Bitter Sweet Symphony: the impact of sugars on autoimmunity

Kissel, T.

LETTER

Genetic predisposition (HLA-SE) is associated with ACPA IgG variable domain glycosylation in the pre-disease phase of RA

Theresa Kissel[†], Tineke J. Wesemael [†], Anders Lundquist, Heidi Kokkonen, Atsushi Kawakami, Mami Tamai, Dirkjan van Schaardenburg, Manfred Wuhrer, Tom W.J. Huizinga, Hans U. Scherer, Diane van der Woude, Solbritt Rantapää-Dahlqvist, René E.M. Toes

[†] These authors contributed equally to this work as co-first authors.

Ann Rheum Dis. 2022 Jan;81(1):141-143.

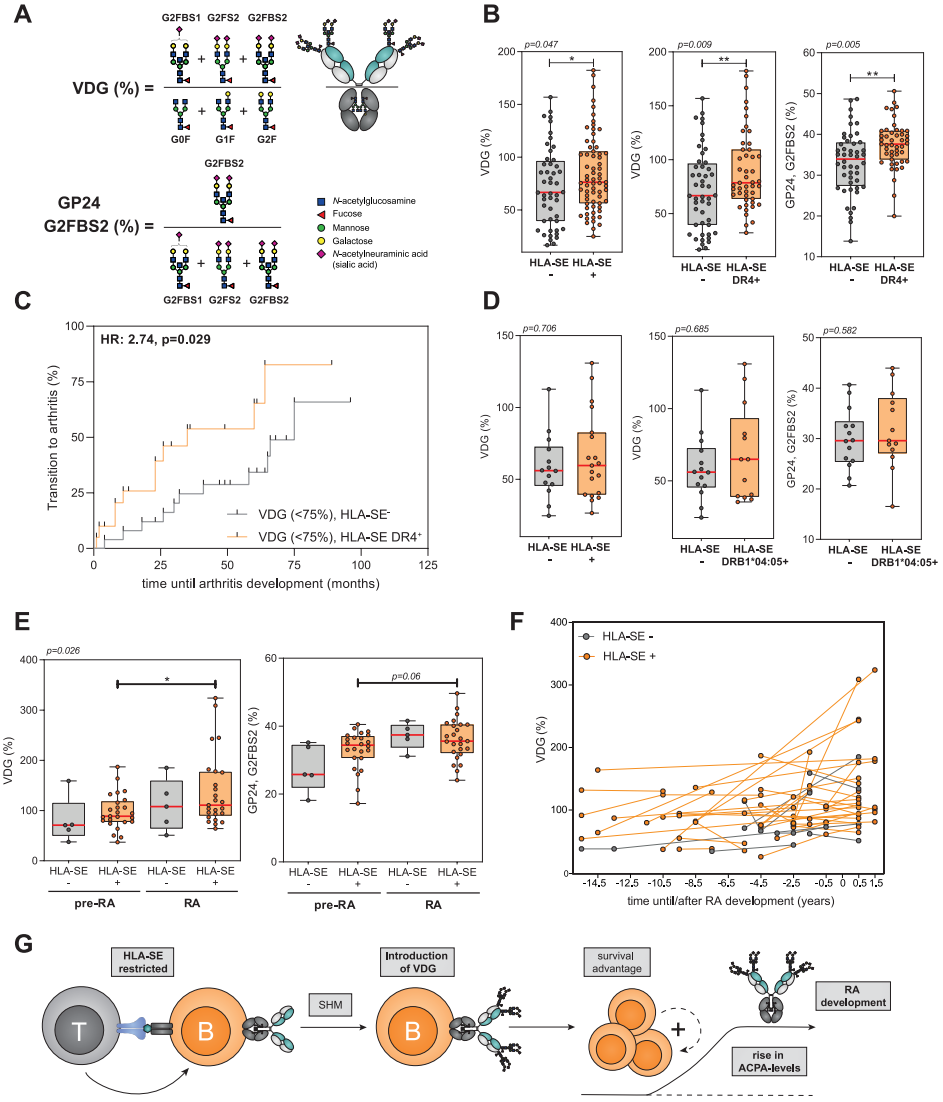
In addition to Fc glycans, Immunoglobulin G (IgG) can carry N-linked glycans in the variable domain. The abundant presence of disialylated variable domain glycans (VDGs) is a special feature of ACPA IgG and possibly other autoantibodies. The introduction of glycosylation sites is mediated by somatic hypermutation (SHM), a T-cell dependent process[1]. The high frequency of glycosylation sites does not correlate with the number of SHM, pointing towards a selective advantage of B cells expressing variable domain glycosylated ACPAs[2]. Previously, we observed that ACPA IgG VDGs are already present in the phase preceding the onset of RA and predictive for disease development[3]. In addition, we provided first evidence that the human leukocyte antigen (HLA) shared epitope (SE) alleles, the most prominent genetic risk factor for ACPA-positive RA, are associated with the presence of VDGs on ACPA IgG pre-disease[4]. Hence, variable domain glycosylation could possibly explain the contribution of HLA-SE restricted T cells in the maturation of the ACPA response. Building upon these results, we now hypothesized that HLA-SE alleles may not be associated with ACPA-positivity as such, but with the specific presence of variable domain glycosylated ACPA IgG, a favorable factor for the development of this multifactorial disease.

To substantiate our hypothesis, we expanded the set of pre-symptomatic individuals (n = 228) and RA patients (n = 126) from Sweden and analyzed two additional cohorts comprising ACPA-positive Dutch individuals with arthralgia (n = 239) and ACPA-positive healthy Japanese individuals (n = 58) (Table S1). We determined the presence/percentage of ACPA IgG VDGs using liquid chromatography[5] and assessed associations with HLA-SE (supplementary materials and methods). In particular, we focused on the most prominent glycan peak (GP24) found on top of the variable domain[1], which carries a bisecting *N*-acetylglucosamine and two terminal sialic acids (G2FBS2) (Figure 1A). ACPA IgG VDGs were, with a median of 58%, already abundantly expressed in healthy individuals (Table S1), in contrast to conventional IgG molecules that yield 12% of VDGs[6]. VDG ($p = 0.047$) and GP24 ($p = 0.003$) were significantly higher in HLA-SE$^+$ Dutch individuals with arthralgia compared to the HLA-SE-negative group (Figure 1B, Table S2 and S3). HLA-SE DR4$^+$ (HLA-DRB1*04:01, *04:04, *04:05, *04:08 and *04:10 alleles) individuals showed the strongest increase in VDGs ($p = 0.009$) and GP24 ($p = 0.005$) compared to HLA-SE-negative individuals (Figure 1B). Even though, we observed a strong correlation between VDGs and ACPA levels (Figure S1), we could not identify an association between ACPA levels and HLA-SE ($p = 0.66$) (Table S4). Moreover, in line with our hypothesis, the association between HLA-SE and GP24 remained significant after correcting for ACPA levels in a multivariable analysis (HLA-SE: $p = 0.03$; HLA-SE DR4$^+$: $p = 0.07$) (Table S3), indicating that HLA-SE primarily associates with abundantly variable domain glycosylated ACPA IgG.

Interestingly, individuals with an "incomplete" VDG (lower than the median of 75%) (Table S1) were more prone to transition to RA, if they were HLA-SE DR4$^+$ (HR: 2.74, $p = 0.029$) (Figure 1C). Conceivably, HLA-SE restricted T cells increase SHM and hence the formation of glycosylation sites, impacting on a subsequent increased risk to develop disease[3]. Likewise, although underpowered and statistically not significant, VDGs and GP24 were numerically increased in the healthy ACPA-positive

individuals from Japan, mainly in the HLA-SE DRB1*04:05$^+$ group, the predominant HLA-SE alleles in this population (Figure 1D). The association between HLA-SE and increased VDG percentages was not present in the Swedish data-set, possibly because all individuals transition to RA (Table S5). However, the findings replicated our previous data, as HLA-SE alleles associated with the presence of ACPA IgG VDGs in the pre-RA phase, after correcting for ACPA-positivity (OR: 1.998, p = 0.040) (Table S7). No association was found between HLA-SE and ACPA in a reciprocal analysis (i.e. after correcting for the presence of VDGs) (OR: 0.620, p = 0.254) (Table S8). Similar to our preceding analyses, this correlation was only found pre-disease as we could not identify a link between HLA-SE and VDGs in established disease (OR: 0.305, p = 0.269) (Table S9), likely because most ACPA IgG carry an abundant amount of VDGs by then (Table S1). Thus, in the phase preceding RA, HLA-SE alleles are associated with ACPAs harboring elevated amounts of VDGs. Additionally, HLA-SE$^+$ individuals showed a significant increase in VDGs towards disease-onset (matched paired analysis, Figures 1, E and F).

Figure 1. Percentage of ACPA IgG variable domain glycosylation (VDG) and glycan peak 24 (GP24, G2FBS2) in HLA-SE$^-$ and HLA-SE$^+$ individuals. (**A**) Formulas to calculate the percentage of ACPA IgG VDG and the most common complex-type disialylated glycan peak found on top of the variable domain, GP24. ACPA IgG were captured, glycans released using PNGase F, 2-AA labelled, HILIC SPE purified and analyzed using UHPLC. The formulas presented are based on the abundance of the liquid chromatography determined Fc-glycan traits G0F, G1F, G2F and VD-glycan traits G2FBS1, G2FS2 and G2FBS2. The respective glycans and their locations on the antibody molecule are schematically illustrated. Agalactosylated (G0), monogalactosylated (G1), digalactosylated (G2), fucose attached to the core GlcNAc (F), bisecting GlcNAc (B), monosialylated (S1), disialylated (S2). Blue square: *N*-acetylglucosamine (GlcNAc), green circle: mannose, yellow circle: galactose, red triangle: fucose, pink diamond: *N*-acetylneuraminic acid. (**B**) ACPA IgG$^+$ individuals with arthralgia from The Netherlands (Amsterdam). Increased ACPA IgG VDG of HLA-SE$^+$ (n = 67) compared to HLA-SE$^-$ (n = 48) individuals. Significantly higher ACPA IgG VDG and GP24 in HLA-SE DR4$^+$ (n = 47) individuals. (**C**) ACPA IgG$^+$ individuals with arthralgia from The Netherlands (Amsterdam) with a VDG < 75% (n = 49). ACPA IgG$^+$ arthralgia individuals with a VDG lower than 75% are more prone to transition to disease and transition earlier, if they are HLA-SE DR4$^+$ (HR: 2.74 (95% CI: 1.07 to 7.00), p-value: 0.029). (**D**) ACPA IgG$^+$ symptom-free healthy individuals from Japan (Nagasaki). Statistically not significant trend for an increased percentage of ACPA IgG VDG and GP24 in HLA-SE$^+$ (n = 19) particularly HLA-SE DRB1*04:05 (n = 13) healthy individuals compared to the HLA-SE$^-$ (n = 14) group. (**E**), (**F**). Matched pairs of samples from pre-symptomatic individuals and RA patients from Sweden (Umeå) (n = 59). HLA-SE$^+$ individuals show a significant increase in their percentage ACPA IgG VDG and GP24 towards disease-onset. HLA-SE$^+$ pre-symptomatic individuals (n = 24) show already high VDG levels up to 15 years before RA-onset. (**G**). Graphical illustration of concluding hypothesis. HLA-SE restricted T cells give help to ACPA IgG-expressing B cells, which results in SHM and the introduction of N-linked glycan sites, and consequently VDG (associations between HLA-SE and VDG). These ACPA IgG VDG$^+$ B cells expand leading to a rise in ACPA levels and ultimately towards disease development. Mann Whitney U tests or linear regression analysis were performed between non-paired and Wilcoxon signed rank test between matched paired samples. The comparison of the survival curves was performed using a Mantel-Cox test. Significant differences are denoted by * or ** and the respective p-values are represented. ►



Hence, the data presented support a concept in which HLA-SE restricted T cells stimulate the introduction of glycosylation sites in ACPA-expressing B cells, an event taking place before the development of ACPA-positive disease. HLA-SE can thus be considered as an "accelerating factor" causing the abundant expression of VDGs on ACPA IgG (Figure 1G). Our data also provide an explanation for why HLA-SE do not associate with ACPAs in healthy individuals[7,8] as these are not yet abundantly glycosylated in their variable domains. The fact that all ACPA IgG are heavily glycosylated in established RA, explains why HLA-SE associate with ACPAs in this disease-stage and emphasizes the possibility that VDGs serve as an important "hit" involved in the unrestrained expansion of the RA-specific autoreactive B-cell response.

Acknowledgments

The authors would like to thank the Department of Biobank Research at Umeå University, Västerbotten Intervention Programme, the Northern Sweden MONICA study and the County Council of Västerbotten for providing data and samples. We would like to thank Dr. Jan Wouter Drijfhout (LUMC, Leiden, The Netherlands) for providing the CCP2 peptide.

Funding

This work has been financially supported by ReumaNederland 17-1-402 and 08-1-34 (to R.E.M.T.), the IMI-funded project RTCure 777357 (to T.W.J.H.), ZonMw TOP 91214031 (to R.E.M.T.), Target-to-B LSHM18055-SGF (to R.E.M.T.), the European Research Council (ERC) advanced grant AdG2019-884796 (to R.E.M.T.), the Swedish Research Council VR 2017-00650 (to S.R-D.) as well as the King Gustaf V's 80-Year Fund, the King Gustaf V's and Queen Victoria's Fund and the Swedish Rheumatism Association (to S.R-D.).

Conflict of interest

H.U.S., T.W.J.H. and R.E.M.T. are mentioned inventors on a patent on ACPA IgG V-domain glycosylation.

Author contributions

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Conceptualization: T.K., T.J.W., and R.E.M.T. Methodology: T.K. Software: T.K., T.J.W., A.L., H.K. Investigation: T.K., T.J.W., A.L., H.K. Visualization: T.K., T.J.W. Supervision: D.v.W., S.R.-D. and R.E.M.T. Writing—original draft: T.K., T.J.W., D.V.W., S.R.-D. and R.E.M.T. Writing—review and editing: A.L., H.K., A.K., M.T., D.v.S., M.W., T.W.J.H., H.U.S.

References

1 Hafkenscheid, L., Bondt, A., Scherer, H. U., Huizinga, T. W., Wuhrer, M., Toes, R. E., 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. 16: 278-287.

2 Vergroesen, R. D., Slot, L. M., Hafkenscheid, L., Koning, M. T., van der Voort, E. I. H., Grooff, C. A., et al., 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* 2018. 77: 956-958.

3 Hafkenscheid, L., de Moel, E., Smolik, I., Tanner, S., Meng, X., Jansen, B. C., et al., N-Linked Glycans in the Variable Domain of IgG Anti-Citrullinated Protein Antibodies Predict the Development of Rheumatoid Arthritis. *Arthritis Rheumatol* 2019. 71: 1626-1633.

4 Kissel, T., van Schie, K. A., Hafkenscheid, L., Lundquist, A., Kokkonen, H., Wuhrer, M., et al., On the presence of HLA-SE alleles and ACPA-IgG variable domain glycosylation in the phase preceding the development of rheumatoid arthritis. *Ann Rheum Dis* 2019. 78: 1616-1620.

5 Jansen, B. C., Hafkenscheid, L., Bondt, A., Gardner, R. A., Hendel, J. L., Wuhrer, M., et al., HappyTools: A software for high-throughput HPLC data processing and quantitation. *PLoS One* 2018. 13: e0200280.

6 Kasermann, F., Boerema, D. J., Ruegsegger, M., Hofmann, A., Wymann, S., Zuercher, A. W., et al., Analysis and functional consequences of increased Fab-sialylation of intravenous immunoglobulin (IVIG) after lectin fractionation. *PLoS One* 2012. 7: e37243.

7 Hensvold, A. H., Magnusson, P. K., Joshua, V., Hansson, M., Israelsson, L., Ferreira, R., et al., Environmental and genetic factors in the development of anticitrullinated protein antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. *Ann Rheum Dis* 2015. 74: 375-380.

8 Terao, C., Ohmura, K., Ikari, K., Kawaguchi, T., Takahashi, M., Setoh, K., et al., Effects of smoking and shared epitope on the production of anti-citrullinated peptide antibody in a Japanese adult population. *Arthritis Care Res (Hoboken)* 2014. 66: 1818-1827.

9 Bos, W. H., Wolbink, G. J., Boers, M., Tijhuis, G. J., de Vries, N., van der Horst-Bruinsma, I. E., et al., Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. *Ann Rheum Dis* 2010. 69: 490-494.

10 Kawashiri, S. Y., Tsuji, Y., Tamai, M., Nonaka, F., Nobusue, K., Yamanashi, H., et al., Effects of cigarette smoking and human T-cell leukaemia virus type 1 infection on anti-citrullinated peptide antibody production in Japanese community-dwelling adults: the Nagasaki Islands Study. *Scand J Rheumatol* 2020: 1-4.

11 Arnett, F. C., Edworthy, S. M., Bloch, D. A., McShane, D. J., Fries, J. F., Cooper, N. S., et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988. 31: 315-324.

12 Kokkonen, H., Brink, M., Hansson, M., Lassen, E., Mathsson-Alm, L., Holmdahl, R., et al., Associations of antibodies against citrullinated peptides with human leukocyte antigen-shared epitope and smoking prior to the development of rheumatoid arthritis. *Arthritis Res Ther* 2015. 17: 125.

13 Jansen, B. C., Bondt, A., Reiding, K. R., Scherjon, S. A., Vidarsson, G. and Wuhrer, M., MALDI-TOF-MS reveals differential N-linked plasma- and IgG-glycosylation profiles between mothers and their newborns. *Sci Rep* 2016. 6: 34001.

14 Hafkenscheid, L., de Moel, E., Smolik, I., Tanner, S., Meng, X., Jansen, B. C., et al., N-linked glycans in the variable domain of ACPA-IgG predict the development of rheumatoid arthritis. *Arthritis Rheumatol* 2019.

Supplemental Tables

Table S1. Descriptive cohort data

	Cohort 1 Pre-symptomatic individuals, Sweden (Umeå) N = 228
Sex (female), n(%)	145 (64)
Age (years), mean (SD)	52.2 (9.4)
Time until symptom-onset of RA (years), median (Q3 - Q1)	4.7 (5.9)
Time until arthritis development (months), median (Q3 - Q1)	
Arthritis/RA during follow up, n (%)	228 (100)
ACPA-positive, n (%)	168 (73.7)
ACPA levels (AU/mL), median (Q3 - Q1)	126.7 (455.5)
HLA-SE positive, n(%)	165/227 (72.7)
VDG positive, n (%)	105 (46.1)
VDG%, median (Q3 - Q1)	97.4 (53.5)
GP24%, median (Q3 - Q1)	34.3 (9.5)

VDG = variable domain glycosylation

GP24 = glycan peak 24, most abundant disialylated VDG peak

ACPA titer/levels were measured as previously described[4,9,10]

Table S2. Linear regression analysis of individuals with arthralgia from The Netherlands (Amsterdam) with VDG percentages as dependent variable.

Arthralgia	**univariable (simple)**			
	B	**95%CI**		**sign.**
		lower	**upper**	
HLA-SE	13.96	0.20	27.71	0.047
HLA-SE DR4	17.98	4.50	31.45	0.009
Sex (female)	0.80	-14.96	16.57	0.92
ACPA levels	20.66	17.68	23.64	< 0.001
Age (years)	0.32	-0.31	0.96	0.32

ACPA levels are log transformed

Table S3. Linear regression analysis of individuals with arthralgia from The Netherlands (Amsterdam) with GP24 percentages as dependent variable.

Arthralgia	**univariable (simple)**			
	B	**95%CI**		**sign.**
		lower	**upper**	
HLA-SE	3.97	1.37	6.57	0.003
HLA-SE DR4	3.73	1.14	6.32	0.005
Sex (female)	-0.003	-3.05	3.04	0.99
ACPA levels	3.19	2.46	3.92	< 0.001
Age (years)	-0.02	-0.14	0.11	0.76

ACPA levels are log transformed

Cohort 1 (Continued) RA patients, Sweden (Umeå) N = 126	Cohort 2 Individuals with arthralgia, The Netherlands (Amsterdam, Reade) N = 239	Cohort 3 Healthy, symptom-free individuals, Japan (Nagasaki) N = 58
78 (61)	185 (77)	38 (66)
59.7 (9.3)	48.3 (11.6)	66.7 (9.9)
-	-	-
	18.22 (30.7)	-
-	137 (57.3)	9 (15.5)
125 (98.4)	239 (100)	58 (100)
592.9 (725.3)	103 (708.5)	35.8 (79)
94/126 (74)	74/128 (73)	22/39 (56.4)
116 (92.1)	211 (87.9)	48 (83.8)
94.2 (50.8)	75.3 (48.9)	58.1 (35.6)
32.1 (9.2)	35.6 (8.98)	32.1 (9.8)

	multivariable (multiple)				multivariable (multiple)		
B	95%CI		sign.	B	95%CI		sign.
	lower	upper			lower	upper	
3.47	-5.26	12.19	0.43				
				6.24	-2.41	14.88	0.16
20.45	17.41	23.48	< 0.001	20.22	17.19	23.25	< 0.001

	multivariable (multiple)				multivariable (multiple)		
B	95%CI		sign.	B	95%CI		sign.
	lower	upper			lower	upper	
2.41	0.31	4.50	0.03				
				1.96	-0.15	4.06	0.07
3.04	2.31	3.77	< 0.001	3.05	2.31	3.79	< 0.001

Table S4. Linear regression analysis of ACPA-positive individuals with arthralgia (The Netherlands) with log-transformed ACPA levels as dependent variable.

Arthralgia	univariable (simple)				multivariable (multiple)			
	B	95% CI		sign.	B	95% CI		sign.
		lower	upper			lower	upper	
HLA-SE	0.18	0.61	0.97	0.66	0.09	-0.24	0.43	0.58
HLA-SE DR4	0.34	0.46	1.12	0.40				
Sex (female)	-0.11	-1.00	0.77	0.80				
Age (years)	-0.02	-0.06	0.01	0.28				
VDG%	0.030	0.026	0.034	< 0.001	0.030	0.026	0.034	< 0.001
GP24%	0.13	0.10	0.15	< 0.001				

VDG = variable domain glycosylation

GP24 = glycan peak 24, most abundant disialylated VDG peak

Table S5. Linear regression analysis of pre-symptomatic individuals from Sweden (Umeå) with VDG percentages as dependent variable.

Pre-RA	univariable (simple)			
	B	95% CI		sign.
		lower	upper	
HLA-SE	-9.536	-27.771	8.699	0.302
Sex (female)	2.747	-12.609	18.103	0.724
ACPA+/-	28.047	-7.129	63.222	0.117
ACPA levels (AU/mL)	0.019	0.010	0.029	0.000
Age (years)	-0.230	-1.026	0.566	0.568
Time until RA development (years)	0.533	-1.383	2.449	0.582

Table S6. Linear regression analysis of pre-symptomatic individuals from Sweden (Umeå) with ACPA levels as dependent variable.

Pre-RA	univariable (simple)			
	B	95% CI		sign.
		lower	upper	
HLA-SE	102.733	-67.058	272.523	0.234
VDG+/-	529.780	394.936	664.624	0.000
VDG%	6.574	3.222	9.927	0.000

VDG = variable domain glycosylation

multivariable (multiple)				multivariable (multiple)				multivariable (multiple)			
B	95% CI		sign.	B	95% CI		sign.	B	95% CI		sign.
	lower	upper			lower	upper			lower	upper	
0.02	-0.42	0.45	0.93								
				0.04	-0.30	0.38	0.82	0.12	-0.31	0.55	0.57
				0.030	0.026	0.035	< 0.001				
0.12	0.09	0.15	< 0.001					0.12	0.09	0.15	< 0.001

multivariable (multiple)				multivariable (multiple)			
B	95% CI		sign.	B	95% CI		sign.
	lower	upper			lower	upper	
-9.388	-27.498	8.721	0.306	-11.20	-28.286	5.877	0.196
27.859	-7.313	63.031	0.119				
				0.020	0.010	0.030	0.000

multivariable (multiple)				multivariable (multiple)			
B	95% CI		sign.	B	95% CI		sign.
	lower	upper			lower	upper	
148.491	-167.840	464.822	0.354	34.621	-117.981	187.222	0.655
				527.869	391.351	664.387	0.000
6.735	3.363	10.108	0.000				

Table S7. Logistic regression analysis of pre-symptomatic individuals from Sweden (Umeå) with detectable vs. non detectable VDG profiles.

Pre-RA	univariable (simple)			
	OR	95% CI		sign.
		lower	upper	
HLA-SE	1.675	0.920	3.049	0.091
Sex (female)	0.735	0.428	1.263	0.265
ACPA+/-	16.176	6.157	42.500	< 0.001
ACPA levels (AU/mL)	1.003	1.002	1.004	< 0.001
Age (years)	1.001	0.974	1.030	0.918
Time until RA development (years)	0.241	-0.126	0.607	0.196

VDG = variable domain glycosylation

Table S8. Logistic regression analysis of pre-symptomatic individuals from Sweden (Umeå) with detectable vs. non detectable (< 25 AU/mL) ACPA.

Pre-RA	univariable (simple)			
	OR	95% CI		sign.
		lower	upper	
HLA-SE	0.851	0.433	1.671	0.639
Sex (female)	0.735	0.428	1.263	0.265
VDG+/-	16.176	6.157	42.500	< 0.001
VDG%	1.025	0.993	1.059	0.121
Age (years)	1.006	0.974	1.038	0.730
Time until RA development (years)	1.271	1.165	1.387	< 0.001

VDG = variable domain glycosylation

Table S9. Logistic regression analysis of RA patients from Sweden (Umeå) with detectable vs. non detectable VDG profiles.

RA	univariable (simple)			
	OR	95% CI		sign.
		lower	upper	
HLA-SE	0.305	0.037	2.504	0.269
Sex (female)	1.091	0.292	4.082	0.897
ACPA+/-	0.000	0.000		0.999
ACPA levels (AU/mL)	1.003	1.000	1.005	0.027
Age (years)	0.981	0.913	1.054	0.595

	multivariable (multiple)				multivariable (multiple)		
OR	95% CI		sign.	OR	95% CI		sign.
	lower	upper			lower	upper	
1.520	0.758	3.048	0.238	1.998	1.034	3.863	0.040
				17.211	6.500	45.572	< 0.001
1.004	1.002	1.005	< 0.001				

	multivariable (multiple)				multivariable (multiple)		
OR	95% CI		sign.	OR	95% CI		sign.
	lower	upper			lower	upper	
1.027	0.105	10.064	0.981	0.620	0.273	1.410	0.254
				26.842	8.688	82.932	< 0.001
1.025	0.993	1.059	0.125				
1.058	0.855	1.309	0.603	1.35	1.208	1.509	< 0.001

	multivariable (multiple)				multivariable (multiple)		
OR	95% CI		sign.	OR	95% CI		sign.
	lower	upper			lower	upper	
0.331	0.039	2.811	0.311	0.297	0.036	2.446	0.259
				0.000	0.000		0.999
1.003	1.000	1.005	0.029				

Supplemental Figure

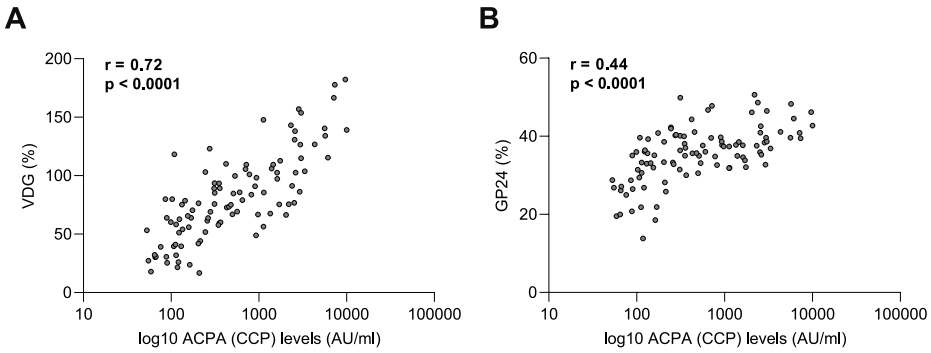


Figure S1. Correlation between ACPA levels (log transformed) and (A) ACPA IgG variable domain glycosylation (VDG) or (B) disialylated VDG (GP24, G2FBS2) in individuals with arthralgia from The Netherlands (Amsterdam). Correlation was assessed with Pearson correlation.

Supplemental Materials and Methods

Patient and public involvement – Patients were involved in this study by donating blood when attending population surveys or health checkups.

Study cohorts – The study presents data from three different cohorts. Cohort 1 includes Swedish individuals diagnosed with RA later in life and sampled at the Medical Biobank of Northern Sweden prior to symptom-onset [median (IQR) pre-dating time: 4.7 (5.9) years; n = 228; 168 ACPA-positive] and after diagnosis of RA (n = 126, 125 ACPA-positive). All RA patients fulfilled the 1987 ARA classification criteria[11]. Cohort 2 presents data from Dutch ACPA-positive individuals with arthralgia with up to 10 years of follow-up for arthritis development sampled at the rheumatology outpatient clinic in the Amsterdam area of The Netherlands. 137 individuals (57.3%) developed arthritis during follow-up. Individuals from cohort 3 were selected as ACPA-positive healthy individuals at risk of RA, which were part of the Nagasaki Island Study performed in Japan. The prospective cohort study was based on resident health checkups. All ACPA-positive individuals were free of RA-specific symptoms (joint complaints) at the time-point of sampling and followed up to three years. 9 individuals (15.5%) developed RA during follow up. Descriptive cohort information is presented in Table S1.

Ethical considerations – All participants have given their written informed consent and the Regional Ethical Review Board Committees approved the studies.

Statistical analyses – Statistical calculations were performed using SPSS for Windows version 27.0 (IBM Corp., NY, USA) or STATA (V.16.1; STATA Corp, College Station, Texas USA). Continuous data were analyzed using non-parametric methods (Mann-Whitney´s U-test and Wilcoxon signed-rank test) and parametric tests when appropriate. All p-values are two-sided and $p < 0.05$ was considered as statistically significant. Univariable (simple) and multivariable (multiple) logistic regression analyses were performed for cohort 1 (Swedish pre-symptomatic individuals) and detectable vs. non detectable VDG profiles or ACPA. Univariate and multivariate linear regression analyses were performed for cohort 1 and 2 (Dutch ACPA-positive individuals with arthralgia) and the percentages of ACPA IgG VDG and GP24 as well as for ACPA levels as dependent variables. Associations with HLA-SE were assessed. The unstandardized coefficient (B) represents the mean change in the response given a one unit change in the predictor. Correlation between ACPA levels (log transformed) and percentages of VDG was assessed with Pearson correlation.

Laboratory analyses – Anti-citrullinated protein antibodies (ACPAs) were analyzed in serum samples using anti-cyclic citrullinated peptide (CCP) enzyme linked immunosorbent assays (ELISAs) and HLA genotyping was performed as previously described[9,10,12]. In short, ACPAs were detected in samples of pre-symptomatic individuals and RA patients from Sweden (cohort 1) using a CCP2

ELISA from Euro-Diagnostica (Malmö, Sweden) with a positivity cut-off set at 25 AU/mL. HLA-DRB1 genotyping was performed using PCR sequence-specific primers from the DR low-resolution and DRB1*04 subtyping kits (Olerup SSP AB, Saltsjöbaden, Sweden). The SE genes were defined as HLA-DRB1*01:01, 04:01, *04:04, *04:05 and *04:08[12]. The CCP-status from the Dutch ACPA-positive individuals with arthralgia (cohort 2) were determined using a second-generation anti-CCP ELISA (Axis Shield, Dundee UK). The cut-off level was set at 5 AU/ml. HLA-DRB1 genotyping was performed by sequenced-based, high resolution typing (Department of Immunogenetics, Sanquin, Amsterdam). The SE alleles were defined as HLA-DRB1*01:01, *01:02, *04:01, *04:04, *04:05, *04:08, *04:10 and *10:01[9]. ACPAs from healthy individuals from Japan (cohort 3) were detected by a CLEIA (STACIA MEBLux test CCP; MBL, Nagoya, Japan). The cut-off level was < 4.5 U/mL. HLA-DRB1 genotyping was performed with the WAKFlow HLA typing kit (Wakunaga Pharmaceutical, Osaka, Japan) based on the reverse sequence-specific oligonucleotide probes method coupled with the xMAP technology designed for the use with the Luminex system (Luminex Japan, Tokyo, Japan)[10].

ACPA IgG capturing and VDG analysis – ACPA IgG capturing, total glycan release, glycan labeling, UHPLC analysis and data processing were performed as previously described[1,4,5,13]. In brief, ACPAs were isolated from 25 μl of serum samples using NeutrAvidin Plus resin (Thermo Fisher Scientific) coupled with 0.1 μg/μl CCP2-biotin. Following, IgG isolation was performed using FcXL affinity beads (Thermo Fisher Scientific). N-linked glycans were released from captured ACPA IgG using PNGase F and subsequently labeled with 2-aminobenzoic acid (2-AA, Sigma Aldrich) and 2-picoline borane (2-PB, Sigma Aldrich). The glycans were purified via hydrophilic interaction liquid chromatography-solid phase extraction (HILIC SPE) using GHP membrane filter plates (Pall Life Science). 2-AA labelled and HILIC SPE purified N-linked glycans were analyzed using a Dionex Ultimate 3000 (Thermo Fisher Scientific) instrument, a FLR fluorescence detector set with excitation and emission wavelengths of 330 and 420 nm and a Acquity UHPLC BEH Glycan column (1.7 μm 2.1 mm × 100 mm; Waters, Milford, MA). Separation was performed as previously published[4]. The generated chromatograms were aligned and separated into 24 glycan peaks (GP) using Chromeleon version 7.1.2.1713 (Thermo Fisher Scientific). The glycan composition in each chromatographic peak was defined as previously described[5]. Calibration was performed using HappyTools version 0.0.2 and the elution time of five calibrant peaks (G0F, G1F, G2F, G2FBS1 and G2FBS2). Integration and baseline correction were performed using the following settings: *general settings*: start t = 8 min, end t = 30 min; *peak detection settings*: minimum intensity: 0.05, sigma value: 2.0; *calibration settings*: minimum peaks: 4, minimum S/N: 1; *quantitation settings*: datapoints: 100, baseline order: 1, background window: 1.0, MT slice points: 5. The amount of N-glycans in each peak was expressed as the total integrated area under the curve (AUC). The cut-off was based on the AUC of all blank and 8 ACPA-negative healthy donor samples, excluding low outliers (below Q_1−1.5 × IQR) and high outliers (above Q_3+1.5 × IQR). The percentage VDG was calculated based on the following formula: [(G2FBS1+G2FS2+G2FBS2)/(G0F+G1F+G2F) × 100][5,14] and the percentage of GP24 as follows: [G2FBS2/(G2FBS1+G2FS2+G2FBS2) × 100].