doi:10.1016/j.jmb.2005.02.021 J. Mol. Biol. (2005) 347, 935–948

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Domain Study of Bacteriophage P22 Coat Protein and Characterization of the Capsid Lattice Transformation by Hydrogen/Deuterium Exchange

**Sebyung Kang and Peter E. Prevelige Jr***

*Department of Biochemistry and Molecular Genetics and Department of Microbiology University of Alabama at Birmingham, Birmingham AL 35294, USA*

Viral capsids are dynamic structures which undergo a series of structural transformations to form infectious viruses. The dsDNA bacteriophage P22 is used as a model system to study the assembly and maturation of icosahedral dsDNA viruses. The P22 procapsid, which is the viral capsid precursor, is assembled from coat protein with the aid of scaffolding protein. Upon DNA packaging, the capsid lattice expands and becomes a stable virion. Limited proteolysis and biochemical experiments indicated that the coat protein consists of two domains connected by a flexible loop.

To investigate the properties and roles of the sub-domains, we have cloned them and initiated structure and function studies. The N-terminal domain, which is made up of 190 amino acid residues, is largely unstructured in solution, while the C-terminal domain, which consists of 239 amino acid residues, forms a stable non-covalent dimer. The N-terminal domain adopts additional structure in the context of the C-terminal domain which might form a platform on which the N-terminal domain can fold.

The local dynamics of the coat protein in both procapsids and mature capsids was monitored by hydrogen/deuterium exchange combined with mass spectrometry. The exchange rate for C-terminal domain peptides was similar in both forms. However, the N-terminal domain was more flexible in the empty procapsid shells than in the mature capsids. The flexibility of the N-terminal domain observed in the solution persisted into the procapsid form, but was lost upon maturation. The loop region connecting the two domains exchanged rapidly in the empty procapsid shells, but more slowly in the mature capsids. The global stabilization of the N-terminal domain and the flexibility encoded in the loop region may be a key component of the maturation process.



*Keywords:* bacteriophage P22; capsid maturation; hydrogen/deuterium exchange; mass spectrometry

**Corresponding author*

## Introduction

Biological processes are often performed by large macromolecular complexes.[1] The assembly of these complexes must be tightly regulated to avoid premature assembly or aggregation. In order to control the assembly, the protein subunits are often synthesized in an unassociable form and subsequently switched to an associable form either through the use of effector molecules or through interactions with the growing structure itself.[2]

Viral capsids are one well-known example of dynamic supramolecular structures that are self-assembled and undergo a series of controlled structural transformations to reach an infectious form. The *Salmonella typhimurium* bacteriophage P22 has been studied extensively as a model for icosahedral capsid assembly and as a prototype for a morphogenesis of mammalian viruses.[3] The multi-step assembly pathway of bacteriophage P22 is well characterized biochemically and

Abbreviations used: cryoEM, cryo electron microscopy; ESI, electrospray ionization; FT-ICR MS, Fourier transform-ion cyclotron resonance mass spectrometry; TOF, time of flight; Q-TOF, quardrupole time of flight; MALDI-MS, matrix assisted laser desorption ionization; ppm, part per million.

E-mail address of the corresponding author: prevelig@uab.adu

0022-2836/$ - see front matter 

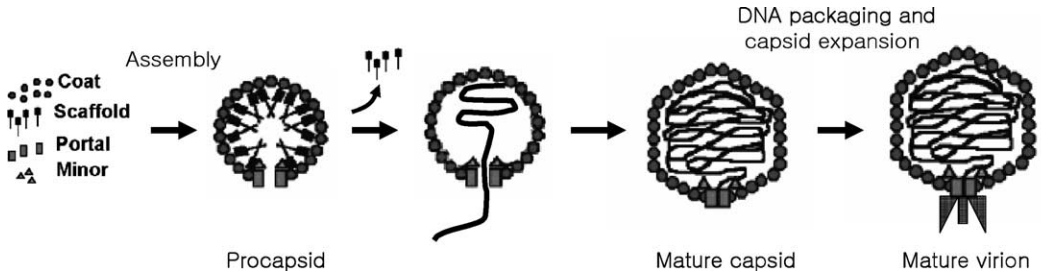


**Figure 1.** Assembly pathway of bacteriophage P22. The meta-stable procapsids are formed by co-assembly of 420 copies of coat protein monomer, 300 copies of scaffolding protein, 12 copies of portal protein and a few copies of several minor proteins. As concatomeric phage DNA is packaged through the portal vertex, the procapsids are transformed into the mature capsids which are 10% larger in diameter and more stable than procapsids. Subsequently, the portal ring is closed and tail attachment occurs.

genetically[4] (Figure 1). In the first step, the procapsid, a $T=7$ meta-stable viral capsid precursor with a diameter of 580 Å, is assembled from 420 copies of the 47 kDa coat protein with the aid of approximately 300 copies of the 33 kDa scaffolding protein and minor proteins which are required for infectivity. The scaffolding protein is not required for coat protein folding, but plays a key role in assembly fidelity and activity.[5,6] The concatomeric phage DNA is packaged into the procapsid in a headful manner through the action of terminase proteins fueled by ATP hydrolysis. Concomitant with DNA packaging, the meta-stable procapsid is transformed into the mature capsid and the scaffolding protein is released. The structural transformation results in a 10% increase in shell diameter,[7] a pronounced angular appearance and an increase in capsid stability.[8]

Although the secondary structure of the P22 coat protein subunit is only slightly changed during assembly,[9–11] the stability and compactness of the coat protein subunits are substantially increased.[11] Limited proteolysis studies have indicated that the coat protein monomer consists of two protease stable domains connected *via* a flexible loop region.[12] This loop region is susceptible to protease in the monomeric coat protein and in procapsids, but is protease-resistant in mature capsids. These observations suggested that domain movement involving hinge bending mediates assembly and shell maturation.[7,10,11] Cryo electron microscopy (cryoEM)-based reconstructions of the procapsid and mature capsid at ~9 Å have revealed the rearrangement and movement of structural elements.[13] The rearrangement of structural elements was localized and tilting and bending of large structural elements was observed.

Although X-ray crystallography and cryoEM provide high-resolution structural information, they cannot follow structural dynamics, nor can they monitor continuous conformational changes of macromolecules. Amide protons in proteins undergo exchange with solvent protons and the exchange rates depend on solvent accessibility and protein stability.[14] Solvent shielding and intra- or intermolecular hydrogen bonding are directly related to protein structure, protein integrity and macromolecular structure formation.[15] Therefore the amide hydrogen exchange rate is a sensitive probe of protein structure, dynamics and stability. NMR spectroscopy has typically been used to measure individual hydrogen/deuterium exchange rates to obtain high resolution mapping of protein dynamics.[16–18] However, NMR spectroscopy is limited to relatively small proteins due to the spectral complexity of large proteins and complexes. Mass spectrometry makes it possible to apply this technique to large protein complexes such as viruses.[19–21] The incorporation of deuterons into proteins results in a mass increase which can be monitored by mass spectrometry. To obtain localized structural information, the protein can be proteolytically fragmented.[22,23] Recently, mass spectrometry-based hydrogen/deuterium exchange has been used to investigate structural features accompanying protein folding and unfolding,[24–27] functionally different forms of protein,[28] and protein–protein interactions.[20,24,29]

In this study, we have cloned each individual domain of the P22 coat protein and studied the structural and functional properties of the isolated domains. We have also investigated the local stabilities of the procapsids and the mature capsids with hydrogen/deuterium exchange combined with mass spectrometry to address the conformational changes accompanying maturation in detail.

## Results

### Cloning and purification of the N and C-terminal domains

Previous limited proteolysis studies of monomeric coat protein suggested it consisted of two domains connected by a flexible loop region.[12] In order to characterize the individual domains, we have cloned the N-terminal (residues 1–190) and the C-terminal domain (residues 191–429) into a pET-3a-based expression vector. When expressed in *Escherichia coli*, both the N and C-terminal domains formed inclusion bodies. To purify each domain the inclusion bodies were denatured, diluted and

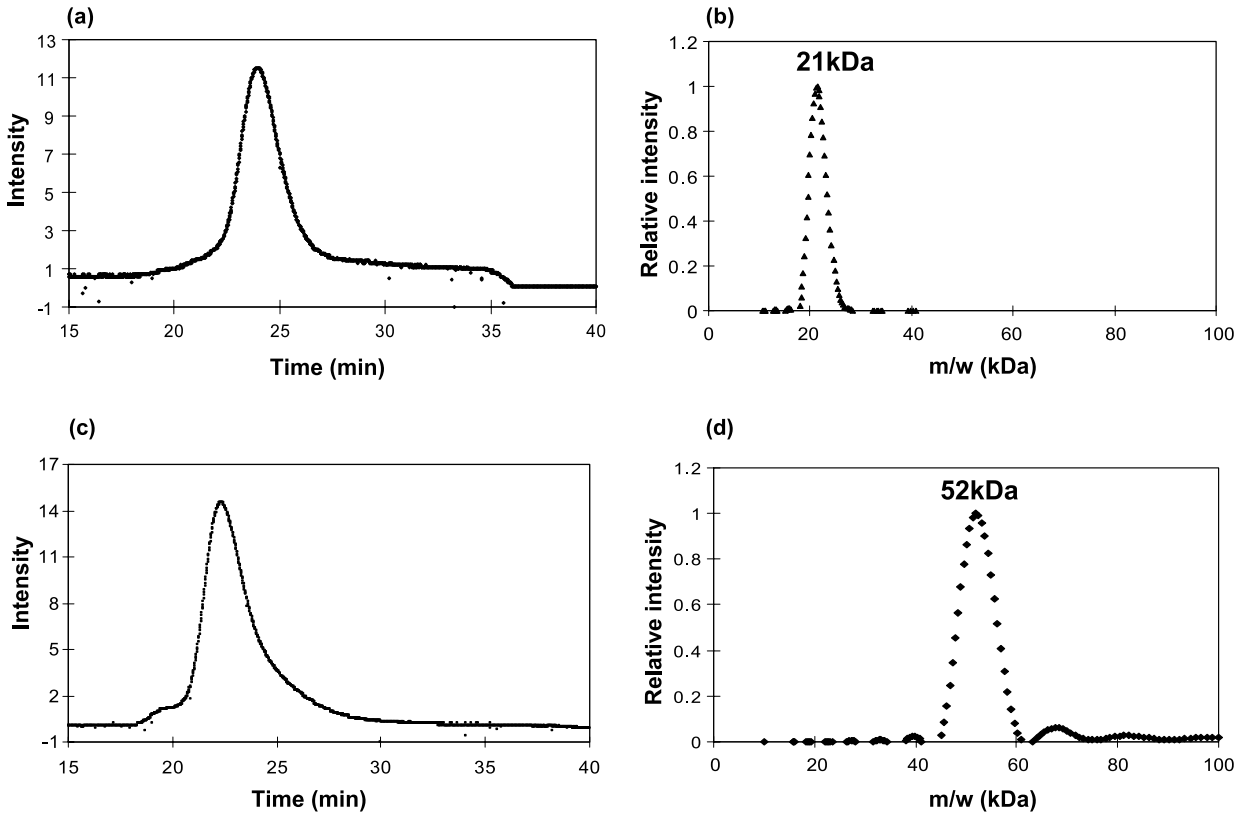


**Figure 2.** SEC and LS of N and C-terminal domains. (a) The refolded N-terminal domain (0.5 ml at 0.4 mg/ml) was loaded onto a 10 mm×300 mm Superdex 75 (Amersham Bioscience) size exclusion column and eluted at a rate of 0.5 ml/minute and detected by refractive index. (b) The molecular mass distribution of the N-terminal domain was determined by simultaneous measurement of refractive index and light scattering at 90°. Molecular mass and mass distributions were established using the software, which is provided by the manufacturer (Discovery32). (c) and (d) are same as (a) and (b) but for the C-terminal domain. The molecular mass of the N-terminal domain was 21 kDa and the C-terminal domain was 52 kDa. The values of $M_w/M_n$ were 1.012 and 1.054 for the N and the C-terminal domains, respectively.

refolded by dialysis following a procedure similar to that used to refold the intact coat protein monomer from procapsids.[6] As a control, the intact coat protein was cloned into the same vector, expressed, and purified and found to have the same activity profile as coat monomer obtained from dissociation of procapsids.

### The C-terminal domain forms stable non-covalent dimer in solution

The refolded N and C-terminal domains were analyzed by size exclusion chromatography in combination with light scattering. Both the N and C-terminal domains eluted as single peaks (Figure 2(a) and (c)). Although they eluted with similar retention times, the molecular mass of the N-terminal domain determined by static light scattering was 21 kDa, while the molecular mass of the C-terminal domain was 52 kDa indicating that it was eluting as a dimer (Figure 2(b) and (d)). The ratio of $M_w/M_n$ which represents polydispersity was 1.012 and 1.054 for the N and C-terminal domains, respectively, indicating the peaks were largely homogeneous. The C-terminal domain contains a single cysteine (residue 404). To determine whether the dimer was the result of cysteine oxidation, both non-reducing and reducing SDS-PAGE analysis was performed. The C-terminal domain electrophoresed as a monomer in both gels (data not shown). Taken together these data suggest that the C-terminal domain formed stable non-covalent dimers in solution.

### The N-terminal domain is highly disordered in solution

In order to study the integrity of the tertiary structure of the domains, the fluorescence spectra of the intact coat monomer and the individual N and C-terminal domains were recorded under both native and denaturing conditions. The coat protein monomer contains six tryptophan residues which are evenly distributed between the two domains. Under native conditions, both the coat protein monomer and the C-terminal domain dimer displayed fluorescence emission maxima at 338 nm, and high fluorescence intensity (Figure 3(a)). Under denaturing conditions the emission maxima of the coat protein monomer and the C-terminal domain

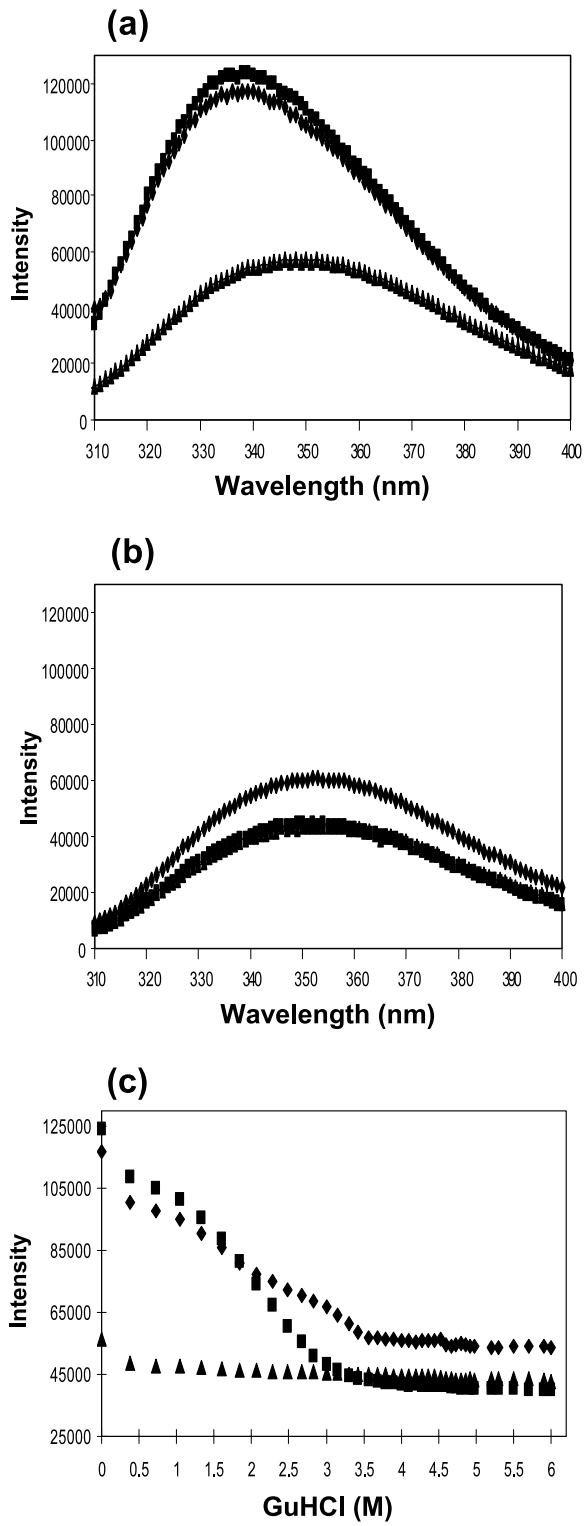


**Figure 3.** Fluorescence spectra of the N-terminal domain, the C-terminal domain and coat protein monomer. (a) Fluorescence spectra of the N-terminal domain (triangle), the C-terminal domain (square) and coat protein monomer (diamond) were obtained in buffer B at 25 °C using a ISS PC1 photon counting spectrofluometer ($\lambda_{exc}$ = 280 nm) in 1 cm path length cells. Protein concentrations were 5 μM each. (b) Fluorescence spectra of denatured coat protein monomer and individual domains were obtained in buffer B supplemented with 6 M GuHCl. (c) The change in fluorescence intensity during GuHCl-induced denaturation as monitored by fluorescence emission intensity. The final protein concentration in all cases was 5 μM. Emission intensities of coat protein monomer (diamond) and the C-terminal domain (square) were recorded in buffer B at 338 nm, and the N-terminal domain (triangle) was recorded at 349 nm.

dimer shifted to 352 nm and the intensity of emission peak was dramatically reduced (Figure 3(b)). These results indicate that the tryptophan residues of the coat protein monomer and the C-terminal domain dimer were buried and relatively inaccessible under native conditions, but became solvent exposed and therefore red-shifted as these proteins underwent denaturation.

In contrast, the N-terminal domain displayed an emission maximum at 349 nm and relatively low fluorescence intensity under native conditions. There was relatively little change in emission maximum (from 349 nm to 352 nm) and fluorescence intensity under denaturing conditions (Figure 3(a) and (b)). These results suggest that the tryptophan residues of the N-terminal domain are solvent exposed under both native and denaturing conditions and suggest that the N-terminal domain might be disordered in solution.

The stability of the coat protein monomer and the individual domains was monitored by fluorescence spectral intensity during GuHCl-induced denaturation (Figure 3(c)). Both the intact coat protein and the C-terminal domain dimer displayed GuHCl-induced denaturation whereas the fluorescence intensity of the N-terminal domain was essentially unchanged from 0 M to 6 M GuHCl.

The CD spectrum of the N-terminal domain had the characteristic shape of random coil structure; an ellipticity increase at 210–230 nm and concomitant decrease at 190–205 nm (Figure 4(a)). This result agrees well with the results from fluorescence spectroscopy, and suggests that the N-terminal domain is highly disordered in solution. The C-terminal domain dimer retained secondary structure and its CD spectra were concentration-independent consistent with the gel chromatography data indicating that the C-terminal domain dimers were stable in solution (Figure 4(a)).

To determine whether the isolated domains adopt structure in the context of the intact protein, the CD spectra of the individual domains were compared to that of intact coat protein (Figure 4(b)). The combined CD spectrum of the two domains did not correspond to the spectrum of intact coat monomer, indicating conformational differences between the free domains and the domains in the context of the whole protein. The individual domains, as well as an equimolar mixture of the two, were unable to assemble into procapsids *in vitro* (data not shown). The CD spectra and *in vitro* assembly results imply that the N-terminal and the C-terminal domain act cooperatively in folding and assembly.

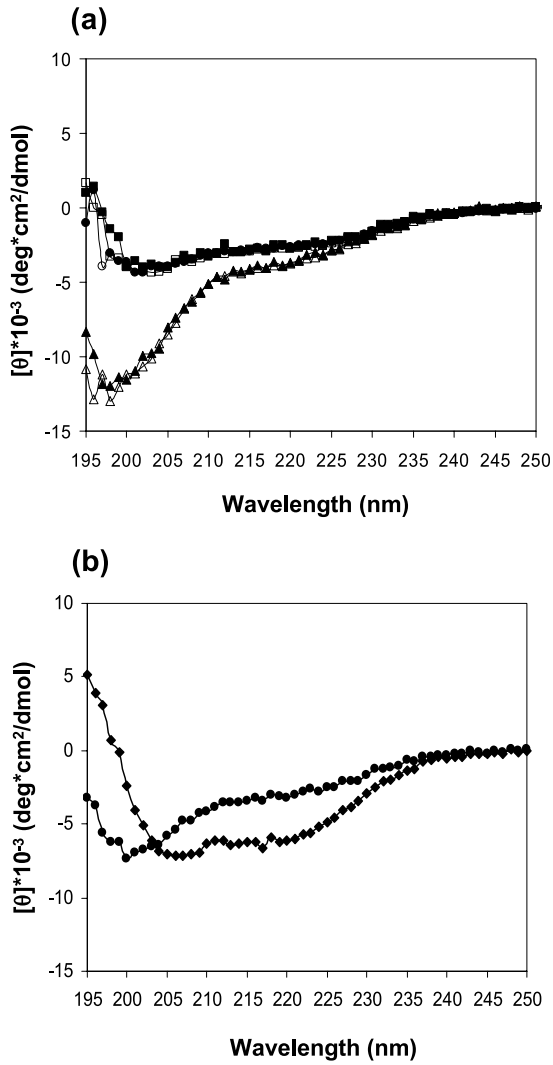


**Figure 4.** Circular dichroism spectra of the N-terminal domain, the C-terminal domain and coat protein monomer. (a) CD spectra of the N-terminal domain and the C-terminal domain were measured in 25 mM $PO_4$ and 25 mM NaCl (pH 7.6) using 1 mm path length cells. Protein concentrations of the N-terminal domain were 0.1 mg/ml (filled triangle) and 0.2 mg/ml (open triangle). Protein concentrations of the C-terminal domain were 0.1 mg/ml (filled square), 0.2 mg/ml (open square) and 0.4 mg/ml (open circle). (b) CD spectra of the N-terminal and the C-terminal domains were combined and weight-averaged (circle), and the CD spectrum of the intact coat protein was measured under same condition as the isolated domains (diamond).

## Hydrogen/deuterium exchange and peptic fragmentation

To determine if the flexibility shown in the N-terminal domain in solution is evident in the procapsids, we performed hydrogen/deuterium exchange combined with mass spectrometry on both the empty procapsid shells and the mature capsids. To insure that any observed differences in exchange rates between the two forms were due to alteration in protein/protein interactions, rather than protein/DNA interactions, the mature form of the virion was produced using a mutant which allows the DNA to leak out after packaging. Previous studies have demonstrated that the exchange rate of DNA within the capsid is identical with that of free DNA, indicating that the capsid lattice does not present a barrier to $^2H_2O$.[30] To initiate exchange, the empty procapsid shells or the mature capsids were tenfold diluted into $^2H_2O$ and incubated at room temperature. At various times, the exchange reactions were sampled and quenched at low pH (pH 2.5) then flash frozen in liquid nitrogen for subsequent analysis by electrospray ionization-time of flight (ESI-TOF) mass spectrometry. To disrupt the capsids it was necessary to include 8 M urea in the quench reactions. The dissociated monomers were digested with pepsin, the peptic fragments were identified by exact mass matching with Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR MS) and MS/MS sequencing, and the extent of deuteration quantified by ESI-TOF mass spectrometry. Peptic fragments covered ~85% of the coat protein primary sequence (Figure 5(a)).

Typically ESI produces multiply charged ions (in positive ionization mode), and any given peptide displays multiple peaks due to the random incorporation of naturally occurring isotopes (primarily $^{13}C$). Replacement of the amide proton by deuterium causes a 1 Da mass increase and shifts the isotopic distribution to higher $m/z$ (Figure 5(b)). The spectra in Figure 5(b) demonstrate the time-dependent incorporation of deuterium into the peptide spanning residues 129–145 for both the empty procapsid shells and the mature capsids. In the empty procapsid shells this peptide exchanged rapidly whereas in the mature capsid the same peptide exchanged more gradually. This result suggests that peptide 129–145 becomes less exposed as the procapsid undergoes maturation to the capsid.

The mass spectroscopy-based approach provides exchange data for peptides distributed throughout the entire protein. To obtain a model independent estimate of exchange rate, we integrated the area under the progress curves from zero time to 32 hours and calculated the ratio of the areas for the empty procapsid shells and the mature capsids profiles (Figure 6, inset). This index provides a model independent estimate of the overall exchange, for example zero exchange would lead to an integrated area value of zero, whereas immediate exchange would lead to the largest integrated area value.

The peptides derived from the N-terminal domain (residues 1–145) exchanged more rapidly in the empty procapsid shells than in the mature capsids, whereas the peptides of the C-terminal domain (residues 207–274, 295–391 and 414–429) exchanged at almost the same rate, and were quite protected in both capsid forms (Figure 6). Thus, the C-terminal domain seems to form the core of the

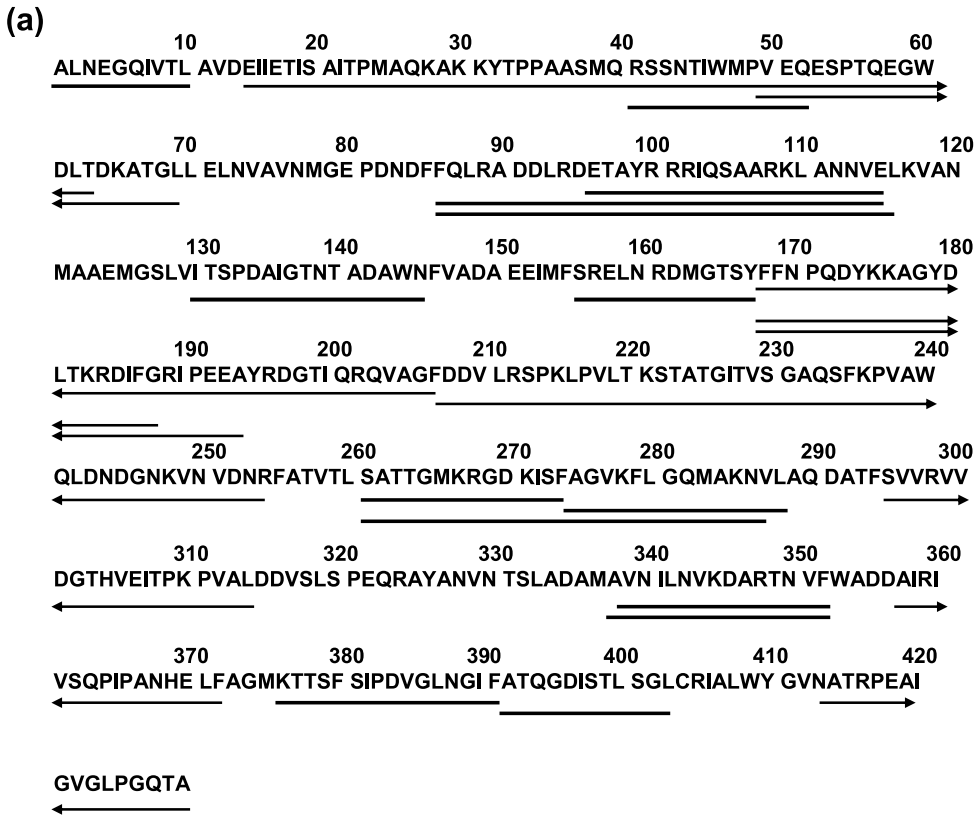


**Figure 5** (*legend opposite*)

structure and only slightly change conformation as maturation occurs, whereas in the procapsid form the N-terminal domain partially retains the flexibility seen in solution and undergoes additional stabilization upon maturation. The global increase in protection in the N-terminal domain might arise from the creation of extensive interactions between neighboring subunits in the mature capsids as seen in the cryoEM-based reconstructions. The loop region (156–207) showed large differences between the empty procapsid shells and the mature capsids (Figure 6), suggesting the loop region undergoes a dramatic rearrangement resulting in new contacts and enhanced protection upon maturation.

To obtain a quantitative estimate of the rate of exchange, the amount of deuterium incorporated at each time point was determined by calculating centroids of the isotopic distribution and plotted against the exchange time (Figure 7). The reproducibility in determining the centroid is typically $>0.1$ Da. The progress curves for each individual peptide were fit to a three component exponential model by non-linear least squares method (Figure 7). The average back exchange was estimated as 30% based on calibration studies with fully deuterated peptides. In this model, the total number of exchangeable hydrogen atoms, $N$, is divided into three groups, $a$ (fast), $b$ (intermediate) and $c$ (slow), with exchange rate constants $k_1(>1\ \mathrm{min}^{-1})$, $k_2(0.01\ \mathrm{min}^{-1}$–$1\ \mathrm{min}^{-1})$ and $k_3(<0.01\ \mathrm{min}^{-1})$, respectively.

$$D = N - a\exp(-k_1 t) - b\exp(-k_2 t) - c\exp(-k_3 t)$$

The centroids of the peptide 129–145 isotopic distribution peaks were plotted against exchange time (Figure 7(a)).

Many of peptides, for example peptide 295–314, showed quite similar exchange patterns in both the empty procapsid shells and the mature capsids (Figure 7(b)). However, peptides from the loop region showed very different patterns of exchange between the empty procapsid shells and the mature capsids (Figure 7(c) and (d)). The loop region peptide 168–207 exchanged very quickly in the empty procapsid shells (~80% of exchangeable amino acid residues exchanged within three minutes), but was well protected from exchange in the mature capsids (~70% of exchangeable amino acid residues remained protected for several days) (Figure 7(c)). This result suggests that the loop region (peptide 168–207) in the empty procapsid shells is solvent exposed, but becomes inaccessible in the mature capsid. The adjacent peptide 156–167 also showed significantly different exchange patterns between the empty procapsid shells and mature capsids (Figure 7(d)). The peptide 156–167 in the empty procapsid shells exchanged gradually whereas in the mature capsids it was also well protected. The differences in the exchange patterns in the loop region imply that the loop region is bent or rearranged and makes new contacts during the maturation procedure.

The cryoEM-based reconstructions suggested

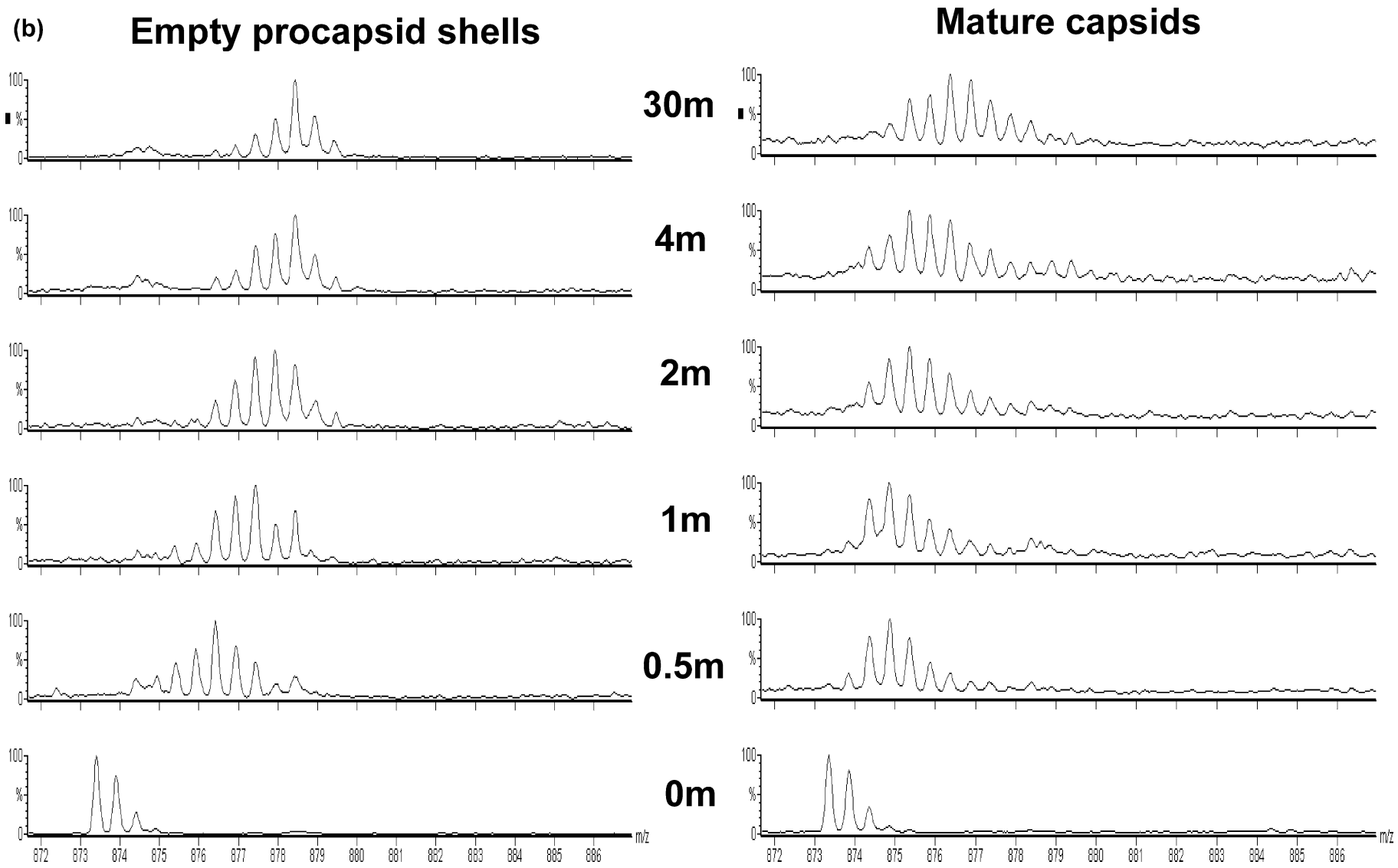


**Figure 5.** Amino acid sequences of P22 capsid coat protein and H/$^2$H exchange profile for peptide residues 126–145. (a) Amino acid sequence of P22 coat protein. The observed peptic fragments of P22 capsid coat protein are underlined. The peptic fragments, which were identified by a combination of exact mass matching and MS/MS sequencing, span ~85% of the protein primary sequence. (b) Mass spectra of the peptic fragment corresponding to residues 126–145 from the empty procapsid shells (left panel) and the mature capsids (right panel) at various exchange times. In each panel, the bottom spectrum represents the un-exchanged control.

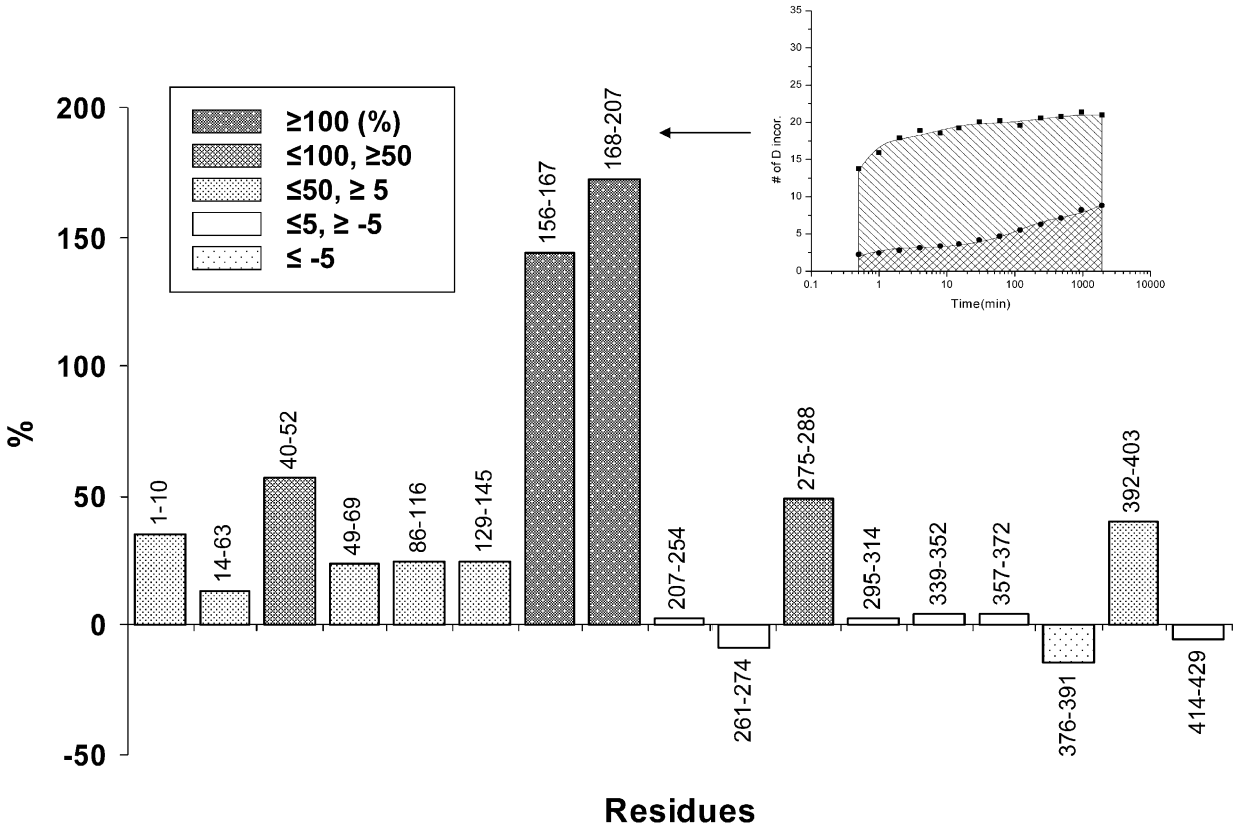


**Figure 6.** Relative exchange comparison between the empty procapsid shells and the mature capsids. The area under the exchange progress curves between zero hours to 32 hours for each peptide was integrated (and shown in inset) and the ratio of the areas for the empty procapsid shells to the mature capsids was calculated and plotted.

that mature capsids have more extensive inter-subunit interactions than the procapsids, but the series of rearrangements in which new contacts are created may cause a slight local exposure of the other sites after maturation. There was only one peptide, the peptide 376–391, which exchanged more rapidly in the mature capsids than in the procapsid (Figure 7(e)).

## Discussion

The proteins that comprise viral capsids have to truly be protean. They need sufficiently conformational flexibility to meet the demands of quasi-symmetry in the assembled capsid. They need to be capable of assembling a stable capsid without error, and in some cases the capsid need to be capable of both assembly and disassembly. Many viral capsids can undergo structural transformations in response to changes in environment such as pH or ionic strength.[21,31] Frequently the infectious viral capsid is assembled in a series of steps involving structural transformations.[32–35] In the case of the dsDNA containing phage, the capsid undergoes a pronounced conformational change accompanying DNA packaging.[3] This suggests that the subunits have sufficient structure for self-assembly but retain sufficient flexibility to afford concerted structural transformations.

Three lines of evidence suggest the presence of multiple domains in P22 coat protein. Two distinct transitions have been observed by fluorescence during the pressure-induced unfolding of the coat protein,[36] renaturation of the coat protein monomer from GuHCl indicated the presence of two relatively stable domains in the coat protein monomer,[37] and limited proteolysis studies have indicated that the coat protein monomer consists of two approximately equal domains connected by a flexible loop region.[12] We initiated biochemical studies to investigate properties of the individual domains and relate them to the stability of the coat protein in the procapsid and capsid lattices. The N-terminal domain is largely unstructured in solution, while the C-terminal domain forms stable non-covalent dimers. The N-terminal domain does adopt additional structure in the context of the C-terminal domain suggesting the C-terminal domain forms a platform on which the N-terminal domain can fold.

Hydrogen/deuterium exchange studies can provide information about protein structure and stability. Previous time-resolved Raman spectroscopy monitoring of hydrogen/deuterium exchange in procapsid and capsid lattices likened

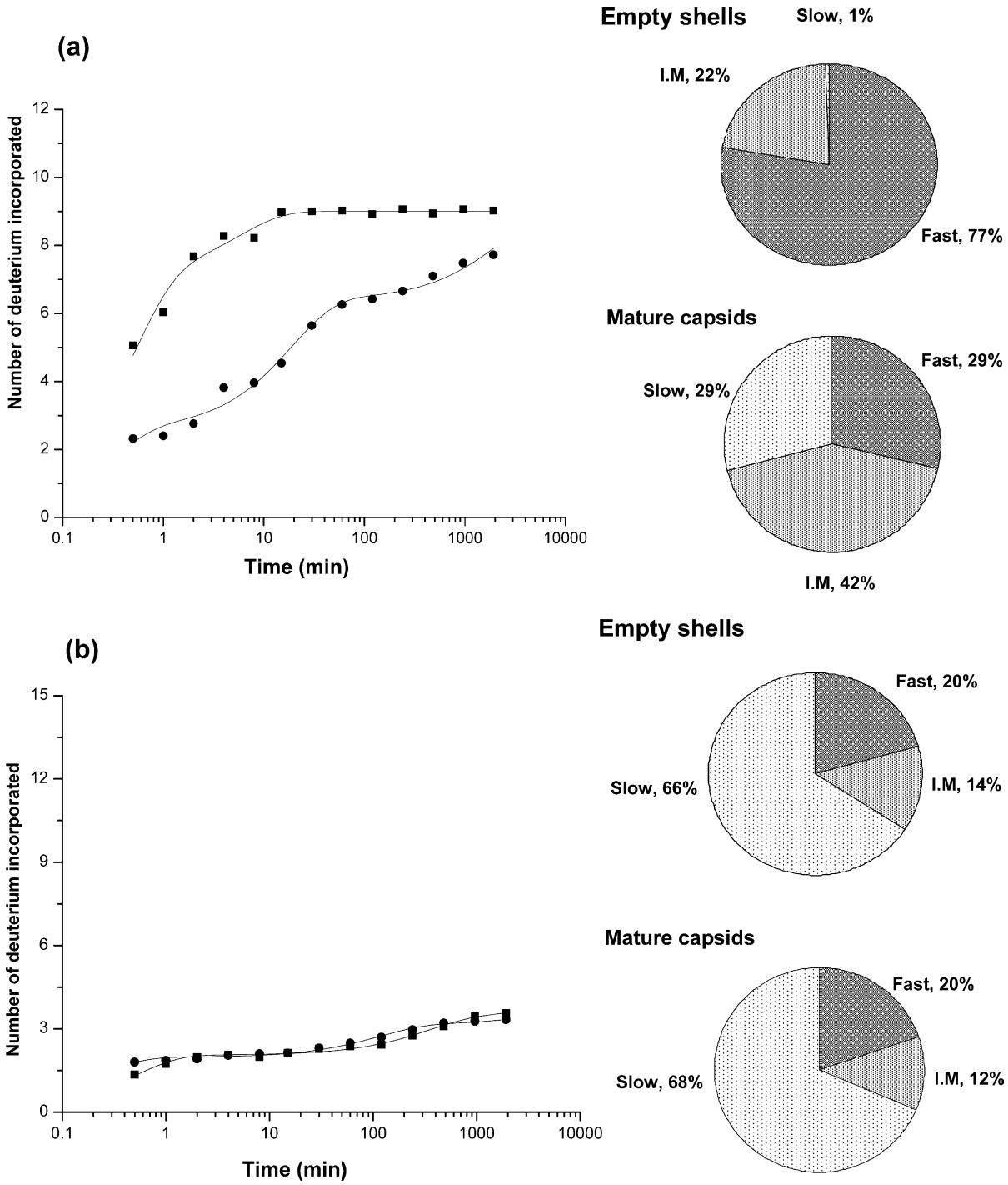


**Figure 7** (*legend follows part (e) of this figure*)

the procapsid form to a late folding intermediate and also demonstrated an increase in exchange protection during capsid maturation.[11] However, these studies were unable to localize the changes to a particular region within the sequence. The use of enzymatic digestion and mass spectrometry to analyze hydrogen/deuterium exchange experiments allows region-specific information to be obtained even on large supramolecular structures.[19–23] Hydrogen/deuterium exchange studies of P22 capsids have been performed in our laboratory using MALDI-MS.[19] However, the need to use denaturant to disrupt capsids resulted in data of limited signal to noise, and limited sequence coverage. The use of ESI-TOF/MS and reverse phase column-based desalting allowed us to increase the coverage of the coat protein sequence in the peptic fragments to ~85% and extend hydrogen/deuterium exchange experiments.

These studies revealed that the increase in exchange protection first detected by Raman spectroscopy was localized primarily in the N-terminal

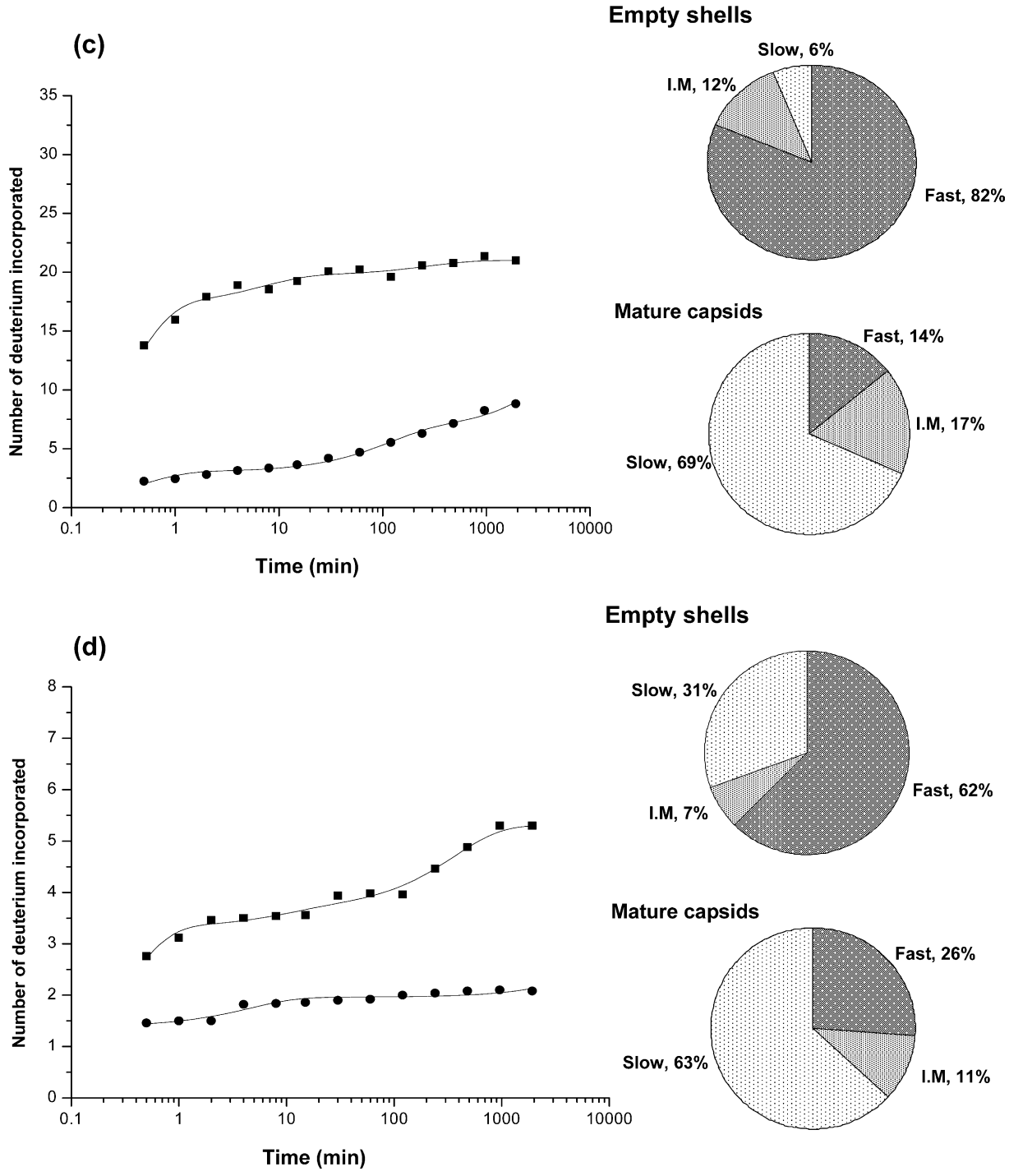


**Figure 7** (*legend opposite*)

domain of the coat protein and suggested that the flexibility observed in solution persisted into the procapsid state but was subsequently lost upon maturation. In contrast there seems to be little change in the structure or stability of the C-terminal region upon maturation. CryoEM-based reconstructions of procapsids and mature capsids at ~9 Å have revealed the existence of more extensive interactions among subunits in the mature capsids than in the procapsid, and have characterized the rearrangement of structural elements *via* rigid body type twists and movements.[13] The exchange data suggest that those types of rearrangements and movements might mainly occur in the N-terminal domain.

It is possible that the N-terminal domain plays an important role in solution as well. To assemble properly the coat protein needs to form stable intersubunit interactions but these interactions need to occur in a controlled fashion, facilitated by the presence of scaffolding protein. Kinetic studies of assembly *in vitro* have shown that the assembly

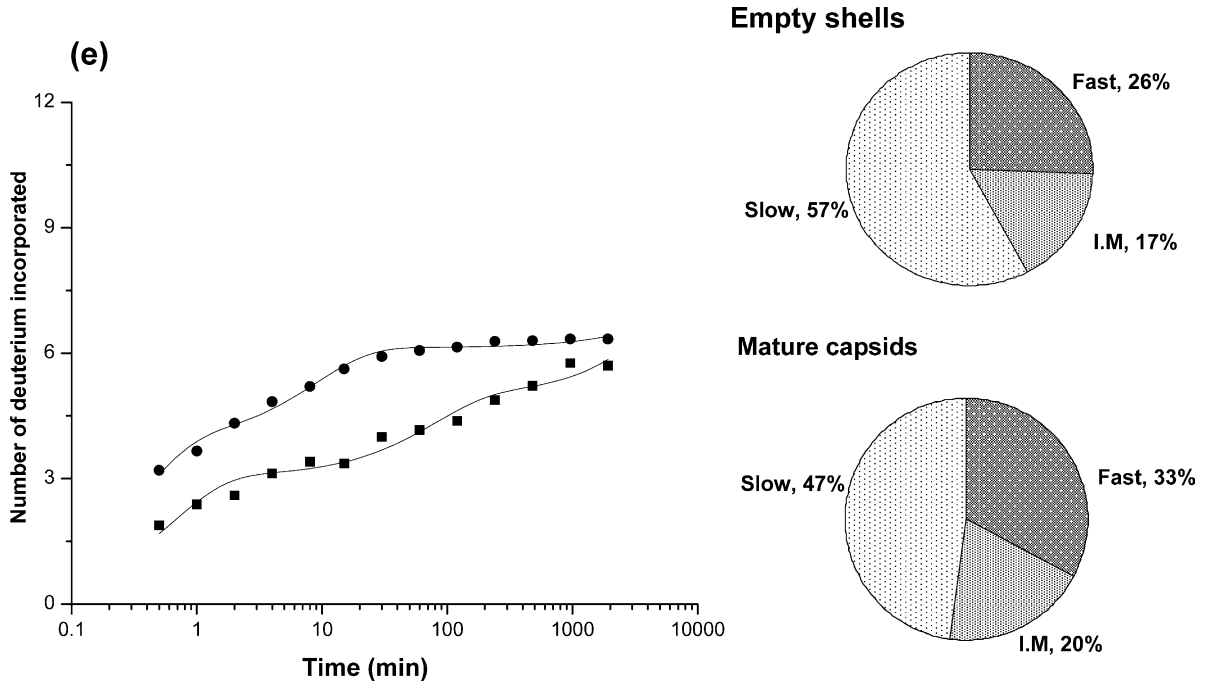


**Figure 7.** Plots of deuterium incorporation at different time points. Plots (left panel) of the number of deuterium atoms incorporated *versus* the exchange period for peptic fragments from the empty procapsid shells (square) and the mature capsids (circle). The continuous line represents the fit obtained by multi-exponentially fitting the exchange data. The pie charts (right panel) represent the relative contributions of the three components fast, intermediate and slow with exchange rate constants $k_1$ ($>1$ min$^{-1}$), $k_2$ (0.01 min$^{-1}$–1 min$^{-1}$) and $k_3$ ($<0.01$ min$^{-1}$), respectively. Peptic fragments (a) 129–145, (b) 295–314, (c) 168–207, (d) 156–167, and (e) 376–391.

active form of the coat protein is a monomer.[5,38] However, all of the temperature-sensitive coat protein mutants examined form small stable oligomers, typically dimers or trimers that are unable to assemble. It is possible that these mutations arise from inappropriate interactions between the C-terminal domains of partially folded species. Of the 18 reported temperature-sensitive (*ts*) mutations in the coat protein[39] 14 lie within the C-terminal domain. Given that the N-terminal domain remains unstructured in solution in the absence of the C-terminal domain one possibility is that the C-terminal domain acts as a template surface to facilitate folding and a properly folded N-terminal domain is required to mask C-terminal domain interacting interfaces. Scaffolding protein would enable the interactions and the conformational changes in the N-terminal domain accompanying expansion would "cement" the lattice in place. This general model is consistent with the observation that the bulk of the stability for the folded form of the coat protein is derived from inter-subunit contacts.

The hydrogen/deuterium exchange profiles of the loop region changed dramatically (Figures 6 and 7(c) and (d)) in going from the empty capsid shells to mature capsids. It exchanged rapidly in the empty procapsid shells, but was well protected in the mature capsids. This result implies that the loop region is exposed to solvent in the empty procapsid shells, but buried in the mature capsids. These data are consistent with the proteolysis studies that the loop region which is susceptible to protease in the procapsids becomes completely resistant.[12] It had previously been observed that cleavage of the loop region facilitated expansion. This suggests that despite the increased protection observed in the loop region upon maturation and the fact that the expanded form is more thermodynamically stable than the procapsid form restructuring the loop region is an energetically unfavorable event. The driving force required to surmount the high activation energy barrier is provided *in vivo* by ATP-driven DNA packaging[3] and *in vitro* by heat or chemical agents such as SDS and urea.[11,40]

A similar transformation has been observed for the bacteriophage HK97. Although there is low sequence similarity ($<20\%$) between the coat proteins of HK97 and P22, the general folding of two coat proteins is quite similar.[13] In the cryoEM studies of HK97, the E-loop of HK97 which protrudes outside of the capsid in the prohead II, becomes bent and K169 in the E-loop knob cross-links to N356 of a neighboring subunit in the head II.[41,42] In addition to rigid body twisting and movement, the bending of E-loop is a major difference in the transformation from prohead II to head II. Interestingly, K166 which is in the E-loop is susceptible to trypsin in the prohead II, but not in head II.[43] Limited proteolysis and hydrogen/deuterium exchange experiments on the P22 capsids suggest that residues 157–207, the loop region participates in the maturation process similar to the E-loop of HK97. Thus, the loop region of P22 may

protrude outside of the procapsids like that of HK97 in the prohead II, and bend to create new inter-subunit contacts in forming stable mature capsids.

## Materials and Methods

### Cloning and purification of the individual domains

The plasmids encoding N and C-terminal regions of the P22 coat protein were constructed by polymerase chain reaction subcloning from a pET-3a plasmid which encoded the genes for coat, scaffolding and portal proteins. The primer pairs used were CGGGTAGCATATGGCTTTGAACGAAGGTCA and CGCGGATCCTTAAATACGCCCGAAGATGTC for the 570 base-pair N-terminal domain and GGAATCCATATGCCTGAAGAAGCATACCGAG and CGCGGATCCTTACGCAGTCTGACCAGGC (NdeI and BamHI sites underlined) for the 717 base-pair C-terminal domain. These amplified regions were digested with NdeI and BamHI, ligated into the pET-3a vector, used to transform $CaCl_2$-treated competent *E. coli* strain BL21 and selected for ampicillin resistance and screened for the appropriate insert size by restriction digestion. Candidate colonies were tested for expression and verified by DNA sequencing. The individual domains were over-expressed in *E. coli* grown in LB medium by induction with 2 mM IPTG. Cells were lysed by repeated freeze–thaw cycles, treated with lysozyme and DNase I, and the pellets were collected by centrifugation at 12,000***g*** for 45 minutes. The pellets were resuspended with washing buffer (100 mM Tris–HCl, 5 mM EDTA, 5 mM DTT, 2 M urea, 2% (v/v) Triton X-100, pH 7.0), and centrifuged at 26,000***g*** for 30 minutes. This procedure was repeated twice. To remove detergent and denaturant, the pellets were washed twice with washing buffer without Triton X-100 and urea. The individual domains were solubilized with extracting buffer (50 mM Tris–HCl, 5 mM EDTA, 5 mM DTT, 8 M GuHCl), and clarified by centrifugation at 100,000***g*** for one hour and the supernatant was reserved. The extracted N and C-terminal domains were refolded by dialysis against buffer B (50 mM Tris–HCl, 25 mM NaCl, 2 mM EDTA) at 4 °C.

### Size exclusion chromatography and light scattering

The molecular mass determination of the refolded N and C-terminal domains was performed by size exclusion chromatography in combination with light scattering. The refolded N and C-terminal domains (0.4 mg/ml each) were loaded onto a 10 mm × 300 mm Superdex 75 (Amersham Bioscience) size exclusion column and eluted with buffer B at a rate of 0.5 ml/minute. A PDI2000DLS light scattering detector (Precision Detectors Inc.) equipped with Water molecules 2410 refractometer was used to determine the molecular mass. Simultaneous measurement of refractive index and light scattering allowed for molecular mass determination of non-covalent complexes. Molecular mass and mass distributions were established using the software which is provided by the manufacturer (Discovery32).

### Fluorescence spectroscopy

Fluorescence spectra were obtained in buffer B at 25 °C using an ISS PC1 photon counting spectrofluorometer ($\lambda_{exc}$ = 280 nm) in 1 cm path length cells. Equilibrium denaturation experiment was performed by supplementing the buffer B with increasing concentrations of GuHCl while maintaining a final protein concentration of 5 μM.

### Circular dichroism (CD) spectroscopy

CD spectra were recorded at 20 °C on AVIV 62DS spectrometer in the spectral range 195–250 nm with 2 nm bandwidth. All samples were measured in 25 mM $PO_4$ and 25 mM NaCl (pH 7.6) using 0.1 cm path length cells.

### Preparation of empty procapsid shells and mature capsids

The procapsids and the mature capsids were prepared using *Salmonella typhimurium* strain DB7000 infected with P22 strain $2^{-}am/13^{-}am$ and $4^{-}am/13^{-}am$, respectively, as described[6] and further purified by sucrose gradient centrifugation. The empty procapsid shells were prepared by repeated extraction of scaffolding protein with 0.5 M GuHCl at 4 °C. Purified empty procapsid shells and mature capsids were stored in buffer B at 4 °C.

### Hydrogen/deuterium exchange experiments

The empty procapsid shells or the mature capsids were exchanged by tenfold dilution into $^2H_2O$ to reach 90% final concentration at 21 °C, pH 7.6 and sampled at zero time, 30 seconds, one, two, four, eight, 15 and 30 minutes, one, two, four, eight, 16 and 32 hours. The exchange was quenched by the addition of formic acid to 1% (v/v) and 8 M urea to dissociate empty procapsid shells and mature capsids completely. The quenched samples were immediately flash frozen and stored at −80 °C until analysis. The samples were thawed rapidly, mixed with an equal volume of pepsin (~20 μM finally) and digested for two minutes on the ice. Digested samples (~80 pmol) were loaded directly onto a C4 trap (Michrom BioResources, Inc.) which replaced the loading loop allowing for rapid washing with water to avoid introducing urea into the ESI source. Following washing, the injector was switched to allow flow into the ESI source. The injection valve, C4 trap and tubing were submerged completely in a 0 °C ice bath. The peptides were rapidly eluted with a 5–95% (v/v) acetonitrile gradient (36 μl/minute). Exchange mass analyses were performed on an ESI-TOF mass spectrometer (Micromass LCT). The amount of deuterium incorporated was determined by calculating the centroid of the isotopic distributions. The eluted peptides were identified by exact mass measurement with 9.4 FT-ICR MS (IonSpec) equipped with dual spray system and MS/MS sequencing with Q-TOF API-US (Micromass). All mass-based assignments agreed with the theoretical mass to within 10 ppm. Because pepsin cuts non-specifically, MS/MS sequencing was performed on all peptides with favorable signal to noise. No discrepancies between mass-based and MS/MS assignments were found.

Progress curves for individual peptides were fitted to a sum of three exponentials derived from the exchange rate expression:[14]

$$D = N - a\exp(-k_1 t) - b\exp(-k_2 t) - c\exp(-k_3 t)$$

where $D$ is the observed number of deuterium atoms incorporated at time $t$ and $N$ is the total number of exchangeable protons. $a$, $b$ and $c$ represent the number of amino acid residues that exchange with rate constant $k_1$(fast, $>1.0\ \mathrm{min}^{-1}$), $k_2$(intermediate, $0.01–1.0\ \mathrm{min}^{-1}$) and $k_3$(slow, $<0.01\ \mathrm{min}^{-1}$), respectively.

## Acknowledgements

The authors thank Drs A. Hawkridge and D. Muddiman of the W. M. Keck center for FT-ICR Mass Spectrometry Laboratory, Mayo Clinic for assistance in identifying the peptic fragments. This work was supported by NIH grant GM47980 (P.E.P.).

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.02.021

## References

1. Albert, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, **92**, 291–294.
2. Casper, D. L. (1980). Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* **32**, 103–138.
3. King, J. & Chiu, W. (1997). In *Structural Biology of Viruses* (Chiu, W., Burnett, R. & Carcea, R., eds), pp. 288–351, Oxford University Press, New York.
4. King, J. & Casjens, S. (1974). Catalytic head assembling protein in virus morphogenesis. *Nature*, **251**, 112–119.
5. Casjens, S. & King, J. (1974). P22 morphogenesis I: catalytic scaffolding protein in capsid assembly. *J. Supramol. Struct.* **2**, 202–224.
6. Prevelige, P. E., Jr, Thomas, D. & King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunit into icosahedral shells *in intro. J. Mol. Biol.* **202**, 743–757.
7. Prasad, B. V. V., Prevelige, P. E., Jr, Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993). Three-dimensional transformation of capsid associated with genome packaging in a bacterial virus. *J. Mol. Biol.* **231**, 65–74.
8. Galisteo, M. L. & King, J. (1993). Conformational transformations in the protein lattice of phage P22 procapsids. *Biophys. J.* **65**, 227–235.
9. Prevelige, P. E., Jr, Thomas, D., King, J., Towse, S. A. & Thomas, G. J., Jr (1990). Conformational states of the bacteriophage P22 capsid subunit in relation to self-assembly. *Biochemistry*, **29**, 5626–5633.
10. Prevelige, P. E., Jr, Thomas, D., Aubrey, K. L., Towse, S. A. & Thomas, G. J., Jr (1993). Subunit conformational changes accompanying bacteriophage P22 capsid maturation. *Biochemistry*, **35**, 4619–4627.
11. Tuma, R., Prevelige, P. E., Jr & Thomas, G. J. (1998). Mechanism of capsid maturation in a double-stranded DNA virus. *Proc. Natl Acad. Sci. USA*, **95**, 9885–9890.
12. Lanman, J., Tuma, R. & Prevelige, P. E., Jr (1999). Identification and characterization of the domain structure of bacteriophage P22 coat protein. *Biochemistry*, **38**, 14614–14623.
13. Jiang, W., Li, Z., Zhang, Z., Baker, M. L., Prevelige, P. E., Jr & Chiu, W. (2003). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nature Struct. Biol.* **10**, 131–135.
14. Englander, S. W. & Kallenbach, N. R. (1984). Hydrogen exchange and structural dynamics of protein and nucleic acids. *Quart. Rev. Biophys.* **16**, 521–655.
15. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* **20**, 4–14.
16. Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR. *Nature*, **335**, 700–704.
17. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.
18. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
19. Tuma, R., Coward, L. U., Kirk, M. C., Barnes, S. & Prevelige, P. E., Jr (2001). Hydrogen-deuterium exchange as a probe of folding and assembly in viral capsids. *J. Mol. Biol.* **306**, 389–396.
20. Lanman, J., Lam, T. T., Barnes, S., Sakalian, M., Emmett, M. R., Marshall, A. G. & Prevelige, P. E., Jr (2003). Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectroscopy. *J. Mol. Biol.* **325**, 759–772.
21. Wang, L., Lane, L. C. & Smith, D. L. (2001). Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci.* **10**, 1234–1243.
22. Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522–531.
23. Smith, D. L., Deng, Y. & Zhang, Z. (1997). Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass Spectrosc.* **32**, 135–146.
24. Miranker, A., Robinson, C. V., Radford, S. E., Apli, R. T. & Dobson, C. M. (1993). Detection of transient protein folding population by mass spectrometry. *Science*, **262**, 896–900.
25. Arrington, C. B. & Robertson, A. D. (2000). Correlated motions in native proteins from MS analysis of NH exchange: evidence for a manifold of unfolding reactions in ovomucoid third domain. *J. Mol. Biol.* **300**, 221–232.
26. Deng, Y. & Smith, D. L. (1998). Identification of unfolding domains in large proteins by their unfolding rates. *Biochemistry*, **37**, 6256–6262.
27. Deng, Y. & Smith, D. L. (1999). Rate and equilibrium constants for protein unfolding and refolding determined by hydrogen exchange-mass spectrometry. *Anal. Biochem.* **276**, 150–160.
28. Wang, S. S., Tobler, S. A., Good, T. A. & Fernandez, E. J. (2003). Hydrogen exchange mass spectrometry analysis of β-amyloid peptide structure. *Biochemistry*, **42**, 9507–9514.
29. Lanman, J. & Prevelige, P. E., Jr (2004). High-sensitivity mass spectrometry for imaging subunit interactions: hydrogen/deuterium exchange. *Curr. Opin. Struct. Biol.* **14**, 181–188.
30. Reilly, K. E. & Thomas, G. J., Jr (1994). Hydrogen exchange dynamics of the P22 virion determined by time-resolved Raman spectroscopy. Effects of chromosome packaging on the kinetics of nucleotide exchanges. *J. Mol. Biol.* **241**, 68–82.
31. Lata, R., Conway, J. F., Cheng, N., Duda, R. L.,

Hendrix, R. W., Wikoff, W. R. *et al*. (2000). Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell*, **100**, 253–263.
32. Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R. & Steven, A. C. (1996). The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly. *J. Mol. Biol.* **263**, 447–462.
33. Turner, B. G. & Summers, M. F. (1999). Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32.
34. Canady, M. A., Tihova, M., Hanzlik, T. N., Johnson, J. E. & Yeager, M. (2000). Large conformational changes in the maturation of a simple RNA virus, nudaurelia capensis omega virus (NωV). *J. Mol. Biol.* **299**, 573–584.
35. Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H. & Fuller, S. D. (1997). Intermediates in the assembly pathway of the double-stranded RNA virus φ6. *EMBO J.* **16**, 4477–4487.
36. Prevelige, P. E., Jr, King, J. & Silva, J. L. (1994). Pressure denaturation of the bacteriophage P22 coat protein and its entropic stabilization in icosahedral shells. *Biophys. J.* **66**, 1631–1641.
37. Teschke, C. M. & King, J. (1993). Folding of the phage P22 coat protein *in vitro*. *Biochemistry*, **32**, 10839–10847.
38. Fuller, M. T. & King, J. (1982). Assembly *in vitro* of bacteriophage P22 procapsids from purified coat and scaffolding subunits. *J. Mol. Biol.* **156**, 633–665.
39. Gordon, C. L. & King, J. (1994). Genetic properties of temperature-sensitive folding mutants of the coat protein of phage P22. *Genetics*, **136**, 427–438.
40. Earnshaw, W., Casjens, S. & Harrison, S. C. (1976). Assembly of the head of bacteriophage P22: X-ray diffraction from heads, proheads and related structure. *J. Mol. Biol.* **104**, 387–410.
41. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. & Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*, **289**, 2129–2133.
42. Conway, J. F., Wikoff, W. R., Cheng, N., Duda, R. L., Hendrix, R. W., Johnson, J. E. & Steven, A. C. (2001). Virus maturation involving large subunit rotations and local refolding. *Science*, **292**, 744–748.
43. Hendrix, R. W. & Duda, R. L. (1998). Bacteriophage HK97 head assembly: a protein ballet. *Advan. Virus Res.* **50**, 235–288.

***Edited by M. F. Summers***

*(Received 4 January 2005; received in revised form 4 February 2005; accepted 4 February 2005)*