Backbone Dynamics of a Model Membrane Protein: Assignment of the Carbonyl Carbon ^{13}C NMR Resonances in Detergent-Solubilized M13 Coat Protein†

Gillian D. Henry, Joel H. Weiner, and Brian D. Sykes*

Medical Research Council Group in Protein Structure and Function and Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

Received December 19, 1986

ABSTRACT: The major coat protein of the filamentous bacteriophage M13 is a 50-residue amphiphilic polypeptide which is inserted, as an integral membrane-spanning protein, in the inner membrane of the *Escherichia coli* host during infection. ^{13}C was incorporated biosynthetically into a total of 23 of the peptide carbonyls using labeled amino acids (alanine, glycine, lysine, phenylalanine, and proline). The structure and dynamics of carbonyl-labeled M13 coat protein were monitored by ^{13}C nuclear magnetic resonance (NMR) spectroscopy. Assignment of many resonances was achieved by using protease digestion, pH titration, or labeling of the peptide bond with both ^{13}C and ^{15}N. The carbonyl region of the natural-abundance ^{13}C NMR spectrum of M13 coat protein in sodium dodecyl sulfate solution shows approximately eight backbone carbonyl resonances with line widths much narrower than the rest. Three of these more mobile residues correspond to assigned peaks (glycine-3, lysine-48, and alanine-49) in the individual amino acid spectra, and another almost certainly arises from glutamic acid-2. A ninth residue, alanine-1, also gives rise to a very narrow carbonyl resonance if the pH is well above or below the pK_a of the terminal amino group. These data suggest that only about four residues at either end of the protein experience large-amplitude spatial fluctuations; the rest of the molecule is essentially rigid on the time scale of the overall rotational tumbling of the protein–detergent complex. The relative exposure of different regions of detergent-bound protein was monitored by limited digestion with proteinase K. The N-terminal hydrophilic region was very readily removed by the enzyme, and a stable core particle was isolated which contained the hydrophobic region and most of the C-terminal segment. Comparable spectra and digestion patterns were obtained when the protein was solubilized in sodium deoxycholate, suggesting that the coat protein binds both amphiphiles in a similar fashion.

Nuclear magnetic resonance (NMR)[1] spectroscopy, a powerful technique in the study of protein structure and dynamics, has yet to find widespread application in the area of membrane proteins. Such proteins are generally limited in quantity and insoluble in the absence of lipid or detergent. The major coat protein (gene 8 protein) of the filamentous coliphage M13 is an attractive integral membrane protein by virtue of its structural simplicity (it spans the lipid bilayer only once), small size (50 residues), and ease of preparation. In addition, isotopic labels such as ^{2}H, ^{13}C, ^{15}N, and ^{19}F can be incorporated biosynthetically, thus allowing selected nuclei to be studied by NMR in the absence of a strong background signal. To determine the mobility of the protein backbone, the peptide carbonyls of detergent-solubilized coat protein have been investigated by ^{13}C NMR spectroscopy both in natural abundance and after selective incorporation of ^{13}C-labeled amino acids.

M13 coat protein is an amphiphilic polypeptide (Figure 1) which plays a structural role in the mature phage (Banner et al., 1981) but is inserted as an integral protein in the inner membrane of the host *Escherichia coli* cell during infection (Smilowitz et al., 1972). New coat protein, which is processed by cleavage of a 23-residue leader sequence (Chang et al., 1977; Sugimoto, 1978), also inserts into the inner membrane where it is rapidly dispersed with the original (Smilowitz, 1974). Assembly of progeny phage occurs in the inner membrane without lysis of the host cell. The amino acid sequence of M13 coat protein (Asbeck et al., 1969; Nakashima & Konigsberg, 1974) contains a 19-residue uncharged hydrophobic central core (Figure 1). The N-terminal 20 residues possess a net negative charge, whereas the shorter C-terminal segment is basic. M13 coat protein (or the identical protein from the closely related phages fl or fd) has been the subject of numerous biophysical studies in intact virus, phospholipid dispersions, and detergent micelles [e.g., see Knippers and Hoffmann-Berling (1966), Nozaki et al. (1976, 1978), Williams and Dunker (1977), Opella et al. (1980), Gall et al. (1981), Cross and Opella (1982, 1985), Dettman et al. (1982, 1984), Valentine et al. (1985), Wilson and Dahlquist (1985), and Henry et al. (1986a,b)].

M13 coat protein was labeled biosynthetically with 1-^{13}C-enriched alanine, glycine, lysine, phenylalanine, and proline, a total of 23 residues which represent all 3 domains. The [^{13}C]carbonyl label provides a convenient probe of the backbone, allowing a number of chemically similar sites along the length of the polypeptide chain to be compared. The relaxation behavior of a carbonyl carbon is complex (including both dipole–dipole and chemical shift anisotropy terms); thus,

†Supported by the Medical Research Council of Canada (MRC Group in Protein Structure and Function and MRC Grant MT5838) and the Alberta Heritage Foundation for Medical Research. Paper 1 in the series is by Henry et al. (1986a).

*Address correspondence to this author at the Department of Biochemistry, University of Alberta.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; SDS, sodium dodecyl sulfate; DFP, diisopropyl fluorophosphate; Me_4Si, tetramethylsilane; τ_c, overall rotational correlation time of the protein–detergent complex; ppm, parts per million; DOC, deoxycholate.

0006-2960/87/0426-3619$01.50/0

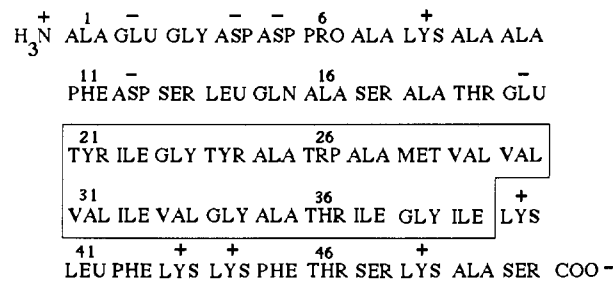

FIGURE 1: Sequence of M13 coat protein according to Asbeck et al. (1969) and Nakashima and Konigsberg (1974). The hydrophobic sequence is indicated.

quantitation of internal motions at a single frequency is difficult. Nevertheless, the line widths of carbonyl resonances in M13 coat protein provide a good qualitative picture of backbone mobility as line width is a function of the effective correlation time of a given nucleus. The carbonyl/carboxyl region of the natural-abundance ^{13}C NMR spectrum of sodium dodecyl sulfate (SDS)-solubilized coat protein contains approximately eight very narrow resonances attributable to backbone carbonyls. These nuclei are undergoing additional motions on a faster time scale than the overall tumbling of the protein–detergent micelle. Assignment of ^{13}C spectra obtained from protein specifically labeled at individual carbonyl sites has enabled the more mobile regions of the protein to be defined; in keeping with our previous model of the protein in both SDS and deoxycholate micelles (Henry et al., 1986a), all of the narrow resonances appear to arise from residues at the termini. The assignment of individual carbonyl resonances also forms a basis for the interpretation of hydrogen exchange data [see Henry et al. (1987)] obtained indirectly from the H/D isotope shift of the ^{13}C NMR resonances of the carbonyl carbons. The relative exposure of different regions of SDS- and DOC-bound coat protein was monitored by digestion with proteinase K. Comparable spectra and digestion patterns are obtained in both detergents, suggesting that the coat protein associates with both SDS and deoxycholate micelles in a similar way.

Experimental Procedures

Materials

Escherichia coli strains KA197 (CGSC 5243, Hfr, *thi* 1, *phe* A27, *rel* A1, λ⁻), G11a1 (CGSC 5168, Hfr, *rel* A1, *ilv*-229, *met* B1, *amp* A1), and AT2457 (CGSC 4507 Hfr, *thi* 1, *gly* A6, *rel* A1, λ⁻, *spo* T1) were obtained from Dr. B. Bachmann, Coli Genetic Stock Centre, Yale University School of Medicine. DL-[1-^{13}C]Alanine (99%), DL-[1-^{13}C]proline (90%), DL-[1-^{13}C]phenylalanine (99%), DL-[1-^{13}C]lysine (99%), [1-^{13}C]glycine (99%), and [^{15}N]leucine (99%) were from MSD Isotopes (Pointe Claire, Dorval, Quebec). Sodium dodecyl sulfate (electrophoretic grade) was obtained from Bio-Rad (Richmond, CA) or ICN Biomedicals Canada Ltd. (Montreal, Quebec), and sodium deoxycholate was from Sigma Chemical Co. (St. Louis, MO). The enzymes carboxypeptidase A (DFP treated) and proteinase K (protease type XI) were purchased from Sigma.

Methods

Growth of Labeled Phage. M13 was labeled with phenylalanine, lysine, and proline as described in Henry et al. (1986) using *E. coli* strain KA197. Alanine-labeled protein was prepared similarly except that a 0.1 g L^{-1} sample of each of the other 19 amino acids was included in the culture medium. Glycine was incorporated by using the auxotrophic strain AT2457, and 0.05 g L^{-1} each of uracil, thymine, and cytosine and 0.1 g L^{-1} adenine and guanosine (the base, guanine, is insoluble in the growth medium) were also included to suppress purine biosynthesis, which requires glycine. Doubly labeled [1-^{13}C]lysine/[^{15}N]leucine phage was prepared by using strain G11a1 in the presence of 0.08 g L^{-1} DL-[1-^{13}C]-lysine, 0.04 g L^{-1} each of L-[^{15}N]leucine and L-methionine, and 0.1 g L^{-1} each of isoleucine and valine.

Preparation of Samples and NMR Spectroscopy. Coat protein containing detergent micelles (deoxycholate or SDS) were prepared as described by Henry et al. (1986a) and concentrated by ultrafiltration. D_2O was added to a final concentration of 5–10%. Protein concentrations, determined by the absorbance at 280 nm using an $A^{1\%}_{280nm}$ value of 16.6, were typically between 0.7 and 1.0 mM (except where indicated).

^{13}C NMR spectra were recorded at 75.45 MHz at room temperature (23 °C) on a Nicolet NT 300 WB spectrometer using 3.5 cm^3 of sample contained within a 12-mm tube. Protons were decoupled by using the MLEV16 decoupling sequence (Levitt & Freeman, 1981). Chemical shifts are quoted relative to Me_4Si and measured with respect to internal dioxane at 67.37 ppm (Shindo et al., 1978).

Enzyme Digestions. Proteinase K digestions were performed at pH 9.0 by using an initial enzyme:protein ratio of 1:850 (w/w). The enzyme was dissolved in water and the concentration determined by the absorbance at 280 nm using an $A^{1\%}_{280nm}$ value of 14.2 (Ebeling et al., 1974); solutions were used immediately. More enzyme was added, when necessary, as described in the text. Carboxypeptidase A digestions were carried out at pH 8.0 using an enzyme:protein ratio of approximately 1:30 (w/w).

Results

M13 coat protein, like most membrane proteins, forms high molecular weight aggregates in aqueous solution in the absence of lipid or detergent (Knippers & Hoffmann-Berling, 1966; Cavalieri et al., 1976; Nozaki et al., 1978). Detergent-solubilized protein was used in these experiments since the protein–detergent micelles have a much shorter overall rotational correlation time (τ_c) than lipid vesicles and the line widths are correspondingly narrower. Both SDS and deoxycholate have been used by ourselves and others to solubilize the coat protein (Hagen et al., 1978; Cross & Opella, 1979, 1980; Wilson & Dahlquist, 1985), and its structure (from circular dichroism and NMR experiments) is thought to be similar in both detergents (Nozaki et al., 1976; Henry et al., 1986a). However, SDS was emphasized in this series of experiments because the ^{13}C NMR line widths of the carbonyl carbons correspond to the values predicted from the known overall rotational correlation time τ_c; many resonances in DOC-bound coat protein appear unusually broad, possibly due to conformational heterogeneity. Coat protein exists as a dimer in SDS (and deoxycholate) solutions (Makino et al., 1975). Under the conditions of these experiments, about 60 molecules of SDS are bound per dimer (Makino et al., 1975), giving it an effective molecular weight of $\simeq$27 000. Assuming isotropic tumbling, τ_c was previously determined to be approximately 1.1×10^{-8} s (Henry et al., 1986a).

Carbonyl and carboxyl ^{13}C NMR resonances are clustered together in a region of about 12 ppm at the low-field region (171–183 ppm from Me_4Si) of the spectrum. Although all nonprotonated carbons are expected to have narrow line widths, Figure 2a shows the natural-abundance carbonyl/carboxyl spectrum of M13 coat protein to display approximately 13 very narrow resonance superimposed upon a broader

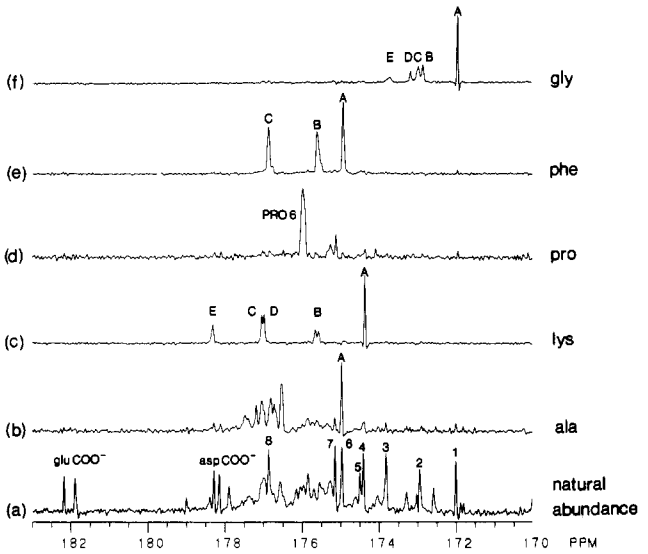

FIGURE 2: Low-field (carbonyl/carboxyl) region of the 75.45-MHz ^{13}C NMR spectrum of M13 coat protein in SDS micelles, pH 9.0. 32K data points were collected with a sweep width of ±8064 Hz. The carrier frequency was placed at 98 ppm. Resolution was enhanced by weighting the free induction decay by multiplication with a double-exponential function. (a) Natural-abundance spectrum, 80 000 scans, 5.5 mM protein. The glutamate and aspartate carboxyl peaks are tentatively assigned, and the remaining narrow resonances are labeled 1–8. (b) [1-^{13}C]Alanine label, 45 000 scans, 1.0 mM protein. (c) [1-^{13}C]Lysine label, 33 000 scans, 0.7 mM protein. (d) [1-^{13}C]Proline label, 33 000 scans, 0.9 mM protein. (e) [1-^{13}C]-Phenylalanine label, 32 000 scans, 0.7 mM protein. (f) [1-^{13}C]Glycine label, 30 000 scans, 0.8 mM protein.

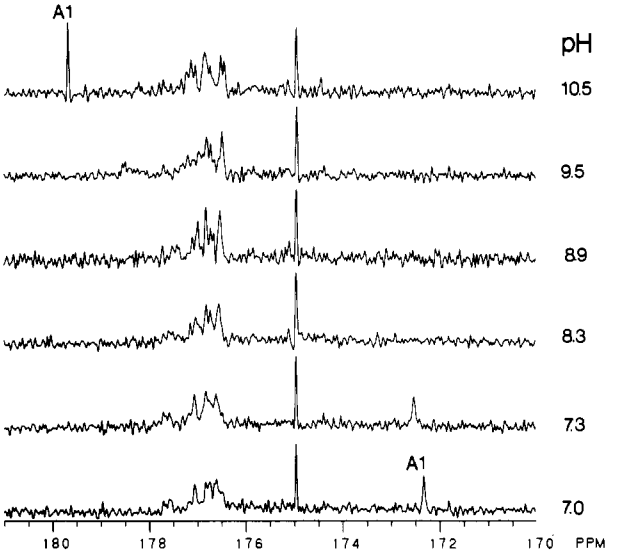

FIGURE 3: Assignment of alanine-1; pH titration of [1-^{13}C]alanine-labeled coat protein in SDS micelles. 10 000 scans were averaged in each spectrum, conditions otherwise as in Figure 2. The pK_a of the terminal amino group is 8.8.

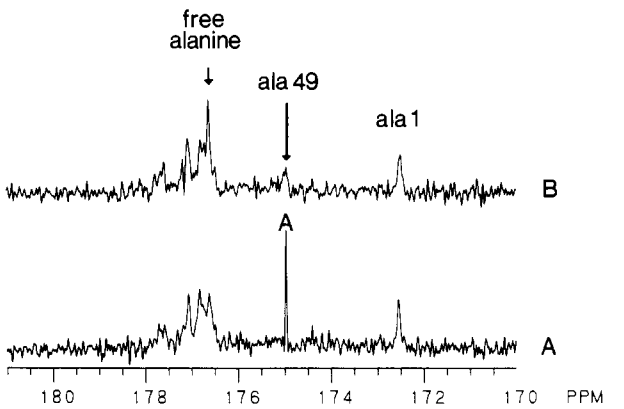

FIGURE 4: Assignment of alanine-49 (peak A) by carboxypeptidase A digestion. (A) [1-^{13}C]Alanine-labeled coat protein in SDS micelles, pH 7.3. (B) After ≃3-h digestion with 1:30 (w/w) carboxypeptidase A.

envelope. This means that several carbon atoms are experiencing significantly more motional freedom than the rest. Likely candidates for the mobile nuclei include the two glutamic acid and three aspartic acid side chain carboxyls. Carboxyl carbons tend to resonate downfield of the carbonyls, with glutamyl residues generally well downfield of aspartyl residues (Keim et al., 1973a,b; Shindo et al., 1978). The two resonances at ≃182 ppm and the three at ≃178 ppm have therefore been assigned tentatively to glutamic and aspartic acids, respectively (Figure 2). This leaves a total of eight resonances (numbered 1–8 in Figure 2a), one of which may belong to glutamine-15 but the rest must originate from the protein backbone.

Individual spectra corresponding to coat protein labeled biosynthetically at the carbonyl carbon with alanine (10 residues), lysine (5 residues), proline (1 residue), phenylalanine (3 residues), and glycine (4 residues) are shown in Figure 2b–f. The resonances in each spectrum are labeled A, B, C, etc. for assignment purposes and scaled relative to the tallest peak in each. Peak heights between spectra are therefore not directly comparable. Three peaks are seen in the five labeled proteins (alanine A, lysine A, and glycine A) which clearly correspond to the narrow resonances of the natural-abundance spectrum (peaks 6, 4, and 1, respectively).

Assignments. (*a*) *Alanine.* There are 10 alanine residues in M13 coat protein; 6 of these occur in the N-terminal hydrophilic region, 3 in the hydrophobic core, and the tenth (alanine-49) is the penultimate residue (Figure 1). The alanine label did not incorporate very well, possibly due to decarboxylation during growth, and natural-abundance peaks are apparent just above base line (Figure 2b). An additional peak, which is seen at either the high-field or the low-field end of the carbonyl region (depending on pH), is observed to be very narrow at pH values <6.5 or >10.5 (Figure 3). This is alanine-1 which titrates over a range of almost 8 ppm according to the protonation state of the N-terminal amino group (pK_a = 8.8). At pH values between 6.5 and 10.5, alanine-1 is subject to exchange broadening; for example, at pH 9.0 in Figure 2b, it is almost invisible. The rest of the spectrum is comparatively pH insensitive. Thus, alanine-1 is an additional mobile carbonyl which is not apparent in Figure 2 (a or b). Peak A, the narrow upfield resonance of Figure 2b, is readily assigned to alanine-49 by carboxypeptidase A digestion (Figure 4). Carboxypeptidase A removes serine-50 and alanine-49 sequentially from the C-terminus but does not cleave the basic residue, lysine-48. A new resonance, corresponding to free alanine, is observed at 176.8 ppm. The other eight alanine residues are clustered together around 177 ppm. They are not well resolved and have not been assigned individually. Nevertheless, the line widths suggest that the rest of the alanines, notably alanines-7, -9, and -10, are much more motionally restricted than alanine-1 or alanine-49.

(*b*) *Proline.* The single proline residue, proline-6 (Figure 2d), is clearly not one of the narrow resonances of the natural-abundance spectrum. It is actually composed of two closely overlapping resonances [shown more clearly in Figure 4 of Henry et al. (1987)]. This "doublet" character is not readily explained, but a single Lorentzian line is observed above 35 °C, below pH 4 and above pH 10. The phenomenon of

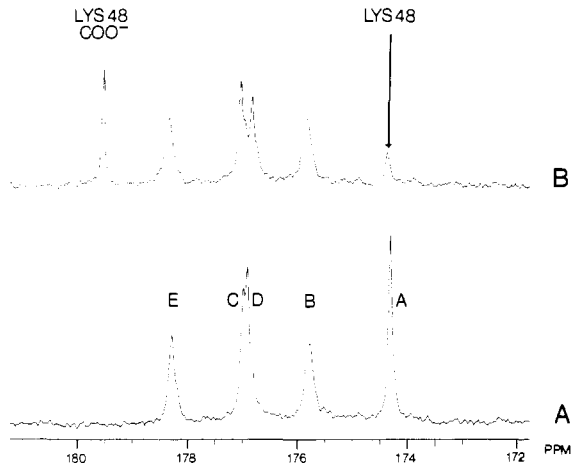

FIGURE 5: Assignment of lysine-48 (peak A) by carboxypeptidase A digestion. (A) [1-^{13}C]lysine-labeled coat protein in SDS micelles, pH 7.5, with 18% D_2O. (B) After ≃3-h digestion with 1:30 (w/w) carboxypeptidase A. The newly exposed lysine-48 carboxyl is the most downfield peak in spectrum (B). The upfield "shoulder" on peaks C, D, and E is an isotope effect due to slowly exchanging deuterium on the adjacent amide nitrogen [see Henry et al. (1987)]. Coat protein was 0.7 mM, and 10000 scans were averaged. Line broadening is 2 Hz.

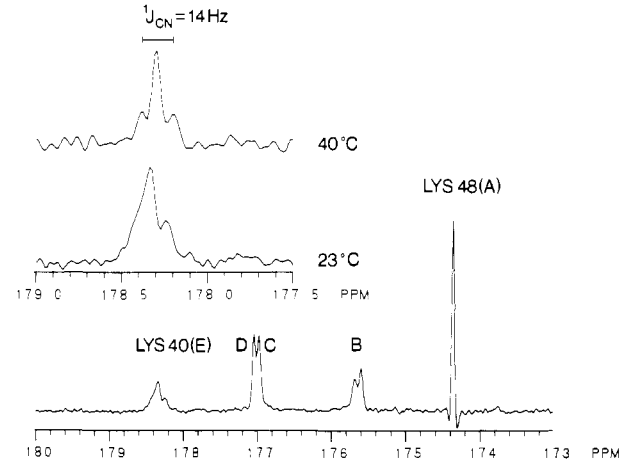

FIGURE 6: Assignment of lysine-40 in [1-^{13}C]lysine/[^{15}N]leucine-labeled coat protein. The lysine-40/leucine-41 peptide carbonyl ($^{13}CO^{15}NH$) is *J*-coupled to ^{15}N, resulting in a doublet ($^1J_{CN}$ = 14 Hz) which superimposes the residual $^{13}CO^{14}NH$ singlet of peak E (lysine-40).

"double peaks" is not restricted to proline but is observed under similar conditions for the carbonyl carbons of lysine-8, phenylalanine-45 (see later), leucine-14, leucine-41, tyrosine-21, and tyrosine-24 (data not shown) and may be related to the dimeric nature of the coat protein (G. D. Henry, J. H. Weiner, and B. D. Sykes, unpublished results).

E. coli cells readily catabolize excess proline to form glutamic acid (Frank & Ranhand, 1964), and this is the likely origin of the additional peaks observed about 1 ppm upfield of the proline resonance. This would also account for the relatively poor signal to noise ratios obtained with proline-labeled protein. The chemical shift of the narrow upfield peak titrates over a pH range of 3–5, and its shift at pH 9 corresponds exactly to that of peak 7 of the natural-abundance spectrum (Figure 2a). As alanine-1 has been shown to be mobile (as is glycine-3; see later), this resonance almost certainly arises from the backbone carbonyl of glutamic acid-2.

(*c*) *Lysine.* M13 coat protein possesses five lysine residues, one in the N-terminal region (lysine-8) and a cluster of four at the C-terminus which is probably involved in binding DNA. As with alanine-49, carboxypeptidase A digestion revealed the sharp upfield peak A to be lysine-48 (Figure 5). The newly exposed lysine carboxyl resonates at 179.5 ppm, downfield of the rest of the spectrum.

Peak E (Figure 2c) was assigned to lysine-40 by double labeling of the peptide bond lysine-40–leucine-41 with ^{13}C (carbonyl) and ^{15}N (amide). Directly bonded ^{15}N (spin $^1/_2$) splits the carbonyl resonance into a doublet, whereas ^{14}N (spin 1)–^{13}C couplings are not observed because of the rapid quadrupolar relaxation of ^{14}N. As ^{15}N-labeled amino acids do not incorporate readily into proteins unless extensive precautions are taken against deamination (Griffey et al., 1985), a significant number of ^{13}C–^{14}N bonds still remain. Nevertheless, a doublet ($^1J_{CN}$ = 14 Hz) is seen to superimpose peak E in Figure 6. The ^{13}C-^{15}N coupling is more apparent at 40 °C than at room temperature (see inset). A small isotope shift on ^{13}C by the bonded nitrogen nucleus results in asymmetry of the doublet about the ^{14}N-bonded singlet [see Hansen (1983)].

The remainder of the lysine resonances were assigned by limited digestion with proteinase K. This enzyme, which remains active in SDS solutions, is a broad-specificity protease which displays a preference for peptide bonds on the C-terminal side of hydrophobic residues (Ebeling et al., 1974). It has been shown by using [3-^{13}C]alanine-labeled coat protein that the hydrophilic regions (i.e., including alanine-18 and alanine-49) may be readily removed by limited treatment with proteinase K, leaving a core particle in which the resonance of the three alanine methyl carbons of the hydrophobic core (residues 25, 27, and 35) are unaffected (Henry et al., 1986a). With the use of this alanine methyl label as a marker for the hydrophobic core (see Figure 7A), a mixture of [3-^{13}C]alanine- and [1-^{13}C]lysine-labeled protein was digested with limited quantities of proteinase K. A time course of the digestion is shown for the methyl and carbonyl regions in panels A and B, respectively, of Figure 7. Within the time taken to collect the first spectrum (1.41 h), peak B has disappeared to be replaced by a single peptide resonance at 174.2 ppm. The rest of the spectrum is unaltered. By comparison, large changes occur in the alanine spectrum, where most of the resonances arise from the N-terminal region (alanines-1, -7, -9, -10, -16, and -18). This suggests that resonance B is probably lysine-8. Further digestion results in the slow disappearance of the lysine-48 resonance and the appearance of a downfield peak, corresponding to a carboxylic acid. This indicates that alanine and serine have been cleaved from the C-terminus, but lysine-48 remains, becoming the new C-terminal amino acid (compare with carboxypeptidase digestion in Figure 5). This cleavage results in a small shift in resonance C, although lysines D and E appear unaffected. Gradual changes are seen in the alanine spectrum over the same period. After about 20 h, when no further changes had occurred, the small peptides were removed from the undigested core by gel filtration on Sephadex G25, and the core region was reconcentrated and examined (Figure 7A,B, spectrum q). Only the three central core alanine residues remain, whereas four of the five lysines are intact, confirming the correspondence of peak B and lysine-8. Peak C, which shifts upfield during the digestion, was tentatively assigned to lysine-44 and peak D, which does not shift, to lysine-43. However, these adjacent residues, which are barley resolved, display similar properties, and the distinction is therefore not generally relevant. Further addition of large quantities of proteinase K to the core peptide did result in digestion of the lysine residues, but only at a rate comparable to the three alanines, suggesting total destruction of the core peptide.

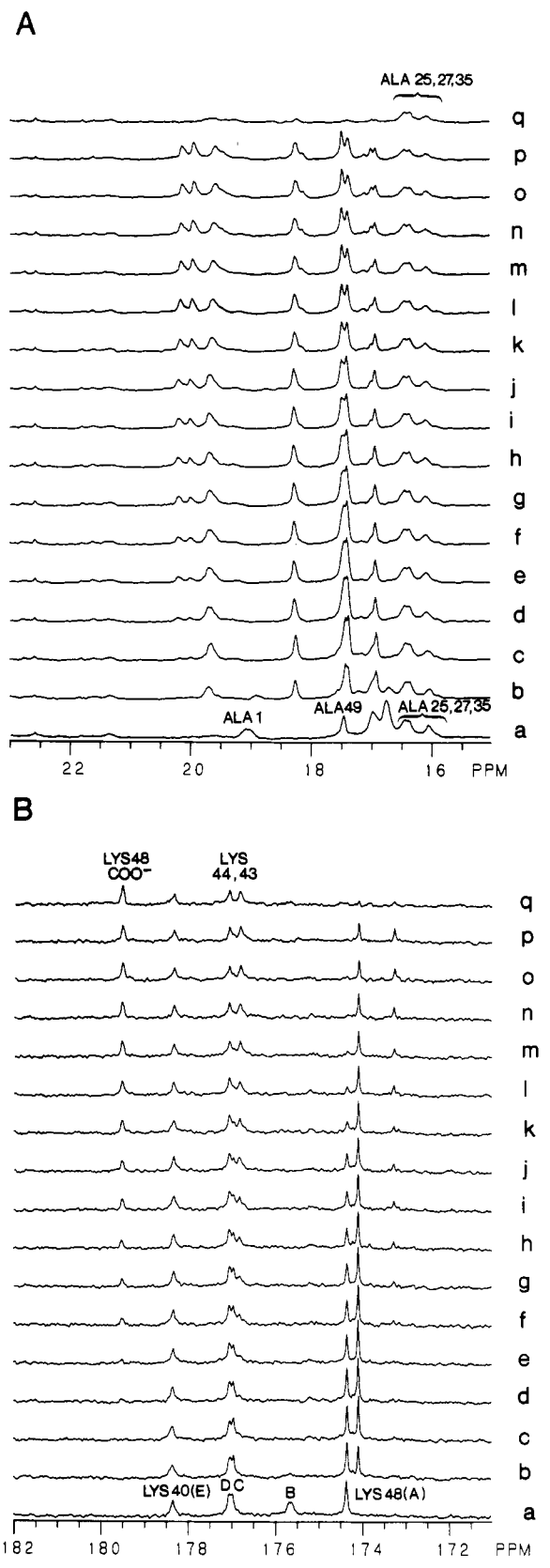

FIGURE 7: Time course of proteinase K digestion of [1-^{13}C]lysine/[3-^{13}C]alanine-labeled coat protein in SDS micelles, pH 9.0, showing (A) methyl and (B) carbonyl regions of the spectrum. The acquisition parameters were as in Figure 2; 5000 transients were averaged (1.41 h per spectrum) except after gel filtration (10000) where the protein was less concentrated. (a) No enzyme; (b–d) 1:850 w/w proteinase K; (e–k) further addition of 1:850 (w/w) proteinase K; (l–p) further addition of 1:280 proteinase K; (q) core particle after gel filtration on Sephadex G25 to remove small peptides. A line broadening of 2 Hz was used for all spectra.

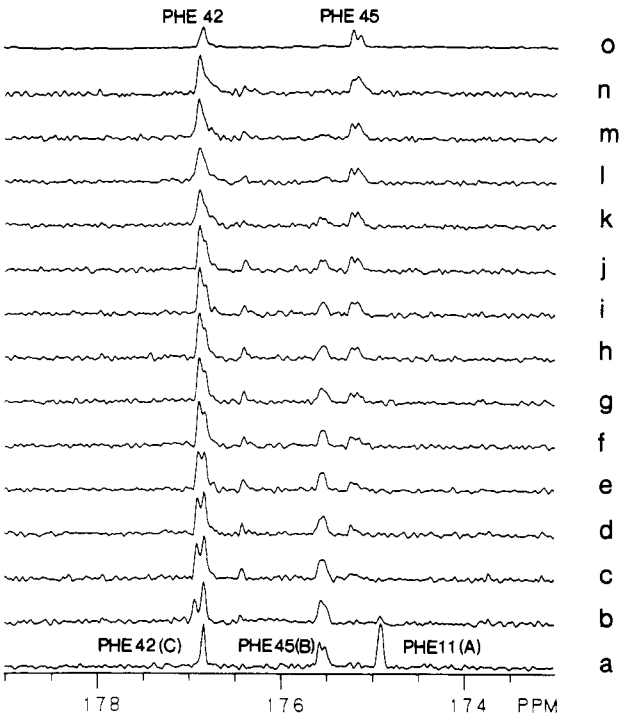

FIGURE 8: Carbonyl region of [1-^{13}C]phenylalanine-labeled coat protein in SDS micelles, pH 9.0, throughout a time course of proteinase K digestion. 5000 scans (1.41 h) were collected for each spectrum; conditions otherwise as in Figure 2. (a) No enzyme; (b–j) 1:850 w/w proteinase K, 1.41-h delay between spectra; (k–n) further addition of 1:280 w/w proteinase K; (o) protein after gel filtration to remove small peptides. Resolution was enhanced by multiplication of the free induction decay by a double-exponential function.

The five lysine residues have thus been assigned, and the proteinase K resistant region of the protein has been shown to extend at least from threonine-19 as far as lysine-48. The N-terminal hydrophilic region, unlike the C-terminus, appears to be very susceptible to proteinase K.

(*d*) *Phenylalanine.* The three phenylalanine residues (Figures 1 and 2e) occur in the N-terminal domain (phenylalanine-11) and flanking the two adjacent lysines at the C-terminus (phenylalanines-42 and -45). On the basis of the success of the lysine assignment, the phenylalanine spectrum can be completely assigned by using proteinase K (Figure 8). An alanine methyl label (not shown) was included for control. Peak A, like lysine-8 under similar conditions, is cleaved almost immediately (Figure 8b) and is assigned to phenylalanine-11. A new peak appears which partially overlaps peak C. Peak B, which is split into two resonances, shifts upfield on a time scale corresponding to the exposure of lysine-48 at the C-terminus, whereas peak C shifts downfield slightly. Peak B was thus assigned with reasonable certainty to phenylalanine-45 and peak C to phenylalanine-42. After gel filtration, the core particle was shown to retain peaks B and C at equal intensity as anticipated (Figure 8o).

(*e*) *Glycine.* The glycine spectrum (Figure 2f) contains another narrow resonance (peak A), which can be identified with peak 1, the most upfield resonance in the entire natural-abundance carbonyl spectrum. Of the four glycine residues, three are in the hydrophobic domain, and the fourth is glycine-3. Digestion with proteinase K (not shown) in an experiment analogous to those described previously produces no apparent change in the spectrum although the very narrow resonance (peak A) is removed by gel filtration and therefore assigned to glycine-3. The three glycines of the hydrophobic core result in four peaks, one of which (E) is quite broad. This is not thought to be due to scrambling of the label because

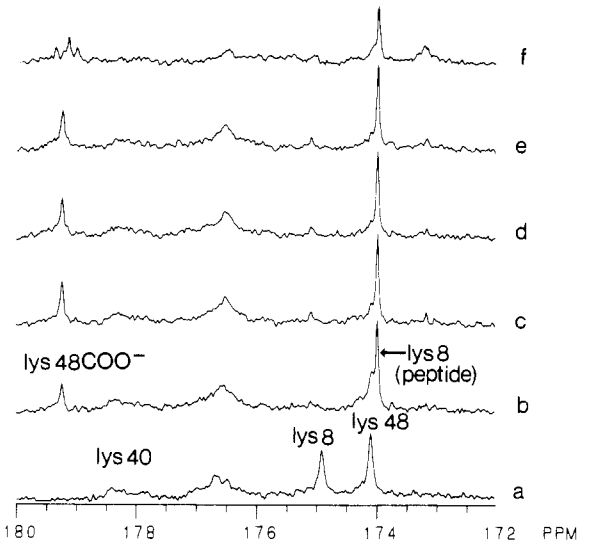

FIGURE 9: Carbonyl region of [1-^{13}C]lysine-labeled coat protein solubilized in sodium deoxycholate throughout a time course of proteinase K digestion. 5000 scans were averaged, conditions otherwise as in Figure 2. (a) Undigested coat protein, showing assignments; (b–e) 1:850 w/w proteinase K; (f) after extensive digestion with 1:280 w/w proteinase K; lysines-40, -43, -44, and -48 have been degraded.

only glycine is observed in a protein hydrolysate. Furthermore, the total peak area is almost exactly 4 times that of glycine-3 if the spectrum is recorded under conditions of full relaxation and the decoupler gated on only during acquisition to avoid buildup of an NOE. The most likely contaminant, serine, was also avoided by growing the cells on a mutant strain which lacks serine hydroxymethyltransferase. This glycine presents another example of a single residue experiencing more than one environment (see above).

Digestion of Deoxycholate-Solubilized Coat Protein with Proteinase K. In addition to providing a useful tool in resonance assignment, proteinase K digestion has allowed us to monitor the protection afforded to various regions of the protein by the bound detergent. As SDS is thought to bind to hydrophilic as well as hydrophobic regions in some proteins, an experiment exactly analogous to that described in Figure 7 was performed in deoxycholate, a detergent which generally binds only to hydrophobic regions in proteins (Robinson & Tanford, 1975). A similar experiment was performed by Woolford and Webster (1975) using radiolabeled coat protein. The C-1 lysine spectrum obtained in deoxycholate (Figure 9) is very similar to that in SDS (Figure 3), and the assignments (to be presented elsewhere) correspond exactly. The deoxycholate spectrum, however (with the exception of lysine-48), is considerably broader, with apparent line widths approaching 40 Hz for lysine-40. This is not due to aggregation, as such particles have been shown to have similar overall correlation times (Henry et al., 1986), but probably results from conformational heterogeneity, perhaps in the packing of the protein and detergent to form the micelle.

Proteinase K digestion in deoxycholate proceeds more rapidly than in SDS, but the overall digestion pattern is otherwise very similar. Lysine-8 is removed in the time taken to acquire the first spectrum, and a new narrow resonance almost superimposes lysine-48, which is already somewhat diminished (Figure 9b). At a slightly slower rate, lysine-48 becomes exposed at the C-terminus, resulting in a downfield carboxyl resonance. Extensive aggregation (not seen in SDS) is apparent during the time course of digestion in deoxycholate, as evidenced by the gradual increase in line width of various resonances. A core particle, apparently identical with that obtained in SDS, can be isolated by gel filtration (not shown). Further digestion (Figure 9f) eventually removes the remaining lysine residues at a rate comparable to those of the buried alanine residues (25, 27, and 35).

Discussion

A study of the carbonyl carbons of a protein can potentially yield much valuable information concerning the dynamics of the protein backbone. Although the α-carbons could deliver similar information, the ^{13}C-carbonyl-labeled amino acid precursors are generally easier to synthesize and more readily available commercially. It is also important that they resonate well away from most detergent or lipid resonances. The narrow line widths of the nonprotonated carbonyl carbons lead to well-resolved spectra and have also enabled us to use the isotopically shifted resonances (^{13}CONH and ^{13}COND forms) observed in 50% H_2O/D_2O mixtures as a measure of amide exchange rates (Hawkes et al., 1978; Henry et al., 1986b, 1987). However, although the α-carbons are relaxed almost entirely by dipolar interactions with the attached proton (Oldfield et al., 1975), the carbonyl carbons are relaxed both by the dipolar mechanism (with adjacent αCH and NH protons) and by chemical shift anisotropy. In fact, the latter mechanism may be dominant for carbonyl carbons at 75 MHz (Norton et al., 1977). This makes it difficult to quantitate their motional behavior from relaxation data obtained at a single frequency. The line width of a resonance (which is inversely related to T_2, the spin–spin relaxation time) is a monotonic function of correlation time. Nuclei which are undergoing motions in addition to the overall micelle tumbling (and on a faster time scale) will experience a shorter "effective" correlation time and give rise to narrow resonances. A qualitative description of backbone motions can therefore be obtained from the line widths. Most backbone carbons in a folded protein are expected to be rigid; that is, no motions of significant amplitude occur with a time period shorter than τ_c ($\simeq 10^{-8}$ s in this case), so that most carbonyl carbon resonances in a protein would be expected to be of similar width. Large-amplitude motions which are slower than this are possible, but these will not be detected by the NMR experiment.

The natural-abundance spectrum of M13 coat protein contains approximately eight motionally narrowed backbone carbonyl resonances. Three of these have been identified: glycine-3, lysine-48, and alanine-49, and peak 7 almost certainly corresponds to glutamic acid-2. Alanine-1, which is exchange broadened to the point of invisibility at pH 9, also gives a very narrow resonance if the pH is well above or below its pK_a. Thus, most, if not all, of the mobility arises from the extreme ends of the molecule. Peaks 2, 3, 5, and 8 of the natural-abundance spectrum have yet to be assigned. One of these may be the side chain of glutamine-15, and another, probably peak 8 (judging by chemical shift), must originate from the C-terminal carboxyl, serine-50. As phenylalanine-45 and proline-6 are clearly not mobile residues, and the transition between mobility and rigidity cannot be abrupt, the best candidates for the remaining resonances are aspartic acid-4, aspartic acid-5, threonine-46, and serine-47. The terminal regions of the protein are quite highly charged and therefore less likely to fold into an ordered structure.

This view of the protein, in which all the rapid motions occur at the termini, is in good agreement with our previous experiments in which motions of the $\alpha\beta$ axis of [3-^{13}C]alanine-labeled coat protein were quantitated (Henry et al., 1986a). Modeling of order parameters (Lipari & Szabo, 1982) on a diffusion-in-a-cone model led to semiangles of 58° and

38° for alanines-1 and -49, respectively, whereas the remaining alanines were rigid. These results also correlate well with the results of Cross and Opella (1980), who pointed out approximately eight narrow α-carbon resonances in the ^{13}C NMR spectrum of SDS-solubilized coat protein. Circular dichroism experiments (Nozaki et al., 1976, 1978) (which give similar spectra for M13 coat protein in SDS, deoxycholate, or phospholipid vesicles) suggest approximately 20% of the protein (10 residues) is in the "random" coil form.

All of the narrow resonances are at the high-field end of the carbonyl region and correspond most closely to the shifts observed in model peptides (Keim et al., 1973a,b, 1974). Some residues, e.g., phenylalanine and lysine, show substantial downfield shifts for residues closer to the hydrophobic core; this may correspond either to the strength of H bonding or to the hydrophobicity of the environment, or quite possibly to both effects combined. Shifts in the opposite direction are observed for alanine methyl resonances (see Figure 7A, for example).

Proteinase K, in addition to its usefulness in assignment experiments, provides information on the exposure of the protein. This enzyme has been used to determine the orientation of coat protein in phospholipid vesicles (Bayer & Feigenson, 1985). Although SDS binds to water-soluble proteins in a ratio of 1.4 g/g of protein (Reynolds & Tanford, 1971), there is much evidence to suggest that it has very different interactions with the small hydrophobic membrane protein (Nozaki et al., 1978). Proteinase K digestion readily removes the N-terminal hydrophilic segment whereas the hydrophobic core and part of the C-terminal hydrophilic region are protected from the enzyme. Experiments with [1-^{13}C]-tyrosine-labeled coat protein (G. D. Henry, J. H. Weiner, and B. D. Sykes, unpublished data) show that tyrosine-21 is not removed from the core particle, suggesting the cleavage site to be threonine-19 or glutamic acid-20. Very similar results are obtained in deoxycholate, which normally binds only to the hydrophobic regions of proteins (Robinson & Tanford, 1979). Circular dichroism (Nozaki et al., 1976) and NMR experiments (e.g., with C-1 lysine-labeled protein) suggest that the protein structure is basically similar in both detergents. Although the N-terminus is readily susceptible to proteinase K, the C-terminus survives digestion almost intact; in fact, lysines-40, -43, and -44 are not digested until the whole core particle breaks down. It is possible that lysine-rich regions are not good substrates for proteinase K; however, this region of the protein (in DOC micelles) is not readily susceptible to trypsin either (Woolford & Webster, 1978; G. D. Henry, J. H. Weiner, and B. D. Sykes, unpublished observations). It is unlikely that the positive charges on lysine could be buried, but they might be stably complexed with the negatively charged detergent. Alternatively, this region of the protein may represent a particularly stable region of secondary structure.

In summary, the [^{13}C]carbonyl-labeled M13 coat protein has provided good evidence for a model in which large-scale motions are possible only at the N- and C-termini. The rest of the molecule is rigid on a time scale of 10^{-8} s. Hydrogen exchange measurements, derived indirectly from the carbonyl resonances, are described in Henry et al. (1987).

ACKNOWLEDGMENTS

We thank G. McQuaid for upkeep of the NT 300WB spectrometer, Dr. S. J. Opella for providing many preprints prior to publication, and for many discussions with him and his research group, and Dr. J. D. J. O'Neil for many helpful discussions.

Registry No. SDS, 151-21-3; sodium deoxycholate, 302-95-4.

REFERENCES

Asbeck, V. F., Beyreuther, K., Kohler, H., von Wettstein, G., & Braunizer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem. 350*, 1047–1056.

Banner, D. W., Nave, C., & Marvin, D. A. (1981) *Nature (London) 289*, 814–816.

Bayer, R., & Feigenson, G. W. (1985) *Biochim. Biophys. Acta 815*, 369–379.

Cavalieri, S. J., Goldthwaite, D. A., & Neet, K. E. (1976) *J. Mol. Biol. 102*, 713–722.

Chang, C. N., Blobel, G., & Model, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 361–365.

Cross, T. A., & Opella, S. J. (1979) *J. Supramol. Struct. 11*, 139–145.

Cross, T. A., & Opella, S. J. (1980) *Biochem. Biophys. Res. Commun. 92*, 478–484.

Cross, T. A., & Opella, S. J. (1981) *Biochemistry 20*, 290–297.

Cross, T. A., & Opella, S. J. (1982) *J. Mol. Biol. 159*, 543–549.

Cross, T. A., & Opella, S. J. (1985) *J. Mol. Biol. 182*, 367–381.

Dettman, H. D., Weiner, J. H., & Sykes, B. D. (1982) *Biophys. J. 37*, 243–251.

Dettman, H. D., Weiner, J. H., & Sykes, B. D. (1984) *Biochemistry 23*, 705–712.

Ebeling, W., Hennrich, H., Klockow, M., Metz, H., Orth, H. D., & Lang, H. (1974) *Eur. J. Biochem. 47*, 91–97.

Frank, L., & Ranhand, B. (1964) *Arch. Biochem. Biophys. 107*, 325–331.

Gall, C. M., Cross, T. A., DiVerdi, J. A., & Opella, S. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 101–105.

Hagen, D. S., Weiner, J. H., & Sykes, B. D. (1979) *Biochemistry 18*, 2007–2012.

Hansen, P. E. (1983) *Annu. Rep. NMR Spectrosc. 15*, 105–234.

Henry, G. D., Weiner, J. H., & Sykes, B. D. (1986a) *Biochemistry 25*, 590–598.

Henry, G. D., O'Neil, J. D. J., Weiner, J. H., & Sykes, B. D. (1986b) *Biophys. J. 49*, 329–331.

Henry, G. D., Weiner, J. H., & Sykes, B. D. (1987) *Biochemistry* (following paper in this issue).

Keim, P., Vigna, R. A., Marshall, R. C., & Gurd, F. R. N. (1973a) *J. Biol. Chem. 248*, 6104–6113.

Keim, P., Vigna, R. A., Morrow, J. S., Marshall, R. C., & Gurd, F. R. N. (1973b) *J. Biol. Chem. 248*, 7811–7818.

Keim, P., Vigna, R. A., Nigen, A. M., Morrow, J. S., & Gurd, F. R. N. (1974) *J. Biol. Chem. 249*, 4149–4165.

Knippers, R., & Hoffmann-Berling, H. (1966) *J. Mol. Biol. 221*, 281–292.

Levitt M. H., & Freeman, R. (1981) *J. Magn. Reson. 43*, 502–507.

Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546–4559.

Makino, S., Woolford, J. L., Tanford, C., & Webster, R. (1975) *J. Biol. Chem. 250*, 4327–4337.

Nakashima, Y., & Konigsberg, W. (1974) *J. Mol. Biol. 88*, 598–600.

Norton, R. S., Clouse, A. O., Addleman, R., & Allerhand, A. (1977) *J. Am. Chem. Soc. 99*, 79–83.

Nozaki, Y., Chamberlain, B. K., Webster, R. E., & Tanford, C. (1976) *Nature (London) 259*, 335–337.

Nozaki, Y., Reynolds, J. A., & Tanford, C. (1978) *Biochemistry 17*, 1239-1246.
Oldfield, E., Norton, R. S., & Allerhand, A. (1975) *J. Biol. Chem. 250*, 6368-6380.
Opella, S. J., Cross, T. A., DiVerdi, J. A., & Sturm, C. F. (1980) *Biophys. J. 32*, 531-548.
Reynolds, J. A., & Tanford, C. (1970) *Proc. Natl. Acad. Sci. U.S.A 66*, 1002-1007.
Robinson, N. C., & Tanford, C. (1975) *Biochemistry 14*, 369-378.
Shindo, H., Egan, W., & Cohen, J. S. (1978) *J. Biol. Chem. 253*, 6751-6755.
Smilowitz, H. (1974) *J. Virol. 13*, 94-99.
Smilowitz, H., Carson, J., & Robbins, P. W. (1972) *J. Supramol. Struct. 1*, 8-18.
Sugimoto, K., Sugisaki, H., & Takanami, M. (1977) *J. Mol. Biol. 110*, 487-507.
Valentine, K. G., Schneider, D. M., Leo, G. C., Colagno, L. A., & Opella, S. J. (1986) *Biophys. J. 49*, 38-40.
Williams, R. W., & Dunker, A. K. (1977) *J. Biol. Chem. 252*, 6253-6255.
Wilson, M. L., & Dahlquist, F. W. (1985) *Biochemistry 24*, 1920-1928.
Woolford, J. L., & Webster, R. E. (1975) *J. Biol. Chem. 250*, 4333-4339.

Backbone Dynamics of a Model Membrane Protein: Measurement of Individual Amide Hydrogen-Exchange Rates in Detergent-Solubilized M13 Coat Protein Using ^{13}C NMR Hydrogen/Deuterium Isotope Shifts†

Gillian D. Henry, Joel H. Weiner, and Brian D. Sykes*

Medical Research Council Group in Protein Structure and Function and Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

Received December 19, 1986

ABSTRACT: Hydrogen-exchange rates have been measured for individual assigned amide protons in M13 coat protein, a 50-residue integral membrane protein, using a ^{13}C nuclear magnetic resonance (NMR) equilibrium isotope shift technique. The locations of the more rapidly exchanging amides have been determined. In D_2O solutions, a peptide carbonyl resonance undergoes a small upfield isotope shift (0.08–0.09 ppm) from its position in H_2O solutions; in 1:1 H_2O/D_2O mixtures, the carbonyl line shape is determined by the exchange rate at the adjacent nitrogen atom. M13 coat protein was labeled biosynthetically with ^{13}C at the peptide carbonyls of alanine, glycine, phenylalanine, proline, and lysine, and the exchange rates of 12 assigned amide protons in the hydrophilic regions were measured as a function of pH by using the isotope shift method. This equilibrium technique is sensitive to the more rapidly exchanging protons which are difficult to measure by classical exchange-out experiments. In proteins, structural factors, notably H bonding, can decrease the exchange rate of an amide proton by many orders of magnitude from that observed in the freely exposed amides of model peptides such as poly(DL-alanine). With corrections for sequence-related inductive effects [Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150–158], the retardation of amide exchange in sodium dodecyl sulfate solubilized coat protein has been calculated with respect to poly(DL-alanine). The most rapidly exchanging protons, which are retarded very little or not at all, are shown to occur at the N- and C-termini of the molecule. In the N-terminal region, up to and including aspartic acid-12, retardations no more than about 20-fold are observed; these exchange rates are much more rapid than the values determined for regions of stable secondary structure in other proteins. Proceeding inward from the C-terminus, by contrast, reveals a steep and progressive increase in retardation of the exchange rate from threonine-46 to leucine-41 in a manner suggestive of the fraying end of a helical segment. A model of the detergent-solubilized coat protein is constructed from these H-exchange data which is consistent with circular dichroism and other NMR results.

Measurement of the exchange rates of backbone amide protons in a protein has been recognized for many years to be a potential source of structural and dynamic information. The initial H-exchange experiments on model polypeptides (Linderstrom-Lang, 1955; Linderstrom-Lang & Schellman, 1959) in fact provided much of the initial impetus for the development of current concepts of conformational fluctuations in proteins. Although the interpretation of H-exchange data is still controversial (Woodward et al., 1982; Englander & Kallenbach, 1984), the rates of exchange of amide protons with the solvent are generally considered to be heavily influenced by H bonding, for example, amide exchange rates were recently used to corroborate the solution structure of the *lac* repressor head piece as determined by nuclear magnetic res-

† Supported by the Medical Research Council of Canada (MRC Group in Protein Structure and Function and MRC Grant MT5838) and the Alberta Heritage Foundation for Medical Research. Paper 2 in the series is by Henry et al. (1987).

* Address correspondence to this author at the Department of Biochemistry, University of Alberta.

0006-2960/87/0426-3626$01.50/0