Bioorganic Chemistry **27,** 172–188 (1999)
Article ID bioo.1998.1122, available online at http://www.idealibrary.com on IDEAL®

Enediyne Biosynthesis and Self-Resistance: A Progress Report

Jon S. Thorson,*[,1] Ben Shen,†[,2] Ross E. Whitwam,* Wen Liu,†[,‡] Yuan Li,‡ and Joachim Ahlert*

**Laboratory for Biosynthetic Chemistry, Molecular Pharmacology and Therapeutics Program, Memorial Sloan-Kettering Cancer Center and the Sloan-Kettering Division, Graduate School of Medical Sciences, Cornell University, 1275 York Avenue, Box 309, New York, New York 10021; †Department of Chemistry, University of California, One Shields Avenue, Davis, California 95616; and ‡Institute of Medicinal Biotechnology, The Chinese Academy of Medical Sciences, Beijing, 100050 China*

Received August 3, 1998

The enediyne antitumor antibiotics are appreciated for their novel molecular architecture, remarkable biological activity, and fascinating mode of action, and many have spawned considerable interest as anticancer agents in the pharmaceutical industry. Of equal importance to these astonishing properties, the enediynes also offer a distinct opportunity to study the unparalleled biosyntheses of their unique molecular scaffolds and what promises to be unprecedented modes of self-resistance to highly reactive natural products. Elucidation of these aspects should unveil novel mechanistic enzymology and may provide access to the rational biosynthetic modification of enediyne structure for new drug leads, the construction of enediyne overproducing strains, and eventually lead to an enediyne combinatorial biosynthesis program. This article reviews the published enediyne biosynthetic labeling and blocked mutant studies and provides a brief account of efforts in the Shen laboratory to elucidate the route to C-1027 biosynthesis and efforts in the Thorson laboratory to elucidate the calicheamicin biosynthetic cascade.

INTRODUCTION

The enediyne antitumor antibiotics are appreciated for their novel molecular architecture, remarkable biological activity, and fascinating mode of action (*1–4*). To date, at least 13 members of this unique family have been discovered, all of which fall roughly into two categories. The first are the chromoprotein enediynes, all postulated to possess a novel 9-membered ring chromophore core structure and require a specific associated protein for chromophore stabilization, while the second class contains a 10-membered ring system and lacks any additional stabilization factors. Among the known enediynes, many have spawned considerable interest as anticancer agents in the pharmaceutical industry. In particular, calicheamicin-antibody conjugates were

[1] To whom correspondence and reprint requests should be addressed. Fax: (212) 717-3066. E-mail: jthorson@sbnmr1.ski.mskcc.org.

[2] To whom correspondence and reprint requests should be addressed. Fax: (530) 752-8995. E-mail: shen@chem.ucdavis.edu.

0045-2068/99 $30.00

recently reported to have successfully completed phase II clinical trials (*5–9*), esperamicin A_1 was reported to be in phase II, and various neocarzinostatin delivery systems, such as the poly(styrene-co-maleic acid) conjugated neocarzinostatin (SMANCS), have shown clinical promise in comparison to alternative chemotherapies (*10–17*). Of equal importance to these astonishing properties, the enediynes also offer a distinct opportunity to study the unparalleled biosyntheses of their unique molecular scaffolds and what promises to be unprecedented modes of self-resistance to highly reactive natural products. This article reviews the published enediyne biosynthetic labeling and blocked mutant studies. In addition, we present a brief account of current efforts in the Shen laboratory to elucidate 9-membered ring chromoprotein enediyne biosynthesis (C-1027) and current efforts in the Thorson laboratory to elucidate 10-membered ring enediyne biosynthesis (calicheamicin).

9-MEMBERED RING CHROMOPROTEIN ENEDIYNES

This first class (Fig. 1) is composed of neocarzinostatin (**1**) from *Streptomyces carzinostaticus* (*18,19*), kedarcidin (**2**) from *Actinomycete* L585-6 (*20,21*), C-1027 (**3**) from *Streptomyces globisporus* (*22–24*), maduropeptin (**4**) from *Actinomadura madurea* (*25*), N1999A2 (**5**) from *Streptomyces* sp. AJ9493 (26), actinoxanthin from *Actinomyces globisporus* (*27*), largomycin from *Streptomyces pluricolorescens* (*28*), auromomycin from *Streptomyces macromomyceticus* (*29–32*), and sporamycin from *Streptosporangium pseudovulgare* (*33*), all of which are believed to possess a novel bicylo[7.3.0.]dodecadiynene chromophore core structure essential for biological activity. In addition, with the exception of **5**, a required apoprotein acts as a stabilizer and

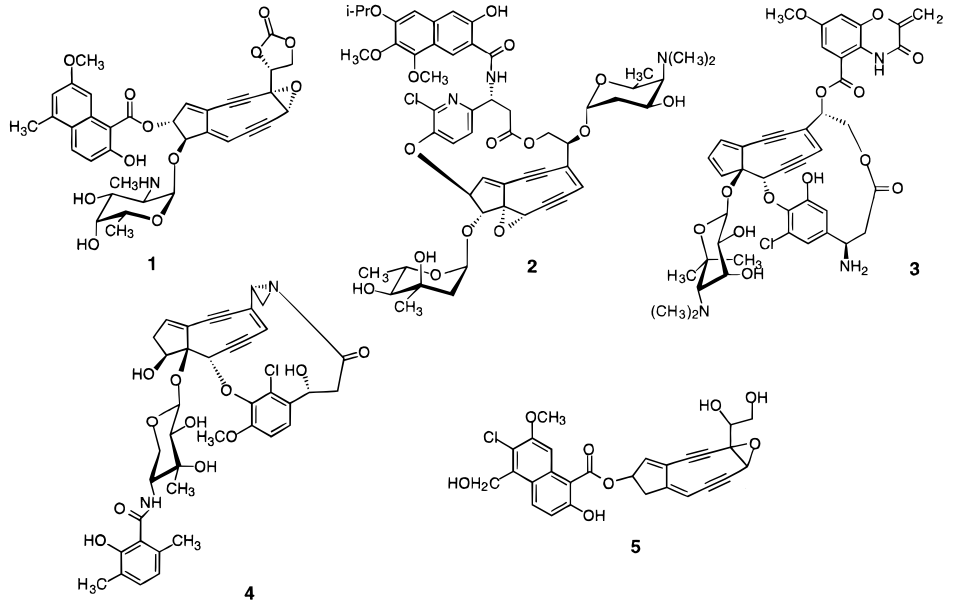

FIG. 1. Bicyclo[7.3.0]enediyne chromophore for neocarzinostatin (**1**), kedarcidin (**2**), C-1027 (**3**), maduropeptin (**4**), and N1999A2 (**5**). The stereochemical determination of **5** has not been reported.

specific carrier for the unstable chromophore and for its transport and interaction with target DNA. Due to chromophore lability, currently only structures **1–5** are available. In contrast, the apoproteins were purified, the amino acid sequences of the apoproteins for neocarzinostatin (NcsA) (*34*,*35*), kedarcidin (Ked) (*36*,*37*), C-1027 (CagA) (*38*), actinoxanthin (AxnA) (*39*), maduropeptin (Mad) (*36*,*40*), and auromomycin (McmA) (*41*) were determined, and the corresponding genes, *ncsA* (*42*), *cagA* (*43*), *axnA* (*44*), and *mcmA* (*45*), were cloned and sequenced.

The apoprotein of the chromoprotein enediynes. With the exception of maduropeptin (*36*,*40*) the apoproteins isolated to date are small, acidic proteins and are highly homologous (Fig. 2). In addition, DNA sequence analysis of the cloned genes revealed they encode preapoproteins with typical leader peptides of 32–34 amino acid residues. Structural information is available for Ked (*46*), NcsA (*47*), McmA (*48*), and AxnA (*49*) and reveals a seven-stranded antiparallel β-barrel domain linked to a subdomain composed of two β-hairpin ribbons. Interestingly, it has been suggested this motif resembles the variable domain of immunoglobins (*36*). Functionally, these apoproteins have also been shown to exhibit a selective protease activity, suggesting the chromoprotein enediynes may actually function *in vivo* via a targeted delivery "two-pronged" attack in which the apoprotein selectively cleaves histones to expose DNA which is

```
            1                                                    50
      CagA  ~MSLRHMSRR ASRFGVVAVA SIGLAAAAQS VAF.AAPAFS VSPASGLSDG
      AxnA  ~MSLRHMSRR ASRFGVVAVA SIGLAAAAQS VAF.AAPAFS VSPASGLSDG
      NcsA  MVPISIIRNR VAKVAVGSAA VLGLAVGFQT PAVAAAPTAT VTPSSGLSDG
      McmA  ~~MLQNTSRF LARAGATVGV AAGLAFSLPA DR.DGAPGVT VTPATGLSNG
       Ked  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~ASAAVS VSPATGLADG
Consensus  ---l---srr --r-gv---a --GLA---q- -a--aap--- V-Pa-GLsdG

            51                                                   100
      CagA  QSVSVSVSGA AAGETYYIAQ CAPVG.GQDA CNPATATSFT TDASGAASFS
      AxnA  QSVSVSVSGA AAGETYYIAQ CAPVG.GQDA CNPATATSFT TDASGAASFS
      NcsA  TVVKVAGAGL QAGTAYDVGQ CAWVDTGVLA CNPADFSSVT ADANGSASTS
      McmA  QTVTVSATGL TPGTVYHVGQ CAVVEPGVIG CDATTSTDVT ADAAGKITAQ
       Ked  ATVTVSASGF ATSTSATALQ CAILADGRGA CNVAEFHDFS L.SGGEGTTS
Consensus  --V-Vs--G- --g--y---Q CA-v--G--a Cn-a-----t -da-G----s

            101                                                  150
      CagA  FVVRKSYTGS T.PEGTPVGS VDCATA..AC NLGAGN.SGL DLGHVALTFG
      AxnA  FVVRKSYAGS T.PEGTPVGS VDCATD..AC NLGAGN.SGL DLGHVALTFG
      NcsA  LTVRRSFEGF L.FDGTRWGT VDCTTA..AC QVGLSDAAGN GPEGVAISFN
      McmA  LKVHSSFQAV VGANGTPWGT VNCKVV..SC SAGLGSDSGE GAAQ.AITFA
       Ked  VVVRRSFTGY VMPDGPEVGA VDCDTAPGGC EIVVGGNTG. EYGNAAISFG

Consensus  --Vr-S--g- --p-Gt--G- VdC-t----C --g-g---G- -----A--F-
```

FIG. 2. Amino acid sequence comparison of the apoproteins for C-1027, CagA (D10827), actinoxanthin, AxnA (D11457), neocarzinostatin (D10996), NcsA (D10996), auromomycin, McmA (D90006), and kedarcidin, Ked (P41249). The leader peptides are shown in italic and the first amino acid of the mature apoproteins are shown in boldface. In the consensus, capital letters represent amino acid residues conserved in all sequences and small letters represent amino acid residues conserved in three of the four leader peptides or four of the five apoproteins. Given in parentheses are nucleotide sequence accession numbers for *cagA*, *axnA*, *ncsA*, and *mcmA*, and protein sequence accession number for Ked.

then cut by the chromophore (*36*,*50*). In addition, the observation that the expression of *cag*A and *mcm*A is constitutive and independent of the chromophore production, assuring that the chromophore will be sequestered, suggests the apoprotein may function in a self-resistance role as well (*43*,*45*).

Neocarzinostatin biosynthesis. While much is known about the apoproteins, the instability of the enediyne chromophores has rendered classical isotope-labeling experiments difficult and thus, the biosynthesis of **1** is the only member of this family examined *in vivo*. Hensens and coworkers established the origin of all the carbons in **1** on the basis of the incorporation of [1-^{13}C]- and [1,2-^{13}C]acetate, L-[*methyl*-^{3}H]methionine, and sodium [^{14}C]carbonate by cultures of *S. carzinostaticus* (Fig. 3a) (*51*). The cyclic carbonate carbon is derived from carbonate and the *N*- and *O*-methyl groups are of methionine origin. Consistent with bacterial type II polyketide synthase (PKS) assembly of aromatic compounds, the naphthoic acid moiety is derived from a hexaketide consisting of six head-to-tail coupled acetate units, which opens the possibility of directly cloning the PKS genes for the naphthoic acid moiety (*52*). The bicyclo[7.3.0]enediyne core appears to be derived from a minimum of eight head-to-tail coupled acetates with the loss of a C-1 or C-2 carbon from two units, respectively. In contrast to the bicyclo[7.3.1] enediyne core system (Fig. 3b and 3c), the two triple bonds of **1** are derived from a pair of intact acetate units, indicating that the 9-membered bicyclo[7.3.0]enediyne core and the 10-membered [7.3.1]enediyne core may be biosynthetically distinct.

At least two mechanisms can be proposed to interpret these isotope labeling results. Hensens and coworkers favor a fatty acid origin of **1**, suggesting C_{18} oleate as a precursor that is shortened by loss of carbons from both ends and is desaturated via

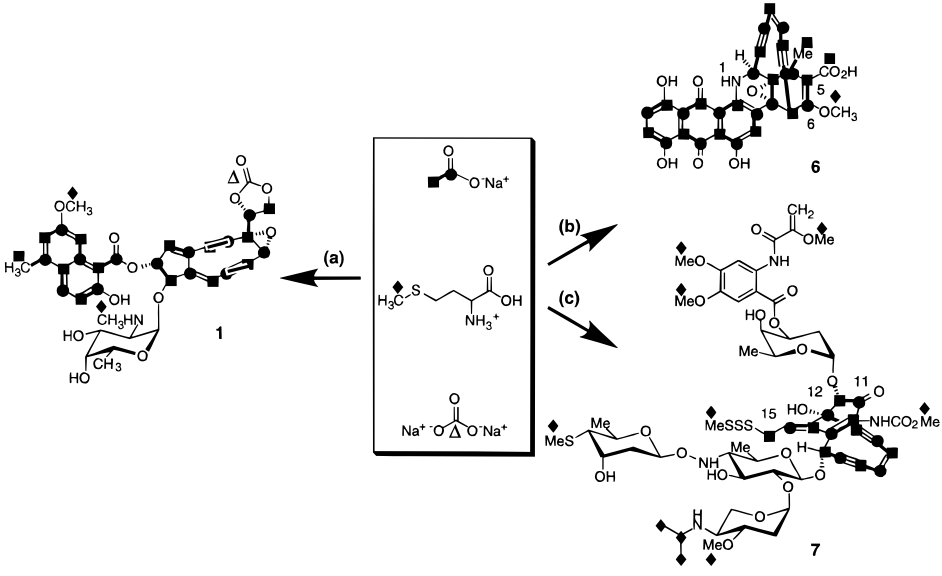

FIG. 3. Incorporation of isotopically-labeled precursors into (a) neocarzinostatin (**1**) by *S. neocarzinostaticus*, (b) dynemicin (**6**) by *M. chersina*, and (c) esperamicin A_1 (**7**) by *A. verrucosospora*.

the oleate-crepenynate pathway. This pathway is well established for polyacetylene biosynthesis, but to date has only been observed in higher plants and fungi (*53*). In this mechanism, formation of an enediyne would require an acetylene-forming enzyme and recently a plant gene encoding such an enzyme was identified (*54*), thus providing an opportunity to scan for an analogous gene in *S. neocarzeostaticus*. Alternatively, it can be envisaged that **1** is of polyketide origin and that an enediyne PKS catalyzes the biosynthesis of a polyunsaturated linear heptaketide intermediate that is subsequently cyclized into the enediyne core structure.

Cloning the genes responsible for C-1027 biosynthesis and resistance. It is very likely that the functions of many genes of the C-1027 cluster could be predicted based solely on sequence comparison since genes for many of the proposed biosynthetic steps have now been cloned and characterized from various streptomycetes. For example, the benzoxazolinate moiety could be derived from pyruvate and anthranilate, the latter of which is of shikimic acid origin. The β-amino acid moiety is likely derived from tyrosine by an aminomutase, which has been characterized from several *Streptomyces* species (*55,56*). Due to the extreme aglycone similarities of **1** and **3**, the labeling studies described above suggest the bicyclo[7.3.0]enediyne core of **3** is also most likely of type I polyketide or fatty acid origin (Fig. 8). While the conversion of a linear heptaketide into the bicyclo[7.3.0]enediyne core must proceed in multiple steps, involving unprecedented reaction mechanisms, polyketide biosynthesis could be readily tested by attempting to clone the putative PKS genes for **3** according to the various methods developed for aromatic polyketide (*82,84*), macrolide (*52*), and polyene (*57*) biosynthesis in streptomycetes.

Our first attempt to clone the biosynthetic genes for **3** employed the random amplified polymorphic DNA technique (*58*). Wild-type *S. globisporus* was treated with acriflavine and five mutants unable to produce **3** were selected for by both an antimicrobial assay (*59*) and an antitumor assay (*60*). Using sets of random primers and genomic DNA from both wild-type *S. globisporus* and mutant strains as templates, PCR amplification revealed a 0.9-kb fragment (designated F2) from the wild-type strain, which was missing in one (AF67) of the five mutants tested. Since this was the only observed difference in the overall PCR analysis, we reasoned that F2 might encode C-1027 biosynthesis genes and a mutation in this region had been introduced in the blocked mutant AF67. Nucleotide sequence analysis of F2 revealed an open reading frame (orf) of 369 bp encoding a 122 amino acid peptide but unfortunately no homologue of this peptide was found in a database search (*61*). Intriguingly, we also failed to detect any apoprotein from the broth of the C-1027 nonproducing mutants, suggesting that the preapoprotein cannot be secreted unless it binds to the chromophore.

Our second approach to cloning biosynthetic genes for **3** took the advantage of the enormous knowledge available for deoxysugar biosynthesis in streptomycetes (*62,63*). We reasoned that the glycosyl moiety **18** of **3** is biosynthesized from glucose as outlined in Fig. 6. Since genes encoding these enzymes are known to be homologous, we adopted the PCR approach to clone the NDP-glucose-4,6-dehydratase gene, expected to encode E_{od}, from *S. globisporus* using primers designed from conserved regions within dehydratase genes from several actinomycetes (*64*). Indeed, a fragment of the predicted size of 500 bp was amplified, whose identity as part of a putative

dehydratase gene was confirmed by sequence analysis. This fragment was subsequently used to screen a *S. globisporus* genomic library, constructed in pOJ446 (*65*,*66*), resulting in the isolation of seven positive cosmid clones. Preliminary restriction mapping of these cosmids indicates they overlap a continuous 55-kb region of the *S. globisporus* chromosome. Within this region, the dehydratase probe was mapped to a 3.0-kb *Bam*HI fragment, which was cloned and sequenced (Fig. 4). Sequence analysis revealed an orf, ORF1 whose deduced product is highly homologous to known NDP-glucose-4,6-dehydratases such as StrE (*67*) and MtmE (*68*), supporting ORF1 to encode the putative E_{od} for the biosynthesis of **18**.

Antibiotic production genes have invariably been found to be clustered in one region of the microbial chromosome, consisting of structural, resistance, and regulatory genes. Consequently, it was anticipated that ORF1 and *cagA*, a putative structural gene and a known resistance gene for **3**, respectively, should be clustered and *S. globisporus* chromosomal walking from the ORF1 and *cagA* genes should lead to the identification of other genes of the C-1027 cluster. Indeed immediately downstream of ORF1 was ORF2 (Fig. 4), whose deduced product belongs to a family of membrane efflux pumps such as LfrA from *Mycobacterium tuberculosis* (*69*) and RifP from *Amycolatopsis mediterranei* (*70*). Since LfrA and RifP confer fluoroquinolone and rifamycin resistance, respectively, by transporting drug out of the cell in exchange for protons, we assign ORF2 as a drug efflux pump, providing *S. globisporus* the second mechanism of resistance to **3**. Although it is tempting to propose that **3** is first sequestered by binding to the preapoprotein CagA to form a complex, which is then transported out of the cell by the efflux pump ORF2 and processed by removing the leader peptide to yield the chromoprotein, we do not have any data to support this hypothesis.

To determine if *cagA* is clustered with the ORF1/2 locus, we next designed primers according to the flanking regions of *cagA* and, remarkably, amplified the *cagA* gene from two of the seven overlapping cosmids. The identity of the PCR product harboring *cagA* was verified by sequencing. Further analysis by Southern hybridization and restriction analysis mapped *cagA* to a 4.2 *Bam*HI fragment, which was located approximately 15 kb from the ORF1/2 locus (Fig. 4). Encouraged by the fact that these two loci are indeed clustered, we sequenced the *cagA* locus and the region between the

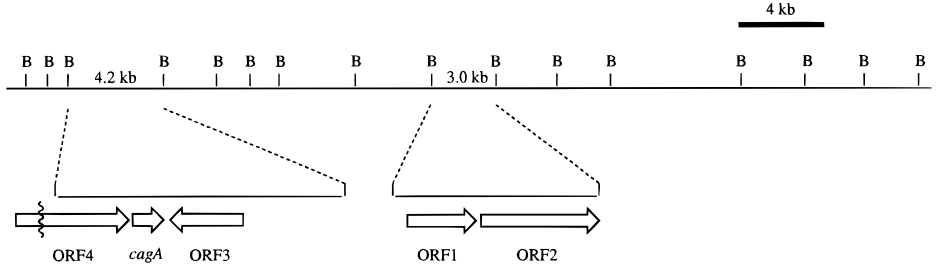

FIG. 4. Tentative restriction map of the putative deoxysugar biosynthesis locus and the resistance locus of the gene cluster for **3** in *S. globisporus*. Only selected restriction sites (B, *Bam*HI) are shown on the map.

cagA and ORF1/2 loci and made putative functional assignments by comparison of the gene products to the protein database. To our delight, flanking *cagA* are ORF3 and an incomplete ORF4 (Fig. 4), whose deduced products are highly homologous to glucose-1-phosphate nucleotidyltransferases and glycosyl aminotransferases, suggesting ORF3 and ORF4 to encode the putative E_p and E_{am} for the biosynthesis of **18**, respectively. Finally, along with several other orfs, ORF5 and an incomplete ORF6 were found between the *cagA* and ORF1/2 loci (data not shown). The deduced products of the latter showed high similarity to known glycosyl epimerases and glycosyltransferases. We propose that ORF5 encodes the putative E_{ep} for the biosynthesis of **18** while ORF6 encodes the putative glycosyltransferase catalyzing the attachment of the fully modified sugar **18** to the enediyne moiety of **3**. Among the putative genes needed for the biosynthesis of **18** (Fig. 6), we now have candidates for all steps except E_{met} and these results, together with ORF2 and *cagA*, provide strong support that we have cloned the putative gene cluster for **3** in *S. globisporus*, consisting of at least five structural genes and two resistance genes.

Finally, we have confirmed that the cloned 55-kb *S. globisporus* DNA encodes C-1027 biosynthesis by direct gene disruption of ORF1. The corresponding *S. globisporus* mutants, in which the putative NDP-glucose-4,6-dehydratase gene has been disrupted, no longer produce C-1027. Chromosomal walking beyond the *cagA* and ORF1/2 loci to clone the entire C-1027 gene cluster and functional analysis of genes encoding C-1027 biosynthesis are currently under investigation.

10-MEMBERED NONCHROMOPROTEIN ENEDIYNES

Calicheamicin γ_1^I (Fig. 5; **9**) from *Micromonospora echinospora* spp. *calichensis* is the most prominent of the 10-membered nonchromoprotein enediynes and its unprecedented molecular architecture in conjunction with its superb biological activity make **9** an excellent target to study biosynthetically. Of the two distinct structural regions within **9** (*71–73*), the aryltetrasaccharide displays a highly unusual series of glycosidic, thioester, and hydroxylamine linkages and serves to deliver the drug to specific tracts (5′-TCCT-3′ and 5′-TTTT-3′) within the minor groove of DNA (*74*), while the aglycone of **9** consists of a highly functionalized bicyclo[7.3.1]tridecadiynene core structure with an allylic trisulfide serving as the triggering mechanism (*75*). Three additional bicyclo[7.3.1]tridecadiynenes are also suspected to be biosynthetically related to **9**. First, the esperamicins (**7**) from *Actinomadura verrucosospora* share an almost identical bicyclic enediyne core (with the exception of an additional hydroxyl group) and have sugar appendages with similarly unusual structural motifs (*76–79*). Second, namenamicin (**8**) was very recently isolated from the thin encrusting orange asician *Polysyncraton lithostrotum* and contains the **9** aglycone with a distinct carbohydrate domain which contributes to the slightly different DNA recognition pattern (*80*). It also should be mentioned that 16 *Micromonospora* species have been isolated from the tissue of *P. lithostrotum*, suggesting **8** is most likely also of microbial origin.

The final member of the 10-membered ring family, dynemicin (**6**) from *Michromonospora chersina*, contains a novel combination of an anthraquinone moiety fused with a 10-membered bridging enediyne ring (*81*). Dynemicin is extremely cytotoxic and causes both single- and double-stranded DNA cuts. In addition, due to common

structural elements, the extensive progress in anthracycline antibiotic biosynthesis should be beneficial in understanding the biosynthesis of **6** (*82–84*).

Dynemicin labeling studies. In a manner similar to the studies focusing upon the biosynthesis of **1** (*51*), [1-^{13}C]- and [1,2-^{13}C]acetate, L-[*methyl*-^{13}C]methionine, and [α-^{15}N,^{13}C]glycine incorporation into **6** (Fig. 3b) revealed that all carbons (with the exception of the C_6 methoxy group) were acetate-derived and the four carbons of both yne moieties originated from separate acetate units (*81,85*). This was in contrast to the ^{13}C-acetate-labeling patterns of **1**, in which the corresponding yne carbons of **1** emanate from single acetates (Fig. 3a), and suggested the biosynthesis of the 9-membered ring and 10-membered ring enediynes may be fundamentally distinct (*51*). That the anthraquinone and enediyne moieties of **6** are not connected by any common acetate units led Tokiwa *et al.* to propose each are synthesized as separate heptaketides with a subsequent acetate addition to give rise to the C_5 carboxylate (*81,85*). The N_1

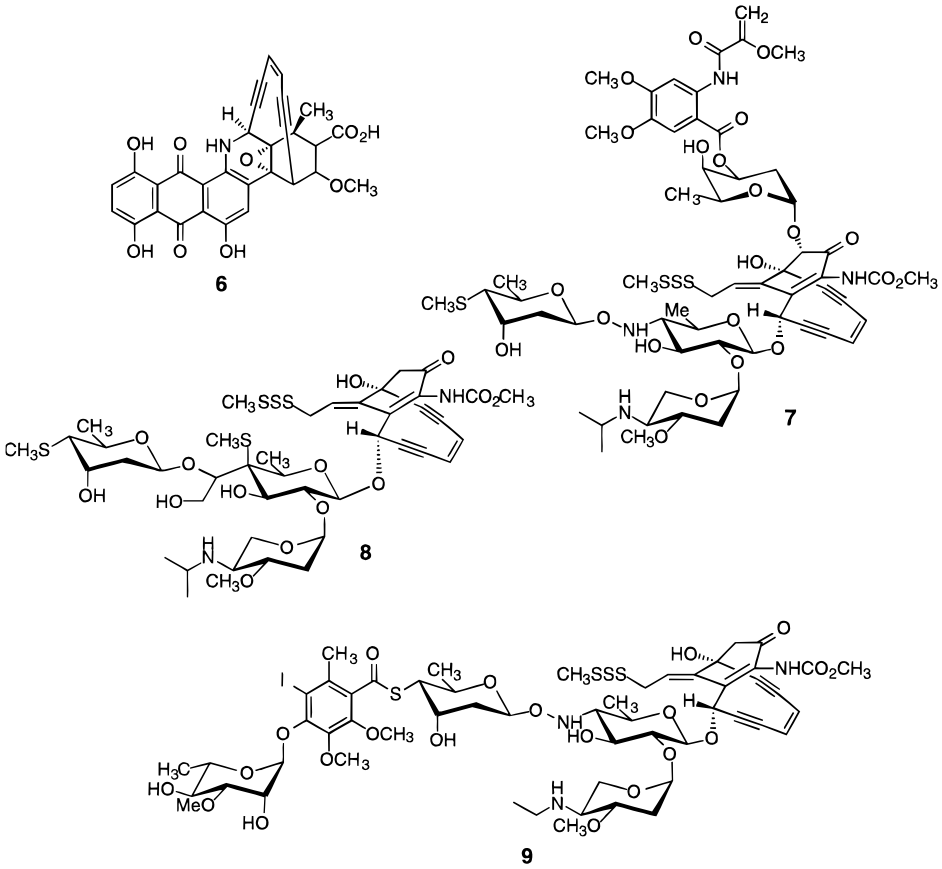

FIG. 5. The structural representation of the naturally occuring 10-membered ring enediynes dynemicin (**6**), esperamicin (**7**), namenamicin (**8**), and calicheamicin γ_1^I (**9**).