**Structure of a Spumavirus Gag central domain reveals a cryptic retroviral capsid**

Neil J. Ball1¶, Giuseppe Nicastro1¶, Moumita Dutta1, Dominic Pollard1&#a, David C. Goldstone1&#b, Marta Sanz-Ramos1, Dirk Lindemann2, Jonathan P. Stoye1,3 William R. Taylor1, Peter B. Rosenthal1, and Ian A. Taylor1\*

1The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London, UK.

2Institute of Virology, Technische Universität Dresden, Fetscherstr. 74, 01307 Dresden, Germany.

3Faculty of Medicine, Imperial College London, London, UK.

#aCurrent address: Faculty of Life Sciences, Imperial College London, London, UK.

#bCurrent address: Laboratory of Structural Biology, School of Biological Sciences, University of Auckland, Auckland, New Zealand.

\*Corresponding author

E-mail: [Ian.Taylor@crick.ac.uk](mailto:Ian.Taylor@crick.ac.uk) (IAT)

¶These authors contributed equally to this work

&These authors also contributed equally to this work

**Abstract (Max 300 words)**

The *Spumaretrovirinae,* or foamyviruses (FVs) are complex retroviruses that infect many species of monkey and ape. Despite little sequence homology, FV and orthoretroviral Gag proteins perform equivalent functions, including genome packaging, virion assembly, trafficking and membrane targeting. However, presently, there is a paucity of structural information with regards FVs and it is unclear how disparate FV and orthoretroviral Gag molecules share the same function. In order to probe the functional overlap of FV and orthoretroviral Gag we have determined the structure of a central domain of Gag from the Prototypic Foamyvirus (PFV). The structure (PFV-CA) comprises two all α-helical domains that although have no sequence similarity share the same core fold as the N- and C-terminal domains of an archetypal orthoretroviral CA protein. Comparison with other orthoretroviruses places PFV-CA structurally closest to that of the CA of gamma-retroviruses and further *in vitro* and functional virological assays show that PFV-CA is required for virus assembly and disassembly. These data provide the first information that relates the Spuma and Orthoretrovirinae and suggests a common ancestor for both lineages contained an ancient CA fold.

**Author Summary (Max 200 words)**

Foamyviruses (FVs) or Spuma-retroviruses derive their name from the cytopathic effects they cause in cell culture. By contrast, infection in humans is benign and FVs have entered the human population through zoonosis from apes resulting in the emergence of Prototypic Foamyvirus (PFV). Like all retroviruses FVs contain *gag*, *pol* and *env* structural genes and replicate through reverse-transcription and host genome integration. Gag, the major structural protein, is required for genome packaging, virion assembly, trafficking and egress. However, although functionally equivalent, FV and orthoretroviral Gag share little sequence homology and it is unclear how they perform the same function. Therefore, to understand more about relationship between FV and orthoretroviral replication we have carried out structural study of PFV Gag. Here we present the structure of a central domain from PFV-Gag and show that despite little sequence similarity it shares the same fold as the CA domain of orthoretroviral Gag. These data provide the first information that relates the Spuma and Orthoretrovirinae Gag and we discuss our findings in terms of evolutionary divergence of Spuma and Orthoretroviral lineages.

**Introduction**

Spuma- or foamy viruses (FVs) are complex retroviruses and constitute the only members of the *Spumaretrovirinae* subfamily within the *Retroviridae* family. They have been isolated from a variety of primate hosts [[1-4](#_ENREF_1)] as well as from cats [[5-7](#_ENREF_5)], cattle [[8](#_ENREF_8)], horses [[9](#_ENREF_9)] and sheep [[10](#_ENREF_10)]. Endogenous FVs have also been described in sloth [[11](#_ENREF_11)], aye-aye [[12](#_ENREF_12)] and coelacanth [[13](#_ENREF_13)]. Prototypic foamyvirus (PFV) is a FV isolated from human sources [[14](#_ENREF_14), [15](#_ENREF_15)]. The PFV genome is highly similar to that of isolates of simian foamy virus from chimpanzee (SFVcpz) and so infection in humans is believed to have arisen through a zoonotic transmission [[16-18](#_ENREF_16)]. Nevertheless, even though FVs are endemic within non-human primates and display a broad host range, human-to-human transmission of PFV has never been detected. Moreover, although in cell culture FV infection causes pronounced cytopathic effects [[19](#_ENREF_19)] infection in humans is apparently asymptomatic [[20-22](#_ENREF_20)] making their usage as vectors for gene therapy an attractive proposition [[23](#_ENREF_23)].

FVs share many similarities with other retroviruses in respect of their genome organisation and life cycle. However, they vary from the *Orthoretrovirinae* in a number of important ways. These include the timing of reverse transcription that occurs in virus producer cells rather than newly infected cells [[24](#_ENREF_24), [25](#_ENREF_25)] and the absence of a Gag-pol fusion protein [[26](#_ENREF_26), [27](#_ENREF_27)]. In addition, the Gag protein remains largely unprocessed in FVs [[28](#_ENREF_28)] whereas within the *Orthoretrovirinae* processing of the Gag polyprotein represents a critical step in viral maturation, producing the internal structural proteins Matrix (MA), Capsid (CA) and Nucleocapsid (NC) found in mature virions. Furthermore, FV Gag lacks the Major Homology Region (MHR) and Cys-His boxes found in orthoretroviral CA and NC, respectively. Despite these profound dissimilarities, the Gag proteins of the two retroviral subfamilies carries out the same functional roles including viral assembly, nucleic acid packaging, transport to and budding through the cytoplasmic membrane of the producer cell as well as trafficking through the cytoplasm of the target cell and uncoating. In addition, FV Gag also contains the determinants for restriction by Trim5 [[29](#_ENREF_29), [30](#_ENREF_30)] that in orthoretroviruses comprises the assembled CA lattice [[31](#_ENREF_31)].

To date, high-resolution X-ray and/or NMR structures of have been reported for MA, CA and NC components of Gag from numerous retroviruses [[32-41](#_ENREF_32)] but for FVs only the structure of the Env-binding N-terminal domain of PFV-Gag has been reported [[42](#_ENREF_42)]. Further structural information with regard to other Gag domains of FVs has remained elusive and is vital for any detailed understanding of how FV Gag fulfils its many functions. Here we report the NMR structure of a two-domain central region from the Gag of PFV that is it is required for the assembly of foamyviruses. Our data reveal that although unrelated at the level of primary sequence, the central domains of spumavirus Gag are structurally related to the N- and C-terminal domains of the CA protein of other retroviruses. Further phylogenetic analysis reveals the PFV Gag central domains are most closely related to that of the gamma retroviruses. These data suggest a common ancestor for retroviral Gag and we propose a role for gene duplication in the evolution of retroviral CA-assembly domains.

**Results**

**Structure of the PFV-Gag central conserved region.**

Alignment of the primary sequences of FV Gag proteins from primate and other mammalian hosts reveals two regions of strong conservation, an N-terminal region corresponding to the Env-binding domain [[42-44](#_ENREF_42)] and the other located centrally (**Fig. 1A**). Within this central region, large sections of highly conserved sequence are present. Therefore, to understand more about the nature of the PFV-Gag central conserved region, the solution structure of PFV-Gag(300-477) was determined using multidimensional heteronuclear NMR spectroscopy. Details of data collection, structure determination and model quality are presented in **Table 1.**

The structure is the only example of a central region of spumaretroviral Gag and comprises nine α-helices packed into two domains, connected by a short 5-residue linker (**Fig 1B**). Superposition of the 20 conformers in the family of structures results in a backbone atom rmsd of 0.3 Å for ordered residues 304-355, 358-477 showing that the structure is well defined except for the N- and C-termini and loop regions (**Fig. S1**). Residues P300-H383 make up the N-terminal domain (PFV-NtDCEN) containing four helices (α1-α4) and the C-terminal domain (PFV-CtDCEN), residues G384-R477, contains the remaining five helices (α5-α9). In PFV-NtDCEN, helices α1-α3 form an antiparallel 3-helix bundle connected to α4 by a long loop that closely tracks one face of α3. In PFV-CtDCEN helices α5-α9 are arranged as a five-helix antiparallel bundle. In both domains, the inner faces of the helices pack to form an extensive hydrophobic core through interaction of apolar sidechains.

Examination of the protein backbone dynamics using 15N NMR relaxation measurements (**Fig. S1**), show that residues within helices α1- α4 and α5-α9 of the PFV-NtDCEN and PFV-CtDCEN exhibit the large and positive heteronuclear NOE (HetNOE) values and have uniform 15N T1 and T2 values indicating a rigid backbone. Additionally, the presence of inter-domain NOEs, together with little variation in the T1/T2 values, suggests the PFV-NtDCEN and PFV-CtDCEN are structurally and dynamically dependent and have a coupled movement. Based on these relaxation rates and assuming an isotropic model, a rotational correlation time (tc) of 14.1 ns for the NtDCEN-CtDCEN di-domain was determined, consistent with a ~ 20 kD globular protein. The residues at the N- and C-termini outside of this core region have lower T1 and higher T2 values, reduced or negative HetNOEs, close to zero 1DNH residual dipolar couplings (RDC) and mainly random coil chemical shifts indicating rapid (psec) internal motion in these terminal regions. In addition, the relaxation data also reveals internal regions of high mobility, including residues G356 to G366 located in the long loop connecting α3-α4, residues G384 to P388 in the NtDCEN-CtDCEN interdomain linker and G432 part of a stretch of highly conserved residues (-P431-**G**-Q-A434-) located in the loop connecting α7-α8 of CtDCEN and in close spatial proximity to a conserved Y/F464-x-x-L-G-L469 motif, at the C-terminus of α9 (**Fig. 1A**) that is required for Gag assembly [[45](#_ENREF_45)]. Together with these relaxation data a number of interdomain NOEs define a largely hydrophobic NtDCEN-CtDCEN interface comprising 550Å2 of buried surface area (**Fig 1B**). Although not extensive in area, there is substantial packing of apolar sidechains from NtDCEN residues on helices α2 and α4 (I326, V375 and F379) with CtDCEN residues (L410, M413 and L414) on helices α5 and α6 (**Fig. 1C**) that contribute to the stability of the interface.

The nine helices present in PFV-Gag(300-477) are α1, Q308-T314; α2, E323-V338; α3, P344-L355; α4, W371-H383; α5, H389-E402; α6, V404-S415; α7, Y419-Y429; α8, Q433-Q445; α9, N449-L467.

**Structural similarity with other retroviral genera**

Structural similarity searches of the PDB with PFV-Gag(300-477), PFV-NtDCEN and PFV-CtdCEN were conducted using both the DALI search engine [[46](#_ENREF_46)] and the SSM server [[47](#_ENREF_47)]. Although only very weak (Q-scores 0.1 - 0.3) these searches did reveal that individual NtDCEN and CtDCEN domains displayed structural similarities with amino- or carboxyl-terminal domains from orthoretroviral CAs (**S1 Table**). Notably, the best alignments were between NtDCEN with orthoretroviral CA-CtD domains and by CtDCEN with orthoretroviral CA-NtD domains. Structural alignments of PFV-NtDCEN with four orthoretroviral CA-CtD structures are shown in **Fig. 2A-D**. It is apparent that the overall topology is conserved with closest match based upon rmsd over all α-carbons with the CA-CtD of the alpha-retrovirus RSV (3G1G). However, all the orthoretroviral CA-CtD structures contain an additional α-helix that inserts between α3 and α4 of NtDCEN (**Fig. 2A-D**). Residues on this helix make interactions at 3-fold interfaces within the assembled capsid in HIV-1 [[48](#_ENREF_48), [49](#_ENREF_49)]. As PFV-NtDCEN and its absence in foamy virus possibly relates to differences in the assembly properties of spuma- and orthoretroviral capsids.

Structural alignments of CtDCEN with the CA-NtD from five classes of retrovirus are shown in **Fig. 3E-I**. These alignments reveal that helices α5 to α9 of PFV-CtDCEN have the same core fold and topological arrangement as the five helical bundle present in the CA-NtD of orthoretroviruses. The closest match based upon rmsd over all α-carbons is with CA-NtD of the gamma-retrovirus MLV (3BP9). Notably, although the core fold aligns well the interspersing loops that connect the secondary structure elements in the orthoretroviral CA-NtD are absent or much shorter in CtDCEN. Therefore given these differences and only relatively weak alignments in order to quantify the degree of similarity we applied a decoy mode approach combined with SSAP analysis [[50](#_ENREF_50), [51](#_ENREF_51)]. These data..... WT from here

* Evolutionary perspective from Willie Taylor (inc. figure 4)?
  + Both FV and orthoretroviruses started out with two 5-helix bundles via a duplication event
  + Divergent evolution based on selective pressure?

**Solution conformation of Foamyvirus Gag-central domains**

To gain insight into the assembly properties of PFV-Gag central domains, PFV-Gag(300-477), together with PFV-Gag(300-387) and PFV-Gag(381-477)] that encompass PFV-NtDCEN and PFV-CtDCEN were analysed by velocity (SV-AUC) and equilibrium (SE-AUC) analytical ultracentrifugation. The experimental parameters, molecular weights derived from the data and statistics relating to the quality of fits are shown in **Table 2**.

SV-AUC analysis of the whole of the conserved region, PFV-Gag(300-477), revealed a sedimentation coefficient (S20,w) of 1.87 (**Fig. 3A**) and derived molar mass of 20.6 kDa demonstrating that PFV-Gag(300-477) is a stable monomer in solution. These observations were confirmed by multispeed SE-AUC at varying protein concentration. The equilibrium distribution from an individual multispeed experiment is presented in **Fig. 3B**. The individual gradient profiles showed no concentration dependency of the molecular weight and fit globally with a single ideal molecular species model, producing weight averaged molecular weight of 20.3 kDa demonstrating the monomeric nature of this PFV central region. SV-AUC analysis of PFV-Gag NtDCEN also revealed this domain to be monomeric in solution with (S20,w) of XXX (**Fig. 2A**) and derived molar mass of YYY kDa By contrast SV-AUC data recorded on PFV-Gag(381-477) produced a sedimentation coefficient continuous distribution function [C(S)], that contained two species, with S20,w of 1.65 and 2.07 (**Fig. 2A**) and were the proportion of the fast 2.07 S, component increased with increasing concentration consistent with monomer-dimer equilibria. The affinity of self-association was analysed by multispeed SE-AUC recorded at varying protein concentration. These data (**Fig. 2B**) are best fit by self-association models where the 11.9 kDa PFV-Gag(381-477) monomers dimerise with an equilibrium association constant of 1.1x106 M-1 (0.9 µM KD). These data are consistent with the distribution of peaks in the C(S) functions derived from SV-AUC data. Moreover, they reveal that whilst the entire PFV-Gag central region is monomeric the CtD, residues has the propensity for self-association.

**Structure of the PFV-Gag CtdCEN homodimer.**

Based upon our solution observations and the significant structural homology of the core of CtDCEN with the structures of CA-NtDs of orthoretroviruses we determined the structure of PFV-Gag(381-477) homodimer. Details of data collection and structure determination are presented in **Table 2**. The structure comprises five-antiparallel α-helices (residues N393-E402, V404-L414, Q420-Y429, Q433-Q445 and Q450-L467) corresponding to α5 to α9 in the PFV-Gag(300-477). This core structure is virtually identical to the equivalent helices in PFV-Gag(300-477) with the exception that α5 is ~2 turns shorter. The CtDCEN dimer interface is largely hydrophobic and comprises 470 Å2 of buried surface. Apolar interactions between residues on α5 and α6 make up the majority of the interface where α6 of one monomer packs against α6 of the other monomer with L410 and M413 making reciprocal interactions (**Fig. S2**). An L410E/M413E double mutant completely abolished dimerisation as determined by SV-AUC. Same residues involved in both interfaces

**Mutational analysis if the NtD-CtD interface**

Mutations were made at V375, L410 and M413 in the PFV-Gag NtDCEN – CtDCEN interface. These were examined for infectivity and where viral particles were produced the morphology and integrity of the particle assessed by cryo-electron microscopy. In addition, the effect of a double L410Q/M413Q on homodimer formation by PFV-Gag CtDCEN was assessed by sedimentation velocity AUC.

**Infectivity**

The capacity for mutants at the central domain dimer interface...

**Discussion**

**The Foamyvirus Gag contains an orthoretroviral CA-NtD domain**

Inspection of the core of the PFV Gag NtDCEN-CtDCEN di-domain revealed unanticipated structural similarities to that of the CA-NtD and CA-CtD of orthoretroviruses. Whilst the core 5 helix bundles are structurally very similar, orthoretroviral CA-NtDs contain additional features located on the capsid outer surface. When compared with B-MLV CA (Mortuza et al., 2008; PDB accession code 3BP9), the PFV Gag central domain lacks the N-terminal β hairpin, the transverse helix (between α6 and α7) and α7 is significantly shorter (**Fig. S6**). The determinants for restriction of N-MLV by Trim5α have been mapped to these additional regions which sit on the ‘top’ surface of the capsid [[52](#_ENREF_52)]. In FV Gag, however, the N-terminal domain has been shown to be responsible for resistance to Trim5α [[42](#_ENREF_42)] and is also situated on the outer surface of the capsid.

* Cryptic CA in PFV is only responsible for capsid assembly

## Potential role of the PFV Gag central domain in capsid disassembly

* Discuss prominent secondary cleavage site of PFV
  + Occurs between R311/S312 (Lehmann-Che et al., 2005)
  + May permit structural rearrangement similar to orthoretroviral maturation
  + *In vitro* cleavage reveals PFV(1-310) is a dimer and (311-477) is a monomer
    - CD-NtD may occlude CD-CtD dimer interface
    - Would destabilise capsid and permit nuclear import...
  + PFV-Gag(300-477) represents initial cleavage state
* PFV-Gag(340-477) represents a cryptic secondary cleavage site fragment - still a monomer

**The YxxLGL motif**

The YxxLGL motif is a conserved feature of the Gag protein in all spumaretroviruses and has been proposed to be important for assembly [[45](#_ENREF_45)]. In the structure of the PFV central domain this motif is found at the C-terminus of CtDCEN (residues Y464-L469). Y464 resides in a hydrophobic pocket and forms part of the core of the antiparallel helical bundle (**Fig. S5**) and there is limited capacity for hydrogen bonding between the hydroxyl group and the peptide backbone. The tyrosine is not completely conserved between different FVs (**Fig. 1A**) but it appears that the minimum requirement is that there is a phenyl group at this position that can be accommodated by the hydrophobic pocket.

Residues 465-469, along with the completely conserved PGQA (residues 431-434; **Fig. 1A**) form a patch on the surface of the PFV Gag central domain (**Fig. S5**). This patch is located on the opposite surface of the protein to the dimer interface and is exposed during self-association. If PFV Gag assembles in a similar fashion to mature orthoretroviral CA [[53](#_ENREF_53)], this would place this patch at the two-fold axis between hexamers. However, if PFV Gag assembles in a similar fashion to immature orthoretroviral CA [[54](#_ENREF_54)], this would permit the YxxLGL motif to pack against the PGQA patch in a neighbouring monomer within the hexamer. In either case, this may explain why altering the YxxLGL motif has severe ramifications for PFV assembly and infectivity.

**Capsid Structure and restriction**

Members of the Trim5α family of restriction factors block infection of cells by HIV-1, as well as other lentiviruses, orthoretroviruses and the foamyviruses The mechanism of this restriction requires interaction of Trim5α with CA component of Gag in the context of an assembled capsid shell [[55](#_ENREF_55)]. However, whilst some of the determinants for orthoretroviral CA recognition have been elucidated [[52](#_ENREF_52), [56](#_ENREF_56)] the molecular determinants for Trim5α restriction of foamyviruses remain largely unmapped. Moreover, given the lack of sequence identity and structural homology between the orthoretroviral and FV Gag it is also unclear how some Trim5α species-types can recognise and restrict such a wide diversity of viruses. Our structural analysis of PFV Gag has now somewhat surprisingly revealed that FVs also contain a CA region that comprises two domains with folds related to the CA-NtD and CtD of orthoretroviral Gag. In the orthoretroviruses the CA-NtDs form an important part of the hexamer that serves as the basic building block for core assembly and as a pathogen-associated molecular pattern (PAMP) that is the target for Trim5α recognition. Given this unexpected similarity between spuma and orthoretroviral gag, our studies now prompt a number of questions. First, are hexamers present in FV virions? More detailed structural studies will be required to answer this question. Second, might such domains form part of the ordered assemblies of other viruses? Perhaps a survey of viral structural proteins present in the PDB will answer this question. Third, might these findings suggest that the capsid binding restriction factors TRIM5α and Fv1 recognize viruses other than Retroviridae containing such domains? It has been puzzling how restriction factors can recognize viruses as diverse as HIV-1 but perhaps a hexameric target is key to recognition. TRIM5α forms a hexamer and Fv1 has a number of shared organisational features. If so, hexamers might represent a pathogen associated molecular pattern present on multiple families of virus.

**Methods**

**Protein Expression and purification**

The DNA sequences coding for PFV-Gag residues 300-477, 300-381, and 381-477 were amplified by PCR from template plasmid pcziGag4 containing the PFV Gag gene. PCR products were inserted into a pET22b expression vector (Novagen) using the NdeI and XhoI restriction sites in order to produce C-terminal His-tag fusions. The correct sequence of expression constructs was verified by automated DNA sequencing (GATC Biotech). His-tagged PFV constructs were expressed in the *E. coli* strain Rosetta 2 (DE3) and purified using Ni-NTA affinity (Qiagen) and size exclusion chromatography (SEC) on Superdex 75 (GE healthcare). For NMR studies proteins were grown in minimal media supplemented with 15NH4Cl, 13C-Glucose and/or 2H2O and purified as previously described. For PFV-Gag(340-482) an ion exchange step (mono Q) was added between the affinity and SEC steps.

**NMR Spectroscopy**

All NMR experiments were carried out at 298 K on Bruker Avance 600-, 700-, 800-, and 950-MHz spectrometers. 1H/2H, 13C 15N-labeled PFV-Gag samples (residues 300-477 and 381-477) were prepared in buffer containing 20mM Tris-HCl, 20 mM NaCl, 0.5 mM TCEP pH 7.0. Protein concentrations for the NMR experiments were ~300 μM for PFV-Gag(300-477) and 1.6 – 2 mM for PFV-Gag CtDCEN . 1H, 13C and 15N resonance assignments for protein backbone were obtained from three-dimensional HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, CBCA(CO)NH, HNCO, HN(CA)CO experiments. For side-chain chemical shift assignments 3D HBHA(CO)NH, CC(CO)NH, H(CCO)NH, (H)CCH-TOCSY, and CCHTOCSY spectra were also acquired. In addition, aromatic side-chain resonances were assigned from the analysis of the 1H-13C HSQC tuned to aromatic carbons, 2D (HB)CB(CGCD)HD, 2D (HB)CB(CGCDCE)HE as well as 3D 13C-edited NOESY-HSQC tuned to aromatic carbons. Inter-proton distance restraints for structural calculations were obtained from 3D 13C-edited NOESY-HSQC and 15N-edited NOESY-HSQC spectra recorded using a 100 ms mixing time. The dimer interface of PFV-Gag(381-477) was identified by intermolecular distance restraints using 13C/15N-filtered 13C-edited NOESY spectra. The 3D-filtered spectra were obtained using an asymmetrically labelled dimer of HPFV-Gag(381-477) prepared by mixing equimolar unlabelled protein with uniformly 13C/15N-labeled protein (1.6 mM total protein concentration). For residual dipolar coupling (RDC) measurements, weakly aligned 15N-labelled samples of PFV-Gag(300-477) (200 µM) and PFV-Gag(381-477) (2 mM) were prepared by the addition of 10mg/ mL filamentous phage Pf1 (ASLA Biotech Ltd, Latvia). Sample instability precluded the measurement of RDCs on PFV-Gag(340-477) samples. 1D NH RDCs were measured using the In-Phase and Anti-Phase method [[57](#_ENREF_57)]. The RDC values were obtained by subtracting the reference value in isotropic solution. All spectral data were processed with NMRPipe [[58](#_ENREF_58)] and analyzed with CARA [[59](#_ENREF_59)].

**Protein structure determination**

The solution structures for PFV-Gag(300-477) and the PFV-Gag CtDCEN dimer were calculated using the program ARIA (Ambigious Restraints for Iterative Assignment v 2.3) [[60](#_ENREF_60)]. Nine iterations of progressive assignment and structure calculation combined with NOE distance restraints, hydrogen bonds, dihedral angle restraints, predicted by the TALOS program [[61](#_ENREF_61)] and RDC measurements were employed in a simulated annealing protocol. For the PFV-Gag CtDCEN homodimer the inter-proton NOE-derived distance restraints present in the filtered NOESY experiments were defined as intermolecular and the corresponding NOEs removed from the 3D 13C-NOESY-HSQC.

Initial structures were used to determine the axial and rhombic components of the alignment tensors with the program MODULE [[62](#_ENREF_62)]. Subsequently, the RDC restraints were added in the final refinement stage of structure calculations. Only data for residues located in rigid secondary structure elements (1H-15N NOE > ~0.75) were employed. A final ensemble of the 20 lowest energy structures derived from 100 calculated structures and refined in an explicit water box in the last iteration was selected. The superimposition of the 20 lowest-energy structures and the ribbon diagram of one representative structure are shown in **Supplementary figure 3**. The quality of the calculated structure ensembles were assessed and validated with the Protein Structure Validation Suite-PSVS [[63](#_ENREF_63)] and Procheck-NMR [[64](#_ENREF_64)]. For the final 20 lowest-energy NMR structures, no distance or torsional angle restraint was violated by more than 0.5 Å or 5°, respectively. Structure determination are summarised in **Table 2**.

**15N Relaxation measurements**

The backbone 15N relaxation parameters of the spin-lattice relaxation time T1, the spin-spin relaxation time T2 and the steady-state heteronuclear 1H-15N NOE relaxation were determined at 25°C on a 700 MHz spectrometer using a 15N-labeled NMR samples for PFV-Gag(300-477). The time delays used for T1 experiments were 10, 50, 100, 200, 400, 500, 750, 1000, and 1400 ms, and those for T2 experiments were 8, 16, 32, 48, 64, 80, 96, 112, 128 and 160 ms. The T1 and T2 relaxation data were obtained by fitting the individual peak intensities using nonlinear spectral lineshape modelling and fitted to single exponential using routines within NMRPipe [[58](#_ENREF_58)]. 1H-15N NOE values were calculated from peak intensity ratios obtained from spectra with and without 1H saturation prior to the 15N excitation pulse.

**Structure alignment and comparisons**

The DALI search engine (<http://ekhidna.biocenter.helsinki.fi/dali_server/>) and protein structure comparison service (SSM) at the European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm/>) were used to search for structural homologues from the PDB. PFV-Gag NtDCEN and CtDCEN were superimposed upon orthoretroviral CA NtD and CtDs using SUPERPOSE [[47](#_ENREF_47)] from the ccp4 program package. The quality of the fit was ranked using the Q-score (Table S1). WT from here??

Structural alignments produced using the SAP program [[51](#_ENREF_51)]. SAP uses a local structural environment based measure of comparison and compared to the raw rmsd measure is less sensitive to local structural variation. However, the similarity scores that are returned must be inverted before being used for clustering. Inversion was calculated using a Gaussian function: exp(-s^2/a), where "s" is the SAP score and "a" (which sets the spread of the bell-curve) was adjusted to give the maximum spread of values. The value a = 500 was used.

Dendrograms were calculated using the hierarchical clustering function "hclust()" in the statistical package "R" (<http://www.r-project.org/>). Dendrograms were calculated based on average, centroid and complete distances and only minor branch length changes were seen with no topological rearrangement. The trees produced using the "complete" option are presented.

**Analytical Ultracentrifugation**

Sedimentation velocity experiments were performed in a Beckman Optima Xl-I analytical ultracentrifuge using conventional aluminium double sector centrepieces and sapphire windows. Solvent density and the protein partial specific volumes were determined as described [[65](#_ENREF_65)]. Prior to centrifugation, samples were prepared by exhaustive dialysis against the buffer blank solution, 20 mM Tris-HCl pH 8, 150 mM NaCl and 0.5 mM TCEP (Tris Buffer). Centrifugation was performed at 50,000 rpm and 293 K in an An50-Ti rotor. Interference data were acquired at time intervals of 180 s at varying sample concentration (0.5-2.0 mg/ml). Data recorded from moving boundaries was analysed in terms of the size distribution functions C(S) using the program SEDFIT [[66-68](#_ENREF_66)].

Sedimentation equilibrium experiments were performed in a Beckman Optima XL-I analytical ultracentrifuge using aluminium double sector centrepieces in an An-50 Ti rotor. Prior to centrifugation, samples were dialyzed exhaustively against the buffer blank (Tris Buffer). After centrifugation for 30 h, interference data was collected at 2 hourly intervals until no further change in the profiles was observed. The rotor speed was then increased and the procedure repeated. Data were collected on samples of different concentrations of PFV-Gag(300-477) and PFV-Gag(381-477) at three speeds and the program SEDPHAT [[69](#_ENREF_69), [70](#_ENREF_70)] was used to determine weight-averaged molecular masses by nonlinear fitting of individual multi-speed equilibrium profiles to a single-species ideal solution model. Inspection of these data revealed that the molecular mass of PFV-Gag(300-477) showed no significant concentration dependency and so global fitting incorporating the data from multiple speeds and multiple sample concentrations was applied to extract a final weight-averaged molecular mass. For PFV-Gag(381-477) the molecular masses showed significant concentration dependency and so global fitting of a monomer-dimer equilibrium model incorporating the data from multiple speeds and multiple sample concentrations was applied to extract the dimerisation association constant (KA).

**Particle release and infectivity assay**

Cell culture supernatants containing recombinant viral particles were generated as described previously [[71](#_ENREF_71)]. Briefly, 293T cells were co-transfected in 10 cm dishes with a Gag expression plasmid (pcziGag4 or PG mutants thereof, as indicated), Env (pcoPE), Pol (pcoPP), and the transfer vector (puc2MD9) at a ratio of 16:1:2:16 using Polyethyleneimine (PEI) reagent and 16 µg DNA total. At 48 h post transfection (p.t.) cell-free viral vector supernatant was harvested using 0.45 μm sterile filters.

For transduction efficiency analysis 2 x 104 HT1080 cells were plated in 12-well plates 24 h before infection. The target cells were incubated with 1 ml of plain cell-free viral supernatant or serial dilutions thereof for four to six hours. Determination of the percentage of eGFP-expressing cells was performed 72 h after infection by flow cytometry analysis and used for titre determination as previously described [[72](#_ENREF_72)]. All transduction experiments were repeated at least three times. To compare the infectivity in repetitive experiments the titre obtained for wild type supernatants in individual experiments was set to an arbitrary value of 100%. The other values were then normalized as percentage of the wild type value.

Viral protein expression in transfected cells and particle-associated protein composition was examined by Western blot analysis. Preparation of cell lysates from one transfected 10-cm cell culture dish was performed by incubation with 0.6 ml lysis buffer for 20 min at 4°C followed by centrifugation through a QIAshredder (Qiagen). All protein samples were mixed with equal volumes of 2xPPPC (100 mM Tris-HCl; pH 6.8, 24 % glycerol, 8 % SDS, 0.2 % Bromophenol blue, 2 % ß-mercaptoethanol) prior to separation by SDS-PAGE using 7.5% polyacrylamide gels. Viral particles were concentrated from cell-free supernatant of transfected 293T cells by ultracentrifugation through a 20% sucrose cushion at 4°C and 25,000 rpm for 3 h in an SW32 rotor. The viral pellet was resuspended in phosphate-buffered saline (PBS). Immunoblotting using polyclonal antisera specific for PFV Gag [[73](#_ENREF_73)] was performed as previously described [[74](#_ENREF_74)]. The chemiluminescence signal was digitally recorded using a LAS3000 imager and quantified using ImageGauge in the linear-range of the sample signal intensities as described previously [[75](#_ENREF_75)].

**Acknowledgements**

We acknowledge the MRC UK biomedical NMR centre for access. The structure coordinates of PFV(300-477) and PFV-Gag CtDCEN were deposited in the Protein Data Bank under the accession codes XXX, YYY and ZZZ. Chemical shift assignments were deposited in the BioMagResBank, accession number MMM

**Author contributions**

N.J.B., G.N., D.P., D.C.G., M.S-R and I.A.T. performed experiments. N.J.B., G.N., D.P., D.C.G., M.S-R., D. L., J.P.S. and I.A.T. contributed to experimental design, data analysis and manuscript writing.

**Competing interests statement**

The authors declare no competing financial interests.

**References**

1. Lecellier CH, Saib A. Foamy viruses: between retroviruses and pararetroviruses. Virology. 2000;271(1):1-8. Epub 2000/05/18. doi: 10.1006/viro.2000.0216

S0042682200902162 [pii]. PubMed PMID: 10814564.

2. Herchenroder O, Renne R, Loncar D, Cobb EK, Murthy KK, Schneider J, et al. Isolation, cloning, and sequencing of simian foamy viruses from chimpanzees (SFVcpz): high homology to human foamy virus (HFV). Virology. 1994;201(2):187-99. Epub 1994/06/01. doi: S0042-6822(84)71285-2 [pii]

10.1006/viro.1994.1285. PubMed PMID: 8184531.

3. Bieniasz PD, Rethwilm A, Pitman R, Daniel MD, Chrystie I, McClure MO. A comparative study of higher primate foamy viruses, including a new virus from a gorilla. Virology. 1995;207(1):217-28. PubMed PMID: 7871729.

4. Broussard SR, Comuzzie AG, Leighton KL, Leland MM, Whitehead EM, Allan JS. Characterization of new simian foamy viruses from African nonhuman primates. Virology. 1997;237(2):349-59. Epub 1997/11/14. doi: S0042-6822(97)98797-3 [pii]

10.1006/viro.1997.8797. PubMed PMID: 9356346.

5. Saib A. Non-primate foamy viruses. Curr Top Microbiol Immunol. 2003;277:197-211. Epub 2003/08/12. PubMed PMID: 12908774.

6. Riggs JL, Oshirls, Taylor DO, Lennette EH. Syncytium-forming agent isolated from domestic cats. Nature. 1969;222(5199):1190-1. Epub 1969/06/21. PubMed PMID: 5788991.

7. Fabricant CG, Rich LJ, Gillespie JH. Feline viruses. XI. Isolation of a virus similar to a myxovirus from cats in which urolithiasis was experimentally induced. The Cornell veterinarian. 1969;59(4):667-72. Epub 1969/10/01. PubMed PMID: 4310934.

8. Malmquist WA, Van der Maaten MJ, Boothe AD. Isolation, immunodiffusion, immunofluorescence, and electron microscopy of a syncytial virus of lymphosarcomatous and apparently normal cattle. Cancer research. 1969;29(1):188-200. Epub 1969/01/01. PubMed PMID: 4974302.

9. Tobaly-Tapiero J, Bittoun P, Neves M, Guillemin MC, Lecellier CH, Puvion-Dutilleul F, et al. Isolation and characterization of an equine foamy virus. J Virol. 2000;74(9):4064-73. Epub 2001/02/07. PubMed PMID: 10756018; PubMed Central PMCID: PMC111920.

10. Flanagan M. Isolation of a spumavirus from a sheep. Aust Vet J. 1992;69(5):112-3. Epub 1992/05/01. PubMed PMID: 1320857.

11. Katzourakis A, Gifford RJ, Tristem M, Gilbert MT, Pybus OG. Macroevolution of complex retroviruses. Science. 2009;325(5947):1512. Epub 2009/09/19. doi: 10.1126/science.1174149. PubMed PMID: 19762636.

12. Han GZ, Worobey M. An endogenous foamy virus in the aye-aye (Daubentonia madagascariensis). Journal of virology. 2012;86(14):7696-8. Epub 2012/05/11. doi: 10.1128/JVI.00650-12. PubMed PMID: 22573860; PubMed Central PMCID: PMC3416287.

13. Han GZ, Worobey M. An endogenous foamy-like viral element in the coelacanth genome. PLoS pathogens. 2012;8(6):e1002790. Epub 2012/07/05. doi: 10.1371/journal.ppat.1002790. PubMed PMID: 22761578; PubMed Central PMCID: PMC3386198.

14. Switzer WM, Bhullar V, Shanmugam V, Cong ME, Parekh B, Lerche NW, et al. Frequent simian foamy virus infection in persons occupationally exposed to nonhuman primates. J Virol. 2004;78(6):2780-9. Epub 2004/03/03. PubMed PMID: 14990698; PubMed Central PMCID: PMC353775.

15. Heneine W, Schweizer M, Sandstrom P, Folks T. Human infection with foamy viruses. Curr Top Microbiol Immunol. 2003;277:181-96. Epub 2003/08/12. PubMed PMID: 12908773.

16. Herchenroder O, Turek R, Neumann-Haefelin D, Rethwilm A, Schneider J. Infectious proviral clones of chimpanzee foamy virus (SFVcpz) generated by long PCR reveal close functional relatedness to human foamy virus. Virology. 1995;214(2):685-9. Epub 1995/12/20. doi: S0042-6822(85)70086-4 [pii]

10.1006/viro.1995.0086. PubMed PMID: 8553577.

17. Achong BG, Mansell PWA, Epstein MA, Clifford P. An unusual virus in cultures from a human nasopharyngeal carcinoma. J Natl Cancer Inst. 1971;46:299-307.

18. Epstein MA. Simian retroviral infections in human beings. Lancet. 2004;364(9429):138-9; author reply 9-40. Epub 2004/07/13. doi: 10.1016/S0140-6736(04)16621-X. PubMed PMID: 15246722.

19. Rethwilm A. Molecular biology of foamy viruses. Med Microbiol Immunol. 2010;199(3):197-207. Epub 2010/05/07. doi: 10.1007/s00430-010-0158-x. PubMed PMID: 20445989.

20. Meiering CD, Linial ML. Historical perspective of foamy virus epidemiology and infection. Clin Microbiol Rev. 2001;14(1):165-76. Epub 2001/01/09. doi: 10.1128/CMR.14.1.165-176.2001. PubMed PMID: 11148008; PubMed Central PMCID: PMC88968.

21. Murray SM, Picker LJ, Axthelm MK, Hudkins K, Alpers CE, Linial ML. Replication in a superficial epithelial cell niche explains the lack of pathogenicity of primate foamy virus infections. J Virol. 2008;82(12):5981-5. Epub 2008/04/11. doi: JVI.00367-08 [pii]

10.1128/JVI.00367-08. PubMed PMID: 18400853; PubMed Central PMCID: PMC2395144.

22. Murray SM, Linial ML. Foamy virus infection in primates. Journal of medical primatology. 2006;35(4-5):225-35. Epub 2006/07/29. doi: JMP171 [pii]

10.1111/j.1600-0684.2006.00171.x. PubMed PMID: 16872286.

23. Lindemann D, Rethwilm A. Foamy virus biology and its application for vector development. Viruses. 2011;3(5):561-85. Epub 2011/10/14. doi: 10.3390/v3050561

viruses-03-00561 [pii]. PubMed PMID: 21994746.

24. Yu SF, Baldwin DN, Gwynn SR, Yendapalli S, Linial ML. Human foamy virus replication: a pathway distinct from that of retroviruses and hepadnaviruses. Science. 1996;271(5255):1579-82. Epub 1996/03/15. PubMed PMID: 8599113.

25. Moebes A, Enssle J, Bieniasz PD, Heinkelein M, Lindemann D, Bock M, et al. Human foamy virus reverse transcription that occurs late in the viral replication cycle. J Virol. 1997;71(10):7305-11. PubMed PMID: 9311807.

26. Löchelt M, Flügel RM. The human foamy virus pol gene is expressed as a Pro-Pol polyprotein and not as a Gag-Pol fusion protein. J Virol. 1996;70(2):1033-40.

27. Enssle J, Jordan I, Mauer B, Rethwilm A. Foamy virus reverse transcriptase is expressed independently from the Gag protein. Proc Natl Acad Sci U S A. 1996;93(9):4137-41. Epub 1996/04/30. PubMed PMID: 8633029; PubMed Central PMCID: PMC39500.

28. Flügel RM, Pfrepper KI. Proteolytic processing of foamy virus Gag and Pol proteins. Curr Top Microbiol Immunol. 2003;277:63-88.

29. Pacheco B, Finzi A, McGee-Estrada K, Sodroski J. Species-specific inhibition of foamy viruses from South American monkeys by New World Monkey TRIM5{alpha} proteins. J Virol. 2010;84(8):4095-9. Epub 2010/02/05. doi: JVI.02631-09 [pii]

10.1128/JVI.02631-09. PubMed PMID: 20130055.

30. Yap MW, Lindemann D, Stanke N, Reh J, Westphal D, Hanenberg H, et al. Restriction of foamy viruses by primate Trim5alpha. J Virol. 2008;82(11):5429-39. PubMed PMID: 18367529.

31. Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, et al. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proc Natl Acad Sci U S A. 2006;103(14):5514-9. PubMed PMID: 16540544.

32. Kingston RL, Fitzon-Ostendorp T, Eisenmesser EZ, Schatz GW, Vogt VM, Post CB, et al. Structure and self-association of the Rous sarcoma virus capsid protein. Structure Fold Des. 2000;8(6):617-28. PubMed PMID: 10873863.

33. Mortuza GB, Haire LF, Stevens A, Smerdon SJ, Stoye JP, Taylor IA. High-resolution structure of a retroviral capsid hexameric amino-terminal domain. Nature. 2004;431(7007):481-5. Epub 2004/09/24. doi: 10.1038/nature02915. PubMed PMID: 15386017.

34. Mortuza GB, Dodding MP, Goldstone DC, Haire LF, Stoye JP, Taylor IA. Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. Journal of molecular biology. 2008;376(5):1493-508. Epub 2008/01/29. doi: 10.1016/j.jmb.2007.12.043. PubMed PMID: 18222469.

35. Cornilescu CC, Bouamr F, Yao X, Carter C, Tjandra N. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. J Mol Biol. 2001;306(4):783-97. PubMed PMID: 11243788.

36. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, et al. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell. 1996;87(7):1285-94. PubMed PMID: 8980234.

37. Jin Z, Jin L, Peterson DL, Lawson CL. Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. J Mol Biol. 1999;286(1):83-93. PubMed PMID: 9931251.

38. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, et al. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science. 1997;278(5339):849-53. PubMed PMID: 9346481.

39. Khorasanizadeh S, Campos-Olivas R, Summers MF. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. J Mol Biol. 1999;291(2):491-505. PubMed PMID: 10438634.

40. Campos-Olivas R, Newman JL, Summers MF. Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. J Mol Biol. 2000;296(2):633-49. PubMed PMID: 10669613.

41. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI. Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science. 1996;273(5272):231-5. PubMed PMID: 8662505.

42. Goldstone DC, Flower TG, Ball NJ, Sanz-Ramos M, Yap MW, Ogrodowicz RW, et al. A unique spumavirus Gag N-terminal domain with functional properties of orthoretroviral matrix and capsid. PLoS pathogens. 2013;9(5):e1003376. Epub 2013/05/16. doi: 10.1371/journal.ppat.1003376. PubMed PMID: 23675305; PubMed Central PMCID: PMC3649970.

43. Cartellieri M, Herchenroder O, Rudolph W, Heinkelein M, Lindemann D, Zentgraf H, et al. N-terminal Gag domain required for foamy virus particle assembly and export. J Virol. 2005;79(19):12464-76. Epub 2005/09/15. doi: 79/19/12464 [pii]

10.1128/JVI.79.19.12464-12476.2005. PubMed PMID: 16160174; PubMed Central PMCID: PMC1211529.

44. Wilk T, Geiselhart V, Frech M, Fuller SD, Flugel RM, Lochelt M. Specific interaction of a novel foamy virus Env leader protein with the N-terminal Gag domain. J Virol. 2001;75(17):7995-8007. PubMed PMID: 11483744.

45. Mannigel I, Stange A, Zentgraf H, Lindemann D. Correct capsid assembly mediated by a conserved YXXLGL motif in prototype foamy virus Gag is essential for infectivity and reverse transcription of the viral genome. J Virol. 2007;81(7):3317-26. Epub 2007/01/19. doi: JVI.01866-06 [pii]

10.1128/JVI.01866-06. PubMed PMID: 17229703; PubMed Central PMCID: PMC1866044.

46. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci. 1995;20(11):478-80.

47. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2256-68. PubMed PMID: 15572779.

48. Byeon IJ, Meng X, Jung J, Zhao G, Yang R, Ahn J, et al. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. Cell. 2009;139(4):780-90. Epub 2009/11/17. doi: S0092-8674(09)01298-7 [pii]

10.1016/j.cell.2009.10.010. PubMed PMID: 19914170; PubMed Central PMCID: PMC2782912.

49. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013;497(7451):643-6. Epub 2013/05/31. doi: 10.1038/nature12162. PubMed PMID: 23719463; PubMed Central PMCID: PMC3729984.

50. Taylor WR. Decoy models for protein structure comparison score normalisation. J Mol Biol. 2006;357(2):676-99. PubMed PMID: 16457842.

51. Taylor WR. Protein structure comparison using iterated double dynamic programming. Protein Sci. 1999;8(3):654-65. PubMed PMID: 10091668.

52. Ohkura S, Goldstone DC, Yap MW, Holden-Dye K, Taylor IA, Stoye JP. Novel escape mutants suggest an extensive TRIM5alpha binding site spanning the entire outer surface of the murine leukemia virus capsid protein. PLoS pathogens. 2011;7(3):e1002011. Epub 2011/04/13. doi: 10.1371/journal.ppat.1002011. PubMed PMID: 21483490; PubMed Central PMCID: PMC3068999.

53. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, et al. X-ray structures of the hexameric building block of the HIV capsid. Cell. 2009;137(7):1282-92. Epub 2009/06/16. doi: S0092-8674(09)00580-7 [pii]

10.1016/j.cell.2009.04.063. PubMed PMID: 19523676; PubMed Central PMCID: PMC2840706.

54. Schur FK, Hagen WJ, Rumlova M, Ruml T, Muller B, Krausslich HG, et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. Nature. 2015;517(7535):505-8. Epub 2014/11/05. doi: 10.1038/nature13838. PubMed PMID: 25363765.

55. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature. 2011;472(7343):361-5. Epub 2011/04/23. doi: 10.1038/nature09976. PubMed PMID: 21512573; PubMed Central PMCID: PMC3081621.

56. Ohkura S, Stoye JP. A comparison of murine leukemia viruses that escape from human and rhesus macaque TRIM5alphas. Journal of virology. 2013;87(11):6455-68. Epub 2013/03/29. doi: 10.1128/JVI.03425-12. PubMed PMID: 23536686; PubMed Central PMCID: PMC3648092.

57. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson. 1998;131(2):373-8. PubMed PMID: 9571116.

58. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6(3):277-93. PubMed PMID: 8520220.

59. Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. Journal of biomolecular NMR. 1995;6(1):1-10. Epub 1995/07/01. doi: 10.1007/BF00417486. PubMed PMID: 22911575.

60. Linge JP, Habeck M, Rieping W, Nilges M. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics. 2003;19(2):315-6. Epub 2003/01/23. PubMed PMID: 12538267.

61. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 1999;13(3):289-302. PubMed PMID: 10212987.

62. Dosset P, Hus JC, Marion D, Blackledge M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. Journal of biomolecular NMR. 2001;20(3):223-31. Epub 2001/08/25. PubMed PMID: 11519746.

63. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007;66(4):778-95. Epub 2006/12/23. doi: 10.1002/prot.21165. PubMed PMID: 17186527.

64. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996;8(4):477-86. PubMed PMID: 9008363.

65. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science: The Royal Society of Chemistry, Cambridge United Kingdom; 1992. p. 90-125.

66. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J. 2002;82(2):1096-111. PubMed PMID: 11806949.

67. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000;78(3):1606-19. PubMed PMID: 10692345.

68. Brown PH, Schuck P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophys J. 2006;90(12):4651-61. PubMed PMID: 16565040.

69. Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, et al. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem. 2004;326(2):234-56. PubMed PMID: 15003564.

70. Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem. 2003;320(1):104-24. PubMed PMID: 12895474.

71. Müllers E, Uhlig T, Stirnnagel K, Fiebig U, Zentgraf H, Lindemann D. Novel functions of prototype foamy virus Gag glycine- arginine-rich boxes in reverse transcription and particle morphogenesis. J Virol. 2011;85(4):1452-63. Epub 2010/11/26. doi: JVI.01731-10 [pii]

10.1128/JVI.01731-10. PubMed PMID: 21106749; PubMed Central PMCID: PMC3028916.

72. Ho YP, Schnabel V, Swiersy A, Stirnnagel K, Lindemann D. A small-molecule-controlled system for efficient pseudotyping of prototype foamy virus vectors. Molecular therapy : the journal of the American Society of Gene Therapy. 2012;20(6):1167-76. Epub 2012/04/05. doi: 10.1038/mt.2012.61. PubMed PMID: 22472951; PubMed Central PMCID: PMC3369295.

73. Stange A, Luftenegger D, Reh J, Weissenhorn W, Lindemann D. Subviral particle release determinants of prototype foamy virus. J Virol. 2008;82(20):9858-69. Epub 2008/08/08. doi: JVI.00949-08 [pii]

10.1128/JVI.00949-08. PubMed PMID: 18684814; PubMed Central PMCID: PMC2566296.

74. Lindemann D, Pietschmann T, Picard-Maureau M, Berg A, Heinkelein M, Thurow J, et al. A particle-associated glycoprotein signal peptide essential for virus maturation and infectivity. J Virol. 2001;75(13):5762-71. Epub 2001/06/08. doi: 10.1128/JVI.75.13.5762-5771.2001. PubMed PMID: 11390578; PubMed Central PMCID: PMC114292.

75. Stange A, Mannigel I, Peters K, Heinkelein M, Stanke N, Cartellieri M, et al. Characterization of prototype foamy virus gag late assembly domain motifs and their role in particle egress and infectivity. J Virol. 2005;79(9):5466-76. Epub 2005/04/14. doi: 79/9/5466 [pii]

10.1128/JVI.79.9.5466-5476.2005. PubMed PMID: 15827161; PubMed Central PMCID: PMC1082757.

**Figure legends**

**Fig. 1. NMR structure of the central domain of PFV Gag.** (**A**) Sequence alignment of foamy virus Gag-central domains from mammals, old and new world monkeys (SFV). Mammalian FVs are abbreviated as follows: BFV, Bovine; EFV, Equine; FFV, Feline. Monkey species are abbreviated as follows: mac, Macaque; agm, African green monkey; spm, Spider monkey; sqm, Squirrel monkey; mar, marmoset. Numbering corresponds to the PFV sequence. (**B**) Cartoon representation of PFV-Gag(300-477) backbone is shown in cyan. The secondary structure elements are numbered sequentially from the amino-terminus and the N- and C-termini are indicated. (**C**) The interface between PFV-Gag NtDCEN and CtDCEN domains.

**Fig. 2. Alignment of PFV-Gag central domain with orthoretroviral CA domains.** Panels **A-D**, are the best-fit 3D structural superimpositions of PFV-Gag(300-383) NtDCEN with the CA-CtD of (**A**) RSV, (**B**) HIV-1, (**C**) BLV, (**D**) HTLV. Panels **E-I**, are the best-fit 3D structural superimpositions of PFV-Gag(385-477) CtDCEN with the CA-NtD of (**E**) RSV, (**F**) HIV-1, (**G**) JSRV, (**H**) HTLV and (**I**) MLV. In all panels, molecules are shown in cartoon representation with helices displayed as cylinders, PFV-Gag is coloured cyan.

**Fig. 3. Evolutionary relatedness of PFV-Gag NtDCEN and CtDCEN domains to orthoretroviral CA domains**. (**A**) Structural similarity based on the inverted SAP score...... WT to carry on

**Fig. 4.** C**onformation and solution oligomeric state of FV-Gag central domains.** (**A**) C(S) distributions derived from sedimentation velocity data recorded from PFV-Gag(300-477) at 1 mg/mL (left panel); PFV-Gag NtDCEN at 2 mg/mL(middle panel) and PFV-CtDCEN 2 mg/mL (right panel). (**B**) Multi-speed sedimentation equilibrium profiles determined from interference data collected on PFV-Gag(300-477) at 49 µM (left panel) and PFV-Gag CtDCEN at 42 µM (right panel). Data was recorded at the speeds indicated. The solid lines represent the global best fit to the data using either a single species or monomer-dimer equilibrium model. The lower panels show the residuals to the fit.

**Fig. 5. NMR structure of PFV-Gag CtDCEN homodimer.** (**A**) Cartoon representation of the structure of the PFV Gag 381-477 dimer. Monomer-A is shown in dark blue and Monomer-B in cyan. The α-helices are labelled as for PFV-Gag(300-477). (**B**) Details of the homodimer interface. Residues that contribute to the interface are shown as sticks. (**C**) C(S) distribution derived from sedimentation velocity data recorded from PFV-Gag CtDCEN L410E/M413E mutant (300-477) at 1 mg/mL (dashed line). The C(S) distribution derived from sedimentation velocity data recorded from wt PFV-Gag CtDCEN at x mg /mL is shown also for comparison (solid line).

**Fig.6. PFV-Gag central domain interface mutants.** (**A**)cEM analysis of PFV-Gag central domain interface mutants. (B) Infectivity and particle budding of central domain mutants.

**Table 1. NMR and refinement statistics for PFV-Gag Central domains**

|  |  |  |  |
| --- | --- | --- | --- |
|  | PFV(381-477) | PFV(340-482) | PFV(300-477) |
| **NMR distance and dihedral constraints** |  |  |  |
| NOE Distance constraints |  |  |  |
| *Total NOE* | 3061 | 3093 | 4140 |
| *Unambiguous* | 2616 | 2823 | 3637 |
| *Intermolecular* | 20x2 |  |  |
| Hydrogen bonds | 31x2 | 41 | 48 |
| Total dihedral angle restraints |  |  |  |
| *φ* | 74x2 | 109 | 113 |
| *Ψ* | 74x2 | 109 | 114 |
| Total RDCs | 41x2 |  | 46 |
|  |  |  |  |
| **Structure statistics** |  |  |  |
| Violations (mean and s.d.) |  |  |  |
| *Distance constraints (>0.5Å)* | 0 | 0 | 0 |
| Deviations from idealised geometry |  |  |  |
| *Bond lengths (Å)* | 0.008 | 0.009 | 0.015 |
| Average pairwise r.m.s. deviation (Å) |  |  |  |
| *Heavy* | 0.6 | 0.7 | 0.6 |

**Table 2. Hydrodynamic parameters of PFV-Gag Central domains**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **PFV-Gag**  **(300-477)** | **PFV-Gag**  **NtDCEN** | **PFV-Gag**  **CTDCENmon** | **PFV-Gag**  **CTDCENdim** |
| ***Parameter*** | | | | |
| ͞v (ml.g-1) | 0.7416 | 0.7415 | 0.7312 | |
| ρ(g.ml-1) | 1.005 | 1.005 | 1.005 | |
| a Mr (Da) | 20,543 | 10,660 | 11,894 | |
| ε280 (M-1.cm-1) | 16,960 | 11,000 | 5,960 | |
| ***Sedimentation velocity*** | | | | |
| Crange (µM) | 24 - 97 | 188 | 42 - 168 | |
| b S20,w (x1013) sec | 1.87 | 1.25 | 1.65 | 2.07 |
| cMw C(S) (kDa) | 20.6 | 10.3 | 14.7 | 20.7 |
| d rmsd C(S) (x10-3) | 3.0 - 5.1 | 5.7 | 5.8 - 7.6 | |
| ***Sedimentation equilibrium*** | | | | |
| Crange (µM) | 24 - 97 |  | 17 - 84 | |
| eKD (µM) | - |  | 0.9 | |
| f Mw (kDa) | 20.3 |  | - | |
| g rmsd (x10-3) | 6.9 - 7.1 |  | 4.8 - 6.0 | |
| h χ2 | 1.98 |  | 1.33 | |

aMolar mass calculated from the protein sequence

bThe S20,w value remained constant across the concentration range tested.

cThe weight averaged molecular weight derived from the best fit C(S) function.

dThe range of the rms deviations observed when data were fitted using a continuous sedimentation coefficient distribution model.

eThe equilibrium dissociation constant calculated from a monomer-dimer self-association model.

fThe weight averaged molecular weight from Global SE analysis using a species analysis model.

gThe range of the rms deviations observed for each multi-speed sample when fitted individually to the appropriate model.

hThe global reduced chi-squared for the global fit to the appropriate model.

**Supporting Information**

**S1 Fig. NMR family of structures PFV Gag(300-477)**

**S2 Fig. NMR relaxation data**

(**A**) xxx (**B**) zzz

**S2 Fig. Infectivity assays**

(**A**) xxx (**B**) yyy

**S3 Fig. 340-477 structure**

**S4 Fig. NMR family of structures PFV-Gag CtDCEN dimer**

**S4 Fig. YxxLGL motif**

Overview with surface patch (inset) conserved aromatic (Y464) in pocket

**S5 Fig. The restriction determinants for MLV CA map to regions absent in the PFV Gag central domain**

Superposition of MLV CA (orange) and α3 – α7 of PFV-Gag(340-482) (cyan). MLV CA is also represented as a molecular surface, with red mapping the residues involved in restriction determination. The helices of the PFV Gag central domain are labelled.

**S1 Table. SSM superpose scores for structural alignments**

**S2 Table. Dirk Quantitative-PCR primers**

**S3 Table. Hydrodynamic analysis of FFV-Gag(1-348)**

**S1 Table Structural alignment scores**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Query | Target | Q-Score | N-align | RMSD (Å) |
| PFV-Gag NtDCEN | RSV | 0.304 | 61 | 2.83 |
| HIV-1 | 0.252 | 58 | 3.18 |
| BLV | 0.245 | 57 | 3.24 |
| HTLV | 0.100 | 60 | 3.05 |
| PFV-GagCtDCEN | MLV | 0.181 | 68 | 2.81 |
| HTLV | 0.176 | 71 | 3.11 |
| HIV-1 | 0.173 | 67 | 2.57 |
| JSRV | 0.149 | 63 | 2.96 |
| RSV | 0.104 | 59 | 3.30 |

**S3 Table Hydrodynamic parameters FFV(1-XXX)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Parameter*** | | | | |
| ͞v (ml.g-1) | 0.7373 | | | |
| ρ(g.ml-1) | 1.004 | | | |
| a Mr (Da) | 39,988 | | | |
| ε280 (M-1.cm-1) | 44,460 | | | |
| ***Sedimentation velocity*** | | | | |
| Crange (µM) | 50 | | | |
| b S20,w (x1013) sec | 3.95 | 4.98 | 7.36 | 10.32 |
| cMw C(S) (kDa) | 52.1 | 73.8 | 132.5 | 220.0 |
| d rmsd C(S) (x10-3) | 6.2 | | | |

1. Molar mass calculated from the protein sequence
2. The S20,w value remained constant across the concentration range tested.
3. The weight averaged molecular weight derived from the best fit C(S) function.
4. The range of the rms deviations observed when data were fitted using a continuous sedimentation coefficient distribution model.