

Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy

Mallika Sastry · Ling Xu · Ivelin S. Georgiev ·
Carole A. Bewley · Gary J. Nabel · Peter D. Kwong

Received: 25 February 2011 / Accepted: 25 April 2011 / Published online: 12 June 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract NMR spectroscopic characterization of the structure or the dynamics of proteins generally requires the production of samples isotopically enriched in ^{15}N , ^{13}C , or ^2H . The bacterial expression systems currently in use to obtain isotopic enrichment, however, cannot produce a number of eukaryotic proteins, especially those that require post-translational modifications such as N-linked glycosylation for proper folding or activity. Here, we report the use of an adenovirus vector-based mammalian expression system to produce isotopically enriched ^{15}N or $^{15}\text{N}/^{13}\text{C}$ samples of an outer domain variant of the HIV-1 gp120 envelope glycoprotein with 15 sites of N-linked glycosylation. Yields for the ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled gp120s after affinity chromatography were 45 and 44 mg/l, respectively, with an average of over 80% isotope incorporation. Recognition of the labeled gp120 by cognate antibodies that recognize complex epitopes showed affinities comparable to the unlabeled protein. NMR spectra, including ^1H - ^{15}N and ^1H - ^{13}C HSQCs, ^{15}N -edited NOESY-HSQC, and 3D HNCO, were of high quality, with signal-to-noise

consistent with an efficient level of isotope incorporation, and with chemical shift dispersion indicative of a well-folded protein. The exceptional protein yields, good isotope incorporation, and ability to obtain well-folded post-translationally modified proteins make this mammalian system attractive for the production of isotopically enriched eukaryotic proteins for NMR spectroscopy.

Keywords Adenovirus vector · Eukaryotic expression · HIV-1 envelope · Isotope enrichment · N-linked glycosylation

Introduction

Structural studies of proteins and their complexes by NMR spectroscopy often require incorporation of ^{15}N , ^{13}C , or ^2H stable isotopes. In the vast majority of examples, *E. coli* expression systems have been used for isotope labeling, because of their ability to provide efficient incorporation of stable isotopes and high levels of protein production. Bacterial protein production systems, however, fail to produce well-folded proteins for numerous protein classes. These include many eukaryotic proteins, particularly those that are secreted, have multiple disulfide bonds, require specific chaperones, require specific prosthetic groups, or are post-translationally modified. Despite extensive efforts with solubility enhancing fusion tags, lower growth temperature, co-expression with eukaryotic chaperones, periplasmic expression, and other methods such as creation of an oxidizing environment in the bacterial cytoplasm, only moderate improvements in success rate have been obtained (Schein and Noteborn 1988; Bessette et al. 1999; Cornelis 2000; Kadokura and Beckwith 2001; Esposito and Chatterjee 2006).

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9506-4) contains supplementary material, which is available to authorized users.

M. Sastry · L. Xu · I. S. Georgiev · G. J. Nabel ·
P. D. Kwong (✉)
Vaccine Research Center, National Institute of Allergy
and Infectious Diseases, National Institutes of Health,
40 Convent Drive, Bethesda, MD 20892-3027, USA
e-mail: pdkwong@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

By contrast, eukaryotic proteins can often be efficiently produced in their native form by mammalian cell lines such as human embryonic kidney (HEK) 293 cells and Chinese hamster ovary (CHO) cells. A number of investigators have attempted to produce isotopically labeled proteins using such mammalian expression systems. In an early study, Hansen and colleagues used a mixture of isotopically labeled amino acids derived from bacterial or algal hydrolysates to obtain 0.5 mM uniformly ^{15}N -labeled urokinase from 0.75 to 1.0 l culture medium from the mouse myeloma Sp2/0 cell line (Hansen et al. 1992). Wyss and colleagues subsequently reported expression of human CD2 with ^{15}N -labeled lysine residues using stably transfected CHO cells and media containing isotopically labeled lysine (Wyss et al. 1993, 1997), and a strategy incorporating a mixture of labeled amino acids has been used to obtain 10 mg/l of uniformly labeled ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled human chorionic gonadotropin from a stably transfected CHO cell line (Lustbader et al. 1996). A similar strategy was used to produce ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled IgG2a antibody from a mouse hybridoma cell line (Shindo et al. 2000), and Werner and colleagues used a subset of ^{15}N and $^{15}\text{N}/^{13}\text{C}$ amino acids (G, K, L, Q, S, T, V, and W) to label rhodopsin expressed from HEK293 cells with an average yield of 2 mg/l purified protein (Werner et al. 2008). More recently, Skelton and colleagues reported the development of a reduced nutrient media formulation for Lec1 cells, a glycosylation deficient CHO cell line, to obtain partially labeled proteins (Skelton et al. 2010). Unfortunately, a majority of the published methods for obtaining isotopically labeled proteins from eukaryotic expression systems require labor intensive optimization of synthetic media, a problem that is often compounded by low yields. As a result, the use of mammalian expression systems has often been limited to amino-acid type-specific labeling (Arata et al. 1994; Klein-Seetharaman et al. 2002, 2004).

A eukaryotic expression system capable of expressing isotopically labeled, well-folded, post-translationally modified proteins at high yield from commercially available media has thus been widely sought. One such system, which uses an adenovirus vector coupled to mammalian expression, was developed for the expression of transgenes in the context of vaccines and gene therapy (Nabel 1999; Barouch and Nabel 2005). Recombinant adenovirus, wherein the E1 region has been deleted, is replication incompetent, and a further deletion of the E3 region allows for insertion of up to 8-kilobases of recombinant transgenes. 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 (Babich et al. 1983; Huang and Schneider 1990). We found

this system capable of expressing the HIV-1 gp120 envelope glycoprotein, which has 9 disulfides and ~ 20 sites of N-linked glycosylation, at a level of ~ 50 mg/l for wild-type and various truncated variants (Zhou et al. 2007; Wang et al. 2009), a level of expression substantially better than achieved by transient transfection or from transformed drosophila cells (Kwong et al. 2010). As precise measurement of the flexibility of HIV-1 gp120 may provide insight into its immunogenicity, we sought to produce an isotopically labeled fragment of HIV-1 gp120 that contained the epitope for broadly neutralizing antibodies and that was sufficiently small and well behaved to be analyzed by NMR spectroscopy. Herein, we report an efficient method to isotopically enrich proteins from a mammalian system that exploits the high level of protein expression obtained from an adenoviral vector. We demonstrate the method for an outer domain variant of HIV-1 gp120, comprising 229 amino acids, 15 potential sites of N-linked glycosylation, and four disulfide bonds. This outer domain variant also contains the initial attachment site of HIV-1 to the CD4 receptor, a site of HIV-1 vulnerability to neutralizing antibodies and therefore of vaccine importance (Zhou et al. 2007, 2010; Wu et al. 2009).

Materials and methods

Restriction enzymes and transfection system

BamH1, Xba1, Sac1, Cla1 were obtained from New England Biolabs. Cre recombinase was obtained from Novagen (Madison, WI). ProFection[®] Mammalian Transfection System was obtained from Promega (Madison, WI).

Media and isotopes

High glucose containing, pyruvate-free Dulbecco's modified Eagle medium (DMEM) with HEPES, NaHCO_3 , dialyzed fetal bovine serum (FBS), trypsin-EDTA, and penicillin/streptomycin were obtained from Invitrogen. Inc (Carlsbad, CA). ^{15}N -CGM6000 (U-98%), $^{15}\text{N}/^{13}\text{C}$ -CGM6000 (U-98%) containing HEPES and NaHCO_3 , EDTA- d_{12} (98%), Tris-D11 (98%), and D_2O (D, 99.8%) were purchased from Cambridge Isotopes Limited (Andover, MA). Sodium pyruvate was purchased from Sigma-Aldrich (St. Louis, MO).

Cell lines and growth in protonated and deuterated media

HEK 293 and A549 cells used in this work were obtained from American Type Culture Collection (Manassas, VA). A549 cells were grown in DMEM with 25 mM HEPES,

10% heat inactivated dialyzed fetal bovine serum (FBS), and 1% penicillin/streptomycin. A549 cells were split to desired density, centrifuged at 360 g for 10 min and resuspended in appropriate media (DMEM, ^{15}N -CGM6000, or $^{15}\text{N}/^{13}\text{C}$ -CGM6000 with and without pyruvate). Cells were recounted prior to plating in 6 well plates at 0.21×10^6 cells/well and incubated at 37°C for 7 days. Aliquots were taken every 24 h and cells counted in triplicate for each media type. For deuterated media, A549 cells were initially adapted to 20% D_2O -containing DMEM, 10% heat inactivated FBS and 1% penicillin/streptomycin (Murphy et al. 1977) for a few days before seeding at 1×10^6 cells/plate (10 cm diameter) in DMEM containing 20, 45 or 70% D_2O . Viable cells were counted every 24 h for 10 days to obtain growth curves.

Recombinant adenovirus vectors

Recombinant adenovirus type 5 (Ad5)-green fluorescent protein (GFP) and Ad5-HIV-1 gp120 outer domain were constructed by homologous recombination as described previously (Aoki et al. 1999). In brief, the shuttle vectors containing GFP or outer domain DNA constructs were recombined with a cosmid which carries Ad5 genomic DNA. The Ad5 genomic DNA lacks the E1 region and has a non-functional E3 region. The recombinations were carried out using Cre recombinase (Novagen, Madison, Wis.). The Ad5 viruses were purified twice using CsCl gradient centrifugation, followed by a desalting column to remove the CsCl. The expression of GFP gene was under the control of an RSV promoter (Ohno et al. 1997). The expression of gp120 outer domain was under the control of a CMV promoter and the construct has a C-terminal HRV3C cleavage site followed by a His₆ purification tag. The aforementioned adenoviral vectors are available upon request from GJN; commercial adenovirus expression systems are also available (such as “ViraPower” from Invitrogen).

Green fluorescent protein expression and purification

The HEK 293 and A549 cells were seeded at 0.8×10^6 cells/well in a six well plate on day-1. The following day, media was replaced with fresh DMEM, ^{15}N CGM6000, $^{15}\text{N}/^{13}\text{C}$ CGM6000 or 70% D_2O containing unlabeled media. 293 cells were transfected with a plasmid containing the gene for GFP using lipofectamine 2000 (Invitrogen) as the transfection agent. A control vector pVRC8400 was also transfected as a negative control. A549 cells were infected with a recombinant adenovirus (rAdGFP) that contained a gene for GFP. The time course of protein expression was followed by harvesting both HEK 293 and

A549 cells at 24 h intervals over a period of 6 days post transfection with plasmid DNA or post infection by recombinant adenovirus. Cells were lysed, and the production of GFP quantified by measuring the fluorescence of crude cell lysates and comparing to a standard curve (Biovision).

A549 adherent cells were maintained in DMEM. Typically, cells were seeded in 15 cm plates at $12\text{--}15 \times 10^6$ cells/plate on day-1 in fresh DMEM containing 10% heat inactivated dialyzed FBS, 1% penicillin/streptomycin and allowed to grow overnight at 37°C. The following day, media was replaced with labeled ^{15}N , $^{15}\text{N}/^{13}\text{C}$ CGM6000 or fresh DMEM. A549 cells were infected with rAdGFP to a final concentration of 2,500 particles/cell. Protein expression was monitored by following the fluorescence signal of GFP; 72–96 h post infection cells were washed with PBS, lifted off with trypsin–EDTA, and pelleted at 360 g rpm for 15', and the cell pellet washed with PBS and stored at -80°C until further use. Cell pellet was resuspended in 10 ml cell lysis buffer (Cell Signaling), placed on ice for 10 min, nutated at 4°C for 1 h and the lysate spun down at 14,000 rpm at 4°C for 40 min. The cleared lysate was immediately loaded onto an anti-GFP affinity column, nutated at 4°C overnight and subsequently washed with PBS. GFP was eluted using 25 mM Tris and 3 M MgCl_2 , pH 7.4. Fractions containing GFP were pooled; concentrated and desalted using a PD10 column, and purity of GFP was analyzed by SDS–PAGE.

HIV-1 gp120 outer domain expression and purification

A549 cells were seeded at $12\text{--}15 \times 10^6$ cells/plate on day-1 in fresh DMEM containing 10% heat inactivated dialyzed FBS, 1% penicillin–streptomycin and allowed to grow overnight at 37°C. The following day, media was replaced with ^{15}N , $^{15}\text{N}/^{13}\text{C}$ CGM6000 or fresh DMEM containing the glycosidase inhibitors kifunensine (12.5 mg/l) and swainsonine (5 mg/l). A549 cells were 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. The culture supernatant was harvested 96–108 h post infection; cell debris was spun down at 365 g for 15 minutes. The culture supernatant was filtered and the outer domain was purified by immobilized nickel- and antibody b12-affinity chromatography. Fractions containing the outer domain were pooled, concentrated, dialyzed against PBS and deglycosylated using EndoH_f followed by size-exclusion chromatography.

Quantification of isotope incorporation

To quantify the extent of isotope incorporation, unlabeled, $^{15}\text{N}/^{13}\text{C}$ -labeled, ^{15}N -labeled and deglycosylated HIV-1

outer domain and GFP were subjected to tryptic digestion followed by MALDI TOF mass spectrometry (see Supplementary methods). A modified algorithm based on the approaches of Kubinyi (1991) was used to compute theoretical estimates of isotopic distributions using isotope masses and abundances described previously (Kubinyi 1991) (Table S3). In this modified algorithm, the ^{15}N abundance was calculated between 0 and 100% at increments of 1%, and for each percent- ^{15}N value, the isotope distribution pattern was computed as follows. The isotope distribution pattern for each element was computed separately and peaks with intensities less than a cutoff threshold were discarded (a cutoff of 10^{-7} was used). The isotope distribution patterns for the different elements were then combined to obtain the overall distribution pattern for the given ^{15}N abundance percentage. ‘Major’ peaks were defined to be peaks at integral distances from the mono-isotopic peak. The intensities for all peaks with masses within a cutoff threshold from a given major peak were added to the intensity of the major peak (a mass cutoff of 0.3 was used), thus forming ‘super-peaks.’ The highest super-peak intensity was set to 100 and all other super-peak intensities were normalized accordingly. Only significant super-peaks with intensities (after normalization) greater than a cutoff threshold were output (a cutoff of 10^{-4} was used); the intensities of all other super-peaks were set to zero. Once the isotope distribution patterns for all ^{15}N abundance percentages were computed, these patterns were compared against the observed mass spectrometry (M/S) spectra. Computed patterns for which the mass of the highest-intensity peak did not match the mass of the highest-intensity peak in the observed M/S spectra were discarded from further consideration. A range of possible incorporation percentages was obtained from the remaining computed patterns. For each of the patterns, the correlation between that pattern and the observed M/S pattern (using the heights of the M/S peaks) was computed using linear regression analysis. For the outer domain peptide, no adjustment for noise was performed on the M/S pattern. Correlation computation for outer domain included the peaks in the mass range 842.5–869.5, spanning the observed M/S peaks for that peptide. A similar grid search was performed for the double labeled outer domain to obtain an estimate of the percentage incorporation of ^{13}C while utilizing the estimate of ^{15}N incorporation from the analysis of the ^{15}N labeled outer domain (see supplemental material). For the GFP peptide, the M/S pattern was adjusted for noise, resulting in a total of seven M/S peaks in the mass range 1,355–1,362. Correlation computation for GFP included the peaks in the mass range 1,347.7–1,378.7, corresponding to the mass range in which all computed significant super-peaks were

observed; missing M/S peaks within this range were set to zero.

Surface plasmon resonance

To characterize the state of GFP, an anti-GFP antibody ab1218 (Abcam) was directly immobilized onto a Biacore CM5 sensor chip to a final surface density of ~ 500 RU. Unlabeled, ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled GFP were used as analytes at concentrations ranging from 3.9 to 62.5 nM, at two-fold serial dilutions. Similarly for the binding of HIV-1 gp120 outer domain antibodies b13 and b12 were directly immobilized onto Biacore CM5 sensor chips to a final surface density of ~ 500 RU. Unlabeled, ^{15}N -, and $^{15}\text{N}/^{13}\text{C}$ -labeled outer domain were used as analytes at concentrations ranging from 3.9 to 250 nM, at two-fold serial dilutions.

NMR spectroscopy

NMR experiments were acquired on a Bruker Avance 900 MHz spectrometer equipped with a 5 mm TCI cryoprobe. Spectra were acquired on the deglycosylated ^{15}N and $^{15}\text{N}/^{13}\text{C}$ labeled HIV-1 gp120 outer domain a ~ 28 kDa protein with a concentration of 0.4 mM in 90% $\text{H}_2\text{O}/10\%\text{D}_2\text{O}$ at a pH of 7.0 and a temperature of 25°C. Standard Bruker pulse sequences were used for ^1H - ^{15}N HSQC, ^1H - ^{13}C HSQC, ^{15}N -edited NOESY-HSQC and the HNCOC experiment (Yamazaki et al. 1994), with States-TPPI quadrature detection. Water suppression was achieved by WATERGATE (Sklenar et al. 1993) and water flip back pulses. The HSQC spectra were acquired with 1,024 complex points in the direct dimension and with spectral widths of 13,550 and 2,919 Hz in the proton and nitrogen dimensions, respectively and 256, points in the nitrogen dimension. The ^1H - ^{13}C HSQC spectra were acquired with 1,024 points in the direct dimension and with spectral widths of 10,776 and 18,108 Hz in the proton and carbon dimension, respectively with 300 points in the carbon dimension. The 3D HNCOC (Kay et al. 1990; Grzesiek and Bax 1992) spectrum was recorded with 512, 32 and 25 complex points in proton, nitrogen and carbon dimensions, with spectral widths of 10,776, 2,919 and 4,528 Hz in the proton nitrogen and carbon dimensions, respectively. Recycle delays of 1 s were used for all experiments. Data sets were processed using NMRPipe (Delaglio et al. 1995). ^{15}N dimensions were extended to 64 complex points with linear prediction. Both ^{15}N and ^{13}C dimensions were apodized with a shifted squared sine-bell apodization function and zero filled to 256 complex points prior to Fourier transformation. All data were analyzed with the program CARI (Keller 2004).

Results

Mammalian expression system using adenoviral and transient transfection vectors

To provide an initial assessment of the ability of adenovirus vectors to produce a folded eukaryotic protein in isotopically labeled media, we tested the expression of GFP, which permitted quantification directly from crude cell lysates by fluorescence. We tested two expression systems, an adenovirus-vectored approach with A549 cells and a transient transfection approach with HEK293 cells, on four different media. These included unlabeled DMEM, two commercially-available CGM6000 media isotopically labeled with ^{15}N or $^{15}\text{N}/^{13}\text{C}$, and unlabeled DMEM made from 70% D_2O (Fig. 1). With all four media, the expression from the adenovirus vector was substantially higher than from transient expression. More notable was the consistent level of expression in all four media, suggesting that expression in labeled media did not substantially alter adenovirus-vectored protein expression. This level of expression was especially surprising for 70% D_2O media as growth of the A549 cells in this media was substantially slower (Fig. S1), and, in the case of 293 cells, resulted in extensive cell death by day four following transfection.

Characterization and quantification of isotopic incorporation in GFP

To quantify the degree of isotope enrichment in the adenovirus vector-expressed GFP, we expressed, purified and characterized GFP from A549 cells infected with recombinant adenovirus containing the GFP gene. Purification by

immune-affinity chromatography gave GFP with purity greater than 95% by SDS-PAGE (Fig. S2). We obtained milligram quantities of GFP from unlabeled, ^{15}N -, and $^{15}\text{N}/^{13}\text{C}$ -labeled media. Surface plasmon resonance analysis with the anti-GFP antibody ab1218 showed similar K_{DS} (0.85–3.9 nM) for labeled and unlabeled samples (Fig. S2). We encountered difficulties obtaining mass spectral data on intact GFP, and since we knew that the assessment of the glycosylated outer domain would require fragmentation, we assessed isotope incorporation using a combination of tryptic digestion and MALDI TOF mass spectrometry. In mass spectra of unlabeled GFP, three peptides with the sequences FSVSGEGEGDATYGGK (1,503.5 amu), TIFFKDDGNYK (1,347.4 amu), SAMPEGYVQER (1,266.3 amu) gave high quality peaks; in mass spectra of the labeled samples, however, overlapping peaks and poor quality data for the 1,503.5 and 1,266.3 peptides limited our analysis to the TIFFKDDGNYK peptide, with the percent incorporation of ^{15}N estimated at $72 \pm 3\%$, and the best fitting computed spectrum showing 74% (Table S1, Fig. S2).

HIV-1 gp120 outer domain: expression and antigenic characterization

Because the full-length HIV-1 gp120 glycoprotein is over 500 amino acids in size, we sought to produce a biologically relevant fragment more conducive to NMR analysis. One potential fragment, an outer domain variant, was initially described by Sodroski and colleagues (Yang et al. 2004) and spanned gp120 residues 252–482 (numbering for gp120 residues follows the standard HXBc2 numbering scheme (Korber et al. 1998)). This fragment contained the initial site of HIV-1-engagement of the CD4 receptor and

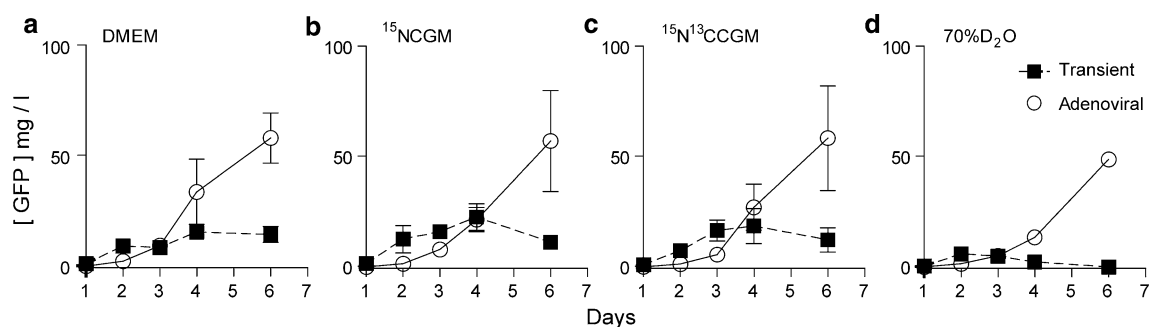


Fig. 1 Time course of GFP expression in transient versus adenoviral expression system. Protein production in an adenoviral expression system with A549 cells is superior to transient transfection with HEK 293 cells. **a** Transient expression of GFP in 293 cells (filled square) versus A549 cells/adenoviral expression (circle) in DMEM. **b** Transient expression of GFP in 293 cells (filled square) versus A549 cells/adenoviral expression (circle) in ^{15}N labeled CGM6000 media. **c** Transient expression of GFP in 293 cells (filled square) versus A549 cells/adenoviral expression (circle) in $^{15}\text{N}/^{13}\text{C}$ labeled CGM6000. **d** Transient expression of GFP in 293 cells (filled square) versus

A549 cells/adenoviral expression (circle) in 70% D_2O containing unlabeled media. In the 70% D_2O containing media, a large percentage of HEK 293 cells are dead five days post transfection. A549 cells, however, do not appear to be as affected as HEK 293 cells by the 70% D_2O containing media: in particular, expression from adenovirus-vectored genes does not appear to be strongly affected by the presence of 70% D_2O containing media, although the cells do not appear to grow (Fig. S1). All measurements were made in duplicate, SEM error bars are displayed for the time course of protein expression

was therefore of substantial vaccine interest. The initial outer domain version, however, showed reduced binding to broadly neutralizing anti-HIV-1 antibodies such as b12 and also contained the flexible third variable loop (V3), which is immunodominant. Various modifications restored b12 affinity (Wu et al. 2009; Kanekiyo et al. 2010; Xu et al. 2010), and we settled on a construct composed of gp120 residues 252–482, with V3 truncation and other modifications to alter relative CD4-binding site immunogenicity or to increase antigenic fidelity to the full length gp120 (Figs. S3–S5).

Adenovirus-vectored expression of this gp120 outer domain in A549 mammalian cells showed yields approaching 50 mg/l after Ni-affinity chromatography and b12-antibody affinity chromatography from unlabeled, ^{15}N -, and $^{15}\text{N}/^{13}\text{C}$ -labeled media (Table 1). N terminal analysis showed the secreted outer domain to start with A-P-R-R-P-V-V, where the first three amino acids were cloning artifacts and R-P-V-V corresponds to residues 252–255 of R2 gp120. SDS-PAGE analysis showed a broad smear of proteins of varying molecular weights, suggestive of heterogeneous N-linked glycosylation, and Endo H cleavage, which truncates the high mannose sugars N-linked glycan to the protein-proximal N-acetyl glucosamine, reduced the smear to a single tight band (Fig. 2a).

Antigenic analysis with three CD4-binding site reactive antibodies, b12, b13 and VRC01, showed similar affinities for glycosylated outer domain expressed in unlabeled, ^{15}N -, and $^{15}\text{N}/^{13}\text{C}$ -labeled media (Table S4). Although the affinity to broadly neutralizing antibody b12 was quite tight (~ 30 nM), the Endo H-deglycosylated protein demonstrated faster off-rates than observed for the full-length gp120 (Fig. 2b). For VRC01, the Endo H-deglycosylated protein failed to bind, perhaps relating to glycan interactions made by VRC01 or indicating that this outer domain variant did not perfectly replicate the antigenic behavior of full-length gp120. Additional modifications to optimize OD are currently being assessed; indeed, a driving rationale behind our NMR analysis of OD is to provide information on structure and flexibility as a means to facilitate its optimization as an immunogen to elicit antibodies like VRC01.

HIV-1 gp120 outer domain: isotopic incorporation

The differential glycosylation indicated that mass spectrometric characterization of isotope incorporation would require analysis of tryptic peptides. Analysis of the sequence for the secreted HIV-1 gp120 outer domain indicated twelve potential tryptic peptides. Of these, nine peptides contained the sequence signature for N-linked glycosylation (N-X-T or N-X-S), one peptide comprised the C terminal His tag, and the two remaining peptides had similar masses (842.54 and 835.43 amu). The heptapeptide TIIVQLR at 842.54 amu, was observed in the mass spectra of outer domain produced in unlabeled media, but the predicted 835.43 amu fragment was absent. The percentage incorporation of ^{15}N for the tryptic peptide TIIVQLR in outer domain was estimated to be $85 \pm 4\%$ (Kubinyi 1991) (Table 1), and the percent incorporation in the best-correlated calculated spectrum was 87% (Fig. 2c).

A similar tryptic digest, MALDI TOF mass spectral analysis on outer domain expressed in $^{15}\text{N}/^{13}\text{C}$ -labeled media showed a much more complex m/z distribution (Fig. 2d). We attribute this to the presence of multiple species with differential amounts of isotope incorporation as well as possible overlap from the predicted 835.43 amu peptide, though we did not observe this peptide in the ^{15}N -labeled samples. When calculating the percent ^{13}C incorporation, we assumed that a similar incorporation of ^{15}N occurs in the double labeled medium, giving an estimated ^{13}C incorporation of $85 \pm 2\%$ with best correlation at 84% ^{13}C incorporation (Table 1 and Fig. 2d).

HIV-1 gp120 outer domain: NMR characterization

One dimensional proton NMR spectra of deglycosylated HIV-1 gp120 outer domain (Fig. 3a) exhibits a well dispersed spectrum 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 of very high quality and exhibits resolved backbone and side-chain amides;

Table 1 Protein expression levels and isotope enrichment levels of HIV-1 gp120 outer domain using an adenovirus vector-based mammalian expression system

	Unlabeled	^{15}N CGM6000	$^{15}\text{N}/^{13}\text{C}$ CGM6000
HIV-1 gp120 outer domain*	34 ± 5 mg/l	45 ± 5 mg/l	44 ± 2 mg/l
^{15}N and ^{13}C isotope incorporation levels [#]	N/A	$85 \pm 4\%$	$85 \pm 4\%$, $85 \pm 2\%$

* Yields are based on the total amount of protein obtained after nickel- and antibody-affinity chromatography. Yields are reported as mean \pm SEM

[#] Percent ranges of isotope incorporation were estimated using the subset of ^{15}N , ^{13}C enrichment levels for which the mass of the highest-intensity computed peak matched the mass of an observed m/z peak: 852.5 amu for ^{15}N heptapeptide and 884.5 amu for $^{15}\text{N}/^{13}\text{C}$ heptapeptide (see Table S1, Fig. 2d, and methods)

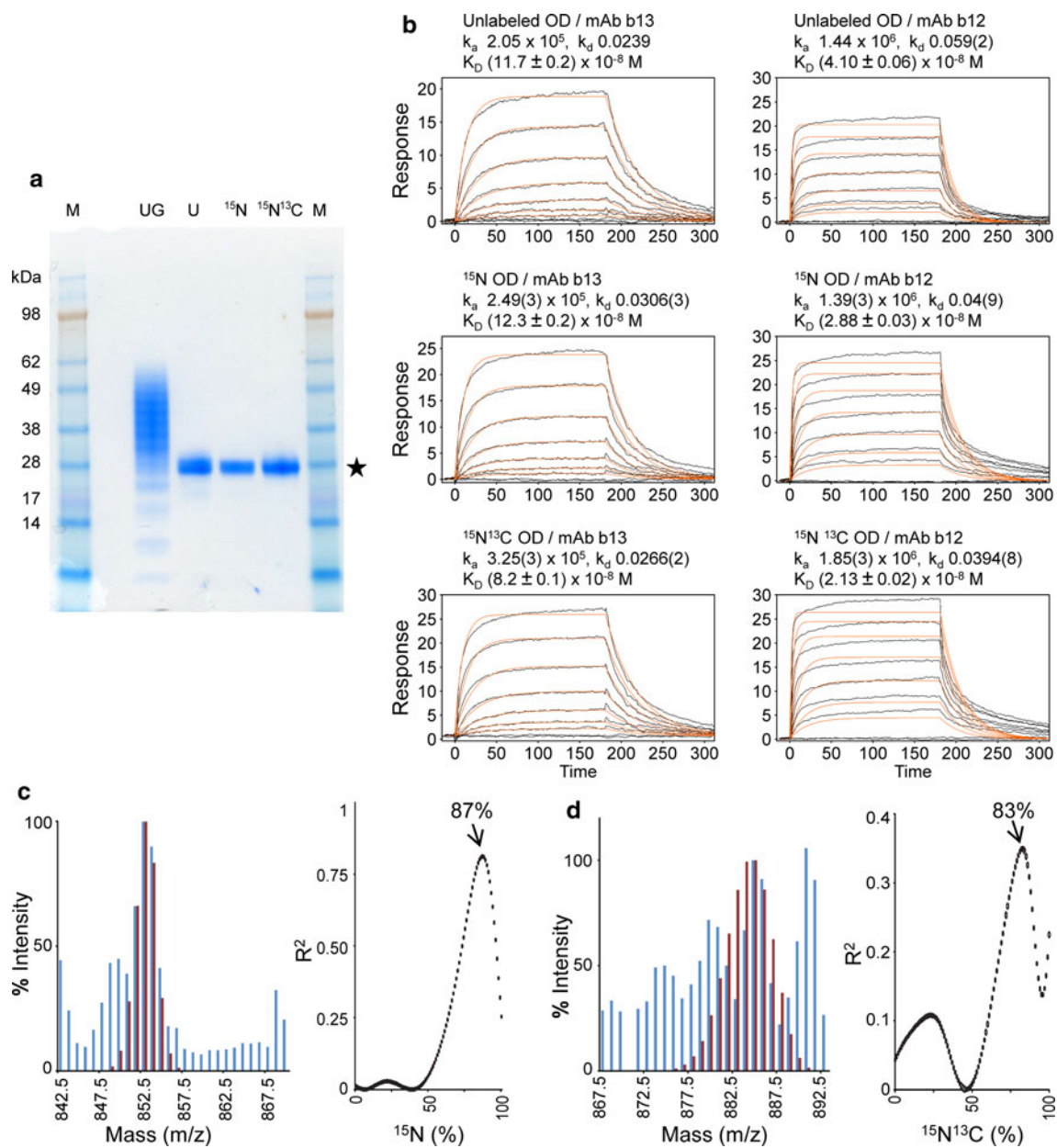


Fig. 2 Characterization of isotopically enriched HIV-1 gp120 Outer domain expressed using the adenoviral expression system. Production of isotopically enriched correctly folded post-translationally modified proteins is feasible using the adenoviral expression system. **a** SDS-PAGE analysis of the HIV-1 gp120 outer domain. Lane UG glycosylated outer domain demonstrates the micro heterogeneity observed in the glycans, Lanes U, ^{15}N , $^{15}\text{N}/^{13}\text{C}$ deglycosylation of unlabeled, ^{15}N , $^{15}\text{N}/^{13}\text{C}$ labeled gp120 outer domain with Endoglycosidase H resulted in a 28 kDa deglycosylated protein that was used for biophysical measurements (gel filtration profiles are shown in Fig. S6). Lane M Molecular weight markers. **b** Surface plasmon resonance analysis of deglycosylated unlabeled, ^{15}N and $^{15}\text{N}/^{13}\text{C}$ labeled binding to monoclonal antibodies b12 and b13 demonstrates that the expressed protein is correctly folded and biologically active. **c** Mass spectral analysis of a tryptic peptide fragment TIIVQLR used to determine % incorporation of ^{15}N . A comparison of experimental

(blue histogram) and computed (maroon histogram) pattern for 87% incorporation of ^{15}N is shown (left panel). The correlation between observed experimental pattern and computed patterns are shown for each percentage incorporation of ^{15}N . **d** Mass spectral analysis of a tryptic peptide TIIVQLR to determine % incorporation of ^{13}C . A comparison of experimental (blue histogram) and computed (maroon histogram) pattern for 84% incorporation of ^{13}C is shown (left panel). Although the maximum of the correlation for ^{13}C incorporation is at 83%, the best fit of the experimental and computational ^{13}C incorporation was estimated to be 84%, which allows the highest computational peak to match one of the experimentally-observed modes (see “Supplementary methods”). The correlations between observed experimental pattern and computed patterns are shown for each percentage incorporation of ^{13}C with a fixed ^{15}N incorporation of 84% (right panel)

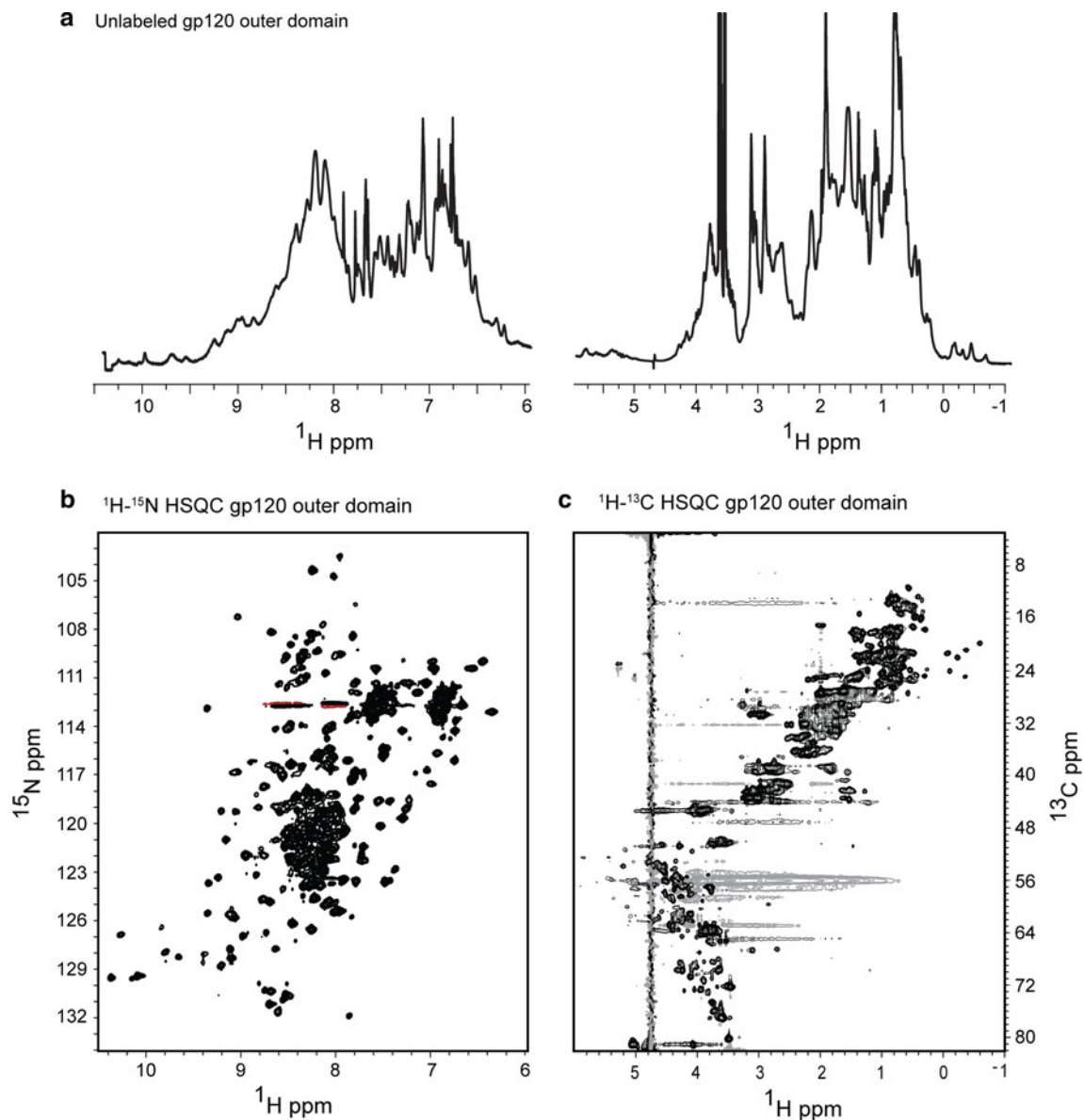


Fig. 3 Characterization of the HIV-1 gp120 outer domain expressed and purified using the adenoviral expression system. One dimensional proton spectra of deglycosylated unlabeled HIV-1 gp120 outer domain acquired at 600 MHz equipped with a room temperature probe at 25°C is shown in (a). ^1H - ^{15}N and ^1H - ^{13}C HSQC acquired at

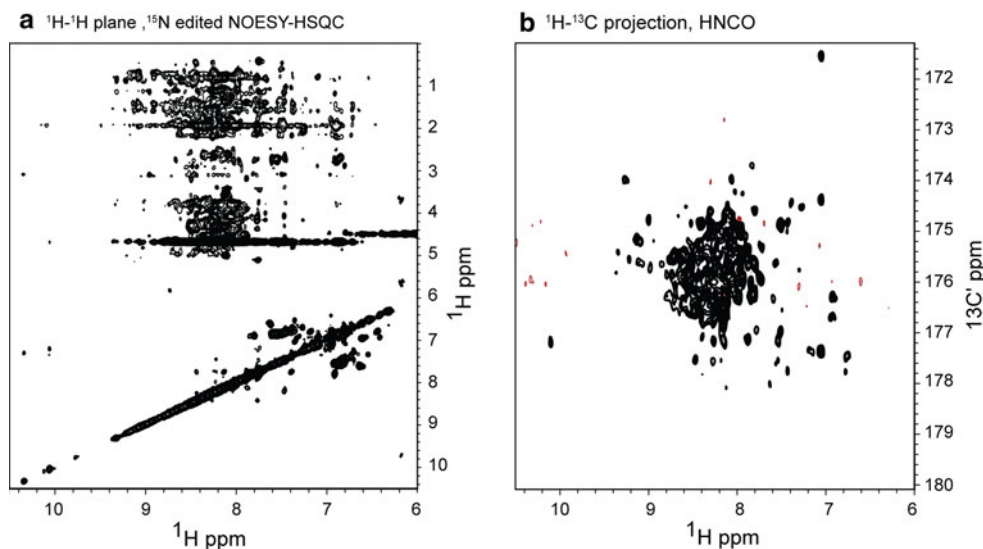
900 MHz and 25°C are shown in (b) and (c), respectively. The outer domain is an example of a secreted disulfide bonded glycoprotein that is functionally active and the excellent signal to noise demonstrates the success of the expression method

remarkably, we are able to account for 188 of the expected 210 cross peaks in this spectrum (Fig. 3b). Some crowding is observed in the central region of the HSQC spectrum (δ_{H} 7.8–8.2 and δ_{C} 118–124 ppm), likely arising from overlap of backbone amide protons, residual *N*-acetyl glucosamine resonances, or the remaining His tag used for purification. The ^1H - ^{13}C HSQC (Fig. 3c) exhibits very good chemical shift dispersion, where numerous H α and oxymethine protons can be seen as well as the presence of many upfield shifted methyl resonances, all of which are indicative of a structured protein. The excellent signal to noise further

demonstrates the success of the adenoviral expression method.

A high quality ^1H - ^1H plane of a ^{15}N edited NOESY-HSQC spectrum (Fig. 4a) shows the feasibility of obtaining sufficient numbers of NOE assignments for structure calculations. Finally, the ^1H - ^{13}C projection of a 3D HNCO spectrum (Fig. 4b) 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. Thus, a good quality H-H plane of a ^{15}N

Fig. 4 Heteronuclear NMR spectroscopy of $^{15}\text{N}/^{13}\text{C}$ HIV-1 gp120 outer domain. A ^1H - ^1H plane of a 3D ^{15}N edited NOESY HSQC acquired on a $\sim 400\ \mu\text{M}$ ^{15}N gp120 outer domain at 900 MHz is shown in (a). In b, a ^1H - ^{13}C projection of a 3D HNCOC spectrum acquired on a $\sim 400\ \mu\text{M}$ $^{15}\text{N}/^{13}\text{C}$ gp120 outer domain at 900 MHz further demonstrates that the adenovirus-vectored mammalian expression system can provide the necessary isotopically enriched samples for heteronuclear NMR spectroscopy



edited NOESY-HSQC (Fig. 4a) as well as a ^1H - ^{13}C projection of a 3D HNCOC spectrum (Fig. 4b) demonstrates that the mammalian expression system can provide the necessary isotopically enriched samples for structural and dynamic characterization of proteins and protein complexes.

Discussion

The relative immunogenicity of a particular portion of a protein is often strongly correlated with its flexibility. For example, seven flexible loops of the adenovirus hexon with less than 1% of virion mass accounts for much of adenovirus immunogenicity (Huang et al. 2005; Roberts et al. 2006); moreover, the immunogenicity of an epitope placed into different scaffolds was found to correlate strongly with the flexibility of the transplanted epitope (Ofek et al. 2010). Thus the rational design of an HIV-1 envelope immunogen should take into account its flexibility.

Unfortunately, such information is not so easily attained. Despite a growing body of biophysical information for the HIV-1 gp120 glycoprotein—from crystal structures of gp120 in complex with various ligands (Kwong et al. 1998; Huang et al. 2005; Zhou et al. 2007, 2010; Chen et al. 2009; Pancera et al. 2010), isothermal titration calorimetry providing measurements of entropic and enthalpic changes (Myszka et al. 2000), and hydrogen deuterium exchange providing a coarse description of conformational stability (Kong et al. 2010)—an atomic-level description of HIV-1 gp120 flexibility has been missing, and NMR spectroscopy is one of the only ways to obtain such information.

One barrier to obtaining NMR spectroscopic information is the production of isotopically labeled HIV-1 gp120.

HIV-1 gp120 has only been expressed from eukaryotic cells, and efficient labeling from such cells has in the past proven problematic. Researchers have used partial or AATS labeling of proteins from mammalian systems to obtain structural analysis of a handful of eukaryotic proteins (Takahashi and Shimada 2010), and a number of research groups have also focused their efforts on developing cost-effective media suitable for mammalian cell growth (Skelton et al. 2010). Our goal was to develop an expression system that produced milligram quantities of isotopically enriched glycoproteins that would allow solution structural characterization. Towards this goal we adapted an adenoviral expression system to obtain post-translationally modified ^{15}N and $^{15}\text{N}/^{13}\text{C}$ labeled proteins. Our approach capitalized on the ability of adenoviruses to transfect mammalian cells with very high efficiency and to overexpress adenoviral genes to obtain isotopically enriched proteins in milligram quantities. We characterized the growth of A549 lung carcinoma cells used in the adenoviral system for protein expression in different growth media such as DMEM, ^{15}N , and $^{15}\text{N}/^{13}\text{C}$ labeled CGM6000. A549 cells exhibited similar growth characteristics in labeled CGM6000 media as compared to unlabeled DMEM (Fig. S1A). NMR analysis of larger proteins (>20–30 kDa) is aided by perdeuteration or random fractional deuteration of carbon-bound hydrogens in a protein, which can in principle be obtained by growth of mammalian cells in D_2O -containing media.

Overall, adenovirus-vector expression in mammalian cells offers an alternative to manual optimization of media components, refolding protocols, and/or fermenter growth that may require additional equipment. The adenoviral expression system is straightforward to implement and general in utility. One drawback relative to other eukaryotic

systems is the time required to obtain recombinant virus (several weeks), which, nevertheless, is substantially less than many other steps of the NMR structure-determination process. At this stage we have only shown incorporation for ^{15}N and ^{13}C , not ^2H . However, the ability of adenovirus to produce GFP in 70% D_2O (Fig. 1) suggests that the requisite deuteration can be accomplished.

Acknowledgments We thank the NMR staff at the New York Structural Biology Consortium for assistance with instrumentation. We also thank Mary Ann Gawinowicz of the Protein Core Facility at Columbia University for N terminal sequencing and mass spectrometry, L. Shapiro for editorial assistance with the manuscript, and T. Waybright and T. Veenstra at the NCI for assistance with mass spectral data acquisition. We also thank the members of the structural biology and structural bioinformatics core sections 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).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Aoki K, Barker C, Danthine X, Imperiale MJ, Nabel GJ (1999) Efficient generation of recombinant adenoviral vectors by Cre-lox recombination in vitro. *Mol Med* 5(4):224–231
- Arata Y, Kato K, Takahashi H, Shimada I (1994) Nuclear magnetic resonance study of antibodies: a multinuclear approach. *Methods Enzymol* 239:440–464
- 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(7):1212–1221
- Barouch DH, Nabel GJ (2005) Adenovirus vector-based vaccines for human immunodeficiency virus type 1. *Hum Gene Ther* 16(2):149–156
- Bessette PH, Åslund F, Beckwith J, Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci USA* 96(24):13703–13708
- Chen L, Kwon YD, Zhou T, Wu X, O'Dell S, Cavacini L, Hessell AJ, Pancera M, Tang M, Xu L, Yang ZY, Zhang MY, 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(5956):1123–1127
- Cornelis P (2000) Expressing genes in different *Escherichia coli* compartments. *Curr Opin Biotechnol* 11(5):450–454
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293
- Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* 17(4):353–358
- Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* (1969) 96(2):432–440
- 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(51):12713–12718
- Huang JT, Schneider RJ (1990) Adenovirus inhibition of cellular protein synthesis is prevented by the drug 2-aminopurine. *Proc Natl Acad Sci USA* 87(18):7115–7119
- Huang C-c, Tang M, Zhang M-Y, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, Kwong PD (2005) Structure of a V3-containing HIV-1 gp120 core. *Science* 310(5750):1025–1028
- Kadokura H, Beckwith J (2001) The expanding world of oxidative protein folding. *Nat Cell Biol* 3(11):E247–E249
- Kanekiyo M, Panet A, Xu L, Boyington JC, Louder MK, Yang Z, Mascola JR, Nabel GJ (2010) Molecular design and biochemical characterization of HIV-1 clade C Env gp120 derivatives. *AIDS Vaccine* 2010, Atlanta
- Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* (1969) 89(3):496–514
- Keller RLJ (2004) The computer aided resonance assignment tutorial. Cantina
- Klein-Seetharaman J, Reeves PJ, Loewen MC, Getmanova EV, Chung J, Schwalbe H, Wright PE, Khorana HG (2002) Solution NMR spectroscopy of [α - ^{15}N]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(6):3452–3457
- Klein-Seetharaman J, Yanamala NVK, Javed 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(10):3409–3413
- Kong L, Huang C-c, Coales SJ, Molnar KS, Skinner J, Hamuro Y, Kwong PD (2010) Local conformational stability of HIV-1 gp120 in unliganded and CD4-bound states as defined by amide hydrogen/deuterium exchange. *J Virol* 84(19):10311–10321
- Korber B, Foley B, Kuiken C, Pillai S, Sodroski J (1998) Numbering positions in HIV relative to HXB2CG. In: Korber B, Kuiken C, Foley B et al (eds) *Human retroviruses and AIDS*. Theoretical biology and biophysics group, Los Alamos, pp 102–111
- Kubinyi H (1991) Calculation of isotope distributions in mass spectrometry. A trivial solution for a non-trivial problem. *Analytica Chimica Acta* 247 (1):107–119
- Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393(6686):648–659
- 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 (2010) Structural biology and the design of effective vaccines for HIV-1 and other viruses. In: Georgiev SV (ed) *National Institute of Allergy and Infectious Disease, NIH, vol 3. Humana*, pp 387–402
- 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(4):295–304
- 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(1):87–94
- Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, Hendrickson WA, Wyatt R, Sodroski J, Doyle ML (2000) Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci USA* 97(16):9026–9031

- Nabel GJ (1999) Development of optimized vectors for gene therapy. *Proc Nat Acad Sci USA* 96(2):324–326
- Ofek G, Guenaga FJ, Schief WR, Skinner J, Baker D, Wyatt R, Kwong PD (2010) Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Nat Acad Sci USA* 107(42):17880–17887
- Ohno T, Yang Z, Ling X, Jaffe M, Nabel EG, Normolle D, Nabel GJ (1997) Combination gene transfer to potentiate tumor regression. *Gene Ther* 4(4):361–366
- Pancera M, Majeed S, Ban YE, Chen L, Huang CC, Kong L, Kwon YD, Stuckey J, Zhou T, Robinson JE, Schief WR, Sodroski J, Wyatt R, Kwong PD (2010) Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc Nat Acad Sci USA* 107(3):1166–1171
- Roberts DM, Nanda A, Havenga MJE, Abbink P, Lynch DM, Ewald BA, Liu J, Thorner AR, Swanson PE, Gorgone DA, Lifton MA, Lemckert AAC, Holterman L, Chen B, Dilraj A, Carville A, Mansfield KG, Goudsmit J, Barouch DH (2006) Hexon-chimaeric adenovirus serotype 5 vectors circumvent pre-existing anti-vector immunity. *Nature* 441(7090):239–243
- Schein CH, Noteborn MHM (1988) Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Nat Biotech* 6(3):291–294
- Shindo K, Masuda K, Takahashi H, Arata Y, Shimada I (2000) Letter to the editor: backbone ^1H , ^{13}C , and ^{15}N resonance assignments of the anti-dansyl antibody Fv fragment. *J Biomol NMR* 17(4):357–358
- Skelton D, Goodyear A, Ni D, Walton W, Rolle M, Hare J, Logan T (2010) Enhanced production and isotope enrichment of recombinant glycoproteins produced in cultured mammalian cells. *J Biomol NMR* 48(2):93–102
- Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-tailored water suppression for ^1H - ^{15}N HSQC experiments optimized to retain full sensitivity. *J Magn Reson Ser A* 102(2):241–245
- Takahashi H, Shimada I (2010) Production of isotopically labeled heterologous proteins in non-*E. coli* prokaryotic and eukaryotic cells. *J Biomol NMR* 46 (1):3–10
- Wang L, Cheng C, Ko S-Y, Kong W-P, Kanekiyo M, Einfeld D, Schwartz RM, King CR, Gall JGD, Nabel GJ (2009) Delivery of human immunodeficiency virus vaccine vectors to the intestine induces enhanced mucosal cellular immunity. *J Virol* 83(14):7166–7175
- 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(1):49–53
- Wu L, Zhou T, Yang Z-y, Svehla K, O'Dell S, Louder MK, Xu L, Mascola JR, Burton DR, Hoxie JA, Doms RW, Kwong PD, Nabel GJ (2009) Enhanced exposure of the CD4-binding site to neutralizing antibodies by structural design of a membrane-anchored human immunodeficiency virus type 1 gp120 domain. *J Virol* 83(10):5077–5086
- Wyss DF, Withka JM, Knoppers MH, Sterne KA, Recny MA, Wagner G (1993) ^1H resonance assignments and secondary structure of the 13.6 kDa glycosylated adhesion domain of human CD2. *Biochemistry* 32 (41):10995–11006
- 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(3):534–542
- Xu L, Boyington JC, Zhou T, Zhu J, Wu L, Louder MK, Akahata W, Kwong PD, Mascola JR, Yang Z, Nabel GJ (2010) Optimization of the outer domain of gp120 as an HIV-1 vaccine immunogen. *Aids Res Hum Retrovir* 26(10):A166–A167
- Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994) A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J Am Chem Soc* 116:11655–11666
- Yang X, Tomov V, Kurteva S, Wang L, Ren X, Gorny MK, Zolla-Pazner S, Sodroski J (2004) Characterization of the outer domain of the gp120 glycoprotein from human immunodeficiency virus type 1. *J Virol* 78(23):12975–12986
- Zhou T, Xu L, Dey B, Hessel 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(7129):732–737
- Zhou T, Georgiev I, Wu X, Yang Z-Y, Dai K, Finzi A, Do Kwon Y, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD (2010) Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* 329(5993):811–817