Probing Protein Structure and Function with an Expanded Genetic Code

Virginia W. Cornish, David Mendel, and Peter G. Schultz*

A general biosynthetic method has been developed which makes it possible to site-specifically incorporate unnatural amino acids with novel properties into proteins. In this approach the codon encoding the amino acid of interest is replaced with the "blank" nonsense codon UAG by oligonucleotide-directed mutagenesis. A suppressor tRNA that recognizes this codon is generated by run-off transcription and then chemically aminoacylated with the desired unnatural amino acid. Addition of the mutagenized gene and the aminoacylated suppressor tRNA to an in vitro extract capable of supporting protein biosynthesis generates a mutant protein containing the unnatural amino acid at the specified position. This methodology has recently been used to study the stability, specificity, and catalytic properties of a number of proteins. In these studies amino acids and analogues possessing altered hydrogen-bonding, electronic, and steric properties and unique backbone conformations have all been site-specifically incorporated into proteins. In addition, photoactivatable amino acids, isotopically labeled amino acids, and amino acids bearing biophysical probes have been inserted site-specifically. This chemistry increases our ability to carry out detailed physical organic studies on this important class of macromolecules.

Keywords: amino acids · mutagenesis · proteins

1. Introduction

Proteins are at the crossroads of virtually every biological process, including signal transduction, catalysis, gene regulation, and the immune response. Yet we still understand little about the mechanisms whereby these biopolymers, composed of twenty simple building blocks, carry out their remarkable range of functions. One important tool for probing the forces that govern protein structure and folding, biomolecular recognition, and catalysis is site-directed mutagenesis. This method makes it possible to substitute a specific amino acid in a protein with any of the other nineteen common amino acids.[1, 2] However, in contrast to studies of small molecules for which analogues with virtually any structural changes can be synthesized, changes in protein structure at a given site are limited to the common amino acids. Ideally, one would like to tailor the steric or electronic properties of an amino acid in a protein to address a specific structure–function question. Desired changes might include modifying the acidity, nucleophilicity, hydrogen-bonding, or hydrophobic properties of an amino acid side chain; altering or restricting the protein backbone conformation; or introducing biophysical probes such as spin labels or isotopic labels at a specific position within a protein. The possibility of substituting such unnatural amino acids, beyond those specified by the genetic code, would greatly expand our ability to manipulate and study protein structure and function and may allow for the generation of new proteins with novel properties.

A number of methods can be used to incorporate unnatural amino acids into proteins. Solid-phase peptide synthesis, particularly recent advances in the segment synthesis-condensation approach,[3] has allowed for the synthesis of small proteins ($\leq$12 kDa) containing novel amino acids in milligram quantities.[4–11] Protein semisynthesis, in which a synthetic peptide is ligated to a protein fragment to produce a full-length protein, has been used to incorporate unnatural amino acids into proteins.[12, 13] This approach is complicated by the need to cleave the protein specifically at the peptide ligation site and by difficulties in coupling the protein and peptide termini selectively. Chemical modification also has been used to introduce a variety of unnatural side chains into proteins including cofactors, spin labels, and oligonucleotides.[14–20] In this method, however, substitutions are largely restricted to simple derivatives of reactive amino acid side chains (e.g. lysine, cysteine, tyrosine) on the protein surface. Alternatively, biosynthetic methods that employ chemically modified aminoacyl-tRNAs have been used to incorporate a number of biophysical probes into proteins synthesized in vitro.[21, 22] This approach is limited, though, because mutations are restricted to derivatives of the natural amino acids and the modified amino acid is substituted at multiple sites. Furthermore, both the modified and natural amino acids are incorporated at the site of interest because of

[*] Prof. P. G. Schultz, V. W. Cornish
Department of Chemistry, University of California
Berkeley, CA 94720 (USA)
Telefax: Int. code +(510)643-6890
Dr. D. Mendel
Lilly Research Laboratories
Eli Lilly and Company
Indianapolis, IN 46285 (USA)

  0570-0833/95/0606-0621 $ 10.00+.25/0

difficulties in removing endogenous aminoacyl-tRNA from the in vitro extract.

Recently, a biosynthetic approach has been developed which for the first time makes possible the site-specific incorporation of a large variety of unnatural amino acids into proteins.[23-25] In this article we describe the methodology as well as its current scope and limitations. In addition, applications are described in which the substitution of unnatural amino acids with novel electronic, steric, and biophysical properties into proteins has provided new insights into the mechanisms of enzymatic catalysis, protein stability, and biomolecular recognition.

2. The Methodology

2.1. Design Considerations

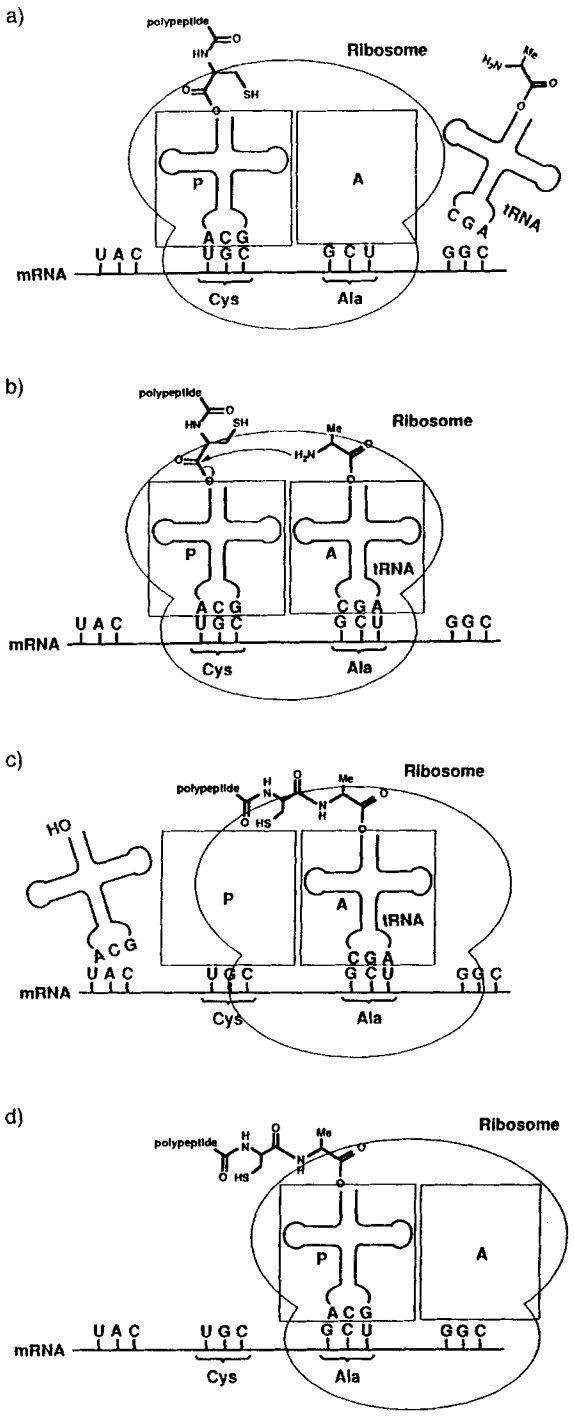
The ability to incorporate unnatural amino acids site-specifically into proteins depends on the following observations: 1) Although three stop codons (UAA, UAG, and UGA) exist, only one is required for the termination of protein synthesis (Scheme 1), leaving two "blanks" in the genetic code (Scheme 2).[26,27] 2) Suppressor tRNAs, tRNAs that insert one of the common amino acids in response to a stop codon, function efficiently in vivo and in vitro.[28] 3) Anticodon–codon recognition is independent of the amino acid at the acceptor stem of the tRNA (the adaptor hypothesis).[29,30] 4) The translational machinery shows relatively broad substrate specificity.[31,32] In light of these facts, the following strategy was developed (Scheme 3). The codon for the amino acid of interest is replaced with the stop codon UAG through standard oligonucleotide-directed mutagenesis.[33] A suppressor tRNA that recognizes this codon is chemically acylated with the desired unnatural amino acid. Addition of the mutagenized gene or mRNA and the aminoacylated suppressor tRNA to an in vitro extract capable of supporting protein biosynthesis generates a mutant protein containing the unnatural amino acid at the desired position.[23,24] The development of this methodology requires the generation of the requisite suppressor tRNA, a general method for acylating the tRNA with unnatural amino acids, and the synthesis of sufficient protein in vitro for meaningful mechanistic or structural studies.

2.2. The Suppressor tRNA

The facile incorporation of an unnatural amino acid into a protein requires a tRNA that is recognized by the protein biosynthetic machinery sufficiently well to suppress a stop codon with high efficiency. This suppressor tRNA, however, cannot be a substrate for any of the aminoacyl-tRNA synthetases present in the in vitro protein synthesis extract. If the suppressor tRNA were recognized by any of the aminoacyl-tRNA synthetases, the tRNA could be subject to proofreading (deacylation of the noncognate amino acid) and/or aminoacylation with one of the common twenty amino acids. This recognition would result in either low suppression efficiencies or incorporation of a common amino acid, as well as the desired unnatural amino acid, in response to the stop codon.

We initially chose to construct an amber suppressor tRNA (a tRNA that recognizes the stop codon UAG) derived from yeast phenylalanyl-tRNA in which residues 34 to 37 have been replaced by 5′-CUAA-3′[34] (in accord with Yarus' rule).[35-37] This suppressor tRNA is known to function efficiently in vitro[34] and is likely to be a poor substrate for the *E. coli* aminoacyl-tRNA synthetases.[38] Alternatively, a suppressor tRNA derived from *E. coli* $tRNA^{Gly}$[39] has been used to suppress UAG codons in short peptides in a rabbit reticulocyte translation system.[25] Both suppressor tRNAs can be readily prepared in relatively large quantities by run-off transcription (Scheme 4).[40,41] Even though these tRNAs are produced in vitro and likely do not contain modified bases,[42] both incorporate unnatural amino acids efficiently and exhibit little readthrough (incorporation of any amino acid other than the desired unnatural amino acid in response to the UAG codon) at optimal Mg^{2+} concentrations.

Peter G. Schultz was born in 1956 in Cincinnati, Ohio, and did both his undergraduate and graduate work at the California Institute of Technology. His thesis work resulted in the design of the first synthetic sequence-selective DNA cleaving molecules. In 1985, after postdoctoral studies at the Massachusetts Institute of Technology, he joined the faculty of the University of California at Berkeley, where he is currently a Professor of Chemistry, a Principal Investigator at the Lawrence Berkeley Laboratory, and an Investigator of the Howard Hughes Medical Institute. His research interests span organic chemistry, molecular biology, and immunology. His research focuses on the study of mechanisms of molecular recognition and catalysis in biological systems and applications to the design of molecules with novel biological (and medicinal) function. His group has contributed to the development of catalytic antibodies, DNA binding proteins, peptides, and oligonucleotides, RNA compounds, and unnatural biopolymers such has oligourethanes and oligosulfones. He has developed methods for the incorporation of unnatural amino acids in proteins and for the manipulation of enzyme specifity through chemical and biological mutagenesis. Schultz has received many awards including the Waterman Award of the National Science Foundation, the Pure Chemistry Award of the American Chemical Society, the Young Investigator Award of the Protein Society, and the Wolf Prize in 1994.



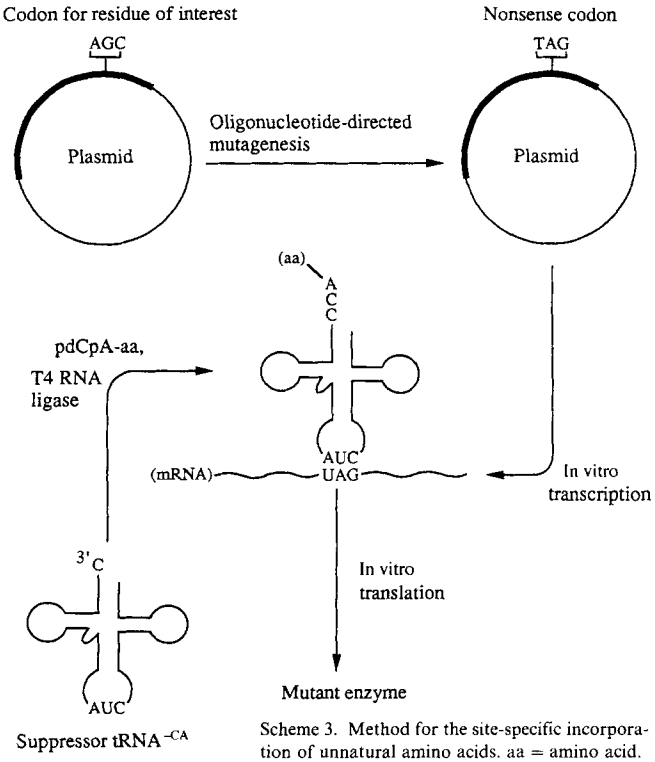
Scheme 1. Schematic representation of protein biosynthesis: a) binding of the aminoacyl-tRNA to the ribosomal complex; b) peptide bond formation; c) dissociation of the deacylated tRNA; d) ribosome translocation.

2.3. Aminoacylation of the tRNA

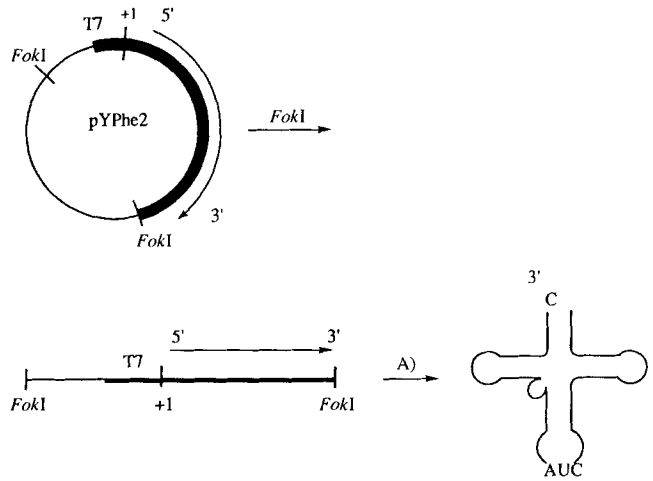
With the suppressor tRNA in hand, a general method for acylation of the tRNA with any unnatural amino acid had to be developed. Enzymatic aminoacylation of the tRNA with aminoacyl-tRNA synthetases is not general because of the high specificity of these enzymes.[43, 44] Direct chemical acylation is probably not practical owing to the large number of reactive sites in the tRNA. Consequently, a modification of a two-step

UUU, UUC	Phe	UCU, UCC, UCA, UCG	Ser	UAU, UAC	Tyr	UGU, UGC	Cys
UUA, UUG	Leu			UAA, UAG	TERM	UGA	TERM
						UGG	Trp
CUU, CUC, CUA, CUG	Leu	CCU, CCC, CCA, CCG	Pro	CAU, CAC	His	CGU, CGC, CGA, CGG	Arg
				CAA, CAG	Gln		
AUU, AUC, AUA	Ile	ACU, ACC, ACA, ACG	Thr	AAU, AAC	Asn	AGU, AGC	Ser
AUG	Met			AAA, AAG	Lys	AGA, AGG	Arg
GUU, GUC, GUA, GUG	Val	GCU, GCC, GCA, GCG	Ala	GAU, GAC	Asp	GGU, GGC, GGA	Gly
				GAA, GAG	Glu		

Scheme 2. The genetic code, TERM = termination codon.

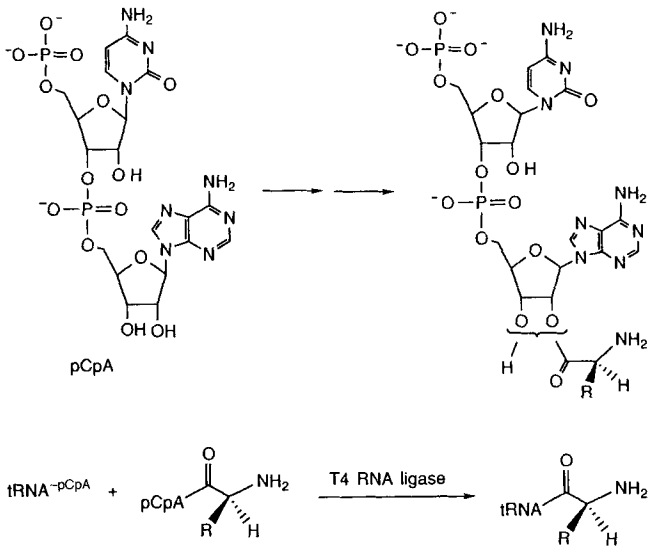


Scheme 3. Method for the site-specific incorporation of unnatural amino acids. aa = amino acid.



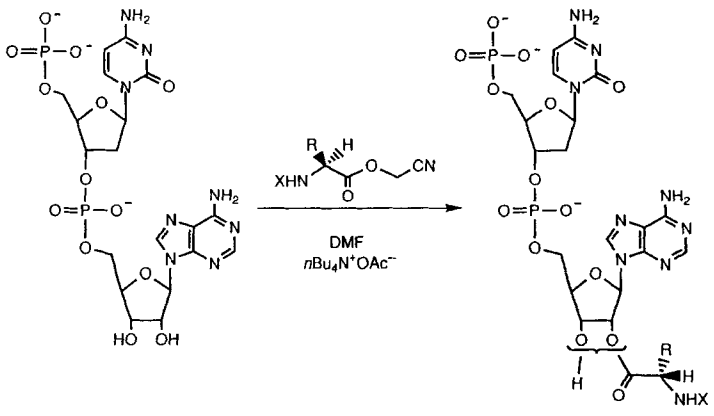
Scheme 4. Preparation of $tRNA^{Phe}_{CUA}$ by run-off transcription. The heavy line represents the DNA sequence coding $tRNA^{-CA}_{CUA}$. A) 1. T7 RNA polymerase, NTPs; 2. enzyme extraction, precipitation with EtOH. One thousand copies of tRNA are obtained per copy of DNA, in other words, roughly 1 mg tRNA per 30 μg digested template DNA.

method originally developed by Hecht et al.[45] was used in which the dinucleotide pCpA is chemically acylated with a N^{α}-protected amino acid and then enzymatically ligated to a truncated tRNA ($tRNA^{-CA}$ missing the terminal dinucleotide pCpA at the 3′-acceptor stem) (Scheme 5).[46] The truncated suppressor tRNA can be generated directly by run-off transcription of a *Fok*I linearized plasmid encoding the tRNA gene.[41]



Scheme 5. Strategy for the chemical aminoacylation of $tRNA^{Phe}_{CUA}$. R = amino acid side chain.

Acylation of the dinucleotide is complicated by side reactions including diacylation and acylation of the exocyclic amino groups of the bases which result in low yields of the desired product. Moreover, the aminoacyl tRNA linkage itself is relatively unstable in the pH range (7–8) required for efficient T4 RNA ligase function.[47, 48] We found that the cyanomethyl ester of an amino acid reacts selectively with the 2′,3′-hydroxyl groups of the ribose to give the monoacylated product, obviating the need to protect the dinucleotide prior to aminoacylation (Scheme 6).[49] In addition, deoxycytidine was substituted for cytidine in the dinucleotide, significantly simplifying the synthesis and eliminating another 2′-OH group without affecting biological activity. The α-amino group of the amino acid and any reactive side chains are protected as the nitroveratryloxy (NVOC) carbamate, ester, or ether[50, 51] prior to acylation of the dinucleotide. These protecting groups can be removed photochemically from the intact aminoacyl-tRNA (after ligation) in high yield under mildly acidic conditions which prevent deacylation of the aminoacyl-tRNA.[49] For light-sensitive amino acids the biphenylisopropyloxycarbonyl (BPOC) group can be used; the BPOC group is cleaved with mild acid from the aminoacyl-pdCpA prior to ligation to $tRNA^{-CA}$.[25, 49] This aminoacylation protocol is relatively straightforward, proceeds in high yield (both the aminoacylation and ligation reactions), and is applicable to amino acids with a wide variety of side chains.[48, 24] Nevertheless, the generation of a nonspecific "aminoacyl tRNA synthetase", either a lipase, mutagenized synthetase, or a catalytic antibody, could significantly simplify this step.[52–54]



Scheme 6. Simplified acylation of pdCpA using an amino acid cyanomethyl ester. R = amino acid side chain, X = protecting group.

2.4. In Vitro Protein Synthesis

Proteins are synthesized in vitro because no general methodology exists for introducing large quantities of the aminoacylated suppressor tRNA into intact, dividing cells.[55] We currently use an *E. coli* in vitro transcription/translation protein synthesis system.[56, 57, 24] This system consists of an *E. coli* S-30 extract which contains all of the proteins and RNAs required for transcription and translation. Nucleotide triphosphates, bacterial tRNAs, phosphoenolpyruvate, pyrophosphatase, a variety of salts (Mg^{2+} and Ca^{2+} concentration can significantly affect protein yields) and cofactors, and the DNA encoding the gene of interest are added to this S-30 extract to reconstitute protein synthesis in vitro. Typically the gene of interest is placed under a strong T7 or bacterial RNA polymerase promoter. The in vitro reaction can be carried out on scales ranging from 30 μL to 5 mL. The requisite S-30 can be prepared reasonably on a 2 L scale, the suppressor tRNA on a 30 mg scale, the aminoacyl-pdCpA on a 0.2 mmol scale, and the plasmid on a 12 mg scale. A major limitation of in vitro protein synthesis remains the relatively small quantities of protein that can be obtained. Regardless of whether the protein is synthesized from plasmid DNA using an *E. coli* transcription/translation system or from mRNA using a rabbit reticulocyte translation system, protein yields greater than $100\,\mu g\,mL^{-1}$ generally have not been achieved.[24] Many factors are known to affect protein expression levels in vivo, such as the strength of the RNA polymerase promoter, the particular ribosome-binding site, codon usage, and mRNA stability and secondary structure.[58, 59] Unfortunately, in vitro protein yields do not necessarily correlate with those obtained in vivo, and we have found no simple and general rules for ensuring high protein expression in vitro.

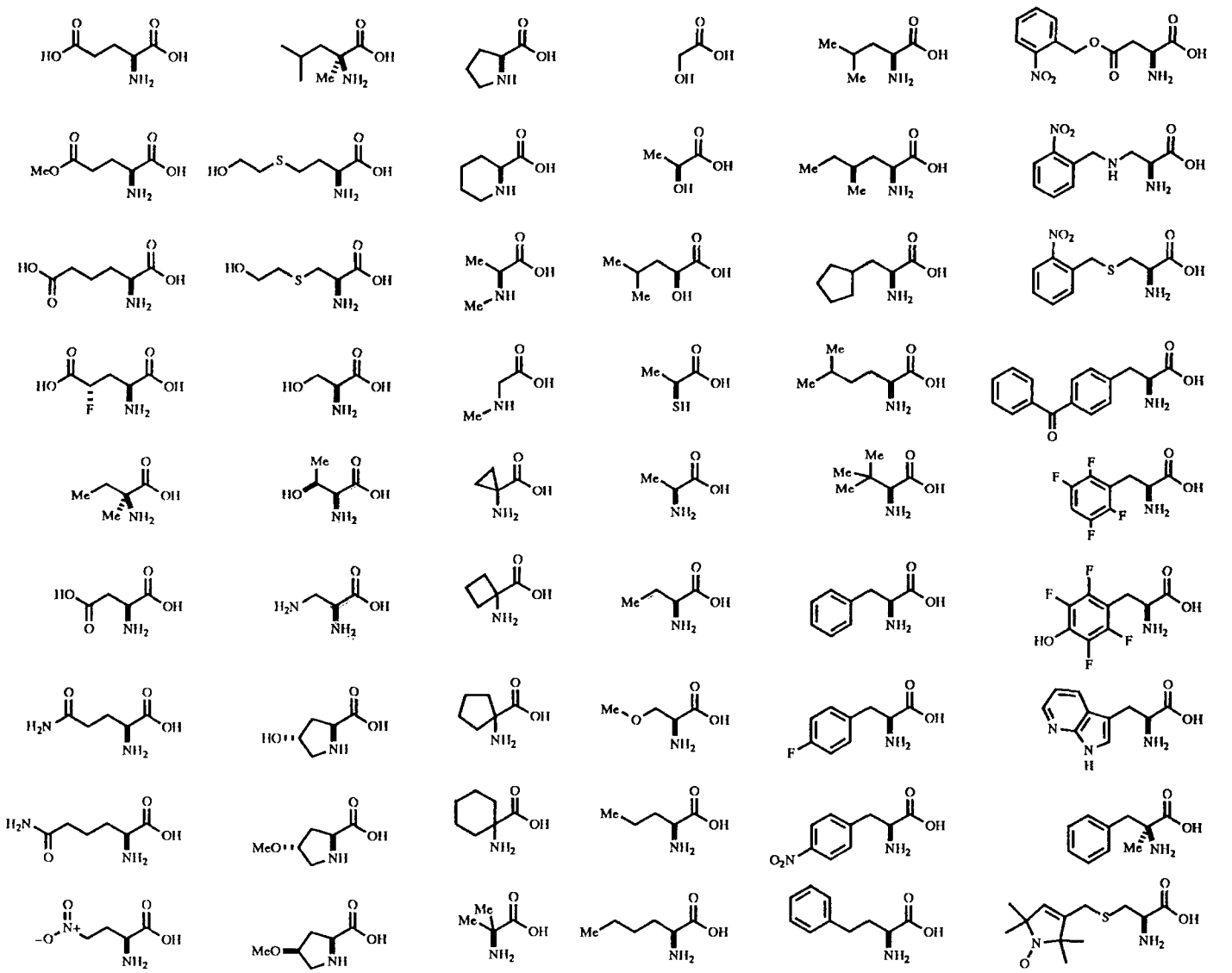
2.5. Scope of the Methodology

The relatively broad substrate specificity of the elongation factors and peptidyl transferase suggested that a wide variety of amino acids could be incorporated into proteins.[31, 32] This is in fact the case—over seventy amino acids have been incorporated at more than twenty sites in a variety of proteins by using the

above methodology (Scheme 7). Conformationally restricted amino acids (e.g. 2,4-methanoproline (not shown), cyclopropylglycine, and α-methyl-L-leucine), amino acids with spin labels and photoaffinity labels as side chains, amino acids with altered pK_A values and hydrogen-bonding properties, "caged" amino acids, α-hydroxy acids, and amino acids with unusual steric properties (e.g. *tert*-butylglycine) have all been incorporated into proteins. Suppression efficiencies can vary widely depending on the nature of the amino acid, ranging from 100% for substitution of L-norleucine for Gln 28 in 434 repressor[60] to 10% for incorporation of α-methyl-L-leucine in T4 lysozyme.[61] Although the suppression efficiency of a given amino acid cannot be predicted a priori, several trends emerge when the identities of the amino acids tested are correlated with their suppression efficiencies.[61, 62] In general, large hydrophobic amino acids such as *p*-benzoyl-L-phenylalanine are inserted with higher efficiency than small amino acids such as alanine and glycine, or charged amino acids such as homoglutamate or ornithine. This correlation is in agreement with reports that large, hydrophobic amino acids such as Trp and Phe bind more tightly to the *E. coli* elongation factor EF-Tu than do other natural amino acids such as Ala, Glu, and Lys.[63]

The stereochemistry of an amino acid also affects suppression efficiency. Although L-amino acids and some α,α-disubstituted amino acids such as α-aminoisobutyric acid (AIB) and cyclopentylglycine can be incorporated into proteins, D-amino acids do not seem to be accommodated by the translational machinery.[61] These stereochemical requirements are in agreement, for the most part, with previous studies on the affinity of a variety of unnatural aminoacyl-tRNAs for *E. coli* EF-Tu and the ribosome.[31] Interestingly, previous experiments, in which the ability of an aminoacyl-tRNA or aminoacyl-CpA to bind to the *E. coli* ribosomal A-site and participate in peptide elongation was studied, suggested that achiral amino acids like AIB could not be used as substrates by the *E. coli* protein biosynthetic machinery.[64–66]

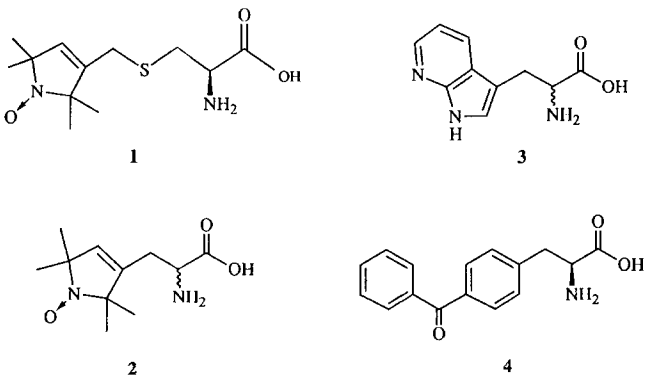
A variety of backbone mutations are accomodated by the translational machinery.[67] *N*-Methyl-L-phenylalanine (rabbit reticulocyte), *N*-methylglycine (*E. coli*) and *N*-methyl-L-alanine



Scheme 7. Examples of amino acids and analogues that have been successfully incorporated into proteins.

(*E. coli*) are site-specifically incorporated into proteins, whereas *N*-ethyl-L-alanine is not.[68] Pipecolinic acid, the six-membered version of proline, is accepted by the *E. coli* system, whereas the ring-contracted proline analogues aziridine-2-carboxylic acid and azetidine-2-carboxylic acid are not. (This may stem from chemical instability rather than from discrimination by the biosynthesis machinery.) However, Tirrell and others[69, 70] have reported the successful biosynthetic incorporation of L-azetidine-2-carboxylic acid into polypeptides. Other types of backbone replacements for α-amino acids that have been examined include β- and γ-amino acids such as β-alanine, γ-aminobutyric acid, and longer variants[71] as well as dipeptides and dipeptide mimetics.[72] To date, attempts to incorporate these amino acids using either *E. coli* or rabbit reticulocyte lysate systems have proven unsuccessful.

The precise structure of an amino acid also influences its suppression efficiency dramatically. For example, even though the spin-labeled amino acids **1** and **2** both contain the 1-oxyl-2,2,5,5-tetramethylpyrrolidine group (Scheme 8), only **1** could



Scheme 8. Structure of biophysical probes tested for incorporation into T4L.

be incorporated in vitro.[62] In addition, suppression efficiencies may vary with the nature of the in vitro protein synthesis system (prokaryotic vs. eukaryotic). While L-4′-(3-trifluoromethyl-3*H*-diazirin-3-yl)phenylalanine was incorporated poorly (<10% efficiency) when an *E. coli* in vitro transcription/translational system was used,[73] the same amino acid was inserted with higher efficiency (>40%) in a rabbit reticulocyte in vitro extract.[74] Finally, in contrast to in vivo results,[35–37] there appears to be little correlation between the codons adjacent to the UAG codon and suppression efficiencies.[62] As more unnatural amino acids are substituted into proteins, this methodology should provide additional insights into the specificity of the translational machinery and the factors governing in vitro suppression efficiency.

2.6. Future Work

Future work will focus on broadening the scope of the methodology by increasing protein yields, by minimizing the effect of the amino acid identity on suppression efficiency, and by developing ways to introduce more than one amino acid into a protein at a time. Efforts to improve protein yields will require improvements in in vitro protein synthesis. Higher protein yields may be achieved by using in vitro systems derived from other sources[55] or alternative approaches such as continuous flow in vitro systems.[75] Alternatively, since the *E. coli* transcription/translation system is active at 37 °C for about only one hour, protein expression levels may be increased by using other lysates such as lysed spheroplasts, capable of supporting protein synthesis for longer periods of time.[76, 77]

Suppression efficiencies may be improved by 1) altering nucleotides in the tRNA molecule involved in tRNA recognition by the elongation factors, the ribosome, or the aminoacyl-tRNA synthetases; 2) utilizing suppressor tRNAs from different species (here tRNA must be "engineered" to eliminate recognition by the corresponding aminoacyl-tRNA synthetase); 3) deleting release factor-1 (*E. coli*), which competes with the suppressor tRNA for recognition of the UAG codon, from the in vitro extract;[78] or 4) modifying the specificity of elongation factor EF-Tu. One would also like to be able to produce proteins containing several unnatural amino acids. A single unnatural amino acid can be incorporated at more than one position in a protein simply by introducing several UAG codons in the coding sequence.[79] Preliminary efforts to insert different amino acids into the same protein have focused on using the other nonsense codons,[80] infrequently occurring codons,[81] or a 65th codon composed of nonstandard nucleotides.[82]

3. Applications

In this section, several applications of the methodology are discussed to illustrate both the scope of the methodology and the kinds of questions that can be addressed when unnatural amino acids can be incorporated into proteins. These examples include the site-specific incorporation of biophysical probes into proteins, the construction of photocaged proteins, and studies of protein stability, enzyme mechanism, and signal transduction.

3.1. Biophysical Probes

Several amino acids that can serve as biophysical probes have been incorporated into proteins through unnatural amino acid mutagenesis (Scheme 8). A novel spin-labeled amino acid **1**, an unnatural fluorescent amino acid **3**, and an efficient cross-linking agent **4** were all introduced site-specifically at several external and internal positions in T4 lysozyme (T4L) with reasonable efficiency.[62] By taking advantage of recent loop-gap resonator technology, the ESR spectrum of approximately 10 μL of a 1 μM solution of purified T4L containing the spin-labeled amino acid **1** was measured. Replacement of one of the tryptophans in T4L by 7-azatryptophan (**3**) resulted in a red shift in the fluorescent emission maxima. In addition, amino acids that form active nitrene, carbene, and ketyl radicals have been successfully incorporated into *ras* p21 and T4L at specific positions.[62, 83] Photoaffinity probes can in principle be used to identify key components involved in complex biochemical processes, for example signal transduction and gene activation.

With considerably more effort, a sufficient quantity of purified T4L specifically labeled at site Ala 82 with [^{13}C]Ala was produced. The proton resonances of the labeled residue were observed selectively, in both the native and denatured states, by using ^{13}C-decoupled ^{1}H NMR spectroscopy (Fig. 1).[84] Finally, enough staphylococcal nuclease (SNase) containing the unnatural amino acid homoglutamate at position Glu 43 was obtained for determination of the mutant protein's structure by X-ray crystallography to a resolution of 2.4 Å.[85] The ability to incorporate probes capable of detecting local structure and dynamics into proteins site-specifically should allow for more precise studies of protein folding and stability, conformational changes in proteins, and the interaction of proteins with other proteins, small molecules, and membranes. For example, it should be possible to carry out site-specific spectroscopy on both the folded and unfolded states of proteins, compare denaturation curves for various regions of a protein, and probe local structural changes that occur during protein translocation.

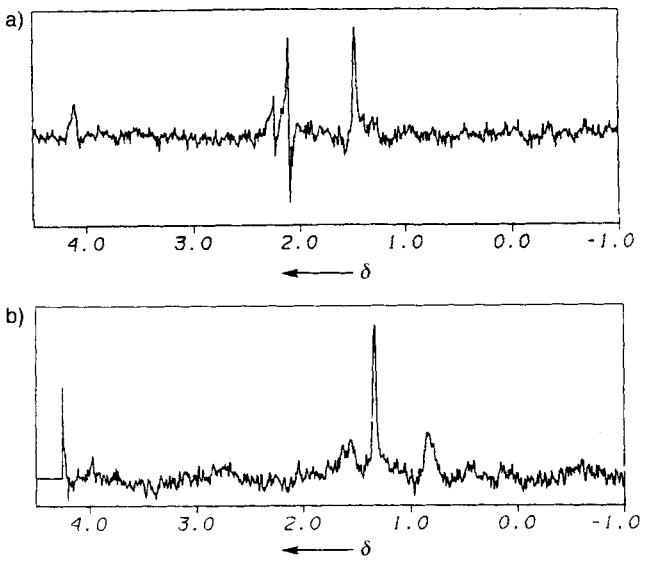
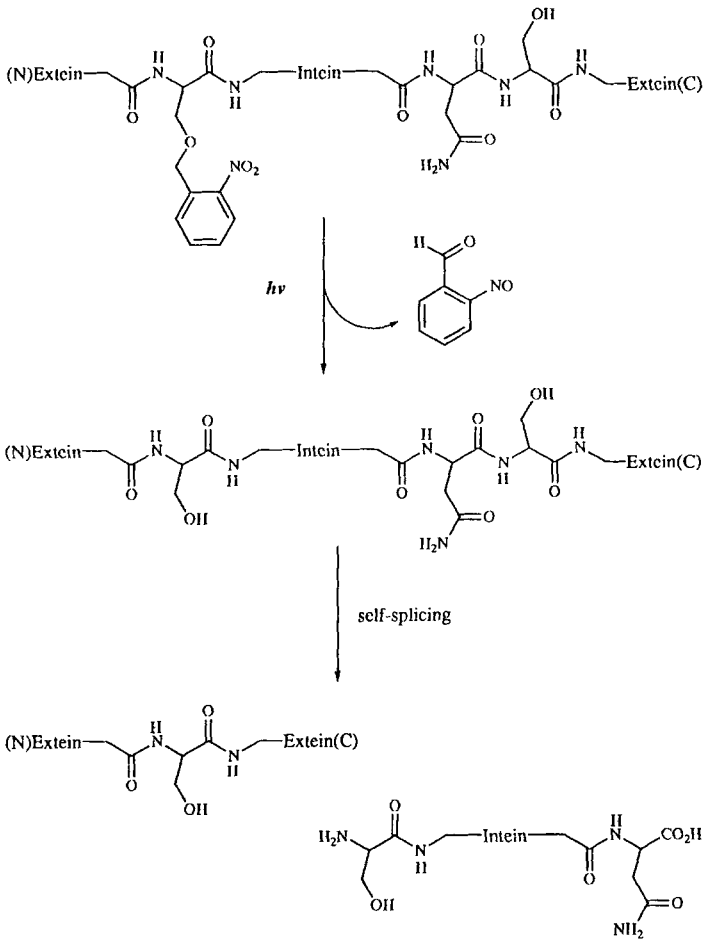


Fig. 1. a) ^{13}C-Decoupled ^{1}H NMR spectrum of the T4L variant Ala 82 → [^{13}C]Ala; b) ^{13}C-decoupled ^{1}H NMR spectrum of the denatured form of the same sample.

3.2. "Caged" Proteins

Low molecular weight "caged" substances, inactive precursor molecules that can be rapidly activated by photolysis, are useful tools for biochemical studies.[86, 87] Mutagenesis with unnatural amino acids provides a unique approach to the construction of photocaged proteins. For example, the unnatural amino acid *ortho*-β-nitrobenzylaspartate was substituted for the active-site residue Asp 20 in T4L.[88] Since Asp 20 is essential for catalytic activity, the resulting protein was inactive. Removal of the protecting group by photolysis of the intact, modified protein restored full catalytic activity.

In a second example, a serine residue thought to be involved in the self-splicing reaction of the protein Vent DNA polymerase was replaced by *O*-nitrobenzylserine (Scheme 9).[89] The resulting caged polymerase did not undergo the self-splicing reaction, allowing the full-length protein to be isolated. Photolysis of the unspliced protein resulted in loss of the nitrobenzyl protecting group and subsequent protein splicing. This result demonstrates unequivocally that protein splicing occurs at the posttranslational level. The ability to construct caged proteins of this sort should make possible a broad range of time-resolved experiments relevant to catalytic mechanism, biomolecular recognition, and protein folding.



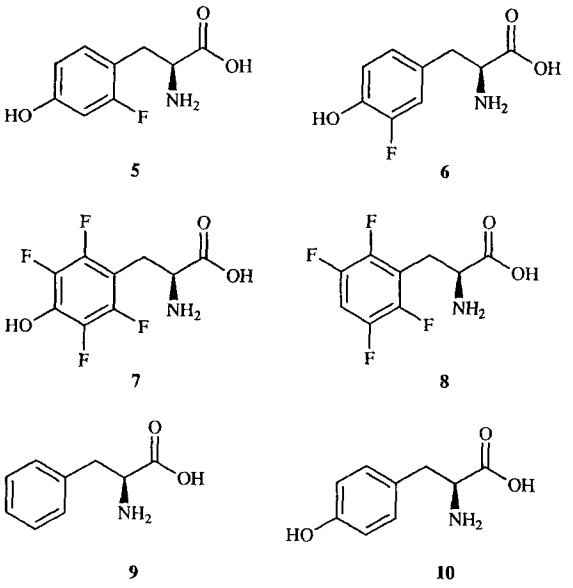
Scheme 9. Photoactivated self-splicing of Vent polymerase using a caged serine analogue. The mechanism of the protein self-splicing mechanism is not known.

3.3. Protein Stability

Mutagenesis with unnatural amino acids has been used to make defined mutations in proteins in the hope of gaining a more precise picture of the forces governing protein stability.[90, 91] An outstanding issue in the literature on protein stability and folding is the role of hydrogen bonding. Although hydrogen bonds are clearly important in determining the secondary and tertiary structure of proteins,[92, 93] the magnitude of their contribution to protein stability has been difficult to assess.[94, 95] Stabilization values for hydrogen bond formation determined from studies with mutants, in which one member of a hydrogen-bonded pair is deleted, vary considerably depending upon the nature and local environment of the hydro-

gen bond.[90, 96] Moreover, a concern in all such studies with mutants is the possibility of introducing additional destabilizing interactions by leaving an unpaired hydrogen bond donor or acceptor and/or altering local solvation and packing interactions (e.g. a Glu-Phe versus a Glu-Tyr pair).

In an effort to determine the degree to which side-chain hydrogen bonding stabilizes the folded state of proteins while minimizing the steric and electronic perturbations associated with deleting one member of a hydrogen-bonded pair, Tyr27 in staphylococcal nuclease (SNase) was replaced with a number of isosteric, fluorinated tyrosine analogues, including 2-fluorotyrosine (**5**), 3-fluorotyrosine (**6**), and tetrafluorotyrosine (**7**) as well as 2,3,5,6-tetrafluorophenylalanine (**8**) and phenylalanine (**9**) (Scheme 10).[97] These mutants were designed to examine



Scheme 10. Structures of the amino acids with altered hydrogen-bonding properties substituted for Tyr27 in SNase.

the effect of increasing the strength of the Tyr27–Glu10 hydrogen bond on protein stability. Denaturation studies of the corresponding mutants revealed a unique free energy correlation between stability constant K_{app} ($\alpha = 0.38$) and the pK_a of the hydroxyl group in the tyrosine analogue in position 27 (Fig. 2). Indeed, substitution of Tyr (pK_a ca. 10) with tetrafluoro-L-tyrosine (pK_a ca. 5) increased the stability of SNase by 2.3 kcal mol^{-1}. This experiment provides strong evidence that side-chain intramolecular hydrogen bonds can preferentially stabilize the folded state of a protein relative to the unfolded state in water. In addition this study effectively extends the use of substituent effects in physical organic chemistry from small molecules to proteins.

A series of mutations were also made in T4L to examine the role of backbone hydrogen bonding in determining protein stability.[68] The surface amino acid Ala82 in T4L was replaced with its hydroxy isostere lactic acid (Fig. 11). Ala82 is at a break between two helices with the NH group and side chain exposed to water. The Ala82 → lactate mutation effectively substitutes a good hydrogen-bond acceptor, the amide carbonyl group, with

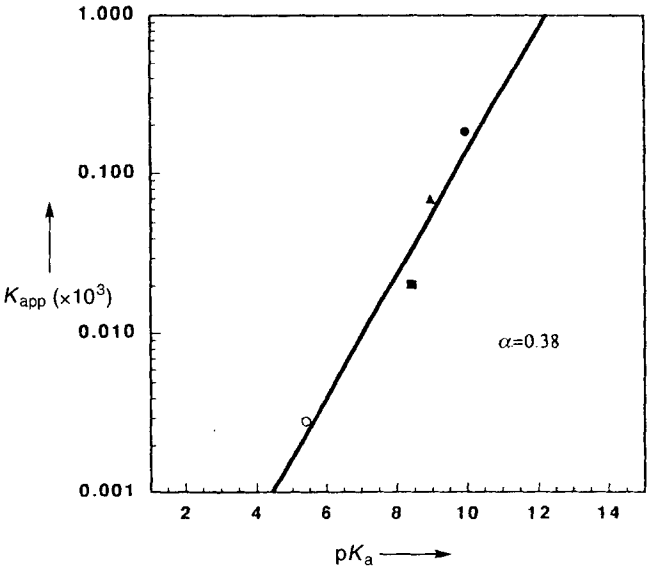
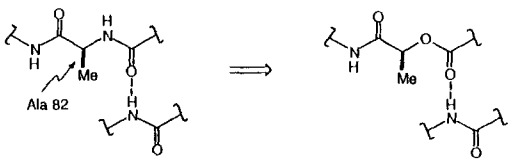


Fig. 2. The relation between the pK_a of the residue at position 27 (Tyr **10** and Tyr analogues **5**–**7**) and the stability (K_{app}) of the SNase. ● = **10**, ▲ = **5**, ■ = **6**, ○ = **7**.

a considerably weaker hydrogen-bond acceptor, the ester carbonyl group.[98] Because both esters and amides exist mainly in the *trans* conformation,[99] the destabilization caused by this substitution (decrease in the melting temperature T_m by 3.7 °C (1.0 kcal mol^{-1})) provides evidence that backbone amide hydrogen-bonding interactions do in fact help to stabilize proteins.



Scheme 11. Substitution of Ala82 in T4L with lactic acid.

In an examination of the importance of packing interactions in the core of a protein,[100–103] replacements were made for Leu133 which lies along the edge of the largest cavity in the interior of T4L (Fig. 3).[104] Previous attempts to stabilize T4L by increasing the packing density of the hydrophobic core by means of a Leu133 → Phe or Ala129 → Val mutation had been unsuccessful, as both mutations disrupt neighboring residues.[104] However, incorporation of the unnatural amino acid (*S,S*)-2-amino-4-methylhexanoic acid (**11**) and (*S*)-2-amino-3-cyclopentylpropanoic acid (**12**), both of which were designed to fill the cavity with minimal strain, stabilizes T4L, increasing the protein's melting temperature by 1.9 °C (0.6 kcal mol^{-1}) and 4.3 °C (1.24 kcal mol^{-1}), respectively.[105] This result demonstrates that amino acids that increase the bulk of buried hydrophobic residues without concomitant introduction of strain can increase protein stability significantly. As there is little difference in surface area between **12** and **11** (Δsurface area = 0.9 Å^2), the difference in stability between the two mutants likely reflects the fact that the cyclic amino acid **12** loses less conformational entropy upon folding than does the acyclic amino acid **11**. Consequently, side-chain conformational

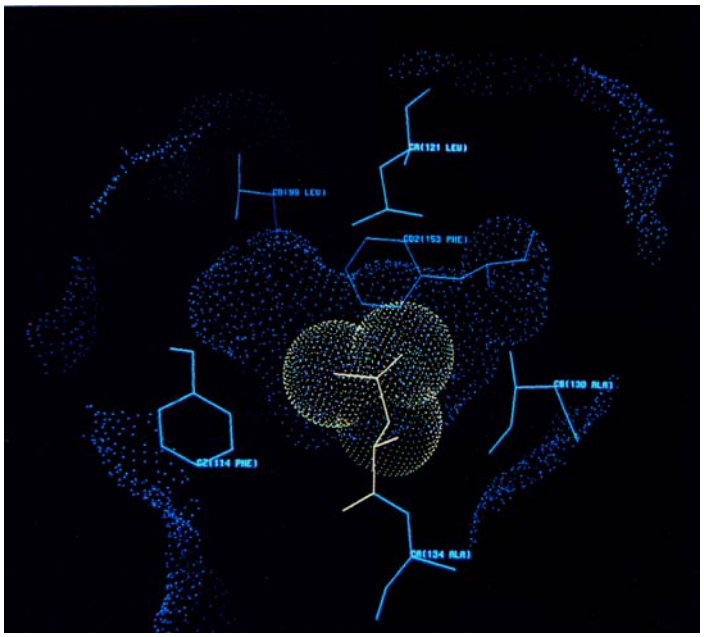
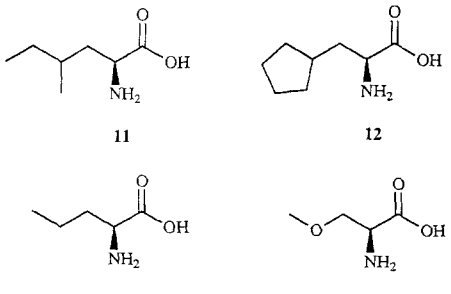


Fig. 3. Graphic representation of the van der Waals surfaces (yellow) of the side chains of Leu133 in T4L.

entropy, as well as packing interactions, plays an important role in determining protein stability.

In a second example, substitution of norvaline (**13**) and *O*-methylserine (**14**) (Scheme 12) for Leu133 allowed for a measurement of the effect of modifying side-chain hydrophobicity in the absence of changes in packing interactions or side-chain entropy.[105] In this case it was found that the difference in the stabilities of the two mutant enzymes relative to that of wild-type T4L corresponds to the difference in the partitioning ratios of the respective *N*-acetyl amide derivatives in octanol–water. This correlation provides further evidence that these partitioning ratios give an accurate measure of solvation effects.



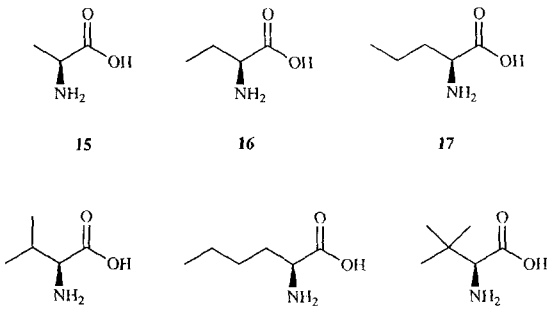
11 **12**

13 **14**

Scheme 12. Structures of the amino acids substituted for Leu113 in T4L.

In another experiment, systematic removal of methyl and methylene groups from the Leu133 side chain was found to lead to a stepwise decrease in protein stability.[105] However, the change was nonlinear; $\Delta\Delta G$ became increasingly large with decreasing side-chain surface area. This trend is reflected accurately in the semiquantitative estimates of the $\Delta\Delta G$ values and results from the net effect of changes in the hydrophobic, entropic, packing, and cavity energy terms.

The effects of incorporating β-branched amino acids were examined by comparing the stabilities of mutant proteins containing the series of branched and linear amino acids **15**–**20** (Scheme 13), at two surface-exposed sites in the middle of two



15 **16** **17**

18 **19** **20**

Scheme 13. β-Branched and linear amino acids incorporated at sites 44 and 68 in T4L.

α-helices in T4L.[106] The natural β-branched amino acids valine, isoleucine, and threonine are thought to destabilize α-helices. In known protein structures these amino acids occur infrequently in α-helices relative to amino acids such as leucine, and they destabilize α-helical peptides and protein α-helices.[107–113] For example, when a series of linear and branched amino acids were incorporated into an α-helical peptide, the β-branched amino acids were found to be destabilizing relative to the linear and γ-branched amino acids.[114] Substitution of *tert*-butylglycine (**20**) for *n*-butylglycine (**19**) or alanine (**15**) destabilizes the peptide by 0.9 kcal mol^{-1}; however, the same substitution either destabilizes T4L by 0.7 kcal mol^{-1} ($\Delta T_m = 2.6\,^{\circ}C$) (site 44) or stabilizes the protein by 0.3 kcal mol^{-1} ($\Delta T_m = 1.0\,^{\circ}C$) (site 68). This difference illustrates the difficulty in delineating simple rules as to which factors stabilize or destabilize proteins given the influence of context effects. In addition, the stabilities and simulated structures of the *tert*-butylglycine mutants have provided insight into the effects of β-branched side chains on α-helix stability in proteins.

The results described underscore the difficulty in interpreting mutagenesis data when more than one property of an amino acid is altered at once. The ability to precisely alter the structure of an amino acid should allow us to better dissect the individual contributions of hydrophobicity, packing, entropy, and cavity formation on protein stability. We are currently extending these studies to hydrogen bonding, and π–π and π–ion interactions in protein interiors.

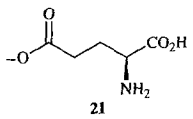
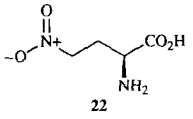
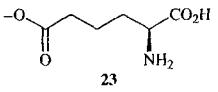
3.4. Enzyme Mechanisms

Mutagenesis with unnatural amino acids also is being used to probe the catalytic mechanisms of the enzymes staphylococcal nuclease, aspartate aminotransferase, methionine aminopeptidase, and ribonucleotide reductase. The enzyme staphylococcal nuclease (SNase) accelerates the hydrolysis of phosphodiester bonds in nucleic acids some 10^{16}-fold over the rate of the uncatalyzed reaction. This enzyme has been the subject of many struc-

tural, mechanistic, and mutagenesis studies aimed at understanding how enzymes can achieve such extraordinary rate enhancements.[115-119] Based on these studies it has been suggested that general base catalysis contributes significantly to the catalytic efficiency of this enzyme. Specifically, it is thought that Glu43 in SNase acts as a general base to activate a water molecule for attack on the phosphodiester backbone of DNA. Glu43 is known to be important for catalysis, because replacement by the natural amino acids Asp and Gln results in a significant loss in activity.[120]

Surprisingly, substitution of Glu43 with either homoglutamate **23** or the nitro analogue (*S*)-4-nitro-2-aminobutyric acid (**22**), which is both isoelectronic and isosteric to glutamate but a much poorer base, yielded mutant enzymes with kinetic constants similar to those of wild-type SNase under normal assay conditions (Table 1).[85] The pH dependence of the catalytic

Table 1. Catalytic efficiences of wild-type staphylococcal nuclease and mutants Glu43 → (*S*)-4-nitro-2-aminobutyric acid and Glu43 → homoglutamate. V_{max} = maximum reaction rate (measured based on the absorption at 260 nm, A_{260}), K_M = Michaelis constant.

Amino acid	V_{max} ($\mu g^{-1} min^{-1}$)	K_M(DNA) ($\mu g mL^{-1}$)	$K_M(Ca^{2+})$ (μM)
 **21**	6.7±0.7	8±3	320±30
 **22**	3.0±0.5	26±8	470±40
 **23**	5.2±0.2	10±2	290±30

efficiencies of these mutants (the pH profiles of the mutants with nitroglutamate **22** and homoglutamate **23** suggest that the nitronate anion is not formed) and the crystal structure of the homoglutamate43 mutant (Fig. 4) suggest that Glu43 may not be acting as a base, but rather may play a structural role, serving as a bidendate hydrogen-bond acceptor to fix the conformation of the neighboring loop. Independent studies by Gerlt et al., in which the loop adjacent to Glu43 was deleted from SNase, have led to the same conclusion.[121]

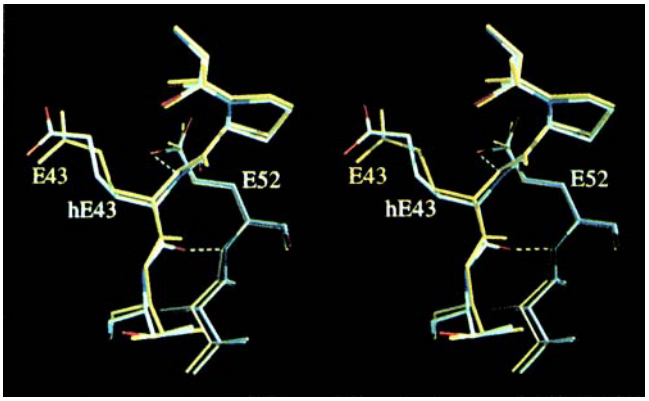


Fig. 4. Comparison of the positions of the side chains of Glu43 and homoglutamate43 as determined by X-ray crystallography for the apo wild-type and homoglutamate43 mutant proteins, respectively (stereoplot).

3.5. Cellular Signal Transduction

Mutagenesis with unnatural amino acids has been used to probe the role of *ras* p21 in cellular signal transduction pathways. Mammalian proteins encoded by the *ras* genes are thought to function as regulators of various signal transduction processes involved in cell growth and differentiation.[122-125] The chemical basis for signal regulation involves cycling of the protein between the inactive guanosine diphosphate (GDP) bound state and the active guanosine triphosphate (GTP) bound state (Fig. 5). Point mutations that result in a decrease in

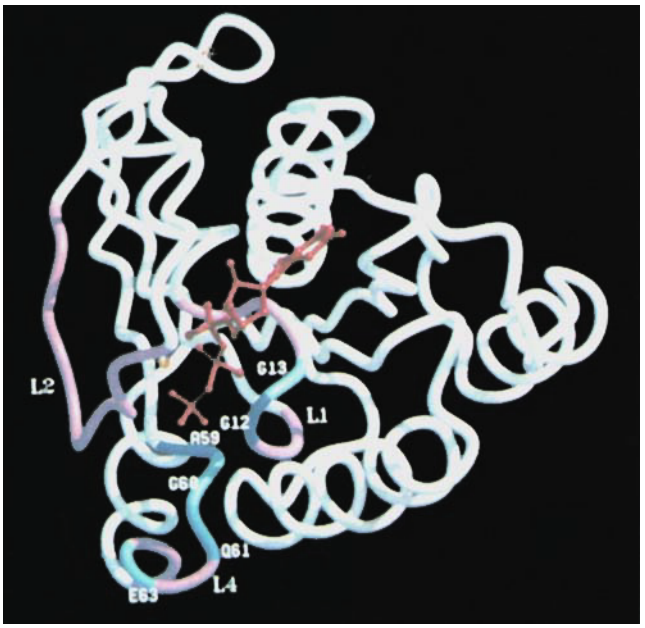
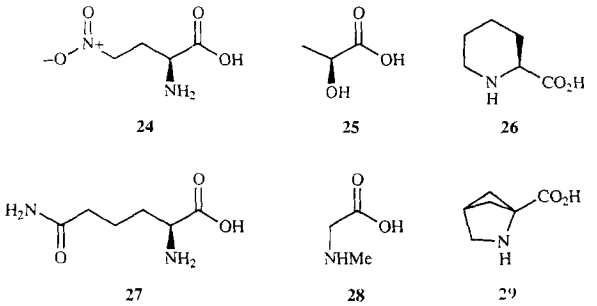


Fig. 5. Model of *ras* p21 with bound GTP [138].

the intrinsic GTPase activity of *ras* p21 or of the GTPase activity stimulated by GTPase activating protein (GAP) are associated with approximately 30% of human cancers.[123, 126-129] In order to better understand the molecular basis by which mutations in *ras* p21 lead to switch inactivation, we have substituted residues in loop L4 (the switch II region), loop L2 (the switch I region), and loop L1 (the phosphate binding loop) with a series of unnatural amino acids.[130-132]

Mutations at Gln61 are among the most common that lead to impaired intrinsic GTPase activity. This residue lies in loop L4 of *ras* p21, a region that undergoes conformational change upon GTP/GDP exchange.[133-140] On the basis of biochemical studies and the three-dimensional X-ray crystal structures of the GTP-, Gpp(CH_2)p-, and Gpp(NH)p-bound forms of *ras* p21, it has been proposed that Gln61 is critical for γ-phosphate binding

and catalysis.[133-141] Specifically, it has been proposed that the γ-carboxamide of Gln 61 either polarizes a water molecule (175) for attack on the γ-phosphate or stabilizes the incipient pentacoordinate transition state. In order to test these mechanistic hypotheses, we replaced Gln 61 with the isoelectronic, isosteric nitro analogue **24** and with homoglutamine **27** (Scheme 14).[130]



Scheme 14. Structures of some unnatural amino acids used to probe the switch function of *ras* p21.

Whereas seventeen natural mutants at position 61 have reduced GTPase activity and are not activated by GAP, both unnatural mutants had GTPase activity similar to that of the wild-type protein and were activated by GAP. The fact that these mutants retained activity suggests that residue 61 plays some other role in catalysis, perhaps serving to distort the bound GTP toward the transition state geometry.

Mutations at Gly 12 of *ras* p21 also result in impaired intrinsic GTPase activity and are commonly associated with oncogenic activation.[123, 142] Gly 12 occurs in a highly conserved type II β-turn, [Gly 11-X-X-X-Gly-Lys-(Ser or Thr)], a phosphate-binding loop found in many nucleotide-binding proteins.[143, 144] Mutation of Gly 12 to any common amino acid other than proline results in diminished GTPase activity.[142] To more fully understand the role of Gly 12 mutants in switch function, we inserted a number of unnatural amino acids and analogues at this site, including lactic acid (**25**, Lac), pipecolic acid (**26**, Pip), and *N*-methylglycine (**28**, MeGly) (Scheme 14).[130] The protein containing pipecolic acid (**26**) at position 12, which was expected to have a much more negative torsion angle ϕ than Gly 12, displayed GTPase activity similar to that of wild-type ras. The *N*-methylglycine (**28**) mutant also had comparable activity (N^{α}-methyl substitution also decreases ϕ values). Whereas the Ala 12 mutant had reduced GTPase activity, the variant with isosteric lactic acid (**25**), in which the backbone amide linkage is replaced by an ester linkage, had normal GTPase activity. The fact that only mutants that can adopt unusual backbone conformations are active suggests that a conformational change in loop L1 may relieve unfavorable side-chain interactions at position 12 in the transition state. However, even though the Pro 12, Pip 12, Lac 12, and MeGly 12 variants had intrinsic GTPase activity similar to that of wild-type *ras* p21, they were not activated by GAP. Moreover, the Gly 13 → Thr, Gly 13 → allothreonine, and Gly 13 → Ser proteins, which have intrinsic GTPase activity two to three times that of wild-type *ras* p21 (attributable to the β-hydroxyl group), were not activated by GAP and were not transforming in a germinal vesicle breakdown assay with *Xenopus laevis* oocytes. One interpretation of these results is that loop L1 can adopt two or more conformations in solution and that its positioning by one or more effector proteins modulates GTPase activity and onocogenic activation. The unusual backbone structures of the Gly 12 and Gly 13 mutants may prevent switch function.

In a third mutagenesis experiment the role of Pro 34 in loop L2 function was probed. The most significant difference between the GTP- and GDP-bound forms of *ras* p21 is in this region (residues 32–40).[133-140] It has been proposed that Pro 34, which is conserved in *ras* p21 and close to the active site, may play a role in controlling the conformation of loop L2, perhaps through a *cis–trans* isomerization of the Pro 34 amide bond.[137] In order to examine the structural and mechanistic role of this residue more precisely, Pro 34 was replaced with 2,4-methanoproline (**29**) (Scheme 14), which is strongly biased toward the *trans* configuration by virtue of the C_{α} substitution.[131] The fact that this mutant has intrinsic and GAP-activated GTPase activity comparable to that of the wild-type protein strongly suggests that the *cis–trans* isomerization of Pro 34 does not play a key role in signal transduction. Similar substitutions may be useful in probing the role of backbone isomerization in protein folding pathways.

4. Conclusion

The ability to incorporate unnatural amino acids into proteins site-specifically makes possible detailed physical studies on this important class of organic compounds. Mutations can be made to probe the precise nature of the contribution of an individual amino acid to protein structure or function, often without dramatic losses in activity. This methodology should provide new insights into and perhaps challenge assumptions about the nature of protein structure and function. In addition, it may be possible to generate mutant proteins with novel functions. The remaining goals are increasing the quantities of protein obtainable and finding ways to bypass the specificity of the protein biosynthetic machinery so that more diverse amino acid structures can be incorporated to further add to the power of this approach.

We are grateful for financial support for this work from the U. S. Office of Naval Research, National Institutes of Health, and the Department of Energy. V. W. C thanks the National Science Foundation for a Predoctoral Fellowship.

Received: April 6, 1994 [A 61 IE]
German version: *Angew. Chem.* **1995**, *107*, 677

[1] M. J. Zoller, M. Smith, *Methods Enzymol.* **1983**, *100*, 468.
[2] J. R. Knowles, *Science* **1987**, *236*, 1252.
[3] D. Y. Jackson, J. Burnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, *266*, 243.
[4] E. Bayer, *Angew. Chem.* **1991**, *103*, 117; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 113.
[5] T. Nakatsuka, T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1987**, *109*, 3808.
[6] F. M. Richards, H. W. Wyckoff in *The Enzymes, Vol. IV* (Ed.: P. D. Boyer), Academic Press, New York, **1971**.
[7] K. Hofmann, H. Bohn, *J. Am. Chem. Soc.* **1966**, *88*, 5914.
[8] E. T. Kaiser, D. S. Lawrence, *Science* **1984**, *226*, 505.
[9] E. T. Kaiser, *Acc. Chem. Res.* **1989**, *22*, 47.
[10] K. Rose, *J. Am. Chem. Soc.* **1994**, *116*, 30.

[11] M. Schnolzer, S. B. H. Kent, *Science* **1992**, *256*, 221; P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *ibid.*, **1994**, *266*, 776.
[12] R. E. Offord, *Protein Eng.* **1987**, *1*, 151.
[13] I. M. Chaiken, *Crit. Rev. Biochem.* **1981**, *11*, 255.
[14] E. T. Kaiser, D. S. Lawrence, S. E. Rokita, *Annu. Rev. Biochem.* **1985**, *54*, 565.
[15] D. S. Wuttke, H. B. Gray, S. L. Fisher, B. Imperiali, *J. Am. Chem. Soc.* **1993**, *115*, 8455.
[16] K. E. Neet, A. Nanci, D. E. Koshland, *J. Biol. Chem.* **1968**, *243*, 6392.
[17] L. Polgar, M. L. Bender, *J. Am. Chem. Soc.* **1966**, *88*, 3153.
[18] *Spin Labeling, Theory and Applications* (Ed.: L. Berliner), Academic Press, New York, **1976**.
[19] S. J. Pollack, G. Nakayama, P. G. Schultz, *Science* **1988**, *242*, 1038.
[20] D. R. Corey, P. G. Schultz, *Science* **1987**, *233*, 1401.
[21] U. C. Krieg, P. Walter, A. E. Johnson, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8604.
[22] J. Brunner, *Annu. Rev. Biochem.* **1993**, *62*, 483.
[23] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182.
[24] J. A. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren, P. G. Schultz, *Methods Enzymol.* **1992**, *202*, 301.
[25] J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Diala, *J. Am. Chem. Soc.* **1989**, *111*, 8013.
[26] F. H. C. Crick, L. Barrett, S. Brenner, R. Watts-Tobin, *Nature (London)* **1961**, *192*, 1227.
[27] A. Brenner, A. O. W. Stretton, S. Kaplan, *Nature (London)* **1965**, *206*, 994.
[28] L. Bossi, J. R. Roth, *Nature (London)* **1980**, *286*, 123.
[29] F. H. C. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138.
[30] F. Chapeville, F. Lipmann, G. von Ehrenstein, B. Weisblum, W. J. Ray, S. Benzer, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1086.
[31] S. Chladek, M. Sprinzl, *Angew. Chem.* **1985**, *97*, 377; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 371.
[32] S. M. Hecht, *Acc. Chem. Res.* **1992**, *25*, 545.
[33] J. R. Sayers, W. Schmidt, F. Eckstein, *Nucleic Acids Res.* **1988**, *16*, 791.
[34] A. Bruce, J. Atkins, N. Wills, O. Uhlenbeck, R. Gesteland, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 7127.
[35] D. Ayer, M. Yarus, *Science* **1986**, *231*, 393.
[36] L. Bossi, *J. Mol. Biol.* **1983**, *164*, 73.
[37] A. Miller, A. Albertini, *J. Mol. Biol.* **1983**, *164*, 59.
[38] Y. Kwok, J. T. Wong, *Can. J. Biochem.* **1980**, *58*, 213.
[39] J. W. Roberts, J. Carbon, *J. Biol. Chem.* **1975**, *250*, 5530.
[40] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783.
[41] C. J. Noren, S. J. Anthony-Cahill, D. J. Suich, K. A. Noren, M. C. Griffith, P. G. Schultz, *Nucleic Acids Res.* **1990**, *18*, 83.
[42] K. M. Harrington, I. A. Nazarenko, D. B. Dix, R. C. Thompson, O. C. Uhlenbeck, *Biochemistry* **1993**, *32*, 7617.
[43] R. Giege, J. D. Puglisi, C. Florentz, *Progress Nucleic Acids Res. Mol. Biol.* **1993**, *45*, 129.
[44] P. Schimmel, *Annu. Rev. Biochem.* **1987**, *56*, 125.
[45] T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade, S. M. Hecht, *Biochemistry* **1984**, *23*, 1468.
[46] S. A. Robertson, C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Nucleic Acids Res.* **1989**, *17*, 9649.
[47] D. Hentzen, P. Mandel, J. P. Garel, *Biochim. Biophys. Acta* **1972**, *281*, 228.
[48] M. C. Griffith, Ph.D. Thesis, University of California, Berkeley, **1990**.
[49] S. A. Robertson, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **1991**, *113*, 2722.
[50] A. Patchornik, B. Amit, R. B. Woodward, *J. Am. Chem. Soc.* **1970**, *92*, 6333.
[51] S. R. Adams, J. P. Y. Kao, R. Y. Tsien, *J. Am. Chem. Soc.* **1989**, *111*, 7957.
[52] D. Kern, R. Giege, J. P. Ebel, *Eur. J. Biochem.* **1972**, *31*, 148.
[53] N. Chinsky, A. L. Margolin, A. M. Klibanov, *J. Am. Chem. Soc.* **1989**, *111*, 386.
[54] P. Kast, H. Hennecke, *J. Mol. Biol.* **1991**, *222*, 99.
[55] *Transcription and Translation: A Practical Approach* (Eds.: B. D. Hames, S. J. Higgins), IRL Press, Oxford, **1984**.
[56] J. Pratt in ref. [55], p. 179.
[57] G. Zubay, *Annu. Rev. Genet.* **1973**, *7*, 267.
[58] D. V. Goeddel (Ed.:), *Methods Enzymol.* **1990**, *185*.
[59] *Maximizing Gene Expression* (Eds.: W. Rezenikoff, L. Gold), Buttersworth, Boston, **1986**.
[60] D. Suich, Ph.D. Thesis, University of California, Berkeley, **1993**.
[61] D. Mendel, J. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **1993**, *115*, 4359.
[62] V. W. Cornish, D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2910.
[63] A. Louie, F. Jurnak, *Biochemistry* **1985**, *24*, 6433.
[64] K. Quiggle, G. Kumar, T. W. Oh, E. K. Ryu, S. Chladek, *Biochemistry* **1981**, *20*, 3480.
[65] A. Bhuta, *Biochemistry* **1981**, *20*, 8.
[66] P. Bhuta, *Biochemistry* **1982**, *21*, 899.
[67] S. Fahnestock, H. Neumann, V. Shashoua, A. Rich, *Biochemistry* **1970**, *9*, 2477.
[68] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197.
[69] M. M. Grant, A. S. Brown, L. M. Corwin, R. F. Troxler, C. Franzblau, *Biochim. Biophys. Acta* **1975**, *404*, 180.
[70] D. A. Tirrell, T. J. Deming, personal communication.
[71] D. Mendel, J. A. Ellman, P. G. Schultz, unpublished.
[72] J. D. Bain, D. A. Wacker, E. E. Kuo, A. R. Chamberlin, *Tetrahedron* **1991**, *47*, 2389.
[73] D. Benson, J. A. Ellman, P. G. Schultz, unpublished.
[74] G. Baldwin, A. Martoglio, C. Zugliani, J. Brunner, *Biochemistry* **1988**, *27*, 7951.
[75] A. S. Spirin, V. I. Baranov, L. A. Ryabova, S. Y. Ovodov, Y. B. Alahhov, *Science* **1988**, *242*, 1162.
[76] O. Landman, S. Spiegelman, *Proc. Natl. Acad. Sci.* **1955**, *41*, 698.
[77] M. Nomura, J. Hosoda, S. Nishimura, *Biochim. Biophys. Acta* **1958**, *29*, 161.
[78] L. G. Kleina, J. M. Masson, J. Normanly, J. Abelson, J. H. Miller, *J. Mol. Biol.* **1990**, *213*, 705.
[79] V. W. Cornish, P. G. Schultz, unpublished.
[80] D. Mendel, P. G. Schultz, unpublished.
[81] M. Chunhua, W. Kudlicki, O. W. Odom, G. Kramer, B. Hardesty, *Biochemistry* **1993**, *32*, 7939.
[82] J. Bain, C. Switzer, A. Chamberlin, S. Benner, *Nature* **1992**, *356*, 537.
[83] J. A. Ellman, D. R. Benson, P. G. Schultz, unpublished.
[84] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schultz, D. E. Wemmer, *J. Am. Chem. Soc.* **1992**, *114*, 7959.
[85] J. K. Judice, T. R. Gamble, E. C. Murphy, A. M. de Vos, P. G. Schultz, *Science* **1993**, *261*, 1578.
[86] B. L. Stoddard, J. L. Bruhnke, D. Koenigs, N. Porter, D. Ringe, G. A. Petsko, *Biochemistry* **1990**, *29*, 8042.
[87] M. Wilcox, R. W. Viola, K. W. Johnson, A. P. Billington, B. C. Carpenter, J. A. McCray, A. P. Guzikowski, G. P. Hess, *J. Org. Chem.* **1990**, *55*, 1585.
[88] D. Mendel, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **1991**, *113*, 2758.
[89] X. Xiong, S. N. Cook, W. E. Jack, L. E. Danley, J. A. Ellman, P. G. Schultz, C. J. Noren, *Cell*, submitted.
[90] K. A. Dill, *Biochemistry* **1990**, *29*, 7133.
[91] B. W. Matthews, *Annu. Rev. Biochem.* **1993**, *62*, 139.
[92] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci., USA* **1951**, *37*, 729.
[93] L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci., USA* **1951**, *37*, 205.
[94] G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Molecules*, Springer, New York, **1991**.
[95] G. D. Rose, R. Wolfenden, *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381.
[96] A. R. Fersht, *Trends Biochem. Sci.* **1987**, *12*, 301.
[97] J. Thorson, E. Chapman, E. C. Murphy, P. G. Schultz, J. K. Judice, *J. Am. Chem. Soc.* **1995**, *117*, 1157.
[98] M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, *J. Org. Chem.* **1983**, *48*, 2877.
[99] K. B. Wiberg, K. E. Laidig, *J. Am. Chem. Soc.* **1987**, *109*, 5935.
[100] K. Yutani, K. Ogasahara, T. Tsugita, Y. Sugino, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4441.
[101] M. Matsumura, W. J. Becktel, B. W. Matthews, *Nature (London)* **1988**, *334*, 406.
[102] J. J. T. Kellis, K. Nyberg, D. Sali, A. R. Fersht, *Nature (London)* **1988**, *333*, 784.
[103] W. A. Lim, R. T. Sauer, *Nature (London)* **1989**, *339*, 31.
[104] M. Karpusas, W. A. Baase, M. Matsumura, B. W. Matthews, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8237.
[105] D. Mendel, J. A. Ellman, Z. Chang, D. L. Veenstra, P. A. Kollman, P. G. Schultz, *Science* **1992**, *256*, 1798.
[106] V. W. Cornish, M. I. Kaplan, D. Veenstra, P. A. Kollman, P. G. Schultz, *Biochemistry* **1994**, *33*, 12022.
[107] P. Y. Chou, G. D. Fasman, *Annu. Rev. Biochem.* **1978**, *47*, 251.
[108] K. T. O'Neil, W. F. DeGrado, *Science* **1990**, *250*, 646.
[109] S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, R. L. Baldwin, *Nature (London)* **1990**, *344*, 268.
[110] P. C. Lyu, M. I. Liff, L. A. Marky, N. R. Kallenbach, *Science* **1990**, *250*, 669.
[111] M. Blaber, X.-J. Zhang, B. W. Matthews, *Science* **1993**, *260*, 1637.
[112] M. Blaber, X.-J. Zhang, J. D. Lindstrom, S. D. Deprot, W. A. Baase, B. W. Matthews, *J. Mol. Biol.* **1994**, *235*, 600.
[113] A. Horowitz, J. M. Matthews, A. R. Fersht, *J. Mol. Biol.* **1992**, *227*, 560.
[114] P. C. Lyu, J. C. Sherman, A. Chen, N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5317.
[115] F. A. Cotton, E. E. J. Hazen, M. J. Legg, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 2551.
[116] P. J. Loll, E. E. Lattman, *Proteins* **1989**, *5*, 183.
[117] D. J. Weber, A. K. Meeker, A. S. Mildvan, *Biochemistry* **1991**, *30*, 6103.
[118] D. J. Weber, E. H. Serpersu, D. Shortle, A. S. Mildvan, *Biochemistry* **1990**, *29*, 8632.
[119] P. W. Tucker, J. E. E. Hazen, F. A. Cotton, *Mol. Cell. Biochem.* **1979**, *23*, 3.

[120] D. W. Hibler, N. J. Stolowich, M. A. Reynolds, J. A. Gerlt, J. A. Wilde, P. H. Bolton, *Biochemistry* **1987**, *26*, 6278.
[121] S. P. Hale, L. B. Poole, J. A. Gerlt, *Biochemistry* **1993**, *32*, 7479.
[122] M. Barbacid, *Annu. Rev. Biochem.* **1987**, *56*, 779.
[123] J. L. Bos, *Cancer Res.* **1989**, *49*, 4682.
[124] J. A. Glomset, M. H. Gelb, C. C. Fransworth, *Trends Biochem. Sci.* **1990**, *15*, 139.
[125] D. Spandidos, *Ras Oncogenes*, Plenum, New York, **1989**.
[126] M. Trahey, M. Trahey, G. Wong, R. Halenbeck, B. Rubinfield, G. A. Martin, M. Ladner, C. M. Long, W. J. Crosier, K. Watt, K. Koths, F. McCormick, *Science* **1988**, *242*, 1697.
[127] M. Trahey, F. McCormick, *Science* **1987**, *238*, 542.
[128] G. A. Martin, D. Viskochil, G. Bollag, P. C. McCabe, W. J. Crosier, H. Haubruck, L. Conroy, R. Clark, P. O'Connell, R. M. Cawthon, M. A. Innis, F. McCormick, *Cell* **1990**, *63*, 843.
[129] G. Xu, B. Lin, K. Tanaka, D. Dunn, D. Wood, R. Gesteland, R. White, R. Weiss, F. Tamanoi, *Cell* **1990**, *63*, 835.
[130] H. H. Chung, D. R. Benson, P. G. Schultz, *Science* **1993**, *259*, 806.
[131] H. H. Chung, D. R. Benson, V. W. Cornish, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10145.
[132] H. H. Chung, D. R. Benson, P. G. Schultz, *J. Am. Chem. Soc.* **1993**, *115*, 6414.
[133] A. M. d. Vos, L. Tong, M. Milgurn, P. Matias, J. Jancarik, S. Noguchi, S. Nishimura, K. Miura, E. Ohtsuka, S.-H. Kim, *Science* **1988**, *239*, 888.
[134] E. F. Pai, W. Kabsch, U. Krengel, K. C. Homes, J. John, A. Wittinghofer, *Nature (London)* **1989**, *341*, 209.
[135] L. Tong, M. V. Milburn, A. M. d. Vos, S.-H. Kim, *Science* **1989**, *245*, 244.
[136] M. V. Milburn, L. Tong, A. M. deVos, A. Brunger, Z. Yamaizumi, S. Nishimura, S.-H. Kim, *Science* **1990**, *247*, 939.
[137] E. F. Pai, U. Krengel, G. A. Petsko, R. S. Goody, W. Kabsch, A. Wittinghofer, *EMBO J.* **1990**, *9*, 2351.
[138] G. G. Prive, M. V. Milburn, L. Tong, A. M. deVos, Z. Yamaizumi, S. Nishimura, S.-H. Kim, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3649.
[139] U. Krengel, I. Schilchting, A. Scherer, R. Schumann, M. Frech, J. John, W. Kabsch, E. F. Pal, A. Wittinghofer, *Cell* **1990**, *62*, 539.
[140] I. Schlichting, S. C. Almo, G. Rapp, K. Wilson, A. Petratos, A. Lentfer, A. Wittinghofer, W. Kabsch, E. F. Pai, G. A. Petsko, R. S. Goody, *Nature (London)* **1990**, *345*, 309.
[141] C. J. Der, T. Finkel, G. M. Cooper, *Cell* **1986**, *44*, 167.
[142] P. H. Seeburg, W. W. Colby, D. J. Capon, D. V. Goeddel, A. D. Levison, *Nature (London)* **1986**, *312*, 71.
[143] N. J. Gay, J. E. Walker, *Nature (London)* **1983**, *301*, 262.
[144] R. K. Wierenga, P. Terpstora, W. G. Hol, *J. Mol. Biol.* **1986**, *187*, 101.