Biochemistry

© Copyright 1990 by the American Chemical Society Volume 29, Number 25 June 26, 1990

Perspectives in Biochemistry

Initiation of mRNA Translation in Prokaryotes†

Claudio O. Gualerzi*,‡,§ and Cynthia L. Pon‡

Laboratory of Genetics, Department of Cell Biology, University of Camerino, Camerino (MC), 62032 Italy, and Max-Planck-Institut für Molekulare Genetik, D-1000 Berlin 33, West Germany

Received February 28, 1990

Initiation of protein synthesis consists of several interrelated steps during which the translation initiation region of the messenger RNA (mRNA TIR)[1] and the initiator tRNA (fMet-tRNA^{Metf}) are selected by the ribosome to form a ternary complex. Formation of the first peptide bond between the P-site-bound fMet-tRNA and the A-site-bound aminoacyl-tRNA specified by the second mRNA codon marks the transition from initiation to elongation. Three protein factors (initiation factors IF1, IF2, and IF3) and at least one GTP molecule are required to ensure the efficiency and fidelity of this process.

Initiation is usually the rate-controlling step of translation; under normal conditions (e.g., with the *lacZ* mRNA), elongation proceeds at a rate of ~12 amino acids/s (Sørensen et al., 1989), while the ribosomes load onto the same mRNA at ~3.2-s intervals (Kennell & Riezman, 1977). That initiation is rate-limiting is also indicated by the fact that translational efficiency can often be dramatically affected by changes in the mRNA that alter the TIRs leaving the coding sequence unmodified, while the reverse is generally not true (Gold, 1988, and references cited therein). Initiation also represents the focal point of many posttranscriptional regulatory mechanisms, so that a thorough understanding of its molecular basis and the identification of the determinants of its fidelity and efficiency are prerequisites for understanding regulation of gene expression and for optimizing the production of proteins and peptides from chimeric vectors.

In this paper, we shall review the structural and functional properties of the components involved in translation initiation (with the exception of the ribosomal subunits) and the mechanism of their mutual interactions, which eventually lead to the formation of the 70S initiation complex. We shall also discuss the mechanism of recognition between ribosomes and mRNA during initiation and the molecular determinants of translational efficiency.

Space limitation does not allow the treatment of translational regulation of gene expression. This subject, however, has been reviewed in several recent articles [e.g., Draper (1987), Stormo (1987), Gold (1988), de Smit and van Duin (1990), Dreyfus and Jacques (1990), and McCarthy and Gualerzi (1990)].

The Initiator tRNA

Escherichia coli contains a main and a minor form of initiator tRNA, namely, tRNA^{Metf1} (~75%) and tRNA^{Metf2} (~25%). Both tRNAs contain 77 nucleotides and differ in the presence of either 7-methyl-G or A at position 47 (Figure 1). The two molecules are encoded by *metZ* (tRNA^{Metf1}) and *metY* (tRNA^{Metf2}): *metZ* is present in two tandemly arranged copies at 61′ on the chromosome, while *metY* is the first gene of the *nusA–infB* operon mapping at 69′ (Nagase et al., 1988, and references cited therein). Although no functional difference between these two forms has been reported so far, the presence of two separate genes is likely to serve some regulatory function. In this connection, it is noteworthy that the *metY* promoter is less sensitive to ppGpp than is the *metZ* promoter (25% vs 65% reduction in transcription at 0.1 mM), suggesting that *metY* constitutively supplies tRNA^{Metf} under conditions leading to the stringent response (Nagase et al., 1988). Two forms of initiator tRNA differing by the inversion of a GC base pair in the T-stem have also been found in *Thermus thermophilus* (Watanabe et al., 1979), but only a single gene for tRNA^{Metf} has been found in *Bacillus subtilis*. The latter is present at 275° of the chromosome within a cluster containing 21 tRNA genes (Green & Vold, 1983).

The initiator tRNA has the same CAU anticodon (Figure 1) as the tRNA^{Met} used in elongation and is aminoacylated

† Dedicated to the fond memory of Professor Heinz-Günter Wittmann, to whom we are grateful for the opportunity of having worked for nearly 20 years on the subject of this review and whose support and encouragement we will miss.

* Address correspondence to this author.
‡ Max-Planck-Institut.
§ University of Camerino.

[1] Abbreviations: IF, initiation factor; SD, Shine–Dalgarno; TIR, translation initiation region; DMS, dimethyl sulfate; CMCT, *N*-cyclohexyl-*N*′-[2-(*N*-methylmorpholinio)ethyl]carbodiimide *p*-toluenesulfonate.

0006-2960/90/0429-5881$02.50/0 © 1990 American Chemical Society

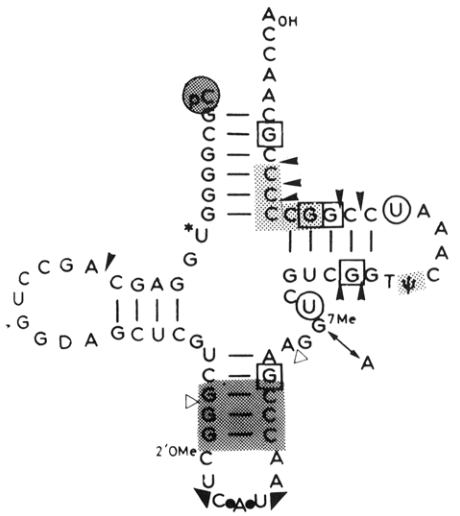

FIGURE 1: Cloverleaf structure of *E. coli* $tRNA^{Metf}$. The symbols for the modified bases are as follows: *U, 4-thiouridylic acid; D, dihydrouridylic acid; 2′OMe, 2′-*O*-methylcytidylic acid; 7MeG, 7-methylguanylic acid; T, thymidylic acid; ψ, pseudouridylic acid. The figure highlights (with dark stippling) two distinct structural features of the initiator tRNA molecule (i.e., the unpaired 5′ pC and the three consecutive GC base pairs of the anticodon stem); due to its peculiar anticodon loop conformation, the phosphodiester bonds indicated by dark triangles, which are accessible in elongator tRNAs, are resistant to S1 nuclease in $tRNA^{Metf}$. The phosphodiester bonds of the anticodon loop digested in both initiator and elongator tRNAs are marked by dark circles. The regions of fMet-$tRNA^{Metf}$ affected by its interaction with IF2 are indicated: sites where cleavage by RNase V_1 is prevented are marked with arrowhead and those where cleavage is enhanced, by open triangles; regions where the phosphates are protected from ethylnitrosourea are lightly stippled; guanines whose N7 reactivity toward DMS is increased are boxed; uracils whose reactivity toward CMCT is reduced are encircled. For further details, see the original literature quoted in the text.

by the same synthetase, which recognizes primarily the bases of the anticodon (Schulman, 1979; Pelka & Schulman, 1986, and references cited therein). Although its 3.5-Å crystal structure is overall very similar to that of an elongator tRNA, such as yeast $tRNA^{Phe}$, $tRNA^{Metf}$ is endowed with unique structural features connected with its special role in protein synthesis (Wakao et al., 1989, and references cited therein). One of these is the presence of three consecutive GC base pairs in the anticodon stem that confer rigidity and regularity to the helix and a particular conformation to the anticodon loop (Figure 1), which targets the initiator tRNA to the ribosomal P-site. Indeed, the progressive substitution of the three GC base pairs weakens binding to the ribosomal P-site while increasing accessibility of the anticodon loop to nuclease S1 (Seong & RajBhandary, 1987a).

Furthermore, the special structure of Met-$tRNA^{Metf}$ allows its specific recognition by N^{10}-formyltetrahydrofolate:Met-tRNA transformylase, which modifies the αNH_2 group (Schulman, 1979). This formylation results in a 6–10-fold stimulation of protein synthesis in *E. coli* extracts (Kung et al., 1979). fMet-tRNA interacts specifically with IF2 by virtue of its blocked αNH_2 group, and a portion of the T-stem and loop is protected from chemical and enzymatic attack in the ensuing complex (Figure 1); the binding of IF2 also results (by a long-range effect) in an increased flexibility of the anticodon arm (Wakao et al., 1989).

Finally, the absence of a Watson–Crick base pair at the end of the amino acid acceptor stem (Figure 1) is responsible for the resistance of fMet-tRNA to the action of peptidyl-tRNA hydrolase and, together with the blocking of the αNH_2 group, for its weak interaction with EFTu–GTP (Schulman, 1979;

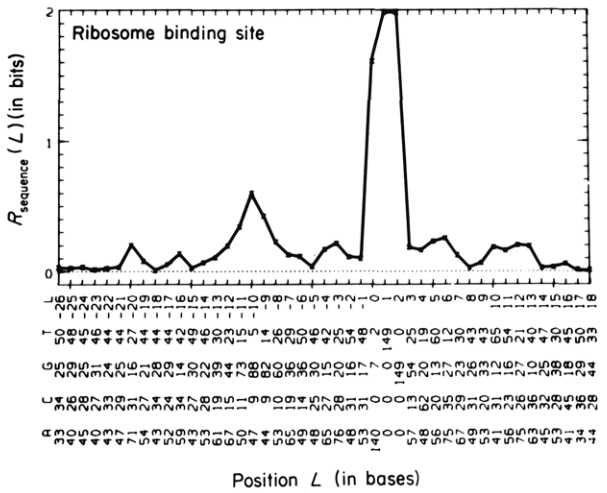

FIGURE 2: Nonrandomness in aligned ribosome-binding sites of mRNAs [from Schneider et al. (1986)].

Tanada et al., 1982; Louie et al., 1984). Single base substitutions resulting in mutants containing a base-paired 5′ nucleotide enable the initiator tRNA to be used in translation elongation (Seong & RajBhandary, 1987b). The unpaired 5′ pC is also likely to be responsible for the restricted mobility of the 3′-terminus observed by EPR spectroscopy (Pscheidt & Wells, 1986).

The Translation Initiation Region (TIR) of mRNA and the Structural Bases for Translational Efficiency

The TIR of most cistrons (≥90%) contains the initiation codon AUG; more rarely GUG (~8%), UUG (~1%), and in only one known case (i.e., *infC*) AUU serve as initiation codons. Almost invariably, the TIRs also contain a 3–9-base-long sequence [the Shine and Dalgarno (SD) sequence] complementary to part or all of the anti-SD sequence 5′-CACCUCCUU-3′ found at the 3′-end of 16S rRNA separated by a spacer of variable length (optimally 5–9 bases) from the initiation triplet (Steitz, 1980; Gren, 1984; Schneider et al., 1986).

Other distinguishing features of mRNA TIRs and their relationship to translational efficiency have been searched for by statistical analyses. In addition to the three elements mentioned above, a loose consensus sequence characterized by a strong bias for "nonstructurogenic" nucleotides was identified by Scherer et al. (1980). Subsequently, Schneider et al. (1986) identified at least five additional regions both upstream and downstream from the initiation codon in which the nucleotide sequence of bona fide translational starts is significantly different from "false" starts (Figure 2); the presence of other primary structure elements has also been correlated with efficient expression (McCarthy et al., 1985; Petersen et al., 1988; Gallie & Kado, 1989; Olins & Rangwala, 1989; Thanaraj & Pandit, 1989). None of the above elements, however, seems to be essential per se for initiation, but their interplay and the resulting TIR secondary and tertiary structures probably determine the efficiency of the recognition by ribosomes and the level of translation.

The importance of TIR primary sequence as a determinant of ribosomal recognition, regardless of its context, clearly emerges from the elegant work of Dreyfus (1988), while the importance of secondary structure has been pointed out in numerous reports [e.g., see Ganoza et al. (1987), Gold (1988), and de Smit and van Duin (1990) and references cited therein]. Only in a minority of the cases, however, have the postulated

Table I: Cloned Bacterial Genes Specifying Initiation Factors Whose Primary Structures Are Known

organism	struct gene	map loc	within operon	factor	amino acids	M_r	isoel pt[i]	ref
E. coli	*infA*	20′	no	IF1	71	8118	10.1	*a*
B. subtilis	*infA*	12°	yes	IF1	72	8213	7.8	*b*
E. coli	*infBα*	69′	yes	IF2α	890	97349	5.9	*c*
E. coli	*infBβ*	69′	yes	IF2β	733	79713	5.6	*c*
B. stearothermophilus	*infB*	?	?	IF2	742	82043	7.0	*d*
S. faecium	*infB*	?	?	IF2	785	86415	7.9	*e*
B. subtilis	*infBα*	145°	?	IF2α	716	78600	5.4	*f*
B. subtilis	*infBβ*	145°	?	IF2β	623	68182	4.8	*f*
E. coli	*infC*	38′	yes	IF3	180	20548	10.3	*g*
B. stearothermophilus	*infC*	?	yes	IF3	171	19678	10.7	*h*

[a] Sands et al., 1987. [b] Boylan et al., 1989. [c] Sacerdot et al., 1984. [d] Brombach et al., 1986. [e] Friedrich et al., 1988. [f] Shazand et al., 1990. [g] Sacerdot et al., 1982. [h] Pon et al., 1989. [i] The values were calculated by using the UWGCG program (Devereux et al., 1984).

secondary structures been experimentally verified. At the risk of oversimplifying the issue, it can be said that the primary structure of the TIR has a 2-fold function: to provide sequence-related features for the specific recognition by ribosomal components and to dictate the (presence or absence of) secondary structures that affect (negatively or positively) the efficiency of initiation.

Especially intriguing is the influence that the nature of the initiation triplet (AUG, GUG, UUG, or AUU) has on the level of translational expression. In most (but not all) cases, changing the rare initiation triplet into the more common AUG results in moderate (Reddy et al., 1985; Khudyakov et al., 1988) to large (Brombach & Pon, 1987) increases of expression. Nonetheless, inspection of the catalogue of genes having the rare initiation triplets strongly argues against the idea that the cell uses these codons to attain a substantial reduction in the level of translation. Instead, a likely reason for the cell to use the rare initiation codons is that they serve as targets for regulatory mechanisms aimed at select genes. This concept is best illustrated by the case of the AUU triplet found in *infC* of both *E. coli* (Sacerdot et al., 1982) and *Bacillus stearothermophilus* (Pon et al., 1989). When this AUU is changed to AUG, in vivo expression of both genes is increased about 40-fold because, confirming Gold's intuition (Gold et al., 1984), autorepression by IF3 is lost (Butler et al., 1987). On the other hand, the expression of mRNAs (different from that of IF3) attains comparable levels in vitro regardless of whether the initiation triplet is AUG or AUU. The IF3 autorepression requires a small excess of free IF3 molecules and is specifically aimed at the dissociation of 30S initiation complexes containing mRNAs with an AUU initiation triplet (A. La Teana, C. L. Pon, and C. O. Gualerzi, unpublished results).

The Initiation Factors: Structural, Evolutionary, and Genetic Aspects

Several bacterial genes encoding IFs have been isolated and cloned; their map locations (when available) and some structural characteristics of their products are listed in Table I.

The three IFs do not share any structural homology and appear to be evolutionarily conserved: IF1 of *E. coli* shares 69% identical residues with IF1 of *B. subtilis*, while IF3 of *E. coli* is 50% identical with IF3 of *B. stearothermophilus*. The case of IF2 is more intriguing; all the sequences known so far display a large degree of homology in the C-terminal two-thirds of the molecule, but their N-terminal parts hardly resemble each other (Figure 3), suggesting that this part may have some species-specific regulatory function probably not related to translation. This idea is supported by the finding that even after long deletions from the 5′-terminus (hatched area in Figure 3), the *infB* genes encode IF2 fragments fully active in all the basic translation activities, while internal deletions (e.g., like that shown in Figure 3) result in the loss of specific translational functions (Pon & Gualerzi, 1988; Severini et al., 1990). In addition to IF2α, *infB* of *E. coli* (and probably of *B. subtilis*) also expresses a shorter protein (IF2β) resulting from a low-rate initiation event at an in-frame GUG codon downstream from the main initiation start site (Plumbridge et al., 1985). Both the mechanism and the significance of this occurrence remain obscure.

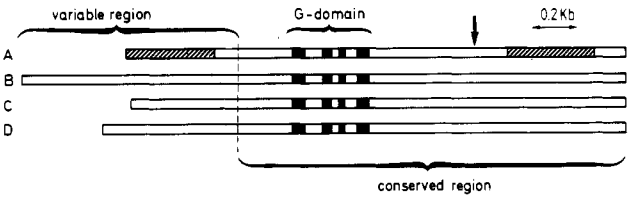

FIGURE 3: Schematic representation of *infB* from (A) *B. stearothermophilus*, (B) *E. coli* (IF2α), (C) *E. coli* (IF2β), and (D) *Streptococcus faecium*. The darkened areas represent the conserved structural elements of the GTP-binding site. The hatched areas represent the deletions introduced by genetic manipulations, and the arrow indicates the point of proteolytic cleavage that separates the G from the C domain (Pon & Gualerzi, 1988; Severini et al., 1990).

Several attempts to crystallize initiation factors have failed so far. However, the small size of IF1 and its exceptionally well-resolved ^{1}H NMR spectrum have prompted an attempt to determine its solution structure by 2D NMR spectroscopy; IF1 was found to possess a complex structure consisting of extensive β-sheet motifs in parallel and antiparallel orientations, three β-turns, and two short α-helices, one near the N-terminus and the other near the middle of the molecule (M. Paci, R. Boelens, and R. Kaptein, personal communication).

IFs–Ribosome Interaction: Thermodynamic, Structural, and Topographical Aspects

Quantitative data on the association constants and stoichiometry of the interactions between IFs and ribosomes (Table II) have been obtained by fluorescence polarization studies (Weiel & Hershey, 1981, 1982; Zucker & Hershey, 1986) and by Airfuge centrifugation (Pon et al., 1985; Celano et al., 1988). The results are in fairly good agreement and indicate that the 30S subunit has a single high-affinity site for each factor. IF1 and IF3 show negligible affinity for 50S subunits and for 70S monomers (they are actually ejected from the 30S subunit upon subunit association). IF2, on the other hand, binds with a fairly high affinity to the large subunit and also to the 70S monomers from which it is released upon GTP hydrolysis.

The interactions with the 30S subunit display different properties. The binding of IF3 is only slightly affected by IF1

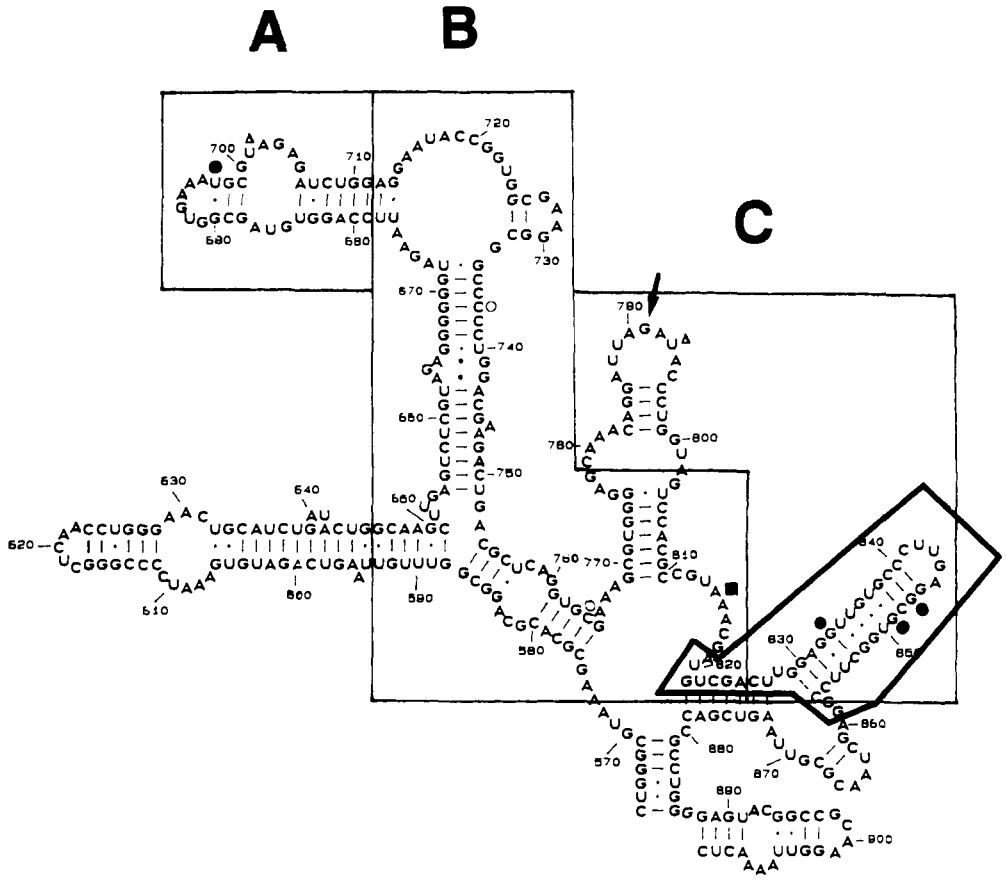

FIGURE 4: Secondary structure of the central region of 16S rRNA. The segment enclosed in the darker frame is that found cross-linked to IF3 by Ehresmann et al. (1986), and the arrow indicates the position of the G → A transition resulting in a decrease of IF3 binding (Tapprich et al., 1989). The sites of enhanced (darkened symbols) or reduced (open symbols) reactivity or cleavage in the presence of IF3 are indicated as follows: hydrolysis by RNase V_1 (○, ●) and reaction with DMS (■) and CMCT (△) [modified from Muralikrishna and Wickstrom (1989)].

and IF2, while the binding of IF1 or IF2 is strongly stimulated by the other two factors. Furthermore, increasing the ionic strength (e.g., from 50 to 150 mM NH_4Cl) reduces 15-fold and 20-fold the K_a of the IF2–30S and IF1–30S complexes, respectively, but decreases only 2-fold the K_a of the IF3–30S complex. All interactions are relatively unaffected by changes in the Mg^{2+} concentration but are drastically weakened by hydrostatic pressure, especially those of IF1 and IF2. It has been suggested that the IF1–30S interaction is an entropy-driven process triggered mainly by the release of counterions from the RNA phosphates and involving a minimum of 2.7–3.6 ion pairs. In the case of IF2 and IF3, ionic and hydrophobic interactions seem to be equally important for ribosomal binding.

Considering the estimated intracellular concentrations of ribosomes (~10 μM) and IFs (~1 μM each) and the K_a's of their interactions, it can be argued that in the cell nearly all the IFs are bound to the native 30S subunits.

The identification of some amino acid residues of the IFs involved in the interactions with the 30S subunit has been obtained by chemical modifications and ^1H NMR spectroscopy (Gualerzi et al., 1986, and references cited therein) and, more recently, by genetic manipulations and site-directed mutagenesis. Thus, Arg 69 plays an important role in the binding of IF1 to the 30S subunit, while the deprotonation of His 29 is apparently involved in the release of the factor from the 30S subunit upon subunit association (Gualerzi et al., 1989). In the case of IF2, two structurally compact and functionally separate domains have been identified: a C-terminal domain containing the fMet-tRNA binding site and a central domain, the G domain, containing the 50S and GTP binding sites as well as the GTPase catalytic center (Spurio et al., 1989).

The topographical localization of the IFs on the ribosome has been investigated by standard protein–protein and protein–RNA cross-linking and, in the case of IF3, visualized by immunoelectron microscopy.

A group of eight ribosomal proteins (S1, S7, S11–13, S18, S19, and S21) has consistently been found cross-linked in significant yield to more than one IF or to the same factor by more than one laboratory, and cross-links between IF1–IF2, IF2–IF3, and IF2–L7/L12 have also been found (Boileau et al., 1983, and references cited therein).

Of the three factors, only IF3 can be efficiently cross-linked to rRNA. The cross-linked factor is found unequally distributed between two 16S rRNA regions that are probably close to each other in the 30S subunit. The major cross-linking site has been localized between residues 819 and 859 in the central domain of the rRNA (Figure 4), while the minor one has been identified between nucleotides 1506 and 1529 (Ehresmann et al., 1986, and references cited therein).

Consistent with this localization is the finding that IF3 reduces the chemical reactivity and enzymatic accessibility of some nucleotides and enhances attack at others in the central (nucleotides 690–850) domain of 16S rRNA. Here three subdomains (A, B, and C of Figure 4) can be distinguished by the effects produced by IF3: domains A and C are either structurally sequestered or stabilized, while domain B is destabilized and more exposed in the presence of the factor (Muralikrishna & Wickstrom, 1989). Finally, mutation of guanine to adenine at position 791 (Figure 4) impairs the association of the 30S subunits with the 50S subunits and increases 10-fold the dissociation rate constant of the IF3–30S complex without altering the on-rate, so that the K_a is reduced by approximately 1 order of magnitude. These results suggest

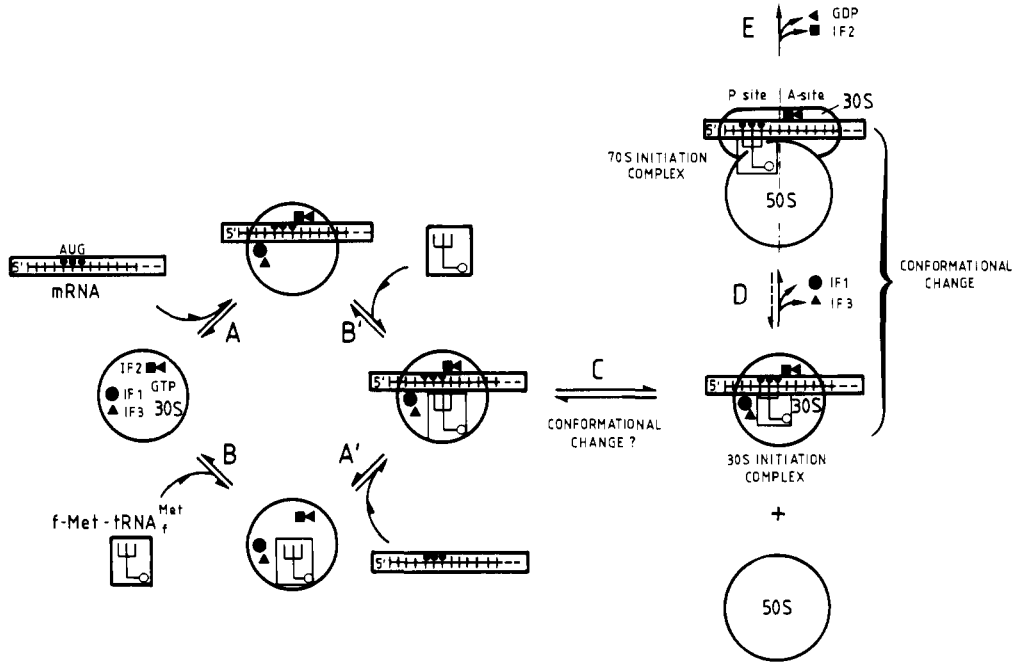

FIGURE 5: Simplified mechanistic model of translation initiation. Steps A, A', B, and B' are in rapid equilibrium. Step C represents the first-order, rate-limiting rearrangement of the ternary preinitiation complex kinetically controlled in both directions by the IFs. Step D represents the virtually irreversible subunit association giving rise to the 70S initiation complex [modified from Gualerzi and Pon (1981)].

Table II: Association Constants and Stoichiometry of 30S–Initiation Factor Interaction[a]

complex	K_a (M^{-1}) [stoichiometry (n)]	
	no other factors present	two other factors present
30S–IF1	0.9×10^6 [1.3][b]	1.0×10^8 [0.8][b]
	0.2×10^7 [0.5][c]	9.1×10^7 [0.5][c]
30S–IF2	3.5×10^7[b]	1.8×10^8 [0.75][b]
	1.5×10^7 [0.7][c]	
	3.7×10^7 [0.7][c,d]	
	5.7×10^7 [0.6][e]	
30S–IF3	4.3×10^7 [1.4][b]	

[a] For references, see text. [b] Determined by fluorescence polarization at 10 mM Mg^{2+} and 50 mM NH_4^+. [c] Determined by Airfuge centrifugation at 7.5 mM Mg^{2+} and 50 mM NH_4^+. [d] In the presence of 1 mM GTP. [e] Determined by fluorescence polarization at 5 mM Mg^{2+} and 100 mM NH_4^+.

that IF3 recognizes a specific structure on the 30S ribosomal subunit that includes G791 to stabilize its interaction (Tapprich et al., 1989).

Combining the above information with what is known concerning the topographical organization of the 30S ribosomal subunit [e.g., see Brimacombe (1988) and Stern et al. (1988)], we can draw a relatively unambiguous albeit low-resolution picture of the IFs' binding sites on the ribosome: IF3, having been found cross-linked to elements of both sides, most likely bridges the cleft between head and platform, while IF1 should be located in the neck region and IF2 more toward the head, between IF1 and IF3. A topographical model in which IF2 binds close to both IF1 and IF3 but the latter two factors are not in contact with each other is consistent with the cross-linking data and also with some of the above-mentioned binding studies (Zucker & Hershey, 1986).

The topographical localization of the IFs is also consistent with their role of influencing the kinetics of codon–anticodon interaction at the P-site (see below), which is believed to occur at the bottom of the cleft (Ofengand et al., 1986). Finally, the localization of IF3 at or near bases (e.g., G791) implicated in subunit association (Herr et al., 1979; Tapprich & Hill, 1986) may underlie the well-known subunit antiassociation activity of IF3 [see Hershey (1987) and references cited therein].

The Pathway of Translation Initiation

The probable events occurring during initiation are depicted in Figure 5. This mechanistic model is suggested by kinetic analyses of the formation of 30S and 70S initiation complexes and is compatible with the relevant data available [for review, see Gualerzi et al. (1988)]. Mutatis mutandis, the proposed mechanism is similar to that suggested by McClure (1985) for initiation of transcription.

When the concentrations of mRNA and fMet-tRNA are saturating, the 30S subunit (bearing one molecule each of the three IFs) binds these ligands in random order to form a ternary preinitiation complex in which the two ligands are not yet interacting. A rate-limiting first-order rearrangement, kinetically controlled by the IFs, promotes codon–anticodon interaction at the ribosomal P-site and the formation of the 30S initiation complex. This complex either dissociates into its original components or binds a 50S subunit, giving rise to the 70S initiation complex. Due to the concomitant ejection of IF1 and IF3, formation of the 70S initiation complex is virtually irreversible, while IF2 promotes the positioning of fMet-tRNA in the ribosomal P-site. The ejection of IF2 from the 70S initiation complex is accompanied by the factor- and ribosome-dependent hydrolysis of GTP.

Function of the Initiation Factors

The molecular basis for IF3 activity is probably a change in the conformational dynamics of the 30S subunits induced by the binding of the factor (Pon et al., 1982, and references cited therein), while the basis for IF2 activity probably resides in its capacity to bind fMet-tRNA, to undergo conformational change, and to induce conformational changes of both initiator tRNA and ribosomes (Pon et al., 1985; Zucker & Hershey, 1986; Canonaco et al., 1989; Wakao et al., 1989; our unpublished observations). The mechanism of action of IF1 remains completely obscure.

The main functions of the IFs are summarized in Table III. As mentioned above, the key function of the IFs is to affect

Table III: Main Properties and Functions of Prokaryotic Initiation Factors

factor	properties and function
IF1	binds to 30S subunits, is ejected upon subunit association, increases the affinity of the 30S for IF2 and somewhat for IF3, and stimulates IF2 and IF3 activities
IF2	G-protein: binds to 30S subunits (optimally with GTP), increases the affinity of the 30S subunit for IF1, and contains a binding site for fMet-tRNA; kinetic effector of 30S (and 70S) initiation complex formation and increases on-rate (favoring binding of aminoacyl-tRNAs with blocked αNH_2 groups) and reduces dissociation rate constants; upon subunit association interacts with 50S subunits activating ribosome-dependent GTPase, positions fMet-tRNA in P-site, and is ejected from ribosomes; kinetic proofreading?
IF3	binds to 30S subunits and is ejected upon subunit association; increases the affinity of the 30S for IF1 and IF2; kinetic effector and fidelity factor of 30S initiation complex formation and increases both on- and off-rates of its formation favoring dissociation of noninitiator aminoacyl-tRNAs; subunit antiassociation factor

kinetically the formation and dissociation of the codon–anticodon interaction at the P-site of the 30S subunit and, ultimately, to influence which and how many 30S initiation complexes enter the elongation cycle after association with the 50S subunit. All three factors (but IF1 only in combination with the other two) stimulate the on-rate of 30S initiation complex formation. The effect of IF2 is much larger with aminoacyl-tRNAs having a blocked αNH_2 group but is observable also with aminoacyl-tRNAs that do not form binary complexes with IF2 off the ribosomes; thus the suggestion that this factor promotes the recognition and ribosomal binding of fMet-tRNA through a carrier mechanism analogous to that of EFTu does not seem to be entirely supported by the experimental evidence (Gualerzi et al., 1986, 1988, and references cited therein).

Furthermore, regardless of the type of aminoacyl-tRNA bound, the dissociation rate of the 30S complexes and the rate of exchange between free and 30S-bound aminoacyl-tRNAs are substantially lowered by IF2. IF3, on the other hand, produces a large increase in both dissociation and exchange rates, especially when the bound aminoacyl-tRNA is different from fMet-tRNA. In conclusion, it appears that IF2 and IF3 accelerate the locking and unlocking of the codon–anticodon interaction at the P-site while favoring the formation of the correct over the incorrect and the dissociation of the incorrect over the correct 30S complexes. The role of IF3 as a fidelity factor had been recognized a long time ago and was traced to the recognition by the factor of some specific feature of the initiator tRNA molecule (Gualerzi & Pon, 1981, and references cited therein) but received little attention until Gold's laboratory demonstrated this activity in toeprinting experiments and provided evidence that IF3 "inspects" the correctness of both the anticodon stem of the tRNA and the P-site codon–anticodon interaction (Hartz et al., 1989) (Figure 6).

The IFs also influence in different ways the equilibrium 70S $\rightleftharpoons$ 30S + 50S: by binding to the free 30S subunits, IF3 shifts the equilibrium to the right, thus providing a continuous supply of native 30S subunits to feed the initiation process; IF1 increases the rate of exchange between ribosomal subunits; and IF2 favors subunit association [see Hershey (1987) and references cited therein].

Finally, as mentioned above, IF2 is endowed with an IF2- and ribosome-dependent GTPase activity. At least three nonmutually exclusive physiological functions for this activity can be envisaged: (a) to promote the ejection of IF2 from the ribosome, (b) to position fMet-tRNA in its final transpeptidation-competent ribosomal binding site, and (c) to increase further the fidelity of initiation by driving a kinetic proofreading mechanism.

Ribosome–mRNA Interaction and Initiation Factor Function

The idea that the IFs, IF3 in particular, promote or stimulate (natural) mRNA–ribosome interaction and influence the

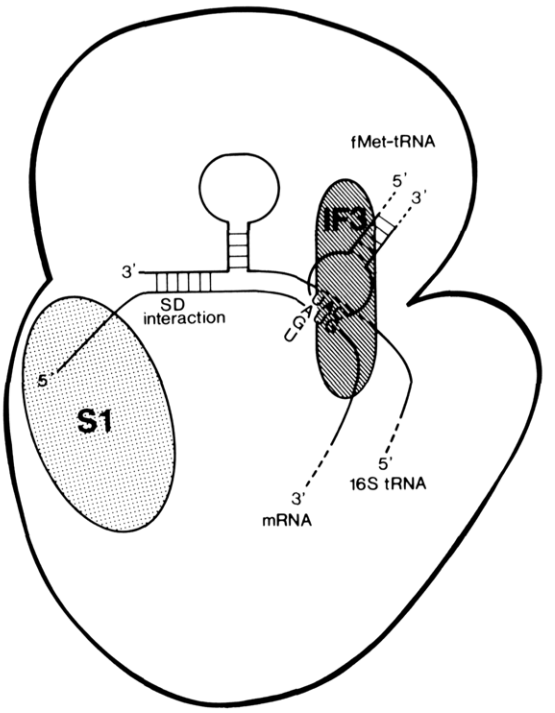

FIGURE 6: mRNA–30S interaction at initiation. This diagram highlights the three main mRNA–ribosome interactions (the SD–anti-SD and the codon–anticodon base pairings and the mRNA–S1 interaction) that dictate translational efficiency. The figure also presents a visual summary of the results obtained in the laboratory of L. Gold concerning the structural elements of initiator tRNA and the portion of the codon–anticodon interaction at the P-site inspected by IF3 to ensure fidelity of translation initiation (see text for details).

SD base pairing remained deeply rooted for a long time, in spite of strong circumstantial evidence to the contrary (Gualerzi & Pon, 1981, and references cited therein).

Direct evidence that neither IF3 nor the other IFs influence the SD interaction and the affinity of the 30S subunit for mRNAs with or without SD sequence has been obtained from several types of binding experiments (Calogero et al., 1988; Canonaco et al., 1989). In agreement with these findings, Laughrea and Tam (1990), using model mDNAs and mRNAs having 4–8-nucleotide-long SD sequences and primary structures similar to the R17 coat protein initiation site, could not detect any effect of the IFs (nor of ribosomal protein S1) on the recognition, binding, or exchange rate of mRNAs on the 30S subunit.

Although the IFs have no detectable effect on the K_a's of the binary complexes between 30S subunits and mRNAs, their presence may well influence the position of the mRNA in its ribosomal binding site. Thus, it has been suggested that, in the absence of IFs, the mRNA preferentially occupies a ribosomal "standby site" corresponding to the region where the SD interaction takes place. In the presence of factors, the mRNA is shifted toward another ribosomal site, possibly closer

to the ribosomal P-site where the IFs exert their kinetic influence on codon–anticodon interaction (Canonaco et al., 1989).

Selection of mRNA TIR by Ribosomes

After recalling that the access of ribosomes to noninitiation regions of the mRNA may be restricted by secondary structure constraints [e.g., see Ganoza et al. (1987)], we can ask what are the positive elements governing mRNA–ribosome interaction at initiation and dictating the fidelity and efficiency of the ensuing translational process.

The ribosome is believed to accommodate the mRNA in a "U-shaped" channel or trough (Evstafieva et al., 1983; Olson et al., 1988), probably constituted by structural elements contributed by several ribosomal proteins and stretches of 16S rRNA, so that each mRNA, depending on its structure, binds to the 30S subunit by means of several alternative, more or less specific interactions with both 16S rRNA and ribosomal proteins. Although our knowledge of ribosome structure is still too rudimentary to allow a detailed understanding of these interactions, at least two 30S proteins (i.e., S1 and S21) have consistently been implicated in mRNA binding, and it has recently been demonstrated that the oligo(U)-containing "translational enhancers" found in some mRNA TIRs may be specifically recognized and bound by S1 (I. Boni, USSR Academy of Science, personal communication).

In addition to these interactions, specific base pairing can direct the mRNA TIR to its ribosomal binding site. Base pairing in vivo between the mRNA SD sequence and anti-SD sequence of the 16S rRNA (Figure 6) has been elegantly demonstrated (Hui & de Boer, 1987; Jacob et al., 1987). In vitro, however, mRNAs with and without the SD sequence are translated at the same rate and with the same dependence on IFs, provided that the mRNA concentration is properly selected. The SD interaction is also mechanistically irrelevant for 30S initiation complex formation and for the selection of the mRNA reading frame (Calogero et al., 1988). In agreement with these data, Melançon et al. (1990) showed that 30S mutants lacking the anti-SD sequence are able to initiate translation of natural mRNAs at the correct start sites and concluded that neither translational efficiency (in vitro) nor selection of translational start is solely controlled by the SD interaction. Due to the competition of several mRNA TIRs for a limited number of 30S subunits, however, the influence of the SD interaction on translational efficiency is probably greater in vivo than in vitro. In conclusion, the available data support the suggestion that the in vivo function of the SD interaction is to ensure millimolar concentrations of a potential initiation triplet near the ribosomal P-site (Calogero et al., 1988).

As mentioned above, additional types of base pairing between mRNA and 16S rRNA have also been suggested and correlated with either regulatory mechanisms (Gold et al., 1984) or high levels of translational expression (Petersen et al., 1988; Gallie & Kado, 1989; Olins & Rangwala, 1989; Thanaraj & Pandit, 1989). The occurrence of any other type of base pairing aside from the SD interaction, however, remains to be demonstrated.

Are these thermodynamic-based mechanisms sufficient to explain the specificity of the mRNA selection process and to account for the different levels of translational efficiency encountered in nature? At least in some cases the affinity of the mRNAs for the ribosome either does not correlate (Calogero et al., 1988) or correlates only qualitatively (Lang et al., 1989) with the level of their expression, indicating that an additional mechanism must exist, possibly superimposed on the first. We suggest that this mechanism is based on the kinetic selection of the "best-fit" 30S initiation complexes from among the multitude of 30S ternary complexes that may form in the cell. Thus, the native 30S subunits may be expected, in principle, to bind any aminoacyl-tRNA (or peptidyl-tRNAs, if these are not completely scavenged by peptidyl-tRNA hydrolase) to any triplet (not necessarily cognate) happening to be in the P-site; the stability of these complexes would vary, however, depending on the nature and stereospecificity of the mRNA–ribosome and codon–anticodon interactions (Potapov, 1982). According to the mechanistic scheme presented above (Figure 5), each potential 30S initiation complex can either dissociate into its individual components or become fixed in a 70S initiation complex and enter the elongation cycle. Thus, to be translationally productive, each interaction must pass through the kinetic screen set by the IFs at step C of the initiation pathway. According to this model, both selection of the correct initiation start and translational efficiency depend on the ratio between the (presumably constant) on-rate of 70S initiation complex formation and the (variable) rate of dissociation of the individual 30S complexes.

In addition to offering a unitary explanation for both fidelity and efficiency of translation initiation, this kinetic model provides a simple rationale for why initiation obligatorily begins with the 30S subunit (Blumberg et al., 1979).

References

Blumberg, B. M., Nakamoto, T., & Kezdy, F. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 251–255.

Boileau, G., Butler, P., Hershey, J. W. B., & Traut, R. R. (1983) *Biochemistry 22*, 3162–3170.

Boylan, S. A., Suh, J. W., Thomas, S. M., & Price, C. (1989) *J. Bacteriol. 171*, 2553–2562.

Brimacombe, R. (1988) *Biochemistry 27*, 4207–4214.

Brombach, M., & Pon, C. L. (1987) *Mol. Gen. Genet. 208*, 94–100.

Brombach, M., Gualerzi, C. O., Nakamura, Y., & Pon, C. L. (1986) *Mol. Gen. Genet. 205*, 97–102.

Butler, J. S., Springer, M., & Grunberg-Manago, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4022–4025.

Calogero, R. A., Pon, C. L., Canonaco, M. A., & Gualerzi, C. O. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6427–6431.

Canonaco, M. A., Gualerzi, C. O., & Pon, C. L. (1989) *Eur. J. Biochem. 182*, 501–506.

Celano, B., Pawlik, R. T., & Gualerzi, C. O. (1988) *Eur. J. Biochem. 178*, 351–355.

de Smit, M. H., & van Duin, J. (1990) *Prog. Nucleic Acids Res. Mol. Biol. 38*, 1–35.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Draper, D. (1987) in *Translational Regulation of Gene Expression* (Ilan, J., Ed.) pp 1–26, Plenum Press, New York.

Dreyfus, M. (1988) *J. Mol. Biol. 204*, 79–94.

Dreyfus, M., & Jacques, N. (1990) *Mol. Microbiol.* (in press).

Ehresmann, C., Moine, H., Mougel, M., Dondon, J., Grunberg-Manago, M., Ebel, J. P., & Ehresmann, B. (1986) *Nucleic Acids Res. 14*, 4803–4821.

Evstafieva, A. G., Shatsky, I. N., Bogdanov, A. A., Semenkov, Y. P., & Vasiliev, V. D. (1983) *EMBO J. 2*, 799–804.

Friedrich, K., Brombach, M., & Pon, C. L. (1988) *Mol. Gen. Genet. 214*, 595–600.

Gallie, D. R., & Kado, C. I. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 129–132.

Ganoza, M. C., Kofoid, E. C., Marlière, P., & Louis, B. G. (1987) *Nucleic Acids Res. 15*, 345–360.

Gold, L. (1988) *Annu. Rev. Biochem. 57*, 199–233.

Gold, L., Stormo, G., & Saunders, R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7061–7065.

Green, C. J., & Vold, B. S. (1983) *Nucleic Acids Res. 11*, 5763–5774.

Gren, E. J. (1984) *Biochimie 66*, 1–29.

Gualerzi, C., & Pon, C. L. (1981) in *Structural Aspects of Recognition and Assembly in Biological Macromolecules* (Balaban, M., Sussman, J. L., Traub, W., & Yonath, A., Eds.) pp 805–826, ISS Press, Rehovot, Israel, and Philadelphia, PA.

Gualerzi, C. O., Pon, C. L., Pawlik, R. T., Canonaco, M. A., Paci, M., & Wintermeyer, W. (1986) in *Structure, Function and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) pp 621–641, Springer-Verlag, New York.

Gualerzi, C. O., Calogero, R. A., Canonaco, M. A., Brombach, M., & Pon, C. L. (1988) in *Genetics of Translation* (Tuite, M. F., Picard, M., & Bolotin-Fukuhara, M., Eds.) pp 317–330, Springer-Verlag, Berlin and Heidelberg, West Germany.

Gualerzi, C., Spurio, R., La Teana, A., Calogero, R., Celano, B., & Pon, C. (1989) *Protein Eng. 3*, 133–138.

Hartz, D., McPheeters, D. S., & Gold, L. (1989) *Genes Dev. 3*, 1899–1912.

Herr, W., Chapman, N. M., & Noller, H. F. (1979) *J. Mol. Biol. 30*, 433–449.

Hershey, J. W. B. (1987) in *Escherichia coli and Salmonella typhimurium—Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., & Umbarger, H. E., Eds.) pp 613–647, American Society for Microbiology, Washington, DC.

Hui, A., & DeBoer, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4762–4766.

Jacob, W. F., Santer, M., & Dahlberg, A. E. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4757–4761.

Kennell, D., & Riezman, H. (1977) *J. Mol. Biol. 114*, 1–21.

Khudyakov, Y. E., Neplyueva, V. S., Kalinina, T. I., & Smirnov, V. D. (1988) *FEBS Lett. 232*, 369–371.

Kung, H. F., Eskin, B., Redfield, B., & Weissbach, H. (1979) *Arch. Biochem. Biophys. 195*, 396–400.

Lang, V., Gualerzi, C., & McCarthy, J. E. G. (1989) *J. Mol. Biol. 210*, 659–663.

Laughrea, M., & Tam, J. (1990) *Biochem. Cell Biol. 67*, 812–817.

Louie, A., Ribeiro, N. S., Reid, B. R., & Jurnak, F. (1984) *J. Biol. Chem. 259*, 5010–5016.

McCarthy, J. E. G., & Gualerzi, C. O. (1990) *Trends Genet. 6*, 78–85.

McCarthy, J. E. G., Schairer, H. U., & Sebald, W. (1985) *EMBO J. 4*, 519–526.

McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171–204.

Melançon, P., Leclerc, D., Destroismaisons, N., & Brakier-Gingras, L. (1990) *Biochemistry 29*, 3402–3407.

Muralikrishna, P., & Wickstrom, E. (1989) *Biochemistry 28*, 7505–7510.

Nagase, T., Ishii, S., & Imamoto, F. (1988) *Gene 67*, 49–57.

Ofengand, J., Ciesiolka, J., Denman, R., & Nurse, K. (1986) in *Structure, Function and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) pp 472–494, Springer-Verlag, New York.

Olins, P. O., & Rangwala, S. H. (1989) *J. Biol. Chem. 264*, 16973–16976.

Olson, H. M., Lasater, L. S., Cann, P. A., & Glitz, D. G. (1988) *J. Biol. Chem. 263*, 15196–15204.

Pelka, H., & Schulman, L. (1986) *Biochemistry 25*, 4450–4456.

Petersen, G. B., Stockwell, P. A., & Hill, F. D. (1988) *EMBO J. 7*, 3957–3962.

Plumbridge, J. A., Deville, F., Sacerdot, C., Petersen, H. U., Cenatiempo, Y., Cozzone, A., Grunberg-Manago, M., & Hershey, J. W. B. (1985) *EMBO J. 4*, 223–229.

Pon, C. L., & Gualerzi, C. O. (1988) in *Gene Expression and Regulation: The Legacy of Luigi Gorini* (Bissell, M., Dehò, G., Sironi, G., & Torriani, A., Eds.) pp 137–150, Elsevier Science Publishers, Amsterdam, The Netherlands.

Pon, C. L., Pawlik, R. T., & Gualerzi, C. (1982) *FEBS Lett. 137*, 163–167.

Pon, C. L., Paci, M., Pawlik, R. T., & Gualerzi, C. O. (1985) *J. Biol. Chem. 260*, 8918–8924.

Pon, C. L., Brombach, M., Thamm, S., & Gualerzi, C. (1989) *Mol. Gen. Genet. 218*, 355–357.

Potapov, A. P. (1982) *FEBS Lett. 146*, 5–8.

Pscheidt, R. H., & Wells, B. D. (1986) *J. Biol. Chem. 261*, 7253–7256.

Reddy, P., Peterkofsky, A., & McKenney, K. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 5656–5660.

Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J. A., Grunberg-Manago, M., & Blanquet, S. (1982) *EMBO J. 1*, 311–315.

Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A., & Grunberg-Manago, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7787–7791.

Sands, J. F., Cummings, H. S., Sacerdot, C., Dondon, L., Grunberg-Manago, M., & Hershey, J. W. B. (1987) *Nucleic Acids Res. 15*, 5157–5168.

Scherer, G. F. E., Walkinshaw, M. D., Arnott, S., & Morre, D. J. (1980) *Nucleic Acids Res. 8*, 3895–3907.

Schneider, T. D., Stormo, G. D., Gold, L., & Ehrenfeucht, A. (1986) *J. Mol. Biol. 188*, 415–431.

Schulman, L. H. (1979) in *Transfer RNA: Structure, Properties and Recognition* (Schimmel, P. R., Söll, D., & Abelson, J. N., Eds.) pp 311–324, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Seong, B. L., & RajBhandary, U. L. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 334–338.

Seong, B. L., & RajBhandary, U. L. (1987b) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8859–8863.

Severini, M., Choli, T., La Teana, A., Pon, C., & Gualerzi, C. (1990) *J. Protein Chem.* (in press).

Shazand, K., Tucker, J., Chiang, R., Stansmore, K., Sperling-Petersen, H. U., Grunberg-Manago, M., Rabinowitz, J. C., & Leighton, T. (1990) *J. Bacteriol. 172*, 2675–2687.

Sørensen, M. A., Kurland, C. G., & Pedersen, S. (1989) *J. Mol. Biol. 207*, 365–377.

Spurio, R., La Teana, A., Severini, M., Gambacurta, A., Choli, T., Falconi, M., Gualerzi, C., & Pon, C. (1989) *Atti Assoc. Genet. Ital. 35*, 349–350.

Steitz, J. A. (1980) in *Ribosomes: Structure, Function, and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., & Nomura, M., Eds.) pp 479–495, University Park Press, Baltimore, MD.

Stern, S., Weiser, B., & Noller, H. F. (1988) *J. Mol. Biol. 204*, 447–481.

Stormo, G. D. (1987) in *Translational Regulation of Gene Expression* (Ilan, J., Ed.) pp 27–49, Plenum Press, New York.

Tanada, S., Kawakami, M., Nishio, K., & Takemura, S. (1982) *J. Biochem. (Tokyo) 91*, 291–299.

Tapprich, W. E., & Hill, W. E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 556–560.
Tapprich, W. E., Goss, D., & Dahlberg, A. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4927–4931.
Thanaraj, T. A., & Pandit, M. W. (1989) *Nucleic Acids Res. 17*, 2973–2985.
Wakao, H., Romby, P., Westhof, E., Laalami, S., Grunberg-Manago, M., Ebel, J. P., Ehresmann, C., & Ehresmann, B. (1989) *J. Biol. Chem. 264*, 20363–20371.
Watanabe, K., Kuchino, Y., Yamaizumi, Z., Kato, M., Oshima, T., & Nishimura, S. (1979) *J. Biochem. (Tokyo) 86*, 893–905.
Weiel, J., & Hershey, J. W. B. (1981) *Biochemistry 20*, 5859–5865.
Weiel, J., & Hershey, J. W. B. (1982) *J. Biol. Chem. 257*, 1215–1220.
Zucker, F. H., & Hershey, J. W. B. (1986) *Biochemistry 25*, 3682–3690.

Articles

Structure of the Aspartic Protease from Rous Sarcoma Retrovirus Refined at 2-Å Resolution†

Mariusz Jaskólski,‡,§ Maria Miller,‡ J. K. Mohana Rao,‡ Jonathan Leis,‖ and Alexander Wlodawer*,‡
Crystallography Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, P.O. Box B, Frederick, Maryland 21701, and Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106
Received December 21, 1989; Revised Manuscript Received March 20, 1990

ABSTRACT: The structure of Rous sarcoma virus protease has been solved by multiple isomorphous replacement in the crystal form belonging to space group $P3_121$, with unit-cell parameters a = 88.95 Å and c = 78.90 Å. The enzyme belongs to the family of aspartic proteases with two identical subunits composing the active homodimer. The noncrystallographic dyad relating these two subunits was identified after preliminary tracing in the MIR map and was used for phase improvement by electron-density averaging. Structure refinement resulted in a model that included 1772 protein atoms and 252 water molecules, with an R factor of 0.144 for data extending to 2 Å. The secondary structure of a retroviral protease molecule closely resembles that of a single domain in pepsin-like aspartic proteases and consists of several β-strands and of one well-defined and one distorted α-helix. The dimer interface is composed of the N- and C-terminal chains from both subunits which are intertwined to form a well-ordered four-stranded antiparallel β-sheet. In each monomer, the catalytic triad (Asp-Ser-Gly) is located in a loop that forms a part of the ψ-structure characteristic to all aspartic proteases. The position of a water molecule between the active-site aspartate residues and the general scheme of H bonding within the active site bear close resemblance to those in pepsin-like aspartic proteases and therefore suggest a similar enzymatic mechanism. The binding cleft over the active site is covered by two flap arms, one from each monomer, which are partially disordered. The retroviral protease dimer has been compared with several enzymes of cellular origin, with chains aligning to an rms deviation of 1.90 Å or better.

Avian retroviruses contain an RNA genome encapsidated in a symmetric icosahedral core that is surrounded by a lipoprotein envelope (Dickson et al., 1984). The viral genes *gag*, *pol*, and *env* are translated as precursor polypeptides which are cleaved to produce the mature form of the proteins found in virions (Dickson et al., 1984). Processing of the *gag* and *pol* gene protein precursors of avian sarcoma/leukosis virus (ASLV), murine leukemia viruses (MuLV), and human immunodeficiency virus (HIV) has been shown to be directed by a viral-encoded protease (Dickson et al., 1984; Katoh et al., 1985; Kramer et al., 1986).

In avian retrovirus-infected cells, the viral-specific protease (PR) is synthesized as part of two large precursor polypeptides, $Pr76^{gag}$ and $Pr180^{gag\text{-}pol}$. The former includes the PR at its C-terminus, and the latter contains all of the *gag* and the *pol* sequences (Dickson et al., 1984) and is produced at $^1/_{20}$ the amount of $Pr76^{gag}$. The ASLV PR gene encodes a 13.5-kDa protein comprised of a single contiguous polypeptide chain of 124 amino acids. The enzyme purified from virions, however, is a dimer that exhibits a substrate specificity for its homologous *gag* and *pol* polyproteins although there are reports of heterologous substrates being cleaved by the PR (Skalka, 1989). Synthetic peptides that represent homologous *gag* and *pol* precursor cleavage sites are specific substrates for the PR and, in contrast to precursor polypeptides, are rapidly cleaved (Kotler et al., 1988, 1989).

The polyprotein cleavage sites have been deduced by alignment of the amino acid sequences at the C- and N-termini of the mature *gag* and *pol* proteins to sequences found in the precursor. Comparison of the sequence around 27 cleavage sites from several retroviruses indicates that each is cleaved at a unique sequence. The only common feature is that the

† Research sponsored in part by the National Cancer Institute, DHHS, under Contract NO1-CO-74101 with ABL, and in part by NCI Grant CA 38046 to J.L.
* To whom correspondence should be addressed.
‡ ABL-Basic Research Program.
§ On leave from Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland.
‖ Case Western Reserve University.

0006-2960/90/0429-5889$02.50/0 © 1990 American Chemical Society