Structure and self-association of the Rous sarcoma virus capsid protein

Richard L Kingston[1†], Tanja Fitzon-Ostendorp[1], Elan Zohar Eisenmesser[1‡], Gisela W Schatz[2], Volker M Vogt[2], Carol Beth Post[1,3] and Michael G Rossmann[1*]

Background: The capsid protein (CA) of retroviruses, such as Rous sarcoma virus (RSV), consists of two independently folded domains. CA functions as part of a polyprotein during particle assembly and budding and, in addition, forms a shell encapsidating the genomic RNA in the mature, infectious virus.

Results: The structures of the N- and C-terminal domains of RSV CA have been determined by X-ray crystallography and solution nuclear magnetic resonance (NMR) spectroscopy, respectively. The N-terminal domain comprises seven α helices and a short β hairpin at the N terminus. The N-terminal domain associates through a small, tightly packed, twofold symmetric interface within the crystal, different from those previously described for other retroviral CAs. The C-terminal domain is a compact bundle of four α helices, although the last few residues are disordered. In dilute solution, RSV CA is predominantly monomeric. We show, however, using electron microscopy, that intact RSV CA can assemble *in vitro* to form both tubular structures constructed from toroidal oligomers and planar monolayers. Both modes of assembly occur under similar solution conditions, and both sheets and tubes exhibit long-range order.

Conclusions: The tertiary structure of CA is conserved across the major retroviral genera, yet sequence variations are sufficient to cause change in associative behavior. CA forms the exterior shell of the viral core in all mature retroviruses. However, the core morphology differs between viruses. Consistent with this observation, we find that the capsid proteins of RSV and human immunodeficiency virus type 1 exhibit different associative behavior in dilute solution and assemble *in vitro* into different structures.

Addresses: [1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA, [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA and [3]Department of Medicinal Chemistry, Purdue University, West Lafayette, IN 47907, USA.

Present addresses: [†]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA and [‡]National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA.

*Corresponding author.
E-mail: mgr@indiana.bio.purdue.edu

Key words: electron microscopy, NMR spectroscopy, retrovirus, self-assembly, X-ray crystallography

Received: **10 February 2000**
Revisions requested: **16 March 2000**
Revisions received: **10 April 2000**
Accepted: **11 April 2000**

Published: **24 May 2000**

Structure 2000, **8**:617–628

0969-2126/00/$ – see front matter

Introduction

Rous sarcoma virus (RSV) is a tumor-generating avian retrovirus. Following its isolation in 1911, it has become one of the most thoroughly studied retroviruses. These viruses are dependent on RNA-directed DNA synthesis for the establishment of infection, via the process of reverse transcription. They are the causative agents of cancer and immune disorders in many species.

All retroviruses possess a gene (*gag*), which encodes the major structural proteins of the virion. The *gag* gene is translated as a polyprotein, which directs the formation and release of spherical viral particles. During or after budding of the virus, the Gag polyprotein is cleaved by the viral protease, generating the internal structural proteins found in the infectious retrovirus. These include the proteins MA (matrix), CA (capsid), and NC (nucleocapsid). Proteolytic cleavage is associated with a structural rearrangement of the virus, termed maturation. The most dramatic structural change seen in the maturing retrovirus is the formation of the viral core, a large assemblage in which a shell of CA encapsidates the genomic RNA and the replicative enzymes of the virus. Within the capsid, NC associates tightly with the RNA, forming a ribonucleoprotein (RNP) complex, which is the template for reverse transcription. The retroviral core is delivered into the cytoplasm of the host cell during the initial stages of infection.

There has been rapid progress in structural studies of the individual components of retroviruses (for reviews focused on human immunodeficiency virus (HIV-1) see [1,2]). Much less is known about the interaction of these components within the virion and the regulation of retroviral assembly and maturation. This paper focuses on the retroviral capsid protein, a molecule comprising two loosely associated, predominantly α helical domains. Structures of the nearly full-length capsid protein, or of the individual domains, have been reported for HIV-1 [3–8], equine infectious anemia virus (EIAV) [9], and human T-cell leukemia virus type 1 (HTLV-1) [10]. These results,

obtained using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, show that there can be large variations in the relative positions of the two domains, with few or no interactions between them [5,7,9,10]. The C-terminal domain contains the most highly conserved amino acid sequence within the retroviral Gag protein, a stretch of 20 residues termed the major homology region (MHR). Although the role of this region is unknown, an essential function in viral replication is implied by its evolutionary conservation in otherwise widely diverged sequences. This supposition has been confirmed by site-directed mutagenesis [11,12].

In the immature virion, CA is embedded within the Gag polyprotein. All immature retroviruses have a similar supramolecular organization. The Gag proteins are arranged radially, with the N termini facing the viral membrane and the C termini in the interior of the particle. Retroviruses are heterogeneous in size and lack any overall symmetry [2]. Although CA forms the exterior shell of the viral core in all mature retroviruses, the core morphology differs among these viruses. The core in HIV-1 and other lentiviruses (e.g. EIAV) has a predominantly conical appearance. The cores of other retroviruses, however, such as RSV (cryo-EM unpublished observations) and murine leukemia virus (MLV), resemble irregular polyhedra.

The sequence immediately downstream of CA is important in directing immature particle formation in RSV and also during maturation of the virus [13]. The proteolytic processing that occurs at the C terminus of CA has been studied in both RSV [14,15] and HIV-1 [16,17]. In both of these retroviruses, CA is separated from the downstream NC by a short amino acid sequence, termed the spacer peptide (SP in RSV, and SP1 or p2 in HIV-1). The spacer is 12 amino acids long in RSV and 14 amino acids long in HIV-1 (with some variation among different isolates). In RSV, deletions of all, or part, of SP do not block particle assembly or release. All such mutants are non-infectious, however, with the resulting particles exhibiting increased size heterogeneity and a reduced stability of the viral core [13,14]. In HIV-1, most studies have shown that deletions or point mutations in this region also result in morphologically aberrant particles of heterogeneous size [16]. Cleavage at the N terminus of the spacer peptide is much slower than cleavage at its C terminus, and an intermediate species CA SP (CA SP1 in HIV-1) transiently accumulates and disappears during the maturation process [15].

We initiated structural studies on RSV CA to help understand the role of CA and the spacer peptide in viral assembly. Crystals of this protein have been described previously, in which the molecule is arranged in pseudo-helical fashion [18]. Although the interactions of CA within these crystals might mimic those found *in vivo*, the crystallography is not straightforward, partly because of crystal disorder. In this paper, we report independent structure determinations of the two domains of RSV CA; the N-terminal domain by X-ray crystallography and the C-terminal domain and spacer peptide by solution NMR spectroscopy. Results of *in vitro* assembly studies on intact CA are also presented.

Results and discussion

The N-terminal domain of RSV CA

Domain structure, the role of the β hairpin, and interactions within the crystal

X-ray diffraction data were collected to 2.05 Å resolution from crystals of the N-terminal domain of CA (residues 1–154) (Table 1). Phases to 2.7 Å resolution were experimentally determined by multiwavelength anomalous diffraction (MAD) methods. The crystal space group is $P2_12_12_1$, with two copies of the molecule in the asymmetric unit. One of the molecules undergoes relatively large displacements from its mean position within the crystal lattice, and the electron density for this copy is less resolved. The crystallographic R factors for the final refined model were 24.7% (R_{work}) and 27.1% (R_{free}). The last eight residues of both $CA_{1–154}$ molecules are disordered and have not been modeled. Comparison with structures of HIV-1 and EIAV CA [7,9] shows that these residues form the flexible linker between the two domains of the intact protein.

Table 1

X-ray diffraction data and model refinement statistics: N-terminal domain of RSV CA.

Data set	Native	MAD
X-ray diffraction data		
X-ray source	APS 14-BM-c	APS 14-BM-d
Outer resolution limit (Å)	2.05	2.70
Redundancy	5.0 (4.7)*	4.0
Completeness (%)	92.5 (89.8)*	98
$R_{measure}$	0.049 (0.335)*	
peak 1		0.037/0.056†
peak 2		0.034/0.064†
inflexion 1		0.036/0.038†
inflexion 2		0.035/0.046†
remote		0.034/0.048†
Model refinement		
R_{work}/R_{free}	24.7/27.1	
Missing regions		
molecule 1/molecule 2	None/6–8, 96–98	
Mean B values		
molecule 1/molecule 2 (Å^2)	36/49	
Number of water molecules	101	
Rmsd from ideal		
bond lengths (Å)	0.005	
bond angles (°)	1.1	
Ramachandran plot		
residues in most favored regions (%)	94	

*Statistics for highest resolution bin (2.11–2.05 Å) in parentheses.
†Centric/acentric observations.

The N-terminal domain comprises seven α helices, together with a short β hairpin at the N terminus, and is similar to that found in other retroviral CAs (Figure 1). The least conserved region within CA is found between the fourth and last helices of the domain (the top left of the molecule in Figure 1), where corresponding helices vary significantly in both length and orientation in the four known structures (RSV, HIV-1, HTLV-1, EIAV). This region encompasses the flexible, proline-rich, cyclophilin-A binding loop found in HIV-1 CA.

The β hairpin in RSV CA (amino acids 1–12) projects away from the main body of the domain and is anchored in place at the N terminus by a salt bridge between the charged N-terminal proline residue and a buried aspartic acid (Figure 1). In this, it closely resembles HIV-1 CA, in which the hairpin (amino acids 1–13) is held in place by equivalent interactions [3–5]. In the better defined copy of RSV $CA_{1–154}$, the electron density for the β hairpin is readily interpretable. In the second copy of this molecule, there is no interpretable electron density for amino acids 6–8, the most distal part of the hairpin. This structural element seems to be quite mobile in both RSV and HIV-1 [3,4,19]. The conservation of the β hairpin among retroviral CAs has been widely anticipated; however, RSV is only the second retrovirus in which this has been experimentally demonstrated. In structure determinations of EIAV and HTLV-1 CA [9,10], the molecules studied did not have wild-type sequences at the N terminus, and the structure of this region was not resolved.

The N-terminal β hairpin, as seen in RSV and HIV-1 CA, can only completely form after proteolysis of the Gag precursor. In HIV-1, it was proposed, largely on the basis of crystallographic results [3], that structural rearrangement at the N terminus following proteolysis allows formation of a new CA–CA interface, facilitating assembly of the capsid in the maturing virus [3,4]. The N-terminal proline, and its bonding partner (aspartic acid or glutamic acid), are conserved among highly diverged retroviral sequences, suggesting that similar events occur in almost all retroviruses. Hence, it was hypothesized that the N terminus of CA functioned as a 'conformational switch' responsible for redirecting assembly of CA in the mature virion. This concept is supported by the *in vitro* assembly studies on purified viral components, although results for HIV-1 [20] and RSV [21] were not entirely consistent. The recent observation that truncated HIV-1 Gag proteins with an 'immature' CA N terminus can be assembled into either tubes (resembling the mature core) or spheres (resembling the immature virion), depending upon solution pH [22], indicates that the switching hypothesis is probably oversimplified. It is not clear to what extent the structure at the N terminus of mature CA is formed within the Gag polyprotein. The β hairpin must be at least transiently disrupted to allow access to the viral protease and, hence, must fold in a semi-autonomous fashion. The N-terminal proline in RSV is followed in the sequence by three β-branched amino acids (Val–Val–Ile), which have a high intrinsic propensity for β sheet formation [23]. This might help direct the formation and stabilization of the hairpin. More generally, there is a strong preference for β-branched amino acids at positions 2 and 3 in the sequences of all retroviral CAs.

The two noncrystallographically related copies of the protein interact through a tightly packed interface, arising from the parallel association of the fourth helix of the domain (Figure 2). The interface exhibits almost perfect twofold rotational symmetry (the two molecules involved are related by a rotation of 178°, with a translation of 0.2 Å along the symmetry axis). Residues at the core of the interface are hydrophobic, but the area of interaction is small, with the molecular surface area of each of the interacting domains reduced by only 1.9% (250 $Å^2$). The domain interaction is mediated by an octahedrally coordinated metal ion (probably Mg^{2+}, which was present at high concentration in the crystallization medium). The metal ion is bound to Asp71 in each molecule and to four coplanar water molecules. This interface is different from the twofold symmetric interfaces previously reported for the N-terminal domain of HIV-1 CA in complex with a monoclonal antibody fragment [5] or with cyclophilin A [3]. The N-terminal β hairpin is found on the exterior of the complex and is not involved in interface formation (Figure 2). This interface could be one of several interactions involved in assembly of the immature or mature particle.

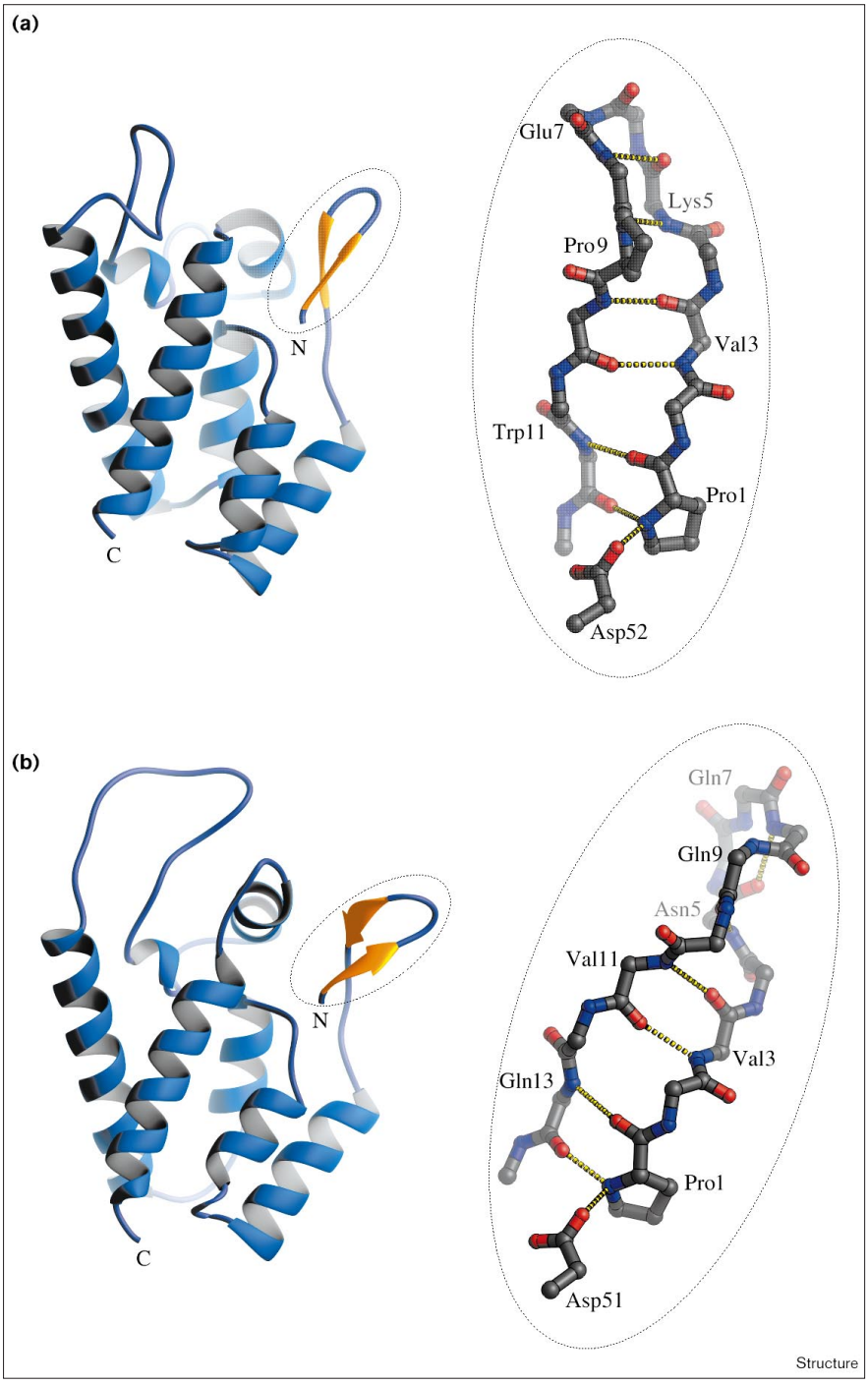
The C-terminal domain of RSV CA

Structure determination by NMR spectroscopy

We overexpressed two constructs in *Escherichia coli*, encompassing the C-terminal domain of CA, and studied them by solution NMR spectroscopy. Both start at residue 155 of CA and contain, in addition to the wild-type sequence, an initiating methionine. Evaluation of the probable domain boundary was made on the basis of proteolysis experiments on HIV-1 CA [4], coupled with alignment of the HIV-1 and RSV CA sequences. Recent structural studies suggest that the flexible polypeptide linking the N- and C-terminal domains begins several residues upstream [7,9,10]. The first construct ($CA_{155–237}$) ends at Ala237, which is the C terminus of the predominant variant of mature RSV CA [15]. The second construct ($CA_{155–249}$) ends at Met249, hence including both the C-terminal domain of CA and the spacer peptide. The constructs show no apparent self-association in dilute solution, as monitored by size-exclusion chromatography (data not shown). This is consistent with results for the intact protein, which behaves in a similar fashion (see below).

The four α helices of the C-terminal domain were readily identified by analysis of their chemical shifts [24] and by

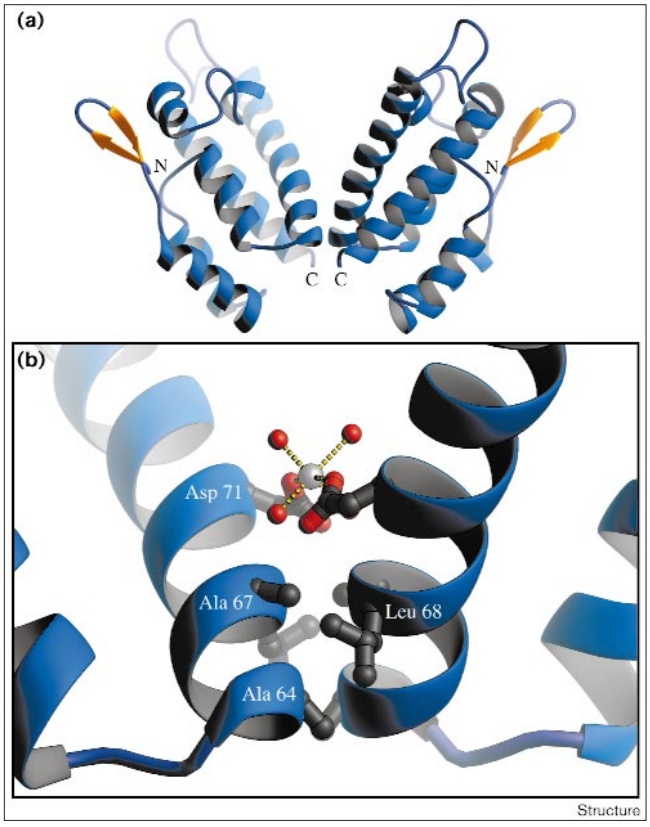
Figure 1



The N-terminal domain of the retroviral capsid protein. **(a)** RSV, **(b)** HIV-1. The β hairpin is shown in detail, in an orientation that is slightly different from that of the whole domain. The boundaries of the helices within the N-terminal domain of RSV CA are: helix **1**, 15–30; **2**, 33–44; **3**, 49–61; **4**, 62–87; **5**, 102–107; **6**, 116–124; **7**, 125–147. Helix definitions incorporate the N- and C-terminal capping residues. Coordinates for the N-terminal domain of HIV-1 CA were taken from the X-ray structure of the molecule in complex with cyclophilin A [3]. The program Ribbons 2.0 [50] was used in the preparation of the figure.

characteristic patterns of short and medium range nuclear Overhauser enhancements (NOEs). The final helix terminates at Arg225. In both the shorter and longer constructs, there are no significant chemical shift deviations from random coil values beyond Pro230. Addition of the spacer peptide to the C terminus of CA slightly perturbs the NMR chemical shifts of the last seven residues (amino acids 230–237) and results in chemical shift heterogeneity, beginning at Pro230. Instead of a single resonance associated with each nucleus, more than one was often observed, although they all clustered around random coil values. The chemical shift heterogeneity is most probably because of *cis–trans* isomerization of Pro230 and Pro246. There is an almost complete absence of inter-residual

Figure 2



The twofold symmetric interaction between CA_{1-154} molecules seen within the crystal. **(a)** The interaction involves the parallel packing of helix 4 from adjacent molecules. Although the distal part of the β hairpin and one of the surface loops are not completely ordered in one copy of the molecule, this is not shown. **(b)** Details of the interface. The sidechain atoms Ala64, Ala67, Leu68 and Asp71 are shown in ball-and-stick representation. Asp71 in both molecules is coordinated to a bridging metal ion, most likely Mg^{2+}. The program Ribbons 2.0 [50] was used in the preparation of the figure.

NOEs in the region C-terminal to Pro230, whether or not the spacer peptide is present.

The magnitudes of the steady-state heteronuclear ^{1}H–^{15}N NOEs were measured to study the mainchain flexibility. The steady-state ^{1}H–^{15}N NOE is highly sensitive to reorientation of the ^{1}H–^{15}N bond vector, with negative, or small positive, NOE values being indicative of fast motion relative to the overall tumbling rate of the molecule. In the shorter construct, the heteronuclear NOEs for the backbone amide groups start to decrease sharply at Arg225, and beyond residue 230 they become negative. The same pattern is exhibited in the longer construct with none of the residues in the spacer peptide exhibiting large positive heteronuclear NOEs. These results indicate that the C terminus of CA and the spacer peptide (residues 230–249) do not possess any persistent secondary or tertiary structure and are flexible in solution. Deletion of the spacer peptide results in no detectable change within the structured region of the C-terminal domain.

Table 2

Structure calculation statistics: C-terminal domain of RSV CA.

NOE-derived distance restraints*	
Ambiguous	460
Unambiguous	1323
Intraresidue ($\|i-j\|=0$)	647
Sequential ($\|i-j\|=1$)	333
Medium range ($1<\|i-j\|<5$)	239
Long range ($\|i-j\|\geq 5$)	104
J-coupling restraints	
$^{3}J(H^{N},H^{\alpha})$ (φ)	59/3†
$^{3}J(N,H^{\beta})$ (χ*)	16/21†
Precision of structural ensemble (19 structures)	
Rmsd from minimized average structure	
All atoms: residues 155–225 (Å)	1.53
Cα atoms: residues 155–225 (Å)	0.86
Agreement of structures with restraints‡	
Distance restraint violations	0
J-coupling restraints (Hz)	0.628§
Chemical shifts	
Cα (ppm)	1.114§
Cβ (ppm)	1.494§
Non-amide proton chemical shifts (ppm)	0.324§

*Restraint lists were not edited to remove structurally insignificant restraints. †Achiral groups/prochiral groups. ‡Mean value for ensemble. §Rmsd between experimental values and values calculated from the NMR structures.

Solution structure of the C-terminal domain

The three-dimensional structure of the C-terminal domain (residues 155–230) was determined from NOE-derived distance restraints, together with 3-bond J-coupling measurements and chemical shift data (see the Materials and methods section). The minimized average structure agrees well with the experimental restraints (Table 2) and has 88% of the amino acids in the most favored region of the Ramachandran plot.

The four α helices of the C-terminal domain form a compact bundle (Figure 3). The N-terminal residues (155–159) have an extended conformation and pack against the first helix of the domain. The MHR encompasses this N-terminal extended strand, the subsequent β turn and much of the first helix. Several of the invariantly conserved amino acids within the MHR are involved in formation of the β turn. Critical to the integrity of this region is a glutamine residue (Gln158 in RSV) that functions as a hydrogen-bond donor and acceptor, bridging the N terminus of helix 1 and the C terminus of helix 2. All tested mutations of this residue in RSV disrupt particle assembly and abolish viral replication [12] (T Cairns and R Craven, unpublished observations). A conservative substitution of the equivalent residue in HIV-1 has a similar effect [11]. Most other elements of the associated hydrogen bonding

network within HIV-1 [6] are present in RSV. It has not been possible to fully define the sidechain conformation of Arg170, which is completely conserved among CA sequences. Nonetheless, there are several indications that this residue adopts a fixed position within the structure (see Materials and methods). The hydrophobic residues within the MHR are found on the interior face of the first helix. Given the potential for compensatory mutations within the hydrophobic core of most proteins, the reason for their conservation remains uncertain.

The spacer peptide appears unstructured: implications for assembly and maturation

The lack of persistent structure at the C terminus of RSV CA and within the spacer peptide is consistent with studies on other retroviral CAs. The last 8–10 amino acids of the HIV-1 and EIAV CAs are disordered in X-ray crystal structures [6–9], and NMR spectroscopy shows that this region lacks structure in HTLV-1 CA [10]. For HIV-1, one of the proteins studied encompassed both the C-terminal domain and the spacer peptide SP1 [8]; however, SP1 was found to be disordered within the crystal. Although a helix overlapping the CA SP1 boundary in HIV-1 has been predicted [16], no direct experimental evidence exists in support of this hypothesis.

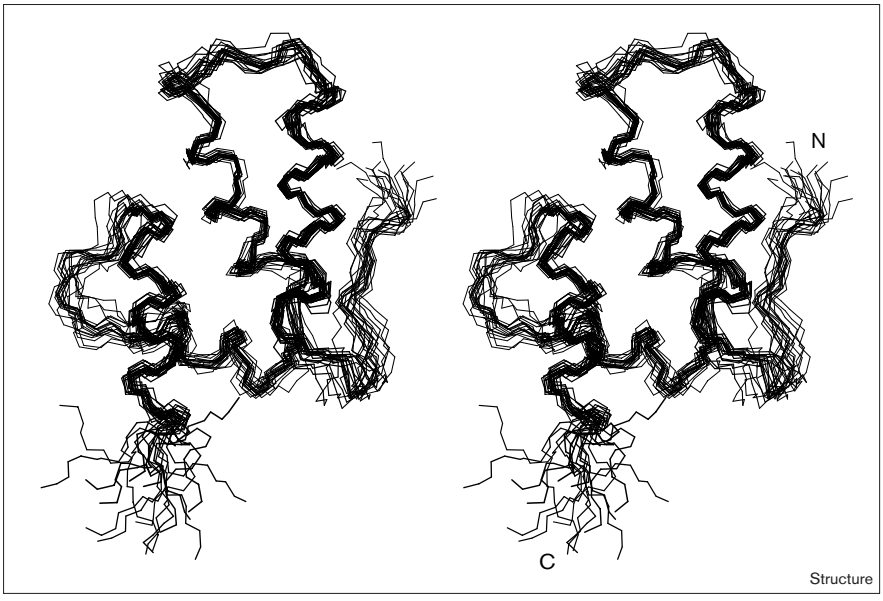
It is unclear if these observations can be related directly to the situation *in vivo*. Proteolytic cleavage might be associated with conformational change, thereby causing the structure of the mature proteins, or transiently observed intermediates, to differ from that embedded in the polyprotein precursor. Within the immature virus, lateral interactions between the Gag polyprotein, mediated by nucleic acid binding, will confine the polypeptide chain at the CA–NC boundary to a very restricted volume, potentially influencing structure in this region. Nevertheless, there is little evidence to suggest that the spacer peptide needs to be structured to fulfill its role in viral assembly and maturation. The assembly of immature particles is a robust process, and the Gag polyprotein in some retroviruses is known to include domains that are unstructured *in vitro* [1]. During maturation, the removal of the spacer peptide from the C terminus of CA appears essential for proper condensation of CA around the RNP complex in RSV and HIV-1 [14,16,17]. The presence of a long, flexible extension at the C terminus of CA, however, could block capsid assembly through a simple steric mechanism.

The majority of retroviruses possess a spacer peptide which separates mature CA and NC; however, it varies widely in length and composition, ranging from five amino acids in EIAV to 25 amino acids in bovine immunodeficiency virus (BIV). In several retroviruses (RSV, HIV-1, BIV), there are multiple cleavage sites within the spacer peptide. Some retroviruses (mouse mammary tumor virus (MMTV) and MLV) lack a spacer peptide altogether. It, therefore, seems unlikely that the spacer peptide has a completely defined and invariant structure in HIV-1 or in any other retrovirus.

Intermolecular association and the C-terminal domain

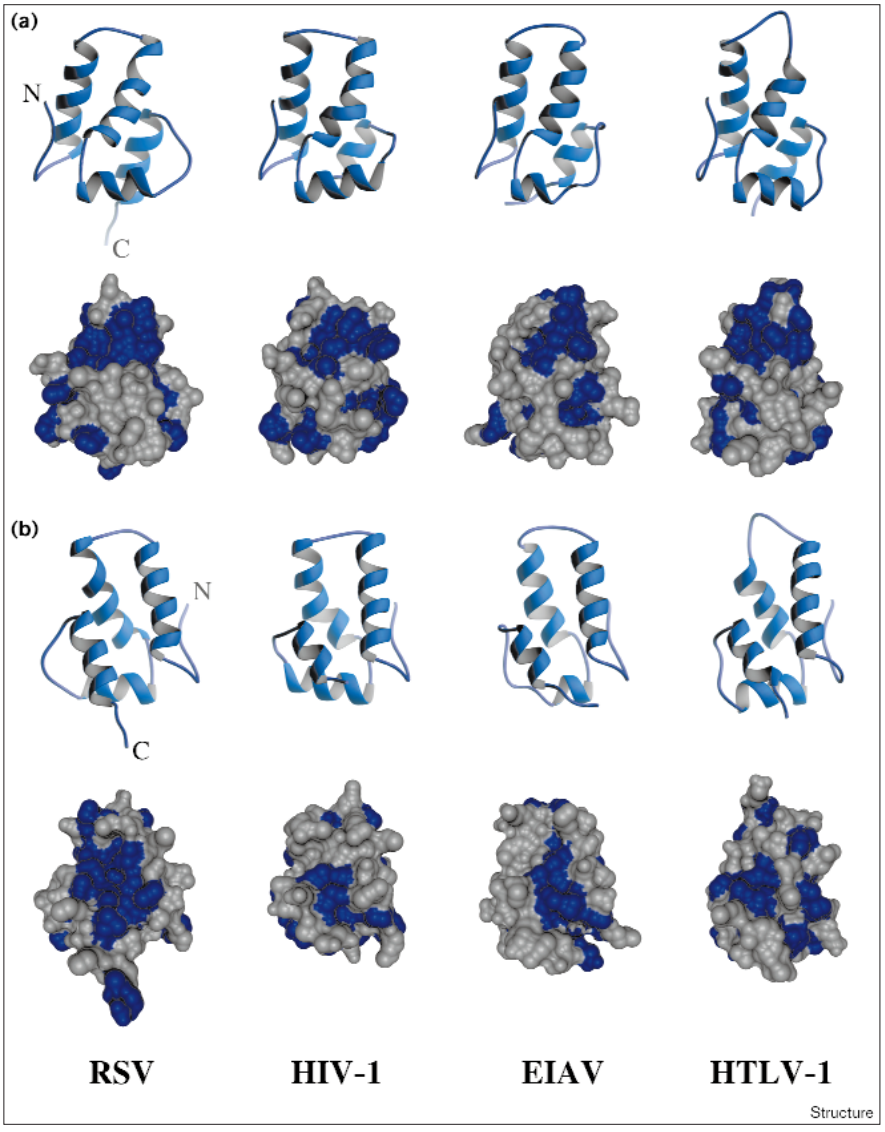
Structures of retroviral CAs have not fully clarified the pathway or outcome of viral assembly. The CA–CA interactions within the immature and mature virus might be quite different, as suggested, for example, by crosslinking studies on MLV [25]. The study of capsid

Figure 3



A stereo-diagram showing 19 conformers representing the solution structure of the C-terminal domain of RSV CA. Only the backbone atoms (N, C, Cα) are displayed. The boundaries of the helices within the C-terminal domain of RSV CA are: helix **1**, 163–177; **2**, 181–197; **3**, 198–208; **4**, 214–226. Helix definitions incorporate the N- and C-terminal capping residues. The program Molscript [51] was used in the preparation of the figure.

Figure 4



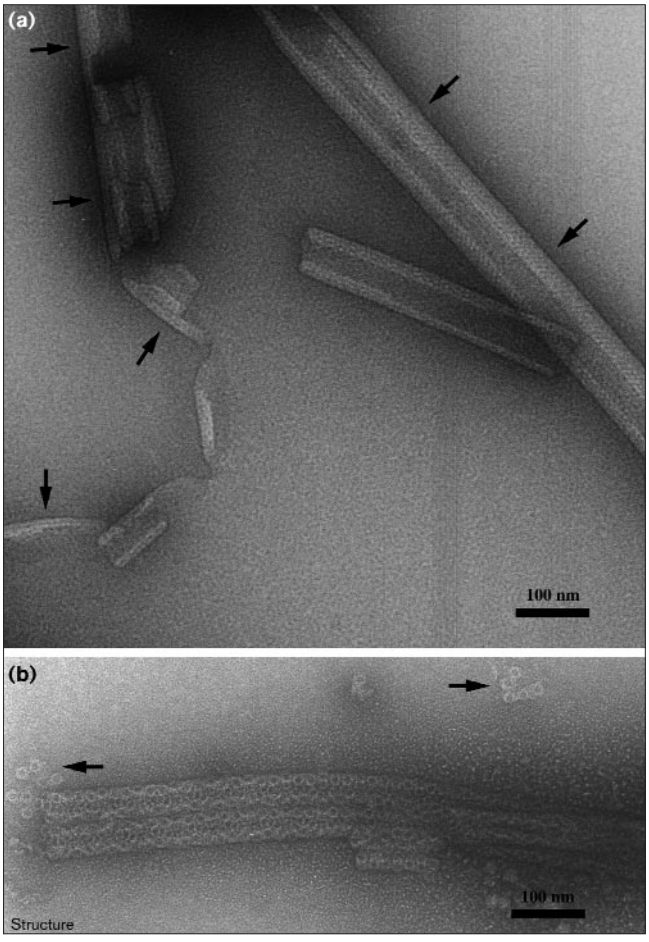
Surface properties of the C-terminal domain. **(a)** Schematic representation of the C-terminal domains of four retroviral CAs (top). Corresponding representations of the molecular surface (bottom). Regions of the molecular surface arising from non-polar amino acids (G,A,P,V,I,L,F,M,C,W [52]) are colored blue. **(b)** This figure displays the same information as in (a) but the domains have been rotated by 180° so that the reverse face of the molecule is visible. For clarity, the N terminus has been truncated at an equivalent point in each structure (Asp155 in RSV, Asp152 in HIV-1, Asn152 in EIAV, Ser135 in HTLV-1). Several residues N-terminal to this position should be considered part of the C-terminal domain and are potentially involved in domain association [6,8]. Ribbon diagrams were prepared using the program Ribbons 2.0 [50], and molecular surfaces were calculated and displayed using the programs MSMS and MSV [53]. Vertices of the triangulated surface were colored by the property of the closest atom. The Protein Data Bank accession codes of the structural models used to generate the figure are 1a43 (HIV-1 [8]), 1eia (EIAV [9]), 1qrj (HTLV-1 [10]).

condensation in the maturing virus is further complicated by the transitory nature of the interactions involved, the potential for cooperative association of the two flexibly linked domains of CA, and the irregularity of the authentic viral core.

The C-terminal domain of HIV-1 CA (the 'dimerization domain' [6,8]) is responsible for the self-association of this protein in dilute solution. The X-ray crystal structure of the C-terminal domain of HIV-1 CA shows that the dimer interface is built around the parallel packing of helix 2 across a twofold rotation axis [6,8]. A similar, but not identical, interaction is seen in a crystal structure of EIAV CA [9]. This dimer might be an intermediate in the capsid assembly pathway, such as occur during assembly of some icosahedral viruses.

The molecular surface of the C-terminal domain of CA is displayed for RSV, and for three other retroviruses, in Figure 4. On the front face of the domain (Figure 4a), a hydrophobic surface patch associated with the N terminus of helix 2 is relatively conserved, consistent with its persistent involvement in the domain interactions seen crystallographically [6–9]. On the reverse face of the domain (Figure 4b), there is another hydrophobic surface patch, which is of largest extent in RSV. The predominant contribution to this region is made by residues within the first and fourth helices of the domain, and its size suggests that it could be involved in intermolecular association. Although the burial of nonpolar residues might drive protein–protein association, the complementarity of the interacting regions will dictate the specific result. Changing the solution pH can profoundly affect the outcome of

Figure 5



Two discrete modes of assembly for RSV CA. **(a)** TEM image of part of a planar array of RSV CA, negatively stained with uranyl acetate. The planar array curls up, forming partially closed tubular structures, at the sheet boundaries (indicated with dark arrows). **(b)** TEM image of tubes of RSV CA, apparently formed from a toroidal oligomer. The specimen has been negatively stained with uranyl acetate. Free oligomers are indicated with dark arrows.

in vitro assembly studies on purified retroviral components [21,22], and, hence, the protonation state of individual amino acids might be critical to assembly.

Self-association of intact RSV CA

In vitro *assembly of RSV CA into regular structures*

Size-exclusion chromatography was used to characterize the behavior of intact RSV CA in dilute solution. Protein concentrations up to 780 μm were studied, at near neutral pH, with no change in behavior observed under reducing or oxidizing conditions. Purified CA eluted in a single peak, at a position consistent with a monomeric associative state. As expected for a monomeric species, the elution time did not depend on the concentration of the protein loaded onto the column. An N-terminally truncated HTLV-1 CA is also monomeric at concentrations up to 1000 μm [10]. In contrast, HIV-1 CA dimerizes with an equilibrium dissociation constant K_d = 10–30 μM, with dimerization apparently mediated by the C-terminal domain alone [6].

The structures of the two independently folded domains of RSV CA, and the behavior of CA in dilute solution [11], give few indications how the molecule might interact to form higher-order structures. Consequently, transmission electron microscopy (TEM) was used to study the *in vitro* assembly of CA. Intact CA (2–10 mg/ml) assembles into regular structures at mildly acidic pH, in the presence of 0.6–1.1 M NaCl. The *in vitro* assembly of HIV-1 CA to form tubular structures occurs under similar conditions [26,27]. At these concentrations, the salt acts to reduce protein solubility.

Two types of higher-order assembly were observed at such conditions (Figure 5). Most commonly, large planar arrays formed, spanning thousands of nanometers. The edges of these arrays are often curled up to form partially closed tubes with a varying external diameter of ~90 nm. The planar sheets were ordered, with Fourier transformation of images giving rise to discrete peaks, arranged on a regular lattice. The diffraction pattern can be indexed on a hexagonal lattice (approximate cell dimensions $a = b = 9.1$ nm, $\gamma = 60°$) or, alternatively, on a 'centered' rectangular lattice (approximate cell dimensions $a = 9.0$ nm, $b = 15.9$ nm, $\gamma = 90°$). Diffraction from well-preserved negatively stained arrays extends to approximately 25 Å resolution. Filtered images of untilted, negatively stained arrays possess approximate *C*2mm plane group symmetry. This would arise in projection if the layer group of the crystal was *C*222. This tentative layer group assignment implies that there are twofold symmetric interfaces that mediate the formation of the array.

The second mode of RSV CA assembly was observed more sporadically. In this case, the CA assembled into toroidal oligomers (mean diameter ~20 nm), possessing a stain-accessible central cavity (Figure 5). Neither mode of assembly described for RSV CA can exactly represent the situation within the virion, however, because the authentic viral core is a closed, irregular structure, that lacks a fixed geometry.

Comparison with previous results

The *in vitro* assembly of HIV-1 CA protein into tubular structures has been reported previously [26–28]. These tubes have a variable diameter of 40–50 nm and exhibit long-range order [28,29]. They appear to have a layered architecture, with the N-terminal domain on the outside of the cylinder and the C-terminal domain on the inside [28]. Morphologically similar tubes can be assembled from CA–NC fusion proteins in the presence of nucleic acids. Under suitable solution conditions, these tubes close off

into conical structures that closely resemble the mature core of HIV-1 [22,29]. Similarities in the helical diffraction patterns of the tubular structures, and the viral core, suggest similarities in their internal organization [2]. Thus, for HIV-1 there are clear parallels between the structures assembled *in vitro* and *in vivo*.

In this paper, two additional modes of assembly have been identified for RSV CA alone. Neither of the resulting structures appears macroscopically similar to the tubes of HIV-1 CA. The planar sheets of RSV CA have rectangular symmetry (tentative layer group assignment *C222*), whereas HIV-1 CA tubes reportedly possess hexagonal symmetry [29]. The tubular structures of RSV CA, built up from toroidal oligomers, are also dissimilar to the thin-walled hollow tubes of HIV-1 CA.

Biological implications

Rous sarcoma virus (RSV) is a tumorigenic avian retrovirus. The emergence of the related human immunodeficiency virus (HIV-1), and the resulting world-wide autoimmunodeficiency syndrome (AIDS) pandemic, has lent a new urgency to the study of such infectious agents. The RSV capsid protein (CA), like all the internal structural proteins of the virus, is first expressed as part of a polyprotein and is later liberated by the viral protease. CA forms the exterior shell of the viral core, which is delivered into the host cell in the initial stages of infection. The core in RSV resembles an irregular polyhedron, differing from the conical appearance of the core in lentiviruses, such as HIV-1. Despite knowledge of the three-dimensional structures of many retroviral components, the understanding of core architecture and assembly is rudimentary.

The structures of the N- and C-terminal domains of RSV CA are presented in this paper and are found to closely resemble those described for other retroviruses. Along with HIV-1, RSV is one of the best characterized molecular genetic systems for the study of retroviruses. Structures of the two domains of CA will be of great importance in interpreting phenotypes resulting from mutagenesis. In common with HIV-1, the N terminus of RSV CA forms a β hairpin, anchored in place by an internal salt bridge. Given that refolding of the N terminus following proteolysis might help direct capsid assembly in the maturing virus, the conservation of this structure across retroviral families is significant and has not been demonstrated previously. The N-terminal domain associates through a twofold symmetric interface within the crystal that does not directly involve the β hairpin.

***In vitro* assembly studies of intact RSV CA are also presented. The molecule is shown to form both tubes, constructed from a toroidal oligomer, and flat sheets. Both modes of assembly occur at mildly acidic pH in the presence of moderate salt concentrations. Image analysis of the planar sheets, which are the predominant assembly product, suggests that twofold symmetric interactions between domains mediate the formation of this structure. Although the *in vitro* assembly of HIV-1 CA into tubular structures has been reported previously, these do not macroscopically resemble the sheets or tubes of RSV CA. With *in vitro* assembly systems for both HIV-1 and RSV CA now described, it might be possible to understand the origin and significance of the different core morphologies of these retroviruses.**

Materials and methods

Structure determination of the N-terminal domain of RSV CA

Protein expression, purification and crystallization

DNA encoding the selected region of the RSV Gag protein (Prague strain C) [30] was amplified using the polymerase chain reaction, subcloned into the pET 3xc expression vector (Novagen), and transformed into the BL21 (DE3) pLysS strain of *E. coli* (Novagen). The expressed protein (CA_{1-154}) comprises the first 154 amino acids of mature CA and has a molecular weight of 16,480 Da.

The transformed cells were grown at 37°C in 2YT media, supplemented with ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml). Protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells harvested 4–5 h after induction. Cells were lysed using a French press. The crude cell lysate was suspended in 20 mM Tris/HCl buffer pH 8.0, containing 50 mM NaCl, 1 mM EDTA, 1 mM DTT, and 1 mM PMSF. Protamine sulfate (final concentration 10 mg/ml) was added to precipitate nucleic acids. Following centrifugation, CA_{1-154} was purified from the supernatant by size-exclusion chromatography. Superdex 75 preparation and analytical grade columns (Pharmacia) were used sequentially.

Selenomethionine-substituted protein was produced by down-regulation of methionine biosynthesis, using the same vector and host strain. Cells were grown in 2YT media until ready for induction. The cells were spun down, washed, and resuspended in M9 minimal media, supplemented with lysine, threonine, valine, phenylalanine, leucine and isoleucine, all at 100 μg/ml, and selenomethionine at 50 μg/ml. After 30 min growth, the cells were induced and harvested 4–5 h later. The incorporation of selenomethionine was confirmed using matrix-associated laser desorption ionisation (MALDI) mass spectrometry.

Crystals of CA_{1-154} were grown by vapor diffusion methods. Crystallization conditions were identified using experiments based on orthogonal arrays [31]. The protein (20–30 mg/ml in 20 mM Tris/HCL buffer, pH 8.0, 50 mM NaCl) was equilibrated at 4°C, against a reservoir solution containing 0.15 M Boric acid/KOH buffer, pH 9.1, 16–22% (w/v) PEG 6000, and 0.7 M $Mg(NO_3)_2$. It was necessary to use seeding procedures to obtain crystals of a useful size. Crystals of the selenomethionine-substituted protein were prepared in the same fashion. In the latter case, however, it was necessary to include the reducing agent Tris-carboxy ethyl phosphine (TCEP) at a concentration of 1–5 mM in order to prevent protein oxidation. Concentrations of all proteins studied in this paper were measured by spectrophotometric methods.

Crystals were transferred to a cryoprotective solution (0.15 M Boric acid/KOH buffer, pH 9.1, 24% (w/v) PEG 6000, 0.7 M $Mg(NO_3)_2$, 15% (v/v) ethylene glycol) and flash frozen in a cold gas stream for X-ray data collection. All diffraction data were collected at 113K. Data integration and scaling were performed with the programs Denzo and Scalepack [32].

A native data set, which was used for the refinement of the structural model, was collected from a single frozen crystal at the Advanced Photon Source (APS) in Argonne, Illinois (station 14-BM-c) (Table 1). The crystal space group is $P2_12_12_1$, with cell dimensions $a = 40.5$ Å, $b = 64.5$ Å, $c = 108.9$ Å. There are two molecules in the asymmetric unit of the crystal, related by an almost perfect twofold rotation. The Matthews coefficient, V_M, is 2.16 Å^3/Da, which corresponds to a solvent content of ~43%. Bragg diffraction from the crystals extends to approximately 2.0 Å resolution, but falls off in an anisotropic fashion.

Crystallographic phase determination by MAD

To phase the X-ray diffraction data, a MAD experiment, based on the selenium K-edge, was carried out at the APS (station 14-BM-d). Data were collected at five wavelengths, four of which were close to the X-ray absorption edge (Table 1). The inverse beam method was used to accumulate the Friedel pairs at each wavelength. Data collected at each wavelength were put on the same relative scale using the program FHSCAL [33]. Estimates for $|F_A|$, the structure factor amplitudes of the anomalously scattering selenium atoms, were derived using an algebraic procedure (RLK, unpublished observations). Direct methods coupled with phase annealing [34] were used to locate nine of ten selenium atoms. Phases were calculated and refined using the program MLPHARE [35]. This yielded a readily interpretable map, which was further improved using solvent flattening and histogram matching with the program DM [36].

Model building and refinement

The program Xfit from the XTALVIEW software package [37] was used for interactive model building. The model was refined with the program CNS [38], using torsion angle molecular dynamics procedures that incorporate a maximum-likelihood target function. Atomic displacements were modeled with an isotropic B value for each atom, with restraints on the B values of neighboring atoms. The contribution of the bulk solvent to the X-ray diffraction was modeled using a flat masking procedure [39]. No noncrystallographic symmetry restraints or constraints were applied during refinement. Five percent of the observed data was randomly selected for calculation of R_{free} and excluded from all refinement procedures. Statistics associated with the final refined model are presented in Table 1.

Structure determination of the C-terminal domain of RSV CA

Sample preparation

The two proteins encompassing the C-terminal domain of CA were expressed from cloned DNA segments as described for the N-terminal domain. One of the expressed proteins ($CA_{155–237}$) comprises the last 83 amino acids of CA, plus an initiating methionine at the N terminus, and has a molecular weight of 9212 Da. A second protein ($CA_{155–249}$) contains, in addition, the 12 amino acid spacer peptide and has a molecular weight of 9485 Da.

Uniformly ^{15}N- and ^{15}N/^{13}C-labeled proteins were prepared by growing the cells in M9 minimal media, containing $^{15}NH_4Cl$ (1 g/l) and $^{12}C_6$-D-glucose (2 g/l) or $^{15}NH_4Cl$ (1 g/l) and $^{13}C_6$-D-glucose (2 g/l), respectively, supplemented with ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml). Protein expression was induced late in log phase growth (A600 ~0.7) by addition of IPTG. The cells were harvested 4 h after induction and lysed either by freeze-thaw cycling or using a French press. The crude cell lysate was suspended in 50 mM sodium phosphate buffer, pH 7.5, containing 50 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol. Protamine sulfate (final concentration 10 mg/ml) was added to precipitate nucleic acids. Following centrifugation, the RSV CA fragments were purified from the supernatant by size-exclusion chromatography. Superdex 200 and Superdex 75 columns (Pharmacia) were used in sequence. Purified proteins were concentrated to 10–20 mg/ml in 50 mM sodium phosphate buffer, pH 4.9, containing 50 mM NaCl, 1 mM EDTA, and 1 mM deuterated dithiothreitol. 10% (v/v) $^2H_20$ was added to the samples before NMR data acquisition.

NMR spectroscopy and data processing

All spectra were recorded at 25°C on a Varian UnityPlus 600 MHz spectrometer, equipped with a z-shielded gradient triple-resonance probe. NMR data were analyzed with the programs NMRPipe [40] and ANSIG [41]. ^{1}H chemical shifts are reported relative to DSS. ^{13}C and ^{15}N chemical shifts were referenced indirectly using standard frequency ratios.

Steady-state heteronuclear ^{1}H–^{15}N NOEs were measured using the pulse sequences of Farrow *et al.* [42], with and without proton saturation. Approximate error estimates were derived from the intensity variation in background regions [42].

Spectral assignments

Sequential backbone assignments were made using a combination of ^{15}N-resolved [^{1}H,^{1}H]-NOESY, HNHA, HNHB[77], and J-correlation experiments (CBCANH, CBCACONH). Assignments of sidechain nuclei were made using a combination of J-correlation experiments (H(CCO)NH, C(CO)NH) and NOE-based approaches. Complete ^{1}H, ^{15}N and ^{13}C chemical shift assignment was achieved for residues 155–230 of the C-terminal domain. Stereospecific assignments have not been made. Chemical shift assignment within the unstructured C-terminal region (231–249) was complicated by chemical shift heterogeneity, resonance overlap, and limited NOE information. Partial assignments, including most of the backbone nuclei, have been made for this region.

Determination of conformational constraints

Distance restraints for the calculation of the 3D structure were derived from ^{13}C- or ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectra (mixing time 150 ms). Distances (D) were estimated using the isolated spin pair approximation, $D_{ij}^{-6} = AI_{ij}$, where A is a calibration factor determined using known distances within the α helices (H$\alpha(i) \rightarrow$ HN(i+3) ~3.4 Å). Upper (U) and lower (L) bounds for the structure calculation were then set to $L = D - 0.12D^2$, $U = D + 0.12D^2$ [43]. Ambiguous and degenerate NOE crosspeaks were treated using 'r^{-6} summation'. The ^{3}J(H^N,H$^\alpha$) and ^{3}J(N,H$^\beta$) coupling constants were derived from the HNHA and HNHB experiments, respectively [44]. Only the longer construct ($CA_{155–249}$) was isotopically enriched with both ^{15}N and ^{13}C, and restraints used in the structure calculation were derived from spectra recorded on this protein.

Calculation of the three-dimensional structure

Starting from an extended strand conformation, structures were calculated using torsion angle dynamics with the program CNS [38]. NOE-derived distances were restrained by a flat-bottomed parabolic function and a simplified repulsive function describing the van der Waals interactions. Restraints on backbone ϕ and ψ torsion angles were incorporated using an empirical relationship between these quantities and the $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts [45]. Direct refinement against vicinal scalar coupling constants was employed [46,47], with Karplus parameters relating coupling constants to torsion angles. Restraints based on ^{1}H chemical shifts were included for the non-amide protons [48].

For the final calculation, 100 structures were generated using torsion angle molecular dynamics. On the basis of agreement with the experimental restraints and geometric ideality 19 these structures were accepted (Figure 3). A mean structure was calculated from this ensemble and regularized by restrained energy minimization. Statistics associated with the final structure calculation are shown in Table 2.

Structure validation

The final model has 88% of the residues in the core region of the Ramachandran plot and the remainder (excepting Ser210) in the additional allowed regions (defined by the program PROCHECK). Ser210 is found in a poorly restrained surface loop connecting helices 3 and 4 of the domain (Figure 3). The final model contains no highly improbable sidechain χ_1,χ_2 choices [49]. It was also verified that most short interproton distances predicted by the model had corresponding NOE crosspeaks.

Arg170, which is an invariantly conserved residue within the MHR of the C-terminal domain, is poorly defined in the current model. Few NOE interactions were observed for the distal region of this sidechain, with exchangeable H^{η} protons. However, there are several indications that this residue actually adopts a fixed position within the structure. Firstly, the steady-state $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ NOE is positive, with a magnitude (0.63) that is only slightly smaller than the backbone ^{1}H–^{15}N NOEs. In comparison, the $^{1}H^{\varepsilon}$–$^{15}N^{\varepsilon}$ NOEs for the other four arginine residues in the C-terminal domain (Arg185, Arg194, Arg206 and Arg225) are less than 0.08. Secondly, the sidechain Hε of Arg170 experiences a strong downfield shift, resonating at 9.86 ppm (cf. the random coil chemical shift of 8.06 ppm), consistent with the involvement of this residue in a strong hydrogen-bonding network. Finally, a weak NOE crosspeak can be identified between Hε of Arg170 and Hα of either residue 166 or 167. This localizes the sidechain to the same region it occupies in the crystallographic structures of HIV-1 and EIAV CA [6,8,9].

Self-association of RSV CA

Behavior of the CA in dilute solution

Bacterial overexpression and purification of intact RSV CA has been described previously [18]. The behavior of the protein in dilute solution was studied by size-exclusion chromatography on a Superdex 75 column (Pharmacia). The protein was suspended in 25 mM EPPS/NaOH buffer, pH 8.0, containing 1 mM EDTA and 50 mM NaCl. The redox potential of the solution was controlled by including reduced and oxidized glutathione. Column runs were performed under both reducing (5.0 mM oxidized glutathione, 0.5 mM reduced glutathione) and oxidizing (0.5 mM oxidized glutathione, 5.0 mM reduced glutathione) conditions, with protein loading concentrations between 1 and 20 mg/ml.

Assembly of CA into higher order structures

Self-assembly of CA was studied using both vapor-diffusion and equilibrium dialysis techniques. In general, sitting-drop vapor-diffusion methods gave the most reproducible results. CA (2–10 mg/ml in 20 mM Tris/HCl buffer, pH 7.5, 1 mM EDTA, 50 mM NaCl) was equilibrated against NaCl solutions buffered at varying pH. To monitor assembly, the samples were applied to glow-discharged carbon/formvar-coated electron microscopy grids. Excess liquid was removed by blotting. The samples were stained with 1% (w/v) uranyl acetate, air-dried, and examined in the electron microscope. Assembly into regular structures occurred at mildly acidic pH, in the presence of moderate (0.6–1.1 M) NaCl concentrations. Typically, the NaCl solutions were buffered with 0.20 M citric acid/KOH pH 4.9, 0.20 M acetic acid/KOH pH 4.9, or 0.20 M malic acid/KOH pH 5.5. The temperature at which the experiments were performed was not critical and ranged between 4 and 37°C.

Electron microscopy and image processing

Images of the specimens were recorded using a Phillips EM420 electron microscope equipped with a lanthanum boride filament, operated at 80–100 kV. Micrographs were recorded on Kodak SO-163 photographic film, at a nominal magnification of 36000x, using a low-dose unit to reduce electron exposure. The micrographs were digitized on a Zeiss-SCAI scanner, with a step size of 7 or 14 μm.

The MRC image processing package was used for image analysis. Defocus values were determined from the spacing of the Thon rings in the computed Fourier transform of each image. For analysis of the planar periodic arrays, maxima in the Fourier transform were manually indexed, and the lattice refined by least-squares methods. Fourier coefficients were then derived directly from the computed transform, by determining the amplitude and phase at the positions predicted by the crystallographic reciprocal lattice.

Supplementary material

Supplementary material including a TEM image of RSV CA, a Fourier transform and projection map is available at http://current-biology.com/supmat/supmatin.htm.

Accession numbers

The atomic coordinates and raw data of the RSV domain structures have been deposited with the Protein Data Bank (accession numbers 1EM9 for the N-terminal domain and 1EOQ for the C-terminal domain).

Acknowledgements

We thank Alan Simpson for assistance with X-ray data collection and for useful discussions; Andy LiWang and Theresa Groesch for assistance with NMR spectroscopy; Terje Dokland, Norm Olson, and Tim Baker for advice on electron microscopy and image processing; Rebecca Craven and Tina Cairns for communicating their results in advance of publication; Bao Vuong for the purification of RSV CA; Swati Joshi for preliminary studies on the *in vitro* assembly of RSV CA; Wendy Breyer for critically reading the manuscript; and Brian Matthews for allowing RLK to complete this work while visiting his laboratory. We are grateful to the staff of the BioCARS beamlines at the Advanced Photon Source for their assistance. We apologize to the authors of the programs mentioned in the text, as well as other omissions to original work on account of space limitations imposed by the journal. This work was supported by National Institutes of Health grants to MGR, VMV and CBP, as well as a Purdue University reinvestment grant and a grant from the Purdue Cancer Center.

References

1. Turner, B.G. & Summers, M.F. (1999). Structural biology of HIV. *J. Mol. Biol.* **285**, 1-32.
2. Wilk, T. & Fuller, S.D. (1999). Towards the structure of the human immunodeficiency virus: divide and conquer? *Curr. Opin. Struct. Biol.* **9**, 231-243.
3. Gamble, T.R., *et al.*, & Hill, C.P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285-1294.
4. Gitti, R.K., Lee, B.M., Walker, J., Summers, M.F., Yoo, S. & Sundquist, W.I. (1996). Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* **273**, 231-235.
5. Momany, C., *et al.*, & Rossmann, M.G. (1996). Crystal structure of dimeric HIV-1 capsid protein. *Nat. Struct. Biol.* **3**, 763-770.
6. Gamble, T.R., *et al.*, & Hill, C.P. (1997). Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* **278**, 849-853.
7. Berthet-Colominas, C., Monaco, S., Novelli, A., Siba, G., Mallet, F. & Cusack, S. (1999). Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *EMBO J.* **18**, 1124-1136.
8. Worthylake, D.K., Wang, H., Yoo, S., Sundquist, W.I. & Hill, C.P. (1999). Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution. *Acta Crystallogr. D* **55**, 85-92.
9. Jin, Z., Jin, L., Peterson, D.L. & Lawson, C.L. (1999). Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. *J. Mol. Biol.* **286**, 83-93.
10. Khorasanizadeh, S., Campos-Olivas, R. & Summers, M.F. (1999). Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *J. Mol. Biol.* **291**, 491-505.
11. Mammano, F., Öhagen, Å., Höglund, S. & Göttlinger, H.G. (1994). Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* **68**, 4927-4936.
12. Craven, R.C., Leure-duPree, A.E., Weldon, R.A.J. & Wills, J.W. (1995). Genetic analysis of the major homology region of Rous sarcoma virus Gag protein. *J. Virol.* **69**, 4213-4227.
13. Krishna, N.K., Campbell, S., Vogt, V.M. & Wills, J.W. (1998). Genetic determinants of Rous sarcoma virus particle size. *J. Virol.* **72**, 564-577.
14. Craven, R.C., Leure-duPree, A.E., Erdie, C.R., Wilson, C.B. & Wills, J.W. (1993). Necessity of the spacer peptide between CA and NC in the Rous sarcoma virus Gag protein. *J. Virol.* **67**, 6246-6252.
15. Pepinsky, R.B., Papayannopoulos, I.A., Chow, E.P., Krishna, N.K., Craven, R.C. & Vogt, V.M. (1995). Differential proteolytic processing leads to multiple forms of the CA protein in avian sarcoma and leukemia viruses. *J. Virol.* **69**, 6430-6438.
16. Accola, M.A., Höglund, S. & Göttlinger, H.G. (1998). A putative α-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J. Virol.* **72**, 2072-2078.
17. Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H. &

Kräusslich, H.G. (1998). Sequential steps in human immunodeficiency virus particle release revealed by alterations of individual Gag polyprotein cleavage sites. *J. Virol.* **72**, 2846-2854.
18. Kovari, L.C., *et al*., & Rossmann, M.G. (1997). Crystals of Rous sarcoma virus capsid protein show a helical arrangement of protein subunits. *Virology* **238**, 79-84.
19. Campos-Olivas, R. & Summers, M.F. (1999). Backbone dynamics of the N-terminal domain of the HIV-1 capsid protein and comparison with the G94D mutant conferring cyclosporin resistance/dependence. *Biochemistry* **38**, 10262-10271.
20. Gross, I., Hohenberg, H., Huckhagel, C. & Kräusslich, H.G. (1998). N-terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. *J. Virol.* **72**, 4798-4810.
21. Campbell, S. & Vogt, V.M. (1997). In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J. Virol.* **71**, 4425-4435.
22. Gross, I., *et al*., & Kräusslich, H.G. (2000). A conformational switch controlling HIV-1 morphogenesis. *EMBO J.* **19**, 103-113.
23. Street, A.G. & Mayo, S.L. (1999). Intrinsic β-sheet propensities result from van der Waals interactions between sidechains and the local backbone. *Proc. Natl Acad. Sci. USA* **96**, 9074-9076.
24. Wishart, D.S. & Sykes, B.D. (1994). The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4**, 171-180.
25. Pepinsky, R.B. (1983). Localization of lipid–protein and protein–protein interactions within the murine retrovirus *gag* precursor by a novel peptide-mapping technique. *J. Biol. Chem.* **258**, 11229-11235.
26. Ehrlich, L.S., Agresta, B.E. & Carter, C.A. (1992). Assembly of recombinant human immunodeficiency virus type 1 capsid protein *in vitro*. *J. Virol.* **66**, 4874-4883.
27. Gross, I., Hohenberg, H. & Kräusslich, H.G. (1997). In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.* **249**, 592-600.
28. Grättinger, M., Hohenberg, H., Thomas, D., Wilk, T., Müller, B. & Kräusslich, H.G. (1999). *In vitro* assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* **257**, 247-260.
29. Ganser, B.K., Li, S., Klishko, V.Y., Finch, J.T. & Sundquist, W.I. (1999). Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80-83.
30. Schwartz, D.E., Tizard, R. & Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. *Cell* **32**, 853-869.
31. Kingston, R.L., Baker, H.M. & Baker, E.N. (1994). Search designs for protein crystallization based on orthogonal arrays. *Acta Crystallogr. D* **50**, 429-440.
32. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
33. Tickle, I.J. (1991). Refinement of single isomorphous replacement heavy-atom parameters in Patterson vs reciprocal space. In *Isomorphous Replacement and Anomalous Scattering. Proceedings of the CCP4 Study Weekend, 25-26 January 1991.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 87-95, Science and Engineering Research Council, Daresbury.
34. Sheldrick, G.M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467-473.
35. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering. Proceedings of the CCP4 Study Weekend, 25-26 January1991.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80-86, Science and Engineering Research Council, Daresbury, Warrington, UK.
36. Collaborative Computational Project Number 4 (1994). The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
37. McRee, D.E. (1999). XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
38. Brünger, A.T., *et al*., & Warren, G.L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
39. Jiang, J.S. & Brünger, A.T. (1994). Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100-115.
40. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.
41. Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. & Laue, E.D. (1994). Solution structure and dynamics of Ras p21•GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* **33**, 3515-3531.
42. Farrow, N.A., *et al*., & Kay, L.E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984-6003.
43. Nilges, M., Macias, M.J., O'Donoghue, S.I. & Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.* **269**, 408-422.
44. Bax, A., *et al*., & Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239**, 79-105.
45. Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. (1995). The impact of direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. B* **106**, 92-96.
46. Garret, D.S., *et al*., & Clore, G.M. (1994). The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson. B* **104**, 99-103.
47. Constantine, K.L., Friedrichs, M.S., Mueller, L. & Bruccoleri, R.E. (1995). J-coupling restraint potentials for nonstereospecifically assigned methylene protons and ensemble-average calculations. *J. Magn. Reson. B* **108**, 176-184.
48. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. (1996). A potential involving multiple proton chemical-shift restraints for nonstereospecifically assigned methyl and methylene protons. *J. Magn. Reson. B* **112**, 79-81.
49. Dunbrack, R.L.J. & Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661-1681.
50. Carson, M. (1997). Ribbons. *Methods Enzymol.* **277**, 493-505.
51. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
52. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. & Zehfus, M.H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838.
53. Sanner, M.F., Olson, A.J. & Spehner, J.C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305-320.