Conformational Stability of P22 Tailspike Proteins Carrying Temperature-Sensitive Folding Mutations†

George J. Thomas, Jr.,*,‡ Renee Becka,‡ Donna Sargent,‡ Myeong-Hee Yu,§ and Jonathan King§

Division of Cell Biology and Biophysics, School of Basic Life Sciences, University of Missouri—Kansas City, Kansas City, Missouri 64110, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received August 2, 1989; Revised Manuscript Received December 29, 1989

ABSTRACT: The thermostable tailspike endorhamnosidase of *Salmonella* phage P22 provides a model system for comparing the role of amino acid sequences in determining the intracellular folding pathway with their role in stabilizing the mature structural protein. Complete Raman band assignments are given here for the native form of the tailspike trimer in aqueous solution. Once correctly folded and assembled, the wild-type and two well-characterized mutant proteins, *tsf*Ile258 → Leu and *tsf*Gly323 → Asp, exhibit the same secondary structure in solution, consisting predominantly of β-strand (56 ± 5%) and turns (17 ± 2%). Raman bands that are sensitive indicators of hydrogen-bonding interactions of tyrosine (phenolic OH) and tryptophan (indole NH) are unchanged between 30 and 80 °C in both wild type and *tsf* mutants. Similarly, Raman bands that are sensitive to changes in the hydrophobic environment of nonpolar side chains exhibit no significant temperature dependence in wild type and *tsf* mutants. In contrast, these conformational features are greatly altered by chemical denaturation of the tailspike with lithium halide and guanidine hydrochloride. In the chemically denatured tailspike, the β-strand structure is substantially converted to irregular or "random coil" conformation. These findings confirm conclusions from physiological studies that the three-dimensional structures of the *tsf* mutants, once stabilized at permissive temperatures, are equivalent to the native structure of the wild type, and this structure is maintained at temperatures far above those that block the folding of the chain into the final native conformation. The mutant sites therefore represent positions in the amino acid sequence where the side chains make important contributions to the stability of folding intermediates but do not contribute significantly to the stabilization of the native structure.

It has long been recognized that the primary structure of a protein can contain all of the information necessary to fold the macromolecule into its native three-dimensional conformation. Recent experiments on ribonuclease A (Udgaonkar & Baldwin, 1988) and cytochrome *c* (Roder et al., 1988) confirm that specific structural intermediates can be identified along the folding pathway. The conformations of these intermediates must also be specified by the amino acid sequence. However, the specific roles of the individual amino acids in directing the conformation of an intermediate, compared with their roles in stabilizing the mature native structure, are not well understood. For the P22 tailspike protein, in vivo folding intermediates can be distinguished from the native protein, making it possible to differentiate effects on the folding pathway from effects on the stability of the native state (Goldenberg & King, 1981; Sargent et al., 1988; Sturtevant et al., 1989).

A large set of temperature-sensitive mutants of the P22 tailspike protein have been isolated and characterized (Smith et al., 1980; King, 1986; Yu & King, 1988). These mutations are a different class than the well-characterized temperature-sensitive mutations of T4 lysozyme, many of which destabilize the native protein (Alber & Matthews, 1987). When released from the ribosome at the permissive temperature (30 °C) the mutant polypeptide chains form the native tailspike through the pathway shown in Figure 1 (Goldenberg & King, 1982). However, when released at a higher restrictive temperature (40 °C), the chains fail to form the native tailspikes (Smith & King, 1981). The folding pathway is blocked at a stage in which the mutant polypeptide chains are not correctly folded or associated (Goldenberg et al., 1983). If maintained at the restrictive temperature (40 °C), these temperature-sensitive folding (*tsf*) polypeptide chains form aggregates of partially or incorrectly folded chains (Haase-Pettingell & King, 1988). However, once correctly folded at permissive temperatures, the native forms of the mutant proteins are as active as wild type. The proposed maturation pathways of wild-type and mutant tailspikes are presumed to be identical at permissive temperatures (Figure 1).

We have begun a systematic study of the Raman spectra of wild-type and mutant tailspikes in order to determine the conformations in solution and to investigate the loci and nature of *tsf* defects (Sargent et al., 1988). To facilitate structural comparisons, we have assigned the principal Raman bands of the so-called fingerprint region (vibrational frequency range: 300–1800 cm^{-1}) and hydrogen stretching region (2500–3500 cm^{-1}) to their side-chain and backbone groups of the tailspike. Of the 40 or so assigned bands, several are diagnostic of specific side-chain environments and/or interactions, and these have been interpreted to provide specific structural information about mutant and wild-type proteins. In many cases these bands exhibit behavior which parallels that observed for the Raman amide I and amide III secondary structure markers.

We have chosen for careful study the wild-type tailspike and three well-characterized *tsf* mutants: tsU24 (Ile258 → Leu), tsH302 (Gly323 → Asp), and tsU38 (Gly435 → Glu). The tsU24 and tsH302 mutants, unlike tsU38, are as robust as wild type in their resistance to thermal inactivation (Sargent et al., 1988; Sturtevant et al., 1989). The tsU24 mutant represents the substitution of one hydrophobic side chain for another. On

† Part 27 in the series Studies of Virus Structure by Raman Spectroscopy. Supported by NIH Grant AI11855 (G.J.T.) and NSF Grant PCM 8402546 (J.K.).

* To whom correspondence should be addressed.

‡ University of Missouri.

§ Massachusetts Institute of Technology.

0006-2960/90/0429-4181$02.50/0

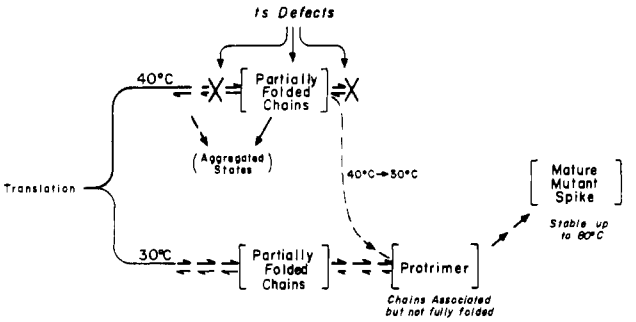

FIGURE 1: Intracellular pathway of maturation of the bacteriophage P22 tailspike. The polypeptide chain released from the ribosome at 30 °C (left) partially folds and associates to form a protrimer intermediate. The thermostable tailspike results from subsequent folding of the protrimer. Temperature-sensitive folding mutants released at 40 °C are blocked at an early stage in chain folding (temperature-sensitive stages) and do not associate to form competent protrimer intermediates at this restrictive temperature. However, if infected cells are shifted to 30 °C, the mutant chains continue through the productive pathway.

the other hand, tsH302 substitutes a large charged residue for a glycine that is thought to be at the surface of the protein (Yu & King, 1988). The possibility existed that the mutants were causing local disturbances in the structure of this large protein but in regions far from the location involved in the rate-limiting step of the unfolding pathway. The current Raman study allows us to address this question in greater detail.

Experimental Procedures

(1) Materials. The wild-type tailspike and mutants tsU24, tsH302, and tsU38 were purified from *Salmonella* cells infected with appropriate phage strains (Yu & King, 1988). The purified tailspikes were stored in the cold at a concentration of 1 mg/mL in Tris buffer (50 mM Na-Tris, pH 7.5 ± 0.1, 25 mM NaCl, 2 mM EDTA, 1 mM β-mercaptoethanol, 1% glycerol) prior to concentrating the solution (≈100 μg/μL) for Raman spectroscopy. Solutions of the native protein that were examined for sulfhydryl Raman bands were prepared without β-mercaptoethanol in the Tris buffer. Further details of sample handling for Raman spectroscopy of P22 tailspikes have been described (Sargent et al., 1988).

For denaturation studies, tailspikes were treated with guanidine hydrochloride (Gdn·HCl), lithium bromide, and lithium chloride to achieve final protein concentrations of approximately 35–50 μg/μL in 6 M Gdn·HCl, 6 M LiCl, or 6 M LiBr. In separate protocols, low concentrations of tailspike in the denaturing solvents (ca. 1–10 μg/μL) were prepared and subsequently concentrated 10-fold by ultrafiltration.

(2) Raman Spectroscopy. Raman spectra were obtained in both single- and multichannel modes with instrumentation described previously (Li et al., 1981; Lamba et al., 1990). Spectra in the region 600–1800 cm^{-1} were recorded on thermostated samples (Thomas & Barylski, 1970) at intervals of 10 °C. Further details are given by Sargent et al. (1988).

The Raman data were corrected for scattering by the solvent and buffer and for the gently sloping background characteristic of aqueous solution spectra of proteins. Signal averaging, base-line corrections, spectral subtraction, and related routines were performed on-line with data acquisition (Li et al., 1981; Lamba et al., 1990). Deconvolution procedures were carried out on IBM microcomputers using software described previously (Thomas & Agard, 1984). For quantitative Raman intensity comparisons in the region 300–1800 cm^{-1}, the data were normalized to the integrated intensity of the 1450-cm^{-1} band. In the 2750–4000-cm^{-1} region the broad water band was used as an intensity standard.

Table I: Raman Frequencies and Assignments of the Bacteriophage P22 Tailspike

frequency (cm^{-1})	intensity[a]	assignment[b]
415	1	skeletal
495	1	skeletal (Val, Gly)
621	1	Phe ring
642	1	Tyr ring
759	2	Trp ring
830	2	Tyr ring
852	3	Tyr ring
878	1	Trp ring
900	S	CH_3 rock (Ala)
936	2	CH_3 rock[c]
958	2	CH_3 rock[c]
1002	8	Phe ring
1015	S	Trp ring
1031	3	Phe ring
1050	S	C–N and C–C stretch[d]
1075	S	C–N and C–C stretch[d]
1100	S	C–C stretch (Ala)
1127	2	C–C stretch
1155	1	CH_3 rock[c]
1175	1	CH_3 rock[c]
1208	3	Tyr ring
1235	5	amide III[e]
1260	S	Tyr ring
1316	3	CH_2 deformation
1339	3	CH_2 deformation; Trp ring
1359	S	Trp ring
1398	2	CH_3 deformation; COO^- stretch (Asp, Glu)
1420	2	CH_2 deformation (Gly)
1450	5	CH_2 deformation
1460	S	CH_3 deformation
1550	2	Trp ring
1577	S	Trp ring
1605	S	Phe ring
1618	5	Tyr ring
1668	10	amide I[e]
2549	0	S–H stretch (Cys, strongly hydrogen-bonded)
2586	0	S–H stretch (Cys, weakly hydrogen-bonded)
2865	5	aliphatic C–H stretch
2935	10	aliphatic C–H stretch
2970	6	aliphatic C–H stretch
3060	1	aromatic C–H stretch

[a] Arbitrary units based upon a 1–10 scale, with 10 assigned to the 1668-cm^{-1} band. S denotes shoulder or satellite to a more intense band. [b] Based upon model compound studies as cited in Thomas et al. (1983). [c] Due mainly to side chains of Val, Leu, and Ile. [d] Due mainly to side chains of aliphatic residues. [e] See Sargent et al. (1988) for additional peaks resolved by Fourier deconvolution in this spectral region.

Results and Interpretation

(1) Raman Spectrum and Assignments. The Raman spectrum in the region 650–1750 cm^{-1} for the wild-type tailspike is shown in Figure 2 (top). Labels indicate Raman frequencies of prominent peaks, for which corresponding assignments are noted along the abscissa. A complete tabulation of the Raman frequencies and assignments is given in Table I. The most intense Raman band is due to the amide I mode and is centered at 1668 cm^{-1}. Other intense Raman bands originate from the amide III mode (centered at 1235 cm^{-1}), ring vibrations of aromatic side chains (Phe at 621, 1002, 1031, and 1602 cm^{-1}; Tyr at 642, 830, 852, 1208, 1260, and 1618 cm^{-1}; Trp at 759, 878, 1550, and 1577 cm^{-1}), and methyl and methylene deformation (1316, 1339, 1450, and 1460 cm^{-1}) and stretching modes (2865, 2935, and 2970 cm^{-1}) associated primarily with aliphatic side chains. Figure 2 also shows that the spectra of mutants tsU24 and tsH302 are virtually indistinguishable from the spectrum of wild-type tailspikes.

(2) Thermostability of Secondary Structure. Previously, we used Raman spectroscopy to monitor the thermostability

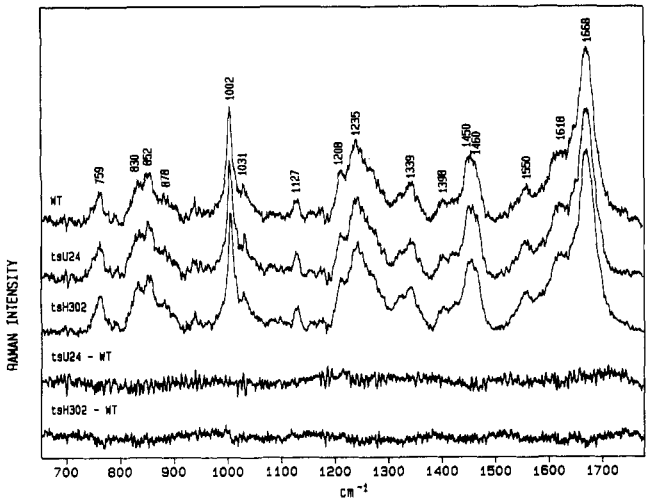

FIGURE 2: Raman spectra in the 650–1750-cm^{-1} region of the bacteriophage P22 tailspike. From top to bottom: wild-type tailspike, mutant tsU24 tailspike, mutant tsH302 tailspike, difference spectrum between wild type and tsU24 mutant, and difference spectrum between wild type and tsH302 mutant. Labels above the spectrum of the wild-type tailspike show frequencies of the prominent Raman peaks. Complete assignments are given in Table I. Spectra over the extended interval 300–1800 cm^{-1} are as shown in Figure 2 of Sargent et al. (1988).

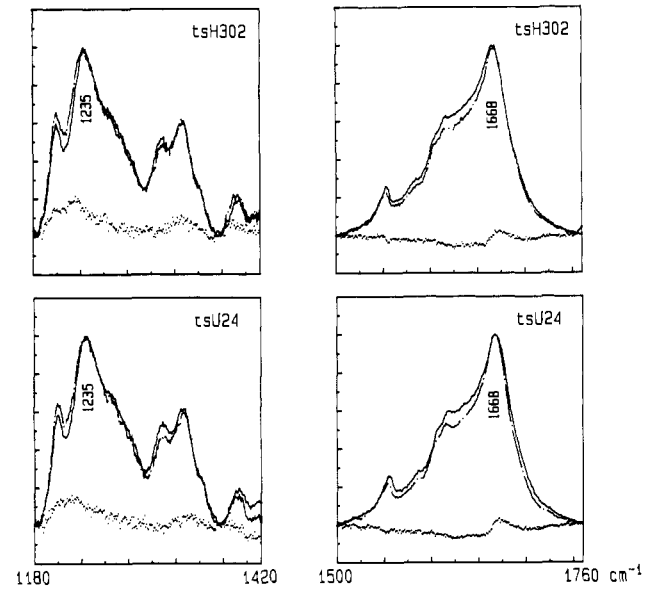

FIGURE 3: Signal-averaged Raman spectra in regions containing amide III (left panels: 1180–1420 cm^{-1}) and amide I (right panels: 1500–1760 cm^{-1}) bands of mutant tailspikes tsU24 (bottom) and tsH302 (top). Experimental data collected from samples at 30 °C (—) and 80 °C (···) are plotted with the computed difference spectrum (···) in each panel and indicate no significant secondary structure change. (See discussion in text.)

of secondary structure of the wild-type and tsU38 mutant tailspikes in aqueous solution (Sargent et al., 1988). The data showed that the wild-type protein was resistant to thermal denaturation up to 80 °C, while the tsU38 mutant was thermostable to 60 °C. We have carried out similar Raman experiments on mutants tsU24 and tsH302. As shown in Figure 3, there are no major changes in the positions or intensities of the principal amide I (1668 cm^{-1}) and amide III (1235 cm^{-1}) peaks in either mutant over the range 30–80 °C, indicating that the secondary structures of both of these mutants are fully as resistant to thermal denaturation as is the wild-type structure. The apparent reduction of Raman intensity in each mutant, observable in Figure 3 as a broad negative difference extending from 1600 to 1650 cm^{-1}, is due

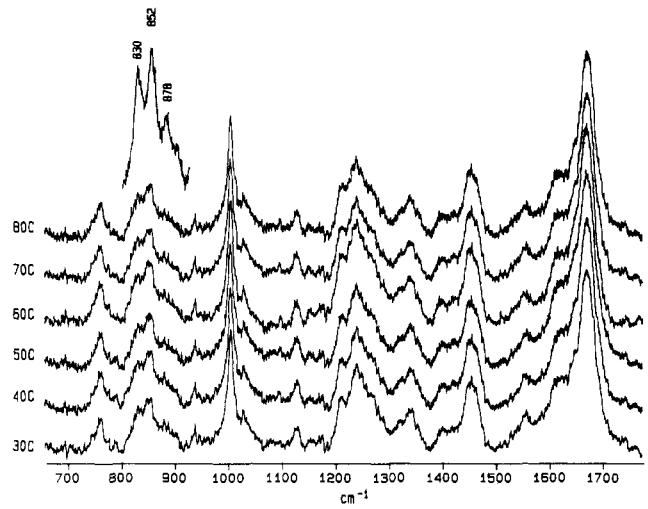

FIGURE 4: Raman spectra in the region 650–1750 cm^{-1} of wild-type tailspike at 10 °C intervals between 30 and 80 °C, showing invariance to temperature of the more prominent bands. The inset shows a high-resolution trace at 80 °C of the region 800–920 cm^{-1}, which includes the characteristic ring modes of tyrosine (830 and 852 cm^{-1}) and tryptophan residues (878 cm^{-1}), as discussed in the text. An identical trace is obtained at 30 °C.

to the uncompensated effect of temperature upon the spectrum of water (Sargent et al., 1988). Similarly, variations in the arbitrary base line of the data in the 1180–1420-cm^{-1} region generate minor artifacts that are apparent between 1180–1230 and 1300–1350 cm^{-1}.

The data of Figure 3 and of Sargent et al. (1988) show that, once matured at the permissive temperature (30 °C), all of the mutants are thermostable to well above the temperature (40 °C) at which the *tsf* defects are manifest.

(*3*) *Thermostability of Side-Chain Interactions and Environments.* We have monitored the Raman bands that originate from well-characterized vibrations of specific amino acid side chains in spectra of both wild-type and mutant tailspikes (tsU24, tsU38, and tsH302) over the temperature range 20 $\leq t$ (°C) $\leq$ 80. As a representative example we show in Figure 4 the Raman data obtained from the wild-type tailspike in the interval 650–1750 cm^{-1}, including bands of tyrosine and tryptophan residues between 800 and 920 cm^{-1} (inset), which are sensitive to hydrogen-bonding interactions of their respective phenolic OH and indole NH groups (Siamwiza et al., 1975; Miura et al., 1988). The data of Figure 4 show clearly that these bands undergo no significant frequency shifts or intensity changes with temperature. Similarly, we have monitored as a function of temperature the following spectral regions: (i) 500–550 cm^{-1}, for evidence of possible disulfide bond formation from oxidation of cysteinyl S–H side chains, (ii) 1000–1050 cm^{-1}, for altered stacking interactions of phenylalanine residues, (iii) 1340–1380 cm^{-1}, for changes in the hydrophobic microenvironment of tryptophan indoles, (iv) 1390–1410 cm^{-1}, for changes in ionization states of aspartate and glutamate carboxyls, and (v) the spectral regions of aliphatic CH deformation (1300–1500 cm^{-1}) and stretching (2800–3000 cm^{-1}) vibrations, for isomerizations of aliphatic side chains. In every case, no significant spectral change could be detected.

(*4*) *Nature of Side-Chain Interactions.* Many bands in the Raman spectrum of the tailspike are broad (Figure 2), which suggests that they contain overlapping components. Closely spaced band components may originate from vibrations of residues of the same kind in slightly different microenvironments of the protein (e.g., exposed and buried indole rings of tryptophan) or from residues of different kinds that generate

nearly identical vibrational frequencies (e.g., aliphatic C–H deformation modes of different types of amino acids, Table I). When the band components are at least partially resolved instrumentally, resolution enhancement techniques can be employed to improve band decomposition for structural interpretation. Previously, we employed Fourier deconvolution methods to identify the probable band components in the spectral intervals 1180–1400 and 1500–1800 cm^{-1}, and we discussed their structural significance, with particular emphasis on the protein secondary structure markers in these intervals, i.e., Raman amide III and amide I bands, respectively (Sargent et al., 1988). Here we employ correlations that have been established between certain Raman markers and residue side-chain environments (Siamwiza et al., 1975; Miura et al., 1988; H. Li and G. J. Thomas, Jr., unpublished results) to comment further on the nature of side-chain interactions in the tailspike.

(*a*) *Tyrosines.* The tailspike subunit contains 27 tyrosines and all of these contribute to the characteristic Raman bands of the phenolic ring at 642, 830, 852, 1208, 1260, and 1618 cm^{-1} (Table I). The relative intensities of the bands at 852 and 830 cm^{-1} are particularly informative. Lord and coworkers (Siamwiza et al., 1975) have shown that the Raman intensity ratio I_{852}/I_{830} is diagnostic of the hydrogen-bonding state of the phenolic OH group. Three distinctive states have been identified. When the OH group is the acceptor of a strong hydrogen bond from a positive donor group, it is observed that $I_{852}/I_{830} = 2.50$. We denote this acceptor state as "A". Tyrosines exclusively in the A state are found in coat subunits of Pf1 and fd virions (Thomas et al., 1983). When the phenolic OH acts as both donor and acceptor of moderately strong hydrogen bonds, as for example when exposed to solvent H_2O molecules or otherwise hydrogen-bonded to both donor and acceptor groups, then $I_{852}/I_{830} = 1.25$. We designate this exposed state as "E". Tyrosines in the E state are frequently observed at the surfaces of proteins (Thomas et al., 1986). Finally, when OH is the donor of a strong hydrogen bond to a negative acceptor group, $I_{852}/I_{830} = 0.30$. We designate this donor state as "D". Tyrosines in the D state have been identified in subunits of several RNA plant viruses (Verduin et al., 1984, Prescott et al., 1985). These quantitative correlations are believed to be accurate to ±5% (Siamwiza et al., 1975).

As shown in Figure 4 [and in Figure 2 of Sargent et al. (1988)], we find that for wild-type and mutant tailspikes $I_{852}/I_{830} = 1.25 \pm 0.07$. Since the Raman bands monitor all tyrosines, the intensity quotient of 1.25 represents an average over hydrogen-bonding states of all 27 Tyr residues per subunit (or 81 per trimer). There are 51 possible configurations of the 27 tyrosines that are consistent with $I_{852}/I_{830} = 1.25 \pm 0.07$, including the possibilities that all are in the E state (E = 27, A = 0, D = 0, for which $I_{852}/I_{830} = 1.25$) and that none are in the E state (E = 0, A = 11, D = 16, for which $I_{852}/I_{830} = 1.20$). The present results are sufficient to rule out all configurations for which A > 11 as well as all configuration for which D > 16. The Raman spectra further indicate that hydrogen-bonding configurations of tyrosines in wild-type and mutant tailspikes are the same. One clear-cut conclusion from these data is that a significant fraction of the tailspike tyrosines have the –OH group acting as a hydrogen-bond acceptor.

(*b*) *Tryptophans.* The seven tryptophan residues of the tailspike subunit account exclusively for the clearly resolved Raman band at 759 cm^{-1} and generate as well a number of other characteristic Raman bands (Table I). The most useful for structural assessments are the moderately intense tryptophan ring mode observed at 878 cm^{-1} (Figure 4) and the barely discernible shoulder at 1359 cm^{-1} (Figure 2). A detailed analysis by Harada and co-workers (Miura et al., 1988) has shown that the Trp ring mode in the 870–884-cm^{-1} interval of the Raman spectrum is diagnostic of the hydrogen-bonding state of the indole ring 1-NH group, while the relative intensity and sharpness of the Trp ring mode near 1360 cm^{-1} are indicative of the hydrophobicity of the indole ring environment. Thus, the conspicuous 878-cm^{-1} Raman band (Figure 4) indicates that the most representative state of tryptophan in the tailspike is such that the indole 1-NH group donates a moderately strong hydrogen bond. This is consistent with the weakness of the 1359-cm^{-1} shoulder (barely detectable in Figure 2), which shows that Trp residues exist in hydrophilic environments. Each hydrophilic environment may consist of water molecules or other polar groups, and different hydrogen-bonding partners are to be expected for different residues (Miura et al., 1988). In short, the Raman spectra of both wild-type and mutant tailspikes suggest that there are few or no tryptophans buried in hydrophobic domains and most likely all tryptophans engage the 1-NH donor in moderately strong hydrogen bonding, such as that which can occur with H_2O molecules or other appropriate acceptors.

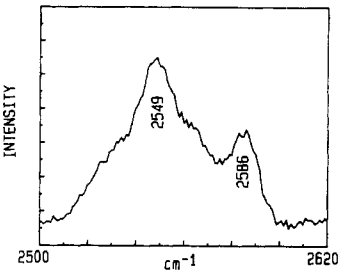

FIGURE 5: Raman spectrum of the tailspike in the region of cysteinyl S–H stretching vibrations, showing the dominance of strongly hydrogen-bonded S–H groups (2549-cm^{-1} band) and a lesser population of weakly or non-hydrogen-bonded S–H groups (2586-cm^{-1} band). See text.

(*c*) *Cysteines.* The tailspike subunit contains eight cysteine residues (Sauer et al., 1982), all of which contribute to a weak but complex band shape between 2500 and 2600 cm^{-1} (Figure 5). The Raman intensities associated with sulfhydryl group stretching vibrations are typically very weak compared with other bond-stretching vibrations and have been greatly amplified in Figure 5. That all cysteines are indeed reduced is confirmed by the absence of any Raman band in the interval 500–550 cm^{-1}, where Raman markers of disulfide bridges normally occur (Lord & Yu, 1970). As noted above and discussed previously (Thomas et al., 1982; Sargent et al., 1988), the cysteines resist oxidation even after the tailspike is thermally denatured at 90 °C. The complex S–H band shape of Figure 5 is composed of a major broad band at 2549 cm^{-1} and a satellite peak at 2586 cm^{-1}. The relative intensity ratio of the 2549- and 2586-cm^{-1} components is roughly 3:1. Assuming that the Raman scattering cross section is the same for all cysteines, the data indicate that six of the eight cysteines per subunit contribute to the broad 2549-cm^{-1} component while two contribute to the 2586-cm^{-1} component.

Typically, in a soluble globular protein, the cysteinyl S–H stretching vibration generates a Raman band near 2570–2580 cm^{-1}. For example, the principal S–H Raman band is observed near 2575 cm^{-1} for both the P22 capsid (gp5) and scaffolding (gp8) subunits (Thomas et al., 1982). For many other viral proteins examined previously, the dominant Raman band from cysteinyl S–H stretching also is centered near 2575 cm^{-1}, although a weak satellite peak is sometimes observed at lower

frequency. [For a survey, see Thomas (1987).] A definitive explanation for the anomalously low frequency (2549 cm^{-1}) of the major S–H band of the tailspike, as well as the rather high frequency (2586 cm^{-1}) of the minor band component, is now being developed (H. Li and G. J. Thomas, Jr., unpublished results). Preliminary results indicate that the 2549-cm^{-1} frequency is diagnostic of strong hydrogen bonding of the S–H donor to a highly electronegative acceptor, such as an amide carbonyl or side-chain C=O acceptor (Asp, Asn, Glu, and Gln are examples of the latter). The 2586-cm^{-1} frequency, on the other hand, is believed to originate from S–H groups that do not participate in such hydrogen bonding.

When the tailspike is transferred from H_2O to D_2O solvent, the S–H groups are not readily exchanged to S–D groups, as judged by the absence of Raman intensity in the 1800–1900-cm^{-1} region, where S–D stretching vibrations normally occur (Thomas et al., 1982). We find no evidence of deuterium exchange of S–H even after 72 h in D_2O. This result indicates that both the strongly hydrogen-bonded sulfhydryls (i.e., those S–H groups generating the band at 2549 cm^{-1}) and the remaining sulfhydryls (generating the band at 2586 cm^{-1}) are inaccessible for solvent-mediated hydrogen exchange. Rigidity of the tailspike tertiary structure may be a major factor disfavoring both the exchange of sulfhydryls at physiological conditions and the formation of disulfide bridges at elevated temperatures.

(5) *Effects of Chemical Denaturants.* The extraordinary resistance of the tailspike to thermal denaturation prompted us to attempt chemical denaturation as a means of examining possible unfolding pathways. Among conventional denaturants, 8 M urea is poorly suited for Raman spectroscopy, and 6 M guanidine hydrochloride (Gdn·HCl) provides an acceptable window in the structurally informative amide III region but not in the complementary amide I region. More suitable for Raman spectroscopy is a 6 M LiBr solution (or 6 M LiCl), which has been employed in Raman denaturation studies of lysozyme (Chen et al., 1974). We attempted to use 6 M Gdn·HCl, 6 M LiBr, and 6 M LiCl as solvents for denaturation of the tailspike protein. All of these efforts were only partially successful. We obtained with each denaturing solvent a white precipitate, suspended in supernatant, from which only scant Raman signals could be collected. The precipitated material was not resolubilized by dialyzing out the chemical denaturants, indicating that it represented denatured and aggregated chains. The data from these specimens are of interest for comparison with the native tailspike.

The Raman spectra of the denatured proteins are shown in Figure 6. In the case of the Gdn·HCl precipitate (Figure 6a), only the strongest Raman bands of the protein are visible, including amide III, which is centered near 1244 cm^{-1}. (The band near 1663 cm^{-1} is due partly to amide I and partly to Raman interference from the denaturant solution.) On the basis of the shifted amide III marker, from 1235 cm^{-1} in the native state to 1244 cm^{-1} in the Gdn·HCl solution, we conclude that the tailspike secondary structure has been greatly altered. The amide III band position as well its large breadth indicate a substantially disordered chain (random coil). This Gdn·HCl-denatured protein is structurally distinct from that produced by thermal denaturation, where virtually no change to the protein secondary structure was observed (Sargent et al., 1988). The tyrosine doublet of the Gdn·HCl-denatured tailspike is also noticeable above the level of background noise in Figure 6a. It is clear that the doublet intensity ratio (I_{852}/I_{830}) has changed considerably from the value of 1.25 observed in the native tailspike (Figure 2) and in the thermally

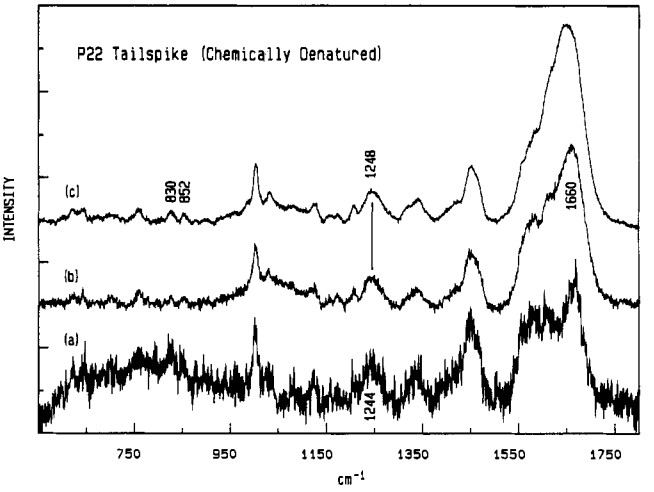

FIGURE 6: Raman spectra of the P22 tailspike protein (wild type) in the following chemical denaturants: (a) 6 M Gdn·HCl solution, (b) 6 M LiCl solution, and (c) 6 M LiBr solution, all at 20 °C. Labels indicate amide bands and tyrosine bands that are greatly altered from their positions and/or relative intensities in the native protein (cf. Figure 2).

denatured tailspike (Sargent et al., 1988).

Examination of the LiBr and LiCl precipitates (Figure 6, spectra b and c, respectively) shows that they are similar to one another yet somewhat different from the Gdn·HCl precipitate (Figure 6a). One of the most striking aspects of these data is the still larger change in the amide III region, where the sharp 1235-cm^{-1} amide III band of the native protein has been replaced by a very broad band centered near 1248 cm^{-1}. This indicates that in the presence of lithium halides a larger portion of the native β-sheet structure is lost and converted to the random coil form. Also for the lithium halide precipitates, the tyrosine doublet undergoes an intensity reversal such that the 852-cm^{-1} component is considerably less intense than the 830-cm^{-1} component; i.e., $I_{852}/I_{830} \approx 0.3$. This change may be interpreted as a net conversion of all tyrosine residues to roles of exclusive hydrogen-bond donor, presumably to the denaturant halide ion acceptor Br^- or Cl^-. Such a structural transition requires unfolding of all domains that contain Tyr residues. Since the 27 Tyr residues are distributed widely through the subunit (Sauer et al., 1982), we may conclude that the subunit is extensively unfolded rather than conservatively unfolded as in the case of thermal denaturation (Sargent et al., 1988). Consistent with this interpretation is the shift of the amide I band center to ca. 1660 cm^{-1} in the denatured tailspike, which is well below the value of 1668 cm^{-1} observed for the native tailspike. This is most apparent for the LiCl suspension (Figure 6b), where the supernatant interference has been reasonably well compensated.

In summary, the degree of denaturation of secondary structure (i.e., loss of β-sheet) is least for thermal perturbation (to 90 °C), intermediate for Gdn·HCl treatment (6 M), and greatest for lithium halide treatment (6 M). The same order is observed with respect to the degree of exposure of tyrosine residues at the surface of the denatured tailspike. Finally, we note that the chemical denaturants employed here do not result in the appearance of significant Raman intensity in the 500–550-cm^{-1} region (data not shown), indicating no significant disulfide bond formation. This may be due in part to the reducing agent (1 mM β-mercaptoethanol) present in the denaturing solutions. However, the 10-fold excess of cysteine over mercaptoethanol SH groups in these solutions suggests that the reducing agent does not play an exclusive role. The denaturations by Gdn·HCl, LiCl, and LiBr, like thermal de-

naturation (Sargent et al., 1988), apparently do not introduce disulfide bridges among or within subunits, consistent with results reported previously for LiBr denaturation of lysozyme (Chen et al., 1974).

(*6*) *Comparison of Observed and Predicted Secondary Structures.* The Raman spectrum does not give the distribution of secondary structures through the amino acid sequence. As an initial approach to assessing those sequences for which β-sheet or alternative structures are statistically favored, we have applied several secondary structure prediction algorithms (Garnier et al., 1978; Chou & Fasman, 1978; Burgess et al., 1974; Nagano, 1977) to the known protein sequence (Sauer et al., 1982). All prediction methods gave qualitative agreement with the Raman results, indicating strand (β-sheet) as the sequence-preferred secondary structure. Depending upon the prediction method, between 40 and 60% of the subunit is predicted to be in the strand conformation. A substantial amount of turns (16–42%) is also predicted by all methods. An assessment of these results with regard to the locus of mapped *tsf* mutations (Villafane & King, 1988) will be considered in a future publication.

Discussion and Conclusions

Since backbone and side-chain vibrations of many of the residues of proteins in solution are represented in the Raman spectra, conformational changes smaller than the overall denaturation transition should be detectable. Specifically, with the experimental protocols employed in the present work, the lower limit of detectability of changes in secondary structure through Raman amide I and amide III profiles of Figure 3 is 2.5 ± 0.3% (Thomas, 1989; Prevelige et al., 1990), which corresponds to about 17 residues of the tailspike subunit. Our measurements show that, within this limit of sensitivity, the mutant proteins (tsH302, tsU24, and tsU38) are not conformationally different from the wild type. Similar limits of sensitivity apply to Raman bands that monitor side-chain interactions.

The assignment and analysis of Raman bands due to specific amino acid side chains have permitted the average hydrogen-bonding environments of tyrosine (phenolic OH), tryptophan (indole NH), and cysteine (thiol SH) residues to be determined for the native tailspike. These amino acids represent about 6% of the sequence and are distributed throughout the polypeptide chain (Sauer et al., 1982). In the native structure, we find the interaction of the typical tyrosine OH group to be that of both hydrogen-bond donor and hydrogen-bond acceptor. We also find that indole NH groups of tryptophan residues donate moderately strong hydrogen bonds to electronegative acceptors. Tryptophans are not buried in hydrophobic environments. The majority of sulfhydryl groups of cysteines donate strong hydrogen bonds, probably to carbonyl acceptors. The S–H groups are sequestered from access to or direct hydrogen bonding with solvent H_2O molecules and are therefore resistant to deuterium exchange in D_2O solvent. These elements of the tailspike native *tertiary* structure are as thermostable as the native *secondary* structure.

Raman spectra show that both Gdn·HCl (6 M) and LiCl or LiBr (6 M) denature the secondary and tertiary structures of the tailspike, the latter denaturants being somewhat more effective than the former. The chemical denaturations produced highly insoluble aggregates rather than soluble unfolded chains at the concentrations employed. In these aggregates, much of the native β-sheet secondary structure was replaced by disordered or random coil conformations. Also, conformation-sensitive Raman markers of the tyrosines indicate that the phenolic hydroxyls donate strong hydrogen bonds to electronegative acceptors, presumably the denaturant halide ions, which suggests a relatively unfolded tertiary conformation for the chemically denatured tailspike aggregates. Earlier studies (Sargent et al., 1988) have shown that the thermal denaturation of the tailspike at 90 °C resulted in conservative changes to both the secondary and tertiary structures despite aggregation of the products of thermal denaturation. Thus, the aggregates resulting from chemical denaturations are structurally distinct from those of thermal denaturation discussed previously.

The present results demonstrate that the median temperature (80 °C) of secondary structure stability in mutants tsU24 (Ile258 → Leu) and tsH302 (Gly323 → Asp) is the same as that observed in the wild-type tailspike. This coincides with the finding of similar melting temperatures as determined by differential scanning calorimetry. A somewhat lesser thermostability (60 °C) was observed previously for the mutant tsU38 (Gly435 → Glu) (Sargent et al., 1988). Secondary structures of the three *tsf* mutants are thus thermostable to well above the temperature (40 °C) at which the folding defects are apparent.

The predicted secondary structure of the subunit sequence is in general accord with the structural features of the soluble tailspike trimer as deduced from the Raman spectrum. This structure is characterized experimentally (Sargent et al., 1988) by a predominance of β-sheet (56 ± 5%) and turns (17 ± 2%). A search for amphipathic β-sheet domains in the subunit sequence by Fourier transform methods (Finer-Moore & Stroud, 1984) was inconclusive. Some domains of amphipathic β-sheet periodicity are revealed, but they appear to be insufficient to account for the high β-sheet content of the native structure. No clear-cut pattern of repetitive amphipathic structure has emerged, such as the β-sheet that has been proposed for the fiber protein of adenovirus (Green et al., 1983). This leads us to suspect that maturation of the β-sheet secondary structure in the native tailspike requires trimerization of the subunit.

There are a number of examples of amino acid sequences that are important for folding or chain association but are subsequently removed from the native structure. These include the registration peptides of collagen and the cleaved sequences of proinsulin and prosubtilisin (Zhu et al., 1989). In these cases, sequences necessary for folding are irrelevant for stability. Although it has generally been assumed that all residues in the native structure make some contribution to stability, comparative sequence analysis (for example, in the hemoglobins) reveals that at many positions in the chain there can be great sequence variation, indicating that the residues involved are not critical for stability. The *tsf* mutants of the P22 tailspike identify residues that are critical for the folding pathway but are not critical to the stability of the native structure. In a previous paper (Sargent et al., 1988) we showed that one of these mutants, tsU38, identified a *tsf* site that also contributed to the stability of the native structure, though to a lesser extent than its apparent contribution to the productive folding pathway. The two mutants studied here, tsU24 and tsH302, are representative of the majority of the *tsf* mutants—they identify sites crucial to chain folding but not of significance to stabilization of the native conformation. The structural analysis indicates that the side chains at these sites make no contribution to the native structure, even though they disrupt the folding pathway. Recognition of the existence of such positions in amino acid sequences, important for folding but irrelevant for final structure, will be important in efforts to decipher the folding information coded in the sequence.

ACKNOWLEDGMENTS

We thank Dr. Janet Finer-Moore and Prof. Robert M. Stroud (University of California, San Francisco) for making available computer programs to analyze protein side-chain amphipathy.

REFERENCES

Alber, T., & Matthews, B. W. (1987) in *Protein Engineering* (Oxender, D. L., & Fox, C. F., Eds.) pp 289–297, Alan R. Liss, New York.

Burgess, A. W., Ponnuswamy, P. K., & Scheraga, H. A. (1974) *Isr. J. Chem. 12*, 239–286.

Chen, M. C., & Lord, R. C. (1974) *J. Am. Chem. Soc. 96*, 4750–4752.

Chen, M. C., Lord, R. C., & Mendelsohn, R. (1974) *J. Am. Chem. Soc. 96*, 3038–3042.

Chou, P. Y., & Fasman, G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Finer-Moore, J., & Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 155–159.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Goldenberg, D., & King, J. (1981) *J. Mol. Biol. 145*, 633–651.

Goldenberg, D., & King, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3403–3407.

Goldenberg, D. P., Smith, D. H., & King, J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 7060–7064.

Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R., & McLachlan, A. D. (1983) *EMBO J. 2*, 1357–1365.

Haase-Pettingell, C., & King, J. (1988) *J. Biol. Chem. 263*, 4977–4983.

King, J. (1986) *Bio/Technology 4*, 297–303.

Lamba, O., Becka, R., & Thomas, G. J., Jr. (1990) *Biopolymers* (in press).

Li, Y., Thomas, G. J., Jr., Fuller, M., & King, J. (1981) *Prog. Clin. Biol. Res. 64*, 271–283.

Lord, R. C., & Yu, N.-T. (1970) *J. Mol. Biol. 50*, 509–524.

Miura, T., Takeuchi, H., & Harada, I. (1988) *Biochemistry 27*, 88–94.

Nagano, K. (1977) *J. Mol. Biol. 109*, 251–274.

Oxender, D. L., & Fox, C. F., Eds. (1987) *Protein Engineering*, Alan R. Liss, New York.

Prescott, B., Sitaraman, K., Argos, P., & Thomas, G. J., Jr. (1985) *Biochemistry 24*, 1226–1231.

Prevelige, P., Thomas, D., King, J., Towse, S., & Thomas, G. J., Jr. (1990) *Biochemistry* (in press).

Roder, H., Elove, A. G., & Englander, S. W. (1988) *Nature 335*, 700–704.

Sandberg, W. S., & Terwilliger, T. C. (1989) *Science 245*, 54–57.

Sargent, D., Benevides, J. M., Yu, M.-H., King, J., & Thomas, G. J., Jr. (1988) *J. Mol. Biol. 199*, 491–502.

Sauer, R. T., Krovatin, W., Poteete, A. R., & Berget, P. B. (1982) *Biochemistry 21*, 5811–5815.

Siamwiza, M. N., Lord, R. C., Chen, M. C., Takamatsu, T., Harada, I., Matsuura, H., & Shimanouchi, T. (1975) *Biochemistry 14*, 4870–4876.

Smith, D. H., & King, J. (1981) *J. Mol. Biol. 145*, 653–676.

Smith, D. H., Berget, P. B., & King, J. (1980) *Genetics 96*, 331–352.

Sturtevant, J. M., Yu, M.-H., Haase-Pettingell, C., & King, J. (1989) *J. Biol. Chem. 264*, 10693–10698.

Thomas, G. J., Jr. (1987) in *Biological Applications of Raman Spectroscopy, Vol. 1, Raman Spectra and the Conformations of Biological Macromolecules* (Spiro, T. G., Ed.) pp 135–201, Wiley-Interscience, New York.

Thomas, G. J., Jr. (1989) in *Proceedings of the Third European Conference on the Spectroscopy of Biological Molecules* (Bertoluzza, A., Fagnano, C., & Monti, P., Eds.) pp 3–10, Società Editrice Esculapio, Bologna, Italy.

Thomas, G. J., Jr., & Barylski, J. (1970) *Appl. Spectrosc. 24*, 463–464.

Thomas, G. J., Jr., & Agard, D. A. (1984) *Biophys. J. 46*, 763–768.

Thomas, G. J., Jr., Li, Y., Fuller, M. T., & King, J. (1982) *Biochemistry 21*, 3866–3878.

Thomas, G. J., Jr., Prescott, B., & Day, L. A. (1983) *J. Mol. Biol. 165*, 321–356.

Thomas, G. J., Jr., Prescott, B., & Benevides, J. M. (1986) *Biomol. Stereodyn. 4*, 227–254.

Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.

Verduin, B. J. M., Prescott, B., & Thomas, G. J., Jr. (1984) *Biochemistry 23*, 4301–4308.

Villafane, R., & King, J. (1988) *J. Mol. Biol. 204*, 607–619.

Yu, M.-H., & King, J. (1988) *J. Biol. Chem. 263*, 1424–1431.

Zhu, X., Ohta, Y., Jordan, F., & Inouye, M. (1989) *Nature 339*, 483–484.