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

Rochelle D. Vergroesen[*,1], Linda M. Slot[*,1], Lise Hafkenscheid[*,1], Marvyn T. Koning[*,2], Ellen I.H. van der Voort[1], Christine A. Grooff[1], George Zervakis[2], Hendrik Veelken[2], Tom W.J. Huizinga[1], Theo Rispens[3], Hans U. Scherer[1], René E.M. Toes[1]

* equal contribution

[1] Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

[2] Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands

[3] Sanquin Research and Landsteiner Laboratory, Academic Medical Center, Amsterdam, The Netherlands

Note:

The following is a postprint version of the manuscript. The official publisher version of the article can be found at https://doi.org/10.1136/annrheumdis-2017-212052

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

Rochelle D. Vergroesen[*,1], Linda M. Slot[*,1], Lise Hafkenscheid[*,1], Marvyn T. Koning[*,2], Ellen I.H. van der Voort[1], Christine A. Grooff[1], George Zervakis[2], Hendrik Veelken[2], Tom W.J. Huizinga[1], Theo Rispens[3], Hans U. Scherer[1], René E.M. Toes[1]
* equal contribution

[1] Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[2] Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands
[3] Sanquin Research and Landsteiner Laboratory, Academic Medical Center, Amsterdam, The Netherlands

Corresponding author: Dr. Hans Ulrich Scherer, Department of Rheumatology, Leiden University Medical Center, P.O. Box 9600, Leiden 2300 RC, The Netherlands; h.u.scherer@lumc.nl, +31-715261832

The majority of patients with rheumatoid arthritis (RA) harbours immunoglobulin G (IgG) antibodies targeting citrullinated protein antigens (ACPA). Recently, we showed that >90% of ACPA-IgG in serum are glycosylated in the variable domain.[1] *N*-linked glycosylation requires a consensus sequence in the protein backbone (N-X-S/T, where X is not proline), which is scarce in germline-encoded Ig variable region genes.[2, 3] Accordingly, hyperglycosylation of ACPA-IgG requires either clonal expansion of B cells expressing B cell receptors (BCR) containing germline-encoded *N*-glycosylation sites or generation of *de-novo* sites through somatic hypermutation (SHM).[4]

Here, we analysed the BCR repertoire of ACPA-expressing B cells to understand the molecular basis of this remarkable glycosylation. ACPA-expressing B cells were sorted as pools (10 cells per pool) from PBMC of 8 ACPA-positive RA patients.[5] ARTISAN PCR-based BCR-sequencing[6] followed by full-length variable region IgG transcript analysis revealed high nucleotide mutation rates in 97 unique ACPA-IgG heavy chains (HC; mean±SD: 52.86±16.73; figure 1A). 81% of these contained one or more *N*-glycosylation sites.

To replicate these findings and to acquire additional information on paired heavy and light chains (LC), Ig transcripts of 87 single cell-sorted ACPA-IgG clones (6 donors) were analysed, again revealing high nucleotide mutation rates in the HC variable region (mean±SD: 48.55±16.05; figure 1B). Significantly lower mutation rates were observed for 31 single cell-sorted tetanus toxoid (TT)-specific clones (mean±SD: 25.15±18.92; figure 1B). TT-specific clones contained no *N*-glycosylation sites, in contrast to 79% of HC and 88% of paired HC/LC sequences from the ACPA-IgG clones. Additionally, both pool- and single-sorted cell sequence analyses revealed similar high nucleotide mutation rates for ACPA-LC (mean±SD: 36.18±15.09 and mean±SD: 34.51±16.79, respectively; not shown). Furthermore, 59% of ACPA-LC contained one or more *N*-glycosylation sites compared to 4-5% of healthy control LC. Further analyses of HC revealed that all sites in pool/single cell-sorted ACPA-IgG clones were introduced by SHM; furthermore, the degree of SHM did not correlate with the frequency of sites (figures 1C and 1D). Moreover, no accumulation of N-P-S/T sites (chosen as reference due to its similarity to N-X-S/T) was observed in ACPA-IgG, in contrast to the *N*-glycosylation tripeptide N-X-S/T. In fact, no N-P-S/T sequences were identified by either ACPA-IgG sequencing approach (pool/single cell-sorted). Finally, we observed a relative increase of sites in the complementarity-

determining region (CDR) 1 and a relative absence in CDR3 compared to healthy controls (figure 2A). Together, these findings indicate that the remarkable frequency of *N*-glycosylation sites is not the result of random accumulation of mutations but of a selective process during maturation of ACPA-expressing B cells. Intriguingly, modelling of the spatial positioning of the sites revealed that most sites are located on the exterior of the antibody molecule (figures 2B-D).

In conclusion, we provide the first in-depth analysis of the presence of *N*-glycosylation sites in the variable region of ACPA-IgG. The distribution pattern of sites across the ACPA-IgG variable domain and the spatial localization of *N*-glycosylation sites on the exterior of the molecule suggest that their function in selection processes is not primarily related to antigen recognition. Our data favour the concept that introduction of *N*-glycosylation sites generates selective advantages which allow ACPA-expressing B cells to escape from classical selection mechanisms in germinal centers. This is in contrast to the selection of B cells against recall antigens, which is primarily driven by affinity for cognate antigens.[7] In fact, the overall low-avidity of secreted polyclonal ACPA-IgG is in line with this hypothesis.[8] Possibly, ACPA-IgG variable domain glycans interact with glycan receptors in the vicinity of the BCR. These glycans are highly sialylated, suggesting siglecs as potential receptors.[2] Thus, these findings and considerations have important implications for understanding citrulline-specific immunity in RA.

Acknowledgements and affiliations

We thank Jan Wouter Drijfhout (LUMC, Leiden, The Netherlands) for providing the CCP2 and arginine control peptides. We thank the Flow cytometry Core Facility of the LUMC for technical support and the Leiden Genome Technology Center (LGTC) for their expertise and processing of sequencing samples. This study was supported by grants from the Dutch Arthritis Foundation (project no. 15-2-402 (to HUS) and no. 12-2-403 (to REMT)), The Netherlands Organisation for Health Research and Development (ZonMW TOP project no. 91214031 (to REMT), ZonMW clinical fellowship project no. 90714509 (to HUS), ZonMW VENI grant no. 91617107 (to HUS), ZonMW Enabling Technology Hotel grant no. 435002030 (to HUS)), the IMI-funded consortium project BeTheCure (contract no. 115142-2), and the Professor Steenhuis Foundation (to MTK).

References

1. Hafkenscheid L, Bondt A, Scherer HU, et al. Structural Analysis of Variable Domain Glycosylation of Anti-Citrullinated Protein Antibodies in Rheumatoid Arthritis Reveals the Presence of Highly Sialylated Glycans. Mol Cell Proteomics. 2017 Feb; 16(2):278-287.
2. van de Bovenkamp FS, Hafkenscheid L, Rispens T, et al. The Emerging Importance of IgG Fab Glycosylation in Immunity. J Immunol. 2016 Feb 15; 196(4):1435-1441.
3. Shakin-Eshleman SH, Spitalnik SL, Kasturi L. The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. J Biol Chem. 1996 Mar 15; 271(11):6363-6366.
4. Dunn-Walters D, Boursier L, Spencer J. Effect of somatic hypermutation on potential N-glycosylation sites in human immunoglobulin heavy chain variable regions. Mol Immunol. 2000 Feb-Mar; 37(3-4):107-113.
5. Kerkman PF, Fabre E, van der Voort EI, et al. Identification and characterisation of citrullinated antigen-specific B cells in peripheral blood of patients with rheumatoid arthritis. Ann Rheum Dis. 2016 Jun; 75(6):1170-1176.
6. Koning MT, Kielbasa SM, Boersma V, et al. ARTISAN PCR: rapid identification of full-length immunoglobulin rearrangements without primer binding bias. Br J Haematol. 2016 Jun 15.
7. Goodnow CC, Vinuesa CG, Randall KL, et al. Control systems and decision making for antibody production. Nat Immunol. 2010 Aug; 11(8):681-688.
8. Suwannalai P, Scherer HU, van der Woude D, et al. Anti-citrullinated protein antibodies have a low avidity compared with antibodies against recall antigens. Ann Rheum Dis. 2011 Feb; 70(2):373-379.

Figure legends

Figure 1. A high degree of somatic hypermutation in ACPA-IgG clones which does not correlate with the frequency of N-glycosylation sites. Pool and single cells were sorted as described.[5] All independent clones are defined as having identical V, D, J genes and CDR3 regions.

(A) Immunoglobulin heavy variable region (IGHV) mutations in 97 ACPA-IgG clones obtained with pool-sequencing (n=8 donors). (B) IGHV mutations in 87 ACPA-IgG clones that were positive in CCP2-ELISA compared to 31 TT-IgG clones obtained from sequencing of cultured single cells (n=8 and n=3 donors, respectively). P-value was calculated using Mann-Whitney U test for unpaired data (**** $p < 0.0001$). (C) Correlation of the number of IGHV mutations with the number of N-glycosylation sites of 97 ACPA-IgG clones. Non-parametric Spearman correlation, $r = 0.10$, $p = 0.32$. (D) Correlation of the number of IGHV mutations with the number of *N*-glycosylation sites of 87 ACPA-IgG single cell-derived clones. Non-parametric Spearman correlation, $r = 0.19$, $p = 0.071$. All *N*-glycosylation sites were introduced by SHM. No association between IGHV-gene usage and number of N-glycosylation sites was observed using either method.

Figure 2. Distribution and spatial localisation of *N*-glycosylation sites in ACPA-IgG clones. (A) Percentage of *N*-glycosylation sites located in framework (FR) 1, CDR1, FR2, CDR2, FR3 and CDR3 regions of IGHV. Distribution of 102 sites in 97 ACPA-IgG clones obtained with pool-sequencing (left panel, black), 87 sites in 87 ACPA-IgG clones obtained with sequencing of cultured single cells (right panel, gray), both compared to 660 sites in 6724 IGHV sequences from 12 healthy donors (V-region matched, red). (B) Structural model of the top view on the antigen binding pocket of ACPA-IgG clones. (C) Front view of ACPA-IgG heavy chain structures containing *N*-glycosylation sites (asparagine residues colored in green). (D) Front view of ACPA-IgG light chain structures containing N-glycosylation sites (asparagine residues colored in red). All models contain variable regions of 58 ACPA-IgG clones with paired heavy and light chain sequences had a confidence score of 100% with a sequence identity of 47.7%±7.96 and protein coverage of 98.55%±0.63.

FIGURE 1

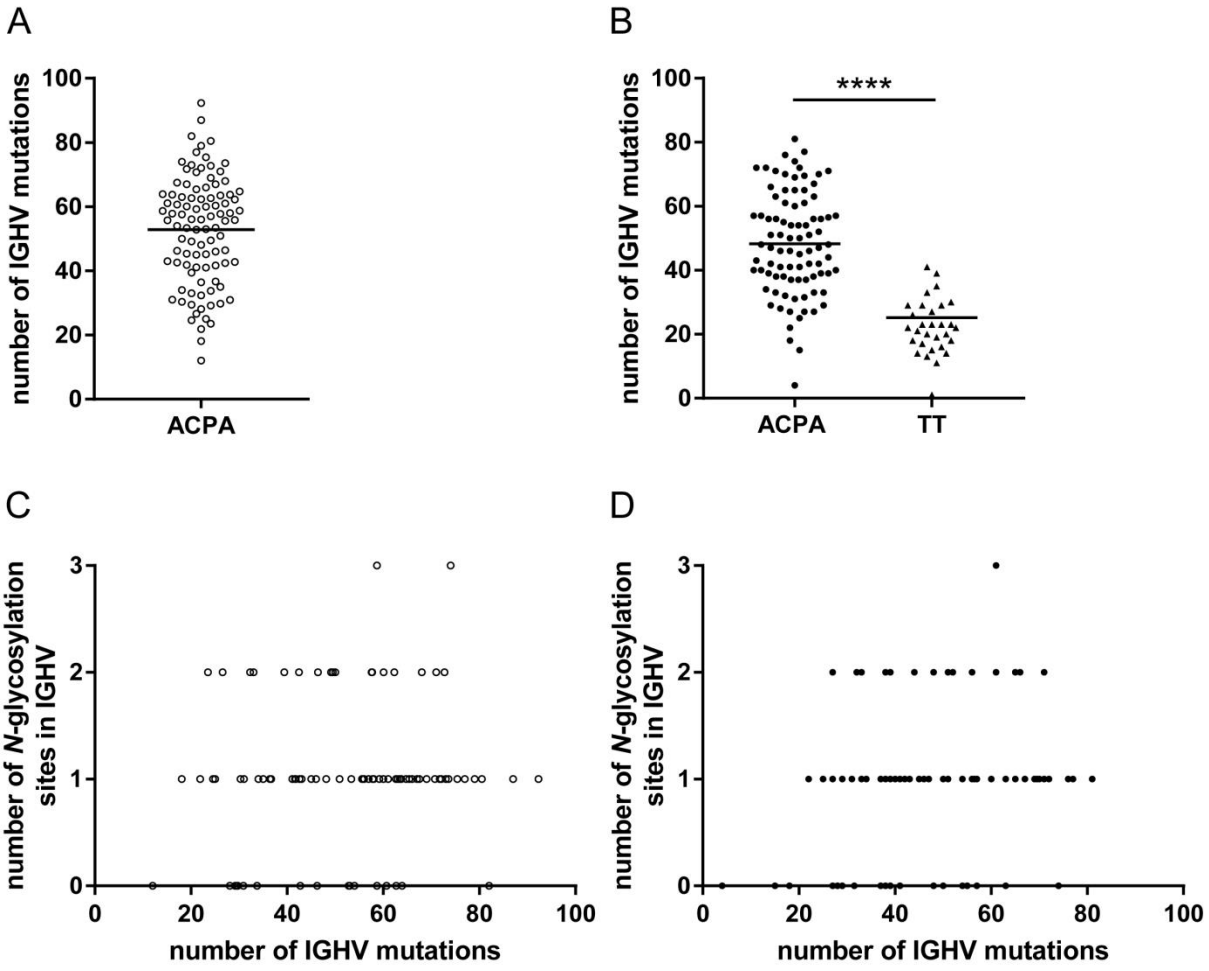
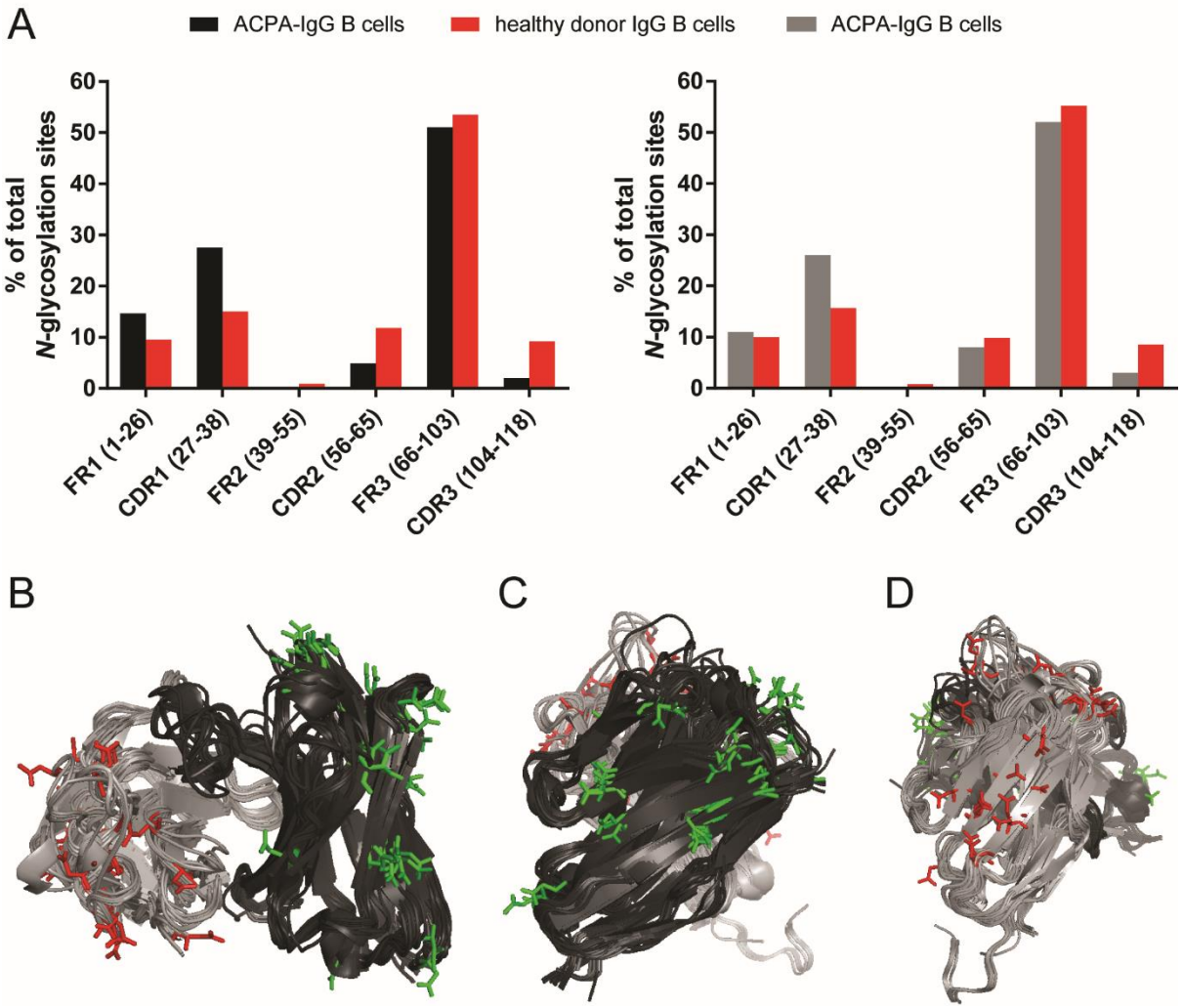

A
number of IGHV mutations
100
80
60
40
20
0
ACPA
B
number of IGHV mutations
****
ACPA
TT
C
number of N-glycosylation sites in IGHV
3
2
1
0
0
20
40
60
80
100
number of IGHV mutations
D
number of N-glycosylation sites in IGHV
number of IGHV mutations

FIGURE 2


A
ACPA-IgG B cells
healthy donor IgG B cells
ACPA-IgG B cells
% of total N-glycosylation sites
60
50
40
30
20
10
0
FR1 (1-26)
CDR1 (27-38)
FR2 (39-55)
CDR2 (56-65)
FR3 (66-103)
CDR3 (104-118)
B
C
D