*Journal of Chromatography*, 512 (1990) 325–335
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 1902

# Peptide mapping of HIV polypeptides expressed in *Escherichia coli*

# Quality control of different batches and identification of tryptic fragments containing residues of aromatic amino acids or cysteine

STAFFAN RENLUND* and ING-MARIE KLINTROT
*Pharmacia LKB Biotechnology, S-751 82 Uppsala (Sweden)*
MICHAEL NUNN and JEFFREY L. SCHRIMSHER
*Pharmacia Genetic Engineering Inc., La Jolla, CA (U.S.A.)*
and
CHRISTER WERNSTEDT and ULF HELLMAN
*Ludwig Institute for Cancer Research, Uppsala Branch, Uppsala (Sweden)*

## ABSTRACT

Peptide mapping was used for the quality control of different batches of the recombinant HIV proteins p24 core and p24–gp41, expressed in *Escherichia coli*. These proteins comprise *gag* and *env* region polypeptides of the virus and may serve as suitable components in the diagnosis of HIV infections. The proteins were digested with trypsin and the mixtures were subjected to peptide mapping to prove batch equivalence of p24–gp41 and to isolate fragments of the p24–gp41 digest that differ from those of the p24 core digest. The proteins were reduced with dithiothreitol and the cysteine residues were derivatized by addition of 4-vinylpyridine. Peptide mapping was performed by means of reversed-phase high-performance liquid chromatography. Batch equivalence was proved by comparison of the maps. Peaks present in one map but not in the other were considered to be due to sequence differences or variability in digestion.

## INTRODUCTION

The outer membrane of the human immunodeficiency virus (HIV) contains a glycoprotein gp41[1–3] which is important in the attachment of the virus to the surface of the host. The core, which carries the RNA, contains a 232-residue polypeptide, p24 (refs. 1–3). A fusion protein, consisting of p24 core and part of gp41, was produced using the recombinant DNA technique (Fig. 1). This polypeptide, called p24–gp41, may be a suitable reagent in the diagnosis of HIV-1 infections.

0021-9673/90/$03.50 

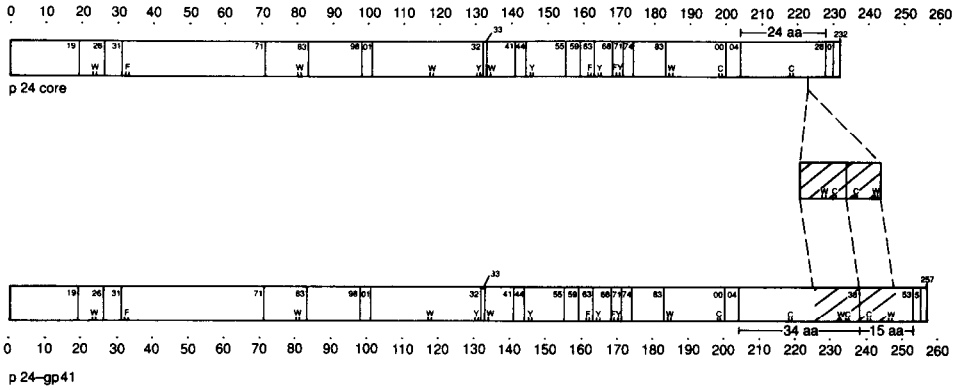


Fig. 1. General structures of p24 core and p24–gp41. The figures in the boxes indicate the positions of the Lys and Arg residues and thus the fragments produced by complete tryptic digestion. The aromatic amino acid residues and the cysteine residues are indicated with one-letter symbols.

Peptide mapping has been used for fragmentation in protein sequencing[4–7] and to detect lot-to-lot variations in recombinant proteins[8,9]. The aim of this study was to perform a quality control of different batches of p24–gp41 by means of peptide mapping of their tryptic digests.

The reduced and unfolded polypeptides were derivatized using 4-vinylpyridine[10,11], generating pyridylethylated cysteine residues with an absorbance maximum at 254 nm. Tentative assignments were made by multiple-wavelength detection at 215, 254 and 280 nm. Sequence analysis confirmed those fragments that differed in the tryptic fingerprints of p24 and p24–gp41.

## EXPERIMENTAL

### *Chemicals and reagents*

Acetic acid (No. 55), acetonitrile (No. 14291), tris(hydroxymethyl)aminomethane (Tris) (No. 8382), EDTA (No 8418), 4-vinylpyridine (No. 808513) and water (LiChrosolv No. 15333) were obtained from E. Merck (Darmstadt, F.R.G.), ammonium hydrogencarbonate (No. 10302) and guanidine hydrochloride (No. 45208) from BDH (Poole, U.K.), triethylamine (TEA) (No. 25108) and trifluoroacetic acid (TFA) (No. 28902) from Pierce (Rockford, IL, U.S.A.), dithiothreitol (DTT) (No. D-0632) and N-tosyl-L-phenylalanyl chloromethyl ketone (TPCK)-treated trypsin (No. T-8642) from Sigma (St. Louis, MO, U.S.A.) and anhydrous calcium chloride (No. 4132) from Mallinckrodt (St. Louis, MO, U.S.A.). All chemicals were of highest available purity.

### *Samples*

The HIV-derived polypeptides p24–gp41 and p24 core were produced using the recombinant DNA technique. p24 core is coded by the *gag* gene of the HIV and contains 232 amino acid residues[1]. p24–gp41 constitutes a fusion of p24 and part of gp41, a glycoprotein found in the outer membrane or envelope of the HIV[1–3], making the non-glycosylated fusion protein 257 residues long.

The purification of the two polypeptides from *Escherichia coli* included

suspension and lysis of the cells in 0.1 mol/l sodium phosphate buffer (pH 7.0) containing 0.17% (w/w) lysozyme, centrifugation of the homogenate, resuspension of the pellet in Tris buffer containing urea and dithiothreitol, centrifugation, stirring of the supernatant with guanidine hydrochloride and DTT and overnight dialysis. The solution was then centrifuged and the supernatant desalted by gel filtration. The p24 polypeptides were obtained from the desalted protein fraction by cation-exchange

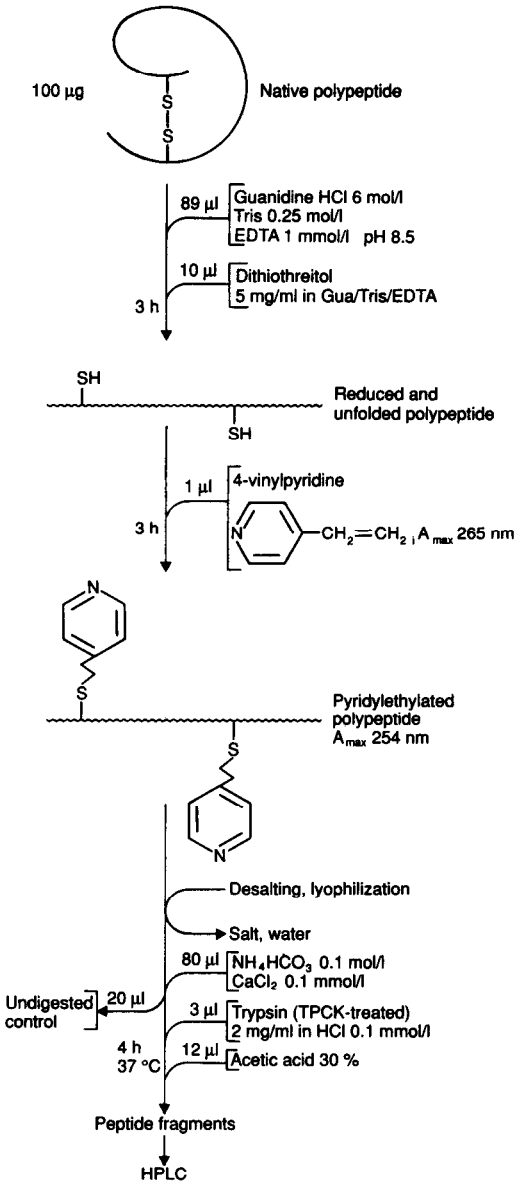


Fig. 2. Protocol used for reduction, derivatization and tryptic digestion of the polypeptides.

chromatography (CM-Sepharose® CL-6B; Pharmacia LKB Biotechnology, Uppsala, Sweden) and dried using a vacuum centrifuge.

*Reduction, derivatization and tryptic digestion*

The polypeptides (4 nmol; 40 μmol/l, containing cysteine residues corresponding to 80 or 160 μmol/l) were reduced with DTT (3 mmol/l) in Tris (0.25 mol/l, pH 8.5) containing guanidine hydrochloride (6 mol/l) and EDTA (0.25 mol/l) and the SH groups were derivatized using 4-vinylpyridine (94 mmol/l) as shown in Fig. 2.

*Desalting.* NAP™-10 columns (Pharmacia LKB Biotechnology) were equilibrated with TEA (20 mmol/l, adjusted to pH 8.35 with acetic acid). Each polypeptide sample was diluted to 1.0 ml with the TEA solution and applied to the column. The polypeptide was eluted with 1.5 ml of TEA solution and dried in a vacuum centrifuge.

*Tryptic digestion.* Trypsin was dissolved in hydrochloric acid (0.1 mmol/l) to an enzyme concentration of 2 μg/μl. The incubation buffer used was ammonium hydrogencarbonate (100 mmol/l) containing calcium chloride (0.1 mmol/l).

The polypeptide (100 μg, 4 nmol) was dissolved in 80 μl of incubation buffer. An aliquot (20 μl; 25 μg) was removed to serve as an undigested control, 3 μl (6 μg) of trypsin solution were added to the remaining 60 μl (75 μg) of peptide sample and 1 μl of incubation buffer was added to the undigested control. Both mixtures were incubated for 4 h at 37°C and stirred every 30 min. The digestion was terminated by adding 12 μl of 30% acetic acid to the digested sample and 4 μl to the undigested control to give a final peptide concentration of 1 μg/μl for both samples.

*Equipment*

All the instruments and software for the chromatographic procedures were from Pharmacia LKB Biotechnology. The equipment included HPLCmanager™ software for instrument and gradient control, a Solvent Conditioner 2156, two HPLC pumps 2248, High Pressure Mixer for solvent delivery and gradient formation and an Autosampler 2157 for sample injection. The column used was a SuperPac™ Pep-S $C_2/C_{18}$ (5 μm, 100 Å) (250 × 4 mm I.D.) equipped with a precolumn (10 × 4 mm I.D.) packed with the Pep-S material. The column temperature was maintained at 38°C by means of Column Oven 2155.

All measurements were monitored at 215, 254 and 280 nm using Variable Wavelength Monitor 2141. Data were converted to Nelson chromatography software for evaluation.

*Peptide mapping: chromatographic procedures*

The solvents, balanced to the same absorbance, were (A) 0.15% TFA in water and (B) 0.144% TFA in acetonitrile–water (60:40). The samples (2 nmol) were eluted at a flow-rate of 1 ml/min in a gradient composed of four linear sections as follows: 0% B for 2 min, 0–25% B in 30 min, 25–35% B in 58 min, 35–60% B in 46 min and 60–100% B in 10 min.

*Amino acid sequence analysis*

Amino acid sequences were determined with an Applied Biosystems (Foster City, CA, U.S.A.) Model 470A protein/peptide sequencer equipped with an on-line detection system, Applied Biosystems Model 120A phenylthiohydantoin (PTH)

analyser. The apparatus was operated according to the manufacturer and the sequencer performed at a *ca.* 90–92% repetitive yield. The initial yield was not determined. PTH standards, including pyridylethyl cystine-PTH, were purchased from Applied Biosystems.

## RESULTS AND DISCUSSION

### *Comparison of different batches of p24–gp41*

The tryptic maps of three batches of p24–gp41 are shown in Fig. 3. The two upper chromatograms (batches 17 and 23) match well, but the lower trace (batch 34) exhibits conspicious differences from the former chromatograms. Batch 34 exhibits three peaks between 120 and 130 min in proportions which are different from those of the other batches. This may be due to variability in digestion.

### *Tentative identification of fragments*

The general structures of p24 core and p24–gp41 are shown schematically in Fig. 1. The boxes indicate the fragments produced by tryptic cleavage. The position of tryptophan (W), tyrosine (Y), phenylalanine (F) and cysteine residues (C) are indicated with one-letter symbols.

The tryptic maps of p24–gp41 monitored at three different wavelengths are shown in Fig. 4. At 215 nm we detect peptide fragments in general, at 254 nm fragments containing pyridylethylated cysteine residues and at 280 nm fragments containing aromatic amino acid residues. Ratios of the peak areas (PA) taken from the different traces are given in Table I. A PA280/PA215 > 0 shows that the fragment contains

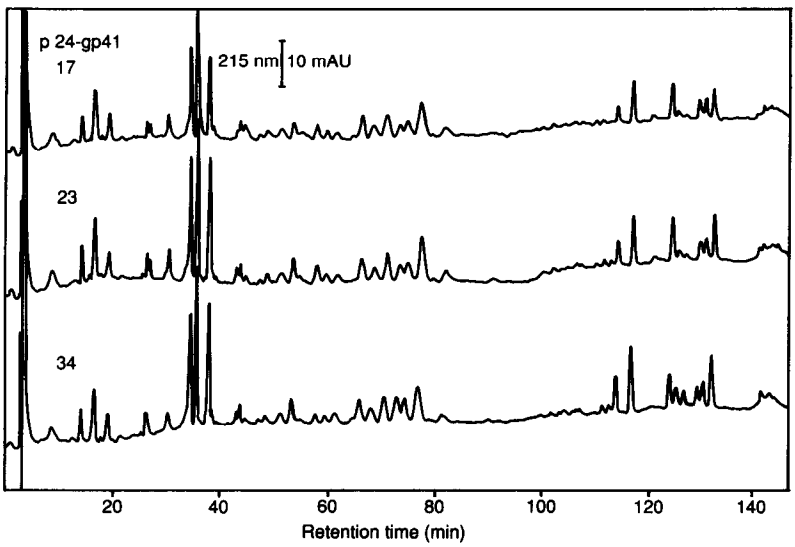


Fig. 3. Tryptic maps of three batches (Nos. 17, 23 and 34) of p24–gp41. The samples (1 nmol) were injected into a SuperPac Pep-S, $C_2/C_{18}$ (5 μm) column (250 × 4 mm I.D.) and eluted in a gradient running from 100% A (0.15% TFA) to 100% B [0.144% TFA in acetonitrile–water (60:40)] in 146 min. The gradient rates of the different segments of the gradient were 0.50, 0.10, 0.33 and 2.4% min acetonitrile, respectively.

TABLE I

HIV POLYPEPTIDES p24–gp41 AND p24 CORE WERE DIGESTED WITH TRYPSIN AND SUBJECTED TO PEPTIDE MAPPING, THE PEAK AREAS (PA) MONITORED AT 215, 254 AND 280 nm WERE CALCULATED AND THE PA280/PA215, PA254/PA215, PA254/PA280 RATIOS WERE CALCULATED FOR THE FRAGMENTS ELUTED AT THE DIFFERENT RETENTION TIMES ($V_t$)

Ratios indicated by > are calculated from peaks which were off-scale in the 254-nm trace. Fragments that differed in the two peptide maps were identified by amino acid sequence analysis (amino acid residues in italics).

| *$V_t$ (min)* | *Polypeptide* | | | | | | *Fragments identified by amino acid sequence analysis* | |
|---|---|---|---|---|---|---|---|---|
| | *p24–gp41* | | | *p24 core* | | | *Residue numbers* | *Amino acid sequence* |
| | *PA280/ PA215* | *PA254/ PA215* | *PA254/ PA280* | *PA280/ PA215* | *PA254/ PA215* | *PA254/ PA280* | | |
| 14.0 | 0 | 0 | – | 0 | 0 | – | | |
| 16.6 | 0.65 | 0.42 | 0.65 | 0.63 | 0.51 | 0.80 | | |
| 19.7 | 0 | 0 | – | 0 | 0 | – | | |
| 26.2 | 0 | 0 | – | 0 | 0 | – | | |
| 30.2 | 1.0 | **2.0** | **2.0** | 0 | 0 | – | | |
| 34.3 | 0 | 0 | – | 0 | 0 | – | | |
| 35.4 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| 36.0 | 0.28 | 0 | 0 | 0 | 0 | – | | |
| 37.6 | 1.1 | 1.1 | 1.0 | 0.92 | 0.91 | 1.0 | | |
| **43.4** | 0.79 | **2.0** | **2.6** | ← No fragment → | | | 239–253 | *LICTTAVPW*NGPGHK |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 44.3 | 0.55 | 0.65 | 1.2 | – | – | – | | |
| 50.3 | 0.20 | 0 | 0 | 0.30 | 0 | – | | |
| 53.0 | 0 | 0 | – | 0 | 0 | – | | |
| 57.5 | 0 | 0 | – | 0 | – | – | | |
| 59.0 | 0 | 0 | – | 0.18 | – | – | | |
| 67.6 | 0.64 | 0.66 | 1.0 | 0.49 | – | | – | |
| 70.0 | 0.95 | 1.0 | 1.1 | 2.1 | >0.34 | – | | |
| 72.5 | 0.94 | 1.1 | 1.2 | 0.93 | >0.20 | >0.21 | | |
| **76.1** | 0.80 | **2.1** | **2.6** | 0.84 | >1.4 | **>1.7** | 184–200 | *NWMTETLLVQNANPDCK* |
| **108** | ← No fragment → | | | 0 | >0.67 | **>2** | 205–228 | *ALGPAATLEEMMTACQGVGGPGHK* |
| 112 | 0.70 | 0.18 | 0.26 | 0.69 | – | – | | |
| 114 | 0.61 | 0 | – | 0.65 | 0 | – | | |
| 116 | 0.62 | 0.61 | 0.98 | 0.56 | >0.05 | >0.09 | | |
| **118** | ← No fragment → | | | 0.45 | >0.10 | >0.22 | 184–200<br>205–228 | *NWMTETLLVQNANP*DCK/<br>*ALGPAATLEEMMTACQGVGGPGHK* |
| **124** | 0.57 | **2.2** | **3.9** | ← No fragment → | | | 205–238 | *ALGPAATLEEMMTACQGVG*GPDQQ LLGIWGCSGK |
| 125 | 0.71 | **2.5** | **3.5** | – | – | – | | |
| 129 | 0 | 0 | – | 0 | 0 | – | | |
| 130 | 0 | 0 | – | 0 | 0 | – | | |
| 132 | 0 | 0 | – | 0 | 0 | – | | |

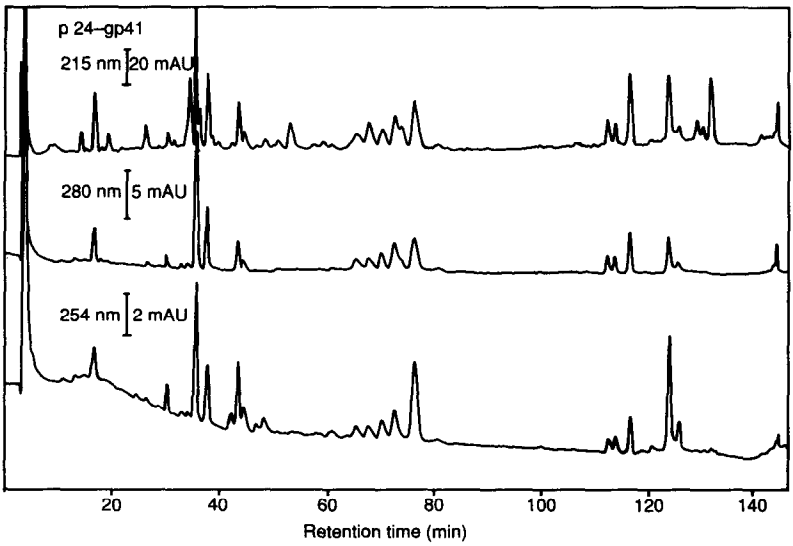


Fig. 4. Tryptic maps of p24–gp41 (2 nmol) monitored at 215, 254 and 280 nm. For chromatographic conditions, see text and Fig. 3.

aromatic amino acid residue(s), while PA254/PA215 or PA254/PA280 > 1.5 indicate that the fragment has a pyridylethylated cysteine content.

In p24–gp41 there should be eleven fragments having no aromatic amino acid residue (Fig. 1). There are ten peaks in the 215-nm trace which have no corresponding peaks in the 280-nm trace and which therefore have PA280/PA215 = 0 (Table I). These peaks should correspond to non-aromatic fragments and the others should represent fragments containing aromatic residues.

In the tryptic digest of p24–gp41 there should be three cysteine-containing fragments, one containing two cysteine residues and two containing one cysteine residue. By comparing the PA280/PA215 and PA254/PA280 ratios in Table I with the corresponding peaks in Fig. 4 we can identify these.

The digest of p24–gp41 gives five peaks having PA254/PA215 or PA254/PA280 ratios > 1.5 (Table I). Three of these peaks are dominant among the five (Fig. 4) (43.4, 76.1 and 124 min) and therefore most likely represent the pyridylethylated cysteine-containing fragments of the p24–gp41 digest. The PA254/PA280 ratio of the peak at 124 min is significantly higher (50%) than that of the others. (Table I). This peak could therefore be predicted to represent the fragment containing two cysteine residues, *i.e.*, the fragment represented by residues 250–238. The fragments eluted at 43.4 and 76.1 min should thus contain one cysteine residue each, *i.e.*, should be the fragments represented by residues 184–200 and 239–253. Without further information we cannot tell which of these peaks represents which fragment. However, it is possible to make a prediction in this respect by comparing the digests of p24–gp41 and p24 core.

*Comparisons of p24–gp41 and p24 core*

The digest of p24 core should produce two pyridylethylated cysteine-containing fragments. One should be common to the digests of both p24 core and p24–gp41

(residues 184–200) and the other should be unique to p24 core (residues 205–228) (Fig. 1). It has already been concluded that the peak at 76.1 min should represent a fragment containing one cysteine residue. Since that peak is common to both maps (Fig. 5), it should represent the fragment containing residues 184–200.

The peak at 43.4 min of the p24–gp41 peptide map should therefore represent the second fragment of the p24–gp41 digest containing only one cysteine residue, *i.e.*, the fragment composed by residues 239–253.

The fragments discussed above, *i.e.*, those eluted at 43.4, 76.1 and 124 min, were identified by amino acid sequence analysis. The predictions of their cysteine contents were thereby confirmed (Table I).

Two further fragments of the p24–gp41 digest (at 30.2 and 125 min) had PA254/PA215 and PA254/PA280 ratios > 1.5 (Table I). The PA254/PA280 ratio of the former indicates a content of one cysteine residue of the fragment, whereas the PA254/PA280 ratio of the latter indicates that the fragment should contain two cysteine residues. The fragments were considered to be incompletely reduced or incompletely digested polypeptide material and were therefore not subjected to sequence analysis.

*Sequencing of non-identical parts of p24 core and p24–gp41*

The digest of p24–gp41 should contain two fragments (residues 205–238 and 239–253) which cannot be produced by p24 core. Candidates are the fragments eluted at 43.4 and 124 min, since they have no corresponding fragments in the digest of p24 core (bottom map in Fig. 5). Their identities were discussed above.

The digest of p24 core should produce one unique 24-residue fragment (residues 205–228) with no corresponding fragment in the digest of p24–gp41. However, the

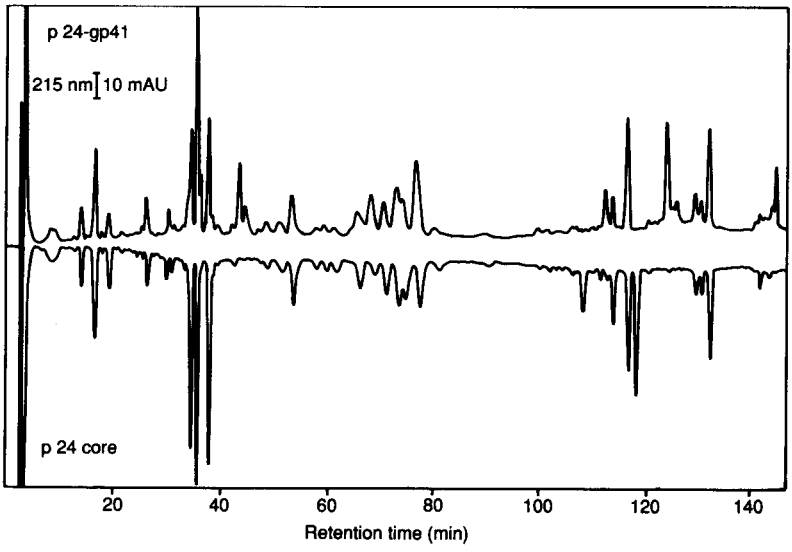


Fig. 5. Tryptic maps of p24–gp41 (2 nmol) and p24 core inverted (2 nmol). For chromatographic conditions, see text and Fig. 3.

tryptic map of p24 core contains two peaks having no corresponding peaks in the map of p24–gp41, one at 108 and the other at 118 min.

Comparisons of the different absorbances at 254 and 280 nm of the fragments eluted with these peaks makes identification possible. The fragment eluted at 118 min absorbs almost equally at both wavelengths (chromatograms not shown) and should therefore contain no pyridylethylated cysteine residue but presumably one (or several) tryptophan residues. This precludes its identity with the fragment composed of residues 205–228, which contains no aromatic amino acid residue but a cysteine residue. However, the fragment eluted at 108 min has a PA280/PA215 ratio = 0 but a PA254/PA280 ratio > 2 and should therefore have no aromatic residue but a pyridylethylated cysteine residue. This fragment is therefore a perfect candidate for the unique 24-residue fragment of the p24 digest (Fig. 1).

Amino acid sequence analysis confirmed these predictions (Table I). The fragment eluted at 118 min turned out to be a double sequence composed of residues 184–200 and 205–228, presumably connected by an unreduced disulphide bridge. Thus, as predicted, the fragment contained no pyridyl group although it contained two cysteine residues and, also as predicted, one tryptophan residue.

## CONCLUSIONS

The three different batches of p24–gp41 gave rise to equivalent peptide maps indicating identical contents of the batches. To prove this sequence analysis of every single fragment produced in the different digests, amino acid analysis or fast atom bombardment mass spectrometry would be required, which is impractical for quality control purposes.

When comparing the structures of different polypeptides it is of great help to be able to identify fragments containing aromatic amino acid residues or cysteine residues. The usefulness in this respect of comparing the absorbances of a peptide at 215 and 280 nm is well known, primarily for the identification of fragments containing tryptophan and to some extent fragments containing tyrosine and phenylalanine. Less known is the ease by which cysteine residues can be detected after reduction and pyridylethylation. We have clearly demonstrated that both the occurrence and amounts of cysteine residues in separate fragments can be predicted by comparing their absorbances at 280 and 254 nm. To us, this way of identifying such fragments seems both economical and efficient.

## REFERENCES

1 L. Ratner, W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo and F. Wong-Staal, *Nature (London)*, 313 (1985) 277.
2 R. M. Thorn, G. A. Beltz, C.-H. Hung, B. F. Fallis, S. Winkle, K.-L. Cheng and D. J. Marciani, *J. Clin. Microbiol.*, 25 (1987) 1207.
3 R. C. Gallo, *Sci. Am.*, 256 (1987) 39.
4 V. M. Ingram, *Methods Enzymol.*, 6 (1963) 831.
5 J. C. Bennett, *Methods Enzymol.*, 11 (1967) 330.
6 H. V. J. Kolbe, F. Jaeger, P. Lepage, C. Roitsch, G. Lacaud, M.-P. Kieny, J. Sabatie, S. W. Brown, J.-P. Lecocq and M. Girard, *J. Chromatogr.*, 476 (1989) 99.

7 R. C. Chloupek, R. J. Harris, C. K. Leonard, R. G. Keck, B. A. Keyt, M. W. Spellman, A. J. S. Jones and W. S. Hancock, *J. Chromatogr.*, 463 (1989) 375.
8 R. L. Garnick, M. J. Ross and C. P. du Mée, in J. Swarbrick and J. C. Boylan (Editors), *Encyclopedia of Pharmaceutical Technology, Absorption of Drugs to Biovailability of Drugs and Bioequivalence*, Vol. 1, Marcel Dekker, New York and Basel, 1988, p. 253.
9 R. L. Garnick, N. J. Solli and P. A. Papa, *Anal. Chem.*, 60 (1988) 2546.
10 J. F. Cavins, and M. Friedman, *Anal. Biochem.*, 35 (1970) 489.
11 M. Friedman, L. H. Krull and J. F. Cavins, *J. Biol. Chem.*, 245 (1970) 3868.