The EMBO Journal vol.15 no.13 pp.3267–3274, 1996

The solution structure of the bovine leukaemia virus matrix protein and similarity with lentiviral matrix proteins

Stephen Matthews[1,2], Michael Mikhailov[3,4], Arsene Burny[5] and Polly Roy[3,4,6]

[1]Department of Biochemistry, Imperial College of Science, Technology and Medicine, University of London, Exhibition Road, South Kensington, London SW7 2AY, [3]Laboratory of Molecular Biophysics and Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, [4]NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR, UK, [5]Department of Molecular Biology, University of Brussels, B1640, Rhode-St-Genese, Belgium and [6]Department of Public Health Sciences, University of Alabama at Birmingham, Birmingham, AL 35294, USA

[2]Corresponding author

In the mature virion, retroviral matrix proteins are found in association with the inner face of the viral membrane. They play a critical role in determining the morphogenesis of virus assembly. We have determined the three-dimensional solution structure of the bovine leukaemia virus (BLV) matrix protein by heteronuclear nuclear magnetic resonance. The protein contains four principal helices that are joined by short, partially structured loops. Despite no sequence similarity with the lentiviruses, the structure shows an intriguing homology with the equivalent protein from the human and simian immunodeficiency viruses. A root-mean-square deviation of 3.78 Å is observed over the backbone atoms of 36 equivalent helical positions. The similarity implies a possible common assembly unit for the matrix proteins of type C retroviruses.

Keywords: bovine leukaemia virus/lentivirus/matrix protein/solution structure

Introduction

The bovine leukaemia virus (BLV) is both structurally and biologically related to the human T-cell lymphotropic viruses (HTLV-I and HTLV-II) and the simian T-cell leukaemia virus (STLV), with many steps in the disease progression being similar (Ina and Gojobori, 1990). For this reason, it has been chosen by several investigators as an animal model for HTLV. HTLV-I infection is associated with diseases ranging from mild immunosuppression to adult T-cell leukaemia and myelopathy spastic paraparesis (Poiesz *et al.*, 1980; Inamura *et al.*, 1988). BLV and HTLV-I share significant homology and form a phylogenic group that is distinct from the lentiviruses. Specifically, the matrix protein (MA) from BLV shares some 30% identity with the equivalent region from HTLV-I but only ~3% identity with the MAs of human and simian immunodeficiency viruses (HIV and SIV). Despite this, both the HTLV/BLV group and the lentiviruses share a similar replicative life cycle. However, whereas HIV destroys CD4 T cells, HTLV can transform CD4 T cells to an unrestricted growth rate (Palker, 1992).

Retroviral replication is initiated by the penetration of the virion core into the host cell. This process is initiated by a series of interactions between the lipoprotein envelope and a cell surface receptor. Following penetration, the single-stranded RNA is converted into double-stranded DNA by the viral reverse transcriptase, which is encoded by the *pol* gene. The provirus intermediate that is formed migrates to the nucleus and is incorporated into the host's genome by the virion-associated integrase. After transcription and transport of the viral mRNA to the cytoplasm, translation of the viral proteins occurs. The gag precursor protein assembles into particles and eventually is cleaved into the principal proteins, which include the MA, the capsid and nucleocapsid proteins. The envelope, which is acquired during budding, is encoded by the *env* gene and contains the surface glycoproteins, comprising an outer domain, a transmembrane segment and a cytoplasmic domain.

In the mature virion, MA is situated on the inner face of the viral envelope, where it is associated with the lipid bilayer in an icosahedral arrangement (Gelderblom *et al.*, 1989; Arnold and Arnold, 1991; Nermut *et al.*, 1994; Spearman *et al.*, 1994). In type C retroviruses, targeting signals within MA are essential for the coherent assembly of virions at the plasma membrane and efficient release of virions. Large deletions of MA redirect gag assembly from a site at the plasma membrane to the endoplasmic reticulum (Fäcke *et al.*, 1993; Gonzalez *et al.*, 1993; Yuan *et al.*, 1993; Zhou *et al.*, 1994). The lentivirus and HTLV/BLV groups both display type C assembly characteristics. In type B and D retroviruses, the gag protein pre-assembles before association with the membrane (Gelderblom, 1991; Wills and Craven, 1991). A particularly well studied example of type D retroviruses is the Mason–Pfizer monkey virus (M-PMV). Interestingly, a single point mutation in the MA of M-PMV enables this D-type virus to assemble in a manner similar to a type C virus (Rhee and Hunter, 1990). These studies imply a key role for MA in determining the steps in retroviral morphogenesis. There is little sequence homology between retroviral MAs, the only common feature being a myristate attached to the amino-terminal glycine.

The first atomic resolution insight into MA function came from the determination of the three-dimensional (3D) solution structure of HIV-1 MA by nuclear magnetic resonance (NMR) (Massiah *et al.*, 1994; Matthews *et al.*, 1994, 1995). HIV-1 MA in solution has an extremely compact fold and contains a triple-stranded, mixed β-sheet together with five principal α-helices and a short 3–10 helix, joined by short loops or regions of extended structure. The structure presents a localized patch of basic residues that is most likely responsible for efficient

targeting to the membrane, by providing a surface for association with the acidic phospholipid headgroups of the inner face of the membrane. This is consistent with the notion that the myristoylated glycine and the basic region form a bipartite autonomous membrane targeting signal, a common feature among the C-type retroviral matrix proteins. Furthermore, results from a recent X-ray crystallographic study of SIV MA indicate that protein trimerization may be the basis of MA assembly (Rao *et al.*, 1995). Here we report the structure of the BLV MA in solution and, despite no sequence similarity, structural homology with the MAs of HIV and SIV. Structurally, this represents a particularly intriguing result inasmuch as the sequences of both HTLV and BLV MAs contain >20% proline residues, compared with only ~3% for HIV and SIV MA. To our knowledge, this is the first report of a structure from the HTLV/BLV retroviral group and not only provides additional evidence in support of the trimer mode of assembly, but has implications for general type-C retroviral assembly and perhaps common evolutionary origins.

Results and discussion

The 109 residue MA from BLV was expressed as a glutathione-*S*-transferase (GST) fusion protein in *Escherichia coli*. The structures were calculated on the basis of 1004 nuclear Overhauser effect (NOE) distance restraints and 33 H bond distance restraints. This represents an average of 14 distance restraints per residue for the well-defined region of the protein. The calculations were executed within the program X-PLOR using a dynamic simulated annealing protocol (Nilges *et al.*, 1988; Brunger, 1993). A total of 30 structures were calculated, of which 17 contained no distance violations greater than 0.5 Å. The sequence-specific backbone ^{1}H assignments were completed using CT-HNCA and CT-HN(CO)CA experiments (Grzesiek and Bax, 1992a). Figure 1 shows an example of strips taken from the CT-HNCA experiment for the helical region between residues 15 and 25. The distance restraints were identified from 3D ^{15}N–^{1}H and ^{13}C–^{1}H nuclear Overhauser (NOESY) heteronuclear multiple quantum coherence (HMQC) spectroscopy (Marion *et al.*, 1989; Bax *et al.*, 1990; Driscoll *et al.*, 1990) and 2D homonuclear NOESY spectra.

Solution structure of BLV MA

The final family of 17 refined structures is shown in Figure 2. The overall root-mean-square (r.m.s.) deviations between the family and mean co-ordinate position are 0.74 ± 0.1 Å for the backbone atoms and 1.18 ± 0.1 Å for all heavy atoms in regions of secondary structure. For all residues between 13 and 85, the r.m.s. deviation values are 0.87 ± 0.1 Å and 1.43 ± 0.1 Å for backbone and all heavy atoms, respectively. The complete set of structural statistics is shown in Table I. The pattern of sequential NOEs across the Xaa–Pro peptide bond and characteristic ^{13}C chemical shifts of the C_β resonance are indicative of a *trans* conformation for all nine proline residues between residues 13 and 85 (Poznanski *et al.*, 1993; Mcinnes *et al.*, 1994). However, some evidence for *cis–trans* isomerization was observed at temperatures close to the unfolding temperature of BLV MA (35°C). ^{1}H–^{15}N HSQC

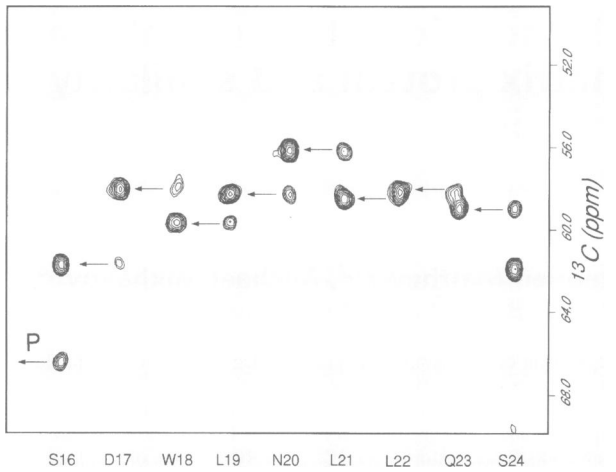


Fig. 1. Representative strips taken from the CT-HNCA spectrum for the helical region spanning residues 16–24. The inter-residue cross-peaks are indicated by the bar and the C_α resonance of Pro15 is labelled P.

spectra and ^{15}N relaxation show that the terminal residues outside the highly structured region of the protein (residues 13–85), which includes the sites of myristoylation and protease cleavage, are extremely mobile in solution. Satellites in the ^{1}H–^{15}N HSQC and CBCA(CO)NH spectra also confirm that the terminal Xaa–Pro peptide bonds experience significant *cis–trans* isomerization. These observations are consistent with the level of accessibility needed for the enzymatic reactions.

The structure consists of four α-helices separated by short, semi-structured loops which give rise to the close packing of adjacent helices. Helix A runs from residues Ser14 to Asn29 through a proline at position 15 within the first turn. The hydrophobic side chains of Trp18, Leu21, Leu22 and Ala25 make contacts with the helix B through Phe37, Leu40 and Ile44. Helix B, Ser32 through Pro33 to Thr51, is extremely hydrophobic and lies within the centre of the protein and contacts all the other helices. Figure 2B shows the structured hydrophobic contacts occurring around this helix. Helix C, spanning residues 57–65, is kinked about the Gly–Gly–Pro sequence and contacts both helices B and D. In particular, Cys65 has numerous NOE cross-peaks to Trp46, Phe70 and Leu77. The C-terminal helix, helix D, runs from Val73 to Leu85 and contains a proline at position 74.

The major structural roles for the proline residues within BLV MA fall into three main categories. First, in the helices A, B and D, a proline residue is present within the first turn. Helix-like conformation was confirmed by the observation of the standard helical sequential NOEs together with NOEs from both delta protons of the proline to neighbouring amides. The inability to donate a hydrogen bond and the fact that the ϕ is permanently constrained to –65° makes proline a particularly common N-terminal stabilizer of helices. Second, the proline residues in the Pro–Xaa–Pro–Xaa–Pro motif between helices A and B add a degree of stiffness to this loop. Third, the Pro–Pro–Gly sequence between residues 66 and 68 provides a turn between helices C and D.

Similarity with HIV and SIV MA

A library of 900 protein 3D structural domains was searched automatically by overlaying the sequences of

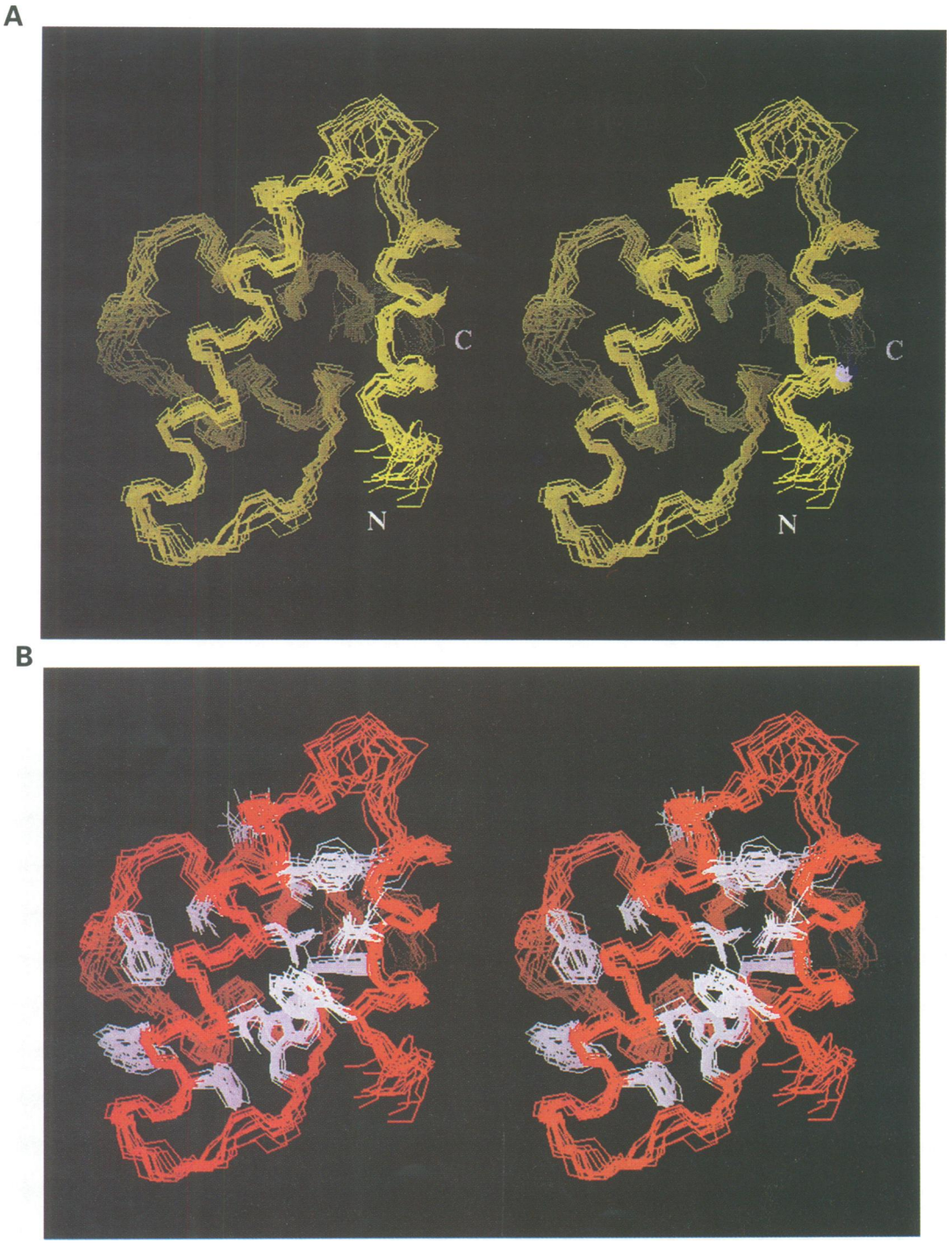


Fig. 2. Stereoview of the superimposition of the 17 refined BLV MA structures. (**A**) The backbone trace for residues 13–85 in BLV MA. (**B**) The backbone trace (in red) together with the principle side chains making hydrophobic contacts (in white).

BLV MA with every fifth residue in each database structure. Residues aligned by this procedure were used to derive an initial fit of BLV MA and the database structure. The initial fit was then refined using the protein structure comparison algorithm described by Russell and Barton (1992). The topology of BLV MA was judged using this method to be closest to that of the folds of HIV-1 and SIV MA. Thirty six equivalent C_α atoms were used to derive the superimposition and gave an r.m.s. deviation of 3.78 Å with the structure of SIV MA (Rao *et al.*, 1995). The recent crystal structure of SIV MA has revealed that it adopts an identical fold to HIV-1 MA, determined by earlier NMR studies. The r.m.s. deviations between the secondary structure backbone atoms of SIV and HIV-1 MA are 1.88 Å for the average Matthews structure (Matthews *et al.*, 1994, 1995) and 3.67 Å for the average Massiah structure of HIV-1 MA (Massiah *et al.*, 1994). In Figure 4, a protein cartoon representation of SIV/BLV MA superimposition is shown together with the sequence alignment. Despite the negligible sequence

Table I. Results from structure calculations for BLV MA

Statistic		<SA>[a]	SA_{ref}[b]
Restraints[c]			
Intraresidue	205	0.0023 ± 0.001	0.019
Medium and short range (i to $1 \leq i \leq 4$)	586	0.0490 ± 0.007	0.046
Long range (i to $i > 4$)	213	0.0490 ± 0.008	0.042
Hydrogen bonds	33	0.0420 ± 0.001	0.074
Final experimental energy terms			
F_{NOE}[d]		67.2 ± 14.2	62.1
F_{L-J}[e]		−253.5 ± 11.7	−233.5
F_{angle}[d]		418.5 ± 14.2	411.4
$F_{improper}$[d]		38.1 ± 5.4	32.5
F_{bond}[d]		52.4 ± 4.5	45.6
R.m.s. deviations from restraints and idealized geometry			
NOE distances[c]		0.045 ± 0.005	0.044
Bonds[c]		0.0061 ± 0.0004	0.0057
Angles[f]		1.04 ± 0.02	1.03
Improper angles[f]		0.58 ± 0.09	0.53

[a]The average r.m.s. deviations for the final 17 structures.
[b]This represents the restrained minimised average structure of the 17 structures.
[c]Ångstroms.
[d]kJ/mol.
[e]Calculated using the full CHARMM energy function (Brooks *et al.*, 1983), kJ/mol.
[f]Degrees.

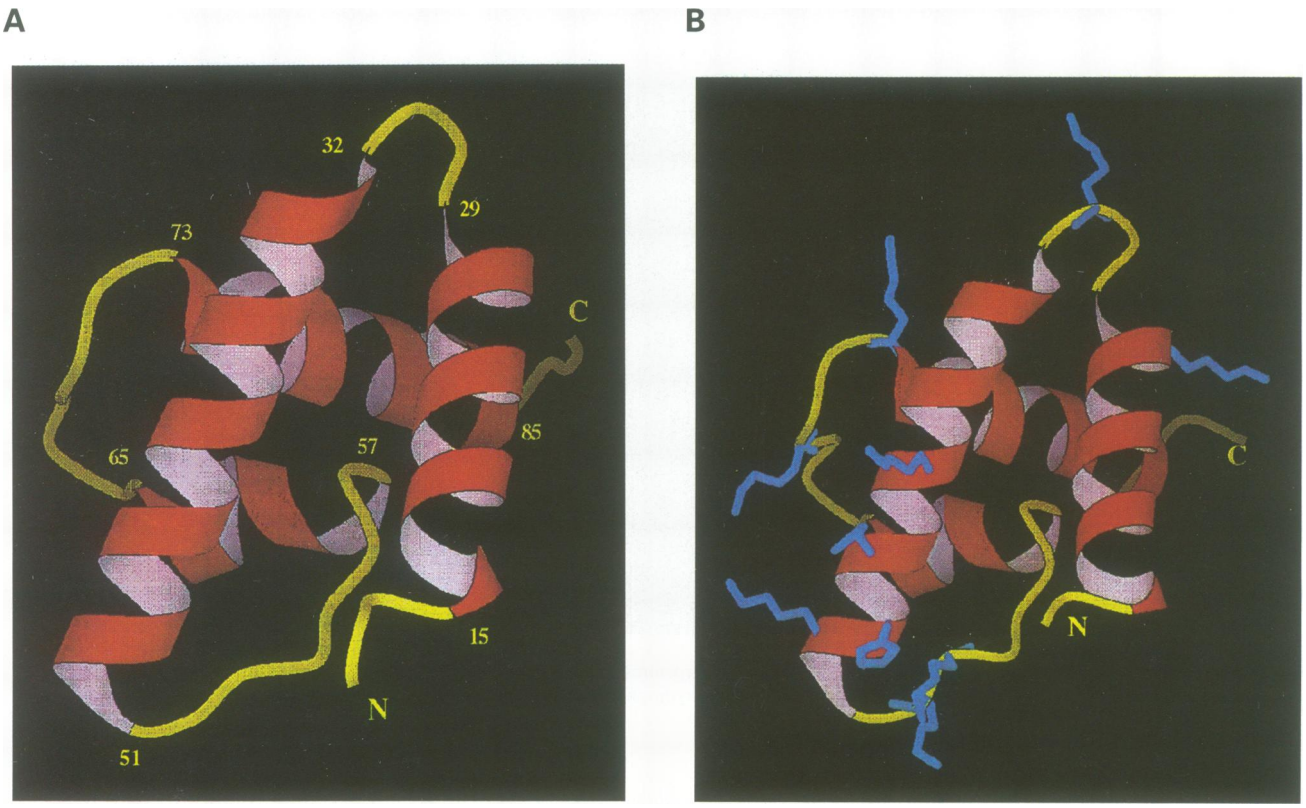
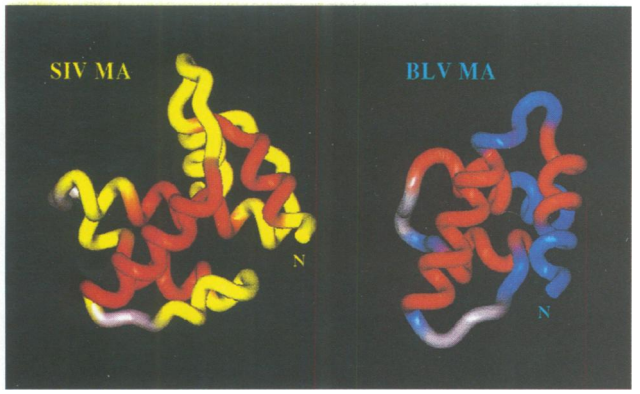


Fig. 3. Schematic representation of the structure of BLV MA. (**A**) A MOLSCRIPT (Kraulis, 1991) diagram representing the backbone trace. The start and stop residues or the helices are numbered. The N- and C-termini are also labelled. For clarity, the orientation of the molecule is kept identical to Figure 2. (**B**) A MOLSCRIPT (Kraulis, 1991) diagram representing the backbone trace together with the side chain bonds of arginine and lysine (in blue). The N- and C-termini are labelled.

identity between the two proteins, the four helices of BLV MA display a similar topology with the first four α-helices of SIV MA. Moreover, the structure of BLV MA also presents an analogous highly basic surface region. In SIV MA, numerous arginine and lysine residues within the N-terminal 30 amino acids form a highly basic region that provides a motif presumably to interact with the acidic phospholipid head groups. This mechanism, in addition to

A



B

```
ma_hiv1h2              MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILG
ma_sivmm251            MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRFGLAESLLENKEGCQKILS
ma_blvau       MGNSPSYNPPAGISPSDWLNLLQSAQRLNPRPSP   SDFTDLKNYIHWFHKTQKKP    W  TFTSGG
ma_blvj        MGNSPSYNPPAGISPSDWLNLLQSAQRLNPRPSP   SDFTDLKNYIHWFHKTQKKP    W  TFTSGG
ma_htlv2 MGQIHGLSPTPIPKAPRGLSTHHWLNFLQAAYRLQPRPSD   FDFQQLRRFLKLALKTPI      WLNPIDYSL
ma_htl1m MGQIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSS   YDFHQLKKFLKIALETPV      WICPINYSL
ma_htl1c MGQIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSS   YDFHQLKKFLKIALETPV      WICPINYSL
ma_htl1a MGQIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSS   YDFHQLKKFLKIALETPA      RICPINYSL

ma_hiv   QLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSSQVSQNY
ma_siv   VLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVKHTEEAKQIVQRHLVVETGTAETMPKTSRPTAPSSGRGGNY
ma_blvau         PASCPPGKFGRVPLVLATL  NEVLSNDEGAPGASA      PEEQPPPYDPPAILPIISEGNRNR
ma_blvj          PTSCPPGRFGRVPLVLATL  NEVLSNEGGAPGASA      PEEQPPPYDPPAVLPIISEGNRNR
ma_htlv2         LASLIPKGYPGRVVEIINILVKNQVSPSAPAAPVPTPICPTTTPPPPPPPSPEAHVPPPYVEPTTTQCF
ma_htl1m         LASLIPKGYPGRVNEILHILIQTQAQPSRPAPPPPSS       PTHDPPDSDPQIPPPYVEPTAPQVL
ma_htl1c         LASLIPKGYPGRVNEILHILIQTQAQPSRPAPPPPSS       STHDPPDSDPQIPPPYVEPTAPQVL
ma_htl1a         LASLIPKGYPGRVNEILHILIQTQAQPSRPAPPPPSS       PTHDPPDSDPQIPPPYVEPTAPQVL
```

Fig. 4. The structural similarity between the MAs from SIV and BLV. (**A**) Protein worms showing the superimposition of SIV and BLV MA. For clarity, the molecules are separated, the orientation of the molecule remains unchanged from superimposition. The equivalent helical positions are shown in red and the residues involved at the trimer interface are shown in grey. (**B**) Sequence alignment for HIV-1, SIV and the HTLV/BLV group based on the structural homology. The bold letters indicate 36 equivalent C_α atom positions. The sequences are labelled with their Swiss-Prot accession numbers.

myristoyl group burial, provides the strength of interaction necessary for anchoring gag to the membrane. In addition to mutagenesis studies, the possible presence of this interaction is established by studies in which basic amino acids have been shown to bind sodium dodecylsulphate via an electrostatic mechanism (Forgacs, 1994). The significantly stronger interaction observed for lysine may account for the increased occurrence of lysine over arginine in type C MAs. In the BLV MA, the basic residues are not concentrated at the N-terminus but are relatively evenly distributed throughout the protein sequence. However, the majority of these residues fall on the same side of the molecule within the 3D structure (Figure 3). This creates a strong dipole across the molecule and a positively charged surface for membrane interaction. There are two principal differences between the structures of BLV and SIV MA. First, SIV and HIV MAs contain a triple-stranded, mixed β-sheet that is not present in BLV MA. This region accommodates the nuclear localization motif (Bukrinsky *et al.*, 1993a,b; Galley *et al.*, 1995) that enables HIV to productively infect non-dividing cells; a distinguishing feature of lentiviruses not shared by oncoviruses. Second, an additional three helices are observed at the C-terminus of SIV but are not seen in BLV MA.

Implications for assembly

In the crystal structure the SIV, molecules are organized around a 3-fold axis to produce a trimer that is proposed to be a fundamental intermediate of gag assembly (Rao *et al.*, 1995). This model has many attractive features. First, electron microscopy has visualized HIV gag in an icosahedral arrangement (Nermut *et al.*, 1994). The assembly unit is a six membered ring that is constructed from two monomers donated by three individual trimers. Second, the orientations of the N- and C-termini within the SIV MA trimer are consistent with the major functions of MA. All the N-terminal residues implicated in membrane targeting and anchoring project from the upper surface with the basic side chains available to bind to the acid phospholipid head groups and the N-terminal myristoyl moiety positioned for burial within the membrane. The C-terminal helix extends from the lower surface and is ideally situated for interaction with the capsid protein, p24. Third, although the inter-monomer interactions within the trimer are specific, the surface area buried upon trimerization is relatively small. This manifests itself as a relatively weak complex that is consistent with the observations of monomeric SIV and HIV MAs in solution (Massiah *et al.*, 1994; Matthews *et al.*, 1994, 1995). The additional driving force necessary for complete virion assembly is provided by oligomerization of the capsid protein (Ehrlich *et al.*, 1992; Dorfman *et al.*, 1994).

In addition to the general helical topology, a striking similarity is observed between the trimer interface loops of SIV MA (shown in grey in Figure 4) and the loops at the helix B–C and C–D junctions in BLV MA. The SIV trimer interface is represented by the loop residues between

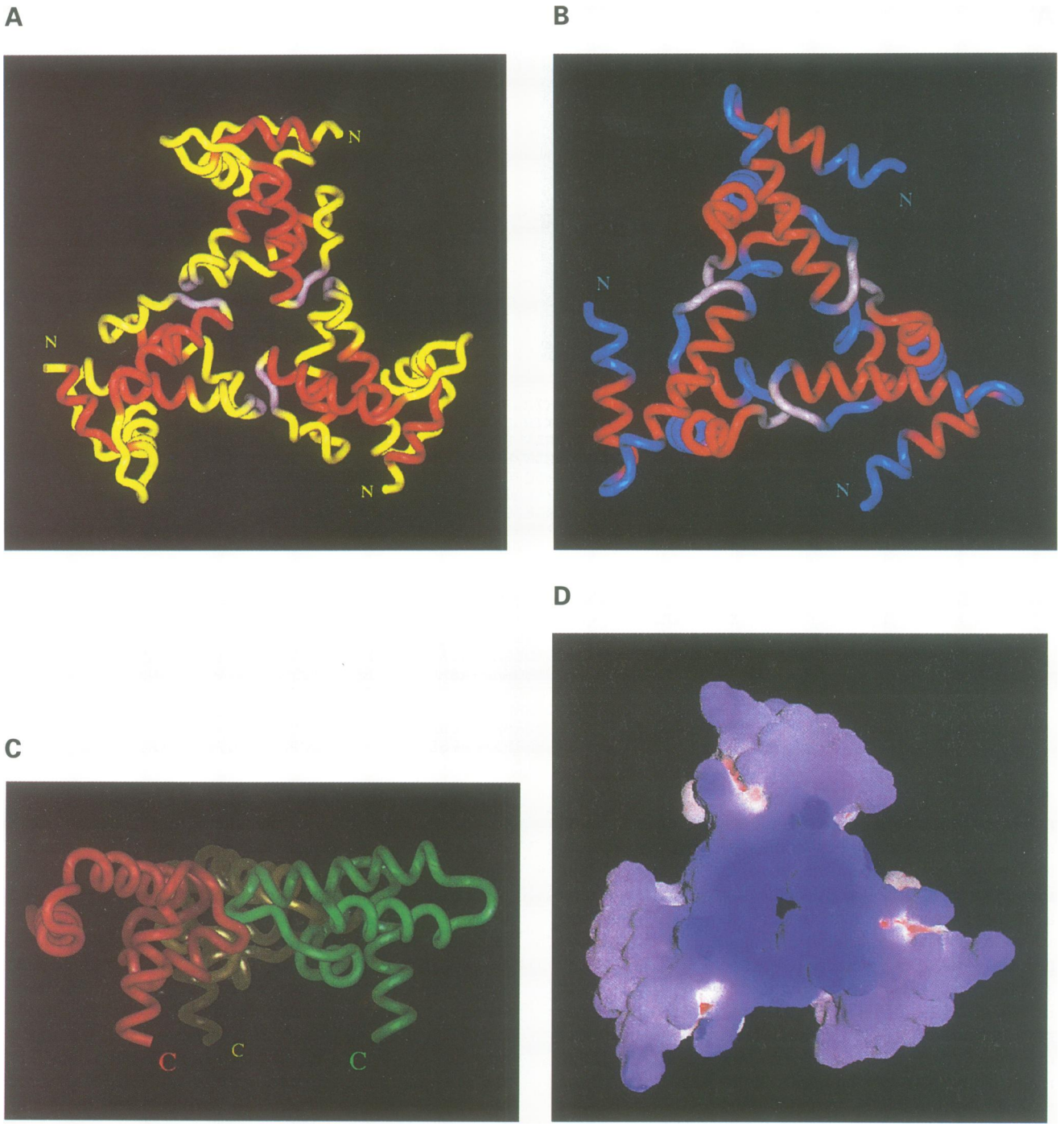


Fig. 5. Representation of the trimer model for BLV MA. (**A**) Protein worm showing the relative orientation of the monomers of SIV MA within the trimer model. The trimer is viewed from above the plane of the membrane. The equivalent helical positions are shown in red and the residues involved at the trimer interface are shown in grey. (**B**) Protein worm showing the relative orientaion of the monomers of BLV MA within the trimer model. The orientation of the trimer and the colour code is the same as in (A). Within the inherent error of the proposed BLV model, the dimensions of the SIV and BLV MA trimers are the same (diameters of ~65 and 60 Å are observed for SIV and BLV). (**C**) A depth-cued protein worm showing the relative orientation of the monomers of BLV MA within the trimer model. The orientation represents a 90° rotation about the *x*-axis of (A) and (B). The colour code is for clarity of chain identity. The N- and C-termini are labelled for the respective monomers. (**D**) A GRASP representation (Nicholls *et al.*, 1991) of the electrosatic properties of the BLV MA trimer. The calculated dipole is perpendicular to the membrane. Blue colour indicates positive charge and red negative.

45–47 and 70–72 (shown in grey in Figures 4 and 5). The main inter-monomer interactions occur as hydrogen bonds between the carbonyl of Gly45 and the amide of Ser72, and between the amide of Ala47 and the carbonyl of Thr70. In the BLV MA structure, the putative trimer interface loops, based on superimposition, span regions around Gln51 and Lys69. We propose that BLV MA can therefore assemble in an analogous fashion to SIV, with main chain hydrogen bonds forming the principal interactions. Main chain contacts between conformationally flexible regions, such as those observed in the SIV MA crystal, are common features in viral assembly (Harrison,

1992; Stuart, 1993; Grimes *et al.*, 1995). Interactions of this type remove any requirement for high degrees of sequence conservation between viruses and rationalize the poor sequence homology observed between the retroviral MAs. Our model for the trimerization of BLV, shown in Figure 5, is based on the superimposition of the loop regions from individual BLV monomers on the molecules of the SIV trimer. Circumstantial evidence endorsing the SIV MA trimer structure is also evident in the BLV MA model. The basic surface observed in BLV MA monomers extends across the upper surface of our trimer model, with the dipole direction perpendicular to the membrane (Figure 5D) and all but one (per monomer) of the basic residues available for interaction. The C-terminus of BLV also projects from the lower surface and is available for interaction with the core (Figure 5C). A tapering of the unit, similar to that for SIV MA, is also observed, with the trimer being ~60 Å in diameter at the membrane interface and 45 Å at the C-terminus (Figure 5C). This enables close packing of neighbouring trimers (Nermut *et al.*, 1994).

The protein sequences of the MA from HTLV I, HTLV II, STLV and BLV are extremely homologous with each other. This implies that any conclusions based on the structure of BLV MA will probably hold true for all strains of HTLV and STLV. Furthermore, we propose that the principle of type C MA design is based largely on one requirement: a trimeric assembly unit mediated by promiscuous main chain hydrogen bonds that presents a highly positively charged surface to the inner face of the membrane and the C-terminus to the core.

Materials and methods

BLV MA expression and purification

The DNA encoding the 109 residues of the BLV MA was amplified using PCR and cloned into the expression vector pGEX2T (Smith and Johnson, 1988). Expression of GST was achieved after induction of the BL21 *E.coli* cells with isopropyl-β-D-thiogalactopyranoside. Harvested cells were resuspended in phosphate buffer at pH 7.4 and disrupted by sonication at 5°C for six 1 min periods. The supernatant was then applied to a glutathione–Sepharose column and eluted with 10 mM glutathione in 50 mM Tris–HCl at pH 8.6. The eluted polyprotein was then cleaved by incubating with thrombin (1 U/1 mg fusion protein) at room temperature for 4 h. The pH was then adjusted to 9.0 and the solution was applied to a Q-Sepharose anion-exchange column. BLV MA eluted before GST with a linear gradient from 0 to 2 M NaCl in 50 mM Tris at pH 9.0. BLV MA was pooled, dialysed against 30 mM NaCl and 15 mM phosphate buffer at pH 6.0 and freeze-dried. For the preparation of ^{15}N- and $^{15}N/^{13}C$-labelled material, the bacteria were grown on minimal media with either ^{15}N-labelled NH_4Cl or both ^{15}N-labelled NH_4Cl and $^{13}C_6$-labelled glucose as the nitrogen and carbon sources.

NMR spectroscopy

For NMR spectroscopy, 10 mg of freeze-dried protein was dissolved in 0.6 ml of H_2O containing 10% D_2O. All NMR spectra were recorded at 500 MHz proton frequency and the temperature was maintained at 293 K throughout the experiments. The sequence-specific backbone ^1H assignments were completed using the CT-HNCA and CT-HN(CO)CA experiments (Grzesiek and Bax, 1992a). Assignment of C_β resonances of the proline resonances were completed with the CBCA(CO)NH experiment (Grzesiek and Bax, 1992b). The secondary structure and ^1H side chain assignments were determined using ^{15}N-^1H NOESY–HMQC and Hartmann–Hahn (HOHAHA) HMQC spectroscopy (Marion *et al.*, 1989; Driscoll *et al.*, 1990). In the final step, ^1H-^{13}C NOESY–HMQC, ^1H-^{13}C-^{13}C-^1H total correlated spectroscopy (HCCH–TOCSY) and HCCH–COSY experiments (Bax *et al.*, 1990) are utilized to assign the ^1H chemical shifts of the remaining amino acid spin systems. This procedure lays out the ground work for the complete interpretation of the ^{15}N and ^{13}C NOESY–HMQC spectra, which enabled distance restraints to be obtained for the full structure calculation. Additional restraints were identified from 2D NOESY (100 ms mixing time) spectra recorded at 750 MHz on a fully exchanged D_2O sample.

Structure calculation

The NOESY cross-peak intensities were measured at both 80 and 150 ms mixing times. The structures were calculated on the basis of 1004 NOE distance constraints and 33 H bond distance restraints using a dynamic simulated annealing protocol executed within the program X-PLOR (Nilges *et al.*, 1988; Brunger, 1993). The NOE restraints were composed of 205 intraresidue, 586 sequential (residue i to residues $1 \leq i \leq 4$) and 213 long range (residue i to residues $i > 4$) connectivities. The distance restraints were categorized into three groups on the basis of estimated NOE cross-peak intensity: strong 2.7, medium 3.8 and weak 5.0 Å. Standard values for pseudo-atom corrections were added to the NOE distance constraints where appropriate. Hydrogen-bonded NH groups were identified in helical regions by the presence of an NH resonances 6 h after dissolving in D_2O. The helical hydrogen bonds are characterized by two restraints r_{NH-O} = 2.1–2.3 Å and r_{N-O} = 2.1–3.3 Å, and were introduced once they could be assigned unambiguously following initial structure calculations.

Acknowledgements

The authors generously acknowledge Professor Iain Campbell for advice and access to NMR spectrometer time. We are extremely grateful to Drs Dave Stuart and ZiHe Rao for access to the co-ordinates of SIV MA. For help with the structural similarity search we thank Dr Geoff Barton. We would also like to thank Jonathan Boyd, Harold Schwalbe, Nick Soffe and Helena Kovacs for useful discussion. This project was supported by the MRC, and S.M. was an MRC career development award recipient. The co-ordinates will be deposited with the Brookhaven protein data bank and are available from S.M. (sjm@bioch.ox.ac.uk).

References

Arnold,E. and Arnold,G.F. (1991) Human immunodeficiency virus structure—implications for antiviral design. *Adv. Virus Res.*, **39**, 1–58.

Bax,A., Clore,G.M. and Gronenborn,A.M.J. (1990) ^1H-^1H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^1H and ^{13}C spectra of ^{13}C enriched proteins *Magn. Resononance*, **88**, 425–431.

Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimisation and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

Brunger,A.T. (1993) *XPLOR Manual Version 3.1.* Yale University Press, New Haven, CT.

Bukrinsky,M.I., Haggerty,S., Dempsey,M.P., Sharova,N., Adzhubei,A., Spitz,L., Lewis,P., Goldfarb,D., Emerman,M. and Stevenson,M. (1993a) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature*, **365**, 666–669.

Bukrinsky,M.I., Sharova,N., McDonald,T.L., Pushkarskaya,T., Tarpley, W.G. and Stevenson,M. (1993b) Association of integrase, matrix and reverse transcriptase antigens of human immunodeficiency virus type I with viral nucleic acids following acute infection. *Proc. Natl Acad. Sci. USA*, **90**, 6125–6129.

Dorfman,T., Bukovsky,A., Ohagen,A., Hogland,S. and Gottlinger,H.G. (1994) Functional domains of the capsid protein of human immunodeficiency virus type 1. *J. Virol.*, **68**, 8180–8187.

Driscoll,P.C., Clore,G.M., Marion,D., Wingfield,P.T. and Gronenborn, A.M. (1990) Complete resonance assignment for the polypeptide backbone of interleukin-1-β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry*, **29**, 3542–3556.

Ehrlich,L.S., Agresta,B.E. and Carter,C.A. (1992) Assembly of recombinant human immunodeficiency virus type 1 capsid protein *in vitro*. *J. Virol.*, **66**, 4874–4883.

Fäcke,M., Janetzko,A., Shoeman,R.L. and Kräusslich,H.-G. (1993) A large deletion in the matrix protein of the human immunodeficiency virus *gag* gene redirects virus particle assembly from the plasma membrane to endoplasmic reticulum. *J. Virol.*, **67**, 4972–4980.

Forgacs,E. (1994) Study of the binding of sodium dodecylsulfate to dibasic amino acids by reversed-phase chromatograhy. *Fresenius J. Anal. Chem.*, **349**, 743–745.

Galley,P., Swingler,S., Aiken,C. and Trono,D. (1995) HIV-1 infection of nondividing cells—C-terminal tyrosine phosphorylation of the viral matrix protein is a key regulator. *Cell*, **80**, 379–388.

Gelderblom,H.R. (1991) Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS*, **5**, 617–637.

Gelderblom,H.R., Ozel,M. and Pauli,G. (1989) Morphogenesis and morphology of HIV—structure–function relations. *Arch. Virol.*, **106**, 1–13.

Gonzalez,S.A., Affrachino,J.L., Gelderblom,H.R. and Burny,A. (1993) Assembly of the matrix protein of simian immunodeficiency virus into virus-like particles. *Virology*, **194**, 548–556.

Grimes,J., Basak,A.K., Roy,P. and Stuart,D. (1995) The crystal-structure of bluetongue virus VP7. *Nature*, **373**, 167–170.

Grzesiek,S. and Bax,A. (1992a) Improved three-dimensional triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Resonance*, **96**, 432–440.

Grzesiek,S. and Bax,A. (1992b) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Resonance*, **99**, 201–207.

Harrison,S.C. (1992) Viruses. *Curr. Opin. Struct. Biol.*, **2**, 293–299.

Ina,Y. and Gojobori,T. (1990) Molecular evolution of human T-cell leukemia virus. *J. Mol. Evol.*, **31**, 493.

Inamura,J., Tsulimoto,A., Ohta,Y., Hirose,S., Shimotohno,K., Miwa,M. and Miyoshi,I. (1988) DNA blotting analysis of human retroviruses in cerebrospinal fluid of spastic paraparesis patients: the viruses are identical to human T-cell leukemia virus type 1 (HTLV-1). *Int. J. Cancer*, **42**, 493–499.

Kraulis,P.J. (1991) Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn, A.M. and Clore,G.M. (1989) Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann–Hahn multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1. *Biochemistry*, **28**, 6150–6156.

Massiah,M.A., Staricg,M.R., Paschall,C., Summers,M.F., Christensen, A.M. and Sundquist,W.I. (1994) The three-dimensional structure of the human immunodeficiency virus type I matrix protein. *J. Mol. Biol.*, **244**, 198–223.

Matthews,S., Barlow,P., Boyd,J., Barton,G., Russell,R., Mill,H., Cunningham,M., Meyers,N., Burns,N., Clark,N., Kingsman,S., Kingsman,A. and Campbell,I. (1994) Structural similarity between the p17 matrix protein of HIV-1 and interferon-γ. *Nature*, **370**, 666–668.

Matthews,S., Barlow,P.N., Clark,N., Kingsman,S., Kingsman,A. and Campbell,I. (1995) The refined solution structure of the HIV-1 matrix protein, p17. *Biochem. Soc. Trans.*, **23**, 725–728.

Mcinnes,C., Kay,C.M., Hodges,R.S. and Sykes,B.D. (1994) Conformational differences between *cis* and *trans* proline isomer of a peptide antigen representing the receptor binding domain of *Pseudomonas aeruginosa* as studied by ^{1}H NMR. *Bioploymers*, **34**, 1221–1230.

Nermut,M.V., Hockley,D.J., Jowett,J.B.M., Jones,I.A., Garreau,M. and Thomas,D. (1994) Fullerene-like organisation of HIV gag-protein shell in virus-like particles produced by recombinant baculovirus. *Virology*, **198**, 288–296.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association—insight from the interfacial and thermodynamic properties of hydrocarbon. *Proteins: Struct. Funct. Genet.*, **11**, 281–296.

Nilges,M., Gronenborn,A.M. and Clore,G.M. (1988) Determination of 3-dimensional structures of proteins by simulated annealing with interproton distance restraints—application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor-2. *Protein Eng.*, **2**, 27–38.

Palker,T.J. (1992) Human T-cell lymphotropic viruses: review and prospects for antiviral therapy. *Antiviral Chem. Chemother.*, **3**, 127–139.

Poiesz,B.J., Ruscetti,F.W., Gazdar,A.F., Bunn,P.A., Minna,J.D. and Gallo,R.C. (1980) Detection and isolation of type C retrovirus particle from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl Acad. Sci. USA*, **77**, 7415–7419.

Poznanski,J., Ejchart,A., Wierzchowski,K.L. and Ciurak,M. (1993) ^{1}H and ^{13}C NMR investigation on *cis–trans* isomerisation of proline peptide bonds and conformation of aromatic side-chains in H-Trp-$(Pro)_n$-Tyr-OH peptides. *Biopolymers*, **33**, 781–795.

Rao,Z.H., Belyaev,A.S., Fry,E., Roy,P., Jones,I.M. and Stuart,D.I. (1995) Crystal structure of SIV matrix antigen and its implications for virus assembly. *Nature*, **378**, 743–747.

Rhee,S.S. and Hunter,E. (1990) A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. *Cell*, **63**, 77–86.

Russell,R.B. and Barton,G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison—assignment of global and residue confidence levels. *Proteins: Struct. Funct. Genet.*, **14**, 309–323.

Smith,D.B. and Johnson,K.S. (1988) Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene*, **67**, 31–40.

Spearman,P., Wang,J., Heyden,N.V. and Ratner,L. (1994) Identification of human immunodeficiency virus type I gag protein domains essential to membrane binding and particle assembly. *J. Virol.*, **68**, 3233–3242.

Stuart,D. (1993) Viruses. *Curr. Opin. Struct. Biol.*, **3**, 167–174.

Wills,J.W. and Craven,R.C. (1991) Form, function, and the use of retroviral gag proteins. *AIDS*, **5**, 639–654.

Yuan,X., Yu,X., Lee,T. and Essex,M. (1993) Mutations in the N-terminal region of human immunodeficiency virus type I matrix protein block intracellular transport of the gag precursor. *J. Virol.*, **67**, 6387–6394.

Zhou,W., Parent,L.J., Willis,J.W. and Resh,M.D. (1994) Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type I gag protein which interacts with acidic phospholipids. *J. Virol.*, **68**, 2556–2569.

Received on December 18, 1995; revised on March 8, 1996