A trimeric structural domain of the HIV-1 transmembrane glycoprotein

Min Lu,[1] Stephen C. Blacklow[1,2] and Peter S. Kim[1]

Infection with HIV-1 is initiated by fusion of cellular and viral membranes. The gp41 subunit of the HIV-1 envelope plays a major role in this process, but the structure of gp41 is unknown. We have identified a stable, proteinase-resistant structure comprising two peptides, N-51 and C-43, derived from a recombinant protein fragment of the gp41 ectodomain. In isolation, N-51 is predominantly aggregated and C-43 is unfolded. When mixed, however, these peptides associate to form a stable, α-helical, discrete trimer of heterodimers. Proteolysis experiments indicate that the relative orientation of the N-51 and C-43 helices in the complex is antiparallel. We propose that N-51 forms an interior, parallel, homotrimeric, coiled-coil core, against which three C-43 helices pack in an antiparallel fashion. We suggest that this α-helical, trimeric complex is the core of the fusion-competent state of the HIV-1 envelope.

[1]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA
[2]Department of Pathology, Brigham & Womens Hospital Boston, Massachusetts 02115, USA

Correspondence should be addressed to P.S.K.

The envelope glycoprotein (Env) of human immunodeficiency virus type 1 (HIV-1) plays a pivotal role in viral infectivity and tropism by mediating the specific, high-affinity attachment of virions to the cell-surface CD4 receptor and the subsequent fusion of viral and cellular membranes[1]. Env is also a major target for the human immune response to HIV-1 infection[2]. The Env protein is synthesized as a protein precursor that is subsequently cleaved to yield two noncovalently associated subunits, gp120 and gp41 (ref. 1). gp120 is a surface glycoprotein responsible for receptor binding[3,4], and gp41 is a transmembrane glycoprotein involved in the membrane-fusion process[5].

Although the functions of the HIV-1 Env protein are becoming better understood, the structure of this molecule is unknown. Some amino acid sequence features of HIV-1 Env bear a striking similarity to influenza haemagglutinin (HA)[6], the best characterized membrane-fusion protein. In both proteins, the ability to mediate membrane fusion is activated by intracellular cleavage of an inactive precursor into two mature glycoproteins[7]. In both HA and Env, the proteolytic cleavage generates a new N terminus on the transmembrane subunit (TM). The hydrophobic amino-acid sequence of this newly exposed N-terminal region is known, in the case of HA, to insert into the target membrane during the fusion process[8,9], and is therefore referred to as the fusion peptide. The hydrophobic N-terminal regions in both TM proteins are followed by sequences containing a 4–3 hydrophobic repeat, predicted to form a coiled coil[10,11].

Many viral fusion proteins are known to undergo a conformational change to become active in mediating membrane fusion. HA, for example, switches from a fusion-inactive structure to a stable fusion-active conformation when exposed to the mildly acidic endosome, leading to exposure of the fusion peptide and virus-cell membrane coalescence[12]. With cell-line adapted isolates of HIV-1, binding of soluble CD4 induces dissociation of gp120 (shedding) from the surface of viral particles[13,14]. Although CD4-induced shedding does not occur with many primary isolates of HIV-1, new antigenic epitopes of gp41 are often exposed upon binding of soluble CD4 (ref. 15). For these reasons, it has been proposed that binding of gp120 to the cellular receptor CD4 results in conformational changes in the HIV-1 Env protein that promote activation of membrane-fusion activity[16–18]. This proposal is further supported by recent observations[19] that soluble CD4 increases the infectivity of primary HIV-1 strains at low concentrations, but inhibits infectivity at higher concentration, presumably as gp120 sheds from the virus (see also ref. 60). No detailed three-dimensional structural information is available to illuminate the nature of this proposed conformational change in gp41.

Here we use protein dissection to identify a proteinase-resistant structure consisting of two peptides derived from a recombinant fragment of the ectodomain (that is, the extraviral portion) of HIV-1 gp41. In isolation, the N-51 peptide, which includes the predicted N-terminal helical region, is predominantly aggregated, and the C-43 peptide, from the C-terminal region, is unfolded. Upon mixing, these two peptides associate preferentially to form a stable, α-helical trimer of heterodimers, each comprising N-51 and C-43. The biophysical properties and sequence characteristics of these peptides lead to a model for the structure of the complex.

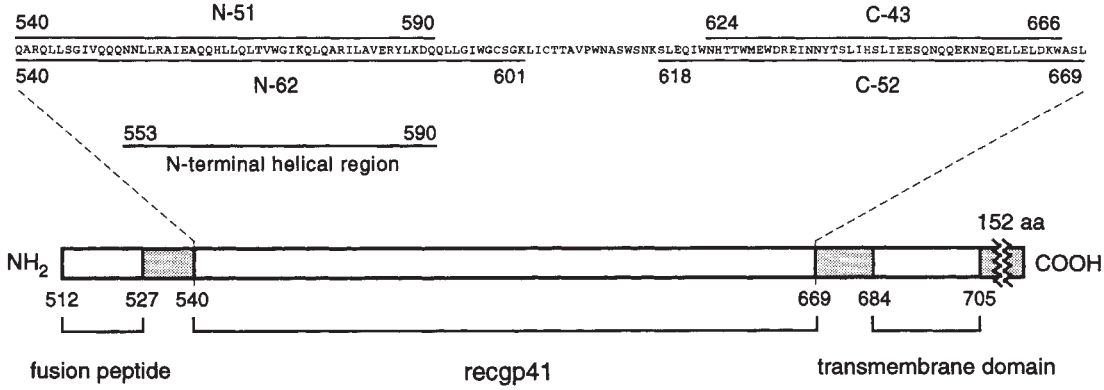

Fig. 1 The gp41 transmembrane protein. Schematic representation of the HIV-1 gp41 protein, illustrating the known functional domains and the recombinant protein fragment of the gp41 ectodomain (recgp41). Expansion above recgp41 shows the amino-acid sequence in single letter code. The highly conserved N-terminal helical region[6,26] is indicated by a bar. The sequences of the proteinase K proteolytic fragments N-51 and C-43, and the tryptic fragments N-62 and C-52 are also indicated.

Proteolysis yields N-51 and C-43

To characterize the structure of the HIV-1 transmembrane protein gp41, we produced recgp41, a bacterially expressed peptide spanning residues 540–669 of gp160 (Fig. 1). This peptide model corresponds to the hydrophilic segment of TM, and lacks the fusion peptide, transmembrane region, and cytoplasmic domain (see Fig. 1). Under physiological conditions, recgp41 forms an insoluble aggregate. However, extensive proteolytic digestion of the recgp41 aggregate with trypsin generates two peptide fragments of similar size (Fig. 2*a*). These fragments correspond to residues 540—601 (N-62) and 618–669 (C-52), respectively (Fig. 1). Digestion of recgp41 with proteinase K produces two shorter peptide fragments: residues 540—590 (N-51) and 624–666 (C-43) (Fig. 1 and 2*b*). The similarity between the tryptic and the proteinase K proteolytic fragments suggests that there is a common proteinase-resistant domain within recgp41. Because proteinase K is a less specific proteinase, the proteolytic fragments N-51 and C-43 may more accurately define the domain structure. Accordingly, we used these peptides for further study.

Specific interaction between N-51 and C-43

The N-51 peptide contains the predicted N-terminal helical region of gp41 (Fig. 1). Although it contains a 4–3 hydrophobic repeat, this sequence is not a typical coiled-coil sequence because hydrophobic residues also occur at the **e** and **g** positions (Fig. 3*a*). Circular dichroism (CD) spectra of N-51 show a strong minimum at 225 nm, which is not typical of α-helical structure (Fig. 3*c*). Moreover, the CD signal of N-51 is concentration-dependent (data not shown). The thermal unfolding transition of N-51 is non-cooperative and irreversible (Fig. 3*d*), and the solution becomes turbid upon heat denaturation. Isolated N-51 is proteinase-resistant (Fig. 4). Collectively, these observations indicate that the N-51 peptide has a strong tendency to aggregate.

The sequence of the C-43 peptide also contains a 4–3 hydrophobic repeat (Fig. 3*b*), which is predicted to

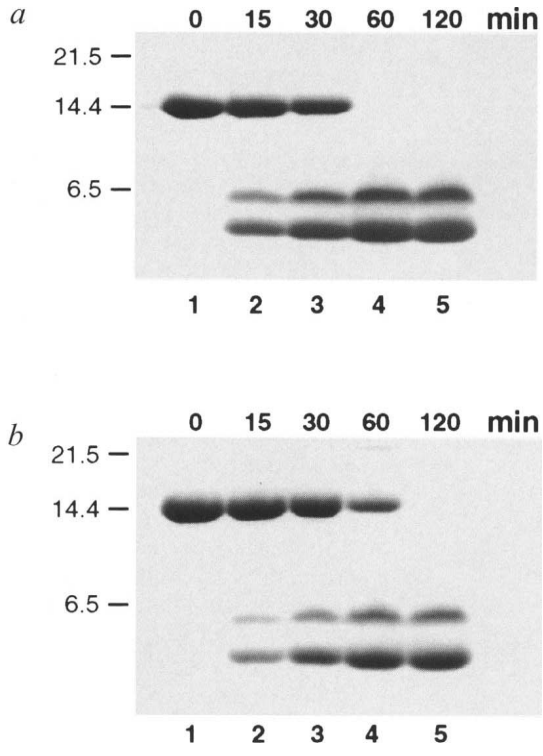

Fig. 2 Proteolysis of the recombinant protein fragment of the gp41 ectodomain. *a*, Trypsin digestion of recgp41 (see Methods). Lanes 1–5, digestion after 0, 15, 30, 60, and 120 min, respectively. The digestion products were analyzed by tricine gel electrophoresis with Coomassie blue staining[59]. The sizes of protein molecular weight markers (in kDa) are indicated at the left. Proteolytic digestion of recgp41 with trypsin produces two major fragments, spanning residues 540–601, plus an N-terminal methionine residue (denoted N-62), and 618–669 (denoted C-52), and a minor fragment, spanning residues 543–601. *b*, Proteinase K digestion of recgp41. The digestion was carried out as described in (*a*). Proteinase K cleaves recgp41 into two fragments, spanning residues 540–590, plus an N-terminal methionine residue (denoted N-51), and 624–666 (denoted C-43).

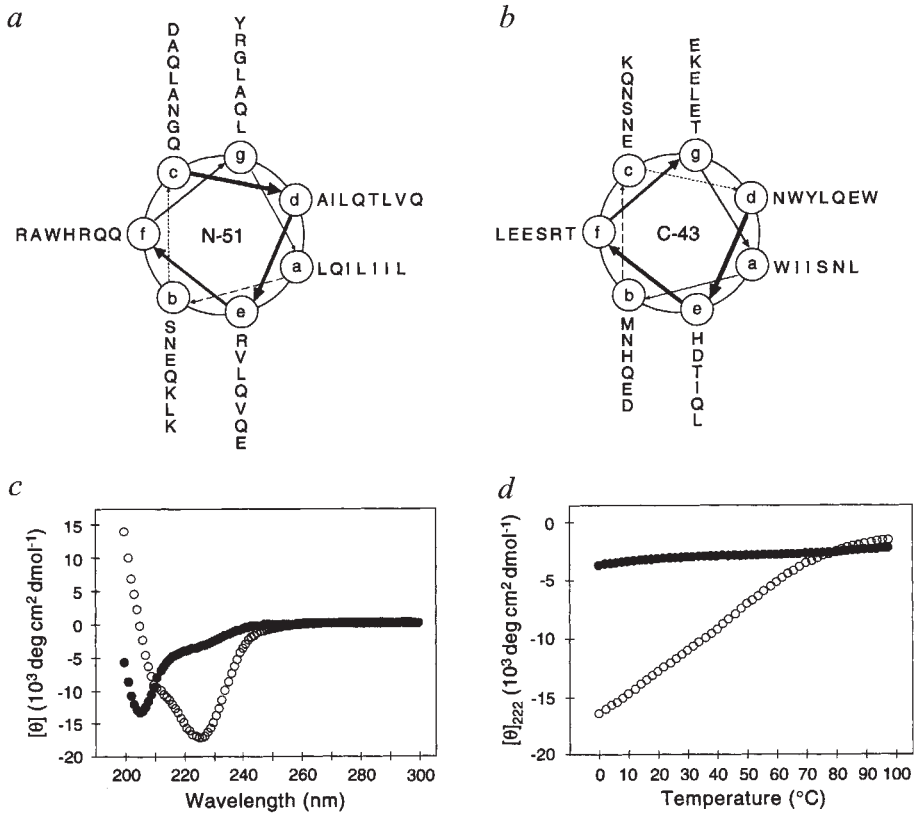

Fig. 3 Lack of a well-defined structure for the isolated N-51 and C-43 peptides. *a*, The sequence of the N-51 peptide is projected onto a helical wheel. The view is from the N terminus. *b*, The sequence of the C-43 peptide is projected onto a helical wheel, as in (a). *c*, The CD spectra for the isolated N-51 (open circles) and C-43 (closed circles) peptides in PBS at 0 °C. *d*, Temperature dependence of the CD signal at 222 nm for N-51 (open circles) and for C-43 (closed circles).

form an isolated coiled coil by the program of Lupas *et al.*[20], but not with the more restrictive algorithm of Berger *et al.*[21]. In agreement with the latter prediction, the isolated C-43 peptide displays little secondary structure, as judged by CD (Fig. 3*c*,*d*). Consistent with the CD results, the isolated peptide is highly susceptible to proteolytic digestion by proteinase K (Fig. 4). The inability of isolated C-43 to form a folded structure (and hence, to be proteinase resistant) suggests that folding (and proteinase resistance) of C-43 is dependent on N-51. Indeed, we find that increasing concentrations of N-51 result in the protection of C-43 from proteolytic digestion, and that maximum proteinase resistance is achieved when the stoichiometry of C-43 to N-51 is 1:1 (Fig. 4).

Trimer of heterodimers

The CD spectrum of an equimolar mixture of the N-51 and C-43 peptides is typical of an α-helix, displaying the characteristic minima at 208 and 222 nm (Fig. 5*a*). The complex is fully helical (Fig. 5*a*) and very stable, unfolding irreversibly with a transition midpoint (apparent T_m) of 90 °C under physiological conditions (Fig. 5*b*). This type of irreversibility is observed in other multimeric proteins, in which aggregation of unfolded molecules interferes with refolding[22]. Sedimentation equilibrium centrifugation of the N-51/C-43 complex at 20 °C indicates definitively that the complex consists of three molecules each of N-51 and C-43: the observed molecular mass is 34,200 M_r, compared to an expected molecular mass of 11,400 M_r for the N-51/C-43 heterodimer (Fig. 5*c*). We conclude that the N-51 and C-43 peptides associate preferentially to form a stable, α-helical, discrete trimer of heterodimers, each comprising N-51 and C-43.

An antiparallel helix orientation

The results of a recent study of the SIV transmembrane protein[23] suggested that the relative orientation of the N-51 and C-43 peptides in the trimeric complex is antiparallel. To test this hypothesis, we made a variant of recgp41, shortened by 12 residues at the N terminus (Fig. 5*d*). Because the N-51/C-43 complex is fully helical, we assume that each peptide is helical. Proteolytic digestion of the recgp41 variant with proteinase K yields two peptide fragments; N-39 (shortened by 12 residues at the N terminus), and C-30 (truncated by 13 residues at the C terminus). These results indicate that the N-51 and C-43 helices are antiparallel to each other in the complex (Fig. 5*d*).

Effect of proline substitutions

To study the functional role of the putative coiled-coil sequence in the N-terminal region of gp41, other workers have introduced single proline substitutions into this region of HIV-1 Env[24,25]. Although these proline substitutions result in Env proteins that are completely defective in mediating membrane fusion, these substitutions have little or no effect on the oligomeric structure of gp120/gp41. Because it is expected that the proline substitutions would disrupt the coiled-coil structure, the phenotype of these mutant viruses has been interpreted to indicate that the N-terminal 4–3 repeat of gp41 does not play a crucial role in Env protein oligomerization[24,25]. Our finding that the trimeric N-51/C-43 complex is highly stable under physiological conditions suggests that the N-terminal helical region of gp41 is involved in Env oligomerization after all, but that single proline substitutions are not necessarily sufficient to disrupt the entire trimeric complex structure. To test this hypothesis, we substituted Leu 566 and Ile 573 of recgp41 with proline[24,25] to yield two mutant recombinant fragments: L566P and I573P.

Digestion of either mutant recgp41 protein with trypsin yields the same N-62 and C-52 tryptic fragments as does wild-type recgp41 (see Fig. 1). Unlike wild-type recgp41, however, both mutants are susceptible to proteinase K digestion, even under mild diges-

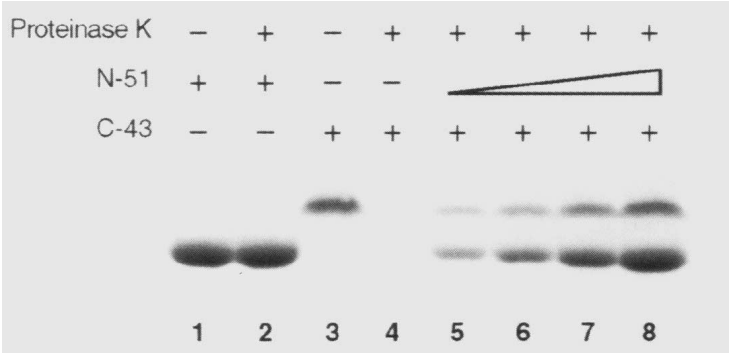

Fig. 4 Protection of C-43 against proteolysis by addition of N-51. Lanes 1, 2: N-51 (40 μM final concentration) before and after incubation with proteinase K, respectively (see Methods). Lanes 3, 4: C-43 (40 μM) before and after, respectively. Lanes 5-8, digestion of C-43 (40 μM) preincubated with 5, 10, 20, and 40 μM of N-51, respectively. The digestion products were analysed as described in Fig. 2*a*.

tion conditions. CD spectra indicate that the trypsin-derived complexes from the proline mutants are less helical than the wild-type complex (approximately 89% helix content for wild-type, 70% for L566P, and 79% for I573P) (Fig. 6*a*). The structure in the mutant complexes also unfolds with lower apparent T_m values (93 °C for wild-type, 74 °C for L566P, and 77 °C for I573P) (Fig. 6*b*). Sedimentation equilibrium experiments indicate that both the wild-type and proline mutant N-62/C-52 complexes are trimeric (Fig. 6*c*). We conclude that the N-terminal helical region is likely to be involved in Env oligomerization and that the previously studied proline mutations in Env do not alter structural features sufficiently to disrupt trimerization. The destabilization effect of the proline mutations observed here may be related to the fusion-defective phenotype[24,25].

Peptide inhibition of syncytium formation

Synthetic peptides corresponding to the N-terminal helical or C-terminal regions of the gp41 ectodomain effectively block HIV-1-induced membrane fusion[26–28]. Our finding that the N-terminal and C-terminal portions of the gp41 ectodomain associate preferentially into a very stable trimeric complex suggests that these peptide inhibitors function by forming complexes with viral gp41. We therefore examined the relative abilities of N-51, C-43, and their complex to inhibit syncytium formation between cells expressing HIV-1 Env and CD4-expressing target cells (see Methods). Fig. 7 shows the inhibition of syncytium formation by N-51, C-43, and their equimolar mixture. The C-43 peptide is the most effective inhibitor, with a 90% inhibition concentration (IC_{90}) of 0.14 μM, whereas the N-51 peptide has an IC_{90} of 1.3 μM. When the two peptides are mixed, the complex inhibits syncytium formation with an IC_{90} of 1.5 μM. Thus, N-51 masks the potent inhibitory activity of C-43, suggesting that C-43 inhibits fusion by interaction with the N-terminal region of viral gp41.

A model for the N-51/C-43 complex

By proteolytic digestion of a recombinant protein fragment, we have identified a stable and proteinase-resistant structure composed of two peptides, N-51 and C-43, that correspond to the N-terminal and C-terminal regions of the HIV-1 gp41 ectodomain. It has previously been proposed that a sequence from the

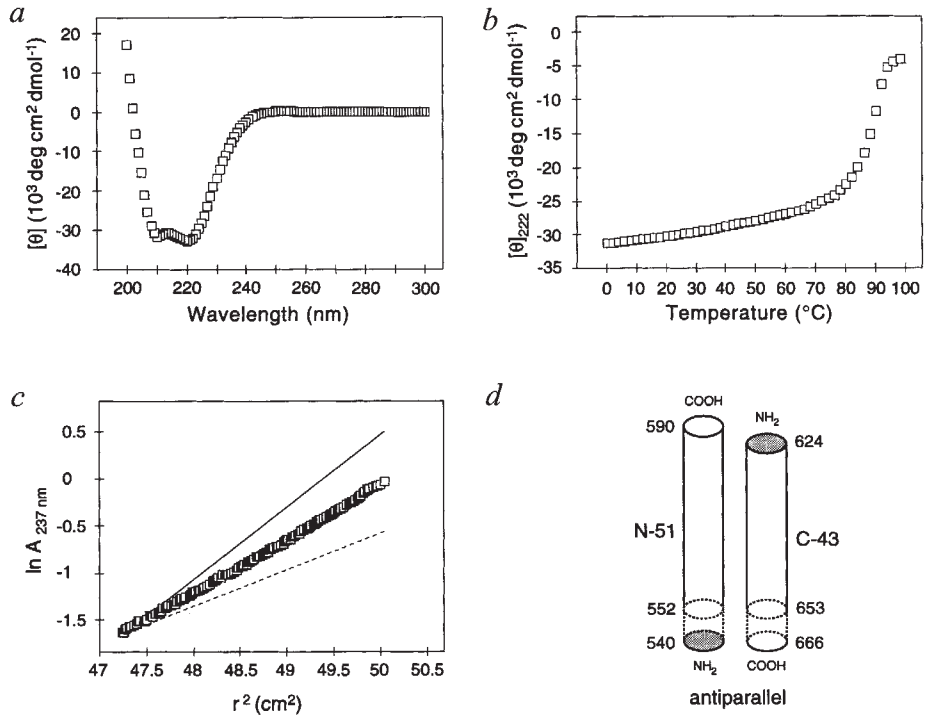

Fig. 5 Folding of the N-51/C-43 complex as a helical trimer of heterodimers. *a*, The CD spectrum of a mixture of N-51 (10 μM) and C-43 (10 μM) in PBS at 0 °C. *b*, The CD signal at 222 nm for the N-51/C-43 complex as a function of temperature. *c*, The molecular mass of the N-51/C-43 complex, as determined by sedimentation equilibrium centrifugation (see Methods). The slope of the plotted data indicates that the N-51/C-43 complex is trimeric. The calculated data for dimeric and tetrameric models are indicated by dashed and solid lines, respectively. *d*, The proteinase K digestion products of recgp41 and its variant form with the twelve-residue N-terminal truncation are shown. Proteolytic digestion of the recgp41 variant with proteinase K produces two fragments with masses of 4749 and 3717 (determined by laser desorption mass spectrometry), corresponding to N-39, spanning residues 552-590, plus an N-terminal methionine residue, and C-30, spanning 624-653, respectively. These results indicate that the N-51 and C-43 peptides associate antiparallel to each other in the complex.

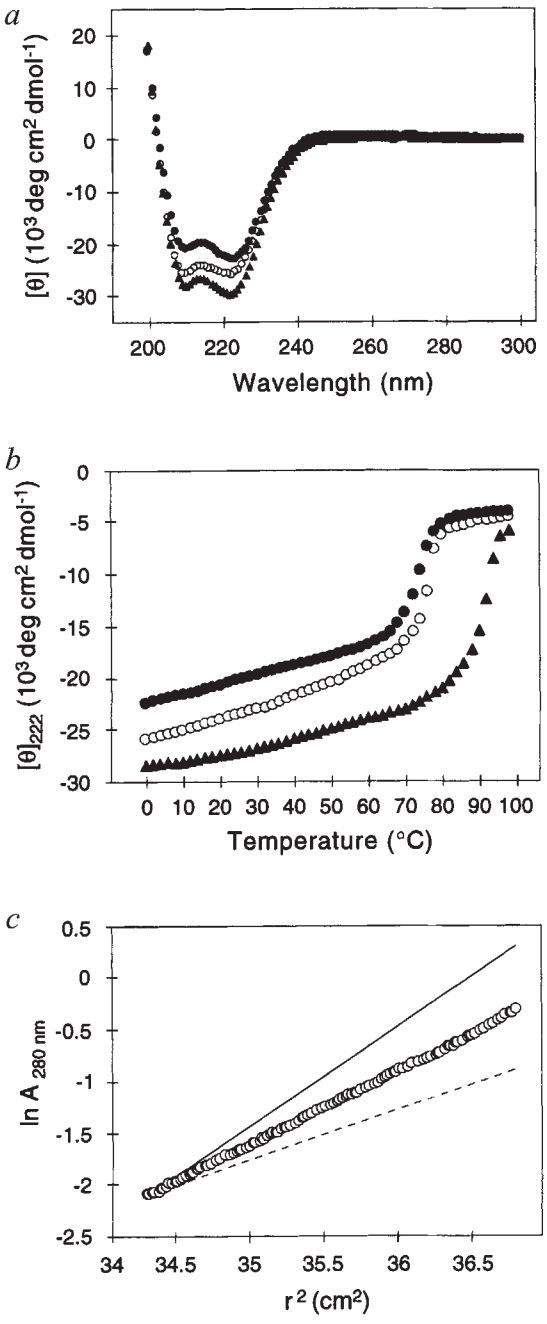

Fig. 6 Effect of proline substitutions on complex formation. *a*,The CD spectra of a mixture of N-62 (10 μM) and C-52 (10 μM) in PBS at 0 °C for wild type (closed triangles), L566P (closed circles), and I573P (open circles). *b*, Temperature dependence of the CD signal at 222 nm for the N-62/C-52 complex for wild type (closed triangles), L566P (closed circles), and I573P (open circles). *c*, The molecular mass of the N-62/C-52 complex for I573P, as determined by sedimentation equilibrium experiments. Data for I573P are consistent with a trimeric model (observed molecular weight of 40,600 M_r). The calculated data for dimeric and tetrameric models are indicated by dashed and solid lines, respectively. The data for wild type and L566P are also consistent with a trimeric model (the observed molecular weight of the N-62/C-52 complex is 42,600 M_r for wild type, and 40,100 M_r for L566P).

N-terminal region of gp41 may form a coiled coil[10,11]. In addition, a synthetic peptide of 38 residues within N-51, corresponding to this predicted helical region (Fig. 1), is helical and self-associating[26]. This finding has been interpreted to indicate the formation of an isolated coiled coil; however, the oligomerization state of this 38-residue peptide is not well-defined[26]. The N-51 peptide studied here is not helical (Fig. 3*c*) but instead forms an aggregate, which is consistent with the failure of coiled-coil prediction algorithms to identify the N-51 sequence as likely to form a coiled coil[20,21]. The C-43 peptide is predicted to form an isolated coiled coil by the algorithm of Lupas *et al.*[20], but not by that of Berger *et al.*[21]. Our results show that isolated C-43 is unstructured. Thus, neither N-51 nor C-43 forms a well-defined structure in isolation. Strikingly, however, when these peptides are mixed, they associate preferentially and fold autonomously as a stable, α-helical, discrete trimer of heterodimers.

What is the nature of the trimeric complex? Our results indicate that the N-51/C-43 complex is fully helical and that the N-51 and C-43 helices are oriented antiparallel to each other in the complex. Although neither peptide is a coiled coil by itself, the sequence features of these peptides indicate that they could form amphiphilic helices. Traditional coiled-coil proteins have a characteristic heptad repeat, $(a.b.c.d.e.f.g)_n$, with hydrophobic residues at positions a and d and predominantly polar residues elsewhere[29,30]. Residues at the a and d positions pack against each other to form the hydrophobic interface in two-, three- and four-stranded coiled coils[31-34]. The sequences of both N-51 and C-43 show a 4–3 hydrophobic repeat pattern at positions a and d (Fig. 3), although neither sequence is predicted to form an isolated coiled coil by the algorithm of Berger *et al.*[21]. The N-51 sequence also contains a preponderance of hydrophobic residues at the e and g positions (Fig. 3*a*). These sequence characteristics, taken together with the biophysical properties of the complex and the previous observations of Wild *et al.*[26], lead us to propose that N-51 forms an interior, parallel, homotrimeric coiled-coil core, in which the e and g positions of adjacent N-51 strands pack against the a and d positions of each C-43 strand to bury the hydrophobic residues of each monomer (Fig. 8). This model is strongly supported by similar findings upon protein dissection of the SIV gp41 ectodomain[23].

HIV-1 inhibition by peptides

There are known synthetic peptide inhibitors of HIV-1 syncytium formation and infectivity that overlap with the N-51 and C-43 sequences[26–28]. Although much is known about the relationships between sequence and activity for these peptide inhibitors[26,28,35–37], the mechanism of inhibition has not been established. Our finding that the N-terminal and C-terminal portions of the gp41 ectodomain associate preferentially into a stable complex suggests strongly that each synthetic peptide inhibits membrane fusion by associating with its endogenous partner within HIV-1 Env, in a dominant-negative manner[38], to disrupt gp41-mediated fusion. This hypothesis is consistent with our observa-

tion that inhibition of syncytium formation by C-43 is markedly reduced when stoichiometric amounts of N-51 are also present. We also suggest that the tendency of N-51 to aggregate in the absence of C-43 may account for the poor inhibitory activity of this peptide, if the aggregates of N-51 are less effective on a molar basis than the monomers. Finally, although it is possible that the inhibitory activity of the N-51/C-43 complex is due to a small fraction of dissociated C-43, it seems likely that an additional mechanism is involved in inhibition of syncytium formation by the N-51/C-43 complex.

Recent studies of a gp41 ectodomain maltose-binding chimeric protein[39] complement our protein dissection studies. When this chimeric protein contains a proline mutation in the N-terminal helical region of gp41, the resulting mutant is a potent inhibitor of HIV-1 fusion and infectivity, although the wild-type protein and a truncated chimeric protein lacking the putative C-terminal helical region do not have inhibitory activity[39]. These results were interpreted to suggest that the proline mutation in the chimeric protein disrupts the association between the C-terminal sequence and the N-terminal region, thereby rendering the C-terminal sequence available for interaction with its target site[39]. It is not clear, however, why the truncated chimeric protein, in contrast to the isolated N-51 peptide studied here, does not have inhibitory activity.

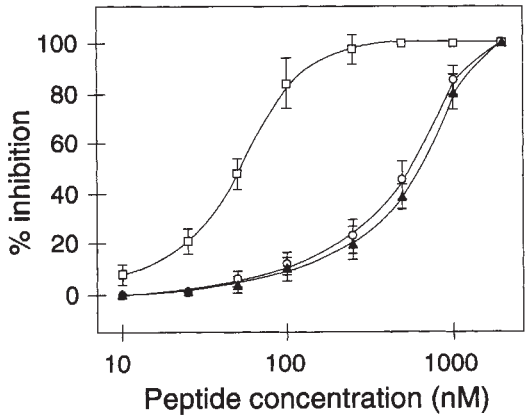

Fig. 7 Inhibition of syncytium formation. Inhibition of syncytium formation between cells expressing HIV-1 Env and CD4-expressing target cells by N-51 (open circles), C-43 (open squares), and an equimolar mixture of N-51 and C-43 (closed triangles), determined as described in Methods. Standard deviations of the mean of quadruplicate samples are indicated by vertical bars.

Oligomeric structure of Env

Attempts to define the oligomeric state of HIV-1 Env have yielded conflicting results. Electron microscopy, chemical cross-linking, and velocity sedimentation studies suggest that gp120/gp41 is likely to exist as either a trimer[40,41] or a tetramer[42–45]. The highly conserved 4–3 hydrophobic repeat in the N-terminal portion of gp41 has been predicted to play a role in oligomer formation and stability[6,10,24]. Our results show that this N-terminal region interacts specifically with the C-terminal region of the gp41 ectodomain to fold as a stable, trimeric domain in isolation. A recent study of the ectodomain of the retrovirus envelope protein from Moloney murine leukemia virus also reveals a trimeric oligomeric structure[46].

Mutational analyses of gp41 have demonstrated that substituting proline for isoleucine or leucine within the N-terminal heptad repeat of gp41 abolishes Env membrane fusion activity, but does not interfere with Env oligomerization[24,25]. This observation has led to the conclusion that the putative coiled-coil domain is required for virus entry and/or fusion but is not necessary for protein oligomerization[24,25]. Our results indicate that the N-51 region is likely to be a crucial component of the oligomerization domain. Our results also indicate clearly that proline substitutions in the N-51 peptide do not prevent trimerization of the α-helical complex, although small, significant structural alterations are observed, as judged by CD. Thus, the phenotype of the mutant Env proteins with single proline substitutions can be reconciled with the suggestion that

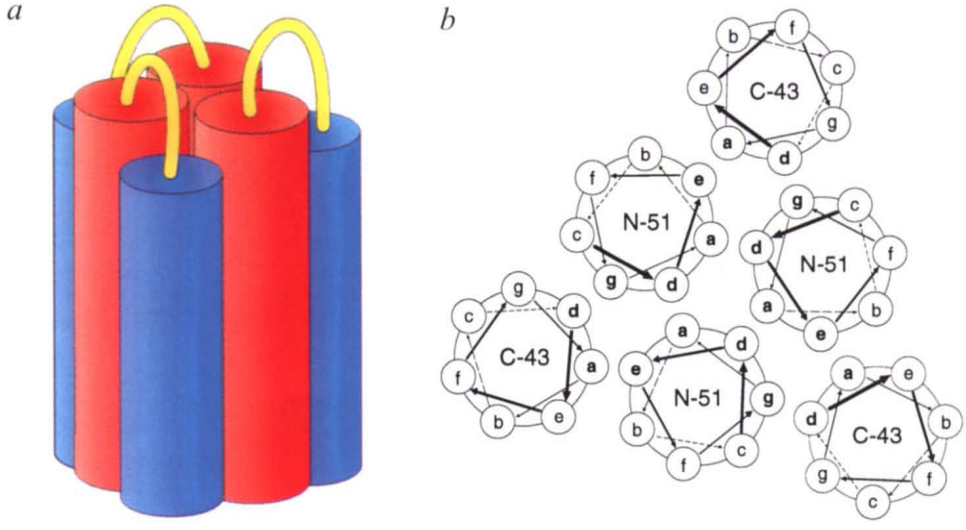

Fig. 8 Model for the putative, α-helical trimer of heterodimers formed by the N-51 and C-43 peptides in the HIV-1 transmembrane protein gp41. *a*, Schematic diagram depicting the N-51 peptide in the center as a parallel, trimeric coiled-coil-like structure, against which are packed three C-43 peptide helices arranged in an antiparallel fashion (see text). *b*, Cross-section of the trimeric structure in (*a*), depicting possible hydrophobic interactions between amino acids at positions **a** and **d** of adjacent N-51 helices, and between amino acids at positions **a** and **d** of single C-43 helices and positions **e** and **g** of adjacent N-51 helices. The helical wheel view is from the top, as depicted in (*a*).

the proline substitutions do not disrupt the entire oligomerization domain, but do introduce structural perturbations which abolish the ability of the Env protein to mediate membrane fusion.

Conformational changes in Env

Flu HA, the best characterized membrane-fusion protein, is known to undergo a dramatic conformational change in order to become fusion-active. Based on the observation that an extended loop region in the native conformation of HA can form a stable, trimeric coiled coil as an isolated peptide, a spring-loaded model for the conformational change of HA was proposed, in which the loop region of the native state is transformed into a coiled coil at lower pH to project the fusion peptide toward the target cell membrane[18,47]. The crystal structure of a large fragment of the acid-induced state of HA confirms many aspects of this proposed model[48].

It has been proposed that the CD4-induced release of gp120 is accompanied by a conformational change from the native to the fusogenic state in HIV-1 Env[16–18]; however, the nature of this conformational change is not yet defined. Because the α-helical, trimeric complex consisting of the N-51 and C-43 peptides (Fig. 8) is a highly stable structure in the absence of gp120, we speculate that this structure represents part of the fusion-competent state of gp41, similar to the low-pH induced structure of HA[18,48]. Our results do not necessarily rule out the possibility that this structure is also found in the native state of gp120/gp41. In either case, the N-terminal sequence is one of the most highly conserved in HIV-1 Env, which otherwise exhibits considerable genetic diversity, even among closely related isolates. It is likely that the coiled-coil core formed by these residues adjacent to the fusion peptide, as in HA, is crucial for HIV-1-mediated membrane fusion.

Methods

Cloning and mutagenesis. Plasmid pHIVhxb2cg, containing the HIV-1 *env* gene from the HXB2 isolate, was kindly provided by D. Sanders and R.C. Mulligan (Whitehead Institute). The 2.7 kb fragment from pHIVhxb2cg was used as a template to amplify the gene coding for residues 540—669 by PCR, with primers encoding a 5' NdeI site and a 3' BamHI site. The amplified 415 bp fragment was subsequently cloned into the NdeI-BamHI sites of a phagemid-T7 expression vector, pAED4 (ref. 49), to yield p41–130. Proline substitutions were introduced individually into p41–130 by oligonucleotide-directed mutagenesis[50]. Standard recombinant DNA techniques were used[51].

Protein expression and purification. *Escherichia coli* BL21(DE3) cells, freshly transformed with an appropriate plasmid, were grown to late log phase. Protein expression was induced by addition of 1 mM IPTG. After another 6 h of growth at 37 °C, the bacteria were harvested by centrifugation, and the pellets were stored frozen. Inclusion bodies were isolated as described[52]. The inclusion bodies were resuspended in 50 mM Tris, pH 8.5, 10 mM NaCl, 8 M urea, 1 mM EDTA, and 0.1 M DTT. The supernatant fraction was then loaded onto a DEAE-Sepharose column, pre-equilibrated with 20 mM Tris, pH 8.5, 10 mM NaCl, 4 M urea, 1 mM EDTA, and 5 mM DTT, and protein was eluted with a linear gradient from 50–1000 mM NaCl. Fractions containing partially purified protein were pooled and dialyzed against 5% acetic acid. After dialysis, the protein was purified to homogeneity by reversed-phase HPLC, using a Vydac C-18 preparative column and a linear gradient of acetonitrile containing 0.1% trifluoracetic acid. The purified protein was subjected to air oxidation in 6 M GuHCl, and the oxidized protein was repurified by HPLC. The identity of each HPLC-purified protein was confirmed by laser desorption mass spectrometry (PerSeptive Biosystems). All observed masses were within 4 AMU of the expected molecular masses.

Proteolysis experiments. For a typical digestion, 2 μg of protein were incubated with 0.2 μg of TLCK-treated trypsin (Sigma) or 0.2 μg of proteinase K (Boehringer Mannheim) at 37 °C for 2 hr in 10 μl of 50 mM Tris-HCl, pH 7.5. Proteolysis was quenched by addition of phenylmethylsulphonyl fluoride (PMSF, Sigma) to a final concentration of 2 mM. The proteolytic fragments were separated and purified by HPLC as described above, and characterized by N-terminal sequencing and mass spectrometry. For proteolysis protection experiments, samples were incubated with 1% proteinase K (wt/wt, proteinase K/peptide) at 37 °C for 2 h in PBS buffer (50 mM sodium phosphate, pH 7, 150 mM NaCl).

CD spectroscopy. CD spectra were recorded on an AVIV Model 62DS CD spectrometer equipped with a thermoelectric temperature controller. The cuvettes used for wavelength spectra and for thermal unfolding studies were 0.1 cm and 1 cm in pathlength, respectively. Peptide concentrations were determined by absorbance at 280 nm in 6 M GuHCl[53]. Each CD sample contained 10 μM peptide in PBS buffer. The CD signal at 222 nm was measured as a function of temperature, and thermal melts were performed in 2 °C steps, with 2 min of equilibration at each temperature, and an acquisition time of 0.5 min. None of the melting curves is reversible. The apparent melting temperature (T_m) was estimated from the maximum of the first derivative, with respect to T^{-1} (Kelvin), of the CD signal at 222 nm (ref. 54). The value of $[\theta]_{222}$ for 100% helicity is estimated to be -33,000 deg cm^2 dmol^{-1} (ref. 55).

Analytical ultracentrifugation. Analytical ultracentrifugation was performed at 20 °C on a Beckman XL-A Optima analytical ultracentrifuge at rotor speeds of 15,000 and 18,000 rpm. Three samples were analysed at heterodimer concentrations of 5 μM, 20 μM and 50 μM in PBS. Samples were dialyzed against the reference buffer (PBS) for 24 h. Data were fit to a single ideal species model. No systematic deviation of the residuals was observed in any concentration set. Specific volumes and solvent densities were calculated as described[56].

Cell culture and fusion assay. CHO[HIVe] (clone 7d2) cells[57] were a generous gift from M. Krieger and MT-2 cells[58] were obtained through the NIH AIDS Research and Reference Reagent Program. Cells were maintained in either (i) Hams F-12 containing 100 U/ml penicillin, 100 μg/ml streptomycin, and 2 mM glutamine (medium A), supplemented with 5% fetal bovine serum (FBS) and 500 μg/ml G418 (CHO[HIVe]); or (ii) RPMI 1640 containing 100 U/ml penicillin, 100 μg/ml streptomycin, and 2 mM glutamine supplemented with 10% FBS (MT-2). CHO[HIVe] cells were plated at 1×10^4 cells well^{-1} in a 96-well dish in medium A plus 5% FBS, and were grown for 48 h. Then, 1×10^5 MT-2 cells were added in the presence and absence of peptide inhibitors. After 24 h incubation at 37 °C, syncytia were counted by microscopic examination at 40× magnification.

Received 7 September; accepted 26 October 1995.

Acknowledgements
We thank P. Harbury and C. Carr for helpful discussions, M. Krieger, S. Acton and C. Doetsch for advice in cell culture, and R. Rutkowski, L. Wu and members of the Kim lab for critical reading and comments on the manuscript. S.C.B. is a recipient of a Howard Hughes Medical Institute Postdoctoral Fellowship for Physicians. This research was supported by the Howard Hughes Medical Institute.

1. Bour, S., Geleziunas, R. & Wainberg, M.A. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. *Micriobiol. Rev.* **59**, 63–93 (1995).
2. Nara, P.L., Garrity, R.R. & Goudsmit, J. Neutralization of HIV-1: a paradox of humoral proportions. *FASEB J.* **5**, 2437–2455 (1991).
3. Klatzmann, D. *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).
4. McDougal, J.S. *et al.* Binding of the HTLV-III/LAV to T4+ T cells by a complex of the 100 K viral protein and the T4 molecule. *Science* **231**, 382–385 (1986).
5. Gallaher, W.R. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* **50**, 327–328 (1987).
6. Gallaher, W. R., Ball, J. M., Garry, R. F., Griffin, M. C. & Montelaro, R. C. A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res. Hum. Retroviruses* **5**, 431-440 (1989).
7. White, J. M. Membrane fusion. *Science* **258**, 917–924 (1992).
8. Stegmann, T., Delfino, J.M., Richards, F.M. & Helenius, A. The HA2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion. *J. biol. Chem.* **266**, 18404–18410 (1991).
9. Tsurudome, M., *et al.* Lipid interactions of the hemagglutinin HA2 NH_2-terminal segment during influenza virus-induced membrane fusion. *J. biol. Chem.* **267**, 20225–20232 (1992).
10. Delwart, E.J., Mosialos, G. & Gilmore, T. Retroviral envelope glycoproteins contain a leucine zipper like repeat. *AIDS Res. Hum. Retroviruses* **6**, 703–706 (1990).
11. Chambers, P., Pringle, C.R. & Easton, A.J. Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J. Gen. Virol.* **71**, 3075–3080 (1990).
12. Hughson, F. M. Structural characterization of viral fusion proteins. *Cur. Biol.* **5**, 265–274 (1995).
13. Moore, J.P., McKeating, J.A., Weiss, R.A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250**, 1139–1142 (1990).
14. Hart, T.K. *et al.* Binding of soluble CD4 proteins to human immunodeficiency virus type 1 and infected cells induces release of envelope glycoprotein gp120. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2189–2193 (1991).
15. Sattentau, Q.J. & Moore, J.P. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.* **174**, 407–415 (1991).
16. Allan, J.S. Receptor-mediated activation of immunodeficiency viruses in viral fusion. *Science* **252**, 1322–1323 (1991).
17. Moore, J.P., McKeating, J.A., Weiss, R.A., Clapham, P.R. & Sattentau, Q.J. Receptor-mediated activation of immunodeficiency viruses in viral fusion. *Science* **252**, 1322–1323 (1991).
18. Carr, C.M. & Kim, P.S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823–832 (1993).
19. Sullivan, N., Sun, Y., Li, J., Hofmann, W. & Sodroski, J. Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T-cell line-passaged human immunodeficiency virus type 1 isolates. *J. Virol.* **69**, 4413–4422 (1985).
20. Lupas, A., VanDyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
21. Berger, B.A., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M. & Kim, P.S. Predicting coiled coils using pairwise residue correlations. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8259–8263 (1995).
22. Jaenicke, R. & Rudolph, R. Refolding and association of oligomeric proteins. *Meth. Enzymol.* **131**, 218–250 (1986).
23. Blacklow, S.C., Lu, M. & Kim, P.S. A trimeric subdomain of the simian immunodeficiency virus envelope glycoprotein. *Biochemistry* **in the press**.
24. Dubay, J.W., Roberts, S.J., Brody, B. & Hunter, E. Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. *J. Virol.* **66**, 4748–4756 (1992).
25. Chen, S.S., Lee, C.N., Lee, W.R., McIntosh, K. & Lee, T.H. Mutational analysis of the leucine zipper-like motif of the human immunodeficiency virus type 1 envelope transmembrane glycoprotein. *J. Virol.* **67**, 3615–3619 (1993).
26. Wild, C.T., Oas, T., McDanal, C.B., Bolognesi, D. & Matthews, T.J. A synthetic peptide inhibitor of human immunodeficiency virus replication: correlation between solution structure and viral inhibition. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10537–10541 (1992).
27. Jiang, S., Lin, K., Strick, N. & Neurath, A.R. HIV-1 inhibition by a peptide. *Nature* **365**, 113 (1993).
28. Wild, C.T., Shugars, D.C., Greenwell, T.K. McDanal, C.B. & Matthews, T. J. Peptides corresponding to a predictive α-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9770–9774 (1994).
29. Hodges, R.S., Sodek, J., Smillie, L.B. & Jurasek, L. Tropomyosin: amino acid sequence and coiled-coil structure. *Cold Spring Harb. Symp. Quant. Biol.* **37**, 299–310 (1972).
30. McLachlan, A.D. & Stewart, M. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. molec. Biol.* **98**, 293–304 (1975).
31. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel, coiled coil. *Science* **254**, 539–544 (1991).
32. Wilson, I.A., Skehel, J.J. & Wiley, D.C. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
33. Harbury, P.B., Zhang, T., Kim, P.S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).
34. Harbury, P.B., Kim, P.S. & Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83 (1994).
35. Jiang, S., Lin, K., Strick, N. & Neurath, A. R. Inhibition of HIV-1 infection by a fusion domain binding peptide from the HIV-1 envelope glycoprotein gp41. *Biochem. Biophys. Res. Commun.* **195**, 533–538 (1993).
36. Neurath, A.R., Lin, K., Strick, N. & Jiang, S. Two partially overlapping antiviral peptides from the external portion of HIV type 1 glycoprotein 41, adjoining the transmembrane region, affect the glycoprotein 41 fusion domain. *AIDS Res. Hum. Retroviruses* **11**, 189–190 (1995).
37. Wild, C.T., Greenwell, T., Shugars, D., Rimsky-Clarke, L. & Matthews, T. The inhibitory activity of an HIV-1 type peptide correlates with its ability to interact with a leucine zipper structure. *AIDS Res. Hum. Retroviruses* **11**, 323–325 (1995).
38. Herskowitz, I. Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219–222 (1987).
39. Chen, C.H., Matthews, T.J., McDanal, C.B., Bolognesi, D.P. & Greenberg, M.L. A molecular clasp in the human immunodeficiency virus (HIV) type 1 TM protein determines the anti-HIV activity of gp41 derivatives: implication for viral fusion. *J. Virol.* **69**, 3771–3777 (1995).
40. Gelderblom, H.R. *et al.* Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation. *Micron. Microsc. Acta* **19**, 41–60 (1988).
41. Weiss, C.D., Levy, J.A. & White, J.M. Oligomeric organization on infectious human immunodeficiency virus type 1 particles. *J. Virol.* **64**, 5674–5677 (1990).
42. Schawaller, M., Smith, G.E., Skehel, J.J. & Wiley, D.C. Studies with crosslinking reagents on the oligomeric structure of the *env* glycoprotein of HIV. *Virology* **172**, 367–369 (1989).
43. Pinter, A. *et al.* Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. *J. Virol.* **63**, 267–279 (1989).
44. Earl, P.L., Doms, R.W. & Moss, B. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proc. natn. Acad. Sci. U.S.A.* **87**, 648–652 (1990).
45. Thomas, D.J. *et al.* gp160, the envelope glycoprotein of human immunodeficiency virus type 1, is a dimer of 125-kilodalton subunits stabilized through interactions between their gp41 domains. *J. Virol.* **65**, 3797–3803 (1991).
46. Fass, D. & Kim, P.S. Dissection of a retrovirus envelope reveals structural similarity to influenza hemagglutinin. *Curr. Biol.* **in the press**.
47. Carr, C.M. & Kim, P.S. Flu virus invasion: halfway there. *Science* **266**, 234–236 (1994).
48. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. Structure of influenza hemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
49. Doering, D. Functional and Structural Studies of a Small f-actin Binding Protein. *PhD thesis*, Massachusetts Institute of Technology, Cambridge, Massachusetts (1992).
50. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meths. Enzymol.* **154**, 367–382 (1987).
51. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).
52. Peng, Z.Y. & Kim, P.S. A protein dissection study of a molten globule. *Biochemistry* **33**, 2136–2141 (1994).
53. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954 (1967).
54. Cantor, C. & Schimmel, P. *Biophysical Chemistry*, Part III. W.H. Freeman and Company, New York, 1131–1132 (1980).
55. O'Shea, E.K., Rutkowski, R. & Kim, P.S. Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538–542 (1989).
56. Laue, T.M., Shah, B.D., Ridgeway, T.M. & Pelletier, S.L. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S.E., Rowe, A.J. & Horton, J.C. eds) 90–125 (Royal Society of Chemistry, Cambridge, 1992).
57. Kozarsky, K. *et al.* Glycosylation and processing of the human immunodeficiency virus type 1 envelope protein. *AIDS* **2**, 163–169 (1989).
58. Harada, S., Koyanagi, Y. & Yamamoto, N. Infection of HTLV-III/LAV in HTLV-1–carrying cells MT-2 and MT-4 and application in a plaque assay. *Science* **229**, 563–566 (1985).
59. Schagger, H. & von Jagow, G. Tricine-sodium dodecylsulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochemistry* **166**, 368–379 (1987).
60. Allan, J.S., Strauss, J. & Buck, D.W. Enhancement of SIV infection with soluble receptor molecules. *Science* **247**, 1084–1088 (1990).