AN EXPLORATORY ANALYSIS OF RADIOIMMUNOASSAYS DATA: ESTABLISHING SUBTYPES OF THE HEPATITIS-B VIRUS

By

M. Anthony Wong, Jack R. Wands, Robert Marciniak, Russell D. Brown and James Shorey

WP 1468 December 1983
AN EXPLORATORY ANALYSIS OF RADIOIMMUNOASSAYS DATA:
ESTABLISHING SUBTYPES OF THE HEPATITIS-B VIRUS

By

M. Anthony Wong, Jack R. Wands, Robert Marciniak,
Russell D. Brown and James Shorey

WP 1468                          December 1983
An Exploratory Analysis of Radioimmunoassays Data: Establishing Subtypes of the Hepatitis-B Virus

by

M. Anthony Wong
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139 U.S.A.

Jack R. Wands, Robert Marciniak, Russell D. Brown
Gastrointestinal Unit, Massachusetts General Hospital
and the Department of Medicine, Harvard Medical School
Boston, MA 02114 U.S.A.

and

James Shorey
Department of Medicine
University of Texas Southwestern Medical School
Dallas, TX 75325 U.S.A.
M. Anthony Wong is Associate Professor, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. Jack R. Wands is Associate Professor and R. D. Brown and R. A. Marciniak are research associates at the Gastrointestinal Unit, Massachusetts General Hospital and the Department of Medicine, Harvard Medical School, Boston, MA 02114. James Shorey is Professor, Department of Medicine, University of Texas Southwestern Medical School and the V. A. Medical Center, Dallas, TX 75325. This work was supported by Grants AA-02666 and CA-35711 from the National Institutes of Health. JRW is the recipient of a Research Career Development Award AA-00048 from the National Institutes of Health.
ABSTRACT

A novel approach has been developed for the analysis of hepatitis B virus (HBV) antigenic structure which creates numerical "signatures" of HBV viral strains. This technique employs high affinity IgM and IgG monoclonal antibodies (anti-HBs) directed toward distinct and separate determinants on hepatitis B surface antigen (HBsAg). Such antibodies have been used to develop sensitive and specific radioimmunoassays for measurement of HBsAg-associated determinants in serum. This approach also involves the use of a statistical technique for aligning the unknown concentrations of HBsAg present in different individuals with the same viral strain. In this paper, we will describe the iterative procedure developed for analyzing these radioimmunoassays data and we will also present the findings of our exploratory analysis. In summary, this approach suggests that HBV is far more heterogeneous than previously recognized by polyvalent anti-HBs antibodies, and the present technique is shown to be useful in epidemiologic studies of HBV transmission. This approach also has broader significance for the study of subtle or major antigenic changes among other viral agents including polio, influenza and herpes since it is not necessary to know the concentration of virus or viral protein in complex protein mixtures.

KEY WORDS: Monoclonal antibodies; radioimmunoassays data; least-squares; alignment procedure; exploratory analysis.
1. INTRODUCTION

The development of monoclonal antibodies by Kohler and Milstein has led to an explosion in biological research. (See, for example, Chisari 1984). In this study, we will focus on the use of monoclonal antibodies to characterize the antigenic structure of the hepatitis B virus (HBV) at the molecular level, and demonstrate with the aid of statistical analysis, that HBV is far more heterogeneous than previously recognized.

We (Wands, Ben-Porath and Wong, 1984) have previously described the production and characterization of monoclonal antibodies to HBV (anti-HBs) and hepatitis B surface antigen (HBsAg). By taking advantage of the unique properties of these antibodies, we were able to construct highly sensitive and specific radioimmunoassays (RIAs) for the measurement of HBsAg-associated epitopes in serum; these RIAs and the statistical data they generate will be described in Section 2. Since each such antibody binds to a distinct and separate antigenic determinant on HBsAg, the antigenic structure of a HBV viral strain is characterized by its binding activities to a panel of monoclonal antibodies over a range of viral concentrations. A model for the functional relationship between binding activity and viral concentration is discussed in Section 3, and a statistical procedure is developed for estimating this functional relationship. This development is necessary since the concentration of virus or viral protein in serum samples cannot be determined because they are complex protein mixtures.

In Section 4, an iterative least-squares procedure for aligning the concentration of samples from the same HBV subtype is described. In addition, we will report on our exploratory analysis of the functional form of the relationship between binding activity and viral concentration.
The biological significance of the present approach to characterizing the antigenic structure of HBV subtypes will be discussed in Section 5. We will also indicate its usefulness in epidemiologic studies of HBV transmission and its value in studying the antigenic composition of other vital agents including polio, influenza, and herpes.
2. THE RADIOIMMUNOASSAYS

The immunization protocols, cell fusion technique, growth and cloning of hybridomas producing anti-HBs have been reported in Wands and Zurawski (1981). The monoclonal anti-HBs antibodies have been characterized with respect to specificity for determinants on HBsAg, antibody class and subclass and affinity for HBsAg-associated determinants. (See, for example, Wands et al., 1981.) Monoclonal antibodies designated 5C3(Ig G2a), 5C11, 2C6, 1C7 and 4E8(Ig G1) and 5D3, 1F8 and 2F11(IgM) were selected for use in this study from a library of monoclonal anti-HBs antibodies because of their following properties: 1) several bind to all known subtypes of HBsAg; 2) some demonstrate qualitative binding differences to major HBsAg subtypes; 3) they recognize distinct and separate determinants on HBsAg; and 4) they possess very high affinity constants (4.8 x 10^9 to 4.0 x 10^11 liters/mole per molecule) for HBsAg-associated epitopes. (For details, see Wands and Zurawski, 1981).

We employed 8 monoclonal "simultaneous sandwich" radioimmunoassays (RIAs) for analysis of HBsAg-associated binding activities in various HBsAg positive serum samples. In brief, polystyrene beads were coated with a monoclonal IgM anti-HB designated 5D3 and the other 8 antibodies, including 5D3, were radiolabeled to a specific activity of 10-12 μCi/μg protein with the Hunter-Bolton reagent (Bolton and Hunter, 1973). For each antibody, the 5D3 coated beads were incubated with serial 10-fold dilutions (10^x; x = 0, 1, ..., 5) of samples (100μl) of a given serum and 100 μl of radiolabeled probe, consisting of approximately 150,000 cpm of the
monoclonal anti-HB, for 4 hours at 45\(^\circ\) C. The beads were washed extensively with distilled water and the radioactivity bound to each bead was measured with a Packard gamma well counter.

The RIAs data generated from a serum sample was, therefore, an 6 x 8 matrix whose (ij)th element was the natural logarithm of the signal-to-noise ratio (\(\log \text{S/N}\), defined as the median cpm bound in the experimental samples of the serum/median cpm bound in samples of a negative control serum), obtained when the \(10^i\)th dilution of the serum and the jth antibody were used. In the present study, we selected 64 individuals from the United States whose serum are representative of known HBsAg subtypes. These specimens have been classified as ayw\(_2\) (subtype 2), ayw\(_3\) (subtype 3), adw\(_4\) (subtype 6) and adw\(_2\) (subtype 7) HBsAg subtypes by RIAs using conventional polyvalent anti-HBs antibodies. (See Shorey, Brown and Wands, 1981.)
3. THE STATISTICAL PROBLEM

Since each monoclonal antibody binds to a distinct and separate antigenic determinant on HBsAg, the antigenic structure of a HBV subtype is characterized by its binding activities to a panel of monoclonal antibodies over a range of viral concentrations. The statistical problem involved in establishing the antigenic structure of a HBV subtype is therefore one of estimating the multiple response functions $f^A (A = 1, 2, \ldots, 8)$ between $\log_e S/N$ and $\log_{10}$ viral concentration ($x$) for the 8 selected antibodies using serum samples from that subtype. It is hypothesized that each $f^A(x)$ is a smooth growth curve which increases monotonely from zero to a unique local maximum at a high value of $x$ and then decreases with further increase in $x$ because of binding interferences (see Figure 1); it should be pointed out, however, that most serum samples do not have very high viral concentrations.

Let $y^A_{i,d}$ be the $\log_e S/N$ value obtained for antibody $A$ ($A = 1, 2, \ldots, 8$) at $\log_{10}$ dilution factor $d$ ($d = 0, 1, \ldots, 5$) of the $i$th serum sample ($i = 1, 2, \ldots, n$) from a HBV subtype. Then the statistical model for the RIAs data can be expressed as follows:

$$y^A_{i,d} = f^A (x_i - d) + e^A_i, \quad d = 0, 1, \ldots, 5; \quad A = 1, 2, \ldots, 8; \quad i = 1, 2, \ldots, n [1]$$

where $x_i$ is the unknown $\log_{10}$ concentration of the $i$th serum sample at $d = 0$, the $e^A_i$'s are the error terms with $E[e^A_i] = 0$ and standard derivation $[e^A_i] = \sigma^A$, and the $f^A$'s are of the same functional form. The $x_i$'s cannot be determined because serum samples are complex protein mixtures. Since the $x_i$'s are unknown, we cannot use standard curve-fitting methods...
to estimate the $f^A(x)$'s directly from the data; this problem is illustrated in Figure 2a where the $y_{i,d}^A$'s obtained for the 14 subtype 6 samples are plotted against d, ignoring the variations of the $x_i$'s. We have therefore developed an iterative least-squares procedure for aligning the concentrations of serum samples from the same subtype. The alignment is effected by first setting $x_1 = 0$, and then estimating the other $x_i$'s ($i = 2, 3, ..., n$) using the statistical model given in Equation [1]. The aligned data are subsequently used to estimate the $f^A(x - x_1)$'s which collectively characterizes the antigenic structure of a HBV subtype. This iterative procedure will be described in the next section.
4. AN ITERATIVE ALIGNMENT PROCEDURE

The proposed procedure for aligning the unknown $x_i$'s ($i = 1, 2, \ldots, n$) of serum samples from a HBV subtype consists of the following steps:

**STEP 1 (Initial Alignment):** Select a "leader" whose undiluted $\log_{10}$ viral concentration $x_1$ is set to be zero. Using the 8 piecewise linear functions defined by the $\log_e S/N$ values obtained for this leader at $x = 0, 1, \ldots, 5$, the estimates of the other $x_i$'s ($i = 2, 3, \ldots, n$) on this selected scale are then obtained using a least squares method.

**STEP 2 (Curve Fitting):** Estimate the response functions $f^A (A = 1, 2, \ldots, 8)$ from the aligned data using exploratory regression techniques. The overall sum of squared residuals is also computed.

**STEP 3 (Updating Alignment):** Using the $f^A$'s obtained in STEP 2, the estimates of $x_i$'s ($i = 1, 2, \ldots, n$) are updated. If the resulting decrease in the overall sum of squared residuals is less than a specified threshold, stop; otherwise, go to STEP 2.

A detailed description of the proposed alignment procedure is given in the following, but it should be noted that this procedure only provides estimates of the $(x_i - x_1)$'s, and not the $x_i$'s themselves.

4.1 Initial Alignment Step

The aim of the procedure is to align the unknown viral concentrations of the serum samples from a HBV subtype. It is equivalent to estimating the $(x_i - x_1)$'s ($i = 1, 2, \ldots, n$), and if we set $x_1$ to be zero, then we are only estimating the $x_i$'s ($i = 2, \ldots, n$). In this step, the objective is to find initial estimates of the $x_i$'s.
Since from [1], \( y_{i,d}^A = f^A(x_i - d) + e_i^A \) and each \( f^A \) is a smooth growth curve increasing monotonely from zero to a unique local maximum, the following 8 piecewise linear functions, \( f^A_p \) \((A = 1, 2, \ldots, 8)\), defined by \( y_{i,d}^A \) \((d = 0, 1, \ldots, 5; A = 1, 2, \ldots, 8)\) are approximations of the \( f^A \)'s:

\[
f^A_p(x) = a_d^A + b_d^A x, \quad d < x < d + 1; \quad d = 0, 1, 2, \ldots
\]

where \( a_d^A \) and \( b_d^A \) is respectively the intercept and slope of the line between \( (d, y_{i,d}^A) \) and \( (d+1, y_{i,d+1}^A) \), and \( y_{i,d}^A = 0 \) for \( d > 6 \) because \( 10^6 \) or higher dilutions of practically all serum samples would contain effectively no viral protein. The \( x_i \)'s are then estimated by using a least squares method. Specifically, for each \( i \), we find the value of \( x_i \) to minimize the sum of squared residuals

\[
SSR_i = \sum_{A=1}^{8} \sum_{d=0}^{5} [y_{i,d}^A - f^A_p(x_i - d)]^2
\]

Since interpolation is preferred to extrapolation, the "leader" sample is the one with the most \( \max_i y_{i,0}^A \) values among the 8 possible maxima. Using the \( f^A \)'s given in [2], it can be shown that if \( j < x < j + 1 \) \((j = 0, 1, 2, \ldots, 5)\), then the least squares estimate of \( x_i \) is given by

\[
x_{i,j} = j + \sum_{A=1}^{8} \sum_{d=j}^{j+5} b_d^A y_{i,d-j}^A - \sum_{A=1}^{8} \sum_{d=j}^{j+5} a_d^A b_d^A - \sum_{A=1}^{8} \sum_{d=j}^{j+5} (b_d^A)^2 (d-1)
\]

By computing \( x_{i,j}, j = 0, 1, 2, \ldots, 5 \), and their associated \( SSR_i \) values, the least squares estimate of \( x_i \) can be found by identifying the \( x_{i,j} \) with the smallest \( SSR_i \) value.
It should be noted that, instead of the $f_{p}^{A}$'s, we could have used a smooth spline function (see, for example, Winsberg and Ramsay, 1982); but the least squares estimate of the $x_{i}'s$ would then have to be obtained by a numerical optimization algorithm. We decide to use the picewise linear functions because our objective here is to find a set of reasonably good initial estimates of the $x_{i}'s$. Moreover, an estimation method more resistant than the least squares procedure could have been used; instead, $y_{1,d}^{A}$ values associated with large least-squares residuals are identified and removed from all subsequent analysis. The aligned data for the 14 subtype 6 serum samples are shown in Figure 2b, where the $\log_{e} S/N$ values are plotted against the negative aligned $\log_{10}$ viral concentration ($-x$).

4.2 Curve-Fitting Step

With the aligned data, a better approximation of the $f_{p}^{A}$'s than the $f_{p}^{A}$'s can be obtained by pooling the sample data from the same HBV subtype. Our objective is therefore to fit a function $f_{n}^{A}$ to the aligned data for monoclonal antibody A, and the $f_{n}^{A}$'s should be of the same functional form. By examining Figure 2b, it can be observed that the variability of $\log_{e} S/N$ is not constant across the range of viral concentrations; as expected from the experimental procedure, the variability is high when the concentration is low or when $\log_{e} S/N$ is close to zero. Hence, we use a weighted least squares procedure, in which $\log_{e} S/N$ values close to zero carry very little weight, to find a function that would fit the aligned data well.

Several functional forms including the quadratic, cubic, Gamma and Gaussian curves have been fitted to the aligned data for all known HBV subtypes, but plots of the residuals obtained for these curves indicate that they do not
fit the data well. As suggested by residuals analysis, the best-fitting function is the logistic curve of the form:

\[
\hat{f}_n^A(x) = \exp\left[3 \sum_{k=0}^{\frac{3}{8}} \beta_k x^k\right]/(1 + \exp\left[3 \sum_{k=0}^{\frac{3}{8}} \beta_k x^k\right])
\]  

where \(\hat{f}_n^A(x) = (\hat{f}_n^A(x) + 1)/9\). This function fits the monotonely increasing part of the aligned data particularly well, but it cannot be used for extrapolation purposes. Moreover, it should be noted that the cubic term is not necessary to fit the aligned data for most HBV subtypes; but for those whose serum samples have high viral concentrations, this extra term is needed to capture the transition at the peak of the growth curves. The final fitted curves and the aligned data for subtype 2, 3, 6 and 7 are shown in Figure 3-6 respectively, and they are not very different from the plots obtained after this intermediate step. The overall sum of squared residuals

\[
\sum_{i=1}^{n} \sum_{A=1}^{8} \sum_{d=0}^{5} [y_{i,d} - \hat{f}_n^A(x_{\bar{i}}d)]^2
\]

is also computed for the samples from each subtype.

4.3 Updating Step

With the \(\hat{f}_n^A\)'s, the \(x_i\)'s, \(i=1, 2, \ldots, n\) are re-estimated in turn by finding the value of \(x\) that would minimize

\[
\text{SSR}_i = \sum_{A=1}^{8} \sum_{d=0}^{5} [y_{i,d} - \hat{f}_n^A(x_{\bar{i}}d)]^2
\]

Since the updated estimates will be in the neighborhood of the previous estimates and \(\text{SSR}_i\) is a continuous function of \(x_i\), the new estimates can be obtained numerically. It has been found that the solutions after the first few iterations are practically the same and hence in order to avoid excessive computing, a threshold value \(T\) is specified so that when the difference in
\[ \sum_{i=1}^{n} \text{SSR}_i \text{ is not greater than } T, \text{ the iterative procedure stops; otherwise,} \]

the iterations will continue by going back to STEP 2.

Although the developed procedure is unduly dependent on the selected leader, our immunoassays data are so well behaved that this dependency does not create a problem in the present application. We have also examined the variability of the estimated relative viral concentrations of samples from a HBV subtype by using a jack-knife procedure. By leaving out one serum sample at a time, \( n \) sets of estimates of relative concentrations are obtained by the iterative procedure. Again, since our data align very well, these solution sets are found to be very similar. Finally, we would like to point out that this procedure can be generalized to the case where the serum samples have very high viral concentrations and can also be expanded to include other monoclonal anti-HBs antibodies.
5. DISCUSSION

We have described a novel approach, "signature" analysis, for the study of hepatitis B viral antigenic structure. The cardinal feature of this approach included: 1) the use of high affinity anti-HBs monoclonal antibodies, 2) the construction of RIAs specific for different epitopes on HBsAg, and 3) a statistical technique for aligning the unknown concentrations of HBsAg present in different individuals of the same subtype. The estimated binding curves which describe the functional relationships between $\log_e S/N$ and viral concentration for a panel of selected monoclonal antibodies are characteristic for any HBV viral strain and the composite profile defines the antigenic composition of HBV.

"Signature" analysis suggests that HBV shows far more antigenic heterogeneity than previously recognized by polyvalent anti-HBs antibodies. First with respect to classic HBsAg subtypes, we find the greatest antigenic diversity when comparing the ad (subtype 6 and 7) to ay (subtype 2 and 3) classical subtype group. (See, for example, Figure 7a; note that this and all subsequent figures are obtained by aligning the subtypes which are being compared.) This diversity is in accord with recent sequence data (Fujiyama et al., 1983) showing substantial changes in amino acid composition between HBsAg subtypes adw and ayw. In constrast, when the ad and ay domains are the same, and the difference between the w specificities are examined, signature analysis demonstrates remarkable antigenic homogeneity among the w domains (see, for example, Figure 7b).

Further evidence of the antigenic diversity of HBV is supported by the demonstration of subgroups within the "classic" HBsAg subtypes. For example, ayw may now be divided into at least two separate subgroups. In
the United States, an additional subgroup (see Figure 8) has been found to be distinct from the larger ayw\textsubscript{3} (subtype 3) group. Moreover, we have found that Australian aborigines are infected with a ayw\textsubscript{3} strain which differs substantially from ayw\textsubscript{3} groups or subgroups in the United States with respect to epitope composition and density. Therefore, "signature" analysis may permit studies of the genetic evolution of HBS in various parts of the world. Indeed, preliminary studies of the ayw\textsubscript{3} subtype from Africa and the Far East suggest additional viral strains distinct from those found in the United States and Australia. Furthermore, recent investigations suggest that vertical transmission of HBV from mother to child or horizontal spread within families breeds true. That is, the RIAs data obtained for serum samples from all family members infected with HBV including mother and child align very well. Therefore, "signature" analysis will be useful in epidemiologic studies of HBV transmission.

Finally we believe that "signature" analysis may also be of value in studying the antigenic composition of other viral agents. For example, such an approach may be useful in the characterization of antigenic structure of polio, influenza, herpes, and arenavirus groups and other viral agents propagated in tissue culture. Since "signature" analysis does not require that the concentration of viral protein in complex protein mixtures be known, study of antigenic similarity or differences among other viruses would seem possible.
REFERENCES


Fig. 1. A sketch of the expected relationship between \(\log_e S/N\) and \(\log_{10}\) viral concentration in serum.
UNALIGNED SUBTYPE 6 (adw₄)

Fig. 2a. Plotting log S/N against log₁₀ dilution factor for the 14 unaligned subtype 6 samples.
SUBTYPE 6 \((\text{adw}_4)\)

Fig. 2b. Plotting \(\log \frac{S}{N}\) against negative aligned \(\log_{10}\) viral concentration (-x) for the 14 aligned subtype 6 samples.
Fig. 3. Plotting log S/N against negative aligned log$_{10}$ viral concentration (-x) for the 5 aligned subtype 2 samples. The fitted functions are also shown.
Fig. 4. Plotting log(e) against negative aligned log(viral concentration) for the 8 subtype 3 samples. The fitted functions are also shown.
SUBTYPE 6 (adw₄)

Fig. 5. Plotting log S/N against negative aligned log₁₀ viral concentration (-x) for the 14 subtype 6 samples. The fitted functions are also shown.
Fig. 6. Plotting log S/N against negative aligned log_{10} viral concentration (-x) for the 7 subtype 7 samples. The fitted functions are also shown.
Fig. 7a. Comparing the ad (subtype 6: adw₄) to the ay (subtype 3: ayw₃) classical subtype group.

Fig. 7b. Demonstrating antigenic homogeneity among the w domain (subtype 6: adw₄ and subtype 7: adw₂)
Fig. 8. Demonstrating the existence of subgroups in classical subtype 3 (ayw₃).