B cells in Multiple Sclerosis

- on idiotopes and antigen presentation

Doctoral Thesis by

Rune Alexander Høglund

Department of Neurology, Akershus University Hospital

Institute of Clinical Medicine, University of Oslo

2020

AKERSHUS UNIVERSITY HOSPITAL

© Rune Alexander Høglund, 2020

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-647-8

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Table of Contents

1. Summary

The role of B cells in pathophysiology of multiple sclerosis has been of great interest to wide array of researchers after the introduction of B cell depleting therapies demonstrated significant effects on central nervous system inflammation. In particular memory B cells have been suggested as an important subset. In multiple sclerosis patients both T and B cells aggregate in perivascular cuffs and the meninges of the central nervous system and are associated with areas of demyelination causing disability of the patients. Clonal expansion of T and B cells intrathecally suggests specific responses are driving the inflammation, but collaboration between the cell types have been suggested to be dysregulated. Identifying the disease driving agent or process is therefore a major goal. Identifying ways to target the pathogenic cells is another.

In this thesis, I discuss the potential of idiotopes as drivers of this dysregulated T-B collaboration. Idiotopes are epitopes derived from immunoglobulin variable regions, and thus the B cell receptors themselves. It has previously been demonstrated that T cells specific to idiotopes can initiate and drive pathological immune responses. In multiple sclerosis idiotope-specific T cells were demonstrated in two patients, suggesting such a mechanism could be relevant to investigate further. Due to the vast diversity of immunoglobulin repertoires, particularly among mutated variable regions, this has been a challenge to pursue in more patients. The work presented herein addresses this, by using bioinformatic prediction tools to identify potentially antigenic idiotopes in multiple sclerosis patients.

By using neural network prediction tools built by collaborators, we were able to identify key areas of multiple sclerosis patient immunoglobulin heavy chain variable regions with predicted high affinity for human leukocyte antigen class II molecules. These were associated with areas of high likelihood for endosomal processing. The predictions were further investigated and validated using *in vitro* assays in order to identify key factors in immunoglobulin degradation and in order to identify autologous, idiotope-specific T cells. Our findings suggest that multiple sclerosis patients have a repertoire of idiotope-specific T cells, responding to immunoglobulin heavy chain variable region peptides. All in all, the results suggest idiotopes participate in the dysregulated T-B collaboration occurring in multiple sclerosis.

This thesis further addresses how idiotope-driven T-B collaboration fits with current and previous knowledge of multiple sclerosis immunopathology and how this aligns with our current understanding of therapeutic mechanisms of action. Finally, I discuss the implications of our findings in both healthy immune regulation and potentially in dysregulation occurring in other auto-immune diseases.

2. Acknowledgements

The work in this thesis was performed at the Department of Neurology, Akershus University hospital in the period 2016-2019, as a PhD student at the Institute of Clinical Medicine, Faculty of Medicine, University of Oslo. It was funded mainly by a PhD grant from the Norwegian Research Council, with additional contributions from Akershus University Hospital, Fritz and Ingrid Nilsen endowment and unrestricted research awards from Biogen and Novartis.

I want to extend my deepest gratitude to Trygve Holmøy, my main supervisor, who has been nothing but inclusive and encouraging ever since our first email exchanges discussing project options. Your scientific insights never seize to impress. I have thoroughly enjoyed our discussions, writing (and re-writing) manuscripts or grant applications, and our meetings in social and scientific contexts. Your door is always open, and it has always felt as a door of opportunity. I am grateful for and also in awe of the lab-prowess of Andreas Lossius, my co-supervisor, who has provided invaluable support in every aspect of my scientific life. I'm lucky to have you by my side in the clinic and the lab. And Silje Bøen Torsetnes, my other co-supervisor, who has been a supporting and inspiring lab-partner and friend during my PhD period.

This thesis has relied heavily on collaborators close and afar. First and foremost, Robert D. Bremel and E. Jane Homan have been invaluable team-players during the entirety of the PhD period. Detailed discussions at every hour of the day (and night) due to the time difference have been very much appreciated. Previous research group colleagues Alina Tomescu-Baciu and Egil Røsjø, current colleagues Frode Vartdal and Justyna Polak all contributed with valuable input and constructive discussions during our group meetings. Additionally, I'm thankful for the work performed by Jorunn N. Johansen prior to my arrival, as my work would not have been possible without hers. I'm grateful to Jūratė Šaltytė Benth, who provided excellent and friendly guidance in advanced statistical methods. Finally, I thank Bjarne Bogen for peeks into his deep insights and wisdom regarding idiotope-driven T-B collaboration.

I would also like to thank the Department of Neurology, Akershus University Hospital. In particular the department head Tormod Fladby for his support of our work, and the previous and current heads of junior doctors, Kashif Faiz and Jannicke Koldeus-Falch for allowing me to keep one foot in the clinical world. Akershus University Hospital is also privileged to have excellent lab facilities at EpiGen, headed by Anna Frengen, who has been a pleasure to work with.

Additionally, I would like to thank the Faculty of Medicine, University of Oslo for allowing me to participate in PhD studies, all the way from the Medical Student Research Program ("*Forskerlinjen*"), and through the rest of the PhD program. I strongly encourage the Faculty to

continue recruitment of young medical students into research. I am also thankful for the thorough mid-term evaluation by Yvonne Böttcher.

Bedside to bench research like this would never happen without the willingness of patients to participate in our studies. I am indebted to every patient who actively chose to participate in our work, whose contributions has been invaluable.

And to my parents, Liv and Erland, thank you for always supporting me and my pursuit of knowledge.

Finally, thank you Marianne, my beautiful wife. You are the love of my life and have been supporting and inspiring me all the way. Together, we have had some amazing years, enriched by our children, Annabel Marie and Jakob Alexander, who bring us immense joy.

Immunology is complex and so is the brain, not to mention life in general. Attempting to understand neuroimmunology is to me the ultimate challenge in my professional life and a mission obviously incomplete. Balancing it all would not have been possible without the support of my family and guidance of my peers. For this I am forever grateful.

Rune Alexander Høglund
December 2019

3. Abbreviations

ADA	antidrug antibody
ANOVA	analysis of variance
APC	antigen presenting cell
BCR	B cell receptor
CCR	C-C chemokine receptor
CD	cluster of differentiation
CDR	complementarity determining region
CFSE	Carboxyfluorescein succinimidyl ester
CIS	clinically isolated syndrome
CLIP	Ii remnant peptide
CNS	central nervous system
CSF	cerebrospinal fluid
CXCR	C-X-C chemokine receptor
DC	dendritic cells
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EBNA	Epstein Barr nuclear antigen
EBV	Epstein Barr virus
FC	frequency class
Foxp3	Forkhead box P3
FW	framework region
GILT	gamma-interferon inducible thiol reductase
GWAS	genome wide association study
HLA	human leukocyte antigen
IC50	half maximal inhibitory concentration
IEDB	immune epitope database
IFN	interferon
Ig	immunoglobulin (M/G/E/A/D)
IGH	immunoglobulin heavy chain
IGK/L	Immunoglobulin light (kappa/lambda) chain
Ii	invariant chain
IL	interleukin
IMGT	international immunogenetics information system
IvIg	intravenous immunoglobulin
m/cTEC	medullary/cortical thymic epithelial cells
M/DMF	mono/dimethyl fumarate
mAb	monoclonal antibody
MHC	major histocompatibility complex
MRI	magnetic ressonance imaging
MS	multiple sclerosis
nLCMS	nanoliquid chromotography mass spectrometry
OCB	oligoclonal bands
OIND	other inflammatory neurological diseases
PAGE	polyacrylamide gel electrophoresis
PBMC	peripheral blood mononuclear cells
PCR	polymerase chain reaction
pHLA	peptide:HLA
PPMS	primary progressive MS
RNA	ribonucleic acid
RRMS	relapsing remitting MS
SDS	Sodium dodecyl sulfate
SLE	systemic lupus erythematosus
SPMS	secondary progressive MS
TCEM	T cell exposed motif
TCR	T cell receptor
TdT	terminal deoxynucleotidyl transferase
TGF	transforming growth factor
TNF	tumour necrosis factor
Treg	regulatory T cell
VDJ	Variable Diversity Joining genes

4. List of papers

Paper I - Høglund R.A., Lossius A., Johansen J.N., Homan J., Benth J.Š., Robins H., Bogen B., Bremel R.D. and Holmøy T.
In Silico Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis. Frontiers in immunology. **2017**;8:1255
doi: 10.3389/fimmu.2017.01255

Paper II - Høglund R.A.*, Torsetnes S.B.*, Lossius A., Bogen B., Homan E.J., Bremel R.D. and Holmøy T.
Human cysteine cathepsins degrade immunoglobulin G in vitro in a predictable manner. International Journal of Molecular Sciences. **2019**;20(19):4843
doi: 10.3390/ijms20194843
*Shared first-authorship

Paper III - Høglund R.A., Bremel R.D., Homan E.J., Torsetnes S.B, Lossius A. and Holmøy T.
$CD4^+$ T cells in the blood of MS patients respond to predicted epitopes from B cell receptors found in spinal fluid. Manuscript, **2020**.

Paper IV - Høglund R.A., Polak J., Vartdal F., Holmøy T. and Lossius A.
B-cell composition in the blood and cerebrospinal fluid of multiple sclerosis patients treated with dimethyl fumarate. Multiple sclerosis and related disorders. **2018**;26:90-5.
doi: 10.1016/j.msard.2018.08.032.

5. Introduction

5.1 The immune system

The human innate and adaptive immune systems are results of millions of years of evolution resulting in a vast array of cells, molecules and mechanisms designed to protect us from pathogens and other foreign challenges the body may encounter (1, 2). The innate immune system is our first line of defense, and consists of a limited set of barriers, and germline encoded cells and proteins capable of initiating immune responses rapidly based on limited pattern recognition systems discriminating self from non-self (3). The adaptive immune system consists of two main cellular lineages, T and B cells, capable of recognizing, adapting- and responding to, and remember a vast repertoire of antigenic determinants present on pathogens, albeit after a slower primary response (1). The main leukocytes of these two systems are all derived from bone marrow resident multipotent hematopoietic stem cells (Figure 1). Importantly, the systems do not work in isolation, but rather as a symphony orchestra while responding to foreign threats, or causing auto-immune conditions (3).

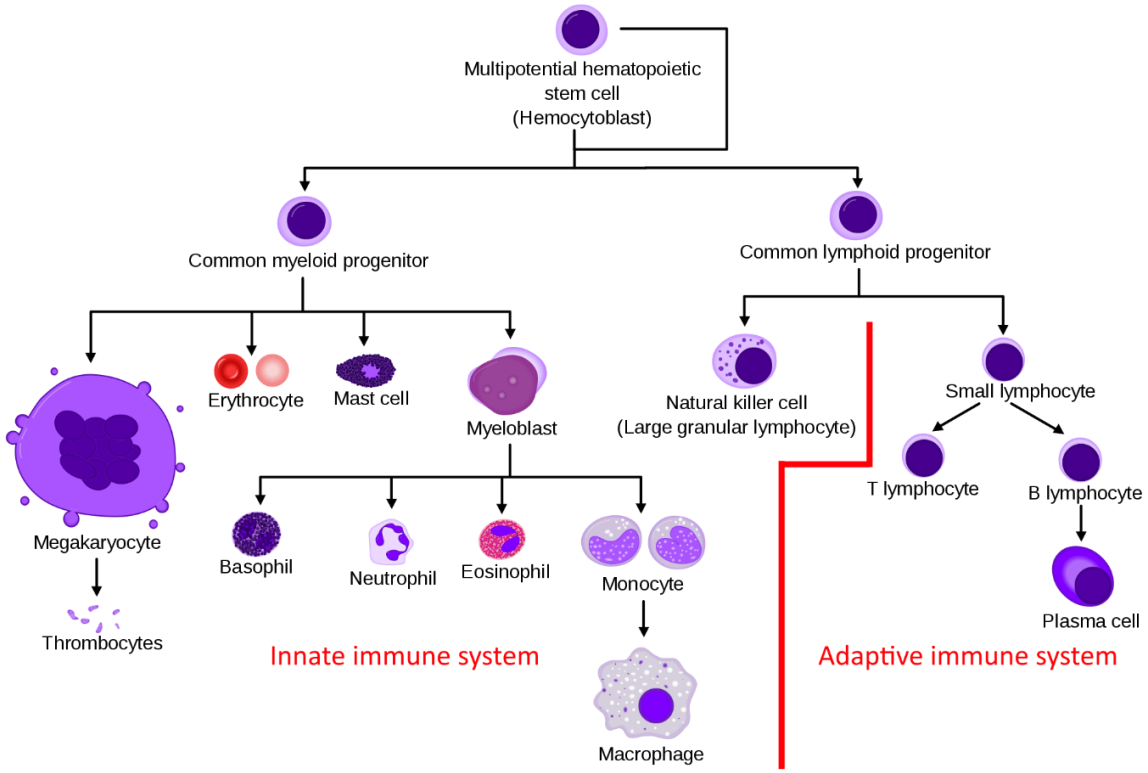


Figure 1. The hematopoiesis of the human immune system
All leukocytes, and even erythrocytes and thrombocytes, are descendants of common multipotent hematopoietic stem cells. The leukocytes are classified as members of the adaptive or innate immune system.
(After M. Häggström and A. Rad, 2009. Accessed 09.08.2019,
url: https://commons.wikimedia.org/wiki/File:Hematopoiesis_simple.svg. Modifications in red.
Licensed as CC BY-SA 3.0, https://creativecommons.org/licenses/by-sa/3.0/)

5.2 B cell development and the B cell receptor

B cells develop from lymphoid progenitors to mature (naïve) B cells in the bone marrow. During development from multipotent stem cells to mature B cells, the cells assemble what is known as the B cell receptor (BCR) (4). This receptor consists of two identical heavy chains (IGH) and two identical kappa or lambda light chains (IGK/L) with variable and constant regions (Figure 2). During development, immature B cells undergo several complex steps to assemble the variable regions (reviewed in (5), (6) and (4)), including sequential deoxyribonucleic acid (DNA) recombination of Variable (V), Diversity (D) and Joining (J) genes in heavy chain and afterwards V and J genes in light chain, as well as junctional insertions or deletion of nucleotides between the genes (7). The B cells avoid dual BCR expression by allelic exclusion of the non-utilized IGH and IGK or IGL genes (3). Final receptors are either immunoglobulin (Ig)M or IgD, due to alternative splicing of heavy chain RNA, causing the variable region to be paired with either μ or δ constant exons (Figure 2) (3).

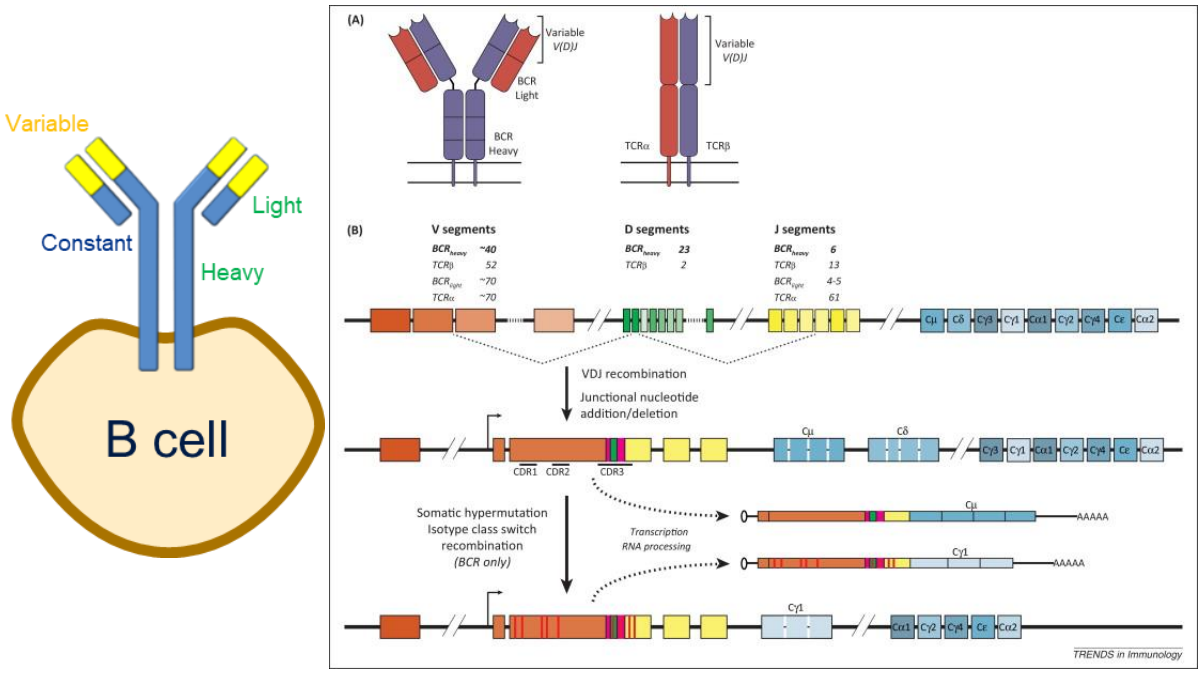


Figure 2. B cell receptor and diversification of antigen receptor repertoires.
Left panel: Mature B cells express B cell receptors (BCR) made up by heavy and light chains with variable and constant regions. **Right panel:** The BCR and T cell receptor (TCR) develop through similar mechanisms, including V(D)J recombination, junctional insertions and deletions. BCRs can additionally undergo somatic hypermutation as well as isotype class switch after exiting the bone marrow. (Note: Various estimates of number of V, D and J genes are reported, numbers in this reprinted figure may conflict more recent publications.) **Right panel:** Reprinted from Trends in Immunology Vol 35, Issue 12, J.J.A. Calis, B.R. Rosenberg, "*Characterizing immune repertoires by high throughput sequencing: strategies and applications",* pages 581-590 (8), with permission from Elsevier (© 2014).

Humans have at least 51 functional IGH V-, 25 D and 6 J genes. Similarly, IGK and IGL repertoires consists of 30-40 V and 4-5 J functional genes (4, 9). The random combination of these,

along with junctional diversity is the basis of a calculated potential diversity of up to 10^{13} theoretical BCR variants (8). Considering there are less than 10^9 circulating naïve B cells in the peripheral blood (5 L) of adults (10), the likelihood of identifying two identical BCR from such a repertoire in a single blood sample (8 mL), would be miniscule.

A broad diversity of BCRs ensures immune competence against a vast array of pathogens, but it also potentially allows reactivity towards autoantigens. Thus, a process known as central tolerance is necessary to drastically reduce the number of autoreactive B cells exiting the bone marrow (reviewed in (11)). Upon assembling the heavy and light chain, forming a functional BCR, the B cell may cease any further alterations and continue maturation if it receives sufficient signaling. This is called positive selection. If the BCR instead does not receive sufficient signaling (incomplete receptor, unsuccessful heavy-light pairing) or receives too strong a signal through the BCR (indicating autoreactivity in the bone marrow environment), it may continue light-chain rearrangements (receptor editing) (12, 13) or undergo apoptosis (clonal deletion) (14). During receptor editing, the B cell attempts to rearrange the light chain until it finds a functional and non-autoreactive heavy-light chain pairing, and by doing so may escape apoptosis (15). As B cells progresses from immature to mature the proportion of autoreactive cells drop, but not completely. When they exit the bone marrow, some mature naïve B cells are still autoreactive (multiple studies, summarized and presented jointly in (16)), and are subjected to peripheral tolerance inducing mechanisms.

Because of central tolerance and other mechanisms, the final variation of BCRs is not entirely random and not as rich as theoretically possible. Studies have shown that the usage of IGHV/D/J-, IGKV/J- and IGLV/J genes is heavily biased towards certain genes, and that these biases are surprisingly consistent across individuals (reviewed in (17)). A great example, relevant to this thesis is the biased usage of IGHV genes, where IGHV3 family genes represents nearly half of the peripheral blood BCR repertoire, followed by IGHV1 (approx. 20%) and IGHV4 (approx. 15%) (18). This biased family pattern was later shown to persist in naïve and memory B cell subsets, albeit with some changes at gene level (9). Additionally, the enzyme controlling nucleotide insertions, terminal deoxynucleotidyl transferase (TdT), has a G-nucleotide preference causing a bias towards G/C insertions, and A/T rich sequences seems to be more susceptible to deletions (19). All of these processes occur already early in fetal life (20), even the IGHV3 bias is observed as early. The bias is not limited to humans, as similar usage of IGHV3-like genes was also observed in other jawed vertebrates, implying a possible evolutionary role (21).

The mechanisms described above restricts the BCR repertoire somewhat, and diversity of mature naïve BCR repertoires was imputed to approximately match the number of circulating naïve B cells ($<10^9$) (9). Still, BCR variability does not end after bone marrow development but is rather expanded during a process called somatic hypermutation, addressed in section 5.4.

5.3 T cells – receptors and tolerance

Similar to the B cells and BCRs, T cells have their own distinct T cell receptors (TCRs). The receptor structure is, however, different and made up by αβ or γδ[1] dimers (Figure 2). The αβ T cells are divided into cluster of differentiation (CD)4^+ or $CD8^+$ subsets, the former has TCRs recognizing peptides on human leukocyte antigen (HLA) class II and the latter on HLA class I presented by antigen presenting cells (APC). T cells start their development from lymphoid progenitors in the bone marrow, but emigrate to the thymus as immature thymocytes for final maturation (3). While in the thymus, the TCR variable regions are assembled in a similar fashion to the BCR, including VDJ recombination and nucleotide insertions and deletions. Just like in B cells, receptor rearrangements are done sequentially, first the V, D and J genes of the TCR β chain are assembled while the thymocytes are in the subcapsular zone, and secondly the cells move to the thymic cortex, where V and J genes of the α chain are assembled (3). Due to high numbers of V, D and J genes the potential repertoire is immense. However, it is severely limited by tolerance-inducing mechanisms including positive and negative selection[2].

As the thymocytes move into the thymic cortex, they express both CD4 and CD8 (double positive cells), but this changes after the first step of selection processes in the thymus (22). Thymocytes interact with APCs known as cortical thymic epithelial cells (cTECs), with specialized proteolytic pathways (including "thymoproteasome" subunit β5t in mice and humans (23, 24) and endosomal cathepsin V/L2 in humans and L in mice (25, 26)), generating "private" antigens (22). Thymocytes that interact and recognize HLA-I presented peptides on cTECs continue differentiation to single positive $CD8^+$ T cells, and thymocytes that recognize HLA-II presented peptides become $CD4^+$ T cells. Single positive cells move on from this positive selection to the thymic medulla, while thymocytes without functional TCRs at this stage die from neglect (27).

Negative selection of autoreactive thymocytes occur both in the cortex (28, 29) and the medulla by dendritic cells (DCs) and medullary TECs (mTECs), and is an essential process to avoid autoimmunity (22). The mTECs are jointly capable of expressing and presenting a vast array of tissue restricted antigens, made possible by expression of autoimmune regulator protein (AIRE), causing promiscuous expression of genes and thus proteins that otherwise would not be presented in the thymus (30). In addition, these cells are able to "spread" the antigens to DCs in the microenvironment (31), causing a certain redundancy to the system. The DCs are also the main presenters of exogenously derived antigens, either acquired in the periphery or sampled from the serum (32). Unlike the cTECs, mTECs and DCs utilize conventional proteolytic pathways and enzymes to generate peptides for HLA

[1] γδ T cells have different properties than the more conventional αβ T cells and will not be discussed further in this thesis.

[2] Most research available on T cell tolerance stems from mouse models primarily, but several findings have later been confirmed for humans. Below is a short summary of tolerance inducing mechanisms based on knowledge from both mice and humans.

presentation, and thus the peptide repertoire presented by mTECs reflect what T cells may encounter in the periphery (22). Finally, the thymus is also home to a limited number of B cells with an antigen presenting phenotype, also expressing AIRE, that likely participate in negative selection of T cells (33, 34). The presentation of autoantigens has three possible outcomes for the thymocytes: 1) Cells with low auto reactivity in the medulla may exit the thymus as naïve $CD4^+$ or $CD8^+$ cells. 2) Cells with TCRs interacting too strongly to presumed autoantigens undergo apoptosis. 3) Cells that not quite belong in either of these groups may become what is known as regulatory T cells (Treg), a repertoire of T cells with tolerogenic functions. The mechanism by which these cells arise in the thymus is still up for debate, but seems to be a result of both TCR to peptide:HLA (pHLA) affinity and the pHLA avidity (density) in the thymus (reviewed in (35)).

The estimated number of unique TCRs in circulating T cells ($<10^7$) do not cover the estimated spectrum of possible epitopes they need to recognize ($>10^{12\text{-}15}$) (36-38). However, the T cell receptor only interacts with parts of the epitopes, the T cell exposed motifs (TCEM[3]) (39), as others are "hidden" in the HLA groove (40, 41). Despite tolerance, T cells are inherently cross-reactive, potentially cross-reacting to thousands of epitopes (36, 42, 43), although limited by the repertoire of peptides that fit and bind to self-HLA in vivo.

In circulation and upon antigen experience, naïve $CD4^+$ T cell may differentiate further depending on the conditions of activation. Subsets can be identified by expression of certain surface markers, cytokines or transcription factor, and have different functions (44). For instance, the Th1 and Th17 subsets are considered pro-inflammatory, due to expression of tumor necrosis factor (TNF), interferon gamma (IFN-γ) or interleukin (IL-)17 (45), while Tregs[4] with suppressive capabilities are recognized by expression of transcription factor forkhead box (Fox)p3 and transforming growth factor beta (TGF-β) (46).

5.4 B cells as antigen presenting cells – T-B collaboration and affinity maturation

In humans, mature, naïve B and T cells enter the blood after exiting the bone marrow or thymus. The cells find their way to the spleen, lymph nodes or mucosal lymphoid tissue only to be activated upon meeting their cognate antigens (4, 47). B cells are exceptional APCs, capable of rapidly capturing and internalizing external antigens with their BCR, degrade these in the endolysosomal system and present them to $CD4^+$ T cells on HLA class II molecules (47-49), steps leading to what is known as T-B cell collaboration (Figure 3) (50). Such T-B collaboration is the basis of clonal selection theory (51) and a cornerstone among immunological paradigms today (52). As each T- or B

[3] See methods section 8.3.3 for more on TCEM

[4] Peripheral induction of Tregs is also possible, but distinction from those generated in the thymus is challenging (46).

lymphocyte only has one version of a TCR or BCR, only cells (clones) with cognate receptors recognize the antigen and are activated.

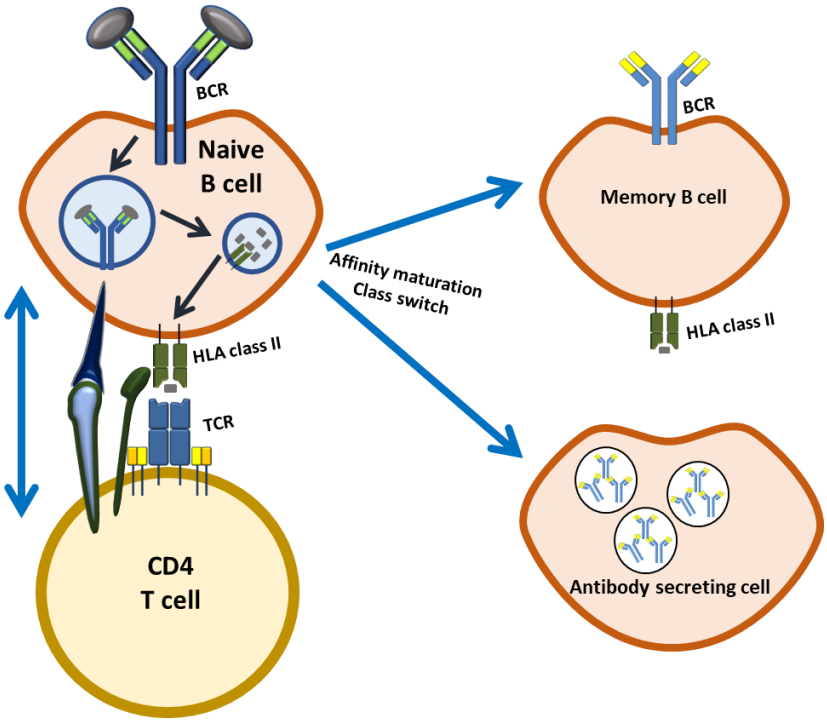


Figure 3. T-B cell collaboration

B cells are professional antigen presenting cells, presenting proteins recognized by their BCR and digested in the endolysosomal system on HLA class II molecules for $CD4^+$ T cells. Upon interacting with specific T cells, a cascade of intercellular interactions and signalling is initiated, triggering maturation and/or activation of both cell types. For the B cells, this includes affinity maturation and class switch, as well as differentiation to antibody secreting effector cells or memory cells.

The endolysosomal systems of B cells and other professional APCs are specialized to generate epitopes to bind and be presented on HLA class II molecules. To achieve this, the enzymes and process involved can neither be too destructive, nor too restrictive when it comes to degradation (53). Mixtures of cysteine, serine and aspartyl cathepsins (A, B, C, D, E, F, G, H, K, L, O, S, V, W and X), legumain and reducing enzyme gamma-interferon inducible thiol reductase (GILT) ensures this in professional APCs (47, 54). Upon BCR:antigen ligation, antigen and BCR are brought into the endolysosomal compartments of the B cells rapidly, and approximately 1000x more efficiently than simple endocytosis (47). At gradually increasing acidity they are subject to the endolysosomal proteases. Simultaneously, HLA class II with bound invariant chain (Ii) enters the same compartments and prepares to receive antigenic epitopes, by gradually trimming Ii to CLIP (Ii remnant peptide) and exchange it for a potential antigenic epitope at the proper pH (47). Upon ligation, the final pHLA class II complex is brought to the cell surface for presentation[5]. The exact fate of BCRs in the

[5] For the sake of simplicity HLA-DM and HLA-DO are omitted from this abbreviated recount of endolysosomal events. This does not imply in any way they are irrelevant. Also, the term MIIC (endolysosomal compartments with MHCII) is sometimes used inconsequentially and is avoided here (47, 55).

endolysosomal systems remains poorly described, even though it was shown fragments of BCR are presented on HLA class II (see section 5.5).

When the B cell presents antigens to cognate $CD4^+$ T cells in lymph nodes[6] or other lymphoid tissue, it may receive "help", inducing further maturation processes in germinal centers (56). The signals are received through costimulatory receptors including CD80, CD86 and CD40, that interact with the T cell alongside the pHLA:TCR interaction (3). The amount of help depends on the amount of antigen processed and presented in addition to presence of cognate T cells, and thus the B cells with receptors of better affinity for the relevant antigen receive more help (57, 58). Upon enough stimuli the B cells may undergo isotype switching[7], exchanging constant regions of IgM and IgD with more specialized isotypes and subtypes, including either $IgG_{1\text{-}4}$, IgE or $IgA_{1\text{-}2}$, each with different immunological attributes and/or structure (3, 60). Simultaneously, activation-induced cytidine deaminase (AID) mediates a localized mutational activity within the IGHV region (61, 62). Since further T cell help is still dependent on presentation of antigen, the B clones with the mutations increasing affinity, are preferentially selected for causing expansion of B cells with high affinity BCRs. These processes are called class switch recombination and affinity maturation, and results in either antibody secreting plasma blasts, plasma cells (effector cells) or memory B cells (Figure 3) (63). Upon new antigenic stimuli, these populations may either neutralize the antigen (antibodies secreted from plasma cells) or be reactivated to induce new clonal expansion (memory cells) (63). Finally, because of the mutations introduced into the already diverse IGHV regions, there can be near infinite possible variations of the BCR.

5.5 Idiotypes and idiotopes

The term idiotype originated when researchers found unique antigenic determinants on immunoglobulins that were not allotypes (variations in the constant chains), capable of generating anti-idiotypic antibodies (64, 65). The idiotype of an antibody or BCR corresponds to the variable regions (66). The term was made famous and became widely used upon Niels Jerne's immune network theory, claiming the immune system is a functional network of idiotypic and anti-idiotypic antibodies (67). For this he was he was awarded a split Nobel Prize in Physiology in 1984 (68). Later however, the subject fell out of favor and is often frowned upon today as a theory without substance (69, 70).

Idiotopes on the other hand, are epitopes derived from the variable regions (71), with potential to induce a T cell response (72-75). Such idiotopes may be presented by B cells spontaneously or during an activated state such as antigen driven response (76-80). Due to the variability of immunoglobulin variable regions, the list of potential idiotopes in any individual is near endless. Still, it seems T cells are largely tolerant to germline IGHV sequences (74, 81), while mutational activity

[6] In lymph nodes, these are known as T follicular helper cells (Tfh)
[7] B cell activation and/or class switch may occur independently from T cell help as well (3, 59).

can break tolerance (74). It was suggested mutations could increase major histocompatibility complex (MHC) affinity, or generate MHC presented epitopes unknown to the T cell repertoire. Such loss of tolerance represents a potential for error in classic T-B collaboration, as the BCR and bound antigen are both subject to the same degradation pathways in a near 1:1 molecular ratio, and it was further suggested it could lead to inappropriate T-B collaboration (Figure 4) (82, 83). This was later confirmed in mouse models, where interaction between idiotope-specific T cells and B cells with immunogenic idiotopes initiated germinal center reactions, isotype class switch and production of IgG towards self-antigens caused states of autoimmunity in the mice (80, 84-86). Recently it was shown that several human B cell lymphomas present idiotope peptides on their HLA class II surface molecules (87, 88), and the mutations involved have encouraged use of the term "neoantigens" when referring to potential immunogenic idiotopes[8] (88, 89). Even though presentation of idiotopes on HLA/MHC class II molecules have been thoroughly shown across multiple cell lines, human tissues and mouse models (76, 88, 90, 91), it remains unclear where and how these are processed intracellularly. Two paths have been proposed for B cells, one is that immunoglobulin chains are retained in endoplasmic reticulum/Golgi apparatus could be the source (82); the other is the classic endolysosomal pathway the BCR enters upon antigen ligation (49, 77), as was recently elegantly demonstrated to occur in naïve B cells in a mouse model (92).

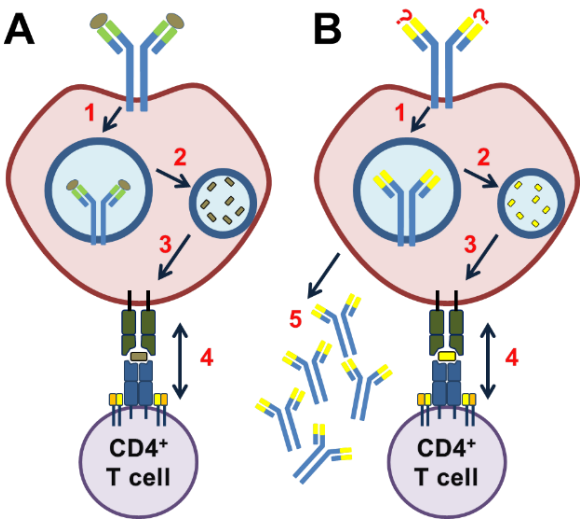


Figure 4. Idiotope-driven T–B cell collaboration. *"Unlike classical T–B cell collaboration (A) (…), idiotope-driven T–B cell collaboration (B) is an unlinked response. A BCR of any specificity (including self) is brought into the endosomal pathway (1), the BCR processed by endosomal proteases (2) and fragments from the variable region presented on HLA class II molecules (3). An idiotope-specific CD4+ T cell may help the B cell in a non-linked mechanism (4). All of steps 1–4 must occur for idiotope-driven T–B cell collaboration to take place and may result in differentiation of B cells into immunoglobulin G (IgG) secreting cells (5)".* Modified figure and text from Paper I – Høglund et al. (2017) Front. Immunol. 8:1255) (93). CC-BY 4.0. (https://creativecommons.org/licenses/by/4.0/)

Because the immunoglobulin repertoires are immense, research into idiotope-driven T-B collaboration has for long been restricted to cell- or mouse models with limited T and B cell repertoires. These models are excellent for describing cell-cell interactions and allow good control of the experiment, but only partly describe the complexity of a full-scale immune system. The past decade has spawned high-throughput technology capable of mapping our immune repertoires with

[8] Typically used regarding B-cell malignancies.

increasing depth and accuracy (38, 94, 95), allowing deeper investigation into how idiotope-driven responses may influence the systems. Bremel and Homan developed bioinformatic prediction tools specifically to investigate the immunogenicity of sequenced IGHV repertoires, and found that certain areas of IGHV are associated with a predicted higher affinity to HLA class II molecules, and further suggested tolerance could be determined by the occurrence of TCEM[9] (39). Tools like these may allow identification and classification of potential immunogenic idiotopes in large IGHV repertoires.

5.6 Brain immunosurveillance

The human nervous- and immune systems, products of evolution, are both "*highly sophisticated"* and *"specifically destined for interaction with the environment"* (96). Neuroimmunologists might also state they were destined to interact with each other. The mammalian central nervous system (CNS) has lymphatic drainage, capable of transporting immune cells and/or antigens, which was identified in the brain, and described anatomically and functionally by two independent groups in 2015 (97, 98), and later also for the medulla (99). In the brain, the main lymphatic vessels run along the dural venous sinus, communicating with cerebrospinal fluid (CSF) in the subarachnoid space, and drains to the deep cervical lymph nodes (100). Prior to this the brain was generally considered a more or less immune privileged site, even though prior observations indicated otherwise (101). Peripheral immune cells enter the CNS compartments mainly in one of three ways (Figure 5): 1) into CSF through the choroid plexus capillaries (102-104); 2) partly across the blood-brain-barrier, into perivascular (Virchow-Robins) spaces; or 3) across the blood-meningeal-barrier (105). During homeostasis, this is limited to a few memory $CD4^+$ T cells and monocyte-derived macrophages entering through choroid plexus epithelium and patrolling the CSF and perivascular spaces, together monitoring the CNS for foreign substances in addition to the parenchymal microglia (100). In order to invade the CNS parenchyma, cells must additionally traverse the *glia limitans*, but it seems few peripheral immune cells do so in healthy individuals (100). A recent study has identified tissue resident memory $CD4^+$ and $CD8^+$ T cells in normal appearing white matter of recently deceased individuals, which may have entered during inflammatory episodes (106).

[9] See methodological section 8.3.3 for further discussion on TCEM.

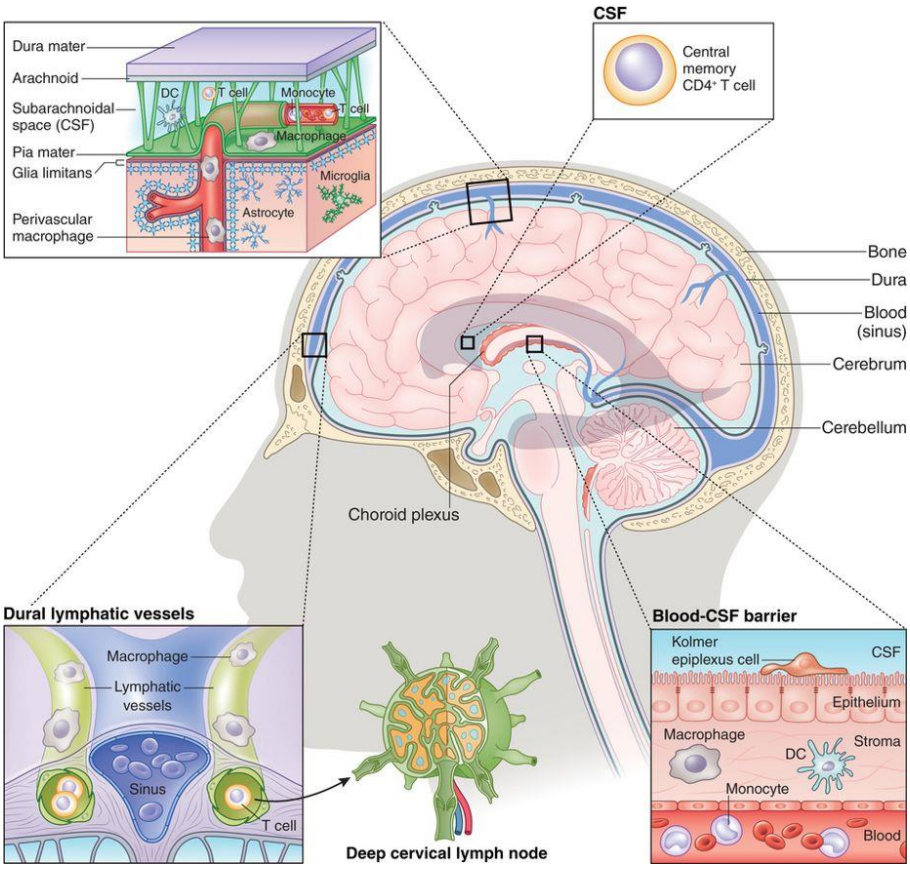


Figure 5. CNS immunosurveillance

"Scheme of the nondiseased brain, depicting anatomical structures and cells involved in ensuring tissue integrity. (…) CSF is produced in the choroid plexus (which has a blood–CSF barrier), bathes the brain, contains T cells and flows both in the parenchyma and in the subarachnoid area, which comprises arteries and the perivascular space. Whereas the CSF drains back to blood circulation, immune cells and proteins in CSF may be drained primarily through meningeal lymphatic structures to reach deep cervical lymph nodes (…)"

Figure and abbreviated text by Prinz and Priller (100). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, [Nature Neuroscience, 20, 136–144 (2017) doi:10.1038/nn.4475], © 2017.

5.7 Multiple sclerosis

In multiple sclerosis (MS) a combination of environmental, genetic and stochastic factors causes the immune system to launch attacks on the CNS, causing loss of neurological functions of heterogeneous nature (107) and thereby limiting the patient's capability to interact with the environment. MS is widely considered a "*chronic inflammatory, demyelinating and neurodegenerative disease affecting the CNS*" (108).

The Norwegian nationwide prevalence was estimated to be 203 / 100,000 (109), possibly increasing (110), and local studies have estimated incidences ranging from 8 to 11 / 100,000, affecting females over twice as frequent as males (111-113). These numbers are among the highest in the world (114, 115), and as the disease onset typically hits at around 20-40 years of age, MS is an important

cause for non-traumatic neurologic disability in young adults (108, 116). Because MS is a lifelong disease, it therefore represents a considerable socioeconomic burden (116).

Clinically, MS has classically been classified as either relapsing remitting (RRMS) or primary progressive (PPMS) phenotypes at onset (107). The former represents near 90% of the patients, who experience a disease with repeated subacute, inflammatory insults (relapses), causing neurological deficits that correspond to the lesions' locations, followed by part- or full recovery of function. The latter presentation is more uncommon, and presents as slow, progressive loss of function, without the classic relapses. After years of active disease, relapsing forms may progress into secondary progressive MS (SPMS) (107). Recently investigators have argued for a different, and perhaps more clinically relevant classification of progressive disease, taking into account the inflammatory activity and deeming the disease as active or non-active, with or without progression (117).

Both relapsing and progressive MS have the same core diagnostic criteria: Dissemination in time and space, as determined clinically and/or radiologically (magnetic resonance imaging, MRI), with additional para-clinical examinations (oligoclonal bands [OCB], dissemination in time only) as described by the 2017 revised McDonald criteria (118). A diagnosis of RRMS can be established after two clinical relapses with distinct locations and separation in time, or one clinical relapse and objective evidence of more lesions on MRI of varying age[10] or presence of oligoclonal bands as a measure of time dissemination. Progressive MS can be established if one year of disability progression is accompanied by two out of the following: One or more T2-hyperintense lesion at typical locations[11], two or more T2-hyperintense lesions in the spinal cord, or demonstration of OCB in CSF (118). Importantly, in both subtypes the clinical history and examination is the most important to establish a diagnosis.

Risk factors for MS include genetic and environmental factors. Recently the International Multiple Sclerosis Genetics Consortium (IMSGC) published a comprehensive genetic map of risk genes associated with MS (119), describing more than 200 risk loci accounting for up to 48% of genetic contribution to MS. By using gene ontology annotation, researchers have further found the majority of these genes are associated with peripheral immune cells and microglia (119, 120). Of the peripheral immune cells, risk genes are expressed in T cells, B cells, monocytes and NK cells, reaffirming[12] potentially important roles for all of these in MS immunopathology. Still, the strongest genetic risk contribution are HLA genes, and in particular HLA-DRB1*15:01 with odds ratio (OR) ~3 for MS for carriers (119, 121, 122), which has caused extensive investigations into potential disease specific epitopes that could drive a pathogenic T-B collaboration. On the other hand, some HLA class

[10] MRI criteria utilizes location and characteristics of lesions found in T2 sequences or as contrast (Gadolinium) enhancing on T1 during acute inflammation with loss of blood brain barrier integrity.
[11] Lesions are typically periventricular, cortical and/or juxtacortical or infratentorial. Lesions in the spinal cord are typically "short", unlike those of neuromyelitis optica (108).
[12] Potential pathological roles for these cells were previously described by other studies, see section 5.8.

I alleles are protective, including HLA-A*02:01 in particular (OR ~0.6) (123). Of note, IGHV/DJ genes and immunoglobulin allotype genes were not investigated fully in these genome wide association studies (GWAS) for technical reasons. Even so, the genetics does not fully explain the risk for MS. In monozygotic twins the concordance rate is only 25% (124), implying an environmental and/or stochastic contribution to risk. The strongest environmental risk factors associated with MS are smoking, low vitamin D levels, Epstein Barr virus (EBV) infection and obesity (125). Interestingly, having more than one of these risk factors these factors seemingly cause a combined extra increase in risk to develop MS, and they are all associated with effects on the immune system (126).

5.8 The immunopathology of multiple sclerosis

The immunopathology of MS is complex, and our understanding incomplete. Several forms of studies have contributed to our knowledge of the immune systems' contribution to pathology. Besides the previously mentioned GWAS, histopathological studies have shown leukocytes accumulate in perivascular spaces around post-capillary venules in the CNS, as well as meningeal infiltrates on the cortical surface in all forms of the disease (127-130). Upon investigation, these accumulations are dominated by $CD8^+$ T cells and B cells, followed by macrophages and $CD4^+$ T cells (127), and inflammatory disruption of the blood brain barrier allows cells to enter into parenchyma (131). In and around these accumulations there is demyelination, oligodendrocyte death, and neurodegenerative changes including axonal loss corresponding to lesions observed on MRI (132, 133). Immunogenetics studies have further demonstrated that the T and B cell populations from meninges, parenchyma and CSF are clonally expanded and related (134-136), and dynamically exchange with the blood and deep cervical lymph nodes (137-141). In the case of B cells, CSF IgG (found as OCBs) can be matched to IGHV transcripts sequenced from B cells in the CSF of MS patients (134, 140). Despite the cellular accumulations and cell clonality findings indicating some form of ongoing T-B collaboration, researchers have not been able to sort out what initiates this inflammatory response, and/or what sustains it afterwards (132).

There is no lack of theories of what drives MS. Most of the evidence above points towards an autoimmune disease mechanism. However, a common autoantigen for all patients with MS has yet to be established, for both T cells (142, 143) and B cells (143). This makes MS an outlier in a wider group of autoimmune inflammatory diseases of the CNS, including anti-neuronal encephalitides (144) (i.e. anti-NMDA[13] receptor encephalitis), anti-MOG[13] encephalomyelitis (145) or neuromyelitis optica (146) (anti-AQP4[13] associated disease). Some argue there may be methodological reasons for failing to find the autoantigens, others argue there may be more than one to find.

[13] NMDA - N-methyl-D-aspartate, MOG- myelin oligodendrocyte glycoprotein, AQP4 – aquaporin 4.

Others again have suggested infectious cause for disease. And due to the epidemiological association, EBV has been studied thoroughly in this regard. EBV infects naïve, germinal center- and possibly memory[14] B cells, and allow transformation from naïve/germinal center B cells to memory B cells, including class switch and somatic hypermutation (147). The virus remains latent within the memory B cell repertoire *in vivo* upon infection. Some have suggested viral reactivation within the CNS causes disease (148), a claim that remains controversial and contended to this day (148, 149). Other have suggested EBV contributes by inducing cross reactivity (150) or by expanding repertoires of autoreactive B cells (Pender's hypothesis) (151). However, as with autoantigens, experimental verifications remain inconclusive (152).

5.8.1 Idiotope-driven T-B collaboration in multiple sclerosis

Based on the evidence provided by genetics, histopathology, *in vitro* and *ex vivo* studies etc. (section 5.8), as well as lessons learnt from how the immune system responds to effective therapies (section 5.9), pathogenic T-B collaboration seems to be a key element of MS immunopathology. Given the intimate relationship between the BCR and antigen during ligation and B cell activation, a potential role for idiotopes in autoimmune disease seems likely. The subject has previously been investigated in mice, which developed systemic lupus erythematosus (SLE)-like disease upon idiotope-driven T-B collaboration (80, 84). Idiotopes has further been suggested as potential drivers or contributors to the pathology human SLE (153), rheumatoid arthritis (154), microscopic polyangiitis (155) and B cell lymphomas (87, 88, 156).

Inspired by the lack of a common autoantigen in MS patients and by investigations into idiotope-driven T-B mechanisms in other diseases, Holmøy et al. attempted to stimulate peripheral blood mononuclear cells (PBMC) taken from MS patients with autologous CSF IgG, and found that 14/21 responded with proliferation of HLA-DR restricted T cells, more than what was the case for myelin basic protein, autologous serum IgG (5/21) or with control patients with other inflammatory neurological diseases (4/17 and 3/17 for CSF and serum IgG, respectively) (157). This spawned follow-up studies demonstrating that CSF derived DR restricted T cell clones recognized specific mutated idiotopes within CSF derived monoclonal antibody (mAb) variable regions, causing secretion of inflammatory cytokines (158, 159). Further it was shown how such idiotope-specific $CD4^+$ T cells could induce oligodendrocyte apoptosis (160). These experiments provided evidence of a potential role idiotope-driven T-B collaboration in MS, and a coherent hypothesis of potential disease-related mechanisms was described in 2009 (161). According to this, B cell IGHV mutations acquired during affinity maturation create immunogenic idiotopes that are presented to T cells, and cause idiotope-specific T cells to drive and sustain the intrathecal synthesis of IgG in MS. The theory is generally compatible with missing general autoantigens, HLA risk associations, immunogenetic observations,

[14] Shown to occur *in vitro*, possibly also applies *in vivo*.

types of cells involved, mechanism of action for available therapies (see below), but was unable to fully answer the reason for CNS localization (161). The theory was additionally hard to either verify or falsify in multiple individuals because of the vast variation of T and B cell repertoires.

5.9 Therapeutic aspects for multiple sclerosis

The therapeutic landscape for MS has changed drastically over the past 20 years, ranging from only a few options (glatiramer acetate and beta-interferons) available for clinicians initially, to a full range spectrum of therapeutic agents (162). These include[15] the random copolymer glatiramer acetate (Copaxone®, Copemyl®), beta-interferons (Rebif®, Plegridy®, Avonex®, Betaferon®), anti-lymphoproliferative teriflunomide (Aubagio®), dimethyl fumarate (Tecfidera®), S1P-1 receptor blocker fingolimod (Gilenya®), cytotoxic adenosine analogue cladribine (Mavenclad®), type II topoisomerase inhibitor mitoxantrone (Novantrone®)[16], the anti-CD20 mAbs rituximab (Rixathon®, Mabthera®) and ocrelizumab (Ocrevuz®), the anti-integrin natalizumab (Tysabri®) and the anti-CD52 alemtuzumab (Lemtrada®) (Figure 6). In addition autologous hematopoietic stem cell transplantation is an option in severe and therapy-resistant MS (163).

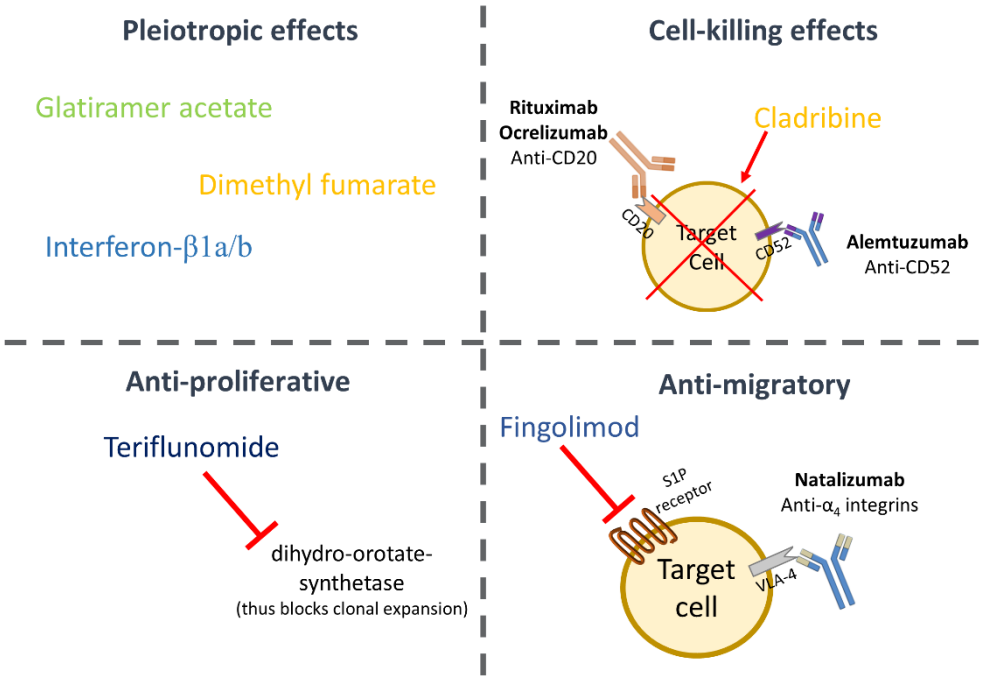


Figure 6. Available disease modifying therapies have various effects on the immune system
The therapies used in MS varies from random copolymers, receptor/enzyme blockers, cytokine analogues and toxic nucleosides to specific monoclonal antibodies. While some therapeutic agents have several effects on more than one cell type, other are more specific in their mechanisms of action (162, 164). S1P – Sphingosine-1-phosphate, VLA-4 - very late antigen 4 (α4β1-integrin)

[15] List limited to those available to Norwegian clinicians.
[16] Rarely used because of potentially severe side effects upon accumulated dose.

In addition to the changes in sheer availability, there has also been a shift in treatment philosophy. Most clinicians previously followed a step-wise approach using low-risk and low-efficacy drugs primarily, followed by escalation to stronger agents upon treatment failure, that also have increased risk of severe adverse effects (165). More recently, clinicians have started to appreciate the need to treat early and efficiently to achieve better long term for the patients (evidence reviewed in (166) and (167)).

Despite evidence that these drugs have several different mechanisms of action on the immune system (164), it has also been argued that one common mechanism may explain the effect of all. Baker et al. suggested memory B cells are the primary target of all available therapeutic agents in one way or another (168), and further argued for an entirely memory B cell centric pathophysiology (169). The relative depletion or inhibition of *blood* memory B cells also corresponded to the level of inflammatory activity (168), while also affecting numerous other immune cell subsets in more than one direction. Therapies that have failed clinical trials either did not affect or actually stimulated the very same subset (170), further strengthening the argument for their pathological role. However, a definite mechanism of how the memory B cells induce and maintain the inflammation seen in MS has remained unclear.

Even though preventive therapies have evolved, the way we treat acute relapses have been standing relatively still. Patients are upon diagnosis of acute relapses offered oral or intravenous steroids, causing faster but not necessarily more complete remission (171). If steroids fail and symptoms are severe, patients may be offered plasma exchange (172). At least one study is investigating a new agent to treat relapses (173).

6. Aims

This thesis aimed to address the relevance of idiotope-driven T-B collaboration in MS and identify strategies to find and potentially target pathogenic B cells contributing to this type of cell interactions.

- As the immune repertoire, even in the CSF of MS patients, is quite immense, identifying the relevant idiotopes have remained a challenge. For idiotope-driven T-B collaboration to occur and potentially generate pathogenic responses we hypothesized the following steps had to occur: First, the immunoglobulin or BCR has to be internalized and processed in the endolysosomal system. Second, resulting peptides has to bind to the patient's HLA class II molecules. And third, HLA presented idiotopes need to trigger a T cell response; the T cells cannot be tolerant to the idiotope. *A primary objective was thus to assess bioinformatic tools for identifying idiotopes fulfilling these criteria.*

- Detailed knowledge of BCR or immunoglobulin degradation is lacking in literature. Our predictive models indicated a possible relevant contribution by cysteine cathepsins in degrading IGHV. As HLA presentation of idiotopes depends on intracellular processing, *we aimed to understand how key proteases participate in degradation of IgG, as a model for the BCR.*

- Idiotope-specific T cells had been identified previously, using time consuming setups that did not allow initial phenotyping of reactive cells. Guided by our bioinformatic prediction tools *we sought to identify whether more patients do have similar idiotope-specific responses, classify these and confirm whether the bioinformatic methods were capable of correctly identifying them.*

- The relevance of idiotopes in MS rests on the premise that B cells have undergone affinity maturation to diversify the BCRs, making them potential targets for intolerant T cells in the CNS. Recent studies have argued memory B cells in circulating blood are major targets of all available treatment options for MS. *We sought to clarify whether the same is true intrathecally, for a commonly used oral disease modifying treatment option (dimethyl fumarate).*

7. Summary of papers

7.1 Paper I

By utilizing a combination of different bioinformatic prediction models, including predictions for HLA affinity and cathepsin processing trained using publicly available datasets, we investigated the potential immunogenicity of BCR IGHV repertoires in patients with MS, compared to patients with other inflammatory neurological diseases (OIND). Key findings were high predicted affinities of idiotopes to disease-associated HLA-DRB1*15:01, and a generally high predicted affinity for complementarity determining region (CDR) 3 derived idiotopes for HLA-DR and DQ in general. A similar variability was not found for HLA class I molecules. Additionally, we identified areas in the framework region (FW) with higher affinity and that these were associated with high probability of cleavage by cysteine cathepsins. Different IGHV families were predicted to differential degradation, due to structural differences. IGHV4 in particular was predicted vulnerable to cathepsin S. Average rarity of TCEM in the BCRs were found to be higher in MS patients than in OIND patients in some IGHV positions. By combining these outputs, we found that 42% of highly transcribed IGHV sequences in MS patients have at least one idiotope with high predicted HLA-DR affinity, high probability of cathepsin cleavage and contain rare TCEM.

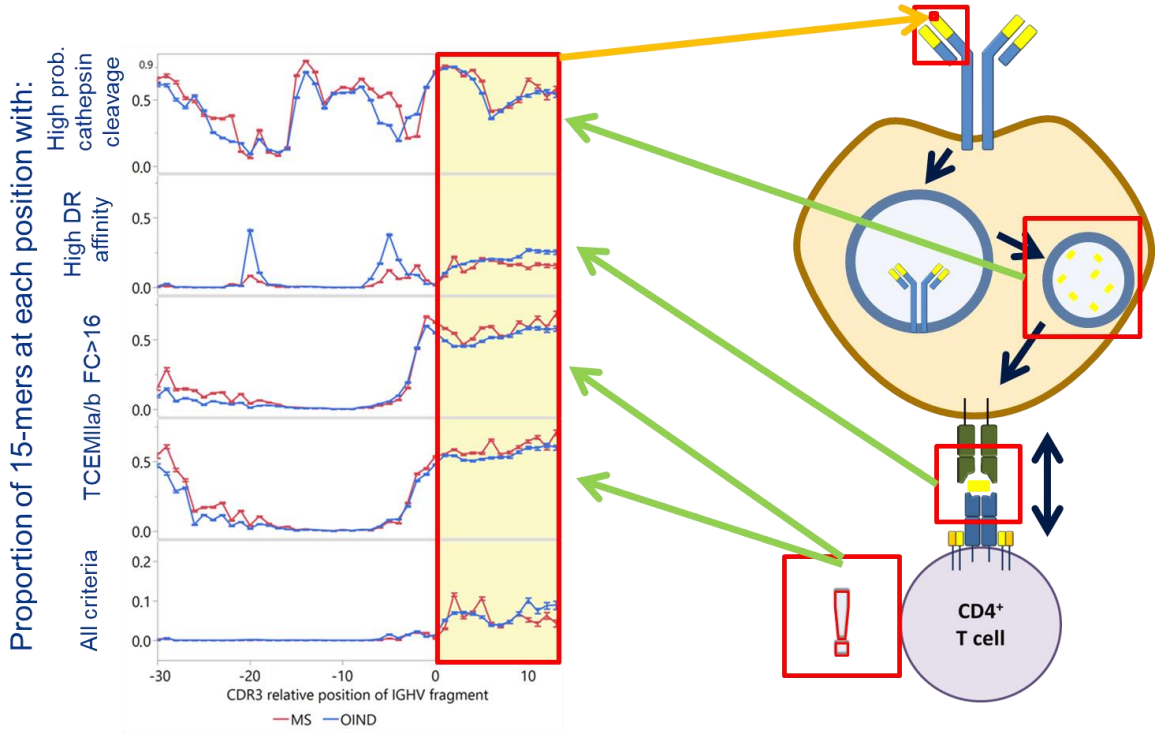


Figure 7. Graphical summary – Paper I

"Proportion of immunoglobulin heavy chain variable (IGVH) fragments that were predicted to have the potential to engage in idiotope-driven T–B cell collaboration. (…)" Each mode of prediction is highlighted by T-B collaboration figure. *"The upper panels show the proportion of fragments at each complementarity determining region 3 (CDR3) relative position that fulfils each criterion. The lower panel shows the proportion that fulfils all criteria. Nearly all fragments inhabiting all three features occur in the CDR3 region (yellow shading)."* Modified figure and text from Paper I – Høglund et al. (2017) *Front. Immunol.* 8:1255 (93). CC-BY 4.0 https://creativecommons.org/licenses/by/4.0/)

7.2 Paper II

In this paper we sought to validate the cathepsin predictions made in paper I. This was done by first assessing the accuracy of cathepsin cleavage prediction models[17] by using monomeric CNS proteins that also are known or potential substrates for cathepsins. Cathepsins S, L or B were mixed individually with substrates and sampled after several time points under similar conditions as the dataset used for training the models. We found that higher predicted probability of cleavage was clearly associated with higher occurrence of cleavages. Secondly, we assessed whether the predictions also were as accurate for tetrameric IgGs, and found accuracy was reduced. We hypothesized this reduction may be a methodological issue, as the nanoliquid chromatography mass spectrometry (nLCMS) used may not detect all cleavages in the larger IgGs. Finally, we described in detail the IgG degradation patterns by the three cathepsins, identifying a varying activity across different acidities, and confirmed that these cathepsins degrade variable regions differently, while simultaneously degrading the constant regions similarly across six different IgGs.

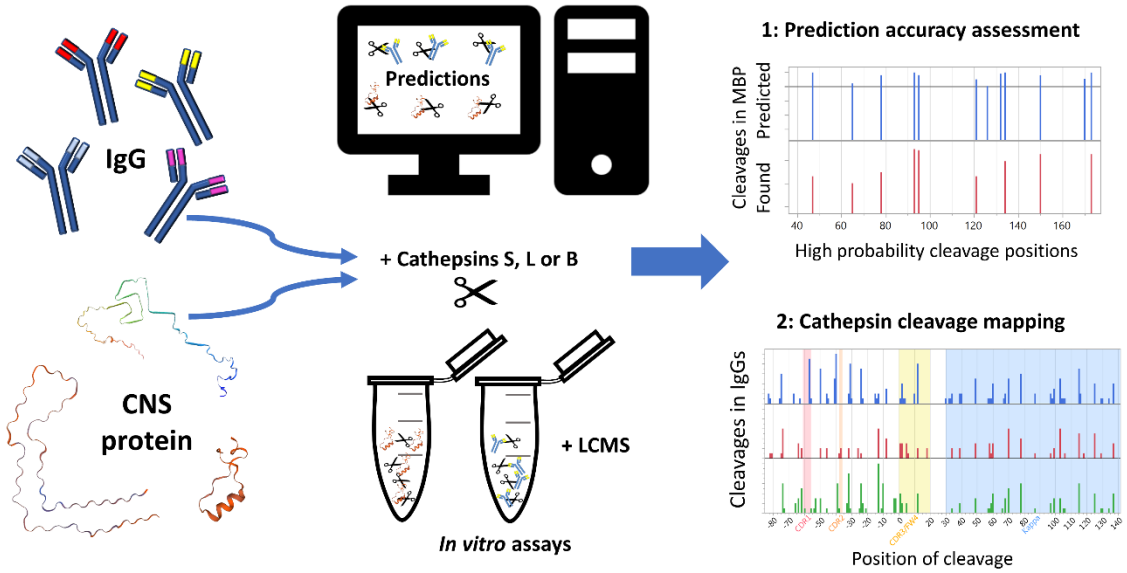


Figure 8. Graphical summary – Paper II
Monoclonal IgGs and recombinant CSF proteins were subject to *in silico* and *in vitro* processing to predict and detect cleavage activity by cysteine cathepsins. Predicted results were compared to actual mass-spectrometry detected cleavages, and cleavages were mapped within the proteins. (Figure from Paper II - Høglund et al (**2019**), *Int. J. Mol. Sci.*, *20*(19), 4843. (174). CC-BY 4.0. (https://creativecommons.org/licenses/by/4.0/)

[17] Same model as used in paper I, trained on publicly available datasets.

7.3 Paper III

We utilized the predictive models described in papers I and II, including predicted HLA-DR affinities and cathepsin cleavage, to guide selection of potentially immunogenic idiotopes and negative controls from the CSF IGHV repertoire described in paper I. Nine of the patients had available cryopreserved PBMC, sampled simultaneously as the lumbar puncture. The predicted idiotope peptides were synthesized and used to identify T cells specific for these in a flow cytometry-based activation assay, detecting expression of CD154 (CD40L) as the designated activation marker after stimulating the PBMC for 12 hours. We identified idiotope-specific memory T cells, frequently expressing C-C chemokine receptor (CCR) 6, in all patients assessed, in some patients we detected multiple robust responses. The idiotope peptides generating a response were derived from CDR3 related peptides, predicted to be released by cathepsins S or B expressed in B cells, and were associated with mutations that could influence affinity or cathepsin cleavage. The findings indicate that these MS patients all have idiotope-specific memory T cells, capable of entering the CNS.

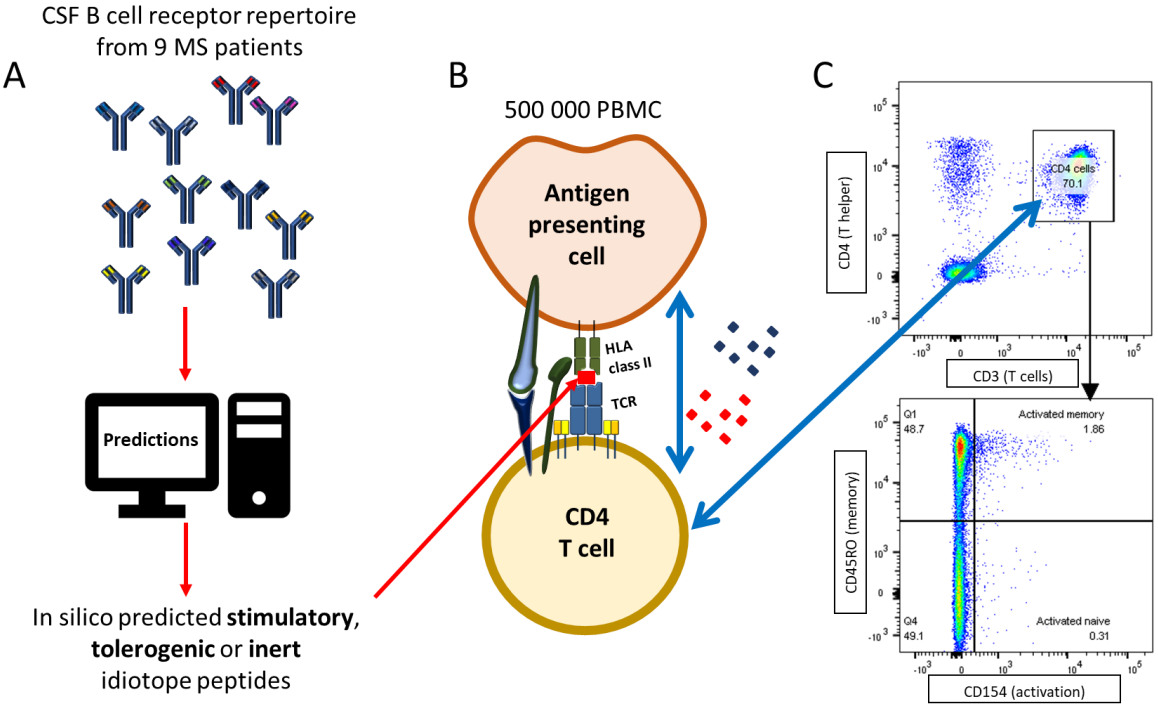


Figure 9 Graphical summary – Paper III

"Flow cytometry based idiotope-specific T cell activation assay. A) IGHV amino acid sequences from nine MS patients were run through predictive models to identify likely antigenic idiotopes based on HLA class II affinity, cathepsin cleavage and frequency classification (FC) of T cell exposed motifs (TCEM). B) 500,000 PBMC were stimulated with synthetic idiotope peptides predicted to be stimulatory, tolerogenic or inert as well as positive and negative controls for 12 hours in presence of anti-CD40 antibodies. B cells or other professional APCs with idiotope peptides bound to their HLA class II receptor may activate cognate $CD4^+$ T cells. C) $CD4^+CD45RO^+$ memory T cells specifically activated by idiotope peptides were detected by surface expression of CD154, upregulated upon TCR stimulation. The example shows a detected memory T cell response to an idiotope peptide." Figure and text from Paper III - Høglund et al., manuscript (**2020)**

7.4 Paper IV

We compared the B cell populations in blood and CSF of patients treated with either dimethyl fumarate (DMF) or other alternatives (glatiramer acetate, beta-interferons or no therapy) in a cross-sectional, explorative study. We found a reduced population of memory B cells in the blood during DMF therapy, thereby confirming other studies finding the same, and further found the reduction correlated with treatment duration. In CSF the absolute count of mononuclear cells was significantly lower in DMF treated than the others, and there was also a disproportionate decrease in plasmablasts. The study thus supports a potential B cell depleting mechanism of DMF but does not answer whether the effect of the drug is mainly on circulating or intrathecal B cells. As the study was explorative with a limited number of included patients, further studies to confirm our findings are necessary.

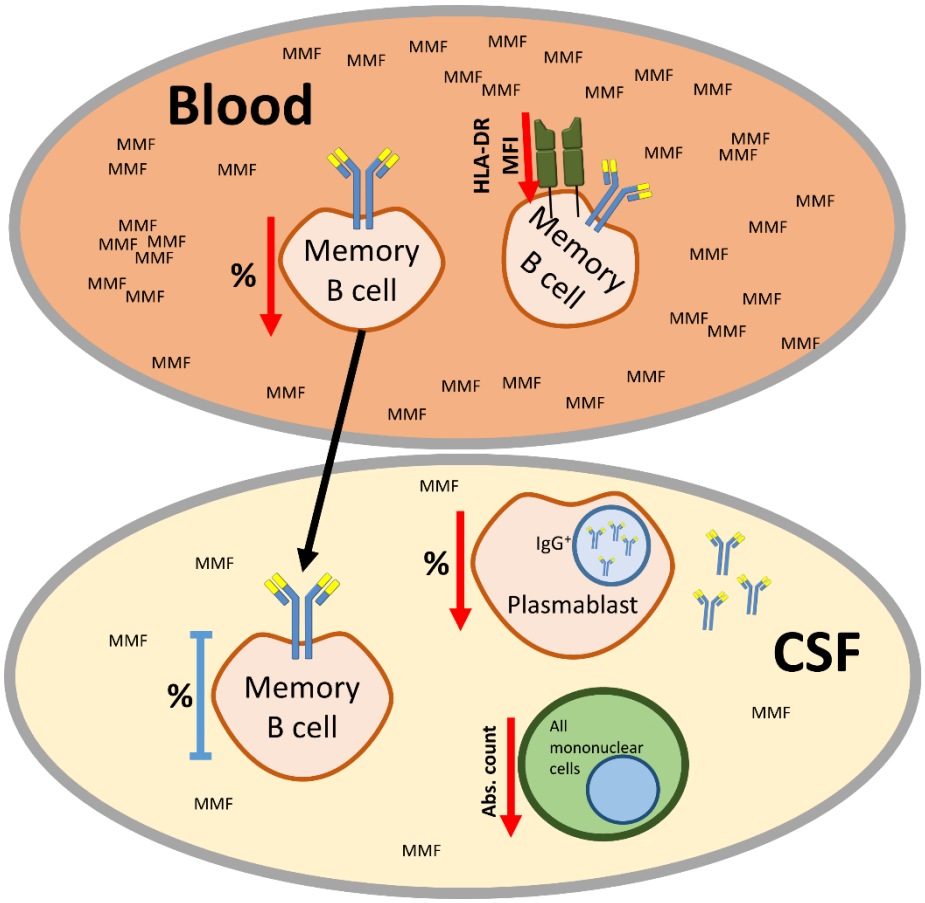


Figure 10: Graphical summary – Paper IV
Changes detected with flow cytometry of CSF and blood B cells in DMF-treated MS patients compared to control group. Figure also illustrates differences in blood and CSF concentration of DMF metabolite monomethyl fumarate (MMF). A modified figure was first presented on the poster P1214 *"Dimethyl fumarate alters the composition of B cells in the blood and cerebrospinal fluid of patients with multiple sclerosis", (Høglund et al.)* presented at the ECTRIMS congress, Berlin, Oct 12th, 2018.

8. Methodological considerations

8.1 Study populations (Papers I, III, IV)

Paper I and III both utilize the same base study population, consisting of patients recruited at Akershus University Hospital and Oslo University Hospital. Recruitment and collection of patient materials was completed prior to the start of this PhD period. Patients either had clinically isolated syndrome (CIS) or RRMS according to the 2010 McDonalds criteria (175), or other inflammatory neurological disease (OIND). Patients with CIS were later confirmed to progress to RRMS. Three of the MS patients had received immunomodulating treatment prior to inclusion into the study, the others were treatment naïve. Patients donated both CSF and blood used to isolate mononuclear cells or PBMC, respectively.

Paper IV investigated a different cohort of 28 MS patients previously recruited at the above-mentioned hospitals, participating in a different study (176). All patients had MS according to the 2010 McDonalds criteria (175), one was classified as SPMS, while the others were RRMS patients. Patients were divided into groups based on the mode of treatment.

8.2 Immunosequencing immunoglobulin variable regions (Papers I, III)

Immunosequencing of IGHV (utilized in paper I and III) was performed prior to the start of this project (140). Available MS and OIND patient samples from both PBMC and CSF were sequenced using Adaptive Biotechnologies Immunoseq protocol (94, 140), designed to sequence the FW3, CDR3 and part of FW4 region, while also minimizing polymerase chain reaction (PCR) bias with an optimized mixture of primers and computational adjustment of residual bias. The protocol utilized cDNA from ribonucleic acid (RNA) transcripts in the patient samples and could thus give indication of IGHV transcript abundance. The result was IGHV sequences spanning 110-130 bp, which further were analyzed using international immunogenetics information system (IMGT)/High V-quest (177) to allow translation to amino acid structure, imputed mutations and classification by IGHV family. The sequences did not cover FW1, CDR1, FW2 or CDR2, nor did they cover any light chain variable regions.

Subsequent assessment of IGHV immunogenicity thus had certain limitations, including a systematic bias towards FW3 or CDR3 derived idiotopes, and negligence for idiotopes in other IGHV or IGHK/L regions that could not be assessed. Still, this area has the highest variability (as reviewed in sections 5.2 and 5.4) and was perhaps the most relevant to investigate. As we had indication of transcript abundance, this could also factor into selection of idiotopes. Additionally, the possibility of PCR and/or sequencing errors could cause false positive or negative results in later analysis, but in this case several of the frequent transcripts were also confirmed to match with CSF IgG in the patients

(140). Throughout the thesis, the amino acid marking peptide starts or cleavage positions in variable regions are mainly given as number of amino acids relative to the cysteine of CDR3.

8.3 Bioinformatics and machine learning for *in silico* prediction models (Papers I, II, III)

Machine learning is a term covering all methods machines (computers) learn from data. This includes supervised learning like classification, regression and ranking, unsupervised learning like clustering and principal components analysis and reinforcement learning (178). Neural networks can be used for supervised learning, but unlike linear and logistic regression or other classic functions, the neural network learns without being told what shape the data should have (the function it should follow). The neural network learns from the data, and teaches itself what shape the function has, and is often termed deep learning. As nature is complex, the need for complex models is immense in medicine, and deep learning neural networks provide new ways forward (179, 180). Examples of use include classification and pattern recognition of images in radiology (181) and more relevant for my topic, prediction of epitopes in immunology (a comprehensive list of examples can be found in (182)). Recently, powerful and open algorithms utilizing mainstream graphic processor units (i.e. Google's Tensorflow, https://www.tensorflow.org/) have become available to the public, allowing anyone with sufficient knowledge to train advanced neural networks in their office or at home.

8.3.1 Neural network prediction of HLA affinity

The predicted peptide HLA affinity is an important tool for limiting selection to relevant epitopes to assess *in vitro*. As briefly mentioned above, several tools of predicting epitopes exist and all include prediction of HLA affinity using neural networks or other predictive models. For this thesis we have collaborated with Robert D. Bremel and E. Jane Homan in ioGenetics LLC. They had developed an HLA affinity prediction model utilizing principal component analysis of amino acid physical properties data (Figure 11), and known affinities expressed as ln(IC50)[18] for input (183, 184). The affinities were procured from the immune epitope database (IEDB) (185). During analysis the first three principal components explained 90% of the variance and was used for further analysis. The first component was correlated with hydrophobicity or polarity, the second with size and the third possibly with electric charge (184). As their system was built using a commercially available software (JMP©), predictions could be performed on desktop computers instead of supercomputers. However, the procedure does not limit itself to this platform, and may be utilized in context of other frameworks as well.

[18] Affinity for HLA molecules is typically reported as IC50 (half maximal inhibitory concentration), a quantitative measure indicating how much of the substance is needed to inhibit binding by 50% in competition binding assays. Lower values indicate higher affinity. Values used for training, and thus for output, were natural logarithms (ln) of IC50.

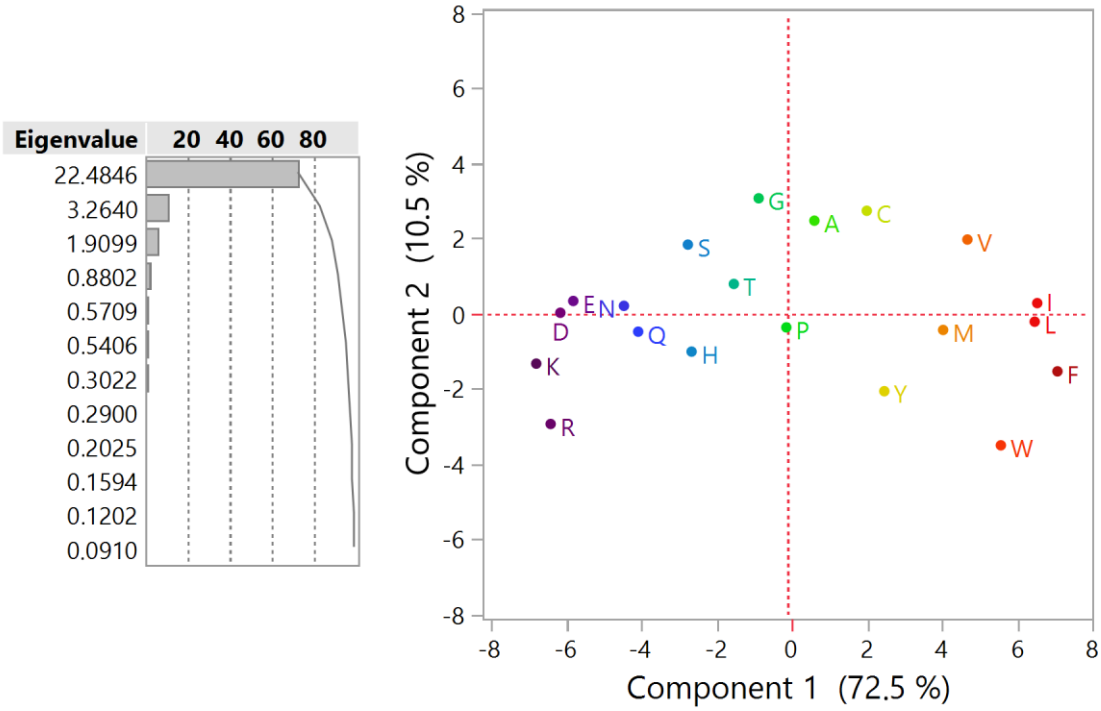


Figure 11. Principal components of amino acid physiochemical properties

All twenty amino acids have distinct physiochemical properties. Based on 31 data points for each amino acid, Bremel and Homan performed principal components analysis and found that the first 11 principal components (y-axis) explained 99% of the variability (x-axis), as shown in the left panel. The first three explain approximately 90%, and the right panel shows the relationship between the first two, explaining 83% of variability. Both affinity and cathepsin cleavage prediction models utilized the first three principal components as input values for training and predictions.

8.3.2 Neural network prediction of cathepsins cleavage

In protease cleavage assays, eight amino acids, four before and four after a potential cleavage site are typically labelled as P4P3P2P1 | P1'P2'P3'P4', where | is the cleavage site in the cleavage site octamer and the numbers indicate the relative distance from it. Some enzymes have clear preferences for certain amino acids in certain positions; Trypsin cleaves when lysine or arginine is in position P1 (186) while legumain prefers aspartic acid or asparagine in P1 (187). Cysteine cathepsins however, are much less specific and rather promiscuous when it comes to amino acid preferences (examples can be found in the MEROPS[19] database (188)). Predicting these enzymes' activity is thus tougher than for specific enzymes and require models that can be trained without fully knowing the enzymes' substrate specificity.

Similar to how HLA-affinity prediction algorithms were developed, ioGenetics also developed a way to utilize principal component analysis of amino acid physical properties data (Figure 11) to predict cathepsins cleavage probability (189), a method that was further validated in paper II. Importantly and unlike the HLA-affinity models using continuous IC50 values, the input and output for this model were binary, using cleave or non-cleaved as input to estimate a probability between 0

[19] MEROPS is a comprehensive database of peptidases and their activities.

and 1 that cleavage would occur. Instead of using the typical 9-mer (for HLA-I) or 15-mers (HLA-II) as input peptides, the algorithm used cleavage site octamers. However, the actual neural network ensembles used for prediction were limited to assessing the cleavage dipeptide P1P1' due to technical limitations.

The models were trained with data from proteomic identification of protease cleavage site (PICS) assays, where trypsin or GluC predigested proteome libraries were digested with cathepsins S, L or B and assessed using nLCMS to identify cathepsins cleavage sites (190). As GluC and trypsin have dissimilar cleavage site preferences, bias introduced by pre-cleavage was reduced, but not eliminated from the training set. Additionally, the input dataset did not cover every possible cleavage site octamer or even all possible P1P1' dipeptides, and trained models could be limited by this as well.

8.3.3 T cell exposed motifs

In the context of a peptide-HLA complex, only parts of the bound peptide will be "visible" to interacting T cells, while other parts will be "hidden" in the HLA groove. The exposed parts, the TCEM, are different for HLA-I and HLA-II bound peptides. Using data from Rudolph et al (40) and Calis et al (41), three patterns of TCEM were deduced by Bremel and Homan (39). TCEM I include residues 4, 5, 6, 7, 8 of a 9-mer in the context of HLA-I, and TCEM IIa and TCEM IIb includes residues 2, 3, 5, 7, 8 and -1, 3, 5, 7, 8 of a 15-mer with a 9-mer core in the context of HLA-II (Figure 12). Importantly, TCEMIIa/b were constructed with data from presentation on HLA-DR, and interpretations in context of other HLA-types should be mindful of this.

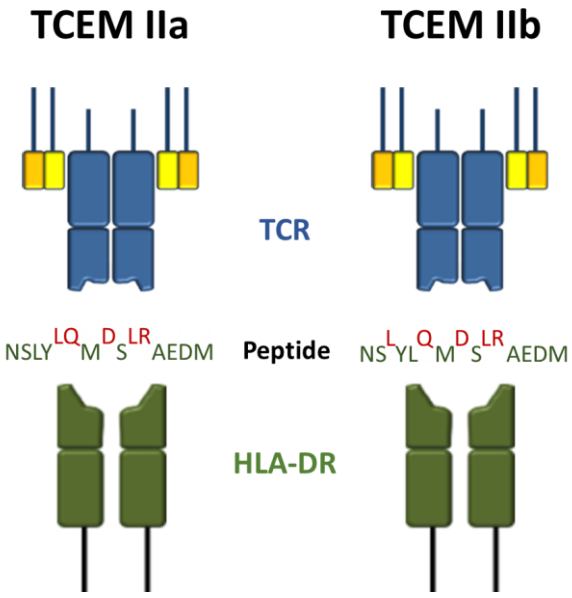


Figure 12. T cell exposed motifs
"Two T cell exposed motifs (TCEM) in context of peptide:HLA-DR binding. TCEM IIa consists of amino acids 2,3,5,7,8 and TCEM IIb of -1,3,5,7,8 in a 9-mer core of 15-mers (-3,-2,-1,1,2,3,4,5,6,7,8,9,+1,+2,+3). The non-linear 5-mer TCEM are the deduced sequences T cell receptors (TCR) may interact with, as the other amino acid residues remain hidden in the HLA-groove. There are theoretically 3.2 million (20^5) of each type."
Figure and text from Paper III – Høglund et al., manuscript (2020).

The frequency of TCEM occurrences within the human proteome (UniProt without IGHV (191)), human microbiome (from the National Institute of Health Human Microbiome Project (192)) and for two different IGHV databases (DeWitt et al. (9), and Johansen et al. (140)) were calculated

and expressed as frequency classes (FCs), as first shown by Bremel and Homan (39). This is a reverse log2 scale, where FC 0 ($1/2^0=1$) indicates the motif occurs in every IGHV sequence, and FC 21 ($1/2^{21}$) indicates it occurs approx. once every 2 million IGHV sequences. For proteome and microbiome occurrences, Johnson SI transformation of log2 values were used.

Unlike previous calculations that utilized full length IGHV sequences (39), FCs were for this work calculated using databases with only part of the IGHV sequence (9, 140), but with a much larger repertoire (see section 8.2). This caused the FC range to extend, ranging from 0 – 22, as opposed to 0 – 16 previously.

8.4 *In vitro* cathepsin cleavage (Paper II)

The *in vitro* cleavage assays utilized were primarily designed to mimic the conditions for the training dataset (190), in order to both assess actual cathepsin activity and cleavage prediction accuracy. These conditions included an acidic environment, reducing agent dithiothreitol (DTT), a cathepsin and a substrate mixed at human body temperature, representing a very simplified model of the human endolysosomal system (see section 5.4). An important difference from the training set was that we utilized whole proteins, not predigested, as substrates to assess cathepsin activity. As we only had predictive models available for cysteine cathepsins S, L and B, we limited testing to these even though assessing other cathepsins could be equally relevant. We set up conditions with differing pH and timing in order to detect potential qualitative differences in activity over time or acidity. Cathepsins S, L and B are all key enzymes in professional APCs (47, 193, 194).

Using therapeutic mAbs (IgG) alone as validation tools for cathepsins cleavage prediction had issues. By utilizing multiple IgG with highly similar constant regions and dissimilar variable regions we could demonstrate reproducibility as constant regions were expected to be degraded similarly across IgG. However, the regions also would serve as redundant regions in statistical validation. Additionally, the large size of IgGs could potentially influence results (see section 8.5). Thus, the choice to use CNS proteins, in addition to IgG, as substrates had justification. Cysteine cathepsins have been detected in resident brain cells including microglia in addition to traditional APCs and also have implied roles in neurodegenerative disorders and aging (195). Substrates synuclein has been investigated in Parkinson's disease, Tau in Alzheimer's disease and myelin basic protein in MS (195, 196). The choices were therefore made to include these four proteins/isoforms with non-redundant known sequences that could be encountered by B cells or other APCs within the same tissue or lymph drain area, and also was available as purified, recombinant versions.

8.5 Nano liquid chromatograph mass spectrometry (Paper II)

We utilized nLCMS to detect the vast array of peptides that may result from *in vitro* cathepsins cleavage assays. During this process, digested proteins are run through liquid

chromatography separation, before the fragments are analyzed by mass spectrometry, allowing very reproducible, sensitive detection of complex peptide mixes using small volumes and low concentrations (197). Results have to be analyzed using bioinformatic software, to interpret and align the results according to their position in the original protein, allowing a qualitative assessment of where the peptides originated from in the substrates.

An important and relevant limitation of nLCMS is the inability to detect larger, undigested fragments, typically of size >40 amino acids. In our setup, where the goal was to detect cleavage sites inflicted by cathepsins, this weakness could lead to an experimental bias of detecting only cleavages occurring in closer proximity to ends or cleavage "hot-spots", a bias that could increase with increasing protein size, or in proteins wholly or partly resistant to degradation.

8.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (Paper II)

Techniques involving sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) are commonly used to denature and separate proteins based on size and charge (198, 199). Proteins disulfide bonds are reduced with inclusion of DTT in the Laemmli buffer, and the sample heated in presence of an anionic detergent (SDS) to denature them and enable gel separation (199). Alternatively, samples can be run without DTT for "non-reducing" conditions. The objective in this paper was to identify residual substrate (IgG) after cleavage with cathepsins, and to assess the effectiveness of DTT at various pH. We utilized commercially available Laemmli buffer (Bio-Rad, CA, USA), DTT (Sigma-Aldrich, MO, USA) and 4-20% Criterion™ TGX gels (Bio-Rad). Gels were stained using Coomassie blue G-250 (Bio-Rad) after electrophoresis run to visualize bands.

The cathepsins cleavage assays already incorporated DTT in smaller concentrations than required for SDS-PAGE, thus reducing conditions were introduced prior to SDS-PAGE experiments. We therefore ran three negative controls, one from the original substrate batch in non-reducing conditions and the cleavage assay negative controls (no cathepsins) under both reducing and non-reducing conditions. By qualitatively assessing the bands arising from reduction of the tetramer IgG (light chain is approx. 25 kDa, heavy chain approx. 50 kDa and full size 150 kDa) we could assess effectiveness of both cathepsins cleavage and DTT reduction at various pH.

Tweaking gel gradients and concentrations, as well as choice of staining method can influence sensitivity at various protein sizes. Importantly, this particular setup was designed to complement the nLCMS results, by detecting larger residual bands, and not smaller oligopeptides that resulted from cleavage.

8.7 T cell activation assays (Paper III)

8.7.1 Selection and use of idiotope peptides

We sought to identify $CD4^+$ T cells responding to predicted idiotopes from CSF BCRs of MS patients. Idiotopes were selected based on the three assumptions required for idiotopes to generate immune responses (processing, presentation and intolerance), using the bioinformatic models developed for paper I. Potential 15-mer idiotopes used for the assay was divided into three categories.

- Idiotopes with high predicted HLA-DR affinity, high probability of cleavage and high FC → deemed "stimulatory".
- Idiotopes with high predicted HLA-DR affinity, high probability of cleavage and low FC → deemed "tolerogenic".
- Idiotopes with low predicted HLA class II affinity, low probability of cleavage with either low or high FC → deemed "inert".

The first two categories were idiotopes private to the individual and represent the presumed positive ones, as T cells responsive to processed and HLA-presented idiotopes seemed likely. However, it had been suggested that frequently occurring motifs could generate tolerance among the T cells (39), and thus the split by FC. The third category was a fixed set of idiotopes selected across patients and representing a control group, with no expected responses from T cells.

The predictive models used for selection were imperfect. Determining cathepsin cleavage probability utilized a liberal approach with relatively low cut-off probabilities at each end, also allowing predicted cuts adjacent to the core 15-mer. The lowered cut-offs were consistent with cleavages observed also at sites with lower probability in Paper II. On the other end the model was limited by only assessing probabilities for one cathepsin at a time. We picked idiotopes cut by *either* cathepsin S, L or B *at both ends +- 3 amino acids.* Thus, the assumptions do not account for mid-peptide cleavage, but this was intentional, as HLA-II groove binding could potentially protect from degradation (200). Second: This model did not include peptides with predicted cut by one cathepsin on one end, and a different on the other, as could be suspected to occur from other data (88). Third, the models did not include all relevant proteases. For instance, expression data indicates cathepsin H in particular could be relevant, as well as (but not limited to) cathepsins A, D, E, and legumain. Fourth, Cathepsin B is also a carboxypeptidase, which potentially can influence results by carboxy-end trimming post mid-protein cleavage.

Before use, the idiotope peptides had to be dissolved. The endosomal compartment upon activated cell status is a reducing environment, as GILT expression is upregulated to allow breaking disulfide bonds of antigens (201, 202). Loss of GILT expression disrupts antigen presenting capabilities of APCs (203). However, in our assay we could not be certain the peptides would enter the

endosomal pathway, but instead possibly bind to surface HLA class II molecules directly. To avoid oxidation that would likely not occur *in vivo*, we opted to avoid usage of dimethyl sulfoxide (DMSO) to dissolve the peptides, which is a known oxidative agent (204). We also dissolved the peptides immediately prior to use, as prolonged exposure to basic solutions may be oxidative as well (as per the manufacturer Mimotopes' instructions).

8.7.2 Detecting specific T cells

Peptide-specific $CD4^+$ T cells were detected utilizing a CD154 (CD40L) surface expression assay. This assay utilizes the upregulation of CD154 on T cells as an early (6-18 hours) response after antigenic stimuli. Conditions of the assay were designed according to previous optimization studies (205, 206). As CD154 is quickly downregulated upon interaction with CD40 on APCs we utilized anti-CD40 to block this. A total of 500,000 PBMCs from each patient were stimulated for 12 hours with either idiotope peptides, a viral peptide pool (Epstein Barr nuclear antigen (EBNA)-1) and anti-CD3/CD28 as positive controls, or human insulin peptide pool or no stimulation as negative controls (see also Figure 9).

Unlike T cell proliferation assays utilizing carboxyfluorescein succinimidyl ester (CFSE) or 3H-thymidin-incorporation, CD154 detection method allows immediate phenotyping of specific cells. This may reduce the risk of including cells that undergo bystander activation and avoids potential phenotype changes induced by continuous activation and cell divisions (207). An alternate approach of the same assay is to utilize intracellular staining of CD154 (205), however the required Brefeldin A interfered with surface expression of other relevant markers and resulted in inaccurate phenotyping in pilot experiments. Use of HLA tetra-/penta/dexa-mers would perhaps be an even better option (208), but because of variability in HLA class II alleles among the patients, this also represented a substantial increase in cost and was discarded.

8.7 Flow cytometry panels (Papers III, IV)

Multicolour flow cytometry is a cornerstone method in cellular phenotyping. By staining cells with specific antibodies conjugated to fluorescent substrates, they can be sorted by population based on surface and/or intracellular protein expression profiles (209). This allows not only classification of cells by subset, but also by function.

The flow cytometry performed for paper III were all done using a FACS Canto II (BD Biosciences), with a three-laser setup (407 nm - violet, 488 nm – blue and 633 nm - red), with eight available detectors. Fluorochromes were chosen by assigning bright fluorochromes to presumed low-expressed markers (i.e. CD154, CCR6, CXCR3) and moderate to dim fluorochromes to markers with distinct expression (i.e. CD3, CD4, CD8, CD14). Fluorochrome compensation was calculated automatically prior to acquiring data, using a commercially available kit (Invitrogen Ultra

Compbeads). Due to severe spill-over between the two violet laser detectors, we chose to only utilize one, and designated our priority marker (CD154) to this. The dimmest fluorochrome (APC-H7) was assigned as dump channel, including markers for CD14, CD8 and Near-IR Live/Dead kit. Gating for populations with unclear delineations in these experiments were set according to fluorescence minus one (FMO) samples (CCR6, CXCR3, CD45RO) or negative controls (CD154). The full panel, including descriptions of filters and lasers used can be found in Table 1.

Table 1 – Surface markers and fluorochromes used in Paper III

Laser	BP Filter	LP	Fluorochromes	Marker	Marker Brightness	Clone (cat. #)
407 nm (Violet)	510/50	502	**Not used**	-	-	-
	450/50		**BV421**	**CD154**	Brightest	TRAP1 (BD 563886)
488 nm (Blue)	585/42	556	**PE**	**CXCR3**	Bright	11A9 (BD 560619)
	780/60	735	**PE-Cy7**	**CD45RO**	Bright	UCHL1 (BD 337168)
		670 655	**PerCP-Cy5.5**	**CD4**	Moderate	SK3 (BD 566316)
	530/30	502	**FITC**	**CD3**	Moderate	UCHT1 (BD 555916)
633 nm (Red)	660/20		**APC**	**CCR6**	Bright	1C6 (BD 550633)
	780/60	735	**APC-H7** **LIVE/DEAD™ Fixable Near-IR**	**CD14** **CD8**	Dim (Dump)	MφP9 (BD 560180) SK1 (BD 560179) (L-34959)

The fluorochrome panel for paper IV was primarily designed to detect and phenotype IgG^+ naïve and memory B cells, as well as plasma cells and plasmablasts in blood and CSF. The markers included anti-CD19, anti-CD27, anti-CD38, anti-CD138, anti-HLA-DR, Ki-67 and anti-IgG (B cell phenotyping), CD3 and CD14 (dump channel), and finally also markers for IgG allotypes. Thus, the panel did not include markers for IgD that could discriminate class-switched and unswitched memory cells[20] (176). The flow-cytometry data was acquired using an LSRFortessa (BD Biosciences), with four lasers (405-, 488-, 561- and 640-nm).

8.8 Ethics and data management

The collection of patient materials for papers I, III and IV was approved by the Regional Ethical Committee South East (2009/23 S-041 43a) and the local Data Protection Officers at Akershus

[20] IgG^- cells could potentially be B cells with other isotypes.

University Hospital and Oslo University Hospital. Paper II did not use patient materials. All patients recruited gave their oral and written informed consent before being included into the studies.

In papers I, III and IV patients underwent lumbar puncture during inclusion of the study. Lumbar puncture is a standard part of the diagnostic workup before setting the diagnosis MS or differentials. It is a relatively uncomplicated procedure performed by the clinician but is not without potential side effects. The most frequent is spinal headache, which is of inconvenience to the patient and rarely dangerous. Much less frequently, more serious infections or bleedings may occur. Most included patients were included during diagnostic workup, and only underwent one lumbar puncture. Some volunteered to undergo lumbar puncture specifically for the study. The materials collected for papers I, III and IV have been subject to multiple studies, thereby maximizing the output from patient contributions. The Ethical Committee approved these procedures.

Some experiments (Papers III and IV) required the use of samples stored in a research biobank. These were securely stored in non-identifying, coded containers at the Department of Immunology and Transfusion Medicine, University of Oslo, administered by senior research group personnel. The group has approval for this biobank from the Norwegian Directorate of Health.

All sensitive data from these studies are coded and safely stored at secure servers (UIO, TSD-Services for Sensitive Data). Patient identification keys are stored on password-protected and encrypted disk-drives, locked away in a physical safe to which only authorized members of the research group have access.

8.9 Statistics

This thesis analyzed data in various forms and shapes, and the variation in statistical tests reflects this. Throughout the thesis, either JMP® (SAS Institute, Cary, NC, USA) or STATA v 14.1 (StataCorp LLC, TX, USA) were used for analyzing data, performing statistical tests and generating graphical outputs.

In all studies, the significance level vas set to 5%. In cases with two groups, few samples and non-normal distributions (Paper IV) we utilized non-parametric 2-sample Wilcoxon exact test[21]. To assess differences across multiple groups (Paper I and II) we utilized analysis of variance (ANOVA), and when variance was not homogenous, we used Welch's test. Differences between groups were determined with Tukey-Kramer HSD. In some cases, the distributions were not following normal distribution, but using Welch's ANOVA was still deemed adequate after consulting with a collaborating statistician, due to sample sizes. Thus, a non-parametric alternative like Kruskal-Wallis did not yield results different from Welch's ANOVA.

[21] Sometimes referred to as Wilcoxon Rank Sum test or Mann-Whitney U test.

In papers I and III we had to resort to more complex models. In paper I we sought to compare the predictive outputs after analyzing IGHV repertoires from MS and OIND patients. Various comparisons were performed, including FW3 vs CDR3, MS vs OIND at each CDR3 relative position and highly transcribed sequences vs sequences with low transcription at each CDR3 relative position. To achieve these tests, the datasets were sorted and split accordingly, as described in detail in Paper I. However, the data was multilayered as each patient had varying amounts of transcripts, some with more clones than others and we also were uncertain of intra-transcript dependencies[22], we had to utilize models that controlled for this. The solution was utilizing multilevel linear mixed models to also assess cluster effects on the data. As some of the data required multiple testing, adjustment with the Benjamini-Hochberg procedure was utilized (210). The false discovery rate was set at 20%, as this primarily was an exploratory study with the intention to provide insights into further studies.
Paper III also utilized a simpler mixed-model to analyze chemokine expression differences between unactivated and activated T cells within the same sample. This helped to control for differences that could be driven by individuals with multiple T cell responses, drowning those with fewer responses. All use of mixed-models were done in collaboration with a collaborating statistician.

In both paper I, II and III standardization procedures were used to improve comparability. Paper I and III utilized Johnson SI standardization to bring HLA-affinity prediction values on a comparable scale, as ln (IC50) values may vary across HLA-types and alleles. Johnson SI transforms data to make the mean zero, thereby allowing "ranking" of the affinity predictions, which was particularly useful in paper III when selecting idiotope peptide candidates. Paper II instead used z-standardization of number of observed cleavages to adjust for differences in protein sizes (for CNS proteins) and/or concentrations used (for IgGs) in the experiments

[22] At each CDR3 relative position, the previous and next 15-mer overlapped with 14 amino acids.

9 Discussion

9.1 Memory B cells as therapeutic targets in multiple sclerosis

Interest in the B cells role in MS pathophysiology has had a massive surge since the introduction of anti-CD20 therapies for MS (211-213), as is evident from several recent and thorough reviews on the matter (168, 169, 214-217). Before this, investigations into the multiple potential roles of T cells dominated the field (217). Still, B cells have also previously been a recognized part of MS immunopathology, much due to the presence OCB in spinal fluid of patients with MS (218, 219) making it a hallmark for the disease and even included in the most recent diagnostic criteria (118), but also due to discovery of ectopic B cell follicles resembling germinal centers in MS patients' meninges (129). Potentially pathological B cells are believed to exchange dynamically between blood and CSF (134, 137-140), but cells expressing required chemokine receptors are more likely to enter the CNS or CSF when the blood-brain barrier is not disrupted by inflammation (216). Memory B cells are enriched for activated leukocyte cell adhesion molecule (ALCAM) important for their CNS entry, in addition to the chemokine receptors (220). Memory B cells are also depleted or functionally impaired in blood upon using currently available therapies for MS (Figure 13) (168, 221), supporting a key pathological role for this subset of B cells. Unlike naïve cells, memory B cells, does not rapidly repopulate the peripheral blood after depletion (222).

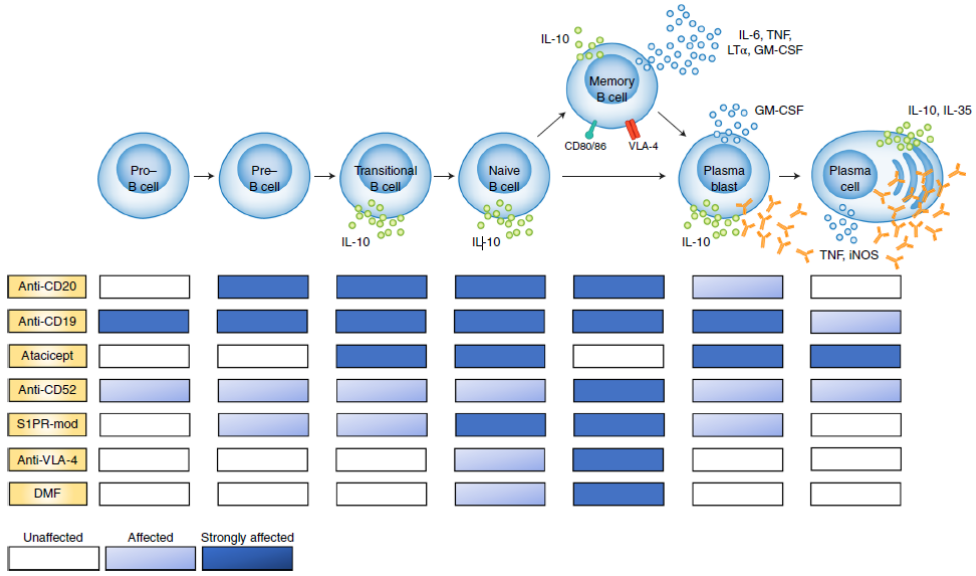


Figure 13. "*Different therapies applied in MS preferentially target distinct cell stages along the B cell lineage.* *Schematic representation of the stages of B cell maturation and differentiation (…). Therapies used or considered in patients with MS are shown in orange rectangles at left. Blue rectangles indicate B cell–lineage stages thought to be affected by the different treatments (…). Atacicept, which is the only one of these therapies that exacerbates rather than limits MS attacks, is also the only one that relatively spares memory B cells. These results from in vivo patient exposures to various immunological interventions support the view that key pathogenic B cells involved in the development of new MS attacks reside in the memory B cell pool. iNOS, inducible nitric oxide synthase."* Figure and abbreviated text reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, from "*Reassessing B cell contributions in multiple sclerosis*" by Li, R., Patterson, K.R. & Bar-Or, A. *Nat Immunol* 19, 696–707 (221). © 2018.

Despite the positive effects of anti-CD20 therapies, depleting B cell numbers in both blood and CSF (223), rituximab has been shown to only cross the blood brain barrier in very small quantities (224). Rituximab molecular size and constant regions shares high similarity to other cell depleting mAbs used for MS (alemtuzumab, ocrelizumab) (sequences available in Paper II), indicating blood to CSF transfer of these likely resembles that of rituximab. For mAbs, any intrathecal effect may differ negatively from blood because of lacking complement factors or effector cells (225). Still, intrathecal effects of other available therapies on B cell subsets have remained poorly described. In paper IV we sought to illuminate this for one of the small molecular therapies, DMF, that had previously been shown to reduce memory B cell numbers in the periphery (226-229). The active metabolite of DMF, monomethyl fumarate had been shown to cross the blood-brain barrier and reach approximately 15% of plasma concentration (230), indicating a potential of intrathecal effect in addition to systemic, a trait shared with the more potent cladribine (231). Our study confirmed the previously demonstrated effects of DMF on blood B cells and further demonstrated that the effects also translates into the CSF, as we observed a significant drop in mononuclear cells and relative drop in plasmablasts. The relative number of memory cells were unchanged, but as the total mononuclear count dropped, the absolute count of memory B cells likely was reduced as well, although not as much as the plasmablasts (Paper IV).

Besides memory B cells, plasmablasts and plasma cells are the other main subsets of B cells in CSF of MS patients (232). Studies have shown these are clonally related and persist over time along with oligoclonal bands in the CSF (134, 233, 234). Importantly, the specificity of antibodies, plasma cells and plasmablasts seemed to be not one, but included both viruses (235) and autoantigens related to oligodendrocytes, astrocytes and neurons (236-238), intracellular debris (239) or simply unknown. Even though studies indicate there likely is an antigen-driven response, these findings seem incompatible with there being a single B cell autoantigen that defines MS. Some of the studies could be compatible with the hypothesis of idiotope-driven pathology, driven by autoantigen:BCR ligation, where B cells recognize an autoantigen, while T cells recognize a BCR idiotope (92, 161).

Our understanding of the pathological roles of memory B cells in MS remains incomplete (170, 221). Among several proposed roles in pathological interactions with the CNS (reviewed in (221)), antigen presentation and T-B collaboration are likely key in triggering inflammation. Recently it was shown that memory B cells causes brain homing memory $CD4^+$ T cells to autoproliferate, and that this activation is inhibited in individuals receiving B cell depleting therapies, indicating a dysregulated T-B collaboration in MS patients (240). However, our understanding of the classic T-B collaboration may be flawed, as it may be more complex than how it typically is presented in textbooks (92). Below I will present evidence gathered by us and others on idiotope-presentation and related T cell responses that may improve our understanding of T-B collaboration in MS, but also potentially in other auto-immune diseases with B cells as key players.

9.2 Using bioinformatics to learn about the immunoglobulin variable regions

Throughout this thesis, models and predictions of T-B collaboration are used to predict and describe how these cells interact with each other. In this regard, it is important to remember that these models are products of their training sets, in this case publicly available datasets on HLA affinities (IEDB, (185)), enzyme activities (190) and TCEM frequencies (9, 39). Interpretation should always be mindful of the limits of the training sets, as these are transferred as limits to the model[23]. Good examples are how the cathepsin prediction models does not include all enzymes present in endolysosomal systems[24] (47), that TCEMs were imputed based on presentation on HLA-DR, not –DP or –DQ (39), or that HLA-affinity predictions were limited to the first three principal components of amino acid physiochemical properties for computational reasons. The observed patterns described may disappear, change or become clearer with renewed or more extensive training datasets, and we therefore rely on validating experiments or support in existing literature in order to draw conclusions from the predicted patterns. The models used may be utilized to analyze large quantities of aligned sequences and look for patterns, as we have done in Paper I and III, or single proteins as we did in Paper II. In order to understand the strengths of and weaknesses of the former for analyzing the immunoglobulin variable regions, I'll start by using examples from the literature on the latter.

Anti-drug antibodies (ADA) towards therapeutic mAbs may become an issue in patients treated with repeated infusions, as ADA may neutralize or remove the effect of mAb therapy (241). Such generation of ADA towards mAbs is dependent on T-B collaboration, and the epitopes are likely from the variable regions of the mAbs (242). For some humanized and chimeric mAbs the epitopes were mapped to confirm this (243, 244). In natalizumab, a mAb used for blocking cell extravasation, a single epitope was identified as responsible for ADA responses, and the epitope was localized in the FW2-CDR2 of the light chain using both HLA class II elution experiments and T proliferation assays (243). An important part of the work was use of NetMHCIIspan 3.2 (245), another affinity prediction neural network model also trained using IEDB data (185), however the authors did not include predictions of cathepsin processing. To illustrate some key points, I have used data from Paper II and the publication describing the epitope and utilized ioGenetics predictions models for illustration (Figure 14).

[23] Some limits were also discussed in the Methodological considerations section.
[24] Of particular interest is cathepsin H, an aminopeptidase expressed in B cells.

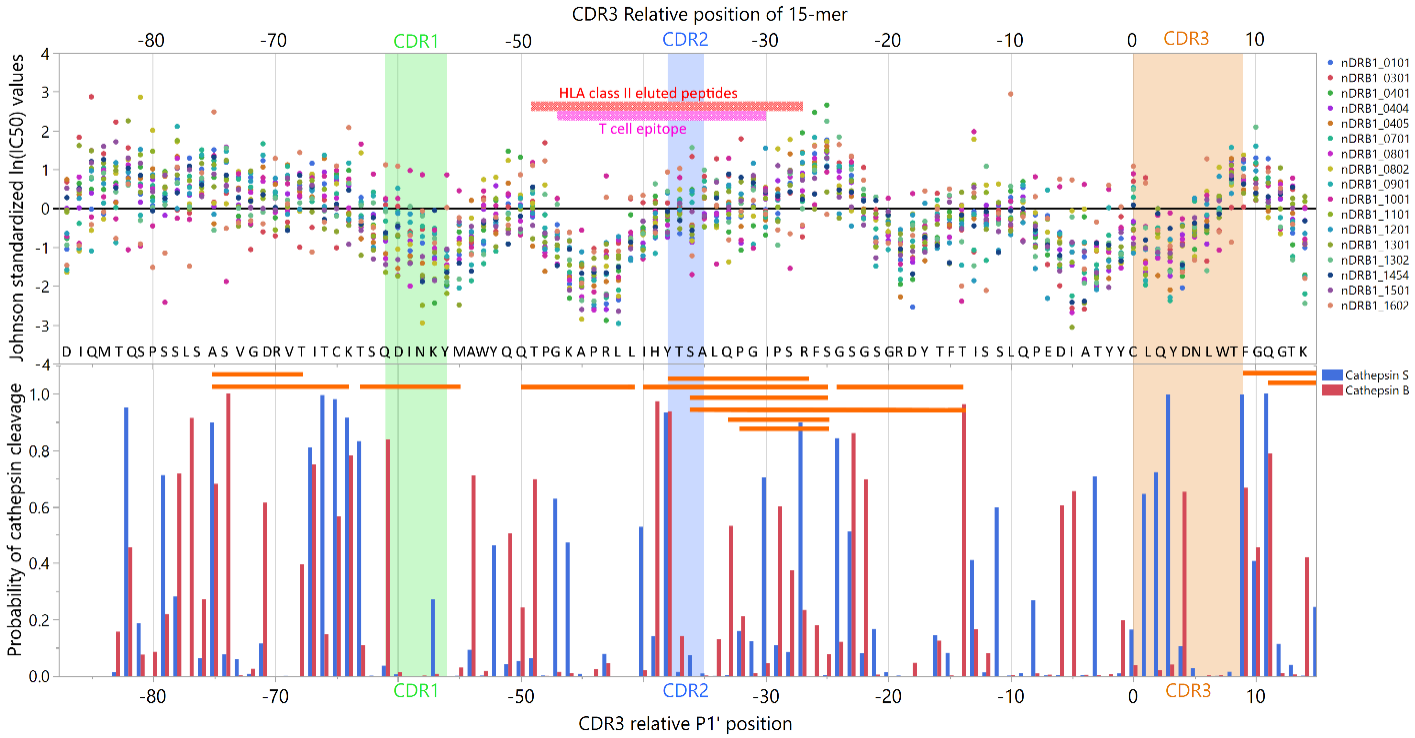


Figure 14. Natalizumab light chain processing, presentation and recognition

Upper panel shows standardized affinity predictions for various HLA-DR alleles for all possible 15-mers in natalizumab light chain (lower value indicates higher affinity). Layered over are the observed presented HLA-class II "consensus" peptide (red) spanning position -49 to -37 of the light chain, and the T cell epitope generating proliferation (purple) (243). Lower panel shows the predicted probabilities of cleavage by cathepsin S or B. Layered over are the unique peptides (orange) found after incubating Cathepsin S with natalizumab at pH 6[25] (Paper II). CDR regions are labelled according to IMGT standards, and the positions given as relative to the cysteine of CDR3.

From this data we may learn several features. One, the light chain of natalizumab has several areas of predicted high affinity and high probability of cleavage, yet only one area was identified upon HLA class II elution experiments (243). Thus, predictions alone cannot tell us all, but perhaps limit investigations to a few areas of interest. Interestingly, areas of higher probability for cleavage "flank" areas of high predicted affinity. Two, cathepsin S alone with natalizumab seemingly destroys the relevant epitope, yet it is presented as whole on natalizumab-specific EBV-transformed B clones after incubation. This may seem confusing, but is in agreement with literature describing how HLA class II binding may protect from epitope destruction (200). Thus, combining affinity and cleavage prediction metrics may be useful for epitope prediction. Three, if we compare these findings to which peptides were found after incubating DCs with infliximab or rituximab, we find the very same pattern of HLA class II presentation, matching prediction patterns, and for rituximab also T cell responsiveness towards epitopes of the same location (244). These examples are only of three different immunoglobulin light chain variable regions. Albeit not human variable regions, one might sense there

[25] No light chain peptides were found at pH 6 after incubation with cathepsin B. Unfortunately, we did not run cathepsin B experiments at pH 4 and 5 for natalizumab.

are certain patterns to how immunoglobulin variable regions are processed and presented, and that these patterns may be picked up by prediction models if we interpret them wisely. Also, as these were mAbs and not BCRs, they may not necessarily reflect how BCRs are processed in B cells.

9.3 On IgG and BCR processing and presentation, *in silico, in vitro* and *in vivo*

In paper II we described how key cysteine cathepsins, known to be present in professional APCs (193), degrade monoclonal IgG in patterns determined by their structure *in vitro*. Due to the high similarity in structure to the BCR (246), we can impute this might be similar for the BCR as well. This makes particular sense upon BCR ligation, as there is no simple explanation to how endosomal cysteine cathepsins or GILT could differentiate between antigen and BCR[26]. However, as our study did not address this directly, so we could not say this for certain. While it is not frequently discussed in literature, there can be little doubt that the BCR indeed can be and is processed by B cells (76, 82, 92). Among the first peptides eluted from MHC class II were peptides derived from constant regions of IgG2a (78). It is clear presentation of immunoglobulin derived peptides occurs constitutively in certain cell lines and B cell malignancies (72, 76, 87, 88) and upon BCR ligation (92), but whether the intracellular location and mode of degradation are the same is not clear.

Available scientific literature may provide some answers. For instance, in a study investigating the HLA class II peptide repertoire after loading DCs with polyclonal intravenous immunoglobulin (IvIg), they found massive presentation of peptides derived from both heavy and light variable regions in particular (247). This indicated endosomal processing of exogenous IgG. In the same experiment, elution was performed using unloaded PBMCs as well, and the authors found preserved presentation of idiotopes from the variable regions[27]. The PBMCs were not sorted, so it's unclear whether the eluted idiotopes came from came from B cells or other APCs (247). Another group utilized lymphomas with completely sequenced heavy and light variable regions and found these also present immunoglobulin variable region peptides in a surprisingly similar manner (87, 88). Others again have eluted HLA class II peptides from APC rich cell populations, and also found idiotope peptides from the very same IGHV locations in cells from the thymus (90), bronchoalveolar lavage samples (91) and synovia of patients with rheumatoid arthritis (248). In the immune epitope database (IEDB), immunoglobulin fragments are a frequent finding in HLA class II elution datasets (185), but CDR3 idiotopes are frequently missing because immunosequencing rarely is performed in parallel.

Our *in silico* predictions and *in vitro* experiments indicate cathepsins S and B in particular may be important for degrading both heavy and light chains of IgG and BCRs. The fact that idiotope peptides from the heavy chain FW3 and light chain FW2-CDR2 locations described above were not frequently found in our peptides dataset after cathepsin cleavage *in vitro* can be explained by how

[26] Occam's razor
[27] From both heavy (FW3) and light chain (FW2-CDR2 area (see section 9.2))

binding to HLA-DR can protect key epitopes from degradation (200). Not surprisingly, our data from Paper I and paper II indicated the areas in question (i.e. heavy chain idiotope peptides starting approx. -20 to -18 amino acids from cysteine of CDR3) are high affinity areas, and also associated with predicted cathepsin cleavage sites at both the amino-and carboxy-ends (Figure 15), thus matching the observations of presented idiotopes described above.

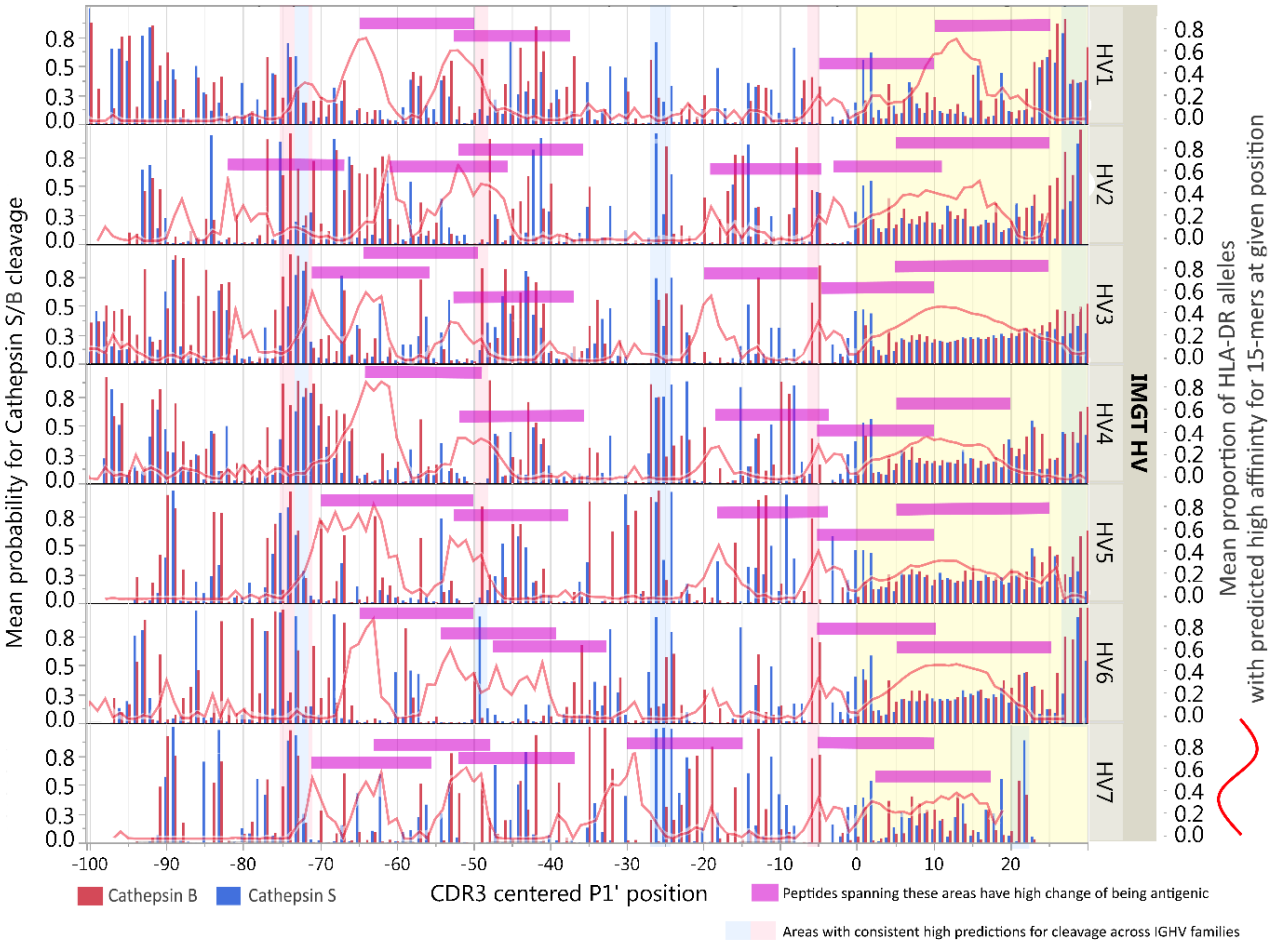


Figure 15. Areas with high affinity and high probability of cleavage in IGHV.
Mean predicted probability for cleavage by cathepsins S and B (bars, left y-axis), overlayered with the proportion of HLA-DR alleles with predicted high affinity binders (red line, right y-axis) at the given CDR3-relative position (x-axis). Areas with conserved high prediction for cleavage are indicated vertically. Areas with conserved high affinity peptides are indicated horizontally. The figure was first presented on the poster *"Immunogenicity of immunoglobulin variable regions",* by Høglund et al. at the 27th European Charcot meeting, and is based on data from Paper II (174) and Bremel and Homan (39).

In paper III, three of the idiotope peptides generating memory T cell responses spanned this specific FW3 area, indicating T cells are exposed to these sequences in vivo[28]. Based on the available HLA class II elution datasets, it seems endogenous and exogenous IgG are degraded and presented similarly, and the pattern of degradation is compatible with the predicted roles of HLA class II binding and available cathepsin processing predictions. These findings feed into the question already asked

[28] Although, cross reactivity cannot be excluded.

many years ago: When the BCR is degraded and presented *in vivo*, how is the T-B collaboration regulated to withstand such presentation (74, 82, 83, 85, 249)?

9.4 A question of tolerance

Presentation of IGHV derived idiotopes seemingly poses a regulatory "problem" for the immune system (74, 82, 83, 249). The results from papers I-III, considering the other research described, suggest it is a common occurrence. Yet not all humans develop B cell intolerance, thus the immune system must have found a way out of the conundrum. It was hypothesized that germline sequences would be associated with T cell tolerance (reviewed in (249)). This is compatible with theories of central tolerance (35), as both thymic B cells, and DCs that potentially present circulating immunoglobulin is present (22, 33). Because mTECs do not undergo VDJ recombination[29], they likely do not participate in tolerization towards immunoglobulin variable regions. With this in mind, T cell tolerization is likely boosted due to placental transfer of IgG from maternal blood (250). Even though the naïve IGHV repertoire is vast, the presented IGHV peptide-repertoire would effectively be limited to areas of high affinity to the individual's HLA class II molecules and corresponding cathepsin activity (Figure 15). Thus, presented idiotope peptides from germline (non-CDR3) variable regions, that effectively have low TCEM FC values (39), would be associated with tolerance. As we could identify memory T cell responses towards such idiotope peptides (low FC, but 3 out of 5 were mutated FW3 peptides) in paper III, we can speculate that deletion as a central tolerance mechanism is not complete, and that perhaps thymus derived or peripherally induced Tregs are important for tolerance towards IGVH. In line with this, FW3 idiotope peptides were eluted from HLA class II molecules from human thymus (90), but in transgenic mice clonal deletion of idiotope-specific T cells was only achieved after introducing high doses of monoclonal IgG containing a well described idiotope (81). If we apply the avidity model of central tolerance (35), the diversity of IGHV combined with the constant presentation of certain IGHV regions by a limited thymic cell repertoire, is perhaps enough to generate a low to intermediated density of pHLA presentation of IGHV, potentially causing some degree of thymic Treg induction.

What about the CDR3 region then? Our data from paper I, which confirmed findings in an independent dataset (39), indicated that the CDR3 region has consistently a predicted high affinity for HLA-DR and -DP alleles in particular (see Supplementary Figure S6 in paper I). This is noteworthy by itself, as the randomness of VDJ recombination and somatic mutations imply affinity should stay around the average. For HLA-A and -B alleles the CDR3 idiotope affinities were predicted so, but for HLA-DR it was consistently predicted to be high. We have not analyzed the overall impact of mutations on this metric, but in Paper III we suggest HLA-affinity increase may have been a factor

[29] Not to our knowledge.

influencing mutations in idiotope peptides with memory T cell responses[30]. Due to the high variation of CDR3 we cannot expect a similar tolerization towards idiotope peptides derived from this region, as is evident also from results in Paper III where most of the memory T cell responses observed were towards CDR3-derived idiotope peptides. *In vivo* however, this does not necessarily translate to intolerance towards the full immunoglobulin molecule, as multiple peptides from the same IGHV may be released in presented simultaneously upon degradation (87, 88). Thus, potentially stimulatory CDR3 fragments may have to compete for HLA binding with other idiotope peptides from high affinity regions (Paper I and Figure 15). If the B cells present idiotope peptides, the presence of tolerogenic idiotope peptides may be enough to quench a pathogenic immune response towards CDR3 idiotopes *in vivo*.

9.5 The IGHV as a regulatory unit

Viewing immunoglobulins as regulatory for the immune system in nothing new (251). Immunoglobulins or BCRs may contribute by diversifying the TCR repertoire (252). Use of IvIg[31] in various inflammatory disorders have been shown to induce expansion of peripheral pTregs[32] *in vivo* (253) and *in vitro (254, 255)* and in mouse models of inflammatory diseases (256-258). Several mechanisms of action have been proposed[33], including various pathways of inducing pTregs (259). Some have suggested that the mechanism is simply impairing antigen presenting capacity of APCs, as IvIg potentially outcompete other antigens for presentation (260), but that does not explain the expansion of Tregs. Of particular interest is how one group suggested the Treg induction may be caused by epitopes in the constant region, namely Tregitopes[34] (261, 262), constant region derived peptides predicted and confirmed *in vitro* to have high affinity for human HLA-DR alleles and to activate Tregs (261). The term has later been expanded to apply to epitopes from all parts of IgGs (262), and it was suggested that Tregitopes indeed is the active ingredient of IvIg (263). However, incubating human DCs with IvIg causes massive presentation of idiotopes from the variable regions and only rarely the originally described Tregitopes (247), seemingly casting doubt whether constant region Tregitopes are the effectors in humans. In this context, both B cells that are cognate to foreign IgG and DCs that accumulate IgG through Fc-receptor uptake can participate in idiotope-driven activation of cognate T cells.

As idiotopes from the variable regions are the main presented peptides by DCs with various HLA-DR alleles incubated with IvIg (247), a finding in line with what was otherwise described in sections 9.2 and 9.3, it is also likely idiotope peptides are the main mediators of pTreg expansion

[30] Idiotope-driven clonal selection.
[31] Therapeutic IvIg consists mainly of IgG, with subclass-distribution near natural distribution.
[32] Also known as iTreg – inducible Treg
[33] Not all are discussed here.
[34] Originally described as hTregitope 289 and hTregitope 167, based on their location in the sequence.

observed. These observations are consistent with what has previously been predicted for IGHV peptide-TCR interactions (39). In line with this, it was found that variable region idiotopes, including two from the region spanning -20 to -4 before CDR3, were among the peptides inducing the largest variance in $CD4^+$ T cell proliferation assays upon tetanus toxoid stimulation (247). Suppression was not significantly different from the control, while 40 nmol/mL IvIg was. However, the relevant idiotope peptides were only tested at 4 nmol/mL and IvIg at 4 nmol/mL was not significantly suppressive either (247). Interestingly, the authors also suggested the lack of constant region Tregitope presentation upon endosomal degradation, could be due to overly active degradation of this region, and this is indeed what we found when we examined how IgG is cleaved by cathepsin S, L and B at various pH (Paper II), cathepsins which also are expressed in DCs (47, 193, 194).

Findings that were shown in Paper I, but not discussed, were changes to TCEM FC and HLA-affinity in FW3. From Figure 8 in Paper I, the FC of MS patient FW3 region idiotopes on average seems higher than OIND, and the differences are higher than what was seen in CDR3. Likewise, differences between blood and CSF IGHV FW3 idiotopes are higher in MS patients than in OIND, when it comes to TCEM FC (Figure 7 in paper I). FC is associated with mutations, as rarer variations introduced by changes causes increase in FC. Thus, we impute mutations are causing the rise in FC, which again is consistent with previous investigations into mutations in these particular sequences (140). Surprisingly, the proportions of predicted high affinity idiotopes for HLA-DR alleles at positions -20 and -5 relative to cysteine of CDR3 were predicted substantially lowered in MS patients. Further subgrouping showed that this difference was not driven entirely by differences in IGHV gene family usage, but was present for both IGHV3 and IGHV4, the families dominating the repertoires (Figure 16). This might suggest that mutations in the FW3 region could cause loss of affinity towards HLA-DR alleles, thereby reducing a possible regulatory response as described above. Somewhat conflicting with this idea was how some of the idiotope peptides with T cell responses in paper III were indeed from the FW3, but these must have been among the sequences with retained high affinity for HLA-DR[35]. Whether this phenomenon is specific for MS or any antigen-driven response is not clear and should be evaluated further.

[35] Because selection criteria were predicted -1.5 Johnson standardized affinity or higher affinity.

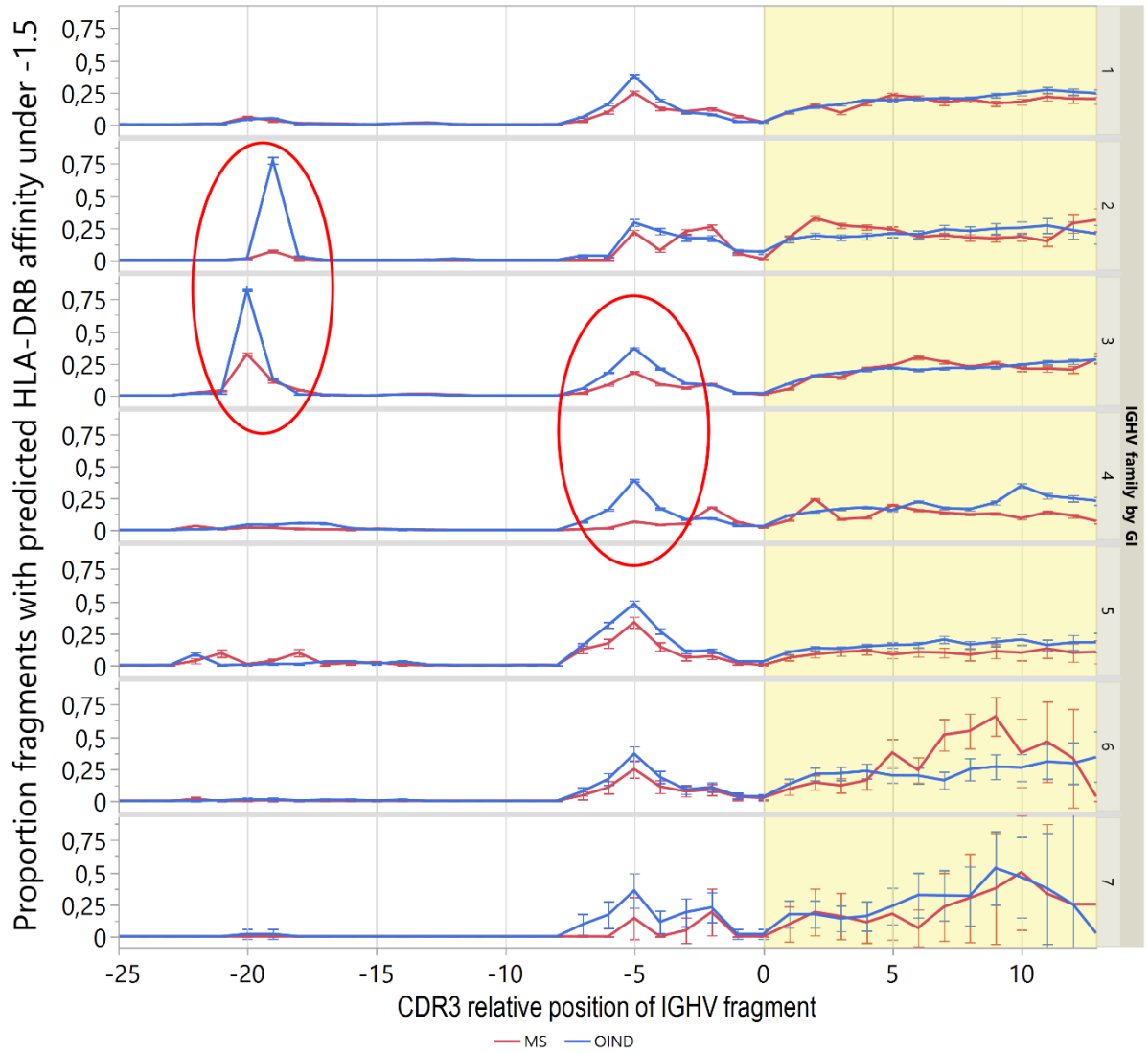


Figure 16. Predicted HLA-DR affinity for CSF derived idiotopes in MS and OIND patients
Proportion of IGHV idiotopes at each given CDR3-relative position with a Johnson SI standardized predicted affinity for patient specific HLA-DR alleles below -1.5 (high affinity binders). Data are the same as HLA-DR data from Figure 10 of paper I (93), but split by sequence IGHV family. Red circles mark areas of notable difference. Error bars indicate 95% confidence intervals, no formal mixed model test of FW3 region differences between MS and OIND have been performed.

9.6 Reevaluation of idiotope-driven T-B collaboration hypothesis in multiple sclerosis

The original hypothesis of idiotope-driven T-B collaboration in MS was rather simple: memory B cells (having undergone affinity maturation) enter the CNS and randomly encounter idiotope-specific T cells that provide help upon interacting with idiotopes presented on HLA class II molecules, potentially generating an inflammatory response (157-161). Potentially immunogenic idiotopes could have been generated during previous affinity maturation processes, during classic T-B collaboration.

In paper I we described how idiotopes from the IGHV CDR3 region were highly likely to be antigenic in MS patients, consistent with the hypothesis. However, the same basic pattern was also observed for patients with OIND, and high affinity of CDR3 idiotopes for HLA-DR alleles had also been established in an unrelated dataset (39). We have further shown that predicted affinity and cleavage patterns in FW3 also show a consistent pattern (Figure 15). This may indicate the increase in affinity may be a general phenomenon, and not necessarily related to MS. While previous studies indicated CSF IgG from MS patients triggered T cells to proliferate more than in OIND patients (157) and that idiotope peptides from these could induce oligodendrocyte death (160), it could still be memory $CD4^+$ T cell responses towards IGHV idiotope peptides like those observed in Paper III, could also be found in OIND or other patients. The presence of idiotope-specific T cells does not necessarily translate to *in vivo* immunogenicity. Still, based on newer knowledge of how IgG is processed and presented in endolysosomal systems (247) (Paper II) and how the T cell responsiveness were mapped to specific idiotopes on CSF IgGs (158, 159), we can argue the observed proliferation likely was due to presentation of idiotope peptides.

The original hypothesis further argued that immunogenic idiotopes may result from mutations causing increased affinity, which is consistent with our observation in paper III where most idiotope peptides that generated memory T cell responses were associated with mutations with a predicted positive effect on HLA-DR presentation. It is also consistent with predictions made about the immunogenicity of the CDR3 region in paper I. However, the biggest prediction differences between MS and OIND patients were not observed in CDR3, but rather in FW3 (Figure 16). If we consider the potential regulatory role of IGHV, as described above, it could also be that mutations in FW3 causes loss of affinity for HLA-DR, thereby also potentially causing a loss of regulatory potential in the same IGHV sequences. Such a loss could clear the way for presentation of immunogenic idiotopes from the CDR3 region, and provide an explanation for the previously observed differences in observed T cell proliferation upon stimulation with CSF IgG (157), and perhaps also generally impaired Treg responses observed in MS patients (264). It could also provide a link to EBV infection, as EBV is known to introduce mutations into infected B cells (161, 265). Perhaps mutations into FW3 regions reduce regulatory control of EBV infected memory B cells, thereby creating an immune evasion mechanisms for the virus, while simultaneously generating a susceptibility to auto-immune diseases like MS and SLE (152). For humans such a mechanism could lead to improved responses to infectious disease and an evolutionary advantage, forming a human-virus symbiosis. This could also serve as an idiotope-extension of Pender's hypothesis[36] (151), and a systematic investigation into EBV mutations' influence on prediction metrics could further illuminate this.

[36] See section 5.8

Genomic risk studies support antigen-presenting functions are associated with MS risk, and the T and B cells involved are enriched for these genes (119). However, due to incomplete knowledge of genomic variation within the IGHV genes, the GWAS studies may not fully have elucidated risk contribution of this region (266). If indeed idiotope presentation is always part of conventional T-B collaboration, as suggested recently (92), this may be a huge hole in knowledge of genetic risk of MS. Genetic interactions on protein level may apply, as we found that MS risk allele HLA-DRB1 15*01 were among the DR alleles with highest predicted affinity for CSF derived IGHV idiotopes in general (93). It is known that some IGHV genes are associated with more effective binding to certain microbial antigens, thereby being influenced by evolutionary pressure (266). In a similar manner it is not unlikely that some IGHV genes may be associated with CNS auto-antigen affinity in MS, as certain IGHV genes are associated with other auto-immune disease (267, 268). Idiotope-driven T-B collaboration could provide a bridge between HLA class II presenting, auto-reactive B cells that are driven by idiotope-specific T cells, an unlinked immune response, as has been seen in mouse models (84, 86). Somewhat contradicting possible IGHV gene risk association is the finding that central B cell tolerance mechanisms seem intact in MS, and that it is instead defects in peripheral tolerance that generates auto-reactive B cells (264). This last finding is also in line with the observed clinical response to B cell depleting agents, as it is mainly the memory repertoire that remains depleted long term, while new naïve cells enter the blood much quicker (168). The studies presented herein therefore does not sufficiently explain the reason for a CNS location of the idiotope-driven T-B collaboration, but the tools provided may help in investigating this further.

9.7 A revised T-B collaboration to explain multiple sclerosis and other autoimmune diseases

In light of evolutionary theory, consistent traits observed in biology should have a reasonable explanation in evolution. Selection of HLA-types and likely also IGHV genes are results of generations of selection pressure by infections mainly (269). The predicted interactions between HLA class II molecules and IGHV regions in endolysosomal systems, described by Homan and Bremel (39), and re-evaluated in this thesis suggest a possible co-evolution of these gene families due to idiotope-driven T-B collaboration. Based on the observations described above, including a premise of idiotopes being presented in context of BCR ligation and T-B collaboration (92), I postulate the following regarding idiotope-driven T-B collaboration in context of B-cell mediated auto-immune diseases:

1. **The IGHV CDR3 is always immunogenic**[37]
   - Tolerance may occur, but the immune system is not built for CDR3 tolerance (249).

[37] These premises have been discussed previously by others than me (249).

2. **The IGHV germline non-CDR3 is always tolerogenic[37].**
   - Tolerance is more likely, non-CDR3 IGHV has predicted preserved high affinity areas and cathepsin cleavage spots, as well as more preserved TCEM.
   -The preserved areas may be due to evolution, possibly to ensure IGHV tolerance and may be enough to generate central and/or peripheral tolerance.
3. **If a B cell is exposed to foreign antigen, the combined immunogenicity of CDR3 and the foreign antigen causes adequate T cell responses, thereby maintaining dual memory.**
   - T cells should not be tolerant to foreign antigens. Upon ligation, the BCR is degraded and presented along with the antigen (92).
   - Given 1) T cells may respond to both.
   - Given 2) some of these peptides may also initiate counterbalancing idiotope-specific Treg responses.
4. **If a B cell is exposed to self-antigen, the combined tolerogenicity of non-CDR3 IGHV and self-antigen is enough to quench an immune response in healthy individuals.**
   - T cells should be tolerant to self-antigens. Central tolerance may cause clonal deletion of self-antigen specific T cells (35).
   - Tregs specific for either self-antigen or non-CDR3 IGHV could in healthy individuals quench a pathogenic immune response, despite immunogenicity of CDR3.
5. **Balance in 4) may be shifted due to unfortunate changes in non-CDR3 IGHV.**
   - If HLA class affinity for non-CDR3 IGHV idiotopes is lost or reduced, responses given 3) and 4) may be unbalanced.
   - Mutations also may cause loss of T cell tolerance upon HLA-presentation[37].
   - Heritable traits (i.e. certain IGHV and HLA gene combinations) may contribute.
   - Unbalanced responses may potentially cause auto-immune disease.

Our findings support ways T cells may be largely tolerant towards the naïve repertoire of IGHV, as some defined IGHV regions are predicted to always be presented, allowing for idiotope-specific regulatory T cells to develop (33, 35, 249), or perhaps be deleted (81, 270). However, it also opens up possibilities how loss of tolerance may occur. For instance: Introduction of mutations that change affinity and/or cathepsins activity may shift or suppress regulatory T cell activity. Mutations in CDR3 may introduce new epitopes, with higher affinity for HLA, outcompeting regulatory ligands, potentially causing autoimmunity, as predicted in paper I and followed up on in paper III. Of note, the

discovery of $CCR6^+$ idiotope-specific T cells[38] for FW3 idiotope peptides in paper III does not falsify this, as $CCR6^+$ Tregs have also been described (271).

If we consider the postulates a joint hypothesis, it should also be testable (272). While the evidence described previously both predict and show that IGHV idiotopes are presented frequently by B cells and other APCs, it has not been proven definitely that this "always" occurs during T-B collaboration. This would be an important first step and is potentially falsifiable. However, doing so would be a monumental effort, as it theoretically would require simultaneous single cell sequencing and immunopeptidomic analysis of B cells extracted from relevant tissues. With time and technology improvements we may eventually be capable. Given consistent presentation, the next step would be to identify idiotope-specific pro-inflammatory $CD4^+$ T cells and Tregs in healthy individuals (Postulates 1) and 2)), which should be achievable with current technology (273, 274). Finally, 3), 4) and 5) could be investigated using bioengineered BCRs/IgGs containing alterations in non-CDR3 or CDR3 that suppress presentation presumed regulatory or stimulatory idiotopes. Such alterations may prove difficult to introduce experimentally into BCRs, as changes may change stability or affinity of the BCR, thereby prohibiting ligation and efficient presentation.

Understanding these mechanisms is likely important for our understanding of autoimmune diseases' immunopathology. In MS specifically, it seems all the prerequisites for idiotope-driven T-B collaboration are present, including predicted cleavage and high affinity of CDR3 associated epitopes (Paper I), presence of idiotope-specific T cells in the periphery (Paper III), and mutated non-CDR3 regions that reduce predicted affinity of potentially regulatory idiotopes. The disease may therefore continue to serve as a model to idiotope-driven T-B collaboration further, but other autoimmune diseases should be probed as well. In addition, investigating other species than mice and humans for possible idiotope-driven mechanisms preserved through evolution might be a feasible approach to identify undescribed mechanisms.

10 Conclusion

The work of this thesis confirms it is possible to predict antigenic idiotopes from human IGHV using high-throughput immunosequencing combined with HLA-affinity and cathepsin cleavage neural network prediction models. In doing so it also further supports the hypothesis that a dysregulated idiotope-driven T-B collaboration is occurring in the CNS of MS patients, which is consistent with current knowledge of genetic and environmental risk for MS and also current insights into the most relevant immune cells and how these operate in MS. Following these findings, the thesis also encourages further investigation into general idiotope-driven T-B collaboration.

[38] Frequently, $CCR6^+$ $CD4^+$ T cells are associated with at Th17 phenotype.

11 Future perspectives

Idiotope-driven T-B collaboration may be more extensive and impact the immune system in more ways than what comes across in classic immunology textbooks. But having read this thesis, perhaps it is easier to accept the existence of idiotope-specific T cells *in vivo*, considering the combination of affinity predictions, cathepsin processing and antigenic attributes that IGHV possess, as described above and in Papers I-III. To some, all this may seem unreasonably complex. But consider this: Idiotope-presentation during T-B collaboration is indeed a simple explanation; what else could happen to BCRs in the endosomal system during B cell activation?

Importantly the work described herein only uses approximated models for antigen presentation. In particular the list of enzymes involved could and should be extended upon future work. Others have utilized minimal models for endosomal systems, limited to cathepsins S, B and H, as well as HLA-DR and –DM (200), which could be a great example to follow for investigating degradation of IgG and BCRs further. In such a system HLA binding would provide protection from cleavage, and potentially confirm the areas of IGHV with likely presentation, after endosomal processing. Such a model could also allow for specific inhibiting agents to be utilized, that would otherwise not be usable in a cellular system, and also the system could allow generation of datasets for improved training of neural network epitope predictors.

The described patterns of increased affinity in FW3 and CDR3 of IGHV should also be investigated further, to perhaps support the hypothesis of idiotope-affinity as a factor in T-B collaboration. For instance, datasets utilizing known HLA class II types in conjunction with presorted naïve and memory immune repertoires (9, 275), could provide materials to analyze differences between naïve and memory sets *in silico.* Likewise, any affinity changes in other parts of IGHV should be investigated in light of potential regulatory potential.

The patterns described may also be of interest to pharmaceutical companies and researcher developing therapeutic mAbs, as assessing and amending immunogenicity based on these at early stages using *in silico* methods of mAb variable regions perhaps may contribute to cheaper development, better long-term treatment responses and perseverance. Knowledge presented here, of which regions are presented consistently and how mAbs are degraded is therefore useful for therapeutic mAb development.

It remains to be demonstrated that idiotope-driven T-B collaboration occurs within the CNS in MS patients. If technological advancements and tissue availability allows it, this should be investigated in perivascular and meningeal T and B cell aggregates seen in MS. Also, the role of EBV as a potential idiotope deregulator, introducing mutations into both non-CDR3 and CDR3 IGHV regions should be investigated.

As previously described, the search for idiotopes in patients in this project was limited by the small area of IGHV sequenced in the original dataset. In any attempt to replicate or expand upon this work colleagues should utilize datasets including full BCRs (heavy and light chains), as can be achieved from single cell sequencing (276). Identification of any T cells responsive to idiotopes, should be followed up by tracking of these T cells into the CNS using TCR sequencing technology (141) to confirm the potential of these interacting. The CD154 activation assay allows sorting of reactive T cells in flow cytometry cell sorters, thus providing a population of cells to sequence from PBMC, and to search for in the CSF[39]. Finally, a systematic approach to identify possible Tregs specific for IGHV framework regions should be attempted.

[39] Alternatively: HLA class II tetramers could also be used similarly.

12 References

1. Cooper MD, Alder MN. The Evolution of Adaptive Immune Systems. Cell. 2006;124(4):815-22. doi:10.1016/j.cell.2006.02.001.

2. Boehm T. Evolution of Vertebrate Immunity. Current Biology. 2012;22(17):R722-R32. doi:10.1016/j.cub.2012.07.003.

3. Chaplin DD. Overview of the immune response. Journal of Allergy and Clinical Immunology. 2010;125(2, Supplement 2):S3-S23. doi:10.1016/j.jaci.2009.12.980.

4. Abbas AK, Lichtman AHH, Pillai S. Cellular and Molecular Immunology E-Book. 9th ed. Philadephia, USA: Elsevier Health Sciences; 2017. 600 p.

5. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302(5909):575-81. doi:10.1038/302575a0.

6. Schatz DG, Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nature Reviews Immunology. 2011;11:251. doi:10.1038/nri2941.

7. Desiderio SV, Yancopoulos GD, Paskind M, Thomas E, Boss MA, Landau N, et al. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature. 1984;311(5988):752-5. doi:10.1038/311752a0.

8. Calis JJA, Rosenberg BR. Characterizing immune repertoires by high throughput sequencing: strategies and applications. Trends in immunology. 2014;35(12):581-90. doi:10.1016/j.it.2014.09.004.

9. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. A Public Database of Memory and Naive B-Cell Receptor Sequences. PloS one. 2016;11(8):e0160853. doi:10.1371/journal.pone.0160853.

10. Morbach H, Eichhorn EM, Liese JG, Girschick HJ. Reference values for B cell subpopulations from infancy to adulthood. Clinical & Experimental Immunology. 2010;162(2):271-9. doi:10.1111/j.1365-2249.2010.04206.x.

11. Nemazee D. Mechanisms of central tolerance for B cells. Nature Reviews Immunology. 2017;17:281. doi:10.1038/nri.2017.19.

12. Tiegs SL, Russell DM, Nemazee D. Receptor editing in self-reactive bone marrow B cells. The Journal of experimental medicine. 1993;177(4):1009-20. doi:10.1084/jem.177.4.1009.

13. Gay D, Saunders T, Camper S, Weigert M. Receptor editing: an approach by autoreactive B cells to escape tolerance. The Journal of experimental medicine. 1993;177(4):999-1008. doi:10.1084/jem.177.4.999.

14. Nemazee DA, Burki K. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. Nature. 1989;337(6207):562-6. doi:10.1038/337562a0.

15. Halverson R, Torres RM, Pelanda R. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. Nat Immunol. 2004;5(6):645-50. doi:10.1038/ni1076.

16. Wardemann H, Nussenzweig MC. B-Cell Self-Tolerance in Humans. Advances in Immunology. 95: Academic Press; 2007. p. 83-110.

17. Jackson K, Kidd M, Wang Y, Collins A. The Shape of the Lymphocyte Receptor Repertoire: Lessons from the B Cell Receptor. Frontiers in immunology. 2013;4(263)doi:10.3389/fimmu.2013.00263.

18. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. Journal of immunology (Baltimore, Md : 1950). 2010;184(12):6986-92. doi:10.4049/jimmunol.1000445.

19. Gauss GH, Lieber MR. Mechanistic constraints on diversity in human V(D)J recombination. Mol Cell Biol. 1996;16(1):258-69. doi:10.1128/mcb.16.1.258.

20. Rechavi E, Somech R. Survival of the fetus: fetal B and T cell receptor repertoire development. Semin Immunopathol. 2017;39(6):577-83. doi:10.1007/s00281-017-0626-0.

21. Schroeder HW, Jr., Wang JY. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(16):6146-50. doi:10.1073/pnas.87.16.6146.

22. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nature reviews Immunology. 2014;14(6):377-91. doi:10.1038/nri3667.

23. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, et al. Regulation of CD8+ T cell development by thymus-specific proteasomes. Science (New York, NY). 2007;316(5829):1349-53. doi:10.1126/science.1141915.

24. Tomaru U, Ishizu A, Murata S, Miyatake Y, Suzuki S, Takahashi S, et al. Exclusive expression of proteasome subunit {beta}5t in the human thymic cortex. Blood. 2009;113(21):5186-91. doi:10.1182/blood-2008-11-187633.

25. Bromme D, Li Z, Barnes M, Mehler E. Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization. Biochemistry. 1999;38(8):2377-85. doi:10.1021/bi982175f.

26. Tolosa E, Li W, Yasuda Y, Wienhold W, Denzin LK, Lautwein A, et al. Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. The Journal of clinical investigation. 2003;112(4):517-26. doi:10.1172/jci18028.

27. Surh CD, Sprent J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. Nature. 1994;372(6501):100-3. doi:10.1038/372100a0.

28. Daley SR, Hu DY, Goodnow CC. Helios marks strongly autoreactive CD4+ T cells in two major waves of thymic deletion distinguished by induction of PD-1 or NF-kappaB. The Journal of experimental medicine. 2013;210(2):269-85. doi:10.1084/jem.20121458.

29. Stritesky GL, Xing Y, Erickson JR, Kalekar LA, Wang X, Mueller DL, et al. Murine thymic selection quantified using a unique method to capture deleted T cells. Proceedings of the National Academy of Sciences. 2013;110(12):4679-84. doi:10.1073/pnas.1217532110.

30. Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nature Immunology. 2001;2(11):1032-9. doi:10.1038/ni723.

31. Klein L, Hinterberger M, von Rohrscheidt J, Aichinger M. Autonomous versus dendritic cell-dependent contributions of medullary thymic epithelial cells to central tolerance. Trends in immunology. 2011;32(5):188-93. doi:10.1016/j.it.2011.03.002.

32. Bonasio R, Scimone ML, Schaerli P, Grabie N, Lichtman AH, von Andrian UH. Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. Nat Immunol. 2006;7(10):1092-100. doi:10.1038/ni1385.

33. Yamano T, Nedjic J, Hinterberger M, Steinert M, Koser S, Pinto S, et al. Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. Immunity. 2015;42(6):1048-61. doi:10.1016/j.immuni.2015.05.013.

34. Frommer F, Waisman A. B Cells Participate in Thymic Negative Selection of Murine Auto-reactive CD4+ T Cells. PloS one. 2010;5(10):e15372. doi:10.1371/journal.pone.0015372.

35. Klein L, Robey EA, Hsieh C-S. Central CD4+ T cell tolerance: deletion versus regulatory T cell differentiation. Nature Reviews Immunology. 2019;19(1):7-18. doi:10.1038/s41577-018-0083-6.

36. Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunology Today. 1998;19(9):395-404. doi:10.1016/S0167-5699(98)01299-7.

37. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A Direct Estimate of the Human αβ T Cell Receptor Diversity. Science (New York, NY). 1999;286(5441):958-61. doi:10.1126/science.286.5441.958.

38. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. Blood. 2009;114(19):4099-107. doi:10.1182/blood-2009-04-217604.

39. Bremel RD, Homan EJ. Frequency Patterns of T-Cell Exposed Amino Acid Motifs in Immunoglobulin Heavy Chain Peptides Presented by MHCs. Frontiers in immunology. 2014;5:541. doi:10.3389/fimmu.2014.00541.

40. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. Annual review of immunology. 2006;24:419-66. doi:10.1146/annurev.immunol.23.021704.115658.

41. Calis JJ, de Boer RJ, Kesmir C. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. PLoS computational biology. 2012;8(3):e1002412. doi:10.1371/journal.pcbi.1002412.

42. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. Cell. 1995;80(5):695-705. doi:10.1016/0092-8674(95)90348-8.

43. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. J Biol Chem. 2012;287(2):1168-77. doi:10.1074/jbc.M111.289488.

44. Riedhammer C, Weissert R. Antigen Presentation, Autoantigens, and Immune Regulation in Multiple Sclerosis and Other Autoimmune Diseases. Frontiers in immunology. 2015;6(322)doi:10.3389/fimmu.2015.00322.

45. Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, et al. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. The Journal of experimental medicine. 2005;201(2):233-40. doi:10.1084/jem.20041257.

46. Dhamne C, Chung Y, Alousi A, Cooper L, Tran D. Peripheral and Thymic Foxp3+ Regulatory T Cells in Search of Origin, Distinction, and Function. Frontiers in immunology. 2013;4(253)doi:10.3389/fimmu.2013.00253.

47. Adler LN, Jiang W, Bhamidipati K, Millican M, Macaubas C, Hung SC, et al. The Other Function: Class II-Restricted Antigen Presentation by B Cells. Frontiers in immunology. 2017;8:319. doi:10.3389/fimmu.2017.00319.

48. Yuseff MI, Pierobon P, Reversat A, Lennon-Dumenil AM. How B cells capture, process and present antigens: a crucial role for cell polarity. Nature reviews Immunology. 2013;13(7):475-86. doi:10.1038/nri3469.

49. Lanzavecchia A. Antigen-specific interaction between T and B cells. Nature. 1985;314(6011):537-9. doi:10.1038/314537a0.

50. Mitchison NA. T-cell–B-cell cooperation. Nature Reviews Immunology. 2004;4(4):308-12. doi:10.1038/nri1334.

51. Burnet FMS. The clonal selection theory of acquired immunity. Nashville: Vanderbilt University Press; 1959.

52. Cohn M, Av Mitchison N, Paul WE, Silverstein AM, Talmage DW, Weigert M. Reflections on the clonal-selection theory. Nature Reviews Immunology. 2007;7(10):823-30. doi:10.1038/nri2177.

53. Yates RM, Hermetter A, Taylor GA, Russell DG. Macrophage Activation Downregulates the Degradative Capacity of the Phagosome. Traffic. 2007;8(3):241-50. doi:10.1111/j.1600-0854.2006.00528.x.

54. Hsing LC, Rudensky AY. The lysosomal cysteine proteases in MHC class II antigen presentation. Immunological reviews. 2005;207:229-41. doi:10.1111/j.0105-2896.2005.00310.x.

55. Neefjes J. CIIV, MIIC and other compartments for MHC class II loading. European journal of immunology. 1999;29(5):1421-5. doi:10.1002/(sici)1521-4141(199905)29:05<1421::Aid-immu1421>3.0.Co;2-c.

56. Jacob J, Kelsoe G, Rajewsky K, Weiss U. Intraclonal generation of antibody mutants in germinal centres. Nature. 1991;354(6352):389-92. doi:10.1038/354389a0.

57. Gitlin AD, Shulman Z, Nussenzweig MC. Clonal selection in the germinal centre by regulated proliferation and hypermutation. Nature. 2014;509:637. doi:10.1038/nature13300.

58. Shulman Z, Gitlin AD, Weinstein JS, Lainez B, Esplugues E, Flavell RA, et al. Dynamic signaling by T follicular helper cells during germinal center B cell selection. Science (New York, NY). 2014;345(6200):1058-62. doi:10.1126/science.1257861.

59. Macpherson AJ, Gatto D, Sainsbury E, Harriman GR, Hengartner H, Zinkernagel RM. A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. Science (New York, NY). 2000;288(5474):2222-6. doi:10.1126/science.288.5474.2222.

60. Vidarsson G, Dekkers G, Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. Frontiers in immunology. 2014;5(520)doi:10.3389/fimmu.2014.00520.

61. Eisen HN, Siskind GW. Variations in Affinities of Antibodies during the Immune Response*. Biochemistry. 1964;3(7):996-1008. doi:10.1021/bi00895a027.

62. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. Cell. 2000;102(5):553-63. doi:10.1016/S0092-8674(00)00078-7.

63. Tomayko MM, Allman D. What B cell memories are made of. Current opinion in immunology. 2019;57:58-64. doi:10.1016/j.coi.2019.01.003.

64. Oudin J, Michel M. [A new allotype form of rabbit serum gamma-globulins, apparently associated with antibody function and specificity]. C R Hebd Seances Acad Sci. 1963;257:805-8.

65. Sirisinha S, Eisen HN. Autoimmune-like antibodies to the ligand-binding sites of myeloma proteins. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(12):3130-5. doi:10.1073/pnas.68.12.3130.

66. Mäkelä O, Karjalainen K. Inherited Immunoglobulin Idiotypes of the Mouse. Immunological reviews. 1977;34(1):119-38. doi:10.1111/j.1600-065X.1977.tb00370.x.

67. Jerne NK. Towards a network theory of the immune system. Ann Immunol (Paris). 1974;125c(1-2):373-89.

68. The Nobel Prize in Physiology or Medicine 1984 [press release]. Nobelprize.org: Nobel Media AB. Accessed:06.11.2019, url:https://www.nobelprize.org/prizes/medicine/1984/press-release/

69. Cohn M. The Wisdom of Hindsight. Annual review of immunology. 1994;12(1):1-62. doi:10.1146/annurev.iy.12.040194.000245.

70. Lemke H. Immune Response Regulation by Antigen Receptors' Clone-Specific Nonself Parts. Frontiers in immunology. 2018;9:1471. doi:10.3389/fimmu.2018.01471.

71. Jorgensen T, Bogen B, Hannestad K. T helper cells recognize an idiotope located on peptide 88-114/117 of the light chain variable domain of an isologous myeloma protein (315). The Journal of experimental medicine. 1983;158(6):2183-8.

72. Bogen B, Malissen B, Haas W. Idiotope-specific T cell clones that recognize syngeneic immunoglobulin fragments in the context of class II molecules. European journal of immunology. 1986;16(11):1373-8. doi:10.1002/eji.1830161110.

73. Bogen B, Jorgensen T, Hannestad K. T helper cell recognition of idiotopes on lambda 2 light chains of M315 and T952: evidence for dependence on somatic mutations in the third hypervariable region. European journal of immunology. 1985;15(3):278-81. doi:10.1002/eji.1830150313.

74. Eyerman MC, Zhang X, Wysocki LJ. T cell recognition and tolerance of antibody diversity. Journal of immunology (Baltimore, Md : 1950). 1996;157(3):1037-46.

75. Eyerman MC, Wysocki L. T cell recognition of somatically-generated Ab diversity. The Journal of Immunology. 1994;152(4):1569-77.

76. Weiss S, Bogen B. B-lymphoma cells process and present their endogenous immunoglobulin to major histocompatibility complex-restricted T cells. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(1):282-6.

77. Snyder CM, Zhang X, Wysocki LJ. Negligible Class II MHC Presentation of B Cell Receptor-Derived Peptides by High Density Resting B Cells. The Journal of Immunology. 2002;168(8):3865-73. doi:10.4049/jimmunol.168.8.3865.

78. Rudensky AY, Preston-Hurlburt P, Al-Ramadi BK, Rothbard J, Janeway CA. Truncation variants of peptides isolated from MHC class II molecules suggest sequence motifs. Nature. 1992;359(6394):429-31.

79. Chicz RM, Urban RG, Gorga JC, Vignali DA, Lane WS, Strominger JL. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. The Journal of experimental medicine. 1993;178(1):27-47. doi:10.1084/jem.178.1.27.

80. Munthe LA, Os A, Zangani M, Bogen B. MHC-restricted Ig V region-driven T-B lymphocyte collaboration: B cell receptor ligation facilitates switch to IgG production. Journal of immunology (Baltimore, Md : 1950). 2004;172(12):7476-84. doi:10.4049/jimmunol.172.12.7476.

81. Bogen B, Dembic Z, Weiss S. Clonal deletion of specific thymocytes by an immunoglobulin idiotype. The EMBO journal. 1993;12(1):357-63.

82. Bogen B, Weiss S. Processing and presentation of idiotypes to MHC-restricted T cells. International reviews of immunology. 1993;10(4):337-55. doi:10.3109/08830189309061709.

83. Zhang X, Smith DS, Guth A, Wysocki LJ. A Receptor Presentation Hypothesis for T Cell Help That Recruits Autoreactive B Cells. The Journal of Immunology. 2001;166(3):1562-71. doi:10.4049/jimmunol.166.3.1562.

84. Aas-Hanssen K, Funderud A, Thompson KM, Bogen B, Munthe LA. Idiotype-specific Th cells support oligoclonal expansion of anti-dsDNA B cells in mice with lupus. Journal of immunology (Baltimore, Md : 1950). 2014;193(6):2691-8. doi:10.4049/jimmunol.1400640.

85. Snyder CM, Aviszus K, Heiser RA, Tonkin DR, Guth AM, Wysocki LJ. Activation and tolerance in CD4(+) T cells reactive to an immunoglobulin variable region. The Journal of experimental medicine. 2004;200(1):1-11. doi:10.1084/jem.20031234.

86. Munthe LA, Corthay A, Os A, Zangani M, Bogen B. Systemic autoimmune disease caused by autoreactive B cells that receive chronic help from Ig V region-specific T cells. Journal of immunology (Baltimore, Md : 1950). 2005;175(4):2391-400. doi:10.4049/jimmunol.175.4.2391.

87. Khodadoust MS, Olsson N, Chen B, Sworder B, Shree T, Liu CL, et al. B-cell lymphomas present immunoglobulin neoantigens. Blood. 2019;133(8):878-81. doi:10.1182/blood-2018-06-845156.

88. Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. Nature. 2017;543(7647):723-7. doi:10.1038/nature21433.

89. Papaioannou D, Strothmeyer A-M, Dühren-von Minden M, Keppler-Hafkemeyer A, Zirlik K, Mikesch K, et al. Evidence for idiotype-directed immunosurveillance is restricted to follicular lymphoma and attributable to somatic hypermutation. Haematologica. 2015;100(4):e143-e6. doi:10.3324/haematol.2014.111252.

90. Collado JA, Alvarez I, Ciudad MT, Espinosa G, Canals F, Pujol-Borrell R, et al. Composition of the HLA-DR-associated human thymus peptidome. European journal of immunology. 2013;43(9):2273-82. doi:10.1002/eji.201243280.

91. Heyder T, Kohler M, Tarasova NK, Haag S, Rutishauser D, Rivera NV, et al. Approach for Identifying Human Leukocyte Antigen (HLA)-DR Bound Peptides from Scarce Clinical Samples. Molecular & cellular proteomics : MCP. 2016;15(9):3017-29. doi:10.1074/mcp.M116.060764.

92. Huszthy PC, Gopalakrishnan RP, Jacobsen JT, Haabeth OAW, Løset GÅ, Braathen R, et al. B cell receptor ligation induces display of V-region peptides on MHC class II molecules to T cells. Proceedings of the National Academy of Sciences. 2019:201902836. doi:10.1073/pnas.1902836116.

93. Høglund RA, Lossius A, Johansen JN, Homan J, Benth JS, Robins H, et al. In Silico Prediction Analysis of Idiotope-Driven T-B Cell Collaboration in Multiple Sclerosis. Frontiers in immunology. 2017;8:1255. doi:10.3389/fimmu.2017.01255.

94. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nature communications. 2013;4:2680. doi:10.1038/ncomms3680.

95. Lossius A, Johansen JN, Vartdal F, Holmøy T. High-throughput sequencing of immune repertoires in multiple sclerosis. Annals of clinical and translational neurology. 2016;3(4):295-306. doi:10.1002/acn3.295.

96. Lemke H, Tanasa RI, Trad A, Lange H. Benefits and burden of the maternally-mediated immunological imprinting. Autoimmunity reviews. 2009;8(5):394-9. doi:10.1016/j.autrev.2008.12.005.

97. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. Nature. 2015;523(7560):337-41. doi:10.1038/nature14432.

98. Aspelund A, Antila S, Proulx ST, Karlsen TV, Karaman S, Detmar M, et al. A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules. The Journal of experimental medicine. 2015;212(7):991-9. doi:10.1084/jem.20142290.

99. Ma Q, Decker Y, Müller A, Ineichen BV, Proulx ST. Clearance of cerebrospinal fluid from the sacral spine through lymphatic vessels. The Journal of experimental medicine. 2019;216(11):2492-502. doi:10.1084/jem.20190351.

100. Prinz M, Priller J. The role of peripheral immune cells in the CNS in steady state and disease. Nature neuroscience. 2017;20(2):136-44. doi:10.1038/nn.4475.

101. Sandrone S, Moreno-Zambrano D, Kipnis J, van Gijn J. A (delayed) history of the brain lymphatic system. Nature medicine. 2019;25(4):538-40. doi:10.1038/s41591-019-0417-3.

102. Reboldi A, Coisne C, Baumjohann D, Benvenuto F, Bottinelli D, Lira S, et al. C-C chemokine receptor 6–regulated entry of TH-17 cells into the CNS through the choroid plexus is required for the initiation of EAE. Nature Immunology. 2009;10(5):514-23. doi:10.1038/ni.1716.

103. Kivisäkk P, Tucky B, Wei T, Campbell JJ, Ransohoff RM. Human cerebrospinal fluid contains CD4+ memory T cells expressing gut- or skin-specific trafficking determinants: relevance for immunotherapy. BMC Immunology. 2006;7(1):14. doi:10.1186/1471-2172-7-14.

104. Svenningsson A, Hansson GK, Andersen O, Andersson R, Patarroyo M, Stemme S. Adhesion molecule expression on cerebrospinal fluid T lymphocytes: evidence for common recruitment mechanisms in multiple sclerosis, aseptic meningitis, and normal controls. Annals of neurology. 1993;34(2):155-61. doi:10.1002/ana.410340210.

105. Ransohoff RM, Engelhardt B. The anatomical and cellular basis of immune surveillance in the central nervous system. Nature reviews Immunology. 2012;12(9):623-35. doi:10.1038/nri3265.

106. Smolders J, Heutinck KM, Fransen NL, Remmerswaal EBM, Hombrink P, Ten Berge IJM, et al. Tissue-resident memory T cells populate the human brain. Nature communications. 2018;9(1):4593. doi:10.1038/s41467-018-07053-9.

107. Compston A, Coles A. Multiple sclerosis. Lancet (London, England). 2008;372(9648):1502-17. doi:10.1016/s0140-6736(08)61620-7.

108. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nature Reviews Disease Primers. 2018;4(1):43. doi:10.1038/s41572-018-0041-4.

109. Berg-Hansen P, Moen SM, Harbo HF, Celius EG. High prevalence and no latitude gradient of multiple sclerosis in Norway. Multiple sclerosis (Houndmills, Basingstoke, England). 2014;20(13):1780-2. doi:10.1177/1352458514525871.

110. Grytten N, Torkildsen Ø, Myhr KM. Time trends in the incidence and prevalence of multiple sclerosis in Norway during eight decades. Acta Neurol Scand. 2015;132(199):29-36. doi:10.1111/ane.12428.

111. Benjaminsen E, Olavsen J, Karlberg M, Alstadhaug KB. Multiple sclerosis in the far north--incidence and prevalence in Nordland County, Norway, 1970-2010. BMC Neurol. 2014;14:226. doi:10.1186/s12883-014-0226-8.

112. Grytten N, Aarseth JH, Lunde HM, Myhr KM. A 60-year follow-up of the incidence and prevalence of multiple sclerosis in Hordaland County, Western Norway. Journal of neurology, neurosurgery, and psychiatry. 2016;87(1):100-5. doi:10.1136/jnnp-2014-309906.

113. Simonsen CS, Edland A, Berg-Hansen P, Celius EG. High prevalence and increasing incidence of multiple sclerosis in the Norwegian county of Buskerud. Acta Neurol Scand. 2017;135(4):412-8. doi:10.1111/ane.12615.

114. Federation MSI [Internet]. Atlas of MS 2013: mapping multiple sclerosis around the world. Mult Scler Int Fed. 2013:1-28. Accessed 01.11.2019, url: https://www.msif.org/wp-content/uploads/2014/09/Atlas-of-MS.pdf

115. Wallin MT, Culpepper WJ, Nichols E, Bhutta ZA, Gebrehiwot TT, Hay SI, et al. Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology. 2019;18(3):269-85. doi:10.1016/S1474-4422(18)30443-5.

116. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, et al. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol. 2011;21(10):718-79. doi:10.1016/j.euroneuro.2011.08.008.

117. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-86. doi:10.1212/wnl.0000000000000560.

118. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. The Lancet Neurology. 2018;17(2):162-73. doi:10.1016/s1474-4422(17)30470-2.

119. Consortium IMSG. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science (New York, NY). 2019;365(6460):eaav7188. doi:10.1126/science.aav7188.

120. Parnell GP, Booth DR. The Multiple Sclerosis (MS) Genetic Risk Factors Indicate both Acquired and Innate Immune Cell Subsets Contribute to MS Pathogenesis and Identify Novel Therapeutic Opportunities. Frontiers in immunology. 2017;8:425. doi:10.3389/fimmu.2017.00425.

121. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9. doi:10.1038/nature10251.

122. Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens. 1991;38(1):1-15. doi:10.1111/j.1399-0039.1991.tb02029.x.

123. Goris A, Pauwels I, Gustavsen MW, van Son B, Hilven K, Bos SD, et al. Genetic variants are major determinants of CSF antibody levels in multiple sclerosis. Brain : a journal of neurology. 2015;138(Pt 3):632-43. doi:10.1093/brain/awu405.

124. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC. Twin concordance and sibling recurrence rates in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(22):12877-82. doi:10.1073/pnas.1932604100.

125. Simpson S, Taylor BV, van der Mei I. The role of epidemiology in MS research: Past successes, current challenges and future potential. Multiple Sclerosis Journal. 2015;21(8):969-77. doi:10.1177/1352458515574896.

126. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nature Reviews Neurology. 2016;13:25. doi:10.1038/nrneurol.2016.187.

127. Machado-Santos J, Saji E, Tröscher AR, Paunovic M, Liblau R, Gabriely G, et al. The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells. Brain : a journal of neurology. 2018;141(7):2066-82. doi:10.1093/brain/awy151.

128. Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, Schmidbauer M, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain : a journal of neurology. 2009;132(5):1175-89. doi:10.1093/brain/awp070.

129. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. Brain pathology (Zurich, Switzerland). 2004;14(2):164-74. doi:10.1111/j.1750-3639.2004.tb00049.x.

130. Magliozzi R, Howell O, Vora A, Serafini B, Nicholas R, Puopolo M, et al. Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. Brain : a journal of neurology. 2007;130(Pt 4):1089-104. doi:10.1093/brain/awm038.

131. Minagar A, Alexander JS. Blood-brain barrier disruption in multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2003;9(6):540-9. doi:10.1191/1352458503ms965oa.

132. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nature reviews Immunology. 2015;15(9):545-58. doi:10.1038/nri3871.

133. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mörk S, Bö L. Axonal Transection in the Lesions of Multiple Sclerosis. New England Journal of Medicine. 1998;338(5):278-85. doi:10.1056/nejm199801293380502.

134. Obermeier B, Lovato L, Mentele R, Bruck W, Forne I, Imhof A, et al. Related B cell clones that populate the CSF and CNS of patients with multiple sclerosis produce CSF immunoglobulin. Journal of neuroimmunology. 2011;233(1-2):245-8. doi:10.1016/j.jneuroim.2011.01.010.

135. Skulina C, Schmidt S, Dornmair K, Babbe H, Roers A, Rajewsky K, et al. Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(8):2428-33. doi:10.1073/pnas.0308689100.

136. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CCA, Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. Journal of Immunology. 1999;163(9):5133-44.

137. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. Science translational medicine. 2014;6(248):248ra107. doi:10.1126/scitranslmed.3008879.

138. von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. The Journal of clinical investigation. 2012;122(12):4533-43. doi:10.1172/jci63842.

139. Palanichamy A, Apeltsin L, Kuo TC, Sirota M, Wang S, Pitts SJ, et al. Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis. Science Translational Medicine. 2014;6(248):248ra106-248ra106. doi:10.1126/scitranslmed.3008930.

140. Johansen JN, Vartdal F, Desmarais C, Tutturen AE, de Souza GA, Lossius A, et al. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: Equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. Clinical immunology (Orlando, Fla). 2015;160(2):211-25. doi:10.1016/j.clim.2015.06.001.

141. Lossius A, Johansen JN, Vartdal F, Robins H, Jurate Saltyte B, Holmøy T, et al. High-throughput sequencing of TCR repertoires in multiple sclerosis reveals intrathecal enrichment of EBV-reactive CD8+ T cells. European journal of immunology. 2014;44(11):3439-52. doi:10.1002/eji.201444662.

142. Hohlfeld R, Dornmair K, Meinl E, Wekerle H. The search for the target antigens of multiple sclerosis, part 1: autoreactive CD4+ T lymphocytes as pathogenic effectors and therapeutic targets. The Lancet Neurology. 2016;15(2):198-209. doi:10.1016/s1474-4422(15)00334-8.

143. Hohlfeld R, Dornmair K, Meinl E, Wekerle H. The search for the target antigens of multiple sclerosis, part 2: CD8+ T cells, B cells, and antibodies in the focus of reverse-translational research. The Lancet Neurology. 2016;15(3):317-31. doi:10.1016/s1474-4422(15)00313-0.

144. Dalmau J, Graus F. Antibody-Mediated Encephalitis. The New England journal of medicine. 2018;378(9):840-51. doi:10.1056/NEJMra1708712.

145. Reindl M, Waters P. Myelin oligodendrocyte glycoprotein antibodies in neurological disease. Nature Reviews Neurology. 2018doi:10.1038/s41582-018-0112-x.

146. Wingerchuk DM, Banwell B, Bennett JL, Cabre P, Carroll W, Chitnis T, et al. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. Neurology. 2015;85(2):177-89. doi:10.1212/WNL.0000000000001729.

147. Küppers R. B cells under influence: transformation of B cells by Epstein–Barr virus. Nature Reviews Immunology. 2003;3(10):801-12. doi:10.1038/nri1201.

148. Peferoen LAN, Lamers F, Lodder LNR, Gerritsen WH, Huitinga I, Melief J, et al. Epstein Barr virus is not a characteristic feature in the central nervous system in established multiple sclerosis. Brain : a journal of neurology. 2010;133(5):e137-e. doi:10.1093/brain/awp296.

149. Serafini B, Rosicarelli B, Veroni C, Mazzola GA, Aloisi F. Epstein-Barr virus-specific CD8 T cells selectively infiltrate the multiple sclerosis brain and interact locally with virus infected cells: clue for a virus-driven immunopathological mechanism. Journal of Virology. 2019:JVI.00980-19. doi:10.1128/jvi.00980-19.

150. Vaughan JH, Riise T, Rhodes GH, Nguyen MD, Barrett-Connor E, Nyland H. An Epstein Barr virus-related cross reactive autoimmune response in multiple sclerosis in Norway. Journal of neuroimmunology. 1996;69(1):95-102. doi:10.1016/0165-5728(96)00069-0.

151. Pender MP. Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases. Trends in immunology. 2003;24(11):584-8.

152. Ascherio A, Munger KL. EBV and Autoimmunity. In: Münz C, editor. Epstein Barr Virus Volume 1: One Herpes Virus: Many Diseases. Cham: Springer International Publishing; 2015. p. 365-85.

153. Williams WM, Staines NA, Muller S, Isenberg DA. Human T cell responses to autoantibody variable region peptides. Lupus. 1995;4(6):464-71. doi:10.1177/096120339500400608.

154. van Schooten WC, Devereux D, Ho CH, Quan J, Aguilar BA, Rust CJ. Joint-derived T cells in rheumatoid arthritis react with self-immunoglobulin heavy chains or immunoglobulin-binding proteins that copurify with immunoglobulin. European journal of immunology. 1994;24(1):93-8. doi:10.1002/eji.1830240115.

155. Peen E, Malone C, Myers C, Williams RC, Jr., Peck AB, Csernok E, et al. Amphipathic variable region heavy chain peptides derived from monoclonal human Wegener's anti-PR3 antibodies stimulate lymphocytes from patients with Wegener's granulomatosis and microscopic polyangiitis. Clinical and experimental immunology. 2001;125(2):323-31.

156. Zangani MM, Frøyland M, Qiu GY, Meza-Zepeda LA, Kutok JL, Thompson KM, et al. Lymphomas can develop from B cells chronically helped by idiotype-specific T cells. The Journal of experimental medicine. 2007;204(5):1181-91. doi:10.1084/jem.20061220.

157. Holmøy T, Vandvik B, Vartdal F. T cells from multiple sclerosis patients recognize immunoglobulin G from cerebrospinal fluid. Multiple sclerosis (Houndmills, Basingstoke, England). 2003;9(3):228-34. doi:10.1191/1352458503ms906oa.

158. Holmøy T, Fredriksen AB, Thompson KM, Hestvik AL, Bogen B, Vartdal F. Cerebrospinal fluid T cell clones from patients with multiple sclerosis: recognition of idiotopes on monoclonal IgG secreted by autologous cerebrospinal fluid B cells. European journal of immunology. 2005;35(6):1786-94. doi:10.1002/eji.200425417.

159. Hestvik AL, Vartdal F, Fredriksen AB, Thompson KM, Kvale EO, Skorstad G, et al. T cells from multiple sclerosis patients recognize multiple epitopes on Self-IgG. Scandinavian journal of immunology. 2007;66(4):393-401. doi:10.1111/j.1365-3083.2007.01955.x.

160. Hestvik AL, Skorstad G, Vartdal F, Holmøy T. Idiotope-specific CD4(+) T cells induce apoptosis of human oligodendrocytes. Journal of autoimmunity. 2009;32(2):125-32. doi:10.1016/j.jaut.2009.01.004.

161. Holmøy T, Vartdal F, Hestvik AL, Munthe L, Bogen B. The idiotype connection: linking infection and multiple sclerosis. Trends in immunology. 2010;31(2):56-62. doi:10.1016/j.it.2009.11.001.

162. Tintore M, Vidal-Jordana A, Sastre-Garriga J. Treatment of multiple sclerosis - success from bench to bedside. Nat Rev Neurol. 2019;15(1):53-8. doi:10.1038/s41582-018-0082-z.

163. Muraro PA, Martin R, Mancardi GL, Nicholas R, Sormani MP, Saccardi R. Autologous haematopoietic stem cell transplantation for treatment of multiple sclerosis. Nature Reviews Neurology. 2017;13:391. doi:10.1038/nrneurol.2017.81.

164. Martin R, Sospedra M, Rosito M, Engelhardt B. Current multiple sclerosis treatments have improved our understanding of MS autoimmune pathogenesis. European journal of immunology. 2016;46(9):2078-90. doi:10.1002/eji.201646485.

165. Myhr KM. [Pharmacological treatment of multiple sclerosis]. Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke. 2010;130(5):490-2. doi:10.4045/tidsskr.09.1126.

166. Comi G, Radaelli M, Soelberg Sørensen P. Evolving concepts in the treatment of relapsing multiple sclerosis. The Lancet. 2017;389(10076):1347-56. doi:10.1016/S0140-6736(16)32388-1.

167. Trojano M, Tintore M, Montalban X, Hillert J, Kalincik T, Iaffaldano P, et al. Treatment decisions in multiple sclerosis — insights from real-world observational studies. Nature Reviews Neurology. 2017;13(2):105-18. doi:10.1038/nrneurol.2016.188.

168. Baker D, Marta M, Pryce G, Giovannoni G, Schmierer K. Memory B Cells are Major Targets for Effective Immunotherapy in Relapsing Multiple Sclerosis. EBioMedicine. 2017;16:41-50. doi:10.1016/j.ebiom.2017.01.042.

169. Baker D, Pryce G, Amor S, Giovannoni G, Schmierer K. Learning from other autoimmunities to understand targeting of B cells to control multiple sclerosis. Brain : a journal of neurology. 2018;141(10):2834-47. doi:10.1093/brain/awy239.

170. Baker D, Pryce G, James LK, Schmierer K, Giovannoni G. Failed B cell survival factor trials support the importance of memory B cells in multiple sclerosis. European journal of neurology. 2019doi:10.1111/ene.14105.

171. Lattanzi S, Cagnetti C, Danni M, Provinciali L, Silvestrini M. Oral and intravenous steroids for multiple sclerosis relapse: a systematic review and meta-analysis. Journal of neurology. 2017;264(8):1697-704. doi:10.1007/s00415-017-8505-0.

172. Rolfes L, Pfeuffer S, Ruck T, Melzer N, Pawlitzki M, Heming M, et al. Therapeutic Apheresis in Acute Relapsing Multiple Sclerosis: Current Evidence and Unmet Needs-A Systematic Review. J Clin Med. 2019;8(10)doi:10.3390/jcm8101623.

173. Imatinib for Multiple Sclerosis (MS) Relapses [Internet]. clinicaltrials.gov. 2018 [cited 31.10.2019]. Available from: https://clinicaltrials.gov/ct2/show/record/NCT03674099.

174. Høglund RA, Torsetnes SB, Lossius A, Bogen B, Homan EJ, Bremel R, et al. Human Cysteine Cathepsins Degrade Immunoglobulin G In Vitro in a Predictable Manner. International Journal of Molecular Sciences. 2019;20(19):4843. doi:10.3390/ijms20194843.

175. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302. doi:10.1002/ana.22366.

176. Lossius A, Tomescu-Baciu A, Holmøy T, Vedeler CA, Rosjo E, Lorentzen AR, et al. Selective intrathecal enrichment of G1m1-positive B cells in multiple sclerosis. Annals of clinical and translational neurology. 2017;4(10):756-61. doi:10.1002/acn3.451.

177. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic acids research. 2008;36(Web Server issue):W503-8. doi:10.1093/nar/gkn316.

178. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning: Mit Press; 2012.

179. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436. doi:10.1038/nature14539.

180. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. New England Journal of Medicine. 2019;380(14):1347-58. doi:10.1056/NEJMra1814259.

181. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep Learning: A Primer for Radiologists. Radiographics. 2017;37(7):2113-31. doi:10.1148/rg.2017170077.

182. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. Journal of Biomedical Informatics. 2015;53:405-14. doi:10.1016/j.jbi.2014.11.003.

183. Bremel RD, Homan EJ. An integrated approach to epitope analysis II: A system for proteomic-scale prediction of immunological characteristics. Immunome research. 2010;6:8. doi:10.1186/1745-7580-6-8.

184. Bremel RD, Homan EJ. An integrated approach to epitope analysis I: Dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. Immunome research. 2010;6:7. doi:10.1186/1745-7580-6-7.

185. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic acids research. 2019;47(D1):D339-d43. doi:10.1093/nar/gky1006.

186. Baird JTT, Craik CS. Chapter 575 - Trypsin. In: Rawlings ND, Salvesen G, editors. Handbook of Proteolytic Enzymes (Third Edition): Academic Press; 2013. p. 2594-600.

187. Barrett AJ, Chen J-M. Chapter 518 - Animal Legumain. In: Rawlings ND, Salvesen G, editors. Handbook of Proteolytic Enzymes (Third Edition): Academic Press; 2013. p. 2309-14.

188. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic acids research. 2014;42(Database issue):D503-9. doi:10.1093/nar/gkt953.

189. Bremel RD, Homan EJ. Recognition of higher order patterns in proteins: immunologic kernels. PloS one. 2013;8(7):e70115. doi:10.1371/journal.pone.0070115.

190. Biniossek ML, Nagler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. Journal of proteome research. 2011;10(12):5363-73. doi:10.1021/pr200621z.

191. Consortium U. UniProt: a hub for protein information. Nucleic acids research. 2015;43(D1):D204-D12. doi:10.1093/nar/gku989.

192. The NIHHMPWG, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al. The NIH Human Microbiome Project. Genome Research. 2009;19(12):2317-23. doi:10.1101/gr.096651.109.

193. Honey K, Rudensky AY. Lysosomal cysteine proteases regulate antigen presentation. Nature reviews Immunology. 2003;3(6):472-82. doi:10.1038/nri1110.

194. Kramer L, Turk D, Turk B. The Future of Cysteine Cathepsins in Disease Management. Trends in Pharmacological Sciences. 2017;38(10):873-98. doi:10.1016/j.tips.2017.06.003.

195. Stoka V, Turk V, Turk B. Lysosomal cathepsins and their regulation in aging and neurodegeneration. Ageing research reviews. 2016;32:22-37. doi:10.1016/j.arr.2016.04.010.

196. Beck H, Schwarz G, Schröter CJ, Deeg M, Baier D, Stevanovic S, et al. Cathepsin S and an asparagine-specific endoprotease dominate the proteolytic processing of human myelin basic protein in vitro. European journal of immunology. 2001;31(12):3726-36. doi:10.1002/1521-4141(200112)31:12<3726::AID-IMMU3726>3.0.CO;2-O.

197. Cross TG, Hornshaw PH. Can LC and LC-MS ever replace immunoassays? . JOURNAL OF APPLIED BIOANALYSIS. 2016;2(4):108-16. doi:10.17145/jab.16.015.

198. Laemmli UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 1970;227(5259):680-5. doi:10.1038/227680a0.

199. Smith BJ. SDS Polyacrylamide Gel Electrophoresis of Proteins. In: Walker JM, editor. Proteins. Totowa, NJ: Humana Press; 1984. p. 41-55.

200. Kim A, Hartman IZ, Poore B, Boronina T, Cole RN, Song N, et al. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. Nature communications. 2014;5(1):5369. doi:10.1038/ncomms6369.

201. Hastings KT, Cresswell P. Disulfide reduction in the endocytic pathway: immunological functions of gamma-interferon-inducible lysosomal thiol reductase. Antioxidants & redox signaling. 2011;15(3):657-68. doi:10.1089/ars.2010.3684.

202. Shimabukuro-Vornhagen A, Zoghi S, Liebig TM, Wennhold K, Chemitz J, Draube A, et al. Inhibition of protein geranylgeranylation specifically interferes with CD40-dependent B cell activation, resulting in a reduced capacity to induce T cell immunity. Journal of immunology (Baltimore, Md : 1950). 2014;193(10):5294-305. doi:10.4049/jimmunol.1203436.

203. Haque MA, Li P, Jackson SK, Zarour HM, Hawes JW, Phan UT, et al. Absence of γ-Interferon–inducible Lysosomal Thiol Reductase in Melanomas Disrupts T Cell Recognition of Select Immunodominant Epitopes. The Journal of experimental medicine. 2002;195(10):1267-77. doi:10.1084/jem.20011853.

204. Yiannios CN, Karabinos JV. Oxidation of Thiols by Dimethyl Sulfoxide. The Journal of Organic Chemistry. 1963;28(11):3246-8. doi:10.1021/jo01046a528.

205. Frentsch M, Arbach O, Kirchhoff D, Moewes B, Worm M, Rothe M, et al. Direct access to CD4+ T cells specific for defined antigens according to CD154 expression. Nature medicine. 2005;11(10):1118-24. doi:10.1038/nm1292.

206. Meier S, Stark R, Frentsch M, Thiel A. The influence of different stimulation conditions on the assessment of antigen-induced CD154 expression on CD4+ T cells. Cytometry A. 2008;73(11):1035-42. doi:10.1002/cyto.a.20640.

207. Høglund RA, Hestvik AL, Holmøy T, Maghazachi AA. Expression and functional activity of chemokine receptors in glatiramer acetate-specific T cells isolated from multiple sclerosis patient receiving the drug glatiramer acetate. Human immunology. 2011;72(2):124-34. doi:10.1016/j.humimm.2010.10.016.

208. Novak EJ, Liu AW, Gebe JA, Falk BA, Nepom GT, Koelle DM, et al. Tetramer-guided epitope mapping: rapid identification and characterization of immunodominant CD4+ T cell epitopes from complex antigens. Journal of immunology (Baltimore, Md : 1950). 2001;166(11):6665-70. doi:10.4049/jimmunol.166.11.6665.

209. Maciorowski Z, Chattopadhyay PK, Jain P. Basic Multicolor Flow Cytometry. Current Protocols in Immunology. 2017;117(1):5.4.1-5.4.38. doi:10.1002/cpim.26.

210. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300. doi:10.2307/2346101.

211. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. The New England journal of medicine. 2017;376(3):221-34. doi:10.1056/NEJMoa1601277.

212. Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. The New England journal of medicine. 2008;358(7):676-88. doi:10.1056/NEJMoa0706383.

213. Bar-Or A, Grove RA, Austin DJ, Tolson JM, VanMeter SA, Lewis EW, et al. Subcutaneous ofatumumab in patients with relapsing-remitting multiple sclerosis. The MIRROR study. 2018;90(20):e1805-e14. doi:10.1212/wnl.0000000000005516.

214. Sospedra M. B cells in multiple sclerosis. Current opinion in neurology. 2018;31(3):256-62. doi:10.1097/wco.000000000000563.

215. Wekerle H. B cells in multiple sclerosis. Autoimmunity. 2017;50(1):57-60. doi:10.1080/08916934.2017.1281914.

216. Blauth K, Owens GP, Bennett JL. The Ins and Outs of B Cells in Multiple Sclerosis. Frontiers in immunology. 2015;6(565)doi:10.3389/fimmu.2015.00565.

217. Sabatino JJ, Pröbstel A-K, Zamvil SS. B cells in autoimmune and neurodegenerative central nervous system diseases. Nature Reviews Neuroscience. 2019doi:10.1038/s41583-019-0233-2.

218. Kabat EA, Moore DH, Landow H. An electrophoretic study of the protein components in cerebrospinal fluid and their relationship to the serum proteins. Journal of Clinical Investigation. 1942;21(5):571-7. doi:10.1172/JCI101335.

219. Yahr MD, Goldensohn SS, Kabat EA. Further studies on gamma globulin content of cerebrospinal fluid in multiple sclerosis and other neurological diseases. Annals of the New York Academy of Sciences. 1954;58(5):613-24. doi:10.1111/j.1749-6632.1954.tb54099.x.

220. Michel L, Grasmuck C, Charabati M, Lécuyer M-A, Zandee S, Dhaeze T, et al. Activated leukocyte cell adhesion molecule regulates B lymphocyte migration across central nervous system barriers. Science translational medicine. 2019;11(518):eaaw0475. doi:10.1126/scitranslmed.aaw0475.

221. Li R, Patterson KR, Bar-Or A. Reassessing B cell contributions in multiple sclerosis. Nature Immunology. 2018;19(7):696-707. doi:10.1038/s41590-018-0135-x.

222. Thompson SA, Jones JL, Cox AL, Compston DA, Coles AJ. B-cell reconstitution and BAFF after alemtuzumab (Campath-1H) treatment of multiple sclerosis. Journal of clinical immunology. 2010;30(1):99-105. doi:10.1007/s10875-009-9327-3.

223. Cross AH, Stark JL, Lauber J, Ramsbottom MJ, Lyons JA. Rituximab reduces B cells and T cells in cerebrospinal fluid of multiple sclerosis patients. Journal of neuroimmunology. 2006;180(1-2):63-70. doi:10.1016/j.jneuroim.2006.06.029.

224. Petereit HF, Rubbert-Roth A. Rituximab levels in cerebrospinal fluid of patients with neurological autoimmune disorders. Multiple sclerosis (Houndmills, Basingstoke, England). 2009;15(2):189-92. doi:10.1177/1352458508098268.

225. Komori M, Lin YC, Cortese I, Blake A, Ohayon J, Cherup J, et al. Insufficient disease inhibition by intrathecal rituximab in progressive multiple sclerosis. Annals of clinical and translational neurology. 2016;3(3):166-79. doi:10.1002/acn3.293.

226. Smith MD, Martin KA, Calabresi PA, Bhargava P. Dimethyl fumarate alters B-cell memory and cytokine production in MS patients. Annals of clinical and translational neurology. 2017;4(5):351-5. doi:10.1002/acn3.411.

227. Longbrake EE, Cantoni C, Chahin S, Cignarella F, Cross AH, Piccio L. Dimethyl fumarate induces changes in B- and T-lymphocyte function independent of the effects on absolute lymphocyte count. Multiple sclerosis (Houndmills, Basingstoke, England). 2017:1352458517707069. doi:10.1177/1352458517707069.

228. Li R, Rezk A, Ghadiri M, Luessi F, Zipp F, Li H, et al. Dimethyl Fumarate Treatment Mediates an Anti-Inflammatory Shift in B Cell Subsets of Patients with Multiple Sclerosis. Journal of immunology (Baltimore, Md : 1950). 2017;198(2):691-8. doi:10.4049/jimmunol.1601649.

229. Lundy SK, Wu Q, Wang Q, Dowling CA, Taitano SH, Mao G, et al. Dimethyl fumarate treatment of relapsing-remitting multiple sclerosis influences B-cell subsets. Neurology(R) neuroimmunology & neuroinflammation. 2016;3(2):e211. doi:10.1212/nxi.0000000000000211.

230. Edwards K, Penner N, Rogge M, Sheikh S, Zhu B. A pharmacokinetic study of delayed-release dimethyl fumarate to evaluate cerebrospinal fluid penetration in patients with secondary progressive multiple sclerosis. September; London2016.

231. Liliemark J. The clinical pharmacokinetics of cladribine. Clin Pharmacokinet. 1997;32(2):120-31. doi:10.2165/00003088-199732020-00003.

232. Cepok S, Rosche B, Grummel V, Vogel F, Zhou D, Sayn J, et al. Short-lived plasma blasts are the main B cell effector subset during the course of multiple sclerosis. Brain : a journal of neurology. 2005;128(Pt 7):1667-76. doi:10.1093/brain/awh486.

233. Colombo M, Dono M, Gazzola P, Chiorazzi N, Mancardi G, Ferrarini M. Maintenance of B lymphocyte-related clones in the cerebrospinal fluid of multiple sclerosis patients. European journal of immunology. 2003;33(12):3433-8. doi:10.1002/eji.200324144.

234. Tomescu-Baciu A, Johansen JN, Holmøy T, Greiff V, Stensland M, de Souza GA, et al. Persistence of intrathecal oligoclonal B cells and IgG in multiple sclerosis. Journal of neuroimmunology. 2019;333:576966. doi:10.1016/j.jneuroim.2019.576966.

235. Jarius S, Eichhorn P, Franciotta D, Petereit HF, Akman-Demir G, Wick M, et al. The MRZ reaction as a highly specific marker of multiple sclerosis: re-evaluation and structured review of the literature. Journal of neurology. 2017;264(3):453-66. doi:10.1007/s00415-016-8360-4.

236. Blauth K, Soltys J, Matschulat A, Reiter CR, Ritchie A, Baird NL, et al. Antibodies produced by clonally expanded plasma cells in multiple sclerosis cerebrospinal fluid cause demyelination of spinal cord explants. Acta Neuropathol. 2015;130(6):765-81. doi:10.1007/s00401-015-1500-6.

237. Rivas JR, Ireland SJ, Chkheidze R, Rounds WH, Lim J, Johnson J, et al. Peripheral VH4+ plasmablasts demonstrate autoreactive B cell expansion toward brain antigens in early multiple sclerosis patients. Acta Neuropathol. 2017;133(1):43-60. doi:10.1007/s00401-016-1627-0.

238. Liu Y, Given KS, Harlow DE, Matschulat AM, Macklin WB, Bennett JL, et al. Myelin-specific multiple sclerosis antibodies cause complement-dependent oligodendrocyte loss and demyelination. Acta Neuropathol Commun. 2017;5(1):25. doi:10.1186/s40478-017-0428-6.

239. Brandle SM, Obermeier B, Senel M, Bruder J, Mentele R, Khademi M, et al. Distinct oligoclonal band antibodies in multiple sclerosis recognize ubiquitous self-proteins. Proceedings of the

National Academy of Sciences of the United States of America. 2016;113(28):7864-9. doi:10.1073/pnas.1522730113.

240. Jelcic I, Al Nimer F, Wang J, Lentsch V, Planas R, Jelcic I, et al. Memory B Cells Activate Brain-Homing, Autoreactive CD4+ T Cells in Multiple Sclerosis. Cell. 2018;175(1):85-100.e23. doi:10.1016/j.cell.2018.08.011.

241. De Groot AS, Scott DW. Immunogenicity of protein therapeutics. Trends in immunology. 2007;28(11):482-90. doi:10.1016/j.it.2007.07.011.

242. Waldmann H. Human monoclonal antibodies: the residual challenge of antibody immunogenicity. Methods in molecular biology (Clifton, NJ). 2014;1060:1-8. doi:10.1007/978-1-62703-586-6_1.

243. Cassotta A, Mikol V, Bertrand T, Pouzieux S, Le Parc J, Ferrari P, et al. A single T cell epitope drives the neutralizing anti-drug antibody response to natalizumab in multiple sclerosis patients. Nature medicine. 2019;25(9):1402-7. doi:10.1038/s41591-019-0568-2.

244. Hamze M, Meunier S, Karle A, Gdoura A, Goudet A, Szely N, et al. Characterization of CD4 T Cell Epitopes of Infliximab and Rituximab Identified from Healthy Donors. Frontiers in immunology. 2017;8(500)doi:10.3389/fimmu.2017.00500.

245. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018;154(3):394-406. doi:10.1111/imm.12889.

246. Reth M. Antigen Receptors on B Lymphocytes. Annual review of immunology. 1992;10(1):97-121. doi:10.1146/annurev.iy.10.040192.000525.

247. Sorde L, Spindeldreher S, Palmer E, Karle A. Tregitopes and impaired antigen presentation: Drivers of the immunomodulatory effects of IVIg? Immun Inflamm Dis. 2017;5(4):400-15. doi:10.1002/iid3.167.

248. Seward RJ, Drouin EE, Steere AC, Costello CE. Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. Molecular & cellular proteomics 2011;10(3):M110.002477-M110. doi:10.1074/mcp.M110.002477.

249. Bogen B, Ruffini P. Review: to what extent are T cells tolerant to immunoglobulin variable regions? Scandinavian journal of immunology. 2009;70(6):526-30. doi:10.1111/j.1365-3083.2009.02340.x.

250. Palmeira P, Quinello C, Silveira-Lessa AL, Zago CA, Carneiro-Sampaio M. IgG placental transfer in healthy and pathological pregnancies. Clinical & developmental immunology. 2012;2012:985646. doi:10.1155/2012/985646.

251. Dwyer JM. Immunoglobulins in autoimmunity: history and mechanisms of action. Clin Exp Rheumatol. 1996;14 Suppl 15:S3-7.

252. João C. Immunoglobulin is a highly diverse self-molecule that improves cellular diversity and function during immune reconstitution. Medical Hypotheses. 2007;68(1):158-61. doi:10.1016/j.mehy.2006.05.062.

253. Tjon AS, Tha-In T, Metselaar HJ, van Gent R, van der Laan LJ, Groothuismink ZM, et al. Patients treated with high-dose intravenous immunoglobulin show selective activation of regulatory T cells. Clinical and experimental immunology. 2013;173(2):259-67. doi:10.1111/cei.12102.

254. Kessel A, Ammuri H, Peri R, Pavlotzky ER, Blank M, Shoenfeld Y, et al. Intravenous immunoglobulin therapy affects T regulatory cells by increasing their suppressive function. Journal of immunology (Baltimore, Md : 1950). 2007;179(8):5571-5. doi:10.4049/jimmunol.179.8.5571.

255. Zhang G, Wang Q, Song Y, Cheng P, Xu R, Feng X, et al. Intravenous immunoglobulin promotes the proliferation of CD4(+)CD25(+) Foxp3(+) regulatory T cells and the cytokines secretion in patients with Guillain-Barre syndrome in vitro. Journal of neuroimmunology. 2019;336:577042. doi:10.1016/j.jneuroim.2019.577042.

256. Massoud AH, Guay J, Shalaby KH, Bjur E, Ablona A, Chan D, et al. Intravenous immunoglobulin attenuates airway inflammation through induction of forkhead box protein 3-positive regulatory T cells. J Allergy Clin Immunol. 2012;129(6):1656-65.e3. doi:10.1016/j.jaci.2012.02.050.

257. Tha-In T, Metselaar HJ, Bushell AR, Kwekkeboom J, Wood KJ. Intravenous immunoglobulins promote skin allograft acceptance by triggering functional activation of CD4+Foxp3+ T cells. Transplantation. 2010;89(12):1446-55. doi:10.1097/TP.0b013e3181dd6bf1.

258. Ephrem A, Chamat S, Miquel C, Fisson S, Mouthon L, Caligiuri G, et al. Expansion of CD4+CD25+ regulatory T cells by intravenous immunoglobulin: a critical factor in controlling experimental autoimmune encephalomyelitis. Blood. 2008;111(2):715-22. doi:10.1182/blood-2007-03-079947.

259. Maddur MS, Kaveri SV, Bayry J. Circulating Normal IgG as Stimulator of Regulatory T Cells: Lessons from Intravenous Immunoglobulin. Trends in immunology. 2017;38(11):789-92. doi:10.1016/j.it.2017.08.008.

260. Paquin Proulx D, Aubin E, Lemieux R, Bazin R. Inhibition of B cell-mediated antigen presentation by intravenous immunoglobulins (IVIg). Clinical immunology (Orlando, Fla). 2010;135(3):422-9. doi:10.1016/j.clim.2010.01.001.

261. De Groot AS, Moise L, McMurry JA, Wambre E, Van Overtvelt L, Moingeon P, et al. Activation of natural regulatory T cells by IgG Fc-derived peptide "Tregitopes". Blood. 2008;112(8):3303-11. doi:10.1182/blood-2008-02-138073.

262. Cousens LP, Tassone R, Mazer BD, Ramachandiran V, Scott DW, De Groot AS. Tregitope update: Mechanism of action parallels IVIg. Autoimmunity reviews. 2013;12(3):436-43. doi:10.1016/j.autrev.2012.08.017.

263. De Groot AS, Cousens L, Mingozzi F, Martin W. Tregitope peptides: the active pharmaceutical ingredient of IVIG? Clinical & developmental immunology. 2013;2013:493138. doi:10.1155/2013/493138.

264. Kinnunen T, Chamberlain N, Morbach H, Cantaert T, Lynch M, Preston-Hurlburt P, et al. Specific peripheral B cell tolerance defects in patients with multiple sclerosis. The Journal of clinical investigation. 2013;123(6):2737-41. doi:10.1172/jci68775.

265. Souza TA, Stollar BD, Sullivan JL, Luzuriaga K, Thorley-Lawson DA. Influence of EBV on the Peripheral Blood Memory B Cell Compartment. The Journal of Immunology. 2007;179(5):3153-60. doi:10.4049/jimmunol.179.5.3153.

266. Watson CT, Breden F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. Genes Immun. 2012;13(5):363-73. doi:10.1038/gene.2012.12.

267. Iversen R, Roy B, Stamnaes J, Høydahl LS, Hnida K, Neumann RS, et al. Efficient T cell–B cell collaboration guides autoantibody epitope bias and onset of celiac disease. Proceedings of the National Academy of Sciences. 2019;116(30):15134-9. doi:10.1073/pnas.1901561116.

268. Raposo B, Dobritzsch D, Ge C, Ekman D, Xu B, Lindh I, et al. Epitope-specific antibody response is controlled by immunoglobulin VH polymorphisms. The Journal of experimental medicine. 2014;211(3):405-11. doi:10.1084/jem.20130968.

269. Flajnik MF. A cold-blooded view of adaptive immunity. Nature Reviews Immunology. 2018;18(7):438-53. doi:10.1038/s41577-018-0003-9.

270. Lauritzsen GF, Hofgaard PO, Schenck K, Bogen B. Clonal deletion of thymocytes as a tumor escape mechanism. International journal of cancer. 1998;78(2):216-22. doi:10.1002/(SICI)1097-0215(19981005)78:2<216::AID-IJC16>3.0.CO;2-8.

271. Rivino L, Gruarin P, Haringer B, Steinfelder S, Lozza L, Steckel B, et al. CCR6 is expressed on an IL-10-producing, autoreactive memory T cell population with context-dependent regulatory function. The Journal of experimental medicine. 2010;207(3):565-77. doi:10.1084/jem.20091021.

272. Bains W. How to write up a hypothesis: the good, the bad and the ugly. Medical Hypotheses. 2005;64(4):665-8. doi:10.1016/j.mehy.2004.10.003.

273. Noyan F, Lee YS, Zimmermann K, Hardtke-Wolenski M, Taubert R, Warnecke G, et al. Isolation of human antigen-specific regulatory T cells with high suppressive function. European journal of immunology. 2014;44(9):2592-602. doi:10.1002/eji.201344381.

274. Bacher P, Scheffold A. Flow-cytometric analysis of rare antigen-specific T cells. Cytometry A. 2013;83(8):692-701. doi:10.1002/cyto.a.22317.

275. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. Data from: A public database of memory and naive B-cell receptor sequences. Dryad Data Repository; 2016.

276. Lindeman I, Emerton G, Mamanova L, Snir O, Polanski K, Qiao SW, et al. BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. Nat Methods. 2018;15(8):563-5. doi:10.1038/s41592-018-0082-3.

Paper I

I

frontiers in Immunology

ORIGINAL RESEARCH
published: 02 October 2017
doi: 10.3389/fimmu.2017.01255

In Silico Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis

Rune A. Høglund[1,2], Andreas Lossius[1,3], Jorunn N. Johansen[3], Jane Homan[4], Jūratė Šaltytė Benth[2,5], Harlan Robins[6], Bjarne Bogen[2,3,7], Robert D. Bremel[4†] and Trygve Holmøy[1,2†]*

[1] Department of Neurology, Akershus University Hospital, Lørenskog, Norway, [2] Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [3] Faculty of Medicine, Department of Immunology and Transfusion Medicine, University of Oslo and Oslo University Hospital Rikshospitalet, Oslo, Norway, [4] EigenBio LLC, Madison, WI, United States, [5] Health Services Research Unit, Akershus University Hospital, Lørenskog, Norway, [6] Adaptive Biotechnologies, Seattle, WA, United States, [7] Centre for Immune Regulation, University of Oslo, Oslo, Norway

Memory B cells acting as antigen-presenting cells are believed to be important in multiple sclerosis (MS), but the antigen they present remains unknown. We hypothesized that B cells may activate CD4+ T cells in the central nervous system of MS patients by presenting idiotopes from their own immunoglobulin variable regions on human leukocyte antigen (HLA) class II molecules. Here, we use bioinformatics prediction analysis of B cell immunoglobulin variable regions from 11 MS patients and 6 controls with other inflammatory neurological disorders (OINDs), to assess whether the prerequisites for such idiotope-driven T–B cell collaboration are present. Our findings indicate that idiotopes from the complementarity determining region (CDR) 3 of MS patients on average have high predicted affinities for disease associated HLA-DRB1*15:01 molecules and are predicted to be endosomally processed by cathepsin S and L in positions that allows such HLA binding to occur. Additionally, complementarity determining region 3 sequences from cerebrospinal fluid (CSF) B cells from MS patients contain on average more rare T cell-exposed motifs that could potentially escape tolerance and stimulate CD4+ T cells than CSF B cells from OIND patients. Many of these features were associated with preferential use of the IGHV4 gene family by CSF B cells from MS patients. This is the first study to combine high-throughput sequencing of patient immune repertoires with large-scale prediction analysis and provides key indicators for future *in vitro* and *in vivo* analyses.

Keywords: multiple sclerosis, idiotope, B cell, T cell, bioinformatics, immunoglobulin heavy variable, immunosequencing, immunoglobulin

OPEN ACCESS

Edited by:
Zsolt Illes,
University of Southern Denmark
Odense, Denmark

Reviewed by:
Lindsay B. Nicholson,
University of Bristol,
United Kingdom
Chandirasegaran Massilamany,
National Institutes
of Health (NIH), United States
ADI Vaknin-Dembinsky,
Hadassah Medical Center, Israel

****Correspondence:***
Rune A. Høglund
r.a.hoglund@medisin.uio.no

†Joint senior authorship.

Specialty section:
This article was submitted to Multiple Sclerosis and Neuroimmunology, a section of the journal Frontiers in Immunology

Received: *26 May 2017*
Accepted: *20 September 2017*
Published: *02 October 2017*

Citation:
Høglund RA, Lossius A, Johansen JN, Homan J, Benth JŠ, Robins H, Bogen B, Bremel RD and Holmøy T (2017) In Silico Prediction Analysis of Idiotope-Driven T–B Cell Collaboration in Multiple Sclerosis. Front. Immunol. 8:1255. doi: 10.3389/fimmu.2017.01255

INTRODUCTION

Multiple sclerosis (MS) is a chronic inflammatory, demyelinating, and neurodegenerative disease of the central nervous system (CNS), thought to be mainly mediated by the immune system (1). Although T cells as mediators of disease have been investigated thoroughly over the years, recent trials of B cell targeted therapies (i.e., rituximab and ocrelizumab) point to these cells as equally

Abbreviations: APC, antigen-presenting cell; BCR, B cell receptor; CDR, complementarity determining region; mAb, monoclonal antibody; c/mTEC, cortical/medullary thymic epithelial cell; CI, confidence interval; CNS, central nervous system; CSF, cerebrospinal fluid; FC, Frequency class; FW, framework; HLA, human leukocyte antigen; IC_{50}, the half maximal inhibitory concentration; IgG, immunoglobulin G; IGHV, immunoglobulin heavy variable; MHC, major histocompatibility complex; MS, multiple sclerosis; OIND, other inflammatory neurological disease; TCEM, T cell-exposed motif.

important contributors (2, 3). Notably, depleting B cells in the periphery has a substantial effect within the CNS (4). It is also possible that other approved therapies for MS act by depleting or prohibiting CD19$^+$, CD27$^+$ memory B cells from invading the CNS (5). In MS, B cell immunoglobulin heavy chain variable (IGHV) repertoires suggest that clonally expanded plasma cells in the brain and cerebrospinal fluid (CSF) are derived from peripheral B cells that have matured in cervical lymph nodes (6–8). Hence, it seems that peripheral memory B cells play an important role in MS immunopathology.

The mechanisms by which memory B cells induce pathology could involve antibody production or secretion of cytokines (9). However, B cell-depleting therapies targeting CD20 ameliorate disease before reducing immunoglobulin G (IgG) production (10), which is therefore not likely their main mechanism. Whereas the discovery of intrathecal Ig production in the CNS is an old one (11), a common antigenic determinant has yet to be discovered. However, B cells in MS lesions (12) and CSF (13–17) have evidently undergone somatic hypermutation indicating T cell help, suggestive of a possible antigen being involved with B cells as antigen-presenting cells (APC) (18). We have proposed an alternative hypothesis to explain how T–B cell collaboration in absence of a common antigen can result in intrathecal IgG production (19). It was shown that B cells present endogenously processed variable region fragments (idiotopes) on major histocompatibility complex (MHC) class II molecules (20, 21). T cells can specifically recognize this idiotope–MHC complex, resulting in a T cell response (22, 23). Such an interaction between idiotope$^+$ B cells that present idiotope-MHCII and idiotope-specific CD4$^+$ T cells is named idiotope-driven T–B collaboration (21, 24, 25) (**Figure 1**). An important feature of idiotope-driven T–B collaboration is that unlike conventional antigen-linked T–B collaboration (26, 27), it is unlinked in the sense that while the B cell can recognize any (self) antigen, the T cell recognizes a different antigen (idiotope-MHCII). Thus, B cells of theoretically any specificity, including self-specificity, can be helped by idiotope-specific CD4$^+$ T cells to develop into IgG producing plasma cells (25). Consistent with this idea, endogenous idiotopes were eluted from MHC II molecules on B cells (28, 29), and idiotope-driven T–B cell collaboration has been shown to drive the development of autoimmune disease in transgenic mice (30).

Extending idiotope driven T–B collaboration to humans, we have previously demonstrated that human leukocyte antigen (HLA)-DR restricted CD4$^+$ T cells from blood and CSF of MS patients can recognize multiple idiotopes within the complementarity determining region 3 (CDR3) and mutated framework (FW) regions on autologous CSF IgG (31–33), showing that MS patients have a repertoire idiotope-matched T–B cell pairs. Idiotope-specific CD4$^+$ T cells specifically recognized idiotopes presented by autologous Epstein Barr virus transformed CSF B cells, suggesting that B cells can process and present their endogenous idiotopes on HLA class II molecules (33), and they are also induced to kill oligodendrocytes upon activation (34).

Further large-scale investigations into this mechanism in MS have been hampered by overwhelming numbers of possible IGHV region idiotopes. High-throughput sequencing now offers a possibility to characterize the immune repertoire in unprecedented depth and detail (35) and has triggered a rapid growth of bioinformatic approaches for diagnostic and research purposes (36), including methods to assess possible immunogenicity of B cell variable region sequences (37).

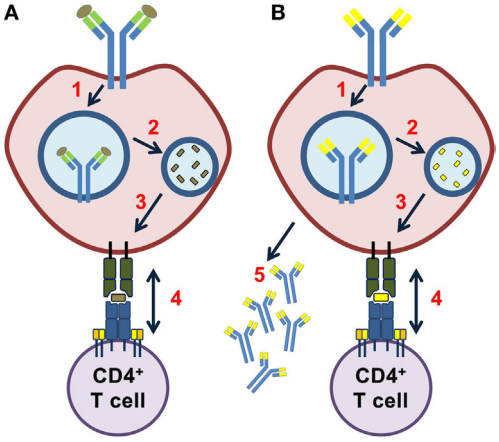


FIGURE 1 | Idiotope driven T–B cell collaboration. In a classical T–B cell collaboration **(A)**, an exogenous antigen bound to the B cell receptor (BCR) is brought into the endosomal pathway (1), processed by proteases (2), and fragments of the antigen presented on human leukocyte antigen (HLA) class II molecules (3). A CD4$^+$ T cell specifically recognizing this exogenous antigen provides help to the B cell (4). **(B)** In idiotope driven T–B cell collaboration, a BCR of any specificity (including self) is brought into the endosomal pathway (1), the BCR processed by endosomal proteases (2) and fragments from the variable region presented on HLA class II molecules (3). An idiotope-specific CD4$^+$ T cell may help the B cell in a non-linked mechanism (4). All of steps 1–4 must occur for idiotope-driven T–B cell collaboration to take place and may result in differentiation of B cells into immunoglobulin G (IgG) secreting cells (5).

There are several prerequisites for idiotope-driven T–B cell collaboration: the idiotopes would need to undergo endosomal processing; the processed idiotope fragments must have sufficient affinity for HLA class II molecules; and they must be sufficiently rare to avoid T cell tolerance (**Figure 1**). In this article, we combined high-throughput sequencing of the B cell receptor (BCR) transcriptome with *in silico* prediction analysis to assess whether these prerequisites exist in the intrathecal compartment of MS patients.

MATERIALS AND METHODS

Patients

In this study, we included 11 relapsing-remitting MS patients and six patients with other inflammatory neurological disorders (OINDs), recruited at Akershus University Hospital and Oslo University Hospital. All patients (except MS-11) and the procedures for PBMC isolation, RNA extraction and cDNA preparation have been described previously (7). The cDNA sequencing was performed by Adaptive Biotechnologies using the immunoSEQ level assay (38), with resulting 130bp sequences spanning the IGH-VDJ region.

Multiple sclerosis patients were either treatment naive or treated with first-line therapies (MS-2, MS-3, and MS-4), while OIND patients were untreated at the time of lumbar puncture. All MS patients and one OIND patient had oligoclonal IgG bands in CSF. A summary of patient and sample characteristics are shown in S1 in Supplementary Material.

All participants provided written informed consent for participation. The study was approved by the Committee for Research Ethics in the South-Eastern Norwegian Healthy Authority (REK Sør-Øst S-04143a), the Norwegian Social Science Data Services (no. 11069) and the review boards at AHUS and OUS.

Genotyping for HLA-A, HLA-B, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1 was performed with four-digit resolution at the Department of Immunology and Transfusion Medicine at Oslo University Hospital, by utilizing a combination of sequence-specific primer- and sequencing based typing technologies. For some patients (MS-1, 3, 5, 8, OIND-1, 5, and 6), we used the strong linkage disequilibrium with HLA-DPB1 to deduce their likely DPA1 alleles (39). HLA types are shown in **Table 1**.

Preparation of Datasets

After removing non-productive sequences, IGHV amino acid sequences were deduced using the ImMunoGeneTics (IMGT) database and the IMGT/High-V-Quest analysis tool (version 3.3.4) (40, 41). This analysis identified additional non-productive sequences that were removed. IGHV transcripts comprising more than 0.5% of total reads within each compartment were designated "highly transcribed." Finally a single FASTA file containing all the IGHV sequences with tagged information consisting of patient code, compartment, frequency rank and tag describing whether it was highly transcribed was prepared. These sequences are deposited online at http://doi.org/10.6084/m9.figshare.5035703.

An extensive public dataset of IGHV nucleotide sequences from three healthy individuals (42, 43) was obtained online (http://datadryad.org/resource/doi:10.5061/dryad.35ks2). The corresponding IGHV amino acid sequences were deduced according to IMGT standards, and used for further analysis.

The compiled patient dataset of IGHV amino acid sequences was submitted to EigenBio (WI, USA) for processing and prediction analysis. Each sequence was given a unique general identifier (gi), and sequences with exact matching amino acid sequence within each patient were identified and given a clonal identifier for statistical purposes. Every possible 15-mer and 9-mer (denoted as IGHV fragments) were derived from each IGHV amino acid sequence, and indexed according to their N-terminus CDR3-relative position as determined by IMGT standards (44), designating cysteine 104 at the start of CDR3 as position 0. This indexing process resulted in extensive databases of overlapping IGHV fragments offset by one single amino acid, and provided a basis for systematic comparison of fragments in the FW3 and CDR3 regions across MS and OIND patients, and healthy individuals.

Cathepsin Cleavage Probabilities

Peptidase cleavage by cathepsins S, L, and B was predicted with neural network models developed using datasets from Biniossek et al. (45), by methods described previously (46). In short, all IGHV sequences were converted into sequential octamers using the P4P3P3P1-P1′ P2′ P3′ P4′ convention with the scissile bond designated as the bond between amino acid 4 and amino acid 5 designated P1P1′. The neural network ensembles each produce a probability of cleavage of the scissile bond ranging from 0 (uncleaved) to 1 (cleaved). An ensemble median cleavage probability of >0.8 was set as the prediction threshold for the analyses.

Endosomal enzymes digest proteins into peptides of varying lengths and a peptide 15-mer is commonly presented on HLA

TABLE 1 | Patients HLA types.

	HLA class I alleles		HLA class II alleles				
Patient	**HLA-A**	**HLA-B**	**DRB1**	**DQA1**	**DQB1**	**DPA1**	**DPB1**
MS-1	02:01 + 02:01	15:01 + 44:02	13:02 + 15:01	01:02 + 01:02	06:02 + 06:04	01:03/05 + 01:03/05[a]	04:01 + 04:01
MS-2	02:01 + 02:01	07:02 + 27:05	01:01 + 15:01	01:01 + 01:02	05:01 + 06:02	01:03 + 01:03	04:01 + 04:02
MS-3	03:01 + 25:01	07:02 + 18:01	15:01 + 15:01	01:02 + 01:02	06:02 + 06:02	01:03/05 + 01:03/05[a]	04:01 + 04:01
MS-4	03:01 + 24:02	07:02 + 35:01	01:01 + 15:01	01:01 + 01:02	05:01 + 06:02	01:03 + 01:03	02:01 + 20:01
MS-5	02:01 + 30:01	13:02 + 44:03	07:01 + 15:01	01:02 + 02:01	02:02 + 06:02	01:03/05 + 01:03/05[a]	04:01 + 04:01
MS-6	03:01 + 32:01	07:02 + 08:01	03:01 + 13:01	01:03 + 05:01	02:01 + 06:03	01:03 + 02:01	01:01 + 04:01
MS-7	24:02 + 69:01	35:01 + 44:02	11:04 + 15:01	01:02 + 05:05	03:01 + 06:02	01:03 + 01:03	04:01 + 04:02
MS-8	01:01 + 31:01	07:02 + 15:01	15:01 + 15:01	01:02 + 01:02	06:02 + 06:02	01:03 + 01:03[a]	02:01 + 02:01
MS-9	25:01 + 31:01	18:01 + 44:03	07:01 + 08:01	02:01 + 04:01	02:02 + 04:02	01:03 + 01:03	02:01 + 04:01
MS-10	03:01 + 29:02	35:01 + 45:01	01:01 + 04:05	01:01 + 03:03	03:02 + 05:01	01:03 + 01:03	04:01 + 04:02
MS-11	02:01 + 02:01	15:01 + 40:01	04:04 + 11:01	03:01 + 05:05	03:01 + 03:02	01:03 + 02:02	02:01 + 05:01
OIND-1	01:01 + 68:01	08:01 + 35:03	03:01 + 15:01	01:02 + 05:01	02:01 + 06:02	02:01/02/03:02 + 01:03[a]	01:01 + 03:01
OIND-2	01:01 + 02:05	08:01 + 50:01	03:01 + 07:01	02:01 + 05:01	02:01 + 02:02	01:03 + 01:03	04:02 + 104:01
OIND-3	01:01 + 01:01	08:01 + 08:01	03:01 + 03:01	05:01 + 05:01	02:01 + 02:01	02:01 + 02:01	01:01 + 14:01
OIND-4	02:01 + 02:01	15:01 + 27:05	01:01 + 04:01	01:01 + 03:01	03:02 + 05:01	01:03 + 01:03	04:01 + 04:02
OIND-5	03:01 + 68:01	07:02 + 44:02	07:01 + 15:01	01:02 + 02:01	02:02 + 06:02	01:03/03:01/04:01 + 02:01[a]	04:02 + 11:01
OIND-6	01:01 + 01:01	08:01 + 08:01	03:01 + 03:01	05:01 + 05:01	02:01 + 02:01	02:01/02/03:02 + 1:03/03:01/04:01[a]	01:01 + 04:02

[a]*Deduced by DPB1 linkage disequilibrium.*

class II (47). However, the HLA class II molecules display peptides of widely varying lengths. In an effort to simulate this process, a "fuzzy logic" system was devised where peptide excision probability was determined by examining the simultaneous cleavage probabilities from the N-terminus minus three amino acids to the C-terminus plus three amino acids. Thus, the process generates predicted excised peptides ranging from 15 to 21 amino acids. The cutoffs for this "fuzzy logic" excision prediction were intentionally lowered to avoid under-prediction. If the maximum probability of cathepsin cleavage at either terminal was ≥0.5 and simultaneously had a probability of ≥0.25 on the other end, that would lead to an "*Excision*" call.

HLA Affinity Predictions

For each CSF-derived 9- and 15-mer peptide, we predicted the affinity for 37 HLA class I and 28 class II molecules using previously described models (37, 48, 49). The neural network ensembles used for affinity predictions were developed using public datasets of allelic affinities [half-maximal inhibitory concentration (IC_{50}) units] from http://www.iedb.org (downloaded June 2012). The alleles for which affinities were predicted are shown in S2 in Supplementary Material. Predicted affinities are either presented as the natural logarithm of IC_{50} [$\ln(IC_{50})$] or a Johnson SI standardized value of $\ln(IC_{50})$ within patient and compartment. Standardization was performed to bring the predicted values onto an equal scale for comparative and/or illustrative purposes. In previous publications using such predictions, standardizations were performed within protein, as all peptides within a protein may compete with each other in terms of HLA affinity (37, 46). For the current publication, the IGHV sequences were shorter, and there is also a possibility of HLA affinity competition between different IgG molecules, hence the overall within patient standardization.

T Cell-Exposed Motifs (TCEMs)

In a peptide-HLA (pHLA) complex, some amino acids of the peptide will be exposed to T cells (TCEM), and others will be oriented inwards toward the HLA molecule (groove-exposed motif). For a 15-mer peptide in a pHLA complex, we numbered the amino acid residues from −3 to 12. Utilizing the work of Rudolph et al. (50) and Calis et al. (51), as previously described (37) three different types of non-continuous TCEM were deduced and designated TCEM I (amino acid residues 4, 5, 6, 7, 8 of a 9-mer), TCEM IIa (2, 3, 5, 7, 8 of a 15-mer with a core 9-mer), and TCEM IIb (−1, 3, 5, 7, 8 of a 15-mer with a core 9-mer). The latter two are relevant for HLA class II predictions, the former for HLA class I predictions. Analysis of the models of Rudolph et al. (50) indicate that TCEM IIa or TCEM IIb motifs occur in approximately equal proportions of TCR:MHC class II structures. All possible TCEM patterns were identified for all IGHV fragments.

Rare IGVH sequences may escape tolerance and be stimulatory under the right circumstances (52). TCEM frequencies are unequally distributed in the IGHV region and also elsewhere in the proteome as some motifs occur far more frequently than others (37). Their frequency in the IGHV region can be assessed by assigning a frequency class (FC), which is a reverse log2 scale where FC 0 ($1/2^0$) corresponds to "occurring in every IGHV sequence" and FC 21 ($1/2^{21}$) to "occurring once every approx. 2 million sequence" (37). In this study, we considered a TCEM with FC above 16 (once every 65,536 sequence) as rare.

To compare the mean FC of our patients TCEMs, we compiled a FC classification system using a public database of 37 million unique BCRs spanning the FW3 and CDR3 from memory and naive B cells from three healthy donors published by Dewitt et al. (42, 43), consistent with a previously published classification based on 56,000 sequences from Genbank (37). The IGVH sequences of this database are of the same length and were established with similar technology as those from our patients, thereby minimizing technical or disease-related bias. All TCEMs occurring in the dataset were assigned a FC class based on their mean patient and compartment-specific −log2 frequencies.

The FC as a measure of presumed likelihood for IGVH sequences to escape tolerance does not take into account the possible occurrence of similar TCEMs elsewhere in the human proteome or in the gut microbiome. While the human proteome is common for patients, and can to some degree be accounted for, the gut microbiome displays variations across populations and ages (53). For this publication, we used databases of TCEM occurrences in the human proteome (assembled from UniProt, with removal of Ig variable regions) (54) and microbiome (from NIH Human Microbiome Project Reference Genomes database) (55) as described previously (56), by searching for all 3.2 million possible variations of each TCEM.

Each TCEM occurs at a characteristic frequency in proteomes. These frequencies were then normalized to a zero mean unit variance scale using Johnson SI scale transformation of log2 frequency values.

Validation of Prediction Analyzes

We have previously derived two monoclonal antibodies from CSF B cells of two MS patients (CSF mAbs), and demonstrated that one idiotope from each of these mAbs (pMS1 and pMS2) was both processed, presented on HLA class II molecules and recognized by cloned $CD4^+$ T cells *in vitro* (32, 33). These were therefore suitable for validation of the prediction analyzes. Cathepsin cleavage, HLA affinity, and TCEM occurrence were predicted as for the main dataset. The FC was calculated using a previously described dataset comprising the complete IGVH region (37, 56). Idiotope pMS1-VH1 was presented on DRB1*13:02 and pMS2-VH3 on DRB1*13:01 encoded HLA molecules. As these have identical amino acid sequences, the affinity for DRB1*13:02 was predicted for both peptides.

Statistics

All predictive models were built by EigenBio using JMP® software version 12.1/13.0 (SAS Institute, Cary, NC, USA), by script processing. STATA v 14.1 (StataCorp LLC, TX, USA) and JMP® 12.1 were used for statistical analyses. Plots were created in JMP® 12.1. All plots displaying CDR3 relative positions are cropped to include ~99% of the IGHV fragments.

For bioinformatics processing purposes, to avoid end-effects in various algorithms, we added a standard immunoglobulin signal peptide and three amino acid sequence ("DTR") to the

beginning, and a 26 amino acid sequence derived from the IgG1 constant region ("GTLVTVSSASTKGPSVFPLAPSSKST") at the end of each IGHV sequence. Parts of these were retained as described below for the comparative statistical analyses. Our IGHV fragments were indexed by the position of the first cysteine determining the start of the CDR3 region, and changes due to mutations in the CDR3 region could influence both TCEM and affinity predictions at indexed positions even prior to position 0.

For statistical testing, we created three subsets of the IGHV sequences from our patients. For comparison of differences in mean FC and mean HLA affinities between MS and OIND patients within the CDR3, we used a subset limited to fragments with approximately half of the amino acids of the IGHV-fragment within the CDR3. This subset contained fragments starting at indexed position −7, and ended in the position where the fragment would contain eight amino acids of the added constant region. A second subset containing FW3 (spanning approximately amino acid 73–104 by IMGT numbering) and CDR3 regions were compared within MS and OIND patients. For this purpose, we used the patients dataset, with the addition of the "DTR" amino acid sequence at the start and "GTL" amino acid sequence at the end, and defined fragments as being influenced by CDR3 changes similarly as above, with a cutoff at position −7. In the third subset, we used a similar approach as for the second subset, and included the blood-derived sequences for comparisons between blood and CSF. No extra amino acid sequences were attached to the sequences of this dataset.

In the first subset, the differences between patients and controls, between low and high FC within patients and controls, and between IGHV4 and the other IGHV families were assessed by estimating the multilevel mixed effects model for each outcome variable. A multilevel approach was chosen because the data exhibit a three-level hierarchical structure, with levels for patient, relative CDR3 position, and clone. The intraclass correlation coefficient was calculated to assess the cluster effect on each level. The cluster effect was highest at patient-level for all variables. Clone-level demonstrated negligible or no cluster effect and was therefore not taken into account. Adjustment for cluster effect on CDR3 relative position-level caused convergence problems. Therefore, the differences between the categories were assessed by estimating a linear mixed model with fixed effects for factor variable and random intercepts for patients at each relative CDR3 position separately. Benjamini-Hochberg adjustment for multiple testing was applied within each outcome variable with acceptable false discovery rate (FDR) set to 20% (57).

In the second subset, the cluster effect on protein clone-level was also negligible or zero, and hence ignored. The cluster effect on patient- or relative position-level or both were present for most variables. A linear mixed model with fixed effect for factor defining the region was estimated. Random effects for either patients or relative position were included in the model. For variables with cluster effect on both levels, random effects for relative position nested within the patient were assessed, but as these were negligible only the models with single random effect were estimated.

No cluster effect on protein clone-level was found in the third subset. Intrapatient correlations were close to zero or not present. Consequently, comparisons between blood and CSF within patients and controls as well as within IGHV families for patients and controls separately were performed by independent samples *t*-test at each relative CDR3 position, and *p*-values were further adjusted for multiple testing by Benjamini-Hochberg procedure. As this was an exploratory study also aiming to guide further studies on the proposed mechanism, the FDR was set at 20%.

Finally, generalized linear models with random effects for patients were estimated to assess the differences in number of IGHV sequences containing fragments meeting a set of idiotope criteria between MS and OIND groups, as well as between highly transcribed IGHV sequences and other sequences.

The results are reported as mean differences between the groups with the corresponding 95% confidence intervals (CIs) and *p*-values.

RESULTS

IGHV Transcripts

From all patients a total of 1,812,920 IGHV amino acid sequences were deduced after removing non-productive transcripts, with a mean of 3,552 (95% CI 1,399–5,706) sequences obtained from CSF and 125,180 (95% CI 78,262–172,100) from blood. The mean CDR3 lengths of CSF sequences were 15.55 amino acids (95% 15.5–15.6, $n = 25{,}556$) for MS patients and 15.25 amino acids (95% CI 15.20–15.29, $n = 34{,}833$) for controls ($p < 0.001$, independent samples *t*-test). For blood, the corresponding lengths were 15.48 (95% CI 15.48–15.49, $n = 1{,}551{,}154$) for MS and 16.03 (95% CI 16.01–16.04, $n = 201{,}377$) for controls ($p < 0.001$, independent samples *t*-test). The IGHV gene family usage is shown in Table S3 in Supplementary Material. As reported previously by us and others (7, 58), there was a preferential use of IGHV4 in CSF from MS patients.

Most IGHV sequences and transcripts used in this project were previously published (7). However, in the present study we also included exceedingly rare sequences for the purpose of creating a thorough database of TCEM, resulting in a higher total number than previously described (7).

Cathepsin Cleavage

Due to the IGHV4 bias in CSF among the MS patients, we first investigated whether the predicted pattern of cathepsin cleavage differed across IGHV families. An analysis of variance by IGHV family yielded significant variations (Welch ANOVA F(6, 951.32/951.49/949.68) for cathepsins S, L, and B, respectively, $p < 0.0001$ for all). Cathepsin S was predicted to preferentially cleave IGHV4 derived sequences, cathepsin L was predicted preferentially to cleave those from IGHV5, and cathepsin B to preferentially cleave those from IGHV7. Interestingly, cathepsin S was also predicted to cleave IGHV3 sequences least efficiently (**Figure 2**; Table S4 in Supplementary Material).

As CDR3 is most diverse and therefore hypothesized to be the main source of immunogenic idiotopes, we further investigated whether the cathepsins were likely to release CDR3 fragments (**Figure 2B**). All three cathepsins displayed a similar overall pattern of cleavage sites in FW3 and at the start of CDR3.

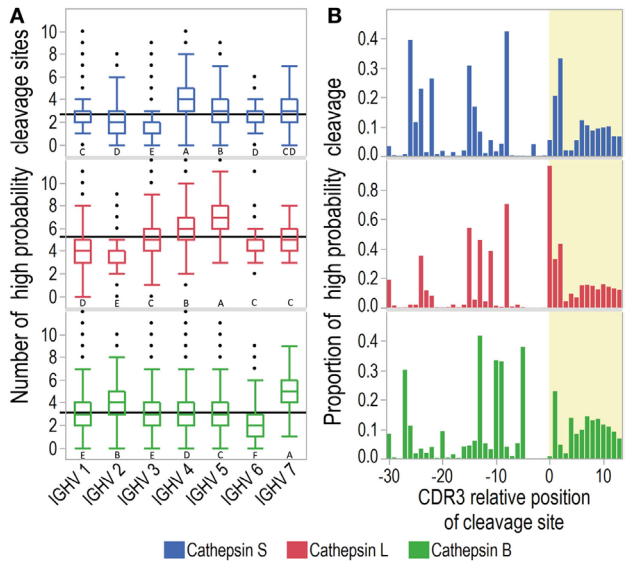


FIGURE 2 | Prediction of cathepsin cleavage. Cathepsin cleavage sites in immunoglobulin heavy chain variable (IGHV) transcripts were predicted with neural-net models. **(A)** Mean summarized numbers of predicted cleavage sites (>0.8 probability for cleavage) for transcripts of all IGHV families are shown as solid black lines and distributions as outlier box plots with whiskers covering first and third quartile ±1.5*(interquartile range). For each cathepsin, IGVH families not connected by the same letter are significantly different (Tukey Kramer HSD). **(B)** The proportion of transcripts with >0.8 probability of cleavage for each complementarity determining region 3 (CDR3) relative position. The CDR3 relative position aligned with the cleavage site at P1-P1′. CDR3 is marked with yellow shading.

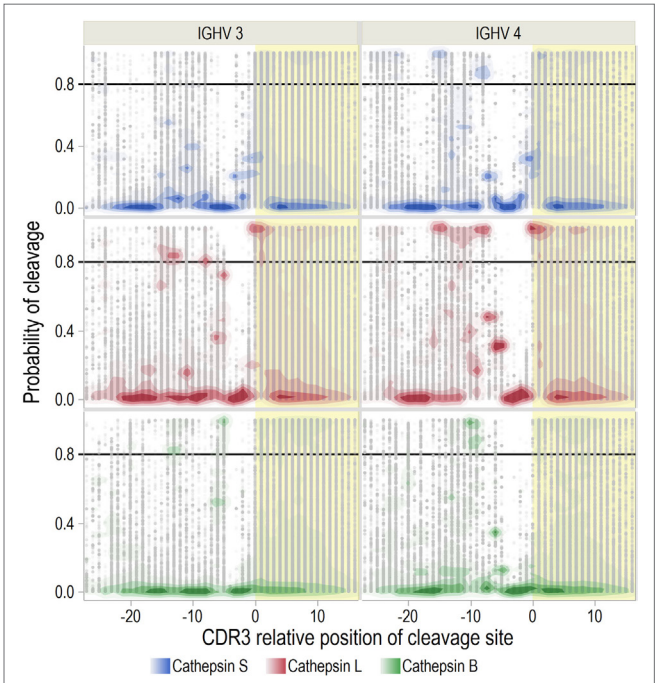


FIGURE 3 | Distribution of cleavage probabilities by complementarity determining region 3 (CDR3) relative position. Cathepsin cleavage sites in immunoglobulin heavy chain variable (IGHV) sequences were predicted with neural-net models. The distributions of predicted probabilities for cleavage (range 0–1) are shown for IGHV3 and 4 (see S5 in Supplementary Material for IGHV1, 2, and 5–7). Sites with predicted probability >0.8 are considered to have high probability for cleavage. The CDR3 region is marked with yellow shading and the CDR3 relative position is aligned with the cleavage site at P1-P1′.

Notably, cathepsin L was predicted to cleave almost all IGHV sequences before or after the cysteine marking the start of CDR3. Cathepsins S and B showed less pronounced peaks for cleavage at the same positions.

We next investigated whether these patterns differ across IGHV families (**Figure 3**; S5 in Supplementary Material). The previously identified hotspot for predicted cathepsin L activity at the CDR3 start was consistently found for all IGHV families. For cleavage of IGHV4 by cathepsin L and S, there were also two hotspots in FW3. No hotspot was identified for cathepsin S for IGHV3. As cathepsins S and L are endopeptidases that recognize octamers within the peptide, cleavage at these hotspots would effectively block other cleavages in the immediate vicinity. Therefore, these hotspots probably represent the most likely cleavage sites.

HLA Affinities

Different HLA molecules may display different binding affinities for IGHV fragments. Analyzing all CSF IGHV fragments together, fragments from CDR3 had consistently higher predicted affinities for HLA-DR and -DP than fragments from FW3 (Figure S6 in Supplementary Material). While DRB1*15:01 was among the DR molecules with the highest mean standardized affinities for CDR3-derived fragments, the same was not true for the linked DQA1*01:02-DQB1*06:02 among DQ molecules. In general, the predicted patterns of affinities for different HLA class II molecules were similar between MS and OIND patients (not shown).

We next investigated the IGVH sequences from CSF for binding affinity to MS-associated HLA molecules (**Figure 4**). CDR3 fragments from MS patients had higher predicted affinity compared to FW3 fragments for DRB1*15:01 and to a lesser extent for DQA1*01:02-DQB1*06:02, but lower affinity for A*02:01 ($p < 0.001$ for all, S7 in Supplementary Material). After correcting for intraclass correlations at patient-level, there were no significant differences between the MS and OIND patients in predicted affinity for these HLA molecules for any IGHV fragment in vicinity of the CDR3. Similarly, no significant differences were detected for MS or OIND patients when comparing highly transcribed to other IGHV sequences (data not shown). However, IGHV4 family fragments from MS and OIND patients had higher predicted affinity for DRB1*15:01 than other IGHV fragments at almost every position within the CDR3 (Table S8B in Supplementary Material). For DQA1*01:02-DQB1*06:02 and A*02:01, the results were similar to those for DRB1*15:01, except for IGHV4 where both higher and lower mean affinities were predicted within the CDR3 (Tables S8C,D in Supplementary Material).

Because each patient did not carry all HLA alleles, we extracted the affinity predictions for those carried by each individual. For heterozygous patients, we used the allele with the

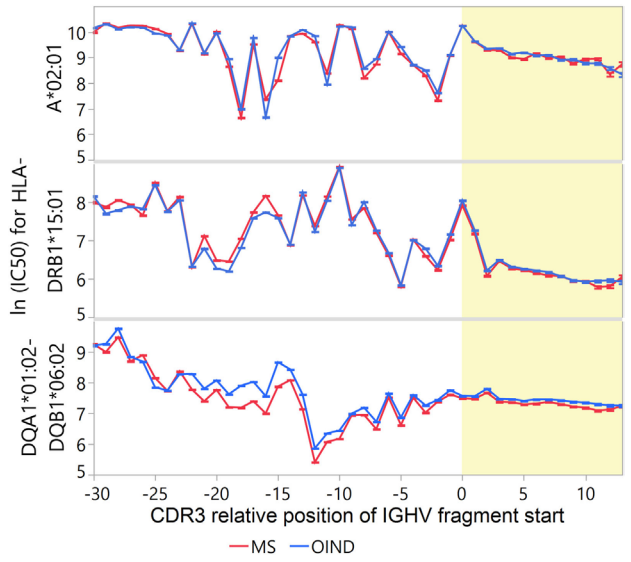


FIGURE 4 | Immunoglobulin heavy chain variable (IGHV) fragment affinities for multiple sclerosis (MS)-associated human leukocyte antigen (HLA) molecules. Binding affinities of cerebrospinal fluid (CSF) IGHV fragments were predicted for HLA-A*02:01, HLA-DRB1*15:01, and HLA-DQA1*01:02-DQB1*06:02 with neural-net models. Mean Johnson SI standardized values of ln(IC_{50}) were calculated for each CDR3 relative position, with low values indicating higher affinity. Yellow shade indicates the CDR3 region. Each error bar is constructed using a 95% confidence interval of the mean.

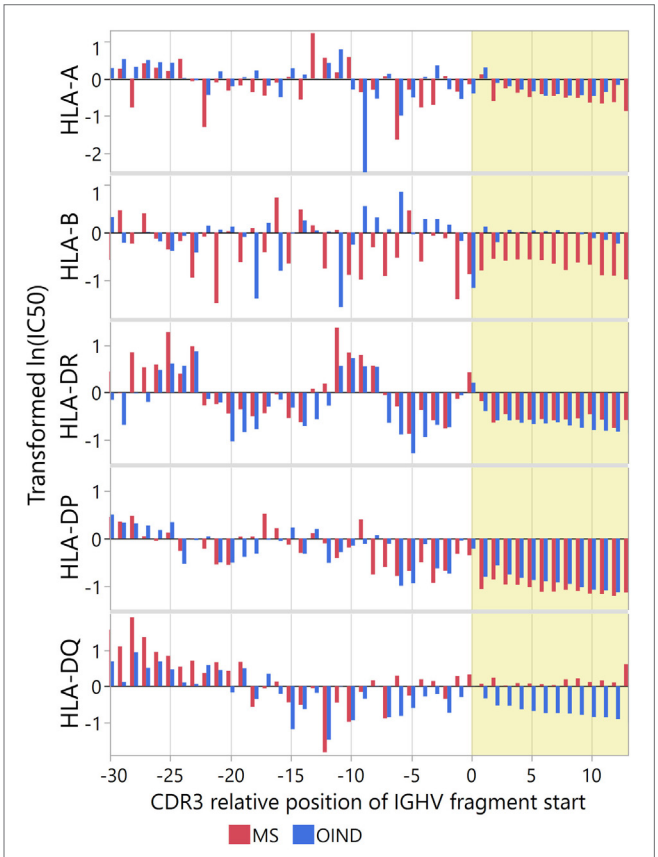


FIGURE 5 | Immunoglobulin heavy chain variable (IGHV) fragment affinities for patient-specific human leukocyte antigen (HLA) molecules. Binding affinities of IGHV fragments were predicted for patient-specific HLA A, B, DP, DR, and DQ molecules (listed in **Table 1**). Results are presented as mean Johnson SI standardized values of ln(IC_{50}) for each complementarity determining region 3 (CDR3) relative position, with low values indicating high affinity. For heterozygous patients with two sets of predictions for one HLA molecule, we used the lowest standardized ln(IC_{50}). Yellow shade indicates the CDR3 region.

highest predicted affinity (**Figure 5**). In general, the standardized predicted affinities were very similar for DR and DP, with high predicted affinity in CDR3 for both MS and OIND patients. For DQ on the other hand, only OIND patients followed this pattern. This most likely reflects that MS patients more frequently carry DQA1*01:02-DQB1*06:02, which we previously showed did not display the highest affinities within the CDR3.

TCEM Patterns

From all possible IGHV fragments (n = 52,566,906) we identified TCEM I, IIa and IIb and created databases of TCEM occurrences. We identified approximately 1.5 million unique TCEM of each pattern from a theoretical maximum of 3.2 million. These overlapped >98% with the TCEMs in the dataset derived from healthy individuals by DeWitt et al. (42), which could therefore be used for frequency classification in our dataset (**Table 2**).

Different occurrences of TCEM between populations and compartments could point to a selection process. The results of cluster analysis performed on pairwise correlations for the summarized occurrences of TCEM IIa in blood and CSF for each patient group are shown in **Figure 6**. The TCEM patterns of IGHV sequences from CSF of MS patients differed from those in blood and also from those of the OIND patients. There were two notable exceptions; TCEM from CSF of OIND-4 clustered consistently with that in CSF of MS patients. This patient was the only OIND patient with oligoclonal IgG bands (OCB) and IGHV4 dominance in the CSF. The other exception was MS-6, from whom the TCEM in CSF clustered with that in blood.

TABLE 2 | Number of unique TCEM identified in each dataset.

	Patient dataset[a]	Healthy BCR dataset	Overlap (%)[b]
TCEM I	1,448,203	2,747,840	98.64
TCEM IIa	1,439,194	2,736,223	98.62
TCEM IIb	1,403,685	2,710,199	98.64

[a]*Combination of blood and CSF derived IGHV fragments from both MS and OIND patients.*

[b]*Percentage of TCEM identified in the patients, also identified in the DeWitt et al. healthy BCR set.*

Notably, MS-6 was one of two MS patients with dominant IGHV3 use in CSF. TCEM I and IIb displayed similar patterns as TCEM IIa (data not shown). Corresponding results were found using standardized (z transformed) TCEM occurrences (S9 in Supplementary Material).

We next performed cluster analysis on pairwise correlations of TCEM occurrences by compartment and IGHV family.

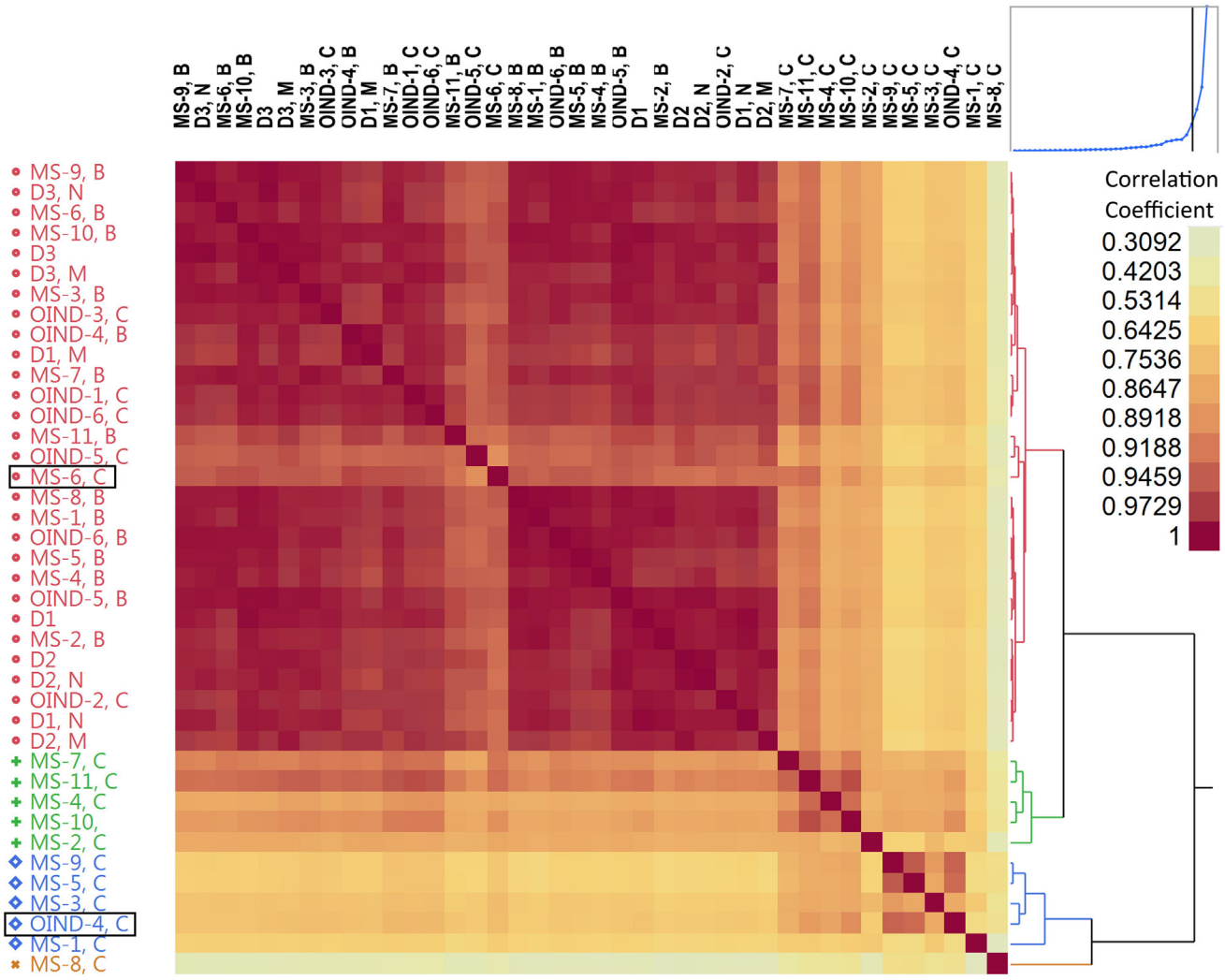


FIGURE 6 | Hierarchical cluster analysis of T cell-exposed motifs (TCEMs) by patient and compartment. The occurrences of all TCEM IIa in immunoglobulin heavy chain variable (IGHV) fragments were identified and summarized by patient and compartment in patients with multiple sclerosis (MS, MS 1–11), other inflammatory neurological disorders (OINDs, OIND 1–6), and in three healthy individuals (D1–3). Pairwise correlation coefficients were calculated for all possible pairs, and displayed as hierarchical cluster matrix. Blood samples from all individuals cluster together with OIND cerebrospinal fluid (CSF) samples (red), while MS CSF generally cluster differently (blue, green and orange). The two exceptions (MS-6 C and OIND-4 C) are marked with boxes. For D1-3, *N* denotes naive B cells from blood, and M denotes mature B cells from blood. For MS and OIND, B at the end of the identification code denotes blood and C denotes CSF.

This revealed clustering on IGHV family level (S10 in Supplementary Material), implying that either dominance of IGHV4 or relative lack of IGHV3 could be driving the differences in TCEM patterns. This also suggests that differences in number of IGHV sequences were not driving the TCEM clustering, as could be suspected because CSF samples generally clustered together. Similar results were obtained when analyzing samples by IGHV family, patient-ID and compartment. Again the samples clustered mainly by IGHV family and secondly by compartment (data not shown).

Frequency Classifications

As CSF IgG is more mutated than IgG from blood (7, 12, 15, 59), the TCEMs of IGHV sequences from CSF could be rarer than those from blood. We therefore compared FC distributions in blood vs. CSF (**Figure 7**; Tables S11A–F in Supplementary Material). Although most pronounced among the MS patients, the mean FC of TCEMs from the CDR3 was significantly higher in CSF at nearly every CDR3 relative positions for all three TCEM for both patient groups. The mean FC was also significantly higher in CSF for each IGHV family analyzed separately (data not shown).

Our analyzes predicted that CDR3 fragments were both likely to be released by cathepsins and to bind HLA class II molecules with higher affinities than FW3 fragments. Hence, we analyzed whether the CDR3 region generated more rare TCEMs. The CDR3 regions of CSF from both MS and OIND patients generated on average more rare TCEM than FW3 ($p < 0.001$ for

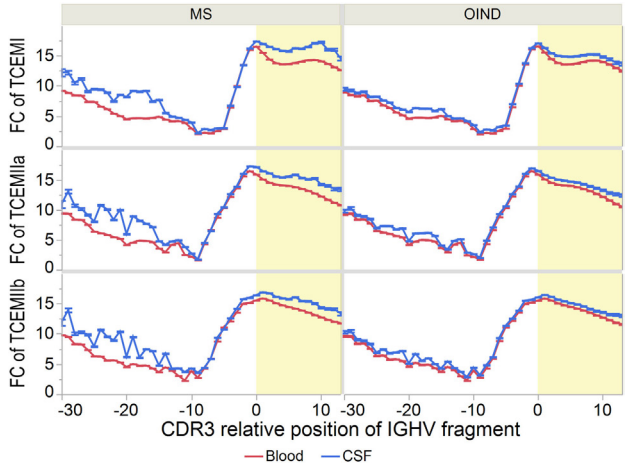


FIGURE 7 | Frequency of T cell-exposed motifs (TCEMs)—differences between blood and cerebrospinal fluid (CSF). Mean frequency class (FC) of TCEMs [−log2 values with 95% confidence interval (CI)] in immunoglobulin heavy chain variable (IGHV) fragments were calculated for all complementarity determining region 3 (CDR3) relative positions and compared between blood and CSF. High FC indicates rare TCEMs. In both multiple sclerosis (MS) and other inflammatory neurological disorders (OINDs) patients, fragments from CSF carry significantly more rare TCEMs than fragments from blood ($p < 0.001$, mixed-model comparisons, see Tables S11A–C in Supplementary Material for details). Yellow shading indicates CDR3.

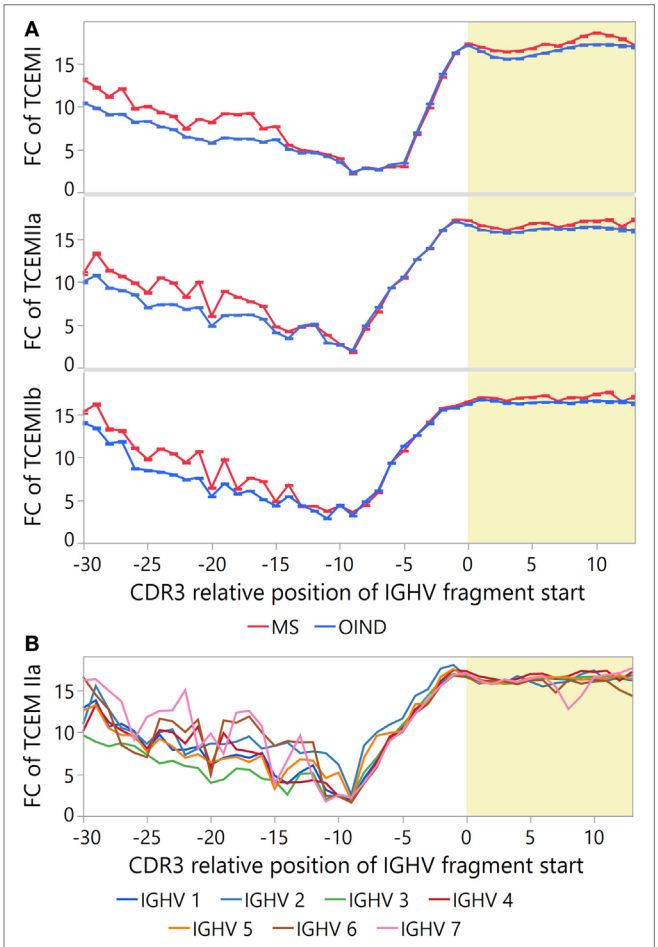


FIGURE 8 | Frequency of T cell-exposed motifs (TCEMs) in immunoglobulin heavy chain variable (IGVH) fragments from cerebrospinal fluid (CSF)—differences between patient groups and IGVH families. Mean frequency class (FC) of TCEMs [−log2 values with 95% confidence interval (CI)] in CSF IGHV fragments were calculated for all CDR3 relative positions, and differences are shown **(A)** between multiple sclerosis (MS) and other inflammatory neurological disorders (OINDs), and **(B)** by the IGHV families. High FC indicates rare TCEMs. Yellow shading indicates CDR3.

all TCEMs), probably driven by greater diversity in the CDR3 (Table S12 in Supplementary Material).

Because rare motifs are expected to be most likely to escape tolerance (52), we also tested whether CDR3 fragments from MS patients contain on average rarer motifs than those from OIND patients. After correcting for intra-class correlations in a multilevel hierarchical mixed model and multiple testing, the MS patients had significantly higher FC than the OIND patients at several CDR3 positions (**Figure 8A**; Tables S13B–D in Supplementary Material). This was most evident for TCEM I and IIa.

As IGHV4 family bias was found in MS CSF samples (Table S3 in Supplementary Material), we compared the FC of sequences carrying IGHV4 to other IGHV families (**Figure 8B**). Differences in FC between IGHV families were most pronounced prior to position −7. For CDR3, we found statistically higher FC at nearly all positions for sequences carrying IGHV4 than for sequences carrying any other IGHV family (Tables S13E–G in Supplementary Material). Thus, it seems that IGHV4 possess all the assumed prerequisites for T cell stimulation.

Summarized Attributes

To identify IGVH sequences most likely to engage in idiotope-driven T–B cell collaboration, we devised an "idiotope score" identifying IGHV fragments with both high patient-specific HLA-DRB1 standardized affinity [$\ln(IC_{50}) < -1.5$]; TCEM II (a or b) FC > 16; and predicted fuzzy cut "Excision" by cathepsin S (**Figure 9**). Although many IGHV fragments fulfill one of these criteria relatively few fulfill all, and these almost exclusively reside in the CDR3 region. In a few peaks in the CDR3, almost 10% of the IGHV fragments fulfilled all criteria. Among the MS patients, the highly transcribed CDR3 fragments in CSF were generally most likely to carry all these traits (**Figure 10**). Moreover, in MS patients highly transcribed IGHV sequences were more likely than others to carry at least one IGHV fragment with these attributes (OR = 1.39, $p = 0.01$ unadjusted model). However, after adjusting for cluster effect on patient-id-level the difference was no longer significant (OR = 1.24, $p = 0.11$). For OIND patients, we found no significant difference in unadjusted (OR = 1.20, $p = 0.38$) or adjusted models (OR 1.25, $p = 0.29$). Among highly transcribed IGHV sequences, 42% from the MS patients and 34% from the OIND patients carried at least one fragment fulfilling all criteria (OR = 1.41, $p = 0.16$; adjusted on patient-id-level: OR = 1.40, $p = 0.27$).

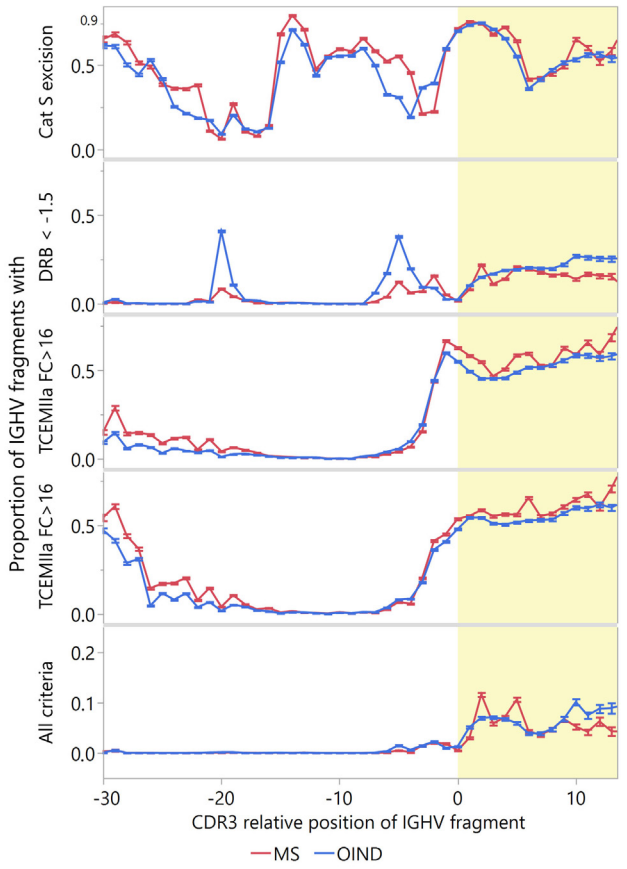


FIGURE 9 | Proportion of immunoglobulin heavy chain variable (IGVH) fragments that could engage in idiotope-driven T–B cell collaboration. The idiotope score identifies IGHV fragments most likely to engage in idiotope-driven T–B cell collaboration in context of human leukocyte antigen (HLA)-DR molecules [excised: predicted 15-mer release by cathepsin S; high affinity: patient-specific Johnson standardized ln(IC_{50}) for HLA-DRB1 < −1.5; rare: frequency class (FC) of T cell-exposed motif (TCEM) > 16]. The upper panels show the proportion of fragments at each complementarity determining region 3 (CDR3) relative position that fulfills each criterion. The lower panel shows the proportion that fulfills all criteria. Nearly all fragments inhabiting all three features occur in the CDR3 region (yellow shading). Error bars indicate 95% confidence interval of the mean.

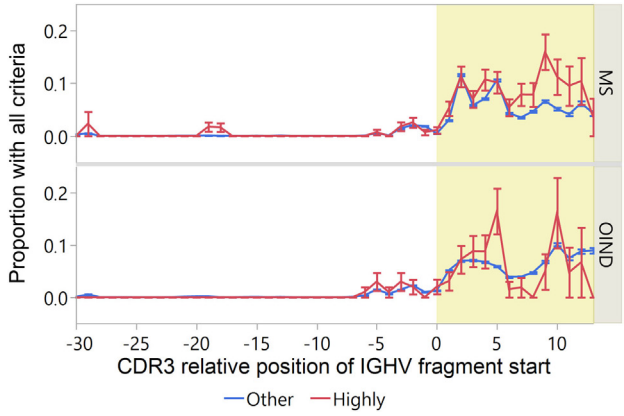


FIGURE 10 | Fragments from highly transcribed immunoglobulin heavy chain variable (IGVH) sequences from multiple sclerosis (MS) patients are most likely to meet the criteria for idiotope-driven T–B cell collaboration. The proportion of IGHV fragments meeting the requirements for idiotope driven T–B collaboration [excised: predicted 15-mer release by cathepsin S; high affinity: standardized human leukocyte antigen (HLA)-DRB affinity < −1.5; rare: frequency class (FC) of T cell-exposed motif (TCEM) > 16] is displayed by their complementarity determining region 3 (CDR3) relative position. CDR3 region yellow shaded. Error bars indicate 1 SEM.

TCEM in the Human Proteome and Gut Microbiome

By plotting the mean Johnson standardized frequencies of TCEM found in the human proteome and gut microbiome in a similar fashion as the immunoglobulin FC scale, we found that the CDR3 regions generate TCEM that generally occur rarely in both the gut microbiome and the human proteome (S14 and S15 in Supplementary Material). There was a significantly lower standardized mean occurrence of TCEM in the CDR3 than FW3 for both gut microbiome ($p < 0.001$) and the human proteome ($p < 0.001$). Hence it seems that somatic hypermutation and recombination in the CDR3 is capable of generating TCEM that are rare, both in the healthy IGHV repertoire and in the human proteome and gut microbiome.

Validation

To validate the epitope prediction, we performed prediction analysis of the V regions of two CSF mAbs previously derived from CSF B cells of two MS patients (32). We have previously demonstrated *in vitro* that each VH region of these CSF mAbs carry an idiotope (pMS1-VH1 and pMS2-VH3) that was processed by APCs, and presented on DRB1*13:01/13:02 molecules to $CD4^+$ T cells which specifically recognized the particular idiotope (32, 33). As shown in **Figure 11** prediction analysis anticipated that both peptides were likely to be cleaved at positions allowing for presentation by the relevant HLA-DR molecule, and that they would bind their restriction element (DRB1*13:01/13:02) with high affinity. The FCs had to be calculated for the complete VH region and are therefore not directly comparable to those used for the complete dataset, but nevertheless shows that the TCEM associated with these idiotopes are rare.

DISCUSSION

We have hypothesized that idiotope-driven T–B cell collaboration may drive the intrathecal immune response in MS (19, 33). Although proof of concept studies have provided some evidence compatible with this hypothesis (31–33), the immense diversity of the immune repertoire have previously precluded further analyses. Here we combined high-throughput sequencing of IGHV transcripts with *in silico* prediction analyses to assess whether the requirements for such T–B cell collaboration exist. Our findings indicate that idiotopes from the CDR3 regions of MS patients on average have high affinities for disease associated HLA-DRB1*15:01 molecules and are predicted to be endosomally processed by cathepsin S and L in positions that allows such HLA binding to occur. Additionally, CDR3 sequences from

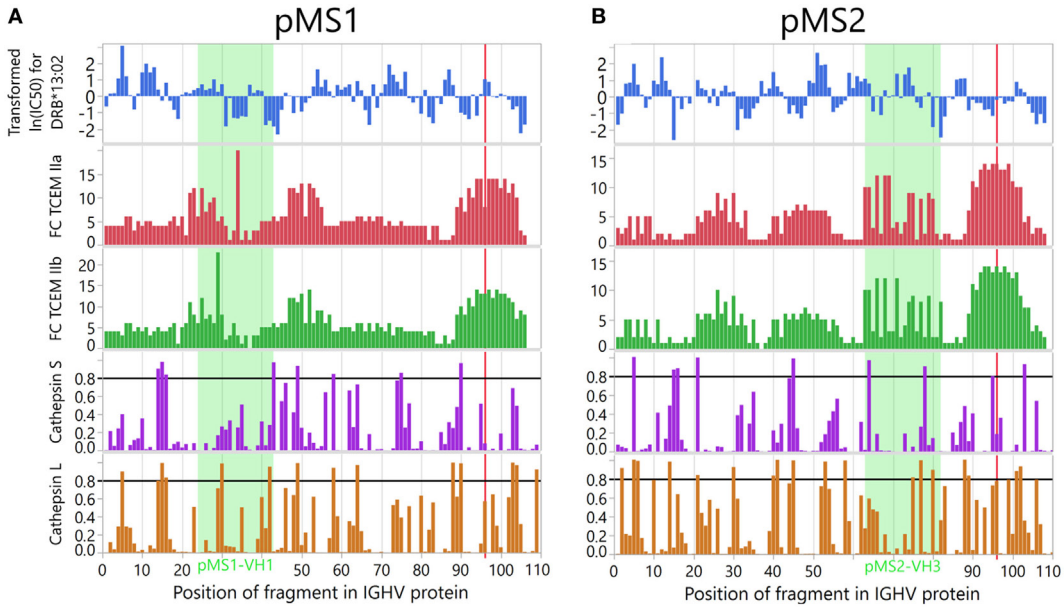


FIGURE 11 | *In vitro* validation of *in silico* prediction analyses. The peptides **(A)** pMS1-VH1 and **(B)** pMS2-VH3 are derived from V regions of immunoglobulin G (IgG) produced by cerebrospinal B cells of two multiple sclerosis (MS) patients and have been shown *in vitro* to be processed and presented on human leukocyte antigen (HLA) DRB1*13:01/13:02 molecules to idiotope-specific CD4+ T cells. Cathepsin cleavage, HLA affinity and T cell-exposed motif (TCEM) occurrence were predicted as for the main dataset, for each position within the immunoglobulin heavy chain variable (IGHV) transcript. The frequency classes (FCs) were obtained using a dataset comprising the full length VH region, and are therefore not directly comparable to those calculated for the main dataset. The cysteine at the start of complementarity determining region 3 (CDR3) is marked with a red line.

CSF B cells from MS patients contain on average rarer TCEM that could potentially stimulate non-tolerant CD4+ T cells, than corresponding sequences from OIND patients. Many of these features were associated with the previously described IGHV4 gene family bias (7, 58, 59) of CSF B cells in MS, indicating a possible explanation for this previously unexplained predominance. The IGHV gene family distribution with IGHV3-gene family dominance in blood also correlated with previously published results from healthy individuals (37, 42, 60).

Cathepsins S and L are essential for antigen presentation on MHC class II molecules (61, 62), including both processing of invariant chain (Ii) and preparation of protein fragments expressed on MHC class II molecules (63). Cathepsin S was shown to be particularly important for endosomal processing in B cells, dendritic cells and macrophages, while cathepsin L was essential for the Ii chain processing in cortical thymic epithelial cells and in macrophages (64, 65). Both these cathepsins are therefore likely relevant for endosomal processing of idiotopes. Cathepsin S was also implicated in MS by early expression studies suggesting a possible disease association (66), but this has not been replicated in more recent GWAS studies (67). The cathepsin S (*CTSS*) gene was further reported to be associated with treatment responses of both glatiramer acetate and IFN-beta (68). It was shown that cathepsin S could cleave myelin basic protein as a possible mechanism of action (69), and a cathepsin S-like helminthic protease was efficient in cleaving IgG (70). Even earlier publications showed that human lysosomal proteases can cleave IgG at acidic and to a lesser extent at neutral pH (71). Cathepsin S and L have similar cleavage patterns in general (72). Our cleavage predictions are in line with these findings and indicate that cathepsin S especially may be important for endosomal processing of BCRs in a way that allows idiotopes of the CDR3 region to be released.

Three observations point to CDR3 as crucial for generation of idiotopes capable of stimulating CD4+ T cells: First, cathepsins S and L both displayed high probability of cleavage around the CDR3 start; second, these cleavage spots were immediately followed by regions with high predicted affinity for HLA class II molecules; and third, the same region was associated with relatively high mean FC values for all TCEM, implying high likelihood for T cell stimulation.

It has been shown that IGHV fragments from endogenous IgG are indeed processed and presented on MHC II molecules (20). However, while the effects of cathepsin S and L on foreign antigens and Ii-chain have been well characterized (73), there have been no specific studies to our knowledge on how human cathepsins S and L act on Ig variable regions in endosomal conditions. The predictive models were built on data from proteomic identification of cleavage sites assays (45), where cleavage sites are readily available in preprocessed polypeptide cocktails (45, 46). Native IgG molecules, on the other hand, contain disulfide bonds that could interfere with cathepsin activity (74). Future studies addressing these questions would be important for validating our predictions.

B cells are likely to play a role as APCs in MS (18), but which antigens they present are unclear. B cells are capable of processing

and presenting their own Ig, as demonstrated almost 30 years ago in mice (20, 22), and recently shown to occur in large scale on HLA class II molecules in humans with mantle cell lymphoma (75). Such presentation would normally be enhanced by BCR stimulation, which activates B cells and induces proper antigen presentation potential (76). An alternative way for the B cell to upregulate their HLA class II expression and antigen presenting potential, was shown in mice to occur through CD40 stimulation in the thymus (77), as thymus B cells upregulated AIRE, HLA class II and CD80 in a CD40 dependent (and BCR independent) mechanism. It is not known whether B cells from the CNS or CSF of MS patients express AIRE.

Our results of predicted high affinity for HLA class II molecules for CDR3-derived fragments are consistent with findings in the Genbank IGHV dataset from healthy donors, including the different patterns observed for HLA-DQ (37). Notably peptides derived from the CDR3/FW3 region studied here were recently shown to be extensively presented on HLA class II in human mantle cell lymphoma B cells, and also recognized by idiotope-specific $CD4^+$ T cells (75). Interestingly, for the single lymphoma patient with available HLA data and T cell specificity (75), our predictions confirmed high affinity of the eluted VH peptide to HLA-DRB1*04:01 molecules [predicted $\ln(IC_{50}) = 3.69$, SD = 0.69]. Along with our results, this implies that the diversity of this region either generates on average higher affinity peptides, or that some idiotopes are selected for their affinity to HLA class II. As our sequences did not span the whole Ig variable region, we were unable to compare affinities for peptides derived from CDR3 and FW3 with those derived from CDR1, −2 and other FW regions. However, the recent results from human mantle cell lymphoma suggest that the most relevant part of the Ig molecule for idiotope-driven T–B cell collaboration were included in this study (75).

The concept of idiotope-driven T–B cell collaboration is founded on an idea of lack of tolerance for the IGHV region, but it is not fully known to which extent such tolerance occurs (52). Central T cell tolerance is mediated through positive and negative selection within the thymus, with help of cortical and medullary thymic epithelial cells (mTECs) as well as dendritic cells (78). During negative selection T cells are exposed to self-peptides by mTECs with help of promiscuous gene expression regulated by the autoimmune regulator AIRE protein, leading to either clonal deletion or induction of regulatory T cells (78). However, V (D) J recombinations of the IGHV genes only occur in B cells, and mTECs presumably are unable to present the huge number of idiotopes resulting from this process. Yet it was shown that T cells are likely tolerant to germline-encoded (non-mutated) IGHV regions (23, 79). It is possible that this could be mediated through circulating Ig, as very high concentrations of monoclonal Ig can induce tolerance through clonal deletion in the thymus (80, 81). Recent studies have found that both naive and class-switched B cells in the thymus of mice are of peripheral origin and capable of AIRE-induced antigen presentation (77, 82), even without BCR stimulation. This could provide another explanation as to how B cells can generate tolerance in the thymus, but the studies did not investigate to what extent such B cells present their own IGHV regions. Also, only a few B cells are present in the thymus at any time (77), providing a relatively small pool of BCRs to generate tolerance.

Our study utilizes TCEM as a model for how TCR interact with pHLA. Rudolph et al. described how only a few amino acid residues of the peptide in a pHLA complex interact with TCRs (50). These observations were then used to deduce the atomic contacts of motifs exposed to T cells (37). As HLA class II TCEMs are non-linear, matching TCEM can appear in context of both high and low affinity peptides. The TCEM model is applicable to any protein in a pHLA complex, and we expect TCEM occurring in the human proteome to be associated with tolerance if they are presented in context of HLA (56). A logarithmic scale of TCEM frequency classification (FC system) was developed and described for an IGHV repertoire from healthy individuals, and it was shown that each TCEM has a characteristic frequency of use. Some TCEM occur very frequently in IGHV regions (low FC), while others are incredibly rare (high FC) (37). The observation that some TCEM are present in every single, second, fourth, etc., IGHV sequence could possibly explain why relatively few thymic B cells may induce central tolerance for a substantial proportion of the IGHV repertoire (37). In agreement with previous findings of high diversity in transcribed CDR3 regions from CSF B cells (7), we found here that the CDR3 sequences from CSF contained on average quite rare TCEM, compatible with high likelihood of escaping tolerance. Moreover, the finding that CSF B cells have higher mean FC (more rare) TCEM than blood B cells is compatible with the notion that B cells are selected into the intrathecal compartment for their ability to stimulate idiotope-specific T cells. Finally, our comparison of TCEM in the IGHV sequences and the human proteome and gut microbiome found that CDR3-derived IGHV fragments more frequently carried TCEM that were rare in both these compartments, again suggesting that they would be more likely to escape tolerance. It was previously shown that neither the gut microbiome nor the human proteome cover the entirety of TCEM diversity (56), allowing for occurrence of many TCEM unique to the IGHV regions.

In agreement with the recently discovered lymph drainage of the CNS (83), B cells in the brain of MS patients seem to mature in cervical lymph nodes (8), and B cells in the CSF are clonally related to those in blood (7, 59). B cells may also proliferate within ectopic lymphoid follicles within the CNS of MS patients with long-standing disease (84). Hence, maturation necessary for idiotope-driven T–B cell collaboration could occur both in the periphery and within the CNS (8, 59). Such maturation may increase CDR3 variability and influence any of the parameters investigated in this study, for instance mutations that generate rare TCEM could also influence HLA affinity or cathepsin cleavage.

There are several limitations to this study. The number of included patients was quite low. Most importantly the findings *in silico* study needs to be further validated *in vitro.* In this study, the controls were restricted to OIND patients. To address whether the proposed mechanism drives the intrathecal synthesis of oligoclonal IgG in MS, MS patients without evidence of this phenomenon could be informative. Such patients are however rare, and can be hard to identify as the absence of two or more OCB bands by routine methods not necessarily rules out intrathecal synthesis of oligoclonal Ig (85–88). Moreover, the

proposed hypothesis does not exclude that individuals without intrathecal synthesis of oligoclonal IgG have a repertoire of idiotope-matched T–B cell pairs, but rather that a break of immune tolerance against self-IgG leads to dysregulated idiotope-driven T–B cell collaboration (19). This corresponds to other candidate autoantigens in MS, as myelin-specific T cells in the blood of healthy individuals are a frequent finding (89). Moreover, T cell responses against self-IgG is not unique for MS, but have previously been shown in patients with other inflammatory diseases such as systemic lupus erythematosus (90–92), granulomatosis with polyangiitis (Wegener's granulomatous) (93), and rheumatoid arthritis (94).

CONCLUSION

The overwhelming complexity of the immune repertoire calls for novel approaches to chart the interactions between immune receptors. This is the first work combining high-throughput sequencing of the IGHV transcriptome with *in silico* predictions analysis for T cell activation in a human disease. We predict that the three proposed prerequisites (successful endosomal processing, high HLA class II affinity and sufficiently rare TCEM) for idiotope-driven T–B cell collaboration are likely to occur in the CDR3 region in the CSF of MS patients, with as many as 42% of the highly transcribed IGHV sequences possess at least one segment with these features.

ETHICS STATEMENT

All participants provided written informed consent for participation. The study was approved by the Committee for Research Ethics in the South-Eastern Norwegian Healthy Authority (REK Sør-Øst S-04143a), the Norwegian Social Science Data Services (no. 11069) and the review boards at AHUS and OUS.

AUTHOR CONTRIBUTIONS

RH contributed with dataset preparations, statistical designs, analysis and interpretation of the data, and drafting the manuscript. AL contributed by sample acquisition, interpreting the data, and writing the manuscript. JJ contributed with collection and preparation of samples and provided the immunosequence patient database and revising the manuscript. JH contributed by interpreting the data and revising the manuscript. JB contributed by designing and performing statistical experiments, the interpretation of these and revising the manuscript. BB contributed by developing the concept and design of the study, as well as revising the manuscript. HR contributed by design of immunosequencing techniques as well as revising the manuscript. RB contributed by designing the bioinformatics algorithms, preparing datasets, interpreting the data and revising the manuscript. TH contributed by designing the study, interpreting the data, and writing the manuscript.

ACKNOWLEDGMENTS

We would like to thank all patients that have participated in this study. We also thank Marte K. Viken at the Department of Immunology and Transfusion Medicine, Faculty of Medicine, University of Oslo and Oslo University Hospital, Rikshospitalet, Oslo, Norway, for the kind help with HLA typing. The authors also would like to thank Alleen Hager for assistance in computational processing. Some of the data appearing in this manuscript was first published as an abstract and poster (P435) at ECTRIMS 2016, London.

FUNDING

The study was supported with a grant from the Norwegian Research Council (FriMedBio) (grant/project number 250864/F20) and Akershus University Hospital internal strategic funds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fimmu.2017.01255/full#supplementary-material.

REFERENCES

1. Compston A, Coles A. Multiple sclerosis. *Lancet* (2008) 372(9648):1502–17. doi:10.1016/s0140-6736(08)61620-7
2. Hauser SL. The Charcot Lecture | beating MS: a story of B cells, with twists and turns. *Mult Scler* (2015) 21(1):8–21. doi:10.1177/1352458514561911
3. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med* (2017) 376(3):221–34. doi:10.1056/NEJMoa1601277
4. Hohlfeld R, Meinl E. Ocrelizumab in multiple sclerosis: markers and mechanisms. *Lancet Neurol* (2017) 16(4):259–61. doi:10.1016/S1474-4422(17)30048-0
5. Baker D, Marta M, Pryce G, Giovannoni G, Schmierer K. Memory B cells are major targets for effective immunotherapy in relapsing multiple sclerosis. *EBioMedicine* (2017) 16:41–50. doi:10.1016/j.ebiom.2017.01.042
6. Obermeier B, Mentele R, Malotka J, Kellermann J, Kumpfel T, Wekerle H, et al. Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis. *Nat Med* (2008) 14(6):688–93. doi:10.1038/nm1714
7. Johansen JN, Vartdal F, Desmarais C, Tutturen AE, de Souza GA, Lossius A, et al. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. *Clin Immunol* (2015) 160(2):211–25. doi:10.1016/j.clim.2015.06.001
8. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med* (2014) 6(248):248ra107. doi:10.1126/scitranslmed.3008879
9. Kinzel S, Weber MS. B cell-directed therapeutics in multiple sclerosis: rationale and clinical evidence. *CNS Drugs* (2016) 30(12):1137–48. doi:10.1007/s40263-016-0396-6
10. Cross AH, Stark JL, Lauber J, Ramsbottom MJ, Lyons JA. Rituximab reduces B cells and T cells in cerebrospinal fluid of multiple sclerosis patients. *J Neuroimmunol* (2006) 180(1–2):63–70. doi:10.1016/j.jneuroim.2006.06.029
11. Kabat EA, Moore DH, Landow H. An electrophoretic study of the protein components in cerebrospinal fluid and their relationship to the serum proteins. *J Clin Invest* (1942) 21(5):571–7. doi:10.1172/JCI101335

12. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CCA, Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol* (1999) 163(9):5133–44.
13. Owens GP, Bennett JL, Lassmann H, O'Connor KC, Ritchie AM, Shearer A, et al. Antibodies produced by clonally expanded plasma cells in multiple sclerosis cerebrospinal fluid. *Ann Neurol* (2009) 65(6):639–49. doi:10.1002/ana.21641
14. Winges KM, Gilden DH, Bennett JL, Yu X, Ritchie AM, Owens GP. Analysis of multiple sclerosis cerebrospinal fluid reveals a continuum of clonally related antibody-secreting cells that are predominantly plasma blasts. *J Neuroimmunol* (2007) 192(1–2):226–34. doi:10.1016/j.jneuroim.2007.10.009
15. Owens GP, Ritchie AM, Burgoon MP, Williamson RA, Corboy JR, Gilden DH. Single-cell repertoire analysis demonstrates that clonal expansion is a prominent feature of the B cell response in multiple sclerosis cerebrospinal fluid. *J Immunol* (2003) 171(5):2725. doi:10.4049/jimmunol.171.5.2725
16. Monson NL, Brezinschek H-P, Brezinschek RI, Mobley A, Vaughan GK, Frohman EM, et al. Receptor revision and atypical mutational characteristics in clonally expanded B cells from the cerebrospinal fluid of recently diagnosed multiple sclerosis patients. *J Neuroimmunol* (2005) 158(1–2):170–81. doi:10.1016/j.jneuroim.2004.04.022
17. Obermeier B, Lovato L, Mentele R, Bruck W, Forne I, Imhof A, et al. Related B cell clones that populate the CSF and CNS of patients with multiple sclerosis produce CSF immunoglobulin. *J Neuroimmunol* (2011) 233(1–2):245–8. doi:10.1016/j.jneuroim.2011.01.010
18. Blauth K, Owens GP, Bennett JL. The ins and outs of B cells in multiple sclerosis. *Front Immunol* (2015) 6:565. doi:10.3389/fimmu.2015.00565
19. Holmoy T, Vartdal F, Hestvik AL, Munthe L, Bogen B. The idiotype connection: linking infection and multiple sclerosis. *Trends Immunol* (2010) 31(2):56–62. doi:10.1016/j.it.2009.11.001
20. Weiss S, Bogen B. MHC class II-restricted presentation of intracellular antigen. *Cell* (1991) 64(4):767–76. doi:10.1016/0092-8674(91)90506-T
21. Weiss S, Bogen B. B-lymphoma cells process and present their endogenous immunoglobulin to major histocompatibility complex-restricted T cells. *Proc Natl Acad Sci U S A* (1989) 86(1):282–6. doi:10.1073/pnas.86.1.282
22. Bogen B, Malissen B, Haas W. Idiotope-specific T cell clones that recognize syngeneic immunoglobulin fragments in the context of class II molecules. *Eur J Immunol* (1986) 16(11):1373–8. doi:10.1002/eji.1830161110
23. Eyerman MC, Zhang X, Wysocki LJ. T cell recognition and tolerance of antibody diversity. *J Immunol* (1996) 157(3):1037–46.
24. Bogen B, Weiss S. Processing and presentation of idiotypes to MHC-restricted T cells. *Int Rev Immunol* (1993) 10(4):337–55. doi:10.3109/08830189309061709
25. Munthe LA, Os A, Zangani M, Bogen B. MHC-restricted Ig V region-driven T-B lymphocyte collaboration: B cell receptor ligation facilitates switch to IgG production. *J Immunol* (2004) 172(12):7476–84. doi:10.4049/jimmunol.172.12.7476
26. Mitchison NA. The carrier effect in the secondary response to hapten-protein conjugates. II. Cellular cooperation. *Eur J Immunol* (1971) 1(1):18–27. doi:10.1002/eji.1830010104
27. Lanzavecchia A. Antigen-specific interaction between T and B cells. *Nature* (1985) 314(6011):537–9. doi:10.1038/314537a0
28. Rudensky A, Preston-Hurlburt P, al-Ramadi BK, Rothbard J, Janeway CA Jr. Truncation variants of peptides isolated from MHC class II molecules suggest sequence motifs. *Nature* (1992) 359(6394):429–31. doi:10.1038/359429a0
29. Chicz RM, Urban RG, Gorga JC, Vignali DA, Lane WS, Strominger JL. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J Exp Med* (1993) 178(1):27–47. doi:10.1084/jem.178.1.27
30. Munthe LA, Corthay A, Os A, Zangani M, Bogen B. Systemic autoimmune disease caused by autoreactive B cells that receive chronic help from Ig V region-specific T cells. *J Immunol* (2005) 175(4):2391–400. doi:10.4049/jimmunol.175.4.2391
31. Holmoy T, Vandvik B, Vartdal F. T cells from multiple sclerosis patients recognize immunoglobulin G from cerebrospinal fluid. *Mult Scler* (2003) 9(3):228–34. doi:10.1191/1352458503ms906oa
32. Holmoy T, Fredriksen AB, Thompson KM, Hestvik AL, Bogen B, Vartdal F. Cerebrospinal fluid T cell clones from patients with multiple sclerosis: recognition of idiotopes on monoclonal IgG secreted by autologous cerebrospinal fluid B cells. *Eur J Immunol* (2005) 35(6):1786–94. doi:10.1002/eji.200425417
33. Hestvik AL, Vartdal F, Fredriksen AB, Thompson KM, Kvale EO, Skorstad G, et al. T cells from multiple sclerosis patients recognize multiple epitopes on self-IgG. *Scand J Immunol* (2007) 66(4):393–401. doi:10.1111/j.1365-3083.2007.01955.x
34. Hestvik AL, Skorstad G, Vartdal F, Holmoy T. Idiotope-specific CD4(+) T cells induce apoptosis of human oligodendrocytes. *J Autoimmun* (2009) 32(2):125–32. doi:10.1016/j.jaut.2009.01.004
35. Boyd SD, Crowe JE Jr. Deep sequencing and human antibody repertoire analysis. *Curr Opin Immunol* (2016) 40:103–9. doi:10.1016/j.coi.2016.03.008
36. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol* (2015) 36(11):738–49. doi:10.1016/j.it.2015.09.006
37. Bremel RD, Homan EJ. Frequency patterns of T-cell exposed amino acid motifs in immunoglobulin heavy chain peptides presented by MHCs. *Front Immunol* (2014) 5:541. doi:10.3389/fimmu.2014.00541
38. Larimore K, McCormick MW, Robins HS, Greenberg PD. Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* (2012) 189(6):3221–30. doi:10.4049/jimmunol.1201303
39. Hollenbach JA, Madbouly A, Gragert L, Vierra-Green C, Flesch S, Spellman S, et al. A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. *Immunogenetics* (2012) 64(8):559–69. doi:10.1007/s00251-012-0615-3
40. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol* (2012) 882:569–604. doi:10.1007/978-1-61779-842-9_32
41. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* (2008) 36(Web Server issue):W503–8. doi:10.1093/nar/gkn316
42. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. A public database of memory and naive B-cell receptor sequences. *PLoS One* (2016) 11(8):e0160853. doi:10.1371/journal.pone.0160853
43. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. *Data from: A Public Database of Memory and Naive B-Cell Receptor Sequences*. Dryad Data Repository (2016). doi:10.5061/dryad.35ks2
44. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* (2009) 37(Database issue):D1006–12. doi:10.1093/nar/gkn838
45. Biniossek ML, Nagler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J Proteome Res* (2011) 10(12):5363–73. doi:10.1021/pr200621z
46. Bremel RD, Homan EJ. Recognition of higher order patterns in proteins: immunologic kernels. *PLoS One* (2013) 8(7):e70115. doi:10.1371/journal.pone.0070115
47. Chicz RM, Urban RG, Lane WS, Gorga JC, Stern LJ, Vignali DAA, et al. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* (1992) 358(6389):764–8. doi:10.1038/358764a0
48. Bremel RD, Homan EJ. An integrated approach to epitope analysis II: a system for proteomic-scale prediction of immunological characteristics. *Immunome Res* (2010) 6:8. doi:10.1186/1745-7580-6-8
49. Bremel RD, Homan EJ. An integrated approach to epitope analysis I: dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. *Immunome Res* (2010) 6:7. doi:10.1186/1745-7580-6-7
50. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* (2006) 24:419–66. doi:10.1146/annurev.immunol.23.021704.115658
51. Calis JJ, de Boer RJ, Kesmir C. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. *PLoS Comput Biol* (2012) 8(3):e1002412. doi:10.1371/journal.pcbi.1002412
52. Bogen B, Ruffini P. Review: to what extent are T cells tolerant to immunoglobulin variable regions? *Scand J Immunol* (2009) 70(6):526–30. doi:10.1111/j.1365-3083.2009.02340.x

53. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* (2012) 486(7402):222–7. doi:10.1038/nature11053
54. Consortium U. UniProt: a hub for protein information. *Nucleic Acids Res* (2015) 43(D1):D204–12. doi:10.1093/nar/gku989
55. The NIHHMPWG, Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al. The NIH human microbiome project. *Genome Res* (2009) 19(12):2317–23. doi:10.1101/gr.096651.109
56. Bremel RD, Homan EJ. Extensive T-cell epitope repertoire sharing among human proteome, gastrointestinal microbiome, and pathogenic bacteria: implications for the definition of self. *Front Immunol* (2015) 6:538. doi:10.3389/fimmu.2015.00538
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* (1995) 57(1):289–300. doi:10.2307/2346101
58. Owens GP, Kannus H, Burgoon MP, Smith-Jensen T, Devlin ME, Gilden DH. Restricted use of VH4 germline segments in an acute multiple sclerosis brain. *Ann Neurol* (1998) 43(2):236–43. doi:10.1002/ana.410430214
59. von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest* (2012) 122(12):4533–43. doi:10.1172/jci63842
60. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* (2010) 184(12):6986–92. doi:10.4049/jimmunol.1000445
61. Nakagawa TY, Rudensky AY. The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation. *Immunol Rev* (1999) 172:121–9. doi:10.1111/j.1600-065X.1999.tb01361.x
62. Villadangos JA, Bryant RA, Deussing J, Driessen C, Lennon-Dumenil AM, Riese RJ, et al. Proteases involved in MHC class II antigen presentation. *Immunol Rev* (1999) 172:109–20. doi:10.1111/j.1600-065X.1999.tb01360.x
63. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat Rev Immunol* (2015) 15(4):203–16. doi:10.1038/nri3818
64. Honey K, Rudensky AY. Lysosomal cysteine proteases regulate antigen presentation. *Nat Rev Immunol* (2003) 3(6):472–82. doi:10.1038/nri1110
65. Hsing LC, Rudensky AY. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol Rev* (2005) 207:229–41. doi:10.1111/j.0105-2896.2005.00310.x
66. Haves-Zburof D, Paperna T, Gour-Lavie A, Mandel I, Glass-Marmor L, Miller A. Cathepsins and their endogenous inhibitors cystatins: expression and modulation in multiple sclerosis. *J Cell Mol Med* (2011) 15(11):2421–9. doi:10.1111/j.1582-4934.2010.01229.x
67. Mahurkar S, Suppiah V, O'Doherty C. Pharmacogenomics of interferon beta and glatiramer acetate response: a review of the literature. *Autoimmun Rev* (2014) 13(2):178–86. doi:10.1016/j.autrev.2013.10.012
68. Foti Cuzzola V, Palella E, Celi D, Barresi M, Giacoppo S, Bramanti P, et al. Pharmacogenomic update on multiple sclerosis: a focus on actual and new therapeutic strategies. *Pharmacogenomics J* (2012) 12(6):453–61. doi:10.1038/tpj.2012.41
69. Beck H, Schwarz G, Schroter CJ, Deeg M, Baier D, Stevanovic S, et al. Cathepsin S and an asparagine-specific endoprotease dominate the proteolytic processing of human myelin basic protein in vitro. *Eur J Immunol* (2001) 31(12):3726–36. doi:10.1002/1521-4141(200112)31:12<3726::AID-IMMU3726>3.0.CO;2-O
70. Kong Y, Chung YB, Cho SY, Kang SY. Cleavage of immunoglobulin G by excretory-secretory cathepsin S-like protease of *Spirometra mansoni* plerocercoid. *Parasitology* (1994) 109(Pt 5):611–21. doi:10.1017/S0031182000076496
71. Fehr K, LoSpalluto J, Ziff M. Degradation of immunoglobulin G by lysosomal acid proteases. *J Immunol* (1970) 105(4):973–83.
72. Kirschke H. Cathepsin S A2. In: Salvesen G, Rawlings ND, editors. *Handbook of Proteolytic Enzymes*, Chap. 413. London: Academic Press (2013). p. 1824–30.
73. van Kasteren SI, Overkleeft HS. Endo-lysosomal proteases in antigen presentation. *Curr Opin Chem Biol* (2014) 23:8–15. doi:10.1016/j.cbpa.2014.08.011
74. Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* (2012) 4(1):17–23. doi:10.4161/mabs.4.1.18347
75. Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* (2017) 543(7647):723–7. doi:10.1038/nature21433
76. Lanzavecchia A. Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. *Annu Rev Immunol* (1990) 8:773–93. doi:10.1146/annurev.iy.08.040190.004013
77. Yamano T, Nedjic J, Hinterberger M, Steinert M, Koser S, Pinto S, et al. Thymic B cells are licensed to present self antigens for central T cell tolerance induction. *Immunity* (2015) 42(6):1048–61. doi:10.1016/j.immuni.2015.05.013
78. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol* (2014) 14(6):377–91. doi:10.1038/nri3667
79. Bogen B, Jorgensen T, Hannestad K. T helper cell recognition of idiotopes on lambda 2 light chains of M315 and T952: evidence for dependence on somatic mutations in the third hypervariable region. *Eur J Immunol* (1985) 15(3):278–81. doi:10.1002/eji.1830150313
80. Bogen B, Dembic Z, Weiss S. Clonal deletion of specific thymocytes by an immunoglobulin idiotype. *EMBO J* (1993) 12(1):357–63.
81. Lauritzsen GF, Hofgaard PO, Schenck K, Bogen B. Clonal deletion of thymocytes as a tumor escape mechanism. *Int J Cancer* (1998) 78(2):216–22. doi:10.1002/(SICI)1097-0215(19981005)78:2<216:AID-IJC16>3.0.CO;2-8
82. Yamano T, Steinert M, Klein L. Thymic B cells and central T cell tolerance. *Front Immunol* (2015) 6:376. doi:10.3389/fimmu.2015.00376
83. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. *Nature* (2015) 523(7560):337–41. doi:10.1038/nature14432
84. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. *Brain Pathol* (2004) 14(2):164–74. doi:10.1111/j.1750-3639.2004.tb00049.x
85. Jarius S, Eichhorn P, Franciotta D, Petereit HF, Akman-Demir G, Wick M, et al. The MRZ reaction as a highly specific marker of multiple sclerosis: re-evaluation and structured review of the literature. *J Neurol* (2017) 264(3):453–66. doi:10.1007/s00415-016-8360-4
86. Hassan-Smith G, Durant L, Tsentemeidou A, Assi LK, Faint JM, Kalra S, et al. High sensitivity and specificity of elevated cerebrospinal fluid kappa free light chains in suspected multiple sclerosis. *J Neuroimmunol* (2014) 276(1–2):175–9. doi:10.1016/j.jneuroim.2014.08.003
87. Poyraz T, Kaya D, Idiman E, Cevik S, Karabay N, Arslan D, et al. What does an isolated cerebrospinal fluid monoclonal band mean: a tertiary centre experience. *Neurology* (2015) 84(14 Suppl):P5.247.
88. Sandberg-Wollheim M, Vandvik B, Nadj C, Norrby E. The intrathecal immune response in the early stage of multiple sclerosis. *J Neurol Sci* (1987) 81(1):45–53. doi:10.1016/0022-510X(87)90182-1
89. Zhang J, Markovic-Plese S, Lacet B, Raus J, Weiner HL, Hafler DA. Increased frequency of interleukin 2-responsive T cells specific for myelin basic protein and proteolipid protein in peripheral blood and cerebrospinal fluid of patients with multiple sclerosis. *J Exp Med* (1994) 179(3):973–84. doi:10.1084/jem.179.3.973
90. Williams WM, Staines NA, Muller S, Isenberg DA. Human T cell responses to autoantibody variable region peptides. *Lupus* (1995) 4(6):464–71. doi:10.1177/096120339500400608
91. Blank M, Shoenfeld Y. The story of the 16/6 idiotype and systemic lupus erythematosus. *Isr Med Assoc J* (2008) 10(1):37–9.
92. Dayan M, Segal R, Sthoeger Z, Waisman A, Brosh N, Elkayam O, et al. Immune response of SLE patients to peptides based on the complementarity determining regions of a pathogenic anti-DNA monoclonal antibody. *J Clin Immunol* (2000) 20(3):187–94. doi:10.1023/A:1006685413157
93. Peen E, Malone C, Myers C, Williams RC Jr, Peck AB, Csernok E, et al. Amphipathic variable region heavy chain peptides derived from monoclonal human Wegener's anti-PR3 antibodies stimulate lymphocytes from patients with Wegener's granulomatosis and microscopic polyangiitis. *Clin Exp Immunol* (2001) 125(2):323–31. doi:10.1046/j.1365-2249.2001.01482.x

94. van Schooten WC, Devereux D, Ho CH, Quan J, Aguilar BA, Rust CJ. Joint-derived T cells in rheumatoid arthritis react with self-immunoglobulin heavy chains or immunoglobulin-binding proteins that copurify with immunoglobulin. *Eur J Immunol* (1994) 24(1):93–8. doi:10.1002/eji.1830240115

Conflict of Interest Statement: RB and JH hold equity in EigenBio, the company responsible for designing the bioinformatics models used in this project. HR is an equity holder in Adaptive Biotechnologies. The two companies are independent. All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Høglund, Lossius, Johansen, Homan, Benth, Robins, Bogen, Bremel and Holmøy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers

Supplementary Material

In silico prediction analysis of idiotope-driven T-B cell collaboration in multiple sclerosis

Rune A. Høglund*[1,2], Andreas Lossius[1,3], Jorunn N. Johansen[3], Jane Homan[4], Jūratė Šaltytė Benth[5], Harlan Robins[6], Bjarne Bogen[2,3,7], Robert Bremel[4†] and Trygve Holmøy[1,2†]

***Correspondence:**
Rune A. Høglund, MD
Clinical Neuroscience group, Department of Neurology, Akershus University Hospital, Postbox 1000, N-1478 Lørenskog, Norway.
Phone: +47.67963957; E-mail: r.a.hoglund@medisin.uio.no

S1 - Subject characteristics at time of sample collections

ID	Sex	Age	Diagnosis	Disease duration (months)[a]	Previous or ongoing treatment	Number of relapses	Time since relapse (months)	OCB[b]	CSF cell count[c]	Albumin ratio	IgG index
MS-1	F	29	RR-MS[d]	14	None	1	14	+	3	4	0.85
MS-2	F	33	RR-MS	156	IFN β 1a (48 months)	5	60	+	1	5.1	0.81
MS-3	F	36	RR-MS	66	GA (27 months)	4	33	+	3	3.8	1.7
MS-4	F	63	RR-MS	23	IFN β 1b (2 months)	1	23	+	1	3.6	0.68
MS-5	F	27	RR-MS	6	None	4	4	+	2	3.8	1.2
MS-6	F	33	RR-MS[d]	23	None	1	23	+	3	3.5	0.55
MS-7	M	33	RR-MS	48	None	3	3	+	10	17	0.9
MS-8	F	42	RR-MS	64	None	2	20	+	1	5.5	0.94
MS-9	F	34	RR-MS	2	None	1	2	+	25	6	2.93
MS-10	F	42	RR-MS	35	None	2	18	+	11	3.2	1.06
MS-11	F	20	RR-MS	13	None	2	1	+	10*	4.8*	1.2*
OIND-1	M	45	Aseptic meningitis	1	None	N/A	N/A	-	17	8	0.48
OIND-2	M	28	Aseptic meningoencephalitis	0	None	N/A	N/A	-	40	18.7	0.64
OIND-3	M	39	Neurosarcoidosis	13	None	N/A	N/A	-	75	31	0.65
OIND-4	M	42	Polyradiculitis	7	None	N/A	N/A	+	10	8	0.88
OIND-5	F	45	Aseptic meningitis	0	None	N/A	N/A	-	14	3	0.48
OIND-6	M	48	Neurosarcoidosis	16	None	N/A	N/A	-	20	26	0.46

F: female; M: male; RR-MS: relapse-remitting multiple sclerosis; OIND: other inflammatory neurological disease; IFN: interferon; GA: glatiramer acetate

[a] since first symptom, [b] Oligoclonal bands (OCB) positive indicates presence of >2 CSF specific IgG bands on isoelectric focusing, [c] number of mononuclear cells per microliter CSF, [d] Patients MS-1 and MS-6 had clinically isolated syndrome at inclusion, but later developed definite RR-MS. *values are from a 1 month prior sampling. The sample from which we acquired cells for sequencing was stopped after 2 mL due to accidental bleeding.

S2 – List of human leukocyte antigen (HLA) alleles included in the prediction analyses

HLA class I		HLA class II		
A	**B**	**DRB**	**DP**	**DQ**
01:01	07:02	1*01:01	DPA1*01:03-DPB1*02:01	DQA1*01:01-DQB1*05:01
02:01	08:01	1*03:01	DPA1*01:03-DPB1*04:02	DQA1*01:02-DQB1*06:02
02:02	15:01	1*04:01	DPA1*01:03-DPB1*04:01	DQA1*03:01-DQB1*03:02
02:03	15:03	1*04:04	DPA1*02:01-DPB1*01:01	DQA1*04:01-DQB1*04:02
02:06	18:01	1*04:05	DPA1*02:01-DPB1*05:01	DQA1*05:01-DQB1*02:01
03:01	27:05	1*07:01	DPA1*03:01-DPB1*04:02	DQA1*05:01-DQB1*03:01
11:01	35:01	1*08:02		
23:01	40:01	1*09:01		
24:02	40:02	1*11:01		
24:03	44:02	1*12:01		
26:01	44:01	1*13:02		
29:02	51:01	1*15:01		
30:01	53:01	3*01:01		
30:02	54:01	3*02:02		
31:01	57:01	4*01:01		
32:01	58:01	5*01:01		
33:01				
68:01				
68:02				
69:01				

S3 – IGHV characteristics

	Unique IGHV (N)		IGHV family usage (%)												
			CSF							Blood					
ID	CSF	Blood	1	2	3	4	5	6	7	1	2	3	4	5	6	7
MS-1	98	238,951	30.6	3.1	33.7	31.6	0.0	1.0	0.0	25.3	5.2	48.2	12.0	6.8	0.8	1.7
MS-2	241	264,483	38.6	5.4	40.7	10.8	3.7	0.4	0.4	22.5	3.4	54.2	14.3	4.9	0.7	0.0
MS-3	417	118,547	30.5	9.8	13.2	45.8	0.0	0.5	0.2	19.9	4.2	46.1	20.4	8.3	1.1	0.0
MS-4	109	235,356	6.4	9.2	39.4	37.6	5.5	1.8	0.0	27.2	3.9	41.1	13.7	13.6	0.5	0.0
MS-5	1,409	157,163	4.1	10.1	13.8	70.7	0.6	0.4	0.4	29.1	4.8	40.5	13.5	9.8	0.6	1.7
MS-6	3,409	110,489	23.3	3.8	46.3	18.5	5.8	2.2	0.0	24.9	4.5	43.0	16.3	9.8	1.5	0.1
MS-7	2,753	117,643	22.8	5.5	31.4	34.1	4.2	2.0	0.0	21.5	6.0	42.8	21.5	7.1	1.0	0.1
MS-8	1,432	35,456	3.1	5.0	12.2	79.1	0.3	0.1	0.3	27.1	5.1	46.8	12.5	5.7	1.1	1.6
MS-9	14,413	158,631	2.9	9.0	13.2	74.4	0.3	0.1	0.0	22.0	4.4	46.8	16.6	8.0	0.6	1.6
MS-10	617	105,514	10.7	5.5	38.9	40.7	3.9	0.3	0.0	21.7	3.4	51.4	16.0	6.3	1.2	0.0
MS-11	658	8,921	12.5	6.1	39.7	34.3	6.1	0.3	1.1	9.4	1.8	58.2	16.0	11.9	2.0	0.7
OIND-1	10,247	-	15.1	3.1	53.4	21.9	5.4	1.1	0.0	-	-	-	-	-	-	-
OIND-2	4,115	-	20.6	3.5	55.1	15.3	3.7	0.7	1.0	-	-	-	-	-	-	-
OIND-3	8,283	-	20.8	3.4	47.1	22.7	5.2	0.7	0.1	-	-	-	-	-	-	-
OIND-4	5,091	5,978	4.8	2.6	22.5	67.4	2.0	0.7	0.0	19.9	4.9	52.0	15.2	6.2	1.7	0.1
OIND-5	266	122,444	14.7	2.3	48.9	18.4	15.0	0.4	0.4	21.6	3.9	51.2	14.4	8.0	0.9	0.0
OIND-6	6,831	72,955	17.1	3.2	51.3	22.4	5.4	0.5	0.0	28.8	3.7	44.8	15.4	6.5	0.8	0.0

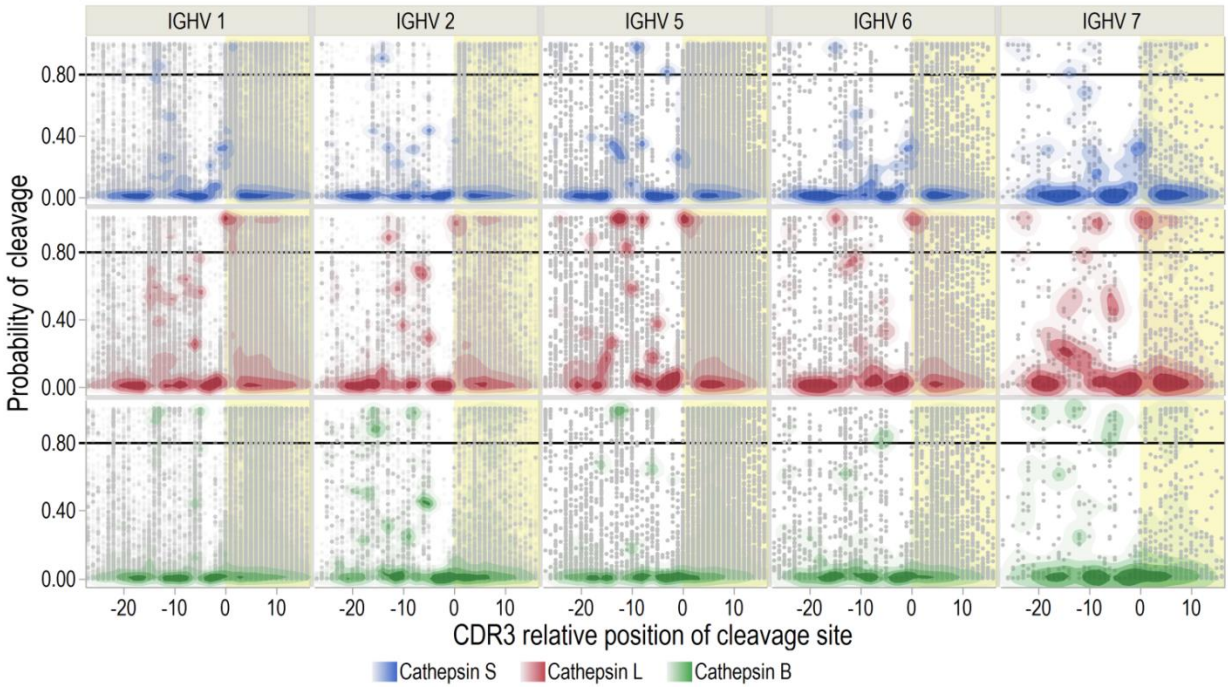
S4 – Cathepsin cleavage of IGHV transcripts

		Cathepsin S			Cathepsin L			Cathepsin B		
	Number	Mean	LCI	UCI	Mean	LCI	UCI	Mean	LCI	UCI
IGHV 1	7921	2.63	2.60	2.66	3.83	3.80	3.86	3.08	3.04	3.11
IGHV 2	3035	2.39	2.34	2.44	3.52	3.47	3.58	3.97	3.91	4.02
IGHV 3	21865	1.54	1.52	1.56	4.78	4.76	4.80	3.04	3.02	3.06
IGHV 4	24966	3.80	3.78	3.82	6.25	6.24	6.27	3.18	3.16	3.20
IGHV 5	2098	3.40	3.34	3.46	7.41	7.35	7.47	3.50	3.43	3.56
IGHV 6	430	2.41	2.28	2.54	4.59	4.45	4.72	2.26	2.11	2.41
IGHV 7	74	2.76	2.44	3.07	5.04	4.71	5.37	4.89	4.53	5.25

L/UCI – Lower and upper 95% confidence interval

S5 – Distribution of cleavage probabilities

Supplementary Figure S5

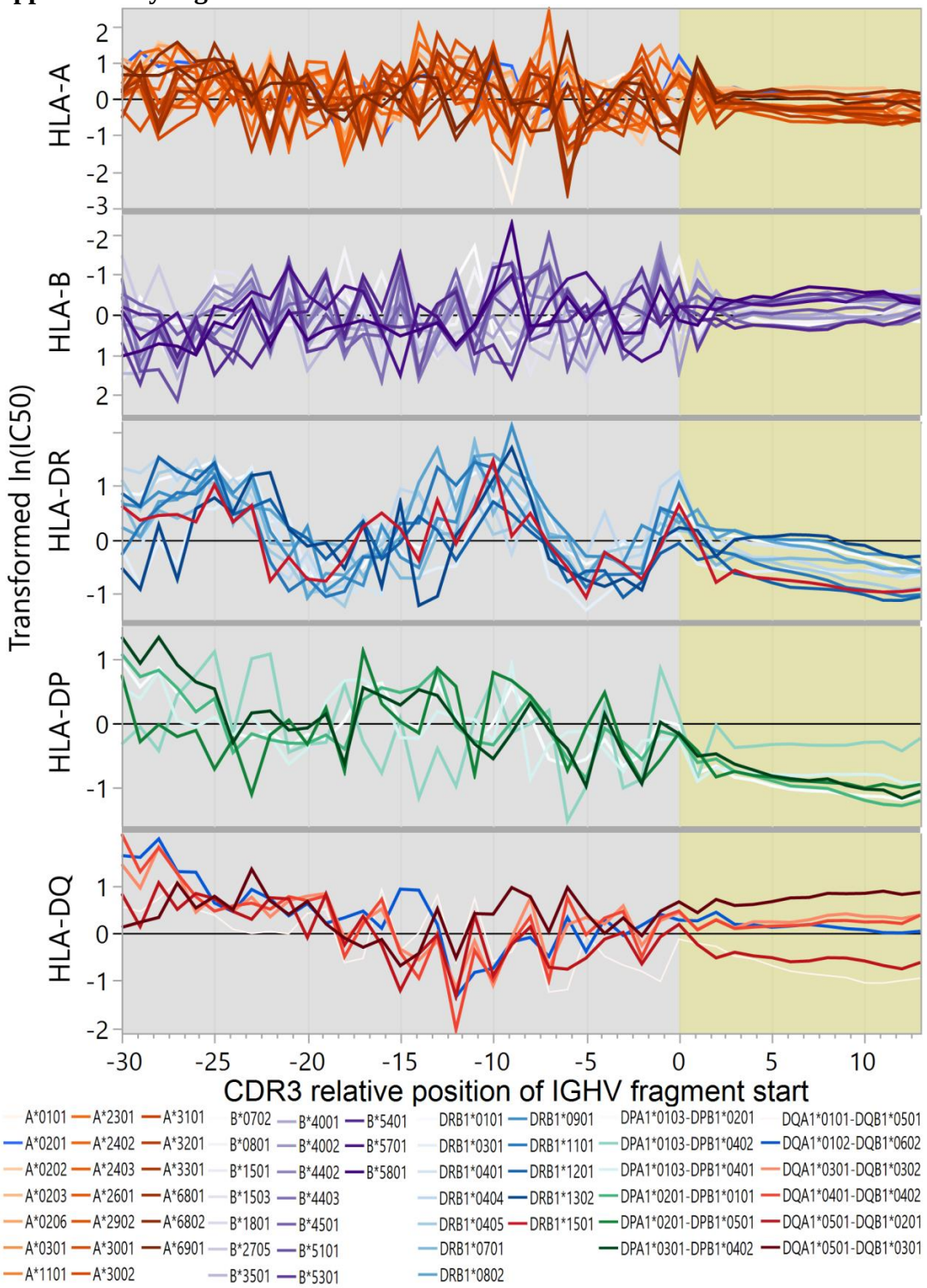


Cathepsin cleavage sites in immunoglobulin heavy variable (IGHV) transcripts were predicted with neural-net models. The distribution of predicted probabilities for cleavage (range 0-1) for Cathepsin S, L and B are shown for IGHV families 1, 2 and 5-7. The cut-off lines are set at 0.8, above which is considered a high probability for cleavage. The complementarity determining region (CDR) 3 region is marked with yellow shading and the CDR3 relative position is aligned with the cleavage site at P1-P1'.

CDR3 is marked with shading. Cathepsin L shows high predicted probability of cleavage at the CDR3 start for all IGHV families. All cathepsins have predicted several hotspots in the framework (FW) 3 region.

S6 - IGHV fragment affinities for HLA class I and II molecules

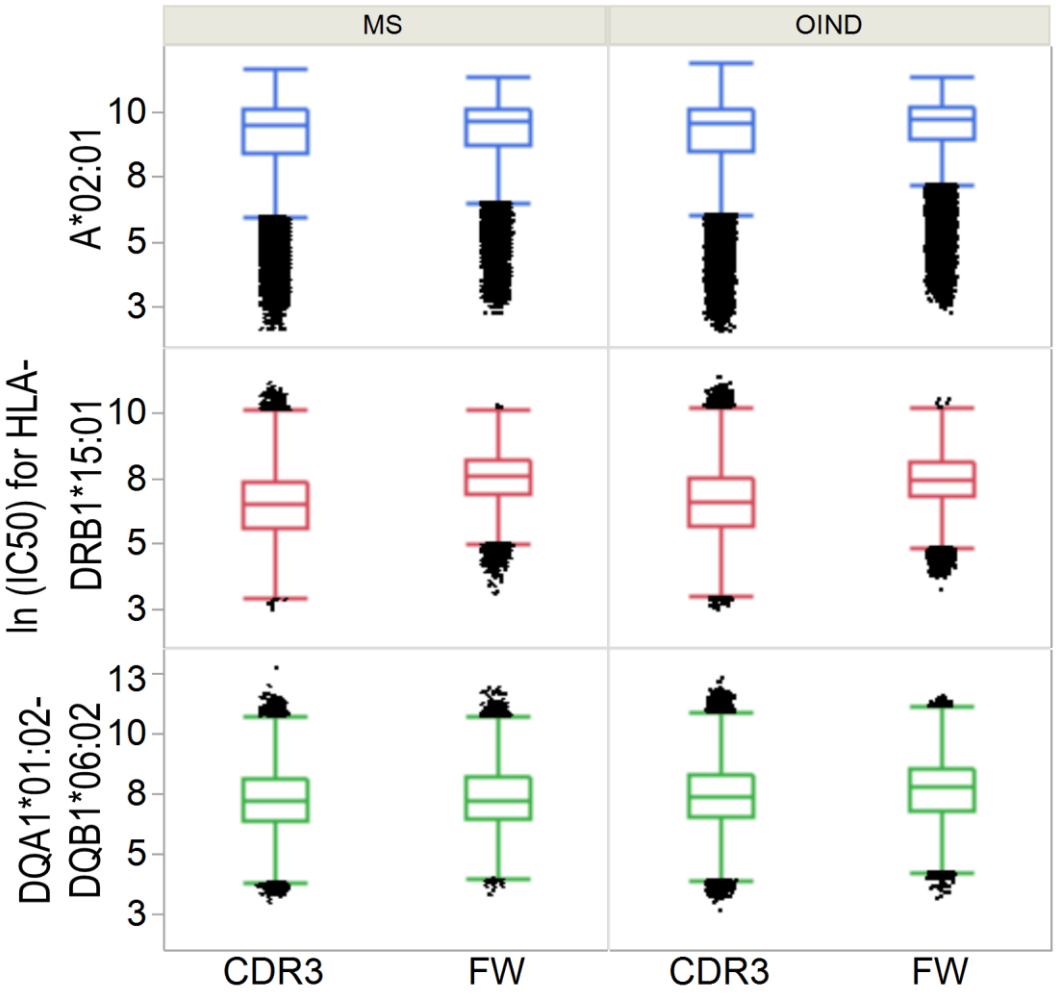
Supplementary Figure S6



Binding affinities of CSF IGHV fragments from all patients were predicted for HLA molecules listed in Supplementary Materials S2. Results are presented as mean Johnson SI standardized values of ln (IC_{50}) calculated for each CDR3 relative position (N-terminus of a 15-mer in the case of HLA class II and a 9-mer in the case of HLA class I), bringing the different predictions to a comparable scale. Low values indicate higher affinity. Yellow shade indicates the CDR3 region.

S7 - Mean ln (IC_{50}) affinities for selected HLA molecules

Supplementary Figure S7



Binding affinities of IGHV fragments were predicted for HLA-A*02:01, HLA-DRB1*15:01 and HLA-DQA1*01:02-DQB1*06:02 with neural-net models. We compared the CDR3 vs FW3 region by splitting the transcripts at CDR3 relative position -7. Mean predicted ln (IC_{50}) by region (CDR3 and FW3) and by disease are shown as outlier box plots with whiskers covering 1st and 3rd quartile -/+ 1.5*(interquartile range). Table S7 shows the adjusted means used for statistical testing.

Supplementary Table S7

		HLA DRB1*15:01	HLA DQA1*01:02-DQB1*06:02	HLA A*02:01
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	-	4.38	-
	CDR3 relative position	3.76	1.74	1.11
	transcript clone-level	-	-	-
	Mean difference	0.96	0.07	-0.21
MS	LCI	0.95	0.07	-0.22
	UCI	0.97	0.08	-0.2
	p-value	<0.001	<0.001	<0.001
		Adjusted for cluster effect: CDR3 relative position	Unadjusted for cluster effect	Adjusted for cluster effect: CDR3 relative position
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	-	-	0.7
	CDR3 relative position	3.1	2.49	-
	transcript clone-level	-	-	-
	Mean difference	0.86	-0.75	0.21
OIND	LCI	0.85	-0.76	0.20
	UCI	0.86	-0.74	0.21
	p-value	<0.001	<0.001	<0.001
		Unadjusted for cluster effect	Adjusted for cluster effect CDR3 relative position	Unadjusted for cluster effect

ICC – Intra class correlation, L/UCI – Lower and Upper Confidence interval

S8 – Statistical tests of affinity for disease associated HLA molecules

Supplementary Table S8a – Intra class correlation on patient level

DQA1_0102-DQB1_0602		DRB1_1501		A_0201	
RelPos	ICC (%)	RelPos	ICC (%)	RelPos	ICC (%)
-7	21,0	-7	11,2	-7	28,0
-6	20,0	-6	9,3	-6	6,7
-5	19,8	-5	10,4	-5	18,4
-4	7,9	-4	17,8	-4	13,5
-3	23,6	-3	9,4	-3	20,1
-2	8,1	-2	9,7	-2	25,3
-1	15,9	-1	9,2	-1	14,6
0	8,2	0	9,9	0	5,4
1	6,5	1	7,7	1	14,8
2	19,6	2	13,8	2	10,9
3	10,1	3	8,9	3	12,4
4	14,2	4	7,0	4	9,3
5	10,6	5	10,1	5	14,9
6	15,9	6	12,9	6	7,8
7	19,3	7	15,5	7	11,6
8	13,8	8	16,3	8	11,1
9	20,6	9	13,8	9	15,4
10	14,1	10	14,7	10	26,3
11	19,2	11	16,2	11	17,0
12	12,3	12	13,2	12	8,4
13	13,6	13	12,3	13	14,0
14	12,8	14	14,6	14	15,3
15	15,3	15	17,0	15	12,7
16	14,6	16	20,2	16	11,4
17	15,1	17	18,2	17	17,2
18	13,4	18	18,1	18	14,7
19	20,1	19	22,9	19	21,3
20	20,7	20	24,2	20	22,0
21	18,9	21	22,7	21	21,4
22	20,0	22	20,6	22	22,2
23	38,5	23	36,6	23	25,8
24	21,9	24	17,5	24	34,0
25	38,2	25	31,5	25	17,9
26	19,7	26	22,2	26	36,7
27	43,3	27	44,1	27	17,7
28	13,5	28	35,0	28	32,2
29	70,1	29	52,4	29	59,2
30	0,0	30	66,0	30	0,0

ICC=Intra class correlation

Supplementary Table S8b – Mean difference in predicted affinity (ln (IC50) for HLA-DRB1*15:01 for IGHV4 fragments compared to other IGHV families

DRB1_1501	IGHV_family 4 vs others				Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,122	0,000	-0,135	-0,011	*	*
-6	-0,104	0,000	-0,118	-0,090	*	*
-5	-0,066	0,000	-0,079	-0,053	*	*
-4	0,006	0,467	-0,011	0,023		
-3	0,022	0,028	0,002	0,041	*	*
-2	-0,079	0,000	-0,098	-0,059	*	*
-1	-0,136	0,000	-0,158	-0,114	*	*
0	-0,074	0,000	-0,096	-0,053	*	*
1	-0,065	0,000	-0,090	-0,041	*	*
2	-0,128	0,000	-0,152	-0,104	*	*
3	-0,034	0,006	-0,058	-0,010	*	*
4	-0,007	0,564	-0,031	0,017		
5	-0,043	0,000	-0,067	-0,020	*	*
6	-0,064	0,000	-0,088	-0,041	*	*
7	-0,113	0,000	-0,136	-0,090	*	*
8	-0,095	0,000	-0,118	-0,073	*	*
9	-0,107	0,000	-0,129	-0,084	*	*
10	-0,096	0,000	-0,117	-0,074	*	*
11	-0,098	0,000	-0,121	-0,074	*	*
12	-0,076	0,000	-0,101	-0,051	*	*
13	-0,042	0,001	-0,067	-0,016	*	*
14	0,046	0,001	0,019	0,073	*	*
15	-0,044	0,003	-0,073	-0,015	*	*
16	-0,084	0,000	-0,118	-0,050	*	*
17	-0,108	0,000	-0,149	-0,067	*	*
18	-0,069	0,010	-0,122	-0,017	*	*
19	-0,010	0,761	-0,071	0,052		
20	0,158	0,000	0,080	0,236	*	*
21	0,044	0,504	-0,085	0,173		
22	0,192	0,037	0,012	0,373	*	*
23	0,405	0,002	0,150	0,660	*	*
24	-0,622	0,067	-1,289	0,044	*	
25	-0,112	0,714	-0,711	0,487		
26	0,491	0,195	-0,252	1,235		
27	-0,137	0,841	-1,476	1,203		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S8c – Mean difference in predicted affinity (ln (IC50) for HLA-DQA1*01:02-DQB1*06:02 for IGHV4 fragments compared to other IGHV families

DQA1_0102-DQB1_0602	IGHV_family 4 vs others				Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,414	0,000	-0,432	-0,397	*	*
-6	-0,083	0,000	-0,103	-0,062	*	*
-5	-0,163	0,000	-0,186	-0,140	*	*
-4	-0,071	0,000	-0,094	-0,049	*	*
-3	-0,069	0,000	-0,094	-0,045	*	*
-2	0,031	0,011	0,007	0,055	*	*
-1	0,059	0,000	0,036	0,083	*	*
0	0,099	0,000	0,077	0,120	*	*
1	-0,001	0,944	-0,024	0,023		
2	0,063	0,000	0,039	0,086	*	*
3	0,130	0,000	0,108	0,152	*	*
4	0,048	0,000	0,026	0,070	*	*
5	0,067	0,000	0,046	0,088	*	*
6	0,156	0,000	0,134	0,178	*	*
7	0,010	0,343	-0,011	0,031		
8	0,056	0,000	0,034	0,078	*	*
9	0,108	0,000	0,085	0,131	*	*
10	-0,062	0,000	-0,087	-0,037	*	*
11	-0,041	0,003	-0,068	-0,014	*	*
12	0,054	0,000	0,025	0,084	*	*
13	0,040	0,014	0,008	0,071	*	*
14	-0,056	0,002	-0,091	-0,020	*	*
15	0,024	0,240	-0,016	0,063		
16	0,063	0,008	0,017	0,110	*	*
17	-0,057	0,067	-0,118	0,004	*	
18	-0,047	0,199	-0,118	0,025		
19	-0,156	0,000	-0,239	-0,073	*	*
20	-0,421	0,000	-0,527	-0,316	*	*
21	-0,143	0,083	-0,304	0,019	*	
22	-0,451	0,000	-0,689	-0,213	*	*
23	-0,601	0,000	-0,908	-0,294	*	*
24	0,842	0,033	0,066	1,618	*	*
25	0,286	0,487	-0,521	1,092		
26	-0,702	0,128	-1,607	0,203	*	
27	0,195	0,814	-1,426	1,816		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

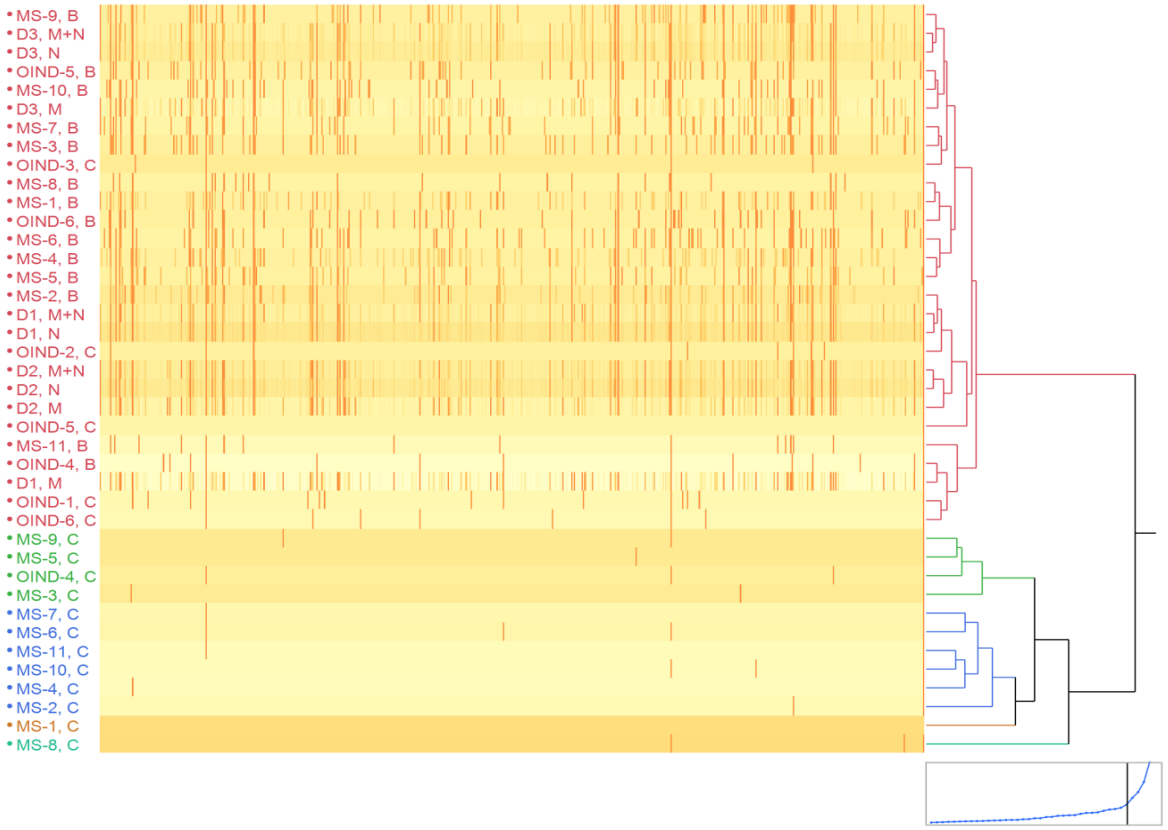
Supplementary Table S8d – Mean difference in predicted affinity (ln (IC50) for HLA-A*01:02 for IGHV4 fragments compared to other IGHV families

A_0201	IGHV_family 4 vs others				Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,514	0,000	-0,521	-0,507	*	*
-6	-0,059	0,000	-0,067	-0,052	*	*
-5	-0,203	0,000	-0,222	-0,185	*	*
-4	-0,104	0,000	-0,132	-0,076	*	*
-3	0,003	0,860	-0,026	0,032		
-2	-0,077	0,000	-0,109	-0,045	*	*
-1	-0,009	0,412	-0,029	0,012		
0	0,054	0,000	0,042	0,066	*	*
1	0,039	0,000	0,024	0,055	*	*
2	0,060	0,000	0,037	0,083	*	*
3	0,116	0,000	0,090	0,142	*	*
4	0,013	0,373	-0,016	0,042		
5	-0,013	0,436	-0,046	0,020		
6	-0,043	0,007	-0,075	-0,012	*	*
7	-0,052	0,003	-0,086	-0,018	*	*
8	0,101	0,000	0,064	0,138	*	*
9	-0,051	0,019	-0,094	-0,009	*	*
10	-0,363	0,000	-0,408	-0,318	*	*
11	-0,181	0,000	-0,228	-0,134	*	*
12	-0,041	0,127	-0,093	0,012		
13	0,004	0,892	-0,055	0,063		
14	-0,343	0,000	-0,410	-0,275	*	*
15	-0,235	0,000	-0,308	-0,161	*	*
16	-0,024	0,567	-0,105	0,057		
17	-0,012	0,827	-0,117	0,094		
18	-0,287	0,000	-0,420	-0,154	*	*
19	-0,704	0,000	-0,867	-0,550	*	*
20	-0,008	0,936	-0,200	0,185		
21	-0,149	0,276	-0,417	0,119		
22	-0,458	0,036	-0,885	-0,030	*	*
23	0,308	0,285	-0,256	0,871		
24	0,440	0,545	-0,986	1,866		
25	0,202	0,794	-1,313	1,717		
26	-0,207	0,827	-0,063	1,649		
27	-1,440	0,426	-4,983	2,104		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

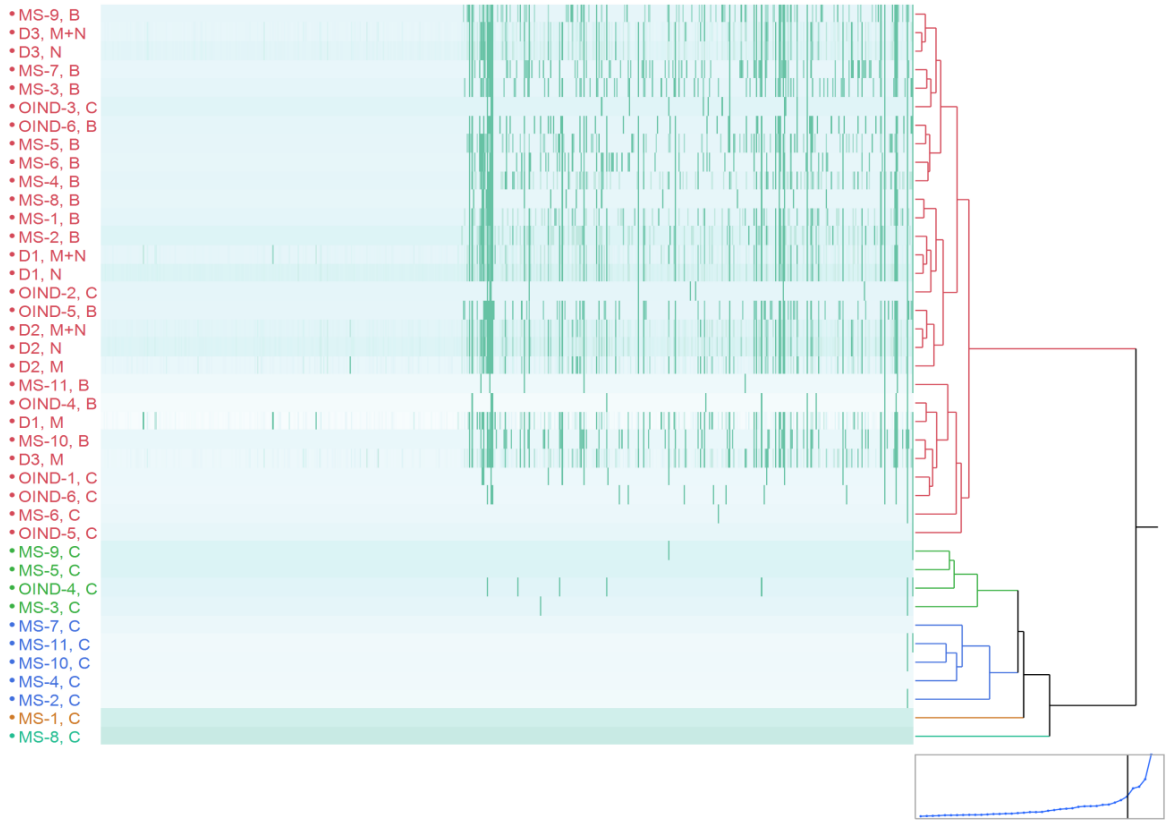
S9 – Hierarchical clustering of Z-standardized TCEM occurrences

Supplementary figure S9a - Hierarchical clustering of Z-standardized TCEMI

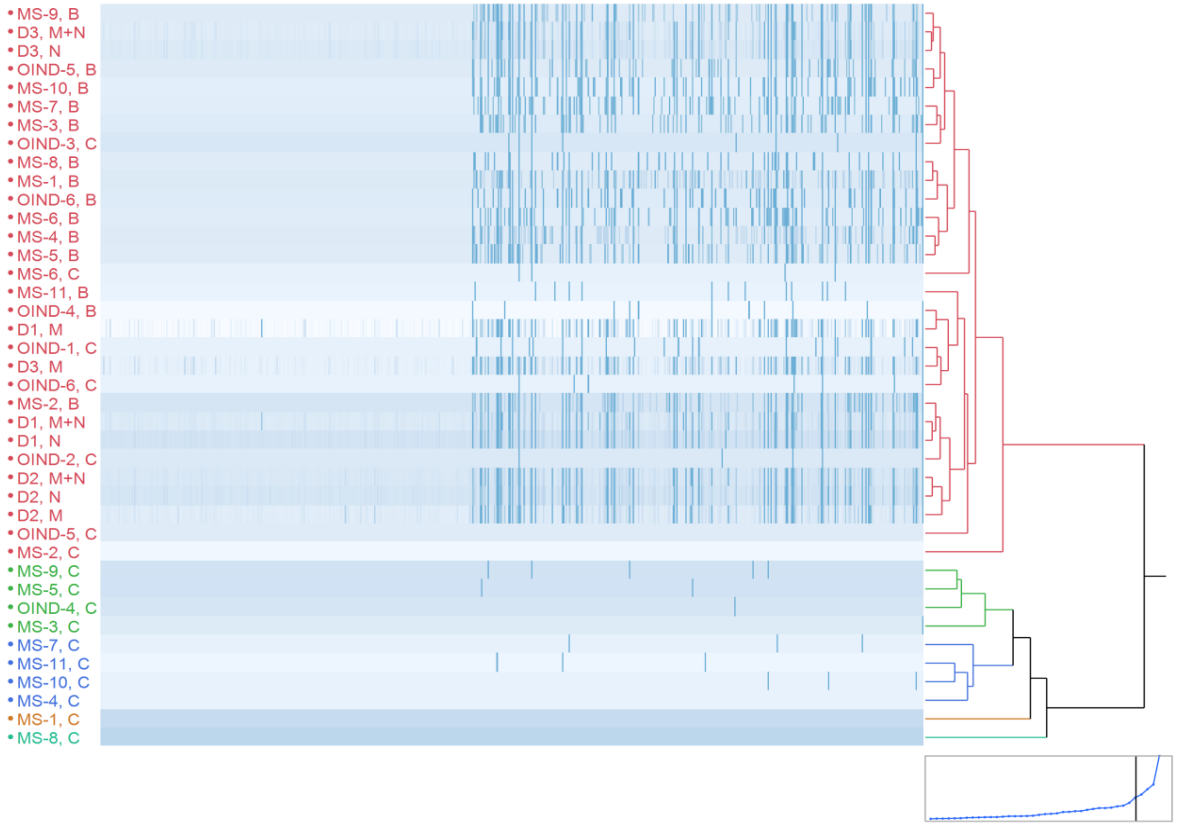


The occurrences of all TCEM (x axis) in IGHV transcripts were identified and summarized by patient and compartment in patients with MS (MS 1-11), and OIND (OIND 1-6) and three healthy individuals (D1-3). The figures S9a-c are the results of hierarchical clustering analysis with Wards' method of the Z-standardized occurrences within patient and compartment for all three TCEM patterns. N=Naïve B cells from blood, M=Mature B cells from blood. In each patient B at the end of the identification code denotes blood and C denotes CSF.

Supplementary Figure S9b - Hierarchical clustering of Z-standardized TCEMIIa

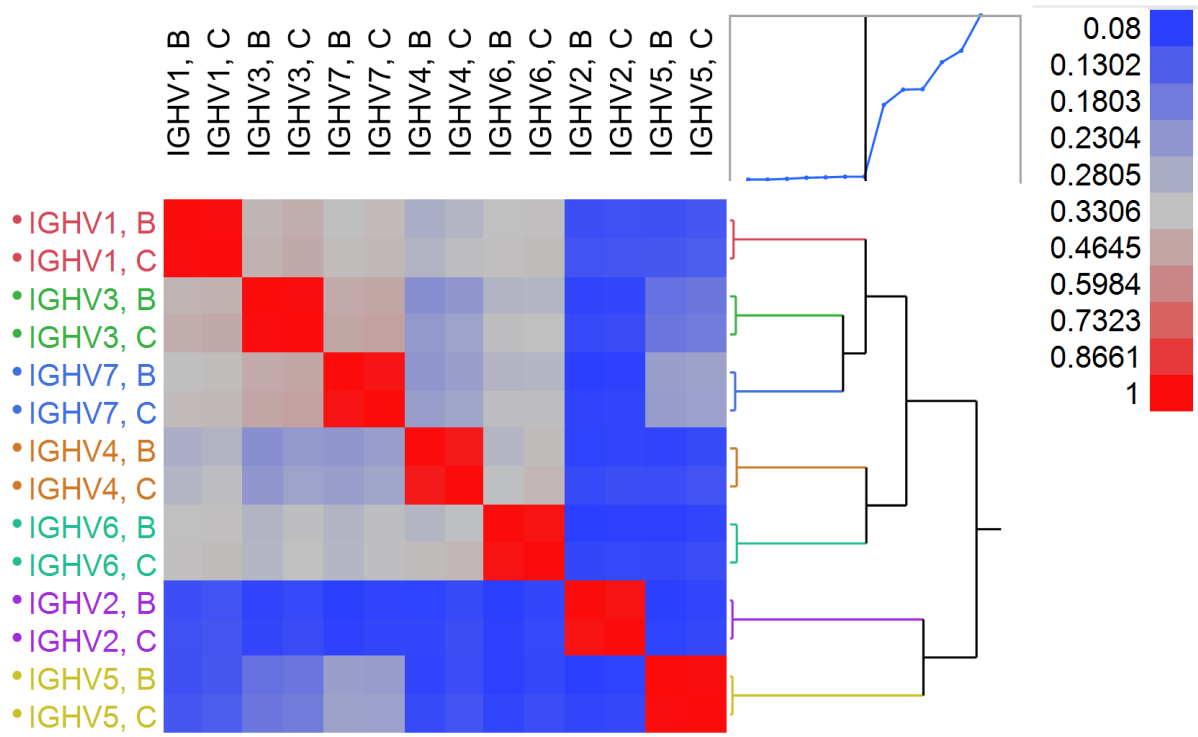


Supplementary Figure S9c - Hierarchical clustering of Z-standardized TCEMIIb



S10 – Hierarchical clustering of pairwise correlations of T cell exposed motif occurrences within each compartment and IGHV family

Supplementary Figure S10



We identified all possible T cell exposed motifs (TCEM) IIa in IGHV transcripts and counted their occurrences in blood and cerebrospinal fluid (CSF) and within each IGHV family. B denotes blood- and C denotes CSF derived IGHV transcripts. Hierarchical clustering analysis with Wards' method of pairwise correlation coefficients shows that clustering occurs on IGHV family level. The pairwise metric analysis can only be computed if a TCEM is present in both samples and the observed correlations are thus not attributable to differences in sample size.

S11 – Comparison of mean FC in blood and CSF at CDR3 relative positions

Supplementary Table S11a - Mean difference of TCEMI frequency class between blood and CSF in MS patients

	MS; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	**SIGN adj**	**SIGN**
-32	-3,850	0,000	-4,492	-3,208	*	*
-31	-2,857	0,000	-3,503	-2,211	*	*
-30	-2,976	0,000	-3,416	-2,535	*	*
-29	-3,348	0,000	-3,596	-3,101	*	*
-28	-2,060	0,000	-2,215	-1,905	*	*
-27	-2,765	0,000	-2,883	-2,648	*	*
-26	-1,736	0,000	-1,835	-1,636	*	*
-25	-2,126	0,000	-2,203	-2,049	*	*
-24	-2,839	0,000	-2,902	-2,775	*	*
-23	-2,752	0,000	-2,813	-2,691	*	*
-22	-2,026	0,000	-2,078	-1,973	*	*
-21	-3,521	0,000	-3,567	-3,475	*	*
-20	-3,707	0,000	-3,754	-3,660	*	*
-19	-4,565	0,000	-4,612	-4,517	*	*
-18	-4,416	0,000	-4,462	-4,371	*	*
-17	-4,616	0,000	-4,661	-4,571	*	*
-16	-2,726	0,000	-2,768	-2,683	*	*
-15	-2,825	0,000	-2,868	-2,782	*	*
-14	-1,094	0,000	-1,134	-1,053	*	*
-13	-0,788	0,000	-0,823	-0,753	*	*
-12	-0,546	0,000	-0,582	-0,511	*	*
-11	-0,510	0,000	-0,540	-0,480	*	*
-10	-0,988	0,000	-1,023	-0,953	*	*
-9	-0,148	0,000	-0,176	-0,119	*	*
-8	-0,635	0,000	-0,668	-0,603	*	*
-7	-0,562	0,000	-0,592	-0,531	*	*
-6	-0,430	0,000	-0,465	-0,394	*	*
-5	-0,102	0,000	-0,141	-0,063	*	*
-4	-0,308	0,000	-0,350	-0,267	*	*
-3	0,141	0,000	0,102	0,181	*	*
-2	0,040	0,026	0,005	0,074	*	*
-1	-0,239	0,000	-0,272	-0,205	*	*
0	-0,854	0,000	-0,894	-0,815	*	*
1	-1,478	0,000	-1,526	-1,431	*	*
2	-2,088	0,000	-2,143	-2,033	*	*
3	-2,336	0,000	-2,397	-2,274	*	*
4	-2,251	0,000	-2,320	-2,183	*	*
5	-2,130	0,000	-2,207	-2,054	*	*
6	-2,466	0,000	-2,550	-2,383	*	*
7	-2,123	0,000	-2,214	-2,033	*	*
8	-2,218	0,000	-2,322	-2,114	*	*
9	-2,783	0,000	-2,909	-2,656	*	*
10	-3,140	0,000	-3,309	-2,971	*	*
11	-2,367	0,000	-2,587	-2,147	*	*
12	-2,887	0,000	-3,150	-2,625	*	*
13	-1,932	0,000	-2,312	-1,551	*	*
14	-4,330	0,000	-4,865	-3,796	*	*
15	-2,550	0,000	-3,360	-1,740	*	*
16	-0,559	0,252	-1,515	0,397	*	
17	-0,140	0,819	-1,339	1,059		
18	1,353	0,043	0,044	2,663	*	*
19	0,259	0,740	-1,269	1,786		
20	0,014	0,987	-1,681	1,709		
21	-1,433	0,140	-3,338	0,471	*	
22	-2,417	0,011	-4,270	-0,563	*	*
23	-2,338	0,012	-4,170	-0,506	*	*
24	-2,162	0,023	-4,021	-0,303	*	*
25	-2,441	0,007	-4,224	-0,659	*	*
26	-1,876	0,039	-3,656	-0,097	*	*
27	-1,440	0,116	-3,233	0,353	*	
28	-1,449	0,121	-3,279	0,382	*	
29	-1,054	0,349	-3,260	1,153		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S11b - Mean difference of TCEMIIa frequency class between blood and CSF in MS patients

	MS; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	**SIGN adj**	**SIGN**
-32	-3,289	0,000	-4,021	-2,557	*	*
-31	-2,234	0,000	-2,851	-1,616	*	*
-30	-1,511	0,000	-1,970	-1,052	*	*
-29	-3,668	0,000	-3,937	-3,400	*	*
-28	-2,333	0,000	-2,487	-2,178	*	*
-27	-1,838	0,000	-1,964	-1,713	*	*
-26	-1,645	0,000	-1,742	-1,553	*	*
-25	-1,676	0,000	-1,743	-1,608	*	*
-24	-4,679	0,000	-4,750	-4,607	*	*
-23	-4,231	0,000	-4,292	-4,170	*	*
-22	-2,875	0,000	-2,928	-2,822	*	*
-21	-4,787	0,000	-4,845	-4,729	*	*
-20	-1,834	0,000	-1,875	-1,793	*	*
-19	-4,385	0,000	-4,429	-4,340	*	*
-18	-3,327	0,000	-3,374	-3,280	*	*
-17	-2,876	0,000	-2,919	-2,832	*	*
-16	-2,534	0,000	-2,575	-2,493	*	*
-15	-1,206	0,000	-1,239	-1,173	*	*
-14	-1,261	0,000	-1,297	-1,225	*	*
-13	-0,566	0,000	-0,603	-0,528	*	*
-12	-0,514	0,000	-0,551	-0,478	*	*
-11	-1,337	0,000	-1,366	-1,309	*	*
-10	-0,545	0,000	-0,577	-0,514	*	*
-9	-0,140	0,000	-0,166	-0,114	*	*
-8	-0,214	0,000	-0,256	-0,173	*	*
-7	0,136	0,000	0,096	0,176	*	*
-6	-0,620	0,000	-0,652	-0,587	*	*
-5	-0,110	0,000	-0,144	-0,077	*	*
-4	-0,479	0,000	-0,510	-0,447	*	*
-3	-0,276	0,000	-0,311	-0,242	*	*
-2	-0,435	0,026	-0,471	-0,400	*	*
-1	-0,796	0,000	-0,833	-0,759	*	*
0	-1,253	0,000	-1,296	-1,211	*	*
1	-1,381	0,000	-1,427	-1,335	*	*
2	-1,583	0,000	-1,632	-1,534	*	*
3	-1,349	0,000	-1,399	-1,298	*	*
4	-1,342	0,000	-1,396	-1,287	*	*
5	-1,580	0,000	-1,641	-1,519	*	*
6	-2,081	0,000	-2,150	-2,012	*	*
7	-1,813	0,000	-1,893	-1,733	*	*
8	-1,923	0,000	-2,022	-1,829	*	*
9	-2,523	0,000	-2,641	-2,404	*	*
10	-1,992	0,000	-2,150	-1,833	*	*
11	-2,352	0,000	-2,556	-2,147	*	*
12	-2,304	0,000	-2,458	-2,061	*	*
13	-2,652	0,000	-3,005	-2,299	*	*
14	-3,966	0,000	-4,465	-3,468	*	*
15	-2,262	0,000	-3,023	-1,502	*	*
16	-1,865	0,000	-2,766	-0,964	*	*
17	-2,553	0,819	-3,686	-1,421		
18	-0,609	0,334	-1,845	0,626		
19	-2,626	0,000	-4,043	-1,209	*	*
20	-1,362	0,096	-2,966	0,242	*	
21	-2,670	0,004	-4,465	-0,875	*	*
22	-3,396	0,000	-5,149	-1,644	*	*
23	-1,345	0,122	-3,052	0,361	*	
24	-2,776	0,001	-4,402	-1,150	*	*
25	-1,154	0,119	-2,606	0,298	*	
26	-0,779	0,297	-2,243	0,685		
27	-1,672	0,023	-3,113	-0,230	*	*
28	-1,985	0,018	-3,628	-0,341	*	*
29	-1,542	0,129	-3,535	0,451	*	

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S11c - Mean difference of TCEMIIb frequency class between blood and CSF in MS patients

	MS; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-32	-3,289	0,000	-3,992	-2,586	*	*
-31	-2,475	0,000	-3,067	-1,883	*	*
-30	-2,094	0,000	-2,494	-1,694	*	*
-29	-4,395	0,000	-4,672	-4,117	*	*
-28	-1,738	0,000	-1,901	-1,576	*	*
-27	-2,019	0,000	-2,143	-1,896	*	*
-26	-2,498	0,000	-2,589	-2,408	*	*
-25	-1,625	0,000	-1,693	-1,557	*	*
-24	-4,417	0,000	-4,489	-4,345	*	*
-23	-3,981	0,000	-4,035	-3,927	*	*
-22	-3,309	0,000	-3,366	-3,253	*	*
-21	-5,093	0,000	-5,153	-5,033	*	*
-20	-1,681	0,000	-1,718	-1,644	*	*
-19	-4,501	0,000	-4,549	-4,452	*	*
-18	-1,346	0,000	-1,393	-1,299	*	*
-17	-2,696	0,000	-2,738	-2,655	*	*
-16	-2,884	0,000	-2,923	-2,845	*	*
-15	-1,130	0,000	-1,162	-1,098	*	*
-14	-2,290	0,000	-2,328	-2,252	*	*
-13	-0,331	0,000	-0,368	-0,294	*	*
-12	-1,248	0,000	-1,284	-1,212	*	*
-11	-1,436	0,000	-1,464	-1,407	*	*
-10	-0,453	0,000	-0,489	-0,418	*	*
-9	-0,852	0,000	-0,883	-0,821	*	*
-8	-0,136	0,000	-0,177	-0,096	*	*
-7	0,048	0,007	0,013	0,083	*	*
-6	-0,618	0,000	-0,650	-0,586	*	*
-5	0,345	0,000	0,309	0,382	*	*
-4	-0,492	0,000	-0,522	-0,462	*	*
-3	-0,642	0,000	-0,676	-0,608	*	*
-2	-0,766	0,000	-0,800	-0,732	*	*
-1	-0,870	0,000	-0,906	-0,834	*	*
0	-0,858	0,000	-0,898	-0,817	*	*
1	-1,031	0,000	-1,074	-0,988	*	*
2	-1,165	0,000	-1,210	-1,119	*	*
3	-1,082	0,000	-1,129	-1,034	*	*
4	-1,191	0,000	-1,243	-1,139	*	*
5	-1,300	0,000	-1,358	-1,242	*	*
6	-1,892	0,000	-1,957	-1,827	*	*
7	-1,568	0,000	-1,641	-1,495	*	*
8	-1,892	0,000	-1,978	-1,807	*	*
9	-1,926	0,000	-2,028	-1,823	*	*
10	-1,589	0,000	-1,722	-1,455	*	*
11	-1,840	0,000	-2,009	-1,671	*	*
12	-2,088	0,000	-2,286	-1,889	*	*
13	-1,587	0,000	-1,873	-1,299	*	*
14	-3,032	0,000	-3,438	-2,625	*	*
15	-1,574	0,000	-2,196	-0,952	*	*
16	-1,892	0,000	-2,639	-1,146	*	*
17	-1,760	0,000	-2,715	-0,805	*	*
18	-0,395	0,458	-1,437	0,648		
19	-3,145	0,000	-4,384	-1,906	*	*
20	-1,018	0,165	-2,455	0,420	*	
21	-1,919	0,234	-3,579	-0,260	*	
22	-2,848	0,001	-4,445	-1,251	*	*
23	-1,473	0,063	-3,028	0,081	*	
24	-2,702	0,000	-4,180	-1,224	*	*
25	-0,709	0,334	-2,147	0,729		
26	-0,838	0,262	-2,303	0,627	*	
27	-1,679	0,016	-3,048	-0,310	*	*
28	-1,583	0,037	-3,071	-0,095	*	*
29	-1,763	0,046	-3,495	-0,031	*	*

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S11d - Mean difference of TCEMI frequency class between blood and CSF in OIND patients

	OIND; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-33	-0,624	0,092	-1,350	0,102	*	
-32	-0,161	0,388	-0,526	0,205		
-31	-0,147	0,388	-0,480	0,186		
-30	-0,452	0,000	-0,688	-0,216	*	*
-29	-0,737	0,000	-0,919	-0,555	*	*
-28	-0,554	0,000	-0,675	-0,434	*	*
-27	-0,602	0,000	-0,700	-0,505	*	*
-26	-0,707	0,000	-0,797	-0,617	*	*
-25	-0,657	0,000	-0,734	-0,581	*	*
-24	-0,828	0,000	-0,892	-0,765	*	*
-23	-0,871	0,000	-0,930	-0,812	*	*
-22	-0,762	0,000	-0,813	-0,711	*	*
-21	-0,942	0,000	-0,988	-0,897	*	*
-20	-1,134	0,000	-1,179	-1,089	*	*
-19	-1,610	0,000	-1,657	-1,563	*	*
-18	-1,491	0,000	-1,537	-1,446	*	*
-17	-1,567	0,000	-1,612	-1,522	*	*
-16	-1,059	0,000	-1,101	-1,017	*	*
-15	-1,129	0,000	-1,172	-1,087	*	*
-14	-0,535	0,000	-0,574	-0,496	*	*
-13	-0,433	0,000	-0,467	-0,399	*	*
-12	-0,402	0,000	-0,436	-0,367	*	*
-11	-0,241	0,000	-0,271	-0,212	*	*
-10	-0,623	0,000	-0,658	-0,589	*	*
-9	-0,338	0,000	-0,366	-0,310	*	*
-8	-0,570	0,000	-0,602	-0,538	*	*
-7	-0,531	0,000	-0,561	-0,501	*	*
-6	-0,655	0,000	-0,690	-0,620	*	*
-5	-0,520	0,000	-0,557	-0,483	*	*
-4	-0,536	0,000	-0,575	-0,497	*	*
-3	-0,331	0,000	-0,368	-0,294	*	*
-2	-0,229	0,000	-0,261	-0,196	*	*
-1	-0,225	0,000	-0,256	-0,193	*	*
0	-0,494	0,000	-0,530	-0,457	*	*
1	-0,750	0,000	-0,794	-0,706	*	*
2	-0,952	0,000	-1,004	-0,901	*	*
3	-1,193	0,000	-1,250	-1,135	*	*
4	-1,352	0,000	-1,416	-1,288	*	*
5	-1,268	0,000	-1,339	-1,198	*	*
6	-1,212	0,000	-1,291	-1,134	*	*
7	-1,128	0,000	-1,216	-1,040	*	*
8	-1,064	0,000	-1,165	-0,963	*	*
9	-1,022	0,000	-1,145	-0,900	*	*
10	-0,847	0,000	-0,998	-0,696	*	*
11	-0,876	0,000	-1,065	-0,686	*	*
12	-0,883	0,000	-1,118	-0,649	*	*
13	-1,196	0,000	-1,482	-0,911	*	*
14	-1,540	0,000	-1,892	-1,187	*	*
15	-1,673	0,000	-2,101	-1,244	*	*
16	-2,399	0,000	-2,909	-1,888	*	*
17	-2,861	0,000	-3,485	-2,237	*	*
18	-2,872	0,000	-3,586	-2,157	*	*
19	-1,986	0,000	-2,793	-1,178	*	*
20	-1,823	0,000	-2,712	-0,933	*	*
21	-1,282	0,011	-2,272	-0,292	*	*
22	-1,440	0,009	-2,519	-0,362	*	*
23	-2,000	0,001	-3,146	-0,854	*	*
24	-1,579	0,010	-2,775	-0,383	*	*
25	-1,608	0,004	-2,693	-0,522	*	*
26	-0,314	0,596	-1,477	0,849		
27	-0,261	0,665	-1,443	0,922		
28	-0,495	0,437	-1,748	0,758		
29	-0,090	0,903	-1,542	1,361		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S11e - Mean difference of TCEMIIa frequency class between blood and CSF in OIND patients

	OIND; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	**SIGN adj**	**SIGN**
-32	-0,218	0,309	-0,637	0,202		
-31	-0,071	0,672	-0,401	0,259		
-30	-0,561	0,000	-0,808	-0,314	*	*
-29	-0,877	0,000	-1,088	-0,666	*	*
-28	-0,710	0,000	-0,833	-0,588	*	*
-27	-0,509	0,000	-0,613	-0,405	*	*
-26	-0,715	0,000	-0,802	-0,628	*	*
-25	-0,457	0,000	-0,524	-0,389	*	*
-24	-1,076	0,000	-1,146	-1,006	*	*
-23	-1,318	0,000	-1,377	-1,259	*	*
-22	-1,107	0,000	-1,158	-1,055	*	*
-21	-1,485	0,000	-1,542	-1,429	*	*
-20	-0,630	0,000	-0,670	-0,591	*	*
-19	-1,402	0,000	-1,447	-1,357	*	*
-18	-1,146	0,000	-1,192	-1,100	*	*
-17	-1,240	0,000	-1,282	-1,197	*	*
-16	-0,956	0,000	-0,996	-0,916	*	*
-15	-0,403	0,000	-0,435	-0,370	*	*
-14	-0,476	0,000	-0,511	-0,441	*	*
-13	-0,632	0,000	-0,669	-0,595	*	*
-12	-0,617	0,000	-0,653	-0,580	*	*
-11	-0,457	0,000	-0,485	-0,429	*	*
-10	-0,489	0,000	-0,520	-0,459	*	*
-9	-0,302	0,000	-0,328	-0,275	*	*
-8	-0,591	0,000	-0,631	-0,551	*	*
-7	-0,468	0,000	-0,505	-0,431	*	*
-6	-0,516	0,000	-0,547	-0,484	*	*
-5	-0,303	0,000	-0,334	-0,272	*	*
-4	-0,396	0,000	-0,426	-0,366	*	*
-3	-0,252	0,000	-0,285	-0,220	*	*
-2	-0,436	0,000	-0,469	-0,403	*	*
-1	-0,449	0,000	-0,484	-0,414	*	*
0	-0,616	0,000	-0,655	-0,576	*	*
1	-0,706	0,000	-0,750	-0,663	*	*
2	-0,839	0,000	-0,886	-0,792	*	*
3	-0,839	0,000	-0,886	-0,792	*	*
4	-0,823	0,000	-0,875	-0,771	*	*
5	-0,707	0,000	-0,765	-0,649	*	*
6	-0,797	0,000	-0,863	-0,731	*	*
7	-0,804	0,000	-0,883	-0,725	*	*
8	-0,790	0,000	-0,884	-0,695	*	*
9	-0,945	0,000	-1,061	-0,828	*	*
10	-1,106	0,000	-1,250	-0,961	*	*
11	-1,229	0,000	-1,480	-1,050	*	*
12	-1,626	0,000	-1,845	-1,460	*	*
13	-1,910	0,000	-2,178	-1,643	*	*
14	-2,337	0,000	-2,674	-2,000	*	*
15	-2,358	0,000	-2,767	-1,949	*	*
16	-3,288	0,000	-3,776	-2,799	*	*
17	-3,002	0,000	-3,605	-2,400	*	*
18	-2,456	0,000	-3,144	-1,769	*	*
19	-2,759	0,000	-3,545	-1,973	*	*
20	-2,323	0,000	-3,182	-1,464	*	*
21	-1,379	0,000	-2,316	-0,441	*	*
22	-2,031	0,000	-2,969	-1,092	*	*
23	-1,228	0,024	-2,293	-0,163	*	*
24	-1,328	0,014	-2,387	-0,269	*	*
25	-0,638	0,216	-1,652	0,375	*	
26	-0,426	0,405	-1,429	0,578		
27	0,124	0,823	-0,964	1,212		
28	-0,157	0,807	-1,414	1,101		
29	-0,186	0,786	-1,536	1,163		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

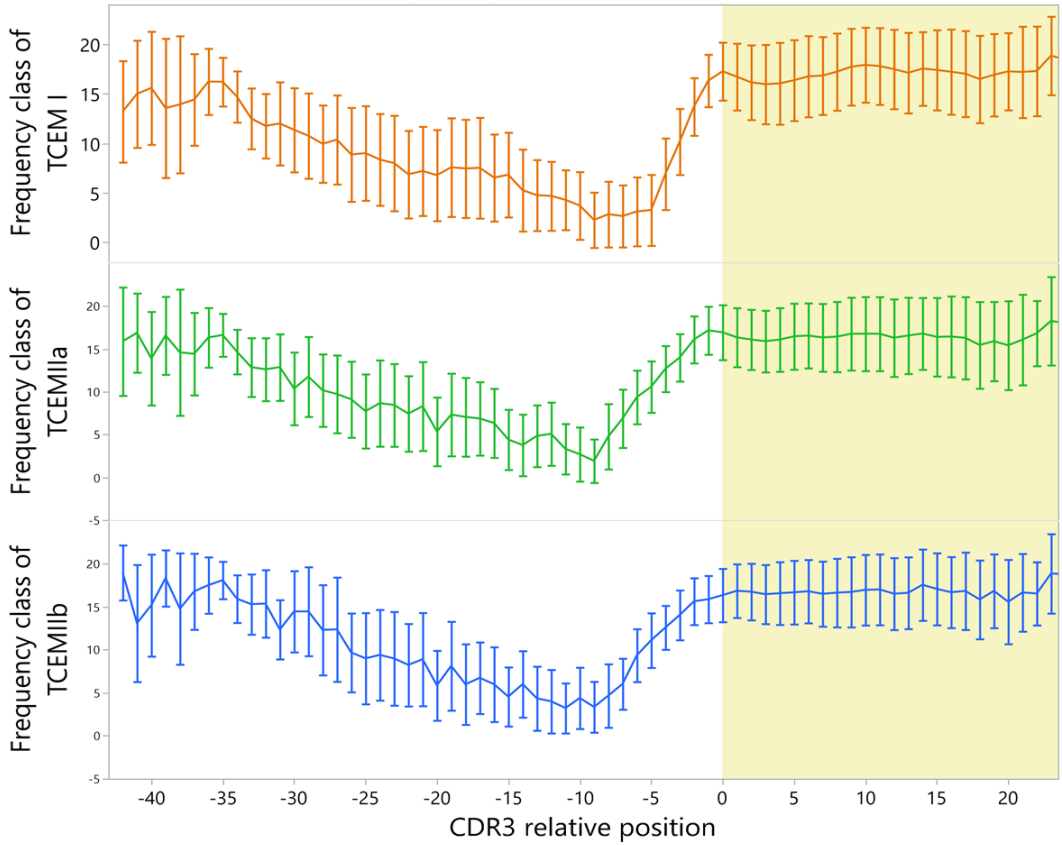
Supplementary Table S11f - Mean difference of TCEMIIb frequency class between blood and CSF in OIND patients

	OIND; Blood vs CSF	mean(0)-mean(1)			Benj.-Hoch.	Estimated
RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-32	-0,092	0,649	-0,491	0,306		
-31	0,115	0,474	-0,200	0,430		
-30	-0,685	0,000	-0,903	-0,466	*	*
-29	-0,867	0,000	-1,086	-0,648	*	*
-28	-0,698	0,000	-0,828	-0,567	*	*
-27	-0,466	0,000	-0,570	-0,362	*	*
-26	-0,778	0,000	-0,862	-0,695	*	*
-25	-0,555	0,000	-0,620	-0,489	*	*
-24	-1,006	0,000	-1,076	-0,936	*	*
-23	-1,065	0,000	-1,118	-1,012	*	*
-22	-1,152	0,000	-1,206	-1,098	*	*
-21	-1,653	0,000	-1,711	-1,595	*	*
-20	-0,554	0,000	-0,590	-0,518	*	*
-19	-1,493	0,000	-1,541	-1,444	*	*
-18	-0,703	0,000	-0,748	-0,657	*	*
-17	-1,081	0,000	-1,122	-1,040	*	*
-16	-0,696	0,000	-0,735	-0,658	*	*
-15	-0,574	0,000	-0,605	-0,542	*	*
-14	-0,907	0,000	-0,945	-0,870	*	*
-13	-0,377	0,000	-0,414	-0,341	*	*
-12	-0,743	0,000	-0,779	-0,707	*	*
-11	-0,578	0,000	-0,605	-0,550	*	*
-10	-0,553	0,000	-0,588	-0,517	*	*
-9	-0,417	0,000	-0,447	-0,386	*	*
-8	-0,566	0,000	-0,604	-0,527	*	*
-7	-0,259	0,000	-0,291	-0,226	*	*
-6	-0,481	0,000	-0,512	-0,450	*	*
-5	-0,274	0,000	-0,309	-0,241	*	*
-4	-0,364	0,000	-0,392	-0,335	*	*
-3	-0,326	0,000	-0,358	-0,295	*	*
-2	-0,534	0,000	-0,565	-0,503	*	*
-1	-0,472	0,000	-0,506	-0,439	*	*
0	-0,535	0,000	-0,573	-0,497	*	*
1	-0,600	0,000	-0,641	-0,560	*	*
2	-0,611	0,000	-0,655	-0,568	*	*
3	-0,622	0,000	-0,668	-0,576	*	*
4	-0,668	0,000	-0,718	-0,619	*	*
5	-0,634	0,000	-0,688	-0,579	*	*
6	-0,719	0,000	-0,781	-0,657	*	*
7	-0,701	0,000	-0,773	-0,629	*	*
8	-0,613	0,000	-0,697	-0,530	*	*
9	-0,705	0,000	-0,807	-0,604	*	*
10	-0,870	0,000	-0,992	-0,748	*	*
11	-0,952	0,000	-1,101	-0,803	*	*
12	-1,341	0,000	-1,521	-1,161	*	*
13	-1,421	0,000	-1,640	-1,203	*	*
14	-1,841	0,000	-2,114	-1,568	*	*
15	-1,873	0,000	-2,208	-1,539	*	*
16	-2,479	0,000	-2,879	-2,079	*	*
17	-2,305	0,000	-2,810	-1,799	*	*
18	-2,083	0,000	-2,659	-1,507	*	*
19	-2,798	0,000	-3,472	-2,124	*	*
20	-1,725	0,000	-2,506	-0,943	*	*
21	-1,429	0,000	-2,258	-0,600	*	*
22	-1,734	0,000	-2,561	-0,907	*	*
23	-0,574	0,224	-1,500	0,352	*	
24	-1,043	0,035	-2,010	-0,076	*	*
25	0,437	0,371	-0,523	1,398		
26	-0,014	0,978	-0,981	0,953		
27	-0,198	0,711	-1,246	0,850		
28	-0,459	0,421	-1,581	0,663		
29	-0,219	0,714	-1,398	0,959		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

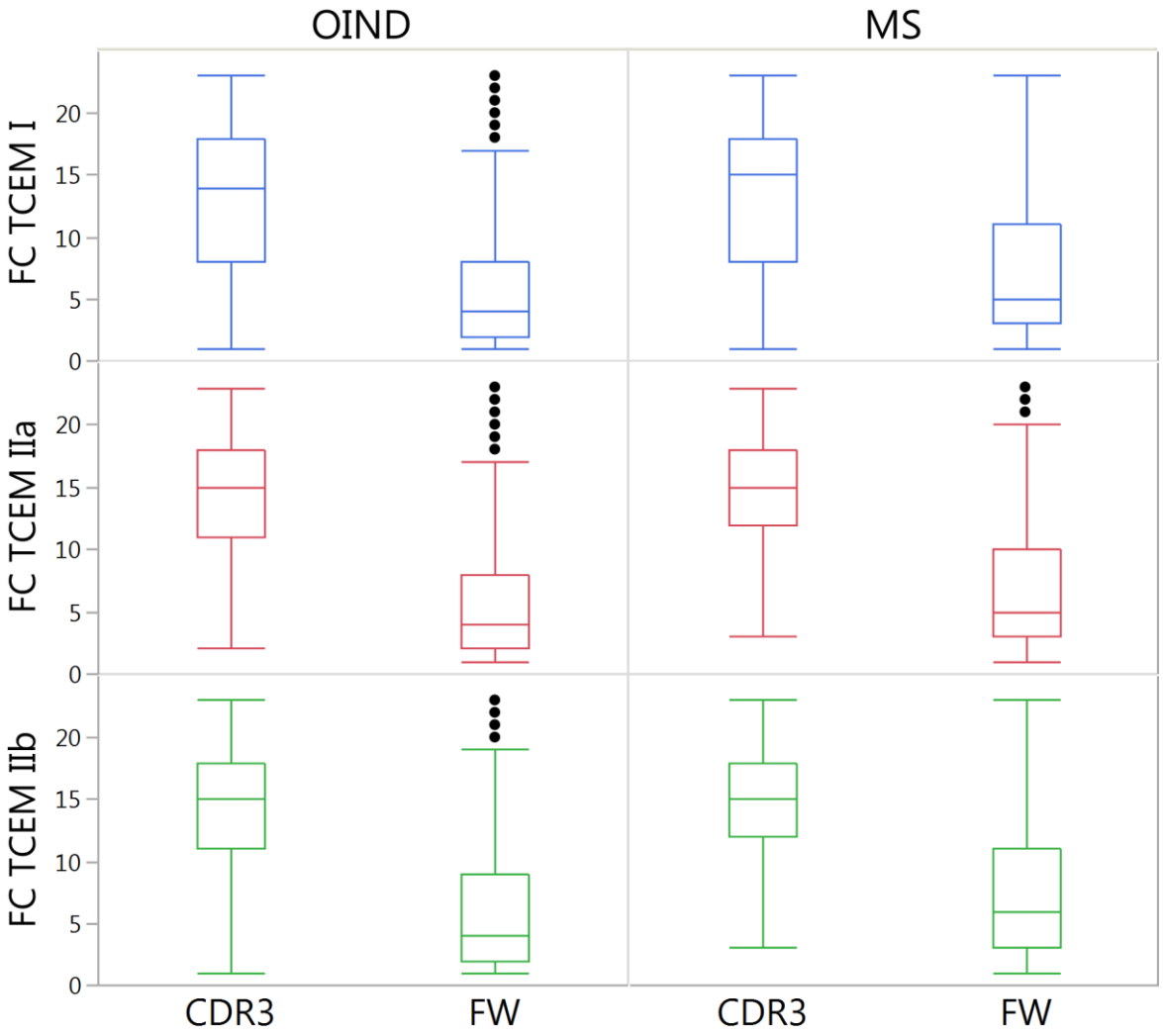
S12 – Comparison of mean FC in the CDR3 vs FW3

Supplementary Figure S12a –Mean frequency of T cell exposed motifs in IGHV transcripts.



We identified all TCEM in our patients CSF IGHV fragments, and used TCEM occurrences in three healthy individuals to generate a –log2 based FC for each TCEM. Mean frequency class (FC) of all patients in the dataset is displayed for all CDR3 relative positions. Low values represent frequent TCEM. Each error bar is constructed using 1 standard deviation from the mean. The CDR3 relative position refers to first N-terminus amino acid of a 15-mer in the case of TCEM II and a 9-mer in the case of TCEM I.

Supplementary Figure S12b –Mean frequency class of CDR3 vs FW3 derived fragments in IGHV transcripts



We compared the mean FC of CDR3- vs FW3-derived fragments by splitting the transcripts at CDR3 relative position -7. Mean FC by region (CDR3 and FW3) and by disease are shown as outlier box plots with whiskers covering 1st and 3rd quartile -/+ 1.5*(interquartile range). Supplementary Table S12 shows the adjusted means used for statistical testing

Supplementary Table S12 - TCEM Frequency class mean difference in FC in FW3 vs CDR3

		FC for TCEM I	FC for TCEM IIa	FC for TCEM IIb
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	-	-	-
	CDR3 relative position	-	-	-
	transcript clone-level	-	-	-
MS	Mean difference	-5.81	-7.99	-7.46
	LCI	-5.83	-8.01	-7.49
	UCI	-5.78	-7.97	-7.44
	p-value	<0.001	<0.001	<0.001
		Unadjusted for cluster effect	Unadjusted for cluster effect	Unadjusted for cluster effect
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	1.2	1.19	1.27
	CDR3 relative position	-	6.53	5.82
	transcript clone-level	-	-	-
OIND	Mean difference	-6.75	-8.83	-8.31
	LCI	-6.77	-8.84	-8.32
	UCI	-6.73	-8.81	-8.29
	p-value	<0.001	<0.001	<0.001
		Unadjusted for cluster effect	Unadjusted for cluster effect	Unadjusted for cluster effect

ICC – Intra class correlation, L/UCI – Lower and Upper Confidence interval

S13 – Statistical tests of mean FC of TCEMs

Supplementary Table S13a – Intra class correlation of FC on patient level

TCEM_I	
RelPos	ICC (%)
-7	23,6
-6	21,4
-5	19,2
-4	14,2
-3	11,0
-2	14,9
-1	20,9
0	12,0
1	12,0
2	14,4
3	16,2
4	11,7
5	16,0
6	20,2
7	10,8
8	10,8
9	9,9
10	10,9
11	8,4
12	10,7
13	8,3
14	12,4
15	10,5
16	14,1
17	12,7
18	11,3
19	16,6
20	17,9
21	6,9
22	29,1
23	26,3
24	30,4
25	37,8
26	40,0
27	29,2
28	26,3
29	0,0

TCEM_IIa	
RelPos	ICC (%)
-7	15,8
-6	22,5
-5	21,8
-4	13,8
-3	10,3
-2	11,4
-1	8,2
0	21,6
1	9,1
2	11,5
3	13,9
4	14,7
5	18,0
6	9,7
7	10,5
8	10,7
9	13,2
10	14,0
11	8,1
12	7,2
13	9,6
14	8,0
15	11,2
16	12,8
17	13,4
18	10,1
19	9,0
20	32,7
21	16,5
22	21,0
23	24,3
24	25,3
25	29,2
26	25,0
27	0,0
28	31,6
29	57,8
30	0,0

TCEM_IIb	
RelPos	ICC (%)
-7	13,6
-6	19,9
-5	15,5
-4	14,6
-3	24,2
-2	13,6
-1	12,4
0	15,0
1	6,4
2	11,4
3	18,4
4	17,5
5	16,9
6	14,1
7	18,8
8	15,0
9	11,0
10	10,2
11	12,1
12	7,0
13	9,5
14	10,5
15	12,3
16	13,6
17	12,9
18	8,2
19	14,1
20	28,4
21	19,0
22	13,9
23	28,8
24	29,2
25	32,5
26	27,6
27	10,3
28	36,5
29	55,0
30	0,0

ICC=Intra class correlation

Supplementary Table S13b – Mean difference in FC of TCEMI motifs between MS and OIND patients

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	0,636	0,163	-0,257	1,529		
-6	0,381	0,411	-0,528	1,290		
-5	0,031	0,943	-0,809	0,871		
-4	0,006	0,984	-0,596	0,608		
-3	-0,397	0,038	-0,772	-0,022	*	*
-2	-0,441	0,063	-0,906	0,025	*	
-1	-0,342	0,318	-1,012	0,328		
0	-0,019	0,928	-0,425	0,388		
1	0,336	0,134	-0,103	0,775		
2	0,690	0,013	0,144	1,237	*	*
3	0,874	0,008	0,233	1,516	*	*
4	0,674	0,003	0,233	1,116	*	*
5	0,697	0,039	0,036	1,358	*	*
6	0,896	0,032	0,077	1,715	*	*
7	0,447	0,016	0,085	0,808	*	*
8	0,440	0,008	0,115	0,765	*	*
9	0,254	0,101	-0,050	0,557		
10	0,277	0,074	-0,027	0,582	*	
11	0,239	0,033	0,019	0,459	*	*
12	0,322	0,014	0,064	0,580	*	*
13	0,212	0,042	0,008	0,417	*	*
14	0,339	0,009	0,084	0,595	*	*
15	0,165	0,178	-0,075	0,406		
16	0,166	0,322	-0,163	0,495		
17	0,084	0,571	-0,207	0,376		
18	0,070	0,623	-0,208	0,348		
19	0,075	0,702	-0,308	0,457		
20	0,154	0,413	-0,215	0,523		
21	0,094	0,522	-0,195	0,384		
22	0,809	0,028	0,087	1,530	*	*
23	0,406	0,269	-0,314	1,126		
24	0,235	0,510	-0,464	0,935		
25	-0,583	0,261	-1,600	0,433		
26	-0,148	0,847	-1,654	1,358		
27	-0,016	0,987	-1,994	1,962		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S13c – Mean difference in FC of TCEMIIa motifs between MS and OIND patients

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,118	0,717	-0,755	0,520		
-6	0,258	0,568	-0,629	1,145		
-5	0,067	0,873	-0,757	0,892		
-4	-0,229	0,287	-0,651	0,192		
-3	-0,031	0,853	-0,360	0,298		
-2	-0,034	0,854	-0,394	0,326		
-1	0,241	0,052	-0,002	0,485		
0	0,611	0,146	-0,213	1,434		
1	0,348	0,034	0,027	0,670	*	*
2	0,462	0,030	0,046	0,879	*	*
3	0,141	0,640	-0,449	0,730		
4	-0,099	0,762	-0,741	0,543		
5	0,593	0,000	0,533	0,654	*	*
6	0,260	0,177	-0,117	0,636		
7	0,044	0,839	-0,385	0,473		
8	0,437	0,023	0,059	0,815	*	*
9	0,104	0,697	-0,419	0,627		
10	0,394	0,125	-0,109	0,896		
11	0,340	0,005	0,100	0,580	*	*
12	0,041	0,756	-0,217	0,299		
13	0,474	0,000	0,239	0,708	*	*
14	0,219	0,080	-0,026	0,465		
15	0,211	0,202	-0,113	0,534		
16	0,375	0,036	0,025	0,726	*	*
17	-0,045	0,838	-0,480	0,390		
18	0,099	0,569	-0,241	0,439		
19	-0,097	0,576	-0,436	0,243		
20	-0,205	0,743	-1,433	1,022		
21	0,271	0,379	-0,333	0,875		
22	0,641	0,064	-0,036	1,318		
23	-0,024	0,962	-1,016	0,968		
24	0,354	0,470	-0,606	1,315		
25	-0,207	0,760	-1,535	1,121		
26	0,708	0,317	-0,679	2,095		
27	-1,000	0,220	-2,598	0,598		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S13d – Mean difference in FC of TCEMIIb motifs between MS and OIND patients

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,041	0,865	-0,515	0,432		
-6	0,247	0,517	-0,500	0,994		
-5	-0,261	0,369	-0,831	0,308		
-4	-0,173	0,432	-0,605	0,259		
-3	0,394	0,385	-0,495	1,283		
-2	0,167	0,443	-0,260	0,595		
-1	0,288	0,131	-0,086	0,663		
0	0,383	0,152	-0,140	0,906		
1	0,140	0,215	-0,081	0,361		
2	0,371	0,068	-0,027	0,770		
3	-0,066	0,867	-0,843	0,710		
4	-0,139	0,720	-0,897	0,619		
5	0,586	0,092	-0,096	1,268		
6	0,414	0,134	-0,128	0,957		
7	0,359	0,373	-0,431	1,150		
8	0,703	0,006	0,201	1,206	*	*
9	0,303	0,134	-0,094	0,701		
10	0,263	0,145	-0,090	0,616		
11	0,244	0,242	-0,165	0,653		
12	0,059	0,628	-0,180	0,298		
13	0,439	0,000	0,208	0,671	*	*
14	0,089	0,600	-0,244	0,422		
15	0,136	0,480	-0,242	0,515		
16	0,303	0,152	-0,111	0,718		
17	-0,130	0,953	-0,439	0,413		
18	-0,024	0,879	-0,334	0,286		
19	0,085	0,757	-0,451	0,620		
20	-0,105	0,848	-1,177	0,967		
21	0,129	0,722	-0,581	0,839		
22	0,590	0,041	0,025	1,154		*
23	-0,188	0,775	-1,478	1,101		
24	-0,330	0,654	-1,770	1,110		
25	-0,498	0,585	-2,284	1,287		
26	0,811	0,336	-0,843	2,465		
27	-1,635	0,041	-3,203	-0,067		*

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S13e – Mean difference in FC of TCEMI motifs between IGHV4 and other IGHV family fragments

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,954	0,000	-1,010	-0,897	*	*
-6	-1,188	0,000	-1,250	-1,125	*	*
-5	-0,890	0,000	-0,955	-0,824	*	*
-4	-0,679	0,000	-0,746	-0,613	*	*
-3	-0,709	0,000	-0,771	-0,647	*	*
-2	-0,567	0,000	-0,621	-0,513	*	*
-1	-0,414	0,000	-0,463	-0,366	*	*
0	-0,040	0,153	-0,094	0,015		
1	0,152	0,000	0,090	0,214	*	*
2	0,156	0,000	0,086	0,225	*	*
3	0,306	0,000	0,233	0,379	*	*
4	0,188	0,000	0,113	0,262	*	*
5	0,401	0,000	0,329	0,473	*	*
6	0,483	0,000	0,414	0,551	*	*
7	0,232	0,000	0,168	0,295	*	*
8	0,273	0,000	0,213	0,334	*	*
9	0,188	0,000	0,129	0,246	*	*
10	0,292	0,000	0,236	0,349	*	*
11	0,309	0,000	0,251	0,367	*	*
12	0,217	0,000	0,157	0,277	*	*
13	0,037	0,250	-0,026	0,099		
14	0,042	0,178	-0,019	0,103		
15	-0,002	0,942	-0,069	0,064		
16	-0,067	0,090	-0,144	0,010	*	
17	0,200	0,000	0,110	0,291	*	*
18	-0,092	0,118	-0,207	0,023	*	
19	0,071	0,284	-0,059	0,202		
20	0,037	0,652	-0,123	0,197		
21	-0,039	0,770	-0,298	0,221		
22	0,011	0,955	-0,373	0,395		
23	-0,166	0,516	-0,667	0,335		
24	0,036	0,937	-0,862	0,935		
25	0,318	0,530	-0,675	1,312		
26	0,184	0,815	-1,353	1,721		
27	-0,112	0,958	-4,307	4,082		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

Supplementary Table S13f – Mean difference in FC of TCEMIIa motifs between IGHV4 and other IGHV family fragments

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,890	0,000	-0,953	-0,827	*	*
-6	-0,070	0,017	-0,127	-0,013	*	*
-5	-0,530	0,000	-0,585	-0,475	*	*
-4	-0,192	0,000	-0,241	-0,142	*	*
-3	-0,654	0,000	-0,705	-0,603	*	*
-2	-0,301	0,000	-0,352	-0,250	*	*
-1	0,187	0,000	0,135	0,239	*	*
0	0,140	0,000	0,081	0,199	*	*
1	0,187	0,000	0,123	0,251	*	*
2	0,117	0,000	0,051	0,183	*	*
3	0,022	0,533	-0,046	0,089		
4	0,110	0,002	0,042	0,179	*	*
5	0,308	0,000	0,239	0,377	*	*
6	0,299	0,000	0,232	0,366	*	*
7	0,142	0,000	0,075	0,210	*	*
8	0,102	0,003	0,033	0,170	*	*
9	0,132	0,000	0,063	0,201	*	*
10	0,311	0,000	0,242	0,381	*	*
11	0,029	0,414	-0,041	0,100		
12	-0,093	0,014	-0,167	-0,019	*	*
13	0,059	0,142	-0,020	0,137	*	
14	0,155	0,000	0,073	0,237	*	*
15	0,023	0,611	-0,065	0,110		
16	0,084	0,099	-0,016	0,185	*	
17	-0,136	0,036	-0,264	-0,009	*	*
18	0,071	0,372	-0,085	0,227		
19	0,632	0,000	0,443	0,820	*	*
20	-0,158	0,197	-0,398	0,082		
21	-0,123	0,484	-0,467	0,221		
22	0,491	0,050	-0,001	0,982	*	
23	-0,784	0,029	-1,487	-0,082	*	*
24	-0,142	0,850	-1,614	1,329		
25	0,347	0,674	-1,270	1,963		
26	-0,465	0,694	-2,779	1,849		
27	-0,839	0,738	-5,744	4,067		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

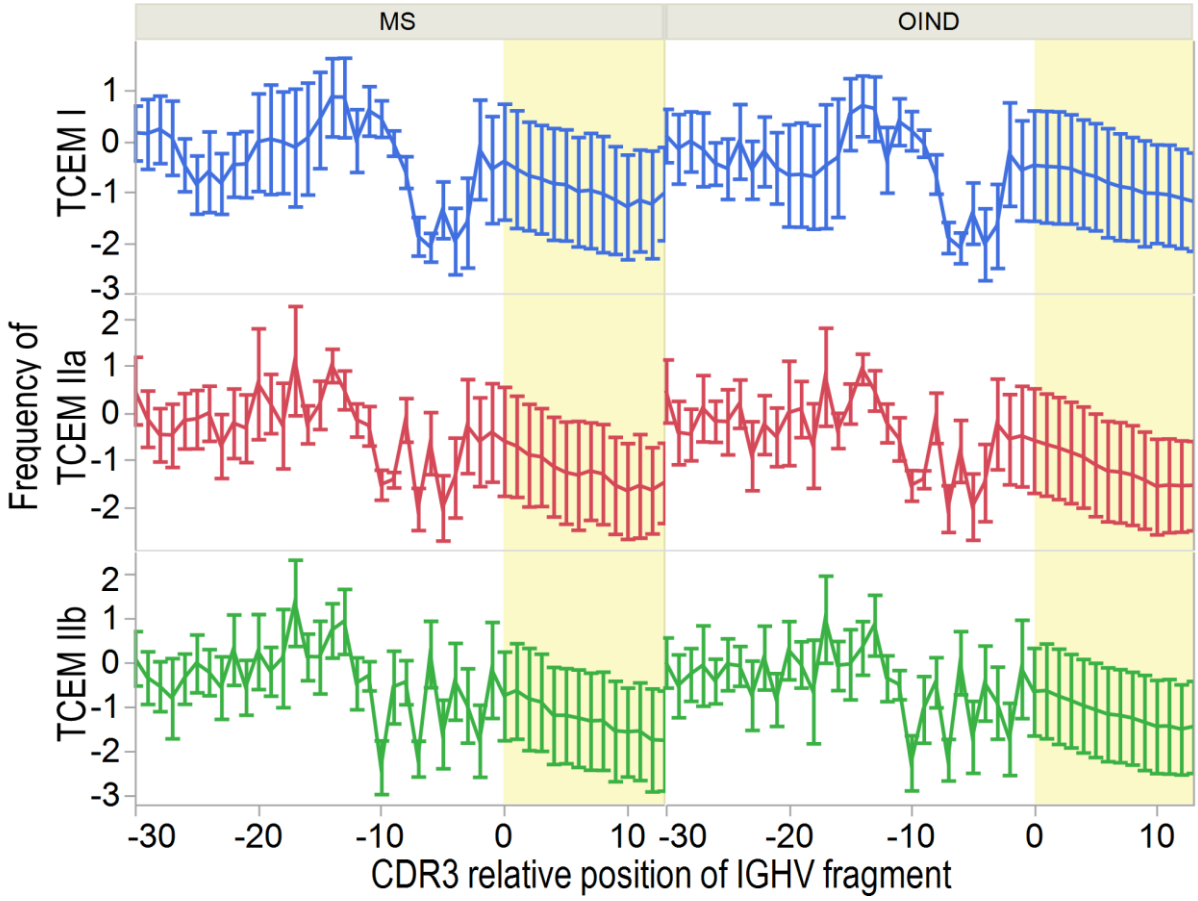
Supplementary Table S13g – Mean difference in FC of TCEMIIb motifs between IGHV4 and other IGHV family fragments

RelPos	Mean diff	p-value	LCI	UCI	SIGN adj	SIGN
-7	-0,486	0,000	-0,541	-0,431	*	*
-6	-0,209	0,000	-0,266	-0,152	*	*
-5	-1,436	0,000	-1,493	-1,378	*	*
-4	-0,107	0,000	-0,155	-0,060	*	*
-3	-0,450	0,000	-0,503	-0,397	*	*
-2	0,162	0,000	0,112	0,212	*	*
-1	0,075	0,004	0,024	0,126	*	*
0	-0,378	0,000	-0,436	-0,321	*	*
1	0,107	0,000	0,049	0,165	*	*
2	-0,054	0,084	-0,116	0,007	*	
3	-0,121	0,000	-0,185	-0,057	*	*
4	0,177	0,000	0,110	0,243	*	*
5	0,199	0,000	0,132	0,266	*	*
6	0,320	0,000	0,255	0,385	*	*
7	0,285	0,000	0,219	0,351	*	*
8	0,299	0,003	0,232	0,366	*	*
9	0,178	0,000	0,112	0,245	*	*
10	0,150	0,000	0,082	0,218	*	*
11	-0,046	0,194	-0,115	0,023		
12	-0,089	0,015	-0,161	-0,018	*	*
13	-0,040	0,306	-0,118	0,037		
14	0,439	0,000	0,358	0,520	*	*
15	-0,069	0,130	-0,158	0,020	*	
16	0,038	0,471	-0,065	0,142		
17	0,140	0,033	0,011	0,268	*	*
18	-0,212	0,010	-0,374	-0,050	*	*
19	0,475	0,000	0,267	0,684	*	*
20	-0,359	0,005	-0,610	-0,108	*	*
21	-0,734	0,000	-1,093	-0,376	*	*
22	0,472	0,101	-0,092	1,036	*	
23	-1,195	0,003	-1,977	-0,414	*	*
24	0,419	0,653	-1,410	2,248		
25	1,108	0,276	-0,885	3,101		
26	-1,142	0,390	-3,748	1,464		
27	-1,344	0,591	-6,249	3,532		

LCI=Lower confidence interval. UCI=upper confidence interval. ICC=Intra class correlation
Black stars indicate significant differences. Red stars indicate significant differences after correction for multiple testing (20% FDR, Benjamini-Hochberg)

S14 – TCEM occurrences in the gut microbiome vs IGHV fragments

Supplementary Figure S14 – TCEM occurrence in the gut microbiome assigned to patients IGHV TCEMs



We compared the Johnson standardized occurrence of TCEM in gut microbiome of CDR3- vs FW3-derived fragments by splitting the transcripts at CDR3 relative position -7. Mean Johnson standardized occurrence by IGHV region and by disease are shown as outlier box plots with whiskers covering 1st and 3rd quartile -/+ 1.5*(interquartile range). Supplementary Table S14 shows the adjusted means used for statistical testing. Figure S14 shows the mean Johnson standardized occurrence in gut microbiome in IGHV transcripts by their CDR3 relative positions (first N-terminus amino acid of a 15-mer in the case of TCEM II and a 9-mer in the case of TCEM I). Each error bar is constructed using 1 standard deviation from the mean.

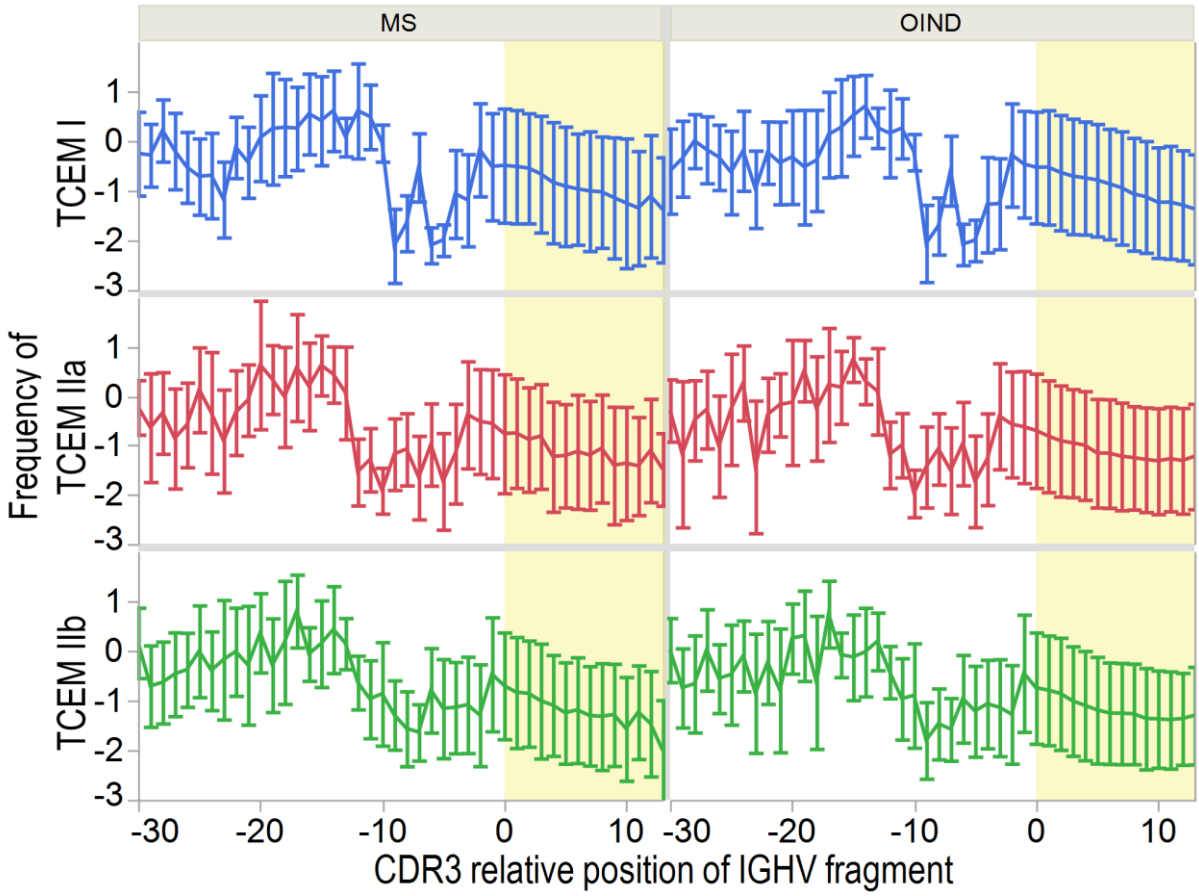
Supplementary Table S14 - Gut microbiome TCEM frequencies in FW3 vs CDR3

		Johnson SI transformed TCEM I frequency	Johnson SI transformed TCEM IIa frequency	Johnson SI transformed TCEM IIb frequency
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	1.03	1.23	-
	CDR3 relative position	2.84	4.12	3.76
	Transcript clone level	-	-	-
MS	Mean difference	1.78	0.67	0.53
	LCI	1.76	0.66	0.52
	UCI	1.79	0.68	0.54
	p-value	<0.001	<0.001	<0.001
		Adjusted for cluster effect CDR3 relative position-level	Adjusted for cluster effect CDR3 relative position-level	Adjusted for cluster effect CDR3 relative position-level
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	0.34	-	-
	CDR3 relative position	-	3.53	-
	Transcript clone-level	-	-	-
OIND	Mean difference	0.84	0.81	0.72
	LCI	0.84	0.80	0.72
	UCI	0.85	0.81	0.72
	p-value	<0.001	<0.001	<0.001
		Adjusted for cluster effect Patient-level	Unadjusted for cluster effect	Unadjusted for cluster effect

ICC – Intra class correlation, L/UCI – Lower and Upper Confidence interval

S15 – TCEM occurrences in the human proteome vs IGHV fragments

Supplementary Figure S15 – Mean Johnson standardized TCEM counts in the human proteome assigned to patients IGHV fragments TCEMs



We compared the Johnson standardized occurrence of TCEM in human proteome of CDR3- vs FW3-derived fragments by splitting the transcripts at CDR3 relative position -7. Mean Johnson standardized occurrence by IGHV region and by disease are shown as outlier box plots with whiskers covering 1st and 3rd quartile -/+ 1.5*(interquartile range). Supplementary Table S15 shows the adjusted means used for statistical testing. Figure S15 shows the mean Johnson standardized occurrence in human proteome in IGHV transcripts by their CDR3 relative positions (first N-terminus amino acid of a 15-mer in the case of TCEM II and a 9-mer in the case of TCEM I). Each error bar is constructed using 1 standard deviation from the mean

Supplementary Table S15 - Human proteome TCEM frequencies in FW3 vs CDR3

		Johnson SI transformed TCEM I frequency	Johnson SI transformed TCEM IIa frequency	Johnson SI transformed TCEM IIb frequency
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	1.49	-	-
	CDR3 relative position	2.71	3.17	3.86
	transcript clone-level	-	-	-
MS	Mean difference	0.76	0.14	0.25
	LCI	0.75	0.13	0.24
	UCI	0.76	0.15	0.26
	p-value	<0.001	<0.001	<0.001
		Adjusted for cluster effect CDR3 relative position	Adjusted for cluster effect CDR3 relative position	Adjusted for cluster effect CDR3 relative position
		ICC (%)	ICC (%)	ICC (%)
	Patient-level	0.34	-	0.35
	CDR3 relative position	-	2.69	3.2
	transcript clone-level	-	-	-
OIND				
	Mean difference	0.66	0.29	0.20
	LCI	0.65	0.28	0.19
	UCI	0.66	0.30	0.20
	p-value	<0.001	<0.001	<0.001
		Unadjusted for cluster effect	Adjusted for cluster effect CDR3 relative position	Adjusted for cluster effect CDR3 relative position

ICC – Intra class correlation, L/UCI – Lower and Upper Confidence interval

Paper II

II

International Journal of
Molecular Sciences

MDPI

Article

Human Cysteine Cathepsins Degrade Immunoglobulin G In Vitro in a Predictable Manner

Rune Alexander Høglund [1,2,3,*,†], Silje Bøen Torsetnes [1,2,†], Andreas Lossius [1,2,4], Bjarne Bogen [4], E. Jane Homan [5], Robert Bremel [5] and Trygve Holmøy [1,3]

1 Department of Neurology, Akershus University Hospital, 1478 Lørenskog, Norway; silje.torsetnes@medisin.uio.no (S.B.T.); andreas.lossius@medisin.uio.no (A.L.); trygve.holmoy@medisin.uio.no (T.H.)
2 Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital and University of Oslo, 1478 Lørenskog, Norway
3 Institute of Clinical Medicine, University of Oslo, 0372 Oslo, Norway
4 Department of Immunology and Transfusion Medicine, Faculty of Medicine, University of Oslo, 0372 Oslo, Norway; bjarne.bogen@medisin.uio.no
5 ioGenetics LLC, Madison, WI 53704, USA; jane_homan@iogenetics.com (E.J.H.); robert_bremel@iogenetics.com (R.B.)
* Correspondence: r.a.hoglund@medisin.uio.no
† These authors contributed equally to this work.

Received: 9 September 2019; Accepted: 27 September 2019; Published: 29 September 2019

Abstract: Cysteine cathepsins are critical components of the adaptive immune system involved in the generation of epitopes for presentation on human leukocyte antigen (HLA) molecules and have been implicated in degradation of autoantigens. Immunoglobulin variable regions with somatic mutations and random complementarity region 3 amino acid composition are inherently immunogenic. T cell reactivity towards immunoglobulin variable regions has been investigated in relation to specific diseases, as well as reactivity to therapeutic monoclonal antibodies. Yet, how the immunoglobulins, or the B cell receptors, are processed in endolysosomal compartments of professional antigen presenting cells has not been described in detail. Here we present in silico and in vitro experimental evidence suggesting that cysteine cathepsins S, L and B may have important roles in generating peptides fitting HLA class II molecules, capable of being presented to T cells, from monoclonal antibodies as well as from central nervous system proteins including a well described autoantigen. By combining neural net models with in vitro proteomics experiments, we further suggest how such degradation can be predicted, how it fits with available cellular models, and that it is immunoglobulin heavy chain variable family dependent. These findings are relevant for biotherapeutic drug design as well as to understand disease development. We also suggest how these tools can be improved, including improved machine learning methodology.

Keywords: cathepsin; endosome; endolysosome; protease; B cell; antigen presenting cell; bioinformatics; in silico model; protease cleavage prediction

1. Introduction

The endosomal system of antigen presenting cells (APCs) is home to cysteine cathepsins (B, C, F, H, K, L, O, S, V, W and X), serine cathepsins (A and G), aspartyl cathepsins (D and E), legumain (asparagine endopeptidase, AEP), and gamma-interferon inducible thiol reductase (GILT) [1–3]. The expression differs with maturation and activation status of APCs [1,4]. B cells are increasingly investigated as APCs for $CD4^+$ T cells, as their APC functions have been connected to disease pathophysiology [1,5–7]. Upon activation, the B cell receptor (BCR) and bound antigen are internalized, the antigen is degraded

in the endolysosomal system and the resulting fragments of both may bind to major histocompatibility complex (MHC) class II for presentation on the cell surface [1,8–10].

Cysteine cathepsins have previously been implied in aging and neurodegenerative disorders and have been detected in microglia, astrocytes, or neurons, in addition to traditional APCs [11]. Further, neuroinflammation is increasingly investigated in what is traditionally considered neurodegenerative disorders [12]. Several central nervous system (CNS) proteins are associated with disease and are either known or potential targets for cysteine cathepsins, including amyloid beta and Tau (Alzheimer's disease), alpha-synuclein (Parkinson's disease), and myelin basic protein (MBP, multiple sclerosis) [11].

Monoclonal antibody (mAb) drugs are increasingly being used and developed as therapy for cancer, inflammatory, autoimmune, and other diseases [13,14]. They consist of immunoglobulins (Igs) with constant regions of varying isotypes and allotypes, and variable antigen binding regions of either mouse, human, or humanized origin, which make them inherently immunogenic [15,16]. Development of antibodies towards mAbs is dependent on degradation of Igs by B cells and T cell help [17]. Similar mechanisms have been demonstrated in mice models [18,19].

Observed immunogenicity of therapeutic mAbs could not fully be explained by human leukocyte antigen (HLA)-affinity and T cell epitope predictions alone [20]. Upon internalization of BCR-Ig complexes human GILT allows reduction of disulfide bonds [21–23], and endolysosomal proteases likely participate in further degradation of the Igs [24]. In murine bone-marrow derived APCs (non-B cells), cathepsins B and S were important for degrading $F(ab')_2$, after internalization via the FcγR [25]. We have previously described how cysteine cathepsins S, L and B were predicted to cleave human Ig variable regions in specific patterns, and suggested specific roles for these in the degradation of Igs and possibly BCRs allowing presentation of potentially immunogenic fragments on HLA class II [26,27]. Of these cathepsins, S and B are well expressed in B cells, while all three are expressed in monocytes and microglia [1,4,28,29].

While processing and presentation of internalized antigens are frequently investigated, the fate of BCRs upon activation remains poorly described. Still, it has been demonstrated in mice that B cells process and present fragments from their own BCRs on surface MHC class II molecules [10,30,31]. More recently such presentation was found to be extensive in human B cell lymphomas [32,33]. It is likely that cysteine cathepsins degrade both antigen and BCR alike. As Igs and BCRs share common structures [34], understanding degradation of Igs including mAbs could improve our understanding of BCR fragment presentation on HLA class II.

As with other antigens, understanding processing and presentation of Ig requires estimates of processing in the endo-lysosome compartment. Here we present in silico and in vitro experimental validation for cathepsins activity prediction models using CNS proteins including a well described autoantigen (MBP), as well as six therapeutic mAbs. The results suggest that cysteine cathepsins S, L and B effectively degrade both CNS proteins and immunoglobulin G (IgG) in specific and predictable patterns in acidic and reducing conditions simulating endolysosomal compartments.

2. Results

2.1. Prediction Platform Validation: In Silico Evaluations

Cathepsin peptidases have the ability to cleave many different cleavage site octamers (CSOs) and each enzyme family has activity on substrates that is strongly dependent on the amino acids upstream and downstream of the scissile bond. During the development of the prediction platform it was found that a single general scheme that encompassed all (i.e., 400) different scissile bond dipeptides was not achievable. Thus, an approach was developed wherein each unique P1P1′ scissile bond dipeptide has its own set of neural network (NN) ensembles; each scissile dipeptide in a protein is computed with a neural network ensemble specific for that dipeptide and each cathepsin has several hundred different ensembles. Although the in silico cross-validation of our prediction model platform had previously

demonstrated an approximately 90% true positive and 10% false positive rate [35], in an effort to simplify the process here we additionally compared the accuracy to a different machine learning model (support vector machine—SVM) (Supplementary Figure S3), used for binary prediction models such as cleave/no cleavage in the case of cathepsins. In this evaluation, the scissile bond-specific NN ensembles out-performed the SVM in predicting the number of cleavages, indicating that the original NN model is adequately suited for protease cleavage prediction.

2.2. Prediction Platform Validation: In Vitro Findings Compared to In Silico Predictions on CNS Proteins

Although training of the NN ensembles employed the best practices available for the task, the size of the training sets is small in comparison to those typically used for large scale artificial intelligence and machine learning. The accuracy of the NN models for full sized proteins had not previously been assessed and the original training set comprised fragments of proteins of partially digested human cells [36]. As different mAbs contain largely similar protein structures, using these alone for validation would cause redundancy in testing. Therefore, we tested the validity of the predictions for full size CNS proteins that may be degraded by cells expressing the cysteine cathepsins (recombinant myelin basic protein [rMBP]-2, rMBP-6, Tau, or α-synuclein), using in vitro experiments at pH 6, as described in the Method section. To evaluate quality of samples, peptides by sample were clustered using Ward's method (Supplementary Figure S4), showing high similarity between samples with the same protein and cathepsin and no or very few peptides detected in negative controls (30 h incubation). This indicates both lack of impurities or cross-contamination, and sparse spontaneous degradation. The peptide size distribution from different incubation times (Figure 1) indicated substantial cathepsin induced cleavage of the substrates already after 6 h. All cathepsins generated peptides of comparable lengths, ranging from 6 to 45 amino acids, with more than 40% falling into an HLA class II fitting range of 11–20 amino acids after 24 h of cleavage.

Next, we sought to compare predictions to observed cleavage of the CNS proteins. We quantified and standardized the number of observed cleavages at every CSO after 24 h of incubation with either cathepsin S, L or B using the nano-liquid chromatography mass spectrometry (nLCMS results). The CSOs for all proteins were combined into a single dataset, along with the prediction model cleavage probabilities for the same CSOs. All CSOs were classified by their cleavage probability into grouped ranges (0–0.19, 0.2–0.39, 0.4–0.59, 0.6–0.79, and 0.8–1) and the groups were compared to identify any correlation between cleavage probability and standardized cleavage observations (Figure 2). Of note, CSOs with a low predicted cleavage probability (<0.20) vastly outnumber the other binned groups and reflect the combinatorial effects of the flanking amino acids. The neural net model performed well for cathepsin S and L predictions, as higher predicted probability for cleavage was associated with higher number of cleavages. Also, for over 63% of CSOs with the highest probabilities of cleavage, we observed at least one cleavage. The cathepsin B model underperformed, with a relatively high number of cleavages observed when not predicted (0–0.2 probability). This could possibly be related to its joint endo- and carboxypeptidase capabilities [37]. Such a property will inherently influence the prediction accuracy. This phenomenon is illustrated in Supplementary Figure S5, where the observed number of cleavages for rMBP-2 is plotted by relative maximum distance to a high predicted probability (>0.8) cleavage site. A slight curve-shift to the left could be observed for cathepsin B, but not for L or S, consistent with possible combined endo- and carboxypeptidase activity of cathepsin B.

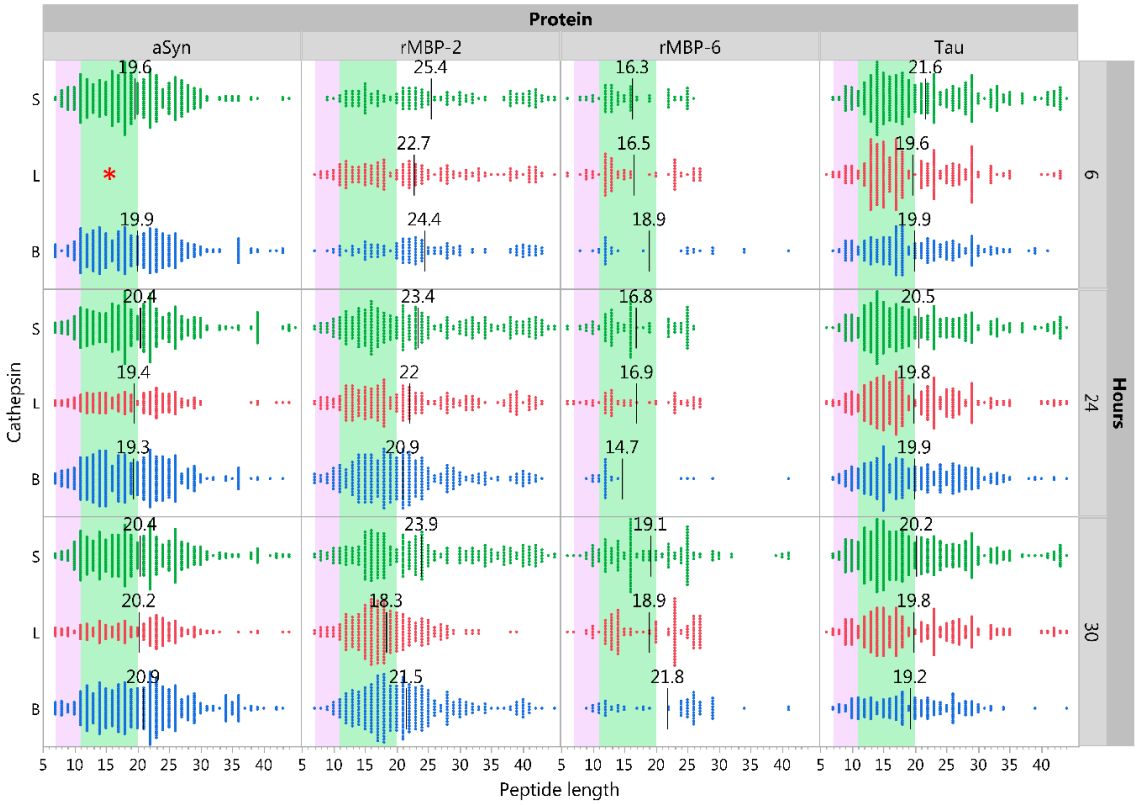


Figure 1. Peptide lengths resulting from in vitro cathepsin digestion of central nervous system proteins. Distribution of peptide lengths after digestion of alpha-synuclein (aSyn), recombinant myelin basic protein (rMBP) isoforms 2 and 6, and tau with either cathepsin B, L, or S at 6, 24, or 30 h at pH 6. Each data point represents one identified peptide at the given time point. Black lines with annotations indicate the mean size of peptides. Purple and green areas indicate peptide sizes fitting HLA class I and II, respectively. * aSyn 6-h sample for cathepsin L was lost due to technical error.

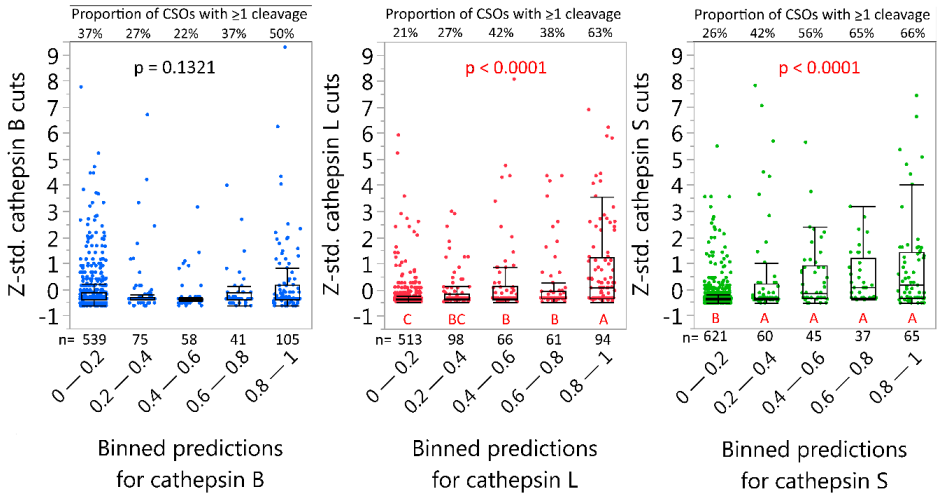


Figure 2. Comparison of predicted and observed cleavage of CNS proteins. All potential cleavage site octamers (CSOs) within alpha-synuclein, recombinant myelin basic protein isoforms 2 and 6, and tau were binned into ranges of 0.2 based on the predicted cleavage probability (X-axis). Intra-protein z-standardized number of observed cuts after 24 h at corresponding CSOs are depicted on the Y-axis. The *p*-values indicate Welch ANOVA significance for cathepsin B/L/S (F(4, 1.53/13.03/12.24)) and differing letters indicate binned groups that have significant difference in mean number of observed cleavages (Tukey–Kramer, HSD). Whiskers are outlier box-plots.

2.3. Cysteine Cathepsins Degrade Immunoglobulins In Vitro

As the NN models performed adequately on peptide cocktails (in silico tests) as well as full sized proteins (in vitro tests), it seemed likely that our previous predicted effects of cathepsins on Igs or BCRs could be relevant [26]. To examine if these cathepsins efficiently degraded Igs, we followed the same procedure as described above for CNS proteins, mixing the mAbs rituximab, natalizumab, alemtuzumab, adalimumab, ocrelizumab, or infliximab individually with each cathepsin at pH 6. Unlike for the CNS proteins, cathepsin S yielded significantly more nLCMS detectable IgG peptides than cathepsins L or B (Figure 3A). The size distributions of IgG peptides were compatible with both HLA class I and II grooves and did not seem to vary much between the different mAbs (Figure 3B). This indicates that single cathepsins can generate IgG fragments for presentation on HLA, and that the cathepsin S, known to be expressed in B cells more than cathepsins L [1], is superior in this function at pH 6.

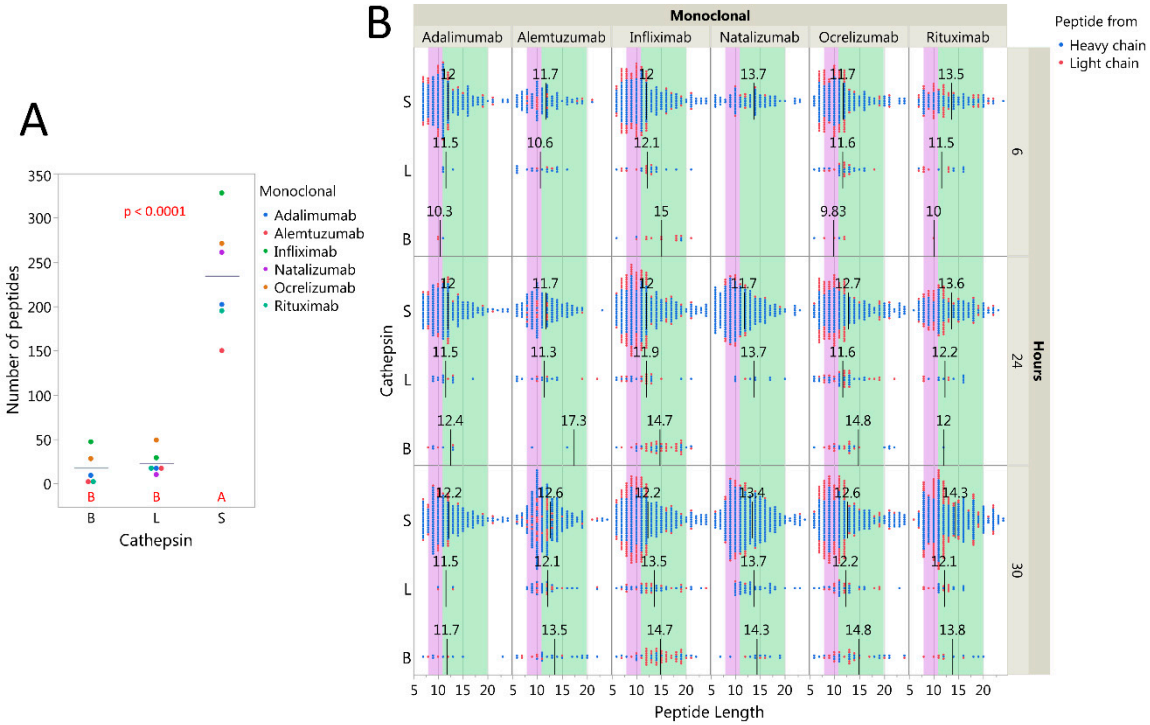


Figure 3. Digestion of monoclonal antibodies at pH 6 by cathepsins S, L and B. Detected peptides after digestion of 1200 nM alemtuzumab, rituximab, natalizumab, or 2400 nM adalimumab, infliximab, or ocrelizumab with either cathepsin B, L, or S at 6, 24, or 30 h at pH 6. (**A**) Cathepsin S yields significantly more detectable peptides than cathepsins L and B at pH 6, after 24 h of incubation. The bars indicate average number of peptides detected. Significance as determined by ANOVA testing and Tukey–Kramer HSD (different red letters indicate significant difference between groups). (**B**) Distribution of peptide lengths (x-axis). Each data point represents one identified peptide at the given time point. Black lines with annotations indicate the mean size of peptides. Purple and green areas indicate peptide sizes fitting HLA class I and II, respectively. The length range is cropped to display 99% of the peptides.

As the size distribution of IgG peptides were compatible with HLA presentation, we went on to investigate from which regions these peptides were derived, focusing mainly on cathepsin S. The primary protein structures of the heavy and light chains of all six mAbs (Table S1) were utilized to align the identified peptides to the corresponding amino- and carboxy-end cleavage locations. Figure 4A,C display a relatively fixed pattern of degradation for constant regions of both heavy and light chains. A small cleavage location shift was observed for natalizumab heavy chain, due to the inherent sequence difference between IgG4 and IgG1. Interestingly, the heavy constant 2 regions

seemed to be most sensitive to cleavage across the mAbs. Thus, the cathepsins demonstrated a capability of cleaving a variety of CSOs consistently across several mAbs. Cleavages observed for the variable regions contrasts this, as patterns differed between the mAbs (Figure 4A,C). A notable difference was the higher number of observed cleavages and cleavage positions in heavy chains for the chimeric infliximab and rituximab compared to the other mAbs, which carry humanized or human variable regions (Figure 4C,D).

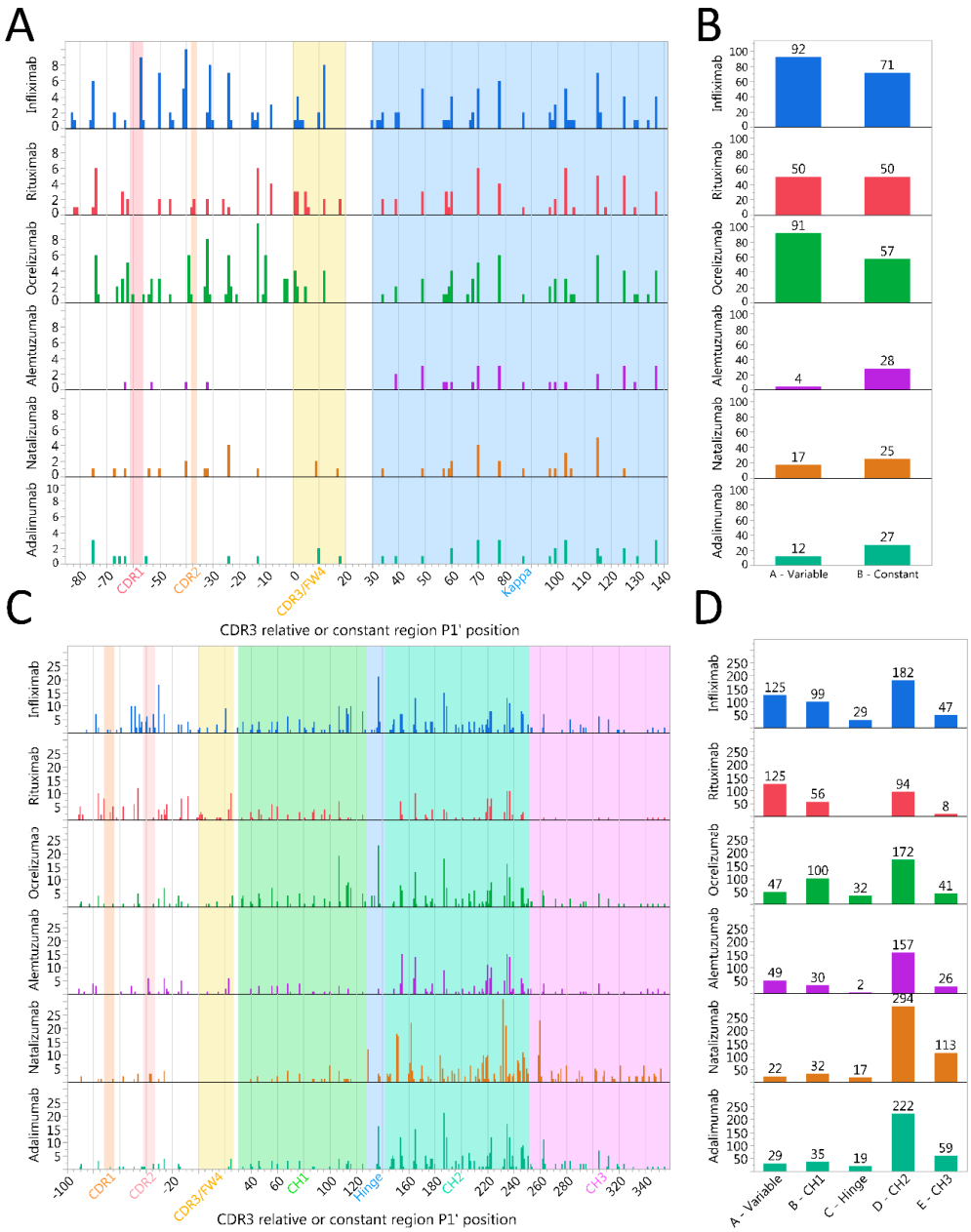


Figure 4. Observed pattern of cathepsin S cuts in monoclonal antibodies. The monoclonal antibodies adalimumab, alemtuzumab, infliximab, natalizumab, ocrelizumab, and infliximab were incubated with cathepsin s for 24 h at pH 6. Non-standardized number of observed cuts in light (**A**/**B**) and heavy chains (**C**/**D**) identified by nano-liquid chromatography mass spectrometry. Cuts are presented by their location in sequence (**A**/**C**) or summarized by region (**B**/**D**). For alignment purposes, the variable region position is assigned by the relative position of P1′ in the cleavage site octamer to the cysteine (0) of CDR3. The constant regions are aligned to start at position 30.

2.4. Neural Net Prediction Accuracy for Immunoglobulin Cathepsin Cleavage

Based on the above results cleavage within the variable region is likely important for the immunogenicity of therapeutic mAbs. As the model can be used to individually assess the likelihood for such cleavage, we assessed the peptide distribution qualitatively, compared to predicted cleavage sites for alemtuzumab heavy chain variable and constant region 2 (Figure 5). Notably, many peptides seem to be derived from longer fragments and either start or end at a predicted cleavage site, but not necessarily a site with high probability of cleavage (>0.8). A larger pool of unique peptides was detected after 30 h than after 6 h (Figures 3B and 5).

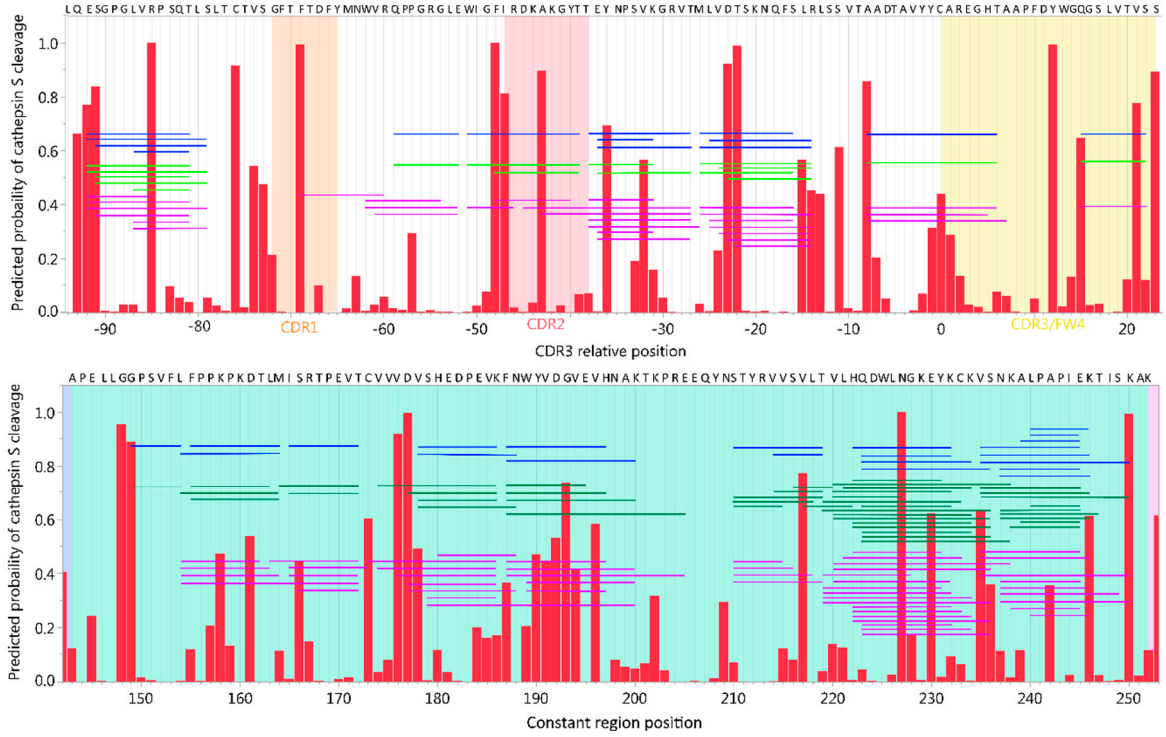


Figure 5. Detected peptides overlap with predicted cleavage sites. Predicted cleavage probability (x-axis) by cathepsin S in variable (upper panel) and constant heavy 2 (CH2) (lower panel) region of alemtuzumab. The vertical bars indicate the predicted position of P1′ of a P1-P1′ cleavage bond, and thus the first amino acid after a cut. Horizontal bars each indicate unique peptides detected starting at a P1′ and ending at a P1, as identified by nLCMS after 6 (blue), 24 (green), and 30 (purple) hours.

As with the CNS proteins, we further tested statistically the predictive models' accuracy for Ig variable region cleavage at pH 6 in a binned analysis. Cathepsin S predictions performed well, with high cleavage probability being associated with higher number of cleavages but were not as accurate as for the CNS proteins (Figure 6). For instance, only 45–50% of high probability cleavage sites had at least one cleavage observation. In addition, peptides found from the shorter IgG light chain seemingly fit better with predictions than heavy chain. This and the patterns shown in Figure 5 indicated that longer fragments resulting from incomplete cleavage, with lengths exceeding nLCMS method limitation, potentially remained undetected. Not surprisingly, the accuracy for cathepsins B and L was not as good as with cathepsin S, given fewer peptides on which to base the analysis (Supplementary Figure S6).

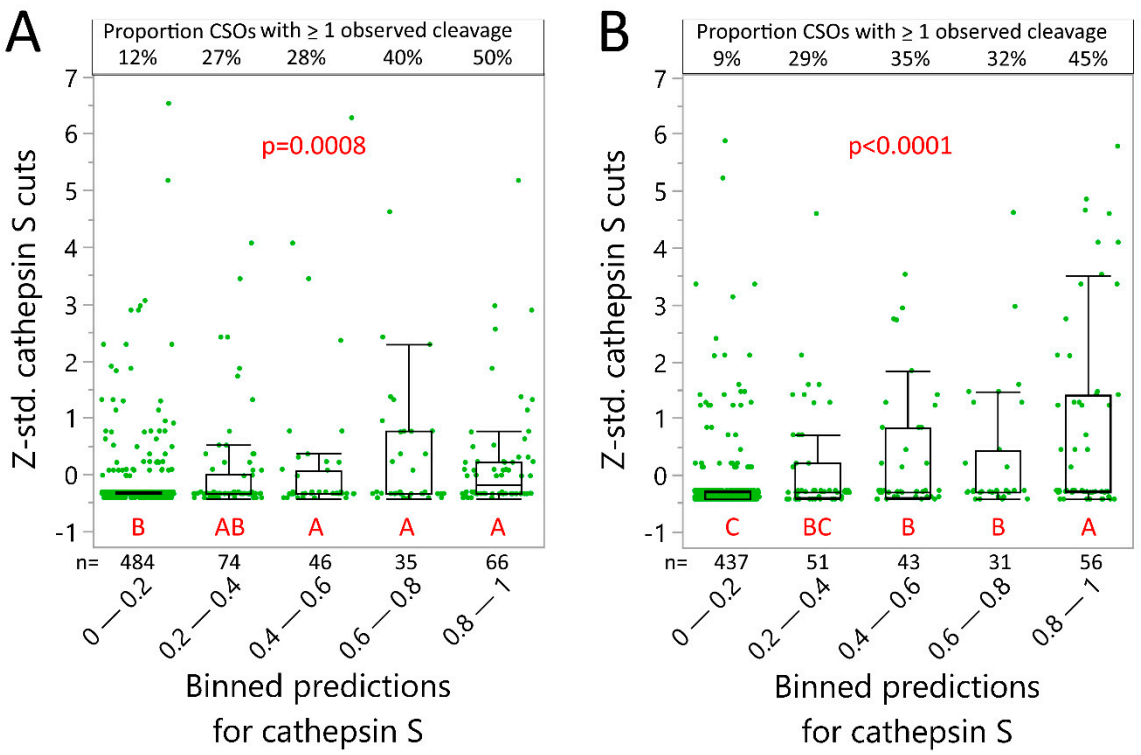


Figure 6. Evaluation of cleavage accuracy for monoclonal antibody variable regions. Cleavage probability by cathepsin S for all possible cleavage site octamers (CSOs) within (**A**) heavy and (**B**) light chain variable regions of rituximab, infliximab, ocrelizumab, natalizumab, alemtuzumab, and adalimumab were binned into ranges of 0.2 (X-axis). Intra-chain z-standardized number of observed cuts after 24 h at pH 6 are depicted on the Y-axis. *p*-values indicate Welch ANOVA significance (F(4, 5.16/9.05) for heavy and light respectively), and differing red letters indicate significant differences between groups (Tukey–Kramer, HSD). Whiskers are outlier box-plots.

2.5. Influence of pH on Cathepsin Activity

The pH optimum for cathepsins differs. Moreover, DTT reducing efficiency wanes at low pH [38], offering less reduction of IgG disulfide bonds that also could influence degradation patterns. We therefore further tested digestion by cathepsins S, L and B at pH 4 and 5 in addition to pH 6 (using only 1:100 enzyme to substrate ratios). Cathepsins L and B generated more peptides at lower pH, while pH had little influence on peptide yield for cathepsin S (Figure 7). Similar results were obtained for infliximab (Figure S8). The cathepsins also showed a relatively conserved cleavage pattern across multiple pHs for adalimumab, best illustrated by cathepsin S (Supplementary Figure S7) due to its preserved activity at pH 6. However, cathepsin L and B also display high levels of similarities when comparing pH 4 to pH 5 results (Supplementary Figure S7). As predictive models were built using datasets generated at pH 6, we did not test prediction accuracy at pH 4 and 5.

To assess whether DTT activity in fact was reduced at low pH, we performed sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) assays to assess residual mAb multimer structures. Negative samples were run with both reducing and non-reducing running buffers to account for reduction occurring in cleavage assay as well as the SDS-PAGE assay. It is evident that DTT activity was far more potent at pH 6 than pH 4, as several larger structures remained intact at pH 4 (Figure S9A). These likely reflected various combinations of heavy- and light chains sized 75 kDa (heavy + light), 100 kDa (2× heavy), 125 kDa (2× heavy + light), and 150 kDa (full IgG). Another observation was that bands around 50 kDa (heavy) and 25 kDa (light) were still abundant at pH 6, indicating incomplete degradation of these even in presence of reducing conditions, both for cathepsin incubated samples and negative control samples. The loss of multimeric structures was

time dependent, as demonstrated for ocrelizumab at pH 5 in Supplementary Figure S9B. Apart from cathepsin L at pH 4, only small differences were observed between cathepsin samples and negative controls for all pHs, implying that a considerable amount of heavy and light chains remained intact even after cathepsins processing.

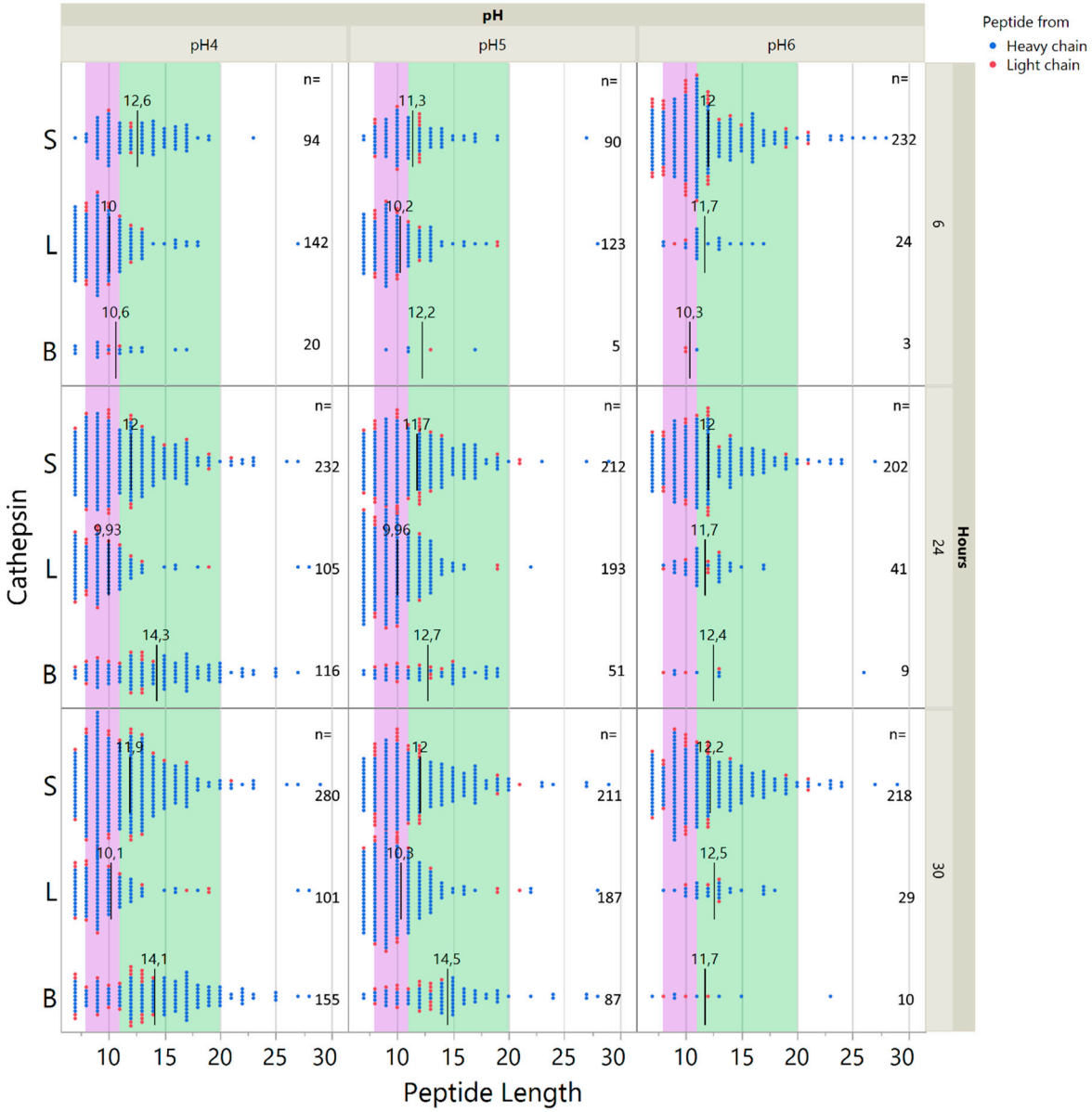


Figure 7. Adalimumab digestion by cathepsin S, L and B at pH 4, 5, and 6. Distribution of peptide lengths after digestion of 2400 nM adalimumab with either cathepsin B, L, S at 6, 24, or 30 h at pH 4, 5, or 6. Each data point represents one identified peptide at the given time point. Black lines with annotations indicate the mean size of peptides. Purple and green areas indicate peptide sizes fitting HLA class I and II, respectively. (Note: For pH 6, the data for cathepsins S and B are the same as in Figure 3B).

The cathepsins require reduction by e.g., DTT for activation but are also capable of auto-catalytic activation at acidic pH [37,39,40]. We observed that cathepsin activity was present despite loss of DTT efficiency, which indicated that the cathepsins most likely were auto-catalytically activated at acidic pH. Also, results indicate that cathepsins did not fully degrade the IgGs, which is compatible with a limited proteolytic activity for optimal generation of epitopes [41].

2.6. Immunoglobulin Heavy Variable Gene Family Determines Different Cleavage Patterns

Differences in amino acids patterns between the different IGHV families could likewise modulate the cathepsin cleavage patterns and thus be critical for immunogenicity of therapeutic mAbs. Previous data indicated that the immunoglobulin heavy variable (IGHV) family may dictate differences in degradation [26], and our findings here confirmed that such differences may be predicted to some extent. Thus, we sought to identify differences and/or similarities by using a previously assembled Ig variable region library [42], and plotting mean predicted cleavage probabilities for all CSOs using the C-terminal cysteine of CDR3 as an alignment to coordinate the relative position of P1′ (Figure 8). The mean probability of cleavages for cathepsin S clearly demonstrated different patterns of degradations by IGHV family, although some features were preserved. Notably, at CDR3 relative position −26, there was a preserved high probability for a cathepsin S cleavage site across all IGHV families, that was also consistently identified for all mAbs assessed with cathepsin S in vitro (Figures 4 and 8). In addition, a less pronounced but consistent increase in probability for cathepsin S cleavage across IGHV families was observed at the beginning of CDR3 (Figure 8). IGHV 3 had the lowest predicted cathepsin S probabilities for cleavage in the framework 3 region, consistent with our previous findings [26].

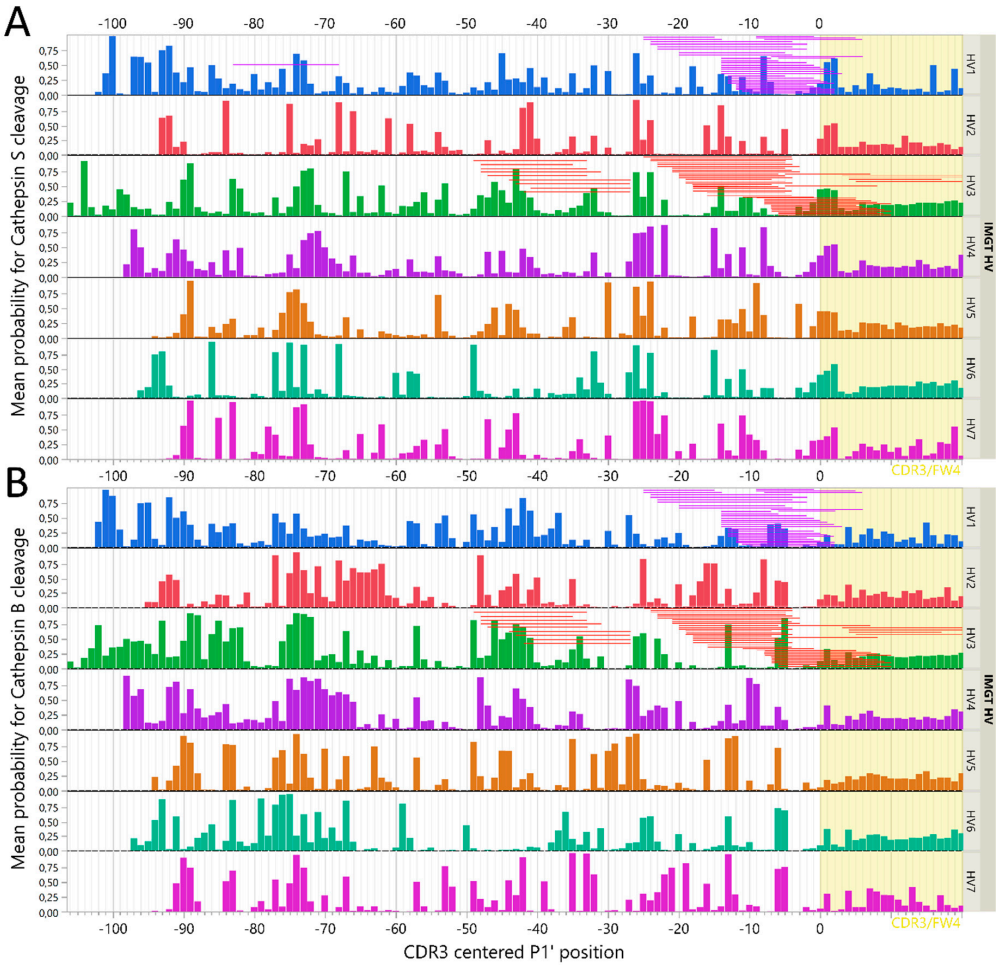


Figure 8. Predicted cleavage patterns of IGHV families using GenBank IGHV set. Approximately 16,000 curated IGHV sequences were divided by their V-family and analyzed with the cathepsin cleavage models. All possible cleavage site octamers (CSO) for each IGHV were aligned according the relative position of P1′ to the CDR3 (yellow) region cysteine (x-axis). Mean predicted probability for CSO cleavage at each position by cathepsin S (**A**) or cathepsin B (**B**) is shown on the y-axis. Superimposed are aligned IGHV peptides described by Khodadoust et al. (31), eluted from HLA class II of mantle cell lymphoma: MCL065 (purple) and MCL052 (red).

To investigate the validity of these assessments, we further compared the cathepsin B and S cleavage prediction data with those reported from IGHV-derived peptides eluted from HLA class II on lymphoma cells from two patients with mantle cell lymphoma [33]. As the full IGHV sequences of these clones were not available, we assembled the most complete IGHV sequence possible from each cell line (MCL052 and MCL065) and aligned them using the International Immunogenetics Information System (IMGT) database standards and assigned the assembled sequences to an IGHV family [43]. Then, the peptides were aligned according to the CDR3-relative position and compared to the observed cleavage pattern with predicted output of GenBank sequences as well as the mAb cleavage assays (Figure 8). Notably, many identified cleavage sites from the lymphoma IGHV peptides could be explained by either cathepsin B or S activity. For instance, cleavage around CDR3-relative position −26 for MCL052 (IGHV3) is compatible with cathepsin S protease activity, as it is evident from both the GenBank set and observed cuts in IGHV3-carrying mAbs (ocrelizumab and adalimumab). Likewise, consistent cleavage around position −5 can be explained by cathepsin B activity for IGHV3. Notably, several predicted cleavage sites confirmed by our in vitro studies were not identified in these peptides, possibly indicating a protective role of HLA class II binding. Compatible with this, IGHV 15-mers starting around CDR-3 relative positions −40, −20, and −5, as well as within the CDR3, was previously predicted to have high affinity for HLA-DR molecules [26,42].

3. Discussion

We hypothesized that CNS proteins and Ig variable regions are degraded in predictable patterns by cysteine cathepsins S, L and B in endolysosomal compartments of APCs [26]. Here, we have demonstrated such degradation patterns in vitro, showing how these cathepsins all degrade CNS proteins and IgGs into peptides sized to fit in HLA class II under conditions resembling the endolysosomal compartments. Further, we have validated in silico neural net models that can predict the pattern of such proteolysis.

The endolysosomal compartments are acidic and reducing [3], allowing proteases to degrade most foreign and self-proteins. Cathepsin L and S have both been attributed key importance in degrading class II-associated invariant chain peptide (CLIP) and preparing MHC class II for antigen binding, as well as antigen processing in general [1,2]. Several cathepsins are found in CNS cells [28], and cathepsin S and B in particular have suggested roles in neurodegenerative diseases [44]. It has been shown that cathepsin S has an important role in degradation of MBP [45], and we identified several peptides investigated for their potential immunogenicity ($MBP_{13–32}$, $MBP_{131–155}$, and $MBP_{146–170}$) [46], or associated cleavage sites, after cleavage with cathepsins L or S. Another variant, $MBP_{83–99}$, was both predicted and found to be destroyed by cathepsins S and L, as has also been described previously [45].

With heterogeneous degradation cleavage patterns [36] and presence in antigen presenting cells [1], a potential role for cathepsins S, L and B in degrading diverse Igs seemed likely. In this study, we confirmed that these cathepsins cause IgGs to be degraded in a pattern determined by their structure, as is evident from a fixed degradation pattern of constant region, and differing patterns in the variable regions.

Therapeutic mAbs are generally designed to minimize immunogenicity [47], yet anti-drug antibodies remain problematic. IgG antibodies make up the majority of anti-drug antibodies [48], and generation of such antibodies requires T cell help [17,19]. Due to the diversity of variable regions of heavy and light chains, we assume that the immunogenic T cell epitopes are derived from the variable regions, and several tools exist to make predictions to find them [49]. However, as data on Ig processing has been lacking, assumptions on processing are frequently absent in these tools. Here, we showed that cathepsins expressed by B cells efficiently generate epitopes from IGHV regions. Interestingly, chimeric antibody heavy chain variable regions were particularly prone to degradation, possibly contributing to their higher immunogenicity [20]. Parallel to this, it was shown that peptides introduced into human heavy constant 2 regions were more effectively presented on MHC II in mice

than peptides inserted into the other domains [50], consistent with the observed higher number of cleavages within this region (Figure 4C).

We and others have previously suggested that mutations in the IGHV region could break T-cell tolerance towards B cell receptors in vivo, leading to autoimmune disease [27,51–53]. Any small change, be it introduced by mutation or by design, could influence cathepsin cleavage patterns, and thus which IGHV peptides are presented. We have further attempted to model likelihood of such T-cell responsiveness to IGHV variable regions, using a combination of HLA class II affinity and cleavage by either cathepsin S, L or B [26]. However, the results of this study were based on in vitro experiments, that are not necessarily directly comparable to full-scale intracellular processing, and do not encompass the full complexity of the endolysosomal compartments. The intracellular machinery resulting in HLA class II presentation is intricate, involving a suitable cell activation state, endosomal environment, multiple cathepsins, GILT, HLA class II, and HLA-DM [54]. Protection from digestion by HLA class II binding may be particularly relevant. Nevertheless, several studies have published epitope libraries (www.iedb.org, [55]) demonstrating that peptides from Igs and/or BCRs are presented frequently on different APC's HLA class II molecules [56–59], and a few also performed IGHV sequencing to achieve an optimal search database [32,33]. Interestingly, IGHV peptides derived from dendritic cells loaded with therapeutic intravenous Igs [57] share similarities with IGHV peptides derived from self BCR in lymphomas [32,33], suggesting a similar mechanism of degradation.

It has been suggested that predicted cathepsin cleavage patterns did not explain HLA class II eluted IGHV peptides from the lymphomas [32]. This assessment may not have accounted for differential degradation of the IGHV families, nor the predicted high affinity for HLA-DR molecules of peptides in the framework 3 region [26,42]. We found that several HLA class II eluted IGHV peptides could be explained by either cathepsin B or S (Figure 8). Likewise, another group eluted HLA class II bound peptides from DCs incubated with infliximab or rituximab, and found several peptides compatible with both our predicted pattern and our observed peptides after cleavage with individual cathepsins [60]. Cleavage sites not explained by cathepsins described here, are likely the result of other endosomal proteases, including cathepsin H, as demonstrated for other substrates in more complex in vitro models [54], or legumain cleaving aspartic or asparagine bonds [61,62].

Based on the nLCMS results alone, one could presume that the IgGs were completely degraded by cathepsins, particularly at lower pH values. Yet, SDS-PAGE experiments unveiled a significant amount of heavy and light chains with relatively high molecular weights remaining after in vitro cathepsin processing. Additionally, cathepsin degradation may potentially have rendered some larger fragments that were not detected by gel analysis, due to differences in size and/or cleavage position. Even with the high sensitivity of a mass spectrometer, it is not possible to detect every cleavage site due to detection restrictions of the nLCMS instrument (typically 6–40 amino acid peptides). We also assume that identification of degradation close to a free carboxyl- or amino-end will be somewhat overestimated compared to that in the middle of large structures. This will skew the nLCMS output, and potentially explain a poorer prediction accuracy for the heavy chains. In complete endolysosomal systems of APCs, these restrictions may not apply, as different cathepsins likely work in tandem under reducing and increasingly acidic conditions to ensure proper degradation. In vitro models including multiple cathepsins [54], or unbiased HLA-elution assays accounting for both processing and HLA binding [33,63], can generate training sets further improving cleavage accuracy prediction of neural net models.

Cathepsin-generated epitopes are likely important for eliciting anti-drug antibodies, and the knowledge of these mechanisms is therefore important in the design of future therapeutic mAbs. Specific insight into B cell expressed cathepsin degradation of IgGs, as shown here, can supplement traditional epitope-mapping tools.

4. Methods

4.1. Cathepsin Cleavage Predictions

It is common practice to consider the amino acid contacts in a CSO, comprising ± 4 amino acids from the scissile bond, as the peptide contact region of a peptidase [64]. Cleavage occurs between amino acids 4 and 5 of the CSO. We have previously described the conversion of amino acid sequences into matrices of principal components of the physical properties of the amino acids as the input layer of neural networks [65,66]. For this study prediction of cleavage probability for cathepsin S, L and B were done with neural network models as described previously [26,35], trained using proteome derived-peptide library datasets from Biniossek et al. [36]. The method used was analogous to one used to predict peptide affinities for HLA class I and II [65,66]. In brief, neural net ensembles for each cathepsin were trained using principal components of amino acid physical properties of the CSO (Figures S1 and S2) to predict the cleavage probability of a peptide bond P1-P1′ of any P4P3P2P1-P1′P2′P3′P4′ octamer. Amino acid sequences were converted to 3-row matrices using the first three principal components that comprises approximately 90% of the variance in a range of different physical properties commonly used in structural biology [66]. The output of the neural networks ranged between 0 (low-) and 1 (high) probability for cleavage. Pseudo code for the training process is given in Figure S2 and derivation of the activation functions of the neural networks were done with the "Neural" platform of JMP® (SAS Institute, Cary, NC, USA). As the input is the primary amino acid sequence of proteins, once derived, the activation functions can be used to make predictions of any protein divided into sequential potential CSO. More details can be found in the supplemental section of an earlier publication [35], but are similar to those in common use in artificial intelligence modeling.

Predictions for cathepsin S, L and B cleavage were computed for every potential CSO in all substrates described below, as well as for 16.000 IGHV sequences previously curated from GenBank [42]. For IGHV sequences, family was assigned according to IMGT [43], and CSO P1′ positions were indexed and aligned by their relative position to the cysteine marking the beginning of CDR3 (position 0).

4.2. Cathepsins and Substrates of the In Vitro Cleavage Assays

In this study we used recombinant human cathepsin S (UniprotKB P25774), L (UniprotKB P07711), and B (UniprotKB P07858) from R&D Systems (Biotechne, Minneapolis, MN, USA). The two types of substrates in this study were proteins derived from the central nervous system and therapeutic mAbs. The CNS proteins were rMBP isoform 2 (Uniprot KB P02686-2) and a variant of isoform 6 (P02686-6) (LSBio, Seattle, WA, USA); recombinant alpha synuclein (α-synuclein) isoform 1 (P37840-1, rPeptide, Watkinsville, GA, USA), and microtubule-associated protein tau (Tau) isoform Tau-F (P10636-8, rPeptide, Watkinsville, GA, USA). The therapeutic mAbs were alemtuzumab (Lemtrada®, Genzyme, Cambridge, MA, USA), natalizumab (Tysabri®, Biogen, Cambridge, MA, USA), rituximab (Rixathon®, Sandoz, Holzkirchen, Germany), ocrelizumab (Ocrevus®, Roche, Basel, Switzerland), adalimumab (Humira®, Abbvie, North Chicago, IL, USA), and infliximab (Inflectra®, Pfizer, New York, NY, USA).

The cathepsins were mixed with substrate at ratios of 1:100 or 1:300 (*w*/*w*) in 50 mM sodium phosphate, 200 mM NaCl, 5 mM EDTA, and 4 mM dithiothreitol (DTT) at pH 4, 5, or 6. Substrate concentrations were 1200 or 2400 nM, to extensively exceed the mass spectrometry detection limit. Samples were incubated at 37 °C, on a shaker plate at 300 rpm for up to 30 h. Aliquots were incubated for 6, 24, or 30 h, and immediately frozen at −20 °C to stop catabolic activity. For each substrate, a negative control without cathepsin was run parallel to the experiments.

4.3. Nano Liquid Chromatography Mass Spectrometry and Related Software for Data Processing

All machines, equipment and software used for nLCMS were from Thermo Fisher Scientific (Waltham, MA, USA) unless otherwise stated.

The instrument performing nLC separation was a nano EasyLC1000, equipped with Accucore 150-C4 pre- and analytical columns (0.3 × 5 mm and 0.075 × 150 mm) used in a vented 2-column setup.

Mobile phases (MPs) were 0.1% formic acid in H_2O (MPA) and 0.1% formic acid in acetonitrile (MPB). Loading solution was also MPA and a volume of 5 uL of sample was injected at the flowrate 3 µL/min for every analysis. The mass spectrometer (MS) acquisition was turned on after injection and during acquisition the analytical flow rate was constant at 400 nL/min, being initially isocratic with 1% MPB for 1 min, before MPB was ramped up from 1% to 50% in 10 min and then from 50% to 70% in 2 min.

Mass spectrometry was performed with a QExactive Orbitrap with a heated electrospray ionization source operated at +2 kV. Data was acquired in a data-dependent manner by the following parameters: resolution of 70,000 in MS and 17500 in MS/MS, scan range from 350–1350 *m/z*, AGC target of 1e6 for MS, and 1e6 for MS/MS, top 7 selected for fragmentation, dynamic exclusion of 5 s, and exclusion of unknown charge.

Method setup and data acquisition was controlled by the Xcalibur™ software (version 2.2), while data processing and identification of peptides was performed using MaxQuant version 1.6.1.0 with the built in Andromeda search engine (freeware available at maxquant.org) [67]. Peptide false discovery rate was set to 0.01, and a mass tolerance of 5 ppm and 25 ppm was used in MS and MS/MS, respectively. Cleavage specificity was set to unspecific and methionine oxidation, N-terminal acetylation and asparagine deamidation were used as variable modifications, though no modified variants were detected.

The full sequences for substrates were acquired from the Uniprot database [68], and from the IMGT 2D/3Dstructure mAb-database or patent filings [69]. In some cases, the constant regions were imputed from existing literature [16]. The mAb sequences used are listed in Supplementary Table S1.

4.4. SDS-PAGE

SDS-PAGE was run under reducing or non-reducing conditions with one µg of select samples to assess residual IgG fragments with sizes exceeding the optimal nLCMS detection range (about >40 amino acid length). We utilized 4–20% Criterion™ TGX (Bio-Rad, Hercules, CA, USA) gels and Laemmli sample buffer (Bio-Rad, Hercules, CA, USA) with or without 50mM 1,4-dithiothreitol (DTT, Sigma-Aldrich, St. Louis, MO, USA) prepared according to the manufacturer's instructions. Gels were stained with Coomassie blue G-250 (Bio-Rad, Hercules, CA, USA) and photographed using ChemiDoc™ XRS+ with Image Lab™ software version 6.0.0 (Bio-Rad, Hercules, CA, USA).

4.5. Statistics

All statistical analyses and graphics were performed in JMP® Pro 14.1 (SAS Institute, Cary, NC, USA). For statistical testing of number of cleavages across proteins, intra-protein z-standardization of observed cleavage frequencies was used to improve comparability across substrates of differing lengths and/or concentrations. The accuracy of the models was only evaluated statistically for conditions emulating the conditions of the training sets (35). Unless otherwise stated, figures depict the cleavage position as the P1′ of the CSO, indicating the position of the first amino acid after cleavage. For graphic output purposes the position of P1′ was assigned relative to the cysteine at the start of CDR3 for variable regions and constant regions were aligned to start at position 30.

5. Conclusions

Using mass spectrometry proteomics techniques, we have demonstrated that NN ensembles derived using the principal components of physical properties of amino acids flanking the scissile bond, can predict in vitro proteolysis of both CNS proteins and mAbs by cathepsins S, L and B. While the constant regions of Igs follow a highly reproducible pattern of degradation, variable regions display differing patterns that are related to their IGHV family structure. This knowledge may be essential for understanding immune responses against both endogenous Igs and BCRs as well as therapeutic mAbs. As NN training is an ongoing process the CSO peptides in this study will enable re-training of the NN, improving their accuracy. These results further suggest that directed efforts towards expanding the

knowledge base regarding the specificity and expression patterns of peptidases involved in antigen presentation is warranted.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/19/4843/s1. Proteomics data from cathepsin cleavage assays can be found at: https://doi.org/10.6084/m9.figshare.9777725.

Author Contributions: Conceptualization, R.A.H., S.B.T., A.L. and T.H.; Data curation, R.A.H., S.B.T., E.J.H. and R.B.; Formal analysis, R.A.H. and S.B.T.; Funding acquisition, R.A.H., S.B.T. and R.T.H.; Investigation, R.A.H., S.B.T., A.L. and R.B.; Methodology, R.A.H., S.B.T., B.B., E.J.H. and R.B.; Project administration, R.T.H.; Resources, S.B.T., E.J.H., R.B. and R.T.H.; Software, E.J.H. and R.B.; Supervision, R.T.H.; Visualization, R.A.H.; Writing—original draft, R.A.H. and S.B.T.; Writing—review & editing, A.L., B.B., E.J.H., R.B. and R.T.H.

Funding: This study was funded by the Norwegian Research Council (grant 250864/F20), Akershus University Hospital internal strategic funding and an unrestricted research grant awarded by Novartis Norway.

Acknowledgments: We would like to thank Tormod Fladby for supporting the study. We also thank Jørgen Jahnsen at the Department of Gastroenterology, Akershus University Hospital for kindly supplying infliximab, and Bjørn Allan Moum at the Department of Gastroenterology, Oslo University Hospital for kind donation of adalimumab for our work. CNS protein models in the graphical abstract were built using SWISS-MODEL [70]. The models are for illustration only.

Conflicts of Interest: R.B. and E.J.H. hold equity in ioGenetics LLC, the company responsible for designing the bioinformatics models used in this project. RAH, AL, TH have all received speakers' honoraria, unrestricted research grants and/or participated in advisory boards for Biogen, Merck, Novartis, Roche and Sanofi Genzyme. SBT and BB declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

APC	antigen presenting cell
BCR	B cell receptor
CDR3	complementarity determining region 3
CNS	Central nervous system
CSO	cleavage site octamer
HLA	human leukocyte antigen
Ig	immunoglobulin
IgG	immunoglobulin G
IGHV	immunoglobulin heavy variable
mAb	monoclonal antibody
MHC	major histocompatibility complex
NN	neural network
nLCMS	nano liquid chromatography mass spectrometry
rMBP	recombinant myelin basic protein
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SVM	support vector machine

References

1. Adler, L.N.; Jiang, W.; Bhamidipati, K.; Millican, M.; Macaubas, C.; Hung, S.C.; Mellins, E.D. The Other Function: Class II-Restricted Antigen Presentation by B Cells. *Front. Immunol.* **2017**, *8*, 319. [CrossRef]
2. Hsing, L.C.; Rudensky, A.Y. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol. Rev.* **2005**, *207*, 229–241. [CrossRef] [PubMed]
3. van Kasteren, S.I.; Overkleeft, H.S. Endo-lysosomal proteases in antigen presentation. *Curr. Opin. Chem. Biol.* **2014**, *23*, 8–15. [CrossRef]
4. Shimabukuro-Vornhagen, A.; Zoghi, S.; Liebig, T.M.; Wennhold, K.; Chemitz, J.; Draube, A.; Kochanek, M.; Blaschke, F.; Pallasch, C.; Holtick, U.; et al. Inhibition of protein geranylgeranylation specifically interferes with CD40-dependent B cell activation, resulting in a reduced capacity to induce T cell immunity. *J. Immunol.* **2014**, *193*, 5294–5305. [CrossRef] [PubMed]
5. Hauser, S.L. The Charcot Lecture | beating MS: A story of B cells, with twists and turns. *Mult. Scler.* **2015**, *21*, 8–21. [CrossRef] [PubMed]

6. Mathias, A.; Perriard, G.; Canales, M.; Soneson, C.; Delorenzi, M.; Schluep, M.; Du Pasquier, R.A. Increased ex vivo antigen presentation profile of B cells in multiple sclerosis. *Mult. Scler. J.* **2016**, *23*, 802–809. [CrossRef] [PubMed]
7. Hoydahl, L.S.; Richter, L.; Frick, R.; Snir, O.; Gunnarsen, K.S.; Landsverk, O.J.B.; Iversen, R.; Jeliazkov, J.R.; Gray, J.J.; Bergseng, E.; et al. Plasma Cells Are the Most Abundant Gluten Peptide MHC-expressing Cells in Inflamed Intestinal Tissues From Patients With Celiac Disease. *Gastroenterology* **2019**, *156*, 1428–1439.e1410. [CrossRef] [PubMed]
8. Honey, K.; Rudensky, A.Y. Lysosomal cysteine proteases regulate antigen presentation. *Nat. Rev. Immunol.* **2003**, *3*, 472–482. [CrossRef]
9. Avalos, A.; Ploegh, H. Early BCR Events and Antigen Capture, Processing, and Loading on MHC Class II on B Cells. *Front. Immunol.* **2014**, *5*, 777–780. [CrossRef]
10. Weiss, S.; Bogen, B. B-lymphoma cells process and present their endogenous immunoglobulin to major histocompatibility complex-restricted T cells. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 282–286. [CrossRef] [PubMed]
11. Stoka, V.; Turk, V.; Turk, B. Lysosomal cathepsins and their regulation in aging and neurodegeneration. *Ageing Res. Rev.* **2016**, *32*, 22–37. [CrossRef]
12. Prinz, M.; Priller, J. The role of peripheral immune cells in the CNS in steady state and disease. *Nat. Neurosci.* **2017**, *20*, 136–144. [CrossRef] [PubMed]
13. Wootla, B.; Denic, A.; Rodriguez, M. Polyclonal and Monoclonal Antibodies in Clinic. In *Human Monoclonal Antibodies: Methods and Protocols*; Steinitz, M., Ed.; Humana Press: Totowa, NJ, USA, 2014; pp. 79–110.
14. Grilo, A.L.; Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* **2019**, *37*, 9–16. [CrossRef] [PubMed]
15. Waldmann, H. Human monoclonal antibodies: The residual challenge of antibody immunogenicity. *Methods Mol. Biol.* **2014**, *1060*, 1–8. [CrossRef] [PubMed]
16. Jefferis, R.; Lefranc, M.-P. Human immunoglobulin allotypes: Possible implications for immunogenicity. *mAbs* **2009**, *1*, 332–338. [CrossRef] [PubMed]
17. De Groot, A.S.; Scott, D.W. Immunogenicity of protein therapeutics. *Trends Immunol.* **2007**, *28*, 482–490. [CrossRef]
18. Andersen, T.K.; Huszthy, P.C.; Gopalakrishnan, R.P.; Jacobsen, J.T.; Fauskanger, M.; Tveita, A.A.; Grødeland, G.; Bogen, B. Enhanced germinal center reaction by targeting vaccine antigen to major histocompatibility complex class II molecules. *npj Vaccines* **2019**, *4*, 9. [CrossRef] [PubMed]
19. Jacobsen, J.; Haabeth, O.-A.W.; Tveita, A.A.; Schjetne, K.W.; Munthe, L.A.; Bogen, B. Naive Idiotope-Specific B and T Cells Collaborate Efficiently in the Absence of Dendritic Cells. *J. Immunol.* **2014**, *192*, 4174–4183. [CrossRef] [PubMed]
20. De Groot, A.S.; Martin, W. Reducing risk, improving outcomes: Bioengineering less immunogenic protein therapeutics. *Clin. Immunol.* **2009**, *131*, 189–201. [CrossRef]
21. Su, Y.; Carey, G.; Marić, M.; Scott, D.W. B Cells Induce Tolerance by Presenting Endogenous Peptide-IgG on MHC Class II Molecules via an IFN-γ-Inducible Lysosomal Thiol Reductase-Dependent Pathway. *J. Immunol.* **2008**, *181*, 1153–1160. [CrossRef]
22. Hastings, K.T.; Cresswell, P. Disulfide reduction in the endocytic pathway: Immunological functions of gamma-interferon-inducible lysosomal thiol reductase. *Antioxid Redox Signal.* **2011**, *15*, 657–668. [CrossRef] [PubMed]
23. Santoro, L.; Reboul, A.; Kerblat, I.; Drouet, C.; Colomb, M.G. Monoclonal IgG as antigens: Reduction is an early intracellular event of their processing by antigen-presenting cells. *Int. Immunol.* **1996**, *8*, 211–219. [CrossRef]
24. Fehr, K.; LoSpalluto, J.; Ziff, M. Degradation of immunoglobulin G by lysosomal acid proteases. *J. Immunol.* **1970**, *105*, 973–983.
25. Driessen, C.; Lennon-Dumenil, A.M.; Ploegh, H.L. Individual cathepsins degrade immune complexes internalized by antigen-presenting cells via Fcgamma receptors. *Eur. J. Immunol.* **2001**, *31*, 1592–1601. [CrossRef]
26. Hoglund, R.A.; Lossius, A.; Johansen, J.N.; Homan, J.; Benth, J.S.; Robins, H.; Bogen, B.; Bremel, R.D.; Holmoy, T. In Silico Prediction Analysis of Idiotope-Driven T-B Cell Collaboration in Multiple Sclerosis. *Front. Immunol.* **2017**, *8*, 1255. [CrossRef] [PubMed]

27. Holmoy, T.; Vartdal, F.; Hestvik, A.L.; Munthe, L.; Bogen, B. The idiotype connection: Linking infection and multiple sclerosis. *Trends Immunol.* **2010**, *31*, 56–62. [CrossRef] [PubMed]
28. Wu, C.; Orozco, C.; Boyer, J.; Leglise, M.; Goodale, J.; Batalov, S.; Hodge, C.L.; Haase, J.; Janes, J.; Huss, J.W.; et al. BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **2009**, *10*, R130. [CrossRef] [PubMed]
29. Rock, R.B.; Hu, S.; Deshpande, A.; Munir, S.; May, B.J.; Baker, C.A.; Peterson, P.K.; Kapur, V. Transcriptional response of human microglial cells to interferon-gamma. *Genes Immun.* **2005**, *6*, 712–719. [CrossRef]
30. Bogen, B.; Weiss, S. Processing and presentation of idiotypes to MHC-restricted T cells. *Int. Rev. Immunol.* **1993**, *10*, 337–355. [CrossRef]
31. Weiss, S.; Bogen, B. MHC class II-restricted presentation of intracellular antigen. *Cell* **1991**, *64*, 767–776. [CrossRef]
32. Khodadoust, M.S.; Olsson, N.; Chen, B.; Sworder, B.; Shree, T.; Liu, C.L.; Zhang, L.; Czerwinski, D.K.; Davis, M.M.; Levy, R.; et al. B-cell lymphomas present immunoglobulin neoantigens. *Blood* **2019**, *133*, 878–881. [CrossRef] [PubMed]
33. Khodadoust, M.S.; Olsson, N.; Wagar, L.E.; Haabeth, O.A.W.; Chen, B.; Swaminathan, K.; Rawson, K.; Liu, C.L.; Steiner, D.; Lund, P.; et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* **2017**, *543*, 723–727. [CrossRef] [PubMed]
34. Reth, M. Antigen Receptors on B Lymphocytes. *Ann. Rev. Immunol.* **1992**, *10*, 97–121. [CrossRef] [PubMed]
35. Bremel, R.D.; Homan, E.J. Recognition of higher order patterns in proteins: Immunologic kernels. *PLoS ONE* **2013**, *8*, e70115. [CrossRef] [PubMed]
36. Biniossek, M.L.; Nagler, D.K.; Becker-Pauly, C.; Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J. Proteome Res.* **2011**, *10*, 5363–5373. [CrossRef] [PubMed]
37. Mort, J.S. Chapter 406—Cathepsin B. In *Handbook of Proteolytic Enzymes (Third Edition)*; Rawlings, N.D., Salvesen, G., Eds.; Academic Press, Elsevier: London, UK, 2013; pp. 1784–1791.
38. Han, J.C.; Han, G.Y. A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal. Biochem.* **1994**, *220*, 5–10. [CrossRef] [PubMed]
39. Kirschke, H. Chapter 413—Cathepsin S. In *Handbook of Proteolytic Enzymes (Third Edition)*; Rawlings, N.D., Salvesen, G., Eds.; Academic Press, Elsevier: London, UK, 2013; pp. 1824–1830.
40. Kirschke, H. Chapter 410—Cathepsin L. In *Handbook of Proteolytic Enzymes (Third Edition)*; Rawlings, N.D., Salvesen, G., Eds.; Academic Press, Elsevier: London, UK, 2013; pp. 1808–1817.
41. Yates, R.M.; Hermetter, A.; Taylor, G.A.; Russell, D.G. Macrophage Activation Downregulates the Degradative Capacity of the Phagosome. *Traffic* **2007**, *8*, 241–250. [CrossRef] [PubMed]
42. Bremel, R.D.; Homan, E.J. Frequency Patterns of T-Cell Exposed Amino Acid Motifs in Immunoglobulin Heavy Chain Peptides Presented by MHCs. *Front Immunol.* **2014**, *5*, 541. [CrossRef]
43. Alamyar, E.; Duroux, P.; Lefranc, M.-P.; Giudicelli, V. IMGT® Tools for the Nucleotide Analysis of Immunoglobulin (IG) and T Cell Receptor (TR) V-(D)-J Repertoires, Polymorphisms, and IG Mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. In *Immunogenetics: Methods and Applications in Clinical Practice*; Christiansen, F.T., Tait, B.D., Eds.; Humana Press: Totowa, NJ, USA, 2012; pp. 569–604.
44. Lowry, J.R.; Klegeris, A. Emerging roles of microglial cathepsins in neurodegenerative disease. *Brain Res. Bull.* **2018**, *139*, 144–156. [CrossRef]
45. Beck, H.; Schwarz, G.; Schroter, C.J.; Deeg, M.; Baier, D.; Stevanovic, S.; Weber, E.; Driessen, C.; Kalbacher, H. Cathepsin S and an asparagine-specific endoprotease dominate the proteolytic processing of human myelin basic protein in vitro. *Eur. J. Immunol.* **2001**, *31*, 3726–3736. [CrossRef]
46. Bielekova, B.; Sung, M.-H.; Kadom, N.; Simon, R.; McFarland, H.; Martin, R. Expansion and Functional Relevance of High-Avidity Myelin-Specific $CD4^+$ T Cells in Multiple Sclerosis. *J. Immunol.* **2004**, *172*, 3893. [CrossRef]
47. Tiller, K.E.; Tessier, P.M. Advances in Antibody Design. *Annu. Rev. Biomed. Eng.* **2015**, *17*, 191–216. [CrossRef]
48. Baker, M.P.; Jones, T.D. Identification and removal of immunogenicity in therapeutic proteins. *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 219–227.

49. Jawa, V.; Cousens, L.P.; Awwad, M.; Wakshull, E.; Kropshofer, H.; De Groot, A.S. T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. *Clin. Immunol.* **2013**, *149*, 534–555. [CrossRef] [PubMed]
50. Flobakk, M.; Rasmussen, I.B.; Lunde, E.; Frigstad, T.; Berntzen, G.; Michaelsen, T.E.; Bogen, B.; Sandlie, I. Processing of an Antigenic Sequence from IgG Constant Domains for Presentation by MHC Class II. *J. Immunol.* **2008**, *181*, 7062. [CrossRef] [PubMed]
51. Bogen, B.; Ruffini, P. Review: To what extent are T cells tolerant to immunoglobulin variable regions? *Scand J. Immunol.* **2009**, *70*, 526–530. [CrossRef] [PubMed]
52. Munthe, L.A.; Corthay, A.; Os, A.; Zangani, M.; Bogen, B. Systemic autoimmune disease caused by autoreactive B cells that receive chronic help from Ig V region-specific T cells. *J. Immunol.* **2005**, *175*, 2391–2400. [CrossRef] [PubMed]
53. Munthe, L.A.; Os, A.; Zangani, M.; Bogen, B. MHC-restricted Ig V region-driven T-B lymphocyte collaboration: B cell receptor ligation facilitates switch to IgG production. *J. Immunol.* **2004**, *172*, 7476–7484. [CrossRef] [PubMed]
54. Kim, A.; Hartman, I.Z.; Poore, B.; Boronina, T.; Cole, R.N.; Song, N.; Ciudad, M.T.; Caspi, R.R.; Jaraquemada, D.; Sadegh-Nasseri, S. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat. Commun.* **2014**, *5*, 5369. [CrossRef]
55. Vita, R.; Mahajan, S.; Overton, J.A.; Dhanda, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343. [CrossRef]
56. Collado, J.A.; Alvarez, I.; Ciudad, M.T.; Espinosa, G.; Canals, F.; Pujol-Borrell, R.; Carrascal, M.; Abian, J.; Jaraquemada, D. Composition of the HLA-DR-associated human thymus peptidome. *Eur. J. Immunol.* **2013**, *43*, 2273–2282. [CrossRef]
57. Sorde, L.; Spindeldreher, S.; Palmer, E.; Karle, A. Tregitopes and impaired antigen presentation: Drivers of the immunomodulatory effects of IVIg? *Immun. Inflamm. Dis.* **2017**, *5*, 400–415. [CrossRef]
58. Heyder, T.; Kohler, M.; Tarasova, N.K.; Haag, S.; Rutishauser, D.; Rivera, N.V.; Sandin, C.; Mia, S.; Malmström, V.; Wheelock, Å.M.; et al. Approach for Identifying Human Leukocyte Antigen (HLA)-DR Bound Peptides from Scarce Clinical Samples. *Mol. Cell Proteom.* **2016**, *15*, 3017–3029. [CrossRef]
59. Seward, R.J.; Drouin, E.E.; Steere, A.C.; Costello, C.E. Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. *Mol. Cell Proteom.* **2011**, *10*, M110.002477–M002110.002477. [CrossRef]
60. Hamze, M.; Meunier, S.; Karle, A.; Gdoura, A.; Goudet, A.; Szely, N.; Pallardy, M.; Carbonnel, F.; Spindeldreher, S.; Mariette, X.; et al. Characterization of CD4 T Cell Epitopes of Infliximab and Rituximab Identified from Healthy Donors. *Front. Immunol.* **2017**, *8*, 1–11. [CrossRef] [PubMed]
61. Dall, E.; Brandstetter, H. Structure and function of legumain in health and disease. *Biochimie* **2016**, *122*, 126–150. [CrossRef] [PubMed]
62. Manoury, B.; Hewitt, E.W.; Morrice, N.; Dando, P.M.; Barrett, A.J.; Watts, C. An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature* **1998**, *396*, 695–699. [CrossRef] [PubMed]
63. Graham, D.B.; Luo, C.; O'Connell, D.J.; Lefkovith, A.; Brown, E.M.; Yassour, M.; Varma, M.; Abelin, J.G.; Conway, K.L.; Jasso, G.J.; et al. Antigen discovery and specification of immunodominance hierarchies for MHCII-restricted epitopes. *Nat. Med.* **2018**, *24*, 1762–1772. [CrossRef] [PubMed]
64. Rawlings, N.D.; Waller, M.; Barrett, A.J.; Bateman, A. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **2014**, *42*, D503–D509. [CrossRef]
65. Bremel, R.D.; Homan, E.J. An integrated approach to epitope analysis II: A system for proteomic-scale prediction of immunological characteristics. *Immunome Res.* **2010**, *6*, 8. [CrossRef] [PubMed]
66. Bremel, R.D.; Homan, E.J. An integrated approach to epitope analysis I: Dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. *Immunome Res.* **2010**, *6*, 7. [CrossRef] [PubMed]
67. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367. [CrossRef]
68. Consortium, T.U. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2018**, *47*, D506–D515. [CrossRef] [PubMed]

69. Adams, C.W.C.F.A., (Mountain View, CA, 94043, US), Chan, Andrew C. (1201 Cloud Avenue, Menlo Park, CA, 94025, US), Crowley, Craig W. (151 Durazno Way, Portola Valley, CA, 94028, US), Lowman, Henry B. (400 San Juan Avenue, P.O. Box 2556 El Granada, CA, 94018, US), Nakamura, Gerald R. (1529 Portola Drive, San Francisco, CA, 94127, US), Presta, Leonard G. (1900 Gough Street, #206 San Francisco, CA, 94109, US) IMMUNOGLOBULIN VARIANTS AND USES THEREOF. 2004. Available online: https://worldwide.espacenet.com/publicationDetails/originalDocument?CC=WO&NR=2004056312A2&KC=A2&FT=D&ND=&date=20040708&DB=&locale=# (accessed on 15 November 2018).
70. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Supplementary materials

Human cysteine cathepsins degrade immunoglobulin G *in vitro* in a predictable manner

Rune A. Høglund[1,2,3†], Silje Bøen Torsetnes[1,2†], Andreas Lossius[1,2,4], Bjarne Bogen[4], E. Jane Homan[5], Robert Bremel[5], Trygve Holmøy[1,3]

[1]Department of Neurology, Akershus University Hospital, Norway

[2]Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital and University of Oslo, Lørenskog, Norway

[3]Institute of Clinical Medicine, University of Oslo, Norway

[4]Department of Immunology and Transfusion Medicine, Faculty of Medicine, University of Oslo, Norway

[5]ioGenetics LLC, Madison, Wisconsin, USA

†These authors contributed equally

Table S1 Included immunoglobulins for the study

general identifier[1]	monoclonal	Isotype	Chain	Allotype	V family[2]	Full sequence (VH-CH)	Length
7602	infliximab	IgG1k	L†	Km3	5	DILLTQSPAILSVSPGERVSFSCRASQFVGSSIHWYQQRTNGSPRLLIKYASESMSGIPSRFSGSGSGTDFTLSINTVESEDIADYYCQQSHSWPFT FGSGTNLEVKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVY ACEVTHQGLSSPVTKSFNRGEC	214
			H†	G1m17	6/3	EVKLEESGGGLVQPGGSMKLSCVASGFIFSNHWMNWVRQSPEKGLEWVAEIRSKSINSATHYAESVKGRFTISRDDSKSAVYLQMTDLRTEDTG VYYCSRNYYGSTYDYWGQGTTLTVSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVV TVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDG VEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDI AVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	449
7609	rituximab	IgG1k	L	Km3	4	QIVLSQSPAILSASPGEKVTMTCRASSSVSYIHWFQQKPGSSPKPWIYATSNLASGVPVRFSGSGSGTSYSLTISRVEAEDAATYYCQQWTSNPP TFGGGTKLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVY ACEVTHQGLSSPVTKSFNRGEC	213
			H	G1m17,1	1/1	QVQLQQPGAELVKPGASVKMSCKASGYTFTSYNMHWVKQTPGRGLEWIGAIYPGNGDTSYNQKFKGKATLTADKSSSTAYMQLSSLTSEDSAV YYCARSTYYGGDWYFNVWGAGTTVTVSAASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLS SVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFY PSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	451
8005	alemtuzumab	IgG1k	L	Km3	1	DIQMTQSPSSLSASVGDRVTITCKASQNIDKYLNWYQQKPGKAPKLLIYNTNNLQTGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCLQHISRPRT FGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVY ACEVTHQGLSSPVTKSFNRGEC	214
			H	G1m17,1	4	QVQLQESGPGLVRPSQTLSLTCTVSGFTFTDFYMNWVRQPPGRGLEWIGFIRDKAKGYTTEYNPSVKGRVTMLVDTSKNQFSLRLSSVTAADTA VYYCAREGHTAAPFDYWGQGSLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSS VVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYP SDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	451
4irz	natalizumab	IgG4k	L†	Km3	1	DIQMTQSPSSLSASVGDRVTITCKTSQDINKYMAWYQQTPGKAPRLLIHYTSALQPGIPSRFSGSGSGRDYTFTISSLQPEDIATYYCLQYDNLWTF GQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYA CEVTHQGLSSPVTKSFNR	210
			H†	-	1	QVQLVQSGAEVKKPGASVKVSCKASGFNIKDTYIHWVRQAPGQRLEWMGRIDPANGYTKYDPKFQGRVTITADTSASTAYMELSSLRSEDEAVY YCAREGYYGNYGVYAMDYWGQGTLVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSL SSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYGPPCPSCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYV DGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYP SDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK	450
2H7.v16	ocrelizumab	IgG1k	L	Km3	1	DIQMTQSPSSLSASVGDRVTITCRASSSVSYMHWYQQKPGKAPKPLIYAPSNLASGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQWSFNPP TFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKV YACEVTHQGLSSPVTKSFNRGEC	471
			H	G1	3	EVQLVESGGGLVQPGGSLRLSCAASGYTFTSYNMHWVRQAPGKGLEWVGAIYPGNGDTSYNQKFKGRFTISVDKSKNTLYLQMNSLRAEDTAV YYCARVVYYSNSYWYFDVWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSL SSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGF YPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	232
7860	adalimumab	IgG1k	L†	Km3	1	DIQMTQSPSSLSASVGDRVTITCRASQGIRNYLAWYQQKPGKAPKLLIYAASTLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYCQRYNRAPY TFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKV YACEVTHQGLSSPVTKSFNRGEC	214
			H†	G1m17,3 ,2	3	EVQLVESGGGLVQPGRSLRLSCAASGFTFDDYAMHWVRQAPGKGLEWVSAITWNSGHIDYADSVEGRFTISRDNAKNSLYLQMNSLRAEDTAV YYCAKVSYLSTASSLDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSS VVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYP SDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK	451

[1] Identifier in IMGT mAb database or patent filings

[2] Closest (mouse/)human family, determined by IMGT V-quest.

† Imputed constant region sequence based on allotype described in literature.

Figures S1 Artificial neural network perceptron topology

The perceptron used for prediction of cleavage has three layers: an input layer consisting of the vectors of the first three principal components of the amino acids in the octamer binding site; a hidden layer consisting of eight nodes with symmetry to the octamer binding site; and a single output layer, which is the cleavage prediction. A hyperbolic tangent activation function was used for all interconnections within the perceptron structure.

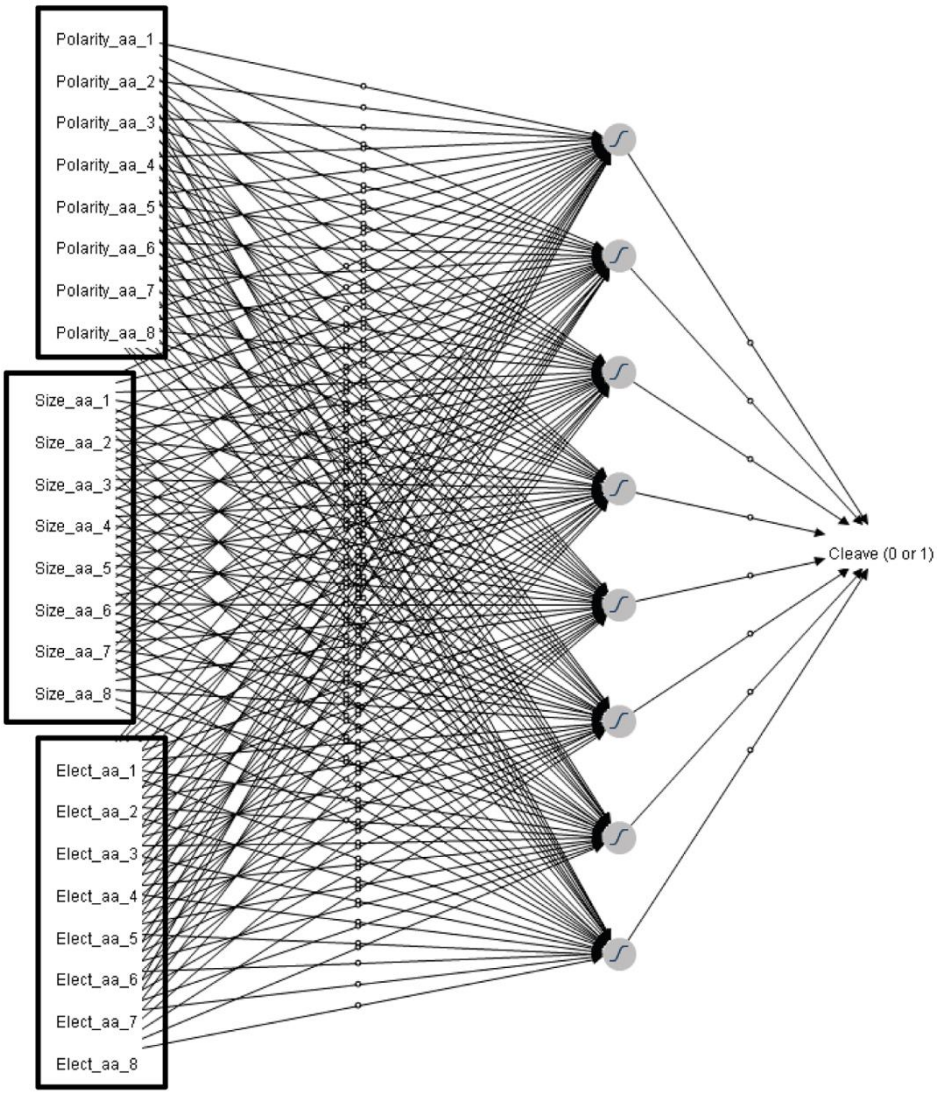
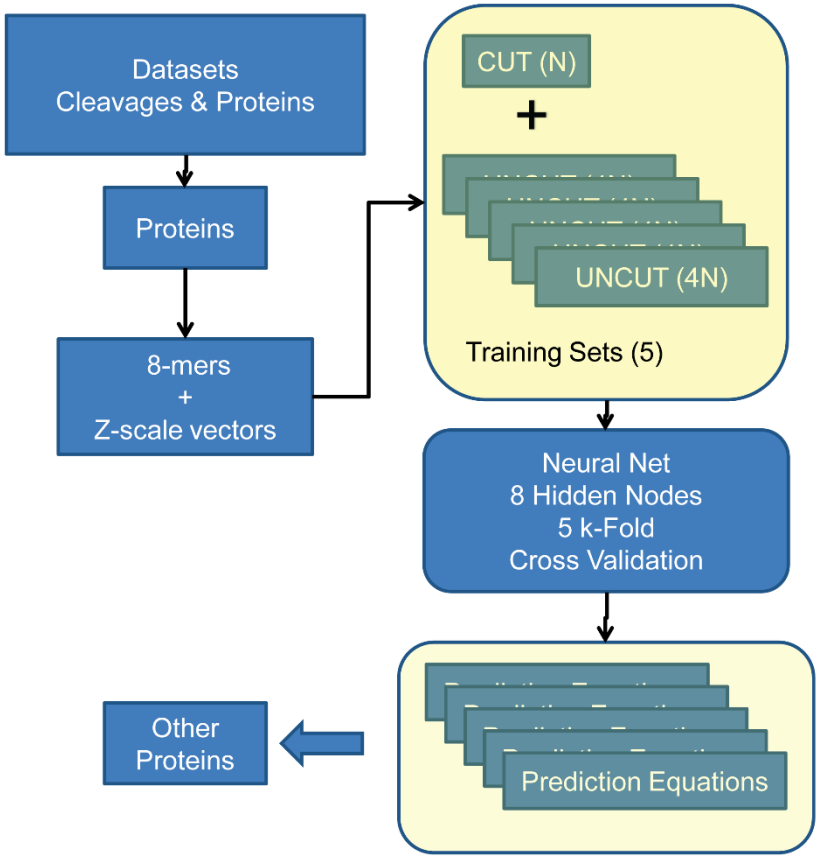


Figure S2 Data flow



Experimentation showed that a common prediction scheme for all scissile bonds was not achievable with the tools available. Thus, as an alternative a separate neural network prediction ensemble for each scissile bond dipeptide was developed. Of the 400 theoretical dipeptides there were for cathepsin (Cat)S 342, for CatB 272 and for CatL 255 dipeptides in the data set of Biniossek et al (1). The data sets of the cathepsins used had partially overlapping scissile bond preferences and each of the different cathepsins had a subset that were not cleaved. Biniossek et al (1) also indicated a preference for certain amino acids in the P2 position that is also seen in the MEROPS database (https://www.ebi.ac.uk/merops/). Consistent with that observation there were also partially overlapping P2P1 preferences between the different cathepsins. Potential scissile bonds without training sets are coded as missing values (not as zero). It was also found that the same scissile bond dipeptide occurred in many more uncleaved CSO than in those that were cleaved. Thus, a bagging process was developed where random training sets were assembled for each dipeptide and each of the training sets contained 5 times as many uncleaved CSO as cleaved CSO. These training sets were used in a 5k-Fold cross validation. The following pseudocode outlines the basic data assembly and data processing activity that were used to collate the training cleaved peptides

and their cohorts from the same protein set but that were not cleaved. This results in this article were derived with a probabilistic neural network and used a 5-kfold process cross validation process, others could be used. We have also used the process successfully with recursive partitioning and support vector machines.

Bootstrap Aggregating ("bagging") Training Set Assembly
Input:

1. Cleaved octomers derived from proteomic cleavage data sets
2. Uncleaved octomers from same proteins derived from octomer-windowing the protein sequences by single amino acid displacement of intact proteins in the proteomic cleavage data sets.

Output:

1. Training sets of matched ratios of cleaved:uncleaved octomers for each amino acid (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y) found at the P1 position of the CSO (<= 20)
2. Training sets of matched ratios of cleaved:uncleaved octomers for each amino acid (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y) found at the P1' position of the CSO (<= 20)
3. Training sets of matched ratios of cleaved:uncleaved octomers sets for all P1-P1' combinations (AA … YY) (<= 400)

Preliminary:

1. Create a dataset of singleton octomers from the cleaved and uncleaved data sets from the protein datasets downloaded from the repository.
2. Remove the cleaved octomers from the total set (the downloaded sets of intact proteins will also contain the cleaved octomers as well and they must be removed).

Repeat
For Each ***E***nsemble ***T***raining ***C***ohort (***ETC***)
Process 1:
For Each P1 anchor amino acid (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y)
For each cleaved octomer with matching P1
Select 4 non-cleaved octomers with matching P1 at random =>) ETC_P1(A … Y)
End For
End For
Process 2:
For Each P1' anchor amino acid (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y)
For each cleaved octomer with matching P1'
Select 4 non-cleaved octomers with matching P1 at random => ETC_P1'(A … Y)
End For
End For
Process 3:
For Each P1-P1' dipeptide (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y)as combinatorial pairs
For each cleaved octomer with matching P1-P1'
Select 4 non-cleaved octomers with matching P1-P1' at random => ETC_P1-P1'(AA … YY)
End For
End For
End For
Until (*ETC* == 5)

Bootstrap Aggregating ("bagging") Classifier Construction
Input:
Multiple ***ETC*** (*e.g.* 5 per (A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y) for P1 and P1')
Output:
Multiple (*e.g.* 5) ensembles from each of 5 *ETC* = 25 total classifiers for each P1, P1' and P1-P1'

Note: This process results in a total 50 discriminant equations for each potential scissile bond (25 for the P1 and 25 for the P1' side). As training is done independently concordant predictions of cleavage probability provide added confidence in the results for any particular scissile bond. The average probability of the 25 member equation ensembles is used as the prediction metric.

Repeat
Repeat with each ***ETC***
Build classifier* discriminant equations using 5-kfold cross validation each from a different random starting point
Until(N repeats == 5)

Until (N Classifiers per ***ETC*** == 5)

Classifier use for prediction of cleavage probability within protein sequences
For each sequential pair of amino acids in a protein sequence

1. Compute average probabilities using 25 equation ensembles from the P1 side of the scissile bond that match the particular amino acid pair
2. Compute average probabilities using 25 equation ensembles from the P1' side of the scissile bond that match the particular amino acid pair

Optionally (if 1 and 2 are discordant):

3. Compute average probabilities using 25 equation ensembles for the P1-P1' dipeptide pair

End For

Figure S3 Comparing neural net to support vector machine

Comparison of the performance of a probabilistic neural network (NN) and a support vector machine (SVM) as binary classifiers for predicting cleavage of human cathepsin L. The cleavage site octamers in the peptide training sets had either an alanine or a glycine at position P1 (a) and (c) Glycine at P1 Total of the cleaved trainer peptides was 222 (indicated by the blue horizontal line). Cleaved peptides were paired for training with 5 un-cleaved random cohorts with 888 peptides in each set (indicated by red horizontal line). (b) and (d) alanine at P1. Total of the cleaved trainer peptides was 111 (blue horizontal line). Cleaved peptides were paired for training with 5 un-cleaved random cohorts with 444 peptides in each set (red horizontal line).

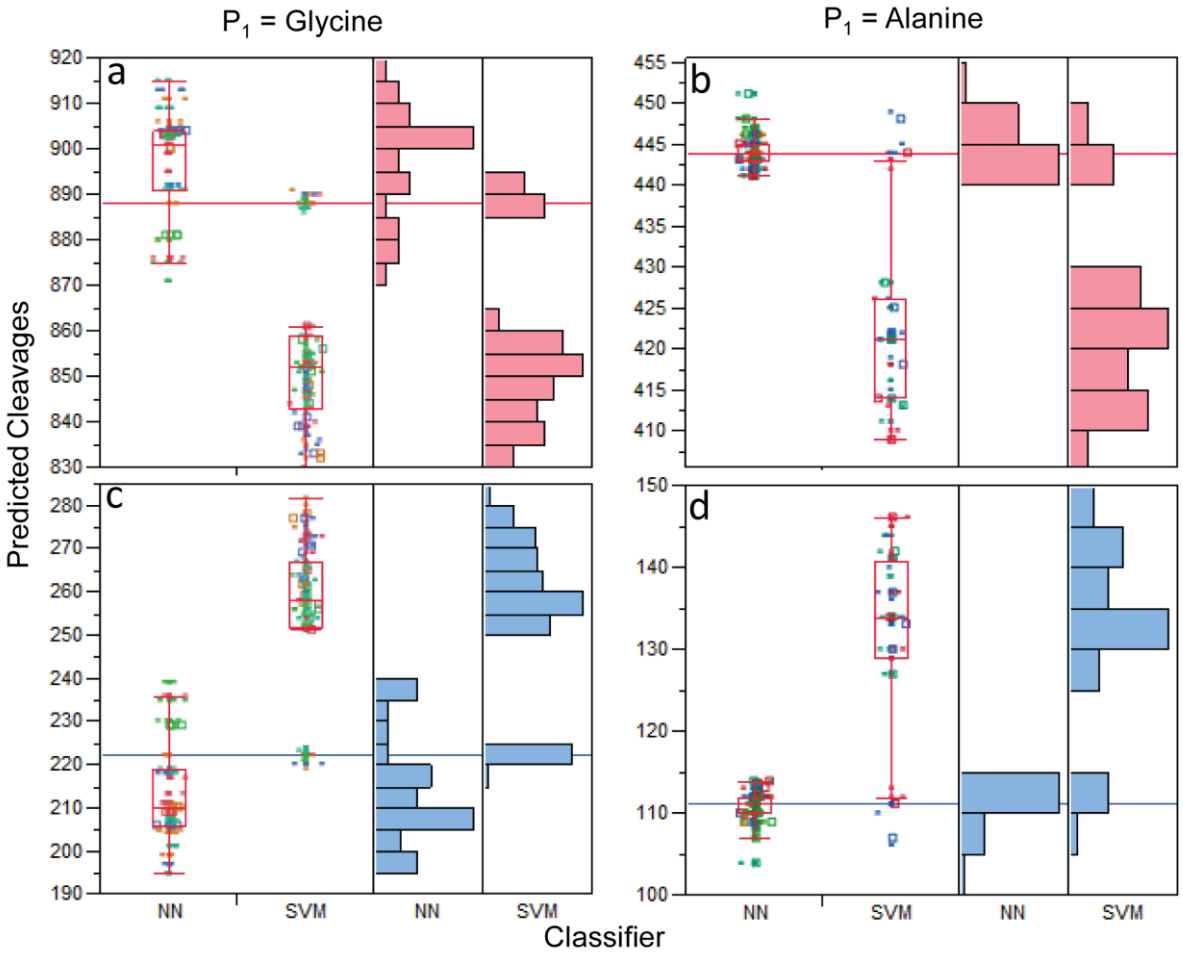


Figure S4 Hierarchical cluster of peptide occurrences by sample

Alpha-synuclein (aSyn), recombinant myelin basic protein (rMBP) isoforms 2 and 6, and tau digested by cathepsin S, L or B at pH 6 for 6, 24 or 30 hours were analyzed using LC-MS to identify resulting peptides. The occurrences of the peptides were compared using pairwise hierarchical clustering of standardized values, by method of Ward.

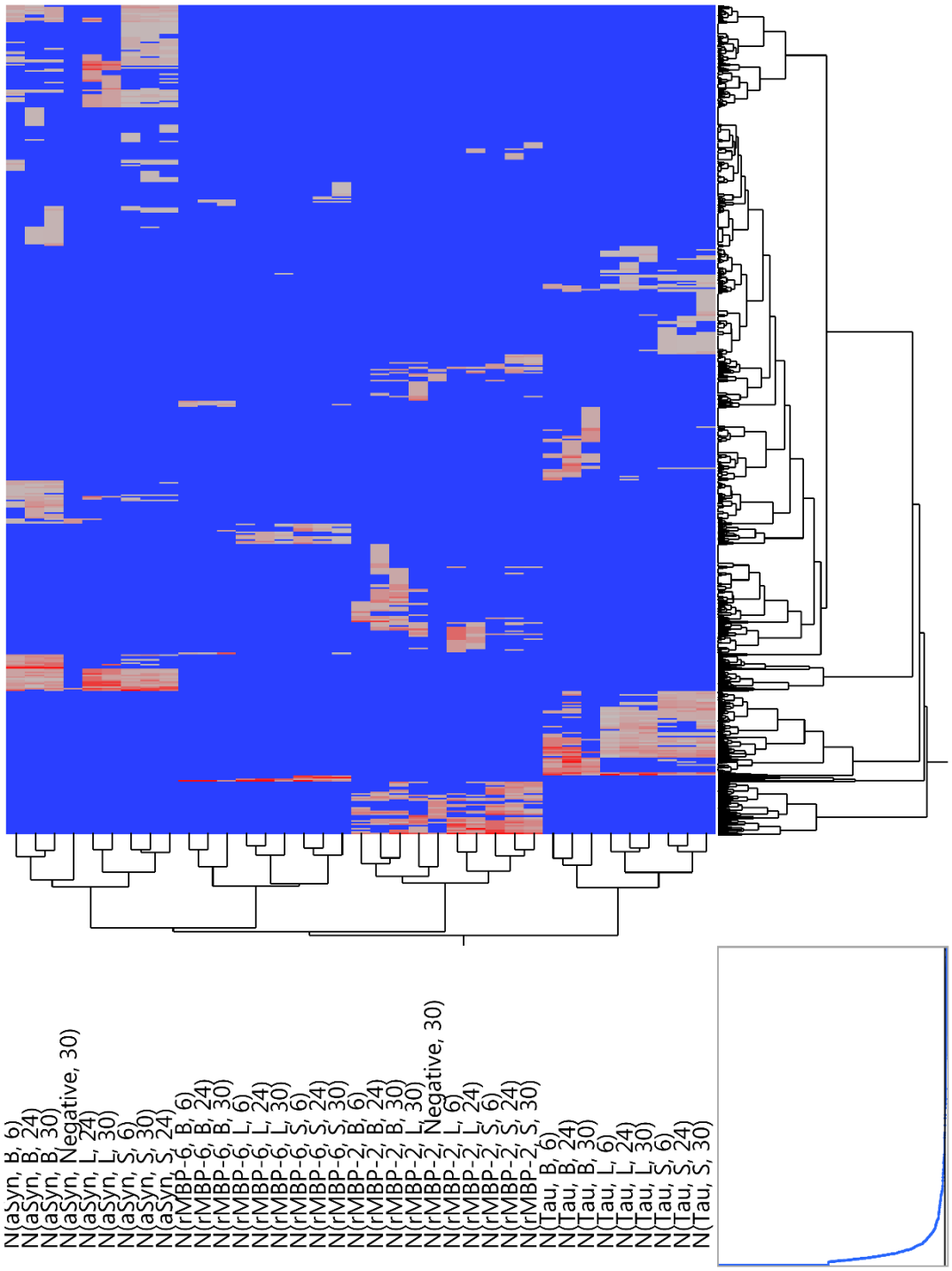


Figure S5 Possible carboxypeptidase activity for cathepsin B

The maximum distance from a high-probability cleavage site (p>0.79) intra protein (rMBP-2) was plotted (X-axis) against the observed number cleavages at the given positions. Each point indicates number of cleavages associated with a given distance, and the lines are smoothed averages.

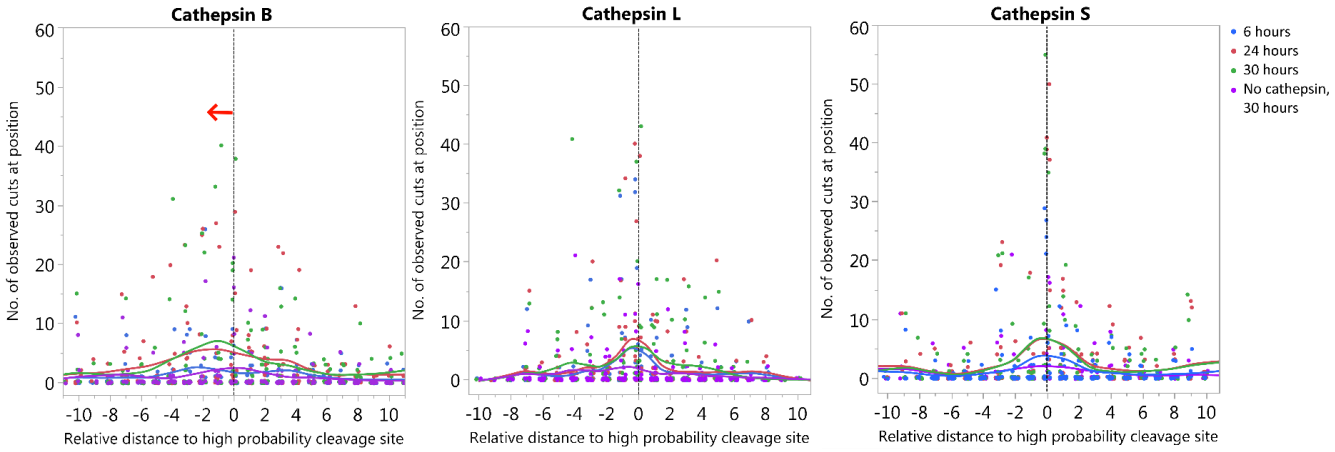


Figure S6 Evaluation of cleavage accuracy for immunoglobulin variable regions

Cleavage probability for all possible cleavage site octamers (CSOs) within A) heavy and B) light chain variable regions of rituximab, infliximab, ocrelizumab, natalizumab, alemtuzumab and adalimumab were binned into ranges of 0.2 (X-axis). Intra-chain z-standardized number of observed cuts after 24 hours at pH 6 are depicted on the Y-axis. P-values indicate Welch ANOVA significance, and differing letters indicate significant differences between groups (Tukey-Kramer, HSD).

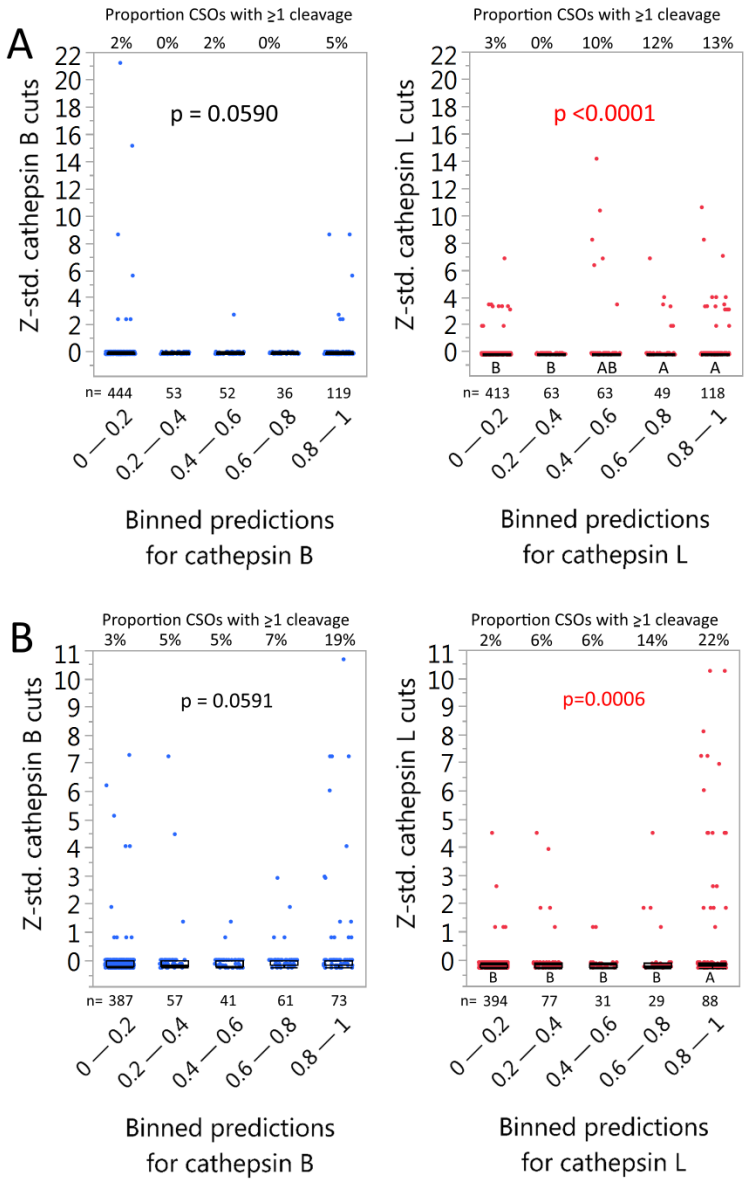
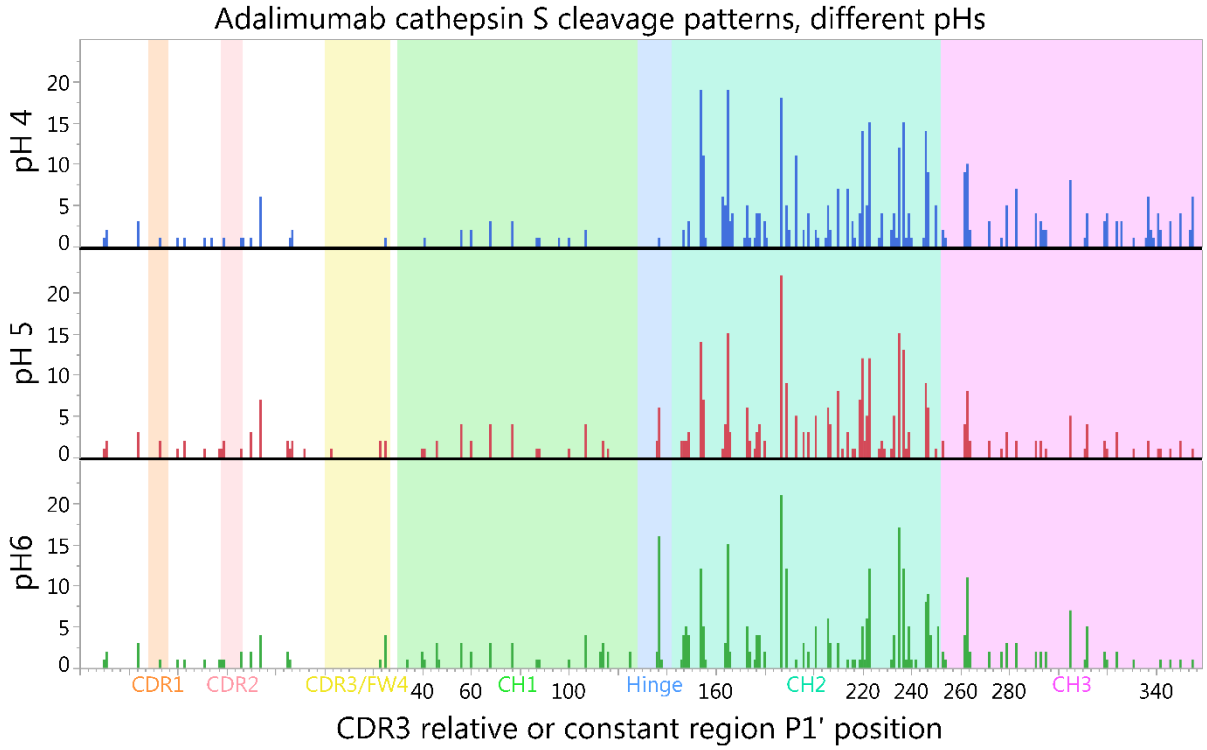


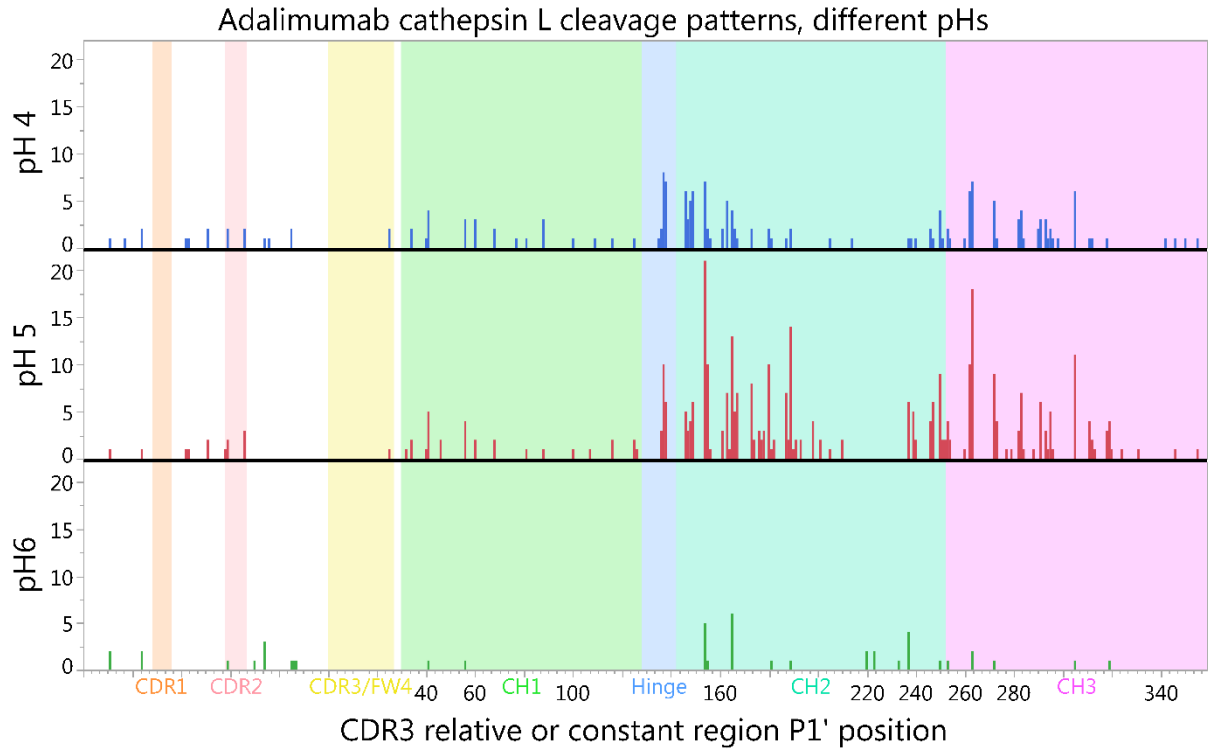
Figure S7 Adalimumab digestion pattern at different pH

Non-standardized number of observed cuts with LC/MS within the Ig light chain after incubating adalimumab with a) cathepsin S, b) cathepsin L and c) cathepsins B for 24 hours at pHs 4, 5 or 6. For alignment purposes, the position is assigned by the relative position of P1' in the cleavage site octamer to the cysteine (0) of CDR3. The constant region is aligned to start at position 30.

A)



B)


Adalimumab cathepsin L cleavage patterns, different pHs
pH 4
pH 5
pH6
0
5
10
15
20
CDR1
CDR2
CDR3/FW4
CH1
Hinge
CH2
CH3
40
60
100
160
220
240
260
280
340
CDR3 relative or constant region P1' position

C)

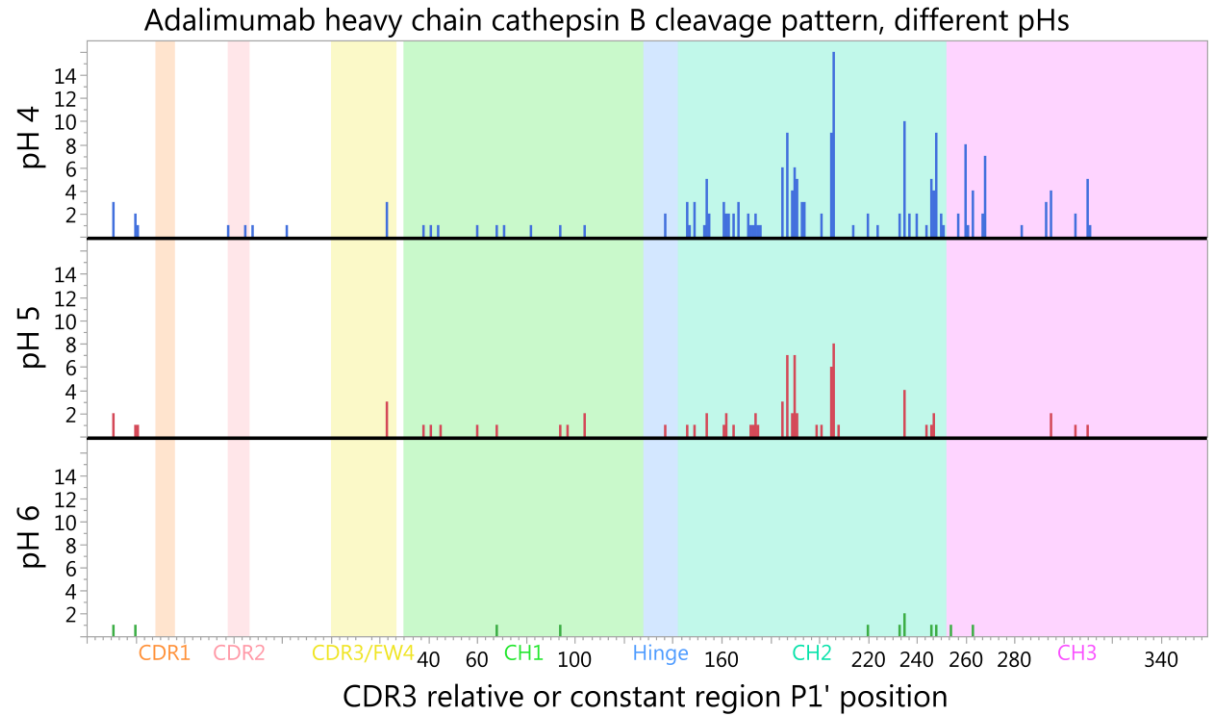

Adalimumab heavy chain cathepsin B cleavage pattern, different pHs
pH 4
pH 5
pH 6
2
4
6
8
10
12
14
CDR1
CDR2
CDR3/FW4
CH1
Hinge
CH2
CH3
40
60
100
160
220
240
260
280
340
CDR3 relative or constant region P1' position

Figure S8 Infliximab digestion by cathepsin S, L and B at pH 4, 5 and 6.

Distribution of peptide lengths after digestion of 2400 nM infliximab with either cathepsin B, L, S at 6, 24 or 30 hours at pH 4, 5 or 6. Each datapoint represents one identified peptide at the given time point. Black lines with annotations indicate the mean size of peptides. Purple and green areas indicate peptide sizes fitting HLA class I and II, respectively.

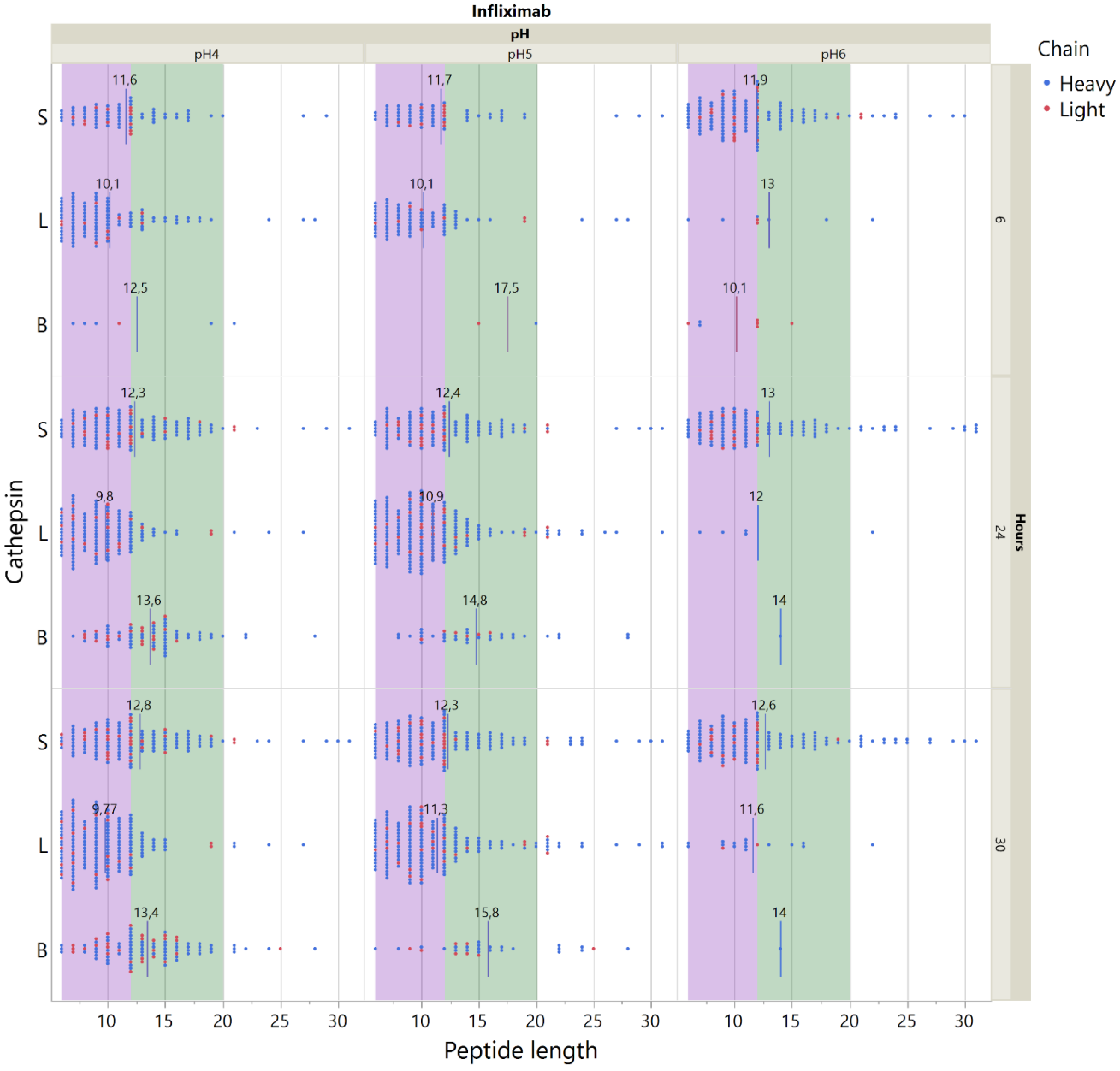
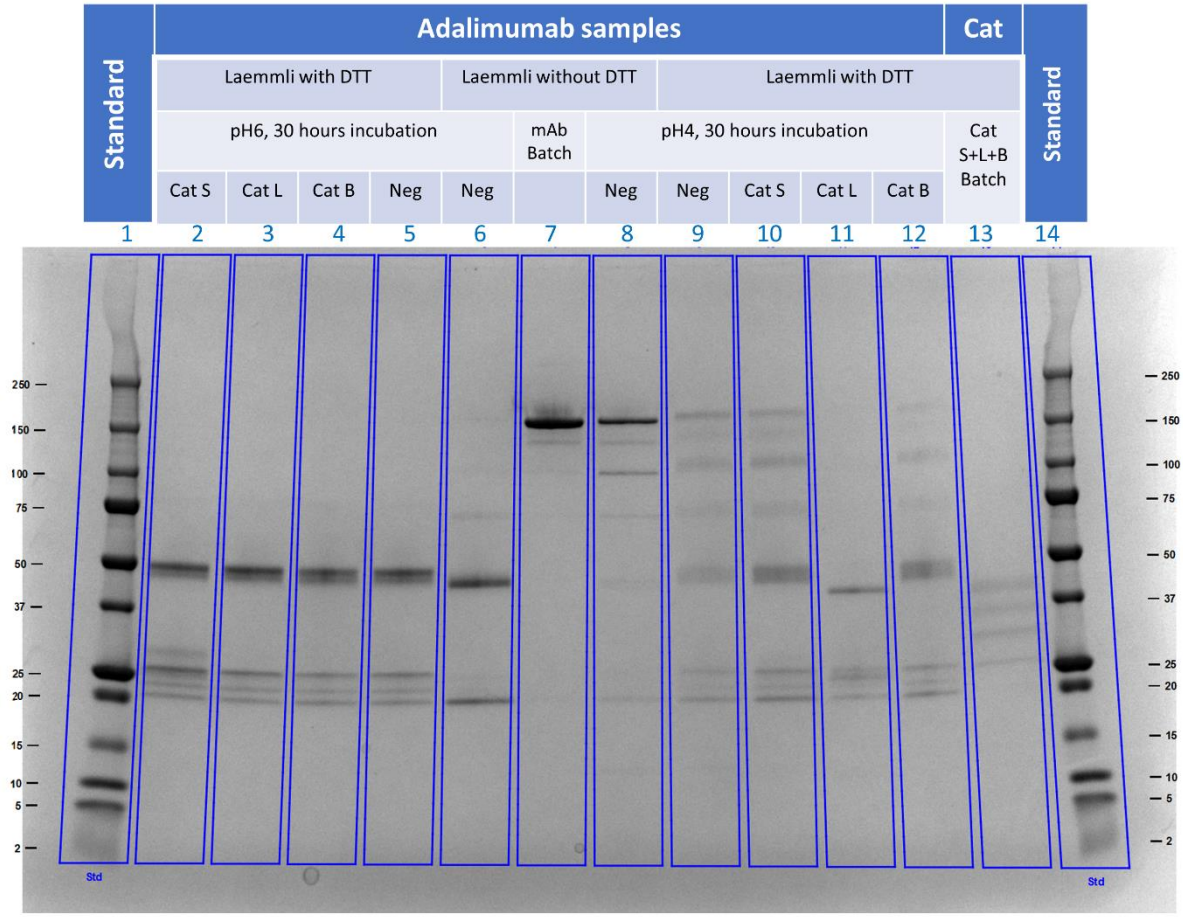


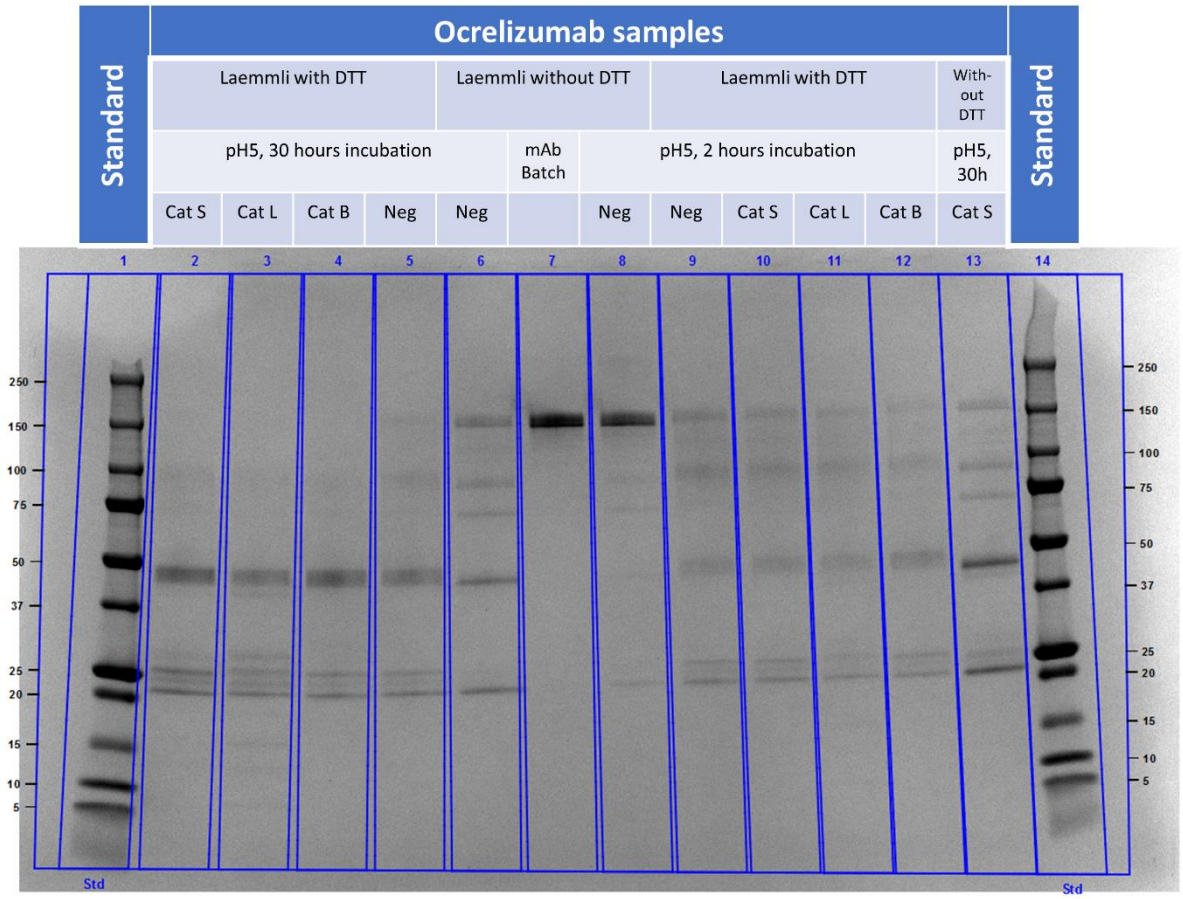
Figure S9 SDS-PAGE of digested IgGs

One µg of adalimumab (A) or ocrelizumab (B) digested by cathepsins S, L or B for 2 or 30 hours in presence 4 mM DTT of per well, were stained with Coomassie blue. Adalimumab and ocrelizumab samples incubated with 4 mM DTT alone (Neg) and unmanipulated batch samples were used as controls. Findings in lane 11 (Cathepsin L, pH 4, 30 hours) were replicated using both adalimumab and ocrelizumab samples (not shown). A weaker band with the same location was also found with cathepsin L at pH 5 (Figure S8B).

A)



B)


Standard
Ocrelizumab samples
Laemmli with DTT
Laemmli without DTT
Laemmli with DTT
With-out DTT
pH5, 30 hours incubation
mAb Batch
pH5, 2 hours incubation
pH5, 30h
Cat S
Cat L
Cat B
Neg
Neg
Neg
Neg
Cat S
Cat L
Cat B
Cat S
Standard
1 2 3 4 5 6 7 8 9 10 11 12 13 14
250 150 100 75 50 37 25 20 15 10 5
Std
Std

References

1. Biniossek ML, Nagler DK, Becker-Pauly C, Schilling O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *Journal of proteome research* (2011) **10**(12):5363-73. Epub 2011/10/05. doi: 10.1021/pr200621z. PubMed PMID: 21967108.

Paper III

III

CD4$^+$ T cells in the blood of MS patients respond to predicted epitopes from B cell receptors found in spinal fluid

Rune A. Høglund[1,2,3], Robert Bremel[4], E. Jane Homan[4], Silje Bøen Torsetnes[1,3], Andreas Lossius[1,2,5†] and Trygve Holmøy[1,2†]

[1]Department of Neurology, Akershus University Hospital, Lørenskog, Norway

[2]Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[3]Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital and University of Oslo, Lørenskog, Norway

[4]ioGenetics LLC, Madison, WI, USA

[5]Institute of Basic Medical Sciences, Department of Molecular Medicine, University of Oslo, Oslo, Norway

† These authors contributed equally

*** Correspondence:**
Rune A. Høglund
r.a.hoglund@medisin.uio.no

Keywords: B cell, T cell, multiple sclerosis, idiotope, T-B collaboration, IGHV, epitope prediction, Th17

Abstract

B cells are important pathogenic players in multiple sclerosis (MS), but their exact role is not known. We have previously demonstrated that B cells from cerebrospinal fluid (CSF) of MS patients can activate T cells that specifically recognize antigenic determinants (idiotopes) from their B cells receptors (BCRs). The aim of this study was to evaluate these findings in MS patients by combining *in silico* prediction models and *in vitro* assays to identify antigenic idiotopes of immunoglobulin heavy-chain variable (IGHV) gene repertoires.

CSF IGHV repertoires from MS patients were sequenced and translated. To guide selection of potential antigenic idiotopes, we utilized *in silico* predicted HLA-DR affinity, endosomal processing, as well as transcript frequency from nine MS patients. Idiotopes with predicted low affinity and low likelihood of cathepsins cleavage were inert controls. Peripheral blood mononuclear cells from these patients were stimulated with the selected idiotope peptides in presence of anti-CD40 for 12 hours. T cells were then labeled for activation status with anti-CD154 antibodies and CD3$^+$CD4$^+$ T cells phenotyped as memory (CD45RO$^+$) or naïve (CD45RO$^-$), with potential for brain migration (CXCR3 and/or CCR6 expression). Anti-CD14 and -CD8 were utilized to exclude monocytes and CD8$^+$ T

cells. Unstimulated cells or insulin peptides were negative controls, and EBNA-1 peptides or CD3/CD28 beads were positive controls.

CD4$^+$ T cells from all nine MS patients were activated by idiotope peptides with predicted high HLA-DR affinity and high likelihood of cathepsin cleavage, whereas no robust responses were seen towards predicted inert peptides. Responses were mainly observed towards peptides affiliated with the CDR3 region. Activated memory CD4$^+$ T cells expressed the chemokine receptor CCR6, affiliated with a Th17 phenotype and allowing passage into the central nervous system.

This *in vitro* study suggests that MS patients have a memory T cell repertoire capable of recognizing frequent BCR idiotopes found in endogenous CSF, and that these T cells express chemokine receptors allowing them to reach the CSF B cells presenting these idiotopes. It further indicates that antigenic properties of BCR idiotopes can be identified *in silico* using HLA affinity and endosomal processing predictions.

1 Introduction

Multiple sclerosis (MS) is a chronic, inflammatory disease, likely initiated or sustained by the adaptive immune system (1). B cells have recently been attributed a particularly important role, as removing these from circulation efficiently dampens inflammation within the central nervous system (CNS) (2-4). The exact role for the B cells is still unclear and could involve antigen presentation, antibody production or cytokine secretion (5). The memory subset of B cells seems to be particularly relevant, as these are targets for depletion, reduction or inhibition by several effective MS therapeutic agents (6). The fact that central B cell tolerance mechanisms remain functional in MS, in contrast to type 1 diabetes or rheumatoid arthritis, also argues for a particular role for memory B cells (7).

We and others have previously proposed that memory B cells, due to accumulated variation within the immunoglobulin variable regions, may have B cell receptors (BCRs) that themselves could contain T cell antigens capable of triggering autoimmune diseases (8-12). Proof-of-concept studies demonstrated that MS patients have T cells specific for antigenic determinants within the variable regions of the immunoglobulins (idiotopes), and that these may induce inflammatory responses and proliferation (13-16). Such mechanisms allow B cells of various specificities to receive help from T cells specific for an unlinked antigen, an idiotope, and has been shown in mouse models to induce immunoglobulin class switching and cause production of auto-antibodies triggering auto-immune disease (12, 17, 18). An analogous immune response is the generation of anti-drug antibodies to therapeutic monoclonal antibodies (mAbs), where T cell epitopes were mapped to the variable regions (19-22). Variable region of mAbs may be chimeric or only partly humanized, providing additional potentially immunogenic idiotopes.

Immunosequencing technology has progressed, and the sheer magnitude of potential idiotopes to assess in patients is impossible to perform *in vitro*. We reasoned there are key steps necessary for idiotope-driven T-B collaboration to occur, including successful endolysosomal processing of the BCR, sufficient affinity for human leukocyte antigen (HLA) class II molecules, and lack of T cell tolerance. *In silico* models based on these assumptions suggest that nearly half of cerebrospinal fluid (CSF) BCR variable regions from MS patients harbor potential antigenic idiotopes (9). These models

included prediction of HLA-DR affinities (23, 24), likelihood of endosomal processing by cysteine cathepsins (25, 26) and modelling of tolerance likelihood based on T cell exposed motifs (TCEM) (9, 27). It has previously been suggested that frequently occurring TCEM in variable regions (i.e. germline framework motifs) could be tolerogenic, while rare motifs (i.e. complementarity determining region [CDR]3 or motifs resulting from mutations) potentially could be stimulatory to T cells (10, 27). Thymocytes could be exposed to frequent immunoglobulin heavy chain variable (IGHV) TCEM in the thymus by thymic B cells (28), or by dendritic cells sampling serum immunoglobulins (29, 30).

As MS is a chronic, inflammatory disease of the CNS, we expected that relevant T cells have a memory phenotype with capacity to migrate into the CNS. Here we present *in silico* and *in vitro* evidence for presence of such CD4$^+$ memory T cells in blood of nine MS patients, responding to idiotope-peptides from BCR IGHV regions present in endogenous CSF.

2 Methods

2.1 Patients

In this study, we investigated patient materials collected from nine relapsing-remitting MS (RRMS) patients from whom we have immunosequenced the CSF IGVH repertoire (9), and from whom we had collected peripheral blood mononuclear cells (PBMC) in parallel with the CSF cells. Demographic and disease characteristics are described in Supplementary Table 1. The nine patients had on average 1079 (SD=1213) translated IGHV sequences, which comprised 30-45 amino acids covering part of the framework region 3 (FW3), the entire CDR3 and part of FW4 (http://doi.org/10.6084/m9.figshare.5035703). All participants provided written informed consent before participating. The study was approved by the Regional Committee for Research Ethics in South-Eastern Norwegian Health Authority (REK Sør-Øst S-04143a).

2.2 In silico parameters for predicting antigenic properties of IGVH idiotopes

We utilized the previously assembled CSF IGHV amino acid sequence dataset (9, 31) as a source of idiotopes. The sequences were private to the individual, and every IGHV sequence was split into all possible 15-mers. To pre-select potential antigenic or inert idiotopes, we utilized three *in silico* parameters: Firstly, the Johnson SI normalized neural net predicted HLA class II affinity was determined (as described in detail previously (9, 23, 24)). Secondly, the likelihood of endosomal processing by key cathepsins S, L or B (9, 25, 26), was evaluated by using "fuzzy logic" as described previously (9). This method accounts for how HLA class II molecules bind peptides of varying lengths, by allowing predicted cleavage outside the bounds of the core 15-mer and lowering predicted cleavage probability cut-offs to increase sensitivity. Thirdly, the occurrence of T cell exposed motifs (TCEM) IIa or IIb within the potential idiotopes was determined (9, 27). TCEM IIa/b are non-linear pentamers within 15-mer peptides (Supplementary Figure 1), and frequency was calculated based on occurrence in two different datasets (27, 32) to account for intra-dataset bias and include TCEM from the full IGHV region. Peptides predicted to have T cell antigenic properties were those with the highest predicted HLA-DR affinities as well as predicted cleavage with any of the cathepsins selected

among the patient's own IGHV sequences. The predicted inert idiotope peptides had opposite attributes, also including HLA-DQ and HLA-DP predictions, selected from the full IGHV dataset. Predicted antigenic idiotopes were further split by the frequency class (FC) of TCEM, where those with rare TCEM (high FC, motifs occurring less than once in every 131,072 (2^{17}) IGHV or rarer) were believed to have the highest potential of generating a stimulatory response, and lower FC (TCEM occurring in every 128 IGHV or more frequently) implicated a higher likelihood of T cell tolerance. From idiotopes fulfilling these criteria, the ones from the most abundant transcripts in CSF were chosen. Duplicates among the top candidates were not included. In cases without enough sequences fulfilling the criteria, the affinity limit was adjusted until enough sequences could be included. The formal criteria are given in Table 1.

2.3 Idiotope peptides

The selected idiotope peptides (Supplementary Table 2), were synthesized by Mimotopes (Australia) to a minimum purity of 70%, with an average purity of 90% supplied. Aliquots of 0.1 mg were dissolved separately immediately prior to use in T cell activation assays with either Milli-Q water, 0.1 % acetic acid, 0.1% ammonia, 20 % acetonitrile or 8 % dimethyl formamide to a batch concentration of 800 µM idiotope peptide. The first solvent tried was always Milli-Q water, but if the peptide was not fully solved then either acetic acid or ammonia was added dependent on the predicted chemical property of each peptide. The last resorts to dissolve the peptides were either acetonitrile or dimethyl formamide. Dimethyl-sulfoxide (DMSO) was not used to avoid oxidative loss of cysteine, tryptophan and methionine rich-sequences, which frequently occur in the IGHV regions. Compatibility of all these solvents was verified by sustained responses to EBNA-1 peptides in assays using PBMC from healthy donors (not shown).

2.4 T cell activation assay

Activation assays were performed according to previously described and optimized protocols (33, 34), with a few modifications (Figure 1). Cryopreserved PBMC were thawed and immediately centrifuged at 400 × g and washed twice in RPMI 1640 Glutamax (Thermo Fisher Scientific, Germany) containing 10% heat inactivated human serum (BioWest, USA). Cells were resuspended to a final concentration of 2.5×10^6 cells/mL, mixed with 1 µg/mL anti-CD40 (clone G28.5, BioXcell, USA) and plated onto a 96-well U bottom plate. In each well, 500,000 PBMC were stimulated with either 10 µM idiotope peptide (Mimotopes), 1 µg/mL EBNA-1 HLA class II pool (Miltenyi Biotec, Germany), 1 µg/mL insulin peptide pool (Milteny Biotec), 80,000 Anti-CD3/CD28 beads (Thermo Fisher Scientific) or an equivalent amount of Milli-Q water. Experiments were performed in technical duplicates. All wells were thoroughly pipette-mixed and incubated for 12 hours at 37 °C and 5% CO_2.

2.5 Flow cytometry

After incubation, cells were transferred to a 96-well V bottom plate, centrifuged at 400 × g and washed with PBS (made in-house). Prior to labelling with antibodies, cells were incubated with fixable Near-IR Live/Dead kit (Thermo Fisher Scientific) to exclude dead cells. After wash, cells

were resuspended in PBS with 0.5% FBS (BioRad, Germany) and 2 mM EDTA and labelled according to the manufacturers recommendations with BV421 anti-CD154, FITC anti-CD3, PerCP-Cy5.5 anti-CD4, PE-Cy7 anti-CD45RO, PE anti-CXCR3 and APC anti-CCR6, as well as APC-H7 anti-CD14 and APC-H7 anti-CD8 for dump channel (all BD Biosciences, USA). The fluorochrome panel is described in detail in Supplementary Table 3. After labelling and washing twice, cells were analyzed using a FACS Canto II flow cytometer (BD Biosciences) with a three laser and 4-2-2 detector set-up. Compensation was performed using Ultra Comp beads (Thermo Fisher Invitrogen) according to the manufacturer's instructions. Fluorescence minus one (FMO) controls were used to determine gate borders. All reported results are means of technical duplicates. In two cases there were not enough cells to complete all samples (MS-2 and MS-7).

2.6 Statistics

To assess differences in chemokine expression in reactive vs non-reactive cells in samples with clear T cell responses to idiotope peptide, we applied a full factorial mixed model using activation status as fixed effect and patient subject as random effect. Analysis and all figures were created using JMP® 14 (SAS Institute, USA).

3 Results

3.1 Idiotope peptide panels

Demographic and disease characteristics of the nine patients included in the study are shown in Supplementary Table 1. For each patient these we selected a panel of ten predicted stimulatory and five predicted tolerogenic idiotope peptides. Additionally, a common panel of seven predicted inert idiotope peptides was utilized (Supplementary Table 2). We have previously shown that TCEM rarity value was associated with peptide locations in IGHV (9, 27). As expected, the predicted stimulatory peptides mapped mainly to CDR3 and the predicted tolerogenic peptides to FW3 (Figure 2). As inert idiotope peptides were selected to have either high or low TCEM frequency, their locations were mixed approximately 50-50.

3.2 Identification of idiotope-specific T cells

We moved on to identify specific T cell responses. We classified responses three times as high as the same individual's unstimulated control (background) as positive (Figure 3). In all assessed MS patients, we identified robust responses towards CD3/CD28 beads (mean 45% of memory CD4$^+$ T cells, range 23-81%). As expected, no patient responded towards insulin. Responses towards EBNA-1 peptides varied and were only classified as positive in 3/9 patients, indicating either that they did not have CD4$^+$ T cells specific for EBNA-1 or that the assay was incapable of detecting these.

All patients had idiotope-specific T cells towards at least one predicted antigenic idiotope peptide (Figure 3). We observed T cell responses against both predicted stimulatory and tolerogenic peptides, and only one weak response against a predicted inert peptide. Some of the most robust responses were seen in the tolerogenic peptide group. To exclude that our results were results of random

activation, we also replicated the experiment in two patients with enough cryopreserved PBMC (MS 7 and MS 11, Supplementary Figure 2), and found comparable responses across two experiments.

As described by others (35, 36) the proportion of CD45RO$^+$ memory cells among CD4$^+$ T cells varied between the MS patients (mean 50%, range 32-68% in unstimulated samples). To maximize comparability among individuals we therefore chose to use activated cells as a proportion of CD4$^+$CD45RO$^+$ memory cells for further analysis. However, the responses were not limited to the memory T cells exclusively; in some patients' naïve cells responded towards the same idiotope peptides (Figure 3), possibly indicating a lack of tolerance towards these. Examples included MS-2, MS-3, MS-6 and MS-10, where the first three all had low proportions of memory T cells. We further analyzed the ratio of percentage of activated memory: activated naïve cells (Figure 4). As expected, a higher proportion of responding CD4$^+$ T cells were observed in the memory compartment than in the naïve compartment for nearly all antigenic idiotope peptides with observed T cell responses, with the notable exceptions of MS-6 and MS-10.

In order to further characterize the idiotope peptides that elicited CD4$^+$ T cell responses, we labelled them using metadata from the IGHV they were derived from and information on cathepsins cleavages prediction (Figure 5). Interestingly, the 24 of 26 idiotope peptides that generated T cell responses were derived from cathepsins S or B cuts, but not from cathepsin L alone. In contrast to cathepsin L, cathepsins S and B are typically expressed in B cells (37). In addition, the idiotope peptides generating responses *in vitro* were most frequently found near the CDR3, regardless of being predicted stimulatory or tolerogenic *in silico.*

3.3 Idiotope-specific T cells are enriched for CCR6, but not CXCR3

Pro-inflammatory T cells may gain access to the CNS by interaction between CCR6 or CXCR3 and their respective ligands (38, 39). We therefore investigated the expression of these among idiotope-specific T cells, compared to non-specific cells within the same sample (Figure 6). We found that the idiotope-specific T cells were enriched for CCR6$^+$ cells, but not for CXCR3$^+$ cells. This is compatible with a CNS homing phenotype of these idiotope-specific T cells and is line with what has been found among CD4$^+$ memory T cells in CSF of MS patients previously (40).

3.4 Antigenic idiotope peptides carry somatic mutations

Our main hypothesis was that the combination of VDJ recombination, nucleotide insertions or deletions, and somatic mutations within the IGHV generate potentially antigenic idiotopes. Particularly the latter may cause loss of tolerance during a clonal selection process. We therefore utilized IMGT V-Quest (41) to reanalyze the IGHV sequences and identify these characteristics within the idiotope peptides (Figure 7). Not surprisingly, all except two (in MS-11) idiotope peptides carried mutations either within the 15-mer (potentially changing HLA affinity) or in immediate vicinity (potentially affecting cathepsin activities).

To check whether the somatic mutations influenced the predictions for cathepsin cleavage, HLA-DR affinity and TCEM FC, we identified corresponding germline sequences with the closest relative

IGH-VDJ genes as given by IMGT V-quest (41) and compared predicted outcomes for mutated and germline sequences. We then ran the *in silico* predictions to assess whether the mutations influenced predicted HLA-DR, probability for cathepsin S or B cleavage or rarity of TCEM of idiotope peptides. The predicted inert peptides were not selected in a patient specific manner and the patient's HLA type could not have contributed to mutation selection, they were therefore left out of this analysis. As expected, mutations were associated with rarer TCEM (higher FC). We found that approximately half of the antigenic peptides carried mutations with net positive effects on the predictions (increased affinity or cleavage likelihood). Six had mixed changes, and five had imputed negative changes for antigen presentation. Note that even the last group of peptides had HLA-DR affinities sufficiently high to be included in the initial analysis of potentially antigenic idiotope peptides, even though the mutations lowered their predicted HLA-DR affinities compared to the corresponding germline encoded sequence.

4 Discussion

We have suggested that the inflammation observed in MS can be initiated and/or maintained by idiotope-specific T cells driving an unlinked T-B collaboration within the CNS (8, 15). Here we demonstrate how neural network epitope prediction models identify potentially antigenic idiotopes from CSF B cells in nine out of nine assessed MS patients. We further show that these antigenic idiotopes are associated mainly with the CDR3, predicted cleavage with B cell specific cathepsins and with mutations in the IGHV region.

In previous work we identified idiotopes within the IGHV regions generating HLA-DR restricted responses in two MS patients (14, 15). These idiotopes were also associated with mutations and were capable of triggering T cells to destroy oligodendrocytes (13). Development of high throughput sequencing techniques allowed us to assess multiple MS-patient CSF IGHV-repertoires (31), and advances in epitope prediction models including protease cleavage probabilities allowed identification of multiple idiotopes potentially capable of HLA-DR presentation as well as release by cathepsin cleavage (9). In this study, these models could identify idiotopes generating T cell responses in all patients, but the hit rate was still relatively low given the selection criteria (19% among peptides with predicted high affinity and high probability of cleavage). As we only tested 500,000 PBMC (n=2) for each peptide, our findings nevertheless suggest a high precursor frequency of idiotope-specific CD4$^+$ T cells. This is in line with previous observations in MS patients (14-16). Thus, PBMC from 14 of 21 of MS patients responded towards autologous CSF immunoglobulin G (IgG), whereas only four responded to myelin basic protein (MBP) and five to autologous serum IgG (16). In limiting dilution assays the frequency of PBMC responding to a DRB1*1301-restricted idiotope derived from a mutated IGHV framework region 2 from a CSF B cell clone was $1:2x10^4$, while $<1:10^6$ PBMC responded to the corresponding germline-encoded peptide or a MBP peptide (MBP 85–99) suggested to be immunogenic in patients (15). This HLA-DR restriction fits with the prediction models, as DR alleles exhibited the most consistent affinity pattern for IGHV idiotopes (9). HLA-DRB1*15:01 in particular was among those with highest predicted affinity for FW3 and CDR3 derived idiotopes, providing a potential link to inherited risk associated with HLA-DR alleles

observed in MS (42). Our results are also in line with the previous observation of several antigenic idiotopes on individual CSF IgG molecules (14).

The high responder rates cause questions as to what is and is not normal in a functional, normal immune system, but our assay was not designed to answer this. We have, however, previously shown that some patients with other inflammatory diseases also have CD4$^+$ T cell responses to self IgG from CSF, although much more rarely than among MS patients (16). While some rare idiotope-specific T cells responses may be undetectable in our experimental setup, it could also indicate the model or selection criteria can be improved. For instance, by including both light and heavy chain variable regions while also considering all B cell specific cathepsins simultaneously. Additionally, inclusion of HLA-DQ and -DP could be attempted as well, as -DQ and -DP alleles may have contributed to antigen presentation in these experiments in addition to the predicted role for -DR. Lastly, the IGHV sequences obtained by immunosequencing are rather short and do not include the framework 2 region, where we have previously mapped an antigenic idiotope (15).

We have recently shown that cathepsins S and B, endosomal enzymes expressed in B cells (37), are capable of degrading IgG variable regions into peptides sized to fit HLA class II molecules (26). In B cells with BCR recognizing anti-BCR IgG it was shown that idiotopes from both were presented on major histocompatibility complex (MHC)-II in mice, indicating the BCR and its cognate antigen follow the same pathway of degradation (43). It was thus not surprising to see that 24/25 idiotope peptides with predicted high HLA DR affinity, and that also elicited memory CD4$^+$ T cell responses *in vitro*, were predicted to be released by either cathepsins S or B, but not cathepsin L alone. In the blood, the latter is expressed mainly in monocytes (44). We have also observed that B cells from the CSF of MS patients have antigen presenting phenotypes, expressing cathepsins S, H and B, as well as HLA-DR (J. Polak et al., unpublished observations). As a minimal system including cathepsins S, B and H, as well as HLA-DR and -DM was sufficient to generate a diverse HLA class II presented antigen repertoire (45), it is likely these also are key cathepsins in the idiotope-driven response.

A total of 21 of 26 idiotope peptides generating memory CD4$^+$ T cell responses were associated with the CDR3, consistent with our previous suggestions that this region may be most likely to induce such responses due to the combined events of VDJ-recombination, nucleotide insertions and deletions, and somatic hypermutation (8, 9). The memory phenotype of these cells indicated that they were antigen experienced. This supports a concept of a general dysregulated T-B cell collaboration response in MS, and is in line with current belief that antigen presentation may be a core role for B cells in MS immunopathology (46) and the current knowledge of genetic risk contribution involving antigen presentation (42). However, such a response is not necessarily specific for MS, as a seemingly random mutational activity could generate a similar response in any individual under unfortunate circumstances. For instance, it was shown that B-cell presentation of IGHV idiotopes are common occurrences in multiple B cell lymphomas (47, 48), suggesting a potential role in malignant disease. In fact, presentation of BCR idiotopes seems to be a common occurrence upon antigen stimulation in a mouse model (43), and FW3 idiotopes can be eluted from HLA class II from human thymus (49), synovial tissue of rheumatoid- and Lyme arthritis (50), bronchoalveolar lavage samples (51) and in dendritic cells loaded with IvIg (52). The phenomenon is not limited to HLA-DR but was

found to occur with HLA-DQ in EBV transformed cell lines as well (53). Further, idiotope-specific T cells have been identified in both SLE (54, 55) and rheumatoid arthritis (56). Thus, although idiotope-specific CD4$^+$ T cells seem to be enriched in MS patients compared to controls (16), idiotope-driven T-B collaboration may be a general feature of immune regulation (10, 57).

Some of our observations conflict with previous theories. For instance it was previously suggested frequently occurring TCEM in variable regions could predict tolerance toward the idiotope (9, 10), as thymocytes would be exposed in the thymus to induce either regulatory T cells (Tregs) or deletion (58). In line with this it has been shown that high concentrations of monoclonal IgG can induce central and peripheral tolerance in various mouse models (reviewed in (10)). More recently it was shown that repeated exposure to idiotopes caused induction of Foxp3$^+$ idiotope-specific T cells in mice (43). However, our observations indicate that at least some of the predicted tolerogenic idiotope peptides had escaped central tolerance mechanisms. In fact, some of the most robust responses were observed towards peptides with frequently occurring TCEM. This observation does not exclude the possibility of peripheral tolerance induction. Unfortunately, we did not have access to PBMC collected in parallel with the CSF B cells to assess idiotope-specific Tregs, as Tregs are generally very rare and may require enrichment procedures to identify (59, 60). It is possible that TCEM frequency alone is not sufficient to predict tolerance. If HLA affinity and TCEM were combined in a single variable (61), a more realistic image of what the T cells are exposed to could be made clear. It has previously been shown that only certain areas of the distinct IGHV families exhibit consistent increase in predicted HLA affinity (9, 27), and several of these seem to correspond to matching increases in probability for cleavage by cathepsins S and B (26). It is also possible that the low frequency or function of Tregs in MS patients (62, 63) contributes to increased immunogenicity of idiotopes, as capacity to suppress is lowered. More recently, it was shown that memory B cells are capable of auto-activating CXCR3$^+$CCR6$^+$ T cells in an HLA-DR dependent manner. These authors suggested the protein RASGRP2 was responsible (63), but idiotopes could be alternate candidates.

We were not able to directly study intrinsic T-B collaboration. We therefore sought to identify whether MS patients have a repertoire of memory T cells that match endogenous CSF IGHV idiotopes, implying such collaboration had occurred previously and possibly was still ongoing. Previous studies have already established the possibility of direct idiotope-specific T-B collaboration in MS (13, 14). Importantly our study lacked a control group to compare the T cell responses. A fitting control group would be patients with other neurological inflammatory disorders, as used in our previous work (9, 16). Unfortunately, we have no cryopreserved PBMC collected in parallel with CSF B cells from these patients, who have now been treated for several years with immunosuppressive and immunomodulatory drugs precluding experiments on fresh PBMC.

5 Conclusion

By combining high-throughput immunosequencing of CSF B cell repertoires with *in silico* epitope prediction models and *in vitro* activation assays, we were able to identify idiotope-specific memory T cells expressing CCR6 in nine out of nine assessed MS patients. The majority of these idiotope peptides were associated with the CDR3 region, were predicted to have high probability of cleavage

by cathepsins S or B and had mutations that influenced affinity or cleavage predictions. This supports the concept that MS patients have a circulating repertoire of memory T cells capable of invading the CNS, and that cathepsin cleavage plays a role for shaping of the idiotope-specific T cell repertoire.

6 Conflict of Interest

RDB and EJH hold equity in ioGenetics LLC, the company responsible for designing the bioinformatics models used in this project. All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

RAH contributed with experimental design, performing the experiments, analysis and interpretation of the data, and drafting the manuscript. RDB contributed by designing the bioinformatics algorithms, preparing datasets, interpreting the data and revising the manuscript. EJH contributed by interpreting the data and revising the manuscript. SBT contributed with experimental design and revised the manuscript. AL contributed by designing the experiments, interpreting the data, and writing the manuscript. TH contributed with conceptualizing the study, interpreting the data, and writing the manuscript.

8 Funding

The study was supported with a grant from the Norwegian Research Council (grant number 250864/F20), Akershus University Hospital internal strategic funds, the Fritz and Ingrid Nilsen endowment and an unrestricted research award from Biogen Norway.

9 Acknowledgments

We are grateful for the patients' sample contributions for this study. We also thank Jūratė Šaltytė Benth for advice on mixed model statistical tests.

10 Abbrievations

BCR – B cell receptor
CCR – CC chemokine receptor
CD – cluster of differentiation
CDR – complementarity determining region
CNS – central nervous system
CSF – cerebrospinal fluid
CXCR – CXC chemokine receptor
EBNA – Epstein Barr nuclear antigen
FC – frequency class
FW – framework region
HLA – human leukocyte antigen

IgG – immunoglobulin G
IGHV – immunoglobulin heavy chain variable
MBP – myelin basic protein
MHC – major histocompatibility complex
MS – multiple sclerosis
PBMC – peripheral blood mononuclear cells
TCEM – T cell exposed motif

11 References

1. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. *Nature reviews Immunology* (2015) **15**(9):545-58. Epub 2015/08/08. doi: 10.1038/nri3871. PubMed PMID: 26250739.

2. Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *The New England journal of medicine* (2008) **358**(7):676-88. Epub 2008/02/15. doi: 10.1056/NEJMoa0706383. PubMed PMID: 18272891.

3. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung HP, Hemmer B, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *The New England journal of medicine* (2017) **376**(3):221-34. Epub 2016/12/22. doi: 10.1056/NEJMoa1601277. PubMed PMID: 28002679.

4. Bar-Or A, Grove RA, Austin DJ, Tolson JM, VanMeter SA, Lewis EW, et al. Subcutaneous ofatumumab in patients with relapsing-remitting multiple sclerosis. *The MIRROR study* (2018) **90**(20):e1805-e14. doi: 10.1212/wnl.0000000000005516.

5. Wekerle H. B cells in multiple sclerosis. *Autoimmunity* (2017) **50**(1):57-60. Epub 2017/02/09. doi: 10.1080/08916934.2017.1281914. PubMed PMID: 28166681.

6. Baker D, Marta M, Pryce G, Giovannoni G, Schmierer K. Memory B Cells are Major Targets for Effective Immunotherapy in Relapsing Multiple Sclerosis. *EBioMedicine* (2017). Epub 2017/02/06. doi: 10.1016/j.ebiom.2017.01.042. PubMed PMID: 28161400.

7. Kinnunen T, Chamberlain N, Morbach H, Cantaert T, Lynch M, Preston-Hurlburt P, et al. Specific peripheral B cell tolerance defects in patients with multiple sclerosis. *The Journal of clinical investigation* (2013) **123**(6):2737-41. Epub 2013/05/17. doi: 10.1172/jci68775. PubMed PMID: 23676463; PubMed Central PMCID: PMC3668812.

8. Holmoy T, Vartdal F, Hestvik AL, Munthe L, Bogen B. The idiotype connection: linking infection and multiple sclerosis. *Trends in immunology* (2010) **31**(2):56-62. Epub 2009/12/08. doi: 10.1016/j.it.2009.11.001. PubMed PMID: 19962346.

9. Hoglund RA, Lossius A, Johansen JN, Homan J, Benth JS, Robins H, et al. In Silico Prediction Analysis of Idiotope-Driven T-B Cell Collaboration in Multiple Sclerosis. *Frontiers in immunology* (2017) **8**:1255. Epub 2017/10/19. doi: 10.3389/fimmu.2017.01255. PubMed PMID: 29038659; PubMed Central PMCID: PMC5630699.

10. Bogen B, Ruffini P. Review: to what extent are T cells tolerant to immunoglobulin variable regions? *Scandinavian journal of immunology* (2009) **70**(6):526-30. Epub 2009/11/13. doi: 10.1111/j.1365-3083.2009.02340.x. PubMed PMID: 19906193.

11. Munthe LA, Corthay A, Os A, Zangani M, Bogen B. Systemic autoimmune disease caused by autoreactive B cells that receive chronic help from Ig V region-specific T cells. *Journal of*

immunology (Baltimore, Md : 1950) (2005) **175**(4):2391-400. Epub 2005/08/06. doi: 10.4049/jimmunol.175.4.2391. PubMed PMID: 16081810.

12. Munthe LA, Os A, Zangani M, Bogen B. MHC-restricted Ig V region-driven T-B lymphocyte collaboration: B cell receptor ligation facilitates switch to IgG production. *Journal of immunology (Baltimore, Md : 1950)* (2004) **172**(12):7476-84. Epub 2004/06/10. doi: 10.4049/jimmunol.172.12.7476. PubMed PMID: 15187126.

13. Hestvik AL, Skorstad G, Vartdal F, Holmoy T. Idiotope-specific CD4(+) T cells induce apoptosis of human oligodendrocytes. *Journal of autoimmunity* (2009) **32**(2):125-32. Epub 2009/03/03. doi: 10.1016/j.jaut.2009.01.004. PubMed PMID: 19250800.

14. Hestvik AL, Vartdal F, Fredriksen AB, Thompson KM, Kvale EO, Skorstad G, et al. T cells from multiple sclerosis patients recognize multiple epitopes on Self-IgG. *Scandinavian journal of immunology* (2007) **66**(4):393-401. Epub 2007/09/14. doi: 10.1111/j.1365-3083.2007.01955.x. PubMed PMID: 17850583.

15. Holmoy T, Fredriksen AB, Thompson KM, Hestvik AL, Bogen B, Vartdal F. Cerebrospinal fluid T cell clones from patients with multiple sclerosis: recognition of idiotopes on monoclonal IgG secreted by autologous cerebrospinal fluid B cells. *European journal of immunology* (2005) **35**(6):1786-94. Epub 2005/05/03. doi: 10.1002/eji.200425417. PubMed PMID: 15864781.

16. Holmoy T, Vandvik B, Vartdal F. T cells from multiple sclerosis patients recognize immunoglobulin G from cerebrospinal fluid. *Multiple sclerosis (Houndmills, Basingstoke, England)* (2003) **9**(3):228-34. Epub 2003/06/20. doi: 10.1191/1352458503ms906oa. PubMed PMID: 12814167.

17. Aas-Hanssen K, Funderud A, Thompson KM, Bogen B, Munthe LA. Idiotype-specific Th cells support oligoclonal expansion of anti-dsDNA B cells in mice with lupus. *Journal of immunology (Baltimore, Md : 1950)* (2014) **193**(6):2691-8. Epub 2014/08/17. doi: 10.4049/jimmunol.1400640. PubMed PMID: 25127856.

18. Jacobsen J, Haabeth O-AW, Tveita AA, Schjetne KW, Munthe LA, Bogen B. Naive Idiotope-Specific B and T Cells Collaborate Efficiently in the Absence of Dendritic Cells. *The Journal of Immunology* (2014) **192**(9):4174-83. doi: 10.4049/jimmunol.1302359.

19. Hamze M, Meunier S, Karle A, Gdoura A, Goudet A, Szely N, et al. Characterization of CD4 T Cell Epitopes of Infliximab and Rituximab Identified from Healthy Donors. *Frontiers in immunology* (2017) **8**(500). doi: 10.3389/fimmu.2017.00500.

20. Cassotta A, Mikol V, Bertrand T, Pouzieux S, Le Parc J, Ferrari P, et al. A single T cell epitope drives the neutralizing anti-drug antibody response to natalizumab in multiple sclerosis patients. *Nature medicine* (2019) **25**(9):1402-7. doi: 10.1038/s41591-019-0568-2.

21. Jawa V, Cousens LP, Awwad M, Wakshull E, Kropshofer H, De Groot AS. T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. *Clinical Immunology* (2013) **149**(3, Part B):534-55. doi: https://doi.org/10.1016/j.clim.2013.09.006.

22. Harding FA, Stickler MM, Razo J, DuBridge RB. The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. *MAbs* (2010) **2**(3):256-65. Epub 2010/04/20. doi: 10.4161/mabs.2.3.11641. PubMed PMID: 20400861; PubMed Central PMCID: PMC2881252.

23. Bremel RD, Homan EJ. An integrated approach to epitope analysis II: A system for proteomic-scale prediction of immunological characteristics. *Immunome research* (2010) **6**:8. Epub

2010/11/04. doi: 10.1186/1745-7580-6-8. PubMed PMID: 21044290; PubMed Central PMCID: PMC2991286.

24. Bremel RD, Homan EJ. An integrated approach to epitope analysis I: Dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. *Immunome research* (2010) **6**:7. Epub 2010/11/04. doi: 10.1186/1745-7580-6-7. PubMed PMID: 21044289; PubMed Central PMCID: PMC2990731.

25. Bremel RD, Homan EJ. Recognition of higher order patterns in proteins: immunologic kernels. *PloS one* (2013) **8**(7):e70115. Epub 2013/08/08. doi: 10.1371/journal.pone.0070115. PubMed PMID: 23922927; PubMed Central PMCID: PMC3726486.

26. Høglund RA, Torsetnes SB, Lossius A, Bogen B, Homan EJ, Bremel R, et al. Human Cysteine Cathepsins Degrade Immunoglobulin G In Vitro in a Predictable Manner. *International Journal of Molecular Sciences* (2019) **20**(19):4843. PubMed PMID: doi:10.3390/ijms20194843.

27. Bremel RD, Homan EJ. Frequency Patterns of T-Cell Exposed Amino Acid Motifs in Immunoglobulin Heavy Chain Peptides Presented by MHCs. *Frontiers in immunology* (2014) **5**:541. Epub 2014/11/13. doi: 10.3389/fimmu.2014.00541. PubMed PMID: 25389426; PubMed Central PMCID: PMC4211557.

28. Yamano T, Nedjic J, Hinterberger M, Steinert M, Koser S, Pinto S, et al. Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. *Immunity* (2015) **42**(6):1048-61. doi: 10.1016/j.immuni.2015.05.013.

29. Lauritzsen GF, Hofgaard PO, Schenck K, Bogen B. Clonal deletion of thymocytes as a tumor escape mechanism. *International journal of cancer* (1998) **78**(2):216-22. Epub 1998/10/01. doi: 10.1002/(SICI)1097-0215(19981005)78:2<216::AID-IJC16>3.0.CO;2-8. PubMed PMID: 9754655.

30. Bonasio R, Scimone ML, Schaerli P, Grabie N, Lichtman AH, von Andrian UH. Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. *Nat Immunol* (2006) **7**(10):1092-100. Epub 2006/09/05. doi: 10.1038/ni1385. PubMed PMID: 16951687.

31. Johansen JN, Vartdal F, Desmarais C, Tutturen AE, de Souza GA, Lossius A, et al. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: Equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. *Clinical immunology (Orlando, Fla)* (2015) **160**(2):211-25. Epub 2015/06/10. doi: 10.1016/j.clim.2015.06.001. PubMed PMID: 26055752.

32. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. *Data from: A public database of memory and naive B-cell receptor sequences*. Dryad Data Repository (2016).

33. Meier S, Stark R, Frentsch M, Thiel A. The influence of different stimulation conditions on the assessment of antigen-induced CD154 expression on CD4+ T cells. *Cytometry A* (2008) **73**(11):1035-42. Epub 2008/09/13. doi: 10.1002/cyto.a.20640. PubMed PMID: 18785645.

34. Frentsch M, Arbach O, Kirchhoff D, Moewes B, Worm M, Rothe M, et al. Direct access to CD4+ T cells specific for defined antigens according to CD154 expression. *Nature medicine* (2005) **11**(10):1118-24. Epub 2005/09/28. doi: 10.1038/nm1292. PubMed PMID: 16186818.

35. Wu Q, Wang Q, Mao G, Dowling CA, Lundy SK, Mao-Draayer Y. Dimethyl Fumarate Selectively Reduces Memory T Cells and Shifts the Balance between Th1/Th17 and Th2 in Multiple Sclerosis Patients. *The Journal of Immunology* (2017) **198**(8):3069-80. doi: 10.4049/jimmunol.1601532.

36. Longbrake EE, Ramsbottom MJ, Cantoni C, Ghezzi L, Cross AH, Piccio L. Dimethyl fumarate selectively reduces memory T cells in multiple sclerosis patients. *Multiple Sclerosis Journal* (2016) **22**(8):1061-70. doi: 10.1177/1352458515608961. PubMed PMID: 26459150.

37. Adler LN, Jiang W, Bhamidipati K, Millican M, Macaubas C, Hung SC, et al. The Other Function: Class II-Restricted Antigen Presentation by B Cells. *Frontiers in immunology* (2017) **8**:319. Epub 2017/04/08. doi: 10.3389/fimmu.2017.00319. PubMed PMID: 28386257; PubMed Central PMCID: PMC5362600.

38. Axtell RC, Steinman L. Gaining entry to an uninflamed brain. *Nature Immunology* (2009) **10**(5):453-5. doi: 10.1038/ni0509-453.

39. Müller M, Carter S, Hofer MJ, Campbell IL. Review: The chemokine receptor CXCR3 and its ligands CXCL9, CXCL10 and CXCL11 in neuroimmunity – a tale of conflict and conundrum. *Neuropathology and Applied Neurobiology* (2010) **36**(5):368-87. doi: 10.1111/j.1365-2990.2010.01089.x.

40. Reboldi A, Coisne C, Baumjohann D, Benvenuto F, Bottinelli D, Lira S, et al. C-C chemokine receptor 6–regulated entry of TH-17 cells into the CNS through the choroid plexus is required for the initiation of EAE. *Nature Immunology* (2009) **10**(5):514-23. doi: 10.1038/ni.1716.

41. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic acids research* (2008) **36**(Web Server issue):W503-8. Epub 2008/05/27. doi: 10.1093/nar/gkn316. PubMed PMID: 18503082; PubMed Central PMCID: PMC2447746.

42. Consortium IMSG. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science (New York, NY)* (2019) **365**(6460):eaav7188. doi: 10.1126/science.aav7188.

43. Huszthy PC, Gopalakrishnan RP, Jacobsen JT, Haabeth OAW, Løset GÅ, Braathen R, et al. B cell receptor ligation induces display of V-region peptides on MHC class II molecules to T cells. *Proceedings of the National Academy of Sciences of the United States of America* (2019):In press.

44. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology* (2009) **10**(11):R130. doi: 10.1186/gb-2009-10-11-r130.

45. Kim A, Hartman IZ, Poore B, Boronina T, Cole RN, Song N, et al. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nature communications* (2014) **5**(1):5369. doi: 10.1038/ncomms6369.

46. Sospedra M. B cells in multiple sclerosis. *Current opinion in neurology* (2018) **31**(3):256-62. Epub 2018/04/10. doi: 10.1097/wco.000000000000563. PubMed PMID: 29629941.

47. Khodadoust MS, Olsson N, Chen B, Sworder B, Shree T, Liu CL, et al. B-cell lymphomas present immunoglobulin neoantigens. *Blood* (2019) **133**(8):878-81. Epub 2018/12/14. doi: 10.1182/blood-2018-06-845156. PubMed PMID: 30545830; PubMed Central PMCID: PMC6384186.

48. Khodadoust MS, Olsson N, Wagar LE, Haabeth OAW, Chen B, Swaminathan K, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* (2017) **543**(7647):723-7. doi: 10.1038/nature21433. PubMed PMID: 28329770.

49. Collado JA, Alvarez I, Ciudad MT, Espinosa G, Canals F, Pujol-Borrell R, et al. Composition of the HLA-DR-associated human thymus peptidome. *European journal of immunology* (2013) **43**(9):2273-82. Epub 2013/05/31. doi: 10.1002/eji.201243280. PubMed PMID: 23719902.

50. Seward RJ, Drouin EE, Steere AC, Costello CE. Peptides presented by HLA-DR molecules in synovia of patients with rheumatoid arthritis or antibiotic-refractory Lyme arthritis. *Molecular & cellular proteomics : MCP* (2011) **10**(3):M110.002477-M110. Epub 2010/11/16. doi: 10.1074/mcp.M110.002477. PubMed PMID: 21081667.

51. Heyder T, Kohler M, Tarasova NK, Haag S, Rutishauser D, Rivera NV, et al. Approach for Identifying Human Leukocyte Antigen (HLA)-DR Bound Peptides from Scarce Clinical Samples. *Molecular & cellular proteomics : MCP* (2016) **15**(9):3017-29. Epub 2016/07/24. doi: 10.1074/mcp.M116.060764. PubMed PMID: 27452731.

52. Sorde L, Spindeldreher S, Palmer E, Karle A. Tregitopes and impaired antigen presentation: Drivers of the immunomodulatory effects of IVIg? *Immun Inflamm Dis* (2017) **5**(4):400-15. Epub 2017/06/01. doi: 10.1002/iid3.167. PubMed PMID: 28560793; PubMed Central PMCID: PMC5691310.

53. Bergseng E, Dorum S, Arntzen MO, Nielsen M, Nygard S, Buus S, et al. Different binding motifs of the celiac disease-associated HLA molecules DQ2.5, DQ2.2, and DQ7.5 revealed by relative quantitative proteomics of endogenous peptide repertoires. *Immunogenetics* (2015) **67**(2):73-84. Epub 2014/12/17. doi: 10.1007/s00251-014-0819-9. PubMed PMID: 25502872; PubMed Central PMCID: PMC4297300.

54. Williams WM, Staines NA, Muller S, Isenberg DA. Human T cell responses to autoantibody variable region peptides. *Lupus* (1995) **4**(6):464-71. Epub 1995/12/01. doi: 10.1177/096120339500400608. PubMed PMID: 8749569.

55. Dayan M, Segal R, Sthoeger Z, Waisman A, Brosh N, Elkayam O, et al. Immune response of SLE patients to peptides based on the complementarity determining regions of a pathogenic anti-DNA monoclonal antibody. *Journal of clinical immunology* (2000) **20**(3):187-94. Epub 2000/08/15. doi: 10.1023/a:1006685413157. PubMed PMID: 10941826.

56. van Schooten WC, Devereux D, Ho CH, Quan J, Aguilar BA, Rust CJ. Joint-derived T cells in rheumatoid arthritis react with self-immunoglobulin heavy chains or immunoglobulin-binding proteins that copurify with immunoglobulin. *European journal of immunology* (1994) **24**(1):93-8. Epub 1994/01/01. doi: 10.1002/eji.1830240115. PubMed PMID: 8020576.

57. Jerne NK. Towards a network theory of the immune system. *Ann Immunol (Paris)* (1974) **125c**(1-2):373-89. Epub 1974/01/01. PubMed PMID: 4142565.

58. Klein L, Robey EA, Hsieh C-S. Central CD4+ T cell tolerance: deletion versus regulatory T cell differentiation. *Nature Reviews Immunology* (2019) **19**(1):7-18. doi: 10.1038/s41577-018-0083-6.

59. Noyan F, Lee YS, Zimmermann K, Hardtke-Wolenski M, Taubert R, Warnecke G, et al. Isolation of human antigen-specific regulatory T cells with high suppressive function. *European journal of immunology* (2014) **44**(9):2592-602. Epub 2014/07/06. doi: 10.1002/eji.201344381. PubMed PMID: 24990119.

60. Bacher P, Scheffold A. Flow-cytometric analysis of rare antigen-specific T cells. *Cytometry A* (2013) **83**(8):692-701. Epub 2013/06/22. doi: 10.1002/cyto.a.22317. PubMed PMID: 23788442.

61. Xu-Monette ZY, Li J, Xia Y, Crossley B, Bremel RD, Miao Y, et al. Immunoglobulin somatic hypermutation has clinical impact in DLBCL and potential implications for immune checkpoint blockade and neoantigen-based immunotherapies. *Journal for ImmunoTherapy of Cancer* (2019) **7**(1):272. doi: 10.1186/s40425-019-0730-x.

62. Venken K, Hellings N, Thewissen M, Somers V, Hensen K, Rummens JL, et al. Compromised CD4+ CD25(high) regulatory T-cell function in patients with relapsing-remitting multiple sclerosis is correlated with a reduced frequency of FOXP3-positive cells and reduced FOXP3 expression at the single-cell level. *Immunology* (2008) **123**(1):79-89. Epub 2007/09/28. doi: 10.1111/j.1365-2567.2007.02690.x. PubMed PMID: 17897326; PubMed Central PMCID: PMC2433271.

63. Fletcher JM, Lonergan R, Costelloe L, Kinsella K, Moran B, O'Farrelly C, et al. CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. *Journal of immunology (Baltimore, Md : 1950)* (2009) **183**(11):7602-10. Epub 2009/11/18. doi: 10.4049/jimmunol.0901881. PubMed PMID: 19917691.

12 Tables

Table 1: Criteria for *in silico* prediction of antigenic properties

	Stimulatory idiotopes	Tolerogenic idiotopes	Inert idiotopes
Johnson SI normalized predicted HLA affinity	< -1.5[1,2]	< -1.5[1,2]	> 1[1,2,,3]
Cathepsin cleavage probability	Cleaved by S or L or B (liberal)	Cleaved by S or L or B (liberal)	Neither S nor L nor B (strict)
TCEM IIa/b frequency class	≥17	≤7	Both high and low
N=	10 / patient	5 / patient	Same 7 for all

1 Patient specific, based on HLA-DR/DQ/DP typing

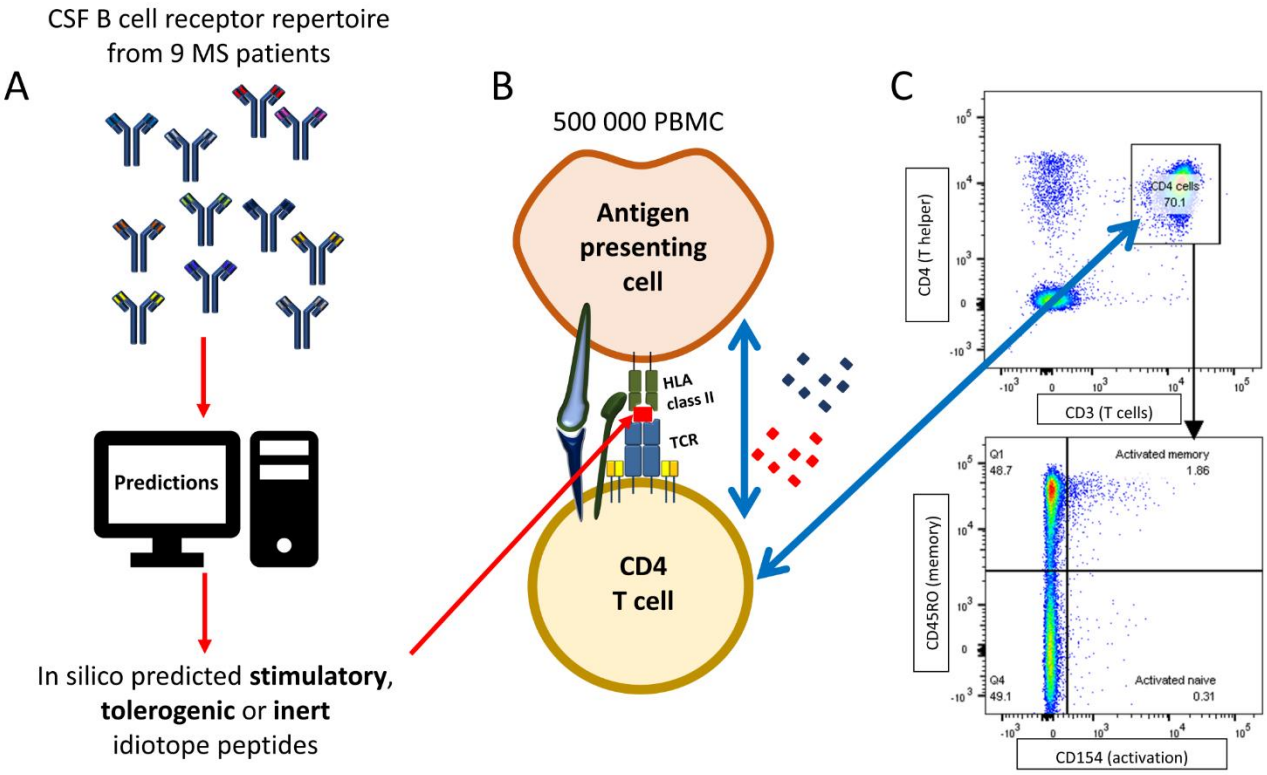
2 If too few sequences fulfilled all criteria, this was adjusted until enough sequences could be included

3 Selected from the full pool of patient IGHV sequences. The available predicted HLA affinity had to be low for all the patients' HLA class II variants (incl. DQ and DP in addition to DR).

4 T cell exposed motif frequency scale is inverse logarithmic. Frequency class 17 indicates the motif occurs no more frequently than once every 2^{17} IGHV sequence.

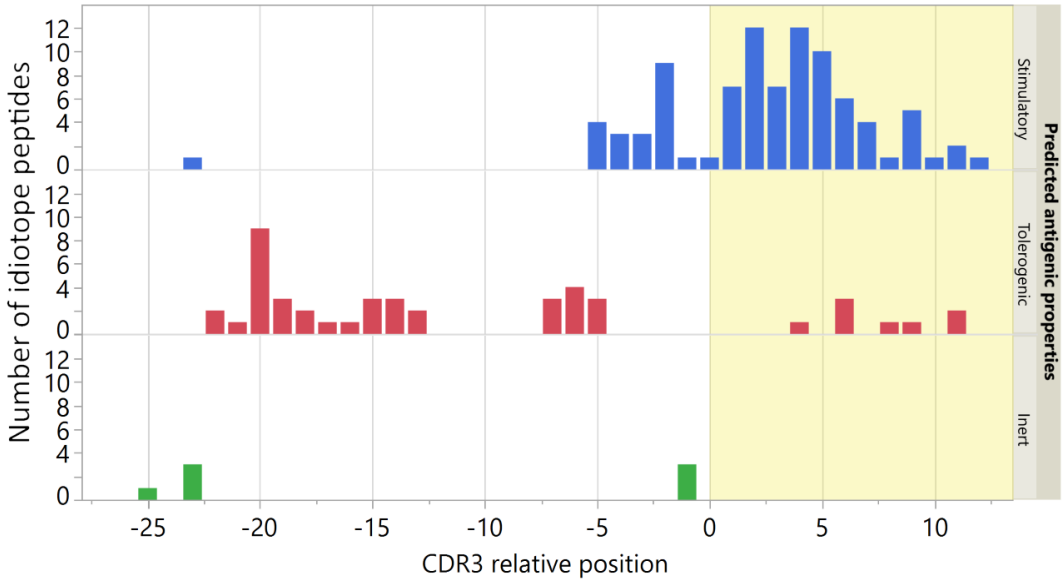
13 Figures

Figure 1. Flow cytometry based idiotope-specific T cell activation assay



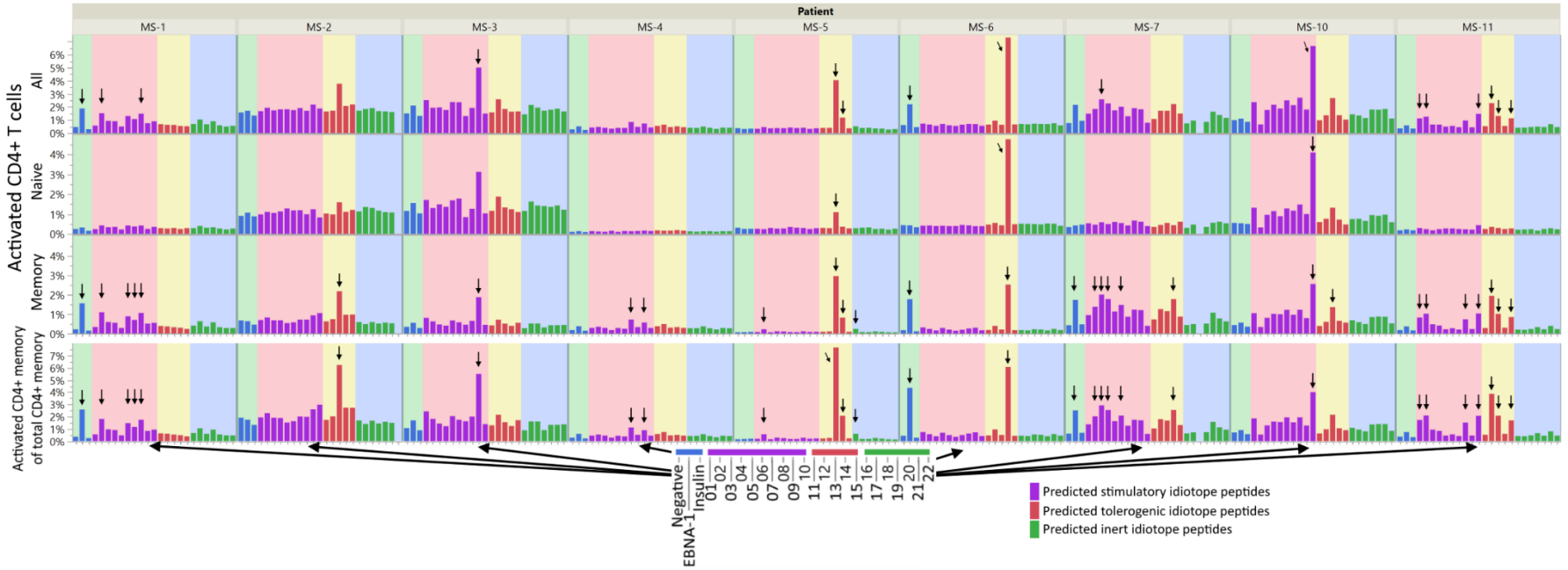
A) IGHV amino acid sequences (mean 1079 [SD=1213] per patient) from nine MS patients were run through predictive models to identify likely antigenic idiotopes based on HLA class II affinity, cathepsin cleavage and frequency classification (FC) of T cell exposed motifs (TCEM). B) 500,000 PBMC were stimulated with synthetic idiotope peptides predicted to be stimulatory, tolerogenic or inert as well as positive and negative controls for 12 hours in presence of anti-CD40 antibodies. B cells or other professional APCs with idiotope peptides bound to their HLA class II receptor may activate cognate $CD4^+$ T cells. C) $CD4^+CD45RO^+$ memory T cells specifically activated by idiotope peptides were detected by surface expression of CD154, upregulated upon TCR stimulation. The example shows a detected memory T cell response to idiotope peptide 12 in patient MS-11.

Figure 2. Location of origin for idiotope peptides



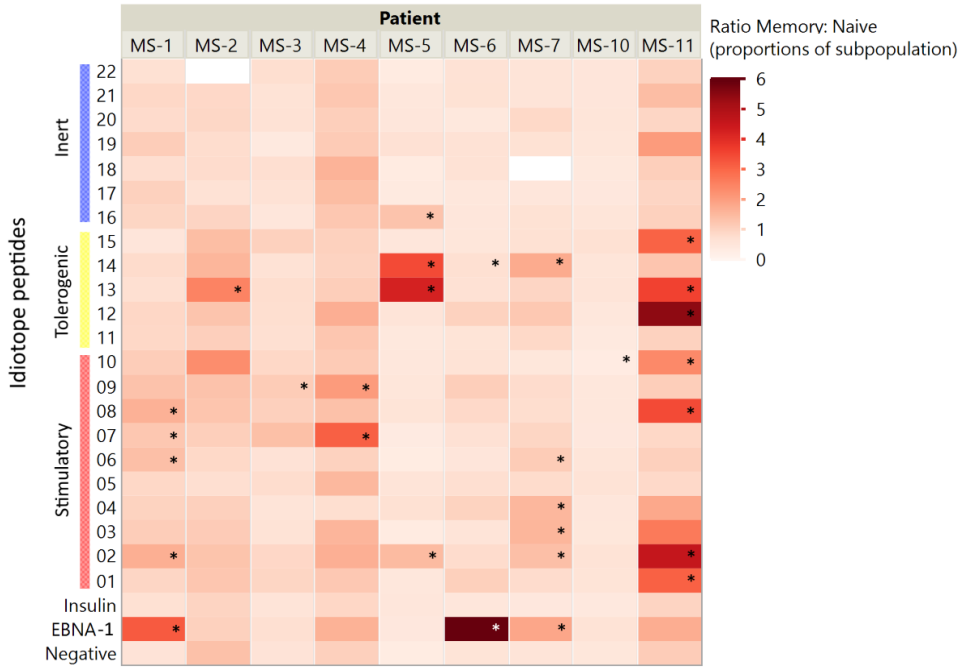
The CDR3 relative position was determined by the location of the first amino acid in each 15-mer in the original IGHV sequence. The seven predicted inert peptides were used in all nine patients. Bar colors indicate predicted antigenic properties and yellow shading indicates the CDR3 region.

Figure 3. CD4$^+$ T cell responses against idiotope peptides.



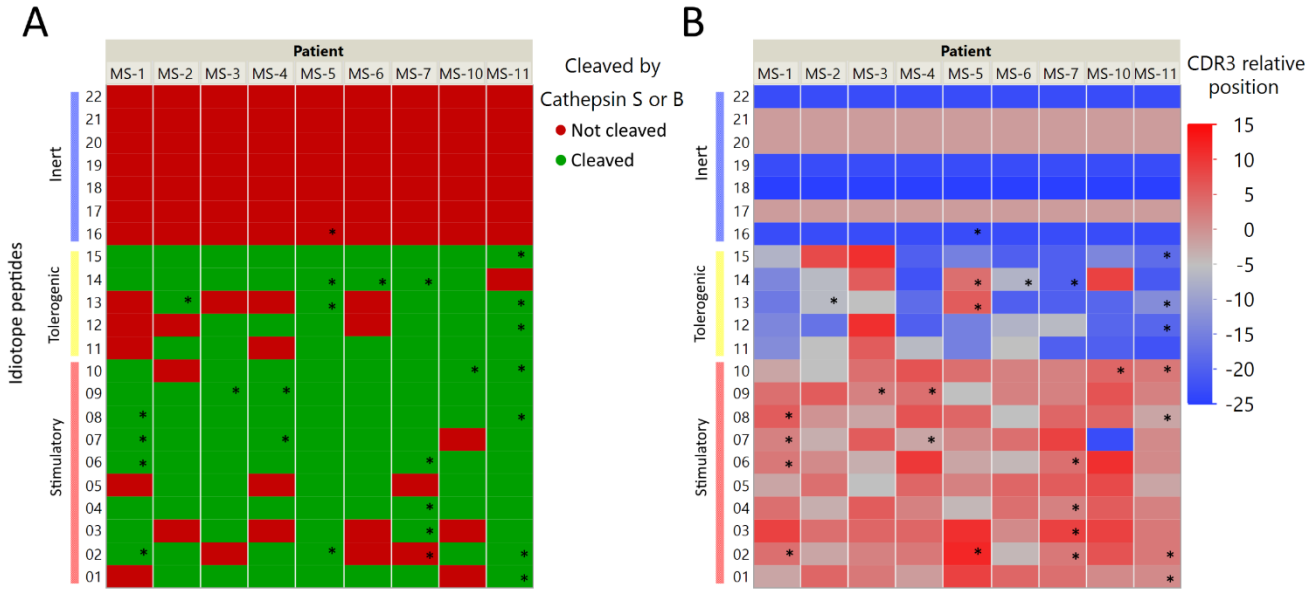
A total of 500,000 PBMC were left unstimulated, or stimulated with EBNA-1 peptide mix, insulin peptide mix, anti CD3/CD28 beads (not shown) or one of 22 idiotope peptides for 12 hours in presence of anti-CD40 antibodies and analyzed by flow cytometry. We gated on CD3$^+$CD4$^+$CD8$^-$ T cells and assayed for the activation marker CD154 among all CD4$^+$ cells, CD45RO$^+$ memory- or CD45RO$^-$ naïve cells. Activated cells are presented as proportions of all CD4$^+$ cells (upper three panels) or proportion of memory cells (lower panel). Responses were deemed positive (arrows) if the proportion of CD154$^+$ cells were 3x higher than in unstimulated (negative) wells.

Figure 4. Memory to naïve activation ratios



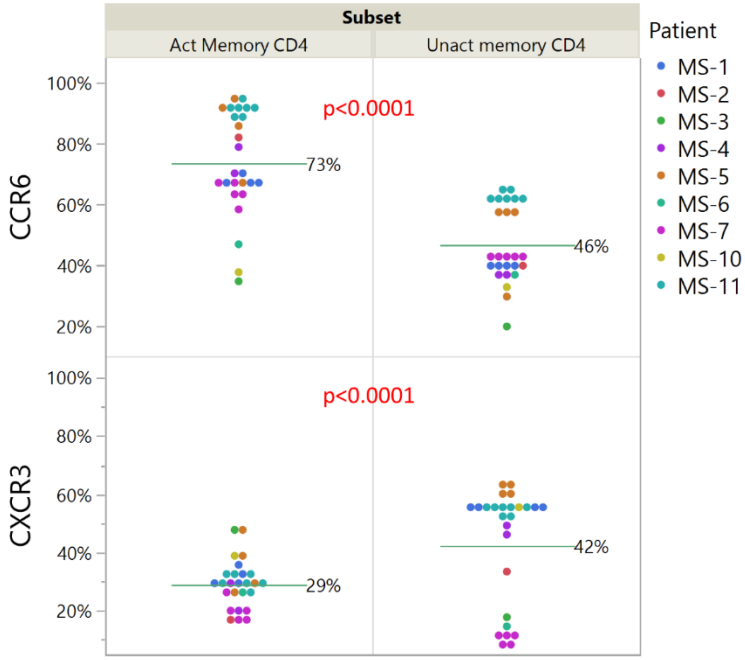
The ratios of activated CD4⁺ cells (% of CD45RO⁺ memory: % of CD45RO⁻ naïve cells) were assessed for all idiotope peptides. A higher ratio indicates a higher proportion of responder cells among memory CD4+ T cells than among naïve cells. The idiotope peptides that elicited memory T cell responses three times higher than the unstimulated control are marked with *.

Figure 5. IGHV Localization and predicted possibility of cathepsin processing of antigenic idiotope peptide panel



Idiotope peptides were labelled according to their cathepsin S/B cleavage prediction (A) or position of origin within the IGHV sequence (B). Idiotope peptides that generated a memory T cell response are labelled with *.

Figure 6. Relative expression of chemokine receptors CCR6 and CXCR3 on idiotope specific CD4$^+$ T cells.



Expression among activated (CD154$^+$) or un-activated (CD154$^-$) memory (CD45RO$^+$) CD4$^+$ T cells responding towards predicted idiotope peptides. P-values are results of full factorial mixed model, differences shown are un-adjusted values.

Figure 7. Effect of somatic mutations on predicted HLA DR affinity, cleavage possibility and frequency classification of T cell exposed motifs.

Patient	Idiotope peptide	IGHV sequence (idiotope peptide)	Prediction changes from germline sequence: HLA-DR affinity	Prediction changes from germline sequence: Cleavage probability	TCEMIIa/b FC
MS-1	02	RADDTAVYYCARE*RDVT*ISGAII*WL*DYYMDV----------WGKGTLVTVS	+ ÷	–	14/16 to 19/19
MS-1	06	RAEDTAFYYCTKD*RLGWWEVA*YDFWSGS*N*DY--------WGRGTLVTVS	– –	÷	19/22
MS-1	07	NSLRGEDTAVYYCA*REVYY*HFWSG*NVGGF*DP----------WGQGTLVTVS	– +	–	13/15 to 20/22
MS-1	08	SDDTAVYFCAR*VNSGGF*HYDHN*GGVPFYM*DV------ WGEGTLVTVS	+ ÷	+	16/18 to 20/21
MS-2	13	KTLYLQMNSLRPEDTAIYYCARD*LFMIL SPD*----------WGQGTLVTVS	+ –	+	6/6
MS-3	09	TMTNMDPEDTATYFCAR*Y FYH*ITAYYY*AIDY*----------WGQGTLVTVS	+ +	+[3]	13/13 to 17/17
MS-4	07	FTVSRDGAQNSLYLHMNSLRPEDTAVYYCARLAFRTVDY----------WGPGTLVTVS	÷ ÷	–	20/18 to 17/17
MS-4	09	NEFSLKLSSVTVADTAVYYCARG*SFVI*VTGYY*YS*DAFDI----------WGQGTLVTVS	– –	–	15/18 to 17/18
MS-5	02	TMTNMDPVDTATYYCAR*MEDWYYYDNGVYGRWAPYFF*DY-WGQGTLVTVS	– –	–	17/17
MS-5	13	SKSQFSLRLSSVTAADTAVYYCAR*SQG*TGSYYYSYYMDV----------WGKGTLVTVS	– ÷	–	5/7
MS-5	14	DTSKNQFSLKLTSVTAADTAVYYCARGYSSGYYYYGMDV----------WGQGTLVTVS	÷ ÷	+	5/7
MS-5	16[1]	SVDTSKNEFSLKVTSVTAADSAVYYCAR*YRYGGDP IFDS*----------WGQGTLVTVS	n.a	n.a.	n.a.
MS-6	14	VTISEDTSKNQLSLRLSPVTAADTAVYYCAR*LRRSH*VDF----------WGHGTLVTVS	+ +	+ ÷	7/6
MS-7	02	SVSLHMKKLRVEDTAVYYCARE*LITFGTINVNWQFTN*DY----------WGRGTLVTVS	+ ?[2]	–	14/19 to 18/19
MS-7	03	SVSLHMKKLRVEDTAVYYCARE*LITFGTINVNWQFTNDY*----------WGRGTLVTVS	– ?[2]	+	19/19
MS-7	04	DTSRNQFSLKLTSLTAADTALYYCA*RIWRKALVTVYFHD*----------WGQGTLVTVS	+ ?[2]	+	17/17
MS-7	06	ELSRLRSDDTAVYYCARE*NTYYDFWRAVSPHKY*NYGLDV--WGQGTLVTVS	– ?[2]	+	17/17
MS-7	14	KNTLFLQMNSLRVEDTAVYYCARD*ADD*YYGSGSY*RY*FDY----------WGQGTLVTVS	+ ?[2]	+	1/3 to 5/2
MS-10	10	AKNSLYLQMNSLRAEDTAVYYCAR*GI*HGDY*HYRLY*FFDN----------WGQGTLVTVS	– +	+	17/15 to 20/19
MS-11	01	KNQISLKLSVTAADTAVYYCAR*GGRWLLQIG*YYYGMDV----------WGQGTLVTVS	– –	–	18/19
MS-11	02	SKNHFSLKLSSVTATDTAVYYCAR*RDIVLVPAADAYDI*----------WGQGTLVTVS	+ +	+ ÷	14/15 to 19/19
MS-11	08	DTSKSQVVLTMTNLDPVDTATYYCAR*QVYTFN*WFNWFDP----------WGQGTLVTVS	÷ ÷	–[3]	13/13 to 21/20
MS-11	10	STNQFSLKLTSVTAADTAVYYCAR*LLSVRKSWLSGW*FDP----------WGQGTLVTVS	– ÷	–	19/20
MS-11	12	TPSIDKSKNQFSLKLTFVTAADTAVYYCAR*LGGGAY*FDY----------WGQGTLVTVS	+ +	+ ÷	4/4 to 7/7
MS-11	13	KNQISLKLSVTAADTAVYYCAR*GGRWLLQIG*YYYGMDV----------WGQGTLVTVS	+ +	+ ÷	3/1
MS-11	15	ADKSISTAYLQWSSLKASDTAMYYCAR*PHYYDSLDAFDI*----------WGQGTLVTVS	– –	–	6/6
IGHV region:		Framework(FW)3 CDR3 FW4 Constant			

Each idiotope peptide (red) that generated memory T cell responses was aligned within their original IGHV sequence, according to the position of the first cysteine (yellow) of CDR3. Mutations (underlined) and *insertions* (italic) were identified using IMGT V-quest. The IGHV sequences were compared to imputed germline variants to identify changes in predicted outcomes caused by mutations. Change in affinity for patients' two DR alleles was determined by >0.1 difference in ln(IC50) value and >0.1 change in probability for cathepsin S or B cleavage at either side of idiotope peptide ± 3 amino acids. Changes are depicted as + (higher), ÷ lower and − unchanged. Green indicates imputed net improved-, yellow mixed- and red indicates net negative effect for antigen presentation. TCEM FC: T cell exposed motif frequency class

[1]Peptide not in patient's CSF IGHV repertoire. [2]Missing HLA prediction for one allele. [3]These also had much lower probability of intra-peptide destruction by cathepsins

Supplementary Materials for

$CD4^+$ T cells in the blood of MS patients respond to predicted epitopes from B cell receptors found in spinal fluid

Supplementary Table 1. Subject characteristics at time of sample collection

ID	Sex	Age	Diagnosis	Disease duration (months)[a]	Previous or ongoing treatment	Number of relapses	Time since relapse (months)	OCB[b]	CSF cell count[c]	Albumin ratio	IgG index	CSF IGHV Sequences (N)
MS-1	F	29	RR-MS[d]	14	None	1	14	+	3	4	0.85	98
MS-2	F	33	RR-MS	156	IFN β 1a (48 months)	5	60	+	1	5.1	0.81	241
MS-3	F	36	RR-MS	66	GA (27 months)	4	33	+	3	3.8	1.7	417
MS-4	F	63	RR-MS	23	IFN β 1b (2 months)	1	23	+	1	3.6	0.68	109
MS-5	F	27	RR-MS	6	None	4	4	+	2	3.8	1.2	1409
MS-6	F	33	RR-MS[d]	23	None	1	23	+	3	3.5	0.55	3409
MS-7	M	33	RR-MS	48	None	3	3	+	10	17	0.9	2753
MS-10	F	42	RR-MS	35	None	2	18	+	11	3.2	1.06	617
MS-11	F	20	RR-MS	13	None	2	1	+	10[e]	4.8[e]	1[e]	658

F: female; M: male; RR-MS: relapse-remitting multiple sclerosis; OCB: CSF-specific oligoclonal IgG bands; CSF: cerebrospinal fluid; IGHV: Immunoglobulin heavy variable; IFN: interferon; GA: Glatiramer acetate

[a] Since first symptom. [b] OCB positive indicates presence of >2 CSF specific IgG bands on isoelectric focusing. [c] Number of mononuclear cells per microliter CSF.

[d] Patients MS-1 and MS-6 had clinically isolated syndrome at inclusion, but later developed definite RR-MS. [e] Values are from one month prior sampling. The sample from which we acquired cells for sequencing was stopped after 2 mL due to accidental bleeding

Supplementary Table 2. Selected idiotope peptides

Patient	Peptide	Group	Idiotope peptide
MS1	1	Predicted stimulatory	YYCARERDVTISGAI
MS1	2	Predicted stimulatory	RDVTISGAIIWLDYY
MS1	3	Predicted stimulatory	SGAIIWLDYYMDVWG
MS1	4	Predicted stimulatory	NERWLRTVAPLDSWG
MS1	5	Predicted stimulatory	YYCARFPRAFREDWF
MS1	6	Predicted stimulatory	DRLGWWEVAYDFWSG
MS1	7	Predicted stimulatory	REVYYHFWSGNVGGF
MS1	8	Predicted stimulatory	GGFHYDHNGGVPFYM
MS1	9	Predicted stimulatory	ADIVLISAASPFDYW
MS1	10	Predicted stimulatory	YYCARMRGSYRYFFD
MS1	11	Predicted tolerogenic	MNSLRAEDTAVYYCA
MS1	12	Predicted tolerogenic	ELRSLRADDTAVYYC
MS1	13	Predicted tolerogenic	YMELRSLRSDDTAVY
MS1	14	Predicted tolerogenic	ELRSLRSDDTAVYYC
MS1	15	Predicted tolerogenic	DDTAVYYCATDADIV
MS2	1	Predicted stimulatory	GMSPFYYFYDMDVWG
MS2	2	Predicted stimulatory	YFCARARDYVPFYFG
MS2	3	Predicted stimulatory	RDYVPFYFGIEVWGQ
MS2	4	Predicted stimulatory	TYYCGHSLFRDLLSS
MS2	5	Predicted stimulatory	RGVVIPMDGGFDYWG
MS2	6	Predicted stimulatory	AKCPHWGNSWYAPFD
MS2	7	Predicted stimulatory	LYYCAIAWGFWSTYY
MS2	8	Predicted stimulatory	CAIAWGFWSTYYPFY
MS2	9	Predicted stimulatory	YEFWSDYYPAPHNWF
MS2	10	Predicted stimulatory	TAVYYCALLRNYYFD
MS2	11	Predicted tolerogenic	TAVYYCARLRGVVIP
MS2	12	Predicted tolerogenic	FFLQWSSLKASDTAM
MS2	13	Predicted tolerogenic	DTAIYYCARDLFMIL
MS2	14	Predicted tolerogenic	DTAVYYCAKDLWYYD
MS2	15	Predicted tolerogenic	GWSYYYYYGMDVWGQ
MS3	1	Predicted stimulatory	ASVLWFGVRGSYFDY
MS3	2	Predicted stimulatory	RGGNTMVWGLFITSD
MS3	3	Predicted stimulatory	NTMVWGLFITSDSYA
MS3	4	Predicted stimulatory	EGFGVIILGPIDYWG
MS3	5	Predicted stimulatory	TATYFCAWTPTAYWR
MS3	6	Predicted stimulatory	TYFCAWTPTAYWRFE
MS3	7	Predicted stimulatory	EGFGVILLGPIDYWG
MS3	8	Predicted stimulatory	YFCARYFYHITAYYY
MS3	9	Predicted stimulatory	RYFYHITAYYYAIDY
MS3	10	Predicted stimulatory	FYHITAYYYAIDYWG
MS3	11	Predicted tolerogenic	LGSQYYYYGMDVWGR
MS3	12	Predicted tolerogenic	ICYRYYYYGMDVWGQ

MS3	13	Predicted tolerogenic	TAVYYCARPLGRVRG
MS3	14	Predicted tolerogenic	LGSHYYYYGMDVWGR
MS3	15	Predicted tolerogenic	LCYKYYYYGMDVWGQ
MS4	1	Predicted stimulatory	YCAREVNLKWELLVF
MS4	2	Predicted stimulatory	EVNLKWELLVFDAFD
MS4	3	Predicted stimulatory	NLKWELLVFDAFDIW
MS4	4	Predicted stimulatory	RHREWLRYRGFDYWG
MS4	5	Predicted stimulatory	STQLWWGLDFGSWGQ
MS4	6	Predicted stimulatory	LGGVIVPQIVDPWGQ
MS4	7	Predicted stimulatory	YYCARLAFRTVDYWG
MS4	8	Predicted stimulatory	VTYYDILTGFQPKFG
MS4	9	Predicted stimulatory	SFVIVTGYYYSDAFD
MS4	10	Predicted stimulatory	YYDFWSGPGHIDYWG
MS4	11	Predicted tolerogenic	DTAVYYCATGLFYEI
MS4	12	Predicted tolerogenic	KNSFFLQMNSLRAAD
MS4	13	Predicted tolerogenic	QFSLRLSSVTAADTA
MS4	14	Predicted tolerogenic	KSKNQFSLKLTSLTA
MS4	15	Predicted tolerogenic	KKTLYLQMNSLKTED
MS5	1	Predicted stimulatory	YDNGVYGRWAPYFFD
MS5	2	Predicted stimulatory	GVYGRWAPYFFDYWG
MS5	3	Predicted stimulatory	PSCYNRNYYFHGLDV
MS5	4	Predicted stimulatory	AVYYCTTVGHMGYFY
MS5	5	Predicted stimulatory	SSEWELMMIVDYWGQ
MS5	6	Predicted stimulatory	YSCARLVIFGMVIID
MS5	7	Predicted stimulatory	ARLVIFGMVIIDNVP
MS5	8	Predicted stimulatory	IFGMVIIDNVPLNWF
MS5	9	Predicted stimulatory	TATFYCAHVWPGYTY
MS5	10	Predicted stimulatory	WPGYTYGYPNNWLDP
MS5	11	Predicted tolerogenic	LKLRSVTAADTAVYF
MS5	12	Predicted tolerogenic	LKLRSVTATDTAFYY
MS5	13	Predicted tolerogenic	TGSYYYSYYMDVWGK
MS5	14	Predicted tolerogenic	YSSGYYYYGMDVWGQ
MS5	15	Predicted tolerogenic	LNLRSVTAADTAVYF
MS6	1	Predicted stimulatory	SSSLYLYYYSMDVWG
MS6	2	Predicted stimulatory	AMYFCTREGLFPRPF
MS6	3	Predicted stimulatory	ARDFYGCRGDKCHLT
MS6	4	Predicted stimulatory	PYYYDTTVMDWFDPW
MS6	5	Predicted stimulatory	GLLVLQGWGWAYDYW
MS6	6	Predicted stimulatory	AVYYCVSADTFYYYY
MS6	7	Predicted stimulatory	RKFYGAVLQMTFHLW
MS6	8	Predicted stimulatory	TAVYYCVCWAGWLVA
MS6	9	Predicted stimulatory	RLGYSYGPRWWFDPW
MS6	10	Predicted stimulatory	RSEQWLTTTEYFQHW
MS6	11	Predicted tolerogenic	TAVYYCARARGWFGL
MS6	12	Predicted tolerogenic	DDTAVYYCARVWWDQ

MS6	13	Predicted tolerogenic	QNSVYLQMDSLRAED
MS6	14	Predicted tolerogenic	ADTAVYYCARLRRSH
MS6	15	Predicted tolerogenic	KNSLYLQMNSLRTED
MS7	1	Predicted stimulatory	RGAWLTNDYYTYYGL
MS7	2	Predicted stimulatory	ELITFGTINVNWQFT
MS7	3	Predicted stimulatory	TINVNWQFTNDYWGR
MS7	4	Predicted stimulatory	RIWRKALVTVYFHDW
MS7	5	Predicted stimulatory	YYDFWSGNPDRFDYW
MS7	6	Predicted stimulatory	NTYYDFWRAVSPHKY
MS7	7	Predicted stimulatory	GGYIAWGPKKHYYYG
MS7	8	Predicted stimulatory	RLPTIWARNPNFHYY
MS7	9	Predicted stimulatory	KVADTLAVRLPYFDC
MS7	10	Predicted stimulatory	RSHPNFYLGELSFEG
MS7	11	Predicted tolerogenic	KNTLYLQMNSLRPED
MS7	12	Predicted tolerogenic	DTAIYYCARDRLLWF
MS7	13	Predicted tolerogenic	KNTMFLQMNSLRVED
MS7	14	Predicted tolerogenic	KNTLFLQMNSLRVED
MS7	15	Predicted tolerogenic	KNTLFLQMNSLRAED
MS10	1	Predicted stimulatory	AKAVRMQLWLFGSWG
MS10	2	Predicted stimulatory	NGRFLEWFPLYYFDY
MS10	3	Predicted stimulatory	RFLEWFPLYYFDYWG
MS10	4	Predicted stimulatory	NYDILTGFYLASLEL
MS10	5	Predicted stimulatory	DILTGFYLASLELID
MS10	6	Predicted stimulatory	TGFYLASLELIDSWG
MS10	7	Predicted stimulatory	DNAMDILYLQVNSLR
MS10	8	Predicted stimulatory	RINLWTAMPAGGPGL
MS10	9	Predicted stimulatory	NLWTAMPAGGPGLND
MS10	10	Predicted stimulatory	HGDYHYRLYFFDNWG
MS10	11	Predicted tolerogenic	KSMLYLQMNSLRVED
MS10	12	Predicted tolerogenic	NSLYLQMDSLRAEDM
MS10	13	Predicted tolerogenic	NSLYLQMNSLRAEDM
MS10	14	Predicted tolerogenic	TTSYYYFYYMDVWGK
MS10	15	Predicted tolerogenic	KLSFVTAADTAVFYC
MS11	1	Predicted stimulatory	ARGGRWLLQIGYYYG
MS11	2	Predicted stimulatory	RRDIVLVPAADAYDI
MS11	3	Predicted stimulatory	RLARELILGPEYYYY
MS11	4	Predicted stimulatory	RFDIATTVPLGFDYW
MS11	5	Predicted stimulatory	YYCVRLVPKRTATLH
MS11	6	Predicted stimulatory	ARVAAWWLAHGTSDS
MS11	7	Predicted stimulatory	VRDWYRWFGDTGDDY
MS11	8	Predicted stimulatory	YYCARQVYTFNWFNW
MS11	9	Predicted stimulatory	RQVYTFNWFNWFDPW
MS11	10	Predicted stimulatory	LLSVRKSWLSGWFDP
MS11	11	Predicted tolerogenic	KSKNQFSLKLTFVTA
MS11	12	Predicted tolerogenic	NQFSLKLTFVTAADT

MS11	13	Predicted tolerogenic	LISVTAADTAVYYCA
MS11	14	Predicted tolerogenic	AKSLLYLQMNSLRAE
MS11	15	Predicted tolerogenic	TAYLQWSSLKASDTA
MS10	16	Predicted inert	DTSKNEFSLKVTSVT
MS6	17	Predicted inert	FCTRVGDRRHYGGNS
MS6	18	Predicted inert	TADKSTRTAYMELSG
MS5	19	Predicted inert	DRSKNQFSLKLSSVT
MS4	20	Predicted inert	YCARDGRREQLVPNS
MS7	21	Predicted inert	YCARDNSNWTRGSGF
MS11	22	Predicted inert	DRSKNQFSLKVTSVT

Colors correspond to bars in Figure 3

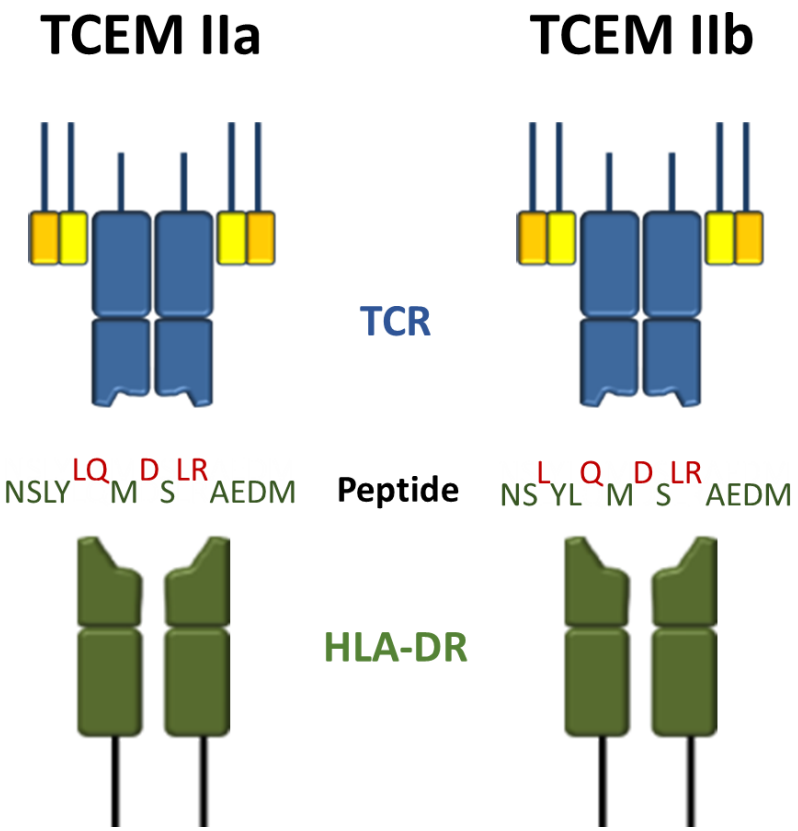
Supplementary Table 3. Surface markers and fluorochromes used on FACS Canto II

Laser	BP Filter	LP	Fluoro-chromes	Marker	Comment	Clone (cat. #)
407 nm (Violet)	450/50		**BV421**	**CD154**	T cell activation	TRAP1 (BD 563886)
488 nm (Blue)	585/42	556	**PE**	**CXCR3**	Chemokine rec.	11A9 (BD 560619)
	780/60	735	**PE-Cy7**	**CD45RO**	Memory T cells	UCHL1 (BD 337168)
		670 655	**PerCP-Cy5.5**	**CD4**	CD4 T cells	SK3 (BD 566316)
	530/30	502	**FITC**	**CD3**	T cells	UCHT1 (BD 555916)
633 nm (Red)	660/20		**APC**	**CCR6**	Chemokine rec.	1C6 (BD 550633)
	780/60	735	**APC-H7** **LIVE/DEAD™ Fixable Near-IR**	**CD14** **CD8**	Monocytes CD8 T cells Dead cells (All: dump channel)	MφP9 (BD 560180) SK1 (BD 560179) (L-34959)

BP: Bandpass; LP: longpass

Supplementary Figure 1. T cell exposed motifs IIa and IIb

Two T cell exposed motifs in context of peptide:HLA-DR binding, TCEM IIa and IIb, were deduced as described previously (1). TCEM IIa consists of amino acids 2,3,5,7,8 and TCEM IIb of -1,3,5,7,8 in a 9-mer core of 15-mers (-3,-2,-1,1,2,3,4,5,6,7,8,9,+1,+2,+3). The non-linear 5-mer motifs are the deduced sequences T cell receptors (TCR) may interact with, as the other amino acid residues remain hidden in the HLA-groove. There are theoretically 3.2 million (20^5) of each type, and their frequency of occurrence in immunoglobulin heavy chain variable regions may be calculated using a reverse logarithmic scale (Occurrences / $2^{n)}$, thereby designating a frequency class (FC=n) to each TCEM variant.

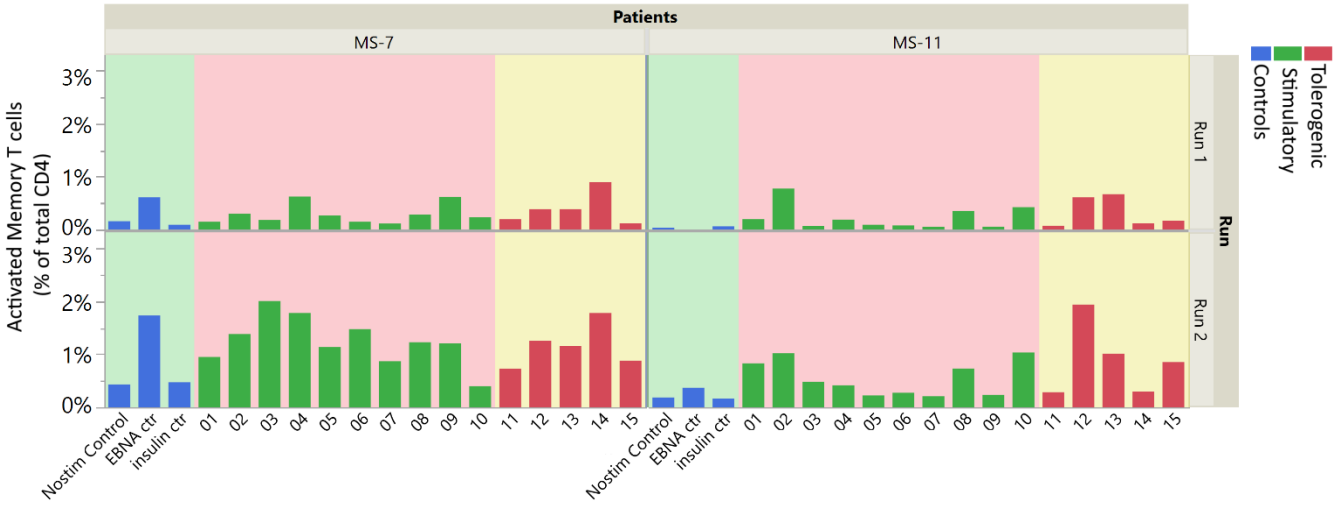


Reference

1. Bremel RD, Homan EJ. Frequency Patterns of T-Cell Exposed Amino Acid Motifs in Immunoglobulin Heavy Chain Peptides Presented by MHCs. Frontiers in immunology. 2014;5:541.

Supplementary Figure 2. Activated memory cells ($CD45RO^+CD154^+$) among $CD4^+$ T cells in replicate experiments

Responses towards idiotope peptides were assessed twice within the same lot of PBMC material in two patients. A total of 500,000 PBMCs in technical duplicates from the same lot were stimulated with EBNA-1 or insulin peptides mixes, CD3/CD28 activation beads, nothing or idiotope peptides for 16 (run 1) or 12 (run 2) hours in presence of anti-CD40 antibodies. $CD3^+CD4^+CD8^-$ T cells were identified with flow cytometry, and activation of specific cells assessed among $CD45RO^+$ memory cells using the marker CD154. Similar or even increased responses were observed during the second run. Run 2 is the same as shown in Figure 3 in the main manuscript. Background colors are for visual purposes only.



Paper IV

IV

Multiple Sclerosis and Related Disorders 26 (2018) 90–95

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

journal homepage: www.elsevier.com/locate/msard

B-cell composition in the blood and cerebrospinal fluid of multiple sclerosis patients treated with dimethyl fumarate

Rune A. Høglund[a,b,*], Justyna Polak[b,c], Frode Vartdal[c], Trygve Holmøy[a,b], Andreas Lossius[a,c,*]

[a] Department of Neurology, Akershus University Hospital, Lørenskog, Norway
[b] Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[c] Department of Immunology and Transfusion Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
Multiple sclerosis
B cell
Dimethyl fumarate
Memory B cell
Cerebrospinal fluid
Plasmablast

ABSTRACT

Background: B cells may contribute to the immunopathogenesis of multiple sclerosis (MS). Dimethyl fumarate (DMF) has recently been shown to reduce the frequency of memory B cells in blood, but it is not known whether the drug influences the cellular composition in the cerebrospinal fluid (CSF).
Methods: A cross-sectional study examining the cellular composition in blood and cerebrospinal fluid (CSF) from 10 patients treated with DMF and 18 patients receiving other disease modifying drugs or no treatment.
Results: Patients treated with DMF had reduced proportions of memory B cells in blood compared to other MS patients ($p = 0.0007$), and the reduction correlated with treatment duration ($r_s = -0.75$, $p = 0.021$). In the CSF, the absolute number of mononuclear cells were significantly lower in DMF-treated patients compared to the other patients ($p = 0.023$), and there was a disproportionate decrease of plasmablasts ($p = 0.031$).
Conclusion: The results of this exploratory study support a B-cell mediated mechanism of action for DMF in both blood and CSF.

1. Introduction

B cells have an important role in multiple sclerosis (MS), as is evident from several successful clinical trials of B-cell depletion (Hauser et al., 2017; Kappos et al., 2011; Bar-Or et al., 2008). Several potential mechanisms have been proposed, including antigen presentation, cytokine secretion, and antibody production (Lehmann-Horn et al., 2017). In MS, the proportion of immunoglobulin class-switched $CD27^+$ B cells is increased in the cerebrospinal fluid (CSF) compared to the peripheral blood (Eggers et al., 2017), and in particular the frequency of plasmablasts expressing CD38 and CD138 have been shown to be elevated (Cepok et al., 2005; Kowarik et al., 2014). Studies have demonstrated that immunoglobulin class-switched B cells in the CSF are related to B cells in the brain parenchyma (Obermeier et al., 2011), cervical lymph nodes (Stern et al., 2014), and peripheral blood (von Budingen et al., 2012; Johansen et al., 2015), indicating a dynamic exchange between these compartments.

Dimethyl fumarate (DMF) is an oral disease-modifying treatment (DMT) approved for treating relapsing-remitting MS (RRMS). The mechanism of action is not fully clarified (Deeks, 2016), but includes a reduction of peripheral blood $CD19^+CD27^+$ memory B cells (Smith et al., 2017; Longbrake et al., 2017; Li et al., 2017; Lundy et al., 2016). Upon intake, DMF is rapidly metabolized to monomethyl fumarate (MMF), which crosses the blood-brain barrier in animal models and in humans, however only at 15% of plasma concentration (Edwards et al., 2016; Penner et al., 2016).

The penetration of MMF into the CSF, and the exchange of B cells between the periphery, CSF and the central nervous system (CNS), suggest the possibility that DMF may exert a direct or indirect effect on intrathecal B cells. To address this, we examined blood and CSF B-cell subsets from MS patients treated with DMF and patients receiving other or no immunomodulatory treatment. We found that while DMF treatment is associated with a decrease of $CD19^+CD27^+$ memory B cells in the periphery, it is associated with a significant reduction of $CD19^+CD27^+CD38^+$ plasmablasts in CSF.

Abbreviations: CSF, cerebrospinal fluid; DMT, disease modifying therapy; IgG, immunoglobulin G; M/DMF, mono-/dimethyl fumarate; MFI, median fluorescence intensity; MS, multiple sclerosis; RRMS, relapsing remitting MS; TFM, teriflunomide

* Corresponding authors at: Department of Neurology, Akershus University Hospital, PO Box 1000, Lørenskog 1478, Norway.
E-mail addresses: r.a.hoglund@medisin.uio.no (R.A. Høglund), andreas.lossius@rr-research.no (A. Lossius).

https://doi.org/10.1016/j.msard.2018.08.032
Received 26 April 2018; Received in revised form 22 August 2018; Accepted 30 August 2018
2211-0348/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Demographic and clinical characteristics of patients.

Treatment	DMF	Other	Statistical test
Participants	10	18	–
Age (years)	41 (27–58)	47.5 (25–57)	$p = 0.75$[a]
Sex (M/F)	4/6	5/13	$p = 0.68$[b]
Disease duration (months)	47.5 (10–153)	102 (2–283)	$p = 0.10$[a]
Treatment duration (months)	8 (2–13)	All: 35 (3–152)[c] GA: 79 (18–152) TFM: 12.5 (3–27)	$p = 0.0012$[a]
EDSS score	1.25 (1–3.5)	1.25 (0–5.5)	$p = 0.78$[a]
Number of relapses	1.5 (1–5)	3 (1–6)	$p = 0.11$[a]
Time since last relapse (months)	18.5 (2–118)	5 (1–164)	$p = 0.60$[a]
Albumin ratio[d]	4.4 (3.6–9.24)	4.6 (2.3–6.9)	$p = 0.67$[a]
IgG index[d]	0.7 (0.5–2.1)	1.2 (0.5–1.9)	$p = 0.74$[a]

Numbers given as median (range). DMF: dimethyl fumarate. EDSS: expanded disability status scale. GA: glatiramer acetate. MRI: magnetic resonance imaging. TFM: teriflunomide.
[a] Wilcoxon 2-sample exact test.
[b] Fishers exact test.
[c] untreated patients excluded.
[d] compared between DMF treated and no-treated.

2. Materials and methods

2.1. Patients

MS patients ($n = 28$) were recruited at the Departments of Neurology at Akershus University Hospital and Oslo University Hospital; other data based on the same cohort have been previously published (Lossius et al., 2017). All patients met the 2010 McDonalds criteria for MS (Polman et al., 2011); 1 patient was classified as secondary progressive MS and the remaining as RRMS. All patients had intrathecal synthesis of immunoglobulin G (IgG) at the time of diagnosis, detected as oligoclonal IgG bands by isoelectric focusing. Lumbar puncture was performed either as a part or outside of diagnostic investigations, reflected by the variation in disease duration (Table 1). The study was approved by the Regional Ethical Committee South East (2009/23 S-04143a). All patients gave written informed consent before inclusion.

2.2. Cell subset analysis

For each individual included in this study, blood, serum and CSF samples were collected once, during the same consultation. Blood and CSF mononuclear cells were isolated as described previously, counted using a Neubauer improved hemocytometer, and immediately processed for flow cytometry, which was performed within two hours after sample collection (Johansen et al., 2015; Lossius et al., 2017). For B-cell subset identification we used fluorochrome-conjugated antibodies as described (Lossius et al., 2017), including anti-human CD19, CD27, CD38, CD138, IgG, Ki-67, and HLA-DR, and additionally CD3 and CD14 for dump-channel purposes (BD Biosciences). Paired cell samples from CSF and blood were stained for surface antigens, and after fixation and permeabilization following the manufacturer's instructions (Fixation/Permeabilization Solution Kit, BD Biosciences), the cells were stained for Ki-67. The flow cytometry dataset was reanalyzed for the present study using FlowJo Version 10.2 (FlowJo, LLC). Of note, flow cytometry data from one DMF-treated and one untreated patient were excluded due to technical reasons.

2.3. Quantification of IgG and albumin

For each patient, blood and CSF were collected simultaneously. After centrifugation and removal of cells for immediate flow cytometry analysis, the serum and CSF supernatants were frozen at −80 °C. The supernatants were thawed and analyzed as a single batch within 36 months after the collection and freezing of the first patient sample. The collection and freezing procedures followed the published consensus protocol for the standardization of cerebrospinal fluid collection and biobanking (Teunissen et al., 2009). IgG and albumin in CSF were quantified by nephelometry (BN ProSpec Systems, Siemens). IgG and albumin in serum were determined by turbidimetry and colorimetry, respectively (Vitros 5.1 FS, Ortho Clinical Diagnostics).

2.4. Statistics

Statistical analyses were performed in JMP® pro 12.1 (StataCorp, LLC). Due to non-normality of data, non-parametric 2-sample Wilcoxon exact tests were used to compare groups unless otherwise specified. The significance level was set at 5%, and the tests were two-sided. No correction for multiple testing was performed. Results are presented as median [range]. Figures were made in FlowJo and JMP pro 12.1.

3. Results

3.1. Patients

Ten patients received treatment with DMF at inclusion, whereas 18 received other DMTs or no treatment (7 patients received glatiramer acetate, 4 received teriflunomide, and 7 received no treatment). Clinical and demographic data at the time of CSF and blood (including serum) collection are shown in Table 1. With the exception of treatment durations, which was shorter for the patients treated with DMF than other DMTs, there were no significant differences in clinical characteristics between DMF-treated patients and the other patients . All DMF-treated patients had normal total lymphocyte count before treatment start, with a median value of 1.7×10^9/L [1.2–2.6], and 2 patients developed lymphopenia during DMF treatment ($< 0.9 \times 10^9$/L).

3.2. DMF-treatment is associated with lower numbers of mononuclear cells in the CSF

DMF-treated patients had lower total CSF mononuclear cell counts compared to the other patients (560/mL [80–3000] vs 680/mL [264–6736], $p = 0.023$). Similar results were obtained when comparing DMF-treated patients to those not receiving any treatment (Fig. S1 A). The CSF mononuclear cell count in DMF-treated patients tended to correlate inversely with the treatment duration, but this was not statistically significant (Spearman r_s −0.6, $p = 0.07$).

3.3. DMF reduces the frequency of memory B cells in blood and plasmablasts in the CSF

Previous studies have shown that DMF reduces the number of memory B cells in blood (Smith et al., 2017; Longbrake et al., 2017; Li et al., 2017; Lundy et al., 2016), typically identified as $CD19^+CD27^+$ cells. To identify these cells in blood and CSF, a gating strategy based on available data was devised (Fig. 1). Confirming previous studies, we found that DMF-treatment was associated with a relative reduction in $CD19^+CD27^+CD38^-$ memory B cells in blood (12.8% [2.2–25.2] in DMF-treated vs 26.6% [6.9–46.6] in the other MS patients, $p = 0.0007$, Fig. 2A). Moreover, the reduction correlated strongly with treatment duration ($r_s = -0.75$, $p = 0.021$, Fig. 2B). There were no significant differences between the two groups in the proportions of $CD19^+CD27^+CD38^+$plasmablasts in blood (Fig. 2A). In the CSF, in contrast, there was a selective depletion of $CD19^+CD27^+CD38^+$ plasmablasts among DMF-treated patients (0% [0–13.6] vs 5.4% [0–69.8], $p = 0.031$, Fig. 2A), whereas the relative proportion of $CD19^+CD27^+CD38^-$ memory B cells was similar in both groups

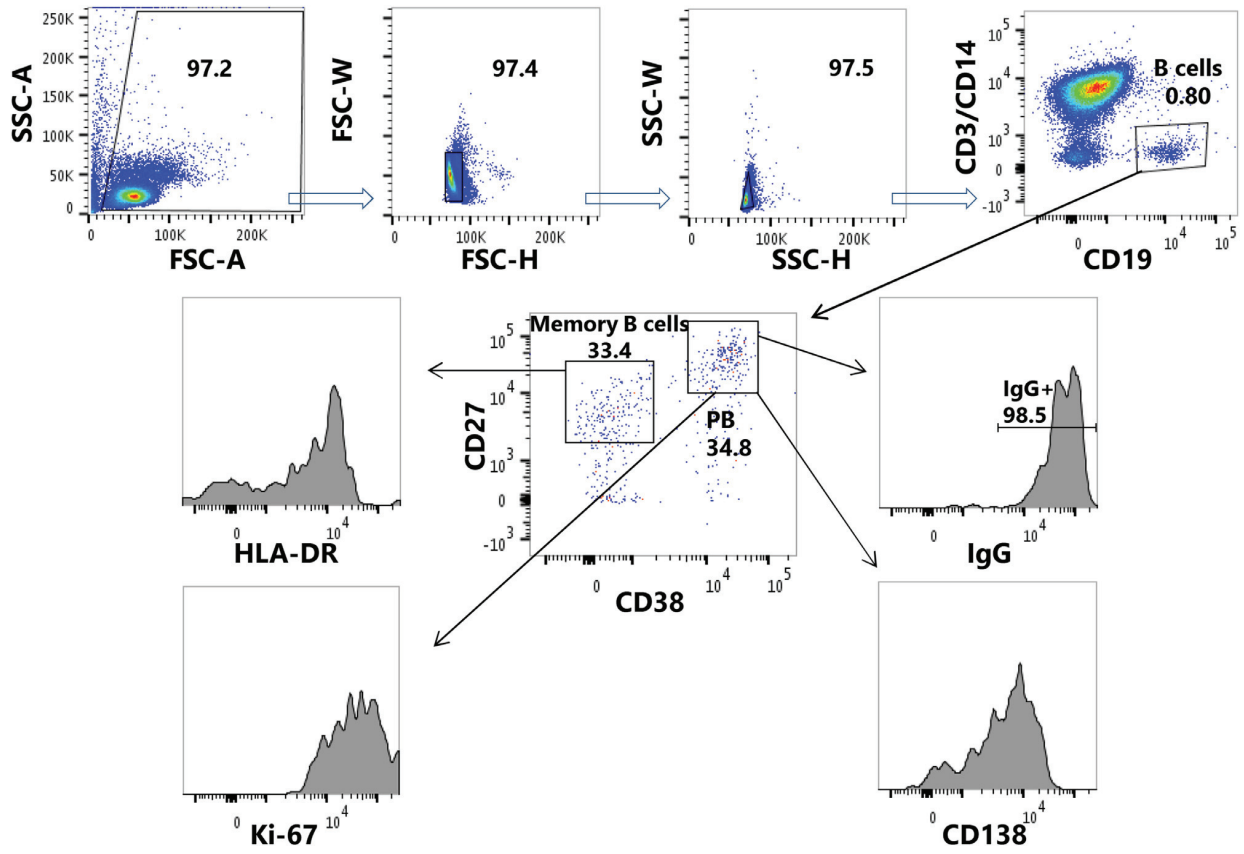


Fig. 1. Gating strategy. We devised a gating strategy to identify single lymphocytes expressing CD19, CD27, CD38, CD138, HLA-DR, IgG and/or Ki-67. Cells expressing CD3 or CD14 were excluded in a dump channel.

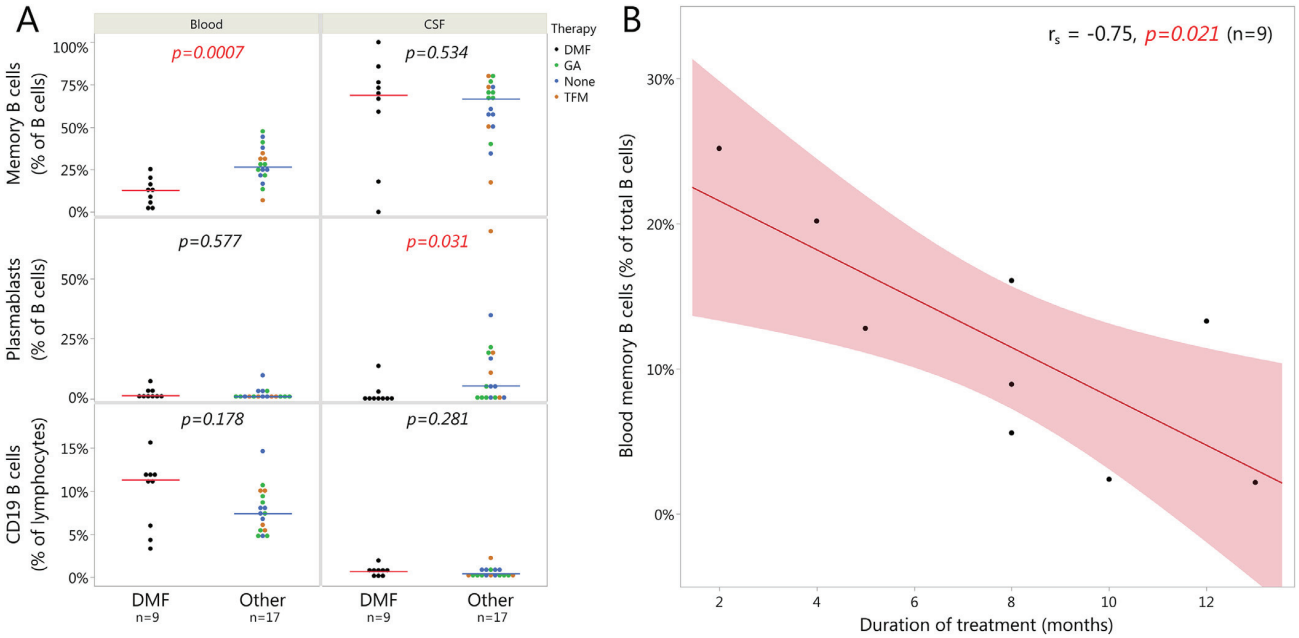


Fig. 2. DMF preferentially reduces the frequency of memory B cells in blood and plasmablasts in the CSF. (A) Proportion of $CD19^+CD27^+CD38^-$ memory B cells and $CD19^+CD27^+CD38^+$ plasmablasts of $CD19^+$ B cells, and proportion of $CD19^+$ B cells of lymphocytes in blood and CSF. Each symbol represents the cell frequencies in a given individual, and the bars depict the median. (B) Proportion of memory B cells in DMF treated patients, in linear fit with treatment duration (boundaries depict 95% confidence interval). DMF, dimethyl fumarate; GA, glatiramer acetate; TFM, teriflunomide.

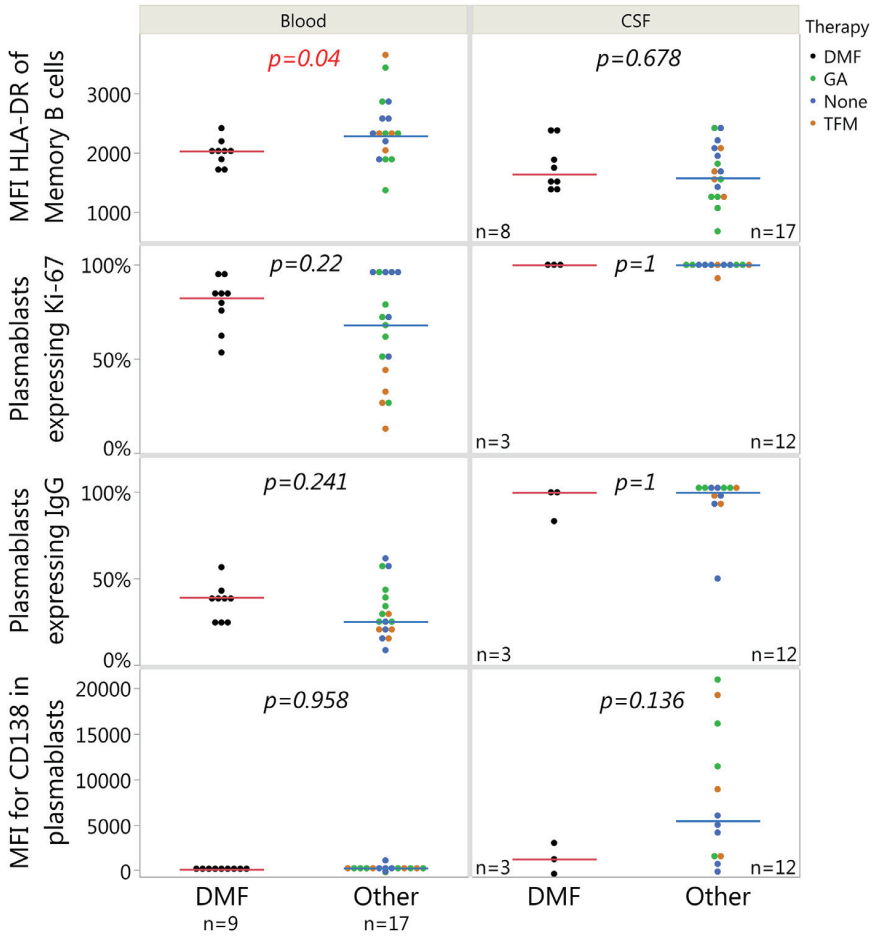


Fig. 3. DMF reduces HLA-DR expression among memory B cells in blood. MFI of HLA-DR for $CD19^+CD27^+CD38^-$ memory B cells, proportion of $CD19^+CD27^+CD38^+$ plasmablasts expressing Ki-67 and IgG, and MFI of CD138 for plasmablasts. Each symbol represents the cell frequencies or MFI in a given individual, and the bars depict the median. DMF, dimethyl fumarate; GA, glatiramer acetate; TFM, teriflunomide.

(68.9% [0–100] in DMF-treated vs 66.7% [17.4–80] in other MS-patients, $p = 0.534$, Fig. 2A). There were no differences between the two groups in the proportion of total $CD19^+$ B cells in blood or CSF (Fig. 2A). Including only DMF-treated and untreated patients in the analyses yielded similar results (Fig. S1 B).

3.4. DMF reduces HLA-DR expression among memory B cells in blood

To further characterize the B-cell subsets, we examined antigen-presenting potential (HLA-DR), IgG expression, proliferation status (Ki-67), and the expression of the differentiation marker CD138. In patients treated with DMF, we found that blood $CD19^+CD27^+CD38^-$ memory B cells express lower levels of the HLA class II molecule HLA-DR (median fluorescence intensity (MFI) 2025 [1702–2410] vs 2279 [1360–3649], $p = 0.04$, Fig. 3). The majority of blood plasmablasts expressed Ki-67 and IgG, and there were no significant differences between the patient groups (Fig. 3). We found overall low levels of CD138 expression among blood plasmablasts.

In the CSF, the HLA-DR expression among $CD19^+CD27^+CD38^-$ memory B cells was similar in both groups (MFI 1634 [1337–2391] vs 1570 [666–2414], $p = 0.68$). Virtually all intrathecal $CD19^+CD27^+CD38^+$ plasmablasts in both patient groups expressed Ki-67 and the majority expressed IgG (100% [83.3–100] vs 100% [50–100], Fig. 3). In most patients in both groups we found that CSF plasmablasts expressed moderate to high levels of CD138 (MFI 1223 [−392–2998] vs MFI 5437 [−142–20,911]). There were no significant differences for these variables between the patient groups (Fig. 3).

3.5. DMF treatment did not influence intrathecal IgG production

Since DMF-treated patients exhibited preferential reductions of $CD19^+CD27^+CD38^+$ antibody-secreting cells in CSF, we compared the intrathecal IgG production among DMF-treated patients ($n = 10$) with that among untreated patients ($n = 7$). There were no differences in IgG or albumin levels or indices of these, also indicating a similar blood–brain barrier integrity (Table 1).

4. Discussion

This cross-sectional study demonstrates that while DMF-treatment causes a time-dependent reduction in $CD19^+CD27^+CD38^-$ memory B cells in blood, it leads to a depletion of $CD19^+CD27^+CD38^+$ plasmablasts in the CSF. Our findings in blood are in agreement with recent publications (Smith et al., 2017; Longbrake et al., 2017; Li et al., 2017; Lundy et al., 2016), and the present work demonstrates how the peripheral effects of DMF translate into the CSF. These data point to a possible B-cell mediated mechanism of action for DMF, in addition to the previously discussed pluripotent immunomodulatory effects (Deeks, 2016).

The almost immediate effect of anti-CD20 treatment in MS is thought to be mediated by a reduction of B cells participating in antigen presentation and/or cytokine production (Funaro et al., 2016; Kinzel

and Weber, 2016). Accordingly, the intrathecal IgG production does not seem to be affected by the treatment in the short run (Hauser et al., 2017; Cross et al., 2006). Memory B cells express HLA class II molecules and are potent antigen-presenting cells (Kurosaki et al., 2015). Interestingly, we found that blood $CD19^+CD27^+CD38^-$memory B cells express lower levels of HLA-DR in DMF-treated than in untreated MS patients, suggesting that memory B cells in these patients are less efficient in presenting antigens. One could speculate that a reduced amount of memory B cells, with additionally reduced expression of HLA class II molecules, may result in less maturation and development of $CD19^+CD27^+CD38^+$antibody-secreting cells in the CSF. This possibility is supported by the dynamic exchange of B cells between the CSF and periphery (Obermeier et al., 2011; Stern et al., 2014; von Budingen et al., 2012; Johansen et al., 2015). However, a not mutually exclusive possibility is that the observed alterations in the composition of CSF B cells could be due to a local intrathecal effect of MMF.

Despite depletion of $CD19^+CD27^+CD38^+$ plasmablasts in the CSF of DMF-treated patients, we did not observe a reduced intrathecal IgG production. This could be due to persistent secretion from end-differentiated plasma cells protected within tertiary lymphoid tissue in the meninges of MS patients (Pikor et al., 2016). In line with this thought, we found that all antibody-secreting cells in the CSF express high levels of Ki-67, compatible with recently derived proliferating plasmablast. Interestingly, a similar persistence of intrathecal IgG synthesis has been observed in patients receiving anti-CD20 therapies (Hauser, 2015). In accordance with the present findings, no difference in circulating IgG, IgA, or IgM in blood of DMF treated patients has been observed, indicating that end differentiated plasma cells could be protected from the effects of the drug also in the periphery in a short-term perspective (Longbrake et al., 2017).

The strength of this study is the comparison of CSF B cell subsets between two groups of MS patients similar in clinical and demographic characteristics. This was made possible by slight different treatment strategies between the two clinical departments contributing to the study. Samples from both hospitals were transported directly to the same external laboratory and processed identically, and the findings are not influenced by differences in inflammatory activity. Although we cannot exclude that differences in treatment durations could have influenced the results, treatment durations were long enough to allow full immunological effect in the majority of patients receiving DMF and other DMTs.

This study has several limitations. First, the number of patients is small. Given the risk of lumbar puncture headache, as well as rare but more serious risks such as infection, bleeding and herniation, we cannot perform lumbar puncture solely for the purpose of studying drug effects. While properly understanding the mechanism of action of DMF in MS patients is important, such studies must due to the above reasons be part of studies on disease mechanisms, which was the primary aim of this study (Lossius et al., 2017). This is therefore an exploratory study, the *p*-values were not corrected for multiple testing, and the results need to be reproduced in an independent study. Second, the control group is heterogeneous from a treatment standpoint, which potentially could confound our results. Nevertheless, as the sub-analysis with untreated controls ($n = 7$) demonstrated similar results, and lymphocyte depletion and reduced memory B cell population in blood are analogue to previous reports (Smith et al., 2017; Longbrake et al., 2017; Li et al., 2017; Lundy et al., 2016), we believe our findings have merit. While reduction in blood plasmablasts similar to what we observed in CSF have not been extensively described, a recent interim analysis of the PROCLAIM study demonstrated a decline in blood plasmablasts that had a delayed onset compared to memory B cell depletion (von Hehn et al., 2017). This delay, in addition to the relatively low proportions of circulating plasmablasts in blood compared to CSF of MS patients can potentially have hidden this phenomenon in studies with shorter timeframes. Lastly, as our data did not contain markers for IgD, we were unable to differentiate between class-switched and unswitched memory B cells. A recent publication limited to blood found that both populations are affected similarly, with a corresponding increase in $CD27^-IgD^+$ naïve B cells (Nakhaei-Nejad et al., 2018). New studies as to how these subpopulations are distributed in the CSF could also enlighten whether the effects of DMF are mainly peripheral or intrathecal.

5. Conclusion

This is the first study on the effects of DMF treatment on B cells in CSF, suggesting that while memory B cells are reduced in blood, plasmablasts are preferentially depleted in CSF. These findings, which need to be verified in an independent study, support the hypothesis that DMF mediates positive effects through B-cell modulation.

Funding

The study was supported by the South-Eastern Norway Regional Health Authority (grant 2016079) and the Norwegian Research Council (grant 250864/F20).

Conflict of interests

RH has received speaker honoraria from Merck and Roche, and unrestricted research grants from Biogen and Novartis. TH has received unrestricted research grants or speaker honoraria from Biogen, Merck, Novartis, Sanofi Genzyme and Teva. AL has received speaker honoraria from Roche, and unrestricted research grants from Sanofi Genzyme. JP and FV declare no conflict of interest.

Author contributions

Study concept and design: A.L., F.V., R.H., T.H.; data acquisition and/or analysis: A.L., J.P., R.H.; drafting the manuscript A.L., R.H.. All authors revised the manuscript for intellectual content.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2018.08.032.

References

Bar-Or, A., Calabresi, P.A., Arnold, D., et al., 2008. Rituximab in relapsing-remitting multiple sclerosis: a 72-week, open-label, phase I trial. Ann. Neurol. 63, 395–400. https://doi.org/10.1002/ana.21363.

von Budingen, H.C., Kuo, T.C., Sirota, M., et al., 2012. B cell exchange across the blood-brain barrier in multiple sclerosis. J. Clin. Invest. 122, 4533–4543. https://doi.org/10.1172/jci63842.

Cepok, S., Rosche, B., Grummel, V., et al., 2005. Short-lived plasma blasts are the main B cell effector subset during the course of multiple sclerosis. Brain J. Neurol.y 128, 1667–1676. https://doi.org/10.1093/brain/awh486.

Cross, A.H., Stark, J.L., Lauber, J., et al., 2006. Rituximab reduces B cells and T cells in cerebrospinal fluid of multiple sclerosis patients. J. Neuroimmunol. 180, 63–70. https://doi.org/10.1016/j.jneuroim.2006.06.029.

Deeks, E.D., 2016. Dimethyl fumarate: a review in relapsing-remitting MS. Drugs 76, 243–254. https://doi.org/10.1007/s40265-015-0528-1.

Edwards, K., Penner, N., Rogge, M., et al., 2016. A pharmacokinetic study of delayed-release dimethyl fumarate to evaluate cerebrospinal fluid penetration in patients with secondary progressive multiple sclerosis. ePosters EP1501. Mult. Scler.s J. 22 (3_suppl), 706–827. https://doi.org/10.1177/1352458516663067.

Eggers, E.L., Michel, B.A., Wu, H., et al., 2017Eggers. Clonal relationships of CSF B cells in treatment-naive multiple sclerosis patients. JCI Insight 2 (22), e92724. https://doi.org/10.1172/jci.insight.92724.

Funaro, M., Messina, M., Shabbir, M., et al., 2016. The role of B cells in multiple sclerosis: more than antibodies. Discov. med. 22, 251–255.

Hauser, S.L., Bar-Or, A., Comi, G., et al., 2017. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. N. Engl. J. Med. 376, 221–234.

Hauser, S.L., 2015. The Charcot lecture | beating MS: a story of B cells, with twists and turns. Mult. Scler. 21, 8–21. https://doi.org/10.1177/1352458514561911.

von Hehn, C., Mehta, D., Prada, C., et al., 2017. Interim results of an open-label study to assess the effects of delayed-release dimethyl fumarate on lymphocyte subsets and immunoglobulins in patients with relapsing-remitting multiple sclerosis. In: Poster 380 presented at American Academy of Neurology - 69th Annual Meeting, Session P5,

MS and CNS Inflammatory Disease, April 27. Accessed online May 25, 2018 at http://submissions.mirasmart.com/Verify/AAN2017/Submission/Temp/radFD3DA.pdf
Johansen, J.N., Vartdal, F., Desmarais, C., et al., 2015. Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. Clin. immunol. 160, 211–225. https://doi.org/10.1016/j.clim.2015.06.001.
Kappos, L., Li, D., Calabresi, P.A., et al., 2011. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. Lancet 378, 1779–1787. https://doi.org/10.1016/s0140-6736(11)61649-8.
Kinzel, S., Weber, M.S., 2016. B Cell-directed therapeutics in multiple sclerosis: rationale and clinical evidence. CNS Drugs 30, 1137–1148. https://doi.org/10.1007/s40263-016-0396-6.
Kowarik, M.C., Grummel, V., Wemlinger, S., et al., 2014. Immune cell subtyping in the cerebrospinal fluid of patients with neurological diseases. J. Neurol. 261, 130–143. https://doi.org/10.1007/s00415-013-7145-2.
Kurosaki, T., Kometani, K., Ise, W., 2015. Memory B cells. Nat. Rev. Immunol. 15, 149–159. 2015/02/14, doi:10.1038/nri3802.
Lehmann-Horn, K., Kinzel, S., Weber, M., 2017. Deciphering the Role of B Cells in multiple sclerosis—towards specific targeting of pathogenic function. Int. J. Mol. Sci. 18, 2048.
Li, R., Rezk, A., Ghadiri, M., et al., 2017. Dimethyl fumarate treatment mediates an anti-inflammatory shift in Bcell subsets of patients with multiple sclerosis. J. Immunol. 198, 691–698. https://doi.org/10.4049/jimmunol.1601649.
Longbrake, E.E., Cantoni, C., Chahin, S., et al., 2017. Dimethyl fumarate induces changes in B- and T-lymphocyte function independent of the effects on absolute lymphocyte count. Mult. Scler. 1352458517707069, doi:10.1177/1352458517707069.
Lossius, A., Tomescu-Baciu, A., Holmoy, T., et al., 2017. Selective intrathecal enrichment of G1m1-positive B cells in multiple sclerosis. Ann. Clin. Transl.l Neurol. 4, 756–761. https://doi.org/10.1002/acn3.451.
Lundy, S.K., Wu, Q., Wang, Q., et al., 2016. Dimethyl fumarate treatment of relapsing-remitting multiple sclerosis influences B-cell subsets. Neurol.(R) Neuroimmunol. Neuroinflamm. 3, e211. https://doi.org/10.1212/nxi.0000000000000211.
Nakhaei-Nejad, M., Barilla, D., Lee, C.-H., et al., 2018. Characterization of lymphopenia in patients with MS treated with dimethyl fumarate and fingolimod. Neurol. – Neuroimmunol. Neuroinflamm. 5. https://doi.org/10.1212/nxi.0000000000000432.
Obermeier, B., Lovato, L., Mentele, R., et al., 2011. Related B cell clones that populate the CSF and CNS of patients with multiple sclerosis produce CSF immunoglobulin. J. Neuroimmunol. 233, 245–248. https://doi.org/10.1016/j.jneuroim.2011.01.010.
Penner, N., Zhu, B., Woodward, C., et al., 2016. Penetration of dimethyl fumarate into cerebrospinal fluid and brain: a pharmacokinetic and tissue distribution study in monkeys. Poster Session 1, P672. Mult. Scler. J.l 22 (3_suppl), 88–399. 277, doi:10.1177/1352458516663081.
Pikor, N.B., Prat, A., Bar-Or, A., et al., 2016. Meningeal tertiary lymphoid tissues and multiple sclerosis: a gathering place for diverse types of immune cells during CNS autoimmunity. Front. immunol. 6. Mini Review, doi:10.3389/fimmu.2015.00657
Polman, C.H., Reingold, S.C., Banwell, B., et al., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology 69, 292–302. 2011/03/10, doi:10.1002/ana.22366
Smith, M.D., Martin, K.A., Calabresi, P.A., et al., 2017. Dimethyl fumarate alters B-cell memory and cytokine production in MS patients. Ann. clin. transl neurol. 4, 351–355. https://doi.org/10.1002/acn3.411.
Stern, J.N., Yaari, G., Vander Heiden, J.A., et al., 2014. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. Sci. Transl. Med. 6, 248ra107. https://doi.org/10.1126/scitranslmed.3008879.
Teunissen, C.E., Petzold, A., Bennett, J.L., et al., 2009. A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking. Neurology 73 (22), 1914–1922. https://doi.org/10.1212/WNL.0b013e3181c47cc2.