Journal of Bioenergetics and Biomembranes, Vol. 20, No. 1, 1988

MINI-REVIEW

Expression of the *unc* Genes in *Escherichia coli*

John E. G. McCarthy[1]

Received June 4, 1987

Abstract

The *unc* (or atp) operon of *Escherichia coli* comprises eight genes encoding the known subunits of the proton-translocating ATP synthase (H^+-ATPase) plus a ninth gene (*uncI*) of unknown function. The subunit stoichiometry of the H^+-ATPase ($\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1 a_1 b_2 c_{10-15}$) requires that the respective *unc* genes be expressed at different rates. This review discusses the experimental methods applied to determining how differential synthesis is achieved, and evaluates the results obtained. It has been found that the primary level of control is translational initiation. The translational efficiencies of the *unc* genes are determined by primary and secondary mRNA structures within their respective translational initiation regions. The respective rates of translation are matched to the subunit requirements of H^+-ATPase assembly. Finally, points of uncertainty remain and experimental strategies which will be important in future work are discussed.

Key Words: *E. coli unc* operon; H^+-ATPase; subunit stoichiometry; gene expression; codon usage; translational initiation; Shine–Dalgarno sequence; recombinant DNA.

Introduction

The proton-translocating ATP synthase (H^+-ATPase) of *Escherichia coli* plays a pivotal role in the interconversion of the free energy of a proton electrochemical gradient and the free energy of ATP synthesis/hydrolysis. It belongs to a group of enzymes manifesting homologous structures and functions that is widely represented in nature (see, e.g., Amzel, 1981; Senior, 1979; Nelson, 1981; Fillingame, 1981). The *E. coli* H^+-ATPase is chromosomally encoded and is constitutively synthesized to reach between 1 and 2%

[1]Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-3300 Braunschweig, West Germany.

0145-479X/88/0200-0019$06.00/0

of the total cellular protein in wild type cells (von Meyenburg *et al.*, 1984b). Purification and characterization of the enzyme revealed that it is composed of eight different types of subunit (Friedl *et al.*, 1979; Foster and Fillingame, 1979; Schneider and Altendorf, 1982). The soluble F_1 part of the enzyme, which alone has ATP hydrolase activity, comprises five types of subunit (α, β, γ, δ, and ε), whereas the membrane-integrated component F_0 comprises three types of subunit (*a*, *b*, and *c*).

Mapping experiments indicating the location of the genes encoding the subunits of the enzyme began more than sixteen years ago (Butlin *et al.*, 1971; Downie *et al.*, 1979). Most of the so-called *unc* genes[2] could be shown by means of complementation experiments to map at about 83 min on the *E. coli* chromosome (Downie *et al.*, 1979; Gibson, 1983). Indeed, all of the *unc* genes could be isolated as a group on defective derivatives of bacteriophage λ bearing *E. coli* chromosomal DNA from the *ori* region (Miki *et al.*, 1978; von Meyenburg *et al.*, 1979).

Sequencing of DNA from these phages subsequently revealed the complete primary structures of the *unc* genes together with those of the flanking regions on the chromosome (see the reviews of von Meyenburg *et al.*, 1982a; Futai and Kanazawa, 1983; and Walker *et al.*, 1984). Combining the DNA sequencing information with protein sequencing data and also comparing it with the sequences of homologous genes in other organisms allowed elucidation of the structure and organization of the genes encoding the eight different subunits of the H^+-ATPase (see especially Walker *et al.*, 1984). These genes lie, together with a ninth gene called gene 1 or *uncI*, in an operon (Fig. 1). A role has yet to be assigned to the *uncI* gene, which, at least under the experimental conditions described so far, is not essential to the assembly or function of the H^+-ATPase (von Meyenburg *et al.*, 1982b; Gay, 1984). A striking aspect of the operon structure is that the genes encoding the F_0 membrane sector of the H^+-ATPase, and the genes encoding the relatively soluble F_1 sector, respectively, are arranged in two groups, as shown in Fig. 1, which also correlates the nomenclature for the genes with that for the polypeptides. This organization of the genes (and even their order) into F_0 and F_1 groups is to various extents conserved in other bacteria which have F_1F_0-type H^+-ATPases (Cozens and Walker, 1987).

A particularly remarkable aspect of the structure of the *E. coli* H^+-ATPase (and also of its counterparts in other bacteria, mitochondria, and chloroplasts) is the stoichiometry of the component subunits. The relative molar quantities of the *E. coli* subunits estimated on the basis of radioactivity

[2]Mutations in the *unc* genes do not necessarily lead to an "uncoupled" state and this gene name is therefore misleading in strict bioenergetic terms. For historical reasons, however, *unc*, rather than the more appropriate *atp* (see Hansen *et al.*, 1981; Walker *et al.*, 1984), is used in this review.

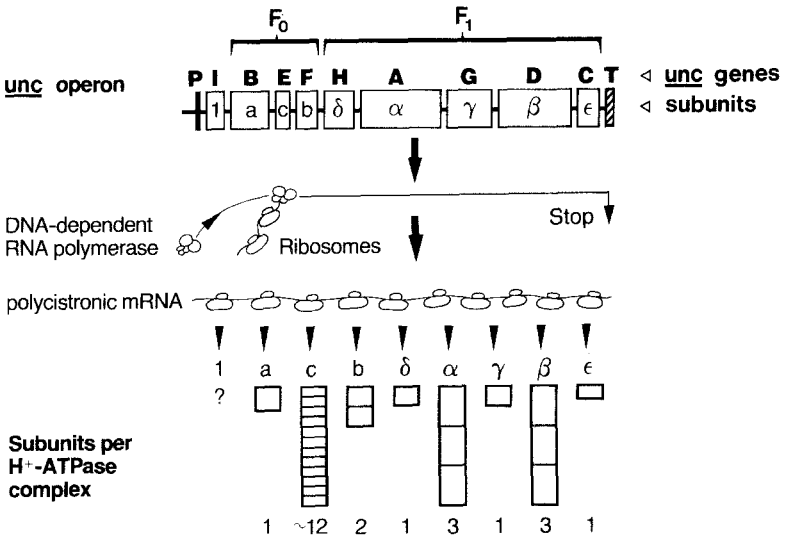

Fig. 1. Expression of the *unc* genes. The major route seems to involve transcription of a long polycistronic message bearing all nine cistrons, which are then translated at different rates. The relative amounts of each subunit synthesized in this way match the H^+-ATPase stoichiometry quite closely so that there is little wastage.

incorporation studies are $\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1 a_1 b_2 c_{10-15}$ (Foster and Fillingame, 1982; von Meyenburg *et al.*, 1982a; see also Lünsdorf *et al.*, 1984), whereas the respective *unc* genes are present in single copies. This raises a question of general relevance to the expression of genes in *E. coli*, especially of those genes that are organized in operons. Are these relative amounts of subunits matched by differential rates of subunit synthesis and if so, how is this differential expression achieved? This is the major theme dealt with in this article. Brief reference will also be made to the factors controlling (over)-production of H^+-ATPase subunits in *E. coli*, and to the effects observed as a result of it. Further aspects of the structure and function of the individual *unc* genes and of their products have been dealt with elsewhere (see, e.g., Futai and Kanazawa, 1983; Walker *et al.*, 1984).

Possible Levels of Control

The simplest mode of synthesis of the H^+-ATPase subunits would involve no control to accommodate the relative quantities required of the different subunits. Those protein molecules exceeding the required amounts would simply linger in the cell until finally being degraded. This very wasteful mode of synthesis would be atypical for *E. coli*, and we now know

that it does not provide an adequate description of expression of the *unc* operon. Indeed the mode of expression of the *unc* operon is much more interesting than that, as has become evident from examination of its transcription and translation.

The initial indication as to the mode of transcription of the *unc* genes took the form of polarity effects induced by the insertion of phage Mu. Thus the insertion of Mu phage into *uncB*, for example, disrupted transcription of all of the downstream genes (Gibson *et al.*, 1978; compare also von Meyenburg *et al.*, 1982b). Sequencing data subsequently revealed the presence of a clearly promoterlike signal upstream of the start of *uncI* (Gay and Walker, 1981). That this is the major promoter for transcription of the *unc* operon could be confirmed by S1 nuclease mapping, which localized the start point of transcription as 73 base pairs upstream of the *uncI* reading frame (Nielsen *et al.*, 1984; Porter *et al.*, 1983). At least one further, weaker promoter seems to be present in the *uncB* gene (von Meyenburg *et al.*, 1982a; Nielsen *et al.*, 1984). However, the efficiency of transcriptional initiation within this gene is estimated to be about one-tenth of that directed by the major *unc* promoter (Nielsen *et al.*, 1982).

The electrophoretic separation of gene transcripts extracted from *E. coli* and their subsequent blotting onto nitrocellulose (Northern blotting) allowed the identification of the mRNA species generated by transcription of the *unc* operon (Jones *et al.*, 1983). Hybridization experiments were performed with radioactively labelled (nick-translated) stretches of DNA derived from different regions of the *unc* operon. A major transcript could be detected which hybridized with both promoter-proximal and promoter-distal probes and which was also long enough ($\approx$7000 bases) to have been initiated at the mapped major start point and terminated at the ϱ-independent terminator located downstream of *uncC*. A relatively minor *unc* mRNA species of 5 to 6 kb was also detected. The hybridization data indicated that this latter species could theoretically have been initiated within the *uncI* gene and would be long enough to end within, or just after, the *uncD* cistron, perhaps as the result of termination there. It is, however, uncertain how the minor mRNA species is formed. These results indicated that the *unc* operon is transcribed primarily to yield one mRNA species bearing all of the *unc* cistrons, which is in turn translated to produce all of the H^+-ATPase subunits (Fig. 1). A striking aspect of the Northern blots presented by Jones *et al.* (1983) is that there is relatively little sign of degradation of the *unc* mRNA species, which both appear as remarkably sharp hybridization bands. This observation is relevant to the mode of degradation of *unc* mRNA, a point that will be returned to later.

It should be emphasized here that little is known about the (transcriptional) regulation of the *unc* operon in response to changes in the physiological

state of the cell. Jones *et al.* (1983) found that the same two mRNA species described above were present in apparently the same relative amounts in both aerobically and anaerobically grown cells. They also examined the expression of the plasmid-borne *galK* gene transcribed from the major *unc* promoter inserted before it. That there was no apparent variation in galactokinase synthesis under aerobic and anaerobic conditions indicated that the *unc* promoter activity also remained invariant under the tested growth conditions.

Is there any mechanism at the mRNA level that could be responsible for differential synthesis of H^+-ATPase subunits? Premature transcriptional termination and/or differential mRNA degradation could theoretically alter the relative concentrations and/or half-lives of specific *unc* cistrons. The influence of premature termination would be expected to be limited to the possible imposition of a gradient of expression (polarity) decreasing from the promoter down to the terminator. On the other hand, increased rates of degradation of specific regions, perhaps preceded by selective processing of the polycistronic mRNA, could in principle provide a means of controlling the relative concentrations of any of the cistrons. The experiments of Jones *et al.* however, did not reveal the presence of any processing or degradation products, although it should be remembered that these can be difficult to detect. A further relevant observation was made by pulse-labelling the H^+-ATPase subunits synthesized in cells at different time points after the addition of rifampicin, which blocks initiation of transcription. It could be demonstrated that all of the *unc* cistrons have similar functional half-lives (von Meyenburg *et al.*, 1984a). This result indicated that differential control is unlikely to be exercised upon the degradation of specific cistrons, and is reminiscent of the data reported for the morphogenetic region of bacteriophage λ.[3]

It was in fact investigations of the translational efficiency of the *unc* operon that first directly identified a capacity for differential expression of the individual genes.

Differential Translational Efficiency Within the *unc* Operon

Studies of the expression of plasmid-borne *unc* genes *in vitro* and in minicells provided the first indications that the H^+-ATPase subunits are

[3]The morphogenetic genes of bacteriophage λ are transcribed from the p'_R promoter to yield a very long (over 20 kbp) polycistronic mRNA. The amounts of the individual proteins synthesized per cell differ in some cases by a ratio of greater than 100 : 1. There is no preferential chemical or functional decay of the cistrons on the mRNA, which indicates that control is exercised at the translational level (Ray and Pearson, 1974, 1975).

synthesized at different rates (Downie *et al.*, 1980; Brusilow *et al.*, 1982). Brusilow *et al.* (1982) found that the profile of expression of most of the *unc* genes was roughly in accordance with the specific requirements for each of the different subunits in enzyme assembly. A similar pattern of expression rates was also observed when transcription was initiated from promoters other than the major *unc* promoter. These data led to the suggesion (Brusilow *et al.*, 1982) that there could be control of *unc* gene expression at the translational level. Further experiments using an *in vitro* system, moreover, also showed that *uncI* is expressed at about the same rate as *uncB* and *uncG* (Brusilow *et al.*, 1983). There were however, discrepancies between the minicell and *in vitro* systems. Especially the last three genes of the *unc* operon were apparently more poorly expressed *in vitro* than in minicells (Brusilow *et al.*, 1982).

How can the efficiency of translation be controlled? One conceivable point of control is the polypeptide elongation process, the rate of which can be influenced by the sequence of the mRNA being translated (see, e.g., Pedersen, 1984; Varenne *et al.*, 1984). It is generally observed that highly expressed genes contain many fewer codons corresponding to minor isoaccepting species of tRNA than do poorly expressed genes (Ikemura, 1981a,b; Grosjeans and Fiers, 1982). This general correlation between the expression rates of genes and their codon usages (but see, e.g., Holm, 1986) has been cited as indirect evidence for a role of codon structure (and codon context) in determining the relative rate of translation. It has been argued that highly expressed genes contain more of those codons that allow anticodon interactions of intermediate energies which are in turn postulated to be conducive to rapid turnover (Grantham *et al.*, 1981). The same line of argument has also been applied to the *unc* operon, in which the codon usages of the highly expressed genes (e.g., *uncE*, *uncA*, and *uncD*), as opposed to the poorly expressed ones, are consistent with the generally established trend (Hansen *et al.*, 1981; Kanazawa *et al.*, 1984; Walker *et al.*, 1984). It is of course implicit in this explanation of differential synthesis that elongation is rate controlling in translation. However, as outlined below, this hypothesis finds no support in the results of more detailed investigations of *unc* gene translation.

It is useful to define the translational initiation region (TIR) of a gene in functional terms as that region of the mRNA primarily responsible for determining not only the position, but also the efficiency, of translational initiation (Fig. 2; see also Steitz, 1979; Gold *et al.*, 1981). This region includes the translational start codon and the Shine–Dalgarno region, which shows complementarity to the 3′ end of the 16S ribosomal RNA, as well as other elements including the N-terminal region of the structural gene itself. Our appreciation of the extent and nature of the TIR, and of the function of the sequence elements within it, has developed slowly over the last ten years. The

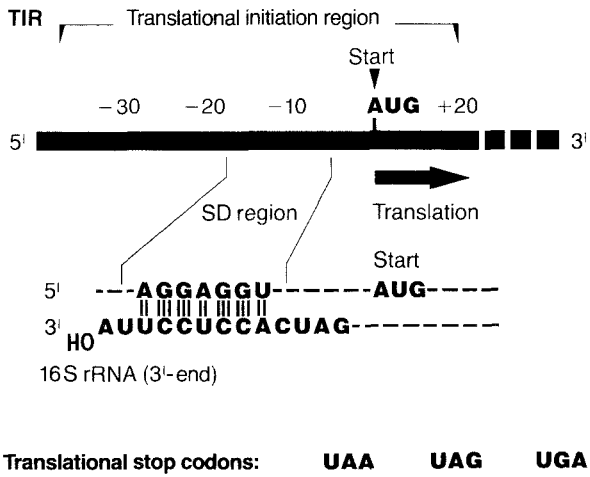

Fig. 2. This diagram defines the translational initiation region of a cistron on the mRNA. It indicates the approximate position of the Shine–Dalgarno (SD) region relative to the AUG start codon. The 3′ end of the 16S rRNA shows (a variable degree of) homology to the SD region.

approximate form presented schematically in Fig. 2 has been shaped primarily by the results of sequence comparisons and of numerous investigations of structure–function relationships performed with various mRNA species. A major limitation is our lack of understanding of the mode of interaction of the mRNA with the ribosome and with the other factors involved in translation.

The sequences of the *unc* TIRs are compared in Table I. In most cases the sequence upstream of the start codon includes the end of the previous cistron. In the cases of *uncE*, *uncF*, and *uncG* there are relatively long intercistronic, noncoding regions. It is not known how the bases shown are spatially organized with respect to each other *in vivo*. Predictions can, however, be made of possible base pairing interactions using algorithms tested on, for example, RNA phage genomes, ribosomal RNA, and transfer RNA (see, e.g., Tinoco *et al.*, 1973; Zuker and Stiegler, 1981). Attention has been drawn to the computer-predicted formation of local secondary structures in the TIRs of *unc* genes (Brusilow *et al.*, 1982; see Fig. 3). In fact local secondary structures are predicted to exist throughout the *unc* operon (the nature of longer-range interactions and their significance can only be guessed at). The existence of local secondary structure specifically involving the TIR, however, could have a special significance. It might render specific recognition elements, such as the Shine–Dalgarno region and/or the translational initiation codon, more inaccessible to components of the initiation complex (i.e., make specific interactions energetically more unfavorable). This was

Table I. TIRs of the *unc* Genes[a]

Source	Sequence
16S rRNA	^{3}AUUCCUCCACUAGGU----
Scherer *et al.*	UU.UUAAAAAUUAAGGAGGUAUAUUAUGAAAAAAAAUUAAAAAACU
uncB	UGUAAUUAACAACAAAGGGUA*AAAGGCAUCAUGGCUUCAGAAAAUAUGAC
uncE	CGUUUUAACUGAAACAAACUGGAGACUGUCAUGGAAAACCUGAAUAUGGA
uncF	AGAACGUUAACUAAAUAGAGGCAUUGUGCUGUGAAUCUUAACGCAACAAU
uncH	UGUCGCUGAACUGUA*AGGAGGGAGGGGCUGAUGUCUGAAUUUAUUACGGU
uncA	GACGUCUUGCAGUCUUA*AGGGGACUGGAGCAUGCAACUGAAUUCCACCGA
uncG	GGCAGGCCGCAAGGCAUUGAGGAGAAGCUCAUGGCCGGCGCAAAAGACAU
uncD	UUA*AACAGGUUAUUUCGUAGAGGAUUUAAGAUGGCUACUGGAAAGAUUGU
uncC	AAAACUUUA*ACGCCUUAAUCGGAGGGUGAUAUGGCAAUGACUUACCACCU

[a]The TIRs (from −30 to +20) of the eight *unc* genes encoding known H^+-ATPase subunits are shown. Shown also, for comparison, is the "consensus" sequence of Scherer *et al.* (1980) deduced from a range of TIRs (not selected on the basis of their efficiencies), which indicates preferred bases at the given positions. Above that is the sequence of the 3′end of the 16S rRNA of *E. coli* which shows complementarity to the Shine–Dalgarno regions on the mRNA. The TIRs are lined up with respect to their start codons (in the case of *uncB* the gene start is uncertain; see Walker *et al.*, 1984). The Shine–Dalgarno regions and upstream stop codons are underlined and marked with an asterisk, respectively.

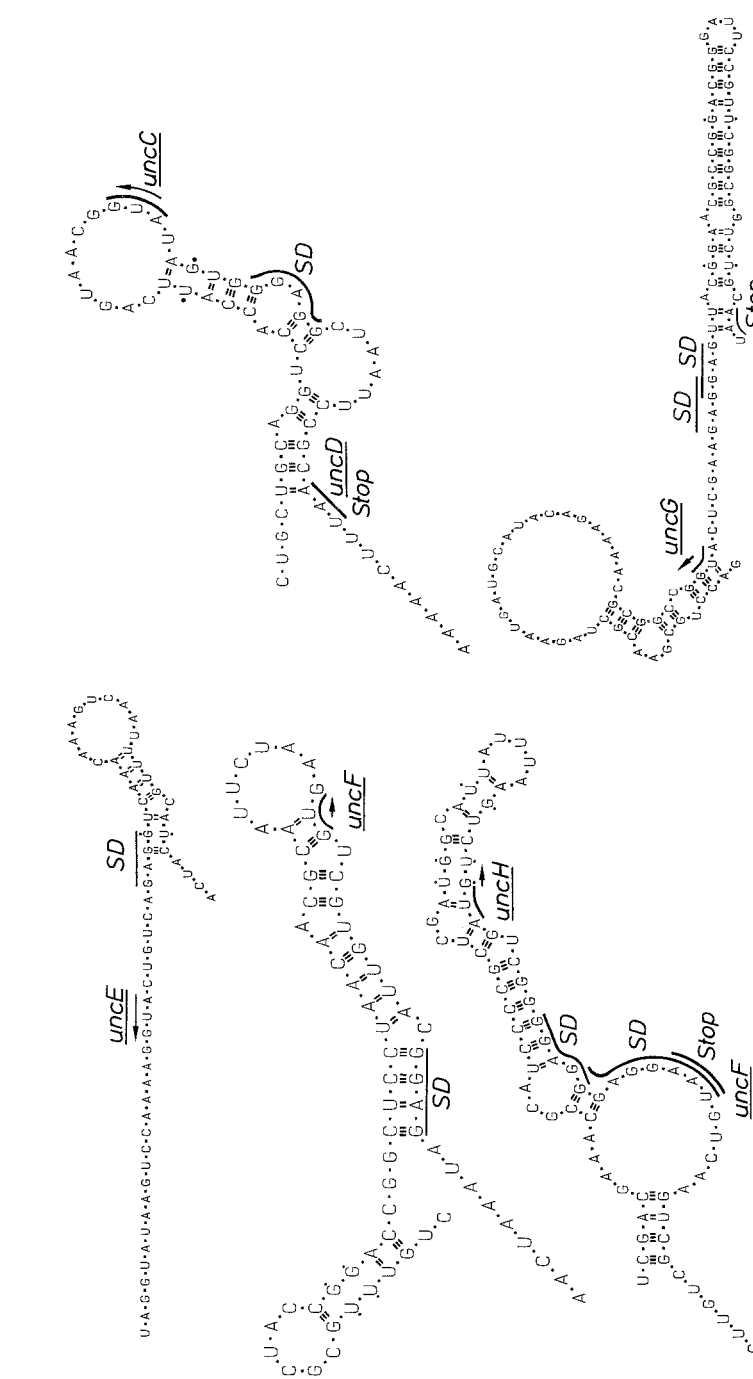

Fig. 3. Predicted secondary structures for the TIRs of *uncE*, *uncF*, *uncH*, *uncG* and *uncC* plotted according to the programme of Zuker and Stiegler (1981). In each case the start codon and area of interaction with the 3′ end of the 16S rRNA (SD region) are indicated. In some cases the stop codon of the upstream cistron is also marked. The ΔG^0 of formation at 37°C (in kcal) of each of the structures shown was calculated to be: *uncE*, −0.4; *uncF*, −7.0; *uncH*, −11.8; *uncG*, −20.5; *uncC*, −6.0.

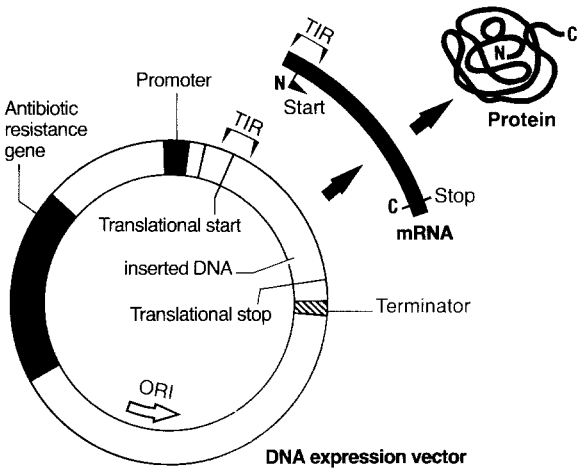

Fig. 4. Simplified scheme of expression from an expression vector, showing the transcriptional promoter and terminator on the plasmid, the TIR and stop codon of the inserted DNA, and the N- and C-termini of the synthesized protein.

indeed proposed as an explanation of the relatively poor levels of expression of *uncF*, *uncH*, and *uncG* (Brusilow *et al.*, 1982).

The above considerations prompt some more specific questions. Do the respective TIRs of the *unc* genes support different efficiencies of translation? Are these efficiencies well matched to the subunit requirements of H^+-ATPase assembly? How does TIR structure determine the efficiency of initiation?

More detailed investigations of the translation of the *unc* genes are most readily performed using plasmids (expression vectors) bearing suitable transcriptional promoter and terminator signals (Fig. 4). Restriction sites located between these two types of signal can be used for the insertion of parts of the *unc* operon. The inclusion of a constant marker gene (e.g., that encoding galactokinase) behind the cloning site(s) but in front of the terminator can provide a means of indirectly monitoring transcriptional activity. A range of (manipulated) DNA fragments can be inserted into vectors of this form. Thus the effects of alterations in TIR sequences upon the relative levels of specific protein (and mRNA) molecules can be investigated.

The Translational Efficiencies of *uncE*, *uncF*, and *uncH*

The approach described was employed in a study of that region of the *unc* operon showing the greatest variation in terms of physical representation

of specific genes in the assembled H^+-ATPase complex (McCarthy *et al.*, 1984, 1985). DNA fragments bearing *uncE*, *uncF*, and *uncH* were inserted into expression plasmids behind either the *trp* + *lac* fusion promoter *tac* (in the vector pDR540; Russell and Bennett, 1982) or the bacteriophage λ promoter p_L (in the vector pJLf101; McCarthy *et al.*, 1985). Transcription from the *tac* promoter in the derivatives of pDR540 ran through the inserted *unc* sequences and continued through the *galK* gene lying behind the cloning site. The respective rates of synthesis of the subunits *c*, *b*, and δ directed by the cloned *unc* sequences could be compared with each other and with expression from the *galK* gene. Experiments performed *in vivo* and *in vitro* indicated that the translational rates directed by the TIRs of *uncE*, *uncF*, and *uncH*, respectively, are in a ratio to each other that compares favorably with the stoichiometry of the corresponding subunits in the assembled H^+-ATPase.

Particularly remarkable was the highly efficient rate of translational initiation directed by the *uncE* TIR. Deletions of the upstream sequence were generated progressively using Bal31 exonuclease, and their effects upon the expression of *uncE* were measured. The results indicated that the full efficiency of the *uncE* TIR was dependent upon the presence of the sequence including at least 20 bases 5′ of the Shine–Dalgarno sequence, i.e., at least −30 with respect to the start codon (see Table I to locate this position). Removal of the bases as far as −11 from the start codon, i.e., still leaving the Shine–Dalgarno region intact, resulted in a reduction in synthesis of subunit *c* by more than a factor of 10. Similar results were obtained with both types of expression vector, which indicated that the effects of progressively removing bases from the 5′ end of the *uncE* TIR could not simply be attributed to the influence of any plasmid-borne structures.

Thus much of the intercistronic sequence lying between *uncB* and *uncE* serves to enhance[4] the efficiency of translation of *uncE* relative to the other *unc* genes. The whole of the intercistronic sequence (from −50 to +1 of *uncE*) was isolated as a fragment from the *unc* operon or copied in the form of a synthetic oligodeoxyribonucleotide (whereby alterations to the sequence could also be introduced; see later). It was then placed in front of the start codons of a range of other genes derived either from *E. coli* or other organisms, including *E. coli unc* genes, *sucC* and *sucD* and the human cDNA sequences encoding interferon β and interleukin 2 (McCarthy *et al.*, 1986; Schauder *et al.*, 1987). Estimations of the translational efficiencies of these gene-TIR

[4]It is, however, inappropriate to refer to this sequence as an "enhancer" of translation, first of all because it lies structurally and functionally at one extreme of a spectrum of TIRs and cannot be cleanly set apart from the others, and secondly because of the already established use of this word in the field of eukaryotic gene expression.

combinations carried by various expression plasmids revealed that the *uncE* TIR sequence acts generally to promote efficient translational initiation, i.e., in combination with many different N-terminal structural gene sequences.

TIR sequences of other genes that are highly expressed (efficiently translated) in *E. coli* were found which show homology to that part of the *uncE* TIR upstream of the start codon (see, e.g., McCarthy *et al.*, 1985, 1986). The overall pattern common to these TIRs comprises a pyrimidine (U)-rich region followed by an interrupted purine (mainly A)-rich region (see the *uncE* TIR in Table I). Moreover, at least the U-rich part of this pattern is evident in the TIRs of an even wider range of genes (especially among bacteriophages; see, e.g., Jay *et al.*, 1982).

The *uncE* TIR is predicted to show little tendency to form secondary structure (Fig. 3) and is therefore expected to have a relatively "open" conformation. Both this "openness" and the particular sequence of the *uncE* TIR must be responsible for the high frequency of initiation by ribosomes. A number of different RNA–RNA and/or RNA–protein interactions could be involved in achieving efficient initiation. The mRNA could theoretically interact with sequences outside of the anticodon of the initiator tRNA (see, e.g., Ganoza *et al.*, 1985) or with parts of the 16S rRNA outside of the Shine–Dalgarno region. Relevant mRNA–protein interactions could involve a protein factor or ribosomal subunit (for example S1, which has a high affinity for polypyrimidine stretches; Goelz and Steitz, 1977). Recent investigations of the effects of specific substitutions in the *uncE* TIR provide support for the hypothesis that the U-rich region is important to the attainment of the highest rates of translation (Schauder and McCarthy, unpublished data).

The two genes lying immediately after *uncE* in the operon are much more poorly expressed, both within the operon structure and when cloned individually (Brusilow *et al.*, 1982; McCarthy *et al.*, 1985). When *uncF* was expressed in its normal position downstream of *uncE* (on a plasmid), its translation was partly dependent upon the translation of *uncE* (McCarthy *et al.*, 1985). The extremely low rate of expression of the individually cloned *uncH* gene relative to expression of *uncE*, *uncF*, and the *galK* gene on the expression plasmid also suggested that *uncH* translation is partly dependent upon translation of *uncF*. The simplest interpretation of these data is that *uncE*, *uncF*, and also probably *uncH*, are (albeit weakly) translationally coupled. One explanation of this phenomenon is that ribosomes at the end of the upstream cistron help to improve the recognition of, and initiation at, the downstream TIR for other ribosomes by destabilizing secondary structure within it. On the other hand, to what extent ribosomes can "re-initiate," at a downstream cistron after completing the previous one, is in general not clear. Translational coupling to various degrees has been observed in a number of other operons [for example, ribosomal protein operons (Lindahl

and Zengel, 1982), the *trp* operon (Oppenheim and Yanofsky, 1980), and the *thr* operon (Little *et al.*, 1986)]. How the "tightness" of coupling is related to the length and nature of the intervening (or overlapping) sequence between two cistrons is still not understood (compare, e.g., Dennis, 1984; Little *et al.*, 1986).

Klionsky *et al.* (1986) engineered two types of alteration in the *uncF* TIR by means of *in vitro* mutagenesis in order to investigate the relationship between structure and function. In the first case they substituted the two Gs at -3 and -5, respectively (see Table I), with two Cs, thereby reducing the stability of the secondary structure predicted for the TIR (a structure very similar to that shown in Fig. 3).[5] In the second case they changed the GUG start codon to an AUG. The mutated *uncF* genes were inserted as part of a fragment also bearing *uncE* and *uncH* into pBR322, and expression from these constructs was compared with the wild type sequence in a minicell system. In both instances the rate of synthesis of *uncF* was increased relative to that of the wild type by a factor of 2–3, and the rate of synthesis of *uncH* was increased by a factor of 1–2 (as would be expected if the two genes are translationally coupled, compare McCarthy *et al.*, 1985). Interpretation of the first experiment is complicated by the fact that the G/C substitutions eliminate the start codon of a three codon reading frame overlapping with the *uncF* cistron. It is not clear to what extent the expression of *uncF* is influenced by the presence of this "pre-cistron" as opposed to the presence of base-pairing within the TIR. In previous experiments (Simoni, 1984) the expression of *uncH* and of the genes lying downstream of it could be increased by a factor 1.5 *in vitro* by adding a poly-G oligomer. It was suggested that the oligomer disrupted the G-C rich secondary structure in the *uncH* TIR (see Fig. 3), thereby facilitating the access of ribosomes. This mechanism of action remains however, to be substantiated. The result of the second mutagenesis experiment is consistent with the observation made with other systems that the substitution of AUG for a GUG start codon can allow more efficient translation (Klionsky *et al.*, 1986; but see later).

In further experiments synthetic oligodeoxyribonucleotides bearing the *uncE* TIR sequence from -50 to -1, or from -50 to -10, were inserted at various distances in front of *uncF* and *uncH* borne separately on expression plasmids (Fig. 5). The effects of the introduction of this sequence upon expression were investigated both *in vitro* and *in vivo* (McCarthy, unpublished results). The relative amounts of *uncF* and *uncH* mRNA species produced by the different constructs were checked by Northern- and dot-blotting so that

[5] This figure highlights one of the problems with predicting secondary structure, namely that more than one possible structure within a given energy range can often be predicted for a given RNA sequence. The structure given here for *uncG*, for example, differs from that in Brusilow *et al.*, 1982.

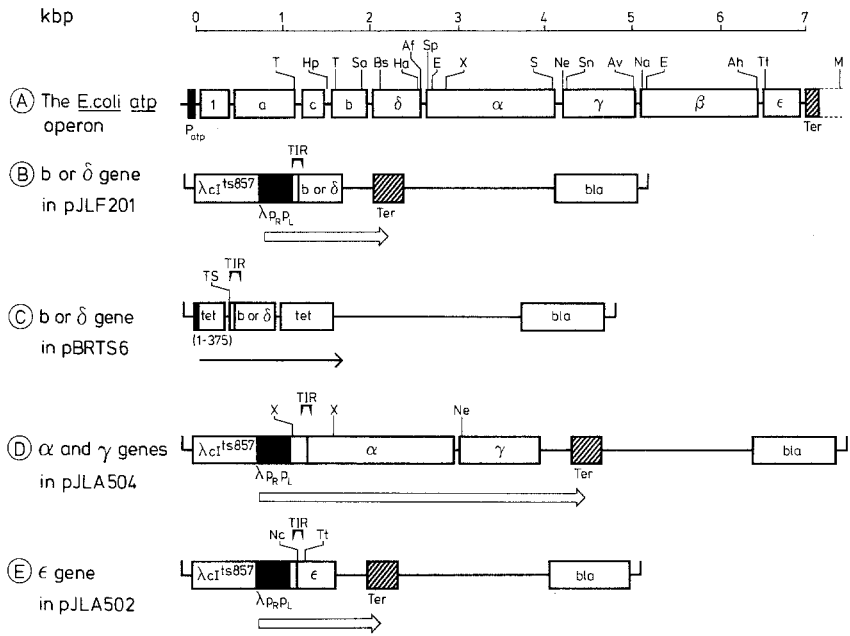

Fig. 5. Examples of expression vector derivatives constructed using *unc* genes. *uncF* and *uncH* were recombined with various TIRs and inserted either behind the bacteriophage λ promoters p_R and p_L in the expression vector pJLF201 or behind the *tet* promoter in pBRTS6. In the latter case, an operon fusion was effectively formed in which the genes were inserted behind the N-terminal part of the *tet* gene followed by six translational stop codons (in all three reading frames). *uncA* and *uncG* were inserted together on one DNA fragment into pJLA504. The latter vector bears the *uncE* TIR sequence from − 50 to − 1 just upstream of its SphI cloning site. *uncA* has an SphI site directly at its translational start codon. *uncC* was inserted into a similar vector (pJLA502), either with its own complete TIR, or so that the *uncE* TIR sequence was fused onto its start codon. The partial restriction map of the *unc* operon shows the positions of restriction sites for the following enzymes: Af, AflII; Ah, AhaIII; Av, AvaI; Bs, BstEII; E, EcoRI; Ha, HaeII; Hp, HpaI; M, MluI; Na, NarI; Ne, NaeI; S, SalI; Sa, Sau3A; Sn, SnaBI; Sp, SphI; T, TaqI; Tt, Tth111I.

the effects specifically related to translation could be assessed. The presence of the *uncE* TIR sequence induced an up to tenfold increase in the efficiency of translation of both genes, the exact effect depending on its position relative to the translational start codon. The expression of two of these TIR-gene recombinant constructs is compared with expression from *uncF* and *uncH* together with their respective natural TIRs in Fig. 6. In all of the *uncF* constructs the GUG start codon was retained, and in one of the most efficiently translated TIR–*uncH* combinations, the usual AUG had been changed to a GUG. Apparently, the presence of this start codon does not prevent changes elsewhere in the TIR from inducing a substantial net increase in translational efficiency.

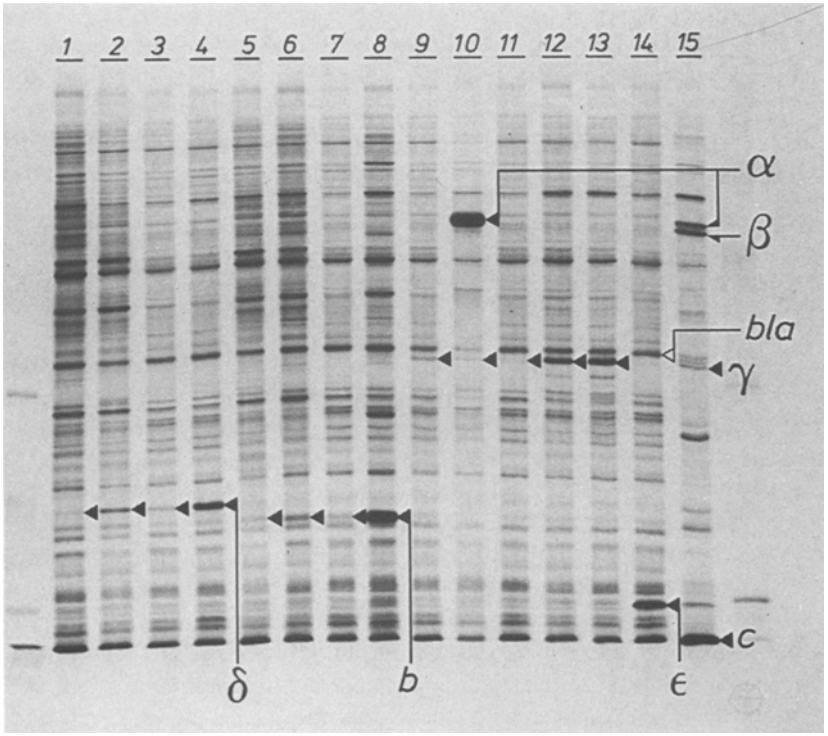

Fig. 6. A "pulse-chase" labelling experiment performed with whole cells of LM1241, an *unc* deletion mutant of *E. coli* (Δ*uncIBEFHAGD* 702). The cells were grown in minimal medium and given a 100-sec pulse of ^{35}S-L-methionine, followed by a 100-sec chase of unlabelled L-methionine. The fluorograph shows the results obtained with cells containing the following plasmids: lane 1, pBRTS6 + *uncH* with its normal TIR; lane 2, pBRTS6 + *uncH* with the *uncE* TIR sequence upstream of its start codon; lanes 3 and 4, as 1 and 2, but in pJLf201; lane 5, pBRTS6 + *uncF* with its normal TIR; lane 6, pBRTS6 + *uncF* with the *uncE* TIR sequence; lanes 7 and 8, as 5 and 6, but in pJLf201; lane 9, *uncG* with its normal TIR in pJLf504; lane 10, *uncA* and *uncG* with its normal TIR in pJLA504 as shown in Fig. 5; lane 11, the same construct as lane 10, except that the TIR of *uncA* was removed (see text); lanes 12 and 13, two different combinations of *uncG* plus the *uncE* TIR sequence in pJLA504; lane 14, *uncC* with its own TIR in pJLA502; lane 15, pBJC1888 from K. von Meyenburg, which directs the synthesis of all eight H^+-ATPase subunits. The plasmids bearing λ promoters were expressed after induction at 42°C; derivatives of pBRTS6 were expressed at 37°C.

Thus manipulations in the TIR sequences upstream of *uncF* and *uncH*, respectively, are sufficient to allow large increases in the expression of these genes. This enhanced expression is attributable (compare Klionsky *et al.*, 1986) to an increased efficiency (frequency) of translational initiation brought about by the introduction of DNA from the *uncE* TIR. The relative efficiency of translation of the recombinant constructs was only very approximately

(inversely) correlated with the stability of the secondary structures predicted for the TIR gene combinations. This type of analysis is, however, very difficult to make because of the lack of an independent means of assessment of the effects of changes in base sequence. These changes included, for example, variations in the position and sequence of the Shine–Dalgarno region.

The Expression of *uncA*, *uncG*, *uncD* and *uncC*

Differing translation rates of the genes downstream of *uncH* are evident both *in vitro* and *in vivo* (Brusilow *et al.*, 1982; McCarthy, unpublished data; see, e.g., Fig. 6). *uncG* cloned by itself on an expression plasmid is poorly translated when initiation is controlled by its natural TIR (Figs. 5 and 6). There is potential for relatively stable secondary structure in the *uncG* TIR which could negatively influence initiation (Fig. 3). The efficiency of its translation is apparently little affected by translation of the *uncA* cistron when this is located in its normal position upstream of it. The low level of translational coupling is particularly striking when *uncA* and *uncG* are inserted into the expression vector pJLA504 using the SphI site at the beginning of *uncA* (Fig. 5). In this case the natural sequence upstream of *uncA* is replaced by the *uncE* TIR sequence with the result that *uncA* is very efficiently translated and expressed to levels of up to 50% of total cellular protein. This hardly seems to affect expression of *uncG*, which is still translated at about the same level as when it is present alone with its natural TIR (and the last 120 bp of *uncA*) on the same vector (Fig. 6). However, when almost all of *uncA* is present upstream of *uncG* except that the *uncA* TIR has been removed (so that *uncA* cannot be translated), *uncG* is more poorly expressed, possibly because of enhanced rates of premature transcriptional termination within the long, untranslated *uncA* sequence (Fig. 6).

Introduction of part of the *uncE* TIR upstream of the *uncG* start codon also led to increased expression of this gene (Fig. 6). By combining the *uncE* TIR sequence with parts of the *uncG* TIR in various ways (using synthetic oligodeoxyribonucleotides) it was possible to generate a range of TIRs with different primary and (predicted) secondary structures, differing by varying degrees from the native *uncG* TIR (see Fig. 3). The translational efficiency of *uncG* could be increased by more than a factor 10.

uncC cloned individually with its natural TIR is not as poorly expressed *in vivo* as *uncF*, *uncG*, or *uncH* (Fig. 6; and compare Brusilow *et al.*, 1982). It could nevertheless still be more efficiently translated in combination with the *atpE* TIR sequence (Fig. 5; expression data not shown). A higher translational efficiency directed by the natural TIR could possibly compensate for

underrepresentation of a functional *uncC* cistron at the mRNA level (see, e.g., earlier). The efficiency of *uncC* translation is in fact sufficiently high to lead to the synthesis of more than one copy of subunit ε per H^+-ATPase complex. The *uncC* TIR is predicted not to form highly stable secondary structure (Fig. 3).

Overexpression of the *unc* Genes

It has already been described how individual *unc* genes can be overexpressed using multicopy plasmids, especially where the efficiency of translation has been increased by manipulation of the TIRs. High levels of overexpression of the total *unc* operon, maintaining the normal ratios of expression between the component genes, can be achieved simply by allowing it to be transcribed from the bacteriophage λp_R promoter on a high copy number plasmid (von Meyenburg *et al.*, 1984b; see also Fig. 6). H^+-ATPase could be produced to reach 18–23% of total cellular protein. This level of overproduction strongly inhibited division and growth of *E. coli* K12 and resulted in the formation of membrane cisterns and vesicles within the cells. More detailed investigations of the effects of overexpression (von Meyenburg *et al.*, 1985) showed that it is primarily overproduction of subunit *a* that is responsible for the observed effects (compare Kanazawa *et al.*, 1984). Overproduction of subunit *a*, either alone or in combination with one or more of the other subunits, not only inhibits growth but also protein synthesis, as well as partially collapsing the protonmotive force. The simplest explanation of this is that subunit *a* can exercise ionophoric activity when overproduced (von Meyenburg *et al.*, 1985). The overproduction of the other *unc* subunits, even in some cases (*c* and α) up to levels of 40–50% of total cellular protein, is tolerated much better by *E. coli* (von Meyenburg *et al.*, 1985; McCarthy, unpublished observations).

The overexpression of the *unc* genes achieved using a strong promoter on a high copy number plasmid is to varying degress limited by the efficiency of translation (see Fig. 6). The large increases in expression resulting from the alterations in the *unc* TIRs already described are most logically explained by improved efficiencies of translational initiation. This argues against control being exercised by translational elongation in the chromosomally located *unc* operon. In general terms, elongation acts probably only rarely to control the overall rate of translation of chromosomal genes since the rate of ribosomal translocation is usually fast in comparison with the frequency of ribosomal initiation (compare, e.g., Talkad *et al.*, 1976; Kennel and Riezman, 1977; Maaloe, 1979; Pedersen, 1984). However, the elongation rate could theoretically become limiting when the initiation efficiency is increased (by,

e.g., manipulating the TIR). Moreover, restrictions on elongation could be imposed as the result of excessive draining of tRNA pools where genes are present on multicopy plasmids and the efficiency of transcription (and of translational initiation) is high. The above might explain why, despite various manipulations performed with their TIRs, the *unc* genes showing codon usages more typical of poorly expressed *E. coli* genes (*uncF*, *uncH*, and *uncG*) could not be overexpressed to the same degree as *uncE* and *uncA* (Fig. 6).

Conclusions and Future Prospects

The investigations described in this review have provided an initial picture of the control of gene expression in the *E. coli unc* operon. The primary mode of control involves an interplay of different mechanisms at the translational level. Thus the different efficiencies of initiation promoted by the individual TIRs, and to a limited extent translational coupling, determine the overall rates of translation. Closer examination of the TIRs has revealed the presence of structures that both enhance (especially in the *uncE* TIR) and reduce (e.g., in the *uncF*, *uncH*, and *uncG* TIRs) the efficiency of initiation (or of translational coupling). It still remains to be determined to what extent primary, as opposed to secondary (or tertiary), structure in many of these TIRs is rate controlling, and indeed how this rate control comes about.

The codon usage patterns of the *unc* operon remain unexplained. There may be growth conditions (as yet unidentified) under which codon usage influences *unc* gene expression. However, given the results described in this review, it must be assumed that the rate of translational elongation does not control expression of the chromosomal *unc* genes. A full discussion of the possible significance of codon usage (and codon context) is outside the scope of this review (see, e.g., Kurland, 1987; Trifonov, 1987).

No evidence has been obtained that feedback control is exercised by H^+-ATPase subunits. Investigations of the rates of translation of the plasmid-borne *unc* genes were, for example, performed *in vitro* and *in vivo* in both unc^+ and unc^- backgrounds (McCarthy, unpublished data; see Fig. 6). There were no apparent differences in the characteristics of expression, irrespective of the stage of accumulation of H^+-ATPase subunits (synthesized either under the direction of chromosomal or plasmid-borne *unc* genes). However, further experiments involving fusions of *unc* TIRs onto a marker gene (*lacZ*) need to be performed to confirm this result. The influence of other types of effectors has also not been ruled out. Numerous examples of RNA binding proteins that can inhibit or enhance translation are known (see, e.g., Kozak, 1983; Mahajna *et al.*, 1986; Miller *et al.*, 1987). The influence of an mRNA-binding

protein on *unc* gene expression could have been missed, especially if this protein was constitutively expressed.

The modes of synthesis and degradation of the polycistronic mRNA are not fully settled. There is particular uncertainty about the promoter-distal part of the operon; there may, for example, be reduced amounts of functionally active *uncC* mRNA. Further investigations of the physical concentrations and functional activities of specific parts of the *unc* mRNA are therefore required. The influence of premature transcriptional termination events and of mRNA degradation processes upon *unc* expression need to be more precisely defined (compare Stanssens *et al.*, 1986; Brawerman, 1987). The mode of degradation of the mRNA is of particular interest; how do progressive $3' \rightarrow 5'$ exonucleolytic degradation and endonucleolytic cleavages, respectively, participate in the overall process (compare Cannistraro *et al.*, 1986; Newbury *et al.*, 1987)?

All of these lines of investigation are currently underway. The results from them should tell us more about control in the *unc* operon as well as helping us to understand other systems in *E. coli.*

Acknowledgments

I would like to thank W. Sebald, H. U. Schairer, and K. von Meyenburg for discussions in the past. O. Michelsen and K. von Meyenburg kindly provided the plasmids pBJC1888 and CM2443, as well as the *E. coli* strain LM1241. I am also grateful to D. Lincoln for help with the computing, and to A. Eberle for assistance with the preparation of this manuscript.

References

Amzel, M. (1981). *J. Bioenerg. Biomembr.* **13**, 109–121.
Brawermann, G. (1987). *Cell* **48**, 5–6.
Brusilow, W. S. A., Klionsky, D. J., and Simoni, R. D. (1982). *J. Bacteriol.* **151**, 1363–1371.
Brusilow, W. S. A., Porter, A. C. G., and Simoni, R. D. (1983). *J. Bacteriol.* **155**, 1265–1270.
Butlin, J. D., Cox, G. B., and Gibson, F. (1971). *Biochem. J.* **124**, 75–81.
Cannistraro, V. J., Subbarao, M. N., and Kennell, D. (1986). *J. Mol. Biol.* **192**, 257–274.
Cozens, A. L., and Walker, J. E. (1987). *J. Mol. Biol.* **194**, 359–383.
Dennis, P. P. (1984). *J. Biol. Chem.* **259**, 3202–3209.
Downie, J. A., Gibson, F., and Cox, G. B. (1979). *Annu. Rev. Biochem.* **48**, 103–131.
Downie, J. A., Langman, L., Cox, G. B., Yanofsky, C., and Gibson, F. (1980). *J. Bacteriol.* **143**, 8–17.
Fillingame, R. H. (1981). *Curr. Top. Bioenerg.* **11**, 35–106.
Foster, D. L., and Fillingame, R. H. (1979). *J. Biol. Chem.* **254**, 8230–8236.
Foster, D. L., and Fillingame, R. H. (1982). *J. Biol. Chem.* **257**, 2009–2015.
Friedl, P., Friedl, C., and Schairer, H. U. (1979). *Eur. J. Biochem.* **100**, 175–180.
Futai, M., and Kanazawa, H. (1983). *Microbiol. Rev.* **47**, 285–312.
Ganoza, M. C., Marliere, P., Kofoid, E. C., and Louis, B. G. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 4587–4591.

Gay, N. J. (1984). *J. Bacteriol.* **158**, 820–825.
Gay, N. J., and Walker, J. E. (1981). *Nucleic Acids Res.* **9**, 3919–3926.
Gibson, F. (1983). *Proc. R. Soc. London Ser. B* **215**, 1–18.
Gibson, F., Downie, J. A., Cox, J. B., and Radik, J. (1978). *J. Bacteriol.* **134**, 728–736.
Goelz, S., and Steitz, J. A. (1977). *J. Biol. Chem.* **252**, 5177–5179.
Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., and Stormo, G. (1981). *Annu. Rev. Microbiol.* **35**, 365–403.
Grantham, R., Gautier, D., Gouy, M., Jacobzone, M., and Mercier, R. (1981). *Nucleic Acids Res.* **9**, 43–74.
Grosjean, H., and Fiers, W. (1982). *Gene* **18**, 199–209.
Hansen, F. G., Nielsen, J., Riise, E., and von Meyenburg, K. (1981). *Mol. Gen. Genet.* **183**, 463–472.
Holm, L. (1986). *Nucleic Acids Res.* **14**, 3075–3087.
Ikemura, T. (1981a). *J. Mol. Biol.* **146**, 1–21.
Ikemura, T. (1981b). *J. Mol. Biol.* **151**, 389–409.
Jay, E., Seth, A. K., Rommens, J., Sood, A., and Jay, G. (1982). *Nucleic Acids Res.* **10**, 6319–6329.
Jones, H. M., Brajkovich, C. M., and Gunsalus, R. P. (1983). *J. Bacteriol.* **155**, 1279–1287.
Kanazawa, H., Kiyasu, T., Noumi, T. and Futai, M. (1984). *J. Bacteriol.* **158**, 300–306.
Kennell, D., and Riezman, H. (1977). *J. Mol. Biol.* **114**, 1–21.
Klionsky, D. J., Skalnik, D. G., and Simoni, R. D. (1986). *J. Biol. Chem.* **261**, 8096–8099.
Kozak, M. (1983). *Microbiol. Rev.* **47**, 1–45.
Kurland, C. G. (1987). *Trends Biochem. Sci.* **12**, 126–128.
Lindahl, L., and Zengel, J. M. (1982). *Adv. Genet.* **21**, 53–121.
Little, S., Campbell, K., Hyde, S., and Robinson, M. K. (1986). *Abstracts Microbe* **86**, P.B, 21–10.
Lünsdorf, H., Ehrig, K., Friedl, P., and Schairer, H. U. (1984). *J. Mol. Biol.* **173**, 131–136.
Maaloe, O. (1979). In *Biological Regulation and Development* (Goldberger, R. F., ed.), Plenum Press, New York, pp. 487–542.
Mahajna, J., Oppenheim, A. B., Rattray, A., and Gottesman, M. (1986). *J. Bacteriol.* **165**, 167–174.
McCarthy, J. E. G., Schairer, H. U., and Sebald, W. (1984). EBEC Rep. 3, pp. 587, 588.
McCarthy, J. E. G., Schairer, H. U., and Sebald, W. (1985). *EMBO J.* **4**, 519–526.
McCarthy, J. E. G., Sebald, W., Gross, G., and Lammers, R. (1986). *Gene* **41**, 201–206.
Miki, T., Hiraga, S., Nagata, T., and Yura, T. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 5099–5103.
Miller, E. S., Karam, J., Dawson, M., Trojanowska, M., Gauss, P., and Gold, L. (1987). *J. Mol. Biol.* **194**, 397–410.
Nelson, N. (1981). *Curr. Top. Bioenerg.* **11**, 1–34.
Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D., and Higgins, C. F. (1987). *Cell* **48**, 297–310.
Nielsen, J., Jorgensen, B. B., Hansen, F. G., Petersen, P. E., and von Meyenburg, K. (1982). EBEC Rep. 2, pp. 611–612.
Nielsen, J., Jorgensen, B. B., von Meyenburg, K., and Hansen, F. G. (1984). *Mol. Gen. Genet.* **193**, 64–71.
Oppenheim, S. D., and Yanofsky, C. (1980). *Genetics* **95**, 785–795.
Pedersen, S. (1984). *EMBO J.* **3** 2895–2898.
Porter, A. G. G., Brusilow, W. S. A., and Simoni, R. D. (1983). *J. Bacteriol.* **155**, 1271–1278.
Ray, P. N., and Pearson, M. L. (1974). *J. Mol. Biol.* **85**, 163–175.
Ray, P. N., and Pearson, M. L. (1975). *Nature (London)* **253**, 647–650.
Russell, D. R., and Bennett, G. N. (1982). *Gene* **20**, 231–243.
Schauder, B., Blöcker, H., Frank, R., and McCarthy, J. E. G. (1987). *Gene* **52**, 279–283.
Scherer, G. F. E., Walkinshaw, M. D., Arnott, S., and Morré, D. J. (1980). *Nucleic Acids Res.* **8**, 3895–3907.
Schneider, E., and Altendorf, K. (1982). *Eur. J. Biochem.* **126**, 149–153.
Senior, A. E. (1979). In *Membrane Proteins in Energy Transduction* (Capaldi, R. A., ed.), Marcel Dekker, New York, pp. 233–276.

Simoni, R. D. (1984). In *H^+-ATPase (ATP Synthase): Structure, Function, Biogenesis. The F_0F_1 Complex of Coupling Membranes* (Papa, S., Altendorf, K., Ernster, L., and Packer, L., eds.), ICSU Press,, Adriatica Editrice, Bari, Italy, pp. 77–88.
Stanssens, P., Remaut, E., and Fiers, W. (1986). *Cell* **44**, 711–718.
Steitz, J. A. (1979). In *Biological Regulation and Development* (Goldberger, R. F., ed.), Plenum Press, New York, pp. 349–399.
Talkad, V., Schneider, E., and Kennell, D. (1976). *J. Mol. Biol.* **104**, 299–303.
Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973). *Nature New Biol.* **246**, 40–41.
Trifonov, E. N. (1987). *J. Mol. Biol.* **194**, 643–652.
Varenne, S., Buc, I., Lloubes, R., and Lazdunski, C. (1984). *J. Mol. Biol.* **180**, 549–576.
von Meyenburg, K., Hansen, F. G., Riise, E., Bergmans, H. E. N., Meijer, M., and Messer, W. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 121–128.
von Meyenburg, K., Jorgenson, B. B., Nielsen, J., Hansen, F. G., and Michelsen, O. (1982a). *Tokai J. Exp. Clin. Med. Suppl.* **7**, 23–31.
von Meyenburg, K., Jorgenson, B. B., Nielsen, J., and Hansen, F. G. (1982b). *Mol. Gen. Genet.* **188**, 240–248.
von Meyenburg, K., Nielsen, J., Jorgensen, B. B., Michelsen, O., Hansen, F. G., and Van Deurs, B. (1984a). EBEC Rep. 3, pp. 67–68.
von Meyenburg, K., Jorgensen, B. B., and van Deurs, B. (1984b). *EMBO J.* **3**, 1791–1797.
von Meyenburg, K., Jorgensen, B. B., Michelsen, O., Sorensen, L., and McCarthy, J. E. G. (1985). *EMBO J.* **4**, 2357–2363.
Walker, J. E., Saraste, M., and Gay, N. J. (1984). *Biochim. Biophys. Acta* **768**, 164–200.
Zuker, M., and Stiegler, P. (1981). *Nucleic Acids Res.* **9**, 133–148.