

Humoral Immunity Profiling of Subjects with Myalgic Encephalomyelitis Using a Random Peptide Microarray Differentiates Cases from Controls with High Specificity and Sensitivity

Sahajpreet Singh¹ · Phillip Stafford² · Karen A. Schlauch^{3,4} · Richard R. Tillett⁴ · Martin Gollery⁵ · Stephen Albert Johnston² · Svetlana F. Khaiboullina^{1,6} · Kenny L. De Meirleir¹ · Shanti Rawat¹ · Tatjana Mijatovic⁷ · Krishnamurthy Subramanian¹ · András Palotás^{6,8} · Vincent C. Lombardi^{1,9}

Received: 3 June 2016 / Accepted: 29 November 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Myalgic encephalomyelitis (ME) is a complex, heterogeneous illness of unknown etiology. The search for biomarkers that can delineate cases from controls is one of the most active areas of ME research; however, little progress has been made in achieving this goal. In contrast to identifying biomarkers that are directly involved in the pathological process, an immunosignature identifies antibodies raised to proteins expressed during, and potentially involved in, the pathological process. Although these proteins might be unknown, it is possible to detect antibodies that react to these proteins using random peptide arrays. In the present study, we probe a custom 125,000 random 12-mer peptide microarray with sera from 21 ME cases and 21 controls from the USA and Europe and used these data to develop a diagnostic signature. We further used these peptide sequences to potentially uncover the naturally occurring candidate antigens to which these antibodies may specifically react with *in vivo*. Our analysis

revealed a subset of 25 peptides that distinguished cases and controls with high specificity and sensitivity. Additionally, Basic Local Alignment Search Tool (BLAST) searches suggest that these peptides primarily represent human self-antigens and endogenous retroviral sequences and, to a minor extent, viral and bacterial pathogens.

Keywords Antibody · Chronic fatigue syndrome · Immunosignature · Myalgic encephalomyelitis · Peptide array

Introduction

Myalgic encephalomyelitis (ME), also commonly referred to as chronic fatigue syndrome or ME/CFS, is a heterogeneous illness characterized by a number of physical symptoms and comorbid conditions including neurocognitive dysfunction,

Sahajpreet Singh and Phillip Stafford contributed equally to this work.

✉ András Palotás
palotas@asklepios-med.eu

✉ Vincent C. Lombardi
vclombardi@nvcbcr.org

¹ Nevada Center for Biomedical Research,
1664 N Virginia St. MS 0552, Reno, NV 89557-0552, USA

² The Biodesign Institute Center for Innovations in Medicine at
Arizona State University, Tempe, AZ, USA

³ Department of Biochemistry and Molecular Biology,
University of Nevada, Reno, NV, USA

⁴ Nevada INBRE Bioinformatics Core, University of Nevada,
Reno, NV, USA

⁵ Tahoe Bioinformatics, Incline Village, Reno, NV, USA

⁶ Kazan Federal University, Kazan, Russian Federation

⁷ R.E.D. Laboratories, Zellik, Belgium

⁸ Asklepios-Med (private medical practice and research center),
Kossuth Lajos sgt. 23, Szeged 6722, Hungary

⁹ Department of Pharmacology, University of Nevada, Reno,
School of Medicine, Reno, NV, USA

systemic inflammation, innate immune activation, and gastrointestinal abnormalities [1]. Current estimates suggest that as many as 2.5 million individuals suffer from ME, with an annual productivity loss in excess of US\$9 billion in the USA alone, underscoring the importance of ME as a major public health concern both economically and socially.

Exactly what causes ME is unknown at this time; however, a number of potential triggers are associated with the development of the disease including physical trauma, emotional distress, infection, and chemical exposure [2–4]. Familial studies and genetic screening studies indicate that a genetic predisposition also plays an important role in the pathophysiology of ME [5–8]. Presently, there are no unique physical symptoms or reproducible biomarkers that can delineate this disease. For this reason, a diagnosis can only be made when an individual meets a series of inclusion and exclusion criteria, typically through a lengthy and expensive diagnostic process [9, 10]. Although the search for potential biomarkers has been one of the most active areas of ME research, little progress has been made in achieving this goal. Here, we report the progress in applying the immunosignature technology to this problem.

While not universally prevalent, a number of clinical observations such as natural killer (NK) cell dysfunction, viral reactivation, and inflammatory cytokine production have been consistently reported in the ME literature over the years and support an organic basis for this disease [11–15]. However, the mechanisms responsible for these observations remain elusive, but, if identified, this knowledge would lead to a greater understanding of ME pathology, potentially leading to effective treatment options.

Antibodies are glycoproteins, produced by B lymphocytes (B-cells) and plasma cells, in response to foreign molecules (antigens), such as those found in bacteria and viruses. As the central component of humoral immunity, they limit the spread of infection by binding to and neutralizing the pathogen or by activating other adaptive immune responses. B-cells also produce antibodies directed against self-antigens, but they are normally removed in the bone marrow early in their development. Although on rare occasion, however, this system fails, leading to autoimmunity.

Identifying the antigens to which antibodies react with during the course of a disease may lead to a greater understanding of the humoral immune response associated with that disease. For instance, ascertaining a dominant epitope may help in the development of an effective vaccine or identifying reactive antigens that are homologous to self-proteins may reveal autoimmune pathology [16]. Immunosignatures (IMS) are screens of serum antibodies using random peptide arrays, and this technique has been used successfully to identify biomarkers in diseases that are difficult to diagnose such as cancer, valley fever, and Alzheimer's disease [17–19].

In the present study, we utilized a microarray consisting of 125,000 random peptide sequences to screen the serum of healthy control subjects and those who present with symptoms

consistent with a diagnosis of ME. Our data identified an IMS that accurately delineated ME cases from controls with 92.9% specificity and 97.6% sensitivity. Additionally, Basic Local Alignment Search Tool (BLAST) searches suggest that these peptides have sequence homology primarily to human self-antigens and endogenous retroviral sequences, but also to a minor extent, viral and bacterial antigens. This proof-of-concept study potentially represents the first step toward a specific and sensitive diagnostic for ME and also may provide important knowledge regarding the pathophysiology of the disease.

Materials and Methods

Study Subjects

For this study, a total of 42 subjects were recruited from across the USA and Europe. ME cases consisted of 11 US and 10 European subjects and controls consisted of 12 US and 10 European subjects. Informed consent was obtained from each participant according to a human subjects protocol approved by the University of Nevada Biomedical Institutional Review Board (protocol B12-031). The cases identified as having ME were physician diagnosed and met Carruthers et al.'s criteria for ME as well as 1994 Fukuda et al.'s criteria for CFS [9, 10, 20].

Microarray

Serum samples from respective cases and controls were diluted 1:1 in glycerol and stored at -20°C until analyzed. The 125,000 random 12-mer peptide microarrays were manufactured according to the methods of Leguti et al. [21] and blocked in 0.5% BSA (Sigma, St. Louis, MO) and $1\times$ PBS, pH 7.2. Samples were diluted to 1:1000 in $1\times$ PBS, 0.5% BSA, and 0.05% Tween 20 pH 7.2 and exposed to the microarrays for 1 h at 37°C with gentle agitation. After 1 h, the arrays were washed in distilled water $3\times$ and incubated with 4 nM of Alexa Fluor 555 conjugated goat anti-human IgG (H & L) and 5 nM of Alexa Fluor 647 conjugated goat anti-human IgM heavy chain (Thermo Fisher) for 1 h at room temperature, then washed $3\times$ in distilled water and $1\times$ in 90% isopropyl alcohol, and dried in a centrifuge. Slides were scanned on an Innopsys 1100 scanner at 0.5- μm resolution, and TIFF images were aligned in GenePix 6.0.

Data Analysis

Peptide expression data were subjected to the following quality control steps. First, array images were evaluated for clearly identifiable spatial variation, including streaks and bubbles. Peptide array background values were subtracted from signal values in both Cy3 and Cy5 channels using simple

background subtraction. Before normalization, peptides with an incidence of high background values were filtered. Specifically, peptides having more than 50% incidence in either channel of negative background-corrected values (signal background) were excluded. The remaining raw values were normalized first within each array via the median method and then between arrays using the Aquantile method, with the *limma* package in R [22].

Normalized data were then averaged across replicated peptides and replicated samples. Peptides were again filtered after normalization and averaging for high incidence of low signal intensities with respect to background intensities. (These are seen as missing values in the data, as normalization includes a logarithmic transform that is not applicable to negative values.) Specifically, any peptide having more than 25% missing values for either cohort was excluded.

This final data set (103,385 peptides) was analyzed using the data mining algorithm Random Forest [23] in a progressive stepwise process of reduction using each respective peptide sequence as the predictive variable and subject status (ME case or control) as the target variable. For each iteration, 5000 random decision trees were built using one half the square root of N with a minimal of two parental nodes at each branch. Small classes were upweighted to equal the size of the largest target class and “out of bag” testing with replacement was employed to test the model. In the first step, the top 30% of peptides were selected and rescreened; then, the top 40% of peptides were rescreened. In the final step, multiple iterations were performed systematically, removing the least contributing peptides until the signature did not improve.

In order to potentially identify the biological antigens to which the synthetic random peptides represent, the penultimate iteration, consisting of 233 peptides, was searched against viral, bacterial, human, and endogenous retroviral proteins, each derived from the National Center for Biotechnology Information (NCBI) nr database using the ncbi-blast+ BLASTP protein sequence similarity search tool (v. 2.4.0). The virus protein database was produced by filtering nr for virus species with human hosts as recorded at NCBI Taxonomy. Similarly, the bacterial protein database was generated by restriction of nr to the subset of bacterial species identified within the PATRIC database to be associated with human hosts (<http://www.patricdb.org>). The human protein database contained those found in NCBI RefSeq. The HERVd protein database was generated by the combination of nr proteins self-identified in human endogenous retroviral lineages with a set of human endogenous retrovirus (HERV)-like proteins reported as proteins of *Homo sapiens* origin. BLAST parameters were set as follows: wordsize 2, window_size 15, threshold 16, PAM30 scoring matrix, gapopen 9, gapextend 1, evalue 1000, maximum reported alignments per high scoring pair (HSP) of query/subject (max_hsps) 1, and minimum query coverage by HSP percent (qcov) 34.

Additional BLAST output format options were set to record NCBI taxonomic identifiers (taxids) of proteins and the BLAST traceback operations (btop), a text string that encodes the alignment, mismatch, and gap information. Hits lacking any ungapped subalignment of five or more amino acid identities were identified using btop information and excluded from the analysis set. Species and genus taxa of subject proteins were mapped to each protein from the reported taxids with ETE Toolkit (<http://etetoolkit.org>; v3.0.0b35); a Python framework for phylogenetic tree analysis. In order to limit biasing as a result of protein size, we implemented a simple metric adjustment (Adj.), whereby the number of amino acids in a given protein was divided by the number of peptides having homology to that protein. Potentially conserved peptide motifs were investigated using the multiple sequence alignment tool Clustal X [24].

Results

Classification by Random Forest

In order to test whether differences exist between the antibody profiles of ME cases and controls, analysis was carried out using the Random Forest (RF) classification algorithm. The RF algorithm uses an ensemble of unpruned classification or regression trees produced through bootstrap sampling of the training data set and random feature selection in tree generation. Prediction is made by a majority vote of the predictions of the ensemble. The strength of the analysis was evaluated by out of bag sampling with replacement of the original data. RF is an attractive method since it handles both discrete and continuous data, it accommodates and compensates for missing data, and it is invariant to monotonic transformations of the input variables. The RF algorithm is well suited for peptide microarray analysis in that it can handle highly skewed values well and weighs the contribution of a given peptide according to its relatedness with others.

Through multiple iterations of RF processing, we identified a signature of 25 peptides that was able to identify ME cases from controls with 92.9% specificity and 97.6% sensitivity (Table 1). Each peptide was ranked according to its contribution to the signature, with the top peptide being ranked at 100 and subsequent peptides ranked relative to this peptide. The relative contribution of these 25 peptides and their sequence are given in Fig. 1. We conclude that, at least based on the analysis with this small sample set, IMS can distinguish ME from non-ME samples.

Homology BLAST Search and Sequence Alignment

In order to potentially identify the biological antigens to which the synthetic random peptides may represent, we developed

Table 1 Results of 21 ME cases and 21 controls each screened for reactivity with IgG and IgM

Actual class	Total class	Percent class	Predicted classes	
			Control <i>N</i> = 44	ME case <i>N</i> = 44
Control	42	97.62%	41	1
ME case	42	92.86%	3	39
Total	84			
Average		95.24%		
Overall % correct		95.24%		
Specificity		92.86%		
Sensitivity/recall		97.62%		
Precision		93.18%		
F1 statistic		95.35%		

an analytical pipeline and used this to search the 233 peptides from our penultimate RF iteration against the human proteome. Additionally, we have previously reported that gastrointestinal plasmacytoid dendritic cells (pDCs) produce proteins that are consistent with human endogenous retroviral sequences [25]; therefore, we also included these sequences in our search. Finally, we used this pipeline to search for homology to bacterial and viral antigens of pathogens known to infect humans.

Our initial analysis identified over 5000 human protein sequences that met our search criteria. When filtered to limit those proteins that were identified by four or more random peptides, this number was reduced to 166 proteins. In an attempt prevent overrepresentation of larger

proteins, which have a greater chance of having homologous sequences to a given random peptide, we used the simple metric of dividing the number of amino acids of the protein by the number of peptides that were homologous to that protein. The top 30 human proteins, adjusted for size, are given in Table 2. Among the likely most relevant human proteins identified in this search were proteins involved in mitochondrial function (AMACR, ETFDH, SLC25A40), lipid metabolism (AGK, ACOXL, CEL SEC23A), neurological function (APBA3, ASIC1, GABRB3, STAC), and immune responses (CD274, LGMN, MX1, MX2).

Previous studies have proposed that HERV elements may be associated with neurological diseases including multiple sclerosis [26], amyotrophic lateral sclerosis [27], and schizophrenia [28]. With this, and our previous studies in mind, we included HERV sequences in our homology search. Nine HERV sequences were identified with sequence homology to at least two of our top 233 random peptide sequences; the most relevant HERV sequence showed homology to seven of the 233 peptides (Table 3). Importantly, the seven sequences were not randomly represented throughout the HERV sequence but largely converged the same position in the protein, as revealed by Clustal X alignment (Fig. 2). Further analysis showed that this conserved motif is well represented in 40 of the 233 random peptides (Fig. 3), suggesting that this motif significantly contributes to the observed IMS.

Immunoreactivity to a given synthetic random peptide may be the result of cross-reactivity to pathogen-derived antigens encountered during an infection. To explore this possibility, we surveyed our top 233 random peptides against the proteomes of bacteria and viruses known to infect humans (Table 4). As before, the proteins were filtered to limit those that were hit by multiple peptides; however, the threshold was reduced to three peptides. When adjusted for protein size, the six most significant viral proteins with sequence homology to our random peptides were the gp120 protein of HIV (six hits); followed by the polyprotein of GB virus Ccpz (three hits); the envelope glycoprotein I of Human herpesvirus 2 (four hits); the phosphoprotein of canine distemper virus (four hits); the RNA-dependent RNA polymerase, rodent paramyxovirus (three hits); and finally, the outer capsid protein of porcine rotavirus C (three hits). When adjusted for protein size, the most significant bacterial peptides with sequence homology to our top random peptides were a hypothetical protein from *Serratia marcescens* (four hits), a diaminopimelate aminotransferase from *Paenibacillus senegalensis* (five hits), the peptidase M16 of *Anaerofustis stercorihominis* (three hits), the type IV secretion protein Rhs of *Hafnia alvei* (three hits), and the SusC/RagA family TonB-linked outer membrane protein of *Bacteroides nordii* (three hits).

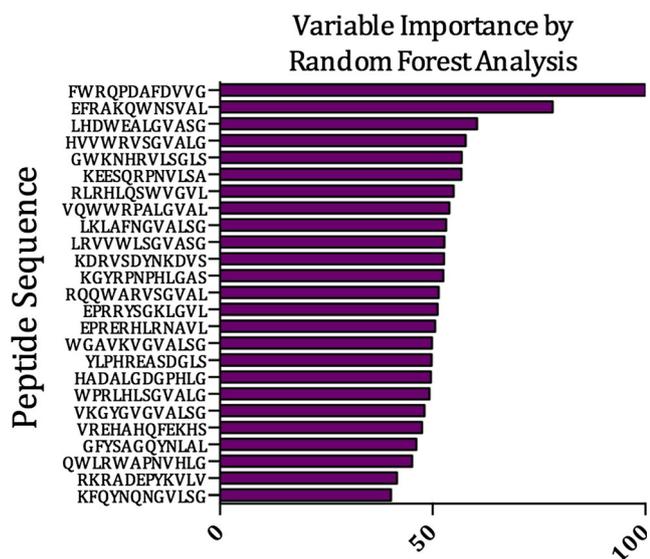
**Fig. 1** Random forest prediction. Horizontal bars represent the relative importance that each random peptide contributes to the final diagnostic signature

Table 2 Number of peptides homologous to a respective human protein sequences

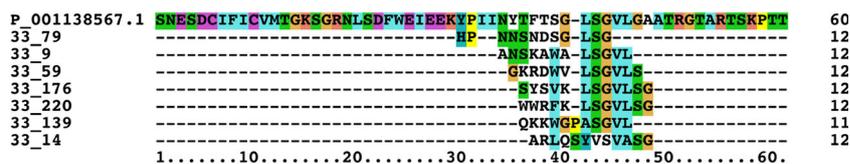
Peptides	Accession	Symbol	Description	Length	Adj.
4	Q8NAP1	GATS	Stromal antigen 3 opposite strand	163	40.8
13	NP_033173	SEC23A	Sec23 homolog A	765	58.8
10	AKI70819	ETFDH	Electron transfer flavoprotein Dehydrogenase	617	61.7
9	NP_073150	EBF2	Early B-cell factor 2	575	63.9
5	AAH27322	SLC25A40	Solute carrier family 25 Member 40	338	67.6
5	EAX02872	ARMCX4	Armadillo repeat containing, X-linked 4	360	72
4	EAW58763	CD274	CD274 molecule	290	72.5
7	NP_056530	PLA2G3	Phospholipase A2 group III	509	72.7
7	Q6PJ69	TRIM65	Tripartite motif containing 65	517	73.9
4	NP_001103408	C6orf136	Chromosome 6 open reading frame 136	315	78.75
4	CAG46638	HMOX2	Heme oxygenase 2	316	79
7	NP_004877	APBA3	Amyloid beta precursor Protein binding family A member 3	575	82.1
9	P19835	CEL	Carboxyl ester lipase	753	83.7
7	NP_001171517	MX1	MX dynamin-like GTPase 1	662	94.6
4	ABQ59031	AMACR	Alpha-methylacyl-CoA racemase	382	95.5
4	NP_003140	STAC	SH3 and cysteine-rich domain	402	100.5
6	NP_001307526	DDX5	DEAD-box helicase 5	614	102.3
4	NP_060708	AGK	Acylglycerol kinase	422	105.5
5	NP_001086	ASIC1	Acid sensing ion Channel subunit 1	528	105.6
7	NP_758441	CTAGE1	Cutaneous T-cell Lymphoma-associated antigen 1	745	106.4
5	CAG33352	CCT2	Chaperonin containing TCP1 subunit 2	535	107
4	CAG33687	LGMN	Legumain	433	108.3
10	NP_006217	PLCL1	Phospholipase C-like 1	1095	109.5
5	NP_000449	HNF1B	HNF1 homeobox B	557	111.4
4	NP_068712	GABRB3	Gamma-aminobutyric acid type A receptor Beta3 subunit	473	118.25
6	P20592	MX2	MX dynamin-like GTPase 2	715	119.7
7	NP_060868	CACNA2D3	Calcium voltage-gated channel auxiliary subunit alpha2delta 3	1091	121.2
5	P30825	SLC7A1	Solute carrier family 7 member 1	629	125.8
4	NP_004557	PFKFB3	6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	520	130
4	XP_011509718	ACOXL	Acyl-CoA oxidase-like	547	136.8

Numerous other human pathogens were identified that contained homologous sequences to our random peptides but were excluded because of our adjusting metric (data not shown).

Table 3 Number of peptides homologous to respective human endogenous retroviral sequences

Peptides	Accession	Symbol	Description	Length	Adj.
7	NP_001138567	HHLA1	HERV-H LTR-associating protein 1 precursor	531	75.8
3	P61566	ERVK-24	Endogenous retrovirus group K member 24	588	196
3	AAY87455	ERVK-6	Env type 1, partial	603	201
4	AAM81188	HRV-5	Pol protein, partial	863	215.7
3	P61570	ERVK-25	Endogenous retrovirus group K member 25	661	220.3
3	P61565	ERVK-21	Endogenous retrovirus group K member 21	698	232.6
2	P60507	ERVFC1	Endogenous retrovirus group FC1 env polyprotein	584	292
2	Q14264	ERV3-1	Endogenous retrovirus Group 3 member 1	604	302
2	ABB52637	ERV PABLB-1	Endogenous retrovirus group PABLB member 1	665	332.5

Fig. 2 Clustal X alignment of seven random peptides homologous to HERV-H LTR-associating protein 1 precursor



Discussion

In this report, we present a “proof-of-concept” study where random peptide arrays show utility in delineating ME cases from healthy controls. The ultimate goal of this work is the development of a non-subjective clinical tool for diagnosing patients with ME. To this end, we utilized a random peptide array, which has previously produced IMS for other chronic and complicated diseases that are difficult to diagnose such as cancer, valley fever, and Alzheimer’s disease [17–19]. Given the complexity and the size of the data set, we elected to use the machine learning data mining algorithm Random Forest to identify potential candidates that may lead to a diagnostic signature. Using out of bag testing with replacement, our model was able to predict cases and controls with 92.9% specificity and 97.6% sensitivity using 25 peptides. Finally, we developed an analytical pipeline to BLAST these peptides against the human, HERV, virus, and bacterial proteomes for sequence homology, in order to explain the underpinnings of

the IMS. This pilot study supports the premise that immunosignatures represent a viable approach to achieve our overarching goal of developing a diagnostic tool and further potentially identify naturally occurring antigens to these antibodies.

A number of studies have attempted to identify a reproducible biomarker for ME [29–33]. In particular, serum or plasma cytokine and chemokine analyses have shown promise, in that several investigators, including our group, have reported clear differences when comparing ME cases to healthy controls [11, 15, 34–36]. Additionally, cytokine differences may provide valuable information regarding the pathophysiology of the disease. For instance, previous studies have suggested that ME is characterized by a Th2 shift [37, 38], an observation that may explain the prevalence of persistent viral infections associated with this disease [39]. However, most cytokines that are produced in response to innate immune activation, as is seen with ME, are not consistently expressed [40]. In contrast, serum antibodies are much more stable. For example, most IgG subclasses have half-lives of more than 20 days [41].

Immunosignatures have been used successfully to provide understanding to the pathophysiology of chronic diseases. For example, Restrepo et al. reported that plasma antibodies from subjects with Alzheimer’s disease (AD) could be used to provide an IMS that can distinguish AD cases from non-AD controls reproducibly over time [42]. It was also shown that eight of the 50 signature random peptides have the ability to react with antibodies initially raised against native amyloid- β [19], a protein shown to be significantly involved in Alzheimer’s disease [43].

These observations raise the possibility that an IMS may provide clues to the pathophysiology of ME. However, in contrast to AD, there are no proteins known to be ubiquitous in the pathological process of ME with which to test. Although we have identified a number of peptides that accurately identify ME cases from controls, divining naturally occurring homologous peptides is challenging. An antibody typically covers approximately 15 amino acids of its cognate epitope; however, only about five or six amino acids contribute to the ΔG° of antigen/antibody binding [44, 45] and these amino acids may not necessarily be contiguous. Therefore, the natural antigen is likely to be very short and may contain gaps in the primary sequence. In an attempt to overcome these obstacles, we developed a custom analytical pipeline to conduct BLAST searches against the NCBI human protein

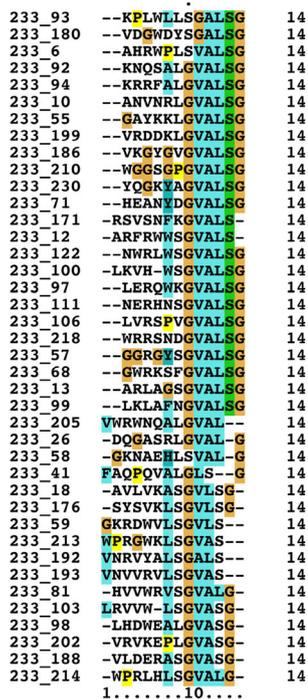


Fig. 3 Clustal X alignment. Clustal X alignment of 40 random peptides showing the largely conserved motif of GVALSG

Table 4 Number of peptides homologous to a respective viral and bacterial protein sequences

Peptides	Accession	Symbol	Description	Length	Adj.
Viruses					
6	CAD87195	Q70QU7	gp120 protein, human immunodeficiency virus 1	198	33
3	ADL29714	E0ADF8	Polyprotein GB virus Ccpz	242	80.7
4	AMB66322	P06764	Envelope glycoprotein I human herpesvirus 2	372	93
4	AJO72800	P06940	Phosphoprotein canine distemper virus	507	126.8
3	BAO04373	U6C7N8	RNA-dependent RNA polymerase rodent Paramyxovirus	549	183
3	AJL35076	A0A0C5B2I7	Outer capsid protein Porcine rotavirus C	733	244.3
Bacteria					
4	WP_033638106	None	Hypothetical protein [<i>Serratia marcescens</i>]	284	71
5	WP_010273745	None	diaminopimelate aminotransferase [<i>Paenibacillus senegalensis</i>]	397	79.4
3	WP_007050391	B1C8L1	Peptidase M16 [<i>Anaerofustis stercorihominis</i>]	422	140.7
3	WP_004091469	G9Y4S5	Type IV secretion protein Rhs [<i>Hafnia alvei</i>]	644	214.7
3	WP_007483701	I9H2D3	SusC/RagA family TonB-linked outer membrane protein [<i>Bacteroides nordii</i>]	1046	348.7

database. In addition to annotated human proteins, we also queried against HERV proteins as well as proteins from bacteria and viruses known to infect humans. As it was probable that very large proteins would be more likely to be identified over smaller proteins by random chance, we implemented a metric to adjust for this issue. Interestingly and somewhat unexpectedly, we identified a number of protein hits in the human database. Previous studies by Fluge et al. showed that anti-CD20 B-cell depleting drug, rituximab, showed efficacy in treating subjects with ME [46, 47]. Although the mechanism responsible for this observation remains to be elucidated, it does suggest that self-reactive antibodies may contribute to the pathophysiology of this disease. Indeed, several studies have reported self-reactive antibodies in subjects with ME [48–50] so our results are potentially consistent with this supposition.

Previously, we reported that gut-associated pDCs in subjects with ME were immunoreactive to antibodies that react with endogenous retroviral proteins [25]. Other studies have also reported retroviral protein sequences in subjects with neurological and autoimmune disease; however, the meaning of these observations has yet to be resolved [28, 51, 52]. Nonetheless, these sequences, if uniquely expressed in a disease state, may prove to be a useful biomarker. We thus included the HERVd databases in our query space and observed seven of our significant peptide-1 displayed sequence homology to the HERV-H LTR-associating protein 1 precursor (HHLA1). If dispersed through the protein, the probability of seven random peptides hitting this sequence would be exceedingly small. However, upon further examination, it was discovered that all seven peptides represented a largely conserved sequence (LSGVLS) in the HERV protein. A similar and overlapping conserved sequence motif (GVALSG) was

observed in at least 40 of the 233 top peptides identified by our RF analysis. This observation raises two important issues. First, the discovery of this conserved peptide motif may represent a critical discovery in resolving the pathophysiology of ME, assuming it is confirmed in other cohorts and it is shown to be unique to ME. Secondly, because this motif is short and present in many pathogens, we cannot say with absolute certainty that we have identified the naturally occurring antigen that gave rise to the antibodies that react with this motif. When this sequence is considered in isolation, we have observed it within several other proteins, in particular, the bacteria genus *Burkholderia* and also in the human protein calcium voltage-gated channel protein CACNA2D3 (Table 2). Further studies will be required to identify with greater certainty the native antigen to this conserved motif.

Lastly, we BLASTed the 233 random peptides identified by RF against the proteomes of viruses and bacteria known to infect humans. The most prevalent viral hit was to the gp120 protein of human immunodeficiency virus 1. This sequence was homologous to the conserved motif; therefore, it is likely the result of cross-reactivity to antibodies raised to HERVs or another similar sequence. Of the bacterial hits that have been previously associated with ME, the type IV secretion protein Rhs of *H. alvei* was identified in our search. *H. alvei* is Gram-negative, facultative anaerobic intestinal bacteria. A previous study by Maes et al. reported that ME cases have elevated serum IgA and IgM antibodies, to *H. alvei*, that likely result from intestinal bacterial translocation [53]. Given the diversity of gastrointestinal bacteria, it may be possible that antibodies raised to translocated bacterial products could potentially cross-react with self-proteins. Indeed, a number of studies have identified gastrointestinal comorbidity and/or an altered gut microbiome as a potential associating factor with ME

[54–58]. However, at this point, it has yet to be determined if these observations are cause or effect.

In conclusion, the data presented in this report represents a proof-of-concept study that random peptide arrays show utility in delineating ME cases from healthy controls. Additionally, our study has identified a conserved peptide motif that is preferentially recognized by serum antibodies in a large number of ME cases over that of healthy controls. This study warrants further investigations using additional ME cohorts as well as cohorts of subjects with other chronic diseases with overlapping symptomology.

Acknowledgements These studies were supported by awards from the National Institutes of Health (R01AI078234) and the National Institute of General Medical Sciences (GM103440), from the National Institutes of Health, and from funds donated by the ME/CFS community. We are additionally grateful to all our study subjects, whose selfless participation made this research possible.

Compliance with Ethical Standards

Informed consent was obtained from each participant according to a human subjects protocol approved by the University of Nevada Biomedical Institutional Review Board (protocol B12-031).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Klimas NG, Koneru AO (2007) Chronic fatigue syndrome: inflammation, immune function, and neuroendocrine interactions. *Curr Rheumatol Rep* 9:482–487
- Bell KM, Cookfair D, Bell DS, Reese P, Cooper L (1991) Risk factors associated with chronic fatigue syndrome in a cluster of pediatric cases. *Rev Infect Dis* 13(Suppl 1):S32–S38
- Racciatti D, Vecchiet J, Ceccomancini A, Ricci F, Pizzigallo E (2001) Chronic fatigue syndrome following a toxic exposure. *Sci Total Environ* 270:27–31
- Jason LA, Porter N, Hunnell J, Rademaker A, Richman JA (2011) CFS prevalence and risk factors over time. *J Health Psychol* 16:445–456. doi:10.1177/1359105310383603
- Albright F, Light K, Light A, Bateman L, Cannon-Albright LA (2011) Evidence for a heritable predisposition to chronic fatigue syndrome. *BMC Neurol* 11:62. doi:10.1186/1471-2377-11-62
- Schlauch KA, Khaiboullina SF, De Meirleir KL, Rawat S, Petereit J, Rizvanov AA, Blatt N, Mijatovic T et al (2016) Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Transl Psychiatry* 6:e730. doi:10.1038/tp.2015.208
- Rajeevan MS, Dimulescu I, Murray J, Falkenberg VR, Unger ER (2015) Pathway-focused genetic evaluation of immune and inflammation related genes with chronic fatigue syndrome. *Hum Immunol* 76:553–560. doi:10.1016/j.humimm.2015.06.014
- Buchwald D, Herrell R, Ashton S, Belcourt M, Schmalting K, Sullivan P, Neale M, Goldberg J (2001) A twin study of chronic fatigue. *Psychosom Med* 63:936–943
- Carruthers BM, van de Sande MI, De Meirleir KL, Klimas NG, Broderick G, Mitchell T, Staines D, Powles AC et al (2011) Myalgic encephalomyelitis: International Consensus Criteria. *J Intern Med* 270:327–338. doi:10.1111/j.1365-2796.2011.02428.x
- Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A (1994) The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group *Annals of internal medicine* 121:953–959
- Fletcher MA, Zeng XR, Barnes Z, Levis S, Klimas NG (2009) Plasma cytokines in women with chronic fatigue syndrome. *J Transl Med* 7:96. doi:10.1186/1479-5876-7-96
- Kerr JR, Tyrrell DA (2003) Cytokines in parvovirus B19 infection as an aid to understanding chronic fatigue syndrome. *Curr Pain Headache Rep* 7:333–341
- Nakamura T, Schwander SK, Donnelly R, Ortega F, Togo F, Broderick G, Yamamoto Y, Chermiack NS et al Cytokines across the night in chronic fatigue syndrome with and without fibromyalgia. *Clinical and vaccine immunology : CVI*. doi:10.1128/CVI.00379-09
- Patarca R (2001) Cytokines and chronic fatigue syndrome. *Ann N Y Acad Sci* 933:185–200
- Khaiboullina SF, DeMeirleir KL, Rawat S, Berk GS, Gaynor-Berk RS, Mijatovic T, Blatt N, Rizvanov AA et al (2015) Cytokine expression provides clues to the pathophysiology of Gulf War illness and myalgic encephalomyelitis. *Cytokine* 72:1–8. doi:10.1016/j.cyto.2014.11.019
- Richer J, Johnston SA, Stafford P (2015) Epitope identification from fixed-complexity random-sequence peptide microarrays. *Molecular & cellular proteomics : MCP* 14:136–147. doi:10.1074/mcp.M114.043513
- Chapoval AI, Legutki JB, Stafford P, Trebukhov AV, Johnston SA, Shoikhet YN, Lazarev AF (2015) Immunosignature: serum antibody profiling for cancer diagnostics. *Asian Pac J Cancer Prev* 16:4833–4837
- Navalkar KA, Johnston SA, Woodbury N, Galgiani JN, Magee DM, Chicacz Z, Stafford P (2014) Application of immunosignatures for diagnosis of valley fever. *Clinical and vaccine immunology : CVI* 21:1169–1177. doi:10.1128/CVI.00228-14
- Restrepo L, Stafford P, Magee DM, Johnston SA (2011) Application of immunosignatures to the assessment of Alzheimer's disease. *Ann Neurol* 70:286–295. doi:10.1002/ana.22405
- Carruthers B, Jain AK, De Meirleir K, Peterson D, Klimas N, Lerner A, Bested A, Flor-Henry P et al (2003) Myalgic encephalomyelitis/chronic fatigue syndrome: clinical working case definition, diagnostic and treatment protocols. *J Chronic Fatigue Syndrome* 11:1–12
- Legutki JB, Zhao ZG, Greving M, Woodbury N, Johnston SA, Stafford P (2014) Scalable high-density peptide arrays for comprehensive health monitoring. *Nat Commun* 5:4785. doi:10.1038/ncomms5785
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3. doi:10.2202/1544-6115.1027
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32. doi:10.1023/A:1010933404324
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403–405
- De Meirleir K, Khaiboullina S, Fremont M, Hulstaert J, Rizvanov A, Palotás A, Lombardi V (2013) Plasmacytoid dendritic cells in the duodenum of individuals diagnosed with myalgic encephalomyelitis are uniquely immunoreactive to antibodies to human endogenous retroviral proteins in vivo 27
- Dolei A, Garson JA, Arru G, Clerici M, Germa R, Marche PN, Perron H (2014) Multiple sclerosis-associated retrovirus and related human endogenous retrovirus-W in patients with multiple sclerosis. *J Neuroimmunol* 266:87–88. doi:10.1016/j.jneuroim.2013.11.009

27. Li W, Lee MH, Henderson L, Tyagi R, Bachani M, Steiner J, Campanac E, Hoffman DA et al (2015) Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med* 7:307ra153. doi:10.1126/scitranslmed.aac8201
28. Karlsson H, Schroder J, Bachmann S, Bottmer C, Yolken RH (2004) HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. *Mol Psychiatry* 9:12–13. doi:10.1038/sj.mp.4001439
29. Fukuda S, Nojima J, Motoki Y, Yamaguti K, Nakatomi Y, Okawa N, Fujiwara K, Watanabe Y et al (2016) A potential biomarker for fatigue: oxidative stress and anti-oxidative activity. *Biol Psychol* 118:88–93. doi:10.1016/j.biopsycho.2016.05.005
30. Frith J, Zalewski P, Klawe JJ, Pairman J, Bitner A, Tafil-Klawe M, Newton JL (2012) Impaired blood pressure variability in chronic fatigue syndrome—a potential biomarker. *QJM: monthly journal of the Association of Physicians* 105:831–838. doi:10.1093/qjmed/hcs085
31. Fletcher MA, Rosenthal M, Antoni M, Ironson G, Zeng XR, Barnes Z, Harvey JM, Hurwitz B et al (2010) Plasma neuropeptide Y: a biomarker for symptom severity in chronic fatigue syndrome. *Behav Brain Funct* 6:76. doi:10.1186/1744-9081-6-76
32. Steinau M, Unger ER, Vernon SD, Jones JF, Rajeevan MS (2004) Differential-display PCR of peripheral blood for biomarker discovery in chronic fatigue syndrome. *J Mol Med* 82:750–755. doi:10.1007/s00109-004-0586-4
33. Vernon SD, Unger ER, Dimulescu IM, Rajeevan M, Reeves WC (2002) Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome. *Dis Markers* 18:193–199
34. Broderick G, Fuite J, Kreitz A, Vernon SD, Klimas N, Fletcher MA (2010) A formal analysis of cytokine networks in chronic fatigue syndrome. *Brain Behav Immun* 24:1209–1217. doi:10.1016/j.bbi.2010.04.012
35. Broderick G, Katz BZ, Fernandes H, Fletcher MA, Klimas N, Smith FA, O’Gorman MR, Vernon SD et al (2012) Cytokine expression profiles of immune imbalance in post-mono-nucleosis chronic fatigue. *J Transl Med* 10:191. doi:10.1186/1479-5876-10-191
36. Smylie AL, Broderick G, Fernandes H, Razdan S, Barnes Z, Collado F, Sol C, Fletcher MA et al (2013) A comparison of sex-specific immune signatures in Gulf War illness and chronic fatigue syndrome. *BMC Immunol* 14:29. doi:10.1186/1471-2172-14-29
37. Craddock TJ, Fritsch P, Rice MA Jr, del Rosario RM, Miller DB, Fletcher MA, Klimas NG, Broderick G (2014) A role for homeostatic drive in the perpetuation of complex chronic illness: Gulf War illness and chronic fatigue syndrome. *PLoS One* 9:e84839. doi:10.1371/journal.pone.0084839
38. Torres-Harding S, Sorenson M, Jason LA, Maher K, Fletcher MA (2008) Evidence for T-helper 2 shift and association with illness parameters in chronic fatigue syndrome (CFS). *Bull IACFS ME* 16:19–33
39. Bansal AS, Bradley AS, Bishop KN, Kiani-Alikhan S, Ford B (2012) Chronic fatigue syndrome, the immune system and viral infection. *Brain Behav Immun* 26:24–31. doi:10.1016/j.bbi.2011.06.016
40. Stringer EA, Baker KS, Carroll IR, Montoya JG, Chu L, Maecker HT, Younger JW (2013) Daily cytokine fluctuations, driven by leptin, are associated with fatigue severity in chronic fatigue syndrome: evidence of inflammatory pathology. *J Transl Med* 11:93. doi:10.1186/1479-5876-11-93
41. Brekke OH, Sandlie I (2003) Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat Rev Drug Discov* 2:52–62. doi:10.1038/nrd984
42. Restrepo L, Stafford P, Johnston SA (2013) Feasibility of an early Alzheimer’s disease immunosignature diagnostic test. *J Neuroimmunol* 254:154–160. doi:10.1016/j.jneuroim.2012.09.014
43. Sadigh-Eteghad S, Sabermarouf B, Majidi A, Talebi M, Farhoudi M, Mahmoudi J (2015) Amyloid-beta: a crucial factor in Alzheimer’s disease. *Med Princ Pract* 24:1–10. doi:10.1159/000369101
44. Prasad L, Sharma S, Vandonselaar M, Quail JW, Lee JS, Waygood EB, Wilson KS, Dauter Z et al (1993) Evaluation of mutagenesis for epitope mapping. Structure of an antibody-protein antigen complex. *J Biol Chem* 268:10705–10708
45. Schwarz FP, Tello D, Goldbaum FA, Mariuzza RA, Poljak RJ (1995) Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies. *European journal of biochemistry/FEBS* 228:388–394
46. Fluge O, Bruland O, Risa K, Storstein A, Kristoffersen EK, Sapkota D, Naess H, Dahl O et al (2011) Benefit from B-lymphocyte depletion using the anti-CD20 antibody rituximab in chronic fatigue syndrome. A double-blind and placebo-controlled study. *PLoS One* 6:e26358. doi:10.1371/journal.pone.0026358
47. Fluge O, Mella O (2009) Clinical impact of B-cell depletion with the anti-CD20 antibody rituximab in chronic fatigue syndrome: a preliminary case series. *BMC Neurol* 9:28. doi:10.1186/1471-2377-9-28
48. Tanaka S, Kuratsune H, Hidaka Y, Hakariya Y, Tatsumi KI, Takano T, Kanakura Y, Amino N (2003) Autoantibodies against muscarinic cholinergic receptor in chronic fatigue syndrome. *Int J Mol Med* 12:225–230
49. Konstantinov K, von Mikecz A, Buchwald D, Jones J, Gerace L, Tan EM (1996) Autoantibodies to nuclear envelope antigens in chronic fatigue syndrome. *J Clin Invest* 98:1888–1896. doi:10.1172/JCI118990
50. Wheatland R (2005) Chronic ACTH autoantibodies are a significant pathological factor in the disruption of the hypothalamic-pituitary-adrenal axis in chronic fatigue syndrome, anorexia nervosa and major depression. *Med Hypotheses* 65:287–295. doi:10.1016/j.mehy.2005.02.031
51. Sicut J, Sutkowski N, Huber BT (2005) Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. *J Rheumatol* 32:1821–1831
52. Moyes DL, Goris A, Ban M, Compston A, Griffiths DJ, Sawcer S, Venables PJ (2008) HERV-K113 is not associated with multiple sclerosis in a large family-based study. *AIDS Res Hum Retrovir* 24:363–365. doi:10.1089/aid.2007.0196
53. Maes M, Mihaylova I, Leunis JC (2007) Increased serum IgA and IgM against LPS of enterobacteria in chronic fatigue syndrome (CFS): indication for the involvement of gram-negative enterobacteria in the etiology of CFS and for the presence of an increased gut-intestinal permeability. *J Affect Disord* 99:237–240. doi:10.1016/j.jad.2006.08.021
54. Hanevik K, Kristoffersen EK, Sornes S, Morch K, Naess H, Rivenes AC, Bodtke JE, Hausken T et al (2012) Immunophenotyping in post-giardiasis functional gastrointestinal disease and chronic fatigue syndrome. *BMC Infect Dis* 12:258. doi:10.1186/1471-2334-12-258
55. Morris G, Berk M, Carvalho AF, Caso JR, Sanz Y, Maes M (2016) The role of microbiota and intestinal permeability in the pathophysiology of autoimmune and neuroimmune processes with an emphasis on inflammatory bowel disease type 1 diabetes and chronic fatigue syndrome. *Curr Pharma Des* (in press)
56. Navaneetharaja N, Griffiths V, Wileman T, Carding SR (2016) A role for the intestinal microbiota and virome in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)? *J Clin Med* 5:55. doi:10.3390/jcm5060055
57. Jackson ML, Butt H, Ball M, Lewis DP, Bruck D (2015) Sleep quality and the treatment of intestinal microbiota imbalance in chronic fatigue syndrome: a pilot study. *Sleep Sci* 8:124–133. doi:10.1016/j.slsci.2015.10.001
58. Fremont M, Coomans D, Massart S, De Meirleir K (2013) High-throughput 16S rRNA gene sequencing reveals alterations of intestinal microbiota in myalgic encephalomyelitis/chronic fatigue syndrome patients. *Anaerobe* 22:50–56. doi:10.1016/j.anaerobe.2013.06.002