

# STUDIES ON THE PRINCIPLES THAT GOVERN THE FOLDING OF PROTEIN CHAINS

Nobel Lecture, December 11, 1972

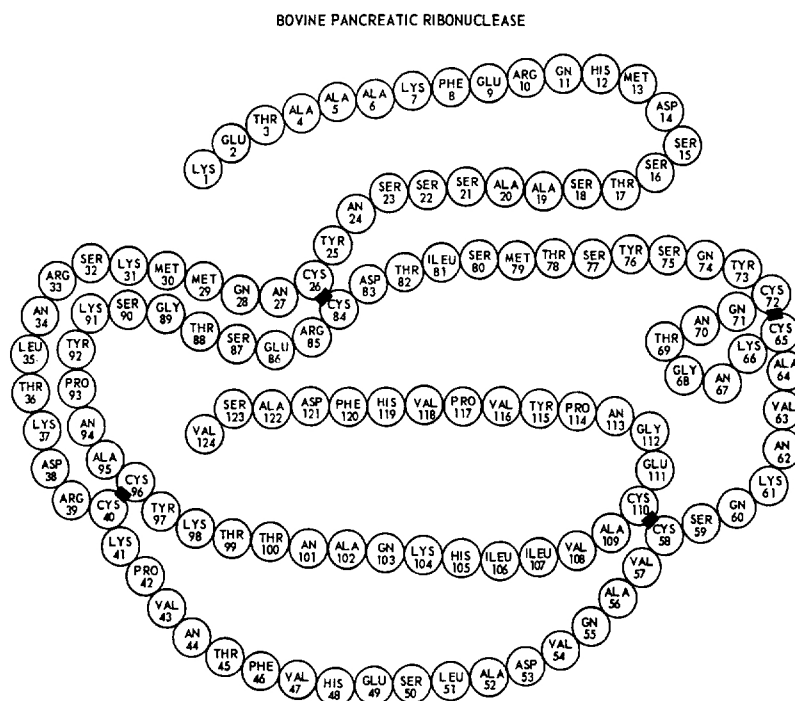
by

CHRISTIAN B. ANFENSEN

National Institutes of Health

Bethesda, Maryland

The telegram that I received from the Swedish Royal Academy of Sciences specifically cites "... studies on ribonuclease, in particular the relationship between the amino acid sequence and the biologically active conformation . . ." The work that my colleagues and I have carried out on the nature of the process that controls the folding of polypeptide chains into the unique three-dimensional structures of proteins was, indeed, strongly influenced by observations on the ribonuclease molecule. Many others, including Anson and Mirsky (1) in the '30s and Lumry and Eyring (2) in the '50s, had observed and discussed the reversibility of denaturation of proteins. However, the true elegance of this consequence of natural selection was dramatized by the ribonuclease work, since the refolding of this molecule, after full denaturation by reductive cleavage of its four disulfide bonds (Figure 1), required that only one of the 105



possible pairings of eight sulfhydryl groups to form four disulfide linkages take place. The original observations that led to this conclusion were made together with my colleagues, Michael Sela and Fred White, in 1956–1957 (6). These were in actuality, the beginnings of a long series of studies that rather vaguely aimed at the eventual total synthesis of the protein. As we all know, Gutte and Merrifield (7) at the Rockefeller Institute, and Ralph Hirschman and his colleagues at the Merck Research Institute (8), have now accomplished this monumental task.

The studies on the renaturation of fully denatured ribonuclease required many supporting investigations (9, 10, 11, 12) to establish, finally, the generality which we have occasionally called (13) the “thermodynamic hypothesis”. This hypothesis—states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc.) is the one in which the Gibbs free energy of the *whole system* is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a *given environment*. In terms of natural selection through the “design” of macromolecules during evolution, this idea emphasized the fact that a protein molecule only makes stable, structural sense when it exists under conditions similar to those for which it was selected—the so-called physiological state.

After several years of study on the ribonuclease molecule it became clear to us, and to many others in the field of protein conformation, that proteins devoid of restrictive disulfide bonds or other covalent cross linkages would make more convenient models for the study of the thermodynamic and kinetic aspects of the nucleation, and subsequent pathways, of polypeptide chain folding. Much of what I will review will deal with studies on the flexible and convenient staphylococcal nuclease molecule, but I will first summarize some of the older, background experiments on bovine pancreatic ribonuclease itself.

#### SUPPORT FOR THE “THERMODYNAMIC HYPOTHESIS.”

An experiment that gave us a particular satisfaction in connection with the translation of information in the linear amino acid sequence into native conformation involved the rearrangement of so-called “scrambled” ribonuclease (12). When the fully reduced protein, with 8 SH groups, is allowed to reoxidize under denaturing conditions such as exist in a solution of 8 molar urea, a mixture of products is obtained containing many or all of the possible 105 isomeric disulfide bonded forms (Schematically shown at the bottom right of Figure 2). This mixture is essentially inactive—having on the order of 1 % the activity of the native enzyme. If the urea is removed and the “scrambled” protein is exposed to a small amount of a sulfhydryl group-containing reagent such as mercaptoethanol, disulfide interchange takes place and the mixture eventually is converted into a homogeneous product, indistinguishable from native ribonuclease. This process is driven entirely by the free energy of conformation that is gained in going to the stable, native structure. These experi-

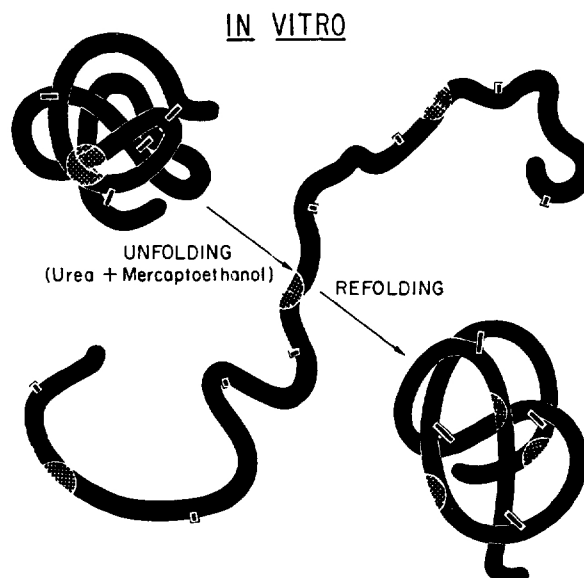


Fig. 2.  
Schematic representation of the reductive denaturation, in 8 molar urea solution containing 2-mercaptoethanol, of a disulfide-cross linked protein. The conversion of the extended, denatured form to a randomly cross linked, "scrambled" set of isomers is depicted in the lower right portion of the figure.

ments, incidentally, also make unlikely a process of obligatory, progressive folding during the elongation of the polypeptide chain, during biosynthesis, from the  $\text{NH}_2$ - to the  $\text{COOH}$ -terminus. The "scrambled" protein appears to be essentially devoid of the various aspects of structural regularity that characterize the native molecule.

A disturbing factor in the kinetics of the process of renaturation of reduced ribonuclease, or of the "unscrambling" experiments described above, was the slowness of these processes, frequently hours in duration (11). It had been established that the time required to synthesize the chain of a protein like ribonuclease, containing 124 amino acid residues, in the tissues of a higher organism would be approximately 2 minutes (14, 15). The discrepancy between the *in vitro* and *in vivo* rates led to the discovery of an enzyme system in the endoplasmic reticulum of cells (particularly in those concerned with the secretion of extracellular,  $\text{SS}$ -bonded proteins) which catalyzes the disulfide interchange reaction and which, when added to solutions of reduced ribonuclease or to protein containing randomized  $\text{SS}$  bonds, catalyzed the rapid formation of the correct, native disulfide pairing in a period less than the requisite two minutes (16, 17). The above discrepancy in rates would not have been observed in the case of the folding of non-crosslinked structures and, as discussed below, such motile proteins as staphylococcal nuclease or myoglobin can undergo virtually complete renaturation in a few a seconds or less.

The disulfide interchange enzyme subsequently served as a useful tool for the examination of the thermodynamic stability of disulfide-bonded protein

structures. This enzyme, having a molecular weight of 42,000 and containing three half-cystine residues, one of which must be in the SH form for activity (18, 19), appears to carry out its rearranging activities on a purely random basis. Thus, a protein whose SS bonds have been deliberately broken and reformed in an incorrect way, need only be exposed to the enzyme (with its essential half-cystine residue in the pre-reduced, SH form) and interchange of disulfide bonds occurs until the native form of the protein substrate is reached. Presumably, SS bonds occupying solvent-exposed, or other thermodynamically unfavorable positions, are constantly probed and progressively replaced by more favorable half-cystine pairings, until the enzyme can no longer contact bonds because of steric factors, or because no further net decrease in conformational free energy can be achieved. Model studies on ribonuclease derivatives had shown that, when the intactness of the genetic message represented by the linear sequence of the protein was tampered with by certain cleavages of the chain, or by deletions of amino acids at various points, the added disulfide interchange enzyme, in the course of its "probing", discovered this situation of thermodynamic instability and caused the random reshuffling of SS bonds with the formation of an inactive cross-linked network of chains and chain fragments (e.g., (20)). With two naturally occurring proteins, insulin and chymotrypsin, the interchange enzyme did, indeed, induce such a randomizing phenomenon (21). Chymotrypsin, containing three SS-bonded chains, is known to be derived from a single-chained precursor, chymotrypsinogen, by excision of two internal bits of sequence. The elegant studies of Steiner and his colleagues subsequently showed that insulin was also derived from a single-chained precursor, proinsulin (Figure 3), which is converted to

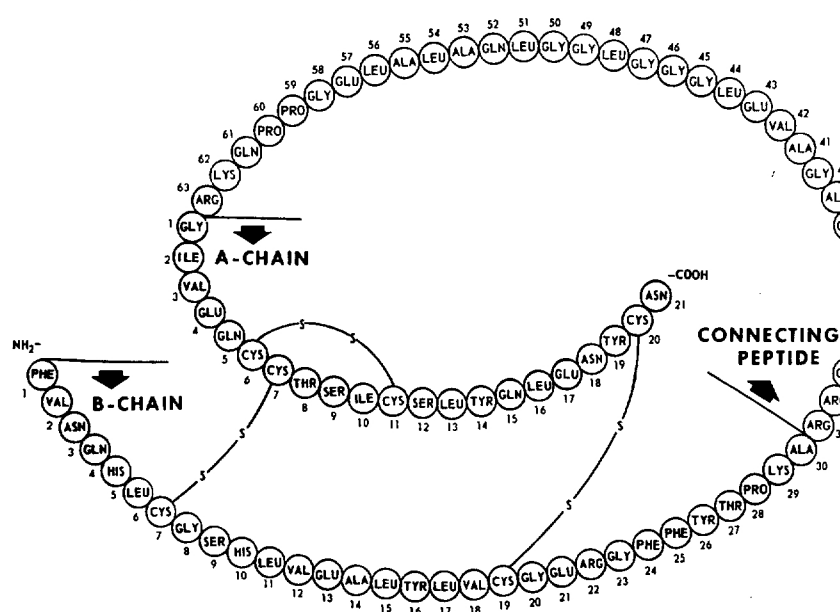


Fig. 3.

The structure of porcine proinsulin (R. E. Chance, R. M. Ellis and W. W. Bromer, *Science*, 161, 165 (1968).

the two-chained form, in which we normally find the active hormone, by removal of a segment from the middle of the precursor strand after formation of the 3 SS bonds (22). In contrast, the multichained immune globulins are *not* scrambled and inactivated by the enzyme, reflecting the fact that they are normal products of the disulfide bonding of 4 preformed polypeptide chains.

#### FACTORS CONTRIBUTING TO THE CORRECT FOLDING OF POLYPEPTIDE CHAINS.

The results with the disulfide interchange enzyme discussed above suggested that the correct and unique translation of the genetic message for a particular protein backbone is no longer possible when the linear information has been tampered with by deletion of amino acid residues. As with most rules, however, this one is susceptible to many exceptions. First, a number of proteins have been shown to undergo reversible denaturation, including disulfide bond rupture and reformation, after being shortened at either the  $\text{NH}_2$ - or  $\text{COOH}$ -terminus (23). Others may be cleaved into two (24, 25, 26), or even three, fragments which, although devoid of detectable structure alone in solution, recombine through noncovalent forces to yield biologically active structures with physical properties very similar to those of the parent protein molecules. Richards and his colleagues (24) discovered the first of these recombining systems, ribonuclease-S (RNase-S), which consists of a 20 residue fragment from the  $\text{NH}_2$ -terminal end held by a large number of noncovalent interactions to the rest of the molecule, which consists of 104 residues and all four of the disulfide bridges. The work by Wyckoff, Richards and their associates on the three-dimensional structure of this two-fragment complex (27) and on the identification of many of the amino acid side chains that are essential for complementation is classical, as are studies by Hofmann (28) and Scoffone (29) and their colleagues on semi-synthetic analogues of this enzyme derivative. Studies in our own laboratory (30) showed that the 20 residue "RNase-S-peptide"

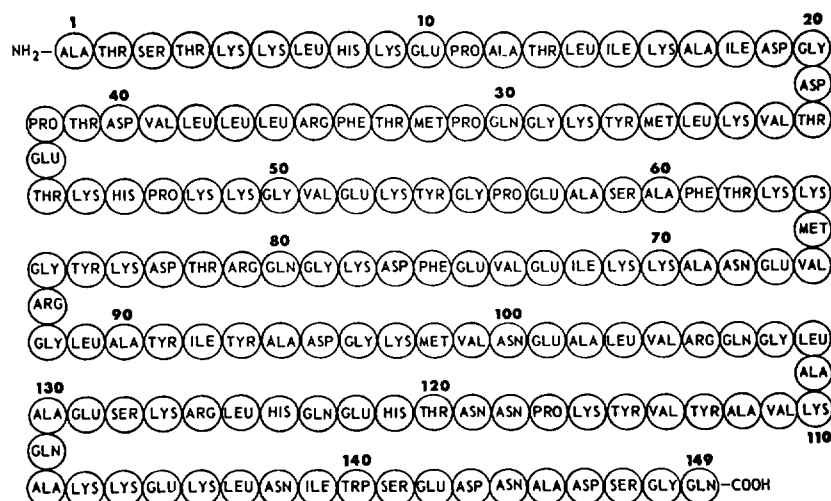


Fig. 4.  
Covalent structure of the major extracellular nuclease of *Staphylococcus aureus* (32, 33).

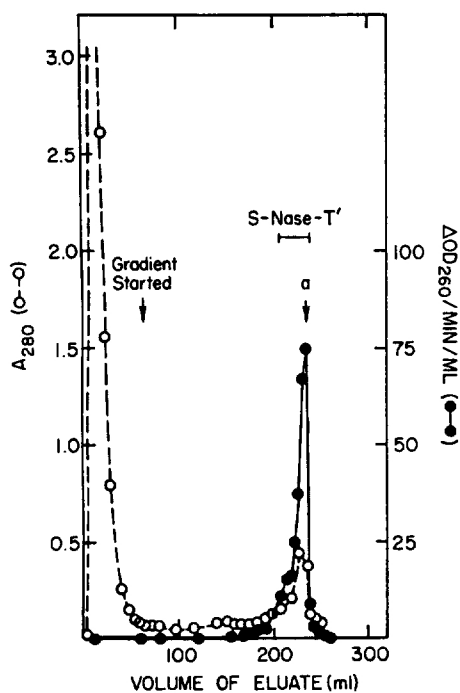


Fig. 5.  
Isolation of semisynthetic nuclease-T  
on a phosphocellulose column follow-  
ing "functional" purification by trypsin  
digestion in the presence of calcium  
ions and thymidine-3'5'-diphosphate  
(41).

fragment could be reduced by 5 residues at its COOH-terminus without loss of enzymic activity in the complex, or of its intrinsic stability in solution.

Other examples of retention of native structural "memory" have been found with complexing fragments of the staphylococcal nuclease molecule (25, 31). This calcium-dependent, RNA and DNA cleaving enzyme (Figure 4) consists of 149 amino acids and is devoid of disulfide bridges and SH groups (32, 33). Although it exhibits considerable flexibility in solution as evidenced by the ready exchange of labile hydrogen atoms in the interior of the molecule with solvent hydrogen atoms (34), only a very small fraction of the total population deviates from the intact, native format at any moment. Spectral and hydrodynamic measurements indicate marked stability up to temperatures of approximately 55°. The protein is greatly stabilized, both against hydrogen exchange (34) and against digestion by proteolytic enzymes (35) when calcium ions and the inhibitory ligand, 3'5'-thymidine diphosphate (pdTp), are added. Trypsin, for example, then only cleaves at very restricted positions—the loose aminoterminal portion of the chain and a loop of residues that protrudes out from the molecule as visualized by X-ray crystallography. Cleavage occurs between lysine residues 5 and 6 and, in the sequence -Pro-Lys-Lys-Gly- (residues 47 through 50), between residues 48 and 49 or 49 and 50 (25). The resulting fragments (6—48) and (49—149) or (50—149), are devoid of detectable structure in solution (36). However, as in the case of RNase-S, when they are mixed in stoichiometric amounts regeneration of activity (about 10 %) and of native structural characteristics occurs (the complex is called nuclease-T). Nuclease-T has now been shown (37a) to be closely isomorphous with native nuclease (37b). Thus the cleavages

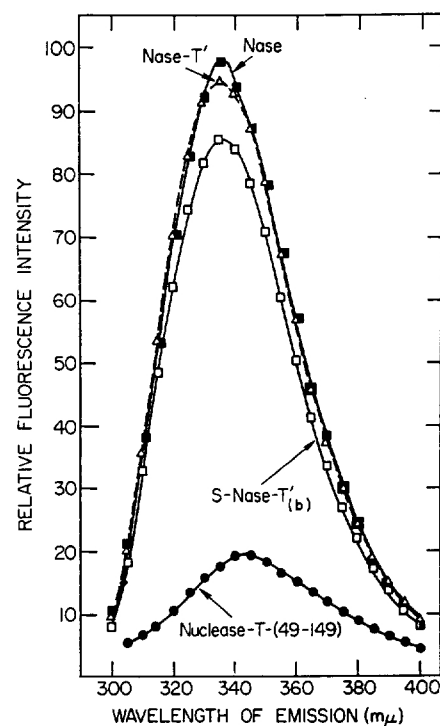


Fig. 6.  
Use of fluorescence measurements to determine the relative hydrophobicity (presumably reflecting "nativeness" in the case of nuclease) of the molecular environment of the single tryptophan residue in this protein (39, 41).

and deletions do not destroy the geometric "sense" of the chain. Recently it was shown that residue 149 may be removed by carboxypeptidase treatment of nuclease, and that residues 45 through 49 are dispensable, the latter conclusion the result of solid phase-synthetic studies (38) on analogues of the fragment, (6—47).

Earlier studies by David Ontjes (39) had established that the rapid and convenient solid-phase method developed by Merrifield (40) for peptide synthesis could be applied to the synthesis of analogues of the (6—47) fragment of nuclease-T. The products, although contaminated by sizeable amounts of "mistake sequences" which lack an amino acid residues due to slight incompleteness of reaction during coupling, could be purified by ordinary chromatographic methods to a stage that permitted one to make definite conclusions about the relative importance of various components in the chain. Taking advantage of the limited proteolysis that occurs when nuclease is treated with trypsin in the presence of the stabilizing ligands, calcium and pdTp, Chaiken (41) was able to digest away those aberrant synthetic molecules of (6—47) that did not form a stable complex with the large, native fragment, (49—149). After digestion of the complex, chromatography on columns of phosphocellulose (Figure 5) yielded samples of semisynthetic nuclease-T that were essentially indistinguishable from native nuclease-T. For example, the large enhancement of fluorescence of the single tryptophan residue in nuclease (located at position 140 in the (50—149) fragment) upon addition of the native (6—49) fragment was also shown when, instead, synthetic (6—47) peptide isolated from semisynthetic nuclease-T that had been purified as described above was added (Figure 6).

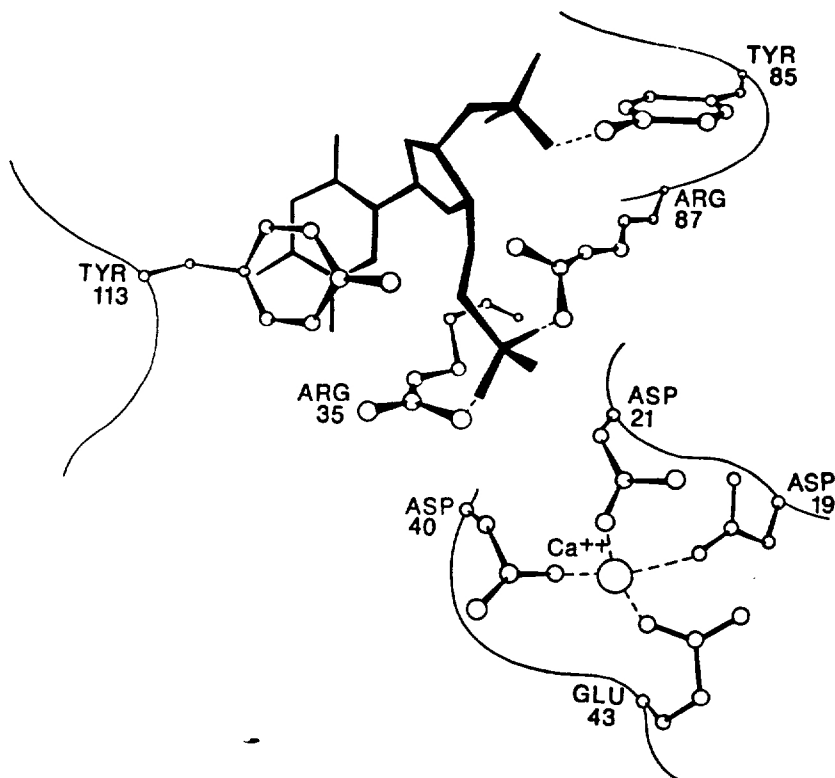


Fig. 7.

Amino acid residues in the sequence of nuclease that are of particular importance in the catalytic activity and binding of substrate and calcium ions (42).

The dispensability, or replaceability, of a number of residues to the stability of the nuclease-T complex was established by examining the fluorescence, activity, and stability to enzymatic digestion, of a large number of semisynthetic analogues (42). As illustrated in Figure 7, interaction with the calcium atom required for nuclease activity normally requires the participation of four dicarboxylic amino acids. Although the *activities* of complexes containing synthetic (6—47) fragments in which one of these had been replaced with an asparagine or glutamine residue were abolished (with one partial exception— asparagine at position 40), three dimensional structure and complex stability was retained for the most part. Similarly, replacement of arginine residue 35 with lysine yielded an inactive complex, but nevertheless one with strong three dimensional similarity to native nuclease-T.

A second kind of complementing system of nuclease fragments (31) consists of tryptic fragment (1—126) and a partially overlapping section of the sequence, (99—149), prepared by cyanogen bromide treatment of the native molecule (shown schematically in Figure 8). These two peptides form a complex with about 15 % the activity of nuclease itself which is sufficiently stable in the presence of pdTp and calcium ions to exhibit remarkable resistance to digestion by trypsin. Thus, many of the overlapping residues in the complex,



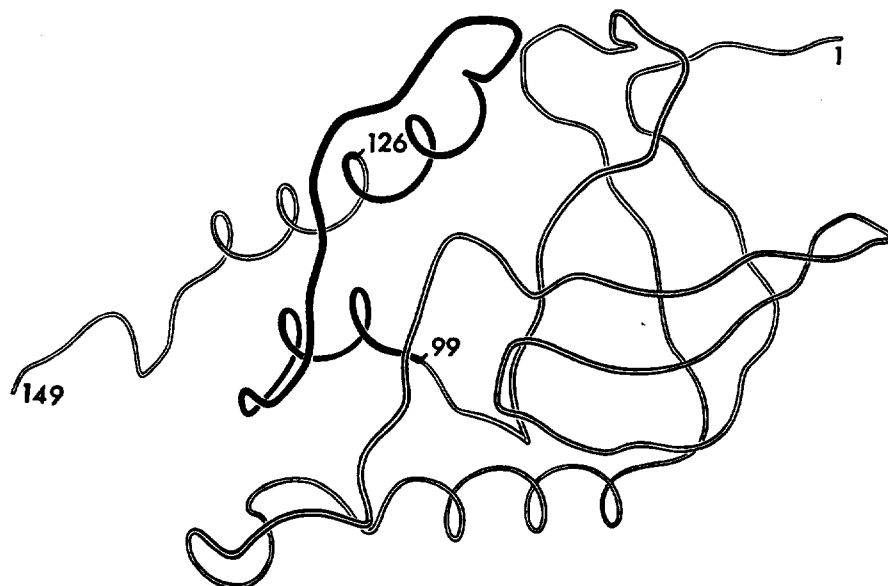


Fig. 8.

A schematic view of the three dimensional structure of staphylococcal nuclease (37b, 53).

(1—126): (99—149), may be “trimmed” away with the production of a derivative, (1—126): (111—149). Further degradation of each of the two components, the former with carboxypeptidases A and B and the latter with leucine aminopeptidase, permits the preparation of (1—124): (114—149) which is as active, and as structurally similar to native nuclease, (as evidenced by estimates of hydrodynamic, spectral, and helical properties) as the parent, undegraded complex. A number of synthetic analogues of the (114—149) sequence have been prepared (43), which also exhibit activity and “native” physical properties when added to (1—126). I will discuss below the manner in which these complexing fragments have been useful in devising experiments to study the processes of nucleation and folding of polypeptide chains.

#### MUTABILITY OF INFORMATION FOR CHAIN FOLDING.

Biological function appears to be more a correlate of macromolecular geometry than of chemical detail. The classic chemical and crystallographic work on the large number of abnormal human hemoglobins, the species variants of cytochrome *c* and other proteins from a very large variety of sources, and the isolation of numerous bacterial proteins after mutation of the corresponding genes have made it quite clear that considerable modification of protein sequence may be made without loss of function. In those cases where crystallographic studies of three-dimensional structure have been made, the results indicate that the geometric problem of “designing,” through natural selection, molecules that can subserve a particular functional need can be solved in many ways. Only the *geometry* of the protein and its active site need be conserved, except, of course, for such residues as actually participate in a unique way in a catalytic or regulatory mechanism (44). Studies of model systems have led

to similar conclusions. In our own work on ribonuclease, for example, it was shown that fairly long chains of poly-D, L-alanine could be attached to eight of the eleven amino groups of the enzyme without loss of enzyme activity (45). Furthermore, the polyalanylated enzyme could be converted to an extended chain by reduction of the four SS bridges in 8 M urea and this fully denatured material could then be reoxidized to yield the active, correctly folded starting substance. Thus, the chemistry of the protein could be greatly modified, and its capacity to refold after denaturation seemed to be dependent only on *internal* residues and not those on the outside, exposed to solvent. This is, of course, precisely the conclusion reached by Perutz and his colleagues (46), and by others (47) who have reviewed and correlated the data on various protein systems. Mutation and natural selection are permitted a high degree of freedom during the evolution of species, or during accidental mutation, but a limited number of residues, destined to become involved in the internal, hydrophobic core of proteins, must be carefully conserved (or at most replaced with other residues with a close similarity in bulk and hydrophobicity).

#### THE COOPERATIVITY REQUIRED FOR FOLDING AND STABILITY OF PROTEINS.

The examples of non-covalent interaction of complementing fragments of proteins quoted above give strong support to the idea of the essentiality of cooperative interactions in the stability of protein structure. As in the basic rules of languages, an incomplete sentence frequently conveys only gibberish. There appears to exist a very fine balance between stable, native protein structure and random, biologically meaningless polypeptide chains.

A very good example of the inadequacy of an incomplete sequence comes from our observations on the nuclease fragment, (1—126). This fragment contains all of the residues that make up the active center of nuclease. Nevertheless, this fragment, representing about 85 % of the total sequence of nuclease, exhibits only about 0.12 % the activity of the native enzyme (48). The further

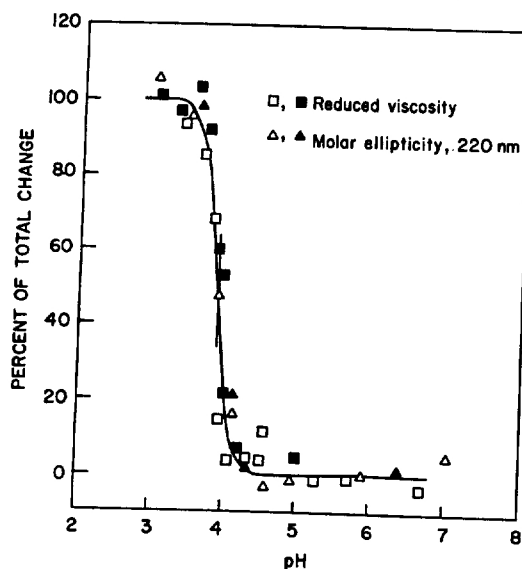


Fig. 9. Changes in reduced viscosity and molar ellipticity at 220 nm during the acid-induced transition from native to denatured nuclease.

□ and ■, Reduced viscosity; △ and ▲, molar ellipticity at 220 nm. □ and △, Measurements made during the addition of acid; ■ and ▲, measurements made during the addition of base.

A. N. Schechter, H. F. Epstein and C. B. Anfinsen, unpublished results.

addition of 23 residues during biosynthesis, or the addition, *in vitro*, of residues 99–149 as a complementing fragment (31), restores the stability required for activity to this unfinished gene translation.

The transition from incomplete, inactive enzyme, with random structure, to competent enzyme, with unique and stable structure, is clearly a delicately balanced one. The sharpness of this transition may be emphasized by experiments of the sort illustrated in Figure 9. Nuclease undergoes a dramatic change from native globular structure to random disoriented polypeptide over a very narrow range of pH, centered at pH 3.9. The transition has the appearance of a “two-stage” process—either all native or all denatured—and, indeed two-state mathematical treatment has classically been employed to describe such data. In actuality, it has been possible to show, by NMR and spectrophotometric-experiments (49), that one of the 4 histidines and one tyrosine residue of the 7 in nuclease become disoriented before the general and sudden disintegration of organized structure. However, such evidences of a stepwise denaturation and renaturation process are certainly not typical of the bulk of the cooperatively stabilized molecule.

The experiments in Figure 9, involving measurements of intrinsic viscosity and helix-dependent circular dichroism, are typical of those obtained with most proteins. In the case of nuclease, not only is the transition from native to denatured molecule during transfer from solution at pH 3.2 to 6.7 very abrupt, but the process of renaturation occurs over a very short time period. I will not discuss these stop-flow kinetic experiments (50) in detail in this lecture. In brief, the process can be shown to take place in at least two phases; an initial rapid nucleation and folding with a half-time of about 50 milliseconds and a second, somewhat slower transformation with a half time of about 200 milliseconds. The first phase is essentially temperature independent (and therefore possibly entropically driven) and the second temperature dependent.

#### NUCLEATION OF FOLDING.

A chain of 149 amino acid residues with two rotatable bonds per residue, each bond probably having 2 or 3 permissible or favored orientations, would be able to assume on the order of  $4^{149}$  to  $9^{149}$  different conformations in solution. The extreme rapidity of the refolding makes it essential that the process take place along a limited number of “pathways”, even when the statistics are severely restricted by the kinds of stereochemical ground rules that are implicit in a so-called Ramachandran plot. It becomes necessary to postulate the existence of a limited number of allowable initiating events in the folding process. Such events, generally referred to as nucleations, are most likely to occur in parts of the polypeptide chain that can participate in conformational equilibria between random and cooperatively stabilized arrangements. The likelihood of a requirement for cooperative stabilization is high because, in aqueous solution, ionic or hydrogen bonded interactions would not be expected to compete effectively with interactions with solvent molecules and anything less than a sizeable nucleus of interacting amino acid side chains would probably have a very short lifetime. Furthermore, it is important to stress that the

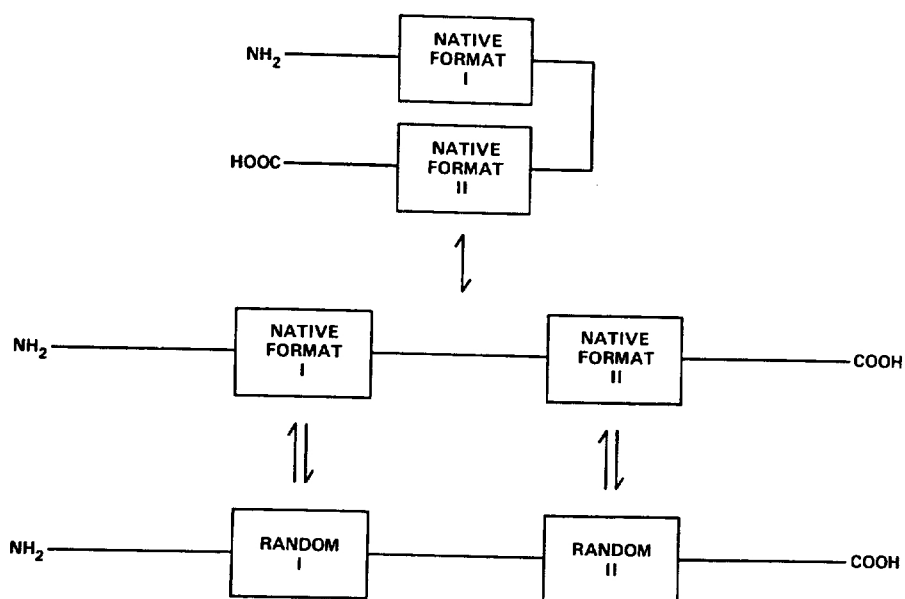


Fig. 10.

How protein chains might fold (see the text for a discussion of this fairly reasonable, but subjective proposal).

amino acid sequences of polypeptide chains designed to be the fabric of protein molecules only make functional sense when they are in the three dimensional arrangement that characterizes them in the *native protein structure*. It seems reasonable to suggest that portions of a protein chain that can serve as nucleation sites for folding will be those that can “flicker” in and out of the conformation that they occupy in the final protein, and that they will form a relatively rigid structure, stabilized by a set of cooperative interactions. These nucleation centers, in what we have termed their “native format”, (Figure 10) might be expected to involve such potentially self-dependent substructures as helices, pleated sheets or beta-bends.

Unfortunately, the methods that depend upon hydrodynamic or spectral measurements are not able to detect the presence of these infrequent and transient nucleations. To detect the postulated “flickering equilibria” and to determine their probable lifetimes in solution requires indirect methods that will record the brief appearance of individual “native format” molecules in the population under study. One such method, recently used in our laboratory in a study of the folding of staphylococcal nuclease and its fragments, employs specific antibodies against restricted portions of the amino acid sequence (51).

Figure 8 depicts the three dimensional pattern assumed by staphylococcal nuclease in solution. Major features involving organized structure are the three-stranded antiparallel pleated sheet approximately located between residues 12 and 35, and the three alpha-helical regions between residues 54—67, 99—106 and 121—134. Antibodies against specific regions of the nuclease molecule were prepared by immunization of goats with either polypeptide

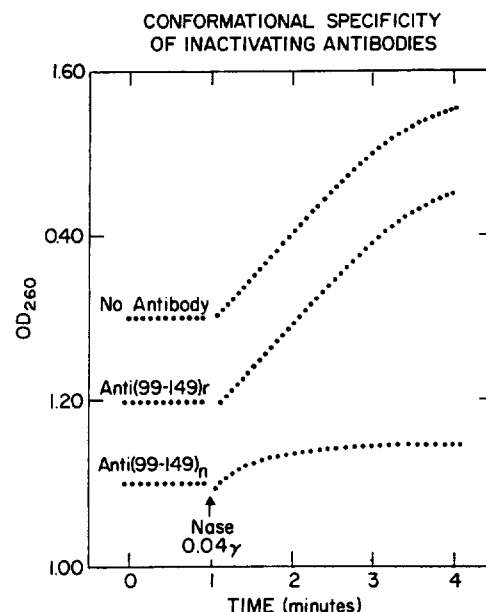


Fig. 11.  
Inhibition of nuclease activity by anti- $[(99-149)_r]$ , and lack of inhibition by anti- $[(99-149)_r]$  made against the peptide  $(99-149)$ , presumably in a random conformation (51).

fragments of the enzyme or by injection of the intact, native protein with subsequent fractionation of the resulting antibody population on affinity chromatography columns consisting of agarose bearing the covalently attached peptide fragment of interest (51, 52). In the former manner there was prepared, for example, an antibody directed against the polypeptide, residues 99–149, known to exist in solution as a random chain without the extensive helicity that characterizes this portion of the nuclease chain when present as part of the intact enzyme. Such an antibody preparation is referred to as anti-(99–149)<sub>r</sub>, the subscript indicating the disordered state of the antigen.

When, on the other hand, a fraction of anti-native nuclease serum, isolated on an agarose-nuclease column, was further fractionated on agarose-(99–149), a fraction was obtained which was specific for the sequence (99–149) but presumably only when this bit of sequence occupied the “native format”. This latter conclusion is based on the observation that the latter fraction, termed anti-(99–149)<sub>n</sub> (the subscript “n” referring to the native format) exhibited a strong inhibitory effect on the enzymic activity of nuclease whereas anti-(99–126)<sub>r</sub> or anti-(99–149)<sub>r</sub> were devoid of such an effect. (see Figure 11). This conclusion was further supported by the observation that the conformation-stabilizing ligands, pdTp and calcium ions, showed a marked inhibitory effect on the precipitability of nuclease by anti-(1–126)<sub>r</sub> and anti-(99–149)<sub>r</sub> but had little effect, if any on such precipitability by anti-(1–149)<sub>n</sub> (51). This finding reinforced the idea that many of the antigenic determinants recognized by the anti-*fragment* antibodies are present only in the “unfolded” or “non-native” conformation of nuclease. Analysis of the reaction between anti-(99–149) and nuclease could be shown by measurements of changes in the

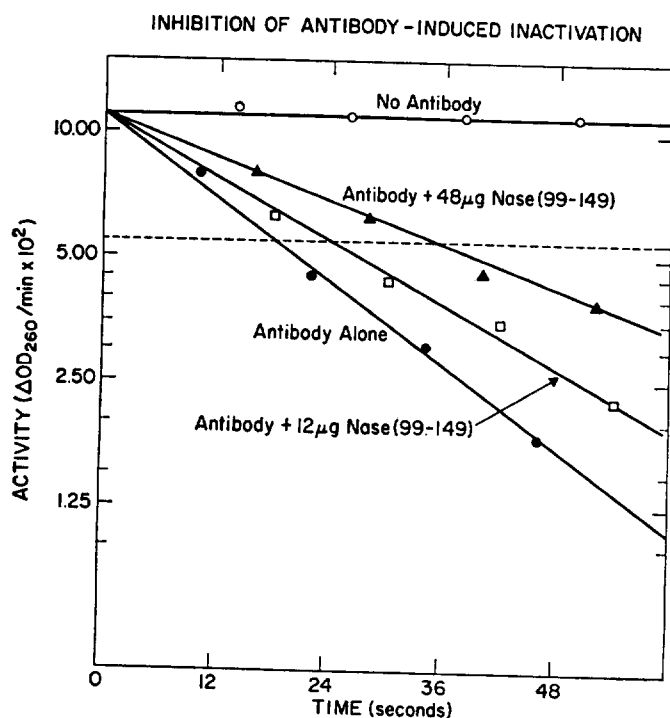


Fig. 12.

Semilogarithmic plot of activity *vs.* time for assays of 0.05  $\mu\text{g}$  of nuclease in the presence of:  $\circ - \circ$  no antibody,  $\bullet - \bullet$  6  $\mu\text{g}$  of anti-(99-126)<sub>n</sub>,  $\square - \square$  6  $\mu\text{g}$  of anti-(99-126)<sub>n</sub> plus 12  $\mu\text{g}$  of (99-149) and  $\blacktriangle - \blacktriangle$  6  $\mu\text{g}$  of anti-(99-126)<sub>n</sub> plus 48  $\mu\text{g}$  of (99-149). The dotted line represents one-half of the initial activity.

kinetics of inhibition of enzyme activity (Fig. 12), to be extremely rapid:  $k_{\text{off}}$ , on the other hand, is negligibly small.

The system may be described by two simultaneous equilibria, the first concerned with the "flickering" of fragment (99-149), which we shall term "P", from random to native "format", and the second with the association of anti-(99-149)<sub>n</sub>, which we shall term, simply, "Ab" with fragment P in its native format: i.e.  $P_n$ .

$$P_r \rightleftharpoons P_n \quad K_{\text{conf.}} = \frac{[P_n]}{[P_r]}$$

$$Ab + P_n \rightleftharpoons AbP_n \quad K_{\text{assoc.}} = \frac{[AbP_n]}{[Ab] \cdot [P_n]}$$

$$K_{\text{conf.}} = \frac{[AbP_n]}{K_{\text{assoc.}} \cdot [Ab] \cdot [P_r]}$$

Two equilibria involving fragment (99-149) of nuclease with the corresponding equilibrium-constant expressions.

The amount of unbound antibody in the second equilibrium may be estimated from measurements of the kinetics of inactivation of the digestion of denatured

Table 1

Studies of the equilibrium between the peptide fragment (99—149) in its random form [(99—149)<sub>r</sub>] and in the form this fragment assumes in the native structure of nuclease [(99—149)<sub>n</sub>]

Abbreviations: P, fragment (99—149); Ab, antibody.

$$K_{\text{conf.}} = \frac{[\text{AbP}_n]}{K_{\text{assoc.}} \cdot [\text{Ab}] \cdot [\text{P}_T]}$$

[Ab] <sub>total sites</sub> (μM)	[P <sub>T</sub> ] (μM)	t <sub>1/2</sub> (s)	[Ab] <sub>free sites</sub> (μM)	[Ab] <sub>bound sites</sub> (μM)	K <sub>conf.</sub>	% of P <sub>T</sub> as P <sub>n</sub>
0.076	0	18	0.076	0	—	—
0.076	0.65	20	0.068	0.0080	2.20 × 10 <sup>-4</sup>	0.022
0.076	2.0	24	0.057	0.019	2.02 × 10 <sup>-4</sup>	0.020
0.076	2.6	27	0.051	0.025	2.29 × 10 <sup>-4</sup>	0.023
0.076	7.8	35	0.039	0.037	1.47 × 10 <sup>-4</sup>	0.015
0.076	6.5	33	0.042	0.034	1.51 × 10 <sup>-4</sup>	0.015
K <sub>conf.</sub> = (2.0 ± 0.4) × 10 <sup>-4</sup>						

DNA substrate by a standard amount of nuclease added to the preincubated mixture of fragment (99—149) and anti-(99—149)<sub>n</sub>. Making the assumption that the affinity of anti-(99—149)<sub>n</sub> for (99—149)<sub>n</sub> in its folded (P) form is the same as that determined for this antigenic determinant in native nuclease, the value for the term K<sub>conf.</sub> may be calculated from measureable parameters. A series of typical values shown in Table 1, suggests that approximately 0.02 % of fragment (99—149) exists in the native format at any moment. Such a value, although low, is very large relative to the likelihood of a peptide fragment of a protein being found in its native format on the basis of chance alone.

Empirical considerations of the large amount of data now available on correlations between sequence and three dimensional structure (54), together with an increasing sophistication in the theoretical treatment of the energetics of polypeptide chain folding (55) are beginning to make more realistic the idea of the *a priori* prediction of protein conformation. It is certain that major advances in the understanding of cellular organization, and of the causes and control of abnormalities in such organization, will occur when we can predict, in advance, the three dimensional, phenotypic consequences of a genetic message.

#### BIBLIOGRAPHY

1. Anson, M. L., *Advan. Protein Chem.* 2, 361 (1945).
2. Lumry, R. and Eyring, H., *J. Phys. Chem.* 58, 110 (1954).
3. Hirs, C. H. W., Moore, S. and Stein, W. H., *J. Biol. Chem.* 235, 633 (1960).
4. Potts, J. T., Berger, A., Cooke, J. and Anfinsen, C. B., *J. Biol. Chem.* 237, 1851 (1962).
5. Smyth, D. G., Stein, W. H. and Moore, S., *J. Biol. Chem.* 238, 227 (1963).
6. Sela, M., White, F. H. and Anfinsen, C. B., *Science* 125, 691 (1957).
7. Gutte, B. and Merrifield, R. B., *J. Biol. Chem.* 246, 1922 (1971).
8. Hirschmann, R., Nutt, R. F., Veber, D. F., Vitali, R. A., Varga, S. L., Jacob, T. A., Holly, F. W. and Denkwalter, R. G., *J. Am. Chem. Soc.* 91, 507 (1969).
9. White, F. H., Jr. and Anfinsen, C. B., *Ann. N. Y. Acad. Sci.* 81, 515 (1959).

10. White, F. H., Jr., *J. Biol. Chem.* **236**, 1353 (1961).
11. Anfinsen, C. B., Haber, E., Sela, M. and White, F. H., Jr., *Proc. Nat. Acad. Sci. U. S.* **47**, 1309 (1961).
12. Haber, E. and Anfinsen, C. B., *J. Biol. Chem.* **237**, 1839 (1962).
13. Epstein, C. J., Goldberger, R. F. and Anfinsen, C. B., *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963).
14. Dintzis, H. M., *Proc. Nat. Acad. Sci. U. S.* **47**, 247 (1961).
15. Canfield, R. E. and Anfinsen, C. B., *Biochemistry* **2**, 1073 (1963).
16. Goldberger, R. F., Epstein, C. J. and Anfinsen, C. B., *J. Biol. Chem.* **238**, 628 (1963).
17. Venetianer, P. and Straub, F. B., *Biochim. Biophys. Acta* **67**, 166 (1963).
18. Fuchs, S., DeLorenzo, F. and Anfinsen, C. B., *J. Biol. Chem.* **242**, 398 (1967).
19. DeLorenzo, F., Goldberger, R. F., Steers, E., Givol, D. and Anfinsen, C. B., *J. Biol. Chem.* **241**, 1562 (1966).
20. Kato, I. and Anfinsen, C. B., *J. Biol. Chem.* **244**, 5849 (1969).
21. Givol, D., DeLorenzo, F., Goldberger, R. F. and Anfinsen, C. B., *Proc. Nat. Acad. Sci. U. S.* **53**, 766 (1965).
22. Steiner, D. F., *Trans. N. Y. Acad. Sci. Ser. II* **30**, 60 (1967).
23. Anfinsen, C. B., *Developmental Biology Supplement* **2**, 1 (1968), Academic Press Inc. U.S.A., 1968.
24. Richards, F. M., *Proc. Nat. Acad. Sci. U. S.* **44**, 162 (1958).
25. Taniuchi, H., Anfinsen, C. B. and Sodja, A., *Proc. Nat. Acad. Sci. U. S.* **58**, 1235 (1967).
26. Kato, I. and Tominaga, N., *FEBS Letters* **10**, 313 (1970).
27. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. and Richards, F. M., *J. Biol. Chem.* **245**, 305 (1970).
28. Hofmann, K., Finn, F. M., Linetti, M., Montibeller, J. and Zanetti, G., *J. Amer. Chem. Soc.* **88**, 3633 (1966).
29. Scoffone, E., Rocchi, R., Marchiori, F., Moroder, L., Marzotto, A. and Tamburro, A. M., *J. Amer. Chem. Soc.* **89**, 5450 (1967).
30. Potts, J. T., Jr., Young, D. M. and Anfinsen, C. B., *J. Biol. Chem.* **238**, 2593 (1963).
31. Taniuchi, H. and Anfinsen, C. B., *J. Biol. Chem.* **246**, 2291 (1971).
32. Cone, J. L., Cusumano, C. L., Taniuchi, H. and Anfinsen, C. B., *J. Biol. Chem.* **246**, 3103 (1971).
33. Bohnert, J. L. and Taniuchi, H., *J. Biol. Chem.* **247**, 4557 (1972).
34. Schechter, A. N., Moravsek, L. and Anfinsen, C. B., *J. Biol. Chem.* **244**, 4981 (1969).
35. Taniuchi, H., Moravsek, L. and Anfinsen, C. B., *J. Biol. Chem.* **244**, 4600 (1969).
36. Taniuchi, H. and Anfinsen, C. B., *J. Biol. Chem.* **244**, 3864 (1969).
- 37a. Taniuchi, H., Davies, D. and Anfinsen, C. B., *J. Biol. Chem.* **247**, 3362 (1972).
- b. Arnone, A., Bier, C. J., Cotton, F. A., Hazen, E. E., Jr., Richardson, D. C., Richardson, J. S. and Yonath, A., *J. Biol. Chem.* **246**, 2302 (1971).
38. Sanchez, G. R., Chaiken, I. M. and Anfinsen, C. B., *J. Biol. Chem.*, 1973, in press.
39. Ontjes, D. and Anfinsen, C. B., *J. Biol. Chem.* **244**, 6316 (1969).
40. Merrifield, R. B., *Science* **150**, 178 (1965).
41. Chaiken, I. M., *J. Biol. Chem.* **246**, 2948 (1971).
42. Chaiken, I. M. and Anfinsen, C. B., *J. Biol. Chem.* **246**, 2285 (1971).
43. Parikh, I., Corley, L. and Anfinsen, C. B., *J. Biol. Chem.* **246**, 7392 (1971).
44. Fitch, W. M. and Margoliash, E., *Evolutionary Biology*, **4**, 67 (1970), Edited by Th. Dobzhansky, M. K. Hecht and W. C. Steere, Appleton-Century-Crofts, New York, 1970.
45. Cooke, J. P., Anfinsen, C. B. and Sela, M., *J. Biol. Chem.* **238**, 2034 (1963).
46. Perutz, M. F., Kendrew, J. C. and Watson, H. C., *J. Mol. Biol.* **13**, 669 (1965).
47. Epstein, C. J., *Nature* **210**, 25 (1966).
48. D. Sachs, H. Taniuchi, A. N. Schechter and A. Eastlake, unpublished work.
49. Epstein, H. F., Schechter, A. N., and Cohen, J. S. *Proc. Nat. Acad. Sci. U. S.*, **68**, 2042 (1971).



50. Epstein, H. F., Schechter, A. N., Chen, R. F. and Anfinsen, C. B., *J. Mol. Biol.* *60*, 499 (1971).
51. Sachs, D. H., Schechter, A. N., Eastlake, A. and Anfinsen, C. B., *Proc. Nat. Acad. Sci. U. S.*, *69*, 3790, (1972).
52. Sachs, D. H., Schechter, A. N., Eastlake, A. and Anfinsen, C. B., *J. Immunol.* *109*, 1300 (1972).
53. Sachs, D. H., Schechter, A. N., Eastlake, A. and Anfinsen, C. B., *Biochemistry*, *11*, 4268 (1972).
54. Anfinsen, C. B. and Scheraga, H., *Adv. in Prot. Chem.*, Vol. 27, 1973, in preparation.
55. H. A. Scheraga, *Chemical Reviews*, *71*, 195 (1971).