HHS Public Access

Author manuscript

*Proteins*. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

*Proteins*. 2022 January ; 90(1): 309–313. doi:10.1002/prot.26204.

# Structural evidence that MOAP1 and PEG10 are derived from retrovirus/retrotransposon Gag proteins

**Katarzyna Zurowska**, **Ayaan Alam**, **Barbie K. Ganser-Pornillos**, **Owen Pornillos**

Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia, USA

## Abstract

The Gag proteins of retroviruses play an essential role in virus particle assembly by forming a protein shell or capsid and thus generating the virion compartment. A variety of human proteins have now been identified with structural similarity to one or more of the major Gag domains. These human proteins are thought to have been evolved or "domesticated" from ancient integrations due to retroviral infections or retrotransposons. Here, we report that X-ray crystal structures of stably folded domains of MOAP1 (modulator of apoptosis 1) and PEG10 (paternally expressed gene 10) are highly similar to the C-terminal capsid (CA) domains of cognate Gag proteins. The structures confirm classification of MOAP1 and PEG10 as domesticated Gags, and suggest that these proteins may have preserved some of the key interactions that facilitated assembly of their ancestral Gags into capsids.

### Keywords

crystal structure; capsid-like protein; major homology region; structural conservation

## Introduction

The Gag proteins of retroviruses and retrotransposons assemble into virus particles or capsids which organize the enzymatic reactions that reverse transcribe their encoding RNA into DNA and integrate the DNA copy into the host cell's genome. A number of human proteins have now been identified that are structurally related to retrovirus or retrotransposon Gag[1–4]. These proteins (which we term here "DGags") are thought to have acquired new cellular functions or have been "domesticated" over evolutionary time, following some ancient integration events[5]. DGag proteins are largely understudied, but are annotated as having functional links to diverse cellular processes such as transcriptional regulation, apoptosis, differentiation and neurological signaling[5]. DGags are also implicated in human disease and identified as markers for neurological disorder and cancer[6–8].

Gag proteins contain three major structure-function modules, called MA, CA and NC (reviewed in ref.[9]). MA (matrix) binds to lipid bilayers and targets assembling retrovirus particles to the plasma membrane. CA (capsid) consists of two independently-folding N-

Correspondence: opornillos@virginia.edu, bpornillos@virginia.edu.

terminal and C-terminal domains (NTD and CTD), which mediate assembly of a protein shell compartment or capsid. NC (nucleocapsid) contains one or more zinc knuckles that bind nucleic acids and package the retroviral or retrotransposon RNA within the capsid. Over evolutionary time, different DGags have retained various combinations of these modules, presumably in concordance with the various selective pressures that delineated their new functions. For example, the SCAN protein family is a large group of DGags (71 identified members in human) that have retained CA-CTD-like and NC-like domains but have lost all the other Gag domains[5,10]. SCAN proteins have re-purposed the original oligomerization activity of CA-CTD and nucleic acid binding activity of NC to now function as dimeric transcription factors. Another example is the mammalian Arc protein (and its paralogs in *Drosophila*), which retained not only the structural folds of the NTD and CTD CA domains but also the full set of quaternary interactions across these domains (i.e., NTD-NTD, NTD-CTD and CTD-CTD) that allow them to assemble into virus-like or capsid-like particles[4,11,12]. The assembled Arc particles are proposed to transfer RNA from cell to cell, much like a real retrovirus[13,14]. This ability to assemble into a virus-like compartment represents a novel mechanism of sub-cellular organization and signal transmission.

We asked whether other putative DGag proteins have also retained the ability to form virus-like capsids, and here focus on the human orthologs of MOAP1 (modulator of apoptosis 1) and PEG10 (paternally expressed gene 10). Along with Arc, MOAP1 and PEG10 were identified in a genome-wide study to potentially have retained CA-like assembly modules with both the NTD and CTD[5]. However, neither structural nor functional confirmation that MOAP1 and PEG10 are CA-like is available. We now report here the X-ray crystal structures of the CA CTD-like domains of MOAP1 and PEG10, and confirm that these proteins are indeed structural homologs of retrovirus and retrotransposon Gag proteins.

## Methods

The MOAP1 and PEG10 CA-like C-terminal domains (which we term here MOAP1-CTD and PEG10-CTD, respectively) were each expressed in *E. coli* BL21(DE3) cells through IPTG induction, and initially purified on Ni-NTA resin (Qiagen). MOAP1-CTD contained a $His_6$-SUMO leader sequence, which was removed by cleavage with Ulp1 protease. PEG10-CTD contained a non-cleavable $His_6$ leader sequence. Both proteins were further purified to homogeneity by using ion exchange and size exclusion chromatography, and exchanged into buffer (10 mM Tris, pH 8.0, 75 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine (TCEP)) during the sizing step.

Crystals were grown at 20 °C in hanging drops by mixing equal volumes of precipitant and protein solution (8–10 mg/mL). MOAP1-CTD crystallized in 0.1 M Tris, pH 7.5, 20% (*w*/*v*) PEG 3,350, 3% (*v*/*v*) hexanediol. Selomethionine-labeled PEG10-CTD crystallized in 0.1 M HEPES, pH 7.5, 25% (*w*/*v*) PEG 3,350, 0.2 M NaCl. The crystals were cryoprotected with 25% (*v*/*v*) glycerol or ethylene glycol and flash-frozen in liquid nitrogen. Diffraction data were collected at SER-CAT beamlines 22ID and 22BM at the Advanced Photon Source. Indexing, peak integration and scaling were performed with HKL2000[15]. MOAP1-CTD was initially solved through single anomalous diffraction (SAD) methods from a low-resolution data set (5 Å) derived from an osmium soak of a P321 crystal, which allowed building of

a partial backbone model. The partial model was then used to phase the high-resolution (1.85 Å) P1 crystal reported here by using molecular replacement. PEG10-CTD was solved by SAD methods from a high-resolution (1.9 Å) data set collected from a selenomethionine-labeled crystal. Structure refinement and model building were performed using the tools in PHENIX[16] and Coot[17]. Figures were prepared with PyMol (Delano Scientific). Structure statistics are provided in Table 1. Coordinates and structure factors are deposited at the RCSB PDB database with ID codes 7LGC (MOAP1-CTD) and 7LGA (PEG10-CTD).

## Results and Discussion

The primary sequences of cognate Gags and DGags are widely divergent, and pair-wise comparisons cannot be used to precisely delineate domain boundaries in MOAP1 or PEG10. We therefore used secondary structure predictions to guide the design of a series of test constructs that spanned the putative CA-like NTDs and CTDs (separately and in tandem) for overexpression in *E. coli*, with the expectation that soluble, purifiable and biochemically well-behaved constructs are properly folded. Among the well-behaved fragments are constructs that putatively corresponded to MOAP1-CTD (residues 243–352) and PEG10-CTD (residues 161–252) (Fig. 1A). These were purified to homogeneity, crystallized and then structures were solved and refined against high resolution X-ray diffraction data (Table 1).

As illustrated in Figure 1, both MOAP1-CTD (Fig. 1B) and PEG10-CTD (Fig. 1C) consist of five α-helices that pack into the same tertiary fold. Despite the lack of readily recognizable sequence identity to retroviruses or retrotransposons, both structures are clearly related to the CA-CTDs of HIV-1 (a retrovirus; Fig. 1D) and Ty3 (a retrotransposon; Fig. 1E). We therefore conclude that MOAP1 and PEG10 are, indeed, Gag-like proteins. The N-terminal helices (colored in purple) of MOAP1-CTD and PEG10-CTD are more elongated compared to the corresponding helices in HIV-CTD and other retrovirus CA-CTDs but similar to that in Ty3-CTD. The C-terminal helices (colored in green) are significantly extended in MOAP1-CTD and PEG10-CTD compared to retroviruses and retrotransposons. Interestingly, in some retroviruses this region can also adopt extended helical configurations to mediate oligomerization and higher-order assembly of immature capsid forms[18–22].

An important function of CA-CTD in retrovirus and retrotransposon Gag proteins is to mediate inter-subunit dimerization interactions during capsid assembly[23] (Fig. 2). The dimer interface consists of the first (colored in purple) and third (pink) helices, as well as portions of the strand/turn/helix element that connects these two helices. This dimerization interaction is very weak and with few exceptions is not detectable in solution, but readily manifests in crystal packing contacts and in assembled capsids or capsid-like particles. Correspondingly, both the purified MOAP1-CTD and PEG10-CTD proteins were also monomeric in solution, even at concentrations approaching 1 mM (data not shown). However, analysis of the crystal lattice packing interactions indicated that both proteins may have crystallized as dimers. In particular, the PEG10-CTD crystals contained two copies of a non-crystallographic dimer in the P21 unit cell (Fig. 2B). This non-crystallographic dimer is remarkably similar to the HIV-1 CA-CTD dimers observed in solution[24] and in assembled HIV-1 capsids[25] (Fig. 2C) and the CA-CTD dimer in Ty3 capsids[26] (Fig. 2D); the close

correspondence is illustrated by superposition of the PEG10-CTD and Ty3-CTD structures as dimeric units (Fig. 2E). There is, therefore, a high likelihood that PEG10 has retained this oligomerization activity. In the case of MOAP1-CTD, the P1 unit cell also contained two monomers that pack in a dimeric fashion (Fig. 2A) that is superficially similar to but distinct from the PEG10-CTD and viral CA-CTD dimers. The putative MOAP1-CTD dimer interface consists only of the N-terminal helix.

Retrovirus CA-CTDs also contain a 20-residue segment called the MHR (major homology region), which is so-called because it is the most highly conserved primary sequence motif in Gag[27] (Fig. 3A). The MHR sequence was historically thought to be highly conserved in retrotransposons as well, but this turned out not to be the case[26]. In all known retrovirus CA-CTD structures, the MHR adopts a tightly packed strand-turn-helix configuration (Fig. 3C). As first described for HIV-CTD, this configuration is stabilized by a network of hydrogen bonds and salt bridges between glutamine, glutamate and arginine sidechains that appeared to be absolutely conserved within the MHR motif[23]. Additionally conserved positions are occupied by large hydrophobic sidechains – typically phenylalanine, tyrosine or leucine – which form part of the hydrophobic core of the CTD. We found that MOAP1-CTD contains all the absolutely conserved and additionally conserved residues in its MHR motif (Fig. 3A). These residues form the same network of interactions seen in retrovirus CA-CTDs (Fig. 3B). Remarkably, the three-dimensional configuration of the MOAP1-CTD MHR backbone is essentially identical to that of HIV-CTD (Fig. 3E). In contrast, the MHR motif of PEG10-CTD lacks the glutamate and arginine sidechains that generate the hydrogen bond and salt bridge network (Fig. 3A). Although the PEG10-CTD MHR still adopts a strand-turn-helix fold (likely because the conserved hydrophobic sidechains are in place) (Fig. 3D), the three-dimensional backbone configuration is significantly divergent from those of MOAP1-CTD and HIV-CTD (Fig. 3E).

In conclusion, the results of our structural studies support the classification of MOAP1 and PEG10 as domesticated Gag proteins. Close examination of the structural details reveal that these proteins may have preserved key features – namely, an intact MHR and/or the ability to dimerize – which, in cognate Gag proteins, are functionally critical for virion morphogenesis and capsid assembly[23,28]. We speculate that the conservation of these structural features in these DGags may indicate conservation of the assembly function as well. Further experiments are underway to test this hypothesis.

## Acknowledgements

We thank Yueping Wan, Kaneil Zadrozny, Jacint Sanchez and Jonathan Wagner for technical assistance. This study was funded by NIH grant R01-AI129678 (BKG-P and OP).

## References

1. Ivanov D, Stone JR, Maki JL, Collins T, Wagner G. Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. Mol Cell. 2005;17(1):137–143. [PubMed: 15629724]
2. Peterson FC, Hayes PL, Waltner JK, et al. Structure of the SCAN domain from the tumor suppressor protein MZF1. J Mol Biol. 2006;363(1):137–147. [PubMed: 16950398]

3. Liang Y, Huimei Hong F, Ganesan P, et al. Structural analysis and dimerization profile of the SCAN domain of the pluripotency factor Zfp206. Nucleic Acids Res. 2012;40(17):8721–8732. [PubMed: 22735705]
4. Zhang W, Wu J, Ward MD, et al. Structural basis of arc binding to synaptic proteins: implications for cognitive disease. Neuron. 2015;86(2):490–500. [PubMed: 25864631]
5. Campillos M, Doerks T, Shah PK, Bork P. Computational characterization of multiple Gag-like human proteins. Trends Genet. 2006;22(11):585–589. [PubMed: 16979784]
6. Okabe H, Satoh S, Furukawa Y, et al. Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. Cancer Res. 2003;63(12):3043–3048. [PubMed: 12810624]
7. Bi R, Kong LL, Xu M, et al. The Arc Gene Confers Genetic Susceptibility to Alzheimer's Disease in Han Chinese. Mol Neurobiol. 2018;55(2):1217–1226. [PubMed: 28108859]
8. Watson KM, Gardner IH, Byrne RM, et al. Differential Expression of PEG10 Contributes to Aggressive Disease in Early Versus Late-Onset Colorectal Cancer. Dis Colon Rectum. 2020;63(12):1610–1620. [PubMed: 33149023]
9. Freed EO. HIV-1 assembly, release and maturation. Nat Rev Microbiol. 2015;13(8):484–496. [PubMed: 26119571]
10. Sander TL, Stringer KF, Maki JL, Szauter P, Stone JR, Collins T. The SCAN domain defines a large family of zinc finger transcription factors. Gene. 2003;310:29–38. [PubMed: 12801630]
11. Cottee MA, Letham SC, Young GR, Stoye JP, Taylor IA. Structure of *Drosophila melanogaster* ARC1 reveals a repurposed molecule with characteristics of retroviral Gag. Sci Adv. 2020;6(1):eaay6354. [PubMed: 31911950]
12. Erlendsson S, Morado DR, Cullen HB, Feschotte C, Shepherd JD, Briggs JAG. Structures of virus-like capsids formed by the Drosophila neuronal Arc proteins. Nat Neurosci. 2020;23(2):172–175. [PubMed: 31907439]
13. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T. Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. Cell. 2018;172(1–2):262–274.e211. [PubMed: 29328915]
14. Pastuzyn ED, Day CE, Kearns RB, et al. The neuronal gene Arc encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. Cell. 2018;172(1–2):275–288.e218. [PubMed: 29328916]
15. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997;276:307–326.
16. Adams PD, Afonine PV, Bunkóczi G, et al. The Phenix software for automated determination of macromolecular structures. Methods. 2011;55(1):94–106. [PubMed: 21821126]
17. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486–501. [PubMed: 20383002]
18. Schur FK, Dick RA, Hagen WJ, Vogt VM, Briggs JA. The structure of immature virus-like Rous sarcoma virus Gag particles reveals a structural role for the p10 domain in assembly. J Virol. 2015;89(20):10294–10302. [PubMed: 26223638]
19. Wagner JM, Zadrozny KK, Chrustowics J, et al. Crystal structure of an HIV assembly and maturation switch. Elife. 2016;5:e17063. [PubMed: 27416583]
20. Schur FK, Obr M, Hagen WJ, et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. Science. 2016;353(6298):506–508. [PubMed: 27417497]
21. Qu K, Glass B, Doležal M, et al. Structure and architecture of immature and mature murine leukemia virus capsids. Proc Natl Acad Sci U S A. 2018;115(50):E11751–E11760. [PubMed: 30478053]
22. Dick RA, Xu C, Morado DR, et al. Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. PLoS Pathog. 2020;16(1):e1008277. [PubMed: 31986188]
23. Gamble TR, Yoo S, Vajdos FF, et al. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science. 1997;278(5339):849–853. [PubMed: 9346481]
24. Byeon IJ, Meng X, Jung J, et al. Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. Cell. 2009;139(4):780–790. [PubMed: 19914170]

25. Mattei S, Glass B, Hagen WJ, Kräusslich HG, Briggs JA. The structure and flexibility of conical HIV-1 capsids determined within intact virions. Science. 2016;354(6318):1434–1437. [PubMed: 27980210]

26. Dodonova SO, Prinz S, Bilanchone V, Sandmeyer S, Briggs JAG. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. Proc Natl Acad Sci U S A. 2019;116(20):10048–10057. [PubMed: 31036670]

27. Wills JW, Craven RC. Form, function, and use of retroviral gag proteins. AIDS. 1991;5(6):639–654. [PubMed: 1883539]

28. Mammano F, Ohagen A, Höglund S, Göttlinger HG. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. J Virol. 1994;68(8):4927–4936. [PubMed: 8035491]

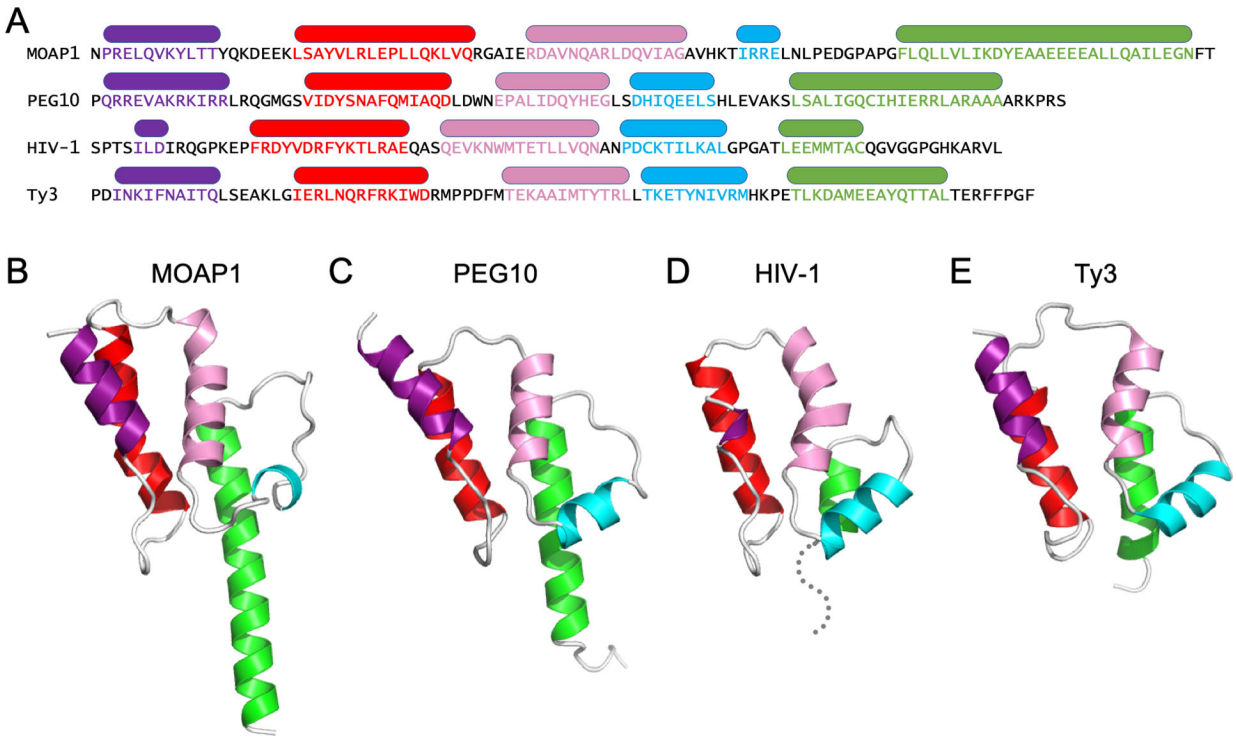


**Figure 1.**

A, Primary sequence and secondary structures of the CA-like CTDs of MOAP1 and PEG10, along with those of CA-CTDs from the retrovirus HIV-1 and retrotransposon Ty3. B, Crystal structure of MOAP1-CTD. C, Crystal structure of PEG10-CTD. D, Crystal structure of the mature form of the CA-CTD of HIV-1 (PDB ID: 2A43[23]). E, CryoEM structure of the mature form of the CA-CTD of Ty3, as found in assembled capsids (PDB ID: 6R23[26]).

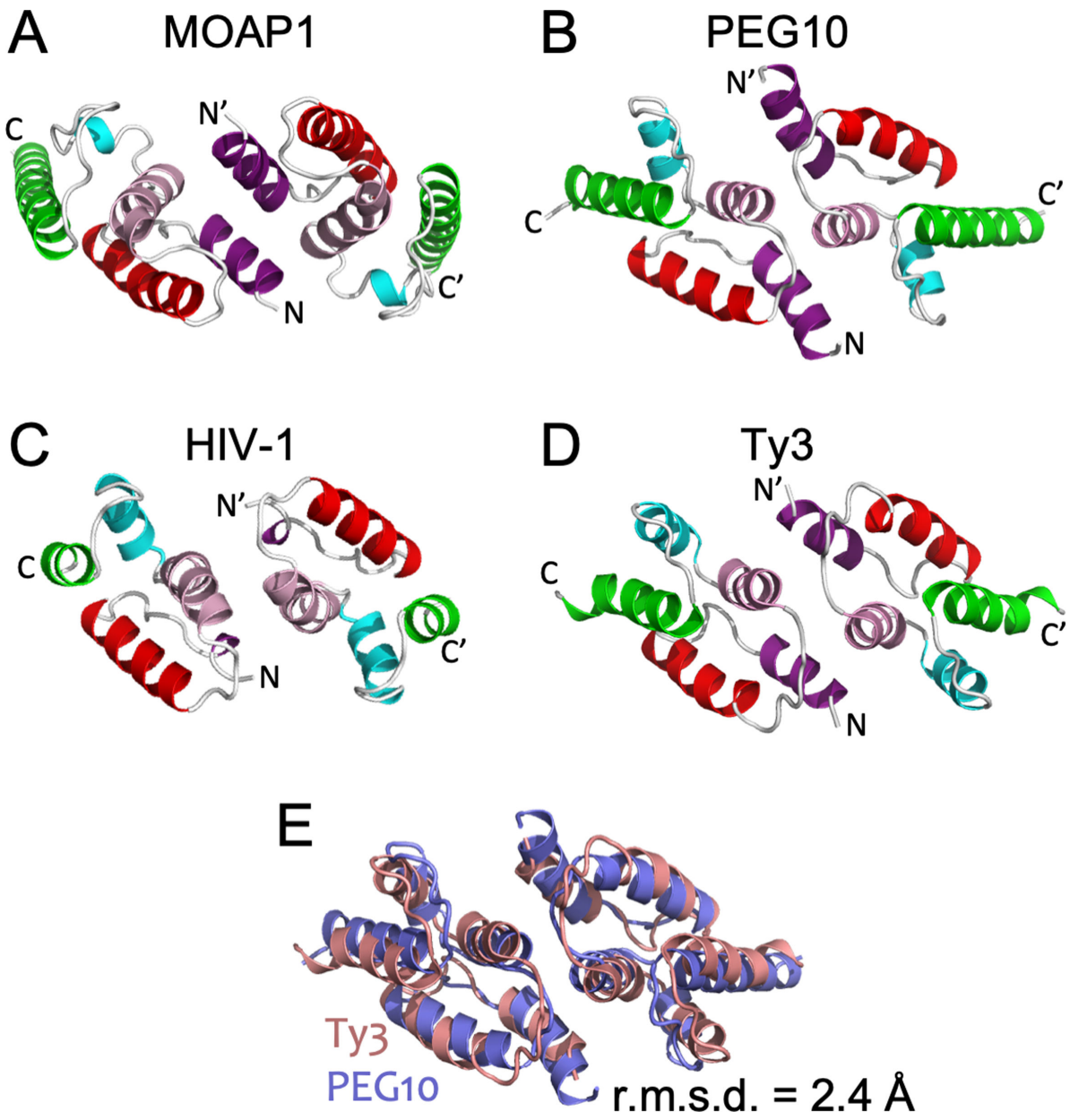


**Figure 2.**

A, Crystallographically observed dimer of MOAP1-CTD. B, Crystallographically observed dimer of PEG10-CTD. C, HIV-1 CA-CTD dimer in solution (PDB ID: 2KOD[24]). D, Ty3 CA-CTD dimer in assembled capsids (PDB ID: 6R23[26]). In A-D, N-terminal and C-terminal ends are indicated, with the two monomers distinguished by an apostrophe (‘). E, Backbone-guided superposition of the PEG10-CTD and Ty3-CTD structures, which were superimposed as dimeric units. r.m.s.d. = root mean square deviation over equivalent Cα atoms.

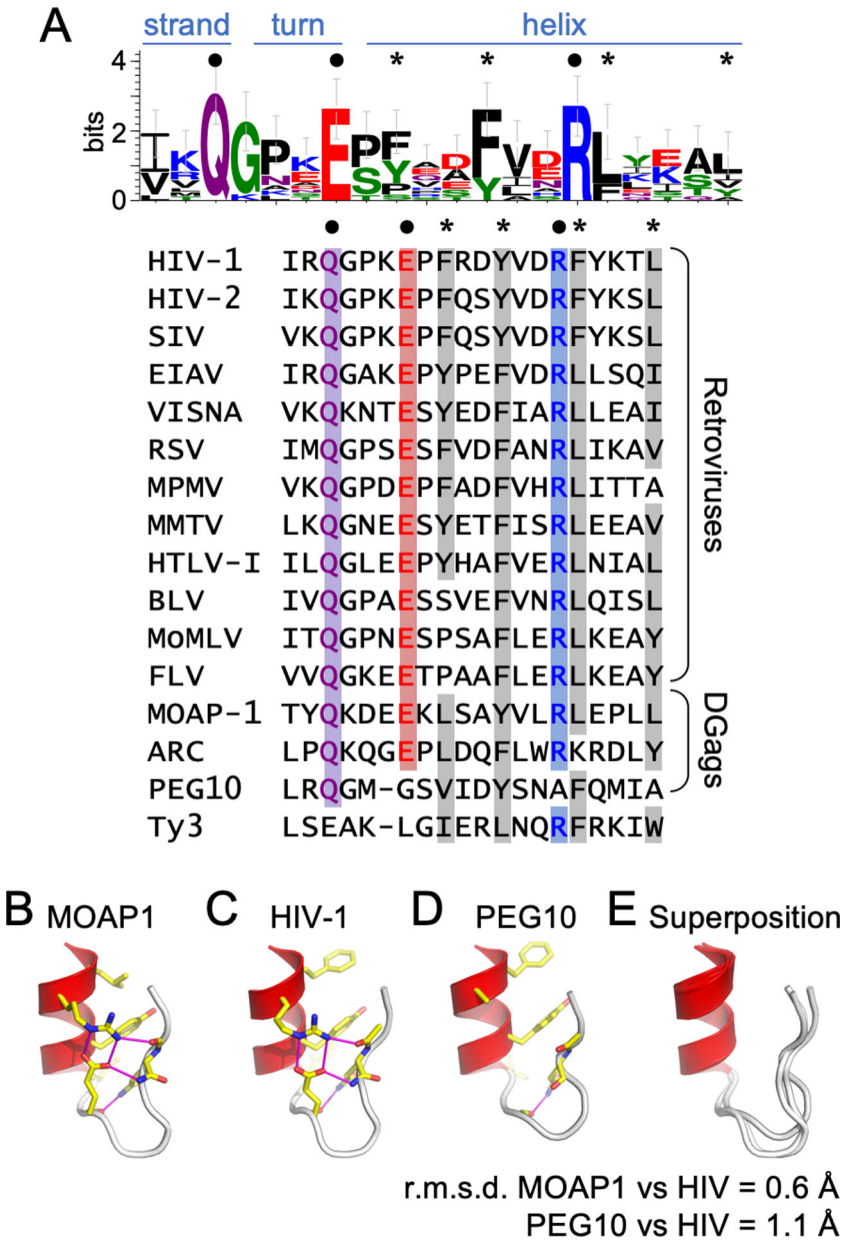


**Figure 3.**

A, Primary sequence conservation in the MHR (major homology region) of retroviruses. •, absolutely conserved polar and charged residues that comprise the hydrogen bond and salt bridge network; *, additionally conserved hydrophobic residues that stabilize the strand-turn-helix MHR fold. B, The MHR of MOAP1-CTD. C, The MHR of PEG10-CTD. D, The MHR of HIV-CTD (PDB ID: 2A43[23]). E, The MHR backbone configurations of MOAP1-CTD and HIV-CTD are highly similar, whereas PEG10-CTD is more divergent. r.m.s.d. = root mean square deviation over equivalent backbone atoms.

**Table 1.**

Crystallographic statistics.

| | MOAP1-CTD | PEG10-CTD |
|---|---|---|
| PDB ID | 7LGC | 7LGA |
| Diffraction | | |
| Beamline | APS 22ID | APS 22BM |
| Wavelength, Å | 1.000 | 0.979 |
| Space group | P1 | P21 |
| Unit cell | | |
| *a b c*, Å | 34.0, 34.1, 57.7 | 51.60, 70.29, 56.89 |
| α β γ, ° | 97.7, 93.9, 94.6 | 90, 98.1, 90 |
| Resolution range, Å | 50–1.85 (1.88–1.85) | 50–1.90 (1.93–1.90) |
| $R_{merge}/R_{pim}$ | 0.08 (0.80) / 0.04 (0.60) | 0.09 (0.68) / 0.03 (0.39) |
| CC½ | 0.999 (0.536) | 0.992 (0.644) |
| Mean I/σ<I> | 18.3 (1.3) | 19.6 (1.4) |
| Completeness, % | 94.8 (65.5) | 98.4 (81.2) |
| Average redundancy | 4.8 (1.7) | 7.2 (3.5) |
| Wilson B-factor, Å$^2$ | 18.81 | 16.32 |
| Refinement | | |
| Resolution range | 30.16–1.85 (1.97–1.85) | 41.90–1.90 (1.93–1.90) |
| No. of unique reflections | 19,264 (1,707) | 48,308 (404) |
| No. of reflections in free set | 893 (89) | 3,441 (33) |
| $R_{work}/R_{free}$, % | 20.7/25.5 (25.8/29.9) | 18.6/20.9 (29.6/34.0) |
| No. of nonhydrogen atoms | | |
| protein | 1,600 | 2,735 |
| water | 172 | 225 |
| Average B-factor, Å$^2$ | | |
| protein | 24.54 | 25.27 |
| water | 29.15 | 28.67 |
| Coordinate deviations | | |
| bond lengths, Å | 0.006 | 0.006 |
| bond angles, ° | 0.757 | 0.763 |
| Ramachandran statistics | | |
| favored, % | 98.5 | 99.7 |
| outliers, % | 0 | 0 |
| MolProbity clashscore | 0.62 | 0.74 |

Values in parenthesis are for the highest resolution shell.