallization of the BAR1•Lys–amidine-GM3g complex revealed a striking,  
549 seemingly LOG-specific arrangement in the 3D-structure of the *holo* complex (**Fig 5h,I**  
550 and **Fig S12a**). Notably, consistent with design, the longer length of the LOG moiety  
551 allowed the GM3g to ‘reach through’ a seemingly flexibly-engaged CDRL3 region to  
552 engage key residues in CDRH2, and, also, to some extent CDRH1, leaving the part of  
553 the groove formed by CDRH3, CDRL1 and CDRL2 unoccupied. The antibody binding  
554 pocket is largely hydrophobic in character.

555

556 The crystal contained two complete copies of the complex which are largely identical  
557 (rmsd of light chain 0.5 Å). In both the electron density is well ordered for all three  
558 sugar rings and the amidine of Lys–amidine-GM3g, but less well ordered for the ‘reach  
559 through’ lysine side-chain. As a seemingly key ‘foothold’ the indole of Trp57<sub>H</sub> stacks  
560 against the alpha-face of the Gal sugar of GM3g to create a classical pi-CH interaction  
561 (**Fig 5i** and **Fig S12c**) found in diverse so-called carbohydrate modules (CBMs) (40,  
562 41). This is supported by binding of the tip Neu5Ac sugar of GM3g, which makes five  
563 hydrogen bonds to BAR1 backbone, including a striking bidentate interaction of its C-  
564 1 carboxylate with amide nitrogen atoms of Ala58<sub>H</sub> and Val59<sub>H</sub> but notably there is no  
565 charge-driven interaction. Several highly coordinated water molecules (W) also

566 contribute to binding, as well as a hydrophobic pi-CH interaction with Phe37<sub>H</sub>. The  
567 reducing-end Glc of GM3g also makes hydrogen bonds to three water molecules, two  
568 of which bridge to the protein (including W3 which bridges to Tyr99<sub>L</sub>, Asn63<sub>H</sub> and  
569 galactose) but only three direct van der Waal contacts with the protein. The amidine  
570 linkage of Lys–amidine-GM3g makes hydrogen bonds to the protein (Tyr97<sub>L</sub>) and to a  
571 water molecule (W5) that bridges to the glucose and, intriguingly, a cation-pi  
572 interaction with Tyr37<sub>L</sub> confirming a contribution from the longer amidine linker to  
573 binding. The aliphatic side chain of the lysine makes van der Waal contacts with  
574 Tyr97<sub>L</sub>.

575

576 To probe specific contributions to binding, including the ‘foothold’ Trp57<sub>H</sub>, we probed  
577 the residues lining the binding site of BAR-1 through Ala-scanning mutagenesis. uSTA  
578 protein NMR allowed us to look at the modulation of the binding pose adopted by Lys–  
579 amidine-GM3g. Strikingly, whilst alterations of lining residues Phe37<sub>H</sub>, Val59<sub>H</sub>,  
580 Tyr103<sub>H</sub> and Tyr99<sub>K</sub> retained residual binding in an ELISA format (**Fig S14**), their  
581 interaction surfaces were all essentially similar (**Fig 5j**). By contrast, no interactions at  
582 all were observed between Lys–amidine-GM3g and ‘foothold’ mutant Trp57Ala, further  
583 emphasizing its key role (**Fig S14c**). Specifically, when exciting the ‘ligand only’  
584 sample at 8 ppm, small residual signal is seen in the STD spectrum. This was found  
585 to be of identical magnitude to the spectrum of the Trp57<sub>H</sub> BAR-1 mutant, revealing  
586 that there was no detectable binding between ligand and protein.

587

588 Together, these structural and biophysical analyses highlight the key residues  
589 important in driving binding. These residues were notably conserved amongst the

590 *IGHV2* subgroup-containing clones that we had validated for GM3g binding, with  
591 particularly high sequence similarity in their V<sub>H</sub>-encoded CDRH1 and CDRH2 loops  
592 (**Fig 5k**), and are also consistent with our mutagenic and structural analyses. These  
593 data also showed that involvement of CDRH3 in ligand binding was limited, which  
594 aligns not only with our structural analyses but also with the observation from our  
595 broad set of B cells isolated against the ligand and their tolerance of exceptionally  
596 diverse CDRH3.

## 597 Discussion

598 For self-glycans, there is a heavy incentive to skew the naïve B cell repertoire to avoid  
599 the presence of self or self-like glycan-reactive B cells to prevent generation of  
600 autoreactive antibodies (1, 8), as supported by evidence of anti-glycan responses  
601 associated with various autoimmune conditions (42, 43). Notably, previous studies  
602 have failed to reliably raise high-titre antibodies responses against GM3 using  
603 conventional autologous formulations (21, 26, 44).

604

605 The LOG modular format has potential advantages compared with immunisation with  
606 autologous GM3, namely: i) the docking of the sugar to a peptidic carrier allows for  
607 associated T cell help, and ii) non-native presentation of otherwise immunorecessive  
608 TACAs via a bespoke chemical linker may bypass the tolerogenic constraints that  
609 prevent antibodies being raised against native glycan presentations in endogenous  
610 glycoconjugates. Our discovery that GM3g-specific IgG responses were readily  
611 mounted in a mouse (predominantly by the *IGHV2* subgroup) reveals that the LOG  
612 modular format of self-glycans can access a subset of naïve B cells that native  
613 presentations of the same glycan do not.

614

615 We have rationalised the lack of native glycoconjugate cross-reactivity by combining  
616 immunogenetic, structural, biochemical and biophysical-based analyses. The  
617 structure of the BAR-1 Fab with Lys–amidine-GM3g reveals that that the sugar portion  
618 is recognised by the CDR1 and CDR2 loops in the  $V_H$  domain. Intriguingly, the  
619 recognition of the galactose and sialic acid sugars closely resembles (**Fig S13**) the  
620 arrangement seen for a hexasaccharide binding antibody (45). Although showing a



621 different ‘reach through’ orientation, there is striking conservation of several features  
622 of the final tip glycan interactions: primarily a classical pi-CH interaction between the  
623 CH beta-face of Gal and the indole ring of Trp (Asensio et al., 2013), the recruitment  
624 of bridging water molecules, a bidentate hydrogen bond between the carboxylate of  
625 the terminal sialic acid with the protein backbone and the engagement of the methyl  
626 of the *N*-acetyl by an aromatic residue. We suggest this serves as generic ‘foothold’  
627 (focused on the Trp) in the antibody, and by implication BCR, for these two sugars.  
628 With LOGs, the linker of Lys–amidine-GM3g reaches through to the sugar-binding  
629 pocket by spanning the V<sub>H</sub>V<sub>L</sub> interface to be recognised by the CDR3 loop of the V<sub>L</sub>  
630 domain. In native GM3 the trisaccharide has a glycosidic link in close proximity to the  
631 branched arrangement of the ganglioside (displaying stearic acid and sphingosine). A  
632 ‘reach through’ mode observed for Lys–amidine-GM3g would therefore be unavailable  
633 to ‘self’ GM3. Close embedding of GM3 into the membrane creates prohibitive van der  
634 Waal clashes for antibody engagement.

635

636 The surprising immunogenicity of the LOG amidine-GM3g is, we hypothesise, a  
637 function of the set of V(D)J configurations tolerated by the BAR-1 clonal class. The  
638 binding of a naïve B cell receptor to an epitope is dependent on the combinatorial  
639 effect of the distinct V(D)J configuration of the cell. However, the structural  
640 dimensionality imparted by the germline configuration — a function of the unique  
641 combinatorial gene segment composition — is, we show here, drastically reduced in  
642 this LOG-raised clonal class. These clones tolerate highly diverse sets of D<sub>H</sub> and J<sub>H</sub>  
643 segments, and our structural characterisation demonstrates that the corresponding  
644 CDRH3 has no substantial contribution to glycan recognition.

645

646 This suggests a clone-by-clone basis under which these anti-TACA antibody elicitation  
647 approaches ought to be considered. Taken together, our data now suggest that future  
648 approaches necessitate analyses of how a given glycoconjugate might engage  
649 TACA:linker specific B cell clonotype, using comprehensive structural and biochemical  
650 analysis of clonal reactivity, consideration of increased scope for SHM-directed affinity  
651 maturation, and determination of native-recognition in relation to context-dependent  
652 TACA-specific recognition.

653

654 Use of different protein platforms (e.g. HEL-[ $\alpha$ -amidine-GM3g]<sub>6</sub> and gp120-[ $\alpha$ -amidine-  
655 GM3g]<sub>16</sub>) led to the same germline configurational response against the GM3g self-  
656 glycan moiety. This was also associated with a clear modulation of the germinal centre  
657 experience the clones had undergone with respect to the amount of follicular help  
658 offered by the different protein backbones. There was a direct relationship between  
659 the amount of T cell help detected and the ensuing antigen-specific B cell IgG  
660 response. This in turn suggests future application using heterologous boosters: yet  
661 more T-immunodominant backbones conducive to greater somatic hypermutation  
662 rates may then lead to clonal affinity maturational outcomes. Moreover, heterologous  
663 immunisation strategies based on TACAs presented in the context of systematically  
664 varied LOGs conjugated to different chemical linkers might exploit affinity maturation  
665 processes to 'walk' clones towards native glycan reactivity. This is unlike classical  
666 germline-targeted approaches, which use isolated and highly mutated antibodies of  
667 known functional effect as a template germline clonal class (15, 47). However, in the  
668 instance described here, it is not known whether 'up-mutation' of the BAR-1 class can  
669 move towards a functional effect to yield native GM3g recognition.

670

671 Finally, the explicit demonstration also of the presence of germline-encoded lectin-like  
672 motifs (48–50) present in the murine BCR germline is striking. This not only challenges  
673 the dogma associated with the perceived poor immunogenicity of glycans (51) but may  
674 also provide an explanation for the greatly divergent views and results that have in the  
675 past been obtained from immunisations with glycoconjugates. Not only may this be a  
676 consequence of conjugate presentation format (e.g. ‘parallel’ *versus* ‘orthogonal’ or  
677 shorter *versus* longer linkage), as we argue here, but may also be a consequence of  
678 the restricted clonotypic response that we have discovered here. It may be that only  
679 upon engagement of the correct glycoconjugate or glyco-epitope would a large  
680 proportion of naïve B cells be activated by using appropriate ‘predisposed’ germline  
681 BCRs, thus improving the frequency of B cell activation events *in vivo* and explaining  
682 the relatively high titres of anti-GM3g antibodies elicited after a limited immunisation  
683 regimen. We therefore propose that the logical design of the entire conjugate and not  
684 just, for example, the glycan as has been typical, is important to properly exploit these  
685 rare, correct engagement events in the effective design of future immunogens.  
686

## 687 **Methods**

### 688 *Generation of synthetic GM3 liposomes*

689 All lipids were from Avanti Polar lipids. Lipid mixtures (as indicated in the results) in  
690 chloroform at 1 mg/mL were dried under the flow of nitrogen, rehydrated with buffer  
691 (150 mM NaCl, 10 mM HEPES, 2 mM CaCl<sub>2</sub>) and vortexed to form multilamellar  
692 vesicles. Then, suspension of the multilamellar vesicles were sonicated at power 3,  
693 duty cycle 40% for 10 mins using Branson Sonifier 250. Liposomes were stored under  
694 nitrogen atmosphere.

695

### 696 *Recombinant protein expression*

697 Proteins were expressed recombinantly in-house using the HEK 293Freestyle  
698 expression system (Life Technologies). Cells were transfected using PEI Max  
699 (Polysciences) and relevant expression vector. Proteins were purified from  
700 supernatant using either Protein A agarose beads (Life Technologies), or  
701 immunoaffinity chromatography (D1.3 for HEL and 2G12 for gp120) where columns  
702 were prepared with AminoLink Plus resin (Life Technologies), both used according to  
703 the manufacturer's instruction. Purified protein was tested for endotoxin contamination  
704 prior to immunisation. These analyses were conducted using either a HEK293T TLR4-  
705 CD14-MD2 IL-18 reporter line (Invivogen) or the RAW-Blue Cell assay (Invivogen).  
706 These readouts were acquired according to the manufacturer's protocol. Protein  
707 preparations where the reporter endotoxin readouts were less than the < 0.125 ng/mL  
708 LPS control was considered clean.

709

### 710 *Synthesis of LOGs*

711 Details of chemical synthesis and characterisation are outlined in **Document S1**.

712 **HEL-[amidine-GM3g]<sub>6</sub>**: GM3g-SCN **1** (404 mg, 587  $\mu$ mol) was activated in sodium  
713 methoxide solution (20 mM in 29 ml, 1.0 eq. of CH<sub>3</sub>ONa) by following the standard  
714 protocol (as shown in the preparation of GM3g-imidate **7**). After stirring for 4 days at  
715 room temperature, THF (87 mL) or ether (29 mL) was added to precipitate the sugar.  
716 The white solid was separated by centrifugation, the supernatant was discharged and  
717 the white residue was then dried under vacuum before being used immediately for  
718 protein modification.

719  
720 The precipitated sugar was dissolved in PBS buffer (2.8 mL, pH = 7.4). A fresh solution  
721 of protein in PBS buffer (0.7 mL, 5 mg/mL, pH = 7.4) was added (final concentration  
722 of HEL was 1 mg/mL) and the mixture was incubated at 25 °C for 12 h (checked by  
723 SDS-PAGE and LC\_MS if necessary). The reaction was desalted by PD-10 column  
724 twice (note: desalting is not sufficient to completely remove sugar. Excess sugar was  
725 in the post fraction; GM3g-CN **1** could be recovered by purification). Dialysis was  
726 subsequently performed in PBS buffer at 4 °C (4 h  $\times$  2 and 12 h  $\times$  1). The solution was  
727 concentrated, sterilized and stored at 4 °C. Concentration of HEL-[amidine-GM3g]<sub>6</sub>  
728 was analyzed by BCA assay (7.13 mg/mL, 0.6 mL, endotoxin free-PBS buffer, pH =  
729 7.4). Notably, modifications using [amidine-Le<sup>x</sup>] were prepared in an essentially  
730 identical manner.

731  
732 **HEL-[aminoalkyl-GM3g]<sub>6</sub>**: A mixture of HEL (11ul, 220ug, 20mg/ml in PBS, pH =  
733 7.4), GM3gOCH<sub>2</sub>CHO (41.56 ul, 30 mg/ml in H<sub>2</sub>O, 20 eq./lysine, 6\*lysine), freshly  
734 prepared NaBH<sub>3</sub>CN solution (23.19 ul, 5 mg/ml in H<sub>2</sub>O, 20 eq./lysine, 6\*lysine), topped  
735 to 220ul with H<sub>2</sub>O (144.25 ul, final protein concentration was 1mg/ml) was incubated

736 in 37 °C for 24 h without shaking. Then solution was immediately dialyzed to PBS at  
737 4 °C.

738

739 **HEL-[~~-amide~~-GM3g]<sub>6</sub>**: 15.83 ul of **S-Short-NHS** (8 eq./lysine, 60mg/ml in DMSO,  
740 freshly prepared) was added into HEL (50ul, 1mg, 20mg/ml in PBS, pH = 7.4) solution,  
741 mixed and incubated for 3 h at room temperature. Following immediate desalting into  
742 water, protein was concentrated to 2 mg/ml and checked with LC-MS. To a solution of  
743 dimeric GM3g **28** (1 mg) in water (75.6 ul), 1.54 ul of TCEP solution (1.0 eq, 0.5 M in  
744 water, freshly prepared from TCEP (free acid) solid, neutralized by 3 eq. of NaOH)  
745 was added and, after incubation at room temperature for 30 mins, transferred it into  
746 287 ul of iodo-HEL solution (574 ug, 2mg/ml, 6.4 eq. GM3g-SH/Lysine) above. 172 ul  
747 of sodium borate (100 mM, pH = 8.5) was added, topped to 1mg/ml (protein solution)  
748 with 37 ul of water. This mixture was incubated at RT for 3 hours (LC-MS checking)  
749 before being dialysed against PBS (pH = 7.4), concentrated, sterilized, and analyzed  
750 by BCA assay.

751

## 752 *SDS-PAGE*

753 Proteins were evaluated via SDS-PAGE. Samples were ran on precast NuPAGE Bis-  
754 Tris gels (Life Technologies) with 1X MOPS buffer (Life Technologies) according to  
755 the manufacturer's instructions. Proteins were stained using InstantBlue Coomassie  
756 protein stain (Abcam). Molecular weights were manually estimated according to the  
757 Novex Sharp protein marker. To obtain sharper bands on native N-linked glycosylated  
758 proteins, such as gp120, sample was pre-treated with PNGase F (New England  
759 Biolabs) according to the manufacturer's instructions.

760

761 *Primary amine ELISA*

762 Following the conjugation of the synthetic glycoconjugates to a carrier protein, relative  
763 free amines were contrasted to estimate the degree of lysine modification. This  
764 protocol was adapted (52). Briefly, protein samples (5 µg) in 10 µL PBS were mixed  
765 in 40 µL 0.1M sodium bicarbonate buffer. 5% solution of 2,4,6-trinitrobenzenesulfonic  
766 acid (TNBSA) was diluted 1:500 in the bicarbonate buffer, and 25 µL of this mixture  
767 was added to the protein sample. After 2 h incubation at 37°C. 25 µL 10% SDS and  
768 12.5 µL of 1M HCl was added. The absorbance at 335 nm was measured using a  
769 Spectramax M5 spectrophotometer.

770

771 *Animal experimentation*

772 Wild-type pathogen-free 6-week-old BALB/c mice were purchased from Charles River.  
773 Animals were monitored daily and provided standard chow and water *ad libitum*.  
774 Immunisation schedules are outlined in the results, and mice were bled periodically  
775 from the tail vein. Animals were sacrificed via a rising CO<sub>2</sub> gradient and subsequent  
776 cervical dislocation schedule 1 procedure.

777

778 *ELISA assays*

779 Samples were serially diluted and incubated on antigen-coated and blocked  
780 SpectraPlate-96 HB (PerkinElmber) plates. Antibodies were detected with either anti-  
781 mouse IgG-HRP (STAR120P, Bio-Rad), anti-mouse IgG1-HRP (STAR132P, Bio-  
782 Rad), anti-mouse IgG2a-HRP (STAR133P, Bio-Rad), anti-mouse IgM-HRP (II/41, BD  
783 Bioscience), or anti-human IgG-HRP (Jackson ImmunoResearch). ELISAs were

784 developed using 1-Step Ultra TMB ELISA substrate (Life Technologies), with the  
785 reaction being terminated with 0.5M H<sub>2</sub>SO<sub>4</sub>. For competition ELISAs, sample was pre-  
786 incubated with the coating antigen for 1 h, before adding the competitor for an  
787 additional hour for a new equilibrium to be reached. Detection and development were  
788 subsequently conducted, as per the direct ELISA protocol. Cytokine ELISAs were  
789 performed using commercially available kits. IL-2, IL-4 and IFN- $\gamma$  in supernatant was  
790 measured according to the manufacturer's protocol (all Life Technologies).

791 Optical densities were measured at 450 and 570 nm on the Spectramax M5 plate  
792 reader (Molecular Devices). After background subtraction, logistic dose-response  
793 curves were fitted in GraphPad Prism. Endpoint titres were determined as the point at  
794 which the best-fit curve reached an OD<sub>450-570</sub> value of 0.01, a value which was always  
795 > 2 standard deviations above background.

796

#### 797 *Intracellular cytokine analysis*

798 Whole splenocyte suspensions (5 x 10<sup>5</sup> cells in 200  $\mu$ L in a flat-bottom 96-well plate)  
799 were stimulated *in vitro* with 10  $\mu$ g/mL antigen in cRPMI for 16 h. For the final 6 h,  
800 5  $\mu$ g/mL brefeldin A (Biolegend) was added to all wells to suspend ER–Golgi trafficking  
801 and block cytokine secretion. Cells were subsequently washed with PBS with ice cold  
802 2mM EDTA and stained with TruStain FcX Plus (Biolegend) and LIVE/DEAD Fixable  
803 Blue on ice for 30 minutes. Surface markers were subsequently stained on ice for 20  
804 minutes: anti-mouse CD3-PE (1:200, 17A2, Biolegend), anti-mouse CD4-APC (1:200,  
805 RM4-5, Biolegend) and anti-mouse CD8-AF700 (1:200, RPA-T8, Biolegend).  
806 Following treatment with fixation and permeabilization buffers (Biolegend), cytokine was  
807 stained for using anti-mouse IFN- $\gamma$ -PE/DAZZLE 954 (1:100, XMG1.2, Biolegend)—



808 this was conducted on ice for 40 minutes. Cells were washed twice with FACS buffer  
809 (PBS with 2% FCS and 0.05% sodium azide; FB) before acquiring data on the BD  
810 Fortessa X-20 (BD Bioscience).

811

### 812 *Vaccine-specific B cell isolation*

813 Antigen probes were synthesised via modification using NHS-esterified biotin, AF647  
814 or AF488 protein modification kits, as per the manufacturer's instructions (Life  
815 Technologies). Successful modification was confirmed by both SDS-PAGE and  
816 subsequent fluorescent gel scanning on a ChemiDoc (Bio-Rad), as well as via mass  
817 spectrometry.

818 Immunised BALB/c mice were immunised 4 weeks prior to B cell isolation. Single cell  
819 suspensions were generated from the spleen, inguinal lymph nodes and bone marrow  
820 (femur and tibia). Fc receptors were blocked and stained with LIVE/DEAD Fixable  
821 Blue, as outlined earlier. The following surface stain cocktail was prepared: anti-mouse  
822 F4/80-PE (1:200, BM8, Biolegend), anti-mouse Gr-1 (1:200, RB6-8C5, Biolegend),  
823 anti-mouse CD3-PE (1:200, 17A2, Biolegend), anti-mouse CD4-PE (1:200, RM4-5,  
824 Biolegend), anti-mouse CD8-PE (1:200, RPA-T8, Biolegend), anti-mouse B220-  
825 eFluor450 (1:100, RA3-6B2, BD Biosciences), anti-mouse IgD-AF700 (1:200, 11-  
826 26c.2a, Biolegend), anti-mouse IgM-PE/Cy7 (1:200, R6-60.2, BD Biosciences), anti-  
827 mouse IgG1-FITC (1:200, A85-1, BD Biosciences), anti-mouse IgG2a/2b-FITC (1:200,  
828 R2-40, BD Bioscience), antigen probes as indicated in the results (10 µg/mL). Cells  
829 were stained on ice for 1 h. Cells were washed with FB and sorted immediately on a  
830 BD FACS Aria Fusion (BD Biosciences). LIVE/DEAD<sup>-</sup>DUMP-B220<sup>mid-hi</sup>IgD<sup>-</sup>  
831 (IgM<sup>+</sup>/IgG<sup>+</sup>)Ag<sup>+</sup> B cells were singly sorted into MicroAmp Optical 96-Well PCR Plates

832 (Life Technologies) containing 5  $\mu$ L 1X TCL buffer supplemented with 1% 2-ME.  
833 Immediately following sorting, plates were centrifuged at 1,500 g for 1 minute. Plates  
834 were stored at -80°C until use.

835

### 836 *B cell receptor variable region recovery*

837 Recovery of the antigen-specific B cell receptor variable regions was carried out,  
838 adapted from previous publications (53, 54). We are happy to share a detailed step-  
839 by-step protocol upon request. Briefly, single cell lysates were thawed on ice and RNA  
840 was captured on RNAClean XP beads (Beckman Coulter), subsequently washing with  
841 70% ethanol. RNA was eluted and cDNA libraries were synthesised using SuperScript  
842 III (Life Technologies) with random primers (Life Technologies). VH and VK regions  
843 were recovered using the first PCR primer sets (**Table S1**) and Q5 polymerase. VH  
844 amplicons were purified and sequenced using 5' MsVHE. These sequences were  
845 used to determine B clonality.

846 To confirm the recovered sequences were truly antigen-specific, antibodies were  
847 synthesised recombinantly. To incorporate the variable regions into an expression  
848 vector, vector-overlapping adapters were incorporated via PCR (**Table S1**), and the V  
849 regions were inserted into cut backbone vectors (heavy chain: FJ475055; kappa  
850 chain: FJ75056) via Gibson reaction (New England Biolabs). Successful clones were  
851 prepared, and vector products were transiently transfected into HEK 293Freestyle  
852 cells, as previously outlined.

853

### 854 *Immunogenetic analyses*

855 Immunogenic analyses were performed on the VH regions of successfully recovered  
856 clones (**Table S2**). Sequences were screened against the mouse germline gene  
857 segment repertoire using the Immunogenetics Information System (IMGT;  
858 [https://www.imgt.org/IMGT\\_vquest/input](https://www.imgt.org/IMGT_vquest/input)). These outputs were used to determine both  
859 clonality (i.e. their inferred V(D)J configuration) and the SHM rates. Sequences that  
860 returned either no result (indicative of poor sequence read quality) or was unproductive  
861 (e.g. premature stop codon) was excluded from our analysis. Clonal lineage trees were  
862 determined using GCTree (55) and rendered on Adobe Illustrator.

863

#### 864 *Glycan array*

865 Glycan arrays were custom printed on a MicroGridII (Digilab) using a contact  
866 microarray robot equipped with StealthSMP4B microarray pins (Telechem) as  
867 previously described (38). Briefly, samples of each glycan were diluted to 100  $\mu$ M in  
868 150 mM Na<sub>3</sub>PO<sub>4</sub> buffer, pH 8.4. Aliquots of 10  $\mu$ l were loaded in 384-well plates and  
869 imprinted on NHS-activated glass slides (SlideH, Schott/Nexterion), each containing  
870 6 replicates of each glycan. Remaining NHS-ester residues were quenched by  
871 immersing slides in 50 mM ethanolamine in 50 mM borate buffer, pH 9.2, for 1 hr.  
872 Blocked slides were washed with water, centrifuged dry, and stored at -20 °C until use.  
873 Serum samples were diluted 1:200 in PBS + 0.05% Tween-20 and applied directly to  
874 the array for 1h incubation. Following 1h, samples were rinsed from the array surface  
875 by dipping 4 x each in PBS-Tween, PBS and deionized water, respectively. Washed  
876 arrays were reprobed with anti-mouse-IgG-AlexaFluor488 (10ug/mL) for 1h  
877 incubation. Following secondary incubation, arrays were washed again by dipping 4 x  
878 each in PBS-Tween, PBS and deionized water, respectively, and dried by

879 centrifugation. Dried slides were scanned for 488 signal on an Innoscan 1100AL  
880 scanner (Innopsys) and signal intensities were calculated using Mapix (Innopsys) and  
881 graphed using Excel (Microsoft).

#### 882 *Surface plasmon resonance*

883 SPR was performed using a Biacore T200 instrument. GM3g-IME was immobilised  
884 onto an S-CM5 sensor chip as previously described (39). For analysis of Fab binding,  
885 serial dilutions were sequentially injected at a flow rate of 10  $\mu$ L/minute. An anti-c-  
886 Myc Fab (clone: 9E10) was used as a negative control.

887

#### 888 *Isothermal titration calorimetry*

889 Affinities of Fab BAR1 for the Amidine-GM3g and Lys–amidine-GM3g were measured  
890 by isothermal titration calorimetry using an automated PEAQ-ITC instrument  
891 (MicroCal) at 25 °C. Titrations were performed using 18  $\times$  2  $\mu$ L injections of 200-300  
892  $\mu$ M of the polysaccharide into 20-30  $\mu$ M of the protein in PBS buffer. The heats of  
893 dilution measured from injection of the ligands into the buffer were subtracted, and  
894 titration curves were fitted with one-site binding model.

895

#### 896 *Universal standard transfer analysis*

897 All NMR experiments were recorded at 298K on a 950-MHz spectrometer with Bruker  
898 Avance III HD console and 5-mm TCI CryoProbe, running TopSpin 3.6.1 and using a  
899 SampleJet. All ligands in this work were first assigned using selective 1D Hartmann-  
900 Hahn TOCSY and HSQC experiments.

901

902 The uSTA experiments were either recorded with the same stddiffesgp.2 as previously  
903 described (39), or a pseudo 3D version that used an inputted file vdlst to increment  
904 the saturation times. The number of points were set to 32768 or 65536 and sweep  
905 width to 16.05ppm for an acquisition time of 2.150s and 4.300s. All spectra were  
906 processed using nmrPipe within the uSTA workflow as previously described (39),  
907 resulting in 'transfer efficiencies' that quantify the strength of the saturation transfer  
908 and inform on the binding pose in the complex.

909  $K_D$  and  $k_{off}$  rates for amidine-lysine/WT BAR-1 complex were obtained by repeating  
910 the uSTA analysis over a range of protein/ligand concentrations and globally analysing  
911 the build-up curves for the NAc proton (**Fig S11**) as described previously (39). The  
912 interaction surface for the X-ray data was calculated from the structure using a  $\langle 1/r^6 \rangle$   
913 expectation value between each proton in the ligand, and all protons in the protein, as  
914 described previously.

915

916 In Lys–amidine-GM3g, we would anticipate a range of  $R_1$  relaxation times which could  
917 affect the transfer efficiencies. To address this, we measured the  $R_1$  and  $R_2$  relaxation  
918 rates of each proton of the ligand and developed a correction that allowed us to rescale  
919 the transfer efficiencies to account to variations in the relaxation rate. The adjustments  
920 to the interaction surfaces by performing this operation were modest (**Fig S11**).

921

922  $R_1$  and  $R_2$  relaxation rates were recorded using bespoke pulse sequences derived  
923 from the Bruker t1ir and cpmg sequences, with water suppression achieved by using  
924 excitation sculpting from the zgesgp sequence. The  $R_1$  experiment employed the  
925 zgesgp phase cycle had no cycle on the inversion pulse. The  $R_2$  experiment employed  
926 the zgesgp phase cycle with a y, -y pulse sequence on the CPMG refocusing pulse,

927 which was performed before the water suppression sequence to avoid interference. In  
928 the final measurements, the interscan delay was set to 5s to allow significant relaxation  
929 of protons. Both relaxation spectra were recorded with 8 transients and 4 dummy  
930 scans per increment, 65536 acquisition points and a sweep width of 15.96 ppm  
931 (950MHz) for an acquisition time of 2.163 s. Spectra to obtain  $R_1$  were acquired with  
932 13 delays: 5, 0.001, 0.05, 0.1, 0.25, 0.5, 0.8, 1, 1.5, 2, 3, 4 and 5s. Spectra to obtain  
933  $R_2$  were recorded with 12 delays using a spin-echo time of 800us (2x400us) per cycle  
934 and 0, 400, 40, 80, 120, 160, 200, 240, 280, 320, 360 and 400 cycles. The 90-degree  
935  $^1\text{H}$  times were calibrated manually.

936

937 To perform the correction of the transfer efficiencies, we first took the fitting parameters  
938 obtained from the full  $K_D$  analysis of Lys–amidine-GM3g / BAR-1 complex interaction.  
939 The  $R_1$  rate for the NAc proton in the ligand ( $0.37\text{ s}^{-1}$ ) was essentially identical to that  
940 measured using the  $R_1$  experiment ( $0.4\text{ s}^{-1}$ ), providing independent support for our  
941 analysis. We then simulated the transfer efficiencies expected as we systematically  
942 vary  $R_1$  and  $R_2$ . The expected transfer efficiency was largely invariant of  $R_2$ , but could  
943 vary by a factor of 2 as  $R_1$  varies by one order of magnitude. We used these curves to  
944 interpolate the expected transfer efficiency as both  $R_1$  and  $R_2$  tend to zero, thus to a  
945 reasonable first approximation, removing variation in ligand relaxation between atoms  
946 from the uSTA measurement. The  $R_1$  correction was more significant than the  $R_2$   
947 correction. The largest variation of  $R_1$  in the dataset was a factor of 3.5, ranging from  
948  $0.4\text{ s}^{-1}$  (NAc) to  $1.4\text{ s}^{-1}$ , leading to only modest adjustments of the transfer efficiency.

949

950 *X-ray crystallography*

951 BAR-1 Fab was loaded onto a gel filtration Superdex 200 column (GE Healthcare) in  
952 10 mM Tris-HCl, pH 7.5, 150 mM NaCl. Co-crystals appeared at 20 °C after a few  
953 weeks from a hanging drop of 0.1  $\mu$ L of protein solution (19 mg/mL with 1.2 mM Lys–  
954 amidine-GM3g) with 0.1  $\mu$ L of reservoir solution containing 30% (w/v) PEG MME 2000,  
955 0.1 M potassium thiocyanate in vapor diffusion with reservoir. Crystals were frozen  
956 with the same solution containing 16% glycerol and 4 mM Lys–amidine-GM3g . Data  
957 were collected at the Diamond light source oxfordshire (beamlines I24). Data were  
958 processed with XIA2 (56–60). Structure has been solved by molecular replacement  
959 using PHASER and pdb file 6ug7 (45), and the structure was builded with Autobuild  
960 program, refined with REFINE of PHENIX with NCS restraints (61) and adjusted with  
961 COOT (62). Coordinates and topologies of ligands were generated by PRODRG (63)  
962

963 Structures were refined at 1.9 Å. Two Fab molecules are present in the asymmetric  
964 unit (H/L and A/B). The two molecules are very similar (rmsd of 0.4852 Angstroems  
965 for 213 residues). Final refinement statistics are given in Table 1. Atomic coordinates  
966 and structure factors have been deposited in the Protein Data Bank (8BJZ). The  
967 quality of all structures was checked with MOLPROBITY (64). The Ramachandran  
968 statistics are as follows: 98% favoured and 2% allowed.

969

#### 970 *Data processing and statistical evaluation*

971 Flow cytometry data was evaluated in FlowJo V.10.8. Graphs were generated in  
972 GraphPad Prism V9.4 and using GCTree (55), and later edited in Adobe Illustrator for  
973 aesthetics. Statistical analysis was conducted either in GraphPad Prism V.9.4 or in  
974 RStudio V.4.1. Chao1 estimates were performed using the RStudio iNEXT package

975 (65). Statistical test details are provided in the results, figures and associated figure  
976 legends.

977

## 978 **Acknowledgments**

979 This work was supported by the Rosetrees Trust  
980 Interdisciplinary Award ID2020/100023. Upgrades of 600-MHz and 950-MHz NMR  
981 spectrometers were funded by the Wellcome Trust (grant ref: 095872/Z/10/Z) and the  
982 Engineering and Physical Sciences Research Council (grant ref: EP/R029849/1),  
983 respectively, and by the University of Oxford Institutional Strategic Support Fund, the  
984 John Fell Fund, and the Edward Penley Abraham Cephalosporin Fund. This project  
985 has received funding from the European Research Council (ERC) under the European  
986 Union's Horizon 2020 research and innovation programme (grant agreement  
987 101002859). E.S. is supported by Swedish Research Council Starting Grant (2020-  
988 02682). QJS is a Jenner Vaccine Institute Investigator and a James Martin School  
989 Senior fellow. The Chemistry theme at the Rosalind Franklin Institute is supported by  
990 the EPSRC (V011359/1 (P)).

991

## 992 **Author contributions**

993 Conceptualization of project: LPD, TS, BGD, QJS

994 Methodology: LPD, XX, AK, LM, CLB, HS, CG, RS, KL, ES, JP, AJB, JN, TS, BGD,  
995 QJS

996 Investigation: LPD, XX, AK, LM, CJB, RM, HS, CG, RS, KL, RAR, ES

997 Funding acquisition: BD, QJS

998 Project administration: BD, QJS



- 999 Supervision: ES, JP, AJB, JN, BGD, QJS
- 1000 Writing – original draft: LPD
- 1001 Writing – review & editing: LPD, BGD, QJS
- 1002

1003 **Main Figure Legends**

1004 **Figure 1: Immunogenicity of semi-synthetic, non-native GM3g-based LOGs in**  
1005 **mice.**

1006 **(a)** Native mammalian ganglioside and TACA, GM3. **(b)** 'Tag-and-modify' approaches  
1007 to chemically coupling GM3g to lysine sidechains of diverse protein carriers. **(c)**  
1008 Immunisation schedule. **(d)** Terminal LOG-specific IgG endpoint titers of immunized  
1009 mice. **(e,f)** Position of lysine residues on HEL (PDB: 193L) and their select substitution  
1010 to arginine ('-') in mutant set. **(g)** Evaluation of autologous LOG-specific IgG post-  
1011 immunisation. **(h)** gp120-[–amidine-Le<sup>x</sup>] LOG design. **(i)** Terminal endpoint titers. Data  
1012 were compared via Dunn's test. HEL mutant Ig titres were clustered into low (0–1),  
1013 medium (2–4) and high (5–6) glycan occupancy and pairwise compared.

1014

1015 **Figure 2: Evaluation of the B cell clonality raised against the semi-synthetic**  
1016 **amidine-GM3g LOG.**

1017 **(a)** Immunisation schedule. **(b, c)** Antigen-specific sorting strategy on pre-gated IgD<sup>-</sup>  
1018 B cells. **(d)** B cell clonality as inferred from the VH sequences. **(e)** Location the B cells  
1019 were recovered. **(f–h)** VH gene segment utilisation. **(i)** Alignment of enriched *IGHV2*  
1020 gene segments. **(j,k)** Distribution of common V-genes from B cells that bound either  
1021 the GM3g- or HEL-specific probes. **(l)** Odds ratios were calculated for the proportion  
1022 of share V<sub>H</sub>-genes as a proxy for germline restrictiveness. Statistical analysis was  
1023 conducted parametrically on the log of the odds ratio. **(m)** Supernatant from  
1024 representative *IGHV2*-origin B cells were screened against gp120-[–amidine-GM3g]<sub>16</sub>  
1025 via ELISA. **(c–h)** Representative data from Mouse 1.

1026

1027 **Figure 3: Protein backbone determines germinal center experience and clonal**  
1028 **maturation of GM3g LOG-specific B cell.**

1029 WT 6-week-old BALB/c mice were primed with 10  $\mu$ g gp120-[ $\alpha$ -amidine-GM3g]<sub>16</sub> +  
1030 20  $\mu$ g MPLA and culled 4-weeks post-prime. **(a)** Antigen probe sorting strategy on pre-  
1031 gated IgD<sup>-</sup> B cells and **(b)** probe binding specificities of sorted events. **(c)** Terminal  
1032 serum antibody reactivity was determined via ELISA. **(d,e)** IGHV gene utilisation. **(f)** B  
1033 cell clonality was inferred according to the iGL VH sequences. **(g)** Percentage of clonal  
1034 family members of isolated B cells with respect to protein backbone. **(h)** Chao1  
1035 estimates for mice immunized with either HEL- or gp120-based GM3g LOGs. **(i)**  
1036 Location of isolated antigen-specific B cells. **(j)** VH mismatches compared to the iGL  
1037 with respect to the compartment the B cells were isolated from. **(k)** Distributions of  
1038 SHM rates with respect to protein backbone. **(l–n)** Evaluation of the Tfh population in  
1039 draining iLNs 7 days post-prime. **(o)** Example clonal family phylogeny. **(p)** Relative  
1040 EC50 values of recombinant GM3g-specific clonal family titrated against HEL-[ $\alpha$ -  
1041 amidine-GM3g]<sub>6</sub>. **(b,f,i,j)** Representative data from mouse 1. Data were compared via  
1042 non-parametric Kruskal-Wallis and post-hoc Dunn's test.

1043

1044 **Figure 4: Glycan specificity, linker engagement and context dependency of**  
1045 **amidine-GM3g LOG-raised antibodies.**

1046 **(a)** Antiserum from gp120-[ $\alpha$ -amidine-GM3g]<sub>16</sub>-immunized mice was screened for  
1047 reactivity against a mammalian-derived glycan library. **(b–d)** A competition ELISA was  
1048 conducted to evaluate the polyclonal antibody dependency on component and  
1049 adjacent ligand segments, as well as the antibody tolerance of alternative linker  
1050 formats. Polyclonal reactivity against **(e)** a negative ceramide control and **(f)** native  
1051 presentations of SiaLac on GM3 was evaluated via direct ELISA using serum raised

1052 in gp120-[–amidine-GM3g]<sub>16</sub> antiserum. **(g)** These analyses were additionally  
1053 screened using a set of purified GM3g-specific recombinant monoclonal antibodies  
1054 later characterised. Data were compared using a post-hoc Dunn’s test.

1055

1056 **Figure 5: Structural, biophysical and biochemical characterization of BAR-1**  
1057 **reveals germline-encoded lectin-like GM3g-engaging motifs.**

1058 **(a)** The binding kinetics of the BAR-1 Fab was evaluated via SPR using an amidine-  
1059 GM3g-coated chip. **(b)** ITC was conducted with a soluble reductionist LOG mimic, Lys-  
1060 amidine-GM3g. **(c)** Competition ELISA using BAR-1 soluble ligands. **(d)**  $K_D$  and  $k_{off}$  for  
1061 the amidine/Bar-1 complex were determined using uSTA NMR measurements  
1062 repeated over titrated proton/ligand concentrations and analysing data from the NAc  
1063 proton signal. **(e)** NMR uSTA analysis of the BAR-1:Lys–amidine-GM3g binding in  
1064 solution. **(f)** Truncation of Lys–amidine-GM3g to GM3g and its tip disaccharide  
1065 (Neu5Ac-Gal; Siagal) notably leads to a readjustment of GM3g focused even more  
1066 upon the ‘foothold’ interaction of Gal. **(g)** Example raw uSTA data illustrating residue  
1067 level precision that reveals the subtle readjustment of Lys–amidine-GM3g to GM3g  
1068 effect, here analysed via the H-5”, H-8” and H-9” positions. **(h)** Side view of the 1.9-Å  
1069 x-ray structure of the BAR-1 Fab bound to the Lys-amidine-GM3g. Both heavy and  
1070 light chain CDRs are marked. Binding side of GM3g-amidine-Lys. **(i)** Residues within  
1071 4.0 Å of the polysaccharide are displayed and hydrogen bonds are shown as black  
1072 broken lines. Water is marked in red. Lys–amidine-GM3g is shown as sticks. **(k)** Logo  
1073 plots of the CDRH residues from isolated GM3g-specific *IGHV2* subgroup-encoded B  
1074 cells.

1075

1076 **References**

- 1077 1. J. S. New, R. G. King, J. F. Kearney, Manipulation of the glycan-specific  
1078 natural antibody repertoire for immunotherapy. *Immunol Rev.* **270**, 32–50  
1079 (2016).
- 1080 2. J. Lübbers, E. Rodríguez, Y. van Kooyk, Modulation of Immune Tolerance via  
1081 Siglec-Sialic Acid Interactions. *Front Immunol.* **9**, 2807 (2018).
- 1082 3. J. Müller, L. Nitschke, The role of CD22 and Siglec-G in B-cell tolerance and  
1083 autoimmune disease. *Nat Rev Rheumatol.* **10**, 422–428 (2014).
- 1084 4. F. Y. Avci, D. L. Kasper, How bacterial carbohydrates influence the adaptive  
1085 immune system. *Annu Rev Immunol.* **28**, 107–130 (2010).
- 1086 5. F. Y. Avci, X. Li, M. Tsuji, D. L. Kasper, Carbohydrates and T cells: a sweet  
1087 twosome. *Semin Immunol.* **25**, 146–151 (2013).
- 1088 6. O. T. Avery, W. F. Goebel, CHEMO-IMMUNOLOGICAL STUDIES ON  
1089 CONJUGATED CARBOHYDRATE-PROTEINS : II. IMMUNOLOGICAL  
1090 SPECIFICITY OF SYNTHETIC SUGAR-PROTEIN ANTIGENS. *J Exp Med.*  
1091 **50**, 533–550 (1929).
- 1092 7. R. Rappuoli, E. De Gregorio, P. Costantino, On the mechanisms of conjugate  
1093 vaccines. *Proc Natl Acad Sci U S A.* **116**, 14–16 (2019).
- 1094 8. L. P. Deimel, X. Xue, Q. J. Sattentau, Glycans in HIV-1 vaccine design -  
1095 engaging the shield. *Trends Microbiol* (2022), doi:10.1016/j.tim.2022.02.004.
- 1096 9. W. B. Struwe, E. Chertova, J. D. Allen, G. E. Seabright, Y. Watanabe, D. J.  
1097 Harvey, M. Medina-Ramirez, J. D. Roser, R. Smith, D. Westcott, B. F. Keele,  
1098 J. W. J. Bess, R. W. Sanders, J. D. Lifson, J. P. Moore, M. Crispin, Site-  
1099 Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble  
1100 Trimeric Immunogen. *Cell Rep.* **24**, 1958–1966 (2018).

- 1101 10. X. Wei, J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-  
1102 Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A.  
1103 Nowak, B. H. Hahn, P. D. Kwong, G. M. Shaw, Antibody neutralization and  
1104 escape by HIV-1. *Nature*. **422**, 307–312 (2003).
- 1105 11. C. O. Barnes, H. B. Gristick, N. T. Freund, A. Escolano, A. Y. Lyubimov, H.  
1106 Hartweiger, A. P. West, A. E. Cohen, M. C. Nussenzweig, P. J. Bjorkman,  
1107 Structural characterization of a highly-potent V3-glycan broadly neutralizing  
1108 antibody bound to natively-glycosylated HIV-1 envelope. *Nat Commun*. **9**,  
1109 1251 (2018).
- 1110 12. T. Schoofs, C. O. Barnes, N. Suh-Toma, J. Golijanin, P. Schommers, H.  
1111 Gruell, A. P. J. West, F. Bach, Y. E. Lee, L. Nogueira, I. S. Georgiev, R. T.  
1112 Bailer, J. Czartoski, J. R. Mascola, M. S. Seaman, M. J. McElrath, N. A. Doria-  
1113 Rose, F. Klein, M. C. Nussenzweig, P. J. Bjorkman, Broad and Potent  
1114 Neutralizing Antibodies Recognize the Silent Face of the HIV Envelope.  
1115 *Immunity*. **50**, 1513–1529 (2019).
- 1116 13. T. Zhou, A. Zheng, U. Baxa, G.-Y. Chuang, I. S. Georgiev, R. Kong, S. O’Dell,  
1117 S. Shahzad-UI-Hussan, C.-H. Shen, Y. Tsybovsky, R. T. Bailer, S. K. Gift, M.  
1118 K. Louder, K. McKee, R. Rawi, C. H. Stevenson, G. B. E. Stewart-Jones, J. D.  
1119 Taft, E. Waltari, Y. Yang, B. Zhang, S. S. Shivatare, V. S. Shivatare, C.-C. D.  
1120 Lee, C.-Y. Wu, J. C. Mullikin, C. A. Bewley, D. R. Burton, V. R. Polonis, L.  
1121 Shapiro, C.-H. Wong, J. R. Mascola, P. D. Kwong, X. Wu, A Neutralizing  
1122 Antibody Recognizing Primarily N-Linked Glycan Targets the Silent Face of  
1123 the HIV Envelope. *Immunity*. **48**, 500-513.e6 (2018).
- 1124 14. K. J. Doores, Z. Fulton, M. Huber, I. A. Wilson, D. R. Burton, *J Virol*, in press,  
1125 doi:10.1128/JVI.01110-10.

- 1126 15. J. M. Steichen, D. W. Kulp, T. Tokatlian, A. Escolano, P. Dosenovic, R. L.  
1127 Stanfield, L. E. McCoy, G. Ozorowski, X. Hu, O. Kalyuzhniy, B. Briney, T.  
1128 Schiffner, F. Garces, N. T. Freund, A. D. Gitlin, S. Menis, E. Georgeson, M.  
1129 Kubitz, Y. Adachi, M. Jones, A. A. Mutafyan, D. S. Yun, C. T. Mayer, A. B.  
1130 Ward, D. R. Burton, I. A. Wilson, D. J. Irvine, M. C. Nussenzweig, W. R.  
1131 Schief, HIV Vaccine Design to Target Germline Precursors of Glycan-  
1132 Dependent Broadly Neutralizing Antibodies. *Immunity*. **45**, 483–496 (2016).
- 1133 16. D. J. Leggat, K. W. Cohen, J. R. Willis, W. J. Fulp, A. C. deCamp, O.  
1134 Kalyuzhniy, C. A. Cottrell, S. Menis, G. Finak, L. Ballweber-Fleming, A.  
1135 Srikanth, J. R. Plyler, T. Schiffner, A. Liguori, F. Rahaman, A. Lombardo, V.  
1136 Philiponis, R. E. Whaley, A. Seese, J. Brand, A. M. Ruppel, W. Hoyland, N. L.  
1137 Yates, L. D. Williams, K. Greene, H. Gao, C. R. Mahoney, M. M. Corcoran, A.  
1138 Cagigi, A. Taylor, D. M. Brown, D. R. Ambrozak, T. Sincomb, X. Hu, R. Tingle,  
1139 E. Georgeson, S. Eskandarzadeh, N. Alavi, D. Lu, T.-M. Mullen, M. Kubitz, B.  
1140 Groschel, J. Maenza, O. Kolokythas, N. Khati, J. Bethony, S. Crotty, M.  
1141 Roederer, G. B. Karlsson Hedestam, G. D. Tomaras, D. Montefiori, D.  
1142 Diemert, R. A. Koup, D. S. Laufer, M. J. McElrath, A. B. McDermott, W. R.  
1143 Schief, Vaccination induces HIV broadly neutralizing antibody precursors in  
1144 humans. *Science (1979)*. **378**, eadd6502 (2023).
- 1145 17. O. Haji-Ghassemi, R. J. Blackler, N. Martin Young, S. V Evans, Antibody  
1146 recognition of carbohydrate epitopes†. *Glycobiology*. **25**, 920–952 (2015).
- 1147 18. K. O. Lloyd, C. M. Gordon, I. J. Thampoe, C. DiBenedetto, Cell Surface  
1148 Accessibility of Individual Gangliosides in Malignant Melanoma Cells to  
1149 Antibodies Is Influenced by the Total Ganglioside Composition of the Cells<sup>1</sup>.  
1150 *Cancer Res.* **52**, 4948–4953 (1992).

- 1151 19. S. Cavdarli, S. Groux-Degroote, P. Delannoy, Gangliosides: The Double-Edge  
1152 Sword of Neuro-Ectodermal Derived Tumors. *Biomolecules*. **9** (2019),  
1153 doi:10.3390/biom9080311.
- 1154 20. T. Shimizu, M. Nagane, M. Suzuki, A. Yamauchi, K. Kato, N. Kawashima, Y.  
1155 Nemoto, T. Maruo, Y. Kawakami, T. Yamashita, Tumor hypoxia regulates  
1156 ganglioside GM3 synthase, which contributes to oxidative stress resistance in  
1157 malignant melanoma. *Biochim Biophys Acta Gen Subj*. **1864**, 129723 (2020).
- 1158 21. P. O. Livingston, G. Ritter, M. J. Calves, Antibody response after immunization  
1159 with the gangliosides GM1, GM2, GM3, GD2 and GD3 in the mouse. *Cancer*  
1160 *Immunol Immunother*. **29**, 179–184 (1989).
- 1161 22. Y. Pan, P. Chefalo, N. Nagy, C. Harding, Z. Guo, Synthesis and immunological  
1162 properties of N-modified GM3 antigens as therapeutic cancer vaccines. *J Med*  
1163 *Chem*. **48**, 875–883 (2005).
- 1164 23. R. J. Bitton, M. D. Guthmann, M. R. Gabri, A. J. L. Carnero, D. F. Alonso, L.  
1165 Fainboim, D. E. Gomez, Cancer vaccines: an update with special focus on  
1166 ganglioside antigens. *Oncol Rep*. **9**, 267–276 (2002).
- 1167 24. M. D. Guthmann, R. J. Bitton, A. J. L. Carnero, M. R. Gabri, G. Cinat, L.  
1168 Koliren, D. Lewi, L. E. Fernandez, D. F. Alonso, D. E. Gómez, L. Fainboim,  
1169 Active specific immunotherapy of melanoma with a GM3 ganglioside-based  
1170 vaccine: a report on safety and immunogenicity. *J Immunother*. **27**, 442–451  
1171 (2004).
- 1172 25. F. Estevez, A. Carr, L. Solorzano, O. Valiente, C. Mesa, O. Barroso, G. V  
1173 Sierra, L. E. Fernandez, Enhancement of the immune response to poorly  
1174 immunogenic gangliosides after incorporation into very small size  
1175 proteoliposomes (VSSP). *Vaccine*. **18**, 190–197 (1999).



- 1176 26. P. O. Livingston, G. Y. Wong, S. Adluri, Y. Tao, M. Padavan, R. Parente, C.  
1177 Hanlon, M. J. Calves, F. Helling, G. Ritter, Improved survival in stage III  
1178 melanoma patients with GM2 antibodies: a randomized trial of adjuvant  
1179 vaccination with GM2 ganglioside. *Journal of Clinical Oncology*. **12**, 1036–  
1180 1044 (1994).
- 1181 27. S. Yu, Q. Wang, J. Zhang, Q. Wu, Z. Guo, Synthesis and Evaluation of Protein  
1182 Conjugates of GM3 Derivatives Carrying Modified Sialic Acids as Highly  
1183 Immunogenic Cancer Vaccine Candidates. *Medchemcomm*. **2**, 524–530  
1184 (2011).
- 1185 28. V. Mata-Haro, C. Cekic, M. Martin, P. M. Chilton, C. R. Casella, T. C. Mitchell,  
1186 The vaccine adjuvant monophosphoryl lipid A as a TRIF-biased agonist of  
1187 TLR4. *Science (1979)*. **316**, 1628–1632 (2007).
- 1188 29. L. C. James, D. S. Tawfik, The specificity of cross-reactivity: promiscuous  
1189 antibody binding involves specific hydrogen bonds rather than nonspecific  
1190 hydrophobic stickiness. *Protein Sci*. **12**, 2183–2193 (2003).
- 1191 30. E. A. Padlan, Anatomy of the antibody molecule. *Mol Immunol*. **31**, 169–217  
1192 (1994).
- 1193 31. Y. C. Lee, C. P. Stowell, M. J. Krantz, 2-Imino-2-methoxyethyl 1-  
1194 thioglycosides: new reagents for attaching sugars to proteins. *Biochemistry*.  
1195 **15**, 3956–3963 (1976).
- 1196 32. J. M. Chalker, G. J. L. Bernardes, B. G. Davis, A “tag-and-modify” approach to  
1197 site-selective protein modification. *Acc Chem Res*. **44**, 730–741 (2011).
- 1198 33. I. Green, W. E. Paul, B. Benacerraf, Hapten carrier relationships in the DNP-  
1199 PLL foreign albumin complex system: induction of tolerance and stimulation of  
1200 cells in vitro. *J Exp Med*. **127**, 43–53 (1968).

- 1201 34. N. C. Peters, D. H. Hamilton, P. A. Bretscher, Analysis of cytokine-producing  
1202 Th cells from hen egg lysozyme-immunized mice reveals large numbers  
1203 specific for “cryptic” peptides and different repertoires among different Th  
1204 populations. *Eur J Immunol.* **35**, 56–65 (2005).
- 1205 35. A. Chao, Nonparametric Estimation of the Number of Classes in a Population.  
1206 *Scandinavian Journal of Statistics.* **11**, 265–270 (1984).
- 1207 36. L. Mesin, A. Schiepers, J. Ersching, A. Barbulescu, C. B. Cavazzoni, A.  
1208 Angelini, T. Okada, T. Kurosaki, G. D. Victora, Restricted Clonality and Limited  
1209 Germinal Center Reentry Characterize Memory B Cell Reactivation by  
1210 Boosting. *Cell.* **180**, 92-106.e11 (2020).
- 1211 37. O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C.  
1212 Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel,  
1213 I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong, J.  
1214 C. Paulson, Printed covalent glycan array for ligand profiling of diverse glycan  
1215 binding proteins. *Proceedings of the National Academy of Sciences.* **101**,  
1216 17033–17038 (2004).
- 1217 38. W. Peng, R. P. de Vries, O. C. Grant, A. J. Thompson, R. McBride, B.  
1218 Tsogtbaatar, P. S. Lee, N. Razi, I. A. Wilson, R. J. Woods, J. C. Paulson,  
1219 Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched  
1220 Human-type Receptors, Conferring Potential for Increased Avidity. *Cell Host*  
1221 *Microbe.* **21**, 23–34 (2017).
- 1222 39. C. J. Buchanan, B. Gaunt, P. J. Harrison, Y. Yang, J. Liu, A. Khan, A. M.  
1223 Giltrap, A. Le Bas, P. N. Ward, K. Gupta, M. Dumoux, T. K. Tan, L. Schimaski,  
1224 S. Daga, N. Picchiotti, M. Baldassarri, E. Benetti, C. Fallerini, F. Fava, A.  
1225 Giliberti, P. I. Koukos, M. J. Davy, A. Lakshminarayanan, X. Xue, G.

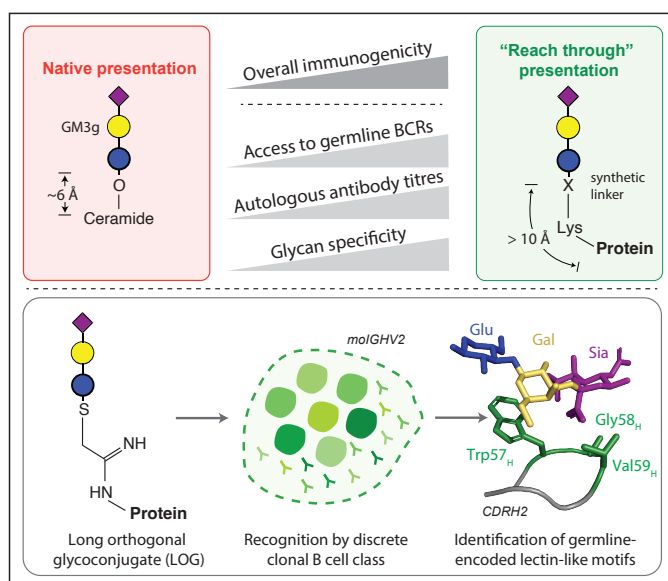
- 1226 Papadakis, L. P. Deimel, V. Casablanco-Antràs, T. D. W. Claridge, A. M. J. J.  
1227 Bonvin, Q. J. Sattentau, S. Furini, M. Gori, J. Huo, R. J. Owens, C. Schaffitzel,  
1228 I. Berger, A. Renieri, null null, J. H. Naismith, A. J. Baldwin, B. G. Davis,  
1229 Pathogen-sugar interactions revealed by universal saturation transfer analysis.  
1230 *Science* (1979). **377**, eabm3125 (2022).
- 1231 40. B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B.  
1232 Henrissat, The Carbohydrate-Active EnZymes database (CAZy): an expert  
1233 resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
- 1234 41. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat,  
1235 The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids*  
1236 *Res.* **42**, D490–D495 (2014).
- 1237 42. N. Dotan, R. T. Altstock, M. Schwarz, A. Dukler, Anti-glycan antibodies as  
1238 biomarkers for diagnosis and prognosis. *Lupus.* **15**, 442–450 (2006).
- 1239 43. J. S. Temme, D. L. Butler, J. C. Gildersleeve, Anti-glycan antibodies: roles in  
1240 human disease. *Biochemical Journal.* **478**, 1485–1509 (2021).
- 1241 44. K. Ding, T. Ekberg, J. Zeuthen, S. Teneberg, K.-A. Karlsson, A. Rosén,  
1242 Monoclonal antibody against a lactose epitope of glycosphingolipids binds to  
1243 melanoma tumour cells. *Glycoconj J.* **10**, 395–405 (1993).
- 1244 45. C. Soliman, J. X. Chua, M. Vankemmelbeke, R. S. McIntosh, A. J. Guy, I.  
1245 Spendlove, L. G. Durrant, P. A. Ramsland, The terminal sialic acid of stage-  
1246 specific embryonic antigen-4 has a crucial role in binding to a cancer-targeting  
1247 antibody. *Journal of Biological Chemistry.* **295**, 1009–1020 (2020).
- 1248 46. J. L. Asensio, A. Ardá, F. J. Cañada, J. Jiménez-Barbero, Carbohydrate–  
1249 Aromatic Interactions. *Acc Chem Res.* **46**, 946–954 (2013).

- 1250 47. A. Escolano, H. B. Gristick, M. E. Abernathy, J. Merckenschlager, R. Gautam,  
1251 T. Y. Oliveira, J. Pai, A. P. J. West, C. O. Barnes, A. A. Cohen, H. Wang, J.  
1252 Golijanin, D. Yost, J. R. Keeffe, Z. Wang, P. Zhao, K.-H. Yao, J. Bauer, L.  
1253 Nogueira, H. Gao, A. v Voll, D. C. Montefiori, M. S. Seaman, A. Gazumyan, M.  
1254 Silva, A. T. McGuire, L. Stamatatos, D. J. Irvine, L. Wells, M. A. Martin, P. J.  
1255 Bjorkman, M. C. Nussenzweig, Immunization expands B cells specific to HIV-1  
1256 V3 glycan in mice and macaques. *Nature*. **570**, 468–473 (2019).
- 1257 48. J. Topin, M. Lelimosin, J. Arnaud, A. Audfray, S. Pérez, A. Varrot, A. Imberty,  
1258 The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is  
1259 a Determinant for Recognition by Pathogen Lectins. *ACS Chem Biol*. **11**,  
1260 2011–2020 (2016).
- 1261 49. A. Gabba, A. Bogucka, J. G. Luz, A. Diniz, H. Coelho, F. Corzana, F. J.  
1262 Cañada, F. Marcelo, P. V Murphy, G. Birrane, Crystal Structure of the  
1263 Carbohydrate Recognition Domain of the Human Macrophage Galactose C-  
1264 Type Lectin Bound to GalNAc and the Tumor-Associated Tn Antigen.  
1265 *Biochemistry*. **60**, 1327–1336 (2021).
- 1266 50. F. Tobola, M. Lelimosin, A. Varrot, E. Gillon, B. Darnhofer, O. Blixt, R. Birner-  
1267 Gruenberger, A. Imberty, B. Wiltschi, Effect of Noncanonical Amino Acids on  
1268 Protein–Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate  
1269 Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS*  
1270 *Chem Biol*. **13**, 2211–2219 (2018).
- 1271 51. P. O. Livingston, G. Ragupathi, Cancer vaccines targeting carbohydrate  
1272 antigens. *Hum Vaccin*. **2**, 137–143 (2006).

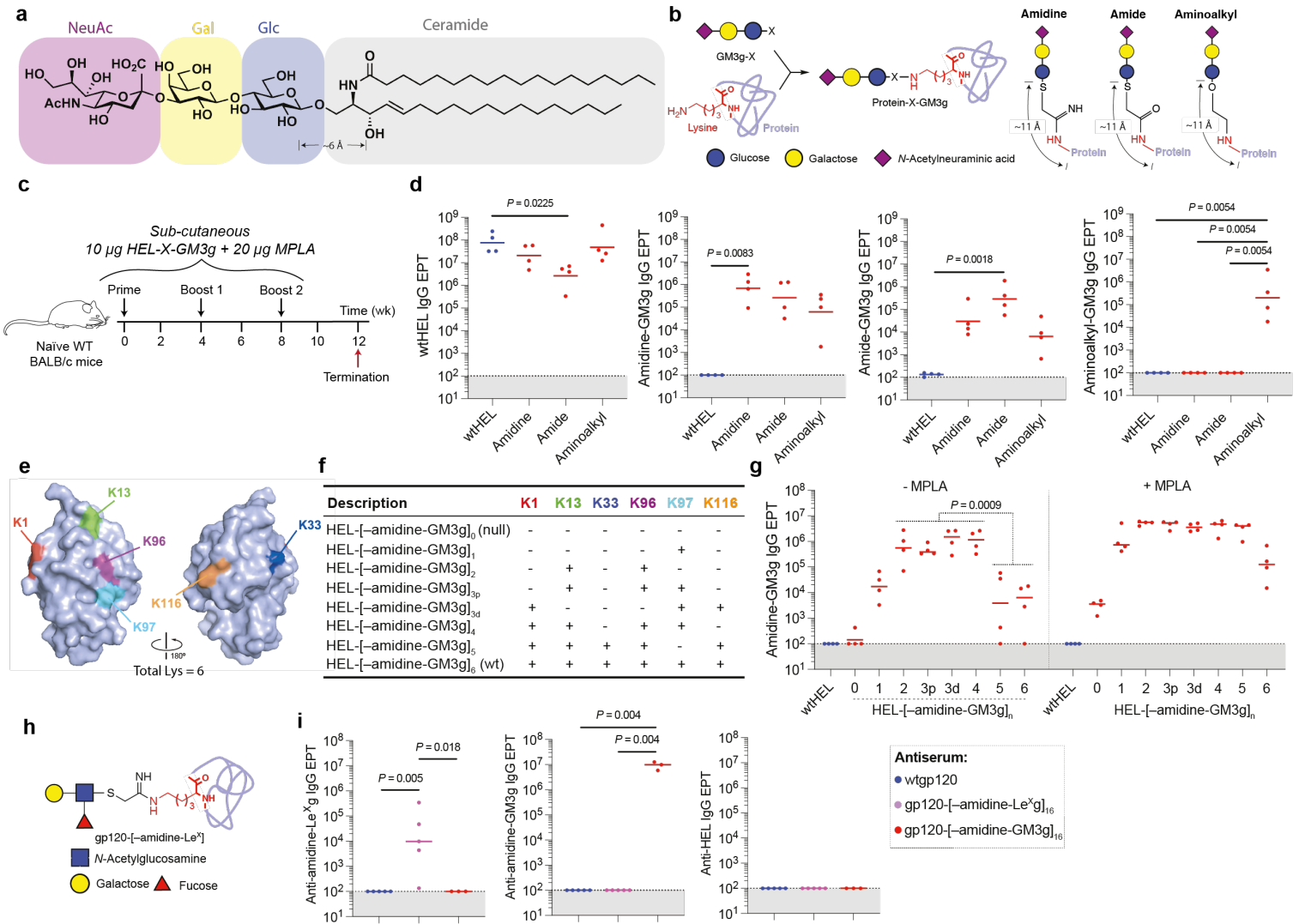
- 1273 52. G. T. Hermanson, "Chapter 3 - The Reactions of Bioconjugation" in  
1274 *Bioconjugate Techniques (Third Edition)*, G. T. Hermanson, Ed. (Academic  
1275 Press, Boston, Third Edit., 2013), pp. 229–258.
- 1276 53. I. Y. Ho, J. J. Bunker, S. A. Erickson, K. E. Neu, M. Huang, M. Cortese, B.  
1277 Pulendran, P. C. Wilson, Refined protocol for generating monoclonal  
1278 antibodies from single human and murine B cells. *J Immunol Methods*. **438**,  
1279 67–70 (2016).
- 1280 54. C. Viant, A. Escolano, S. T. Chen, M. C. Nussenzweig, Sequencing, cloning,  
1281 and antigen binding analysis of monoclonal antibodies isolated from single  
1282 mouse B cells. *STAR Protoc*. **2**, 100389 (2021).
- 1283 55. W. S. DeWitt III, L. Mesin, G. D. Victora, V. N. Minin, F. A. Matsen IV, Using  
1284 Genotype Abundance to Improve Phylogenetic Inference. *Mol Biol Evol*. **35**,  
1285 1253–1265 (2018).
- 1286 56. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr D Biol*  
1287 *Crystallogr*. **62**, 72–82 (2006).
- 1288 57. W. Kabsch, XDS. *Acta Crystallogr D Biol Crystallogr*. **66**, 125–132 (2010).
- 1289 58. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R.  
1290 Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J.  
1291 McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R.  
1292 J. Read, A. Vagin, K. S. Wilson, Overview of the CCP 4 suite and current  
1293 developments. *Acta Crystallogr D Biol Crystallogr*. **67**, 235–242 (2011).
- 1294 59. Z. Zhang, N. K. Sauter, H. van den Bedem, G. Snell, A. M. Deacon,  
1295 Automated diffraction image analysis and spot searching for high-throughput  
1296 crystal screening. *J Appl Crystallogr*. **39**, 112–119 (2006).

- 1297 60. G. Winter, xia2: an expert system for macromolecular crystallography data  
1298 reduction. *J Appl Crystallogr.* **43**, 186–190 (2010).
- 1299 61. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll,  
1300 B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B.  
1301 K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D.  
1302 Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev,  
1303 L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure  
1304 determination using X-rays, neutrons and electrons: recent developments in  
1305 Phenix. *Acta Crystallogr D Struct Biol.* **75**, 861–877 (2019).
- 1306 62. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta*  
1307 *Crystallogr D Biol Crystallogr.* **60**, 2126–2132 (2004).
- 1308 63. A. W. Schüttelkopf, D. M. F. van Aalten, PRODRG: a tool for high-throughput  
1309 crystallography of protein–ligand complexes. *Acta Crystallogr D Biol*  
1310 *Crystallogr.* **60**, 1355–1363 (2004).
- 1311 64. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J.  
1312 Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity : all-atom  
1313 structure validation for macromolecular crystallography. *Acta Crystallogr D Biol*  
1314 *Crystallogr.* **66**, 12–21 (2010).
- 1315 65. T. C. Hsieh, K. H. Ma, A. Chao, iNEXT: an R package for rarefaction and  
1316 extrapolation of species diversity (Hill numbers). *Methods Ecol Evol.* **7**, 1451–  
1317 1456 (2016).
- 1318
- 1319

## Graphical Abstract —

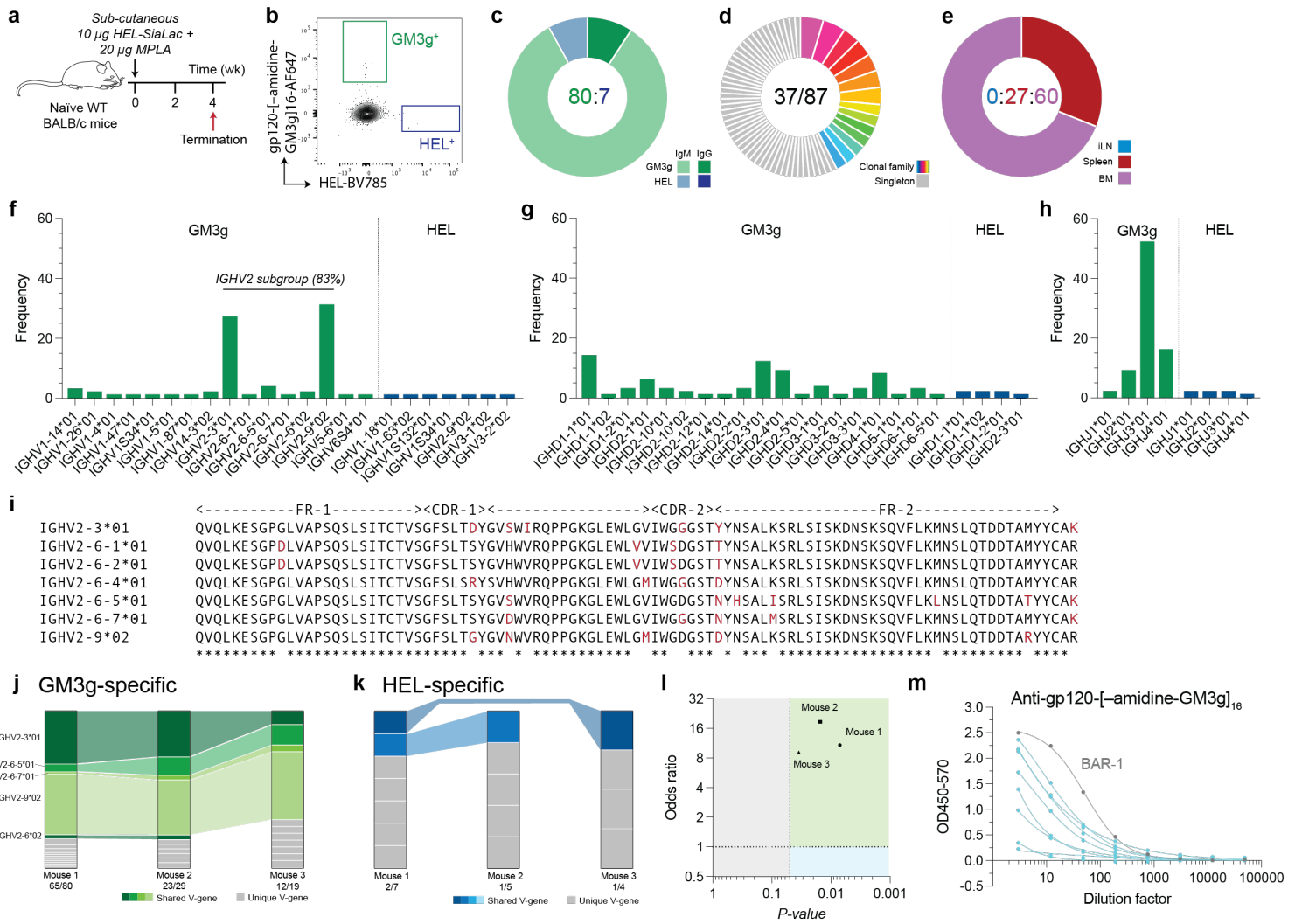


**Figure 1: Immunogenicity of semi-synthetic, non-native GM3g-based LOGs in mice.**

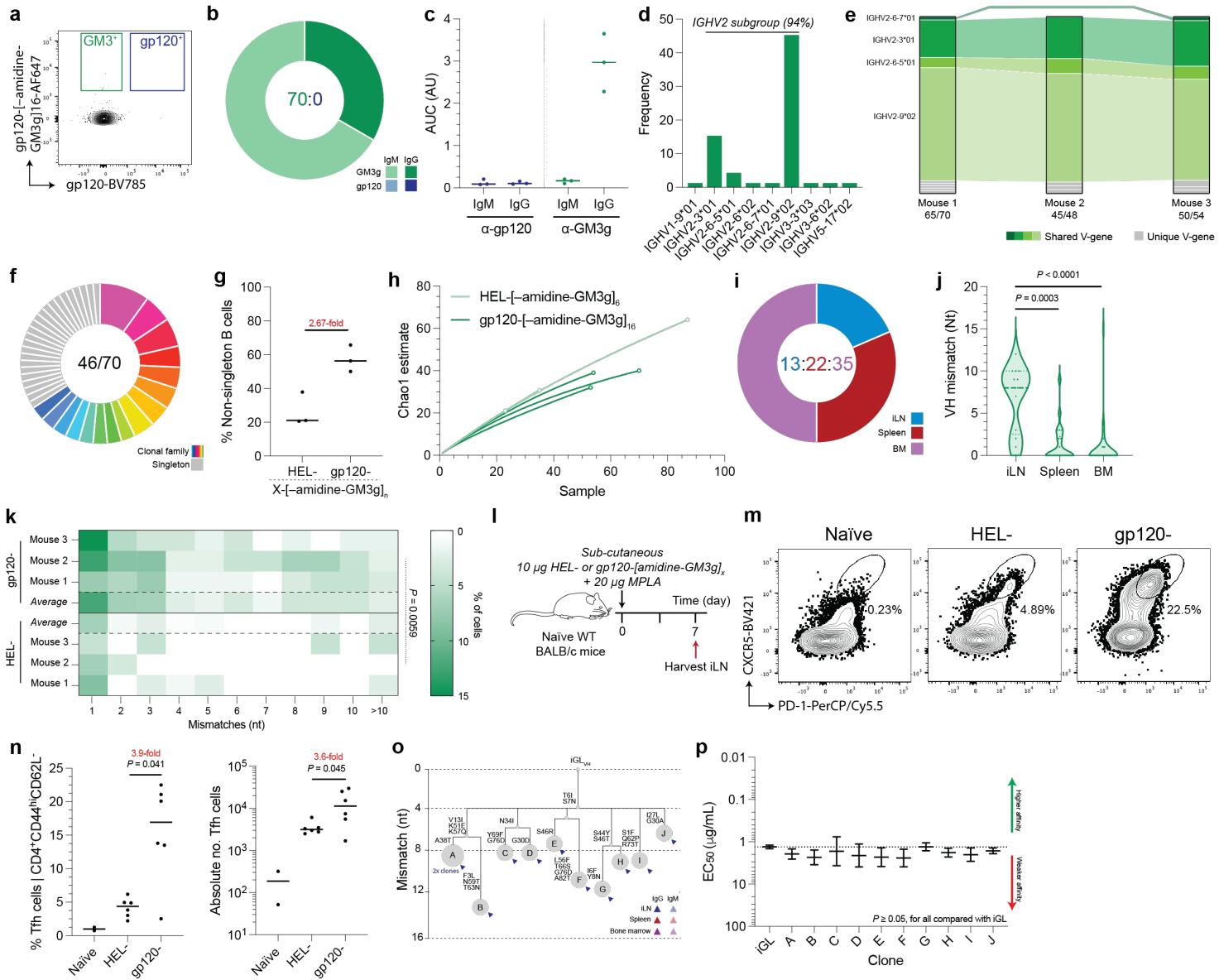




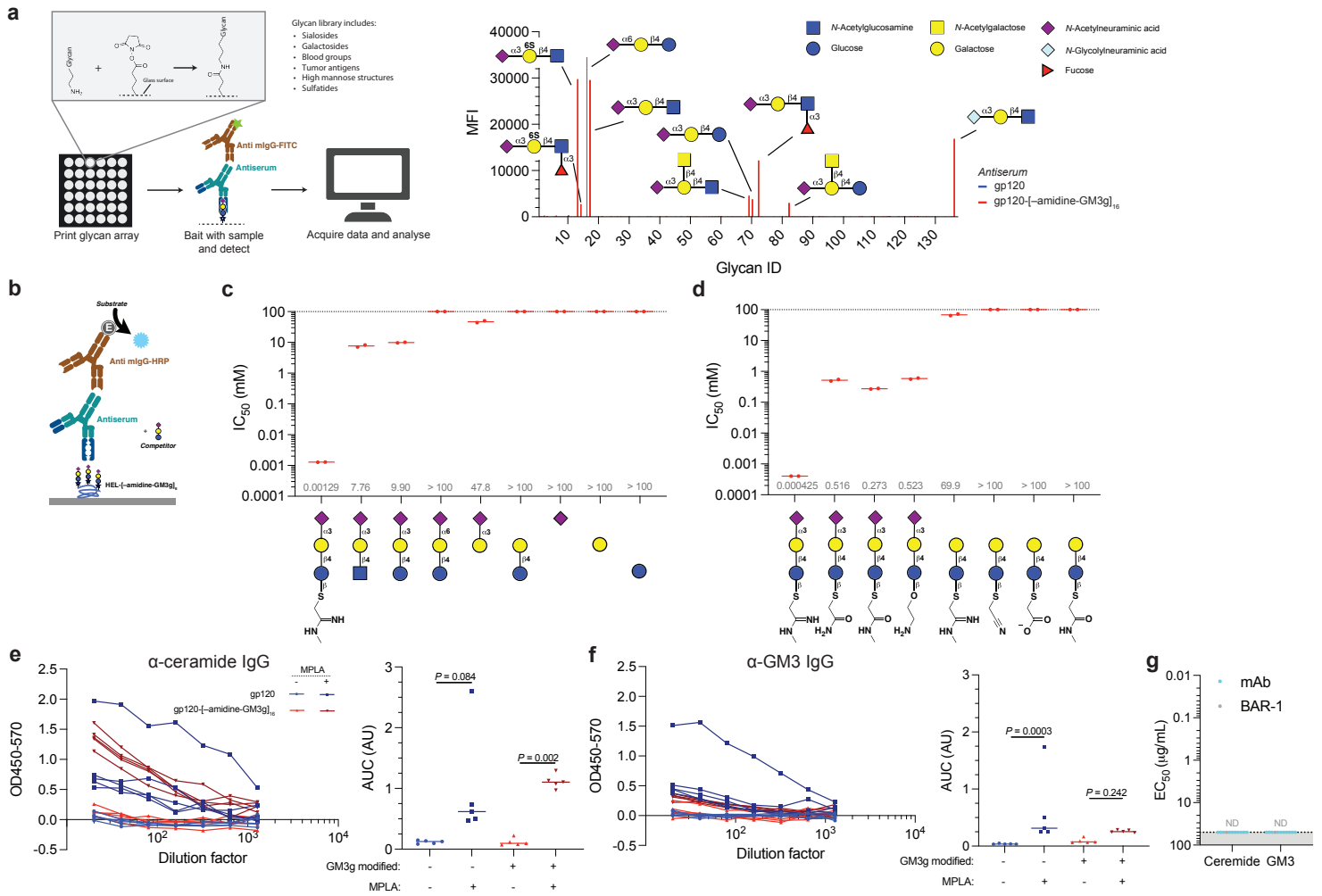
**Figure 2: Evaluation of B cell clonality raised against the novel semi-synthetic amidine-GM3g LOG.**



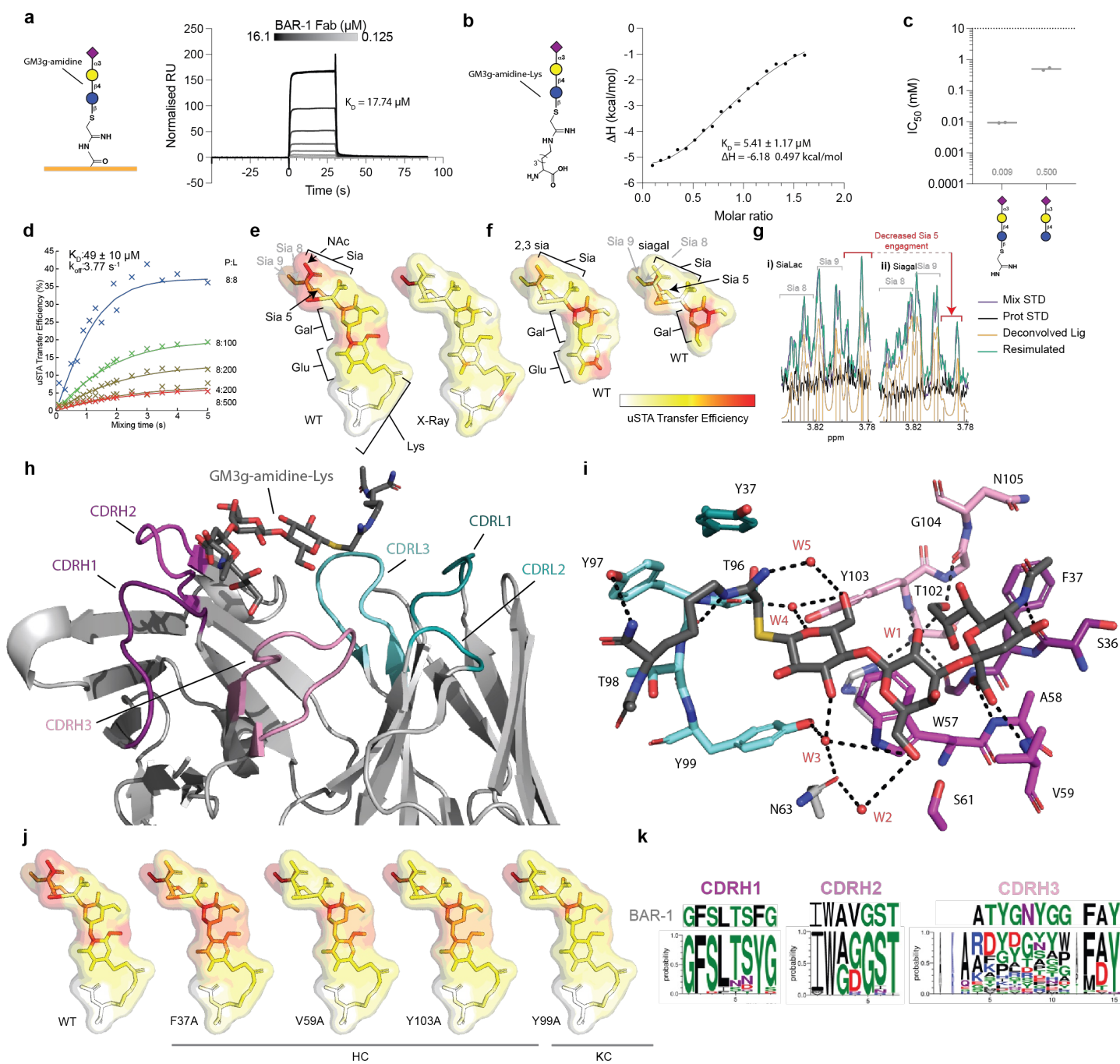
**Figure 3: Protein backbone determines germinal centre experience and clonal maturation of GM3g LOG-specific B cells.**

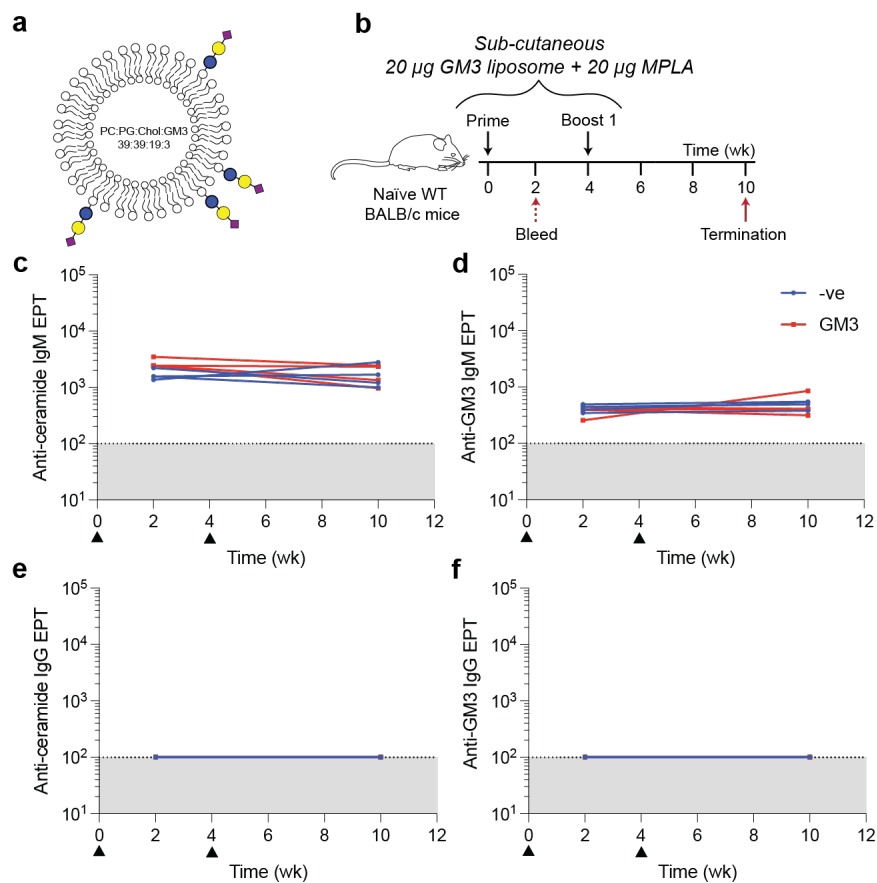


**Figure 4: Glycan specificity, linker engagement and context dependency of amidine-GM3g LOG-raised antibodies.**



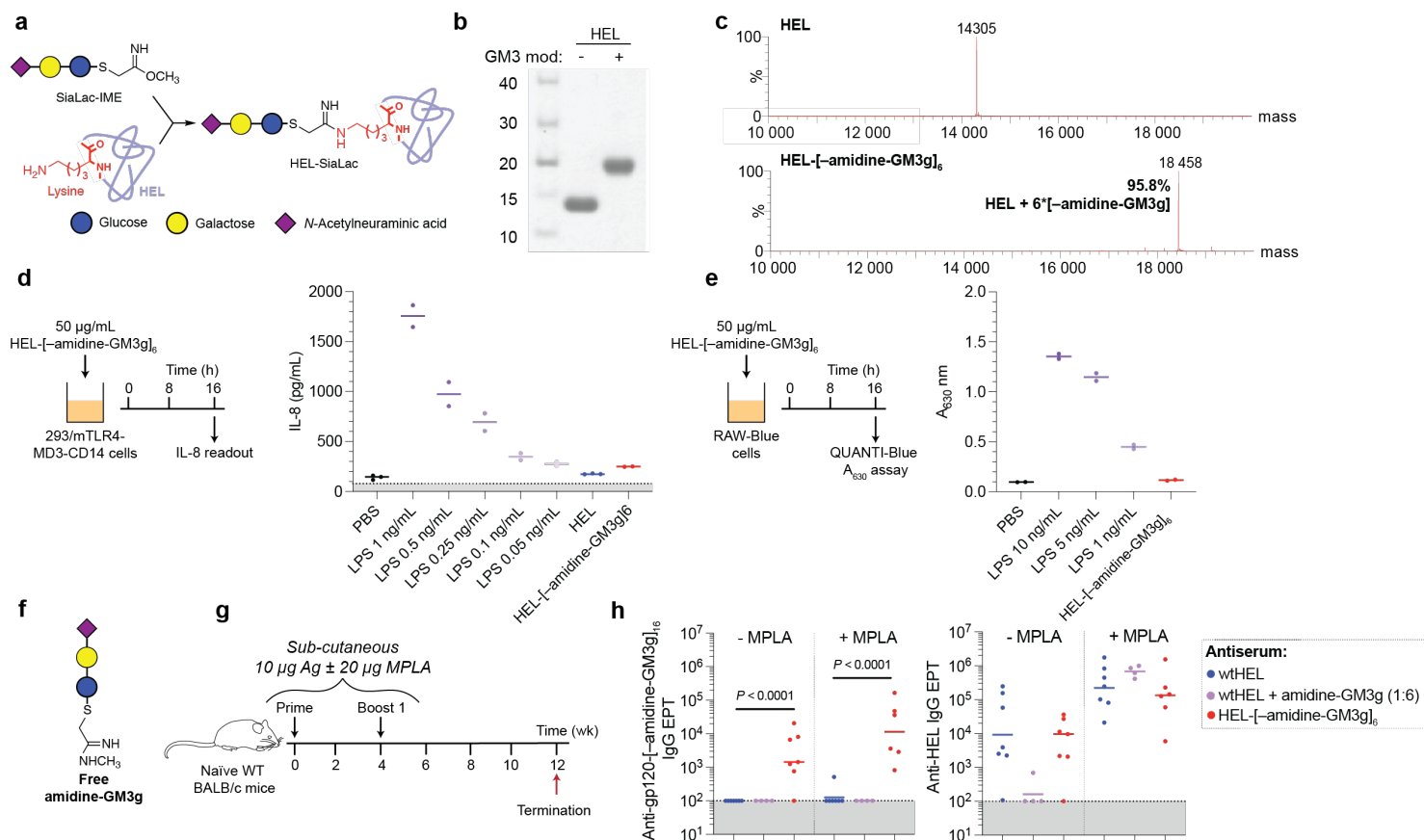
**Figure 5: Structural and biochemical characterisation of prototypical mAb, BAR-1, reveals critical glycan engaging motifs.**





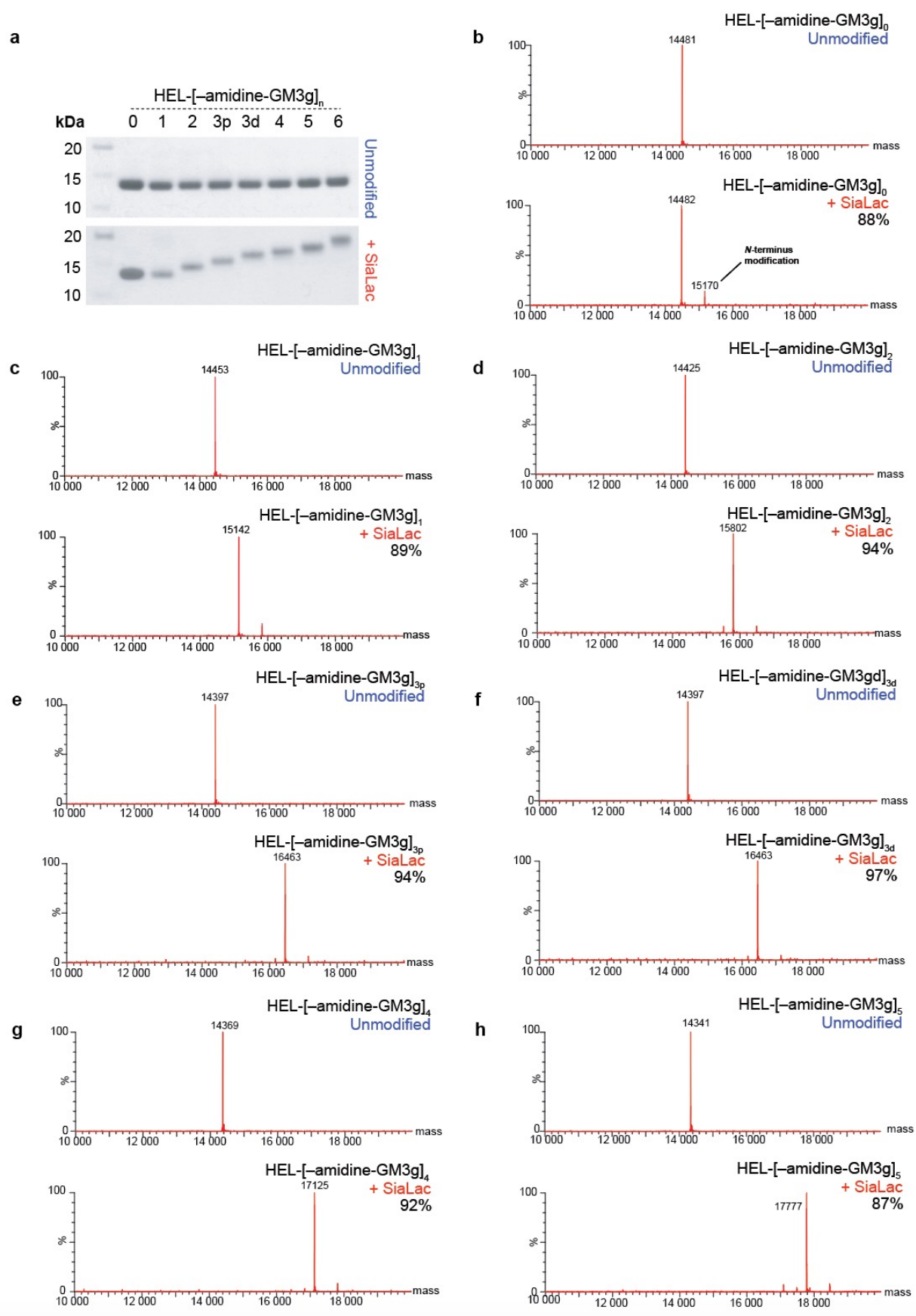
**Figure S1: Immunorecessiveness of GM3 liposomes in mice.**

(a) Liposomes were synthesized both with and without GM3. (b) Immunization schedule. (c-f) Serum IgM and IgG reactivity was screened via direct ELISA against both ceramide and GM3 over the immunisation period.



**Figure S2: Chemical and immunological characterisation of the amidine-GM3g LOG.**

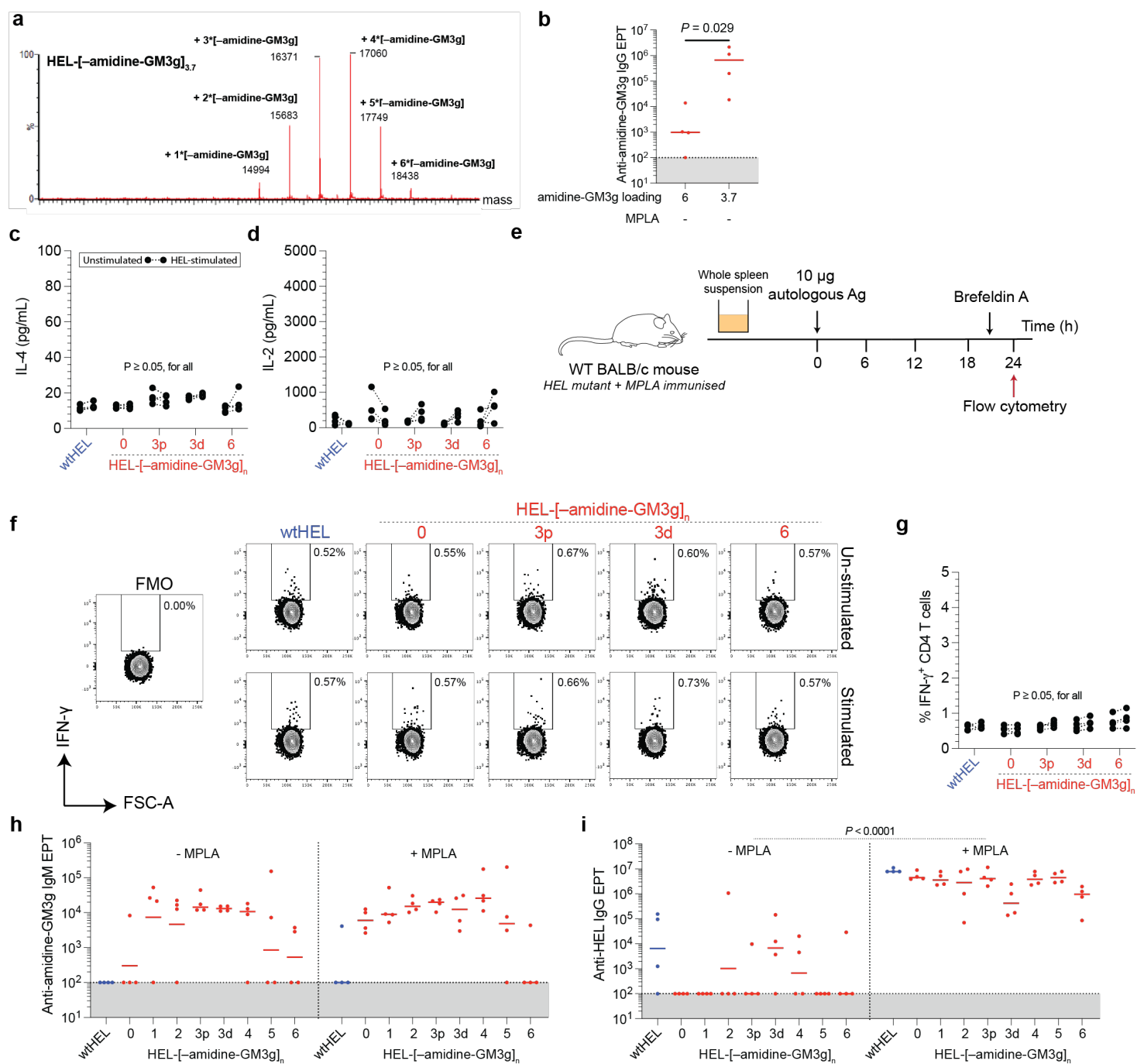
(a) Overview of HEL-[amidine-GM3g]<sub>6</sub> synthesis. (b) SDS-PAGE of HEL following [-amidine-GM3g] conjugation. (c) Mass spectra of HEL-[amidine-GM3g]<sub>6</sub> sample. (d) LPS contamination was tested via incubating 293/mTLR4-MD3-CD14 cells with HEL-[amidine-GM3g]<sub>6</sub>. IL-8 production was evaluated via ELISA. (e) Broader endotoxin contamination was screened using RAW-Blue cells. (f) Free amidine-GM3g design. (g) Immunisation schedule. (h) Terminal IgG endpoint titres. Data were compared via Tukey's post-hoc multiple comparison test.



**Figure S3: Chemical characterization of amidine-GM3g-modified HEL mutants.**

**(a)** SDS-PAGE of the purified HEL mutants and their GM3g-modified counterparts. **(b-h)** Mass spectra of the modified HEL mutant products. Refer to Figure 1f for K->R mutation code.

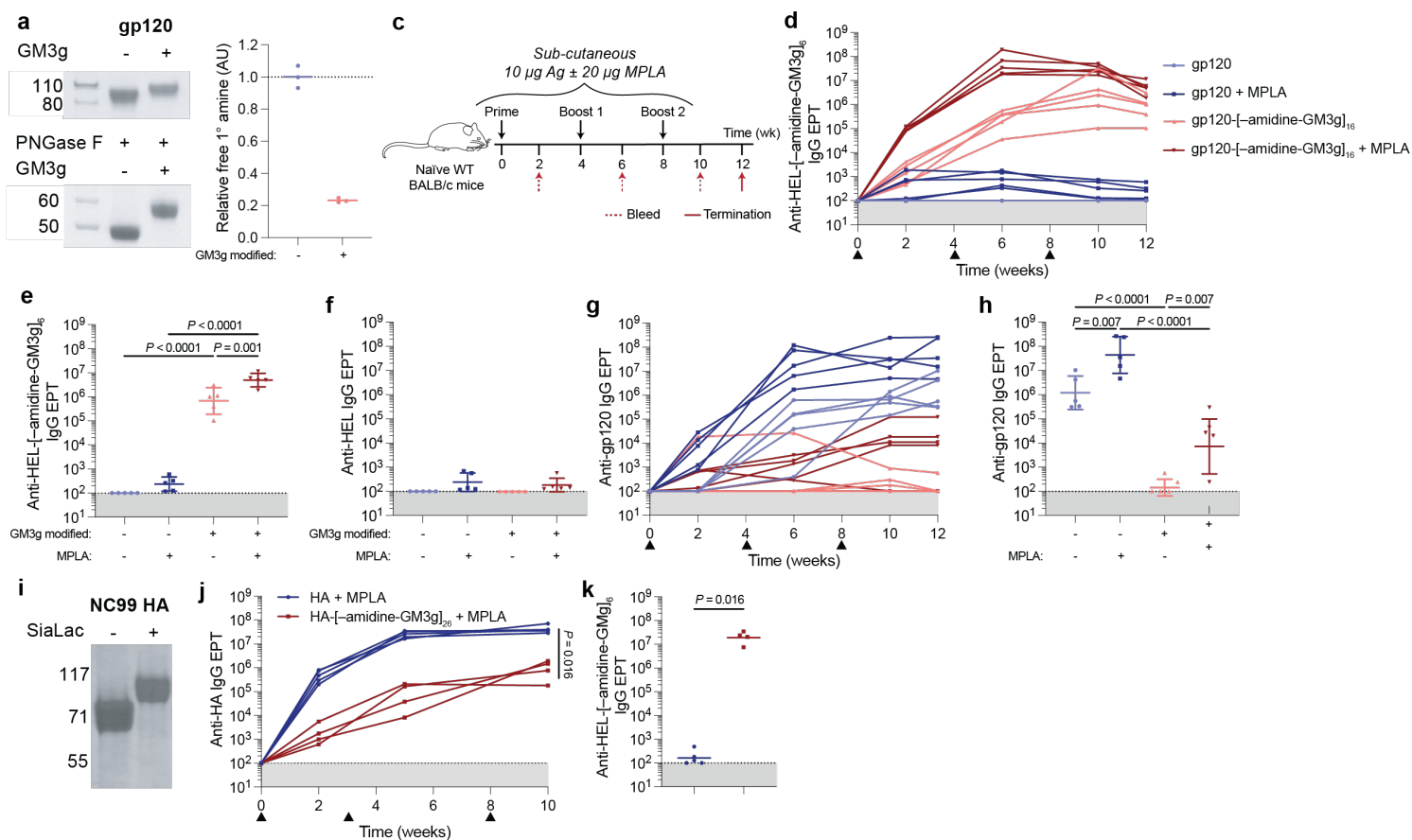




**Figure S4: Dissecting whether modifications to the HEL protein backbone implicate Th responses.**

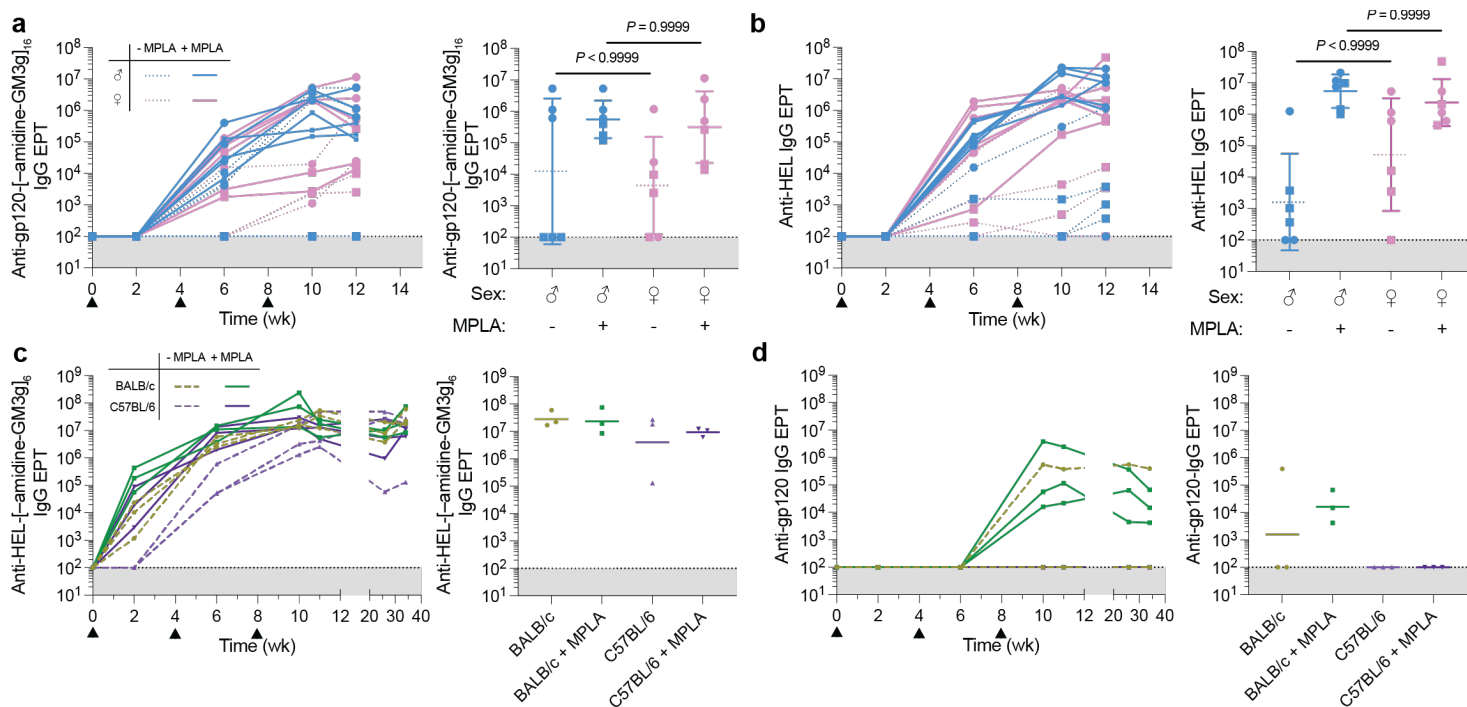
(a) Mass spectra of wtHEL partially modified with amidine-GM3g, producing HEL-[amidine-GM3g]<sub>3,7</sub>. (b) Terminal gp120-[amidine-GM3g]<sub>16</sub>-reactive IgG endpoint titers of mice primed and boosted with HEL-[amidine-GM3g]<sub>3,7</sub>. (c,d) Whole splenocytes of animals immunized with amidine-GM3g-modified HEL mutants were stimulated *in vitro* for 72 h and cytokine release in supernatant was screened. (e-g) Intracellular IFN-γ was detected via flow cytometry on pre-gated CD4<sup>+</sup> cells. (h) IgM endpoint titres were screened against gp120-[amidine-GM3g]<sub>16</sub> two-weeks post-prime. (i) Terminal IgG-specific IgG. Data were compared using Dunn's tests, except (i) where two-way ANOVA contrasted adjuvant and sugar loading effects.





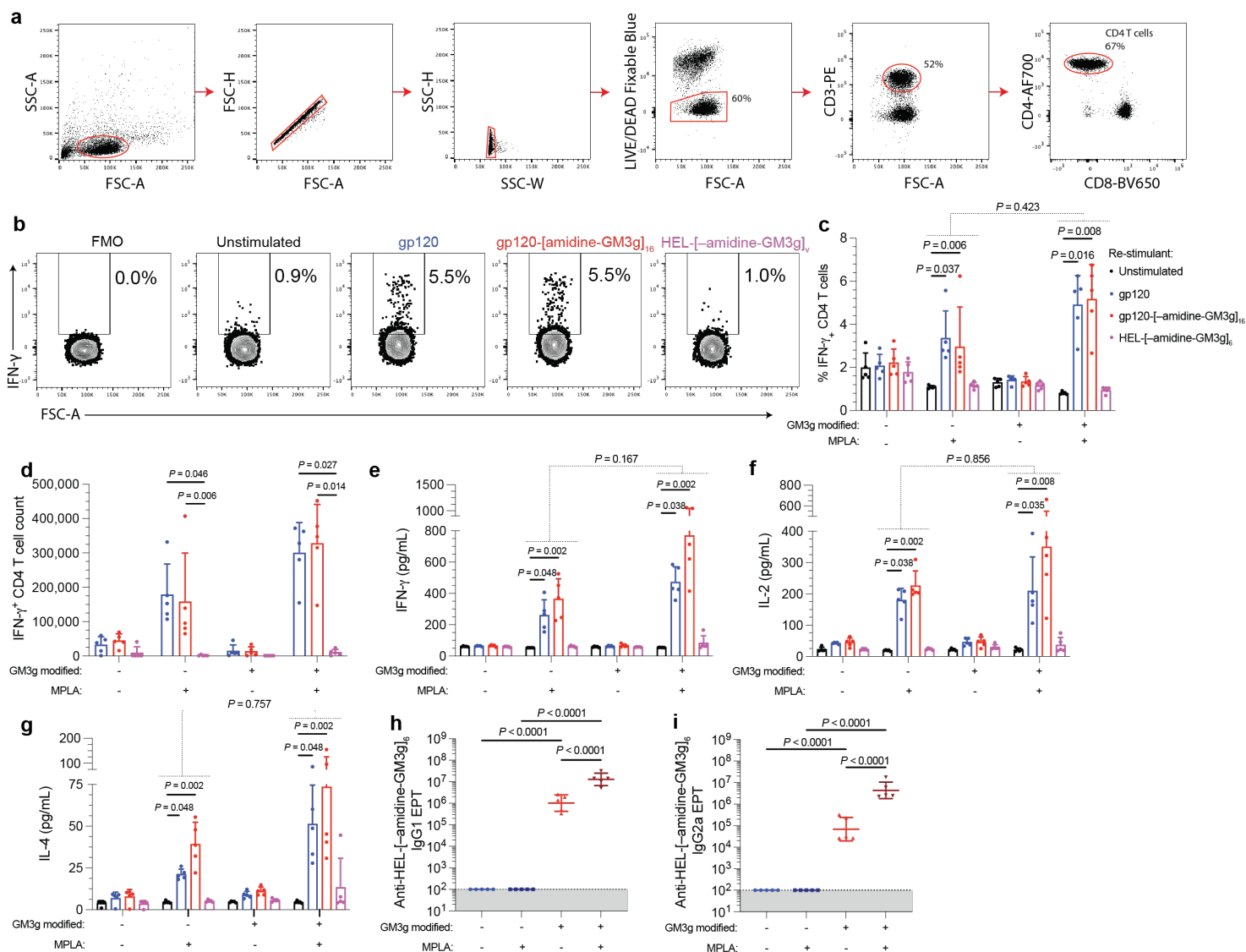
**Figure S5: LOG-specific antibody responses occur against amidine-GM3g across multiple protein carrier proteins.**

(a) SDS-PAGE of amidine-GM3g modified and unmodified proteins pre- and post-PNGase F treatment. (b) Free amine ELISA post-LOG modification. (c) Immunisation schedule. (d–h) Longitudinal or terminal serum IgG endpoint titres against LOG-specific and protein carrier constructs in animals immunised with gp120-[amidine-GM3g]<sub>16</sub>. Data were evaluated using a post-hoc Tukey's test. (i) SDS-PAGE of IAV-derived H1N1 (NC99) HA post-GM3g modification. (j,k) Longitudinal and terminal IgG endpoint titres. Data were evaluated using Mann-Whitney tests.



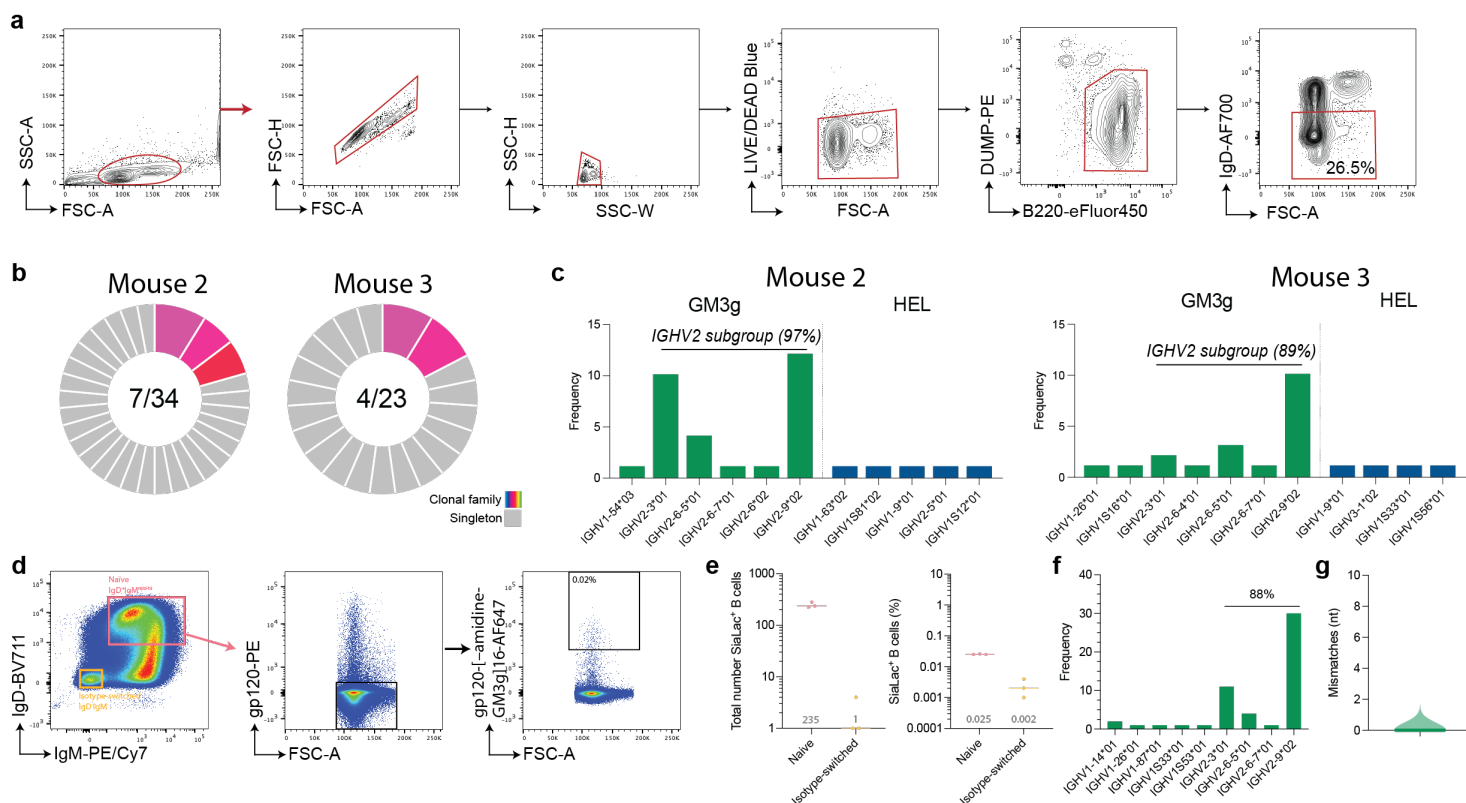
**Figure S6: Evaluating the sex and murine background effects on the anti-amidine-GM3g LOG response.**

(a,b) Male and female WT BALB/c mice were immunised three times (▲) with 10 µg HEL-[amidine-GM3g]<sub>6</sub> ± 20 µg MPLA. Both LOG and protein backbone-specific serum IgG endpoint titres were determined both longitudinally and at the terminal timepoint. Data were compared via Dunn's multiple comparison test. (c,d) BALB/c and C57BL/6 mice were immunised with 10 µg gp120-[amidine-GM3g]<sub>16</sub> ± 20 µg MPLA. Serum IgG endpoint titres against antigen components, LOG and protein backbone, were measured.



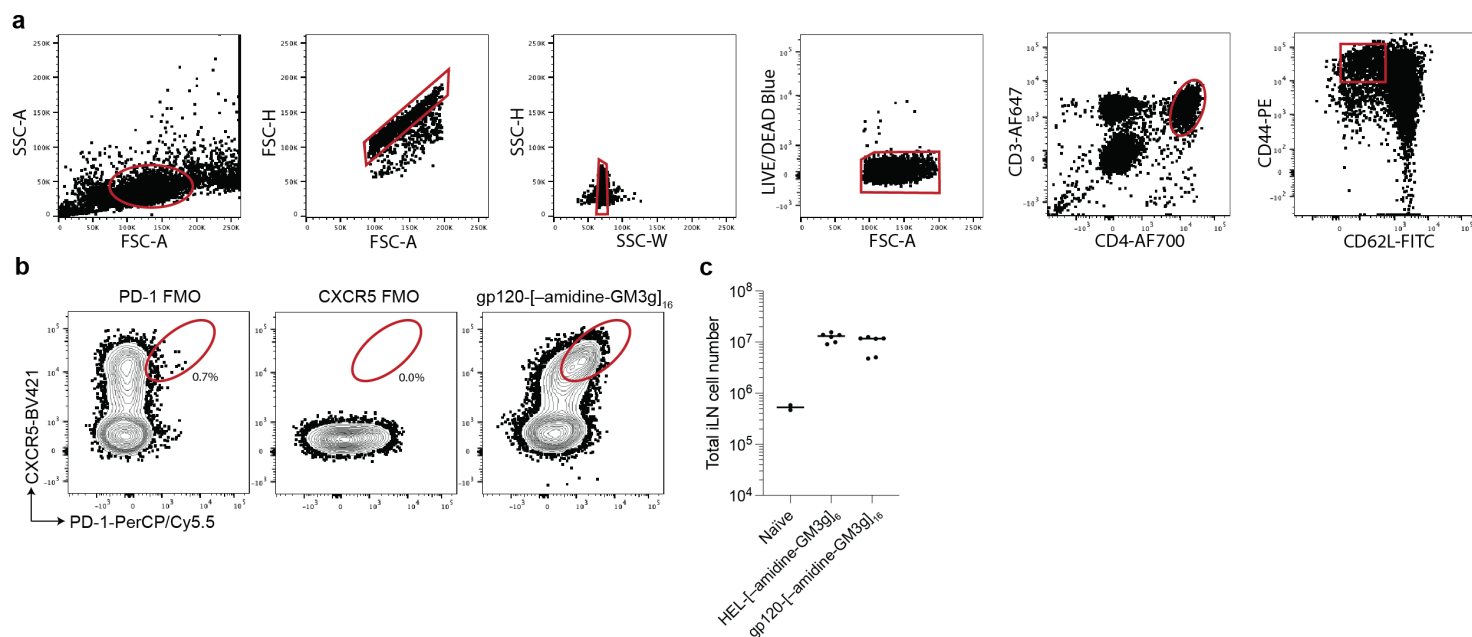
**Figure S7: Th cell recall responses in mice immunized with gp120-[amidine-GM3g]<sub>16</sub>.**

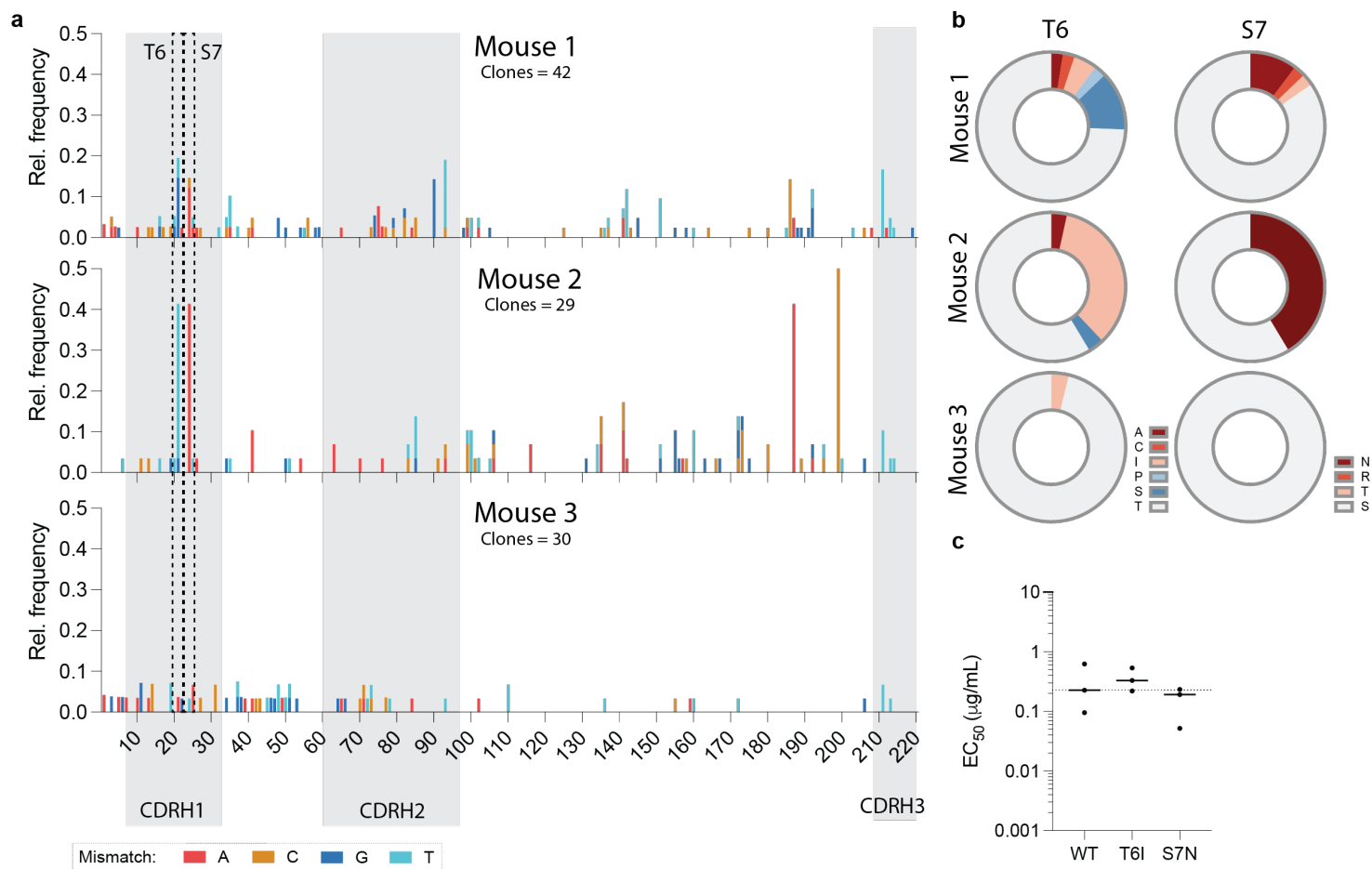
(a-d) Intracellular cytokine staining was performed on splenocytes of immunised mice, restimulated *in vitro* with different protein antigens, as indicated. IFN- $\gamma$  production among CD4<sup>+</sup> T cells was compared between vaccination and restimulatory conditions. (e-g) Cytokine release was similarly compared in splenocytes restimulated for 72 h via ELISA. (h,i) Serum LOG-specific IgG subclass endpoint titres were measured via ELISA. Data were compared pairwise via Tukey's post-hoc test. Establishment of an interaction effect between vaccination and restimulatory conditions were determined via two-way ANOVA.



**Figure S8: Clontyping of HEL-[amide-GM3g]<sub>6</sub>-immunized mice.**

(a) Gating strategy for IgD<sup>+</sup> B cells. (b) Clonal family clustering and (c) *IGHV* gene-segment utilisation in mice primed with HEL-[amide-GM3g]<sub>6</sub>. (d) Gating strategy to identify the antigen-specific naive B cell population from splenocytes. (e) The absolute number and percentage of [amide-GM3g]<sup>+</sup> B cells. (f,g) Heavy chain V-regions were recovered and sequence-validated from one mouse, confirming their clontypic origins and GC inexperience.

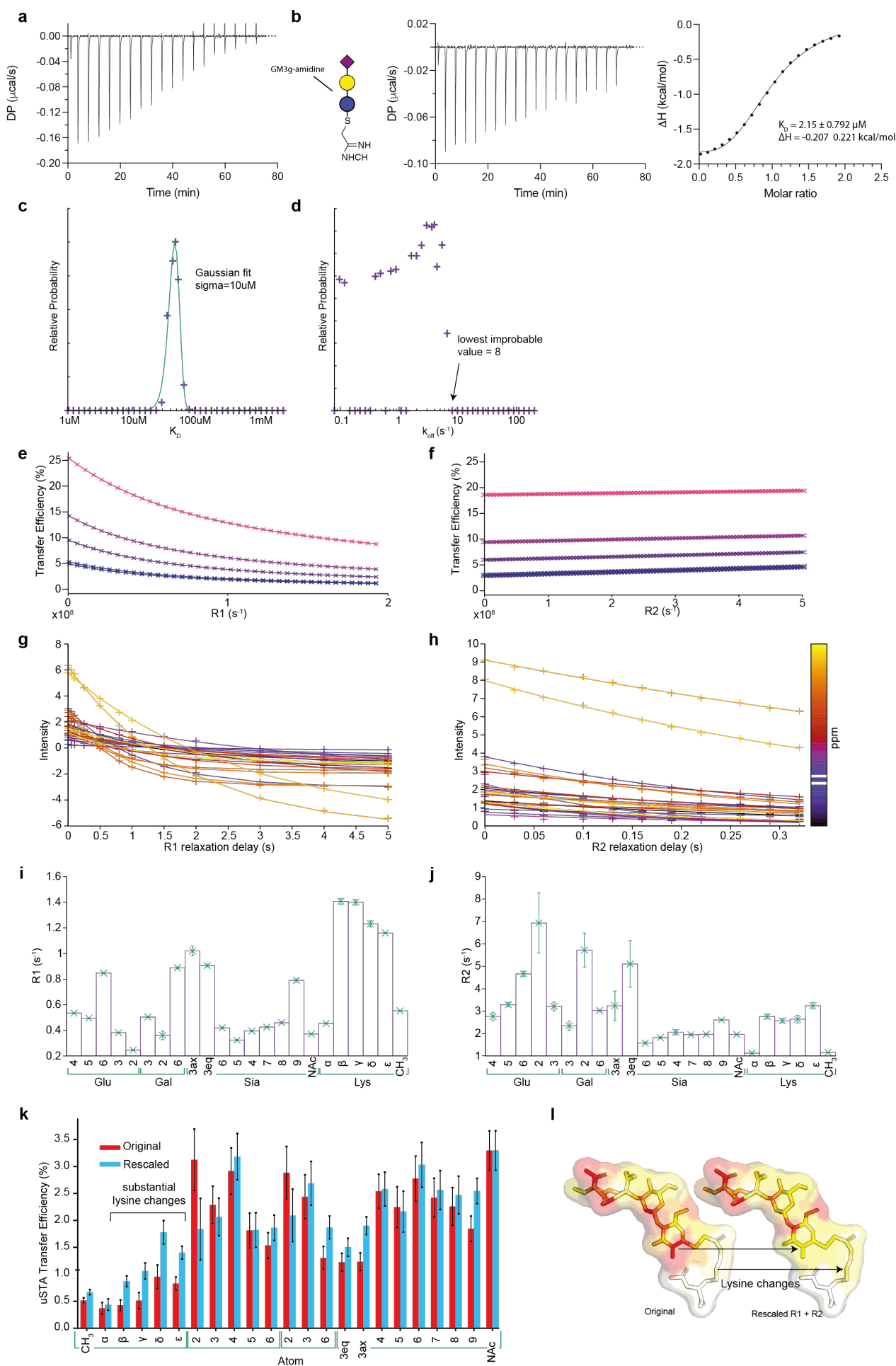




**Figure S10: Mutation frequencies observed across the gp120-[–amidine-GM3g]<sub>16</sub>-raised *IGHV2* subgroup population.**

**(a)** Manhattan plot of the nucleotide mismatches from all isolated *IGHV2*-origin GM3g-binding B cell raised against the gp120-[–amidine-GM3g]<sub>16</sub> LOG. **(b)** Substitutional implications at mutation hotspot codons, where the wild-type encodes T6 and S7. **(c)** The most common substitutions were mutated into the WT BAR-1 sequence and their relative binding against gp120-[–amidine-GM3g]<sub>16</sub> was compared via ELISA.

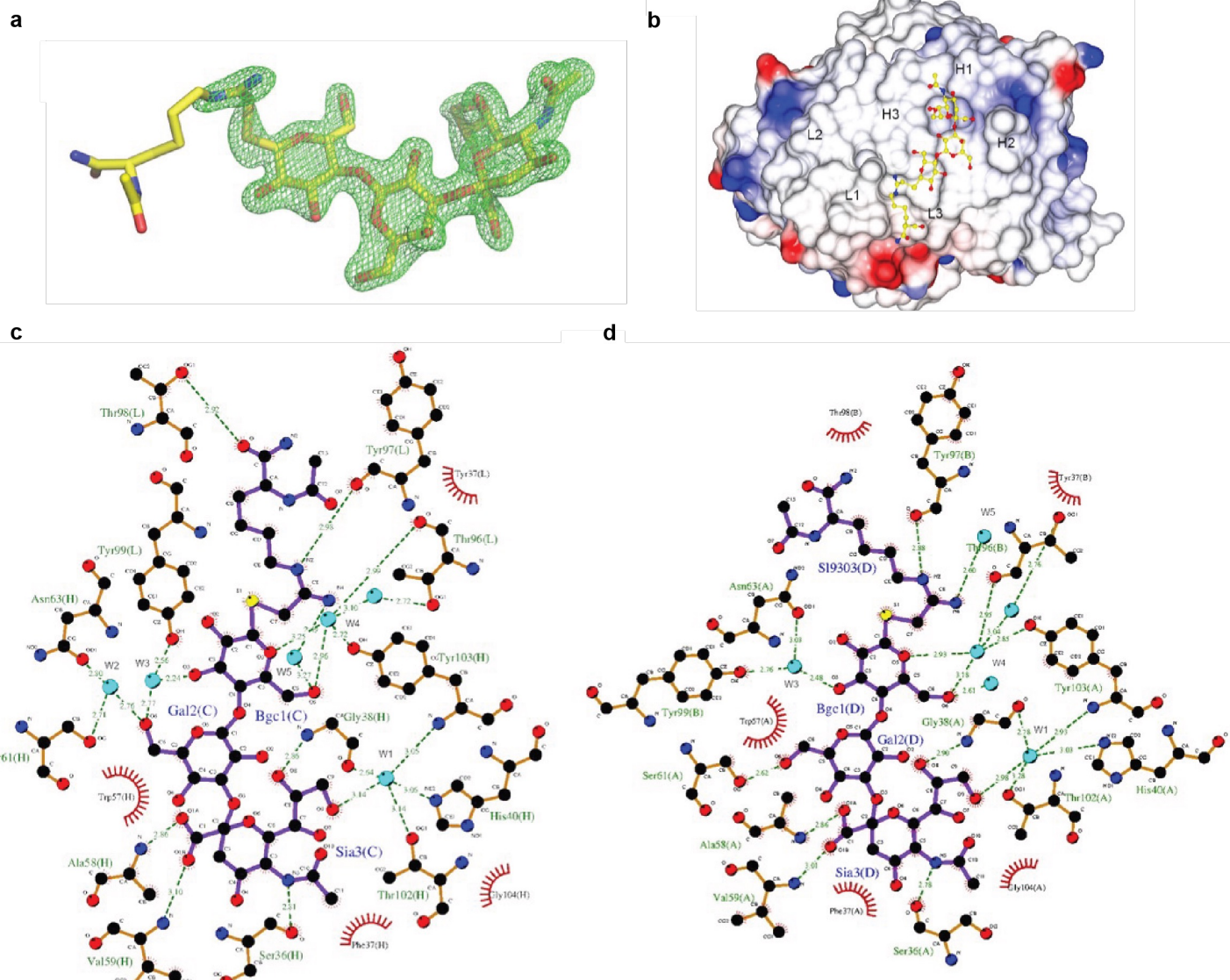
**Figure S11: Biophysical characterisation and uSTA analysis of BAR-1.**



**Figure S11: Details of Lys–C(NH)NH-GM3g•BAR1 complex by uSTA NMR.**

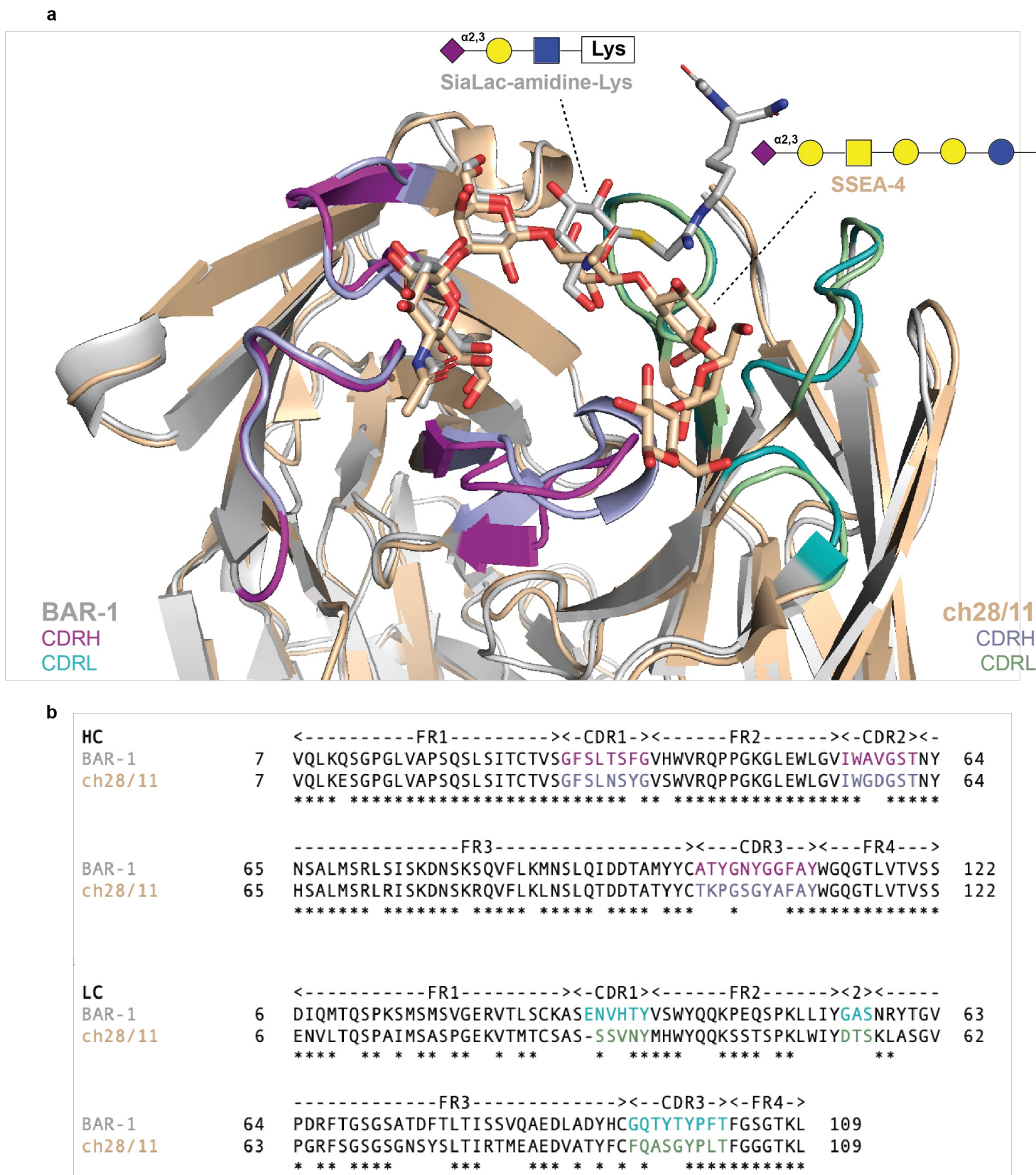
**(a)** Raw titration data of BAR-1 Fab against Lys–amidine-GM3g. **(b)** ICT performed against amidine-GM3g. **(c)** Results of the Bloch-McConnell fitting of BAR-1 with Lys–C(NH)NH-GM3g. **(d)** These reveal good quality fits of the data. Iteratively changing and fixing the  $K_D$  value, refitting the data and following the variation in the probability of the model being correct ( $\exp(-\chi^2/2)$ ) allows construction of an error surface. To an excellent approximation, the variation in the fitted  $K_D$  follows a gaussian distribution **(e)**. Performing the same analysis on the  $k_{off}$  parameter resulted in a non-central distribution, indicating that in this case, while  $K_D$  is well determined,  $k_{off}$  is not. The distribution is reasonably interpreted by a log-normal distribution, resulting in the most probable value being  $3.77 \text{ s}^{-1}$  but with asymmetric error bars,  $+4 \text{ s}^{-1}$ ,  $-2 \text{ s}^{-1}$ . The distribution can be interpreted as placing a limit on  $k_{off}$ , such that  $k_{off} < 8 \text{ s}^{-1}$ . **(f,g)**  $R_1$  and  $R_2$  relaxation rates were obtained for each proton in amidine-lysine. The variation in relaxation rates approximately by a factor of 3, prompted us to consider the effects of this on the transfer efficiency. Notably, the  $R_1$  determined from the  $K_D$  analysis for the NAc proton ( $0.37 \text{ s}^{-1}$ ) was consistent with the value measured directly and independently ( $0.4 \text{ s}^{-1}$ ) supporting the quantitative uSTA analysis. **(h,i)** The simulated parameters from the  $K_D$  analysis in **c** were used to simulate the variation in transfer efficiency as a function of  $R_1$  and  $R_2$ , revealing almost no variation with  $R_2$ , but a modest variation with  $R_1$ . **(j)** These curves were interpolated using a biexponential function for  $R_1$  and a linear function for  $R_2$ , and were used to provide a rescaling factor to adjust the transfer efficiencies of each atom to the value expected if relaxation was identical to the NAc proton. The largest correction was for the lysine delta proton ( $R_1 1.4 \text{ s}^{-1}$ ) which was furthest from the NAc  $R_1$  ( $0.4 \text{ s}^{-1}$ ). In this extreme case, the correction to the transfer efficiency was a factor of 2. **(k)** The original and rescaled interaction surfaces for Lys–C(NH)NH-GM3g. The overall pattern observed is largely invariant of the rescaling, with some positions varying more than others. The main conclusions drawn from inspection of the surface, that the NAc methyl group and the sialic acid moiety dominate the interaction, that protons in all GM3g sugars are important, and that the lysine does not contribute substantially to the interaction are independent of the relaxation correction. In the manuscript, all interaction surfaces shown have had the transfer efficiencies 'corrected' using this method.





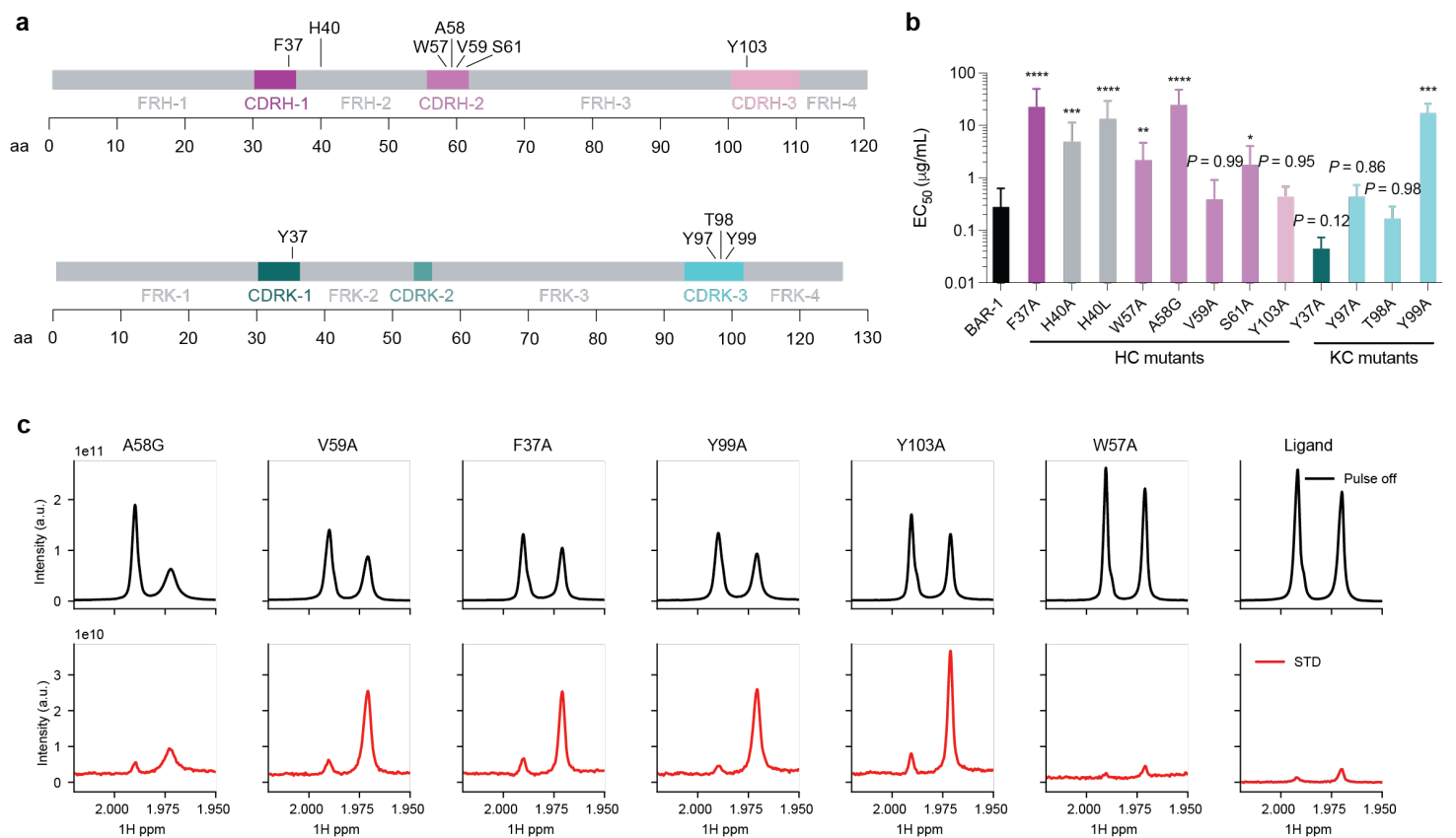
**Figure S12: Details of the X-ray structure of Lys-C(NH)NH-GM3g-BAR1.**

**(a)**  $F_0-F_c$  electron density omit map at  $3\sigma$  around SiaLac-amidine-Lys molecule. SiaLac-amidine-Lys is shown as sticks with carbon atoms coloured in yellow, nitrogen in dark blue and oxygen in red. **(b)** Surface of the binding side of BAR-1/SiaLac-amidine-Lys complex structure. The surface of Bar-1 is colored by electrostatic charges calculated in CCP4MG (red for negative potential, white for neutral and blue for positive). SiaLac-amidine-Lys is shown as sticks with the carbon in yellow. CDR loops have been labelled. **(c,d)** Ligplot diagrams illustrating BAR-1/siaLac-amidine-Lys interactions for chain H/L and A/B. Covalent bonds of the polysaccharide and the protein residues are in purple and brown sticks, respectively. Hydrogen bonds are represented by green dashed lines and hydrophobic contacts are shown as red semi-circles with radiating spokes



**Figure S13: Comparison between BAR-1 and ch28/11.**

(a) Structural alignment of the BAR-1 x-ray structure bound to Lys–C(NH)NH–GM3g and that of ch28/11 to SSEA-4. (b) Sequence alignment of the two structures. Highlight similar SiaGal-recognition pattern as indicated by similar CDRH1/2 motifs.



**Figure S14: Alanine scanning of BAR-1.**

**(a)** Sequence schematic of BAR-1 and select residues targeted for mutagenesis. **(b)** ELISA EC<sub>50</sub> binding was compared against gp120-[ $\alpha$ -amidine-GM3g]<sub>16</sub> binding ( $n = 4$ ). Data were compared via Tukey's post-hoc multiple comparison test. *P*-value denotations: '\*\*\*\*'  $P < 0.0001$ , '\*\*\*'  $P < 0.001$ , '\*\*'  $P < 0.01$  and '\*'  $P < 0.05$ . **(c)** 'Pulse off' 1D NMR (black) and saturation transfer difference (STD) spectra for the various BAR-1 mutants considered, showing specifically the distinctive NAc methyl groups that terminate the Lysine moiety (Left hand peak) and the Sialic acid (Right hand peak).