STRUCTURAL BASIS OF STABILITY OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) CAPSID

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

By

Anna Gres

Dr. John J. Tanner, Dissertation Supervisor

Dr. Stefan G. Sarafianos, Dissertation Supervisor

DECEMBER 2017

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

STRUCTURAL BASIS OF STABILITY OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) CAPSID

Presented by Anna Gres

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

Professor John J. Tanner

Professor Stefan G. Sarafianos

Professor Lesa J. Beamer

Professor Kent S. Gates

To my parents Tadevush and Natallia Hres

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my mentor Dr. Sarafianos who was a great advisor and taught me how to think critically, communicate science more efficiently, and collaborate with others. He is a great example for me, and I will be more than happy to continue learning from him if I have the opportunity. I also want to thank my other mentor Dr. John J. Tanner for his honest opinion, critical comments, and professional insights. I am very grateful to committee members, Dr. Lesa J. Beamer and Dr. Kent S. Gates for their time, constructive criticism and valuable input during my committee meetings and Structural Biology Group meetings.

Special thanks to our collaborators Dr. Eric Freed, Dr. Emico Urano and Dr. Mariia Novikova from NIH; Dr. Veneet KewalRamani and Dr. Guangai Xue from NIH; Dr. Donald Burke-Agüero and Dr. Margaret Lange from MMI Department of the University of Missouri; Dr. Arthur Olson, Dr. Stefano Forli and Dr. Pierrik Craveur from the Scripps Research Institute; Dr. Christopher Aiken and Jiong Shi from Department of Pathology, Immunology & Microbiology of the Vanderbilt University; Dr. Kalus Schulten, Dr. Juan Perilla and Chaoyi Xu from Department of Physics & Beckman Institute of the University of Illinois at Urbana-Champaign; Dr. Peijun Zhang and Xiaofeng Fu from Department of Structural Biology of the University of Pittsburgh. These collaborations have been vital for the progress of my research.

This research was supported in part by the National Institutes of Health (AI112417, AI120860, and GM103368 to SGS). I thank Jay Nix of ALS beamline 4.2.2 for assistance with data collection. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under

Contract No. DE-AC02-05CH11231. I am also grateful to Ruslan Sanishvili and all the staff of APS Sector 23 (GM/CA-CAT) for helpful discussions regarding data collection, processing, refinement, and validation strategies. This research used resources of the Advanced Photon Source (APS), a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

Throughout the years, I had the chance to work with many people in the lab. Of all the people, I want to thank first Dr. Karen Kirby who is an excellent teacher and person. Karen taught me various crystallographic techniques and procedures; her help was immense especially during the beginning of my graduate studies in HIV-1 capsid-related research. Karen's patience is an incredible example for me. I also thank her for her generous help in organizing the lab functions in the best way possible. I was also fortunate to work with Dr. Elefterios Michailidis who helped me to get started with my project. I was lucky to share the office space with Dr. Zhe Li, who has been a great companion and immense support with crystallography-related projects.

I am very grateful to Dr. Dandan Liu, Tom Laughlin and Emily Conrod for their great help with cloning of various HIV-1 capsid mutants. I am extremely thankful to Grace Yang and students Rohit Rao, Emily Conrod, Catie Vornholt, Victoria Grill for their great help with expression and purification of HIV-1 capsid mutants.

I also want to thank all the former and current graduate students, Dr. Yee Tsuey Ong, Dr. Tanya Ndongwe, Andrew Huber, Jennifer Wolf, Mary Casey, Maritza Puray Chavez, Obi Ukah, Jackie Flores, Maria Mpoftsi, Seongmi Kim, Vincent Yapo and Mahmoud Farghali, as well as research faculty Dr. Philip Tedbury, Dr. Kamal Singh, Juan

Ji and Rama Koppisetti for working together, exchanging ideas and reagents, and for the great lab environment.

I want to thank my advisor Dr. Natalia Loginova, colleagues and friends from Belarus, Dr. Tatsiana Kavalchuk, Dr. Maryna Savko, Dr. Volha Vaitkevich, and Maryna Lipay, for their full support and kind words of advice. I also thank all the friends I have in the United States, who helped me during all these years to feel like home, and especially my international group of friends Dariya Tsyrenzhapova, Nana Naskidashvili, Iuliia Alieva, Salome Chitorelidze, Zivile Raskauskaite, Katalin Toth, Sera Chiuchiarelli, Fernanda Plucani Amaral, Kateryna Schroeder, Fridah Mubichi, Eleni Galata, Elizavet Levogianni, Yulia Innokentieva, Tatiana Pronina, Marina Materikina, Darima Butitova, Armen Kazarian, Andrick Payen, Veniamin Katykhov, Selim Sukhtaiev, Alex Magnier, David Copeland, and many more. I want to say many thanks to Jerry Brightwell from Chemistry department and to administrative personnel at LSC for their support and help.

I dedicated my dissertation thesis to my parents Tadevush and Natallia Hres. My father, Tadevush, passed away shortly before I had to move to the United States. However, he had always been very proud of me and my achievements. My mother, Natallia, has been so far away in Belarus but still, she is so close to every step in my life. Her immense support is beyond words in happy and arduous times. I am also grateful to all other members (aunts, uncles, cousins, nephews, and nieces) of our big family for their continuous support and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
ARV	Antiretroviral
BD	Benzodiazepine
BM	Benzimidazole
BVM	Bevirimat
CA	Capsid Protein
C-A1	Coumercyn A1
CMOS	Complementary Metal Oxide Semiconductor
CPSF6 or CF-Im-68	Cleavage and Polyadenylation Specificity Factor 6
Cryo-EM	Cryo-Electron Microscopy
CypA	Cyclophilin A
DNA	Deoxyribonucleic Acid
cDNA	Complementary DNA
FDA	Food and Drug Administration
FG	Phenylalanine/Glycine
Gag	Group-specific Antigen
Group M	Major Group
Group N	New Group
Group O	Outlier Group
Group P	Pending the Identification of Further Human Cases
SU gp120	Surface Glycoprotein

TM gp41	Transmembrane Glycoprotein
HAART	Highly Active Antiretroviral Therapy
HIV-1	Human Immunodeficiency Virus Type 1
HIV-2	Human Immunodeficiency Virus Type 2
ICAM-1	Intercellular Adhesion Molecule 1
IN	Integrase
InSTIs	Integrase Strand Transfer Inhibitors
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
LRAs	Latency-Reversing Agents
MA	Matrix
MW	Molecular Weight
Mx2 or MxB	Human Myxovirus Resistance 2
bNAbs	Broadly Neutralizing Antibodies
NC	Nucleocapsid
Nef	Negative Factor
NFAT	Nuclear Factor of Activated T-cells
NIH	National Institutes of Health
NMR	Nuclear Magnetic Resonance
NNRTIs	Non-Nucleoside Reverse Transcriptase Inhibitors
NPCs	Nuclear Pore Complexes
NRTIs	Nucleoside/Nucleotide Reverse Transcriptase Inhibitors
dNTP	Deoxynucleoside Triphosphate

NLS	Nuclear Localization Signal
Nup153	Nucleoporin 153
Nup358/RanBP2	Nucleoporin 358
OD 600nm	Optical Density at 600 nm
orf	Open Reading Frame
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PIC	Pre-Integration Complex
PR	Protease
PIs	Protease Inhibitors
PrEP	Pre-Exposure Prophylaxis
PTEFb	Positive Transcription Elongation Factor-b
RNA	Ribonucleic Acid
gRNA	genomic RNA
RT	Reverse Transcriptase
RTC	Reverse Transcription Complex
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SRP	Signal Recognition Particle
TEM	Transmission Electron Microscopy
TMV	Tobacco Mosaic Virus
TNPO3 or TRN-SR2	Transportin 3
TRIM5α	Tripartite Motif-containing Protein 5α

TRIMCyp	TRIM5 α and CypA fusion protein
Vif	Virion Infectivity Factor
Vpr	Viral Protein R
WT	Wild Type

LIST OF PUBLICATIONS

1. **Gres AT**, Kirby KA, KewalRamani V, Tanner JJ, Pornillos O, Sarafianos SG. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349 (6243)**, 99-103 (2015).

2. **Gres AT**, Yang Q, Kirby KA, Sarafianos SG. Characterization of HIV-1 capsid interactions with host cell factors. *To be submitted in Acta Crystallographica Section D,* 2017. *Manuscript in preparation.*

3. **Gres AT**, Liu D, Kirby KA, Yang Q, Shi J, Aiken C, Fu X, Peijun Z, Francis A, Melikyan GB, Sarafianos SG. X-ray structures of P38A, P38A/T216I, E45A and E45A/R132T reveal the structural mechanism of HIV-1 capsid stability. *To be submitted in Nature Communications,* 2017. *Manuscript in preparation.*

4. Novikova M, Balasubramaniam M, Kudchodkar S, Soheilian F, **Gres AT**, Yang Q, Kirby KA, Fontana J, Steven AC, Sarafianos SG, Freed EO. Identification of a novel element in HIV-1 CA critical for assembly and maturation. *To be submitted in Journal of Virology,* 2017. *Manuscript in preparation.*

5. Craveur P, **Gres AT**, Liu D, Yang Q, Kirby KA, Sarafianos SG, Olson AJ. Novel CA_{NTD}-CA_{NTD} interaction essential for HIV-1 mature capsid stability: tracking the Achilles heel of the capsid core in the pentamer interfaces. *To be submitted in Nature Communications,* 2017. *Manuscript in preparation.*

LIST OF PRESENTATIONS

* Presenting co-author is underlined

1. **Gres AT***, Liu D, Kirby KA, Perilla JR, Schulten K, Shi J, Aiken C, Tanner JJ, Fu X, Zhang P, Sarafianos SG. May 2016. Crystal structures of E45A, E45A/R132T, P38A, and P38A/T216I HIV-1 capsid proteins show how mutations alter capsid stability. *National Meeting on Retroviruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. (Oral Presentation)

2. **Gres AT**, Kirby KA, Liu D, Laughlin T, Rao R, Coonrod E, KewalRamani V, Tanner JJ, Novikova M, Urano E, Freed EO, Pornillos O, Sarafianos SG. Dec 2015. Crystal structures of native and mutant HIV-1 capsid proteins reveal molecular details of interactions with ligands and structural basis of capsid stability. *Meeting of HIV Interaction and Viral Evolution Center,* La Jolla, CA, USA. (Oral Presentations)

3. **Gres AT**, Kirby KA, Liu D, Puray-Chavez M, KewalRamani V, Tanner JJ, Sarafianos SG. May 2015. Crystal structures of the native full-length HIV-1 capsid protein with CPSF6 and PF74 reveal changes at the inter-hexamer interfaces. *National Meeting on Retroviruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. (Oral Presentation)

4. **Gres AT**, Liu D, Kirby KA, Yang Q, Perilla JR, Schulten K, Shi J, Aiken C, Tanner JJ, Fu X, Zhang P, Sarafianos SG. Jun 2016. Crystal structures of P38A, P38A/T216I, E45A and E45A/R132T HIV-1 capsid proteins highlight the plasticity of HIV-1 capsid. *Structural Biology Related to HIV/AIDS*, National Institutes of Health, Bethesda, MD, USA. (Poster Presentation)

5. Novikova M, Balasubramaniam M, Kudchodkar S, Soheilian F, **Gres AT**, Kirby KA, Fontana J, Steven AC, Sarafianos SG, Freed EO. Jun 2016. Identification of a novel element in HIV-1 CA critical for assembly and maturation. *Structural Biology Related to HIV/AIDS*, National Institutes of Health, Bethesda, MD, USA. (Poster Presentation)

6. Xue G, Goh SL, Yu HJ, **Gres A**, Lee K, Sarafianos SG, Luban J, KewalRamani VN. Jun 2016. HIV-1 interaction with CypA regulates the use of FG-nucleoporins for nuclear entry. *Structural Biology Related to HIV/AIDS*, National Institutes of Health, Bethesda, MD, USA. (Poster Presentation)

7. Novikova M, Balasubramaniam M, Kudchodkar S, Soheilian F, **Gres AT**, Kirby KA, Fontana J, Steven AC, Sarafianos SG, Freed EO. May 2016. Identification of a novel element in HIV-1 CA critical for assembly and maturation. *National Meeting on Retroviruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. (Oral Presentation)

8. Xue G, Goh SL, Yu HJ, **Gres A**, Lee K, Sarafianos SG, Luban J, KewalRamani VN. May 2016. HIV-1 interaction with CypA regulates the use of FG-nucleoporins for

nuclear entry. *National Meeting on Retroviruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. (Poster Presentation)

9. Novikova M, Balasubramaniam M, Kudchodkar S, Soheilian F, **Gres AT**, Kirby KA, Sarafianos SG, Freed EO. Feb 2016. The PPIP122-125 Motif in HIV-1 CA is an essential assembly and maturation element. *Conference on Retroviruses and Opportunistic Infections (CROI)*, Boston, MA, USA. (Oral Presentation)

10. **Gres AT**, Kirby KA, Liu D, Laughlin T, Rao R, Coonrod E, KewalRamani V, Tanner JJ, Novikova M, Urano E, Freed EO, Pornillos O, Sarafianos SG. Dec 2015. Crystal structures of native and mutant HIV-1 capsid proteins reveal molecular details of interactions with ligands and structural basis of capsid stability. *Meeting of HIV Interaction and Viral Evolution Center,* La Jolla, CA, USA. (Poster Presentations)

11. **Gres AT**, Kirby KA, KewalRamani V, Tanner JJ, Pornillos O, Sarafianos SG. Jul 2015. Crystal structures of native and mutant HIV-1 capsid proteins reveal molecular details of interactions with ligands and structural basis of capsid stability. *Annual Meeting of American Crystallographic Association*, Philadelphia, PA, USA. (Poster Presentation)

12. **Gres AT**, Kirby KA, Liu D, Laughlin TG, KewalRamani V, Tanner JJ, Pornillos O, Sarafianos SG. Jun 2015. Crystal structures of native and mutant HIV-1 capsid proteins reveal molecular details of interactions with ligands and structural basis of capsid stability. *Structural Biology Related to HIV/AIDS*, National Institutes of Health, Bethesda, MD, USA. (Oral Presentation)

13. **Gres AT**, Kirby KA, KewalRamani V, Tanner JJ, Pornillos O, Sarafianos SG. Apr 2015. Crystal structures of the native full-length HIV-1 capsid protein reveal conformational variability. *Missouri Life Sciences Week,* CS Bond Life Sciences Center, Columbia, MO, USA. (Poster Presentation)

14. **Gres AT**, Kirby KA, Hachiya A, Michailidis E, Pornillos O, Sugiura W, Lee K, KewalRamani V, Tanner JJ, Sarafianos SG. Aug 2014. Native hexameric full-length HIV-1 capsid: crystal structure and drug targeting. *23 Congress and General Assembly of the International Union of Crystallography*, Montreal, Canada. (Oral Presentation)

15. **Gres AT**, Kirby KA, Hachiya A, Michailidis E, Pornillos O, Sugiura W, Tanner JJ, Sarafianos SG. Jun 2014. Crystal structure of the native hexameric full-length HIV-1 capsid. *Structural Biology Related to HIV/AIDS*, Bethesda, MD, USA. (Poster Presentation)

16. **Gres AT**, Kirby KA, Hachiya A, Michailidis E, Pornillos O, Sugiura W, Tanner JJ, Sarafianos SG. May 2014. Crystal structure of the native hexameric full-length HIV-1 capsid. *National Meeting on Retroviruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. (Oral Presentation)

FOOTNOTES

The content of the dissertation by Anna Gres is based on one published research paper and four manuscripts in preparation. The contribution of co-authors are as follows:

Chapter 2 – X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability

Chapter 2 is based on a research paper published in *Science* in 2015. The authors and title are as follows: **Gres AT**, Kirby KA, KewalRamani V, Tanner JJ, Pornillos O, Sarafianos SG. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349 (6243)**, 99-103 (2015).

ATG performed protein expression, purification, crystallographic experiments and was responsible for writing the manuscript. KAK helped initially with cryoprotection, data collection, processing, initial stages of refinement, model building and was heavily involved in manuscript editing. KWR, JJT, OW significantly contributed to the discussion and manuscript editing. SGS oversaw all aspects of the project and edited the manuscript.

Chapter 3 – Characterization of HIV-1 capsid interactions with host cell factors and pharmacological ligands

Chapter 3 is based on a manuscript that is in preparation (to be submitted in *Acta Crystallographica Section D*). The authors and title are as follows: **Gres AT**, Yang Q, Kirby KA, Sarafianos SG. Characterization of HIV-1 capsid interactions with host cell factors. (2017)

ATG performed the experiments and was responsible for writing the manuscript. QY performed protein expression and purification. KAK was involved in manuscript editing. SGS oversaw all aspects of the project and edited the manuscript.

Chapter 4 – X-ray crystal structures of P38A, P38A/T216I, E45A and E45A/R132T reveal structural mechanism of HIV-1 capsid stability

Chapter 4 is based on a research paper currently in preparation (to be submitted in *Nature Communications*). The authors and title are as follows: **Gres AT**, Liu D, Kirby KA, Yang Q, Shi J, Aiken C, Fu X, Peijun Z, Francis A, Melikyan GB, Sarafianos SG. X-ray crystal structures of P38A, P38A/T216I, E45A and E45A/R132T reveal the structural mechanism of HIV-1 capsid stability. (2017)

ATG performed crystallization experiments, collected diffraction data and performed the complete analysis the data. DL performed cloning experiments. QY performed the majority of the protein expression and purification. JS and CA shared mutant viruses. XF and PZ performed assembly assays and described their results. ACF and GBM performed CypA-DsRed loss assay and described their results. KAK and SGS supervised the project. ATG primarily wrote the paper. All authors discussed the results and implications and commented on the manuscript.

Chapter 5 – Identification of a novel element in HIV-1 CA critical for assembly and maturation

Chapter 5 is based on a manuscript that is in preparation (to be submitted in *Journal of Virology*). The authors and title are as follows: Novikova M, Balasubramaniam M,

Kudchodkar S, Soheilian F, **Gres AT**, Yang Q, Kirby KA, Fontana J, Steven AC, Sarafianos SG, Freed EO. Identification of a novel element in HIV-1 CA critical for assembly and maturation. (2017)

ATG performed crystallization experiments, collected diffraction data and performed the complete analysis the crystallographic data. QY performed protein expression and purification. KAK and SGS supervised the crystallographic aspect of the project. MN, MB, SK, FS performed cell-based and electron microscopy assays. JF, ACS, EOF supervised the cell-based side of the project. MN discussed the results of virologic experiments. ATG wrote results and discussions of the crystallographic experiments. All authors discussed the results and implications and commented on the manuscript.

Chapter 6 – Novel CA_{NTD}-CA_{NTD} interaction essential for HIV-1 mature capsid stability: tracking the Achilles heel of the capsid core in the pentamer interfaces

Chapter 6 is based on a manuscript that is in preparation (to be submitted in *Nature Communications*). The authors and title are as follows: Craveur P, **Gres AT**, Liu D, Yang Q, Kirby KA, Sarafianos SG, Olson AJ. Novel CA_{NTD}-CA_{NTD} interaction essential for HIV-1 mature capsid stability: tracking the Achilles heel of the capsid core in the pentamer interfaces. (2017)

PC and ATG wrote the manuscript. PC performed molecular modeling and computational analysis. ATG performed transmission electron microscopy, crystallization experiments, collected diffraction data and performed the complete analysis the data. DL performed cloning experiments. QY performed protein expression and purification. AJO and SGS supervised the project. KAK was involved in manuscript editing.

STRUCTURAL BASIS OF STABILITY OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) CAPSID

Anna Gres

Dr. John J. Tanner, Dissertation Supervisor

Dr. Stefan G. Sarafianos, Dissertation Supervisor

ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of acquired immunodeficiency syndrome (AIDS). Since its discovery in early 1980, many advances have been made in the prevention and management of HIV/AIDS. One of the most important ones has been antiretroviral therapies (ART). Currently, there are more than 30 drugs approved for the treatment of HIV-1 infection. Used as combinations, they efficiently suppress viral loads and reduce AIDS-related deaths. Despite the advances, there is no cure and infection can eventually lead to fatal results. Hence, there is a need for new antivirals with novel mechanisms of action, favorable resistance and low toxicity profiles, which will offer more therapeutic options in the clinic.

The HIV-1 capsid protein (CA) has been increasingly viewed as an attractive therapeutic target as it plays a critical role in multiple steps of the virus life cycle, including uncoating, reverse transcription, nuclear entry, integration site selection, and assembly. Moreover, it is also necessary for shielding the deoxyribonucleic acid (DNA) product of reverse transcription from immune surveillance of the target cell. HIV-1 CA interacts with host factors including Cyclophilin A (CypA), cleavage and polyadenylation specific factor

6 (CPSF6), and nucleoporin Nup153. It is synthesized as a central part of the Gag polyprotein, which is cleaved during maturation to give matrix (MA), CA, nucleocapsid (NC), and several other peptides. Of ~5,000 Gag molecules in the immature HIV-1 virion, only ~1,500 CA molecules assemble into the mature capsid core that encloses the viral ribonucleic acid (RNA) and enzymes. The core comprises ~250 hexamers and 12 pentamers that allow the formation of the closed shape. Notably, all these diverse functions and interactions are regulated by the structure of a single, remarkably flexible protein, CA, which has been the main focus of this thesis.

Over the past 25 years, several X-ray, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM) structures of HIV-1 CA fragments or engineered variants have been solved. However, none of these provide the complete set of molecular details of the critical CA-CA contacts that govern capsid stability, which is at the heart of its biological functions.

In this study, we aimed to solve crystal structures of HIV-1 CA, including the elusive structure of the native full-length HIV-1 CA, in the space group P6, which allows the building of the hexameric biological unit. Our findings describe novel interactions between CA monomers related by 6-fold symmetry within a hexamer (intra-hexamer) and by 3-fold and 2-fold symmetry between neighboring hexamers (inter-hexamer). These structures help elucidate how CA builds a hexagonal lattice, the foundation of the mature capsid. Moreover, they demonstrate that the intra- and inter-hexamer interfaces are malleable and can change through an adaptable hydration layer. Disruption of this layer by crystal dehydration treatment alters inter-hexamer interfaces and condenses CA packing. The structures reveal a remarkable plasticity which explains the polymorphism observed

in virions. They also establish our experimental system to be the most relevant for the study of CA interactions in a native context.

We have used this system to obtain crystal structures of CA in complex with either a CA-binding antiviral (PF74) or a host factor peptide (CPSF6 or Nup153). Previous structures of CA fragments or engineered CA hexamer in complex with PF74, CPSF6 or Nup153 have provided valuable information on CA-host factor interactions, but they lack information on inter-hexamer interfaces. Hence, despite extensive work the details of CA interactions with host factors and pharmacological ligands that regulate the HIV life cycle as well as the structural basis of CA stability that determines virus infectivity are not well understood. Our findings reveal novel information about the changes at the hexamer-hexamer interfaces, thus providing structural insights into how those ligands can affect uncoating and assembly.

An enormous research effort has been invested over the years to determine the phenotypes of natural and artificial mutations in HIV proteins, including CA. Those studies have been significant for defining various CA functions and mapping them to the N-terminal (CA_{NTD}) and C-terminal (CA_{CTD}) domains. Additionally, several mutations have been identified to be useful in studies aimed at dissecting how CA performs its essential nonstructural functions. However, there is currently no structural information on any biologically relevant CA mutants (not counting the cross-linked CA constructs). Such information will reveal mechanisms of stabilization of the capsid core and may be applicable to other pathogenic viruses.

Thus, using our experimental setup, we have studied three groups of CA mutants. Initially, we focused on mutations in CA altering core stability and impairing viral

infectivity that suggests that core of optimal stability and its proper uncoating are critical for productive infection. P38A and E45A CA mutations have been reported to destabilize or hyperstabilize the capsid, resulting in non-infectious virus. Compensatory mutations T216I and R132T partially rescue infection from the defects associated with P38A and E45A, respectively. We performed crystallographic analysis of CA proteins bearing P38A, P38A/T216I, E45A, and E45A/R132T mutations, evaluated their assembly competence *in vitro*, and estimated core stability and uncoating. Mutant phenotypes are not a consequence of the main structural rearrangements in CA. Structural analysis suggests mutual electrostatic repulsion between pairs of glutamic (E45) and aspartic (D51) acids forced into proximity by the CA structure provides an environmentally-sensitive switch which can control the state of assembly and disassembly of the capsid. The mutant structures highlight additional rearrangements that may affect host factor recognition and trafficking of small-molecules across the capsid shell.

Furthermore, early studies determined that mutations introduced into the loop between CA helices H6 and H7 (residues 122-125, Pro-Pro-Ile-Pro, or PPIP motif) to be lethal or caused decrease in infectivity; however, they were not studied in detail. To address the role of the PPIP motif in the mature capsid, we have crystallized CA proteins bearing P122A, I124A, T58A/I124A, T58S/T107I/P122A, V11I/T58A/P122A, and V11I/T58A/I124A mutations. The structures revealed subtle structural rearrangements not only at the sites of mutations but also at the inter- and intra-hexamer interfaces, as well as at the host factor binding sites. This implies that mutations may alter the stability of the mature capsid core and/or host factor binding and recognition. Hence, the H6-H7 loop of the

HIV-1 CA is a new structural element essential for inter-hexamer contacts in the immature Gag and mature capsid lattices.

Finally, statistical analysis of the intermolecular interactions between two monomers either within a hexamer or pentamer in the published capsid models revealed the hydrogen bond (H-bond) between the glutamic acid at position 28 and lysine in position 30 (E28~K30) being more abandoned at the intra-pentamer, rather than intra-hexamer interfaces. To investigate if it may contribute to counterbalancing the electrostatic destabilization observed in the pentamers, we tested purified mutant CA proteins, harboring R18A, E28A and R18A/E28A mutations, for the cylinder formation *in vitro* and solved their crystal structures. The results indirectly support the presence of E28~K30 H-bond predominantly in the pentamers. This interaction may regulate pentamer stability and may be essential for the proper CA assembly, stability, and uncoating of the HIV-1 capsid core.

Collectively, the structures highlight that CA plasticity is a key factor for its stability and how challenging it is to fully understand the effect of even a single mutation on this highly flexible protein. Our studies unravel the structural basis of core stability, which affects multiple steps in the virus life cycle. Moreover, they provide unique information on how CA structure controls interactions with host factors and small molecule CA-targeting antivirals.

Chapter 1. General introduction of HIV

1.1. HIV/AIDS

Acquired immune deficiency syndrome (AIDS) first came to public attention in 1981 and represented an acquired immunodeficiency, which invariably led to the demise of the infected individual. Thus, whatever the transmissible agent, the immune system could not eliminate the infection. The human immunodeficiency virus (HIV) was identified as the causative agent two years later. The provirus was not sequenced until 1985. Since that time, AIDS has claimed more than 25 million lives. Although the annual number of AIDS-related deaths has dropped, the epidemic continues to have a substantial effect on certain countries and high-risk groups (*1*).

Global HIV incidence reached its peak in 1997, at 3.3 million new infections (*2*). Annual incidence has stayed relatively constant at about 2.6 million per year since 2005, after a period of fast decline between 1997 and 2005 (*2*). The number of people living with HIV/AIDS has been steadily increasing and reached 38.8 million in 2015. At the same time, HIV/AIDS mortality has been declining at a steady pace, from a peak of 1.8 million deaths in 2005 to 1.2 million deaths in 2015 (*2*). There is substantial heterogeneity in the levels and trends of HIV/AIDS across countries. Although many countries have experienced decreases in HIV/AIDS mortality and annual new infections, other countries have had slowdowns or increases in rates of change in annual new infections (*2*).

In 2015, 1.8 million (75.4 %) new HIV infections were in sub-Saharan Africa, with large proportions in western, southern, and eastern sub-Saharan Africa (Figure 1-1, A). South Asia accounted for 8.5 % of new infections per year (*2*), southeast Asia for 4.7 %

and East Asia for 2.3 % (Figure 1-1, A). The highest rates of infection were in southern Africa, with more than 1 % of the population per year becoming infected in Botswana, Lesotho, and Swaziland (Figure 1-1, B). The highest estimated incidence rates in Europe were recorded in Russia, and in Asia were recorded in Cambodia (Figure 1-1, B). In the Americas, only Belize, Guyana, and Haiti had rates of more than 50 per 100,000 people (Figure 1-1, B).

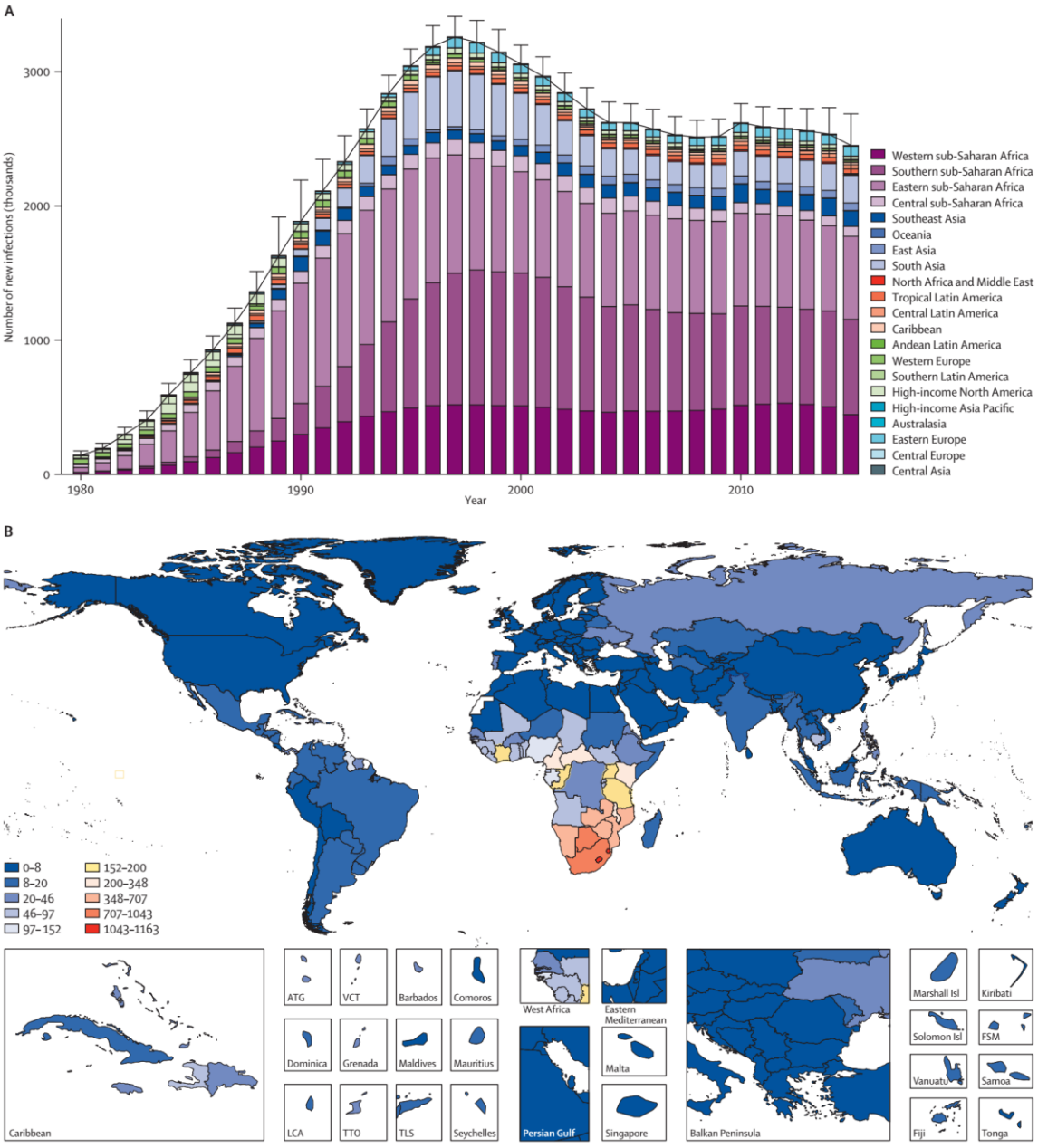


Figure 1-1. The incidence of new HIV infections from 1980 to 2015, and HIV incidence in 2015.

(**A**) A global number of new HIV infections by region. Bars show the mean number of estimated new infections within a given year. (**B**) A separate color represents regions. HIV incidence by country. Incidence was calculated as cumulative new cases of HIV throughout the year divided by the total population at the mid-year. Rates are per 100,000 people. This figure was copied from (*2*) (License CC-BY).

1.2. HIV genus and species

HIV is a part of the *Retroviridae* family (*3*), which consists of a vast number of diverse viruses (Table 1-1). The hallmark of the family is its unique replicative strategy. Upon entry into host cells, the viral genomic RNA (gRNA) is reverse transcribed into complementary DNA (cDNA) and subsequently integrated into the host chromosome. This integrated viral DNA, named as a provirus, functions as the template for viral genes transcription and translation, resulting in the generation of progeny virions. HIV is a member of the lentivirus genus that is characterized by a long latency period, progressive infection in which the virus evades the immune response of the host (*4*), and the unique ability to replicate in non-dividing cells.

The HIV epidemic arose after zoonotic infections with simian immunodeficiency viruses from African primates; bushmeat hunters were probably the first group to be infected with HIV (*5*). HIV-1 was transmitted from apes and HIV-2 from sooty mangabey monkeys.

Table 1-1. Retrovirus taxonomy.

Family	Subfamily	Genus	Species
Retroviridae	Orthoretrovirinae	Alpharetrovirus	Avian carcinoma Mill Hill virus 2
			Avian leukosis virus
			Avian myeloblastosis virus
			Avian myelocytomatosis virus 29
			Avian sarcoma virus CT10
			Fujinami sarcoma virus
			Rous sarcoma virus
			UR2 sarcoma virus
			Y73 sarcoma virus
		Betaretrovirus	Jaagsiekte sheep retrovirus
			Langur virus
			Mason-Pfizer monkey virus
			Mouse mammary tumor virus
			Squirrel monkey retrovirus
		Deltaretrovirus	Bovine leukemia virus
			Primate T-lymphotropic virus 1
			Primate T-lymphotropic virus 2
			Primate T-lymphotropic virus 3
		Epsilonretrovirus	Walleye dermal sarcoma virus
			Walleye epidermal hyperplasia virus 1
			Walleye epidermal hyperplasia virus 2
		Gammaretrovirus	Chick syncytial virus
			Feline leukemia virus
			Finkel-Biskis-Jinkins murine sarcoma virus
			Gardner-Arnstein feline sarcoma virus
			Gibbon ape leukemia virus
			Guinea pig type-C oncovirus
			Hardy-Zuckerman feline sarcoma virus
			Harvey murine sarcoma virus
			Kirsten murine sarcoma virus
			Moloney murine sarcoma virus
			Murine leukemia virus
			Porcine type-C oncovirus
			Reticuloendotheliosis virus
			Snyder-Theilen feline sarcoma virus
			Trager duck spleen necrosis virus
			Viper retrovirus
			Woolly monkey sarcoma virus
		Lentivirus	Bovine immunodeficiency virus
			Caprine arthritis encephalitis virus
			Equine infectious anemia virus
			Feline immunodeficiency virus
			Human immunodeficiency virus 1
			Human immunodeficiency virus 2
			Puma lentivirus
			Simian immunodeficiency virus
			Visna/maedi virus

Family	Subfamily	Genus	Species
	Spumaretrovirinae	Spumavirus	African green monkey simian foamy virus
			Bovine foamy virus
			Equine foamy virus
			Feline foamy virus
			Macaque simian foamy virus
			Simian foamy virus

Four groups of HIV-1 exist and represent three separate transmission events from chimpanzees (M, N, and O), and one from gorillas (P). Groups N, O, and P are restricted to West Africa. Group M, which is the cause of the global HIV pandemic, started about 100 years ago. It consists of nine subtypes (or clades): A–D, F–H, J, and K (Figure 1-2). Subtype C predominated in Africa and India and accounted for 48 % of cases of HIV-1 in 2007 worldwide (*6*). Subtype B predominates in Western Europe, the Americas, and Australia. Circulating recombinant subtypes (CRFs) are becoming more common. In addition to CRFs, unique recombinant forms (URFs) have been detected in individuals. Numerous CRFs, URFs, and inter-subtype recombinant viruses are acknowledged, although classifications and definitions of HIV strain often change.

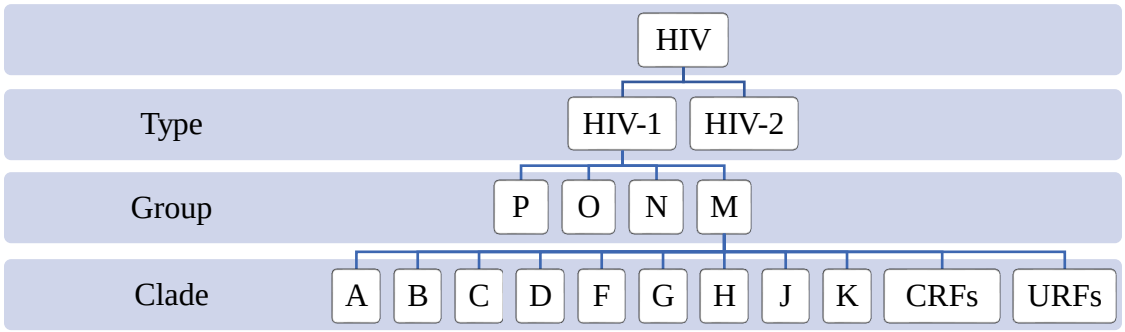


Figure 1-2. HIV types and subtypes.

The marked genetic diversity of HIV-1 is a consequence of the error-prone function of reverse transcriptase (RT), rapid turnover rates, and the persistent nature of the virus, which results in a high mutation rate (*7*). The high rate of HIV-1 evolution enables the virus to elude the body's immune control and better resist drug therapy, which makes producing an effective vaccine problematic. HIV-2 is largely confined to West Africa and causes a similar illness to HIV-1, but immunodeficiency progresses more slowly, and HIV-2 is less transmissible (*5*).

1.3. HIV-1 genome organization

HIV-1 is an enveloped, single-stranded, positive-sense RNA virus with a diameter ~120 nm. The HIV-1 genome is shown schematically in Figure 1-3. It contains three open reading frames (orfs): *gag* – the orf that encodes precursor 55 kDa Gag (group specific antigen) of the major structural proteins of the virus interior MA, CA, NC and p6; *pol* – the orf that encodes enzymatic activities of the virus PR, RT, RNAse H and IN; and *env* – the frame that encodes precursor of the viral envelope proteins gp41 and gp120, which are responsible for receptor binding. HIV-1 is a complex retrovirus and, hence, encodes further six regulatory proteins that enhance and control the replication of the virus (Nef, Rev, Tat, Vif, Vpr, Vpu). The *pol* orf overlaps with that of *gag* and is expressed *via* frame-shifting event that produces Gag-Pol protein.

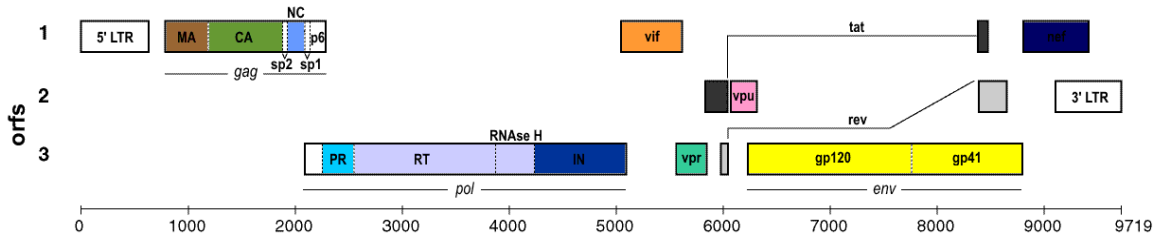


Figure 1-3. Map of the HIV-1 genome.

The HIV-1 RNA genome contains *gag, pol, env, tat, vif, vpu, vpr, rev,* and *nef* genes. Both ends of RNA are flanked by long terminal repeats (LTRs). Rectangles represent open reading frames. *Gag* codes for the Gag precursor polyprotein, which is processed by HIV-1 protease during maturation into the matrix (MA), capsid (CA), nucleocapsid (NC), and p6 proteins. *Pol* codes for protease (PR), reverse transcriptase (RT), and integrase (IN), which are produced as the Gag-Pol precursor polyprotein (a result of ribosome frameshift near the 3' end of *gag*). *Env* codes for the gp160 precursor polyprotein, which is cleaved into gp120 and gp41. Tat and Rev result from spliced exons, shown in black and gray rectangles, respectively. This figure was adapted from (*8*).

Table 1-2 provides a summary of the features of the 15 proteins produced by HIV-1 (*9, 10*).

Table 1-2. Functions of HIV-1 proteins.

Gene	Protein	Functions
gag	Matrix (MA), p17	Plasma membrane targeting of Gag and Gap-Pol for virion assembly; incorporation of envelope glycoproteins; post-entry events
	Capsid (CA), p24	Virion core structure and assembly
	Nucleocapsid (NC), p7	Virion packaging of genomic RNA; RNA chaperone; virion assembly; coats gRNA inside the core
	p6	Initiates incorporation of Vpr; promotes virion budding
pol	Protease (PR), p15	Proteolytic processing of Gag and Gag-Pol during maturation
	Reverse transcriptase (RT), p66/p51	cDNA synthesis; RNaseH domain degrades gRNA
	Integrase (IN), p31	Covalent insertion of viral cDNA into the host DNA
env	Surface glycoprotein (SU), gp120	Binds cell-surface CD4 receptors and CCR5 or CXCR4 co-receptors; mediates virus attachment and entry
	Transmembrane glycoprotein (TM), gp41	Contains fusion peptide; mediates membrane fusion upon virus entry
nef	Negative factor (Nef), p27	Downregulation of CD4 and major histocompatibility complex (MHC) class I molecules; T-cell activation; moderate enhancer of viral infectivity; blocks apoptosis; pathogenicity determinant
rev	Regulator of virion, Rev (p19)	Binds to Rev responsive element (RRE) and promotes nuclear export of viral mRNAs
tat	Trans-activator of transcription, Tat (p16/p14)	Binds to trans-activating response element (TAR) and activates transcription initiation and elongation from the long-terminal repeat (LTR) promoter
vif	Virion infectivity factor (Vif), p23	Suppresses APOBEC3G/APOBEC3F, host factors that deaminate DNA:RNA duplex and inhibit infection
vpr	Viral protein R (Vpr), p14	Moderate enhancer of post-entry infectivity; G2/M cell-cycle arrest
vpu	Viral protein U, Vpu (p16)	CD4 downregulation in the endoplasmic reticulum; induces virion release from host cell surface by counteracting action of host factor BST2/tetherin

1.4. HIV-1 virion

Like all viruses, HIV-1 comprises proteins that are unique to the virus and serve various functions regarding viral replication. A typical newly released HIV-1 particle (Figure 1-4) contains ~2,500 Gag molecules, arranged radially in a roughly spherical array with their N-termini facing outward and their C-termini inward (*11*).

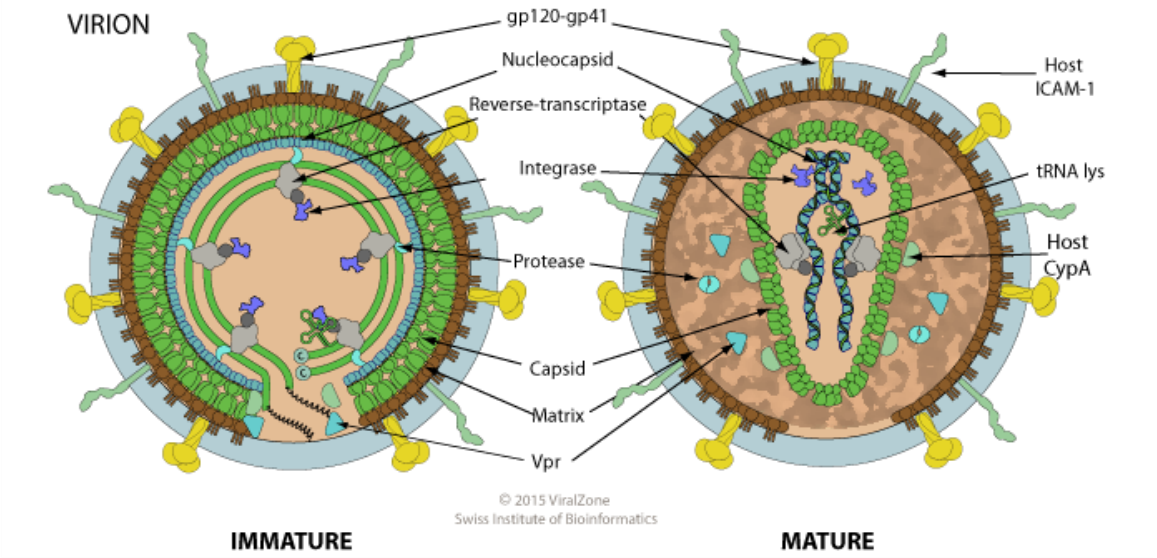


Figure 1-4. Organization of the immature and mature HIV-1 virions.

The HIV-1 virion contains a cell-derived lipid bilayer that incorporates envelope glycoproteins gp120 and gp41 trimers. The structural proteins of HIV-1 (matrix, capsid, nucleocapsid) are derived from the Gag polyprotein. Gag molecules assemble at the plasma membrane to form immature virions. The viral protease is activated during maturation, and it cleaves Gag to generate a set of new proteins, which then reassemble. In the mature virion, two single-stranded genomic RNAs coated with nucleocapsid are encapsidated in a fullerene-cone shell made up of capsid and surrounded by a sphere shell consisting of matrix. Reverse transcriptase and integrase are responsible for replication of the viral genome inside a host cell. Virions also contain other viral and host cell proteins. This figure was copied from http://viralzone.expasy.org/all_by_species/5182.html.

Particle interior is occupied by genomic RNA (gRNA) of the virus (a pair of identical RNA molecules joined in a dimeric form), along with ~20 molecules of tRNA (*12*) and other small cellular RNAs (*13, 14*), and the virus-coded enzymes (PR, RT, IN). The Gag molecules are associated with a lipid bilayer that is derived from the plasma membrane of the virus-producing cell. SU gp120 and TM gp41 form trimers that project through the bilayer (Figure 1-4, Table 1-3).

After the immature particle is released from the cell, the Gag polyprotein is cleaved by viral PR into a series of cleavage products. These cleavage events are termed "maturation" of the particle, and the products include MA, CA, NC, p6, and the spacer peptides sp1 and sp2. Maturation results in a reorganization of the internal structure of the

virion and is required for the infectivity of the particle. NC is complexed with RNA in the interior of the mature particle. Other viral (Vif, Vpr, Nef) and cellular proteins (including CypA, actin, cofilin) are present in smaller amounts.

Table 1-3. Virion composition.

Nucleic acids		
Viral RNA (genome)	2 copies	
tRNA Lys	20 copies (*12*): ~8 $tRNA_3^{lys}$, and 12 $tRNA_{1,2}^{lys}$	
Host signal recognition particle (SRP) RNA/7SL (*13, 14*)	?	
Spliced viral mRNA (*15-17*)	?	
Host U6 spliceosomal RNA, host polymerase III RNA (*15, 16, 18*)	?	
Surface proteins		
gp120	~12-30 (*19, 20*)	
gp41	~12-30 (*19, 20*)	
Host intercellular adhesion molecule 1 (ICAM-1) (*21*)	?	
Proteins inside the virion		
	Immature virion	**Mature virion**
Gag	~2,500 (1400 (*19*), 2,500 (*11*) to 5,000 (*22*))	-
MA	-	~2,500 (relative to Gag)
CA	-	one or two capsids composed of 1,000-1,500 CA proteins (*23*)
NC	-	~2,500 (relative to Gag)
p6	-	~2,500 (relative to Gag)
Gag-Pol	~120 (120 (*24*) to 200 (*25*))	-
PR	-	~120 (relative to Gag-Pol)
RT	-	~120 (relative to Gag-Pol)
IN	-	~120 (relative to Gag-Pol)
Vif	1-150 (*26, 27*)	1-150 (*26*)
Vpr	16-300 (14-18 (*28*), or up to 300 (*29*))	16-300 (14-18 (*28*), or up to 300 (*29*))
Nef	-	~5-10 (*30, 31*)
Host components		
Actin (*32*)	~250-350	~250-350
CypA (*33*)	~230	~230
Cofilin	~50-250	~50-250
Ezrin (*32*)	~50	~50
Moesin (*32*)	~50	~50

1.5. Life cycle and replication

The HIV replication cycle (Figure 1-5) is a complex process that includes:

- Attachment – virion attaches to the target host cell by interactions between SU gp120 and CD4 (the cell surface HIV-1 receptor) and CCR5 or CXCR4 co-receptors.
- Penetration – following binding, TM gp41 undergoes a conformational change that promotes virus-cell membrane fusion, thereby allowing entry of the core, which undergoes partial disassembly or uncoating.
- Reverse transcription – RT synthesizes double-stranded (ds) viral DNA from single-stranded viral RNA.
- Nuclear trafficking and entry – viral nucleoprotein complex is transported to the nucleus.
- Integration – IN catalyzes integration of the dsDNA into the host chromosome and the DNA is repaired.
- Transcription – viral transcripts are expressed from the promoter located in the 5' long terminal repeat (LTR), with Tat greatly enhancing the rate of transcription. A set of spliced and genomic-length RNAs are transported from the nucleus to the cytoplasm, where they can be translated or packaged. This is regulated by Rev.
- Translation – viral mRNAs are translated in the cytoplasm, and the Gag and Gag-Pol polyproteins become localized to the cell membrane. The *env* mRNA is translated at the endoplasmic reticulum (ER).
- Assembly – the virus particle is assembled from the Gag and Gag-Pol polyproteins, Vif, Vpr, Nef, and the gRNA, and an immature virion begins to bud from the cell

surface. To provide SU gp120 and TM gp41 proteins for the outer membrane coat during budding, the Env polyprotein (gp160) must first be released from complexes with CD4, which is coexpressed with gp160 in the ER. Vpu assists this process by promoting CD4 degradation. Gp160 is then transported to the cell surface, where again it must be prevented from binding CD4. Nef promotes endocytosis and degradation of surface CD4.

- Budding – new virus coated with SU gp120, and TM gp41 exits the host cell.
- Maturation – viral PR cuts Gag and Gag-Pol polyproteins into individual components. The mature virion is then ready to infect the next cell.

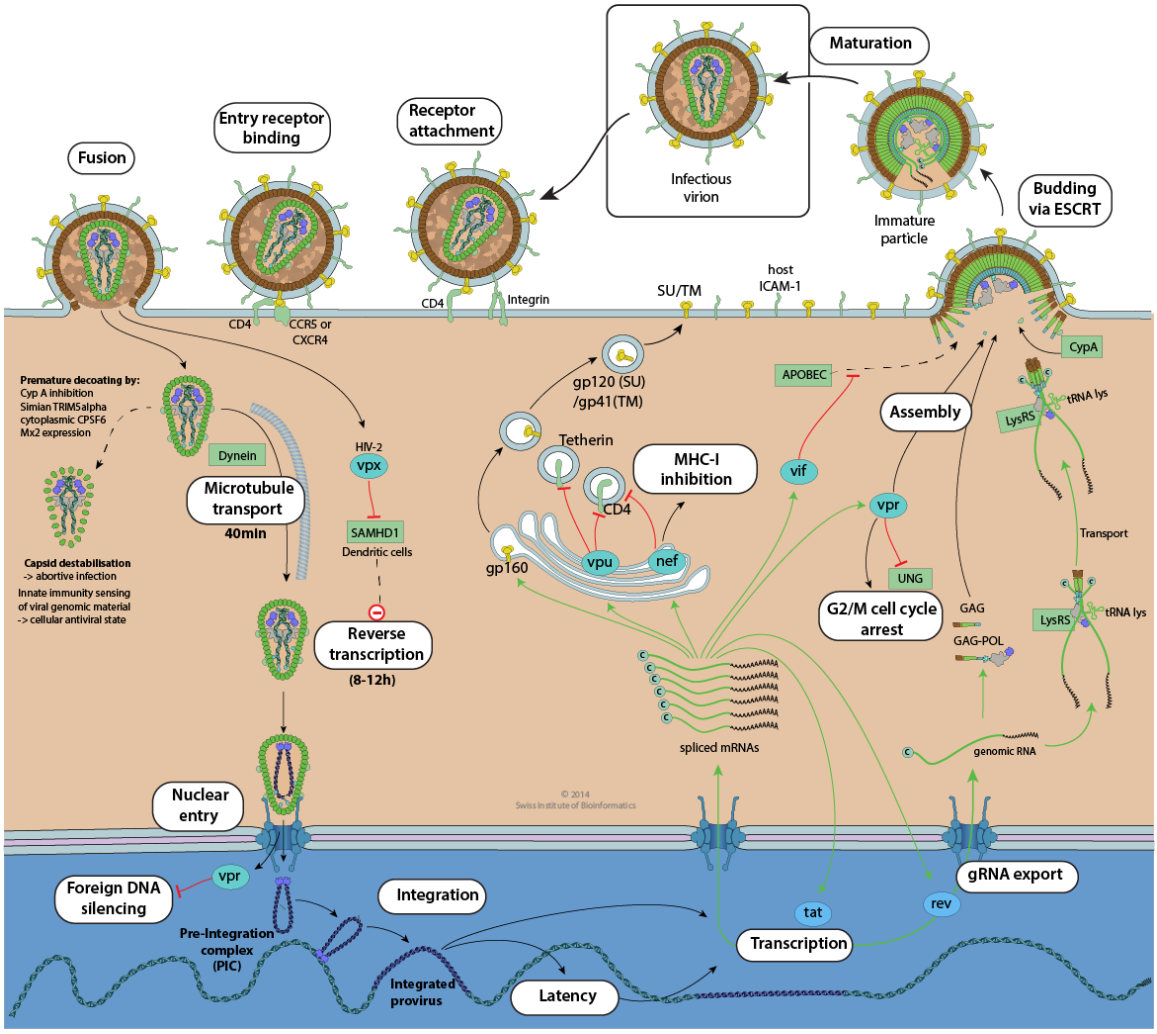


Figure 1-5. HIV replication cycle.

The HIV-1 life cycle in a host cell involves the following main stages. (1) Entry of HIV-1 involves binding of the virus to a cell through the interaction of gp120 with the CD4 receptor and CCR5 or CXCR4 co-receptor. This leads to fusion of the viral envelope with the cellular membrane, allowing the release of viral core into the cytoplasm. (2) As the viral core disassembles, reverse transcription takes place and results in the synthesis of double-stranded DNA. (3) The pre-integration complex forms and is transported into the nucleus for insertion of viral DNA into the host DNA. (4) HIV-1 RNAs are transcribed using the host transcription machinery. (5) RNAs are transported into the cytoplasm to generate viral proteins and precursor proteins. (6) Two copies of HIV-1 RNA and several viral proteins assemble near the cellular membrane, and a new virion is released by budding. (7) The new virion with a spherical core undergoes maturation *via* the proteolytic activity of HIV-1 PR, allowing CA proteins to rearrange and form the conical-shaped core. This figure was copied from http://viralzone.expasy.org/all_by_species/5096.html.

1.6. HIV transmission and pathogenesis

1.6.1. HIV tropism

HIV-1 can infect a number of cells, including T lymphocytes, monocytes/macrophages, dendritic cells, and, in the central nervous system, microglia. All these cell types express the CD4 glycoprotein (*34*), which serves as the receptor for HIV-1 and HIV-2. Efficient HIV entry into the target cells is mediated by sequential interactions between HIV envelope proteins (gp120 and gp41) with the CD4 receptor and a co-receptor, typically either β-chemokine receptor type 5 (CCR5), or α-chemokine receptor type 4 (CXCR4) (*35, 36*).

Understanding the viral property that we refer to as tropism has evolved. The earliest classification was based on the ability of the cultured virus to replicate efficiently in T-cell or macrophage cultures: the T-tropic/M-tropic designation. When the co-receptors were discovered, it became apparent that the primary basis for the tropism was the ability of the virus to enter cells using CXCR4, CCR5, or either as a coreceptor. Viruses that use CCR5 but not CXCR4 are R5; those that use CXCR4 but not CCR5 are X4; and those that can use either CCR5 or CXCR4 are referred to as R5X4 or dual (*37*).

Interestingly, co-receptor preference has been associated with other viral traits. Specifically, CXCR4-tropic viruses have been associated with increased host cell pathogenicity and more rapid progression of disease (*37*).

1.6.2. HIV transmission

HIV-1 is transmitted by sexual contact across mucosal surfaces (rectum, vagina, urethra, mouth, eyes), by maternal-infant exposure (pregnancy, delivery, breastfeeding),

and by percutaneous inoculation (injuries from contaminated sharps, sharing of needles or syringes, transfusions from infected donors). A sufficient quantity of viruses must be transferred to infect a person with HIV. The virus cannot live long outside of the human body; it cannot be transmitted *via* saliva, tears, sweat, feces, urine.

From 2005 to 2015, HIV newly infected globally 2.5-2.6 million people per year (*2*). The major route of transmission is sexual, which accounts for ~70 % of HIV infections (*38*). The remainder is largely attributed to percutaneous inoculation, including needle sharing, contaminated blood transfusion, and health-care-related accidents, as well as mother-to-child transmission that takes place at the placenta *via* contaminated maternal blood (*38*).

Notably, the efficiency of HIV-1 transmission can be modulated by other factors, including sexually transmitted diseases, pregnancy, and circumcision.

1.6.3. HIV infection

HIV-1 infection is usually initiated by a single virion infecting a single target cell at the portal of entry (*38*). A typical time course of infection can be divided into four stages (*39*):

1. Primary infection (~1-2 weeks). The virus freely replicates and spreads from the initial site of infection to other tissues and organs that become replication sites. Viremia is undetectable; there is no immune response or symptoms of infection.
2. Acute (or primary) infection phase (~2-4 weeks). It is characterized by relatively high levels of viremia (>10^7 copies of viral RNA per milliliter of blood), and large fractions of infected $CD4^+$ T-cells in blood and lymph nodes. This phase is often accompanied

by “flu-like” symptoms, including fever and enlarged lymph nodes. The immune response begins to appear, both in the form of antibodies against all viral proteins, and a $CD8^+$ T-cell response against HIV-1 antigens expressed on infected cells. As a result, large numbers of activated $CD4^+$ T-cells are generated, providing targets for viral replication. At the end of the acute phase, the level of viremia sharply declines due to partial control by the immune system and exhaustion of activated target cells.

3. Chronic infection (or “clinical latency”) (~1-20 years). This phase is characterized by a constant or slowly increasing level of viremia (1-100,000 copies/ml) and steady (~1,000 cells/ml), or gradually falling levels of $CD4^+$ T-cells. Usually, patients in this phase are asymptomatic and unaware of being infected. Large numbers of $CD4^+$ T-cells become infected and die every day.
4. Progression from HIV to AIDS. Finally, the number of$CD4^+$ T-cells declines to the point (~200 cells/ml) at which immune control of adventitious infectious agents can no longer be maintained, and opportunistic infections begin to appear. Control of the HIV-1 infection is lost, the level of viremia rises during the AIDS phase, culminating in the death of the infected patient. Untreated HIV-1 infection is one of the most uniformly lethal infectious diseases known, with a mortality rate over 95 %.

1.7. HIV treatment and prevention

1.7.1. Antiretroviral therapy

Highly active antiretroviral therapy (HAART) is the current treatment for HIV-1 infection. HAART involves taking a combination of HIV medicines (called an HIV regimen) every day. According to the World Health Organization 2016 guidelines,

HAART should be initiated in everyone living with HIV at any CD4 cell count. Additionally, the use of daily oral pre-exposure prophylaxis (PrEP) is recommended as a prevention choice for people at substantial risk of HIV infection as part of "a combination prevention approach".

Enormous progress has been made in reducing HIV deaths, especially in low-income countries, by expanding preventative measures for mother-to-child transmission and HAART programs. Enhancement in medication regimens has transformed the perception that HIV infection leads to a fatal disease, which was a near certainty before the implementation of HAART. Today, infected people instead have a long-term chronic condition. However, only 41 % of people living with HIV were receiving HAART in 2015 (*2*).

More than 30 antiretroviral (ARV) agents have been approved by the Food and Drug Administration (FDA) for HIV-1 treatment (Table 1-4). The majority of these agents are reverse transcriptase inhibitors – nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). Other classes of approved inhibitors include integrase strand transfer inhibitors (InSTIs), protease inhibitors (PIs), fusion inhibitors and entry inhibitors.

Despite the treatment success of ARV therapy, it is not a cure for HIV or AIDS. Patients must receive the treatments continuously for life due to HIV-1's ability to establish latency *via* integration in the host genome or face rapid rebound of the replicating virus if the treatment is interrupted. Long-term treatment ultimately can lead to the manifestation of drug-resistant viral variants. HAART decreases the probability of the emergence of drug-resistant HIV-1. However, the evolution of drug resistance can become problematic

when a virologic failure occurs due to noncompliance to the drug regimen, poor drug tolerability, and drug interactions among ARV agents (*40*).

Moreover, prolonged treatment may lead to diabetes, heart disease, liver disease, and various forms of cancer. Antiretroviral therapy is also expensive. Therefore, new classes of antiretroviral agents and new treatments targeting latent HIV-1 reservoirs are necessary.

Table 1-4. Drugs used in the HIV infection treatment.

Class	Brand	Generic Name	Approved
Multi-class combination products	Atripla	efavirenz, emtricitabine and tenofovir disoproxil fumarate	2006
	Complera	emtricitabine, rilpivirine, and tenofovir disoproxil fumarate	2011
	Evotaz	atazanavir sulfate, combicistat	2015
	Prezcobix	cobicistat, darunavir ethanolate	2015
	Stribild	elvitegravir, cobicistat, emtricitabine, tenofovir disoproxil fumarate	2012
Nucleoside reverse transcriptase inhibitors (NRTIs)	Combivir	lamivudine and zidovudine	1997
	Emtriva	emtricitabine, FTC	2003
	Epivir	lamivudine, 3TC	1995
	Epzicom	abacavir and lamivudine	2004
	*Hivid	zalcitabine, dideoxycytidine, ddC	1992
	Retrovir	zidovudine, azidothymidine, AZT, ZDV	1987
	Trizivir	abacavir, zidovudine, and lamivudine	2000
	Truvada	tenofovir disoproxil fumarate and emtricitabine	2004
	Videx EC	enteric coated didanosine, ddI EC	2000
	Videx	didanosine, dideoxyinosine, ddI	1991
	Viread	tenofovir disoproxil fumarate, TDF	2001
	Zerit	stavudine, d4T	1994
	Ziagen	abacavir sulfate, ABC	1998
Non-nucleoside reverse transcriptase inhibitors (NNRTIs)	Edurant	rilpivirine	2011
	Intelence	etravirine	2008
	Rescriptor	delavirdine, DLV	1997
	Sustiva	efavirenz, EFV	1998
	Viramune (Immediate Release)	nevirapine, NVP	1996
	Viramune XR (Extended Release)	nevirapine, NVP	2011

Class	Brand	Generic Name	Approved
Protease inhibitors (PIs)	*Agenerase	amprenavir, APV	1999
	Aptivus	tipranavir, TPV	2005
	Crixivan	indinavir, IDV	1996
	*Fortovase	saquinavir	1997
	Invirase	saquinavir mesylate, SQV	1995
	Kaletra	lopinavir and ritonavir, LPV/RTV	2000
	Lexiva	fosamprenavir Calcium, FOS-APV	2003
	Norvir	ritonavir, RTV	1996
	Prezista	darunavir	2006
	Reyataz	atazanavir sulfate, ATV	2003
	Viracept	nelfinavir mesylate, NFV	1997
Fusion inhibitors	Fuzeon	enfuvirtide, T-20	2003
Entry inhibitors – CCR5 co-receptor antagonist	Selzentry	maraviroc	2007
HIV integrase strand transfer inhibitors (InSTIs)	Isentress	raltegravir	2007
	Tivicay	dolutegravir	2013
	Vitekta	elvitegravir	2014

* no longer marketed
Information assessed on Jan 6, 2017, from FDA website:
http://www.fda.gov/ForPatients/Illness/HIVAIDS/Treatment/ucm118915.htm

1.7.2. Vaccines

The vaccine has proved to be a successful approach for preventing infectious diseases caused by some viruses including smallpox virus, poliovirus, human papilloma virus, and hepatitis B virus (*41*). Despite early optimism and enormous efforts in HIV-1 vaccine development, after more than three decades there are no licensed HIV preventative or therapeutic vaccines.

Effective vaccines stimulate protective immunity similar to that which occurs during natural infection. The ideal HIV-1 vaccine should elicit broad adaptive immune responses that are effective against all HIV-1 strains. Unprecedented challenges face the development of an effective HIV vaccine, including extreme genetic diversity of HIV, uncertainty about the mechanism by which an HIV vaccine might confer protection, and difficulty in the development of antigens that are highly immunogenic (*42*).

To date, among more than 218 HIV vaccine trials, only five vaccines have advanced to Phase IIb and III clinical trials, including the VAX003, VAX004, Step/Phambili, RV144, and HVTN505 trials (*43*). Those trials utilized the following therapeutic HIV vaccine approaches:

- inactivated whole virus depleted of gp120;
- single or multiple HIV antigens administered as DNA;
- autologous dendritic cells;
- viral vectors, e.g., poxviruses (canarypox ALVAC-HIV, vCP1452, vCP1433, fowlpox, MVA), adenoviruses (Ad5).

The VAX003 and VAX004 vaccines induced non-neutralizing antibodies and showed no significant reduction in the HIV-1 acquisition. The Phambili trials failed to show a protective effect. The Step data revealed that vaccine-recipients with pre-existing immunity to adenovirus vector serotype and/or who were uncircumcised had an increased risk of HIV-1 acquisition. The most recent vaccine regimen, HVTN505, failed to have any significant effect on HIV-1 acquisition or viral load. Only one candidate vaccine, RV144, has significantly reduced HIV-1 acquisition, even though the efficacy was limited (60 % at 12 months but 31 % at 42 months) (*42-45*).

Although multiple therapeutic vaccines have failed in the past, Tat vaccines and broadly neutralizing antibodies (bNAbs) are amongst the latest approaches for exploration. The HIV-1 transactivator of transcription (Tat) is an essential HIV virulence factor, which is crucial for virus gene expression, replication, transmission, and disease progression. Interestingly, anti-Tat antibodies are uncommon in natural infection and, when present, correlate with the asymptomatic state and lead to lower or no disease progression (*46*). In Phase I, preventative and therapeutic Tat vaccination has been shown to be safe and immunogenic. Results of the Phase II trials of Tat vaccine demonstrated immune cell restoration, reduction of immune activation, and reduced HIV-1 DNA viral load (*46*). Phase III trials are being designed to confirm the efficacy of the Tat vaccine.

Broadly neutralizing antibodies offer potentially not only prevention, but also therapy and cure. In early-phase clinical trials, VRC01 and another human monoclonal antibody, 3BNC117 (*47, 48*), reduced viral load in HIV-1-infected individuals not on HAART. However, monotherapy is likely to be insufficient to control infection, and antibody-drug or antibody-antibody combinations are expected to be required for complete viremic control.

Despite the lack of successful HIV vaccine, our understanding of viral immunology has progressed remarkably. Even if only partially effective, therapeutic vaccines may be valuable for HIV-infected individuals as a treatment intensification on the immune system and viral reservoir when combined with HAART. Vaccination at the time of primary infection or early in the course of infection has been proposed to limit reservoir establishment and promote viral eradication, analogous to the rationale for early HAART initiation.

1.7.3. Latency, reservoirs, and potential cure

Despite the great successes of HAART in the control of HIV replication and significant reduction of morbidity and mortality, it is unable to cure HIV, and lifelong treatment is needed. HIV can persist in patients on HAART because of long-lived resting memory T-cells (*49*), periodic homeostatic proliferation of cells that sustain the latent viral reservoir, residual viral replication, inadequate penetration of antiretroviral drugs into tissues where the virus replicates, and persistence of HIV in sanctuary sites (such as the gastrointestinal tract, lymphoid tissue, and the CNS) (*50*).

Latency is defined as the integration of HIV DNA into the host genome in the absence of virus production. Latency can be established *in vitro via* direct infection of resting CD4 T-cells in the presence of specific chemokines, or after reversion of an activated infected T-cell to a resting state. Latency has been shown *in vivo* in central and transitional memory T-cells (*49*) and naive T-cells. Latency can also be established in monocyte-macrophages and astrocytes (*51*), but the importance of these cells to virus persistence in patients on antiretroviral therapy is unclear.

The question of why HIV establishes a state of latent infection is of considerable interest. It is a reversible nonproductive state of infection of individual cells. For some viruses, *e.g.*, herpes family, it is an important mechanism for immune evasion. It is not so clear that latency serves this function for HIV. HIV replicates actively throughout the infection in untreated individuals. The rapid evolution of escape mutants provides the principal mechanism by which the virus avoids immune responses. It has been suggested that latency is a “hard-wired” feature of the regulation of HIV gene expression that evolved as a “bet-hedging strategy” to allow the virus to be successfully transmitted across mucosal

barriers, with subsequent reactivation once the initially infected cells reach a tissue site that is more favorable for viral replication (*52, 53*). If this hypothesis is correct, the form of latency involved is likely different from the one that allows persistence of the virus during HAART (*54*).

The simplest explanation for the existence of the latent reservoir is that HIV latency is a consequence of viral tropism for activated CD4 T-cells (*55*). T-cell activation results in a gradual upregulation of the CCR5 co-receptor, increase of deoxynucleoside triphosphate (dNTP) and release of sequestered forms of the host transcription factors NF-κB, nuclear factor of activated T-cells (NFAT), and positive transcription elongation factor-b (PTEFb) – all of which play a major role in HIV gene expression. Infection of activated CD4 T-cells results in rapid reverse transcription, integration, viral gene expression, and virus production, generally followed by the death of the cells, usually in one to two days. On the contrary, infection of resting CD4 T-cells is hindered by the absence of CCR5, low levels of dNTPs maintained by the dNTP triphosphohydrolase SAMHD1 (*56*), and by a cell death pathway which is triggered by innate immune recognition of reverse transcription intermediates (*57*). Infection of both activated and resting CD4 T-cells results in cell death. However, infection of activated $CD4^+$ T cells that are transitioning back to a resting state, while the cells are still permissive for reverse transcription and integration of the viral genome but not for high-level viral gene expression, may allow the establishment of a stable state of latent infection in resting $CD4^+$ T cells (*55*). This is a rare event, consistent with the extremely low frequency of latently infected cells *in vivo* (~$1:10^6$). Viewed in this light, HIV latency is an unfortunate accident of viral tropism. Whether this or other explanations for the origins of HIV latency

are correct, there is no doubt that latency is established in all infected individuals and that it serves as a barrier to HIV cure.

Much interest surrounds the discovery of either a functional cure (long-term control of HIV without antiretroviral therapy) or a sterilizing cure (complete elimination of all HIV-infected cells) (*51*). Hopes that a cure might be possible have been raised by a case report of a man who underwent stem cell transplants for leukemia (referred to as the Berlin patient), and an infant who started antiretroviral therapy soon after delivery (referred to as the Mississippi baby).

The best-documented report of cure is the Berlin patient, a man with HIV on antiretroviral therapy who had acute myeloid leukemia and received two bone marrow transplants from a donor with a homozygous defect in CCR5, a key co-receptor needed by HIV for cell entry. Shortly after transplantation, the patient ceased antiretroviral therapy and minimum, or no virus has been detected in plasma or tissue for more than six years (*58*). This case has inspired the development of gene therapy to eliminate CCR5 in patient-derived T-cells and stem cells with new technologies such as zinc finger nucleases that can reduce CCR5 expression (*59*).

The Mississippi baby was born to an HIV-infected mother and began HAART within 30 h of birth (*50*). The infant successfully showed suppression of plasma viremia to below the limit of detection and continued HAART for 18 months, at which point she was lost to follow-up and discontinued medications. Remarkably, she remained aviremic for more than two years without receiving HAART, raising hopes that she might have completely eradicated HIV. Unfortunately, her plasma viremia ultimately rebounded to above 10,000 copies/ml, and HAART was reinitiated. It is believed that the extremely early

initiation of HAART played a key role in the protracted period of aviremia in the absence of drugs by minimizing the size of the persistent HIV reservoir. It is also probable that the latent viral reservoir is less prone to spontaneous reactivation and/or the spread of infection is less efficient in the context of the immature immune system of an infant or child.

Furthermore, 1-15 % of HIV-infected individuals who began HAART during the acute-early phase of infection, subsequently controlled their plasma viremia for prolonged periods of time after discontinuation of therapy (*50*). Considering that plasma viremia typically rebounds within a few weeks after the cessation of HAART in the majority of HIV-infected individuals who initiate treatment during the chronic phase of infection, this finding highlights potential benefits of, and reasons for, promoting the early initiation of HAART.

The most widely discussed approach for eliminating the reservoir is the "shock and kill" approach, in which small-molecule latency-reversing agents (LRAs) are used to induce HIV gene expression in the hopes that the infected cells will then die as a result of viral cytopathic effects and/or natural or induced HIV-specific immune responses. *In vitro*, many compounds can activate HIV production from latency, including histone deacetylase inhibitors, methylation inhibitors, activators of NF-kB such as prostratin, and other compounds including disulfiram. Despite impressive activity in various *in vitr*o models, most LRAs, have weak activity in *ex vivo* studies using resting CD4 T-cells from patients on HAART. No reduction in the reservoir has been demonstrated in clinical trials, but there is evidence for increases in cell-associated HIV RNA and slight transient increases in plasma HIV RNA with certain inhibitors (*54*).

Another approach to the elimination of infected cells involves the use of antibodies directed at the HIV envelope protein, which should be expressed on the cell surface following reversal of latency. In particular, there has been interest in bNAbs that can recognize a vast range of HIV isolates. While these antibodies may be difficult to induce by vaccination, there is hope that passive infusion of bNAbs could contribute to the killing of latently infected cells if done in conjunction with LRA treatment. Together, these findings suggest that it will be possible to reverse HIV latency *in vivo* (*54*).

1.8. HIV-1 CA as a drug target

1.8.1. HIV-1 CA structure

The 26-kD capsid protein (CA) is initially translated as the central domain of the 55-kD Gag polyprotein, where it functions in viral assembly and in packaging the cellular protein CypA (*60*). Upon budding, Gag is processed by the viral protease to produce MA (132 residues), CA (231 residues), and NC (54 residues), as well as other smaller peptides (Figure 1-6, A). The cleaved proteins subsequently undergo a dramatic morphological rearrangement – maturation – to create the infectious viral particle. The MA protein lines the inner surface of the viral membrane, while the ~1,500 CA molecules form a distinctive fullerene-like conical core structure (capsid core or core) that encloses the NC/RNA complex at the center of the virion (Figure 1-6, B). Upon infection, the capsid core must subsequently disassemble or rearrange to allow reverse transcription of the RNA genome and active transport of the pre-integration complex (PIC) into the nucleus.

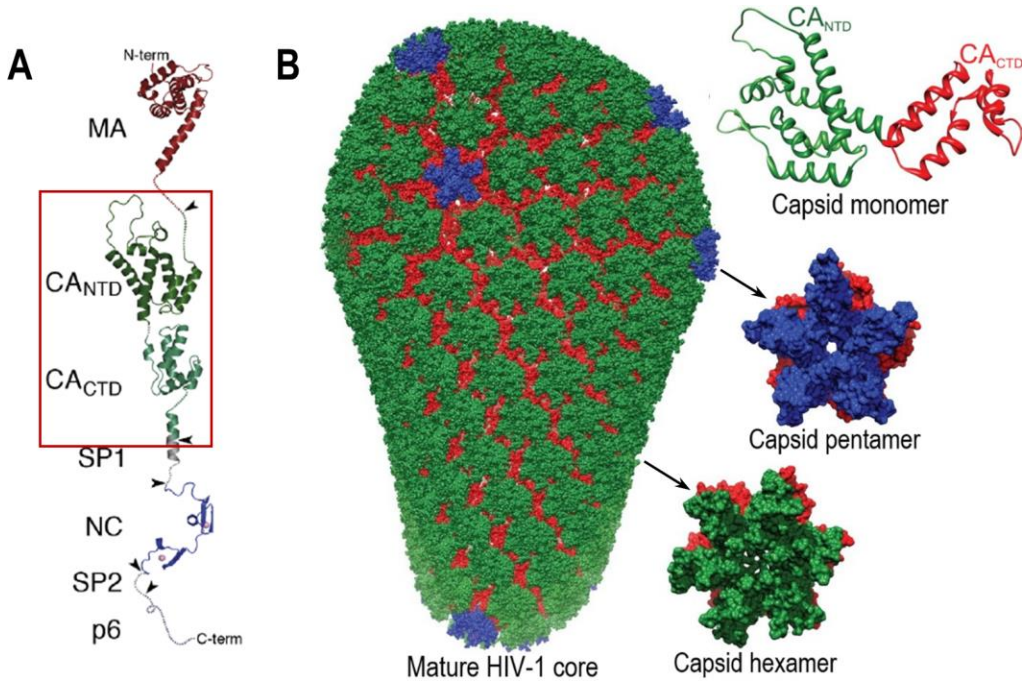


Figure 1-6. HIV-1 capsid.

(**A**) Structure of Gag polyprotein with CA in red box (arrows indicate HIV-1 protease cleavage sites), adapted from Ganser-Pornillos *et al.* (*61*). (**B**) Model of the HIV-1 core consisting of CA hexamer (green CA_{NTD}s, red CA_{CTD}s) and pentamer rings (blue CA_{NTD}s, red CA_{CTD}s), adapted from Deshmukh *et al.* (*62*).

The structure of CA and relevant CA assemblies, including the core or CA tubes, have been studied extensively over the past ~25 years. The CA protein is highly helical with two independently folded domains, the N-terminal domain (CA_{NTD}) and C-terminal domain (CA_{CTD}), which are connected by a flexible inter-domain linker (*62-66*).

CA_{NTD} consists of β-hairpin, five long α-helices (H1, H2, H3, H4, and H7) and two shorter ones (H5 and H6). The long helices form a coiled coil-like structure. The nonpolar faces form the interior of the super-helix and provide hydrophobic interactions between them. Neighboring helices are anti-parallel, except for helices H1 and H3, which are parallel. There is minimal contact between helices H3 and H4, leaving a small gap in the five-stranded coiled coil-like structure (*60, 66-69*). Mutations and deletions in the CA_{NTD} typically give rise to virions that assemble and bud, but they are noninfectious (*70-72*).

Many of these mutant viruses have aberrant capsid morphologies. Another important function of CA_{NTD} is to bind CypA, which results in packaging of ~200 CypA copies into the virion (*60, 73, 74*).

CA_{CTD} has an ovoid fold that comprises a 3_{10}-helix, and an extended strand followed by four α-helices (H8, H9, H10, and H11). The N-terminus of H10 is linked to the C-terminus of H11 by a disulfide bond between Cys198 and Cys218. This domain also contains a conserved stretch of 20 amino acids (residues 153-172), termed the major homology region (MHR). The MHR forms a compact strand-turn-helix motif that packs against the C-terminal end of H9 (*66, 75-78*). The CA_{CTD} appears to function primarily as an assembly domain and is required for Gag oligomerization, CA dimerization (*75*), and viral assembly (*70, 71, 79*). The MHR, which is conserved in onco- and lentiviruses and the yeast retrotransposon Ty-3, is essential for viral replication as different MHR mutations block viral replication at distinct stages – including assembly, maturation, and early steps of infectivity (*79*). The correlation between proper core assembly and viral infectivity suggests that HIV-1 CA plays essential roles during both early (entry, uncoating, nuclear trafficking, and entry) and late steps (assembly and maturation) of the replication cycle.

Purified HIV-1 CA can assemble into hollow cylindrical particles *in vitro* without the addition of any other viral and cellular protein or nucleic acid. Assembly occurs at neutral pH and requires high protein and salt concentrations, indicating that weak hydrophobic interactions mediate particle formation (*80*). The cylindrical structures observed *in vitro* have a diameter of approximately 55 nm but exhibit a tubular shape and are of various length (*81*).

In mature HIV particles, CA forms a cone-shaped capsid shell with an average length of ~103 nm, an average diameter of ~52 nm at the broad end, and ~25 nm on the tapered end (*82*). It has been shown that a mixture of cones and cylinders (ratios as high as 2:3) forms spontaneously upon incubation of a CA-NC fusion protein with a purified 1400-nucleotide HIV-1 RNA template in the presence of salt (*83*). The synthetic cones were capped at both ends, and many appeared strikingly similar to authentic HIV-1 cores. Although asymmetric, the cores often seemed highly regular, suggesting that they might be constructed from a regular underlying lattice.

It was proposed that HIV-1 cores form cones using the same symmetry principles as carbon in the fullerene. The viral core assembles on a curved p6 lattice with five pentameric defects arrayed at the narrow end (Figure 1-6, B). Because a total of 12 pentameric defects are required to create a closed object from a hexagonal lattice, the wide end of the cone will be capped with an additional seven pentameric defects (*83*). The precise distribution of the pentamers will determine the shapes of the caps. The driving force for closing the conical lattice could be saturation of all possible CA interfaces, as has also been observed for carbon cones.

The cones and tubes seem to be closely related, as they form simultaneously, and individual particles can be identified in which a tube turns into a cone at a point of declination. These *in vitro* assemblies appear to be analogs to authentic HIV-1 capsids because both conical and tubular capsids are also observed in virions (at a ratio of ~20:1). Thus, structural analyses of the CA assemblies formed *in vitro* are relevant for understanding the organization of the viral capsid (*84*).

Cryo-electron microscopy (cryo-EM) reconstructions and molecular modeling have revealed general domain relationships in the hexagonal lattice and helped build models of the mature HIV-1 core (*65, 77, 83-85*). The core comprises ~250 CA hexamers and 12 pentamers that allow the shell to close (Figure 1-6, B). Six CA molecules associate laterally to form the hexamer unit *via* CA_{NTD}s. The CA_{CTD}s are located adjacent to and beneath the CA_{NTD}s, and form homodimers that connect neighboring hexamers. The CA hexamers have an exterior diameter of ~90 Å, and the protein shell is ~60 Å.

The mature HIV-1 capsid lattice comprises different types of CA-CA interfaces (*65, 77, 84-90*): a six-fold symmetric CA_{NTD}-CA_{NTD}, an intra- and inter-molecular CA_{NTD}-CA_{CTD}, an inter-hexamer 2-fold and 3-fold CA_{CTD}-CA_{CTD}. The CA-CA interactions at various interfaces are of critical importance for the proper formation of the core and multiple stages of the retroviral life cycle (*64, 65, 75, 77, 91-101*).

High-resolution structural analyses helped reveal some details of these interactions. X-ray and NMR structures of isolated CA_{NTD} (*69, 102, 103*) and CA_{CTD} (*75-78*) domains or full-length CA monomers, dimers or complexes with large antibody fragments (*62, 63, 66-68*) have provided a tertiary structure for HIV-1 CA and vital details for interactions at the 2-fold interface.

Structures of engineered cross-linked CA hexamers or pentamers elucidated the interactions between CA protomers in the hexamer and pentamer building blocks (*64, 95, 104*). A large number of water-mediated hydrogen bonds between the polar side chains and backbone atoms were shown to be around the hydrophobic center of the CA_{NTD}-CA_{NTD} interface. These water-mediated hydrogen bonds can reposition themselves to accommodate the variations between slightly different pentamer and hexamer assemblies

with a conserved hydrophobic core at the CA_{NTD}-CA_{NTD} interface. Also, the charged residues are positioned closer to the center of the pentamer assembly, implying that electrostatic interactions may be the switch to form the quasi-equivalent assembly. By comparing the structures of different hexameric crystals, they showed that the structural differences could be recapitulated by varying relative orientations between CA_{NTD}s and CA_{CTD}s without changing the structure of individual domains. In this process, it appears that different dimer interfaces are needed to induce the variable curvature on the capsid surface. Although engineered mutations improved the biophysical aggregation properties of the crosslinked hexamers, they also significantly altered key interactions at the 2-fold interface and eliminated interactions at the 3-fold interface. Hence, despite advances in the field of CA structural biology, the molecular details of the CA-CA interactions that govern multiple steps of the virus life cycle remained insufficiently described.

1.8.2. HIV-1 CA-targeting antivirals

Following viral infection and fusion of the viral and cellular membranes, the core undergoes controlled uncoating, which gives rise to the initiation of downstream replication steps, specifically, reverse transcription and nuclear entry (*103, 105*). Proper capsid assembly is required for the production of infectious virions. Importantly, mutations that are expected to affect CA-CA interactions can have a profound effect on the uncoating, reverse transcription, nuclear entry, and assembly processes (*64, 65, 75, 77, 91, 92, 94-101, 106, 107*). Hence, CA is increasingly viewed as an attractive therapeutic target (*98, 108, 109*).

An increasing number of small molecules and peptide-based compounds have been reported to affect various aspects of CA function. Antiviral activity is mediated by disrupting CA-CA interactions in the immature Gag lattice, the mature CA core, or both. Those CA-targeting antivirals have been shown to bind at least four different sites of CA_{NTD} and CA_{CTD}.

Site 1 (Figure 1-7, blue) was the first one to be targeted by small-molecule CAP-1 (*103, 110*). Since the initial discovery of CAP-1, several other small molecules that inhibit the assembly and functions of HIV-1 CA have been discovered including multiple CAP-1 derivatives (*111*), as well as compounds bearing acylhydrazone (*112-114*), benzodiazepine (BD) (*115, 116*) and benzimidazole (BM) chemotypes (*116, 117*).

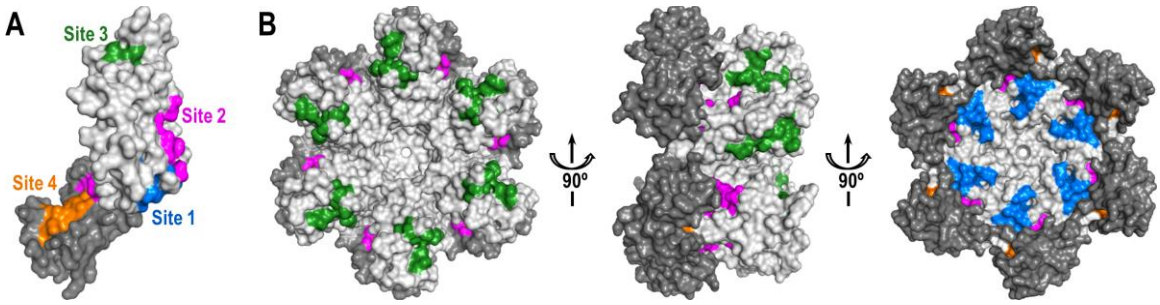


Figure 1-7. Binding sites of CA-targeting compounds to HIV-CA.

(**A**) Distinct binding sites of CA-targeting antivirals: Site 1 – blue (PDB ID: 4E91), Site 2 – pink (PDB ID: 4U0E), Site 3 – green (PDB ID: 4E91), and Site 4 – orange (PDB ID: 2BUO). (**B**) Three orthogonal views of binding sites in the context of CA hexamer.

Site 2 (Figure 1-7, pink) is the binding site of cellular proteins cleavage and polyadenylation specificity factor subunit 6 (CPSF6) and nuclear pore complex protein Nup153 (Nup153) (*118-120*). It has been targeted by PF74 (*120, 121*) and pyrrolopyrazolone analogs, e.g., BI-2 (*119, 122*). Those compounds seem to act similarly: they destabilize HIV-1 capsid core in the fate-of-capsid experiments, increase CA-NC assembly, and stabilize preassembled CA-NC complexes as well as prevent CPSF6 and

Nup153 binding to the HIV-1 core (*123, 124*). However, PF74 and BI-2 also differ. PF74 is active during both early and late stages of viral replication, while BI-2 is active only during early post-entry events (*119*). PF74 acts at a step before reverse transcription while BI-2 has no impact on reverse transcription, but acts before nuclear import of the PIC (note that at low concentrations PF74 has been shown to block infection independently from reverse transcription) (*125*). Although BI-2 and PF74 share the same binding site on CA_{NTD}, in the context of the hexameric capsid, BI-2 does not interact across the CA_{NTD}-CA_{CTD} interface. Different affinities for capsid between these two compounds may explain these results or suggest that PF74 and BI-2 display different mechanisms of action on HIV-1 replication (*119*).

Additionally, this site is proposed to be the binding site for coumercyn A1 (C-A1), an antibiotic targeting gyrase B, which impairs HIV-1 integration in a capsid-dependent way without affecting reverse transcription (*126, 127*). It was suggested that C-A1 perturbs the sequence of nuclear events leading to correct viral integration by inducing greater uncoating and affecting the PIC structure bound to Nup153 (*127*).

Site 3 (Figure 1-7, green) has been targeted by compounds of the benzimidazole scaffold (*128, 129*) which have been reported to compete with the binding of interferon-inducible myxovirus resistance protein MxB.

Site 4 (Figure 1-7, orange), a hydrophobic cavity in the CA_{CTD}, has been targeted by a number of peptides and small molecules identified *via* screening or rationally designed (*101, 130-139*). Those compounds have been shown to destabilize the CA_{CTD}-CA_{CTD} 2-fold interface; however, they remain comparatively weak binders.

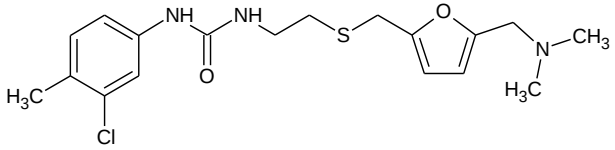
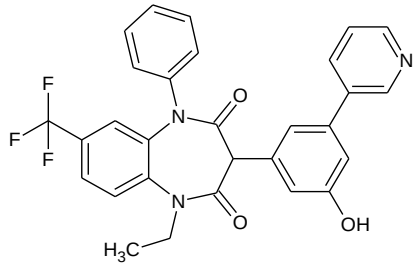
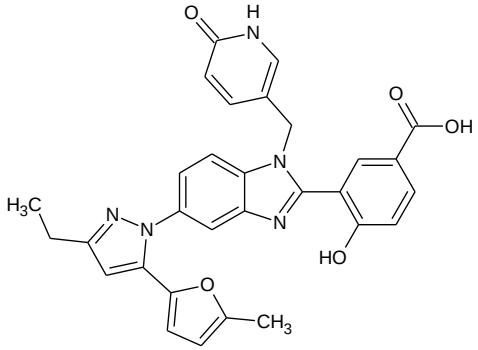
One of the most successful attempts to target Gag so far involves blocking maturation of the viral core. Bevirimat (BVM) is the first compound reported to inhibit HIV-1 replication by interfering with the final CA-sp1 cleavage step of Gag processing, which results in the accumulation of CA-sp1 precursor (p25) and immature non-infectious viral cores (*140, 141*). BVM has been shown to bind the CA-sp1 junction, which prevents the PR from cleaving it (*142, 143*). Because incomplete processing at the CA-sp1 junction strongly interferes with virion maturation and particle infectivity (*144, 145*), even a partial block to cleavage at this site can elicit a potent antiviral effect. An additional feature of maturation inhibitors that may contribute to their antiviral activity is their ability to stabilize the immature Gag lattice (*146*).

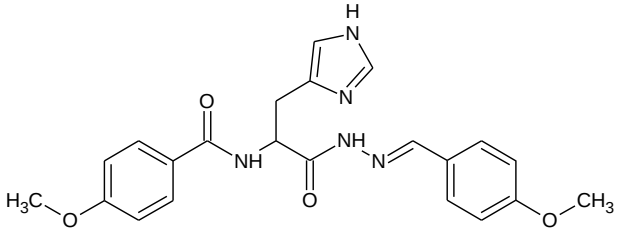
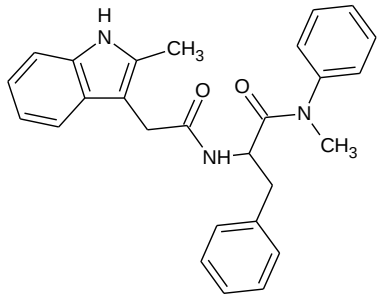
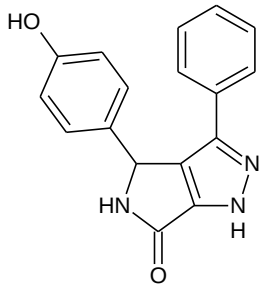
The *in vitro* results described above were followed by successful phase I and phase II clinical trials, demonstrating safety and efficacy in a small number of patients (*147, 148*). However, the larger phase IIb trials revealed a high baseline drug resistance to BVM (*149*). A significant fraction (~50 %) of treated patients had pre-existing resistant viruses that had significantly reduced BVM sensitivity (*6*). It is the only capsid effector that was clinically tested, however, due to the prevalence of resistance-conferring polymorphisms, BVM was discontinued as a potential therapeutic agent (*150, 151*).

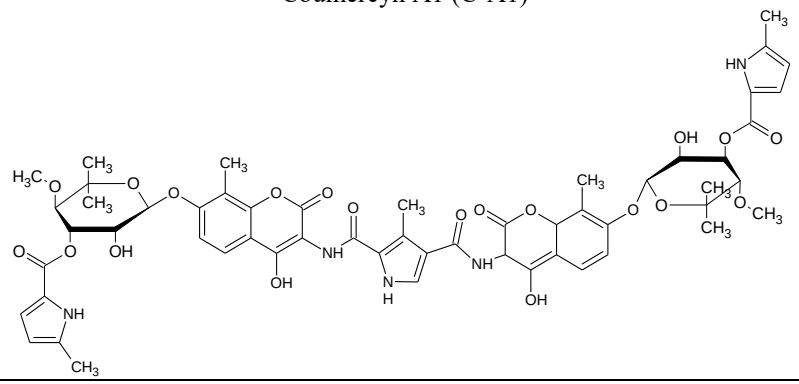
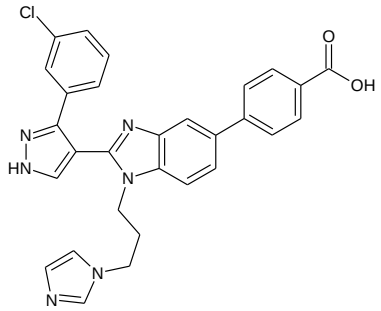
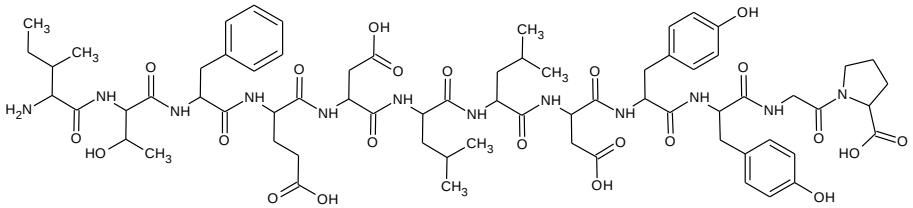
In addition to the CA-targeted antivirals described above, a number of effectors conferring mostly different phenotypes have been reported to interfere with CA assembly (Table 1-5). Taken together, the available examples demonstrate that various organic molecules can bind either in pockets located close to inter-subunit interfaces, or in other sites in the mature HIV-1 capsid, and sterically or allosterically interfere with CA assembly and/or stability and HIV-1 infection. Problems of low potency, low solubility and/or low

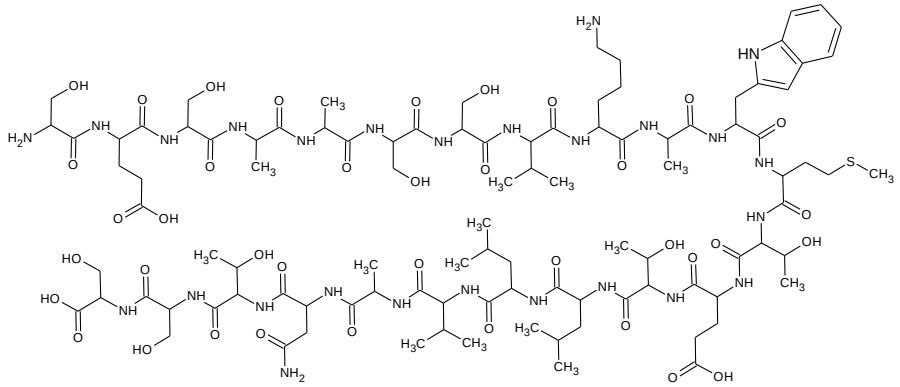
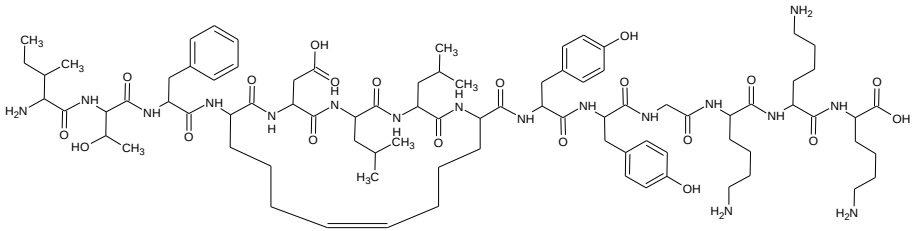
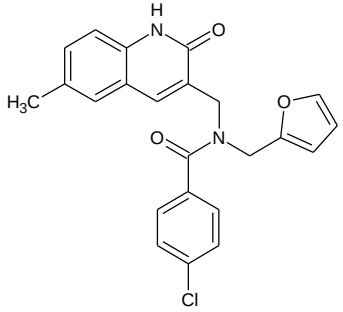
metabolic stability remain to be addressed, but the results obtained hold great promise for the development of anti-HIV-1 drugs based on organic compounds targeting CA.

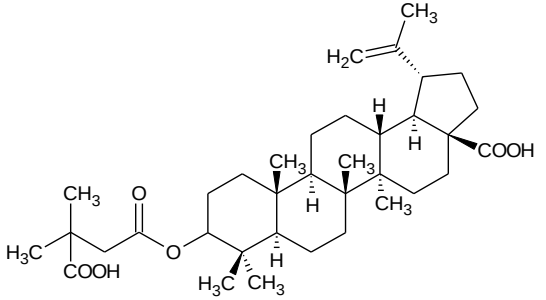
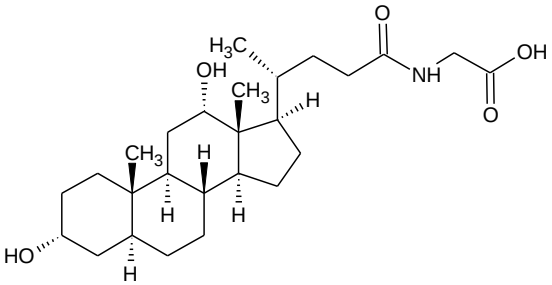
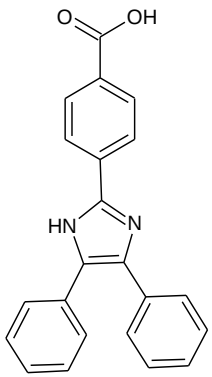
Table 1-5. Compounds targeting CA with demonstrated activity against HIV-1.

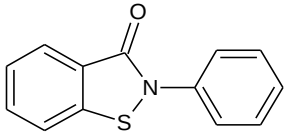
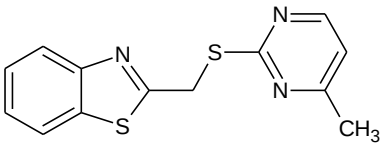
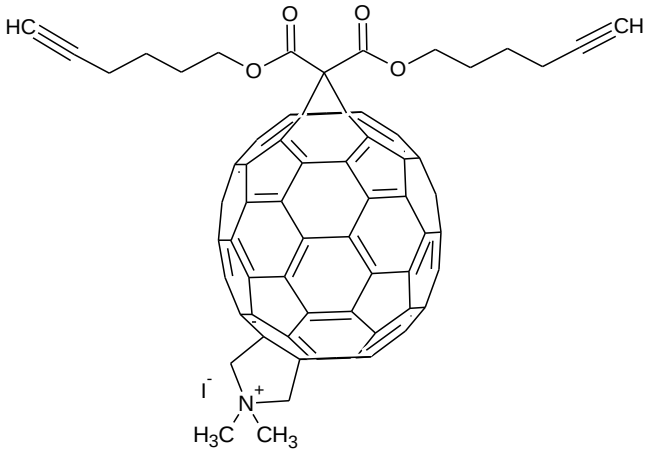
Compound	Mechanisms of action	Binding site
Site 1		
CAP-1 	Inhibits capsid assembly during viral maturation, and interferes to some extent with normal Gag–Gag interactions during assembly of the immature particle (*110, 152*).	CA_{NTD}, helices 1, 2, 4 and 7
BD 3 	Blocks Gag assembly and virion release (*115, 116*).	CA_{NTD}, helices 1, 2, 4, and 7
BM 4 	Inhibits assembly of mature conical capsids (*115-117*).	CA_{NTD}, helices 1, 2, 4 and 7

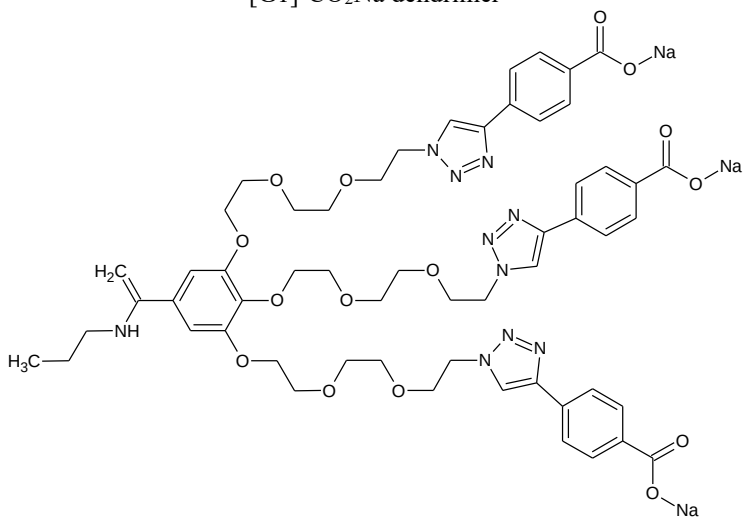
Compound	Mechanisms of action	Binding site
Compound 8b 	Inhibits capsid assembly (*113*).	CA_{NTD}, helices 1, 2, 4 and 7 (predicted by docking)
Site 2		
PF-3450074 (PF74) 	Enhances the rate of CA multimerization *in vitro*, accelerates capsid dissociation in cells, inhibits reverse transcription, prevents CPSF6 and Nup153 binding to the HIV-1 core (*119-121, 123, 125*).	CA_{NTD}-CA_{CTD}, helices 3, 4, 5, 7 and 8
BI-2 	Stabilizes the viral capsid *in vitro*, destabilizes the HIV-1 core during infection, prevents CPSF6 and Nup153 binding to the HIV-1 (*119, 122, 124*).	CA_{NTD}, helices 3, 4, 5 and 7

Compound	Mechanisms of action	Binding site
Coumercyn A1 (C-A1) 	Perturbs the sequence of nuclear events leading to correct viral integration (*126, 127*).	CA_{NTD}-CA_{CTD}, helices 3, 4, 5, 7 and 8 (predicted by docking)
Site 3		
BM 4 	Assembly and antiviral effects (*116, 128*).	CA_{NTD}, helices 4, 6 and 7
Site 4		
CAI 	Inhibits assembly of mature- and immature-like particles *in vitro* (*131, 132*).	CA_{CTD}, helices 8, 9 and 11

Compound	Mechanisms of action	Binding site
CAC1M 	Inhibits dimerization, *in vitro* CA assembly and infection (*134, 137*)	CA_{CTD}, helices 8, 9 and 11
NYAD-13 	Inhibits mature and immature HIV particle assembly (*135, 136*).	CA_{CTD}, helices 8, 9 and 11
NYAD6 	Disrupts the formation of mature-like particles (*138*).	CA_{CTD}, helices 8, 9 and 11 (predicted by docking)

Compound	Mechanisms of action	Binding site
Beviramat (PA-457) 	Blocks HIV-1 maturation by inhibiting CA-sp1 cleavage (*140, 142, 143*).	CA_{CTD}-sp1 junction helix
GDC 	Interferes with the Gag assembly (*153*).	CA_{NTD}, helices 1, 2, 4 and 7; CA_{CTD}
I-XW-053 	Disrupts CA assembly *in vitro*, inhibits reverse transcription, indirectly pointing to dysfunction in the uncoating process (*154*).	CA_{NTD}, helix 2; CA_{CTD}, helix 8 (predicted by docking)

Compound	Mechanisms of action	Binding site
Ebselen 	Increases the rate of CA multimerization *in vitro*, stabilizes the mature capsid, thus, negatively impacting the uncoating process (*108*).	Covalent Cys 198, Cys 218
BMMP 	Inhibits assembly of purified CA, destabilizes the core post-entry (*155*).	Unknown
Fullerene 	Blocks Gag and Gag-Pol processing (*156*).	Unknown

Compound	Mechanisms of action	Binding site
[G1]-CO_2Na dendrimer 	Hamper the *in vitro* assembly of the HIV-1 capsid (*157*).	CA_{CTD}-CA_{CTD}
Recombinant ankyrin $Ank^{GAG}1D4$	Interferes with the Gag assembly and budding pathway at the plasma membrane inhibits HIV-1 production (*158*).	CA_{NTD}, residues 1-110

1.9. Cellular factors interacting with HIV-1 CA

1.9.1. CypA

Cyclophilin A (CypA) is a host peptidyl prolyl isomerase that has been known to interact with Gag for more than 20 years (*73, 159, 160*). CypA is incorporated into virus particles, although the more biologically relevant interaction seems to occur in the cytoplasm of the target cell, where CypA can promote infection in some cell types (*161-164*). CypA incorporation is inhibited by immunosuppressive drug cyclosporin A (CsA, a ligand for CypA) (*165, 166*). When CsA disrupts CypA-Gag interactions, the resulting viruses are less infectious (*73, 160*). However, CsA inhibition of CypA is incomplete and is cell type-dependent (*167, 168*). Viruses deficient in CypA contain a full complement of viral proteins, but fail to synthesize viral DNA (*165*). Virions lacking CypA are mainly of an immature type and lack condensed core structures (*169, 170*).

CypA facilitates trans/cis interconversion of peptidylprolyl bonds to catalyze slow folding reactions of proteins (*171*). CypA binds to a particular conserved proline-rich motif (PX5PX2PX4P) in HIV-1 CA (central domain of Gag) (*73*) through its hydrophobic pocket (*172, 173*). The proline motif forms a loop extending away from the main body of CA (*60, 69*) and CypA acts on the peptide bond between residues Gly89 and Pro90 (*174*). CypA binding, its rotamase activity, and the structure of CA are all essential for virus infectivity (*175*).

In the context of an assembled CA lattice, it is tempting to speculate that CypA-induced conformational changes in CA may provide the mechanistic basis for disassembly of the capsid core. However, data supporting this hypothesis are contradictory, with one

study observing *in vitro* destabilization of CA–NC complexes (*176*), and a different study finding that CypA has the opposite effect, stabilizing viral cores *in vitro (177)*.

In summary, it is hard to assign a single, critical role to CypA; it seems to be apparent that CypA plays an essential part either during uncoating or during the step of infection that occurs concurrently with it.

1.9.2. TRIM5α and TRIMCyp

Tripartite motif-containing protein 5α (TRIM5α) is one of the ~70 tripartite motif (TRIM)-containing family proteins (*178*). TRIM5α consists of N-terminal RING domain, B-box type 2 domains connected to the central coiled-coil domain (*179*). In the case of TRIM5, the coiled coil drives the formation of dimers. The nature of the C-terminal domain can vary widely among proteins of the TRIM family (*178*). TRIM5α also has B30.2 (SPRY) C-terminal domain, which harbors the recognition site for HIV capsid.

Interestingly, TRIM5α proteins are unable to restrict retroviruses that are found in the same host species, but they potently inhibit retroviruses that are found in other species (*180*). Thus, TRIM5α can confer broad innate immunity to retrovirus infection in primate cells, and is likely to be an important natural barrier to cross-species retrovirus transmission (*180, 181*).

As discussed above, a number of lentivirus capsids bind to the abundant host cell chaperone protein CypA *via* an exposed peptide loop on the surface of the assembled capsid. Even though the precise role of this interaction is not entirely clear, both the sequence of this loop and the CypA can affect the sensitivity of HIV-1 to TRIM5α (*182-184*). In owl monkeys (*185*) and some macaques, retrotransposition events have placed

CypA complementary DNA into the TRIM5 locus. As a result, a chimeric gene is expressed as a TRIM5-CypA fusion protein (TRIMCyp), where CypA domain replaces the SPRY domain. In general, as predicted, TRIMCyp proteins are potent inhibitors of lentiviruses whose capsids bind CypA.

The mechanisms by which TRIM5 proteins act to block retroviral infection are not completely understood. Nevertheless, TRIM5α and TRIMCyp bind directly to HIV-1 capsid. Following viral entry into the target cells, TRIM5α and TRIMCyp accelerate capsid fragmentation, therefore disrupting the architecture of reverse transcription complex (RTC) and inhibiting the synthesis of viral cDNA (*186*). Recently, cryo-electron microscopy analyses showed a propensity of purified TRIM5α to assemble into hexagonal lattices that can interact in an ordered, polyvalent manner with preformed hexagonal lattices of HIV-1 CA. In any case, the TRIM lattice must presumably be distorted from ideal geometries to accommodate the irregularly curved surfaces of retroviral capsids (*187*). Thus, slight mismatches between the interacting lattices will tend to create discontinuities in the extended CA lattice that could contribute to the accelerated capsid dissociation that accompanies restriction.

1.9.3. Mx2

Human myxovirus resistance 1 (Mx1 or MxA) and myxovirus resistance 2 (Mx2 or MxB) are members of the IFN-inducible dynamin-like guanosine triphosphatase (GTPase) superfamily that plays roles in cellular processes that require membrane remodeling, such as vesicular transport and cytokinesis (*188*). Human Mx1, which shares 63 % amino acid identity with Mx2, is a well-documented inhibitor against a variety of

RNA and DNA viruses, including flu and influenza-like togaviruses, bunyaviruses, rhabdoviruses, and thogoto, coxsackie, and hepatitis B viruses (*188*). However, little was known about the antiviral activity of its close relative Mx2. Only recently, the Mx2 was described as a new HIV-specific restriction factor (*189, 190*). Unlike the majority of restriction factors, Mx proteins are highly conserved in vertebrates. Most mammals have two Mx genes due to an ancient duplication event that resulted in the current Mx1 (or MxA) and Mx2 (or MxB) paralogs (*191*). Like many restriction factors, the Mx1 and Mx2 expression are induced by type I (α/β) and type III (λ) interferons, but not by other cytokines. However, they are not expressed constitutively and are not induced directly by viruses, unless the infection triggers an interferon response.

Similar to dynamins, Mx proteins are characterized by an N-terminal GTPase domain (G), a middle domain (MD), and a C-terminal GTPase effector domain (GED) (*192*). Unlike Mx1, Mx2 exists in two isoforms (long and short), which can display different subcellular localizations. The long isoform contains a nuclear localization signal (NLS) in the N-terminus, which is required for anti-HIV activity (*129, 189, 190, 193, 194*).

The exact molecular mechanism by which Mx2 blocks HIV remains largely unknown. It has been shown that MxB interacts with *in vitro* assembled HIV-1 capsid-nucleocapsid (CA-NC) complexes (*129*). At the moment, there are several proposed mechanisms for Mx2 antiviral activity such as defects in uncoating (*129*), in nuclear import (*189, 190*), or in the integration process by a CypA-dependent mechanism (*189, 193*). Future studies need to elucidate which of these steps are targeted by Mx2 (*195*).

1.9.4. CPSF6

Cleavage and polyadenylation specificity factor 6 (CPSF6 also known as CF-Im-68) is an mRNA-processing protein that shuttles between the nucleus and cytoplasm (*125, 196*). CPSF6 shows three distinct domains: an N-terminal RNA binding domain containing the RNP1 (residues 124–131) and RNP2 (83–88) motifs, a proline-rich middle part (208–398, 47 % prolines), and a C-terminus, consisting mainly of arginine residues alternating with glutamate, aspartate, and serine residues (SR)-rich domain (489–551) (*197*). This domain organization is strongly reminiscent of that found in SR proteins and SR protein–related polypeptides involved in pre-mRNA splicing.

SR-rich nuclear-localization signal maintains a predominately nuclear steady state localization (*125*), although the protein shuttles dynamically between the nucleus and cytoplasm (*111*). Notably, it has been found that a C-terminally truncated form of mouse CPSF6 ($CPSF6_{1–358}$, which is missing SR-rich domain and is therefore predominately cytoplasmic) potently restricts HIV-1 and SIV infection, but not infection by other retroviruses (*198*). This suggests a role for the CPSF6 in the ability of primate lentiviruses to infect nondividing cells. Cells expressing $CPSF6_{1–358}$ retain standard reverse transcription, but show the impaired nuclear entry of HIV-1 (*198*).

$CPSF6_{1–358}$-mediated restriction mapped to HIV-1 CA, as a single amino acid substitution within CA (N74D) bypasses this restriction (*198*). It has also been demonstrated that CPSF6 stabilizes viral cores *in vivo* (*199*) and CA–NC tubes *in vitro* (*176*). Moreover, crystal structures of HIV-1 CA complexes (either N-terminus or cross-linked CA) with $CPSF6_{313-327}$ peptide revealed details of interactions in the binding pocket (*118*), which is formed by intermolecular interactions between the CA_{CTD} and CA_{NTD} of

neighboring CA molecules (*119, 200*). In summary, it seems likely that full-length CPSF6 has a role in mediating the nuclear import of the PIC, although this has yet to be formally demonstrated (*201*).

1.9.5. TNPO3

Three genome-wide small interfering RNA screens aiming to identify cellular factors required for HIV-1 infection identified transportin 3 (TNPO3 also known as TRN-SR2) as a component facilitating late stages of HIV-1 infection (*202-204*). TNPO3 is a β-karyopherin, which governs the nuclear import of specific cargoes, including of SR-rich proteins (*205*). TNPO3 is a solenoid-type protein composed of 20 α-hairpin motifs (two antiparallel α-helices joined by a short linker) known as HEAT repeats, with its N- and C-terminal arches facing each other (*206, 207*).

Early studies proposed a direct interaction between TNPO3 and viral integrase (*208*), subsequent genetic evidence additionally suggested a functional interplay between TNPO3 and viral capsid (*199, 209, 210*).

Initial studies showed that TNPO3 depletion reduces the amount of viral integration, but does not diminish viral reverse transcription (*177, 202, 208, 210, 211*). Further studies demonstrated that TNPO3 depletion inhibits 2-LTR circle formation (*199*).

Remarkably, although TNPO3 knockdown potently inhibited infectivity of wild-type HIV-1, infection by the 14 CA mutants (including E45A, Q63A/Q67A and N74D that were also reported to be resistant to $CPSF6_{1-358}$,) out of a panel of 27 HIV-1 CA mutants was unaffected by TNPO3 depletion (*199, 210*).

Collectively, these studies suggest that TNPO3 participates in the nuclear translocation of the viral PIC, which might be attributable to one of two mechanisms. First, TNPO3 may affect HIV-1 nuclear import directly by binding to CA. Given the independence of the N74D mutant on TNPO3 depletion, this may occur at the same CPSF6-binding site. Alternatively, TNPO3 may mediate the proper nucleoplasmic localization of other proteins relevant to the nuclear import of the PIC, such as CPSF6 (*199*), as suggested by its ability to regulate the localization of SR-rich proteins (*205*).

1.9.6. Nup358

Nuclear pore complexes (NPCs) are large (~60 MDa in yeast and ~120 MDa in vertebrates) assemblies that mediate nucleocytoplasmic transport in the eukaryotic cells. They are built from 34 distinct proteins – 31 soluble nucleoporins (Nups) and three integral membrane proteins of the pore membrane domain (Poms) (*212, 213*). Nups assemble in multiple copies to form an eight-fold rotationally symmetric complex (*214*). In addition to its central role, the NPC has various other essential functions, such as chromatin organization, regulation of transcription, and DNA repair (*213*).

Two components of NPCs, nucleoporin 358 (Nup358; also known as RanBP2) and Nup153, were identified in genome-wide screens for cellular factors required for HIV-1 infection (*202-204*).

Nup358 is a component of nuclear pore filaments, which project towards the cytoplasm from the NPC and, like many other NPC channel proteins, has phenylalanine/glycine (FG) repeats that form a hydrophobic meshwork that regulates traffic of molecules above ~40 kDa across the pore (*212, 215*).

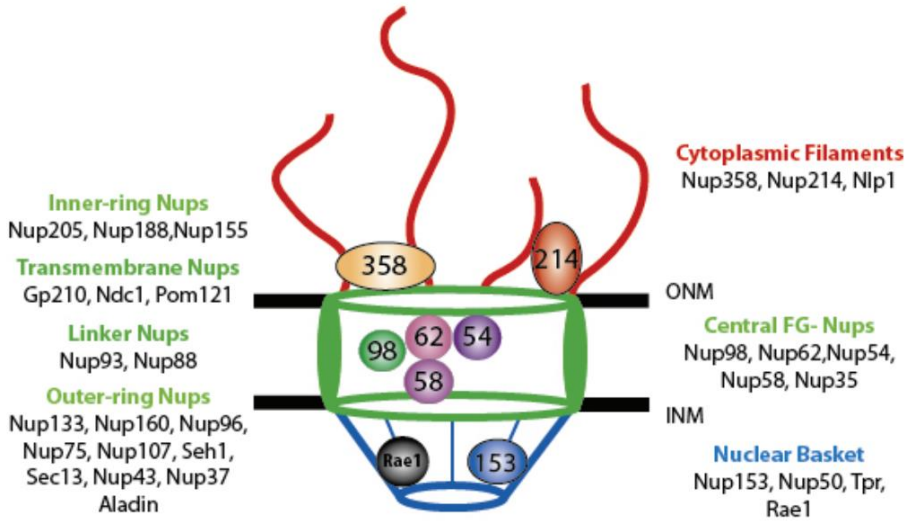


Figure 1-8. The nuclear pore complex (NPC).

Schematic representation of a cross-section view of an NPC as it spans the inner (INM) and outer (ONM) nuclear membranes. Vertebrate Nups are listed based on their localization within the NPC. The central core (green) is composed of the inner-ring Nups, transmembrane Nups, linker Nups, outer-ring Nups and the phenylalanine-glycine (FG)-Nups. The location of select nucleoporins within the NPC is shown. This figure was copied from (*216*).

The polypeptide chain of Nup358 contains 3224 residues and can approximately be divided into several distinct regions: an N-terminal ~830-residue α-helical region (*217*) followed by four RanGTP binding domains, eight consecutive zinc finger motifs, an E3 ligase domain, and a C-terminal CypA homology domain (*218*). These domains are connected by predicted unstructured regions containing FG repeats.

Several roles have been proposed for Nup358 involving cell cycle control, nuclear export, and transportin/importin dependent nuclear import (*212*). Additionally, Nup358 is a cofactor for HIV-1 replication. Nup358 depletion reduces HIV-1 infectivity, 2-LTR circle formation, and integration, but leaves the level of viral reverse transcription unchanged (*219*). Moreover, Nup358 depletion leads to mistargeting of HIV integration site, suggesting that Nup358 depletion prevents the efficient nuclear import of the RTC (*220-222*).

Recently, Nup358 has been shown to contain a C-terminal CypA-homology domain (Nup358Cyp) that interacts with the HIV-1 CA (*220, 223, 224*) and induces its isomerization (*223*), leading to the hypothesis that Nup358-mediated isomerization of CA causes core uncoating. However, other studies confirmed that Nup358 plays a major role in the HIV-1 life cycle, albeit one that is independent of the Nup358Cyp (*225*), leaving the mechanism by which Nup358 engages HIV-1 unclear.

1.9.7. Nup153

Nup153 is also present in the NPC and is conceptually similar to Nup358 (*213*). However, Nup153 is localized to the nuclear side of the NPC, nuclear basket, although several studies suggest the three major domains of Nup153 are localized to different regions of the NPC (*226-229*).

The primary structure of Nup153 can be roughly divided into three regions (*229*). N-terminal region contains a nuclear envelope targeting cassette, a nuclear pore associating region, and an RNA binding domain (*230*). The central domain consists of four zinc finger motifs that share the highest similarity with zinc fingers of Nup358. Together, those domains of Nup153 and Nup358 define the so-called RanBP2-type zinc finger class that is unique with respect to the repetitive arrangement of four to eight zinc fingers, and also shares conserved, nonstructural residues (*231*). Finally, C-terminal portion encompasses ~30 irregularly spaced degenerate pentapeptide repeats xFxFG [xxxFG, xFxFx, xFxFG, where x is any amino acid, F is phenylalanine, and G is glycine]. Different classes of FG-containing repeats are found in many Nups and often represent binding sites for nuclear

import and export receptors (*232, 233*). In addition to participating in the formation of NPCs, Nup153 plays a pivotal role in cell cycle regulation and apoptosis (*234*).

The Nup153 expression has been shown to be required for HIV-1 infection with the major functional determinant of Nup153 dependency to be HIV-1 CA protein (*235*). A direct interaction between the CA N-terminal domain and Nup153 C-terminal domain has been demonstrated (*236*). Notably, Nup153 relies on the same binding pocket in CA that is required for interactions with other nuclear-import pathway factors, including CPSF6 and TNPO3 (*237*). Crystal structure of assembled HIV-1 CA hexamers with Nup153 peptide (residues 1407-1423) revealed that the binding pocket identified for CPSF6 peptide (residues 313–327) also accommodates Nup153. Both Nup153 and CPSF6 use a linear binding motif to interact with hexameric CA, anchored around a core F but otherwise mediate different contacts (*119*).

Current data suggest that Nup153 is likely necessary for trafficking the HIV-1 PIC through the nuclear pore and into the nucleus (*203, 235*). The viral nucleoprotein complex is likely to initially dock to the NPC by engaging Nup358 through its Nup358Cyp, though FG motifs, may also participate (*225*). While intact HIV-1 cores are too large to enter the NPC, partially disassembled CA cores may come far enough for the remaining CA to be accessible to FG domains of Nup153. CA interaction with Nup153 may serve two distinct roles during viral infection. First, Nup153 may be responsible for physically translocating the PIC by engaging CA molecules that may associate with it (*236, 237*). Second, CA interaction with Nup153 may be required to finish uncoating of the viral core at the NPC and prime the PIC for nuclear import (*236, 237*). Additional experiments are necessary to distinguish between these two possibilities (*201*).

1.9.8. Model of HIV-1 nuclear trafficking and entry

Despite intense study, the details of HIV-1 and other lentiviral nuclear import mechanisms remain incompletely understood. Based on the collective data, the following working model has been proposed (Figure 1-9) (*105, 236*). The initial steps of uncoating likely occur shortly after entry (*238, 239*), while the final events may occur at the NPC (*240*).

Based on recent studies (*241-243*), the model suggests that CypA prevents premature uncoating, thus facilitating normal reverse transcription, subsequent recruitment of the host cell proteins TNPO3, CPSF6, Nup358, Nup153, thereby leading to translocation of PICs through the NPCs (Figure 1-9, A) and integration.

Furthermore, alteration of this pathway by CA mutations or CypA or CPSF6 depletion results in premature uncoating and exposure of viral DNA, thus allowing detection by DNA sensors, e. g. cGAS (*242*), and production of soluble type 1 interferon (*241*), which can induce Mx2-mediated inhibition (*190*) of nuclear entry (Figure 1-9, B). Moreover, TNPO3, Nup153, or Nup358 knockdown results in altered integration site selection away from gene-dense regions of chromatin.

Finally, CA-targeting antivirals and simian restriction factors, such as TRIM5α and TRIMCyp, accelerate uncoating, therefore disrupting reverse transcription and inhibiting the synthesis of viral cDNA (Figure 1-9, C).

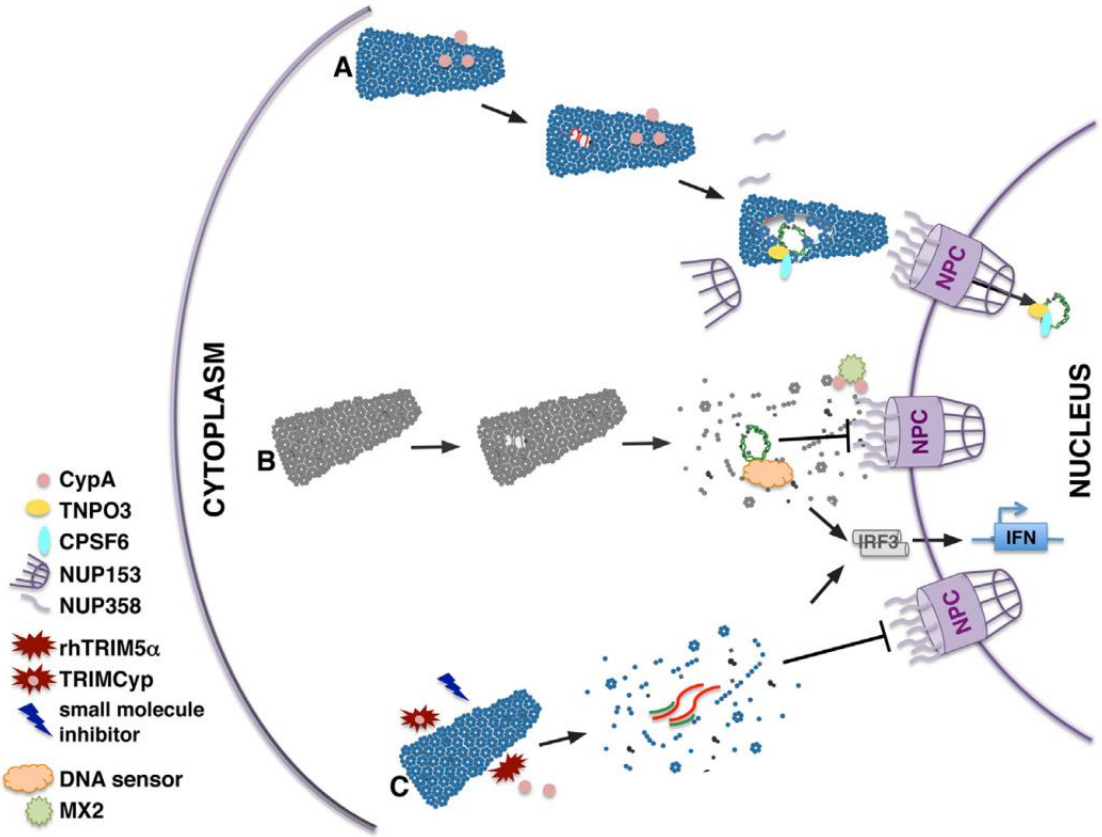


Figure 1-9. Model of HIV-1 nuclear trafficking and entry.

(**A**) Uncoating likely involves multiple steps and utilizes CypA to protect the viral DNA from immune surveillance. Nuclear entry is facilitated by CA interactions with host cell factors, including TNPO3, CPSF6, Nup153, and Nup358. (**B**) CA mutants, CypA or CPSF6 depletion, induces premature uncoating resulting in viral DNA exposure and innate immune response stimulation, which induces expression of Mx2, a host factor that prevents viral DNA nuclear entry. (**C**) Uncoating is accelerated by small molecule antivirals and primate restriction factors, TRIM5α and TRIMCyp, resulting in a reduction of reverse transcription and little nuclear import of viral DNA. This figure was copied from (*105*).

Chapter 2. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability

Note: This chapter is based on a research paper that was published in *Science* in 2015. The contributions of co-authors are described in the "Footnotes."

A. Introduction

The mature capsid of HIV-1 is formed from a single capsid protein (CA) containing N-terminal (CA_{NTD}, residues 1-145) and C-terminal (CA_{CTD}, residues 150-231) domains connected by a flexible linker region (*62-65, 95*). The capsid contains ~250 CA hexamers and 12 CA pentamers. CA hexamers comprise six CA_{NTD}s held together by CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} contacts between adjacent CAs, while the six CA_{CTD}s engage in inter-hexamer interactions (*62-65, 95*). CA-CA interactions affect capsid structural integrity and infectivity (*64, 65, 75, 92, 95, 96, 99, 106, 244*). Following viral entry, the capsid undergoes controlled disassembly (uncoating), which seems coordinated with reverse transcription (*92, 96*). Antivirals targeting CA (*110, 121, 122, 132, 138, 152*) include PF-3450074 (PF74), which has a bimodal mechanism of action (*119, 124, 125, 200*): at lower concentrations (nanomolar to ~2 μM) it competes with binding of host factors CPSF6 and Nup153, affecting nuclear entry. At higher concentrations, PF74 blocks uncoating and reverse transcription (*119, 123-125, 200, 237*). Crystal structures of PF74 with CA_{NTD} (CA_{NTD}-$_{PF74}$) (*121*) or cross-linked CA hexamers (CA_{XL}-$_{PF74}$) (*119, 200*) have shown that PF74 binds at the same site as CPSF6 and Nup153. However, the structural mechanism by which therapeutically relevant high concentrations of PF74 affect uncoating remains incompletely defined.

Cryo-electron microscopy (cryo-EM) studies have helped build informative models of the mature HIV-1 capsid (*65, 85, 95*). X-ray and NMR structures of CA domains or CA monomers or dimers have described some interactions at interfaces (*62-65, 76, 95, 132, 133*). Structures of cross-linked CA (CA_{XL}) elucidated the CA interactions in hexamers and pentamers (*64, 95*). However, engineered mutations left inter-hexamer interactions that govern virus uncoating and assembly insufficiently described.

B. Materials and methods

Protein expression and purification

Full-length, wild-type HIV-1 CA was cloned without fusion tags in pET11a plasmid provided by C. Tang. Protein was expressed in *E. coli* BL21(DE3)RIL as previously described and purified by ammonium sulfate precipitation followed by anion exchange chromatography (*245*).

Crystallization, cryoprotection, and dehydration

Screening for CA crystallization was done in 96-well trays with an Art Robbins Crystal Gryphon robotic crystallization system (Art Robbins Instruments, Sunnyvale, CA) using commercial (Hampton Research, Aliso Viejo, CA; Qiagen Inc, Valencia, CA; Rigaku Reagents, Bainbridge Island, WA) crystallization screening kits. CA crystals grew at 18 $^{\circ}$C in drops containing CA (<5 mg/ml), PEG 3350, sodium iodide, and sodium cacodylate. Hexagonal plate-like crystals appeared after five days, and crystal growth was completed in over two weeks. Crystals were briefly soaked in a solution containing 20 % glycerol before cryo-cooling in liquid nitrogen. Crystal dehydration was performed by serial

transfer of the protein crystal to drops containing cryoprotective agent and increasing concentrations of PEG 3350 (*246*).

HIV-1 CA complex with PF74

Native CA crystals, grown as described above, were soaked with PF74 (synthesized at the Chemical Biology Laboratory, NCI, as previously reported (*118*)) for approximately 24 h, briefly soaked in a solution containing 20 % glycerol for cryoprotection, and cryo-cooled in liquid nitrogen.

Data collection and structure determination

Data were collected on a Rigaku R-AXIS IV++ image plate system at the University of Missouri-Columbia and on a CMOS detector at Advanced Light Source (ALS) beamline 4.2.2, Lawrence Berkeley National Laboratory. More than ten datasets were collected (2.4-3.5 Å) and processed using XDS (*247*) or HKL3000 (*248*). The data were examined for the presence of systematic absences. However, no characteristic patterns were observed. Thus, the crystals were indexed in hexagonal space group P6 with unit cell dimensions ranging a = 87-93 Å, b = 87-93 Å, and c = 56-59 Å, and one CA molecule contained in the asymmetric unit. No twinning was present, as determined by either POINTLESS (*249*) or XTRIAGE (*250*). Space group and twinning were also verified in ZANUDA (*251*). The phase problem of *CA* (2.4 Å) was solved using single-wavelength anomalous diffraction (SAD) and of *dCA* and CA_{PF74} using molecular replacement, with the 3H47 and *CA* as starting models, respectively. For SAD, substructure solution, phasing, density modification, preliminary model building, iterative model building and refinement were carried out using SHELX C/D/E, MLPHARE, DM, SOLVE/RESOLVE,

BUCCANEER and REFMAC (*251, 252*) in HKL3000 (*248*). For molecular replacement, initial phases were solved *via* PHASER (*251*). Several rounds of iterative model building and refinement were carried out using Coot (*253*) and PHENIX (*250*), REFMAC (*251, 252*), or PDBREDO (http://www.cmbi.ru.nl/pdb_redo/), respectively. Coordinates and molecular topologies for PF74 were generated in PRODRG (*251*). The position of PF74 in the structure was built into a difference Fourier map calculated before inclusion of the respective structural elements in refinement. Structure validation of final models was performed with MOLPROBITY (http://molprobity.biochem.duke.edu/). Accessible and buried surface areas were calculated using PISA (*251*). The figures showing structural information were generated in PyMOL (http://www.pymol.org/) and CCP4MG (*251*).

C. Results and discussion

Native HIV-1 capsid structure

We crystallized native full-length CA (Figure 2-1, A) and solved its structure (*CA*) in space group P6 with one molecule/asymmetric unit (Figure 2-1, B, Table 1-1). CA subunits from neighboring hexamers are related by 2- and 3-fold crystallographic symmetry (Figure 2-1, C and D). The *CA* structure is in general agreement with the 9Å cryo-EM maps of the flattened CA hexagonal lattice (*85*) (Figure 2-2, A) and tubes (*65*) (Figure 2-2, B). The *CA* fold is also in agreement with crystal and NMR structures of full-length CA (*62-64, 85, 95*), CA_{NTD} and CA_{CTD} (*69, 76, 103, 121, 132, 152*). Key interactions between CA_{NTD} and CA_{CTD} likely to stabilize the capsid are *intra-hexamer* (CA_{NTD}-CA_{NTD}, CA_{NTD}-CA_{CTD}) contacts between adjacent CAs around the 6-fold axis, *inter-hexamer*

(CA_{CTD}-CA_{CTD}) contacts at the 2- and 3-fold axes (Figure 2-1, D) (*64, 65*), and *intra-subunit* (CA_{NTD}-CA_{CTD}) contacts.

Table 2-1. Summary of X-ray data collection and refinement statistics.

	CA	***dCA***	***CA_{PF74}***
Data collection			
X-ray source	R-Axis IV++	ALS 4.2.2	ALS 4.2.2
Software	HKL3000	XDS	XDS
Space group	P6	P6	P6
Unit cell dimensions			
a, b, c (Å)	92.3 92.3 57.3	87.1 87.1 55.9	92.2 92.2 57.0
α, β, γ (°)	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0
ASU content	1	1	1
Wavelength (Å)	1.5418	1.000137	1.000111
Resolution range (Å)[a]	20.0–2.4 (2.5–2.4)	60.0–2.8 (3.0–2.8)	60.0–2.7 (2.8–2.7)
R_{merge}	0.080 (>1)	0.100 (>1)	0.080 (>1)
R_{meas}	0.097 (>1)	0.110 (>1)	0.084 (>1)
R_{pim}	0.038 (0.385)	0.034 (0.425)	0.025 (0.535)
$\langle I/\sigma I \rangle$	19.0 (2.3)	19.8 (1.8)	24.4 (1.8)
$CC_{1/2}$ (%)	99.9 (67.6)	99.9 (70.4)	99.9 (71.4)
Completeness (%)	99.5 (97.6)	99.7 (98.2)	99.5 (96.5)
Redundancy	10.9 (9.9)	10.8 (9.8)	11.2 (10.7)
Mosaicity	0.17	0.79	0.16
Refinement			
Resolution (Å)	19.0–2.4	45.0–2.8	45.0–2.7
No. total reflections	117831	65579	85991
No. unique reflections	10834	6091	7709
No. test reflections[b]	475	277	354
R_{work} / R_{free}	22.2 / 24.9	20.8 / 25.0	21.0 / 24.1
No. atoms	1764	1709	1731
Protein	1680	1685	1663
Ligand/Ion	9	18	40
Water	75	6	28
Wilson B-factor (Å^2)	33.2	64.9	70.9
Average B-factors (Å^2)	47.7	76.0	87.9
Protein	48.2	76.1	88.9
Ligand/Ion	57.4	71.0	65.4
Water	36.1	51.3	63.8
RMS deviations			
Bond lengths (Å)	0.004	0.008	0.012
Bond angles (°)	0.81	1.06	1.45
MolProbity Statistics[c]			
All atom clash score	0.89	2.66	2.65
Rotamer outliers (%)	0	0	0
Cβ deviations >0.25 Å	0	0	0
Ramachandran[c]			
Favored region (%)	98	98	98
Outliers (%)	0	0	0
PDB accession code	4XFX	4XFY	4XFZ

[a] Values in parentheses are for highest-resolution shell; [b] random selection; [c] values obtained from MOLPROBITY.

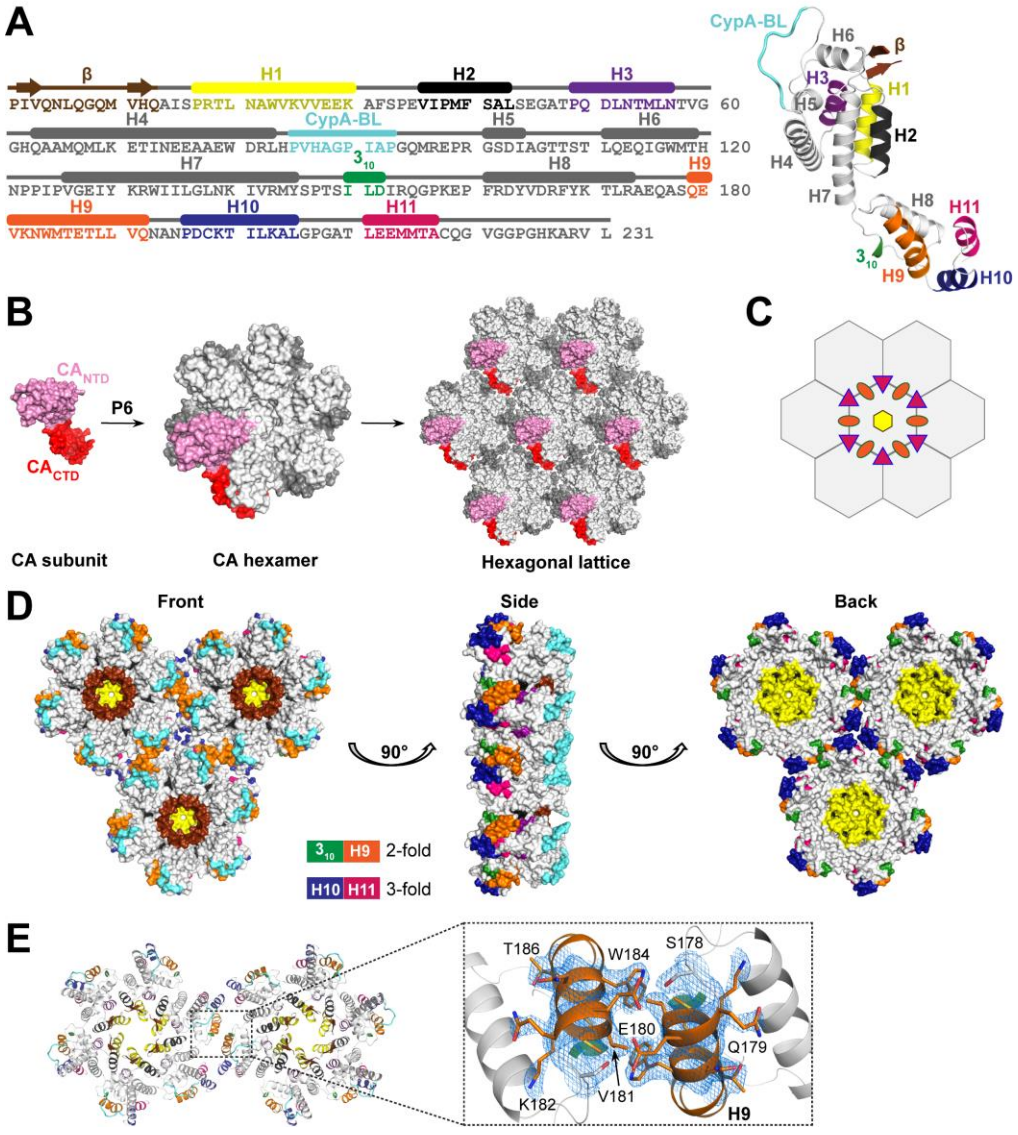


Figure 2-1. Crystal structure of native CA.

(**A**) Secondary structure and ribbon diagram of native *CA*. The CA_{NTD} comprises β-hairpin (1-13, brown), CypA-BL (85-93, cyan), and seven α-helices: H1 (17-30; yellow), H2 (36-43; black), H3 (49-57; purple), and H4 (63-83), H5 (101-104), H6 (111-119), and H7 (126-145) in gray. The CA_{CTD} comprises 3_{10}-helix (150-152; green), and four α-helices: H8 (161-173; gray), H9 (179-192; orange), H10 (196-205; blue), and H11 (211-217; pink). (**B**) Application of 6-fold crystallographic symmetry generates the *CA* hexagonal lattice (a single native *CA* molecule is shown in surface view representation; pink CA_{NTD}, red CA_{CTD}). (**C**) The six *CA* subunits in a hexamer are related by 6-fold crystallographic symmetry (yellow hexagon); *CA* subunits from neighboring hexamers are related by 2-fold (orange ovals) and 3-fold (pink triangles) crystallographic symmetry, shown at the inter-hexamer interfaces. (**D**) Orthogonal views of three native *CA* hexamers colored as in **A**. The hexamers are stabilized by interactions at the 6-fold (brown β-hairpin, yellow H1, black H2, and purple H3), 2-fold (green 3_{10} and orange H9), and 3-fold interfaces (blue H10 and pink H11). (**E**) Inter-hexamer interactions at the 2-fold interface. Interpretable electron density is now observed for all residues at the 2-fold interface of *CA* (2.4 Å; 2Fo–Fc; σ=1.2) including residues 176-187, which were previously disordered in crosslinked hexamer structures.

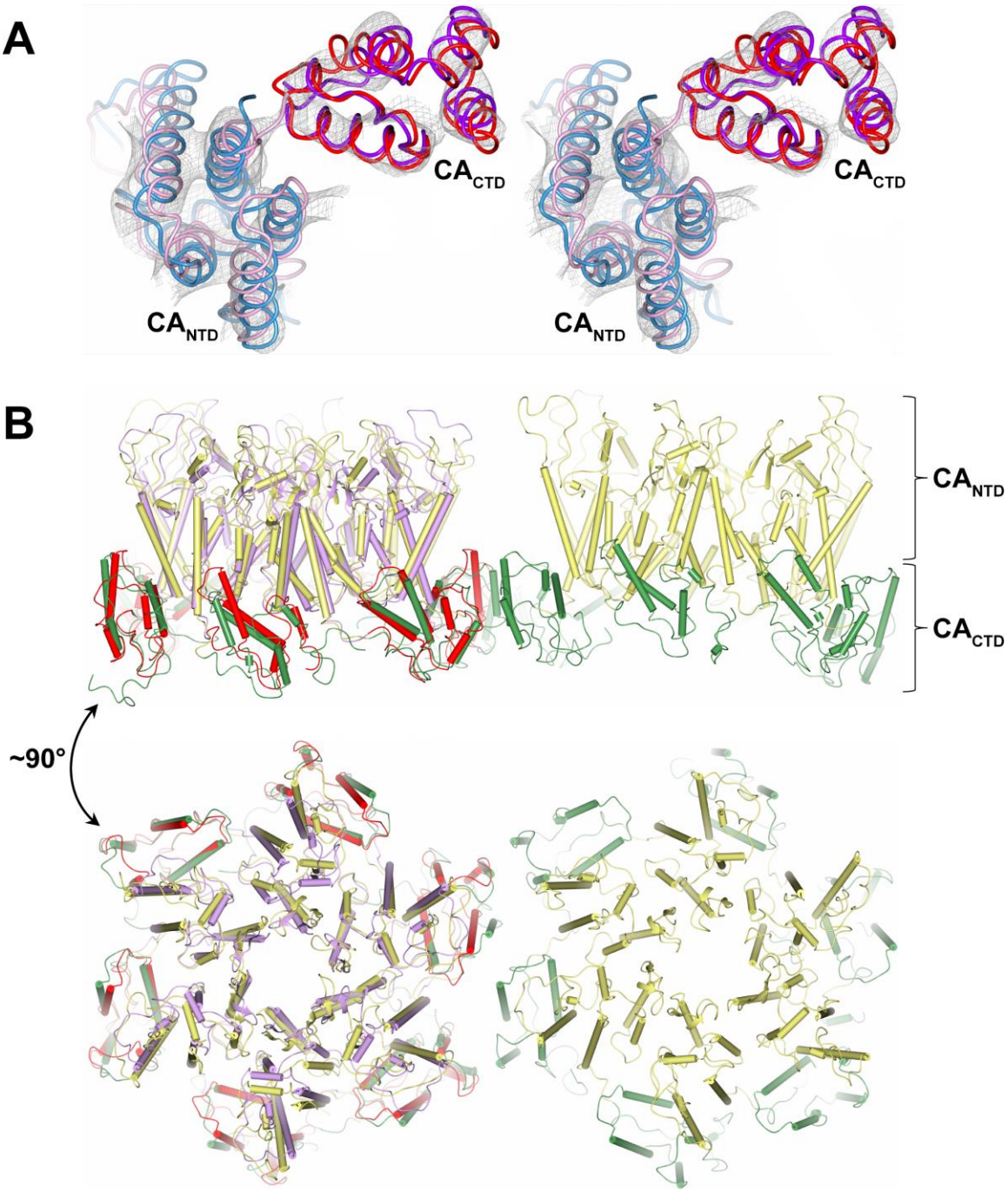


Figure 2-2. Comparison of CA crystal structure and cryo-EM models.

(**A**) Stereo view of superposed coordinates of *CA* (pink CA_{NTD}; red CA_{CTD}) and cryo-EM model of the flattened CA hexagonal lattice (PDB ID: 3DIK) (light blue CA_{NTD}; purple CA_{CTD}). Also shown is the 9 Å cryo-EM map (gray mesh) that corresponds to the 3DIK coordinates. (**B**) Two orthogonal views of a *CA* hexamer (pink CA_{NTD}s; red CA_{CTD}s) superposed with two adjacent hexamers from the 9 Å cryo-EM model of CA tubes (PDB ID: 3J34) (yellow CA_{NTD}s; green CA_{CTD}s). All superpositions were calculated by fitting the α-carbons of residues 1-221.

The inter-hexamer interactions at the 2-fold are clearly defined in *CA* (Figure 2-1, E). They involve multiple residues and water molecules (Figure 2-2, A). These contacts are reminiscent of, but different than, those in CA_{CTD} structures (Figure 2-4, A, Table 2-2) (*76, 132*). They also differ significantly from the original cross-linked hexameric (CA_{XL}, PDB ID: 3H47) (Figure 2-4, B, Table 2-2) and other CA structures containing W184A and M185A (*84*) at the dimerization interface (*64, 119, 200*).

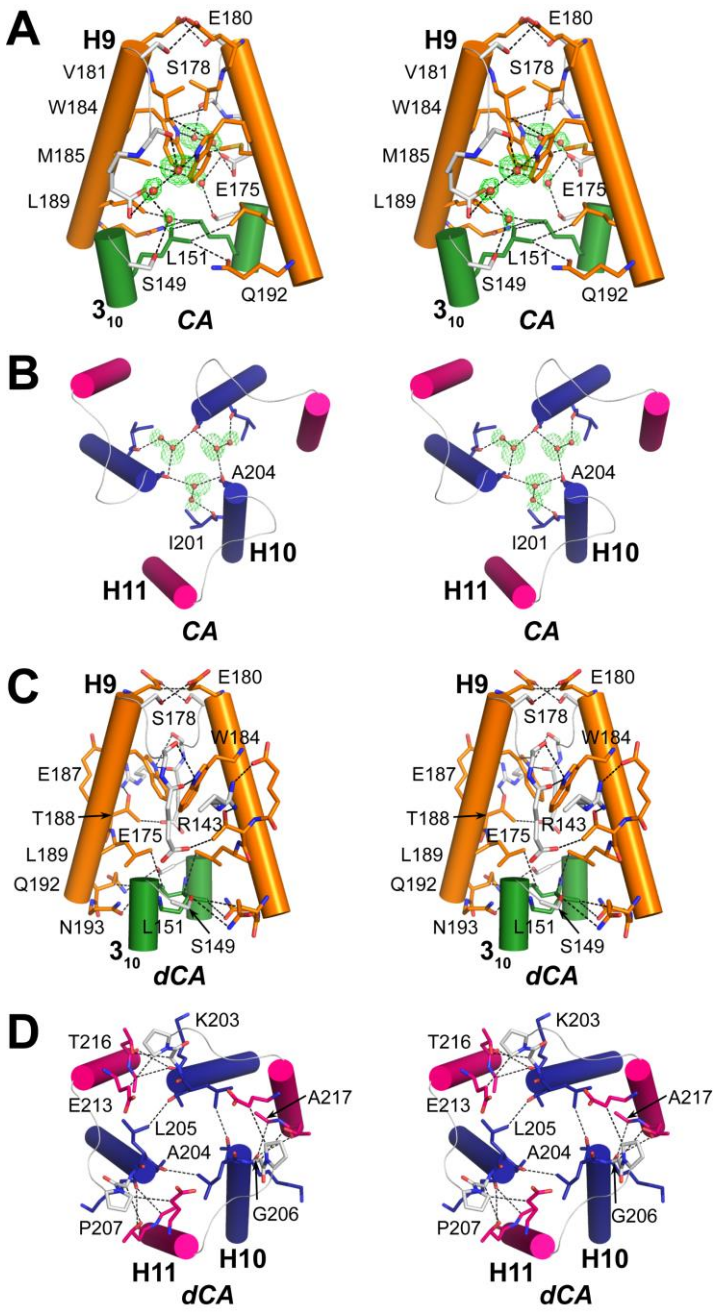


Figure 2-3. Inter-hexamer interactions at the 2-fold and 3-fold interfaces.

Stereo views of CA_{CTD} regions that are related by 2-fold (**A** and **C**) or 3-fold symmetry (**B** and **D**) from *CA* (**A** and **B**), and *dCA* (**C** and **D**). The 2-fold interface comprises helices H9 and 3_{10} in **A** and **C**. The 3-fold interface comprises helices H10 and H11 in **B** and **D**. Ordered water molecules (spheres) at the 2-fold (**A**) and 3-fold (**B**) interfaces of *CA* are modeled in 2.4 Å simulated annealing omit Fo–Fc electron density maps at σ=2.5 (green mesh). B-factors of refined waters were between 30-50 $Å^2$ and matched well with those of interacting atoms from S149, E175, Q176, W184, I201 and A204. Helices are shown in cartoon representation; black dashed lines between residues (in sticks) indicate that they are within ~4 Å. No water molecules are present at the 2-fold and 3-fold of *dCA* (**C** and **D**).

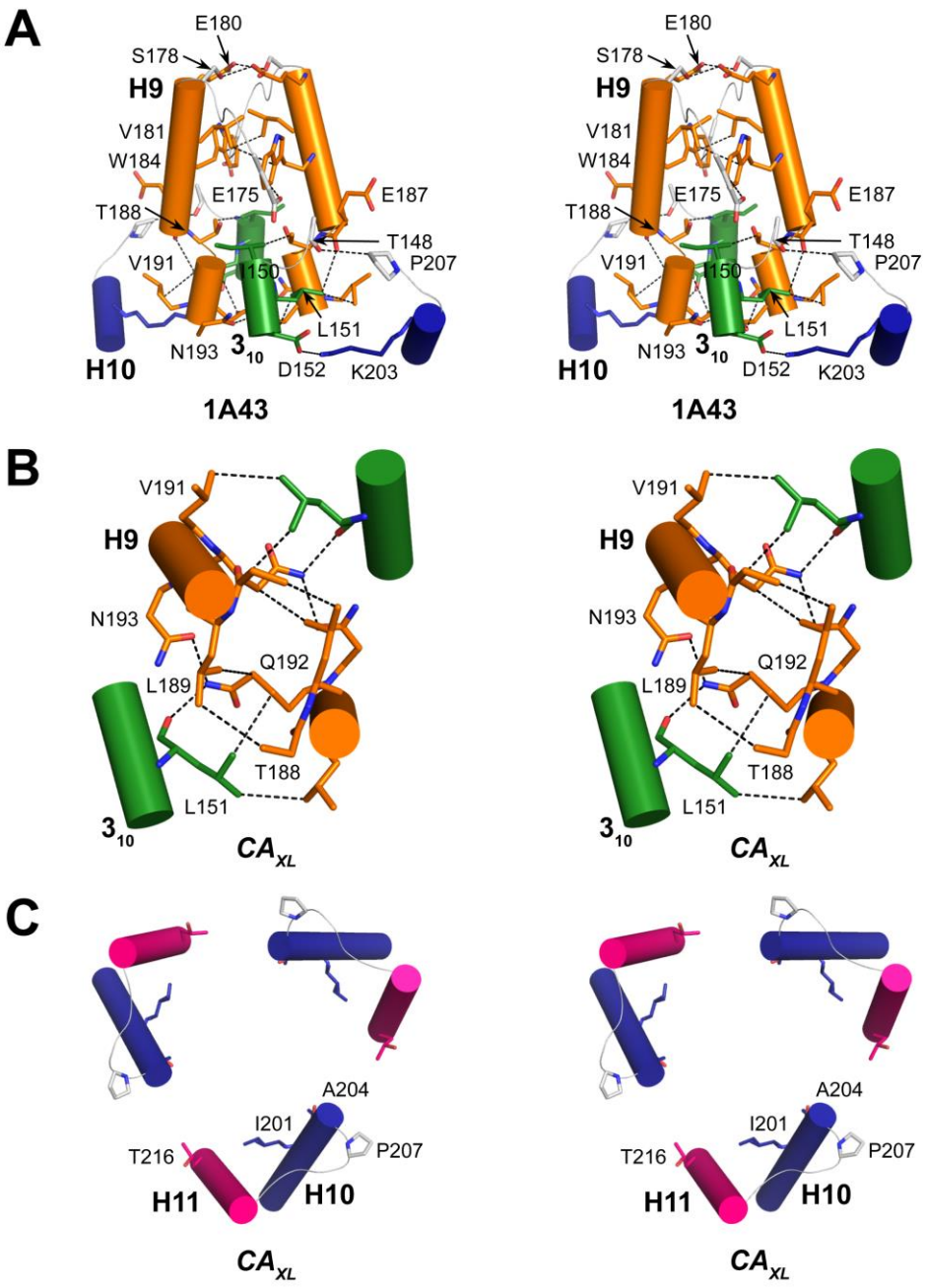


Figure 2-4. Interactions at the 2-fold and 3-fold interfaces of CA_{XL} and CA_{CTD} structures.

(**A**) Stereo view at the 2-fold interface between two CA_{CTD} subunits in the crystal structure of the CA_{CTD} fragment of CA (PDB ID: 1A43). Similar interactions were also observed in the related CA_{CTD} NMR structure (Table 2-2, and PDB ID: 2KOD). Stereo view of CA_{CTD} regions that are related by 2-fold (**B**) and 3-fold (**C**) from *CA_{XL}*. The 2-fold interface comprises helices H9, H10, and 3_{10} and the 3-fold helices H10 and H11. Black dashed lines indicate distances of up to ~4 Å.

Table 2-2. CA_{CTD}-CA_{CTD} interactions at the 2-fold interface and relevant mutagenesis studies[a].

CA	*dCA*	CA_{PF74}	CA_{XL}	1A43	2KOD[b]	2BUO	Location	Effect on virus	Possible function/contacts	Reference
	143 143						Helix 7	R143A: destabilization of capsid, reduction in the rate of capsid disassembly, reduction in virus replication	cofactor selective mutant	(*92, 97, 119*)
				148 148 148 148		148	Bend, loop	T148A: wild type-like behavior; T148I: reduction in infectivity, but production of viable CA mutants	linker between the CA_{NTD} and CA_{CTD}	(*96, 244*)
149[c]	149					149	Bend, loop	S149A: decrease in uncoating activity and core stability; S149C, S149G: reduction in infectivity, but production of viable CA mutants	linker between the CA_{NTD} and CA_{CTD}	(*96, 100, 244*)
				150			3_{10}-helix	I150A: destabilization of CA_{CTD}–CA_{CTD} association; I150V: reduction in infectivity, but production of viable CA mutants	linker between the CA_{NTD} and CA_{CTD}	(*96, 101*)
151 151 151	151 151 151 151		151 151 151	151 151 151 151	151	151 151 151	3_{10}-helix	L151A: production of noninfectious virions that assemble unstable cores with aberrant morphology, and synthesize almost no viral DNA, destabilizing mutation; L151I, L151Q: production of non-viable HIV-1 CA mutants	linker between the CA_{NTD} and CA_{CTD} inter-subunit contacts, critical for dimerization	(*91, 96, 101*)
151	151	151	151	151	151					
				152 152		152	3_{10}-helix	D152A: reduction in infectivity, no obvious changes in morphology of either immature or mature particles	periphery of the 2-fold interface, favorable	(*92*)

CA	*dCA*	*CA*$_{PF74}$	*CA*$_{XL}$	1A43	2KOD[b]	2BUO	Location	Effect on virus	Possible function/contacts	Reference
				154			Loop	R154A: prevention of dimerization	hydrophilic contacts periphery of the 2-fold interface	(*101*)
				173		173	Helix 8	R173I: production of non-viable HIV-1 CA mutants	periphery of the 2-fold interface	(*96*)
175[c]	175 175			175	175	175 175	Bend, loop	E175A: increase in homodimerization affinity	critical for dimerization	(*101*)
						176	Loop	Q176A: reduction in infectivity, no obvious changes in morphology of either immature or mature particles	periphery of the 2-fold interface, favorable hydrophilic contacts	(*92*)
176[c]	176	176				176 176				
					177	177	Loop	A177S: production of viable HIV-1 CA mutants	periphery of the 2-fold interface	(*96*)
178	178	178		178	178	178	Loop	S178A: decrease in uncoating activity and core stability, produces non-infectious virus; increase in homodimerization affinity	phosphorylation of this residue is essential for viral infectivity and uncoating	(*100, 101*)
180 180 180	180 180 180	180 180 180		180 180 180 180	180 180 180 180	180 180 180	Helix 9	E180A: reduction in infectivity, no obvious changes in morphology of either immature or mature particles; increase in homodimerization affinity	periphery of the 2-fold interface, favorable hydrophilic contacts	(*101*)
				181	181	181	Helix 9	V181A: destabilizing mutation	inter-subunit contacts critical for dimerization	(*101*)
181			181	181	181	181 181				

CA	*dCA*	*CA*$_{PF74}$	*CA*$_{XL}$	1A43	2KOD[b]	2BUO	Location	Effect on virus	Possible function/contacts	Reference
184[c]			184	184	184	184	Helix 9	W184A causes Gag and capsid assembly defects, production of noninfectious virus; reduction of CA dimerization and intermolecular Gag–Gag interactions *in vitro,* diminishment of immature particle production *in vivo* and abolishment of capsid dimerization; W184L, W184R: production of non-viable HIV-1 CA mutants	principle 2-fold interface of the capsid protein	(*75, 92, 93, 96*)
				184	184	184				
184	184	184			184	184				
	184			184	184					
184		184			184	184				
184		184			184	184				
184		184			184	184				
185		185			185	185	Helix 9	M185A causes Gag and capsid assembly defects, production of noninfectious virus; reduction of CA dimerization and intermolecular Gag–Gag interactions *in vitro,* diminishment of immature particle production *in vivo* and abolishment of capsid dimerization; M185I: reduction in infectivity, production of viable CA mutants	principle 2-fold interface of the capsid protein	(*75, 92, 96*)
						185				
	187			187			Helix 9	E187A: no effect on viral infectivity, increase in homodimerization affinity E187V: reduction in infectivity, production of viable CA mutants	favorable hydrophilic contacts	(*96, 133*)
				187						
	188			188			Helix 9, loop	T188A: destabilizing mutation	inter-subunit hydrophobic contacts	(*101*)
			188	188	188					
	188		188	188		188				
				188						
189	189		189		189	189	Helix 9			(*101, 254*)

CA	*dCA*	*CA*$_{PF74}$	*CA*$_{XL}$	1A43	2KOD[b]	2BUO	Location	Effect on virus	Possible function/contacts	Reference
			189					L189A: various degrees of defects in HIV-1 assembly and release, defects in Gag processing	Gag-membrane binding, multimerization, and folding inter-subunit contacts critical for dimerization	
			191	191			Helix 9	V191A: no significant effect on dimerization affinity	inter-subunit hydrophobic contacts	(*101*)
192	192	192	192	192	192		Helix 9	Q192A: reduction in infectivity, no obvious changes in morphology of either immature or mature particles; increase in homodimerization affinity	periphery of the 2-fold interface, favorable hydrophilic contacts	(*92, 98*)
	192		192	192						
			192	192						
	193		193	193			Loop	N193A: no significant effect on dimerization affinity, destabilizing mutation	favorable hydrophilic contacts	(*98, 101*)
			203	203			Turn	K203A: no obvious defects in assembly, maturation, or transmembrane gp41 and reverse transcriptase packaging, yet significant reduction in infectivity; alteration of the intrinsic stability of the viral capsid; K203C/Q219C: production of noninfectious particles	stabilization of the interface, periphery of the 2-fold/3-fold interfaces, favorable hydrophilic contacts	(*65, 77, 92, 93, 97, 99*)
				207			Turn	P207A: reduction of infectivity, particle production and presence of cones similar to wild-type virus; 207C/216C: spontaneous dimer- and trimerization of CA, reduction in infectivity	periphery of the 2-fold/3-fold interfaces	(*77, 92, 97, 99*)

CA	*dCA*	*CA*$_{PF74}$	*CA*$_{XL}$	1A43	2KOD[b]	2BUO	Location	Effect on virus	Possible function/contacts	Reference
						229	Loop			
						229				

[a] Residues within up to ~4 Å distance are shown in the same color.

[b] Additional interactions at the CA_{CTD}-CA_{CTD} 2-fold interface are reported in the NMR CA_{CTD} dimer structure (PDB ID: 2KOD) (M144/K203, M144/G206, M144/P207, Y145/E187, Y145/T188, Y145/V191, Y145/Q192, Y145/K199, Y145/K203, S146/E187, S146/T188). Residues M144, Y145, and S146 are part of the CA_{NTD} in the context of full-length CA proteins (*CA*, *dCA*, *CA*$_{PF74}$, and *CA*$_{XL}$). They participate in the extended hydrogen bonding network giving rise to the intra-hexamer CA_{NTD}-CA_{CTD} interface of the capsid.

[c] Residue is involved in water-mediated interactions.

Table 2-3. CA_{CTD}-CA_{CTD} interactions at the 3-fold interface and relevant mutagenesis studies[a].

CA	*dCA*	*CA_{PF74}*	Location	Effect on virus replication	Possible function/contacts	Reference
201[b]			Helix 10	I201A: various degrees of defects in HIV-1 assembly, release, and Gag processing; I201V/L: retains infectivity and capsid stability; I201D: production of non-infectious virions that contain unstable or abnormal cores	part of the central hydrophobic core; Gag-membrane binding, multimerization, and folding	(*65, 96, 254*)
	203 203		Turn	K203A: no obvious defects in assembly, maturation, or transmembrane gp41 and reverse transcriptase packaging, yet significant reduction in infectivity; alteration of the intrinsic stability of the viral capsid; K203R: decrease in replication fitness; K203C/Q219C: production of non-infectious particles	stabilization of the interface, periphery of the 2-fold/3-fold interfaces, favorable hydrophilic contacts	(*65, 77, 92, 93, 97, 99*)
204[b]		204 204	Turn	A204V/L: retains infectivity and capsid stability; A204D: production of noninfectious virions that contain unstable or abnormal cores; A204C: exhibits higher assembly efficiency, showing enhanced capsid stability and reduced infectivity; A204G: reduction in infectivity, production of viable CA mutants	part of the central hydrophobic core	(*65, 96*)
	204	204				
	205	205	Bend	L205V/I: retains infectivity and capsid stability; L205D: are non-infectious and contain unstable or abnormal cores	part of the central hydrophobic core	(*65*)
	206		Loop			
	207		Turn	P207A: reduction in infectivity, particle production and presence of cones similar to wild-type virus; P207C/T216C: spontaneous cross-linking resulted in dimer and trimer forms of CA, reduction in infectivity	periphery of the 2-fold/3-fold interfaces	(*77, 92, 97, 99*)
	213		Helix 11	E213A/Q: alteration of the intrinsic stability of the viral capsid and impairment of HIV-1 infectivity	stabilization of the 3-fold interface	(*65, 77*)
	216 216		Helix 11	T216A: reduction in infectivity, production of viable CA mutants; T216S: decrease in replication fitness;	periphery of the 3-fold interface	(*77, 94, 96, 99*)

CA	*dCA*	*CA*$_{PF74}$	Location	Effect on virus replication	Possible function/contacts	Reference
				P207C/T216C: spontaneous dimer- and trimerization of CA, reduction in infectivity		
	217		Helix 11	A217V: production of non-viable HIV-1 CA mutants		(*96*)

[a] Residues within up to ~4 Å distance are shown in the same color.
[b] Residue is involved in water-mediated interactions.

Cryo-EM/modeling studies proposed that inter-hexamer interactions are hydrophobic (*65, 76*). The higher resolution *CA* structure reveals details of these contacts (Figure 2-3, B, Table 2-3) that include hydrophilic, water-mediated interactions at the 2- and 3-fold interfaces (Figure 2-3, A and B). These waters engage in H-bond interactions with either side-chains of conserved residues (S149, E175, W184 at the 2-fold), or main-chain carbonyls (Q176 at the 2-fold; I201, A204 at the 3-fold) (Figure 2-3, A and B). Thus, the closest inter-subunit distance at the 3-fold of *CA* is ~6 Å *versus* ~15 Å in CA_{XL} (Figure 2-5, A and C). Inter-hexamer interface waters contribute to capsid stabilization. Of 450 assigned waters/hexamer (75/monomer), 30 are at the 2- and 3-fold interfaces (three waters for each of the six 2-folds and two waters for each of the six 3-folds/hexamer). As there are ~250 hexamers/capsid, thousands of water molecules at the inter-hexamer interfaces should significantly contribute to capsid stability.

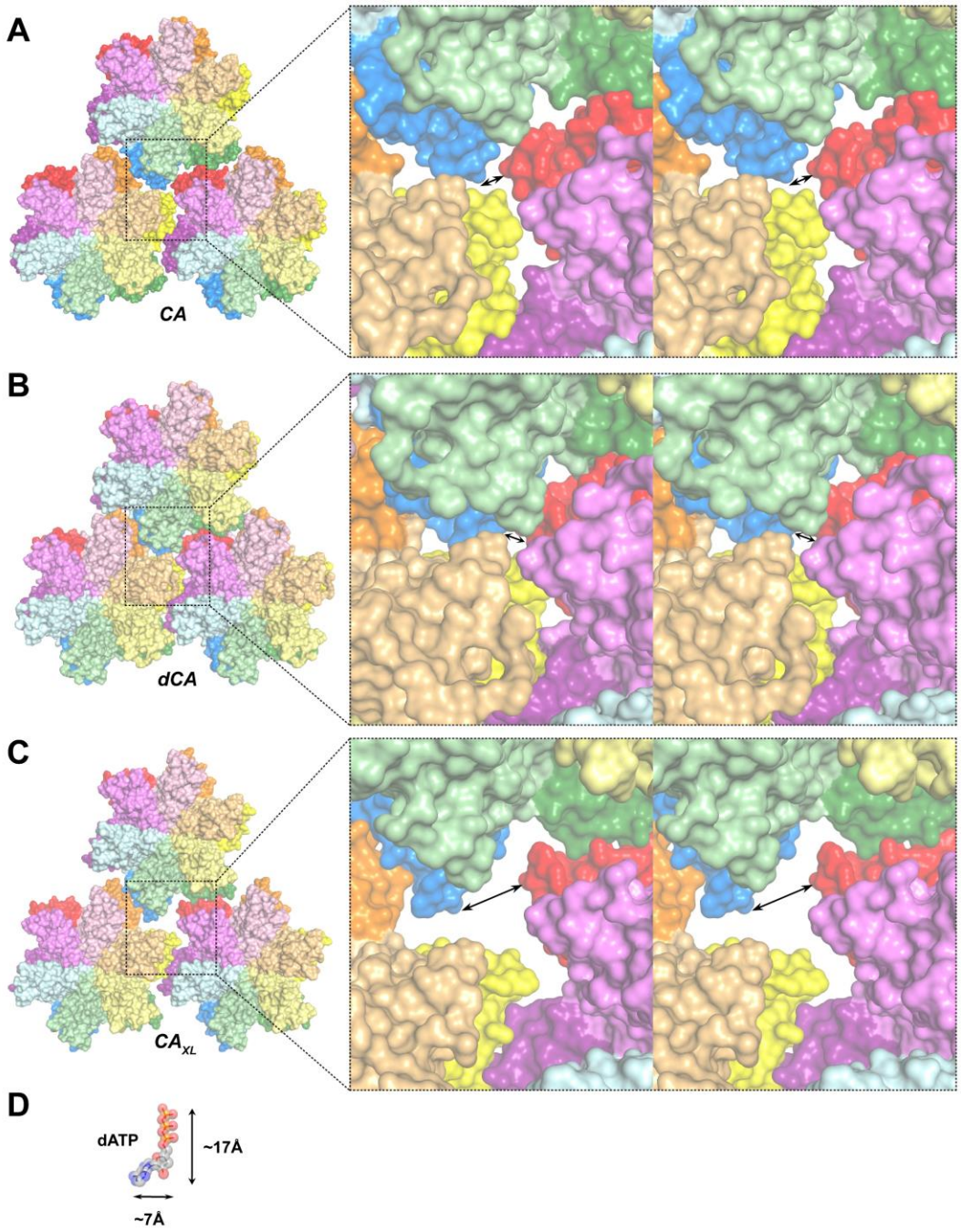


Figure 2-5. Crystal packing of HIV-1 CA in different crystal structures.

Neighboring hexamers in the lattices of *CA* (**A**), *dCA* (**B**), and *CA*$_{XL}$ (**C**) are shown in surface representation and also in enlarged stereo views. CA$_{NTD}$s and the corresponding CA$_{CTD}$s are colored by the same colors, with CA$_{CTD}$s being darker (light cyan *vs.* blue, light green *vs.* green, light yellow *vs.* yellow, light orange *vs.* orange, light pink *vs.* red, and violet *vs.* purple). (**D**) A nucleoside triphosphate (dATP) is shown in space-filling form as a reference relative to the gaps at the 3-fold interfaces. Arrows in insets of A, B, and C correspond to ~6 Å, ~3 Å, and ~15 Å.

To investigate the role of this hydration layer, we used controlled dehydration (*246*) that alters the water content of crystals. CA crystal dehydration (*dCA*) contracted the unit-cell dimensions by ~3-6 % (Table 2-1). Superposition of *CA* on *dCA* reveals conformational rearrangement in the relative orientation of CA_{NTD} and CA_{CTD}, imparted by a hinge-like motion with the linker region as a pivot point (Figure 2-6). These changes correlate with packing differences at the 2- and 3-fold (Figure 2-3, A *vs*. C, B vs. D; Figure 2-5, A *vs.* B). Hence, *dCA* hexamers (Figure 2-5, B) arrange tighter (~3 Å closest distance) than CA_{XL} (Figure 2-5, C) and *CA* (Figure 2-5, A) hexamers, creating up to 22 and 15 additional contacts in the *dCA* 2-and 3-fold (Table 2-2, Table 2-3) interfaces. Solvent-accessible area calculations reveal that *dCA* has ~150 % and ~200 % more buried surface than *CA* and CA_{XL} at the 2-fold, and ~400 % more than *CA* at the 3-fold (Table 2-4). The 3-fold interface interactions are between K203-T216, K203-A217 (main-chain atoms), and A204-L205, G206-T216, and P207-E213 (side-chains) (Table 2-3, Figure 2-3, D).

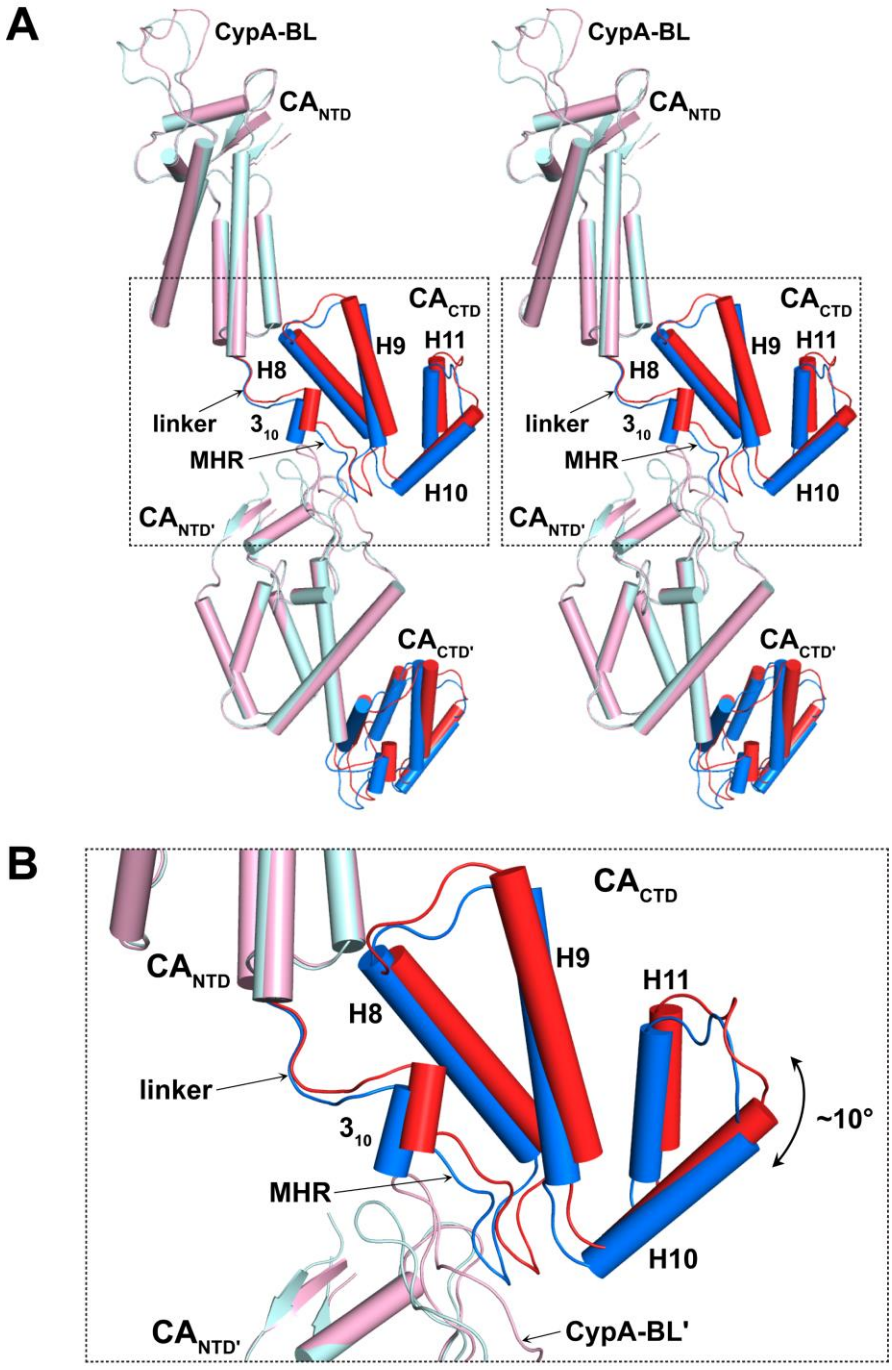


Figure 2-6. Inter-domain orientations and crystal contacts of CA and dCA structures.

(**A**) Stereo view of superposed *CA* (pink CA_{NTD} and red CA_{CTD}) and *dCA* (cyan CA_{NTD} and blue CA_{CTD}) subunits (alignment based on the CA_{NTD}s) from adjacent lattice layers (second lattice layer elements are marked with a prime symbol, *e.g.,* $CA_{NTD'}$). The crystal contacts involve the major homology region (MHR) from the CA in the top lattice layer and the CypA-BL from the CA in the bottom lattice layer. These crystal contacts together with changes in CA_{NTD}-CA_{CTD} orientation upon crystal dehydration are shown in (**B**).

Table 2-4. Buried surface area calculated for CA_{CTD}–CA_{CTD} interfaces in CA structures.

Structure	Buried surface area, Å^2	
	2-fold interface	3-fold interface
CA	906.4 (3.6 %)	277.5 (0.7 %)
dCA	1416.1 (5.5 %)	1161.3 (3.0 %)
CA_{PF74}	845.9 (3.4 %)	459.6 (1.2 %)
CA_{XL}	685.2 (2.9 %)	0 (0.0 %)
PDB ID: 1A43	1844.5 (18.5 %)	NA
PDB ID: 2KOD	1758.1 (13.6 %)	NA
PDB ID: 2BUO	997.8 (8.1 %)	NA

Sequence alignment (Figure 2-7) shows that almost all 2- and 3-fold residues are entirely conserved (205, 206, 213, 216; 207 is highly conserved). Consistent with our data, A204 substitutions result in non-infectious virions with unstable/abnormal cores (*65, 92*). We also observe variations in the CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} intra-hexamer contacts (Table 2-5) that alter H-bond networks and water-mediated or hydrophobic interactions (Figure 2-8, A). Mutagenesis studies also confirm the importance of interface residues for core morphology and stability, DNA synthesis, and infectivity (Table 2-2, Table 2-3) (*64, 75, 91-93, 96, 99, 133, 254*).

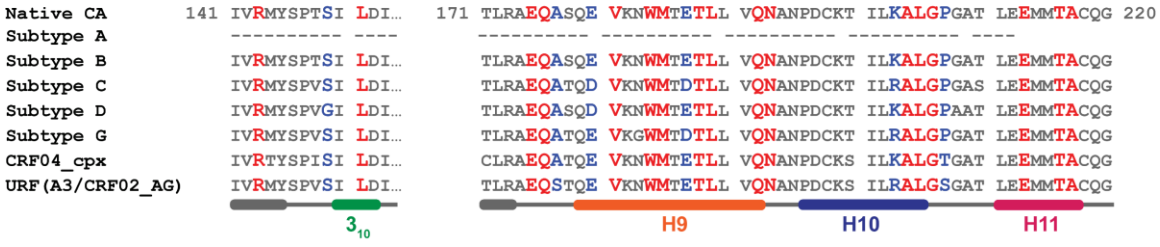


Figure 2-7. Sequence alignment of CA_{CTD} regions among HIV-1 group M subtypes.

The sequence used for crystallization experiments described here (*CA*, *dCA*, and *CA_{PF74}*) is labeled 'Native CA.' Fully conserved residues at the 2-fold and 3-fold interfaces of multiple HIV-1 subtypes (A, B, C, D, G, CRF04_cpx, URF(A3/CRF02_AG) are highlighted in red. Highly conserved residues at the 2-fold and 3-fold interfaces are in blue.

Table 2-5. Intra-hexamer CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} interactions.

Residues participating in hydrogen bonding networks are shown in red. Interacting residues present in all compared structures are highlighted in blue. All interactions are among domains from neighboring CA subunits in the same hexamer.

CA	*dCA*	*CA_{PF74}*	*CA_{XL}*
CA_{NTD}–CA_{NTD}			
V3 / H12	-	-	-
Q4 / V11	Q4 / V11	-[a]	Q4 / V11
Q4 / H_2O / H12[b]	Q4 / H12	-[a]	Q4 / H_2O / H12
R18 / P17	R18 / P17	-	-
R18 / R18[c]	R18 / R18[c]	R18 / R18[c]	R18 / R18[c]
T19 / P17	T19 / P17	T19 / P17	T19 / P17
K30 / E28	K30 / E28	-	K30 / H_2O / E28
-	-	-	K30 / H_2O / T58
-	-	-	K30 / H_2O / G60
E35 / N57	E35 / N57	E35 / N57	E35 / N57
E35 / H_2O / T58	E35 / T58	E35 / T58	E35 / H_2O / T58
E35 / G60	E35 / G60	E35 / G60	E35 / G60
P38 / H_2O / T54	-	P38 / T54	P38 / H_2O / T54
P38 / N57	P38 / N57	P38 / N57	P38 / N57
-	-	P38 / T58	-
M39 / V24	M39 / V24	M39 / V24	-
M39 / T58	M39 / T58	M39 / T58	M39 / T58
-	-	-	A42/ H_2O / I15
A42 / L20	A42 / L20	A42 / L20	A42 / L20
A42 / T54	A42 / T54	A42 / T54	A42 / T54
L43 / L20	-	-	L43 / L20
-	L43 / P17	-	-
-	-	E45 / H12	-
E45 / A14	E45 / A14	E45 / A14	C45 / C14[d]
E45 / I15	-	E45 / I15	-
E45 / Q50	E45 / Q50	E45 / H_2O / Q50	C45 / H_2O / Q50
E45 / D51	-	E45 / H_2O /D51	-
-	-	E45 / H_2O /T54	-
CA_{NTD}–CA_{CTD}			
-	-	-	R162 / H_2O / M144
R162 / Y145	R162 / Y145	R162 / Y145	R162 / H_2O / Y145
-	-	-	R162 / H_2O / S146
V165 / A64	V165 / A64	V165 / A64	V165 / A64
D166 / H62	D166 / H62	D166 / H62	D166 / H_2O / H62
D166 / H_2O / Q63	D166 / Q63	D166 / H_2O / Q63	D166 / H_2O / Q63
D166 / A64	D166 / A64	D166 / A64	D166 / A64

CA	*dCA*	*CA$_{PF74}$*	*CA$_{XL}$*
D166 / H_2O / A65	-	-	D166 / H_2O / A65
D166 / H_2O / Y145	-	-	D166 / H_2O / Y145
Y169 / Q63	Y169 / Q63	Y169 / Q63	Y169 / Q63
Y169 / Q67	Y169 / Q67	Y169 / Q67	Y169 / Q67
-	-	-	K170 / G60
-	K170 / Q63	-	K170 / Q63
R173 / N57	R173 / N57	R173 / N57	R173 / N57
R173 / V59	R173 / V59	R173 / V59	R173 / V59
-	-	-	R173 / H_2O / G60
R173 / E63	R173 / E63	R173 / E63	R173 / E63
-	Q179 / K70	-[e]	-
-	N183 / Q67	-	-
T210 / E71	T210 / E71	T210 / E71	T210 / E71
T210 / E75	-	T210 / E75	-
-	-	L211 / A64	-
L211 / Q67	L211 / Q67	L211 / Q67	-
L211 / M68	-	L211 / M68	L211 / M68
L211 / E71	L211 / E71	L211 / E71	L211 / E71
E212 / M68	-	E212 / M68	E212 / M68
E212 / H_2O / E71	E212 / E71	-	E212 / E71
E212 / K140	E212 / K140	E212 / K140	E212 / K140
E212 / H_2O / R143	-	E212 / H_2O / R143	-
E212 / M144	E212 / M144	E212 / M144	E212 / M144
M215 / A64	-	-	M215 / A64
M215 / M68	M215 / M68	M215 / M68	M215 / M68
-	M215 / M144	-	-
-	-	M215 / Y145	M215 / Y145
T216 / M144	T216 / M144	T216 / M144	-
	Q219 / M144	-	Q219 / M144

[a] Residue Q4 was not modeled in *CA$_{PF74}$*.
[b] X / H_2O / Y refers to water-mediated interaction between amino acids X and Y.
[c] Same interacting residue pairs, but different conformations
[d] Engineered crosslinked hexamer.
[e] Residue K70 in *CA$_{PF74}$* interacts with PF74.

A notable difference between *CA* and *dCA* involves interactions of R143 in CA_{NTD} with E175 or Q176 in CA_{CTD} helix H8 within the same subunit (Figure 2-8, A and B). The R143-Q176 interaction in *CA* is replaced with R143-E175 in *dCA*. This change repositions helix H9, leads to loss of the W184-bound water, and alters the H9-H9' 2-fold inter-

hexamer interactions (Figure 2-8, B). Alternative interactions of R143 with CA_{CTD} were observed in engineered unliganded (PDB ID: 3H4E) or ligand-complexed structures (PDB-ID: 4U0E), crystallized in orthorhombic (*64*) or hexagonal space groups, respectively (*119*), highlighting its functional importance, also supported by virological studies (*92, 119*).

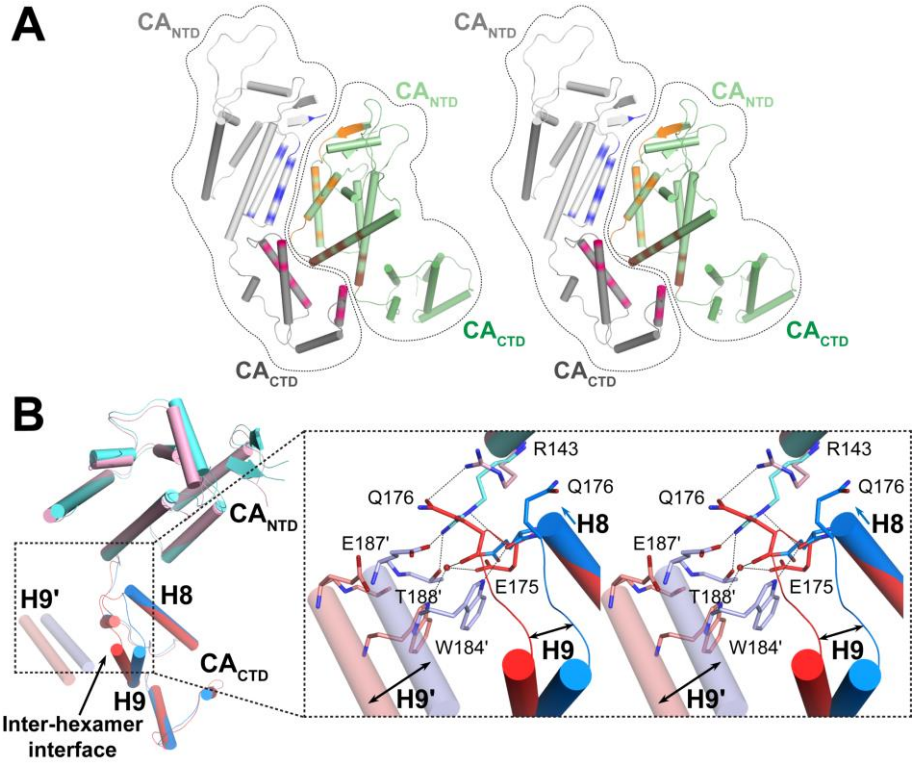


Figure 2-8. Changes at the intra-hexamer interfaces.

(**A**) Stereo view of intra-hexamer CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} inter-subunit interfaces. Two neighboring CA subunits are shown as cartoons outlined for clarity (CA_{NTD}s in lighter colors than the corresponding CA_{CTD}s). Sites of varying interactions among *CA*, *dCA*, CA_{XL}, and CA_{PF74} are marked in blue/orange for CA_{NTD}-CA_{NTD} interfaces or pink/brown for CA_{CTD}-CA_{NTD} interfaces. (**B**) Intra-subunit rearrangement linked to changes at the 2-fold interface. Enlarged stereo view of the boxed region shows shifts in the position of H9 helices in neighboring subunits (marked with prime symbols). Least squares superposition (residues 143–174, 192–219) of *dCA* (cyan CA_{NTD}, blue CA_{CTD}) on *CA* (pink CA_{NTD}, red CA_{CTD}). Crystal dehydration results in slight extension of helix H8 (small blue arrow), interaction of R143 with main-chain E175 carbonyl instead of Q176, and repositioning of helix H9 (black arrow). In *dCA*, W184' from the H9' helix (light blue for *dCA* and pink for *CA*) forms a hydrogen bond with main-chain E175 carbonyl from a neighboring subunit, whereas in *CA,* W184' interacts with Q176 and side-chain of E175 through water-mediated contacts. Moreover, in *dCA* R143 also interacts with E187' and T188' from the neighboring subunit, thus, becoming a part of the 2-fold interface. Black dashed lines connect residues or waters that interact through hydrogen bonds.

HIV-1 capsid structure in complex with PF74

Capsid plasticity may allow therapeutic intervention by stabilizing non-productive uncoating and assembly intermediates or disrupting the inter-hexamer hydration layer. Residues at inter-hexamer interfaces are conserved among clades (*92, 96, 255*) providing a high genetic fragility that makes CA an attractive antiviral target (*96*). Among >20 CA-targeting compounds, PF74 is the most potent antiviral (*110, 121, 132, 138, 152*). To address how PF74 alters inter-hexamer interactions affecting capsid stability and reverse transcription, we solved the crystal structure of PF74 with native CA (CA_{PF74}). PF74 binds across two CA monomers in CA_{PF74} similar to recent cross-linked structures (*119, 200*) (Figure 2-9, Figure 2-10, A), but with the indole moiety arranged differently than in $CA_{NTD\text{-}PF74}$ (Figure 2-9, C). In addition to the interactions with CA_{NTD} helices H3, H4, H5, and H7 observed in $CA_{NTD\text{-}PF74}$ *(121),* PF74 also contacts CA_{CTD} H8 and H9 from a neighboring subunit in $CA_{XL\text{-}PF74}$ and CA_{PF74} (*119, 120*). In $CA_{XL\text{-}PF74}$, H9 mutations W184A and M185A leave part of the H9-H9' 2-fold interface disordered and there are no interactions at the neighboring 3-fold region. These interactions are now observed in CA_{PF74}. We also observe differences in PF74 contacts with CA_{CTD} residues at or near dimerization helix H9 (Y169 and K182 in CA_{PF74}, *vs.* Q176, S178, and E179 in $CA_{XL\text{-}PF74}$) (Figure 2-9, B) (*119, 120*). Interestingly, PF74 binding results in subtle structural changes at the remote 3- and 2-fold regions (Figure 2-9, D, Figure 2-10, B, and C). These variations are reminiscent of allosteric changes in the CA_{CTD} complex with antiviral peptide CAI binding near the CA_{CTD}-CA_{CTD} interface (*132*). Specifically, we observe changes in hydrophobic and water-mediated interactions at the CA_{PF74} 2-fold affecting the buried surface area (Table 2-2Table 1-1 and Table 2-4). Also, changes of the CA_{PF74} H10 helices (rmsd ~1 Å with *CA*) lead to

their convergence at the 3-fold, with the distance between A204 main-chain oxygens changing from 5.5 to 3.6 Å. In turn, this change leads to a displacement of the water molecules seen in *CA* (Figure 2-9, D and Figure 2-10, C) and a 150 % increase in buried surface area at the 3-fold of CA_{PF74} (Table 2-3 and Table 2-4). Hence, the CA_{PF74} structure provides insights into the mechanism by which high concentrations of PF74 change intra- and inter-hexamer interactions and affect core stability (Table 2-2, Table 2-3, and Table 2-5).

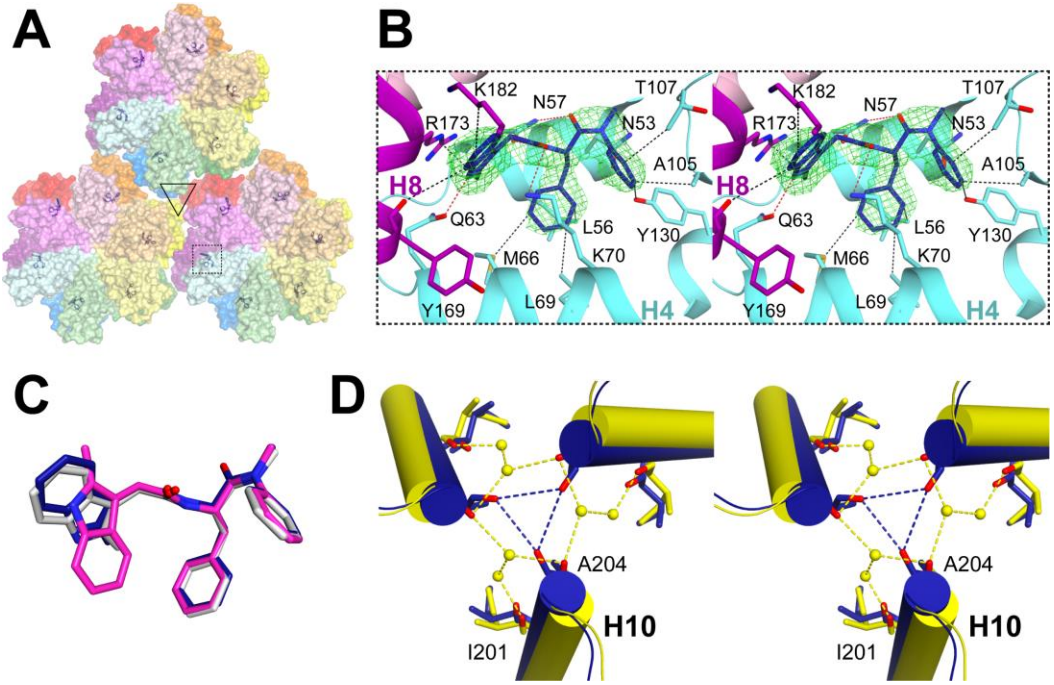


Figure 2-9. Effects of PF74 on HIV-1 CA structure.

(**A**) PF74 binding at the CA_{NTD}-CA_{CTD} interfaces of neighboring subunits within a CA hexamer. Top view of CA_{PF74} hexamers (side view is shown in Figure 2-10, A). There is one PF74 molecule bound to every CA subunit at sites that are distant to the 3-fold interface (black triangle). CA_{NTD}s and the corresponding CA_{CTD}s are colored by the same colors (light and dark, respectively). (**B**) Close-up stereo view of the PF74 binding site (small box in (**A**)). CA_{NTD} and a CA_{CTD} of a neighboring subunit bind PF74 (magenta and cyan ribbons). H9 is omitted for clarity. PF74 is modeled in a 2.7 Å simulated annealing omit map (σ=2.5). (**C**) PF74 conformations in CA_{PF74} (blue), $CA_{XL\text{-}PF74}$ (PDB-ID: 4U0E; PF74 in grey), and $CA_{NTD\text{-}PF74}$ (PDB-ID: 2XDE; PF74 in magenta). PF74 in 4QNB is almost identical to 4U0E and omitted for clarity. (**D**) Stereo view of the 3-fold interface of CA_{PF74} superposed onto *CA* (aligned on residues 1-219). Helices H10 of CA_{PF74} and *CA* are in yellow and blue. *CA* waters are shown as yellow spheres; no waters were present in CA_{PF74}. Black dashed lines indicate interactions within ~4 Å and red dashed lines indicate H-bond interactions.

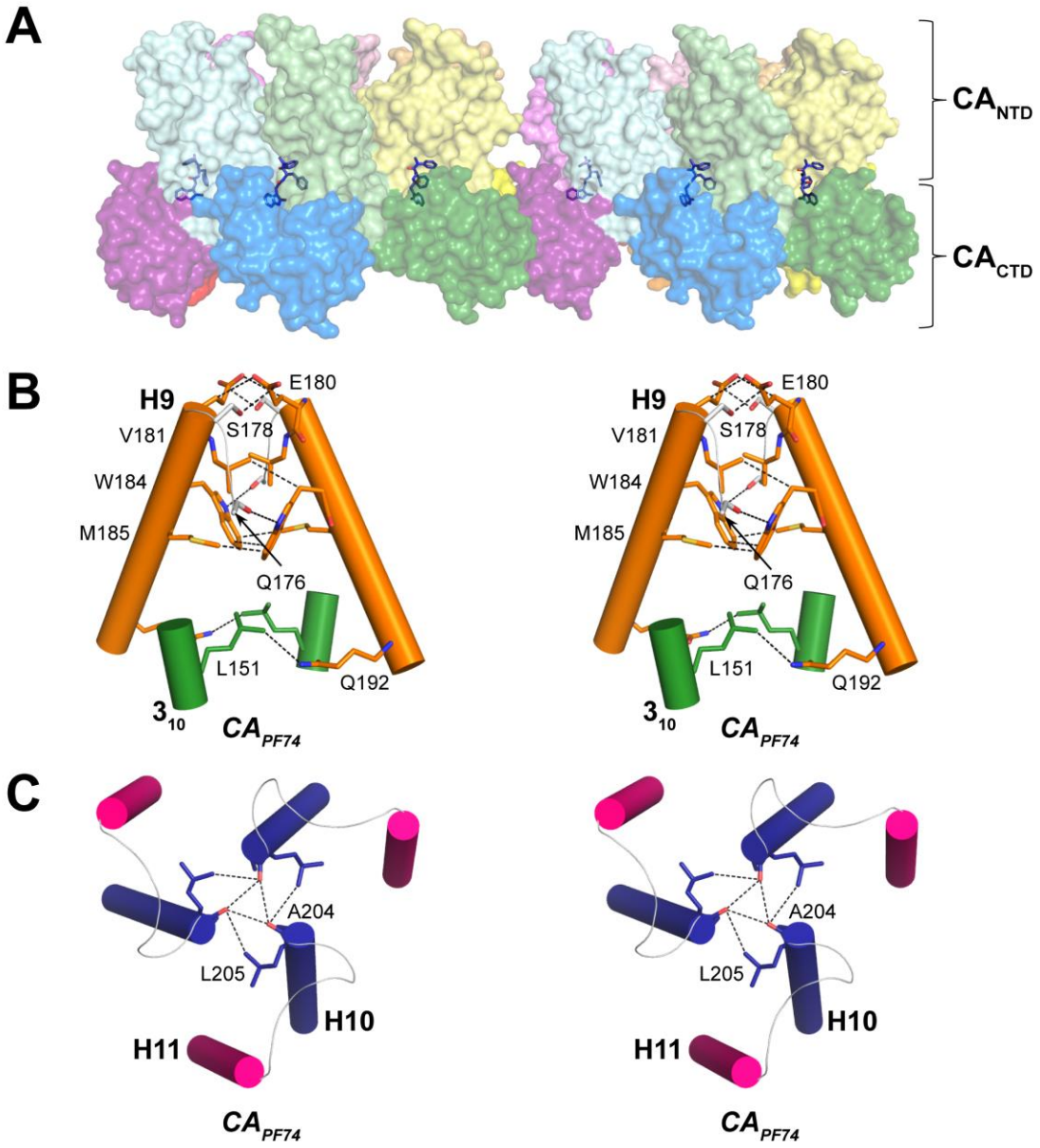


Figure 2-10. Binding site of PF74 and atomic interactions at the inter-hexamer interfaces of *CA_{PF74}* structure.

(**A**) Binding of PF74 at the CA_{NTD}-CA_{CTD} interfaces of neighboring monomers within a CA hexamer: top view of *CA_{PF74}* hexamers shown in Figure 2-9, A. Stereo views of CA_{CTD} regions that are related by 2-fold (**B**) or 3-fold (**C**) symmetry in *CA_{PF74}*. The 2-fold interface comprises helices H9 and 3_{10} and the 3-fold helices H10 and H11. Black dashed lines indicate distances of up to ~4 Å.

Implications for viral life cycle

Extensive structural heterogeneity has been reported for CA monomers in the Bovine Leukemia Virus CA hexamers (*256*). Similarly, comparison of the present *CA*, *dCA*, and CA_{PF74} HIV-1 structures also suggest a CA conformational plasticity, indicating that capsid structural variability is a common feature among multiple retroviruses. Variability in inter- and intra-hexamer contacts is also proposed in HIV-1 capsid molecular models (*65, 85*).

The nature of CA structural variability goes beyond mere side-chain rearrangement. A likely key structural determinant essential for the stabilization of variable interfaces is the presence of structured waters observed at strategic regions (*64*), including the 2- and 3-fold interfaces (Figure 2-3, A and B). This hydration layer could function as an extension of CA structure, contributing to surface complementarity among flexible CAs. It is likely adaptable and facilitates nearly isoenergetic structural rearrangements contributing to quasi-equivalent structural variability.

Similar to previous reports (*64*) we observe changes in the relative orientation of CA_{NTD} and CA_{CTD} (~10º rotation in Figure 2-6) that may cause a tilt among neighboring hexamers and contribute to the surface curvature of the capsid. Different inter-hexamer tilt angles would be anticipated at the narrow and broad ends of the capsid or between laterally or longitudinally positioned hexamers. Given the asymmetric nature of capsid, it is likely that no two CAs are identical in an HIV-1 capsid (*65, 85*). Hence, CA plasticity allows a wide range of conformations that contribute to the structural variability of the HIV-1 asymmetric core.

CA pliability allows interactions of capsids with multiple host factors. For HIV-1, these include TRIM5α, CPSF6, MxB, cyclophilin, Nup153, and Nup358 (*105, 257*). Such diverse interactions afford a functional versatility reminiscent of a Swiss Army knife. Moreover, the variability in subunit distances at the 3-fold (~3-15 Å, Figure 2-5) may affect permeability to deoxynucleoside triphosphates (footprint diameter ≤10 Å). This may be relevant to reverse transcription, although nucleotide trafficking may also occur through imperfections in the malleable core structure. Extensive variability may become an Achilles heel for HIV by providing opportunities for pocket targeting, core destabilization, or stabilization of non-productive structural intermediates.

Chapter 3. Characterization of HIV-1 capsid interactions with host cell factors and pharmacological ligands

Note: This chapter is based on a manuscript currently in preparation. The contributions of co-authors are described in the "Footnotes."

A. Introduction

The HIV-1 capsid protein (CA) performs multiple roles in viral replication (*61*). Synthesized as a central domain of Gag polyprotein, CA facilitates virus assembly into the spherical, immature viral particle. Upon proteolytic maturation of Gag, released CA proteins reassemble into the mature capsid, or core, which encapsulates the viral genome and associated enzymes. CA contains highly helical N-terminal (CA_{NTD}) and C-terminal (CA_{CTD}) domains connected by a flexible inter-domain region. The core is organized in a primarily hexagonal lattice of ~250 CA hexamers interrupted by 12 CA pentamers that help close the capsid at vertices (*83, 84*). Each CA hexamer and pentamer comprises CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} intra-hexamer contacts, and the hexagonal lattice is formed by 2-fold and 3-fold CA_{CTD}-CA_{CTD} inter-hexamer interactions (*64, 65, 77, 85, 95*).

Following infection and fusion of the viral and cellular membranes, the capsid core disassembles through a controlled process known as uncoating, which seems to be coordinated and linked to reverse transcription (*92, 96, 97*). There is an increasing body of evidence that the CA structure and its interactions with host proteins are critical for reverse transcription. Moreover, CA is genetically and functionally implicated in nuclear import of the provirus. It mediates interactions with cellular transport factors, including cleavage and polyadenylation specific factor 6 (CPSF6), which affects the capsid core stability and

cloaks the DNA reverse transcription product from exposure to DNA-sensing host factors that can induce antiviral immune response (*118, 258*). CA also interacts with an increasing number of host factors – including TNPO3, Nup358/RanBP2, Nup153, MxB, and Cyclophilin A – affecting the efficiency of viral DNA nuclear entry and the site of integration (*129, 161, 162, 189, 190, 193, 209, 210, 220, 235-237, 257, 259-264*).

Structures of CA_{NTD} (*118*) or engineered CA hexamer (*119, 200*) in complex with peptides from CPSF6 or Nup153 have provided valuable information on CA-host factor interactions, but they lack information on inter-hexamer interfaces. The presence of only small peptides from the host factors also significantly limits our understanding of their effects on uncoating, reverse transcription, and assembly. Hence, despite extensive structural work, the details of CA interactions with host factors that regulate the HIV life cycle, as well as the structural basis of CA stability that determines virus infectivity, are not well understood.

Given its involvement in multiple critical processes, CA is an attractive therapeutic target (*98, 109, 265-267*). A number of antivirals are known to bind the CA_{NTD} or CA_{CTD} domains (*98, 110, 121, 131, 132, 138, 152, 268*). Among them, PF-3450074 (PF74) blocks both the assembly and the uncoating steps of the viral life cycle (*121, 138*). It accelerates capsid dissociation in cells, inhibits reverse transcription, prevents CPSF6 and Nup153 binding to the HIV-1 core (*119, 121, 123, 125, 200*), and enhances the rate of CA multimerization *in vitro*. PF74 binds helices H3, H4, H5, H7, and H8 at the CA_{NTD}-CA_{CTD} intra-hexamer interface (*119, 200*). NYAD6 disrupts the formation of mature-like particles (*138*). It is predicted to bind helices H8, H9, and H11 of the CA_{CTD}. In the Sarafianos laboratory, a chemical library of compounds was screened, and a small molecule, 18E8,

which interferes with CA multimerization was identified. It showed broad antiretroviral activity in cell-based assays due to binding to HIV-1 CA. This was revealed by experiments for the selection of drug resistance, in which the A105T mutation in CA conferred resistance to the compound. Time-of-inhibitor addition experiments showed that 18E8 targets an early step in the HIV replication cycle, after reverse transcription and before integration. Hence, CA structure provides multiple opportunities for specific targeting of various steps of the life cycle and can offer novel strategies for therapeutic intervention.

Here, we have solved crystal structures of native CA in complex with specific peptides of the CPSF6 and Nup153 proteins. These structures confirm that cellular factors make interactions both within and between CA subunits, encompassing part of the CA_{NTD}-CA_{CTD} intra-hexamer interface. Additionally, binding of CPSF6 and Nup153 results in multiple subtle structural changes at the 2-fold and 3-fold inter-hexamer interfaces, which may affect stability and integrity of the HIV-1 core. Moreover, we have characterized the effect of pharmacological ligands (PF74, NYAD6, 18E8) that have antiviral properties through binding their CA target on its assembly. All compounds significantly alter the morphology of CA assemblies, producing short tubes, cones, sheets, wide tubes or spheres. Hence, this study helps to unveil further the relation between interactions that control core stability and multiple biological events.

B. Materials and methods

Production of HIV-1 CA, peptides, and compounds

HIV-1 CA was cloned without fusion tags in the pET11a plasmid (*268*). Protein was expressed in *E. coli* BL21(DE3)RIL as previously described and purified by

ammonium sulfate precipitation followed by anion exchange chromatography (*245, 268*). $CPSF6_{313-324}$ 12-mer peptide (PVLFPGQPFGQP) from the host cell factor CPSF6 was synthesized by Biomatik USA (https://www.biomatik.com/). $Nup153_{1407-1423}$ 17-mer peptide (TNNSPSGVFTFGANSST) from the host cell factor Nup153 was synthesized by Biomatik USA (https://www.biomatik.com/). PF74 was synthesized at the Chemical Biology Laboratory, NCI, as previously reported (*118*). NYAD6 was synthesized at the Laboratory of Molecular Modeling & Drug Design, Lindsley F. Kimball Research Institute of the New York Blood Center, as previously reported (*138*). 18E8 was purchased from Maybridge (http://www.maybridge.com/).

Crystallization of HIV-1 CA

CA crystals the grew at 18 °C in hanging drops, containing ~5 mg/ml of protein, 6-9 % PEG 3350, 2-6 % glycerol, sodium iodide, and sodium cacodylate. Hexagonal plate-like crystals appeared after five days, and crystal growth was complete in over two weeks.

HIV-1 CA complexes with host cell factors

Native CA crystals, grown as described above, were soaked with CPSF6 and Nup153 peptides (final concentration 1 mM) for approximately 24 h, briefly dipped in paraffin oil for cryoprotection and cryo-cooled in liquid nitrogen.

Data collection and structure determination

Data were collected on a Dectris Eiger-16m detector at the Advanced Photon Source (APS) beamline 23 ID-B, Argonne National Laboratory, and on a CMOS detector at Advanced Light Source (ALS) beamline 4.2.2, Lawrence Berkeley National Laboratory.

Datasets were processed using XDS (*247*). The data were examined for the presence of systematic absences. However, no characteristic patterns were observed. Thus, the crystals were indexed in hexagonal space group P6 with one CA molecule per asymmetric unit. No twinning was present, as determined by POINTLESS (*249*) or XTRIAGE (*250*). Space group and twinning were also verified in ZANUDA (*251*). The phase problem of CA was solved using molecular replacement, with the native CA (PDB ID: 4XFX) as a starting model. Initial phases were solved *via* PHASER (*251*). Several rounds of iterative model building and refinement were carried out using Coot (*253*) and PHENIX (*250*), REFMAC (*251, 252*), or PDBREDO (http://www.cmbi.ru.nl/pdb_redo/), respectively. Structure validation of final models was performed with MOLPROBITY (http://molprobity.biochem.duke.edu/). Accessible and buried surface areas were calculated using PISA (*251*). The figures showing structural information were generated in PyMOL (http://www.pymol.org/).

Morphology studies using TEM

Capsid protein assembly either by itself (150 µM) or in the presence of dimethyl sulfoxide (DMSO, 4 %) or pharmacological ligand (PF74, NYAD6, 18E8; 400 µM) was triggered by high salt concentration (final concentration 1 M) in buffer containing 25 mM Tris, pH 8.1. The samples were incubated for 1 h at 37 °C and prepared for transmission electron microscopy (TEM). Each sample (5 µl) was adsorbed during 5 min on the grids coated with colloidal carbon and made hydrophilic by glow-discharge (Pelco Easiglow Glow Discharge) for 45 sec. Excess fluid was removed, and the grids were washed with water for 2 min and fixed for 5 min in a drop of 2 % uranyl acetate. The dried grids were

loaded into a sample-holder before visualization with a Jeol 1400 transmission electron microscope, Electron Microscopy Core, University of Missouri, with a magnification of 20,000× or 40,000×.

C. Results and discussion

Crystallographic analysis of CA in complex with CPSF6 and Nup153 peptides

To understand how host cell factors CPSF6 and Nup153 interact with native CA we determined crystal structures with the peptide ligands. It has been previously shown that an FG-containing (phenylalalnine-glycine) 15-mer peptide from CPSF6 (*118, 119, 200*) or 17-mer from Nup153 is sufficient for binding to HIV-1 CA (*119, 235, 237*). The two peptides ($CPSF6_{313–324}$ with sequence PVLFPGQPFGQP and $Nup153_{1407–1423}$ with sequence TNNSPSGVFTFGANSST) were synthesized and used for crystallographic studies.

We have obtained crystal structures for both complexes, CA with $CPSF6_{313–324}$ (CA_{CPSF6}) and CA with $Nup153_{1407–1423}$ (CA_{Nup153}), in hexagonal space group (P6) (Table 3-1, Figure 3-1). The interpretable electron density is observed for both CPSF6 or Nup153 (Figure 3-1, C, F). Both peptides bind across two CA monomers in a hexamer (Figure 3-1, A, Figure 3-2, A), which is similar to the binding mode observed in the structures of crosslinked CA hexamer in complex with $CPSF6_{313–327}$ (PDB IDs: 4U0A and 4WYM; labeled $CA_{XL\text{-}CPSF6}$) or $Nup153_{1407–1423}$ (PDB IDs: 4U0C; labeled $CA_{XL\text{-}Nup153}$) (Figure 3-2, B, C) (*119, 200*). Collectively, the structures confirm that both binding pockets are part of a large protein-protein interface that is present only in assembled CA.

Table 3-1. Summary of data collection and refinement statistics.

	CA_{CPSF6}	CA_{Nup153}
Data collection		
X-ray source	APS 23 ID-B	ALS 4.2.2
Software	XDS	XDS
Space group	P6	P6
a, b, c (Å)	92.7 92.7 58.0	92.5 92.5 58.3
α, β, γ (°)	90.0 90.0 120.0	90.0 90.0 120.0
ASU content	1	1
Wavelength (Å)	1.033203	1.000031
Resolution range (Å)[a]	47.0–2.5 (2.6–2.5)	36.2–2.4 (2.5–2.4)
R_{merge}	0.063 (0.700)	0.065 (0.689)
R_{meas}	0.069 (0.774)	0.068 (0.723)
R_{pim}	0.029 (0.326)	0.021 (0.220)
$\langle I/\sigma I \rangle$	18.1 (2.7)	24.7 (3.3)
$CC_{1/2}$ (%)	99.8 (58.3)	99.9 (90.1)
Completeness (%)	99.9 (100)	97.2 (100)
Redundancy	5.6 (5.5)	10.6 (10.8)
Mosaicity	0.13	0.17
Refinement		
Resolution (Å)	47.0–2.5	36.2–2.4
No. total reflections	55693	115488
No. unique reflections	9956	10941
No. test reflections[b]	598	630
R_{work} / R_{free}	24.2 / 27.3	23.4 / 27.4
No. atoms	1835	1806
Protein	1794	1742
Ligand/Ion	10	9
Water	31	55
Wilson B-factor (Å^2)	47.2	47.5
Average B-factors (Å^2)	75.2	71.4
Protein	75.3	71.7
Ligand/Ion	96.3	83.1
Water	62.4	59.8
RMS deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.53	0.54
MolProbity Statistics[c]		
All atom clash score	3.34	3.73
Rotamer outliers (%)	0	0
Cβ deviations >0.25 Å	0	0
Ramachandran[c]		
Favored region (%)	98	99
Outliers (%)	0	0
PDB accession code	6AY9	6AYA

[a] Values in parentheses are for highest-resolution shell; [b] random selection; [c] values obtained from MOLPROBITY.

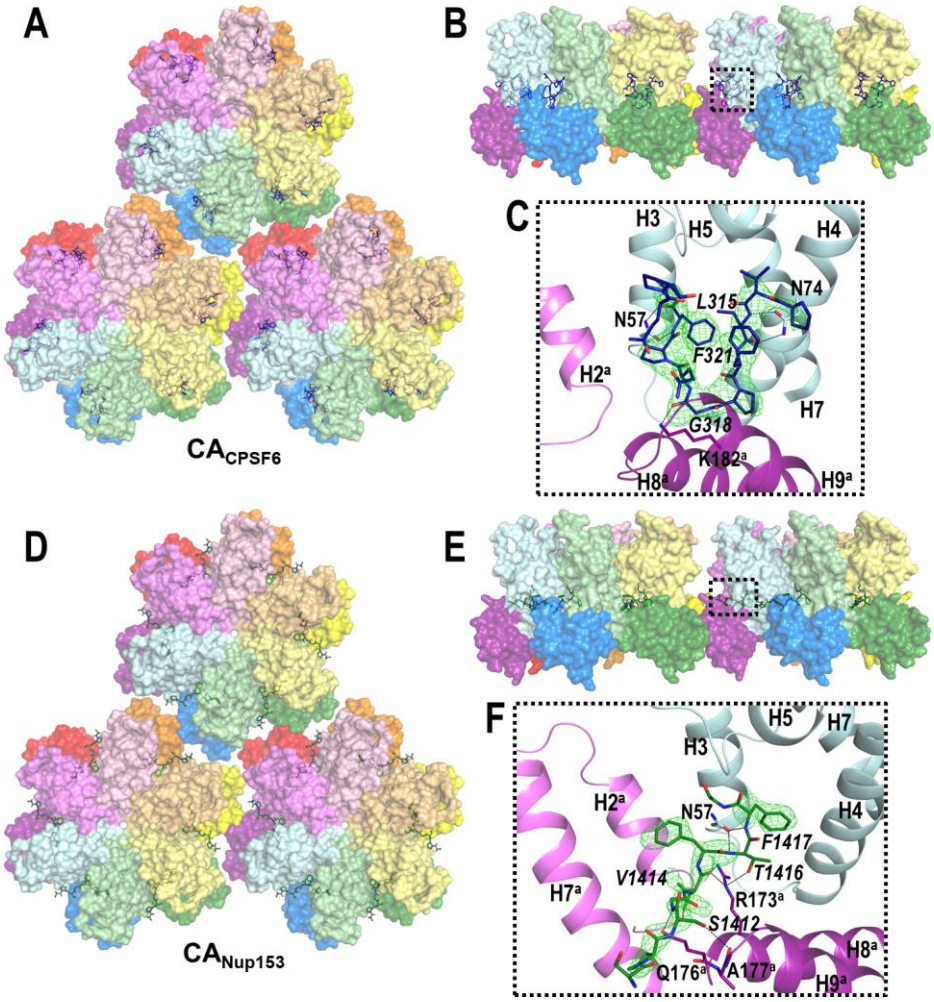


Figure 3-1. Binding sites of CPSF6 and Nup153 peptides.

(**A-C**) CPSF6 (blue sticks) binds at the CA_{NTD}-CA_{CTD} interfaces of neighboring subunits within CA hexamer. (**A**) Top view of CA_{CPSF6} hexamers, a side view is shown in (**B**). (**C**) Close-up view of the CPSF6 binding site (small box in (**B**)) with CA_{NTD}s and CA_{CTD} shown in ribbons. CPSF6 is modeled in a 2.5 Å simulated annealing composite omit map (σ=2.5). (**D-F**) Nup153 (green sticks) binds at the CA_{NTD}-CA_{CTD} interfaces of neighboring CA subunits. (**D**) Top view of CA_{Nup153} hexamers, a side view is shown in (**E**). (**F**) Close-up view of the Nup153 binding site (small box in (**E**)) with CA_{NTD}s and CA_{CTD}s shown in ribbons. Nup153 is modeled in a 2.4 Å simulated annealing composite omit map (σ=2.5). CA_{NTD}s and the corresponding CA_{CTD}s are colored by the same colors (light and dark, respectively). Black dashed lines indicate H-bond interactions with involved residues shown in sticks and labeled. Principal host factor residues are labeled in italic text.

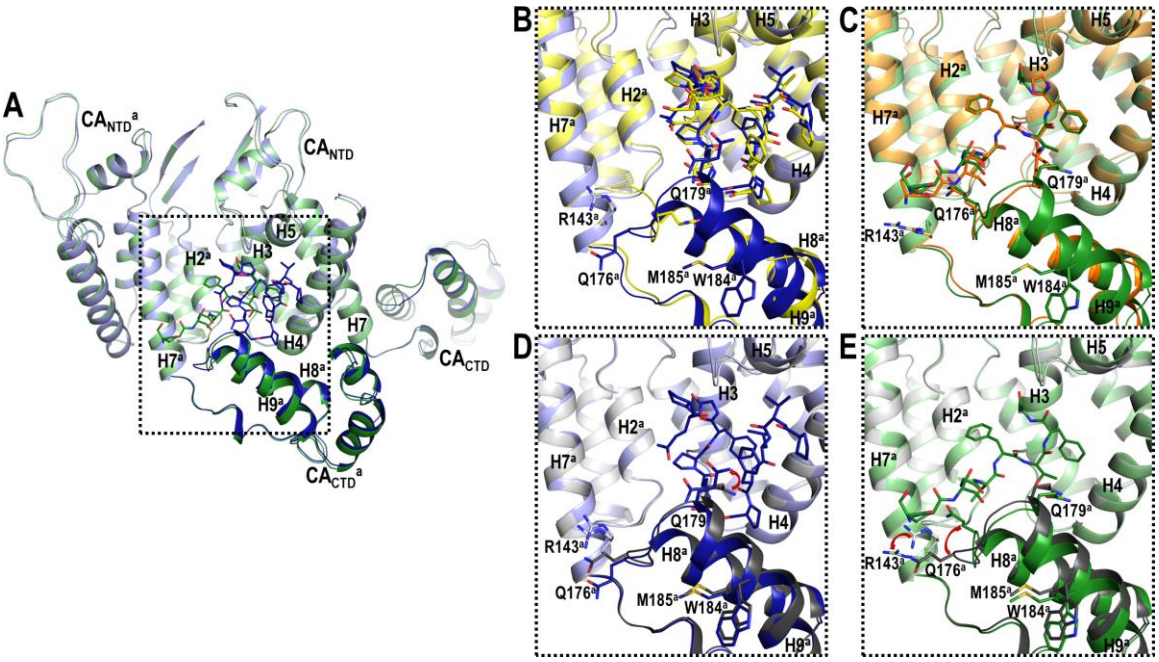


Figure 3-2. CPSF6 and Nup153 bind a multi-subunit interface in a hexamer.

(**A**) Two monomers from the CA_{CPSF6} structure are shown in light blue (CA_{NTD}s), and blue (CA_{CTD}s) superposed with the CA_{Nup153} in light green (CA_{NTD}s) and green (CA_{CTD}s). The helices that comprise the complete binding site are labeled. (**B**) An enlarged view of two monomers from the CA_{CPSF6} superposed with the corresponding ones in the $CA_{XL\text{-}CPSF6}$ (CA_{NTD}s in light yellow and CA_{CTD}s in yellow). (**C**) An enlarged view of two monomers from the CA_{Nup153} superposed with the corresponding ones in the $CA_{XL\text{-}Nup153}$ (CA_{NTD}s in light orange and CA_{CTD}s in orange). (**D**) Two monomers from the CA_{CPSF6} superposed with native CA (PDB ID: 4XFX; CA_{NTD}s in white and CA_{CTD}s in gray). (**E**) Two monomers from the CA_{Nup153} superposed with native CA. The side-chains of residues R143, Q176, Q179, W184, and M185 are shown in sticks. The conformational changes of the specific residues are highlighted by the red arrows.

The complete binding site is formed by CA_{NTD} helices H3, H4, H5 and H7 of one CA monomer (referred to as the 'first site') and $CA_{NTD}{}^{a}$ helices $H2^{a}$ and $H7^{a}$ and $CA_{CTD}{}^{a}$ helices $H8^{a}$ and $H9^{a}$ from an adjacent monomer (referred to as the 'second site') (Figure 3-2, A). As observed in the $CA_{XL\text{-}CPSF6}$ and $CA_{XL\text{-}Nup153}$ (*118*), CPSF6 forms a more compact, almost cyclized structure, while Nup153 (residues 1410–1417) forms an extended linear conformation that runs across the face of helices $H2^{a}$ and $H7^{a}$ (Figure 3-3, A-C). The CPSF6 peptide in the CA_{CPSF6} forms similar interactions as the $CA_{XL\text{-}CPSF6}$ within both binding sites. However, Q179 interacts with residue *F316* of CPSF6 instead of

forming hydrogen bonds as observed in $CA_{XL\text{-}CPSF6}$ (Figure 3-2, B). Hydrogen bonds (H-bonds) are found between CA residues N57, N74, K182[a] and CPSF6 residues *F321*, *L315*, and *G318*, respectively (Figure 3-1, C).

The Nup153 peptide in the CA_{Nup153} structure forms subtly different interactions when compared to $CA_{XL\text{-}Nup153}$ (*119*). It interacts with CA residues N53, L56, N57, M66, K70, G106, T107, and Y130 within the first binding site and residues P34[a], I37[a], P38[a], S41[a], N139[a], R143[a], R173[a], Q176[a] and A177[a] within the second binding site (Figure 3-2, C). The following H-bonds are observed between CA residues N57, R173[a], Q176[a], and A177[a] and Nup153 residues *T1416*, *F1417*, *V1414*, *T1416*, and *S1412* (Figure 3-1, F). The observed differences may be explained by incorporation of CA_{CTD} helices H8 and H9 into the host factor binding interface, a region that is structurally dynamic in the context of crosslinked CA hexamers.

Comparison of CA_{CPSF6} and CA_{Nup153} to unliganded native CA revealed that the overall fold of the structures is essentially the same (rmsd 0.4084 and 0.4301 Å) (Figure 3-2, D, E). Notably, in the CA_{CPSF6}, residue Q179[a] assumes different conformation than in the uncomplexed CA to avoid a steric clash with *F316* of the CPSF6 (Figure 3-2, D), while in the CA_{Nup153} R143[a] and Q176[a] rearrange to accommodate Nup153 (Figure 3-2, E).

The host factor interface incorporates helix H9 of the CA_{CTD}, which is part of the 2-fold inter-hexamer interface. In $CA_{XL\text{-}CPSF6}$ and $CA_{XL\text{-}Nup153}$, H9 mutations W184A and M185A leave part of the H9-H9[a] 2-fold interface disordered and no interactions are observed at the adjacent 3-fold region (helices H10 and H11). These interactions are now present in the CA_{CPSF6} and CA_{Nup153} (Figure 3-3).

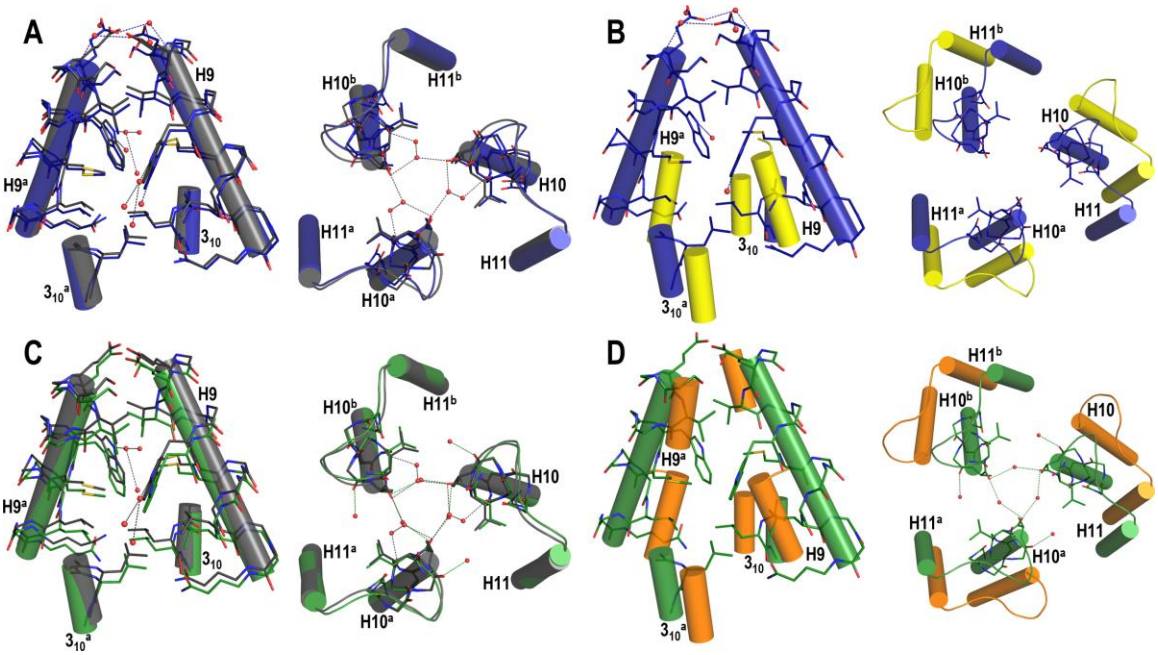


Figure 3-3. CPSF6 and Nup153 binding affect the 2-fold and 3-fold inter-hexamer interfaces.

(**A**) Native CA (gray) superposed with CA_{CPSF6} (blue). (**B**) CA_{CPSF6} superposed with $CA_{XL\text{-}CPSF6}$ (yellow). (**C**) Native CA superposed with CA_{Nup153} (green). (**D**) CA_{Nup153} superposed with $CA_{XL\text{-}Nup153}$ (orange). Water molecules are shown as spheres. Dashed lines indicate water-mediated interactions. Superscripts indicate different subunits.

CPSF6 and Nup153 binding results in subtle structural changes at the 2-fold and 3-fold regions (Figure 3-3, A, C). Specifically, we observe changes in hydrophobic and water-mediated interactions, resulting in the decrease in the buried surface area of up to ~15 % at the 2-fold and ~50 % at the 3-fold interfaces (Table 3-2).

Table 3-2. Buried surface area calculated for interfaces in the CA_{CPSF6} and CA_{Nup153} structures.

Structure	Buried surface area (BSA), Å²		
	Intra-hexamer interface	Inter-hexamer interface	
	CA_{NTD}-CA_{NTD} CA_{NTD}-CA_{CTD}	CA_{CTD}-CA_{CTD} 2-fold	CA_{CTD}-CA_{CTD} 3-fold
CA_{CPSF6}	2254.4 (8.6 %)	842.3 (3.2 %)	136.2 (0.3 %)
CA_{Nup153}	2308.5 (9.0 %)	782.9 (3.0 %)	175.5 (0.5 %)
CA	2237.7 (8.8 %)	906.4 (3.6 %)	277.5 (0.7 %)

Morphology studies of the effects of pharmacological ligands on CA assembly

In mature HIV-1 particles, CA forms a conical capsid core with an average length of ~103 nm, an average diameter of ~52 nm at the broad end, and ~25 nm on the tapered end (*82*). Purified HIV-1 CA assembles *in vitro* into hollow cylinders highly heterogeneous length with a diameter of approximately 55 nm (*81*). Upon incubation of a capsid-nucleocapsid fusion protein with a purified HIV-1 RNA template, a mixture of cones and cylinders is formed (*83*).

To evaluate the effects of the pharmacological ligands (Figure 3-4) on the assembly competence of CA, we incubated CA protein in the presence of DMSO and compounds (PF74, NYAD6, 18E8) at high salt concentration and analyzed cylinder formation in TEM.

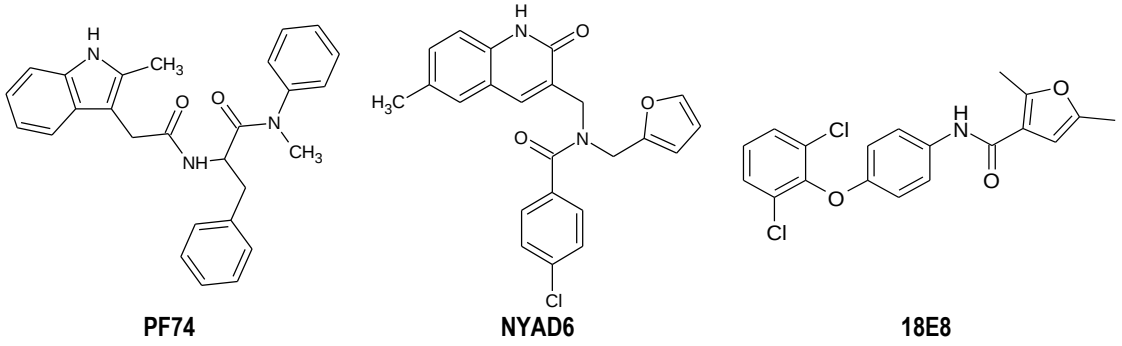


Figure 3-4. Chemical structures of selected pharmacological ligands.

As expected, CA protein assembles into long hollow cylinders (or tubes), with DMSO having minimal effect on the morphology of the observed assemblies (Figure 3-5, A). Consistent with previous reports, in the presence of PF74 CA assembles into the short tubes and cones (*269*). CA assembly in the presence of NYAD6 was greatly damaged, resulting in the occasional formation of short tubes (Figure 3-5, B) (*138*). Notably, in the presence of 18E8 CA can form cylinders, as well as assemblies of aberrant morphologies: sheets, spheres, wide tubes (Figure 3-5, C).

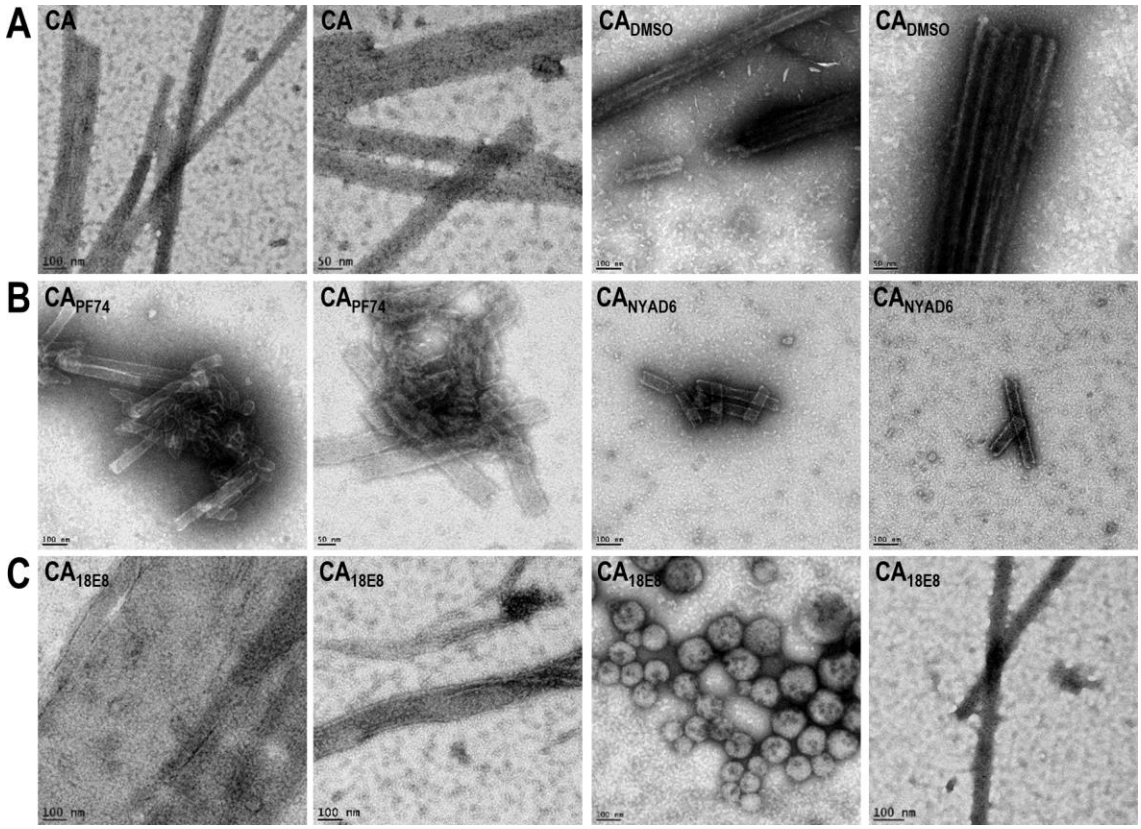


Figure 3-5. Effects of pharmacological ligands on CA assembly.

(**A-C**) TEM analysis of CA assemblies in the presence of DMSO or ligands, as indicated. Projection images were recorded from the corresponding samples; scale bars 100 nm and 50 nm.

To determine the binding sites for the NYAD6 and 18E8, we have tried to obtain crystal structures in complex with native or crosslinked CA. Unfortunately, all attempts to soak or co-crystallize NYAD6 and 18E8 with CA turned out to be unsuccessful, presumably due to low occupancy of the ligand in the binding site.

Implications for the mechanism of uncoating

Early studies based on high variability assays with reverse transcription complexes (RTCs) and pre-integration complexes (PICs) had suggested that CA was not associated with the product of reverse transcription and had led to the hypothesis that CA structure is important only for early viral replication. However, more recent studies on the kinetics of

uncoating and characterization of the restriction factor TRIMCypA (*238, 270*) demonstrate that reverse transcription facilitates uncoating. However, the exact timing of uncoating or disassembly of the capsid lattice is still being debated.

CPSF6 and Nup153 binding to the CA_{NTD}-CA_{CTD} interface that exists only in the context of assembled CA hexamer provides an argument for classifying CPSF6 and Nup153 as cellular factors that bind retroviral capsid before its complete disassembly. This observation supports a model wherein the capsid remains intact (or at least partially intact) immediately before docking at the nuclear pore.

In summary, our study confirmed that host proteins CPSF6 and Nup153, in addition to CA-targeting antiviral PF74, make contacts with two neighboring subunits within CA hexamer. The structures revealed the binding of host proteins additionally affects 2-fold and 3-fold interfaces between neighboring hexamers. This information may be instrumental for future studies of the mechanisms by which CA-interacting cellular factors affect viral uncoating, nuclear import, and viral restriction. Moreover, small-molecule CA-targeting antivirals that interfere with interactions at the CA-CA interfaces within the assembled core may offer novel avenues for therapeutic intervention.

Chapter 4. X-ray crystal structures of P38A, P38A/T216I, E45A and E45A/R132T reveal structural mechanism of HIV-1 capsid stability

Note: This chapter is based on a research paper currently in preparation that will be submitted in *Nature Communications*. The contributions of co-authors are described in the "Footnotes."

A. Introduction

During maturation of human immunodeficiency virus type 1 (HIV-1), capsid proteins (CA) assemble into a conical core (or capsid) surrounding the viral genome. Formation of a core of optimal stability is a strict requirement for efficient HIV-1 infection (*92, 93, 97*). Following delivery into the cytoplasm, the HIV-1 core undergoes controlled disassembly (uncoating), which seems coordinated with productive reverse transcription and cloaking the DNA product from immune surveillance of the target cell (*242, 271*).

The structure of CA and its effects on core stability are critical for uncoating, reverse transcription, nuclear entry, integration site selection, and assembly (*64, 65, 75, 92, 95, 96, 99, 106, 244*). CA folds into two distinct domains connected by a linker: N-terminus (CA_{NTD}), composed of seven α-helices and a β-hairpin, and C-terminus (CA_{CTD}), composed of a 3_{10}-helix and four α-helices (*62-65, 95*).

Purified HIV-1 CA can spontaneously assemble *in vitro* into tubes and cones that recapitulate the CA-CA interactions of authentic viral capsids (*83-85, 93*). The cores comprise ~250 CA hexamers and 12 CA pentamers (*83-85, 93*). Hexagonal lattice is the foundation of the mature capsid. It is stabilized by intra-hexamer (CA_{NTD}-CA_{NTD} and

CA_{NTD}-CA_{CTD} contacts between six adjacent CAs in a hexamer), inter-hexamer (2-fold and 3-fold CA_{CTD}-CA_{CTD} interactions between two or three adjacent CA hexamers, respectively), as well as intra-protomer (CA_{NTD}-CA_{CTD} contacts in the CA monomer) interfaces (*62-65, 85, 95, 268*).

The enormous research effort has been invested over many years to determine the phenotypes of natural and artificial mutations in HIV proteins, including CA (*272*). Those studies have been significant for defining various CA functions and mapping them to the CA_{NTD} and CA_{CTD} (*92, 93, 97, 102*). Additionally, a number of mutations have been identified to be useful in studies aimed at dissecting how CA performs its essential nonstructural functions (*92*). Among them, the most studied mutations are P38A and E45A. They were not obviously defective in assembly, maturation, or viral proteins packaging yet the infectivity was significantly reduced (*92*). The reduction of infectivity has been linked to altered core stability: P38A mutation destabilized viral cores, while E45A resulted in hyperstable capsids (*97*). Moreover, respective second site compensatory mutations, T216I and R132T, have been selected (*273*). Those mutations rescued infectivity impairment exhibited by the original mutation without correction of the intrinsic viral capsid stability defect (*273*). The available data are summarized in Table 4-1. Structural assessment of the effects of P38A, E45A, and R132T mutations has been limited to analyzing chemical shift changes by NMR spectroscopy of the corresponding purified N-terminal domain CA proteins. However, the exact structural effect of the mutations in the context of full-length CA and capsid assemblies remained unclear.

Table 4-1. Available biological data for CA WT and mutants.

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
			In vitro			
Isolation of HIV-1 cores by centrifugation *via* a detergent layer into a sucrose gradient; yield of cores, %	~10 %	~2 %	~3 %	~19 %	~22 %	Cores recovered from the P38A/T216I and E45A/R132T exhibited CA levels that were similar to those of the corresponding single mutants. Suppressor mutations do not correct the aberrant intrinsic stability of the P38A and E45A mutant capsids (*97, 273*).
Disassembly of purified HIV-1 cores; recovery, % of the total CA released from the cores during incubation	~50 % CA released from the cores	ND[a]	ND	~16 % CA released from the cores	~21 % CA released from the cores	The poor recovery of core-associated CA from the P38A mutant particles precluded its analysis. Cores isolated from E45A and E45A/R132T both exhibited slower uncoating (*97, 273*).
Exogenous reverse transcription (RT); % wild-type HIV-1 activity	100 %	~106 %	ND	~96 %	ND	No significant defects in exogenous RT activity were observed for the CA mutants. Quantities of active RT enzyme incorporated in the mature particles are similar to those in WT HIV-1 (*97*).
Viral RNA content; % wild-type HIV-1 activity	100 %	~86 %	ND	~102 %	ND	No significant reductions in viral RNA packaging were observed. Quantities of viral RNA incorporated in the mature particles are similar to those in WT HIV-1 (*97*).

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
Endogenous RT; % wild-type HIV-1 activity	100 %	~184 %	ND	~115 %	ND	E45A mutant synthesized quantities of DNA similar to that of wild-type HIV-1. P38A was enhanced by up to twofold relative to WT virions. CA mutations do not markedly impair RNA packaging or formation of a functional ribonucleoprotein complex within the virion (*97*).
Turbidity assay; rate of CA assembly, min	~10 min	~3-4 min	~2 min	~2 min	~3-4 min	All mutant proteins exhibited accelerated assembly. The effects of mutations on CA assembly *in vitro* were not correlated with biological phenotypes of the corresponding CA mutant viruses (*273, 274*).
Assembly competence; tube formation as determined by TEM	++	++	ND	++	ND	Cylinder formation was similar to CA WT at all protein concentrations (15, 10, and 5 mg/ml) tested (*92, 93*).
Atomic force microscopy (AFM) analysis; point stiffness, N/m	Assembled tubes ~0.052 N/m; isolated cores ~0.097 N/m	ND	ND	Assembled tubes ~0.153 N/m; isolated cores ~0.152 N/m	Assembled tubes ~0.146 N/m	E45A mutation elevates capsid stiffness in comparison with that of the CA WT. The stiffness of E45A/R132T mutant CA assemblies is similar to that of the E45A mutant (*275*).
Cell-based assays						
Gag expression in transfected 293T	Normal	Normal	ND	Normal	ND	Altered Gag expression or stability does not account for any of the phenotypes (*92*).
Gag processing, particle release, and	Normal	Normal	ND	Normal	ND	None of the mutations blocked Gag processing or grossly

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
viral protein packaging						affected the virion stoichiometry (*92*).
Capsid assembly; presence of conical capsids in pelleted virions as determined by TEM	+	+	ND	+	ND	Mutations supported normal or almost normal capsid assembly (*92, 93*).
Replication in CEM cells	Replicated	Failed to replicate	Replicated with delay	Failed to replicate	Replicated with delay	Compensatory T216I and R132T mutations partially restore the ability of the corresponding P38A and E45A mutant viruses to replicate in CEM cells (*97, 273*).
Single-round reporter assay; relative infectivity, % CA WT	100 %	~3 %	~36 %	~4 %	~27 %	T216I and R132T markedly enhanced the infectivity of P38A and E45A. The second-site mutations relieve a defect in a step of the virus replication cycle before integration (*93, 97, 273*).
HIV-1 RT in target cells*	Normal	Competent for efficient RT; exhibited unusually rapid kinetics	ND	Severe defects at both early and late stages of RT	ND	P38A was competent for efficient RT, but exhibited unusually rapid kinetics, with DNA synthesis peaking several hours earlier than for WT HIV-1. E45A exhibited severe defects at both early and late stages of RT. The impaired RT probably reflects a specific difference in viral core stability (*97*).
HIV-1 RT in target cells*; copies of HIV-1 DNA	~60,000 Normal	~5,000 Impaired RT	~30,000 Partially rescued	~65,000 Normal RT	~32,000 ND	Impaired infectivity of P38A is a result of its reduced RT capacity. T216I partially rescued the impaired RT

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
						exhibited by P38A. E45A mutation impairs later stages of infection, after nuclear entry (*273*).
Early and late RT*; fold change 15 min post-infection	Reference	ND	ND	Two fold higher than CA WT	ND	More rapid accumulation of early RT products was observed for E45A compared to CA WT. Late RT products did not increase significantly during 155 min time course (*264*).
Viral RNA 5-ethynyl uridine (EU) staining kinetics; beginning of decay, min post-infection	50 min	ND	ND	Staining increased in 15-25 min, followed by decline at 35 min	EU staining resembled CA WT	The capsid of E45A HIV-1 dissociated early after infection. The E45A phenotype was partially reversed with the addition of the R132T mutation (*276*).
Capsid permeability to antibodies and the fluorescent dye	Little RNA staining; almost no staining of NC	ND	ND	Viral RNA and NC staining	ND	Both the small molecule dye and the anti-NC antibodies could penetrate the E45A cores (*276*).
Cell cycle dependence**; relative infectivity in arrested *vs.* control cells, fold change	<0.5 fold reduction	<0.5 fold reduction	<0.5 fold reduction	8 fold reduction	<0.5 fold reduction	P38A and P38A/T216I mutants behaved similarly to WT HIV-1. R132T corrects the selective impairment of infection of nondividing cells associated with the E45A mutant virus (*273, 277*).
Sensitivity of viruses to PF74 inhibition in single-cycle infection assays**; IC50, μM	~0.26	Hypersensitive	ND	Resistant	~0.38	E45A exhibited reduced sensitivity to inhibition by PF74 relative to WT HIV-1. P38A was more sensitive, exhibiting a greater reduction of infection at both low and high PF74 concentrations. The restored

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
						sensitivity of the E45A/R132T virus to PF74 suggests that the R132T mutation partially reverses the E45A-induced uncoating defect in target cells (*123, 177, 273*).
Extent of PF74 binding; % of WT	100	~210	ND	~100	ND	P38A bound approximately twice as much of the PF74 as WT HIV-1 particles, potentially contributing to the increased sensitivity to PF74. The compound bound to E45A to an extent comparable to that of the WT, which suggests that altered sensitivity is likely due to increased capsid stability (*123*).
The ability to saturate TRIM5 restriction in monkey cells**	Ability to abrogate restriction *in trans*	Impaired	Rescued	Similar to WT	ND	The efficient trans-abrogation of TRIM5 restriction requires particles with a stable capsid. T216I restores the ability of P38A particles to abrogate restriction. Thus, T216I mutation prevents premature disassembly of the P38A mutant core in target cells, thereby relieving its impaired ability to interact with restriction factors (*273, 278, 279*).
CPSF6-358 restriction of HIV-1 infection**	Restricted	Retained sensitivity	ND	Resistant in dividing cells, but strongly restricted in the growth-arrested cells	ND	CPSF6-358 interaction with incoming HIV-1 cores impairs productive interactions with uncoating or transport factors (*119, 198*).

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
Sensitivity to restriction by TRIM-Nup153	Restricted	Retained sensitivity	ND	Less effective at restricting	ND	Some CA disassembly may be needed for interaction with Nup153 (*119, 237*).
HIV-1 sensitivity to Nup85, **Nup107**, **Nup133**, **Nup153**, **Nup155**, **Nup160**, and **Nup358 depletions****	HIV-1 infection was decreased in cells depleted of Nup155, Nup160, or Nup358	ND	ND	Less sensitive to Nup depletions	ND	E45A HIV-1 mutant interacts inefficiently with the nuclear pore complex (*198, 235, 237*).
Effect of CypA knockdown or cyclosporine treatment on infectivity in Nup153 depleted cells**	Relatively unaffected	ND	ND	5- to 20-fold increase in infectivity	ND	Perturbation of CypA binding dictated the sensitivity to Nup153 depletion. It appears the amount of CypA bound to the HIV-1 core can dictate whether the pre-integration complex undergoes downstream processes requiring Nup153, perhaps by altering the dynamics of uncoating (*235*).
HIV-1 dependence on TNPO3 for infection**	Dependent	Partially dependent	ND	Not affected	Partially dependent	Infection by the E45A/R132T mutant exhibited TNPO3 dependence between those of E45A and the WT. The second-site suppressor mutant restores TNPO3 dependence of infection (*177, 198, 210*).
Structural assessment						
Nuclear magnetic resonance (NMR); chemical shift change, ppm	Reference spectrum	<0.43 Dispersed over a wider region (V36, I37, A38, M39, K30-S33, W23-E28, M55, L136, V142)	ND	0.8; Local (A45, G46)	0.3-1; Local (A45, G46, I129, T132, W133, L136)	The spectra of all the mutants are very similar to that of CA WT, demonstrating conservation of the overall global fold. The effects of the P38A may result from subtle

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
						changes in the overall structure and its involvement in the intra-hexamer interactions. The effects of the E45A and R132T mutations arise from the change in chemical nature of the substituted amino acid (*273*).
This study						
In vitro						
Assembly competence; cylinder formation as determined by cryo-EM	Long tubes	No tubes	Long tubes	Short tubes and cones	Long tubes	Altered assembly morphologies were reverted to long tubes as the CA WT by the second compensatory mutations.
Pelleting assay	++	-	++	+++	++	Second site mutations reversed assembly efficiencies to levels similar to the CA WT.
Cell-based assay						
CypA-DsRed loss assay; core stability	Reference	Less stable	Less stable	More stable	More stable than CA WT, but less stable than E45A	E45A/R132T is less stable than E45A, but is significantly more stable than CA WT. P38A and P38A/T216I are slightly different. Collectively, compensatory mutations do not fully correct the intrinsic stability defects imposed by P38A and E45A mutations.
Structural assessment						
X-ray crystallography	Reference structure	Affected residues: P1, H12, L20, E28, E29, K30, A31, F32, S33, P34, E35, V36, I37, A38, M39, S41, A42,	In addition to residues affected by P38A: T200, I201, L202, K203, A204, L205, G206,	Affected residues: P1, H12, A45, E128, R132, Q50, D51	Affected residues: P1, H12, A45, E128, T132, Q50, D51	E45 is engaged in both attractive and repulsive ionic interactions with D51 from the neighboring subunit. Mutation of E45 to A directly relieves electrostatic repulsion, resulting in

Assay	CA WT	P38A	P38A/T216I	E45A	E45A/R132T	Interpretations, references
		E45, T54, E128, R132, R143, M144, Y145, R162, Q176	P207, G208, M215, I216, A217, Q219, G220, V221			stabilization of the E45A CA hexamer and the core. In the P38A mutant, the network of interactions around E45 is alerted leading to increase in the electrostatic repulsion between E45 and D51, which leads to the destabilization of the P38A CA hexamers and the core. R132T and P216I, can partially offset the effect of primary mutations. R132T partially restores the overall net charge of CA, while T216I stabilizes inter-hexamer interactions.

[a] ND – no data; * – these studies show contradictory results; ** – the infectivity of P38A and E45A mutant viruses are significantly impaired.

Thus, we have focused our analysis on P38A, P38A/T216I, E45A, and E45A/R132T CA mutants. To elucidate the effects of compensatory mutations, we evaluated assembly competence of CA mutant proteins *in vitro*; estimated core stability and uncoating using novel CypA-DsRed loss assay; and solved corresponding crystal structures.

We discerned that P38A CA was incapable of forming tubular structures and E45A assembled into short tubes and cones. Compensatory mutations T216I and R132T, respectively, rescued the ability of mutant proteins to assemble into long hollow cylinders. However, CypA-DsRed loss assay revealed that P38A and P38A/T216I underwent rapid uncoating, while E45A and E45A/R132T disassembled much slower than CA WT. Thus, the second site suppressor mutations did not fully correct the defects imposed by primary mutations.

Our structural analysis of full-length CA proteins bearing corresponding mutations suggests that mutant phenotypes are not a consequence of major structural rearrangements in HIV-1 CA. Induced changes prompt that the residues in positions 45 (glutamic acid) and 51 (aspartic acid) are the key players controlling the stability of the capsid core and its disassembly. Moreover, the structures highlight possible structural rearrangements that may affect host factor recognition and trafficking across the capsid shell.

B. Materials and methods

Design, expression, and purification of CA mutants

P38A, P38A/T216I, E45A, E45A/R132T CA mutants were based on a pET11a construct (*268*). Mutations were introduced using overlap extension PCR cloning and

verified by DNA sequencing. Mutant P38A, P38A/T216I, E45A, E45A/R132T CA proteins were expressed and purified as previously described (*245, 268*).

Crystallization of CA mutants

Crystals of the P38A, P38A/T216I, E45A, E45A/R132T CA grew at 18 °C in hanging drops, containing 2-5 mg/ml of protein, 2-14 % PEG 3350, 2-6 % glycerol, sodium iodide, and sodium cacodylate. Hexagonal plate-like crystals appeared after five days, and crystal growth was completed in over two weeks. Crystals were briefly soaked in 20 % glycerol or paraffin oil before cryo-cooling in liquid nitrogen.

Data collection and structure determination

Data were collected on a MAR CCD or Pilatus3 6M detectors at the Advanced Photon Source, Sector 23. Datasets were collected and processed using XDS (*247*). The data were examined for the presence of systematic absences. However, no characteristic patterns were observed. Thus, the crystals were indexed in hexagonal space group P6 with one CA molecule in the asymmetric unit. No twinning was present, as determined by either POINTLESS (*249*) or XTRIAGE (*250*). Space group and twinning were also verified in ZANUDA (*251*). The phase problem of CA was solved either using single-wavelength anomalous diffraction (SAD) or molecular replacement, with the native CA (PDB ID: 4XFX) as a starting model. For SAD, substructure solution, phasing, density modification, model building, and refinement were carried out using SHELX C/D, SOLOMON, PARROT, BUCCANEER and REFMAC in CRANK-2 (*251*). For molecular replacement, initial phases were solved *via* PHASER (*251*). Several rounds of iterative model building and refinement were carried out using Coot (*253*) and PHENIX (*250*), REFMAC (*251,*

252), or PDBREDO (http://www.cmbi.ru.nl/pdb_redo/), respectively. Structure validation of final models was performed with MOLPROBITY (http://molprobity.biochem.duke.edu/). Accessible and buried surface areas were calculated using PISA (*251*). The figures showing structural information were generated in PyMOL (http://www.pymol.org/).

Pelleting assay

CA wild type (WT) and mutants (P38A, P38A/T216I, E45A, and E45A/R132T) were assembled at 2 mg/ml (80 µM) in buffer containing 1 M sodium chloride and 50 mM Tris hydrochloride pH 8.0 at 37 °C for 1 hr. 5 µl samples were withdrawn from the reaction mixtures and immediately used for cryo-EM analysis. The remaining sample was pelleted at 21,000×g with an Eppendorf centrifuge 5417R for 30 min at 4 ºC and supernatants (S) and pellets (P) were mixed with 4× LDS loading buffer (Invitrogen) supplemented with 10 mM dithiothreitol (DTT), without boiling, were loaded on 10 % SDS-PAGE and stained with Coomassie Blue.

Morphology studies of CA mutants

The fresh assembled samples (4 µl) were applied to the carbon side of a glow discharged perforated Quantifoil grid (Quantifoil Micro Tools, Jena, Germany). The grids were then manually blotted with a filter paper from the backside to remove the excess fluid, and plunge-frozen in liquid ethane using a homemade gravity plunger. For cryo-EM imaging, the frozen grids were loaded into a cryo-holder (Gatan Inc., Pleasanton, CA), inserted into a Tecnai F20 transmission electron microscope (FEI, Inc., Hillsboro, OR) and imaged with a 4k×4k charge-coupled device camera (Gatan). Low dose (~20 $e^{-}/Å^{2}$)

projection images were recorded at a nominal magnification of 50,000× with a pixel size of 2.26 Å and underfocus values ranging from 3.0 to 5.0 µm. The low magnification images were recorded at the magnification 5,000×.

CypA-DsRed loss assay

HIV-1 viruses bearing P38A, P38A/T216I, E45A, and E45A/R132T mutations were produced in 293T cells by incorporating INsfGFP and CypA-DsRed. Viruses were bound to poly-l-lysine treated coverglass. Images were taken later for each CA mutant and WT. The virus was permeabilized with 100 µg/ml Saponin for 1min followed by wash 1× in Dulbecco's phosphate-buffered saline (dPBS). The solution was replaced with 200 µl dPBS and 4-fields of view were imaged at room temperature. Total numbers of integrase (IN) spots and CypA-DsRed spots were determined for each condition. The % core retention was calculated for the first time point for each CA-mutant by using the respective intact virus (pre-Saponin) as a measure. Core stability was determined by plotting the loss of CypA-DsRed spots over-time with respect to the initial number of spots. INsfGFP spots remained constant and served as a reference marker. Cyclosporine A (CsA) 5 µM was added at t = 21 min. The background immature particles were subtracted for all the time points.

C. Results and discussion

Crystallographic analysis of P38A and P38A/T216I CA mutants

We have crystallized and solved the structures of full-length CA bearing P38A, P38A/T216I, E45A or E45A/R132T substitutions in the wild type (WT) CA background

(Figure 4-1). Similar to the CA WT, the crystal structures of CA mutants have been solved in space group P6 with one molecule/asymmetric unit (Table 4-2).

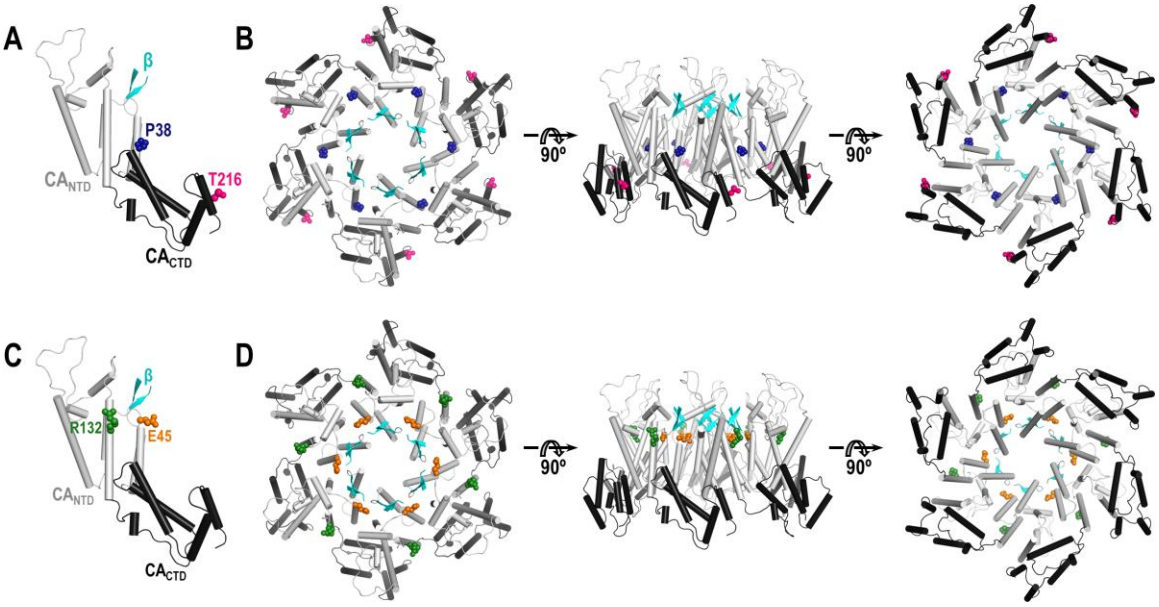


Figure 4-1. Mutation sites in the wild type full-length HIV-1 capsid protein (CA WT).

Locations of mutation sites P38 (dark blue spheres) and T216 (pink spheres) in CA WT monomer (**A**) and hexamer (**B**) shown with alternate orthogonal views. Locations of mutation sites E45 (orange spheres) and R132 (green spheres) in CA WT monomer (**C**) and hexamer (**D**). CA_{NTD}s are in light gray, CA_{CTD}s in black, N-terminal β-hairpin in light blue.

Table 4-2. Summary of data collection and refinement statistics.

	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
Data collection					
X-ray source	APS 23 ID-B	APS 23 ID-D	APS 23 ID-D	APS 23 ID-B	APS 23 ID-D
Software	XDS	XDS	XDS	XDS	XDS
Space group	P6	P6	P6	P6	P6
Unit cell dimensions					
a, b, c (Å)	92.1 92.1 57.5	92.2 92.2 57.7	87.6 87.6 56.5	92.5 92.5 57.8	92.4 92.4 57.7
α, β, γ (°)	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0
ASU content	1	1	1	1	1
Wavelength (Å)	1.0332	1.03319	1.0332	1.0332	1.03319
Resolution range (Å)[a]	46.6–2.4 (2.5–2.4)	46.8–2.6 (2.7–2.6)	37.9–2.5 (2.6–2.5)	46.8–2.2 (2.3–2.2)	46.8–2.0 (2.1–2.0)
R_{merge}	0.065 (>1)	0.071 (0.804)	0.144 (0.807)	0.084 (>1)	0.054 (0.992)
R_{meas}	0.068 (>1)	0.075 (0.850)	0.162 (0.916)	0.089 (>1)	0.057 (>1)
R_{pim}	0.022 (0.344)	0.024 (0.273)	0.073 (0.428)	0.027 (0.417)	0.018 (0.395)
$\langle I/\sigma I \rangle$	19.6 (2.1)	15.8 (1.7)	9.1 (2.0)	17.6 (1.9)	23.3 (1.9)
$CC_{1/2}$ (%)	99.8 (71.6)	99.8 (79.1)	99.2 (64.3)	99.9 (62.8)	100 (58.9)
Completeness (%)	99.5 (96.7)	99.8 (98.4)	99.3 (95.9)	99.9 (99.2)	99.9 (99.3)
Redundancy	10.1 (9.6)	9.8 (9.3)	4.8 (4.4)	11.3 (10.7)	9.8 (7.0)
Mosaicity	0.14	0.10	0.23	0.13	0.07
Refinement					
Resolution (Å)	46.6–2.4	46.8–2.6	34.6–2.5	47.0–2.2	47.0–2.0
No. total reflections	110167	86275	41802	160668	186695
No. unique reflections	10893	8751	8643	14249	19126
No. test reflections[b]	631	521	422	692	918
R_{work} / R_{free}	22.8 / 25.6	20.0 / 24.1	20.3 / 24.6	20.3 / 22.3	19.7 / 21.8
No. atoms	1748	1752	1799	1844	1868
Protein	1702	1706	1724	1732	1708
Ligand/Ion	13	12	15	14	15
Water	33	34	60	98	145
Wilson B-factor (Å^2)	67.4	77.5	42.1	47.2	40.4
Average B-factors (Å^2)	87.6	99.9	69.6	63.4	58.1
Protein	87.7	100.2	69.8	63.2	58.2
Ligand/Ion	103.1	118.6	78.7	85.2	66.9

	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
Water	76.4	76.5	61.3	62.8	56.8
RMS deviations					
Bond lengths (Å)	0.002	0.009	0.002	0.007	0.007
Bond angles (°)	0.497	1.27	0.456	1.062	1.106
MolProbity Statistics[c]					
All atom clash score	4.98	2.92	4.90	1.43	2.61
Rotamer outliers (%)	0	0	0	0	0
Cβ deviations >0.25 Å	0	0	0	0	0
Ramachandran [c]					
Favored region (%)	99	99	98	98	98
Outliers (%)	0	0	0	0	0
PDB accession code	6B2G	6B2H	6B2I	6B2J	6B2K

[a] Values in parentheses are for highest-resolution shell; [b] random selection; [c] values obtained from MOLPROBITY.

The structures of P38A and P38A/T216I are very similar to that of CA WT (rmsd 0.3213 and 0.4081 Å), demonstrating conservation of the overall global fold of the protein (*64, 268*). P38A is the mutation in the middle of the helix H2, which, together with helices H1 and H3, forms the 18-helix barrel at the center of the hexamer. Consistent with the previous structural assessment by NMR (*273*), the P38A exhibited overall subtle changes that were dispersed over a wide region (Figure 4-2). The affected residues are located proximal to the site of mutation in helices H1 (L20, E28, E29, K30), H2 (P34, E35, V36, I37, A38, M39, S41, A42) and H3 (T54) (Figure 4-2, B). Moreover, the changes extended downstream to the site of mutation, altering the preceding loop between helices H1 and H2 (A31, F32, S33) (Figure 4-2, D). Further changes at the end of helix H7 (R143, M144, Y145) were mediated primarily *via* F32 in the affected loop, which in turn remodeled the loop between helices H8 and H9 (Q176) (Figure 4-2, D). Furthermore, additional rearrangements are observed at the beginning of helix H8 (R162) in the neighboring subunit due to the changes in H7. Finally, there are changes upstream to the mutation site that alter network of interactions between E45 (the loop between H2 and H3), E128 (H7), R132 (H7) of one subunit and P1 (β-hairpin), H12 (β-hairpin), Q50 (H3), D51 (H3) of the neighboring one in a hexamer (Figure 4-2, C). Interestingly, this network of interactions controls the conformation β-hairpin (*280*).

Thus, the P38A mutation initiates a cascade of subtle changes altering network of intra-protomer, as well as CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} intra-hexamer interactions, affecting three different subunits in a hexamer (Figure 4-2, A). Those changes could be related to a general effect of the structural adjustment for the mutation in CA. These observations together with a subtle decrease in calculated buried surface area (BSA) at CA-

CA interfaces (Table 4-3) suggest that P38A mutation has a more general effect such as an overall loosening of the CA structure.

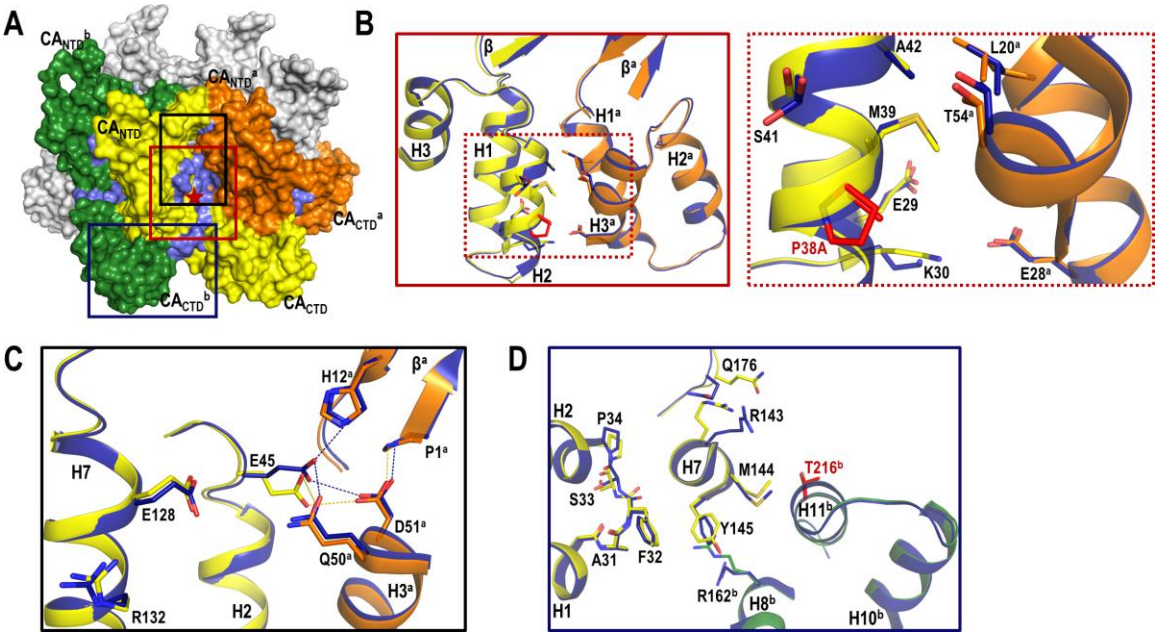


Figure 4-2. Structural changes associated with P38A mutation.

(**A**) CA hexamer is shown in surface view representation with three CA monomers colored in green, yellow and orange, and the other three in gray. Mutation site is marked with a red star. The residues affected by P38A mutation are shown in light blue. (**B-D**) Superposition of CA WT (three neighboring subunits colored in green, yellow and orange) and P38A (in blue). Specific residues affected by P38A mutation (in red) are shown as sticks.

Table 4-3. Buried surface area calculated for interfaces in CA mutant structures.

Structure	Buried surface area (BSA), Å^2			
	Intra-hexamer interface	Inter-hexamer interface		
	CA_{NTD}-CA_{NTD} CA_{NTD}-CA_{CTD}	CA_{CTD}-CA_{CTD} 2-fold	CA_{CTD}-CA_{CTD} 3-fold	CA_{NTD}-CA_{NTD} 3-fold
P38A	2270.4 (8.7 %)	887.4 (3.4 %)	237.9 (0.6 %)	-
P38A/T216I	2402.8 (9.1 %)	935.6 (3.5 %)	586.2 (1.5 %)	-
E45A[a]	2315.8 (8.7 %)	1452.5 (5.4 %)	1418.1 (3.5 %)	143.4 (0.4 %)
E45A[b]	2443.2 (9.3 %)	818.1 (3.1 %)	194.7 (0.5 %)	-
E45A/R132T	2139.7 (8.4 %)	843.2 (3.3 %)	131.4 (0.3 %)	-
CA WT	2237.7 (8.8 %)	906.4 (3.6 %)	277.5 (0.7 %)	-

Similar rearrangements are observed in the P38A/T216I structure due to P38A mutation (Figure 4-3, A-C). Second site mutation, T216I, is located in helix H11, which is

spatially proximal to H7, one of the areas affected by P38A (Figure 4-2, D). T216I results in subtle rearrangements close to mutation site in helix H11 (M215, I216, A217, Q219, G220, V221), as well as affects helix H10 (I201, L202, K203, A204, L205) and the loop between H10 and H11 (G206, P207, G208) (Figure 4-3, D). As a result, there is ~150 % increase in the BSA at the 3-fold CA_{CTD}-CA_{CTD} inter-hexamer interface. Moreover, ~5 % increase in the BSA is observed at the other CA-CA interfaces (Table 4-3). Thus, T216I induces further subtle rearrangements that may lead to slight stabilization of the CA hexamers as well as interactions between them, thereby partially offsetting the destabilizing effect of the P38A.

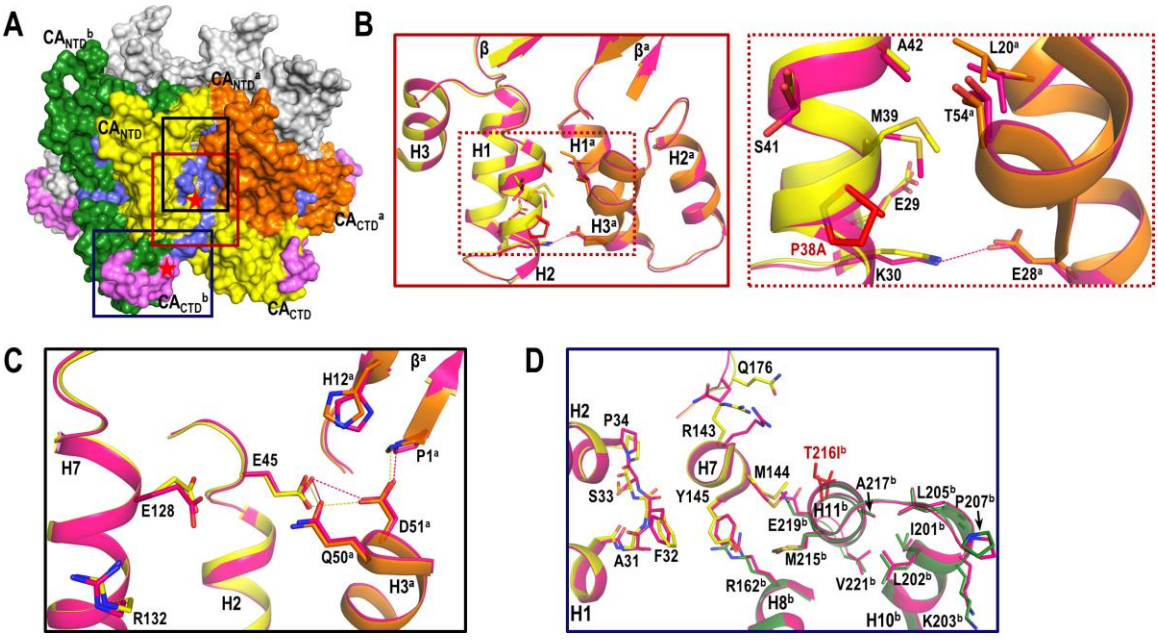


Figure 4-3. Structural changes associated with P38A/T216I mutations.

(**A**) CA hexamer is shown in surface view representation with three CA monomers colored in green, yellow and orange, and the other three in gray. Mutation sites are marked with red stars. The residues affected by mutations are shown in light blue (P38A) and light pink (T216I). (**B-D**) Superposition of CA WT (three neighboring subunits colored in green, yellow and orange) and P38A/T216I (in pink). Specific residues affected by P38A and T216I mutations (in red) are shown as sticks.

Crystallographic analysis of E45A and E45A/R132T CA mutants

Surprisingly, E45A mutant crystallizes in the same space group, but with two different unit cell dimensions: a=b=87.6 Å, c=56.5 Å (labeled E45A^{a}) and a=b=92.5 Å, c=57.8 Å (labeled E45A^{b}) (Table 4-2). Both structures reveal the overall global fold to be essentially the same as in CA WT. Superpositions of E45A^{a} and E45A^{b} with CA WT result in larger deviations of atomic positions (rmsd 1.4429 and 0.8113 Å, respectively) in comparison to those observed for P38A and P38A/T216I.

The ~5 % difference in the unit-cell dimensions between E45A^{a} and E45A^{b} may be explained by the repositioning of the loop between helices H8 and H9 (~3 Å movement). This movement primarily affects adjacent helices H8 and H9 (Figure 4-4) but is also translated to the helix 3_{10}, which is part of the 2-fold interface together with H9, as well as H10 and H11 that comprise the 3-fold inter-hexamer interface (Figure 4-4). This is reminiscent of the rearrangement previously observed upon crystal dehydration of the CA WT (*268*). As a result, the hexamers in E45A^{a} arrange tighter than in CA WT or E45A^{b} (Figure 4-5) forming extended interactions (Table 4-4) at the 2-fold and 3-fold CA_{CTD}-CA_{CTD} inter-interfaces that are distant to the location of the mutation. Solvent-accessible area calculations reveal that E45A^{a} has ~160 % and ~510 % more buried surface than CA WT at the 2-fold and 3-fold, respectively (Table 4-3). Notably, novel inter-hexamer interface is formed between N-terminal domains along the 3-fold symmetry axis (BSA 143.4 Å^2 or ~0.4 %) that involves R82 of the E45A^{a} CA_{NTD}s (Figure 4-5). Thus, the presence of this tighter structure, additionally stabilized by CA_{NTD}-CA_{NTD} 3-fold interface, may explain the hyperstabilization observed for the HIV-1 E45A capsid.

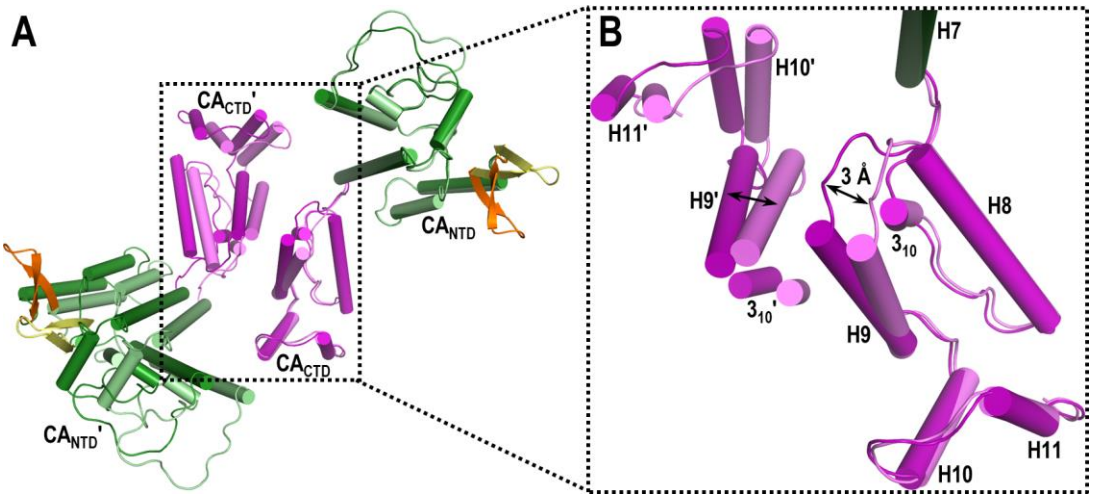


Figure 4-4. Intra-subunit rearrangement linked to changes at the inter-hexamer interfaces.

(**A**) Least squares superposition (residues 17–143) of E45A[a] (light green CA_{NTD}s, light purple CA_{CTD}s) on E45A[b] (green CA_{NTD}s, purple CA_{CTD}). (**B**) An enlarged view of the boxed region shows changes in the position of helices H9, 3_{10}, H10 and H11 in neighboring subunits (marked with prime symbols).

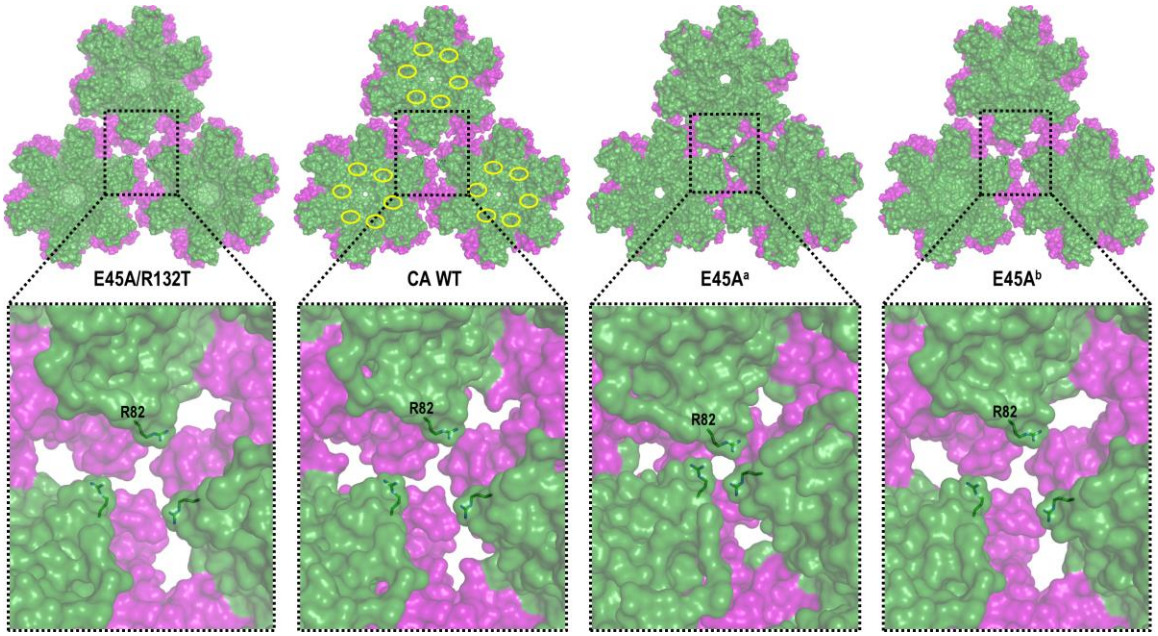


Figure 4-5. Arrangement of hexamers in the crystal structures of CA WT, E45A, and E45A/R132T mutants.

Neighboring hexamers in the lattices of CA WT, E45A[a], E45A[b] and E45A/R132T are shown in surface representation and also in enlarged views. CA_{NTD}s are colored in green, CA_{CTD}s in purple. R82 is labeled and shown in sticks. E45A mutation sites are outlined in yellow ovals.

Table 4-4. Intra- (CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD}) and inter- (2-fold and 3-fold CA_{CTD}-CA_{CTD}) hexamer interactions.

Residues participating in hydrogen bonding networks are shown in pink. Interacting residues present in all compared structures are highlighted in blue. All interactions are among domains from neighboring CA subunits.

CA WT	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
CA_{NTD}-CA_{NTD} intra-hexamer interface					
V3 / H12	V3 / H12	V3 / H12	-	V3 / H12	V3 / H12
-	-	-	-	V3 / Q13	V3 / Q13
-	-	-	Q4 / N5	-	-
-	-	-	Q4 / Q7	-	-
Q4 / V11	Q4 / V11	Q4 / V11	-	Q4 / V11	Q4 / V11
Q4 / H12	Q4 / H12	Q4 / H12	-	Q4 / H12	Q4 / H12
Q4 / H2O / H12	-	-	-	-	-
-	-	-	-	N5 / V11	-
-	-	-	-	N5 / L6	-
-	-	-	-	L6 / L6	-
-	-	-	-	L6 / Q7	-
-	-	Q7/Q9	-	Q7 / Q9	-
-	-	-	-	G8 / Q9	-
R18 / P17	-	R18 / P17	-	R18 / P17	R18 / P17
R18 / R18	R18 / R18	R18 / R18	R18 / R18	R18 / R18	R18 / R18
T19 / P17	T19 / P17	T19 / P17	T19 / P17	T19 / P17	T19 / P17
-	-	-	N21 / A22	-	-
K30 / E28	-	K30 / E28	K30 / E28	-	-
-	-	-	-	-	K30 / H_2O / T58
-	-	-	-	-	K30 / H_2O / G60
E35 / N57	-	E35 / N57	-	E35 / N57	E35 / N57
E35 / T58	-	E35 / T58	E35 / T58	E35 / T58	E35 / T58

CA WT	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
E35 / H_2O / T58	E35 / H_2O / T58	-	-	E35 / H_2O / T58	E35 / H_2O / T58
-	-	-	E35 / V59		
E35 / G60	-	E35 / G60	E35 / G60	E35 / G60	E35 / G60
P38 / H_2O / T54	-	-	-	-	-
P38 / N57	A38 / N57	A38 / N57	P38 / N57	P38 / N57	P38 / N57
P38 / T58	A38/T58	A38/T58	P38 / T58		
M39 / V24	-	-	M39 / V24	M39 / V24	M39 / V24
M39 / T58	M39 / T58	M39 / T58	-	M39 / T58	M39 / T58
-	-	-	A42/ H_2O / I15	-	-
A42 / L20	A42 / L20	A42 / L20	A42 / L20	A42 / L20	A42 / L20
A42 / T54	A42 / T54	A42 / T54	A42 / T54	A42 / T54	A42 / T54
L43 / L20	-	-	L43 / L20	-	L43 / L20
-	-	-	L43 / P17	-	-
E45 / H_2O / H12	E45/H12	-	-	-	-
E45 / A14	E45 / A14	E45 / A14	A45 / A14	A45 / A14	A45 / A14
E45 / I15	E45 / I15	E45 / I15	-	-	-
E45 / Q50	E45 / Q50	E45 / Q50	-	-	-
E45 / D51	E45 / D51	E45 / D51	-	-	-
CA_{NTD}-CA_{CTD} intra-hexamer interface					
R162 / H_2O / M144	-	-	R162 / M144	-	-
R162 / Y145	R162 / Y145	R162 / Y145	R162 / Y145	R162 / Y145	R162 / Y145
V165 / A64	V165 / A64	V165 / A64	V165 / A64	V165 / A64	V165 / A64
D166 / H62	D166 / H62	D166 / H62	D166 / H62	D166 / H62	D166 / H62
D166 / Q63	D166 / Q63	D166 / Q63	D166 / Q63	D166 / Q63	D166 / Q63
D166 / H_2O / Q63	D166 / H_2O / A63	D166 / H_2O / A63	-	D166 / H_2O / A63	D166 / H_2O / A63
D166 / A64	D166 / A64	D166 / A64	D166 / A64	D166 / A64	D166 / A64
D166 / H_2O / A65	D166 / H_2O / A65	-	D166 / H_2O / A65	D166 / H_2O / A65	D166 / H_2O / A65

CA WT	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
D166 / H_2O / Y145	D166 / H_2O / Y145	-	D166 / H_2O / Y145	D166 / H_2O / Y145	D166 / H_2O / Y145
Y169 / Q63	Y169 / Q63	Y169 / Q63	Y169 / Q63	Y169 / Q63	Y169 / Q63
Y169 / Q67	Y169 / Q67	Y169 / Q67	Y169 / Q67	Y169 / Q67	Y169 / Q67
-	K170 / Q63	-	K170 / Q63	K170 / Q63	K170 / Q63
R173 / H_2O / L56	-	-	-	-	-
R173 / N57	R173 / N57	R173 / N57	R173 / N57	R173 / N57	R173 / N57
R173 / V59	R173 / V59	R173 / V59	R173 / V59	R173 / V59	R173 / V59
R173 / E63	R173 / E63	R173 / E63	R173 / E63	R173 / E63	R173 / E63
R173 / H_2O / E63	-	-	R173 / H_2O / E63	R173 / H_2O / E63	R173 / H_2O / E63
-	Q179 / H_2O / Q67	-	Q179 / Q67	-	-
-	-	-	Q179 / K70	-	-
-	-	-	N182 / Q67	-	-
T210 / E71	T210 / E71	T210 / E71	T210 / E71	T210 / E71	T210 / E71
T210 / E75	T210 / E75	-	-	-	-
L211 / A64	L211 / A64	L211 / A64	L211 / A64	L211 / A64	L211 / A64
L211 / Q67	L211 / Q67	L211 / Q67	L211 / Q67	L211 / Q67	L211 / Q67
L211 / M68	L211 / M68	L211 / M68	L211 / M68	L211 / M68	-
L211 / E71	L211 / E71	L211 / E71	L211 / E71	L211 / E71	L211 / E71
E212 / M68	E212 / M68	E212 / M68	-	E212 / M68	E212 / M68
E212 / H_2O / E71	-	-	-	E212 / H_2O / E71	E212 / H_2O / E71
E212 / K140	E212 / K140	-	E212 / K140	E212 / K140	E212 / K140
-	E212/R143	E212/R143	-	E212 / R143	E212/R143
E212 / H_2O / R143	-	-	-	-	-
E212 / M144	E212 / M144	E212 / M144	E212 / M144	E212 / M144	E212 / M144
M215 / A64	-	-	M215 / A64	M215 / A64	-
M215 / M68	M215 / M68	M215 / M68	M215 / M68	M215 / M68	M215 / M68
-	M215 / M144	M215 / M144	M215 / M144	-	M215 / M144

CA WT	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
M215 / Y145	M215 / Y145	-	M215 / Y145	-	M215 / Y145
-	-	-	-	-	-
-	Q219 / M144	Q219 / M144	Q219 / M144	-	Q219 / M144
-	-	-	-	Q219 / H_2O / M144	-
CA_{CTD}-CA_{CTD} 2-fold inter-hexamer interface					
-	-	-	R143 / W184	-	-
-	-	-	R143 / E187	-	-
-	-	-	R143 / T188	-	-
-	-	-	S149 / Q192	-	-
-	-	-	I150 / L189	-	-
L151 / L151	L151 / L151	-	L151 / L151	-	-
L151 / L189	L151 / L189	L151 / L189	L151 / L189	L151 / L189	L151 / L189
L151 / Q192	L151 / Q192	L151 / Q192	L151 / Q192	L151 / Q192	L151 / Q192
-	-	-	L151 / N193	-	-
-	-	-	E175 / W184	-	-
E175 / H_2O / W184	-	-	-	-	-
-	-	-	E175 / Q192	-	-
Q176 / W84	-	-	-	-	-
Q176 / H_2O / W184	-	-	-	-	-
-	-	-	A177 / W184	-	-
S178 / E180	S178 / E180	S178 / E180	S178 / E180	-	-
E180 / E180	E180 / E180	E180 / E180	E180 / E180	E180 / E180	E180 / E180
-	-	-	-	E180 / V181	E180 / V181
V181 / W184	V181 / W184	V181 / W184	V181 / W184	V181 / W184	V181 / W184
W184 / W184	-	-	-	W184 / W184	W184 / W184
W184 / M185	W184 / M185	W184 / M185	W184 / M185	W184 / M185	W184 / M185

CA WT	P38A	P38A/T216I	E45A[a]	E45A[b]	E45A/R132T
CA_{CTD}-CA_{CTD} 3-fold inter-hexamer interface					
-	-	-	I201 / A204	-	-
-	-	-	K203 / T216	-	-
-	-	-	K203 / A217	-	-
-	-	-	K203 / G220	-	-
-	-	A204 / A204	-	-	-
A204 / H_2O / A204	A204 / H_2O / A204	A204 / H_2O / A204	-	A204 / H_2O / A204	A204 / H_2O / A204
-	-	-	A204 / L205	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	P207 / E212	-	-
-	-	-	P207 / E213	-	-
-	-	-	P207 / T216	-	-

The E45A primarily remodels network of interactions proximal to the site of mutation (Figure 4-6). It involves residues in the loop between helices H2 and H3 (S44, A45, G46), helix H7 (E128, K130, R132) of one subunit, as well as residues in helix H3 (Q50, D51) and β-hairpin (P1, H12, Q13) of the neighboring subunit in a hexamer (Figure 4-6, A, B, E, F). As a result, there is a ~10 % increase in the BSA at the intra-hexamer interfaces of E45A[b]. Additional subtle changes are observed at the inter-hexamer interfaces resulting in ~10 % and ~30 % decrease in the BSA at the 2-fold and 3-fold, respectively (Table 4-3). Notably, this site has also been altered by the P38A mutation (Figure 4-6, A-D).

Similar changes are observed in the E45A/R132T structure (Figure 4-6, G). Second site mutation, R132T, located in the helix H7 in the vicinity of E45A, partially restores extended water mediated interactions impaired by E45A. Additionally, there is a small decrease in the BSA at the intra- and inter-hexamer interfaces (Table 4-3). Moreover, R132T partially restores the overall net charge of CA, which is consistent with HIV-1 capsid stability being sensitive to changes in pH and ionic strength of the media (Figure 4-6). This suggests that the R132T mutation may at least partially decrease the stability of the hexamers and hexagonal lattice.

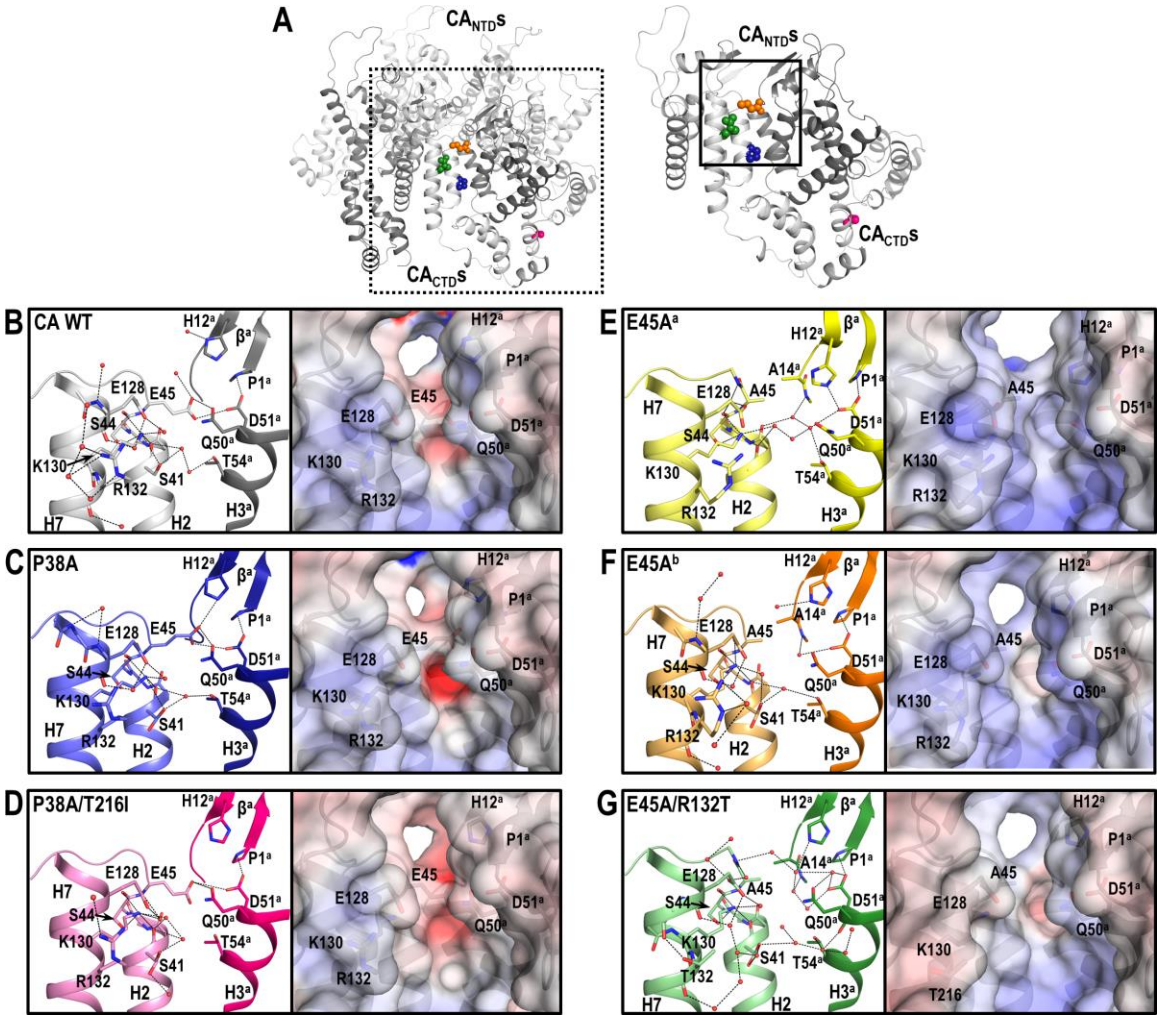


Figure 4-6. The effects of mutations on polar and water-mediated contacts around residue 45 and the salt bridge between P1 and D51.

(**A**) The orientation of CA WT hexamer with the side view of one representative monomer (white) and its interaction with the adjacent subunit (gray) outlined in dashed line and enlarged. Locations of mutation sites P38 (blue), T216 (pink), E45 (orange), and R132 (green) are shown as spheres. The effects of mutations in the region around residue 45 and the salt bridge between P1 and D51 (enlarged views of the boxed region, solid line) are shown in (**B-G**). Polar and water-mediated contacts in CA WT (**B**), P38A (**C**), P38A/T216I (**D**), E45A^{a} (**E**), E45A^{b} (**F**), and E45A/R132T (**G**). Black dashed lines represent putative hydrogen bonds with waters shown as red spheres. Selected side chains are shown explicitly and labeled. Surface representation of respective views is colored according to the electrostatic potential from −10 kBT/e (red) to +10 kBT/e (blue).

Comparison of the E45A^{a} and E45A^{b} with resolved β-hairpins revealed that they assume two opposite conformations that differ by ~11 Å (as measured by the displacement of Q7 Cα): 'open' in the E45A^{a}, and 'closed' in the E45A^{b} (Figure 4-7, A). Consistent with

the previous report based on the comparison of all available CA crystal structures (*280*), the two β-hairpin conformations are the result of a pivoting movement of ~39° about the N-terminal proline (P1). In the E45A^{b} structure where β-hairpin assumes 'closed' conformation, P1 forms a salt-bridge with D51 (*102*) (Figure 4-7, B). Conversely, in the E45A^{a} where β-hairpin assumes 'open' conformation, D51 also participates in a second salt-bridge interaction with H12. Moreover, six R18 in the middle of the E45A^{a} and E45A^{b} hexamers similarly adopt 'open' and 'closed' conformations (Figure 4-7, C and D), resulting in the pores with a diameter of ~11 Å and 5 Å, respectively.

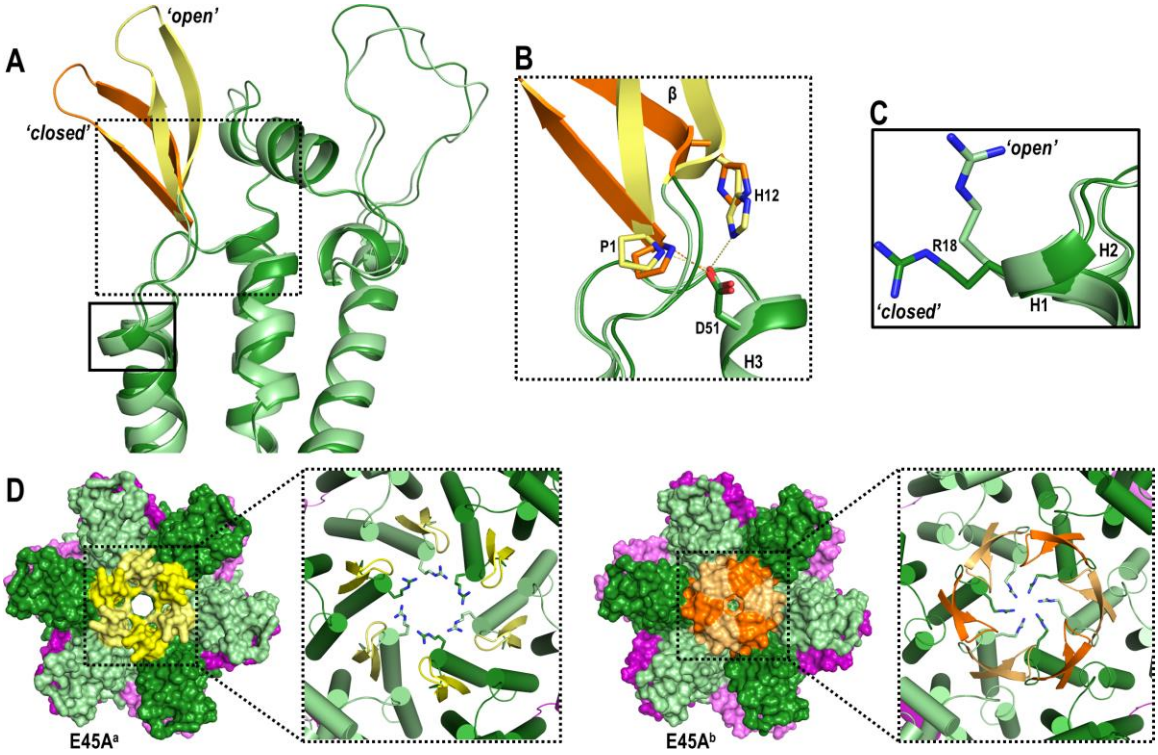


Figure 4-7. Changes in conformations of β-hairpin and R18.

(**A**) Superposition of N-terminal domains from E45A^{a} (CA_{NTD} in light green, β-hairpin in light yellow) and E45A^{b} (CA_{NTD} in green, β-hairpin in orange). (**B**) Detailed view of the boxed region (dashed line) shows that 'open' (light yellow) and 'closed' (orange) β-hairpin conformations are the result of the hydrogen-bond network around P1, H12, and D51. (**C**) Detailed view of the boxed region (solid line) shows that R18A toggles between 'open' and 'closed' state. (**D**) β-hairpin conformations dictate the presence of a pore at the six-fold axis. E45A^{a} and E45A^{b} hexamers are shown in surface view representation with CA_{NTD}s in light green and green, CA_{CTD}s in light purple and purple. β-hairpin in E45A^{a} structure is shown in light yellow and

yellow, while in E45A[b] in light orange and orange. The boxed region shows detailed view at the 6-fold axis in cartoon with R18 shown as sticks.

Thus, the effect of P38A and T216I, E45A and R132T mutations on CA structure is extremely complex affecting not only the sites of mutation but also modulating interactions within and between the CA hexamers.

Morphology studies using cryo-EM

It is known that the CA protein is capable of different assembly pathways *in vitro* depending on the protein concentration, pH, or ionic strength of the solution (*80, 81, 281*). The sensitivity of the higher-order structures to such factors indicates that polar interactions on the multimer surface regulate their formation (*81*). Previous studies of assembly properties of CA mutants revealed that P38A and E45A assembled with efficiencies similar to that of the CA WT (*93*) at protein concentrations 15, 10, and 5 mg/ml.

To evaluate the effects of compensatory mutations on assembly competence of CA mutants, we incubated CA proteins at a lower protein (2 mg/ml) and high salt (1 M) concentrations and analyzed cylinder formation in cryo-electron micrographs (cryo-EM). Under these conditions, the CA protein carrying P38A mutation does not assemble into tubular structures, but tubular assembly was rescued by the compensatory mutation T216I (Figure 4-8, A, B). The E45A CA mutant assembles into short tubes and cones (Figure 4-8, A, B), and again this assembly morphology was reverted to long tubes as the CA WT by the second compensatory mutation R132T. These results are consistent with the pelleting assay (Figure 4-8, C). The E45A assembles more efficiently compared to the CA WT and other CA mutants, while the assembly of P38A is largely impaired. These observations are consistent with characterizing P38A as an unstable, and E45A as a hyperstable, CA mutants

(*97*). Interestingly, under the conditions of this assay, compensatory mutations T216I and R132T, respectively, reverted the effect of the primary mutation on the assembly competence of mutant proteins.

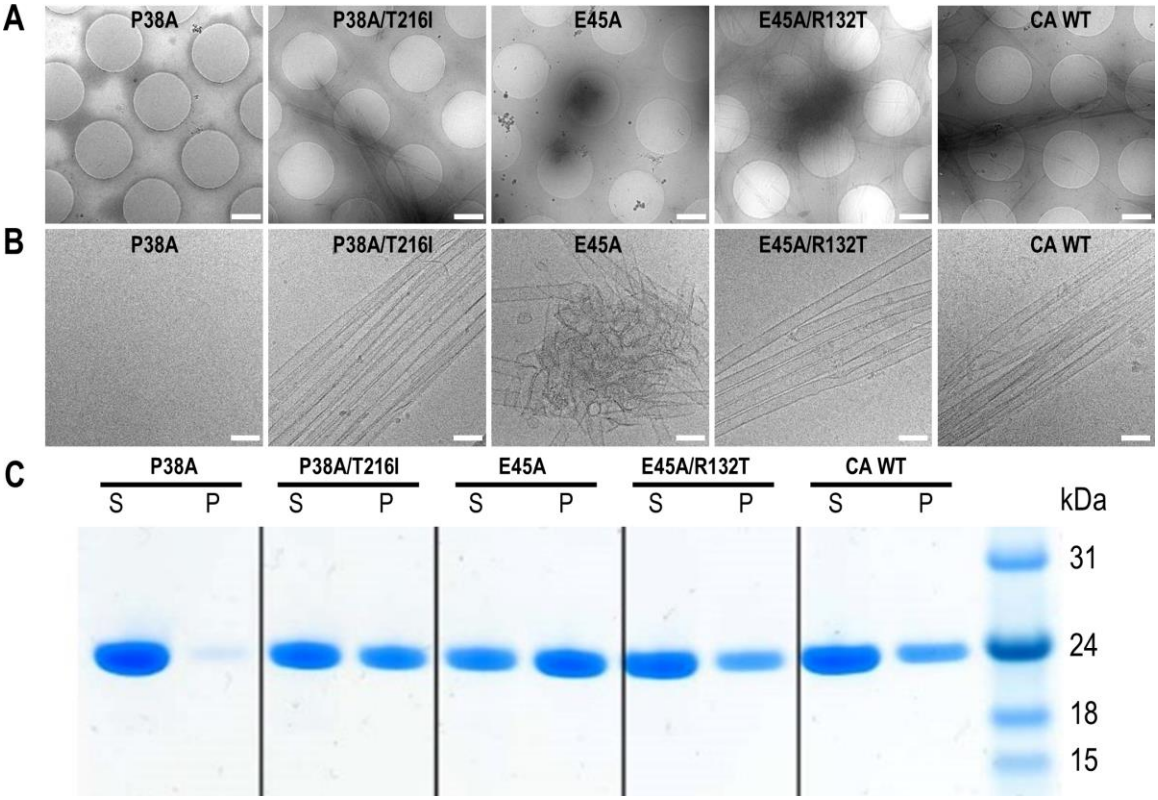


Figure 4-8. Effects of capsid mutations on assembly.

(**A-B**) Cryo-EM analysis of CA mutant assemblies. Projection images were recorded at low (**A**) and high (**B**) magnifications from the corresponding samples as indicated. Scale bars, 1 μm in (**A**), and 100 nm in (**B**), respectively. (**C**) Pelleting assay for CA mutant assemblies. Four CA mutants and CA WT are labeled. 'S' and 'P' stand for the supernatant and pellet from each sample. Protein products are visualized by Coomassie Blue staining. Molecular weight markers are labeled on the right.

Evaluation of core stability using CypA-DsRed loss assay

Biochemical studies have shown that detergent-treated purified HIV-1 capsids tend to disassemble in a temperature-dependent manner and point mutations in CA can strongly modulate core stability (*97*). Cores from P38A CA could not be recovered, presumably due to their reduced stability. Cores from E45A CA yielded greater quantities, were more

stable, and disassembled more slowly than CA WT (*97*). Notably, cores recovered from the double mutants P38A/T216I and E45A/R132T, behaved similarly to those of the corresponding single mutants (*273*).

Recently, a novel strategy to visualize HIV-1 uncoating using a fluorescently tagged oligomeric form of a capsid-binding host protein cyclophilin A (CypA-DsRed) – which is specifically packaged into virions through the high-avidity binding to CA – has been developed (*282*). Single virus imaging revealed that CypA-DsRed remained associated with cores after permeabilization/removal of the viral membrane and that CypA-DsRed and CA were lost concomitantly from the cores *in vitro* and living cells. The rate of loss was modulated by the core stability and was accelerated upon the initiation of reverse transcription.

To evaluate the effects of P38A, P38A/T216I, E45A, and E45A/R132T CA mutations on the uncoating and core stability, we employed CypA-DsRed loss assay. CypA-DsRed was lost from the permeabilized E45A particles much slower than from CA WT (Figure 4-9, A, B), whereas virtually all CypA puncta disappeared immediately after permeabilization of the P38A viruses (Figure 4-9, C). Notably, for the E45A mutation, the second site suppressor mutation (R132T) does significantly decrease the core stability, but its stability is still much greater than that of the CA WT. Consistent with previous reports (*273, 283, 284*), the second site suppressor mutations do not fully correct the stability defects imposed by E45A or P38A.

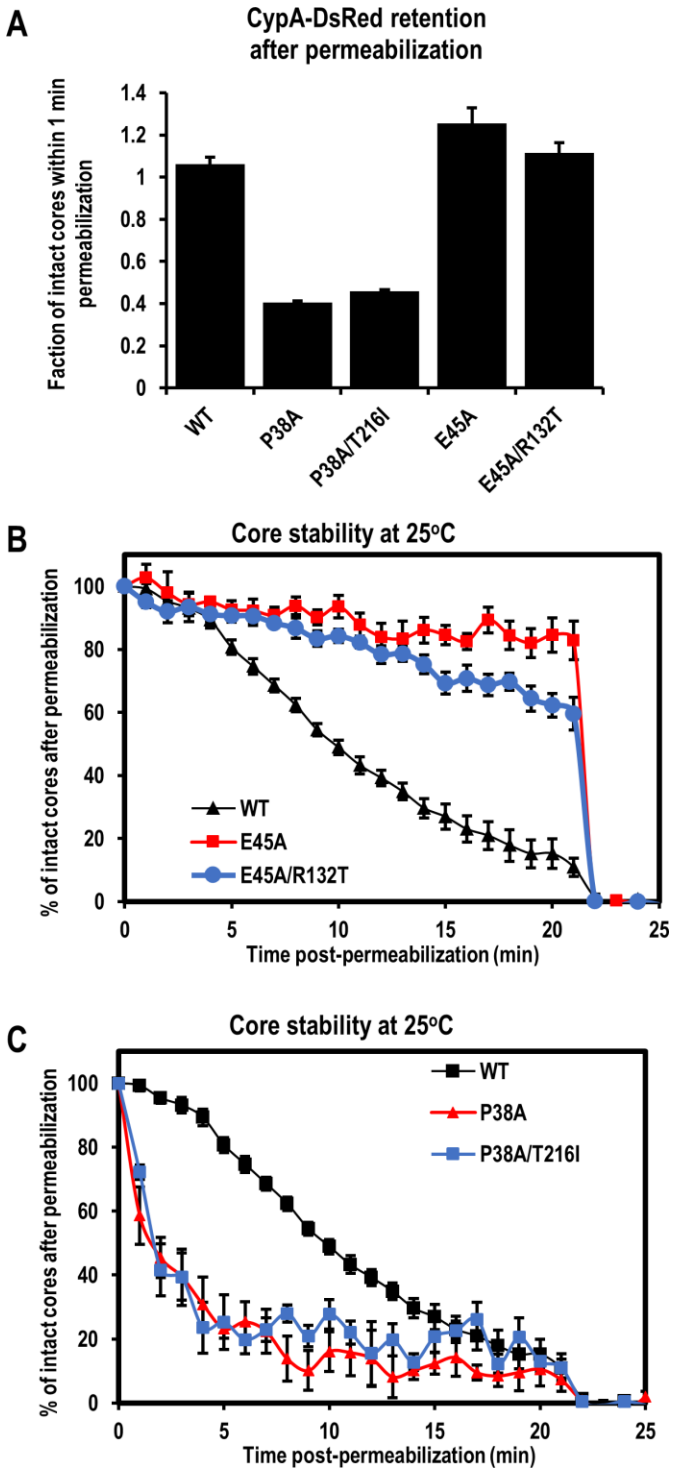


Figure 4-9. Effects of capsid mutations on core stability.

(**A**) A snapshot of core stability right after permeabilization. (**B-C**) show the kinetics of a one-step loss pfCypA-DsRed from cores over time.

Implications for the structural mechanism of HIV-1 capsid stability and uncoating

Uncoating, or disassembly of the viral capsid, is a critical, yet poorly understood step of the retroviral life cycle. In the present work, we observed that the T216I and R132T mutations can at least partially correct *in vitro* assembly defects imposed by P38A and E45A, respectively, without correcting the altered capsid stability imposed by the original mutations. As suggested previously (*273*), one possible explanation for the discrepancy would be that the available assays, even improved over time, are limited in their ability to detect subtle changes in capsid stability. Alternatively, the suppressor mutations alter interactions with host factors that participate in HIV-1 uncoating in target cells.

The core stability is intrinsically connected to the strength of the interactions between the building blocks and between the monomers within each building block, with the latter having a larger impact on the CA stability (*275*). Our structural analysis of P38A, P38A/T216I, E45A, and E45A/R132T CA proteins revealed no major structural rearrangements in HIV-1 CA.

The P38A mutation by itself, or present together with T216I, initiates a cascade of subtle changes that result in remodeling network of CA-CA interactions in a hexamer, which lead to a decrease in its stability (Figure 4-2). T216I, being a mutation at the 3-fold inter-hexamer interface, induces further subtle rearrangements that may result in stabilization of interactions between hexamers, thereby partially offsetting the effect of the P38A (Figure 4-3). Oppositely, E45A primarily remodels network of polar and water-mediated interactions proximal to the site of mutation. The second site mutation, R132T,

located in the vicinity of E45A, partially restores impaired water mediated interactions (Figure 4-6). Notably, the common feature for all those mutants is altered interactions around the residue 45.

Additionally, P38A (Figure 4-10) and, to a greater extent, E45A (Figure 4-11) mutations alter surface electrostatic potential in CA hexamers. Compensatory mutations T216I and R132T, respectively, may partially restore the overall net charge of CA. Those local changes in surface electrostatic potential could explain differences in susceptibility to PF74, interactions with host factors, and as a result dependence of the infection on the cell cycle (for details see Table 4-1).

The T216I mutation has been shown to reverse the impaired ability of the P38A mutant to abrogate TRIMCyp- and TRIM5α-mediated restriction of HIV-1 (Table 4-1) (*273*). Recently, it has been demonstrated that TRIM5α binding involves multiple capsid molecules along the 2-fold and 3-fold inter-hexamer interfaces. Thus, structural changes at the 3-fold interface in P38A/T216I may explain the rescue of the ability to abrogate TRIM5 restriction.

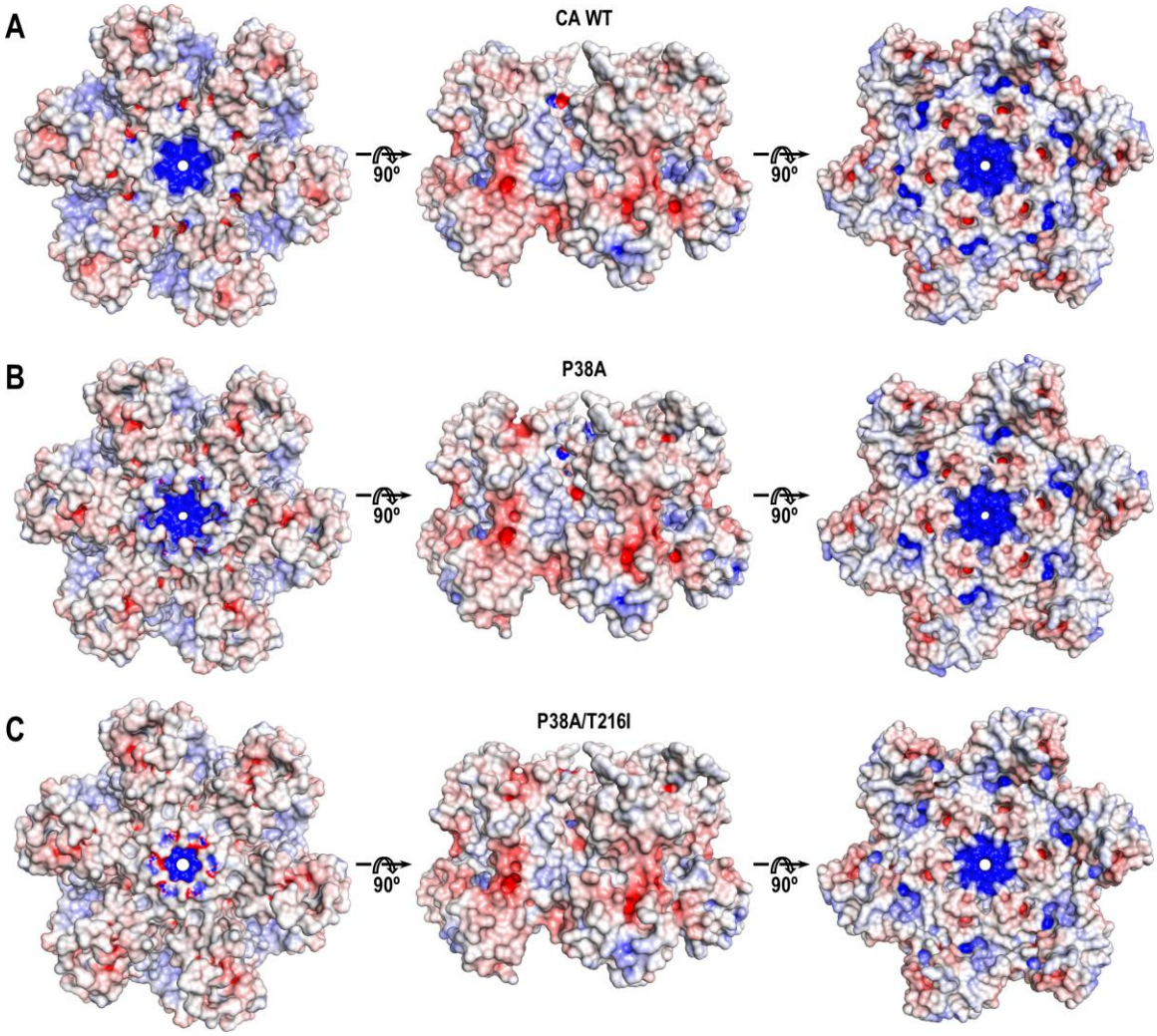


Figure 4-10. Surface electrostatic potential of CA WT, P38A, and P38A/T216I.

Surface representation of CA WT (**A**), P38A (**B**), and P38A/T216I (**C**) hexamers with alternate orthogonal views colored according to the electrostatic potential from −10 kBT/e (red) to +10 kBT/e (blue).

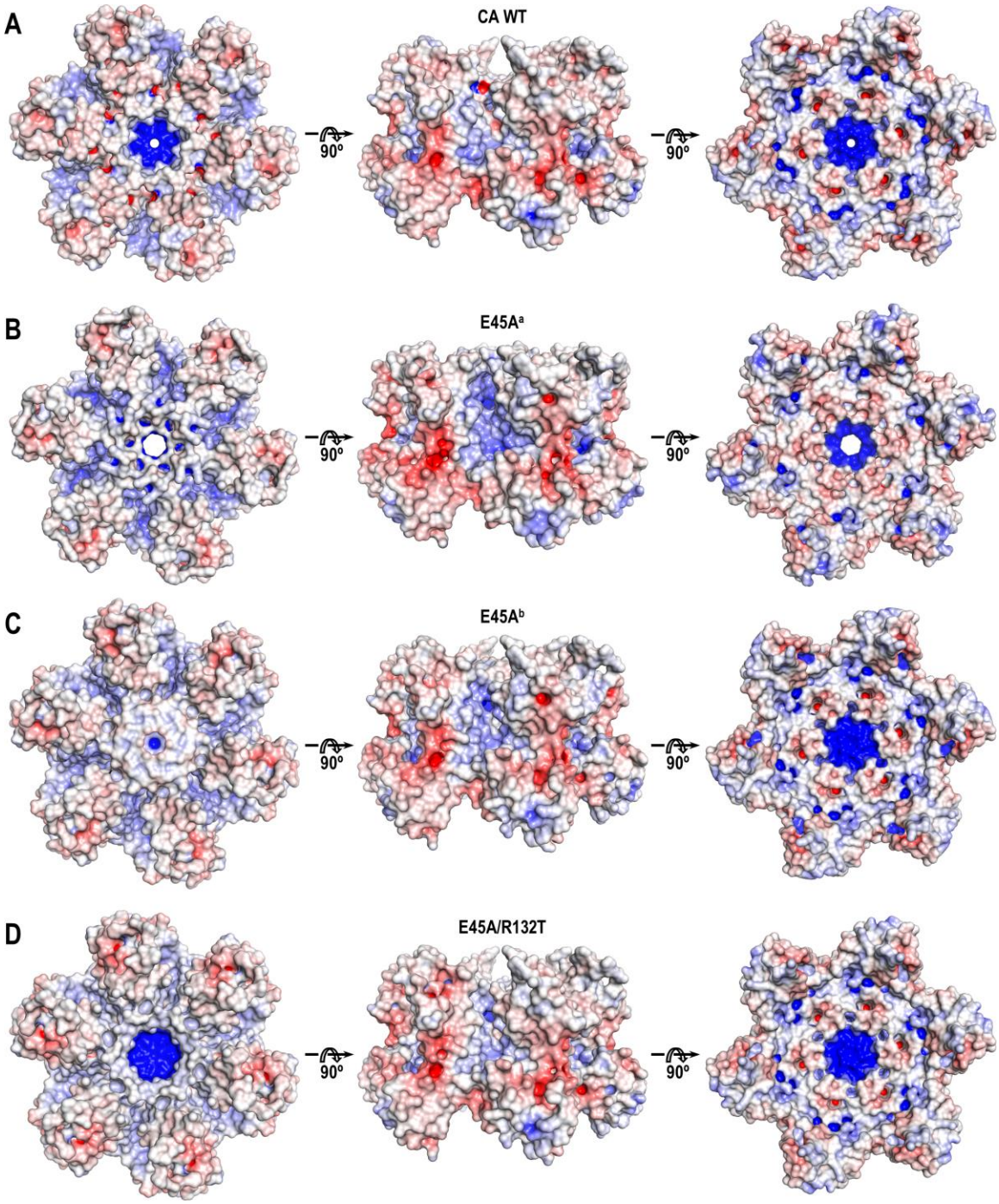


Figure 4-11. Surface electrostatic potential of CA WT, E45A, and E45A/R132T.

Surface representation of CA WT (**A**), E45A[a] (**B**), E45A[b] (**C**), and E45A/R132T (**D**) hexamers with alternate orthogonal views colored according to the electrostatic potential from −10 kBT/e (red) to +10 kBT/e (blue).

Surprisingly, the E45A mutant has been crystallized in the two different crystal forms with the overall fold of the protein being essentially the same, but one structure being tighter (E45A^{a}) than the other (E45A^{b}). The tighter structure (E45A^{a}) was additionally stabilized by the novel 3-fold inter-hexamer interface formed between N-terminal domains of the protein. This may explain the hyperstabilization of the E45A capsid.

Interestingly, the two structures are characterized by the two different β-hairpin conformations. In the E45A^{a} β-hairpin assumes 'open' conformation, creating a chamber along the 6-fold axis in a hexamer, which culminates in a ring of six positively charged arginine side chains from residue 18 (R18). The R18 residues adopt different conformations in the two structures as well. Similar to β-hairpin, R18 assumes 'open' conformation in the tighter E45A^{a} structure, giving rise to a pore in the middle of a hexamer (Figure 4-7). The diameter of the pore (~11 Å) should be sufficient to allow transit of negatively charged small molecules (including deoxynucleoside triphosphates (dNTPs)) to the interior of HIV-1 E45A capsid.

The existence of the pore has been previously proposed based on comparing all of the available CA crystal structures (*280*) with resolved β-hairpins. Primary residues responsible for the hairpin movement were identified as P1, D51, and H12, with the protonation state of H12 being crucial in determining which arrangement was favored and therefore suggesting that it is pH regulated (*280*). Our two E45A structures indicate that β-hairpin can assume two opposite conformations under very similar conditions at pH 6.5, confirming maximum pore flexibility under physiological conditions.

It has been previously observed that E45A cores appear to be more permeable to fluorescent dyes than CA WT cores (*276*). Moreover, E45A HIV-1 exhibited a rapid viral

RNA decay profile and a more rapid accumulation of early reverse transcripts. As a result, it was concluded that the E45A capsid dissociated early after infection (*276*). Our new data provide evidence that pores can exist in the intact E45A capsid that may allow an enhanced influx of various small molecules, including dyes and dNTPs, which explain the observed phenotypes.

It has been previously proposed that the highly hydrated character of the CA is compatible with the quasi-equivalent switching mechanism because water molecules should be particularly adept at repositioning to accommodate altered orientations in hydrogen bonding and side chain packing geometries (*64, 95*). This hypothesis is supported by the two different crystal structures of the E45A that differ substantially in the structured water molecules content. The E45A^{a} (resolution 2.5 Å) with tighter interactions has 60 waters modeled, while the E45A^{b} structure (resolution 2.2 Å) has 98 waters; resolution by itself may not account for such significant difference in hydration.

Our crystal structures may also help explain the assembly phenotype of the E45A mutant. The presence of short tubes and cones indicates that E45A accommodates curvature more easily than CA WT, and occasionally facilitates the incorporation of pentamers into the assembling hexagonal lattice. It has been proposed that electrostatic forces control switching between the pentamer and hexamer (*95*). As indicated above, the point of closest approach occurs between R18 at the top of helix H1, which is well accommodated in the hexamer. Pentamer formation, however, brings charged R18 at the center of the ring in closer proximity than in a hexamer, creating a stronger electrostatic repulsion. Relief of the repulsion is expected to favor pentamer formation; indeed, the

ability of R18 to assume 'open' conformation in the E45A[a] promotes assembly of the curved particles.

In the crystal structure of CA WT, glutamic acid E45 is engaged in both attractive and repulsive ionic interactions with aspartic acid D51 from the neighboring subunit. Mutation of E45 to alanine directly relieves electrostatic repulsion between those two residues, thus resulting in significant stabilization of the E45A CA hexamer and subsequently of the viral core itself. Elimination of these forces also allowed us to capture two opposite orientations of β-hairpin under essentially the same conditions. The effect of the P38A mutation is a lot more complex. In addition to altering the intra-hexamer interfaces proximal or downstream to the site of mutation, it also employs the E45 residue, altering the network of interactions that presumably increases the electrostatic repulsion between E45 and D51, which leads to the destabilization of the P38A CA hexamers and the core.

The second site suppressor mutations, R132T and P216I, can partially offset the effect of primary E45A and P38A mutations. R132T partially restores the overall net charge of CA, while T216I stabilizes inter-hexamer interactions. Moreover, consistent with the previous hypothesis (*268*), structured water molecules also contribute to the capsid stabilization. The hyperstable E45A mutant has at least twice as many modeled water molecules than the unstable P38A mutant.

Thus, we suggest that the residues E45 and D51 are the key players controlling the conformation of the β-hairpin, the presence of the pore, as well as the stability of the core and its disassembly. Mutual electrostatic repulsion between pairs of carboxylate groups forced into proximity by the CA structure provides an environmentally-sensitive switch,

active under physiological conditions, that can control the state of assembly and disassembly (or uncoating) of the viral capsid. This mechanism is additionally supported by site-directed mutagenesis data, as replacing carboxylate group (D51) by the corresponding amide (N) results in capsids that exhibit significantly increased stability compared to CA WT (*285*). Moreover, D51N formed long tubular structures *in vitro* comparable to CA WT, both in terms of external diameter and length of the tubes (*286*).

This structural mechanism is reminiscent of the role of carboxylate groups in the life cycle of the tobacco mosaic virus (TMV) (*287*). When a TMV virion enters the target cell, it experiences significant changes in calcium ions concentration and pH. Upon removal of protons and calcium ions from carboxyl-carboxylate and carboxylate-phosphate pairs, the electrostatic repulsion between their side chains in adjacent subunits drives the disassembly of the TMV coat protein.

Collectively, the structures support our hypothesis that CA plasticity is a key factor for its stability, and how challenging it is to fully understand the effect of even a single mutation on this highly flexible protein. The possibility of capsid stability regulation through changes in pH, dNTP recruitment, and DNA synthesis, provides a model whereby DNA synthesis is coordinated with uncoating to cloak the viral genome from cytoplasmic DNA sensing.

Chapter 5. Identification of a novel element in HIV-1 CA critical for assembly and maturation

Note: This chapter is based on a manuscript that is currently in preparation. The contributions of co-authors are described in the "Footnotes."

A. Introduction

The assembly and maturation of HIV-1 and other retroviruses are driven by the viral structural polyprotein, Gag, which consists of several independently folded domains matrix (MA), capsid (CA), and nucleocapsid (NC), connected by flexible linkers (*107*). MA domains direct Gag to the plasma membrane, CA domains mediate protein-protein interactions, while NC domains bind the RNA. Upon assembly at the cellular membrane, Gag makes a spherical, immature capsid shell around the viral genome. During maturation, Gag is cleaved by the viral protease, which results in the rearrangement of newly released CA proteins into the mature capsid core and formation of the infectious virions. CA has two independently folded domains: N-terminal (CA_{NTD}) and C-terminal (CA_{CTD}), connected by a short linker. Due to the multiple roles of CA in the virus life cycle, it has been increasingly viewed as an attractive therapeutic target. A number of compounds targeting CA and inhibiting HIV have already been identified (*267*), including CAP, BD, BM, BI, and PF74 (*110, 116, 121, 122*).

The arrangement of CA in immature HIV-1 is very different from the one in mature HIV-1 capsid core (Figure 5-1) (*143, 288*). The mature capsid consists of ~250 hexamers and 12 pentamers that allow forming the closed shape. The mature CA hexamer comprises a central ring formed by the six adjacent CA_{NTD}s, and an external ring of CA_{CTD}s that link

hexamers together (Figure 5-1, A) (*64, 95*). Hexagonal lattice is the foundation of the mature capsid. It is stabilized by intra-hexamer CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} contacts, while inter-hexamer interactions are mediated *via* 2-fold and 3-fold CA_{CTD}-CA_{CTD} contacts (*289*).

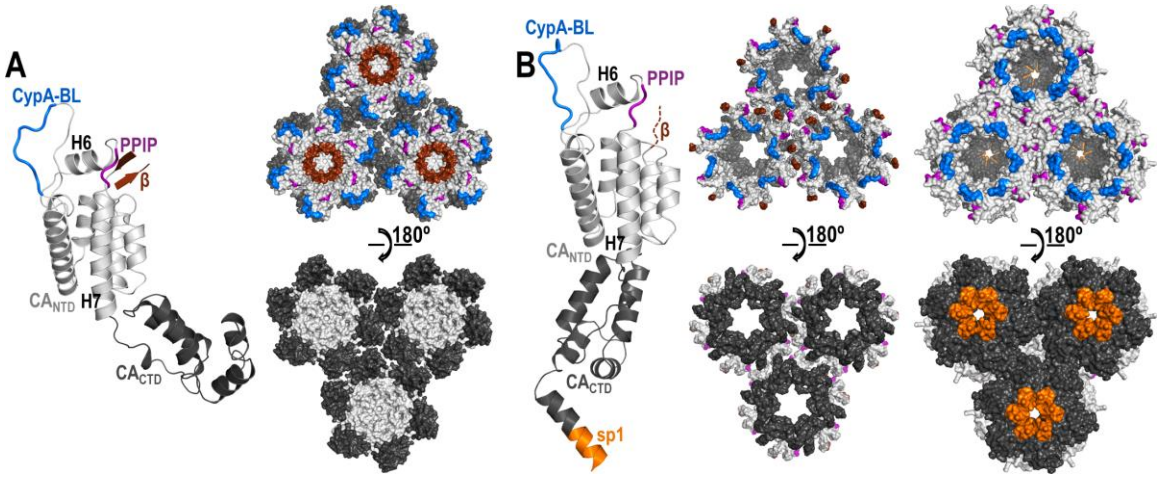


Figure 5-1. HIV-1 mature and immature hexagonal lattices.

(**A**) Mature HIV-1 CA lattice (PDB ID: 4XFX). CA protein is shown in ribbon with CA_{NTD} in white and CA_{CTD} in gray. CypA-binding loop (CypA-BL) is highlighted in blue, PPIP motif in purple, and β-hairpin in brown. Two opposite views of neighboring hexamers in the mature CA lattice are shown in surface. (**B**) Immature HIV-1 Gag lattice. The central domain of Gag polyprotein is shown in ribbon with CA_{NTD} in white, CA_{CTD} in gray, and sp1 in orange (PDB ID: 5L93). Residues that comprise β-hairpin in the mature CA (brown dashed line) form an extended strand in Gag and are disordered in the inspected structures. Two opposite views of neighboring hexamers in the immature Gag lattices from PDB ID: 4USN (model includes only Cα tracing) and PDB ID: 5L93 are shown in surface.

Unlike the mature capsid, the immature HIV-1 Gag shell is a more smoothly curving sphere that is interrupted by large discontinuities (*290*). Likewise, the Gag shell has hexagonal paracrystalline lattice symmetry. Oppositely, the immature CA hexamer consists of a central ring formed by the six adjacent CA_{CTD}-sp1, and an external ring formed of CA_{NTD}s that link hexamers together (Figure 5-1, B) (*142, 143, 288*). The lattice is stabilized by intra-hexamer CA_{CTD}-CA_{CTD}, and, to some extent, CA_{NTD}-CA_{NTD} contacts, while inter-hexamer interactions are mediated *via* homodimeric and homotrimeric CA_{NTD}-

CA_{NTD} interfaces. A homodimeric interface is formed between helices H1 from two adjacent hexamers. A homotrimeric, three-helix bundle interface, is formed by helices H2 from neighboring hexamers (*288*). Notably, unlike in the mature capsid, there is no large CA_{NTD}-CA_{CTD} interface.

Over the years enormous research effort has been invested in determining the phenotypes of natural and artificial mutations in HIV proteins, including CA (*272*). Those studies have been significant for defining various CA functions and mapping them to the CA_{NTD} and CA_{CTD} (*92, 93, 97, 102*). Early studies determined that mutations introduced into the loop between helices H6 and H7 (Pro-Pro-Ile-Pro, or PPIP motif) were lethal or caused decrease in infectivity; however, they were not studied in detail (*291*). Notably, this motif is highly conserved and is proximal to the homotrimeric interface of the immature Gag lattice (Figure 5-1). However, its role in the assembly and maturation of HIV-1 is poorly understood.

B. Materials and methods

Protein expression and purification

P122A, I124A, T58A/I122A, V11I/T58A/P122A, V11I/T58A/I124A, and T58S/T107I/P122A CA were cloned without fusion tags in pET11a plasmid, and provided by M. Novikova. Mutant proteins were expressed and purified as previously described (*268*).

Crystallization of CA mutants

Crystals of the P122A, I124A, T58A/I122A, V11I/T58A/P122A, V11I/T58A/I124A, and T58S/T107I/P122A CA grew at 18 $^{\circ}$C in hanging drops, containing

5-10 mg/ml of protein, 6-9 % PEG 3350, 2-6 % glycerol, sodium iodide, and sodium cacodylate. Hexagonal plate-like crystals appeared after five days, and crystal growth was completed in over two weeks. Crystals were briefly soaked in 20 % glycerol or paraffin oil before cryo-cooling in liquid nitrogen.

Data collection and structure determination

Data were collected on a CMOS detector at Advanced Light Source (ALS) beamline 4.2.2, Lawrence Berkeley National Laboratory or on a MAR CCD or Dectris Eiger-16m detector at the Advanced Photon Source (APS) beamline 23 ID-B, Argonne National Laboratory. Datasets were collected and processed using XDS (*247*). The data were examined for the presence of systematic absences; however, no characteristic patterns were observed. Thus, the crystals were indexed in hexagonal space group P6 with one CA molecule in the asymmetric unit. No twinning was present, as determined by either POINTLESS (*249*) or XTRIAGE (*250*). Space group and twinning were also verified in ZANUDA (*251*). The phase problem of CA was solved either by single-wavelength anomalous diffraction (SAD), or molecular replacement, with the native CA (PDB ID: 4XFX) as a starting model. For SAD, substructure solution, phasing, density modification, model building, and refinement were carried out using SHELX C/D, SOLOMON, PARROT, BUCCANEER, and REFMAC in CRANK-2 (*251*). For molecular replacement, initial phases were solved *via* PHASER (*251*). Several rounds of iterative model building and refinement were carried out using Coot (*253*) and PHENIX (*250*), REFMAC (*251, 252*), or PDBREDO (http://www.cmbi.ru.nl/pdb_redo/), respectively. Structure validation of final models was performed with MOLPROBITY

(http://molprobity.biochem.duke.edu/). Accessible and buried surface areas were calculated using PISA (*251*). The figures showing structural information were generated in PyMOL (http://www.pymol.org/) (*251*).

C. Results and discussion

Summary of biological data

We performed alanine-scanning mutagenesis of the PPIP motif and showed that mutations P122A and I124A significantly impaired virus release, particle infectivity, and replication in T-cell lines (Table 5-1). Also, using electron microscopy and cryo-electron tomography techniques, we demonstrated that these two mutants displayed obvious defects in the formation of Gag lattice in both protease-containing and protease-deficient virions. In contrast, mutations P123A and P125A were reasonably well tolerated. To better understand the role of the PPIP motif in CA structure and function, we selected for compensatory mutations that rescue the functional defects caused by the P122A and I124A substitutions. Two mutations, V11I and T58A, when present together, rescue the defects imposed by both P122A and I124A such that V11I/T58A/P122A and V11I/T58A/I124A triple mutants are assembly competent, infectious, and replicate with near-WT kinetics. One of the mutants, T58S/T107I/P122A, displayed nearly WT levels of infectivity and WT particle morphology (Table 5-1); however, this mutant replicated with a significant delay in Jurkat cells. We showed that this mutant virus is hypersensitive to the CA-targeting antiviral PF74.

Table 5-1. Infectivity of the CA mutants.

	Infectivity, %	MT4 replication	Jurkat	PF74
P122A	1-6	-	-	
I124A	1-5	-	-	
T58A/I124A	40-80	+	+ (2-8 days delay)	
T58S/T107I/P122A	70-100	+	+ (~10 days delay)	more sensitive
V11I/T58A/P122A	40-65	+	+ (2-6 days delay)	
V11I/T58A/I124A	40-80	+	+ (2-6 days delay)	

Crystallographic analysis

To determine whether the above-mentioned mutations alter the structure of the mature HIV-1 CA lattice, we crystallized CA proteins bearing P122A, I124A, T58A/I124A, T58S/T107I/P122A, V11I/T58A/P122A, or V11I/T58A/I124A mutations. The structures have been solved in the P6 space group with one molecule/asymmetric unit (Table 5-2). Overall, the mutant structures are similar to the CA WT (rmsd: 0.394 Å for P122A; 0.400 Å for I124A; 0.452 Å for T58A/I124A; 0.381 Å for T58S/T107I/P122A; 0.313 Å for V11I/T58A/P122A; 0.406 Å for V11I/T58A/I124A), indicating that mutations do not disrupt the fold of the CA protein.

Table 5-2. Summary of data collection statistics and refinement.

	P122A	I124A	T58A/I124A	V11I/T58A/P122A	V11I/T58A/I124A	T58S/T107I/P122A	T58S/T107I/P122A$_{PF74}$
Data collection							
X-ray source	APS 23 ID-B	ALS 4.2.2	ALS 4.2.2	ALS 4.2.2	ALS 4.2.2	APS 23 ID-B	ALS 4.2.2
Software	XDS	XDS	XDS	XDS	XDS	XDS	XDS
Space group	P6	P6	P6	P6	P6	P6	P6
Unit cell dimensions							
a, b, c (Å)	92.7 92.7 57.8	92.6 92.6 58.0	91.7 91.7 57.7	92.5 92.5 57.7	91.9 91.9 57.9	92.4 92.4 57.5	93.0 93.0 57.6
α, β, γ (°)	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0
ASU content	1	1	1	1	1	1	1
Wavelength (Å)	1.033203	1.000031	1.000031	1.000032	1.000031	1.033760	1.000031
Resolution range (Å)[a]	46.9–2.3 (2.4–2.3)	58.0–2.4 (2.5–2.4)	45.8–2.6 (2.7–2.6)	57.7–2.4 (2.5–2.4)	46.0–2.8 (2.9–2.8)	46.7–2.4 (2.5–2.4)	46.5–2.8 (2.9–2.8)
R_{merge}	0.060 (0.633)	0.113 (0.750)	0.107 (>1)	0.078 (0.781)	0.109 (>1)	0.071 (>1)	0.104 (0.932)
R_{meas}	0.066 (0.699)	0.119 (0.797)	0.112 (>1)	0.082 (0.829)	0.115 (>1)	0.075 (>1)	0.110 (0.995)
R_{pim}	0.028 (0.293)	0.036 (0.267)	0.034 (0.466)	0.025 (0.270)	0.035 (0.345)	0.026 (0.490)	0.034 (0.341)
$\langle I/\sigma I \rangle$	14.9 (2.6)	15.8 (2.6)	18.3 (1.6)	20.7 (2.4)	18.0 (2.1)	17.7 (1.6)	19.1 (1.9)
$CC_{1/2}$ (%)	99.8 (83.3)	99.8 (80.8)	99.9 (57.7)	99.8 (81.5)	99.9 (75.4)	99.9 (49)	99.8 (71.4)
Completeness (%)	99.9 (100)	97.2 (96.8)	99.5 (96.0)	98.2 (98.1)	97.3 (97.3)	94.1 (95.2)	95.9 (86.5)
Redundancy	5.7 (5.6)	10.6 (8.4)	10.6 (9.7)	10.5 (8.8)	10.6 (9.9)	8.3 (8.1)	10.2 (7.8)
Mosaicity	0.15	0.19	0.36	0.29	0.20	0.21	0.24
Refinement							
Resolution (Å)	46.9–2.3	58.0–2.4	45.8–2.6	36.1–2.4	46.0–2.8	46.7–2.4	46.5–2.8

	P122A	I124A	T58A/I124A	V11I/T58A/P122A	V11I/T58A/I124A	T58S/T107I/P122A	T58S/T107I/P122A$_{PF74}$
No. total reflections	72287	115437	91517	114114	73644	86620	72267
No. unique reflections	12735	10933	8617	10917	6969	10448	7089
No. test reflections[b]	704	643	512	629	383	500	460
R_{work} / R_{free}	24.0 / 26.8	18.9 / 23.6	19.0 / 24.7	22.8 / 26.6	21.3 / 26.0	22.4 / 24.3	22.8 / 27.9
No. atoms	1747	1856	1726	1768	1729	1749	1743
Protein	1682	1679	1661	1681	1683	1693	1676
Ligand/Ion	13	19	17	16	16	15	46
Water	52	158	48	71	30	41	21
Wilson B-factor (Å^2)	56.5	30.7	45.9	43.1	48.5	51.5	52.1
Average B-factors (Å^2)	77.8	49.2	70.5	61.1	76.1	79.6	67.4
Protein	78.0	49.2	70.8	61.4	76.5	79.8	68.2
Ligand/Ion	79.7	62.3	79.3	73.9	82.6	90.8	48.1
Water	69.9	47.3	55.4	51.1	47.3	67.2	44.5
RMS deviations							
Bond lengths (Å)	0.002	0.007	0.009	0.004	0.007	0.002	0.003
Bond angles (°)	0.51	1.06	1.31	0.64	1.08	0.48	0.57
MolProbity Statistics[c]							
All atom clash score	2.07	1.47	1.50	5.63	1.18	4.11	3.50
Rotamer outliers (%)	0	0	0	0	0	0	0
Cβ deviations >0.25 Å	0	0	0	0	0	0	0
Ramachandran[c]							

	P122A	I124A	T58A/I124A	V11I/T58A/P122A	V11I/T58A/I124A	T58S/T107I/P122A	T58S/T107I/P122A$_{PF74}$
Favored region (%)	99	99	98	99	98	98	98
Outliers (%)	0	0	0	0	0	0	0
PDB accession code	6AXR	6AXW	6AXX	6AXS	6AXY	6AXT	6AXV

[a] Values in parentheses are for highest-resolution shell; [b] random selection; [c] values obtained from MOLPROBITY.

Rather, mutations result in the structural rearrangement in the proximity or at the site of mutation. Specifically, the crystallographic analysis revealed that P122A and I124A result in changes subtle variations in the PPIP motif and induce subtle repositioning of residues 92-96 of the proximal CypA-binding loop (CypA-BL) (Figure 5-2).

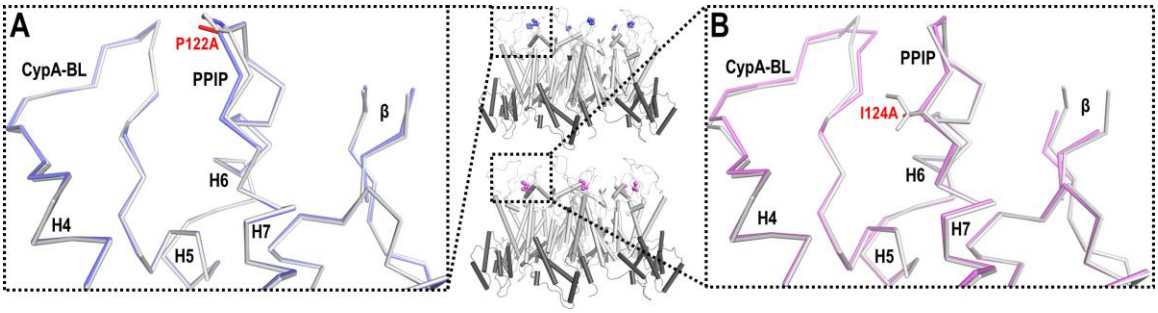


Figure 5-2. P122A and I124A subtly affect PPIP motif and CypA- binding loop.

CA WT hexamers (CA_{NTD}s in white, CA_{CTD}s in gray) with mutation sites (in spheres) are shown in cartoon. Small boxes show close-up views of changes induced by (**A**) P122A (in light blue) and (**B**) I124A (in light pink) mutations. The mutant proteins (shown Cα tracing) are superposed with CA WT (in white).

Analysis of the crystal structures of T58A/I124A, T58S/T107I/P122A, V11I/T58A/P122A, and V11I/T58A/I124A, revealed that T58A and T58S mutations altered conformations of L20 and M39 from neighboring CAs, thus affecting CA_{NTD}-CA_{NTD} intra-hexamer interface (Figure 5-3). Notably, residue M39 in helix H2 has been previously identified to be crucial for CA assembly and capsid formation (*92, 93*).

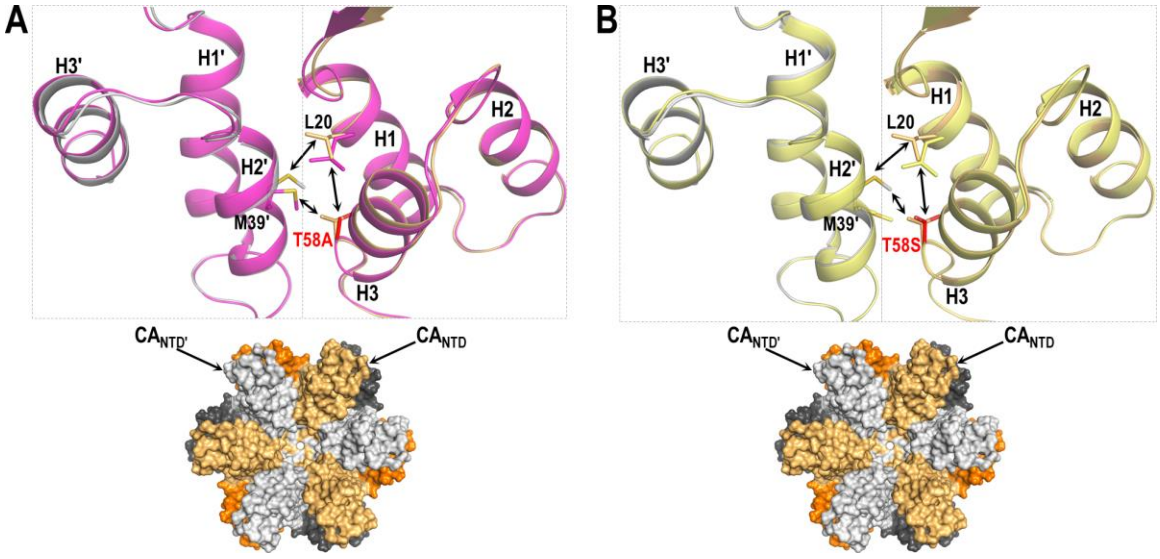


Figure 5-3. T58A in T58A/I124A, V11I/T58A/I124A, and V11I/T58A/P122A or T58S in T58S/T107I/P122A mutants alter the network of inter-subunit interactions in a hexamer.

CA WT hexamers (alternate CA_{NTD}s in white and light orange, CA_{CTD}s in gray and orange) are shown in surface. (**A**) An enlarged view of changes induced by T58A mutation. Representative T58A/I124A mutant (in pink) superposed with CA WT (in white and light orange). Two neighboring CA subunits are outlined for clarity, and their elements are marked with prime symbols. (**B**) An enlarged view of changes induced by T58S mutation. T58S/T107I/P122A mutant (in light yellow) superposed with CA WT. Affected residues are in sticks with sites of mutations highlighted in red.

V11I/T58A/P122A and V11I/T58A/I124A mutant structures suggest that V11I mutation may help stabilize the β-hairpin (Figure 5-4). However, this region is rather flexible in our structures. Thus, complete information of the V11I effects on the mature CA is hard to discern. Notably, V11I mutation is proximal to the PPIP motif in the context of the immature Gag lattice (Figure 5-1). Thus, it may help to overcome the defects imposed by P122A and I124A on the immature HIV.

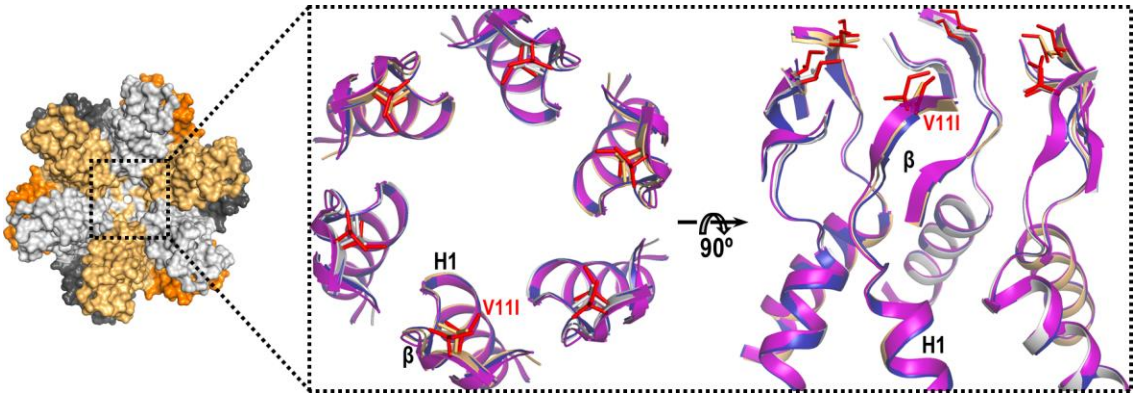


Figure 5-4. V11I mutation may affect β-hairpin in V11I/T58A/I124A and V11I/T58A/P122A.

CA WT hexamer is shown in surface with alternate CA_{NTD}s colored in white and light orange and CA_{CTD}s in gray and orange. The boxed region shows orthogonal enlarged views of the six β-hairpins. V11I is shown in sticks.

Finally, examination of T58S/T107I/P122A structure, suggests that PF74 susceptibility is likely affected by changes in the PF74 binding pocket, directly *via* T107I and/or indirectly *via* the T58S mutation (Figure 5-5, A). To address this and explain the increased sensitivity of the triple mutant to PF74, we have solved the additional crystal structure of the complex. The strong, interpretable electron density was observed for the compound (Figure 5-5, B). As established previously for the CA WT in complex with PF74, the CA-targeting antiviral binds across two CA monomers in a hexamer. The increased sensitivity of T58S/T107I/P122A towards PF74 may be explained by additional hydrophobic interactions between I107 and phenyl ring of the compound.

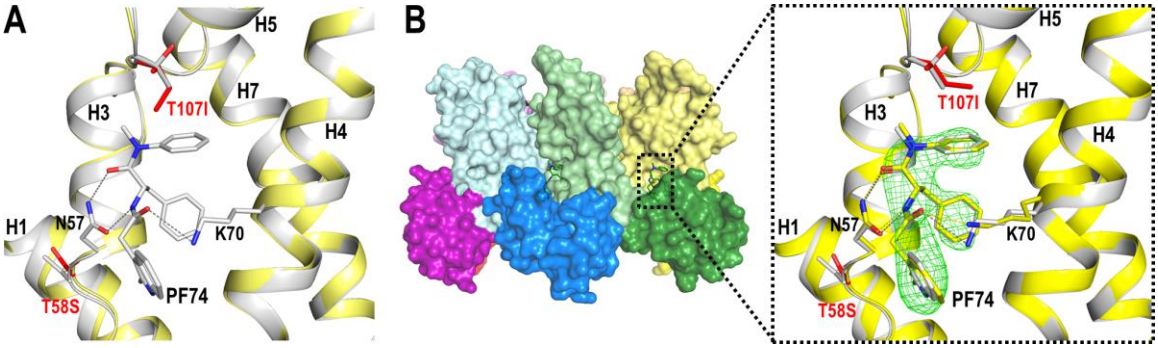


Figure 5-5. Structure of T58S/T107I/P122A and its PF74 complex.

(**A**) An enlarged view of the PF74-binding site. T58S/T107I/P122A mutant (in light yellow) superposed with CA WT complex with PF74 (in white). Dashed lines indicate H-bond interactions with involved residues shown in sticks and labeled. Mutation sites are highlighted in red. (**B**) T58S/T107I/P122A CA hexamer in complex with PF74 is shown in surface representation with CA_{NTD}s and the corresponding CA_{CTD}s colored by the same colors (light and dark, respectively). The boxed region shows an enlarged view of the PF74 binding site. T58S/T107I/P122A in complex with PF74 (in yellow) superposed with CA WT PF74 complex. PF74 is modeled in a 2.8 Å simulated annealing composite omit map (σ=3.0).

Interestingly, we have noticed that residue Q176 in the P122A mutant assumes different conformation when compared to CA WT. Its conformation is similar to the one observed for CA WT in complex with host cell factor nucleoporin Nup153 (Figure 5-6, A-C). Notably, in all other mutants, despite some flexibility, the Q176 conformation is similar to the one in CA WT. This observation, together with subtle changes in the CypA-binding loop suggest that mutations may alter interactions with host cell factors.

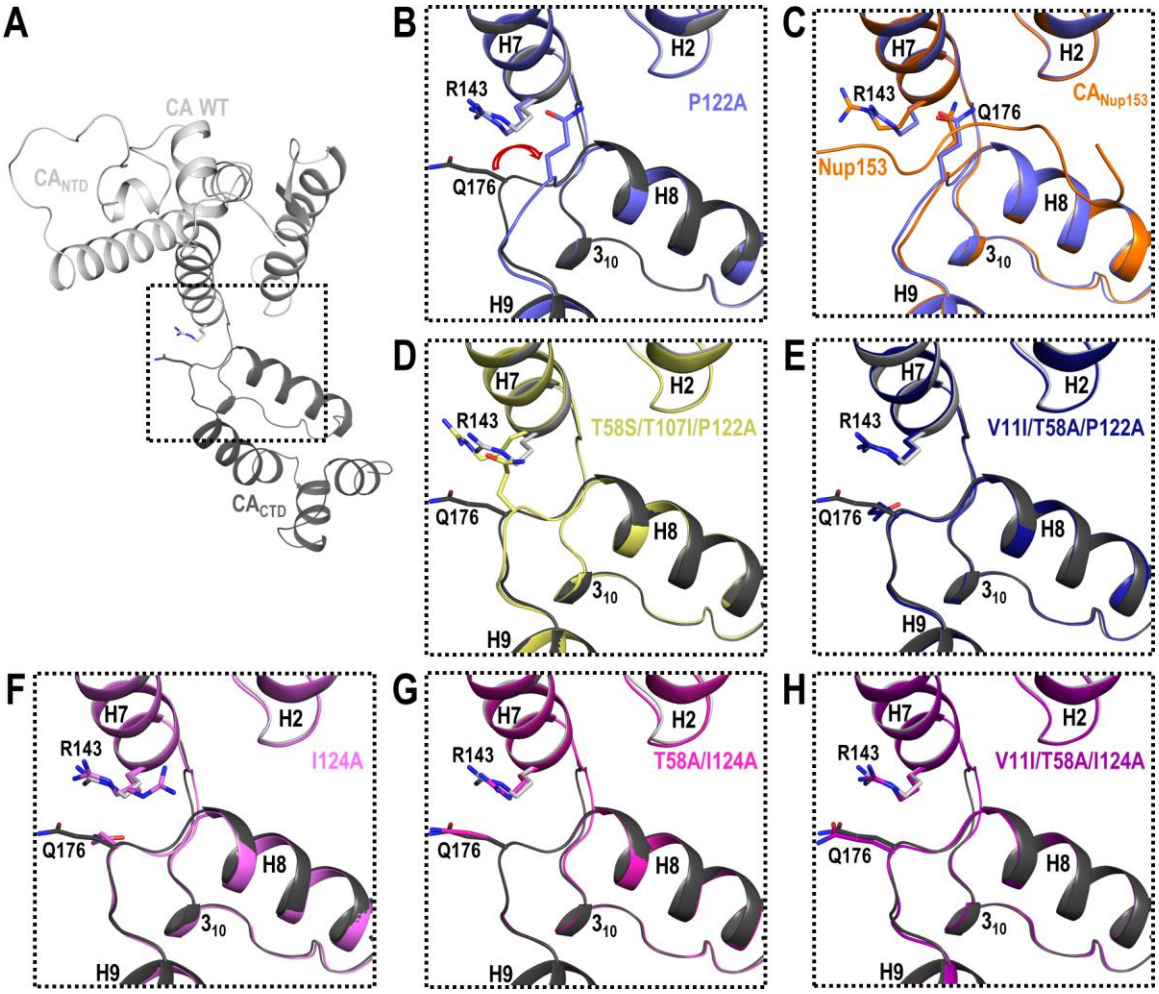


Figure 5-6. Conformational changes of R143 and Q176 at the CA_{NTD}-CA_{CTD} intra-protomer interface.

(**A**) CA WT (CA_{NTD} in white, CA_{CTD} in gray) is shown in ribbon. Side chains of R143 and Q176 are shown as sticks. (**B-D**) Enlarged views of the boxed region in (**A**); P122A (in light blue) superposed with CA WT or CA WT in complex with Nup153 (in orange). Red arrow highlights repositioning of Q176. (**D-H**) Enlarged views of the indicated CA mutants superposed with CA WT (dashed box in (**A**)).

In addition to the localized changes, mutations induce subtle remote variations at the CA-CA interfaces. There is up to ~10 % decrease in the buried surface area at the intra-hexamer interfaces (CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD}). Moreover, mutations also result in variations of up to ~10 % at the 2-fold and ~40 % at the 3-fold inter-hexamer interfaces (Table 5-3). The observed differences imply that mutations may alter the stability of the mature capsid core.

Table 5-3. Buried surface area calculated for interfaces in CA mutant structures.

Structure	Buried surface area (BSA), Å^2		
	Intra-hexamer interface	Inter-hexamer interface	
	CA_{NTD}-CA_{NTD} CA_{NTD}-CA_{CTD}	CA_{CTD}-CA_{CTD} 2-fold	CA_{CTD}-CA_{CTD} 3-fold
P122A	2182.3 (8.6 %)	864.3 (3.4 %)	200.7 (0.5 %)
I124A	2054.5 (8.0 %)	844.0 (3.3 %)	171.6 (0.4 %)
T58A/I124A	2040.5 (8.2 %)	924.3 (3.7 %)	193.2 (0.5 %)
V11I/T58A/P122A	2180.7 (8.6 %)	864.4 (3.4 %)	133.2 (0.3 %)
V11I/T58A/I124A	2235.1 (8.8 %)	974.2 (3.8 %)	276.3 (0.7 %)
T58S/T107I/P122A	2198.0 (8.6 %)	822.1 (3.2 %)	265.5 (0.7 %)
T58S/T107I/P122A$_{PF74}$	2210.2 (8.7 %)	806.4 (3.2 %)	132.3 (0.3 %)
CA WT	2237.7 (8.8 %)	906.4 (3.6 %)	277.5 (0.7 %)

In summary, all the changes imposed by mutations of interest on the mature HIV-1 CA are very subtle. This implies that PPIP mutations primarily affect immature Gag, rather than mature CA lattice. Hence, the H6-H7 loop of the HIV-1 CA is a novel structural element critical for assembly and maturation.

Chapter 6. Novel CA_{NTD}-CA_{NTD} interaction essential for HIV-1 mature capsid stability: tracking the Achilles heel of the capsid core in the pentamer interfaces

Note: This chapter is based on a research paper currently in preparation that will be submitted in *Nature Communications*. The contributions of co-authors are described in the "Footnotes."

A. Introduction

HIV-1 life cycle

The HIV-1 capsid protein (CA) plays crucial roles during early, and late stages of the viral life cycle (*201*). Following viral and host cell membrane fusion, the capsid core is released into the cytoplasm. The core contains and protects the viral proteins and RNA genome from degradation. Reverse transcription occurs in the core and is tightly linked to the capsid disassembly, or uncoating. This leads to the import of the double-stranded (ds) viral DNA into the host cell nucleus, where it is integrated into the host DNA. Upon release of new virions from the host cells, the CA, being a central domain of the Gag polyprotein, promotes its assembly into a spherical structure that underlies the viral membrane (*290*). During maturation, HIV protease cleaves Gag, resulting in the release of CA and several other proteins and peptides. Multiple copies of CA condense around the viral genome forming the mature capsid core (*23*).

Capsid structure

CA consist of a sequence of 231 amino acids, which fold into two distinct domains: the N-terminal domain (CA_{NTD}) and C-terminal domain (CA_{CTD}), connected by a short linker. During maturation, CA molecules assemble to form the characteristic fullerene-cone shaped structure of the mature HIV-1 capsid core (Figure 6-1), which comprises hexamers and pentamers. Several models of the core are available, all composed of ~200 hexamers, and exactly 12 pentamers, which are required to form the closed shape (*83, 84*).

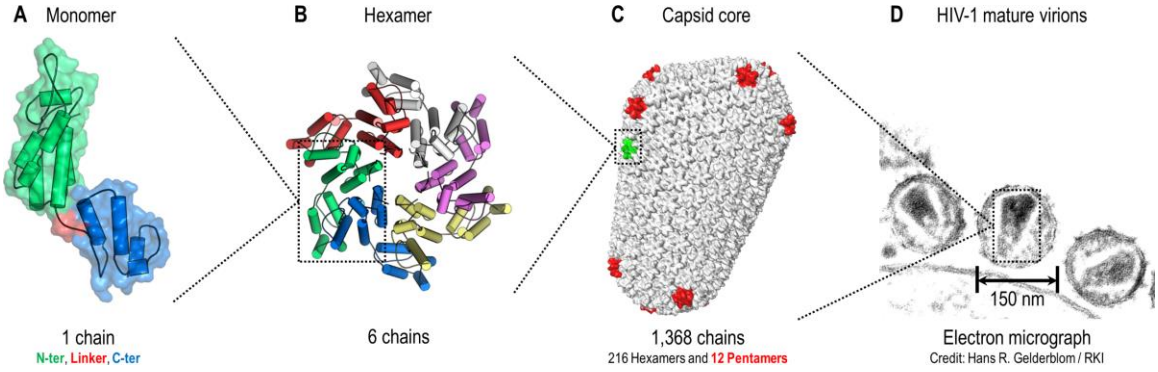


Figure 6-1. Structure and assembly of the HIV-1 mature capsid core.

(**A**) The capsid protein (CA) is a monomer composed of 2 domains connected by a flexible linker, (**B**) which assembles with other copies to form either hexamers or pentamers. (**C**) Hundreds of hexamers and exactly 12 pentamers interact to form the mature capsid core. (**D**) In the mature virions, this compartment encloses viral RNA and proteins, and houses reverse transcription.

Capsid uncoating – state of the art

Upon entry of the HIV-1 into the host cell, the capsid core is released into the cytoplasm where it dissociates in connection with reverse transcription in a poorly understood manner (*105, 236, 292*). Several models of uncoating have been proposed (*201*), each supported by different experimental evidence. These models mainly differ in the timing of uncoating, which occurs either immediately after entry (*293, 294*), or during the nuclear trafficking (*220, 222, 238, 239*), or when the capsid reaches the nuclear

envelope (*240*). It is challenging to describe a comprehensive picture of the uncoating as it takes place in conjunction with other processes (such as reverse transcription (*238, 240*)), and involves interactions with viral partners (matrix, reverse transcriptase, integrase, vpr (*293, 294*)), host partners (Cyclophilin A (CypA) (*220*), microtubules (*295*), dynein (*296*), kinesin (*297*), cleavage and polyadenylation specificity factor 6 (CPSF6) (*241*), transportin 3 (TNPO3) (*259*), nuclear pore proteins Nup358, and Nup153 (*222*)), regulated in a spatiotemporal fashion.

Capsid core stability

Moreover, uncoating also depends on capsid structure, its flexibility, and stability (*97, 298*). Intermolecular CA_{NTD}-CA_{NTD} interactions are essential for the formation and stabilization of the CA_{NTD} rings observed in hexamer (*64*) and pentamer (*95*) assemblies. Additionally, CA_{NTD}-CA_{CTD} contacts occur between two juxtaposed monomers within a hexamer or pentamer (*64, 95*). In contrast, the CA_{CTD}-CA_{CTD} interactions are observed between monomers of neighboring hexamers (or pentamers). The malleability of these CA_{CTD}-CA_{CTD} interfaces, bringing together hundreds of hexamers and 12 pentamers, enables the formation of curved mature cores (*64, 90*).

Recent NMR studies (*299*) have shown that major variations in different *in vitro* CA assembly morphologies (tubes, sheets, and spheres) involve minor variations in the molecular structures of ordered segments. The study suggests that different morphologies are attributable to changes in conformational distributions within disordered segments, inter-molecular CA_{CTD}-CA_{CTD} 2-fold interface, as well as intra-molecular helix-helix packing.

Additionally, it has been proposed that the narrow end of the capsid cone is a weak point involved in the process of disassembly. Indeed, this area has a higher concentration of pentamers, which are less stable than hexamers due to the greater electrostatic repulsion of six R18 residues that form the middle of the CA_{NTD} ring (*95*).

Capsid as a drug target

Mutagenesis studies highlighted the extreme genetic fragility of CA, especially of the CA_{NTD} helices that appear to be more sensitive than CA_{CTD} helices towards mutations that might perturb the structure or stability of the CA_{NTD} itself or the CA hexamer (*96*).

The multifunctional nature of CA in the virus replication cycle has led to increasing drug discovery efforts using CA as a target. As a result, a number of binders have been identified (*267*) that either interfere with viral maturation by disrupting core assembly (CAP, BD and BM compounds (*110, 116*)), or impact the uncoating process (BI compounds (*122*)), while others affect both steps (PF74(*121*)), in addition to inhibiting reverse transcription. The effect of ligand binding on viral uncoating has been extensively studied, however, it is not clear whether the antiviral effect is the result of the increase or decrease in stability of CA assemblies. Moreover, the binding of compounds to the capsid core may additionally prevent interactions with host protein partners required for proper uncoating, as observed for BI-2 and PF74, whose shared binding sites overlap with those for CPSF6 and Nup153 (*119, 124*).

Identification of interactions involved at CA interfaces is a key step in the development of new antivirals targeting stability of CA assemblies. Here we present evidence pointing towards the importance of the novel CA_{NTD}-CA_{NTD} interaction at the

pentamer interface. This information could be essential for the future development of new drugs targeting mature CA assemblies.

B. Materials and methods

Design, expression, and purification of CA mutants

CA R18A, E28A, and R18A/E28A mutants were based on a pET11a construct (*268*). Mutations were introduced using overlap extension PCR cloning and verified by DNA sequencing. Mutant CA proteins were expressed and purified as previously described (*245, 268*).

TEM assembly of CA mutants

Native CA and mutants (R18A, E28A, and R18A/E28A) were assembled at 150 μM in buffer containing 1 M NaCl and 50 mM Tris-HCl pH 8.1 at 37 ^{o}C for 1 hr. Five microliters of each sample were adsorbed during 5 min on grids coated with colloidal carbon and made hydrophilic by glow-discharge for 45 sec. After this, excess fluid was removed, and grids were washed with water, fixed in a drop of 2 % uranyl acetate and dried before visualization using a JEOL JEM 1400 transmission electron microscope with a magnification of 2,500×, 20,000× and 40,000×. Each reaction was repeated at least three times.

Crystallization of CA mutants

Crystals of the CA mutant proteins grew at 18 $^{\circ}$C in hanging drops, containing 2-5 mg/ml of protein, 6-9 % PEG 3350, 2-6 % glycerol, sodium iodide, and sodium cacodylate. Hexagonal plate-like crystals appeared after five days, and crystal growth was

completed in over two weeks. Crystals were briefly dipped in paraffin oil before cryo-cooling in liquid nitrogen.

Data collection and structure determination

Data were collected on a Dectris Eiger-16m detector at the Advanced Photon Source (APS) beamline 23 ID-B, Argonne National Laboratory. Datasets were collected and processed using XDS (*247*). The data were examined for the presence of systematic absences; however, no characteristic patterns were observed. Thus, the crystals were indexed in hexagonal space group P6 with one CA molecule in the asymmetric unit. No twinning was present, as determined by either POINTLESS (*249*) or XTRIAGE (*250*). Space group and twinning were also verified in ZANUDA (*251*). The phase problem of CA was solved using molecular replacement, with the native CA (PDB ID: 4XFX) as a starting model. Initial phases were solved *via* PHASER (*251*). Several rounds of iterative model building and refinement were carried out using Coot (*253*) and PHENIX (*250*), REFMAC (*251, 252*), or PDBREDO (http://www.cmbi.ru.nl/pdb_redo/), respectively. Structure validation of final models was performed with MOLPROBITY (http://molprobity.biochem.duke.edu/). Accessible and buried surface areas were calculated using PISA (*251*). The figures showing structural information were generated in PyMOL (http://www.pymol.org/).

C. Results and discussion

Differences in the dimer interactions in the hexamer and pentamer assemblies

Currently, two models of the entire capsid core are available in the Protein Data Bank (PDB); PDB ID: 3J3Q and 3J3Y (*65*). They were built using a combination of

computational techniques (molecular dynamics flexible fitting, molecular modeling, and all-atom molecular dynamics simulations), and based on several data types (cryo-electron microscopy (cryo-EM) maps, cryo-electron tomography (cryo-ET) images, and X-ray structures). Using these models, we statistically analyzed the inter-molecular interactions between two monomers either within hexamer or pentamer interfaces (Figure 6-2). One interaction emerged to be significantly more frequent at the pentamer rather than hexamer interfaces. The hydrogen bond (H-bond) between the glutamic acid in position 28 and lysine in position 30 (E28~K30) was found in 86.7 % of dimers comprising pentamer interfaces, in comparison to 24.2 % of those at the hexamer interfaces (Table 6-1, A). This H-bond is positioned at the bottom of a pocket between two N-terminal domains of the two consecutive CA monomers (Figure 6-3); more specifically, it involves the end of the two helices 1 (H1 and H1'). Here we use the term 'NDI-pocket' to refer to this N-terminal domains interface pocket.

Table 6-1. Cα-Cα distances between E28 and K30 in the models and cryo-ET reconstructions.

H-bond interactions were identified in the model structures using a distance threshold of 3.35Å between the oxygen atoms of the E28 side-chain and the nitrogen atom of the K30 side chain.

A)		**Hexamer**	**Pentamer**
Model, PDB IDs: 3J3Q and 3J3Y	Number of dimers	2412	120
	% dimers with E28~K30 H-bond	24.2 %	86.7 %
	Cα-Cα distance (Å)	14.45 ($^{+}$/. 2.02)	11.91 ($^{+}$/. 0.98)
	Δ (Hexamer - Pentamer) (Å)	2.54	
B)		**Hexamer**	**Pentamer**
Cryo-ET, PDB IDs: 5MCX → 5MCZ, 5MD0 → 5MD9, 5MDA → 5MDG	Number of dimers	100	5
	Cα-Cα distance (Å)	13.64 ($^{+}$/. 0.34)	10.05 ($^{+}$/. 0.04)
	Δ (Hexamer - Pentamer) (Å)	3.59	

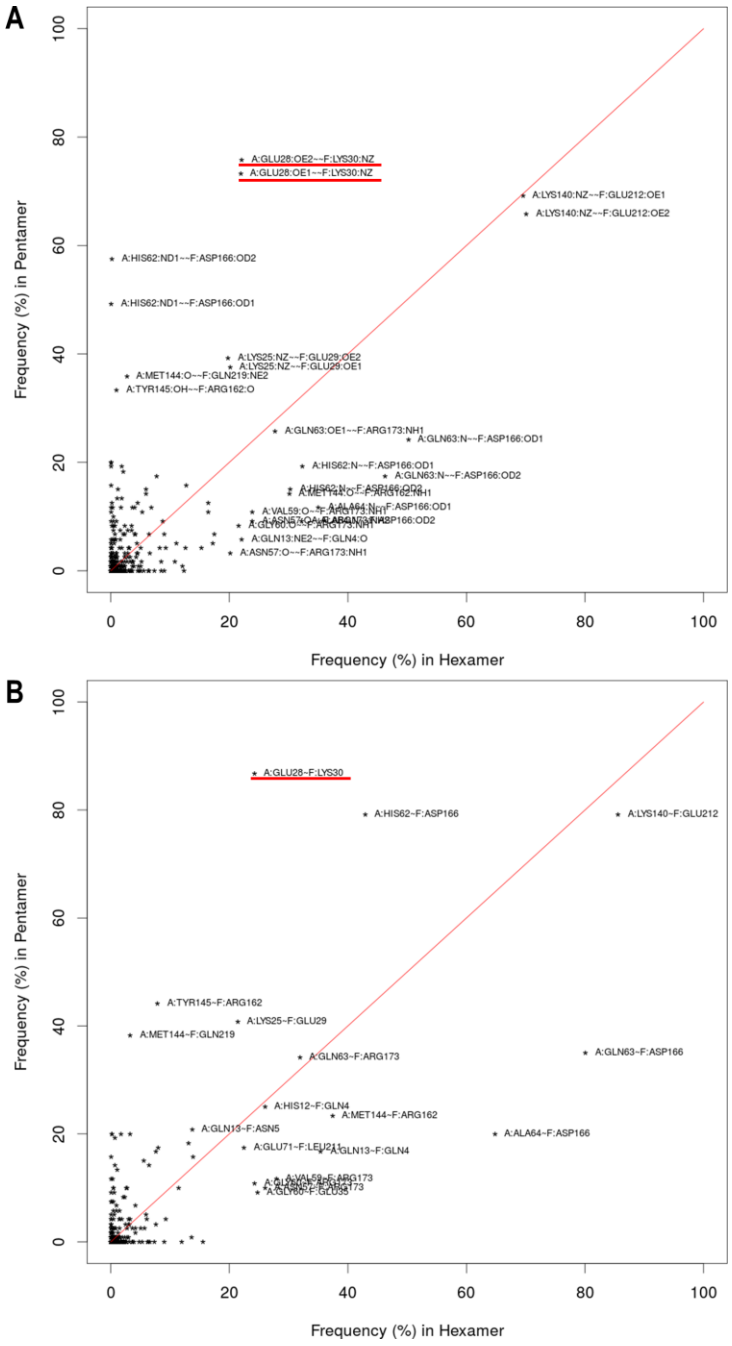


Figure 6-2. Comparison of H-bond interaction frequencies at the CA monomer-monomer interfaces in the context of a hexamer or pentamer (structures from PDB IDs: 3J3Q & 3J3Y).

For both plots, the *x* axis refers to the H-bond frequencies observed in the context of a hexamer (out of 2412 monomer-monomer interfaces), and the *y* axis – in the context of a pentamer (out of 112 interfaces). Only interactions observed in at least 20 % dimers (either in hexamers or pentamers) are labeled. Labels underlined in red correspond to the E28~K30 H-bond. (**A**) Each point refers to H-bond between one atom from monomer A and one atom from monomer F. (**B**) Each point refers to interactions between one residue from monomer A and the other one from monomer F, regardless if it involves one or more H-bond interactions.

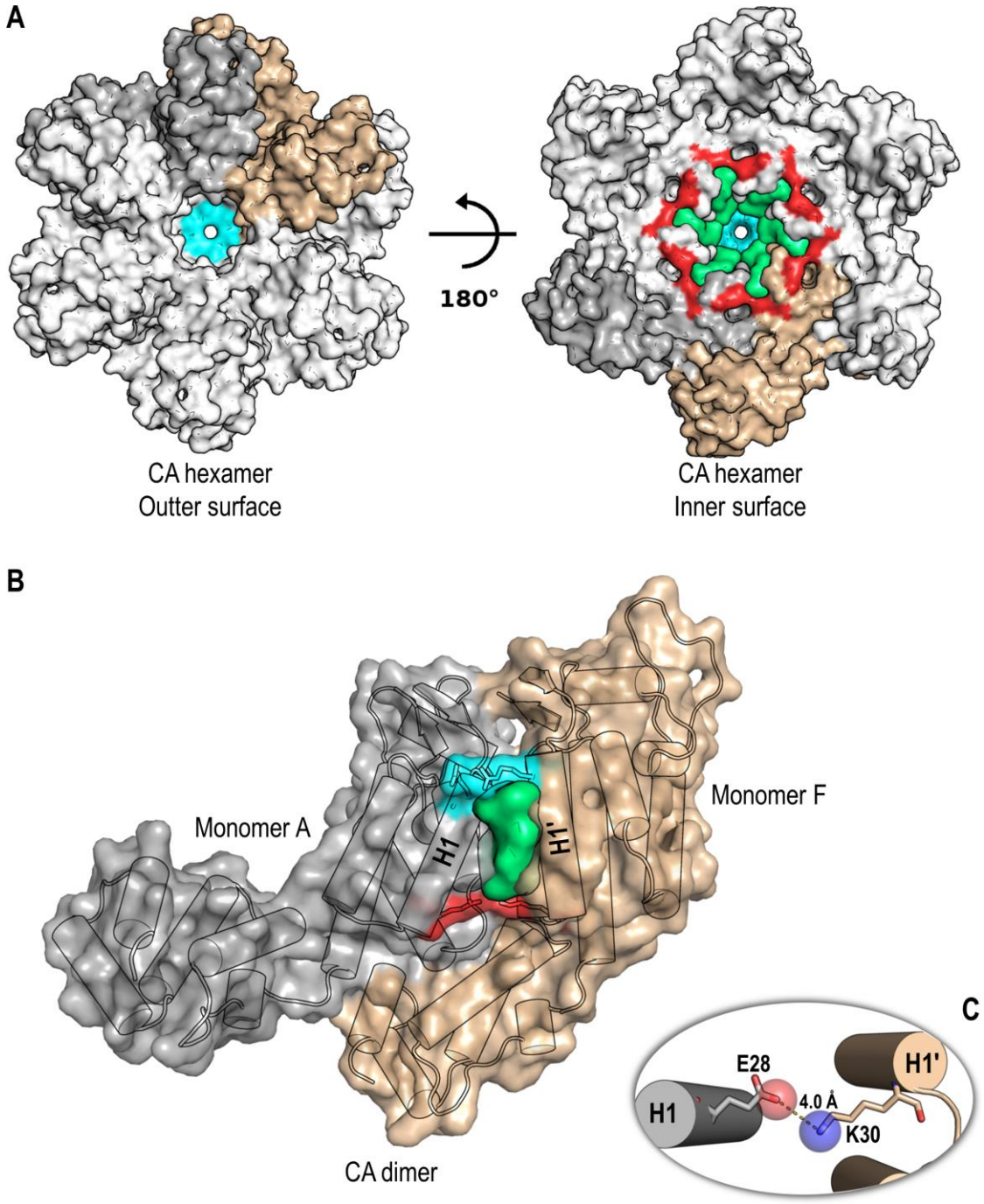


Figure 6-3. The N-terminal domains interface pocket (NDI-pocket).

(**A**) The NDI-pocket (in green) occupies the inner surface of the capsid core, and symmetrically encircles the 6-fold axis of the hexamer (5-fold in the context of a pentamer). (**B**) The pocket is defined by the CA_{NTD}-CA_{NTD} interface of the two neighboring monomers (monomer A and F), trapped between two helices H1 and H1'. It is capped by the two R18 side chains (in cyan) at the top, and E28 from one monomer and K30 from the other (in red) at the bottom. (**C**) In the X-ray structure of the native capsid (PDB ID: 4XFX), the E28 and K30 are too far to form a direct H-bond interaction.

More recently, Mattei *et al.* solved 20 sub-nanometer cryo-electron tomography structures from intact virions (*90*), giving access to the backbone coordinates from which the impact of the core curvature could be observed. The Cα distances between E28 and K30 show stronger differences between pentamer and hexamer interfaces than the ones measured in the two models (Table 6-1, B).

Interestingly, in the available X-ray structure of the mature wild-type capsid (CA WT) protein (PDB ID: 4XFX) (*268*), this interaction is not observed. In fact, it is not observed in any of the CA crystal structures available in the PDB, *i. e.* dehydrated, mutated, or bound to a ligand or peptide. The bond between E28 and K30 is only described as an indirect interaction mediated by water molecules in the X-ray structure of the cross-linked hexamer (PDB ID: 3H47) (*64*) (Supp. Data, Table 5 from ref (*268*)). Two reasons could explain the absence of this interaction in the crystal structures: first, most of them (except one) correspond to the CA hexamer, not pentamer, assemblies, and, second, the crystal lattice is flat, lacking the characteristic curvature of the core itself. This curvature is expected to impact the orientation of CA_{NTD}s (*90*), and interactions within the CA_{NTD} helix bundle (*299*). Thus, we can expect that it may affect the CA_{NTD}-CA_{NTD} interface between neighboring monomers. Moreover, the curvature is supposed to be more pronounced in the presence of pentamers (*65, 95*).

The only available X-ray structure of a pentamer is a crosslinked one (PDB ID: 3P05) (*95*). The engineered disulfide bond takes place at the top of the NDI-pocket, between the position 21 of the H1 of one monomer, and the position 22 of the H1' in the neighboring monomer. This constraint brings E28 and K30 closer at the end of the helices, leading to the reorientation of their side-chains to avoid steric clashes (Figure 6-4).

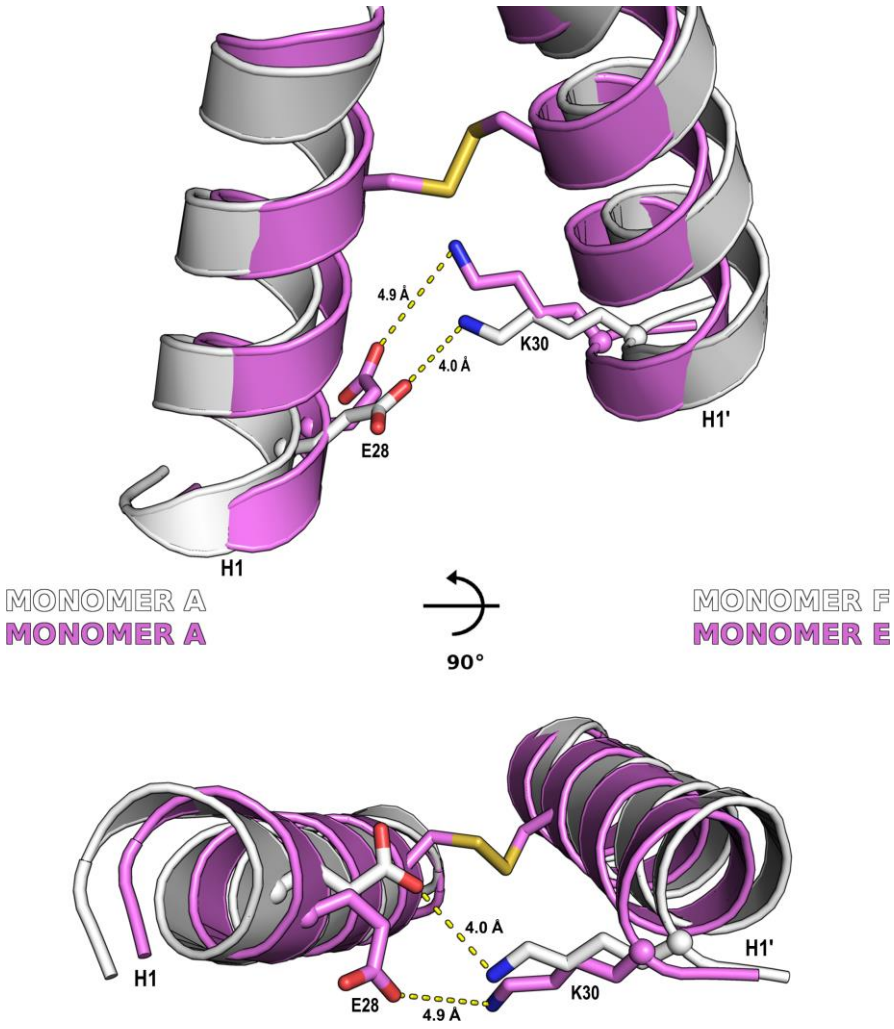


Figure 6-4. E28 and K30 side chain orientations in the cross-linked pentamer compared to the CA WT hexamer.

No interaction is observed between E28 and K30 in the crystal structure of the CA hexamer (PDB ID: 4XFX; in white). Both side chains are too far to form a proper H-bond. In the only available crystal structure of the CA pentamer (PDB ID: 3P05; in pink), the neighboring CA_{NTD}s are cross-linked with the engineered disulfide bond (in yellow) between position 21 in the H1, and position 22 in the H1'. This constraint brings closer the two residues E28 and K30 in comparison to the hexamer, however, the E28~K30 H-bond is not observed. The E28 and K30 side chains rearrange to avoid steric clashes. Notably, other reorientations of E28 and K30 side chains could be observed in the five CA_{NTD} interfaces in the pentamer crystal structure.

Based on this structural observation, we hypothesized that the E28~K30 H-bond stabilizes the CA_{NTD}-CA_{NTD} interface in the context of pentamer assemblies. This could be a specific interaction, which counterbalances the greater electrostatic repulsion observed between R18 at the 5-fold in the pentamer compared to the 6-fold in the hexamer.

E28, K30, and R18 are highly conserved

E28, K30, and R18 are highly conserved positions, corresponding to 99.4 %, 98.8 % and 99.4 % sequence identity across the subtype reference sequences available in the Los Alamos National Laboratory, HIV mutation browser: http://hivmut.org (*272*). There are only three subtypes, 03GH173_06 (GenBank: AB286852.1), nx2 (GenBank: HM067748.1), and A32989 (GenBank: AF408630.1), where those positions are substituted with K30R, E28D/K30E, and R18K respectively (Figure 6-5).

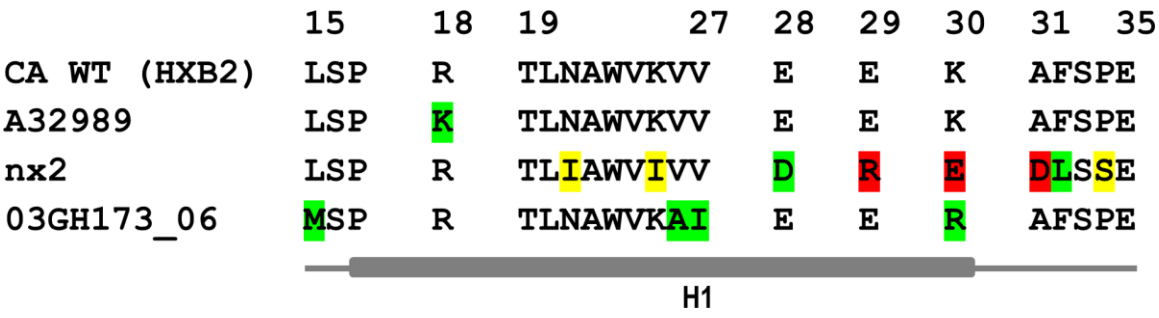


Figure 6-5. Sequence alignment of A32989, nx2 and 03GH173_06 subtypes (positions 18 through 38) to CA WT (HXB2).

Substitutions to similar amino acids are highlighted in green. Substitutions accompanied with the charge change is shown in red. Substitutions to non-related amino acids are in yellow.

Mutagenesis and infectivity studies

Previous mutagenesis studies and *in vivo* infectivity assays have shown the importance of these positions for the virus replication. Single point substitution of K30 with shorter, oppositely charged or not charged amino-acids, K30E and K30N, resulted in the production of non-infectious particles (*96*). Interestingly, this study provides cryo-EM images of purified virions of the K30N CA mutant. The observed morphologies of mutant virions were described as "*indistinguishable from the WT virus*" (*96*). However, this interpretation is ambiguous when taking into account additional images kindly provided

by the authors of the original study (Figure 6-6) since the core is hard to discern in the micrographs for both WT and K30N viruses.

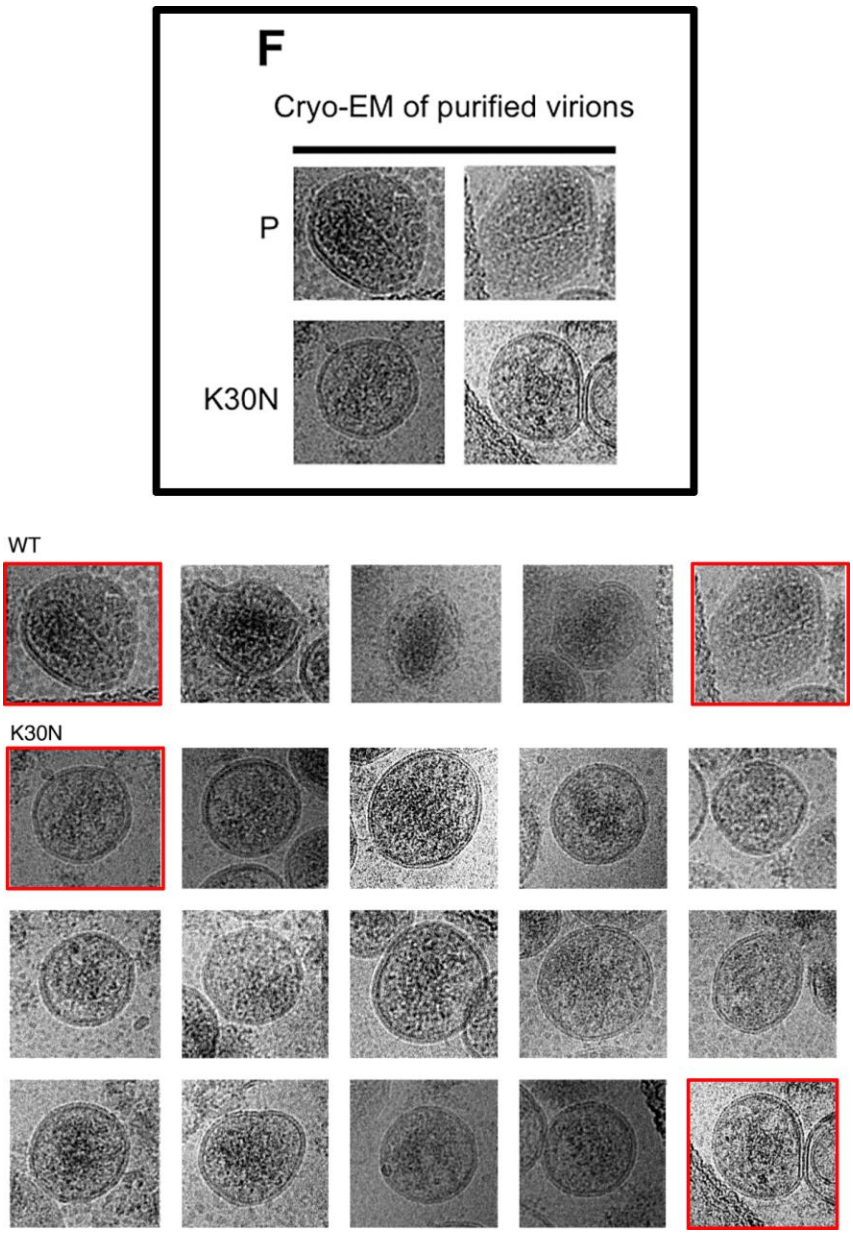


Figure 6-6. Cryo-EM images of the purified virions of the CA WT and K30N mutant.

Cryo-EM images presented in the **Figure 7 panel F from *Rihn, S. J. et al. PLoS Pathog 2013*** are framed in black at the top of the page (*96*). The micrographs below correspond to the additional data kindly shared by the authors. The images used in the **Figure 7 panel F** are highlighted in red. According to the study, K30N mutants exhibited "*efficient particle formation*" and "*morphologies that were indistinguishable from WT virus*." However, in our opinion, it is difficult to dissect the effect of the K30N mutation on the core assembly in the virions, as in both WT and mutant sets of images the core is hard to discern.

The E28A/E29A and R18A/N21A double mutants similarly resulted in the production of non-infectious virus. Studies of the E28A/E29A viral particle morphology using transmission electron microscopy (TEM) showed normal levels of particle release. However, the core formation was eliminated (*92, 93*), highlighting that the reduction or loss of infectivity is directly linked to the lack of the mature conical capsid. The primary defect in R18A/N21A assembly appeared to be a production of abnormally shaped and multiple capsids, but this mutation also reduced the overall efficiency of capsid assembly, suggesting an additional slight Gag assembly defect.

Assembly and TEM analyses of recombinant CA proteins

The CA WT is known to assemble *in vitro* into long hollow tubular structures, homogeneous in width with an external diameter of ~55 nm, but highly heterogeneous in length. The tubes are composed of CA hexamers and resemble authentic viral capsids (*80, 81, 84*).

Previous study has shown that E28A/E29A reduced, but did not eliminate, CA assembly *in vitro* (*93*), producing cylinders similar to the CA WT at higher protein concentrations (~15 mg/ml), with severe attenuation of their production at the lower protein concentration (~5 mg/ml).

Furthermore, R18A and R18A/N21A have been previously shown to impact the morphology of the *in vitro* CA assemblies by shifting the observed phenotypes to spheres, cones, spirals, and short capped cylinders (*93*). This phenotype has been explained by an increased frequency of pentamer incorporation into the assembling hexamer lattice in the presence of R18A mutation. The R18 brings together charged side chains, which creates a

positively charged channel in the middle of the CA_{NTD} rings (*280*). In the pentamer assembly, the charges are brought closer than in the hexamer, resulting in stronger electrostatic repulsion, which disfavors pentamers relative to hexamers (*95*).

Further studies revealed that elimination of the charge by replacement of R18 with a large hydrophobic residue (*i. e.*, V, I, L, F) favored pentamer incorporation and induced assembly of spheres presumably due to the stabilizing effect of hydrophobic contacts (*85, 93, 95*). With variation of the assembly conditions, those mutants also yielded cylinders, cones, and very large spheres up to ~2.5 mm in diameter, which collapsed and flattened (*85*) upon application to the EM-grid.

To address our hypothesis that E28~K30 H-bond may contribute to counterbalancing the electrostatic destabilization of the pentamers, we tested purified mutant CA proteins, harboring R18A, E28A and R18A/E28A mutations, for the cylinder formation *in vitro*.

CA WT (used as a control) and CA mutant proteins were assembled *in vitro* and analyzed by negative-stain TEM. Consistent with previous reports, under experimental conditions CA WT formed hollow cylinders (or tubes) homogeneous in diameter ~45–60 nm, but extremely heterogeneous in length. As expected, R18A mutant assembled efficiently, forming sheets and highly variable in diameter spheres (ranging from 30 nm to 220 nm, occasionally achieving ~400 nm), *i. e.,* phenotypes, which reveal the presence of pentamer assembly *in vitro* (Figure 6-7). Notably, shortening the R18 side chain (R18G and R18A) should neither affect the overall structure of the CA protein nor impact its ability to form hexamer assemblies (*280*), since they are present in the observed morphologies.

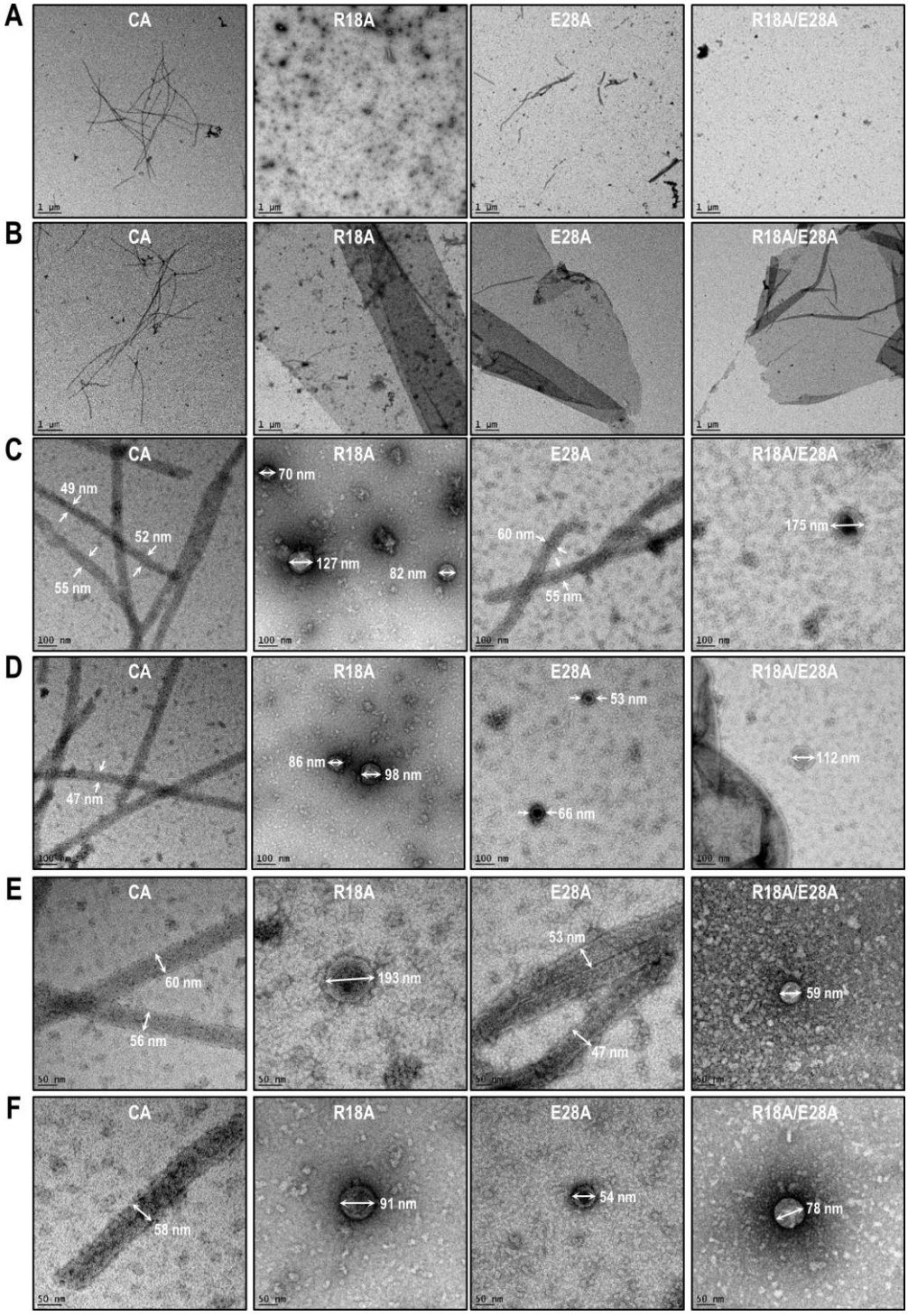


Figure 6-7. Effects of R18A, E28A, and R18A/E28A mutations on CA assembly.

(**A-F**) TEM analysis of CA mutant assemblies. Projection images were recorded at 2,500× (**A-B**), 20,000× (**C-D**), and 40,000× (**E-F**) magnifications from the corresponding samples as indicated. Scale bars, 1 µm in (**A-B**), 100 nm in (**C-D**), and 50 nm in (**E-F**), respectively.

Hollow tubes were observed during the E28A assembly, although, their formation was attenuated in comparison to the CA WT (Figure 6-7). Additionally, this mutant formed sheets and spheres (diameters ~50–130 Å). These results suggest that, even though E28A has detrimental effects on CA assembly *in vitro*, the E28~K30 interaction is not required for the hexamer assembly.

To probe if pentamer interactions could be observed *in vitro* without the putative E28~K30 H-bond interaction, we included R18A/E28A double mutant in our analysis. The R18A/E28A assembly efficiency was reduced, but not eliminated, resulting in altered morphologies: sheets and small number spheres (diameters ranging between 30 and 210 nm) (Figure 6-7). While R18A mutation was primarily responsible for the phenotype, E28A significantly decreased the assembly efficiency of the R18A/E28A CA protein. Residues R18 and E28 outline the NDI-pocket (Figure 6-3, B) without forming a direct interaction (the distance between them ~14Å). Thus, the observed differences in assembly could not be related to the removal of a direct contact between R18 and E28. Taken together, these data support our hypothesis that E28~K30 H-bond can contribute to the stabilization of pentamer assembly.

Crystallographic analysis of CA mutant proteins

To address if the above-mentioned mutations alter the structure of the CA protein, we have crystallized CA proteins harboring R18A, E28A, and R18A/E28A mutations. The structures have been solved in the P6 space group with one molecule/asymmetric unit (Table 6-2).

Table 6-2. Summary of X-ray data collection and refinement statistics.

	R18A	E28A	R18A/E28A
Data collection			
X-ray source	APS 23 ID-B	APS 23 ID-B	APS 23 ID-D
Software	XDS	XDS	XDS
Space group	P6	P6	P6
Unit cell dimensions			
a, b, c (Å)	92.9 92.9 58.1	92.7 92.7 58.0	92.1 92.1 58.3
α, β, γ (°)	90.0 90.0 120.0	90.0 90.0 120.0	90.0 90.0 120.0
ASU content	1	1	1
Wavelength (Å)	1.033203	1.033203	1.033203
Resolution range (Å)[a]	47.1–2.2 (2.2–2.1)	47.0–2.2 (2.3–2.2)	47.0–2.3 (2.4–2.3)
R_{merge}	0.079 (0.842)	0.056 (0.639)	0.075 (0.930)
R_{meas}	0.087 (0.930)	0.062 (0.711)	0.083 (>1)
R_{pim}	0.037 (0.391)	0.026 (0.308)	0.035 (0.428)
$\langle I/\sigma I \rangle$	12.5 (1.8)	16.2 (2.3)	13.6 (1.8)
$CC_{1/2}$ (%)	99.7 (63.3)	99.7 (77.9)	99.6 (57.9)
Completeness (%)	99.8 (97.8)	99.8 (98.6)	98.6 (98.1)
Redundancy	5.6 (5.4)	5.6 (5.1)	5.6 (5.6)
Mosaicity	0.14	0.18	0.12
Refinement			
Resolution (Å)	47.1–2.1	47.0–2.2	47.0–2.3
No. total reflections	95453	82207	71522
No. unique reflections	16949	14720	12726
No. test reflections[b]	936	854	713
R_{work} / R_{free}	21.7 / 24.6	21.3 / 24.6	20.6 / 24.5
No. atoms	1815	1834	1793
Protein	1687	1701	1686
Ligand/Ion	16	18	14
Water	112	115	93
Wilson B-factor (Å^2)	35.6	45.6	39.0
Average B-factors (Å^2)	57.3	66.0	61.3
Protein	57.4	66.1	61.3
Ligand/Ion	68.2	81.8	71.3
Water	54.6	61.7	59.0
RMS deviations			
Bond lengths (Å)	0.002	0.002	0.003
Bond angles (°)	0.514	0.481	0.564
MolProbity Statistics[c]			
All atom clash score	2.66	2.63	2.36
Rotamer outliers (%)	0	0	0
Cβ deviations >0.25 Å	0	0	0
Ramachandran [c]			
Favored region (%)	99	99	99
Outliers (%)	0	0	0
PDB accession code	5W4O	5W4P	5W4Q

[a] Values in parentheses are for the highest-resolution shell; [b] random selection; [c] values obtained from MOLPROBITY.

Overall, the mutant structures are very similar to the CA WT (rmsd: 0.3495 Å for R18A; 0.4054 Å for E28A; 0.4632 Å for R18A/E28A) (Figure 6-8).

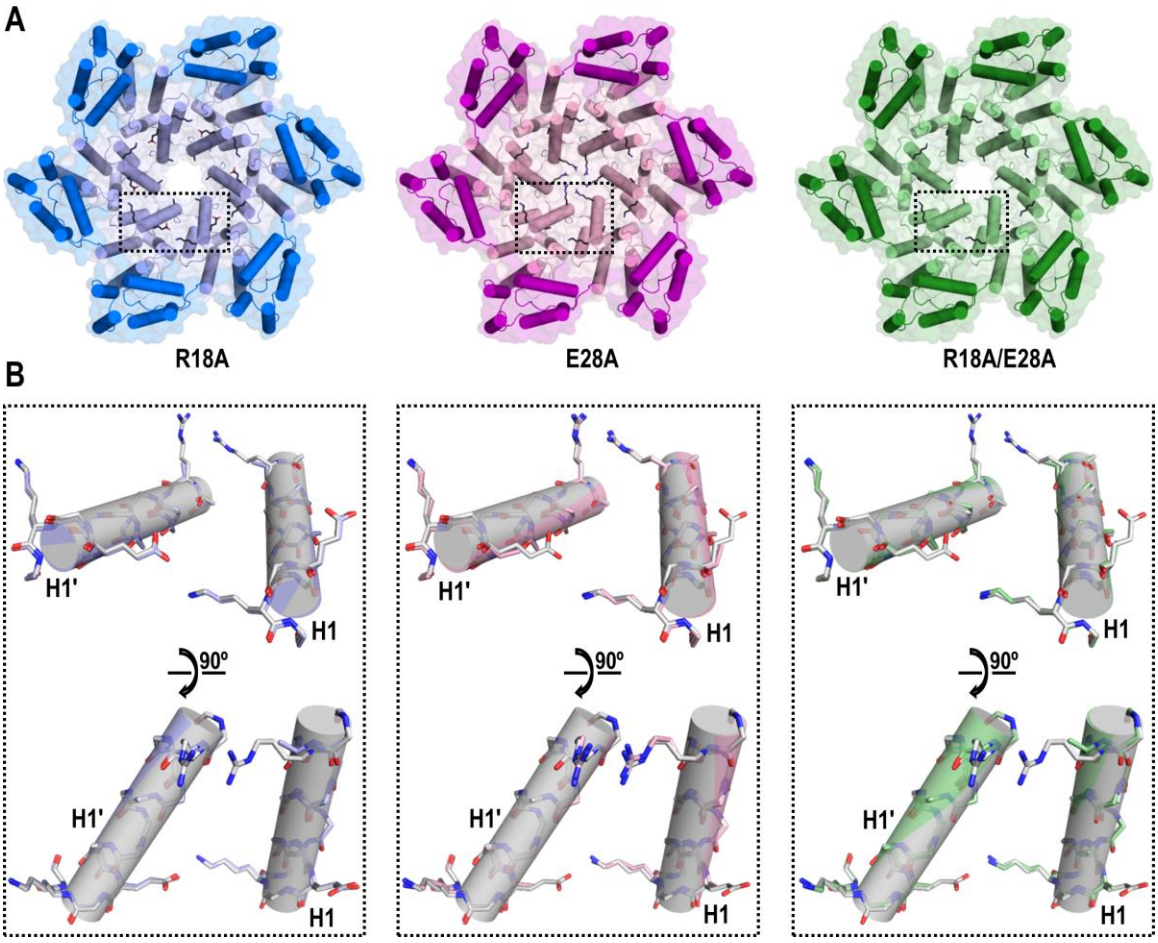


Figure 6-8. X-Ray structures of R18A, E28A, and R18A/E28A CA mutants compared to CA WT (PDB ID: 4XFX).

(**A**) HIV-1 CA mutants R18A (CA_{NTD}s in light blue, CA_{CTD}s in blue), E28A (CA_{NTD}s in light pink, CA_{CTD}s in purple), and R18A/E28A (CA_{NTD}s in light green, CA_{CTD}s in green) form a hexamer that is very similar to CA WT. The side-chains for the positions of interest R18 or A18, E28 or A28, and K30 are shown as black sticks. (**B**) Two helices 1 (H1 and H1'), encompassing the NDI-pocket, are very similar to the CA WT (shown in white). The side chains are shown for positions 18, 28, and 30 only.

The side chains of all mutated residues were solvent exposed and were not involved in significant intra- or intermolecular interactions in the CA crystal structures, thereby minimizing the chance of significant structural perturbations beyond the mutation site. This observation is further supported by the buried surface area calculations (Table 6-3). R18A

mutation in both single and double mutant results in ~10 % decrease in the buried surface area at the intra-hexamer interface (CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD}), while E28A has no significant effect. Interestingly, in addition to the localized changes and alteration of the intra-hexamer interface, those mutations cause remote subtle side chain rearrangements that result in 5-10 % decrease of the buried surface area at the 2-fold, and 30-60 % decrease at the 3-fold inter-hexamer interfaces.

Table 6-3. Buried surface area calculated for interfaces in CA mutant structures.

Structure	Buried surface area, Å^2		
	Intra-hexamer interface CA_{NTD}-CA_{NTD} CA_{NTD}-CA_{CTD}	Inter-hexamer interface 2-fold CA_{CTD}-CA_{CTD}	3-fold CA_{CTD}-CA_{CTD}
R18A	2002.0 (7.7 %)	856.6 (3.3 %)	105.9 (0.3 %)
E28A	2272.9 (8.7 %)	836.6 (3.2 %)	117.0 (0.3 %)
R18A/E28A	2041.9 (8 %)	829.3 (3.2 %)	192.9 (0.5 %)
CA WT	2237.7 (8.8 %)	906.4 (3.6 %)	277.5 (0.7 %)

Notably, in the R18A crystal structure, there is an indirect water-mediated H-bond interaction between E28 and K30, which was similarly observed in the cross-linked hexamer (PDB ID: 3H47), while E28A, R18A/E28A, and CA WT structures lack this interaction. Summarizing, the R18A, E28A, and R18A/E28A X-ray structures demonstrate that mutations neither affect the folding of the CA protein nor dramatically alter the CA hexamer assembly. Consequently, the attenuation of *in vitro* assembly for the R18A/E28A double mutant is likely explained by an impact of the E28A mutation on the formation of the pentamer interface.

Implications for structural determinants of HIV-1 capsid stability and assembly

The residues R18, E28, and K30, are highly conserved among numerous known HIV-1 subtypes. Mutagenesis studies have shown their importance for the production of infectious virus and highlighted their impact on the core assembly *in virions*. In the available models of the HIV-1 cores, E28 and K30 preferentially form a H-bond at the CA_{NTD}-CA_{NTD} interface of pentamers. This H-bond is not observed in the available CA X-ray structures. Analysis of R18A, E28A, and R18A/E28A assembly *in vitro*, supported by the corresponding crystal structures, provides indirect evidence that E28A mutation impacts the pentamer, not hexamer, stability and/or assembly *in vitro*.

Table 6-4 summarizes the obtained results and suggested interpretations based on the comparison of the CA proteins assembly *in vitro*.

Table 6-4. Viral phenotypes of HIV-1 CA mutations.

Protein	Location of mutation	*In vitro* assembly	Phenotype	Interpretation
CA WT	-	+++		Tubes are formed of hexamers; pentamer incorporation is a rare event *in vitro*.
R18A	Helix 1	+++	 + sheets	Stabilizes pentamers, as a result, the frequency of pentamer incorporation increases; does not severely affect the ability to form hexamers (supported by X-ray structure).
E28A	Helix 1	++	 +  + sheets	Formation of tubes attenuated; has subtle effect on hexamers (supported by X-ray structure).
R18A/E28A	Helix 1	+	 + sheets	Decreases the assembly of hexamers and pentamers.

As previously indicated, E28A attenuates the formation of tubes but does not abolish the hexamer assembly. R18A increases the ability to incorporate pentamers upon

assembly, without affecting the ability to form hexamers. The significant reduction in assembly efficiency observed for the R18A/E28A could be explained by the impact of the mutations on either the formation of hexamers or pentamers.

Nonetheless, one observation remains difficult to explain: why no tubes are observed for the R18A/E28A mutant if these mutations do not significantly affect the intra-hexamer interface? One explanation could be that the R18A favors the formation of pentamers, which are unstable due to the absence of H-bond between A28 and K30, even though this interaction is not required for the assembly of pentamers. Another explanation is the destabilizing effect of mutations on remote inter-hexamer interfaces, which are essential for the proper CA assembly.

Collectively, the data support our hypothesis that the E28~K30 H-bond exists predominantly in the pentamer assemblies and is substantial for the counterbalancing their intrinsic instability imposed by electrostatic repulsion between R18 at the 5-fold. This interaction may regulate pentamer stability and may be essential for the proper CA assembly, stability, and uncoating of the HIV-1 capsid core.

Chapter 7. Summary

Over the past 25 years, considerable effort has focused on describing the structural details of the retroviral capsid. However, detailed molecular interactions between HIV-1 CA hexamers have been elusive in the context of a native protein. In this study, we solved the structure of the native full-length CA that provides structural details of interactions at the inter- and intra-hexamer interfaces that control stability of the hexagonal lattice. The CA structure also revealed a hydration layer that is adaptable and modulates interactions among CA molecules. We have demonstrated that disruption of this layer by crystal dehydration treatment alters inter-hexamer interfaces and distorts CA packing, highlighting an inherent structural variability of the capsid core.

In all previous structures, interactions at the 3-fold inter-hexamer interface are entirely missing, and those at the 2-fold are strongly disrupted by engineered mutations. Our system has established a novel structural framework for the study of native CA in the presence of all residues that participate in the critical inter-hexamer interactions that control capsid stability and multiple biological events.

CA has an intrinsic plasticity that provides biological and evolutionary benefits by contributing towards successful completion of complex steps in the virus life cycle. Moreover, CA has been reported to interact with multiple host cell factors, including CPSF6 (*118, 198, 264*), TNPO3 (*209, 210, 259*), MxB (*129, 189, 190, 193, 260-263*), Nup153 (*237*), Nup358 (*220*), and CypA (*161, 162, 220, 257, 264*). Due to its involvement in multiple critical processes, CA is an attractive therapeutic target (*98, 265*) with several antivirals known to bind at the CA_{NTD} or CA_{CTD} domains, including NYAD6 (*138*) and PF74 (*123*). Our laboratory recently discovered 18E8, a CA-targeting antiviral that blocks

infectivity without affecting completion of reverse transcription or assembly. Hence, CA structure opens up new opportunities to specifically target diverse steps of the viral life cycle and can offer novel strategies for therapeutic intervention.

Despite the availability of crystal structures of CA_{NTD} (*118, 121*) and cross-linked CA hexamers (*119, 120*) in complex with antiviral PF74 or host factors CPSF6 and Nup153, the mechanism by which they affect uncoating is not well understood. Our study confirmed previous findings that CPSF6, Nup153, and PF74 make contacts across two subunits in a CA hexamer. Additionally, they revealed that effector binding results in subtle changes at the 2-fold and 3-fold interfaces between neighboring hexamers. Hence, the structures provided new insights into the mechanism by which PF74, CPSF6 and Nup153 change intra- and inter-hexamer interactions that affect core stability. Moreover, we evaluated the effects of the pharmacological ligands (PF74, NYAD6, and 18E8) on the CA assembly *in vitro*. CA protein assembles into long tubes. The presence of PF74 leads to the formation of short tubes and cones, NYAD6 greatly diminishes the CA assembly giving short tubes, while 18E8 leads to the production of aberrant morphologies in addition to long tubes. This information will be instrumental for future studies of the mechanisms by which CA-interacting cellular factors and antivirals affect viral uncoating, nuclear import, viral restriction, and assembly.

Furthermore, we have used this system to study the effects of three groups of CA mutations. First, we focused on the panel of mutations (P38A, E45A, P38A/T216I, E45A/E45A/R132T) reported to alter capsid stability and impair infectivity, suggesting that optimal core stability and its proper uncoating is critical for productive infection (*97, 273*). Consistent with previous reports, we observed that the T216I and R132T mutations

could at least partially correct *in vitro* assembly defects imposed by P38A and E45A, respectively, without correcting the altered capsid stability imposed by the original mutations. Based on the crystallographic analysis of these CA mutants, we suggest that the residues E45 and D51 are the key players controlling the conformation of the β-hairpin, the presence of the pore at the 6-fold intra-hexamer interface, as well as the stability of the capsid core and its disassembly. Mutual electrostatic repulsion between pairs of carboxylate groups forced into proximity by the CA structure provides an environmentally-sensitive switch, active under physiological conditions, that can control the state of assembly and disassembly (or uncoating) of the viral capsid.

Furthermore, early studies determined that mutations introduced into the loop between CA helices H6 and H7 (PPIP motif) were lethal or caused decrease in infectivity, however, they were not studied in detail (*291*). We have crystallized CA proteins bearing P122A and I124A in the PPIP motif, as well as T58A/I124A, T58S/T107I/P122A, V11I/T58A/P122A, and V11I/T58A/I124A CA mutant proteins that rescued the HIV-1 from the defects imposed by P122A and I124A. Mutations did not disrupt the fold of the CA protein; rather they affected the sites of mutations and proximal regions. Additionally, subtle variations observed at the inter- and intra-hexamer interfaces and host factor binding sites imply that mutations may alter the stability of the mature capsid core and/or host factor binding and recognition.

Finally, based on statistical analysis of the interactions between two monomers either within a hexamer or pentamer in the published capsid models, we noticed the H-bond between E28 and K30 (E28~K30) being more abandoned at the intra-pentamer, rather than intra-hexamer interfaces. We hypothesized that it might contribute to

counterbalancing the electrostatic destabilization observed in the pentamers. To address this, we tested purified mutant CA proteins, harboring R18A, E28A, and R18A/E28A mutations, for the cylinder formation *in vitro* and solved their crystal structures. Collectively, the results suggest that mutations neither affect the fold of the CA protein nor dramatically alter the CA hexamers, indirectly supporting our hypothesis that the E28~K30 H-bond exists predominantly in the pentamers. This interaction may regulate pentamer stability and may be essential for the proper CA assembly, stability, and uncoating of the HIV-1 capsid core.

In summary, the structures identified in this study support our hypothesis that the structural malleability of HIV-1 CA is a key determinant in its diverse functional roles and interactions with host factors. The possibility of capsid stability regulation through changes in pH, dNTP recruitment, DNA synthesis, and host factor binding, supports a model whereby DNA synthesis is coordinated with uncoating to cloak the viral genome from cytoplasmic DNA sensing.

BIBLIOGRAPHY

1. J. A. Moss, HIV/AIDS Review. *Radiol Technol* **84**, 247-267; quiz p 268-270 (2013).

2. G. H. Collaborators *et al.*, Estimates of global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2015: the Global Burden of Disease Study 2015. *Lancet HIV* **3**, e361-387 (2016).

3. I. C. o. T. o. Viruses. (London, UK, 2016), vol. 2017.

4. O. Narayan, J. E. Clements, Biology and pathogenesis of lentiviruses. *J Gen Virol* **70 (Pt 7)**, 1617-1639 (1989).

5. P. M. Sharp, B. H. Hahn, Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine* **1**, a006841 (2011).

6. J. Hemelaar *et al.*, Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* **25**, 679-689 (2011).

7. D. L. Robertson, P. M. Sharp, F. E. McCutchan, B. H. Hahn, Recombination in HIV-1. *Nature* **374**, 124-126 (1995).

8. K. Hryckiewicz, M. Bura, A. Kowala-Piaskowska, B. Bolewska, I. Mozer-Lisewska, HIV RNA splicing. *HIV & AIDS Review* **10**, 61-64 (2011).

9. C. M. Swanson, M. H. Malim, SnapShot: HIV-1 proteins. *Cell* **133**, 742, 742 e741 (2008).

10. A. D. Frankel, J. A. Young, HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**, 1-25 (1998).

11. L. A. Carlson *et al.*, Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe* **4**, 592-599 (2008).

12. Huang Y *et al.*, Incorporation of excess wild-type and mutant tRNA(3Lys) into human immunodeficiency virus type 1. *J. Virol.* **68**, 7676-7683 (1994).

13. A. A. Onafuwa-Nuga, A. Telesnitsky, S. R. King, 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA* **12**, 542-546 (2006).

14. C. Tian, T. Wang, W. Zhang, X. F. Yu, Virion packaging determinants and reverse transcription of SRP RNA in HIV-1 particles. *Nucleic Acids Res* **35**, 7288-7302 (2007).

15. L. Houzet *et al.*, HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res* **35**, 2695-2704 (2007).

16. D. Muriaux, J. Mirro, D. Harvin, A. Rein, RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci U S A* **98**, 5246-5251 (2001).

17. C. Liang, J. Hu, R. S. Russell, M. Kameoka, M. A. Wainberg, Spliced human immunodeficiency virus type 1 RNA is reverse transcribed into cDNA within infected cells. *AIDS Res Hum Retroviruses* **20**, 203-211 (2004).

18. K. E. Giles, M. Caputi, K. L. Beemon, Packaging and reverse transcription of snRNAs by retroviruses may generate pseudogenes. *Rna-a Publication of the Rna Society* **10**, 299-307 (2004).

19. P. Zhu *et al.*, Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc Natl Acad Sci U S A* **100**, 15812-15817 (2003).

20. E. Chertova *et al.*, Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *Journal of virology* **80**, 9039-9052 (2006).

21. M. R. Tardif, M. J. Tremblay, Presence of Host ICAM-1 in Human Immunodeficiency Virus Type 1 Virions Increases Productive Infection of CD4+ T Lymphocytes by Favoring Cytosolic Delivery of Viral Material. *Journal of virology* **77**, 12299-12309 (2003).

22. J. A. Briggs *et al.*, The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **11**, 672-675 (2004).

23. J. A. Briggs, T. Wilk, R. Welker, H. G. Krausslich, S. D. Fuller, Structural organization of authentic, mature HIV-1 virions and cores. *The EMBO journal* **22**, 1707-1715 (2003).

24. S. K. Lee, M. Potempa, R. Swanstrom, The choreography of HIV-1 proteolytic processing and virion assembly. *The Journal of biological chemistry* **287**, 40867-40874 (2012).

25. T. Wilk, S. D. Fuller, Towards the structure of the human immunodeficiency virus: divide and conquer? *Current opinion in structural biology* **9**, 231-243 (1999).

26. M. A. Khan *et al.*, Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *Journal of virology* **75**, 7252-7265 (2001).

27. S. Kao *et al.*, Human Immunodeficiency Virus Type 1 Vif Is Efficiently Packaged into Virions during Productive but Not Chronic Infection. *Journal of virology* **77**, 1131-1140 (2003).

28. S. P. Singh *et al.*, Epitope-tagging approach to determine the stoichiometry of the structural and nonstructural proteins in the virus particles: amount of Vpr in relation to Gag in HIV-1. *Virology* **268**, 364-371 (2000).

29. B. Muller, U. Tessmer, U. Schubert, H. G. Krausslich, Human immunodeficiency virus type 1 Vpr protein is incorporated into the virion in significantly smaller amounts than Gag and is phosphorylated in infected cells. *Journal of virology* **74**, 9727-9731 (2000).

30. R. Welker, H. Kotler, H. R. Kalbitzer, H. G. Krausslich, Human immunodeficiency virus type 1 Nef protein is incorporated into virus particles and specifically cleaved by the viral proteinase. *Virology* **219**, 228-236 (1996).

31. A. Kotov, J. Zhou, P. Flicker, C. Aiken, Association of Nef with the human immunodeficiency virus type 1 core. *Journal of virology* **73**, 8824-8830 (1999).

32. D. E. Ott *et al.*, Cytoskeletal proteins inside human immunodeficiency virus type 1 virions. *Journal of virology* **70**, 7734-7743 (1996).

33. E. K. Franke, H. E. H. Yuan, J. Luban, Specific Incorporation of Cyclophilin-a into Hiv-1 Virions. *Nature* **372**, 359-362 (1994).

34. P. J. Maddon *et al.*, The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348 (1986).

35. Y. Feng, C. C. Broder, P. E. Kennedy, E. A. Berger, HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor. *Science* **272**, 872-877 (1996).

36. H. Choe *et al.*, The β-Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-1 Isolates. *Cell* **85**, 1135-1148 (1996).

37. E. Coakley, C. J. Petropoulos, J. M. Whitcomb, Assessing chemokine co-receptor usage in HIV. *Current Opinion in Infectious Diseases* **18**, 9-15 (2005).

38. G. M. Shaw, E. Hunter, HIV transmission. *Cold Spring Harbor perspectives in medicine* **2**, (2012).

39. J. Coffin, R. Swanstrom, HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harbor perspectives in medicine* **3**, a012526 (2013).

40. E. J. Arts, D. J. Hazuda, HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine* **2**, a007161 (2012).

41. G. J. Nabel, Designing tomorrow's vaccines. *N Engl J Med* **368**, 551-560 (2013).

42. Y. Z. Cohen, R. Dolin, Novel HIV vaccine strategies: overview and perspective. *Ther Adv Vaccines* **1**, 99-112 (2013).

43. J. K. Mann, T. Ndung'u, HIV-1 vaccine immunogen design strategies. *Virol J* **12**, 3 (2015).

44. G. E. Gray, F. Laher, E. Lazarus, B. Ensoli, L. Corey, Approaches to preventative and therapeutic HIV vaccines. *Current opinion in virology* **17**, 104-109 (2016).

45. S. Y. Shin, Recent update in HIV vaccine development. *Clin Exp Vaccine Res* **5**, 6-11 (2016).

46. A. Cafaro *et al.*, Development of a novel AIDS vaccine: the HIV-1 transactivator of transcription protein vaccine. *Expert Opin Biol Ther* **15 Suppl 1**, S13-29 (2015).

47. M. Caskey *et al.*, Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522**, 487-491 (2015).

48. Z. Zhang, S. Li, Y. Gu, N. Xia, Antiviral Therapy by HIV-1 Broadly Neutralizing and Inhibitory Antibodies. *Int J Mol Sci* **17**, (2016).

49. D. Finzi *et al.*, Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature medicine* **5**, 512-517 (1999).

50. T. W. Chun, S. Moir, A. S. Fauci, HIV reservoirs as obstacles and opportunities for an HIV cure. *Nat Immunol* **16**, 584-589 (2015).

51. G. Maartens, C. Celum, S. R. Lewin, HIV infection: epidemiology, pathogenesis, treatment, and prevention. *The Lancet* **384**, 258-271 (2014).

52. I. M. Rouzine, A. D. Weinberger, L. S. Weinberger, An evolutionary role for HIV latency in enhancing viral transmission. *Cell* **160**, 1002-1012 (2015).

53. B. S. Razooky, A. Pai, K. Aull, I. M. Rouzine, L. S. Weinberger, A hardwired HIV latency program. *Cell* **160**, 990-1001 (2015).

54. J. D. Siliciano, R. F. Siliciano, Recent developments in the effort to cure HIV infection: going beyond N = 1. *J Clin Invest* **126**, 409-414 (2016).

55. T. W. Chun *et al.*, *In vivo* fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nature medicine* **1**, 1284-1290 (1995).

56. H. M. Baldauf *et al.*, SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nature medicine* **18**, 1682-1687 (2012).

57. G. Doitsh *et al.*, Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* **505**, 509-514 (2014).

58. D. R. Kuritzkes, Hematopoietic stem cell transplantation for HIV cure. *J Clin Invest* **126**, 432-437 (2016).

59. L. Li *et al.*, Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. *Mol Ther* **21**, 1259-1269 (2013).

60. T. R. Gamble *et al.*, Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid. *Cell* **87**, 1285-1294 (1996).

61. B. K. Ganser-Pornillos, M. Yeager, W. I. Sundquist, The structural biology of HIV assembly. *Current opinion in structural biology* **18**, 203-217 (2008).

62. L. Deshmukh *et al.*, Structure and dynamics of full-length HIV-1 capsid protein in solution. *Journal of the American Chemical Society* **135**, 16133-16147 (2013).

63. S. Du *et al.*, Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *Journal of molecular biology* **406**, 371-386 (2011).

64. O. Pornillos *et al.*, X-ray structures of the hexameric building block of the HIV capsid. *Cell* **137**, 1282-1292 (2009).

65. G. Zhao *et al.*, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643-646 (2013).

66. C. Berthet-Colominas *et al.*, Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *The EMBO journal* **18**, 1124-1136 (1999).

67. A. J. Prongay *et al.*, Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. *Proc Natl Acad Sci U S A* **87**, 9980-9984 (1990).

68. C. Momany *et al.*, Crystal structure of dimeric HIV-1 capsid protein. *Nature structural biology* **3**, 763-770 (1996).

69. R. K. Gitti *et al.*, Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* **273**, 231-235 (1996).

70. A. S. Reicin *et al.*, Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity. *Journal of virology* **69**, 642-650 (1995).

71. A. von Poblotzki *et al.*, Identification of a region in the Pr55gag-polyprotein essential for HIV-1 particle formation. *Virology* **193**, 981-985 (1993).

72. C. T. Wang, E. Barklis, Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants. *Journal of virology* **67**, 4264-4273 (1993).

73. E. K. Franke, H. E. Yuan, J. Luban, Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **372**, 359-362 (1994).

74. J. Colgan, H. E. H. Yuan, E. K. Franke, J. Luban, Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization. *Journal of virology* **70**, 4299-4310 (1996).

75. T. R. Gamble, Structure of the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein. *Science* **278**, 849-853 (1997).

76. D. K. Worthylake, H. Wang, S. Yoo, W. I. Sundquist, C. P. Hill, Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta crystallographica. Section D, Biological crystallography* **55**, 86-92 (1999).

77. I. J. Byeon *et al.*, Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* **139**, 780-790 (2009).

78. D. Ivanov *et al.*, Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. *Proc Natl Acad Sci U S A* **104**, 4353-4358 (2007).

79. T. Dorfman, A. Bukovsky, A. Ohagen, S. Hoglund, H. G. Gottlinger, Functional Domains of the Capsid Protein of Human-Immunodeficiency-Virus Type-1. *Journal of virology* **68**, 8180-8187 (1994).

80. I. Gross, H. Hohenberg, H. G. Krausslich, *In vitro* assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *European journal of biochemistry / FEBS* **249**, 592-600 (1997).

81. L. S. Ehrlich, B. E. Agresta, C. A. Carter, Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *Journal of virology* **66**, 4874-4883 (1992).

82. R. Welker, H. Hohenberg, U. Tessmer, C. Huckhagel, H. G. Krausslich, Biochemical and Structural Analysis of Isolated Mature Cores of Human Immunodeficiency Virus Type 1. *Journal of virology* **74**, 1168-1177 (2000).

83. B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, Assembly and analysis of conical models for the HIV-1 core. *Science* **283**, 80-83 (1999).

84. S. Li, C. P. Hill, W. I. Sundquist, J. T. Finch, Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409-413 (2000).

85. B. K. Ganser-Pornillos, A. Cheng, M. Yeager, Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* **131**, 70-79 (2007).

86. J. Lanman *et al.*, Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat Struct Mol Biol* **11**, 676-677 (2004).

87. J. Lanman *et al.*, Identification of Novel Interactions in HIV-1 Capsid Protein Assembly by High-resolution Mass Spectrometry. *Journal of molecular biology* **325**, 759-772 (2003).

88. J. M. Grime *et al.*, Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nature communications* **7**, 11568 (2016).

89. M. Tsiang *et al.*, A trimer of dimers is the basic building block for human immunodeficiency virus-1 capsid assembly. *Biochemistry* **51**, 4416-4428 (2012).

90. S. Mattei, B. Glass, W. J. Hagen, H. G. Krausslich, J. A. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437 (2016).

91. J. Y. Jiang *et al.*, The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology* **421**, 253-265 (2011).

92. U. K. von Schwedler, K. M. Stray, J. E. Garrus, W. I. Sundquist, Functional Surfaces of the Human Immunodeficiency Virus Type 1 Capsid Protein. *Journal of virology* **77**, 5439-5450 (2003).

93. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, Assembly properties of the human immunodeficiency virus type 1 CA protein. *Journal of virology* **78**, 2545-2552 (2004).

94. S. Manocheewa, J. V. Swain, E. Lanxon-Cookson, M. Rolland, J. I. Mullins, Fitness costs of mutations at the HIV-1 capsid hexamerization interface. *PloS one* **8**, e66065 (2013).

95. O. Pornillos, B. K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid. *Nature* **469**, 424-427 (2011).

96. S. J. Rihn *et al.*, Extreme genetic fragility of the HIV-1 capsid. *PLoS pathogens* **9**, e1003461 (2013).

97. B. M. Forshey, U. von Schwedler, W. I. Sundquist, C. Aiken, Formation of a Human Immunodeficiency Virus Type 1 Core of Optimal Stability Is Crucial for Viral Replication. *Journal of virology* **76**, 5667-5677 (2002).

98. R. Bocanegra, A. Rodriguez-Huete, M. A. Fuertes, M. Del Alamo, M. G. Mateu, Molecular recognition in the human immunodeficiency virus capsid and antiviral design. *Virus research* **169**, 388-410 (2012).

99. X. Meng *et al.*, Protease cleavage leads to formation of mature trimer interface in HIV-1 capsid. *PLoS pathogens* **8**, e1002886 (2012).

100. P. Wacharapornin, D. Lauhakirti, P. Auewarakul, The effect of capsid mutations on HIV-1 uncoating. *Virology* **358**, 48-54 (2007).

101. M. del Alamo, J. L. Neira, M. G. Mateu, Thermodynamic dissection of a low affinity protein-protein interface involved in human immunodeficiency virus assembly. *The Journal of biological chemistry* **278**, 27923-27929 (2003).

102. U. K. von Schwedler *et al.*, Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *The EMBO journal* **17**, 1555-1568 (1998).

103. B. N. Kelly *et al.*, Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(N)(133-278). *Biochemistry* **45**, 11257-11266 (2006).

104. O. Pornillos, B. K. Ganser-Pornillos, S. Banumathi, Y. Hua, M. Yeager, Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus. *Journal of molecular biology* **401**, 985-995 (2010).

105. Z. Ambrose, C. Aiken, HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* **454-455**, 371-379 (2014).

106. B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, Assembly Properties of the Human Immunodeficiency Virus Type 1 CA Protein. *Journal of virology* **78**, 2545-2552 (2004).

107. B. K. Ganser-Pornillos, M. Yeager, O. Pornillos, Assembly and architecture of HIV. *Advances in experimental medicine and biology* **726**, 441-465 (2012).

108. S. Thenin-Houssier *et al.*, Ebselen, a Small-Molecule Capsid Inhibitor of HIV-1 Replication. *Antimicrob Agents Chemother* **60**, 2195-2208 (2016).

109. P. R. Tedbury, E. O. Freed, HIV-1 Gag: An Emerging Target for Antiretroviral Therapy. *Current topics in microbiology and immunology*, (2015).

110. C. Tang *et al.*, Antiviral inhibition of the HIV-1 capsid protein. *Journal of molecular biology* **327**, 1013-1020 (2003).

111. M. D. Ruepp *et al.*, Mammalian pre-mRNA 3' end processing factor CF I m 68 functions in mRNA export. *Mol Biol Cell* **20**, 5211-5223 (2009).

112. B. Tian *et al.*, Synthesis and antiviral activities of novel acylhydrazone derivatives targeting HIV-1 capsid protein. *Bioorganic & medicinal chemistry letters* **19**, 2162-2167 (2009).

113. Y. Jin *et al.*, SAR and molecular mechanism study of novel acylhydrazone compounds targeting HIV-1 CA. *Bioorg Med Chem* **18**, 2135-2140 (2010).

114. P. Prevelige. (2007), chap. PCT/US2006/041284.

115. L. D. Fader *et al.*, Discovery of a 1,5-dihydrobenzo[b][1,4]diazepine-2,4-dione series of inhibitors of HIV-1 capsid assembly. *Bioorganic & medicinal chemistry letters* **21**, 398-404 (2011).

116. C. T. Lemke *et al.*, Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site within the N-terminal domain of the viral CA protein. *Journal of virology* **86**, 6643-6655 (2012).

117. M. Tremblay *et al.*, Inhibition of HIV-1 capsid assembly: optimization of the antiviral potency by site selective modifications at N1, C2 and C16 of a 5-(5-furan-2-yl-pyrazol-1-yl)-1H-benzimidazole scaffold. *Bioorganic & medicinal chemistry letters* **22**, 7512-7517 (2012).

118. A. J. Price *et al.*, CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS pathogens* **8**, e1002896 (2012).

119. A. J. Price *et al.*, Host Cofactors and Pharmacologic Ligands Share an Essential Interface in HIV-1 Capsid That Is Lost upon Disassembly. *PLoS pathogens* **10**, e1004459 (2014).

120. A. Bhattacharya *et al.*, Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc Natl Acad Sci U S A* **111**, 18625-18630 (2014).

121. W. S. Blair *et al.*, HIV capsid is a tractable target for small molecule therapeutic intervention. *PLoS pathogens* **6**, e1001220 (2010).

122. L. Lamorte *et al.*, Discovery of novel small-molecule HIV-1 replication inhibitors that stabilize capsid complexes. *Antimicrob Agents Chemother* **57**, 4622-4631 (2013).

123. J. Shi, J. Zhou, V. B. Shah, C. Aiken, K. Whitby, Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *Journal of virology* **85**, 542-549 (2011).

124. T. Fricke, C. Buffone, S. Opp, J. Valle-Casuso, F. Diaz-Griffero, BI-2 destabilizes HIV-1 cores during infection and Prevents Binding of CPSF6 to the HIV-1 Capsid. *Retrovirology* **11**, (2014).

125. K. Peng *et al.*, Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *Elife* **3**, e04114 (2014).

126. L. Vozzolo *et al.*, Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein. *The Journal of biological chemistry* **285**, 39314-39328 (2010).

127. N. Y. Chen *et al.*, HIV-1 capsid is involved in post-nuclear entry steps. *Retrovirology* **13**, 28 (2016).

128. N. Goudreau *et al.*, Novel inhibitor binding site discovery on HIV-1 capsid N-terminal domain by NMR and X-ray crystallography. *ACS chemical biology* **8**, 1074-1082 (2013).

129. T. Fricke *et al.*, MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1. *Retrovirology* **11**, 68 (2014).

130. M. Niedrig *et al.*, Inhibition of infectious human immunodeficiency virus type 1 particle formation by Gag protein-derived peptides. *J Gen Virol* **75 (Pt 6)**, 1469-1474 (1994).

131. J. Sticht *et al.*, A peptide inhibitor of HIV-1 assembly in vitro. *Nat Struct Mol Biol* **12**, 671-677 (2005).

132. F. Ternois, J. Sticht, S. Duquerroy, H. G. Krausslich, F. A. Rey, The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. *Nat Struct Mol Biol* **12**, 678-682 (2005).

133. V. Bartonova *et al.*, Residues in the HIV-1 capsid assembly inhibitor binding site are essential for maintaining the assembly-competent quaternary structure of the capsid protein. *The Journal of biological chemistry* **283**, 32024-32033 (2008).

134. R. Bocanegra *et al.*, Rationally designed interfacial peptides are efficient *in vitro* inhibitors of HIV-1 capsid assembly with antiviral activity. *PloS one* **6**, e23877 (2011).

135. S. Bhattacharya, H. Zhang, A. K. Debnath, D. Cowburn, Solution structure of a hydrocarbon stapled peptide inhibitor in complex with monomeric C-terminal

domain of HIV-1 capsid. *The Journal of biological chemistry* **283**, 16274-16278 (2008).

136. H. Zhang *et al.*, A cell-penetrating helical peptide as a potential HIV-1 inhibitor. *Journal of molecular biology* **378**, 565-580 (2008).

137. M. T. Garzon *et al.*, The dimerization domain of the HIV-1 capsid protein binds a capsid protein-derived peptide: a biophysical characterization. *Protein Sci* **13**, 1512-1523 (2004).

138. F. Curreli *et al.*, Virtual screening based identification of novel small-molecule inhibitors targeted to the HIV-1 capsid. *Bioorg Med Chem* **19**, 77-90 (2011).

139. H. T. Zhang *et al.*, Dual-acting stapled peptides target both HIV-1 entry and assembly. *Retrovirology* **10**, (2013).

140. F. Li *et al.*, PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc Natl Acad Sci U S A* **100**, 13555-13560 (2003).

141. J. Zhou *et al.*, Small-Molecule Inhibition of Human Immunodeficiency Virus Type 1 Replication by Specific Targeting of the Final Step of Virion Maturation. *Journal of virology* **78**, 922-929 (2003).

142. J. M. Wagner *et al.*, Crystal structure of an HIV assembly and maturation switch. *Elife* **5**, (2016).

143. F. K. Schur *et al.*, An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science*, (2016).

144. M. A. Checkley, B. G. Luttge, F. Soheilian, K. Nagashima, E. O. Freed, The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation. *Virology* **400**, 137-144 (2010).

145. B. Muller *et al.*, HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *The Journal of biological chemistry* **284**, 29692-29703 (2009).

146. P. W. Keller, C. S. Adamson, J. B. Heymann, E. O. Freed, A. C. Steven, HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice. *Journal of virology* **85**, 1420-1428 (2011).

147. D. E. Martin, R. Blum, J. Doto, H. Galbraith, C. Ballow, Multiple-dose pharmacokinetics and safety of bevirimat, a novel inhibitor of HIV maturation, in healthy volunteers. *Clin Pharmacokinet* **46**, 589-598 (2007).

148. P. F. Smith *et al.*, Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-o-(3',3'-dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus infection. *Antimicrob Agents Chemother* **51**, 3574-3581 (2007).

149. K. Van Baelen *et al.*, Susceptibility of human immunodeficiency virus type 1 to the maturation inhibitor bevirimat is modulated by baseline polymorphisms in Gag spacer peptide 1. *Antimicrob Agents Chemother* **53**, 2185-2188 (2009).

150. E. Seclen *et al.*, High prevalence of natural polymorphisms in Gag (CA-SP1) associated with reduced response to Bevirimat, an HIV-1 maturation inhibitor. *AIDS* **24**, 467-469 (2010).

151. J. Verheyen *et al.*, High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates. *AIDS* **24**, 669-673 (2010).

152. B. N. Kelly *et al.*, Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. *Journal of molecular biology* **373**, 355-366 (2007).

153. A. Lampel *et al.*, Targeting the Early Step of Building Block Organization in Viral Capsid Assembly. *ACS chemical biology* **10**, 1785-1790 (2015).

154. S. Kortagere *et al.*, Inhibiting early-stage events in HIV-1 replication by small-molecule targeting of the HIV-1 capsid. *Journal of virology* **86**, 8472-8481 (2012).

155. E. Urano *et al.*, Novel postentry inhibitor of human immunodeficiency virus type 1 replication screened by yeast membrane-associated two-hybrid system. *Antimicrob Agents Chemother* **55**, 4251-4260 (2011).

156. E. Castro *et al.*, Characterization of New Cationic N,N-Dimethyl[70]fulleropyrrolidinium Iodide Derivatives as Potent HIV-1 Maturation Inhibitors. *J Med Chem* **59**, 10963-10973 (2016).

157. R. Domenech *et al.*, Dendrimers as potential inhibitors of the dimerization of the capsid protein of HIV-1. *Biomacromolecules* **11**, 2069-2078 (2010).

158. S. Nangola *et al.*, Antiviral activity of recombinant ankyrin targeted to the capsid domain of HIV-1 Gag polyprotein. *Retrovirology* **9**, 17 (2012).

159. J. Luban, K. L. Bossolt, E. K. Franke, G. V. Kalpana, S. P. Goff, Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73**, 1067-1078 (1993).

160. M. Thali *et al.*, Functional association of cyclophilin A with HIV-1 virions. *Nature* **372**, 363-365 (1994).

161. T. Hatziioannou, D. Perez-Caballero, S. Cowan, P. D. Bieniasz, Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. *Journal of virology* **79**, 176-183 (2005).

162. E. Sokolskaja, D. M. Sayah, J. Luban, Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *Journal of virology* **78**, 12800-12808 (2004).

163. N. A. Kootstra, C. Munk, N. Tonnu, N. R. Landau, I. M. Verma, Abrogation of postentry restriction of HIV-1-based lentiviral vector transduction in simian cells. *Proc Natl Acad Sci U S A* **100**, 1298-1303 (2003).

164. G. J. Towers *et al.*, Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. *Nature medicine* **9**, 1138-1143 (2003).

165. D. Braaten, E. K. Franke, J. Luban, Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *Journal of virology* **70**, 3551-3560 (1996).

166. E. K. Franke, J. Luban, Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the Gag-cyclophilin A interaction. *Virology* **222**, 279-282 (1996).

167. A. De Iaco, J. Luban, Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. *Retrovirology* **11**, 11 (2014).

168. Y. Li, A. K. Kar, J. Sodroski, Target cell type-dependent modulation of human immunodeficiency virus type 1 capsid disassembly by cyclophilin A. *Journal of virology* **83**, 10951-10962 (2009).

169. L. B. Kong *et al.*, Cryoelectron microscopic examination of human immunodeficiency virus type 1 virions with mutations in the cyclophilin A binding loop. *Journal of virology* **72**, 4403-4407 (1998).

170. D. N. Streblow, M. Kitabwalla, M. Malkovsky, C. D. Pauza, Cyclophilin a modulates processing of human immunodeficiency virus type 1 p55Gag: mechanism for antiviral effects of cyclosporin A. *Virology* **245**, 197-202 (1998).

171. F. X. Schmid, Protein folding: Prolyl isomerases join the fold. *Current Biology* **5**, 993-994 (1995).

172. D. Braaten, H. Ansari, J. Luban, The hydrophobic pocket of cyclophilin is the binding site for the human immunodeficiency virus type 1 Gag polyprotein. *Journal of virology* **71**, 2107-2113 (1997).

173. M. Schutkowski *et al.*, Extended binding sites of cyclophilin as revealed by the interaction with HIV-1 Gag polyprotein derived oligopeptides. *FEBS Letters* **394**, 289-294 (1996).

174. B. R. Howard, F. F. Vajdos, S. Li, W. I. Sundquist, C. P. Hill, Structural insights into the catalytic mechanism of cyclophilin A. *Nature structural biology* **10**, 475-481 (2003).

175. C. Aberham, S. Weber, W. Phares, Spontaneous mutations in the human immunodeficiency virus type 1 gag gene that affect viral replication in the presence of cyclosporins. *Journal of virology* **70**, 3536-3544 (1996).

176. T. Fricke, A. Brandariz-Nunez, X. Wang, A. B. Smith, 3rd, F. Diaz-Griffero, Human cytosolic extracts stabilize the HIV-1 core. *Journal of virology* **87**, 10587-10597 (2013).

177. V. B. Shah *et al.*, The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. *Journal of virology* **87**, 422-432 (2013).

178. S. Nisole, J. P. Stoye, A. Saib, TRIM family proteins: retroviral restriction and antiviral defence. *Nature reviews. Microbiology* **3**, 799-808 (2005).

179. M. H. Malim, P. D. Bieniasz, HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harbor perspectives in medicine* **2**, a006940 (2012).

180. T. Hatziioannou, D. Perez-Caballero, A. Yang, S. Cowan, P. D. Bieniasz, Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proc Natl Acad Sci U S A* **101**, 10774-10779 (2004).

181. T. Hatziioannou *et al.*, Generation of simian-tropic HIV-1 by restriction factor evasion. *Science* **314**, 95 (2006).

182. L. Berthoux, S. Sebastian, E. Sokolskaja, J. Luban, Cyclophilin A is required for TRIM5{alpha}-mediated resistance to HIV-1 in Old World monkey cells. *Proc Natl Acad Sci U S A* **102**, 14849-14853 (2005).

183. Z. Keckesova, L. M. Ylinen, G. J. Towers, Cyclophilin A renders human immunodeficiency virus type 1 sensitive to Old World monkey but not human TRIM5 alpha antiviral activity. *Journal of virology* **80**, 4683-4690 (2006).

184. M. Stremlau *et al.*, Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A* **103**, 5514-5519 (2006).

185. D. M. Sayah, E. Sokolskaja, L. Berthoux, J. Luban, Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569-573 (2004).

186. M. Stremlau *et al.*, The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).

187. B. K. Ganser-Pornillos *et al.*, Hexagonal assembly of a restricting TRIM5alpha protein. *Proc Natl Acad Sci U S A* **108**, 534-539 (2011).

188. B. H. Kim, A. R. Shenoy, P. Kumar, C. J. Bradfield, J. D. MacMicking, IFN-inducible GTPases in host cell defense. *Cell Host Microbe* **12**, 432-444 (2012).

189. M. Kane *et al.*, MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563-566 (2013).

190. C. Goujon *et al.*, Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559-562 (2013).

191. O. Haller, P. Staeheli, M. Schwemmle, G. Kochs, Mx GTPases: dynamin-like antiviral machines of innate immunity. *Trends Microbiol* **23**, 154-163 (2015).

192. S. Gao *et al.*, Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* **35**, 514-525 (2011).

193. Z. Liu *et al.*, The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe* **14**, 398-410 (2013).

194. C. Goujon *et al.*, Transfer of the amino-terminal nuclear envelope targeting domain of human MX2 converts MX1 into an HIV-1 resistance factor. *Journal of virology* **88**, 9017-9026 (2014).

195. M. Colomer-Lluch, L. S. Gollahon, R. Serra-Moreno, Anti-HIV Factors: Targeting Each Step of HIV's Replication Cycle. *Curr Hiv Res* **14**, 175-182 (2016).

196. S. Dettwiler, C. Aringhieri, S. Cardinale, W. Keller, S. M. Barabino, Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *The Journal of biological chemistry* **279**, 35788-35797 (2004).

197. U. Rüegsegger, D. Blank, W. Keller, Human Pre-mRNA Cleavage Factor Im Is Related to Spliceosomal SR Proteins and Can Be Reconstituted *In Vitro* from Recombinant Subunits. *Molecular Cell* **1**, 243-253 (1998).

198. K. Lee *et al.*, Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7**, 221-233 (2010).

199. A. De Iaco *et al.*, TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology* **10**, 20 (2013).

200. A. Bhattacharya *et al.*, Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proceedings of the National Academy of Sciences*, 201419945 (2014).

201. E. M. Campbell, T. J. Hope, HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nature reviews. Microbiology* **13**, 471-483 (2015).

202. A. L. Brass *et al.*, Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921-926 (2008).

203. R. Konig *et al.*, Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49-60 (2008).

204. H. Zhou *et al.*, Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495-504 (2008).

205. N. Kataoka, J. L. Bachorik, G. Dreyfuss, Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol* **145**, 1145-1152 (1999).

206. G. N. Maertens *et al.*, Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci U S A* **111**, 2728-2733 (2014).

207. V. G. Tsirkone *et al.*, Structure of transportin SR2, a karyopherin involved in human disease, in complex with Ran. *Acta Crystallogr F Struct Biol Commun* **70**, 723-729 (2014).

208. F. Christ *et al.*, Transportin-SR2 imports HIV into the nucleus. *Curr Biol* **18**, 1192-1202 (2008).

209. L. Krishnan *et al.*, The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *Journal of virology* **84**, 397-406 (2010).

210. A. De Iaco, J. Luban, Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus. *Retrovirology* **8**, 98 (2011).

211. L. Zhou *et al.*, Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. *PLoS pathogens* **7**, e1002194 (2011).

212. C. Strambio-De-Castillia, M. Niepel, M. P. Rout, The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol* **11**, 490-501 (2010).

213. A. Hoelz, E. W. Debler, G. Blobel, The structure of the nuclear pore complex. *Annu Rev Biochem* **80**, 613-643 (2011).

214. E. Hurt, M. Beck, Towards understanding nuclear pore complex architecture and dynamics in the age of integrative structural analysis. *Curr Opin Cell Biol* **34**, 31-38 (2015).

215. A. A. Labokha, A. Fassati, Viruses challenge selectivity barrier of nuclear pores. *Viruses* **5**, 2410-2423 (2013).

216. V. Le Sage, A. J. Mouland, Viral subversion of the nuclear pore complex. *Viruses* **5**, 2019-2042 (2013).

217. S. A. Kassube *et al.*, Crystal structure of the N-terminal domain of Nup358/RanBP2. *Journal of molecular biology* **423**, 752-765 (2012).

218. J. Wu, M. J. Matunis, D. Kraemer, G. Blobel, E. Coutavas, Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *The Journal of biological chemistry* **270**, 14209-14213 (1995).

219. R. Zhang, R. Mehla, A. Chauhan, Perturbation of host nuclear membrane component RanBP2 impairs the nuclear import of human immunodeficiency virus -1 preintegration complex (DNA). *PloS one* **5**, e15620 (2010).

220. T. Schaller *et al.*, HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS pathogens* **7**, e1002439 (2011).

221. K. E. Ocwieja *et al.*, HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS pathogens* **7**, e1001313 (2011).

222. F. Di Nunzio *et al.*, Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. *PloS one* **7**, e46037 (2012).

223. K. Bichel *et al.*, HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. *Retrovirology* **10**, 81 (2013).

224. D. H. Lin, S. Zimmermann, T. Stuwe, E. Stuwe, A. Hoelz, Structural and functional analysis of the C-terminal domain of Nup358/RanBP2. *Journal of molecular biology* **425**, 1318-1329 (2013).

225. A. M. Meehan *et al.*, A cyclophilin homology domain-independent role for Nup358 in HIV-1 infection. *PLoS pathogens* **10**, e1003969 (2014).

226. S. Krull, J. Thyberg, B. Bjorkroth, H. R. Rackwitz, V. C. Cordes, Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Mol Biol Cell* **15**, 4261-4277 (2004).

227. B. Fahrenkrog *et al.*, Domain-specific antibodies reveal multiple-site topology of Nup153 within the nuclear pore complex. *Journal of structural biology* **140**, 254-267 (2002).

228. N. Pante, R. Bastos, I. McMorrow, B. Burke, U. Aebi, Interactions and three-dimensional localization of a group of nuclear pore complex proteins. *J Cell Biol* **126**, 603-617 (1994).

229. T. C. Walther *et al.*, The nucleoporin Nup153 is required for nuclear pore basket formation, nuclear pore complex anchoring and import of a subset of nuclear proteins. *The EMBO journal* **20**, 5703-5714 (2001).

230. K. S. Ullman, S. Shah, M. A. Powers, D. J. Forbes, The nucleoporin nup153 plays a critical role in multiple types of nuclear export. *Mol Biol Cell* **10**, 649-664 (1999).

231. M. M. Higa, S. L. Alam, W. I. Sundquist, K. S. Ullman, Molecular characterization of the Ran-binding zinc finger domain of Nup153. *The Journal of biological chemistry* **282**, 17090-17100 (2007).

232. K. Ribbeck, D. Gorlich, The permeability barrier of nuclear pore complexes appears to operate *via* hydrophobic exclusion. *The EMBO journal* **21**, 2664-2671 (2002).

233. S. Shah, S. Tugendreich, D. Forbes, Major binding sites for the nuclear import receptor are the internal nucleoporin Nup153 and the adjacent nuclear filament protein Tpr. *J Cell Biol* **141**, 31-49 (1998).

234. J. R. Ball, K. S. Ullman, Versatility at the nuclear pore complex: lessons learned from the nucleoporin Nup153. *Chromosoma* **114**, 319-330 (2005).

235. K. A. Matreyek, A. Engelman, The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *Journal of virology* **85**, 7818-7827 (2011).

236. K. A. Matreyek, A. Engelman, Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses* **5**, 2483-2511 (2013).

237. K. A. Matreyek, S. S. Yucel, X. Li, A. Engelman, Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS pathogens* **9**, e1003693 (2013).

238. A. E. Hulme, O. Perez, T. J. Hope, Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc Natl Acad Sci U S A* **108**, 9975-9980 (2011).

239. A. E. Hulme, Z. Kelley, D. Foley, T. J. Hope, Complementary assays reveal a low level of CA associated with nuclear HIV-1 viral complexes. *Journal of virology*, (2015).

240. N. J. Arhel *et al.*, HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *The EMBO journal* **26**, 3025-3037 (2007).

241. J. Rasaiyaah *et al.*, HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* **503**, 402-405 (2013).

242. X. Lahaye *et al.*, The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity* **39**, 1132-1142 (2013).

243. X. Lahaye *et al.*, Nuclear Envelope Protein SUN2 Promotes Cyclophilin-A-Dependent Steps of HIV Replication. *Cell Rep*, (2016).

244. J. Jiang *et al.*, The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology* **421**, 253-265 (2011).

245. J. Lanman, J. Sexton, M. Sakalian, P. E. Prevelige, Jr., Kinetic analysis of the role of intersubunit interactions in human immunodeficiency virus type 1 capsid protein assembly in vitro. *Journal of virology* **76**, 6900-6908 (2002).

246. B. Heras, J. L. Martin, Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta crystallographica. Section D, Biological crystallography* **61**, 1173-1180 (2005).

247. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132 (2010).

248. W. Minor, M. Cymborowski, Z. Otwinowski, M. Chruszcz, HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta crystallographica. Section D, Biological crystallography* **62**, 859-866 (2006).

249. P. Evans, Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **62**, 72-82 (2006).

250. P. D. Adams *et al.*, PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography* **58**, 1948-1954 (2002).

251. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242 (2011).

252. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367 (2011).

253. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132 (2004).

254. A. Joshi, K. Nagashima, E. O. Freed, Mutation of dileucine-like motifs in the human immunodeficiency virus type 1 capsid disrupts virus assembly, gag-gag interactions, gag-membrane binding, and virion maturation. *Journal of virology* **80**, 7939-7951 (2006).

255. S. Dimonte, M. Babakir-Mina, S. Aquaro, HIV-1 B-subtype capsid protein: a characterization of amino acid's conservation and its significant association with integrase signatures. *Virus genes* **48**, 429-437 (2014).

256. G. Obal *et al.*, Conformational plasticity of the native retroviral capsid revealed by X-ray crystallography. *Science*, (2015).

257. L. Hilditch, G. J. Towers, A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Current opinion in virology* **4**, 32-36 (2014).

258. K. Lee *et al.*, HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. *Journal of virology* **86**, 3851-3860 (2012).

259. J. C. Valle-Casuso *et al.*, TNPO3 is required for HIV-1 replication after nuclear import but prior to integration and binds the HIV-1 core. *Journal of virology* **86**, 5931-5936 (2012).

260. I. Busnadiego *et al.*, Host and viral determinants of Mx2 antiretroviral activity. *Journal of virology* **88**, 7738-7752 (2014).

261. J. L. Fribourgh *et al.*, Structural Insight into HIV-1 Restriction by MxB. *Cell Host Microbe* **16**, 627-638 (2014).

262. J. Kong *et al.*, Characterization of the amino-terminal domain of Mx2/MxB-dependent interaction with the HIV-1 capsid. *Protein & cell* **5**, 954-957 (2014).

263. K. A. Matreyek *et al.*, Host and viral determinants for MxB restriction of HIV-1 infection. *Retrovirology* **11**, 90 (2014).

264. Z. Ambrose *et al.*, Human immunodeficiency virus type 1 capsid mutation N74D alters cyclophilin A dependence and impairs macrophage infection. *Journal of virology* **86**, 4708-4714 (2012).

265. P. E. Prevelige, Jr., New approaches for antiviral targeting of HIV assembly. *Journal of molecular biology* **410**, 634-640 (2011).

266. W. Muranyi, S. Malkusch, B. Muller, M. Heilemann, H. G. Krausslich, Super-resolution microscopy reveals specific recruitment of HIV-1 envelope proteins to viral assembly sites dependent on the envelope C-terminal tail. *PLoS pathogens* **9**, e1003198 (2013).

267. S. Thenin-Houssier, S. T. Valente, HIV-1 Capsid Inhibitors as Antiretroviral Agents. *Curr Hiv Res* **14**, 270-282 (2016).

268. A. T. Gres *et al.*, STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349**, 99-103 (2015).

269. R. Hadravova, M. Rumlova, T. Ruml, FAITH - Fast Assembly Inhibitor Test for HIV. *Virology* **486**, 78-87 (2015).

270. D. Perez-Caballero, T. Hatziioannou, F. Zhang, S. Cowan, P. D. Bieniasz, Restriction of human immunodeficiency virus type 1 by TRIM-CypA occurs with rapid kinetics and independently of cytoplasmic bodies, ubiquitin, and proteasome activity. *Journal of virology* **79**, 15567-15572 (2005).

271. L. Bulli *et al.*, Complex Interplay between HIV-1 Capsid and MX2-Independent Alpha Interferon-Induced Antiviral Factors. *Journal of virology* **90**, 7469-7480 (2016).

272. N. E. Davey *et al.*, The HIV mutation browser: a resource for human immunodeficiency virus mutagenesis and polymorphism data. *PLoS Comput Biol* **10**, e1003951 (2014).

273. R. Yang *et al.*, Second-site suppressors of HIV-1 capsid mutations: restoration of intracellular activities without correction of intrinsic capsid stability defects. *Retrovirology* **9**, 30 (2012).

274. C. C. Douglas, D. Thomas, J. Lanman, P. E. Prevelige, Jr., Investigation of N-terminal domain charged residues on the assembly and stability of HIV-1 CA. *Biochemistry* **43**, 10435-10441 (2004).

275. R. Ramalho, S. Rankovic, J. Zhou, C. Aiken, I. Rousso, Analysis of the mechanical properties of wild type and hyperstable mutants of the HIV-1 capsid. *Retrovirology* **13**, 17 (2016).

276. H. Xu *et al.*, Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. *Retrovirology* **10**, 70 (2013).

277. M. Yamashita, O. Perez, T. J. Hope, M. Emerman, Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS pathogens* **3**, 1502-1510 (2007).

278. B. M. Forshey, J. Shi, C. Aiken, Structural requirements for recognition of the human immunodeficiency virus type 1 core during host restriction in owl monkey cells. *Journal of virology* **79**, 869-875 (2005).

279. G. Zhao *et al.*, Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS pathogens* **7**, e1002009 (2011).

280. D. A. Jacques *et al.*, HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*, (2016).

281. S. Campbell, V. M. Vogt, Self-assembly *in vitro* of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *Journal of virology* **69**, 6487-6497 (1995).

282. A. C. Francis, M. Marin, J. Shi, C. Aiken, G. B. Melikyan, Time-Resolved Imaging of Single HIV-1 Uncoating *In Vitro* and in Living Cells. *PLoS pathogens* **12**, e1005709 (2016).

283. A. E. Hulme, Z. Kelley, E. A. Okocha, T. J. Hope, Identification of capsid mutations that alter the rate of HIV-1 uncoating in infected cells. *Journal of virology* **89**, 643-651 (2015).

284. S. Jun *et al.*, Direct visualization of HIV-1 with correlative live-cell microscopy and cryo-electron tomography. *Structure* **19**, 1573-1581 (2011).

285. B. Leschonsky, C. Ludwig, K. Bieler, R. Wagner, Capsid stability and replication of human immunodeficiency virus type 1 are influenced critically by charge and size of Gag residue 183. *J Gen Virol* **88**, 207-216 (2007).

286. S. Abdurahman, M. Youssefi, S. Hoglund, A. Vahlne, Characterization of the invariable residue 51 mutations of human immunodeficiency virus type 1 capsid protein on *in vitro* CA assembly and infectivity. *Retrovirology* **4**, 69 (2007).

287. G. Stubbs, Tobacco mosaic virus particle structure and the initiation of disassembly. *Philos Trans R Soc Lond B Biol Sci* **354**, 551-557 (1999).

288. F. K. Schur *et al.*, Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. *Nature* **517**, 505-508 (2015).

289. A. T. Gres *et al.*, X-Ray Structures of Native HIV-1 Capsid Protein Reveal Conformational Variability. *Science*, Submitted (2015).

290. J. A. Briggs *et al.*, Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A* **106**, 11090-11095 (2009).

291. T. Hatziioannou, S. Cowan, U. K. Von Schwedler, W. I. Sundquist, P. D. Bieniasz, Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid. *Journal of virology* **78**, 6005-6012 (2004).

292. C. Aiken, Viral and cellular factors that regulate HIV-1 uncoating. *Curr Opin HIV AIDS* **1**, 194-199 (2006).

293. M. D. Miller, C. M. Farnet, F. D. Bushman, Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *Journal of virology* **71**, 5382-5390 (1997).

294. A. Fassati, S. P. Goff, Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *Journal of virology* **75**, 3626-3635 (2001).

295. Y. Sabo *et al.*, HIV-1 induces the formation of stable microtubules to enhance early infection. *Cell Host Microbe* **14**, 535-546 (2013).

296. P. Pawlica, L. Berthoux, Cytoplasmic dynein promotes HIV-1 uncoating. *Viruses* **6**, 4195-4211 (2014).

297. Z. Lukic, A. Dharan, T. Fricke, F. Diaz-Griffero, E. M. Campbell, HIV-1 uncoating is facilitated by dynein and kinesin 1. *Journal of virology* **88**, 13613-13625 (2014).

298. B. M. Forshey, U. von Schwedler, W. I. Sundquist, C. Aiken, Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* **76**, 5667-5677 (2002).

299. J. X. Lu, M. J. Bayro, R. Tycko, Major Variations in HIV-1 Capsid Assembly Morphologies Involve Minor Variations in Molecular Structures of Structurally Ordered Protein Segments. *The Journal of biological chemistry* **291**, 13098-13112 (2016).

VITA

Anna Tadeushevna Gres was born to Tadevush and Natallia Hres on April 26th, 1988 in Lida, Grodno region, Belarus. Her interest in science dates back to early childhood. She always excelled in natural sciences and received numerous awards in chemistry in high school. She graduated with honors from Belarusian State University in June 2010, where she studied pharmaceutical chemistry within five years from 2005. Majoring in bioinorganic chemistry, she earned M.S. diploma in June 2011.

Years of extensive study at the university inspired her to undertake a greater challenge in pursuing a doctorate in chemistry. In August 2012, Anna T. Gres started her graduate studies at the University of Missouri-Columbia. During the course of her studies, Anna received multiple honors and awards, including *The Uta von Schwedler Prize for Retrovirology, 2016* (awarded annually at the "Retroviruses" meeting, Cold Spring Harbor, NY, with the purpose of honoring the accomplishments of a distinguished graduate student as he/she completes a dissertation in retrovirology), *The ICDD Ludo Frevel Crystallography Scholarship, 2016* (awarded annually by the International Center for Diffraction Data (ICDD) to support the education and research program of promising graduate students in crystallography-related fields), *Linus Pauling Poster Prize, 2015* (awarded annually at the American Crystallographic Association Annual Meeting to the best graduate or undergraduate poster presentations), as well as travel awards from the University of Missouri and American Crystallographic Association. In September 2017 she successfully completed her Ph.D. in Chemistry (emphasis area Biochemistry) under the supervision of Dr. Stefan G. Sarafianos and Dr. John J. Tanner.