

Chapter 11

Mammalian Expression of Isotopically Labeled Proteins for NMR Spectroscopy

Mallika Sastry, Carole A. Bewley, and Peter D. Kwong

Abstract NMR spectroscopic characterization of biologically interesting proteins generally requires the incorporation of $^{15}\text{N}/^{13}\text{C}$ and/or ^2H stable isotopes. While prokaryotic incorporation systems are regularly used, mammalian ones are not: of the nearly 9,000 NMR macromolecular structures currently deposited in the Protein Data Bank, only a handful (<0.5%) were solved with proteins expressed in mammalian systems. This low number of structures is largely a reflection of the difficulty in producing uniformly labeled, mammalian-expressed proteins. This is unfortunate, as many interesting proteins require mammalian cofactors, chaperons, or post-translational modifications such as N-linked glycosylation, and mammalian cells have the necessary machinery to produce them correctly. Here we describe recent advances in mammalian expression, including an efficient adenoviral vector-based system, for the production of isotopically enriched proteins. This system allows for the expression of mammalian proteins and their complexes, including proteins that require post-translational modifications. We describe how this system can produce isotopically labeled ^{15}N and ^{13}C post-translationally modified proteins, such as the outer domain of HIV-1 gp120, which has 15 sites of N-linked glycosylation. Selective amino-acid labeling is also described. These developments should reduce barriers to the determination of NMR structures with isotopically labeled proteins from mammalian expression systems.

Abbreviations

CHO	Chinese Hamster Ovary cells
HEK	Human Embryonic Kidney cells
BHK21	Baby Hamster Kidney cells
HIV-1	Human Immunodeficiency Virus Type 1
RSV	Rous Sarcoma Virus
CMV	Cytomegalovirus
PEI	Polyethyleneimine
ITR	Inverted terminal repeat

M. Sastry (✉) • P.D. Kwong
Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health,
40 Convent Drive; Building 40, Room 2613B, Bethesda, MD 20892-3027, USA
e-mail: sastrym@mail.nih.gov

C.A. Bewley
Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive
and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

BGHpA	Bovine growth hormone polyadenylation signal
SPR	Surface plasmon resonance
PBS	Phosphate buffered saline
ATCC	American Type Culture Collection
DMEM	Dubelco's modified eagle media
FBS	Dialyzed fetal bovine serum
CAR	Coxsackie-and Adenovirus Receptor.

11.1 Introduction

NMR spectroscopic characterization of proteins generally requires the incorporation of ^{15}N , ^{13}C , or ^2H stable isotopes. In most cases, full length proteins or their individual domains are expressed and purified from prokaryotic expression systems, and a number of these methods are discussed in Chap. 1 of this book. An analysis of the macromolecular structures deposited in the PDB (December 2011 release) reveals that over 99% of proteins studied by NMR spectroscopy are expressed prokaryotically (Table 11.1). A large number of biologically interesting proteins, however, require eukaryotic cofactors, chaperons or post-translational modifications such as glycosylation for proper folding and activity, and bacterial expression systems may lack the necessary cellular machinery to produce correctly folded functional proteins that are suitable for NMR spectroscopy. Why not express and purify post-translationally modified proteins from eukaryotic sources? Traditionally, eukaryotic systems such as insect, lower eukaryotes, and mammalian cells produce low yields of the desired protein, are time consuming, and/or are prohibitively expensive. Recent advances in expression techniques, such as multi-host vectors and the use of codon-optimized genes, have however reduced barriers to the evaluation and optimization of gene expression in different hosts. Furthermore, use of bioreactors and advances in cell culture techniques (such as the development of large scale mammalian transient expression systems) have allowed milligram- to gram-scale production of proteins suitable for structural analysis. Indeed, recombinant proteins produced from mammalian cells such as Chinese Hamster Ovary (CHO) and Human Embryonic Kidney (HEK) cells are routinely used in cryo-EM and X-ray crystallographic studies. An analysis of the PDB reveals that ~98% of the entries that used proteins obtained from mammalian expression systems are for structures solved by X-ray crystallography.

Despite these advances the production of isotopically enriched proteins suitable for NMR spectroscopy has been stymied by the absence of a suitable expression system as well as the cost of labeled media. We recently reported the adaptation of an efficient adenoviral vector-based mammalian expression system for the production of isotopically enriched proteins [1]. The high yields obtained from the adenoviral coupled mammalian expression system balanced the cost of labeled media and allowed for the production of post-translationally modified proteins suitable for NMR.

This chapter provides a summary of different systems used for mammalian expression and explicit details for the adenoviral vector-based expression system. We compare the adenoviral vector-based expression system to other methods – through a case study of a variant of the HIV-1 gp120 outer domain comprising 230 amino acids, 15 potential sites of *N*-linked glycosylation, and four disulfide bonds – and extend our previous characterization of this expression system to include selective amino-acid labeling.

11.2 Overview of Mammalian Expression

Recent development of eukaryotic systems including yeast, insect, and mammalian cells and associated vectors – which allows for the transient, inducible or constitutive expression of a target gene – has allowed researchers to investigate the possibility of overexpressing proteins in eukaryotic systems [2–4].

Table 11.1 Protein Data Bank^a entries solved by solution NMR spectroscopy

Experimental method	Total	Protein	Mammalian expression	Isotopically enriched mammalian expression
Solution NMR spectroscopy	9,210	8,052	8	4

^aDecember 2011 release

Mammalian cells have the necessary co-factors, chaperons and cellular machinery to produce correctly folded post-translationally modified functional proteins. These systems are currently used to produce proteins for therapeutic purposes and use primarily Chinese Hamster Ovary (CHO), Human Embryonic Kidney (HEK), Baby Hamster Kidney (BHK21), human fibrosarcoma (HT1080) and human lymphoma (Namalwa) cells. We describe expression methods in the following sections and provide details for production of isotopically enriched proteins using the HIV-1 gp120 outer domain as an example from the adenoviral vector-based mammalian expression system.

11.2.1 *Transient Expression in Mammalian Cells*

Transient expression of a target gene is defined as the temporary production of a protein. Transfection is achieved by well-established techniques wherein plasmid DNA containing the target gene is mixed with calcium phosphate, a cationic lipid, or other reagents [5] such as Lipofectamine (Fig. 11.1a). DNA of interest is introduced into the cells either by transiently opening pores or by fusion of liposomes with the cell membrane. Transfection can also be achieved by electroporation [6]. The factors governing gene expression in eukaryotes are similar to those in the prokaryotic systems discussed in preceding chapters. A strong promoter such as the SV40 early promoter, the Rous Sarcoma Virus (RSV) promoter or the cytomegalovirus (CMV) very early promoter can be used to increase protein expression [7]. Transient transfection of HEK 293E cells with polyethyleneimine (PEI) mediated transfections has been used to obtain milligram to gram quantities of secreted proteins [8, 9]. Transient transfection can be achieved in adherent as well as suspension cells, with the possibility of scale up [10, 11]. These proteins are functionally active and have been used in biophysical and crystallographic studies [12–14]. This method is simple to implement and requires very little infrastructure. In principle, protein expression from adherent cells using ¹⁵N, ¹⁵N/¹³C labeled media is feasible although potentially prohibitively expensive. Lack of suitable labeling media for suspension cells complicates production of labeled proteins from transient expression. As a result a number of investigators have turned to constitutive expression in mammalian cells.

11.2.2 *Constitutive Expression in Mammalian Cells*

Plasmid DNA containing the target gene and a strong promoter (SV40, RSV or CMV) is co-transfected with another plasmid containing a selectable marker (e.g. hygromycin, tetracycline, neomycin or G418) [15, 16]. The mammalian cells expressing the selectable marker are screened for drug resistance and a clonal cell line that expresses the protein of interest is established (Fig. 11.1b). Protein expression is dependent on the toxicity of the heterologous protein, the promoter used and the position of gene integration. A number of research groups have attempted to produce isotopically labeled proteins from mouse or Chinese Hamster Ovary (CHO) cell lines using either a mixture of algal/bacterial hydrolysates [17] or a mixture of labeled amino acids [18–21]. Unfortunately these methods result in low yields and can be prohibitively expensive. There are only four reports in the protein data bank

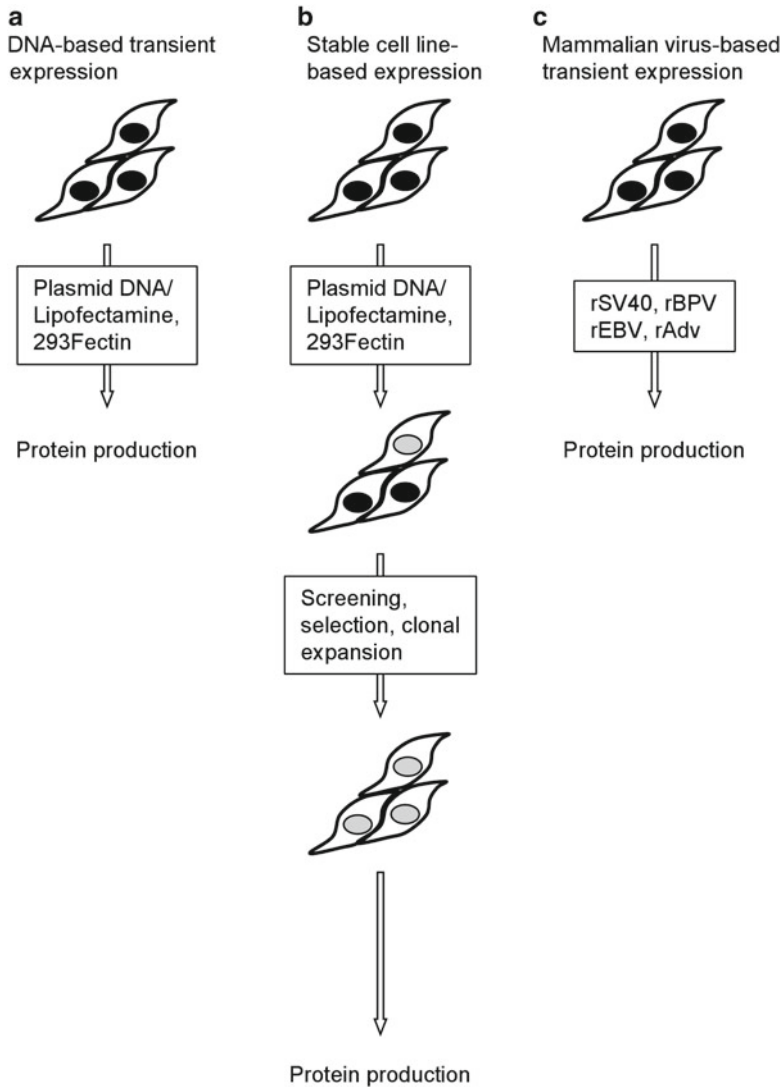


Fig. 11.1 Mammalian expression systems used for protein production. An overview of expression systems used to obtain proteins from (a) transient, (b) stable cell line based mammalian expression and (c) mammalian viral vectors based transient expression

(1URK, 1GYA, 1KLA, 1AH1) for structures obtained using heteronuclear NMR experiments from mammalian cell lines (Table 11.1) [22–25]. Researchers have thus focused on partial [26] or amino-acid type-specific labeling [27–29] of proteins from mammalian expression systems.

11.2.3 Transient Expression Using Mammalian Viruses

SV40, poxviruses, herpesviruses, papillomaviruses and adenoviruses have been used as expression vectors [15, 30–33]. Viral vectors wherein nonessential genes are substituted for a foreign gene [34] as well as vectors resulting in defective viral genomes can be used for transient protein expression by

infecting a wide range of mammalian cells (Fig. 11.1c). Protein expression is driven by viral promoters and, at the expense of production of cellular proteins, can increase the overall yield of the heterologous protein. The following section focuses on the adenoviral expression system that we have adapted to produce isotopically enriched glycoproteins.

11.3 Recombinant Adenoviruses as a Tool to Obtain Isotopically Labeled Proteins

Adenoviruses are double-stranded DNA viruses. Their genome comprises ~36 kb of linear double-stranded DNA [35]. Adenoviral genome that includes the E1-E4 region is transcribed early in the life cycle and is involved in the replication of the virus. Deletion of the E1 region renders the virus replication incompetent and thus safe to use as a gene delivery vector. A further deletion of the E3 region allows for insertion of up to 8-kilobases of recombinant transgenes into the E1 region [32, 36, 37]. An E1/E3 deleted virus can thus be used to deliver target genes with very high efficiency to many different cell types. A very high efficiency of transfection and specific translational discrimination between viral and cellular mRNA together facilitate the exceptional expression of adenovirus-vectored proteins from mammalian cells [38, 39].

11.3.1 Design and Composition of Recombinant Adenoviral Vectors

A schematic outline of the design and composition of an adenoviral vector is shown in Fig. 11.2. Specifically, the adenoviral cosmid (pVRC1194) consists of 9.2–100 m.u (map units) of the adenoviral genome, along with a deletion in the E3 region and a loxP site at 9.2 m.u. The shuttle vector (pVRC1290) into which the target gene is cloned, contains the adenoviral 5' inverted terminal repeat (ITR), a 0–1 m.u packaging signal followed by the target gene, the bovine growth hormone polyadenylation signal (BGHpA), a loxP site and 9.2–16.1 m.u of the adenoviral genome [40]. The shuttle plasmid and the adenoviral cosmid are linearized and recombined *in vitro* using Cre recombinase (Novagen-EMD Biosciences, Madison, Wisconsin) to obtain recombinant adenoviral genome. Recombinant adenoviral DNA can be purified by standard methods [41, 42]. The recombinant adenoviral DNA obtained from the Cre-Lox recombination reaction contains the target gene flanked by adenoviral sequences. These flanking sequences consist of DNA packaging signal as well as the adenoviral genome to reconstitute defective adenoviruses. Since generation of recombinant adenovirus can be a lengthy process, prior to embarking on recombinant adenoviral production, construct design and functional properties of the target protein were tested using transient expression in either HEK293 adherent or suspension cells.

11.3.2 Generation of Adenoviruses Containing Gene of Interest

To produce recombinant adenovirus containing the target gene, recombinant Ad5 DNA obtained from the Cre-Lox reaction is transfected into helper 293 cells (Fig. 11.2) using calcium phosphate transfection methodology. Since the E1 region of the adenoviral DNA is deleted, viral production can only take place in mammalian cells such as HEK293 with endogenous E1 proteins that can complement the defect in the recombinant adenovirus [44]. HEK293 cells transfected with recombinant adenoviral

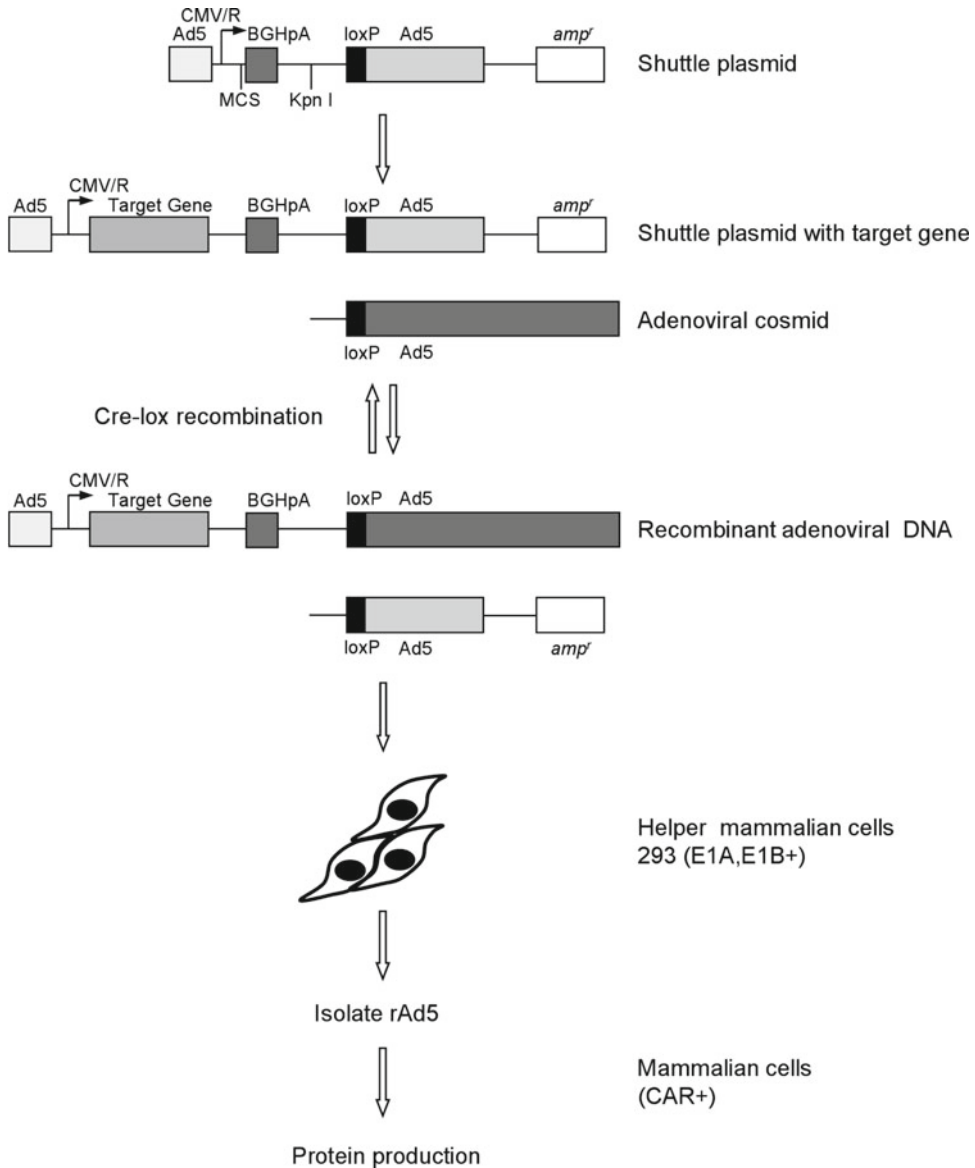


Fig. 11.2 Schematic overview of the adenoviral based mammalian expression system. The target gene is cloned into a shuttle plasmid (pVRC1194) using the restriction sites present in the multiple cloning site (*MCS*); linearized adenoviral cosmid DNA (pVRC1290) and shuttle plasmid are then recombined *in vitro* with Cre-Lox recombinase to obtain recombinant adenoviral genome (Aoki and Nabel [40]). The recombinant adenoviral type 5 DNA is transfected into helper mammalian cells and recombinant adenovirus is isolated using well established methods [41, 42]. Target protein production is achieved by infecting CAR⁺ mammalian cells. There are reports in the literature that Ad3 entry may be mediated by CD46 positive cells [43]

DNA are monitored for plaque formation for up to 10 days. Upon plaque formation, cells are harvested, the recombinant virus is isolated and protein expression of the desired gene tested by infecting A549 lung carcinoma cells with the crude virus. Protein expression of the target gene can be monitored and confirmed by western blots. Once protein expression is confirmed, the crude supernatant or cell lysate

can be further evaluated for the presence of a correctly folded protein of interest by surface plasmon resonance (SPR) analysis [45, 46] or equivalent technique. Once protein expression is confirmed, the crude virus preparation is used to infect low passage HEK293 cells to produce recombinant adenovirus. Specifically, HEK293 cells are seeded at 2.0×10^7 cells per 15 cm plate 24 h prior to infection. The cells should be ~90% confluent at the time of infection. Cells are harvested ~30 h post infection with the crude virus. Harvested cells are spun down at $311 \times g$ for 10 min, the cell pellet is washed twice with phosphate buffered saline (PBS) and subsequently resuspended in 10 mM Tris-HCl pH 8.0.

11.3.3 Isolation and Purification of Recombinant Adenoviruses

Recombinant virus is released from the cells by three cycles of freezing in a dry ice/ethanol bath and thawing at 37°C. The cell lysate is spun down at $552 \times g$ for 10 min. The Ad5 virus present in the supernatant is further purified twice using CsCl gradient centrifugation, followed by a desalting column to remove the CsCl [41, 47–49]. The purified adenovirus is quantitated and aliquoted aseptically; these aliquots can be stored at –20°C and used as needed for protein expression. The purified adenovirus is an infectious agent and each researcher should consult their institution's guidelines for handling biohazardous material.

11.3.4 Growth Characteristics of Cell Lines

HEK293 and A549 cells can be obtained from American Type Culture Collection (ATCC). Primary HEK cells are the best hosts for the replication of human adenoviruses [50]. A549 cells are lung carcinoma cells; these cells grow in monolayers and can grow in Dubelco's modified eagle media (DMEM) supplemented with 10% heat inactivated, dialyzed fetal bovine serum (FBS) and 1% penicillin/streptomycin. Recombinant adenoviruses can infect a large number of cell types that express the Coxsackie-and Adenovirus Receptor (CAR) protein. Because, adenoviruses also increase glycolysis in continuous cell lines and thereby induce the cells to produce large quantities of acid [51] the DMEM is buffered with 25 mM Hepes to maintain a neutral pH during cell culture. As in the case of all eukaryotic cells, A549 cells require a sterile environment and must be kept from overgrowing and losing viability.

Depending on the post-translational modifications present in a given target protein, a researcher may need to investigate protein expression in different cell lines; each of these may have specific requirements for growth media. As a result, understanding cell growth in the expression media to be used is one of the first steps in obtaining isotopically labeled proteins. For our case study A549 lung carcinoma cells are brought up in DMEM, ^{15}N -CGM6000, or $^{15}\text{N}/^{13}\text{C}$ -CGM6000 media supplemented with 10% heat inactivated, dialyzed FBS and 1% penicillin/streptomycin and seeded in six well plates at 0.2×10^6 cells/well and incubated at 37°C for 6 days. Aliquots are taken every 24 h and cells counted for each media type. Perdeuteration or fractional deuteration of amino acid side chains is essential for the study of large proteins [52]. This can be achieved in principal by the growth of mammalian cells in deuterated media. A549 cells can be readily adapted to 20% D_2O -containing DMEM, 10% heat inactivated FBS and 1% penicillin/streptomycin [53]. The adapted cells are seeded at 1×10^6 cells/10 cm plate in DMEM containing 20, 45 or 70% D_2O . Viable cells are counted every 24 h for 6 days to obtain growth curves. Once it has been established that the cells are viable in the chosen growth media for the necessary time period, the researcher may test protein expression in the chosen cell line.

11.3.5 Protein Expression

Once the growth characteristics and media requirements of the favored cell line have been established, small scale protein expression can be initiated in six well plates. This is necessary prior to embarking on large scale protein production to assess the expression conditions as well as proper folding and activity of the desired protein. Expression of glycoproteins from mammalian cells can result in heterogeneous glycosylation due to variation in the occupancy as well as the nature of glycan. The glycosylation pattern obtained from mammalian cells can be high-mannose, hybrid or complex in nature. In our case study, we used a combination of kifunensine, a potent inhibitor of α -mannosidase I [54], as well as swainsonine, a potent inhibitor of α -mannosidase II [55], to obtain Endoglycosidase H-sensitive high mannose glycans [56, 57].

For small scale protein expression using the adenoviral expression system in six-well plates, A549 adherent cells are routinely maintained in DMEM. Typically, cells are seeded at 0.8×10^6 cell/well on Day 1 in fresh DMEM containing 10% heat inactivated dialyzed FBS, 1% penicillin/streptomycin and allowed to grow overnight at 37°C and 5% CO₂. The following day (Day 2), media is replaced with labeled ¹⁵N, ¹⁵N/¹³C CGM6000 or fresh DMEM containing 10% heat inactivated dialyzed FBS and 1% penicillin/streptomycin. A549 cells are then infected with recombinant adenovirus (rAd) to a final concentration of 2,500 particles/cell. In general, a high multiplicity of infection is used so that all cells in the culture are synchronously infected. Protein expression of a cytoplasmic or secreted protein is monitored 72–96 h (Days 4–5) post infection either by harvesting cells or the culture media, respectively, and testing by a method of choice such as surface plasmon resonance, immunoprecipitation or western blotting.

For large scale production of unlabeled or isotopically enriched glycoprotein, A549 cells are seeded at 12–15 $\times 10^6$ cells/15 cm plate on Day 1 in fresh DMEM containing 10% heat inactivated dialyzed FBS, 1% penicillin-streptomycin and allowed to grow overnight at 37°C and 5% CO₂. The following day (Day 2), media is replaced with ¹⁵N, ¹⁵N/¹³C CGM6000 or fresh DMEM containing 10% heat inactivated dialyzed FBS, 1% penicillin/streptomycin as well as the glycosidic inhibitors kifunensine (12.5 mg/L) and swainsonine (5 mg/L). A549 cells are infected with recombinant adenovirus containing the gene for HIV-1 gp120 outer domain (rAdOD) 1–2 h later to a final concentration of 2,500 particles/cell. For secreted proteins, the culture supernatant is harvested 96–108 h post infection; cell debris is spun down at 365 \times g. The culture supernatant is filtered aseptically through a 0.2 μ m filter and, if necessary, the cell free supernatant can be concentrated five to tenfold by tangential flow filtration [58] prior to protein purification. The expressed protein (HIV-1 gp120 outer domain) is purified by immobilized nickel- and antibody b12-affinity chromatography. Fractions containing the outer domain are pooled, concentrated, dialyzed against PBS and deglycosylated using EndoH_f followed by size-exclusion chromatography to obtain the deglycosylated protein [1]. For cytoplasmic proteins; culture supernatant is removed and adherent A549 cells are harvested following treatment with Trypsin-EDTA. Cells are pelleted at 300 \times g, washed twice with PBS and subsequently lysed with cell lysis buffer (Cell Signaling) and the protein of interest purified by the method of choice.

11.3.6 Selective Labeling of Specific Amino Acids

Mammalian expression systems as well as insect cells are currently used to obtain amino-acid type specific labeling of proteins [27, 59, 60]. A similar approach is possible in the adenoviral mediated expression system. The methodology to produce selectively labeled protein is similar to that described in Sect. 11.3.5. CGM-6750 containing ¹⁵N labeled glycine with all other amino acids and media components unlabeled can be used to produce protein specifically labeled with ¹⁵N glycine (Fig. 11.5a). We did observe minor peaks in the glycine enriched sample possibly due to scrambling to serine or cysteine residues. CGM-6750 containing ¹⁵N labeled valine with all other media components unlabeled

can be used to produce protein selectively labeled with ^{15}N valine (Fig. 11.5b) We did observe minor peaks in valine enriched spectra possibly due to scrambling to isoleucine or leucine residues. We estimate yields of 60 and 49 mg/l of pure glycosylated ^{15}N glycine and ^{15}N valine enriched outer domain proteins, respectively. These yields balance out the high cost of the labeled media and allow for the production of selectively labeled proteins.

11.3.7 Quantification of Isotope Incorporation

Typically quantification of isotopic incorporation is obtained from MALDI TOF spectra of an enriched protein. However for glycoproteins such as the HIV-1 gp120 outer domain the inherent heterogeneity in glycosylation makes such an analysis very complicated. Thus, a glycoprotein is subjected to proteolytic digestion to identify peptide fragments devoid of N/O linked glycans followed by mass spectrometric analysis to obtain the isotope incorporation levels [1, 61].

11.4 NMR Characterization of Expressed Protein

Initially, the unlabeled full length or a target protein domain expressed either transiently or from the adenoviral expression system can be characterized using 1D or 2D NMR spectroscopy. Thus, in our example the HIV-1 gp120 outer domain obtained from adenovirus expression system exhibits a one dimensional ^1H NMR spectrum (Fig. 11.3a) that is well dispersed with resolved, upfield-shifted methyl protons as well as a relatively well dispersed amide region, indicative of a well folded protein. The ^1H - ^{15}N HSQC spectra of the outer domain is also of very high quality and exhibits resolved backbone and side-chain amides (Fig. 11.3b). The ^1H - ^{13}C HSQC (Fig. 11.3c) exhibits very good chemical shift dispersion along with upfield shifted methyl resonances, indicative of a structured protein.

To assess uniform $^{15}\text{N}/^{13}\text{C}$ incorporation necessary for assignment of backbone and side chain atoms, two data sets were recorded. In Fig. 11.4a, a high quality ^1H - ^{13}C plane of a 3D HNCACB spectrum acquired at 900 MHz and 25°C shows correlations from the amide protons to the C_α and C_β carbons and demonstrates the uniform enrichment of C_α and C_β , a critical experiment for structural characterization of a protein by NMR spectroscopy. The ^1H - ^{13}C projection of a 3D HNCO spectrum (Fig. 11.4b) showing correlations between the amide NH protons and the backbone CO resonances demonstrates conclusively that the adenoviral expression system provides sufficient isotope enrichment to allow acquisition of triple resonance experiments to obtain full backbone and side chain resonance assignments. In many instances, especially for large proteins, protein-protein complexes or membrane proteins, selective labeling of specific amino acids along with perdeuteration or random fractional deuteration becomes necessary to study specific interactions or complement assignments in extremely crowded regions of 3D spectra. Selectively labeled outer domain spectra enriched with ^{15}N glycine and valine (Fig. 11.5a, b) provide further proof that the adenoviral expression system is suitable for obtaining isotopically enriched proteins for structural characterization of proteins and protein complexes by NMR spectroscopy.

11.5 Conclusions

Characterization of the structural and dynamic properties of post-translationally modified proteins by NMR spectroscopy has been hindered by the absence of suitable expression systems to obtain isotopically enriched proteins. Recent advances in molecular biology and cell culture techniques

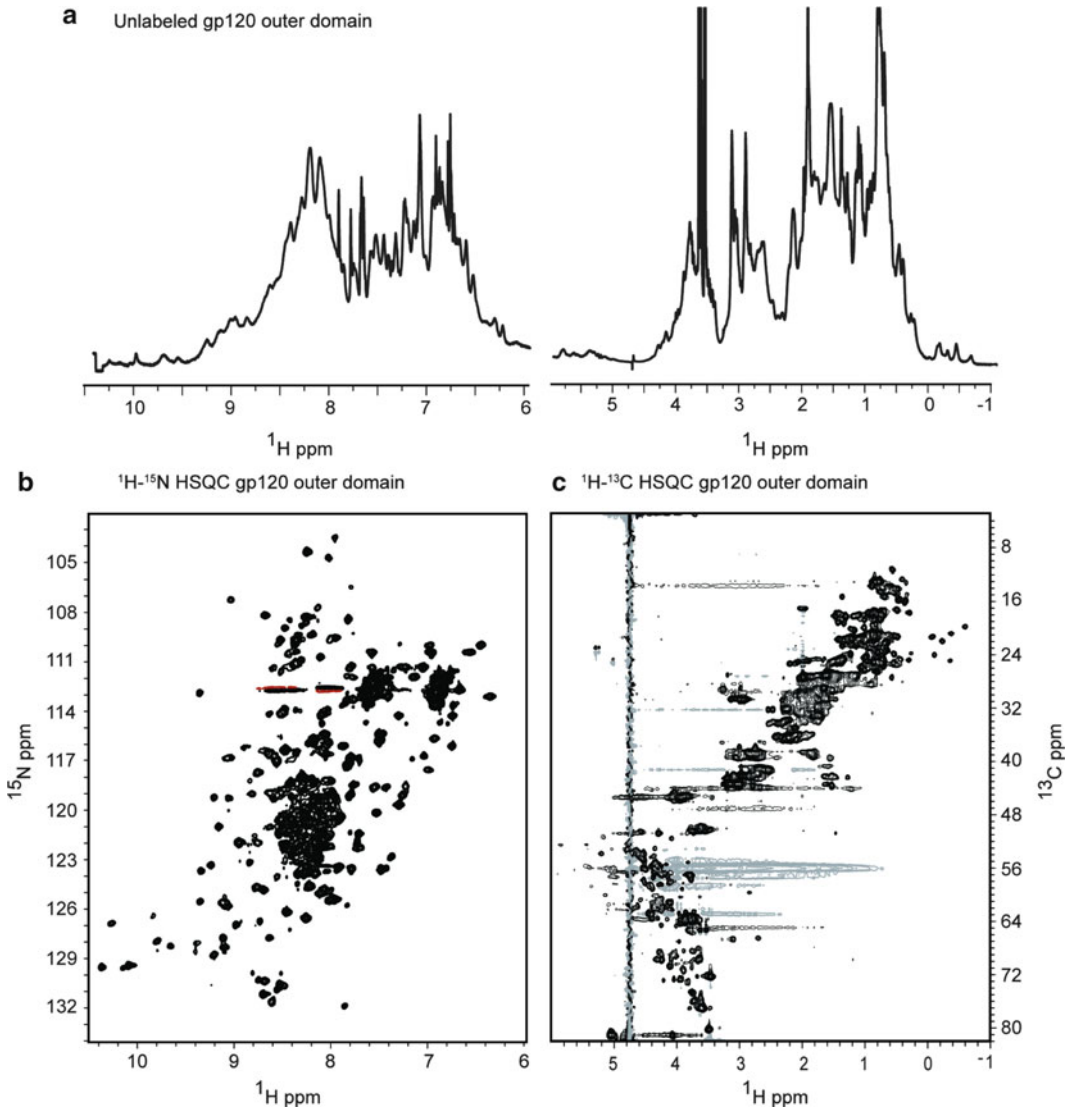


Fig. 11.3 HIV-1 gp120 outer domain expressed and purified from the adenoviral expression system is a functionally active and well folded protein. One dimensional proton spectra of unlabeled deglycosylated HIV-1 gp120 outer domain acquired at 600 MHz and 25°C is shown in panel A. ^1H - ^{15}N and ^1H - ^{13}C HSQC of $^{15}\text{N}/^{13}\text{C}$ labeled outer domain acquired at 900 MHz and 25°C are shown in (b) and (c), respectively (Reproduced from Sastry et al. [1]. With permission from Springer)

have revolutionized recombinant protein production from mammalian expression systems. In this chapter we have described an adenoviral based mammalian expression system that exploits the high level of protein expression obtained from an adenoviral vector when coupled to lung carcinoma cells. This system was developed for the expression of transgenes in the context of vaccines and gene therapy [62] The methodology described in this chapter to obtain uniform as well as selective amino-acid labeled proteins should reduce the barrier to structure determination of post-translationally modified proteins and their complexes by NMR spectroscopy.

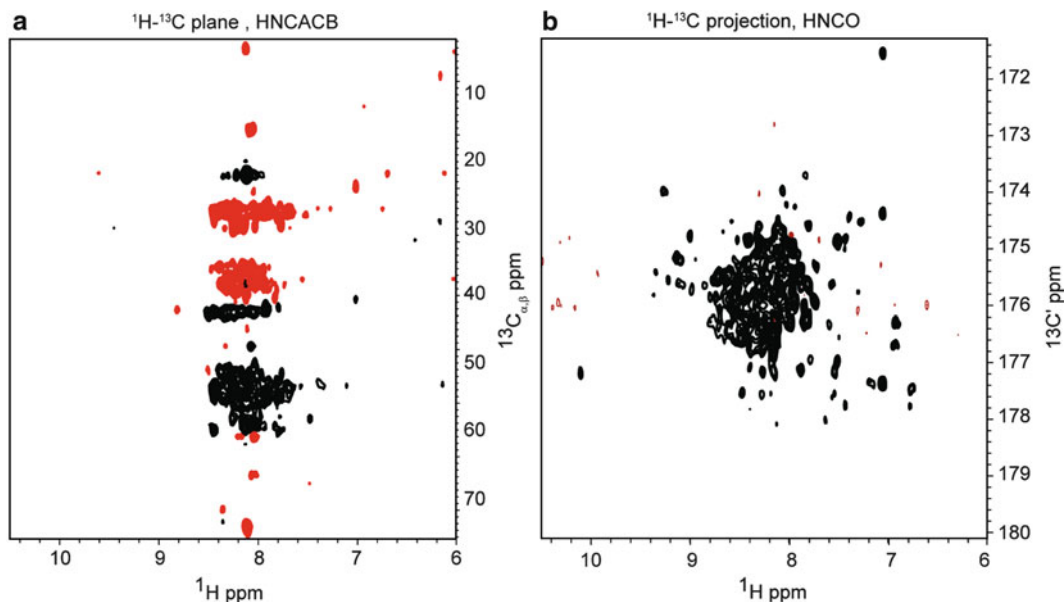


Fig. 11.4 HIV-1 gp120 outer domain expressed and purified from the adenoviral expression system is a functionally active and well folded protein. (a) A ^1H - ^{13}C $C_{\alpha}C_{\beta}$ plane of a 3D HNCACB experiment acquired on a $\sim 400 \mu\text{M}$ $^{15}\text{N}/^{13}\text{C}$ gp120 outer domain sample at 900 MHz and 25°C showing correlations from the amide NH to $C_{\alpha}C_{\beta}$ atoms demonstrates the uniform enrichment of C_{α} and C_{β} . The downfield shifted amides are relatively weak and are not observed in the 2D plane of the HNCACB experiment. (b) A ^1H - ^{13}C projection of a 3D HNCO experiment further demonstrates that the adenovirus-vectored mammalian expression system can provide the necessary isotopically enriched samples for heteronuclear NMR spectroscopy (Adapted from Sastry et al. [1]. With permission from Springer)

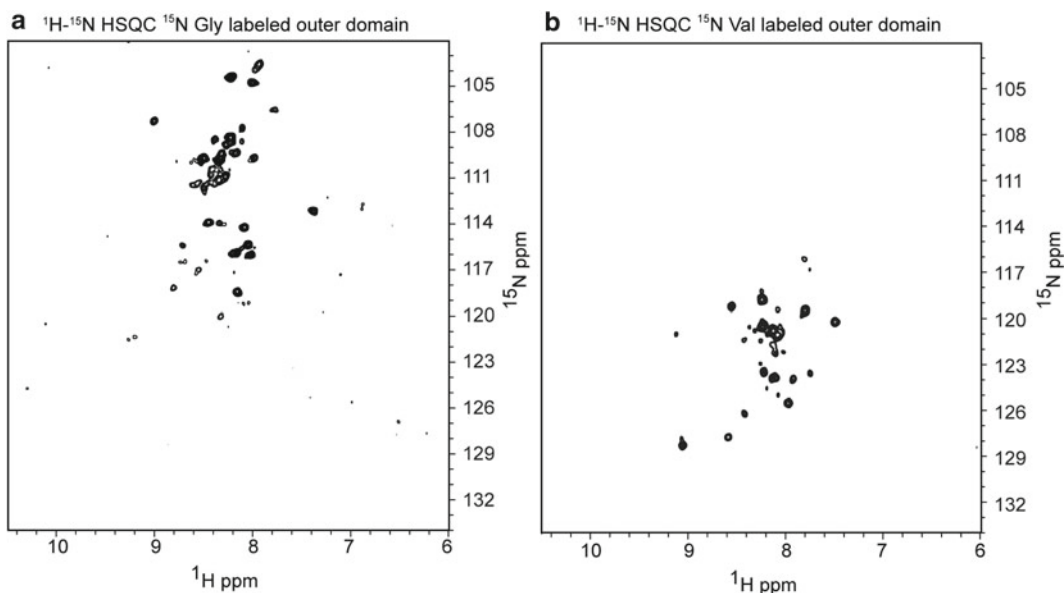


Fig. 11.5 Single amino acid labeling of HIV-1 gp120 outer domain from the adenoviral expression system is feasible. (a) ^1H - ^{15}N HSQC spectra of HIV-1 gp120 outer domain selectively labeled with ^{15}N glycine acquired at 900 MHz and 25°C and (b) ^1H - ^{15}N HSQC spectra of HIV-1 gp120 outer domain selectively labeled with ^{15}N valine acquired at 700 MHz and 25°C . The ^1H - ^{15}N HSQC of selectively labeled outer domain spectra exhibit excellent signal to noise and thus demonstrate that the adenoviral system can be used to obtain selectively labeled proteins

11.6 Materials

We report materials used for selective amino-acid labeling. Materials for summarized experiments can be found in the respective primary publications [1, 40].

Specifically labeled ^{15}N Valine-CGM6750 and ^{15}N Glycine-CGM6750 with all other amino acids and components unlabeled were obtained from Cambridge Isotope Laboratories, Inc (Andover, MA). High glucose containing DMEM with Hepes and NaHCO_3 was obtained from Life Technologies. Kifunensine and swainsonine were obtained from Enzo Life Sciences. Nickel-NTA was obtained from Qiagen Inc.

Acknowledgments We thank the NMR staff at the New York Structural Biology Consortium for assistance with instrumentation and data acquisition. We also thank the members of the Structural Biology Section and Structural Bioinformatics Section at the Vaccine Research Center for insightful comments and discussions. Support for this work was provided by the Intramural Program of the NIH (NIAID and NIDDK). 900 MHz spectrometers were purchased with funds from NIH, USA, the Keck Foundation, New York State, and the NYC Economic Development Corporation.

11.7 Appendix

A list of reagents used in the adenoviral based mammalian expression system

Reagents	Source
DNA vectors	
pVRC1194	Available upon request from Dr. G. Nabel
pVRC1290	Available upon request from Dr. G. Nabel
Restriction enzymes	
BamHI	New England Biolabs Inc.
XbaI	New England Biolabs Inc.
SacI	New England Biolabs Inc.
ClaI	New England Biolabs Inc.
Recombinases and proteases	
Cre recombinase	Novagen-EMD Biosciences Inc.
HRV3C protease	Novagen-EMD Biosciences Inc.
Transfection kits and reagents	
Mammalian transfection system	Promega
Lipofectamine	Life Technologies
Cell lines	
HEK293	ATCC
A549	ATCC
Cell culture reagents	
DMEM with glucose, NaHCO_3 , Hepes	Life Technologies
Fetal bovine serum	Life Technologies
Dialyzed fetal bovine serum	Life Technologies
Penicillin/Streptomycin	Life Technologies
Trypsin/EDTA	Life Technologies
Kifunensine	Enzo Life Sciences
Swainsonine	Enzo Life Sciences
Sodium pyruvate	Sigma-Aldrich
BE6000	Cambridge Isotope Laboratories, Inc.
^{15}N CGM6000	Cambridge Isotope Laboratories, Inc.

(continued)

(continued)

Reagents	Source
¹⁵ N/ ¹³ C CGM6000	Cambridge Isotope Laboratories, Inc.
¹⁵ N Glycine-CGM6750	Cambridge Isotope Laboratories, Inc.
¹⁵ N Valine-CGM6750	Cambridge Isotope Laboratories, Inc.
EDTA-d ₁₂ (98%)	Cambridge Isotope Laboratories, Inc.
Tris-D11 (98%)	Cambridge Isotope Laboratories, Inc.
D ₂ O (99.8%)	Cambridge Isotope Laboratories, Inc.
Protein purification	
Nickel-NTA	Qiagen Inc
Protease inhibitor cocktail-EDTA free	Roche Diagnostics
Cell lysis buffer	Cell Signaling

References

- Sastry M, Xu L, Georgiev IS, Bewley CA, Nabel GJ, Kwong PD (2011) Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy. *J Biomol NMR* 50:197–207
- Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J Struct Biol* 172:55–65
- Kriz A, Schmid K, Baumgartner N, Ziegler U, Berger I, Ballmer-Hofer K, Berger P (2010) A plasmid-based multigene expression system for mammalian cells. *Nat Commun* 1:120
- Trowitzsch S, Klumpp M, Thoma R, Carralot J-P, Berger I (2011) Light it up: highly efficient multigene delivery in mammalian cells. *Bioessays* 33:946–955
- Kingston RE, Chen CA, Okayama H (2001) Calcium phosphate transfection. In: *Current protocols in immunology*, vol 31. Wiley, New York, pp 10.13.1–10.13.9
- Potter H, Heller R (2010) Transfection by electroporation. In: *Current protocols in molecular biology*, vol 92. Wiley, New York, pp 9.3.1–9.3.10
- Xia W, Bringmann P, McClary J, Jones PP, Manzana W, Zhu Y, Wang S, Liu Y, Harvey S, Madlansacay MR, McLean K, Rosser MP, MacRobbie J, Olsen CL, Cobb RR (2006) High levels of protein expression using different mammalian CMV promoters in several cell lines. *Protein Expr Purif* 45:115–124
- Coleman TA, Parmelee D, Thotakura NR, Nguyen N, Bürgin M, Gentz S, Gentz R (1997) Production and purification of novel secreted human proteins. *Gene* 190:163–171
- Backliwal G, Hildinger M, Chenuet S, Wulhfard S, De Jesus M, Wurm FM (2008) Rational vector design and multipathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. *Nucleic Acids Res* 36:e96
- Hopkins RF, Wall VE, Esposito D (2012) Optimizing transient recombinant protein expression in mammalian cells. In: Hartley JL (ed) *Protein expression in mammalian cells: methods and protocols*, vol 801, *Methods in molecular biology*. Humana Press, New York, pp 251–268
- Geisse S, Fux C (2009) Recombinant protein production by transient gene transfer into mammalian cells. In: Burgess RR, Deutscher MP (eds) *Methods in Enzymology*, vol 463. Academic Press, San Diego, pp 223–238
- Zhou T, Xu L, Dey B, Hessell AJ, Van Ryk D, Xiang S-H, Yang X, Zhang M-Y, Zwick MB, Arthos J, Burton DR, Dimitrov DS, Sodroski J, Wyatt R, Nabel GJ, Kwong PD (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* 445:732–737
- Chen L, Kwon YD, Zhou T, Wu X, O'Dell S, Cavacini L, Hessell AJ, Pancera M, Tang M, Xu L, Yang Z-Y, Zhang M-Y, Arthos J, Burton DR, Dimitrov DS, Nabel GJ, Posner MR, Sodroski J, Wyatt R, Mascola JR, Kwong PD (2009) Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science* 326:1123–1127
- Kwong PD, Nabel GJ, Acharya P, Boyington JC, Chen L, Hood C, Kim A, Kong L, Kwon YD, Majeed S, McLellan J, Ofek G, Pancera M, Sastry M, Shah AC, Stuckey J, Zhou T (2011) Structural biology and the design of effective vaccines for HIV-1 and other viruses. In: Georgiev S (ed) *National Institute of Allergy and Infectious Disease, NIH*. Humana, New York, pp 387–402
- Southern P, Berg P (1982) Transformation of mammalian cell to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J Mol Appl Genet* 1:327–341
- Chen C, Okayama H (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* 7:2745–2752

17. Hansen AP, Petros AM, Mazar AP, Pederson TM, Rueter A, Fesik SW (1992) A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. *Biochemistry* 31:12713–12718
18. Wyss DF, Dayie KT, Wagner G (1997) The counterreceptor binding site of human CD2 exhibits an extended surface patch with multiple conformations fluctuating with millisecond to microsecond motions. *Protein Sci* 6:534–542
19. Wyss DF, Withka JM, Knoppers MH, Sterne KA, Recny MA, Wagner G (1993) Proton resonance assignments and secondary structure of the 13.6 kDa glycosylated adhesion domain of human CD2. *Biochemistry* 32:10995–11006
20. Lustbader JW, Birken S, Pollak S, Pound A, Chait BT, Mirza UA, Ramnarain S, Canfield RE, Brown JM (1996) Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures. *J Biomol NMR* 7:295–304
21. Shindo K, Masuda K, Takahashi H, Arata Y, Shimada I (2000) Backbone ¹H, ¹³C, and ¹⁵N resonance assignments of the anti-dansyl antibody Fv fragment. *J Biomol NMR* 17:357–358
22. Hansen AP, Petros AM, Meadows RP, Nettesheim DG, Mazar AP, Olejniczak ET, Xu RX, Pederson TM, Henkin J, Fesik SW (1994) Solution structure of the amino-terminal fragment of urokinase-type plasminogen activator. *Biochemistry* 33:4847–4864
23. Hinck AP, Archer SJ, Qian SW, Roberts AB, Sporn MB, Weatherbee JA, Tsang MLS, Lucas R, Zhang B-L, Wenker J, Torchia DA (1996) Transforming growth factor β 1: three-dimensional structure in solution and comparison with the X-ray structure of transforming growth factor β 2. *Biochemistry* 35:8517–8534
24. Wyss D, Choi J, Li J, Knoppers M, Willis K, Arulanandam A, Smolyar A, Reinherz E, Wagner G (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* 269:1273–1278
25. Metzler WJ, Bajorath J, Fenderson W, Shaw SY, Peach R, Constantine KL, Naemura J, Leyte G, Lavoie TB, Mueller L, Linsley PS (1997) Solution structure of human CTLA 4 and delineation of a CD80/CD86 binding site conserved in CD28. *Nat Struct Biol* 4:527–531
26. Werner K, Richter C, Klein-Seetharaman J, Schwalbe H (2008) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* 40:49–53
27. Arata Y, Kato K, Takahashi H, Shimada I (1994) Nuclear magnetic resonance study of antibodies: a multinuclear approach. *Method Enzymol* 239:440–464
28. Klein-Seetharaman J, Yanamala NVK, Javeed F, Reeves PJ, Getmanova EV, Loewen MC, Schwalbe H, Khorana HG (2004) Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad Sci USA* 101:3409–3413
29. Klein-Seetharaman J, Reeves PJ, Loewen MC, Getmanova EV, Chung J, Schwalbe H, Wright PE, Khorana HG (2002) Solution NMR spectroscopy of [α -¹⁵N]lysine-labeled rhodopsin: the single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci USA* 99:3452–3457
30. Mulligan RC, Howard BH, Berg P (1979) Synthesis of rabbit [β]-globin in cultured monkey kidney cells following infection with a SV40 [β]-globin recombinant genome. *Nature* 277:108–114
31. Warnock JN, Daigre C, Al-Rubeai M (2011) Introduction to viral vectors. In: Merten O, Al-Rubeai M (eds) *Viral vectors for gene therapy: methods and protocols*, vol 737, *Methods in molecular biology*. Humana Press, New York, pp 1–25
32. Gluzman Y, Reichl H, Scolnick D (1982) Helper-free adenovirus type-5 vectors. In: Gluzman Y (ed) *Eukaryotic viral vectors*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 187–192
33. Howley PM, Sarver N, Law M-F (1983) Eukaryotic cloning vectors derived from bovine papillomavirus DNA. In: Wu R, Grossman L, Moldave K (eds) *Methods in Enzymology*, vol 101. Academic Press, pp 387–402
34. Waehler R, Russell SJ, Curiel DT (2007) Engineering targeted viral vectors for gene therapy. *Nat Rev Genet* 8:573–587
35. Berk AJ (1986) Adenovirus promoters and E1A transactivation. *Annu Rev Genet* 20:45–79
36. Shenk T (1996) Adenoviridae: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM (eds) *Fields virology*. Lippincott-Raven, Philadelphia, pp 2111–2148
37. Berkner K, Sharp P (1982) Preparation of adenovirus recombinant using plasmids of viral DNA. In: Gluzman Y (ed) *Eukaryotic viral vectors*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 193–198
38. Babich A, Feldman LT, Nevins JR, Darnell JE Jr, Weinberger C (1983) Effect of adenovirus on metabolism of specific host mRNAs: transport control and specific translational discrimination. *Mol Cell Biol* 3:1212–1221
39. Huang JT, Schneider RJ (1990) Adenovirus inhibition of cellular protein synthesis is prevented by the drug 2-aminopurine. *Proc Natl Acad Sci USA* 87:7115–7119
40. Aoki K, Barker C, Danthinne X, Imperiale MJ, Nabel GJ (1999) Efficient generation of recombinant adenoviral vectors by Cre-lox recombination in vitro. *Mol Med* 5:224–231
41. Moore D, Dowhan D (2002) Purification and concentration of DNA from aqueous solutions. In: *Current protocol in molecular biology*, vol 59. Wiley, New York, pp 2.1.1–2.1.10
42. Voytas D (2001) Agarose gel electrophoresis. In: *Current protocol in molecular biology*, vol 51. Wiley, New York, pp 2.5A.1–2.5A.9

43. Hall K, Blair Zajdel ME, Blair GE (2009) Defining the role of CD46, CD80 and CD86 in mediating adenovirus type 3 fiber interactions with host cells. *Virology* 392:222–229
44. Graham FL, Smiley J, Russell WC, Nairn R (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* 36:59–72
45. Rich RL, Myszkka DG (2010) Grading the commercial optical biosensor literature—Class of 2008: ‘The Mighty Binders’. *J Mol Recognit* 23:1–64
46. Raghavan M, Bjorkman PJ (1995) BIAcore: a microchip-based system for analyzing the formation of macromolecular complexes. *Structure* 3:331–333
47. Duffy AM, O’Doherty AM, O’Brien T, Strappe PM (2005) Purification of adenovirus and adeno-associated virus: comparison of novel membrane-based technology to conventional techniques. *Gene Ther* 12:S62–S72
48. Tan R, Li C, Jiang S, Ma L (2006) A novel and simple method for construction of recombinant adenoviruses. *Nucleic Acids Res* 34:e89
49. Chillon M, Alemany R (2011) Methods to construct recombinant adenovirus vectors. In: Merten O, Al-Rubeai M (eds) *Viral vectors for gene therapy: methods and protocols*, vol 737, *Methods in molecular biology*. Humana Press, New York, pp 117–138
50. Horwitz MS (1996) Adenoviruses. In: Fields BN, Knipe DM, Howley PM (eds) *Fields virology*. Lippincott-Raven, Philadelphia, pp 2149–2171
51. Fisher TN, Ginsberg HS (1957) Accumulation of organic acids by HeLa cells infected with type 4 adenovirus. *Proc Soc Exp Biol Med* 95:47–51
52. LeMaster DM (1990) Deuterium labelling in NMR structural analysis of larger proteins. *Q Rev Biophys* 23:133–174
53. Murphy J, Desai C, Giaretti W, Kendall F, Nicolini C (1977) Experimental results on mammalian cells growing in vitro in deuterated medium for neutron-scattering studies. *J Cell Sci* 25:87–94
54. Elbein AD, Tropea JE, Mitchell M, Kaushal GP (1990) Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J Biol Chem* 265:15599–15605
55. Elbein AD (1991) Glycosidase inhibitors: inhibitors of N-linked oligosaccharide processing. *FASEB J* 5:3055–3063
56. Kong L, Sheppard NC, Stewart-Jones GBE, Robson CL, Chen H, Xu X, Krashias G, Bonomelli C, Scanlan CN, Kwong PD, Jeffs SA, Jones IM, Sattentau QJ (2010) Expression-system-dependent modulation of HIV-1 envelope glycoprotein antigenicity and immunogenicity. *J Mol Biol* 403:131–147
57. Magnelli P, Bielak A, Guthrie E (2012) Identification and characterization of protein glycosylation using specific endo- and exoglycosidases. In: Hartley JL (ed) *Protein expression in mammalian cells: methods and protocols*, vol 801, *Methods in molecular biology*. Humana Press, New York, pp 189–211
58. van Reis R, Brake JM, Charkoudian J, Burns DB, Zydney AL (1999) High-performance tangential flow filtration using charged membranes. *J Membr Sci* 159:133–142
59. Anglister J, Frey T, McConnell HM (1984) Magnetic resonance of a monoclonal anti-spin-label antibody. *Biochemistry* 23:1138–1142
60. Gossert A, Hinniger A, Gutmann S, Jahnke W, Strauss A, Fernández C (2011) A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *J Biomol NMR* 51:449–456
61. Truhlar SME, Cervantes CF, Torpey JW, Kjaergaard M, Komives EA (2008) Rapid mass spectrometric analysis of ¹⁵N-Leu incorporation fidelity during preparation of specifically labeled NMR samples. *Protein Sci* 17:1636–1639
62. Nabel GJ (1999) Development of optimized vectors for gene therapy. *Proc Natl Acad Sci USA* 96:324–326