Identification and Characterization of the Domain Structure of Bacteriophage P22 Coat Protein†

Jason Lanman, Roman Tuma,‡ and Peter E. Prevelige, Jr.*

Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35205

Received July 2, 1999; Revised Manuscript Received September 9, 1999

ABSTRACT: The bacteriophage P22 serves as a model for assembly of icosahedral dsDNA viruses. The P22 procapsid, which constitutes the precursor for DNA packaging, is built from 420 copies of a single coat protein with the aid of stoichiometric amounts of scaffolding protein. Upon DNA entry, the procapsid shell expands and matures into a stable virion. It was proposed that expansion is mediated by hinge bending and domain movement. We have used limited proteolysis to map the dynamic stability of the coat protein domain structures. The coat protein monomer is susceptible to proteolytic digestion, but limited proteolysis by small quantities of elastase or chymotrypsin yielded two metastable fragments (domains). The N-terminal domain (residues 1−180) is linked to the C-terminal domain (residues 205−429) by a protease-susceptible loop (residues 180−205). The two domains remain associated after the loop cleavage. Although only a small change of secondary structure results from the loop cleavage, both tertiary interdomain contacts and subunit thermostability are diminished. The intact loop is also required for assembly of the monomeric coat protein into procapsids. Upon assembly, coat protein becomes largely protease-resistant, baring cleavage within the loop region of about half of the subunits. Loop cleavage decreases the stability of the procapsids and facilitates heat-induced shell expansion. Upon expansion, the loop becomes protease-resistant. Our data suggest the loop region becomes more ordered during assembly and maturation and thereby plays an important role in both of these stages.

Viral assembly pathways, consisting of a sequence of protein associations, are essential for the viral lifecycle and thus constitute a novel class of antiviral drug targets (*1*, *2*). The morphogenesis of a viral capsid from multiple copies of a single or a few coat proteins represents a genetically efficient way for building large assemblies from smaller subunits (*3*). Therefore, well-characterized viral systems can serve as models for understanding the assembly of other complex biological machinery and can provide the mechanistic insight required for de novo design of macromolecular complexes (*4*). However, in a fashion similar to the protein folding problem, the detailed rules governing "the assembly code" are still unknown.

The assembly pathway of bacteriophage P22 is one of the best characterized paradigms for macromolecular assembly (*5*). In particular, it serves as an assembly model for double-stranded DNA (dsDNA) viruses with a two-step pathway. In the first step, multiple copies of a single coat protein (47 kDa) and three minor proteins, which are required for infectivity, polymerize with the aid of a scaffolding protein (33 kDa), into a $T = 7$ procapsid. In the second step, the viral genome is packaged into the procapsid, and the capsid undergoes maturation (Figure 1) (*6*). During maturation, the coat protein shell undergoes expansion, leading to a 10% increase in diameter (*7*) and to stabilization of the capsid (*8*). Lattice expansion accompanying DNA packaging appears to be a common theme in many dsDNA-containing viruses including the bacteriophage (*9*) and herpesviruses (*10*). Although the secondary structure of the P22 coat protein subunit is only slightly affected by expansion (*11*−*13*), the dynamic stability of the coat protein subunit is substantially increased (*14*). It has been proposed that domain movement by hinge bending is the mechanism of maturation (*7*, *12*, *14*). Although it is possible to delineate two domains of electron density in the high-resolution cryoelectron microscopy map of the procapsid, the chemical identity of these domains remains unknown.

In the present work, we report the identification and structural characterization of the proposed domains and the connecting loop region of P22 coat protein. Additionally, we present evidence that the intact loop region plays an important role in assembly and maturation.

MATERIALS AND METHODS

Procapsid Shell Preparation. Procapsid shells were produced by coexpression of coat protein and scaffolding protein in *E. coli* (BL21 DE3) at 37 °C. Procapsids were purified from cell pellets using a protocol described in ref *15*. The scaffolding protein was extracted from the procapsids using 0.6 M GuHCl[1] (guanidine hydrochloride) in buffer B (25 mM Tris-HCl, 25 mM NaCl, and 2 mM EDTA, pH 7.6) (*15*). Coat protein shells were separated from the scaffolding

† This work was supported in part by NSF DBI-9726698 and NIH GM47980 (P.E.P.).

* To whom correspondence should be addressed at the Department of Microbiology, University of Alabama at Birmingham, BBRB 416/6, 845 19th St. S., Birmingham, AL 35294-2170. Telephone: (205) 975-5327. Fax: (205) 975-5479. E-mail: prevelig@uab.edu.

‡ Present address: Department of Biosciences and Institute of Biotechnology, University of Helsinki, Biocenter 2 PL 56, FIN-00014, Finland.

10.1021/bi9915420 CCC: $18.00
Published on Web 00/00/0000 PAGE EST: 9.4

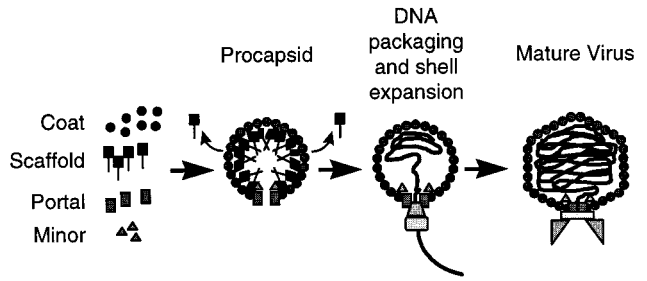

FIGURE 1: Assembly pathway of bacteriophage P22. Multiple copies of a single coat protein polymerize with the aid of a scaffolding protein to form a $T = 7$ procapsid. Three minor proteins are also required for infectivity. The viral dsDNA is packaged into the procapsid. Upon DNA packaging, the procapsid undergoes a morphological change (maturation) in which the capsid diameter increases by 10% and becomes more polyhedral.

protein on a sucrose gradient (5−20%) containing 0.5 M GuHCl in buffer B. The shells were collected by centrifugation (35 000 rpm, 42.1 Beckman rotor, 4 °C, 1 h) and resuspended in buffer B.

Preparation of Monomeric Coat Protein. Shells were dissociated in 6 M GuHCl. Coat protein was prepared by refolding from 6 M GuHCl (1 mg/mL, 4 °C) by dialyzing against buffer B containing 5 mM β-mercaptoethanol. Aggregated and assembled protein was removed by centrifugation (40 000 rpm, 42.1 Beckman rotor, 4 °C, 2 h), and the supernatant was loaded onto an anion exchange column (Pharmacia HiTrap Q) and eluted with a NaCl gradient (coat protein monomer eluted at 0.2 M). Monomeric coat protein was further purified by size-exclusion chromatography in buffer B (TSK 3000, TosoHaas).

Protease Digestion and Mapping. Monomeric coat protein was digested using elastase at 4 °C with an enzyme-to-substrate ratio of 1:1000. Aliquots were quenched at different time intervals (5, 10, 20, 30, 60, and 120 min) with PMSF, and analyzed by 13% SDS−PAGE.

Fragments resulting from a 4 h digestion were separated on a 13% SDS−PAGE and then transferred to a PVDF membrane (BIO-RAD Sequi-Blot) in 10 mM CAPS [3-(cyclohexylamino)-1-propanesulfonic acid] buffer, pH 11. The N-terminal sequences of the transferred bands were obtained by automated Edman degradation. Another aliquot of the digestion product was analyzed using MALDI-TOF (matrix-assisted laser desorption ionization-time-of-flight) mass spectrometry.

Assembled coat protein (procapsid shells and expanded shells) were subjected to similar protease digestion, details of which are provided within the figure legends.

Preparation of Cleaved Monomer. The cleaved monomer was prepared by proteolysis with elastase at 4 °C for 8 h using an enzyme-to-substrate ratio of 1:100. SDS−PAGE analysis demonstrated that this deliberate high enzyme-to-substrate ratio resulted in the nearly complete cleavage of the monomer. At this high enzyme concentration, the recovery of fragments was not quantitative, presumably because some fraction of the fragments was completely digested. The reaction was quenched with 1 mM PMSF and Pefa-Bloc (Boehringer Mannheim), and the resulting associated N- and C-terminal fragments (N/C complex, see Results) were separated from smaller proteolytic fragments by size-exclusion chromatography as described above. The cleaved monomer was kept on ice and used within a few hours of preparation.

Spectroscopy and Thermal Denaturation. The secondary structures of the N/C complex and the full-length coat protein monomer were compared using circular dichroism (CD). Spectra for these proteins were recorded using an AVIV 62DS spectrometer in the spectral range 190−250 nm at 5 °C in a 1 mm path length cell. Multiple scans were averaged and normalized (*16*). The coat protein concentration was determined spectrophotometrically using an extinction coefficient of 4.45×10^4 L mol^{-1} cm^{-1} (*1*).

The thermal denaturation of both the full-length coat protein and the N/C complex was monitored by CD. For spectra collected between 10 and 25 °C, the temperature was raised in 5 °C increments. For the spectra recorded between 28 and 70 °C, the increment was 2 °C. All temperature-resolved spectra were measured in 25 mM PO_4 and 25 mM NaCl, pH 7.6, using a stirred 1 cm path length cuvette.

Fluorescence spectroscopy was performed on dilute solutions of the intact coat protein and the associated fragments in buffer B at 25 °C using a Shimadzu 1501 fluorometer ($\lambda_{exc} = 280$ nm, 10 nm slits).

Data Analysis. Singular value decomposition (SVD) was applied to the ensemble of spectra to determine the number of base spectra required because visual inspection of the spectra indicated that unfolding of the coat protein involved more than two states. The intact coat protein spectra required three base spectra. A constrained rotation was applied because the base spectra obtained from SVD are not physically meaningful. Any spectrum (R) at temperature t can be represented by linear combination of the three basis spectra (factors), $U1$, $U2$, and $U3$, with their cognate "concentration" factors, $P1_t$, $P2_t$, and $P3_t$.

$$R_t = P1_t U1 + P2_t U2 + P3_t U3 \quad (1)$$

Similarly, the same spectrum can be expressed by linear combination of the physically meaningful spectra corresponding to the folded (F), unfolding intermediate (I), and aggregated species (A):

$$Y_t = B_t F + C_t I + D_t A \quad (2)$$

where B_t, C_t, and D_t are their respective mole fractions. The spectra of the folded (F) and aggregated species (A) were determined experimentally as averages over the plateau regions (temperature range 10−34 and 46−60 °C, respectively). These spectra together with nonnegativity of the molar fractions served as constraints in the rotation of the SVD factors, which was performed by a least-squares fitting protocol (*17*, *18*).

SVD of the N/C complex spectra revealed four base components. Spectra of the folded and aggregated conformations were established as averages over the plateau regions (10−20 and 56−70 °C). One intermediate (I_1) was stable and well represented by spectra in the region of 36−40 °C. The spectrum of the remaining intermediate (I_2) and the molar fractions were determined as described above.

[1] Abbreviations: GuHCl, guanidine hydrochloride; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; MALDI-TOF, matrix-assisted laser desorption ionization-time-of-flight; PMSF, phenylmethylsulfonyl fluoride; N/C complex, associated N- and C-terminal fragments; CD, circular dichroism; SVD, singular value decomposition.

Heat-Induced Expansion. Cleaved shells for expansion studies were prepared by digestion with elastase (1:20 enzyme-to-substrate ratio) for 1 h at 30 °C to prevent premature expansion of the cleaved shells. Heat-induced expansion of procapsid and cleaved procapsid shells was performed at 35, 40, 45, and 50 °C, and the products were separated on a 1.2% agarose gel as previously described (*14*). To determine the degree of digestion in the expanded and unexpanded shells, the agarose gel bands were excised and extracted using a kit for DNA extraction (Amicon) and analyzed on 10−20% Tris−Tricine SDS−PAGE (BioRad Laboratories).

RESULTS

Domain Structure of the Coat Protein

Coat Protein Monomer Consists of Two Domains. The domains of soluble proteins are usually linked by flexible loops. Such loops are accessible to limited proteolysis, which produces a pattern of discrete proteolytic fragments that can be used to delineate the domain structure and location of the linking loops. We have applied limited proteolysis to study the domain structure of the P22 coat protein. Proteolysis was performed at 4 °C using low enzyme-to-substrate ratios (1:1000 to 1:100) due to the limited stability of the coat protein at higher temperatures. Pancreatic elastase, a serine protease, was chosen because it provides the desired low sequence specificity and a preference for longer substrates, which restricts the cleavage to solvent-exposed flexible loops (*19*).

The time course of proteolysis revealed the formation of two metastable fragments (Figure 2a). These fragments were similar in size and appeared to be approximately half the size of coat protein (47 kDa). Fragments could arise from either a stepwise cleavage (e.g., the larger fragment serves as a precursor to the smaller one) or the proteolysis of a single flexible loop. The kinetics of fragment production (Figure 2b) revealed that both fragments appeared at the same rate. The decrease of intact coat protein concentration was roughly equivalent to the combined increase in the fragment populations. These data favor the latter scheme of a single cleavage site located between two domains of the coat protein.

We analyzed the fragments using N-terminal sequencing and mass spectrometry to identify the proteolytic site. The first five residues of the fragment band (ALNEG) were identical to those of the coat protein N-terminus with the N-terminal methionine removed (*20*). The first five residues of the upper fragment band were YRDGT, which were located downstream of the proteolytic site A195-Y196. The C-terminal fragment had a mass of 25 011 ± 30 Da. The predicted mass for the C-terminal fragment, if it was not cleaved further, is 25 010 Da. Thus, the upper fragment band constitutes the intact C-terminal half of the protein. The mass of the N-terminal fragment was 21 658 ± 30 Da, matching the predicted mass of 21 634 Da. Longer incubations with elastase led to a secondary cleavage after Q204 as shown by N-terminal sequencing (Figure 2d). Based on its specificity, elastase is expected to prefer to cleave after V205 and A206 (*19*). The observed cleavage after Q204 may reflect the inaccessibility of V205 and A206 as a result of steric hindrance.

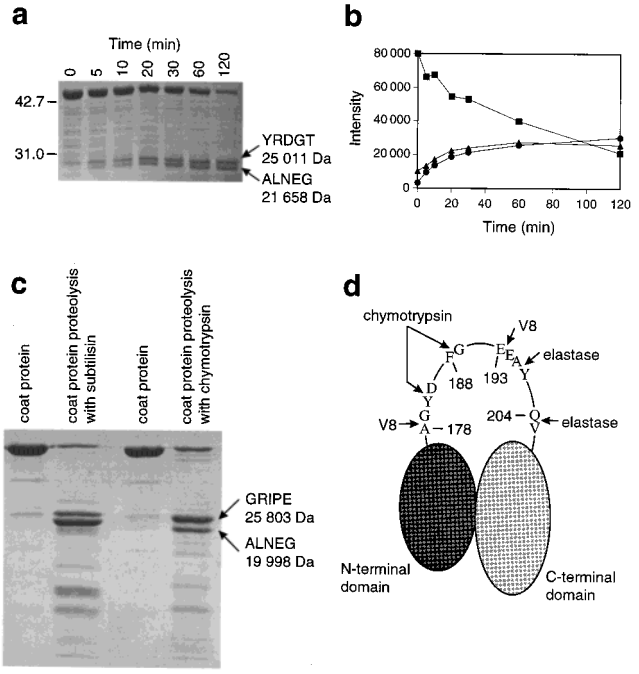

FIGURE 2: Limited proteolysis of P22 coat protein. (a) Time course of coat protein (0.5 g/L) proteolysis with elastase (1:1000 enzyme-to-substrate ratio) at 4 °C. Aliquots were quenched and analyzed on 13% SDS−PAGE. Molecular masses (MALDI-TOF) and N-terminal sequences are shown for the two major proteolytic products. (b) A plot of the relative integrated band intensities from panel a for the N-terminal fragment (solid circles), C-terminal fragment (solid triangles), and intact coat protein (solid squares). (c) Coat protein proteolysis with subtilisin and chymotrypsin (using conditions identical to those used for digestion with elastase). Reactions were quenched with PMSF, and samples were analyzed with 13% SDS−PAGE. (d) Preferred cleavage sites for elastase, chymotrypsin, and bacterial V8 protease were mapped to an accessible loop region (residues 180−205) between the N-terminal domain (21 kDa) and the C-terminal domain (25 kDa).

If elastase is sterically hindered from cleaving in the N-terminal region and C-terminal region, then other proteases should also be sterically hindered. To test proteolytic protection, we used subtilisin, chymotrypsin, and bacterial V8 protease, which have different specificities than elastase. Proteolysis of coat protein with subtilisin produced two fragments, similar in size to the ones found by proteolysis with elastase (Figure 2c). Heterogeneity in the N- and C-termini of the proteolytic fragments arising from the nonspecific nature of subtilisin made it difficult to map the precise location of the cleavage sites. However, proteolysis with chymotrypsin also produced two fragments that were similar in size to the fragments found by proteolysis with elastase or subtilisin (Figure 2a,c). The first five residues of the smaller fragment were ALNEG, which corresponds to the N-terminus of coat protein. The mass of the N-terminal fragment was 19 998 ± 30 Da. Cleavage after Y180 is predicted to generate a fragment with a mass of 20 007 Da. This value is in good agreement with the observed value. The larger fragment appropriately starts with the sequence GRIPE which is located in the middle of the protein (residues 189−194). The measured mass of this fragment was 25 802 ± 30 Da, which compares favorably with a mass of 25 763 Da predicted for a fragment spanning residues 189 to the end of the protein (residue 430). The mass of the excised

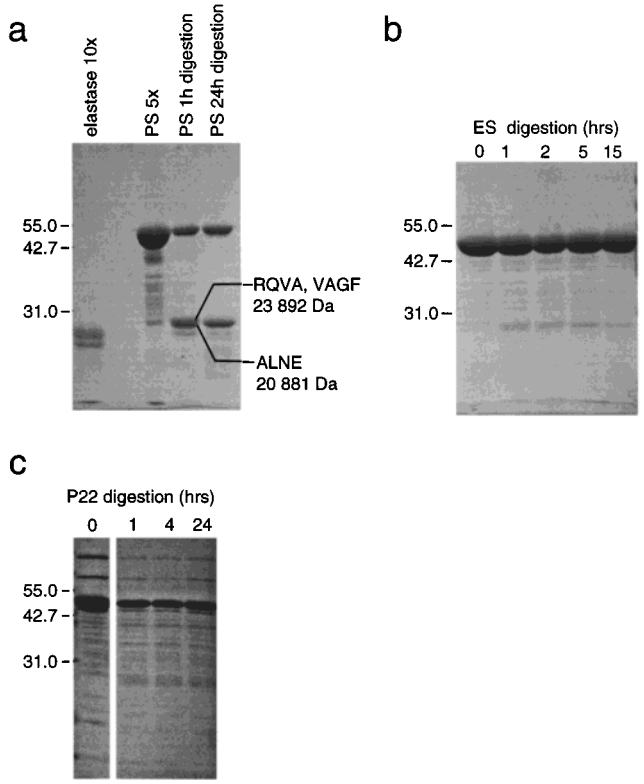

FIGURE 3: Proteolysis of assembled coat protein. (a) Procapsid shells (PS) (2 mg/mL) were digested with elastase (1:20 enzyme-to-substrate ratio) for 1 and 24 h at 35 °C. Reaction products were separated on 13% SDS−PAGE. Elastase was run as a control at 10 times the concentration used in digestion, and the procapsid shells (PS) were run at 5 times the concentration used (PS 5×) during digestion. Bands were transferred onto a PVDF membrane and subjected to 4 cycles of N-terminal sequencing. An aliquot of the 24 h digestion was also subjected to MALDI-TOF mass analysis. (b) Heat-expanded shells (2 mg/mL) were subjected to elastase digestion (1:10 enzyme-to-substrate ratio) for 1−15 h at 35 °C, and products were analyzed on 13% SDS−PAGE. (c) Mature virus (2 mg/mL) was subjected to elastase digestion (1:10 enzyme-to-substrate ratio) for 1−24 h at 35 °C, and products were analyzed on 13% SDS−PAGE.

flexible loop encompassing residues 180−188 (delineated by the two cleavage sites of chymotrypsin) fully accounts for the lower combined mass of these fragments with respect to that of the intact coat protein. Thus, the cleavage sites of both chymotrypsin and elastase are within an accessible loop region of about 25 amino acids (residues 180−205) that connects the N- and C-terminal domains. These results are supported by cleavage with bacterial V8 protease [P1 site specificity Glu preferred over Asp, Phe, or Ala (*19*)], which cleaved after residues A178 and E193 within the loop region (data not shown). Interestingly, no cleavage after D209, 210 (V8) or F208 (chymotrypsin) was detected, suggesting that the region immediately following the loop is protected.

The Two-Domain Structure of Coat Protein Is Preserved within the Procapsid Shell. The domain structure of the coat protein within the procapsid was examined using limited proteolysis. The enzyme-to-substrate ratio was increased to 1:20, and the digestion was performed at 35 °C because no detectable cleavage was achieved under the conditions used for coat protein monomer. One hour digestion with elastase produced two fragments with molecular masses between 20 and 25 kDa (Figure 3a). About half of the coat protein in the procapsid shells remained intact, as determined by densitometry of the stained bands. Incubation for 24 h did not produce further fragmentation but decreased the apparent molecular mass of the larger fragment. Two N-terminal sequences were recovered from the larger fragment (RQVA, VAGF). Both of these are located within the flexible loop region and indicate cleavage from residues 202−204. A single N-terminal sequence (ALNE) was recovered from the fragment band, indicating that this fragment corresponds to the N-terminal half of the protein.

Mass spectrometry on the products of the 24 h digestion identified fragments of several masses. A mass of 23 883 ± 30 Da was assigned to the C-terminal fragment starting with residue V205 (predicted mass 23 892 Da). A smaller peak with mass 24 832 ± 30 likely represented another C-terminal fragment that starts with residue R197 (predicted mass 24 847 Da). A measured mass of 20 879 ± 30 Da corresponded to the N-terminal fragment termination at F188. Taken together, elastase cleaves preferentially after F188, Y196, Q202, and Q204, all of which are located within the flexible loop of the monomer (cf. Figure 2d). In contrast to the monomer, approximately half of the coat protein subunits within the procapsid shell could be digested. Similar results were obtained using pancreatic chymotrypsin and subtilisin. Thus, about half of the coat protein subunits within the procapsids retain their flexible loop, albeit in a less accessible form than in the monomer.

Apart from the flexible loop region, the procapsid coat protein subunit itself is greatly protected against proteolysis. Only a few additional protease-sensitive sites were identified after several days of digestion with elastase, chymotrypsin, and subtilisin (data not shown).

The Hinge Region Becomes Protected upon Expansion. Maturation of the procapsid shell leads to stabilization (*8*) and increased protection of the polypeptide chain against hydrogen deuterium exchange (*14*). It is of interest to map the regions that contribute to these changes. It is particularly desirable to know whether the hinge region remains flexible or undergoes a conformational change. Flexibility of the hinge within the in vitro heat-expanded shells was probed by exhaustive digestion with several proteases. Representative results of such exhaustive digestion by elastase are shown in Figure 3b. The heat-expanded shells were completely resistant to proteolysis with elastase and other proteases (chymotrypsin, subtilisin, and glutamyl protease V8; data not shown). Similarly, extensive proteolytic attack left the mature, DNA-containing, capsids intact (Figure 3c). Thus, the loop region undergoes stabilization during maturation, becoming fully protected in the mature state.

Structure and Stability of the Elastase-Cleaved Coat Protein Monomer

Fragments of Coat Protein Monomer Remain Associated. In an effort to separate the fragments so we could characterize them individually, digested coat protein was loaded onto an anion exchange column. The predicted isoelectric points of the two fragments differ (4.4 for the N-terminal versus 8.1 for the C-terminal fragment); however, both fragments coeluted with the intact coat protein. This observation suggested that the fragments remained associated and had a net charge (at pH 7.6) similar to the intact coat protein.

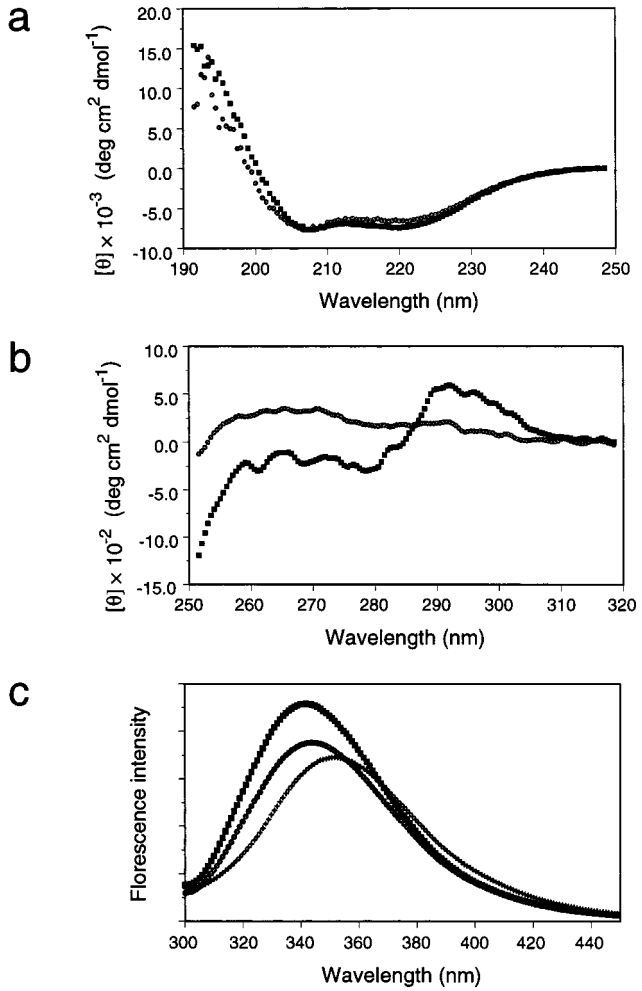

FIGURE 4: Comparison of optical spectra of the intact coat protein monomer and the N/C complex. (a) The far-UV circular dichroism (CD) spectra of the intact coat protein monomer (solid squares) and N/C complex (solid gray circles) were collected from 50 mM Tris buffer solutions, 25 mM NaCl, pH 7.6 and 10 °C, with protein concentrations at 0.2 mg/mL. (b) The near-UV CD spectra of the intact coat protein monomer (solid squares) and N/C complex (solid gray circles) were collected in the same buffer as above, protein concentrations 0.1 mg/mL. (c) Fluorescence spectra of the intact coat protein monomer (solid squares), the N/C complex (solid gray circles), and coat protein unfolded in 6 M GuHCl (crosses). Protein concentrations were 0.031 mg/mL in 50 mM Tris buffer, 25 mM NaCl, pH 7.6, 25 °C.

Therefore, the digested coat protein was analyzed using size-exclusion chromatography and SDS−PAGE to verify that the fragments remained associated. The N- and C-fragments coeluted, presumably as a complex, at the same position as the intact protein. Finally, native gel electrophoresis demonstrated that the N- and C-fragments comigrated to a position typical of intact coat protein (data not shown). For these reasons, we refer to the cleaved coat protein as the N/C complex.

The Associated Coat Protein Fragments Retain Secondary and Most Tertiary Structure. CD and fluorescence spectra of the N/C complex were compared to those of the intact monomer to delineate the role of the flexible hinge in stabilization of secondary and tertiary structure. Comparison of the far-UV CD spectra of the intact coat protein monomer with those of the N/C complex revealed only small changes in secondary structure (Figure 4a). The observed ellipticity decrease at 200 nm and concomitant increase at 220−225 nm indicated the conversion of small amounts of α-helical structure into random coil. In the near-UV, the decreased Cotton effect of aromatic side chains in the cleaved coat protein suggested that the six tryptophan residues have lost some of their asymmetric environment upon cleavage of the loop region (Figure 4b).

The intrinsic tryptophan fluorescence is a sensitive gauge of the tertiary structure of the coat protein (*21, 22*). At room temperature (25 °C), the intact coat protein is well folded with most tryptophan residues sequestered from solvent as manifested by the maximum of the fluorescence emission peak at 338 nm (Figure 4c). Upon denaturation, the coat protein emission peak is reduced in intensity and shifts to 352 nm (Figure 4c). The fluorescence intensity and emission maximum of the N/C complex are intermediate between these forms (344 nm) and indicated at least partial exposure of tryptophan residues to solvent.

The Associated Coat Protein Fragments Are Less Stable than Intact Coat Protein. The thermal denaturation of the intact coat protein and of the N/C complex was followed by CD in order to examine the stabilization effect of the loop region. The intact coat protein displayed an unfolding transition between 38 and 42 °C. As the temperature was increased, the protein aggregated as witnessed by an increase in β-sheet and the appearance of visible aggregates. The unfolding transition could not be monitored directly because the spectra of the aggregated and unfolded forms overlapped at intermediate temperatures. Singular value decomposition (SVD) was used to determine that three components were sufficient to represent all temperature-resolved spectra. As two of the basis spectra were established experimentally (corresponding to the native and the aggregated forms, respectively), the spectrum of a transient unfolding intermediate (populated only at temperatures between 35 and 45 °C) was obtained by rotation of the SVD-based factors using a least-squares method. The three spectra and their mole fractions are shown in Figure 5a,b. The spectrum corresponding to the native form exhibits negative peaks at 208 and 222 nm, diagnostic of α-helical content (Figure 5a). The native structure unfolds cooperatively between 38 and 42 °C with a midpoint at 39 °C (Figure 5b). As a result, the completely unfolded intermediate is partially populated in the transition region. This intermediate most likely served as a precursor for irreversible aggregation that proceeded rapidly above 40 °C (Figure 5b). The spectrum of the aggregate indicates conversion of the unfolded protein into a β-strand conformation (Figure 5a).

The same method was used to determine that four components were required to represent the temperature-resolved spectra of the associated N/C complex. Spectra of the native state, a partially unfolded intermediate, and the fully unfolded ensemble were well determined experimentally. The spectrum of a metastable aggregate was obtained by rotation of the SVD factors (Figure 5c).

The individual spectra at each temperature were fit by linear combination of the constituent spectra, and the fractional composition was determined. The midpoint for the unfolding of the N/C complex was 30 °C (Figure 5d), significantly lower than that of the intact coat protein (39 °C). The first transition was marked by a loss of α-helical secondary structure, and was less cooperative than that of the intact coat protein. Unfolding resulted in the population

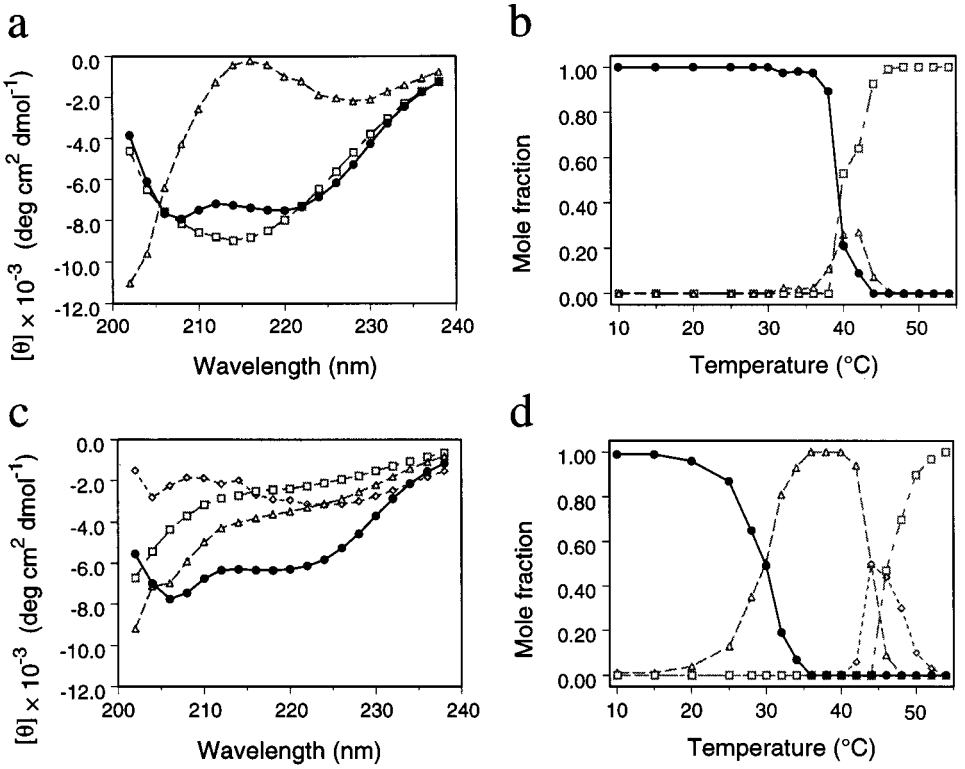

FIGURE 5: Temperature stability of the intact and cleaved coat protein. (a) The far-UV CD spectra corresponding to the native (solid circles), unfolding intermediate (determined by SVD and fitting) (open triangles), and unfolded/aggregated (open squares) species of the intact coat protein monomer and (b) their respective mole fractions. Panels c and d show similar data for the N/C complex: native (solid circles), first unfolding intermediate (open triangles), second unfolding/aggregated intermediate, determined by SVD and fitting (open diamonds), and completely unfolded (open squares) species.

of an intermediate state, with substantial amounts of β-strand secondary structure (Figure 5c). This intermediate is stable up to 41 °C, whereupon it cooperatively converts into an aggregate and/or the completely unfolded form (Figure 5d). Thus, cleavage of the loop not only destabilizes the protein, suggesting the presence of stabilizing interdomain interactions in the native state, but also alters the unfolding pathway.

Structure and Stability of the Elastase-Cleaved Procapsid

Procapsid Shells with Cleaved Hinge Retain Structural Integrity. The protease resistance of coat protein fragments within the procapsid shell suggests that the fragments remain associated with the shell. The structural integrity of shells harboring cleaved subunits was further examined by sedimentation through a 5−20% sucrose gradient followed by SDS−PAGE. Cleaved shells sedimented to the position typical of the intact procapsid shells (Figure 6a,b). Additionally, a fraction of intact shells that migrated slightly faster than the untreated intact control was found in the protease-treated preparation (Figure 6a). The ratio of cleaved to intact subunits within the cleaved fraction was approximately 2:1. None of the cleaved material was found at the top of the gradient (the faint band recovered in the top fraction is due to cleavage in situ by the elastase), confirming that cleaved shells retain structural integrity. Because the cleavage site remains accessible in procapsids, it seems likely that the hinge region is located on the surface of the procapsid shell.

Procapsid Shells with Cleaved Hinge Bind Scaffolding Protein. Intact procapsid shells but not the expanded shells retain the capacity to bind scaffolding protein (*23*). Thus, scaffolding protein binding offers a sensitive probe of the lattice structure with respect to maturation (shell expansion). The cleaved procapsid shells were mixed with a slight molar excess (4:3) of the scaffolding protein and incubated at 20 °C. The reentry of scaffolding protein was monitored by the time-dependent increase in the turbidity of the solution as previously described (*24*) (data not shown). A similar increase was observed for both the intact and cleaved procapsid shells. After completion of the reentry (4 h), the reaction mixtures were separated using sucrose gradient and SDS−PAGE (Figure 6c,d). The cleaved shells accumulated a similar amount of scaffolding protein (Figure 6d) as the intact control (Figure 6c).

Procapsid Shells with a Cleaved Hinge Are Less Stable. Although the intact hinge region is not important for the structural integrity of the procapsid shell, it may contribute to its stability. In the procapsid, the coat protein subunit is stabilized against denaturation by intersubunit contacts (*25, 26*), and cleavage of the loop may affect folding or stability. The fluorescence emission spectra of the intact and elastase-digested procapsid shells are compared in Figure 7a. In contrast to the monomer, the cleaved coat protein subunits within shells exhibit only a slight red shift of the fluorescence emission band, suggesting that despite loop cleavage the tertiary structure remains stabilized by the intersubunit contacts in the procapsid.

The stabilities of the intact and cleaved procapsid shells were monitored by fluorescence during GuHCl-induced disassembly. Conditions which dissociate the procapsid result in the immediate denaturation of the coat protein subunit; therefore, the relative stability of assembled shells can be monitored by the red shift in fluorescence emission ac-

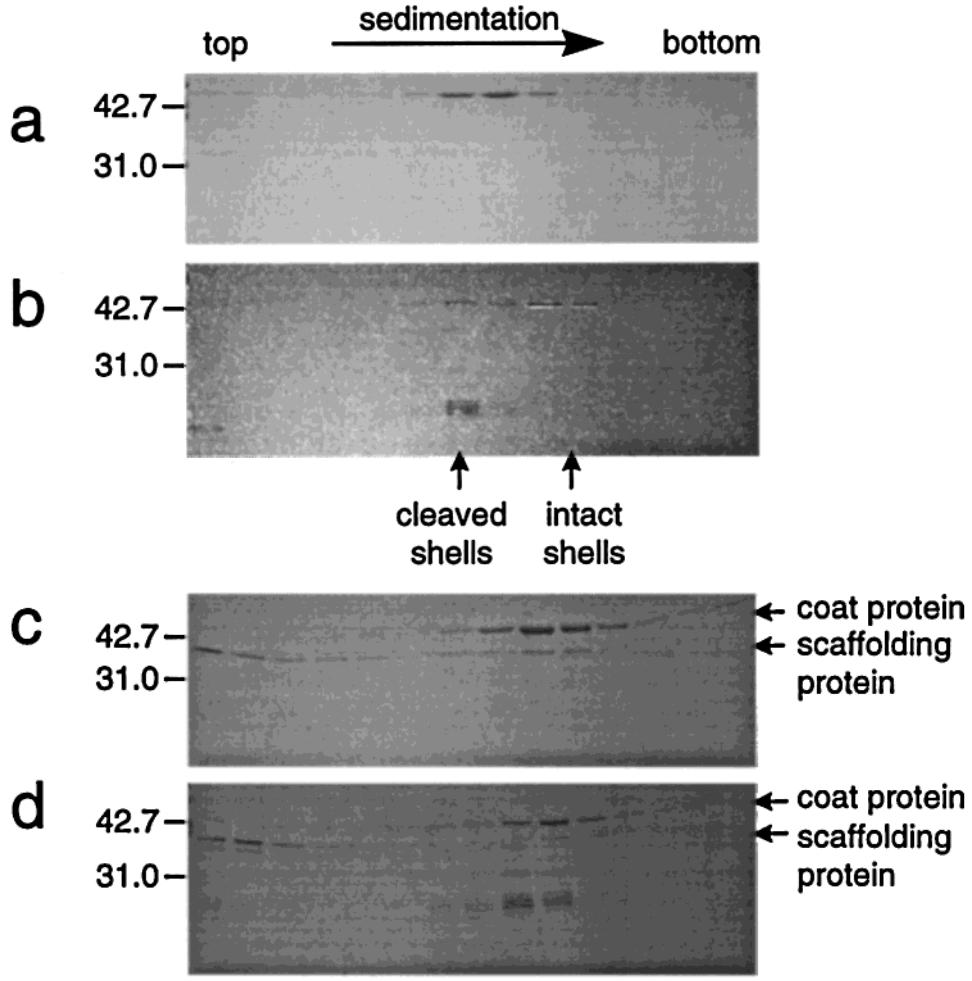

FIGURE 6: Comparison of procapsid shells with hinge-cleaved procapsid shells. Procapsid shells were proteolyzed with elastase (1:10 enzyme-to-substrate ratio) at 30 °C for 1 h. Empty shells were either separated on a 5−20% sucrose gradient or subjected to scaffolding re-entry and then separated on the gradient. Fractions from the gradient were run on 10−20% gradient SDS−PAGE. (a) Untreated empty procapsid shells. (b) Elastase-cleaved empty shells. (c) Untreated procapsid shells after re-entry of scaffolding protein. (d) Elastase-cleaved procapsid shells after re-entry of the scaffolding protein.

companying subunit denaturation. The shift in the center of spectral mass as a function of GuHCl concentration was used to compare the stability of intact and cleaved procapsid shells (*27*)(Figure 7b). The cleaved shells dissociated at significantly lower GuHCl concentration ($c_{1/2}$ = 1.25 M) than the intact control ($c_{1/2}$ = 1.50 M) and in a less cooperative manner. Thus, the cleaved shells are less stable, and the intact loop region increases the cooperativity of the assembly. These results are similar to those reported above for the cleaved monomer coat protein.

Activity of the Cleaved Coat Protein

It is possible the loop region plays an important role in both assembly and maturation because it becomes progressively more protease-resistant following assembly and maturation. The cleaved monomer coat protein and procapsids were tested for assembly and expansion in vitro, respectively.

Assembly of the Cleaved Monomer. Procapsids can be assembled in vitro by mixing coat protein monomer with scaffolding protein (*15, 28*). Chymotrypsin-cleaved coat protein monomer was mixed with an excess of scaffolding protein and incubated at 20 °C for 4 h. The products of assembly of the cleaved and intact coat proteins were separated on a sucrose gradient. In contrast to the intact coat protein control, almost all the cleaved coat protein remained at the top of the gradient; less than 5% of the digested complex co-assembled with the residual intact coat protein. Unlike the intact procapsids, this assembly product bound negligible amounts of the scaffolding protein, which is indicative of aberrant assemblies (*24*). Even less of the elastase- and subtilisin-cleaved coat protein assembled, presumably due to a lesser amount of residual intact coat protein. Thus, the intact loop region is indispensable for procapsid assembly in vitro.

Expansion of the Cleaved Procapsid Shell. Maturation of the P22 capsid is accompanied by shell expansion which can be mimicked in vitro by heating (*8*). It was shown previously that heat-induced expansion is an exothermic process governed by a high activation energy (*8, 14*). It is likely that the loop region may affect the energetics of expansion because it affects the stability of both the monomer and the procapsid shell. Therefore, we have examined the effect of hinge cleavage on the kinetics of expansion. Procapsid shells were cleaved with elastase for 1 h at 30 °C and then subjected to heat-induced expansion at different temperatures for specific time periods (Figure 8a).

Due to their larger hydrodynamic radius, expanded shells migrate more slowly than do procapsids on agarose gels

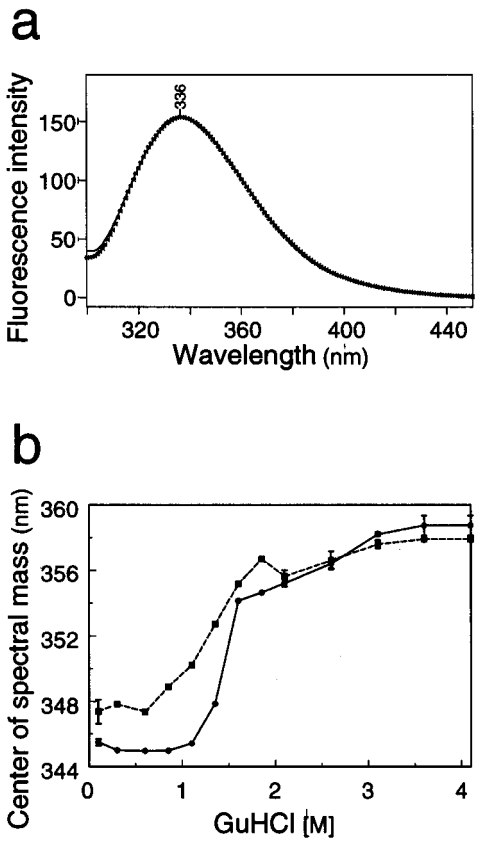

FIGURE 7: Comparison of structural integrity and stability using fluorescence. (a) Fluorescence emission spectra of the intact (solid line) and elastase-cleaved (crosses) procapsids at 25 °C, 0.01 mg/mL shells in 50 mM Tris and 25 mM NaCl. (b) Stability of intact and cleaved procapsid shells monitored by fluorescence during GuHCl-induced disassembly. Center of spectral mass = $\Sigma_i \lambda_i I^{F_i}$ is reported as a function of GuHCl concentration for the intact procapsids (solid circles) and elastase-cleaved procapsids (solid squares).

(Figure 8a). Uncleaved procapsids did not significantly expand at 40 °C even after 8 h, while approximately 50% of the cleaved shells did expand. At 50 °C, the cleaved procapsids expanded almost an order of magnitude faster than the intact shells (note difference in time scale in Figure 8a). Despite such a rapid rate of expansion, a fraction of the cleaved shells did not expand even after a long period of heating at 50 °C. The structural basis for this heterogeneity is not clear. Two shell populations were also observed for intact shells [(*27*) and unpublished data of R.T. and P.E.P.].

Following the 50 °C treatment (21 min), the agarose gel bands corresponding to the expanded and nonexpanded cleaved shells were excised and analyzed on SDS−PAGE (Figure 8b). The coat protein fragments were found exclusively in the upper band, which corresponds to the expanded shells (Figure 8b, lane E). In contrast, the lower band, which corresponds to the nonexpanding population of procapsids, contained only intact coat protein (Figure 8b, lane P). Thus, the nonexpanding population of procapsid shells harbors protected hinge regions.

The apparent activation energy for expansion of cleaved shells was determined from the temperature dependence of the expansion rate (Figure 8c). The activation energy of 97 kcal/mol (at the 35−50 °C temperature region) is similar to that of the intact procapsid shell which was determined in a higher temperature range, 60−70 °C (*14*).

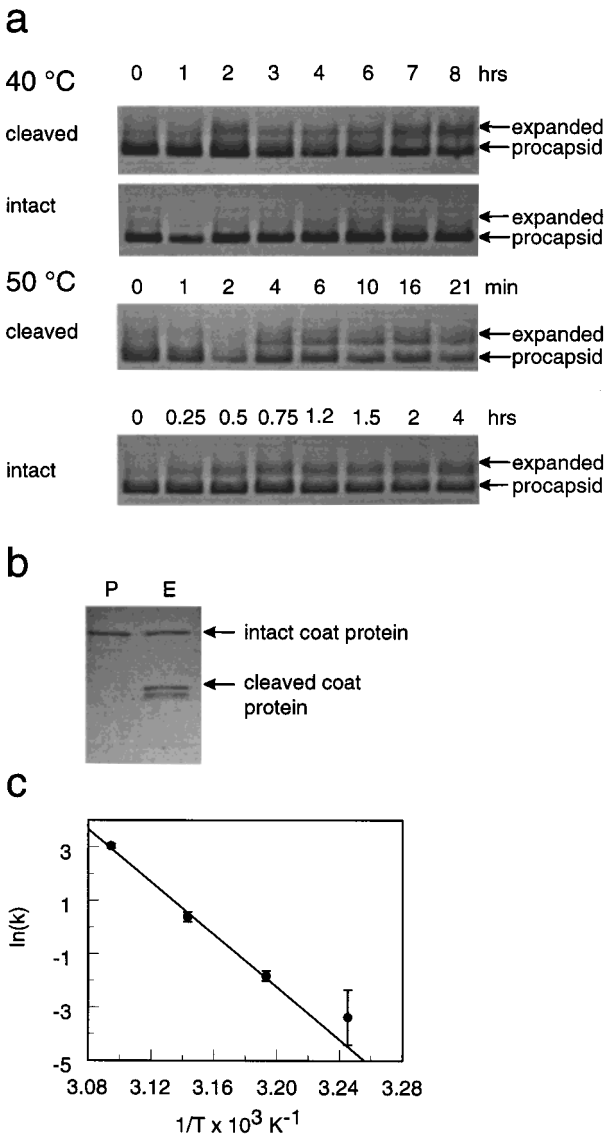

FIGURE 8: Expansion of elastase-cleaved procapsid shells. (a) Cleaved and uncleaved procapsid shells were heat-expanded at 40 and 50 °C for the specified time periods. The products were separated on a 1.2% agarose gel and detected with Coomassie blue. Due to their larger hydrodynamic radius, expanded shells migrate slower through the gel. (b) Agarose bands corresponding to the nonexpanding fraction (lane P) and the expanded fraction (lane E) of the elastase-cleaved shells were cut out, extracted, and separated on 10−20% gradient SDS−PAGE. (c) Intensities from stained agarose gels were used to determine expansion rates at different temperatures for elastase-cleaved shells. The resulting rates are presented in a semilogarithmic (Arrhenius) plot. The solid line represents a weighted least-squares fit of the experimental rates with apparent activation energy $E_a = 97 \pm 10$ kcal/mol shells.

DISCUSSION

Domain Structure. In the present study, we have localized two putative domains of the P22 coat protein. Each of the domains (21 and 25 kDa) is large enough to form a "jelly-roll" β-barrel structure (typically 18−24 kDa) that is characteristic of other icosahedral viruses (*29*). Like the subunits of these viruses, P22 coat protein is rich in β-strand secondary structure (*11, 30*).

The two domains are connected via a flexible hinge encompassing 25 amino acids. In agreement with its suggested surface location, the loop region constitutes the most

polar region within the coat protein sequence (*31*). The N- and C-domains remain associated via noncovalent interactions after cleavage of the hinge in the monomer. Cleavage of the hinge does not lead to unfolding of the monomer (secondary structure) but affects tertiary structure (reflected by tryptophan environment) and stability of the protein.

Polypeptide Disorder and Dynamics in Assembly and Maturation. The loop region is readily accessible in the monomer. Although the loop region is also accessible to proteolysis in procapsid shells, the rate of digestion is extremely slow at low temperatures. The loop becomes more accessible at elevated temperatures, suggesting that local unfolding is required to make the loop accessible. These motions are further constrained in the mature form, which is completely protease-resistant. These data are consistent with the observation that the hydrogen/deuterium exchange protection of this region increases during assembly and maturation (manuscript in preparation). Thus, the loop region becomes progressively less accessible during assembly and maturation. Conformational ordering and restriction of the main chain dynamics during P22 assembly and maturation have been observed using spectroscopic techniques (*14*). Raman spectroscopy demonstrated that assembly of the monomer into procapsids is accompanied by the ordering of disordered regions into the β-strand (R.T., P.E.P., and George J. Thomas, Jr., unpublished data) and that the subunit becomes further stabilized during maturation. While the Raman technique cannot assign the changes to a particular region within the sequence, the experiments described above suggest that at least some of the disorder-to-order transition is associated with amino acids within and in the vicinity of the loop.

Conformational Heterogeneity within the Hinge Region of the Procapsid Lattice. The proteolysis experiments identified two classes of procapsid shells. The first class exhibits a protease-susceptible loop whereas the second class is completely protease-resistant. The protease-resistant class did not expand at temperatures below 60 °C. However, the efficiency of expansion of these particles could be increased without changing the apparent activation energy by addition of 0.5−1 M urea (R.T. and P.E.P., unpublished). Such a moderate concentration of urea would be expected to promote local unfolding, presumably within the loop region, and may shift the population distribution toward the "expansion-active" precursor by allowing for loop flexibility. Conformational heterogeneity of large macromolecular assemblies was previously proposed based on the apparent concentration independence of pressure dissociation (*32*).

Conformational Flexibility of the Hinge Region Is Essential for Expansion. The apparent activation energy (97 kcal/mol) for expansion of cleaved shells over the 30−50 °C temperature range is similar to that of the intact shells at high temperature [90 kcal/mol, (*14*)]. At low temperature, expansion of the intact shells is extremely slow despite being governed by a lower activation energy (about 40 kcal/mol). When accelerated by urea, the activation energy is about 90 kcal/mol (R.T. and P.E.P., unpublished results), similar to that seen when the loop is proteolytically removed. Therefore, we suggest that flexibility of the loop region controls only the first step in expansion, which is rate-limiting at low temperatures in the absence of urea (35−55 °C). The slow rate, despite the low activation energy, is likely due to the fact that constraint of the loop is contributing primarily to the entropy of activation rather than the enthalpy of activation. Thus, the rate of this step exhibits a relatively low temperature dependence (activation energy) (*14*). At higher temperature, or in the presence of urea, the entropic barrier is diminished, and the high activation barrier of approximately 90 kcal/mol is dominated by the enthalpy of activation. This barrier likely corresponds to the proposed partial dissociation of the domains during expansion (*14*). This same barrier governs the expansion of the cleaved shells at low temperature due to the diminished contribution of the absent loop region to the entropy of activation.

Model for Maturation. With these data we have developed the following model of domain movement to explain the maturation process. First we propose that rigid body movement of the two domains is likely to account for the conformational change accompanying loop cleavage of the monomer coat protein, because the hydrodynamic properties of the cleaved coat protein are similar to those of the intact coat protein (reflected by size-exclusion chromatography). This domain movement is also likely to be responsible for the drastically reduced assembly activity of the cleaved coat protein. It is likely that the relaxation of domains in the N/C complex precludes activation of the coat protein for assembly because loop cleavage does not directly perturb the scaffolding binding site (at least within the procapsid shell). Thus, the loop provides an essential conformational constraint for the coat protein causing it to maintain its assembly-competent conformation.

Second, we propose that during maturation the loop region becomes more ordered. Thus, the nonexpanding shells (approximately 50%) are likely to have the loop region in an "ordered" mature conformation and require local unfolding of this stable structure in order to undergo expansion. Therefore, local disorder or flexibility within the loop region is essential for shell expansion. In this regard, it is interesting to note that of 18 randomly distributed temperature-sensitive mutations in the coat protein, only 1, A174N, interferes with the ability of the wild type protein to form phage in a coinfection (*31*). This mutation lies immediately adjacent to the protease-defined loop region. The dominant phenotype may arise from the formation of mixed procapsids containing both wild-type and mutant subunits which are unable to expand properly due to decreased flexibility in the mutant subunit. Interestingly, a glycine substitution at the same locus, A174G, does not produce a dominant phenotype.

REFERENCES

1. Teschke, C. M., King, J., and Prevelige, P. E., Jr. (1993) *Biochemistry 32*, 10658−10665.
2. Prevelige, P. E., Jr. (1998) *Trends Biotechnol. 16*, 61−65.
3. Crick, F. H. C., and Watson, J. D. (1956) *Nature 177*, 473−475.
4. Douglas, T., and Young, M. (1998) *Nature 393*, 152−155.
5. Prevelige, P. E., Jr., and King, J. (1993) *Prog. Med. Virol. 40*, 206−221.
6. King, J., and Casjens, S. (1974) *Nature 251*, 112−119.
7. Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J., and Chiu, W. (1993) *J. Mol. Biol. 231*, 65−74.
8. Galisteo, M. L., and King, J. (1993) *Biophys. J. 65*, 227−235.

9. King, J., and Chiu, W. (1997) in *Structural Biology of Viruses* (Chiu, W., Burnett, R., and Carcea, R., Eds.) pp 288−351, Oxford University Press, Inc., New York.
10. Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R., and Steven, A. C. (1996) *J. Mol. Biol. 263*, 447−462.
11. Prevelige, P. E., Jr., Thomas, D., King, J., Towse, S. A., and Thomas, G. J., Jr. (1990) *Biochemistry 29*, 5626−5633.
12. Prevelige, P. E., Jr., Thomas, D., Aubrey, K. L., Towse, S. A., and Thomas, G. J., Jr. (1993) *Biochemistry 32*, 537−543.
13. Tuma, R., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (1996) *Biochemistry 35*, 4619−4627.
14. Tuma, R., Prevelige, P. E., Jr., and Thomas, G. J., Jr. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9885−9890.
15. Prevelige, P. E., Jr., Thomas, D., and King, J. (1988) *J. Mol. Biol. 202*, 743−757.
16. Yang, J. T., Wu, C.-S. C., and Martinez, H. M. (1986) in *Methods in Enzymology*, pp 208−269, Academic Press, New York.
17. Paatero, P., and Tapper, U. (1993) *Chemom. Intell. Lab. Syst. 18*, 183−194.
18. Paatero, P., and Tapper, U. (1994) *Environmetrics 5*, 111−126.
19. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (1998) *Handbook of proteolytic enzymes*, Academic Press, San Diego.
20. Flinta, C., Persson, B., Jörnvall, H., and von Heijne, G. (1986) *Eur. J. Biochem. 154*, 193−196.
21. Teschke, C. M., and King, J. (1993) *Biochemistry 32*, 10839−10847.
22. Teschke, C. M., and King, J. (1995) *Biochemistry 34*, 6815−6826.
23. Greene, B., and King, J. (1994) *Virology 205*, 188−197.
24. Greene, B. (1995) in *Biological sciences*, p 244, MIT Press, Cambridge, MA.
25. Prevelige, P. E., Jr., King, J., and Silva, J. L. (1994) *Biophys. J. 66*, 1631−1641.
26. Foguel, D., Teschke, C. M., Prevelige, P. E., Jr., and Silva, J. L. (1995) *Biochemistry 34*, 1120−1126.
27. de Sousa, P. C., Tuma, R., Prevelige, P. E. J., Silva, J. L., and Foguel, D. (1999) *J. Mol. Biol. 287*, 527−538.
28. Prevelige, P. E., Jr., Thomas, D., and King, J. (1993) *Biophys. J. 64*, 824−835.
29. Rossmann, M. G., and Johnson, J. E. (1989) *Annu. Rev. Biochem. 58*, 533−573.
30. Thomas, G. J., Jr., Li, Y., Fuller, M. T., and King, J. (1982) *Biochemistry 21*, 3866−3878.
31. Gordon, C. L., and King, J. (1994) *Genetics 136*, 427−438.
32. Silva, J. L., Foguel, D., Da Poian, A. T., and Prevelige, P. E. (1996) *Curr. Opin. Struct. Biol. 6*, 166−175.

BI9915420