Potential of a unique antibody gene signature to predict conversion to clinically definite Multiple Sclerosis

Elizabeth M. Cameron, Sade Spencer, Jonathan Lazarini, Christopher T. Harp, E. Sally Ward, Mark Burgoon, Gregory P. Owens, Michael K. Racke, Jeffrey L. Bennett, Elliot M. Frohman, and Nancy L. Monson

Abstract

We identified a unique antibody gene mutation pattern (i.e., “signature”) in CSF B cells from MS patients not present in control populations. Prevalence of the signature in CSF B cells of patients at risk to develop MS predicted conversion to MS with 91% accuracy in a small cohort of clinically isolated syndrome patients. If confirmed, signature prevalence would be a novel genetic diagnostic tool candidate for patients with early demyelinating disease of the central nervous system.

Keywords

Multiple sclerosis; B lymphocytes; Antibodies; Gene rearrangement; Mutational signature

1. Introduction

B cells have historically been implicated in the pathogenesis of multiple sclerosis (MS) since elevated central nervous system (CNS) gamma immunoglobulins were first described in MS patients in the 1940s (Kabat et al., 1948). Additional evidence of B cell involvement in MS pathogenesis is extensive (reviewed in Antel and Bar-Or, 2006; Duddy and Bar-Or, 2006; Franciotta et al., 2008; Frohman et al., 2006; Martin Medel and Monson, 2007; McLaughlin and Wucherpfennig, 2008; Owens et al., 2006)) and has been recently substantiated by the efficacy of rituximab (Rituxan), a B cell depleting antibody, in a cohort of patients with relapsing remitting MS (RRMS) (Hauser et al., 2008). Furthermore, Rituxan and intravenous immunoglobulin, drugs that solely affect B cells or their antibody products, have been reported to decrease severity of disease in MS patients refractory to benefit with corticosteroids, interferon-beta, and mitoxantrone (Achiron, 2008; Leussink et al., 2008; Stuve et al., 2005; Tselis et al., 2008).
Several groups investigating the role of B cells in MS have hypothesized that the distribution of genes used to generate antibodies in B cells from the cerebrospinal fluid (CSF) and brain lesions of MS patients are different from expected distributions. Indeed, the distributions are different in some cases, particularly with regard to a family of variable heavy chains (VH4), which are significantly increased in frequency compared to expected distributions (Baranzini et al., 1999; Colombo et al., 2000; Harp et al., 2007; Monson et al., 2005; Owens et al., 1998; Owens et al., 2003; Owens et al., 2007; Qin et al., 1998; Ritchie et al., 2004). Additionally, MS CSF B cells show extensive clonal expansion and high mutational frequencies in the CSF B cell pool from this population of patients (Baranzini et al., 1999; Colombo et al., 2000; Monson et al., 2005; Owens et al., 2003; Qin et al., 1998; Ritchie et al., 2004), and the antibodies these cells produce bind to neuroantigens (Kolln et al., 2006; Lambracht-Washington et al., 2007). In contrast, VH4 expressing B cells in the periphery of healthy donors (Brezinschek et al., 1995; Brezinschek et al., 1997), MS patients (Owens et al., 2007), and VH4 expressing B cells in the CSF of patients with other neurological diseases (OND) are present at expected frequencies (Table 1 and (Harp et al., 2007)). Since antibody gene mutation patterns are influenced by antigen driven selection, we hypothesized that VH4 expressing CSF-derived B cells of MS patients would harbor antibody gene mutation patterns that would be distinct from VH4 expressing peripheral B cells derived from healthy controls. To address this contention, we characterized antibody gene mutations in a VH4 subdatabase extracted from the parent heavy chain antibody database consisting of 373 CSF-derived B cells from 11 patients with definite MS. Our analysis revealed a unique pattern of antibody gene replacement mutations in CSF B cells from MS patients that was not prevalent in antibody gene repertoires from CSF B cells of OND patients. Furthermore, prevalence of this conspicuous signature in B cell antibody repertoires from patients with a first inflammatory demyelinating episode (a clinically isolated syndrome; CIS) can predict conversion to clinically definite MS (CDMS) within 3–18 months after initial sampling.

2. Materials and methods

2.1 Patient description

CSF was collected from ten RRMS patients, one PPMS patient (M484), three patients with other neurological diseases (OND341, ataxia; OND758, headache, and OND116, chronic inflammatory demyelinating polyneuropathy), and two patients with one demyelinating event suggestive of MS (i.e. Clinically Isolated Syndrome (CIS)) at UT Southwestern Medical Center (UTSWMC) (Harp et al., 2007; Monson et al., 2005) in accordance with the UTSWMC Institutional Review Board (IRB). CSF was collected from nine patients with CIS at University of Colorado Denver (UCD) as previously described (Bennett et al., 2008) in accordance with the UCD IRB. The CIS patients had a single episode of demyelination (optic neuritis, brainstem or spinal cord syndrome), and the majority had multiple lesions on MRI satisfying the dissemination in space criterion of the McDonald criteria. None of the patients had received immunomodulatory agents for at least one month prior to lumbar puncture. A second relapse confirming a multiple sclerosis diagnosis had not occurred at the time of sample acquisition, thus not fulfilling the dissemination in time criterion (McDonald et al., 2001; Polman et al., 2005). Subsequent diagnosis of definite MS was made using the revised McDonald criteria (Polman et al., 2005). Conversion to definite MS was not revealed to the antibody sequence analysis team until after signature score predictions had been calculated.

2.2 MS and CIS B Cell Antibody Database Generation

At UTSWMC, antibody repertoires were generated from CD19+ CSF B cells using single cell PCR as previously described (Harp et al., 2007; Monson et al., 2005). The MSCSF database consists of antibody rearrangements from 373 CD19+ CSF B cells from 10 RRMS and 1 PPMS...
patient recruited at UT SWMC. The CISC SF database consists of antibody rearrangements from 304 CD19+ CSF B cells from 10 CIS patients (ON4–8 did not have a CD19+ CSF B cell antibody repertoire) and 228 CD138+ CSF plasma cells from 7 CIS patients (CIS132, CIS429, ON4–10 and ON3–4 did not have CD138+ CSF plasma cell antibody repertoires). To clarify, antibody repertoires from CIS patients at UCD were generated from both single CD19+ CSF B cells and single CD138+ CSF plasma cells (Bennett et al., 2008), while antibody repertoires from CIS patients at UT SWMC were generated from single CD19+ CSF B cells only. Since the resultant databases (CIS CD19+ CSF from UT SWMC, CIS CD19+ and CIS CD138+ CSF from UCD) were similar in mutational frequency, variable heavy chain (VH) gene family usage, and heavy chain Joining segment (JH) usage, the two databases were combined for analysis (Supplemental Table 1).

2.3 Control B Cell Antibody Database Generation

The healthy control peripheral blood (HCPB) antibody database has been used in multiple studies (Brezinschek et al., 1997; Brezinschek et al., 1998; Dorner et al., 1997; Dorner et al., 1998a; Dorner et al., 1998b; Dorner et al., 1998c; Farner et al., 1999; Hansen et al., 2000; Harp et al., 2007; Monson et al., 2005; Monson et al., 2000) and consists of 348 CD19+ or CD19+IgM+ peripheral B cells from two healthy control donors. The memory HCPB antibody database (mHCPB) consists of 205 sequences from the HCPB antibody database that contain 4 or more mutations (less than 98% homology to the germline sequence, n=123) combined with sequences from a HCPB antibody database generated from class-switched IgD−CD27+ memory B cells (n=82) (Tian et al., 2007) (Genbank 535266–535274, 535324–535368, 535381–535408, and 535416–535418). As expected, the class-switched IgD-CD27+ memory B cell database had a higher percentage of mutated codons that resulted in a replacement than the mHCPB database (compare 64.7% vs 70.3%, p=0.002 by χ² test). The OND CD19+ CSF antibody database consists of 65 sequences. UCD and UT SWMC cell isolation and IgH amplification was performed similarly. All sequences were re-confirmed by our laboratory using IgBlast (those obtained from UCD and from GenBank) (http://www.ncbi.nlm.nih.gov/igblast/), and only codons 24–93 were considered in the analysis.

2.4 Mutation analyses

Frequency of replacement mutations (RF) was calculated as the number of replacement mutations at each codon position divided by the total number of replacement mutations in each VH4 subdatabase and displayed as a percentage. The MSCSF database contains 373 sequences with 2,475 replacement mutations, and the CISC SF database contains 302 CD19+ and 226 CD138+ sequences with 4081 replacement mutations (2,052 in CD19+ and 2,029 in CD138+). The HCPB database contains 348 sequences with 1,086 replacement mutations, and the mHCPB database contains 205 sequences with 1,857 replacement mutations. The OND CSF database contains 65 sequences with 482 replacement mutations, and the MSPB database contains 156 sequences with 392 replacement mutations. In total, 1,675 sequences and 10,373 replacement mutations were analyzed in this manner. Table 1 contains VH4 sequence numbers and Table 2 legend contains number of VH4 replacement mutations. Codon domains and numbers were defined by Kabat (Kabat et al., 1983), and Tomlinson in V-base (vbase.mrc-cpe.cam.ac.uk/), respectively.

2.5 Statistical strategy for signature identification

Codons included in the signature were identified using three criteria. First, we identified codons that had statistically different RF values in the MSCSFVH4 database compared to HCPBVH4 by Goodness of Fit test where the expected frequency is the RF calculated in HCPBVH4. Twenty-four codons passed this criterion. Next, codon positions that had an RF in both the
MSCSF\textsubscript{VH4} and HCPB\textsubscript{VH4} databases that was less than the average $\pm$ 2 S.D. of the memory HCPB\textsubscript{VH4} subdatabase were excluded. Thus, since the average $\pm$ S.D. RF of the memory HCPB\textsubscript{VH4} database was 0.68 $\pm$ 0.59, any individual codon RF less than 1.86 in both databases was excluded. Fourteen codons passed this additional criterion. Eight of these 14 codons (31B, 32, 40, 56, 57, 60, 81, and 89) were defined as “hot” since the RF at that codon position within the MSCSF\textsubscript{VH4} database was statistically higher compared to the HCPB\textsubscript{VH4} database. Six of these 14 codons (30, 43, 52, 77, 82 and 82a) were defined as “cold” since the RF at that codon position within the MSCSF\textsubscript{VH4} database was statistically less compared to the HCPB\textsubscript{VH4} database. Two of the 6 “cold” codons (52 and 82a) were excluded because the RF value in the MSCSF\textsubscript{VH4} database at that codon position was significantly higher than 1.86 (the average $\pm$ 2 S.D. of the memory HCPB\textsubscript{VH4} subdatabase). The overall signature consequently consisted of codons 30, 31B, 32, 40, 56, 57, 60, 77, 81, 82, and 89. This analysis was not biased by differences in the prevalence of particular codons (31B in particular), as individual VH4 gene frequencies in MSCSF were similar to HCPB by $\chi^2$ test using a Bonferroni corrected p-value of 0.004 (data not shown).

2.5 Statistical computation of the signature score

Signature scores were generated by calculating Z-scores for the RF values at the 6 codons within the signature (31B, 40, 56, 57, 81 and 89) that had the most significant difference in RF compared to HCPB\textsubscript{VH4} at each codon position. The Z-score formula is: (RF at codon X minus the average RF in HCPB\textsubscript{VH4}) divided by the standard deviation of the average RF in HCPB\textsubscript{VH4}. For example, the average RF in HCPB\textsubscript{VH4} within the 6 signature codons was 1.6 $\pm$ 0.9 and so an RF of 4.4 at codon 31B would be assigned a score of 3.1 ($Z$ score=$\frac{(4.4-1.6)}{0.9}$). Individual z-scores at each of the 6 codon positions were then added to generate the composite signature z-score. The average composite signature score in the MSCSF\textsubscript{VH4} database was 10.9 $\pm$2.0 and so any signature score of an individual CIS patient above 6.8 (average $\pm$ 2 S.D.) was predicted to convert to CDMS. Of note, both the ONDCSF\textsubscript{VH4} signature score (at 4.5), and the MSPB\textsubscript{VH4} score (at 2.0) were below the threshold for MS conversion. CD19$^+$ CSF B cell and CD138$^+$ CSF plasma cell mutation positions both contributed to each CIS patient’s signature score, while the MSCSF\textsubscript{VH4} signature scores were only composed of CD19$^+$ CSF B cells.

2.6 VH4 structure

A human VH4–30.4 antibody structure described in (Guddat et al., 1993) was obtained from the Protein Data Bank (www.rcsb.org) under the identification moniker 1MCO, and adapted using the RasMol program (mc2.cchem.berkeley.edu/Rasmol/) to highlight codons within the designated signature of the heavy chain variable region.

3. Results

The 51 antibody heavy chain variable genes are subdivided into 7 different families ((Cook and Tomlinson, 1995) (http://www.ncbi.nlm.nih.gov/igblast/)), and it has been well-established that peripheral blood B cells from healthy donors utilize VH antibody genes most often from the VH3 family (“HCPB” in Table 1 and (Brezinschek et al., 1995; Brezinschek et al., 1997; Huang et al., 1992; Kraj et al., 1997; Wardemann et al., 2003; Yurasov et al., 2005)). In contrast, it has been reported by us and others that B cells in the CSF of MS patients often utilize VH4 antibody genes more frequently than those in the VH3 family (“MSCSF” in Table 1 and (Baranzini et al., 1999; Colombo et al., 2000; Harp et al., 2007; Monson et al., 2005; Owens et al., 1998; Owens et al., 2003; Owens et al., 2007; Qin et al., 1998; Ritchie et al., 2004)). The CISCSF antibody database consisting of CD19$^+$ B cells only had a similar frequency of B cells that utilize VH4 family genes in comparison to HCPB (26.2% vs. 21.8%, $p=0.20$ by $\chi^2$ test) (Table 1); in contrast, when CD138$^+$ plasma cells were included, the CISCSF
had a higher frequency of B cells that utilize VH4 family genes in comparison to HCPB (35.0% vs. 21.8%, $p=0.00001$ by $\chi^2$ test) (Table 1). Some individual CIS patient CSF B cell antibody repertoires were enriched for B cells utilizing VH4 family genes in comparison to the random expected frequency, as reported previously (Bennett et al., 2008). CSF-derived B cell antibody repertoires from patients with Other Neurological Diseases (OND) were not enriched for VH4-expressing CSF B cells in comparison to HCPB (23.1% vs. 21.8%, $p=0.83$ by $\chi^2$ test) or mHCPB (23.1% vs. 16.6%, $p=0.24$ by $\chi^2$ test), indicating that VH4 over-expression in the CSF of MS patients was not due to bias in the ability of VH4 expressing B cells to enter the CNS.

### 3.1 Identification of codons within MSCSF that are enriched for replacement mutations

Since VH4 expressing B cells are enriched in the CSF of MS patients, we hypothesized that mutational analysis would reveal a pattern (i.e. “signature”) of antibody gene replacement mutations that is unique to VH4 expressing B cells from the CNS of MS patients in comparison to HCPB. In order to test this hypothesis, the percentage of replacement mutations (RF) at each codon within the VH4 subdatabase extracted from the parent database (MSCSF$_{VH4}$) was determined and compared to the RF at each codon position within the VH4 subdatabase extracted from the parent HCPB database (HCPB$_{VH4}$). Replacement frequencies were used so that only those mutations resulting in an amino acid change would be considered. In addition, codon amino acid replacement can result from 1, 2, or 3 nucleotide changes within the codon, and so replacement frequencies limit bias based on the number of nucleotides in a codon that are mutated to generate a replacement. Hot spots were defined as those codon positions within MSCSF$_{VH4}$ with a statistically higher RF at a particular codon position in comparison to HCPB$_{VH4}$ (Table 2). Using this approach, 8 codon positions (31B, 32, 40, 56, 57, 60, 81, and 89) were identified that have a total RF value in MSCSF$_{VH4}$ (25.0%) that was statistically higher than in HCPB$_{VH4}$ (12.6%) ($p=0.001$ by $\chi^2$ test). Cold spots were defined as those codon positions within MSCSF$_{VH4}$ with a statistically lower RF at a particular codon position in comparison to HCPB$_{VH4}$ (Table 2). Four codons (30, 43, 77 and 82) were identified as cold spots that have a total RF value in MSCSF$_{VH4}$ (5.1%) that was statistically less than in HCPB$_{VH4}$ (8.5%) ($p=0.001$ by $\chi^2$ test).

Individual MS patient RFs within the 8 hot spot codons of the signature ranged from 22.5 to 34.1 percent (data not shown), indicating that some individual patient MSCSF repertoires had a greater enrichment of replacements at these 8 codon positions than others. Also, the variability of RF values within the 8 hot spot codons of the signature in individual VH4 genes in MSCSF$_{VH4}$ ranged from 14.5 to 36 percent (data not shown), indicating that some individual VH4 genes had a greater enrichment of replacements at these 8 hot spot codon positions than others. Previous analysis had identified codon 56 as a replacement mutation hotspot in HCPB (Dorner et al., 1997; Dorner et al., 1998a), which intensified as a hot spot in MSCSF$_{VH4}$ since a significantly greater percentage of replacement mutations were found in MSCSF$_{VH4}$ at codon 56 compared to HCPB$_{VH4}$. Of note, there was a 7.0 fold increase in replacement accumulation at codon 31B in the MSCSF$_{VH4}$ database in comparison to HCPB$_{VH4}$ that is likely due to the use of this codon by only a subset of VH4 genes (4–30, 4–31, 4–39 and 4–61). When the analysis was restricted to those B cells expressing VH4 genes that contain codon 31B, there was a 3.1-fold increase in RF of the MSCSF$_{VH4}$ database compared to HCPB$_{VH4}$ ($p<0.001$).

An example of a signature-enriched VH4 antibody gene rearrangement from a CSF-derived B cell of an MS patient is provided in Figure 1. Of note, 5 of the 8 hot spot codons of the signature retained higher RF values in MSCSF compared to the memory HCPB database (31B, 40, 56, 57, and 60), emphasizing that the signature does not simply reflect enrichment of memory B cells in the CSF.
3.2 Potency of signature score to predict development of clinically definite MS

We reasoned that prevalence of the signature would allow us to identify patients at risk to develop MS who subsequently convert to CDMS. Current criteria for diagnosis of MS requires dissemination of lesions both in time and space (Barkhof et al., 1997; Polman et al., 2005; Tintore et al., 2000). When MRI lesions alone are not sufficient to confirm diagnosis, CSF abnormalities can be used to meet the criteria of dissemination in space (Polman et al., 2005; Siritho and Freedman, 2009). Risk of conversion to clinically definite MS in patients who have had a single demyelinating event is 50 to 90% if the patient has an abnormal MRI (Beck et al., 2003; Brex et al., 2002; Cole et al., 1998; O’Riordan et al., 1998; Soderstrom et al., 1998), but of those patients with a normal MRI, up to 29% had oligoclonal bands and converted to CDMS (Cole et al., 1998).

In order to test whether signature prevalence could predict conversion to MS, we generated CSF B cell repertoires from patients who had one demyelinating event that placed them “at risk” to develop MS. Such patients are typically diagnosed with CIS. CD19+ B cell and CD138+ plasma cell repertoires from the CSF of two CIS patients at UTSWMC and nine patients at UCHSC were generated and analyzed for RF values within the 6 codons of the signature defined in the MSCSFVH4 database that had the most significant difference in RF compared to HCPB-VH4 at each codon position (codons 31B, 40, 56, 57, 81 and 89). RF values were combined using a signature score that accounts for RF variance as described in Materials and Methods. The average signature score in the MSCSFVH4 database was 10.9 ±2.0 (range 7.6 to 11.9), and so any individual CIS patient score that was 6.8 (average signature score – 2 S.D.) or higher was predicted to develop MS (Figure 2 and Table 3). Notably, the signature score from a pool of VH4 expressing CSF B cells of 3 OND patients was 4.5, and the signature score from a pool of VH4 expressing peripheral blood B cells of 3 CDMS patients was 2.0, and thus did not reach the 6.8 signature score threshold. Also, signature scores based on CD19+ B cell sequences only (in the patients where this was possible) did not change predictions based on signature score. This was expected since there is significant overlap in the antibody gene repertoires of CD19+ B cells and CD138+ plasma cells from the CSF of the same patients (Martin Mdel and Monson, 2007; Ritchie et al., 2004), suggesting that the memory B cell pool present in the CSF is the reservoir for differentiation of plasma cells in the CSF.

As indicated in Table 3, prediction of conversion to CDMS using the antibody gene signature score was accurate in 8 of 8 CIS patients that converted to CDMS. Lack of signature prevalence also accurately predicted that 2 of 2 patients who had recently experienced a first demyelinating event (ON3-1 and ON4–10) would not develop CDMS, and indeed, have not developed CDMS up to 2 years after initial sampling. One additional patient who had recently experienced a first demyelinating event (ON3–4) had a high signature score (11.3), but had not converted to CDMS at the two-year follow-up. The antibody gene signature yielded a sensitivity of 100%, specificity of 67%, positive predictive value of 89%, negative predictive value of 100%, and accuracy of 91%, as defined by others applying the McDonald Criteria to identify CIS patients that would convert to MS (Dalton et al., 2002). Most patients in this cohort converted to CDMS within 3–6 months of repertoire sampling, although in the case of CIS132, conversion to CDMS was not confirmed until 17 months after antibody repertoire sampling (Table 3). MRI, OCB and VH4/VH2 bias are also useful in assessing probability of MS conversion (Bennett et al., 2008; Freedman et al., 2005; Frohman et al., 2003; Korteweg et al., 2006; Paolino et al., 1996; Soderstrom et al., 1998), but were not considered in calculating the signature score.

4. Discussion

The intense somatic hypermutation accumulation in MSCSFVH4 enabled us to identify a unique antibody gene signature enriched for replacement accumulation at codons 31B, 32, 40, 56, 57,
60, 81 and 89 that was not observed in HCPB}_VH4 or ONDCSF}_VH4. Of note, any residual effect of naïve B cells on the RF calculation was minimized by tabulating only those sequences with mutations resulting in amino acid replacements. This approach minimized bias in the signature that may have reflected enrichment of mutation accumulation in CNS derived B cells (which are mostly memory and thus have high mutation rates) compared to peripheral B cells (which are mostly naïve and thus have low mutation rates). In addition, 5 of the 8 hot codons of the signature retain higher RF values in the MSCSF database compared to the memory HCPB database. Finally, if signature score reflected enrichment of mutation accumulation due to the repertoire’s high memory representation, then all signature scores from the CISCSF antibody repertoires should have been high since they were all heavily enriched for memory B cells. This was not the case, since CISCSF repertoires ON3-1 and ON4–10, despite being heavily enriched for memory B cells (with mutation frequencies of 5.2% and 6.7%, respectively), had signature scores below the threshold of 6.8 (ON3-1 score=6.4, ON4–10 score=2.2).

It was compelling to investigate whether the antibody gene signature may be of value to identify CIS patients who would subsequently develop MS, since early and accurate diagnosis of MS is of tantamount importance in clinical care (Stuve et al., 2008). Signature prevalence could be used to identify patients who would be diagnosed with CDMS within 3–18 months of experiencing their first demyelinating event. Of note, patient ON3–4 had a signature score that indicated this patient would convert to CDMS (score=11.3), but did not demonstrate a lesion load by MRI, banding by OCB, or VH4/VH2 bias, and had not developed CDMS up to 2 years after sampling was performed (Table 3). It will be interesting to determine whether this patient is diagnosed with CDMS over time. It is also important to note that the majority of patients in this cohort already had evidence of MS risk as indicated by positive MRI and OCB. Patient ON4–7, however, did not present with brain lesions by MRI, but had a signature score that indicated this patient would be diagnosed with CDMS (score=10.2). Indeed, this patient did convert to CDMS within 5 months of CSF B cell antibody repertoire sampling, and provides a reasonable example of how signature prevalence may predict CDMS diagnosis in patients that either do not present with brain lesions by MRI, or who are not evaluated by MRI at this stage of their disease. It will be interesting to determine whether the combination of MRI and signature prevalence would be useful in predicting MS conversion. Signature prevalence may also provide an evaluation mechanism to identify the most appropriate patient candidates to receive B cell depletion therapies, for example. Certainly this is a priority since a recent investigation demonstrated significant efficacy of Rituxan in RRMS patients (Hauser et al., 2008).

Given the urgency for the early identification of MS and the rapid initiation of disease modifying therapy, presentation of a molecular signature in the CSF B cells of CIS patients who develop MS may provide a unique tool for identifying at-risk individuals. However, wide implementation of the current form of this approach would be problematic since the AGS scores presented here were generated using a specialized technique that is labor intensive (single cell PCR) and requires fresh CSF for sampling. Developing other approaches to generate AGS data that maintains accuracy, but does not require a specialized laboratory to perform, is attainable and of paramount importance.

In addition, many early MS patients have atypical clinical presentations or unremarkable MRI scans, and patients with alternative inflammatory conditions may mimic idiopathic demyelinating disease. In these circumstances, the advent of a molecular diagnostic signature would increase diagnostic sensitivity and specificity. Investigating the utility of the antibody gene signature in such patients is ongoing in our laboratory.

The presence of a mutational signature among clonally expanded VH4 germline antibodies in MSCSF may be helpful in understanding disease pathogenesis. For example, the VH4 germline
mutational signature may be the direct result of antigen targeting in the humoral immune response. Therefore, determining the antigen specificity of signature-enriched antibodies from CSF B cells of patients with definite MS and CIS is one of the first steps towards dissecting whether signature-enriched B cells have the potential to participate in MS pathogenesis. Of note, 5 of the 8 signature codons (31B, 32, 56, 57 and 60) we identified as having a unique accumulation of amino acid replacements in MSCSF_{VH4} are predicted to have direct antigen contact since they reside in CDRs (Supplemental Figure 1). Dissecting the relative contribution of replacement mutations at each of these signature codons as well as those outside of the antigen-binding region will address the impact of both codon classifications (direct and indirect antigen binding capacity) on antigen binding affinity.

In summary, a unique signature of antibody gene replacement mutations was identified in the MSCSF_{VH4} database that is not observed in healthy control peripheral blood or CSF-derived VH4-expressing B cells from patients with other neurological diseases. Prevalence of the signature was accurate in identifying CIS patients that would convert to CDMS, but needs to be tested on a larger cohort of patients at both high and low risk to develop MS. Identifying the antigen specificity of signature-enriched CSF B cells from these patients may also reveal a unique group of antigens that are central to initiation of humoral autoimmunity in the CNS. It is likely that the MS-specific VH4 antibody gene signature provides both a new focus of investigation to further elucidate the role of B cells and their antibody products in MS and a new candidate molecular diagnostic tool for MS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the patients who consented to sampling for this study. Special thanks to Dr. Olaf Stüve for critical review of this manuscript. Bonnie Darnell, Angie Mobley and Elizabeth Curry are thanked for their technical expertise in cell sorting, and Tanya Hendricks and Timothy Ahearne for repertoire analysis support.

This study was supported by grants from the National Institutes of Health (NIH) to NLM (RO1 NS 40993) and MKR (RO1 NS 37513, RO1 AI 47133, and K24 NS 44250), the National Multiple Sclerosis Society (NMS) to NLM (RG3267) and JLB (RG3908), the Yellow Rose Foundation (NLM and MKR), the Wadsworth Foundation (to NLM) and Howson funds (to NLM). EMC and CH were supported by NIH NRSA5 T32 AI 005284-28 from NIAID. The authors have no conflicting financial interests.

References


Kabat, EA.; Wu, TT.; Bilofsky, H.; Reid-Miller, M.; Perry, H. Sequences of Proteins of Immunological Interest. United States Department of Health and Human Services; Washington, D.C: 1983.


**Figure 1. Example of VH4 comparison**

A VH4-30.4 sequence is listed as the germline configuration (allele 01) and compared to a patient CD19⁺ B cell sequence. The germline protein conversion and the changes made by replacement mutations in the patient sequence are noted. Signature codons are boxed, with the dashed boxes demarcating cold spots, and the solid boxes demarcating hot spots. CDRs as defined by Kabat (Kabat et al., 1983) are shaded.

---

**Codon**

**IgH4-30.4** Germline DNA

<table>
<thead>
<tr>
<th>Codon</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>31A</th>
<th>31B</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
</tr>
</thead>
<tbody>
<tr>
<td>M125-391 DNA</td>
<td>GTC</td>
<td>TCT</td>
<td>GGT</td>
<td>GGC</td>
<td>TCC</td>
<td>ATC</td>
<td>GTC</td>
<td>AGT</td>
<td>GGT</td>
<td>AGT</td>
<td>CGT</td>
<td>GAT</td>
<td>GAT</td>
<td>CAC</td>
<td>TGG</td>
<td>AGT</td>
</tr>
<tr>
<td>Germline Protein</td>
<td>V</td>
<td>S</td>
<td>G</td>
<td>G</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>G</td>
<td>D</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>H</td>
</tr>
</tbody>
</table>

**IgH4-30.4** Germline DNA

<table>
<thead>
<tr>
<th>Codon</th>
<th>38</th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>51</th>
<th>52</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>M125-391 DNA</td>
<td>CGC</td>
<td>CAG</td>
<td>CCC</td>
<td>CCA</td>
<td>GGG</td>
<td>AAG</td>
<td>GGC</td>
<td>CTG</td>
<td>GAG</td>
<td>TGG</td>
<td>ATT</td>
<td>GGG</td>
<td>TAC</td>
<td>ATC</td>
<td>TAT</td>
<td>TAC</td>
</tr>
<tr>
<td>Germline Protein</td>
<td>R</td>
<td>Q</td>
<td>P</td>
<td>P</td>
<td>G</td>
<td>K</td>
<td>G</td>
<td>L</td>
<td>E</td>
<td>W</td>
<td>I</td>
<td>G</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

**IgH4-30.4** Germline DNA

<table>
<thead>
<tr>
<th>Codon</th>
<th>54</th>
<th>55</th>
<th>56</th>
<th>57</th>
<th>58</th>
<th>59</th>
<th>60</th>
<th>61</th>
<th>62</th>
<th>63</th>
<th>64</th>
<th>65</th>
<th>66</th>
<th>67</th>
<th>68</th>
<th>69</th>
</tr>
</thead>
<tbody>
<tr>
<td>M125-391 DNA</td>
<td>AGT</td>
<td>GGG</td>
<td>AGC</td>
<td>ACC</td>
<td>TAC</td>
<td>TAC</td>
<td>AAC</td>
<td>CCG</td>
<td>TCC</td>
<td>CTC</td>
<td>AAG</td>
<td>AGT</td>
<td>CGA</td>
<td>GTT</td>
<td>ACC</td>
<td>ATA</td>
</tr>
<tr>
<td>Germline Protein</td>
<td>S</td>
<td>G</td>
<td>S</td>
<td>T</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>P</td>
<td>S</td>
<td>L</td>
<td>K</td>
<td>S</td>
<td>R</td>
<td>V</td>
<td>T</td>
<td>I</td>
</tr>
</tbody>
</table>

**IgH4-30.4** Germline DNA

<table>
<thead>
<tr>
<th>Codon</th>
<th>70</th>
<th>71</th>
<th>72</th>
<th>73</th>
<th>74</th>
<th>75</th>
<th>76</th>
<th>77</th>
<th>78</th>
<th>79</th>
<th>80</th>
<th>81</th>
<th>82</th>
<th>82A</th>
<th>82B</th>
<th>82C</th>
</tr>
</thead>
<tbody>
<tr>
<td>M125-391 DNA</td>
<td>TCA</td>
<td>GTA</td>
<td>GAC</td>
<td>ACG</td>
<td>TCC</td>
<td>AAG</td>
<td>AAC</td>
<td>CAG</td>
<td>TCC</td>
<td>CTC</td>
<td>CTG</td>
<td>AAG</td>
<td>CTG</td>
<td>AAG</td>
<td>GGC</td>
<td>AGA</td>
</tr>
<tr>
<td>Germline Protein</td>
<td>S</td>
<td>V</td>
<td>D</td>
<td>T</td>
<td>S</td>
<td>K</td>
<td>N</td>
<td>Q</td>
<td>F</td>
<td>S</td>
<td>L</td>
<td>T</td>
<td>K</td>
<td>S</td>
<td>V</td>
<td>I</td>
</tr>
</tbody>
</table>

**IgH4-30.4** Germline DNA

<table>
<thead>
<tr>
<th>Codon</th>
<th>83</th>
<th>84</th>
<th>85</th>
<th>86</th>
<th>87</th>
<th>88</th>
<th>89</th>
<th>90</th>
<th>91</th>
<th>92</th>
<th>93</th>
<th>94</th>
</tr>
</thead>
<tbody>
<tr>
<td>M125-391 DNA</td>
<td>ACT</td>
<td>GCC</td>
<td>GCA</td>
<td>GAC</td>
<td>AGC</td>
<td>GCC</td>
<td>GTG</td>
<td>TAT</td>
<td>TAC</td>
<td>TGT</td>
<td>GCC</td>
<td>AGA</td>
</tr>
<tr>
<td>Germline Protein</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>T</td>
<td>A</td>
<td>V</td>
<td>Y</td>
<td>Y</td>
<td>C</td>
<td>A</td>
<td>R</td>
</tr>
</tbody>
</table>
Figure 2. Signature score in individual MS and CIS patients
Signature scores were generated by calculating Z-scores for the RF values at the 6 codons within the signature (31B, 40, 56, 57, 81 and 89). Individual z-scores at each of the codon positions were compiled to generate the composite signature z-score. MS patient signature scores are shown as black circles (●), CIS patient signature scores that resulted in prediction of CDMS are black squares (■), and CIS patient signature scores that resulted in prediction of unlikely to convert to definite MS are in open squares (□). The average composite signature score in the MSCSF\textsubscript{VH4} database was 10.9 ±2.0 (black line) and so any signature score of an individual CIS patient above 6.8 (average – 2 S.D.; threshold shown as red line) was predicted to convert to CDMS. For reference, ONDCSF\textsubscript{VH4} group signature score was 4.5, and MSPB\textsubscript{VH4} signature score was 2.0.
<table>
<thead>
<tr>
<th>B cell source</th>
<th>Expected Frequency</th>
<th>HCPB&lt;sup&gt;b&lt;/sup&gt;</th>
<th>mHCPB&lt;sup&gt;c&lt;/sup&gt;</th>
<th>MSCSF&lt;sup&gt;d&lt;/sup&gt;</th>
<th>CISCSF&lt;sup&gt;e&lt;/sup&gt;</th>
<th>CISCSF&lt;sup&gt;f&lt;/sup&gt;</th>
<th>ONDCSF&lt;sup&gt;g&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>By gene frequency&lt;sup&gt;j&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt; IgD&lt;sup&gt;-&lt;/sup&gt; CD27&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt; CD138&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD19&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>VH1</td>
<td>21.6</td>
<td>15.5</td>
<td>10.7</td>
<td>27.1&lt;sup&gt;h&lt;/sup&gt;,&lt;sup&gt;i&lt;/sup&gt;</td>
<td>9.9</td>
<td>8.9</td>
<td>20.0</td>
</tr>
<tr>
<td>VH2</td>
<td>5.9</td>
<td>2.0</td>
<td>2.0</td>
<td>1.6</td>
<td>5.3&lt;sup&gt;h&lt;/sup&gt;</td>
<td>7.0</td>
<td>3.1</td>
</tr>
<tr>
<td>VH3</td>
<td>43.1</td>
<td>55.2</td>
<td>62.0</td>
<td>32.4&lt;sup&gt;h&lt;/sup&gt;,&lt;sup&gt;i&lt;/sup&gt;</td>
<td>56.3</td>
<td>47.5</td>
<td>46.2&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>VH4</td>
<td>21.6</td>
<td>21.8</td>
<td>16.6</td>
<td>34.3&lt;sup&gt;h&lt;/sup&gt;,&lt;sup&gt;i&lt;/sup&gt;</td>
<td>26.2&lt;sup&gt;i&lt;/sup&gt;</td>
<td>35.0</td>
<td>23.1</td>
</tr>
<tr>
<td><strong>Total Sequences</strong></td>
<td>51</td>
<td>348</td>
<td>205</td>
<td>373</td>
<td>302</td>
<td>528</td>
<td>65</td>
</tr>
<tr>
<td><strong>Number of Donors</strong></td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>

Abbreviations: VH, variable heavy; HCPB, healthy control peripheral blood; mHCPB, memory healthy control peripheral blood; MSCSF, multiple sclerosis cerebrospinal fluid; CISCSF, clinically isolated syndrome cerebrospinal fluid; ONDCSF, other neurological disease cerebrospinal fluid.

<sup>a</sup>Values provided in percent

<sup>b</sup>The HCPB group includes CD19<sup>+</sup> B cell antibody sequences from healthy controls BF1 (n=67) and BF2 (n=281).

<sup>c</sup>The mHCPB group includes CD19<sup>+</sup> B cell antibody sequences from healthy controls BF1 (n=18) and BF2 (n=105) with 4 or more mutations (less than 98% homology to germline) and IgD<sup>-</sup>CD27<sup>+</sup> B cell antibody sequences from healthy controls HA (n=9), HB (n=44), HC (n=26), and HE (n=3) (Tian et al., 2007).

<sup>d</sup>The MSCSF group includes CD19<sup>+</sup> sequences from MS patients M125 (n=101), M199 (n=19), M354 (n=6), M368 (n=49), M376 (n=8), M484 (n=9), M522 (n=71), M584 (n=85), M875 (n=21), M217 (n=1), and M887 (n=3).

<sup>e</sup>The CISCSF group includes CD19<sup>+</sup> sequences from CIS patients CIS132 (n=19), CIS429 (n=57), CIS3-1 (n=24), ON3-1 (n=23), ON3-3 (n=39), ON3-4 (n=28), ON3-5 (n=35), ON4-7 (n=17), ON4-10 (n=31), and ON5-2 (n=29).

<sup>f</sup>In addition to those CD19<sup>+</sup>sequences listed in 5, this group includes CD138<sup>+</sup> sequences from CIS patients CIS3-1 (n=76), ON3-1 (n=45), ON3-3 (n=12), ON3-5 (n=44), ON4-7 (n=20), ON4-8 (n=17), ON5-2 (n=12).

<sup>g</sup>The ONDCSF group includes CD19<sup>+</sup> B cell antibody sequences from OND patients OND341 (n=32), OND758 (n=19), and OND116 (n=14).

<sup>h</sup>Significantly different from HCPB frequency

<sup>i</sup>Significantly different from mHCPB frequency

<sup>j</sup>Expected frequency from (Cook and Tomlinson, 1995)
Table 2

Percentage of Replacement Mutations in Each Signature Codon

<table>
<thead>
<tr>
<th>Codon</th>
<th>Location</th>
<th>MSCSF\text{VH}_4</th>
<th>HCPB\text{VH}_4</th>
<th>mHCPB\text{VH}_4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RF</td>
<td>RF</td>
<td>p-value(^d)</td>
</tr>
<tr>
<td>31B(^e)</td>
<td>CDR1</td>
<td>3.5(^b)</td>
<td>0.5(^b)</td>
<td>7.0</td>
</tr>
<tr>
<td>32</td>
<td>CDR1</td>
<td>2.3</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>40(^e)</td>
<td>FR2</td>
<td>2.7</td>
<td>1.0</td>
<td>2.7</td>
</tr>
<tr>
<td>56(^a,e)</td>
<td>CDR2</td>
<td>5.5</td>
<td>3.0</td>
<td>1.8</td>
</tr>
<tr>
<td>57(^e)</td>
<td>CDR2</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>60</td>
<td>CDR2</td>
<td>2.4</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>81(^e)</td>
<td>FR3</td>
<td>4.7</td>
<td>3.0</td>
<td>1.5</td>
</tr>
<tr>
<td>89(^e)</td>
<td>FR3</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.0</td>
<td>12.6</td>
<td>2.0</td>
</tr>
<tr>
<td>30</td>
<td>FR1</td>
<td>2.0</td>
<td>4.0</td>
<td>0.5</td>
</tr>
<tr>
<td>43</td>
<td>FR2</td>
<td>0.9</td>
<td>2.0</td>
<td>0.5</td>
</tr>
<tr>
<td>77</td>
<td>FR3</td>
<td>1.5</td>
<td>2.5</td>
<td>0.6</td>
</tr>
<tr>
<td>82</td>
<td>FR3</td>
<td>0.7</td>
<td>2.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>

|       |          | 5.1              | 8.5             | 0.5              | 0.001            | 7.4           | 0.7           | 0.01           |

Abbreviations in table: CDR, complementary determining region; FR, framework; RF, replacement frequency; MSCSF, multiple sclerosis cerebrospinal fluid; HCPB, healthy control peripheral blood; mHCPB, memory HCPB; NS, not significant

\(^a\) Previously published replacement hotspot (Dorner et al., 1997; Dorner et al., 1998a)

\(^b\) As defined by Kabat (Kabat et al., 1983)

\(^c\) “Hotspot Total” is the total RF within codons 31B, 32, 40, 56, 57, 60, 81 and 89. “Coldspot Total” is the total RF within codons 30, 43, 77 and 82. 199, 337 and 965 replacements respectively were included in this analysis for HCPB\text{VH}_4, mHCPB\text{VH}_4, and MSCSF\text{VH}_4.

\(^d\) Comparing HCPB\text{VH}_4 or mHCPB\text{VH}_4 to MSCSF\text{VH}_4 RFs at each codon position using \(\chi^2\) goodness-of-fit where expected frequency is the RF calculated in HCPB\text{VH}_4 and mHCPB\text{VH}_4 respectively.

\(^e\) Codon used in the calculation of signature score
### Table 3

**CIS Patient Summary and Signature Score Predictions**

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Time to LP (^a)</th>
<th>MRI Brain Lesions</th>
<th>OCB</th>
<th>CD19/CD138 (b)</th>
<th>CD19 VH4 bias (c)</th>
<th>CD138 VH4 bias (d)</th>
<th>Signature Score (e)</th>
<th>Prediction based on signature score (f)</th>
<th>Definite MS</th>
<th>Time to MS Diagnosis (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIS132</td>
<td>1</td>
<td>GD(^e)</td>
<td>Yes</td>
<td>10/NA</td>
<td>No</td>
<td>NA</td>
<td>12.1</td>
<td>CDMS</td>
<td>Clinical</td>
<td>18</td>
</tr>
<tr>
<td>CIS429</td>
<td>1</td>
<td>GD(^e)</td>
<td>Yes</td>
<td>96/NA</td>
<td>No</td>
<td>NA</td>
<td>15.0</td>
<td>CDMS</td>
<td>Clinical</td>
<td>3</td>
</tr>
<tr>
<td>CIS3-1</td>
<td>2</td>
<td>WML</td>
<td>Yes</td>
<td>24/76</td>
<td>Yes</td>
<td>Yes</td>
<td>15.5</td>
<td>CDMS</td>
<td>MRI</td>
<td>3</td>
</tr>
<tr>
<td>ON3-5</td>
<td>4</td>
<td>GD(^e)</td>
<td>Yes</td>
<td>39/13</td>
<td>No</td>
<td>No</td>
<td>11.3</td>
<td>CDMS</td>
<td>Clinical</td>
<td>3</td>
</tr>
<tr>
<td>ON4–7</td>
<td>3</td>
<td>WML</td>
<td>Yes</td>
<td>172/20</td>
<td>No</td>
<td>Yes</td>
<td>12.8</td>
<td>CDMS</td>
<td>Clinical</td>
<td>5</td>
</tr>
<tr>
<td>ON4–8</td>
<td>1.5</td>
<td>WML</td>
<td>Yes</td>
<td>NA/8</td>
<td>NA</td>
<td>Yes</td>
<td>9.6</td>
<td>CDMS</td>
<td>Clinical</td>
<td>5</td>
</tr>
<tr>
<td>ON5-2</td>
<td>1</td>
<td>GD(^e)</td>
<td>Yes</td>
<td>29/12</td>
<td>Yes</td>
<td>Yes</td>
<td>7.9</td>
<td>CDMS</td>
<td>Clinical</td>
<td>3</td>
</tr>
<tr>
<td>ON3-1</td>
<td>10</td>
<td>WML</td>
<td>Yes</td>
<td>23/45</td>
<td>No</td>
<td>No</td>
<td>6.4</td>
<td>No</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>ON4–10</td>
<td>1.25</td>
<td>WML</td>
<td>No</td>
<td>3/NA</td>
<td>NA</td>
<td>No</td>
<td>2.2</td>
<td>No</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>ON3–3</td>
<td>1</td>
<td>GD(^e)</td>
<td>Yes</td>
<td>26/NA</td>
<td>No</td>
<td>No</td>
<td>11.3</td>
<td>CDMS</td>
<td>Clinical</td>
<td>5</td>
</tr>
<tr>
<td>ON4–10</td>
<td>4</td>
<td>None(^g)</td>
<td>No</td>
<td>26/NA</td>
<td>No</td>
<td>No</td>
<td>11.3</td>
<td>CDMS</td>
<td>Clinical</td>
<td>5</td>
</tr>
</tbody>
</table>

Abbreviations in table: CIS, Clinically Isolated Syndrome; ON, Optic Neuritis; LP, lumbar puncture; GD, gadolinium enhancing lesion positive; WML, white matter lesions by T2; OCB, oligoclonal bands; CDMS, clinically definite MS; NA, not applicable

\(^a\)CIS132 and CIS429 were generated at UTSWMC; the remaining patient CSF B cell repertoires were generated at UCD.

\(^b\)Months from first demyelinating event to LP.

\(^c\)Values given are number of unique sequences in CD19 repertoire/CD138 repertoire; family usage can be found in (Bennett et al., 2008; Harp et al., 2007). CIS429 was a 62 y.o. male first presenting with optic neuritis and the repertoire had 2% VH1 usage, 0% VH2, 77% VH3, and 19% VH4.

\(^d\)Bias was considered significantly different from random frequency (Cook and Tomlinson, 1995) or expected frequency in HCPB (Brezinschek et al., 1997; Brezinschek et al., 1998; Dörner et al., 1997; Dörner et al., 1998a; Dörner et al., 1998b; Dörner et al., 1998c; Farmer et al., 1999; Hansen et al., 2000; Monson et al., 2000). VH2 bias was also observed in CD19 repertoires from ON3–3, ON3–5 and ON4–7.

\(^e\)Signature score was calculated as outlined in materials and methods, and uses both CD19 and CD138 sequences. The average score among MS patients is 10.9 ±2.0.

\(^f\)Months from first demyelinating event to MS diagnosis.

\(^g\)One spinal cord lesion was observed by T2 weighted MRI.