Biochemistry

 Volume 29, Number 41 October 16, 1990

Perspectives in Biochemistry

Hazards and Their Exploitation in the Applications of Molecular Biology to Structure–Function Relationships†

Paul Schimmel

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

Received May 24, 1990; Revised Manuscript Received June 21, 1990

The applications of molecular biology to enzymology and to protein and nucleic acid structure–function relationships are not always straightforward. Early work on structure–function relationships of proteins and nucleic acids utilized chemical modifications and affinity labeling to assign functions to specific amino acids or nucleotide bases and used limited proteolysis or nuclease digestions to generate fragments that encoded structural domains. Typically, manipulations that yielded altered nucleic acids or proteins, and analyses of functional consequences of alterations, were done in vitro. Recombinant DNA technology facilitates any desired nucleotide or amino acid substitution and, in principle, production of any desired fragment. However, the production of altered molecules and at least some of their functional analyses requires in vivo systems. Here a large set of incompletely characterized or unknown variables can lead to unintended and unsuspected changes in structures and activities.

These complications include novel associations between host-cell-encoded and plasmid-encoded molecules, recombination between plasmid and host genome sequences that generate sequence variants of the plasmid-encoded molecules, idiosyncratic effects on activity of sequences that are exposed in the creation of protein or nucleic acid fragments, and translational misreading of codon replacements. Failure to analyze for these and other possibilities in the investigation of products that are expressed from recombinant molecules can lead to serious errors of interpretation. At the same time, some of these phenomena are now sufficiently well understood that they can be exploited to advance further the understanding of structure–function–activity relationships.

Significance of Genetic Backgrounds for Protein Analysis

The Potential for Novel Interactions. Proteins or nucleic acids that are subjected to mutagenesis as part of investigations of their structure–function relationships are typically expressed from a multicopy plasmid in a suitable host organism such as *Escherichia coli* or the yeast *Saccharomyces cerevisiae*. Functional activity in vivo usually involves a test for complementation of a genetic defect in the host cell genome. The failure of a plasmid-encoded point or deletion mutant to complement is prima facie evidence that the altered region in the mutant species is important for activity (although further studies would have to show that the mutation did not cause a defect in folding and/or result in rapid degradation). Conversely, complementation of the growth defect by a recombinant deletion fragment would suggest that the deleted residues are dispensable for activity and that the residual fragment encodes the active site.

In vivo complementation tests are the most rapid way to screen large numbers of variants so as to establish a functional road map of the molecule. Such analyses are generally followed by detailed biochemical studies of specific mutants and fragments that are expressed to high level from recombinant plasmids. The simple interpretation of the complementation of a growth defect by a plasmid-encoded point or deletion mutant is that the mutant protein retains the essential catalytic residues (Jasin et al., 1983).

However, this interpretation can be wrong. The reason has to do with the mutant allele on the chromosome that is used for the complementation analysis and the possibility that any complementation phenotype can be specific to that allele (Jasin et al., 1984). Most commonly, the genetic background consists of a point or deletion mutation in the chromosome that results in production of an inactive host cell protein or protein fragment. The disabled protein is implicitly assumed to be biochemically silent, but there are many examples that demonstrate otherwise. Because in vitro assays of activity are often done in extracts that contain the overexpressed protein, the artifacts that occur in vivo are recapitulated in vitro and thereby reinforce their credibility.

Early genetic and biochemical analyses of defined systems revealed novel interactions that result in biological activity as

†Supported by Grants GM15539 and GM23562 from the National Institutes of Health.

0006-2960/90/0429-9495$02.50/0

a result of associations between molecules that individually are inactive. More sophisticated recent applications of cloning and mutagenesis have extended considerably the understanding of these protein complexes. Novel complexes can form in vivo and in vitro and thus can obscure the interpretation of activity that results from expression of a recombinant protein in a host cell that encodes a defective point or deletion mutant of the same protein. Because some of the complexes that give rise to activity are specific to particular mutants or fragments, the delineation of a catalytic region in a protein or RNA can be confounded if it is based on comparing apparent activities of members of a set of recombinant products.

Example of a Novel Protein Complex That Has Potentially Misleading Biological Activity in Vivo and in Vitro. One example is the complementation of chromosomal point mutants of *E. coli* alanine tRNA synthetase by plasmid-encoded pieces of the enzyme that have no activity whatsoever. This enzyme is an α_4 tetramer of identical polypeptides of 875 amino acids (Putney et al., 1981a,b; Schimmel, 1990). It catalyzes the two-step aminoacylation reaction:

$$\text{enzyme} + \text{alanine} + \text{ATP} \rightarrow \text{enzyme·alanyl-AMP} + \text{PP}_i$$

$$\text{enzyme·alanyl-AMP} + \text{tRNA} \rightarrow \text{alanyl-tRNA} + \text{AMP} + \text{enzyme}$$

In the first reaction a tightly bound aminoacyl adenylate is formed and in the second reaction the adenylate reacts with the 3′-end of the tRNA to form the aminoacyl linkage. The adenylate synthesis reaction can be assayed independently of the overall aminoacylation reaction, thus providing two convenient measures of enzyme activity. An amino-terminal fragment of 368 amino acids is a monomer and has the full adenylate synthesis activity of the native enzyme, but does not charge tRNA (Putney et al., 1981a; Regan et al., 1987). A monomeric fragment of 461 amino-terminal residues (fragment 461N) can charge tRNA, although the catalytic activity is reduced. Thus, although the native enzyme is a tetramer, monomeric fragments have sites for each of the substrates and have catalytic activity.

The chromosomal mutant alleles Gly674 → Ala (G674A) and Gly677 → Ala (G677A) confer a temperature-sensitive phenotype that prevents growth at 42 °C (Theall et al., 1977; Jasin et al., 1985). The activity of these mutant proteins is severely reduced at the permissive temperature of 30 °C, although it is sufficient to sustain cell growth. At 42 °C the reduced activity cannot support cell division. As expected, introduction of plasmids that encode fragment 461N results in complementation of the growth defect (Jasin et al., 1983; Ho et al., 1985).

In an extension of these experiments, complementation of the mutant alleles was observed with recombinant plasmids that encode large internal deletions in the essential amino-terminal catalytic domain. Thus, while they complement the mutant alleles, these proteins are completely devoid of catalytic activity. The region within the inactive proteins that is required for complementation was subsequently mapped to a resolution of approximately 100 amino acids (Jasin et al., 1984).

The explanation for the complementation paradox is that the inactive deletion proteins bind to the G674A and G677A mutant polypeptides and activate their catalytic sites. The deletion proteins that complement have in common a domain that extends from Glu700 to Glu808. This segment is required for subunit interactions in the native (tetrameric) protein (Jasin et al., 1983). Activation of the G677A mutant polypeptide by the addition of an inactive deletion fragment that encodes Glu700–Glu808 was demonstrated in vitro to occur through formation of heterooligomers. The binding of the inactive deletion fragments to the mutant polypeptides is believed to elevate the defective k_{cat} of the G674A and G677A proteins, with little effect on substrate K_m parameters (Jasin et al., 1984, 1985).

The detailed explanation for this effect will only emerge when a high-resolution structure is available. Had not the catalytic domain of alanine tRNA synthetase previously been mapped by deletion analysis, the complementation results could have incorrectly designated those inactive pieces that complement the G674A and G677A mutations as ones that encode the catalytic site. At the same time, the ambiguity could have been eliminated if a genetic background that avoided the possibility of formation of heterooligomers was used to test the activity of the deletion mutants.

Additional Novel Complexes. While the above example with alanine tRNA synthetase is unusual in that pieces that are not essential for activity can activate the catalytic site of a mutant enzyme by formation of a complex, at least the potential for novel protein associations was evident in early genetic and chemical modification studies. Thus, Zamenhof and Villarejo (1962) described the in vivo association of two enzymatically inactive fragments of β-galactosidase to generate catalytic activity. Complementation is achieved by association of a small N-terminal α-fragment with a second larger protein that has a deletion in the N-terminal region of the protein [see review by Zabin and Fowler (1980)]. Complementation in vivo and in vitro by formation of heterodimers of different inactive mutant subunits of dimeric alkaline phosphatase was described by Garen and Garen (1963) and Schlesinger and Levinthal (1963), and an association between different mutant forms of *Neurospora crassa* glutamate dehydrogenase was demonstrated by Fincham and Coddington (1963). Jackson and Yanofsky (1969) restored tryptophan synthetase activity in vitro by forming functional dimers of point mutant and deletion fragments of the enzyme's α-chain. Also, detailed physical studies characterized the interaction of the amino-terminal 20 amino acid S-peptide of bovine pancreatic ribonuclease A with the remaining 104 amino acid (residues 21–124) S-protein to generate active monomeric enzyme from the reassociated chains [see Richards and Wyckoff (1971) and Kim and Baldwin (1982)]. Similar studies with staphylococcal nuclease (Anfinsen et al., 1971; Taniuchi et al., 1986), cytochrome b_5 reductase (Strittmatter et al., 1972), human pituitary growth hormone (Li & Bewley, 1976), and thioredoxin (Slaby & Holmgren, 1979; Holmgren & Slaby, 1979) demonstrated that individual protein fragments could combine in vitro to reconstitute catalytic activity.

Unlike alkaline phosphatase, glutamate dehydrogenase, and tryptophan synthetase, which normally are oligomeric proteins with intersubunit interactions, ribonuclease A and several other examples are monomers. In these instances it is clear that intrachain cohesive interactions alone are sufficient to reassemble active enzyme from fragment constituents. In the case of S-peptide, the fragment alone does not constitute a stable structural domain but rather is organized by virtue of integration with the S-protein in a way that recreates the active-site region. Nonetheless, associations between covalently separated, stable structural domains of a monomeric enzyme can also reconstitute enzyme activity. For example, *Salmonella typhimurium* alanine racemase can be split by proteolysis into nonoverlapping 28- and 11-kilodalton stable domains that spontaneously associate to reconstitute enzyme activity (Galakatos & Walsh, 1987). Thus, peptides and domains that are derived from protein monomers can associate to recon-

stitute "holoenzyme" somewhat like the way intersubunit interactions reconstitute a multimeric protein.

Even unsplit monomeric proteins can form oligomeric species when the reacting partners have redundant sequences. An early example was from Crestfield et al. (1963a,b), who reported that carboxymethylation of either His12 (to give 3-carboxymethyl-His12) or His119 (to give 1-carboxymethyl-His119) in pancreatic ribonuclease A resulted in 93% and 100% inactivation, respectively, for hydrolysis of the test substrate cytidine cyclic 2′,3′-phosphate. Crestfield et al. proposed that the chain was folded in three dimensions to bring the two histidines into close proximity at the active site so that modification of either of these essential histidines would inactivate the protein. Subsequently, the three-dimensional structure revealed in detail the organization of these histidines and of other side chains such as Lys41 at the active site (Richards & Wyckoff, 1971). However, the most novel aspect of this chemical modification work was not the technically challenging task of mapping active-site residues. Most significant was the observation that lyophilization of the carboxymethyl-His12 species with the carboxymethyl-His119 variant resulted in recovery of 45% of the specific activity of native ribonuclease A toward cytidine cyclic 2′,3′-phosphate. No activity was obtained upon lyophilization of the carboxymethyl-His119 derivative by itself, and 7–21% activity was observed with the same treatment of the carboxymethyl-His12 species. The authors postulated that partially unfolded chains refolded in a way that enabled heterooligomers to form in which an active site was reconstructed with an unmodified His12 donated by one chain and an unmodified His119 donated by the other. Although the enzyme is normally a monomer, the conditions of the reconstitution were imagined to cause unfolding so that normal intrachain packing interactions were replaced by their interchain counterparts to form artificial dimers and/or higher order oligomers.

This example and the aforementioned examples illustrate that novel associations that may affect the complementation phenotype in vivo of recombinant proteins are not restricted to the intracellular environment. Because novel associations can be recapitulated in vitro, assays of activities of recombinant enzyme variants in extracts can be hazardous. If structure–function relationships of any protein are studied by analysis of mutants, then such mutants ideally should be expressed in a genetic background that is devoid of any allele of the protein other than the one that is expressed from the recombinant plasmid.

Even if the genetic background does not encode any allele of that expressed from a recombinant plasmid, there is still the possibility of an active complex formed with an unrelated molecule. For example, the barely detectable activity of the amino-terminal M15 deletion mutant of β-galactosidase is dramatically elevated by association with an antibody (Accolla & Celada, 1976). Presumably the antibody stabilizes an active conformation and/or serendipitously provides side chains that assist catalysis. In a different example, Harper et al. (1988) tested the effect of the "S-peptide" counterpart of human angiogenin on activation of the S-protein of bovine pancreatic ribonuclease A. While human angiogenin has 35% sequence identity to human pancreatic ribonuclease A and catalyzes limited digestion of 18S and 28S ribosomal RNA, it does not cleave the usual pancreatic RNase substrates including cytidine cyclic 2′,3′-phosphate. An alignment of the bovine pancreatic RNase 20 amino acid S-peptide with its amino-terminal counterpart in angiogenin shows only four amino acid identities, including His12. In spite of the lack of sequence similarity to the bovine enzyme's S-peptide, the angiogenin S-peptide associates with the inactive bovine pancreatic RNase S-protein to generate RNase A like specificity and activity that, for some substrates such as cytidine cyclic 2′,3′-phosphate, approaches 50% of that of native RNase A.

Use of Null Alleles

Formation of heterooligomers and their potentially misleading phenotypes in vivo and in vitro can only be avoided by use of genetic backgrounds that encode null alleles of the gene of interest. A true null allele results from ablation of the gene from the chromosome and is not simply a point mutant that has no detectable (null) activity. However, it is sometimes technically difficult to ablate all sequences from a gene, and the gene fragment(s) that remain(s) may be transcriptionally active. In these instances, there is the potential for novel complexes between plasmid-encoded and chromosome-encoded gene products, even with the use of a "null" allele.

Gene-specific deletions and disruptions have been created in *E. coli* (Jasin & Schimmel, 1984; Winans et al., 1985; Toth & Schimmel, 1986; Shevell et al., 1988) and in *S. cerevisiae* (Rothstein, 1983; Winston et al., 1984) by site-specific recombination with linear DNA fragments that encode chromosomal DNA sequences which flank both sides of the gene of interest. In one approach, a gene for a selectable marker replaces the target gene and is inserted between the flanking chromosomal sequences of the linear fragment. Reciprocal recombination between the chromosomal target site and the homologous sequences at the ends of the linear fragment results in replacement of the target gene with that of the selectable marker. Southern blot analysis is used to confirm the change at the chromosomal site (Jasin & Schimmel, 1984). Additional procedures for gene alteration and inactivation are described by Gutterson & Koshland (1983) and Lukacsovich and Venetianer (1989).

If the product encoded by the gene is essential for catalytic activity, then the null allele will be lethal, so that specialized procedures are required to express plasmid-encoded mutants in cells that harbor a null allele on the chromosome. Lethality accompanying gene deletion is avoided in yeast by using diploid strains that retain one copy of the wild-type allele when the other is deleted from the sister chromosome. The activities of mutant proteins such as deletions of *S. cerevisiae* glutamine tRNA synthetase were tested by transformation of diploid strains (that harbor the null allele on one chromosome) with plasmids that encode the deletions. The transformants are then sporulated to give haploid cells with the autonomously replicating plasmid. The viable haploid spores are those which harbor plasmids that provide the essential function (Ludmerer & Schimmel, 1987).

For bacteria such as *E. coli* the function lost by a deletion in the monoploid chromosome must be provided by a maintenance plasmid that can be removed under selective conditions. Deletion of *alaS* or of *glyS* from the chromosome was accomplished in cells which have a "temperature-sensitive" maintenance plasmid that encodes *alaS* and *glyS*, respectively (Jasin & Schimmel, 1984; Toth & Schimmel, 1986). The maintenance plasmid has a temperature-sensitive replicon that prevents replication at the restrictive temperature of 42 °C (Meacock & Cohen, 1979). Thus, growth of cells that have an *alaS* deletion and the maintenance plasmid is temperature-sensitive.

Rescue of the growth phenotype is tested by introduction of a second plasmid that encodes a mutant. The second, test plasmid is compatible with the maintenance plasmid and can

replicate at the restrictive temperature. After a period at the restrictive temperature, the maintenance plasmid is lost and the test plasmid is the sole potential source of enzyme activity. If the mutant encoded by the test plasmid has sufficient activity, then cell growth can be resumed at a lower temperature, such as 30 or 37 °C. This is particularly useful for large-scale fermentations to produce bulk quantities of a mutant enzyme that are uncontaminated with an allelic version of the same protein.

The aforementioned C-terminal fragments of alanine tRNA synthetase that complement the *G674A* and *G677A* mutant alleles do not rescue growth of the *alaS* null strain (Jasin et al., 1984). Thus, this strain provided a rigorous screen for the activities of 18 deletions of the proteins and of a variety of point mutants. The same strain has also been the host cell for production and purification of those mutants that have activity sufficient to sustain cell growth.

Potential Artifacts from Recombination of Plasmid-Encoded Sequences with a Chromosomal Locus

Particularly when cell growth is done under selective pressure, recombination between plasmid- and chromosome-encoded sequences can occur in a region of DNA sequence homology [reviewed in Orr-Weaver and Szostak (1985) and Smith (1988)]. When selective pressure favors the change caused by recombination, then cells that harbor the altered molecules will gradually become predominant in the population.

Recombination can repair point mutations on the one hand and large deletions on the other. This can occur via a "single-crossover" integrative homologous recombination event which inserts the plasmid into the chromosome so as to combine two defective sequences into one that is operationally active. For example, complementation of *alaS* mutant alleles is observed with plasmids that encode inactive molecules, through integrative homologous recombination at the mutant chromosomal locus to create reconstructed genes that direct synthesis of active molecules (Jasin et al., 1985).

Instead of integrative recombination of plasmid sequences into the chromosome, complementation of mutant alleles can also occur by a reciprocal exchange from a "double crossover" that generates an altered plasmid- and chromosome-encoded species. (Alternatively, recombination may occur by gene conversion, which is a nonreciprocal transfer of information so that both plasmid and chromosome sequences are now identical.) Thus, a plasmid-encoded internal deletion of approximately 560 codons from *alaS* expresses a protein that is inactive. But when grown under selective pressure in the *G674A* or *G677A* temperature-sensitive mutant background, the multicopy plasmid (that had the internal deletion) acquires chromosomal sequences through a recombination event that completely repairs the deletion (Jasin et al., 1985). Thus, the resulting protein had some activity while the one that was originally expressed had none.

If a plasmid-encoded deletion acquires some of the missing protein sequences, then the change in protein size can easily be detected by molecular weight analysis of the plasmid-expressed protein. In contrast, the exchanges by reciprocal recombination of limited portions of genes can result in a plasmid-encoded species whose sequence difference from the starting protein is more subtle and gives a protein of unaltered size.

The three known pathways for recombination in *E. coli* are RecBCD, RecE, and RecF. These pathways are not independent and all three require the *recA* gene product [Clark, 1973; reviewed in Smith (1988)]. Because *recA* is not essential for growth under many conditions that are commonly used, the most straightforward way to minimize problems caused by recombination is to use a *recA*$^-$ host for the production and in vivo complementation analysis of mutant proteins that are encoded by plasmids that have homologous sequences on the chromosome. (In fungal systems like the yeast *S. cerevisiae*, there is no single gene counterpart to *recA* that can be inactivated so as to prevent mitotic recombination.) In addition, chromosomal deletion alleles can be used that minimize the chance of recombination with sequences from the plasmid that encode the gene that is deleted from the chromosome. However, if these null alleles have residual chromosomal sequences from the gene itself or its flanking sequences, then integrative recombination with homologous sequences in the plasmid may occur in a *recA*$^+$ background and, in effect, "repair" the chromosomal null allele (Toth & Schimmel, 1986).

Dampening Effects of C-Terminal Sequences That Complicate the Delineation of Functional Domains

While proteases are still used to isolate fragments of proteins that represent stable structural domains, gene deletions that express protein fragments from recombinant plasmids offer a far greater repertoire of fragments than the few that can be obtained with proteases. However, the experience with serrendipitously obtained stable and active proteolytic fragments belies the complications that arise with recombinant DNA constructs which introduce stop codons into protein coding sequences at any desired location or which create internal deletions that join distal elements in the sequence. Fragments that are generated by proteases define structural units, but arbitrary deletions can interrupt domains and introduce idiosyncratic conformational and functional changes. These changes can confuse the delineation of domains that are mapped by gene deletions.

The lack of synthesis of a stable product from some gene constructions is a potential source for misinterpretation of a lack of activity in vivo (e.g., complementation of a null allele) or in vitro (e.g., by direct assays of extracts) that has a trivial cause. Evidence for lack of stable quantities of a recombinant protein can be obtained by Western blot analysis with polyclonal antibodies of protein extracts that are resolved by electrophoresis (Burnette et al., 1981) or by visualization by ^{35}S-labeling of plasmid-encoded proteins that are synthesized in maxicells (Sancar et al., 1979). However, because subtle differences in sequence or size can produce major differences in the amount of stable protein that is detectable [cf. Jasin et al. (1983)], each mutant must be separately investigated to determine the relative amount that is produced.

While protein degradation or mRNA instability provides a trivial explanation for a lack of activity, the inactivity of a *stable* protein fragment does not necessarily mean that an essential catalytic site has been removed. This is presumably because some deletions may interrupt a domain and consequently alter the spatial relationships within the residual catalytic core that is still present. For example, the adenylate synthesis activity of glycine tRNA synthetase can be obscured in a deleted *glyS* protein that retains the active site for adenylate synthesis.

Glycine tRNA synthetase is an $\alpha_2\beta_2$ tetramer (Ostrem & Berg, 1970). The subunits have sizes of 303 and 689 amino acids, respectively (Webster et al., 1983). Both subunits are required for the aminoacyl adenylate synthesis and aminoacylation activities (McDonald et al., 1980). A set of C-terminal deletions in the β-chain have been coexpressed with the α-chain and examined for glycyl adenylate synthesis ac-

tivity in vitro (Toth & Schimmel, 1990). A deletion of 104 amino acids from the C-terminus of the β-chain (to give β575) has demonstrable glycyl adenylate synthesis in vitro. A deletion of 178 amino acids gives a β511 chain that has little or no detectable adenylate synthesis activity. This would suggest that residues essential for adenylate synthesis are located between positions 511 and 575 in the protein. However, deletion of an additional 128 amino acids gives β383, which has activity comparable to that of the wild-type, full-length β-subunit. Both β511 and β383 are synthesized in maxicells in comparable amounts and have the expected molecular weights. Thus, their differences in activities do not appear to be a question of amount, but rather a masking of activity in β511 by a conformational effect that is idiosyncratic to that protein.

The masking of catalytic activity by C-terminal sequences not present in the wild-type protein is further illustrated with fragment 461N of alanine tRNA synthetase, which encodes an active amino-terminal fragment of 461 amino acids (Ho et al., 1985). In spite of the in vivo and in vitro activity of 461N, the fusion of 14 in-frame codons derived from the plasmid vector pBR322 to the C-terminus of 461N yields fragment 461Nt, which neither complements the *alaS* null strain nor has any detectable aminoacylation activity in vitro. Both 461N and 461Nt can be overexpressed and purified to homogeneity and, when resolved by SDS gel electrophoresis, are clearly distinguished from each other by their molecular weight difference. Thus, there is a true loss of specific activity and not of protein stability upon fusion of 14 extraneous amino acids to fragment 461N.

Extraneous sequences that are joined to the protein of interest sometimes occur for technical reasons of convenience associated with the nature of the gene construction. The results with alanine tRNA synthetase show that these sequences can obscure the catalytic activity that is encoded by the rest of the protein. In contrast, the β-chain deletions of glycine tRNA synthetase demonstrate that the dampening effect of C-terminal segments is not restricted to foreign sequences. Consequently, only those fragments which *retain* activity can be unambiguously interpreted.

Synonymous Codon Replacements To Check for Spurious Activities from Translational Misreading

Serious errors of interpretation are possible when a low activity for an enzyme results from substitution of an active-site amino acid. Low activities are of particular interest because they can be instructive for the calculation of intrinsic strain or proximity effects imposed on substrate molecules that are bound at the active site. Thus, if replacement of what was believed to be an essential nucleophile like serine with a nonnucleophile like alanine or glycine results in a reduction in activity of 10^3–10^4-fold, then the residual activity might be ascribed to active-site-induced strain on the substrate that facilitates a reaction even without the putative critical nucleophile.

However, translational error frequencies of 10^{-3}–10^{-4} have been observed, and this in itself is sufficient to result in an amino acid misincorporation that yields active molecules in admixture with a completely inactive major species that contains no errors (Edelman & Gallant, 1977; Yarus, 1979; Parker & Friesen, 1980; Bouadaloun et al., 1983; Toth & Schimmel, 1988). A specific example in which the the codon for the active-site Ser68 of β-lactamase was replaced with a GGC glycine codon was recently described (Toth & Schimmel, 1988; Schimmel, 1990). Only cysteine is believed to be able to substitute functionally for Ser68, although the Cys68 mutant has approximately 2% of the activity of the wild-type enzyme (Fisher et al., 1980; Dalbadie-McFarland et al., 1982; Sigal et al., 1984). Yet the Ser68 → Gly (GGC) replacement results in a β-lactamase activity that is sufficient to confer ampicillin resistance on cells which encode the mutant and that can be measured easily in vitro by a standard assay for hydrolysis of penicillin G. However, this activity was shown to be not due to the Gly68 mutant, which has no detectable activity whatsoever (Toth & Schimmel, 1988).

The GGC glycine codon is inferred to be misread by the AGU/C-reading $tRNA^{Ser}$ isoacceptor at a frequency of about 0.1%, so that the Ser68 → Gly (GGC) mutant is a mixture of a Ser68 species with the predominant Gly68 mutant. This $tRNA^{Ser}$ isoacceptor has a GCU anticodon (and is therefore designated as $tRNA^{Ser/GCU}$) that is complementary to the last two bases of the GGC glycine codon. The misreading by this tRNA implies a G–U wobble interaction at the first position of the codon–anticodon interaction and was the first observation of an A–G wobble interaction at this position with a wild-type tRNA.

The problem of contamination of an inactive species with small amounts of an active species that differs by just one amino acid is not easy to solve by biochemical methods. For example, there is no convenient way to purify the resulting Ser68 species away from the Gly68 mutant, especially because there is no difference in net charge between the two molecules. Alternatively, the GGC glycine codon can be replaced with the GGA synonym that, in addition to an unusual first-position G–U wobble interaction, is not complementary to the third base of the anticodon of the $tRNA^{Ser/GCU}$ isoacceptor. Also, the GGA codon cannot be read by any of the three other serine tRNA isoacceptors by less than two-base misreading. Cells that harbor the Ser68 → Gly (GGA) mutant are sensitive to ampicillin and have no detectable activity in vitro for the hydrolysis of penicillin G (Toth & Schimmel, 1988). Thus, the β-lactamase activity observed for the Gly68 mutant is allele-specific, as expected for an example of translational misreading. There is no need to attempt a laborious separation of active from inactive species just to establish this point.

Mismatches can occur in principle at any of the three base-pairing interactions of the codon–anticodon interaction. This means that, before a detailed mechanistic interpretation is made, translational misreading has to be ruled out as a source of low activity associated with a mutant protein. Where it is technically feasible, active-site titrations and measurements of pre-steady-state kinetic parameters can compare the concentrations of active sites with the independently measured protein concentration and thereby address the problem at the biochemical level (Leatherbarrow et al., 1985). But the most rapid test for translational misreading is simply to replace the codon of interest with a synonym and determine whether the activity of the mutant protein is thereby affected.

"Critical" Residues Suggested by Chemical Modification and Sequence Comparisons

Crestfield et al. (1963a,b) concluded that His119 of ribonuclease A was essential for catalytic activity because its specific carboxymethylation resulted in complete enzyme inactivation. While this conclusion happened to be correct, it was based only on modification experiments, and there now are several examples where complete inactivation of an enzyme by modification has not revealed a residue that is essential for activity. These examples include among others His291 of *Rhodospirillum rubrum* ribulose carboxylase (Niyogi et al., 1986), Tyr160 of the catalytic subunit of *E. coli* aspartate transcarbamoylase (Robey & Schachman, 1984), Cys148 of

E. coli lac permease (Trumble et al., 1984; Viitanen et al, 1985; Sarkar et al., 1986), Tyr248 (Gardell et al., 1985; Hilvert et al., 1986; Auld et al., 1989) and Tyr198 (Gardell et al., 1987; Auld et al., 1989) of carboxypeptidase A, and Cys395 of the β-chain of *E. coli* glycine tRNA synthetase (Profy & Schimmel, 1986). In these instances, the complementary use of site-directed mutagenesis showed that replacement of the critical residue with another amino acid did not result in enzyme inactivation.

Similarly, the replacement of the conserved and therefore potentially essential Cys35 in *Bacillus stearothermophilus* tyrosyl-tRNA synthetase with either serine or glycine does not result in loss of function (Winter et al., 1982; Wilkinson et al., 1983). In another example, Harper and Vallee (1988) investigated the effect on ribonucleolytic activity of substitutions of Asp116 in angiogenin. This amino acid is conserved in all known pancreatic ribonucleases, and the crystal structure of bovine pancreatic ribonuclease A shows that the carboxyl side chain hydrogen bonds to an imidazole nitrogen of the active-site His119 and thereby may orient the imidazole group during catalysis (Richards & Wyckoff, 1971; Wodak et al., 1977; Campbell & Petsko, 1987; Wlodawer et al., 1988). However, D116N, D116A, and D116H mutant angiogenins are active and have enhanced ribonucleolytic activity toward tRNA, and the D116H mutant even has enhanced angiogenic activity.

The specific amino acid that is substituted for the potentially "essential" residue can markedly influence the results that are obtained. Thus, NEM modification of Cys395 of the β-chain of *E. coli* glycine tRNA synthetase completely abolishes the aminoacylation activity. However, Cys395 is not essential for activity because the C395A mutant is almost as active as the wild-type enzyme. On the other hand, the activity of a C395Q mutant is reduced by over an order of magnitude. This analysis suggests that NEM modification of Cys395 causes enzyme inactivation by steric effects and also suggests that a more detailed understanding of the role of a specific amino acid cannot be achieved without construction and investigation of many replacements.

Applications That Exploit Some of the Hazards

While there are many ways to introduce complications and artifacts in applications of mutagenesis to structure–function relationships, the complicating factors can also be the means to important applications. Thus, Wente and Schachman (1987) systematically investigated the dependence of activity of the catalytic trimer of aspartate transcarbamoylase on the presence of specific amino acids at or near the subunit interfaces. They showed that, for example, trimers of S52H (Ser52 → His) or K84Q mutant proteins each have less than 0.01% of the activity of the wild-type catalytic trimer, but when the mutant proteins are combined into hybrid trimers, the activity rises almost 10^5-fold. In a related approach, Distefano et al. (1990) used Cys → Ala replacements in the dimeric mercuric reductase to investigate the role of two pairs of cysteines (C135, C140 and C558, C559) for reduction of Hg(II). Coexpression of the A135A140C558C559 mutant with the C135C140A558A559 variant conferred Hg resistance that was 10^4-fold greater than that for cells which expressed only one of the mutant proteins. In vitro studies confirmed the presence of active heterodimers. Thus, the active site is at the subunit interface and requires a C135C140 pair from one subunit and a C558C559 pair from the other.

The association of inactive polypeptides to give active heterooligomers has its opposite counterpart in dominant negative mutants that form heterooligomers with wild-type chains and thereby inactivate them [cf. Herskowitz (1987)]. Early examples include the *lacI* and *recA* systems (Davis & Jacob, 1968; Mueller-Hill et al., 1968; Geisler & Weber, 1976; Yarranton & Sedgwick, 1982). More recently, peptide interference has been used to disrupt subunit assemblies by formation of peptide–subunit complexes that block subunit–subunit interactions. Examples include peptide displacement of subunit interactions in herpes simplex virus ribonucleotide reductase (Cohen et al., 1986; Dutia et al., 1986; Paradis et al., 1988), peptide blocking of the SV40 T-antigen–retinoblastoma protein complex (DeCaprio et al., 1989), and peptide interference of assemby of a viral promoter-specific transcription complex (Haigh et al., 1990). These examples suggest therapeutic applications of peptide interference. They also show that, in applications of mutagenesis to structure–function relationships, residual fragments of a protein (e.g., encoded by a presumed null allele on the chromosome or by extrachromosomal DNA) can potentially block as well as stimulate activity through formation of novel protein complexes.

While recombination between plasmid and chromosomal sequences can introduce sequence alterations that substantially affect biological activity and its interpretation, recombination can also be useful for mapping and cloning spontaneous mutations in an undefined location of a gene of interest. In a marker rescue experiment, gene fragments on plasmids are individually introduced into a host strain that bears a mutation that restricts growth. Under selective pressure, reciprocal recombination between plasmid and chromosomal sequences replaces a region of the chromosome with its counterpart on the plasmid. The part in common to those DNA segments that rescue growth defines the location of the mutation. By an extension of this kind of analysis, the mutation itself can be cloned directly from the chromosome and then sequenced (Jasin et al., 1985).

Concluding Remarks

Although screening for the activities of a large number of deletion and point mutants is facilitated by their introduction into a growth-restricted test strain (e.g., one that harbors a null allele), complementation of the growth defect per se is a poor indicator of the degree of activity. For example, there is a large range (perhaps 100-fold) of activity that can support growth of an *alaS* null strain (Ho et al., 1985). Also, there are examples where mutant tRNA amber suppressors are reported to be active by virtue of suppression in vivo of an amber mutation that restricts cell growth on the one hand (McClain et al., 1988) and yet show no detectable aminoacylation in vitro on the other (Hou et al., 1988; Park et al., 1989; Shi et al., 1990). However, in this case the amber suppression in vivo is weak and presumably represents a level of charging of the mutant tRNAs that is below limits that are detectable in vitro. Unfortunately, the ease of some of the in vivo genetic screens has sometimes obscured the lesson that analytical conclusions about structure–function relationships can only be drawn from detailed studies of defined molecules and reactions in vitro.

To remove all ambiguities of interpretations, the molecules under investigation must be purified to homogeneity, even if it means chromatographic separation of a protein from an allelic version that differs by only one amino acid, as has been achieved with triosephosphate isomerase (Nickbarg et al., 1988). If phenotypes observed in vivo and in extracts are not reproduced with purified molecules, then one uncharacterized interaction or more may occur in vivo and/or in extracts and these interactions might explain the discrepancy.

When the behavior of a purified species parallels that observed in vivo, there is still the possibility of translational misreading which produced sequence variants that have not been resolved by chromatography. Activity may then be due to trace amounts of a novel variant. While there are ways to deal with this possibility (see above), there are in principle additional complications that can result from expression of recombinant molecules in vivo. These include posttranscriptional or posttranslational modifications that are idiosyncratic to some recombinant constructs and not to others. Thus, an intron in a *S. cerevisiae* tyrosine tRNA ochre and a leucine tRNA amber suppressor is needed for a U35 → ψ modification and a C35 → 5-methyl-C modification, respectively (Johnson & Abelson, 1983; Strobel & Abelson, 1986). Recent examples with *E. coli* isoleucine and *S. cerevisiae* aspartate tRNAs demonstrate the substantial effect on posttranscriptional modifications on the specificity of charging of these tRNAs in vitro (Muramatsu et al., 1988; Perret et al., 1990). It is reasonable to assume that more of these kinds of possibilities have not been discovered only because they have not specifically been investigated. For this reason, the applications of molecular biology to structure–function analysis are likely to produce additional pitfalls that have yet to be described.

ACKNOWLEDGMENTS

Dr. Jonathan Burbaum made helpful suggestions and comments on the manuscript.

REFERENCES

Accolla, R. S., & Celada, F. (1976) *FEBS Lett. 67*, 299–302.
Anfinsen, C. B., Cuatrecasas, P., & Taniuchi, H. (1971) *Enzymes 4*, 177–204.
Auld, D. S., Riordan, J. F., & Vallee, B. L. (1989) in *Metal Ions in Biological Systems* (Sigal, H., & Sigel, A., Eds.) pp 359–394, Marcel Dekker, Inc., New York.
Bouadaloun, F., Donner, D., & Kurland, C. G. (1983) *EMBO J. 2*, 1351–1356.
Burnette, W. N. (1981) *Anal. Biochem. 112*, 195–203.
Campbell, R. L., & Petsko, G. A. (1987) *Biochemistry 26*, 8579–8584.
Clark, A. J. (1973) *Annu. Rev. Genet. 7*, 67–86.
Cohen, E. A., Gaudreau, P., Brazeau, P., & Langelier, Y. (1986) *Nature 321*, 441–443.
Crestfield, A. M., Stein, W. H., & Moore, S. (1963a) *J. Biol. Chem. 238*, 2413–2420.
Crestfield, A. M., Stein, W. H., & Moore, S. (1963b) *J. Biol. Chem. 238*, 2421–2428.
Davis, J., & Jacob, F. (1968) *J. Mol. Biol. 36*, 413–417.
DeCaprio, J. A., Ludlow, J. W., Lynch, D., Furukawa, Y., Griffin, J., Piwnica-Worms, H., Huang, C.-M., & Livingston, D. M. (1989) *Cell 58*, 1085–1095.
Distefano, M. D., Moore, M. J., & Walsh, C. T. (1990) *Biochemistry 29*, 2703–2713.
Dlabadie-McFarland, G., Cohen, L. W., Riggs, A. D., Morin, C., & Richards, J. H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6409–6413.
Dutia, B. M., Frame, M. C. Sukba-Sharpe, J. H., Clark, W. N., & Marsden, H. S. (1986) *Nature 321*, 439–441.
Edelmann, P., & Gallant, J. (1977) *Cell 10*, 131–137.
Fincham, J. R. S., & Coddington, A. (1963) *J. Mol. Biol. 6*, 361–373.
Fisher, J., Belasco, J. G., Khosla, S., & Knowles, J. R. (1980) *Biochemistry 19*, 2895–2901.
Galakatos, N. G., & Walsh, C. T. (1987) *Biochemistry 26*, 8475–8480.
Gardell, S. J., Craik, C. S., Hilvert, D., Urdea, M. S., & Rutter, W. J. (1985) *Nature 317*, 551.
Gardell, S. J., Hilvert, D., Barnett, J., Kaiser, E. T., & Rutter, W. J. (1987) *J. Biol. Chem. 262*, 576.
Garen, A., & Garen, S. (1963) *J. Mol. Biol. 7*, 13–22.
Geisler, N., & Weber, K. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 3103–3106.
Gutterson, N. I., & Koshland, D. E., Jr. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4894–4898.
Haigh, A., Greaves, R., & O'Hare, P. (1990) *Nature 344*, 257–259.
Harper, J. W., & Vallee, B. L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7139–7143.
Harper, J. W., Auld, D. S., Riordan, J. F., & Vallee, B. L. (1988) *Biochemistry 27*, 219–226.
Herskowitz, I. (1987) *Nature 329*, 219–222.
Hilvert, D., Gardell, S. J., Rutter, W. J., & Kaiser, E. T. (1986) *J. Am. Chem. Soc. 108*, 5298.
Ho, C., Jasin, M., & Schimmel, P. (1985) *Science 229*, 389–393.
Holmgren, A., & Slaby, I. (1979) *Biochemistry 18*, 5591–5599.
Hou, Y.-M., & Schimmel, P. (1988) *Nature 333*, 140–145.
Jackson, D. A., & Yanofsky, C. (1969) *J. Biol. Chem. 244*, 4539–4546.
Jasin, M., & Schimmel, P. (1984) *J. Bacteriol. 159*, 783–786.
Jasin, M., Regan, L., & Schimmel, P. (1983) *Nature 306*, 441–447.
Jasin, M., Regan, L., & Schimmel, P. (1984) *Cell 36*, 1089–1095.
Jasin, M., Regan, L., & Schimmel, P. (1985) *J. Biol. Chem. 260*, 2226–2230.
Johnson, P. F., & Abelson, J. (1983) *Nature 302*, 681–687.
Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459.
Leatherbarrow, R. J., Fersht, A. R., & Winter, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7840–7844.
Li, C. H., & Bewley, T. A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1476–1479.
Ludmerer, S. W., & Schimmel, P. (1987) *J. Biol. Chem. 262*, 10807–10813.
Lukacsovich, T., & Venetianer, P. (1989) *Mol. Gen. Genet. 218*, 364–366.
McClain, W. H., Chen, Y.-M., Foss, K., & Schneider, J. (1988) *Science 242*, 1681–1684.
McDonald, T., Breite, L., Pangburn, K. L. W., Hom, S., Manser, J., & Nagel, G. M. (1980) *Biochemistry 19*, 1402–1409.
Meacock, P. A., & Cohen, S. A. (1979) *Mol. Gen. Genet. 174*, 135–147.
Mueller-Hill, B., Crapo, L., & Gilbert, W. (1968) *Proc. Natl. Acad. Sci. U.S.A. 59*, 1259–1264.
Muramatsu, T., Nishikawa, K., Nemoto, F., Kucino, Y., Nishimura, S., Miyazawa, T., & Yokoyama, S. (1988) *Nature 336*, 179–181.
Nickbarg, E. B., Davenport, R. C., Petsko, G. A., & Knowles, J. R. (1988) *Biochemistry 27*, 5948–5960.
Niyogi, S. K., Foote, R. S., Mural, R. J., Larimer, F. W., Mitra, S., Soper, T. S., Machanoff, R., & Hartman, F. C. (1986) *J. Biol. Chem. 261*, 10087.
Ostrem, D. L., & Berg, P. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 1967–1974.
Paradis, H., Gaudreau, P., Brazeau, P., & Langelier, Y. (1988) *J. Biol. Chem. 263*, 16045–16050.

Parker, J., & Friesen, J. D. (1980) *Mol. Gen. Genet. 177*, 439–445.
Perret, V., Garcia, A., Grosjean, H., Ebel, J.-P., Florentz, C., & Giege, R. (1990) *Nature 344*, 787–789.
Profy, A. T., & Schimmel, P. (1986) *J. Biol. Chem. 261*, 15474–15479.
Profy, A. T., & Schimmel, P. (1988) *Prog. Nucleic Acid Res. Mol. Biol. 35*, 1–26.
Putney, S. D., Sauer, R. T., & Schimmel, P. R. (1981a) *J. Biol. Chem. 256*, 198–204.
Putney, S. D., Royal, N. J., de Vegvar, H. N., Herlihy, W. C., Biemann, K., & Schimmel, P. (1981b) *Science 213*, 1497–1501.
Regan, L., Bowie, J., & Schimmel, P. (1987) *Science 235*, 1651–1653.
Richards, F. M., & Wyckoff, H. W. (1971) *Enzymes 4*, 647–806.
Robey, E. A., & Schachman, H. K. (1984) *J. Biol. Chem. 259*, 11180.
Rothstein, R. J. (1983) *Methods Enzymol. 101*, 202–211.
Sancar, A., Hack, A. H., & Rupp, W. D. (1979) *J. Bacteriol. 137*, 692–693.
Sarkar, H. K., Menick, D. R., Viitanen, P. V., Poonian, M. S., & Kaback, H. R. (1986) *J. Biol. Chem. 261*, 8914.
Schimmel, P. (1989) *Acc. Chem. Res. 22*, 232–233.
Schimmel, P. (1990) *Adv. Enzymol. Relat. Areas Mol. Biol. 63*, 233–270.
Schlesinger, M. J., & Levinthal, C. (1963) *J. Mol. Biol. 7*, 1–12.
Shevell, D. E., Abou-Zamzam, A. M., Demple, B., & Walker, G. C. (1988) *J. Bacteriol. 170*, 3294–3296.
Shi, J.-P., Francklyn, C., Hill, K., & Schimmel, P. (1990) *Biochemistry 29*, 3621–3626.
Sigal, I. S., DeGrado, W. F., Thomas, B. J., & Petteway, S. R., Jr. (1984) *J. Biol. Chem. 259*, 5327–5332.
Slaby, I., & Holmgren, A. (1979) *Biochemistry 18*, 5584–5591.
Smith, G. R. (1988) *Microbiol. Rev. 52*, 1–28.
Strittmatter, P., Barry, R. E., & Corcoran, D. (1972) *J. Biol. Chem. 247*, 2768–2775.
Strobel, M. C., & Abelson, J. (1986) *Mol. Cell. Biol. 6*, 2663–2673.
Taniuchi, H., Parr, G. R., & Juillerat, M. A. (1986) *Methods Enzymol. 131*, 185–217.
Theall, G., Low, K. B., & Soll, D. (1977) *Mol. Gen. Genet. 156*, 221–227.
Toth, M. J., & Schimmel, P. (1986) *J. Biol. Chem. 261*, 6643–6646.
Toth, M. J., & Schimmel, P. (1990) *J. Biol. Chem. 265*, 1000–1004.
Toth, M. J., Murgola, E. J., & Schimmel, P. (1988) *J. Mol. Biol. 201*, 451–454.
Trumble, W. R., Viitanen, P. V., Sarkar, H. K., Poonian, M. S., & Kaback, H. R. (1984) *Biochem. Biophys. Res. Commun. 119*, 860–867.
Viitanen, P. V., Menick, D. R., Sarkar, H. K., Trumble, W. R., & Kaback, H. R. (1985) *Biochemistry 24*, 7628–7635.
Webster, T. A., Gibson, B. W., Keng, T., Biemann, K., & Schimmel, P. (1983) *J. Biol. Chem. 258*, 10637–10641.
Wente, S. R., & Schachman, H. K. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 31–35.
Winans, S. C., Elledge, S. J., Krueger, J. H., & Walker, G. C. (1985) *J. Bacteriol. 161*, 1219–1221.
Winston, F., Chumley, F., & Fink, G. R. (1983) *Methods Enzymol. 101*, 211–228.
Wlodawer, A., Svensson, L. A., Sjolin, L., & Gilliland, G. L. (1988) *Biochemistry 27*, 2705–2717.
Wodak, S. Y., Liu, M. Y., & Wyckoff, H. W. (1977) *J. Mol. Biol. 116*, 855–875.
Yarranton, G. T., & Sedgwick, S. G. (1982) *Mol. Gen. Genet. 185*, 99–104.
Yarus, M. (1979) in *Transfer RNA: Structure, Properties and Recognition* (Schimmel, P., Soll, D., & Abelson, J., Eds.) pp 501–515, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Zabin, I., & Fowler, A. V. (1980) in *The Operon* (Miller, J. H., & Reznikoff, W. S., Eds.) pp 89–121, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Zamenhof, P. J., & Villarejo, M. (1972) *J. Bacteriol. 110*, 171–178.