Molecular Immunology 67 (2015) 625–635

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Peptide motif analysis predicts lymphocytic choriomeningitis virus as trigger for multiple sclerosis

Charissa Hogeboom

Independent Researcher, Albany, CA 94706, USA

ARTICLE INFO

Article history:
Received 26 June 2015
Accepted 31 July 2015
Available online 28 August 2015

Keywords:
Multiple sclerosis
DRB1*1501
Lymphocytic choriomeningitis virus (LCMV)
Interferon
Immune regulation
Molecular mimicry

ABSTRACT

The etiology of multiple sclerosis (MS) involves both genetic and environmental factors. Genetically, the strongest link is with HLA DRB1*1501, but the environmental trigger, probably a virus, remains uncertain. This investigation scans a panel of proteins from encephalitogenic viruses for peptides homologous to the primary autoantigen from myelin basic protein (MBP), then evaluates candidate peptides against a motif required for T cell cross-reactivity and compares viral prevalence patterns to epidemiological characteristics of MS. The only peptide meeting criteria for cross-reactivity with MBP was one from lymphocytic choriomeningitis virus (LCMV), a zoonotic agent. In contrast to current candidates such as Epstein–Barr virus, the distribution of LCMV is consistent with epidemiological features of MS, including concentration in the temperate zone, higher prevalence farther from the equator, and increased prevalence in proximity to regions of peak MS incidence, while lack of person-to-person transmission is consistent with low MS concordance across monozygotic twins. Further, LCMV blocks induction of type I interferon (IFN). Hypothetically this would dysregulate immune processes in favor of proinflammatory pathways as well as upregulating HLA class II and providing more binding sites for autoantigen. The combination of molecular mimicry with virally-induced immune dysregulation has the potential to explain aspects of autoimmunity not addressed by either mechanism alone.

1. Introduction

Multiple sclerosis (MS) is a chronic immune-mediated neurological disease that affects approximately 1.3 million people worldwide. It is generally accepted that both genetic and environmental factors are involved, but disease etiology remains poorly understood after decades of research. Recent advances have identified a number of genetic risk factors, with HLA DRB1*1501 haplotype continuing to provide the strongest association (Gourraud et al., 2012; Wu et al., 2010; Qiu et al., 2011; Link et al., 2012; Nolan et al., 2012). Yet, with approximately 9% of the world's population carrying this allele (Solberg et al., 2008) and a global MS prevalence of 0.03% (World Health Organization, 2008), the proportion of DRB1*1501-positive individuals who develop MS is low, on the order of 0.3%. Among environmental factors, a viral agent has been postulated to instigate immune recognition of self-antigens (Fujinami, 2001), hypothetically by means of a mechanism termed "molecular mimicry" (Olson et al., 2001). Under this model, structural similarities between a viral peptide and a self-peptide cause activation of autoreactive T cells. Although a range of viruses has been considered over the past 50 years, including poliovirus, measles, rabies, herpes family viruses, mumps, canine distemper, and retroviruses (Kakalacheva et al., 2011), the identity of a causative virus remains elusive.

A popular candidate for an etiologic role in MS is Epstein–Barr virus (EBV), as antibodies against its nuclear antigen-1 (EBNA1) have been found in MS patients (Kakalacheva et al., 2011). However, the distribution of EBV worldwide does not provide a good fit to the epidemiology of MS. MS is concentrated in the temperate zone, with highest prevalence in Europe, Canada, the United States, and Australia (World Health Organization, 2008). Within the temperate zone, MS shows a gradient of prevalence that increases with latitude (Simpson et al., 2011) and is interspersed with regional pockets of exceptionally high prevalence. Although some of these differences can be explained by genetic factors, MS concordance across monozygotic twin pairs is low, ranging from 13% to 31%, based on studies in Canada, the United States, the British Isles, Finland, and Italy (Willer et al., 2003; Islam et al., 2006; Mumford et al., 1994; Kuusisto et al., 2008; Ristori et al., 2006; Sadovnick et al., 1993). This suggests that the environmental trigger for MS in genetically susceptible individuals is somewhat uncommon or has low infectivity. In contrast, the EBV seropositivity rate in adults is in

E-mail address: Charissa.Hogeboom@outlook.com

http://dx.doi.org/10.1016/j.molimm.2015.07.041
0161-5890/

excess of 90% (Kakalacheva et al., 2011). Further, exposure to EBV occurs earlier in life among children in developing countries, with universal seroconversion by age 3–4, whereas infection in developed countries often is delayed until adolescence (Hjalgrim et al., 2007). The high seroprevalence of antibodies against EBV, together with earlier exposure in countries with lower MS prevalence, are not fully consistent with the latitudinal gradient and low twin concordance rates seen for MS. In fact, some researchers suggest that EBV is a marker of chronic brain inflammation rather than causative per se (Castellazi et al., 2014).

The ideal candidate for an infectious trigger interacting with DRB1*1501 under the molecular mimicry hypothesis would satisfy both molecular biological and epidemiological criteria. The infectious agent would contain a peptide that binds to HLA in a similar way as the self-antigen and lies in a similar configuration. The bound peptide would activate the same T cell clones as those that recognize the self-antigen. The infectious agent would be somewhat uncommon or have low infectivity in order to explain the low MS concordance across monozygotic twin pairs. The agent would be most prevalent in the temperate geographic zone. Its distribution would be consistent with the latitudinal gradient observed for MS. It would show higher prevalence or infectivity in regions that report elevated incidence or prevalence of MS.

The aim of this investigation was to predict the viral peptide most closely matching these criteria. A set of proteins from viruses capable of causing encephalitis was scanned for regions of sequence similarity to myelin basic protein (MBP) residues 85–99. The peptides with highest sequence homology were then compared to MBP 85–99 on five scales representing characteristics predictive of protein binding and configuration. The highest scoring viral peptides were evaluated for similarity to a binding motif that has been determined experimentally to activate MBP-reactive T cell clones from MS patients with DRB1*1501 haplotype. Finally, the plausibility of the top predicted virus was evaluated through a review of its epidemiology and a comparison to the prevalence patterns observed for MS.

2. Materials and methods

2.1. General

All computations were done with custom programs written in the R language (Hornik, 2014).

2.2. Viral proteins

A list of viruses capable of causing encephalitis was generated from review of medical reference texts. Encephalitogenic viruses endemic to equatorial regions were excluded as unlikely to be causative, since MS is most prevalent in the temperate zone. Protein sequences derived from the viral capsid or envelope or previously observed to be antigenic were selected for testing. Protein sequences were obtained from the UniProt database.

2.3. Reference proteins

Viral homology scores were contrasted with scores from two other comparators. As a negative control for the MBP immunogen, an arbitrary 15-mer from serum albumin (ALB), residues 152–166, was selected. As a negative control for the viral proteins chosen for testing, a set of 17 control proteins was created by generating a chain of randomly selected amino acids whose frequency of occurrence was similar to the amino acid composition of proteins in general. The random proteins were matched for length with the viral proteins.

2.4. Homology Search

The segment of MBP that constitutes the immunogenic portion was defined as MBP residues 85–99, based on experimental evidence provided by Wucherpfennig et al. (1994) and Hausmann et al. (1999). Viral protein sequences were scanned for regions of high homology with this peptide. As it is unknown whether high homology is required for the entire peptide length, homology was computed for a range of windows on the MBP 85–99 sequence. All possible windows of length 4–15 that covered the region MBP 89–92, the primary anchor region of the MBP immunogen bound to DRB1*1501, were included. This resulted in 40 windows on the MBP immunogen.

Homology scores were computed using the BLOSUM50 blocks substitution matrix (Henikoff and Henikoff, 1992) and were based on alignments without gaps. For each of the 40 windows, each candidate viral protein was marched through all possible alignments with the MBP sequence that fell within the window. Thus, for a protein of length n and a window width k, a total of $(n - k + 1)$ homology scores were computed. For each viral protein, the peptide location resulting in the maximum score for each window was saved. A limited comparison of results using BLOSUM62 showed that conclusions were not dependent on the BLOSUM matrix used (data not shown).

To allow comparison of homology scores across window locations and widths, scores were normalized to window length by averaging. Average homology scores were scaled to percent of maximum, defined as the homology score obtained when comparing the windowed segment of the MBP immunogen with itself, using the value –1 as minimum of the range. The distribution of homology scores was compared across peptide source (viral or random) using a Wilcoxon rank sum statistic at level 0.05. A Simes test was used to control for multiple comparisons. Since the windows were overlapping and nested, the statistics used are considered exploratory rather than indicating true statistical significance. The program and datasets used to conduct this analysis online in a set of six Supplementary files. Please refer to the README.txt file for usage of the remaining 5 items (homology.txt, MBP.csv, albumin.csv, BLOSUM50.csv, and protein_panel.csv).

2.5. Peptide characteristics profiles

Viral peptides were compared to MBP 85–99 on five scales used to predict protein conformation or binding. These were surface accessibility, antigenicity, flexibility, hydrophobicity, and hydrophilicity (Janin et al., 1978; Welling et al., 1985; Karplus and Schulz, 1985; Eisenberg et al., 1984; Parker et al., 1986). For each scale, a profile for MBP 85–99 was defined as the sequence of scale scores for the 15 residues. The MBP profile was then compared with that of each candidate homologous viral peptide by summing the positionwise squared deviations in scale scores. Dividing by the peptide length, each homologous viral peptide was given a "mean squared error" (MSE) value that reflected the degree to which the profile of the viral peptide deviated from that of MBP on the given scale. The MSE values were converted to percents by dividing by the maximum possible deviation from MBP that could have been obtained using that scale.

For reference, a set of microbial peptides shown experimentally to be capable of activating a DRB1*1501-restricted MBP-reactive T cell clone (Hausmann et al., 1999) are also included as positive controls.

2.6. Motif for peptide binding and cross-reactivity

A motif was built to codify the criteria a peptide must meet in order to cross-react with MBP in binding to HLA DRA-DRB1*1501 and activating the same T cell clones. These criteria were