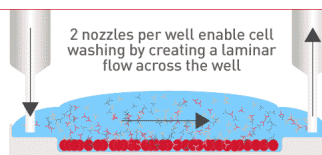


Check out how Laminar Wash systems replace centrifugation completely in handling cells



See How It Works



Biased N-Glycosylation Site Distribution and Acquisition across the Antibody V Region during B Cell Maturation

This information is current as of March 12, 2019.

Jana Koers, Ninotska I. L. Derksen, Pleuni Ooijevaar-de Heer, Benjamin Nota, Fleur S. van de Bovenkamp, Gestur Vidarsson and Theo Rispens

J Immunol published online 8 March 2019

<http://www.jimmunol.org/content/early/2019/03/05/jimmunol.1801622>

Supplementary Material <http://www.jimmunol.org/content/suppl/2019/03/05/jimmunol.1801622.DCSupplemental>

Why *The JI*? [Submit online.](#)

- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

**average*

Subscription Information about subscribing to *The Journal of Immunology* is online at: <http://jimmunol.org/subscription>

Permissions Submit copyright permission requests at: <http://www.aai.org/About/Publications/JI/copyright.html>

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at: <http://jimmunol.org/alerts>



Biased *N*-Glycosylation Site Distribution and Acquisition across the Antibody V Region during B Cell Maturation

Jana Koers,* Ninotska I. L. Derksen,* Pleuni Ooijevaar-de Heer,* Benjamin Nota,[†] Fleur S. van de Bovenkamp,*[‡] Gestur Vidarsson,[§] and Theo Rispens*

Abs can acquire *N*-linked glycans in their V regions during Ag-specific B cell responses. Among others, these *N*-linked glycans can affect Ag binding and Ab stability. Elevated *N*-linked glycosylation has furthermore been associated with several B cell-associated pathologies. Basic knowledge about patterns of V region glycosylation at different stages of B cell development is scarce. The aim of the current study is to establish patterns of *N*-glycosylation sites in Ab V regions of naive and memory B cell subsets. We analyzed the distribution and acquisition of *N*-glycosylation sites within Ab V regions of peripheral blood and bone marrow B cells of 12 healthy individuals, eight myasthenia gravis patients, and six systemic lupus erythematosus patients, obtained by next-generation sequencing. *N*-glycosylation sites are clustered around CDRs and the DE loop for both H and L chains, with similar frequencies for healthy donors and patients. No evidence was found for an overall selection bias against acquiring an *N*-glycosylation site, except for the CDR3 of the H chain. Interestingly, both IgE and IgG4 subsets have a 2-fold higher propensity to acquire Fab glycans compared with IgG1 or IgA. When expressed as rmAb, 35 out of 38 (92%) nongermline *N*-glycosylation sites became occupied. These results point toward a differential selection pressure of *N*-glycosylation site acquisition during affinity maturation of B cells, which depends on the location within the V region and is isotype and subclass dependent. Elevated Fab glycosylation represents an additional hallmark of T_H2-like IgG4/IgE responses. *The Journal of Immunology*, 2019, 202: 000–000.

B cells are essential in adaptive immunity by recognizing and neutralizing a large variety of pathogen-associated Ags via membrane-bound receptors or secreted Abs. To generate a diverse repertoire of Abs, structural diversity is initially generated by the rearrangement of variable (V), diversity (D), and joining (J) genes, followed by somatic hypermutation (SHM) and class-switch recombination. SHM is critical for Ab maturation by optimizing Ag recognition via accumulation of point mutations in the V regions (V[D]J genes) of both H and L chains. An additional layer of Ab diversification arises from the glycosylation of *N*-glycosylation sites. *N*-glycosylation sites are consensus tripeptides consisting of asparagine, followed by any amino acid except

proline, and either serine or threonine (N-X-S/T; X ≠ P). The presence of such a motif does not guarantee its glycosylation and is therefore usually referred to as a potential *N*-glycosylation site. In this article, we will use the term *N*-glycosylation site to indicate these motifs, unless indicated otherwise. Because ~15% of serum IgG has V regions that carry *N*-linked glycans (1, 2), and only few germline (GL) alleles carry *N*-glycosylation consensus tripeptides (3), the majority of V region glycosylation is, therefore, the result of SHM.

Glycosylation can have profound effects on protein function. In particular, IgG Fc glycosylation can directly modulate C1q and FcγR binding (4–8), and associations between glycosylation status and autoimmune disease have been demonstrated (9). Ag-specific IgG responses also have been shown to result in skewing toward altered fucosylation and galactosylation in certain infectious and alloimmune diseases (10–13). Less is known about the V region glycans, so-called Fab glycans, but several studies have reported a role for Fab glycans not only in modulating Ag binding (14–19) but also in increasing Ab stability (20) or extending in vivo Ab *t*_{1/2} (15, 21). Recently, evidence was provided for differential positive or negative selection for Fab glycosylation during Ag-specific Ab responses (19), which could in part be attributed to a direct effect of Fab glycans on Ag binding. Furthermore, altered levels of Fab glycosylation have been associated with pregnancy (1) as well as with several pathophysiological conditions, including follicular lymphoma (22), rheumatoid arthritis (RA) (23–25), and primary Sjögren syndrome (SS) (26, 27), and suggested for systemic lupus erythematosus (SLE) (28). In other words, differential selection pressure toward accumulation of Fab glycans is observed during various immune responses. However, the mechanisms behind these different layers of selection for or against Fab glycosylation and the extent of possible functional consequences are currently not known.

We previously showed that the distribution of *N*-glycosylation sites within the H chain V regions of memory B cells is not

*Department of Immunopathology, Sanquin Research and Landsteiner Laboratory, Amsterdam University Medical Center, University of Amsterdam, 1066 CX Amsterdam, the Netherlands; [†]Department of Research Facilities, Sanquin Research and Landsteiner Laboratory, Amsterdam University Medical Center, University of Amsterdam, 1066 CX Amsterdam, the Netherlands; [‡]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; and [§]Department of Experimental Immunohematology, Sanquin Research and Landsteiner Laboratory, Amsterdam University Medical Center, University of Amsterdam, 1066 CX Amsterdam, the Netherlands

ORCIDs: 0000-0001-7500-5760 (B.N.); 0000-0001-5621-003X (G.V.).

Received for publication December 12, 2018. Accepted for publication February 1, 2019.

This work was supported by Landsteiner Foundation for Blood Transfusion Grant 1626.

Address correspondence and reprint requests to Jana Koers, Sanquin Research, Plesmanlaan 125, 1066 CX Amsterdam, the Netherlands. E-mail address: j.koers@sanquin.nl

The online version of this article contains supplemental material.

Abbreviations used in this article: BM, bone marrow; GL, germline; H-CDR3, CDR3 of the H chain; MG, myasthenia gravis; PB, peripheral blood; RA, rheumatoid arthritis; SHM, somatic hypermutation; SLE, systemic lupus erythematosus; SRA, Sequence Read Archive; SS, Sjögren syndrome; UMI, unique molecular identifier; V_H, V gene of the H chain.

Copyright © 2019 by The American Association of Immunologists, Inc. 0022-1767/19/\$37.50

random but is largely dependent on the biased distribution of positions where a single nucleotide mutation would suffice to introduce an *N*-glycosylation site, so-called “progenitor sites” (Fig. 1A, Table I) (19). Further insights in the regulation of the loss and gain of *N*-glycosylation sites in Ab V regions may aid in understanding the selection for Fab glycosylated Abs during Ab responses in various (patho)physiological conditions. In this study, we expand previous investigation of patterns and acquisition tendencies of *N*-glycosylation sites within H and L chain Ab V regions during B cell maturation of healthy individuals and two B cell–associated diseases, SLE and myasthenia gravis (MG).

Materials and Methods

Data sets

For this study, we included data sets from three open-source efforts. From the first data set (Sequence Read Archive [SRA]: PRJNA338795) (29), nonprocessed sequence reads from peripheral blood (PB) memory B cells (CD19⁺CD27⁺/reads from IgM, IgG, IgA, κ , λ) and naive B cells (CD19⁺CD27[−]/reads from IgM) of four healthy donors and nine MG patients were collected. From the second data set (SRA: PRJEB18926) (30), nonprocessed sequence reads from PB and bone marrow (BM) mononuclear cells (reads from IgG1-4, IgA, IgM, IgE) of six allergic, but otherwise healthy, donors were obtained. In case of seasonal allergens, samples were taken out of season, and in all cases, subjects were free of symptoms at the time of sample collection. From the third data set (SRA: SRP057017) (31), nonprocessed sequence reads from PB memory B cells (CD19⁺IgD[−]CD27⁺/reads from IgM, IgG, IgA) and naive B cells (CD19⁺IgD[−]CD27[−]/reads from IgM) of two healthy donors and seven SLE patients were collected.

Data processing

A total of 13 million raw paired-end reads were assembled, quality trimmed, and filtered using Pandaseq (32) and in-house Perl and Awk scripts (sequences with a Q value <30 or unidentified nucleotides were excluded from further analysis). Isotypes were assigned using C region primer matching [primers sequences are previously described (29–31)], and a random subset of ~20,000 sequences per isotype per sample was used for V region GL alignment and mutation analysis of unique productive reads using IMGT/HighV-QUEST (33). IgM sequences were subdivided into naive IgM with zero mutation in the V gene of the H chain (V_H), thus corresponding to the GL sequence, and memory IgM with ≥1 mutation(s) in the V_H. The memory B cell repertoire was represented by PB memory IgM, IgG, and IgA and BM IgG and IgA. Biological replicates of PB and BM memory IgM, IgG, and IgA and BM IgE sequences were similar; however, this was not the case for PB IgE (30). The latter were therefore excluded from most analyses, and IgE sequences were not included in the combined memory repertoire analysis. Very few sequences of PB IgE were available for one of six donors, and these were therefore excluded from the analysis (30). Sequences of the four IgG subclasses were processed as described above. Sequences from two biological replicates for five healthy donors were available from both PB and BM IgG1-4 B cells. One of six donors had essentially no IgG4 sequences and was excluded from the analysis. In one of the studies, next-generation sequencing was carried out using unique molecular identifiers (UMIs) (29). When using UMI-guided read processing, as described previously (29), error rates decreased with ~1 error per sequence, illustrating the improved sequence quality when using UMI-adopted sequencing approaches. However, there was no appreciable difference in frequency distributions of *N*-glycosylation sites between the different types of processing.

Analysis of *N*-glycosylation sites

N-glycosylation site (N-X-S/T, X≠P) mapping and matching to V region progenitor sites was carried out essentially, as previously described, using either a combination of custom VBScripts and SQL queries in Microsoft Access (19) or custom R scripts. Additional data processing was also carried out using custom R scripts. *N*-glycosylation sites within the CDR3 of the H chain (H-CDR3) were subdivided according to their position in either the left junction, which comprises sites at position 105 up to the start of the D gene, or the right junction, which comprises sites from the end of the D gene up to position 117, and for

designation of a D gene site, the entire nonameric sequon should fall within the IMGT D3-REGION. Sequences missing the D gene were excluded from further analysis. To analyze mutation propensities at progenitor sites, we subdivided progenitor sites as being type I, IIa, and IIb, as described in Results. In brief, for each type of progenitor site, one or two of the three possible mutations at a specific nucleotide position result in a glycosylation site, and those can be compared with the other one or two possible mutations at that same nucleotide position that do not result in an *N*-glycosylation site. For each unique progenitor site, frequencies of each of the three possible mutations were determined and compared. *N*-glycosylation sites that emerged through multiple mutations within the triplet were excluded, and therefore potential cases of codependent mutations were not considered in this analysis. In addition, framework region (FR) one was excluded from the analysis as most 5′ primers annealed to this region. Only progenitor sites for which >1500 sequence reads were available were included in the analysis.

Ab production and purification

rmAbs were produced and purified as previously described (19). In brief, synthetic V_H and V gene of the L chain constructs were separately cloned into pcDNA3.1 expression vectors containing the corresponding constant genes and cotransfected in HEK293F cells. After a 5-d incubation at 37°C in humidified 8% CO₂, Abs were purified from culture supernatant using protein G Sepharose. For the generation of Abs with GL *N*-glycosylation sites, we used clone mAb63 for IGHV4-34 (34). For IGHV1-8, we expressed an IGHV1-2 clone (anti-ADL 1.2) (35) with and without a D81N mutation, which makes the clone homologous to an IGHV1-8 gene [i.e., TSTR(N/D)TSISTA versus TMTRDTSISTA]. Similarly, for IGHV5-10-1, we expressed guselkumab (IGHV5-51) with and without an R66N mutation, making the sequence homologous to IGHV5-10-1 [i.e., DSDT(N/R)YSPSFQ versus DSYTRYSPSFQ].

Detection of *N*-linked glycans at V regions of mAbs

Detection of *N*-glycosylation site occupancy was performed as described previously (19). In brief, glycosylation of rAbs was analyzed by SDS-PAGE and lectin ELISAs, using *Sambucus nigra* agglutinin (Vector Laboratories) for the detection of sialic acid. All mAbs that showed increased m.w. by SDS-PAGE also scored positive in the *S. nigra* agglutinin lectin ELISA.

Statistical tests

Statistical comparisons were performed using two-tailed Student *t* test, one-way ANOVA, or Wilcoxon signed-rank test. Measures for correlation used Pearson or Spearman *r* test. Differences were considered statistically significant when the *p* values were <0.05 (**p* < 0.05, ***p* < 0.01, and ****p* < 0.001). Statistical analyses were performed using GraphPad Prism 7.04.

Results

Distributions, but not frequencies, of non-GL *N*-glycosylation sites are similar among the different memory B cell repertoires

The frequency distributions of non-GL *N*-glycosylation sites (Table I) in naive and memory B cell populations were evaluated for 12 healthy individuals (~1.5 million sequence reads, Supplemental Table I). The frequencies of *N*-glycosylation sites were lowest for naive IgM, followed by IgM memory and κ and λ L chains, and highest for switched memory B cells (Fig. 1B). Interestingly, the frequency of sites was similar for IgG and IgA but significantly higher in BM IgE. The biological replicates of PB IgE yielded completely dissimilar results (30), hence the large SD, and were therefore not taken into account for subsequent analyses. Despite differences in absolute frequencies, the overall distribution of non-GL *N*-glycosylation sites across the H chain V regions was fairly similar for the different PB and BM memory B cell populations and clustered around the CDRs and the DE loop (Fig. 1C). The main exception are *N*-glycosylation sites within FR4 position 125 that were significantly more frequent in IgA compared with the other memory repertoires. *N*-glycosylation sites found in V regions

Table I. Description of V region *N*-glycosylation site subsets

Subset	Description
GL site	(Potential) <i>N</i> -glycosylation site present in GL arrangement of a V gene allele
Non-GL site	All (potential) <i>N</i> -glycosylation sites not corresponding to GL sites. These are the result of SHM or, within the CDR3, can also result from junctional diversity.
Progenitor site	Sequon in the GL that can become an <i>N</i> -glycosylation site with a single nucleotide mutation (see Fig. 4C). Most glycosylation sites in the memory repertoire correspond to these. Also resulted from SHM.
Nonprogenitor site	Other non-GL <i>N</i> -glycosylation sites. These include those within the CDR3 that resulted from junctional diversity as well as those that resulted from two or more mutations during SHM.

of memory B cell L chain sequences ($n = 4$) showed similar clustering of *N*-glycosylation sites around the CDRs and DE loop as observed for H chain sequences (Fig. 1C, κ and λ). Of note, between individual donors, the frequency of sites found at a specific position varies considerably for all repertoires.

Nonprogenitor sites found within and outside of the CDR3

In the V and J genes, the majority of *N*-glycosylation sites within memory rearranged sequences correspond to “progenitor sites” (Table I) (19) ranging from 77 to 88% for H chains and 79 and 88% for κ and λ L chains, respectively. The remaining nonprogenitor sites can be subdivided as either randomly introduced within the CDR3 V-D and D-J/V-J junctions during V(D)J recombination or is the result of two or more mutations. Nonprogenitor sites found in the H chains also have the tendency, although less pronounced, to cluster around the CDRs and DE loop (Supplemental Fig. 1A). Naive B cells carry, besides GL sites (not shown in Fig. 1, see below), *N*-glycosylation sites within V-D and D-J junctions of the H-CDR3 (Fig. 1C, naive IgM), with significantly higher tendency to cluster within V-D junctions (Fig. 1D). The frequency of *N*-glycosylation sites within the D genes is low in the naive repertoire but increases significantly in memory B cells. Despite these changes, the total frequency of sites within the H-CDR3 does not increase upon B cell maturation (see below). Of class-switched B cells found in BM, IgA-expressing B cells have fewer *N*-glycosylation sites within the D gene compared with IgG.

Negative selection for GL *N*-glycosylation sites in the memory repertoire

Three *N*-glycosylation sites are encoded in the GL IGHV repertoire (alleles IGHV1-8, IGHV4-34, and IGHV5-10-1). These were analyzed separately for their frequency in both the naive and memory repertoires (Fig. 2A). In memory B cells, the frequencies of GL sites at position 81 (IGHV1-8) and position 57 (IGHV4-34) are significantly lower than those in naive IgM B cells, both because the V_H gene allele frequencies are lower and, furthermore, because a substantial part of the GL sites is reverted to nonglycosylation sites by SHM. This trend was less pronounced for position 66 in the less-prevalent allele IGHV5-10-1 (which was observed in only 4 out of 12 donors). Similar patterns were observed for BM-derived switched memory B cells (Fig. 2B). Interestingly, GL site IGHV1-8 was retained in most cases for BM IgE (although the allele frequency did decrease compared with naive IgM B cells), whereas the site disappeared because of

mutations in at least 40% of the cases for IgG or IgA. V gene of the L chain alleles containing GL sites (IGKV5-2, IGLV3-12, IGLV3-22, IGLV5-37) were observed infrequently ($\leq 0.06\%$) and were not further analyzed.

Signs of negative selection for acquisition of *N*-glycosylation sites in the H-CDR3

Next, we evaluated the relationship between acquiring *N*-glycosylation sites and the accumulation of nonsilent somatic mutations in the H chain V regions (Fig. 3A). The percentage of sequences that carry non-GL *N*-glycosylation sites was highly dependent on the number of mutations and increases rapidly up to 15 mutations, after which a plateau is established (Fig. 3B, 3C). Furthermore, nonprogenitor sites within V_H and J_H genes accumulate steadily but more slowly with increasing number of mutations (Fig. 3D). By contrast, the percentage of sequences carrying *N*-glycosylation sites within the H-CDR3 is essentially invariant with respect to the number of somatic mutations (Fig. 3D). This implies a negative selection bias for glycosylation sites within the H-CDR3 compared with the overall V region.

Progenitor site and mutation propensity dictate the *N*-glycosylation sites present in the memory repertoire

Next, we examined to which extent the frequency of acquired *N*-glycosylation sites at progenitor site positions depends on the frequency of the progenitor sites. Fig. 4A shows frequencies of progenitor sites, grouped per position for most commonly used V_H gene families, with corresponding *N*-glycosylation sites found in the memory repertoire. Certain progenitor sites are frequent but rarely mutate into an actual *N*-glycosylation site (e.g., position 56 in IGHV3) or the other way around (less-common progenitor sites that relatively frequently mutate into actual *N*-glycosylation sites, e.g., position 86 in IGHV4). For other positions, both the progenitor site and the *N*-glycosylation site are frequently found (e.g., position 90 in IGHV4). In other words, acquiring *N*-glycosylation sites during SHM is not solely dictated by the presence of a progenitor site because the observed mutation frequencies differ substantially between positions. Nevertheless, a moderate correlation was found between the number of progenitor sites (i.e., allele frequency) and the number of *N*-glycosylation sites (Fig. 4B; Pearson $r = 0.55$, $p < 0.001$).

To analyze the propensity for mutation toward *N*-glycosylation sites in more detail, we compared progenitor site mutations resulting in *N*-glycosylation sites to the other two possible mutations at the same nucleotide position. For the majority of progenitor sites (designated type I; an example is shown in Fig. 4C), only one specific “replacement” mutation results in an *N*-glycosylation site (Rglyc), whereas the two other possible mutations correspond to other nonsilent or replacement mutations (Ro). Analysis of all unique type I nonameric progenitor sites across V_H gene families revealed a large variation in mutation frequencies for both glycan-introducing and other replacement mutations across progenitor sites (Fig. 4D). A moderate to strong correlation was observed between either the Rglyc and the most frequent other mutation (Ro1; $r = 0.69$, $p < 0.0001$) and the least frequent other mutation (Ro2; $r = 0.55$, $p < 0.0001$) at each particular site. On average, the Rglyc mutation was slightly more frequent than Ro2 (Fig. 4E, median = 1.51, $p < 0.0001$) and about three times less frequent as Ro1 (median = 0.30, $p < 0.0001$). This means that the overall propensity to acquire a glycan-introducing mutation falls within the range of acquiring another replacement mutation at that position but with a wide range of ~0.01- to 100-fold for individual positions. Furthermore, the relative propensity for glycan-introducing mutations (relative to the other mutations, i.e., Ro1

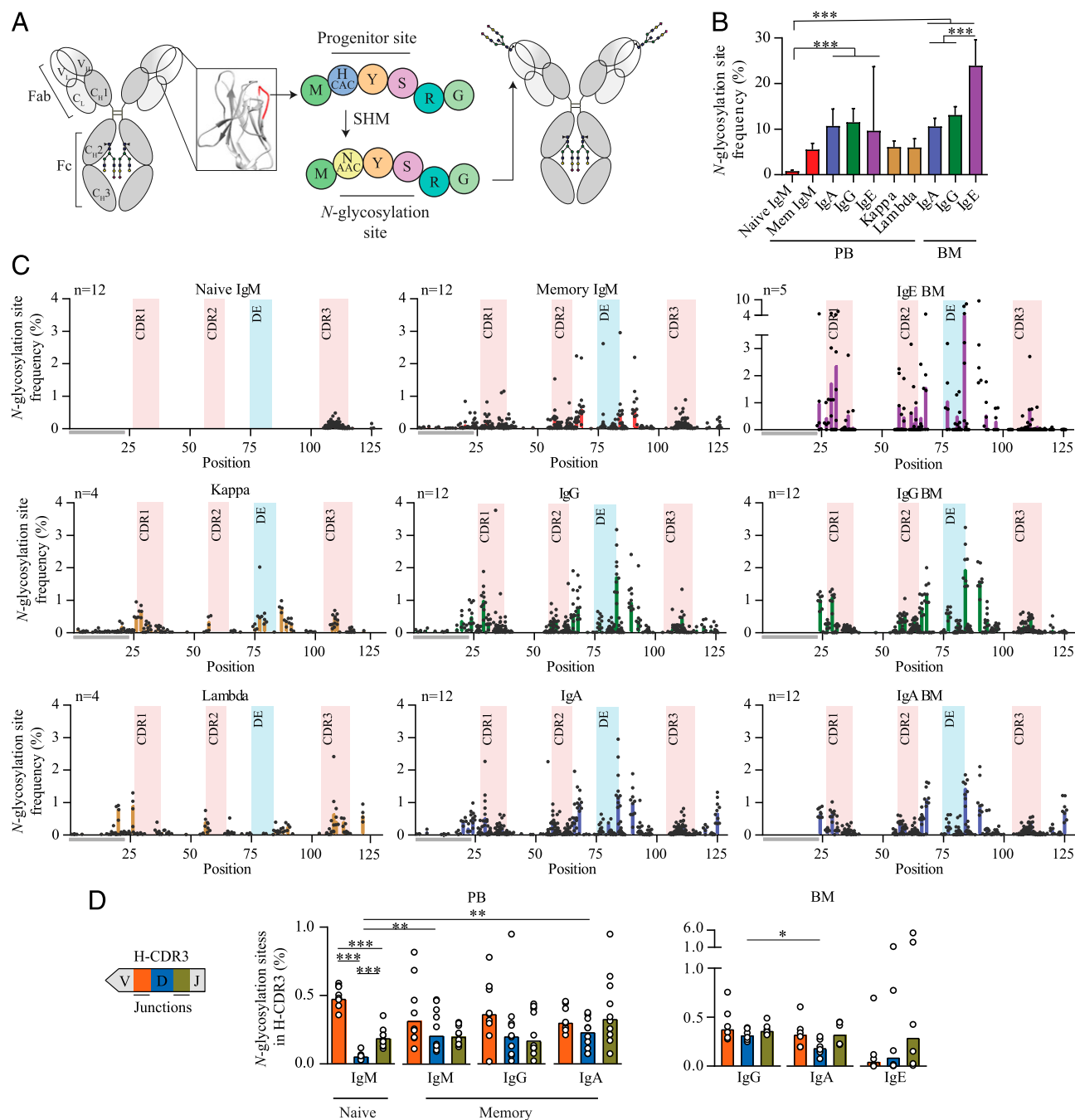


FIGURE 1. Frequency distribution of non-GL N-glycosylation sites across Ab V region sequences for healthy individuals. **(A)** Schematic representation of N-glycosylation site acquisition by SHM at a progenitor site in the Ab V region. **(B)** The frequency of N-glycosylation sites (in percent) per isotype in PB and BM B cells. Bars indicate medians, error bars $\pm$ SD, with the number of individuals indicated in **(C)**. **(C)** The frequency (in percent) distributions of non-GL N-glycosylation sites across the V region are shown for naive (IgM) and memory PB B cells (IgM, IgG, IgA, κ , and λ) and switched memory BM B cells (IgG, IgA, IgE). The number of individuals is indicated in the upper left corner of each frequency distribution. Each dot represents a single donor; bars represent medians. Horizontal gray bars indicate FR1, for which coverage is incomplete (see *Materials and Methods*). Sites within the CDR3 at positions 111.1–112.1 are collapsed into position 111. **(D)** Frequency (in percent) of N-glycosylation sites across junctions and D genes in the H-CDR3 for PB and BM B cells. Each dot represents a single donor; bars represent median. (B and D) *** p < 0.001, ** p < 0.01, * p < 0.05, one-way ANOVA.

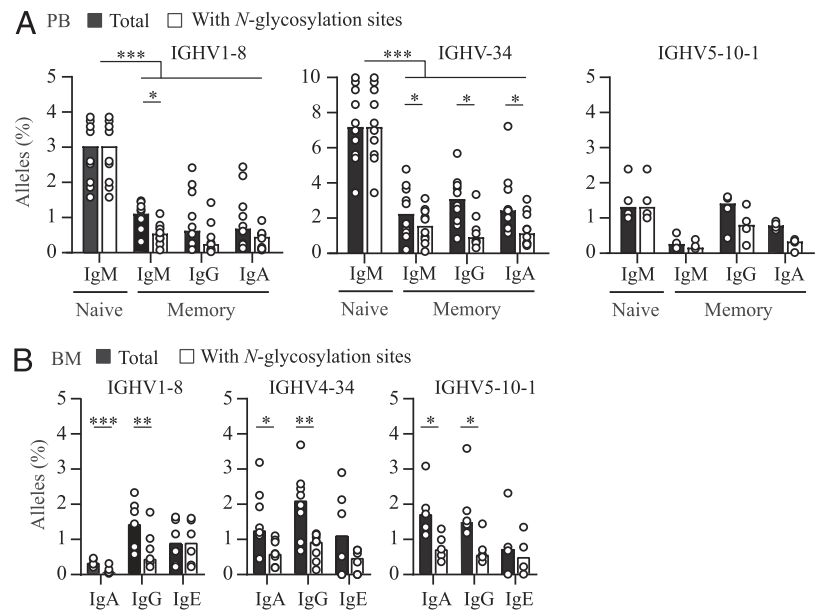
and Ro2) observed for the CDRs is similar for the FRs, despite a higher overall mutation frequency for the CDRs (Fig. 4F, 4G). Similar trends were observed for two other less-frequent groups of progenitor site type IIa/b mutations (examples are shown in Supplemental Fig. 1B), for which two out of three mutations at a single nucleotide position result in a glycosylation site (Supplemental Fig. 1C). Taken together, these results indicate that

there is no pronounced selection pressure against acquiring N-glycosylation sites within the V region during B cell maturation.

T_H2-associated Abs have increased levels of N-glycosylation sites

Because we observed a 2-fold higher frequency of non-GL N-glycosylation sites for IgE in comparison with the other

FIGURE 2. Loss of GL *N*-glycosylation sites in memory B cell repertoires. **(A)** GL site and corresponding allele frequencies (in percent) in naive and memory PB B cells for 12 healthy individuals. The three GL sites correspond to positions 81, 57, and 66 in IGHV1-8, IGHV4-34, and IGHV5-10-1, respectively. The latter allele was only observed in 4 out of 12 donors. **(B)** GL site and corresponding allele frequencies (in percent) in memory BM B cells for six healthy individuals. Bars represent medians, each dot represents one donor. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, one-way ANOVA.



switched isotypes, we also examined *N*-glycosylation site frequencies for the IgG subclasses, including IgG4, which is also associated with T_H2 -like Ab responses. We previously observed an increased frequency of *N*-glycosylation sites in PB IgG4 V region sequences, albeit in a small data set (19). In this article, we confirm and extend this observation for five individuals for which IgG subclass-specific sequences were available for both PB and BM IgG B cells. Both PB and BM IgG4 Abs have significantly higher frequencies of *N*-glycosylation sites in comparison with the other IgG subclasses (Fig. 5A) and similar to BM IgE. The elevated number of *N*-glycosylation sites in IgG4 and IgE could not be explained by elevated levels of SHM (Fig. 5B). Furthermore, there were no obvious differences in V_H allele usage that could clarify these differences in *N*-glycosylation site frequencies

(data not shown). Taken together, the elevated levels of *N*-glycosylation sites for both IgE and IgG4 indicate an altered selection pressure for the V region Ab repertoire for T_H2 -like Ab responses.

Bulk B cell *N*-glycosylation patterns in autoimmune patients resembles those of healthy individuals

Next, we assessed the acquisition of V region *N*-glycosylation sites in the general Ab repertoire of patients with autoantibody-mediated immune diseases (i.e., SLE [$n = 7$] and MG [$n = 9$]). *N*-glycosylation site frequency, distribution, and repertoire composition of bulk memory B cell repertoires of these patients were very similar to those of healthy individuals (distributions not shown; Supplemental Fig. 2A–D). In other words, there does

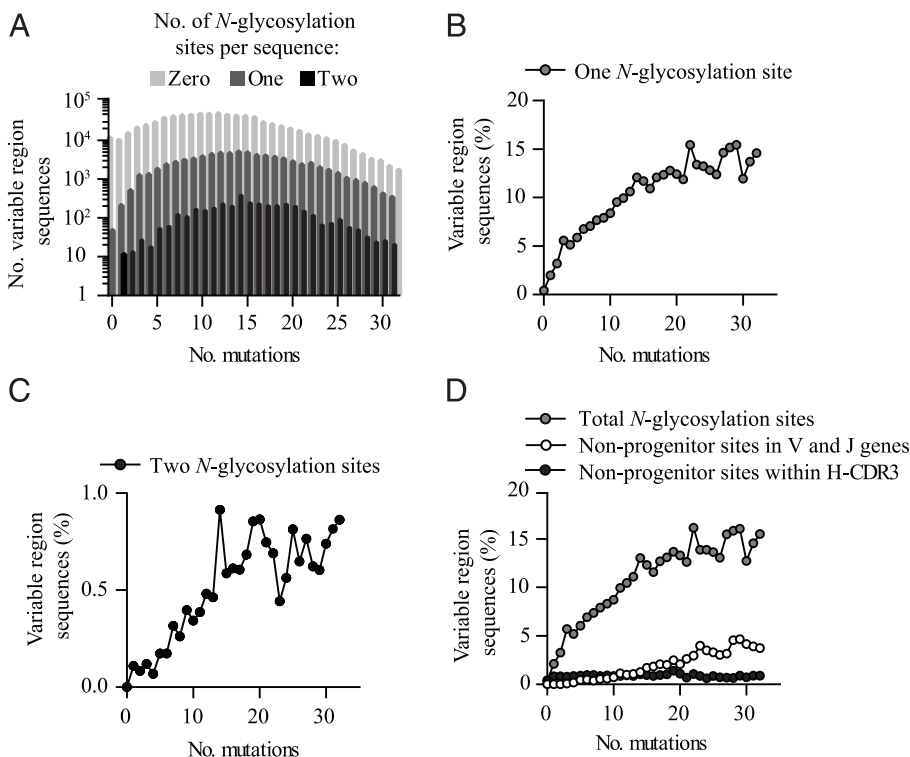


FIGURE 3. Negative selection pressure for non-GL *N*-glycosylation sites in H-CDR3. **(A)** Number of V_H sequences containing zero, one, or two *N*-glycosylation sites versus the number of nonsilent mutations found in the sequence. Sequences of 12 healthy individuals for memory IgM, IgG, and IgA were collapsed as their individual patterns showed high resemblance. **(B)** and **(C)** Percentage of V_H sequences acquiring one (B) or two (C) *N*-glycosylation sites in the V region versus the number of nonsilent mutations. **(D)** Percentages of sequences with non-GL *N*-glycosylation sites in the complete V region, nonprogenitor sites across V and J genes, and within the H-CDR3 versus the number of nonsilent somatic mutations.

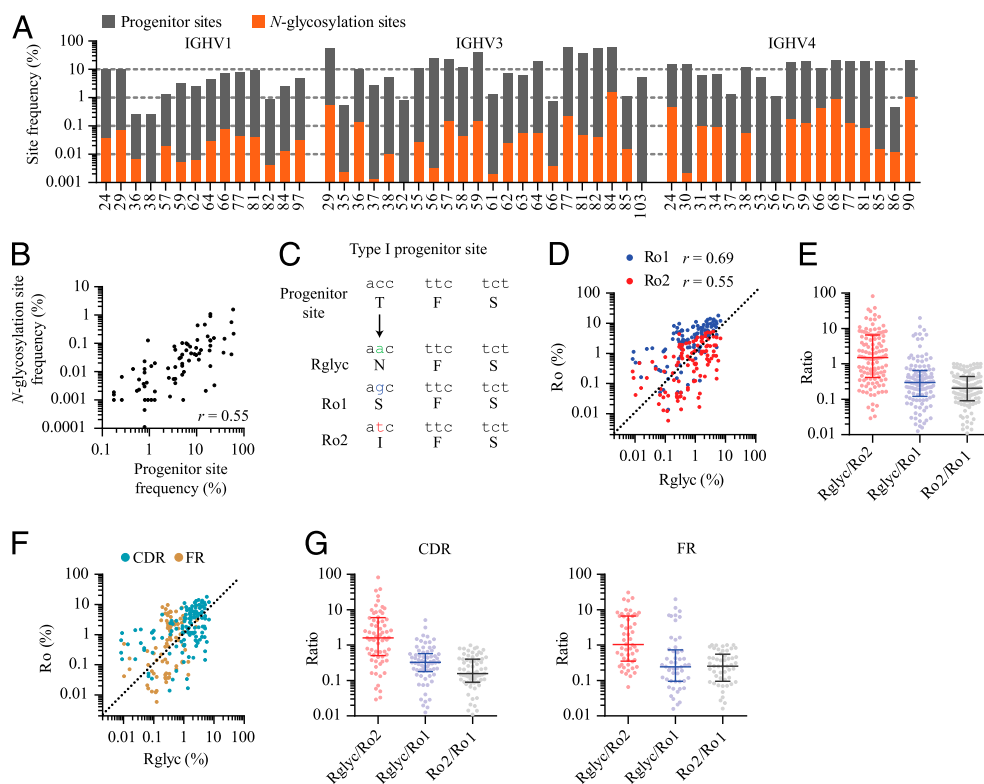


FIGURE 4. The *N*-glycosylation site repertoire is dictated by progenitor site frequency and site-specific mutation propensities. **(A)** Frequencies (in percent) of progenitor sites (gray) versus acquired *N*-glycosylation sites (orange), grouped according to V_H family for 12 healthy individuals (less frequently used V_H families are not shown; nonmatched sites and progenitor sites with a frequency <0.2% are excluded). **(B)** Correlation between progenitor site and *N*-glycosylation site frequencies (in percent). **(C)** Example of a type I GL nonameric progenitor site nucleotide motif. **(D and E)** Scatterplot and ratios of glycan-introducing mutations (Rglyc) stratified to either of the two other replacement mutations (Ro1 and Ro2, the most and least frequent other mutation, respectively) for all type I progenitor sites. Spearman *r* and Wilcoxon signed-rank test. **(F and G)** As in (D) and (E) but stratified according to the progenitor site being in the CDR or FR.

not appear to be gross alterations in Fab glycosylation patterns in these B cell-associated autoimmune diseases.

Abs with N-glycosylation sites at progenitor sites are frequently glycosylated

The presence of an *N*-glycosylation site does not guarantee that sites become occupied with glycans because this is also influenced by the nature of the core amino acid, adjacent amino acids, and the overall folding of the protein (36–38). In this study, we examined the actual glycosylation tendencies for both GL and non-GL V region *N*-glycosylation sites. First, to investigate if the negative selection for GL sites in the memory repertoire may be attributed to glycosylation of these sites, Abs corresponding to GL alleles IGHV1-8, IGHV4-34, and IGHV5-10-1 were expressed, and glycosylation of the three GL *N*-glycosylation sites was analyzed using gel electrophoresis. Abs with Fab glycans exhibit an increased m.w. compared with their nonglycosylated counterparts (Fig. 6A). We observed that the GL site at position 57 in IGHV4-34 is not glycosylated, whereas GL sites at position 81 in IGHV1-8 and 66 in IGHV5-10-1 do become fully and partially glycosylated, respectively.

Next, we investigated the glycosylation of 38 non-GL *N*-glycosylation sites located at progenitor site positions in 27 different recombinant expressed mAbs [including previously described (18, 19, 34–37, 39–41) and newly examined; for details see Supplemental Table II]. From these *N*-glycosylation sites, 35 out of 38 (92.1%) become at least partially occupied by glycans as confirmed by gel electrophoresis, lectin ELISA, or mass spectrometry (Fig. 6B and Supplemental Fig. 2E, 2F). In addition, we did not observe preferential glycosylation of

mAbs that carry the N-X-S sequon compared with N-X-T (90.5% and 94.1%, respectively). These results suggest that in general, the acquisition of *N*-glycosylation sites located on progenitor sites during SHM will lead to expression of Abs with V region glycans. A major exception is the progenitor sites at position 57 in V_H family IGHV4, which are homologous to the GL site in IGHV4-34. Introduction of this *N*-glycosylation site into two different IGHV4 alleles (IGHV4-4 and IGHV4-39) did not lead to the introduction of a glycan (Fig. 6C). In summary, although not every position across the V region has been tested, these results suggest that the majority of the *N*-glycosylation sites introduced by SHM during B cell maturation have the potential to become glycosylated.

Discussion

Glycosylation of Ab V regions stands out as an additional layer of repertoire diversification, being dependent on the primary amino acid structure as well as subsequent events, resulting in attachment and further processing of glycans. In this study, we analyzed in detail the patterns and biases of *N*-glycosylation sites across naive and memory B cell subsets. With the exception of the H-CDR3, we did not find evidence for an overall selection bias against the introduction of *N*-glycosylation sites in the V region. This is noteworthy given that introduction of a large bulky glycan may be a disruptive event in terms of folding or Ag binding.

Interestingly, we observed an almost 2-fold higher propensity for both IgE and IgG4 to acquire glycans compared with IgG1 or IgA, indicating that T_H2-like responses develop a skewed V region repertoire.

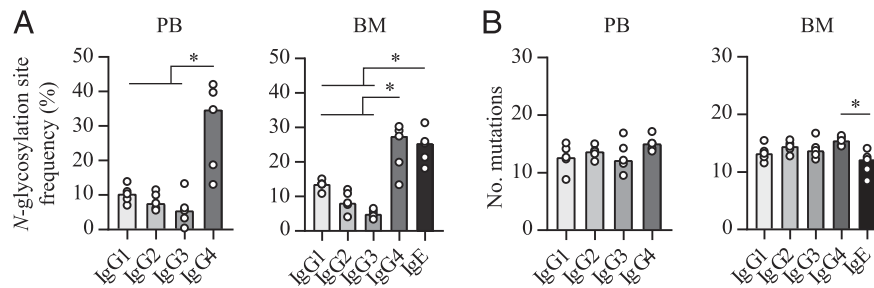


FIGURE 5. Enrichment of non-GL *N*-glycosylation sites in T_H2 -associated Abs IgG4 and IgE. **(A)** Frequency (in percent) of V_H sequences containing *N*-glycosylation sites for PB and BM IgG subclasses and BM IgE sequences for five healthy individuals. Each dot represents a single donor, bars represent medians. **(B)** The number of replacement mutations within the sequences described in (A). Each dot represents a single donor ($n = 5$), bars represent medians. $*p < 0.05$, one-way ANOVA.

The 2-fold increase of acquired *N*-glycosylation sites for IgE and IgG4 could not be explained by increased levels of somatic mutations in these isotypes nor by a biased V gene usage. IgG4 and IgE responses have in common that they are often produced during so-called T_H2 responses, characterized by IL-4 and IL-5, the former of which is known to induce class switching to both IgG4 and IgE (38). Although glycans can positively contribute to Ag binding (14–18), it is unlikely that Ags that provoke a T_H2 -like response would generally be better recognized by Abs carrying glycans. Alternatively, there may be (endogenous) sugar-binding lectins involved in T_H2 -skewed responses that pose a selective survival advantage to B cells with Fab glycosylated B cell receptors. For IgE, there are indications for impaired germinal center responses (39, 40), which tentatively might provide the environment that positively selects for enhanced levels of Fab glycosylation. IgG4 Ab responses are associated with tolerance, which make it feasible that Fab glycans contribute to a tolerogenic phenotype by dampening Ag binding or interaction with lectins on other immune cells and thereby exerting immune modulatory properties. Of note, we have recently observed elevated levels of Fab glycosylation on serum IgG4 from patients with IgG4-related disease (44). A disease characterized, among others, by abnormally high

levels of infiltrating IgG4⁺ plasma cells into one or more affected organs. The potential consequences of elevated Fab glycosylation levels of IgG4 and IgE remain to be investigated.

It is relevant to point out that the biases observed in this study do not solely reflect selection. First, the nonrandom clustering of progenitor sites in the GL repertoire (19) already predetermines to a substantial degree at which positions *N*-glycosylation sites are introduced. Second, the intrinsic mutability of a GL gene varies substantially depending on its sequence, and the distribution of so-called “mutation hot- and cold-spot motifs” influence overall mutation patterns during SHM (45, 46). Furthermore, because the overall patterns for acquiring *N*-glycosylation sites at particular positions reflect the average of a multitude of individual, Ag-specific B cell responses, it is not possible to directly link such patterns to Ag-driven selection, similar to other amino acid substitution profiles (47). For example, the negative bias for *N*-glycosylation site acquisition in the H-CDR3 in general does not exclude the occasional positive contribution of an H-CDR3 glycan to Ag binding. In contrast, it is feasible that alterations in glycosylation patterns may, for example, associate with certain physiological or pathological conditions. In this regard, the hyper Fab glycosylation observed for RA-specific autoantibodies

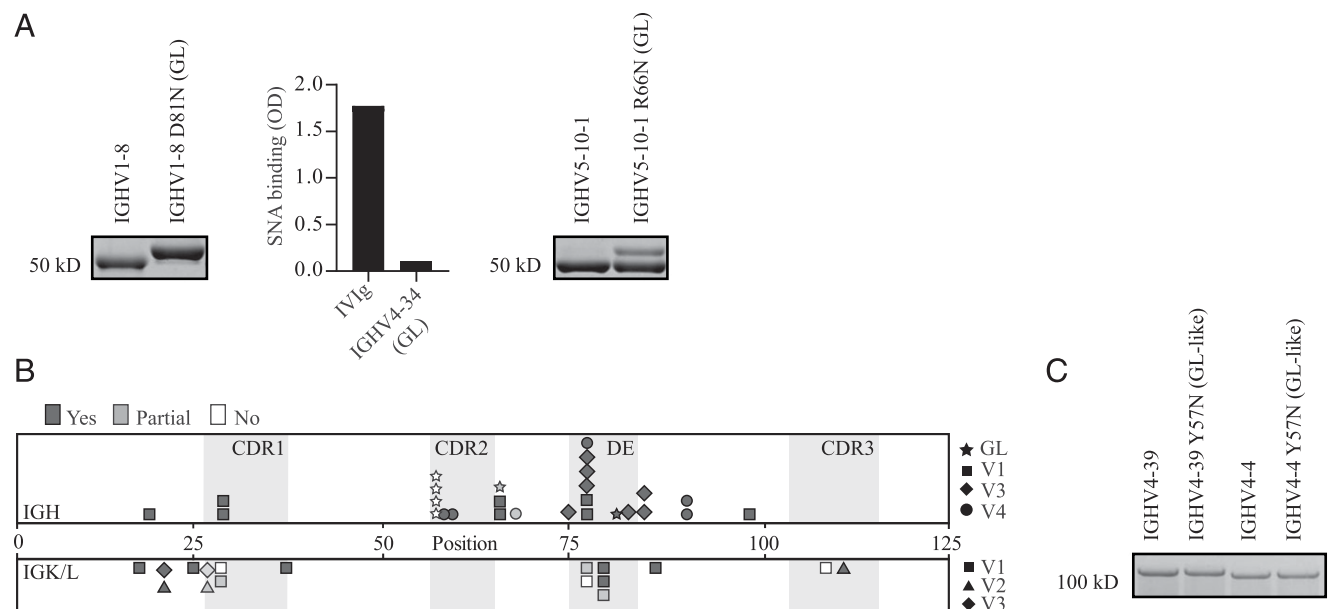


FIGURE 6. Glycosylation of *N*-glycosylation sites in mAbs. **(A)** Gel electrophoresis of GL *N*-glycosylation sites. Left, Reduced SDS-PAGE of full-length IgG1 IGHV1-8, H chain at 50 kDa. Middle, Lectin ELISA of F(ab')₂ IGHV4-34. Right, Reduced SDS-PAGE of full-length IgG1 IGHV5-10-1, H chain at 50 kDa. **(B)** Schematic overview of the glycosylation of 38 *N*-glycosylation sites (in 27 different mAbs) at various progenitor site positions. Glycosylation was confirmed by gel electrophoresis, lectin ELISA, or mass spectrometry as detailed in Supplemental Table II. **(C)** Gel electrophoresis of IGHV4-34 GL site replacement in IGHV4-39 and IGHV4-4. Nonreduced SDS-PAGE of F(ab')₂ fragments, at 100 kDa.

(>90%) (36) or the increased frequencies of *N*-glycosylation sites in parotid gland-derived B cells of primary SS patients (26, 36) represent an important research topic. Nevertheless, bulk serum IgG in RA patients does not appear to be aberrantly Fab glycosylated. In this article, we show that this is also true for the bulk memory repertoire of SLE and MG patients. Thus, despite the presence of relatively large amounts of autoantibodies in some patients, most of the circulating Abs are not autoreactive. Therefore, a differential glycosylation pattern of these autoantibodies cannot be revealed from bulk repertoire analysis. Elevated levels of Fab glycosylation have been suggested as a general feature of autoimmunity (28). However, the observed frequencies of *N*-glycosylation sites within bulk B cell repertoires of autoimmune disease patients fall well within the range of those observed in this study for healthy individuals. Therefore, aberrant Fab glycosylation may be mostly restricted to disease-specific autoreactive B cells, for reasons yet to be unraveled. Altered autoreactivity of Fab glycosylated Abs poses a potential mechanism. A recent paper described positive selection for Fab glycosylated autoantibodies in SS patients due to enhanced autoantigen binding (48). However, both enhanced and diminished binding was observed for autoantibodies associated with RA (36), and Fab glycans have also been suggested to contribute to redemption of autoreactivity (49).

Bulky glycan structures in suboptimal positions may interfere with Ab folding and function. Therefore, one might expect that the introduction of *N*-glycosylation sites is restricted. We observed that *N*-glycosylation sites readily emerge throughout the V region with similar distributions of sites among the different Ab classes that clustered around the Ag-binding regions. A striking exception was an IgA-restrictive site at position 125. It would be of interest to find out if this site becomes glycosylated and whether it serves a purpose in IgA physiology. IgA Abs associated with the intestinal lumen may benefit from extensive glycosylation as it can offer protection for the many residing proteases (50). For one region in particular, H-CDR3, we observed that relatively few sequences acquired *N*-glycosylation sites, irrespective of the accumulation of mutations. Interestingly, of all CDRs, H-CDR3 is most important for Ag recognition and is the epicenter of most Ag–Ab contacts (51). It is therefore likely that introduction of a glycan in the H-CDR3 is a highly unfavorable event, which can explain the local negative selection for *N*-glycosylation sites.

Somatic mutations can introduce but also remove *N*-glycosylation sites. For GL *N*-glycosylation sites, the preservation of the sites in the memory repertoire was an unfavorable event, both in terms of selection against V_H alleles carrying GL sites as well as substantial loss of sites due to SHM. In the case of IGHV1-8, it is possible that the occurring negative selection is attributed to the presence of a glycan on the GL *N*-glycosylation site. By contrast, the GL site in IGHV4-34 is not glycosylated; thus glycosylation can, therefore, not account for the negative selection pressure. However, the IGHV4-34 allele has been associated with autoreactivity and is therefore negatively selected in general (52). The main contribution of GL sites to the comprehensive V region *N*-linked glycan repertoire of the memory B cell population derives from IGHV1-8. With taking into account the infrequent occurrence of IGHV5-10-1, we estimate that GL sites account for ~5% of the glycans found on serum IgG and IgA.

It should be kept in mind that beyond the presence or absence of a glycan, its structural profile can vary substantially from one *N*-linked glycan to another (53). The essential first event during glycosylation is the attachment of a high-mannose structure to the asparagine residue of the *N*-glycosylation site during Ab translation and folding in the endoplasmic reticulum. The

presence of any glycan, regardless of its precise structure, depends on this first step. Hereafter, substantial multistep glycan trimming and build-up takes place in the Golgi, with a wealth of possible outcomes (9). Transfer of *N*-linked glycans to prototypic *N*-glycosylation sites does not always occur and is, among others, constrained by the accessibility of the *N*-glycosylation site for glycan-attaching enzymes. In this study, we show that the majority of *N*-glycosylation sites introduced at progenitor sites by SHM are glycosylated (Fig. 6C), which suggest that progenitor sites in general are located at positions that allow facile glycosylation. In summary, this study identified additional factors that shape the Ab V region *N*-glycosylation repertoire. These results indicate that a differential selection pressure for glycans during B cell responses exists, with a profound skewing during T_H2 -like responses.

Disclosures

The authors have no financial conflicts of interest.

References

- Bondt, A., Y. Rombouts, M. H. J. Selman, P. J. Hensbergen, K. R. Reidling, J. M. W. Hazes, R. J. Dolhain, and M. Wührer. 2014. Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol. Cell. Proteomics* 13: 3029–3039.
- van de Bovenkamp, F. S., L. Hafkenscheid, T. Rispen, and Y. Rombouts. 2016. The emerging importance of IgG Fab glycosylation in immunity. *J. Immunol.* 196: 1435–1441.
- Lefranc, M. P. 2011. IMGT, the international ImmunoGeneTics information system. *Cold Spring Harb. Protoc.* 2011: 595–603.
- Sondermann, P., R. Huber, V. Oosthuizen, and U. Jacob. 2000. The 3.2-Å crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature* 406: 267–273.
- Ferrara, C., S. Grau, C. Jäger, P. Sondermann, P. Brunker, I. Waldhauer, M. Hennig, A. Ruf, A. C. Rufer, M. Stihle, et al. 2011. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc gammaRIII and antibodies lacking core fucose. *Proc. Natl. Acad. Sci. USA* 108: 12669–12674.
- Quast, I., C. W. Keller, M. A. Maurer, J. P. Giddens, B. Tackenberg, L. X. Wang, C. Münz, F. Nimmerjahn, M. C. Dalakas, and J. D. Lünemann. 2015. Sialylation of IgG Fc domain impairs complement-dependent cytotoxicity. *J. Clin. Invest.* 125: 4160–4170.
- Peschke, B., C. W. Keller, P. Weber, I. Quast, and J. D. Lünemann. 2017. Fc-galactosylation of human immunoglobulin gamma isotypes improves clq binding and enhances complement-dependent cytotoxicity. *Front. Immunol.* 8: 646.
- Dekkers, G., L. Treffers, R. Plomp, A. E. H. Bentlage, M. de Boer, C. A. M. Koeleman, S. N. Lissenberg-Thunnissen, R. Visser, M. Brouwer, J. Y. Mok, et al. 2017. Decoding the human immunoglobulin G-glycan repertoire reveals a spectrum of fc-receptor- and complement-mediated-effector activities. *Front. Immunol.* 8: 877.
- Dekkers, G., T. Rispen, and G. Vidarsson. 2018. Novel concepts of altered immunoglobulin G galactosylation in autoimmune diseases. *Front. Immunol.* 9: 553.
- Kapur, R., I. Kustiawan, A. Vestreheim, C. A. M. Koeleman, R. Visser, H. K. Einarsdottir, L. Porcelijn, D. Jackson, B. Kumpel, A. M. Deelder, et al. 2014. A prominent lack of IgG1-Fc fucosylation of platelet alloantibodies in pregnancy. *Blood* 123: 471–480.
- Kapur, R., L. Della Valle, M. Sonneveld, A. Hipgrave Ederveen, R. Visser, P. Ligthart, M. de Haas, M. Wührer, C. E. van der Schoot, and G. Vidarsson. 2014. Low anti-RhD IgG-Fc-fucosylation in pregnancy: a new variable predicting severity in haemolytic disease of the fetus and newborn. *Br. J. Haematol.* 166: 936–945.
- Wang, T. T., J. Sewatanon, M. J. Memoli, J. Wrammert, S. Bournazos, S. K. Bhaumik, B. A. Pinsky, K. Chokephaibulkit, N. Onlamoon, K. Pattanapanyasat, et al. 2017. IgG antibodies to dengue enhanced for FcγRIIIA binding determine disease severity. *Science* 355: 395–398.
- Ackerman, M. E., M. Crispin, X. Yu, K. Baruah, A. W. Boesch, D. J. Harvey, A. S. Dugast, E. L. Heizen, A. Ercan, I. Choi, et al. 2013. Natural variation in Fc glycosylation of HIV-specific antibodies impacts antiviral activity. *J. Clin. Invest.* 123: 2183–2192.
- Wright, A., M. H. Tao, E. A. Kabat, and S. L. Morrison. 1991. Antibody variable region glycosylation: position effects on antigen binding and carbohydrate structure. *EMBO J.* 10: 2717–2723.
- Coloma, M. J., R. K. Trinh, A. R. Martinez, and S. L. Morrison. 1999. Position effects of variable region carbohydrate on the affinity and in vivo behavior of an anti-(1→6) dextran antibody. *J. Immunol.* 162: 2162–2170.
- Wallick, S. C., E. A. Kabat, and S. L. Morrison. 1988. Glycosylation of a VH residue of a monoclonal antibody against alpha (1→6) dextran increases its affinity for antigen. *J. Exp. Med.* 168: 1099–1109.
- Tachibana, H., J. Y. Kim, and S. Shirahata. 1997. Building high affinity human antibodies by altering the glycosylation on the light chain variable

- region in N-acetylglucosamine-supplemented hybridoma cultures. *Cyto technology* 23: 151–159.
18. Leibiger, H., D. Wüster, R. D. Stigler, and U. Marx. 1999. Variable domain-linked oligosaccharides of a human monoclonal IgG: structure and influence on antigen binding. *Biochem. J.* 338: 529–538.
 19. van de Bovenkamp, F. S., N. I. L. Derksen, P. Ooijevaar-de Heer, K. A. van Schie, S. Kruithof, M. A. Berkowska, C. E. van der Schoot, H. IJpeert, M. van der Burg, A. Gils, et al. 2018. Adaptive antibody diversification through N-linked glycosylation of the immunoglobulin variable region. *Proc. Natl. Acad. Sci. USA* 115: 1901–1906.
 20. van de Bovenkamp, F. S., N. I. L. Derksen, M. J. van Breemen, S. W. de Taeye, P. Ooijevaar-de Heer, R. W. Sanders, and T. Rispens. 2018. Variable domain N-linked glycans acquired during antigen-specific immune responses can contribute to immunoglobulin g antibody stability. *Front. Immunol.* 9: 740.
 21. Goletz, S. A., D. Czyk, and L. Stoeckl, inventors; Glycotope GMBH, assignee. 2012. Fab-Glycosylated Antibodies. International Application No. PCT/EP2011/063791, Publication No. WO 2012/020065, 2011Oct. 8.
 22. Radcliffe, C. M., J. N. Arnold, D. M. Suter, M. R. Wormald, D. J. Harvey, L. Royle, Y. Mimura, Y. Kimura, R. B. Sim, S. Inogès, et al. 2007. Human follicular lymphoma cells contain oligomannose glycans in the antigen-binding site of the B-cell receptor. *J. Biol. Chem.* 282: 7405–7415.
 23. Youngs, A., S. C. Chang, R. A. Dwek, and I. G. Scragg. 1996. Site-specific glycosylation of human immunoglobulin G is altered in four rheumatoid arthritis patients. *Biochem. J.* 314: 621–630.
 24. Vergoosen, R. D., L. M. Slot, L. Hafkenscheid, M. T. Koning, E. I. H. van der Voort, C. A. Grooff, G. Zervakis, H. Veelken, T. W. J. Huizinga, T. Rispens, et al. 2018. B-cell receptor sequencing of anti-citrullinated protein antibody (ACPA) IgG-expressing B cells indicates a selective advantage for the introduction of N-glycosylation sites during somatic hypermutation. *Ann. Rheum. Dis.* 77: 956–958.
 25. Lloyd, K. A., J. Steen, K. Amara, P. J. Titcombe, L. Israelsson, S. L. Lundström, D. Zhou, R. A. Zubarev, E. Reed, L. Piccoli, et al. 2018. Variable domain N-linked glycosylation and negative surface charge are key features of monoclonal ACPA: implications for B-cell selection. *Eur. J. Immunol.* 48: 1030–1045.
 26. Hamza, N., U. Hershberg, C. G. M. Kallenberg, A. Vissink, F. K. L. Spijkervet, H. Bootsma, F. G. Kroese, and N. A. Bos. 2015. Ig gene analysis reveals altered selective pressures on Ig-producing cells in parotid glands of primary Sjögren's syndrome patients. *J. Immunol.* 194: 514–521.
 27. Visser, A., M. E. Doorenspleet, N. de Vries, F. K. L. Spijkervet, A. Vissink, R. J. Bende, H. Bootsma, F. G. M. Kroese, and N. A. Bos. 2018. Acquisition of N-glycosylation sites in immunoglobulin heavy chain genes during local expansion in parotid salivary glands of primary sjögren patients. *Front. Immunol.* 9: 491.
 28. Visser, A., N. Hamza, F. G. M. Kroese, and N. A. Bos. 2018. Acquiring new N-glycosylation sites in variable regions of immunoglobulin genes by somatic hypermutation is a common feature of autoimmune diseases. *Ann. Rheum. Dis.* 77: e69.
 29. Vander Heiden, J. A., P. Stathopoulos, J. Q. Zhou, L. Chen, T. J. Gilbert, C. R. Bolen, R. J. Barohn, M. M. Dimachkie, E. Cialfoni, T. J. Broering, et al. 2017. Dysregulation of B Cell repertoire formation in myasthenia gravis patients revealed through deep sequencing. *J. Immunol.* 198: 1460–1473.
 30. Levin, M., F. Levander, R. Palmason, L. Greiff, and M. Ohlin. 2017. Antibody-encoding repertoires of bone marrow and peripheral blood—a focus on IgE. *J. Allergy Clin. Immunol.* 139: 1026–1030.
 31. Tipton, C. M., C. F. Fucile, J. Darce, A. Chida, T. Ichikawa, I. Gregoret, S. Schieferl, J. Hom, S. Jenks, R. J. Feldman, et al. 2015. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat. Immunol.* 16: 755–765.
 32. Masella, A. P., A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld. 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13: 31.
 33. Alamyar, E., V. Giudicelli, S. Li, P. Duroux, and M.-P. Lefranc. 2012. IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (Ig) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* DOI: 10.4172/1745-7580.1000056.
 34. Harindranath, N., I. S. Goldfarb, H. Ikematsu, S. E. Burastero, R. L. Wilder, A. L. Notkins, and P. Casali. 1991. Complete sequence of the genes encoding the VH and VL regions of low- and high-affinity monoclonal IgM and IgA1 rheumatoid factors produced by CD5+ B cells from a rheumatoid arthritis patient. *Int. Immunol.* 3: 865–875.
 35. Van Schouwenburg, P. A., S. Kruithof, C. Votsmeier, K. van Schie, M. H. Hart, R. N. de Jong, E. E. van Buren, M. van Ham, L. Aarden, G. Wolbink, et al. 2014. Functional analysis of the anti-adalimumab response using patient-derived monoclonal antibodies. *J. Biol. Chem.* 289: 34482–34488.
 36. Rombouts, Y., A. Willemze, J. J. B. C. van Beers, J. Shi, P. F. Kerkman, L. van Toorn, G. M. C. Janssen, A. Zaldumbide, R. C. Hoeven, G. J. M. Puijn, et al. 2016. Extensive glycosylation of ACPA-IgG variable domains modulates binding to citrullinated antigens in rheumatoid arthritis. *Ann. Rheum. Dis.* 75: 578–585.
 37. Tachibana, H., K. Seki, and H. Murakami. 1993. Identification of hybrid-type carbohydrate chains on the light chain of human monoclonal antibody specific to lung adenocarcinoma. *Biochim. Biophys. Acta* 1182: 257–263.
 38. Ishizaka, A., Y. Sakiyama, M. Nakanishi, K. Tomizawa, E. Oshika, K. Kojima, Y. Taguchi, E. Kandil, and S. Matsumoto. 1990. The inductive effect of interleukin-4 on IgG4 and IgE synthesis in human peripheral blood lymphocytes. *Clin. Exp. Immunol.* 79: 392–396.
 39. van Schie, K. A., P. Ooijevaar-De Heer, S. Kruithof, C. Plasencia, T. Jurado, D. Pascual Salcedo, J. F. Brandse, G. R. A. M. d'Haens, G. Wolbink, and T. Rispens. 2017. Infusion reactions during infliximab treatment are not associated with IgE anti-infliximab antibodies. *Ann. Rheum. Dis.* 76: 1285–1288.
 40. van de Stadt, L. A., P. A. van Schouwenburg, S. Bryde, S. Kruithof, D. van Schaardenburg, D. Hamann, G. Wolbink, and T. Rispens. 2013. Monoclonal anti-citrullinated protein antibodies selected on citrullinated fibrinogen have distinct targets with different cross-reactivity patterns. *Rheumatology* 52: 631–635.
 41. Della Valle, L., S. E. Dohmen, O. J. Verhagen, M. A. Berkowska, G. Vidarsson, and C. van der Schoot. 2014. The majority of human memory B cells recognizing RHD and tetanus residues in IgM⁺ B cells. *J. Immunol.* 193: 1071–1079.
 42. Aalberse, R. C., C. Lupinek, V. Siroux, R. Nadif, J. Just, J. Bousquet, R. Valenta, and T. A. Platts-Mills. 2018. sIgE and sIgG to airborne atopic allergens: coupled rather than inversely related responses. *Allergy* 73: 2239–2242.
 43. Yang, Z., M. J. Robinson, X. Chen, G. A. Smith, J. Taunton, W. Liu, and C. D. C. Allen. 2016. Regulation of B cell fate by chronic activity of the IgE B cell receptor. *Elife* 5: e21238.
 44. Culver, E. L., F. S. van de Bovenkamp, N. I. L. Derksen, J. Koers, T. Cargill, E. Barnes, L. A. Neef, C. A. M. Koelman, R. C. Aalberse, M. Wührer, and T. Rispens. 2018. Unique patterns of glycosylation in immunoglobulin subclass G4-related disease and primary sclerosing cholangitis. *J. Gastroenterol. Hepatol.* DOI: 10.1111/jgh.14512.
 45. Wei, L., R. Chahwan, S. Wang, X. Wang, P. T. Pham, M. F. Goodman, A. Bergman, M. D. Scharff, and T. MacCarthy. 2015. Overlapping hotspots in CDRs are critical sites for V region diversification. *Proc. Natl. Acad. Sci. USA* 112: E728–E737.
 46. Saini, J., and U. Hershberg. 2015. B cell variable genes have evolved their codon usage to focus the targeted patterns of somatic mutation on the complementarity determining regions. *Mol. Immunol.* 65: 157–167.
 47. Sheng, Z., C. A. Schramm, R. Kong, J. C. Mullikin, J. R. Mascola, P. D. Kwong, and L. Shapiro. NISC Comparative Sequencing Program. 2017. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front. Immunol.* 8: 537.
 48. Koelsch, K. A., J. Cavett, K. Smith, J. S. Moore, S. D. Lehoux, N. Jia, T. Mather, S. M. S. Quadri, A. Rasmussen, C. E. Kaufman, et al. 2018. Evidence of alternative modes of B cell activation involving acquired Fab regions of N-glycosylation in antibody-secreting cells infiltrating the labial salivary glands of patients with Sjögren's syndrome. *Arthritis Rheumatol.* 70: 1102–1113.
 49. Sabouri, Z., P. Schofield, K. Horikawa, E. Spierings, D. Kipling, K. L. Randall, D. Langley, B. Roome, R. Vazquez-Lombardi, R. Rouet, et al. 2014. Redemption of autoantibodies on anergic B cells by variable-region glycosylation and mutation away from self-reactivity. *Proc. Natl. Acad. Sci. USA* 111: E2567–E2575.
 50. Kilian, M., J. Reinholdt, H. Lomholt, K. Poulsen, and E. V. Frandsen. 1996. Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence. *APMIS* 104: 321–338.
 51. Xu, J. L., and M. M. Davis. 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 13: 37–45.
 52. Pugh-Bernard, A. E., G. J. Silverman, A. J. Cappione, M. E. Villano, D. H. Ryan, R. A. Insel, and I. Sanz. 2001. Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance. *J. Clin. Invest.* 108: 1061–1070.
 53. Anumula, K. R. 2012. Quantitative glycan profiling of normal human plasma derived immunoglobulin and its fragments Fab and Fc. *J. Immunol. Methods* 382: 167–176.

Supplemental Table I. Overview of total read counts after data processing per donor per antibody isotype

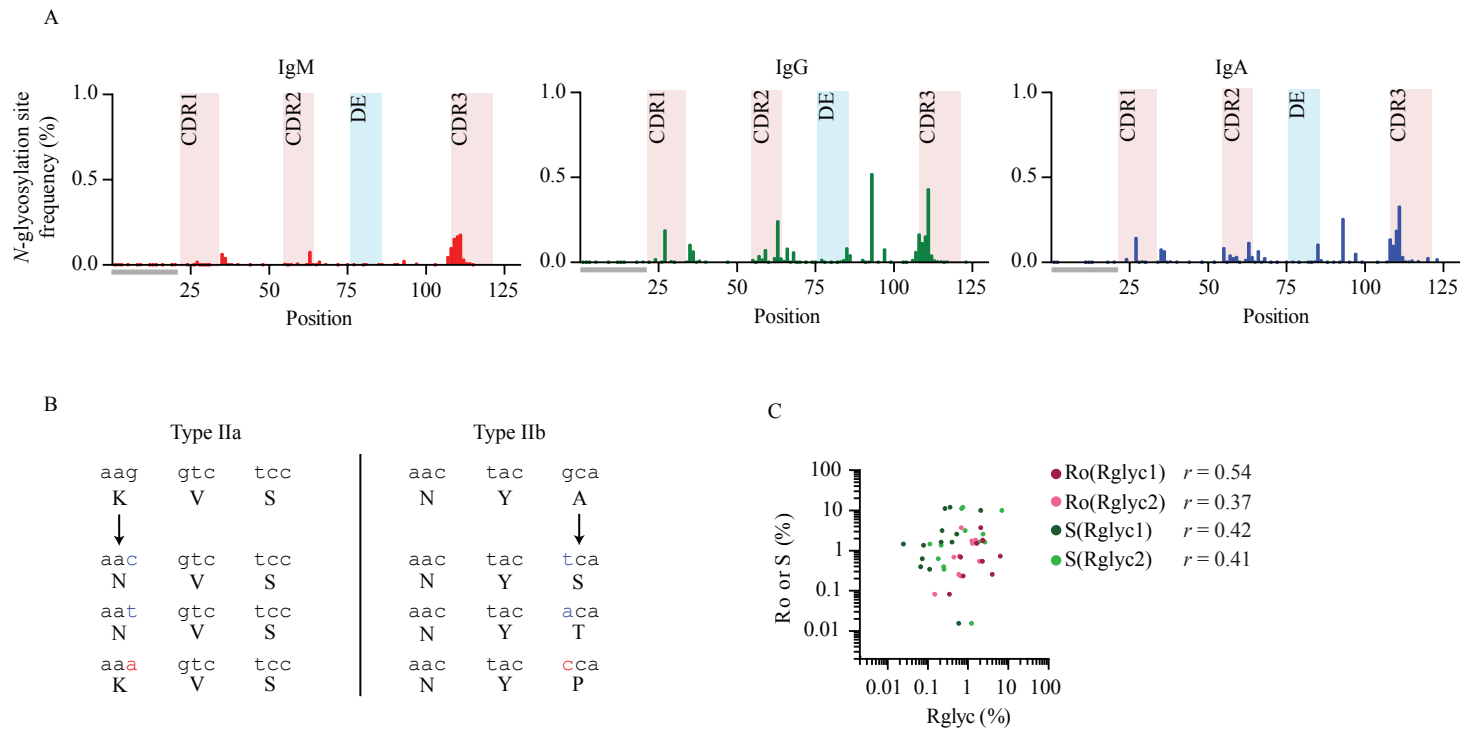
Ref	B cell source	Isotype	Samples per isotype	No. reads
Van der Heiden ²⁹	PB	IgM n	HD07_M ; HD09_M ; HD10_M ; HD13_M	1786 ; 4509 ; 2812 ; 2324
	PB	IgM m	HD07_M ; HD09_M ; HD10_M ; HD13_M	10226; 8511 ; 6845 ; 7759
	PB	IgG	HD07_M ; HD09_M ; HD10_M ; HD13_M	4503 ; 3978 ; 5191 ; 9570
	PB	IgA	HD07_M ; HD09_M ; HD10_M ; HD13_M	10914 ; 7740 ; 6189 ; 10613
	PB	Kappa	HD07_M ; HD09_M ; HD10_M ; HD13_M	9023 ; 9366 ; 8876; 10258
	PB	Lambda	HD07_M ; HD09_M ; HD10_M ; HD13_M	8582 ; 6668 ; 6833 ; 6725
	PB	IgG	MK08_M ; MK02_M ; MK03_M ; AR03_M ; AR04_M ; AR02_M ; AR05_M ; MK04_M ;	3340 ; 5140 ; 1379 ; 5366 ; 5010 ; 7722 ; 4287 ; 1512
	PB	IgA	MK08_M ; MK02_M ; MK03_M ; AR03_M ; AR04_M ; AR02_M ; AR05_M ; MK04_M ; MK05_M	6927 ; 9039 ; 4394 ; 8848; 12289 ; 10166 ; 15470 ; 4951 ; 6568
Levin ³⁰	PB	IgM n	Donor 1 – 6 PB sample 1	7069 ; 8011 ; 4693 ; 8911 ; 3406 ; 6492
	PB	IgM n	Donor 1 – 6 PB sample 2	7786 ; 8847 ; 5561 ; 8025 ; 3194 ; 7156
	PB	IgM m	Donor 1 – 6 PB sample 1	8184 ; 7561 ; 11541 ; 7801 ; 14804 ; 9454
	PB	IgM m	Donor 1 – 6 PB sample 2	8062 ; 8748 ; 13114 ; 7045 ; 13979 ; 10800
	PB	IgG	Donor 1 – 6 PB sample 1	12936 ; 11704 ; 13400 ; 14515 ; 13399 ; 11144
	PB	IgG	Donor 1 – 6 PB sample 2	14200 ; 14268 ; 11433 ; 11473 ; 12273 ; 12598
	PB	IgA	Donor 1 – 6 PB sample 1	11836 ; 11111 ; 13272 ; 12409 ; 13533 ; 10667
	PB	IgA	Donor 1 – 6 PB sample 2	12974 ; 10491 ; 12791 ; 11433 ; 13256 ; 16633
	PB	IgE	Donor 1 – 5 PB sample 1	9144 ; 3803 ; 4822 ; 2463 ; 8312
	PB	IgE	Donor 1 – 5 PB sample 2	10296 ; 7436 ; 9218 ; 6914 ; 6456
	BM	IgG	Donor 1 – 6 BM sample 1	16475 ; 15642 ; 16572 ; 15687 ; 15763 ; 17338
	BM	IgG	Donor 1 – 6 BM sample 2	15979 ; 15624 ; 17693 ; 16377 ; 17374 ; 16019
	BM	IgA	Donor 1 – 6 BM sample 1	12128 ; 11732 ; 11315 ; 12516 ; 12145 ; 11256
	BM	IgA	Donor 1 – 6 BM sample 2	12441 ; 12569 ; 11427 ; 12867 ; 12586 ; 13631
	BM	IgE	Donor 1 – 6 BM sample 1	13241 ; 12477 ; 11765 ; 9620 ; 12638 ; 10802
	BM	IgE	Donor 1 – 6 BM sample 2	13114 ; 13609 ; 12454 ; 8462 ; 11743 ; 10835
Tipton ³¹	PB	IgG	SLE1-126 ; SLE2-178 ; SLE3-292 ; SLE4-2037; SLE5-262 ; SLE7-730 ; SLE6-161	13190 ; 12187 ; 14998 ; 14904 ; 15388 ; 14763; 12006
	PB	IgA	SLE1-126 ; SLE2-178 ; SLE3-292 ; SLE4-2037; SLE5-262 ; SLE7-730 ; SLE6-161	13939 ; 11689 ; 12702 ; 14379 ; 13133 ; 12723; 12715
	BM	IgA	D1 ; D2	9213 ; 11663
	BM	IgG	D1 ; D2	10805 ; 11734

PB, peripheral blood; BM, bone marrow; IgM n, IgM naive; IgM m, IgM memory

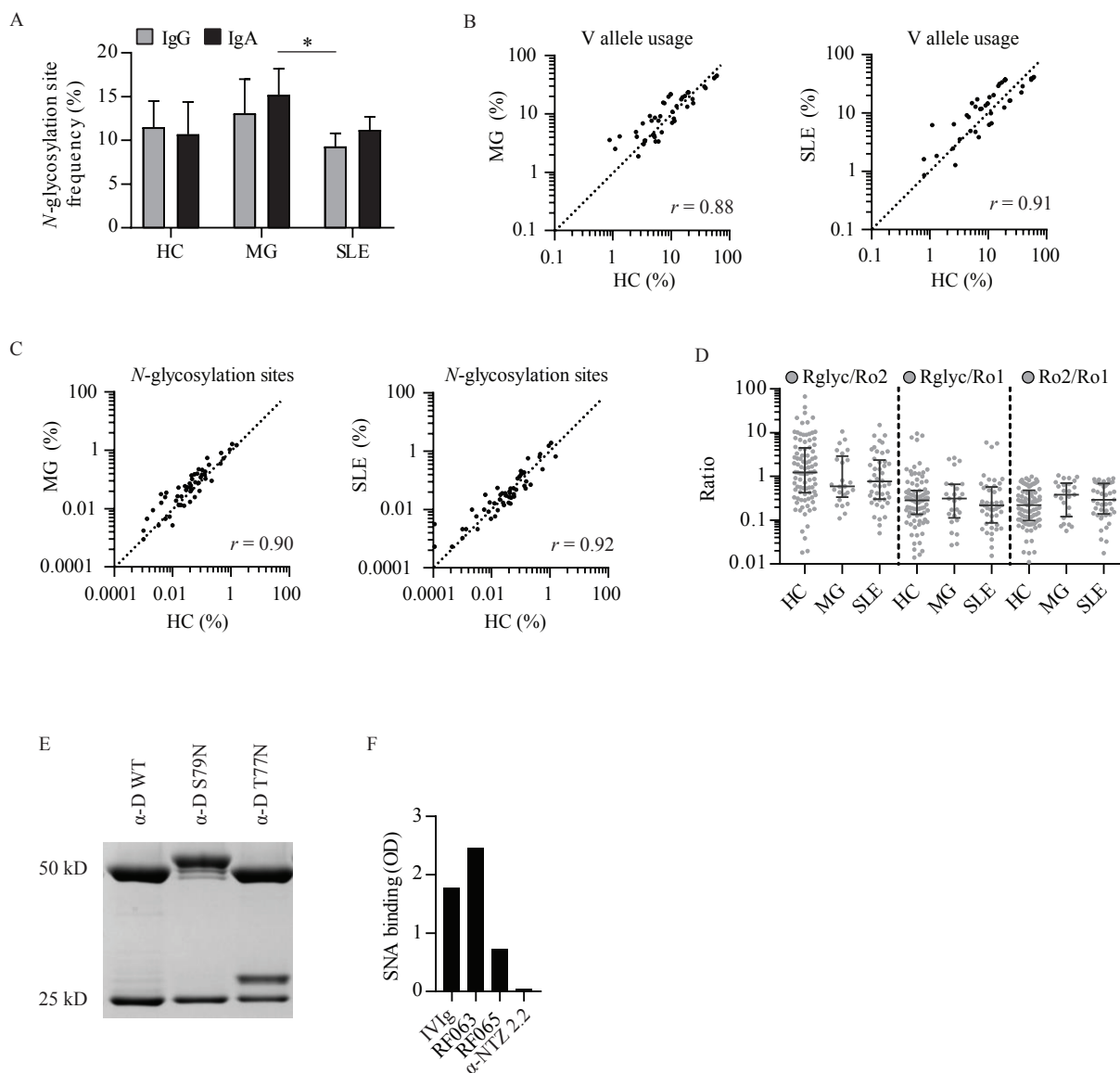
Supplemental Table II. Glycosylation of non-germline *N*-glycosylation sites corresponding to progenitor sites in the variable domain of monoclonal antibodies.

	Position	Allele	Occupied	Clone	<i>N</i> -glycosylation site
Heavy chain	20	IGHV1-69*01	Yes	RA10 ³⁴	NVS
	29	IGHV1-18*01	Yes	α -IFX 1.3 ¹⁹	NFT
	29	IGHV1-2*02	Yes	α -cFib 1.2 ³⁴	NVT
	58	IGHV3-21*02	Yes	RA11 ³⁴	NRS
	59	IGHV3-9*01	Yes	ADL G63S ¹⁹	NSS
	66	IGHV1-69*01	Yes	α -IFX 2.1 ¹⁹	NYT
	66	IGHV1-18*01	Yes	α -ADL 1.3 ¹⁹	NYT
	68	IGHV4-39*01	Partial	RF065	NYT
	75	IGHV3	Yes	CBGA1 ¹⁸	NNS
	77	IGHV1-3*01	Yes	α -ADL 2.1 ¹⁹	NFT
	77	IGHV3-53*03	Yes	RA8 ³⁴	NIS
	77	IGHV3-9*01	Yes	ADL T77N ¹⁹	NIS
	77	IGHV4-39*01	Yes	α -ADL 2.6 ¹⁹	NIS
	77	IGHV1-3*01	Yes	α -ADL 2.10 ¹⁹	NIT
	77	IGHV3-30*01	Yes	α -D T77N	NIS
	82	IGHV3-9*01	Yes	ADL K84T ¹⁹	NAT
	84	IGHV3-30*01	Yes	α -IFX 2.3 ¹⁹	NNT
	84	IGHV3-9*01	Yes	ADL K84N ¹⁹	NNS
	90	IGHV4-38-2*02	Yes	RF063	NLT
	90	IGHV4-34*01	Yes	α -ADL 2.7 ¹⁹	NLT
	97	IGHV1-2*02	Yes	α -cFib 1.2 ³⁴	NDT
Light chain	18	IGLV1-51*01	Yes	α -cFib 1.1 ³⁴	NVT
	22	IGLV3-25*02	Yes	RA8 ³⁴	NCS
	22	IGKV4-1*01	Yes	RA14 ³⁴	NCT
	25	IGLV1-51*01	Yes	mAb-C5 ³⁵	NSS
	26	IGKV3-20*01	Partial	α -ADL 2.10 ¹⁹	NRS
	26	IGKV2-28*01	Yes	RA15 ³⁴	NQS
	28	IGKV1-33*01	Partial	α -ADL 2.9 ¹⁹	NIS
	28	IGKV1-33*01	No	α -NTZ 2.2	NIS
	37	IGKV1-27*01	Yes	ADL L39S ¹⁹	NYS
	77	IGKV1-33*01	Partial	α -ADL 2.9 ¹⁹	NGS
	77	IGKV1-33*01	No	α -NTZ 2.2	NGS
	79	IGKV1-33*01	Yes	α -ADL 2.2 ¹⁹	NGS
	79	IGKV1-27*01	Yes	ADL S79N ¹⁹	NGS
	79	IGKV1-39*01	Partial	α -D S79N	NGS
	86	IGKV1-27*01	Yes	ADL D86N ¹⁹	NFT
	108	IGKV1-27*01	No	ADL A114T ¹⁹	NRT
	110	IGLV2-23*02	Yes	α -IFX 1.4 ¹⁹	NIT

ADL, Adalimumab; IFX, Infliximab; NTZ, Natalizumab; RF, Rheumatoid Factor; GSK, Guselkumab; RA, Rheumatoid arthritis; cFib, citrullinated fibrinogen; H, heavy chain; L, light chain



Supplemental Figure 1. Frequency distribution of non-progenitor sites across antibody variable region for healthy individuals. **(A)** Distributions are shown for memory PB B cells (IgM, IgG and IgA). Bars represent medians from multiple donors ($n=12$). Horizontal grey bars indicates FWR1, for which coverage is incomplete. Sites within the H-CDR3 at positions 111.1 – 112.1 are collapsed into position 111. **(B)** Mutation propensities in Type IIa and IIb progenitor sites. Examples of a Type IIa and IIb germline nonameric progenitor site nucleotide motif **(C)** Scatterplot of the glycan-introducing mutations (Rglyc1, Rglyc2) versus either the silent (S) or replacement (Ro) mutation. Correlations were observed both between Rglyc mutations and silent (S) mutations ($r_s = 0.75$, $p = 0.003$) and non-silent (Ro) mutations ($r_s = 0.56$, $p = 0.007$). The propensity for Rglyc mutations over S mutations was 4-fold lower (median = 0.22, $p < 0.0001$), and slightly higher for Ro mutations (median Rg/Ro = 1.29, $p < 0.04$). Spearman's r , $P < 0.0001$ and Wilcoxon signed rank test.



Supplemental Figure 2. Frequency distribution of non-germline *N*-glycosylation sites across antibody variable region sequences for MG and SLE patients. **(A)** The frequency of *N*-glycosylation sites (%) for IgG and IgA PB B cells of healthy individuals (HC; n=12) and disease subjects (MG; n=9, SLE; n=7). Bars indicate medians, error bars $\pm$ S.D. **(B,C)** Correlation of progenitor sites and acquired glycosylation sites between HC and disease subjects for PB derived memory B cells. **(D)** Ratios of Rglyc versus Ro mutations compared for HC and disease subjects. **(E)** Glycosylation assessment of five monoclonal antibodies with *N*-glycosylation sites. Gel-electrophoresis of α -D monoclonal antibodies. **(F)** Lectin ELISA for IVIg, two rheumatoid factor (RF) monoclonal antibodies RF063 and RF065 and α -natalizumab (NTZ) 2.2.