Nucleic Acids Research, 1993, Vol. 21, No. 17 **4019–4023**

Identification and characterization of *E.coli* ribosomal binding sites by free energy computation

Tamar Schurr, Eyal Nadir and Hanah Margalit*
Department of Molecular Genetics, Hebrew University, Hadassah Medical School, Jerusalem 91010, Israel

Received June 1, 1993; Revised and Accepted July 22, 1993

ABSTRACT

Sequences upstream from translational initiation sites of different *E.coli* genes show various degrees of complementarity to the Shine–Dalgarno (SD) sequence at the 3′ end of the 16S rRNA. We propose a quantitative measure for the SD region on the mRNA, that reflects its degree of complementarity to the rRNA. This measure is based on the stability of the rRNA-mRNA duplex as established by free energy computations. The free energy calculations are based on the same principles that are used for folding a single RNA molecule, and are executed by similar algorithms. Bulges and internal loops in the rRNA and mRNA are allowed. The mRNA string with maximum free energy gain upon binding to the rRNA is selected as the most favorable SD sequence of a gene. The free energy value that represents the SD region provides a quantitative measure that can be used for comparing SD sequences of different genes. The distribution of this measure in more than 1000 *E.coli* genes is presented and discussed.

INTRODUCTION

Recognition of translational initiation sites by ribosomes involves several features of mRNA. These include the initiation codon and nucleotides in its vicinity, and the mRNA secondary structure (reviewed in 1–4). In particular, there is compelling evidence of involvement of a short sequence upstream from the translational start site, in ribosome binding, by base pairing with the 3′ end of 16S rRNA (3). This complementarity was demonstrated first by Shine and Dalgarno (5), and the sequence at the 3′ end of 16S rRNA, GAUCACCUCCUUA, or its complement on the mRNA are known as the Shine–Dalgarno region (hereinafter, the SD region). Different genes contain different sub-sequences of the SD region upstream from their translational start codons. It was demonstrated that the length of the SD sub-sequence and its location relative to the translational start codon affect the efficiency of translation initiation (1–3,6).

When examining the nucleotides upstream from the translational start positions of *E.coli* genes, an unequivocal identification of the SD sequence is not always obvious. In many cases, several sub-sequences of the SD region can be identified, enabling different base-pairings with the rRNA. From sequence examination alone it is difficult to compare the potential of the different mRNA strings for binding the SD sequence on the rRNA. In previous studies on ensembles of *E.coli* translational initiation sites, the SD sequences were located by searching for strings of at least 3 consecutive nucleotides (7,8). However, it is not clear that strings of consecutive nucleotides are the most favorable for interaction, and the selection of such a string is somewhat subjective. We were interested in a measure that would enable the selection of an optimal mRNA SD region based on a quantitative criterion. Our assumption is that the interaction that takes place between the rRNA and mRNA is the one that is energetically most stable, namely, the interaction with maximal free energy gain. We propose the use of thermodynamical considerations for selection of a preferred SD sequence among the various possibilities upstream from the AUG. Moreover, as the SD sequence is identified and defined by its free energy value, this measure enables a quantitative comparison among SD sequences of different genes.

We calculate the free energies (kcal/mol) for all possible duplexes between the 16S rRNA 3′ end and a region of 21 nucleotides upstream from the start codon (9–13). The choice of that region is based on experimental evidence identifying the sequence that is protected against nuclease attack in an initiation complex, as spanning 21 nucleotides upstream and 14 nucleotides downstream from the translational start position (14,15). Computational sequence analyses based on information theory also identified this span of nucleotides as having a high information content (16,17). The mRNA sequence with the maximal free energy gain upon binding with the rRNA is selected as the SD sequence. We demonstrate a difference between the distribution of this measure upstream from translational initiation codons and upstream of internal AUG's in *E.coli* genes, and examine the genes with extreme values of free energy.

METHODS

Our database includes 1159 sequences of complete genes (from start to stop codon) fetched from *E.coli* genome database, ecogen5 (18,19). For each gene, we use a region of 21 nucleotides

* To whom correspondence should be addressed

upstream from the translational start position. The calculations are done for the SD region of the mRNA molecule in its state just prior to ribosome binding, assuming it is momentarily not involved in any intra secondary structure and is available for pairing with the rRNA. For the rRNA SD, we use the 13 nucleotides at the 3′ end of the 16S rRNA (see above). These are thought to be unpaired and available for annealing with the mRNA (20).

We have developed a computer program that generates all possible sub-sequences in the −21 to −1 region relative to the translational start codon, and evaluates their potential for pairing with the SD sequence at the 3′ end of the 16S rRNA. For each such pairing, the program calculates the free energy value of the rRNA-mRNA duplex. We use the experimentally measured free energy values of Freier et al. (12), and apply the improvements generated by Jaeger et al. (13) after optimizing the penalties for gaps. In general, the considerations used in pairing between rRNA and mRNA are similar to those used for the folding of a single RNA molecule (10−13). In case of a mismatch, only internal loops and bulges are allowed (no hairpin loops are allowed in either the mRNA or rRNA). Dangling ends are not included in the free energy calculation. GU base pairs are allowed only in internal positions. The free energy loss by duplex initiation, 3.4 kcal/mol (12), is taken into account.

The sub-sequence that yields the lowest free energy value upon binding is selected as the SD sequence. When several sub-sequences yield the same minimal free energy value, one of them is randomly chosen. Such sub sequences are found in 15% of the studied genes. Usually, these sub sequences differ only in minor components, such as a bulge location (for example, see figure 1). For each gene the program lists the nucleotides participating in the optimal pairing of the mRNA with the rRNA and the free energy value for that pairing.

Computer facilities

The program is written in Fortran 77 and run on VAX 9000-210 (VMS V5.5-2).

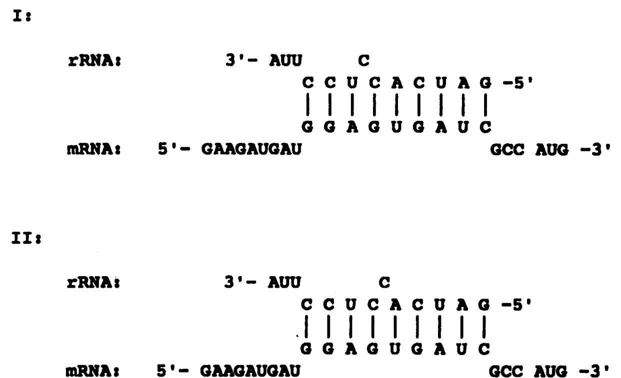

Figure 1. Two energetically equal options for pairing of the mRNA upstream from the hisD gene with rRNA SD. These options differ in their bulge location only (the 7th cytosine in the first pairing option and the 8th in the second, when counting from the 3′ end of the rRNA). G = −9 kcal/mol.

rRNA \ mRNA	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	
G	0	0	0	0	2	1	0	1	3	0	3	4	3	1	5	19	10	39	19	18	26	154
A	0	0	0	4	3	2	4	4	3	3	5	7	3	8	28	27	35	24	25	39	13	237
U	0	0	6	3	3	7	5	6	4	3	13	11	18	29	35	47	46	52	64	36	18	406
C	0	6	3	3	8	6	7	4	3	16	11	17	30	34	49	46	72	64	48	20	11	458
A	5	2	3	8	7	8	5	5	18	12	20	32	24	48	47	79	88	41	24	39	8	493
C	2	3	7	3	5	3	6	15	11	22	40	48	60	79	90	71	53	26	40	9	7	600
C	2	6	3	5	5	6	15	8	24	52	66	78	134	149	137	79	17	19	5	7	3	820
U	2	2	7	5	8	14	8	27	56	68	90	160	178	176	113	23	22	10	3	5	2	979
C	1	5	10	12	15	11	20	51	63	81	156	177	176	122	26	18	7	0	5	2	1	959
C	1	11	13	11	20	22	41	61	61	121	148	153	119	22	18	9	1	1	2	1	0	836
U	6	7	11	20	22	26	45	38	78	92	115	91	18	12	8	1	1	2	1	0	0	594
U	2	5	15	11	12	25	27	46	55	72	65	12	10	5	1	0	1	1	0	0	0	365
A	2	5	7	8	18	17	30	25	37	34	5	9	3	0	0	0	1	0	0	0	0	201
	23	52	85	93	128	148	213	291	416	576	737	799	776	685	557	419	324	279	236	176	89	7102

Figure 2. Frequency of pairs of mRNA-rRNA nucleotides involved in binding according to the free energy calculations. For each mRNA sequence, the string of nucleotides with the lowest free energy value upon binding to the rRNA SD is considered. The rRNA is defined by its exact sequence. The mRNA nucleotides involved in binding are defined by their position relative to the AUG translational start codon, and by their complementary nucleotide on the rRNA. Note that G can pair either with C or a U.

RESULTS AND DISCUSSION

For each sequence (21 nucleotides long) in the database, the program generates all possible pairings between the mRNA and rRNA that comply with the above considerations, and calculates the free energy of the duplex. On the average, 133383 combinations per sequence (s.d. = 87739) have been generated and evaluated. The string of mRNA nucleotides with the lowest free energy value upon binding with the rRNA has been selected. Thus, the SD region of each gene is represented by a sequence and a free energy value of its pairing with a defined sub-sequence of the rRNA.

Figure 2 describes the frequency of pairs of mRNA-rRNA nucleotides involved in the optimal pairing, as indicated by the free energy considerations. The average number of paired nucleotides per sequence is 6.3. As demonstrated in the last column, the most frequent rRNA nucleotides engaged in binding are CCUCCU. These are the nucleotides that are usually looked for visually, when attempting to identify SD sequences. The region mostly involved in binding on the mRNA is located between −12 to −7 relative to the translational start position. The average distance between the last mRNA nucleotide participating in the pairing (at the 3′ end) and the translational start codon is 6.4 nucleotides. There is only moderate correlation between the free energy values of the selected SD sequences and their location relative to the translational start position ($r = 0.45$).

The distribution of free energy values for the 1159 genes in the database is demonstrated in Figure 3, in comparison with the distribution of free energy values calculated similarly for 1000 regions upstream from randomly selected internal AUG codons. The two distributions are clearly different ($p < 0.0001$ by a t-test on the mean values of the two distributions). In general, the sequences of regions upstream from the spurious initiation codons show higher free energy values than the sequences upstream from actual translational start codons. Also, the percentage of sequences with no energetic gain upon pairing with the rRNA, i.e., with no defined SD region, is higher in the spurious sequences. This result supports the validity of the free energy value as a quantitative measure for the quality of a Shine−Dalgarno sequence.

For 1121 out of 1159 genes there is at least one sub-sequence of nucleotides that yields a free energy gain when pairing with the rRNA. The average free energy value per gene is −4.9 kcal/mol with a standard deviation of 2.4 kcal/mol. The genes with lowest free energy values, hence, best SD sequences, are listed in Table I. Note that an optimal sequence with a perfect

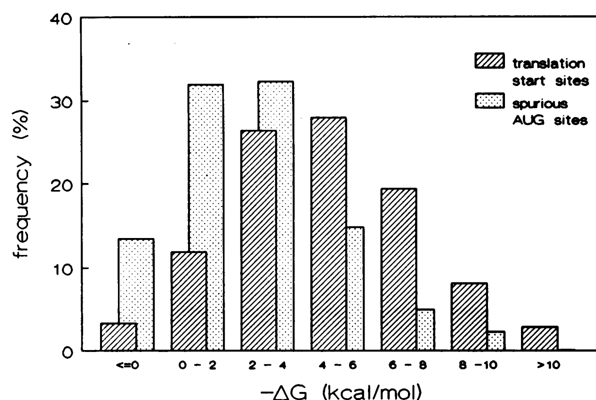

Figure 3. Distribution of free energy values for all 1159 mRNA sequences in the database compared to the distribution of free energy values for 1000 mRNA sequences upstream from internal AUG's.

match to the SD sequence gives a free energy value of −19.5 kcal/mol, while the best score for *E.coli* genes in the database is −14.3 kcal/mol. When examining this group of genes no obvious common denominator can be seen. They are spread all along the genome and vary functionally.

For the other 38 genes (listed in Table II) there is no string of nucleotides that gives a free energy gain when pairing with the rRNA. This list, as that in Table I, contains genes that vary in their function and location along the chromosome. However, 45% of these sequences (17 genes) are of unknown function, in comparison to 17% genes with unknown function in the whole database. The absence of an SD region in these sequences may suggest that some of them are not actual genes, as most of them were determined in *E.coli* genome as open reading frames (21). The sequences with known function in this list have been further analyzed to try and find an alternative mechanism that would enable pairing with the rRNA. Two possibilities have been considered:

a) An SD sequence further upstream from the AUG. As the distance between the SD sequence and the AUG was shown to be important and to influence translation level (1−3), the search for a distant SD region has been supplemented by examining the potential for stem-loop formation in the interval between its location and the AUG. Existence of such a stem-loop secondary structure could bring the distant SD sequence closer to the AUG, enabling it to function in a regular fashion. An example for such a structure can be seen in gene 38 of bacteriophage T4 (3). Using the program described above, all pairing options with the SD rRNA have been checked in a region of 100 nucleotides upstream from the AUG, and the string with lowest free energy value upon binding has been selected. Some of the new upstream subsequences have lower free energy values than the average. No common distance from the translational start site could be identified. Stem-loop structures downstream from the distant SD region have been searched for, by using the FOLD program of the GCG package (22), but consistent such structures could not be identified.

b) An alternative common pattern in the vicinity of the AUG, different from SD region, that would be complementary to other regions of the rRNA. Standard sequence comparison programs (22) have been used to search for such complementarities.

Table I. List of genes with lowest free energy values (< $^{-10}$ kcal/mol).

name	location	energy	function
yabB	1.924	−11.3	−
murF	2.031	−12.3	D-alanine:D-alanine-adding enzyme
speE	2.919	−11.5	spermidine synthase = putrescine aminopropyltransferase
lpxA	4.502	−10.4	UDP-N-acetylglucosamine acetyltransferase
betA	7.105	−10.2	choline dehydrogenase, a flavoprotein
bioC	17.567	−10.6	biotin biosynthesis; reaction prior to pimeloyl CoA
clpA	19.961	−10.6	ATP-dependent clpA protease, ATP-binding, regulatory sununit
dmsC	20.395	−12.9	anaerobic dimethyl sulfoxide reductase chain C
hyaA	22.258	−10.2	hydrogenase-1 small subunit
ndh	25.184	−11.0	respiratory NADH dehydrogenase
icdE	25.687	−12.1	isocytrate dehydrogenase, NADP+ specific chromosomal fragment
topA	28.655	−11.2	DNA topoisomerase I, omega protein
pspC	29.487	−10.9	shock protein activates psp operon expression
narV	33.061	−11.6	cryptic nitrate reductase II, gamma subunit
fdnI	33.393	−10.5	formate dehydrogenase-N, cytochrome B556(Fdn) subunit
hdhA	36.515	−10.2	NAD-dependent 7-alpha-hydroxysteroid dehydrogenase
manX	40.950	−11.1	mannose phosphotransferase system, protein II-A(III)
cheY	42.379	−10.4	response regulator for chemotactic response (cheA sensor)
proW	60.442	−10.2	high-affinity transport system for glycine betaine & proline
ygdC	63.810	−10.1	−
ansB	66.749	−10.1	cytoplasmic L-asparaginase II; isozyme
rpsE	74.179	−10.5	30S ribosomal subunit protein S5
argD	75.120	−10.4	acetylornithine delta-aminotransferase
pckA	76.083	−13.2	phosphoenolpyruvate carboxykinase
atpC	84.391	−10.2	membrane-bound ATP synthase, F1 sector, epsilon-subunit
atpH	84.483	−11.4	membrane-bound ATP synthase, F1 sector, delta-subunit
yjbA	91.380	−11.1	−
phnI	93.095	−10.3	utilization of phosphorus-containing compounds
phnC	93.198	−11.1	utilization of phosphorus-containing compounds
yjfB	95.886	−14.3	−
argI	96.464	−12.6	ornithine cabamoyltransferase
fecD	97.211	−10.6	citrate-dependent iron transport, membrane-bound protein
fimH	98.032	−10.4	minor fimbrial subunit, adhesin

Included are the genes that appear in the right most category of the histogram in Figure 2. Names of genes and locations on *E.coli* chromosome (in centisomes) are consistent with the file Ecogen5, that includes an ordered database of *E.coli* sequences (18,19). The function description is taken from a recent compilation by M. Riley (32). Free energy values are calculated as described in the Methods section. '−' indicates that the gene's function is unknown.

However, no such pattern could be identified that is common in more than a few sequences. Complementarity between the rRNA and mRNA in other regions was proposed before (23−26). In particular, a sequence located downstream from the translational start position and termed the downstream box was proposed as an additional mechanism for ribosome binding in highly expressed genes (24), and was shown to compensate for the lack of a 5′ leader region in the cI gene of the phages lambda and HKO22 (27). No significant appearance of the downstream box could be identified in the sequences listed in Table II.

Some genes could be translationally coupled to upstream adjacent genes with an energetically favorable SD region. A potential translational coupling requires that the involved genes be in close proximity. We have screened the genes that are located

Table II. List of genes with highest free energy values (>= 0 kcal/mol).

name	location	function
nhaA	0.383	Na+/H antiporter activity
apaH	1.096	diadenosine tetraphosphatase
fhuC	3.637	hydroxamate-dependent iron uptake, cytoplasmic membrane component
ybcA	12.205	–
ybeB	14.520	–
nadA	16.961	quinolinate synthetase, A protein
galE	17.123	UDP-galactose 4-epimerase
bioF	17.542	7-keto-8-aminopelargonic acid synthetase
rimK	19.286	ribosomal protein S6 midification protein
yceD	24.774	–
phoQ	25.540	sensor for phoP, histidine protein kinase
tpr	27.710	a protamine like protein
dicA	35.449	regulator of dicB
ydiA	38.444	–
btuD	38.558	vitamin B12 transport, membrane-associated protein
yebB	41.923	–
pgsA	42.935	phosphatidylglycerophosphate synthetase
fliM	43.524	flagelar biosynthesis
menD	51.208	menaquinone biosynthesis
yfdA	53.478	–
yfeA	54.264	–
nadB	58.236	quinolinate synthetase, B protein
ygfA	65.776	–
yggC	66.172	–
parC	68.136	topoisomerase IV subunit A
parE	68.344	topoisomerase IV subunit B
ygiA	68.468	–
ygiB	68.470	–
pinO	74.367	–
yhjA	79.916	–
glyQ	80.192	glycine tRNA synthetase, alpha chain
yibB	81.686	–
spoU	82.355	–
rnpA	83.663	RNase P, protein component
yihA	87.311	–
yijA	89.779	–
rts	89.964	pantothenate kinase
rimI	99.325	modification of 30S ribosomal subunit protein S18

Included are the genes that appear in the left most category of the histogram in Figure 2. Names, locations on chromosome, and function of gene are as in Table I.

upstream from and adjacent to the genes listed in table II, to identify genes with stop codons that are in close proximity to the start codons of the genes with no SD region. A distance not greater than 35 bases, approximately the size that is covered by the ribosome, has been considered as close. In 7 out of the 38 genes such potential translational coupling could be identified; this frequency is not greater than expected, based on the frequency of such proximal genes in the whole database. Also, a translational enhancer, consisting of UUUUUUUU, was identified in the *rnd* gene of *E.coli* and was proposed to play a role in translation of genes with a weak SD region (28). Such a U_8 signal could not be detected in this group of genes.

It is conceivable that an SD region with the potential to bind more favorably with the rRNA could bring about higher translation levels of its gene. In the absence of consistent quantitative data on the level of translation of different genes it is impossible to examine its relationship to the SD measure defined here. In a previous study (25), there was an attempt to relate the quality of the SD region with the level of expression of different genes, as documented in reference 29. Thanaraj and Pandit (25) chose consecutive nucleotides for the SD region and calculated its free energy of pairing with the rRNA based on the free energy values of the dinucleotides (12). They claimed that highly expressed genes show lower free energy values than poorly expressed genes, and that they contain additional regions complementary to the rRNA. The SD measure presented here uses more accurate free energy calculations and demonstrates similar relationships for these groups of genes. Thus, the genes listed in previous studies as highly expressed (25,29) have lower free energy values than the genes listed as poorly expressed. However, examining free energy values of poorly and highly expressed genes, without relating them to the average value in the whole database, might be misleading. We find that the average free energy values of these two groups of genes do not deviate significantly from the mean free energy value of all genes. Therefore, at least based on that sample (25), it seems that the stability of the duplex between the rRNA and the Shine–Dalgarno region, is not a distinctive measure between highly and poorly expressed genes. Also, high levels of expression are due to many factors and do not necessarily reflect high translation levels, implying that the information on level of expression may not be sufficient. To account for the correlation between the SD measure proposed here and the level of translation, direct measurements regarding the level of translation of different genes are needed. However, even if such data existed, we would not expect a linear relationship between the level of translation and the SD measure (3,6). Dreyfus (30) has cloned small DNA fragments of actual ribosome binding sites as well as intergenic and coding regions to a ribosome binding selection vector. All authentic start sites, except for two without an SD region, showed some level of translation. In other sequences no translation at all could be identified, despite the presence of an SD sequence spaced appropriately relative to the start codon. From that study, it is clear that ribosome binding is determined by several factors dictated by the sequences in the vicinity of the SD sequence, that still have to be revealed. The SD region is an important determinant for translation initiation, but it acts in concert with other determinants. Other factors like the initiation codon, the spacing between the SD sequence and the initiation codon, and the native secondary structure of the mRNA (31) play a role in determining the translation initiation efficiency. A combination of all these factors probably determines translation rate (1–4,6,30,31). At this stage we cannot predict that the genes listed in Table I are more efficiently translated than other genes, or that the genes listed in Table II are less efficiently translated, but only to document that they have better or worse SD sequences according to our quantitative measure.

The SD measure presented here is based not only on consecutive nucleotides of the SD region, but also allows for gaps in the mRNA or rRNA, introducing bulges or internal loops to the paired region. The inclusion of gaps enables longer sub-sequences of the SD region to be identified. If the free energy gain for the additional base pairs compensates for the penalty of a bulge or an internal loop, the longer sequences have the potential to interact favorably with the rRNA, and there is no justification for excluding these possibilities from consideration *a priori*. About half of the sequences in our database are found to contain SD sequences predicted to interact more favorably with the rRNA when bulges or internal loops are introduced. The free energy calculations are done as accurately as possible, according to the same rules that are assumed for folding of a single RNA molecule. This measure can be used for the selection of the most favorable SD string in a given sequence, and enables a quantitative comparison between SD regions of different genes.

With future accumulation of data regarding translation rates, it should be possible to combine the quantitative measure proposed here with other sequence dependent factors that determine translation efficiency, to predict the level of translation, based on sequence data alone.

REFERENCES

1. Gold,L. and Stormo,G. (1987) In Neidhart,F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (eds.), *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. ASM, Washington, Vol. II, pp. 1302–1307.
2. Stormo,G.D. (1986) In Reznikoff,W. and Gold,L. (eds.), Maximizing Gene Expression. Butterworths Publishers, Boston. pp.195–224.
3. Gold,L. (1988) *Ann. Rev. Biochem.*, 57, 199–233.
4. de Smith,M.H. and van Duin,J. (1990) *Prog. Nucleic Acids Res. Mol. Biol.*, 38, 1–35.
5. Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci.*, 71, 1342–1346.
6. Ringquist,S., Shinedling,S., Barrick,D., Green,L., Binkley,J., Stormo,G. and Gold,L. (1992) *Mol. Microbiol.*, 6, 1219–1229.
7. Stormo,G.D, Schneider,T.D. and Gold,L. (1982) *Nucleic Acids Res.*, 10, 2971–2996.
8. Gren,E.J. (1984) *Biochimie*, 66, 1–29.
9. Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Croters,D.M. and Gralla,J. (1973) *Nature New Biol.*, 246, 40–41.
10. Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, 9, 133–148.
11. Zuker, M. (1989) *Science*, 244, 48–52.
12. Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci.*, 83, 9373–9377.
13. Jaeger,J.A., Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci.*, 86, 7706–7710.
14. Steitz,J.A. (1969) *Nature*, 224, 967–964.
15. Hartz,D., McPheeters,D.S., Traut,R. and Gold,L. (1988) *Methods in Enzymology*, 164, 419–425. 16. Schneider,TD, Stormo,GD, Gold,L and Ehrenfeucht,A. (1986) *J. Mol. Biol.*, 188, 415–431.
17. Rudd,K.E, and Schneider,T.D. (1992) In Miller,J. (ed.), A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, New-York. pp. 17.19–17.45.
18. Rudd,K.E., Miller,W., Werner,C., Ostell,J., Tolstoshev,C. and Satterfield,S.G. (1991) *Nucleic Acids Res.*, 19, 637–647.
19. Rudd,K.E. (1992) In Miller,J. (ed.), A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, New-York. pp. 2.3–2.43.
20. Noller,H.F. (1984) *Annu. Rev. Biochem.*, 53, 119–162.
21. Rudd,K.E. *personal communication.*
22. Genetic Computer Group (1991) Program manual for the GCG package. Version 7, april 1991, Madison, Wisconsin.
23. Peterson,G.B., Stockwell,P.A. and Hill,D.F. (1988) *EMBO J.*, 7, 3957–3962.
24. Sprengart,M.L., Fatscher,H.P. and Fuchs,E. (1990) *Nucleic Acids Res.*, 18, 1719–1723.
25. Thanaraj,T.A. and Pandit,M.W. (1989) *Nucleic Acids Res.*, 17, 2973–2985.
26. Olins,P.O. and Rangwala,S.H. (1989) *J. Biol. Chem.* 264:16973–16976.
27. Shean,C.S. and Gottesman,M.E. (1992) *Cell*, 70, 513–522.
28. Zhang, J. and Deutcher, M.P. (1992) *Proc. Natl. Acad. Sci.* 89:2605–2609.
29. Gouy,M. and Gautier,C. (1982) *Nucleic Acids Res.*, 10, 7055–7074.
30. Dreyfus,M. (1988) *J. Mol. Biol.*, 204, 79–94.
31. de Smith,M.H. and van Duin,J. (1990) *Proc. Natl. Acad. Sci.*, 87, 7668–7672.
32. Riley,M. (1993) *Micorobiol. Rev.*, In Press.