Limited Proteolysis and Active-site Studies of the First Multienzyme Component of the Erythromycin-producing Polyketide Synthase*

(Received for publication, October 14, 1993, and in revised form, December 29, 1993)

Jesús F. Aparicio‡§, Patrick Caffrey‡, Andrew F. A. Marsden‡, James Staunton§, and Peter F. Leadlay¶

From the Cambridge Centre for Molecular Recognition, and Departments of Biochemistry and Organic Chemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

The domain structure of the 6-deoxyerythronolide B synthase 1 component of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea has been investigated using limited proteolysis and active-site labeling. Trypsin, elastase, endoproteinase Glu-C, and endoproteinase Arg-C were used to cleave the multienzyme, and the sizes of the resulting fragments were assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The location of fragments within the primary structure was established by N-terminal sequence analysis. The cleavage pattern followed domain boundaries previously predicted on the basis of sequence alignments, but many predicted interdomain regions were not cleaved, even under the harsh conditions used. Initial proteolysis generated three large fragments: an N-terminal fragment (about 60 kDa) housing an acyltransferase-acyl carrier protein di-domain; a central fragment (about 90 kDa) containing a ketosynthase-acyltransferase di-domain; and a C-terminal fragment (about 220 kDa) containing the remaining six domains of the multienzyme, including the third acyltransferase. The intact multienzyme behaves as a dimer of molecular mass 660 kDa on gel filtration; and the C-terminal fragment remains dimeric. However, the N-terminal and central fragments appear to be monomeric species. After proteolysis of the multienzyme, the N-terminal di-domain was found to be specifically labeled after incubation with [1-14C]propionyl-CoA, providing the first evidence for its proposed role as a "loading domain" for the propionate starter unit. In contrast, the other two fragments were specifically acylated by [1-14C]methyloxyacyl-CoA, indicating that both the other two acyltransferases remain enzymatically active after proteolysis.

Erythromycin A is a clinically important macrolide antibiotic produced by Saccharopolyspora erythraea (1, 2). Its biosynthesis involves the initial construction of a polyketide chain from one molecule of propionyl-CoA and 6 molecules of methylmalonyl-CoA to form the 14-membered macrolide ring of the first isolable intermediate, 6-deoxyerythronolide B by a mechanism that resembles the biosynthesis of fatty acids (1, 3). Sequencing of the structural genes encoding the chain-building polyketide synthase, 6-deoxyerythronolide B synthase (DEBS), has revealed that it is composed of three type I (multifunctional) polyketides (DEBS 1, DEBS 2, and DEBS 3) (4–7) each of which apparently catalyzes two of the six cycles of chain extension required to produce 6-deoxyerythronolide B. Thus, DEBS 1 would catalyze the binding of the propionyl-CoA starter unit and the first two cycles of chain extension. The domain structure for each DEBS multienzyme has been predicted (6, 7) on the basis of their striking similarity with the known organization of vertebrate fatty acid synthases (8, 9) and of the occurrence, at predicted domain boundaries, of unusual alanine-, proline-, and charged residue-rich sequences (7) that have been implicated as inter-domain linker regions in other multi-domain enzymes (10, 11). However, direct evidence for the topology and domain structure of the DEBS multienzymes is required in order to understand how the growing polyketide chain moves between the 30 or so active sites. Limited proteolysis (12, 13) has been used here, in combination with specific active-site radiolabeling, to study the first multienzyme of the erythromycin-producing polyketide synthase (DEBS 1) and to provide the first structural information on any bacterial antibiotic-producing polyketide synthase.

EXPERIMENTAL PROCEDURES

Materials

1-1-Tosylamido-2-phenylethyl chloromethyl ketone-treated bovine trypsin, 1-chloro-3-tosylamido-7-amino-2-heptanone-treated chymotrypsin, thiglycolic acid (nercaptatoic acid), and 3-cyclohexylamino-1-propanesulfonic acid buffer were purchased from Sigma; high range M, standards for SDS-PAGE and phenyl-Superose, Superdex 75, Seprose 6, and Seprose 12 columns were from Pharmacia LKB Biotechnology Inc.; Elastase, endoproteinase Arg-C, and endoproteinase Glu-C were from Boehringer Mannheim; ProBlott polyvinylidene difluoride membranes and sequencing reagents were from Applied Biosystems Inc., Foster City, CA. All other reagents were from commercial sources and of analytical grade. DEBS 1 protein was purified from S. erythraea as previously described (14). [1-14C]Methyloxyacyl-CoA was prepared from [1-14C]methylmalonate propionate (DuPont-NEN, 51 Ci/mmol) via a 1-14C]propionyl-CoA intermediate, using acetyl-CoA synthetase and transcarboxylase.

Limited Proteolysis of DEBS 1

Trypsin—Purified DEBS 1 was incubated with 1-1-tosylalanyl-2-phenylethyl chloromethyl ketone-treated trypsin at an enzyme/substrate ratio of 1/350 (w/w) in 1 mM EDTA, 2 mM dithiothreitol, 50 mM Tris-HCl buffer, pH 7.5, 20% glycerol (buffer A). Reactions were carried out at 30 °C for various times and terminated by heating at 100 °C for 3 min in electrophoresis sample buffer unless otherwise indicated. The reaction

* This work was supported by project grants from the Science and Engineering Research Council (United Kingdom). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of a bursary from the European Community.

§ To whom correspondence should be addressed: Dept. of Biochemistry, University of Cambridge, Tennis Court Rd., Cambridge CB2 1QW, UK. Tel.: 44-223-335666; Fax: 44-223-335346.

1 The abbreviations used are: DEBS, 1,6-deoxyerythronolide B synthase 1; ACP, acyl carrier protein; AT, acyltransferase; KR, β-ketoreductase; KS, β-ketoacyl ACP synthase; PAGE, polyacrylamide gel electrophoresis.

tion products were separated by SDS-PAGE and either examined after staining with Coomassie Blue R-250 or transferred to ProBlott membranes for N-terminal sequencing.

Elastase—DEBS 1 was treated at an elastase/substrate ratio of 1/200 (w/w) in buffer A. Reactions were performed, terminated, and the products analyzed as described above.

Endoprotease Glu-C—An enzyme/substrate ratio of 1/40 in buffer A was used. Reactions were performed at 37 °C, and the mixtures treated as described above.

Endoprotease Arg-C—We used a proteinase/substrate ratio of 1/30 (w/w) in buffer A. Reactions were performed at 37 °C and treated as described above. In some cases the conditions of proteolysis were modified as indicated.

SDS-PAGE and Determination of N-terminal Sequence

SDS-PAGE was performed in 5, 7, or 10% polyacrylamide gels (15). After electrophoresis, gels were stained either with Coomassie Brilliant Blue R-250 or with silver (16). Alternatively gels were transferred to ProBlott polyvinylidene difluoride membranes according to Matsudaira (17) and subjected to Edman degradation on an Applied Biosystems 477A pulsed-liquid protein sequencer. Molecular weights of protein fragments were evaluated by using calibration curves from measurements with commercially available standard proteins. Protein concentrations were determined by the methods of Bradford (18) and Lowry et al. (19).

Acetylation of DEBS 1

Labeling was carried out at 0 °C, after limited proteolysis, using (2RS) [1-14C]methylmalonyl-CoA or (1-14C)propionyl-CoA as described elsewhere (20).

Fragment Separation under Non-denaturing Conditions—The products of proteolysis of DEBS 1 were separated from each other by microfase protein liquid chromatography using a Smart system (Pharmacia). Proteolysis of DEBS 1 was performed as described above, stopped by rapid freezing on dry ice/acetone, and the reaction mixture was stored at -80 °C.

Hydrophobic Interaction Chromatography—The proteolysis mixture was applied to a phenyl-Superose (Pharmacia) column at 1 M (NH4)2SO4 in buffer A. The column was then washed with 250 μl of the same buffer and the products eluted with a linear gradient of decreasing salt at a flow rate of 50 μl/min at 4 °C.

Size Exclusion Chromatography—The proteolysis mixture was applied onto a Superdex 10, Superose 12, or Superose 6 (Pharmacia) column previously equilibrated with buffer A containing 150 mM NaCl, and chromatography was carried out at a flow rate of 40 μl/min at 4 °C. Superose 12 and a Superose 6 column were calibrated for molecular mass determination using the following proteins: thyroglobulin, ferritin, catalase, aldolase, and bovine serum albumin. The Superdex 75 column was calibrated for molecular mass determination using as standards bovine serum albumin, ovalbumin, carbonic anhydrase, myoglobin, and cytochrome c.

RESULTS

Limited Proteolysis of DEBS1

DEBS 1 was digested with either trypsin, elastase, endoprotease Glu-C, or endoprotease Arg-C at various molar ratios at either 30 or 37 °C as described under “Experimental Procedures” for various lengths of time, and the resulting fragmentation patterns are shown in Fig. 1. Fragment patterns were unaltered when SDS-PAGE was carried out in the presence of dithiothreitol or 2-mercaptoethanol.

Trypsin—The fragmentation pattern generated by trypptic cleavage is shown in Figs. 1B and 2. The initial cleavage introduced by trypsin resulted in the release of a 57-kDa fragment (T6), and a 219-kDa fragment (T5) as the first stable products (see Fig. 1F). The size on SDS-PAGE and the N-terminal sequence of T6 showed that it comprises the di-domain AT1-ACP1 thought to be responsible for loading the propionyl starter unit. Similarly, the N-terminal sequence of T5 indicated that this fragment comprises the domains KR1-ACP2-KS2-AT3-KR2-ACP3. As the hydrolysis continued, an N-terminal piece of the loading domain was removed leading to the generation of a 56-kDa loading domain fragment (T1). Several other fragments appeared concurrent with the disappearance of both the remaining DEBS 1 protein and the transiently produced 300-kDa species generated by the release of T0, namely T10 (94 kDa comprising AT2-KS1) and T6 (168 kDa comprising ACP2-KS2-AT3-KR2-ACP3).

As the time of incubation was increased further fragments were generated. T2 (51 kDa) and T3 (49 kDa) have the same N terminus at Val'500 and contain the fragment KR1-ACP2-KS2-AT3-KR2-ACP3.

DEBS 1 domain organization as revealed by limited proteolysis. A, DEBS 1 map. Each activity is represented by a rectangle whose length is proportional to the proposed length of the “domain” (6). Linker regions are also shown in proportion. Numbers in the proposed domains indicate their distance from to the N terminus of the protein. The ruler indicates the residue number within the primary structure of the protein. B, DEBS 1 fragmentation with trypsin. The representation is proportional. The N-terminal sequence of each fragment is included; C, the same with elastase; D, the same with endoprotease Glu-C; E, the same with endoprotease Arg-C. Asterisks denote fragments stable under the harshest conditions used (see “Experimental Procedures”). Thickened lines indicate the first fragments generated. Dashed lines indicate minor species.

Fig. 1. DEBS 1 domain organization as revealed by limited proteolysis. A, DEBS 1 map. Each activity is represented by a rectangle whose length is proportional to the proposed length of the “domain” (6). Linker regions are also shown in proportion. Numbers in the proposed domains indicate their distance from to the N terminus of the protein. The ruler indicates the residue number within the primary structure of the protein. B, DEBS 1 fragmentation with trypsin. The representation is proportional. The N-terminal sequence of each fragment is included; C, the same with elastase; D, the same with endoprotease Glu-C; E, the same with endoprotease Arg-C. Asterisks denote fragments stable under the harshest conditions used (see "Experimental Procedures"). Thickened lines indicate the first fragments generated. Dashed lines indicate minor species.
the N-terminal end of fragment T5 (Ser^{1443}), dividing the molecule into three independent parts, namely the loading di-domain (AT1-ACP1), the di-domain comprising KS1 and AT2 (fragment T10), and a hexa-domain (fragment T5) that includes the rest of the protein. The same fragmentation pattern was observed in the absence of dithiothreitol or when the concentration of glycerol in the reaction mixture was reduced to 10%, although the reaction was then much faster (data not shown).

Elastase—The fragmentation pattern generated by elastase cleavage of DEBS 1 is shown in Fig. 1C. Interestingly, initial cleavage between the ACP2 domain and the C-terminal KS domain (KS2) at the Ala^{2025}-Thr^{2026} bond yielded two large fragments, E1 (212 kDa) and E2 (158 kDa).

Digestion with 10-fold more elastase generated a complicated proteolytic pattern (Fig. 1C) as a result of the more complete degradation of the N-terminal fragment, with the C-terminal half remaining stable under these conditions.

Endoproteinase Glu-C—The fragmentation pattern resembles the tryptic fragmentation but with three new fragments. So, G8 that includes KS2-KR2-ACP3, G5 that comprises KS1-AT2-KR1-ACP2 (see Fig. 1D) and G11, a 47-kDa fragment that constitutes the AT1 domain.

Endoproteinase Arg-C—The fragmentation pattern with endoproteinase Arg-C was different to that obtained with trypsin (Fig. 1E). The initial cleavage introduced by this protease takes place at Arg^{1442}-Ser^{1443}, cutting the molecule into two large fragments: A5, a 151-kDa fragment containing AT1-ACP1-KS1-AT2, and A1, a 219 kDa fragment whose N terminus is at Ser^{1443}, and contains KR1-ACP2-KS2-AT3-KR2-ACP3. Only when harsher conditions were used (ratio of proteinase/DEBS 1 (1:20 w/w)) were other fragments obtained (see Fig. 1E). In contrast to the results of the tryptic fragmentation, no fragment was found equivalent to T10, but this is probably due to the bigger “loading domain” generated by this proteinase in comparison with the others, which would make the KS1-AT2 fragment smaller and perhaps less structured. Limited proteolysis was also attempted using chymotrypsin, but no stable fragments were observed.

Acylation of DEBS 1

DEBS 1 is rapidly acylated by the substrates methylmalonyl-CoA and propionyl-CoA (20). To determine which domains were being acylated, various proteolytic digests of DEBS 1 were labeled as indicated under “Experimental Procedures.” When a tryptic digest was treated with [1-^{14}C]propionyl-CoA, the radioactivity was found specifically associated with the T0 or T1 fragments (Fig. 2B), indicating that at least one of the components of the loading domain was acylated. With the endoproteinase Glu-C digest, in which an individual AT1 is obtained (fragment G11), analysis showed that the [1-^{14}C]propionyl-CoA labeling was found specifically within AT1 and in none of the fragments lacking this AT domain.

In contrast, when the same digests were treated with [1-^{14}C]methylmalonyl-CoA, labeling was observed in T5, T6, T7, T8, and T10 (Fig. 2B). This indicates that any fragment containing an AT domain other than AT1 could be labeled. The labeling observed in T10 is significant, because this di-domain contains only KS1 and AT2. This labeling was always weaker than that obtained in other fragments, perhaps indicating that the ACP domains might play a role in the stability of this acyl intermediate.

Labeling of other proteinase digests was fully consistent with
the labeling of AT1 by [1-14C]propionyl-CoA and of AT2 and AT3 by [1-14C]methylmalonyl-CoA. No labeling was observed when [2-14C]malonyl-CoA was used as a substrate (not shown).

**Separation of the Proteolytic Fragments**

Tryptic digestion of DEBS 1 resulted in the initial generation of three main fragments, T1, T10, and T5 (see above), which were separated using a phenyl-Sepharose column (see "Experimental Procedures"). When a Superdex 75 gel filtration column was used, the relative elution volume of fragments T1 and T10 indicated that they behave as monomers since their predicted sizes, 56 and 94 kDa, respectively (also determined by SDS-PAGE), were very similar to the observed ones (62 ± 5 and 95 ± 4 kDa). The fact that T1 and T10 were also separated from each other, and not associated with T5 or T6, indicates that under the conditions used, they behave as independent fragments. T2 and T3 (KR1 fragments) were also monomeric, with a native molecular weight of approximately 53,000.

Similar results were obtained upon gel filtration of an endoproteinase Glu-C digest of DEBS 1. Both G2 and G3 were also independent domains, and G3 behaved as a monomer with a molecular mass of 65 kDa. Analogously, G11 and G10 (AT1 and KR1 fragments respectively) were also independent and monomeric, with apparent molecular masses of 48 and 54 kDa, respectively.

On a Superose 12 filtration column, fragments T0, T1, T10, T2, T3, G3, G11, and G10 all behaved again as independent monomers. However, fragments T5 (219 kDa) and T6 (168.5 kDa) both showed a native molecular weight of approximately 460,000. Although this is not within the optimal separation range, it indicates that they behave as dimers. Fragments T7 (161.5 kDa) and T8 (156 kDa) both eluted from this column as a globular protein with a molecular mass of 240 kDa would do, while fragments G5 (154 kDa) or G8 (159 kDa) showed a native molecular weight of approximately 280,000. Whether those fragments were independent or not, could not be decided because of the similarity of their sizes.

Gel filtration on Superose 6 allowed the fractionation of the largest fragments. For example, fragments G2 (220.5 kDa) and G4 (194 kDa) both showed an apparent molecular mass of 440 kDa while T6 (168.5 kDa) showed a molecular mass of approximately 400 kDa and G5 and G8 a mass of 320 kDa. Comparison of these values with the predicted sizes clearly indicated that these fragments are either dimeric or highly asymmetric monomers. Importantly, fragments T7 and T8 showed a molecular weight of 360,000 using this column, which could indicate that they are dimers as well. Fragments E1 (212 kDa) and E2 (159 kDa) also eluted independently and behaved as dimers with molecular masses of 460 and 300 kDa, respectively. From the position of elution of DEBS 1 from this column it appeared to have a molecular mass of 660 kDa. All determinations were performed at least four times and with different enzyme preparations.

**DISCUSSION**

The recent purification to near-homogeneity of the three constituent multienzymes of the erythromycin-producing polyketide synthase from *S. erythraea* (14) has provided the first opportunity to explore the structure and assembly of these unusual multifunctional polypeptides. In general, such multifunctional proteins are composed of independently folded, compact domains connected by linker regions (21, 22) which are readily cleavable by limited treatment with proteolytic enzymes (23). From X-ray diffraction studies, it has been demonstrated (24) that segments of the polypeptide chain of high flexibility, as determined by mean atomic displacements, are correlated with those regions that are vulnerable to limited proteolysis. Evidence, many interdomain loops are located at the surface of the protein and adopt conformations that facilitate the interaction of proteases with the polypeptide substrate.

Proteolytic studies on DEBS 1 were carried out using four different proteases, chosen for their relatively broad primary specificity, so that accessibility to potential target sites, rather than the specificity of the protease itself, is the factor likely to determine the sites of cleavage. N-terminal sequencing of the fragments generated allowed them to be placed within the context of the known primary structure of DEBS 1 (6), allowing the rapid identification of the regions particularly susceptible or resistant to proteolysis. The initial cleavages observed indicated that the DEBS 1 molecule (370 kDa) could be divided into three large parts. For example, trypsin cleaves the protein initially into three fragments of approximately 57, 94 and 219 kDa, with cutting sites at Ala1448 and Ser1449. The smallest (N-terminal) fragment comprises the di-domain AT1-ACP1, which is proposed to function as a loading domain for the propionate starter unit. The central fragment comprises KS1 and AT2, and the large C-terminal fragment contains the rest of the domains originally present in DEBS 1. This overall fragmentation pattern was confirmed by the results of digestion with endoproteinase Glu-C. As with trypsin, two main sites of cleavage were observed. Cleavage occurred at Thr2466, two positions away from the N terminus of T10, and at Arg1431, 12 residues away from Ser1443. Both linker regions were also proteolyzed when the protein was treated with elastase, but harsher conditions were required to release the N-terminal loading di-domain. With endoproteinase Arg-C, only the second linker region was cut, unless harsh conditions were used, when a second cleavage was observed within the first ketosynthase domain KS1, releasing a larger loading domain than with the other enzymes. Remarkably, the C-terminal part of the molecule remained very stable and highly resistant to hydrolysis by any of the proteases used (no cleavage was observed between KS2 and AT3, and under harsh conditions were cuts observed within some other parts of this fragment).

Sequence comparisons have been used to predict the boundaries of the individual domains present in DEBS 1 (6, 7), and the results presented here are broadly consistent with those predictions: only a few minor cuts were found inside the predicted extent of some domains (data not shown). Moreover, one of the two main linker regions, the C-terminal part of the loading domain, is flanked by two regions rich in alanine, proline, aspartic and glutamic acids (i.e., 532AEALAAGTE and 555EAAPGEPVA), which are thought to provide flexibility (7, 25). In the same way, the main point of cleavage with elastase (Ala1952-Thr2026) is located just at the C-terminal end of another putative linker region (2011AEAQAPA) (7).

The three main fragments obtained after digestion with trypsin (see Fig. 1B) were readily separated by gel filtration under non-denaturing conditions, indicating that the fragments are independently folded and that they do not interact strongly with each other. The N-terminal and central di-domains behaved on gel filtration as monomeric species, while the large C-terminal fragment was apparently dimeric. The individual KR1 fragments generated by trypsin or by endoproteinase Glu-C cleavage are also monomeric. DEBS 1 itself behaves on gel filtration as a homodimer of molecular mass 660 kDa, so these results taken together suggest that the dimer is held together by only strong interactions within the C-terminal half of the molecule. This is the first evidence of any kind for the structure in solution of a bacterial polyketide synthase.

After digestion with trypsin, the N-terminal acyltransferase-acyl carrier protein di-domain of DEBS 1 was specifically acylated by [1-14C]propionyl-CoA, providing the first evidence that this portion of DEBS 1 functions as a "loading domain" for the
propionate starter unit, as previously proposed (6), and confirming that the acyltransferase remains active after limited proteolysis. We have recently shown (20, 26) that all of the purified DEBS multienzymes are transiently acylated by the substrate methylmalonyl-CoA, presumably at the active sites of each of the six acyltransferases proposed to load methylmalonyl-CoA extender units onto the enzyme. The finding that, after trypsin digestion, both the central ketosynthase-acyltransferase di-domain and the large C-terminal fragment (which contains the third acyltransferase activity of DEBS 1) were specifically labeled by [1-14C]methylmalonyl-CoA, similarly provided evidence that both of these acyltransferases retain activity after limited proteolysis, and supports the assumption that the acyltransferases function in the loading of methylmalonyl-CoA extender units (4, 6, 7, 20, 26).

The idea of a compact, relatively proteinase-resistant C-terminal half of the DEBS 1 molecule, involved in inter-subunit interactions, receives additional support from the finding that tryptic fragments T6 and T5, and endoproteinase Glu-C fragments G5 and G8, have native molecular masses appropriate for dimeric species. The results do not as yet allow discrimination between the two possibilities of parallel and antiparallel arrangement of the polypeptides within the DEBS 1 dimer. Further work is required to establish the detailed topology of the enzyme, and thus to gain insight into how the growing polyketide chain is efficiently transferred to each active site in turn.

REFERENCES