TO ASSEMBLE, OR NOT TO ASSEMBLE:
THE INITIATION OF RETROVIRAL CAPSID ASSEMBLY

A Dissertation in
Microbiology and Immunology

by
John Gerard Purdy

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2009
The dissertation of John Gerard Purdy was reviewed and approved* by the following:

Rebecca C. Craven  
Associate Professor of Microbiology and Immunology  
Dissertation Adviser  
Chair of Committee

Sarah K. Bronson  
Associate Professor of Cellular and Molecular Physiology

Ira J. Ropson  
Associate Professor of Biochemistry and Molecular Biology

David J. Spector  
Professor and Distinguished Educator of Microbiology and Immunology

John W. Wills  
Distinguished Professor of Microbiology and Immunology

Richard J. Courtney  
Professor, Chair and Distinguished Educator of Microbiology and Immunology

*Signatures are on file in the Graduate School
ABSTRACT

During the complex multistep process of retroviral maturation the immature Gag lattice is cleaved apart prior to the formation of the mature capsid shell. Proper maturation and capsid assembly are required for infectivity. The capsid protein (CA) contains two domains [N-terminal domain (NTD) and C-terminal domain (CTD)] and self-assembles to enclose the genomic RNA and other viral proteins. Retroviral capsids are highly polymorphic, exhibiting a wide-range of structures, and were previously modeled as a lattice consisting of 12 CA pentamers unequally distributed among 150 - 300 hexameric rings. Three CA-CA interfaces form between assembled subunits in the mature capsid: NTD-NTD, NTD-CTD, and CTD-CTD. Very little is known about the mechanism and pathway of assembly that form the mature capsid despite being attractive targets for antiviral therapy. A full understanding of events occurring during maturation, including capsid assembly will require further advances such as: 1. establishing how the cleavage products of Gag reorganize to stimulate capsid assembly during maturation, 2. defining how the CA subunits self-assemble, 3. characterizing the CA pentamer and its interaction with the previously characterized hexamer, and 4. visualizing the capsid lattice at an atomic resolution to guide inhibitor development. Understanding capsid assembly and structure will help to elucidate how the capsid interacts with other viral components to provide necessary functions during the early stages of viral replication.

Ancillary information suggests that the assembly of the capsid is nucleated at a specific location in the maturing virion following Gag cleavage. The presumptive nucleation event is explored in greater detail in this thesis using protein from Rous sarcoma virus (RSV) and human immunodeficiency virus (HIV) and a novel in vitro system that, at least for RSV, recapitulates many features of capsid assembly occurring in the virus. Results clearly demonstrate that RSV CA assembles via a nucleation-driven mechanism in vitro. Assembly was demonstrated to be electrostatically controlled and multivalent anionic salts were
demonstrated to induce CA to form structures resembling authentic mature capsids. A complex as small as a CA dimer acted as a seed for assembly was demonstrated by isolating intermediates of the capsid assembly pathway. Biochemical evidence combined with genetic analyses supports the proposal that a similar nucleation step happens in the maturing virion.

The conserved major homology region (MHR) in the CTD was shown to be required for proper initiation of capsid assembly. Lethal MHR mutations were suppressed by secondary substitutions located in the NTD and CTD that increased the propensity of CA to assemble in vitro. The results are consistent with CA undergoing structural transformations during maturation. Proteolytic cleavage of Gag results in the accumulation of a transient CA-SP (spacer peptide, C-terminal of CA) intermediate. CA-SP enhanced the kinetics of in vitro assembly. The data suggest that CA-SP participates in nucleation of capsid assembly and the MHR participates in an early step of assembly.

Identifying icosahedral particles formed by multimerization of RSV CA protein permitted the development of a pseudo-atomic model for the retroviral CA pentamer. The data are the first description of all three mature CA-CA interactions (NTD-NTD, NTD-CTD, CTD-CTD) forming in pentamers and hexamers. Although further research is needed to understand the polymorphic nature of retroviral capsids, the results provide a major step forward in understanding the architecture of mature capsids at the atomic level. The data further support a model whereby initiation of capsid assembly involves a tripartite interaction between two NTDs and one CTD. The model invokes a critical participatory role not previously described for the interdomain NTD-CTD interaction in forming a functional retroviral capsid.
TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................................. ix
LIST OF TABLES ..................................................................................................................... x
LIST OF ABBREVIATIONS ...................................................................................................... xi
ACKNOWLEDGMENTS ............................................................................................................. xii

CHAPTER I: LITERATURE REVIEW .......................................................................................... 1
A. Introduction to and Aims of the Thesis ............................................................................... 2
B. Defining a Viral Capsid and its Functions ....................................................................... 5
  Definitions and Nomenclature used in this Thesis .............................................................. 5
  General Functions of Viral Capsids .................................................................................... 6
    Genome Packaging and Protection ..................................................................................... 6
    Moving the Genome: Capsid mediated Receptor-Binding and Intracellular Movement ......................................................................................................................... 9
    Release of the Packaged Genome: Metastability and Disassembly ............................... 10
    Capsid Participation in Genome Replication .................................................................... 11

C. Principles of Viral Capsid Morphology .......................................................................... 12
  Icosahedral Capsids ........................................................................................................... 12
  Helical Capsids .................................................................................................................. 17
  Irregular / Complex Capsids ............................................................................................... 18
  Notes on classifying capsids into a morphological group .................................................. 19

D. More than just an assembly line: Understanding Capsid Assembly ............................... 20
  Capsid Protein Self Assembly .............................................................................................. 20
  Assembly Intermediates ...................................................................................................... 21
  Assisted Assembly: Scaffolding Proteins ........................................................................... 22
  A Case Study: Bromoviral Capsid Function, Structure, and Assembly ............................ 23

E. Summary of the Three Types of Capsid Morphologies: Functions and Assembly ............ 26

F. Introduction to Retroviruses and their Replication .......................................................... 27
  Classification and Genetic Organization of Retroviruses .................................................. 27
  Early events of Replication: Viral Entry to DNA Integration ........................................... 29
  Late events of Replication: Virion Production ................................................................... 37

G. Functional and Structural roles of CA during Retroviral Replication .............................. 39
  Genetic Analysis of CA reveals a role of the Capsid during Early Events of Replication ........................................................................................................................................ 42
  Role of CA in Disassembly and in Recognition by Host Restriction Factors .......... 43
  Structural Role of CA in Immature Particles .................................................................... 45
    CA-CA interactions in Immature Particles ....................................................................... 46
      NTD Models for Immature Hexameric Interface .......................................................... 46
      MHR (major homology region) Domain-Swap for Gag Dimerization ....................... 49
      Spacer Peptide Oligomerization in Immature Particles ............................................. 50
    Fullerene Model for the Morphology of Mature Capsids ............................................ 56
    CA-CA interactions in the Mature Retroviral Capsids .................................................... 60
      NTD-NTD Hexamers of Capsids ................................................................................... 60
      CTD-CTD Dimeric Interface ........................................................................................ 63
      NTD-CTD Interdomain Interface ................................................................................ 63
Pentamers (Has anyone seen a pentamer?) .......................................................... 67
Maturation Switches and the “Hallmarks” of Mature Capsids ............................... 67
H. Modeling in situ Retroviral Capsid Assembly .................................................. 68
  Defining a Nucleation-driven Assembly Mechanism .......................................... 69
  Development of an Adequate in vitro Model for in situ Capsid Assembly .......... 73
CHAPTER II
RETROVIRAL CAPSID ASSEMBLY:
A ROLE FOR THE CA DIMER IN INITIATION .................................................... 76
ABSTRACT ............................................................................................................... 77
INTRODUCTION ..................................................................................................... 78
MATERIALS AND METHODS ............................................................................... 80
  Protein expression and purification ................................................................. 80
  In vitro capsid assembly .................................................................................. 82
  Electron microscopy (EM) ............................................................................... 82
  Critical concentration determination ............................................................. 83
  Multi-angle light scattering (MALS) ............................................................... 83
  Circular dichroism spectroscopy and intrinsic tryptophan fluorescence ............ 83
RESULTS ............................................................................................................... 84
  Effects of salts on CA assembly in vitro ......................................................... 84
  Synergy of phosphate concentration and total ionic strength ......................... 88
  Characterization of CA oligomers .................................................................... 91
  Assembly activity of the CA dimer .................................................................... 97
  HIV CA assembly in phosphate ....................................................................... 100
DISCUSSION .......................................................................................................... 101
  Electrostatic control of capsid assembly ........................................................ 104
  Role of Multivalent Anions in Capsid Assembly ............................................. 105
  Potential role of a CA dimer in nucleating assembly ....................................... 107
ACKNOWLEDGEMENTS ...................................................................................... 108
CHAPTER III
CRITICAL ROLE OF CONSERVED HYDROPHOBIC RESIDUES OF THE
MAJOR HOMOLOGY REGION IN MATURE RETROVIRAL CAPSID
ASSEMBLY ............................................................................................................. 112
ABSTRACT .......................................................................................................... 113
INTRODUCTION .................................................................................................... 114
MATERIALS AND METHODS .......................................................................... 118
  Protein expression and purification ............................................................... 118
  In vitro capsid assembly ................................................................................. 119
  Electron microscopy (EM) .............................................................................. 119
RESULTS ............................................................................................................. 120
  MHR mutations do not affect CA structure and stability ......................... 120
  Sodium phosphate induced in vitro RSV CA multimerization ...................... 123
  F167Y and L171V mutant proteins lack the ability to properly assemble, but
  assembly is restored by intradomain and interdomain suppression ............ 130
  D155Y MHR mutant protein retains partial multimerization ability ............. 139
  Lethal E162Q mutation is assembly competent in vitro ............................... 140
Defective assembly ability of L171V and F167Y is rescued by the addition of WT protein ......................................................... 141
Suppressor mutations increased the sensitivity of CA to sodium phosphate ... 147

DISCUSSION ........................................................................ 147
In vitro assembly of RSV CA .................................................. 150
A role for the MHR in controlling capsid assembly ....................... 151
MHR mutations and suppressors act upon an early event of CA assembly... 152
CA structural flexibility, assembly and capsid morphology .................. 153
Implications for the development of drug resistance to antiviral therapies .... 155

ACKNOWLEDGMENTS .......................................................... 156

CHAPTER IV
SPACER PEPTIDE FOLLOWING CA PARTICIPATES IN IN VITRO
INITIATION OF CAPSID ASSEMBLY OF ROUS SARCOMA VIRUS BUT NOT
HUMAN IMMUNODEFICIENCY VIRUS ........................................... 157

ABSTRACT ........................................................................... 158
INTRODUCTION ..................................................................... 159

MATERIALS AND METHODS .................................................. 163
Protein expression and purification .............................................. 163
In vitro assembly and data analysis ............................................ 163
Electron Microscopy (EM) ....................................................... 164

RESULTS ............................................................................... 164
Native CA Proteins Migrate According to their Predicted Mass .......... 164
In vitro Assembly of RSV CA-S and Co-assembly of CA and CA-S ......... 165
RSV SP Alters in vitro Assembly .............................................. 168
In vitro Assembly of HIV CA-SP1 ............................................ 169
Initiation of in vitro Assembly is Increased by RSV SP ..................... 178
CA-SP can Nucleate Mature CA and CA-S Assembly ..................... 178

DISCUSSION ......................................................................... 181
RSV CA-SP as a Nucleus Complex for Capsid Assembly .................. 187
RSV SP and HIV SP1 are Structurally Dynamic Peptides .................. 188
RSV Maturation and Early Steps of Capsid Assembly ...................... 189
Understanding Maturation in HIV and in vitro Assembly of HIV CA and
CA-SP1 .................................................................................. 190
Similarity with Capsid Maturation of Non-retroviruses ..................... 191
Further test for in vitro Defining SP in assembly of CA-SP .................. 192

ACKNOWLEDGEMENTS ....................................................... 193

CHAPTER V
A MISSING LINK IN RETROVIRAL CAPSIDS:
VISUALIZING THE RSV CA PENTAMER .................................... 194

ABSTRACT ........................................................................... 195
INTRODUCTION ..................................................................... 196

MATERIALS & METHODS, RESULTS and DISCUSSION ............. 198
Isolation and characterization of a 60mer CA structure ..................... 198
Icosahedral symmetry of RSV 60mer particles ............................... 198
Pseudo-atomic model of the RSV CA Pentamer .............................. 203
RSV Pentamer vs. HIV Hexamer .............................................. 209


RSV Hexamer........................................................................................................210
Increasing the Resolution for the RSV Pentamer and Hexamer Model........218
Structures Related to the in vitro Assembled T=1 Particles .........................221
Previous Descriptions of 20 nm Particles formed in vitro.........................221
Support for the Fullerene Model of Retroviral Capsids ............................222
ACKNOWLEDGEMENTS.................................................................................223

CHAPTER VI
DISCUSSION..........................................................................................................224
Summary of Data Chapters..................................................................................225
Model for the Initiation of Capsid Assembly in situ .......................................227
Disassembly of Gag and de novo Assembly of Capsid .................................233
Identifying the Dimer Seed Interface: Evidence for the Interdomain Tripartite
Interaction .............................................................................................................234
Role of the MHR during Initiation and the Effects of the Suppressors on CA
Assembly ...............................................................................................................235
The Role of CA-SP in the Initiation of Capsid Assembly ...............................237
The Role of the $\beta$-hairpin during Assembly Initiation ................................240
Does Pentameric Formation depend on the $\beta$-hairpin? ...............................241
Alternative Dimeric Seed Interface of RSV CA Multimerization .................242
Nested Capsids and Multilamellar Capsids: Evidence of an Alternative
Nucleation Site .......................................................................................................243
Retroviral Polyhedrons and Icosahedral Symmetry .......................................243
Determinants of Retroviral Capsids Polymorphism .......................................244
Summary of and Alternative Approaches to the Proposed Experiments .........245
Final Thoughts and Future Directions ..............................................................248
LIST OF REFERENCES .......................................................................................250
LIST OF FIGURES

Fig 1.1. Capsid assembly and functions of a “typical” virus.................................7
Fig 1.2. The three types of capsids.................................................................13
Fig 1.3. Variability in the hexavalent position of icosahedral capsids......................15
Fig 1.4. Assembly of bromoviruses.................................................................24
Fig 1.5. Replication of retroviruses.................................................................30
Fig 1.6. Immature provirion..........................................................................32
Fig 1.7. Structure of mature retroviral virion.....................................................34
Fig 1.8. Structure of mature retroviral CA protein.............................................40
Fig 1.9. Model of RSV immature hexamer.......................................................47
Fig 1.10. Structures of the CA CTD dimers.....................................................52
Fig 1.11. Model for the role of SP (spacer peptide) oligomerization of Gag..............54
Fig 1.12. Fullerene model of the lattice of capsids in the mature virion...................57
Fig 1.13. The mature CA NTD hexamer.........................................................61
Fig 1.14. NTD-CTD interface model.............................................................65
Fig 1.15. Current models of in situ assembly...................................................70
Fig 2.1. Multivalent anion-induced multimerization of monomeric RSV CA..........85
Fig 2.2. Effects of protein concentration, pH, and ionic strength on CA protein assembly...89
Fig 2.3. Low levels of sodium phosphate induce the formation of assembly seeds........92
Fig 2.4. Identification and characterization of CA oligomers...............................95
Fig 2.5. Assembly seeding by the addition of the oligomeric CA.........................98
Fig 2.6. Assembly of HIV CA.................................................................102
Fig 2.7. Electrostatic control of CA assembly................................................110
Fig 3.1. Ribbon diagram of the monomeric structure of RSV CA.........................121
Fig 3.2. Far UV analysis and protein stability of WT and mutant CA proteins........124
Fig 3.3. Sodium phosphate-induced assembly of purified WT and D52A proteins
Fig 3.4. Determination of *in vitro* assembly phenotypes of the lethal-suppressor series of mutations
Fig 3.5. Examination of assembled mutant proteins by EM
Fig 3.6. Assembly of L171V-A38V series of mutant proteins examined by differential centrifugation
Fig 3.7. Rescue of F167Y non-assembly phenotype by WT protein
Fig 3.8. Two lethal CTD CA mutations are defective at nucleation
Fig 3.9. Sensitivity of WT and suppressor mutant proteins to sodium phosphate
Fig 4.1. Schematic summary of CA-SP cleavage during maturation
Fig 4.2. Migration patterns purified protein
Fig 4.3. Kinetics of RSV CA, CA-S and CA-SP assembly
Fig 4.4. EM analysis of protein formed in sodium phosphate conditions
Fig 4.5. Analysis of HIV CA and CA-SP1 multimerization
Fig 4.6. Visualization of CA and CA-S co-assembled structures
Fig 4.7. Dependence of the rate of assembly on protein concentration
Fig 4.8. Co-assembly of RSV CA protein species
Fig 4.9. EM analysis of structures formed by co-assembly of RSV CA and CA-SP
Fig. 5.1. Size exclusion chromatograph (SEC) of I190V protein
Fig. 5.2. RSV CA isometric particles
Fig. 5.3. Pseudo-atomic model of the RSV CA pentamer
Fig. 5.4. Ribbon diagram of RSV CA pentamer
Fig. 5.5. Model for mature RSV CA interfaces
Fig. 5.6. Model for the interaction between capsomers in the icosahedrons
Fig. 5.7. Model for the mature RSV CA hexamer
Fig. 5.8. Charged-residues in the NTD of the pentamer
Fig 6.1. Illustrative model for initiation of RSV capsid assembly……………………………………228

Fig. 6.2. Tripartite interaction as a pseudo-atomic resolution model for assembly initiation……………………………………………………………………………………..230

Fig. 6.3. In vitro assembly of P65Q mutant suppressor CA protein……………………………..238
LIST OF TABLES

TABLE 3.1 Phenotypes of WT and mutant viruses in cells..............................117
TABLE 3.2 Thermodynamic parameters for the unfolding of WT and mutant CA........126
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala, A</td>
<td>alanine</td>
</tr>
<tr>
<td>Arg, R</td>
<td>arginine</td>
</tr>
<tr>
<td>Asp, D</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CA</td>
<td>capsid protein</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>Cryo-EM</td>
<td>cryo-electron microscopy</td>
</tr>
<tr>
<td>Cys, C</td>
<td>cysteine</td>
</tr>
<tr>
<td>DEAE</td>
<td>diethylaminoethyl</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>Env</td>
<td>envelope glycoprotein</td>
</tr>
<tr>
<td>env</td>
<td>envelope gene of retroviruses</td>
</tr>
<tr>
<td>EM</td>
<td>electron microscopy</td>
</tr>
<tr>
<td>Gag</td>
<td>group specific antigen</td>
</tr>
<tr>
<td>gag</td>
<td>gene encoding Gag</td>
</tr>
<tr>
<td>Gdn-HCl</td>
<td>guanidine hydrochloride</td>
</tr>
<tr>
<td>Gln, Q</td>
<td>glutamine</td>
</tr>
<tr>
<td>Glu, E</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>Gly, G</td>
<td>glycine</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HMW</td>
<td>high molecular weight</td>
</tr>
<tr>
<td>Ile, I</td>
<td>isoleucine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>IP</td>
<td>inositol phosphates</td>
</tr>
<tr>
<td>LMW</td>
<td>low molecular weight</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>Lys, K</td>
<td>lysine</td>
</tr>
<tr>
<td>MA</td>
<td>matrix</td>
</tr>
<tr>
<td>MALS</td>
<td>multi-angle laser light scattering</td>
</tr>
<tr>
<td>Met, M</td>
<td>methionine</td>
</tr>
<tr>
<td>MHR</td>
<td>major homology region</td>
</tr>
<tr>
<td>MLV</td>
<td>murine leukemia virus</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocapsid</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NTD</td>
<td>N-terminal domain</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>Phe, F</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>Pol</td>
<td>polymerase</td>
</tr>
<tr>
<td>pol</td>
<td>polymerase gene</td>
</tr>
<tr>
<td>PR</td>
<td>protease</td>
</tr>
<tr>
<td>pro</td>
<td>protease gene</td>
</tr>
<tr>
<td>Pro, P</td>
<td>proline</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RSV</td>
<td>Rous sarcoma virus</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate – polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ser, S</td>
<td>Serine</td>
</tr>
<tr>
<td>SP</td>
<td>spacer peptide</td>
</tr>
<tr>
<td>SU</td>
<td>surface</td>
</tr>
<tr>
<td>T</td>
<td>triangulation number</td>
</tr>
<tr>
<td>Trp, W</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Tyr, Y</td>
<td>tyrosine</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

Both science and life are truly collaborative endeavors. As such my successes in both science and life have been made possible by many other people who have been willing to assist me. I am grateful to my advisor, Dr. Rebecca Craven. Under her guidance I gained many experiences that pressed me to become a better scientist, teacher and scholar. Dr. Craven’s encouragement provided me with the perseverance that allowed me to breakthrough my scientific failures, and her excitement for new scientific understandings helped propel me to my successes.

My gratitude is also extended to the members of my committee: Drs. Sarah Bronson, Ira Ropson, David Spector, and John Wills. Each of my committee members taught me how to approach not only science but also life as a scientist. Many of my successes are due, in part, because of their generosity in sharing their time and thoughts with me. I am also thankful to the faculty and staff of the Departments of Microbiology and Immunology and Biochemistry and Molecular Biology for their support, encouragement and guidance. Thank you to Dr. Richard Courtney for providing an open and collaborative environment for all of the graduate students to learn.

I am grateful to Roland Myers for the many long afternoons spent looking on a small screen in a dark room and for not thinking that I was too crazy when I suggested that we give a name to the scope. Drs. Alasdair Steven, Giovanni Cardone, and Naiqian Cheng of the NIH are gratefully acknowledged for their work on the cryo-EM imaging and reconstruction of some of “my little guys.” They were wonderful collaborators and provided great insight and understanding to my project.

To Parvez and Tam-Linh, and the members of the Wills, Parent and Courtney lab, thank you for putting up with me, a job worthy of many thanks. Kristen Rennoll-Bankert is also acknowledged for her contributions to my project as part of her SURIP research.
I am indebted to Drs. John Flanagan and Maria Bewley for opening up their lab and offices to me, which consequently expanded my graduate student experiences well beyond anything I ever hoped for or imagined. Both John and Maria are extraordinary scientists that hold a high standard for experimental evidence and the scientist that interpret that evidence. They have been an invaluable resource for me and much of the success I found during my research is due in large part to their guidance. It has been a real pleasure to work with John and I truly enjoyed every awkward congratulatory “high-five” that I received. Thanks to all the members of my adopted lab: Brian Tash, Shorena Nadaraia, Dan DeCotiis, Kathy Griffin, and Nikki (a.k.a. Nicole) Gudleski for all of their assistance over the years and for making the lab a fun and interesting place to work. Many of our conversations will hopefully never be forgotten (and repeated).

For bringing life to south central PA and grad school, I thank the many friends that I have made (too many to name individually, so they are grouped): my core group, the micro crew, the volleyball people, and the biochem crew. Carl Orff and John Williams are also acknowledge for composing Carmina Burana and the Dual of the Fates (respectively) that provided the proper “musical accompaniment” to my time spent on my research and thesis.

My deepest gratitude goes to Brian. His support, patience, and understanding amaze me every day. Brian, along with Peri and Zeus Patrick makes everything worthwhile.

And finally: “Good Luck” to the future researchers that will have the great fortune to study CA and retroviral capsid assembly.
CHAPTER I

LITERATURE REVIEW
A. Introduction to and Aims of the Thesis

Viruses are infectious macromolecular complexes consisting of proteins and nucleic acids; many viruses have a protein shell called the capsid that encloses the genome. During the replication of viruses the capsid participates in essential functions such as protecting the genome, assisting entry into an uninfected cell and assisting in intracellular movement of viral particles. For some viruses, including retroviruses, the capsid is required for successful genome replication. Specifically, retroviral particles lacking a discernable capside failed to properly reverse transcribe the genomic RNA into DNA, a requisite step for infectivity.

Some of the functional aspects of a viral capsid are dependent on the capsid being a fully closed shell with a distinct structure (morphology). Most viruses have capsids with regular symmetries, either icosahedral or helical. Well studied icosahedral and helical capsids provide a framework for defining the function and structure of many viral capsids (sections B and C). Much less is known about capsids that lack regular symmetry, classified as irregular/complex, since their unique nature makes them less amenable to structural determination.

The capsids of retroviruses lack overall icosahedral or helical symmetry (33, 54, 58, 164, 198, 235, 518). The capsid morphologies of retroviruses are diverse, ranging from irregular polyhedral to cone-like to tubular (164, 474). Furthermore, capsids of a single species of retrovirus are polymorphic, showing a large particle to particle variability. For example, no two Rous sarcoma virus (RSV) or human immunodeficiency virus type-1 (HIV) capsids can be perfectly superimposed (33, 50, 58, 198). Retroviral capsids, despite this large amount of structural heterogeneity, have similar functions during viral replication suggesting their functionality is not rigorously dependent upon the overall morphology. Thus functional retroviral capsids, unlike viruses with icosahedral or helical capsids, have a range of structural morphologies.
Examination of the morphology of authentic capsids in mature retroviruses, mostly those of HIV, at a low to moderate level of resolution by cryo-electron tomography has provided limited information about the molecular architecture of the capsid shell and no information into the functional role of the capsid during replication (33, 50, 58). The lack of experimental tools to structurally examine the capsid at an atomic level hinders understanding retroviral capsid polymorphism. Alternatively, the building block of retroviral capsids, the CA protein, has been studied at an atomic level by both crystallography and nuclear magnetic resonance. Models based on *in vitro* assembled CA propose a capsid is formed by a lattice consisting of both CA pentamers and hexamers (section G). However, gaps in understanding how CA assembles a polymorphic capsid prevent building a pseudo-atomic model of an authentic capsid based on the high-resolution structure of CA and the low resolution structure of the native capsids.

Electron tomographic studies of mature virions provided only information on the overall shape of the capsids and have failed to reach a high enough resolution to observe capsomers or the CA subunits (33, 50, 58). Visualization of CA forming certain interactions presumed to exist in the native polymorphic capsids was achieved by *in vitro* assembly; although the *in vitro* conditions have been inefficient at inducing CA to form all the interactions found in the native capsid structures (163, 276). In vitro studies have also failed to clearly establish how the cleavage products of Gag reorganize during maturation to prompt the CA subunits to self-assemble into the capsid. Thus, any information regarding the pathway of capsid assembly is purely speculative. In vitro assembly of purified CA and tomographic examinations of mature retroviral virions provide no direct evidence concerning the assembly process of capsids; however those studies led to the proposal that assembly is initiated by a specific oligomeric complex (33, 50, 58, 117, 259). This idea guided the formation of the central hypothesis examined in this thesis: initiation of retroviral capsid assembly is dependent on a nucleating complex of CA.
To analyze retrovirus capsid assembly and morphology, an investigation into the assembly pathway was undertaken in parallel with structural studies. Methods to examine capsid assembly within the maturing retroviral virion (in situ) are currently lacking and previous in vitro methods utilizing HIV CA failed to efficiently produce structures resembling authentic capsids (125, 127, 162, 164, 179, 259, 276). Thus, RSV CA was selected as a model system since previous genetic analysis identified CA mutants which may stimulate assembly providing a more adequate in vitro assembly system (48). Also, limited in vitro studies showed RSV CA assembling a greater amount of biologically relevant structures in vitro than HIV CA (234, 297). Much of the work presented in this thesis builds upon previous in vitro examination RSV CA performed by Campbell et al. (64), Kovari et al. (244), Kingston et al. (234), and Mayo et al. (297). Additionally the extensive amount of genetic information for RSV CA (48, 61, 95, 96, 251, 283, 430, 431) provides a biological framework and test for any new structural information derived from in vitro assembly of CA.

As a way to analyze the presumptive nucleation event, in vitro conditions were defined to induce purified RSV CA to form all of the known CA-CA subunit interactions of the capsid. By exploiting this system, a number of new findings were made (chapters II-V). The in vitro multimerization of CA clearly demonstrates a nucleation-driven process and provides the first description of intermediates of retroviral capsid assembly. Also a novel role for the highly conserved major homology region (MHR) of CA in the initiation of capsid assembly is presented. The assembly of cleavage intermediates of Gag shows that an immature form of CA stimulates the in vitro initiation of capsid assembly. In collaboration with Alasdair Steven the visualization of the building blocks of retroviral capsids (i.e. pentameric and hexamers complexes of CA) provides strong evidence in support of the pentamer-hexamer model previously proposed for authentic capsids (162, 210). All the interactions presumably found in native retroviral capsids occurred in vitro for RSV CA as evident by the formation of both pentamers and hexamers. Finally, a model for the initiation
of retroviral capsid assembly based on the biochemical and structural evidence described here and supported by previous genetic data is discussed (Chapter VI). The results provide a foundation for understanding capsid assembly, polymorphism and function during viral replication.

### B. Defining a Viral Capsid and its Functions

**Definitions and Nomenclature used in this Thesis**

Short definitions for certain terms used throughout this thesis related to how capsids are formed and participate in infection are provided here. *Capsids* are protein shells that enclose the nucleic acids of a virus. Some viruses have a membrane envelope surrounding the capsid while others are nonenveloped. For numerous viruses the capsid shell is called a coat or core.

Capsids are built from small building blocks called *protomers* (Fig. 1.1). These building blocks may be one or more coat, core, or capsid protein(s) with the exact terminology specific to each virus. Each monomer of protein in a capsid is often referred to as a *subunit*. Repeating morphological units visualized on electron micrographs of capsids are called *capsomers*. In many cases a capsomer is either a pentameric or hexameric complex of capsid protein and may appear as a distinct cluster of projections or knobs on the capsid surface. For many viruses, including retroviruses, capsids are required to form an infectious viral particle or the *virion*.

The term *nucleocapsid* is variously used throughout the literature and often refers to the association of capsid protein with the viral nucleic acids and, sometimes, is used interchangeably with the term capsid. In addition to the capsid (CA) protein, retroviruses contain a RNA-binding protein called *nucleocapsid* (NC). In the context of retroviruses, *nucleocapsid* refers only to the NC protein. The capsid shell in the mature retroviruses
encloses the NC—RNA complex. The retroviral capsid along with the NC-coated RNA and other viral components form the \textit{core} of the mature virion. Since the majority of this thesis concerns retroviruses, the word \textit{nucleocapsid} is used only in reference to retroviral NC protein.

Retroviruses are classified into three groups according to their sequence relatedness and replicative strategies. The three groups are Orthoretroviruses, Supmaretroviruses, and unclassified retroviruses. A detailed description of each is beyond the scope of this thesis. Since both RSV and HIV are Orthoretroviruses, this group of viruses will be the primary retroviruses discussed. For simplicity, the term \textit{retrovirus} in this thesis will refer to viruses belonging to the Orthoretroviruses group.

\textbf{General Functions of Viral Capsids}

In the virion or during the course of viral replication the capsid may perform multiple functions (Fig. 1.1). The capsid may package, protect, transport and support synthesis of the genomic material. Examples of the possible mechanisms by which capsids function will be described below and when needed for increased clarity specific viruses will be discussed.

\textit{Genome Packaging and Protection}

A single virion can be exposed to multiple environments with varying temperature, pH, osmotic pressure, and host antiviral agents as the genome is transmitted from an infected cell to a new host. In order for the virus to infect a new host, the genomic DNA or RNA must be stable in intracellular and extracellular environments. When the capsid is not present or is weakened the genomic material has a greater likelihood of being physically or chemically damaged and more easily degraded (discussed in (76, 149, 211, 542, 543)). Thus the capsid
Fig 1.1. Capsid assembly and functions of a “typical” virus. (a) Capsid formation is the result of self-assembly of viral capsid protomers, or building blocks. A capsid is made of many copies of protein subunits which are organized into morphologic units visualized by electron microscopy call capsomers. (b) The capsid shell protects the viral genome (i.e. by blocking nucleases from accessing the genome) and may aid in binding to and movement in a host cell. The capsid must package a newly synthesized genome and in some cases assist with specific steps of genome replication.
shell provides a protective coat for the genomic nucleic acids of viruses from harsh environments.

To protect the genome, the nucleic acids must be packaged inside of the capsid shell. The genomic material can either be packaged into preformed capsids or the capsid may assemble around the genome. To facilitate packaging of the genome into a preformed empty shell, several viruses have a portal complex at one vertex of the capsid shell (69, 82, 156, 157, 329, 421). Portal complexes are molecular motors that pack the nucleic acids into a relatively small volume (133, 231, 374, 421, 469). Alternatively the capsid may assemble around the genome. For example, satellite tobacco mosaic virus capsid protomers (coat protein dimers) bind the genomic RNA concurrent with RNA synthesis (262-264). As more protomers bind, intersubunit interactions establish the structure of the capsid (264, 408). Once capsid assembly and genome packaging is complete the particle forms a stable complex which is critical to protecting and transferring the genome to a new host.

Moving the Genome: Capsid mediated Receptor-Binding and Intracellular Movement

Successful viral infection requires attachment to and subsequent release of the viral genome into an uninfected cell. For some viruses, particularly nonenveloped viruses, attachment and entry are achieved by specialized domains in their capsids. The receptor-binding capacity of capsids is closely related to its structure. Some viruses attach to the cell via a site on the surface of the capsid (186, 271, 345, 390). For other viruses, such as rhinoviruses, the receptor-binding site is located in crevices or canyons in the capsid (31, 320). This observation led to the canyon hypothesis; accordingly, burying the receptor-binding site below the exterior surface of the capsid prevents antigenic recognition of the conserved residues that specify binding to the cellular receptor (389). Enveloped viruses, such as retroviruses, attach to host cells independently of their capsids (section F).
After receptor binding, some viruses enter the host cell where the viral capsid may direct intracellular migration as the particle proceeds to the site of replication (113, 173, 378). Herpesviruses provide a good example of capsid-assisted intracellular migration of the viral genome. Upon infecting a cell, the herpesviral genome is transported to the nucleus inside of the intact capsid (114, 336, 422, 423, 428). After reaching the nucleus the viral genome is released from the capsid shell. Inside the nucleus, synthesized viral genomes are packaged into newly assembled capsids; the intact genome-capsid complex is then exported from the nucleus and transported to the final site of assembly (143, 306, 422, 423).

The intracellular transportation of the viral capsids of herpesviruses as well as numerous other viruses requires microtubules, dynein and/or other cellular machinery (114, 378, 414, 449). The viral capsid may interact directly with the cellular machinery (113, 118, 378, 449). However, the interaction between capsids and the cellular transport machinery may in some cases be indirect (i.e. mediated by a capsid-associated protein) (385, 485). In either case, direct or indirect interaction, the capsid shell remains intact to assist the movement of the viral genome.

Release of the Packaged Genome: Metastability and Disassembly

Capsids are metastable: stable enough to protect and transport the genome but weak enough to release or eject the genome at the right time and location during infection. Structural changes in capsids may be induced by receptor-binding and irreversibly prime the capsid for disassembly (uncoating) and prime the capsid to release the genome (119, 186). Receptor-binding induced conformational changes and accompanying weakening of the capsid has been noted for non-enveloped viruses (119, 170, 174, 201, 425, 426, 487). Similarly, host-induced structural changes in the virion of enveloped viruses have been noted; however, structural changes to the capsid lattice have not been observed (260, 294, 303, 522).
In many cases the exact mechanisms of uncoating (i.e. the molecular changes to the capsid or to the genomic material) are unknown, however, tobacco mosaic virus provides a good example of the molecular steps involved in disassembly (76, 447, 481). Upon entering a plant cell, protons and calcium ions are lost from the coat protein and genomic RNA, resulting in repulsion between negatively-charged residues of the coat protein and the RNA phosphate groups. Consequently, ~20 coat protein subunits located near the 5’ end of the RNA disassociate from the virion, exposing a translational start codon. Complete uncoating is achieved when the RNA is translated by the ribosome (307, 415, 501). Thus, synthesis of the new genomic RNA is also coupled with capsid disassembly.

*Capsid Participation in Genome Replication*

Some viral capsids participate in genome replication prior to releasing the genome. For example, in hepatitis B virus a pre-genomic viral RNA is packaged into the capsid prior to being reverse transcribed into the genomic DNA (356, 450). In this case, synthesis of the genome occurs prior to exiting from an infected cell. Certain hepatitis B capsid mutants prevent proper genome replication, demonstrating that the capsid protein is critical for proper reverse transcription (239, 331, 407, 524). Similarly, some mutations in CA of retroviruses disrupt genomic synthesis suggesting that reverse transcription in retroviruses also occurs in a capsid shell, as discussed later (section G) (61, 96, 145, 290, 477). The capsid in these two reverse transcribing viruses may ensure proper DNA synthesis by providing a protected environment or by organizing the reverse transcribing complex. The exact mechanism by which the capsid participates in the reverse transcription step in either hepatitis B virus or retroviruses is unknown; however, clearly some capsids have a functional role in the viral replication beyond protection and transportation of the genomic material.
C. Principles of Viral Capsid Morphology

Viral capsids are vessels or containers that, among other functions, packages and protects the viral genome. Vessels with a wide variety of shapes could perform these functions, however, for the most part, only a limited number of capsid structures have been noted for a great number of viruses. Viral capsids can be classified into three basic types of symmetry: icosahedral, helical or irregular/complex (Fig. 1.2) (142, 238, 287, 365, 368).

Icosahedral Capsids

The structures of many closed spherical viral particles are modeled according to icosahedral principles as originally described by Caspar and Klug (76, 98). A regular icosahedron is a polyhedron that has 20 equilateral triangular faces, with 5 of the faces meeting at each of the 12 vertices (Fig. 1.2A). An icosahedral particle has 2-, 3- and 5-fold symmetry. The 2-fold axis means that the capsid is identical when rotated 180º; 3-fold when rotated 120º; and 5-fold when rotated 72º. Each of the 12 vertices is located at the 5-fold axis. The morphological unit of icosahedral capsids are either pentavalent or hexavalent capsomeres (76, 98).

The simplest icosahedral viruses have 60 identical protein subunits repeated throughout the capsid with equivalent interactions. Icosahedral capsids may have shells larger than 60 subunits but the subunits of these capsids lack the constraint of strict equivalence. These similar but not identical intersubunit interactions are called quasi-equivalent (76). As such, icosahedral viral capsids have a total number of subunits equal to certain multiples of 60 (i.e. 60T subunits). T, the triangulation number, describes the spatial relationship of pentamers in the capsid lattice. Icosahedral capsids contain 12 pentavalent (5-fold) structures plus 10(T-1) hexavalent structures (76). For example, a T=1 capsid will have 12 pentamers plus 10(1-1) or 0 hexamers for a total of 60 subunits and a T=3 will have 12 pentamers plus
Fig 1.2. The three types of capsid structures. (a) A T= icosahedron showing the 2-, 3-, and 5-fold axis of symmetry. The pentamer and hexamer outlined. The virus particle is a bromovirus (PDB: 1js9) and the image was made using Chimera. (b) An example of a helical capsid, tobacco mosaic virus (60). (c) The morphologies of complex/irregular capsid widely vary. Shown here is an acidianus bottle virus (188). Please note that viruses are not shown in the same scale.
Fig. 1.3. Variability in the hexavalent position of icosahedral capsids. (a) The T=7 papilloma virus capsid is shown. The capsid consists of 72 pentamers, of which 60 are found in the hexavalent position. Instead of containing 420 subunits as is predicted for a T=7 capsid it contains only 360 subunits. The particle was formed by recombinant L1 protein and visualized by cryo-electron microscopy (cryo-EM) as described in (309). (b) The hexavalent position capsomers of paramecium bursa chlorella virus consist of three Vp54 capsid protein monomers (shown on the right). The T=169 icosahedral shell contains 12 pentagonal caps (yellow) and 20 triangular units (red, green, blue, purple, orange). The pentagonal caps contain 30 trimeric capsomers and each of the triangular units contains 60 trimeric capsomers arranged in hexameric symmetry (326).
20 hexamers for total of 180 subunits. The value of $T$ can be any number in the series 1, 3, 7, 13, 19, 21... (as defined by $h^2 + hk + k^2$, where $h \geq 1$ and $k \geq 0$).

The icosahedral model for capsid morphology was inspired by the architectural work of Buckminster Fuller (75, 76). Structural data from bushy stunt virus and poliovirus provided early experimental support for the icosahedral model (74, 140). Currently, the model proposed by Caspar and Klug is supported by the visualization of over 130 icosahedral structures of viral capsids (72).

However some viruses have found unique solutions, not originally predicted by Caspar and Klug, to arrive at the same foundation of icosahedral symmetry and many icosahedral capsids are structurally diverse. For example, hexavelent and pentavalent capsomeres do not necessarily need to be a hexameric or pentameric complex of protein (Fig. 1.3) (32, 280, 326, 380, 466, 514). Thus, the simple model originally proposed by Caspar and Klug fails to predict certain aspects of morphology for some capsids.

**Helical Capsids**

Helical capsids are formed by repeated packing of protein subunits in a spiral array. Helical capsids appear as short or long (up to $\mu$m in length) rods or tubes (Fig. 1.2). These capsids can consist of several thousand identical protein subunits forming a sheath around the genomic DNA or RNA. Similar to quasi-equivalently packed icosahedrons the subunits in the helical arrangement are nearly identical (76, 491). For example tobacco mosaic viral capsids have about 2000 coat-protein subunits, each of which interact with three nucleotides of the genomic RNA and six neighboring coat subunits, however, only three non-identical intersubunit interactions are formed (76, 237, 325, 400).

Protein subunit interactions with the viral genome are important for the helical capsids of tobacco mosaic virus but this is not necessarily true for all helical capsids.
Bacteriophage inoviruses have rod-shaped capsids that enclose and protect the DNA genome, but the capsid lacks the extensive protein subunit—nucleic acid interactions described above for tobacco mosaic virus (491-493). For both inoviruses and tobacco mosaic virus, the helical nature of the virion is only due to the coat protein subunits and is not dependent on the genome (265, 455).

**Irregular / Complex Capsids**

The majority of viral capsids have either icosahedral or helical symmetry; however some viruses have irregular or complex morphotypes. These unique capsids are asymmetric structures that lack overall icosahedral or helical symmetry (Fig. 1.1c). In many regards this category is the “other / miscellaneous” group; that is, the structures are not easily identified as icosahedral or helical. This category contains highly diverse structures including morphotypes that are unique to a single virus; and thus, each capsid in this category may represent a novel architecture, biochemical mechanism for assembly, or functional property. Viruses whose capsids are classified as complex include: retroviruses, poxviruses, acidianus bottle-shaped virus, sulfolobus neozealandicus droplet-shaped virus, and sulfolobus spindle-shaped virus (14, 40, 58, 99, 162, 164, 176, 188, 189, 197, 235, 238, 254, 342, 366-368, 377, 384, 427, 474, 507, 518).

Similar to icosahedral and helical capsids, irregular/complex capsids are formed by assembly of one or more protein subunits. However, in irregular/complex capsids the packing of each subunit is not limited to a few quasi-equivalent interactions. Instead a far greater number of subunit conformational interactions or intersubunit interactions appear to underline the architecture of irregular/complex capsids.

The asymmetrical nature of viral capsids in this category limits the ability to structurally define the subunit interactions by traditional methods that rely on examining
many presumably similar structures. However in the case of retroviruses the advancement of genetic, structural and biochemical tools have shed considerable light on how complex capsids are built and how the capsid contributes to the viral lifecycle. However, many questions remain unanswered: What guides the assembly process, and what controls morphology? Why do the irregular/complex capsids have such unique architecture? How do the subunits interact to form the capsids? Do the irregular structures influence the functional properties of the capsids? The goal of sections F-H of this chapter and the following chapters of this thesis is to examine the assembly, structure and function of retroviral capsids.

**Notes on classifying capsids into a morphological group**

The exact classification of viral capsids into the three categories has varied (142, 238). As noted above, the definition of irregular/complex capsids used in this thesis is an asymmetric capsid that lacks discernable icosahedral or helical symmetry. This definition is similar to one used by Haring et al. to describe novel morphotypes not previously observed (188). However some previous reviews have described viral capsids with localized regions of icosahedral symmetry but lack uniform symmetry as complex icosahedrals. For example, herpesviruses and many phage viruses have a portal complex located at one axis. Similarly the capsids of mimivirus, T4 phage, and reoviruses have been described by some as complex (80, 81, 93, 177, 212, 266, 509, 510, 525). Data presented in this thesis concerns both icosahedral and irregular/complex viral capsids, however in this document “complex” icosahedrons are simply classified as icosahedral capsids. Importantly, the irregular/complex category is necessary when a comprehensive review of capsid morphotypes includes non-icosahedral or non-helical capsids.
D. More than just an assembly line: Understanding Capsid Assembly

Viral capsids are large multi-subunit structures that have very distinct morphologies, but how does a protein subunit “know” to adopt a proper conformation or interaction with a neighboring subunit? Overall, the pathway of an assembling capsid is complicated but well regulated in vivo. Examination of the assembly process has been carefully studied for some viruses by utilizing the ability of capsid protein subunits to self-assemble in vitro. Many viral capsids can form “spontaneously” from purified components and modeling of this apparent spontaneous process has provided conceptual ideas on the assembly process (73, 79, 86, 152, 166, 187, 220, 277, 322, 331, 402, 420, 483).

Capsid Protein Self Assembly

The self-assembly process is driven by conditions that favor the stability of the interactions between the capsid subunits (370, 542, 543). In vitro assembly can be controlled by multiple conditions including pH, temperature, total ionic strength, types of ions present (i.e. addition of Mg$^{2+}$ or PO$_4^{3-}$), oxidative state, and protein concentration. Varying these conditions may allow for the dissection of the assembly mechanism by allowing for biochemical, genetic and biophysical characterization of protein complexes that are intermediates of the self-assembly process. Importantly, in vitro assembly of some viral capsids has identified intermediates in the self-assembly process. Characterization of these intermediates allows for the determination of the way the subunit will be added to the growing capsid. The next subunit to be added is done so according to the local rules model (35, 411, 434). This model proposes that the conformation adapted by an assembling subunit is dependent on the conformation and relative position of its neighboring subunit(s). As such, in a growing capsid shell the self-assembly process is dependent on a set of “local rules”
rather than on larger or more global restraints. Accordingly, only a few simplified rules must be followed to form large multipart structures from many subunits.

Kinetic examinations of in vitro assembling capsids offer supporting evidence for the local rules model (129, 131, 158, 232, 298, 434, 544). Although to show biological relevance for an in vitro assembly system, evidence must be provided to demonstrate that the structures formed in vitro are similar to authentic native capsids. For several viruses, the self-assembled shells formed in vitro are highly similar or identical to those formed in vivo (20, 149, 214, 253, 408, 459). For some viruses, infectious virions can be assembled in vitro from purified protein and genomes (20, 149). However, some in vitro assembly conditions fail to recapitulate in vivo capsid formation (4, 148, 214, 253, 321, 436, 451, 459, 545). In this case the protein subunits form structures not observed in the virion or fail to form all the intersubunit interactions occurring in vivo suggesting the pathway of assembly in vitro does not fully replicated authentic in vivo assembly.

Assembly Intermediates

The local rules model highlights the importance of describing assembly intermediates in determining the assembly pathway. However, isolation of intermediate structures along the pathway of assembly is experimentally difficult due their low stability. Intermediates are quickly consumed by growing capsids under pro-assembly conditions, and under non-assembly conditions intermediates disassociate into smaller more stable units. However, in some cases such as tobacco mosaic virus or bromoviruses, intermediates can be isolated (59, 77, 78, 199, 199, 217, 331, 443, 467, 502, 545, 546).

Characterization of intermediates of the assembly pathway led to a stepwise description of capsid formation of polio virus. A viral protease cleaves a pre-cursor P1 polyprotein into three coat proteins that remain associated in a structural unit (41). The
structural units pentamerize on lipids and direct the formation of the icosahedral capsid (12, 375, 376). Packaging of the genome stimulates a final cleavage of one of the coat protein subunits resulting in an infectious polio particle (13, 139, 208).

For many other viruses (including retroviruses) in vitro and in vivo assembly occur too rapidly to isolate intermediates and appears as an all-or-none process. Previous attempts to isolate intermediates along the assembly path of retroviral capsids failed (258). Nonetheless in chapter II isolated multimeric RSV CA protein complexes were determined to be capsid assembly intermediates. Similar to polio virus, isolation and characterization of assembly intermediates may help generate a model for the pathway of retroviral capsid assembly.

**Assisted Assembly: Scaffolding Proteins**

The self-assembly process sometimes involves scaffolding proteins that help catalyze capsid assembly by providing a structural platform (115, 206, 232, 233, 370, 437, 470). For example, the icosahedral T4 phage capsid assembles in multiple steps and involves a scaffold protein (gp20) and five different capsid proteins, including two major capsid protein subunits (gp23 and gp24) (124, 206, 437). Polymerization of the subunits starts from an initiation complex of gp20 and portal proteins (311, 312, 437, 516). If no portal protein is present then tubular structures form and the initiation of capsid assembly is only slightly affected. In the absence of gp20 scaffolding protein, initiation of assembly is defective demonstrating that self-assembly, specifically initiation of assembly, may involve more than just capsid protein subunits.
A Case Study: Bromoviral Capsid Function, Structure, and Assembly

An in depth understanding of capsid assembly for many viruses has yet to be determined, but fortunately, many aspects of bromoviral capsid assembly have been defined. Bromoviruses form T=3 icosahedral capsids that package and transport the positive-sense single-stranded genomic RNA (148, 229, 229, 335). Careful kinetic and structural analyses showed that the protomer for assembly of bromoviral capsids are coat protein dimers and the pentavalent and hexavalent capsomeres can be thought of as a pentamer of dimers and a hexamer of dimers, respectively (Fig. 1.4) (4, 195, 214, 253, 532, 545).

The coat protein exclusively forms T=3 capsids in vivo (253). However in vitro self-assembly results in the formation of other organized structures in addition to the T=3 shells (4, 148, 321, 459, 545). Most obvious were the presence of tubes and two other classes of icosahedral capsids: T=1 and pseudo T=2 particles (214, 253, 451, 459, 545). The T=2 particles are termed “pseudo” since a T number of 2 is disallowed (see section C). Structural determination of the pseudo T=2 particles demonstrates how the intersubunit interactions can govern the overall morphology.

Each class of observed icosahedral particles contains 12 pentamers of dimers but the pentavalent structures of the pseudo T=2 particles are linked via a different architecture (253, 459). In this aberrant structure a new subunit interface forms by an alteration in the intersubunit interaction at the two-fold axis (Fig. 1.4b). This observation suggests the T=2 particles assemble via the association of pre-formed stable pentamers of dimers while the T=3 particles assemble by the addition of free protomer subunits (214, 253, 459, 545).

Since purified bromoviral coat protein assembled structures other than the T=3 particles, the in vitro assembly pathway is likely different than the in vivo pathway (214, 545). Investigating this concept led to the discovery of a regulatory step following the formation of a pentamer of dimers (214, 543). Accordingly, if a free protomer is added to a
Fig. 1.4. Assembly of bromoviruses. (a) The protomer for bromoviral capsid assembly is a dimeric form of the coat protein. The pentamer of dimers forms from the A (blue) and B (red) subunits while the hexamer of dimers contains three A/B and three C/C (green) protomers. The ribbon diagrams of the A, B and C subunits are also shown in similar orientations with two residues, Y49 and F119, highlighted in blue to demonstrate the flexible nature of bromoviral coat protein. The figure was generated using PDB 1cwp (429) and PDB 1js9 (285). Also note that Fig. 1.2a shows a native T=3 bromoviral capsid (b) Three types of icosahedral shells (T=1, 2 and 3) formed by in vitro assembled coat protein (green). The orientation of the capsomers around the 5-fold axis is shown each capsid. Note that in the T=2 capsid shell two pentamer of dimers associate by an antiparallel interaction along the 2-fold axis. The figure was adapted from (214).
pentamer of dimers then a T=3 structure is generated, whereas if two pentamer of dimers self-associate then a pseudo T=2 particle will form (214, 545). Thus the early events, a step as early as the association of the 6th subunit, in the assembly pathway are a critical determinant of the capsid morphology. The above discussion of bromoviruses provides an illustrative example of how understanding the capsid assembly pathway is critical to understanding the overall morphology of the capsid shell and the role of the capsid during viral replication.

E. Summary of the Three Types of Capsid Morphologies: Functions and Assembly

The morphology of the capsid affects some of the properties of the resulting particle and its functional capabilities, such as assembly and genome packaging. Icosahedral geometry provides an efficient method for enclosing a maximum volume in a closed particle, thus allowing for a greater amount of space to package the viral genomes. However, the genome must be small enough to fit in a pre-determined volume. Irregular/complex capsids may also package the genome into a limited amount of space. At the other end of the spectrum, helical capsids assemble around the genome without the requirement to package the genome into a specified volume and thus, given enough capsid protein subunits, could package an indefinitely large genome (76).

Some basic predictions of the viral replication cycle may be derived from the morphology of its capsids. For example, assembly of a helical capsid is likely unidirectional or bidirectional (21, 60, 76, 237, 284, 343) whereas, icosahedral and complex capsid may simultaneously assemble in multiple dimensions (263, 264, 408). Similar notions regarding the disassembly and release of the viral genome can be inferred from the morphology. Viruses with icosahedral or complex capsids may release or eject the genomic material without complete breakdown of the shell. While complete release of the genome from a helical capsid, such as tobacco mosaic virus, requires each subunit of the capsid to be disassembled. These generalities and predictions based on morphology assist experimental
design but any conclusion regarding the actual replication cycle must be based on experimental evidence beyond a simple description of the capsid structure.

As demonstrated above by bromoviruses, a single type of capsid protein can form multiple different morphological types suggesting that the same principles of capsid protein self-assembly with quasi-equivalent interactions can explain all three types of capsid morphology (75, 76, 162, 164). Furthermore, the results presented and discussed in this thesis show retroviral CA protein self-assembling into capsids with a complex morphology as well as icosahedral and helical structures (Chapters II-V).

F. Introduction to Retroviruses and their Replication

To provide a framework for understanding the structural and functional roles of retroviral capsids during viral replication, the replication cycle of retroviruses is briefly covered in this section. Retroviruses are enveloped viruses that contain two positive-sense RNA strands. Retroviruses have a common genome organization, virion structure and replication cycles including reverse transcribing the RNA genome into DNA and subsequent integration of the genomic DNA into the host genome (454, 474). Retroviruses infect a diverse range of hosts and are the causative agents of both oncogenic and immunodeficiency diseases.

Classification and Genetic Organization of Retroviruses

The family Retroviridae can be subdivided into three groups (Orthoretrovirinae, Spumaretrovirinae, and unclassified Retrovirinae) according to the International Committee on Taxonomy of Viruses. The Orthoretroviruses (simply referred to as retroviruses in this thesis) are further classified according to sequence relatedness into 6 genera: Alpharetrovirus,
Betaretrovirus, Deltaretrovirus, Epsilonretrovirus, Gammaretrovirus, and lentivirus. Rous sarcoma virus (RSV) is an Alpharetrovirus and human immunodeficiency virus (HIV) is a lentivirus.

In addition to sequence relatedness, retroviruses can be distinguished based on the site of their assembly and their mature core morphology (103, 474). Type A particles (Intracisternal A-type particles, type II human endogenous retrovirus) are retroviruses that bud into and accumulate within the endoplasmic reticulum, and have an immature particle morphology. Assembly of immature viruses of type B (mouse mammary tumor virus) and type D (Mason-Pfizer monkey virus) occurs within the cytoplasm prior to transport to plasma membrane. Once released from the cell the viral core takes on a mature morphology (discussed in the following section). Type B cores appear roundish and are eccentrically placed; whereas, type D cores are tubular or bar-shaped. Assembly of a type C (Rous sarcoma virus, murine leukemia virus) particle occurs at the plasma membrane. Their cores are roundish with noticeable angular regions. HIV assembly similarly takes place at the plasma membrane, but the mature cores of HIV are cone-shaped.

Each retroviral particle has two copies of a positive-stranded RNA genome. The genome contain three conserved coding regions in their genomes: gag, pro-pol and env (474). gag encodes the Gag polyprotein and pro-pol encodes the replicative enzymes: protease, reverse transcriptase, and integrase. The translation of Pol occurs as a result of a -1 reading-frame shift at the end of gag, resulting in ~ 5% of Gag being expressed as a Gag-Pol polyprotein (207). In most retroviruses pro is considered part of pol, except for certain avian retroviruses such as RSV where it is part of gag-pro. The viral glycoproteins are encoded by env. The conserved genes are found in a common order: gag-pro-pol-env. All retroviral Gag polyprotein contains three major domains: MA (matrix), CA (capsid) and NC (nucleocapsid). Less conserved regions may also be found in Gag, such as p2 / p10 and spacer peptides (SP)
that flank the RSV CA domain (474). The genome of some retroviruses such as HIV codes for additional accessory proteins.

**Early events of Replication: Viral Entry to DNA Integration**

During infection the Env glycoproteins are responsible for binding the virus to the cell receptors and, subsequently, mediating fusion of the viral and cellular membranes (Fig. 1.5) (23, 121, 299, 354, 435). The Env glycoprotein undergoes large conformational changes during the process of attachment and entry (90). Each glycoprotein is derived from the Env polyprotein. Env contains two subunits: a glycosylated surface envelope protein (SU) and a transmembrane envelope protein (TM) (Fig. 1.6 and 1.7). Alternatively, virions may enter a cell by endocytosis; however, these particles may be rendered non-infectious (134, 291, 403, 404, 445, 530).

What happens immediately following viral entry is less well understood. The core of the mature virion enters the cell, initiates reverse transcription, undergoes some form of uncoating, and delivers the genomic material to the nucleus (Fig. 1.5). In the cytoplasm, the positive-strand genomic RNA is converted to DNA by reverse transcriptase (RT) as part of a reverse transcribing complex. These complexes include the NC-coated viral RNA and associated tRNA primer, RT, integrase and likely other viral components such as CA, and potentially cellular proteins (136, 183). RT is the only viral protein required for *in vitro* reverse transcription but proper *in vivo* DNA synthesis requires other viral factors including CA (44, 45, 61, 96, 145, 382, 430, 431, 461, 477). Synthesis of viral DNA occurs via an elaborate series of steps starting with initiation by a cellular tRNA primer bound to the genomic RNA (169, 191). Negative-strand DNA is synthesized until reaching the 5’ cap of the RNA and continues at the 3’ end of the genome following a “jumping” event termed strand-transfer. Complementation between the newly synthesized negative-strand viral DNA
Fig 1.5. Replication of retroviruses. Mature virions attach via cell surface receptors and entry in the cell. Reverse transcription, uncoating and migration to the cellular nucleus leads to formation of a preintegration complex. The proviral DNA is inserted into the host genome. New viral transcripts are exported to the cytoplasm. Gag, Gag-Pol, and Env proteins are synthesized and traffic the site of immature assembly where the genomic RNA is packaged. Gag and cellular factors coordinate budding of the provirion. Activation of protease leads to the cleavage of Gag and Gag-Pol and the generation of a mature virion during maturation.
Fig. 1.6. Immature provirion. The structure of the immature particle is shown. Env timers line the outer edge of the particle while Gag and Gag-Pol multimerization forms the interior. The domains of Gag-Pol are shown below: MA (matrix), p2/p10, CA (capsid), SP (spacer peptide), NC (nucleocapsid), PR (protease), and Pol (reverse transcriptase and integrase). The figure was made with assistances from Jared Spidel and previous members of the Craven and Wills lab.
Fig. 1.7. Structure of mature retroviral virion. (a) A cartoon depiction of the mature virion structure. The CA forms the capsid shell, NC coats the genomic RNA and MA remains associated with the lipid bilayer. Authentic RSV capsids are included to show the diversity of the mature capsid structure. Cryo-EM tomographic images shows RSV (b, yellow) and HIV (c, red) virions. Colored capsids were made by single particle image reconstructions. The arrowhead in upper RSV panel and right hand panel of HIV shows nested capsid (capsid within a capsid). The images appear in (33, 58).
and the 3’ end of the RNA facilitates strand-transfer. In addition to the DNA polymerase domain, RT also has a RNase H domain for degrading the viral RNA during reverse transcription. Initiation of positive-strand DNA synthesis occurs via a RNA primer that is not degraded by RNase H. Completion of positive-strand DNA requires a second strand transfer jump. Once the final steps of reverse transcription are completed and the genome has been transported to the nucleus, the genomic DNA is integrated into the host genome by action of the viral integrase enzyme and host cellular proteins (87, 88, 472, 520).

The process of reverse transcription may introduce mutations into the genome. RT lacks proof-reading activity and is a low fidelity enzyme resulting in mutation rates higher than many cellular or viral polymerases (28-30, 351, 498). In cell culture, the RSV mutation rate is $1.4 \times 10^{-4}$ errors per nucleotide, a rate similar to that of its RT enzyme (27-29, 270). In a single host, both a high mutation and replication rate may result in a genetically heterogeneous population of virus, or a viral quasispecies (83, 84, 89, 245, 364). Although some mutations may result in a decrease in infectivity fitness, the genetic diversity may allow survival under positive selection factors by evading immune recognition or antiviral drug inhibitors (89, 346, 395). The high error rate of RT may also allow poorly infectious mutant viruses to regain infectivity by creating a second-site suppressing mutation (48, 431, 460, 517). Analysis of second-site suppressors for poorly infectious CA mutants has proven useful in understanding capsid assembly (Chapter III). In addition to the errors made by RT during reverse transcription, mutations may also be introduced by the host polymerase during transcription of new viral RNA from the integrated viral DNA during late events of replication.
Late events of Replication: Virion Production

Expression of new viruses starts with transcription of genomic RNA from the integrated provirus via a host RNA polymerase and is promoted by the viral LTR (long terminal repeat). Viral RNA is exported from the nucleus for use in translation and packaging into nascent particles. Translation leads to the production of viral proteins including Gag polyproteins and Env glycoproteins (454).

Gag and Gag-Pol molecules are targeted to the nucleus and to the plasma membrane (97, 396, 406, 454). About 5% of Gag exists as Gag-Pol in the released particle (207). Gag and Gag-Pol associate with the membrane via a membrane binding domain in MA and in some cases by myristoylation of the second Gly residue of Gag (9, 56, 100, 101, 172, 348, 348, 383, 433, 538, 539). The process by which Gag associates with the glycoprotein spikes is poorly understood but Env proteins migrate to lipid microdomains where a nascent particle will bud (272, 339). Gag and Gag-Pol multimerization, binding of the NC domain to the genomic RNA, accumulation of the viral components at the plasma membrane, and interactions with cellular factors result in the assembly and release of new particles.

More than 500 host proteins may be required during the course of HIV infection (49, 533) of which some of the best characterized are proteins involved in the budding process (255, 300, 315, 316, 412, 453, 478, 486, 529). At the site of assembly the late domain motif in Gag recruits cellular proteins that are part of the ESCRT (endosomal sorting complex required for transport) machinery (255, 300, 315, 316, 352, 353, 412, 453, 478, 486, 529). Interactions among viral and host factors facilitates the completion of budding and the release of an immature particle.

Each immature particle is spherical with a 100 – 200 nm diameter; the mean diameter of RSV particles is 125 nm and of HIV is 145 nm (Fig. 1.7) (50, 52, 58, 474, 518). The immature particle lacks icosahedral symmetry but contains a lattice of hexameric complexes.
formed by Gag (51, 505, 518). The interior portion of immature RSV, HIV and murine leukemia virus appears similar (52, 55, 518). In each particle Gag and Gag-Pol are radially aligned with the N-terminal region associating with the lipid layer and the C-termini pointing towards the center (Fig. 1.6) (52, 55, 518). Embedded in the cell-derived lipid bilayer are ~80 trimeric Env spikes (58, 146, 541). Since most retroviruses do not require Env protein to bud, released particles have a large heterogeneity in number of surface spikes ranging from 0 - 120 (58, 146, 541).

Post-budding, the viral protease (PR) dimerizes and cleaves itself out of Gag-Pro or Gag-Pol (454, 474). The mechanism by which PR is activated is currently unknown. However once active, PR cleaves the Gag and Gag-Pol polyproteins resulting in the loss of the immature lattice structure and prompting a reorganization of the internal core structure in a process termed maturation. During maturation, the structures of the trimeric Env complexes are unchanged, MA remains associated with the viral membrane and NC remains bound to the RNA. The most noticeable changes during maturation occur at the core of the virus (Fig. 1.7). With the assistance of the newly mature NC protein, the RNA undergoes a conformational switch to become a more stabilized dimer and a critical structural element to the mature virus (138, 153-155, 313, 324). In RSV, immediately following budding CA plus a C-terminal spacer peptide (CA-SP) is excised out of Gag but then over the next 3 – 6 hours is differentially cleaved into two forms: CA (ending at Ala 237) and CA-S (ending at Met 240) (34, 355). At some point during maturation, the CA protein self-assembles to generate a shell around the genomic RNA, NC, and the RT / integrase replicative enzymes. The capsid shell along with the other viral proteins and genomic RNA is referred to as the viral core. Once proper maturation is complete a virus particle become less stiff and less resistant to detergent treatment and is capable of infecting a new cell (42, 43, 240, 241, 440, 446, 494).

CA self-assembly is thought to drive capsid assembly in situ. The concentration of CA in a mature virion is 3-5 mM (52, 54, 293, 475), however only 50-80% of CA participates
in capsid assembly (33, 54, 58). Mature retroviruses contain a capsid shell with irregular/complex morphology. RSV capsids are best described as irregular polyhedra (58, 234). A tendency to form cone-shaped capsids is observed for HIV and other lentiviruses (33, 50). However, other retroviral capsids, including RSV, are highly polymorphic (no two capsids are identical) and other structures can be found including tubular, angular, and roundish shapes (33, 50, 54, 58, 474, 518).

Although the capsids of retroviruses have variable morphologies the three-dimensional structure of their mature protein subunit, CA, is conserved. CA is divided into two distinctly folded domains: the N-terminal NTD and the C-terminal CTD that are connected by a flexible linker of ~ 5-10 residues (Fig. 1.8) (15, 37, 67, 159, 160, 171, 210, 228, 234, 310, 317-319, 504). Both domains of CA are highly α-helical. The NTD consists of an N-terminal β-hairpin followed by multiple α-helices, while the CTD has 4 α-helices. The β-hairpin forms only after PR cleavage during maturation (476). The newly liberated N-terminal Pro residue of CA forms a salt bridge with a conserved Asp residue located at the N-terminus of helix 3 of the NTD to stabilize the β-hairpin structure. Although the β-hairpin only forms in mature CA, this structural element of CA is not directly involved in any of the subunit interfaces formed in the capsid. However, the β-hairpin does appear necessary for capsid assembly (47, 171, 476). Despite the conserved secondary and tertiary structures, the primary structure of retroviral CA is diverse except in the N-terminal region of the CTD in an area termed the major homology region (MHR) (96).

**G. Functional and Structural roles of CA during Retroviral Replication**

The functional role of CA in retroviruses goes beyond the formation of a capsid shell around the genome and replicative enzymes. The residues of CA participate in multiple steps during the replication of retroviruses including assembly of both the immature and mature
Fig 1.8. Structure of mature retroviral CA protein. The ribbon diagram of the monomeric structure of mature RSV CA is depicted. The NTD and CTD are connected by a flexible linker represented by a dashed line. The conserved major homology region (MHR) is color orange. The figure was generated using PDB codes: 1em9 (NTD) and 1d1d (CTD).
particle. The mature capsid is required for early events of replication following attachment and entry including reverse transcription. Many of these roles have been identified by genetic analysis or by studies examining antiviral agents.

**Genetic Analysis of CA reveals a role of the Capsid during Early Events of Replication**

Mutagenesis in multiple retroviruses shows CA functioning in both assembly and replication steps. Interpreting phenotypes of CA mutant particles is complicated by its involvement in multiple steps of replication (i.e. both immature and mature assembly and early steps of infection). CA mutants of retroviruses and Ty3 (reverse transcribing yeast retrotransposon) particles can be classified into one of three general categories: 1) WT-like infectivity, 2) non-infectious, blocked particle release, and 3) non-infectious, efficient particle production (10, 16, 17, 61, 96, 116, 117, 122, 141, 145, 160, 202, 261, 283, 290, 341, 371, 381, 397, 410, 431, 442, 461, 477, 482). Non-infectious particles may either be lethal or poorly infectious with undetectable replication. Disruption of immature particle formation (i.e. budding defective mutants) is likely due to disruption of critical Gag-Gag multimerization interfaces or mislocalization of intracellular Gag. Many non-infectious mutants are efficiently released but defective during early replicative stages. Most fail to produce viral DNA despite containing active RT and a functional viral RNA template. These particular mutants reveal the essential nature of the capsid to reverse transcription *in vivo* (upon entering a cell). Since reverse transcription also failed in CA mutants when stimulated in isolated particles (*in situ*) the capsid is thought to be essential to reverse transcription regardless of any other function(s) during the early events of replication.

The mechanistic role of individual CA molecules or of the maturely formed capsid in the formation or activity of the reverse transcription complex is undefined. A strong correlation between loss of infectivity and loss of a capsid structure was found in the reverse
transcription-defective HIV CA mutants (164, 477). Unfortunately, poor visualization of both wild-type (WT) and mutant RSV capsids by thin-section electron microscopy (EM) has limited a similar evaluation. So at best, studies examine the affects of CA mutations on the morphology of the densely-staining RSV viral core and on the particle stability under detergent treatment (48, 61, 96, 164, 477).

Further limiting a through investigation of the RSV CA mutants is the lack of an in vitro system for capsid assembly. Such a system would allow for the determination of the effect of the mutant on capsid assembly irrespective of a function in immature assembly. However, HIV CA is the only retroviral protein shown to assemble efficiently into structures in vitro; although, most of these structures fail to form all the subunit interactions found in authentic capsids (117, 125, 127, 164, 179, 259, 276) demonstrating the need for a more adequate way to in vitro assemble CA from multiple different retroviruses.

Overall, the genetic studies in both RSV and HIV clearly show that CA is critical to the release of immature particles, assembly of mature capsids, and early events of replication including reverse transcription. However the genetic studies oftentimes fail to explain how the altered residue disrupts these or any other biological process. Comprehensive models for the structures of immature and mature particles are needed to explain the phenotypes of some of the CA mutants and their role in particle assembly.

Role of CA in Disassembly and in Recognition by Host Restriction Factors

After entry into the cell and prior to DNA integration the capsid shell is thought to disassemble (112, 136, 137, 145, 165, 445). Disassembly appears to be an essential step to the retroviral replication, but little is known about the molecular events occurring post-entry. Shortly after receptor-mediated entry the capsid shell is at least partially lost. Some experiments in HIV failed to identify CA in the complexes undergoing reverse transcription
(137, 175). However, other studies, particularly in murine leukemia virus (MLV), detected at least a small amount of CA protein associated with active reverse transcribing complexes suggesting some but not all CA is completely dissociated (57, 112, 135, 136, 222, 223, 302, 308, 530).

CA dissociation from the core may be a spontaneous process driven by dilution of the CA protein upon virus-cell fusion. Such a spontaneous process may be due to weak interactions at some of the CA-CA interfaces or charge repulsion by CA subunits (104, 117, 145, 480). Conversely, capsid disassembly in vivo may be an active process involving cellular factors (18, 184, 328, 334, 480). Regardless of the disassembly mechanism, infectivity of a virion is dependent upon a capsid with optimal stability (112, 145). Either hyper-stable or hypo-stable CA mutants are defective during the initial stages of infectivity further suggesting a role of the capsid shell during early events of replication.

Capsid stability is also important for recognition by certain host restriction factors (144, 416). Restriction factors are antiviral proteins, microRNAs, and enzymes (38, 94, 190, 203, 224, 246, 268, 281, 289, 334, 350, 358, 405, 419, 439, 444, 456, 515, 519). The target of multiple anti-retroviral restriction factors is the capsid region (39, 192, 218, 413). Notably, the Fv1 restriction factor of MLV is dependent upon only a single residue at position 110 of CA (107, 247, 317, 344, 357). Furthermore, the mammalian TRIM5α restriction factor targets CA (109-111, 209, 357, 445). Following entry into a cell, TRIM5α accelerates capsids disassembly causing a premature loss of CA from the core suggesting capsid disassembly is a regulated process required for infectivity.

Collectively the studies of mutant viruses and mechanisms of host restriction factors indicate viral capsids are essential to early events of replication including proper disassembly and reverse transcription. Any additional functions of CA or the capsid in a cell await further investigation but experiments show CA or the capsid may act as a co-factor for nuclear import (273, 274, 512, 513, 527). The mechanisms by which a capsid functions during these
critical steps of early replication are unknown. However, without a doubt CA functions as critical interacting structural units in both the immature and mature particles. Understanding the mechanism(s) by which CA or capsids participate in the replication of retroviral virions may be derived from a comprehensive understanding of the structural roles of CA.

**Structural Role of CA in Immature Particles**

The only viral protein required for the intracellular assembly and release of an immature virus-like retroviral particle is the Gag polyprotein (454, 489, 499, 500). The domains of the Gag polyprotein (MA, CA and NC) are important for the assembly of an immature particle. Approximately 1500 (for RSV) to 5000 (for HIV) Gag molecules orchestrate assembly and release of an immature spherical particle from infected cells (52, 349, 475). The stoichiometries listed above were based on scanning transmission EM; however other methods, albeit ones with a greater amount of error, typically report fewer Gag molecules (267, 361, 540). These values are based upon assuming a uniform densely packed particle, however more recent studies show regions of the immature provirions where the Gag lattice is disordered, suggesting 1500 to 5000 Gag molecules is an overestimate (51, 505).

Gag-Gag interactions are critical to the proper formation of an immature particle in both RSV and HIV (215, 288, 463, 484, 523). Gag molecules pack together in a hexagonal lattice with ~80 - 90 Å between the center of two neighboring hexamers and an interior diameter (of the hexameric ring) of ~80 Å (52, 55, 496, 505, 518). However the Gag lattice is not a uniformly closed structure but is better described as a patchwork of both disordered and order regions of hexagonal symmetry (51, 71, 505, 518). Since Gag does not form a closed lattice but instead forms an incomplete network of hexamers (51, 505), pentamers are not required to form the spherical morphology. As a domain within Gag, the immature CA region forms multiple intermolecular Gag-Gag interactions (5, 8, 52, 55, 65, 251, 496, 505, 518).
CA-CA interactions in Immature Particles

NTD Models for Immature Hexameric Interface

The regions of hexagonal symmetry observed in the immature particle have been modeled as being formed predominately by the NTD of CA. Three-dimensional reconstruction of immature HIV provirions revealed areas of clear hexagonal order (51, 505). Modeling the known structures of the domains of Gag into the low resolution reconstructed particle placed the NTD of CA in a hexameric interface (51, 505). However at this point in time, this model is best considered a working illustrative example since the structure of full-length Gag is unknown and the resolution of the reconstructed images are too low for unambiguous fitting of the individual domains of Gag.

The model for the NTD-mediated immature hexamers is supported by other higher resolution models derived from purified fragments of Gag (317, 327). In RSV, the 25 residues of the p10 domain immediately upstream of the CA and the NTD residues are critical for immature Gag hexamer organization (Fig. 1.9). Purified protein that contained 25 amino acids of p10 and the NTD residues of CA (stopping at L146) crystallized as a dimer (327). An anti-parallel dimer interface was formed by 17 of the 25 p10 residues and the first 14 residues of CA. Mutagenesis and cross-linking experiments showed that these residues are involved in the formation of immature hexamers rather than Gag dimers (Fig. 1.9b) (360). As predicted by this model, the p10 domain is essential to the formation of normal immature spherical particles (5, 120, 216, 251). Thus, at least for RSV and potentially murine leukemia virus and other Gammaretroviruses (269), part of the Gag hexameric lattice is likely formed by the NTD of immature CA and the residues immediately upstream of CA.

Only the first 14 residues are drastically different in the immature and mature forms of RSV CA. The N-terminal cleavage of CA allows a β-hairpin structure to form (Fig. 1.9a).
Fig. 1.9. Model of RSV immature hexamer. (a) Molecular model of a single p10 / immature CA NTD hexameric subunit of the predicted hexamer interface (PDB 1p7n). Also shown is the mature NTD of CA that lacks the p10 extension (PDB 1em9). Please note that the majority of the immature and mature NTD have similar structures except the immature lacks the β-hairpin. The “merged” structure was made in UCSF Chimera by using the match-align structure comparison tool. (b) A schematic model of the p10-NTD immature hexamer interface. On the bottom one subunit is represented the immature structure shown in part a. The p10 region is shown as a line that interacts with the neighboring subunit to form the hexameric ring. This figure was adapted from a similar one generated by Phillips et al. (360).
The β-hairpin structure is required for mature assembly (317, 476). The structure of the remainder of the NTD (E15 – L146) is similar between the immature and mature models (327, 476).

An additional model for immature NTD hexamerization is provided by the structural model derived from crystallization of MLV protein (317). In this model, the first three NTD helices form a hexameric ring. In support of the MLV model, the immature NTD (lacking the β-hairpin) of CA from RSV, HIV and MLV form hexameric arrays on a lipid monolayer (22, 161, 295-297, 301). The lipid layer mimics Gag assembly at the plasma membrane. Although conclusive results have yet to confirm the exact nature of hexamerization in the immature particle, the consistent evidence using a variety of techniques to examine diverse retroviruses provides overwhelming support for the conclusion that the NTD forms hexamers in the immature lattice.

**MHR (major homology region) Domain-Swap for Gag Dimerization**

A possible explanation for the role of the CTD, particularly the highly conserved MHR residues, in Gag assembly and maturation was suggested by a crystal structure of a dimer interface formed by HIV CTD protein (Fig. 1.10a) (205). Prior to the crystallographic results, Gag dimerization via the CTD of CA was predicted based on the structural homology between Gag and mammalian SCAN domain dimers (204, 236). SCAN domains function as protein-protein interaction motifs (123, 497). Homology modeling of the HIV CTD with the SCAN domain of a cellular protein predicted that deletion of the A177 residue would shift the dimerization equilibrium of the CTD (204, 205, 394). The cellular protein was ZNF174, a human transcription factor, which contains a zinc-finger and is homologous to Gag.

As predicted, the A177 deleted (ΔA177) CTD easily dimerized in solution and crystal structures revealed a “domain-swap” conformation in which the protein molecules exchange
identical structural elements (Fig. 1.10) (70, 204, 205, 394). In HIV, the interface in the
ΔA177-CTD dimer was formed by swapping the first two helices between two molecules.
The conserved MHR residues are found in the domain-swapped interface (Fig. 1.10b). The
interface is largely hydrophobic and involves the conserved F164, F167 and L171 MHR
residues (numbered according to RSV). The polar residues of the MHR (Q158, E159, E162,
R167; RSV numbering) as well as some less conserved residues on the following helix either
participate directly in the domain-swap or an extensive hydrogen-bond network that may
provide stability to the dimer (205).

Gag is homologous with SCAN domains because of the C-terminal zinc fingers in
NC, thus the domain-swap CTD dimer is thought to exist only in the immature virus (51,
204). The CTD domain-swap dimer is inconsistent with in vitro assembled mature CA
structures (163), further supporting the CTD domain-swap as a dimer interface in only the
immature particle. However, cryo-electron tomographic and cryo-EM analyses of immature
particles have been of insufficient resolution to confirm the domain-swap model (51, 52, 55,
496, 505, 518).

Spacer Peptide Oligomerization in Immature Particles

In the Gag polyprotein, a spacer peptide (SP; also termed SP1 in HIV) of 12 – 15
residues separates the CA and NC domains (128, 193, 194, 243, 249, 305, 355, 465). SP is
important to proper immature particle formation and is a determinant of the spherical-nature
of assembled particles (51, 95, 180, 182, 225, 248, 505). In both RSV and HIV, the SP
residues are highly flexible with a small propensity for adopting an α-helical structure in
solution (314, 330). Characterization of mutant particles suggests SP along with the C-
terminal part of CA form an immature assembly domain in both RSV and HIV (95, 248, 278,
355). As such, SP was suggested to be a determinant of particle stability and morphology
whereas those of lentiviruses are typically conical, and some of the betaretroviruses are cylindrical or bar-shaped (33, 50, 53, 58, 235, 474, 490, 518). Three independent studies (two examining HIV and one focused on RSV), each examining the native structures of ~ 100 mature capsids, found that no two capsids are perfectly congruent or superimposably alike (33, 50, 58). In a single viral stock, particles with well-formed closed capsids are observed along with other irregular structures such as tubes, nested capsids (a capsid within another capsid), or empty capsid. Butan et al. classified the polymorphic capsids of RSV into four categories: angular (irregular pentagons or hexagons with sharply define angles, including “coffin” shaped cores), continuous curvature (cone and lozenge geometries), tubular (with both open and closed caps), and “other” cores (58). The “other” category includes particles with multi-shelled capsids, multiple capsids per virion, nested capsids or those lacking a discernable core.

Given this large diversity of structures, does a virion need a capsid to be infectious? If so, what does a capsid of an infectious virion look like? To answer the first question: a properly formed capsid is likely required for infectivity, as several studies have strongly correlated CA mutations that disrupted the capsid morphology with a loss of infectivity (145, 164, 461, 477). Furthermore, CA mutations which do not greatly alter infectivity also do not grossly alter capsid morphology (145, 164, 461, 477). The second question is more difficult to answer. The large heterogeneity of structures formed in situ suggests a range of structures are functional. However, the putative infectious HIV particles are thought to have a cone-shaped capsid since this is the predominant morphotype observed (33, 50). RSV particles have angular or continuous curvature capsids (58). These capsids appear closed and to have a dense core (filled with RNA) (58).

In addition to the capsid, the specific infectivity level of a particle is determined by a wide range of factors such as functioning glycoprotein spikes or host factors. The infectivity of virus particles is widely thought to be very low; however an amyloid peptide of semen-
Fig. 1.10. Structures of the CA CTD dimers of HIV. (a) Schematic drawing that shows the design of a domain-swap dimerization interaction between two three-helix proteins. (b) Ribbon diagram representing the structure of the CTD domain-swap dimer that is predicted to exist in the immature particle. (c) The mature CTD-CTD dimer interface. In each case the MHR is colored orange and each monomer is either blue or gray. The figure were made using PDB 2ont (b) and 1a43 (c) (205, 504).
Fig. 1.11. Schematic model for the role of SP (spacer peptide) oligomerization of Gag. According to the Wright et al. model (505) SP participates in the Gag hexamerization in the immature particle by forming a 6 helix bundle beneath the hexameric ring formed by the NTD of CA. The relative size of the NTD, CTD and SP are not drawn to scale.
derived prostatic acidic phosphatase greatly enhances HIV infectivity (230, 323, 386). This peptide, SEVI (semen-derived enhancer of virus infection), functions via its polycationic characteristic and likely alters the attachment and entry steps (386). Thus a properly formed functioning capsid is only one of many variables determining infectivity.

**Fullerene Model for the Morphology of Mature Capsids**

Mature retroviral capsids lack uniform symmetry, preventing the development of a high resolution model of the authentic capsid structures. However, using data from *in vitro* assembled purified HIV protein, a model to describe the overall morphology of the mature capsid lattice was proposed by Ganser *et al.* (164). Similar to the models of icosahedral capsids, this model was also based on the architectural work of Buckminster Fuller after *in vitro* assembled HIV CA-NC protein formed cones were found to have similar fullerene geometry to those of elemental carbon (147, 162, 164, 210, 252). Fullerenes are closed carbon-cage structures consisting of only pentagonal and hexagonal rings (147).

According to the Ganser *et al.* model, retroviral capsids are formed by 150 – 300 hexamers arrayed on a lattice that is closed by the insertion of 12 pentamers (Fig. 1.12). The total number of hexamers in the lattice determines the size of the capsid (33, 50, 58, 164, 198). The overall capsid shape is determined by the distribution of pentamers among the hexamers and the polymorphism observed in authentic capsids can be modeled by uneven placement of the pentamers. An asymmetric but bipolar distribution of pentamers (5 at the narrow end and 7 at the wide end) models the cone-shaped capsids observed in HIV virions. A more even but random distribution of the pentamers explains the irregular polyhedral capsids of RSV (164, 198).
Fig 1.12. Fullerene model of the lattice of capsids in the mature virion. A linked chain diagram representing the placement of pentamers (blue or red) in the hexameric lattice for several capsid morphologies including the irregular polyhedra and fullerene cones observed in RSV and HIV virions. Images reproduced from (198) and (164). The fully-closed polymorphic polyhedrons are formed by 150-300 hexamers and 12 pentamers or approximately 800-1800 CA subunits.
The fullerene model of retroviral capsid morphologies is supported by *in situ* assembled cones (33, 54, 276). The fullerene model predicts a 5 + 7 cone will have a ~19° cone angle if the 5 pentamers at the narrow end are symmetrically distributed (162). Consistent with this prediction the observed cone angle of authentic HIV capsids ranged from ~17.5° to 24° (33, 50, 54, 242). Also in support of the fullerene model, *in vitro* assembled CA protein from HIV, RSV and MLV formed hexameric interactions under various conditions (161, 234, 276, 317, 319).

Some of the experimental evidence of the hexameric lattice was derived from EM analysis of HIV CA assembled *in vitro* into tubes (276). The tubes formed a lattice composed solely of linked hexameric rings. Positioning the higher resolution structure of each domain of mature CA into the tubular hexameric lattice provided one of the first molecular models of the capsid: the exterior surface is composed of NTD hexameric rings connected together by CTD-CTD interactions. The CTD is located towards the interior of the viral core. As noted above, the capsid is closed by the insertion of 12 pentameric points to induce curvature in the hexameric lattice (162, 210, 276). However, the lack of experimental data for CA forming pentamers results in a significant gap in supporting evidence for the fullerene model.

*In vitro* assembly of mutant (R18L) HIV CA resulted in the formation of spherical structures providing further support for the initial molecular model derived from the *in vitro* assembled tubes (163). The spheres collapsed upon application to an EM grid and behaved as a well-ordered crystal providing a higher resolution image than the tubular assembled CA protein. EM crystallographic analysis confirmed the NTD hexameric and the CTD dimeric interfaces. A third interface, between the NTD-CTD was also visualized.
CA-CA interactions in the Mature Retroviral Capsids

NTD-NTD Hexamers of Capsids

Similar to the hexamers of the immature lattice, the hexameric rings of the mature lattice are formed by NTD-NTD interactions (163, 276, 319). Mature NTD hexamers have been observed in HIV CA protein assembled in vitro and in MLV NTD protein crystals. The mature hexamers are slightly larger than the immature hexamers (~90 Å diameter compared to ~80 Å) suggesting a different packing arrangement of the 6 NTDs. The 90 Å diameter closely matching the packing in mature particles of HIV and RSV (54, 58).

The mature hexameric interface is mediated by the first three helices of the NTD (Fig. 1.13). The hexameric ring is formed by helices 1 and 2 of a monomer interacting with helix 3 of a neighboring NTD towards the center of the ring. This model is consistent with alteration of infectivity when certain residues located in helix 1-3 are mutated (16, 17, 48, 145, 457, 461, 477, 495). In support of this hexameric model, the use of hydrogen:deuterium exchange implicated NTD helices 1-3 formed an interface upon in vitro assembly of CA protein and in mature virus-like particles (256-258).

The above HIV CA hexamer model is based on fitting of NMR solution structures of the NTD into the EM density map of in vitro assembled R18L CA mutant protein (163). Slight deviation from the EM density occurs at the β-hairpin and helices 4, 5 and 7 (163). These regions of poor fit may reflect an assembly induced change in CA or be due to regions of low resolution in the EM map (163). Overall, the above results from HIV and MLV provide strong evidence for NTD helices 1-3 forming the hexameric lattice in mature virion and support previous models of NTD hexamerization (210, 276).
Fig 1.13. The mature CA NTD hexamer. The hexameric ring found in the mature capsid is formed by interactions between the first three helices of the NTD of CA. For clarity only the CTD of the red subunit is shown. The figure was made using the pseudo-atomic resolution model for HIV CA described by Ganser-Pornillos et al. (163).
**CTD-CTD Dimeric Interface**

According to the molecular model of mature capsids, the hexamers are connected to each other, and presumably pentamers, via linkages mediated by CTD dimerization (276). The CTD of HIV readily dimerizes in solution and under crystallization conditions (37, 160, 292, 310, 388, 462, 504). There are five different dimer structures based on crystallography (37, 160, 205, 462, 504). In three of these the dimer interface is formed by parallel packing of the second helix, termed the dimerization helix (Fig. 1.10b). Of the five, only the structure reported by Worthylake et al. is a reasonable fit to the EM density of in vitro assembled CA (163, 504).

In the Worthylake et al. crystal structure, hydrophobic interactions, salt bridges, and intermolecular hydrogen bonds help stabilize the dimer. The hydrophobic dimer interface includes the burial of two residues located on the second helix: M185 and W184 (504). Consistent with the crystal structure, mutating W184 and M185 to Ala residues destroyed dimerization and caused a loss of infectivity (160, 256).

The dimerization affinity of the CTD is similar to that of the full-length HIV CA protein (160, 521). This observation has led to the assumption that dimerization of full-length CA is mediated by a CTD-CTD interaction but this has yet to be conclusively demonstrated. Unlike HIV, the CTD of other retroviruses does not readily dimerize in vitro and fail to support the results obtained by HIV (228, 234, 318).

**NTD-CTD Interdomain Interface**

An NTD-CTD interdomain interaction was originally predicted by a genetic study of RSV that showed NTD mutations restore a replication defect caused by CTD mutations (48). The interface was biochemically and structurally mapped in HIV (163, 256, 257). Chemical
cross-linking showed that the NTD-CTD interface forms intermolecularly (between two CA subunits) as opposed to intramolecularly (256). Initial mapping suggested the interacting region involves helix 4 in the NTD and the loop between the first two helices in the CTD (256, 257).

Analysis of hexamers from 2D crystals of assembled HIV R18L CA mutant spheres confirmed the initial mapping results (163). The intermolecular interdomain interaction occurs by placing below each subunit of the NTD hexamer the CTD of the neighboring subunit (Fig. 1.13). The N-terminal region of helix 4 of the NTD inserts into a groove in the CTD (163). One side of the groove is formed by CTD helices 1 and 2 while the other side is formed by the connecting loop between helices 3 and 4 of the CTD (Fig 1.14). NTD helix 7 and the loop connecting helix 3 to helix 4 are in close proximity to CTD helices 1 and 4 providing additional but somewhat minor interdomain interactions between two subunits.

The NTD-CTD interdomain interaction is not present in the immature virus and forms only during maturation as determined by comparing hydrogen:deuterium exchange in immature and mature particles lacking Env (257). The function of the interdomain interface has yet to be elucidated but it is thought to provide stability to the hexameric lattice (163, 165, 452). Two HIV assembly inhibitors disrupt the NTD-CTD interface, highlighting the importance of this interaction to the replication of the virus (163, 226, 441, 457, 462).

NMR analysis revealed Ala substitutions at positions 184 and 185, mutations originally designed to disrupt CTD dimerization, abolishes the ability of the second helix to form properly (6, 7, 503). The region showing the greatest increase in dynamics in the W184A/M185A mutant protein is involved in the CTD-CTD and NTD-CTD interfaces (6, 7, 160, 163, 276, 503, 503, 526), suggesting both CTD-CTD and NTD-CTD interfaces are disrupted. Thus, the observed loss of infectivity caused by the mutations is a result of a failure to form or stabilize these mature interfaces. The residues in CA are highly cooperative (i.e. one residue may influence many other residues), as illustrated by the NMR studies, and
Fig. 1.14. NTD-CTD interface model. The CTD lies underneath the neighboring NTD (as shown in Fig. 1.13 using the same color scheme). Intermolecular interdomain interaction is mediated by NTD helix 4 inserting into a groove created by CTD helix 1 and 2 and by the loop connecting CTD helices 3 and 4. The figure is based on the HIV CA model described by Ganser-Pornillos et al. (163).
function in ways difficult to fully understand without a complete understanding of capsid assembly and an atomic-level model for the mature capsid.

**Pentamers (Has anyone seen a pentamer?)**

According to the fullerene model, originally proposed by Ganser *et al.* and similarly proposed by others, of the retroviral capsid architecture, each closed shell incorporates 12 pentamers among the hexamers in the lattice (162, 164, 198, 210, 276). The above description of the three mature interfaces (NTD-NTD, NTD-CTD, and CTD-CTD) explains how a hexamer forms and interacts with neighboring hexameric rings in a lattice. However, the existence of pentamers and the requirement for them to introduce curvature into the hexameric lattice has only remained a mathematical conjecture based on fullerene geometry (162, 198). A lack of experimental evidence for pentamerization of CA since the Ganser *et al.* proposal has left some, over the past decade, to question the existence of pentamers and the validity of the fullerene model. In this respect, the pentamer can be thought of as the “missing link” in capsid assembly. The ability of CA to form pentameric interactions is examined in chapter V.

**Maturation Switches and the “Hallmarks” of Mature Capsids**

Distinct structural switches have been proposed to participate in the transformation of an immature particle into a mature virion during maturation of RSV and HIV. These maturation switches presumably control alternative conformations of the proteins that determine the immature and mature morphology of particles. The switches have been identified based on structural differences between the immature and mature particles. Maturation switches may also control particle morphology by favoring either an immature
spherical structure or a mature polyhedral/tubular structure. The switches pertinent to this thesis are: the N-terminal β-hairpin structure that forms after CA is cleaved out of Gag, the MHR domain-swap that is predicted to exist only in the immature particle, and the SP peptide following CA that forms a 6-helix bundle and controls spherical immature particle formation. Although these maturation switches have been structurally (1, 178, 180, 204, 205, 314, 330, 476, 504) and genetically (2, 46, 61, 96, 164, 248, 278, 279, 477) studied very little is known about how they function during maturation and capsid assembly. The contribution, if any, of the maturation switches to the assembly of the mature capsid is examined in chapters III-IV.

Additionally, other “hallmarks” differentiate the mature capsid shell from the immature Gag lattice. The NTD-CTD interface only forms in the mature capsid shell and is not present in immature particles (51, 55, 257, 258, 505, 518). Also, the Gag lattice only contains hexameric complexes of protein while the mature shell is predicted to contain both hexamers and pentamers (51, 162, 164, 198, 505). A comprehensive model of maturation that incorporates the role of the structural switches and demonstrates how the “hallmarks” of the mature capsid form is lacking.

H. Modeling in situ Retroviral Capsid Assembly

Although there is no comprehensive model to explain the steps of maturation and capsid assembly following release of the immature particle and protease (PR) activation, some concepts for capsid assembly have been proposed based on the structural differences of CA when it is assembled in immature versus mature particles (51, 178, 180, 317, 476). As such, the maturation switches have been suggested to govern assembly by driving conformational changes within CA during maturation (205, 257, 476, 505, 518). However, a mechanistic understanding of these structural features in inducing mature capsid assembly is lacking.

68
Other concepts for capsid assembly are based on interpretations of mature virion structures. Three independent tomographic studies of mature retroviruses proposed that capsid assembly starts at a specific location in the virion (Fig. 1.15) (33, 50, 58). However, the details of each proposed site vary. Briggs et al. proposed assembly initiates at the narrow end of the HIV conical capsid in proximity to the virion membrane (50), while Benjamin et al. suggested assembly starts at the broad end (33). Butan et al. described the putative nucleation site as being near the Env endodomain and involving uncleaved or partially cleaved Gag (58). Importantly, all three studies agree that assembly involves breaking down the Gag lattice and that assembly of a capsid starts with a nucleating-complex.

**Defining a Nucleation-driven Assembly Mechanism**

Assembly of retroviral capsids by a nucleation-dependent pathway has not been formally tested. However, initiation of assembly by first forming a nucleating-seed has been observed for multiple cellular macromolecular complexes (150, 464, 488, 528). For example, formation of both actin and microtubule filaments are dependent on a rate-limiting nucleation step (26, 108, 132, 150, 196, 479). Assembly of large filaments requires a small number of subunits to initially multimerize as a stable aggregate. Once the stable aggregate, or nucleus, is formed new subunits can rapidly be added to the growing complex. Oligomers smaller than the nucleus can spontaneously form but lack stability and readily dissociate, creating a kinetic barrier to assembly. The kinetic barrier is observed as a lag time during the assembly reaction. If pre-formed nuclei are added to the assembly reaction, the initial lag phase can be bypassed. As the subunits in the reaction are consumed by the growing filaments, the free subunits in the solution fall below a concentration required to nucleate additional growth. The concentration of free subunits left in the solution is called the critical concentration.
Fig. 1.15. Current models of *in situ* assembly. Three studies have suggested that capsid assembly proceeds through a nucleation following protease cleavage of Gag. Briggs *et al.* (50) proposed that HIV capsid assembly is initiated at the narrow end of the capsid, while Benjamin *et al.* (33) suggest that assembly starts at the broad end. Butan *et al.* (58) proposed that RSV capsid assembly is initiated by uncleaved or partially cleaved Gag near an Env glycoprotein spike.
Similar to actin and microtubule filament multimerization, the assembly of some viral capsids also has a rate-limited nucleation step. Assembly of viral capsids can be difficult to understand since proper assembly involves 60 to thousands of subunits multimerizing to form a specific structure, and in some cases requires other factors such as the viral genome. Using a combination of theoretical and experimental work the assembly of multiple viral capsids has been demonstrated to occur in multiple distinct steps and is dependent upon the formation of a nucleating-seed (60, 73, 130, 195, 264, 282, 338, 370, 379, 408, 411, 467, 468, 528, 545, 547).

In the bromoviral case study (section D), assembly is nucleated by a dimer of pentamers (195, 545). Based on the structure of the bromoviral capsids, a hexamer of dimers was originally proposed as a nucleating-complex (213, 429). Subsequent kinetic and modeling analyses clearly showed a stable pentamer of dimers seeds capsid assembly (195, 545). This example illustrates that conclusions about the kinetic process of capsid assembly derived solely on interpretations of static structures is potentially erroneous. Likewise, conclusions about the assembly pathway and nucleation of retroviral capsids based only on tomographic studies (33, 50, 58) may be misleading.

Careful defining of a nucleation-driven assembly pathway is an arduous task. To determine if capsid assembly is dependent upon an initiating complex, the polymerization process must be distinguished from a linear polymerization pathway. In linear (isodesmic) assembly the association rate for all subunits is equal, whereas, a nucleated assembly process involves a rate-limiting step (151). These two mechanisms are differentiated by three criteria: 1) observation of a lag period, 2) requirement of a critical protein concentration for assembly, and 3) pre-formed nuclei act as seeds for assembly (151). The previous description of microtubule and actin polymerization provides a good example of an assembly system that has met each criterion.
Development of an Adequate in vitro Model for in situ Capsid Assembly

Since methods to examine assembly of capsids in a maturing virion are lacking, development of an in vitro assembly system to recapitulate in situ events of maturation was necessary to dissect capsid assembly. For an in vitro assembly system to be adequate, CA must form pentameric and hexameric interactions and assemble structures with morphologies similar to those of native viral capsids. The system must also induce CA to form the three mature interfaces (NTD-NTD, NTD-CTD, and CTD-CTD) predicted to exist in authentic capsids. Numerous studies have attempted to achieve similar goals but were met with failure (16, 17, 24, 106, 117, 163, 178-180, 234, 256, 258, 259, 276, 318). Most of these experiments utilized HIV CA, which easily multimerizes in vitro but fails to efficiently form structures resembling authentic cone-shaped capsids (117, 125, 179, 256, 259, 276). One previous study, however, showed purified protein that contained RSV CA residues formed both small isometric spheres and helical tubes (297). RSV CA may have a greater likelihood than HIV CA of assembling both pentamers and hexamers since presumably the RSV spheres contain pentamers similar to those in icosahedral capsids (76) and the tubes contain hexamers comparable to those in HIV tubes (276). During preliminary studies, the fortuitous discovery that multivalent anions induce RSV CA to form capsid-like structures allowed for the collection of the data presented in chapters II-V. Overall, the results demonstrate RSV CA is a better model system than HIV CA for in vitro assembly of retroviral capsids.

The RSV CA assembly system was especially amenable to examining the initiation of capsid assembly including allowing the evaluation of the nucleation-dependency of capsid assembly. The three nucleation-dependent criteria discussed were examined. The first two criteria were previously shown by in vitro assembly of HIV CA (24, 117, 259) but attempts to examine the third criterion, the addition of pre-formed seeds to bypass the lag phase, remains to be shown. Attempts to isolate intermediates of the assembly pathway of HIV CA have
failed (256, 258, 259). By utilizing conditions inducing the multimerization of purified RSV CA, experiments in chapter II test all three criteria for the nucleation-dependency of in vitro capsid assembly, including isolating and characterizing assembly intermediates that act as assembly-seeds.

Chapter II focuses on in vitro assembly of capsids from purified mature CA protein (i.e. completely folded), although the in situ assembly of capsids is likely dependent on maturation switches and Gag cleavage products. As CA is liberated from the immature Gag lattice, the maturing protein undergoes conformational changes to adopt the structure of mature CA. The in vitro assembly system developed was used to investigate some of these proposed conformational changes. The influence of the conserved MHR (major homology region) residues on capsid assembly was examined since the MHR is hypothesized to act as a switch via a domain-swap arrangement during maturation (Chapter III). Similarly the influence of a maturation switch at the C-termini of CA, the SP (spacer peptide), on the assembly of capsids was also analyzed (Chapter IV). The results of chapters II-IV define factors allowing RSV CA subunits to self-assemble in vitro in a way not previously explored. However, a full understanding of retroviral capsid assembly, structure, and function will require further advances including characterizing the CA pentamer and its interaction with hexamers; an achievement also made possible using in vitro assembled RSV CA (Chapter V).

The following chapters contain results that provide critical information necessary to build a comprehensive model for the initiation of retroviral capsid assembly and to build models for the structures of the native retroviral capsids. The data provide a foundation for understanding some aspects of retroviral maturation and the polymorphic nature of capsids. The results answer critical unmet questions and may ultimately support understanding the molecular architecture of capsids and the functions of the capsid upon entering an uninfected cell. Additionally, the work in the following chapters provide novel analyses applicable to the
broader field of viral capsid assembly, including understanding the assembly pathway and structure of a capsid that is both polyhedral and polymorphic.
CHAPTER II

RETROVIRAL CAPSID ASSEMBLY:

A ROLE FOR THE CA DIMER IN INITIATION

Modified from:

John G. Purdy, John M. Flanagan, Ira J. Ropson, and Rebecca C. Craven

ABSTRACT

In maturing retroviral virions CA protein assembles to form a capsid shell that is essential for infectivity. The structure of two folded domains (NTD and CTD) of CA is highly conserved among various retroviruses, and the capsid assembly pathway, although poorly understood, is thought to be conserved as well. In vitro assembly reactions with purified CA proteins of the Rous sarcoma virus (RSV) were used to define factors that influence the kinetics of capsid assembly and analyze the underlying mechanisms of assembly. CA multimerization was triggered by multivalent anions providing evidence that in vitro assembly is an electrostatically-controlled process. In vitro assembly was a well behaved nucleation-driven process that led to formation of structures with morphologies similar to those found in virions. Isolated CA dimers when mixed with monomeric protein acted as efficient seeds for assembly, eliminating a lag phase that is characteristic of the monomer-only reaction. This demonstrates for the first time the purification of an intermediate on the assembly pathway. Differences in the intrinsic tryptophan fluorescence of monomeric protein and the assembly-competent dimer suggest the involvement of the NTD in formation of the dimer. Overall the data provide clear support for a model whereby capsid assembly within the maturing virion is dependent upon formation of a specific nucleating complex that involves a CA dimer.
INTRODUCTION

A functional retroviral core structure forms during a complex multi-staged process in which assembly and egress of an immature particle, coordinated by the Gag polyprotein, occurs first. The action of the virally-encoded protease results in the reorganization of the internal virion constituents into a mature core structure that consists of RNA bound by viral proteins within a capsid shell (454). Characterization of mature retroviral capsids by cryo-electron tomography has led to the prediction that capsid formation is initiated by the establishment of a unique, small complex of CA (capsid protein), perhaps in contact with other viral constituent(s) (33, 50, 58). Further growth of the shell proceeds by the addition of monomers or small oligomers of CA that are liberated during proteolysis of Gag. In vitro CA assembly studies utilizing the proteins of the human immunodeficiency virus (HIV, a lentivirus) provides ancillary support for a nucleation-dependent initiation of capsid assembly (24, 117). However, no identified structures that represent intermediates, including any nucleating complex, along the capsid assembly pathway have been isolated.

Capsids of both HIV and RSV exhibit a dramatic degree of polymorphism, i.e., particle-to-particle variation in morphology (33, 50, 58). This inherent variability prevents the determination of a high resolution structure of authentic retroviral capsids by any conventional structural technique that involves averaging images of many presumably identical structures. Instead, the analysis of more regular, in vitro-assembled CA arrays has yielded a consensus model of protein organization in the core (163, 276). The capsids are best described as protein lattices organized on the same geometric principles as fullerene carbon structures, with subunits arranged in hexamers interspersed with pentamers (162, 164). A lack of regular spacing between pentamers results in a range of capsid morphologies and distinguishes retroviral cores from the icosahedral capsids found in numerous other viruses (162, 198, 332). The shape of an individual retroviral capsid is governed by the arrangement
of the 12 pentamers that punctuate the hexameric lattice (162, 198, 276). An asymmetric but bipolar distribution (5 + 7) of pentamers can explain the cone-shaped cores that predominate in HIV. A more random distribution is consistent the irregular polyhedra of RSV.

Among retroviral CA proteins, the structure of each of the two α-helical domains (the N-terminal NTD and C-terminal CTD), as well as the intersubunit contacts that form the protein shell, are highly conserved (6, 37, 160, 163, 171, 210, 228, 234, 256, 276, 310, 318, 319, 462, 503). The first three NTD helices establish the intrahexameric interfaces (163, 276, 319). Neighboring capsomers are joined via dimeric contacts between CTDs mediated by the second helix of the interacting domains (163, 276, 504). An NTD-CTD interdomain contact was originally inferred by a genetic study of RSV (48) and has since been supported by biochemical, biophysical and structural studies with HIV (163, 256, 257). The requirement of the NTD-CTD interface for infectivity is confirmed by two HIV inhibitors that bind to CA in the region involved in the interdomain contact (226, 441, 457, 462).

The assembly pathway that produces the CA-CA interactions in the maturing virion has not been defined. Previous kinetic examination of in vitro assembling HIV CA resulted in sigmoidal curve consistent with a rate-limited, nucleation-driven process (24, 259). Typically such nucleation-driven events are distinguished mechanistically from linear (isodesmic) polymerization, in which the association rate for all subunits is equal, by three criteria – the existence of a lag period preceding detectable complex formation, the requirement for a critical protein concentration, and the involvement of a nucleating oligomer that seeds further assembly (151). In the case of in vitro assembly of HIV CA, the lag period and critical concentration were described (24, 259). These experiments failed to show direct evidence of a nucleating complex and the structures generated in vitro are related to, but not representing, mature capsids suggest that multimerization of HIV CA may be a poor model for in situ assembly of retroviruses (179, 256, 276).

As an alternative to HIV CA assembly, the initiation of RSV CA assembly in vitro is
presented below. Under experimental conditions in which protein assembly is triggered by multivalent anions such as phosphate, the three criteria of nucleation-driven polymerization are met. Small CA oligomers ranging in size from dimer to tetramer were purified and shown to stimulate the assembly of monomeric CA, documenting that an oligomer as small as a dimer participates in nucleation. This work provides the strongest evidence to date that capsid assembly in situ is likewise dependent upon formation of a unique nucleating species and describes in vitro conditions that may be used to define intermediate steps of the assembly pathway.

MATERIALS AND METHODS

Protein expression and purification

RSV CA protein was expressed using a pET-24(+) plasmid in which RSV CA sequence from Prague C sequence spanning from Pro1 to Ala237 was inserted using the NdeI and HindIII sites. The natural NdeI site existing in the NTD of CA was removed without altering the primary sequence of the CA protein. A vector-derived initiating Met residue was placed before Pro1, however removal of the Met residue from the expressed protein was confirmed by mass spectrometry. The CA protein was expressed in Escherichia coli (E. coli) BL21 DE3 cells using the auto-induction system developed by Studier (448). The bacteria were grown in 250 mL cultures in 2 L fernback flask at 37º C. Post-saturation of E. coli growth the bacteria were pelleted at 4000 x g for 10 min. The CA protein was purified similarly to a previously published protocol (244). The bacteria pellet was resuspended in buffer containing in 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10% glycerol, 1 mM EDTA, 10 mM DTT, complete protease inhibitor cocktail (Roche) and lysozyme (Sigma). The suspension was sonicated on ice, treated with Benzonase (Novagen), and clarified by centrifugation at 21,000 x g for 30 min at 4º C. Soluble material was precipitated with 35%
ammonium sulfate and centrifugation at 21,000 x g for 30 min at 4° C. The pellet was resuspended in 10 mL of 20 mM Tris-HCl pH 7.5, 100 mM NaCl and 1 mM EDTA and dialyzed against 4 L of the same buffer. Insoluble material was removed by centrifugation and the soluble fraction was loaded onto a column packed with diethylaminoethyl beads (DEAE, EMD Bioscience) equilibrated with buffer A (50 mM Tris-HCl pH 7.5, 1 mM DTT and 0.1 mM EDTA). CA does not bind DEAE under these conditions. The column was washed with buffer A and the flow through was collected and concentrated by precipitation with 50% ammonium sulfate and centrifugation at 21,000 x g for 10 min at 4° C. The protein pellet was resuspended in ~1.8 mL and dialyzed against 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.1 mM ETDA. CA was further purified by size exclusion chromatography (Superdex 75) and the fractions containing monomeric CA were pooled and concentrated to approximately 10 mg/ml using iCON Concentrators (Pierce). Protein that eluted at an apparent mass larger than monomer, termed high molecular weight (HMW) were pooled and frozen at -20° C until needed. CA was monitored at each step of purification by SDS-PAGE and staining with Coomassie Blue. The purity and mass of CA was analyzed by SDS-PAGE (using both Coomassie Blue and SyproRuby) and mass spectrometry.

For most experiments protein from three or more independent preparations were used. The CA oligomers were isolated by pooling and concentrating the HMW fraction from ~20 independent protein preparations. HIV CA protein (p24) was similarly purified using the pWISP98-85 plasmid (521). The following extinction coefficient values were used for the RSV CA: 24980 M⁻¹cm⁻¹ for WT, L171V, and 26470 M⁻¹cm⁻¹ for F167Y (167). An extinction coefficient value of 33585 M⁻¹cm⁻¹ was used for HIV CA. The structural image of CA shown in Fig. 2.4c was made using Chimera (359).
In vitro capsid assembly

CA protein was induced to assemble by the addition of specified salt, typically at a concentration of 500 mM. Capsid assembly was monitored by optical density at 450 nm (OD$_{450}$) using a SpectraMax Plus384 spectrophotometer (Molecular Devices). UV-Transparent 96-well microplates (BD Falcon and Corning) were used for 100 µl reactions and UV-Transparent half area 96-well microplates (Corning) were used for 50 µl reactions. In general, one reading per min was performed and to circumvent capsids from settling at the bottom of the well the reactions were mixed for 3 sec prior to each reading. Typically the protein concentrations used for most assembly reactions were 80 µM (2 mg/ml) to 160 µM (4 mg/ml). CA protein was induced to assemble by the addition of salts. For most reactions sodium phosphate pH 8.0 was utilized. When other salts were used the pH was controlled by dissolving the salt in 20 mM Tris-HCl pH 8.0. Determination of the lag time was done using the modified Gompertz equation (548). Unless otherwise noted all assembly reactions were performed at room temperature. For clarity, only one in twenty data points is shown on most of the graphs and only a representative graph is shown from multiple repeated experiments.

Electron microscopy (EM)

Capsids were assembled as described above except the reactions were mixed for 3 sec every 2.5 min. Upon reaching the maximum OD$_{450}$ the products of the in vitro assembly reactions were removed from the spectrophotometer and allowed to settle overnight at 4 ºC prior to being analyzed by EM. The samples were applied to a formvar-coated grid, washed three times with water or 0.1 M KCl, stained with 2% uranyl acetate and visualized using a 120 kV TEM (Phillips) operating at 60 kV. Phosphotungstic acid was used in place of uranyl acetate for samples containing sodium sulfate.
Critical concentration determination

The assembly reactions were carried out as described above. The final sodium phosphate concentration for assembly reactions using RSV CA was 500 mM and 300 mM for experiments containing HIV CA protein. The protein concentration varied from 10 to 160 µM. The reactions were allowed to proceed for an additional 3 to 5 hours after reaching the maximum OD$_{450}$ to ensure completion of assembly. Evaporation during the long incubations was avoided by covering the 96-well plate with sealing tape. The samples were centrifuged in an ultracentrifuge at 66,000 x g for 30 min at 20 °C. The protein concentration remaining in the supernatant was determined by absorbance at 280 nm.

Multi-angle light scattering (MALS)

Determination of protein mass was performed by SEC coupled MALS at room temperature. Oligomeric protein (HMW fractions) from multiple protein preps was pooled and concentrated. The protein was applied to an analytical-grade Superdex 200 HR 10/300 GL column (Amersham Pharmacia) equilibrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA, and 0.1% sodium azide. The scattered intensity of the complexes eluted off the column was recorded at 18 angles using a Dawn Heleos II multiangle laser light scattering detector (Wyatt Technology). The protein concentration in each complex was monitored by changes in the refractive index using an Optilab rEX refractive index detector (Wyatt Technology). The molecular weight of each complex in the various chromatographic peaks was determined using the ASTRA software (Wyatt Technology).

Circular dichroism spectroscopy and intrinsic tryptophan (Trp) fluorescence

Protein was diluted in 50 mM sodium phosphate pH 7.5, 50 mM NaCl, and 0.1 mM EDTA to a final concentration of 0.1 mg/ml. A Jasco J-710 spectropolarimeter operating in
the far-UV region of the spectrum was used for circular dichroism. Intrinsic Trp fluorescence was performed using a 290 nm excitation wavelength on a PTI QuantaMaster luminescence spectrometer.

RESULTS

Effects of salts on CA assembly in vitro

The RSV CA protein was purified from *E. coli* using a previously described protocol (244, 373). Large planar arrays and small spheroidal complexes of CA protein were observed following ammonium sulfate precipitation during purification (data not shown). Reasoning that protein assembly was triggered by ammonium sulfate, a turbidimetric assay was employed to test the influence of salts, pH and protein concentration on CA multimerization. Similar turbidimetric assays have been used to study *in vitro* assembly of viral capsids including those of HIV (108, 259, 347, 370, 417). With protein at 160 µM (4 mg/ml), the development of turbidity could be triggered by the addition of certain salts at 500 mM final concentration (Fig. 2.1a). In particular, sodium phosphate, ammonium phosphate, sodium sulfate, and ammonium sulfate, each of which contained a multivalent anionic group at the pH of 8.0 used in this experiment, caused an increase in optical density at 450 nm (OD$_{450}$) (Fig. 2.1a). The phosphate salts were much more effective than the sulfate salts, in contrast to the recognized ability of these ions to cause salting out of proteins, suggesting that phosphate has a specific influence on CA beyond increasing protein insolubility. In contrast, sodium chloride and ammonium chloride at 500 mM failed to trigger any increase in turbidity (Fig. 2.1a) even after an extended incubation of 6 hours (data not shown).

To determine the nature of the light scattering material formed in the presence of multivalent anions, precipitates were examined by negative staining and EM. Each turbid sample contained organized structures confirming that the increase in OD$_{450}$ was due to light
Fig. 2.1. Multivalent anion-induced multimerization of monomeric RSV CA. (a) CA protein was mixed with various salts at pH 8.0, and the resulting turbidity of the protein solution was followed by measurement of light scattering at 450 nm. The final concentrations were 160 µM protein and 500 mM salt. (b) EM analysis of the products formed in the turbid reactions containing phosphate and sulfate salts. (c) EM analysis of non-turbid samples containing chloride salts. All scale bars represent 100 nm. Additional examples of structures form in sodium phosphate assembly conditions are found in Fig. 3.3.
scattering by large structures of assembled CA. The morphotypes observed included angular capsid-like shells (Fig. 2.1b) similar to the irregular polyhedral capsids of native RSV virions (58, 235), as well as small spheres and tubular structures. The large angular, small spheroidal, and tubular shapes have each been observed with some frequency in authentic RSV particles (58) and are consistent with formation of protein arrays in which the CA subunits are packed according to fullerene geometry (162, 198, 276). No clear bias for one type of structure or another was observed with any of the phosphate or sulfate salts. Consistent with the turbidity results, incubation of protein in ammonium chloride failed to induce assembly of any organized CA structures (Fig. 2.1c); similarly, examination of several grids from each of two independent reactions with sodium chloride identified only rare small spheroidal structures (Fig. 2.1c).

The extremely rapid assembly observed with sodium phosphate (Fig. 2.1a) was slowed when the protein concentration was reduced, and sigmoidal kinetics, with a clear lag followed by a rapid rise in turbidity, became evident (Fig. 2.2a). With 500 mM sodium phosphate to trigger assembly, the lag period increased and both the rate of assembly and the maximum turbidity decreased, as the protein concentration was reduced over a range of 120 μM to 10 μM. Turbidity was undetectable in 300 min at the lowest concentration implying that assembly requires some critical concentration to proceed. To estimate this concentration, samples containing 10-120 μM protein were initiated to assemble by the addition of 500 mM sodium phosphate; 3-5 hrs after the maximal OD was reached the reactions were subjected to ultra-centrifugation. The pool of free CA that was resistant to pelleting was determined by absorbance at A_{280} to be ~ 15 μM (data not shown). This clear dependence of assembly upon protein concentration is consistent with a nucleation-driven process (340, 370).
Synergy of phosphate concentration and total ionic strength

Phosphate-driven assembly reactions were also sensitive to pH. The overall reaction kinetics increased as the pH of the sodium phosphate solution was increased from 6.0 to 8.0 (Fig. 2.2b). This observation is consistent with assembly being dependent upon the presence of multivalent ions since sodium phosphate becomes divalent at the higher pH, further supporting the electrostatic control of in vitro multimerization of CA. The effectiveness of 500 mM sodium phosphate in triggering assembly may reflect the contributions of both the multivalent anion as well as the total ionic strength of the reaction solution.

To dissect the relative importance of these two variables, both the phosphate concentration and the total ionic strength (controlled by addition of sodium chloride) were varied systematically. In a typical experiment where sodium phosphate was held constant at 100 mM, pH 8.0, increasing the sodium chloride level caused a progressive shortening of the lag (Fig. 2.2c). Thus, increasing the total ionic strength causes an increase in the assembly kinetics. The combined effects of the two variables are displayed in Fig. 2.2d where each curve illustrates the effect upon the duration of the lag period by increasing phosphate while keeping the total ionic strength constant. At all ionic strengths tested, increasing the sodium phosphate caused a decrease in the duration of the lag phase. Regardless of the total ionic strength, the lower limit at which the assembly begins (i.e., the calculated lag period begins to decrease) is ~ 25-50 mM sodium phosphate. Without sodium phosphate a detectable turbidity occurred, but only after an extremely prolonged incubation (500-600 min) at very high concentrations of sodium chloride (1.5-2.25 M, Fig. 2.2d).

To assess the formation of organized structures at low (~ 50 mM) levels of both sodium phosphate and ionic strength, monomeric CA was dialyzed against 50 - 100 mM sodium phosphate pH 8.0 at 4 ºC for approximately 4 hours. EM examination showed the
Fig. 2.2. Effects of protein concentration, pH, and ionic strength on CA protein assembly. (a) The turbidimetric profile of RSV CA protein ranging from 10 to 160 µM, induced to assemble by addition of 500 mM sodium phosphate at pH 8.0. (b) Influence of pH of the 500 mM sodium phosphate solution on assembly kinetics. The protein concentration was 120 µM. (c) Influence of ionic strength on assembly kinetics in 100 mM sodium phosphate, pH 8.0. The total ionic strength (I, as shown) was adjusted by the addition of sodium chloride. (d) Dependence of lag time on sodium phosphate concentration at various ionic strengths (I); for each curve, the sodium chloride level was decreased as phosphate increased to maintain constant ionic strength. The protein concentration in part c and d was 80 µM. (e) CA structures formed at low concentration of sodium phosphate. Monomeric CA at 400 µM was dialyzed against 50 mM salt at pH 8 and 4 °C until final protein concentration of 280 µM was reach and then examined by EM. Scale bar represents 100 nm.
a

b

c

d

90
presence of a homogenous population of small spheroidal structures with a diameter of less than 20 nm (Fig. 2.2e). No larger tubular or polyhedral structures similar to those shown in Fig. 1b were observed under these conditions. Thus 50 mM sodium phosphate can trigger oligomerization of monomeric CA but the phosphate anion and total ionic strength are synergistic in supporting the formation of larger CA assemblies, including the capsid-like structures described above.

The hypothesis that 50 mM sodium phosphate drives the formation of small CA oligomers was tested by incubating 80 µM monomeric CA in the presence or absence of sodium phosphate at 50 mM and pH 8.0 for one hour before the sodium phosphate concentration was raised to 500 mM (Fig. 2.3). Reactions that contained either 50 or 100 mM sodium phosphate during the first incubation (phase I) showed immediate assembly without an observable lag in the second phase. In contrast reactions that were not exposed to phosphate during pre-incubation developed turbidity only after a clear lag. When the sodium phosphate level was kept constant at 50 mM during the second phase, no increase in OD$_{450}$ occurred. During both phase I and II the total protein concentration was held constant by adding monomeric CA prior to raising the sodium phosphate concentration to 500 mM. The results shown in Fig. 2.3 demonstrates that low levels of sodium phosphate can trigger the formation of nuclei that act without delay to seed the assembly of the monomer in phase 2.

Characterization of CA oligomers

Two distinct populations of CA protein were observed by size exclusion chromatography (SEC) on a Superdex S75 column following purification by ion-exchange chromatography (Fig. 2.4a). The majority of the protein eluted in a single peak with an apparent mass similar to the 25.5 kD expected of a CA monomer (Fig. 2.4a, labeled LMW). The remaining protein eluted as a cluster of peaks of larger masses (Fig. 4a, labeled HMW).
Fig. 2.3. Low levels of sodium phosphate induce the formation of assembly seeds. Monomeric RSV CA was incubated with 0, 50 or 100 mM sodium phosphate, pH 8.0, for 60 min (Phase I), after which the final sodium phosphate concentration was either raised to 500 mM or kept at 50 mM (Phase II). Additional protein was also added to maintain its concentration at 80 μM.
Mass spectrometry revealed that both the LMW and HMW fractions contained a protein of 25.5 kD (specifically 25522 Daltons for LMW and 25508 Daltons for HMW), confirming that the protein was in fact authentic CA and that the vector-derived initiating methionine was removed. Proteins in the LMW and HMW fractions were indistinguishable by circular dichroism, exhibiting the expected high content of α-helix folding and arguing against the possibility that the HMW fraction consisted of grossly misfolded aggregates (data not shown). Previous studies of the RSV CA protein detected no tendency of the protein to dimerize in vitro (67, 234). The features of the purification protocol used here that allowed oligomerization are not clear but may potentially relate to the use of a ZYP-5052 autoinduction growth medium (448).

SEC coupled with dynamic multi-angle light scattering (MAL'S) was used to independently measure the oligomeric state of the proteins found within the S75 HMW fraction (506). The HMW fractions from multiple independent protein purifications were pooled, concentrated and analyzed by analytical SEC (Superdex 200) with in-line MALS (Fig. 2.4b). The protein eluted in four peaks with apparent masses of 22.5 kD (Peak 1 / monomer), 44.6 kD (Peak 2 / dimer), 66.1 kD (Peak 3 / trimer), and 84.8 kD (Peak 4 / tetramer). The pronounced leading edge of peak 4 suggests the presence of an oligomeric form larger than a tetramer. These numbers confirm the existence of discrete CA oligomers in the HMW material. Furthermore, the appearance of monomeric CA in the SEC-S200/MALS profile indicates that the HMW oligomers have a tendency to dissociate. The difference in the mass of the monomer measured by MALS and by MS is likely due to inaccuracies in the determination of protein concentration in each SEC-S200 peak, a variable that often limits the accuracy of MALS (471, 511) especially at the low protein concentrations.

Intrinsic Trp fluorescence was utilized to probe for conformational differences between the monomer and oligomers. In the monomer, two Trp residues (W69 and W73) are located within a hydrophobic pocket near the NTD-CTD interdomain interface (Fig. 2.4c).
Fig. 2.4. Identification and characterization of CA oligomers. (a) Size exclusion chromatography (SEC) on a preparative grade Superdex 75 column shows two distinct populations of protein, high molecular weight (HMW) and low molecular weight (LMW). (b) Analytical SEC and multi-angled light scattering (MALS) analysis of the HMW fraction. The molar mass of each peak was determined: peak 1 (monomer), peak 2 (dimer), peak 3 (trimer), and peak 4 (tetramer plus larger complexes). (c) The four Trp residues within in CA are shown on the structure of the NTD and linker region of RSV CA (PDB: 1d1d). Note that the protein in this model does not contain the β-hairpin structure at the N-terminus. The residues found in a hydrophilic region are red, and residues found within the hydrophobic environment are blue. (d) Intrinsic Trp fluorescence of the protein in the LMW and HMW fractions.
The other two Trp residues of CA lie in the β-hairpin and in the flexible linker connecting the NTD and CTD (67). Both of these Trp residues (W11 and W153) are located in hydrophilic regions of the monomeric protein. The peak fluorescence signal of monomeric CA occurred at 325 nm (Fig. 2.4d), suggesting that residues within a hydrophobic environment, i.e. the W69 and W73, are responsible for the majority of the signal (473). The HMW oligomers, in comparison to the monomer, exhibited a ~30% decrease in overall intensity with no apparent shift in peak wavelength, indicating a change in the environment of the NTD hydrophobic pocket. Whether due to a change in conformation of each monomer of the oligomer or the effects of intersubunit docking, this observation provides suggestive evidence that the NTD is involved in oligomerization.

Assembly activity of the CA dimer

The hypothesis that the HMW oligomeric proteins represent on-pathway assembly intermediates was tested by adding the HMW oligomers to monomeric protein prior to the addition of sodium phosphate. When 50% (40 µM) of the total protein was added as HMW protein, the assembly proceeded very rapidly; the lag characteristic of monomer assembly disappeared (Fig. 2.5a). At lower ratios of HMW oligomers (10 or 25% of the total protein), the assembly occurred in three phases: 1) a rapid phase with an immediate increase in turbidity that is limited by depletion of the oligomers that seeded the initial burst of activity, 2) a slow growth phase and 3) a final growth phase with kinetics more like that of a monomer-only reaction. These results confirm the presence of assembly-competent oligomers that can seed the formation of large CA structures are present in the HMW fraction.

To determine which HMW oligomer was responsible for nucleating assembly, reactions were seeded with protein from peak 2, 3 or 4. When a quarter of the total protein was contributed by peak 2 CA (dimer), assembly proceeded rapidly with no detectable lag.
Fig. 2.5. Assembly seeding by the addition of the oligomeric CA. (a) Assembly kinetics of reactions containing monomeric CA mixed with protein from the HMW fraction. The total protein concentration was held constant at 80 µM and HMW protein was added at 8, 20, and 40 µM. (b-d) Each individual peak identified in Fig 4b was used to seed capsid assembly. Again, the total protein concentration was held at 80 µM. (e) EM micrographs of products of the peak 2 (dimer) seeded reaction (part b) with scale bars representing 100 nm.
similar to that of the 50% HMW reaction (Fig. 2.5b). Reactions seeded with peak 3 or 4 were multiphasic, similar to those set up with small amounts of the mixed oligomers (Fig. 2.5c-d). The structures that formed by seeding with protein from either peak 2, 3 or 4 showed the same morphotypes observed under standard monomer-only assembly conditions although with a somewhat reduced abundance of tubes (Fig. 2.5e). The results clearly show that the CA oligomers can function to nucleate *in vitro* capsid assembly with the dimer being most efficient. The multiphasic behavior observed with peak 3 and peak 4 protein likely reflects the addition of some active oligomers that immediately initiate assembly; sustained progress, however, requires *de novo* formation of additional oligomers from the monomers, hence the intermediate lag.

**HIV CA assembly in phosphate**

HIV CA protein, unlike RSV CA protein, efficiently assembles under conditions of high sodium chloride concentration (1.0 – 2.5 M) (179, 259, 276, 476) but shows a marked preference for tube formation. Examination of the assembly of monomeric HIV CA showed that protein at 80 µM assembled efficiently in 100 mM sodium phosphate (Fig. 2.6a), and the reaction exhibited sigmoidal kinetics with a discernable lag phase similar to that observed with the RSV protein in slightly higher phosphate. HIV CA showed a critical concentration of ~10 µM under 300 mM sodium phosphate assembly conditions (data not shown) which is slightly larger than the previously published concentration for sodium chloride assembly conditions (117). Unlike RSV CA, however, the HIV protein formed almost exclusively long tubular structures which had an average diameter of ~45 nm (Fig. 2.6b-c) and resembled structures reported by others to form in high concentrations of sodium chloride (164, 259, 276, 476). Some of the tubular polymers appeared to be capped (Fig. 2.6c) with structures that resembled the narrow ends of HIV cones. Finally, during SEC purification of the HIV
protein, a small amount of protein eluted more rapidly than the monomeric CA (Fig. 2.6d). This is expected to represent dimers or higher oligomers, analogous to the RSV HMW fraction. However, these oligomers have not been obtained in sufficient quantity to test their ability to seed a monomer reaction.

**Discussion**

The results in this chapter define the parameters that induce the initiation of *in vitro* assembly of capsid-like structures by CA protein and provide important new insights into the potential mechanisms governing capsid formation in the maturing virion. Assembly of the RSV CA can be triggered by the multivalent anions sulfate and phosphate. The resulting structures are diverse in shape and include capsid-like polyhedral structures as well as tubes, small spheres and multilamellar structures. This range of morphologies suggests that phosphate-mediated assembly conditions support the formation of both pentameric and hexameric associations of CA monomers. The broad range of structures, including angular capsid-like structures, observed in this study may be due to either the near neutral pH or the use of sodium phosphate. In contrast, a previous study showed RSV CA assembling a more restricted range of structures (only small spheroidal and tubular forms) using distinctly non-physiological (pH 5.5 or lower with NaCl at 0.6 M or greater) conditions (234). The concentrations of CA used in this study (160 µM or less) were considerably lower than the concentration of CA in a mature virion (3-5 mM). However, concentrations of CA used for the experiments described in this chapter and the following chapters are consistent with concentrations used in other *in vitro* studies of retroviral CA (24, 64, 117, 125, 127, 162, 164, 178, 179, 234, 258, 259, 276, 477).

Phosphate-driven *in vitro* assembly of the RSV CA protein, followed turbidimetrically, provides a highly reproducible system and allows the quantitative
Fig. 2.6. Assembly of HIV CA. (a) Turbidimetric profile of monomeric HIV CA at 80 µM induced to assemble by 100 and 200 mM sodium phosphate or by 1.0 M NaCl, pH 8.0. (b) Assembled protein (at 200 mM sodium phosphate pH 8.0) examined at low magnification (2750x). The scale bar represents 1000 nm. (c) Higher magnification (35500x) of protein assembled at 250 mM sodium phosphate pH 8.0. The scale bar represents 100 nm. Arrows mark the closed ends of tubular structures. (d) SEC using a S75 preparative column shows the presence of HMW oligomers (inset) and LMW monomeric HIV CA protein.
assessment of reaction kinetics and the requirements for the initiation of assembly. The kinetic data obtained demonstrate that \textit{in vitro} assembly of the RSV protein is a well-behaved nucleation-driven process. In particular, the phosphate-driven reactions exhibit a pronounced lag period which indicates the existence of an essential rate-limiting step, a critical protein concentration below which assembly fails to progress, and the ability of pre-formed oligomers to seed the assembly process effectively overcoming the rate-limited step. In this way retroviral capsid assembly shares some similarities with the assembly processes other viruses, such as tobacco mosaic virus, bromoviruses, bacteriophage P22 and polyoma virus (217, 402, 545).

\textit{Electrostatic control of capsid assembly}

The strong requirement for a multivalent anion such as phosphate to stimulate efficient \textit{in vitro} assembly suggests that shielding of basic residues in CA is needed in order to form assembly-competent oligomers that are presumably associated via some combination of the known NTD-NTD, CTD-CTD and NTD-CTD interfaces. Calculations using standard parameters in GRASP (333) of RSV CA reveals a large net positive charge covering most of the surface of the NTD, whereas the CTD contains only small localized patches of both positive and negative charges (Maria Bewley, personal communication). Therefore, phosphate ions at relatively low concentrations may neutralize charge repulsion at one or more key points on the NTD to trigger initiation of capsid assembly (Fig. 2.7a). A higher ionic strength allows for continued growth establishing the full range of contacts needed to build a structure with a similar morphology as mature authentic capsids. This is consistent with altered \textit{in vitro} assembly rates of HIV mutants within a highly charged region of the NTD which independently led to the proposal that this domain is involved in a nucleation step (117).
The assembly of HIV CA is also strongly dependent upon pH and electrostatic influences (104, 105, 117, 125, 127, 259). Assembly of HIV CA can be triggered by sodium phosphate at relatively low concentrations, but unlike the RSV protein, the resulting structures are almost entirely tubes similar to those that assemble in high concentrations of sodium chloride (179, 259, 276). The marked preference for the formation of tubular arrays of proteins rather than the cone-shapes that predominate in the virions suggest that HIV protein favors the formation of hexamers over pentamers in the in vitro conditions studied. Thus, the sodium phosphate assembly conditions remain an insufficient substitute for the environment that is needed to facilitate the formation of HIV CA pentamers and conical capsids.

Role of Multivalent Anions in Capsid Assembly

The inorganic phosphate salts used to trigger assembly in vitro presumably act as a surrogate for other polyanionic compound(s) in the maturing virion. The candidates inside the particle include viral genomic RNA and cellular RNAs, phospholipids and other envelope constituents, and possibly nucleotides or other small molecules such as inositol phosphate. Sodium phosphate induced in vitro assembly, coupled with EM, described in this chapter provides an excellent tool for initial studies to screen how components of the virion alter mature capsid assembly.

RNA directed or assisted capsid assembly in plant and insect viruses have been well studied (149, 265, 408, 409). Could RNA provide the necessary anionic component to initiate retroviral capsid assembly in situ? The major RNA binding protein in the mature virion is NC and it is commonly thought that most of the RNA is coated by this protein. However, in murine leukemia virus a minor amount of CA is found to be associated with RNA by cross-linking (369). Similarly a small amount of CA can be cross-linked with the genomic RNA in mature RSV virions (data not shown and R. Craven unpublished). Preliminary in vitro capsid
assembly in the presence of RNA (either cellular tRNA or *in vitro* transcribed viral RNA) did not grossly alter the duration of the lag phase (data not shown). However, it is premature to exclude a possible contribution that RNA has on *in situ* capsid formation.

A number of other compounds with multiple phosphate groups, including inositol phosphate derivates (IPs), control *in vitro* assembly of HIV Gag (62, 418) and are found in mature HIV particles (11). Inositol pentaphosphate (IP$_5$) and inositol hexaphosphate (IP$_6$) showed the greatest activity for stimulating Gag assembly (62). IP$_s$ binds MA and induces conformational changes in Gag (102, 398, 399). IP$_6$ may also influence the structure of CA in Gag (418) and may bind the CTD of CA (102). To test the role of IP$_6$ in mature assembly, it was mixed with RSV CA prior to addition of sodium phosphate. The kinetics of *in vitro* assembly was stimulated at levels as low as an equimolar ratio of IP$_6$ to CA (data not shown). However, without sodium phosphate IP$_6$ failed to induce turbidity except at extremely high concentrations (data not shown) that were much higher than the ~ 30 – 65 µM intracellular concentration of IP$_6$ (185, 337, 362). Interestingly, inositol hexasulfate had no effect on *in vitro* Gag assembly (62) even though the results in Fig. 2.1a show that in addition to phosphate, sulfate also induced CA assembly. These preliminary results suggest that IP$_6$ may be one of many factors influencing *in situ* capsid assembly.

In addition to RNA and IP$_6$ others have suggested that components found in the viral membrane influence capsid formation (24, 25, 50, 58). Butan *et al.* suggest that the endodomain of Env located on the interior surface of the membrane influences nucleation based on a correlation between capsid morphology with the number of Env spikes on a particle (i.e. particles that had a greater number of spikes also had properly formed angular capsids and those with fewer spikes had improperly formed capsids) (58). Acidic phospholipids in the membrane have also been proposed to regulate initiation or closure of capsids (24, 50).

The organization of the maturing virion with the membrane on the outside and RNA
inside the nascent capsid shell may provide an ideal three-dimensional charge-stratification that serves to facilitate proper in situ assembly. Tomographic studies of RSV and HIV particles have independently argued for a role of virion constituents in addition to the CA protein itself in the initiation of capsid formation during maturation (33, 50, 58). In particular, the model for HIV cone assembly described by Briggs et al. (50) – specifically, assembly beginning with creation of the narrow end of the cone-shaped capsid, continuing with formation of the hexagonal lattice, and finishing with closure of the shell when the growing cone approaches the distant side of the membrane – is appealing because it predicts a role for the membrane in directing pentamer insertion at the two poles. The prediction that these membrane components control capsid assembly may explain why in their absence the in vitro assembled structures have a wider distribution of morphotypes compared to in situ formed capsids: polyhedral shapes are less common and spheres and rods more abundant in vitro than in virions (58, 373).

Potential role of a CA dimer in nucleating assembly

RSV CA protein is capable of existing as a dimer and its tendency to do so in vitro is favored by the presence of multivalent anions (Fig. 2.3 and 2.4). Furthermore, the dimer immediately stimulates assembly when mixed with monomeric protein in the presence of phosphate ions, arguing that the dimer is involved in the rate-limited assembly step detected by these kinetic experiments. In contrast, the HIV CA protein readily dimerizes in vitro (160, 388), but no assembly-active oligomers have been isolated nor has the structure of the dimer of full-length CA been determined. The hexagonal tubes observed by in vitro assembled HIV CA (276) could be the result of multimerization by a CA dimer protomer instead of monomers thus limiting pentamer formation. Whether in the maturing retroviral virion a dimeric form of CA is released from Gag by proteolysis or whether CA exists transiently as a
monomer that must undergo refolding events in its NTD and/or CTD to stimulate dimerization is unknown at present. Either scenario is consistent with the data presented here that indicate a key role of the dimer in initiating capsid assembly.

The precise intermolecular interfaces through which the assembly-competent dimer forms are undefined. However, observations suggest that formation of the either the NTD-NTD or NTD-CTD interface is a critical step that occurs early on the pathway of capsid assembly. The difference in the intrinsic Trp fluorescence of monomers and oligomers suggests that the neighborhood around W69 and W73 (on helix 4 in the hydrophobic core of the NTD) undergoes a slight change of environment upon formation of the assembly-active dimers. Such a shift would not be expected if the sole point of contact in formation of the dimer is via the CTDs. The influence of the CTD on in vitro assembly of capsids will be examined in the next chapter.

Genetic analyses and inhibitor-binding studies with HIV have documented the essential role of the NTD-CTD contacts to capsid functionality (163, 226, 441, 457, 462, 477). However this study is the first to propose an indispensable role for the NTD-CTD interface during initiation of capsid assembly. This work demonstrates that the capsid assembly pathway may be defined via in vitro approaches. In particular, the determination of high-resolution structures of oligomeric intermediates in capsid assembly and further definition of the roles of other virion constituents in the process will be critical for the development of a complete model for capsid assembly during maturation.

ACKNOWLEDGEMENTS

Our grateful appreciation is extended to Roland Meyers of the Penn State College of Medicine core facilities for assistance with EM analyses, to Eric Barklis for supplying the plasmid for expression of HIV CA protein, and to Maria Bewley for helpful discussions and
GRASP modeling. Funding for this research was provided by the NIH grant CA100322 (R.C.C.), the Pennsylvania Department of Health (R.C.C., Leslie J. Parent, and Michael Katzman) and the Penn State Cancer Institute (R.C.C.).
Fig. 2.7. Electrostatic control of CA assembly. (a) A schematic model of three possible dimeric (CA\textsubscript{dimer}) initiating complexes formed under low sodium phosphate levels. Multivalent anions favor the dimerization of CA monomers via NTD-NTD, NTD-CTD or CTD-CTD interactions by shielding specific positively charged residues. Additional salt is required to overcome electrostatic repulsion and allow further multimerization into large multimeric structures. The salt dependence of assembly and Trp fluorescence of dimeric CA suggest a model for the NTD-NTD or NTD-CTD interaction to initiate CA assembly.
CHAPTER III

CRITICAL ROLE OF CONSERVED HYDROPHOBIC RESIDUES OF THE MAJOR HOMOLOGY REGION IN MATURE RETROVIRAL CAPSID ASSEMBLY

Modified from:

John G. Purdy, John M. Flanagan, Ira J. Ropson, Kristen E. Rennoll-Bankert,
and Rebecca C. Craven

ABSTRACT

During retroviral maturation, the CA protein undergoes dramatic structural changes and establishes unique intermolecular interfaces in the formation of the mature capsid shell that are different from the interactions that existed in the immature precursor. The most conserved region of CA, the major homology region (MHR), is implicated in both immature and mature assembly although the precise contribution of the MHR residues to each event is largely undefined. To test the roles of specific MHR residues in mature capsid assembly, an in vitro system was utilized that stimulated RSV CA to form, among other things, structures resembling authentic capsids. The ability of CA to assemble organized structures was destroyed by substitutions of two conserved hydrophobic MHR residues, demonstrating that these MHR residues are required for proper assembly of mature capsids in addition to any role that these amino acids may play in immature assembly. The defect caused by the MHR mutations was identified as an early step in the capsid assembly process. Consistent with a defect in the initiation of assembly, second-site suppressors that restored both infectivity and in vitro assembly to the MHR substitutions stimulated the nucleation of capsid assembly. The results provide strong evidence for a model in which the hydrophobic residues of the MHR control a conformational reorganization of CA that is needed to initiate capsid assembly and suggest that the formation of an interdomain interaction occurs early during maturation.
INTRODUCTION

The retroviral CA protein plays critical structural roles in each of the two distinct stages of virion assembly. When an immature particle is formed by the polymerization of the Gag polyprotein, the N-terminal (NTD) and C-terminal (CTD) CA domains embedded within Gag control packing and assembly (5, 8, 52, 55, 65, 251, 327, 496, 505). Subsequently, processing of Gag by the viral protease initiates a maturation process in which the mature MA (matrix), CA (capsid), and NC (nucleocapsid) proteins are released from Gag. The free CA polymerizes as a capsid shell around the genomic RNA and NC protein creating the core of the mature virion. The maturation events are complex and include disruption of the CA-CA interfaces that held the Gag proteins together, conformational changes within each domain of CA and the formation of new CA-CA interfaces of the mature capsid shell (33, 50, 52, 160, 181, 257, 297, 319, 355, 476, 504, 505, 518).

In spite of limited sequence similarity, the three-dimensional structure of mature CA is highly conserved among retroviruses and consists of the two mostly α-helical domains, NTD and CTD, connected by an interdomain linker (37, 67, 92, 160, 171, 228, 234). After maturation is completed, the final capsid shell consists of a lattice of CA pentamers and hexamers. The hexamers are established by NTD-NTD interactions, which are linked by CTD-CTD dimerization (160, 161, 163, 256, 276, 295, 319, 504). The dimer interface is formed by the dimerization helix, the second helix of the CTD. A third interface, a NTD-CTD interdomain interaction that forms during maturation, was originally predicted by a genetic study of the Rous sarcoma virus (RSV), and subsequently confirmed and mapped by biochemical and structural studies examining in vitro assembled human immunodeficiency virus (HIV) CA protein (48, 163, 256, 257). Mutagenesis in HIV and the binding of retroviral inhibitors have documented that this interdomain interface is essential for capsid integrity and infectivity (145, 164, 226, 441, 457, 462, 477).
The start of the CTD contains the highly conserved MHR motif that participates in the assembly of both the immature and mature virus. Numerous substitutions in the MHR to the three absolutely conserved polar residues and the conserved hydrophobic residues cause severe defects in immature virus assembly and a loss of infectivity, presumably by disrupting key steps in Gag assembly (61, 96, 261, 290, 341, 442). Certain conservative substitutions at many of the same positions, however, appear to compromise the formation of the mature capsid shell instead of Gag assembly (61, 96, 261, 290). In RSV the loss of infectivity caused by such substitutions was traced to an improperly formed core structure, resulting in a failure of genome replication (48, 61).

The conserved residues of the MHR are not involved directly in any of the known intermolecular interfaces in the mature capsid shell but instead contribute to the structure of the CTD through intramolecular interactions (163, 171, 234, 256, 276, 504). Residues downstream of the MHR are, in contrast, known to be important for the CTD-CTD and NTD-CTD interfaces (160, 163, 256, 257, 276, 504). Many second-site suppressors in RSV that restore infectivity to non-infectious CA mutants are located downstream of the MHR motif and in other regions of CA that are implicated in maturation and the formation of the mature capsid (48, 283, 431). This observation suggests that the conserved MHR substitutions perturb capsid formation by affecting the folding during maturation or final structure of the CTD.

A possible explanation for the role of the MHR in Gag assembly and maturation was suggested by a crystal structure of a dimer interface in the HIV CTD (204, 205, 236). A mutation in the CTD caused a relaxation of the bend between the first and second helices allowing dimerization by a “domain-swap” arrangement wherein the structural elements containing the MHR and the following helix form a largely hydrophobic interface between two Gag molecules. The model is supported by the structural homology between the Gag protein and certain mammalian transcription factors that dimerize via this domain-swap
mechanism (204, 236). Cryo-electron tomographic and cryo-electron microscopic analyses of immature particles, however, have been of insufficient resolution to confirm this model (52, 55, 496, 505, 518).

The domain-swap model proposes that certain residues of the MHR may directly mediate an intermolecular interaction in the immature virus. In the mature form, however, the same residues participate in only intramolecular interactions. Accordingly the conserved hydrophobic MHR residues must move from an exposed position (in the domain-swap arrangement) into the interior of the CTD during maturation (160, 163, 205, 504). Such a reorganization of the CTD was previously suggested based on results observed with an anti-MHR antibody that recognized uncleaved Gag protein but not mature CA protein (96). However, the molecular events of capsid assembly occurring during maturation (i.e. the precise nature of the structural transformations, the sequence of events or intermediates in the capsid assembly pathway, and the role of specific residues in controlling these events) have remained largely uncharted. Given the importance of the MHR to the maturation process and the demonstrated potential of maturation inhibitors as effective antiviral therapies (275, 401, 424, 441, 457, 462), an understanding of the molecular details of this process remains a critical unmet need.

The MHR and suppressor mutations provide a powerful tool to allow the genetic and biochemical dissection of events occurring during retroviral maturation. The ability of the MHR to control mature capsid assembly, independent of any role in immature assembly, was examined by selecting MHR mutations which cause a loss of core integrity and infectivity without any obvious consequence on Gag assembly. Overall, the results reveal that certain residues of the MHR, in particular the conserved hydrophobic amino acids, have a critical role during maturation in facilitating an early step in capsid assembly which likely involves interdomain interactions.
Table 3.1 Phenotypes of WT and mutant viruses in cells

<table>
<thead>
<tr>
<th>Virus</th>
<th>Infectivity</th>
<th>Budding Efficiency</th>
<th>RT Efficiency</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(48, 61, 96, 283)</td>
</tr>
<tr>
<td>F167Y</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>(48, 61, 96)</td>
</tr>
<tr>
<td>F167Y/I190V</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(48, 283)</td>
</tr>
<tr>
<td>L190V</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(48, 283)</td>
</tr>
<tr>
<td>L171V</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>(61, 96)</td>
</tr>
<tr>
<td>L171V/A38V</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(48, 283)</td>
</tr>
<tr>
<td>A38V</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(48, 283)</td>
</tr>
<tr>
<td>D155Y/R185W</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(283)</td>
</tr>
<tr>
<td>R185W</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(283)</td>
</tr>
<tr>
<td>E162Q/F193L</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>(283)</td>
</tr>
<tr>
<td>E162Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(61)</td>
</tr>
</tbody>
</table>

The infectivity, budding (release from infected cell) efficiency, and RT (reverse activity in a newly infected cell) were previously characterized. The efficiency of each mutant was compared to the WT (wild-type) virus: + ≥ 60% activity of the WT virus, - < 60% of the WT virus.
MATERIALS AND METHODS

Protein expression and purification

Protein was expressed using a pET-24(+) plasmid in which RSV CA sequence was inserted using the NdeI and HindIII sites. The expressed protein corresponded with RSV Prague C sequence from Pro1 to Ala237 of CA and included a non-native initiating Met residue. Cleavage of the Met from the expressed protein was confirmed by mass spectrometry for the WT and MHR mutant proteins. The D52A and A38V mutations were created by QuikChange site-directed mutagenesis (Stratagene). All other mutations were amplified from previously constructed plasmids (48, 61) and were inserted into the pET-24(+) plasmid containing CA using the PstI and HindIII restriction sites.

The CA protein was expressed in E. coli (BL21 DE3) using the auto-induction system developed by Studier (448) and purified by DEAE and SEC as described in chapter II. Isotopically-labeled protein was produced by bacterial growth in MD-5052 minimal medium supplemented with $^{15}$N NH$_4$Cl. CA was monitored at each step of purification by SDS-PAGE and staining with Coomassie Blue. Expression, solubility and purification of the mutant proteins were similar to that of the WT protein. The following extinction coefficient values were used: 24980 M$^{-1}$cm$^{-1}$ for WT, I190V, L171V, L171V/A38V, A38V, E162Q, E162Q/F193L, F193L, D52A; 26470 M$^{-1}$cm$^{-1}$ for F167Y, F167Y/I190V and D155Y; 31970 M$^{-1}$cm$^{-1}$ for D155Y/R185W; and 30480 M$^{-1}$cm$^{-1}$ for R185W.

Circular dichroism (CD), intrinsic tryptophan (Trp) fluorescence, and unfolding equilibrium

For CD, a Jasco J-710 spectropolarimeter operating in the far-UV region of the spectrum was used for determination of secondary structure. The protein was diluted in 50 mM sodium phosphate pH 7.5, 50 mM NaCl, and 0.1 mM EDTA to a final concentration of 0.1 mg/ml. Intrinsic Trp fluorescence was performed using a 290 nm excitation wavelength
on a PTI QuantaMaster luminescence spectrometer. Protein stability was measured as a function of guanidine hydrochloride (Gdn-HCl) concentration (0 to 5.5 M Gdn-HCl) in 20 mM Tris-HCl pH 7.5 and 0.1 mM EDTA. Analysis was performed by fitting the data to a two state unfolding model as previously described (387).

**In vitro capsid assembly**

Capsid assembly was monitored by optical density (OD) at 450 nm using a SpectraMax Plus384 spectrophotometer (Molecular Devices). Unless otherwise noted, assembly conditions were: 80 µM (2 mg/ml) CA protein in 10 mM Tris-HCl pH 7.5, 75 mM NaCl and 0.05 mM ETDA and was initiated by the addition of 500 mM sodium phosphate pH 7.5 (final concentrations). Each reaction was performed in 100 µl volumes in a 96-well UV-Transparent microplate (BD Falcon). Typically, one minute elapsed between the initiation of assembly and the first reading. The plate was mixed for 5 sec prior to the first reading and for 3 sec prior to each subsequent reading.

The change in OD\textsubscript{450} (ΔOD\textsubscript{450}) was calculated by subtracting the OD\textsubscript{450} value of the in vitro assembly reaction lacking sodium phosphate from the OD\textsubscript{450} value of a parallel reaction containing 500 mM sodium phosphate. For most proteins, the ΔOD\textsubscript{450} graph was a sigmoidal curve resembling a multiphasic reaction with lag, growth and stationary phases. The modified Gompertz equation was used to objectively compare assembly reactions (548). For clarity, only one in ten data points are shown on most of the graphs.

**Electron microscopy (EM)**

The products of the in vitro assembly reactions were examined by EM. Upon reaching the maximum OD\textsubscript{450} the samples were removed from the spectrophotometer and stored at 4° for 12-16 hrs. During storage the assembled protein settled at the bottom of the
tube, allowing for concentration without centrifugation. The settled samples were transferred to a formvar-coated grid, washed with water, stained with 2% uranyl acetate and visualized using a 120 kV transmission electron microscope (TEM, Phillips) operating at 60 kV. The in vitro assembled capsid-like structures retained heavy amounts of stain, thus the contrast of the resulting EM images were adjusted using Adobe Photoshop.

RESULTS

The D155Y, F167Y and L171V MHR mutations examined below (Fig 3.1) each perturbed the mature core integrity as judged by an increased sensitivity to non-ionic detergent and a loss of viral DNA synthesis (48, 61, 96). None of these three mutations caused any overt disruption of Gag functions or immature particle assembly. As a comparison, a phenotypically distinct MHR mutation (E162Q) which alters an absolutely conserved residue was included. In contrast to the other MHR mutations, this substitution reduced particle release from infected cells (61). For all four MHR mutations, second-site suppressors (Fig 3.1) that restore infectivity (48, 283) were also examined in this study. The possibility that the defect in core integrity caused by the MHR mutations might be an indirect consequence of a minor alteration of Gag organization can not be ruled out. Alternatively, these MHR residues may have a direct role in the ability of CA to form the capsid shell. The ability of WT and mutant CA proteins to multimerize in vitro was used to differentiate these two possibilities.

MHR mutations do not affect CA structure and stability

As a preamble to testing the effects of the mutations on capsid assembly, their influence on the secondary structure and stability of monomeric CA was assessed. As
Fig 3.1. Ribbon diagram of the monomeric structure of RSV CA. The locations of the MHR (D155, E162, F167 and L171 colored red) and suppressor (A38, R185, I190 and F193 colored green) residues mutated in this study are shown in a ball and stick fashion. The diagram was made using PBD ID 1em9 (NTD) and 1d1d (CTD); a dashed line represents the interdomain linker. Two views that are rotated by 90° are shown.
expected, the purified monomeric WT protein was highly α-helical as observed by the large negative peaks at 222 nm and 208 nm when examined by CD (67, 234, 432) (Fig 2A). All of the mutant proteins had CD spectra that were indistinguishable from WT, indicating that they were properly folded without any large-scale defects in secondary structure (Fig 3.2A, and data not shown). Next, the effects of the MHR and suppressor substitutions on protein stability were evaluated. WT CA protein was unfolded by guanidine hydrochloride (Gdn-HCl) with a midpoint concentration of ~2.2 M, determined by either the relative ellipticity at 222 nm or the intrinsic Trp fluorescence at 325 nm (Fig. 3.2B and Table 3.1). No increased sensitivity to Gdn-HCl was observed for any of the mutant proteins, indicating a lack of gross destabilization (Table 3.1).

However, the behavior of the D155Y/R185W and R185W proteins deviated from that of WT protein by exhibiting a shoulder at higher concentrations of Gdn-HCl whether followed by CD or intrinsic Trp fluorescence (Fig 3.2B and Table 3.1). Thus, the substitution of a Trp at position 185 appears to have stabilized a partially folded intermediate that was not observed in WT protein or in protein containing only the D155Y substitution (Fig 2B). A similar effect was observed by CD with protein bearing the F193L suppressor without its corresponding MHR mutation (Table 3.1). Intrinsic Trp fluorescence failed to detect this stabilization, presumably due to a lack of a Trp residue in the relevant region.

*Sodium phosphate induced in vitro RSV CA multimerization*

To dissect the importance of particular MHR residues for capsid assembly, a turbidimetric assay was utilized (108, 259, 347, 370, 417). When coupled with EM analysis, the turbidimetric assay allowed for a quantitative and qualitative evaluation of the assembly capabilities of the MHR and suppressor mutant proteins. In this system, monomeric CA
Fig 3.2. Far UV analysis and protein stability of WT and mutant CA proteins. (A) The secondary structure was determined by circular dichroism from 195-255 nm. (B) Protein stability was determined by monitoring the α-helical content of the protein at 222 nm from 0 - 5.5 M guanidine hydrochloride (Gdn-HCl).
Table 3.2. Thermodynamic parameters for the unfolding of WT and mutant CA

<table>
<thead>
<tr>
<th>Protein</th>
<th>Far UV CD</th>
<th>Intrinsic Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Midpoint (M Gdn-HCl)</td>
<td>∆G</td>
</tr>
<tr>
<td>WT</td>
<td>2.24</td>
<td>6.61</td>
</tr>
<tr>
<td>F167Y</td>
<td>2.21</td>
<td>6.97</td>
</tr>
<tr>
<td>F167Y/I190V</td>
<td>2.22</td>
<td>6.16</td>
</tr>
<tr>
<td>I190V</td>
<td>2.22</td>
<td>5.70</td>
</tr>
<tr>
<td>L171V</td>
<td>2.27</td>
<td>4.73</td>
</tr>
<tr>
<td>L171V/A38V</td>
<td>2.05</td>
<td>4.02</td>
</tr>
<tr>
<td>A38V</td>
<td>2.12</td>
<td>5.48</td>
</tr>
<tr>
<td>D155Y</td>
<td>2.26</td>
<td>5.88</td>
</tr>
<tr>
<td>D155Y/R185W</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>R185W</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>E162Q</td>
<td>2.15</td>
<td>5.23</td>
</tr>
<tr>
<td>E162Q/F193L</td>
<td>2.23</td>
<td>6.07</td>
</tr>
<tr>
<td>F193L</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

N.D. The value was undetermined since the data did not fit a two-state [Native → Unfolded] model as described in the Material and Methods. ∆G, change in Gibbs free energy.
Fig 3.3. Sodium phosphate-induced assembly of purified WT and D52A proteins. (A) The assembly reaction of WT and D52A protein was followed by turbidity as described in the Materials and Methods section. Electron microscopy of WT (B-J) and D52A (K) proteins assembled at 500 mM sodium phosphate pH 7.5 at 80 μM; and WT protein (L-N) assembled at 500 mM sodium phosphate pH 8.0 at 160 μM. Scale bars represent 100 nm.
protein at 2 mg/ml (80 µM) was stimulated to assemble by the addition of 500 mM sodium phosphate pH 7.5. After a lag period, assembly was evidenced by an increase in optical density (ΔOD<sub>450</sub>) due to light scattering by the large structures (Fig 3.3A). Under these conditions, the lag time for the WT protein was ~190 min and a maximum optical density was reached after an additional 150 min. The duration of the lag time and the ΔOD<sub>450</sub> reached were dependent upon both sodium phosphate and protein concentration (Chapter II).

Examination of assembled WT CA protein by EM revealed a diversity of structures (Fig 3B-C, and Fig 3L-N), including abundant spheroidal structures and large planar arrays similar to those previously described (234) (Chapter II). The spheroidal structures had a diameter of 16 - 80 nm. The edges of the planar arrays appeared curled to form a tubular structure with a diameter of ~85 nm (Fig 3.3B). Other cylindrical structures appear to be completely formed tubes with a diameter of ~50 nm (Fig 3.3L) rather than curled planar arrays. Both the planar arrays and tubes had varying lengths (200 -1500 nm). Particularly notable was the presence of angular structures (Fig 3.3D-J and 3.3L-N) similar to the irregular polyhedra observed in native RSV capsids (58, 235). Many of the assembled structures were multi-lamellar (Fig 3.3D-I). Such angular structures were not apparently formed under the non-physiological, low pH in vitro conditions used previously (234). A similar range of structures was described for HIV in both virions and in vitro assembled CA protein (33, 50, 54, 179, 259, 276, 476, 490). The shapes and variety of the structures formed at 160 µM CA protein and pH 8.0 (Fig 3.3L-N) appeared identical to ones formed at 80 µM and pH 7.5 (Fig 3.3B-J). The formation of angular capsid-like structures indicates that RSV CA has the ability to form authentic CA-CA interfaces at near neutral pH under the sodium phosphate-induced assembly conditions and supports the validity of this assay to evaluate the effect of CA mutations on capsid assembly.

A mutation expected to disrupt the ability of CA to properly assemble due to a loss of the N-terminal β-hairpin was tested to further validate the in vitro assay (476). Upon cleavage
of CA from Gag, the amino terminus of the protein refolds forming a salt bridge between the terminal Pro residue and a conserved Asp residue (47, 171, 476). Disruption of this salt bridge by an Asp to Ala substitution results in a loss of in vitro CA assembly and infectivity (397, 461, 476). In RSV, the behavior of the D52A mutant CA protein deviated from WT protein by rapidly forming light-scattering material within 15 min (Fig 3.3A). However, the final ∆OD<sub>450</sub> value was lower than for WT protein. No capsid-like structures were formed by the D52A protein, instead, the electron micrographs revealed only dark staining regions of amorphous protein precipitates (Fig 3K), with only a rare inchoately formed cylindrical structure interspersed within the precipitate. These observations are likely due to uncontrolled aggregation of the D52A protein triggered by sodium phosphate, consistent with the phenotype of the analogous D51A mutant in HIV CA (476).

*F167Y and L171V mutant proteins lack the ability to properly assemble, but assembly is restored by intradomain and interdomain suppression*

In the mature structure of CA, F167 and L171 are located proximally to one another within the hydrophobic core of the CTD. Both mutations dramatically inhibited assembly of CA protein in vitro. The F167Y mutant protein formed a small amount of light-scattering material after a greatly lengthened lag time (~550 min) and only amorphous precipitate was observed by EM (Fig 3.4A and 3.5A). Similarly, the L171V protein showed little, if any, assembly in sodium phosphate-induced conditions (Fig 3.4B and 3.5D). These assembly defects imply that the F167 and L171 residues of the MHR are essential for proper capsid assembly.

To further test this idea, the ability of the second-site suppressor mutations I190V and A38V to restore assembly competence to the MHR mutant proteins was examined. The I190V suppressor mutation corrected the in vitro assembly defect of F167Y (Fig 3.4A),
consistent with its effect on virus replication (48). The F167Y/I190V double mutant protein behaved similarly to WT in the turbidimetric assay, having a comparable lag time and maximum ΔOD_{450}. The suppressor of the L171V mutation, the A38V mutation located on helix 2 of the NTD, likewise restored assembly and shortened the long lag phase of L171V protein to a more WT-like level (Fig 3.4B).

The ability of the suppressor mutations to restore assembly to the MHR mutant proteins was also demonstrated by the presence of organized structures in the turbid samples. However, in neither case (F167Y/I190V or L171V/A38V) was the full range of morphologies seen in WT protein reproduced with the double mutant. The F167Y/I190V protein predominantly assembled into spheroids of ~30 nm that often lacked distinct edges and well defined centers (Fig 3.5B). In contrast, only tube-like structures with a diameter of ~50 nm were seen in the assembled L171V/A38V protein (Fig 3.5E). These structures resembled the tubes and planar arrays seen in the WT protein, but many were considerable longer measuring up to ~2 μm in length.

The two suppressor mutations dramatically increased CA multimerization when the MHR substitutions were not present. The I190V protein assembled very quickly with a lag of only a 3.5 min (Fig 3.4A). Like the F167Y/I190V mutant protein, the I190V protein predominately assembled into ~30 nm spheroids (Fig 3.5C). Many of the I190V spheroids were more distinct than those formed by F167Y/I190V protein, with easily discerned edges and centers. The effects of the A38V suppressor mutation were even more striking than those caused by the I190V suppressor. The rise in turbidity by the A38V protein was so rapid that the lag time could not be calculated since the observed curve was not sigmoidal under these conditions (Fig 3.4B). Spheroids, short tubes and capsid-like structures similar to WT were formed by the A38V protein (Fig 3.5F). The results indicate that the I190V and A38V suppressor mutations strongly increase the propensity of CA to multimerize.
Fig 3.4. Determination of *in vitro* assembly phenotypes of the lethal-suppressor series of mutations. F167Y-I190V series (A), L171V-A38V series and F167Y/A38V double mutant (B), D155Y-R185W series (C), and E162Q-F193L series (D). The reactions were stopped when the maximum $\Delta \text{OD}_{450}$ was reached.
Fig 3.5. Examination of assembled mutant proteins by EM. All proteins were assembled using standard conditions (500 mM sodium phosphate pH 7.5 and 80 µM protein). A. F167Y, B. F167Y/I190V, C. I190V, D. L171V, E. L171V/A38V, F. A38V, G. D155Y, H. D155Y/R185W, I. R185W, J. E162Q, K-L. E162Q/F193L, M. F193L, and N. F167Y/A38V. Arrows in panel K (E162Q/F193L) point towards structures that contain or overlap smaller spheroidal structures. The names of the double mutants are abbreviated and do not contain the residue numbers. All scale bars represent 100 nm.
The L171V/A38V double mutant may have assembled more slowly than did the WT protein as indicated by a decreased slope during the growth phase and a lower maximum ΔOD₄₅₀ (Fig 3.4B). However, the abundance of long tubes and lack of other structures (Fig 3.5E) suggests that some of these effects may be due to a lower inherent ability of long tubes to scatter light (36). As an alternative method to follow assembly, a differential centrifugation protocol that separated large assembled products from soluble CA was utilized (Fig 3.6). As predicted from the turbidimetric assay, WT protein was found mostly in the supernatant fraction at 5 and 100 min, in both fractions at 200 min, and mostly in the pellet at 300 min (Fig 3.6). The assembly-defective L171V protein was present only in the supernatant fraction at all time points. In contrast, a detectable amount of L171V/A38V protein was pelletable at early time points, consistent with a WT-like lag time. The amount of pelletable L171V/A38V protein increased slowly with time, continuing to increase at 300 min (Fig 3.6) and beyond (data not shown). Consistent with the turbidity measurements, A38V rapidly assembled into pelletable material (Fig 3.6). This assay confirms that during the growth phase the L171V/A38V protein assembled more slowly than did the WT, even though their lag times were similar. This difference in assembly rates may be due to the drastically different range of particle morphologies formed by these two proteins.

Although the F167 and L171 residues are located near each other, their assembly defects were reversed by secondary mutations located in different domains of CA (Fig 3.1). To test the specificity of suppression, F167Y/A38V protein was tested under in vitro assembly conditions. Assembly of F167Y/A38V protein was especially robust, showing very rapid assembly (Fig 3.4B) and is consistent with the observation that the F167Y/A38V mutant virus is infectious (283). The F167Y/A38V structures appeared multi-lamellar and included angular structures resembling those formed by WT protein (Fig 3.5N). The lack of specificity exhibited by the A38V suppressor and the similar in vitro behavior exhibited by
Fig 3.6. Assembly of L171V-A38V series of mutant proteins examined by differential centrifugation. At 5, 100, 200, and 300 min after initiation of assembly, 5 µl samples were removed from the 100 µl reactions and centrifuged at 18,000 x g for 1 min. The supernatant was removed and the pellet was resuspended in a final volume equivalent to the supernatant fraction. Both fractions were subjected to SDS-PAGE. S, supernant fraction and P, pelleted fraction.
<table>
<thead>
<tr>
<th></th>
<th>5min</th>
<th>100min</th>
<th>200min</th>
<th>300min</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>L171V</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>LV / AV</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>A38V</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th></th>
<th>400min</th>
<th>500min</th>
<th>400min</th>
<th>500min</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>LV / AV</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>
both the F167Y and L171V mutations suggest that F167 and L171 have similar functional roles in capsid assembly.

**D155Y MHR mutant protein retains partial multimerization ability**

The D155 residue is located at the beginning of the MHR near the junction of the interdomain linker and the CTD (Fig 3.1). When tested for effects on *in vitro* assembly capacity, the D155Y substitution behaved like the F167Y and L171V mutations and crippled protein assembly as monitored by turbidity (Fig 3.4C). Assembly was restored to a WT-like pattern by the secondary R185W mutation. As was true of the I190V and A38V suppressor mutants, the R185W substitution in the absence of an MHR mutation dramatically accelerated the formation of light-scattering material upon addition of sodium phosphate (Fig 3.4C). Thus, the three MHR mutations and their suppressors exhibit very consistent effects on *in vitro* assembly.

Unlike F167Y and L171V, the D155Y protein retained some ability to form organized structures (Fig. 3.5G). Small spheroids (in a range of 16-50 nm, mean 30 nm) were readily seen, some of which were observed in chain-like arrays aligned in a roughly cylindrical shape. The larger multi-lamellar or polyhedral structures, commonly seen with the WT protein, were rare (data not shown). The D155Y/R185W protein readily formed larger angular structures (Fig 3.5H) in addition to small spheroids and chain-like arrays of small spheroids similar to the ones observed in the D155Y sample. The R185W mutation by itself favored the formation of large multi-lamellar, angular structures similar to ones seen frequently with the WT protein (Fig 3.5I). Unlike the D155Y and D155Y/R185W proteins, the R185W protein formed tubular structures. Thus, the D155Y mutant possesses a limited ability to multimerize and can be restored to a more WT-like behavior by the R185W secondary substitution.
Lethal E162Q mutation is assembly competent in vitro

The absolutely conserved Glu residue in the MHR (E162 in RSV and E159 in HIV) participates in an extensive intramolecular hydrogen bonding network in both the mature CA and in the domain-swap model for immature Gag assembly (160, 205, 228, 234). In RSV, a Gln substitution at this position resulted in a loss of infectivity (61) which can be suppressed by the F193L second-site mutation (283). The E162Q substitution caused a slight reduction (~40%) in virus release from infected cells (61). The F193L suppressor restored E162Q budding to WT-like levels (283), suggesting that these mutations cause significant impact on Gag function(s).

The E162Q protein readily assembled in vitro (Fig 3.4D) forming structures including the angular multi-lamellar structures similar to those formed by WT protein (Fig 3.5J). In the E162Q/F193L double mutant protein, the F193L suppressor had little effect on in vitro assembly (Fig 3.4D and Fig 3.5K-M), only marginally increasing the turbidity over that achieved by the E162Q protein without improving the lag time. E162Q/F193L protein formed structures that contained or over-lapped smaller spheroidal structures (Fig 3.5K, arrows) and resemble the nested capsids observed in some HIV and RSV particles (33, 50, 58). F193L was the only suppressor tested that failed to stimulate assembly in the context of a double mutant protein (Fig 3.4D). The F193L mutation alone had no detrimental or stimulatory effect on in vitro assembly (Fig 3.4D). Therefore, the phenotypic effects of E162Q and F193L on virus replication are likely not related primarily to alterations in mature capsid assembly, but may be influencing Gag assembly, possibly by altering the domain-swapping of the MHR residues or some unknown mechanism.
Defective assembly ability of L171V and F167Y is rescued by the addition of WT protein

The defect on capsid assembly caused by the F167Y and L171V mutations was further examined by testing the ability of WT protein to rescue the defect(s) of assembly. F167Y and L171V mutations were tested by mixing the WT protein with each of the mutant proteins at equimolar ratios (40 µM each). WT protein by itself at 40 µM exhibited a long lag and limited light-scattering; doubling the protein to 80 µM shortened the lag period and resulted in a maximum turbidity that was roughly twice that of the lower concentration (Fig 3.7A). Adding F167Y mutant protein to the WT protein (each at 40 µM) almost doubled the final turbidity without greatly changing the lag time (388 min for WT vs. 374 min for the mixed sample). Reducing the WT protein to 20 µM and increasing the mutant protein to 60 µM caused an approximately proportional increase in the lag time (500 min for WT vs. 450 min for the mixed sample, data not shown). The mutant protein may stimulate WT assembly by acting as a crowding agent (106). However bovine serum albumin (BSA) protein could not replicate the stimulatory effect of F167Y and actually caused a slight inhibition on WT assembly (Fig 3.7A), arguing against this explanation. The above data suggest that the mutant protein was incorporated into growing structures initiated by WT protein.

The presence of mutant protein in the assembled material was assessed by utilizing $^{15}\text{N}$ labeled protein. Isotopically labeling the F167Y protein with $^{15}\text{N}$ will increase the mass of CA by almost 300 daltons, allowing differentiation of the labeled mutant from unlabeled WT protein by mass spectrometry (MS) (Fig 3.7B-D). Labeling of the F167Y protein had no effect on assembly as determined by turbidity (data not shown). An equimolar ratio of unlabeled WT and $^{15}\text{N}$-labeled F167Y protein were mixed in an assembly reaction, and once the maximum turbidity was reached, the products were centrifuged and washed with 500 mM sodium phosphate. The pellet was extensively dialyzed against buffer (10 mM Tris-HCl pH
7.5) to disassociate the assembled material. Both WT and mutant proteins were found by MS in the recovered material at approximately equal amounts (Fig 3.7D). Altering the molar ratio so that only 25% of the reaction was unlabeled WT protein resulted in a proportional reduction in the amount of WT protein recovered (data not shown). Under similar conditions, $^{15}$N-labeled F167Y protein could not be recovered without the addition of WT protein (data not shown). Furthermore, when WT protein was assembled in the presence of equimolar BSA, no BSA was found by MS (data not shown). This confirms the observation that F167Y mutant protein is capable of co-assembly with WT protein. A parallel series of experiments showed the same behavior with the L171V protein (data not shown).

When mixed with WT protein the mutant CA proteins were co-assembled without altering the lag time, suggesting that the mutants were able to participate in the growth phase but not nucleation. F167Y and L171V mutant proteins were used to further examine the importance of the CTD for the initiation of capsid assembly. The opposing effects on in vitro assembly kinetics of the lethal mutations (an extremely prolonged lag) and second-site compensatory substitutions (rapid assembly kinetics) provides further evidence that the crippling effect of F167Y and L171V may be due to prevention of nucleation of capsid assembly (Fig. 3.4). Action of the lethal mutations on the nucleation step was tested more directly by using the two-phase assembly reaction previously described in chapter II (Fig. 2.3). The mutant proteins were pre-incubated in low phosphate (50 mM) prior to the addition of monomeric WT protein and raising the sodium phosphate level to 500 mM (Fig. 3.8). Both mutant proteins failed to cause the same immediate stimulation of assembly that was seen with the WT protein, confirming earlier interpretations that the integrity of the CTD is critical for the formation of the initiating oligomer. The assembly detected after a considerable lag was likely nucleated by the WT protein added during the second phase. The deficit caused by the F167Y and L171V mutations may be due to influences on either the intradomain (CTD-CTD) and/or the interdomain (NTD-CTD) interface. However, the fact that both of these
Fig 3.7. Rescue of F167Y non-assembly phenotype by WT protein. WT protein was mixed with either F167Y mutant protein (○) or BSA (◊) at equimolar concentrations, 40 µM each (A). Also shown are the turbidimetric profiles of unmixed WT (△), F167Y (▽) and BSA (---) proteins at 40 µM; and WT protein at 80 µM (□). For clarity, only every 20th data point is shown. Mass Spectrometry (MS) analysis of unlabeled WT protein (B) and 15N-labeled F167Y protein (C) that were not mixed together and not subjected to sodium phosphate assembly conditions. 15N-labeled F167Y mutant protein mixed with WT protein at 40 µM each was induced to assemble by addition of sodium phosphate and prepared for MS analysis as described in the text (D).
Fig 3.8. Two lethal CTD CA mutations are defective at nucleation. F167Y and L171V were tested for their influence on assembly initiation by incubating the mutant protein at 50 mM sodium phosphate pH 8.0 and then at 60 min spiking the reaction with additional salt and WT protein, as in Fig 2.3.
lethal CTD mutations can be compensated by a second-site change in the NTD suggests that the interdomain interface is more likely the critical interface formed during nucleation.

Suppressor mutations increased the sensitivity of CA to sodium phosphate

Proteins containing the A38V, R185W and I190V suppressor mutations without a lethal substitution exhibited dramatically increased assembly rates (Fig 3.4A-C, suppressor mutants). The rate of capsid assembly observed with the suppressor mutant proteins was compared to that of the WT protein over a range of sodium phosphate concentrations (0.1 M to 1.0 M). The three suppressor proteins were easily stimulated to assemble at lower concentrations of sodium phosphate than were the WT protein (Fig 3.9). The assembly rates of A38V, R185W and I190V were greater than that of WT at every sodium phosphate concentration tested, whereas the other suppressor (F193L) protein showed a WT-like assembly rate across the entire range. Both the greatly reduced lag time (Fig 4A-C) and the increased sensitivity to sodium phosphate (Fig 3.9) caused by the A38V, R185W and I190V suppressor mutations are consistent with a stimulation of an early step of capsid assembly such as nucleation.

DISCUSSION

The two domains of CA each undergo dramatic structural changes during maturation and establish unique intermolecular interactions in the capsids that are different from those that existed between CA domains of Gag in the immature particle (163, 181, 205, 257, 505). The mechanisms and determinants of these structural transformations are poorly defined. The conserved MHR motif has been amply shown by genetic analyses to be necessary for both immature and mature virus organization (61, 96, 251, 290, 442, 477). These overlapping functions have made it difficult to discern from genetic studies alone the contributions of
Fig 3.9. Sensitivity of WT and suppressor mutant proteins to sodium phosphate. The assembly rate was determined by fitting the growth phase of the turbidity curve to a linear equation and calculating the $\Delta OD/\Delta Time$, where $\Delta OD = \text{final } OD_{450} - \text{initial } OD_{450}$ of the growth phase and $\Delta Time = \text{final time} - \text{initial time}$ of the growth phase. The log of the rate was graphed as a function of sodium phosphate concentration.
specific residues to capsid assembly. By utilizing an *in vitro* approach and a set of well-characterized mutations, certain conserved residues of the MHR were shown to be essential to the formation of mature capsids.

*In vitro assembly of RSV CA*

RSV CA self-assembled *in vitro* into structures of diverse morphologies—spheroidal structures of varying diameters, tubular structures with varying lengths and occasional large structures that have the angular features of authentic RSV cores (58, 235). Examination of authentic RSV particles by cryo-electron tomography documented a similar range of polymorphic structures (58) suggesting that all the CA subunit interactions for *in situ* also formed under the *in vitro* conditions. Several recent studies have suggested that the overall morphology of a capsid in a mature virion may be influenced by additional virion components such as the lipid membrane, Env glycoprotein tails, genomic RNA, or partially cleaved Gag (24, 25, 33, 50, 58). The formation of *in vitro* structures distinctly different from authentic capsids (i.e. long tubes, Fig 3.5E and 3.5M) may reflect the lack of regulation of capsid assembly by such component(s) or reflect an inability of some of the mutant CA proteins from adopting all of the CA-CA interactions that WT CA subunits form. Regardless, the results illustrate the intrinsic ability of RSV CA to form capsid-like structures without the requirement of any other viral component, as shown for HIV CA under high ionic strengthen conditions (179, 259, 276).

The RSV CA assembly reaction had a pattern of an extended lag phase and subsequent growth phase (Figs. 3.3A and 3.4A-D). Similar results were reported for HIV CA multimerization using NaCl-induced assembly conditions (24, 259). This pattern is consistent with a nucleation step followed by elongation and closure steps that were proposed for HIV and RSV capsid assembly *in vivo* (33, 50, 58). The events occurring during the lag period
could not be directly observed, however the use of methods with greater sensitivity than optical light scattering may allow for determining the nucleation step more precisely.

The *in vitro* behavior of mutant CA proteins provides a strong argument that capsid assembly in the maturing virion is also nucleation-driven. Substitutions that alter highly conserved residues in the CTD and cripple capsid integrity in virions also interfere with nucleation of assembly *in vitro* (Fig. 3.8). Second-site compensatory mutations in CA that restore *in situ* capsid formation [Butan and Steven, unpublished] and correct the replication defect in authentic virions (48, 283) also overcome the block to CA nucleation (Fig. 3.4). These opposing effects of lethal and compensatory mutations on *in vitro* nucleation-driven assembly provide strong biochemical evidence in support of the proposal that capsid assembly begins *in situ*, as it does *in vitro*, with a specific nucleation event (33, 50, 58).

**A role for the MHR in controlling capsid assembly**

The lethal F167Y and L171V substitutions caused a complete loss of the initiation of CA assembly *in vitro* (Fig 3.4A-B and Fig 3.5A and D). The conservative F167Y and L171V alterations cause no observable diminishment of immature virus release *in vivo* (48, 61, 96) nor do they interfere with *in vitro* Gag assembly (J. Phillips and V. Vogt, unpublished data). Additionally, the D155Y substitution also limited mature capsid assembly (Fig 3.4C and 3.5G). Although this mutation caused a less severe multimerization defect, signifying that the D155Y mutation affects assembly in some way that is qualitatively different from the F167Y and L171V mutations. In total, these findings argue that certain MHR residues, particularly the conserved hydrophobic residues, have a critical role in the assembly of the mature capsid shell.
MHR mutations and suppressors act upon an early event of CA assembly

The increased lag period of the F167Y and L171V proteins (Fig 3.4A-B) and the ability of these mutants to be incorporated into assembling structures in the presence of WT protein (Fig 3.7A-D and 3.8) led us to hypothesize that the F167 and L171 residues are involved in an early step of capsid assembly. Also consistent with this hypothesis was the improved ability to initiate capsid assembly by three second-site substitutions (A38V, R185W and I190V) that restored infectivity to MHR mutant viruses secondary mutations (Fig 3.4A-B and Fig 3.5B and E) (48, 283). Thus, the stimulation of a nucleation event by these suppressors (Fig 3.4A-C and 3.9) provides a plausible explanation for their ability to suppress the F167Y and L171V mutations and suggests that a block in nucleation is the primary reason for the loss of infectivity observed for these mutant viruses. The lethal F167Y phenotype can also be suppressed by an additional substitution, located in the C-terminal cleavage site of CA, that alters the rate of release of mature CA from Gag (48) further suggesting a role of the MHR in virus maturation.

The ability of the A38V mutation in the NTD to efficiently suppress the assembly defect caused by both the L171V and F167Y mutations in the CTD suggests that an NTD-CTD interaction is essential for the initiation of capsid assembly and may be the oligomerization interface formed in the dimer seeds (Chapter II). The involvement of the NTD-CTD interface in nucleation of capsid assembly is further supported by the isolation of another NTD suppressor (P65Q) of lethal MHR substitutions. The homologous residue of P65 in HIV is located in the NTD-CTD interacting region mapped by in vitro assembly of HIV CA (163, 165, 256, 452), further suggesting that the interdomain interaction forms during the initial steps of capsid assembly.
If the residues of the MHR are involved in an intermolecular interaction in the immature particle but participate in only intramolecular interactions in the mature virion, the CTD must be a structurally dynamic protein (67, 163, 204, 205, 228, 234, 276). Although the flexibility of CA is well documented, little is known about the role of the MHR in determining the structural dynamics of CA. The domain-swap model which predicts that the conserved hydrophobic residues (in RSV: F167, L171, and the neighboring F164) undergo a refolding event during maturation fits the data showing that the hydrophobic MHR mutations disrupt an early event of capsid assembly. Thus, the F167Y, L171V and D155Y MHR mutations are probably altering the ability of the CTD to adopt one or more conformations essential for initiation of capsid formation.

By extension, the suppressors (A38V, I190V and R185W) may correct the defect in assembly by causing an increased propensity of CA to assemble by stabilizing CA in a pro-assembly conformation. The particular locations of the suppressors in CA are consistent with this scenario. The A38 residue lies on the surface of the second helix of the NTD in a location where it may modulate the NTD conformation and have an influence on hexamer formation and/or the NTD-CTD interface. Further support for A38 control of NTD folding is evident by the A38V suppressor acting in a temperature-sensitive fashion (49).

The suppressing mutation at I190 lies on a slight bulge, a conserved feature in the middle of the dimerization helix (67, 228, 234), that is predicted to form upon processing of CA from Gag (205). Although I190 is not directly involved in the dimer interface (19), the I190V substitution may conceivably alter the conformation of the helix, and thus affect CTD dimerization and/or NTD-CTD interaction. The identification of multiple mutations with suppressor capabilities near I190 provides further evidence that control of the conformation of this helix is vital to proper capsid assembly (48, 283, 431).
The R185 residue is located near the “top” of the dimerization helix in a position analogous to a Lys residue which is involved in the NTD-CTD interface in HIV (19, 234, 256). Thus, the R185W suppressor substitution may influence both the CTD-CTD and NTD-CTD interfaces. The increased resistance of the R185W protein to denaturation (Fig. 3.2B and Table 3.1) is consistent with the suggestion that this substitution is acting by stabilizing the CTD domain in an assembly-competent fashion. In the domain-swap model, the R185 is located near a bend between the first two helices of the CTD which has been proposed to undergo a transition during maturation (205, 236), further suggesting that R185W contributes to maturation.

Further support for the plasticity (structural flexibility) of CA is demonstrated by multiple crystal structures each containing differing arrangements of intersubunit contacts (37, 160, 210, 462, 462, 504). Indications of the dynamic behavior of the CTD comes from the high degree of flexibility in the second helix of this domain in an HIV protein that contained an engineered mutation that prevented dimerization (6, 503). During maturation a refolding of the first ~20 residues of the NTD further illustrates the dynamic nature of CA (47, 171, 327, 458, 476). The inherent plasticity of the NTD and CTD is likely due to the various structural roles of CA during multiple stages of the viral replication cycle.

Flexibility in either the structure of the CA monomer or in the intermolecular interactions upon multimerization is needed to build retroviral capsid shells using fullerene principles (33, 162, 164, 276, 332). Therefore, if the MHR mutations and suppressors alter the range of conformations available to the CA monomer, effects on morphology are predicted. Indeed, many of the double mutant and suppressor mutant proteins had a limited range of morphologies, e.g., tubular structures formed by L171V/A38V double mutant or spheroids by the I190V protein.

The tubular structures of L171V/A38V suggest that this mutant preferentially assembles hexamers rather than pentamer. Since NTD-NTD interactions are predicted to
form both the hexamer and pentamer, this shift towards hexamers may be due to an alteration in the NTD-NTD interaction caused by the Ala to Val mutation. However, structures similar to WT were assembled by A38V and the F167Y/A38V mutant proteins suggesting that both hexamers and pentamers are formed. Thus the hexamer formation, and the apparent lack of pentamers, in L171V/A38V protein is likely due to both the A38V and the L171V mutations suggesting that the integrity of the CTD may influence pentamer formation.

The effects of MHR and suppressor mutations on CA structural flexibility, the initiation of capsid assembly, and the morphology of the final products formed in vitro support the idea that the shape of retroviral capsids is determined in part by the flexibility of CA subunits and initiation of capsid assembly (58).

Implications for the development of drug resistance to antiviral therapies

The results from the MHR and suppressor mutations emphasize how genetic alteration of CA can drastically influence protein multimerization and alter virus infectivity. The lethal CA mutations in this study provide valuable tools for probing the structural transformations involved in maturation. Equally important, an understanding of the mechanism whereby suppressor mutants mask a deficit in capsid assembly is likely to have important implications for drug resistance to assembly inhibitors. Two inhibitors, CA-I and CAP-1, block HIV-1 replication by binding to CA (226, 441, 457, 462). Although mutations in CA that alter the binding ability of the inhibitors may confer resistance, the results from the second-site suppressors suggest that drug resistance may also be achieved by mutations in other regions of CA that increase the assembly propensity of the protein.
ACKNOWLEDGMENTS

Carmen Butan, Alasdair Steven, Parvez Lokhandwala, Tam-Linh Nguyen, Rich Kingston, Judy Phillips and Volker Vogt are gratefully acknowledged for providing data prior to publication and for informative discussions. Special acknowledgement goes to J. Brad Bowzard for providing results about the R185W and F193L suppressors prior to publication. The following member of Penn State College of Medicine Core Research facilities provided technical assistance: Roland Myers (EM), Anne Stanley (MS), and Joe Bednarczyk (DNA sequence analysis). This work was funded by grant CA100322 from the National Institute of Health (R.C.C.) and by the Pennsylvania Tobacco Settlement Fund through the PA Department of Health (R.C.C., Michael Katzman, and Leslie Parent.).
CHAPTER IV

SPACER PEPTIDE FOLLOWING CA PARTICIPATES IN IN VITRO INITIATION OF CAPSID ASSEMBLY OF ROUS SARCOMA VIRUS BUT NOT HUMAN IMMUNODEFICIENCY VIRUS
ABSTRACT

During maturation of avian- and lenti-retroviruses the capsid protein (CA) is cleaved in multiple steps by the viral protease. Initially, CA and its C-terminal spacer peptide (SP) are cleaved out of the Gag polyprotein that establishes the immature particle as a CA-SP transient maturation intermediate. Although CA-SP is a relatively stable cleavage intermediate, remaining in the maturing particle hours after initiation of maturation, it is undetectable in fully mature virions. In avian retroviruses such as Rous sarcoma virus (RSV), cleavage of either 9 or 12 amino acids from CA-SP creates two mature CA species, named CA-S and CA, during the late stages of maturation. The reasons for the step-wise cleavage and for the presence of CA and CA-S in mature virions are unknown. However trimming of SP from CA is required to form infectious particles that have stable capsids suggesting CA-SP influences capsid assembly. To investigate the effect, if any, that CA-SP has on mature capsid assembly, the in vitro assembly of all three RSV CA species was examined. RSV CA-SP assembled much more quickly than either CA-S or CA suggesting that SP promotes in vitro capsid assembly. Mixing together the three proteins to recapitulate an environment of the maturing virion, demonstrated that CA-SP stimulates the initiation of assembly of CA and CA-S. However, purified lentiviral human immunodeficiency virus (HIV) CA-SP1 protein failed to show a similar stimulation of capsid assembly. The results obtained from the RSV protein suggest that CA-SP, possibly by cooperating with other features of the maturing virus, participates in the initiation of capsid assembly in virus particles.
INTRODUCTION

Retroviral Gag polyproteins contain at least three common domains: MA (matrix), CA (capsid) and NC (nucleocapsid). Gag of some retroviruses contains intervening residues that form a short, ~10-15 residues, SP (spacer peptide) domain between the C-terminus of CA and the N-terminus of NC. An immature retroviral particle contains Gag hexamers organized in a partially closed lattice (51, 505, 518). After release of an immature particle and during maturation, a viral protease (PR) cleaves Gag to form the mature proteins (MA, CA, NC), a requisite step to infectivity. Early in the maturation process of retroviruses, CA and the C-terminal SP are cleaved out of the Gag (52, 53, 55, 496, 505, 518). Although only a transient intermediate, the resulting CA-SP species is a relatively stable maturation intermediate, i.e. it accumulates shortly after initiation of maturation and remains in the maturing virion for hours as SP is slowly cleaved from CA (95, 250, 355, 465). Prior to completion of maturation, SP must be removed from CA to generate an infectious virion (95, 180, 248, 251, 313, 494).

In RSV, proteolytic processing of CA is a multistep event involving overlapping cleavage recognition sites within SP (Fig 4.1). Removal of 9 residues from the C-terminus of CA-SP results in the formation of CA-S (ending at residue 240), while the cleavage of 12 residues leads to CA (ending at 237). Thus, RSV virions contain two mature forms of capsid protein (CA and CA-S) that are speculated to co-assemble during capsid assembly (355, 508). Pulse-chase experiments show that CA-SP is cleaved into CA and CA-S as early as one hour post budding but may require 5-6 hours for complete cleavage (34, 95).

Deletion of all or part of SP in RSV results in non-infectious particles that exhibit increased size heterogeneity (95, 251, 355). The capsids of these non-infectious mutant particles were found to be less stable under non-ionic detergent treatment, suggesting SP is required to form a proper capsid shell (95, 355). Blocking the removal of SP from CA-SP by genetic alterations to the cleavage-site also results in the loss of infectivity (508). The
Fig 4.1. Schematic summary of CA-SP cleavage during maturation. (a) The domains of RSV Gag are shown with cleavage sites identified by the arrows. A multistep sequence of cleavage occurs following release of the immature particle starting with the formation of a transient CA-SP intermediate. During the next ~6 hrs either 9 or 12 residues are removed by proteolytic processing at the C-terminus of CA-SP, resulting in two mature forms of capsid protein: CA-S and CA, respectively. In the mature virus 60% of the total protein is CA and the remaining form is CA-S. CA-SP is not found in the fully mature virion. (b) Amino acid sequence of C-terminus of CA and the SP domain RSV and HIV. Again the cleavage sites are indicated by arrows.
a

\[ \text{CA}\rightarrow\text{PR cleavage}\rightarrow\text{CA-SP}\rightarrow\text{PR cleavage}\rightarrow\text{CA-S}\rightarrow\text{CA}\]

\text{CA-SP (Concurrent w/ budding)}

\text{CA-S (40%)}

\text{CA (60%)}

(1-6 hrs post budding)

b

RSV: ...GIA AAM SSAIQPLIM

HIV: ...RVL AEAMSQVTNPATM
phenotypes of mutants lacking SP imply that the accumulation of the transient CA-SP intermediate is an essential step for proper assembly of the capsid. Furthermore, cleavage-site mutants that blocked the removal of SP from CA-SP demonstrate that the trimming of SP from CA-SP is required for proper mature capsid assembly and infectivity.

Structural models for retroviral capsids were derived from *in vitro* assembly of purified CA from HIV, RSV, murine leukemia virus, and equine infectious anemia virus (37, 68, 160, 162-164, 171, 210, 256-258, 276, 319, 504). In each model CA-CA interactions in the capsid are located away from the C-termini of CA. Thus any role that the residues at the end of CA including those in SP play in capsid assembly cannot be explained by any current model. Although structural models of capsids fail to explain a role of CA-SP during maturation, a previous genetic study provides support for the proposal that CA-SP influences the initiation of capsid assembly. Specifically, a second-site suppressor of the lethal F167Y mutation in the MHR (major homology region) is suppressed by a S241L substitution in SP (48). The S241L substitution increased the rate by which SP is proteolytically cleaved from CA, suggesting that this mutant is exerting its suppressor-phenotype during maturation. Furthermore, the F167Y mutation causes a defect in initiation of capsid assembly (Chapter III) leading to the hypothesis that SP influences the nucleation event of capsid assembly.

The function of CA-SP during maturation, specifically in proper capsid assembly, is poorly understood due, in large part, to the inability to study the complex processes occurring in a maturing retroviral virion. Since CA-SP is a relatively stable intermediate and co-exists with the mature form(s) of CA during maturation the ability of CA-SP to homoligomerize or to co-assemble with CA-S and CA may provide a snapshot into what occurs during the complex and lengthy process of maturation. The *in vitro* assay utilized in chapters II-III to dissect the early events of RSV capsid assembly provides for a tool to study how the different species of capsid proteins may interact. Similar to CA, the assembly of CA-S and CA-SP of RSV was induced to assemble by sodium phosphate. Furthermore, structures resembling
authentic viral capsids were observed by co-assembly of the two mature forms of capsid proteins (CA and CA-S), supporting previous speculation that both proteins are found in the mature lattice (355, 508). The results showed that CA-SP of RSV, but not CA-SP1 of HIV, assembled with kinetics greater than CA. When co-assembled with CA and CA-S, CA-SP stimulated nucleation of in vitro assembly. The data support a model, at least for RSV, where assembly of the capsid is initiated by CA-SP and growth of the capsid occurs by co-assembly of CA with CA-S.

**MATERIAL AND METHODS**

*Protein expression and purification*

RSV and HIV CA proteins were expressed in *E. coli* using previously described plasmids (373, 521) (Chapter II). RSV CA-S and CA-SP were cloned into the pET-24(+) plasmid by using NdeI and BamHI (CA-S) or HindIII (CA-SP) restriction sites. All RSV sequence is from the Prague C strain. A plasmid for expressing HIV CA-SP1 (p25) was generated by inserting the CA-SP1 sequence from the NL4-3 clone of HIV-1 into the pWISP98-85 CA (p24) expression plasmid using BamHI and NdeI sites. Both RSV and HIV were expressed using autoinduction and purified by ammonium sulfate precipitation, DEAE cation exchange, and SEC as previously described (372, 373) (Chapter II). The following extinction coefficient values were used: 24,980 M$^{-1}$cm$^{-1}$ (RSV CA, CA-S and CA-SP) and 33,585 M$^{-1}$cm$^{-1}$ (HIV CA and CA-SP1).

*In vitro assembly and data analysis*

The optical density at 450 nm (OD$_{450}$) of the assembly reaction was monitored by a spectrophotometer as previously described in chapter II (372, 373). Briefly, CA was stimulated to assemble by the addition of sodium phosphate pH 7.5 or 8.0 in 96-well UV
transparent plates. During assembly, the plate was mixed for 3 sec prior to each read to prevent precipitation of capsids. Assembly rates (ΔOD / Δ time) were calculated by fitting the growth phase of the sigmoidal curve of the assembly kinetics to a linear equation. The reaction order of the assembly was determined by the concentration dependency of the rate of assembly at protein concentrations ranging from 20 – 120 µM. CA assembly results in a highly heterogeneous mixture of structures and the amount of light scattered is limited by the shape of the structure formed (36), so to limit the effect of morphology of the assembled structures on the calculation the assembly rate for only the first 10 min following the lag phase, as determined by the Gompertz equation, was used (370, 548). All graphs were made and analyzed with Graphpad Prism.

*Electron Microscopy (EM)*

Sample and grid preparation for negative staining EM were previously described in chapters II and III. Typically uranyl acetate stain was used but occasionally nano-W (Nano-probes) provided better contrast. Samples were either visualized on a 120 kV Phillips transmission electron microscope (TEM) or a Joel JEM1400 TEM. Images from the Phillips TEM were capture on Kodak film, whereas digital images were capture from the Joel TEM using a CCD camera. In some cases the particles were heavily stained, thus some images were lighted using Adobe Photoshop.

**RESULTS**

*Native CA Proteins Migrate According to their Predicted Mass*

The recombinant RSV CA-SP protein purified from *E. coli* migrated with a smaller apparent mass than the smaller CA and CA-S proteins by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) (Fig 4.2a). The identity of each protein was
confirmed by mass spectrometry (data not shown). The anomalous migration pattern may be due to differences in SDS binding or an alteration in the structure of the protein. Under native (non-SDS denaturing) conditions the three CA variants migrated according to their predicted mass (Fig. 4.2b). Thus, the unusual migration of RSV CA-SP under denaturing conditions is likely due to differences in SDS-binding among the CA species. This anomalous migration pattern observed by E. coli produced proteins is similar to that of protein purified from RSV viral particles (34, 95, 355).

Similar to RSV, a spacer peptide of 14 residues (SP1, also previously termed p2) separates the CA and NC domains in Gag of HIV (Fig. 4.1b). Like RSV, deletion of SP1 or disruption of its cleavage from CA, results in non-infectious particles with aberrant cores (1, 248, 275, 355, 494, 508, 508). The HIV CA and CA-SP1 proteins migrate as expected based on their masses under both denaturing (Fig 4.2a) and non-denaturing conditions (data not shown).

In vitro Assembly of RSV CA-S and Co-assembly of CA and CA-S

The ability of CA-S to assemble was examined using a well characterized in vitro system in which multimerization of CA is induced by sodium phosphate (Chapters II and III). The kinetics of the assembly reaction is that of a typical nucleation-dependent pathway, with an observable lag phase prior to assembly (Fig. 4.3a). The kinetics of CA-S show a similar, but somewhat shorter, lag phase than CA (Fig 4.3a). The turbid samples of CA-S contained structures similar to those previously described to be formed by CA (Fig 4.4a-d).

On average, ~60% of the total capsid protein in a mature RSV virion is CA and the remaining ~40% is CA-S (355). Single particle analysis of authentic mature cores of RSV reported that well formed capsids incorporate ~80% of the total capsid protein (58), implying that CA-S is incorporated into authentic capsids. To test if CA and CA-S can co-assemble,
Fig 4.2. Migration patterns purified protein. (a) Denatured RSV and HIV proteins analyzed by SDS-PAGE using a 15% gel. (b) Migration pattern of native RSV proteins by SEC using a preparative grade Superdex 75 column.
the two proteins were mixed at a ratio of 60:40 of CA to CA-S prior to the addition of sodium phosphate. The mixed in vitro assembly reaction proceeds slightly faster than reactions that contain either CA or CA-S alone, at the same protein concentration (Fig 4.3b), suggesting that CA and CA-S can co-assemble and that CA-S may slightly stimulate initiation of capsid assembly. The final turbidity of the mixed reaction was similar to the levels obtained in the reactions that contained either CA or CA-S alone further supporting the idea that the proteins can co-assemble.

The co-assembled CA and CA-S samples contained tubes and small spheres that were easily observed (Fig 4.6). However the larger angular and multilamellar structures were clumped together and appear very dark (Fig 4.6b-e). The clumping of these structures may be an artifact that occurred during the preparation of the EM grid. Nonetheless, the in vitro results show that the two proteins can co-assemble consistent with the previous speculation that CA-S participates in capsid assembly in situ (355, 508).

**RSV SP Alters in vitro Assembly**

Similar to CA and CA-S, the longer CA-SP protein assembled in sodium phosphate conditions. However, CA-SP rapidly assembled, eliminating a discernable lag phase (Fig 4.3a), suggesting CA-SP is primed for assembly. Tubular, small spherical and capsid-like structures that have distinct angular bends similar to those formed by CA were observed for CA-SP samples (Fig 4.4e-i). CA-SP also formed two additional classes of multilamellar particles that appear larger (~ 40 and ≥ 90 nm) than the previously described 20 nm spherical particles and more round than the capsid-like structures formed by CA (Fig 4.4j-n). Only a small number of these relatively atypical particles were observed. Furthermore, small spheres were associated with tubular structures and, within the limitations of negative staining,
appeared to line the interior of the tubes (Fig 4.4g-i). Similar tubes that contained nested spheres were observed for CA but only very rarely (data not shown).

At near neutral pH the assembly of RSV CA under sodium chloride is non-efficient (Chapter II). However, in high concentrations (2.0 M) of sodium chloride a detectable amount of turbidity was observed for CA-SP but not CA (Fig 4.3c). The turbid material formed in sodium chloride has yet to be examined by EM, but the decrease in phosphate-dependency required to induce assembly is consistent with SP stimulating the initiation of capsid assembly. Other viral factors such as RNA or inositol hexaphosphate, as discussed in chapter II, may similarly stimulate or alter assembly of CA-SP but not CA. Indeed, an initial examination by EM showed that IP₆ may significantly enhance tube formation by CA-SP (Fig. 4.4o).

**In vitro Assembly of HIV CA-SP1**

In contrast to RSV CA and CA-SP, when the kinetics of the assembly of HIV CA was compared to that of CA-SP1, no appreciable differences were noted (Fig 4.5a). Both proteins had similar lag times and reached similar levels of turbidity. A more extensive comparison showed that their assembly rates were equivalent at various concentrations of sodium phosphate (Fig 4.5b). HIV CA forms predominately tubes *in vitro* under sodium phosphate and sodium chloride conditions (164, 179, 276, 372, 476) (Chapter II). CA-SP1 also formed long tubes (Fig 4.5c-h); however, most of the tubes were not as distinctly stained (Fig 4.5d-h). This observation is consistent with a previous study that showed purified CA-SP1 forming tubes under 1.0 M sodium chloride conditions (179). Fewer tubes were observed by EM in CA-SP1 samples compared to CA samples assembled under identical conditions, suggesting the CA-SP1 structures may be less stable. Short bullet-shaped structures were observed in one experiment (Fig 4.5g). These structures have not been previously observed to
Fig 4.3. Kinetics of RSV CA, CA-S and CA-SP assembly. (a) RSV protein at 80 µM was induced to assemble by the addition 500 mM sodium phosphate pH 7.5. (b) CA mixed with CA-S at the 60 to 40 ratio found in the mature virion. The total protein concentration was 120 µM and assembly was induced by 500 mM sodium phosphate pH 7.5. (c) Assembly of RSV CA and CA-SP protein at 120 µM in 2.0 M NaCl.
Fig 4.4. EM analysis of RSV protein assembled by sodium phosphate. (a-d) RSV CA-S assembled by 500 mM sodium phosphate pH 8.0 and examined by EM as described in the Materials and Methods section. (e-n) RSV CA-SP formed under similar conditions as part a-d. (o). RSV CA-SP at 180 µM was assembled by 300 mM sodium phosphate pH 8.0 and 25 µM IP₆ prior examination by EM. All samples except for d and o were stained with uranyl acetate. Samples d and o were stained with nano-W. All scale bars represents 100 nm.
Fig 4.5. Analysis of HIV CA and CA-SP1 multimerization. (a) Turbidity profile of 80 μM HIV CA and CA-SP1 assembled by 250 mM sodium phosphate pH 8.0. (b) Sensitivity of HIV proteins to sodium phosphate pH 8.0. The assembly rate was determined as described in Fig 3.9. (c-h) EM analysis of HIV CA-SP1 protein assembled by sodium phosphate pH 8.0 at either 250 mM (c and h) or 200 mM (d-g). All scale bars represents 100 nm except in h where it represents 1000 nm.
Fig 4.6. Visualization of CA and CA-S co-assembled structures. (a-i) CA and CA-S proteins were mixed at a ratio of 60:40 prior to the addition of 500 mM sodium phosphate pH 8.0. Samples were stained with uranyl acetate as described in the Materials and Methods. All scale bars represent 100 nm.
be formed by HIV CA, suggesting that SP1 may in rare cases alter the overall morphology of the capsid shell.

*Initiation of in vitro Assembly is Increased by RSV SP*

The above results from *in vitro* assembly suggest that CA-SP of RSV, but not HIV, stimulates initiation of capsid assembly *in vitro*. To further test the how nucleation of RSV CA-SP may differ from CA the dependency of the kinetics of assembly on protein-concentration was examined. The rapid assembly kinetics observed by CA-SP was slowed when the protein concentration was reduced (Fig 4.7b). The decrease observed is similar to that observed in the two mature forms of capsid protein (Fig 2.2a and 4.7a). This observation is consistent with assembly of CA-S and CA-SP following a nucleation-driven mechanism similar to CA. An estimate for the nucleus size can be determined by calculating the reaction order from the concentration dependency of the assembly kinetics (130, 340, 370, 545, 547). The rate of the assembly reaction of CA and CA-S was third order, suggesting the rate-limiting step involves a CA trimer (Fig 4.7c). Notably, the CA-SP slope was considerably smaller, suggesting that the initiation of CA-SP assembly, under the conditions tested, may not be following the same pathway as CA and CA-S.

*CA-SP can Nucleate Mature CA and CA-S Assembly*

Based on the increase in assembly kinetics by CA-SP, it was hypothesized that assembly of capsids may be nucleated by CA-SP. To test if CA-SP could initiate CA assembly *in vitro*, the two proteins were mixed together at various ratios, while keeping the total protein concentration constant, prior to addition of sodium phosphate. Consistent with the above hypothesis, adding CA-SP shortened the lag phase of CA assembly (Fig 4.8a). Stimulation of CA assembly was noted for concentration of CA-SP as low as 5% of the total
Fig 4.7. Dependence of the rate of assembly on protein concentration. (a-b) Turbidimetric profile of RSV CA-S (a) and CA-SP (b) proteins ranging from 20 to 100 µM. The samples were induced to assemble by addition of 500 mM sodium phosphate pH 8.0. (c) The ln rate of assembly vs. ln CA concentrations is plotted (points) as described in the Materials and Methods section. The linear slope for each protein is represented by a dashed line and the value of the slope is given below the graph. Calculations for CA were based on the kinetics of assembly shown in Fig 2.2a.
protein. These results suggest that CA-SP homo-oligomerization and potentially hetero-oligomerization between CA-SP and CA may act as seeds assembly of capsids in vitro. Co-assembly of CA and CA-SP resulted in the formation of small spheres, tubes and capsid-like structures similar to those formed by CA (Fig 4.9). Similar to the tubes formed by CA-SP, many of the tubes observed by co-assembly of CA with CA-SP had small spheres associated with them (Fig 4.9b-d, and g). The results demonstrate that self-assembly of CA can be stimulated by CA-SP.

In an RSV particle, all three CA species are present at the same time during maturation (34, 355). Pulse-chase experiments showed that released virions contained similar ratio of all three CA species 3 hours after release from the cell (34). To examine if CA-SP can stimulate capsid assembly at a time point slightly earlier during maturation, the kinetics of co-assembly of 60 µM CA-SP with 30 µM of each CA and CA-S was examined. All three proteins were able to assemble as demonstrated by the final turbidity (Fig 4.8b). The assembly of the sample was robust with a short lag phase similar to sample that contained 120 µM of CA-SP alone. Similar results were found with other ratios of CA-SP to CA and CA-S. The rapidly assembling mixtures of all three CA species have yet to be examined by EM. However, varying the ratio of CA-SP to CA and CA-S in the in vitro assembly conditions may allow for the examination of events that occur shortly after initiation of maturation from those that occur later.

DISCUSSION

The previous observation that a S241L substitution in SP of RSV suppressed a MHR mutation that is defective at initiation of capsid assembly led to the hypothesis that the RSV CA-SP influences the nucleation of assembly. Although CA-SP protein containing the S241L substitution has yet to be tested, the results presented in this chapter demonstrate that the
presence of WT SP at the C-termini of CA increased the kinetics of *in vitro* capsid assembly under sodium phosphate conditions. This observation is consistent with CA-SP stimulating the nucleation event of capsid assembly. Further suggesting that CA-SP is critical to maturation is the observation that formation of infectious particles with properly formed capsids require the transient CA-SP intermediate and the subsequent proteolytic processing of SP from CA-SP (1, 95, 248, 275, 355, 494).

Similar to RSV and HIV, many other retroviruses contain a SP between the C-termini of mature CA and the N-termini of NC. The length of SP ranges from 5 to 25 residues but a few retroviruses, such as murine leukemia virus (MLV), lack a distinct spacer between CA and NC. However, at least for MLV, genetic data suggest that the residues at the C-terminus of CA may function similarly to SP during immature assembly (85). The conserved presence of a spacer peptide between CA and NC among retroviruses may be due to both the necessity of SP to immature assembly and to maturation, including potentially forming an initiating-seed for capsid assembly.

Maturation is marked by a change in the morphology of the interior of the virus: the roughly uniform spherical interior of the immature particle is lost and an asymmetrical-shaped dense-core consisting of a capsid shell is formed. In both RSV and HIV immature particles SP regulates spherical morphology, i.e. deletions or insertions of and mutagenesis in SP result in tubular structures (168, 180, 215, 225). Of note, an examination by EM of the morphology of RSV particles that contained a partial deletion of SP (removal of the last 9 residues of CA-SP) failed to reveal any gross disruption of mature core structure (95). However, any subtle influence on the morphology of the capsid by SP can not be ruled out. To better understand the potential influence of the CA-SP intermediate on capsid morphology *in situ* a cryo-EM tomographic study of RSV particles that contain a full deletion of SP (all 12 residues) is needed. EM examination of HIV particles that had a block of
Fig 4.8. Co-assembly of RSV CA protein species. (a). CA and CA-SP were mixed together at several ratios prior to addition of 500 mM sodium phosphate pH 8.0. The total protein concentration for each sample was 120 µM. (b) All three species were mixed together at the list concentration and assembled under 500 mM sodium phosphate pH 7.5 conditions.
Fig 4.9. EM analysis of structures formed by co-assembly of RSV CA and CA-SP. (a-g) CA and CA-SP were mixed at a 6:1 ratio prior to stimulation of co-assembly by the addition of 500 mM sodium phosphate pH 8.0. Scale bars represent 100 nm.
SP1 cleavage from CA-SP1 (275) suggest that capsid assembly may be initiated but a defect post-initiation may prevent further growth of the capsid. A similar tomographic study of a RSV mutant that blocks SP cleavage from CA-SP may aid in elucidating the function of CA-SP during maturation.

**RSV CA-SP as a Nucleus Complex for Capsid Assembly**

CA-SP not only showed rapid assembly kinetics with a decrease in the lag phase (Fig 4.3a), but it also stimulated both CA and CA-S to assemble (Fig. 4.8). Preliminary experiments showing that the multimerization of CA-SP, unlike CA, is not dependent on multivalent anions is consistent with an increased ability of CA-SP to initiate assembly (Fig. 4.3a). Since CA-SP showed an increase in assembly kinetics, it is predicted that this protein assembles at a lower critical concentration than CA. Small oligomers similar to the dimer–tetramer complexes isolated during purification of CA are also formed by CA-SP, consistent with the hypothesis that CA-SP acts as a seed for capsid assembly (data not shown).

CA-SP formed capsid-like structures and other morphotypes similar to those formed by RSV CA and CA-S, suggesting the same CA inter-subunit interactions are formed by all three CA variants. Since the all three CA species form structures that resemble authentic capsids, it is hypothesized that both CA-S and CA-SP form pentamers and hexamers. The presence of SP at the C-termini of CA failed to drastically shift the distribution in morphology of structures formed by CA suggesting the mature CA-CA interfaces are favored in these in vitro conditions. A few novel structures were formed by RSV CA-SP but not CA or CA-S. The large roundish particles (Fig 4.4j-n) are similar in morphology to immature virions. In these rare particles CA-SP may have an altered subunit packing reflecting the CA-CA interactions in the immature lattice.
RSV SP and HIV SP1 are Structurally Dynamic Peptides

The mechanism by which SP stimulates in vitro assembly is unknown but may involve the structural nature of SP. Although RSV SP and HIV SP1 are highly flexible in solution, the secondary structures of both peptides are predicted to be helical based on the sequence of the amino acids (1). Electron cryo-tomography reconstructed images of immature HIV particles generated a model in which SP1 forms a six-helix bundle assembly domain (51, 505). Similar examination of immature RSV particles also suggests that a similar SP six-helix bundle assembly domain forms in RSV (J. A. Briggs, personal communication). Consistent with this model, SP1 shows a tendency to multimerize in solution (314). In the context of a purified polypeptide consisting of the C-terminal residues of CA, SP1, and NC of HIV, SP1 along with the 12 immediate upstream CA residues and downstream NC residues, was highly flexible with a slight propensity for α-helix formation (330). In a recombinant protein that included SP1 with the last 21 residues of CA and 13 downstream NC residues, the last seven residues of CA, all the residues of SP1, and the first two residues of NC adopted an α-helix form suggesting that the presumptive immature assembly domain consists of CA, SP1 and NC residues (314); however, this protein was placed in a solution containing 30% TFE (trifluoroethanol) which can promote helix formation (227). Thus, at least in Gag, SP likely adopts an α-helical secondary structure and forms an oligomerization domain. Although this model for the immature assembly domain has not been rigorously tested, genetic alterations that are predicted to disrupt the helical nature of SP results in a loss of infectivity and is consistent with the residues of SP adopting an α-helical structure (1, 278, 304).

The structure of SP may change as a result of proteolytic cleavage once CA-SP is liberated from Gag. After cleavage, in the CA-SP intermediate, the C-terminal residues of CA and the residues of SP are likely highly flexible with an inability to adopt any noticeable secondary structure (234, 504). In the monomeric CA-SP intermediate, the structure of CA
consists of two mostly helical domains (N-terminal [NTD] domain and C-terminal domain [CTD]) that are connected by a flexible linker (37, 66-68, 91, 92, 126, 159, 160, 163, 171, 210, 234, 319, 462, 504). The residues of SP are located away from the known mature CA-CA interfaces (37, 160, 163, 210, 256, 257, 276, 317, 319, 327) making it unclear how CA-SP may stimulate the initiation of in vitro RSV capsid assembly. A SP-SP multimerization interface may form by CA-SP under the sodium phosphate conditions examined in this study. Examining the in vitro CA-SP assemblies at a resolution high enough to observe electron-density of a potential SP-SP interface may provide a structural model for how CA-SP stimulates assembly. Alternatively, SP may exert an influence on in vitro assembly by some unidentified long range interaction on the known CA-CA interfaces by altering either the monomer or an early assembly intermediate.

**RSV Maturation and Early Steps of Capsid Assembly**

The molecular steps of maturation, including any influence by CA-SP, are poorly understood. Both the NTD and CTD of CA as well as the residues immediately preceding and following CA participate in formation of the immature Gag hexamer lattice. (51, 205, 225, 297, 505, 518) (J. A. Briggs, personal communication). Upon being liberated from the Gag polyprotein, CA undergoes conformational changes and self-assembles into the capsid shell (180). During early steps of maturation, proteolytic cleavage of the N-termini of CA likely disrupts the immature hexamers and results in a refolding event required for mature capsid assembly (67, 234, 476). Similarly, proteolysis at the SP—NC junction may promote mature assembly by further destabilizing the immature lattice. However many questions remain to be answered about the structural transformations occurring during maturation: Does SP influence the structural conformation of the CTD? Does SP influence the NTD? Does the final cleavage of SP from CA alter the structure of CA or the capsid shell? Why do mature
avian retroviruses contain both CA and CA-S? More experimental evidence will be required to answer such questions.

The results of this *in vitro* study suggest that some incompletely-digested Gag, that contains at least CA-SP, may stimulate the initiation of capsid assembly *in situ* and is consistent with previous proposals that suggest capsid assembly initiates at a specific location in the maturing virion and proceeds by multidirectional growth (21, 24, 33, 50, 58, 372). The presumptive initiating complex has not been identified but was suggested to contain uncleaved or partially-cleaved Gag (58) and a dimeric form of CA (Chapter II). The concentration-dependence of the reaction kinetics suggests that a trimer of CA is the critical nucleating species for CA assembly, but data presented in chapter II clearly demonstrated that a CA dimer is involved in the initiation of capsid assembly. Theoretical modeling of capsid assembly implies that the concentration dependency may overestimate the size of the nucleating complex, suggesting that an alternative method to observe the nucleus is needed (129, 130, 545). Unfortunately, the light scattering method used in this study to follow the reaction kinetics is poor at detecting small complexes and fails to directly monitor the formation of the nucleus. A method with greater sensitivity, such as using laser light or small-angle X-ray scattering, will be necessary to observe the small complexes that initially form following the addition of sodium phosphate. Determination of exact role of CA-SP during maturation will require further analysis. However, based on the results shown in this chapter the nucleating-seed of capsid assembly *in situ* may involve residues of SP. Overall, the data presented here and in chapter II suggest that the rate-limiting step involves an oligomeric complex of CA or CA-SP that is smaller than a pentameric or hexameric capsomer.

*Understanding Maturation in HIV and in vitro Assembly of HIV CA and CA-SP1*

Both HIV CA and CA-SP1 assemble long tubular structures under sodium chloride and sodium phosphate conditions (Fig. 4.3c) (179). Neither of these *in vitro* conditions
stimulates CA or CA-SP1 of HIV to form the cone-shaped capsids predominant in authentic virions. Purified HIV proteins favor the formation of hexameric-containing tubes (276), suggesting that in vitro assembly is unsuccessful at recapitulating a critical step in controlling morphology, specifically the formation of pentamers. Thus, the in vitro assembly of purified HIV protein is a poor model for in situ retroviral capsid assembly. The failure of CA-SP1 to stimulate HIV CA assembly in vitro may simply reflect that this is an inadequate method to examine capsid assembly; or alternatively, may reflect that CA-SP1 of HIV has a different role during maturation than CA-SP of RSV.

Specific blockage of the proteolytic processing of SP1 from the HIV CA-SP1 intermediate is a successful antiviral target and is the mechanism of action by berivimat, the only maturation inhibitor used to treat HIV-infected patients (3, 219, 275, 401, 424, 534-537). Particles that are treated with berivimat or that have a genetic alteration to prevent cleavage of SP1 from CA-SP1 lack normal cone-shaped capsids and infectivity. However an incompletely-formed shell that appears as a crescent-shaped proteinaceous layer near the viral membrane forms in HIV particles that contain CA-SP1, and no CA (1, 275). Some of the improperly formed cores partially encircle electron-dense core material, suggesting CA-SP1 can initiate assembly of a shell but fails to complete capsid assembly during maturation. This observation further supports the hypothesis that SP may be involved in the initiation of capsid assembly.

**Similarity with Capsid Maturation of Non-retroviruses**

The stepwise processing of retroviral capsid proteins is similar to the proteolytic processing observed in several other viruses. Cleavage of polio viral P1 polyprotein is required to form an infectious virion (13, 208). Assembly of infectious bursal disease viral capsids is controlled by cleavage of the VP2 coat protein at four C-terminal sites resulting in the loss the last α-helix (286). The α-helix acts as a transient molecular switch that defines
whether a pentamer or hexamer is formed. The retroviral SP region has also been suggested to act as a switch by controlling spherical and tubular morphology of both released viral particles and in vitro assembled structures (180, 215, 225). However, the residues of SP do not appear to control pentamer and hexamer formation since capsid-like structures that likely contain both capsomers were assembled by CA, CA-S and CA-SP.

Similar to SP, the C-terminal fragment of hepatitis B virus core protein is structurally dynamic. The hepatitis B viral fragment is found in two different states (open or closed) in the assembled capsid (200) and may regulate late steps of genome synthesis and viral release. SP of retroviruses may similarly be found in multiple different states. A better understanding of why RSV CA-SP has a higher propensity to assemble than CA awaits further biochemical and biophysical analyses including the influence of SP on the formation of the CA subunit interactions. The results in this chapter are consistent with CA-SP enhancing the formation or stability of the nucleating complex that initiates assembly of CA and CA-S in RSV virions.

Further test for in vitro Defining SP in assembly of CA-SP

Although the biochemical data presented in this chapter conclusively shows that in vitro RSV capsid assembly is stimulated by CA-SP, little data supports the contention that CA-SP is involved in the nucleation of capsid assembly in situ. The S241L suppressor mutant suggests that CA-SP can overcome a block of capsid assembly in situ. The S241L in vitro assembly is predicted to increase the propensity of CA-SP to assemble, similar to the previously described suppressors (Chapter III). The S241L mutation restores infectivity to the lethal F167Y mutation (48) and is expected to restore in vitro assembly to a F167Y CA-SP mutant protein. Furthermore, S241L CA-SP is hypothesized to restore in vitro assembly to L171V mutant protein (similar to I190V and A38V, Chapter III ((283)). Kinetic and EM analysis of the S241L and double
(F167Y/S241L, L171V/S241L) mutant protein assembly by sodium phosphate is needed. A comparison of the S241L double mutants (F167Y/S241L, L171V/S241L) with CA-SP protein that contains either the F167Y or the L171V may elucidate the mechanism by which S241L is suppressing and the role of the CA-SP intermediate during maturation. Furthermore the in vitro assembly defect caused by the D52A mutant, that lacks the β-hairpin, may be suppressed by SP (297) providing further support that an immature CA-SP intermediate is involved in the initiation of capsid assembly (discussed further in Chapter VI). Purification and analysis of D52A CA-SP is also needed.

ACKNOWLEDGEMENTS

We are grateful to Mike Summers for providing pNL4-3 DNA and to Roland Myers for assisting with EM data collection. We also would like to acknowledge Kristen Rennoll-Bankert for her help in constructing the plasmids to express RSV CA-S and CA-SP and for preliminary protein characterizations.
CHAPTER V

A MISSING LINK IN RETROVIRAL CAPSIDS:

VISUALIZING THE RSV CA PENTAMER

Parts of this chapter were adapted from:
Giovanni Cardone, John G. Purdy, Naiqian Cheng, Rebecca C. Craven, and
Alasdair C. Steven

In this chapter the Materials and Methods, Results and Discussion are combined in a single
section.
Abstract

Mature retroviruses contain a structural shell, called the capsid, which encloses the viral RNA genome, the virally encoded RNA-binding protein NC and replicative enzymes such as reverse transcriptase. The capsid assembles during the process of viral maturation and is composed of the viral CA (capsid) protein. CA contains two folded domains, termed the N-terminal domain (NTD) and the C-terminal domain (CTD), that are connected by a short flexible linker of ~5 residues. Both intradomain (NTD-NTD or CTD-CTD) and interdomain (NTD-CTD) interactions are required to form the capsid. Native structures of mature human immunodeficiency virus (HIV) and Rous sarcoma virus (RSV) capsids are highly polymorphic. HIV capsids predominantly assembled cone-shaped structures, whereas RSV CA tends to assemble irregular polyhedra. The morphology of HIV and RSV capsids can be modeled as a fullerene lattice with a variable number of CA hexamers and 12 CA pentamers. In vitro assembled retroviral hexamers have been studied by x-ray crystallography and by electron microscopy, but the predicted pentamers have not been observed. Analysis by cryo-electron microscopy reveals that in vitro assembled RSV CA forms two classes of icosahedrons, each of which contains 12 pentamers. This chapter contains the first description of the pentameric CA complexes and a pseudo-atomic model for the intradomain and interdomain interactions found in capsids of RSV.
Introduction

Retroviral capsids show dramatic polymorphism, that is to say, that the morphology of each individual capsid appears to be unique. The polymorphic nature prevents their study using techniques based on averaging of many particles or applying icosahedral symmetry that allowed the structural determination of many other viral capsids at an atomic or near-atomic level (<5 Å) by crystallography or three-dimensional cryo-electron microscopy (cryo-EM) reconstruction (391-393, 438). However, single particle analyses of retroviral virions by cryo-electron tomography have achieved moderate levels of resolution (~50 Å) that describe the overall shape of each capsid but fail to observe individual capsomers or subunits (33, 50, 58).

The capsids of retroviruses, including Rous sarcoma virus (RSV), are best described as polyhedrons. Lentiviruses such as human immunodeficiency virus (HIV) have cone-shaped polyhedral capsids. The capsids of both RSV and HIV have been modeled as a lattice shell that contains hexamers and pentamers similar to fullerene polyhedra and cones whose structures have been observed extensively in carbon chemistry (147, 162, 164, 198). Closed fullerene polyhedra and cones require 12 pentons (pentameric declinations) to induce curvature into the hexagonal lattice. Accordingly, the two ends of the HIV conical capsid are modeled as having five pentamers at the narrow end and seven pentamers at the wide end (33, 50, 162, 164, 210). The polyhedral shapes observed in RSV can also be modeled by placing the pentamers more randomly dispersed throughout the hexameric lattice (164, 198). The fullerene cone model correctly predicts the observed cone angles of authentic HIV capsids and those formed by in vitro assembled recombinant protein (33, 50, 162, 164). The subunit building blocks of capsids, the CA protein, of HIV and murine leukemia virus form hexamers under in vitro assembly conditions (163, 276, 317, 319, 363). Furthermore, high resolution x-ray crystallography and nuclear magnetic resonance (NMR) reveal that CA has two domains, the N-terminal NTD and the C-terminal CTD connected by a short flexible linker (67, 92, 159, 160, 210, 228, 234, 317-319, 504, 518). Some of these studies have provided views of
how the CA subunits interact in hexameric lattices. However, no retroviral pentamers have been visualized.

The limited resolutions of tomographic studies prevent atomic level visualizations of authentic capsids of retroviral virions and thus limit the ability to visualize pentamers. Instead, pseudo-atomic molecular models for the architecture of HIV capsids were generated by placing (fitting) high-resolution structures the CA subunit into the lower resolution density maps generated by three-dimensional reconstructions from intact structures formed \textit{in vitro} (163, 276). The pseudo-atomic model of viral capsids shows that a hexagonal lattice is formed by NTD-NTD interactions via the first three helices of CA. Each hexameric ring is connected by CTD-CTD dimerization along its second helix. A third interface, the NTD-CTD interaction, provides intermolecular contact between the two domains. The current model provides information about the hexameric ring and hexamer-hexamer interaction but provides no information regarding the structure of a pentamer or of pentamer-hexamer interaction.

The lack of experimental evidence of CA pentamers has left some to question the existence of the pentamer and the fullerene model of retroviral capsid morphology proposed by Ganser \textit{et al.} (162). If CA pentamers do exist, are the pentameric rings formed by NTD-NTD interactions similar to the hexameric rings? Are pentamers connected to hexamers via the CTD dimerization domain? Do the pentamers contain the NTD-CTD interaction? Visualization of \textit{in vitro} formed CA pentamers may provide the answer to some of these questions.

\textit{In vitro} assembly of RSV CA produces a diversity of structures including structures with morphology similar to authentic capsids (fig 2.1 and 3.3) suggesting the formation of hexamers and pentamers. Numerous structures were small spheroidal particles less than 20 nm in diameter that appeared isometric by negative staining transmission EM (fig 2.1, 2.2e and 3.5c). These spheroidal assemblies are hypothesized, originally by A. C. Steven, to be small icosahedral particles that may be well-suited for three-dimensional cryo-EM
reconstruction. In accordance with this hypothesis the particles should contain 60 (or a multiple of 60) subunits and have a 5-fold axis of symmetry.

Materials & Methods, Results and Discussion

Isolation and characterization of a 60mer CA structure

RSV CA protein was purified by ion exchange and size exclusion chromatography (SEC) as described in chapter II. In addition to the high molecular weight (HMW) peaks identified by SEC (Fig. 2.4a), some preparations of WT protein and all preps of the I190V suppressor mutant protein yielded an additional very large peak of protein eluting at 20 mL on a preparative-grade S75 column (Fig. 5.1). Using SEC coupled with multi-angle light scattering (SEC-MALS) the peak near 20 mL was determined to have a mass that is ~1.3 MD or a ~55-60mer. Unlike the smaller dimers to tetramers previously isolated (Fig. 2.4a), the I190V 55-60mer failed to act as an assembly seed (data not shown) suggesting that the 55-60mer is a structure that has finalized its assembly (i.e. fully closed). When protein from the fraction that eluted at 20 mL was examined by EM, a large number of small isometric ~ 20 nm particles similar to those formed under sodium phosphate conditions were observed. If the particles contain 60 CA subunits and are fully closed they are likely arranged with icosahedral symmetry.

Icosahedral symmetry of RSV 60mer particles

Multiple preparations of both WT and I190V proteins assembled under various concentrations of sodium phosphate (150 mM to 500 mM) at either pH 7.5 or 8.0 were mailed to the Steven lab for analysis. Prior to examination the samples were briefly centrifuged at 1,000 x g at 4 °C to remove the larger tubular and polyhedral structures. The
Fig. 5.1 Size exclusion chromatography (SEC) of I190V protein. WT and I190V mutant proteins were purified in parallel and examined on a SEC S75 preparative grade column. The large oligomers of I190V CA elute around 20 mL (gray line).
Fig. 5.2. RSV CA isometric particles. (a). Cryo-EM analysis of particles formed by 500 mM sodium phosphate. The micrograph was supplied by N. Cheng and A. Steven. Two slightly larger particles are circled (b) Image reconstruction reveals 2-, 3- and 5- fold symmetry consistent with icosahedrons. The pentamer is highlighted on the T=1 structure. (c) A small number of T=3 structures (circled in part a) were used for reconstruction. The hexamer is outlined. Nested within the T=3 shell was a T=1 particle. Part b and c were created by Giovanni Cardone (68).
remaining smaller particles were centrifuged again at ~3,000 x g at 4 °C to concentrate the soluble particles remaining after the first spin. The small ~ 20 nm particles were examined by cryo-EM using a CM200-FEG EM (FEI) with a liquid nitrogen cooling stage as described by Cardone et al. (68). The isometric appearance of the spheroidal particles was confirmed using cryo-EM procedures that do not dry or stain the samples (Fig. 5.2a). Most particles are ~ 17 nm spheres but a few larger spheres (~ 30 nm) were observed. The images of the particles were classified according to size and orientation in which they were embedded in the vitreous ice layer, and analyzed as previously described (68). Averaging and reconstruction of > 2,800 smaller ~ 17 nm spheres showed 2-, 3-, and 5-fold symmetry, consistent with icosahedrons (Fig. 5.2b). The smaller ~ 17 nm spheres are T=1 particles, as originally surmised by Alasdair Steven. The T=1 spheres contain 60 subunits as predicted based on the SEC-MALS results that are arranged in 12 pentameric capsomers.

Pseudo-atomic model of the RSV CA Pentamer

The resolution (~ 10 Å) of the reconstruction was high enough to visualize the 12 pentavalent capsomers but was too low to visualize individual domains and helices of the CA subunits. However, modeling the structures of the domains of CA into the cryo-EM three-dimensional density map, as described in Cardone et al. (68), provides the first pseudo-atomic model for CA subunit interactions in the RSV capsid. The NTD crystal structure (PDB 1em9, (234)) and CTD NMR structure (PDB 1d1d, (67)) correlated best with the cryo-EM density map. Each domain was fitted independently, using an automated rigid-body fitting program (i.e. without allowing permutations to the original crystal or NMR structures), with limited areas where the structure of the NTD or the CTD did not match the cryo-EM density. In the model, all of the α-helices mapped to tubular densities in the cryo-EM
reconstruction (Fig. 2c in (68)). The positions of each domain in the density map were unambiguously identified without overlapping or sterically hindering each other.

Each of the 12 vertices contains five lobular domains extending away from the center (Fig. 5.2). The finger-like protrusions on the surface of the spheres correlates with the NTD. The interior surface of the protein shell is made by the CTD. In each pentamer the CTD of one subunit is located underneath the NTD of the neighboring subunit (Fig. 5.3 and 5.4). The T=1 particles are formed by three CA-CA interfaces, each of which is predicted to occur in the capsid (162, 163, 198, 256, 276).

According to the pseudo-atomic model, the pentameric interface is formed by NTD-NTD interactions (Fig. 5.5a). Helices 1 and 2 of a CA subunit interact with helices 1 and 3 of the neighboring subunit NTD. The 12 pentamers in a particle are connected by CTD-CTD dimerization involving the second CTD helix (Fig. 5.5c and 5.6). In the third interface, the bottom (N-terminus) of helix 4 of the NTD contacts helices 1 and 4 of the CTD (Fig. 5.5b). This interaction occurs in an almost perpendicular fashion between neighboring subunits, thus the NTD-CTD interaction is predominately intermolecular. However, the C-terminus of CTD helix 1 approaches the loop that connects NTD helices 1 and 2 (Fig. 5.5b) and forms an intramolecular interdomain interaction. Thus, the CTD actually docks beneath the intersection of two NTDs. This interdomain interaction and closely neighboring NTD-NTD interaction can be considered an extensive tripartite interface.

The pseudo-atomic model is supported by the dependency of CA multimerization on multivalent anions (Chapter II and III). In the model several positive charges in the interfaces must be neutralized to enable assembly. Additionally, calculations in GRASP (333) showed that a large net positive charge covers most of the NTD and must be neutralized to allow pentamer formation (M. Bewley, personal communication). Phosphate ions likely neutralize specific positive charges in interfaces that form the pentamer and also likely neutralize the
Fig. 5.3. Pseudo-atomic model of the RSV CA pentamer. (a) Top and side views of the model for the retroviral pentamer. One subunit is colored red. Below the pentamer model is a cartoon representation of the subunit interactions. (b) A top view of the pentamer model where the subunits are colored in a rainbow fashion, red to yellow (left). Also show is the model of the pentamer with the CA subunit secondary structure shown as ribbons. Helices 1 and 2 of one subunit interact with helix 3 of the neighboring subunit as shown.
Fig. 5.4. Ribbon diagram of RSV CA pentamer. Three views of the model for the retroviral pentamer: top (outside looking into the capsid), side view, and bottom (inside of the capsid looking out). The NTD and CTD of each monomer are colored the same. The model was made by fitting (docking) the structure of the NTD (PBD 1em9) and the CTD (PBD 1d1d) into the cryo-EM density map of the T=1 icosahedron as described in (68). The images were made using UCSF Chimera software.
surface of the NTD. Thus the pseudo-atomic model is consistent with the biochemistry of CA.

The biology of the virus provides further support for the pseudo-atomic pentamer model. The first helix forms part of the conserved major homology region (MHR) and part of the NTD-CTD interdomain interface. Consistent with the pentamer model, genetic alteration of the MHR results in the loss of infectivity (48, 61, 96, 283) and in vitro CA assembly (Chapter III). The location of the secondary-site suppressing mutations that restore infectivity and capsid assembly to the MHR mutations also provides biological support for the pseudo-atomic model. The suppressing mutations are located in the helices responsible for the CA subunit interactions in the pentamer suggesting modulation of these interfaces alters infectivity and capsid assembly. In particular, the R185W suppressor is found in a region near the CTD-CTD and NTD-CTD interfaces and may reduce the charge-repulsion by removing a positive charge.

**RSV Pentamer vs. HIV Hexamer**

Placing the NTD and CTD structures of CA into the reconstructed icosahedral particles provides the second pseudo-atomic model that visualized all three mature CA interfaces (NTD-NTD, NTD-CTD, and CTD-CTD). The first model was achieved by Ganser-Pornillos et al. using in vitro assembled R18L mutant HIV CA protein and provided a view of only the hexamer (Fig. 1.13) (163). Homologous regions of CA form both the HIV hexamer and the RSV pentamer supporting the conclusion that these models reflect the capsomers found in authentic in situ formed capsids.

Similar to the pentameric rings in the RSV model, the hexameric rings of HIV were formed by NTD-NTD interactions involving helices 1-3. The CTD dimers were mediated by the second CTD helix, similar to the crystal structure solved by Worthylake et al. (163, 504).
In HIV, the dimerization helices are parallel to each other, whereas in RSV the two CTDs form a ~ 45° angle. The interdomain interaction in the HIV hexamer involves helix 4 of the NTD inserting into a groove formed on the CTD surface. The groove is made by CTD helices 1 and 2 on one side and helices 3 and 4 on the other side. Unlike the RSV pentamer little to no intramolecular interaction was found between the NTD and the CTD in the HIV hexamer (163, 363).

Comparison of the pentamers to the hexamers underscores the conservation of the three mature CA-CA interactions. However, are the dissimilarities noted above due to differences between RSV and HIV? Or are they due to differences between pentamers and hexamers? To answer these questions, isolation of HIV CA isometric spheres similar to the RSV T=1 particles was attempted, by varying the conditions (types of salts, protein concentrations, pH and temperature) of assembly, but failed. However, moderate success was found in attempts to visualize mature RSV hexamers.

**RSV Hexamer**

Reconstructions of the larger ~ 30 nm spherical particles visualized by cryo-EM revealed T=3 icosahedrons, consisting of 12 pentamers and 20 hexamers (Fig. 5.2c). A focus was placed on a CA mutant protein, the I190V suppressor, since this mutant produced more of the large ~ 30 nm particles than WT when observed by transmission EM using uranyl acetate negative staining (Fig. 3.5c). However, only ~ 1% of the assembled I190V spheres were actually T=3 particles; the rest appear to be T=1 particles that likely flattened during preparation for transmission EM. The rarity of these particles (only 48 particles were averaged) limited the resolution of the reconstructed image to ~ 22 Å as opposed to the ~ 10 Å resolution achieved with the T=1 particles. In the reconstructed image for the T=3 particles, both pentavalent and hexavalent capsomers are visible but the
Fig. 5.5. Model for mature RSV CA interfaces. The three subunit interfaces found in the pentamer are shown with interacting regions labeled. For reference the β-hairpin is red in (a) and the second CTD helix is green in (c). See the legend for Fig. 5.3 for greater details on how the model and images were generated.
Fig. 5.6. Model for the interaction between capsomers in the icosahedrons. Each pentamer in the icosahedral particles is connected to a neighboring pentamer or hexamer via CTD-CTD dimerization. The capsomer-capsomer interaction is shown by two CA monomers, each from an adjacent capsomer. A cartoon representation of two monomers is also shown.
Fig. 5.7. Model for the mature RSV CA hexamer. A top and side view of the model for the six CA subunits in the mature hexamer. Each CA monomer is shown as a ribbon and one subunit is colored red. The first three helices of the NTD form an 18-helix bundle near the center of the hexamer. Each CTD is located beneath the NTD of the adjacent subunit.
individual helices and side-chains of the residues are not (Fig. 5.2c and 5.7). However, at ~
22 Å the cyro-EM density map of the reconstructed T=3 particle allowed the generation of a
pseudo-atomic model for the RSV hexamer (68). Overall, the interactions in the hexamer are
similar to those in the pentamer model (Fig. 5.7). Each pentamer is connected to a
neighboring pentamer or hexamer via a CTD-CTD interface (Fig. 5.6). However, to
accommodate an additional subunit, the spacing of the NTD-NTD interactions differs
slightly. The angle between helix 4 of the NTD and helix 1 of the CTD in the interdomain
interface is also slightly larger in the T=3 particles than the T=1 particles. Similarly, the angle
between the two CTDs at the dimeric interface is larger (~ 50º) between two hexamers than
two pentamers.

The CA hexamers of RSV and HIV are similar in that all three CA interfaces are
formed by homologous regions of the two CA proteins (68, 163, 363). However, the
hexamers of the RSV icosahedrons are curved whereas those HIV are flat, a difference that
may be due to different grid preparation techniques used prior to cryo-EM analysis (Fig. S3
and S4 in reference (68)). As discussed above for the RSV pentamer, the interdomain
interface in the RSV hexamer is more extensive than that of the HIV hexamer. The CTD
dimer interface between two hexamers of RSV more closely resembles the CTD dimer
interface between two RSV pentamers rather than the interfaced formed by two hexamers of
HIV. The CTD dimer interface modeled in the pentamer and hexamer has recently been
supported by a crystallographic study (19). The results suggest that differences observed
between the RSV pentamer model and the HIV hexamer model are likely due to differences
between RSV and HIV CA. However, the low resolution model of the RSV hexamer and the
lack of a description of the HIV pentamers prevent a stronger conclusion.
Increasing the Resolution for the RSV Pentamer and Hexamer Model

A higher resolution reconstruction may allow for better visualization of the pentamer-hexamer and hexamer-hexamer connections. Understanding how the capsomers interact will provide a basis for modeling the irregular polyhedral structures observed in authentic capsids. Attempts to increase the resolution by enriching for a greater number of T=3 particles using the I190V mutant protein assembled under various sodium phosphate conditions, purified by a SEC-500 column, or nucleated by T=1 icosahedron show some promise (data not shown). Sample visualization by a more powerful electron microscope also shows promise (data not shown). Alternatively a higher resolution for the T=3 or T=1 particles can be obtained by x-ray crystallography. Unlike the polymorphic structures formed in situ and in vitro, the isometric spheroidal particles may crystallize which could allow the determination of the CA subunit interfaces at an atomic level. Such a structural model may potentially lead to the visualization of the side-chain atoms of the residues located in the NTD-NTD, CTD-NTD and CTD-CTD interface as well as any phosphate or water molecules. Preliminary crystallization trials suggest that crystallization of the icosahedrons is a practical approach but may be limited by the low stability of the structures (data not shown, and R. Kingston, personal communication).

Genetic and biochemical techniques may also be used to test the model of the pentamers. The model predicts that positive-charged residues repulse CA subunit packing (Fig. 5.8). Some basic predictions based on this observation can be tested by mutating the positive-charged residues to neutral or negatively-charged amino acids. The mutant protein should have a decrease in CA subunit repulsion and thus assemble more easily in lower concentrations of sodium phosphate than WT CA protein. Furthermore the regions that form the CA subunit interactions may be mapped by hydrogen-deuterium exchange or by a thiol cross-linking method. In the latter case, Cys residues would be engineered at the presumptive
Fig. 5.8. Charged-residues in the NTD of the pentamer. (a) The surface of the positively-charged residues (blue) and negatively-charged residues (red) are shown in the pseudo-atomic model of the RSV pentamer. (b) The surface charges located in the NTD-NTD interface are shown. (c) The surface was removed to reveal the side-chains of the charged-residues highlighted in (b). The two residues (K17 and R21) that line the pentameric ring are labeled. The images in (a-c) are viewed from the outside of the T=1 particle. (d-e) Same as in a-b, respectively, but viewed from inside of the T=1 particle. In each image the CTD was omitted for greater clarity.
interface, a method that has been successfully used to map the immature hexamer interface in RSV (360) and the mature hexamers of HIV (363).

*Structures Related to the in vitro Assembled T=1 Particles*

The T=1 and T=3 isometric particles are the first description of a retroviral CA protein forming icosahedral symmetry. However, rare spherical particles similar to the T=1 particles also form in authentic RSV virions (Fig. 3b of reference. (58)) supporting the conclusion that pentamers form *in situ*. Additionally, the reverse transcribing Ty3 retrotransposon similarly forms icosahedrons inside of yeast cells (254). The observation that WT Ty3 forms T=3, 4, and 7 particles while those that lacked an active protease or reverse transcriptase enzyme form only T=7 particles led to the speculation that the smaller particles are linked to reverse transcription (254). However, it is unlikely that the RSV particles are linked to reverse transcription since the assembled 17 and 30 nm icosahedron are considerably smaller than authentic cores and too small to encapsidate the viral genome. At this point, no functional role is predicted for the T=1 particles in the virus life cycle.

*Previous Descriptions of 20 nm Particles formed in vitro*

Cryo-EM analyses of the ~20 nm particles provide the first description of pentameric and icosahedral symmetry in retroviruses. However, spherical particles similar in size and shape to the T=1 particles were previously described by two groups. In 2000, Kingston *et al.* showed that purified RSV CA formed “toroidal” particles under low pH and NaCl crystallization conditions (234). The toroidal particles were visualized by negative-stain EM; however further analysis of the particles was not performed at that time. Two years later, Mayo *et al.* reported that protein containing RSV CA-S or CA-SP formed ~20 nm spheres (297). Image reconstruction (averaging of ~300 particles) of negative-stained particles that adhered to a lipid layer led the investigators to speculate that the spheres consisted of ~72 CA
subunits or a sphere with 12 pentamers and 2 hexamers. The resolution of the particles failed to show pentameric or hexameric symmetry. It is highly likely that the ~20 nm particles reported by both Kingston et al. and Mayo et al. are T=1 icosahedrons similar to the ones described here. Of note, RSV CA failed to form spheres under the conditions described by Mayo et al. The protein utilized by Mayo et al. lacked the N-terminal β-hairpin suggesting that SP can suppress the defect caused the lack of a β-hairpin in the D52A mutant in the formation of pentamers (discussed further in Chapter VI).

Support for the Fullerene Model of Retroviral Capsids

The examination of in vitro assembled RSV CA icosahedrons completes the “missing link” of retroviral capsids assembly by answering two questions: 1. Does CA have the ability to form a pentamer? and 2. How do the CA subunits interact in the pentamer? The observation and characterization of the T=1 particles provides a final answer to both of these question. Although an understanding of how pentamers form is still lacking, confirmation of pentameric symmetry formed by retroviral CA protein provides strong support for the hypothesis that retroviral capsids adhere to geometries similar to fullerene cones and polyhedra (58, 164, 198, 210, 276). Cryo-tomograms fail to reveal any specific configurations of subunits within in situ formed capsids; however they provide evidence in support of the fullerene model proposed by Ganser et al. (33, 50, 58, 162, 198, 276). Tomographic analysis establishes that retroviral capsids are likely fully closed structures and that the predicted locations of the pentamers are at the capped ends of the cone for HIV and the angular-bends for RSV. Also in support of the fullerene model, the results in this chapter reveal that both retroviral CA pentamers and hexamers can exist in the in the same lattice. However, other questions remain unanswered: What controls pentamer formation? What determines the final location of pentamers within the hexameric lattice? Do pentamers form prior to hexamers in
the assembling capsid? And, how do the subunits interactions govern the polyhedral morphology? Additional *in vitro* and *in situ* examinations is required to answer these questions. Overall the results in this chapter provide a major step towards understanding the molecular architecture of retroviral capsids at the atomic level.

**Acknowledgements**

We are grateful to John Flanagan (PSU, Department of Biochemistry and Molecular Biology) for his assistance and guidance with sample preparation, Roland Meyers for preliminary image visualization (PSU COM, Core Facilities), and J. Bernard Heymann (NIAMS/NIH) for advice on data analysis. The work presented in this chapter was supported by the Intramural Research Program of NIAMS/NIH and the IATAP Program (A.C.S.) and funding from NIH grant CA100322, the Pennsylvania Department of Health, and the Penn State Cancer Institute (R.C.C.).
CHAPTER VI

DISCUSSION
Assembly of a retroviral virion involves two distinct stages. First, the Gag and Gag-Pol polyprotein coordinates the formation and subsequent release of an immature particle. Second, the viral protease cleaves Gag and Gag-Pol into mature MA (matrix), CA (capsid), and NC (nucleocapsid) proteins resulting in a reorganization of the interior of the virus including formation of a capsid during a multistep process termed maturation. The complex sequence of events leading to the transformation of an immature non-infectious particle into a mature infectious virion is poorly understood. Prior to the investigation described in this thesis, no thorough examination of the pathway of capsid assembly, particularly the initiation of assembly, had been reported. The in vitro data presented in the previous chapters studied how the cleavage products of Gag may stimulate capsid assembly and how CA subunits self-assemble to form a capsid, and provided a structural description for the RSV capsomers. The results provide necessary steps toward developing an atomic-level model for the architectures of the polymorphic capsids and may lead to determining the functions of the capsid during the early stages of viral replication.

**Summary of Data Chapters**

The studies presented in this document analyze the second stage of assembly, specifically the initiation of capsid assembly. An adequate in vitro model system for capsid assembly was developed by inducing the self-assembly of CA from the avian retrovirus Rous sarcoma virus (RSV). Capsid-like structures and all subunit interactions found in authentic viral capsids were formed by RSV CA. In vitro multimerization of RSV CA was dependent on multivalent anions, which are likely acting as a substitute for a polyanionic component found in the virion such as RNA or inositol phosphates. The in vitro pathway of capsid assembly was demonstrated to be nucleation-driven, consistent with previous proposals from the analyses of fully mature virions. Isolation of intermediates along the capsid assembly pathway demonstrated that a complex as small as a CA dimer acts as a seed for assembly
(Chapter II). The conserved hydrophobic residues of the MHR (major homology region), that were previously proposed to undergo a conformational change upon cleavage of Gag, were shown to be critical to proper nucleation of capsid assembly providing support for the idea that structural transformations of CA during maturation are necessary for early steps of CA self-assembly (Chapter III). Analyses of mutant CA proteins, combined with previous examination of mutant viral particles, support the theory that a similar nucleation event occurs in maturing virions (Chapter III). Additionally, the Gag cleavage intermediate, CA-SP, stimulated the initiation of in vitro assembly suggesting the nucleating complex in a maturing virion involves an immature form of CA or a segment of Gag (Chapter IV).

The in vitro assembly of RSV CA revealed not only many novel observations regarding the pathway of capsid assembly but also yielded important structural data. The results show the first visualization of a retroviral pentamer, a decade after its existence was originally proposed. A pseudo-atomic model of the RSV pentamer and hexamer was also achieved (Chapter V). These data provide a major step forward in understanding the architecture of mature capsids at the atomic level and their polymorphic nature. The proficient formation of pentamers, hexamers and structures resembling authentic capsids clearly demonstrates that the in vitro assembly of RSV CA is a more adequate model for retroviral capsid assembly than HIV CA (which only efficiently forms hexamers). The above findings, each of which is specific to one of the four preceding chapters, provide novel concepts of retroviral maturation; however, additional information concerning retroviral capsids (their assembly, morphology, and function) can be gained by discussing the results of chapters II-V in aggregate.
Model for the Initiation of Capsid Assembly *in situ*

The results described in this thesis provide critical information required to build a model for the early steps of retroviral capsid assembly during maturation. Provided here is a model for the pathway of capsid assembly *in situ*, starting with the CA-mediated Gag-Gag interactions in the immature particle (Fig. 6.1). Unlike previous models for the initiation of capsid assembly (Fig. 1.15), the model described here contains details of the assembly pathway and the mechanisms that may control initiation of capsid assembly including how CA undergoes transformations during maturation. Gag hexamers are formed by an immature form of the NTD that lacks the β-hairpin structure (reviewed in Chapter I). In RSV, the immature hexamer likely involves the association of the NTD with the preceding residues located in the p10 domain of Gag. The CTD of CA is involved in two different Gag oligomerization motifs. The C-terminal residues of CA, the residues of the following spacer peptide (SP), and some of the N-terminal residues of NC form a six-helix bundle presumed to provide stability to the immature lattice. Additionally, the first two helices of the CTD, including the conserved MHR residues, may form a dimeric interface via a domain-swap mechanism.

Soon after budding, protease (PR) cleaves the CA-SP domains from Gag, disrupting the immature hexamers. The N-terminus of CA refolds into a β-hairpin stabilized by a salt bridge between the first residue of CA (P1) and a conserved Asp (D52 in RSV) located near the “top” of the third NTD helix. The CTD also undergoes conformational changes induced by PR. In the model, cleavage at the SP—NC site destabilizes the SP 6-helix bundle and the domain-swap dimeric interface. As a result the 6-helix bundle is lost as the helices unfold and become structurally flexible/dynamic. Similarly this model suggests that the conserved hydrophobic residues of the MHR undergo a transition from intermolecular contacts in the domain-swap to intramolecular contacts that stabilize the first two CTD helices.
Fig 6.1. Illustrative model of initiation of capsid assembly for RSV. The immature interaction that the involve participation of the CA-domain of Gag is shown prior to protease (PR) digestion (MA and NC domains of Gag are omitted for clarity). Upon PR activity, CA is cleaved out of Gag as a transient CA-SP intermediate. Both the NTD and the CTD of the newly mature CA undergoes conformational changes as described in the text (shown as spheres). Capsid assembly is initiated by a CA-SP dimer that forms via the interdomain (NTD-CTD) tripartite interaction (Fig. 6.2). Continued PR activity results in the trimming of SP from CA-SP to form the mature CA species: CA and CA-S. The addition of both CA and CA-S leads to the formation of pentamers and hexamers. Fully-closed polymorphic polyhedrons with angular bends are formed by 150-300 hexamers and 12 pentamers. Shown is an in vitro formed (gray) and in situ (yellow, (58)) assembled capsid.
Gag Hexamer

p10
NTD
CTD
SP

PR cleavage

NTD
CTD
SP

CA-SP Monomers

Dimeric Seed

NTD-CTD Tripartite Interface

Removal of SP

Addition of CA subunits

CA + CA-S

Capsid Growth

in vitro Capsid

in situ Capsid
Fig. 6.2. Tripartite interaction as a pseudo-atomic resolution model for assembly initiation. In the model two NTDs and a CTD interact to form a region of extensive interactions called the tripartite interaction. One NTD and its CTD interact with the neighboring NTD in the pentamer. (a) Positively charged residues (Arg and Lys) found in the region are shown. In part (b) the MHR mutations are in red and the suppressor mutations are in green. The hydrogen atoms have been removed from the shown side chains for clarity.
Next, assembly of a capsid initiates by a CA-SP dimer forming via an interdomain NTD-CTD tripartite interaction where two neighboring NTDs interact with a CTD (Fig. 6.1-6.2). The tripartite interface is likely formed or stabilized by charge-shielding of positive residues by an unknown polyanionic factor. The conformational changes of the NTD and CTD help to establish the tripartite interaction. At some point, either before or after the formation of the assembly-active dimeric seed, PR cleaves all or part of SP from CA-SP resulting in a pool of mature CA and, in the case of RSV, CA-S. Next, the addition of CA and CA-S subunits to the dimeric seed causes rapid growth of the capsid in multiple directions until being confined, likely by the viral membrane or membrane-associated components. A pentamer likely forms very early and may form prior to a hexamer. The formation of a pentamer introduces a bend in the growing capsid lattice, a requirement to forming a closed capsid shell. The polymorphism exhibited by retroviral capsids is the result of pentamers being randomly dispersed among the hexameric lattice. The factors influencing the formation and placement of the pentamers are unknown but may involve polyanionic components of the viral membrane and the conformation of neighboring CA subunits.

This model is simplified. Other events occur during maturation were not examined in this thesis, such as the condensation of NC and the genomic RNA to form the interior of the core, likely influence capsid assembly in situ. However, the model provides a starting point for more rigorous investigation. The in vitro system developed provides a simple method for initial studies for further testing, such as examining the influence of RNA or viral membrane components (24, 50) on in vitro capsid assembly. Importantly, in vitro and in situ experiments converge on the key aspect of the above illustrative model including that capsid assembly is nucleated by a specific CA complex (21, 24, 33, 50, 54, 58, 117, 259, 372, 373). Additionally, the model provides an explanation for many aspects of maturation and capsid assembly including the formation of the β-hairpin, the existence of the transient CA-SP
species, the influence of the MHR on capsid assembly, and the importance of the NTD-CTD interface. Supporting details of the illustrative model are discussed below.

Disassembly of Gag and de novo Assembly of Capsid

The hexamers formed by Gag in the immature lattice are widely thought to be lost as a result of proteolytic activity during maturation. Thus, capsids are assembled anew (without requirement of the immature Gag lattice) and can be considered a de novo process; many lines of evidence support such a claim. First, monomeric CA has the intrinsic ability to form capsid-like structures without the requirement of any other viral component, including a Gag lattice (Chapter II-III) (125, 127, 179, 259, 276). Second, the occurrence of pentamers in mature capsids can not be explained unless the hexamers of the immature lattice are disassembled since there is no evidence to suggest the immature lattice contains pentameric rings (51, 58, 162, 164, 198, 505). Third, multi-layered shells and nested capsids observed in mature virions also appear compatible with the disassembly of the single layered Gag lattice prior to capsid assembly (33, 50, 53, 54, 58). Additionally, a significant amount of CA (as high as 50%) presumably participating in the immature lattice is not incorporated into the capsid shell, consistent with the break-down of Gag-Gag interactions prior to capsid assembly (54, 55, 58, 475, 496, 518).

Data presented in this thesis are also consistent with de novo assembly of capsids. Specifically, nucleation-driven in vitro assembly is independent of an immature Gag lattice. Additionally, the loss of infectivity caused by the MHR mutations is likely due to their influence on assembly of monomeric CA, specifically the nucleation event, since the mutant proteins were defective for in vitro capsid assembly but lack any observable influence on immature viral assembly. Furthermore the analysis of the interdomain suppressors (Chapter III, and (48, 283, 431)) provides strong evidence to support the idea that de novo capsid assembly occurs since there is no evidence to support an interdomain interaction in the Gag

233
hexamers (51, 257, 505). Although direct evidence to affirm de novo assembly of capsids or any other alternative (such as condensing the immature lattice into a mature shell) is lacking, numerous observations are consistent with de novo synthesis, while little evidence exists to support other pathways.

Identifying the Dimer Seed Interface: Evidence for the Interdomain Tripartite Interaction

The biochemical and genetic data in chapters II-III support that the dimeric-seed of assembly is formed by a tripartite interaction (Fig. 5.3 and Fig. 6.2). First, the NTD undergoes a conformational change upon dimerization as indicated by the decrease in intrinsic Trp fluorescence in the dimeric-seed. The location of the Trp residues in the NTD suggest that the N-terminus of helix 4 of the NTD is altered (Fig. 2.4). The N-terminus of helix 4 of the NTD establishes interactions with the CTD in the tripartite interface. Second, the anion-dependence of CA assembly in vitro (Fig. 2.1) is consistent with the finding of 9 positively-charged residues clustered in the tripartite interface (Fig. 6.1a). The repulsion caused by these positive charges must be overcome to form stable complexes; however only a minor amount of negative charge is supplied by neighboring residues (D24, E41 & E176) suggesting the bulk of the charge neutralization in vitro is due to the phosphate anions added to induce assembly. Formation of the tripartite interaction during initiation provides an explanation for the dependency of in vitro capsid assembly on multivalent anions.

Genetic evidence suggests that the CTD is required for proper initiation of capsid assembly in vitro and in situ. Specifically, the failure of lethal MHR mutants to nucleate assembly (Fig. 3.4 and 3.8) implies that the integrity of the CTD is needed for nucleation. The genetic and biochemical data (Chapters II and III) suggest that the interdomain NTD-CTD interface forms early during self-assembly of RSV CA and imply an active role for the interdomain interface in the initiation of capsid assembly rather than simply providing stability to the formed capsid shell as suggested by others (163, 165, 452).
Genetic and biochemical evidence supports the formation of the tripartite interaction in the assembly-active dimers, however direct structural confirmation supporting such a conclusion is lacking. Chemical cross-linking showed that the dimer is easily cross-linked with lysine-reactive DTSSP (3’,3’-Dithiobis-[sulfosuccinimidylpropionate]). However preliminary attempts to identify the critical dimeric interface by tryptic-digestion and mass spectrometry failed (data not shown). Alternatively, methods such as hydrogen:deuterium exchange or crystallography may yield useful results to identify the dimerization domain in the assembly seeds. Currently these alternative techniques are restricted by the metastable nature of the dimer (i.e. the dimer easily dissociates into monomers).

Role of the MHR during Initiation and the Effects of the Suppressors on CA Assembly

Further support for the illustrative model for initiation of capsid assembly (Fig. 6.1-6.2) is derived from the study of RSV CA mutants. The lethal MHR mutations (F167Y & L171V) abolished the nucleation of in vitro capsid assembly (Chapter III). These residues are located in a hydrophobic pocket critical to the structure of the first two helices of the CTD (160, 234, 504) (Fig. 5.3 and I. Ropson, personal communication). The MHR residues are likely necessary for the NTD-CTD and perhaps the CTD-CTD interactions between neighboring CA subunits in the capsid (68). Importantly, the conserved hydrophobic residues critical to in vitro capsid assembly are the homologous residues found in the domain-swap immature HIV CA dimer (Fig. 1.10). As such these residues are likely to be flexible and to play a key role in switching the CTD from an immature to a mature dimer. Previous characterization of mutant viruses (61, 96) and the biochemical in vitro results (Chapter II) show that the MHR is critical to both immature and mature virion assembly, consistent with a presumed structural role in each step.
Three of the four second site suppressors restored *in vitro* assembly and increased the ability of CA to multimerize (Figs. 3.4 and 3.9). These suppressors appear to overcome the defective nucleation step of *in vitro* capsid assembly caused by the lethal MHR mutations by stimulating the initiation of assembly. The A38V suppressor stimulated *in vitro* assembly (Fig. 3.4) and restored infectivity to L171V and F167Y mutant viruses (48, 283). Position 38 is located on second helix of the NTD where A38V may modulate the NTD-NTD interactions in the tripartite interface (Fig. 5.5b).

The other suppressors (R185W and I190V) examined are located in the CTD and are similarly located near the mature CA-CA interfaces. The R185W substitution may influence both NTD-CTD and CTD-CTD interaction since it is located in the tripartite interface and at the top of the dimerization helix. The ability R185W to act as a suppressor may be due to an increased stability in the entire CTD as implied by the preliminary unfolding studies (Fig. 3.2). This mutation may also act as a suppressor by neutralizing one of the positive charges located in this highly basic interface and increasing the hydrophobicity of the interface. The I190V suppressor is located on the second helix of the CTD and likely influences dimerization (19, 234). The suppressors are located near the CA-CA interfaces and increase the propensity of CA to multimerize likely by modulating mature CA-CA interfaces, especially the tripartite interaction. Consistent with the increased sodium phosphate sensitivity observed in the suppressor proteins (Chapter III), the suppressor proteins likely multimerize into small forms of CA, such as the dimeric-seed, more easily than WT protein, a proposal that can be tested using analytical ultracentrifugation.

One prediction from the model (Fig. 6.1-6.2) is that a mutation in the interdomain interface may increase initiation of capsid assembly and suppress the lethal phenotype of a MHR mutant. One previously identified suppressor (P65Q) is located near the closest contact between the NTD and CTD in the pentamer model and may provide an answer to this question. The P65Q suppressor restored partial infectivity to the lethal R170Q MHR mutation.
Like the other suppressors, the P65Q substitution caused CA to rapidly form turbid material upon the addition of sodium phosphate (Fig. 6.3). However, P65Q failed to assemble into observable structures (data not shown). The turbidity curve of P65Q suggests that this protein may be stimulating the initiation of assembly to a point where the protein is trapped in a state that prevents further assembly (i.e. kinetically trapped or over-initiated) (78, 130, 528, 531, 542, 547). In such a case the intermediates are unable to act as protomers for assembly. Further experiments with P65Q including co-assembly reactions with WT protein will be needed to confirm this over-initiation hypothesis. The preliminary results support the initial conclusion that a suppressing mutation in the NTD-CTD interface increases the nucleation of capsid assembly. The in vitro assembly phenotype of the P65Q mutant suggests that in situ assembly of capsids in P65Q mutant virions is defective; however this appears to not be the case since the mutant virus is infectious (48) suggesting key regulatory factors of in situ capsid assembly are missing in the in vitro conditions.

The Role of CA-SP in the Initiation of Capsid Assembly

The transient CA-SP maturation intermediate clearly stimulated in vitro assembly (Chapter IV). It is difficult to explain how SP stimulates nucleation since the residues of SP are not found in mature capsids. Perhaps the secondary or tertiary folding of the CTD in CA-SP is stabilized under the assembly conditions tested. Such a stabilizing effect may stimulate capsid assembly by enhancing the formation of the CA subunit interactions, including the tripartite interface. In support of this conclusion, the in vitro structures formed by CA-SP are similar to those assembled by CA, and presumably have similar subunit interactions including the interdomain interface.

Alternatively SP may induce a novel CA-CA interface. A new interface may introduce novel geometric constraints on assembly resulting in the formation of
Fig. 6.3 *In vitro* assembly of P65Q mutant suppressor CA protein. The turbidimetric profile for the suppressor mutant, P65Q, shows a rapid but limited rise in optical density. Assembly was performed using 500 mM sodium phosphate pH 8.0.
morphologically distinct structures such as the large round particles shown in Fig 4.3j-n. The presumptive novel interfaces may involve SP-SP homo-oligomerization. These novel interfaces could potentially be identified by cross-linking experiments.

The in vitro results suggest that CA-SP can assemble capsid-like structures in vitro. However, proteolytic processing of CA-SP into the mature CA must occur for infectivity (95, 355, 508). Also blockage of the processing event results in aberrant capsids suggesting CA-SP does not self-associate in situ to form the capsid shell. HIV particles containing CA-SP1 (and no CA) have a partial shell suggesting CA-SP1 can initiate capsid formation but CA is required to correctly assemble a capsid (1, 275).

The Role of the β-hairpin during Assembly Initiation

The β-hairpin is a hallmark structure of mature CA and participates in capsid morphogenesis. The β-hairpin is not directly involved in any CA subunit interface but is located “above” the helices forming the pentameric and hexameric rings. The β-hairpin is required for proper capsid assembly in situ and in vitro. Under sodium phosphate conditions a mutant protein (D52A) lacking the β-hairpin rapidly forms amorphous turbid material. The in vitro assembly phenotype of D52A (Fig. 3.3) is similar to that of P65Q mutant protein (Fig. 6.3) suggesting that the uncontrollable aggregation observed in D52A mutant protein is also due to a defect during initiation of assembly. This hypothesis predicts that the β-hairpin modulates the pathway of capsid assembly by influencing the conformation of the NTD and the formation of the tripartite interaction. Additional support for the β-hairpin modulating the interdomain interaction is provided by a replication deficient mutation located in the β-hairpin being suppressed by a CTD suppressor (431).

A NMR study of HIV CA suggests that folding of the N-terminus upon proteolysis induces changes in the NTD near helices 5 and 6 (458). Thus, the β-hairpin may control other
conformational changes in the NTD to promote CA-CA interactions or stabilize an assembly intermediate such as the nucleating complex. However, supporting evidence for these ideas is lacking since no clear structural model explains the requirement of the β-hairpin for capsid assembly (171, 234, 327, 476). A genetic study in RSV suggests that there is a long-range interaction between the β-hairpin and other regions of CA including the CTD dimerization helix (431). Further genetic studies including isolation of second-site suppressors of mutant CA lacking the β-hairpin may illuminate how the β-hairpin influences capsid assembly.

Does Pentameric Formation depend on the β-hairpin?

Data from the D52A mutant (Chapter III) and RSV CA-SP (Chapter IV) clearly demonstrate that the termini of CA influence capsid assembly. Can the block in assembly caused by the D52A mutant be suppressed by the stimulation of assembly caused by SP? Intriguingly, the sphere that were associated with the tubes formed by CA-SP (fig 4.3g-i) are reminiscent of spheres formed by RSV CA-SP on top of a planar array of lipids as reported by Mayo et al. (297). The spheres forming on the lipid layer lacked the β-hairpin due to an N-terminal his-tag. Spheres were only noted for CA-S and CA-SP and are similar in size and symmetry to the T=1 particles visualized by negative staining (Figs 2 & 7 in reference (297), and Rich Kingston unpublished). These observations suggest that pentamer formation is independent of the β-hairpin when SP is present. This prediction could easily be tested with purified D52A CA-SP protein under the sodium phosphate assembly conditions. If a pentamer forms independently of the β-hairpin then T=1 icosahedral particles should be observed in sodium phosphate-induced assembled D52A CA-SP mutant protein. The experimental results may aid in understanding the role of the β-hairpin and SP during capsid assembly.
Alternative Dimeric Seed Interface of RSV CA Multimerization

A recent study by Bailey et al. showed that RSV CA assembles into structures resembling authentic capsids by the addition of high concentrations of sodium chloride in acidic conditions without the presence of phosphate (19). Similar to the conclusion reached in chapter II, Bailey et al. suggested that initiation of capsid assembly involves a CA dimer. However, the authors proposed CTD dimerization nucleates assembly based on the observation that charge neutralization of two Asp residues in the CTD-CTD interface by a low pH (5.0 or lower) initiated capsid assembly.

Bailey et al. also report the first CTD-CTD dimer crystal structure for RSV. The crystal structure formed at low pH is similar to the CTD dimer interface modeled in chapter V and likely reflects the subunit interface in authentic capsids. This finding substantiates the model originally proposed by Worthylake et al. for the HIV CTD dimer (504).

In contrast to Bailey et al., only small spheroidal structures were observed at moderate levels of sodium chloride in chapter II (Fig. 2.1). Although Bailey et al. focused on the difference in pH (5.0 in the Bailey et al. study versus 7.5 – 8.0 in the current study) as the critical factor, other discrepancies (e.g. differences in sodium chloride concentration) may account for assembly without the presences of multivalent anions. Perhaps even more importantly, the experiments described in this thesis used protein purified by size exclusion chromatography to separate monomeric CA from oligomeric CA. Without this critical step, trace contaminants of CA oligomers may act as seeds for assembly. Although the studies in chapter II-III suggest that the NTD-CTD interface drives assembly, other potential dimeric interfaces (such as the one presented by Bailey et al.) can not be ruled out. Potentially more than one specific interface may nucleate assembly.
Nested Capsids and Multilamellar Capsids: Evidence of an Alternative Nucleation Site

Both in situ and in vitro assembly results in the formation of capsids that enclose other capsids (Fig. 3.5k and 5.2) and that form multi-layered shells (Fig. 2.1 and 3.3) (33, 50, 53, 54, 58, 68, 372, 373). The nested capsids (i.e. capsids within capsids) are rare but were observed in three independent tomographic studies (33, 50, 58). The best example of a characterized nested capsid is the RSV CA T=1 icosahedron enclosed by a T=3 particle (Fig. 5.2c). The nested capsids suggest that the second layer was subsequently added to the initial shell. A loop in the NTD of the nested T=1 particle contacts a loop in the CTD of the outer T=3 layer suggesting that the T=1 spheres act as a template for the assembly of the T=3 layer. This layer-nucleated assembly suggests an alternative pathway of assembly that may account for the multiple capsids and nested capsids observed by tomography (33, 50, 58).

Nested capsids are rare in retroviruses (and thus apparently aberrant). Nested capsids in viral particles may interfere with the formation of an infectious virion. In vitro assembly of CA-SP showed an increased occurrence of spheres nested in or around tubes. As such, SP may provide an alternative nucleation site, suggesting that CA-SP nucleates nested or multi-layers capsids in situ.

Retroviral Polyhedrons and Icosahedral Symmetry

Many diverse viruses have capsids that assemble into regular icosahedrons (Chapter I). Icosahedrons are a specific class of polyhedrons containing 20 equilateral triangular planes. Similar to the icosahedral capsids of other viruses, the polyhedral capsids of retroviruses contain both pentamers and hexamers and local areas of 2-, 3-, and 5-fold symmetry. However retroviral capsids lack the global symmetry found in icosahedral capsids. The extent to which a capsid departs from an icosahedron is reflected by the irregular distribution of the pentamers among the hexamers (162, 198, 332).
Modeling of authentic RSV cores in accordance with the fullerene model of carbon nanotubes (i.e. a lattice with a variable number of hexamers and 12 pentamers) showed local regions of the capsids may have partial icosahedral arrangements (58, 198). Specifically the cap of coffin-shaped polyhedral cores observed by tomographic studies corresponds to a T=17 icosahedron. This modeling is based on ideal shapes, however authentic capsids appear less regular and experimental densities derived from tomograms were a poor fit for icosahedral models (58, 198). Thus retroviral capsids, although categorized as irregular/complex capsids in chapter I, are geometrically related to icosahedral capsids but exhibit a high amount of polymorphism.

**Determinants of Retroviral Capsids Polymorphism**

The data in this thesis provide a foundation for understanding the polymorphic nature of capsids. Greater than a million geometrically distinct polyhedra can be formed by a typical RSV capsid lattice made by 150-300 hexamers and 12 pentamers (198). Do the diverse structures assembled by retroviral CA subunits represent any quasi-equivalency? In contrast to icosahedral capsids built by a small number of quasi-equivalent subunit interactions, a near-continuum of quasi-equivalent conformations of CA has been proposed (198, 332). For icosahedral viruses, such as bromoviruses, quasi-equivalency can be achieved by flexibility in the coat protein (Fig. 1.4). The flexible linker that connects the NTD and CTD may provide variability to the capsid shell (68, 198, 221). Slight variations in the subunit-subunit contacts of the three interfaces (NTD-NTD, NTD-CTD and CTD-CTD) also likely contribute to the broad range of quasi-equivalent interactions formed by CA (68, 198, 363). The results presented in chapter II and V provide strong support for the concept that RSV capsids are irregular polyhedra which can be modeled using fullerene architecture (58, 162, 164, 198). Thus, flexibility in either the structure of the CA monomer or in the intermolecular
interactions formed upon multimerization is needed to build retroviral capsid shells using fullerene principles (33, 162, 164, 276, 332). A better understanding of the flexible nature of CA and of the subunit interactions formed will be needed in order to generate a pseudo-atomic model for a polyhedral capsid lattice.

A recent high resolution crystal structure for the mature hexamer formed by mutant HIV CA shows that the bulk of the NTD-NTD interface is created by hydrophilic contacts (363). An extensive hydrogen-bonding network is formed by well-ordered water molecules between atoms of the polar side chains and protein backbone. The NTD-CTD interface is similarly mediated by hydrophilic interactions (363). The pentamers and hexamers of RSV are likely formed by similar hydrophilic contacts since charge neutralization by phosphates is critical to assembly (Chapter II-V). Overall, HIV CA crystallization showed a slight alteration in the interfaces (i.e. altered orientations in hydrogen-bonding and side chain packing) can be achieved since water molecules can easily be repositioned. Thus, each CA subunit interaction may serve as a point of variability, resulting in a high degree of polymorphism.

Summary of and Alternative Approaches to the Proposed Experiments

Future experimental plans are proposed in the preceding chapters, highlighted here are some of the plans that may yield significant data necessary to understanding CA self-assembly and capsid functions. First, defining the CA-CA subunit interactions in the nucleating assembly-seed or other capsid assembly intermediates (Chapter II) is needed. Analysis of such on-path assembly CA complexes is hindered by the ability to purify such intermediates. Methods to enrich for the purification of CA oligomers during E. coli expressed protein-purification may be possible. However, since the factors critical for the formation and purification of capsid assembly intermediates (dimeric – tetrameric CA
oligomers) characterized in chapter II are unknown an efficient experimental approach cannot be designed. The use of mutant CA that may stabilize a CA oligomeric complex (similar to I190V stabilizing the icosahedron, Chapter V) may be possible. Purification of CA proteins that contain substitutions in the intersubunit interfaces can easily be performed and monitored for an enrichment of potential assembly intermediates. Alternatively, capsid assembly intermediates may be purified from solution of CA self-assembling under multivalent anionic conditions. The low stability of the complexes may be over-come by using chemical cross-linkers. Adding cross-linking reagents during early time points following sodium phosphate induced assembly of CA may yield a high amount of stable capsid assembly intermediates. Purified cross-linked CA complexes may then be studied structurally by crystallography, small-angle X-ray scattering, or potentially by cryo-EM.

Secondly, in chapter III, a strong correlation between in vitro assembly and infectivity was noted for multiple CA mutants. The suppressing CA mutants that restored infectivity to lethal CA substitutions increased the propensity of CA to multimerize. Although the suppressors and double mutant CA proteins had the ability to form capsid in vitro, all of these mutations caused a slight decrease in infectivity (283) suggesting that the mutations may cause a secondary affect on the viral replication cycle. For example, the F167Y/A38V protein efficiently assembles in vitro (Chapter III) and in situ (Butan et al. unpublished) but the mutant virus is only ~70% as infectious as WT (283). Perhaps F167Y/A38V has a disruption in disassembly. Currently an efficient tool to easily examine disassembly of RSV cores in newly infected cells is lacking. Examination of the stability of structures formed in vitro may ascertain if the mutations alter the ability of CA to disassociate. Structures assembled by 500 mM sodium phosphate can be disassembled in vitro by decreasing the phosphate level and protein concentration (data not shown). However this in vitro disassembly assay
needs to be further developed (i.e. examination of the disassembled material by EM, an understanding of how disassembly is influenced by decreasing phosphate versus decreasing protein concentration, influence of other factors including temperature, pH, and sodium chloride level) prior to comparing CA mutants to WT CA.

In Chapter IV, CA-SP and longer fragments of Gag are hypothesized to stimulate the nucleation step of capsid assembly. Previously full-length Gag and truncated forms of RSV Gag (Gag missing the membrane-binding domain (ΔMBD-Gag), CA-SP-NC, p10-NTD) were purified and assembled \textit{in vitro} into immature-like structures (5, 63-65, 288, 327, 360). Thus, the ability of monomeric and multimeric forms of Gag (including structures that form in the presences of RNA or IP₆) to nucleate CA assembly can be straightforwardly tested under sodium phosphate-induced \textit{in vitro} conditions.

In the preceding chapter, pentameric and hexameric complexes of CA are described at 10 – 20 Å resolution. A comparison of HIV hexamers to the RSV structures is described (Chapter V). However attempts to purify HIV pentamers or icosahedrons have failed. Much of the previous attempts focused on WT HIV CA. Mutant HIV CA protein may increase the likelihood of visualizing the HIV pentamer. Mutations at R18, particularly R18L, assembled spheroidal structures under sodium chloride-induced assembly conditions (163). Potentially R18L may form icosahedrons that are suitable for cryo-EM particle reconstruction under sodium phosphate conditions. Alternatively CA-SP1 or CA-SP1-NC may increase pentamer formation. In a previous study, HIV CA-SP1-NC assembled cones by incubation with RNA and high amounts of sodium chloride (162) suggesting an increased efficiency
of pentamer formation when compared to sodium chloride induced assembly of HIV CA. *In vitro* assembly of CA-SP-NC or truncated proteins that remove part of NC may assemble into icosahedrons similar to those described for RSV (Chapter V).

**Final Thoughts and Future Directions**

The polymorphic characteristic of retroviral capsids suggests that infectivity is not limited to virions with only a specific morphology. It is interesting to speculate that the variability in capsid structures is important to the replication of retroviruses but no advantage from their polymorphism is evident. The descriptions of the retroviral pentamer and hexamer in chapter V provide a major step forward in understanding the architecture of mature capsids at the atomic level. However, a structural understanding of how the CA subunits interact to form a high degree of polymorphic capsids awaits further research.

The process by which the CA assembles is also puzzling. Retroviruses, to an extent, tend to adopt certain shapes: cone-like capsids for HIV and irregular polyhedra for RSV. This observation suggests that CA assembly is a favored process as opposed to following a strictly defined pathway. The results described in this thesis build a foundation for understanding the pathway of capsid assembly, particularly the early steps of assembly, and provide a launch pad for future research. What triggers assembly *in situ*? What factors control assembly and morphology? How does the capsid support reverse transcription? These are a few of the many questions still remaining to be answered.

The data in this thesis provide some hints to these questions (such as a polyanionic compound and CA-SP triggering assembly) but no answers. Also the CTD, including the MHR, forms the inner layer of the capsid and may interact with the viral RNA or the reverse transcriptase enzyme to coordinate initiation of reverse transcription. At this time, however, no model provides information regarding how the capsid shell interacts with other core
components or functions during early events of replication including reverse transcription. Answers to the above questions and structurally defining key intermediates of capsid assembly, including the nucleating seed, may provide additional targets for future development of antiviral inhibitors. It is hoped that the information contained in this thesis will provide further questions and inspiration for the future work that is necessary to further understand the biology of retroviruses.
Reference List


37. **Berthet-Colominas, C., S. Monaco, A. Novelli, G. Sibai, F. Mallet, and S. Cusack.** 1999. Head-to-tail dimers and interdomain flexibility revealed by the
crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. EMBO J. 18:1124-1136.


158. **Fuller, M. T. and J. King.** 1980. Regulation of coat protein polymerization by the scaffolding protein of bacteriophage P22. Biophys.J. **32:**381-401.


284


291


Vita

John Gerard Purdy

Education
Ph.D. in Microbiology & Immunology 2003-2009
The Pennsylvania State University, College of Medicine
Hershey, Pennsylvania

B.S. in Biology, magna cum laude 1997-2001
Pacific University, College of Arts and Sciences
Forest Grove, Oregon

Honors and Awards
Participation Support for Oral Presentation

ASV Travel Grant for Oral Presentation

Beta Beta Beta National Biological Honors Society
Pacific University, chapter president (2000-2001)

College of Arts and Sciences Dean’s List
Pacific University, Forest Grove, OR.

Publications


Professional Membership
American Society for Virology, student member since 2005