Volume 307, number 1, 62–65 FEBS 11173 July 1992

Minireview

Prokaryotic polyprotein precursors

Linda Thöny-Meyer[a], August Böck[b] and Hauke Hennecke[a]

[a]*Mikrobiologisches Institut, Eidgenössische Technische Hochschule, ETH-Zentrum, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland* and [b]*Lehrstuhl für Mikrobiologie, Universität München, Maria-Ward-Strasse 1a, D-8000 München 19, Germany*

Received 21 April 1992

Polyproteins have been found only recently in prokaryotes. The four known examples of single bacterial genes encoding precursors that are posttranslationally processed into two mature proteins are addressed here with respect to (i) their genomic arrangement, (ii) the sites of proteolytic processing, (iii) the relevant proteases, (iv) their maturation pathway, and (v) the function of the mature proteins. How these polyproteins may have evolved is also discussed.

Gene-protein relationship; Maturation; Polyprotein; Precursor; Proteolytic processing

1. INTRODUCTION

The concept of a one-to-one relation between genes and proteins was proposed in the early 40's by George Beadle and Edward Tatum in the 'one gene–one enzyme' hypothesis [1] and is still found in textbooks, although slightly modified as the 'one gene–one polypeptide' hypothesis. Yet, it is now well established that there are exceptions to this rule, i.e., a single gene may encode more than one mature protein. Different mechanisms are known to be involved in creating multiple gene products from a single mRNA (we will not discuss the formation of multiple mRNAs from a single gene here). They include in-phase overlapping genes with either readthrough of a termination codon or the use of internal, in-phase translation start sites [2], ribosomal frameshifting [3], and polyprotein processing [4]. Until fairly recently, it was believed that the proteolytic release of functional proteins from polyprotein precursors – the subject of this review – is a hallmark of eukaryotes [5]. In the last few years, however, it has become evident that the phenomenon of polyprotein synthesis and cleavage also exists in different prokaryotes. Here we describe the four bacterial polyprotein precursors known thus far and compare their maturation into two separate, mature products: *Escherichia coli* penicillin G acylase, *Bradyrhizobium japonicum* cytochrome bc_1, *Bacillus subtilis* spore coat proteins and *Bacillus polymyxa* amylase.

Correspondence address: H. Hennecke, Mikrobiologisches Institut, Eidgenössische Technische Hochschule, ETH-Zentrum, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland. Fax: (41) (1) 2620647.

2. *E. coli* PENICILLIN G ACYLASE

The first example of a prokaryotic polyprotein that was discovered concerns the *E. coli* penicillin G acylase, which catalyses the conversion of penicillin G to phenylacetic acid and 6-aminopenicillanic acid, a precursor for semisynthetic penicillins. The enzyme is composed of two dissimilar subunits α and β with molecular masses of 23 and 69 kDa, respectively [6], and is functional only as an $\alpha\beta$ heterodimer. Both subunits are encoded by a single gene (*pac*) [7] whose nucleotide sequence can be translated into a polypeptide of 840 amino acids. This open reading frame encodes a precursor polypeptide that is proteolytically processed in the course of its maturation [8,9]. The four domains of this precursor are depicted in Fig. 1. The first 26 amino acids match a typical signal sequence that is responsible for transport of the protein into the periplasm [10]. The next 209 amino acids represent the α subunit, followed by a 54-amino-acid spacer peptide that is removed during processing. The C-terminal domain is the β subunit consisting of 557 amino acids. Upon sequencing the N and C termini of the mature subunits from the purified proteins the exact processing sites were determined.

Maturation of the acylase precursor was analyzed in vitro and in vivo in order to unravel the processing mechanism [9,11]. The signal peptide-directed translocation of the 98 kDa precursor into the periplasm is followed by a first cleavage event, in which the signal peptide is proteolytically removed, presumably by signal peptidase. The next cleavage occurs at the N terminus of the β subunit, yielding a free 69 kDa β subunit and a 29 kDa α subunit with a C-terminal extension, the former spacer peptide. Finally, the spacer peptide is removed.

Published by Elsevier Science Publishers B.V.

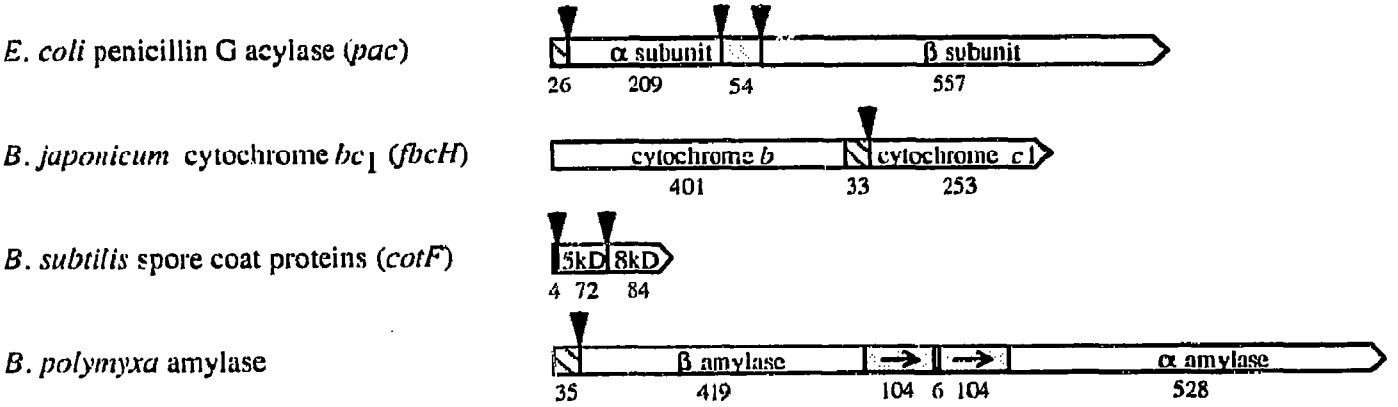

Fig. 1. Arrangement of bacterial polyprotein precursors discussed in this review. The entire primary translation products as predicted from the respective coding regions for the four polyproteins are shown. Moieties with different structural or functional characteristics are separated by vertical lines. The corresponding sizes of these sections are given in amino acids. Vertical arrowheads point to the identified processing sites. Striped sections represent signal peptides, dotted sections represent linker peptides. The horizontal arrows in the linker peptide of the *B. polymyxa* amylase designate a direct repeat. See text for references.

The introduction of mutational alterations into the two subunits and into the spacer peptide helped uncover the requirements of the following structural features for proper maturation: (i) appropriate size of the spacer peptide, (ii) intact C terminus of the polyprotein precursor, and (iii) appropriate conformation of the precursor [11]. It was proposed that the spacer peptide may direct the folding of the precursor into a processing-competent molecule. Studies on unfolding and refolding characteristics of penicillin G acylase supported this idea and suggested that the α peptide constitutes a folding domain for the otherwise relatively insoluble β subunit [12].

The question of whether or not maturation of penicillin acylase requires a special protease was addressed by expressing the *E. coli pac* gene in different Gram-negative bacteria, in all of which posttranslational processing occurred. This suggests that proteolytic cleavage is either catalyzed by protease(s) also present in strains having no *pac* gene or is an autocatalytic reaction [11]. Expression of the *pac* gene in an *E. coli* strain with a leaky outer membrane results in the release of about equal amounts of mature enzyme ($\alpha\beta$ dimer) and of signal peptide-less precursor into the culture medium (C. Keilmann and A. Böck, unpublished data). This allows the conclusion that processing of the precursor occurs in the periplasm and is not coupled to export.

Similar polyprotein precursors for penicillin acylase or related enzymes have been found in other bacteria. The *Kluyvera citrophila* penicillin acylase is also an $\alpha\beta$ heterodimer that is derived from a single precursor with an N-terminal signal sequence and a 54-amino-acid spacer peptide [13]. The protein is homologous to the *E. coli* penicillin acylase with 87% identical amino acids. The establishment of the N- and C-terminal sequences of the mature subunits revealed identical processing sites. It is assumed, therefore, that the *E. coli* and *K. citrophila pac* genes are derived from a common ancestral gene and that the mechanism of maturation is conserved. A mutation changing the glycine in position 21 of the β subunit into glutamic acid gave rise to a *pac* gene-derived precursor polypeptide which could be translocated into the periplasm, but did not undergo maturation, emphasising again the separation between export and processing [14].

Another example is the *Pseudomonas* sp. GK16 cephalosporin acylase [15]. It is synthesized as a 74 kDa precursor, exported into the periplasm owing to its 29-amino-acid signal sequence, which is then cleaved to result in a 70 kDa polypeptide. In the periplasm, posttranslational processing into a 16 kDa α and a 54 kDa β subunit takes place. In this case, only the processing site for release of the β subunit is known. Since the first two N-terminal amino acid residues of the mature *Pseudomonas* and *E.coli* β subunits are identical (Ser, Asn) they may represent structural features of the protease recognition site. From the sizes of the mature subunits it can be speculated that a spacer peptide separates the two subunits in the precursor. However, the details and order of maturation steps of the cephalosporin acylase must be experimentally investigated before conclusions about a processing pathway similar to that of *E. coli* penicillin G acylase can be drawn.

3. *Bradyrhizobium japonicum* CYTOCHROME bc_1

The bacterial ubiquinol-cytochrome *c* oxidoreductase or cytochrome bc_1 complex consists of three subunits, the Rieske iron sulfur protein, cytochrome *b* and cytochrome c_1. These proteins are normally encoded by the three genes *fbcF*, *fbcB* and *fbcC* [16].

In *B. japonicum*, however, this respiratory enzyme complex is encoded by only two genes: *fbcF*, encoding the Rieske protein, and *fbcH*, encoding a 687-amino-

acid precursor polypeptide whose N-terminal moiety later becomes cytochrome *b*, and its C-terminal moiety becomes cytochrome c_1 after proteolytic processing [17]. The characteristics of the *fbcH* primary gene product are shown in Fig. 1. The regions corresponding to apocytochromes *b* (401 amino acids) and c_1 (253 amino acids) are linked by a peptide of 33 amino acids which shows all of the characteristics of a typical signal sequence, except that its location is protein-internal rather than N-terminal. Mature cytochromes *b* and c_1 were detected in Western blots using antibodies directed against synthetic peptides of the *b* and c_1 subunits; however, the putative precursor was not detected, indicating that processing may occur rapidly [18].

N-terminal sequencing of the mature cytochrome c_1 allowed the identification of the cleavage site which perfectly matched a predicted recognition site for signal peptidase (AlaArgAla) at the end of the 33-amino-acid internal signal peptide. Mutational alterations of this site resulted in an unprocessed, 56 kDa FbcH precursor which was immunologically detected both with anti-*b* and anti-c_1 sera [18]. It was thus concluded that (i) the predicted signal peptidase recognition site is a functional processing site in the FbcH precursor and (ii) processing at this position is specific. Interestingly, the unprocessed mutant precursor did not appear to inhibit assembly of the bc_1 complex, nor did it block the covalent binding of heme to the cytochrome c_1 moiety [18]. These two steps in the maturation pathway of the cytochrome bc_1 complex must, therefore, either take place before, or independently of, precursor processing.

Obviously, the protein-internal signal sequence can still direct export of cytochrome c_1 and can be processed properly. However, its subsequent fate is not known; it may either be digested or remain buried in the membrane as a C-terminal appendage of cytochrome *b*.

4. *Bacillus subtilis* CotF PROTEINS

In *B. subtilis* the *cotF* gene encodes a 160-amino-acid precursor for two alkali-soluble polypeptides of 5 and 8 kDa which are components of the spore coat [19]. Fig. 1 shows the arrangement of these polypeptides in the precursor, as deduced from the N-terminal sequence of the two mature proteins. Two cleavage sites were proposed, both immediately after the sequence GluArg. The role of the four N-terminal amino acids is not known. It is interesting to note that another small spore protein, the *cotT* gene product, which is also derived from a precursor, is produced by cleavage after a GluArg sequence [20]. The presumed CotF precursor was not detected in protein gels and is thus probably subject to rapid processing. There is no information about the mechanism and order of the two processing steps or about the putative protease(s) involved.

The small spore coat proteins are most likely structural components of the spore, involved either in the protection of the dormant cell or in facilitating germination. Their initial synthesis in a polyprotein precursor may ensure a one-to-one molar ratio of the mature gene products, which could be necessary for establishing a certain ordered structure.

5. *Bacillus polymyxa* AMYLASE

In *B. polymyxa* a single gene contains in-phase β and α amylase coding sequences in its 5′ and 3′ regions, respectively. It directs the synthesis of a bifunctional precursor (130 kDa) that gives rise to multiform β amylases (70 kDa, 56 kDa and 42 kDa) and a uniform α amylase (48 kDa) after proteolytic processing [21]. Fig. 1 shows the structural features of the amylase precursor. A typical signal sequence is located at the N terminus, obviously directing the protein into the periplasm. The 130 kDa precursor as well as the three β amylases have the same N terminus, suggesting that the first processing event removes the signal peptide [22]. In the middle of the protein a stretch of about 200 amino acids is arranged as a direct repeat of two 104-amino-acid segments [22,23]. The C-terminal part contains the α amylase, whose N terminus is not known. Thus, the additional cleavage sites in the precursor have not yet been identified.

The 130 kDa precursor was not only cleaved into the multiple amylases by *B. polymyxa* extracellular neutral protease (Npr) and intracellular serine protease (Isp) produced in *E. coli*, but also by trypsin, chymotrypsin and subtilisin, indicating that proteolytic cleavage may not be site-specific [24]. A current hypothesis is that the linker peptide with the two repeats may be exposed on the surface of the molecule and, hence, be susceptible to various proteolytic enzymes.

6. CONCLUSIONS

The examples of prokaryotic polyproteins identified so far raise two relevant questions: Firstly, what selective pressure forced the evolution or maintenance of polyproteins, and secondly, did polyproteins evolve by the invention of mechanisms for posttranslational separation of functionally distinct domains in a single protein, or by in-frame fusion of separate genes?

As to the first question, several ideas or speculations can be put forward. (i) Synthesis and posttranslational cleavage of polyproteins may constitute the simplest means of guaranteeing the strictly stoichiometric synthesis of polypeptides. (ii) The formation of a polyprotein can be considered as a way to 'compartmentalise' a multicomponent system, e.g. during the process of translocation into the periplasmic or extracellular space. (iii) Within polyproteins, the different domains may fold interdependently and thereby replace the activity of a chaperone, or, after posttranslational cleavage, increase the local concentrations of the different

components rendering assembly less dependent on diffusion.

As to the second question, it is clear that both mechanisms of evolution of a polyprotein involve mutational alterations in the coding regions and the recruitment of auxiliary functions, such as proteases. In the case of gene fusion one would have to postulate a deletion or rearrangement event which either introduces a pre-existing or creates a new proteolytic cleavage site. This kind of model would fit particularly well to the FbcH precursor which could have been created by an incidental, small deletion, thus joining two adjacent genes for cytochrome *b* and cytochrome c_1 without causing a significant disadvantage, because cleavage of the fused precursor would still occur at the same signal peptidase recognition site provided by precytochrome c_1. The bc_1 polyprotein may therefore be advantageous, as it guarantees the formation of equimolar amounts of the products for complex formation.

The penicillin acylase polyprotein precursor is the only example for which comparable counterparts are known in different bacteria. For several reasons this polyprotein may have been the result of fragmentation of pre-existing structures through the generation of processing sites: (i) the α and β subunits work together as a functional unit and do not show separate activities, (ii) maturation appears to depend on structural features distributed over all parts of the precursor, and (iii) processing is either autocatalytic, or is performed by proteases inherent to diverse bacteria.

By comparing the four examples described here it is striking that the polyproteins share one characteristic: each of them is a precursor for proteins that display functional similarity or contribute to the same activity. Whether or not this is a prerequisite or just a fortuitous coincidence remains to be seen when further examples of prokaryotic polyproteins are identified.

REFERENCES

[1] Beadle, G.W. and Tatum, E.L. (1941) Proc. Natl. Acad. Sci. USA 25, 499–506.
[2] Normark, S., Bergström, S., Edlund, T., Grundström, T., Bengtåke, J., Lindberg, F. and Olsson, O. (1983) Ann. Rev. Genet. 17, 499–525.
[3] Atkins, J.F., Weiss, R.B. and Gesteland, R.F. (1990) Cell 62, 413–423.
[4] Steiner, F.D., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980) Ann. NY Acad. Sci. 343, 1–16.
[5] Bussey, H. (1988) Yeast 4, 17–26.
[6] Böck, A., Wirth, R., Schmid, G., Schumacher, G., Lang, G. and Buckel, P. (1983). FEMS Microbiol. Lett. 20, 135–139.
[7] Mayer, H., Collins, J. and Wagner, F. (1979) in: Plasmids of Medical, Environmental and Commercial Importance (Timmis, K.N. and Pühler, A. Eds.), pp. 459–469, Elsevier/North-Holland Biochemical Press, Amsterdam.
[8] Böck, A., Wirth, R., Schmid, G., Schumacher, G., Lang, G. and Buckel, P. (1983) FEMS Microbiol. Lett. 20, 141–144.
[9] Schumacher, G., Sizman, D., Haug, H., Buckel, P. and Böck, A. (1986) Nucleic Acids Res. 14, 5713–5727.
[10] Oliver, G., Valle, F., Rosetti, F., Gómez-Pedrozo, M. Santamaria, P., Gosset, G. and Bolivar, F. (1985) Gene 40, 9–14.
[11] Sizmann, D., Keilmann, C. and Böck, A. (1990) Eur. J. Biochem. 192, 143–151.
[12] Lindsay, C.D. and Pain, R.H. (1990) Eur. J. Biochem. 192, 133–141.
[13] Barbero, J.L., Buesa, J.M., González de Buitrago, G., Méndez, E., Pérez-Aranda, A. and García, J.L. (1986) Gene 49, 69–80.
[14] Prieto, I., del Carmen Rodiguez, M., Márquez, G., Pérez-Aranda, A. and Barbero, J.L. (1992) Appl. Microbiol. Biotechnol. 36, 659–662.
[15] Matsuda, A. and Komatsu, K.-I. (1985) J. Bacteriol. 163, 1222–1228.
[16] Trumpower, B.L. (1990) Microbiol. Rev. 54, 101–129.
[17] Thöny-Meyer, L., Stax, D. and Hennecke, H. (1989) Cell 57, 683–697.
[18] Thöny-Meyer, L., James, P. and Hennecke, H. (1991) Proc. Natl. Acad. Sci. USA 88, 5001–5005.
[19] Cutting, S., Zheng, L. and Losick, R. (1991) J. Bacteriol. 173, 2915–2919.
[20] Aronson, A.I., Song, H.-Y. and Bourne, N. (1989) Mol. Microbiol. 3, 437–444.
[21] Uozumi, N., Sakurai, K., Sasaki, T., Takekawa, S., Yamagata, H., Tsukagoshi, N. and Udaka, S. (1989) J. Bacteriol. 171, 375–382.
[22] Kawazu, T., Nakanishi, Y., Uozumi, N., Sasaki, T., Yamagata, H., Tsukagoshi, N. and Udaka, S. (1987) J. Bacteriol. 169, 1564–1570.
[23] Rhodes, C., Strasser, J. and Friedberg, F. (1987) Nucleic Acids Res. 15, 3934.
[24] Takekawa, S., Uozumi, N., Tsukagoshi, N. and Udaka, S. (1991) J. Bacteriol. 173, 6820–6825.