Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity

Agnieszka Małecka,[1] Gunhild Trøen,[1] Anne Tierens,[2] Ingunn Østlie,[1] Jędrzej Małecki,[3] Ulla Randen,[1] Sigbjørn Berentsen,[4] Geir E. Tjønnfjord,[5] and Jan M.A. Delabie[2]

[1]Department of Pathology, Oslo University Hospital, Norway; [2]Laboratory Medicine Program, University Health Network and University of Toronto, ON, Canada; [3]Department of Biological Sciences, University of Oslo, Norway; [4]Department of Research and Innovation, Haugesund Hospital, Helse Fonna, Haugesund, Norway; and [5]Department of Haematology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Norway

Correspondence: a.m.malecka@medisin.uio.no/ machynia@yahoo.com
doi:10.3324/haematol.2016.146126

Material and methods

Patients

We collected bone marrow and blood from 27 patients with well-documented primary CAD. This series included 8 patients (CAD-23 to CAD-30) previously published by our group [1], for those cases previously extracted DNA and acquired immunoglobulin heavy chain (*IGH)* sequences were used. Clinical data (collected at diagnosis) were available for most of the patients and included hemoglobin, lactate dehydrogenase (LD), bilirubin and IgM levels, leukocyte, lymphocyte, reticulocyte and thrombocyte counts, cold agglutinin (CA) titer, sex and age at diagnosis. Clinical data are given in Table S1 and S3.

The study was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (number 2012/131). Written informed consent in accordance with this approval was obtained.

Bone marrow and blood sample preparation and Fluorescent activated cell sorting (FACS)

Mononuclear cell suspensions were made from all bone marrow and blood samples using Leucosep® tubes (Greiner Bio-One North America, Inc., Monroe, NC, USA) according to the manufacturer's recommendations. Samples were stored in 10% DMSO in liquid nitrogen until FACS analysis. All blood and bone marrow samples contained monoclonal B cells as determined by prior flow cytometry analysis. Cells were FACS-sorted from either bone marrow or from blood depending on whichever sample volume was the largest to sort as many cells as possible. For FACS analysis, the mononuclear cell suspensions were thawed and washed with PBS supplemented with 0.5 % BSA (PAA laboratories GmbH, Pasching, Austria) and stained

for surface antigens with the following antibodies purchased from Beckman Coulter (Brea, CA, USA): anti-CD45 (clone J.33), anti-CD20 (clone B9E9(HRC20)), anti-CD19 (clone J3-119), anti-CD5 (clone L17F12) and anti-λ and anti-κ polyclonal antibodies purchased from Cytognos (Salamanca, Spain). Antibodies were conjugated to either fluorescein isothiocyanate (FITC), phycoerythrine (Pe), peridinin chlorophyll protein-cy5.5 (PerCP-Cy5.5), phycoerythrine cyanine 7 (PeCy7), Pacific Blue or Krome Orange. After staining, the cell suspensions were incubated for 15 minutes in the dark at room temperature and washed with PBS supplemented with 0.5 % BSA followed by filtration through a 70 μm filter to remove cell clumps.

Stained samples were sorted with high-pressure settings using a FACS Aria Ilu High speed sorter (Becton Dickinson, Franklin Lakes, NJ, USA) equipped with 408nm, 488nm and 633nm lasers. Selection of monoclonal B cells for sorting was performed using Becton Dickinson FACSDiva software, starting with the gating of viable cells using the forward scatter versus side scatter dot plot. Subsequently, CD45 bright, low side scatter events (i.e. lymphocytes) were selected. Then, CD5 positive and CD19 negative events (i.e. T cells) were gated out using a CD5 versus CD19 dot plot leaving only B cells. Finally, monoclonal B cells were separated from the polyclonal B cells using the immunoglobulin (IG) light chain gate, taking advantage of the fact that B-cell clones show either κ or λ IG light chain restriction.

DNA extraction and whole genome amplification

DNA from sorted B cells was extracted using Qiagen AllPrep DNA/RNA Micro kit (Qiagen, Hilden, Germany) according the instructions of the manufacturer. Genomic DNA was subsequently amplified using illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK). The concentration of all

extracted nucleic acid was measured using a NanoDrop 2000 spectrometer (Thermo Scientific, Waltham MA, USA).

Analysis of rearranged *IGH* genes and IG light chain genes

After DNA extraction from sorted B cells, rearranged *IGH* genes were detected using the IGH Somatic Hypermutation Assay v2.0 (Invivoscribe, San Diego, CA, USA). The PCR products were subsequently sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA, USA) and the primers from the IGH Somatic Hypermutation Assay v2.0. In addition, sequencing was repeated with *IGHV* family-specific primers.[2] *IG* light chain genes were amplified by an in-house diagnostic protocol based on Biomed-2 primers[2] and then sequenced. For cases in which rearranged *IGKV3* family genes were detected, additional *IGKV3* family specific primer (designed to acquire longer PCR product) were used, to confirm the finding.[3] All sequences were analyzed using the IMGT database (www.imgt.org).

IG sequences have been deposited in GeneBank, with accession numbers for *IGH* sequences KU560580-KU560602 and for *IGK* sequences KU560603-KU560624.

***MYD88* L265P mutation analysis**

The *MYD88* L265P specific mutation was analyzed using PCR and a SNaPshot mini-sequencing assay (Life Technologies). PCR was carried out using Phusion hot start DNA polymerase (Life Technologies) according to the supplier's instructions with the following PCR primers: 5'-TGC AGG TGC CCA TCA GAA GCG-3' and 5'-CAG ACA GTG ATG AAC CTC AGG ATG C-3'. Then a single nucleotide extension reaction was subsequently performed, according to the instructions of the manufacturer. The extension primers are as follows: 5'-CCC

CCC CCC CAG GTG CCC ATC AGA AGC GAC-3' and 5'-CCT TGT ACT TGA TGG GGA TC-3'. PCR products were fractionated by capillary electrophoresis using a 3100 Genetic Analyzer and GeneMapper v.4.1 Software (Life Technologies).

Re-analysis of previously published protein sequences

Since the nomenclature of IG genes has changed during the last decades and to compare older data with our IG gene sequencing data, previously published IG light chain amino acid sequences were reanalyzed.[4-9] On-line protein analysis software IMGT (http://www.imgt.org/3Dstructure-DB/cgi/DomainGapAlign.cgi) and IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/) were used for this purpose.

Statistical analysis

We used the Pearson correlation, two-tailed test, to unravel connections between molecular features and clinical data (Table 1). For $r > |0.4|$, data were examined manually, and degree-of-freedom (df) and p-values were calculated. Only statistically significant results ($p < 0.05$) are reported.

Supplementary results

Reanalysis of previously published data

Apart from our data presented here, few data have been published with regard to *IGKV* gene usage in CAD. The data available are from older publications reporting only small series of

patients. Therefore, we have reanalyzed previously published *IGKV* gene sequencing data of CAD[5, 7-9] using IMGT and IgBLAST. Reanalysis of published data shows that the *IGKV3-20* gene was used in most CAD cases (12/16). In one case (1/16) the *IGK3-15* gene was used, and in three cases (3/16) other *IGKV* genes were used. For most of the *IGKV3-20* sequences CDR3 sequences were not available. We also reanalyzed data from Li et al.[4] who studied the role of the IG light chain in determining antibody specificity by generating combinatorial antibodies (Table S2). Among other, they cloned the IGH gene from an immortalized human B cell line FS-6 that originates from a CAD patient.[10] The original FS-6 antibody bound I-antigen. The experiment involved pairing of 10 different IG light chains with the original IGH chain from the FS-6 cell line. In the experiment, antibodies with IG light chains obtained from three other CAD patients bound I-antigen by ELISA testing, with IGKV3-20 showing the strongest binding. IG light chains obtained from other individuals showed I-antigen binding exclusively when encoded by IGKV3-20. Of interest, the highly homologous *IGKV3-20* CDR3 sequence found in our study is also largely found in I-antigen binding sequences from Li et al.[4] (Table S2).

Supplementary references

1. Randen U, Troen G, Tierens A, et al. Primary cold agglutinin-associated lymphoproliferative disease: a B-cell lymphoma of the bone marrow distinct from lymphoplasmacytic lymphoma. Haematologica. 2014;99(3):497-504.
2. van Dongen JJ, Langerak AW, Bruggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia. 2003;17(12):2257-2317.
3. Payne K, Wright P, Grant JW, et al. BIOMED-2 PCR assays for IGK gene rearrangements are essential for B-cell clonality analysis in follicular lymphoma. Br J Haematol. 2011;155(1):84-92.
4. Li Y, Spellerberg MB, Stevenson FK, Capra JD, Potter KN. The I binding specificity of human VH 4-34 (VH 4-21) encoded antibodies is determined by both VH framework region 1 and complementarity determining region 3. J Mol Biol. 1996;256(3):577-589.
5. Capra JD, Kehoe JM, Williams RC, Jr., Feizi T, Kunkel HG. Light chain sequences of human IgM cold agglutinins (variable-region subgroups amino-acid sequence-kappa light chain-N-terminal). Proc Natl Acad Sci U S A. 1972;69(1):40-43.
6. Cauerhff A, Braden BC, Carvalho JG, et al. Three-dimensional structure of the Fab from a human IgM cold agglutinin. J Immunol. 2000;165(11):6422-6428.
7. Gergely J, Wang AC, Fudenberg HH. Chemical analyses of variable regions of heavy and light chains of cold agglutinins. Vox Sang. 1973;24(5):432-440.
8. Ruzickova S, Pruss A, Odendahl M, et al. Chronic lymphocytic leukemia preceded by cold agglutinin disease: intraclonal immunoglobulin light-chain diversity in V(H)4-34 expressing single leukemic B cells. Blood. 2002;100(9):3419-3422.
9. Silberstein LE, Jefferies LC, Goldman J, et al. Variable region gene analysis of pathologic human autoantibodies to the related i and I red blood cell antigens. Blood. 1991;78(9):2372-2386.

10. Pascual V, Victor K, Spellerberg M, Hamblin TJ, Stevenson FK, Capra JD. VH restriction among human cold agglutinins. The VH4-21 gene segment is required to encode anti-I and anti-i specificities. J Immunol. 1992;149(7):2337-2344.
11. Milstein C, Neuberger MS, Staden R. Both DNA strands of antibody genes are hypermutation targets. Proc Natl Acad Sci U S A. 1998;95(15):8791-8794.
12. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. J Immunol. 2004;172(6):3382-3384.
13. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta. 1992;1171(1):11-18.
14. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat Immunol. 2001;2(6):530-536.

Table S1. Clinical characteristics, *IGH* and *IGK* sequence analysis (see Table_S1.xls)

CA titer: some laboratories did not report results above titer 2048.

Results from IMGT database analysis: the names of the closest V-GENE, J-GENE, D-GENE of *IGH* and *IGK (IGHV*, *IGHJ*, *IGHD*, *IGKV, IGKJ*), the percentage of identity for *IGHV* and *IGKV* (*IGH* V-REGION identity %, *IGK* V-REGION identity %), AA JUNCTION defined by IMGT/V-QUEST (*IGH* CDR3, *IGK* CDR3).

Mutation in the core region of the N-glycosylation site (NHS) in FR2, or flanking residues only (I and S) and mutation hot spot KLS, located within FR3 as shown in detail in Figure 1.

The most significant findings are highlighted (most frequent *IGKV3-20* gene in blue; stereotyped *IGK* CDR3, mutated N-glycosylation sites and at least 2 mutations in KLS hot spot in bold).

NA: not available; ND: no data.

* Cases previously published by our group [1]

Table S2. Reanalysis of previously published data.[4]

Original light chain name[9]	IMGT gene name (our re-analysis)	Binding to I-antigen (ELISA)- original results[9]	CDR3 compared to the core sequence CQQYGSSPR . . TF
FS-7k	*IGKV1-5*	+ + +	CQQYNSYSR . . TF
FS-6k	***IGKV3-20***	+ + + +	CQQYGRSPR . . TF
RF-SJ3k	***IGKV3-20***	+ +(+)	CQQYGSSLGGYTF
37-2k	***IGKV3-20***	+	CQQYGSSPMY. TF
RF-TSIk	***IGKV3-20***	+	CQQYGSSPG . . TF
RF-TMC-1k	*IGKV3-11*	–	CQHRNNWPPSITF
RF-TS2k	*IGKV3-15*	–	CQQYNNWPP . ITF
91A3k	*IGKV1-33*	–	CQQGNALPR . . TF
HBVIλ	*λ*	–	CQVWDSSSDHVVF
FS-3λ	*λ*	+/–	CQSYDSSLSGYVF

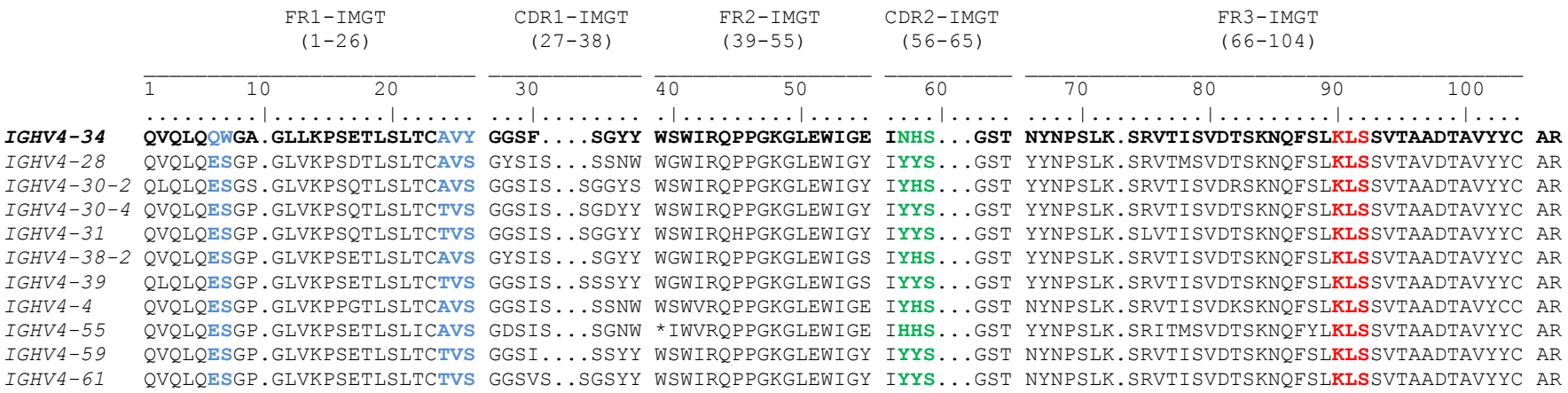
FS3, FS6 and FS7 represent CAD samples. Amino acid differences with respect to core sequences detected in our patients are indicated with red font.

Table S3. Summary of clinical data.

Clinical characteristics	No. of patients	median	range
Hemoglobin (g/dL)	26	9,0	4,9 – 13,9
Leukocytes (10^9/L)	26	6,8	3,1 – 20,7
Lymphocytes (10^9/L)	26	2,8	0,4 – 4,8
Reticulocytes (10^9/L)	24	140,5	76 - 256
Thrombocytes (10^9/L)	26	208,5	108 - 506
CA titer	24	2048*	128 – 256000*
IgM g/L	21	6,9	1,8 – 51,5
LD U/L	26	308,5	222 - 475
Bilirubin µmol/L	26	49,0	10 - 72
Age at diagnosis	25	70	52 - 84

* Some laboratories did not report results above titer 2048.

Figure S1. Comparison of amino acid sequences of *IGH* chains coded by different *IGHV4* family genes.



Legend:
Residues corresponding to the hydrophobic patch of *IGHV4-34* gene required for I-antigen binding (QW and AVY), are marked in blue. Residues corresponding to the core region of the N-glycosylation site in *IGHV4-34* gene (NHS), are marked in green. Residues corresponding to the mutation hot spot, located within FR3 of *IGHV4-34* gene (**KLS**), are marked in red. Germline *IGHV4-34* sequence in bold.

Figure S2. The KLS amino acid sequence in *IGHV4-34* FR3 is coded by a DNA sequence containing five known somatic hypermutation hotspot motifs. The motifs are 2 x R**G**YW, 2 x WR**C**Y and W**A** (R=A/G, Y=C/T, W=A/T).[11-14] Highly mutable nucleotides within these motifs and respective mutations in the sequences are in color (G in blue, C in yellow, A in red).