Comparison of Pf1 and Fd Gene 5 Proteins and Their Single-Stranded DNA Complexes by NMR Spectroscopy and Differential Scanning Calorimetry†

K. G. Davis,‡ S. E. Plyte,‡,§ S. R. Robertson,|| A. Cooper,|| and G. G. Kneale*,‡

Biophysics Laboratories, School of Biological Science, University of Portsmouth, Portsmouth PO1 2DT, U.K., and Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

Received August 9, 1994; Revised Manuscript Received October 24, 1994⊗

ABSTRACT: The Pf1 gene 5 protein forms a large helical nucleoprotein complex (M_r = 3.1 x 10^7) with single-stranded viral DNA, from which a 32 amino acid sequence rich in alanine, proline, and glutamine residues can be removed from the C-terminus by limited proteolysis. Sharp resonances in the ^{1}H NMR spectrum of the Pf1 nucleoprotein complex indicate that the C-teminal region of the protein subunits enjoys remarkable conformational flexibility in the complex. In contrast, the globular N-terminal domain of the protein subunits is rigidly held and does not contribute to the spectrum. The Fd gene 5 protein lacks this C-terminal flexible domain, and no distinct resonances can be observed in the ^{1}H NMR spectrum when this protein is complexed to single-stranded viral DNA. Differential scanning calorimetry shows that the thermal stability of both the Pf1 and Fd gene 5 protein is increased by 8 °C in the complex with DNA, and the transition is highly cooperative. Removal of the C-terminal domain of the Pf1 gene 5 protein subunits has no appreciable effect either on the T_m of the DNA−protein complex or on the cooperative nature of the thermal transition. It is suggested that the C-terminal domain of the Pf1 gene 5 protein acts as a dynamic clamp which kinetically stabilizes the nucleoprotein complex.

Single-stranded DNA binding proteins constitute an important class of DNA binding proteins, playing vital roles in all aspects of the control of the genetic material, including DNA replication, transcription, and translation (Kneale, 1992). This class of protein is well exemplified by the gene 5 proteins of filamentous bacteriophages Fd (Alberts *et al.*, 1972) and Pf1 (Maeda *et al.*, 1982).

The gene 5 protein is one of the key proteins required for the replication of filamentous bacteriophage (Salstrom & Pratt, 1971). It forms the principal component of the intracellular nucleoprotein complex in which ca. 1500 gene 5 proteins subunits are complexed to the single-stranded circular viral DNA forming a regular helical assembly, each protein subunit binding to four nucleotides of the viral DNA (Gray *et al.*, 1982; Kneale & Marvin, 1982). The overall appearance of the Fd and Pf1 nucleoprotein complexes is similar, but more detailed analysis reveals significant differences in structure (Gray *et al.*, 1982; Kneale *et al.*, 1982; Gray, 1989). Moreover, the Pf1 gene 5 protein−viral DNA complex is considerably more resistant to dissociation by salt than the analogous Fd complex (Kneale, 1983).

The gene 5 proteins of Pf1 and Fd both form dimers in solution and contain a high proportion of β structure, with no α helix (Alberts *et al.*, 1972; Day, 1973; Morgan *et al.*, 1989; Carpenter & Kneale, 1991). Crystallographic analysis of the Fd gene 5 protein shows that the two subunits of the dimer are held together by extensive interactions between the two polypeptide chains to form a four-stranded β-barrel (Skinner *et al.*, 1994, Folkers *et al.*, 1994). Despite clear structural and functional similarities, the sequences of the Pf1 and Fd proteins share little overall homology, other than in the region of the DNA binding wing of the protein (residues 12−26 in Fd) which is homologous to the sequence 15−30 in the Pf1 gene 5 protein (Plyte & Kneale, 1991).

The Pf1 gene 5 protein consists of 144 amino acid residues (M_r = 15 400) and is 57 residues longer than the analogous Fd gene 5 protein (M_r = 9700); however, all of the amino acids so far implicated in DNA binding are found in the N-terminal region of the Pf1 gene 5 protein sequence (Tsugita & Kneale, 1985; Plyte & Kneale, 1991), suggesting the possibility of a DNA binding domain similar in size to the gene 5 protein of Fd and that of all other filamentous bacteriophages that have been characterized to date (Kneale, 1992). This has recently been confirmed by limited proteolysis of the Pf1 nucleoprotein complex, showing that 32 residues could be cleaved from the C-terminus of the protein to leave a resistant N-terminal domain that retained the ability to bind to single-stranded DNA (Plyte & Kneale, 1993). Moreover, the resulting complex following limited proteolysis retains the overall helical structure of the native Pf1 complex.

In order to understand the role of the C-terminal domain of the Pf1 gene 5 protein, we have undertaken a comparative study of the native and proteolyzed complexes of Pf1 using ^{1}H NMR spectroscopy and differential scanning calorimetry (DSC). In addition, we have conducted comparable experiments on the Fd nucleoprotein complex. The results show that the C-terminal sequence of the Pf1 gene 5 protein enjoys remarkable flexibility in the complex with viral DNA (M_r = 3 × 10^7) and is responsible for the sharp resonances observed

†This work has been supported by research grants (to G.G.K) from the Leverhulme Trust and BBSRC. The Biological Microcalorimetry facility in Glasgow is funded by SERC (now BBSRC). S.R.R and S.E.P have been supported by SERC (BBSRC) research studentships.

* Author to whom correspondence should be addressed.

‡ University of Portsmouth.

§ Present address: Ontario Cancer Research Institute, Toronto M4X 1K9, Canada.

|| University of Glasgow.

⊗ Abstract published in *Advance ACS Abstracts,* December 1, 1994.

0006-2960/95/0434-148$09.00/0

in the ^{1}H NMR spectrum of the nucleoprotein complex. This sequence does not form part of the globular, folded structure of the protein, and its removal has no effect on the thermal stability of the nucleoprotein complex. The unusual amino acid sequence of the C-terminal domain suggests parallels with the highly flexible sequences found in a variety of multidomain proteins (Williams, 1989).

EXPERIMENTAL PROCEDURES

Preparation of the Pf1 Gene 5 Nucleoprotein Complex. Growth of *Pseudomonas aeruginosa* (strain K), infection with Pf1 phage, lysis of the bacteria, and pelleting of the gene 5 nucleoprotein complex were done according to the procedure of Plyte and Kneale (1993). The resuspended pellet was applied to a Sepharose CL-2B gel filtration column equilibrated with buffer A (10 mM Tris-HCl, 100 mM NaCl, pH 7.5). Those fractions containing complex, as identified by SDS−PAGE (13% acrylamide gels), were pooled, and the material was concentrated by pelleting in a Beckman 70.1Ti rotor at 50 000 rpm, 4 °C for 3 h. The pellet was resuspended in buffer B (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.5) and then reapplied to the phenyl-sepharose fast flow (high sub) hydrophobic interaction column equilibrated with buffer A. Fractions containing pure complex were pooled, and the material was concentrated by pelleting at 50 000 rpm as before. Finally, the pellet was resuspended in the appropriate buffer and the complex clarified at 15 000 rpm for 30 min at 4 °C. The integrity of the protein and DNA components of the purified complex were confirmed by polyacrylamide and agarose gel electrophoresis in the presence of SDS and staining with Coomassie blue and ethidium bromide (respectively).

Preparation of the Fd Gene 5 Nucleoprotein Complex. Growth of *Escherichia coli* (TG1) cells, infection with Fd bacteriophage, and harvesting were performed according to published procedures (Alberts *et al.*, 1983). Cell lysis was achieved as for lysis of the Pf1 infected *P. aeruginosa* cells, followed by similar clarification and DNA removal procedures. The complex was pelleted in a 70.1Ti rotor at 50 000 rpm, at 4 °C in a Beckman L8 ultracentrifuge, and resuspended in buffer B (see above), and the sample was further purified by gel filtration and hydrophobic interaction chromatography as for Pf1 complex.

Limited Proteolysis of Pf1 Gene 5 Nucleoprotein Complex. Purified Pf1 nucleoprotein complex was digested with papain (Papaya Latex, Sigma) at a substrate/enzyme ratio of 1000:1 (w/w) for 1 min at room temperature in buffer A containing 5 mM DTT (Plyte & Kneale, 1994). Both the enzyme and complex were incubated for 15 min at room temperature prior to mixing. The reaction was halted by adding $ZnCl_2$ to give a final concentration of 3 mM, and the partially digested complex was pelleted at 50 000 rpm as above. The UV spectrum of the partially digested complex was similar to that for undigested complex, both showing a maximum at 261 nm. Since no aromatic residues are removed by proteolysis, the extinction coefficient reported for the intact complex was used to estimate concentrations (Kneale, 1983). The M_r of the resistant protein domain was estimated at 12 000, as previously observed (Plyte & Kneale, 1993). Agarose gels stained with ethidium bromide and observed under UV illumination were used to characterize the DNA component of the proteolyzed complex. When loaded under "nondenaturing" conditions, the digested complex ran very slowly, but when loaded in the presence of SDS, a sharper, faster band comigrating with viral DNA was evident under UV illumination, thus confirming the integrity of the proteolyzed complex.

NMR Spectroscopy. For NMR experiments, the pelleted nucleoprotein complex was resuspended in 20 mM $NaDCO_3$/D_2O followed by a second cycle of ultracentrifugation and resuspension in the same buffer, in order to remove as much H_2O as possible. ^{1}H NMR spectra were recorded on a Jeol 270-MHz spectrometer at 20 °C. A total of 70000 scans were accumulated over 10 h for each spectrum.

Differential Scanning Calorimetry. DSC runs followed standard procedures (Cooper & Johnson 1994a,b) using a Microcal MC-2D differential scanning calorimeter at a nominal scan rate of 60 °C h^{-1}, with appropriate equilibration buffer in the reference cell. Both sample and reference solutions were degassed under vacuum with gentle stirring for about 1 min prior to loading. Data were analyzed to give midpoint transition temperatures (T_m) and apparent enthalpies (ΔH_{cal} and ΔH_{VH}) using the Microcal ORIGIN software package.

RESULTS

NMR Spectroscopy of Gene 5 Nucleoprotein Complexes. In the cell, multiple copies of gene 5 protein dimers bind to viral DNA to form a long helical nucleoprotein structure, with a molecular weight of 1.8×10^7 (for Fd) and 3.1×10^7 (for Pf1) (Gray *et al.*, 1982; Kneale & Marvin, 1982). NMR spectra of supramolecular complexes of this size very rarely show clear peaks, due to the extensive line broadening that arises from slow tumbling of the complex in solution. Nevertheless, even very large complexes may possess a certain degree of segmental flexibility that might allow the observation of some structure in the NMR spectrum.

Indeed, the spectrum of the Pf1 nucleoprotein complex shows a considerable number of distinct peaks in the aliphatic region of the ^{1}H NMR spectrum (0−4.5 ppm), many of which are remarkably sharp (20−50 Hz in line width) for a complex of this size, indicating a high degree of mobility for one or more regions of the protein in this complex (Figure 1A). There are no resonances in the spectrum that could be attributed to the DNA component, and thus the DNA must be held quite firmly in the complex. Furthermore, there are no observable resonances in the aromatic region of the spectrum (6−8 ppm), indicating that the aromatic amino acids of the Pf1 gene 5 protein are fully immobilized in the complex with viral DNA. This is consistent with the earlier results from time-resolved anisotropy measurements of the Pf1 nucleoprotein complex, based on the fluorescence of the single tryptophan in the protein (Trp14), which indicated a rotational correlation time of 500 ns or more (Greulich *et al.*, 1985). There is, moreover, a marked lack of resonances in the region 0−1 ppm that would arise from the methyl groups of valine, isoleucine, and leucine, with the possible exception of a small shoulder at 0.9 ppm. Thus none (or virtually none) of these amino acid residues form part of the mobile region of the Pf1 gene 5 protein in the nucleoprotein complex.

Table 1 lists the chemical shifts of the nine most prominant peaks in the ^{1}H NMR spectrum of Pf1 nucleoprotein complex, together with the most likely resonances that

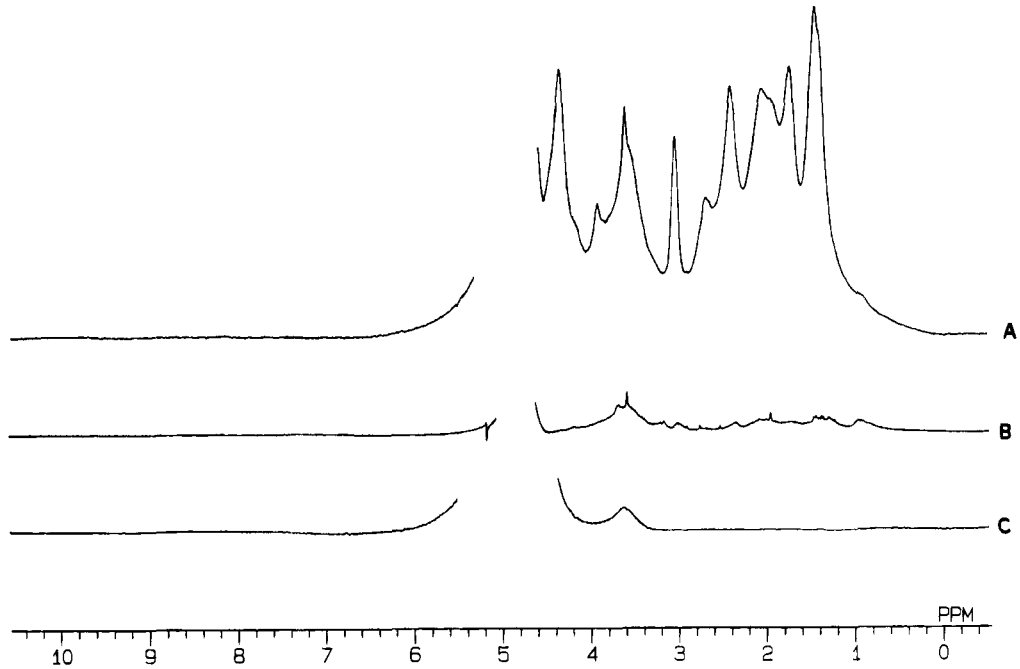

FIGURE 1: 270-MHz ^{1}H NMR spectrum of (a) intact Pf1 gene 5 protein−viral DNA complex (1.7 mg/mL) (b) Pf1 gene 5 protein−viral DNA complex (1.9 mg/mL) following limited proteolysis, and (c) Fd gene 5 protein−viral DNA complex (1.6 mg/mL). All samples were in a buffer containing 20 mM $NaDCO_3/D_2O$ at 20 °C. A total of 70 000 scans were accumulated for each sample. The chemical shifts shown are relative to the water soluble reference 2,2-dimethylsilapentane-5-sulfonic acid.

Table 1. Chemical Shift Values for the Major Peaks in the NMR Spectrum of the Pf1 Gene 5 Protein/DNA Complex

chemical shift (ppm)[a]	proton resonances[b]
1.40 (1.35)	Ala (β-CH_3), Lys (γ-CH_2)
1.70	Lys (β-CH_2), Lys (δ-CH_2), Arg (γ-CH_2)
2.00 (1.90)	Gln (β-CH_2), Arg (β-CH_2)
2.37	Gln (γ-CH_2)
2.65	Asp (β-CH_2)
3.02	Lys (ϵ-CH_2)
3.58 (3.50)	Pro (ring protons)
3.90	Ser (β-CH_2)
4.32	Ala (α-CH), Arg (α-CH), Lys (α-CH)

[a] Figures in parentheses refer to a shoulder on the main peak. [b] Proposed amino acid resonances that contribute to each of the observed peaks on the basis of their expected "random coil" positions.

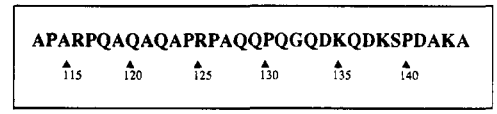

FIGURE 2: C-terminal sequence of the Pf1 gene 5 protein (residues 113−144) that is removed by limited proteolysis of the protein-DNA complex.

contribute to them. In most cases, the resonances assigned to each peak have chemical shifts within 0.05 ppm of their unperturbed ("random coil") values. The 32 amino acid C-terminal sequence of the Pf1 gene 5 protein that is susceptible to proteolysis in the nucleoprotein complex has a striking amino acid composition, being dominated by the eight alanines, eight glutamines, and five prolines that together constitute two-thirds of this sequence (Figure 2). Resonances arising from the side chains of these residues can be identified at 1.40, 2.37, and 3.58 ppm, respectively. The 32-residue C-terminal sequence also contains three lysines, three aspartates, and two arginines. The peak at 1.70 ppm is proposed to originate from the lysine (β- and δ-CH_2) protons and the arginine γ-CH_2 protons. The aspartate β-CH_2 protons are proposed to give rise to the peak at 2.65 ppm. The particularly sharp (20 Hz) peak at 3.02 ppm is very well resolved from other resonances and can be attributed to the lysine ϵ-CH_2 protons. The remaining two amino acid residues that make up the C-terminal sequence of the Pf1 gene 5 protein are Gly132 and Ser139; while the α-CH resonance of glycine cannot be clearly discerned, the β-CH_2 protons of serine could be responsible for the small peak at 3.90 ppm. The chemical shifts of the predominant peaks in the ^{1}H NMR spectrum of the Pf1 nucleoprotein complex can therefore be accounted for by the amino acid composition of the C-terminal 32 amino acid sequence of the gene 5 protein. We propose that the conformational flexibility of the Pf1 nucleoprotein complex arises exclusively from this region of the protein sequence.

The line width of the peak at 3.02 ppm (attributed to the ϵ-CH_2 protons of lysine) is approximately 20 Hz and indicates that the lysine side chains in the C-terminal sequence are extremely mobile. Many of the other peaks appear to be somewhat broader (around 50 Hz), but this may arise from chemical shift heterogeneity. For example, the peak at 1.4−1.5 ppm appears to have two components differing only slightly in chemical shift; they most probably arise from alanine (β-CH_3) and lysine (γ-CH_2) protons, which differ by only 0.07 ppm (~19 Hz) at their "random coil" positions.

NMR Spectroscopy of the Proteolyzed Pf1 Nucleoprotein Complex. In view of the susceptibility of the C-terminal sequence of the Pf1 gene 5 protein to proteolytic cleavage (Plyte & Kneale, 1993), the nucleoprotein complex was trimmed by limited proteolysis and purified for NMR studies. The resulting complex, in which 32 residues are removed from the C-terminus of the protein subunits, shows a strikingly different NMR spectrum from that of the intact complex (Figure 1B). Virtually all the strong resonances disappear, leaving a rather featureless spectrum.

We note, however, that the small peak at 0.9 ppm that can be seen in in the NMR spectrum of the proteolyzed Pf1

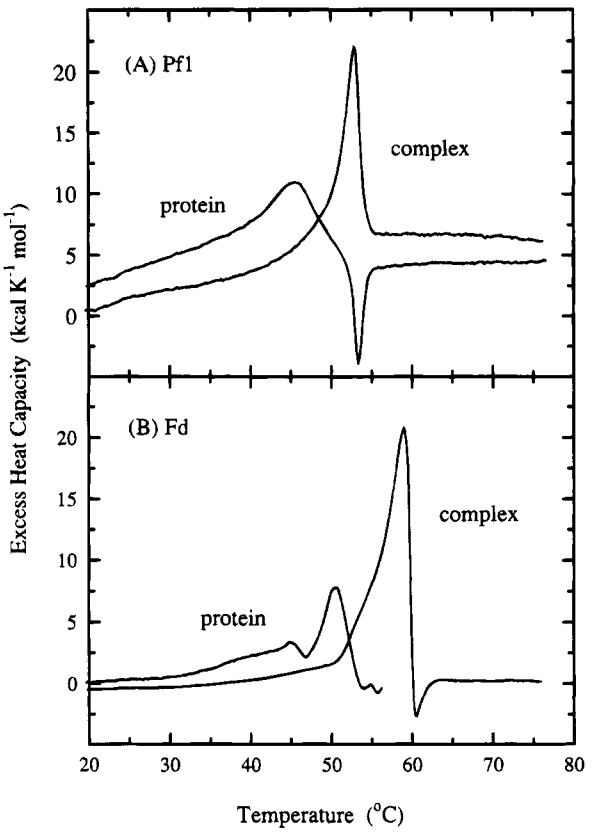

FIGURE 3: DSC thermograms (concentration normalized and corrected for buffer baselines) of (A) Pf1 gene 5 protein and (B) Fd gene 5 protein (each at 1.8 mg/mL) and their nucleoprotein complexes (at 1.8 and 2.0 mg/mL, respectively). Samples were measured in 10 mM Tris-HCl (pH 7.5) with a scan rate of 60 °C/h. Exothermic (negative) responses following the main exothermic transitions are due to irreversible protein aggregation, and no transitions were seen on re-scan of these samples.

complex is also present (as a small shoulder) in the spectrum of the intact complex. Since this peak remains after removal of the C-terminal sequence, it must represent a residue that has some mobility in the core sequence (residues 1−112) of the protein, most probably arising from the methyl group of a valine, leucine, or isoleucine side chain. The experiments on both the intact Pf1 complex and the proteolytically cleaved complex were carried out a number of times, using different preparations of the sample, with essentially the same result. This confirms that the strong NMR signal apparent in the spectrum of the Pf1 nucleoprotein complex arises exclusively from the C-terminal sequence of the protein that is removed by limited proteolysis.

NMR Spectroscopy of the Fd Nucleoprotein Complex. For comparison, ^{1}H NMR spectra were recorded for the Fd nucleoprotein complex. In sharp contrast to the results obtained above, the spectrum of the Fd complex shows only a broad featureless baseline, with the exception of the large peak centered at 4.9 ppm arising from residual H_2O in the sample and the broad peak at 3.6 ppm that arises from water in the vapor phase from within the NMR tube (Figure 1C). Thus there are no regions of the Fd gene 5 protein that show any significant mobility in the complex with viral DNA.

Thermal Stability of Pf1 and Fd Gene 5 Proteins and Their Complexes with Viral DNA. Differential scanning calorimetry (DSC) is a technique that allows detection of the small changes in heat capacity that occur during thermal transitions of biological macromolecules (Cooper & Johnson, 1994a,b; Sturtevant, 1987; Privalov & Potekhin, 1986). In the case of simple, highly soluble, globular proteins, this may lead to detailed thermodynamic data for the unfolding process, for example. In more complex, less reversible systems, one is generally limited to more qualitative comparisons of the thermal stability of related proteins and the investigation of protein domain structure (Privalov, 1982; Johnson *et al.*, 1991; Galisteo *et al.*, 1991). In the case of DNA binding proteins one can also investigate the effects of DNA binding on the stability of the nucleoprotein complex. For this reason we have performed DSC analysis on the Pf1 and Fd gene 5 proteins and their large supramolecular complexes that form with single-stranded viral DNA. In addition, we have investigated the thermal denaturation of the product of limited proteolysis of the Pf1 gene 5 protein−DNA complex.

DSC data for the thermal unfolding of the gene 5 proteins of Pf1 and Fd bacteriophage and their complexes with single-stranded DNA are illustrated in Figure 3. The DSC thermogram comparing the stability of the Pf1 gene 5 protein and its complex with single-stranded DNA is shown in Figure 3A. The free protein undergoes a relatively broad endothermic process, typical of a protein unfolding transition with T_m ~45.4 °C, followed by exothermic processes at higher temperatures characteristic of irreversible aggregation of the unfolded protein. (Re-scan of these samples, and all samples investigated here, after cooling to room temperature in the DSC showed no subsequent transitions.) In comparison, the thermogram of the complex of this protein with viral DNA shows a much sharper transition at somewhat higher temperature (T_m ~53 °C) suggestive of a highly cooperative transition of the stabilised nucleoprotein complex. (Although the shape of these thermograms could possibly be distorted by exothermic events at higher temperatures, these processes do not significantly affect the data analysis provided they occur more slowly and at higher temperatures than the major melting transitions, as they appear to do here in most cases.) The data are consistent with a cooperative thermal unfolding of the entire complex and show no evidence for dissociation of the complex prior to unfolding. The curves were analyzed using the non-2-state algorithm (Microcal ORIGIN) in which deconvolution of the DSC thermogram gives independent estimates of the calorimetric (ΔH_{cal}) and van't Hoff enthalpies (ΔH_{VH}) of the thermal transition. For the simultaneous, cooperative unfolding of oligomeric systems,

$$N_n \rightleftharpoons nU \quad (N = \text{native monomer, } U = \text{unfolded})$$

The ratio $\Delta H_{VH}/\Delta H_{cal}$ gives an estimate of the the size of the cooperative unit, n (Sturtevant, 1987; Cooper & Johnson, 1994a). The estimated ΔH_{cal} for this transition is about +45 kcal mol^{-1}, compared to a much larger apparent van't Hoff enthalpy ΔH_{VH} of about +270 kcal mol^{-1}, suggesting a cooperative unit of about six protein monomers when bound to single-stranded viral DNA under these concentration conditions.

Qualitatively similar results are obtained for the Fd gene 5 protein (Figure 3B), though in this case the protein unfolds at a significantly higher T_m (51 °C) than the Pf1 protein under the same conditions indicating somewhat increased thermal stability for this protein. The Fd nucleoprotein complex is, again, more stable (T_m ~59 °C) and shows similar sharpening of the thermogram compatible with a higher cooperativity and more extensive protein−protein interactions in the

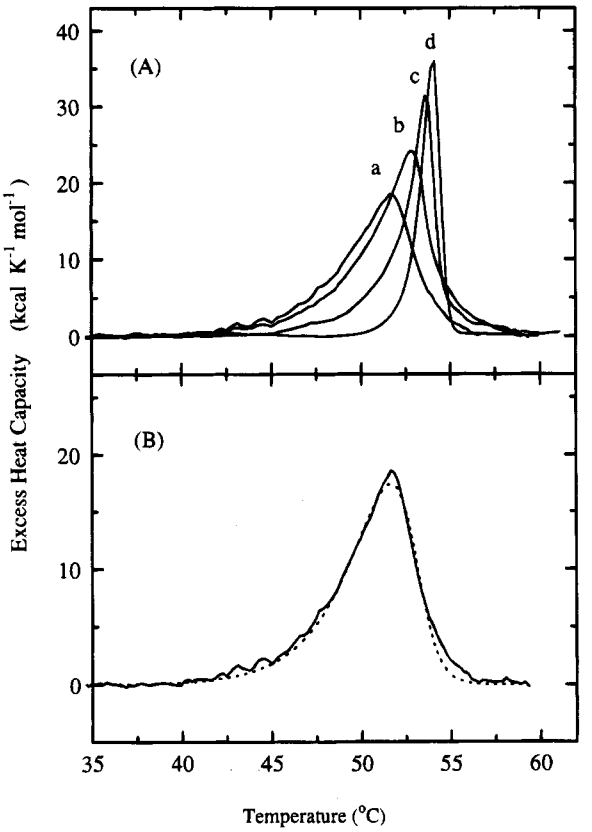

FIGURE 4: (A) Normalized DSC thermograms of Pf1 gene 5 protein–viral DNA complex following limited proteolysis and purification of the complex. Measurements were made under the same buffer conditions as in Figure 3 at protein monomer concentrations of (a) 0.0189, (b) 0.0302, (c) 0.061, and (d) 0.116 mM. (B) Deconvolution of data (a), 0.0189 mM, showing the comparison between experimental data (solid line) and theoretical fit (dotted line) for a model in which four protein monomers (in this case) dissociate and unfold cooperatively from the DNA. The size of this apparent cooperative unit increases at higher concentrations.

supramolecular complex. Despite uncertainties that can arise due to the irreversible nature of all these thermal transitions, it is clear that DNA binding brings about a similar enhancement in thermal stability ($\Delta T_m \sim 8$ °C) and cooperativity for both proteins.

Removal of the C-terminal tail from the Pf1 gene5 protein has only a minimal effect on the DSC of the nucleoprotein complex (Figure 4). The thermal stability of the proteolyzed Pf1 gene 5 protein is very similar to that of the native protein under similar conditions, suggesting that the C-terminal sequence (residues 113–144) does not contribute to the stability of the folded N-terminal domain. However, more detailed experiments on the proteolyzed complex show significant concentration dependence on both the T_m and the shape of the thermograms over a roughly 6-fold concentration range (Figure 4A). This is consistent with the reversible association of protein dimers with the single stranded DNA to form increasingly larger cooperative units as the concentration is increased, similar to DSC observations in other associating systems (Manly *et al.*, 1985; Fukada *et al.*, 1983). In such cases (Figure 4B), DSC data were analyzed in terms of a more complex dissociation–equilibrium model:

$$n\mathrm{N} \rightleftharpoons \mathrm{N}_n \rightleftharpoons n\mathrm{U}$$

in which subunit dissociation occurs simultaneously with thermal unfolding, resulting in the asymmetry observed in the thermograms. Explicit incorporation of the monomer–oligomer equilibrium into this model means that both the shape and T_m of the DSC thermogram will depend on total protein concentration, as observed (Figure 4).

Estimates of the number of cooperative units range from about four protein monomers at low concentrations to 10–20 (or more) at higher protein–DNA concentrations. These figures are also compatible with estimated $\Delta H_{VH}/\Delta H_{cal}$ ratios, which approach 10 or more at higher sample concentrations. Such models are necessarily an oversimplification of the true situation, where protein associations and interactions are occurring on the DNA strands. However, current data, together with the complications arising from irreversibility of the unfolding transitions, do not yet justify more complex analysis.

DISCUSSION

Conformational Flexibility in the C-Terminal Sequence of the Pf1 Gene 5 Protein. The observation of numerous peaks in the NMR spectrum of the Pf1 nucleoprotein complex indicates that certain regions of the protein are remarkably mobile for a macromolecular assembly of this size (3.1×10^7). Following limited proteolysis of the Pf1 nucleoprotein complex, all the major peaks disappear, thus showing that the mobile residues in the native nucleoprotein complex are confined to the 32 amino acid C-terminal sequence of the protein. The lack of chemical shift perturbations in the resonances arising from this sequence are a further indication that it does not form part of the globular folded structure that is characteristic of the N-terminal domain. It is notable that the Fd nucleoprotein complex, which has a helical structure similar to that of Pf1 (Gray *et al.*, 1982), shows no sharp resonances in its NMR spectrum and therefore lacks these mobile regions. High-resolution ^{1}H NMR studies of the Fd gene 5 protein dimer show that the DNA binding wing of the protein (residues 12–26) is more mobile than the remainder of the protein (Folkers *et al.*, 1994). This region is also more disordered in the crystal structure of the protein (Skinner *et al.*, 1994). Nevertheless, it is clear that the mobility of the DNA binding wing in the Fd gene 5 protein does not persist to any measurable extent in the complex with viral DNA.

The C-terminal amino acid sequence of the Pf1 gene 5 protein is dominated by alanine, glutamine, and proline residues, which together account for two-thirds of the sequence. Indeed, residues 113–131 of the Pf1 gene 5 protein sequence consists entirely of these residues interspersed with just two arginines. The sequence at the far C-terminus of the protein (residues 132–144) is somewhat different and contains three lysine residues, three glutamatic acid residues, two prolines, and a single serine and glycine. With the possible exception of glycine, resonances from all these amino acid residues can be observed as relatively sharp lines in the NMR spectrum of the nucleoprotein complex at, or close to, their expected "random coil" values.

Flexible segments of proteins have been observed by NMR in a number of different proteins and supramolecular complexes, although few if any have molecular weights approaching that of the Pf1 nucleoprotein complex (3.1×10^7). Proteins containing such mobile segments include membrane proteins, multienzyme complexes, and nucleopro-

tein complexes; in each case the sequences concerned share similarities in their amino acid composition, being composed predominantly of alanine and proline (and in some cases glutamine), together with variable proportions of acidic and basic amino acid residues.

The pyruvate dehydrogenase (PDH) multienzyme complex of *E. coli* provides an interesting comparison. The ^{1}H NMR spectrum of this complex also shows remarkably sharp peaks, which arise from alanine residues in alanine/proline rich sequences in the linker region between N- and C-terminal domains of the E2p subunit of the enzyme complex (Perham & Roberts, 1981). Likewise, these sharp peaks disappear from the NMR spectrum following limited proteolysis. However, it is notable that there are substantially more of these sharp peaks in the NMR spectrum of the Pf1 nucleoprotein complex, in spite of this complex being almost an order of magnitude larger in molecular mass than the PDH complex. A closer parallel is provided by the related PDH enzyme from *Bacillus stearothermophilus,* since its size is more comparable to that of the Pf1 nucleoprotein complex and its NMR spectrum shows a similar spread of relatively sharp resonances (Duckworth et al., 1982; Packman et al., 1984).

Conformationally flexible sequences have also been reported in a number of RNA viruses. One of the first viruses to be investigated by NMR spectroscopy was tobacco mosaic virus (Jardetzky *et al.*, 1978), where it was shown that the sharp peaks in the ^{1}H NMR spectrum of the capsid protein, which arise from the flexible N-terminus of the protein, also persist in the much larger disc complex. These lines virtually disappear, however, when the protein subunits assemble to form the helical virus (M_r ~4 × 10^6). More recently, it has been shown that a 25-residue sequence in the N-terminal RNA binding domain of CCMV virus enjoys substantial conformational mobility in the free protein, but again this is lost when the intact virus is formed (Van der Graaf *et al.*, 1991). In contrast, the mobility of the C-terminal region of the Pf1 protein persists in the intact nucleoprotein complex. Many of the histones of eukaryotic chromatin have flexible tails which are susceptible to proteolysis; much of this flexibility persists when bound to DNA in the form of nucleosomes (Cary *et al.*, 1981), but whether this mobility is retained when the nucleosomes are assembled into the higher order structure of chromatin has not been established.

Sequences rich in alanine and proline can be found at the N-terminus of the LC1 alkali light chains of myosin, and these also show enhanced mobility by NMR spectroscopy. Although this polypeptide sequence is flexible, it is still conformationally restrained due to the Ala-Pro peptide bonds being restricted to the "trans" conformation (Bandhari *et al.*, 1986). Similar extended structures have been reported for other proline-rich sequences, e.g., in PDH multienzyme complex (Radford *et al.*, 1987) and in the bacterial membrane protein TonB (Evans *et al.*, 1986; Brewer *et al.*, 1990). The function of these X-Pro bonds may be to maintain segments of the polypeptide in an extended conformation, linked by conformationally flexible regions, so that concerted movement of the chains occurs. It is becoming clear that although these structures enjoy a considerable degree of conformational flexibility, they are not truly "random coil".

Mechanistic Implications. What, then, is the function of the C-terminal tail of the Pf1 gene 5 protein? Recent studies on the globular domain of the Pf1 protein produced by limited proteolysis have shown that this domain can bind single-stranded DNA with a similar binding affinity to the intact protein, although the complexes formed with oligonucleotides are kinetically less stable (Plyte & Kneale, 1993). Two possible explanations were proposed: the role of the C-terminal tail of the gene 5 protein could be to clamp the DNA, once it was bound in the DNA binding cleft of the protein; or, alternatively, the tail could participate in protein-protein interactions between adjacent protein dimers. The latter possibility now seems unlikely in view of the DSC results obtained for the proteolyzed complex, which clearly show that the cooperative interactions between protein dimers that accompany single-stranded DNA binding are still maintained in the absence of the C-terminal tail.

Clamping, however, could be achieved if the tail were able loop around the DNA and interact with the same gene 5 protein dimer. The high degree of mobility that is observed in the C-terminal sequence suggests that such a clamp would be a dynamic one, as indeed it would have to be to allow the DNA to access the DNA binding cleft. The kinetic stability of the complex would be increased since the clamp would have to open simultaneously on many subunits before the DNA could be fully released. There is a cluster of charged amino acid residues (three lysine and three glutamic acid residues) found in the last 10 residues at the C-terminus of the protein, which could participate in interactions with other sites on the protein. If the C-terminal domain is involved in protein−protein interactions of the type proposed, these interactions cannot contribute to the thermal stability of the globular domain since the T_m of the main transition is unaffected by removal of this sequence.

Further studies are now in progress in our laboratory on overexpressed cloned domains of the Pf1 gene 5 protein, in an attempt to elucidate the detailed structure and molecular properties of both the globular N-terminal domain and the flexible C-terminal domain of the protein.

ACKNOWLEDGMENT

We thank Chris Turner for skilled assistance with the NMR spectroscopy.

REFERENCES

Alberts, B., Frey, L., & Delius, H. (1972) *J. Mol. Biol.* 68, 139−152.

Bhandari, D. G., Levine, B. A., Trayer, I. P., & Yeadon, M. E. (1986) *Eur. J. Biochem. 160*, 349−356.

Brewer, S., Tolley, M., Trayer, I. P., Barr, G. C., Dorman, C. J., Hannavy, K., Higgins, C. F., Evans, J. S., Levine, B. A., & Wormald, M. R. (1990) *J. Mol. Biol. 216*, 883−895.

Carpenter, M. L., & Kneale, G. G. (1991) *J. Mol. Biol. 217*, 681−689

Cary P. D., Crane-Robinson, C., Bradbury, E. M., & Dixon, G. H. (1981) *Eur. J. Biochem. 119*, 545−551.

Cooper, A., & Johnson, C. M. (1994a) Introduction to Microcalorimetry and Biomolecular Energetics, *Methods in Molecular Biology, Vol. 22: Microscopy, Optical Spectroscopy, and Macroscopic Techniques* (Jones, C., Mulloy, B., & Thomas, A. H., Eds.) Chapter 9, pp 109−124, Humana Press, Totowa, NJ.

Cooper, A., & Johnson, C. M. (1994b) Differential Scanning Calorimetry, *Methods in Molecular Biology, Vol. 22: Microscopy, Optical Spectroscopy, and Macroscopic Techniques*

(Jones, C., Mulloy, B., & Thomas, A. H., Eds.) Chapter 10, pp 125−136, Humana Press, Totowa, NJ.
Day, L. A. (1973) *Biochemistry 12*, 5329−5339.
Duckworth, H. W., Jaenicke, R., Perham, R. N., Wilkie, A. O. M., Finch, J. T., & Roberts, G. C. K. (1982) *Eur. J. Biochem. 124*, 63−69.
Evans, J. S., Levine, B. A., Trayer, I. P., Dorman, C. J., & Higgins, C. F. (1986) *FEBS Lett. 208*, 211−216.
Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konings, R. N. H., & Hilbers, C. W. (1994) *J. Mol. Biol. 236*, 229−246.
Fukada, H., Sturtevant, J. M., & Quiocho, F. A. (1983) *J. Biol. Chem. 258*, 13193−13198.
Galisteo, M. L., Mateo, P. L., & Sanchez-Ruiz, J. M. (1991) *Biochemistry 30*, 2061−2066.
Gray, C. W. (1989) *J. Mol. Biol. 208*, 57−64.
Gray, C. W., Kneale, G. G., Leonard, K. R, Siegrist, H., & Marvin, D. A. (1982) *Virology 116*, 40−52.
Greulich, O., Wijnaendts van Resandt, R., & Kneale, G. G. (1985) *Eur. Biophys. J. 11*, 195−201.
Jardetzky, O., Akasaka, K., Vogel, D., Morris, S., & Holmes, K. C. (1978) *Nature 273*, 564−566.
Johnson, C. M., Cooper, A., & Brown, A. J. P. (1991) *Eur. J. Biochem. 202*, 1157−1164.
Kneale, G. G. (1983) *Biochim. Biophys. Acta 739*, 216−224.
Kneale, G. G. (1992) *Curr. Opin. Struct. Biol. 2*, 124−130.
Kneale, G. G., & Marvin, D. A. (1982) *Virology 116*, 53−60.
Kneale, G. G., & Wijnaendts, R. (1985) *Eur. J. Biochem. 149*, 85−93.
Kneale, G. G., Freeman, R., & Marvin, D. A. (1982) *J. Mol. Biol. 156*, 279−286.
Maeda, K., Kneale, G. G., Tsugita, A., Short, N. J., Perham, R. N., Hill, D. F., & Peterson, G. B. (1982) *EMBO J. 1*, 255−261.
Manly, S. P., Matthews, K. S., & Sturtevant, J. M. (1985)*Biochemistry 24*, 3842−3846.
Morgan, P. J., Harding, S. E., Plyte, S. E., & Kneale, G. G. (1989) *Biochem. Soc. Trans. 17*, 234.
Packman, L. C., Perham, R. N., & Roberts, G. C. K. (1984) *Biochem. J. 217*, 219 −227.
Perham, R. N., & Roberts, G. C. K. (1981) *Biochem. J. 199*, 733−740.
Plyte, S. E., & Kneale, G. G. (1991) *Protein Eng. 4*, 553−560.
Plyte, S. E., & Kneale, G. G. (1993) *Biochemistry 32*, 3623−3628.
Plyte, S. E., & Kneale, G. G. (1994) Limited Proteolysis of DNA−Protein Complexes, *Methods in Molecular Biology, Vol. 30: DNA−Protein Interactions* (Kneale, G. G., Ed.) Chapter 13, pp 161−168, Humana Press, Totowa, NJ.
Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1−104.
Privalov, P. L., & Potekhin, S. A. (1986) *Methods Enzymol. 131*, 4−51.
Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S., & Guest, J. R. (1987) *Biochem. J. 247*, 641−649.
Salstrom, J., & Pratt, D. (1971) *J. Mol. Biol. 61*, 489−501.
Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konings, R. N. H., Wang, A. H.-J., & Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
Sturtevant, J. M. (1987) *Annu. Rev. Phys. Chem. 38*, 463−488.
Tsugita, A., & Kneale, G. G. (1985) *Biochem. J. 228*, 193−199.
Van der Graaf, M., Kroon, G. J. A., & Hemminga, M. A. (1991) *J. Mol. Biol. 220*, 701−709.
Williams, R. J. P. (1989) *Eur. J. Biochem. 183*, 479−497.

BI941836F