Volume 7 Number 4 1979 Nucleic Acids Research

Is UAA or UGA part of the recognition signal for ribosomal initiation?

J.F.Atkins

Howard Hughes Medical Institute at Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

Received 7 July 1979

ABSTRACT

In none of the 92 published prokaryotic sequences is a translation codon preceeded by UAG as the first "termination codon". In most cases the UAA or UGA is close to the initiation codon and may be part of the ribosome recognition signal.

Early work on translation initiation focused on an abortive attempt to find a protein nucleic acid interaction which would explain the initiation specificity observed (review 1). This situation was radically changed by the prediction of Shine and Dalgarno in 1974 (2) that an important feature of initiation is an interaction between the 3' end of the intermediate sized ribosomal RNA and a sequence preceeding the initiation codon. Compelling evidence in favour of the proposal was obtained by the actual isolation of the mRNA-rRNA complex (3). The ultimate proof that this proposal was correct for one prokaryotic initiation has been provided by the recent isolation of mutants affecting the interaction (4). However this is clearly not the whole story as many regions which have the potential for mRNA-rRNA interaction and are followed at the appropriate distance by an AUG or GUG codon are non-functional in initiation. Many functional initiation codons are preceeded by a sequence which does not have extensive complementarity with the 3' end of ribosomal RNA. Another feature of initiation sequences which may affect the Shine Dalgarno interaction is the potential for many of them to form stem and loop structures (recently reviewed in 49). However the role of these potential loops in initiation is unclear. Extensive studies with fusions between the lac initiation region and the λ c1 or λ cro initiation region have so far not revealed the additional features which are important for initiation (52,53).

A re-examination of the sequences surrounding initiation codons reveals a striking fact not previously commented upon. None of the 63 published se-

quences from unrelated prokaryotes have UAG as the first translation terminator upstream of an initiation codon (table 1) and if related sequences are considered the total is 92. Some of these related sequences are similar and some are very different. On a random basis the expected frequency of UAG is 21 and 31 respectively. Several reports have noted that there was a terminator just upstream of the initiation codon, but the extent of this feature has not been appreciated. If the sequence was random then on average one would expect to find one terminator in any frame per seven codons. However examination of the published sequences (table 1) shows that half of the sequences have UAA or UGA within the first three codons, and nearly all of the rest have either within the first seven codons. This puts the majority of the terminator codons between the Shine Dalgarno region and the initiation codon, although some of them are part of that region, and some of them 5' to it. If there is functional significance in the UAA or UGA preceeding the initiation codon, then their distance from the initiation codon might be expected to be more constant. However the distance from the initiation to the Shine Dalgarno region is also variable though less so.

It appears not to be important which of the two, UAA or UGA occurs. The UAA or UGA codon occurs in any of the three frames with respect to the initiation codon, but perhaps there is a preponderance in the zero or +1 frame. There is no apparent distinction between UAG, UAA and UGA either upstream of the first "terminator" or downstream of the initiator, or corollary with the gene termination codon. In table 1 there are 11 UAG, 18 UAA and 10 UGA triplets upstream of the "terminator" preceeding the initiator, and 5 UAG, 19 UAA and 8 UGA triplets downstream of the initiator. In some cases one or two bases in the initiation codon itself form part of a "termination codon". Such terminators are noted in the table and are never UAG. In these cases the next upstream terminator is still nearby and is UAA or UGA. Not quite all the initiators are preceeded by a termination codon, but each of the exceptions may be interesting. The translation initiation codon for Φx174 gene A is very close to a transcription initiation site, and the resulting transcript is unstable (5.6). The proximity of fd gene III to the translation start site has not been published, but there may be at least some similarities with Φx174 gene A (7). The λ repressor (cI) protein can be synthesized from either of two mRNAs originating at different promoters (8). One message contains the sequences preceeding the initiation codon shown in table 1, whereas the other begins at the translation initiation codon directly. The 10 to 20 fold lower level of protein synthesis in the latter case has been presumed to be due to

the lack of mRNA-rRNA base pairing, but it could equally well be due to the absence of a UAA or UGA preceeding the initiation codon. Miller and colleagues (review 1) have studied translational re-starts in the lacI gene. In these cases also UAA or UGA is found as the closest upstream stop codon, however the distances are considerable. Perhaps significantly the level of translation initiation is low.

A considerable number of the known sequences are from polycistronic RNAs but it is very unlikely that the upstream termination codons are merely a fail safe mechanism for termination of the preceeding gene. In MS2 for instance, the maturation gene is the 5' proximal gene, and it has a termination codon as the second upstream codon from the initiation codon (9). The maturation gene is in a different reading frame (10) than the next downstream gene, the coat gene. Nevertheless the UGA which is the second upstream codon from the coat initiator is in phase with the coat gene and not the maturation gene. The coat and synthetase genes are in the same reading frame but there is a double stop codon at the end of the coat gene (11), and no in-phase read-through of these codons to the UGA which preceeds the synthetase gene has ever been detected despite a detailed study (Atkins, Gesteland, and Reid, unpublished).

This examination of "terminator codons" preceeding translation initiation sites was prompted by (1) my studies with J. Steitz and C. Anderson on the effects on protein synthesis of using antibody to release factors (unpublished), (2) the isolation by Ganoza and colleagues (review 12) of a protein which seems to be involved in initiation as well as termination, and (3) the isolation of a mutant, which is not in the structural gene for this protein (12), but which gives suppression of UAA and UGA and not UAG (41). The observations reported here lend support to the general idea suggested by Ganoza and colleagues that termination signals are involved in ribosome recognition (45). However analysis of the sequences specifically limit the signals to UAA and UGA, and make it clear that these signals are in positions where they could be as important as the "Shine Dalgarno interaction" in ribosome recognition. Perhaps then translation initiation involves not only interaction between 16s rRNA and the "Shine Dalgarno sequences", but also the nearby UAA or UGA. From the work of Ganoza it seems likely that a protein is involved in recognition and thus may be part of the initiation complex. However this protein has to be specific for UGA or UAA. Involvement of rRNA (2) in addition to the protein release factors (42,43) in chain termination has been considered, but the situation is still unclear (13).

TABLE 1

Gene	Sequence 5' of initiation codon	Gene initiation codon	Gene stop codon
MS2 maturation[9,10]	UUC CAU UCC UAG GAG GUU UGA CCU	GUG CGA GCU UUU AGU	UAG
MS2 coat[11]	GAG CCC UCA ACC GGA GUU UGA AGC	AUG GCU UCC AAC UUU	UAA
MS2 lysis[11 46,47]	AGG CAA UGC AAG GUC UCC UAA AAG	AUG GAA ACC CGA UUC	UAA
MS2 synthetase[11,31,32]	GCC AUU CAA ACA UGA GGA UUA CCC	AUG UCG AAG ACA ACA	UAG
Qβ maturation[1]	CUG AGU AUA AGA CGA CAU	AUG CCU AAA UUA CCG	
Qβ coat[1]	AAA CUU UGG GUC AAU GUG AUC	AUG GCA AAA UUA GAG	UGA
Qβ synthetase[27,1]	UAA CUA AGG AUG AAA UGC	AUG UCU AAG ACA G	UAA
ϕx174 A[6]	A AAU CUU GGA GGC UUU UUU	AUG GUU CGU UCU UAU	UGA
ϕx174 A*7[6]	U GAG UGU UCA AGA UUG CUG GAG GCC UCC ACU	AUG AAA UCG CGU AGA	UGA
ϕx174 B[6]	UGC UAA AGG UCU AGG AGC UAA AGA	AUG GAA CAA CUC ACU	UGA
ϕx174 C[6]	UAA AUC GAA GUG GAC UGC UGG CGG AAA	AUG AGA AAA UUC GAC	UGA
ϕx174 D[6]	UGU UCA ACC ACU AAU AGG UAA GAA AUC	AUG AGU CAA GUU ACU	UAA
ϕx174 E[6]	CUC GUC GCU GCG UUG AGG CUU GCG UUU	AUG GUA CGC UGG ACU	UGA
ϕx174 F[6]	CUU CGG CCC CUU ACU UGA GGA UAA AUU	AUG UCU AAU AUU CAA	UGA
ϕx174 G[6]	GUA GGU UUU CUG CUU AGG AGU UUA AUC	AUG UUU CAG ACU UUU	UGA
ϕx174 H[6]	UCU CCA GCC ACU UAA GUG AGG UGA UUU	AUG UUU GGU GCU AUU	UAA
ϕx174 J[6]	AAA AUU ACG UGC GGA AGG AGU GAU GUA	AUG UCU AAA GGU AAA	UAA
ϕx174 K[6]	AGA AGU UAA CAC UUU CGG AUA UUU CUG	AUG AGU CGA AAA AUU	UGA
fd I[7]	UUC UUA UUU GGA UUG GGA UAA AUA AAU	AUG GCU GUU UAU UUU	UAA
fd II[7]	CUU UUC UGA UUA UCA ACC GGG GUA CAU	AUG AUU GAC AUG CAU	UAA
fd III[7]	UUG GAG AUU UUC AAC GUG AAA AAA UUA	UUA	UAA
fd IV[7]	CUG UUU CAA UUA AAA AAG GUA AUU CAA	AUG AAA UUG UUA AAU	UAG
fd V[7]	UCU UAA AAU CGC AUA AGG UAA UUC AAA	AUG AUU AAA GUU GAA	UAA
fd VI[7]	CAU ACU GCG UAA UAA GGA GUC UUA AUC	AUG CCA GUU CUU UUG	UAA
fd VII[7]	GUC UGC GCC UCG UUC CGG CUA AGU AAC	AUG GAG CAG GUC GCG	UGA
fd VIII[7]	UUU UAC CCG UUU AAU GGA AAC UUC CUC	AUG AAA AAG UCU UUA	UGA
fd IX[7]	GCU UGG UAU AAU CGC UGG GGG UCA AAG	AUG AGU GUU UUA GUG	UGA
fd X[7]	CUG UUU AAA GCA UUU GAG GGG GAU UCA	AUG AAU AAU UAU GAC	UAA
T7 0.3 site a[1]	AAC UGC ACG AGG UAA CAC AAG	AUG GCU AUG UCU AAC	
T7 0.3 site b[1]	GUA CGA GGA GGA UGA AGA GUA	AUG U	
T7 1.0 polymerase [b]	UUU ACU AAC UGG AAG AGG CAC UAA	AUG AAC ACG AUU AAC	
T7 1.3 ligase[c]	UUU AAC CAA UAG GAG AUA AAC AUU	AUG AUG AAC AUU AAG	
T4 rII B[1]	C CUA AUA AGG AAA AUU	AUG UAC AAU	
E.c. thr leader[22]	AUA AAA AUU ACA GAG UAC ACA ACA UCC	AUG AAA CGC AUU AGC	UGA
E.c. thr A gene[22]	UUU UCG ACC AAA GGU AAC GAG GUA ACA	ACC AUG CGA GUG UUG AAG	
S.t. his leader[24]	UCA AAU GAA UAA GCA UUC AUC GGA AUU	UUU AUG ACA CGC GUU CAA	UAG
S.t. his G[24,d]	CGG UUC AGA CAG GAU AAA GAG GAA CGC	AGA AUG UUA GAC AAC AAC	UGA
S.t. his D[d]	AAG GCG GAA CCC UGU GAU GGA GUA AAG	ACC AUG AGC UUC AAU ACC	UGA
S.t. his C[d]	CGU AAA CGC CCU CAA GGA GCA AGC	AUG AGC ACU GAA AAC	
E.c. phe leader[33]	AAG UCA CUU AAG GAA ACA AAC	AUG AAA CAC AUA CCG	UGA
E.c. phe A[33]	CUU UUU UAU UGA UAA CAA AAA GGC AAC ACU	AUG ACA UCG GAA AAC	
E.c. trp leader[34]	CAC GUA AAA AGG GUA UCG ACA	AUG AAA GCA AUU UUC	
E.c. trp E[34]	UUU UUU GAA CAA AAU UAG AGA AUA ACA	AUG CAA AGA CAA AAA	
E.c. trp A[35,1,e]	UUU GAA AGC ACG AGG GGA AAU CUG	AUG GAA CGC UAC GAA	UAA
S.t. trp B[39]	ACA CUG CGC GCA UAU UAA GGA AAA AAC	AUG ACA ACA CUU CUC	
E.c. trp C[e]	CUG GCG GCA CGA GGG UAA AUG	AUG CAA ACC GUU UUA	
E.c. ara B[1]	UUU UUU GGA UGG AGU GAA ACG	AUG GCG AUU GCA AUU	
E.c. gal E[1]	AUA AGC CUA AUG GAG CGA AUU	AUG AGA GUU CUG GUU	
E.c. gal T[1]	UAU CCC GAU UAA GGA ACG ACC	AUG ACG CAA UUU AAU	
E.c. lac I[36]	AGU CAA UUC AGG GUG GUG AAU	GUG AAA CCA GUA ACG	UGA
E.c. lac Z[36]	GAU AAC AAU UUC ACA CAG GAA ACA GCU	AUG ACC AUG AUU ACG	
β-lactamase[26]	CAA UAA UAU UGA AAA AGG AAG AGU	AUG AGU AUU CAA CAU	UAA
E.c. ribosomal L14[27]	UUG ACA UUA GCG GAG CCU AAA	AUG AUC CAA GAA CAG	
E.c. ribosomal S12[27]	GCU AAA ACC AGG AGC UAU UUA	AUG GCA ACA GUU AAC	
E.c. ribosomal L12[44]	AUU ACC CAA CUU GAG GAA UUU AUA	AUG GCU AAG AAA GUA	UAA
E.c. ribosomal L1 [44]	GGC CUG GUA GUG GAG GAC UAA GAA	AUG GCU AAA CUG ACC	UAA
E.c. ribosomal L10[44]	GGC AAA CAU CCA GGA GCA AAG CUA	AUG GCU UUA AAU CUU	UAA
E.c. ribosomal L7/L12[44]	UUC UGA UAU UCA GGA ACA AUU UAA	AUG UCU AUC ACU AAA	UAA
E.c. RNA polymerase subunit?[44]	CUU GUC AGC GAG CUG AGG AAC CCU	AUG GUU UAC UCC UAU	
E.c. lipoprotein[50,51]	UCA AUC UAG AGG GUA UUA AUA	AUG AAA GCU ACU AAA	UAA
λ cro[18,28]	GGU UGC AUG UAC UAA GGA GGU UGU	AUG GAA CAA CGC AUA	UAA
λ o[28] CUG AGG UCA UUA	CUG GAU CUA UCA ACA GGA GUC AUU	AUG ACA AAU ACA GCA	UAG
λcI [8,29]	UAU CCC UUG CGG UGA UAG AUU UAA CGU	AUG AGC ACA AAA AAG	UGA
λcII[28,18]	UGU UAU CUA AGG AAA UAC UUA CAU	AUG GUU CGU GCA AAC	UGA
λcII (non functional)[18]	GG CCA GUC UAG AAA UAC UUA CAU	AUG GUU CGU GCA AAC	
	← ISZ		

Why is UAG excluded from being part of the "start signal"? The effect of a UAG should be tested by conversion of the "start signal" UAA or UGA by site directed mutagenesis (14). However a mutant of λ with an IS2 insertion just preceeding the cII gene may be instructive in this regard. The insertion interferes with the expression of cII and it causes UAG to be the first terminator upstream of the cII AUG initiation codon (15). In this mutant 4 of the 7 bases in the Shine Dalgarno sequence and all of the bases between the Shine Dalgarno sequence and the initiator AUG remain intact. However since 3 of the 7 bases in the Shine Dalgarno sequence are altered a clear conclusion cannot be drawn from this mutant.

Examination of the known sequences surrounding higher eukaryotic translation initiation sites reveal several differences. UAG is frequently found as the closest "terminator" upstream of the start codon, e.g., in SV40 T and t (16), SV40 VP_2(16), human β globin (17) and turnip yellow mosaic virus coat gene (18). Also the nearest "terminator" codon is frequently much further removed from the initiation codon than in prokaryotes. When more eukaryotic initiation sites are sequenced it will be possible to examine the situation in more detail. However numerous other studies (review 1, 19) also show differences between prokaryotic and eukaryotic initiation. In prokaryotes where there are polycistronic mRNAs the ribosomes may have evolved to recognize an intimate association between termination and initiation (48). Thus the use of a "termination" codon as part of a ribosome recognition signal even when it is unrelated in function to translation termination may not be too surprising.

Sequences surrounding initiation codons. The initiation codon is in italics. The closest upstream "terminator codon" is underlined twice. With ϕx174 gene K, T7 0.3 site b, trp A and the gene for ribosomal protein S12 the closest "terminator codon" (which is UAA or UGA) overlaps the initiation codon but it is underlined by a dotted line. All other terminator codons are also underlined by dotted line. Reference 1 is a review and is not the original source of these references, [a] is G. E. Christie and T. Platt (review 50), [b] is D. McConnell (DNA sequence) and J. Dunn (RNA sequence) unpublished, [c] is J. Dunn, unpublished, [d] is Husson and Barnes, unpublished (DNA sequence) and Kohno, T, unpublished (protein sequence), [e] is Wu and Platt (review 39). The related sequences also examined were from R17 and f2 (review 1), G4 (28) and M13 (P. Van Wezenbeek, T. Hulsegos and J. G. G. Schoenmakers, unpublished results), Salmonella tryptophan operon (40), and the E. coli histidine operon (23). E.c. signifies E. coli and S.t. Salmonella typhimurium.

ACKNOWLEDGMENTS

I thank Ray Gesteland for characteristically helpful suggestions and Joan Steitz and Clelia Ganoza for prior discussions of translation initiation.

REFERENCES

1. Steitz, J. A. (1979) in Biological Regulation and Development, Vol. 1, Goldberger, R. F., Ed., pp. 349-399 Plenum Press, New York.
2. Shine, J. and Dalgarno, L. (1974) Proc. Nat. Acad. Sci. USA 71, 1342-1346
3. Steitz, J. A. and Jakes, K. (1975) Proc. Nat. Acad. Sci. USA 72, 4734-4738
4. Dunn, J. J., Buzash-Pollert, E. and Studier, F. W. (1978) Proc. Nat. Acad. Sci. USA 75, 2741-2745
5. Fujimura, F. K. and Hayashi, M. (1978) in The Single-Stranded DNA Phages, Denhardt, D. T., Dressler, D. and Ray, D. S, Eds., Cold Spring Harbor Laboratory, New York.
6. Sanger, F., Coulson, A. R., Freidmann, R., Air, G. M., Barrell, B.G., Brown, N. L., Fiddes, J. C., Hutchinson III, C. A., Slocombe, P.M., and Smith, M. (1978) J. Mol. Biol. 125, 225-246
7. Schaller, H., Beck, E. and Takanami, M (1978) in the Single-Stranded DNA Phages, Denhardt, D. T., Dressler, D. and Ray, D. S., Eds., Cold Spring Harbor Laboratory, New York
8. Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, T. (1976) Science 194, 156-161
9. de Wachter, R., Vandenberghe, A., Merregaert, J., Contreras, R., and Fiers, W. (1971) Proc. Nat. Acad. Sci. USA 68, 585-589
10. Contreras, R., Ysebaert, M., Min Jou, W. and Fiers, W. (1973) Nature New Biol. 241, 99-101
11. Min Jou, W., Haegeman, G., Ysebaert, M. and Fiers, W. (1972) Nature 237, 82-88.
12. Ganoza, M.C. (1977) Can. J. Biochem 55, 267-281
13. Caskey, C. T., Bosch, L., Konecki, D. S. (1977) J.Biol. Chem. 252, 4435-4437.
14. Taniguchi, T. and Weissmann, C. (1978) J. Mol. Biol. 118, 533-565
15. Rosenberg, M., Court, D., Shimatake, H., Brady, C. and Wulff, D. L. (1978) Nature 272, 414-423.
16. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., and Ysebaert, M. (1978) Nature 273, 113-120
17. Baralle, F. E. (1977) Cell 12, 1985-1995
18. Guilley, J. and Briand, J. P. (1978) Cell 15, 113-122
19. Kozak, M. (1978) Cell 15, 1109-1123
20. Schwarz, E., Scherer, G., Hobom, G. and Kössel, H. (1978) Nature 272, 410-414
21. Belin, D., Hedgpeth, J., Selzer, G. B. and Epstein, P. H. (1979) Proc. Nat. Acad. Sci. USA 76, 700-704
22. Gardner, J. (1979) Proc. Nat. Acad. Sci. USA 76, 1706-1710
23. Di Nocera, P. P., Blasi, F., Di Lauro, R., Frunzio, R., and Bruni, C. B. (1978) Proc. Nat. Acad. Sci. USA 75, 4276-4280
24. Barnes, W. M. (1978) Proc. Nat. Acad. Sci. USA 75, 4281-4285
25. Zurawski, G., Brown, K., Killingly, D. and Yanofsky, C. (1978) Proc. Nat. Acad. Sci. USA 75, 4271-4275
26. Sutcliffe, J. G. (1978) Proc. Nat. Acad. Sci. USA 75, 3737-3741

27. Post, L. E., Arfsten, A. E., Reusser, F. and Nomura, M. (1978) Cell 15, 215-229
28. Godson, G. M., Barrell, B. G., Staden, R. and Fiddes, J. C. (1978) Nature 276, 236-247
29. Humayun, A. (1977) Nucleic Acid Res. 4, 2137-2343
30. Scherer, G. (1978) Nucleic Acid Res. 5, 3141-3155
31. Vandenberghe, A., Min Jou, W. and Fiers, W. (1975) Proc. Nat. Acad. Sci. USA 72, 2559-2562
32. Atkins, J. F. and Gesteland, R. F. (1975) Mol. Gen. Genet. 139, 19-31
33. Zurawski, G., Brown, K., Killingly, D. and Yanofsky, C. (1978). Proc. Nat. Acad. Sci. USA 75, 4271-2475
34. Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C. L., Squires, C. and Yanofsky, C. (1975) Science 189, 22-26
35. Platt, T. and Yanofsky, C. (1975) Proc. Nat. Acad. Sci. USA 72, 2399-2403
36. Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S. (1975) Science 187, 27-35
37. Gesteland, R. F., Wolfner, M., Grisafi, P., Fink, G., Botstein, D. and Roth, J. R. (1976) Cell 7, 381-390
38. Weiner, A. M. and Weber, K. (1973) J. Mol. Biol. 80, 837-855
39. Steege, D. A. and Söll, D. (1979) in Biological Regulation and Development, Vol. 1, Goldberger, R. F., Ed., Plenum Press, New York, pp. 433-485
40. Lee, F., Bertrand, K., Bennett, G. and Yanofsky, C. (1978) J. Mol. Biol. 121, 193-217
41. Phillips, S. L., Schlessinger, D. and Apirion, D. (1969) Cold Spring Harb. Symp. quant. Biol. 34, 499-503
42. Caskey, T., Scolnick, E., Tompkins, R., Goldstein, J., and Milman, G. (1969) Cold Spring Harb. Symp. quant. Biol. 34, 479-491
43. Capecchi, M. R. and Klein, H. A. (1969) Cold Spring Harb. Symp. quant. Biol. 34, 469-477
44. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Nat. Acad. Sci. USA 76, 1697-1701
45. Ganoza, M.C., Fraser, A. R., and Neilson, T. (1978) Biochem. 17, 2769-2775
46. Atkins, J. F., Steitz, J. A., Anderson, C. W., and Model, P. (1979) Cell, in press
47. Beremand, M. and Blumenthal, T. (1979) Cell, in press
48. Platt, T. (1978) in The Operon, Miller, J. H. and Reznikoff, W. S., Eds., pp. 263-302, Cold Spring Harbor Laboratory, New York
49. Selker, E. and Yanofsky, C. (1979) J. Mol. Biol. 130, 135-143
50. Pirtle, R. M., Pirtle, I. L. and Inouye, M. (1978) Proc. Nat. Acad. Sci. USA 75, 2190-2194
51. Pirtle, R. M., Pirtle, I. L. and Inouye, M. (1978) Fed. Proc. 38, 499-499
52. Roberts, T. M., Kacich, R. and Ptashne, M. (1979) Proc. Nat. Acad. Sci. USA 76, 760-674
53. Backman, K. and Ptashne, M. (1978) Cell 13, 65-71