A procedure for the immunoanalysis of samples containing one or more members of a group of cross-reacting analytes

Geoffrey Jones\textsuperscript{a}, Monika Wortberg\textsuperscript{b}, Bruce D. Hammock\textsuperscript{c}, David M. Rocke\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} Graduate School of Management, University of California, Davis, CA 95616, USA  
\textsuperscript{b} BASF AG, Agricultural Research Station, APS/US, D-67144 Limburgerhof, Germany  
\textsuperscript{c} Department of Entomology and Environmental Toxicology, University of California, Davis, CA 95616, USA

Received 12 April 1996; revised 23 July 1996; accepted 30 July 1996

Abstract

The information obtained from a single-antibody immunoassay can be ambiguous when the identity of the analyte is unknown and could be one of a group of compounds all having different affinities for the antibody. If we allow the possibility of mixtures of analytes the difficulty of the situation is aggravated. However, additional information can sometimes be obtained by assaying with a number of different antibodies. We demonstrate a rationale for identifying and quantifying analytes from a group of candidates, and illustrate the application of our method with an example from the class of s-triazine herbicides. Using a four-antibody array, correct identification and accurate quantification were usually achieved for single-analytes samples in the low ppb range. Mixtures of analytes were recognized as such but were more difficult to classify correctly, with some confusion arising between members of subgroups, particularly within the methoxy/methylthio substituted triazines. The mathematical perspective is used to suggest directions for improving the experimental performance of multianalyte immunoassay.

Keywords: Immunoassay; Cross-reactivity; Multianalyte

1. Introduction

In analytical chemistry, the identity of a compound is not usually proven directly: rather one attempts to disprove it under conditions of increasing rigor. When those conditions become sufficiently rigorous we then accept the resulting data as proof of identity. The problem of proof extends to all aspects of analytical chemistry, including immunoassay. Since immunoassays are not commonly coupled to chromatographic systems which might separate the compounds prior to analysis, one has the added complexity of wanting to being able to determine mixtures with detector systems of varying degrees of selectivity for analytes.

The quantitative data obtained from a single immunoassay only have a valid interpretation when either the antibody used is monospecific for a particular analyte or if the sample is known not to contain any possible cross-reactants; otherwise, it is impossible to tell whether the signal obtained is due...
to an amount of the analyte under investigation or a different amount of some other analyte which also has an affinity for the antibody being used. There are many ways to address this problem including the use of a completely different analytical system for validation of the results or the coupling of immunochemical detection methods with chromatographic separation [1]. One could also work on the development of antibodies which are more specific for the analytes of interest using either monoclonal or recombinant technology or careful hapten design and a collection of polyclonal antibodies. However, the use of a library of antibodies to look at the same group of analytes in different ways presents an alternative solution which offers several advantages.

In case of s-triazine herbicides, for example, there are so many possible commercial compounds that it is not economical to make a separate immunoassay for each compound. An array of less specific antibodies that can recognize overlapping classes of materials has the potential to be more efficient than a large batch of separate assays, while retaining the advantages of speed and low cost over conventional systems such as GC–MS.

A number of authors have examined the possibility of using an array of antibodies to enable identification and quantification from within a group of similar analytes. The general approach is discussed by Kauvar [2]. Cheung et al. [3] demonstrate the use of some multivariate statistical methods in analyzing the responses from several antibodies. Karu et al. [4] give an overview and evaluation of various statistical approaches. Wortberg et al. [5] describe the construction and application of an immunoarray in the case of s-triazine herbicides. The preceding deal chiefly with single-analyte samples. Analysis of mixtures of cross-reacting analytes, when the identity of each analyte in the mixture is known, has been demonstrated by Muldoon et al. [6]. A model for such mixture analysis was proposed by Jones et al. [7] and implemented successfully for mixtures of two, three or four analytes [8].

We describe below an assay procedure for samples which might contain one or more of a class of cross-reacting analytes, combining the work by earlier authors on pattern recognition and mixture analysis. First, we present a mathematical and statistical rationale for our procedure. Then we illustrate its application by using a four-antibody assay for eight s-triazine herbicides, including mixtures. Our examples and illustrations use competitive immunoassays with a coating hapten format [8]; the approach, however, is general and easily adapted to other formats.

2. Response paths

First consider the case of two cross-reacting analytes assayed with two different antibodies, with dose-response curves as shown in Fig. 1. In the absence of experimental error, we would know exactly the positions of the calibration curves. Given the responses \((Y_1, Y_2)\) from an unknown sample containing one or the other of the analytes, we could take each candidate analyte in turn and calculate an estimated concentration from each assay. In one case the estimates would be consistent and, in the other, they would not, provided that the analytes showed different patterns of cross-reactivity to the two antibodies. Thus we could identify the analyte as the one which gave a consistent estimate, and this estimate would be the true concentration. Furthermore, we could add more possible analytes and always, provided that the cross-reactivity patterns are sufficiently different, correctly identify the unknown using only two antibodies. Unfortunately experimental variation is unavoidable, so we have to pick the analyte which gives the most consistent estimates. In this section we consider how this consistency should be measured, by taking into account how experi-

![Fig. 1. Responses (optical densities) of two cross-reacting analytes to two different antibodies. The identity of a single-analyte unknown with responses \((Y_1, Y_2)\) is that which gives a consistent concentration estimate \(X\).](image)
mental error impinges on the estimated concentrations. The situation becomes clearer if we represent the assay responses in another fashion.

The two response curves from a single analyte (Fig. 1) can be combined into a single response path as follows: for a given concentration \( x \) of Analyte 1 we can read off from the curves the responses \( (Y_1, Y_2) \) and plot them as a single point in two-dimensional space (see Fig. 2). By varying \( x \) from 0 to \( \infty \) we get the complete response path, given parametrically by the concentration \( x \). This is then repeated for Analyte 2, as shown in Fig. 2. Note that, here, we are assuming common horizontal asymptotes for the two curves from each assay (since they will be jointly estimated from the same microtiter plate), so the response paths begin and end together. Whereas this is a reasonable assumption for competitive assays, it may not be true in general, so that the response paths may not be joined. The overall approach taken here will still be valid, although details of the implementation will change.

Now suppose we have the responses \( (Y_1, Y_2) \) from an unknown single-analyte sample. In the absence of experimental error, this point would have to lie on one or other of the response paths, thus identifying the analyte and enabling quantification. In practice, the point may lie between the two curves, as in Fig. 2; intuitively one would pick the curve nearest the sample point. However, the metric for measuring the distance needs to be chosen appropriately using a statistical model, so that the likelihood of the various possibilities can be evaluated. It may be that the sample point is so far from both response paths that either single analyte is unlikely; in such cases one would suspect a mixture.

Clearly this method will not be reliable for very small or very large concentrations, since here the response paths are close together: one would only be able to ascertain that one had a very small (or very large) concentration of something. Thus, there will be a workable range for the assay (as with single-antibody assays) within which reasonably reliable identification and quantification should be possible. This will depend in part on how far apart the two response paths are, which in turn depends on how different the patterns of cross-reactivity are to the antibodies used.

In practice, we are likely to have more than two analytes in a cross-reacting group, and to require more than two antibodies. We now develop a statistical model for the general case of \( n \) antibodies.

3. The statistical model

We model the individual dose-response curves for single-analyte samples using the four-parameter logistic model [9]. We assume further that the coefficient of variation of the responses \( Y \) from an individual assay is constant, so that \( \log Y \) has a constant standard deviation. There are more sophisticated and flexible ways of incorporating heteroscedasticity [10], but these further complicate the methodology and we have found the log transformation to be adequate. Thus our model is

\[
\log Y_i = \log \left( \frac{A_i - D_i}{1 + \left( \frac{x}{C_i} \right)^{B_i}} + D_i \right) + \epsilon_i, \quad i = 1, \ldots, n,
\]

where \( Y_i \) is the assay response from antibody \( i \), \( x \) the analyte concentration, \( A_i, B_i, C_i, D_i \) the model parameters and \( \epsilon_i \) an error assumed to have a normal distribution with zero mean and standard deviation \( \sigma_i \).
Here, \( A_i \) and \( D_i \) represent, respectively, the assay response at zero and infinite analyte concentration, \( C_i \) the \( IC_{50} \) (the concentration giving 50% inhibition), and \( B_i \) a slope parameter.

The values of \( A, B, C, D \) and \( \sigma \) would, in fact, be estimated by assaying a set of standard concentrations along with the unknowns. We assume here that this estimation is precise, so that the parameter values are known. It is convenient now to write

\[
\log Y_i = f(x) + \epsilon_i, \quad (2)
\]

where \( f(x) \) is a known function as given in Eq. (1). By assumption, \( \log Y_i \) is normally distributed with mean \( f_i(x) \) and standard deviation \( \sigma_i \), so we now find that

\[
\frac{\log Y_i - f_i(x)}{\sigma_i} = \frac{\log Y_i - f_i(x)}{\sigma_i} = (3)
\]

which represents the error distance in terms of multiples of its standard deviation, follows a standard normal distribution. Since the assays with different antibodies can be regarded as statistically independent these variables are independent for each \( i \). Statistical theory then suggests ([11], p.177) that

\[
d_i^2 = \sum_{i=1}^{n} \left( \frac{\log Y_i - f_i(x_i)}{\sigma_i} \right)^2 \sim \chi^2_n, \quad (4)
\]

i.e. the quantity \( d^2 \) as defined here should follow a known distribution: the chi-square distribution with \( n \) degrees of freedom. Furthermore, \( d \) can be regarded as a distance in \( n \)-dimensional space. If in Fig. 2, we re-scale each axis by taking logs and dividing by the estimated \( \sigma \), then \( d \) becomes the ordinary Euclidean distance between the sample point and the curve.

The true value \( x \) is, in fact, not-known, so we estimate it by the value \( X \) which gives the closest distance to the curve. This is found analytically by minimizing over \( x \) the expression for \( d^2 \) in Eq. (4). The resulting minimum distance should approximately follow a \( \chi^2_{n-1} \) distribution (one degree of freedom is lost because of the estimation of \( x \)). Thus tabulated values of the chi-square distribution can be used to decide if the distance from the sample point to a given response path is improbably large. For example, if the distance from the sample point to Analyte 2 in Fig. 2 gave \( d=1.0 \), tables (\( \chi^2 \)) tell us that the probability of obtaining such a result, if the sample were indeed Analyte 2, would be less than 0.01, and we would probably conclude that it was not Analyte 2. If both distances were improbably large, we could next consider the possibility of a mixture.

4. Mixture analysis

The possibility of a multi-analyte sample can be explored using the extended four-parameter logistic model of Jones et al. [7]. The response \( Y_i \) from a binary mixture of analytes with concentrations \( (x_1, x_2) \) is modeled by

\[
\log Y_i = \log \left( \frac{A_i - D_i}{1 + \left( \frac{B_i}{C_i} \right) \left( \frac{B_i}{B_i^*} \right)^{x_i}} \right) + D_i + \epsilon_i, \quad i = 1, \cdots, n, \quad (5)
\]

where \( A_i, B_i, C_i, D_i \) are the parameters of the calibration curve for analyte \( j \) with antibody \( i \), and \( B_i^* \) is the geometric mean of \( B_{1i} \) and \( B_{2j} \). If we use two antibodies, we solve a pair of non-linear simultaneous equations for the estimation of \( (x_1, x_2) \); if there are more than two antibodies, we choose \( (x_1, x_2) \) to minimize \( d^2 \) of Eq. (4), with \( f_i(x) \) replaced by \( f_i(x_1, x_2) \) as in Eq. (5). As before, if we are using the correct pair of analytes for a given sample, \( d^2 \) will follow a chi-squared distribution. Since we are now estimating two extra parameters, \( x_1 \) and \( x_2 \), the appropriate distribution will be \( \chi^2_{n-2} \).

Our proposed analysis, having obtained estimates of the curve parameters using standard concentrations of all analytes in the group, is to first calculate the distance from a sample point to each of the single-analyte response paths. If any of these are plausible when referred to the appropriate chi-square distribution, the corresponding analyte and concentration are considered as plausible determinations of the unknown sample (there may be more than one plausible analyte within the group). If all single analytes give implausible answers, we search all possible binary combinations until a set of plausible solutions is found.

In practice this may produce several possibilities for the composition of the unknown sample, but these
will be ranked in order of plausibility by their $d^2$ value. Strict adherence to the maximum likelihood principle ([11], p.254) would suggest that we accept the solution with the smallest $d^2$, but it is important to be aware of other close alternatives. It may therefore be worthwhile proceeding with the binary mixture analysis even if we obtain a satisfactory solution at the first stage. One could go a stage further and examine possible ternary mixtures, but without a large number of discriminating antibodies this is likely to produce a confused picture since there may be many acceptable combinations of analytes. Further refinement of the experimental procedure, particularly in the careful choice of suitable antibodies, will be necessary for the successful analysis of complex mixtures.

5. Examples

We now illustrate and evaluate this method of analysis using single analytes and binary mixtures chosen from the class of s-triazine herbicides and their metabolites (see Fig. 3). For each unknown sample there are three decisions to be made: is the unknown a single analyte or a mixture, which analytes are present, and at what concentrations?

5.1. Single analyte

Here, we re-analyze the data given in Wortberg et al. [5]. Standard curves were obtained for prometon, atrazine, simazine, cyanazine, hydroxyatrazine, prometryn, terbutryn and deethylatrazine on a single microtiter plate, together with fourteen unknown samples in duplicate (see Fig. 4). Four such plates were treated with four different antibodies: AM7B2.1 [12], K1F4 [13], #2652 [14] and #4653 [15]. This was repeated with another set of four plates using different unknowns, thus giving a total of 28 unknown samples to be determined. The unknowns were all single analytes at concentrations of 0.75, 1.5 or 5 ppb, excepting two samples per plate which were negative controls. A full description of the assay procedure is given in Wortberg et al. [5].

In the analysis, we first assume a single analyte and calculate the minimum distance to each of the eight response paths, using Eq. (4). These distances are referred to a $\chi^2$ distribution (e.g. the 95th percentage point is 7.81) for assessing their likelihood. For example, one sample contained simazine at 5 ppb. The results for this sample are shown in Table 1. The only acceptable possibility is that the sample contains simazine, and the estimated concentration of 4.65 ppb turns out to be quite accurate.

A less successful example is given by the sample containing 0.75 ppb terbutryn. The results for this sample are shown in Table 1. The only acceptable possibility is that the sample contains simazine, and the estimated concentration of 4.65 ppb turns out to be quite accurate.

A less successful example is given by the sample containing 0.75 ppb terbutryn. The results for this sample are shown
in Table 2. None of the single-analyte possibilities is really acceptable, and the distances for prometon, prometryn and terbutryn are similar. (The difficulty of distinguishing between these three with this array of antibodies was noted by Wortberg et al. [5]). We now search through all possible binary combinations, and find that the fit can be improved significantly by including a small amount of hydroxyatrazine, although the resulting distance is still rather large. Our conclusion would probably be that the sample contains either prometon, prometryn or terbutryn, with perhaps a small amount of hydroxyatrazine. The estimate for hydroxyatrazine in the mixture analysis was 0.05 ppb, a seemingly negligible amount, but the antibody here, #4653, was extremely sensitive to hydroxyatrazine ($IC_{50}=0.1$ ppb), enough to reduce the $d^2$ value from 15.97 to 7.64. Close examination of the data revealed that a few adjacent samples had the same characteristic of wanting to add hydroxyatrazine; apparently, a spatial effect on one of the plates was causing a false positive. Spatial trends can sometimes be observed on microplate data [16], variously ascribed to inhomogeneity of the plate material, temperature gradients, misalignment of the plate reader and other effects.

A summary of the results for the 24 positive samples is given in Table 3, showing whether the correct analyte was identified at the first stage, whether the distance statistic at this stage was acceptably small, and whether this could be improved using binary mixture analysis. Out of the 24 samples, 19 succeeded in identifying the correct analyte at the first stage, but in nine of those cases the distance from the model, $d^2$, was unacceptably large and in seven of those the addition of a second analyte (i.e. a binary mixture) gave a significant improvement in fit. Four of the five incorrect identifications arose from confusion between prometon, prometryne and terbutryn. The fifth was actually 0.75 ppb simazine, and the $d^2$ value of 4.19 was acceptable for simazine but a better fit was achieved by assuming cyanazine ($d^2=2.60$).

### 5.2. Binary mixtures

The goal here was to identify and quantify binary mixtures of triazine herbicides out of a pool of eight possible candidates. These herbicides were atrazine, simazine, cyanazine, prometryn, prometon, terbutryn and the two atrazine metabolites hydroxyatrazine and desisopropylatrazine. For this experiment, we used the antibodies AM7B2.1, K1F4, #4652 [15] and #2282 [17]. The plate template and the general assay procedure were as in the single-analyte experiment above. Again two sets of plates were used, giving a total of 28 unknown samples to be determined.
were binary mixtures of 1 ppb of each of two analytes, except for two negative controls and one sample of 1 ppb atrazine only.

Analysis proceeded as above, first assuming a single analyte and comparing the $d^2$'s with a $\chi^2$ distribution, then, if this assumption proved untenable, assuming a binary mixture and comparing with $\chi^2$. We illustrate here with the results from a sample containing 1 ppb simazine and 1 ppb prometon. Single analyte analysis gave the shortest distance to any of the response paths as 29.7, which, when referred to $\chi^2$, has a probability less than 0.00001. Mixture analysis indicates three possible binary mixtures (with acceptably small $d^2$) as shown in Table 4. The rather high estimate for deisopropylatrazine reflects the lower sensitivity to this analyte in our chosen array of antibodies. We would probably conclude that the unknown contained either atrazine and prometon or simazine and prometon. We can illustrate the uncertainty in the estimated concentrations by calculating $d^2$ over an array of values near the estimates and drawing a contour plot as in Fig. 5. The use of such plots for producing confidence regions for the estimates is under investigation.

A summary of the results of the analysis of the 25 mixtures is given in Table 5. As noted by Wortberg et al. [5], the analytes tend to fall into groups with respect to their cross-reactivities: the chloro-s-triazines (atrazine, simazine, cyanazine), the methoxy/methylthio-s-triazines (prometon, prometryn and terbutryn) and hydroxymetabolites (hydroxyatrazine). The dealkylated chloro-s-triazine (deisopropylatrazine) was relatively unreactive with all our antibodies but tended to behave like the first group, and its presence was often masked by the other analytes. As the overall results show, the assay was quite successful at indicating the correct group or groups, but less successful at distinguishing between possible combinations within groups. Thus, for example, a mixture of atrazine and prometryn was identified as cyanazine and prometon, or possibly cyanazine and terbutryn. A mixture of prometon and terbutryn appeared to contain terbutryn only. The estimated concentrations using the correct identities were reasonably accurate, whether or not this was identified as one of the possible answers.

### Table 4
Results for binary mixture analysis of a sample containing 1 ppb simazine and 1 ppb prometon. Only three assumed mixtures are shown; the other 25 combinations gave $p$-values less than 0.0007.

<table>
<thead>
<tr>
<th>Assumed analytes</th>
<th>Conc. (ppb)</th>
<th>$d^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Atrazine</td>
<td>Prometon</td>
<td>0.42</td>
<td>0.67</td>
</tr>
<tr>
<td>Simazine</td>
<td>Prometon</td>
<td>0.67</td>
<td>0.72</td>
</tr>
<tr>
<td>Diatrazine</td>
<td>Prometon</td>
<td>12.76</td>
<td>0.72</td>
</tr>
</tbody>
</table>

### Table 5
Summary of results from assaying 25 binary mixtures showing the outcome and whether the $d^2$ statistic was acceptably low.

<table>
<thead>
<tr>
<th>Correct mixture</th>
<th>Correct group</th>
<th>Single analyte</th>
<th>Incorrect group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable</td>
<td>3</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>$d^2$ too high</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

### 6. Discussion

We have illustrated a methodology for immunoanalysis of samples which might contain one or more from a group of cross-reacting analytes. One obvious application is in cases where a complete set of monospecific antibodies for each of the analytes does not exist. However, this approach...
might be useful even when there are monospecific antibodies available. If samples could be placed into groups using a small number of antibodies with significant cross-reactivities, and individual members of each group could then be identified using a smaller number of more specific antibodies, then the resulting assay procedure could generally use less antibodies and, hence, be more efficient. We could thus have a hierarchical system starting with a single screening assay, then a multiple assay for positive samples sorting them into groups, and a final assay for each group to distinguish between its members.

Our results suggest that the method has potential, but we are still far from reliable identification in every case. Further difficulties could be expected in applying the method to mixtures of more than two analytes, or binary mixtures in which one component was at a much higher concentration than the other. Although some success has already been achieved in this area in cases where the number of candidate analytes is small [8], the difficulty is compounded here by having a large number of possible candidates. In the present situation, our particular antibody array lacked the power to discriminate between certain combinations of analytes and we now consider some reasons for this.

Firstly, the pattern of cross-reactivities has to be sufficiently different for each analyte. We have noted the difficulty in separating prometon, prometryn and terbutryn: this occurs because they all had similar cross-reactivity patterns across the antibodies. We are developing antibodies with more discriminatory power which should increase the utility of our approach.

Secondly, ranges of sensitivity were different. For example, in the first assay the lowest \( IC_{50} \) for deethylatrazine was 12 ppb: for hydroxyatrazine it was 0.1 ppb. The assay could perhaps be improved by decreasing the sensitivity of some of the assays to get a similar dynamic range to each analyte. Another and possibly preferable solution would be to use a dilution series for each unknown.

Thirdly, the size of experimental error (coefficient of variation) can be crucial for multi-analyte analysis. If the individual assays are not very precise this can seriously degrade the performance of the multiple assay. We have noted above a problem with spatial effects, which have the potential to give rise to very misleading conclusions (see also [16]). Precision could be greatly improved if spatial variations could be eliminated. Another factor contributing to low precision is the limited space available on the microtiter plate, so that there has to be a small number of standards for the calibration curves, and few replicates of each unknown. A possible solution is to use separate plates as suggested by Jones et al. [18]. Analytical chemists can certainly encourage manufacturers to work in these directions. As we look into the future of immunoassay technology it is likely that we will see miniaturized immunoassays which among other things will allow more replicates of each assay and many assays to be run on a single plate (see [19]). This will certainly increase precision as computer averaging systems have done in other branches of analytical chemistry such as mass spectrophotometry [20].

Finally, the assumptions made in deriving the distribution of the minimum \( d^2 \) statistic were probably not valid for our data, particularly our assumption that the curve parameters \( A, B, C, D \) and \( \sigma \) are precisely estimated. This would explain why our \( d^2 \) statistic was sometimes too large. The use of separate plates, enabling more standards to be used for each curve, might improve this; otherwise a more complicated statistical argument would be required.

We are currently investigating all of the above possibilities, and hope that some of these approaches can be extended more generally to assist in the interpretation of results from a variety of methods in multi-analyte analysis.

Acknowledgements

This research was supported by NIEHS Superfund 2P42-ES04699, NSF DMS 93-01344, 94-06193 and 95-10511, US EPA CR 819047, USDA Forest Service NAPIAP R8-27, Center for Ecological Health Research CR 819658, NIEHS Center for Environmental Health Sciences IP30-ES05707 and Water Resources Center grant No. W-840. B.D. Hammock is a Burroughs Wellcome Toxicology Scholar. M. Wortberg is a fellow of the Deutsche Forschungsgemeinschaft.
References