The interdigitated β-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization

JASON F. KREISBERG,[1] SCOTT D. BETTS,[2] CAMERON HAASE-PETTINGELL, AND JONATHAN KING

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(RECEIVED August 15, 2001; FINAL REVISION December 6, 2001; ACCEPTED December 6, 2001)

Abstract

The P22 tailspike adhesin is an elongated thermostable trimer resistant to protease digestion and to denaturation in sodium dodecyl sulfate. Monomeric, dimeric, and protrimeric folding and assembly intermediates lack this stability and are thermolabile. In the native trimer, three right-handed parallel β-helices (residues 143–540), pack side-by-side around the three-fold axis. After residue 540, these single chain β-helices terminate and residues 541–567 of the three polypeptide chains wrap around each other to form a three-stranded interdigitated β-helix. Three mutants located in this region — G546D, R563Q, and A575T — blocked formation of native tailspike trimers, and accumulated soluble forms of the mutant polypeptide chains within cells. The substitutions R563Q and A575T appeared to prevent stable association of partially folded monomers. G546D, in the interdigitated region of the chain, blocked tailspike folding at the transition from the partially-folded protrimer to the native trimer. The protrimer-like species accumulating in the G546D mutant melted out at 42°C and was trypsin and SDS sensitive. The G546D defect was not corrected by introduction of global suppressor mutations, which correct kinetic defects in β-helix folding. The simplest interpretation of these results is that the very high thermostability (T_m = 88°C), protease and detergent resistance of the native tailspike acquired in the protrimer-to-trimer transition, depends on the formation of the three-stranded interdigitated region. This interdigitated β-helix appears to function as a molecular clamp insuring thermostable subunit association in the native trimer.

Keywords: Tailspike; endorhamnosidase; protein folding; protein assembly; oligomerization; molecular clamp

The stabilities of fibrous proteins such as coiled-coils and collagen triple helices are associated with the coiling and intertwining of their polypeptide chains. Among globular proteins, the assembly of a number of oligomeric proteins is mediated through the formation of hybrid β-sheets with strands contributed by two or more subunits. Examples include the assembly intermediates formed between the FimC periplasmic chaperone and the FimH adhesin (Choudhury et al. 1999), the tailspike protein of phage P22 (Steinbacher et al. 1994), the adenovirus tail fiber (van Raaij et al. 1999), and numerous 3D domain-swapped dimers (Schlunegger et al. 1997). Various models of protein assembly and aggregation incorporate interchain β-sheet formation, including the polymerization of *Escherichia coli* pilin subunits (Choudhury et al. 1999) and the polymerization of amyloidogenic proteins in the formation of amyloid fibrils (Sunde and Blake 1997).

The P22 tailspike protein (Fig. 1) is a homotrimer of polypeptide chains, each 666 amino acids long. Each sub-

Reprint requests to: Jonathan King, Department of Biology 68-330, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA; e-mail: jaking@mit.edu.; fax: (617) 252-1843.

[1]Present address: Department of Physiology, Box 0444, University of California, 513 Parnassus Ave., San Francisco, California 94143, USA.

[2]Present address: Syngenta Biotechnology, Inc., 3054 Cornwallis Rd., Research Triangle Park, North Carolina 27709, USA.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.3440102.

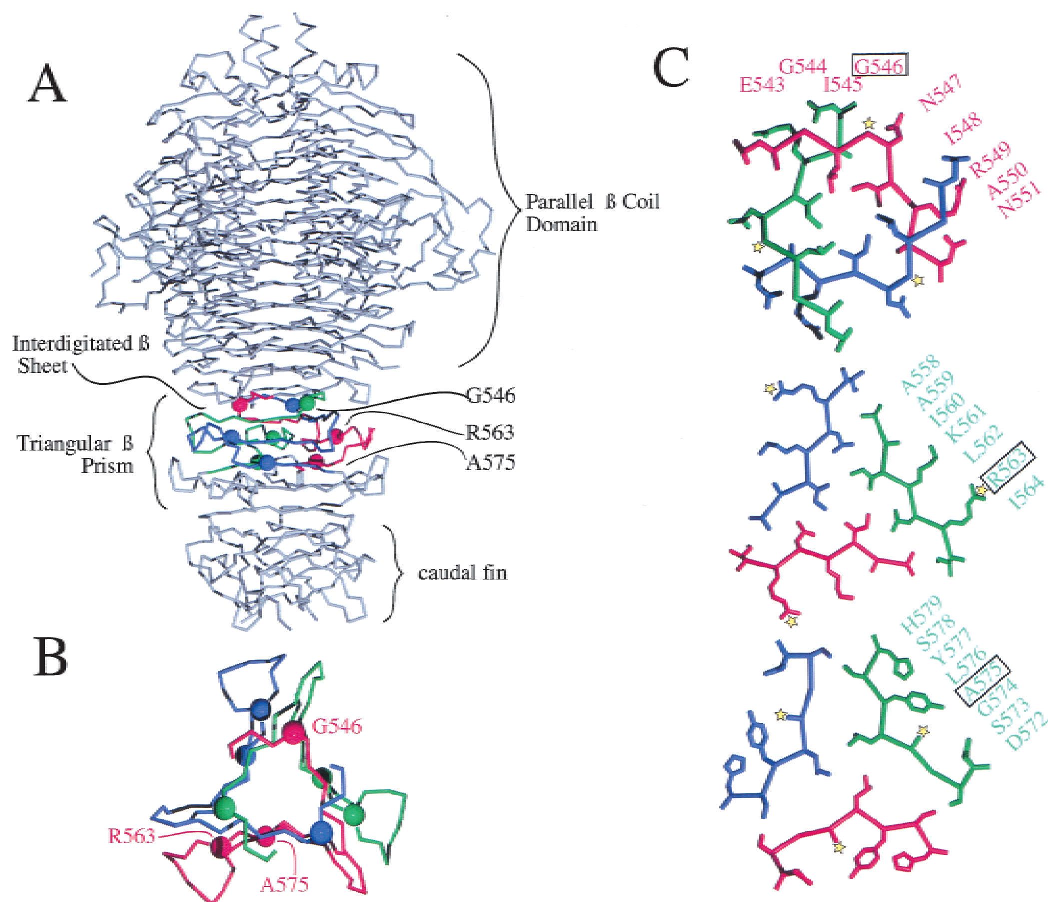


Fig. 1. Crystal structure of P22 tailspike protein. (*A*). α-carbon backbone of the C-terminal fragment (1 tsp) (Steinbacher et al. 1994). Each chain is colored from residues 543–579 (blue, red, and green), which include the oligomeric β-helix domain. The α-carbons of residues Gly 546, Arg 563, and Ala 575 are displayed as spheres with van der Waals dimensions. (*B*) Axial view of the triple β-helix domain from the N terminus toward the C terminus. The coloring scheme is the same as in *A*. (*C*) Axial views of the three stacked coils that form the structure in *B*. Isoleucine and leucine side chains from each subunit can be seen at the center of each coil forming the hydrophobic core of the interdigitated domain. Boxed residues are the sites of substitutions studied in this work. Stars note the positions of these mutations on each chain. Sequences in connecting loops and turns are omitted. The orientation of each view of *C* is the same as in *B*.

unit contains an active site that binds and hydrolyzes the O-antigen lipopolysaccharide of *Salmonella typhimurium*. The active site is located on the surface of the right-handed parallel β-helix domain, an elongated structure of 400 amino acids tightly packed in 13 stacked coils (Baxa et al. 1996). In the trimer, these domains are packed side-by-side and associated through predominantly polar and hydrophilic contacts (Steinbacher et al 1994; Seckler 1998). At the C-terminal end of this threefold region, the three polypeptide chains are intertwined and coiled around a single central axis, forming the braid-like structure shown in color in Figure 1A (Steinbacher et al. 1994). An axial view of this triple β-helical domain (Seckler 1998), is presented in Figure 1B.

In the walls of this interdigitated region, the three polypeptide chain backbones are hydrogen bonded to each other, forming parallel sheets (Steinbacher et al 1994). Each subunit contributes one strand to each of three sheets (Fig. 1B). The side chains pointing in from the interdigitated strands form a single buried hydrophobic core. Thus, in this region of the molecule, its structure moves from having three β-helices, each with their own buried cores, to the interdigitated region with a single buried core formed from side chains contributed by all three chains (Fig. 1C). The pattern of side-chain packing within the core of this three-chain structure is very similar to that observed within the core of the left-handed parallel β-helix, a single-chain conformation found in acyltransferases and other proteins (Kreisberg et al. 2000).

After the interdigitated region, the tailspike strands stop coiling around a central axis and switch to folding back on themselves into anti-parallel β-sheets. In this β-prism structure, the β-strands remain transverse to the long axis of the

trimer, and their side chains still contribute to a single buried hydrophobic core (Kreisberg et al. 2000). This is a very different subunit interface than is found among the single β-helices of residues 143–540 or between the subunits of globular proteins.

Partially folded monomeric and trimeric intermediates to the P22 tailspike have been identified both in vivo and in vitro (Fuchs et al. 1991; King et al. 1996; Betts and King 1999). These folding intermediates are indicated in the upper productive pathway in Figure 2. The earliest folding intermediate shown is the thermolabile single-chain species [I]. This species is partially melted or disordered at temperatures above ~35°C, which is within the upper physiological range of bacterial growth (Haase-Pettingell and King 1988, 1997). At lower, permissive temperatures, this monomeric species folds further and associates to form the partially folded protrimer intermediate. The protrimer is both trypsin sensitive and SDS sensitive, in contrast to the fully folded native conformation (Goldenberg and King 1982; Fuchs et al. 1991). The protrimer can be trapped in the cold and resolved from the native trimer by native polyacrylamide gel electrophoresis (Goldenberg and King 1982 Fuchs et al. 1991; Betts and King 1998; Betts et al. 1999). In native gels and size-exclusion columns, the migration of the partially folded protrimer is retarded compared with the native tailspike, suggesting that it has a greater radius of gyration (Fuchs et al. 1991; Benton et al. 2002)

The native tailspike lacks disulfide bonds. Nonetheless, the protrimer is stabilized by transient interchain disulfide bonds, which are reduced in the formation of the native trimer (Robinson and King 1997; Raso et al. 2001). The mechanism of these coupled oxidation-reduction reactions are not yet known, but the tailspike does contain a C-X-X-C thioredoxin sequence motif comprising cysteines 287 and 290.

Spectroscopic observations indicate that the partially folded monomeric intermediate [I] has a major folded β-helix domain (Fuchs et al. 1991; Danner and Seckler 1993). However, the region that will be interdigitated in the native trimer cannot be in the native fold in the monomeric species.

At high temperatures, [I] partially unfolds or misfolds reversibly to the off-pathway conformation denoted [I*] in the lower part of Figure 2 (Haase-Pettingell and King 1988). Self-association of the [I*] species results in the formation of [I*] multimers that are the early, soluble intermediates on the inclusion body pathway (Speed et al. 1995; Betts et al. 1999). Temperature-sensitive folding (tsf) mutations identify residues that are required for productive tailspike folding in the higher end of the temperature range for bacterial growth, >37°C (Haase-Pettingell and King 1997; Betts et al. 1997; Schuler and Seckler 1998). They appear to act by further destabilizing the thermolabile monomeric intermediate [I] and populating the [I*] aggregating species (Fig. 2). Global suppressor mutation suppress the tsf folding defects by inhibiting off-pathway aggregation (Mitraki et al. 1991b, 1993; Danner and Seckler 1993; Beissinger et al. 1995). All sites of tsf mutations identified to date are restricted to the β-helix domain between residues 143 and 540 (Haase-Pettingell and King 1997; C. Haase-Pettingell and J. King, unpubl.).

Relatively little is known about the amino acid sequences that directly control the folding and maturation of the protrimer to the native trimer. Schwartz and Berget (1989) screened for mutants of the tailspike that interfered with tailspike formation. They isolated three mutants that mapped to amino acid residues within the interdigitated β-helix and triangular prism domain. These mutations blocked the formation of native, SDS-resistant tailspikes but allowed the mutant chains to accumulate in a soluble, but biologically inactive, conformation. The sites of these mutations — G546D, R563Q, and A575T — are indicated by the spheres in Figure 1, A B, and star in C.

The very high stability of the tailspike is only found in the mature trimer and is acheived in the protrimer to mature

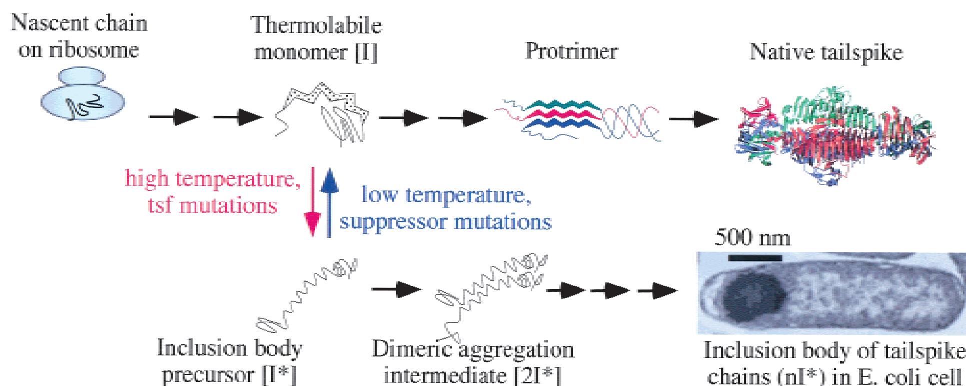


Fig. 2. Competing folding and inclusion body pathways of the P22 tailspike protein. Previously characterized mutants affected the partially folded monomer and its partitioning to the inclusion body pathway (King et al. 1996; Seckler 1998). The mutants described in this work interfere with later steps. Alhough the protrimer intermediate is shown with the β-helices docked as in the native structure, the organization of its β-helices is unclear. The ribosome-bound state of the tailspike chains is described in Clark and King (2001) and the off-pathway intermediates in Speed et al. (1996, 1997). Reproduced from Betts and King (1999) with permission.

trimer transition. Removal of the N-terminal domain still permits formation of a highly stable trimer (Chen and King 1991; Danner and Seckler, 1993; Miller et al. 1998b). However, due to uncertainty over the structure of the protrimer, it was unclear which of the four remaining domains was conferring the high stability to the native trimer; β-helix, interdigitated β-helix, β-prism, or caudal sheets. Characterization of the species accumulated by site-specific substitutions in the interdigitated helix/β-prism region support a model in which formation of the interdigitated and β-prism regions of the trimer provide the sharply increased stability of the native trimer (Sturtevant et al. 1989).

Results

Mutant chains accumulate as soluble non-native species

Tailspike polypeptides containing the single amino acid substitutions G546D, R563Q, and A575T were expressed in *E. coli* using the pET vector expression system. The neutral amino acid substitution G546N was also constructed. Cells were grown at 30°C and harvested 2 h after induction of gene expression with IPTG. The distribution of polypeptide chains between the pellet and supernatant fractions of cell lysates is shown in the SDS gel in Figure 3A. The first two lanes show insoluble pellet and soluble supernatant proteins (between 68 and 200 kD) from host *E. coli* cells containing the pET vector only; these lanes represent the background of host proteins. In the cells expressing the wild-type tailspike gene, tailspikes accumulated primarily as soluble SDS-resistant trimers. About 12% of the wild-type chains formed insoluble inclusion body aggregates; these dissociated in the presence of SDS and migrated as SDS/tailspike chain complexes just above the 68-kD marker in the pellet fraction.

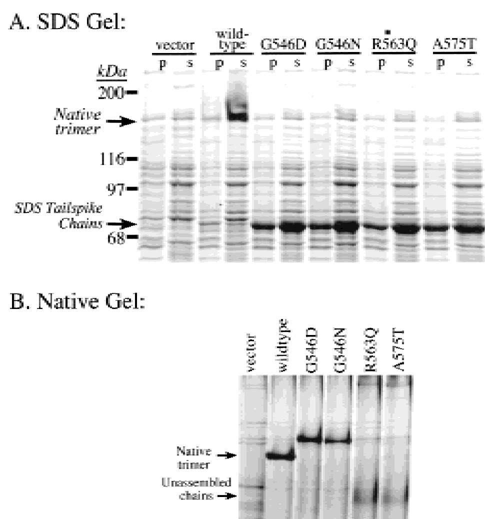


Fig. 3. In vivo expression of tailspike mutants. Recombinant tailspike chains were expressed in *E. coli* for 2 h at 30°C. Pellet and supernatant fractions of cell lysates were analyzed by SDS-PAGE and stained by Coomassie blue. (*A*) SDS gel (7.5% acrylamide) showing pellet (P) and supernatant (S) fractions. Vector fractions are from culture containing pET vector without gene 9 insert. (*B*) Native gel of the supernatant samples (7.5% acrylamide) stained with Coomassie blue.

A majority of mutant tailspike chains in each mutant lysate were soluble, remaining in the supernatant after centrifugation (Fig. 3A). This was in contrast to the behavior of chains carrying tsf substitutions, which efficiently form inclusion bodies (Haase-Pettingell and King 1988, 1997). Although the mutant chains were predominantly soluble, they did not migrate as native trimers, but as SDS/tailspike chain complexes. The fraction of soluble SDS-sensitive chains was 89% for G546D, 80% for G546N, 75% for R563Q, and 60% for A575T. This indicates that the chains accumulating within the *E. coli* cells, had not reached the detergent-resistant trimeric state.

The conformations of soluble mutant tailspike species in the lysate supernatants shown in Figure 3A were analyzed by native gel electrophoresis in Figure 3B. Samples analyzed in Figure 3B are portions of the same supernatant fractions analyzed in Figure 3A. In cells expressing the wild-type tailspike gene, native tailspikes were the predominant soluble species (Figure 3B, wildtype).

Tailspike polypeptides containing the G546D and G546N substitutions accumulated as a single well-resolved species with retarded electrophoretic mobility relative to wild-type tailspikes (Fig. 3B). The similar migration of the G546D and G546N tailspike species shows that the mobility shift is not due to a charge change at position 546, but is most likely attributable to a conformational change induced by the substitutions.

As shown in Figure 3B, tailspike chains containing the R563Q and A575T mutations formed neither native trimers nor protrimer-like species. However, the gel lanes reveal increased staining intensity in the region of the gel in which unassembled, partially folded chains migrate (Betts and King 1998). Western blot analysis confirmed that the rapidly migrating R563Q and A575T mutant species were tailspike chains (data not shown). Minor, very slowly migrating multimeric species also appear in these two lysates, perhaps representing an aggregated state of these chains (Speed et al. 1995).

The mobility of the G546D species is protrimer-like

The G546D species migrated with retarded mobility compared with mature trimers in native gels, a property also associated with the productive protrimer (Goldenberg and King 1982; Goldenberg et al. 1983). To carefully compare its mobility with the wild-type protrimer, wild-type tailspike was refolded in vitro to generate the protrimer intermediate.

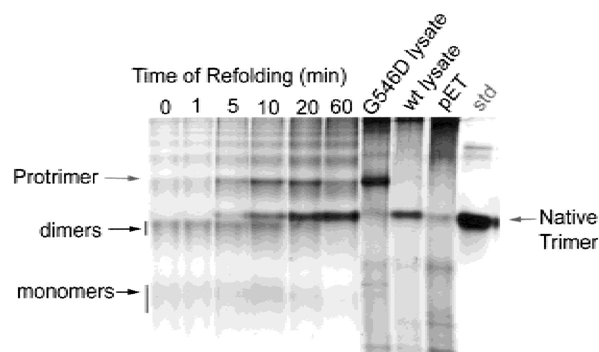


Fig. 4. The G546D species has protrimer-like mobility. Samples of refolding wild-type phage tailspike (*left lanes*) were electrophoresed through a native polyacrylamide gel along side cell lysates to permit comparison of the G546D species with the wild-type protrimer intermediate. These samples were electrophoresed in the absence of added reducing agents. The gels resolve both the productive intermediates and the off-pathway intermediates leading to the aggregated inclusion body state. The *right*-most lane labeled std was loaded with 100 μg/mL-purified tailspike protein. See Materials and Methods for details.

Wild-type and 546D lysates without reducing agent were electrophoresed through the same native gel as the refolded samples (Fig. 4).

The left lanes of the gel display the wild-type refolding samples at various times after dilution out of denaturant. These gels resolve both productive intermediates and multimeric intermediates along the aggregation pathway (Speed et al. 1996, 1997). Early during refolding, the most prominent bands were partially folded monomers and dimers. These bands disappeared with time as they folded and assembled into productive protrimers and native tailspikes, as well as nonproductive higher molecular weight multimeric intermediates along the aggregation pathway.

The four right lanes have the G546D lysate, wild-type lysate, the pET lysate control, and native wild-type standard (100 μg/mL). The native tailspike trimer was the most prominent band in the wild-type (wt) lysate. The lane containing the G546D lysate lacks the native trimer, but displays a prominent band of retarded mobility. The mobility of this major G546D band corresponds very closely to that of the wild-type protrimer, seen most clearly in the 10 and 20 min refolding samples. This indicates that the G546D species is a trapped protrimer-like intermediate.

To be sure that the 546D species was trimeric, lysates were centrifuged through a sucrose gradient, and the fractions analyzed by native gel electrophoresis. The G546D species did sediment similarly to native tailspikes, consistent with a trimeric state (data not shown).

The G546D tailspike species did not behave like wild type when subjected to our standard purification. The standard protocol for native tailspikes involves separation from lysate supernatants by differential centrifugation, salting out, anion-exchange chromatography, and a final polishing step in which the anion-exchange fraction is passed through a hydroxyapatite column. The G546D species, unlike wild-type tailspikes, bound to hydroxyapatite. Elution required denaturing conditions.

Partially purified G546D species from anion-exchange chromotography were not stable at 4°C and converted within hours into multiple multimeric species of different electrophoretic mobilities as well as apparent degradation products. Therefore, subsequent characterization was carried out on the soluble protrimer-like G546D species present in the crude lysates.

The G546D tailspike species is trypsin sensitive

The experiments in Figure 3 revealed that the G546D and G546N tailspike species could be distinguished from wild-type tailspikes on the basis of electrophoretic mobility as well as stability in the presence of SDS. These two properties are characteristics of the protrimer folding intermediate.

The protrimer can also be distinguished from the native tailspike by its susceptibility to trypsin digestion (Goldenberg and King 1982). Therefore, we tested the trypsin sensitivity of the G546D tailspike species to see whether it shared this additional characteristic of the protrimer. Wild-type tailspikes in lysate supernatants were resistant to trypsin digestion at protease concentrations up to 1 mg/mL (Fig. 5A). In contrast, the G546D tailspike species was almost completely degraded within 10 min at a protease concentration of 100 μg/mL, and was completely degraded within 5 min at a protease concentration of 1 mg/mL (Fig. 5B). This result showed that the G546D tailspike species, like the protrimer, is susceptible to trypsin digestion.

The G546D tailspike species is thermolabile

Native tailspikes are thermostable and resist aggregation at or below their melting temperature of 88°C (Sargent et al. 1988; Sturtevant et al. 1989; Thomas et al. 1990; Kreisberg et al. 2000). The G546D tailspike species was incubated at 42°C in order to assess its stability at high temperature. Wild-type tailspikes in lysates from cultures grown at 30°C were unaffected by incubation at 42°C (Fig. 6A, leftmost lanes). In contrast, G546D tailspike species from cells

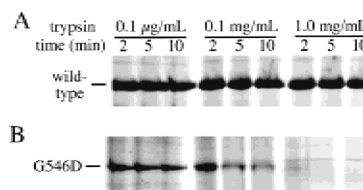
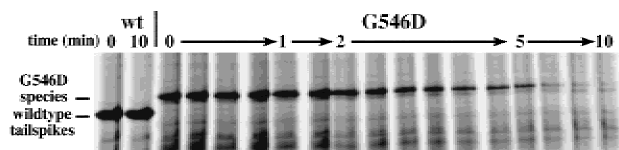


Fig. 5. Trypsin treatment of wild-type and G546D lysate supernatants. Samples were incubated in the presence of trypsin as indicated for up to 10 min, quenched by addition of trypsin inhibitor, and analyzed by nondenaturing gel electrophoresis. See Materials and Methods for details. Coomassie blue-stained gels of wild-type digests (*A*) and G546D digests (*B*).

A. Native Gel:



B. Quantitation:

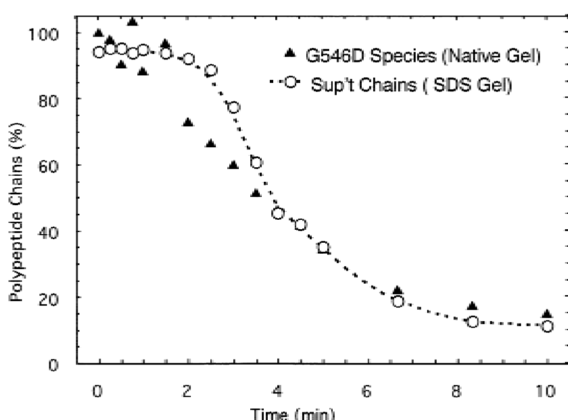


Fig. 6. Heat-induced aggregation of G546D tailspike chains. Lysate supernatants from cultures expressing wild-type and G546D chains were incubated at 42°C for 10 min. Portions of the reaction were chilled at intervals and either directly analyzed by native gel electrophoresis or further fractionated by centrifugation into pellet and supernatant fractions. The chains partitioned from the supernatant to the pellet with time of heating. (*A*) Coomassie blue-stained native gel (9% polyacrylamide) of chilled samples showing the disappearance of the G546D species over time. (*B*) Quantitative analysis of the loss of the G546D species due to aggregation . (▲) The disappearance of G546D species from native gels at 42°C measured by monitoring the staining intensity of the G546D gel band indicated in *A*. (○) Loss of total SDS soluble tailspike chains due to aggregation of the G546D species monitored by SDS-PAGE analysis of pellet and supernatant fractions (see Materials and Methods).

grown at 30°C were depleted from lysate fractions within 10 min during incubation at 42°C (Fig. 6A,B, triangles).

The G546D tailspike chains were not degraded at 42°C but were recovered in the pellet fraction of samples taken during the heat treatment. Figure 6B (○) shows the loss of total soluble G546D tailspike chains from these samples as measured by SDS gel electrophoresis. The chains were not degraded but partitioned into the pellet fractions. We conclude that the G546D tailspike species is in a non-native conformation with increased susceptibility to thermal aggregation with respect to the native conformation.

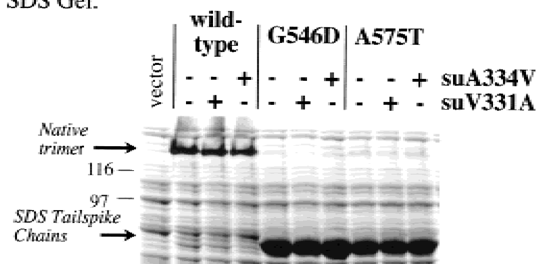
Global suppressors of tsf mutations do not suppress G546D or A575T in vivo at 30°C

The tsf mutations appear to destabilize the junctional folding intermediate [I], resulting in increased partitioning onto the inclusion body pathway. Genetic screens for second-site suppressors of primary tsf mutations repeatedly isolated two intragenic suppressor mutations, V331A and A334V, located in the middle of the β-helix domain (Fane et al. 1991; Mitraki et al. 1991b). These global suppressor mutations correct the folding defects of primary tsf mutations and also improve the folding efficiency of the wild-type protein at high temperatures (Mitraki et al. 1991a, 1993; Beissinger et al. 1995). The suppressors may act by stabilizing the β-helix in the thermolabile single-chain intermediate (Schuler and Seckler 1998).

We tested whether or not the global tsf suppressor mutations could suppress the G546D and A575T mutations and allow accumulation of SDS-resistant tailspikes. The wild-type protein and four double mutants (G546D/V331A, G546D/A334V, A575T/V331A, and A575T/V334A) were expressed in *E. coli* for 3 h at 30°C, and soluble and insoluble fractions of the cell lysates were analyzed by native gel electrophoresis. The results are shown in Figure 7.

The double mutant chains G546D/V331A and G546D/A334V did not form native trimers, but accumulated as soluble SDS/tailspike chains (Fig. 7A, middle lanes), similar to the G546D mutant chains lacking the suppressors. The double mutant A575T/V331A and A575T/A334V chains also failed to form native tailspikes, similar to the A575T

A. SDS Gel:



B. Native Gel:

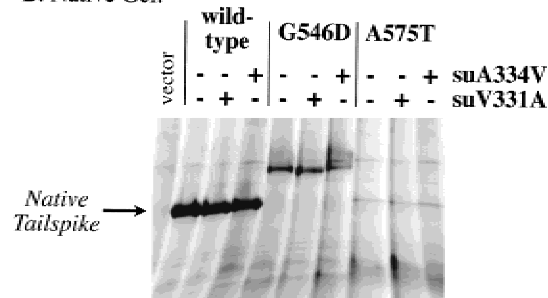


Fig. 7. Global tsf suppressor mutations (su V331A and A334V) do not rescue chains containing the G546D and A575T mutations. Recombinant tailspike chains were expressed in *E. coli* for 3 h at 30°C. Samples were processed as described in Materials and Methods and the supernatant fractions of cell lysates were analyzed by (*A*) SDS-PAGE, and (*B*) native PAGE.

substitution alone (Fig. 7A, right lanes). The mutant lysates were also analyzed by native gel electrophoresis. The G546D chains accumulated as protrimer-like species independently of the presence of the suppressors (Fig. 7B, middle lanes). The A575T chains accumulated as unassembled species, seven also independently of the presence of the suppressors (Fig. 7 B, right lanes). The misfolding of the double mutants in vivo shows that the global tsf suppressors do not correct the defects of these chains containing primary mutations in the triple β-helix domain.

Discussion

The interdigitated β-helix as a molecular clamp

Tailspike maturation involves the conversion of the SDS-sensitive and trypsin-sensitive protrimer folding intermediate to the heat-stable, SDS-resistant, and trypsin-resistant native trimer. The single amino acid substitutions examined here are residues in the triple β-helical domain of the tailspike trimer. In this region, the three polypeptide chain strands are interdigitated so as to be hydrogen bonded to each other and to form a single buried core. The mutants presumably prevent formation of the native trimer structure by interfering with chain association (A575T and R563Q) and protrimer maturation (G546D).

The G546D species had a retarded electrophoretic mobility compared with wild-type tailspikes, dissociated in SDS at room temperature, and was degraded by trypsin. These properties have each been reported previously for the protrimer tailspike folding intermediate (Goldenberg and King 1982). The G546D species also aggregated at 42°C with a half-time of ~4 min. In contrast, wild-type tailspikes resist aggregation at this temperature, as shown in Figure 6, and at temperatures up to 40°C higher (Sturtevant et al. 1989). The existence of single amino acid substitutions in this region that block the formation of the thermostable native trimer suggests that the intertwined region of the trimer acts as a molecular clamp, sharply increasing the stability of the native trimer with respect to the protrimer (Steinbacher et al. 1994; Miller et al. 1998a; Seckler 1998; Kreisberg et al. 2000).

The single buried interfaces between the three chains in this region of the trimer are very different from the hydrophilic interactions at the three β-helix subunit–subunit interfaces. The hydrophobic core of this domain consists of extensive side-chain interactions along the buried surfaces of the subunits as shown in detail in Figure 1C. In the upper rung, the side chains of I545 and I548 from each chain form a six-residue isoleucine ring. In the middle rung, the side chains of I560 and L562 from each subunit are closely packed, and in the lower rung the side chains of L576 from each subunit appear to interlock to form a tight ring.

The strands containing G546 and R563 hydrogen bond with the strands from the other subunits and are clearly part of the interdigitated β-sheet region. The strand containing A575 has folded back to form the antiparallel intrachain β-sheet just below the interdigitated sheet (Fig. 1). As noted above, the single buried core is continuous through this β-prism region.

All three of the mutant sites have their side chains on the solvent-associated surface of the triple β-helix. G546 occurs within a turn in the upper rung, and R563 and A575 occur within β-strands in the middle and lower rungs, respectively. In globular proteins, it is unusual for single amino acid substitutions on the protein surface to have such disruptive effects on folding and/or subunit assembly. In fact, the affect of these substitutions is more likely to be on the monomeric precursor chains, whose conformations in the 540–580 residue region must be quite different from the native trimer.

The N-terminal 1–100 amino acid domain of the tailspike contributes some stability to the native trimer (Chen and King 1991). However, truncated chains with the N terminus removed still form thermostable trimers (Chen and King 1991; Danner and Seckler 1993; Miller et al. 1998b). Seckler and coworkers have characterized isolated β-helix regions. These domains trimerize, but are not thermostable and are in equilibrium with the monomers under all conditions (Miller et al. 1998b).

The interdigitated regions of the chains cannot partner with each other until they are associated correctly in the protrimer. Because the isolated β-helix domain can trimerize, albeit not stably, the protrimer presumably forms through these interactions (Miller et al. 1998b; Schuler and Seckler 1998). The protrimer has a significantly larger molecular radius than the native trimer (Benton et al. 2002). This is most simply explained if the 540–580 regions of the chains have not yet wrapped around each other in the protrimer. If the subunits in the protrimer have already formed the interdigitated β-prism domain, it is difficult to account for the very significant differences between the properties of the protrimer and the stable native trimer.

The partially folded monomeric precursors cannot have their 540–580 chain regions in the native conformation. They must be in some partially folded conformation, perhaps with these regions folded on themselves. In this model, the protrimer-to-trimer transition would involve breaking bonds, as well as forming them, accounting for the high enthalpy of this transition (Danner and Seckler 1993; Seckler 1998).

The polarity of subunit assembly

As shown in Figure 3, the mutations R563Q and A575T each had the same effect on tailspike folding in vivo; monomeric partially folded mutant tailspike chains accumu-

lated in cell lysates. Such species do not accumulate in cells infected with tsf mutants, or to any significant extent in wild-type lysates. The two substitutions allowed folding to proceed to an aggregation-resistant single-chain conformation, but prevented chain association. This result suggests that R563Q and A575T either block or destabilize an early subunit–subunit interaction critical for stable chain association.

The mutations G546D and G546N caused accumulation of an oligomeric tailspike species with conformational properties in common with the protrimer folding intermediate. This substitution thus appears to act after the subunit–subunit interaction that is affected by the R563Q and A575T mutations, blocking conversion of the protrimer to the native conformation. In this model, the protrimer-to-trimer transition represents a wrapping or twisting reaction to generate the interdigitated regions in the native trimer.

One interpretation of these findings is that protrimer assembly does not initate through β-helix association as suggested above, but initiates from the C-terminal end of the chain and proceeds towards the N terminus, in the direction of the subunit β-helix domains. In this model, the R563Q and A575T substitutions would prevent initiation of subunit association. The more N-terminal G546D substitution would not prevent the association reaction, but interfere with the completion of the interdigitation to form the native trimer. The parallel-helix domain appears to be folded in the earliest tailspike intermediates. In this model, the subunit β-helix domains would remain free and not docked with each other until completion of the subunit association.

The partially folded monomer may not be as extended as it is in the native trimer. For example, the C-terminal regions could be docked against the β-helix domain in the partially folded monomer. In this case, the R563Q and A575T substitutions might affect the conformation of such monomeric intermediates, thus blocking productive chain assembly. The N- and C-termini of monomeric β-helix proteins fold back onto their β-helix surfaces in this manner. For example, the last 40 amino acids of pectate lyase C forms a large loop that partially covers the β-helix (Yoder et al. 1993). Similar folded-back termini are observed in other β-helix structures, including pectate lyase E (Lietzke et al. 1994) and rhamnogalacturonase A (1RMG) (Petersen et al. 1997).

The role of disulfide bonds

The protrimer uses temporary disulfide bonds for its stabilization, at least in the higher temperature range of phage growth (Haase-Pettingell et al. 2001). Evidence to date suggests that the most likely candidates for the interchain disulfide bonds in the protrimer are between the C613 and C635 residues that form a ring of six cysteines, with cysteine 613 of one chain closest to cysteine 635 of the neighboring chain (Robinson and King 1997; Haase-Pettingell et al. 2001). Chains carrying serine substitutions at both these cysteines are unable to form native tailspikes (Haase-Pettingell et al. 2001) or protrimer-like species (C. Haase-Pettingell and J. King, unpubl.). In the model described above, disulfide bonds would stabilize the protrimer prior to actual interdigitation and clamp formation.

Unambiguous determination of whether the G546D trimer contains interchain or intrachain disulfide bonds requires additional analysis. However, a protrimer-like species, clearly differentiable from wild type, was recovered independently of the presence of exogenous reducing agent.

An alternate role of the transient disulfide bonds may be to stabilize the conformation of the partially folded monomeric precusor prior to subunit assembly and interdigiatation. The disulfide bonds between the registration domains in procollagen play a similar role in stabilizing precursors prior to triple helix formaiton (Fessler and Fessler 1982; Engel and Prockop 1991; Baum and Brodsky 1999). In this model, the disulfide role may be to keep the precursor strands from interacting incorrectly prior to the correct docking and registration needed for the interdigitation rearrangement.

It is not clear why interchain disulfide bonds are reduced in the maturation to the native trimer. It may be that the interdigitation motions require somewhat extensive rearrangement reactions in the protrimer. Alternatively, the complex interactions between the phage and the cell surface needed for docking and DNA injection may require further molecular motions of the tailspike, which would be constrained by interchain disulfide bonds.

Thermolabile intermediates to a thermostable protein

One of the distinctive features of the tailspike is the contrast between the thermolability of its folding intermediates and the thermostability of its native state (Sturtevant et al. 1989; Miller et al. 1998a, 1998b; Seckler 1998). The interdigitated structure provides a rationale for this, as this feature cannot be formed in early intermediates. In particular the formation of a buried core with contributions from three separate polypeptide chains raises the question of the state of these residues in the partially folded intermediates.

Materials and methods

Materials and DNA sequencing

Gene 9, which encodes the tailspike protein, was amplified from P22 phage DNA by PCR and cloned using the pNOTA vector (5′–3′). The DNA sequence was as reported by Sauer et al. (1982). The clone was designated pNOTA(gene9). The full-length gene 9 insert was subcloned into the expression vector pET11 (Novagen). The resulting construct was designated pET(gene9) and encoded the full-length tailspike polypeptide chain without any extensions

or deletions. Plasmid DNA was purified using QiaFilter Miniprep and Midiprep kits (QIAGEN). DNA sequences were determined using the Sequenase Version 2 plasmid sequencing kit (Stratagene) or by automated DNA sequencing at the M.I.T. Biopolymers Lab.

Gene 9 cloning and mutagenesis

The sequences of the oligonucleotides used for site-directed mutagenesis are shown in Table 1. A silent endonuclease restriction site was introduced with each mutation. All mutations were introduced using the pNOTA9(gene9) vector as mutagenesis template and the QuikChange mutagenesis kit (Stratagene). Mutant clones were identified by restriction mapping and confirmed by DNA sequencing. The *Nhe*I–*Hin*dIII fragment containing the mutated sequence and encoding tailspike residue 275 through the stop codon was subcloned into pET(gene9), replacing the corresponding wild-type sequence. The complete sequence of the subcloned fragment, from *Nhe*I through the stop codon, was confirmed for pNOTA(G546D) and pNOTA(A575T). The DNA sequence of pNOTA(G546N) was confirmed from codon 452 through the stop codon, and the sequence of pNOTA(R563Q) was confirmed from codon 454 through the stop codon. The double mutants G546D/V331A, G546D/A334V, A575T/V331A, and A575T/A334V were constructed using pNOTA(G546D) and pNOTA(A575T) as mutagenesis templates. To ensure that the suppressor was present, the plasmids were sequenced from the *Nhe*I site (residue 375) through the following residue: Ser 597 for G546D/V331A, Leu 419 for G546D/A334V, the stop codon for A575T/V331A, and Pro 409 for A575T/A334V. The presence of each primary mutation was confirmed by restriction mapping.

Tailspike expression in E. coli

Recombinant tailspike polypeptides were overexpressed in *E. coli* BL21(DE3) using the pET vector system (Novagen). Cultures in LB broth supplemented with 50 μg/mL ampicillin (LB-Amp) were inoculated either with colonies or frozen cells and grown overnight at 37°C with shaking. A portion of the overnight culture was diluted 50-fold into fresh LB-Amp medium and incubated at 28 or 30°C with shaking (250 rpm). Recombinant protein expression was induced by addition of IPTG to 400 μM when cultures reached an optical density at 600 nm of ~0.6.

Table 1. *Nucleotide sequences of mutagenic primers*

Mutation	Primer complementary to sense strand	Silent restriction site
G546D	GCTAATTTGGCAGAAGAAGGCCTAGATAA TATCCGCGCTAATAGTTTCGG	*Stu*I
G546N	GCTAATTTGGCAGAAGAAGGCCTAAACAA TATCCGCGCTAATAGTTTCGG	*Stu*I
R563Q	GCGCAGCGATTAAACTGCAGATTCATAAG TTATCAAAGACATTAGATAGCG	*Pst*I
A575T	CAAAGACATTAGATAGCGGAACGTTGTAC AGCCACATTAACGGGGGGG	*Bsr*GI
V331A	CGAACCAGCTATGGGTCAGCCTCGAGCGC CCAGTTTTTACG	*Xho*I
A334V	GCTATGGGTCAGTAAGTAGTGTACAGTTT TTACGTAATAATGGTGGC	*Bsr*GI

After 2 h of further shaking at 30°C, the cells were harvested by centrifugation (7 min, 8000*g*, 4°C) and placed on ice. Cells were resuspended in ice-cold lysis buffer (50 mM Tris-HCl at pH 8.0, 25 mM NaCl, 2 mM EDTA) and stored at −20°C. Frozen suspensions were thawed at room temperature and immediately supplemented with PMSF (1 mM final concentration from 100 mM stock in ethanol) and DTT (2 mM final concentration from 1 M aqueous solution). Lysis was initiated by addition of freshly prepared lysozyme (100 μg/mL final concentration from 2 mg/mL aqueous solution) and incubated at room temperature for 30 min with occasional mixing. Lysed cells were incubated for an additional 20–30 min at room temperature after addition of DNase I (Sigma) (25 μg/mL final concentration from 0.5 mg/mL stock in 0.5 M $MgCl_2$).

Soluble and insoluble fractions of the cell lysates were separated by centrifugation at 4°C for 5 min at 17,000*g* for volumes <1.5 mL and 12,000*g* for higher volumes. Lysate pellets were resuspended in lysis buffer. Lysate fractions were kept on ice and either analyzed immediately or stored at −20°C.

Gel electrophoresis and laser densitometry

The protocol for native gel electrophoresis was as described in Betts et al. (1999). Lysate fractions were prepared for electrophoresis by dilution in one-half volume ice-cold native gel sample buffer (14 mM Tris base, 109 mM glycine, 30% glycerol, 50 mM DTT, and 0.1% bromphenol blue). Samples were kept on ice and electrophoresis was performed in a 4°C cold room using a Tris-glycine running buffer. Samples for SDS-PAGE were mixed with one-half volume SDS sample buffer (188 mM Tris at pH 6.8, 2% SDS, 15% β-mercaptoethanol, and bromphenol blue) and kept at room temperature until electrophoresis. The discontinuous buffer system of King and Laemmli (1971) was used. Coomassie-stained gels were analyzed using a Personal Densitometer SI (Molecular Dynamics) and ImageQuant software. Concentrations of tailspike chains were determined from Coomassie-stained SDS gels using tailspike standards as described previously (Betts and King 1998).

Refolding of the native tailspike

Purified phage wild-type tailspike was unfolded at 2 mg/mL in 5 M Urea, 50 mM citrate (pH 3) for 1 h at room temperature (Seckler et al. 1989). Renaturation was initiated by diluting to 100 μ/mL by a 1/20 dilution in 50 mM Tris (pH 7.6) and 2 mM EDTA. Because some further refolding may occur on ice, the refolding was designed so that all of the samples would terminate at 30-sec intervals. At time 0, the 120-min sample was renatured, then incubated on ice for 30 min to propagate productive intermediates (Betts and King 1998). Samples were then shifted to 20°C to continue refolding. Samples were taken out after 120, 60, 15, 10, 5, 1, and 0 min, put onto ice at 30-sec intervals, and native sample buffer was added. As soon as possible, the refolding samples and 1/20 dilution of wild-type, pET, and 546D lysates were loaded and electrophoresed in the cold through a 7-1/2% native polyacrylamide gel for 1 h at 10 mAmps and 3 h at 15 mAmps. The protein bands were visualized by silver staining (Rabilloud et al. 1988).

Trypsin treatment of lysate supernatants

A stock solution of 10 mg/mL trypsin was prepared in 1% HCl. Trypsin digestions were initiated by mixing 175 μL of lysate supernatant fractions with 20 μL of trypsin in 1% HCl and incubat-

ing at room temperature. Portions of this reaction were quenched at intervals by mixing 50 μL with 10 μL of trypsin inhibitor to a final trypsin:trypsin inhibitor mass ratio of 5:2. Inhibited samples were kept at room temperature for up to 10 min, at which time they were mixed with ice-cold native sample buffer and stored on ice for no longer than 1 h until the initiation of electrophoresis in the cold.

Heat treatment of lysate supernatants

Lysate supernatant fractions (700 μL) were incubated in a circulating water bath at 42°C. At intervals, 45-μL portions were chilled on ice. After the last time point, 16 μL of each chilled sample was mixed with 8 μL of ice-cold 3× native sample buffer and stored on ice until electrophoresis. Additional samples (25 μL) were separated into soluble and insoluble fractions by centrifugation at 4°C for 2 min at 16,000*g*. The supernatant from the centrifugation step was mixed with 12.5 μL of SDS sample buffer and stored at −20°C. The pellet was resuspended in 25 μL of lysis buffer and mixed with 12.5 μL of SDS sample buffer and also stored at −20°C.

Acknowledgments

This research was supported by the National Institutes of Health (GM17,980 to J.K.) and by the National Science Foundation's Engineering Research Center Initiative (8803014). Maria Rodriguez assisted with site-directed mutagenesis. Stephanie Albin assisted with DNA sequencing.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Baum, J. and Brodsky, B. 1999. Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol.* **9:** 122–128.

Baxa, U., Steinbacher, S., Miller, S., Weintraub, A., Huber, R., and Seckler, R. 1996. Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide. *Biophys. J.* **71:** 2040–2048.

Beissinger, M., Lee, S.C., Steinbacher, S., Reinemer, P., Huber, R., Yu, M.H., and Seckler, R. 1995. Mutations that stabilize folding intermediates of phage P22 tailspike protein: Folding in vivo and in vitro, stability, and structural context. *J. Mol. Biol.* **249:** 185–194.

Benton, C.B., Clartk, P.L., and King, J. 2002. Characterization of the protrimer intermediate in the folding pathway of the interdigitated β-helix tailspike protein. *Biochemisty* (in press).

Betts, S. and King, J. 1999. There's a right way and a wrong way: In vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Structure Fold. Des.* **7:** R131–R139.

Betts, S., Haase-Pettingell, C. and King, J. 1997. Mutational effects on inclusion body formation. *Adv. Protein Chem.* **50:** 243–264.

Betts, S., Speed, M., and King, J. 1999. Detection of early aggregation intermediates by native gel electrophoresis and native western blotting. *Methods Enzymol.* **309:** 333–350.

Betts, S.D. and King, J. 1998. Cold rescue of the thermolabile tailspike intermediate at the junction between productive folding and off-pathway aggregation. *Protein Sci.* **7:** 1516–1523.

Chen, B. and King, J. 1991. Thermal unfolding pathway for the thermostable P22 tailspike endorhamnosidase. *Biochemistry* **30:** 6260–6269.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J., and Knight, S.D. 1999. X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **285:** 1061–1066.

Clark, P.L. and King, J. 2001. A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J. Biol. Chem.* **276:** 25411–25420.

Danner, M. and Seckler, R. 1993. Mechanism of phage P22 tailspike protein folding mutations. *Protein Sci.* **2:** 1869–1881.

Engel, J. and Prockop, D.J. 1991. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Chem.* **20:** 137–152.

Fane, B., Villafane, R., Mitraki, A., and King, J. 1991. Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *J. Biol. Chem.* **266:** 11640–11648.

Fessler, L.I. and Fessler, J.H. 1982. Identification of the carboxyl peptides of mouse procollagen IV and its implications for the assembly and structure of basement membrane procollagen. *J. Biol. Chem.* **257:** 9804–9810.

Fuchs, A., Seiderer, C., and Seckler, R. 1991. In vitro folding pathway of phage P22 tailspike protein. *Biochemistry* **30:** 6598–6604.

Goldenberg, D. and King, J. 1982. Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. *Proc. Natl. Acad. Sci.* **79:** 3403–3407.

Goldenberg, D.P., Smith, D.H., and King, J. 1983. Genetic analysis of the folding pathway for the tail spike protein of phage P22. *Proc. Natl. Acad. Sci.* **80:** 7060–7064.

Haase-Pettingell, C. and King, J. 1997. Prevalence of temperature sensitive folding mutations in the parallel β coil domain of the phage P22 tailspike endorhamnosidase. *J. Mol. Biol.* **267:** 88–102.

Haase-Pettingell, C., Betts, S., Raso, S.W., Stuart, L., Robinson, A., and King, J. 2001. Role for cysteine residues in the in vivo folding and assembly of the phage P22 tailspike. *Protein Sci.* **10:** 397–410.

Haase-Pettingell, C. and King, J. 1988. Formation of aggregates from a thermolabile in vivo folding intermediate in P22 tailspike maturation. A model for inclusion body formation. *J. Biol. Chem.* **263:** 4977–4983.

King, J. and Laemmli, U.K. 1971. Polypeptides of the tail fibers of bacteriophage T4. *J. Mol. Biol.* **62:** 465–477.

King, J., Haase-Pettingell, C., Robinson, A.S., Speed, M., and Mitraki, A. 1996. Thermolabile folding intermediates: Inclusion body precursors and chaperonin substrates. *FASEB J.* **10:** 57–66.

Kreisberg, J.F., Betts, S.D., and King, J. 2000. Beta-helix core packing within the triple-stranded oligomerization domain of the P22 tailspike. *Protein Sci.* **9:** 2338–2343.

Lietzke, S.E., Keen, N.T., Yoder, M.D., and Jurnak, F. 1994. The three-dimensional structure of pectate lyase E, a plant virulence factor from *Erwinia chrysanthemi*. *Plant Physiol.* **106:** 849–862.

Miller, S., Schuler, B., and Seckler, R. 1998a. Phage P22 tailspike protein: Removal of head-binding domain unmasks effects of folding mutations on native-state thermal stability. *Protein Sci.* **7:** 2223–2232.

———. 1998b. A reversibly unfolding fragment of P22 tailspike protein with native structure: The isolated beta-helix domain. *Biochemistry* **37:** 9160–9168.

Mitraki, A., Fane, B., Haase-Pettingell, C., and King, J. 1991a. Mutations affecting protein folding and misfolding in vivo. In *Application of enzyme biotechnology* (ed. T. Baldwin and J. Kelly) Plenum Press, New York.

Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J., and King. 1991b. Global suppression of protein folding defects and inclusion body formation. *Science* **253:** 54–58.

Mitraki, A., Danner, M., King, J., and Seckler, R. 1993. Temperature-sensitive mutations and second-site suppressor substitutions affect folding of the P22 tailspike protein in vitro. *J. Biol. Chem.* **268:** 20071–20075.

Petersen, T.N., Kauppinen, S., and Larsen, S. 1997. The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: A right-handed parallel beta helix. *Structure* **5:** 533–544.

Rabilloud, T., Carpentier, G., and Tarroux, P. 1988. Improvement and simplification of low-background silver staining of proteins by using sodium dithionite. *Electrophoresis* **9:** 288–291.

Raso, S.W., Clark, P.L., Haase-Pettingell, C., King, J., and Thomas, Jr., G.J. 2001. Distinct cysteine sulfhydryl environments detected by analysis of Raman S-hh markers of Cys→Ser mutant proteins. *J. Mol. Biol.* **307:** 899–911.

Robinson, A.S. and King, J.A. 1997. Disulfide-bonded intermediate on the folding and assembly pathway of a non-disulfide bonded protein. *Nat. Struct. Biol.* **4:** 450–455.

Sargent, D., Benevides, J.M., Yu, M.H., King, J., and Thomas, Jr., G.J., 1988. Secondary structure and thermostability of the phage P22 tailspike. XX. Analysis by Raman spectroscopy of the wild-type protein and a temperature-sensitive folding mutant. *J. Mol. Biol.* **199:** 491–502.

Sauer, R.T., Krovatin, W., Poteete, A.R., and Berget, P.B. 1982. Phage P22 tail protein: Gene and amino acid sequence. *Biochemistry* **21:** 5811–5815.

Schlunegger, M.P., Bennett, M.J., and Eisenberg, D. 1997. Oligomer formation

by 3D domain swapping: A model for protein assembly and misassembly. *Adv. Protein Chem.* **50:** 61–122.

Schuler, B. and Seckler, R. 1998. P22 tailspike folding mutants revisited: Effects on the thermodynamic stability of the isolated β-helix domain. *J. Mol. Biol.* **281:** 227–234.

Schwarz, J.J. and Berget, P.B. 1989. Characterization of bacteriophage P22 tailspike mutant proteins with altered endorhamnosidase and capsid assembly activities. *J. Biol. Chem.* **264:** 20112–20119.

Seckler, R. 1998. Folding and function of repetitive structure in the homotrimeric phage P22 tailspike protein. *J. Struct. Biol.* **122:** 216–222.

Seckler, R., Fuchs, A., King, J., and Jaenicke, R. 1989. Reconstitution of the thermostable trimeric phage P22 tailspike protein from denatured chains in vitro. *J. Biol. Chem.* **264:** 11750–11753.

Speed, M.A., Wang, D.I., and King, J. 1995. Multimeric intermediates in the pathway to the aggregated inclusion body state for P22 tailspike polypeptide chains. *Protein Sci.* **4:** 900–908.

Speed, M.A., Wang, D.I.C., and King, J. 1996. Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotech.* **14:** 1283–1287.

Speed, M.A., Morshead, T., Wang, D.I., and King, J. 1997. Conformation of P22 tailspike folding and aggregation intermediates probed by monoclonal antibodies. *Protein Sci.* **6:** 99–108.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. 1994. Crystal structure of P22 tailspike protein: Interdigitated subunits in a thermostable trimer. *Science* **265:** 383–386.

Sturtevant, J.M., Yu, M.-H., Haase-Pettingell, C., and King, J. 1989. Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. *J. Biol. Chem.* **264:** 10693–10698.

Sunde, M. and Blake, C. 1997. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50:** 123–159.

Thomas, Jr., G.J., Becka, R., Sargent, D., Yu, M.H., and King, J. 1990. Conformational stability of P22 tailspike proteins carrying temperature- sensitive folding mutations. *Biochemistry* **29:** 4181–4187.

van Raaij, M.J., Mitraki, A., Lavigne, G., and Cusack, S. 1999. A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401:** 935–938.

Yoder, M.D., Keen, N.T., and Jurnak, R. 1993. New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science* **260:** 1503–1507.