JOURNAL OF VIROLOGY, June 1993, p. 3552–3560
0022-538X/93/063552-09$02.00/0

Vol. 67, No. 6

Secondary Structure of gp160 and gp120 Envelope Glycoproteins of Human Immunodeficiency Virus Type 1: a Fourier Transform Infrared Spectroscopic Study

ETIENNE DECROLY, BERNARD CORNET, ISABELLE MARTIN, JEAN-MARIE RUYSSCHAERT, AND MICHEL VANDENBRANDEN*

Laboratoire de Chimie-Physique des Macromolécules aux Interfaces CP206/2, Université Libre de Bruxelles, 1050 Brussels, Belgium

Received 30 October 1992/Accepted 12 March 1993

The secondary structure of the precursor (gp160) of the envelope protein of human immunodeficiency virus type 1 (BH10) and its receptor-binding subunit (gp120) was studied by Fourier-transformed attenuated total reflection spectroscopy. A higher α-helix/β-sheet ratio in the gp120 subunit than in the precursor indicates a structural heterogeneity between the two subunits (gp120 and gp41), in agreement with classical secondary-structure predictions. The secondary structure of gp41 was estimated and compared with existing models. The high α-helical content in gp41 and the dominant β-sheet content in gp120 resemble the distribution in influenza virus hemagglutinin subunits.

Knowledge of the structure of the human immunodeficiency virus (HIV) envelope glycoproteins is essential to the understanding of viral pathogenesis and infectivity at the molecular level. Although the primary sequence has been determined for many strains of HIV, little is known about the molecular structure of HIV envelope proteins.

Specific mutagenesis and epitope mapping have permitted identification of sites involved in CD4 recognition (20, 26, 27, 31), in transmembrane and surface subunit interactions (18), and in the fusion of viral and cellular membranes (8, 9, 20).

However, on the basis of these indirect topological data, establishment of an envelope glycoprotein molecular structure remains highly speculative. A major limiting step in determining the tertiary structure of membrane proteins is the difficulty encountered in obtaining crystals suitable for high-resolution X-ray crystallography and in the present case the fact that the envelope glycoprotein of HIV is made of a transmembrane-surface protein complex. Since crystallographic data about the envelope proteins of HIV are missing, precise information about structural features such as secondary structure could be useful for comparison with proteins of well-known structure. However, the turbidity of membrane systems prevents them from being analyzed by spectroscopic methods using UV or visible light. Infrared (IR) spectroscopy has been shown to be well designed for studying the structure of proteins and peptides in a lipid environment (1, 3, 12, 13, 15, 28).

The scope of the present study is to determine by Fourier transform IR (FTIR) spectroscopy the secondary structure of the gp160 glycoprotein precursor of transmembrane (gp41) and surface (gp120) subunits of the HIV type 1 (HIV-1) envelope glycoprotein in a reconstituted system preserving its CD4-binding capacity. Comparison of the secondary structure of gp160 with the secondary structure of one of its subunits, gp120, provides additional information about the structure of the transmembrane subunit gp41.

* Corresponding author.

MATERIALS AND METHODS

Materials. Detergents octylglucopyranoside (OGP), deoxycholate, and polyoxyethylene-sorbitan monolaurate (Tween 20) were purchased from Sigma Chemicals (O 8001, D 6750, and P 1379, respectively). Triton X-100 was a Merck product (no. 11869).

Lipids DL-α-dimyristoylphosphatidylcholine (DMPC) and cholesterol were from Sigma (P 0888 and C 8378). Tritiated cholesterol (40 to 60 Ci/mmol) was from Amersham (TRK 330).

Soluble recombinant CD4 (sCD4) was a gift of R. Sweet from Smith Kline & French Laboratories, Philadelphia, Pa. gp120 expressed in CHO cells (Celltech lot 10004542/52) was a gift from H. C. Holmes (Medical Research Council [England] AIDS program).

Antibodies OKT4 and OKT4a were from Ortho Diagnostics, anti-gp41 sheep antibody and anti-gp120 sheep antibody were from Biochrom (D7323 and 7324), mouse monoclonal anti-gp120 antibody was from Dupont, and biotinylated sheep anti-mouse antibodies and biotinylated sheep anti-human antibodies were from Amersham (RPN 1021 and RPN 1003, respectively). The pool of human sera containing antibodies against gp160 was a gift from C. Bruck (Smith-Kline-Beecham, Rixensart, Belgium).

Other reagents were streptavidin-peroxidase complex (Amersham, RPN 1051), orthophenylenediamine dihydrochloride (Sigma, P 6787), trazilol (Aprotinin, Bayer), and phenylmethylsulfonyl fluoride (Sigma, P 7626).

Purification of gp160. Recombinant vaccinia virus carrying the HIV (BH10) *env* gene was a gift from C. Bruck (Smith-Kline-Beecham). HIV gp160 envelope glycoprotein was expressed in CV-1 cells by using the vaccinia virus *env* recombinant and purified on lentil-lectin agarose and immunoaffinity columns (unpublished data). All the purification steps were carried out in the presence of 1% OGP. Proteins were eluted from the immunoaffinity column with a 0.150 M KCl–1% OGP solution adjusted to pH 2.5 with HCl instead of conventional elution buffers, which contain carboxylic

acids and strongly interfere with the IR spectrum of proteins and which are difficult to remove completely in subsequent steps. The gp160 showed a single band with only minor gp120 contamination (less than 5%) on a sodium dodecyl sulfate gel.

Formation of liposomes. The purified gp160 (30 μg/ml) was solubilized in 1% OGP–1% deoxycholate–0.1 mM phenylmethylsulfonyl fluoride–100 U of trazilol per ml for 1 h at room temperature. A lipid film was formed by evaporation under nitrogen of a chloroform solution containing DMPC and cholesterol in a 4:3 molar ratio and an aliquot of tritiated cholesterol. Lipids were dispersed in phosphate-buffered saline (PBS) containing 4% OGP; proteins were then added (protein/lipid ratio, 1:1 [wt/wt]), and the mixture was incubated for 40 min at 20°C.

Reconstitution of gp160 into liposomes was accomplished after detergent removal by dialysis against PBS at 4°C for at least 20 h. Protein-free liposomes (control experiment) were formed by using exactly the same procedure except that viral proteins were omitted. For IR measurements, further dialysis was made against distilled water (see below).

CD4 binding to gp160 liposomes. Liposomes were incubated with sCD4 receptor at a ratio of 0.25 μg/μg of gp160 recombinant proteins (2 h at 20°C). gp160 liposomes (100 μl) were mixed with an equal volume of 80% sucrose in PBS in a 4-ml Beckman centrifuge tube. A 30 to 20% continuous sucrose gradient in PBS was overlaid. Centrifugation was carried out at 120,000 × *g* for 16 h at 4°C in a Beckman SW60Ti rotor. gp160 and the gp160-sCD4 complex were detected by enzyme-linked immunosorbent assay (ELISA), and lipids were detected by radioactivity counting (^{3}H-cholesterol). In control experiments, sCD4 was preincubated with anti-CD4 monoclonal antibodies for 1 h at 20°C (5 μg of OKT4 or OKT4a per ml) to determine the specificity of the CD4-gp160 interaction.

Detection of gp160 and the gp160-CD4 complex by ELISA. Ninety-six-well immunoplates (Nunc) were coated overnight at 4°C with 5 μg of anti-gp41 sheep antibody per ml. After saturation with PBS containing 4% newborn calf serum and 1% bovine serum albumin (saturation buffer), the sample (gp160 liposomes in 1% Triton X-100) was added to the coated wells and incubated for 2 h at 20°C or overnight at 4°C. The second antibody (monoclonal anti-gp120, used to detect gp160, or OKT4 anti-CD4, used to detect the gp160-CD4 complex) was then added for 90 min at 37°C. Detection was performed with 50 μl of a streptavidin-peroxidase complex diluted 1:1,000 in saturation buffer (15-min incubation at 37°C, in the dark) after addition of the chromogenic substrate orthophenylenediamine dihydrochloride at 0.4 mg/ml in 0.1 M citrate buffer, pH 4.5, and 0.03% hydrogen peroxide. The optical density at 492 nm was measured within 1 h.

Preparation of sample for FTIR spectroscopy. Samples were dialyzed in a microdialysis apparatus using an 8,000-molecular-weight cutoff membrane (Pierce), first against PBS (24 h), then against distilled water (48 h). The membrane had been washed in distilled water and tested for the absence of released compounds interfering with IR spectra.

Samples were dried at room temperature under a nitrogen stream on a germanium plate (Harrick, EJ2121; 50 by 20 by 2 mm) for attenuated total reflection (ATR) measurement. Only the excess of water was eliminated during the preparation of the sample, with the film remaining in the hydrated state, as monitored by the O—H stretching region of water molecules around 3,300 cm^{-1}. The aperture angle was 45°, yielding 25 internal reflections. The ATR plate had been cleaned by sonication in detergent and rinsed sequentially with distilled water, methanol, and chloroform. It was further cleaned and made more hydrophilic after passage in a plasma cleaner (Harrick, PDC-23G).

The ATR plate was sealed in a universal sample holder (Perkin Elmer 186-0354), and H-D exchange was obtained by flushing D_2O-saturated N_2 for 3 to 5 h at room temperature before measurement. The D_2O-H_2O exchange process was monitored by the change of the O—H stretching band around 3,300 cm^{-1} into the O—D band around 2,490 cm^{-1}. During this period of time, only readily accessible peptide bonds are exchanged. Importantly, the random structure shifts from about 1,655 cm^{-1} to about 1,640 cm^{-1} upon H-D exchange, allowing differentiation of the α-helix from the random structure. The ATR FTIR spectra were recorded on a Perkin-Elmer 1720-X FTIR spectrophotometer equipped with a liquid nitrogen-cooled mercury-cadmium telluride detector. For each spectrum, 64 to 128 scan cycles were averaged; at each cycle, the ratio of the sample spectrum to the background spectrum of a clean germanium reference ATR plate was determined by using a shuttle to move the sample or the reference into the beam. Prior to sample application, the baseline of the clean sample plate was recorded for further spectrum correction. The spectra were recorded at a nominal resolution of 4 cm^{-1} and encoded every 1 cm^{-1}. None of the experimental spectra shown was further smoothed. However, resolution-enhanced spectra were submitted to an apodization procedure during Fourier self-deconvolution (see below), resulting in smoothing of very narrow peaks present in the background of experimental spectra (19). The spectrophotometer was continuously purged with dry air to reduce water vapor absorption in the spectral region of interest.

Secondary-structure determination. The secondary-structure determination procedure, based on the analysis of the amide I′ spectrum (1,600 to 1,700 cm^{-1}), has been described extensively (3, 12, 14). Self-deconvolution of the spectrum was carried out by using a Lorentzian function for the deconvolution and a Gaussian function for the apodization (19). The *k* factor (the resolution enhancement factor) is defined as the ratio of the full width at half height of the deconvoluting Lorentzian curve to the full width at half height of the Gaussian curve used for apodization. For the curve-fitting procedure, the amide I′ band was decomposed into a series of Lorentzian bands assigned to individual secondary-structure components. The initial positions of the Lorentzian bands were those corresponding to maxima most frequently appearing in protein spectra whose resolutions have been enhanced by numerical treatment such as Fourier self-deconvolution, as already described (12, 14). The proportion of a particular secondary structure in the protein was calculated as the sum of the areas of all the Lorentzian bands falling in the frequency domain corresponding to that structure divided by the sum of the areas of all the Lorentzian bands with maxima between 1,689 and 1,615 cm^{-1}. The amide I′ band has been divided into several frequency domains corresponding to different secondary structures. The frequency limits for each structure were first assigned according to data determined theoretically (21) or experimentally (2). These limits were slightly adjusted to obtain a good agreement between the percentage of each secondary structure determined by IR spectroscopy and that determined by X-ray crystallographic data (22) for a set of purified proteins (12).

The frequency limits (cm^{-1}) of the secondary structure domains were as follows: α-helix, 1,661 to 1,647.5; β-sheet, 1,689 to 1,682 and 1,637.5 to 1,615; random, 1,644.5 to

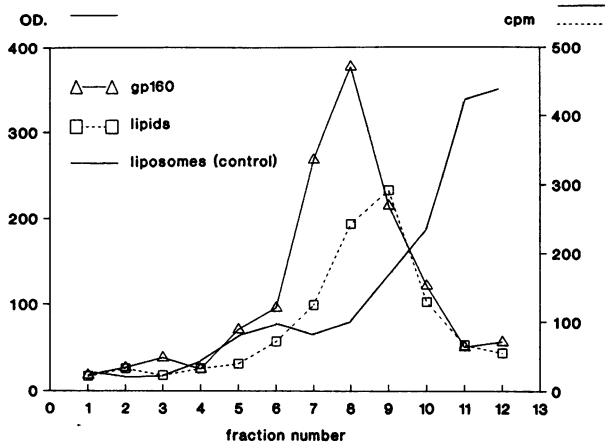

FIG. 1. Centrifugation (120,000 × g for 16 h at 4°C) of gp160 liposomes or free liposomes (control) on a 40 to 2% sucrose gradient. gp160 was detected by ELISA, and lipids were detected by radioactivity counting. Fractions were collected from the bottom of the tube. OD, optical density.

1,637.5; turn, 1,682 to 1,661. The initial frequencies (cm^{-1}) of the Lorentzian bands used to fit the spectrum were 1,695, 1,683, 1,678, 1,672, 1,664, 1,657, 1,648, 1,640, 1,632, and 1,624. Intensities were calculated as two-thirds of the spectrum intensity at the corresponding frequency, and the full width at half height was set at 10 cm^{-1}. Lorentzian bands whose surface contribution to the spectrum did not exceed 25% were rejected. Then, the frequency, intensity, and full width at half height of each Lorentzian band were varied by using a least-squares iterative procedure (12, 14) to fit the experimental spectrum. For a quantitative evaluation of secondary structure, we assumed that the integrated extinction coefficient of the amide group is equal for all types of structures and all types of proteins and that the contribution of side chains in the region of amide I can be neglected. As a matter of fact, these approximations have not impeded the correct estimation (within a small percentage) of the proportion of each secondary structure for a large number of soluble proteins in several approaches essentially identical to ours (2, 4, 29, 30).

RESULTS

Insertion of gp160 into the lipid bilayer of liposomes and CD4 binding. The precursor of the envelope glycoproteins of HIV (gp160) contains (at least) one predicted membrane-spanning segment and aggregates in the absence of detergents. Therefore, the secondary structure of purified gp160 was investigated after insertion into the phospholipid bilayer of liposomes made of DMPC-cholesterol (molar ratio, 4:3). The purified gp160 solubilized in 1% OGP–1% deoxycholate was inserted into the lipid bilayer of liposomes by using the detergent dialysis procedure (see Materials and Methods). After centrifugation on a continuous density gradient (Fig. 1), all the gp160 had comigrated with liposomes (fractions 6 to 10), indicating quantitative reconstitution. As expected, liposomes formed in the absence of proteins move at a lower density (Fig. 1, fractions 10 to 12). The capability of gp160 to bind to CD4 after reconstitution was tested as a further indication that the reconstitution has been performed in conditions that preserve this essential biological property of gp120. After incubation with sCD4, gp160 liposomes were centrifuged in a continuous sucrose gradient. The sCD4-gp160 liposome complex was identified by an ELISA specific for sCD4 bound to gp160 (Fig. 2a). To assess the specificity of the binding, the anti-CD4 antibody OKT4a, which competitively inhibits binding of CD4 to gp160, was preincubated with CD4. OKT4a inhibited CD4 binding to gp160 liposomes (Fig. 2b), whereas OKT4 antibodies (these anti-CD4 antibodies do not compete with gp160-CD4 binding) did not prevent recognition (Fig. 2a), demonstrating that CD4 binding is specific.

Secondary-structure determination. The secondary structure of gp160 and gp120 was evaluated on the basis of the analysis of the FTIR spectra in the 1,700 to 1,600 cm^{-1} region. For proteins, the major contribution (80%) to this spectral region originates from the C=O stretching of the

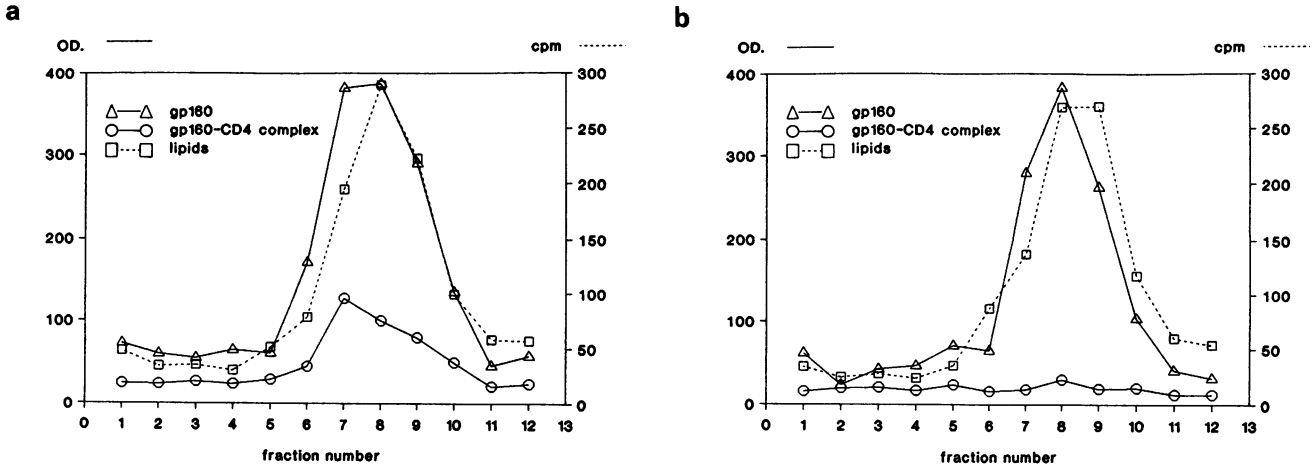

FIG. 2. (a) Binding of CD4 to gp160 liposomes. CD4 was preincubated with OKT4 and mixed with gp160 liposomes as described in Materials and Methods. Unbound CD4 was separated from gp160 liposomes by sucrose gradient centrifugation (120,000 × g for 16 h at 4°C). gp160 and the gp160-CD4 complex were detected by ELISA, and lipids were detected by radioactivity counting (^{3}H-cholesterol). (b) Inhibition of gp160-CD4 binding by OKT4a. CD4 was preincubated with OKT4a prior to mixing with gp160 liposomes as described in Materials and Methods. OD, optical density. Fractions were collected from the bottom of the tube.

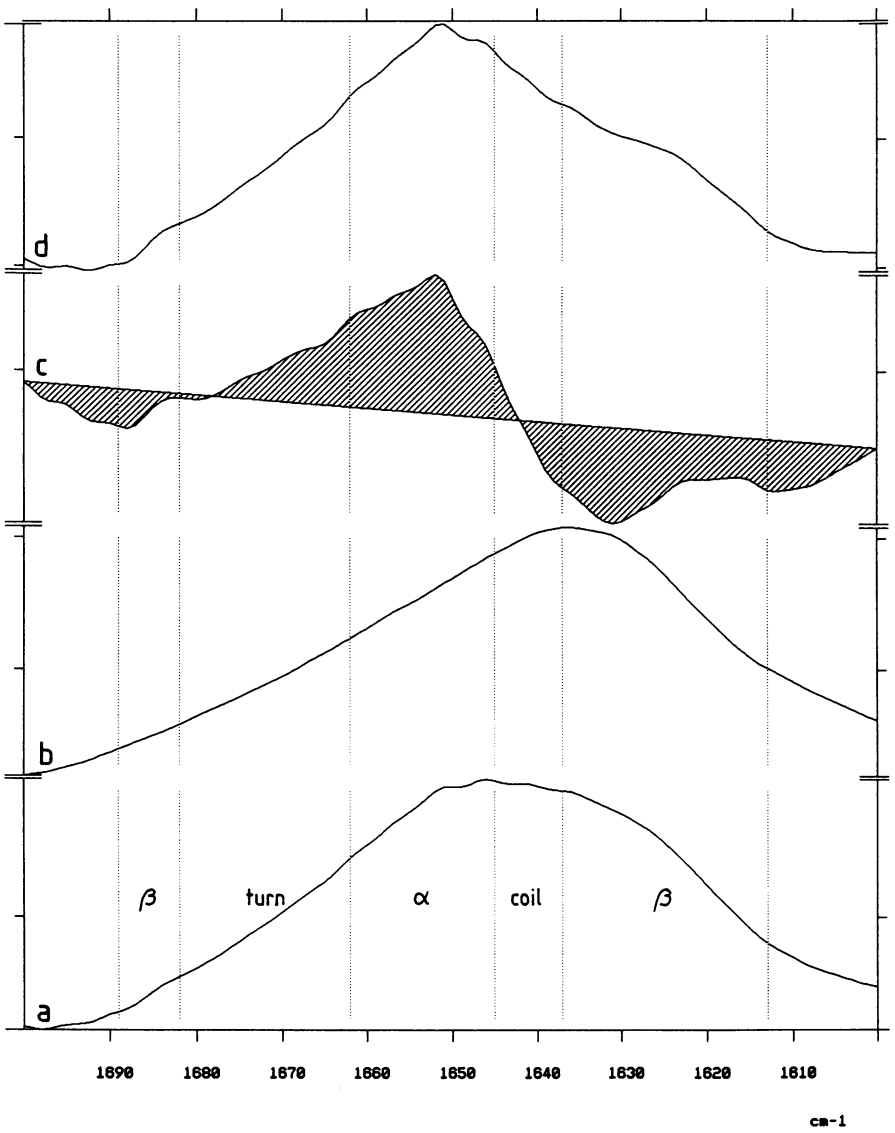

FIG. 3. FTIR spectra (amide I' band, 1,600 to 1,700 cm^{-1}) of gp160 inserted into liposomes (a) and gp120 (b). (c) Difference spectrum for spectra a and b; (d) calculated spectrum for gp41. Spectra are drawn at full scale. Frequency domains corresponding to different secondary structures are delimited by vertical lines. To calculate spectrum c, the areas of spectra a and b were normalized prior to subtraction (subtraction coefficient = 1). Areas corresponding to positive and negative surfaces are hatched, showing their respective contributions. Spectrum d was calculated by the same method as spectrum c except for the subtraction coefficient, which is equal to the gp120/gp160 peptide length ratio (0.60).

peptide amide bond. Its vibrational frequency is very sensitive to hydrogen bonding with other peptide groups or with solvent and reflects the type of secondary structure in which it is involved.

The gp160 shows a broad-band spectrum typical of proteins, with a maximum around 1,645 cm^{-1} and a shoulder in the 1,625 to 1635 cm^{-1} region (Fig. 3, spectrum a). The amide I' band (1,600 to 1,700 cm^{-1}) of gp120 is characterized by an asymmetric broad-band spectrum (Fig. 3, spectrum b), with a maximum around 1,635 cm^{-1} in a frequency domain typical of β-sheet-containing proteins. Comparison of the gp160 and gp120 spectra (Fig. 3, spectra a and b) provides evidence that the two proteins have different secondary-structure contents, with gp120 showing a higher β-sheet spectral component and its precursor, gp160, containing a higher α-helix component together with an important β-sheet contribution. Resolution enhancement by Fourier self-deconvolution separates the spectral components better and further confirms this view (Fig. 4). To appreciate the differences between gp160 and gp120, a difference spectrum (gp160 versus gp120) was generated (Fig. 3, spectrum c). Strong positive deviation in the α-helix region and negative deviation in the β-sheet regions of the difference spectrum suggest that domains of gp160 that are not part of the gp120 subunit (i.e., the gp41 subunit) have a dominant α-helix content (a quantitative estimate will be discussed below). Figure 3 also shows the calculated spectrum (spectrum d) for gp41 after subtraction of the gp120 contribution from the gp160 spectrum, with the subtraction coefficient being equal to the gp120-gp160 polypeptide length ratio.

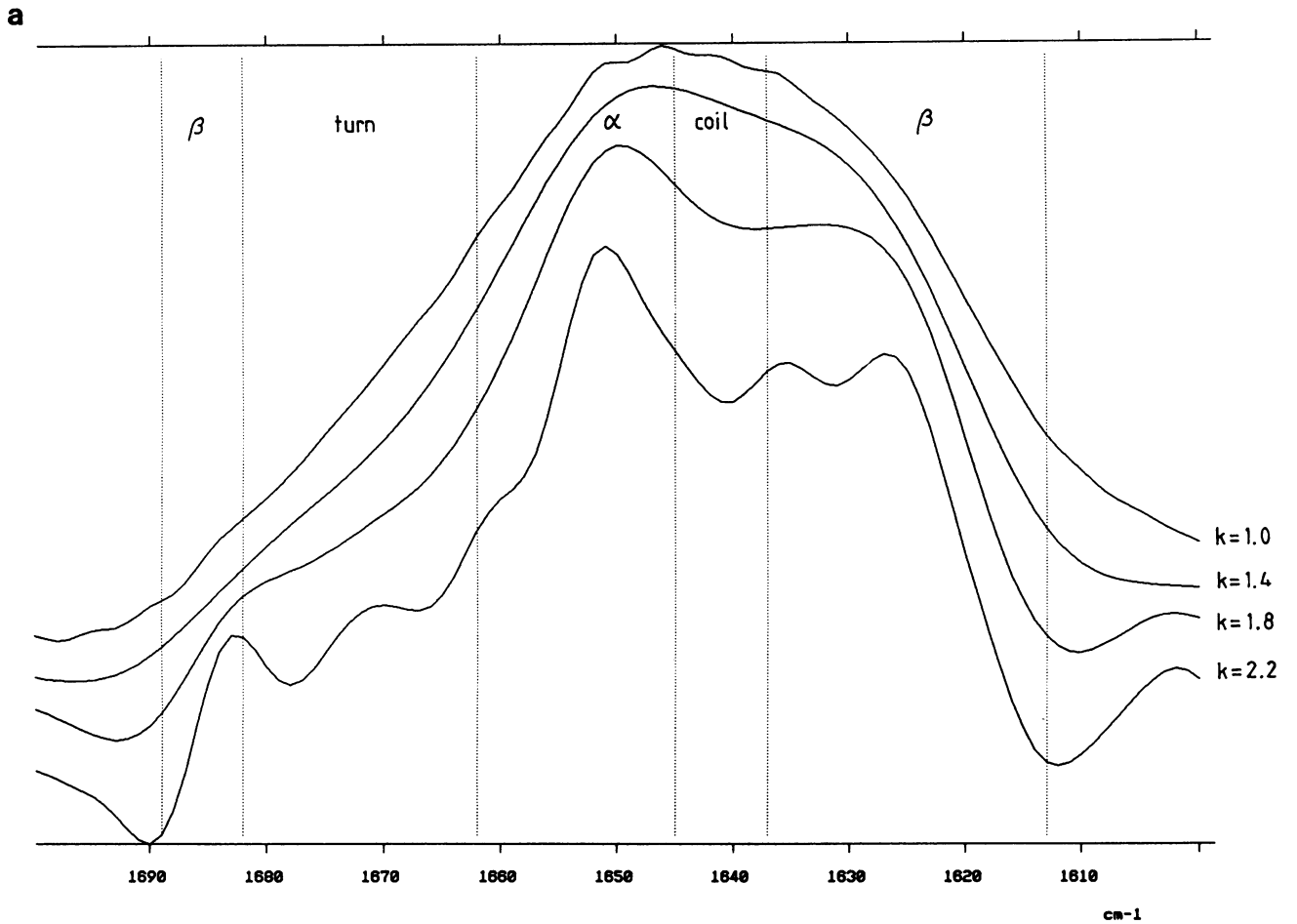

FIG. 4. Enhancement of resolution by Fourier self-deconvolution of the gp160 (a) and gp120 (b) spectra. Spectra are drawn with increasing resolution factors (*k*). Frequency domains corresponding to different secondary structures are delimited by vertical lines.

The secondary structure of gp120 and gp160 was estimated by applying the iterative curve-fitting procedure (see Materials and Methods) to the experimental (untreated) spectra of Fig. 3 (spectra a and b). Figure 5 shows the Lorentzian bands selected by the procedure and the resulting fitting. Table 1 gives the secondary-structure content deduced from the curve fitting. With the data in Table 1, one can calculate a β-sheet/α-helix ratio of 1.75 for gp120 and 1.09 for gp160, indicating a structural heterogeneity inside the envelope glycoprotein and suggesting that a large proportion of α-helix should be associated with gp41. Assuming that the secondary structure of gp120 in the precursor (gp160) and that in the free form are nearly identical, one can deduce the contribution of gp41 to the secondary structure of gp160. The results suggest a high α-helix/β-sheet ratio in gp41 (Table 1).

DISCUSSION

IR spectroscopy has been shown to be useful for analysis of the secondary structure of proteins, especially membrane proteins (3, 12–15, 32). It requires small amounts of protein and is not subject to artifacts due to light scattering of membrane systems such as those observed with circular dichroism measurements.

In a recent study by Goormaghtigh et al. (12), it was shown with a calibrating set of soluble proteins that IR data analyzed by the procedure described in this article give a correct estimation of α-helix and β-sheet structure contents with a standard deviation of 8.6% (n = 28) when the secondary structure parameters of Levitt and Greer (22) for X-ray structures are taken as a reference. The procedure applied to membrane proteins (12) gives results in good agreement with available X-ray data (e.g., for bacteriorhodopsin and porin).

We have estimated the secondary structure of the HIV-1 envelope glycoprotein precursor (gp160) and its gp120 subunit. The two glycoproteins were purified by the same method (immunoaffinity), except that gp160 was first prepurified on a lentil-lectin column (see Materials and Methods), and both recognized to the same extend recombinant sCD4. Moreover, another gp120, produced in a *Drosophila* expression system and purified by ion exchange chromatography gave essentially the same secondary structure as that for the gp120 from CHO cells used in the present study (unpublished data), showing that, in the present case, differences in the expression system and purification protocol do not affect the secondary structure. The estimated purity of protein samples was 95% or higher, and spectra from different purification batches were identical. Thus, the secondary-

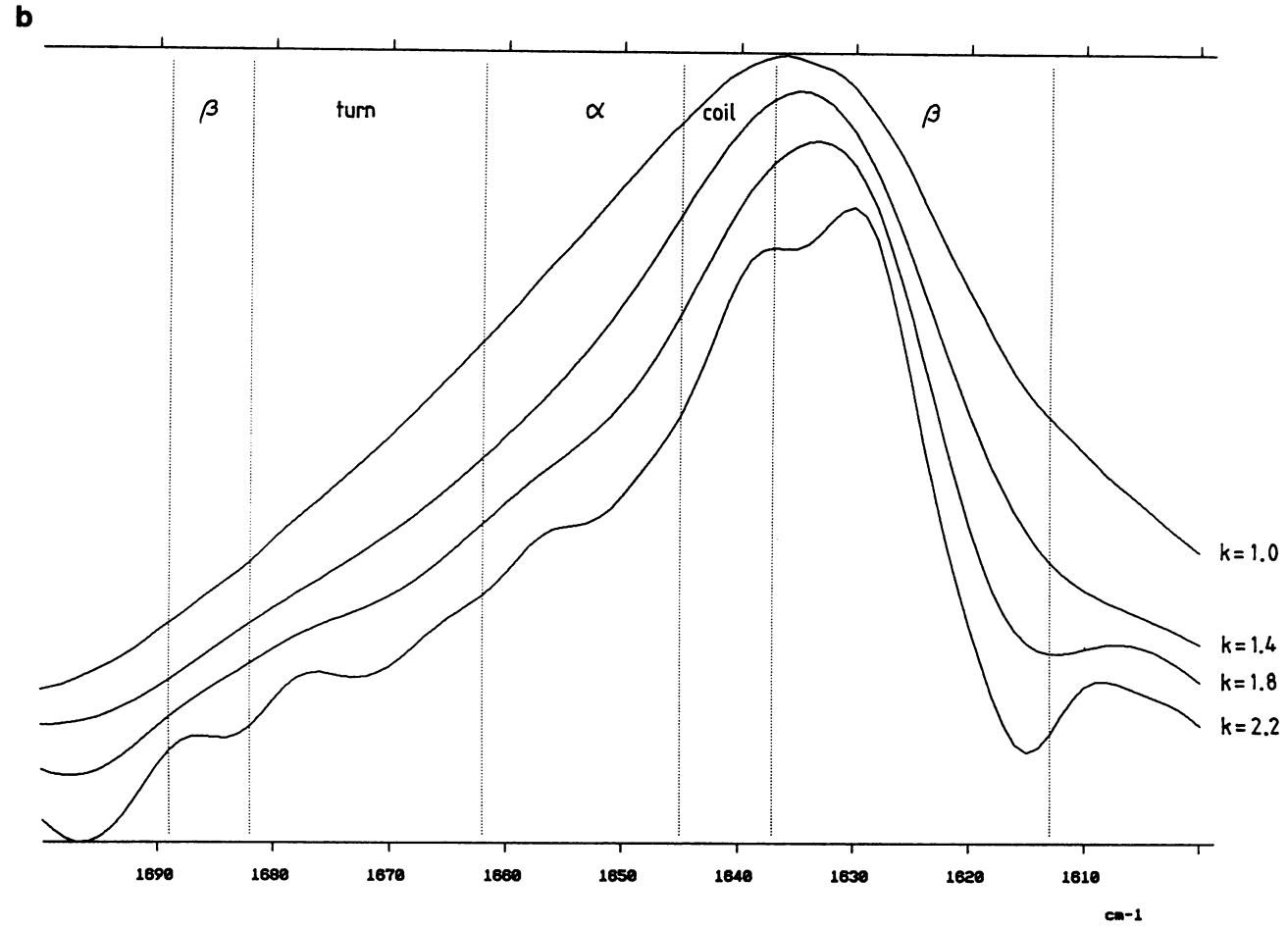

FIG. 4—*Continued.*

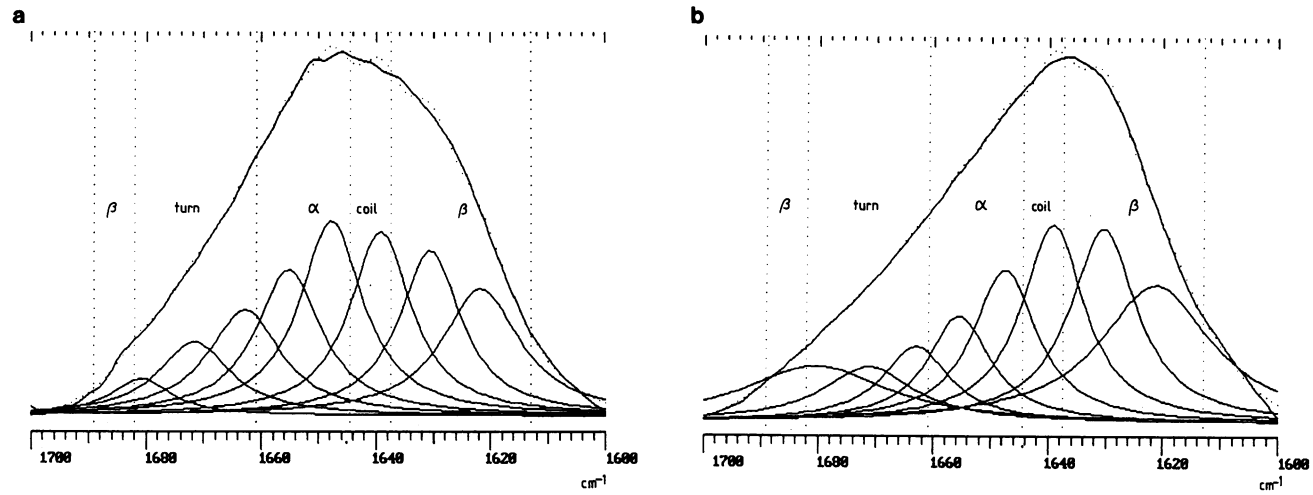

FIG. 5. Curve fitting of the amide I′ spectra of gp160 inserted into liposomes (a) and gp120 (b). Individual Lorentzian components used to fit the experimental curve are shown under the amide I′ spectrum. The resulting curve is shown as a dotted line. Frequency domains corresponding to different secondary structures are delimited by vertical dotted lines.

TABLE 1. Secondary-structure contents of gp160, gp120, and gp41

Protein	Content (%)[a]			
	α-Helix	β-Sheet	Coil	Turn
gp160	32.0	29.4	16.5	22.1
gp120	21.7	37.9	16.4	24.0
gp41	47.4	16.6	16.6	19.2

[a] The data for gp160 inserted into liposomes and gp120 are tentative determinations based on spectra a and b in Fig. 3. The data for the gp41 subunit were calculated by subtracting the contribution of gp120 from the value for gp160 for each type of structure. The subtraction coefficient was estimated from the ratio of the polypeptide length of gp120 to that of gp160.

structure data obtained are expected to reflect the real structure reasonably well.

The envelope protein gp160 was reconstituted into vesicles that were dried onto a germanium plate, a process known to orient bilayers parallel to the surface of the plate (10). Therefore, the possibility that the protein has a preferential orientation relative to the normal axis to the plate exists. This could tend to alter the relative intensities of different structural components because of dichroism effects. However, analysis of spectra of aggregated gp160 (prepared in the absence of lipids), with gp160 being randomly oriented, allows calculation of a secondary structure in agreement with that of the reconstituted gp160, indicating that dichroism effects can be neglected in our system (data not shown).

We have tried to use our secondary-structure data to determine whether gp160 or gp120 can be compared with other proteins on a structural basis. It has been suggested that gp120 displays some homologies with immunoglobulin (Ig) (23) and class II major histocompatibility complex antigen (35) and, as a ligand for CD4, could incorporate the structural features of adhesion molecules and members of the Ig superfamily. Class II major histocompatibility complex constant regions are characterized by their β-strand conformation, which is a major feature of the Ig superfamily (23). IR spectroscopy of extramembrane domains of class II antigens as well as class I antigens, β_2 microglobulin, and IgG has shown (16) an important contribution of β strands to their structure, in agreement with models proposed earlier. The α-helical and β-sheet contents we found for gp120 (21.7 and 37.9, respectively) are close to the values we obtained for IgG and recombinant sCD4 by the same procedure as that used for spectrum analysis (for IgG, $\alpha = 22.3\%$ and $\Sigma\beta = 45.1\%$; for CD4, $\alpha = 20.2\%$ and $\Sigma\beta = 38.1\%$ [unpublished data]). Thus, our data, although not sufficient to support the hypothesis of an Ig-like structure for gp120, are compatible with such a structure.

In Table 1, the contribution of gp41 to the secondary structure of gp160 was calculated assuming that the secondary structure of gp120 in the precursor (gp160) and that in the free form are nearly identical. It seems difficult to imagine that a drastic change in the global secondary structure would occur upon cleavage of gp160 while both gp120 and gp160 were able to bind CD4 and a series of anti-gp120 antibodies in ELISA (not shown). The estimated secondary-structure content of gp41 (Table 1) confirms the high α-helical content of this subunit. Indeed, gp41 contains a large number of segments with α-helical propensity. The 684-to-705 region, which is the most hydrophobic segment of at least 20 residues in the protein, is a putative transmembrane helical segment as predicted by Venable et al. (33; also reference 6), confirmed as a membrane anchor by experimental studies on secretion of recombinant DNA constructs (5, 17). Two amphipathic segments (772 to 790 and 828 to 848) lying in the intracellular part of gp41, although not predicted as purely α-helical by classical algorithms (e.g., Chou-Fasman and Garnier), have extraordinarily high helical hydrophobic moments (6), making them good candidates for α-helix formation, since peptides with such helical hydrophobic moments almost invariably adopt an α-helical conformation in crystal structure (7). These two peptides are thought to form surface-seeking helices or a pair of antiparallel α-helices (33). For the 828-to-855 region, the corresponding synthetic peptide shows a cytolytic activity similar to that of a family of peptides with a common α-helical motif (25). In addition to the above-mentioned putative helices, Gallaher et al. (11) proposed a model based mainly on classical algorithms and experimental data which includes the 557-to-588, 624-to-636, and 651-to-672 segments as additional α-helices. Finally, experiments using synthetic peptides indicate that the N-terminal fusion peptide adopts an α-helical conformation in a hydrophobic environment and a β-sheet structure in water (24).

When all the putative α-helices mentioned above (684 to 705, 772 to 790, and 828 to 855 [33], 557 to 588, 624 to 636, and 651 to 672 [11], and the 16 N-terminal residues) are taken into account, the proportion of α-helix in the gp41 subunit reaches at least 41%, which is not very far from the 47.4% determined by using IR data. A high α-helical content in the gp41 subunit is also suggested if we apply classical secondary-structure prediction methods to the whole gp160 sequence. The Garnier procedure gives 38.7 and 28.8% α-helix contents for the gp41 and gp120 subunits, respectively (with parameters adjusted to fit the structure of proteins with a 20 to 50% α-helix content). The Chou-Fasman procedure gives 32.9 and 25.5% α-helix contents for gp41 and gp120, respectively. Thus, all the models and secondary-structure predictions mentioned above are consistent with our experimental data, showing a higher proportion of α-helix in the gp41 subunit.

Based on secondary-structure prediction, conserved cysteine position, glycosylation sites, immunologically relevant epitopes, and local secondary-structure pattern, Gallaher et al. (11) proposed a general structural model for the extracellular domain of HIV-I gp41 subunit and transmembrane proteins of other retroviruses, templated on the crystal structure of the influenza virus hemagglutinin 2 (HA2) subunit. The model contains the fusion peptide, an amphipathic α-helix extending towards the apex of the glycoprotein that might be involved in homodimer formation, a dominant epitope, an extended α-helix-coil region, an α-helical transmembrane region, and a conserved turn just after the transmembrane region. Our data suggest that the structural homology between the influenza virus HA and the HIV-1 glycoprotein can be extended to the level of their subunits. As shown in the crystal structure (34), the membrane-distal subunit, HA1, has a high content of β-strand structure while the membrane-proximal subunit, HA2, is composed predominantly of α-helix (the contribution of the transmembrane and short intracytoplasmic tail is not taken into account, since the crystal was obtained from the extramembrane portion of HA).

The same kind of balance exists in HIV-1 envelope glycoprotein: gp120 contains a significantly higher amount of β-strand than gp41. It is, therefore, tempting to mention the structural homology between the two glycoproteins, although the viruses do not display the same cell tropism or receptor specificity and are evolutionarily well separated. Such a structural homology together with the possibility of

determining the secondary-structure content of virtually any purified viral glycoprotein subunit in a membrane environment could constitute a more rational basis for future attempts to model the structure of these viral proteins when direct crystallographic data are not available.

ACKNOWLEDGMENTS

We thank E. Goormaghtigh for his constructive remarks on FTIR experiments, C. Bruck (Smith-Kline-Beecham) for continuous help, and H. C. Holmes and the MRC AIDS reagent program for the gp120 CHO (Celltech) gift.

This work was supported by grants from IRSIA (Belgium) and NIH (NIAID-AI-27136-01A1). B.C. is a fellow of the Belgian Fonds National de la Recherche Scientifique (FNRS) and is supported by an FNRS-Hippocrate International grant. M.V. is a Senior Research Assistant of the Belgian Fonds National de la Recherche Scientifique (FNRS).

REFERENCES

1. **Brauner, J. W., R. Mendelsohn, and F. G. Prendergast.** 1987. Attenuated total reflectance Fourier transform infrared studies of the interaction of mellitin, two fragments of mellitin, and δ-hemolysin with phosphatidylcholines. Biochemistry **26**:8151–8158.
2. **Byler, D. M., and H. Susi.** 1986. Calculation of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers **25**:469–487.
3. **Cabiaux, V., R. Brasseur, R. Wattiez, P. Falmagne, J.-M. Ruysschaert, and E. Goormaghtigh.** 1989. Secondary structure of diphtheria toxin and its fragments interacting with acidic liposomes studied by polarized infrared spectroscopy. J. Biol. Chem. **264**:4928–4938.
4. **Castresana, J., A. Muga, and J. L. Arrondo.** 1988. The structure of proteins in aqueous solutions: an assessment of triose phosphate isomerase structure by Fourier-transform infrared spectroscopy. Biochem. Biophys. Res. Commun. **152**:69–75.
5. **Dubay, J., L. Kong, J. Kappes, G. Shaw, B. Hahn, and E. Hunter.** 1988. Mutational analysis of the gp41 glycoprotein. IVth Int. Conf. AIDS, Stockholm, abstr. 1517.
6. **Eisenberg, D., and M. Wesson.** 1990. The most highly amphiphilic α-helices include two amino acid segments in human immunodeficiency virus glycoprotein 41. Biopolymers **29**:171–177.
7. **Eisenberg, D., M. Wesson, and M. Yamashita.** 1988. Applications of hydrophobic moment and atomic solvation parameters (ASPs) to understanding protein folding and ligand binding, p. 8. Am. Assoc. Adv. Sci. Annu. Meet., Boston, Mass., 12–14 February, 1988.
8. **Freed, E. O., D. J. Myers, and R. Risser.** 1990. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. Proc. Natl. Acad. Sci. USA **87**:4650–4654.
9. **Freed, E. O., and R. Risser.** 1990. The role of the HIV envelope glycoproteins in cell fusion and the pathogenesis of AIDS. Bull. Inst. Pasteur **88**:73–110.
10. **Fringeli, U. P., and H. H. Günthard.** 1981. Infrared membrane spectroscopy, p. 270–332. *In* E. Grell (ed.), Membrane spectroscopy. Springer-Verlag, New York.
11. **Gallaher, W. R., J. M. Ball, R. F. Garry, M. C. Griffin, and R. C. Montelaro.** 1989. A general model for the transmembrane proteins of HIV and other retroviruses. AIDS Res. Hum. Retroviruses **5**:431–440.
12. **Goormaghtigh, E., V. Cabiaux, and J. M. Ruysschaert.** 1990. Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. Eur. J. Biochem. **193**:409–420.
13. **Goormaghtigh, E., J. De Meutter, F. Szoka, V. Cabiaux, R. A. Parente, and J.-M. Ruysschaert.** 1991. Secondary structure and orientation of the amphipathic peptide GALA in lipid structures. Eur. J. Biochem. **195**:421–429.
14. **Goormaghtigh, E., and J.-M. Ruysschaert.** 1990. Polarized attenuated total reflection infrared spectroscopy as a tool to investigate the conformation and orientation of membrane components, p. 285–329. *In* R. Brasseur (ed.), Molecular description of biological components, vol. 1. CRC Press, Inc., Boca Raton, Fla.
15. **Goormaghtigh, E., L. Vigneron, M. Kniehbler, C. Lazdunski, and J.-M. Ruysschaert.** 1991. Secondary structure of the membrane-bound form of the pore-forming domain of colicin A. Eur. J. Biochem. **202**:1299–1305.
16. **Gorga, J. C., A. Dong, M. C. Manning, R. W. Woody, W. S. Caughey, and J. L. Strominger.** 1989. Comparison of the secondary structures of human class I and class II major histocompatibility complex antigens by Fourier transform infrared and circular dichroism spectroscopy. Proc. Natl. Acad. Sci. USA **86**:2321–2325.
17. **Haffar, O. K., D. J. Dowenko, and P. Berman.** 1988. Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. J. Cell Biol. **107**:1677–1687.
18. **Helseth, E., U. Olshevsky, C. Furman, and J. Sodrosky.** 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. J. Virol. **65**:2119–2123.
19. **Kauppinen, J. K., D. J. Moffat, D. G. Cameron, and H. H. Mantsch.** 1981. Noise in Fourier self-deconvolution. Appl. Optics **20**:1866–1879.
20. **Kowalski, M., J. Potz, L. Basiripour, T. Dorfman, W. L. Gho, E. Terwilliger, A. Dayton, L. Rosen, W. Haseltine, and J. Sodroski.** 1987. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. Science **237**:1351–1355.
21. **Krimm, S., and J. Bandekar.** 1986. Vibrational spectroscopy and conformational of peptides, polypeptides and proteins. Adv. Protein Chem. **38**:181–364.
22. **Levitt, M., and J. Greer.** 1977. Secondary structure in globular proteins. J. Mol. Biol. **114**:181–239.
23. **Maddon, P. J., S. M. Molineaux, D. E. Maddon, K. A. Zimmerman, M. Godfrey, F. W. Alt, L. Chess, and R. Axel.** 1987. Structure and expression of mouse T4 genes. Proc. Natl. Acad. Sci. USA **84**:9155–9159.
24. **Martin, I., F. Defrise-Quertain, E. Decroly, T. Saermark, R. Brasseur, M. Vandenbranden, and J.-M. Ruysschaert.** 1993. Orientation and structure of the NH2-terminal domain of HIV-1 gp41 peptide in fused and aggregated liposomes. Biochim. Biophys. Acta **1145**:124–133.
25. **Miller, M. A., R. F. Garry, J. M. Jaynes, and R. C. Montelaro.** 1991. A structural correlation between lentiviruses transmembrane protein and natural cytolytic peptides. AIDS Res. Hum. Retroviruses **7**:511–519.
26. **Olshevsky, U., E. Helseth, C. Furman, J. Li, W. Haseltine, and J. Sodroski.** 1990. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. J. Virol. **64**:5701–5707.
27. **Pollard, S. R., W. Meier, P. Chow, J. J. Rosa, and D. C. Wiley.** 1991. CD4-binding regions of human immunodeficiency virus envelope glycoprotein gp120 defined by proteolytic digestion. Proc. Natl. Acad. Sci. USA **88**:11320–11324.
28. **Rothschild, K. J., R. Sanches, and N. A. Clark.** 1982. Infrared absorption of photoreceptor and purple membranes. Methods Enzymol. **88**:696–715.
29. **Surewicz, W. K., and H. H. Mantsch.** 1988. New insight into protein secondary structure from resolution-enhanced infrared spectra. Biochim. Biophys. Acta **952**:115–130.
30. **Susi, H., and M. Byler.** 1986. Resolution-enhanced Fourier transform infrared spectroscopy of enzymes. Methods Enzymol. **130**:290–311.
31. **Syu, W.-J., J.-H. Huang, M. Essex, and T.-H. Lee.** 1990. The N-terminal region of the HIV envelope glycoprotein gp120 contains potential binding sites for CD4. Proc. Natl. Acad. Sci. USA **87**:3695–3699.
32. **Vandenbussche, G., A. Clercx, M. Clercx, T. Curstedt, J. Johansson, H. Jörnvall, and J.-M. Ruysschaert.** 1992. Secondary

structure and orientation of the surfactant protein SP-B in a lipid environment. A Fourier transform infrared spectroscopy study. Biochemistry **31**:9169–9176.

33. **Venable, R. M., R. W. Pastor, B. R. Brooks, and F. W. Carson.** 1989. Theoretically determined three-dimensional structures for amphipathic segments of the HIV-1 gp41 envelope protein. AIDS Res. Hum. Retroviruses **5**:7–22.
34. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature (London) **289**:366–373.
35. **Young, J. A. T.** 1988. HIV and HLA similarity. Nature (London) **333** :215.