REPORT OF A COLLABORATIVE STUDY ON PROPOSED
HIV-1 P24 ANTIGEN REFERENCE REAGENTS

AJ Garrett and AB Heath
National Institute for Biological Standards and Control
Blanche Lane, South Mimms, Potters Bar, Herts EN6 3QG, UK

Introduction

The 24 kilodalton gag gene product, p24, is the major internal structural protein of human immunodeficiency virus type 1 (HIV-1).

Measurements of p24 antigenaemia provide valuable diagnostic and prognostic information on individuals infected with HIV-1 and patients with AIDS and allow assessments of the effectiveness of antiviral therapy in vivo and in vitro.

There is, therefore, a need for an international reference material that can enable reliable comparisons of p24 assays to be made within and between laboratories.

The aim of this collaborative study is to assess the suitability of five candidate preparations of HIV-1 p24 to serve as reference reagents in assays currently used in investigations of HIV infections.

Participants

Eleven laboratories were invited to participate in this study, and results were received from all of them. The laboratories are listed below and are referred to throughout by a code number allocated at random, not necessarily reflecting the order in the list.

Dr J Ball, Regional Virus Laboratory, East Birmingham Hospital, Birmingham B9 5ST
Dr F Barre-Sinoussi, Institute Pasteur, 28 rue du Dr Roux, 75724 Paris, Cedex 15, France
Dr J Eberle, Max von Pettenkofer Institute, 8000 Munchen 2, Germany
Dr AJ Garrett, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG
Dr R Najera-Morrondo, Instituto de Salud Carlos III, 28220 Majadahonda, Spain

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.
Dr F Rae, National Reference Laboratory, Fairfield Hospital, PO Box 65, Yarra Bend Road, Fairfield, Australia

Mr C Ronalds, Division of Virology, Medical College, St Bartholomews Hospital, West Smithfield, London EC1A 7BE

Dr C Schable, Division of HIV/AIDS, Building 1, Room 2367, MS D12, Center for Disease Control, Atlanta, Georgia 30333, USA

Dr R Tedder, Virology Section, University College and Middlesex Medical School, Riding House Street, London W1P 7PN

Dr R Thorstensson, The National Bacteriological Laboratory, Lundagatan, S-105 21 Stockholm, Sweden

Dr G Vercauteren, Institute of Tropical Medicine, Nationalestraat 155, Belgium

Materials

The candidate materials were prepared from p24;

1) Detergent-treated and unfractionated HIV (study code: C)
2) Peptide isolated from detergent-treated HIV (study code: A&F)
3) p24 peptides isolated from recombinant baculovirus and E.coli (study code: B,D,E).

Samples of each material were diluted in human serum albumin and glass-distilled water to give a final concentration of 6mg/ml albumin. Portions, at a nominal weight of 1.0g per ampoule, were added to 4000 ampoules which were then freeze-dried in a Lyomax 2 freeze-drier. Secondary desiccation continued for 6 days.

The serum albumin was negative for HBsAg and anti-HIV. The preparations (study code: A/F & C) derived from virus lysates were shown to have no detectable HIV infectivity.

Five preparations were included in the study, and, labelled with anonymous code letters only, were sent to participants. The preparations were designated the following codes:

<table>
<thead>
<tr>
<th>Study Code</th>
<th>NIBSC Code</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+F</td>
<td>90/636</td>
<td>virus</td>
</tr>
<tr>
<td>B</td>
<td>90/640</td>
<td>rec.baculovirus</td>
</tr>
<tr>
<td>C</td>
<td>90/516</td>
<td>virus</td>
</tr>
<tr>
<td>D</td>
<td>90/514</td>
<td>rec. E.coli</td>
</tr>
<tr>
<td>E</td>
<td>90/512</td>
<td>rec. E.coli</td>
</tr>
</tbody>
</table>

Assay Methods

Participants used either commercially available assay kits from different manufacturers or, in some cases, kits that they had developed themselves. Throughout this report, the kit/method used is referred to by a code letter (a - g).
Study Design

Participants were each sent ampoules of six samples coded A - F. The samples coded A & F were duplicates of the same preparation. They were initially requested to assay all samples concurrently, on three separate occasions, using fresh ampoules of each sample on each occasion, and using the assay method routinely used in their laboratory. After the study had commenced, additional assay kits were made available for use in the study, allowing some laboratories to perform limited numbers of additional assays using different methods.

Laboratories were requested to return raw assay data, in the form of optical densities for dilution series of each preparation. Each dilution step for each preparation was tested in duplicate.

The high cost of the commercial kits limited the number of repeat assays that could be requested, and also the extent of replication within an assay (e.g. no replication on separate plates).

Statistical Methods

All results were analysed as parallel line assays (Finney 1978), relating log of optical density (OD) to log dose (dilution), after selecting areas of the dose-response curve that approximated to the conditions of linearity and parallelism. The usual analysis of variance checks on assay validity were performed.

The potency of preparations B-F were expressed relative to preparation A which was treated as a hypothetical standard.

Variation within laboratories, between assays (using the same method) was expressed as percentage geometric coefficients of variation (%gcv) (Kirkwood 1979).

Variation between laboratories was assessed using a components of variance analysis of variance.

The sensitivity of the assays for detecting p24 antigen was investigated by calculating the estimated titre for each preparation. This was calculated as the maximum dilution for which the preparation was ‘positive’, after taking the mean OD from the two replicates, and comparing to a cutoff value defined by the assay kit being used (e.g. mean OD for negative controls + 0.05). For each assay, the cutoff was taken as that defined by the particular kit being used. Laboratory 8 used its own assay kit, and did not supply a definition of a cutoff or ‘positive’ or ‘negative’. For these assays, the cutoff was taken to be the mean OD for negative controls + 0.05, as this was the commonest definition in use with the commercial kits.

Results

Assay validity:

All assay data were plotted. For many laboratories, the dilution ranges used were such that the dose-response curve was predominantly flat with very low ODs. However, after taking a logarithmic transform of the ODs it was possible to select dilutions which appeared by eye to give adequate
linear and parallel dose response lines. In some cases this meant restricting analysis to the lowest two or three dilution steps of the series. The manufacturers' references were not included in the analysis, as there were some visual suggestions of non-parallelism.

Some assays were invalid on the basis of the assay kit criteria for valid negative control data and cutoff values. The affected assays were Laboratory 3, assay 5, kit 'e'; Laboratory 5, assays 1-3, kit 'f'; Laboratory 7, assay 4, kit 'd'; Laboratory 7, assay 5, kit 'e'; Laboratory 11, assay 10, kit 'd'; These assays were excluded from further analysis.

In addition, some assays were excluded completely (laboratory 9, one assay; laboratory 10, one assay; laboratory 11, three assays), where it was not possible to select a range of dilutions with adequate linear and parallel dose-response lines. In other cases, data for a single preparation was excluded from an assay, if the dose-response for this preparation was inadequate.

For the remaining assays, the usual analysis of variance tests indicated poor parallelism, with 35% (13/37) assays being significantly non-parallel, at a 5% significance level, and 24% (9/37) at the 1% level. However, repeating the analysis for just the coded duplicates A and F, which, since they are the same material, must have parallel dose response lines against log dose, 17% of the assays were still indicated as significantly non-parallel at the .5% level, by the analysis of variance tests. This indicates that there is a problem with the tests, rather than real non-parallelism.

For each assay, the slope of the dose-response line for each preparation over the selected dilution range was expressed as a proportion of that of the standard A. Comparing these proportions across all assays, none was significantly different from 1.0, indicating that none of the preparations was exhibiting consistent non-parallelism.

Looking at the confidence intervals for estimated potency of F relative to A (which must be 1.0, by definition), the estimated potency of F was significantly different from 1.0 for 58% of the assays.

These results for the coded duplicates suggest that the underlying residual variation in the assay is not being adequately estimated by the variation between the duplicate ODs at each dilution step. An underestimation of residual variation would explain the anomalous results from the comparison of the coded duplicates. The underestimation may be due to the assay design, for example duplicate OD's not being the result of completely independent dilution steps, or some plate effect that results in adjacent wells having closer results than would occur with random allocation across a plate.

As a result, the analysis of variance validity tests were ignored, and the selection of dilution ranges, and assessment of assay validity were performed by eye.

It should be emphasised that the above results do not invalidate the assays; only the interpretation of the validity tests.
Relative Potency:

The potency of preparations B to F were calculated relative to preparation A. The results for individual assays are shown in histogram form in figures 1-5. Each box represents the result of an individual assay, and the boxes are labelled with the laboratory code number, and kit/method code.

Figure 5 shows that there is good agreement between laboratories and methods for determining the potency of preparation F relative to A, with a 3.1-fold range in estimates. As these materials are duplicates, this is to be expected. Any variation from a potency of 1.0 will be due to assay to assay variation.

Figure 2 also indicates good agreement between laboratories in the estimation of potency of C relative to A, with a 3.3-fold range.

For the other preparations, B, D and E, the agreement between laboratories and methods is not as good. For preparation B there is a 32.7-fold range of estimates. Figure 1 indicates that for preparation B, all assays carried out with kit 'c' gave substantially lower estimates of potency than all other assays. Two assays from laboratory 11, using kit 'c', were also excluded, as they resulted in very low ODs for preparation B, which could not be included in the parallel line analysis. As a result, they are not included on the figure. Laboratory 8 also has a markedly lower estimated potency from one of their three assays. If the assays from kit 'c' are all excluded, the range of estimates drops to 19.4-fold, the single low estimate from laboratory 8 keeping the range high.

For preparation E, there was a 22.3-fold range of estimates. There is one assay that gives a substantially lower estimate of potency than the others, and again this is with kit 'c'. The other assay shown on the figure using kit 'c' also gives low potency estimates. Eight other assays with kit 'c' were excluded from the parallel line analysis, as a result of very low o.d.'s for preparation E (Lab 11 - 2 assays; Lab 5 - 1 assay; lab 3 - 2 assays; lab 9 - 3 assays), and as a result are not marked on the figure. If the assays from kit 'c' are all excluded, the range of potency estimates drops to 6.4-fold.

The estimated potencies for preparation D (figure 3) are not in as good agreement as for C or F. The range of estimates is 16.8-fold. The estimates do not show a consistent bias for any particular method or laboratory. The results from assays using kit 'c' are in line with the other methods.

Within Laboratory - Between Assay Variation:

The assay to assay variation within laboratories was assessed by calculating the %gcv of potency estimates within each laboratory, for each preparation and method separately. These estimates were then pooled to give an overall %gcv for each preparation. The primary purpose of this analysis was to compare the behaviour of each preparation in the assay system. The low number of assays performed by each laboratory does not allow reliable estimates of within laboratory variation to be presented for each individual laboratory.

The pooled estimates of between assay repeatability (%gcv) are shown in table 1. The value for the coded duplicate F (at 19%) is lower than for
other preparations. Preparation E has a higher figure (44%), whilst the others are all around 30%.

Inspection of the individual assay results suggested that the higher figure for E was the result of high variation in the assays of laboratory 2, caused by an apparently discrepant single assay for E. If assays from laboratory 2 are excluded, the within laboratory repeatability figure for E is in line with other preparations (20 - 30%).

Between Laboratory Variation:

A components of variance analysis of variance was performed. For each laboratory, the analysis was restricted to the assay method for which most assays had been performed. The analysis indicated that for preparations C, E and F, there was no significant additional variation between laboratories, in estimates of relative potency, when compared to the within laboratory assay to assay variation. However, for preparation E, the exclusions of assays from kit 'c' mentioned above were not taken into account by this analysis. The kit 'c' assays shown on the plot were also excluded, as they were not the main assay method for those laboratories. For preparations B and D, there was a significant component of additional variation between laboratories (p < 0.01). The estimated additional between laboratory %gcv's were 123% (preparation B) and 38% (preparation D). The high value for preparation B is due to the low potency estimates from assays with kit 'c'.

Kit Manufacturers References:

Some of the assay kits include references supplied by the manufacturers, calibrated in terms of pg/ml of antigen. For these assays, the parallel line analysis was repeated, including the kit references. Four of the fourteen assays were non-parallel, according to the analysis of variance validity tests. However, the reservations described above apply to the interpretation of these tests. The potency of A was calculated relative to the kit references, and then expressed in ng/ml, using the calibration supplied by the manufacturers for their kit reference. These assays are shown in figure 6. There is some suggestion that the references for kit 'a' are giving lower results than the others, but with the limited number of assays, it is not possible to distinguish consistent kit differences from laboratory or assay to assay variation.

The mean potency of A from all these assays was 10.9 ng/ml, with a range of 2.4 to 26.9 ng/ml.

Assay Sensitivity:

The titres, or maximal dilution at which the preparations were positive, based on the kit criteria, were calculated for each assay and preparation. They are shown in histogram form in figures 7 - 12.

There was considerable variation between the calculated titres, with ranges of 40-fold (D), 64-fold (A, B and F), 128-fold (C) and 160-fold (E).
The assays with kit 'c' gave generally higher titres for preparations A, C and F. For preparations B and E they were generally lower than other assays. For E, two assays from lab 11 and one from lab 5 (all with kit ‘c’) were not positive at any dilution tested, and are not included in the figures. Both these factors contribute to the low relative potencies of B and E, when taking A as a standard.

Discussion

This study has demonstrated that it is possible to assess p24 antigen content using a parallel line assay method. The dilution ranges used in many of the assays included in this study were not ideal. However, by restricting analysis to selected subsets of the dilution ranges, it was possible to obtain adequate linear and parallel dose-response lines for the majority of assays.

Analysis of the data from the coded duplicate preparations A and F indicated that the usual statistical tests of validity, and confidence intervals, were not giving realistic results, due to underestimation of the residual between replicate error, possibly as a result of assay design. As a result, the dilutions and data used for the parallel line assay analysis were selected by eye, and the validity tests ignored. It should be emphasised that these problems do not invalidate the assays, only the interpretation of the usual statistical validity tests.

The extent of replication within each assay was limited by the high cost of the commercially available kits used for the majority of assays.

The potency of the preparations B - F was expressed relative to A, using A as a hypothetical standard. For the coded duplicate F, and for preparation C there was good agreement between laboratories and methods. For the other preparations, agreement was less good. Since A, C and F were derived from tissue culture-grown viruses, and the others from recombinant materials, this demonstrates that better agreement was obtained in virus-virus comparison, than for a virus-recombinant comparison and suggests that a virus-derived preparation may be preferable for use as an International Reference Reagent or Standard.

In particular, one of the commercially available kits (coded ‘c’) gave substantially lower potencies for the recombinant preparations B and E than other kits. In the case of preparation E, the resultant ODs were too low to be included in the analysis for most assays. These low relative potencies were caused by low responses of the kit to preparations B and E, and also a stronger response to the standard A. It would seem that something in the nature of the kit design is giving a differential response to the different materials and, again, suggests that a recombinant based standard may be unsuitable.

The within laboratory - between assay variability was similar for all preparations, with %gcv’s between 20% and 30% (table 1). For preparations A, C, E and F, there was no statistically significant variation between laboratories, over and above that expected from assay to assay variation. The low potencies from kit ‘c’ were not included in this analysis, for preparation E, however. For preparations B and D, the additional variation between laboratories was significant (p < 0.01), and resulted in %gcv’s between laboratories of 123% and 38% respectively. For preparation B, this result is clearly affected by the very low potency estimates from assays with kit ‘c’.
Expressing the potency of A relative to the manufacturers’ kit references, for the fourteen assays that included such references, indicated considerable variation in estimates, ranging from 2.4 ng/ml to 26.9 ng/ml, with a mean of 10.9 ng/ml. There was some suggestion that kit ‘a’ gave lower estimates than other kits, but with such low numbers of assays, it is not possible to separate consistent kit differences from between laboratory or between assay variation.

The sensitivity of the assays for detecting the preparations as "positive", by the criteria specified for the individual kits, varied considerably. The calculated titres (maximal dilution for which the preparation was positive) had ranges from 40-fold to 160-fold.

Conclusions

The use of the parallel line assay method allowed direct quantification of p24 antigen content relative to a standard. Agreement between laboratories and between assay methods was best when potencies of virus-derived candidate preparations were compared. Comparing potencies of recombinant- and virus-derived materials gave poorer agreement.

With one kit, coded ‘c’, there were major discrepancies, relative to other assays, in assessing the potency of two of the recombinant-derived materials B and E.

Subject to satisfactory stability studies, currently in progress, the preparation coded A and F seems suitable to serve as an International Reference Reagent, which we propose should be assigned a unitage of 1000.

Acknowledgements

We are grateful to Dr FA Liberatore, Dupont de Nemours & Co, North Billerica, Massachusetts, 01862, USA; Dr JF Delagneau, Diagnostics Pasteur, 3 Boulevard Raymond Poincare, 92430, Marnes-La-Coquette, France; Mr F Volvovitz, MicroGeneSys Inc, 400 Frontage Road, West Haven, Connecticut, 06516, USA;

Dr MP de Meyer, Innogenetics NV, Kronenburgstraat 45, B-2000 Antwerp, Belgium, for generously donating the materials used in this study, to the Abbott, Dupont and Innogenetics companies for donating some of the assay kits, to Dr P Dawson for advising on and supervising the freeze drying operations and to the participants from the eleven laboratories that contributed data.

References


Table 1

Within Laboratory Repeatability
%gcv's pooled for each preparation

Including laboratory 2:

<table>
<thead>
<tr>
<th>Preparation</th>
<th>%gcv</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>29%</td>
</tr>
<tr>
<td>C</td>
<td>28%</td>
</tr>
<tr>
<td>D</td>
<td>29%</td>
</tr>
<tr>
<td>E</td>
<td>44%</td>
</tr>
<tr>
<td>F</td>
<td>19%</td>
</tr>
</tbody>
</table>

Excluding laboratory 2:

<table>
<thead>
<tr>
<th>Preparation</th>
<th>%gcv</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>31%</td>
</tr>
<tr>
<td>C</td>
<td>19%</td>
</tr>
<tr>
<td>D</td>
<td>30%</td>
</tr>
<tr>
<td>E</td>
<td>21%</td>
</tr>
<tr>
<td>F</td>
<td>19%</td>
</tr>
</tbody>
</table>
Legends to figures

Figures 1-5  Histograms showing the potency of preparations B to F, relative to preparation A.

Figure 6  Histogram showing the potency of preparation A relative to the assay kit manufacturers reference, for assays that included one. Potency calibrated in ng/ml, using kit manufacturer's data.

Figures 7-12  Histograms showing the titre (maximal dilution at which preparation positive) for preparations A to F.

In all figures, each assay is represented by a box, and is labelled with the laboratory code number, and a code letter representing the assay method used.
Fig 1 – Potency of B relative to A
Fig 2 – Potency of C relative to A

Number of Assays

Potency relative to A
Fig 3  Potency of D relative to A

Number of Assays

Potency relative to A
Fig 8 – Titres of B
Fig 10  Titres of D

Number of Assays

Titre

0  2  4  6  8  10  12  14  16  18  20  22  24  26  28  30

8g  8g  8g  7a  7a  7a  7a

3a  3a  3a  3a

11c  11c  11c

11a  11a  11a

10d  10d  10d

5c  5c  5c  5c  5c
Fig 11 – Titres of E

Number of Assays

Titre

10  20  40  80  160  320  640  1280  2560  5120

9c  8g  6d  6a  7a  3c  3a  4a  2b  10d  11a  2b  11d

7c  11b  10d  1e  11a  11b  2b  3d