PROTEINS: Structure, Function, and Genetics 7:99–111 (1990)

Review Article

Understanding Structural Relationships in Proteins of Unsolved Three-Dimensional Structure

Jonathan J. Burbaum, Ruth M. Starzyk, and Paul Schimmel
Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

INTRODUCTION

There are over 17,000 protein sequences in the current GenBank sequence database,[1] and the three dimensional structures of less than 500 of these proteins are known to atomic resolution.[2] The number of new protein sequences continues to grow more rapidly than the number of new structures determined either by X-ray diffraction or nuclear magnetic resonance methods. Unsolved structures include those of proteins that have pharmaceutical applications, either as therapeutic agents in themselves (tissue plasminogen activator, erythropoietin, and the interferons, for example) or as the targets of rationally designed drugs (e.g., the reverse transcriptases of retroviruses). The lack of three-dimensional structures for these and other proteins impedes the development of many important applications.

The three-dimensional structures determined so far have shown that most naturally occurring proteins contain a limited number of structural motifs, which are found in the context of widely different primary sequences. In virtually all proteins, the same fundamental secondary structure motifs, α helices, β sheets, regular turns, etc., are present. These fundamental units are assembled into tertiary structures that can be associated with specific functions, such as binding nucleotides or spanning membranes. Some families of proteins, such as the globins, have significant sequence differences and yet form three-dimensional structures whose α-carbon backbones largely coincide.[3] Thus, a large number of sequences can be built into a limited number of secondary and tertiary structures. Consequently, proteins whose three-dimensional structures are not known can be modelled (at least in part) by analogy with known structures.

To model a protein based on sequence information alone, one or more proteins of known structure must be identified as a starting point. This selection is based on a functional or sequence relationship to the protein of interest. However, similarity of function does not imply similarity in structure; more than one structural motif may perform the same general function. One example, associated with DNA recognition, is the dichotomy of the helix-turn-helix[4] and zinc finger[5] structural motifs. A second criterion for selection of a known structure is a sequence relationship. Similarity (or even identity) in sequence, however, does not imply similarity in structure; identical, short peptides can form more than one secondary structure motif.[6]

In practice, even when both functional and sequence relationships are employed, there are inherent pitfalls. These can be dealt with, at least in part, by experimental tests of the proposed structure. Using site-directed mutagenesis amino acids can be changed or structural motifs can be altered and the roles of particular structural elements in the model can be evaluated. It is the interplay of structural modeling with mutagenesis that is described below.

The structural modeling described here involves members of the aminoacyl-tRNA synthetase family of enzymes. In each cell, there are 20 different aminoacyl-tRNA synthetases (one for each amino acid). These enzymes have a common function, namely, matching an amino acid with its corresponding tRNA. This function is separated into two enzymatic activities: amino acid activation, to form aminoacyl-AMP, and charging, to transfer the aminoacyl group to the tRNA. (Some aminoacyl-tRNA synthetases also possess an editing activity that attenuates mischarging of tRNA.) An additional function, such as regulation of transcription or translation, is also found in some of the enzymes.[7]

Of the 20 enzymes, partial three-dimensional structures are known for only two (the tyrosine enzyme from *Bacillus stearothermophilus*[8,9] and the methionine enzyme from *Esherichia coli*[10–12]), although a third structure (the glutamine enzyme from *E. coli*[13]) has recently been solved. As shown below, the existing sequence information on several of these enzymes, together with structural data on

Received July 11, 1989; revision accepted September 6, 1989.
Ruth M. Starzyk's Present address is Alkermes, Inc., 26 Landsdowne Street, Cambridge, MA 02139.
Address reprint requests to Dr. Paul Schimmel, Department of Biology, M.I.T. 16-437, 77 Massachusetts Ave., Cambridge, MA 02139.

TABLE I. Alignment of Sequences of Aminoacyl-tRNA Synthetases That Contain a Signature Sequence*

Enzyme	Source	Sequence	Reference
Arg-tRNA synthetase	*E. coli*	^{122}PNVAKEMHVGH	15
Gln-tRNA synthetase	*E. coli*	^{34}PEPNGYLHIGH	16
Gln-tRNA synthetase	*H. sapiens*	^{37}PEASGYLHIGH	17
Gln-tRNA synthetase	*S. cerevisiae*	^{258}PEPNGYLHIGH	18
Glu-tRNA synthetase	*E. coli*	^{9}PSPTGYLHVGG	19
Ile-tRNA synthetase	*E. coli*	^{57}PYANGSIHIGH	14
Ile-tRNA synthetase	*S. cerevisiae*	^{47}PFATGTPHYGH	20
Leu-tRNA synthetase	*E. coli*	^{42}PYPSGRLHMGH	21
Leu-tRNA synthetase	*N. crassa*	^{84}PYMNGRLHAGH	22
Leu-tRNA synthetase	*S. cerevisiae*†	^{56}PYPSGALHIGH	23
Met-tRNA synthetase	*E. coli*	^{14}PYANGSIHLGH	24
Met-tRNA synthetase	*S. cerevisiae*	^{205}PYVNNVPHLGN	25
Met-tRNA synthetase	*S. cerevisiae*†	^{22}FYPNAKPHLGH	26
Trp-tRNA synthetase	*B. stearothermophilus*	^{8}IQPSGVITIGN	27
Trp-tRNA synthetase	*E. coli*	^{10}AQPSGELTIGN	28
Trp-tRNA synthetase	*S. cerevisiae*†	^{41}IQPTGCFHLGN	29
Tyr-tRNA synthetase	*B. caldotenax*	^{37}DPTADSLHIGN	30
Tyr-tRNA synthetase	*B. stearothermophilus*	^{37}DPTADSLHIGH	31
Tyr-tRNA synthetase	*E. coli*	^{40}DPTADSLHLGH	32
Tyr-tRNA synthetase	*N. crassa*†	^{102}DPTAPSLHVGH	33
Tyr-tRNA synthetase	*S. cerevisiae*†	^{92}DPTAQSLHLGN	34
Val-tRNA synthetase	*B. stearothermophilus*	^{49}PNVTGKLHLGH	35
Val-tRNA synthetase	*E. coli*	^{42}PNVTGSLHMGH	36
Val-tRNA synthetase	*S. cerevisiae*‡	^{190}PNVTGALHIGH	37

*Sequences are aligned based on the glycine of the HIGH sequence (see text). The first residue of the sequence is numbered, and the conserved prolyl, histidyl, and glycyl residues are boldface. Numbering is relative to the N-terminus of the mature protein (where characterized); otherwise, numbering begins at the initiator methionine of the translated DNA sequence.
†Mitochondrial
‡Mitochondrial and cytoplasmic.

two of them, has allowed building of a model of another synthetase (of divergent sequence and considerably larger size than the basis structures), which has been tested by mutagenesis. These investigations have also given additional insights into structure–function relationships of one synthetase whose structure is partially solved.

NEAR-IDENTITY OF AN ELEVEN AMINO ACID SEGMENT (SIGNATURE SEQUENCE) SUGGESTS THAT ISOLEUCYL-TRNA SYNTHETASE IS STRUCTURALLY RELATED TO AN ENZYME OF KNOWN STRUCTURE

The primary sequences of proteins often contain similarities that suggest historical and/or structural relationships. Such similarity might reasonably be expected of the aminoacyl-tRNA synthetases, because they catalyze the same general reaction. However, only limited similarity has been found. Alignment of sequences to assess such similarity is complicated further by a diversity of quaternary structures and polypeptide chain lengths. Four classes of quaternary structure (α, α_2, α_4, $\alpha_2\beta_2$) have been identified,[7] and sizes of subunits vary from 329 amino acids (Trp-tRNA synthetase [α_2] from *B. stearothermophilus*) to 951 amino acids (Val-tRNA synthetase [α] from *E. coli*).

One notable similarity is a sequence of 11 amino acids in *E. coli* Ile-tRNA synthetase that is nearly identical to a segment of the *E. coli* methionine enzyme.[14] Several other synthetases contain analogous sequences (Table I), although the number of identities in the isoleucine-methionine pair is greatest. This segment is referred to as a "signature sequence" for a subclass of the aminoacyl-tRNA synthetases. As is shown in Table I, a signature sequence element has been identified in 24 enzymes of nine different specificities. The signature sequence is characterized by a C-terminal tetrapeptide (His-hydrophobic-Gly-[His or Asn][1]), referred to as the HIGH sequence. Preceding the HIGH sequence, a peptide of the formula Pro-$(Xaa)_n$-hydrophobic (where n is most often 6) is the consensus.

[1]The exceptions to this pattern are found in enzymes that have substantial similarity in this region to an enzyme that contains the consensus sequence. In all enzymes, the glycine is conserved.

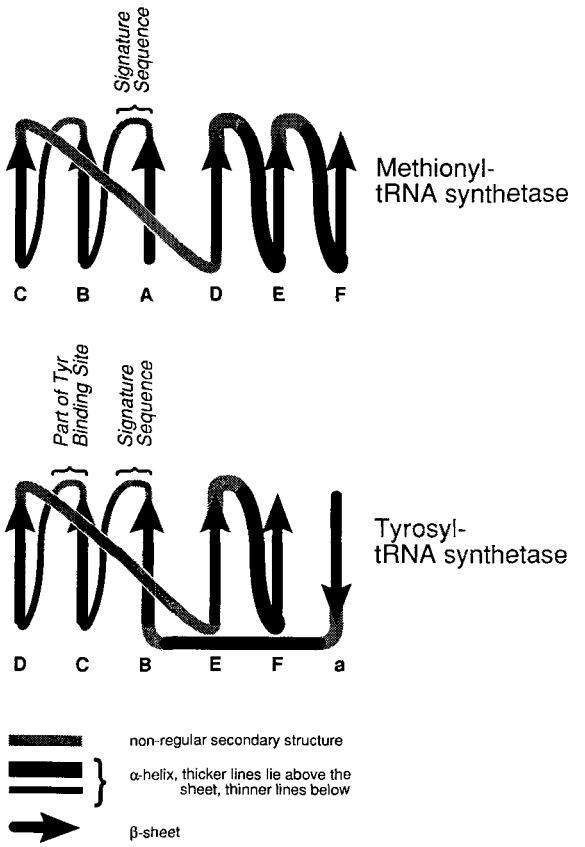

Fig. 1. Schematic representation of nucleotide folds in *E. coli* methionyl-tRNA synthetase and *B. stearothermophilus* tyrosyl-tRNA synthetase. Nucleotide folds, containing six strands of β-sheet and four α-helices, are shown as they occur in the N-terminal domains of two crystal structures of the tRNA synthetases. Solid black lines denote α-helices, and solid arrows indicating the direction of the polypeptide chain denote strands of β-sheet. Regions of nonregular secondary structure are represented by stippled lines. Note the similarities in structure despite a different orientation of the elements of the nucleotide fold. In both representations, the N-terminus is at the base of the arrow labeled A (or a), and the C-terminus is at the tip of the arrow labeled F. Both enzymes also contain a C-terminal domain (not represented). For the tyrosyl-tRNA synthetase, the region of the tyrosine binding site is also indicated. Reproduced from "Evidence from Cassette Mutagenesis for a Structure-Function Motif in a Protein of Unknown Structure" by Clarke et al., with permission of SCIENCE.[56]

SPATIAL DISPOSITION OF SIGNATURE SEQUENCE PEPTIDES IN TWO ENZYMES OF KNOWN STRUCTURE

The secondary structural motifs of the N-terminal domains of Met-tRNA synthetase and Tyr-tRNA synthetase are diagrammed in Figure 1. Both structures can be viewed as Rossmann folds,[38] a structural motif that was originally identified for the dehydrogenases.[39,40] This motif contains six strands of β-sheet and four α-helices in the arrangement of two mononucleotide binding folds (βαβαβ structures). In the high-resolution structures of the tyrosine and methionine enzymes (and in the glutamine structure[13]), the signature sequence peptides occupy the same position in the fold (Fig. 1).

The signature sequence peptides of these two enzymes are superimposable[9,41] and start at the C-terminus of one strand of β-sheet, continue through a loop, and finish at the N-terminus of an α-helix. Because of this remarkable structural identity, it is likely that the signature sequence plays an identical role in these two enzymes. Mutagenesis in the signature sequence region of Tyr-tRNA synthetase by Fersht and coworkers[42] has suggested functions for two of the amino acids of the HIGH sequence. The first histidine is thought to stabilize the γ-phosphoryl group of adenosine triphosphate (ATP) in the transition state of tyrosyl-AMP formation, and the second histidine (or Asn; either variant is fully active) is suggested to bind to the α-phosphate of ATP and of the tyrosyl-adenylate (Tyr-AMP). This functional assignment may explain why the first histidine can be replaced by threonine (as in two of the three Trp-tRNA synthetases listed in Table I), because the OH of threonine would be expected to play the same role as the δ1-NH of histidine. The conserved glycine seems to play no direct role in binding or catalysis but may promote formation of the proper structure; similar, glycine-rich regions are found in other nucleotide binding folds.[43]

From a historical perspective, it is plausible that the subclass of synthetases that includes the methionine and isoleucine enzymes has diverged early in the evolution of modern organisms,[44] possibly at or near the time when the rules of the universal genetic code were systematized. Alternatively, many of the enzymes may have evolved from distinct progenitors. In either case, the historical linkage between Met-tRNA synthetase and Tyr-tRNA synthetase is likely to be weak. Because these two enzymes fold into superimposable structures at the signature sequence, this peptide probably forms the same three-dimensional structure (having the same basic function) in most, if not all, of the synthetases listed in Table I.

SEQUENCE ANALYSIS, AUGMENTED BY PREDICTIONS OF SECONDARY STRUCTURE, SUGGESTS A STRUCTURAL DOMAIN IN ISOLEUCYL-TRNA SYNTHETASE

The limited sequence identities have illuminated structural relationships within the synthetase family. More insight can be obtained from primary sequences by considering substitutions within more broadly defined groups of amino acids, such as replacements of an amino acid with those having similar chemical properties.[45] A natural database of possible substitutions is found in the sequences of proteins that perform identical functions in diverse organisms.[46] This allows the delineation of a "reduced alphabet" for detection of primary sequence similarity of a coarser nature.

When the primary sequences of the methionine

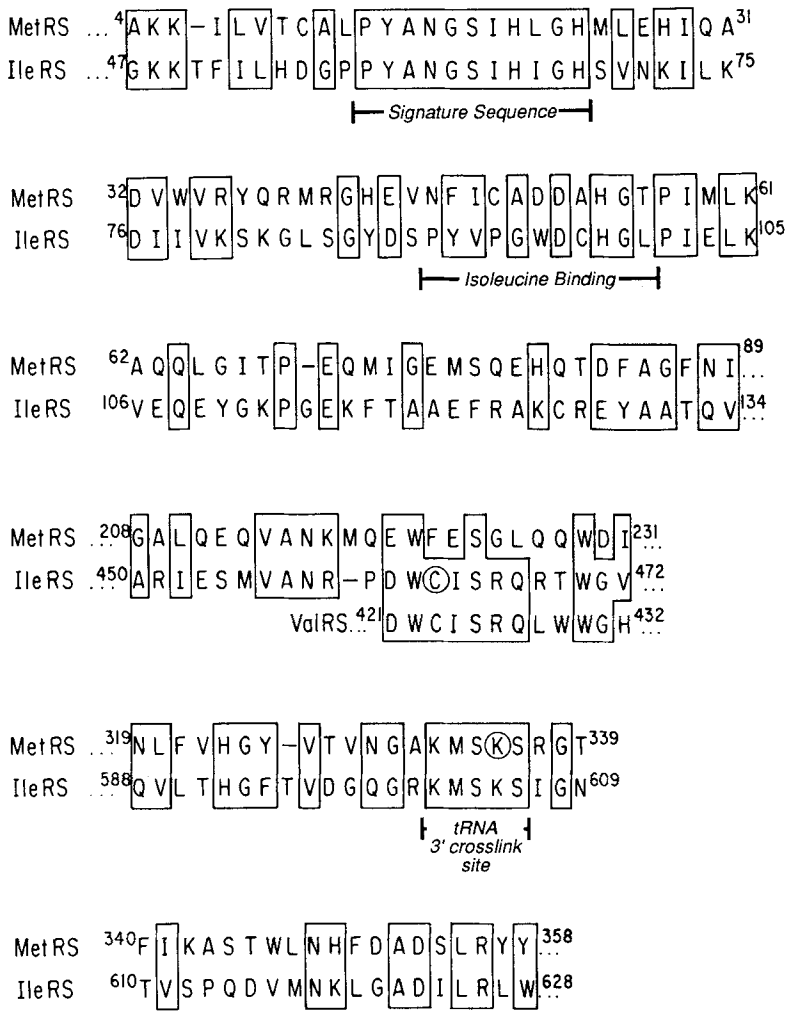

Fig. 2. Reduced alphabet comparison of isoleucyl-tRNA synthetase with the methionine and valine enzymes. Residues 47–134, 450–472, and 588–628 of Ile-tRNA synthetase are aligned with residues 4–89, 208–231, and 319–358 of Met-tRNA synthetase, respectively. A sequence of Val-tRNA synthetase (421–432) that shows some similarity with residues 461–472 of Ile-tRNA synthetase is also shown. The signature sequence (P58 to H68 of the isoleucine enzyme) is highlighted, as is G94, which was identified by cassette mutagenesis as part of the isoleucine binding site. C463 of Ile-tRNA synthetase can be cross linked to an isoleucine analog[47] and is circled. Also circled is K335 of Met-tRNA synthetase, which can be cross-linked to the 3'-end of oxidized Met-tRNA.[71] Reproduced from "Evidence for Dispensable Sequences Inserted into a Nucleotide Fold" by Starzyk et al., with the permission of SCIENCE.[50]

and isoleucine enzymes are aligned using a reduced alphabet (Fig. 2), two regions of significant similarity are identified. A third region can be extended to include Val-tRNA synthetase and includes Cys463 of Ile-tRNA synthetase, which can be affinity labeled with an isoleucine analog.[47] These regions of similarity correspond to particular structural elements of Met-tRNA synthetase. Specifically, the two $\beta\alpha\beta\alpha\beta$ halves of the Rossmann fold of Met-tRNA synthetase are connected by a peptide designated connective polypeptide 1, or CP1. A second segment, called connective polypeptide 2 (or CP2), splits the protein at a second site (between β_D and α_E). The regions of high similarity between the methionine and isoleucine enzymes correspond to secondary structures that form the Rossmann fold, whereas regions of low similarity correspond to the two connective polypeptides. Every Rossmann fold requires a CP1 segment to connect the two halves of the fold, but the length of this segment can vary. In lactate dehydrogenase, for example, this segment is less than 10 amino acids long.[48] In contrast, CP1 is much larger in the Tyr- and Met-tRNA synthetases (approximately 60 and 130 amino acids long, respectively). Sequences that immediately flank CP1 (Fig. 2) have greater than 50% similarity, whereas no significant similarity is found between the two CP1 segments. Furthermore, the predicted CP1 of Ile-tRNA synthetase is about 200 amino acids longer than that of Met-tRNA synthetase.

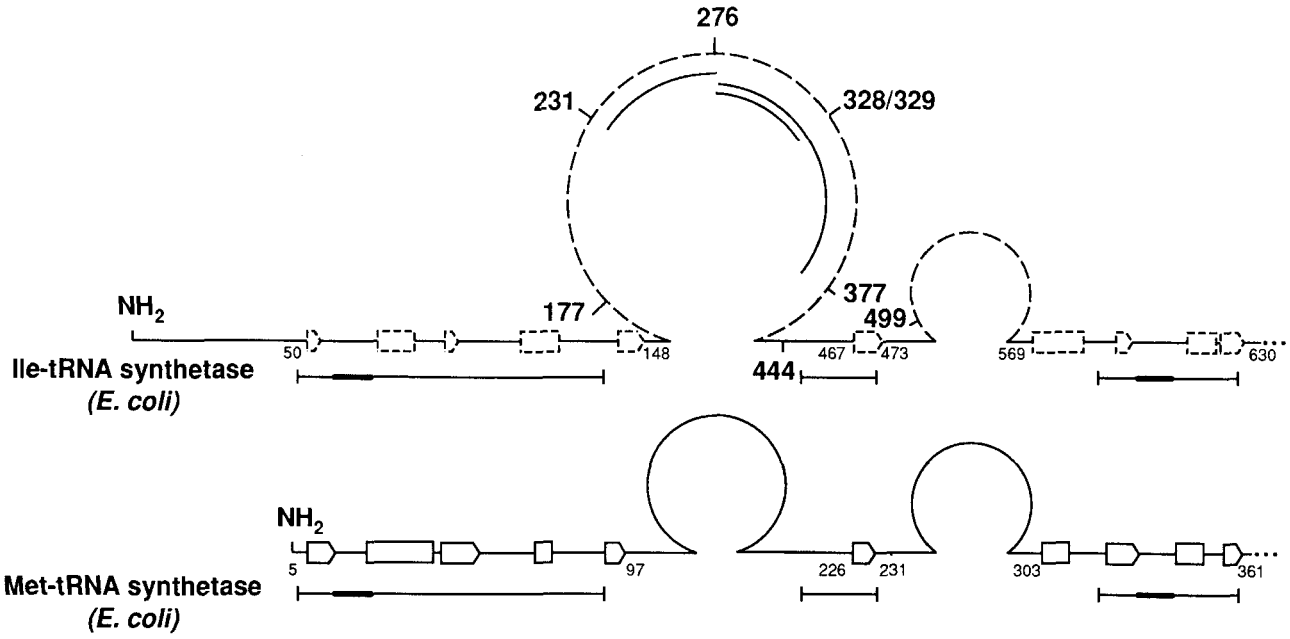

Fig. 3. Alignment of the primary sequences of Ile-tRNA synthetase with Met-tRNA synthetase. Secondary structural elements of the prototypical Rossmann fold of methionyl-tRNA synthetase are indicated as pentagons (β-strands) or rectangles (α-helices). The analogous structural units, as generated by sequence alignment, are also indicated for the isoleucine enzyme. Regions of higher similarity are indicated by the bars, and presumed functional regions (the signature sequence and the tRNA 3′-cross-link site) are indicated by the heavy lines. The large loops represent the connective polypeptides CP1 (nearest the N-terminus) and CP2. In Ile-tRNA synthetase, the segments of CP1 that can be deleted without elimination of enzyme activity are indicated by solid bars, and the associated numbers indicate the boundaries of the deleted segments. Reproduced from "Evidence for Dispensable Sequences Inserted into a Nucleotide Fold" by Starzyk et al., with the permission of SCIENCE.[50]

The similarities of the structural motifs and the location of the signature sequence peptide in the tyrosine and methionine enzymes suggest that common patterns of structure might be found in comparisons of predicted secondary structures. Using this approach, the secondary structure similarities in the signature sequence region (based on an alternating β/α structure as an N-terminal domain) can be extended to synthetases such as Ala-tRNA synthetase. Gly-tRNA synthetase, and Asp-tRNA synthetase of *Saccharomyces cerevisae,* which have no significant similarity to any of the signature sequence variations of Table I.[49] Ile-tRNA synthetase has regions of primary sequence that are similar to segments in Met-tRNA synthetase, and the *predicted* secondary structure elements of the isoleucine enzyme can be aligned with the *known* secondary structure of the methionine enzyme (Fig. 3). Although the structural elements for Ile-tRNA synthetase were predicted empirically from a modified Chou/Fasman program,[50,51] these assignments are strengthened by their association with known structures in Met-tRNA synthetase, in the parts of the primary sequences that flank the CP1 segments.

Taken together, these structural comparisons have allowed us to construct a working model of Ile-tRNA synthetase (Fig. 4). This model includes structural elements that parallel the known structure of the methionyl-tRNA synthetase from *E. coli,* and the secondary structural elements of a Rossmann nucleotide-binding fold are illustrated. Sections of probable functional significance of this predicted structure are noted. A feature of this model is the two connective polypeptides, to which no function has been assigned. In the next section, we summarize evidence that at least part of one of these segments can be deleted without loss of function, suggesting that this segment of the protein contains dispensable sequences.

DEMONSTRATION THAT THE PREDICTED NUCLEOTIDE FOLD OF ISOLEUCYL-TRNA SYNTHETASE CONTAINS DISPENSABLE SEQUENCES

The variation in size of the CP1 segment of the synthetases suggests that some of the large CP1 polypeptide of Ile-tRNA synthetase may not be required for catalytic function. To test for the presence of nonfunctional sequences in this segment, portions of Ile-tRNA synthetase were deleted by removing sequences between restriction sites of a plasmid-borne *ileS* gene that encodes the enzyme. Deletion proteins of the expected sizes were visualized by gel electrophoresis of ^{35}S-labeled plasmid encoded proteins, which were synthesized in a maxicell system,[52] and activities were measured in vitro.

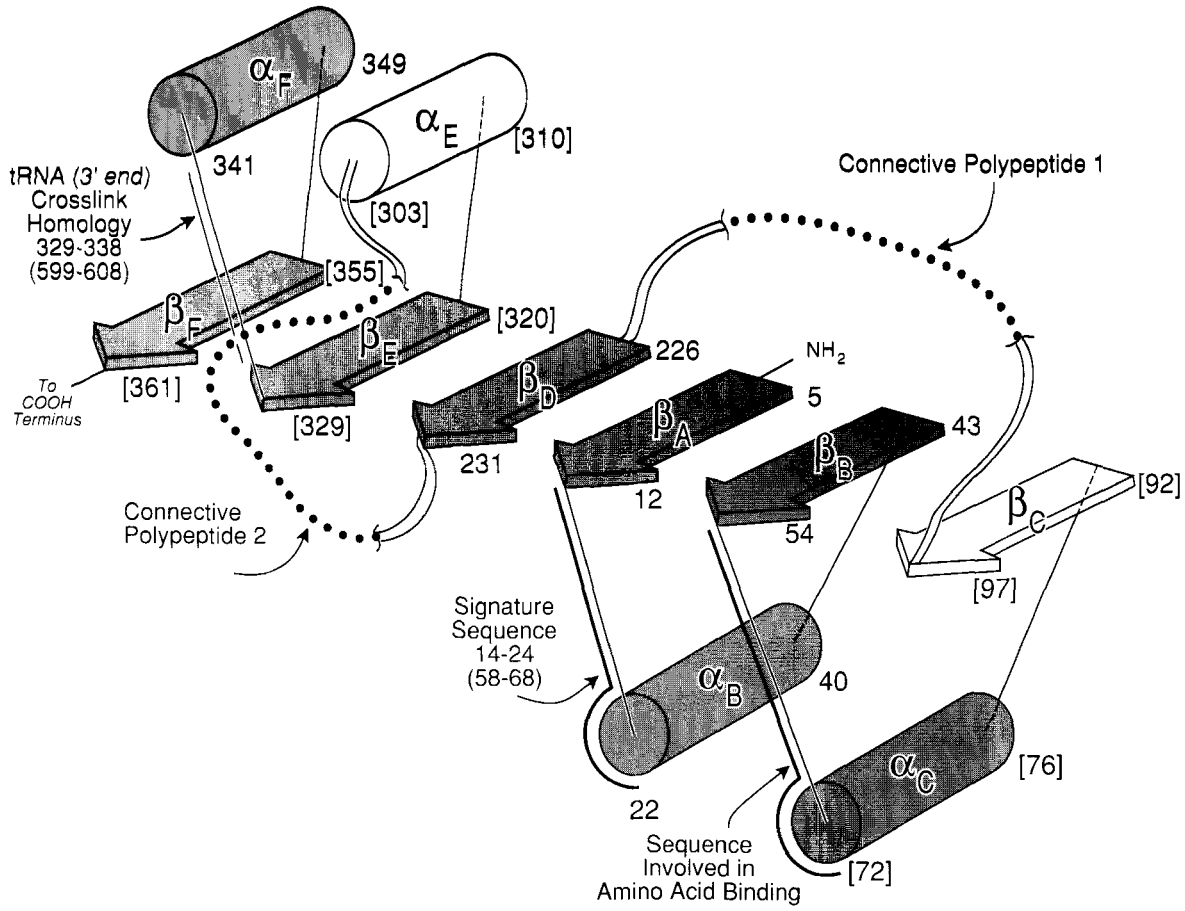

Fig. 4. A three-dimensional model for isoleucyl-tRNA synthetase. The six strands of β-sheet and the four α-helices of a nucleotide fold are represented by arrows and cylinders, respectively. The fold is broken at two points, labeled connective polypeptide 1 (called CP1 in the text) and connective polypeptide 2 (called CP2 in the text). Two regions of extensive structural information, the signature sequence and the tRNA 3'-cross-link site, are labeled. The numbers represent residues in the structure of the methionine enzyme. Comparable residues for the signature sequence and 3'-cross-link site in Ile-tRNA synthetase are given in parentheses. Brackets indicate preliminary assignments for Met-tRNA synthetase. Shading indicates regions of amino acid sequence similarity. Reprinted from "Evidence for Dispensable Sequences Inserted into a Nucleotide Fold" by Starzyk et al., with the permission of SCIENCE.[50]

Some of the deletion proteins retain adenylate synthesis and aminoacylation activities.[50] The segments of CP1 that can be deleted without elimination of activity are diagrammed in Figure 3. Collectively, these deletions encompass 146 of the 316 amino acids that make up CP1. Deletions that either extend into or are entirely within the predicted nucleotide fold on either side of CP1 result in unstable or inactive proteins. These results demonstrate that dispensable sequences are present in Ile-tRNA synthetase and support the structural model for the arrangement of the nucleotide fold and structural elements shown in Figure 4.

Although extensive dispensable sequences at the C-terminal end of Ala-tRNA synthetase,[53] Met-tRNA synthetase,[54] and Tyr-tRNA synthetase[55] have been reported, this was the first observation of dispensable sequences internal to a catalytic domain. The CP1 segment of the synthetases is not completely devoid of function, however (see below). Virtually all the size difference between Ile-tRNA synthetase (939 amino acids) and Met-tRNA synthetase (679 amino acids) is accounted for by the difference in size of the connective polypeptides (about 186 and 24 amino acids for CP1 and CP2, respectively) and the N-terminal extension (43 amino acids) of Ile-tRNA synthetase. The functional consequence (if any) of this size difference is not known, but differences of this type may account for the polymorphism of synthetase polypeptides.

DELINEATING A STRUCTURAL MOTIF IMPORTANT FOR AMINO ACID BINDING

Interspecies comparisons of sequences suggest experiments that can give structural insights. For example, comparative sequence information was used to guide experiments designed to locate part of the isoleucine binding site of Ile-tRNA synthetase.[56]

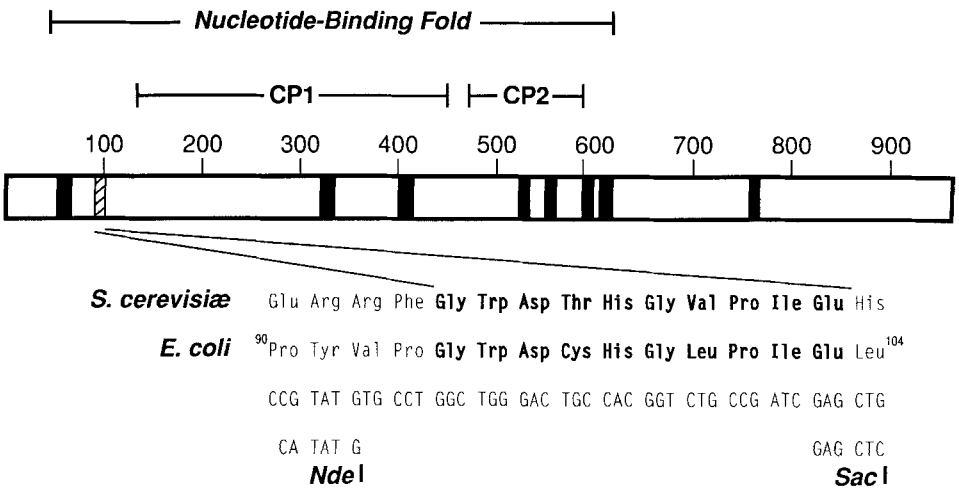

Fig. 5. Alignment of the primary sequences of the Ile-tRNA synthetases from *S. cerevisiae* and *E. coli*. Regions of highest similarity are denoted by the blackened or cross-hatched areas in the alignment. The leftmost highlighted area is the signature sequence, and the cross-hatched one was shown to be important for amino acid binding. The sequence of the cross-hatched region for both the yeast and bacterial enzymes is given, together with the restriction sites that were introduced (with no change in sequence) to produce the cassette. Also, the regions referred to in the text and in other figures, the nucleotide binding fold with its components CP1 (for connective polypeptide 1) and CP2, are indicated. Reprinted from "Evidence from Cassette Mutagenesis for a Structure-Function Motif in a Protein of Unknown Structure" by Clarke et al., with the permission of SCIENCE.[56]

Overall, sequences of Ile-tRNA synthetase from *E. coli* and *S. cerevisiae* have only 27% identity, even if gaps in the sequence alignment are permitted. The dissimilarity between the two sequences is not uniform, however; regions of low similarity are punctuated by regions of higher similarity. The regions of higher similarity are colinear within the two sequences. Nine regions of extended similarity (greater than 70% identity; Fig. 5), which could represent functionally conserved regions, were identified.

The tyrosine binding site of Tyr-tRNA synthetase has been characterized.[57] The tyrosine binding pocket contains amino acids from several regions of the polypeptide chain. One of these—Asp78—is 26 residues beyond the C-terminal end of the signature sequence. The 26 residue spacer consists of an α-helix and a β-strand, so that the signature sequence and the tyrosine binding site are proximal (Fig. 1). (One of the matched peptide sequences of the isoleucine enzymes falls at the same location.) If Ile-tRNA synthetase has an analogous nucleotide binding fold in its N-terminal domain, then this matched region is likely to occupy the same spatial position as the aspartate in the tyrosine binding site of Tyr-tRNA synthetase. This peptide would then be partially responsible for binding isoleucine.

By cassette mutagenesis,[58] the conserved amino acids in the postulated amino acid binding site of Ile-tRNA synthetase were mutated at random. In this experiment, oligonucleotides spanning a small element (41 bp; Fig. 5) of *ile*S were synthesized with intentional errors (at the level of 0.7–2.1 substitutions per fragment). This protocol resulted in a population of point and multiple mutations within a defined region of the synthetic fragment, which were then ligated to the appropriate sites within *ile*S. The ligation mixture was first transformed into an *ile*S$^+$ strain, and the resulting family of mutant plasmids was reintroduced into an auxotrophic strain of *E. coli* by performing 93 independent transformations. Sixty-two of these transformants contain plasmids that encode phenotypically wild-type enzyme, and sequence analysis showed that these collectively have at least 11 independent nucleotide substitutions (Table II). Of the wild-type proteins, ten preserve the original amino acid sequence with silent nucleotide changes and one results in a Cys97→Ser substitution. Because the yeast enzyme encodes Thr at the analogous position, Ser is a plausible alternative. Among the 31 remaining phenotypically mutant plasmids, 28 produced clones with no detectable activity or protein. Of the three remaining transformants, one had an Ile102→Asn substitution (at the ninth position within the identity block). The kinetic parameters for this mutant were affected less than tenfold. The remaining two mutants encode different nucleotide substitutions that result in the same Gly94→Arg replacement (GGC→CGA and GGC→CGC) and result in an enzyme of severely reduced activity.

In the activation reaction, the Gly94→Arg mutant has a K_m for isoleucine that is elevated 6,000-fold, although there is only a fourfold elevation of the K_m for ATP and no significant effect on k_{cat} (Table III). In the charging reaction, the K_m for tRNA is changed less than fourfold. Thus, the effect of the mutation is specific for isoleucine and affects bind-

TABLE II. Sequences of the *Nde*I–*Sac*I Fragment (Extending From Tyr91 to Glu103) of *E. coli* Ile-tRNA Synthetase Mutants Resulting From a Cassette Mutagenesis Experiment (See Text and Fig. 5)*

Type	Sequence of cassette												
Wild type	^{91}Tyr	Val	Pro	Gly	Trp	Asp	Cys	His	Gly	Leu	Pro	Ile	Glu^{103}
	—T	GTG	CCT	GGC	TGG	GAC	TGC	CAC	GGT	CTG	CCG	ATC	G—
Phenotypic wild types†													
Cys97→Ser	Tyr	Val	Pro	Gly	Trp	Asp	**Ser**	His	Gly	Leu	Pro	Ile	Glu
	—T	GTG	CCT	GGC	TGG	GAC	**TCC**	CAC	GGT	CTG	CCC	ATC	G—
Phenotypic mutants													
Gly94→Arg	Tyr	Val	Pro	**Arg**	Trp	Asp	Cys	His	Gly	Leu	Pro	Ile	Glu
	—T	GTG	CCT	**CGA**	TGG	GAC	TGC	CAC	GGT	CTG	CCG	ATC	G—
Gly94→Arg	Tyr	Val	Pro	**Arg**	Trp	Asp	Cys	His	Gly	Leu	Pro	Ile	Glu
	—T	GTG	CCT	**CGC**	TGG	GAC	TGC	CAC	GGT	CTG	CCG	ATC	G—
Ile102→Asn	Tyr	Val	Pro	Gly	Trp	Asp	Cys	His	Gly	Leu	Pro	**Asn**	Glu
	—T	GTG	CCT	GGC	TGG	GAC	TGC	CAC	GGT	CTG	CCG	**AAC**	G—

*The DNA sequence (5′→3′) of this segment of *E. coli* gene is written beneath the translated amino acid sequence of the fragment. Codon and deduced amino acid changes are shown in boldface. (Adapted from Clarke et al.[56] with permission of the publisher.)
†Independent mutations that resulted in no change in amino acid sequence are not shown.

TABLE III. Michaelis-Menten Parameters for Wild-Type and Gly94→Arg Mutant of Ile-tRNA Synthetase*

Enzyme	K_m (Ile) (μM)	K_m (ATP) (mM)	K_m (tRNA) (apparent) (μM)	Relative k_{cat}
Wild type	5	0.4	0.7	1.0
Gly94→Arg mutant	33,000	1.5	0.2	0.85

*The K_m values for ATP and isoleucine were measured with the pyrophosphate exchange assay. The isoleucine concentration for each K_m (ATP) determination was at least threefold above the K_m (Ile) for that enzyme. Relative k_{cat} values were determined by normalizing the pyrophosphate exchange values for V_{max} to the amount of enzyme in crude extracts, as determined by densitometry of Western blots with antibodies to Ile-tRNA synthetase. The apparent K_m for tRNA was determined with purified $tRNA^{Ile}$ at 1 mM ATP and 25 μM isoleucine. (Reproduced from Clarke et al.[56] with permission of the publisher.)

ing of the amino acid. The results support the extended similarity between the isoleucine and tyrosine enzymes proposed earlier and strongly suggest that the isoleucine binding site involves part of the segment that extends from Gly94 to Glu103.

INSERTION OF NEW SEQUENCES INTO THE CATALYTIC DOMAIN OF METHIONYL-TRNA SYNTHETASE

Although at least part of the predicted CP1 of Ile-tRNA synthetase is dispensable for catalysis, those parts that cannot be deleted without loss of activity (see above) may include regions that directly or indirectly affect the amino acid activation reaction. The region designated as CP1 in Met-tRNA synthetase is located in part on the surface.[10–12] This region includes Cys158, which appears to be essential for catalysis; substitution of Leu or Ala at this position completely inactivates synthesis of methionyl adenylate.[59]

Although there is evidence for a role for CP1 in amino acid activation, the results on Ile-tRNA synthetase suggests that the smaller CP1 of Met-tRNA synthetase might be enlarged without loss of activity. To explore this possibility, oligopeptides were inserted at eight locations (Fig. 6) in an 82-amino-acid region of CP1 that extends from Pro130 to Glu212.[59] These locations correspond to restriction sites in the gene (either natural or introduced by mutagenesis), and, at each site, the insertion was performed in two steps. First, a kanamycin resistance marker, flanked on either side by fragments containing multiple restriction sites, was introduced into the CP1 portion of *met*G. Plasmids containing insertions were selected by growth on kanamycin-containing media, and, finally, the kanamycin resistance marker was excised from the gene using sites in the flanking fragments that preserve the continuity of the gene. Depending on the restriction enzyme used in this excision, the gene for the insertion mutant is 5–14 codons larger than the gene for the wild-type enzyme.

All eight insertions are tolerated in the structure; the insertion proteins are clearly visible upon gel electrophoresis of crude extracts. (In contrast, a short deletion in CP1 of Met-tRNA synthetase resulted in a protein that did not accumulate in vivo, and some of the deletion proteins of Ile-tRNA synthetase were somewhat unstable in crude extracts [see above]). Of the eight mutant enzymes, seven have apparent activities in adenylate synthesis, and four of these have more than 20% of the activity of the wild-type protein. The two most active insertion mutants have activities that are virtually identical to that of the wild-type enzyme. These two insertion mutants (14 amino acids inserted at positions 130

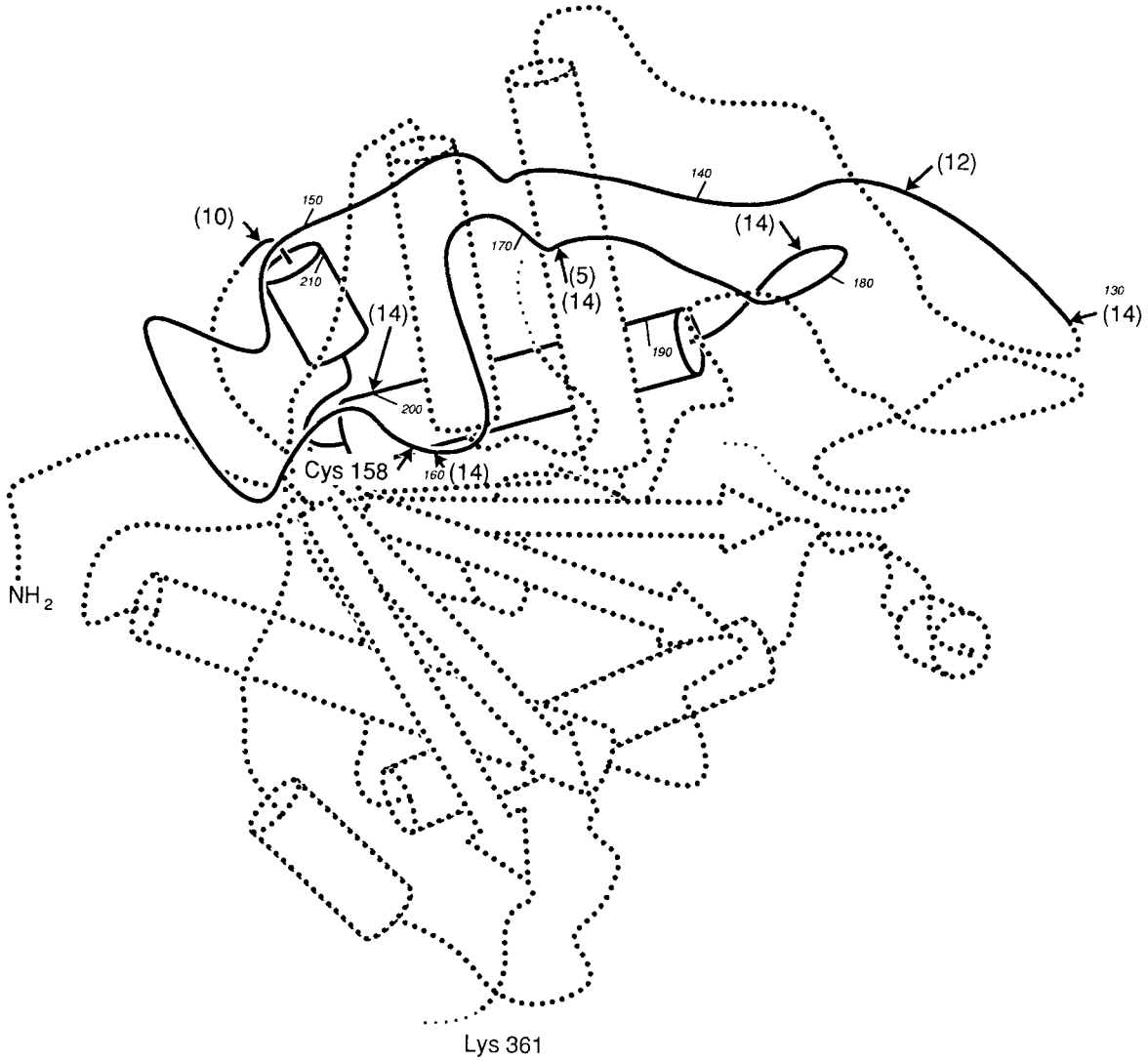

Fig. 6. Three-dimensional representation of the nucleotide fold in *E. coli* methionyl-tRNA synthetase. Arrows indicate β-strands of the prototypical nucleotide fold, and cylinders indicate proposed α-helices. This representation is not exhaustive and is subject to refinement. The region of CP1 that was probed by site-directed insertional mutagenesis is represented as a solid line. In this region, every tenth amino acid is indicated, and arrows denote positions of insertions with their peptide lengths given in parentheses. Reproduced from "Evidence for Dispensable Sequences Inserted into a Nucleotide Fold" by Starzyk et al. [modelled after Brunie et al.[10]], with the permission of SCIENCE.[50]

TABLE IV. Michaelis-Menten Parameters for Wild-Type and Two Insertion Mutants of Met-tRNA Synthetase in the Adenylate Synthesis Reaction*

Enzyme	K_m(Met) (μM)	K_m(ATP) μM	k_{cat} (sec^{-1})	k_{cat}/K_m(Met) (10^{-6} M^{-1} sec^{-1})	k_{cat}/K_m(ATP) (10^{6} M^{-1} sec^{-1})
Wild-type	38	295	83	2.2	0.28
Fourteen residues inserted at					
Pro 130	86	633	94	1.1	0.15
Met 184	61	194	23	0.38	0.12

*The insertion mutants, of 14 amino acids at Pro130 and Met184, were constructed as described in the text. The values reported for k_{cat} were calculated from the known enzyme concentrations and from values of the apparent V_{max}. The concentration range for methionine was 5 to 200 μM and 20–500 μM for ATP (at higher concentrations of ATP, the kinetics were not well behaved). Enzyme concentration was determined by the adenylate burst assay using 1 mM methionine. The enzyme was monomeric under the conditions of the assay as activity was a linear function of protein concentration. Reprinted with permission from "Insertion of New Sequences into the Catalytic Domain of an Enzyme," *Biochemisty* 28:8479–8484, 1989.[59]

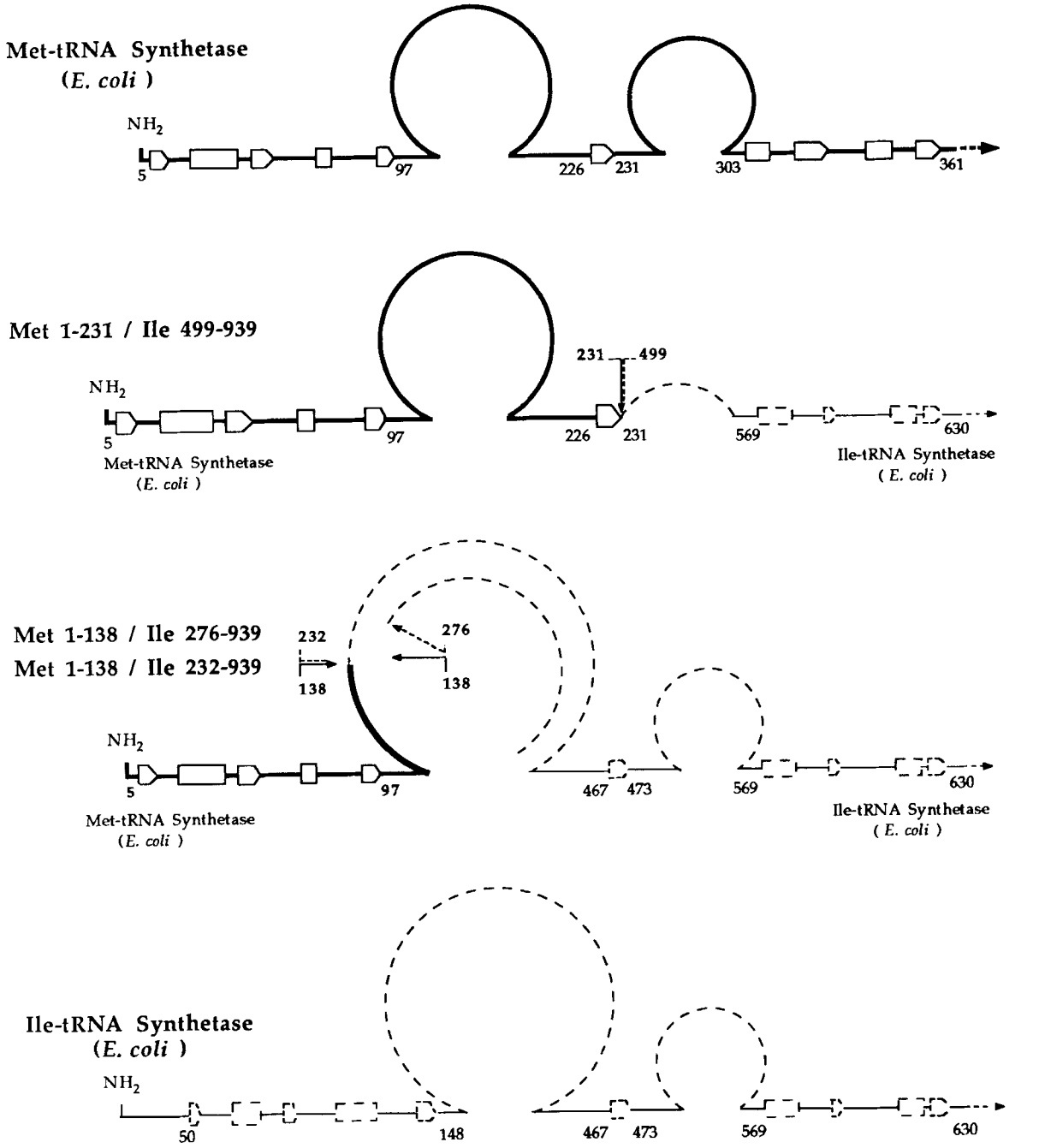

Fig. 7. Diagrammatic representation of three chimeric synthetase constructs. Secondary structure elements of the prototypical Rossmann fold of *E. coli* Met-tRNA synthetase are indicated as solid line rectangles (α-helices) or pentagons (β-strands) along the primary sequence (thick solid line), with boundaries given in small numbers. Dashed rectangles and pentagons within a thin line indicate predictions of secondary structure for Ile-tRNA synthetase that are guided by sequence similarities between the two enzymes. The connective polypeptides CP1 and CP2 are indicated as in Figure 3. Sites of connections between the two synthetases are indicated by arrows with numbers referring to the last amino acid of the methionine enzyme and the first amino acid of Ile-tRNA synthetase. Reprinted from "Construction of Intra-Domain Chimeras of Aminoacyl-tRNA Synthetases" by Starzyk and Schimmel, with the permission of the Journal of Biomolecular Structure & Dynamics.[61]

and 184) were purified, and their kinetic parameters were measured (Table IV). Neither k_{cat} nor K_m (for either substrate) is significantly affected, confirming that, at these two locations in the catalytic structure, oligopeptides of 14 amino acids can be incorporated. Sequence conservation does not appear to be a strong indicator of whether a given insertion will produce an active enzyme, in that these two inser-

tions were made into regions of the sequence that are evolutionarily conserved between the *E. coli* and *S. cerevisiae* enzymes. Interestingly, insertions that have the lowest activities cluster around Cys158 in the primary sequence (a well conserved region) but appear to be well separated in the three-dimensional structure (Fig. 6).

These results with the methionine synthetase may provide for a variety of novel applications. For example, were the inserted sequences properly designed, they could provide binding sites for additional ligands, alter the internal flexibility and "breathing" modes of the active site region,[60] or facilitate the construction of chimeric enzymes. As a first step toward the design of enzymes with novel activities, *intradomain* chimeras were constructed.

CONSTRUCTION OF INTRADOMAIN CHIMERAS OF AMINOACYL-TRNA SYNTHETASES

The studies described above provide a rationale for construction of intradomain hybrids between the Met- and Ile-tRNA synthetases, joined at sequences within the nucleotide fold. Convenient restriction sites in the two connective polypeptides, CP1 and CP2, of the two enzymes allowed us to construct five chimeric genes that encode the N-terminal portion of Met-tRNA synthetase connected to the C-terminal portion of Ile-tRNA synthetase.[61] The location of the connection in three of five intradomain chimeras constructed is illustrated in Fig. 7. Thus the hybrid proteins have a nucleotide-binding fold that contains pieces from both enzymes. Such "intradomain" chimeras, which are joined within a domain, are conceptually distinct from "interdomain" chimeras, which are proteins with connections between whole domains.[62–65] Although packing interactions might not be optimal, intradomain chimeras of synthetases might contain binding sites for ATP and tRNA (the common substrates among synthetases) that use binding interactions from both parent enzymes. By employing a suitable regime of mutagenesis and selection, an inactive, simple chimera might be modified to produce an active synthetase.

Initially, it was necessary to determine whether any of the five chimeric genes could direct synthesis of stable proteins. Three chimeric proteins (illustrated in Fig. 7) accumulate in vivo under maxicell conditions, whereas the other two could not be detected. The stable chimeras, two with connections in CP1 (Met [1–138]–Ile [232–939] and Met [1–138]–Ile [276–939]) and one with connections in CP2 (Met [1–231]–Ile [499–939]) were also detected by immunoblot analysis (with polyclonal antibodies made against isoleucyl-tRNA synthetase) in protein extracts from cells grown under normal (rather than maxicell) conditions. The two additional chimeric genes, which connect codon 126 of *met*G to either codon 232 or codon 276 of *ile*S, encode gene products that do not accumulate in vivo.[61]

In the three stable chimeras, the possibility of a functional "compound" tRNA or ATP binding site can be examined by testing one for tRNA and ATP binding. For the methionine enzyme and other synthetases, some of the determinants for tRNA recognition are located on the C-terminal side of the nucleotide fold.[10,66–68] Given that the hybrid proteins contain a large C-terminal portion of isoleucyl-tRNA synthetase, it will be of interest to determine if any of them specifically bind $tRNA^{Ile}$. Because the amino acid binding site appears to include amino acids from the two sides of the nucleotide fold, functional amino acid binding would not be expected to occur. As expected, these proteins were found to catalyze neither the methionine- nor the isoleucine-dependent pyrophosphate exchange reaction, which measures the first step of the overall aminoacylation reaction.

The genes for the stable chimeric prototypes can be mutagenized and the encoded protein checked for activity by the use of special selection or screening procedures.[69,70] For example, selection could be made for a functional isoleucyl-tRNA synthetase by transformation of a bacterial strain in which *ile*S is deleted from the chromosome. Wild-type *ile*S located on a plasmid that is temperature-sensitive for replication would allow cells to grow at 30°C but not at 42°C. These cells would then be transformed with a second, compatible plasmid encoding a mutated chimeric gene. Some of the transformants that grow at 42°C may contain plasmids with chimeric genes encoding isoleucyl-tRNA synthetase activity.

SUMMARY

The locations of functionally important sequences and general structural motifs have been assigned to Ile-tRNA synthetase. However, a function has not been established for some segments of the protein (e.g., CP1). The method of structural modeling described here cannot establish the details of a 3 Å crystal structure, and, in contrast to a crystal structure, the precision of the model varies according to the extent of a sequence similarity or the functional importance of a region. In Ile-tRNA synthetase, the signature sequence and the flanking regions are likely to be similar in structure to the proteins on which the model is based. For other regions, it may be possible to build a three-dimensional model by connecting well defined regions and refining the positions of the connecting elements by energy minimization.

Structural modelling of this kind must be done cautiously, because the order and orientation of the elements of a structural motif can change in subtle ways. In the case of Tyr-tRNA synthetase, the β-strand nearest the N-terminus is the outermost strand of the nucleotide binding fold; in Met-tRNA

synthetase, the same strand is innermost. Furthermore, the orientation of this strand may be antiparallel (Tyr-tRNA synthetase) or parallel (Met-tRNA synthetase). Because multiple structures that differ in their orientations of structural elements are possible, the structural analogies between proteins should not be naively extrapolated without independent experimental support.

As described above, some regions of proteins tolerate internal deletions and insertions. This provides further experimental support for the practice of allowing for gaps in computer-generated sequence alignments. Nevertheless, because some regions are more tolerant of insertions and deletions than others, the structural and functional significance of a region of broken alignment must be assessed carefully. All gaps in sequence alignments cannot be treated equally, and each must be evaluated within its own context.

In the synthetases of known structure, structural analogy can be used to identify important functional elements. For example, the amino acid binding site of Met-tRNA synthetase might be formed, at least in part, by a peptide that encompasses Ala50; this amino acid aligns with Gly94 of the Ile-tRNA synthetase. This is an example in which results on a protein of unknown structure (Ile-tRNA synthetases) can lead to identification of a potential substrate binding site in a protein of known structure (Met-tRNA synthetase).

REFERENCES

1. GenBank Release 55.0 contains 17,047 loci.
2. The current version of the Brookhaven Protein Data Bank contains approximately 350 unique structures.
3. Bashford, C., Chothia, C., Lesk, A.M. Determinants of a protein fold: Unique features of the globin amino acid sequences. J. Mol. Biol. 196:199–216, 1987.
4. Pabo, C.O., Sauer, R.T. Protein-DNA recognition. Annu. Rev. Biochem. 53:293–321, 1984.
5. Berg, J.M. Potential metal-binding domains in nucleic acid binding proteins. Science 232:485–487, 1986 (Evans, R.M., Hollenberg, S.M. Zinc fingers: Gilt by association. Cell 52:1–3, 1988).
6. Kabsch, W., Sander, C. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. Proc. Natl. Acad. Sci. USA 81:1075–1078, 1984.
7. Schimmel, P. Aminoacyl tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. Annu. Rev. Biochem. 56:125–158, 1987.
8. Bhat, T.N., Blow, D.M., Brick, P., Nyborg, J. Tyrosyl-tRNA synthetase forms a mononucleotide-binding fold. J. Mol. Biol. 158:699–709, 1982.
9. Blow, D.M., Brick, P. Aminoacyl-tRNA Synthetases. In: "Biological Macromolecules and Assemblies," Vol. 2 (Nucleic Acids and Interactive Proteins) Jurnak, F.A., MacPherson, A., eds. New York: John Wiley and Sons, 1985.
10. Brunie, S., Mellot, P., Zelwer, C., Risler, J.-L., Blanquet, S., Fayat, G. Structure-activity relationships of methionyl-tRNA synthetase: Graphics modelling and genetic engineering. J. Mol. Graphics 5:18–21,26, 1987.
11. Zelwer, C., Risler, J.L., Brunie, S. Crystal structure of Escherichia coli methionyl-tRNA synthetase at 2.5 Å resolution. J. Mol. Biol. 155:63–81, 1982.
12. Risler, J.L., Zelwer, C., Brunie, S. Methionyl-tRNA synthetase shows the nucleotide binding fold observed in dehydrogenases. Nature 292:384–386, 1981.
13. Steitz, T. Personal communication.
14. Webster, T., Tsai, H., Kula, M., Mackie, G.A., Schimmel, P. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. Science 226:1315–1317, 1984.
15. Eriana, G., Dirheimer, G., Gangloff, J. Isolation and characterization of the gene coding for *Escherichia coli* arginyl-tRNA synthetase. Nucleic Acids Res. 17:5725–5736, 1989.
16. Hoben, P., Royal, N., Cheung, A., Yamao, F., Biemann, K., Söll, D. *Escherichia coli* glutaminyl-tRNA synthetase. II. Characterization of the glnS gene product. J. Biol. Chem. 257:1644–1650, 1982.
17. Thömmes, P., Fett, R., Schray, B., Kunze, N, Knippers, R. The core region of human glutaminyl-tRNA synthetase homologies with the *Escherichia coli* and yeast enzymes. Nucleic Acids Res. 16:5391–5406, 1988.
18. Ludmerer, S.W., Schimmel, P. Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. J. Biol. Chem. 262:10801–10806, 1987.
19. Breton, R., Sanfacon, H., Papayannopoulos, I., Biemann, K., Lapointe, J. Glutamyl-tRNA synthetase of *Escherichia coli*. Isolation and primary structure of the *gltX* gene and homology with other aminoacyl-tRNA synthetases. J. Biol. Chem. 261:10610–10617, 1986.
20. Englisch, U., Englisch, S., Markmeyer, P., Schischkoff, J., Sternbach, H., Kratzin, H., Cramer, F. Structure of the yeast isoleucyl-tRNA synthetase gene (*ILS1*). DNA sequence, amino acid sequence of proteolytic peptides of the enzyme and comparison of the structure to those of other known aminoacyl-tRNA synthetases. Biol. Chem. Hoppe-Seyler, 368:971–979, 1987.
21. Hartlein, M., Madern, D. Molecular cloning and nucleotide sequence of the gene for *Escherichia coli* leucyl-tRNA synthetase. Nucleic Acids Res. 15:10199–10210, 1987.
22. Benarous, R., Chow, C.M., RajBhandary, U.L. Cytoplasmic leucyl-tRNA synthetase of *Neurospora crassa* is not specified by the *leu-5* locus. Genetics 119:805–814, 1988.
23. Herbert, C.J., Labouesse, M., Dujardin, G., Slonimski, P.P. The *NAM2* proteins from *S. cerevisiae* and *S. douglasii* are mitochondrial leucyl-tRNA synthetases, and are involved in mRNA splicing. EMBO J. 7:473–83, 1988.
24. Dardel, F., Fayat, G., Blanquet, S. Molecular cloning and primary structure of the Escherichia coli methionyl-tRNA synthetase gene. J. Bacteriol. *160:*1115–22, 1984.
25. Walter, P., Gangloff, J., Bonnet, J., Boulanger, Y., Ebel, J.P., Fasiolo, F. Primary structure of the *Saccharomyces cerevisiae* gene for methionyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA 80:2437–2441, 1983.
26. Tzagoloff, A., Vambutas, A., Akai, A. Characterization of *MSM1*, the structural gene for yeast mitochondrial methionyl-tRNA synthetase. Eur. J. Biochem. 179:365–371, 1989.
27. Winter, G.P., Hartley, B.S. The amino acid sequence of tryptophanyl-tRNA synthetase from *Bacillus stearothermophilus*. FEBS Lett. 80:340–342, 1977.
28. Hall, C.V., Van Cleemput, M., Muench, K.H., Yanofsky, C. The nucleotide sequence of the structural gene for *Escherichia coli* tryptophanyl-tRNA synthetase. J. Biol. Chem. 257:6132–6136, 1982.
29. Myers, A.M., Tzagoloff, A. *MSW*, a yeast gene coding for mitochondrial tryptophanyl-tRNA synthetase. J. Biol. Chem. 260:15371–15377, 1985.
30. Jones, M.D., Lowe, D.M., Borgford, T., Fersht, A.R. Natural variation of tyrosyl-tRNA synthetase and comparison with engineered mutants. Biochemistry 25:1887–1891, 1986.
31. Winter, G., Koch, G.L.E., Hartley, B.S., Barker, D.G. The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. Eur. J. Biochem. 132: 383–387, 1983.
32. Barker, D.G., Bruton, C.J., Winter, G. The tyrosyl-tRNA synthetase from *Escherichia coli*. Complete nucleotide sequence of the structural gene. FEBS Lett. 150:419–423, 1982.
33. Akins, R.A., Lambowitz, A.M. A protein required for splicing group I introns in *Neurospora* mitochondria is mito-

chondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell 50:331–345, 1987.
34. Hill, J., Tzagoloff, A. Personal communication.
35. Borgford, T.J., Brand, N.J., Gray, T.E., Fersht, A.R. The valyl-tRNA synthetase from *Bacillus stearothermophilus* has considerable sequence homology with the isoleucyl-tRNA synthetase from *Escherichia coli*. Biochemistry 26: 2480–2486, 1987.
36. Heck, J.D., Hatfield, G.W. Valyl-tRNA synthetase gene of *Escherichia coli* K12. Primary structure and homology within a family of aminoacyl-tRNA synthetases. J. Biol. Chem. 263:868–877, 1988.
37. Jordana, X., Chatton, B., Paz-Weisshaar, M., Buhler, J.-M., Cramer, F., Ebel, J.P., Fasiolo, F. Structure of the yeast valyl-tRNA synthetase gene *(VAS1)* and the homology of its translated amino acid sequence with *Escherichia coli* isoleucyl-tRNA synthetase. J. Biol. Chem. 262:7189–7194, 1987.
38. Brändén, C.I. Relation between structure and function of α/β proteins. Q. Rev. Biophys. 13:317–339, 1980.
39. Rao, S.T., Rossmann, M.R. Comparison of super-secondary structures in proteins. J. Mol. Biol. 76:241–256, 1973.
40. Holbrook, J.J., Liljas, A., Steindel, S.J., Rossmann, M.G. Lactate dehydrogenase. Enzymes 11:191–292, 1975.
41. Blow, D.M., Bhat, T.N., Metcalfe, A., Risler, J.L., Brunie, S., Zelwer, C. Structural homology in the amino-terminal domains of two aminoacyl-tRNA synthetases. J. Mol. Biol. 171:571–576, 1983.
42. Fersht, A.R., Shi, J.-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., Winter, G. Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314:235–238, 1985.
43. Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1:945–951, 1982.
44. Wetzel, R. Aminoacyl-tRNA families and their significance to the origin of the genetic code. Origins Life 9:39–50, 1978.
45. Doolittle, R.F. Similar amino acids: Chance or common ancestry? Science 214:149–159, 1981.
46. Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C. A model of evolutionary change in proteins. In: "Atlas of Protein Sequence and Structure, Vol. 5 Sup. 3." Dayhoff, M.O., ed. 1978:345–362, National Biomedical Research Foundation, Silver Spring, Maryland.
47. Rainey, P., Hammer-Raber, B., Kula, M.R., Holler, E. Modification of L-isoleucyl-tRNA synthetase with L-isoleucyl-bromomethyl ketone. The effect on the catalytic steps. Eur. J. Biochem. 78:239–249, 1977.
48. Rossmann, M.G., Jeffery, B.A., Main, P. Warren, S. The crystal structure of lactic dehydrogenase. Proc. Natl. Acad. Sci. USA 57:515–524, 1967.
49. Webster, T.A., Lathrop, R.H., Smith, T.F. Prediction of a common structural domain in aminoacyl-tRNA synthetases through use of a new pattern-directed inference system. Biochemistry 26:6950–6957, 1987.
50. Starzyk, R.M., Webster, T.A., Schimmel, P. Evidence for dispensable sequences inserted into a nucleotide fold. Science 237:1614–1618, 1987.
51. Ralph, W.W., Webster, T., Smith, T.F. A modified Chou and Fasman protein structure algorithm. Comput. Appl. Biosci. 3:211–216, 1987.
52. Sancar, A., Hack, A.M., Rupp, W.D. Simple method for identification of plasmid-coded proteins. J. Bacteriol. 137: 692–693, 1979.
53. Jasin, M., Regan, L., Schimmel, P. Dispensable pieces of an aminoacyl-tRNA synthetase which activate the catalytic site. Cell 36:1089–1085, 1984.
54. Cassio, D., Waller, J.P. Modification of methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin-modified enzyme. Eur. J. Biochem. 20:283–300, 1971.
55. Waye, M.M.Y., Winter, G., Wilkinson, A.J., Fersht, A.R. Deletion mutagenesis using an "M13 splint": The N-terminal structural domain of tyrosyl-tRNA synthetase *(B. stearothermophilus)* catalyses the formation of tyrosyl adenylate. EMBO J. 2:1827–1829, 1983.
56. Clarke, N.D., Lien, D.C., Schimmel, P. Evidence from cassette mutagenesis for a structure-function motif in a protein of unknown structure. Science 240:521–523, 1988.
57. Brick, P., Blow, D.M. Crystal structure of a deletion mutant of tyrosyl-tRNA synthetase complexed with *tyrosine*. J. Mol. Biol. 194:287–297, 1987.
58. Reidhaar-Olson, J.F., Sauer, R.T. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241:53–57, 1988.
59. Starzyk, R.M., Burbaum, J.J., Schimmel, P. Insertion of new sequences into the catalytic domain of an enzyme. Biochemistry 28:8479–8484, 1989.
60. Bone, R., Silen, J.L., Agard, D.A. Structural plasticity broadens the specificity of an engineered protease. Nature 339:191–195, 1989.
61. Starzyk, R.M., Schimmel, P. Construction of intra-domain chimeras of aminoacyl-tRNA synthetases. J. Biomol. Struct. Dynam. 7:225–234, 1989.
62. Riedel, H., Dull, T.J., Schlessinger, J., Ullrich, A. A chimeric receptor allows insulin to stimulate tyrosine kinase activity of epidermal growth factor receptor. Nature 324: 68–70, 1986.
63. Green, S., Chambon, P. Estradiol induction of a glucocorticoid-responsive gene by a chimeric receptor. Nature 325: 75–78, 1987.
64. Mas, M.T., Chen, C.Y., Hitzeman, R.A., Riggs, A.D. Active human-yeast chimeric phosphoglycerate kinases engineered by domain interchange. Science 233:788–790, 1986.
65. Houghton, J.E., O'Donovan, G.A., Wild, J.R. Reconstruction of an enzyme by domain substitution effectively switches substrate specificity. Nature 338:172–174, 1989.
66. Valenzuela, D., Schulman, L.H. Identification of peptide sequences at the tRNA binding site of *Escherichia coli* methionyl-tRNA synthetase. Biochemistry 25:4555–4561, 1986.
67. Schulman, L.H., Pelka, H., Leon, O. Peptides at the tRNA binding site of the crystallizable monomeric form of *E. coli* methionyl-tRNA synthetase. Nucleic Acids Res. 15: 10523–10530, 1987.
68. Bedouelle, H., Winter, G. A model of synthetase/transfer RNA interaction as deduced by protein engineering. Nature 320:371–373, 1986.
69. Jasin, M., Schimmel, P. Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. J. Bacteriol 159:783–786, 1984.
70. Toth, M.J., Schimmel, P. Internal structural features of *E. coli* glycyl-tRNA synthetase examined by subunit polypeptide chain fusions. J. Biol. Chem. 261:6643–6646, 1986.
71. Hountondji, C., Blanquet, S., Lederer, F. Methionyl-tRNA synthetase from *Escherichia coli:* Primary structure at the binding site for the 3'-end of $tRNA^{fMet}$. Biochemistry 24: 1175–1180, 1985.