Genetic Analysis of P22 Tail Spike Protein Folding

유 명 희
한국과학기술원 유전공학센터

Janathan King
Massachusetts Institute of Technology

Introduction

A fundamental, but often underappreciated aspect of gene expression is the folding of the newly synthesized polypeptide chain into the precisely organized three dimensional protein structure that is required for biological activity.

Since the work of Anfinsen and colleagues (1) it has been known that at least some polypeptide chains, when fully denatured in vitro, can efficiently fold back into their active three dimensional conformation, in the absence of other cellular components. For these chains, the information specifying the 3-D structure is contained in the one-dimensional amino acid sequence.

At present, the 3-D structures of well over 200 native proteins are known to atomic resolution. The amino acid sequence of each of these proteins, and of hundreds more, is also precisely know. Yet it is still not possible to deduce or predict from the inspection of the sequence of a protein of unknown structure precisely how the polypeptide backbone and side chains will be arranged in space. Nor do we understand, for those proteins whose structure has been solved, why the chain takes the conformation it does.

Studies from a number of proteins suggest that some residues are not important in the folding process whereas others are critical. For instance comparison of sequences from homologous proteins, such as globins (2), shows that many residues can be altered without any effect on the native comformations. Also the wide spread existence of amber mutations where multiple amino acids can be inserted and still yield a functional protein (3), indicates that some residues do not contribute to native conformation. On the other hand studies with small peptides like S-peptide of ribonuclease have identified amino acid residues critical critical for leading to and stopping α-helix (4,5).

We have adopted a genetic approach to identifying those residues and local sequences in a polypeptide chain which play an important role on the folding pathway. Our approach has been to isolate and characterize mutants which specifically alter the folding and subunit association pathway of a polypeptide chain, without altering the native protein. Such mutants distinguish residues involved in the kinetic control of conformation from residues involved in the stability and activity of the native protein. This approach is complementary to the efforts to characterize mutations which alter the stability of the mature protein (6,7,8). It is likely that many residues will have roles in both aspects of the functioning of the polypeptide chain. We thought it likely, however, that at least with large proteins, these aspects might be segregated in different local sequences.

In order to distinguish affects of amino acid substitutions on intermediates from affects on the native structure, we have focused on a protein in which folding intermediates and end products are kinetically and energetically well separated. The protein utilized in these studies is the thermostable cell attachment protein of *Salmonella* phage P22 (9).

P22 Tail Spike Protein: Folding and Subuint Assembly

The tail spike is a trimer of polypeptide chains coded for by gene **9** of P22 (9,10). The gene **9** polypeptide chain is 666 amino acids long, without unusual features of amino acid composition or se-

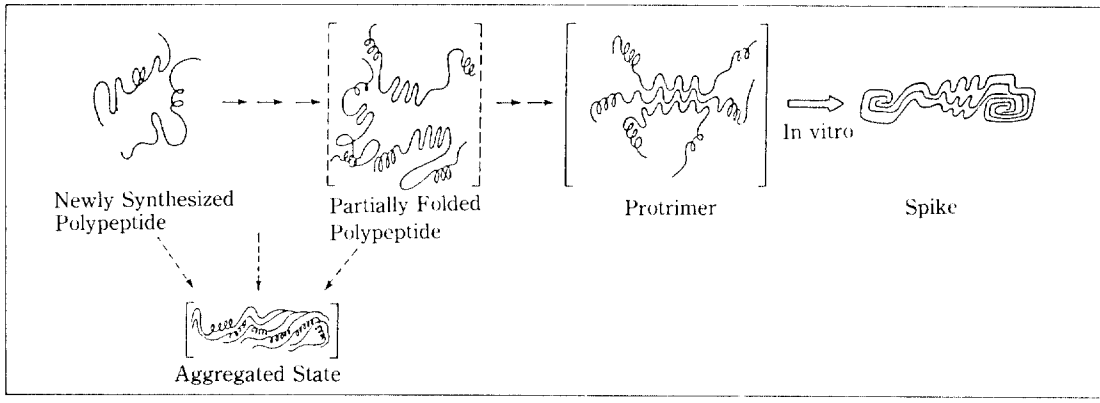

Fig. 1. Tail Spike Folding and Subunit Assembly Pathway.

quence (11). The essential properties of the protein are as follows. One end binds to phage heads in vivo, providing an activity assay (12). The other end is catalytic and is an endorhamnosidase, cleaving the O-antigen of *Salmonella* (13). The protein is a strong antigen and is the target of rabbit antiphage antibodies.

The native protein is very stable, consistent with its role as an external structural protein of the phage. It is heat stable to heat up to 80°C (10). It is resistant to proteases, and is not denatured by sodium dodecyl sulfate up to 60°C, though if heat denatured, the chains then form normal SDS polypeptide chain complexes (10). The tail spike maturation pathway is purely conformational, with no proteolytic cleavage, glycosylation, phosphorylation or other covalent modification known.

Since the native protein is SDS and trypsin resistant, partially folded intermediates can be easily distinguished from the mature spikes by SDS sensitivity or protease sensitivity (14). Studies of the wild type pathway have revealed the following sequence (Fig. 1): the newly synthesized polypeptide chains are released from the ribosome as single chains. They partially fold to a structure that has the specificity for chain-chain recognition. These chains associate into a complex called the protrimer. This kinetic intermediate in the folding and assembly pathway is identified by trapping in the cold (15). The protrimer intermediate has the three chains associated but not fully folded. Note that in this pathway, there is no species corresponding to a native monomer.

The early steps in this pathway are thermolabile. As the temperature is increased, the fraction of chains successfully passing through the pathway decreases (16). The last step, the protrimer to trimer transition has the opposite character. It is inhibited in the cold, permitting trapping, but proceeds at optimal efficiency in the 37-40°C temperature range.

Temperature Sensitive Folding (TSF) Mutations

Since the initial work of Beadle and Tatum (13) genetic approaches have provided a powerful means of analyzing biochemical pathways, primarily through the characterizaton of intermediates accumulating before the step blocked by mutation. To isolate mutant blocked in the folding pathway, we focused on temperature sensitive mutations. We have isolated over 100 perature, but not at high temperature (18).
this thermostable protein. The mutations are distributed over a minimum of 55 sites. They occur over the whole chain, but are more concentrated in the central region. These are conventional ts mutants - they form a plaque at low temrature, but not at high temperature (18).

At permissive temperature the polypeptide chains carrying TSF substitutions proceed through the folding and subunit assembly pathway, and form native tail spikes. Some properties of mutant tail spikes produced at permissive temperature are summurized on Table 1. Though these native proteins carry the TSF amino acid

Table 1. Comparisons of the High Temperature and Low Temperature Forms of the Mutant Polypeptide Chains

	39℃ *ts* Polypeptide Chains	30℃ Mature Protein
Synthesis	Yes	Yes
Head Binding	No	Yes
Endorhamnosidase	No	Yes
Antigenicity	No	Yes
Trypsin Resistance	No	Yes
SDS Resistance	No	Yes
Degradation	No	No

Conclusion: *ts* chains are accumulating in an incompletely or incorrectly folded form.

substitution, they are essentially indistinguishable from wild type. In particular, they are as thermostable as the wile type, requiring temperatures above 80°C for heat denaturation. The substitutions have little effect on the mature state of the protein. In a sense, they represent residues that are not involved in the activities of the mature protein.

At restrictive temperature, the results are very different. The polypeptide chains carrying the mutant amino acid are synthesized at normal rates, and are stable within the cell. However, these chains display none of the activities of the native protein - head binding, endorhamnosidase, absorption, or antigenicity-and do not form the protrimer (14). In fact they are precipitated by antibody made against guanidine denatured tail spikes, indicating that their conformation is closer to the unfolded (14). On cell lysis, and possibily before, these incompletely folded chains associate into large aggreqates.

The high temperature state of the tail spike polypeptide chain is not simply a misfolded species. On shifting infected cells to permissive temperature, active tail spikes are recovered (19). Thus the mutant polypeptide chains accumulating are reversibly related to intermediates in the folding pathway.

The mutant proteins coded by TSF mutations of 33 sites have been studied in some detail. All of them have the same general character; they do not affect the native protein formed at permissive temperature, but they block the formation of the native protein if the polypeptide chain was released from the ribosome at restrictive temperature. Thus, this genetic map is a map of sites in the polypeptide chain which play a critical role in the folding pathway at least at elevated temperature.

Measurements of the overall rate of the tail spike maturation pathway as a function of temperature (Fig. 2) reveal that the mutations do not alter the rates, but rather reduce the yield with increasing temperature (16). This is consistent with either of two general models for the ts defect; (i) the ts substitutions further destabilize an already thermolabile intemediate or (ii) the substitutions speed up a step that levels off the productive pathway, probably to an aggregated state.

Amino Acid Substitutions in the TSF Mutants

We have determined the amino acid substitutions of a number of these mutant proteins by

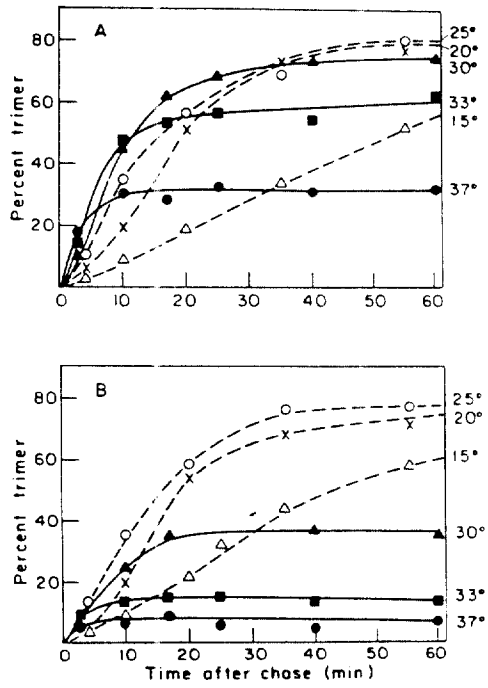

Fig. 2. Formation of Native Tail Spike from Newly Synthesized Polypeptides as a Function of Temperature.

isolating restriction fragments from the phage DNA, and sequencing the fragments.

The genetic map was used to select the fragments isolate (16). A number of mutations have also been sequenced by cloning restriction fragments into M13 and sequencing the insert (20). In all cases where we found nucleotide substitutions, they represented a single amino acid substitution. These are shown in Table 2.

The Effect of Single Amino Acid Substitution on Protein Stability

Single ts amino acid substitutions in many proteins affect protein stability. Previous studies with our ts mutants in the tail spike gene revealed that changes in thermostability could not account for the physiological inability of the polypeptide chain to mature at 40°C. However, in order to examine this in greater detail and to study more subtle secondary effects of the mutations a number of the purified proteins (21) have been examined with microscanning calorimetry in collaboration with J. Sturtevant at Yale. Protein samples were prepared in 20mM KPO_4, pH7.4 at protein concentration 2mg/ml.

Preliminary experiments with wild type and mutant proteins (tsH302, tsmU9, and tsU53) showed i) single amino acid replacements have no substantial effect on either $\triangle$Hcal or $\triangle$Tm; ii) $\triangle$Hvh < $\triangle$Hcal, suggesting that intermediate states are significantly populated; and iii) the Tm for w.t., tsH302, tsmU9, and tsU53 are 89.4°C, 88.8°C, 89.3°C, and 83.5°C, respectively. Note that the Tm of the least stable on (tsU53) is still too high to be responsible for the ts phenotype at 39°C, the temperature at which mutations express their physiological defect.

Discussion

A. The Sites of TSF Mutations

The substitutions are quite diverse, as shown in Table 2. However, two classes stand out: replacements of glycine and replacements of threonines. The threonine substitutions were previously discussed (16), and in this paper only the glycine substitutions are dealt with.

Five glycine residues have been identified as sites of TSF mutations. For all five the substitutions are by bulky charged amino acids. Though it is not suprising that these substitutions would

Table 2. Amino Acid Substitutions of Gene 9 TS Folding Mutants.

Mutation	Residue	Substitution	Local Sequence
tsU 9	177	Gly > Arg	Phe·Ile·Gly·Asp·***Gly***·Asn·Leu·Ile·Phe
tsH 304	244	Gly > Arg	Val·Lys·Phe·Pro·***Gly***·Ile·Glu·Thr·Leu
tsmU 9	281	Gly > Asn	Gly·Leu·Met·Ala·***Gly***·Phe·Leu·Phe·Arg
tsH 302	323	Gly > Asp	Asn·Tyr·Val·Ile·***Gly***·Gly·Arg·Thr·Ser
tsU 38	435	Gly > Glu	Leu·Leu·Val·Arg·***Gly***·Ala·Leu·Gly·Val
tsH 300	235	Thr > Ile	Gly·Tyr·Gln·Pro·***Thr***·Val·Ser·Asp·Tyr
tsU 18	307	Thr > Ala	Asp·Gly·Ile·Ile·***Thr***·Phe·Glu·Asn·Leu
tsH 301	368	Thr > Ile	Thr·Trp·Gln·Gly·***Thr***·Val·Gly·Ser·Thr
tsU 5	227	Ser > Phe	Thr·Leu·Lys·Gln·***Ser***·Lys·Thr·Asp·Gly
tsN 48	333	Ser > Asn	Gly·Ser·Val·Ser·***Ser***·Ala·Gln·Phe·Leu
tsU 19	285	Arg > Lys	Gly·Phe·Leu·Phe·***Arg***·Gly·Cys·His·Phe
tsU 53	382	Arg > Ser	Asn·Leu·Gln·Phe·***Arg***·Asp·Ser·Val·Val
tsH 303, tsU 11	250	Pro > Ser	Glu·Thr·Leu·***Pro***·Pro·Asn·Ala·Lys
tsU 24	258	Ile > Leu	Lys·Gly·Gln·Asn·***Ile***·Thr·Ser·Thr·Leu
tsRAF, tsRH	270	Val > Gly	Glu·Cys·Ile·Gly·***Val***·Glu·Val·His·Arg
ts 9. 1	334	Ala > Val	Ser·Val·Ser·Ser·***Ala***·Gln·Phe·Leu·Arg

cause problems, it was difficult to understand how they could be accommodated into the native protein at permissive temperature. In fact, the native mutant proteins specified by these mutants have altered electrophoretic mobilities and isoelectric points (20), indicating that they are on the protein surface. This allows the charged residues to be accommodated at sites normally occupied by glycines.

The low temperature result shows that there is no steric hine drance in the protein. Therefore, the high temperature defect must be due to a kinetic effect on the folding pathway, as indicated by the mutant phenotypes (19).

Since turns are concentrated at the surface of proteins, and glycines are common at turns, we interpret these sites as representing glycines at surface turns. Indeed Chou-Fasman algorithm (22) predicts ß-turn structure with the local sequences adjacent these substitutions. The charged residue would alter the interaction with the solvent, or perhaps introduce interactions with other parts of the chain, which at elevated temperature interferes with the folding pathway

B. Are the TSF Sites Involved in Chain Folding or Chain Association?

At restrictive temperature, all of the TSF mutant polypeptide chains we have studied are blocked prior to the formation of the protrimer intermediate (14). We have interpreted these and other results as indicating that the mutant amino acid substitutions prevent the chains from achieving the conformation needed for specific chain-chain recognition and association. This kinetic explanation explains why the mutant proteins, once native, are as thermostable as wild type.

An alternative model is that the mutations do not act kinetically, but destabilize a specific domain needed for association. However the distribution of the mutations throughout the polypeptide chain indicates that a very large fraction of the polypeptide chain participates in the "domain"; in fact, the entire chain. Furthermore, since amino acid substitutions at over thirty different sites destabilize this special domain, each one cannot make a very large contribution to its stability. Yet the different mutants all sharply reduce protrimer formation at restrictive temperature.

Experiments on mixed infections with wild type and ts mutants with varied mobility showed that at high temperature mutant chains could not be rescued by the wild type chains. The observations are better for by a model in which the mutations interfere with kinetic steps in the pathway, such that the chain never achieves the correct conformation for association. The substitutions could also affect kinetic steps by speeding up off-pathway interactions.

Reference

1. Anfinsen, C.B. Science. **181,** 223-230.
2. Dickerson, R.E., and Geis, I., (1983) Hemoglobin, The Benjamin/Cummings Publishing Company, Inc.
3. Miller, J.H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H., and schmitz, A. (1979) *J. Mol. Biol.* **131,** 191-222.
4. Bierzynski, A., Kim, P.S. and Baldwin, R.L. (1982) *Proc. Natl. Acad. Sci. USA* **79** 2470-2474.
5. Kim, P.S., and Baldwin, R.L. (1984) *Nature* **307,** 329-334.
6. Shortle, D., (1983) *Gene* **22,** 181-189.
7. Hawkes, R., Grutter, M.G., and Schellman, J. (1984) *J.Mol. Biol.* **175,** 195-212.
8. Alber, T., and Wozniak, J.A., (1985) *Proc. Natl. Acad. Sci. USA* **82,** 747-751.
9. Berget, P.B. and Poteete, A.R. (1980) *J. Virol.* **34,** 234-243.
10. Goldenberg, D.P., Berget, P.B., and King, J., (1982) *J. Biol. Chem.* **257,** 7864-7871.
11. Sauer, R.T., Krovatin, W., Poteete, A.R., and Berget, P.B., (1982) *Biochemistry* **21,** 5811-5815.
12. Israel, J.V., Anderson, T.F., and Levine, M. (1967) *Proc. Natl. Acad. Sci. USA* **57,** 284-291.
13. Iwashita, S. and Kanegasaki, S. (1976) *Eur. J. Biochem.* **65,** 87-94.
14. Goldenberg, D.P., Smith, D.H., and King, J. *Proc. Nat. Acad. Sci. USA* **80,** 7060-7064 (1983).

15. Goldenberg, D. and King, J., (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3403-3407.
16. Yu, M.-H., and King, J., (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6548-6588.
17. Beadle, G.W., and Tatum, E.L., (1941) *Proc. Natl. Acad. Sci. USA* **27,** 499-506.
18. Edgar, R.S. and Lielausis, I. (1964) *Genetics* **49,** 649-662.
19. Smith, D.H., and King, J., (1981) *J. Mol. Biol* **145,** 653-676.
20. Yu, M.-H., and King, J., manuscnipt in preparation.
21. King, J., and Yu, M.-H., Methods in Enzymology: Enzyme structure (in press).
22. Chou, P.Y., and Fasman, G.D. (1978) *Ann. Rev. Biochem.* **47,** 251-276.