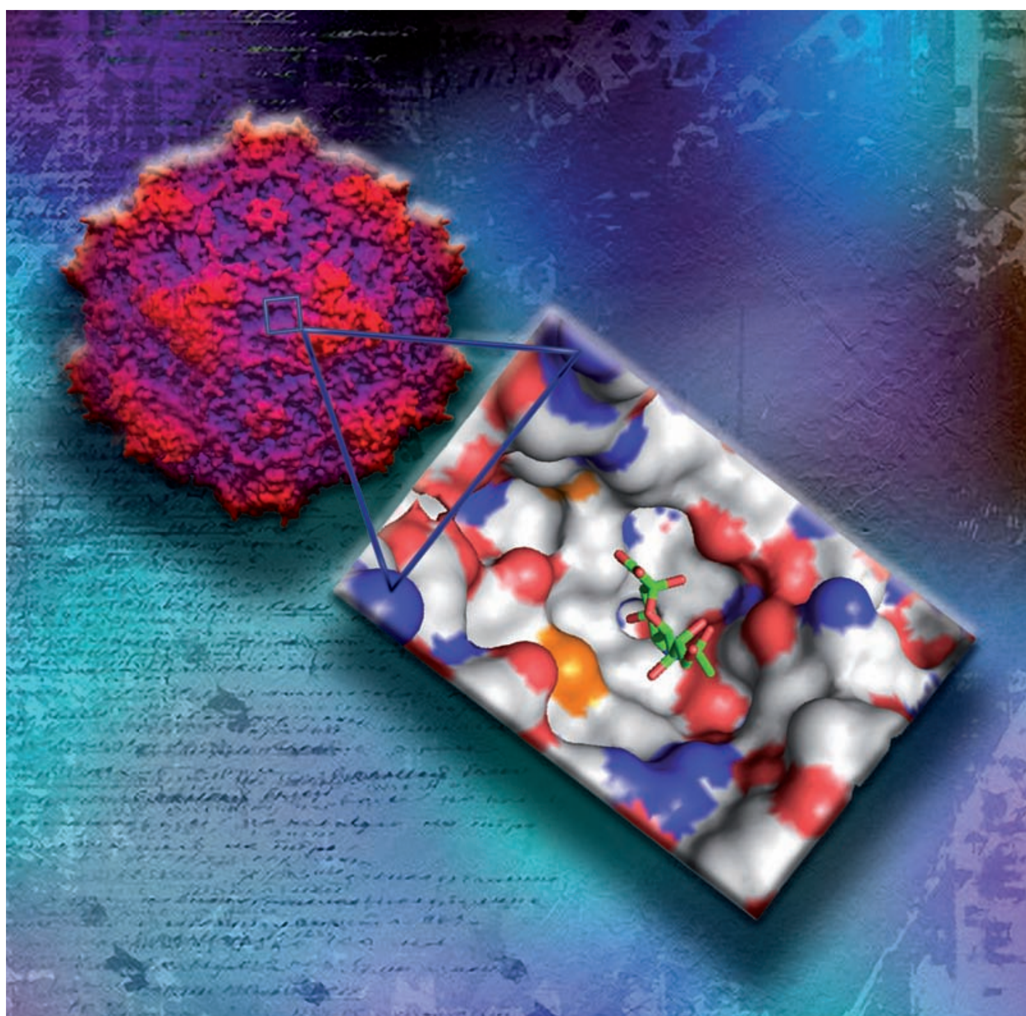
RSC Biomolecular Sciences

Edited by Mavis Agbandje-McKenna and Robert McKenna

Structural Virology



RSC Publishing

Structural Virology

RSC Biomolecular Sciences

Editorial Board:
Professor Stephen Neidle (Chairman), *The School of Pharmacy, University of London, UK*
Dr Marius Clore, *National Institutes of Health, USA*
Professor Roderick E Hubbard, *University of York and Vernalis, Cambridge, UK*
Professor David M J Lilley FRS, *University of Dundee, UK*

Titles in the Series:

1: Biophysical and Structural Aspects of Bioenergetics
2: Exploiting Chemical Diversity for Drug Discovery
3: Structure-based Drug Discovery: An Overview
4: Structural Biology of Membrane Proteins
5: Protein–Carbohydrate Interactions in Infectious Disease
6: Sequence-specific DNA Binding Agents
7: Quadruplex Nucleic Acids
8: Computational and Structural Approaches to Drug Discovery: Ligand–Protein Interactions
9: Metabolomics, Metabonomics and Metabolite Profiling
10: Ribozymes and RNA Catalysis
11: Protein-Nucleic Acid Interactions: Structural Biology
12: Therapeutic Oligonucleotides
13: Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches
14: Nucleic Acid-Metal Ion Interactions
15: Oxidative Folding of Peptides and Proteins
16: RNA Polymerases as Molecular Motors
17: Quantum Tunnelling in Enzyme-Catalysed Reactions
18: Natural Product Chemistry for Drug Discovery
19: RNA Helicases
20: Molecular Simulations and Biomembranes: from Biophysics to Function
21: Structural Virology

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road, Cambridge,
CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247, Email: books@rsc.org
Visit our website at http://www.rsc.org/Shop/Books/

Structural Virology

Edited by

Mavis Agbandje-McKenna and Robert McKenna
Department of Biochemistry and Molecular Biology, University of Florida, FL, USA

RSC Publishing

RSC Biomolecular Sciences No. 21

ISBN: 978-0-85404-171-8
ISSN: 1757-7152

A catalogue record for this book is available from the British Library

© Royal Society of Chemistry 2011

All rights reserved

Apart from fair dealing for the purposes of research for non-commercial purposes or for private study, criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry or the copyright owner, or in the case of reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.

The RSC is not responsible for individual opinions expressed in this work.

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Preface

Viruses can be grouped among the simplest biological systems that have the ability to evolve and adapt to exist in different environments. That is, they have the ability to 'jump' from one host to another, some carrying the necessary molecular machinery to transfer and modify their genetic information from one generation to the next, while others hijack the host machinery to effect the necessary modifications. Because of this innate ability, it would not be unreasonable to state that viruses have most likely infected every life form that has ever existed on our planet, from the simplest single-cell organisms to plants, animals, and humans.

To achieve such biodiversity, viruses have evolved different and efficient strategies for host recognition, internalization, cellular trafficking, genome replication, capsid assembly, genome packaging, release of progeny (for re-infection) and host immune surveillance evasion, to optimize their life cycle in their unique niche. This has resulted in viruses of different shapes and sizes, from simple single-protein spherical or helical assemblages, to multiple complex systems, assembled from hundreds of proteins without/with (enveloped) the incorporation of host lipids. Invariably the viral coat protein(s) (referred to throughout this monograph interchangeably as either CPs or VPs) form some sort of integral protective shell (a viral capsid) around the infectious genomic nucleic acid, which can be single-stranded (ss) DNA, ssRNA, double-stranded (ds) DNA or dsRNA, packaged as single or multiple, linear or circular molecule(s). The packaged viral genome encodes all the required structural CPs/VPs and auxiliary non-structural proteins that are required in combination with host proteins for host infection. The enveloped viruses incorporate their host's lipids as either an internal and/or external envelope during their assembly. For a number of viruses, CP/VP recognition and encapsidation of the genomic nucleic acid is a prerequisite for infectious capsid formation, whereas for others the genome is packaged into preformed capsids via interactions with viral or host encoded proteins. In addition to genome encapsidation and protection during cellular entry and trafficking, the CP/VP can also dictate many other

RSC Biomolecular Sciences No. 21
Structural Virology
Edited by Mavis Agbandje-McKenna and Robert McKenna
© Royal Society of Chemistry 2011
Published by the Royal Society of Chemistry, www.rsc.org

viral functions, including host receptor/vector recognition, transmission and the genomic transduction efficiency during infection.

For spherical viruses, the CP/VP organization in the capsid architecture takes on the form of an icosahedron (a platonic solid with point group symmetry 5.3.2), a regular polyhedron which is assembled from 20 equilateral triangles. This symmetrical shell is a consequence of it consisting of identical (or almost identical) gene products, consistent with the argument that there is insufficient volume inside a virus to accommodate a more complicated protein coding strategy. The exact twofold, threefold and fivefold symmetry of the icosahedron permits the (quasi) equivalent symmetry required to construct structures with 60 or multiples [denoted by a T (triangulation) number] of 60 subunits. This monograph will discuss viruses assembled from the simplest of icosahedral capsids, with $T = 1$ triangulation (assembled from 60 CP/VP subunits), to those with more complicated VP shells assembles and lipid membrane envelopes.

Viruses have been responsible for more human deaths, either through direct infection (such as influenza virus) or infection of crops, than any other known human disease-causing agent. In addition, their ability to package efficiently and deliver genomic material to different living organisms and tissues also makes them attractive vehicles for the delivery of therapeutic genetic material in situations where defective genes lead to disease phenotypes. Thus viruses are the subject of intense scientific study in many different disciplines, including structure biology, in efforts to (i) understand the basic biological processes governing viral infection and (ii) develop treatment strategies, including vaccines, anti-virals and gene delivery vectors.

The use of structure approaches in virology has given insight into the structural basis of assembly, nucleic acid packaging, particle dynamics and interactions with cellular molecules and allowed the elucidation of mechanistic pathways at the atomic and molecular level. Biological processes, such as the life cycle of a virus infection, are governed by numerous intricate macromolecular interactions. The role of the structural virologist is thus to visualize these interactions in three dimensions (3D), to provide a full understanding of these interactions as 'seeing is believing'. These structural characterizations of viruses then provide crucial platforms for the development of treatment and therapeutic strategies (Section 3 of this monograph).

The range of biophysical methods used in structural virology is vast, ranging from hydrodynamic to scattering techniques (Section 1 of this monograph), and have played a fundamental role in our understanding of viral infection in recent years. The method undertaken for a particular study is often dependent on the resolution and type of information desired and also the size and complexity of the macromolecule under investigation, the amount of material available, its solubility in aqueous environments (Chapter 1) and the type of interactions being visualized. For example, for the imaging of whole viruses during infection, confocal microscopy (Chapter 2) and cryo-electron tomography (cryo-ET) (Chapter 4) are applied, which permit studies at molecular resolution. And while both nuclear magnetic resonance (NMR) spectroscopy (Chapter 8) and X-ray crystallography (Chapters 6 and 7) can give atomic resolution detail on protein

backbone and side-chain placement, NMR also provides dynamic (ensemble) information and crystallography provides a 'snapshot' and is often considered static. Solution approaches, such as limited proteolysis combined with mass spectrometry and small-angle scattering approaches (Chapter 3), also provide dynamic information. In cryo-ET and cryo-electron microscopy (cryo-EM) (Chapter 5), macromolecules are frozen in their native state, allowing for discrete selection of dynamic states to be visualized, albeit at lower resolution. Generally, NMR spectroscopy is utilized for small protein molecules that are flexible, X-ray crystallography for medium-sized proteins and complexes that are compact, whereas very large macromolecular assemblages or membranous protein structures are determined by cryo-EM. The largest issue separating cryo-EM and cryo-ET from crystallography, in addition to size and the limitations of crystal formation, is resolution. Cryo-EM has generally been considered a low-resolution technique, giving reconstructions around 15–30 Å, but with advances in sample handling, instrumentation, image processing and model building, near-atomic resolution structures are now being achieved. For cryo-ET the resolution achieveable is still low.

In reality, hybrid approaches, combining NMR, X-ray crystallography and cryo-EM, cryo-ET and solution data, are often adopted, which provides a powerful means of filling gaps which can arise in the structural characterization of large macromolecules. For example, in studies where large viruses cannot be crystallized, subcomponents can be crystallized to obtain high-resolution information, which can then be used to interpret the structure at lower resolution obtained by cryo-EM or cryo-ET. Or atomic structures obtained from homologous viral proteins/virus capsids can be used for 3D homology model building. These approaches permit the pseudo-atomic visualization of interaction interfaces between protein–protein subunits, protein–nucleic acids and protein–lipid in virus capsids and also the visualization of virus capsid–host interactions.

Combined with biochemical, biophysical and molecular biology analysis, structural studies indicate a high degree of fidelity in the steps that result in the assembly of mature infectious virus capsids (Chapter 10). They also show that the fundmental principles governing successful viral capsid assembly, efficient polymerization of CP subunits utilizing specific interface interactions that spontaneously terminate, often employ structural polymorphisms to facilitate the required interactions. Structural virology approaches have also been platforms for the elegant description of the virus infection process, from initial receptor attachment to the interaction of the capsid with host antibodies (Section 2), and provided the targets for therapeutic intervention and improved viral capsid vectors for gene delivery (Section 3).

This monograph is designed to provide a basic introduction to the use of structural virology and its applications in virus research towards functional annotation and is not intended to provide a detailed discussion of approaches utilized.

Mavis Agbandje-McKenna
Robert McKenna

Contents

SECTION 1

Chapter 1 Production and Purification of Viruses for Structural Studies **3**
Brittney L. Gurda and Mavis Agbandje-McKenna

1 Introduction 3
2 Expression Systems 4
Eukaryotic Systems 5
Plant Systems 9
Prokaryotic Systems 10
Cell-free Systems 12
Tissue Samples 12
3 Purification 12
Ultracentrifugation 13
Chromatography 14
4 Example Virus Capsid Production and Purification – Adeno-associated Virus Serotype 1 15
VLP Expression Using the BEVS 16
Production of VLPs in Sf9 Insect Cells 17
Purification of AAV1 VLPs from Infected Sf9 Cells 17
5 Summary 18
6 Acknowledgments 18
References 18

Chapter 2 Microscopic Analysis of Viral Cell Binding, Entry and Infection in Live Cells **22**
Colin R. Parrish

1 Introduction 22
2 Endocytosis, Cytoplasmic Transport and Viral Entry 24

RSC Biomolecular Sciences No. 21
Structural Virology
Edited by Mavis Agbandje-McKenna and Robert McKenna

© Royal Society of Chemistry 2011

Published by the Royal Society of Chemistry, www.rsc.org

3 Virus Labeling for Fluorescence Experiments – Allowing Tracking of Viral Particles, Components and/or Nucleic Acids 24
4 Receptor Attachment and Cell Entry 26
5 Membrane Association and Cell Surface Movement and Uptake 28
6 Receptor Signaling and Endosomal Uptake 31
7 Tracking Endosomal Trafficking of Particles within Live Cells 31
8 Low pH, Membrane Fusion and Other Plasma Membrane or Intra-vesicular Events 36
9 Trafficking of Viral Components Within the Cytoplasm – Role of the Cytoskeleton in the Direct Movement of the Viral Components 37
10 Nuclear Transport and Entry 38
11 Summary and Conclusions 38
References 39

Chapter 3 Probing Viral Capsids in Solution 41
Brian Bothner and Jonathan K. Hilmer

1 Introduction 41
2 Quaternary Dynamics 42
Maturation-associated Dynamics in a Bacteriophage 42
Maturation-associated Dynamics in Small RNA Viruses 44
Structural Transitions 45
3 Solution-phase Equilibrium Dynamics 45
Virus Particles are Dynamic 46
Proteolysis and Mass Analysis 46
4 Methods for Studying Viruses in Solution 48
Spectroscopy 50
Computation 53
Labeling Experiments 54
5 Summary 58
References 58

Chapter 4 Three-dimensional Structures of Pleiomorphic Viruses from Cryo-Electron Tomography 62
Alasdair C. Steven, Giovanni Cardone, Carmen Butan, Dennis C. Winkler and J. Bernard Heymann

1 Introduction 62
2 Cryo-electron Tomography: How It's Done 63
3 Resolution in Cryo-ET 64
4 Features of Pleiomorphic Viruses 68

5 Stowaways or Conscripts? Host Cell Proteins in Virus Particles 71
6 Tomographic Visualization of Cell Entry Events in Subcellular Systems 75
7 Perspective 76
Acknowledgments 78
References 78

Chapter 5 Structure Determination of Icosahedral Viruses Imaged by Cryo-electron Microscopy 81
Robert S. Sinkovits and Timothy S. Baker

1 Introduction 81
2 Image Digitization and Preprocessing 83
Recording Media and Image Digitization 83
Particle Boxing 84
Defocus Estimation 84
3 3D Image Reconstruction 85
Iterative Model-based Refinement and Automation 86
Starting Model/Structure 87
Determining Particle Origins and Orientations: Global and Local Refinement 87
Computing the 3D Reconstruction 90
Estimating the Resolution of the Reconstruction 91
Building a Starting Model from Scratch: the Random Model Computation 93
Hand Determination 94
4 Image Reconstruction Example – PsV-F 95
5 Summary 97
Acknowledgments 97
References 97

Chapter 6 X-ray Crystallography of Virus Capsids 100
Lakshmanan Govindasamy, Mavis Agbandje-McKenna and Robert McKenna

1 Introduction 100
2 Experimental Procedures 101
Crystallization 101
X-ray Data Collection and Processing 104
3 Phase Determination 108
Isomorphous Replacement 108
Molecular Replacement 109
4 Structure Refinement and Model Building 110
5 X-ray Structure Determination Example: AAV4 113

Virus Production, Purification and Crystallization 114
X-ray Diffraction Data Collection 114
Determination of Particle Orientation and Position 115
Phasing, Refinement and Model Building 115
6 Virus Database: VIPER 116
7 Summary 117
8 Acknowledgments 117
References 117

Chapter 7 Structural Studies of Viral Proteins – X-ray Crystallography 123
John Domsic and Robert McKenna

1 Introduction 121
2 Sample Preparation 122
Protein Expression 122
Protein Purification 122
3 Single-crystal Protein Crystallography 124
Crystallization 124
X-ray Data Collection 125
Structure Determination 126
4 Case Examples 127
Influenza Hemagglutinin 127
HIV Envelope Glycoprotein 129
HIV Reverse Transcriptase 130
5 Summary 133
References 133

Chapter 8 Solution NMR Spectroscopy in Characterizing Structure, Dynamics and Intermolecular Interactions of Retroviral Structural Proteins 137
Kang Chen and Nico Tjandra

1 Introduction 135
2 Experimental Methods 137
Sample Preparation 137
Resonance Assignments 138
Solution NMR Structure Determination 139
Spin Relaxation and Dynamics 141
3 Solution Structure and Fast Dynamics 144
HIV-1 Matrix–Capsid^{N} Fragment 145
RSV Capsid 147
EIAV Matrix 148
4 Ligand Interaction and Complex Structure 148
HIV-1 MA Coordinating PI(4,5)P2 150
MoMuLV Nucleocapsid Recognizing Duplex RNA 151
5 Slow and Functional Dynamics 153

PI(4,5)P2 Induced Sub-ms Dynamics on EIAV MA 153
CypA Catalysis on HIV-1 CA^N 154
6 Summary and Discussion 156
References 157

SECTION 2

Chapter 9 Evolution of Viral Capsid Structures – the Three Domains of Life 163
Reza Khayat and John E. Johnson

1 Introduction 163
2 The Adenovirus Lineage 164
The Adenovirus 164
The Bacteriophage PRD1 166
The *Paramecium bursaria* Chlorella Virus Type 1 (PBCV-1) 168
The *Sulfolobus* Turreted Icosahedral Virus (STIV) 169
3 The HK97 Lineage 171
The Bacteriophage HK97 171
The Herpesvirus 172
The *Pyrococcus furiosus* Virus (PfV) 174
The HK97 Fold 175
4 Viral Capsid Evolution 176
References 177

Chapter 10 Mechanisms of Icosahedral Virus Assembly 180
Adam Zlotnick and Bentley A. Fane

1 Introduction 180
2 Assembly in Viruses Without Scaffolding Proteins 181
The Basic Problem 181
A Simplest Case: HBV and Its Implications 183
A Not So Simple Case: CCMV and BMV 184
Assembly-active and -inactive Species: Allostery and Autostery 185
3 Post-nucleation Effects on Assembly: Mutations and Drugs 186
4 Scaffolding Protein-mediated Morphogenesis, P22 a Model System 188
Nucleation and Elongation 188
Portal and Minor Protein Recruitment 190
5 Size Determination and Fidelity 191
Bacteriophage P22 191
Phages P2 and P4, Reprogramming Size Determination 192

6 ϕX174, a Two-scaffolding Protein System 192
Early Assembly, the Internal Scaffolding Protein 193
The Relationship Between the Two Scaffolding Proteins 193
Functions of the External Scaffolding Protein 195
Size Determination and the Minor Vertex Protein H 196
7 Summary 197
References 197

Chapter 11 Mechanisms of Genome Packaging 203
Mark Oram and Lindsay W. Black

1 Introduction 203
2 The Biological Context of Phage DNA Packaging 204
3 Components of the Packaging Machinery 206
The Portal Protein 208
The Terminase Enzyme 209
The DNA Substrate and Packasome Assembly 210
4 Analysis of the Packasome as a Molecular Motor 210
Class 1 dsDNA Packaging 211
Nucleic Acid Packaging Systems of Other Viral Classes 212
5 Fluorescence Approaches to Packaging Motor Dynamics 213
6 Summary 214
Acknowledgments 217
References 217

Chapter 12 Attachment and Entry: Receptor Recognition in Viral Pathogenesis 220
Damian C. Ekiert and Ian A. Wilson

1 Introduction – Viral Receptors Mediate Attachment and Entry 220
2 Cell Surface Receptors 221
3 Viral Receptor Binding Proteins 224
4 Host Range, Tissue Tropism and Transmission 234
5 Summary 238
References 238

Chapter 13 Attachment and Entry: Viral Cell Fusion 243
Rachel M. Schowalter, Everett C. Smith and Rebecca Ellis Dutch

1 Introduction 243
2 Promotion of Membrane Fusion by Viral Fusion Proteins 244

Type I Viral Fusion Proteins 245
Type II Viral Fusion Proteins 247
Type III Viral Fusion Proteins 249
3 Regulation of Fusion Protein Activity 251
Regulation of Type I Fusion Protein Activity 251
Regulation of Type II Fusion Protein Activity 252
4 The Trigger of Fusion Protein Conformational Changes 253
The Low pH Trigger of Fusion 253
Triggering of Fusion Through Receptor Binding 255
5 Summary 256
References 257

Chapter 14 Structural Studies on Antibody–Virus Complexes 261
Thomas J. Smith

1 Introduction 261
2 Antibody Structure and Diversity 261
3 Mechanisms of *In vitro* Antibody-mediated Neutralization of Viruses 263
Aggregation 264
Virion Stabilization 264
Induction of Conformational Changes 264
Abrogation of Cellular Attachment 265
Other *In situ* Effects 265
Significance of *In vitro* Neutralization Mechanisms *In vivo* 266
Antibody–Virus Complexes 266
4 Summary 282
References 283

SECTION 3

Chapter 15 Development of Anti-HIV Drugs 293
Roxana M. Coman and Robert McKenna

1 Introduction: World Human Immunodeficiency Virus (HIV) Epidemic Status 293
2 HIV Genome and Structure 294
3 Viral Life and Replication Cycles 299
4 Antiretroviral (ARV) Therapy and Drug Resistance 301
5 Inhibitors 302
Reverse Transcriptase Inhibitors (RTIs) 302
Protease Inhibitors (PIs) 303
Integrase (IN) Inhibitors 303
Entry and Fusion Inhibitors 303
Highly Active Antiretroviral Therapy (HAART) 305
6 HIV Diversity 306

Groups and Subtypes 306
HIV-1 Non-B Subtypes 308
References 311

Chapter 16 Design of Capsid-binding Antiviral Agents Against Human Rhinoviruses 319
Chuan Xiao, Mark A. McKinlay and Michael G. Rossmann

1 Introduction 319
2 The Common Cold 320
3 Picornavirus Structure 320
4 Human Rhinovirus Receptors 322
5 Anti-HRV Compounds 324
6 Mechanism of Inhibition by Capsid-binding Compounds: Inhibition of Attachment 324
7 Mechanism of Inhibition by Capsid-binding Compounds: Inhibition of Uncoating 324
8 Drug-resistant Mutants 326
9 Computational Analyses 326
10 Development of an Effective Anti-HRV Drug 329
11 Acknowledgments 334
References 334

Chapter 17 Viral Vectors for Gene Delivery 338
David J. Dismuke, Steven J. Gray, Matthew L. Hirsch, Richard Samulski and Nicholas Muzyczka

1 Adenoviral Vectors 338
Capsid Structure and Viral Entry 339
Using Structural Information to Modify Ad Tropism 340
2 Retroviral Vectors 341
Capsid Structure 341
Using Structural Information to Target Lentiviruses 342
3 AAV Vectors 343
Capsid Structure 344
Viral Entry 345
Using Structural Information to Engineer Improved AAV Vectors 347
References 351

Subject Index 358

Section 1

CHAPTER 1

Production and Purification of Viruses for Structural Studies

BRITTNEY L. GURDA AND MAVIS AGBANDJE-MCKENNA

Department of Biochemistry and Molecular Biology, Center for Structural Biology, The McKnight Brain Institute, University of Florida, Gainesville, FL 32610, USA

1 Introduction

Advances in protein production and purification techniques over the past two decades have allowed the structural study of numerous proteins and macromolecular assemblages that would have otherwise been intractable to the necessary approaches (detailed in the following chapters). This chapter focuses on the production and purification of intact viral capsids (particles) with/without genome for structure determination. The production and purification of viral proteins for structure determination by X-ray crystallography and NMR spectroscopy are the subjects of Chapters 7 and 8, respectively. Crystallization is often considered a method of purification and a function of purity, often of a protein or virus capsid, and, as such, sample preparation for structure determination by X-ray crystallography places high demands on sample quality. Screening trials to identify the optimal crystallization conditions also require large quantities of sample compared with the majority of other structure determination approaches discussed in the subsequent chapters of this monograph. Virus samples produced for such analyses also have to be both stable and soluble in their storage buffer since degradation and aggregation are detrimental to the crystallization process. Hence this chapter will focus

RSC Biomolecular Sciences No. 21
Structural Virology
Edited by Mavis Agbandje-McKenna and Robert McKenna
© Royal Society of Chemistry 2011
Published by the Royal Society of Chemistry, www.rsc.org

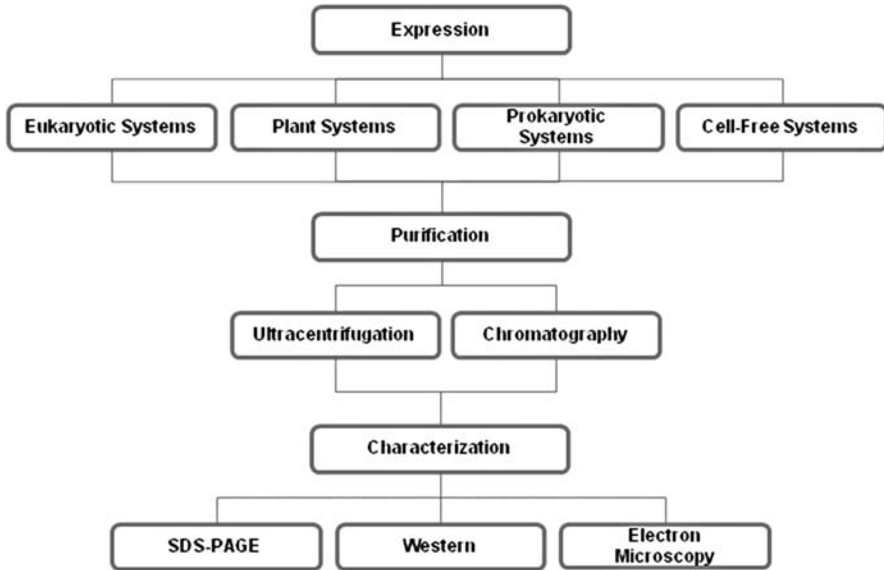


Figure 1 The steps involved in the expression, purification and characterization of virus capsids prior to structural analysis.

on methodologies to produce and purify virus capsids (Figure 1) in quantities suitable for structure determination by X-ray crystallography, with the premise that such a sample would also be suitable for structural or biophysical analysis using other methodologies.

2 Expression Systems

Most viruses are considered hazardous material in their wild-type (wt) infectious form (for information on safe handling and containment of infectious microorganisms and hazardous biological materials, see http://www.cdc.gov/biosafety) and are therefore often studied in a recombinant form. Significant effort has been extended into the development of heterologous expression systems to produce recombinant viral proteins which will assemble into viral capsids. The system selected for use is often dependent on the properties of the viral genes and the environmental requirements of the final product. However, the most important factor to consider is the capacity of the host cells to translate the RNA transcript, to ensure proper folding of the gene product and to sustain the protein(s) expressed in an intact and functional state.[1] Protein expression systems contain at least four general components: (1) the genetic elements necessary for transcription/translation and selection; (2) in vector-based systems, a suitable replicon: plasmid, virus genes, *etc.*; (3) a host strain containing the appropriate genetic traits needed to function with the specific

expression signals and selection scheme; and (4) the culturing conditions for the transformed cells or organisms.[2]

Eukaryotic Systems

Mammalian Cells

Since most viruses currently studied are of human or animal origin, mammalian tissue culture is an ideal source to generate viral capsids for structural studies which are generally aimed at functional annotation. In this system, proper folding is achieved and modifications such as complex glycosylation, phosphorylation, acylation, acetylation and γ-carboxylation are obtained. However, yields can be low, depending on gene product(s), ranging from 0.1 to $100\,\mathrm{mg\,L^{-1}}$ of culture volume. For some of the structural approaches discussed in Section 1 of this monograph, low yields may not be a problem since small amounts of sample are adequate. However, low yields can become problematic in crystallization, especially with a virus that does not have an established crystallization condition. In such a situation, numerous preparation steps may be required to obtain the quantities needed to screen crystallization conditions efficiently. Supplies and reagents can then become expensive, depending on individual cell line requirements. In addition, considerable time and resources can be spent on the construction of a suitable expression system and equally on optimization for suitable yields. In such situations, it is always advisable to seek the expertise of an established molecular biologist before designing new constructs.

Established cell lines and protocols exists for many different tissue systems and, although most of these cell lines are derived from human or mouse tissues, other mammalian cell culture lines are available, such as monkey, raccoon, horse, pig and rabbit. The American Type Culture Collection (ATCC) has over 3400 cell lines from 80 different species, including over 950 cancer cell lines (http://www.atcc.org/). Other cell suppliers include the Health Protection Agency Culture Collections (HPACC; http://www.hpacultures.org.uk/), the German Research Center for Biological Material (DSMZ; http://www.dsmz.de/) and the Riken BioResource Center Cell Bank (Riken; http://www.brc.riken.jp). It is strongly recommended that investigators purchase cell lines from recognized centers such as these listed above to ensure pure, authentic and quality controlled cell lines. The decision to use cells directly from an organism, *i.e.* primary cells or an immortalized cell line, should be based upon requirements of the virus system and available current protocols. As discussed below, there are three main approaches for virus production in mammalian cell lines: (i) infection of permissive cell lines with wt virus, (ii) transfection of cells with plasmid constructs containing viral genome sequences and (iii) viral vector systems which expression heterologous viral genes.

Although the majority of viruses currently studied are obtained from recombinant expression systems (see below), direct infection of cell lines with wt virus can be used to generate suitable quantities of sample for structural studies under certain conditions and for well-characterized viral systems.

For example, the human rhinovirus 3 (HRV3) virion particles used for determining its structure were purified from virus-infected HeLa cells (immortalized human cancer cells). The atomic structure of HRV3 was initially determined to 3 Å,[3] and later refined to 2.15 Å.[4] It was reported that 10–12 L of HeLa cells (at 6–8 $\times 10^5$ cells mL^{-1}) were used to generate the amount of virus necessary to carry out crystallization and structure determination. Echovirus-1, also of the *Picornaviridae* family, was also successfully produced in HeLa cells for its structure determination to $\sim$3.55 Å resolution.[5]

In the use of plasmid constructs, one or more plasmids usually containing capsid proteins alone and, if needed, replication factors, are used to transfect cells, which results in the assembly of virus-like particles (VLPs). Often, another plasmid is added when a packaged gene is desired, *e.g.* reporter gene, or if genome is needed to produce stable virions. Recovered virus can either be purified for structural studies or, if infectious, used to infect permissive cells for continual propagation of virions. As an example, molecular clones containing the capsid sequence of canine parvovirus was used for the transfection of Norden Laboratories feline kidney cells (NLFK)[6] to produce particles for X-ray crystallographic structural studies to 3.2 Å resolution.[7] For the crystallographic structure determination of the immunosuppressive strain of minute virus of mice (MVMi), infectious virions were harvested from plasmid transfected cell lines and subsequently propagated in a permissive cell line to produce virus for crystallization.[8]

The development of heterologous surrogate expression systems for virus capsid production has enabled researchers to overcome the lack of efficient expression in homologous systems for several viruses of interest. As an example, for hepatitis C virus (HCV), a herpes simplex virus-1 (HSV-1)-based amplicon vector system that expresses HCV capsid proteins and the two envelope proteins, E1 and E2, under the HSV-1 IE4 promoter was developed.[9] This system has several advantages; (i) the ability to infect a wide range of cells, without the limitation of transfection efficiency, including primary cells in a quiescent state, (ii) the simplicity of cloning desired genes into amplicons, (iii) the high capacity of incorporation of exogenous sequences in the vector genome and the transfer of high copy numbers of the exogenous gene and (iv) the potential for using amplicons in vaccine design and development.[10] A mini-review has covered HSV amplicons from genomes to engineering.[11] Norovirus is another example of a non-cultivable virus that remained refractory to structural studies due to the lack of a reverse genetics system and a permissive cell line until recent advances. A novel expression strategy, which combined the use of a two baculovirus transactivation system to deliver viral cDNA and an inducible DNA polymerase (pol) II promoter, led to the ability to grow this virus in several cell lines, including HepG2, BHK-21, COS-7 and HEK293T cells.[12,13]

Yeast Cells

Among the microbial eukaryotic host systems, yeasts can combine the advantages of unicellular organisms (*e.g.* ease of genetic manipulation and

growth) with the capabilities of a protein processing typical of eukaryotic organisms (*e.g.*, protein folding, assembly and posttranslational modifications).[14] The majority of recombinant proteins produced in yeast have been expressed using *Saccharomyces cerevisiae*. More commonly referred to as baker's or budding yeast, *S. cerevisiae* was the first eukaryote to have its entire genome sequenced[15] and is still today considered a model organism. A scientific database has been established for *S. cerevisiae* and is available at http://www.yeastgenome.org/. With its biochemistry, basic genetics and cellular biology already well established, this simple eukaryote has become a major tool in answering questions of fundamental biological importance and is a central player in post-genomics research.

Appealing aspects of the yeast expression system are its rapid cell growth (with a doubling time of ~90 min), simple growth media, secretion of recombinant proteins to the medium and glycosylation capability. N-linked glycosylation is minimal with high mannose, but O-linked modifications appear similar to mammalian cells. Phosphorylation, acetylation and acylation are also present. Protein yields are comparable with the baculovirus system (see below) at ~10–200 mg L^{-1} depending on recombinant gene properties. Issues in large-scale protein production involving *S. cerevisiae* appear to be hyperglycosylation and retention in the periplasmic space.[16,17] This ultimately leads to a loss of final protein due to retention and degradation. The search for alternative hosts has led to the use of 'non-conventional' yeasts in expression protocols. The most established examples include *Hansenula polymorpha*, *Pichia pastoris*, *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Pichia methanolica*, *Pichia stipitis*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces bailaii*, *Candida boidinii* and *Schwanniomyces* (*Debaryomyces*) *occidentalis*.[14] These systems are broken down even further into two categories: methyltrophic, *e.g. P. pastoris*, and non-methyltrophic, *e.g. S. cerevisiae*. These categories are based on the fermentation processes involved and generally dictate the promoter that should be used in the experimental design. The choice of yeast host is one of the most important determinants of the success of the entire project, and many reviews debating the subject can be found in the current literature. Generally, the expression of foreign proteins in yeasts consists of (i) cloning of a foreign protein-coding DNA sequence within an expression cassette containing a yeast promoter and transcriptional termination sequences and (ii) transformation and stable maintenance of this DNA in the fusion host.[14] The transformation process is highly dependent on the yeast strain and detailed studies should be conducted in order to achieve high-efficiency transformation.

This system is extensively used for studying biological processes in higher eukaryotes and also allows replication of eukaryotic viruses. The first eukaryotic virus for which replication and genome encapsidatation was conducted in *S. cerevisiae* was brome mosaic virus (BMV), a positive strand RNA [(+)RNA] virus that infects plants.[18,19] The BMV VLPs were subsequently purified for structure-to-function studies using cryo-electron microscopy (cryo-EM) studies.[19] Other (+)RNA viruses that have been successfully replicated in *S. cerevisiae* include the plant viruses tomato bushy stunt virus and carnation

Italian ringspot virus and animal viruses Flock House virus (FHV) and Nodamura virus.[20] Human papillomavirus-16 (HPV-16) VLPs have also been successfully expressed in the yeast system[21] in addition to the bovine papillomavirus-1 (BPV-1).[22,23] The yeast virus L-A was isolated and purified from *S. cerevisiae* and the structure was solved to 3.4 Å resolution.[24]

Insect Cells

Originally isolated from the alfalfa looper (*Autographa californica*) insect, *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) is the most widely used and best characterized baculovirus for recombinant gene expression (a recent review on baculovirus molecular biology is available[25]). The rather large genome ($\sim$134 kbp[26]) can stably accommodate an insertion of $\sim$38kb,[27] making expression of large genes possible. This virus is also known to infect several other insect species including *Spodoptera frugiperda*. The most commonly used insect host cell lines, Sf9 and Sf21AE, are derived from *S. frugiperda* pupal ovarian tissue[28] and the BTI-Tn-5B1-4 line, also known as 'High 5 cells', derived from *Trichoplusia ni* egg cell homogenates.[29] The wt nucleopolyhedrovirus (NPV) produces small inclusion bodies composed of a polyhedron protein which allows for the encapsulation of many virions into a crystalline protein matrix. This protein is expressed in the very late phase of gene expression and is controlled by a very strong promoter, the polydron promoter (a review on baculovirus late expression factors is available[30]). The baculovirus expression vector system (BEVS)[31,32] takes advantage of this very strong polyhedron promoter to drive foreign protein expression. It has also been shown that the non-structural p10 protein is expressed at similar levels in the same very late phase of expression. Both proteins have been shown to be non-essential in the production of baculovirus particles,[33,34] making the replacement of their open reading frame (ORF) ideal for use in foreign gene expression.

The coupling of the very strong polyhedron promoter with a foreign gene-coding region results is the production of high levels of recombinant protein ($\sim$5–200 mg L^{-1}) in a relatively short amount of time using the BEVS. Since the baculovirus genome is generally considered too large to insert the foreign gene of choice by direct ligation, transfer vectors are used. There are many different vectors available for gene insertion, which are variants of a basic design (a review appeared recently[35]). These offer single gene, multiple genes and fusion gene expression. Multiple copies of the promoter can also be engineered into BEVS for the expression of multiple recombinant proteins concurrently in infected cells,[36,37] which permits the assembly of structures that are made up of heterologous proteins, such as viruses.

Advances in experimental design such as a wide variety of transfer vectors, simplified recombinant virus isolation and quantification methods, advances in cell culture technology and commercial availability of reagents have led to the increased use of BEVS for recombinant viral capsid protein production. Belyaev and Roy[37] were able to construct a multiple gene transfer vector which