RNA structural elements for expression in *Escherichia coli*

α_1-Antitrypsin synthesis using translation control elements based on the cII ribosome-binding site of phage λ

Luc-Henri Tessier, Sophie Jallat, Martine Sauvageot, Ronald G. Crystal* and Michael Courtney+

TRANSGENE SA, 11 Rue de Molsheim, 67 Strasbourg, France and **National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20205, USA*

Received 27 August 1986; revised version received 23 September 1986

Analysis of a series of λ *c*II::α_1-antitrypsin (α_1AT) gene fusions of different sizes showed that increased α_1AT expression correlated with the stabilisation of a particular computer-predicted RNA secondary structure. Moreover, significant synthesis of unfused α_1AT was achieved by reconstruction of this conformation to permit interaction between the upstream region of the ribosome-binding site and the first part of the α_1AT coding sequence. This high-level expression was dependent upon certain silent point mutations in the coding sequence, indicating that RNA primary and secondary structure determinants can operate in concert to dictate the efficiency of protein synthesis.

recombinant DNA *Gene fusion* *Point mutation* *λplasmid vector* *Gene cloning* *Emphysema*

1. INTRODUCTION

The distance and sequence composition between the Shine-Dalgarno element and the start codon are important in determining the efficiency of mRNA translation in *E. coli* [1,2]. Moreover, both sequence elements should be free from occluding secondary structure that may block ribosome access [3–5]. It has recently been shown that alterations to the beginning of the coding sequence can also influence gene expression independently of secondary structure effects [6]. Here, the role of a particular secondary structure in determining the efficiency of protein synthesis was investigated. Furthermore, given that ribosomes may recognise structural signals as well as the linear nucleotide sequence, it was attempted to engineer effective expression of a foreign gene by reconstructing the conformation of an endogenous RBS. The model used was the bacteriophage λ *c*II RBS which has been shown to direct efficient expression of CII protein in a multicopy plasmid system [7]. We have used the gene for human α_1AT, a protease inhibitor which has potential clinical application in the therapy of lung disorders characterized by proteolytic tissue damage, for example emphysema [8]. The expression in *E. coli* of functionally active α_1AT and active centre variants of enhanced therapeutic potential have been reported [9,10].

Abbreviations: α_1AT, α_1-antitrypsin; Ap, ampicillin; Δ, deletion; IFN, interferon; IHF, integration host factor; LB, Luria broth; PA, polyacrylamide; PAGE, PA gel electrophoresis; PBS, phosphate-buffered saline; RBS, ribosome-binding site; RID, radial immune diffusion; TE buffer, 2 mM Tris-HCl pH 8, 10 mM EDTA

+ To whom correspondence should be addressed

2. MATERIALS AND METHODS

Restriction enzymes were purchased from Bethesda Research Laboratories, New England Biolabs, Boehringer Mannheim, and Amersham.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

DNA polymerase, polynucleotide kinase, T_4 DNA ligase and calf intestinal phosphatase were purchased from Boehringer Mannheim; S_1 nuclease was from P-L/Pharmacia. The use of these enzymes and protocols for bacterial transformations, DNA preparation and general cloning techniques have been described [11]. Point mutations were obtained using in vitro site-directed mutagenesis essentially by methods already described [12].

Overnight cultures of *E. coli* strain TGE900 [F⁻ *su ilv his bio* (λΔBam *c*Its857 ΔH1)] containing plasmid were grown at 28°C in LB + 100 μg Ap/ml. These were diluted into M9 medium containing 0.2% glucose and amino acid supplements and grown at 28°C to an A_{600} of 0.3 (1.5×10^8 cells per ml). At this point the cultures were divided and parallel samples were grown at 28°C (non-induced) and at 37°C (induced) for 4 h. De novo α_1AT synthesis was analyzed by pulse labelling 200 μl of culture for 1 min with [^{35}S]methionine (Amersham; spec. act. ⩾ 1000 Ci/mmol) to a final concentration of 10 μCi/ml. Labelling was stopped by adding 1 ml cold PBS to the samples which were then centrifuged for 10 min at 10000 × *g*, resuspended in sample buffer and heated at 100°C for 5 min. Aliquots containing 50000 cpm trichloroacetic acid insoluble radioactivity were loaded on SDS-10% PA gels and analyzed by fluorography and autoradiography. ^{14}C-labelled M_r markers were purchased from Bethesda Research Laboratories. 10% PA gel analysis of unlabelled cell extracts was performed to estimate steady-state α_1AT levels. Protein was visualised by Coomassie blue staining and quantitation was achieved by densitometric tracing using a Shimadzu LS-930 scanner.

Cleared cell lysates were prepared as follows: cultures induced as above were harvested by low-speed centrifugation and cells were resuspended in half the volume of the original culture in TE buffer. After sonication the samples were centrifuged at 10000 × *g* for 10 min. Total protein in the cleared lysates was estimated using the Bio-Rad Coomassie blue assay.

The concentration of α_1AT in cleared lysates was measured by radial immune diffusion (Calbiochem) and elastase inhibitory capacity. Elastase assays were performed using 50 ng human neutrophil elastase (Elastin Products) preincubated for 30 min at 23°C with increasing amounts of cell extracts before addition of 2 mM substrate (*N*-succinyl-Ala-Ala-Pro-Val-*p*-nitroanilide; Calbiochem). α_1AT activity was measured by monitoring residual elastase activity at 410 nm.

Possible RNA secondary structures were obtained using a computer algorithm that determined the conformation of minimum free energy [13]. For this analysis a relatively narrow sequence window was used (≤ 200 nucleotides) since for the *c*II RBS efficient translation is achieved with a 150 nucleotide long fragment spanning the initiation codon. This implied that all the elements necessary for translation, including possible secondary structures, were contained within this fragment.

3. RESULTS AND DISCUSSION

Our analysis is based on the structure shown in fig.1A, which is similar to that already proposed for *c*II mRNA [14]. All the sequence elements considered necessary for initiation of translation are contained within this stem-loop conformation. The Shine-Dalgarno sequence (AGGAA) and the initiation codon are largely accommodated within loops, an arrangement thought to be important for optimal translation [4]. The sequence -AAUCAAUUGU- which lies in a loop upstream from the Shina-Dalgarno sequence is homologous to the IHF box [15]. The major role of IHF is in phage λ site-specific recombination but recent reports have suggested an additional function in stimulating *c*II translation [16,17]. It is conceivable that IHF could interact with this sequence element to stimulate translation directly. A striking feature of the *c*II RBS structure is the long double-stranded stem that brings together segment -52 to -39 upstream from the Shine-Dalgarno sequence and 9 to 22, which encodes amino acids 3-7 of the CII protein. We have designed a series of plasmid constructions to investigate the possible importance of this stem structure in the initiation of translation.

A series of gene fusions was constructed with 13, 8, 5 and 3 N-terminal *c*II codons joined to the mature α_1AT coding sequence. Fig.2 shows the expression plasmid and the sequence of the *c*II::α_1AT gene fusions. The original fusion (pTG922) contains, from the 5′-end, 13 codons from *c*II, 4 codons from a polylinker sequence and a mature α_1AT coding sequence lacking only the 5′

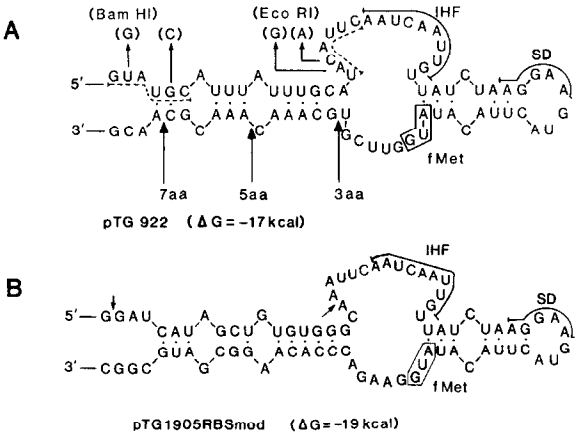

Fig.1. Predicted mRNA secondary structures for the native λ cII ribosome-binding site (RBS) region (A) and a modified RBS/α_1AT expression construction (pTG1905RBSmod) designed to recreate the original structure (B). (A) Structure predicted for λ *c*II mRNA using the algorithm developed by Zuker and Stiegler [13] with an analysis window of 180 nucleotides. Shown (large arrows) are the positions of disruptions of the structure by reducing the λ *c*II::α_1AT fusions to 7, 5 and 3 amino acids. (B) Predicted structure for pTG1905RBSmod. The plasmid was constructed by replacing the segment between the *Bam*HI and *Eco*RI (indicated by arrows) with a sequence complementary to the beginning of the modified α_1AT gene (see text). The α_1AT sequence starts immediately after the start codon (framed).

Glu codon. The 7, 5 and 3 amino acid fusions were derived from pTG922 by replacing the *Nde*I-*Bam*HI fragment with a synthetic DNA duplex of the appropriate sequence. Comparison of the expression of these fusions in *E. coli* allows investigation of the effect of progressively reducing the long stem of the *c*II RBS structure depicted in fig.1A. Each clone was analysed by (i) RID of sonicated culture supernatants, (ii) SDS-PAGE of total cell proteins, and (iii) SDS-PAGE of pulse-labelled newly synthesized proteins. The latter most closely reflects translation efficiency since (i) and (ii) measure steady-state levels and can be influenced by the stability of the protein product. PAGE analysis of total cell proteins demonstrated that polypeptides of the predicted size (43–45 kDa) were synthesized after induction (fig.3A). Densitometric tracing of the gel lanes indicated that 13, 7, 5 and 3 amino acid fusions were expressed at 12, 8.5, 5.5% total cell protein and undetectable levels, respectively (table 1). Measurement of α_1AT activity in cleared sonicated culture supernatants did not reflect this progressive decrease in expression; in each case the same level of activity (3×10^5 units/cell) was measured. This is observed because the bulk of the fusion protein was deposited in insoluble inclusion bodies which pellet on centrifugation of the sonicated samples (table 1).

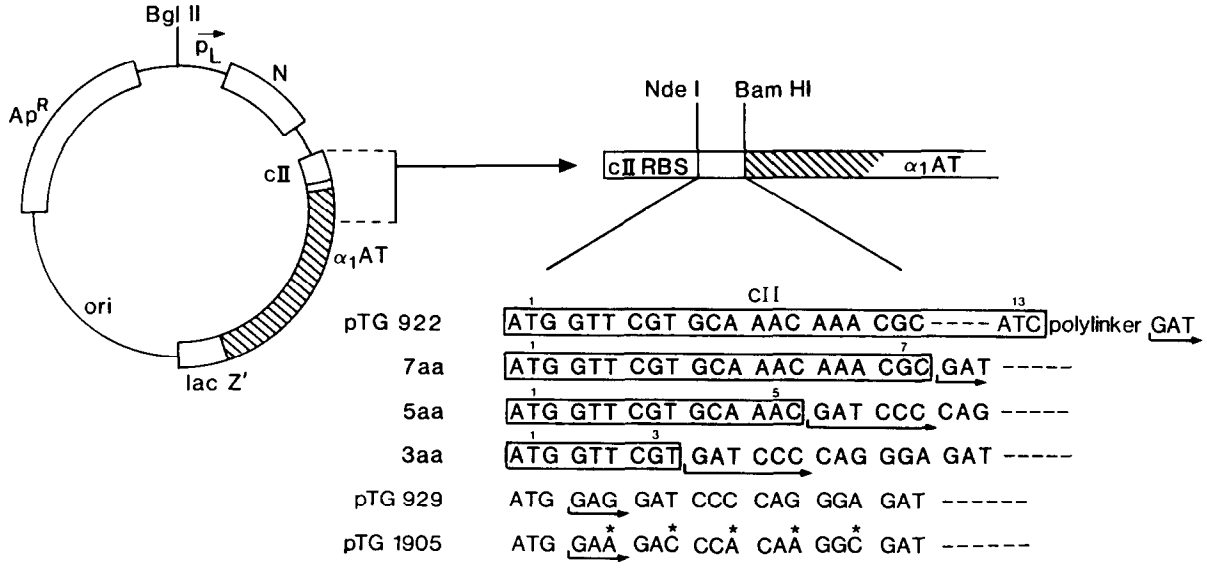

Fig.2. Structure of the *E. coli* expression vector pTG922 [9,10] and the nucleotide sequences of the 5′ portions of the genes studied.

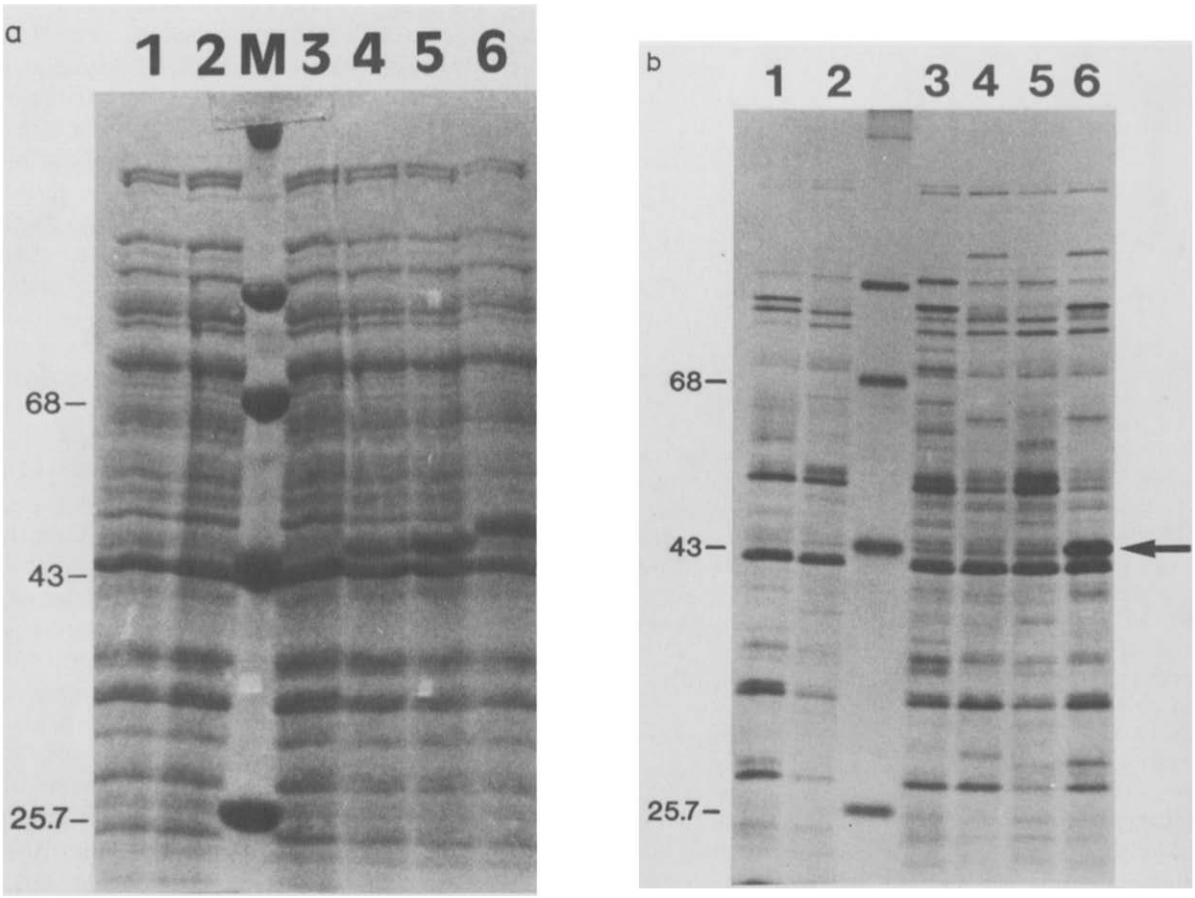

Fig.3. SDS-10% polyacrylamide gel analysis of *E. coli* cell extracts containing α_1AT variants. (A) Total cell lysates. Lanes: 1, 2, control *E. coli* TGE900 cultures containing the parental plasmid pTG908 (expression plasmid with no α_1AT sequences) at 28°C (non-induced) and 37°C (induced); 3,4,5,6, induced cultures containing the 3, 5, 7 and 13 amino acid λ cII: :α_1AT fusions. (B) Extracts from [^{35}S]methionine pulse-labelled cultures. Lanes: 1,2, control cultures containing the pTG908 at 28°C (non-induced) and 37°C (induced), respectively; 3,4,5,6, induced cultures containing pTG929, pTG929RBSmod, pTG1905, pTG1905RBSmod. Molecular mass markers are indicated in kDa.

The decrease in expression levels with reducing size of *c*II fusion follows the increase in calculated free energy of the *c*II RBS secondary structure (−17, −15, −11 and −3 kcal/mol for the 13, 7, 5 and 3 amino acid fusions, respectively). Also the overall shape of the most stable RNA conformation does not change significantly until the fusion is reduced to 3 amino acids (not shown). These data suggest that the efficiency of translation could be directly related to the stability of the *c*II RBS stem and loop structure. When the first codon of mature α_1AT (GAG) is linked directly to the initiation AUG the *c*II RBS stem and loop structure is disrupted and there is no expression (see plasmids pTG929 and pTG1905, fig.2 and table 1). However, if the region 5′ to the IHF loop is modified to permit interaction with the beginning of the α_1AT gene, then it should be possible to reconstruct the original *c*II RBS conformation. To achieve this, unique restriction enzyme sites (for *Eco*RI and *Bam*HI) were generated by site-directed mutagenesis at each end of the long stem, as shown in fig.1A. After treatment with these enzymes synthetic DNA fragments were inserted whose plus strands were partly complementary to nucleotides 9–22 of the α_1AT gene. In this experiment two dif-

Table 1

Analysis of α_1AT expression in the various constructions

Vector	Total α_1AT per cell		% α_1AT of newly synthesized protein	% α_1AT of total soluble protein	% α_1AT of total insoluble protein
	% of total cell protein	Mol/cell			
pTG922 (13 aa)	12	2×10^6	28	1.5	25
7 aa	8.5	1.7×10^6	24	1.5	21
5 aa	5.5	10^6	21	1.5	12
3 aa	–	–	–	–	–
pTG929	–	–	–	–	–
pTG929 RBS mod	–	–	–	–	–
pTG1905	–	–	–	–	–
pTG1905 RBS mod	~1	$\sim 2 \times 10^5$	17	0.5	2

Expression levels relative to total cell protein, newly synthesized protein and total insoluble protein were determined by densitometric scanning of PAGE analyses. Expression with respect to total soluble protein was measured by radial immune diffusion and neutrophil elastase inhibitory capacity. (-) Undetectable α_1AT; aa, amino acid

ferent α_1AT genes were studied; one carried the original cDNA sequence (pTG929), the other being altered to generate silent mutations that brought the sequence closer to the statistically 'preferred' sequence for ribosome binding (pTG1905) [18,19]. The inserted upstream sequences were: 5′-GAUCAUAGCUGUGUGGGCA-3′ (pTG1905RBSmod) and 5′-GAUCAUCCUCCUAGGGACA-3′ (pTG929RBSmod). The sequences of the oligodeoxynucleotides were chosen to allow similarities in the following elements with respect to the original *c*II structure: (i) the overall shape of the structure, (ii) the free energy for the overall structure and, (iii) the position of the interior loops and the size of IHF loop. The structure of the modified RBS for the pTG1905 sequence is shown in fig.1B.

SDS-PAGE of pulse-labelled samples (fig.3B) showed that pTG929 and pTG1905 expressed no detectable α_1AT. In the case of pTG1905 with the modified upstream region (pTG1905RBSmod), there was significant expression (17% of newly synthesized protein) whereas there was no expression with the modified pTG929 (pTG929RBSmod). These results were confirmed by RID and anti-elastase assays of sonicated cultures extracts (table 1). These data suggest that the correct secondary structure together with a favourable primary structure are required for effective protein synthesis. Either feature alone was not sufficient for high-level expression.

Several studies have demonstrated that for efficient translation the Shine-Dalgarno sequence and/or the start codon should be accessible to ribosomes and therefore situated either in a region free from secondary structure elements or in the loop of a hairpin [3,20]. We have shown here that the presence of a particular RNA stem-loop structure may also be important, at least in the case of the *c*II RBS. This conformation could act to maintain the RBS in a spatial configuration favourable for ribosome binding, elongation or perhaps interaction with IHF. Therefore, to construct an expression vector for the production of unfused α_1AT, sequences upstream from the *c*II Shine-Dalgarno element were modified so as to reconstruct the original *c*II RBS structure.

This study also illustrates the importance of the sequence 3′ to the start codon at the level of

primary structure alone. Effective expression of unfused α_1AT was achieved only when silent mutations were introduced in codons 2, 3, 4 and 5 of the α_1AT gene. Thus, for the efficient expression of a foreign gene in *E. coli*, it may be necessary to mutate the nucleotide sequence around the RBS to optimize both the primary and secondary structure of the mRNA.

ACKNOWLEDGEMENTS

We are indebted to Dr J.-P. Lecocq for his support and direction. We also thank Professor P. Chambon and Professor P. Kourilsky for many helpful suggestions, Dr P. Stiegler for advice concerning the computer analysis, F. Daul for the artwork and N. Monfrini for typing the manuscript.

REFERENCES

[1] De Boer, H.A., Hui, A., Comstock, L.J., Wong, E. and Vasser, M. (1983) DNA 2, 231-235.
[2] Shepard, H.M., Yelverton, E. and Goeddel, D.V. (1982) DNA 1, 125-131.
[3] Iserentant, D. and Fiers, W. (1980) Gene 9, 1-12.
[4] Gheysen, D., Iserentant, D., Derom, C. and Fiers, W. (1982) Gene 17, 55-63.
[5] Hall, M.N., Gaday, J., Debarbouillé, M. and Schwartz, M. (1982) Nature 295, 616-618.
[6] Tessier, L.H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M. and Lecocq, J.P. (1984) Nucleic Acids Res. 12, 7663-7675.
[7] Shimatake, H. and Rosenberg, M. (1981) Nature 292, 128-132.
[8] Gadek, J.E. and Crystal, R.E. (1982) in: The Metabolic Basis of Inherited Diseases, 5th edn (Stanbury, J.B. et al. eds) pp.1450-1467, McGraw-Hill, New York.
[9] Courtney, M., Buchwalder, A., Tessier, L.H., Jaye, M., Benavente, A., Balland, A., Kohli, V., Lathe, R., Tolstoshev, P. and Lecocq, J.P. (1984) Proc. Natl. Acad. Sci. USA 81, 669-673.
[10] Courtney, M., Jallat, S., Tessier, L.H., Benavente, A., Crystal, R.G. and Lecocq, J.P. (1985) Nature 313, 149-151.
[11] Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[12] Zoller, M. and Smith, M. (1982) Nucleic Acids Res. 10, 6487-6500.
[13] Zuker, M. and Stiegler, P. (1981) Nucleic Acids Res. 9, 133-148.
[14] Place, N., Fien, K., Mahoney, N.E., Wulff, D.L., Ho, Y.S., De Bouck, C., Rosenberg, M., Shih, N.C. and Gussin, G.N. (1984) J. Mol. Biol. 180, 865-880.
[15] Craig, N.L. and Nash, H.A. (1984) Cell 39, 707-716.
[16] Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I. and Echols, H. (1982) Cell 31, 565-573.
[17] Oppenheim, A.B., Gottesman, S. and Gottesman, M. (1982) J. Mol. Biol. 158, 327-346.
[18] Scherer, G.F.E., Walkinshaw, M.D., Arnott, S. and Morre, D.J. (1980) Nucleic Acids Res. 8, 3895-3907.
[19] Stormo, D., Schneider, T.D. and Gold, L.M. (1982) Nucleic Acids Res. 10, 2971-2996.
[20] Stanssens, P., Remaut, E. and Fiers, W. (1985) Gene 36, 211-223.