Retrovirus envelope domain at 1.7 Å resolution

Deborah Fass[1], Stephen C. Harrison[2] and Peter S. Kim[1]

We report the crystal structure of an extraviral segment of a retrovirus envelope protein, the Moloney murine leukemia virus (MoMuLV) transmembrane (TM) subunit. This segment, which comprises a region of the MoMuLV TM protein analogous to that contained within the X-ray crystal structure of low-pH converted influenza hemagglutinin, contains a trimeric coiled coil, with a hydrophobic cluster at its base and a strand that packs in an antiparallel orientation against the coiled coil. This structure gives the first high-resolution insight into the retrovirus surface and serves as a model for a wide range of viral fusion proteins; key residues in this structure are conserved among C- and D-type retroviruses and the filovirus ebola.

[1]Howard Hughes Medical Institute Whitehead Institute for Biomedical Research Department of Biology Massachusetts Institute of Technology Nine Cambridge Center, Cambridge, Massachusetts, 02142 USA
[2]Howard Hughes Medical Institute Department of Molecular and Cellular Biology Harvard University 7 Divinity Avenue, Cambridge, Massachusetts, 02138 USA

Correspondence should be addressed to P.S.K.

Surface glycoproteins target enveloped viruses to their host cell receptors and mediate fusion of the viral and cellular membranes[1]. The envelope proteins of retroviruses are synthesized as a single chain, which is then proteolytically cleaved into an N-terminal 'surface' subunit (SU) and a C-terminal transmembrane subunit (TM). The SU subunit binds the receptor, while the TM subunit contains the hydrophobic 'fusion peptide' at its N terminus[1] (Fig. 1*a*). To provide a structural framework for studies of retrovirus-mediated membrane fusion, we sought a high-resolution model of a retrovirus TM protein.

Moloney murine leukemia virus (MoMuLV) was chosen for study. MoMuLV SU and TM are linked by a labile disulphide bond that can be stabilized by thiol blocking agents[2], presumably by preventing a free thiol in the envelope complex from initiating thiol-disulphide rearrangement. This rearrangement eliminates the covalent association between SU and TM[2]. The ectodomain of the TM protein, without the fusion peptide, has been shown to fold into a stable structure in the absence of the SU subunit[3]. Furthermore, because the MoMuLV TM subunit lacks glycosylation sites, it can be represented accurately by synthetic and recombinant peptides.

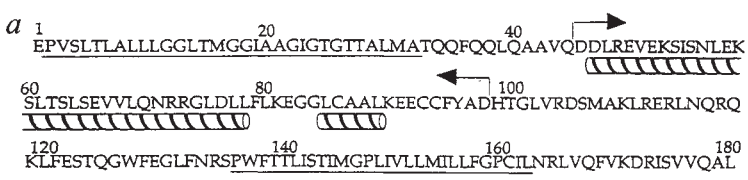
Conserved amino acid sequences in TM

MoMuLV TM (Fig. 1*a*) contains three amino acid sequence motifs that are highly conserved among C- and D-type retroviruses with hosts ranging from mice to humans. The region from residues 43–78 exhibits a 4-3 repeat of hydrophobic amino acids predicted to form a coiled coil[4,5]. An immunosuppressive peptide sequence (ISP), spanning residues 69–85, inhibits cellular immune responses[6,7]. Finally, three cysteine residues form the unvarying pattern CX_6CC from residues 86–94. The ISP and the cysteine motif are also found in envelope proteins the filovirus family[8], which differs morphologically from the retrovirus family and carries an RNA genome of opposite polarity[9].

Previous solution studies[3] demonstrated that a peptide containing the 4-3 hydrophobic repeat, the ISP, and the conserved cysteine motif displays the significant features of the recombinant ectodomain. The peptide, named Mo-55 for its length in residues, is trimeric. In addition, it assumes the identical disulphide bond connectivity as a larger segment lacking only the fusion peptide. Finally, its thermostability differs from that of the larger peptide by only 6 °C. Although the segment of the ectodomain lacking only the fusion peptide does have additional helical residues not present in Mo-55, their structure is lost non-cooperatively on heating[3]. Furthermore, treatment of the recombinant ectodomain with proteinases results in rapid removal of these residues, yielding proteinase resistant fragments similar to Mo-55. The entire ectodomain, containing the fusion peptide, aggregates in aqueous solution, suggesting that the fusion peptides are exposed and are not buried in the core of the structure. Thus, the information required to specify the fold of the stable core of the isolated TM subunit is contained within Mo-55.

Structure determination

Mo-55 was crystallized as described in Methods. The Mo-55 structure (Fig. 1*b*) was determined by multiple isomorphous replacement (MIR) using selenomethionine and thulium derivatives. An initial model built into a 2.4 Å resolution MIR map was improved by rounds of refinement and rebuilding using O[10]. The



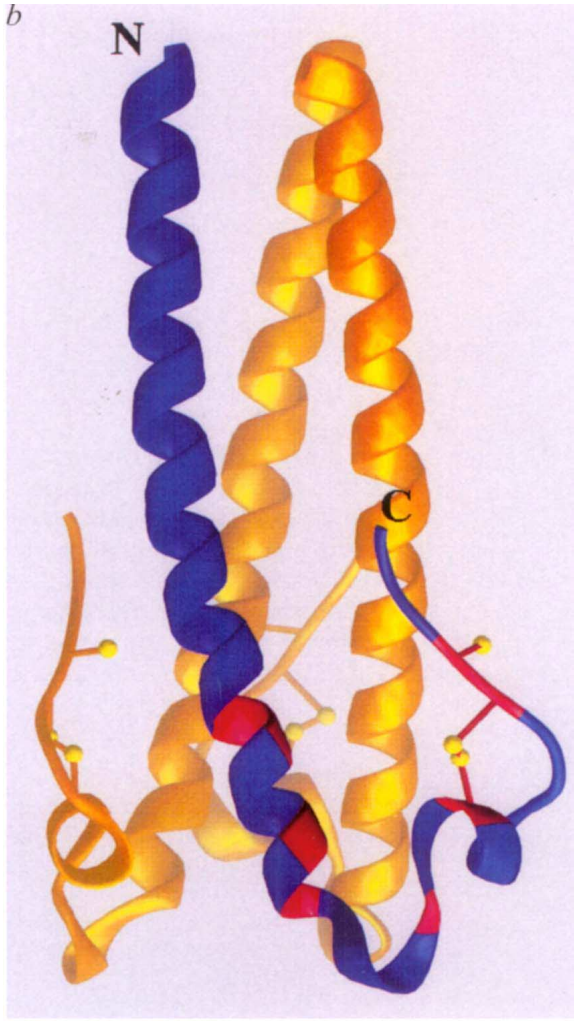


Fig. 1 *a*, Primary structure of MoMuLV protein. TM contains no glycosylation sites. The region between the arrows is Mo-55, the peptide expressed in *Escherichia coli* and crystallized. A hydrophobic and glycine rich sequence of approximately 30 residues at the N terminus of TM (underlined) is characteristic of a fusion peptide which has been shown, for influenza virus, to insert into the target membrane during viral entry[24,25]. Near the C terminus is a second hydrophobic stretch (underlined; residues 136–163) that anchors TM to the virus envelope[26]. Helical regions in the Mo-55 structure are indicated by striped poles under the sequence. *b*, Ribbon diagram of the Mo-55 trimer. Yellow balls indicate sulphur atoms in cysteine side chains, of which two per monomer are in an intramolecular disulphide bond. Residues in red (shown only in one monomer) are conserved among representative C-type retroviruses (MoMuLV, feline leukemia virus, human T-cell leukemia vrius type 1, Gibbon ape leukemia virus and bovine leukemia virus), as well as the D-type retrovirus Mason–Pfizer monkey virus, and the filovirus ebola.

current model includes residues 46–97 of the TM protein, 40 water molecules, and a chloride ion. This model is in good agreement with both the experimental data to 1.7 Å (working *R*-factor 16.9%, free *R*-factor 23.4%) and standard geometry (r.m.s.d. bonds 0.017 Å, r.m.s.d. angles 2.39°). Details of the data collection and refinement are given in Table 1.

MoMuLV TM architecture

The core of Mo-55 is a 33-residue, trimeric coiled coil. Six leucines, two valines, one isoleucine and one asparagine from each monomer are buried at the centre of the coiled coil. The three Asn 71 amide groups of the trimer trap an X-ray scatterer

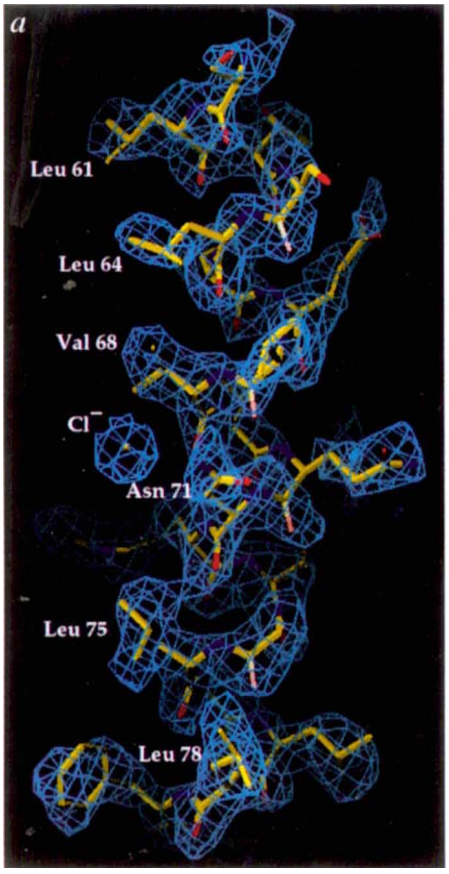
Table 1 Crystallographic and refinement statistics

Data collection and MIR statistics

Parameter	Native	Se	Tm
Resolution (Å)	20.0–1.7	20.0–2.4	20.0–2.4
Completeness (%)	96.0	97.0	97.3
R_{merge}[1]	0.061	0.064	0.054
R_{iso}[2]	—	0.109	0.245
Number of sites	—	1	1
R_c[3]	—	0.74	0.57
Phasing power[4]	—	1.23	2.13

Refinement statistics	
Number of atoms (non-hydrogen)	413
Number of H_2O molecules	40
Number of reflections (working)	5030
R_{cryst}[5]	0.169
Number of reflections (free)	414
R_{free}[6]	0.234
Deviations from ideal geometry	0.017 Å (bonds) 2.39° (angles)

Statistics are reported for crystals of a cysteine-to-alanine mutant at residue 94 in the TM sequence[3].
[1] $R_{merge} = \Sigma\Sigma_j|I_j - \langle I\rangle|/\Sigma|\langle I\rangle|$, where I_j is the intensity measurement for reflection j and $\langle I\rangle$ is the mean intensity for multiply recorded reflections.
[2] $R_{iso} = \Sigma||\boldsymbol{F}_{ph} \pm \boldsymbol{F}_p| - |\boldsymbol{F}_p||/\Sigma|\boldsymbol{F}_p|$, where $\boldsymbol{F}_{ph}$ and $\boldsymbol{F}_p$ are the derivative and native structure factors, respectively.
[3] $R_c = \Sigma||\boldsymbol{F}_{ph} \pm \boldsymbol{F}_p| - |\boldsymbol{F}_{h,c}||/\Sigma|\boldsymbol{F}_{ph} \pm \boldsymbol{F}_p|$ for centric reflections, where $\boldsymbol{F}_{h,c}$ is the calculated heavy-atom structure factor.
[4] Phasing power = $\langle \boldsymbol{F}_h\rangle/E$, where $\langle \boldsymbol{F}_h\rangle$ is the root-mean-square heavy-atom structure factor and E is the residual lack of closure error.
[5] $R_{cryst,\ free} = \Sigma||\boldsymbol{F}_{obs}| - |\boldsymbol{F}_{calc}||/|\boldsymbol{F}_{obs}|$, where the crystallographic and free *R*-factors are calculated using the working and free reflection sets, respectively. The free reflections (about 7% of the total) were chosen before refinement of the initial model and were not used during refinement.



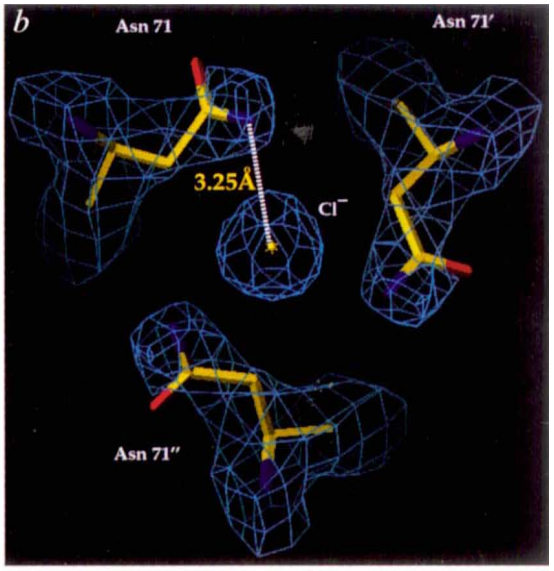


Fig. 2 *a*, A helical segment of the model is enveloped by a 1.2 σ, 2.4 Å MIR map. Residues in the core of the coiled coil are numbered. Only one subunit of the coiled coil is shown. *b*, Asn 71 and symmetry-related residues are viewed along the trimer axis. The distance between the amide nitrogen of Asn 71 and the chloride ion is 3.25 Å. The sum of the van der Waals radii for NH_4^+ (1.43 Å) and Cl^- (1.81 Å) is 3.24 Å (ref. 27).

(Fig. 2*a*, *b*) identified by solution studies as a chloride ion (Table 2). Buried polar interactions have been shown to promote specificity in the folding of model coiled coils[11,12], and the interaction of chloride with the buried Asn 71 residues may prove to be important in MoMuLV.

At the base of the Mo-55 coiled coil, the polypeptide chain reverses direction and forms a short α-helix (residues 85–89) nearly perpendicular to the coiled-coil axis. Hydrophobic clusters are formed by Phe 79, Leu 85, Ala 88 and Leu 89, together with the coiled-coil core residues Leu 75 and Leu 78, as well as Leu 77 and Leu 80 from an adjacent monomer. The C-terminal residues (90–97) of Mo-55 pack primarily against the counterclockwise adjacent monomer, as viewed from the top of the coiled coil. Cys 93 in the extended region forms an intramolecular disulphide bond with Cys 86 in the short helix. Cys 94 has a free thiol.

Implications for viral entry

X-ray crystallographic studies of enveloped virus surface proteins illustrate the remarkable diversity of native envelope protein conformations. Tick-borne encephalitis virus (TBE) E protein is a flat, elongated dimer that lies parallel to the surface of the virus[13], whereas influenza hemagglutintin (HA) is a trimeric 'spike' that extends approximately 150 Å from the viral membrane[14]. These two structures suggest the possibility that each virus family exhibits a unique solution to the membrane fusion problem.

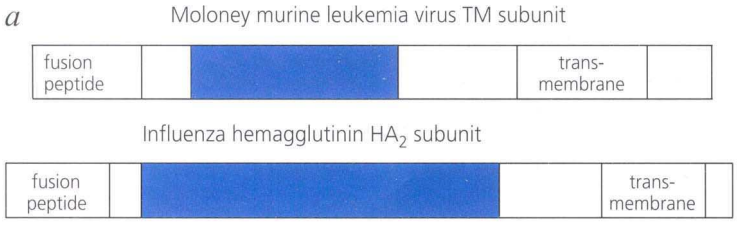
It is striking, therefore, that the structure of Mo-55 shares many features in common with the crystal structure of low pH-converted influenza HA ($TBHA_2$)[15,16] demonstrating that, even in the absence of sequence homology or any obvious evolutionary relationship, diverse viruses may use similar scaffolds for presenting fusion peptides to target cells (Fig. 3). Both $TBHA_2$ and Mo-55 have trimeric coiled coils in similar positions in their primary sequences. In addition, hydrophobic clusters stabilize both the Mo-55 and $TBHA_2$ trimers at the bases of their coiled coils. Finally, the C-terminal residues of both fragments extend back toward the N-termini of the coiled coils, apparently away from the viral membranes. The structural similarities between the envelope proteins of the retrovirus MoMuLV and the well-studied orthomyxovirus influenza open the rich literature on influenza HA (for review see ref. 17) to the question of the retrovirus membrane fusion mechanism.

Within their common structural framework, however, Mo-55 and $TBHA_2$ do contain notable differences. First, the $TBHA_2$ coiled coil spans 65 residues, while that of Mo-55 is only half as long. Furthermore, the Mo-55 coiled coil displays a regular superhelical twist of the order observed in a model trimeric coiled coil[18], and the 4-3 periodicity is preserved throughout the length of the coiled coil. In contrast, the $TBHA_2$ structure contains two regions with an irregular 3-4-4-3 pattern of core residues. Due to these skips, the $TBHA_2$ coiled coil has a pitch of 300–400 Å, while that of Mo-55 is approximately 175 Å. The location of disulphide bonds in the two structures is distinct, and the backbone conformations differ between the hydrophobic clusters. Finally, a short helix packs nearly antiparallel to the coiled coil in $TBHA_2$, a feature absent from the Mo-55 structure.

Table 2 Identification of buried polar interaction

	Midpoint of thermal denaturation	
	Mo-55.3	Asn 71→Ile mutant
none	64 °C	76 °C
NaF (5 mM)	64 °C	—
NaCl (5 mM)	73 °C	76 °C
NaBr (5 mM)	73 °C	—
NaCl (20 mM)	75 °C	—

The temperature midpoint for thermal unfolding (T_m) of Mo-55, but not an asparagine→isoleucine mutant at position 71, is sensitive to the presence of chloride. Chloride and bromide stabilized the Mo-55 structure, but fluoride did not.



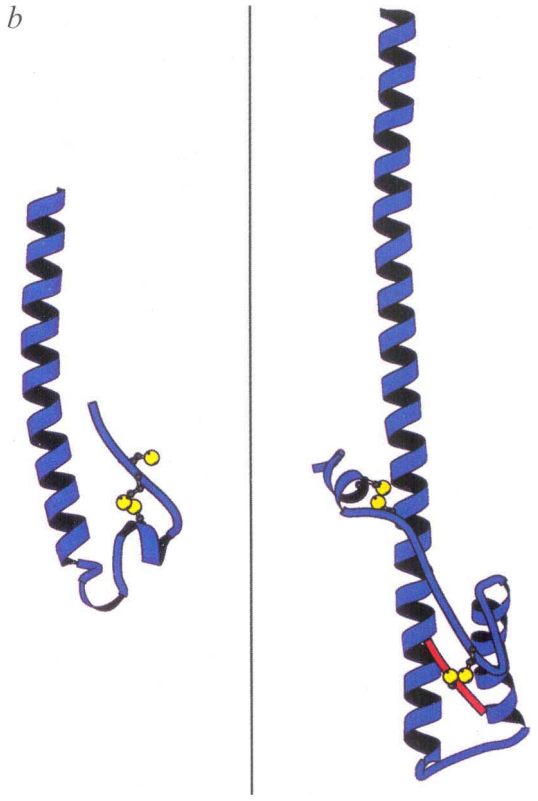


Fig. 3 Comparison of the Mo-55 and influenza $TBHA_2$[15] structures. *a*, A linear map drawn to approximate scale shows the positions of Mo-55 and $TBHA_2$ (blue) within TM and HA_2, respectively. Both crystallographic models are lacking N-terminal residues corresponding to the fusion-peptide region, as well as C-terminal segments that form protease sensitive links[3,16] to the transmembrane regions. *b*, The structures of Mo-55 (left) and $TBHA_2$ (right) are compared (see text and ref. 15) using MOLSCRIPT[28]. $TBHA_2$ is the fragment generated[16] by first cleaving native influenza HA from the viral membrane with the protease bromelain, and then treating this soluble HA with low pH. Low pH causes the receptor binding domains to dissociate from one another[29,30] and the transmembrane polypeptide HA_2 to undergo a dramatic conformational change involving extension of a three-stranded coiled coil[19]. Finally, further proteolysis generates a crystallizable fragment by removal of the dissociated HA_1 domains and the fusion peptides. The $TBHA_2$ structure is thought to be the HA conformation that mediates membrane fusion[15]. Yellow balls in the figure indicate sulphur atoms in cysteine side chains. A short strand from HA_1 is disulphide bonded to HA_2 in the $TBHA_2$ structure and is shown in red. The apparent bend in the Mo-55 helix is an artifact of representing one subunit of a coiled coil in a two-dimensional diagram.

It is likely that the Mo-55 structure represents the core of the fusion active conformation of MoMuLV TM, at a post-binding stage in which SU is no longer required for envelope function. For influenza, the low pH in the endosome leads to displacement of the receptor binding HA_1 domain with a consequent dramatic rearrangement of the transmembrane polypeptide HA_2. We expect that the SU subunit of retroviruses similarly aids in burial of the fusion peptide prior to activation, but the structure of TM in the SU/TM complex is unknown. Whether displacement or shedding of SU in retrovirus envelope proteins leads to a conformational change in TM, perhaps in a 'spring-loaded' manner[19], remains to be determined.

Methods

Crystallization and data collection. Mo-55 was purified as described previously[3]. Crystals in space group $P2_13$ ($a=b=c=53.1$ Å) were grown by vapour diffusion from 1.6 M NaCl, 50 mM KPO_4, pH 5.6, 2.5% PEG 8000. A peptide (called Mo-55.3) with a cysteine-to-alanine mutation at residue 94 formed crystals larger than those of the wild-type peptide, although the space group and unit-cell constants for both forms were identical. Crystals were transferred to 1.6 M NaAc, pH 5.6, 10% PEG 8000 at least 24 hours before data collection. Diffraction data were collected on a Siemens-Xentronics multiwire area detector mounted on an Elliott GX13 X-ray generator. Reflected X-ray intensities were integrated with the program Buddha[20]. Subsequent data analysis was done with CCP4 programs[21].

Phase determination. The Se derivative was prepared by introducing a leucine-to-methionine mutation at residue 47 of TM and incorporating selenomethionine at this position during expression of Mo-55.3[22]. The Tm^{3+} derivative was acquired by soaking crystals of Mo-55.3 in 1.6 M NaAc, pH 5.6, 10% PEG 8000, 50 mM $Tm_2(SO_4)_3$ for 24 hours. The Se and Tm^{3+} sites were located by Patterson methods and refined using the program MLPHARE (CCP4 program suite). Anomalous signal from the Tm^{3+} derivative was included. MIR maps calculated to 2.4 Å revealed connected backbone electron density from residue 47 to residue 96. Density for most of the side chains and many of the carbonyl groups was also clearly visible, facilitating the chain trace.

Structure refinement. Positional and *B*-factor refinements of the initial model were performed using X-PLOR[23], first against data to 2.1 Å. Ten fixed waters were added, and a bulk-solvent correction was applied. Further cycles of rebuilding and refinement were carried out using data to 1.7 Å. Data from wild-type crystals were also collected, and a difference map phased with the refined mutant model was used to position the Cys 94 side chain. The most common cysteine rotamer fit clearly into the largest peak in the difference map. However, extra density remained near the cysteine sulphur atom, indicating partial derivatization. This density was not modelled. The wild-type model was subjected to *B*-factor refinement and bulk-solvent correction, yielding an *R*-factor of 0.190 against data (20–1.8 Å) for the wild-type protein. Refined coordinates will be deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, NY; PDB ID code: 1MOF).

Thermal denaturation. Proteins with cysteine-to-alanine mutations at position 94 (Mo-55.3) were used to eliminate the possibility of intermolecular disulphide bond formation or thiol rearrangement during denaturation. Thermal denaturation was monitored by circular dichroism at 222 nm using 10 μM peptide. Buffer for all melting experiments was 50 mM $NaPO_4$, pH 7.0. CD signal was measured and T_m values were calculated as described previously[3].

Received 15 March; accepted 28 March 1996.

Acknowledgements
We thank J. Berger, S. Gamblin, D. Rodgers and members of the Harvard structural biology laboratory for their patience and generosity with time, equipment, and advice. We also thank T. Ellenberger for protocols for selenomethionine incorporation. Thanks to P. Harbury, B. Schulman, and other members of the Kim laboratory for support and encouragement. D.F. is a National Science Foundation Predoctoral Fellow. This research was supported by the Howard Hughes Medical Institute.

1. Hunter, E. & Swanstrom, R. Retrovirus envelope glycoproteins. *Curr. Top. Microbiol. Immun.* **157**, 187–253 (1990).
2. Pinter, A., Lieman-Hurwitz, J. & Fleissner, E. The nature of the association between the murine leukemia virus envelope proteins. *Virology* **91**, 345–351 (1978).
3. Fass, D. & Kim, P.S. Dissection of a retrovirus envelope protein reveals structural similarity to influenza hemagglutinin. *Curr. Biol.* **5**, 1377–1383 (1995).
4. 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).
5. Delwart, E.L., Mosialos, G. & Gilmore, T. Retroviral envelope glycoproteins contain a "leucine zipper"-like repeat. *AIDS Res. Human Retroviruses* **6**, 703–706 (1990).
6. Cianciolo, G.J., Copeland, T.D., Oroszlan, S. & Snyderman, R. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230**, 453–455 (1985).
7. Ogasawara, M. *et al.* Human INF-γ production is inhibited by a synthetic peptide homologous to retoviral envelope protein. *J. Immunol.* **141**, 614–619 (1988).
8. Volchkov, V.E., Blinov, V.M. & Netesov, S.V. The envelope glycoprotein of Ebola virus contains an immunosuppressive-like domain similar to oncogenic retroviruses. *FEBS Letters* **305**, 181–184 (1992).
9. Feldman, H., Will, C., Schikore, M., Slenczka, W. & Klenk, H.-D. Glycosylation and oligomerization of the spike protein of Marburg virus. *Virology* **182**, 353–356 (1991).
10. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A**47**, 110–119 (1991).
11. 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).
12. Lumb, K.J. & Kim, P.S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochem.* **34**, 8642–8648 (1995).
13. Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C. & Harrison, S.C. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298 (1995).
14. Wilson, I.A., Skehel, J.J. & Wiley, D.C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
15. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
16. Bullough, P.A. *et al.* Crystals of a fragment of influenza haemagglutinin in the low pH induced conformation. *J. Mol. Biol.* **236**, 1262–1265 (1994).
17. Hughson, F.M. Structural characterization of viral fusion proteins. *Curr. Biol.* **5**, 265–274 (1995).
18. Harbury, P.B., Kim, P.S. & Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83 (1994).
19. Carr, C.M. & P.S. Kim, P.S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823–832 (1993).
20. Blum, M., Metcalf, P., Harrison, S.C. & Wiley, D.C. A system for collection and on-line integration of X-ray diffraction data from a multiwire area detector. *J. Appl. Crystallogr.* **20**, 235–242 (1987).
21. *CCP4, a Suite of Programs for Protein Crystallography* (SERC (UK) Collaborative Computing Project No. 4, Daresbury Laboratory, Warrington, 1979).
22. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
23. Brünger, A.T. *X-PLOR (Version 3.1:) A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).
24. Stegmann, T., Delfino, J.M., Richards, F.M. & Helenius, A. The HA_2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion. *J. Biol. Chem.* **266**, 18404–18410 (1991).
25. Tsurudome, M. *et al.* Lipid interactions of the hemagglutinin HA_2 NH_2-terminal segment during influenza virus-induced membrane fusion. *J. Biol. Chem.* **267**, 20225–20232 (1992).
26. Pinter, A. & Honnen, W.J. Topography of murine leukemia virus envelope proteins: characterization of transmembrane components *J. Virol.* **46**, 1056–1060 (1983).
27. Weast, R.C., Ed., CRC Handbook of Chemistry and Physics (CRC Press, Inc. Boca Raton, Florida, 1980), p. F–216.
28. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 924–950 (1991).
29. White, J.M. & Wilson, I.A. Anti-peptide antibodies detect steps in a protein conformational change: low-pH activation of the influenza virus hemagglutinin. *J. Cell. Biol.* **105**, 2887–2896 (1987).
30. Kemble, G.W., Bodian, D.L., Rose, J., Wilson, I.A. & White, J.M. Intermonomer disulfide bonds impair the fusion activity of influenza virus hemagglutinin. *J. Virol.* **66**, 4940–4950 (1992).