Journal of Peptide Science, Vol. 3, 168–180 (1997)

The Structural and Aggregation Properties of the Synthetic C-terminal Half (104-mer) Polypeptide from HIV p24gag Resemble Those of Full-length Protein

HAYDN L. BALL[1,2], A. W. EDITH CHAN[1], WILLIAM A. GIBBONS[3], ANTHONY R. M. COATES[4], and PAOLO MASCAGNI[1]

[1]Italfarmaco Research Centre, Cinisello Balsamo 20092, Milan, Italy
[2]Department of Neurology, University of California, San Francisco, CA 94143-0518, USA
[3]Department of Pharmaceutical Chemistry, School of Pharmacy, University of London, London, WC1N 1AX UK
[4]Department of Medical Microbiology, St George's Hospital Medical School, Cranmer Terrace, London SW17 ORE, UK

Received 22 January 1996

Accepted 9 July 1996

Abstract: The aggregation and structural properties of the synthetic C-terminal half [Ala^{330}, Ala^{350})270–373; 104-mer)] polypeptide from HIV-1 p24gag were studied. In concentrated solutions the synthetic polypeptide aggregated to tetramers which, upon dilution, gave a mixture of monomeric and dimeric species. These results correlated well with the *in vitro* aggregation properties of recombinant p24. The tetrameric form of the synthetic polypeptide had a p*I* which differed by about four units from that of the mixture of monomeric and dimeric species. CD studies indicated that the latter contained, in aqueous solutions, a compact molecule lacking, however, a defined tertiary structure. Addition of MeOH to aqueous solutions of both tetramer and monomer/dimer mixture induced a more defined structure, which was assigned to that of an $\alpha+\beta$ protein in agreement with secondary structure predictions. A model of the dimeric form of the 104-mer, which takes into account the results presented here and those from a study on the specificity of a set of anti-104-mer MoAbs, is presented.

Finally, the results indicated that the structure of the 104-mer in its dimeric form is similar to that adopted by the same sequence when part of full-length p24. © 1997 European Peptide Society and John Wiley & Sons, Ltd.

J. Pep. Sci. 3: 168–180, 1997

(No. of Figures: 9. No. of Tables: 1. No. of Refs: 32.)

Keywords: p24; HIV; synthetic proteins; CD spectroscopy; secondary-structure prediction

Abbreviations: Boc, *tert*-butyloxycarbonyl; CD, circular dichroism; DTNB, 5,5'-dithiobis(2-nitrobenzoic acid); FPLC, fast protein liquid chromatography; HBTU, 2(1H-benzotriazol-l-yl)-1,1,3,3,-tetramethyl uronium; HIV-l, human immunodeficiency virus; HOBT, 1-hydroxybenzotriazole; IEC, ion-exchange chromatography; IEF, isoelectric focusing; MoAb, monoclonal antibody; MOPS, 3-(*N*-morpholino)-propanesulphonic acid; PAM, phenylacetamidomethyl; p*I*, isoelectric point; p24gag, core protein from HIV from processing of the precursor p55; RP-HPLC, reversed-phase high-pressure liquid chromatography; SDS–PAGE, sodium dodecyl sulphate–polyacrylamide gel electrophoresis; TRA, trifluoroacetic acid; TFMSA, trifluoromethane sulphuric acid; 94-mer, fragment 270–363 from HIV-l p24gag protein; 104-mer, fragment [(Ala^{330}, Ala^{350})270–373] from HIV-l p24gag protein.

Address for correspondence: Dr P. Mascagni, Italfarmaco Research Centre, Via dei Lavoratori, 54, Cinisello Balsamo 230092 Milan, Italy.

CCC 1075-2617/97/030168-13

INTRODUCTION

HIV p24gag protein (hereafter referred to as p24) is believed to be self-associated in the HIV core [1] although it is not known yet whether self-association initiates with the precursor of p24 (p55) or whether it takes place after p55 processing. Assembly of p24 is, however, believed to be necessary for virion maturation. Recently, a possible homology between p24 and picornaviral VP2 coat protein has been suggested [2,3]. According to these proposals p24 has the eight-stranded antiparallel β-barrel typical of other viral capsid proteins. These predictions contrast with those which assigned to p24 a predominantly α-helix secondary structure [4,5] and the secondary

structure composition of recombinant protein calculated by CD spectroscopy [6]. In this latter study the CD spectrum of SIV p27 was assigned to an $\alpha+\beta$ structure containing 44% α-helix and 56% β-sheet.

The secondary structure of p24 calculated from CD spectroscopy is likely to be that of a dimeric form of the protein. Thus, self-association, which occurs *in vivo* during virus particle assembly, has been demonstrated *in vitro* by analytical affinity chromatography [7] by showing that, in solution, p24 dimerizes. Furthermore, at concentrations (80 mg/ml) close to the estimated concentration of p24 in the virion, self-association seems to proceed predominantly to dimer with, however, some populations of trimers or tetramers ([7], and references therein).

The chemical synthesis of the C-terminal half-sequence ([Ala^{330}, Ala^{350}]270–373), hereafter referred to as 104-mer, of the HIV p24gag protein and its immunological properties have been reported [8–11]. The synthetic protein was recognized by antibodies in sera of HIV-infected haemophiliacs [11]. Furthermore, a set of seven MoAbs raised against the 104-mer were found to crossreact with viral gag proteins (p55 and/or p24), suggesting that some of the structural features of p24 are maintained in the synthetic protein [9].

In this work the conformational properties of synthetic 104-mer and its ability (if any) to self-associate were studied. Naturally, during the course of the study the question arose as to whether molecules such as the 104-mer can fold into structures contained in the parent, full-length protein. Thus, in the discussion that follows attempts will be made to correlate the structural findings for the synthetic 104-mer with the properties of p24.

MATERIALS AND METHODS

Chemical Synthesis of 104-mer (Synthesis 1)

The sequence of the polypeptide corresponds to that of p24 of HIV-1LAV [12]. During this study findings were published [13] which indicate that the cleavage of p24 from its p55 precursor takes place between residues 363 and 364 ten residues downstream from the initial hypothesis that had guided the synthesis of the 104-mer. The two Cys residues found in the p24 at positions 330 and 350 were replaced by Ala. The peptide was synthesized on the acid-stable PAM resin using an Applied Biosystems (div. Perkin-Elmer, Foster City, CA) 430A synthesizer and improved BOC chemical procedures ([14] and references therein). Details of the synthetic procedures, cleavage and purification of the 104-mer have been described elsewhere [8].

The following treatments were performed on the 104-mer (synthesis 1). After cleavage from the resin the crude polypeptide (approx. 100 mg) was dissolved in 10 ml 6 M guanidine HCl, 50 mM TRIS acetate and 20% 2-mercaptoethanol at pH 8.5 and stirred at room temperature for 2 h to remove residual His di-nitrophenyl protecting groups. The solution was then dialysed (SpectraPor 6, MW cut-off 6000, Spectra Medical Industrial Inc., Los Angeles), against 150 mM ammonium acetate, pH 8.3, for two days at 4 °C to remove smaller molecular weight impurities. The content of the dialysis tubing was then lyophilized.

Chemical Synthesis of 104-mer (Synthesis 2)

An additional synthesis of 104-mer was carried out using more refined synthetic and purification protocols, aimed at preventing the formation of deletion sequences [14–18]. Briefly, single HBTU/HOBt-mediated couplings were used throughout, except Asn, Gln and Arg residues which were double coupled. A capping procedure using *N*-(2-chlorobenzyloxycarbonyloxy) succinimide was performed after the addition of each amino acid to terminate any deletion sequences [17]. The protein was synthesized on PAM resin pre-derivatized with the first amino acid. At the end of the synthesis the N^{α}-terminal protecting group was removed and a reversible lipophilic label introduced as the succinimidyl carbonate [16]. Cleavage of the labelled 104-mer from the resin and removal of side-chain protecting groups was achieved using a low TFMSA/high HF protocol [19] in the presence of scavengers (*p*-cresol, dimethyl sulphide, TFA, TFMSA (4 : 15 : 25: 5) at 0 °C for 2 h and *p*-cresol, *p*-thiocresol, HF (4 : 1 : 50) at 0 °C for 1 h, respectively). HF was removed under vacuum and the cleaved protein precipitated with dry diethyl ether. The solid was filtered, redissolved in 20% acetic acid solution and lyophilized.

Purification of 104-mer (Synthesis 1; Method A)

Dialysed 104-mer was subjected to semi-preparative RP-HPLC (C_4 Vydac, 250 × 9 mm) using a linear gradient 0% to 100% CH_3CN in 360 min. A broad peak eluting between 70 and 80% CH_3CN was collected and the solution lyophilized. The product thus obtained was dissolved in 5 ml of 20 mM TRIS–HCl, 2 mM EDTA, 0.1 M Kcl, pH 5.2 (buffer A) and

loaded onto a preparative size-exclusion chromatography column (TSK GW 2000) equilibrated with the same buffer. The main fraction, which eluted between 10.5 and 12 min, using a flow rate of 2 ml/min, was collected and lyophilized. A 10 mg/ml solution of 104-mer was then further purified by cationic-exchange FPLC (Mono S, Pharmacia Uppsala, Sweden) using buffer A, as for size-exclusion chromatography, and buffer B (buffer A + 1 M KCl) with a gradient of 0 to 100% B in 1.5 h.

Amino acid analysis, SDS–PAGE (20% homogeneous gel, Pharmacia, Uppsala, Sweden) and isoelectric focusing gel electrophoresis (Ampholine Pagplate, pH 3.5–9.5) were used to establish the homogeneity of the purified product. The sequence of the N-terminal 40 residues and the C-terminal 22 residues of the peptide were determined using an ABI 477A amino acid sequencer [8]. Tryptic digestion and FAB mass spectrometry (VGZAB-SE) peptide mapping were used to identify the remainder of the sequence [8].

Purification of 104-mer (Synthesis 1; Method B)

Thirty mg of crude HF cleavage product was dialysed as described above, then dissolved in 40 ml of 6 mM NaCl solution. The pH was adjusted to 7.3 with 0.1 N NaOH and the solution left to stand at 4 °C overnight. One ml of 40% ampholyte (p*I* 3–10) was added to the solution and then transferred to the Rotofor cell. Focusing was allowed to take place for 5.1 h at 4 °C, after which time the contents of each of the 20 cells were harvested. The pH of each cell was recorded and a fraction analysed by analytical RP-HPLC.

The material focusing in the Rotofor cell at pH 5.5 was loaded onto anionic exchange media (TSK-DEAE-5PW, 7.5 × 75 mm) that had been equilibrated with buffer A (10 mM TRIS, pH 7.8). A gradient of 0 to 50% B (buffer A + 0.5 M NaCl) in 30 min was applied to elute the bound material. Anionic exchange FPLC was also performed both at reduced temperature (i.e. 4 °C), by placing the buffer solutions in iced water, and with the addition of 20% MeOH to the solvents. The identity of the product thus obtained was confirmed by amino acid analysis.

Purification of 104-mer (Synthesis 2)

Crude labelled 104-mer was purified by semi-preparative RP-HPLC (C_4 Vydac, 9 × 250 mm; linear gradient 0% to 100% CH_3CN in 360 min; flow rate 2.5 ml min) which permitted the Fmoc-derivatized material to be separated from underivatized impurities [16]. The former was collected and treated for 30 min with 5% triethylamine to remove the Fmoc derivative from the N-terminus [16]. The solution was then reduced in volume, and further purified by RP-HPLC using the same conditions described above.

The purified material was then subjected to ESI-MS (expected 11,205.3, found 11.205).

Secondary Structure Prediction

The sequences corresponding to the 270–363 region of p24 were extracted from the SWISS-PROT database (December 1992 release). They were aligned using a multiple sequence alignment algorithm scoring on sequence homology only [20–22]. The alignment was then divided into shorter segments where the secondary structure elements were considered separately. The decision on where the breaking point of a secondary structure element takes place was based upon the identification within the alignment of gaps and/or insertions which generally identify loop regions. Although the latter often have defined secondary structures, in the following discussion they were assigned to coil, where coil is either a fully unordered structure or an ordered structure which, however, does not contain repeating conformations [23].

This assignment of secondary structure to each segment was based on three different methods, Chou and Fasman [24], Garnier *et al.* [25] and Momany *et al.* [26]. The helical wheel was used to determine an amphiphilic helix. Furthermore, assignments were also made based on sequence homologies with fragments found in proteins with known 3D structure. When two or more of these prediction methods were in agreement, then the structure was used. The secondary structures thus assigned were modelled using the Quanta program (Molecular Simulations, CA, USA), and refined by at least 1000 steps of steepest descendent minimization to visulize the relative orientation of amino acid side chains.

CD Spectroscopy

Secondary structure analysis of the 104-mer was performed on a Jasco J-600 CD spectrometer. The wavelength range examined was 185–250 nm, using a scan speed 50 nm/min and a sensitivity of 10 mdeg. The cell used had a pathlength of 0.1 cm

and a volume of 300 μl. The methanol into buffer titration experiments involved taking CD spectra in 5%, 10%, 20%, 30%, 40% and 50% MeOH. All spectra were baseline subtracted. The observed ellipticity was converted to mean residue weight ellipticity ($[\theta]$/degrees cm^2 sol/dmol). In the case of tretameric p24, its spectrum was measured on a solution directly obtained from cationic-exchange purification. The concentration of polypeptide was calculated by amino acid analysis at the end of the CD experiment and after a desalting procedure. The value thus obtained is therefore approximate (about 0.2 mg/ml), owing to the number of manipulations.

The secondary structure of 104-mer was determined by dissolving 1.82 mg (0.16 μmol) of lyophilized material with p*I* 5.6 in 10 ml of water/MeOH (1:1). 0.5 ml of the 104-mer solution was then placed in the CD cell (pathlength 0.1 cm) and the optical ellipticity recorded at intervals of 0.4 and 0.6 nm. Deconvolution of spectra using the variable selection method [27,28] was then performed to determine the secondary structure composition of the 104-mer.

RESULTS

The first step in this study was to purify synthetic 104-mer using the same protocol as described earlier [8]. The salient points of this purification scheme which bear on the structure of the synthetic polypeptide are discussed in the following section.

Purification of 104-mer (Synthesis 1; Method A)

After an initial fractionation by RP-HPLC and size-exclusion chromatography, the polypeptide was further purified by preparative cationic-exchange chromatography. The use of the latter was suggested by (i) isoelectric focusing gel electrophoresis of crude 104-mer which revealed two main components focusing at approximately 9.3 and 5.6 respectively (lane 1, Figure 1) and (ii) the theoretical p*I* of 104-mer which was calculated to be 8.85.

Thus, a solution containing approximately 10 mg/ml of lyophilized material was loaded onto the column and the material which eluted at 0.13 M salt, collected (Figure 2A). SDS–PAGE (Figure 3A) indicated that the latter had the expected molecular weight and was a single component when reinjected on analytical ion-exchange media (Figure 2B). Further confirmation of its homogeneity was obtained by isoelectric focusing electrophoresis which

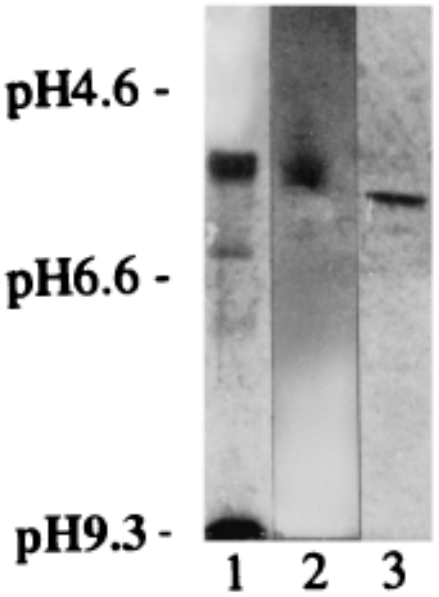

Figure 1 Isoelectric focusing of p24 gag proteins. The analysis was carried out on a PhastSystem using homogeneous polyacrylamide gels with a pH range 3–9: lane 1. crude synthetic 104-mer (synthesis 1); lane 2, purified synthetic 104-mer (synthesis 1; M+D form); lane 3, recombinant full-length p24.

indicated that the polypeptide focused as a single band with a p*I* of about 9.3 (Figure 3B).

Although the material obtained from this purification was a single component by ion-exchange and electrophoresis analysis, during subsequent experiments it was noticed that its chromatographic profile would slowly change as a function of time. In particular, when a solution containing purified 104-mer and stored at −20 °C for three days was subjected to ion-exchange chromatography a new component eluting at the beginning of the salt gradient was observed (Figure 2C). Furthermore, the concentration of this new component increased upon dilution of the solution stored at −20 °C, indicating an equilibrium between either two differently folded proteins or different aggregates of the same species (Figure 2D). Repetition of these dilution experiments using size-exclusion chromatography confirmed the second of the two hypotheses, since the two ion-exchange components eluted as proteins whose apparent molecular weight of 53 and 14 kDa and were therefore consistent with tetrameric and monomeric species (Figure 4). Interestingly, when the latter was isolated and reinjected on the size-exclusion column, a third minor component corresponding to a dimeric form of the 104-mer, was found (labelled 3, Figure 4). Collectively, these

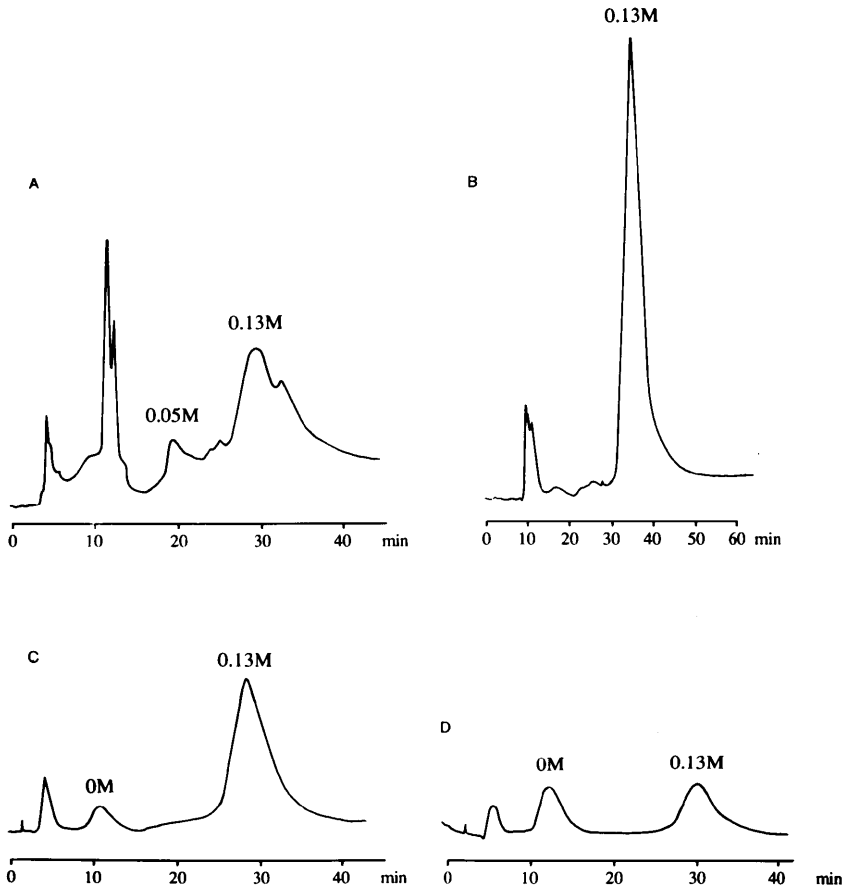

Figure 2 Cationic-exchange chromatography of 104-mer (synthesis 1). Panel A: the peak eluting at a salt concentration 0.13 M and corresponding to the tetrameric form of the polypeptide was collected. Panel B: purified 0.13 M peak, reinjected on the same column. Panel C: as in A after storage at −20°C for three days. Panel D: 1:1 dilution of the protein solution increased the concentration of the early eluting material (cf. C).

results indicated that (i) purification of synthetic 104-mer by the procedure described above had yielded a tetrameric species whose p*I* was 9.3 and (ii) disaggregation to monomeric and dimeric forms followed dilution of solutions containing tetramer. The p*I*s of the monomeric and dimeric species thus obtained were deduced by their chromatographic behaviour to be less than 7.8, where the latter was the pH of the buffer solution used for ion-exchange analysis. Furthermore, these observations raised the question as to whether the component focusing at approximately pH 5.6 in the IEF gel of crude 104-mer (lane 1, Figure 1) was indeed an impurity as initially concluded.

PURIFICATION OF 104-MER (SYNTHESIS 1; METHOD B)

To investigate the nature of the various forms of the peptide and to separate them according to their p*I*, a second purification protocol was used. This was based on the preparative IEF technique in free solution [29]. Accordingly, a 6 mM NaCl solution containing 30 mg crude polypeptide was transferred to the Rotofor cell [29], ampholytes added and then focused. At the end of the run 20 fractions were collected and an aliquot of each analysed by RP-HPLC. In Figure 5A, a diagram of fraction concentration vs. pH indicates that the synthetic material existed in solution in two main forms characterized by p*I*s of about 5.6 and 9.3 respectively. This result correlated well with those from the ion-exchange experiments described previously. Thus, the material with a p*I* of 9.3 was the tetrameric polypeptide (T), while the less intense peak at pH 5.6 was assigned to the mixture of monomeric and dimeric forms (M + D) detected above and whose p*I* had been estimated to be less than 7.8. To characterize this new component further, fraction 8 from the Rotofor purification

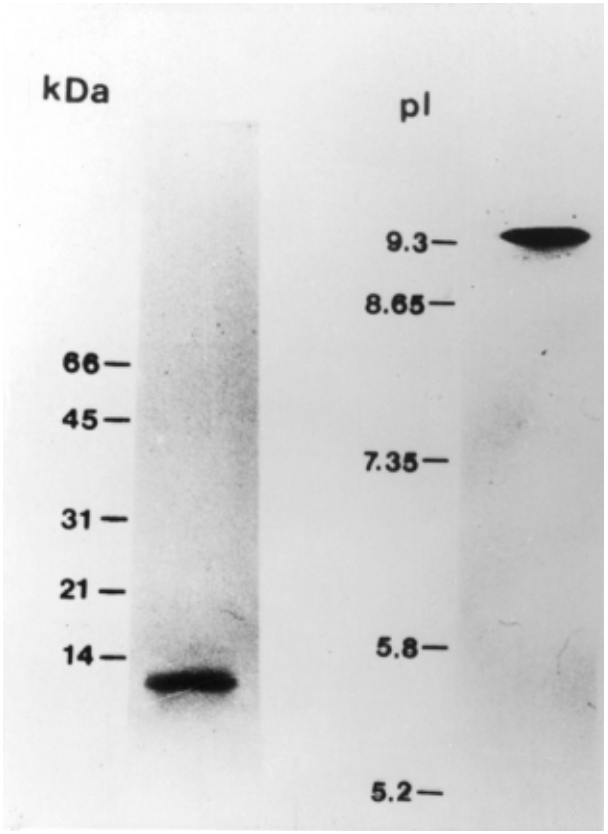

Figure 3 Gel electrophoresis of synthetic 104-mer (synthesis 1) after purification by a combination of RP-HPLC, size-exclusion and cationic-exchange chromatographic procedures. The analysis was carried out on a Phastsystem using a homogeneous (20%) polyacrylamide gel (SDS–PAGE) and a homogeneous gel with pH gradient 3–9: left lane, SDS-PAGE; right lane, IEF.

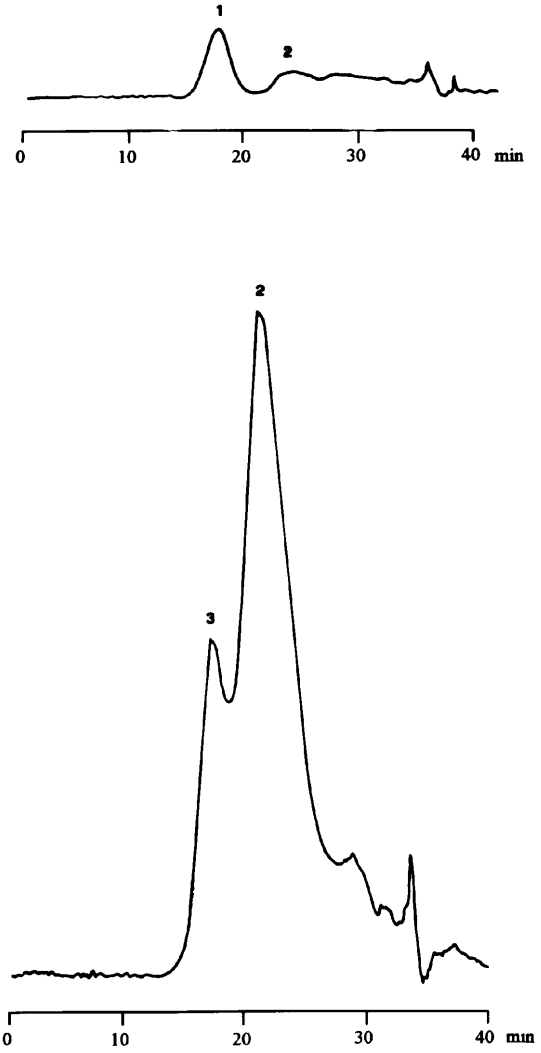

Figure 4 Size-exclusion chromatography of purified 104-mer. Top trace shows tetrameric (labelled 1) and monomeric (labelled 2) forms obtained from the isolation of the material retained on the cationic-exchange column. Bottom trace represent purified peak 2 (Panel A) reinjected on the same column; peak labelled 3 is the dimeric form of 104-mer. Samples were run on a Sephadex 75 column, using buffer containing 20 mM MOPS, 0.5 M NaCl pH 7.0 and a flow rate of 0.5 ml/min. Detection at 280 nm.

(See Figure 5) was subjected to anionic exchange chromatography (Figure 6A). The synthetic polypeptide eluted as a broad and jagged peak at a salt concentration of about 0.1 M. Although these results confirmed the existence of a negatively charged folded species, the complexity of the chromatographic profile was unexpected. Possible explanations to these observations were that fraction 8 contained (i) a mixture of proteins having closely related sequences or (ii) a mixture of proteins having the same sequence but closely related conformations.

Although impurities corresponding to closely related sequences are normally generated during the corse of chemical synthesis and their presence is expected in the crude material, it was felt that this was no longer the case after the protein had been purified using the highly selective IEF technique. Therefore, to prove the existence of multiple conformations of the same protein, ion-exchange chromatography was carried out under a variety of different conditions. These were designed so as to alter the conformational equiolibria that might be present in solution and included temperature variations and the use of MeOH as co-solvent for ion-exchange chromatography.

Decreasing the temperature to 4 °C had a significant effect on the shape of the chromatographic

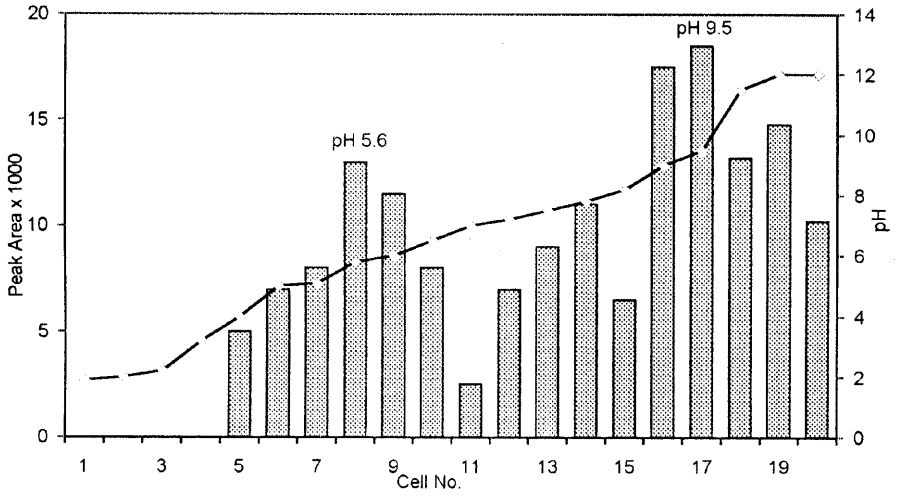

Figure 5 Purification of 104-mer by solution IEF on the Rotofor system. The pH values of the 20 cells collected (right *y*-axis and continuous line) were measured immediately after focusing. The protein content (left *y*-axis and bars) of each fraction was measured by RP-HPLC and expressed as peak area of the material eluting at about 45% CH_3CN on an analytical C4-RP column.

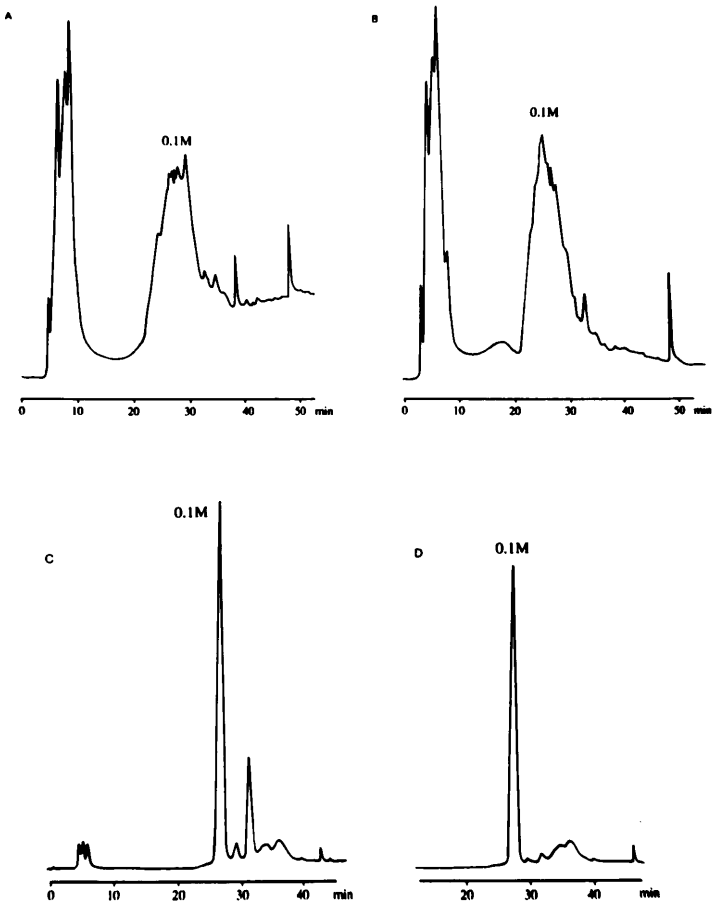

Figure 6 Anionic-exchange chromatography of 104-mer: (A) cell 8 from the Rotofor purification; (B) as A at 4 °C; (C) as A plus 20% MeOH in the elution buffers; (D) as C but after anion-exchange purification using 20% MeOH in the elution buffers.

peak by reducing its width (Figure 6B); however, the most dramatic changes were observed upon addition of MeOH. This induced the progressive narrowing of the chromatographic peak and its collapse into a single sharp line for a MeOH concentration of 20% (Figure 6C). This was isolated and shown to remain a single species when reinjected on the same chromatographic column (Figure 6D).

That the chromatographic results were to be ascribed to conformational changes were confirmed by CD spectroscopy. Thus, the spectrum of fraction 8 in 20% MeOH resembled that of a structure with an appreciable α-helix content which contrasted with a mainly 'random coil' conformation obtained in aqueous solution (see below for discussion).

These results, together with those described in the previous section, indicate that the synthetic polypeptide exists in solution in three different forms, i.e. T, D and M, characterized by p*I*s of 9.3 (T) and 5.6 (M + D mixture). This conclusion was confirmed by the results of a second independent synthesis which was carried out using more advanced chemical and purification protocols (see Materials and methods). Thus, purified material was shown to have the expected mass. When analysed by IEF the purified 104-mer was found to be composed of two species focusing at pHs of 9.3 and 5.6 (data not shown), as described for 104-mer from the first synthesis. These data eliminated the possibility that the latter was the result of an impurity.

CONFORMATIONAL ANALYSIS OF SYNTHETIC 104-MER

To characterize at a structural level the three different forms of 104-mer discussed above, their

Table 1 Alignment of p24 Sequences in the 270–363 Region

```
      270          280       290      300
hiv     ln   kivr m   ysptsi ld irqgpkepfrdyvdrfyktir
viv     lqwvitalrsvrhms  hrpgnp ml vkqkntesyedfiarlleaid
eiv     ms   egikvmi  gkp ka qn irqgakepypefydrllsqik
siv     l   pvk  gdpgas ltgvkqgpdepfadfvhrlittag
olv     l   rsvrhms  hrpgnp ml ikqknsesyedfiarlleaid
biv     lq   aame vaqakhatpgp in ihqgpkepytdfinrlvaale
biv     ll   vlq l       qpwst ivqgpaessvefvnrlqisl

         310       320       330          340
hiv     ae qasqev kn wmtetll  vq nanpd aktil kalgp  aatleemmta
viv     ae pvtdpi kt ylkvtls  yt nastd cqkqmd rtlgtrvqqatveekmqa
eiv     se ghpqei sk fltdtlt  iq nanee crnam rhlrp  edtleekmya
siv     rifgsaeag vd yvkq la  ye nanpa cqaai  rp pr kktdltgyirl
olv     te pvtdpi kt ylkvtls  ft nastd cqkqmd rvlgtrvqqasveekmqa
biv     gm aapett ke yllqhls  id haned cqsil  rplgp  ntpmekkiea
biv     ad nlpdgvlrnplltp lvmqmlt esvskfcrgeasgrg gak  tagl

      350    360
hiv     aqgvggpghkarvl
viv     crdvgsegfkmqll
eiv     crdigttkqkmmll
siv     csdig psyqqgl
olv     crdvgsegfkmqll
biv     crvvgsqkskmqfl
biv     rtigpprmkqpall
```

The seven p24 sequences were extracted from the SWISSPROT database (December 1992). Their codes are as follows; hiv, Human Immunodeficiency Virus; viv, Visna Lentivirus (sheep); eiv, Equine Infectious Anaemia Virus (Horse); siv, Simian Retrovirus (Monkey); olv, Olvine Lentivirus (Sheep); biv, Bovine Immunodeficiency Virus (Cattle); blv, Bovine Leukaemia Virus (Cattle); msv, Moloney Murine Sarcoma Virus (Mouse); rsv, Rous Sarcoma virus (Chicken).

Alignment of the sequences was done using a multiple sequence alignment algorithm scoring on sequence homology only [21–23]. Atom match was done using the Dayhoff matrix.

secondary structure composition was estimated using CD spectroscopy and compared to that obtained using secondary structure prediction algorithms.

Secondary Structure Prediction. Seven sequences corresponding to the 270–363 region of the p24 proteins of HIV and related viruses were extracted from the SWISS-PROT database and then aligned using a multiple sequence alignment algorithm scoring on sequence homology only [20–22]. The proteins in the database (Table 1) thus obtained were about 35% homologous to each other.

The analysis of the secondary structure was carried out by the combined use of three different prediction algorithms [24–26] and sequence homology with proteins of known three-dimensional structure. Assignments were only made when at least two of the above methods were consistent. The results of the prediction are summarized in Table 2 and some of the salient points discussed hereafter.

Table 2 Secondary Structure Prediction of the 270–373 Region of HIV gagp24

Sequence	C–F[a]	GOR[a]	Momany[a]	HW	Assignment
270–276	B	B	B	–	C
277–284	B	C	B	–	B
285–289	H	C	H	–	H
295–300	H	H	H	–	H
301–304	B	C	B	–	B
305–319	H	H	A	A	AH
320–323	B	B	B	–	B
324–329	C	C	C	–	C
330–335	H	B	H	A	AH
336–341	C	H	C	–	C
342–350	H	H	H	A	AH
351–357	C	C	C	–	C
358–363	B	B	B	–	B
364–373	–	–	–	–	C

[a]The sequence of the 104-mer after alignment with homologous regions of other p24 proteins (Table 1) was divided into shorter segments based upon the identification within the alignment of gaps and/or insertions. The prediction of secondary structure was carried out using three different methods: Chou–Fasman (C–F) [24]; GOR [25] and Momany *et al.* [26]. The helical wheel (HW) was used to predict whether a helical region was amphiphilic. An assignment was made when two or more of the three methods used were in agreement. In the case of the 270–276 sequence, although the predictions indicated it to be a β-strand, its structure was assigned to a coil since it is the N-terminal region of synthetic 104-mer.

B = β-strand, H = α-helix, C = coil, A = amphiphilic, AH = amphiphilic helix.

The sequence encompassing amino acids 277–284 was predicted to be a β-strand. Although the highly hydrophobic nature of this region (70% of the residues are hydrophobic) would suggest a location inside the structure of the polypeptide, two anti-104-mer MoAbs [9] directed against different parts of this region and crossreacting with intact p24 indicated that this strand was partially or completely exposed on the surface of the protein.

The 285–300 region is the most conserved region among the seven proteins compared (60% homology with three conserved residues). The 285–291 sequence (IRQGPKE) was predicted to be an amphiphilic helix located at the surface of the protein. This conclusion was in agreement with another experimentally identified anti 104-mer antibody binding site, the conformational epitope 284–289 [9].

The Pro at position 289 provided a slight change in the direction of the helix while the other Pro residue, that at position 292, induced a turn followed by another helix (295–300). Although Pro does occur in the central portion of helices, it is by far the least common residue in this position. Thus, a large loop comprising residues 289 to 294 could be an alternative assignment. A further conformational MoAb which recognized the sequence 288–293 [9] was consistent with this second hypothesis.

The 320–335 was also another conserved region among the seven gag proteins analysed. The first four residues, TLLV, were assigned to a β-strand while the QNANPD sequence (324–329) to an extended structure in agreement with the NMR data on the synthetic (NANP)*n* polypeptide model of *Plasmodium falciparum* circumsporozoite protein [30].

A further conserved region was identified in the amino acid stretch 342–350. Its structure was predicted to be that of a surface amphiphilic helix. Interestingly, a synthetic peptide corresponding to this region has been found to prevent the ability of the p55 gag polyprotein to form virus-like particles [31].

Finally, the 358–363 sequence was assigned to a β-strand. The identification of this region as the binding site for a further MoAb [9] assigned it to a surface strand.

The CD Structure of the M + D Mixture. A CD spectrum was obtained for the M + D mixture in aqueous solution (Figure 7), which had been purified by the Rototfor technique (the spectrum obtained for material from the second synthesis of 104-mer was essentially identical). The spectrum was character-

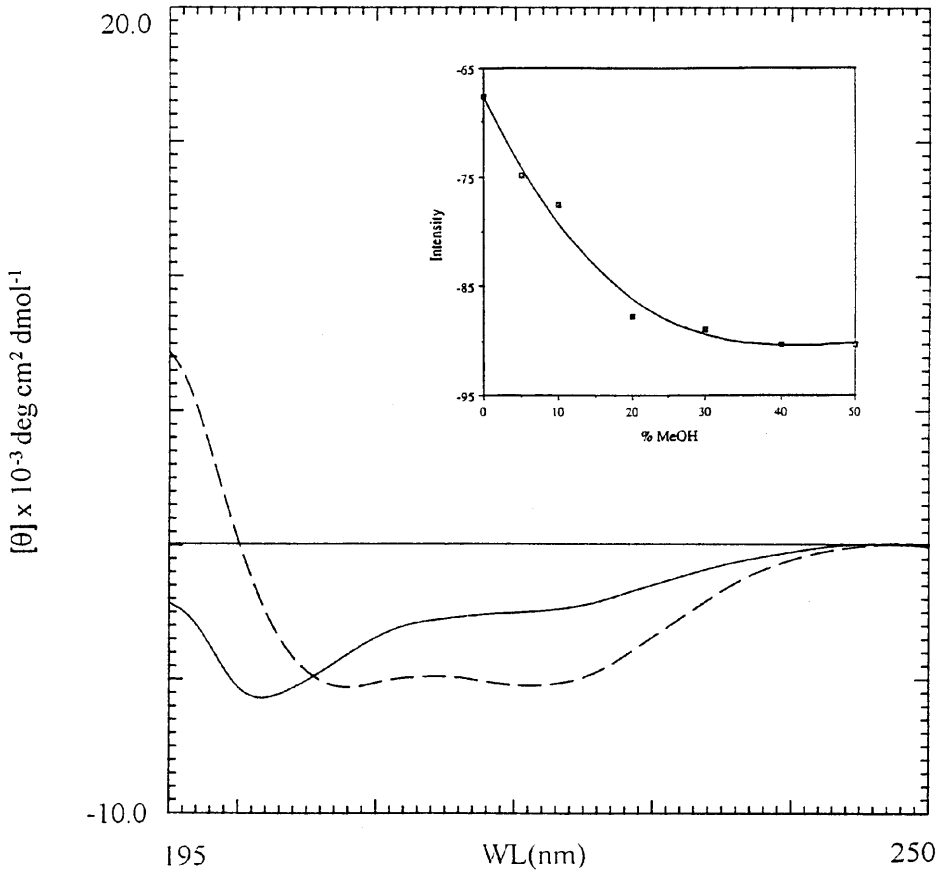

Figure 7 CD spectra of the M + D mixture of 104-mer in 20 mM MOPS, pH 7 (solid line), and in a 1:1 mixture of 20 mM MOPS (broken line), pH 7 and MeOH. Inset: molar ellipticity at 220 nm of the M + D mixture of 104-mer as a function of the percentage of MeOH added to the protein solution.

ized by a negative band at about 200 nm and a second, less intense, negative band at about 222 nm. Although a minimum at 200 nm has been traditionally attributed to the so-called random coil structure, the band at 222 nm, diagnostic for an α-helical contribution, together with the need to interpret the discrepancy between the calculated and found p*I* (8.85 vs. 5.6), seemed to rule out a truly random coil conformation for the M + D mixture.

This conclusion was confirmed by MeOH into buffer titration CD experiments. Thus, the plot of molar ellipticity at 220 nm vs. percentage of MeOH showed a plateau for MeOH concentrations larger than about 30% (Figure 7); this together with an isosbestic point at about 205 nm was consistent with an equilibrium between at least two states. A first state represented by the spectrum at higher concentration of MeOH (Figure 8) and characterized by a predominately α-helical structure and a second state in aqueous solution. The ion-exchange experiments described above, when combined with the observation that the p*I* of the M + D mixture was more than three units less than the calculated value from amino acid composition, seemed to suggest that this second state contained a significant amount of secondary structure and a compact form, but lacked a well-defined tertiary structure.

The secondary structure composition calculated using the variable selection method [21, 22] assigned 39% α-helix and 10% β-structure to the CD spectrum in 50% MeOH and correlated well with the predictions described above (Table 2).

The CD Structure of T. The CD spectrum of an aliquot of T, directly obtained from ion-exchange purification, was similar to that of M + D in buffer (Figure 8). When MeOH was added, changes identical to those observed for M + D were obtained. These results, together with those from the size-exclusion experiments described previously, confirmed the existence of an equilibrium between the three forms of the polypeptide.

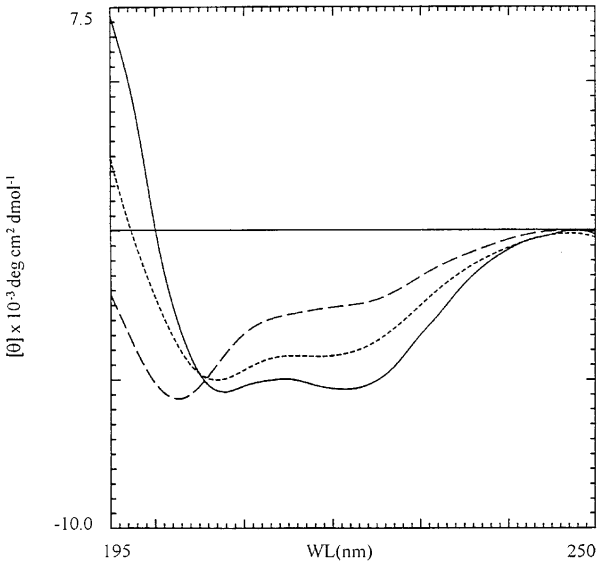

Figure 8 CD spectra of tetrameric 104-mer in 20 mM MOPS, pH 7 (broken line) and tetrameric 104-mer after addition of 30% (dotted line) and 50% MeOH (solid line). The isosbestic point at 205.6 nm corresponds to that found during the MeOH into buffer titration experiments of the M + D mixture (see Figure 7).

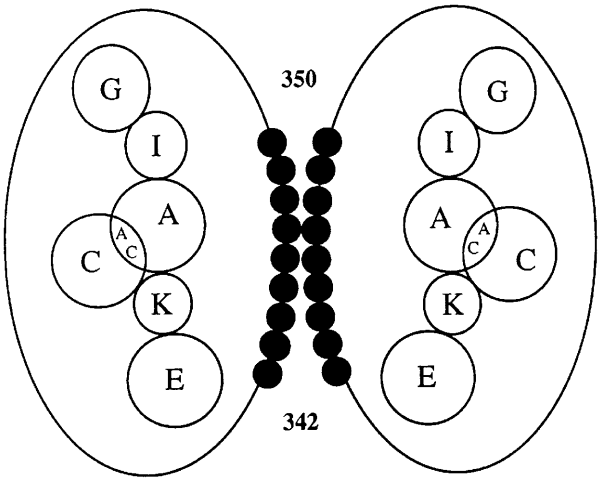

Figure 9 Schematic description of the main structural characteristics of the synthetic 104-mer in its proposed dimeric form. The regions circled and their relative position represent the protein antibody binding sites as defined by anti 104-mer monoclonal antibodies [9]: E = 275–280; K = 281–286; A = 284–289 + 351–356; C = 288 − 293 + 351 − 356; I = 358–363; G = 363–368. The overlapping area (AC) between the A and C circles represents the amino acid region shared by the two discontinuous antibody binding sites. The filled circles are the proposed dimerization region (amino acid sequence 342–340) of the synthetic 104-mer as well as full-length p24.

DISCUSSION

The chromatography experiments described here have highlighted the existence of three different forms of the synthetic 104-mer. These differed from one another in size, i.e. monomer, dimer and tetramer, and also in p*I* values, which were about 5.6 (M + D mixture) and 9.3 (T). The existence of these three forms did not seem to be a direct consequence of the replacement of the two Cys residues at positions 330 and 350 with Ala. Thus, preliminary results (not shown) on a synthetic analogue of 104-mer, which contains the two Cys residues at positions 330 and 350, show that the molecule, after disulphide formation contains two species characterized by p*I*s of approximately 5.5 and 8. Furthermore, concentrated solutions of this analogue, applied to a cationic-exchange chromatography column, were strongly retained by the medium in a fashion. similar to that seen for the 104-mer.

CD spectroscopy as well as ion-exchange studies indicated that the M + D mixture contained, in aqueous solutions, a compact molecule lacking, however, a well-defined tertiary structure. This conclusion was consistent with both MeOH into buffer titration studies and the experimentally derived p*I* of the mixture which was about three units less than the expected value calculated form amino acid composition.

In MeOH/water mixtures the CD structure of M + D was estimated to contain about 40% α-helix contribution, in agreement with secondary structure predictions. Whether the addition of MeOH perturbed the equilibrium between M and D together with inducing more defined secondary and tertiary structures could not be unequivocally determined by our study. Thus, the lack of correlation between molecular size of standard proteins and their retention time in 50% MeOH solutions did not allow a conclusion to be drawn. However, when a sample of M + D kept in 50% MeOH was analysed by size-exclusion chromatography, the proportion of dimer in the M + D mixture increased, suggesting a role of the organic solvent in the equilibrium between the two forms (data not shown).

Dilution experiments as well as CD spectroscopy studies indicated that an equilibrium existed between M, D and T. This equilibrium could be perturbed by varying the peptide concentration or by the addition of MeOH.

Finally, the results shown here for the 104-mer indicated structural analogies with full-length p24

protein. Thus, (i) the p*I* of full-length p24 is approximately 6.25 on IEF gel (lane 3, Figure 1), (ii) full-length p24 has been shown to exist in solution as a dimer [7] and the 104-mer also forms dimers, (iii) the CD spectra of full-length p-24 and that of synthetic 104-mer in MeOH are those of an $\alpha+\beta$ protein and (iv) p24 at high concentrations has a tendency to form tetramers, a characteristic also exhibited by synthetic 104-mer.

In the light of the above observations it seems reasonable to conclude that the conformational properties of the synthetic protein are maintained in full-length p24. Also consistent with this conclusion were the results obtained with anti 104-mer MoAbs which, when used in combination with a set of overlapping peptides, not only allowed the identification of the synthetic protein epitopes, but also suggested that the latter adopts a conformation that resembles that of native p24. Thus, with the exception of one MoAb, all the other six antibodies, which included linear and conformational epitopes, crossreacted with p24 as well as HIV-1 infected cells [9].

Based upon the results presented here and the experimentally derived specificity of anti-104-mer MoAbs, a topology of the 104-mer structure was derived (Figure 9). In this model the surface helix corresponding to the amino acid segment 342–350 represents the sequence through which dimerization of the 104-mer as well as p24 might take place in agreement with the recent observation that a synthetic peptide encompassing amino acids 339–349 of p24 interferes with the self-association processes of p55/p24 [31].

Notes added in proof

During the course of revision of this work a manuscript was published which describes the crystal structure of dimeric HIV-1 p24 complexed with a cognate antibody fragment (32). This study reveals that the C-terminal region of p24 (sequence 152-231 or sequence 284–363 according to the numbering used by us) is disordered in the crystal structure. Thus our conclusion on 104-mer in aqueous solution as lacking a well-defined tertiary structure is substantiated by the X-ray data.

Furthermore the latter reveal that monomer–monomer contacts take place through regions contained in the N-terminal region of p24. Whether additional contacts, like those proposed in our work, involve sequences of the C-terminal region cannot be established, although it is suggested that the C-terminal domain of p24 directs higher order oligomerisation (32 and references therein).

Acknowledgements

The authors wish to thank R. Antolin for the artwork and for technical assistance in editing the manuscript.

References

1. H. R. Gelderblom, E. H. S. Hausman, M. Ozel, G. Pauli and M. A. Koch (1987). Fine structure of HIV and immunolocalisation of structural proteins. *Virology 156*, 171–176.
2. M. G. Rossmann (1988). Antiviral agents targeted to interact with viral capsid proteins and a possible application to human immunodeficiency virus. *Proc. Natl Acad. Sci. USA 85*, 4625–4627.
3. P. Argos (1989). A possible homology between immunodeficiency virus p24 core protein and picornaviral VP2 coat protein: prediction of HIV p24 antigenic sites. *EMBO. 8*, 779–785.
4. A. R. M. Coates, J. Cookson, G. J. Barton, M. J. Zvelebil and M. J. E. Sternberg (1987). AIDS vaccine predictions. *Nature 326*, 549–550.
5. M. J. E. Sternberg, J. B. Geoffrey, M. J. Zvelebil, J. Cookson and A., R. M. Coates (1987). Prediction of antigenic determinants and secondary structures of the major AIDS virus proteins. *FEBS Lett. 218*, 231–237.
6. N. R. Burns, S. Craig, S. R. Lee, M. H. Richardson, N. Stenner, S. E. Adams, S. M. Kingsman and A. J. Kingsman (1990). Purification and secondary structure determination of simian immunodeficiency virus p27. *J. Mol. Biol. 216*, 207–211.
7. S. Rose, P. Hensley, D. J. O'Shannessy, J. Culp, C. Debouck and I. Chaiken (1992). Characterisation of HIV-1 p24 self-association using analytical affinity chromatography. *Proteins: Structure, Function Genetics 13*, 112–119.
8. P Mascagni, D. Y. Sia, A. R. M. coates and W. A. Gibbons (1990). Protein engineering of HIV proteins by total chemical synthesis: The C-terminal 104 residue peptide from gagp24. *Tetrahedron Lett. 31*, 4637–4640.
9. L. R. Haaheim, J. P. Maskell, P. Mascagni and A. R. M. Coates (1991). Fine molecular specificity of linear and assembled antibody binding sites in HIV-1 p24. *Scand. J. Immunol. 34*, 341–350.
10. K. E. Brown, D. C. Shanson, P. Mascagni and A. R. M. Coates (1991). Molecular specificity of two commercial enzyme linked immunosorbent assays for human immunodeficiency virus antigens. *J. Clin. Pathol. 44*, 102–1024.

11. D. Y. Sia, P. Mascagni, A. R. M. Coates and W. A. Gibbons in: *Vaccines for Sexually Transmitted Diseases* A. Meheus and R. E. Spier, Eds, p. 275–280, Butterworth, Sevenoaks. 1989.
12. S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole and M. Alizon (1985). Nucleotide sequence of the AIDS virus. *Cell 40,* 9–11.
13. M. Alizon, S. Wain-Hobson, L. Montagnier and P. Sonigo (1985). Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell 46,* 63–74.
14. S. B. H. Kent (1988). Chemical synthesis of peptides and proteins. *Ann. Rev. Biochem. 57,* 957–989.
15. H. L. Ball and P. Mascagni (1992). Purification of synthetic peptides using reversible chromatographic probes based on the Fmoc molecule. *Int. J. Peptide Protein Res. 40,* 370–379.
16. H. L. Ball, G. Bertolini, S. Levi and P. Mascagni (1994). Purification of synthetic pepties with the aid of reversible chromatographic probes. *J. Chromatogr. A 686,* 73–83.
17. H. L. Ball and P. Mascagni (1995). *N*-(2-Chlorobenzyloxycarbonyloxy) succinimide as a terminating agent for solid-phase peptide synthesis: application to a one-step purification procedure. *Lett. Peptide Sci. 2,* 49–57.
18. H. L. Ball and P. Mascagni (1996). Chemical synthesis and purification of proteins: A methodology. *Int. J. Peptide protein Res.* (in press).
19. J. P. Tam (1987). Synthesis of biologically active transforming growth factor alpha. *Int. J. Peptide Protein Res. 29,* 421–431.
20. S. B. Needleman and C. D. Wunch (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol. 48,* 443–453.
21. M. O. Dayhoff, R. M. Schwartz and B. C. Orcutt in: *Atlas of Protein Sequence and Structure* M. D. Dayhoff, Ed, p. 345–352, National Biomedical Research Foundation, Washington DC 1978.
22. D. F. Feng and R. F. Doolittle (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol. 25,* 351–360.
23. J. S. Richardson and D. C. Richardson in: *Prediction of Protein Structure and the Principles of Protein Conformation.* G. D. Fasman, Ed., p. 1–99, Plenum Press, New York 1989.
24. P. Y. Chou and G. D. Fasman (1974). Prediction of protein conformation. *Biochemistry 13,* 222–245.
25. J. Garnier, D. J. Osguthorpe and B. Robson (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol. 120,* 97–120.
26. F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga (1975). Energy parameters in polypeptides. VII. Geometric parameters, potential atomic charges, non-bonded interactions, intrinsic torsional potential for the naturally occurring aminoacids. *J. Phys. Chem. 79,* 2361–2381.
27. P. Manavalan and W. C. J. Johnson (1983). Sensitivity of circular dichroism to protein tertiary structure class. *Nature 305,* 831–832.
28. P. Manavalan and W. C. J. Johnson (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem. 167,* 76–85.
29. N. B. Egen, M. Bliss, M. Meyersoh, S. M. Owens, L. Arnold and M. Bier (1988). isolation of monoclonal antibodies to phencyclidine from ascites fluid by preparative isoelectric focusing in the Rotofor. *Anal. Biochem. 172,* 488–494.
30. G. Esposito, A. Pessi and A. S. Verdini (1989). H-NMR studies of synthetic polypeptide models of Plasmodium falciparum circumsporozoite protein tandemly repeated sequence. *Biopolymers 28,* 225–246.
31. S. Modrow, A. Poblotzki, R. Wagner, H. R. Gelderblom, G. Pauli, H. Walof and M. Niedrig in: *Peptides 1992,* C. H. Schneider and A. N. Eberle, Eds., p. 133–135, ESCOM. Leiden 1993.
32. C. Momany, L. C. Kovari, A. J. Prongay, W. Keller, R. K. Gitti, B. M. Lee, A. E. Gorbalenya, L. Tong, J. McClure, L. S. Ehrlich, M. F. Summers, C. Carter and M. Rossmann (1996). Crystal Structure of dimeric HIV-1 capsid protein. *Nature Struct. Biol. 3,* 763–770.

