4724 *Biochemistry* **2010,** *49,* 4724–4731
DOI: 10.1021/bi100535k

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Interactions of Protein and Nucleic Acid Components of Hepatitis C Virus As Revealed by Fourier Transform Infrared Spectroscopy[†]

Pedro Carmona*[,‡] and Marina Molina[§]

[‡]*Instituto de Estructura de la Materia (CSIC), Serrano 121, 28006 Madrid, Spain, and* [§]*Departamento de Química Orgánica, Escuela Universitaria de Optica, Arcos de Jalón, s/n, 28037 Madrid, Spain*

Received April 9, 2010; Revised Manuscript Received May 14, 2010

ABSTRACT: The secondary structure of the loop IIId domain in the RNA of hepatitis C virus (HCV) is well-conserved among different genotypes of HCV, which suggests that the nucleocapsid proteins may interact with the genome RNA through this loop structure. Using infrared spectroscopy, we monitored structural changes occurring in HCV core protein and loop IIId upon formation of nucleocapsid-like particles (NLPs). The protein secondary structure of these particles involves β-sheet enrichment in relation to its protein monomer. The phosphodiester backbone vibrations of loop IIId reflect the predominant C3′-endo conformation of the riboses involved in the RNA A-form and reveal the packaging-imposed transition of the said RNA segments toward single-stranded structure within the NLPs. Intermolecular protein−nucleic acid contacts in these particles involve RNA phosphate groups and positively charged amino acid residues such as arginine and lysine. Two-dimensional correlation spectroscopic analysis of the spectra measured in the course of deuteration shows synchronous cross-peaks correlating two bands assigned to guanine and arginine side chain, which is consistent with the presence of guanine−arginine interactions in these NLPs. This is also supported by the kinetically favored formation of NLPs having HCV core protein and guanine-enriched synthetic oligonucleotides. We also found that these NPLs are fully permeable to water molecules.

Hepatitis C virus (HCV)[1] is a major public health concern worldwide. Approximately 2% of the world's population is infected with the virus, which is the major cause of non-A, non-B hepatitis and which often leads to cirrhosis of the liver or hepatocellular carcinoma (*1*).Various approaches have been used to characterize the biological features of HCV structural proteins. These include attempts to purify and identify HCV-like (nucleocapsid-like) particles via expression of core protein alone or together with the other structural proteins, E1 and E2, in *Escherichia coli*, yeast, and animal cells (*2*−*4*), as well as using in vitro systems (*5*−*7*). Nucleocapsid-like particles (NLPs) that are formed by the structural elements of viruses have received considerable attention over the past two decades because they serve as good model systems for the study of the assembly of macromolecular complexes related to HCV. The pathway of NLP assembly can differ depending on the nature of the stabilizing interactions such as protein−protein, protein−nucleic acid, and metal ion- or disulfide bond-mediated interactions (*7*). Protein−protein interactions in NLPs are relatively strong and can result in the formation of stable structures. Purified recombinant core protein and defined structured oligonucleotides have been found to be sufficient for self-assembly of nucleocapsid-like particles in vitro (*5*). These particles have a regular, spherical morphology with a modal distribution of diameters of ∼60 nm. Moreover, NLP formation depends on the 124 highly basic N-terminal residues of the core protein and on oligonucleotides corresponding to the 5′-untranslated region that harbors predicted secondary structural elements (*5*, *8*). These in vitro assembled NLPs were essentially identical in morphology and size and thus offer a means of systematically investigating the structure and function of the HCV core protein and self-assembled nucleocapsids (*5*, *8*, *9*).

However, the specific interactions between the core protein and the viral genomic RNA and the process of assembly into nucleocapsid particles still remain to be demonstrated. Although the availability of natural HCV nucleocapsids for investigation, usually isolated from infected human liver, is severely limited, recombinantly produced HCV proteins and RNA offer abundant systems in which to study the in vitro assembly of the HCV nucleocapsid. Along these lines, a simple, robust, and efficient method for performing extensive studies of the HCV nucleocapsid in vitro has recently been developed (*9*). This method involves a simple and rapid in vitro assay in which the progress of assembly is monitored by measuring an increase in turbidity.

Fourier transform infrared (FTIR) spectroscopy offers advantages for structural studies of viral components and precursor systems. Recent advances in two-dimensional correlation spectroscopy (2DCOS) permit us to probe the structure and dynamics of viral nucleic acid and protein components and their interactions. For spectral analysis, two-dimensional correlation of intensity changes has proven to be useful in enhancing spectral resolution and determining the order of events during a perturbation (*10*−*12*). In this regard, the infrared measurement of H−D exchange in proteins and nucleic acids for subsequent

[†]Financial support from Spanish Ministerio de Ciencia e Innovación (Project CTQ2006-04161/BQU).

*To whom correspondence should be addressed: Instituto de Estructura de la Materia (CSIC), Serrano 121, 28006 Madrid, Spain. Telephone: +34 91 5616800. Fax: +34 91 5645557. E-mail: p.carmona@iem.cfmac.csic.es.

[1]Abbreviations: HCV, hepatitis C virus; HCVc-120, recombinant truncated protein covering the first 120 amino acids of HCV core protein; NLPs, nucleocapsid-like particles.

 Published on Web 05/17/2010

2DCOS analysis has become increasingly important as a means of characterizing the structure of these biomolecules and their intermolecular complexes (*10*, *13*−*15*). From this point of view, we apply here infrared spectroscopy to characterize changes in protein and nucleic acid structure as well as protein−nucleic acid interactions from the unassembled state to the nucleocapsid assembled state. With this aim, we have studied NLPs consisting of the HCV core protein and the loop IIId domain of the HCV RNA 5′-untranslated region which are well conserved among different HCV isolates (*16*, *17*).

MATERIALS AND METHODS

Sample Preparation. The following reverse-phase HPLC-purified oligonucleotide was obtained from Biomedal (Sevilla, Spain): 5′-UAGCCGAGUAGUGUUGGGUCGCGAAAGG-CUU-3′ which constitutes so-called loop IIId of the viral genome (*18*, *19*). The size and integrity of this 31mer RNA were checked by polyacrylamide gel electrophoresis, and its concentration was determined via measurement of the absorbance at 260 nm and using an extinction coefficient of 25 g^{-1} cm^{-1} L.

Recombinant truncated protein covering the first 120 amino acids of the HCV core protein (HCVc-120) was acquired from Milan Analytica AG. It was prepared via expression in *Pichia pastoris* and isolated in a 1.0 mg/mL stock solution in 50 mM phosphate buffer [150 mM NaCl (pH 7.0)]. The protein purity was found to be greater than 95% as determined by PAGE and the A_{260}/A_{280} (0.77) absorbance ratio.

In Vitro Assembly Reactions. Fifty microliters of assembly buffer [1.7 mM magnesium acetate, 100 mM potassium acetate, and 25 mM Hepes (pH 7.4)] containing 8 μg of HCVc-120 core protein was placed in a microcuvette (Eppendorf) to which an equal volume of an assembly buffer solution of RNA was added, and the contents were mixed vigorously to give a 2:1 protein:RNA mole ratio. Another assembly reaction was conducted at a 2:1 protein:RNA ratio using HCV RNA loop IIId and 200 mM NaCl to assess the influence of this salt on NLP formation. Absorbance was monitored with a Perkin-Elmer Lambda 50 spectrophotometer and recorded every 2 s for 10 min. Approximately 8 s elapsed before the first time point was measured.

Electron Microscopy. Samples (~10 μL) collected after in vitro assembly were adsorbed to 400-mesh carbon-Formvar grids (Leica Microsystems) for 5 min. Grids were washed with deionized water and stained for ~2 min with 2.0% phosphotungstic acid (PTA). Grids were then blotted with filter paper, air-dried for at least 5 min, and examined on a Zeiss 910 electron microscope with an acceleration voltage of 80 kV at a magnification of 100000×.

FTIR Experiments. The protein stock solution was brought to a concentration of 30.0 mg/mL in 50 mM phosphate buffer [150 mM NaCl (pH 7.0)] using Microcon centrifugal filter devices (Millipore). This solution was then allowed to dialyze against 50 mM phosphate D_2O buffer [150 mM NaCl (pD 7.1)] for infrared spectroscopic measurements of pure deuterated protein in this buffer. With this aim, the dialysis membrane used had a molecular cutoff of 3500. Infrared spectra were recorded with a Perkin-Elmer 1725X interferometer at 2 cm^{-1} resolution and obtained as averages of 200 scans. Samples of the protein solution were placed in a cell fitted with ZnSe windows and a 12 μm spacer. To compensate for aqueous absorptions, the spectra of the protein-free sample were also measured under the same experimental conditions described above in the same kind of cells. The spectral contributions from residual water vapor in the sample chamber were eliminated using a set of water vapor spectra measured under identical conditions. The subtraction factor was varied until the nonabsorbing region above 1700 cm^{-1} was featureless. The resulting difference spectra were subsequently smoothed with a nine-point Savitsky−Golay function to reduce the noise. Spectral data were treated with GRAMS/AI (Thermogalactic) which includes baseline correction, smoothing, and solvent subtraction.

D_2O for H−D isotopic exchange was acquired from Aldrich, and its minimum isotopic purity was 99.9 at. % D.

For infrared spectroscopic measurements of the pure 31mer RNA, this oligonucleotide was prepared at a concentration of 2% (w/w) and dialyzed against a 50 mM NaCl aqueous solution by using a membrane with a molecular cutoff of 3500. The resulting oligonucleotide solution was then allowed to dialyze against a 50 mM NaCl D_2O solution to deuterate the 31mer RNA. The H_2O and D_2O solvents and NaCl were free from RNAases and DNAases and purchased from Fluka. Approximately 0.5 mg of oligonucleotide in this solution was deposited on a ZnSe window, dried, and placed in an hermetic cell through which a stream of nitrogen, bubbling in a saturated solution of KBr in D_2O, was passed continuously to provide a relative humidity of 82%. The spectrum of the deuterated oligonucleotide film was recorded on said infrared spectrometer at 2 cm^{-1} resolution and obtained as an average of 200 scans. The spectral contribution from residual D_2O vapor in the sample spectrum was eliminated using a set of D_2O vapor spectra measured under identical conditions. The resulting spectra were normalized by using the symmetric PO_2^- stretching band appearing in the 1100−1070 cm^{-1} range. This band is often used as an internal standard for spectral normalization because it is rather insensitive to transitions between the single- and double-stranded forms (*20*).

The in vitro assembly reaction was conducted via addition of 0.1 mL of assembly buffer [1.7 mM magnesium acetate, 100 mM potassium acetate, and 25 mM Hepes (pH 7.4)] containing 80 μg of protein to an equal volume of RNA solution in the same buffer, resulting in a mixed solution with a 2:1 protein:RNA mole ratio. This solution was then deposited on a ZnSe window and allowed to settle overnight at 4 °C, and the supernatant was removed. This assembly reaction was then conducted again on the same ZnSe window, whereby a film of NLPs was formed for subsequent spectral measurements. With this aim, the ZnSe window was placed in a hermetic cell through which a stream of nitrogen, bubbling in a saturated solution of KBr in D_2O, was passed continuously. Spectra were measured every 2 min for the first 90 min while the hydrogen exchange progressed, and they resulted from accumulation of eight scans at 2 cm^{-1} resolution over a range of 4000−700 cm^{-1}. Thirty-two scans were collected for the next spectra measured at time intervals of 5 and 10 min, and the spectrum collected after exchange for 150 min was used as the fully deuterated spectrum. The effects of residual D_2O vapor in the sample spectra were eliminated as described above.

The incorporation of deuterium into the protein backbone was readily followed by monitoring the disappearance of the amide II band. The fraction of the unexchanged amide groups was calculated on the basis of the $\Phi_h(t)$ ratio of the amide II band absorbance at time t and t_0. The amide II band area at any time point was measured after subtraction of the fully deuterated spectrum to eliminate the influence of the guanine 1580 cm^{-1} band. With regard to the H−D exchange of the aqueous solvent within NLPs, it was measured through the area of the δOD_2 band

near 1208 cm^{-1} calculated by subtraction of the spectrum at time t_0 from the spectrum at every time t. The exchanged fraction of solvent [$\theta_d(t)$] is given by the ratio $\theta_d(t) = A_D(t)/A_{D\infty}$, where $A_{D\infty}$ corresponds to the δOD_2 band area of the fully deuterated spectrum.

Two-Dimensional (2D) Correlation Analysis. The infrared spectra measured for H−D exchange were used for 2D correlation analysis. We employed 2D-Pocha, written by Dr. Adachi and Dr. Y. Ozaki (Kwansei-Gakuin University, Tokyo, Japan), which was programmed on the basis of the developed algorithm of generalized 2D correlation spectroscopy (*12*). This two-dimensional (2D) correlation analysis allows the correlation of the dynamic fluctuations of infrared bands in a series of spectra measured at different times. Cross-correlation analysis provides 2D spectra that are defined by two independent frequencies, ν_1 (abscissa) and ν_2 (ordinate). The synchronous 2D correlation spectra of dynamic spectral intensity variations represent the simultaneous occurrence of coincidental changes in spectral intensities measured at ν_1 and ν_2. Correlation peaks appear at both diagonal (autopeaks) and off-diagonal peaks (cross-peaks). Cross-peaks, which can be positive or negative, reflect correlated changes in functional groups within the biomolecular system that occur simultaneously in the same direction (+) or in the opposite direction (−). By contrast, the asynchronous 2D correlation representation is characterized by missing autopeaks and asymmetric cross-peaks which reveal uncorrelated (i.e., out-of-phase) behavior of two bands. The time-dependent changes in the spectral intensities are shown as two-dimensional contours, called synchronous and asynchronous maps (S- and A-maps, respectively), which correlate in-phase (synchronized) and out-of-phase (unsynchronised) intensity changes at two frequencies, respectively. Cross-peaks, which occur in both the S- and A-maps, provide information about the temporal order of the spectral changes. Cross-peaks with the same sign in the S- and A-maps indicate that a spectral change in the ν_1 band (abscissa) occurs predominantly before that of the ν_2 (ordinate) band in the sequential order of time. The temporal order is reversed for opposite signs of the S- and A-map cross-peaks.

RESULTS

Infrared Spectra of Nucleocapsid Components and Structural Integrity of the NLPs. Figure 1A shows the infrared normalized spectra of deuterated NLPs and their protein and RNA constituents. The spectrum of NLPs is complex, consisting of contributions from both viral RNA and protein components. To resolve these components in the 1720−1590 cm^{-1} region, second-derivative spectra have been calculated (Figure 1B). The spectral profiles of HCVc-120 and NLPs in the 1700−1600 cm^{-1} range (Figure 1A) have absorption maxima near 1645 cm^{-1}, what indicates that a major fraction of the protein molecule remains in an unordered structure upon nucleocapsid formation. However, binding of this protein to the 31mer RNA through nucleocapsid formation led to significant changes in the protein second-derivative spectra, namely, intensity increases for the band components near 1635 and 1690 cm^{-1} that can be attributed to β-sheets (*21*−*24*). The fact that this secondary structure is formed to some extent is also supported by the characteristic νNH protein band which is visible during the H−D isotopic exchange process, as described later. The second-derivative band component of NLPs near 1650 cm^{-1} could be ascribed to α-helices, but the unambiguous assignment of this band is difficult because of the potential contribution from deuterated cytosine bases (*20*) of the 31mer RNA (Figure 1B).

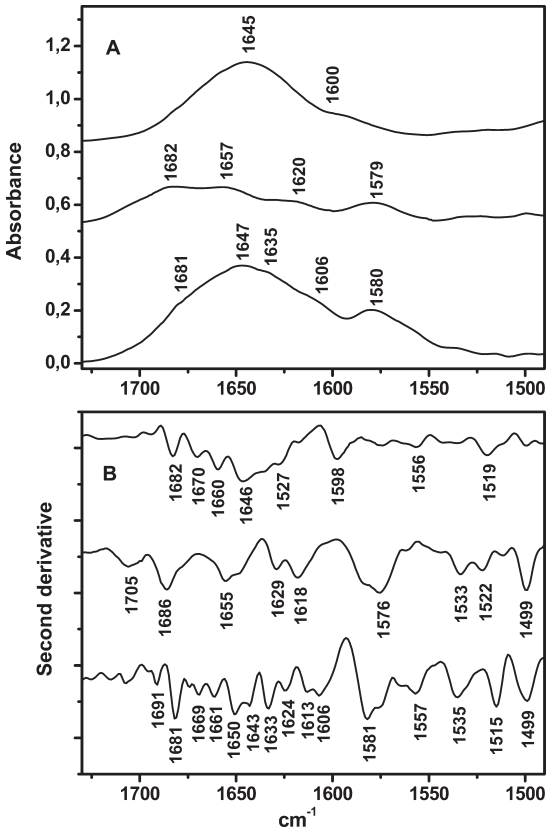

FIGURE 1: (A) Infrared spectra of HCVc-120 in deuterated buffer [30 mg/mL protein in 50 mM phosphate and 150 mM NaCl (pD 7.0)] (top), of a film of deuterated 31mer RNA at 82% relative humidity (middle), and of a deuterated NLP film at 82% relative humidiy (bottom). (B) Second-derivative spectra. The ordinate represents the absorbance/wavenumber^2 (scale divisions of 0.0001 absorbance unit/cm^{-2}).

The 1800−1500 cm^{-1} region of the 31mer RNA consists of absorption bands originating from the in-plane double-bond vibrations of the bases (Figure 1A). The strong band at 1682 cm^{-1} in the spectrum of this oligonucleotide stems from the overlapping contribution of the $\nu C_6{=}O_6$ and $\nu C_4{=}O_4$ vibrations of base-paired guanosines and uracils, respectively (*20*, *25*−*28*). This spectral region also exhibits two bands located at 1657 and 1620 cm^{-1} that are masked by the HCVc-120 core protein in the spectrum of NLPs. The former can be ascribed to overlapping bands generated by free and base-paired cytosine bases as well as by free uracils, and the latter can be attributed to in-plane ring vibrations of adenine and cytosine bases (*20*, *25*−*28*). The 1606 cm^{-1} band in the spectrum of NLPs is unambiguously attributable to arginine residues, particularly to the $\nu_{as}(CN_3D_5^+)$ vibrational mode (*11*, *24*), because the undeuterated adenine and cytosine bands near 1600 cm^{-1} shift toward the 1635−1620 cm^{-1} range upon deuteration (*20*, *25*−*28*). The strongest band in the 1600−1500 cm^{-1} range arises mainly from the guanine C=N stretching vibrations near 1580 cm^{-1} (*20*, *25*, *28*) and is not masked by the spectrum of HCVc-120. The intensity of this band is slightly lower (<10%) in the unpackaged RNA than in NLPs and, accordingly, represents a rather small structural perturbation. This spectral feature can be interpreted in terms of infrared hyperchromism; i.e., this feature represents greater hyperchromic effects in the packaged RNA molecule and signifies an electronically weaker interaction between stacked bases of packaged RNA (*20*, *28*).

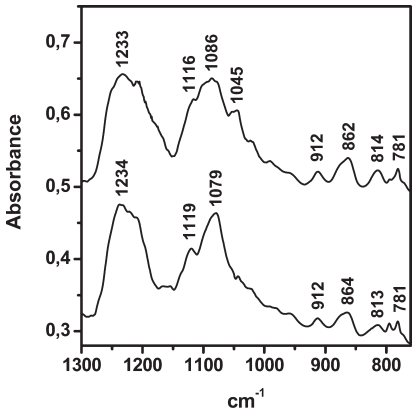

FIGURE 2: Infrared spectrum of a film of deuterated 31mer RNA at 82% relative humidity (top) and of a deuterated NLP film at 82% relative humidity (bottom).

The bands of backbone−sugar vibrations are located at 1500 cm^{-1}. They appear at the expected frequencies for the C3′-endo conformation of riboses characteristic of A-form geometry (*29*, *30*). Thus, the antisymmetric stretching vibration of the PO_2^- group appears near 1235 cm^{-1} (Figure 2) which is indicative of the A-type form having the said C3′-endo sugar puckering. The band at 1086 cm^{-1} has been assigned to the $\nu_s PO_2^-$ mode coupled with the C5′−O5′ vibrations (*31*) and shifts to 1079 cm^{-1} upon nucleocapsid formation. Among the salient different features of protein-free and protein-bound RNA is the said νPO_2^- frequency downshifting generated by interactions of HCVc-120 with phosphate groups in the ribose−phosphate backbone of 31mer RNA, which means that these phosphate groups are major binding sites preferred by this protein. This is also supported by the fact that NLP formation is inhibited by the presence of sodium chloride, as described later. The said significant shift to a lower frequency observed for the phosphodioxy marker of this oligonucleotide probably reflects direct interaction between positively charged side chains (positively charged arginine and lysine residues) and PO_2^-, resulting in the withdrawal of electrons from the PO bonds with attendant reduction in the bond stretching force constant. Thus, these amino acid residues, which constitute ~23.5% of the HCVc-120 protein sequence (*5*), can be a distinctive feature of HCVc-120 related to binding of the 31mer RNA. The 1086 cm^{-1} band is relatively broad because of the heterogeneity of the local environments of phosphate groups, which may reflect structural variety along the ribose−phosphate backbone of the viral RNA. The phosphodiester backbone vibrations coupled to the sugar motions generate the bands at 814 and 862 cm^{-1} in the spectrum of this oligonucleotide which reflect the predominant C3′-endo conformation of the riboses involved in the RNA A-form (*31*, *37*). The intensity of the 814 cm^{-1} band which is indicative of double-strand or base-paired segments is significantly weaker in the spectrum of NLPs, and this spectral change can be interpreted as evidence of the packaging-imposed transition of the said RNA segments toward single-stranded structure within the NLPs.

To assess the integrity or presence of NLPs in the samples containing the HCVc-120 protein and the 31mer RNA, Figure 3 shows a micrograph obtained for NLPs from an assembly reaction liquid sample prepared for infrared spectroscopy. The presence of spherical particles with an average diameter of ~40 nm was observed, which supports the possibility that the protein and RNA spectral changes described above are generated by formation of NLPs.

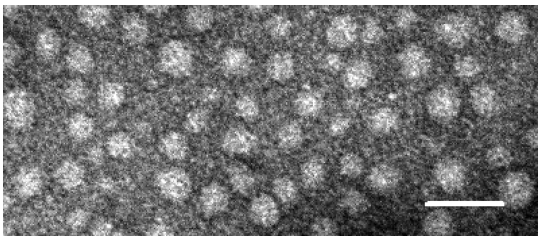

FIGURE 3: Electron micrograph of negatively stained samples of NLPs containing HCVc-120 protein and 31mer RNA at a 2:1 protein:RNA ratio. The bar is 100 nm.

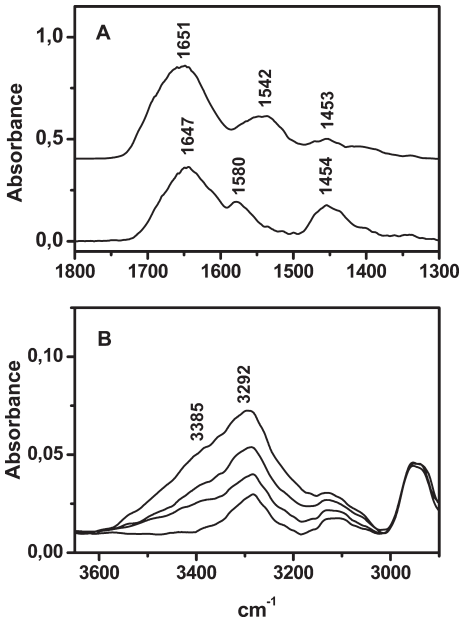

FIGURE 4: (A) Infrared spectra of a film of undeuterated (top) and fully deuterated (bottom) NLPs in the 1800−1300 cm^{-1} region, at deuteration time points of 0 and 150 min. (B) From top to bottom: spectral profile of NLPs in the 3600−3000 cm^{-1} region, at deuteration time points of 50, 80, 100, and 130 min.

Isotopic Exchange Characteristics of the NLPs. A series of spectra were recorded for NLPs as a function of the deuteration time, and the spectra corresponding to 0 and 150 min (fully deuterated NLPs) appear in Figure 4A. While the initial spectrum exhibited a characteristic amide II band at 1542 cm^{-1}, H−D exchange led to a time-dependent isotopic shift of this band from 1542 to 1454 cm^{-1}. Interestingly, the spectral profile in the 3600−3200 cm^{-1} region over the last 90 min of the deuteration process (Figure 4B) reveals the presence of a relatively narrow νNH band component at 3292 cm^{-1} which can be ascribed to an ordered protein backbone. In fact, this absorption maximum falls in the range that can be attributed to the so-called amide A band (3310−3270 cm^{-1}) (*32*, *33*), and the approximately 90 cm^{-1} half-bandwidth that is visible after the 100 min time point suggests the presence of an ordered hydrogen-bonded polypeptide backbone (*34*). This structural assignment is consistent with the protein secondary structure analysis described above through the amide I band indicating the presence of β-sheets in the HCVc-120 protein component of NLPs.

Because H−D exchange is a first-order reaction, the fraction of residual amide protons $\Phi_h(t)$ is expected to display a multiexponential decay corresponding to various groups, i, of amides characterized by a common period T_i

$$\Phi_h(t) = \sum_i a_i \exp(-t/T_i) \quad (1)$$

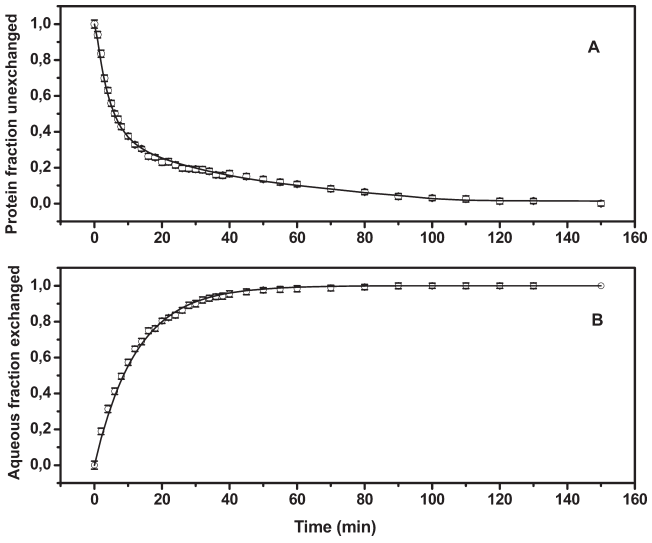

FIGURE 5: (A) Fraction of unexchanged protein amide groups in NLPs vs H−D exchange time. (B) Fraction of deuterated aqueous molecules in NLPs vs H−D exchange time.

where a_i is the proportion of group i. A three-exponential decay model was used for the fitting of the curve in Figure 5A. The T_1, T_2, and T_3 values were computed and found to be 5.1 ± 0.3, 36.6 ± 4, and 108.0 ± 8 min, respectively, and their fitted proportions (a_i) were 0.69 ± 0.031 (a_1), 0.21 ± 0.015 (a_2), and 0.10 ± 0.022 (a_3). The statistical criterion used to ensure the goodness of the three-exponential fit was the χ^2 minimization; i.e., the sum of the squares of the deviations of the theoretical curve from the experimental points is minimized. In addition, the number of three periods is consistent with the fact that, in general, all amide protons in proteins can be divided into three classes (*35*−*37*): (1) fast exchange protons, which are most likely located on the protein surface, on non-hydrogen-bonded amide groups or in regions that are easily solvent accessible, (2) amide protons with intermediate rates located in flexible buried regions, and (3) the slow exchange fraction located in the protein core region or in hydrogen-bonded structures such as β-sheets and helices.

The data depicted in Figure 5B show that ~50 min is sufficient to exchange practically 100% of the solvent molecules, and the intensity of the 1208 cm^{-1} band can be expressed by the following first-order equation

$$\theta_d(t) = 1 - \exp(-t/T) \quad (2)$$

where $\theta_d(t)$ is the exchanged fraction of solvent and the T value is 12.4 ± 0.26 min. Unlike eq 1 which reflects the exponential decay of amide protons, eq 2 results from the exponential growth character of amide group deuteration measured through the 1208 cm^{-1} band of heavy water. Minimization of χ^2 with the monoexponential eq 2 shows that these NLPs are fully permeable to water molecules of the aqueous solvent. In fact, another more slowly exchanging fraction of solvent molecules cannot be detected experimentally. If the penetration of solvent through the NLPs was retarded by the protein shell, another more slowly exchanging class of solvent molecules would be expected and the fitting of the exchange data would require at least two exponential terms in eq 2, one for water molecules outside these particles and another for the interior water. On the other hand, the ribose−phosphate backbone exchanged as rapidly as the solvent. Thus, the rate of intensity decay at 967 cm^{-1} could be fitted to a monoexponential expression (eq 1) with a period T of 12.9 ± 0.4 min. However, deuteration of some nucleobases such as guanine occurs more slowly, as described later in relation to two-dimensional correlation spectroscopy.

DISCUSSION

In this study, we have looked into what structural effect the formation of NLPs may have on the HCVc-120 protein and viral RNA. In addition, we have studied the protein−nucleic acid interactions involved in these NLPs.

Negative-stain electron microscopy analysis of the in vitro assembly of HCVc-120 protein and the 31mer RNA (loop IIId of the viral genome) revealed that rounded particles ranging from 30 to 70 nm in diameter were produced. This is consistent with previous investigations in which NLP formation depended on highly basic N-terminal amino acids 120−124 of the core protein and on the secondary structures within the RNA molecule that was co-incubated with this protein (*3*, *5*). Although it is well-known that the HCVc-120 protein, at low concentrations (< 5 mg/mL), shares the characteristics of native unfolded proteins (*38*), no structural details have been reported in earlier works for this viral protein when bound to the 31mer RNA in the NLPs. From our infrared spectroscopic results, it is apparent that there is β-sheet formation upon assembly of these NLPs. Curve fitting of the amide I band in the spectrum of NLPs after subtraction of the 31mer RNA reveals the presence of two β-sheet band components at 1635 and 1622 cm^{-1} which can be interpreted in terms of intramolecular (12%) and intermolecular (10%) β-sheet formation, respectively, at the expense of unordered structure (*11*, *21*−*24*). The behavior of this protein resembles that of the nucleocapsid proteins of a number of viruses that are reasonably flexible in solution and are subsequently switched to the conformation required for their particular position in the nucleocapsid during assembly. Among the best characterized examples of such a conformational flexibility are the SV40 pentamer (*39*), the coat protein of tomato bushy stunt virus (*40*), the nucleocapsid protein of cauliflower mosaic virus (*41*), and the coat protein of bacteriophage P22 (*42*). The β-sheet formation in the NLPs studied here may suggest nucleocapsid stabilization through three-way junctions of β-sheet-forming protein segments (*43*, *44*).

The antisymmetric PO_2^- stretching mode in the 1250−1210 cm^{-1} region is a sensitive marker band for conformational changes in the ribose−phosphate backbone. It acts as a reporter independent of the bases, allowing one to view only the changes in the ribose−phosphate backbone. The position of this band generated by the 31mer RNA either in the pure state or within the NLPs shows that this single-stranded RNA comprises the A-type form having C3′-endo sugar puckering, and the same can be said for the ribose−phosphate backbone vibration appearing near 814 cm^{-1}. However, despite these similarities, the infrared spectra of packaged and unpackaged RNA are not identical. Thus, a significant difference is reflected in the intensity at 814 cm^{-1} which is lower in the spectrum of NLPs, what can be interpreted as evidence of partial disordering of the 31mer RNA with packaging. This is consistent with the infrared hyperchromism noticed for the 1580 cm^{-1} guanine band on the spectrum of NLPs. With regard to intermolecular protein−nucleic acid contacts in this particle, the downshift in frequency of the $\nu_s PO_2^-$ band upon nucleocapsid formation reflects interactions between RNA phosphate groups and positively charged amino acid residues such as arginine and lysine. Although these protein−nucleic

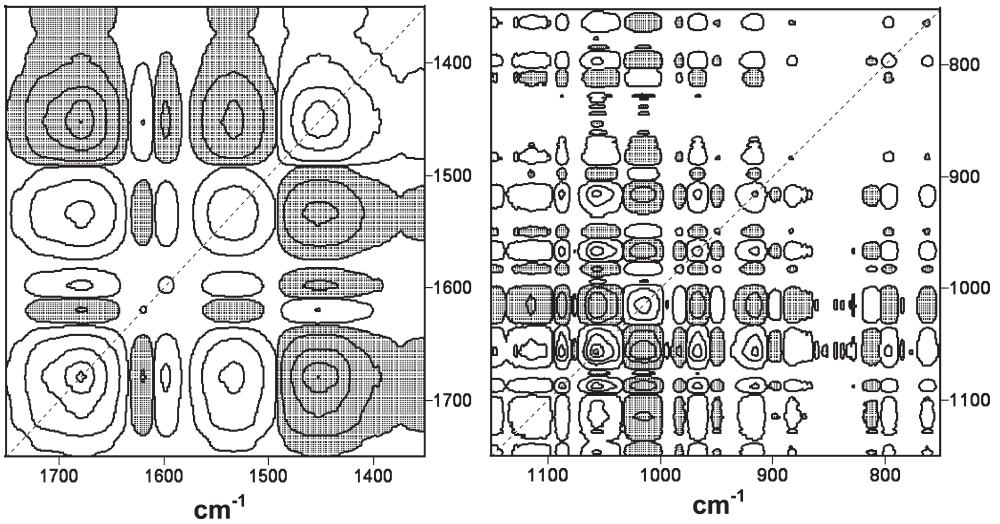

FIGURE 6: Synchronous spectra of NLPs obtained from the 0−10 min time interval of H-D exchange: (left) 1750−1350 cm^{-1} region and (right) 1150−750 cm^{-1} region.

acid contacts are in general nonspecific and ionic, there may be other interactions of the nucleocapsid proteins with the packaging site of the RNA that involve specific interactions and perhaps high-affinity binding. In this regard, it is to be noted that the packaging site in all the known single-stranded RNA viruses is a stem−loop structure (*45*). This fact prompted us to explore other intermolecular interactions using two-dimensional correlation spectroscopy. The synchronous spectra of the 1750−1350 and 1150−750 cm^{-1} regions for the 0−10 min time interval (Figure 6) reveal that the bands near 1598 and 1620 cm^{-1} are positively and negatively correlated (white and dark cross-peaks), respectively, with the band near 1678 cm^{-1}, which can be attributed to deuteration of adenine and cytosine bases during this time interval. At the same time interval, ribose−phosphate backbone deuteration is clearly visible through the diagonal peaks at 1087, 1056, 967, and 915 cm^{-1}. Interestingly, during the 10−90 min time interval (Figure 7), strong positive cross-peaks correlate two bands assigned to guanine (1580 cm^{-1}) and arginine (1603 cm^{-1}). This suggests the presence of guanine−arginine side chain contacts for two reasons: (1) because arginine residues are expected to be visible here because they constitute 18.3% of the amino acid sequence of the HCVc-120 protein (*8*) and (2) because of the kinetically favored formation of NLPs having HCV core protein and guanine-enriched synthetic oligonucleotides. Thus, Table 1 shows the ultraviolet absorbance measured at 350 nm during in vitro assembly of NLPs consistent of HCVc-120 core protein and RNA oligonucleotides, according to the reported in vitro assay in which the progress of nucleocapsid assembly is monitored by ultraviolet spectroscopy (*9*). These results and the electron microscopy analyses of these NLPs show that single-stranded oligonucleotides [5′-$(G)_{25}$-3′, 5′-$(C)_{25}$-3′, 5′-$(A)_{25}$-3′, and 5′-$(U)_{25}$-3′] can be substrates for formation of NLPs, but double-stranded or structured oligonucleotides [5′-$(C)_{15}$-$(G)_{10}$-3′ and 5′-$(A)_{15}$-$(U)_{10}$-3′] seem to be more efficient in triggering the assembly process. Moreover, 5′-$(C)_{15}$-$(G)_{10}$-3′ is one of the most efficient oligonucleotides, what can be attributed to the presence of interactions between arginine side chains and the guanine O(6) and N(7) atoms because arginine residues are frequently found to make hydrogen bonds to these guanine positions (*46*, *47*). Interactions involving guanine base have also been detected by infrared spectroscopy in other protein−nucleic acid biomolecular systems (*48*, *49*). In this regard, a question

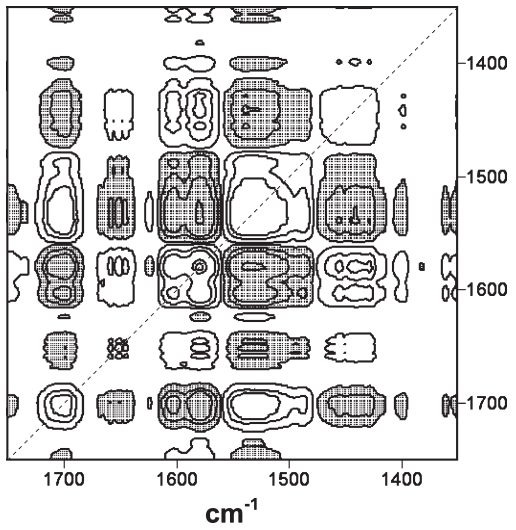

FIGURE 7: Synchronous spectra of NLPs obtained from the 10−90 min time interval of H−D exchange.

about the low efficiency of the 5′-$(G)_{25}$-3′ sequence studied here arises in spite of the fact that guanine is the only base component in this RNA oligonucleotide. This can be explained by considering that this oligonucleotide forms a quadruplex structure (*50*) at room temperature and pH ∼7, whereby the O(6) and N(7) guanine positions are prevented from contacting hydrogen bond donors such as the arginine side chain. On the other hand, HCV loop IIId, whose guanine content is lower than that of the 5′-$(C)_{15}$-$(G)_{10}$-3′ sequence, is the most efficient. From these observations, it appears that in HCV loop IIId there can be in addition to guanine bases some secondary or tertiary structure(s) contributing to the physically efficient interaction with HCVc-120. However, no significant time-dependent absorbance increase was observed in the presence of 200 mM NaCl (Table 1), which reveals the important role played by electrostatic forces in the assembly of NLPs. As proposed in other work (*51*), electrostatic attractions between RNA phosphate groups and protein positively charged side chains provide a driving force for viral assembly. This correlates well with the frequency shifting of the $\nu_s PO_2^-$ mode of RNA phosphate groups upon formation of these NLPs mentioned above. On the other hand, NLP formation is triggered by the

Table 1: Influence of RNA Oligonucleotides on in Vitro Assembly of NLPs Consisting of HCVc-120 Core Protein and RNA at a 2:1 Protein:RNA Mole Ratio

	absorbance (350 nm)	
	5 min	10 min
$5'$-$(C)_{15}$-$(G)_{10}$-$3'$	0.441	0.508
$5'$-$(A)_{15}$-$(U)_{10}$-$3'$	0.387	0.425
$5'$-$(G)_{25}$-$3'$	0.336	0.390
$5'$-$(C)_{25}$-$3'$	0.216	0.272
$5'$-$(A)_{25}$-$3'$	0.362	0.403
$5'$-$(U)_{25}$-$3'$	0.335	0.391
HCV loop IIId	0.468	0.544
HCV loop IIId with 200 mM NaCl	0.002	0.003
without RNA	0.000	0.000

presence of a nucleic acid, because when this is absent no significant increase in absorbance is detected during the absorbance measurement time period (Table 1).

To summarize, our findings show that electrostatic and guanine−arginine side chain interactions are involved in NLP formation in vitro. The HCVc-120 protein not only has a preference for binding guanine-enriched oligoribonucleotides over other oligoribonucleotides but also has a specific interaction with HCV loop IIId. This protein-free oligonucleotide contains ribosyl phosphate groups with conventional A-type backbone geometry, indicative of hairpin formation and base pairing. As a component of the NLPs, the interactions could restrict base stacking of HCV loop IIId and disfavor base pairing within these particles. When compared with the HCVc-120 monomer which is completely unordered, the protein secondary structure of the NLPs is enriched with β-sheets. This structural feature may act as a linker of protein monomers in the formation of these particles, which occurs in other viruses that have been studied (*52*, *53*). Finally, our H−D exchange results show that, despite the absence of visible holes as revealed by electron microscopy, these NLPs are fully penetrated by water molecules, the aqueous penetrability of these particles occurring with H−D exchange in the protein component. This permeability result is consistent with other studies of the permeability of P22 viral capsids to aqueous solvent and permeability of bacteriophage T4 capsids to DNA-specific and protein-specific probes (*54*, *55*).

REFERENCES

1. McLauchlan, J. (2000) Properties of the hepatitis C virus core protein: A structural protein that modulates cellular processes. *J. Viral Hepatol. 7*, 2–14.
2. Baumert, T. F., Ito, S., Wong, D. T., and Liang, T. J. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J. Virol. 72*, 3827–3836.
3. Lorenzo, L. J., Dueñas-Carrera, S., Falcon, V., Acosta-Rivero, N., González, E., de la Rosa, M. C., Menéndez, I., and Morales, J. (2001) Assembly of truncated HCV core antigen into virus-like particles in *Escherichia coli*. *Biochem. Biophys. Res. Commun. 281*, 962–865.
4. Ezelle, H. J., Markovic, D., and Barber, G. N. (2002) Generation of hepatitis C virus-like particles by use of a recombinant vesicular stomatitis virus vector. *J. Virol. 76*, 12325–12334.
5. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S. M., and Watowich, S. J. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *J. Virol. 75*, 2119–2129.
6. Klein, K. C., Dellos, S. R., and Lingappa, J. R. (2005) Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system. *J. Virol. 79*, 6814–6826.
7. Palucha, A., Loniewska, A., Satheshkumar, S., Boguszewska-Chachulska, A. M., Umashankar, M., Milner, M., Haenni, A. L., and Savithri, H. S. (2005) Virus-like particles: Models for assembly studies and foreign epitope carriers. *Prog. Nucleic Acids Res. Mol. Biol. 80*, 135–168.
8. Majeau, N., Gagné, V., Boivin, A., Bolduc, M., Majeau, J. A., Ouellet, D., and Leclerc, D. (2004) The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. *J. Gen. Virol. 85*, 971–981.
9. Fromentin, R., Majeau, N., Laliberté-Gagné, M. E., Boivin, A., Duvignaud, J. B., and Leclerc, D. (2007) A method for in vitro assembly of hepatitis C virus core protein and for screening of inhibitors. *Anal. Biochem. 366*, 37–45.
10. Nabet, A., and Pezólet, M. (1997) Tow-dimensional FT-IR spectroscopy: A powerful method to study the secondary structure of proteins using H-D exchange. *Appl. Spectrosc. 51*, 466–469.
11. Pastrana-Ríos, B., Ocana, W., Ríos, M., Vargas, G. L., Ysa, G., Poynter, G., Tapia, J., and Salisbury, J. L. (2002) Centrin: Its secondary structure in the presence and absence of cations. *Biochemistry 41*, 6911–6919.
12. Noda, I., and Ozaki, Y. (2004) Principle of two-dimensional correlation spectroscopy. In Two-dimensional correlation spectroscopy. Applications in vibrational and optical spectroscopy, p 15, Wiley, Chichester, England.
13. Raussens, V., Ruysschaert, J. M., and Goormaghtigh, E. (2004) Analysis of $^1H/^2H$ exchange kinetics using model infrared spectra. *Appl. Spectrosc. 58*, 68–82.
14. Vigano, C., Smeyers, M., Raussens, V., Schierlinckx, R., Ruysschaert, J. M., and Goormaghtigh, E. (2004) Hydrogen-deuterium exchange in membrane proteins monitored by IR spectroscopy. A new tool to resolve protein structure and dynamics. *Biopolymers 74*, 19–26.
15. Carmona, P., Rodríguez-Casado, A., and Molina, M. (2009) Improving real time measurement of H/D exchange using a FTIR biospectroscopic probe. *Anal. Bioanal. Chem. 393*, 1289–1295.
16. Bukh, J., Purcell, R. H., and Miller, H. (1992) Sequence analysis of the $5'$ non-coding region of hepatitis C. *Proc. Natl. Acad. Sci. U.S.A. 89*, 4942–4946.
17. Bukh, J., Purcell, R. H., and Miller, H. (1994) Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc. Natl. Acad. Sci. U.S.A. 91*, 8239–8243.
18. Honda, M., Rijnbrand, R., Abell, G., Kim, D., and Lemon, S. M. (1999) Natural variation in translational activities of the $5'$ nontranslated RNAs of hepatitis C virus genotype 1a and 1b. Evidence for a long-range RNA-RNA interaction outside of the internal ribosomal entry site. *J. Virol. 73*, 4941–4951.
19. Tanaka, Y., Shimoike, T., Ishii, K., Suzuki, R., Suzuki, T., Ushijima, H., Matsuura, I., and Miyamura, T. (2000) Selective binding of hepatitis C core protein to synthetic oligonucleotides corresponding to the $5'$ untranslated region of the viral genome. *Virology 270*, 229–236.
20. Banyay, M., Sarkar, M., and Gräslund, A. (2003) A library of IR bands of nucleic acids in solution. *Biophys. Chem. 104*, 477–488.
21. Byler, D. M., and Susi, H. (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers 25*, 469–487.
22. Krimm, S., and Bandekar, J. (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem. 38*, 181–363.
23. Surewicz, W. K., Mantsch, H. H., and Chapman, D. (1993) Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry 32*, 389–394.
24. Barth, A., and Zscherp, C. (2002) What vibrations tell us about proteins. *Q. Rev. Biophys. 35*, 369–430.
25. Shimanouchi, T., Tsuboi, M., and Kyogoku, Y. (1964) Infrared spectra of nucleic acids and related compounds. In The Structure and Properties of Biomolecules and Biological Systems. Advances in Chemical Physics (Duchesne, J., Ed.) pp 435−498, Interscience, London.
26. Miles, H. T., and Frazier, J. (1978) Infrared spectroscopy of polynucleotides in the carbonyl region in H_2O solution: A·U systems. *Biochemistry 17*, 2920–2927.
27. Liquier, J., Taillandier, E., Klinck, R., Guittet, E., Gouyette, C., and Huynh-Dinh, T. (1995) Spectroscopic studies of chimeric DNA-RNA and RNA 29-base intramolecular triple helices. *Nucleic Acids Res. 23*, 1722–1728.
28. Sarkar, M., Dornberger, U., Rozners, E., Fritzsche, H., Strömberg, R., and Gräslund, A. (1997) FTIR spectroscopic studies of oligonucleotides that model a triple-helical domain in self-splicing group I introns. *Biochemistry 36*, 15463–15471.
29. Taillandier, E., and Liquier, J. (1992) Infrared spectroscopy of DNA. *Methods Enzymol. 211*, 307–335.
30. Ouali, M., Letellier, R., Sim, J. S., Akhebat, A., Adnet, F., Liquier, J., and Taillandier, E. (1993) Determination of G*G·C triple-helix structure by molecular modeling and vibrational spectroscopy. *J. Am. Chem. Soc. 115*, 4264–4270.
31. Taillandier, E., Liquier, J., and Taboury, J. A. (1985) in Advances in Infrared and Raman Spectroscopy (Clark, R. H., and Hester, R. E., Eds.) Vol. *12*, pp 65−114, Wiley-Heyden, New York.

32. Barth, A. (2007) Infrared spectroscopy of proteins. *Biochim. Biophys. Acta 1767*, 1073–1101.
33. Dave, N., Lórenz-Fonfría, V. A., Leblanc, G., and Padrós, E. (2008) FTIR spectroscopy of secondary-structure reorientation of melibiose permease modulated by substrate binding. *Biophys. J. 94*, 3659–3670.
34. Chirgadze, Y. N., Brazhnikov, E. V., and Nevskaya, N. A. (1976) Intramolecular distortion of the α-helical structure of polypeptides. *J. Mol. Biol. 102*, 781–792.
35. Kim, K. S., Fuchs, J. A., and Woodward, C. K. (1993) Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry 32*, 9600–9608.
36. De Jongh, H. H., Goormaghtigh, E., and Ruysschaert, J. M. (1995) Tertiary stability of native and methionine-80 modified cytochrome c detected by proton-deuterium exchange using on-line Fourier transform infrared spectroscopy. *Biochemistry 34*, 172–179.
37. Li, J., Cheng, X., and Lee, J. C. (2002) Structure and dynamics of the modular halves of *Escherichia coli* cyclic AMP receptor protein. *Biochemistry 41*, 14771–14778.
38. Kunkel, M., and Watowich, S. J. (2004) Biophysical characterization of hepatitis C virus core protein: Implications for interactions within the virus and host. *FEBS Lett. 557*, 174–180.
39. Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L., and Harrison, S. C. (1991) Structure of simian virus 40 at 3.8 Å resolution. *Nature 354*, 278–284.
40. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., and Bricogne, G. (1978) Tomato bushy stunt virus at 2.9 Å resolution. *Nature 276*, 368–373.
41. Cheng, R. H., Olson, N. H., and Baker, T. S. (1992) Cauliflower mosaic virus, a 420 subunite ($T = 7$) multiplayer structure. *Virology 186*, 655–668.
42. Prasad, B. V. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J., and Chiu, W. (1993) Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. *J. Mol. Biol. 231*, 65–74.
43. Olson, A. J., Bricogne, G., and Harrison, S. C. (1983) Structure of tomato bushy stunt virus IV. The virus particle at 2.9 Å resolution. *J. Mol. Biol. 171*, 61–93.
44. Matsuura, K., Murasato, K., and Kimizuka, N. (2005) Artificial peptide-nanospheres self-assembled from three-way junctions of β-sheet-forming peptides. *J. Am. Chem. Soc. 127*, 10148–10149.
45. Prasad, B. V. V., and Prevelige, P. E., Jr. (2003) Viral genome organization. *Adv. Protein Chem. 64*, 219–256.
46. Allers, J., and Shamoo, Y. (2001) Structure-based analysis of protein-RNA interactions using the program ETANGLE. *J. Mol. Biol. 311*, 75–86.
47. Cheng, A. C., Chen, W. W., Fuhrmann, C. N., and Frankel, A. D. (2003) Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol. 327*, 781–796.
48. Malonga, H., Neault, J. F., Arakawa, H., and Tajmir-Riahi, H. A. (2006) DNA interaction with human serum albumin studied by affinity capillary electrophoresis and FTIR spectroscopy. *DNA Cell Biol. 25*, 63–68.
49. Malonga, H., Neault, J. F., and Tajmir-Riahi, H. A. (2006) Transfer RNA binding to human serum albumin: A model for protein-RNA interaction. *DNA Cell Biol. 25*, 393–398.
50. Petrovic, A. G., and Polavarapu, P. (2008) The quadruplex-duplex structural transition of polyriboguanylic acid. *J. Phys. Chem. B 112*, 2245–2254.
51. Van der Schoot, P., and Bruinsma, R. (2005) Electrostatics and the assembly of an RNA virus. *Phys. Rev. E 71*, 061928.
52. Alonso, L. G., García-Alai, M. M., Smal, C., Centeno, J. M., Iacono, R., Castaño, E., Gualfetti, P., and De Prat-Gay, G. (2004) The HPV16 E7 viral oncoprotein self-assembles into defined spherical oligomers. *Biochemistry 43*, 3310–3317.
53. Gaudin, Y., Sturgis, J., Doumith, M., Barge, A., Robert, B., and Ruigrok, R. W. H. (2004) Conformational flexibility and polymerisation of vesicular stomatitis virus matrix protein. *J. Mol. Biol. 274*, 816–825.
54. Reilly, K. E., and Thomas, G. J., Jr. (1994) Hydrogen exchange dynamics of the P22 virion determined by time-resolved Raman spectroscopy. *J. Mol. Biol. 241*, 68–82.
55. Griess, G. A., Khan, S. A., and Serwer, P. (1991) Variation of the permeability of bacteriphage T4. Analysis by use of a protein-specific probe for the T4 interior. *Biopolymers 31*, 11–21.