Immune Responses Are Characterized by Specific Shared Immunoglobulin Peptides That Can Be Detected by Proteomic Techniques*\[S\]

Martijn M. VanDuijn, Lennard J. M. Dekker, L. Zeneyedpour, Peter A. E. Silleviss Smitt, and Theo M. Luider\[1\]

From the Department of Neurology, Erasmus MC, Rotterdam 3000 CA, The Netherlands

In the adaptive immune response, immunoglobulins develop that bind specifically to the antigens to which the organism was exposed. Immunoglobulins may bind to known or unknown antigens in a variety of diseases and have been used in the past to identify novel antigens for use as a biomarker. We propose that the immunoglobulins themselves could also be used as biomarkers in antibody-mediated disease. In this proteomic study, rats were immunized with one of two purified antigens, and immunoglobulins from pre- and postimmune sera were analyzed with nano-LC coupled mass spectrometry. It was found that the two treatment groups could be distinguished based on cluster analysis of the immunoglobulin peptides from the immune sera. In addition, we identified 684 specific peptides that were differentially present in one of the two treated groups. We could find an amino acid sequence for 44% of the features in the mass spectra by combining database-driven and de novo sequencing techniques. The latter were essential for sequence identification, as the more common database-driven approach suffers from a poor representation of immunoglobulins in the available databases. Our data show that the development of immunoglobulins during an immune response is not a fully random process, but that instead selection pressures exist that favor the best binding amino acid sequences, and that this selection is shared between different animals. This finding implies that immunoglobulin peptides could indeed be a powerful and easily accessible class of biomarkers.

When an immune system is exposed to an antigen, it initiates a specific response to counter the perceived threat. Such a response may consist of cellular and humoral components, including the development and expression of immunoglobulin proteins that have affinity for the antigen. Immunoglobulins, or antibodies, are a family of proteins that consist of conserved constant domains that are important in determining the effector function of the antibody and variable domains that contain hypervariable complementarity determining regions (CDR).\[2\]

The CDRs, of which there are three in the immunoglobulin light chain, and three in the heavy chain, form the binding surface for the antigen which evoked the immune response. As such, the CDRs carry the specificity of an immunoglobulin against that antigen.

CDRs form during B-cell development in the immune response. A collection of germline genes form a library of components on which the final immunoglobulin sequence will be based. The immunoglobulin genes rearrange, thus selecting a particular combination of germline genes. The CDR1 and CDR2 regions stem from the germline V-genes and can be further modified by somatic mutations. The CDR3 of the light chain lies at the junction of a V-gene and a J-gene and accumulates more variability due to the possible gene combinations. In addition, the junction between V and J gene can be modified by insertions and deletions. Such variation is enhanced even more in the CDR3 of the heavy chain, as it lies on the junction between three gene segments, V, D, and J. The enhanced diversity in the CDR3 of the heavy chain also makes it the most specific part for antigen recognition (1).

Estimates on the total possible diversity in immunoglobulins vary depending on the assumptions being made, ranging from 10\[12\] to more than 10\[50\] (2, 3). Despite the large range, it is sufficiently clear that sequence variation is large, and based on these numbers one could statistically assume that antibody sequences found in a person must be unique, and that such antibodies will not be found in other subjects. However, evidence is now emerging from the literature that this assumption is incorrect, and that one should also consider that antibodies are subjected to selection pressures after rearrangement and affinity maturation. These pressures drive developing B cells and antibodies in convergent directions when exposed to particular antigens. For some antigens, this has been described as repertoire bias, where particular genes from the germline repertoire are favored in the panel of antibodies that is raised during the immune process (4–7). Moreover, similar CDR3 sequences were found in different subjects in a study with tetanus toxin immunizations in human, and a sequencing project in zebra fish revealed identical CDR3 sequences shared among several fish in an untreated population (8, 9). The occurrence of antigen-specific antibody sequences in multiple individuals opens the possibility that these sequences could be used as biomarkers for clinically interesting immune processes. While immunoglobulins were used before to find antigens for biomarker use, the use of the immunoglobulin sequences themselves as markers for a disease would be a novel method (10). Auto-
immune diseases could be well-suited for such an approach, but also cancer is known to result in an antibody response (11, 12). Thus, specific antibody sequences could be of diagnostic and prognostic use in such diseases. While serum is a logical source of immunoglobulins, additional biofluids may be of interest as well. Cerebrospinal fluid is known to contain immunoglobulins with relevance for diseases like multiple sclerosis and paraneoplastic syndromes, and immunoglobulins are present in other biofluids as well. Immunoglobulins from such compartments may be enriched for a specific antigen as compared with whole IgG from serum, and immunoglobulins from such sources are therefore also of great interest.

To test the feasibility of antibody sequences as biomarkers, we designed a pilot study under controlled and optimized conditions to test whether we can detect and identify specific antibody sequences after immunizations with purified antigens. Two model antigens were selected for this study: dinitrophenol (DNP) as an example of a very small hapten that exposes a minimal diversity of epitopes to the immune system, and HuD, a neuronal RNA-binding protein that is involved in paraneoplastic neurological syndromes (13). Thus, HuD is an example of a clinically relevant protein involved in an antibody-associated pathology. Both antigens were injected repeatedly in a group of outbred rats, and serum samples were taken before and after the immunization series. After purification of IgG, or affinity purification against the appropriate antigen, the immunoglobulins were digested by trypsin and measured by LC-MS, followed by analysis for fragments that overlapped between the samples.

EXPERIMENTAL PROCEDURES

Antigens—HuD with a His<sub>6</sub> tag was expressed in Escherichia coli BL21(DE3)/pLysS cells after induction with IPTG (14). HuD was purified from a cell lysate on an AKTA Prime system by passing it over a nickel column (HiTrap FF), desalting, and further separation on a cation exchange column (HiTrap SP; GE Healthcare, Diegem, Belgium). The product buffer was changed to PBS on a desalting column for the immunizations. DNP was acquired commercially covalently bound to the carrier proteins keyhole limpet hemocyanine (KLH) and bovine serum albumin (BSA) (US Biological, Swampscott, MA) and reconstituted in PBS. A SDS-PAGE analysis of the antigens is shown in supplemental Fig. S1A. Chemicals were from Sigma Aldrich unless noted otherwise.

Immunizations—Rats (5 animals per group) were immunized by Genovac (Genovac GmbH, Freiburg, Germany). Initial immunization with HuD or with DNP-KLH was followed by three boosts and a final bleed. All steps were spaced by 2 weeks. Preimmune sera and final bleed sera were collected for analysis. All immunizations were in Freund’s incomplete adjuvant, to minimize antibodies formed against the bacterial components of the complete adjuvant. Immune sera were tested for affinity with an ELISA assay against recombinant HuD or DNP-BSA. These antigens were coated at 5 μg/ml in PBS on ELISA plates (Mircrolon medium binding, Greiner Bio-One, Wemmel, Belgium), the plates were then blocked in tris buffered saline with 0.05% Tween-20 and 10 mg/ml BSA (blocking buffer). Plates were subsequently incubated with sera and later HRP-conjugated secondary antibodies diluted in blocking buffer. Detection was done with 3,3′,5,5′-tetramethylbenzidine, stopped by adding 1 M HCl, and absorbance was read out at 450 nm. Washes between all steps were under running deionized water.

IgG Purification—IgG was purified from rat preimmune serum using ProteinG affinity columns (Thermo Scientific) according to the manufacturer’s recommendations. For affinity purification of immune sera, purified HuD was immobilized on an AminoLink resin (Thermo Fisher Scientific, Etten-Leur, The Netherlands), as well as BSA. The BSA resin was subsequently incubated with 15 mM dinitrobenzenesulfonic acid in 0.15 M K<sub>2</sub>CO<sub>3</sub> to decorate it with DNP groups (15). Antibodies reactive to the KLH carrier used in the immunizations should not show affinity to the resulting resin. Serum samples were incubated with the affinity resin, and after washing the antibodies were eluted with 1 M propionic acid (pH 2.5). The eluted antibodies were quantified by their A280 (Nanodrop ND-1000, Thermo Scientific, Breda, The Netherlands), and subsequently lyophilized for storage and later analysis. An SDS-PAGE analysis of purified samples is shown in supplemental Fig. S1B.

LC-MS—The immunoglobulin was subjected to in-liquid digestion by trypsin. Samples containing equal amounts of protein were first dissolved in 0.2% Rapigest (Waters Corporation, Milford, MA) in 50 mM NH<sub>4</sub>HCO<sub>3</sub>. The solution was reduced with DTT at 60 °C for 30 min. After cooling, the mixture was alkylated in the dark with iodoacetamide at room temperature for 30 min, and digested overnight with 1:100 (w/w) trypsin mass spectrometry grade (Promega, Madison, WI). Trifluoroacetic acid was added to obtain a final concentration of approx. 0.5% (pH < 2). After 45 min of incubation at 37 °C, the samples were centrifuged at 13,000 rpm for 10 min.

For nano-LC LTQ-Orbitrap mass spectrometry, the sample injection volume was chosen based on the UV absorbance (214 nm) during a test separation on a C18 column. Samples were subsequently injected onto a chromatography system (nanoLC Ultimate 3000; Dionex, Amsterdam, The Netherlands). After preconcentration and washing of the sample on a C18 trap column (1 mm × 300 μm i.d.), peptides were separated on a C18 PepMap column (250 mm × 75 μm internal diameter) (Dionex, Amsterdam, The Netherlands) using a linear 90 min gradient (4–40% acetonitrile/H<sub>2</sub>O 0.1% formic acid) at a flow rate of 300 nL/min. The nanoLC was coupled with a nanospray source of a linear ion trap Orbitrap (LTQ-Orbitrap) mass spectrometer (LTQ Orbitrap XL, Thermo Electron, Bremen, Germany). All samples were measured in a data dependent acquisition mode. Before each run, a blank MS run was performed to monitor system background. The peptide masses are measured in a survey scan with a maximum resolution of 30,000 in the Orbitrap. To obtain a maximum mass accuracy, a prescan is used to keep the ion population in the Orbitrap for each scan approximately the same. During the high-resolution scan in the Orbitrap, the 5 most intense monoisotopic peaks in the spectra were fragmented and measured in the linear ion trap. The fragment ion masses are measured in the linear ion trap to have a maximum sensitivity and the maximum amount of MS/MS data.

Software and Statistics—Label-free quantification of MS data were done with Progenesis LC-MS version 2.5 (Non-linear
Proteomic Analysis of Immunoglobulin Peptides

RESULTS

The immune sera that were obtained showed affinity to the antigens used for the immunizations in an ELISA assay (supplemental Fig. S3). Initial analysis of mass chromatograms obtained from these samples was done with the label-free quantification package Progenesis. LC-MS runs were aligned, and corresponding features were quantified in all samples by means of the area under the peak. The Progenesis dataset consisted of 22,322 features; a Mascot search of the NCBI nr database yielded 2175 amino acid sequences, while an additional Mascot search of a custom database of rat germline immunoglobulin sequences yielded 535 more sequences, and 7066 sequences were found by de novo sequencing with Peaks. In total, we could thus assign a peptide sequence to 9776, or 44%, of the features in our dataset. It should be noted that de novo sequences in particular are prone to contain errors (supplemental Fig. S4). However, alignment of such sequences to the germline database can provide clues about the location of a peptide, even in the presence of several erroneous amino acid assignments.

We first performed an overall analysis to uncover distinguishing properties of the dataset. An unsupervised hierarchical clustering routine in Partek Genomics Suite 6.4 revealed that sera from the immunized animals clustered according to their treatment, except for one DNP sample (Fig. 1A). Thus, the unsupervised clustering suggests that the immunizations have resulted in features that are shared among animals in a treatment group, yet distinguished from the other group. This result was confirmed by a principal com-

Dynamics). Raw abundance data from Progenesis were normalized to the abundance of a set of constant region peptides, which should be equally present in all samples. Each of these peptides was first normalized to the average of all samples for that peptide; the normalization factor was then averaged over the set of constant region peptides, and subsequently used to correct each sample. Cluster analysis was done on these normalized abundances using Partek Genomics Suite (Partek Inc, St. Louis, MO), with Pearsons Dissimilarity as a distance parameter and Ward’s method. Principal component analysis was performed using the covariance dispersion matrix and normalized Eigenvector scaling, also with Partek Genomics Suite. Statistical tests were done using R version 2.10.1, or using Microsoft Excel 2007 where possible. Non-parametric tests were used, as the type of analysis done by Progenesis yields a large number of ties in the Mann-Whitney test. However, one group contains many zero values, which disturbs the otherwise log-normal distribution of the values (supplemental Fig. S2).

First, a Mann-Whitney test was performed on abundances to find masses that significantly differed between the two treatment groups with \( p < 0.01 \). We only included features which emerged after immunization by also requiring \( p < 0.01 \) in a Mann-Whitney-test between treated and preimmune sera, as well as at least a 10-fold increase in average abundance. When one group contains many zero values, \( p \) values increase due to the large number of ties in the Mann-Whitney test. However, we feel that features that are initially absent, but emerge after treatment, are of particular interest. We therefore added features that were present in at least 3 out of 5 animals immunized with one antigen, but absent in the other group, absent in at least 7 of the 9 preimmune sera, and that had at least a 10-fold increase in average abundance in treated animals compared with the preimmune sera. 26% of the peptides that were found fell into this category.

Database searches were done with Mascot 2.2.06 (Matrix Science, London, UK) against the NCBI nr database of July 17 2008 and against a custom rat Ig germline database, with 10 ppm peptide tolerance, 0.5 Da MS/MS tolerance and as large number of ties in the Mann-Whitney test. However, we feel that features that are initially absent, but emerge after treatment, are of particular interest. We therefore added features that were present in at least 3 out of 5 animals immunized with one antigen, but absent in the other group, absent in at least 7 of the 9 preimmune sera, and that had at least a 10-fold increase in average abundance in treated animals compared with the preimmune sera. 26% of the peptides that were found fell into this category.

Although equal amounts of protein were used for the samples based on their \( A_{280} \), we first adjusted the raw abundance data to compensate for spray or ionization deviations and other artifacts that might have occurred during digestion and analysis. This compensation was based on peptides that gave a strong signal and were identified by Mascot as peptides derived from constant domains of either the heavy chain or the kappa or lambda light chains of IgG (supplemental Table S1). Correction factors ranged from about 0.5 to 2 \( \times \), with an average of 1.0 \( \times \).

The immune sera that were obtained showed affinity to the antigens used for the immunizations in an ELISA assay (supplemental Fig. S3). Initial analysis of mass chromatograms obtained from these samples was done with the label-free quantification package Progenesis. LC-MS runs were aligned, and corresponding features were quantified in all samples by means of the area under the peak. The Progenesis dataset consisted of 22,322 features; a Mascot search of the NCBI nr database yielded 2175 amino acid sequences, while an additional Mascot search of a custom database of rat germline immunoglobulin sequences yielded 535 more sequences, and 7066 sequences were found by de novo sequencing with Peaks. In total, we could thus assign a peptide sequence to 9776, or 44%, of the features in our dataset. It should be noted that de novo sequences in particular are prone to contain errors (supplemental Fig. S4). However, alignment of such sequences to the germline database can provide clues about the location of a peptide, even in the presence of several erroneous amino acid assignments.

We first performed an overall analysis to uncover distinguishing properties of the dataset. An unsupervised hierarchical clustering routine in Partek Genomics Suite 6.4 revealed that sera from the immunized animals clustered according to their treatment, except for one DNP sample (Fig. 1A). Thus, the unsupervised clustering suggests that the immunizations have resulted in features that are shared among animals in a treatment group, yet distinguished from the other group. This result was confirmed by a principal com-
ponent analysis. Again, the treatment groups had distinct clustering patterns in this PCA plot (Fig. 1B). We similarly compared samples that all underwent Protein G immunoglobulin purification, but no affinity purification. This facilitates the most correct comparison of treated sera with the preimmune sera in the cluster analysis. The treated sera could be distinguished from each other like before, but also from the preimmune sera (supplemental Fig. S5, A and B).

The abundance data were subsequently subjected to statistical analysis as detailed under “Experimental Procedures.” Briefly, we selected features that differed between groups with \( p < 0.01 \) in a non-parametric test, as well as features that were completely absent in for one antigen, but present in several animals immunized with the other antigen. With those two selection criteria, we found 684 features qualifying as a hit, 186 specific for DNP-treated animals and 498 for the HuD-treated animals. Some examples of peptides that were found are shown in Table 1.

When analyzing large datasets, it is always important to account for the possibility of false discovery. We assessed this by repeatedly scrambling the group assignment of the samples in our dataset (permutation approach). After subsequent analysis of the scrambled data in the same way as the original data, we found on average 15 significant features for DNP and 13 features for HuD. These 28 hits represent the false discovery rate of our approach, which is significantly less than the true number of hits (\( p < 10^{-7} \)) and amounts up to less than 10% of the hits found from the unscrambled dataset. Thus, we expect that 90% of our hits represent true differences between the immunized animals and the control sera. Some features have a higher abundance in preimmune sera than in treated sera: this is not surprising as immune sera were affinity enriched and many sequences may have been removed compared with preimmune sera. Nevertheless, we find that the majority of features were more abundant in treated sera than in the preimmune sera (not shown), and that treated sera can be distinguished from preimmune sera (supplemental Fig. S5, A and B). This suggests that many new antibody sequences were formed after immunization.

For all peptides we determined the most likely origin on the immunoglobulin molecule. All peptides were aligned to a database of rat germline sequences with the Blast application to elucidate their location. It was found that peptides derived from the variable regions were the most numerous in our dataset (Fig. 2), while peptides from the constant domains were responsible for the signals with the highest intensities, on average 2.5-fold more than variable region peptides. Within the variable regions were the most numerous in our dataset (Fig. 2), while peptides from the constant domains were responsible for the signals with the highest intensities, on average 2.5-fold more than variable region peptides. Within the variable
features differentially expressed for the DNP treatment (p 0.09). We looked for a skewed distribution in the germline components analysis. This finding alone shows that the formation of antigen-specific immunoglobulins is not a fully stochastic process, but rather results in reoccurring sequence patterns that are found in most members of the treatment groups. Our goal in this work was to show under optimal conditions whether such characteristic peptides emerge after immunization with a particular antigen. The subjects were homogeneous (but not inbred) set of animals, immunizations were boosted repeatedly to optimize serum levels, a small hapten (DNP) was chosen as one antigen to limit the number of epitopes, while the other antigen (HuD) was more complex but clinically relevant. Also, the antibodies were affinity purified against the antigen used in the immunizations. While a real set of patient samples will almost certainly be more challenging than the current dataset, we felt the current work was important to establish a proof of principle for our approach, before committing to a larger study involving patient material.

It was interesting that we found almost 3 times as many specific peptides for HuD immunized animals than for the DNP treated group. We can only speculate on the cause of this, but it is conceivable that a small single epitope like dinitrophenol permits a smaller variety of antibody sequences to bind than the larger multi-epitope HuD antigen. For the latter, at least two major and several minor epitopes have been described as antigenic regions in patient sera (14, 18).

The specific peptides that emerged from the dataset largely originated from the V-region of the immunoglobulin molecule. Some peptides were assigned to the constant region by our homology analysis, but these peptide sequences resulted from de novo analysis, and the alignment result may well be due to mistakes in the amino acid sequence; constant regions are represented in the database, and peptides from the constant region should have emerged in the Mascot search instead. In the analysis of peptide origin, we also included peptides with a relatively poor alignment score. This was appropriate, as it is known that the majority of peptides are of immunoglobulin origin, and we seek the most probable alignment within an immunoglobulin database. We verified the consequence of our quality threshold by comparing the results to those obtained with a more strin-
Proteomic Analysis of Immunoglobulin Peptides

gent bitscore threshold of 20. Under these criteria, fewer peptides could be assigned to e.g. CDR1, 2 or 3, but the ratio between the different assignments remained similar. Thus, while a permissive bitscore threshold might allow some misalignments, we can present the most probable assessment of more peptides without introducing a systematic error. The analysis of peptide origins also indicated that a repertoire bias occurs after immunization with dinitrophenol. This finding is corroborated by the increased abundance of IGHV1 peptides in DNP-immunized rats, and was also supported by a similar repertoire bias that was described in the mouse (5). Because of the repertoire bias, IGHV1 peptides have some marker value toward the DNP group. However, as these peptides are still expressed in all subjects, their selectivity is poor compared with some of the other peptides we identified, and the practical use of the repertoire bias for the discovery of marker proteins is thus limited.

If the similarities between immunoglobulin peptides within a treatment group stem from a convergent process during maturation, one would expect peptides mutated from the germline to be preferentially present in a particular group. An exact analysis of such statistics is complicated by the fact that the number of mismatches we find is a combination of true mutations and of sequencing mistakes in the peptides, especially in those derived from the more error-prone de novo analysis. In addition, peptides which contain many true mutations are increasingly difficult to align to the germline segment they derived from, and have a higher chance of aligning to an unrelated segment instead. Thus, the analysis of mutation rates must be approached with caution. However, if one assumes a maximum number of unique antibodies in a particular rat to be at most $10^7$, 6 random mutations from germline already generates a potential sequence diversity which exceeds that number of antibodies in a rat (19). Peptides that accumulated so many mutations could therefore be expected to be unique to an individual. Yet, we find peptides like ACKLGSVNVYQQHWR, which are different from germline by 6 amino acids, and are present at high levels in 3 out of 5 DNP-treated animals and none of the other samples (Table 1). In total, we found 93 such peptides that were present exclusively in one treatment group in at least 3 rats. In these peptides, those with an available (de novo) sequence had an average mismatch of 9 amino acids with the best aligning germline sequence. This supports our conclusion that immunoglobulin sequence diversity in a real immune response must be very restricted compared with the potential diversity resulting from random mutations alone.

We investigated whether any sequence motifs or preferred mutations emerged from the peptides that were found to be differentially expressed. Through multiple alignment, a phylogenetic tree was constructed of all differential peptides to reveal similarities in their sequences. Within the diverse population, we found some clusters of antigen-specific peptides with sequence similarity. However, their similarity derived from shared germline sequences, rather than from a recurring mutation from this germline. Thus, based on the available set of sequences we cannot conclude at this time that a particular pattern of mutations occurs in all animals immunized with a particular antigen.

Whereas many peptides were present in multiple samples, the dataset also contains peptides that were found uniquely in only a single animal. We found 172 such peptides, 164 of which were found in one of the immune sera, and the average number of mismatches in this group of peptides was also 9. Although many unique peptides may remain under the detection threshold with our current method, it is compelling that less than 1% of all peptides found in the dataset were unique to a single rat, and the remainder thus was a shared feature among two or more animals.

The part of the immunoglobulin molecule with the most variation is the CDR3 of the heavy chain, which is shaped by a combination of three gene segments, junctional diversity, and mutations. It can therefore be assumed that this CDR3 is particularly important for the affinity of the immunoglobulin molecule, and thus should also contribute to the immunoglobulin peptides that are differentially present in the immunization groups. In our current dataset, we find fewer peptides from CDR3 than from the other CDRs. A possible explanation lies in the fact that we cannot assign all CDR3 peptides as such, because their randomized sequence is too far removed from their germline origin for proper alignment. However, we must also consider the possibility that, for marker purposes, the best immunoglobulin peptides are not necessarily the most unique and random ones. It is conceivable that the most distinctive peptides are the ones that accumulated sufficient mutations to gain antigen specificity, but not so many mutations that convergence becomes less probable. In this case, there would be an optimal number of mutations from the germline sequence to generate a good marker peptide, rather than the maximum number of mutations that yields the best markers. Because of the uncertainty in our analysis of mutations rates, we cannot reliably distinguish between these two hypotheses based on the current data. This would require a large effort based on deep nucleotide sequencing in B-cells, which would provide sequences with a much improved confidence from the samples. Such a study could be coupled to proteomics data to obtain a quantification of immunoglobulin expression. This is beyond the scope of the current work, which aims to show that immunoglobulin peptides are viable for biomarker applications, rather than explore the full sequence diversity of the immunome.

Immunoglobulins may bind to both conformational and linear epitopes. Similarly, both the sequence and three-dimensional structure of the CDRs may be important in determining the antigen specificity and different sequences may result in a similar conformation. Regardless, only a limited set of primary immunoglobulin structures will result in an appropriate binding surface for the target antigen, and we should be able to detect the peptides derived from these primary sequences using our approach.

Our data show many masses that are differentially present in response to treatment, and we can show their origin in the immunoglobulin V-region in many cases. However, for application as a diagnostic tool it is of importance to add further characterization of the masses that are detected. First, the amino acid sequence of all peptides ought to be known for use as a biomarker set. Second, it is preferable to
locate the origin of the peptide in an immunoglobulin molecule, also for mutated peptides, which are difficult to align. It may be possible to improve sequence coverage, quality and annotation with technical innovations. Alternative proteases could supplement trypsin by providing overlapping peptide fragments for de novo sequencing, or fragments of longer length that include more conserved amino acids, and thus could aid the alignment to the germline. Other options for improvement are the use of novel ultra high pressure chromatography techniques that improve resolution and signal intensity, and therefore give better opportunities to determine the sequence of peptides, the partial digestion of immunoglobulins to enrich V-regions and deplete constant regions in the sample, and the use of alternative fragmentation methods such as higher energy collision induced dissociation (HCD) and electron-transfer dissociation (ETD), which facilitate de novo sequencing with fragmentation spectra of superior mass resolution. Finally, it is clearly important to validate marker peptides in larger sample sets originating from different sites. Thus, several aspects may still benefit from optimization of the current work. Nevertheless, we believe that it is convincingly shown that, after an immunization, many shared features arise in the immunoglobulin sequences of subjects receiving this treatment, and thus, that immunoglobulin sequences hold great promise as markers for disease that is associated with an antibody response.

REFERENCES

1. Xu, J. L., and Davis, M. M. (2000) Immunity 13, 37–45
2. Janeway, C. (2005) Immunobiology: The Immune System in Health and Disease, 6th Ed., Garland Science, New York
3. Saada, R., Weinberger, M., Shahaf, G., and Mehr, R. (2007) Immunol. Cell Biol. 85, 323–332
4. Andersen, P. S., Haahr-Hansen, M., Coljee, V. W., Hinnerfeldt, F. R., Varming, K., Bregenholt, S., and Haurum, J. S. (2007) Mol. Immunol. 44, 412–422
5. Loh, D. Y., Bothwell, A. L., White-Scharf, M. E., Imanishi-Kari, T., and Baltimore, D. (1983) Cell 33, 85–93
6. Baranzini, S. E., Jeong, M. C., Butunoi, C., Murray, R. S., Bernard, C. C., and Oksenberg, J. R. (1999) J. Immunol. 163, 5133–5144
7. Sehgal, D., Schiaffella, E., Anderson, A. O., and Mage, R. G. (1998) J. Immunol. 161, 5347–5356
8. Poulsen, T. R., Meijer, P. J., Jensen, A., Nielsen, L. S., and Andersen, P. S. (2007) J. Immunol. 179, 3841–3850
9. Weinstein, J. A., Jiang, N., White, R. A., 3rd, Fisher, D. S., and Quake, S. R. (2009) Science 324, 807–810
10. Sahin, U., Türeci, O., and Pfleiderer, M. (1997) Curr. Opin Immunol. 9, 709–716
11. Kim, R., Emi, M., and Tanabe, K. (2007) Immunology 121, 1–14
12. Leidinger, P., Keller, A., Ludwig, N., Rheinheimer, S., Hamacher, J., Huser, H., Stahle, I., Lenzh, H. P., and Meese, E. (2008) Int. J. Cancer 123, 1631–1636
13. Darnell, R. B., and Posner, J. B. (2003) N. Engl. J. Med. 349, 1543–1554
14. Manley, G. T., Snit, P. S., Dalmou, J., and Posner, J. B. (1995) Ann. Neurol. 38, 102–110
15. Hay, F. C., and Westwood, O. M. (2008) Practical Immunology pp. 115–162, Blackwell Publishing Company, Oxford, UK
16. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BMC Bioinformatics 10, 421
17. Lebranc, M. P., Giudicelli, V., Gistougou, C., Jabado-Michaloud, J., Folch, G., Bellahcine, F., Wu, Y., Genmot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lebranc, G., and Duroux, P. (2009) Nucleic Acids Res. 37, D1006–D1012
18. Sodeyama, N., Ishida, K., Jacek, K. A., Zhang, L., Azuma, A., Yamada, M., Mizusawa, H., and Wada, Y. (1999) J. Neurol. Neurosurg. Psych. 66, 97–99
19. Radbruch, A., Muehlenhaus, G., Luger, E. O., Inamine, A., Smith, K. G., Dörner, T., and Hiepe, F. (2006) Nat. Rev. Immunol. 6, 741–750