(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2005 (08.12.2005)

PCT

(10) International Publication Number
WO 2005/116200 A2

(51) International Patent Classification^7: **C12N 9/10**, C12P 21/06, 23/00

(21) International Application Number: PCT/US2005/015649

(22) International Filing Date: 5 May 2005 (05.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/568,334 5 May 2004 (05.05.2004) US

(71) Applicant *(for all designated States except US)*: **REGENTS OF THE UNIVERSITY OF MINNESOTA** [US/US]; 450 McNamara Alumni Center, 200 Oak Street S.E., Minneapolis, Minnesota 55455-2070 (US).

(71) Applicants and
(72) Inventors: SHERMAN, David, H. [US/US]; 2574 Bedford Rd., Ann Arbor, MI 48104 (US). **MCGARVEY, Nathan** [CA/US]; 4623 Bloomington Avenue S, Minneapolis, MN 55407 (US). **BECK, Zachary, Q.** [US/US]; 434 Third Street, Ann Arbor, MI 48103 (US).

(74) Agent: PARSONS, Angela, M., PH.D.; Fish & Richardson P.C., P.A., 3300 Dain Rauscher Plaza, 60 South Sixth Street, Minneapolis, MN 55402 (US).

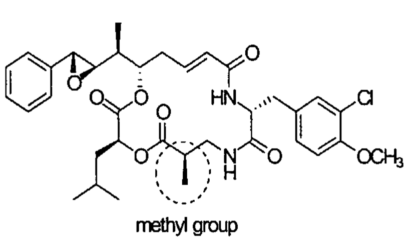
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

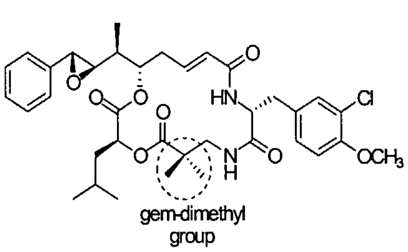
Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACIDS AND POLYPEPTIDES INVOLVED IN THE PRODUCTION OF CRYPTOPHYCIN



Cryptophycin 1
$C_{35}H_{43}ClN_2O_8$
Exact Mass: 654.27
Mol. Wt.: 655.18
Predominant compound produced by Nostoc sp ATCC 53789



Cryptophycin 52
$C_{36}H_{45}ClN_2O_8$
Exact Mass: 668.29
Mol. Wt.: 669.20
Synthetic compound chosen for clinical evaluation

(57) Abstract: The present invention provides polypeptides involved in cryptophycin biosynthesis and the nucleic acid molecules that encode such polypeptides. The nucleic acid molecules and polypeptides of the invention or variants thereof can be used in the methods of the invention to produce cryptophycins.

WO 2005/116200 A2

NUCLEIC ACIDS AND POLYPEPTIDES INVOLVED IN THE PRODUCTION OF CRYPTOPHYCIN

TECHNICAL FIELD

This invention relates to production of cryptophycin, and more particularly to the polypeptides involved in the biosynthesis of cryptophycin and the nucleic acids encoding such polypeptides.

BACKGROUND

Cryptophycins are novel macrolides first isolated from blue-green algae (*Nostoc* sp. GSV22 and *Nostoc* sp. ATCC 53789) and are potent tumor selective cytotoxins *in vivo*. Many syntheses of the major natural products, cryptophycins 1-4, and a wide range of analogs have been published. For example, cryptophycins have been synthesized by a convergent method in which four components, Unit A, Unit B, Unit C, and Unit D (Golakati et al., 1995, *J. Am. Chem. Soc.*, 117(49):12031), are coupled together to form the final product (see, for example, U.S. Patent No. 6,013,626). In other methods, novel semi-synthetic compounds are generated, for example, by converting the epoxide of a natural cryptophycin to a carbon-carbon double bond (see, for example, U.S. Patent Nos. 4,845,085 and 4,845,086). Stereo-selective addition of functional groups is often problematic during chemical synthesis of cryptophycins, however. Therefore, few of the methodologies for cryptophycin syntheses are considered viable or practical on a commercial scale.

SUMMARY

The present invention provides polypeptides involved in cryptophycin biosynthesis and the nucleic acid molecules that encode such polypeptides. The nucleic acid molecules and polypeptides of the invention or variants thereof can be used in the methods of the invention to produce cryptophycins.

In one aspect, the invention provides an isolated nucleic acid molecule that includes a nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to the sequence shown in SEQ ID NO:1 or to a fragment thereof. Such a sequence encodes at least one enzyme involved in biosynthesizing cryptophycin.

The invention further provides for a vector containing such a nucleic acid molecule, and host cells containing such vectors. The invention also provides for cryptophycin or cryptophycin analogues made by such host cells.

In another aspect, the invention provides methods of producing cryptophycin. Such a method generally includes the step of culturing the above-described host cells in the presence of an appropriate substrate and under conditions appropriate for the production of cryptophycin. Such a method can further include the step of purifying the cryptophycin.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30, or to a fragment thereof, wherein the nucleic acid sequence encodes a polypeptide that exhibits functional activity.

The invention further provides for vectors containing such nucleic acid molecules, and host cells containing such vectors. The invention also provides for intermediates in cryptophycin biosynthesis made by such host cells.

The invention further provides a polypeptide encoded by the nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30, or to a fragment thereof. Such polypeptides can have the sequence shown in 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, or 31, respectively.

In still another aspect, the invention provides an isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to nucleotides 9,199 to 10,032 of SEQ ID NO:6, or to a fragment thereof, wherein the nucleic acid sequence encodes a polypeptide that exhibits thioesterase activity under appropriate conditions.

In yet another aspect, the invention provides an isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to the sequence shown in SEQ ID NO:8, or to a fragment thereof, wherein the nucleic acid sequence encodes a polypeptide that exhibits epoxidase activity under appropriate conditions.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% (e.g., 85%, 90%, 95%, 99%, or 100%) sequence identity to the sequence shown in SEQ ID NO:14, or to a fragment thereof, wherein the nucleic acid sequence encodes a polypeptide that exhibits halogenase activity under appropriate conditions.

In another aspect, the invention provides for methods of producing an intermediate in cryptophycin biosynthesis. Such a method includes culturing one or more host cells that contain one or more vectors comprising one or more of the nucleic acid sequences shown in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30 in the presence of one or more appropriate substrates under conditions appropriate for production of an intermediate in cryptophycin biosynthesis.

Representative appropriate conditions include pH, media, temperature, and/or the presence or absence of co-factors. Representative substrates and intermediates in cryptophycin biosynthesis include Cryptophycin 2, 3, 4, 5, 16, and 17 (see Figure 1B).

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the drawings and detailed description, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 shows cryptophycin structures. FIG. 1A is a natural cryptophycin, Cryptophycin 1, and a synthetic cryptophycin, Cryptophycin 52. FIG. 1B illustrates the diversity of natural cryptophycins isolated from *Nostoc* spp.

FIG. 2 is a schematic of the lineage of biologically active cryptophycins.

FIG. 3 is a schematic of the modular structure of the cryptophycins and retro-biosynthesis assembly.

FIG. 4 is a schematic of cosmid pDAM163 and genes identified with relationships to cryptophycin biosynthesis.

FIG. 5 is the nucleotide sequence of the cloned insert of pDAM163 (SEQ ID NO:1).

FIG. 6 is the nucleotide and amino acid sequences of the genes and polypeptides involved in cryptophycin biosynthesis (SEQ ID NOs:2-35).

FIG. 7 shows SEQ ID NOs:36-41, which have 75%, 80%, 85%, 90%, 95%, and 99% sequence identity, respectively, to SEQ ID NO:2.

FIG. 8 is a schematic depicting the predicted cryptophycin assembly line.

FIG. 9 is a schematic demonstrating that the cryptophycin epoxidase (CrpE) has substrate flexibility but a high degree of steroselectivity.

FIG. 10 is a schematic of the synthesis of SNAC substrates.

FIG. 11 is a schematic of cryptophycin thioesterase-catalyzed conversion of substrate 1 to products 2, 3, and 4 in 0.1 M NaPi buffer (pH 8.0) containing 4% DMSO.

FIG. 12 are graphs of cryptophycin thioesterase-catalyzed hydrolysis of substrate 1 using 1.4 μM thioesterase in 50 μL reactions containing 0.1 M NaH_2PO_4 and 4% DMSO.

FIG. 13 is a schematic of cryptophycin thioesterase-catalyzed cyclization and hydrolysis of the *seco*-SNAC-ester of arenastatin and the *seco*-SNAC-ester of the vinyl derivative of arenastatin.

DESCRIPTION OF SEQUENCES

SEQ ID NO:1 is the nucleotide sequence of the cloned insert of pDAM163.

SEQ ID NO:2 is the nucleotide sequence of crpA.

SEQ ID NO:3 is the amino acid sequence of CrpA.

SEQ ID NO:4 is the nucleotide sequence of crpB.

SEQ ID NO:5 is the amino acid sequence of CrpB.

SEQ ID NO:6 is the nucleotide sequence of crpC.

SEQ ID NO:7 is the amino acid sequence of CrpC.

SEQ ID NO:8 is the nucleotide sequence of crpF.

SEQ ID NO:9 is the amino acid sequence of CrpF.

SEQ ID NO:10 is the nucleotide sequence of crpG.

SEQ ID NO:11 is the amino acid sequence of CrpG.

SEQ ID NO:12 is the nucleotide sequence of crpH.

SEQ ID NO:13 is the amino acid sequence of CrpH.

SEQ ID NO:14 is the nucleotide sequence of crpJ.

SEQ ID NO:15 is the amino acid sequence of CrpJ.

SEQ ID NO:16 is the nucleotide sequence of crpM.

SEQ ID NO:17 is the amino acid sequence of CrpM.

SEQ ID NO:18 is the nucleotide sequence of crpN.

SEQ ID NO:19 is the amino acid sequence of CrpN.

SEQ ID NO:20 is the nucleotide sequence of crpP.

SEQ ID NO:21 is the amino acid sequence of CrpP.

SEQ ID NO:22 is the nucleotide sequence of crpU.

SEQ ID NO:23 is the amino acid sequence of CrpU.

SEQ ID NO:24 is the nucleotide sequence of crpV.

SEQ ID NO:25 is the amino acid sequence of CrpV.

SEQ ID NO:26 is the nucleotide sequence of crpX.

SEQ ID NO:27 is the amino acid sequence of CrpX.

SEQ ID NO:28 is the nucleotide sequence of crpY.

SEQ ID NO:29 is the amino acid sequence of CrpY.

SEQ ID NO:30 is the nucleotide sequence of crpZ.

SEQ ID NO:31 is the amino acid sequence of CrpZ.

SEQ ID NO:32 is a nucleotide sequence having 75% sequence identity to SEQ ID NO:2.

SEQ ID NO:33 is a nucleotide sequence having 80% sequence identity to SEQ ID NO:2.

SEQ ID NO:34 is a nucleotide sequence having 85% sequence identity to SEQ ID NO:2.

SEQ ID NO:35 is a nucleotide sequence having 90% sequence identity to SEQ ID NO:2.

SEQ ID NO:36 is a nucleotide sequence having 95% sequence identity to SEQ ID NO:2.

SEQ ID NO:37 is a nucleotide sequence having 99% sequence identity to SEQ ID NO:2.

SEQ ID NO:38 is the sequence of an oligonucleotide.

SEQ ID NO:39 is the sequence of an oligonucleotide.

SEQ ID NO:40 is the sequence of an oligonucleotide.

SEQ ID NO:41 is the sequence of an oligonucleotide.

DETAILED DESCRIPTION

Cryptophycin biosynthesis is accomplished via a mixed Type I PKS/NRPS system. Manipulation of polyketide synthetases (PKSs) and non-ribosomal peptide synthetases (NRPSs) through mutasynthesis, combinatorial biosynthesis, and directed biosynthesis feeding (chemoenzymatic synthesis) has been described for many PKS and NRPS polypeptides. The identification of the corresponding genes allows for these types of approaches with the cryptophycin system. It is possible that altering the PKS enzyme for Unit A formation or the NRPS for Unit B, C, and D formation could generate a wide variety of new cryptophycins. With this invention, it is also possible to incorporate these enzymes in "total synthesis" of cryptophycins to lower the cost and increase the overall yields. For example, the ability of biosynthetic enzymes to exhibit high levels of stereo-chemical control and relaxed substrate specificity, and the sensitivity of the biological and chemical assays for identifying cryptophycins, allow for production of rational "biologically" derived cryptophycins that have superior properties.

Cryptophycins

Figure 1A shows Cryptophycin 1 and Cryptophycin 52. Cryptophycin 52 is nearly identical to Cryptophycin 1, the most active natural compound, except for the presense of gem-dimethyl on the β-alanine unit of Cryptophycin 52 instead of the methyl group on Cryptophycin 1. Figure 1B shows numerous other natural cryptophycins that have been isolated from *Nostoc* spp. A chlorohydrin analog (Cryptophycin 309; see U.S. Publication No. 20020065261 and Figure 2 of the instant application) has been identified and has been shown to be much more active than the current clinical candidate, Cryptophycin 52.

Purification

Routine chromatographic techniques such as high-performance liquid chromatography (HPLC) or thin-layer chromatography (TLC) can be used to purify cryptophycins. See, for example, U.S. Patent No. 5,952,298, which describes specific HPLC conditions for purifying different cryptophycins.

Structure Identification

The structures of cryptophycins can be determined using methodology that is well known to those of skill in the art. Mass spectral analysis can be used, for example. Proton and carbon NMR data obtained from COSY, HMQC, HMBC, and NOESY spectra allows determination of the gross structures of the depsipeptide-type compounds. The presence of the various hydroxy and amino acid units in each compound can be detected by gas chromatographic mass spectral analysis. Total structures, including absolute stereochemistries, can be determined using a combination of chemical degradative and analytical techniques on cryptophycin compounds.

Anti-Fungal Activity

Cryptophycin compounds can be tested against fungal organisms known to be sensitive to such compounds using, for example, a disk-diffusion assay such as a Corbett assay (see, for example, Kemp, 1980, Organic Chemistry, Worth Publishers Inc.). The anti-fungal activity of a cryptophycin is usually correlated with the size of the zone of inhibition (i.e., an area of no microbial growth around an antimicrobial agent in a disk-diffusion test). An organism that can be used to evaluate the anti-fungal activity of a cryptophycin is *Candida albicans*.

Anti-Cancer Activity

The anti-cancer activity of a cryptophycin can be examined using a number of different assays such as cell proliferation assays and cell cycle arrest assays. In addition, cytoskeletal structures such as tubulin can be examined using, for example, immunofluorescence assays. See, for example, U.S. Patent No. 5,945,315.

Cryptophycins can be evaluated for anti-cancer activity against a number of different cell types. For example, murine leukemia cells (e.g., L1210 or P388), murine solid tumor cells (e.g., colon adenocarcinoma 38, pancreatic ductal adenocarcinoma 03, mammary adenocarcinoma M16/M17), human solid tumor cells (e.g., colon CX-1, HCT8, H116, lung H125, mammary MX-1, MCF-7), low malignancy fibroblast cells (e.g., LML), human nasopharyngeal carcinoma cells (e.g., KB), human colon carcinoma cells (e.g., LoVo), and human ovarian carcinoma cells (e.g., SKOV3) can be used to evaluate the anti-cancer activity of a cryptophycin. For example, a disk diffusion assay much like the Corbett assay (Kemp, *supra*) commonly used in antifungal and antibacterial testing can be used to evaluate the anti-cancer activity of a cryptophycin. A zone of inhibition can be correlated with the anti-cancer activity of a cryptophycin.

Nucleic Acids and Polypeptides Involved in Cryptophycin Biosynthesis

Approximately 45 kb of DNA corresponding to the genes predicted to be involved in cryptophycin biosynthesis were cloned into a cosmid designated pDAM163 and sequenced. Figure 4 shows a schematic of pDAM163, while Figure 5 shows the nucleotide sequence of the cloned insert of pDAM163 (SEQ ID NO:1). This cosmid replicated efficiently and stably in well-developed fermentation strains such as *E. coli* B (*E.coli* BL21*pLys*) and *E. coli* K (DH5α) derivatives. Expressing the coding regions contained within pDAM163 can result in the production of cryptophycin in the *E. coli* strains. A variety of microorganisms such as bacteria (e.g., *Escherichia coli*), yeast (e.g., *Pichia pastoris* or *Saccharomyces cerevisiae*), or fungi (e.g., *Neurospora crassa*) that include expression constructs such as pDAM163 or variants thereof can be used to generate cryptophycins.

The components of the biosynthetic pathway are summarized in Table 1, which provides information related to the putative function of each polypeptide.

Table 1. Nucleic Acids and Polypeptides Involved in Cryptophycin Biosynthesis

Designation (SEQ ID NO: Nucleic Acid/Polypeptide)	Length (nt/amino acids)	Putative Function
crpA (SEQ ID NO:2/3)	3690/1229	Polyketide synthetase (PKS)
crpB (SEQ ID NO:4/5)	5832/1943	Nonribosomal Peptide Synthetase (NRPS)
crpC (SEQ ID NO:6/7)	10,032/3343	NRPS
crpF (SEQ ID:NO:8/9)	1353/450	Cytochrome p450 (epoxidase)
crpG (SEQ ID NO:10/11)	885/294	Iron-dependent non-heme hydroxylase
crpH (SEQ ID NO:12/13)	333/110	Aspartate decarboxylase
crpI	-[a]	IS1327 Transposase
crpJ (SEQ ID NO:14/15)	1476/491	Non-heme halogenase
crpK	-[a]	IS892-orf2
crpL	-[a]	IS892-orf1
crpM (SEQ ID NO:16/17)	1383/460	ISRSO13 Transposase
crpN (SEQ ID NO:18/19)	942/313	Benzoyl-CoA reductase/2-Hydroxyglutaryl-CoA dehydratase
crpO	[b]	pvdE type regulator
crpP (SEQ ID NO:20/21)	633/210	Thioredoxin
crpU (SEQ ID NO:22/23)	468/155	N- acetyltransferase
crpV (SEQ ID NO:24/25)	2355/784	Large exoprotein involved in heme utilization
crpW	-[a]	TPR repeat protein
crpX (SEQ ID NO:26/27)	399/132	Chorismate mutase/Prephenate dehydrogenase
crpY (SEQ ID NO:28/29)	645/214	2-Hydroxychromenene-2-carboxylate isomerase
crpZ (SEQ ID NO:30/31)	273/90	3-Dehydroquinate synthase

[a], no open reading frame identified; [b], multiple open reading frames identified.

Nucleic Acid Molecules

The present invention is based, in part, on the identification of nucleic acid molecules that encode polypeptides involved in cryptophycin synthesis. Particular nucleic acid molecules of the invention include the sequences shown in SEQ ID NOs:1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. As used herein, the term "nucleic acid molecule" can include DNA molecules and RNA molecules and analogs of the DNA or RNA molecule generated using nucleotide analogs. A nucleic acid molecule of the invention can be single-stranded or double-stranded, and the strandedness will depend upon its intended use.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NOs: 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. Nucleic acid molecules of the invention include molecules that are at least 10 nucleotides in length and that have at least 75% sequence identity (e.g., at least 80%, 85%, 90%, 95%, or 99% sequence identity) to any of SEQ ID NOs: 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. Nucleic acid molecules that differ in sequence from the nucleic acid sequences shown in SEQ ID NOs: 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30 can be generated by standard techniques, such as site-directed mutagenesis or PCR-mediated mutagenesis. In addition, nucleotide changes can be introduced randomly along all or part of a nucleic acid molecule of the invention, such as by saturation mutagenesis. Alternatively, nucleotide changes can be introduced into a sequence by chemically synthesizing a nucleic acid molecule having such changes.

In calculating percent sequence identity, two sequences are aligned and the number of identical matches of nucleotides or amino acid residues between the two sequences is determined. The number of identical matches is divided by the length of the aligned region (i.e., the number of aligned nucleotides or amino acid residues) and multiplied by 100 to arrive at a percent sequence identity value. It will be appreciated that the length of the aligned region can be a portion of one or both sequences up to the full-length size of the shortest sequence. It also will be appreciated that a single sequence can align with more than one other sequence and hence, can have different percent sequence identity values over each aligned region. It is noted that the percent identity value is usually rounded to the nearest integer. For example, 78.1%, 78.2%, 78.3%, and

78.4% are rounded down to 78%, while 78.5%, 78.6%, 78.7%, 78.8%, and 78.9% are rounded up to 79%. It is also noted that the length of the aligned region is always an integer.

The alignment of two or more sequences to determine percent sequence identity is performed using the algorithm described by Altschul et al. (1997, Nucleic Acids Res., 25:3389-3402) as incorporated into BLAST (basic local alignment search tool) programs, available at ncbi.nlm.nih.gov on the World Wide Web. BLAST searches can be performed to determine percent sequence identity between a nucleic acid molecule of the invention and any other sequence or portion thereof aligned using the Altschul et al. algorithm. BLASTN is the program used to align and compare the identity between nucleic acid sequences, while BLASTP is the program used to align and compare the identity between amino acid sequences. When utilizing BLAST programs to calculate the percent identity between a sequence of the invention and another sequence, the default parameters of the respective programs are used. Sequence analysis of the nucleic acid sequences as performed herein used BLAST version 2.2.8 (updated on February 10, 2004).

The sequences of representative nucleic acids of the invention having 75%, 80%, 85%, 90%, 95%, and 99% sequence identity to SEQ ID NO:2 are shown in Figure 7 (SEQ ID NOs:32-37, respectively). Such sequences can be generated using a computer or by hand. The nucleic acid sequences shown in SEQ ID NOs:32-37 were generated by hand by randomly changing 25 nucleotides out of every 100 nucleotides of SEQ ID NO:2, 2 out of every 10, 15 out of every 100, 1 out of every 10, 5 out of every 100, or 1 nucleotide out of every 100 nucleotides of SEQ ID NO:2, respectively. By “changing,” it is meant that the nucleotide at a particular position is replaced randomly with one of the other three nucleotides. It is apparent to those of ordinary skill in the art that any nucleic acid molecule within the scope of the invention can be generated using the same method described herein (i.e., by similarly changing nucleotides within the sequence of SEQ ID NOs: 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30).

Nucleic acid fragments are included in the invention. Nucleic acid fragments suitable for use in the invention are those fragments that encode a polypeptide having functional activity. These fragments can be called “functional fragments,” although it is understood that it is not the nucleic acid that possesses functionality.

For example, nucleic acid fragments of crpA (SEQ ID NO:2) can be at least 185 nucleotides in length (e.g., 185, 200, 269, 323, 392, 442, 481, 530, 590, 618, 678, 752, 804, 876, 922, 1033, 1146, 1278, 1399, 1478, 1567, 1645, 1712, 1888, 1907, 2061, 2154, 2233, 2357, 2491, 2543, 2677, 2781, 2802, 2944, 3012, 3143, 3257, 3368, 3459, 3548, or 3689); nucleic acid fragments of crpB (SEQ ID NO:4) can be at least 292 nucleotides in length (e.g., 292, 306, 382, 461, 592, 715, 825, 947, 1059, 1172, 1236, 1358, 1496, 1590, 1671, 1774, 1889, 1923, 2047, 2135, 2265, 2346, 2477, 2588, 2667, 2754, 2863, 2954, 3084, 3126, 3278, 3345, 3412, 3551, 3670, 3781, 3890, 3910, 4044, 4123, 4266, 4378, 4423, 4513, 4622, 4783, 4822, 4989, 5002, 5156, 5237, 5368, 5486, 5572, 5691, 5765, or 5831); nucleic acid fragments of crpC (SEQ ID NO:6) can be at least 502 nucleotides in length (e.g., 502, 624, 738, 829, 914, 1026, 1138, 1257, 1318, 1452, 1525, 1637, 1768, 1828, 1987, 2074, 2183, 2294, 2338, 2444, 2557, 2637, 2789, 2816, 2942, 3067, 3178, 3227, 3348, 3459, 3504, 3684, 3759, 3812, 3943, 4005, 4276, 4495, 4658, 4827, 5048, 5276, 5424, 5608, 5877, 6034, 6269, 6447, 6632, 6874, 7006, 7284, 7472, 7647, 7814, 8038, 8246, 8459, 8644, 8888, 9053, 9298, 9436, 9666, 9878, or 10,032); nucleic acid fragments of crpF (SEQ ID NO:8) can be at least 68 nucleotides in length (e.g., 68, 74, 82, 88, 95, 105, 168, 235, 367, 489, 524, 665, 784, 863, 925, 1064, 1138, 1279, or 1352); nucleic acid fragments of crpG (SEQ ID NO:10) can be at least 44 nucleotides in length (e.g., 44, 54, 58, 67, 74, 83, 97, 107, 189, 267, 345, 457, 536, 679, 772, or 884); nucleic acid fragments of crpH (SEQ ID NO:12) can be at least 33 nucleotides in length (e.g., 33, 45, 52, 68, 73, 84, 93, 108, 168, 216, 248, 293, 312, or 332); nucleic acid fragments of crpJ (SEQ ID NO:14) can be at least 74 nucleotides in length (e.g., 74, 106, 187, 254, 304, 379, 467, 522, 592, 667, 714, 781, 859, 911, 978, 1049, 1138, 1273, 1347, 1405, or 1475); nucleic acid fragments of crpM (SEQ ID NO:16) can be at least 69 nucleotides in length (e.g., 69, 136, 216, 362, 486, 592, 647, 781, 844, 919, 1049, 1138, 1274, or 1382); nucleic acid fragments of crpN (SEQ ID NO:18) can be at least 94 nucleotides in length (e.g., 94, 182, 261, 358, 442, 580, 625, 740, 862, or 941); nucleic acid fragments of crpP (SEQ ID NO:20) can be at least 32 nucleotides in length (e.g., 32, 85, 120, 175, 232, 286, 310, 379, 433, 561, or 632); nucleic acid fragments of crpU (SEQ ID NO:22) can be at least 23 nucleotides in length (e.g., 23, 74, 112, 178, 215, 280, 315, 369, 402, or 467); nucleic acid fragments of crpV (SEQ ID NO:24) can be at least 118 nucleotides in length (e.g., 118, 235, 366, 440, 521, 636, 783, 852, 918, 1044, 1168, 1238, 1350, 1448, 1569, 1722, 1838, 1924, 2052, 2167, 2288, or 2354); nucleic acid fragments of crpX (SEQ ID NO:26) can be at least 60 nucleotides in length (e.g., 60, 98, 137, 182, 214, 278, 308, 357, or 398); nucleic acid fragments of crpY (SEQ ID NO:28) can be at least 32 nucleotides in length (e.g., 32, 74, 121, 169, 204, 263, 298, 355, 391, 426, 484, 523, 577, 624, or 644); and nucleic acid fragments of crpZ (SEQ ID NO:30) can be at least 27 nucleotides in length (e.g., 27, 68, 103, 158, 193, 243, or 272). Based on contemporaneous public database searches, such fragments appear not to have more than 85% sequence identify to sequences in the public databases.

As used herein, an "isolated" nucleic acid molecule is a nucleic acid molecule that is separated from other nucleic acid molecules that are usually associated with the reference nucleic acid molecule in the genome. Thus, an "isolated" nucleic acid molecule includes, without limitation, a nucleic acid molecule that is free of sequences that naturally flank one or both ends of the nucleic acid in the genome of the organism from which the isolated nucleic acid molecule is derived (e.g., a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease digestion). Such an isolated nucleic acid molecule is generally introduced into a vector (e.g., a cloning vector, or an expression vector) for convenience of manipulation or to generate a fusion nucleic acid molecule. In addition, an isolated nucleic acid molecule can include an engineered nucleic acid molecule such as a recombinant or a synthetic nucleic acid molecule. A nucleic acid molecule existing among hundreds to millions of other nucleic acid molecules within, for example, a nucleic acid library (e.g., a cDNA, or genomic library) or a portion of a gel (e.g., agarose, or polyacrylamine) containing restriction-digested genomic DNA is not to be considered an isolated nucleic acid.

Isolated nucleic acid molecules of the invention can be obtained using techniques routine in the art. For example, isolated nucleic acids within the scope of the invention can be obtained using any method including, without limitation, recombinant nucleic acid technology, and/or the polymerase chain reaction (PCR). General PCR techniques are described, for example in PCR Primer: A Laboratory Manual, Dieffenbach & Dveksler, Eds., Cold Spring Harbor Laboratory Press, 1995. Recombinant nucleic acid techniques include, for example, restriction enzyme digestion and ligation, which can be used to isolate a nucleic acid molecule of the invention. Isolated nucleic acids of the invention also can be chemically synthesized, either as a single nucleic acid molecule or as a series of oligonucleotides. In addition, isolated nucleic acid molecules of the invention also can be obtained by mutagenesis. For example, an isolated nucleic acid that shares identity with an art known sequence can be mutated using common molecular cloning techniques (e.g., site-directed mutagenesis). Possible mutations include, without limitation, deletions, insertions, substitutions, and combinations thereof.

Vectors containing nucleic acid molecules that encode polypeptides involved in cryptophycin synthesis also are provided by the invention. Vectors, including expression vectors, suitable for use in the present invention are commercially available and/or produced by recombinant DNA technology methods routine in the art. A vector containing a nucleic acid molecule of the invention can have elements necessary for expression operably linked to such a nucleic acid molecule, and further can include sequences such as those encoding a selectable marker (e.g., an antibiotic resistance gene), and/or those that can be used in purification of a polypeptide involved in cryptophycin synthesis (e.g., 6xHis tag).

Vectors containing nucleic acid molecules encoding polypeptides involved in cryptophycin synthesis were deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard Manassas, VA 20110, on ________________, and assigned Accession Numbers __________, ____________, ____________, and ____________. Each deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Elements necessary for expression include nucleic acid sequences that direct and regulate expression of nucleic acid coding sequences. One example of an element necessary for expression is a promoter sequence. Elements necessary for expression also can include introns, enhancer sequences, response elements, or inducible elements that modulate expression of a nucleic acid molecule of the invention. Elements necessary for expression can be of bacterial, yeast, insect, mammalian, or viral origin and vectors can contain a combination of elements from different origins. Elements necessary for expression are described, for example, in Goeddel, 1990, *Gene Expression Technology: Methods in Enzymology*, 185, Academic Press, San Diego, CA. As used herein, operably linked means that a promoter and/or other regulatory element(s) are positioned in a vector relative to a nucleic acid molecule of the invention in such a way as to direct or regulate expression of the nucleic acid molecule. Many methods for introducing nucleic acids into host cells, both *in vivo* and *in vitro*, are well known to those skilled in the art and include, without limitation, calcium phosphate precipitation, electroporation, heat shock, lipofection, microinjection, and viral-mediated nucleic acid transfer.

Another aspect of the invention pertains to host cells into which a vector of the invention, e.g., an expression vector, or an isolated nucleic acid molecule of the invention has been introduced. The term "host cell" refers not only to the particular cell but also to the progeny or potential progeny of such a cell. A host cell can be any prokaryotic or eukaryotic cell. For example, nucleic acid molecules of the invention can be expressed in bacterial cells such as *E. coli*, or in insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Conditions for amplification of a nucleic acid and detection of an amplification product are known to those of skill in the art (see, e.g., *PCR Primer: A Laboratory Manual*, 1995, Dieffenbach & Dveksler, Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; and U.S. Patent Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188). Modifications to the original PCR also have been developed. For example, anchor PCR, RACE PCR, or ligation chain reaction (LCR) are additional PCR methods known in the art (see, e.g., Landegran et al., 1988, *Science*, 241:1077-1080; and Nakazawa et al., 1994, *Proc. Natl. Acad. Sci. USA*, 91:360-364).

Hybridization between nucleic acid molecules is discussed in detail in Sambrook et al. (1989, *Molecular Cloning: A Laboratory Manual*, 2^{nd} Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; Sections 7.37-7.57, 9.47-9.57, 11.7-11.8, and 11.45-11.57). For oligonucleotide probes less than about 100 nucleotides, Sambrook et al. discloses suitable Southern blot conditions in Sections 11.45-11.46. The Tm between a sequence that is less than 100 nucleotides in length and a second sequence can be calculated using the formula provided in Section 11.46. Sambrook et al. additionally discloses prehybridization and hybridization conditions for a Southern blot that uses oligonucleotide probes greater than about 100 nucleotides (see Sections 9.47-9.52). Hybridizations with an oligonucleotide greater than 100 nucleotides generally are performed 15-25°C below the Tm. The Tm between a sequence greater than 100 nucleotides in length and a second sequence can be calculated using the formula provided in Sections 9.50-9.51 of Sambrook et al. Additionally, Sambrook et al. recommends the conditions indicated in Section 9.54 for washing a Southern blot that has been probed with an oligonucleotide greater than about 100 nucleotides.

The conditions under which membranes containing nucleic acids are prehybridized and hybridized, as well as the conditions under which membranes containing nucleic acids are washed to remove excess and non-specifically bound probe can play a significant role in the stringency of the hybridization. Such hybridizations and washes can be performed, where appropriate, under moderate or high stringency conditions. Such conditions are described, for example, in Sambrook et al. section 11.45-11.46. For example, washing conditions can be made more stringent by decreasing the salt concentration in the wash solutions and/or by increasing the temperature at which the washes are performed. In addition, interpreting the amount of hybridization can be affected, for example, by the specific activity of the labeled oligonucleotide probe, by the number of probe-binding sites on the template nucleic acid to which the probe has hybridized, and by the amount of exposure of an autoradiograph or other detection medium.

It will be readily appreciated by those of ordinary skill in the art that although any number of hybridization and washing conditions can be used to examine hybridization of a probe nucleic acid molecule to immobilized target nucleic acids, it is more important to examine hybridization of a probe to target nucleic acids under identical hybridization, washing, and exposure conditions. Preferably, the target nucleic acids are on the same membrane.

A nucleic acid molecule is deemed to hybridize to a nucleic acid of the invention but not to another nucleic acid if hybridization to a nucleic acid of the invention is at least

5-fold (e.g., at least 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 20-fold, 50-fold, or 100-fold) greater than hybridization to another nucleic acid. The amount of hybridization can be quantitated directly on a membrane or from an autoradiograph using, for example, a PhosphorImager or a Densitometer (Molecular Dynamics, Sunnyvale, CA).

Detection of an amplification product or a hybridization complex is usually accomplished using detectable labels. The term "labeled" with regard to an agent (e.g., an oligonucleotide or a polypeptide) is intended to encompass direct labeling of the agent by coupling (i.e., physically linking) a detectable substance to the agent, as well as indirect labeling of the agent by reactivity with another reagent that is directly labeled with a detectable substance. Detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials.

Polypeptides

One aspect of the invention pertains to purified polypeptides involved in cryptophycin synthesis as well as polypeptide fragments, particularly those that possess enzymatic activity (i.e., functional fragments). Predicted amino acid sequences of polypeptides involved in cryptophycin synthesis are shown in SEQ ID NOs:3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

The term "purified" polypeptide as used herein refers to a polypeptide that has been separated or purified from cellular components that naturally accompany it. Typically, the polypeptide is considered "purified" when it is at least 70% (e.g., at least 75%, 80%, 85%, 90%, 95%, or 99%) by dry weight, free from the proteins and naturally occurring molecules with which it is naturally associated. Since a polypeptide that is chemically synthesized is, by nature, separated from the components that naturally accompany it, a synthetic polypeptide is "purified."

Polypeptides involved in cryptophycin synthesis can be purified from natural sources (e.g., a biological sample) by known methods such as DEAE ion exchange, gel filtration, and hydroxyapatite chromatography. A purified polypeptide also can be obtained, for example, by expressing a nucleic acid molecule of the invention in an expression vector. In addition, a purified polypeptide can be obtained by chemical synthesis. The extent of purity of a polypeptide can be measured using any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

In addition to the naturally-occurring polypeptides involved in cryptophycin biosynthesis, the skilled artisan will further appreciate that changes can be introduced into a nucleic acid molecule (e.g., those having the sequence shown in SEQ ID NOs: 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 234, 26, 28, and 30) as discussed herein, thereby leading to changes in the amino acid sequence of the encoded polypeptide. For example, changes can be introduced into nucleic acid coding sequences leading to conservative and/or non-conservative amino acid substitutions at one or more amino acid residues. A "conservative amino acid substitution" is one in which one amino acid residue is replaced with a different amino acid residue having a similar side chain. Similarity between amino acid residues has been assessed in the art. For example, Dayhoff et al. (1978, in *Atlas of Protein Sequence and Structure*, 5(Suppl. 3):345-352) provides frequency tables for amino acid substitutions that can be employed as a measure of amino acid similarity. A non-conservative substitution is one in which an amino acid residue is replaced with an amino acid residue that does not have a similar side chain.

The invention also provides for chimeric or fusion polypeptides. As used herein, a "chimeric" or "fusion" polypeptide includes a polypeptide involved in cryptophycin synthesis operatively linked to a heterologous polypeptide. A heterologous polypeptide can be at either the N-terminus or C-terminus of a polypeptide involved in cryptophycin synthesis. Within a chimeric or fusion polypeptide, the term "operatively linked" is intended to indicate that the two polypeptides are encoded in-frame relative to one another. In a fusion polypeptide, the heterologous polypeptide generally has a desired property such as the ability to purify the fusion polypeptide (e.g., by affinity purification). A chimeric or fusion polypeptide of the invention can be produced by standard recombinant DNA techniques, and can use commercially available vectors.

A polypeptide commonly used in a fusion polypeptide for purification is glutathione S-transferase (GST), although numerous other polypeptides are available and can be used. In addition, a proteolytic cleavage site can be introduced at the junction between a polypeptide and a heterologous polypeptide to enable separation of the two polypeptides subsequent to purification of the fusion polypeptide. Enzymes that cleave such proteolytic sites include Factor Xa, thrombin, or enterokinase. Representative expression vectors encoding a heterologous polypeptide that can be used in affinity purification of a polypeptide involved in cryptophycin synthesis include pGEX (Pharmacia Biotech Inc; Smith & Johnson, 1988, *Gene*, 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ).

Antibodies can be used to detect the presence or absence of polypeptides involved in cryptophycin synthesis. Techniques for detecting polypeptides using antibodies include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. An antibody can be polyclonal or monoclonal, and usually is detectably labeled. An antibody having specific binding affinity for a polypeptide involved in cryptophycin synthesis can be generated using methods well known in the art. The antibody can be attached to a solid support such as a microtiter plate using methods known in the art (see, for example, Leahy et al., 1992, *BioTechniques*, 13:738-743). In the presence of a polypeptide involved in cryptophycin synthesis, an antibody-polypeptide complex is formed.

Detection of a polypeptide-antibody complex is usually accomplished by detectably labeling the antibody. The term "labeled" with regard to an antibody is intended to encompass direct labeling of the antibody by coupling (i.e., physically linking) a detectable substance to the antibody, as well as indirect labeling of the antibody by reactivity with another reagent that is directly labeled with a detectable substance. Detectable substances are described above.

Biosynthesis of Cryptophycin

Figure 3 shows the modular structure of cryptophycins. Cryptophycin biosynthesis is a result of a mixed Type I PKS/NRPS system.

Unit A is a polyketide synthetase derived unit. Incorporation and linkage of unnatural amino acids such as chlorinated methoxy D-tyrosine amino acid (Unit B) and β-methyl β-alanine (Unit C) are consistent with activities of non-ribosomal peptide synthetase domains. The final terminating unit, the rare carboxylic acid of leucine, leucic acid, could be the result of a NRPS system. However, the ester linkage between Unit C and D is not consistent with a peptide bond forming condensation domain of such a system. It is possible that incorporation of this ester occurs by a novel domain as part of a larger NRPS system. Alternatively, incorporation of the ester may be directed by an enzyme that previously has not been described. The generation of the macrocycle to form the core cryptophycin chemical skeleton involves a chain-terminating cyclization step, likely completed by a member of the hydrolase superfamily of enzymes or domains. The lactone formed between Unit A (the hydroxyl group) and Unit D points to a classic thioesterase dependent mechanism. Additional enzymes such as a cytochrome p450-dependent hydroxylase (likely a cryptophycin epoxidase), a non-heme dependent halogenase, o-methyltransferase, and enzymes involved in activation and methylation of the β-carbon of 3-amino propanoic acid are involved in Unit A, B, C, or D synthesis or in final structural components of cryptophycin. Many of these types of enzymes have been previously described from other polyketide and nonribosomal peptide synthetases. For an overview of the predicted pathway of cryptophycin biosynthesis, see Figure 8. See also, Figure 2.

Polyketide synthetase

Based on homology searches of the GenBank database, the nucleotide sequence designated crpA (SEQ ID NO:2) appears to encode a PKS (CrpA; SEQ ID NO:3). Sequence analysis indicated that CrpA contains a PKS domain (positioned at approximately nucleotides 1-450 of SEQ ID NO:2), an acyltransferase domain (positioned at approximately nucleotides 1-220 of SEQ ID NO:2), a dehydrogenase domain (positioned at approximately nucleotides 760-1000 or 860-1000 of SEQ ID NO:2), a ketoreductase domain (positioned at approximately nucleotides 850-1000 of SEQ ID NO:2), and an acyl carrier protein domain.

Polyketides are diverse biologically active molecules with a wide variety of structures. Polyketides are synthesized from 2-carbon units through a series of condensations and subsequent modifications, and occur in many types of organisms including fungi and mycelial bacteria. Polyketide synthetases (PKSs) catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA.

The sequencing of several genes encoding enzymes that produce type 1 modular PKSs has revealed a linear organization of modules, each of which contains the activities needed for one cycle of polyketide chain elongation. The minimal module contains a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) that together catalyze a 2-carbon extension of the chain similar to the condensation of 2-carbon units in the biosynthesis of fatty acids. In PKS polypeptides, the regions that encode enzymatic activities are separated by linker regions, also called scaffold regions. These scaffold regions encode amino acid sequences that space the enzymatic activities at the appropriate distances and in the correct order.

PKS is likely responsible for synthesis of the Unit A region, which is one of the most challenging aspects in the chemical synthesis of cryptophycins. The Unit A portion of the molecule is a dioxadiazacyclo,hexadecenetetrone moiety and represents the beginning polyketide unit (Figure 3).

Non-Ribosomal Peptide Synthetase

Based on homology searches of the GenBank database, the nucleotide sequence designated crpB (SEQ ID NO:4) appears to encode a non-ribosomal peptide synthetase (NRPS) (CrpB; SEQ ID NO:5) involved in production of the Unit B peptide portion of cryptophycin. Sequence analysis indicated that CrpB may contain one or more NRPS domains (positioned at approximately nucleotides 300-950 and 1290-1425 of SEQ ID NO:4), one or more condensation domains (positioned at approximately nucleotides 50-350 and 1475-1780 of SEQ ID NO:4), an adenylation domain, an o-methyltransferase domain (positioned at approximately nucleotides 1000-1200 of SEQ ID NO:4), one or more peptidyl carrier protein domains, an epimerase domain, and one or more acyl CoA synthetase (positioned at approximately nucleotides 525-1000 of SEQ ID NO:4).

Based on homology searches of the GenBank database, the nucleotide sequence designated crpC (SEQ ID NO:6) appears to encode a NRPS (Crp C; SEQ ID NO:7) involved in production of the Units C and D peptide portions of cryptophycin. CrpC also apparently generates a 16-membered peptolide ring during cryptophycin biosynthesis. Sequence analysis indicated that CrpC contains one or more NRPS domains (positioned at approximately nucleotides 250-975, 1350-1600, 1850-2300, and 2950-3100 of SEQ ID NO:6), one or more condensation domains (positioned at approximately nucleotides 1-

300 and 1150-1450 of SEQ ID NO:6), an adenylation domain, one or more peptidyl carrier protein domains, one or more acyl CoA ligase domains (positioned at approximately nucleotides 500-1000 and 1900-2400 of SEQ ID NO:6), one or more acyl CoA synthetase domains (positioned at approximately nucleotides 475-1000 and 1900-2400 of SEQ ID NO:6), and a thioesterase domain.

NRPSs are modular in nature, where a module is usually defined as a segment of the NRPS necessary to catalyze the activation of a specific amino acid and result in the incorporation of that amino acid into a non-ribosomal peptide. A minimal module typically contains three domains: (1) an adenylation domain (about 60 kDa) responsible for selecting and activating an amino acid and transferring the aminoacyl adenylate to a peptidyl carrying center; (2) a thiolation domain, also referred to as a peptidyl carrier protein (8-10 kDa), containing a serine residue that is post-translationally modified with a 4-phosphopantetheine group (Ppant) and acts as an acceptor for the aminoacyl adenylate; and (3) a condensation domain (50-60 kDa), which catalyzes peptide bond-forming chain-translocating steps between an upstream peptidyl-s-Ppant and the downstream aminoacyl-Ppant of the adjacent module. This minimal module for chain extension is typically repeated within a NRPS. A co-linear relationship exists between the number of modules present and the number of amino acids in the final product, with the order of the modules in the synthetase determining the order of the amino acids in the peptide.

Thioesterase Domain

Based on homology searches of the GenBank database, a thioesterase domain is positioned at approximately nucleotide 9,199 to nucleotide 10,032 of CrpC (SEQ ID NO:6).

The cryptophycin thioesterase is likely responsible for the cyclization and release of the cryptophycins from the phosphopantethienyl group of the C-terminal phosphopantetheinyl carrier protein (PCP) of a NRPS. The synthetic methods used for ring closure of cryptophycin thus far limit the scope and ease of derivatization of cyptophycins.

The utility of thioesterase domains as semi-synthetic tools for cylization of synthetic molecules has been demonstrated for gramicidin, epothilone C, and tyrocidine semi-synthesis. See, for example, Wu et al., 2003, *Org. Lett.*, 5:1749; Kohli et al., 2003, *J. Am. Chem. Soc.*, 125:7160; Kohli et al., 2002, *Nature*, 418:658; and Boddy et al., 2003, *J. Am. Chem. Soc.*, 125:3428. Use of the cryptophycin thioesterase for semi-synthesis of cryptophycin provides a new route to synthesis of cyptophycin and its analogues that allows for rapid generation in diversity throughout the entire cyptophycin molecule. Use of a thioesterase domain of the invention to cyclize a cryptophycin chain elongation intermediate (e.g., a *seco*-SNAC- cryptophycin thioester) provides an approach for generating novel cryptophycins.

Cytochrome p450

Based on homology searches of the GenBank database, the crpF nucleic acid sequence (SEQ ID NO:8) appears to encode a cytochrome p450 (CrpF; SEQ ID NO:9), which is likely an epoxidase involved in cryptophycin biosynthesis.

A survey of the structure-activity relationship of cryptophycins has demonstrated the necessity of the epoxide for high-level tubulin depolymerization and anti-proliferative activities toward tumor cells. Opening of the epoxide, however, is one of the major problems encountered in clinical uses of cryptophycins. A new generation of compounds has been synthesized containing a chlorohydrin. Chlorohydrin analogs are generated from cryptophycins containing an epoxide, and act as pro-drugs. Once chlorohydrins are injected into the serum, the compounds are rapidly converted back to the corresponding epoxides.

High-level tubulin depolymerization and anti-proliferative activities toward tumor cells also requires proper stereochemistry of the epoxide group (β epoxide). Synthesis of cryptophycins containing an epoxide often results in a mixture of two diastereomers. One of the diastereomers is usually inactive, thereby requiring reverse-phase HPLC to separate the two compounds. See, for example, Figure 2. In addition to the extra expense and time required for separation, separation of the diastereomers results in a significant loss of starting material.

Using a recombinant cell line expressing an epoxidase or a purified form of an epoxidase could dramatically increase overall yields, eliminate a separation step (e.g., HPLC), and allow more flexibility in synthetic strategies. Since no known natural cryptophycin contains the α-epoxide, the native epoxidase enzyme seems to be highly efficient at generating the desired epoxide diastereomer (see Figure 9). Further, the cryptophycin epoxidase apparently exhibits a high degree of flexibility since it is able to use various substrates (e.g., those having different ring sizes).

Additional Enzymes

Additional enzymes having a variety of functions are involved in cryptophycin biosynthesis. In addition to the PKS, the NRPS, and the epoxidase discussed above, sequence analysis indicated that the following types of enzymes are likely involved in cryptophycin biosynthesis.

Based on homology searches of the GenBank database, crpG (SEQ ID NO:10) appears to encode an iron-dependent non-heme hydroxylase (CrpG; SEQ ID NO:11), which is a member of the γ-butyrobetaine hydroxylase group. Non-heme iron-dependent enzymes generally catalyze a wide variety of O_2 reactions. An iron-dependent non-heme hydroxylase is likely involved in hydroxylation of cryptophycins.

Based on homology searches of the GenBank database, crpH (SEQ ID NO:12) appears to encode an aspartate decarboxylase (CrpH; SEQ ID NO:13). An aspartate decarboxylase (EC 4.1.1.11) is likely involved in production of β-alanine or methyl-β-alanine, which is a precursor for NRPS. See, for example, Williamson & Brown, 1979, *J. Biol. Chem.*, 254:8074-82; and Ramjee et al., 1997, *Biochem. J.*, 323:661-9.

Based on homology searches of the GenBank database, crpI appears to be the remnants of an IS1327 transposition event. The sequences identified as having homology to IS1327 are positioned at approximately nucleotides 9154-8514 of SEQ ID NO:1 (pDAM163). No open reading frame or coding sequences, however, were identified.

Based on homology searches of the GenBank database, crpJ (SEQ ID NO:14) appears to encode a non-heme-dependent, flavin-dependent halogenase (CrpJ; SEQ ID NO:15). See, for example, van Pee & Unversucht, 2003, *Chemosphere*, 52:299-312; and Littlechild, 1999, *Curr. Opin. Chem. Biol.*, 3:28-34. A halogenase is likely involved in chlorination of the Unit B amino acid, o-methyl tyrosine.

Based on homology searches of the GenBank database, crpK appears to be the remnants of an IS892-orf2 transposition event. The sequences identified as having homology to IS892-orf2 are positioned at approximately nucleotides 4730-7039 of SEQ ID NO:1 (pDAM163). No open reading frame or coding sequences, however, were identified.

Based on homology searches of the GenBank database, crpL appears to be the remnants of an IS892-orf1 transposition event. The sequences identified as having homology to IS892-orf2 are positioned at approximately nucleotides 4730-7039 of SEQ ID NO:1 (pDAM163). No open reading frame or coding sequences, however, were identified.

Based on homology searches of the GenBank database, crpM appears to be an ISRSO13 transposase sequence. The identified coding sequence (crpM; SEQ ID NO:16) encodes a polypeptide designated CrpM (SEQ ID NO:17) with unknown function.

Based on homology searches of the GenBank database, crpN (SEQ ID NO:18) appears to encode a non-heme-dependent, iron-dependent hydroxylase (CrpN, SEQ ID NO:19). See, for example, Solomon et al., 2003, *PNAS USA*, 100:3589-94; and Ryle et al., *PNAS USA*, 100:3790-5.

Based on homology searches of the GenBank database, crpO appears to encode a pvdE-type regulator (CrpO). The sequences identified as having homology to a pvdE-type regulator are positioned at approximately nucleotides 786-1768 of SEQ ID NO:1 (pDAM163). A pvdE-type regulator is likely involved in regulating cyptophycin biosynthesis. See, for example, Wilson et al., 2001, *J. Bacteriol.*, 183:2151-5.

Based on homology searches of the GenBank database, crpP (SEQ ID NO:20) appears to encode a thioredoxin (CrpP, SEQ ID NO:21). Thioredoxins are generally reduction/oxidation (redox)-regulatory proteins thought to have anti-apoptotic effects. Thioredoxin is likely involved in redox reactions (e.g., cytochrome p450-dependent hydroxylations) associated with cryptophycin biosynthesis.

Based on homology searches of the GenBank database, crpU (SEQ ID NO:22) appears to encode an N-acetyltransferase (EC 2.3.1.5) (CrpU, SEQ ID NO:23). N-acetyltransferases usually catalyze the transfer of acetyl groups from acetyl-CoA to arylamines.

Based on homology searches of the GenBank database, crpV (SEQ ID NO:24) appears to encode a large exoprotein involved in heme utilization (CrpV, SEQ ID NO:25). A large exoprotein involved in heme utilization may be involved in redox reactions associated with cryptophycin formation (i.e., cytochrome p450-dependent hydroxylations).

Based on homology searches of the GenBank database, crpW appears to encode a tetratricopeptide repeat (TPR) protein (CrpW). A TPR is a 34 amino acid repeated sequence motif found in a number of diverse proteins that may be involved in transcriptional repression, mitochondrial and/or peroxisomal protein transport, cell cycle regulation, protein kinase inhibition, heat shock response, and/or mediating protein-protein interactions. See, for example, Sikorski et al., 1991, *Cold Spring Harbor Symp. Quant. Biol.*, 56:663-73; and Lamb et al., 1995, *Trends Biosci.*, 20:257-9.

Based on homology searches of the GenBank database, crpX (SEQ ID NO:26) appears to encode a chorismate mutase-prephenate dehydrogenase (CrpX, SEQ ID NO:27). A chorismate mutase-prephenate dehydrogenase (EC 1.3.1.12) usually catalyzes the first two steps in the biosynthesis of tyrosine (the chorismate mutase activity) and the conversion of prephenate to p-hydroxyphenylpyruvate in the presence of NAD (the prephenate dehydrogenase activity). A chorismate mutase-prephenate dehydrogenase is likely involved in the production of shikimate-derived PKS starter units in cryptophycin biosynthesis.

Based on homology searches of the GenBank database, crpY (SEQ ID NO:28) appears to encode a 2-hydroxychromene-2-carboxylate isomerase (CrpY, SEQ ID NO:29). A 2-hydroxychromene-2-carboxylate isomerase is involved in the naphthalene catabolic pathway and catalyzes the reaction of 2-hydroxychromene-2-carboxylate into trans-o-hydroxybenzylidenepyruvate. See, for example, Eaton, 1994, *J. Bacteriol.*, 176:7757-62; and Zylstra et al., 1997, *FEMS Microbiol. Lett.*, 153:479-84. A 2-hydroxychromene-2-carboxylate isomerase is likely involved in the production of shikimate-derived PKS starter units in cryptophycin biosynthesis.

Based on homology searches of the GenBank database, crpZ (SEQ ID NO:30) appears to encode a 3-dehydroquinate synthase (CrpZ, SEQ ID NO:31). A 3-dehydroquinate synthase (EC 4.2.3.4) usually catalyzes the cyclization of 3-deoxy-D-arabino-heptulosonic acid 7-phosphate (DAHP) to dehydroquinate. A 3-dehydroquinate synthase may be involved in the production of shikimate-derived PKS starter units.

Combinatorial Techniques and Domain Swapping

It will be apparent to one of skill in the art that any number and/or combination of nucleic acid molecules of the invention (e.g., SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and/or 30) can be joined together to generate a longer nucleic acid molecule (e.g., pDAM163; shown in Figures 4 and 5 and SEQ ID NO:1). In addition, the nucleic acid molecules (SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30) can be manipulated using standard techniques to delete or inactivate activity encoding regions, insert regions from different molecules encoding corresponding activities from the same or different biosynthesis systems, or be otherwise mutated using standard procedures for obtaining genetic alterations. Mutations can be made to the native sequences using conventional techniques such as those described above.

Chemical approaches have lead to highly informative structure-activity relationships. Therefore, the regions suggested for modifications are well defined, particularly in view of the modular-type structure of the PKSs and NRPSs. In addition to approaches that provide mutated polypeptides, it is possible to manipulate entire domains or portions of domains. For example, a domain having a particular activity from one biosynthetic pathway can be exchanged or replaced with a domain having a corresponding activity from a different biosynthetic pathway. Alternatively, a domain having a particular activity from a biosynthetic pathway can be exchanged or replaced with a domain having an unrelated activity from the same or a different biosynthetic pathway.

If replacement of a particular nucleic acid region encoding a host enzyme is to be made, this replacement can be conducted *in vitro* using suitable restriction enzymes and cloning techniques or can be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement region in a donor plasmid and a receptor region in a recipient plasmid. A representative exchange system that involves plasmids that have different temperature sensitivities is described in PCT Publication No. WO 96/40968.

The various nucleic acid molecules involved in cyrptophycin biosynthesis, individually or as a cocktail of such molecules, can be cloned into one or more recombinant vectors. When more than one molecule is cloned together, such elements can be under the control of a single element for expression (e.g., a promoter) or each molecule can be under the control of an element for expression. The nucleotide sequences encoding an enzymatic subunit or a cocktail of such molecules can include flanking restriction sites to allow for the easy deletion and insertion of other molecules or regions of a molecule. In this manner, nucleotide sequences encoding hybrid or chimeric enzymes can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above such as site-directed mutagenesis and PCR.

Expression vectors containing nucleotide sequences encoding a variety of enzymatic activities can be transformed into an appropriate host cell to construct a library. In one approach, a mixture of such vectors is transformed into host cells and the resulting cells plated into individual colonies and selected for successful transformants. Each individual colony represents a colony expressing an enzyme having a particular activity and, ultimately, the ability to produce a particular product. Alternatively, expression vectors can be used individually to transform host cells, which are then assembled into a library. Methods are known for screening a library or isolates from a library for substrate-specificity and/or production of a particular product. Another strategy for preparing a variety of products is by random digestion-religation leading to chimeric domains or modules. A similar such method has been described as a "DNA shuffling method" (see Patten et al., 1997, *Curr. Op. Biotechnol.*, 8: 724-733).

As one non-limiting example, the creation of novel macrolides can be achieved through genetic manipulation of polyketide synthases. The modular nature of polyketide synthases allows for domain exchange between different polyketide synthase genes, resulting in hybrid genes that produce polyketide synthases with altered properties that, in turn, produce modified macrolide structures. Thus, it is possible to control chain length, choice of chain extender unit, degree of β-carbon oxidation level, and stereochemistry. See, for example, PCT Publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358; WO 98/27203; and WO 98/49315; U.S. Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and Fu et al., 1994, *Biochemistry*, 33:9321-9326; McDaniel et al., 1993, *Science*, 262:1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.*, 34(8):881-888.

The application of innovative combinatorial techniques to this type of genetic organization has prompted the generation of novel natural products, by adding, deleting, or exchanging domains or entire modules. See, for example, U.S. Patent Nos. 5,672,491;

5,712,146; 5,830,750; 5,843,718; 5,962,290; and 6,022,731; and Tang et al., 2000, *Science*, 287:640-2). The invention allows for combinatorial biosynthesis technology to produce a diversity of cryptophycin analogues in addition to those cryptophycin analogues produced to date.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1—Cloning and Sequencing the crp gene cluster contained within pDAM163

Primer synthesis and cosmid sequencing was preformed at the University of Minnesota Advanced Genetic Sequencing and Analysis Center-AGAC (St. Paul, MN). Degenerate PCR primers specific for conserved core motifs of peptide synthetase adenylation domains A2 and A8 (Marahiel et al., 1997, *Chem. Rev.*, 97:2651-74) were used and consisted of the following sequences: MTF2' forward primer (5'-GCNGG (ct) GG (ct) GCNTA (ct) GTNCC -3' (SEQ ID NO:38)) and MTR reverse primer (5'-CCNGG (agt) AT (tc) TTNAC (tc) TG-3' (SEQ ID NO:39)) (Neilan et al., 1999, *J. Bacteriol.*, 181:4089-97). Adenylation domain containing DNA fragments of approximately 1100 bp in length were synthesized by PCR using a Hybaid Express PCR thermocycler (30 cycles: 95°C for 1 min, 55°C for 1 min, 72°C for 1 min) with *Nostoc sp* ATCC 53789 genomic DNA as a template. End sequencing of one fragment, pNAM124, using an Applied Biosystems, Inc. ABI3700 sequencer (Foster City, CA) confirmed that the fragment contained an adenylation domain. Prediction of its substrate specificity (aromatic amino acid activating) was determined using methods described previously (Challis et al., 2000, *Chem. Biol.*, 7:211-24). The fragment was radiolabeled using the RadPrime labeling kit (Pharmacia) with [α-^{32}P] dCTP (Amersham) according to the manufacter's directions. The radiolabeled fragment was used to probe the genomic library using standard colony hybridization protocols (Sambrook & Russell, 2000, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press). One cosmid, pDAM163, was selected because it hybridized to the adenylation domain encoding DNA probe contained within pNAM124. The DNA sequence of pDAM163 was obtained by creating a shotgun library of the cosmid within the sequencing vector, pUC18. Sequences obtained were assembled using SeqMan version 5.06 (DNAStar, Madison, WI) and Frameplot 2.3.2 (Ishikawa & Hotta, 1999, *FEMS Microbiol. Lett.*, 174:251-3) used to identify individual open reading frames. The putative functions of the crp biosynthesis genes were assessed by using the open reading frames and their putative protein products versus genes/proteins contained within the GenBank database using BlastN and BlastP.

Example 2—Cloning genes involved in cryptophycin biosynthesis

DNA encoding a putative cryptophycin biosynthetic gene cluster was contained on a cosmid designated pDAM163. pDAM163 DNA was prepared using a Qiagen large construct DNA extraction kit from a 500 mL culture grown overnight at 25°C in LB media containing 50 µg/mL ampicillin.

Example 3—Cryptophycin production

The cosmid, pDAM163, or sub-vectors such as cosmid, plasmids, yeast artifical chromosomes, bacterial artificial chromosomes, or phage vectors containing pDAM163 sequences can be used to biosynthetically prepare cryptophycins in a non-*Nostoc* spp. host. pDAM163 is introduced into an *Escherichia coli* strain that harbors a phosphopantetheinyl transferase gene required for expressing active polyketide synthase and nonribosomal peptide synthetase enzymes. Fermentation of the resulting strain on a large scale, and extracting and detecting cryptophycins are performed as decribed previously (Subbaraju et al., 1997, *J. Nat. Prod.*, 60:302-5; and Golakoti et al., 1994, *J. Am. Chem. Soc.*, 116:4729-37).

Example 4— Cloning strategy of the thioesterase domain

DNA encoding the cryptophycin thioesterase domain is contained at the 3'-end of the 3'-terminal open reading frame of CrpC, which also codes for domains necessary for incorporation of units C and D of cryptophycin. Therefore, truncation of the DNA in the final ORF was necessary in order to isolate the cryptophycin thioesterase. Identification of the DNA encoding the cryptophycin thioesterase was elucidated through use of the NCBI "CDART" program for identification of conserved domains. The "nnpredict" secondary structure prediction program (Kneller et al., 1990, *J. Mol. Biol.*, 214:171) was used to determine the putative secondary structure of the gene product of the putative thioesterase domain and a domain capable of being phosphopantetheinlyated. The forward primer, 5'-ATT TAT CAT ATG GGT TCC GAT TCC GGA GCC GA-3' (SEQ ID NO:40), was designed to a position immediately 3' of a nucleic acid sequence predicted to encode a protein capable of being phosphopantethionylated in a region appearing to lack secondary structure based on the "nnpredict" program results and contained an *Nde*I restriction site. The reverse primer, 5'-AAA TAA GAA TCC TCA TCA TTT TTC CAA TTG ATG GGT-3' (SEQ ID NO:41), was constructed to anneal to the 3' end of the open reading frame and contained a *Bam*HI restriction site.

PCR reactions were performed with 0.1 µL of pDAM163 DNA from the extraction, 1 µM forward primer, 1 µM reverse primer, 1X ExTaq buffer (Takara), 1 µL ExTaq polymerase (Takara), and 1 µM dNTP (Takara) to a final volume of 50 µL with water. The PCR program consisted of 30 cycles of the following amplification conditions: denaturation 1 min at 95°C, 1 min annealing at 50°C, 1.5 min extension at 72°C. PCR fragments corresponding to the desired length were separated on a 1% agarose gel and purified from the gel using a Qiagen gel extraction kit. The PCR fragment was cloned into a pGEM T-Easy vector (Promega) using T-overhang cloning with the pGEM T-Easy kit (Promega).

Clones were transformed into XL-1 Blue competent cells using heat shock protocols as described in the pGEM T-Easy kit. Constructs containing inserts were identified using blue/white screening according to the pGEM T-Easy kit protocol. Five clones containing insert were re-plated and half of the colony was subjected to PCR to verify insert of the desired DNA size using the same PCR condition listed above, with the exception of a 5 min incubation of each clone at 96°C prior to the amplification cycles.

One clone containing the desired size insert was grown in a 2 mL culture overnight in LB media containing ampicillin (50 µg/mL; Research Products International Corp). DNA was purified using a Qiagen mini-prep kit. DNA was submitted for sequencing to the University of Michigan DNA Sequencing Core Lab and sequenced 3 times from the 5' end using the T7 primer binding site and 3 times from the 3' end using the SP6 primer-binding site. DNA from the sequenced clone was ligated into the *Nde*I and *Bam*HI sites in pET28b (Novagen) and transformed into BL21 competent cells using electroporation. All cells were plated on LB plates containing kanamycin (50 μg/mL; Research Products International Corp) and incubated overnight at 37°C. Ten colonies were subjected to PCR verification of the desired DNA insert using the primers and protocols listed above.

Example 5— Expression and purification of the cryptophycin thioesterase domain

A clone containing the desired insert size, as visualized by agarose gel electrophoresis, was grown overnight in 25 mL of 2YT broth (16 g tryptone, 10 g yeast extract, 10 g NaCl) containing 50 μg/mL kanamycin at 37°C. 5 mL of the overnight culture were used to inoculate 1L of 2YT media containing 50 μg/mL kanamycin, which was grown at 37°C. The culture was induced at an OD_{595} of 0.7 with 0.2 mM IPTG and grown overnight at 30°C. Cells were harvested at 5000 *g* for 30 min. The pellet was resuspended in 20 mL 0.1 M sodium phosphate buffer (pH 8) containing 20 mM imidazole and 300 mM NaCl. 4 mg of lysozyme and 2 g sucrose were added to the cell suspension and incubated at room temperature for 30 min until the viscosity of the solution increased. The solution was put on ice and subjected to sonication (5 times for 20 sec) at a level of 6 on the sonicator until the solution became less viscous. The suspension was centrifuged at 17,000 *g* for 1 hour at 4°C.

The supernatant was collected and incubated with 7 mL of Qiagen Ni-Agarose overnight at 4°C. The agarose was then loaded into a column and washed with 10 column volumes of 0.1 M sodium phosphate buffer (pH 8) containing 20 mM imidazole and 300 mM NaCl. The column was washed with 10-column volumes wash buffer containing 50 mM imidazole. Protein was eluted with wash buffer containing 100 mM imidazole. The eluted sample contained ~50 mg of protein as determined using a BioRad Bradford assay kit. Samples were run on a 4-20% SDS-PAGE gel to check for purity. A band corresponding the expected molecular weight was observed at >95% purity. Protein was subjected to a PD-10 column prior to kinetic assays for buffer exchange to 100 mM sodium phosphate buffer (pH 8).

Example 6—Preparation of substrates

Referring to Figure 10, substrate 3 represents the tri-depsipeptide sector of cryptophycin except that the methyl β-alanine residue has been replaced by β-alanine. The remaining functionality has been preserved. The halogenation of the tyrosine residue likely is a tailoring modification, which is performed after thioesterase-mediated cyclization. Therefore, a simple tyrosine methyl ether was employed. The SNAC thioester substrate 3 was prepared from known tri-depsipeptide 1 (Georg et al., *J. Org. Chem.*, 2000, 65:7792-7799) by PyBOP coupling of *N*-acetylcysteamine followed by Boc deprotection with 4 N HCl in 1,4-dioxane to provide 3 as the hydrochloride salt (Figure 10).

Similarly, the Unit A analogs 6 and 9 were prepared by stepwise deprotection of the *t*-butyl ester with TFA containing 1% triethylsilane followed by TBS cleavage with 5% hydrofluoric acid in acetonitrile from known Unit A fragments 4 and 7 (Georg et al., *supra*). PyBOP mediated coupling of subunit 3 with fragments 6 and 9 afforded the *seco*-SNAC-cryptophycin thioester substrates 10 and 11 respectively, which were purified by reverse-phase semi-preparative HPLC (C18, Alltech Econosil 10 x 250 mm, 5 mL/min, 10-100% AcCN/H_2O + 0.1% TFA, 30 minutes).

Example 7—Kinetic characterization of cryptophycin thioesterase activity with a substrate

A standard curve of the cleaved product was determined on a 10-67% acetonitrile/water (0.1% TFA) gradient over 30 min. Cleavage reactions were run for 15 min at 30°C with 1.4 μM cryptophycin thioesterase with substrate concentrations of 0.3125, 0.625, 1.25, 2.5, and 5 mM substrate containing 4% DMSO in 0.1 M NaH_2PO_4 buffer at pH 7, 8, and 8.75. The hydrolyzed version of substrate 3 was monitored in order to determine the rate of hydrolysis for the reactions. All reactions were run in triplicate.

Example 8—Cyclization of cryptophycin substrates

A 1 mL solution containing 100 μM substrate 10 or substrate 11, with 7 μM cryptophycin thioesterase, 0.095 M NaH_2PO_4 buffer (pH 7), and 5% DMSO was incubated for 1 hour at 30°C. Negative control reactions containing all reagents except for the cryptophycin thioesterase were run in parallel. The total contents of each reaction were separated using reverse phase chromatography with a 10-100% gradient (acetonitrile + 0.1% TFA/water + 0.1% TFA) over 37 min on an Alltech Econosil 10 U C18 column with dimensions 250 mm x 4.6 mm. The products were analyzed by electrospray mass spectrometry (ES+). The relative concentration of the products was determined by comparing absorption at 245 nM, which corresponds to the enone functionality contained within each molecule examined.

Example 9—Results

Immediately 5' of the nucleotide sequences encoding the cryptophycin thioesterase are sequences that putatively encode a phosphopantetheinylation domain. The thioesterase domain was, therefore, constructed to begin immediately following the 3' end of DNA predicted to encode the phosphopantetheinylation domain.

The molecular weight of cryptophycin TE was determined to be 35,424 Da by ES+ mass spectrometry and 35,410 by MALDI-TOF mass spectrometry. The calculated average mass for the cryptophycin TE was 35,550.08, and the monoisoptopic mass was determined to be 35527.66. The mass spectrometry determined that the molecular weight of cryptophycin thioesterase corresponds to a thioesterase that is missing its N-terminal methionine. Processing of the N-terminal methionine commonly occurs when proteins are expressed small amino acids adjacent to the N-terminal methionine, such as the glycine that is located adjacent to the N-terminal methionine in the engineered construct.

The cyclized cryptophycins are fairly insoluble in water and, therefore, kinetic characterization of hydrolytic rate of the cryptophycin thioesterase was determined using a substrate modeled after the depsipeptide fragment corresponding to Units B, C and D of Cryptophycin 1.

Characterization of the cryptophycin thioesterase-catalyzed hydrolysis of the substrate 3 was monitored by HPLC. The two hydrolysis products produced by the reaction were determined using ES+ mass spectrometry to be N-acetyl cystamine and molecule 12 (Figure 11).

Initially, the cryptophycin thioesterase was stored in 5% glycerol containing buffer. However, analysis by HPLC/MS of hydrolysis of the substrate 3 with cryptophycin thioesterase containing 5% glycerol revealed that the glycerol adduct was the major product of the reaction with a minor product of the hydrolyzed substrate. Therefore, the expression strain containing cryptophycin thioesterase was recultured and the cryptophycin thioesterase was purified in the absence of glycerol. Subsequent analysis of the cryptophycin thioesterase-catalyzed hydrolysis of the substrate 3 did not reveal a glycerol adduct peak. The generation of the glycerol adduct (molecule 13, Figure 11) warrants caution when determining kinetics using buffers containing glycerol (especially using indirect methods).

The hydrolytic activity of cryptophycin thioesterase was determined for the substrate 3 using steady state kinetic analysis utilizing HPLC analytical methods. Figure 12 outlines the catalytic rate constants for hydrolysis of the substrate 3 with cryptophycin thioesterase at pH 7, pH 8, and pH 8.75.

The ability of the cryptophycin thioesterase to cyclize substrates was examined using *seco*-SNAC-des-epoxy-arenastatin 11 and the des-benzyl derivative of *seco*-SNAC-des-epoxy-arenastatin 10 as substrates (Figure 13).

The partition ratio of cyclization to hydrolysis for the cryptophycin catalyzed reaction with *seco*-SNAC-des-epoxy arenastatin 11 was 5:1, while the partition ratio of cyclization to hydrolysis with the *seco*-SNAC-des-benzyl-des-epoxy-arenastatin 10 was 1:8.3, as determined by HPLC/MS with quantitation of the quantity of enone functionality at 245 nM. Therefore, the cryptophycin thioesterase preferentially cyclized the SNAC thioester of *seco*-des-epoxy-arenastatin over the SNAC thioester of *seco*-des-benzyl-des-epoxy-arenastatin.

The specificity constant (k_{cat}/K_M) for the cryptophycin thioesterase catalyzed hydrolysis of the substrate 3 increased over the pH range from 7 to 8.75 (Figure 12). The increase in the specificity constant was due to an increase in the k_{cat} from pH 7 to pH 8, and a decrease in k_{cat} from pH 8 to pH 8.75. The K_M for the hydrolysis of the substrate 3 decreased slightly from pH 8 to pH 8.75, although the k_{cat} for hydrolysis also decreased, resulting in an overall increase in the specificity constant.

Interestingly, although substrate 3 contains both a thioester bond and an ester bond, hydrolysis occured specifically at the thioester, even after complete hydrolysis of the thioester, indicating a selective preference for that site.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% sequence identity to the sequence shown in SEQ ID NO:1 or to a fragment thereof, wherein said sequence encodes at least one enzyme involved in biosynthesizing cryptophycin.

2. The nucleic acid molecule of claim 1, wherein said nucleic acid sequence has at least 90% sequence identity to the sequence shown in SEQ ID NO:1 or a fragment thereof.

3. The nucleic acid molecule of claim 1, wherein said nucleic acid sequence has at least 95% sequence identity to the sequence shown in SEQ ID NO:1 or a fragment thereof.

4. The nucleic acid molecule of claim 1, wherein said nucleic acid sequence has at least 99% sequence identity to the sequence shown in SEQ ID NO:1 or a fragment thereof.

5. The nucleic acid molecule of claim 1, wherein said nucleic acid sequence is SEQ ID NO:1 (pDAM163).

6. A vector comprising the nucleic acid molecule of claim 1.

7. A host cell comprising the vector of claim 6.

8. A method of producing cryptophycin, comprising the step of:

culturing the host cell of claim 7 in the presence of an appropriate substrate and under conditions appropriate for the production of cryptophycin.

9. The method of claim 8, further comprising the step of:

purifying said cryptophycin.

10. An isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30, or to a fragment thereof, wherein said nucleic acid sequence encodes a polypeptide that exhibits functional activity.

11. The nucleic acid molecule of claim 10, wherein said nucleic acid sequence has at least 90% sequence identity to a sequence selected from said group or fragment thereof.

12. The nucleic acid molecule of claim 10, wherein said nucleic acid sequence has at least 95% sequence identity to a sequence selected from said group or fragment thereof.

13. The nucleic acid molecule of claim 10, wherein said nucleic acid sequence has at least 99% sequence identity to a sequence selected from said group or fragment thereof.

14. The nucleic acid molecule of claim 10, wherein said nucleic acid sequence is selected from said group or fragment thereof.

15. A polypeptide encoded by the nucleic acid sequence of claim 10.

16. The polypeptide of claim 15, wherein said polypeptide has a sequence selected from the group consisting of SEQ ID NOs: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 31.

17. A vector comprising the nucleic acid molecule of claim 10.

18. A host cell containing the vector of claim 17.

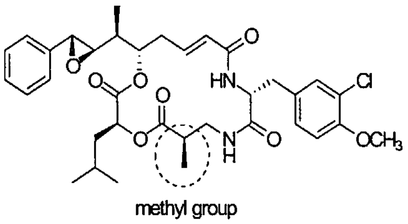
19. An isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% sequence identity to nucleotides 9,199 to 10,032 of SEQ ID NO:6, or to a fragment thereof, wherein said nucleic acid sequence encodes a polypeptide that exhibits thioesterase activity under appropriate conditions.

20. An isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% sequence identity to the sequence shown in SEQ ID NO:8, or to a fragment thereof, wherein said nucleic acid sequence encodes a polypeptide that exhibits epoxidase activity under appropriate conditions.

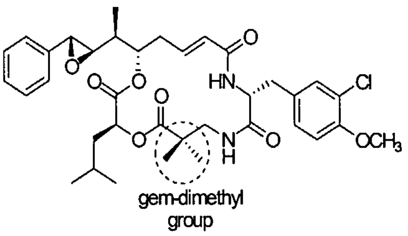
21. An isolated nucleic acid molecule comprising a nucleic acid sequence having at least 85% sequence identity to the sequence shown in SEQ ID NO:14, or to a fragment thereof, wherein said nucleic acid sequence encodes a polypeptide that exhibits halogenase activity under appropriate conditions.

22. The isolated nucleic acid of claim 19, 20, or 21, wherein said appropriate conditions are pH, media, temperature, and/or the presence or absence of co-factors.

FIG. 1A



Cryptophycin 1
$C_{35}H_{43}ClN_2O_8$
Exact Mass: 654.27
Mol. Wt.: 655.18
Predominant compound produced by
Nostoc sp ATCC 53789



Cryptophycin 52
$C_{36}H_{45}ClN_2O_8$
Exact Mass: 668.29
Mol. Wt.: 669.20
Synthetic compound chosen
for clinical evaluation

FIG. 1B

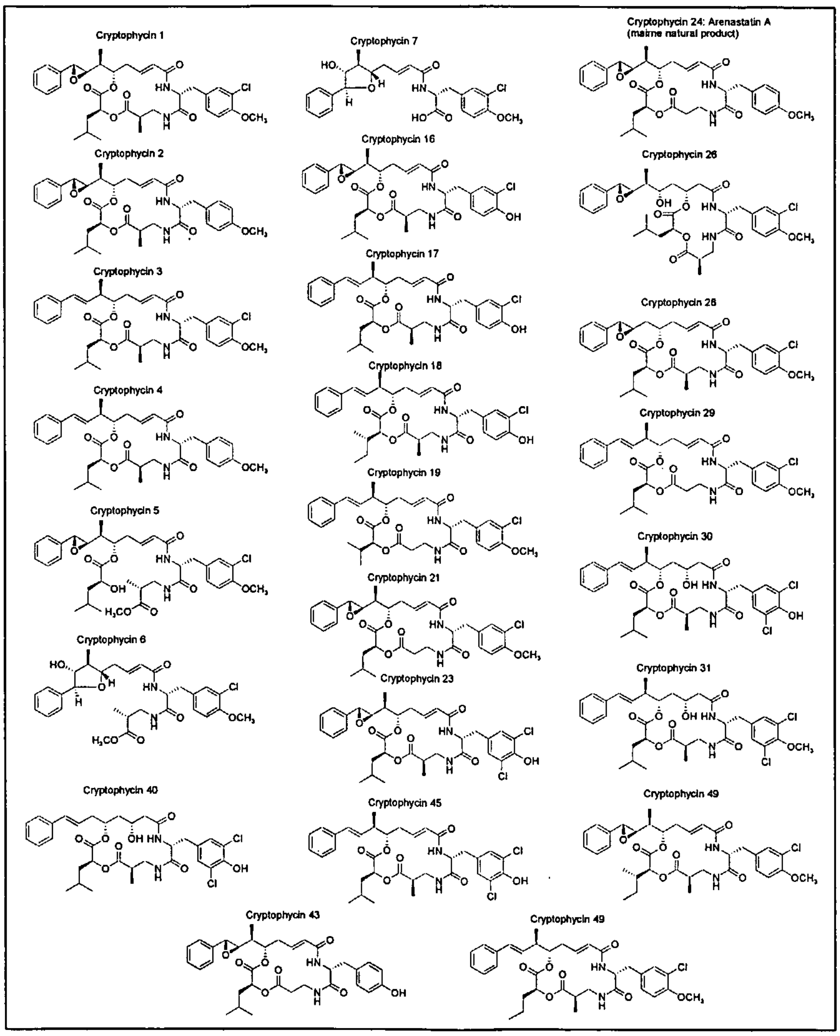

Cryptophycin 1
Cryptophycin 7
Cryptophycin 24: Arenastatin A
(marine natural product)
Cryptophycin 2
Cryptophycin 16
Cryptophycin 26
Cryptophycin 3
Cryptophycin 17
Cryptophycin 28
Cryptophycin 4
Cryptophycin 18
Cryptophycin 29
Cryptophycin 5
Cryptophycin 19
Cryptophycin 30
Cryptophycin 21
Cryptophycin 6
Cryptophycin 23
Cryptophycin 31
Cryptophycin 40
Cryptophycin 45
Cryptophycin 49
Cryptophycin 43
Cryptophycin 49

FIG. 2

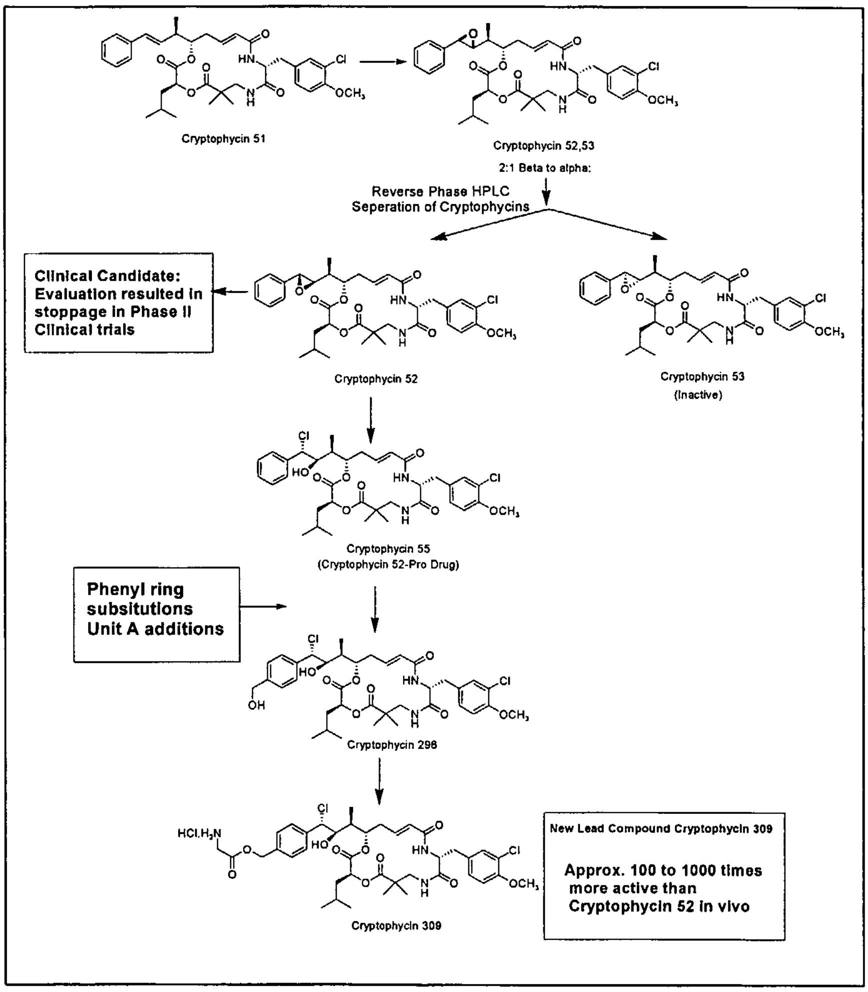

Cryptophycin 51
Cryptophycin 52,53
2:1 Beta to alpha:
Reverse Phase HPLC
Seperation of Cryptophycins
Clinical Candidate:
Evaluation resulted in
stoppage in Phase II
Clinical trials
Cryptophycin 52
Cryptophycin 53
(Inactive)
Cryptophycin 55
(Cryptophycin 52-Pro Drug)
Phenyl ring
subsitutions
Unit A additions
Cryptophycin 296
HCl.H3N
Cryptophycin 309
New Lead Compound Cryptophycin 309
Approx. 100 to 1000 times
more active than
Cryptophycin 52 in vivo
Cl
OCH3
HN
O
H
HO
OH

FIG. 3

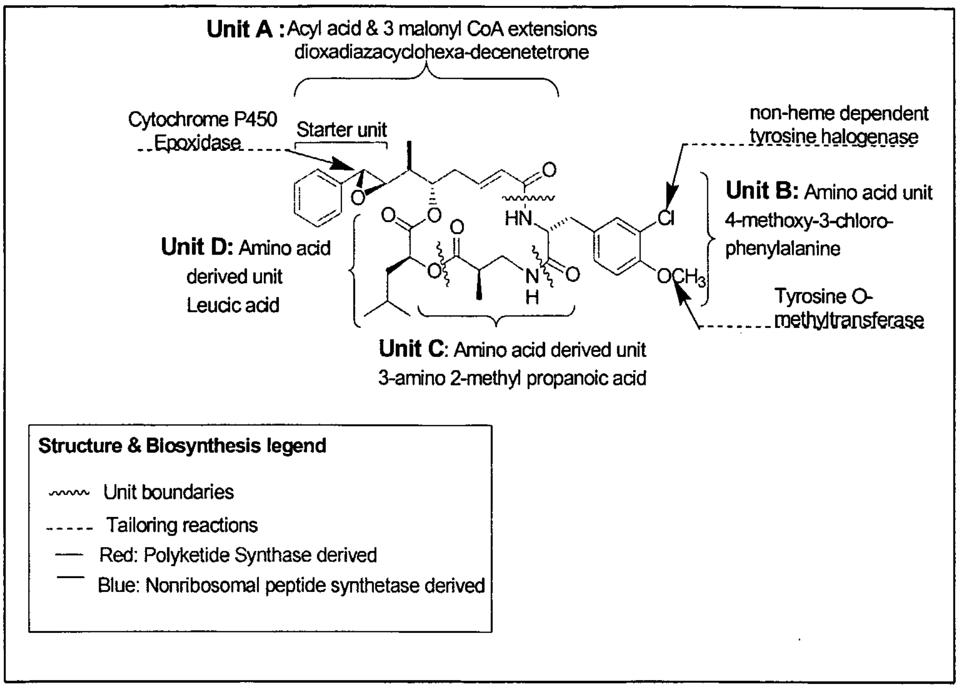
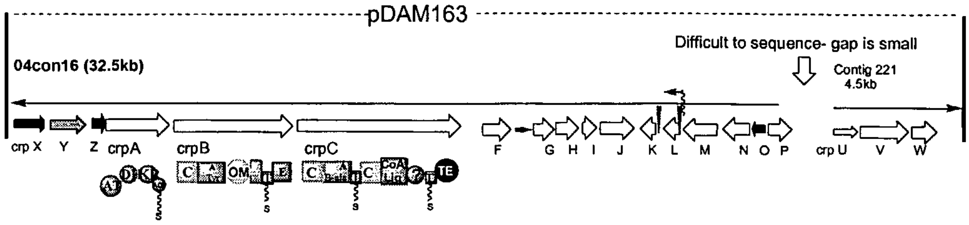

Unit A : Acyl acid & 3 malonyl CoA extensions
dioxadiazacyclohexa-decenetetrone
Cytochrome P450
Epoxidase
Starter unit
non-heme dependent
tyrosine halogenase
Unit B: Amino acid unit
4-methoxy-3-chloro-
phenylalanine
O
HN
Cl
O
O
O
O
O
N
H
O
OCH3
Unit D: Amino acid
derived unit
Leucic acid
Tyrosine O-
methyltransferase
Unit C: Amino acid derived unit
3-amino 2-methyl propanoic acid
Structure & Biosynthesis legend
Unit boundaries
Tailoring reactions
Red: Polyketide Synthase derived
Blue: Nonribosomal peptide synthetase derived

FIG. 4

Crp gene cluster



Crp gene cluster legend

crpA- type I PKS
crpB- NRPS
crpC-NRPS

crpF- cytochrome p450 hydroxylase
crpG- gamma-butyrobetaine hydroxylase
crpH-aspartate decarboxylase
crpI- IS1327 transposase
crpJ- halogenase
crpK-IS892-orf2
crpL-IS892-orf1
crpM- ISRSO13 transposase
crpN- Benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase
crpO- pvdE type regulator?
crpP-Thioredoxin

crpU- acetyltransferase
crpV- large exoprotein involved in heme utilization
crpW- TPR repeat protein (hypothetical)
crpX-chorismate mutase/prephenate dehydrogenase
crpY- 2-hydroxychromenene-2-carboxylate isomerase
crpZ- 3-dehydroquinate synthase

| -inverted repeat

↰ - E.coli consensus promoter

FIG. 5-1

PDAM163
TGATATTAGATGTTTAATATCTTAGGAAATTTTGATAAACAGGAAAGCTTATGACTGTCAGTTGCGAGAACGGTAAAT
TTTTAAACATATTTGCTCAACTTTCTCAACCACCAAATACTTTGCCTTCAGCTTAAAAAGCAAATATACTTCTTATTC
GCTTTGGATATAACGAAAGTTCATATCCCCCTTGTAGCCAGAAGCTTTCGCAAGATCCTTAAAGTTTTGCACGCCGGT
ATACTGATGTCCATTGATAACCCAAGTTGGTACACCTGGGACTTTCGCCGCATTGCACAAGTCTGGGTGGGGATTGAT
ACCTCTCTTATCGCACTCAACTTTAATACTGTCGTTGATTATTTGGTAGGCTTGCTTCCCAAAGATTAACTTTTGTTC
GTGACAGTGAGGACACCACCAAGAAACATATTCTTTTGCCCCGATATACACCAAATGCTTCGCCAAGGCTAGTTCTGC
CTTCCCTGAAGTGGTAGTGATTTCCCAACCGACTCCGGTCTGAGGTCGTTCTTTGGGCAGGAAGAAAATAATCGATGG
GGACTTTTGGTTGTCTGTTTGTTGAGCAATATCTGTCAGTAGTGCAGAGCCAGAACTCGTACCAACACAAATAATGGA
GAGCGTTAACAGTGCAAATAAAAAAGGGTGTTTGATAGCCATATTAAATATATACTTTGATTGAATAGTTGTCGAAGC
TAACAATTGTATGTTTAAGATTGTAGTATTTTGTATAACTGATAATACGAAGCCAGAKRATCCCCATCTTATCTTCGT
CATAATCGAAATTATCATCACCATGTTCTTCATACCACTTTCTTTCTTCCTCCAAAGCGATTTCTCTCTCGATGATTT
TTCTTTCTAGAATTTCACTTTCTAGAATTTCACTTTCAAGCTTTTCTCTTTCAAATCTTTCTCTTTCTATTCTTTCTC
TTCTCTGCCTTTCTTTCTCTTCTCTTTCTCTCTCAAGTTGAATTAATCGCTGCTCCACTCTCTCACAGTAAAAATCCA
AGGCTTTACCAGTGGCTGCTCCGGTAATTGCACCACCAGCCAAAAATATTATTACAGTTTTCCAGCCTTGAATGGGAT
TACCGCTCCACTCCAAACTACCCTCATTGAGCCACACTAGTAGGAAGTAAACGAGAATTCCAATAACAGTATTGATCG
GTGCAAAAACTCTCGGTGTCATCTGGGTTTGACGAAGTATCAGCCATTGCGCCAGGGTGAGAATCACACCAAAGATTG
CCGCAGCAATTCCTGAGAGAATCAGATTACCTGGGAAACTAAACGGATCTACTGTAAATCTAATGATTAGCGTGATAA
TAATCGGGGCAGAGAGCAAAAAGCTCAGAAGTAAGCTAGCGCTAATCGCAAGAAACCAGTAGCGACCGAGATAACCCA
GGCGAAATTCTAACAACGACTCTTGTACTTCCAAACAAAGTCATCCAAAATGCAATACTCGCGGTCAAATAATCAAGC
ATTAGCCGTTGTTTTACCCGTGTGTTGAAGAGGGTATTTTCTAGAATACGCCCAAACAGTATTTTGTTGTTGTGGTGG
CTCTGGTGTGGTGTTTTTTTTTTTTGATAATTGCTCTATGAGATTTTTCAATGAGTTTTATATTTGGTAGTGCTTTAAG
GCTGCAAAATATAAAAAGCAACGAAAAACCCCATCCTTCATACAGGTTGGGGGGATAGATAAAAAACATATATTCCAT
GATGACACAATTCAATATTTGTTCGACTGCTGTCAGCCTAGAGGTGGGGCAATGAACAAACCACCATCCAGACGCAAG
AAAATTACCCCTGCGACATCTGAGGAACCAAAGCTAGCAACTGACCCTGCTCAGGAAAATACTTCTTTGCACGAAAAT
CCAGGGGGAGCAACTATCACGGTGACGGCTGTTGAAGTAACAGATTTGACCCAGGAAGAACAAAGCTTACGCCTGCAT
TTAGAACACCGTGTGGAGAGAGCATTTTTGGAGGCGGGTCAAGCGTTGATGGAGTTGCGGGACAGACGGCTGTACCGT
TCCACGCACCGGACTTTTGAAGAATACTGCCGCGAACGCTTCAATTATAGTCGTGACGCGGCTTACTTGAAGATTTCG
GCTACTGTGGTTTATGAGAATCTTCAAAAGTTTTTGCCGACCATTGGTCGGCAAATTCCAATGCCGACCAACGAACGA
CAATTGCGTTTTTTGGCGAAAGCCGAGTTGGAACCGGCTGTGCAAGCGGATGTATGGCGGCAGGCAGTGGAGCAAGCT
GGCAATAAGATTCCATCCGGTCGCATAGTGAAAGATGTTGTAGATAGGATACGCGAAAGGACGAAAGTACCCAATCCT
TACCACGTTGGGGAGATATGCGTTCTTCTACCCAAAGATAATGCAGACTTGAGAGGTAAAGCGGGTTATTGGGGCGTG
GTCAGCCATGTTGGAGAATACAGTTGTACACTCCAGATATGGGACGGTGACTATACCGTAAAAATCGAACACCTGAAA
TCACTGGAATTACTTGATGAAGATTGCCAATTCATGCAGCAGTTATGTGTGAGGTTACGGCAGTTGCATCAAGTGGAC
AGGCGTGACGAGGCTGTGGATTGGCTGTTGCAGTGGTTGGGGAAACAGGCCAAACCTTATCTGTCATCCTTGCAGTCA
AAGCTGCTGGCGTTTGTTGAGAGAGAGTACAACCTGGTTTGGAAGCAGCAGAAGTGATGAGATAGCTAGTAAACAATA
GGTTAATCCAACAAATACACAATGCAACAATTAACTCATTGCATGAAAGCGGTAAGCGATCGCGGAGGGTCTGGTAGA
GTTGCCATGCTGGAAGGCTTATCGGTTCAAGAAGAAATCTGAGTAGGTCATGGGGAGTGTCCTTTTATAGCCGCCATA
ACCGGACAGTTACCATTTTTCCCTCATGACATAGCACTAAATCTTACCAGCACTTCAAATTAAAGGTAAAGCAGTGCT
AGTCATCAGTCACGATGATAAATATTTCCATTTAGCATCTCGCATTGTAAGGCTGGATTACGGACATCTTAAGTATGA
GTCATGAAAATTATGTATTCCAAACCCGACAACTTACTGCATCCACTGTACCCAATCAGGCGCAGATGTCATCAATTG
ACTAAACTTATCAGTGTAAGTATCGTCAAACTCTAGCATCACTCCCCATCGCTCATCACTCGTGAATCGGAAAATTGG
AACTGAAGCCGATGCAGAGGAACATAACCGCCACAAAGCTGAAGTAAGCGCAGCAGATGATTAGCTCTACGATCCAGC
CCTCTCAATATTGACAAATAGTACACTATGTGAGTTTTCTAAGAAGGTAAGACTAAAACTGCACTTAAGCGCTTATGT
TATCTCCCCTATTTGATGCTTTTGTAGAGGCAAGCCCCGTCAGTGTAATGATGCGAGTCCTAATGGAAAACATTTTTA
ATTCCTCGCGAATGAATCAAATATTTGATACATCAAGCGTTCGCCAATACTCTCAAGAGCTACTGTTTTCGACTCAGG
TGGATTTGATGAGTCTAGTAGTGTGTGGGATGTATCCCTCGGTTCATGCAGCCTATCAGAAGAAGGCAGTGGAGGTAA
GTGTCAGCGCCACAGCGTTATACAACAAACTGCAACGGATTGAACTGCCTGTAAGTCGGGCATTAGTGCATGAGACAG
CATCTGACCTCCAGCAGTTGCTGTTG

FIG. 5-2

ATGTTGAATGTGGAACGCCCCAGTCCTCTAGGAAAACAATATCGGTTGCGGATTGTAGATGGCAGTTGTTTAGCCGGA
ACCGAACGCAGACTAGCAGCGCTGCGCCCCCATGCAGCCAAACCATTACCCGGAAAAACAATCGCCATTCTCGACCCA
GGGACAAAACTGGTGGTTGATGTGATTCCTTGTGAAGACGGTCATTCCCAAGAACGCTCCAAGTTTCATCAGGTTTTG
GCACAAGTGCAACCCCAACAGGTATGGATTGCAGACCGTAACTTTTGTACCGCAGGATTTCTCCATACTATTGCCAAA
CTTGGAGCGTTTTTTGTGATTCGTCAACACGGGGGTTTAGGATACGAGCCTTTTGGTGAGTTACAAGCTGTTGGGTTG
TGCCAAACAGGAACTGTGTTTGAACAACAGGTGGAAATTGTCCATGAGGGAGGGACTTTTCGGTGTCGCCGTATCGTA
GTTAAGTTGACTCGTCCCACCCGTGACCAAGAGTGGGAAATTGCCATTTTTACCAACTTACCACCCACTGACGCAGAC
GGCATTCTGGTGGCACAACTCTATCAAGGGCGGTGGAGTGTGGAAACTTTATTCCAAACTGTGACCCAAAACTTTCAT
GGAGAAATTGAAACCCTAGCTTATCCTAAAGCTGCCTTATTCTCCTACTGCATGGCACTGTCAGCCTACAACCTTTTA
GCGACACTTAAAGCAGTTCTTGGCAGTGTACATGGGGTAGACAAAATCGATATTGGGCTATCCGATTTTTACCTAGTA
GATGATATCCATTCCATCTATCGGGGCATGATGATTGCTATTCCTCCGGTTCATTGGCAATTCTTTGAGGAGTTTACC
AACATTCAGATGGTAGACGTTCTCCAGCATCTAGCAACCAAAGTACATCTCAAATCTTTTCGCAAACACCCCAGAAGT
CCCAAAAAGAAACGACCACCACTCTCTGTTGATGGCAAACATTCCCACTGTTCCACTACTCGAAAGCTCAAGCAATAC
AAAGCAGCTCTTGATGCTATCCCGTGAAGCAATTTCATAAAATATGTTATTTGTCAATATTGAGAGGGCTGGCTCTAC
GATCCTAACGTGGCAAAACTTACTAGAGAAGAGTAAAAATCCTGTAATCTTGACCTTGTAGCGAAATAATGGTGCGAA
AACTTGGCATGAAATTGTCTAAAACCAGAGGCAACATCGTTTGAAGTACTCGATTGTGTTCAAAAAAAATGCCCTTCG
TGCGGTCAAGCAATGTGGAATGAATACAATAATCCTCGACATATAAGAACGTTAAACGGGGTAGTAGAACTACAACTA
AAAATTCGTCGATGTCAAAATAATTCATGTCTGCGGTACAAAAAAGCATATCGACCAGAGCAAGAAGGGTCACTCGCT
CTACCACAAAACGAATTTGGTTTGGATGTGATTTTATAAGGAGCATTACGCTACCAGGAACATAGAAGTGTTCCCCAA
ATACACGCTCACCTCGAATTAAAAGGTATATGTATAAGTCAACGAACGGTCACGCACCTAATTGACAGATATGACGAG
TTACTTTCTTTATGGCTAAAAGACCATAAAAGGTTAAAAGCAATAGTGGCTAATCAAGGACGGGTTATATTAGCGATC
GATGGAATGCAGCCAGAAATTGGACATGAGGTATTATGTATGCTTTGAATCATAATGTGAGAAAAATTTTGATCCATA
AATTAGAAAAAAGTTAACGAAAATTCAGGCTTTCGTGCTAATCAAAAATTAAAACTTTGAATCAAATTATGAGTGAGA
GTTGAAATTCTGAATCATAACTAGAGAATGAGTTGAAAAAACACAATTGGACAAAACTTGCACAAAAAATCCCTGACA
AAATTTCTCTAACTTAAACTTTCAATTCAGAATATGGTTCAAATACCAATGTTATGGTTCAAAACTTTTACACAAGCT
GTTAGGGTAGCATTACTTAGTATTGCTCATGGTTTATGTCATTATCATTACCTGAAGGAAGCAATTAAACCCATATAT
GAGGCGGATCGACATGCAAAAAAGGAATAAAAAAAAAGGTTAGAGGATTACGAGACATTGAATGTAGTGTTGTCAATG
AAGATCAGAAAATGGCGACTATTATTGAAGATTATTGCTCGGCAGTACGTAGTTCTATAACCAATGATGGTCAAACCA
ATTCGCAATTGACAATTCGCAATTCGCAGTTGAATTCAAAGTTAGCTCTGAACCCACCCCTGAATTGAGTCTACTGAT
TTAGAGAATCAGAGTTAGCTCTGAGACCCATTAATTAACAATTCAACAATTAAGTAATTTCTTGTCTTTAATTGCGAA
TTGCGAATTGACAATTGTTTCGGTCATCCACCGTTAGAGGCATCTGGATTAAAGTTACAAGAAAATTTGACGTTGATA
GAACAAAGCTTAGAACGGATGGAAAAAAAGTGCTWTACCACCACCTTTAGTCAACCTAAAATACTGATAGCCAAGGGA
TTATCTGCGACTGTATCTTTATTTTCACTTGTTAGGGTTGCATATCAGTGGGTTGATAAAGCTAGTTATATTCTCAAC
AATAAAATAGCTTTTGATGCTGCTGGAGTCAAACAAAGTTATCAACAACTGTTAACAGAAATGTCCCAACAAAAATAG
AAAGCTGGTACACTGAATACCGCAATCGATAACTTTATAAAAACCACCCATAACTACTGGTCTAGACTTTTTCATTGT
TACGAAATTGAAGATTTTTCCAGAACTAATAATGACTTAGAACATGCTTTTGGTATGTTACGTCATCATCAACGTCGT
TGTACTGGTCGTAAGGTTGCTCCCTCATCCCTCGTTATTCGTGGCTCTGTCAAACTTGCCTGTGCGTAGGCGTAGCCC
GTCGTAGACATCGCTACTAAGCTTCACTCTTTTACCGCATCTGATTTAGCACAAGTTGATATTCATACTTGGCTCGAA
TTACGATCTCAACTGCAAAAACACCACAAAGCCAGAATTGAACAATATCGATTTCTCAGAGACCCCAAGGGTTACTTG
GCTAATTTAGAGAGTCGTCTTCTCTAGTAAGTTTTACCATACTAGGTTTTTCTTGTTCTCAAATCCTGTTGCCATGAC
TCGGATCTTGCAGCTAGATGGTAAGAATTATACCCTAGCTCGCATAGTGCCACTTTCAACCCGACGTTGCAGTTCAGG
TAAGTCCGCTTGTCAATAGGGTTTGAGACGCGCTAACCCTGGGGTATGAAGACTTAAACGATCATGAACAATTACGTC
ATGACAAGATGTTCGTCTTGGCGAGCAGCATCGCATAAATTTTTACCATTTTTAGTATTTCCAGGCTCTAAGTGTGGA
GCAAAGAGTTTCTTTGGAGAGGGATTACCTGTACCAATCCTAATTTTGGCTGAGTTGTAACATATAGACCATCAACTG
CCCCAACTGCCGATAGTATGGAAAATCCCTCGACTGCATTGAAAAACTGGGCATTTGTCTTCATTACCTCGTTACCTT
TTTCCCTTTCGATTGCCAACGCCTGCTTGTCTTGCCTAATGCTGTGGCTTGACTGAATCCGCTGAGACATCTCTGCCA
TAACAAAATCGGCAATCCCCTCCTTAGCGTCCTCCAAGTCCTCGACGCCGGACACCAGATTCAAGAATGCTAGCTTTA
GGTTAGAACTGCCGAAGTCACGGCGACTAAGCCGTTGTGCCTCCCAGAAATAGGAATCCTTGCCACGGTTTTGATCGT
AGAAGGCTGATACGAACACTAAGAAACGCAAATAAGCCTGCCGATAGCTCTGATCGTAGAAAGAAGCAGCTTGTGACT
CAGTCACCTCGC

FIG. 5-3

CACGTATAACACTTGTGATACTGGCTGCGGCTAACAAAGCGCTATAAGTAGCAAGATGCACCCCACTCGATAGTAGGG
GGTCTAGGAAGCAAGCAGCGTCTCCCGATATGAAGTAGGCTGGTCCTGAAAAGGAGTCGGAAGTGTAAGAGTAATCTT
GCTCAACTTTCACGTCTGAGACTAGCTCCCCTAGTGCAACCAGATCCGCTATCAAGGGACACTCTGCAATCGCCTCCA
CGTAGATATCCTTCAAGTTCTTAGTCAGTCTCTCCTTGTAGGTTGACTTATGCATCACTACACCAACGCTCATAATTT
CCTCATCCAAAGGAATTCCCCACACCCAACCATCTGGAATGGAGCCCAAGGCAATCGCACCCGACTGACCTTTAGGTA
GTCTCAAGGCGTTTTTCCAGTACCCCCAGATGCCAACATTCTGGAATACGTCGTGTAGACGGCGGTTTTTCAGATACT
CCGTCGCCATGATCCCAGCACGACCTGAAGCGTCAATCATAAAGTCAAAAGAAATCTCCCCGGTAGTATCATTTGATT
GTGACCAAGTAGCGCTGCGCGGGCGATCGCCATCAAAAGACAACTGGCGAATTTTAGTCCCTTCAAAAACCTTCACAC
CCTGGCTCTTTGAATGCTCTAAAAGCAAGTGGTCGAATTCGTCACGGCGAACTTGGAAGCTGTAGGTGTTGTCCCCCG
TAAGTTCCCCAAAATTGAGGCTCCACTTTTCCGTTCCCCATTCTATGTACGCTCCAGGTTTACGCTGAAAGCCATAAG
CTTCAATTTTCTCGCGTACGCCAAGCAGGTCAAAAATTTCTAAAGCAGAGGGCAAAAGAGATTCCCCAACGTGGTAAC
GCGGGAATACCTCTCGTTCTAACAGCGTTACATCAAAGCCCTCACGAGCCAATAGGGTAGCAGCAGTAGATCCAGAAG
GTCCCCCTCCGATAATTAGAATCTGTGTGGAATTAGGCAGTGTAGACATTGCAGGTTTCTTCTCCAAAAGATACAGTA
TTTTCGCAACAATGGCGGTTGTGTCGCCTAGGGACAAACAATCTGTCTTACTCGTGTGGCATTAAGCGACAACTCCAA
AGATTTCTTTGATAAACTTGGCTTGAGCTACACTGTTCCCCGGCCGATAAAGATACTTCCACTCCCCTTTAACTTGGA
TGTAAGTTTCGTCTACCCGCCATGAATCATTCGTCTGCTTCAAATAAATGAGGACGAATCCGAAATCCAGTTCCAAGC
CGCATTTCAACACCCATCAATTCAGGGTGGAATGATCCACGTCTATGCCTCGCTCCTACATCATCTCCTCCAAGTCCC
AATAGAACAACGAGCGGCAGTACCAACGCACATTAAGCAGGATGATTTCTGGCAAAAGGTGACGCCATTTGAACAGGG
AGCGAGTGGAAATGGAAAGCTAAGAGCCTGTCAAGAAACAACTTTTACAATTTATTATCTAGAAAGCTTACTGAGAAA
GCATTTCTAGCTCAAAAGAGGCAAGGTTGTTACATGACAAGCTCTAAGCATTAACGACAAGGATTTCCTACCCTGCAC
TATCTCACTCAGCTTTTTGCGACACAACCTTTAAAAGCACACTTTTAAAATGCGATGGGCTTACGCCTTACAGGTACT
TGAGAAACTTGTTGTTTTCATCCACGATAATTTTTTTCGGTTCCTTGATTTCGTCAGTTACCTCGAATGCTAGCAATG
TGATAAGGTCTCCTGAATTGACTAGGTGGGCGGAGCCACCGTTCATACAGATTACCCCGGAATTTTCCTCACCTTCTA
GGACATAGGTTTCTAGACGATTACCATTAGTGTTGTCCACCACCATAACCTTTTCACCCGGTAGTATGTCTGCCTTTT
CCATCAGAACTTTGTCTACTGTAATACTTCCGATGTAGTTAACGTTGGCTTCCGTCACCGTCGCTCTGTGAATTTTCG
ACTTCAACATAATACGCATCGTTTCTTTCCTCATGTGGCTTTAAATTTCAATTGAGTTGACTGAGAAATATCTGAGCC
TATATCTATTTGAGATGGCTGATACTTTTTAGCAAATAAACTCAAGTTTTTTGGGGCTATAGAAATACCAAACTTTAA
ATTTATAATATCAGATTGTCCATCAAACCAAAGTCGATTTAGTAGCCTATACTCTTGTTTGGAAAATGCAGTTCTTCC
ATGCAAAACTCTAGTATTATCTACAATAATTATTTGGTTTTGTGCAAGTTTAAAAATTACTTGATTGTCAGGATTATT
TACAAAGTTTTCAAATGATTTAAATGCCGCAAAACTTTTCGATTCAACCGAAACATGAGCTGCATTATCTGCTCTAAA
CCTTACAATAAGCCCAGCATGATGTTCTTCAAAAATAGGTTTAGTTGCTTTTTTATTATCTCTTTTGACTGTAATCGC
ATCAGGATTAAACAAAGTTAACAATCCAACTGGGTTTGTCCGCTTTAGATGTTCATATACCAGCTTGCCATCAATAAG
CTTGGTGAACCCGCCATTTGCAGCAGCAATCTGGCACTGCATTGCCATTACTTTTGGTGGAGTAATTGTGAACGCTCC
ATCCGTATGTAACGATAAATCTGTAGTTGTAGTATTTACATATTCTGGATAACTATCAACAGGACTGATGGGAACAAT
TCCCTGTGAATCAGAATGTTCGTGCTGAATAATTGTTCCAAAATAATCAGACAATTTTAATAAGTTATTCTTAGGTGT
TGCTGAAGGTTCGTGTTCTAGTATTACGAATCCAAACTCATTAAATTTATTTGCCATCTCAGCTTCTTTAGAAACTGG
CATTTCTAATACACTTTTCACTCTAATTATTAGATTTTCAATTTTTATTGAATACATTTTTAATTTTCTCCTGGATGT
CACTGCTCTGCTACAACTAGCTTAATTTTTTTAGAATTCTCATGTTTACAATAAATACGTTTTCAAAAAAAATACTAT
AAGTCTTGCTGATATAGGTTGTAAATCCCTCGATATAGCTAGGTAATAATCAAACATAAAATTAAATGCCTCTATTGC
ATGTTTATTTAGAGCATGATAAGTATGTTTAAACAGAATTATTTACATTCAGTAGACCAAAAAGCTTTCCAAAAGACT
TACCAGCTATTGATTTTCCTTGGGAATGCAATAATCCGCCGATATAAGCTTTTAAGAGGATGTTTAAAAAATTGGGAG
TTGAGTAAAAAATATTTTAAACATCCTCTAATAGTTTAAAATTCAGTTTTTTACAATACAACCATTTTTAATCCACCT
TGAGGATAAATTGAGAAAATATCTCGTACTGGTTTTAAAGGAGGTTTGCCTACCAATTCCAATTGCCAATTCCGCAAA
ATATTAGCCAATACTAACTTCATTTTAAACTGAGCAAATGCCATACCAATGCAAGTTCGGTTACCGCCACCGAAAGGG
AAATACTCATAATTTAAAAATTTATTATCTAGAAAACGTTCTGGCTTAAACTGTTTAGAGTTAGGATATAGTTCTTCC
CGGTGGTGAATTAGATAAATACATGGATAAAGACAAGTTCCTACCTCAAATTGATGACCTCCAATTTCTATTGGCGAT
TTTACAATTCGAGGAAAAGTAGTTAGACCAACTGGATATATTCTCAAGGTTTCAGCACAAACTGCATTGAGATAAGGT
AATTTGCTTATTTCCGTTGGGTCTGGATTATCTCCTAACTCATCTAATTCTTGCAATAACTTGGCTCTTATCTCTGGT
AAGTAATGAATCCAATAATATGCCCATGTTATTGCTGCAGATGTAGTTTCATATCCAGAAAAGATAAGTGTCATTAAC
TCATCTTGCAAC

FIG. 5-4

TCCTCATCTGTCATTTTTCCTCCATTTTCATCTCGTGCTGCCATCAGCATACTGAGGATATCATTGTTGTAATTGTTA
CAATTTTCTCTACGTTCTTTGATTTCTGCAGAAATGATATTTGCAATCTGACGTTGGCAACGTAAAAGATTACCCCAG
GCACTCCAAGAACCCCAGTCTCTTCTAAACACATTGAAGAAAAGAGAGCTAGAAGCAAAGGGATTAGTTATAGTGGAT
ACTATTTGATTAACTATCAATTTGAGTTGTTGATAACGTTCCGTTTTATCTGAACCCAGTAAAACCGTTAACATCGCT
CGCAGCGTAATTTCTTTGACTTCCTTGTAAATAATCAATCTTTGACCAGGTTGCCAATTAGAAGTAACCTGCTTCGTT
GCATGGCATATTAGTTCTCCATAGTTAGATATATTTTGACCATGAAAAGCAGGCATCAGTAGTTTACGCTGTCGTTTA
TGACTACTTCCATCAAGCAAGGTGACGGAATTGTTGCCTAAAAAAAATCCTGCTAAATCGTTAGCTTTAGCTTTTCCA
CTGTCAAAATACTTGTGTTTATCAAAAATTTCTTTTATATCCTTAGGATTACTAATAAGTACTAAAGGTTCAAAACCA
ATAGCTTTGAAGGTAAAAGTGTCTCCATAGCGTGCTCGACACTCTTCCAAAAATTCACAAGGATTATTAAGCCATTGC
AATAAGTTCCACCAAGATGGTGTAGTGGGACCAGGAAGTAATGAGGATTTAGCAGTATTAATCATTTTAGTTATGCTG
GATTTGGCGTAAATTTACTAATTAGACTTTGGACACTATTGAATTTCATTTTTCCAATTGATGGGTTTTCTGTGCTTG
TTCAAGAGATATTTGCATTTGTTGAGCCAATACCTTGACATGAGGCTCACTCAGCATTGAAACATGATTACCCGGAAC
TATATGGATTTCCACTTCTCCATCAGAAAACTGATTCCAACCCCATGTTGGCTCTTGGAAAATGTGAGAATAACTTTC
TTGCTCTGGATTTATCTCCCTCGCACAAAACAAAGTGATTGGAGTTTTATAAGTCTTTTCCGGTTCATACTTAATTTG
ACATTGAGTTTGGAAAACTTGTAATAAACCACGAACAATTTTGATATCTGTTTGAGCAGGCAAAAAACCAACTATTTC
TAACTTTTGCTTGAAATAATTTAATTGTTGCTCCCAAGTTAGAGAAGTTAGAGTTTCATAAGATAAAAATAGATTTTC
TCCAACAATATCTTCAATAACCTCAGCCATTCGACATATCCACTTTGCATTATCCCAGTTAGAAAAATCATTCTGATG
ATTAGCTTGAGAAGTTGGTGCAGGAGTATCTAAAATTCCAACATAAGCAACAGACTTTCCAATAAGTTGTAGTTGATT
CGCCATTTCAAATACTACATGACTGCCAAAGGAATGACCAGCCAAGAAGTAAGGACCAACTGGTTGAACTGTTTGAAT
TGCTTTAATGTGTTGGGAGGCTATTTCTTCAACACTTTTATGAGGTTCGGTTTCACCATCAAGACCTTGTGCTTGTAA
ACCGTATAACGGTTGATTATTTCCAAGATATTGTGCTAAGTGGTGGAAGTAGAGAACATTTCCACCTGCTCCTGGTAC
ACAGAACAAAGGTGGTAATGAACCGTTTTGTTGAATTGGTACTAATGGAGACCAAAGTTCGGCTCCGGAATCGGAACC
AACAAGAAGTGCTAGTCGTTCAATGGTGGGATTTTGAAAAAGAGTGGCTAAAGGTAAATTTTTCTGGAATTGTTGTTG
AATCTCGGACATTAGACGGACAGCTAGAAGGGAATGTCCTCCTAAGCTAAAGAAGTTGTCATGAATACCAATAGAGGG
TAGATTGAGAACTTCTTGGAAAATCTCAACTAACTGACGTTCTGTTTGATTCCGTGGTTTTGTCTGCTCAGAAGTATT
CAAACCTCCATATATAGCAGCAATTTGTTCCCGATTAATTTCTCCTCTTTGAGTAAGGGGTATTTGTTCAAGTTGGAC
AAAGTTAATTTGATTGGGTATCCCAAAGCGATCGTGTAGTTGTAACTCTTGTAAGGAGAGTGCAGCAAGTTCTGGTGT
GGGAGAGGTGAAGTAAGCAGTTAATTTCTGCTTGGGCTGACAATCACGAATCAAATGTTCAACATTTGTTTTAGTTCC
ATCCAATCCGATTAATAGATTATGTTCCGAACCAGATAAAGCTGCTAAAAATGAGTAAAATCCTTGTTGAGGAGTAAT
AATAAAATAGCCCTTAGCACGACTGAGTTCTTGGAATTGATAGCCATGACTTATTCCGGTTTCATTCCACATACTCCA
AGAGCAGCAATAGCTTTGGAAACCGTTTTGTTGTTGATAATCGCTCCATGCTGACTGAAAACTATTTGCTGCACTATA
AGCTGCAACATTGGTTCCTCCAAAGAAACCATTTACAGAACAAAAGTGGACAAATAAAGCATTTTCTTTATCCTTGAG
CAATTGATGCAATACCCAAGTACCGCTAACTTTAGGACGTAAAACAGCAGCGATATTTCCTGGGGTTTCTTTCTCGAT
TGGCGTTTCCTGAATAATCCCAGCCATATGAAATACCCCATCAAGTTGAGTCCTCCATTCTTGTGTTGCTTTTTCTAC
TACCTGTTGTAAACCTACTAAATCACAAATATCTACAGTTTGATAAATTATTGAACCTGGTAGTTTTTCTAATTCTTG
ATACCTCTGCAATTTTGTGCTAGCTTCCTCATTATTATCTTCAATTTGAGTTCTACCAACTAATATTAAATTTGCTTG
ATAATGTTCTAATAAGTACTTTGCAATAACAGTCCCAATTCCTCCAAGCCCTCCTGTAAGTAGATACGTTCCTCCTGG
TAGAATCGGAATTTTTTGTTTTTCCTTAGCAGTCATATCTACTGGTTCCAGACCAGACACAAAACGTTCTCTATTGCG
TATAGCAACTTCCAATTCTTTATCAGCAGAATACAGTTCTTGCCAAATATAACTATTGTTGAGTTCTGGTGCTAATGG
TAAATCTAAATGACGAGTAGTTAACCAAGGCATTTCTTGACTAACAGTTTTAAGTAAGCCTAAAACAGTGGATTTTTC
GGGTTGAATTTTATCTGTGGGATGAACTAATTGGCTTTGATTAGCAATCCATAATAATTTGACTGCTTGCTGTTTGCC
TTGAATTTCTTCTAAAGCTTGTACTAAAAATAGTAAACTGTAAATTCCTTGTTGTTGAGTGGACTCTAAATTTTCCAA
GCTAGAAATTTTTTCAGTCTGCTCGTTGTAGTTCCAAAGATGAAGAATTTGACTAATTACTTGGCTATTTTGCCTCAA
AGAATCAATTAACAAGCGATAGTGTTGTGGATTTCCAGGAACAACAGAATAATGATTTGGGCTAATTTGAGCAAAATT
TGAACCAATAGTAACTTGAGCATATGGTTGAACAGTTTGGGACATTCCTCGGTTATCTTGTTGCCAACCCAAATTATC
TGTAAATATTAGGGTTAAAGTTTTCTGAGAAGAATAATTGAGTAAAGTATTTTTACTTTCTTTAATTTGCCATACTTT
ACGGTAAAACCAGTTGGGAATAGTATTAGTATTATCAAGCAACAAGTCTACACGCTGTCTGAGAGATTTAAACTCACC
ACATTCAAAACGTTGCTTAAGGAGGGAACGTTGAATTTTACCGATGGAAGTTTTGGGAATCAGTTCTTTATCTATGGG
TATTAAATAACTTGGATTTATCCCGCAGTATTTTATAACTTGTTCCCTAACCTTTTTCAAAAGCTCTAATAATTGATT
CTTCTCAGATAC

FIG. 5-5

ATACGGAGTGAAAAAGATTACTAATTCTTCGGTATTATTGCTAGCAACGCAGACTCCACAGGCTGCGGTATAAGAAAC
TTCAACCTCTCCTAATTCTTCAACAACAGCTTCTATTTCATGACTATAATAATTAACTCCATTAATAATAATGATATC
TTTTTGTCGTCCTGTAATCGTTAAGCATCCATCTTTTATAAATCCTAAATCACCTGTATTAAACCAACCATCTTCGGT
AAATGCTTCCTTATTTGCTTTTGGATTTTGATAATAACCAGAAGTAACGGTTAATCCTTTGACCTGAAGTAAACCAAT
TTCACCTTCTGATAATACTTCCATGTCTTGATTGACTATTCTCAGACAAGTACCCCTAATCGGTTTTCCAAGATTTAC
AAAGGAATTATCATCTGAACTTGATAAGAGTGAAAAATTGTCAGAATAAGTAATACCAGAGGAAACCTCAGCCATTCC
CCAAGATGGAGTCATAGCATCCCCAGGTAAGCCAAAGGGAGCAAGTAATTTCAAAAAACGTCTTGCTGTTGCTGCAAC
AATTTGTTCCGCACCATTTAACATCAAGCGAATAGAAGATAAATTCCAATTCTGCTTTTCTATTTCTTGAACAAAATC
ATTAATTAAACTATAAGCAAAGTTAGGAGCAAAAGTAACAGTGACACCAAAAGTATCAATCCAATCCAACCATCTTAA
AGGTTTTTCAATCACTAATTGACTAGTAGCATGAATTTGTTTACATCCTAAATAAATATCCCGGATATGAAAATATAT
TAAACCTGCAACATGGTCTAAGGGCATCCAATTTAAGGTTATATCTTCTGGGGTAAAATTATTCATTTGTATTGAACC
AATAGTCCTACTCAGTAGATTTAAATGGCTCAACTGTACCACCTTAGACATACCTGTACTACCGGAAGTAAGCATGAA
CAGTGCTAAATCTTCTGGTTGGGCATTATAGTAATCTTTATCTGTTGAGAACTTTTGTAAACTTTCAATAGTTTCTAA
CTTAAAGTTGTCGTCATTTAGATTTTGAGACCATTTCTTTAGTTCTGACAATGATTTTTTATCTGTTAAAATCAAAGG
TCTTTCTAACATCTGCCAACTATTTTGTAATTTATTTAGATTGACATTGGGCTGGTCATAGCTTACAGGAATTACAAC
GGGTACGGGAATAAAGCCTCCCAACACACAACCCCAAAAAGCACTAATAAAATCTTTATTTTCTTTTAATTGCAAAAT
AACTTTATCTTGTGGCTTAATTCCCAGTTTTCTGAAGCCACCTAGAATTCTTTGAGCATCTTCTAATAACTGGGCATA
TGATTGAACTTGTTCGGAACCATCAGAGTTAATATAAGTGATTCCTTTGTGAGGAAATTTCCCAGCAGTTTTTTGCAG
CATCTCCCCTAAAGTTTCTGGAGATGATTCTGGAAAGATTAATACTTCTTCGTGGCTGATGGCAGGTGATTTTATTTC
TAATAGGGAACTACTCTCTTTTCCCCTAGCAGTTCTGGGAGTTTCAACTGGAGTAGAACCTTGATTGAAAATAGCTTG
GATTGATGGTAAAAGTTCTTCTAAATGTATCGGAGAAATCGTTTTTACATTTGGCTCAATAAAAACAGCAACTTTATC
AATTTCCGCCTGAGAACCTATTTGTTCTTCCCAAGTGTTAATTAACTCAGAATCAATTATGCTAATAGAAGCTAAACC
TACTTCATCAACTTCTCCAAAACTTGTCAATGGTAAAGCAGATACTGGTACATAGATGCAGGGTAATGGGTATCCAGG
TAACTGAGATTTTAAATAATGGTGTAAAAACTCCCTAGCCCAAGAACCATCTTTGACTACGTAAGCGACTAATTTTTG
ATTGCGTACCATTACATAGCAATCTTCTACCCCTTTCGCTGTTTGTAAAGCTTGTTCAATACGTTGTAGGTTAATTCG
TTGTCCATTAACTGTGACAATTCGATGCTCTTTTCCTAGCAATTCCAGAGAACCATCGACTCGACGACAACCCCATTC
CCCTGTTTTTAATAACTTACCCAGTTGGGTATGTTCTATGAAACTTATAAATTTTTCTGGTTCTGGATGTAACTTGTC
TGGGAGTAAATCGCAATTCCCCAAATAAATTTCTCCTTCTACACTCAAAGGAACTAATTGTTGATGGTTATCTAAAAT
GTAAATTTGTAAATTATTTGAACTCAGGAAATGAACATATTTATCCAGCAGGTAAGTTGTAGCGATGTTATTGTAACT
GTGCAGTACCTGCTCTTGCTCAGAATCTGTGAATAATGGTAATTCACTTATCTTTTGTTGGGGATTTTCTACAATCGC
GCTACACAGATTCTGGAAATGAGCAGTCATGCGCTCAATAGTTGACCCATCAAATAAGTCAGTGTTGTATTCCCATGA
ACCCACTAGTGCTTCGGAAGTTTGCTGCATTGATACTGTTAAATCAAACCGGGCTGTTTCTGTTTGAGAACTCAATAA
ATTAAGGGTCACACCAGGTAATTCTAATTCACCCATGGGTGCATTCTGCAACACAAACATTACCTGGAATAAGGGTGC
ATAACTCAAAGAGCGTTGTGGTTGTAGTACTTCAACTACCTGTTCAAAAGGCACATCCTGATGTTCATAAGCTTCAAG
TGTAGTTTCCCTAACTTGTGCCAGCAAATTCTCAAAACTGGGATTATCTTCAAAACGGGTTTTCAATACCAAAGTATT
GGCAAAAAAGCCAATCAAAGACTCAATTTCACTGCAGTTGCGATTGGCAATGGGTGAACCAATTAAAATATCTAATTG
ACCGCTGTAGCGATAGAGTAAAGTGGCAAACGCTGCGTGCAGGGTCATAAATAAGGTAGTACCCGAGTTCCGAGACAG
GGTTTGCAACTTCTCTTTTAAATCAGTATTTAAACTAAAACTTTGAGTAGTACCCCGGAAAGTTTGCACGGTTGGACG
AGGACGGTCAGTAGGTAATTGTAACAATTCTGGTGCACCCTCTAACTGAGAAAGCCAGTAATTGAGTTGAGTTTCTAG
TACCTTTCCACTTAACCATTGTCTTTGCCAAACTGCAAAGTCTGCATACTGGATTGGTAATTCTGCCAAGGGGGATGG
TTTTCCTGCACTAAAAGCTTGATATAAAGTAGATAGTTCTTGGCTGAATATCCCCATTGACCAACCATCAGAGACAAT
GTGGTGCATCGTCAGTAATAACACATATTCTCTGGCATCTAACTGCAATAAACTACACCTGATTAGTGGTGCAGTTTC
TAAGTCAAAGGGGGTAATTGCTGCAAGTTGTGCTTGTTGGTGAAGGACACTTTCCCGTTCTGTTGCTTCTAGTTGCTG
TAAGTCCGCCACACTGATGTTCATGGTGGCTTCTGGGTGAATTACCTGTATTGGTGTGCCATTCACAGTTCGGAAGCT
GGTGCGTAGTACTTCATGACGGCGGACTATTTCTGATAATGCTTGTTGCAAGGCATTAATATCCAACTTTCCAGTGAC
ACGAATTGCTCCTGGCATGTTATAAGTGGCACTTGACCCTTCAAGTTGGTTGAGGAACCACAACCGGTCTTGTGCAAA
AGATAGGGGTAATTGTTGGTTCTGTGTTCTTGGCTGAATGGGGGGAAGACTTAATGCGCTATTAGTAGTACGTAATTG
GGTTAATGTTTGCTCTAATTGAGCTACAGTGGGAGAGGAAAAGACTGCACTTAGTTCTATTTCTACTTCAAAGGCAAC
TCTGAGTCGGGAAATTAATCGGGTTGCTAGTAGGGAATGTCCTCCCAATTCAAAGAAGTTGTCATGGATTCCAACATT
TTGCACACCTAG

FIG. 5-6

AATAGAAGCGAAGATGTTGGCTATTATTTCTTCACCCGATGTACGTGGTGGGACATATTCATGTTCTCGGCTAATTTC
TCCATCAGGTGCTGGAAGGGCTTTACGGTCTATTTTACCGCTGGGTGTCAACGGTAAGGTGTCTAAGATGACAAAGGC
ACTGGGCATCATGTATTCTGGTAGCTTTTGTTTGAGGAATTCACGCAGGTGATAGGTACTTAGTGATTCATCCTCACA
GACTATGTAGGCTACTAAACGTTTGCTACCTGGAATATCTTCTATGGCAATGACAACGACTTGTTGGATTTGGGGGTG
GGTACTGAGGACTGCTTCGATTTCTCCAAGTTCAATGCGGAAGCCCCGCACCTTCACTTGGTGATCGATACGCCCTAC
AAATTCGATATTACCATCCGGCAGCCAGCGGGCTAGGTCTCCAGTCCTAAACAATCTTTCTGCTTGTGCCTTTTTACT
TTTCCCCCTGTCCTTCACAAACGGGTTGGGGATAAACTTTTCCCGCGTTAACTCGGGCAGATTTAAATACCCTTTTGC
AAGCCCATCTCCGCCGACGTATAGCTCACCCGTAACACCAGGTGGCAAAAGATCGCCATACTTGTCGAGGATGTAAAT
TTGTGTGTTTGAAATCGGTTTTCCAATCGGAATCGTATTTCTTTTCAGATATTCTTCTAGAGCTTCTCCTAAATCTGC
TCTTCGCTCGTCTGCCAACTGCAAATGTATTATTTCTTTATGCAGGACAGCAGTTTCCCTATCCCCTGAGCCACTAGG
AAGATTTTTTAAAGCATCAAGTTTCTCTTTACTTTTTGCTTCAATTTGATTTGCGATTCTCAGTTTGACTTCAAAGCA
TGTAACATCAGCGGCAACTTCTGAAGAGCCGTAGAGATTGAACAATCTGGCAGAGCTGATTTTCTGGTGAAATTCCTT
AGCCAAGGTTAGCGGTAAGACTTCACCGCTGCAAAAGACATATTTGAGATATCGAAGTTTTGTCAGTTGTTGGGGCGC
ATTTTCCAGTATCGCTTTTAATAGCGATGGAACGAGAACAATTCTAGTTACCTTTCGATCGCTCAACAGGCTCATTAG
CCTGGGAATATTGCCCCGTATATCATCTGGAACGATCACAAGGGGAATTCCTTTGAGAAGGGGAGAAAATATTTCCGC
AACATGATCGCCAAAATTGATGGATGTTTTCTGAGAGCAAATCTCATCTGCCCCAAATGGTAGCATTTCCCAGATCCA
ATGCAAGCGATTGACAATGCCGCGAAGCGTGCCGAGAACGGCTTTGGGTTTTCCAGTAGAACCAGACGTATAGATTGC
GTATGCAAGGTCATCCAGTGTTGTTTGCCGATCCAGATTTTCAACGCCTTCCCTAGCAATGACATCCCTATCCCTGTC
CAGGCAAACGATATGGGCATTTTGAGGGGAAATCTTTTCGAGCAGAGGCTGCTGGGTCAATATGATATGCACATTGGA
ATCTTCCTGCATGAACGCCAGTCGTTCTTGCGGATAGTTCGGATCTAACGGCACATATACACCACCAGCTTTAAGTAT
CCCCAACAGTCCTACAATCATATCAATGGAGTATTCTATGCAAATACCCACCAGCACTTCTGGTTTCACTCCCAGAGT
TTGTAGATAATGTGCTAATTGGTTCGCTTTTTGATTTAATTGTTGGTAGGTTGATTGTTCTTCTTCAAACACCACTGC
TATGGAGTTGGGGTTTTTTTCTACCTGTTGCTCAAATAACTGATGAATACATTTATCAGATGGGTAATCCGTTGCAGT
GTTATTCCACTCAACCAATAACTGATGACGTTCTACTTCACTTAATAAAGGTAATTGAGCTACCTTATGTGAAGGATT
TTCCACAATTGCAATTGCTGATAAAACAGTTTGCAGATATCCTAAAATCCACTCAATAGTATTTGAAGAGAAACGAGC
AGTATCGTAACTAATCCTAACTGACAACTTATCCCCAGGAACTGCAACTAAAGTTAGTGGATAATTAGTTTGTTCAAA
AACCTCTATATCACCTAAGTGTAATGAACCTTCTTCATTCAACAAAGAATTATCAATTGGATAATTCTCAAACACCAC
AATGCTCTCAAACAAAGGTATTCCACCTGGTATCTCAGAAGTAGCTTGAATATCAACAAGAGGAGTATAAAAATACTC
TTGTAATTCAACCATTGACTGTTGTATTTTTTGCAACCAAGGTATGAGTTGCTCCTGGGTGGATACTTGTACTCGTAA
GGGAAGGGTGTTAATAAACAGTCCTACCATATTTTCTATCTCAGAGAGGCTAGGAGGACGACCAGAAACAGTCACACC
AAATACTACATCTTTCTCACCACTATAACGACTCAATAGTAAAGCCCAAGCAGCTTGTACTACAGTTGATAAAGTCAC
ATGATGTTGTTGTGCTATATGAAGTAACTTCTGAGTGCATTCAGGGGATAAACTACTTGTTCTCTCCTGATAATCCGC
AGTTTTATACTGTTGCTCTTTCAGAAATTGAGTTTTATCCATTACCAATGGAGTGGGAGCACTAAAACCTTGTAAAGT
TTGTTGCCAAAACTCAATTGCTGCTGATTTGTCTTGAGAATTCAACCAAGCAATATAATCCTGGTAAGGACGTGGTTT
TGGCAATTGGCAATTTTCACCAAGCAGATGTGCTTTATAGAAAATTAAAATTTCTTTAAAAATAATTGATAAACACCA
TCCATCCATAAGGATGTGGTGATGACTCCAGATAAATTTGTAATTATCTTCGCCTAGCCTGACTAACGTACACCGCAT
TAATGGTGCTTGGGATAAGTTAAAACCTTGTTCTCTTTGTGTTTGCAATAATTGTTTTAATTGTTGTTGTTGATCATT
AGAAGAAAGTTCTCGCCAATCAAGAGTATTCCAAGGAACATTAACCTGTTTTAGTACTACTTGTAATGGAGTTTGGCG
ATTTTCCCAAACAAAAAATGTACGTAGAATTGAATGTCTATCTAAAACTTTTTGCCAAGCTCTTTCAAAAGCAGCAAC
ATTGATATTCCCCTTCAAACCCCAGGTCATCTGTTCAAGATATACCCCACTATAAGGTGCATAAAGACTGTGGAACAG
CATCCCTTGTTGCATGGGAGAAAGTGGATAAATTGAAGAGATATTTCTTCTAATTTCTTCGTTGCTCATTGTTCTCTC
TTTTTTTATCTATATTTTTTATATTTACACTATTTGCCCAAGTTTTTTAATAACTCATCAAGTTCTAATTGATTTAAC
TGTGCATCTGGGAAATCACTAGCTGTATATCCAAAACCATTTTCTGACTGGCAATGTTCTATTATTGACTTAATTGCT
TGAATATAGCTTTGTGTCAAATTTTTTACTGTATCATGAGTATGAAAATTACTACTATAAGTCCAATCAATTTGTAAT
TCACCTTCTACCACCAGACTATTAATCTCTAATAGATGGTGACGAGTTTGCTTTGAACTATGATTATCTCCAGTAGAT
TCTGGCGCAAATTTCCAACCCGTTTCCGATTGTATTTGGTCAAATTGTCCTAGGTAGTTAAAACTAATTTCTGGAGTA
GGAATTGTCTGTAGTTTTTGGGTTACAGTAGTATCTTCACACAAGTAACGCAATATACCAAAGCCAATACCACGATGG
GGAATCTCTCGTAATTGTTCTTTAATTGACTTGATAACTTCTGCTGGTTGTTTATCGTCTGGTAATCGCAATAATACT
GGGAATAAACTGGTAAACCAACCTATTGTTCTTGATAAGTCTACATCTGAAAATAGTTCTTCTCTGCCATGTCCTTCT
AGGTCAATTAGT

FIG. 5-7

ACTTTTGAATCTCCCGTCCACTCTGCCAAGGAAACTACTAATGCACTGAGGAGGATATCGTTAATTTGTGTGTTATAA
GCTGAGTTTACTGACCCCAGCAAAGCGCGGGTTTCTTCTGGACTCAATTTCACTCTATAATTAATCGCACTATCAACT
GTTTTTTCTGCTTGAGTGTGAGCAGAATCTAATGGTAGTGGTGTTGTTTCTGACCAAGGTTGGTTGAGCCAATAGTCT
AACTCTTGTTTGATTTTTTCTGATTGTGCATAATTTTTCAATTTCTCTGCCCAATCAATAAATGCTGTTGTTTTCGCA
TTTAGCTGTATTGATTGTTGAGCGATTAGTTGTTGATAGATTGTTTCTAAGTCTGATAGTAAAATTCGCCAACTCACA
CCATCTACTGCTAGGTGATGAATAATAATCAGTAAACGGGCATCAACTTCACTACCTAAGTTAAACATCACCACTTGC
ATTAAAGGTCCCTCTGAGAGGTTTAAACTTGCTTGATATTCCGTGGCGATCTGTGATAAAGCTTGTGGTTGTTCAATG
ACAGGAGTTGATGATAAATCAACTACAGTAAATGCTACGGGATCATCAAAGCCATGGTTTATTTGTTTGTACTCAGAT
GCAACTGATGTGAATCGTAAACGCAGAGCATCGTGATGCTCTAATAATTTTTTCAAGGCTGTTTCGATTAATTCAGTT
TGCAGATGATTGGGAATCTGCAATAAAACTGATTGGTTGTAATGGTGTGCTTCTTGGCTATTTTGTGCAAAGAACCAC
TGTTGAATTGGTGTTAGGGGTGCAACTCCAGTAACTATACCTTGGTTAGCACTGACAGTAACTGTTGTATTGGCTACT
AATGCTAGTTTGGCGATGGTTTGATTTTGGAATATTTGTTTGGGAGTGATTTGTATTCCTAAGTTTTTGGCACGAGAA
ACTACTTGAATACCAAGGATGGAGTCGCCACCAATTTCAAAGAAGTTGTCATGGATGCTGACTTGTTCTTTAAGGAGC
AGTTCWTGCCAAATGTTGGTTAAGATTTGTTCTATTTCTGTGCGTGGTGCGACATATTCATCCTCTCGGCTAACTTCC
CCATCAGGTGCACTTAGGGCTTTGCGGTCTACTTTACCGTTGGGTGTCAACGGTAGGGTGTCTAAGATGACAAAGCTA
GAGGGGAGCATATATTCTGGCAATTTAGACTTTAGGTAGGAGCGCAATTCATTGCTACTCAGTACCTTACGTTCAGGC
TTTGACTTATTAATTGTGTAATCTTCTAGATTATTTTCAGCAAACAATCGTTGATAAAAAGGGTTTTGCTTAAACAAT
TGTTTCCAATGATATGAGATGTATTCAAACTCAATCTCTTTTCCAGTTTTTTTAAAAGACGACCTCGAAGAGTATAA
TCTGGATTCAAATTATTTTCACAAAGCGTTCTCGTACAAACAGCTTCGATGATATCTCCTTCATTTACATAAATTCCT
GGTTCAAACACTGGAAGATAAACTGGTAACCAGCAATGTTCATTTTCTAAAATATCTATACATTCTCCTTCAATTGTG
TGTAAGTTTAATCCCACTGAAAAACCATCTAATCTTCCTGATTTTTCAATAGTTAATTTAATTTGGTGAGTAGATTCT
GTGCTAACAAGCTTGCTAAAGTCTAAATCCTCAAAAACTCCTCGATTGGACAACCAGTTTACTTGATTTAATCCTTTA
ATACATACTCGTAAATCAAAAGGATATCCAACTTGCTCAAATATCTTCTGGGTATAATAACCTGAAACTTTTGTAAAT
TGGGGTTGATTTAGTAATTCATCAGGAAGAGTTACTGCAATAATTTGAGTCACACTTCTTTGGGGAATCATTACACCA
TCTGATTTGAGAAATCTTCTGGCGTTGTTGATAATTACTGCTGCTCCTTCAGATCCACCAATGGGTCCCACAATTTCA
GAAACACATACATCAACTTCTTCTGGTAAGTTGGCTGTAGTAGCGTCTCCATGTATGATTTGAATTTGTTCTGATAAC
CCCAACTCTTGCACGCAAGCTGAAGCTAACTTACTGGTTTGCTCGTCTCTCTCAATTGCGTAGACTTTCTTAGCACCT
GCTTCTGCACAAAATCTGGCTATAATTGCATCCTTGCCCGTGCCAATTTCAACAACTACTTTATCTTTAACCATTTGA
TTAATTGCGACTTGGTAACTCTGGTTTCGACGATGATCATTGGTCATCGCATAGTACAAGAGCTCATCATAAACGTAG
AATTCTGCTACTGAGGGCCAAAGTTCAATTCCTGTCTGGGGCTGGGGATCTTTTTCTTTTGACTGAAGAGGAACTAAA
TATGCTACCAACCGTTTGTGACCCGGAGTATCTTCCCTTTCGGTGACTGCGACTTGCTGTACTTGAGGATGGGTACTC
AGAACTGATTCTATTTCTCCTAGTTCTATGCGGAAACCACGTATTTTCACCTGGTTATCAAGACGACCAAGAAACTCA
ATATTACCATCTGGTAAGTATCGAGCTAAATCTCCAGTTTTATATAGTTTTGATCTGCTATTGAAGGGGTTAGGGATG
AATTTCTCTAAAGTTAATTCCGGTCGGTTGAGGTAACCTCTGGCTAAGCCATAACCTCCGATGTATAATTCTCCGGAT
ACACTTATGGGTACTGGTTCTAAGTGCTTATCTAAGATATAGATTTGGGTGTTTGCAATGGGGCGACCGATAGTAACT
TTCTCGCTACCATGGCTGATTTGAGCCACTGCAGCACCAATAGTAGACTCAGTAGGCCCATAACCATTAAACAAACGA
CGACCAACAGACCACTGATTGGCCAATTCKAMWYTASAAGSWTCCCCTGCCACAATTATCTGACCCAAGGCTGGAAAT
TCATCAGTAGCTAGTACTGCCAGGGCAGAGGGAGGTAACGTAACATGAGTTACACATCTTTCTTGTAAAATTTGCTTT
AAATCCGAACCCGGGATTAACTCAGAAGCTATAGCCAAAATTAGCATTGCTCCAGAAGTCAAAGCGATAAATATTTCC
GAAACTGAAGCATCAAAACTTATAGAAGCAAATTGAAGAACACGACTATTTGGTTCTAGATAAAATAAATTTTTCTGT
GCTTGAATAAGGTTGCACAAAGAAAAATGTTCAATCCCAACCCCCTTGGGAACTCCAGTAGAACCAGAAGTATAAATC
ACATAAGCCAAATTATCTGAACATACCCCAACATCAAGATTCTCCTGACTGTGTTGCTCAATCACTCCCCAATCACTA
TCCAAACAAACCACCTGTGCAGTATGTGACGGCAAAGATTCCAGTAGGGACTTTTGAGCCAACAACACCTCAACACCT
GAATCCGCCAACATATAACTCAACCGTTCTTGGGGATAATTGGGGTCAAGGGGTACATAAGCCCCACCAGCCTTGAGT
ATCCCCAAGAGCCCTACCACCATTTCAAAAGAACGCTCCACGTAAATCCCTACCAGCACCTCTGGTTCGACTCGCAAG
GAAAGCAGGTGATGTGCTAGTTGGTTGGCTTTTTGATTTAATTGTTGGTAGGTTAACTGCTGATTCTCAAATACCACC
GCGACTGCATCCGGTGTTCTCTCTACCTGCTCTTCAAACAATTGATGGATACATTTACTGGGATATTCCCTTGCTGTA
TCATTCCACTCCACCAACAACTGATGCCGTTCTACTTCACTCAATAGGGGTGATTCACTTACCTTTTGTTGAGGATTT
TCCACAATCGCTGACAATAAATTCTGGAAATGACCAGCCATGCGCTCAATGGTTGACTCATCAAACAAGTCAGTGTTG
TACTTAAAAACC

FIG. 5-8

CCAAAAACAGATGAACTCCCCTCCACCATTTCTAAACCTAAATCTAACTGACCTTCCTGTTGAGGTATTTCATAAGGT
TTTATCTTCAATTCTCCCCAATCAACATAGGTTTCTATTTGATTTACAAACAACTTCTGTATATCTTGAGATTTTTGG
AACTGCAGTAGAGAAAAAGAAGCCTGAAAAATCGGCGAACGACTGGGGTCGCGGTGTGGCTGTAGCTTTTCTACCAAT
AGAGCAAATGGGTAATCTTGATGAGCAAGTGCTTCCAATACGGTTTGGCGTACTTGGGCGAGGAAATCTTTGAAACTG
GGATTTCCCGATAAATTTGCTCGCATAACAACAGGATCAACAAAGTAGCCCAAGATCGAAGCAAACTTAGCTTGACTC
CTACCTGAGGTGGGAGAACCGACTAAAATATCCTCCTGGCCTGTGTAACGATACAAAAACACCTGAAAAGTTGCTAAG
AGCATCATGTAAAGTGTTGCTCCCGAGTTTAAAGCCAGCTCCTTGAGTTGCTTAGTGAGCTTGTCAGATAATTTGAAG
TGATGGGAAGCACCATTATAAGTTTTTATCGGTGGTCGCTGTCTTGAGGTTGCTAGGTTTAGTGCTGGCAAATCGCCT
GTCAGTTTTTGCTGCCAGTAGTTCCAGAGTCTTTCCCCTTCAGTCTCCTGCAAAATATTCCTCTGCCAACGAACGTAA
TCTTGGTAAGAATGCTTTAGAGGAGAAAGGGGTGTCTTAAAATCAGCCCATTGTACTTGGTAGAGTTGTGGCAACTCC
TGTATTAACATATCTAAAGACCAGGCATCGCAAGCAATGTGGTGTATGGTTAGCAACAGGACATGTTCTTTCTTGGAA
CGAGTAAACCACCGAACTCGCATAACAGGCCCTCGTTCGAGGTCAAAATATTGTTGATGGCTCTCAATCACTTTCCCT
TTCAGTTCATCTTCACTCCAAGCAGAAGCATCAATTTGCAAGAAATTTAATTCCTGAAAATTATTTACCTGTTGGATT
GACTCAGATCCGAGTTTGGGATAATTTGTACGCAATATCGGATGCCGTTCTATTAGTTTCTCAAATGCCTTTTGCATT
GCTGTAATATCTACTGTTGAGCAAATACGAGCGACAAATGATACGTTATAAGCATGACTTTCTGGTGCTAATTGCCAC
AAAAACCAAAGTGCCCGTTGACCGTAAGAAAGGGGATAGACGTTTAAAATATCTGGGCGATCGCGCAGCAATTGTAAT
ATTTCGGTTTTGTATTGTTTCAGTTGAGCTAATACTAAAGCAGTTGATTCTTCTTGAGGAGCATCGTAACAAAGCCGT
TCGCCCTCACTCCACACTTGCCAACCTTTTATTGAAATATCTTGTAAAAATTCGATTAAATTCATAATTCACCTCTTA
TCCGCTCGTTTTCTTTCCTATTGCTTTGGTAGAGTTGCCCATTATTTTCTGACTCAACTCCTTGATTCTGAGCAACTT
GGCTCAGTTGCTCATTCACTTCAGTGGCTAAATCAACGATACTGATATCTTCTATAAATTTGACTATAGATATATCCA
CGAGCAAGTCAGTTTGAAGCCTATTGTGCAATTCCACAGCCATTAGAGAATCAAGCCCCATAGTGTTCAGGGGCTGTT
GCATATCAATTTGAGAAGTGCTCAAAGAAAGTACTTGAGAAATTTCATCTTTAATGTAAATTATCAAAAGCTTTTCTC
TTTCTCTTGGTAAAGCAGCTTTTAGCTGTTCTAAAAATTCATTGTGCTTTGTCTTTGTTTTGAGGGCTTTTTGCTGTG
ATTTGCTTTCTTTTACCAATTGGGACAGCAATGGTATTTGATTACCAAAACTAAATTGCTCTTGGAACACTGACCATT
GAATTGGTAGGACTCCTACTTGTGGTATGGATTGTTCGAGTAATTGTCCTAGAACCTGCAATCCCTGTTCTGAAGACA
AAAAAGTCATTCCCTTGGACACCATTCTATCTTGATGAGGACTATCCAAATTTGCTGCCATTCCCTCTTGTGCCCATG
GTCCCCAGTTAATGCTCAAGCCAGGTAAACCCATACCCCGTCGATGATGGGCTAAACCATCCATGAAAGCATTAGCAG
CAGCATAATTCCCTTGACCAGGCGAACCCAATATTGAAGCCATAGAGGAAAAACAAACAAAAAAGTCCAAAGGTAGAT
TCTGAGTCAAATTATGCAAATGCCAAGCCCCTTGTACTTTTGGTGCCATCACCTGTGTAAATTTTTCCCAATTCATGT
TTAACAGCAAACCATCATCCAATATCCCAGCAGCATGAATTATTCCTCGTAATGCTGGCAAAGATACTTTGATTGACT
CTATAATTCTTGCCACATTTTCTTGTTGGGAAATATCTCCACACAGGACTAATACTTGCGCTCCTGCCTTCTGTAATT
GTTCAATGGTTTGTTGAGCTTTTGCTGATGGCTGCCTACGTCCGGTAAGTACTAAATATTTGACCCCTTGTTGTACCA
TCCACTCAGCGGTTTTTAACCCCAGTGCTCCCAGACCTCCGGTAATTAAGTAACTGGCTTCGGCTTGGATTAGGTTGT
CTAAAGACTTATATTCTGATAGCTTCAGTTGAAATGGTTGTTGCGAGGAAATTTGTAATCCGGACTGTGTAGATGTAC
TCATTTTTTGTTGCCGCTCTAACCGGGCAACGTGACGTACCCCTTGACAGTAAGCAATTTGGTTTTCATCACCAGGAG
ATAATAGTTCCTCTAACAAAGCAGCTACTGTTTGGGAATCTTCCATAGTTGGATCTAAGTCTAAACACCGGCATTGTA
ATTCCCTATGTTCCTGGGCAATTACTCGACCTAACCCCCATAAAGGTGTTTGTTGGAATTGTATAGGAAGGGACTCAT
TACCCACAGATTGTGAGCCTTGAGTCACTAACCATAATGGGGCACTTTCCATATCTTGATTTTTTACTAAGGCTTGGA
CTAAATGAAGTACGCTGCCACAGCCCAGTTCTTGGGATTTTTGCAACTCCTGTGCCCCAGTCCTTAGTGCTATTGTTG
AGTCCAAACTCCACAGGTGAATAATTCCTCGTAATGGGGGTTGCTGCTCCAAGCTTGATTGCAATAGGTGCAGGAATT
CCTCAGGATGGTTGGGGTTGATTTGATAATGTTGAGATTCTAACTGCTGGTAATTTTCCCCTGGTGTTACTAATATAC
AATGCCAACCTTGTTGTTCTAAGGATTCTACCAGATGTTTGCCTATACCTGTGGGTGGGGAAAACAATAACCAGCTAC
CTGATTTTGTTAAGTCAATTGATTGGTTATGGGGTGAAATTGATTGGGTTTGCCAATGGATTTGATATAACCAATTAT
TAAATTTTGGTTCAATATTACGCAACAAAGCCTCGCGAGAAGTACGTAATAAAGTTAAACCTTCAACTCTTGCTACTA
CTATTCCTTGTTCATCCAATAAACAAACTTTACCGCTCAAAGTTTGTTTATTAGTTTCTGTTGCACCTATCTCTACTT
GAGTCCACAAACTATTACTACCACTCCGATAAATTTGTAGTCGTTTTATTTCCAATGGCAAATAAGTTTCTTGGTTGT
CCGTTTTACCCATAACTGCTGCTAACACCTGGAAGCTAGCATCTAAAAGAATTGGGTGCAGTTGGTATAAAGTTGCAA
CATTCACCTCAGTTTCTGGTAACTGAATTTCACCTAGTGCTTTTCCTTCGCTGTGCCACAGTTGTTTAACGGCTTGGA
AAGAAGAACCGTAATTAAGACCCCATTCTTCAAATTTTTGGTAGAATTCAGTAGGTAATATCTGTTGGTTATACTCGT
CTTTAATCGCTT

FIG. 5-9

TTAAGTTTGTTGTTTCTAATTGGGGGTCTTTATTACCTACTAATATTTTTCCTTCAATATGTAGAATCCATTTAGGTT
CTGAAGAATTAGTGTTTATATCCAAACTGAAAATTTGGAATTTATAGCTTTGTACTAACTGTAAATTTAAAACTATCT
GAATTGTATTAATTTCATCCTTTGATAAAATTAATACTTTTTGGATTGCTATATCTTCTAGGATTAAATCATCTGAAT
TGAATAAAATTGAACCTGCTGCTAAGGCTATTTCCAAGTAAGCTGCTGCTGGGAAAACAGGTTGAGAAAAAACACAGT
GGTGTTGCAGGTAAGTTGGTTGAGAAGCACTAATTTGACATTCAAAACGAATTTGCTGTTCTAAGGCTGCTAAATGTA
ATCTTTGACCGAGTAGAGGGTGAAGATTTTTATGATTTGATAAAAACTGTTTTTGATGTATTAGATTATTATTTGTCT
CAATCCAATAACGTTGCCGTTGAAAGGGATAAGTCGGCAATACTACCTTGCTACGAGAATAATCTTTATCAAACCCTA
ACCAATCAACTTTAACTCCATGCACATATAGTTCAGCCAAACTTTGTAGCATTTGCTGCCAGTCTTCTTGACCTGGTT
TCAAAGAAGGCAACCAAACTCCCACATCTTCTGGCAAGCACTGTCTTCCCATGCCTAACAAAGTTGGTTTGGGTCCAA
TTTCTAAGAAGATGGAATAACCTTCTTGCTGTAATGTGTCCATACTTTGGGCAAATTTCACCGGTTGCCGGACATGAT
TTACCCAATAGCTTGCTGTGGCAATACTATTCTCTGCCCTAGCTCCCGTTACATTTGATACTAATGGAATATTTGGTT
GATTGTAGGTTATTTCTGATGCTACTGCTTCAAAGTCCGCCAACATTGGTTCCATCAAATGTGAATGGAATGCGTGGG
ATACTTGCAGTCGTTTTGTCTTAATGTCTTCTGCTTCTAAGCTATTTTGAACCGCTCCAATTGCTTCTGCCTCACCAG
AAATGACAATGCTTTGGGGTCCGTTAATCGATGCGATCGCTACTTTTTGAGAGTATGGTGCAATTAGTTGATTTACCT
TTTCAATTGAAGCCATTACAGATAACATTTCACCCCCAGAGGGTAACTGTTGCATTAGTCTTCCTCTATGAGCAATCA
GTTTTAAACCATCTTCTAAACTAAATATTCCTGCTACTGTGGCTGCCACATATTCCCCAGCACTATGCCCCATAACCA
CATCCGGTTTTATTCCCCAGGATTCCCATAGTTTATAAAGAGCATATTCTATTGCAAATAAAGCTACTTGGGTATAGG
CGGTTTGAGCTAGGACATTTTCCTGTACTTGAGCGACATCAAGTATTTCTAATAAAGGTTTGTCTAAGTAGTTTTCTA
ATATTTGGGCACATTGATCGCTAGTACCTCTCCATAAAACTTGGTATACGCAACTAATTGCTTCACCAATTCCGGTTG
ATACTTTAATATTCCTGCGTCTACCACCGCAACTATTTTCTTCGGCTTTGTCTCCTCATCTGCCGAAATTACTTGCGC
TAGCGTCGGGTTTTTCAACTCAAATAAATTTTGGGTGAAGTAAATCTCATAGTTAAAAGTAACCGAAACACGTTGATG
AATTAATCTATTTTTTTGCTTGATGTCAACTATCATATTTTTGCAGGTATATCTAAAAGTGCAGTACTATTCAAAGCT
TCAAGGAAAACCTCAACCGAGTGAGAACCATTTACCTTGACTTGATTATTCATGATGAAAAACGGCACGCTGTTGATG
CCATTTAAGCGAGCAAATGCCGATTCAGCAACAACTGTATCAACGACATCGCGATCGTTTAATTGCAACTTTAATTCG
GTAGCATCCATCTGGTATGCTGTACCGATGGCAACAATAACGTTAATATCTCCAATATTCAAACCCTCTTCAAAGTAA
GCTCTATAAATAGCTTCAACGACATCATTTTTTATGTTTGTCGGTGCTAATGCAATCAGTTGGTGAGCAAGCTTAGTA
TTGACAGCCAAACGGATTTTTTCAAAATCTAGCTTAACCCCAGCCGCCTCCCCTGCGCGTTGCGTATAATCAAACATC
TGTTGCATTTCTGGCGCTTTAATGCCTTTTCTATTTTGCATAAAGCTACTAAATTCGTACCCCTCAGCAGGAACAGTA
TCATCCAGAAGAAAGGGATGCCATCGGATATTTACTTCTTGTTCTTGCCATTGTGCCAGTGCATCAAATAGATGTTTT
TTCCCAATTCTGCACCAAGGGCAAACGGTATCATGAAAGATATCTATCAGCATAGTTTTTGTCACTCAAATGCTAATA
TTTGTGTGCATCTGGGGTTTAAAATCTCGTTGCAGAGCCGTTGTATTTAAAGGCTGGAGAAAACTATTAATTTTCTCT
TCAAAAAAACTTTGAGTATTTTCAAACTCTTTAATTAATGCCTCTCTATCCTTCCTAGCCACCAGTCTAGCCAATCGG
CTGTAAGTTTTAGCTAAAAAGCTAATAGCATTACACCTTTCTTCAGTCGCTAGCATAATATCAACGCATAAATTAGGA
TTTTGCGAAAATAAACGTTTTACAATATCAATCTCTTGACGATAGTTAGGAGTTGACATTGTTAAACTCTGCTCTATC
TCTACTCTTGATTGTGCTAAGAAAACACCAAGACTAAATCTACAGAAATGCTGCGTGGCTTGAATAATCACCATCATT
(SEQ ID NO:1)

FIG. 6-1

crpA

ATGGGGCATAGTGCTGGGGAATATGTGGCAGCCACAGTAGCAGGAATATTTAGTTTAGAAGATGGTTTAAAACTGATT
GCTCATAGAGGAAGACTAATGCAACAGTTACCCTCTGGGGGTGAAATGTTATCTGTAATGGCTTCAATTGAAAAGGTA
AATCAACTAATTGCACCATACTCTCAAAAAGTAGCGATCGCATCGATTAACGGACCCCAAAGCATTGTCATTTCTGGT
GAGGCAGAAGCAATTGGAGCGGTTCAAAATAGCTTAGAAGCAGAAGACATTAAGACAAAACGACTGCAAGTATCCCAC
GCATTCCATTCACATTTGATGGAACCAATGTTGGCGGACTTTGAAGCAGTAGCATCAGAAATAACCTACAATCAACCA
AATATTCCATTAGTATCAAATGTAACGGGAGCTAGGGCAGAGAATAGTATTGCCACAGCAAGCTATTGGGTAAATCAT
GTCCGGCAACCGGTGAAATTTGCCCAAAGTATGGACACATTACAGCAAGAAGGTTATTCCATCTTCTTAGAAATTGGA
CCCAAACCAACTTTGTTAGGCATGGGAAGACAGTGCTTGCCAGAAGATGTGGGAGTTTGGTTGCCTTCTTTGAAACCA
GGTCAAGAAGACTGGCAGCAAATGCTACAAAGTTTGGCTGAACTATATGTGCATGGAGTTAAAGTTGATTGGTTAGGG
TTTGATAAAGATTATTCTCGTAGCAAGGTAGTATTGCCGACTTATCCCTTTCAACGGCAACGTTATTGGATTGAGACA
AATAATAATCTAATACATCAAAAACAGTTTTTATCAAATCATAAAAATCTTCACCCTCTACTCGGTCAAAGATTACAT
TTAGCAGCCTTAGAACAGCAAATTCGTTTTGAATGTCAAATTAGTGCTTCTCAACCAACTTACCTGCAACACCACTGT
GTTTTTTCTCAACCTGTTTTCCCAGCAGCAGCTTACTTGGAAATAGCCTTAGCAGCAGGTTCAATTTTATTCAATTCA
GATGATTTAATCCTAGAAGATATAGCAATCCAAAAAGTATTAATTTTATCAAAGGATGAAATTAATACAATTCAGATA
GTTTTAAATTTACAGTTAGTACAAAGCTATAAATTCCAAATTTTCAGTTTGGATATAAACACTAATTCTTCAGAACCT
AAATGGATTCTACATATTGAAGGAAAAATATTAGTAGGTAATAAAGACCCCCAATTAGAAACAACAAACTTAAAAGCG
ATTAAAGACGAGTATAACCAACAGATATTACCTACTGAATTCTACCAAAAATTTGAAGAATGGGGTCTTAATTACGGT
TCTTCTTTCCAAGCCGTTAAACAACTGTGGCACAGCGAAGGAAAAGCACTAGGTGAAATTCAGTTACCAGAAACTGAG
GTGAATGTTGCAACTTTATACCAACTGCACCCAATTCTTTTAGATGCTAGCTTCCAGGTGTTAGCAGCAGTTATGGGT
AAAACGGACAACCAAGAAACTTATTTGCCATTGGAAATAAAACGACTACAAATTTATCGGAGTGGTAGTAATAGTTTG
TGGACTCAAGTAGAGATAGGTGCAACAGAAACTAATAAACAAACTTTGAGCGGTAAAGTTTGTTTATTGGATGAACAA
GGAATAGTAGTAGCAAGAGTTGAAGGTTTAACTTTATTACGTACTTCTCGCGAGGCTTTGTTGCGTAATATTGAACCA
AAATTTAATAATTGGTTATATCAAATCCATTGGCAAACCCAATCAATTTCACCCCATAACCAATCAATTGACTTAACA
AAATCAGGTAGCTGGTTATTGTTTTCCCCACCCACAGGTATAGGCAAACATCTGGTAGAATCCTTAGAACAACAAGGT
TGGCATTGTATATTAGTAACACCAGGGGAAAATTACCAGCAGTTAGAATCTCAACATTATCAAATCAACCCCAACCAT
CCTGAGGAATTCCTGCACCTATTGCAATCAAGCTTGGAGCAGCAACCCCCATTACGAGGAATTATTCACCTGTGGAGT
TTGGACTCAACAATAGCACTAAGGACTGGGGCACAGGAGTTGCAAAAATCCCAAGAACTGGGCTGTGGCAGCGTACTT
CATTTAGTCCAAGCCTTAGTAAAAAATCAAGATATGGAAAGTGCCCCATTATGGTTAGTGACTCAAGGCTCACAATCT
GTGGGTAATGAGTCCCTTCCTATACAATTCCAACAAACACCTTTATGGGGGTTAGGTCGAGTAATTGCCCAGGAACAT
AGGGAATTACAATGCCGGTGTTTAGACTTAGATCCAACTATGGAAGATTCCCAAACAGTAGCTGCTTTGTTAGAGGAA
CTATTATCTCCTGGTGATGAAAACCAAATTGCTTACTGTCAAGGGGTACGTCACGTTGCCCGGTTAGAGCGGCAACAA
AAAATGAGTACATCTACACAGTCCGGATTACAAATTTCCTCGCAACAACCATTTCAACTGAAGCTATCAGAATATAAG
TCTTTAGACAACCTAATCCAAGCCGAAGCCAGTTACTTAATTACCGGAGGTCTGGGAGCACTGGGGTTAAAAACCGCT
GAGTGGATGGTACAACAAGGGGTCAAATATTTAGTACTTACCGGACGTAGGCAGCCATCAGCAAAAGCTCAACAAACC
ATTGAACAATTACAGAAGGCAGGAGCGCAAGTATTAGTCCTGTGTGGAGATATTTCCCAACAAGAAAATGTGGCAAGA
ATTATAGAGTCAATCAAAGTATCTTTGCCAGCATTACGAGGAATAATTCATGCTGCTGGGATATTGGATGATGGTTTG
CTGTTAAACATGAATTGGGAAAAATTTACACAGGTGATGGCACCAAAAGTACAAGGGGCTTGGCATTTGCATAATTTG
ACTCAGAATCTACCTTTGGACTTTTTTGTTTGTTTTTCCTCTATGGCTTCAATATTGGGTTCGCCTGGTCAAGGGAAT
TATGCTGCTGCTAATGCTTTCATGGATGGTTTAGCCCATCATCGACGGGGTATGGGTTTACCTGGCTTGAGCATTAAC
TGGGGACCATGGGCACAAGAGGGAATGGCAGCAAATTTGGATAGTCCTCATCAAGATAGAATGGTGTCCAAGGGAATG
ACTTTTTTGTCTTCAGAACAGGGATTGCAGGTTCTAGGACAATTACTCGAACAATCCATACCACAAGTAGGAGTCCTA
CCAATTCAATGGTCAGTGTTCCAAGAGCAATTTAGTTTTGGTAATCAAATACCATTGCTGTCCCAATTGGTAAAAGAA
AGCAAATCACAGCAAAAAGCCCTCAAAACAAAGACAAAGCACAATGAATTTTTAGAACAGCTAAAAGCTGCTTTACCA
AGAGAAAGAGAAAAGCTTTTGATAATTTACATTAAAGATGAAATTTCTCAAGTACTTTCTTTGAGCACTTCTCAAATT
GATATGCAACAGCCCCTGAACACTATGGGGCTTGATTCTCTAATGGCTGTGGAATTGCACAATAGGCTTCAAACTGAC
TTGCTCGTGGATATATCTATAGTCAAATTTATAGAAGATATCAGTATCGTTGATTTAGCCACTGAAGTGAATGAGCAA
CTGAGCCAAGTTGCTCAGAATCAAGGAGTTGAGTCAGAAAATAATGGGCAACTCTACCAAAGCAATAGGAAAGAAAAC
GAGCGGATAAGAGGTGAATTATGA (SEQ ID NO:2)

FIG. 6-2

CrpA

MGHSAGEYVAATVAGIFSLEDGLKLIAHRGRLMQQLPSGGEMLSVMASIEKVNQLIAPYSQKVAIASINGPQSIVISG
EAEAIGAVQNSLEAEDIKTKRLQVSHAFHSHLMEPMLADFEAVASEITYNQPNIPLVSNVTGARAENSIATASYWVNH
VRQPVKFAQSMDTLQQEGYSIFLEIGPKPTLLGMGRQCLPEDVGVWLPSLKPGQEDWQQMLQSLAELYVHGVKVDWLG
FDKDYSRSKVVLPTYPFQRQRYWIETNNNLIHQKQFLSNHKNLHPLLGQRLHLAALEQQIRFECQISASQPTYLQHHC
VFSQPVFPAAAYLEIALAAGSILFNSDDLILEDIAIQKVLILSKDEINTIQIVLNLQLVQSYKFQIFSLDINTNSSEP
KWILHIEGKILVGNKDPQLETTNLKAIKDEYNQQILPTEFYQKFEEWGLNYGSSFQAVKQLWHSEGKALGEIQLPETE
VNVATLYQLHPILLDASFQVLAAVMGKTDNQETYLPLEIKRLQIYRSGSNSLWTQVEIGATETNKQTLSGKVCLLDEQ
GIVVARVEGLTLLRTSREALLRNIEPKFNNWLYQIHWQTQSISPHNQSIDLTKSGSWLLFSPPTGIGKHLVESLEQQG
WHCILVTPGENYQQLESQHYQINPNHPEEFLHLLQSSLEQQPPLRGIIHLWSLDSTIALRTGAQELQKSQELGCGSVL
HLVQALVKNQDMESAPLWLVTQGSQSVGNESLPIQFQQTPLWGLGRVIAQEHRELQCRCLDLDPTMEDSQTVAALLEE
LLSPGDENQIAYCQGVRHVARLERQQKMSTSTQSGLQISSQQPFQLKLSEYKSLDNLIQAEASYLITGGLGALGLKTA
EWMVQQGVKYLVLTGRRQPSAKAQQTIEQLQKAGAQVLVLCGDISQQENVARIIESIKVSLPALRGIIHAAGILDDGL
LLNMNWEKFTQVMAPKVQGAWHLHNLTQNLPLDFFVCFSSMASILGSPGQGNYAAANAFMDGLAHHRRGMGLPGLSIN
WGPWAQEGMAANLDSPHQDRMVSKGMTFLSSEQGLQVLGQLLEQSIPQVGVLPIQWSVFQEQFSFGNQIPLLSQLVKE
SKSQQKALKTKTKHNEFLEQLKAALPREREKLLIIYIKDEISQVLSLSTSQIDMQQPLNTMGLDSLMAVELHNRLQTD
LLVDISIVKFIEDISIVDLATEVNEQLSQVAQNQGVESENNGQLYQSNRKENERIRGEL (SEQ ID NO:3)

crpB

ATGAATTTAATCGAATTTTTACAAGATATTTCAATAAAAGGTTGGCAAGTGTGGAGTGAGGGCGAACGGCTTTGTTAC
GATGCTCCTCAAGAAGAATCAACTGCTTTAGTATTAGCTCAACTGAAACAATACAAAACCGAAATATTACAATTGCTG
CGCGATCGCCCAGATATTTTAAACGTCTATCCCCTTTCTTACGGTCAACGGGCACTTTGGTTTTTGTGGCAATTAGCA
CCAGAAAGTCATGCTTATAACGTATCATTTGTCGCTCGTATTTGCTCAACAGTAGATATTACAGCAATGCAAAAGGCA
TTTGAGAAACTAATAGAACGGCATCCGATATTGCGTACAAATTATCCCAAACTCGGATCTGAGTCAATCCAACAGGTA
AATAATTTTCAGGAATTAAATTTCTTGCAAATTGATGCTTCTGCTTGGAGTGAAGATGAACTGAAAGGGAAAGTGATT
GAGAGCCATCAACAATATTTTGACCTCGAACGAGGGCCTGTTATGCGAGTTCGGTGGTTTACTCGTTCCAAGAAAGAA
CATGTCCTGTTGCTAACCATACACCACATTGCTTGCGATGCCTGGTCTTTAGATATGTTAATACAGGAGTTGCCACAA
CTCTACCAAGTACAATGGGCTGATTTTAAGACACCCCTTTCTCCTCTAAAGCATTCTTACCAAGATTACGTTCGTTGG
CAGAGGAATATTTTGCAGGAGACTGAAGGGGAAAGACTCTGGAACTACTGGCAGCAAAAACTGACAGGCGATTTGCCA
GCACTAAACCTAGCAACCTCAAGACAGCGACCACCGATAAAAACTTATAATGGTGCTTCCCATCACTTCAAATTATCT
GACAAGCTCACTAAGCAACTCAAGGAGCTGGCTTTAAACTCGGGAGCAACACTTTACATGATGCTCTTAGCAACTTTT
CAGGTGTTTTTGTATCGTTACACAGGCCAGGAGGATATTTTAGTCGGTTCTCCCACCTCAGGTAGGAGTCAAGCTAAG
TTTGCTTCGATCTTGGGCTACTTTGTTGATCCTGTTGTTATGCGAGCAAATTTATCGGGAAATCCCAGTTTCAAAGAT
TTCCTCGCCCAAGTACGCCAAACCGTATTGGAAGCACTTGCTCATCAAGATTACCCATTTGCTCTATTGGTAGAAAAG
CTACAGCCACACCGCGACCCCAGTCGTTCGCCGATTTTTCAGGCTTCTTTTTCTCTACTGCAGTTCCAAAAATCTCAA
GATATACAGAAGTTGTTTGTAAATCAAATAGAAACCTATGTTGATTGGGGAGAATTGAAGATAAAACCTTATGAAATA
CCTCAACAGGAAGGTCAGTTAGATTTAGGTTTAGAAATGGTGGAGGGGAGTTCATCTGTTTTTGGGGTTTTTAAGTAC
AACACTGACTTGTTTGATGAGTCAACCATTGAGCGCATGGCTGGTCATTTCCAGAATTTATTGTCAGCGATTGTGGAA
AATCCTCAACAAAAGGTAAGTGAATCACCCCTATTGAGTGAAGTAGAACGGCATCAGTTGTTGGTGGAGTGGAATGAT
ACAGCAAGGGAATATCCCAGTAAATGTATCCATCAATTGTTTGAAGAGCAGGTAGAGAGAACACCGGATGCAGTCGCG
GTGGTATTTGAGAATCAGCAGTTAACCTACCAACAATTAAATCAAAAAGCCAACCAACTAGCACATCACCTGCTTTCC
TTGCGAGTCGAACCAGAGGTGCTGGTAGGGATTTACGTGGAGCGTTCTTTTGAAATGGTGGTAGGGCTCTTGGGGATA
CTCAAGGCTGGTGGGGCTTATGTACCCCTTGACCCCAATTATCCCCAAGAACGGTTGAGTTATATGTTGGCGGATTCA
GGTGTTGAGGTGTTGTTGGCTCAAAAGTCCCTACTGGAATCTTTGCCGTCACATACTGCACAGGTGGTTTGTTTGGAT
AGTGATTGGGGAGTGATTGAGCAACACAGTCAGGAGAATCTTGATGTTGGGGTATGTTCAGATAATTTGGCTTATGTG
ATTTATACTTCTGGTTCTACTGGAGTTCCCAAGGGGGTTGGGATTGAACATTTTTCTTTGTGCAACCTTATTCAAGCA
CAGAAAAATTTATTTTATCTAGAACCAAATAGTCGTGTTCTTCAATTTGCTTCTATAAGTTTTGATGCTTCAGTTTCG
GAAATATTTATCGCTTTGACTTCTGGAGCAATGCTAATTTTGGCTATAGCTTCTGAGTTAATCCCGGGTTCGGATTTA
AAGCAAATTTTACAAGAAAGATGTGTAACTCATGTTACGTTACCTCCCTCTGCCCTGGCAGTACTAGCTACTGATGAA
TTT

FIG. 6-3

CCAGCCTTGGGTCAGATAATTGTGGCAGGGGAWSCTTSTAYWMTKGAATTGGCCAATCAGTGGTCTGTTGGTCGTCGT
TTGTTTAATGGTTATGGGCCTACTGAGTCTACTATTGGTGCTGCAGTGGCTCAAATCAGCCATGGTAGCGAGAAAGTT
ACTATCGGTCGCCCCATTGCAAACACCCAAATCTATATCTTAGATAAGCACTTAGAACCAGTACCCATAAGTGTATCC
GGAGAATTATACATCGGAGGTTATGGCTTAGCCAGAGGTTACCTCAACCGACCGGAATTAACTTTAGAGAAATTCATC
CCTAACCCCTTCAATAGCAGATCAAAACTATATAAAACTGGAGATTTAGCTCGATACTTACCAGATGGTAATATTGAG
TTTCTTGGTCGTCTTGATAACCAGGTGAAAATACGTGGTTTCCGCATAGAACTAGGAGAAATAGAATCAGTTCTGAGT
ACCCATCCTCAAGTACAGCAAGTCGCAGTCACCGAAAGGGAAGATACTCCGGGTCACAAACGGTTGGTAGCATATTTA
GTTCCTCTTCAGTCAAAAGAAAAAGATCCCCAGCCCCAGACAGGAATTGAACTTTGGCCCTCAGTAGCAGAATTCTAC
GTTTATGATGAGCTCTTGTACTATGCGATGACCAATGATCATCGTCGAAACCAGAGTTACCAAGTCGCAATTAATCAA
ATGGTTAAAGATAAAGTAGTTGTTGAAATTGGCACGGGCAAGGATGCAATTATAGCCAGATTTTGTGCAGAAGCAGGT
GCTAAGAAAGTCTACGCAATTGAGAGAGACGAGCAAACCAGTAAGTTAGCTTCAGCTTGCGTGCAAGAGTTGGGGTTA
TCAGAACAAATTCAAATCATACATGGAGACGCTACTACAGCCAACTTACCAGAAGAAGTTGATGTATGTGTTTCTGAA
ATTGTGGGACCCATTGGTGGATCTGAAGGAGCAGCAGTAATTATCAACAACGCCAGAAGATTTCTCAAATCAGATGGT
GTAATGATTCCCCAAAGAAGTGTGACTCAAATTATTGCAGTAACTCTTCCTGATGAATTACTAAATCAACCCCAATTT
ACAAAAGTTTCAGGTTATTATACCCAGAAGATATTTGAGCAAGTTGGATATCCTTTTGATTTACGAGTATGTATTAAA
GGATTAAATCAAGTAAACTGGTTGTCCAATCGAGGAGTTTTTGAGGATTTAGACTTTAGCAAGCTTGTTAGCACAGAA
TCTACTCACCAAATTAAATTAACTATTGAAAAATCAGGAAGATTAGATGGTTTTTCAGTGGGATTAAACTTACACACA
ATTGAAGGAGAATGTATAGATATTTTAGAAAATGAACATTGCTGGTTACCAGTTTATCTTCCAGTGTTTGAACCAGGA
ATTTATGTAAATGAAGGAGATATCATCGAAGCTGTTTGTACGAGAACGCTTTGTGAAAATAATTTGAATCCAGATTAT
ACTCTTCGAGGTCGTCTTTTAAAAAAAACTGGAAAAGAGATTGAGTTTGAATACATCTCATATCATTGGAAACAATTG
TTTAAGCAAAACCCTTTTTATCAACGATTGTTTGCTGAAAATAATCTAGAAGATTACACAATTAATAAGTCAAAGCCT
GAACGTAAGGTACTGAGTAGCAATGAATTGCGCTCCTACCTAAAGTCTAAATTGCCAGAATATATGCTCCCCTCTAGC
TTTGTCATCTTAGACACCCTACCGTTGACACCCAACGGTAAAGTAGACCGCAAAGCCCTAAGTGCACCTGATGGGGAA
GTTAGCCGAGAGGATGAATATGTCGCACCACGCACAGAAATAGAACAAATCTTAACCAACATTTGGCAWGAACTGCTC
CTTAAAGAACAAGTCAGCATCCATGACAACTTCTTTGAAATTGGTGGCGACTCCATCCTTGGTATTCAAGTAGTTTCT
CGTGCCAAAAACTTAGGAATACAAATCACTCCCAAACAAATATTCCAAAATCAAACCATCGCCAAACTAGCATTAGTA
GCCAATACAACAGTTACTGTCAGTGCTAACCAAGGTATAGTTACTGGAGTTGCACCCCTAACACCAATTCAACAGTGG
TTCTTTGCACAAAATAGCCAAGAAGCACACCATTACAACCAATCAGTTTTATTGCAGATTCCCAATCATCTGCAAACT
GAATTAATCGAAACAGCCTTGAAAAAATTATTAGAGCATCACGATGCTCTGCGTTTACGATTCACATCAGTTGCATCT
GAGTACAAACAAATAAACCATGGCTTTGATGATCCCGTAGCATTTACTGTAGTTGATTTATCATCAACTCCTGTCATT
GAACAACCACAAGCTTTATCACAGATCGCCACGGAATATCAAGCAAGTTTAAACCTCTCAGAGGGACCTTTAATGCAA
GTGGTGATGTTTAACTTAGGTAGTGAAGTTGATGCCCGTTTACTGATTATTATTCATCACCTAGCAGTAGATGGTGTG
AGTTGGCGAATTTTACTATCAGACTTAGAAACAATCTATCAACAACTAATCGCTCAACAATCAATACAGCTAAATGCG
AAAACAACAGCATTTATTGATTGGGCAGAGAAATTGAAAAATTATGCACAATCAGAAAAAATCAAACAAGAGTTAGAC
TATTGGCTCAACCAACCTTGGTCAGAAACAACACCACTACCATTAGATTCTGCTCACACTCAAGCAGAAAAAACAGTT
GATAGTGCGATTAATTATAGAGTGAAATTGAGTCCAGAAGAAACCCGCGCTTTGCTGGGGTCAGTAAACTCAGCTTAT
AACACACAAATTAACGATATCCTCCTCAGTGCATTAGTAGTTTCCTTGGCAGAGTGGACGGGAGATTCAAAAGTACTA
ATTGACCTAGAAGGACATGGCAGAGAAGAACTATTTTCAGATGTAGACTTATCAAGAACAATAGGTTGGTTTACCAGT
TTATTCCCAGTATTATTGCGATTACCAGACGATAAACAACCAGCAGAAGTTATCAAGTCAATTAAAGAACAATTACGA
GAGATTCCCCATCGTGGTATTGGCTTTGGTATATTGCGTTACTTGTGTGAAGATACTACTGTAAC
CCAAAAACTACAGACAATTCCTACTCCAGAAATTAGTTTTAACTACCTAGGACAATTTGACCAAATACAATCGGAAAC
GGGTTGGAAATTTGCGCCAGAATCTACTGGAGATAATCATAGTTCAAAGCAAACTCGTCACCATCTATTAGAGATTAA
TAGTCTGGTGGTAGAAGGTGAATTACAAATTGATTGGACTTATAGTAGTAATTTTCATACTCATGATACAGTAAAAAA
TTTGACACAAAGCTATATTCAAGCAATTAAGTCAATAATAGAACATTGCCAGTCAGAAAATGGTTTTGGATATACAGC
TAGTGATTTCCCAGATGCACAGTTAAATCAATTAGAACTTGATGAGTTATTAAAAAACTTGGGCAAATAG (SEQ ID
NO:4)

CrpB

MNLIEFLQDISIKGWQVWSEGERLCYDAPQEESTALVLAQLKQYKTEILQLLRDRPDILNVYPLSYGQRALWFLWQLA
PESHAYNVSFVARICSTVDITAMQKAFEKLIERHPILRTNYPKLGSESIQQVNNFQELNFLQID

FIG. 6-4

ASAWSEDELKGKVIESHQQYFDLERGPVMRVRWFTRSKKEHVLLLTIHHIACDAWSLDMLIQELPQLYQVQWADFKTP
LSPLKHSYQDYVRWQRNILQETEGERLWNYWQQKLTGDLPALNLATSRQRPPIKTYNGASHHFKLSDKLTKQLKELAL
NSGATLYMMLLATFQVFLYRYTGQEDILVGSPTSGRSQAKFASILGYFVDPVVMRANLSGNPSFKDFLAQVRQTVLEA
LAHQDYPFALLVEKLQPHRDPSRSPIFQASFSLLQFQKSQDIQKLFVNQIETYVDWGELKIKPYEIPQQEGQLDLGLE
MVEGSSSVFGVFKYNTDLFDESTIERMAGHFQNLLSAIVENPQQKVSESPLLSEVERHQLLVEWNDTAREYPSKCIHQ
LFEEQVERTPDAVAVVFENQQLTYQQLNQKANQLAHHLLSLRVEPEVLVGIYVERSFEMVVGLLGILKAGGAYVPLDP
NYPQERLSYMLADSGVEVLLAQKSLLESLPSHTAQVVCLDSDWGVIEQHSQENLDVGVCSDNLAYVIYTSGSTGVPKG
VGIEHFSLCNLIQAQKNLFYLEPNSRVLQFASISFDASVSEIFIALTSGAMLILAIASELIPGSDLKQILQERCVTHV
TLPPSALAVLATDEFPALGQIIVAGXXXXXELANQWSVGRRLFNGYGPTESTIGAAVAQISHGSEKVTIGRPIANTQI
YILDKHLEPVPISVSGELYIGGYGLARGYLNRPELTLEKFIPNPFNSRSKLYKTGDLARYLPDGNIEFLGRLDNQVKI
RGFRIELGEIESVLSTHPQVQQVAVTEREDTPGHKRLVAYLVPLQSKEKDPQPQTGIELWPSVAEFYVYDELLYYAMT
NDHRRNQSYQVAINQMVKDKVVVEIGTGKDAIIARFCAEAGAKKVYAIERDEQTSKLASACVQELGLSEQIQIIHGDA
TTANLPEEVDVCVSEIVGPIGGSEGAAVIINNARRFLKSDGVMIPQRSVTQIIAVTLPDELLNQPQFTKVSGYYTQKI
FEQVGYPFDLRVCIKGLNQVNWLSNRGVFEDLDFSKLVSTESTHQIKLTIEKSGRLDGFSVGLNLHTIEGECIDILEN
EHCWLPVYLPVFEPGIYVNEGDIIEAVCTRTLCENNLNPDYTLRGRLLKKTGKEIEFEYISYHWKQLFKQNPFYQRLF
AENNLEDYTINKSKPERKVLSSNELRSYLKSKLPEYMLPSSFVILDTLPLTPNGKVDRKALSAPDGEVSREDEYVAPR
TEIEQILTNIWXELLLKEQVSIHDNFFEIGGDSILGIQVVSRAKNLGIQITPKQIFQNQTIAKLALVANTTVTVSANQ
GIVTGVAPLTPIQQWFFAQNSQEAHHYNQSVLLQIPNHLQTELIETALKKLLEHHDALRLRFTSVASEYKQINHGFDD
PVAFTVVDLSSTPVIEQPQALSQIATEYQASLNLSEGPLMQVVMFNLGSEVDARLLIIIHHLAVDGVSWRILLSDLET
IYQQLIAQQSIQLNAKTTAFIDWAEKLKNYAQSEKIKQELDYWLNQPWSETTPLPLDSAHTQAEKTVDSAINYRVKLS
PEETRALLGSVNSAYNTQINDILLSALVVSLAEWTGDSKVLIDLEGHGREELFSDVDLSRTIGWFTSLFPVLLRLPDD
KQPAEVIKSIKEQLREIPHRGIGFGILRYLCEDTTVTQKLQTIPTPEISFNYLGQFDQIQSETGWKFAPESTGDNHSS
KQTRHHLLEINSLVVEGELQIDWTYSSNFHTHDTVKNLTQSYIQAIKSIIEHCQSENGFGYTASDFPDAQLNQLELDE
LLKNLGK (SEQ ID NO:5)

crpC

ATGAGCAACGAAGAAATTAGAAGAAATATCTCTTCAATTTATCCACTTTCTCCCATGCAACAAGGGATGCTGTTCCAC
AGTCTTTATGCACCTTATAGTGGGGTATATCTTGAACAGATGACCTGGGGTTTGAAGGGGAATATCAATGTTGCTGCT
TTTGAAAGAGCTTGGCAAAAAGTTTTAGATAGACATTCAATTCTACGTACATTTTTTGTTTGGGAAAATCGCCAAACT
CCATTACAAGTAGTACTAAAACAGGTTAATGTTCCTTGGAATACTCTTGATTGGCGAGAACTTTCTTCTAATGATCAA
CAACAACAATTAAAACAATTATTGCAAACACAAAGAGAACAAGGTTTTAACTTATCCCAAGCACCATTAATGCGGTGT
ACGTTAGTCAGGCTAGGCGAAGATAATTACAAATTTATCTGGAGTCATCACCACATCCTTATGGATGGATGGTGTTTA
TCAATTATTTTTAAAGAAATTTTAATTTTCTATAAAGCACATCTGCTTGGTGAAAATTGCCAATTGCCAAAACCACGT
CCTTACCAGGATTATATTGCTTGGTTGAATTCTCAAGACAAATCAGCAGCAATTGAGTTTTGGCAACAAACTTTACAA
GGTTTTAGTGCTCCCACTCCATTGGTAATGGATAAAACTCAATTTCTGAAAGAGCAACAGTATAAAACTGCGGATTAT
CAGGAGAGAACAAGTAGTTTATCCCCTGAATGCACTCAGAAGTTACTTCATATAGCACAACAACATCATGTGACTTTA
TCAACTGTAGTACAAGCTGCTTGGGCTTTACTATTGAGTCGTTATAGTGGTGAGAAAGATGTAGTATTTGGTGTGACT
GTTTCTGGTCGTCCTCCTAGCCTCTCTGAGATAGAAAATATGGTAGGACTGTTTATTAACACCCTTCCCTTACGAGTA
CAAGTATCCACCCAGGAGCAACTCATACCTTGGTTGCAAAAAATACAACAGTCAATGGTTGAATTACAAGAGTATTTT
TATACTCCTCTTGTTGATATTCAAGCTACTTCTGAGATACCAGGTGGAATACCTTTGTTTGAGAGCATTGTGGTGTTT
GAGAATTATCCAATTGATAATTCTTTGTTGAATGAAGAAGGTTCATTACACTTAGGTGATATAGAGGTTTTTGAACAA
ACTAATTATCCACTAACTTTAGTTGCAGTTCCTGGGGATAAGTTGTCAGTTAGGATTAGTTACGATACTGCTCGTTTC
TCTTCAAATACTATTGAGTGGATTTTAGGATATCTGCAAACTGTTTTATCAGCAATTGCAATTGTGGAAAATCCTTCA
CATAAGGTAGCTCAATTACCTTTATTAAGTGAAGTAGAACGTCATCAGTTATTGGTTGAGTGGAATAACACTGCAACG
GATTACCCATCTGATAAATGTATTCATCAGTTATTTGAGCAACAGGTAGAAAAAAACCCCAACTCCATAGCAGTGGTG
TTTGAAGAAGAACAATCAACCTACCAACAATTAAATCAAAAAGCGAACCAATTAGCACATTATCTACAAACTCTGGGA
GTGAAACCAGAAGTGCTGGTGGGTATTTGCATAGAATACTCCATTGATATGATTGTAGGACTGTTGGGGATACTTAAA
GCTGGTGGTGTATATGTGCCGTTAGATCCGAACTATCCGCAAGAACGACTGGCGTTCATGCAGGAAGATTCCAATGTG
CATATCATATTGACCCAGCAGCCTCTGCTCGAAAAGATTTCCCCTCAAAATGCCCATATCGTTTGCCTGGACAGGGAT
AGGGATGTCATTGCTAGGGAAGGCGTTGAAAATCTGGATCGGCAAACAACAC

FIG. 6-5

TGGATGACCTTGCATACGCAATCTATACGTCTGGTTCTACTGGAAAACCCAAAGCCGTTCTCGGCACGCTTCGCGGCA
TTGTCAATCGCTTGCATTGGATCTGGGAAATGCTACCATTTGGGGCAGATGAGATTTGCTCTCAGAAAACATCCATCA
ATTTTGGCGATCATGTTGCGGAAATATTTTCTCCCCTTCTCAAAGGAATTCCCCTTGTGATCGTTCCAGATGATATAC
GGGGCAATATTCCCAGGCTAATGAGCCTGTTGAGCGATCGAAAGGTAACTAGAATTGTTCTCGTTCCATCGCTATTAA
AAGCGATACTGGAAAATGCGCCCCAACAACTGACAAAACTTCGATATCTCAAATATGTCTTTTGCAGCGGTGAAGTCT
TACCGCTAACCTTGGCTAAGGAATTTCACCAGAAAATCAGCTCTGCCAGATTGTTCAATCTCTACGGCTCTTCAGAAG
TTGCCGCTGATGTTACATGCTTTGAAGTCAAACTGAGAATCGCAAATCAAATTGAAGCAAAAAGTAAAGAGAAACTTG
ATGCTTTAAAAAATCTTCCTAGTGGCTCAGGGGATAGGGAAACTGCTGTCCTGCATAAAGAAATAATACATTTGCAGT
TGGCAGACGAGCGAAGAGCAGATTTAGGAGAAGCTCTAGAAGAATATCTGAAAAGAAATACGATTCCGATTGGAAAAC
CGATTTCAAACACACAAATTTACATCCTCGACAAGTATGGCGATCTTTTGCCACCTGGTGTTACGGGTGAGCTATACG
TCGGCGGAGATGGGCTTGCAAAAGGGTATTTAAATCTGCCCGAGTTAACGCGGGAAAAGTTTATCCCCAACCCGTTTG
TGAAGGACAGGGGGAAAAGTAAAAAGGCACAAGCAGAAAGATTGTTTAGGACTGGAGACCTAGCCCGCTGGCTGCCGG
ATGGTAATATCGAATTTGTAGGGCGTATCGATCACCAAGTGAAGGTGCGGGGCTTCCGCATTGAACTTGGAGAAATCG
AAGCAGTCCTCAGTACCCACCCCCAAATCCAACAAGTCGTTGTCATTGCCATAGAAGATATTCCAGGTAGCAAACGTT
TAGTAGCCTACATAGTCTGTGAGGATGAATCACTAAGTACCTATCACCTGCGTGAATTCCTCAAACAAAAGCTACCAG
AATACATGATGCCCAGTGCCTTTGTCATCTTAGACACCTTACCGTTGACACCCAGCGGTAAAATAGACCGTAAAGCCC
TTCCAGCACCTGATGGAGAAATTAGCCGAGAACATGAATATGTCCCACCACGTACATCGGGTGAAGAAATAATAGCCA
ACATCTTCGCTTCTATTCTAGGTGTGCAAAATGTTGGAATCCATGACAACTTCTTTGAATTGGGAGGACATTCCCTAC
TAGCAACCCGATTAATTTCCCGACTCAGAGTTGCCTTTGAAGTAGAAATAGAACTAAGTGCAGTCTTTTCCTCTCCCA
CTGTAGCTCAATTAGAGCAAACATTAACCCAATTACGTACTACTAATAGCGCATTAAGTCTTCCCCCCATTCAGCCAA
GAACACAGAACCAACAATTACCCCTATCTTTTGCACAAGACCGGTTGTGGTTCCTCAACCAACTTGAAGGGTCAAGTG
CCACTTATAACATGCCAGGAGCAATTCGTGTCACTGGAAAGTTGGATATTAATGCCTTGCAACAAGCATTATCAGAAA
TAGTCCGCCGTCATGAAGTACTACGCACCAGCTTCCGAACTGTGAATGGCACACCAATACAGGTAATTCACCCAGAAG
CCACCATGAACATCAGTGTGGCGGACTTACAGCAACTAGAAGCAACAGAACGGGAAAGTGTCCTTCACCAACAAGCAC
AACTTGCAGCAATTACCCCCTTTGACTTAGAAACTGCACCACTAATCAGGTGTAGTTTATTGCAGTTAGATGCCAGAG
AATATGTGTTATTACTGACGATGCACCACATTGTCTCTGATGGTTGGTCAATGGGGATATTCAGCCAAGAACTATCTA
CTTTATATCAAGCTTTTAGTGCAGGAAAACCATCCCCCTTGGCAGAATTACCAATCCAGTATGCAGACTTTGCAGTTT
GGCAAAGACAATGGTTAAGTGGAAAGGTACTAGAAACTCAACTCAATTACTGGCTTTCTCAGTTAGAGGGTGCACCAG
AATTGTTACAATTACCTACTGACCGTCCTCGTCCAACCGTGCAAACTTTCCGGGGTACTACTCAAAGTTTTAGTTTAA
ATACTGATTTAAAAGAGAAGTTGCAAACCCTGTCTCGGAACTCGGGTACTACCTTATTTATGACCCTGCACGCAGCGT
TTGCCACTTTACTCTATCGCTACAGCGGTCAATTAGATATTTTAATTGGTTCACCCATTGCCAATCGCAACTGCAGTG
AAATTGAGTCTTTGATTGGCTTTTTTGCCAATACTTTGGTATTGAAAACCCGTTTTGAAGATAATCCCAGTTTTGAGA
ATTTGCTGGCACAAGTTAGGGAAACTACACTTGAAGCTTATGAACATCAGGATGTGCCTTTTGAACAGGTAGTTGAAG
TACTACAACCACAACGCTCTTTGAGTTATGCACCCTTATTCCAGGTAATGTTTGTGTTGCAGAATGCACCCATGGGTG
AATTAGAATTACCTGGTGTGACCCTTAATTTATTGAGTTCTCAAACAGAAACAGCCCGGTTTGATTTAACAGTATCAA
TGCAGCAAACTTCCGAAGCACTAGTGGGTTCATGGGAATACAACACTGACTTATTTGATGGGTCAACTATTGAGCGCA
TGACTGCTCATTTCCAGAATCTGTGTAGCGCGATTGTAGAAAATCCCCAACAAAAGATAAGTGAATTACCATTATTCA
CAGATTCTGAGCAAGAGCAGGTACTGCACAGTTACAATAACATCGCTACAACTTACCTGCTGGATAAATATGTTCATT
TCCTGAGTTCAAATAATTTACAAATTTACATTTTAGATAACCATCAACAATTAGTTCCTTTGAGTGTAGAAGGAGAAA
TTTATTTGGGGAATTGCGATTTACTCCCAGACAAGTTACATCCAGAACCAGAAAAATTTATAAGTTTCATAGAACATA
CCCAACTGGGTAAGTTATTAAAAACAGGGGAATGGGGTTGTCGTCGAGTCGATGGTTCTCTGGAATTGCTAGGAAAAG
AGCATCGAATTGTCACAGTTAATGGACAACGAATTAACCTACAACGTATTGAACAAGCTTTACAAACAGCGAAAGGGG
TAGAAGATTGCTATGTAATGGTACGCAATCAAAAATTAGTCGCTTACGTAGTCAAAGATGGTTCTTGGGCTAGGGAGT
TTTTACACCATTATTTAAAATCTCAGTTACCTGGATACCCATTACCCTGCATCTATGTACCAGTATCTGCTTTACCAT
TGACAAGTTTTGGAGAAGTTGATGAAGTAGGTTTAGCTTCTATTAGCATAATTGATTCTGAGTTAATTAACACTTGGG
AAGAACAAATAGGTTCTCAGGCGGAAATTGATAAAGTTGCTGTTTTTATTGAGCCAAATGTAAAAACGATTTCTCCGA
TACATTTAGAAGAACTTTTACCATCAATCCAAGCTATTTTCAATCAAGGTTCTACTCCAGTTGAAACTCCCAGAACTG
CTAGGGGAAAAGAGAGTAGTTCCCTATTAGAAATAAAATCACCTGCCATCAGCCACGAAGAAGTATTAATCTTTCCAG
AATCATCTCCAGAAACTTTAGGGGAGATGCTGCAAAAAACTGCTGGGAAATTTCCTCACAAAGGAATCACTTATATTA
ACTCTGATGGTT

FIG. 6-6

CCGAACAAGTTCAATCATATGCCCAGTTATTAGAAGATGCTCAAAGAATTCTAGGTGGCTTCAGAAAACTGGGAATTA
AGCCACAAGATAAAGTTATTTTGCAATTAAAAGAAAATAAAGATTTTATTAGTGCTTTTTGGGGTTGTGTGTTGGGAG
GCTTTATTCCCGTACCCGTTGTAATTCCTGTAAGCTATGACCAGCCCAATGTCAATCTAAATAAATTACAAAATAGTT
GGCAGATGTTAGAAAGACCTTTGATTTTAACAGATAAAAAATCATTGTCAGAACTAAAGAAATGGTCTCAAAATCTAA
ATGACGACAACTTTAAGTTAGAAACTATTGAAAGTTTACAAAAGTTCTCAACAGATAAAGATTACTATAATGCCCAAC
CAGAAGATTTAGCACTGTTCATGCTTACTTCCGGTAGTACAGGTATGTCTAAGGTGGTACAGTTGAGCCATTTAAATC
TACTGAGTAGGACTATTGGTTCAATACAAATGAATAATTTACCCCAGAAGATATAACCTTAAATTGGATGCCCTTAG
ACCATGTTGCAGGTTTAATATATTTTCATATCCGGGATATTTATTTAGGATGTAAACAAATTCATGCTACTAGTCAAT
TAGTGATTGAAAAACCTTTAAGATGGTTGGATTGGATTGATACTTTTGGTGTCACTGTTACTTTTGCTCCTAACTTTG
CTTATAGTTTAATTAATGATTTTGTTCAAGAAATAGAAAAGCAGAATTGGAATTTATCTTCTATTCGCTTGATGTTAA
ATGGTGCGGAACAAATTGTTGCAGCAACAGCAAGACGTTTTTGAAATTACTTGCTCCCTTTGGCTTACCTGGGGATG
CTATGACTCCATCTTGGGGAATGGCTGAGGTTTCCTCTGGTATTACTTATTCTGACAATTTTCACTCTTATCAAGTT
CAGATGATAATTCCTTTGTAAATCTTGGAAAACCGATTAGGGGTACTTGTCTGAGAATAGTCAATCAAGACATGGAAG
TATTATCAGAAGGTGAAATTGGTTTACTTCAGGTCAAAGGATTAACCGTTACTTCTGGTTATTATCAAAATCCAAAAG
CAAATAAGGAAGCATTTACCGAAGATGGTTGGTTTAATACAGGTGATTTAGGATTTATAAAAGATGGATGCTTAACGA
TTACAGGACGACAAAAAGATATCATTATTATTAATGGAGTTAATTATTATAGTCATGAAATAGAAGCTGTTGTTGAAG
AATTAGGAGAGGTTGAAGTTTCTTATACCGCAGCCTGTGGAGTCTGCGTTGCTAGCAATAATACCGAAGAATTAGTAA
TCTTTTTCACTCCGTATGTATCTGAGAAGAATCAATTATTAGAGCTTTTGAAAAAGGTTAGGGAACAAGTTATAAAAT
ACTGCGGGATAAATCCAAGTTATTTAATACCCATAGATAAAGAACTGATTCCCAAAACTTCCATCGGTAAAATTCAAC
GTTCCCTCCTTAAGCAACGTTTTGAATGTGGTGAGTTTAAATCTCTCAGACAGCGTGTAGACTTGTTGCTTGATAATA
CTAATACTATTCCCAACTGGTTTTACCGTAAAGTATGGCAAATTAAAGAAAGTAAAAATACTTTACTCAATTATTCTT
CTCAGAAAACTTTAACCCTAATATTTACAGATAATTTGGGTTGGCAACAAGATAACCGAGGAATGTCCCAAACTGTTC
AACCATATGCTCAAGTTACTATTGGTTCAAATTTTGCTCAAATTAGCCCAAATCATTATTCTGTTGTTCCTGGAAATC
CACAACACTATCGCTTGTTAATTGATTCTTTGAGGCAAAATAGCCAAGTAATTAGTCAAATTCTTCATCTTTGGAACT
ACAACGAGCAGACTGAAAAAATTTCTAGCTTGGAAAATTTAGAGTCCACTCAACAACAAGGAATTTACAGTTTACTAT
TTTTAGTACAAGCTTTAGAAGAAATTCAAGGCAAACAGCAAGCAGTCAAATTATTATGGATTGCTAATCAAAGCCAAT
TAGTTCATCCCACAGATAAAATTCAACCCGAAAAATCCACTGTTTTAGGCTTACTTAAAACTGTTAGTCAAGAAATGC
CTTGGTTAACTACTCGTCATTTAGATTTACCATTAGCACCAGAACTCAACAATAGTTATATTTGGCAAGAACTGTATT
CTGCTGATAAAGAATTGGAAGTTGCTATACGCAATAGAGAACGTTTTGTGTCTGGTCTGGAACCAGTAGATATGACTG
CTAAGGAAAAACAAAAAATTCCGATTCTACCAGGAGGAACGTATCTACTTACAGGAGGGCTTGGAGGAATTGGGACTG
TTATTGCAAAGTACTTATTAGAACATTATCAAGCAAATTTAATATTAGTTGGTAGAACTCAAATTGAAGATAATAATG
AGGAAGCTAGCACAAAATTGCAGAGGTATCAAGAATTAGAAAAACTACCAGGTTCAATAATTTATCAAACTGTAGATA
TTTGTGATTTAGTAGGTTTACAACAGGTAGTAGAAAAAGCAACACAAGAATGGAGGACTCAACTTGATGGGGTATTTC
ATATGGCTGGGATTATTCAGGAAACGCCAATCGAGAAAGAAACCCCAGGAAATATCGCTGCTGTTTTACGTCCTAAAG
TTAGCGGTACTTGGGTATTGCATCAATTGCTCAAGGATAAAGAAAATGCTTTATTTGTCCACTTTTGTTCTGTAAATG
GTTTCTTTGGAGGAACCAATGTTGCAGCTTATAGTGCAGCAAATAGTTTTCAGTCAGCATGGAGCGATTATCAACAAC
AAAACGGTTTCCAAAGCTATTGCTGCTCTTGGAGTATGTGGAATGAAACCGGAATAAGTCATGGCTATCAATTCCAAG
AACTCAGTCGTGCTAAGGGCTATTTTATTATTACTCCTCAACAAGGATTTTACTCATTTTTAGCAGCTTTATCTGGTT
CGGAACATAATCTATTAATCGGATTGGATGGAACTAAAACAAATGTTGAACATTTGATTCGTGATTGTCAGCCCAAGC
AGAAATTAACTGCTTACTTCACCTCTCCCACACCAGAACTTGCTGCACTCTCCTTACAAGAGTTACAACTACACGATC
GCTTTGGGATACCCAATCAAATTAACTTTGTCCAACTTGAACAAATACCCCTTACTCAAAGAGGAGAAATTAATCGGG
AACAAATTGCTGCTATATATGGAGGTTTGAATACTTCTGAGCAGACAAAACCACGGAATCAAACAGAACGTCAGTTAG
TTGAGATTTTCCAAGAAGTTCTCAATCTACCCTCTATTGGTATTCATGACAACTTCTTTAGCTTAGGAGGACATTCCC
TTCTAGCTGTCCGTCTAATGTCCGAGATTCAACAACAATTCCAGAAAAATTTACCTTTAGCCACTCTTTTTCAAAATC
CCACCATTGAACGACTAGCACTTCTTGTTGGTTCCGATTCCGGAGCCGAACTTTGGTCTCCATTAGTACCAATTCAAC
AAAACGGTTCATTACCACCTTTGTTCTGTGTACCAGGAGCAGGTGGAAATGTTCTCTACTTCCACCACTTAGCACAAT
ATCTTGGAAATAATCAACCGTTATACGGTTTACAAGCACAAGGTCTTGATGGTGAAACCGAACCTCATAAAAGTGTTG
AAGAAATAGCCTCCCAACACATTAAAGCAATTCAAACAGTTCAACCAGTTGGTCCTTACTTCTTGGCTGGTCATTCCT
TTGGCAGTCATGTAGTATTTGAAATGGCGAATCAACTACAACTTATTGGAAAGTCTGTTGCTTATGTTGGAATTTTAG
ATACTCCTGCAC

FIG. 6-7

CAACTTCTCAAGCTAATCATCAGAATGATTTTTCTAACTGGGATAATGCAAAGTGGATATGTCGAATGGCTGAGGTTA
TTGAAGATATTGTTGGAGAAAATCTATTTTTATCTTATGAAACTCTAACTTCTCTAACTTGGGAGCAACAATTAAATT
ATTTCAAGCAAAAGTTAGAAATAGTTGGTTTTTTGCCTGCTCAAACAGATATCAAAATTGTTCGTGGTTTATTACAAG
TTTTCCAAACTCAATGTCAAATTAAGTATGAACCGGAAAAGACTTATAAAACTCCAATCACTTTGTTTTGTGCGAGGG
AGATAAATCCAGAGCAAGAAAGTTATTCTCACATTTTCCAAGAGCCAACATGGGGTTGGAATCAGTTTTCTGATGGAG
AAGTGGAAATCCATATAGTTCCGGGTAATCATGTTTCAATGCTGAGTGAGCCTCATGTCAAGGTATTGGCTCAACAAA
TGCAAATATCTCTTGAACAAGCACAGAAAACCCATCAATTGGAAAAATGA (SEQ ID NO:6)

CrpC

MSNEEIRRNISSIYPLSPMQQGMLFHSLYAPYSGVYLEQMTWGLKGNINVAAFERAWQKVLDRHSILRTFFVWENRQT
PLQVVLKQVNVPWNTLDWRELSSNDQQQQLKQLLQTQREQGFNLSQAPLMRCTLVRLGEDNYKFIWSHHHILMDGWCL
SIIFKEILIFYKAHLLGENCQLPKPRPYQDYIAWLNSQDKSAAIEFWQQTLQGFSAPTPLVMDKTQFLKEQQYKTADY
QERTSSLSPECTQKLLHIAQQHHVTLSTVVQAAWALLLSRYSGEKDVVFGVTVSGRPPSLSEIENMVGLFINTLPLRV
QVSTQEQLIPWLQKIQQSMVELQEYFYTPLVDIQATSEIPGGIPLFESIVVFENYPIDNSLLNEEGSLHLGDIEVFEQ
TNYPLTLVAVPGDKLSVRISYDTARFSSNTIEWILGYLQTVLSAIAIVENPSHKVAQLPLLSEVERHQLLVEWNNTAT
DYPSDKCIHQLFEQQVEKNPNSIAVVFEEEQSTYQQLNQKANQLAHYLQTLGVKPEVLVGICIEYSIDMIVGLLGILK
AGGVYVPLDPNYPQERLAFMQEDSNVHIILTQQPLLEKISPQNAHIVCLDRDRDVIAREGVENLDRQTTLDDLAYAIY
TSGSTGKPKAVLGTLRGIVNRLHWIWEMLPFGADEICSQKTSINFGDHVAEIFSPLLKGIPLVIVPDDIRGNIPRLMS
LLSDRKVTRIVLVPSLLKAILENAPQQLTKLRYLKYVFCSGEVLPLTLAKEFHQKISSARLFNLYGSSEVAADVTCFE
VKLRIANQIEAKSKEKLDALKNLPSGSGDRETAVLHKEIIHLQLADERRADLGEALEEYLKRNTIPIGKPISNTQIYI
LDKYGDLLPPGVTGELYVGGDGLAKGYLNLPELTREKFIPNPFVKDRGKSKKAQAERLFRTGDLARWLPDGNIEFVGR
IDHQVKVRGFRIELGEIEAVLSTHPQIQQVVVIAIEDIPGSKRLVAYIVCEDESLSTYHLREFLKQKLPEYMMPSAFV
ILDTLPLTPSGKIDRKALPAPDGEISREHEYVPPRTSGEEIIANIFASILGVQNVGIHDNFFELGGHSLLATRLISRL
RVAFEVEIELSAVFSSPTVAQLEQTLTQLRTTNSALSLPPIQPRTQNQQLPLSFAQDRLWFLNQLEGSSATYNMPGAI
RVTGKLDINALQQALSEIVRRHEVLRTSFRTVNGTPIQVIHPEATMNISVADLQQLEATERESVLHQQAQLAAITPFD
LETAPLIRCSLLQLDAREYVLLLTMHHIVSDGWSMGIFSQELSTLYQAFSAGKPSPLAELPIQYADFAVWQRQWLSGK
VLETQLNYWLSQLEGAPELLQLPTDRPRPTVQTFRGTTQSFSLNTDLKEKLQTLSRNSGTTLFMTLHAAFATLLYRYS
GQLDILIGSPIANRNCSEIESLIGFFANTLVLKTRFEDNPSFENLLAQVRETTLEAYEHQDVPFEQVVEVLQPQRSLS
YAPLFQVMFVLQNAPMGELELPGVTLNLLSSQTETARFDLTVSMQQTSEALVGSWEYNTDLFDGSTIERMTAHFQNLC
SAIVENPQQKISELPLFTDSEQEQVLHSYNNIATTYLLDKYVHFLSSNNLQIYILDNHQQLVPLSVEGEIYLGNCDLL
PDKLHPEPEKFISFIEHTQLGKLLKTGEWGCRRVDGSLELLGKEHRIVTVNGQRINLQRIEQALQTAKGVEDCYVMVR
NQKLVAYVVKDGSWAREFLHHYLKSQLPGYPLPCIYVPVSALPLTSFGEVDEVGLASISIIDSELINTWEEQIGSQAE
IDKVAVFIEPNVKTISPIHLEELLPSIQAIFNQGSTPVETPRTARGKESSSLLEIKSPAISHEEVLIFPESSPETLGE
MLQKTAGKFPHKGITYINSDGSEQVQSYAQLLEDAQRILGGFRKLGIKPQDKVILQLKENKDFISAFWGCVLGGFIPV
PVVIPVSYDQPNVNLNKLQNSWQMLERPLILTDKKSLSELKKWSQNLNDDNFKLETIESLQKFSTDKDYYNAQPEDLA
LFMLTSGSTGMSKVVQLSHLNLLSRTIGSIQMNNFTPEDITLNWMPLDHVAGLIYFHIRDIYLGCKQIHATSQLVIEK
PLRWLDWIDTFGVTVTFAPNFAYSLINDFVQEIEKQNWNLSSIRLMLNGAEQIVAATARRFLKLLAPFGLPGDAMTPS
WGMAEVSSGITYSDNFSLLSSSDDNSFVNLGKPIRGTCLRIVNQDMEVLSEGEIGLLQVKGLTVTSGYYQNPKANKEA
FTEDGWFNTGDLGFIKDGCLTITGRQKDIIIINGVNYYSHEIEAVVEELGEVEVSYTAACGVCVASNNTEELVIFFTP
YVSEKNQLLELLKKVREQVIKYCGINPSYLIPIDKELIPKTSIGKIQRSLLKQRFECGEFKSLRQRVDLLLDNTNTIP
NWFYRKVWQIKESKNTLLNYSSQKTLTLIFTDNLGWQQDNRGMSQTVQPYAQVTIGSNFAQISPNHYSVVPGNPQHYR
LLIDSLRQNSQVISQILHLWNYNEQTEKISSLENLESTQQQGIYSLLFLVQALEEIQGKQQAVKLLWIANQSQLVHPT
DKIQPEKSTVLGLLKTVSQEMPWLTTRHLDLPLAPELNNSYIWQELYSADKELEVAIRNRERFVSGLEPVDMTAKEKQ
KIPILPGGTYLLTGGLGGIGTVIAKYLLEHYQANLILVGRTQIEDNNEEASTKLQRYQELEKLPGSIIYQTVDICDLV
GLQQVVEKATQEWRTQLDGVFHMAGIIQETPIEKETPGNIAAVLRPKVSGTWVLHQLLKDKENALFVHFCSVNGFFGG
TNVAAYSAANSFQSAWSDYQQQNGFQSYCCSWSMWNETGISHGYQFQELSRAKGYFIITPQQGFYSFLAALSGSEHNL
LIGLDGTKTNVEHLIRDCQPKQKLTAYFTSPTPELAALSLQELQLHDRFGIPNQINFVQLEQIPLTQRGEINREQIAA
IYGGLNTSEQTKPRNQTERQLVEIFQEVLNLPSIGIHDNFFSLGGHSLLAVRLMSEIQQQFQKNLPLATLFQNPTIER
LALLVGSDSGAELWSPLVPIQQNGSLPPLFCVPGAGGNVLYFHHLAQYLGNNQPLYGLQAQGLDGETEPHKSVEEIAS
QHIK

FIG. 6-8

AIQTVQPVGPYFLAGHSFGSHVVFEMANQLQLIGKSVAYVGILDTPAPTSQANHQNDFSNWDNAKWICRMAEVIEDIV
GENLFLSYETLTSLTWEQQLNYFKQKLEIVGFLPAQTDIKIVRGLLQVFQTQCQIKYEPEKTYKTPITLFCAREINPE
QESYSHIFQEPTWGWNQFSDGEVEIHIVPGNHVSMLSEPHVKVLAQQMQISLEQAQKTHQLEK (SEQ ID NO:7)

crpF
ATGATTAATACTGCTAAATCCTCATTACTTCCTGGTCCCACTACACCATCTTGGTGGAACTTATTGCAATGGCTTAAT
AATCCTTGTGAATTTTTGGAAGAGTGTCGAGCACGCTATGGAGACACTTTTACCTTCAAAGCTATTGGTTTTGAACCT
TTAGTACTTATTAGTAATCCTAAGGATATAAAAGAAATTTTTGATAAACACAAGTATTTTGACAGTGGAAAAGCTAAA
GCTAACGATTTAGCAGGATTTTTTTAGGCAACAATTCCGTCACCTTGCTTGATGGAAGTAGTCATAAACGACAGCGT
AAACTACTGATGCCTGCTTTTCATGGTCAAAATATATCTAACTATGGAGAACTAATATGCCATGCAACGAAGCAGGTT
ACTTCTAATTGGCAACCTGGTCAAAGATTGATTATTTACAAGGAAGTCAAAGAAATTACGCTGCGAGCGATGTTAACG
GTTTTACTGGGTTCAGATAAAACGGAACGTTATCAACAACTCAAATTGATAGTTAATCAAATAGTATCCACTATAACT
AATCCCTTTGCTTCTAGCTCTCTTTTCTTCAATGTGTTTAGAAGAGACTGGGGTTCTTGGAGTGCCTGGGGTAATCTT
TTACGTTGCCAACGTCAGATTGCAAATATCATTTCTGCAGAAATCAAAGAACGTAGAGAAAATTGTAACAATTACAAC
AATGATATCCTCAGTATGCTGATGGCAGCACGAGATGAAAATGGAGGAAAAATGACAGATGAGGAGTTGCAAGATGAG
TTAATGACACTTATCTTTTCTGGATATGAAACTACATCTGCAGCAATAACATGGGCATATTATTGGATTCATTACTTA
CCAGAGATAAGAGCCAAGTTATTGCAAGAATTAGATGAGTTAGGAGATAATCCAGACCCAACGGAAATAAGCAAATTA
CCTTATCTCAATGCAGTTTGTGCTGAAACCTTGAGAATATATCCAGTTGGTCTAACTACTTTTCCTCGAATTGTAAAA
TCGCCAATAGAAATTGGAGGTCATCAATTTGAGGTAGGAACTTGTCTTTATCCATGTATTTATCTAATTCACCACCGG
GAAGAACTATATCCTAACTCTAAACAGTTTAAGCCAGAACGTTTTCTAGATAATAAATTTTTAAATTATGAGTATTTC
CCTTTCGGTGGCGGTAACCGAACTTGCATTGGTATGGCATTTGCTCAGTTTAAAATGAAGTTAGTATTGGCTAATATT
TTGCGGAATTGGCAATTGGAATTGGTAGGCAAACCTCCTTTAAAACCAGTACGAGATATTTTCTCAATTTATCCTCAA
GGTGGATTAAAAATGGTTGTATTGTAA (SEQ ID NO:8)

CrpF
MINTAKSSLLPGPTTPSWWNLLQWLNNPCEFLEECRARYGDTFTFKAIGFEPLVLISNPKDIKEIFDKHKYFDSGKAK
ANDLAGFFLGNNSVTLLDGSSHKRQRKLLMPAFHGQNISNYGELICHATKQVTSNWQPGQRLIIYKEVKEITLRAMLT
VLLGSDKTERYQQLKLIVNQIVSTITNPFASSSLFFNVFRRDWGSWSAWGNLLRCQRQIANIISAEIKERRENCNNYN
NDILSMLMAARDENGGKMTDEELQDELMTLIFSGYETTSAAITWAYYWIHYLPEIRAKLLQELDELGDNPDPTEISKL
PYLNAVCAETLRIYPVGLTTFPRIVKSPIEIGGHQFEVGTCLYPCIYLIHHREELYPNSKQFKPERFLDNKFLNYEYF
PFGGGNRTCIGMAFAQFKMKLVLANILRNWQLELVGKPPLKPVRDIFSIYPQGGLKMVVL (SEQ ID NO:9)

crpG
ATGTATTCAATAAAAATTGAAAATCTAATAATTAGAGTGAAAAGTGTATTAGAAATGCCAGTTTCTAAAGAAGCTGAG
ATGGCAAATAAATTTAATGAGTTTGGATTCGTAATACTAGAACACGAACCTTCAGCAACACCTAAGAATAACTTATTA
AAATTGTCTGATTATTTTGGAACAATTATTCAGCACGAACATTCTGATTCACAGGGAATTGTTCCCATCAGTCCTGTT
GATAGTTATCCAGAATATGTAAATACTACAACTACAGATTTATCGTTACATACGGATGGAGCGTTCACAATTACTCCA
CCAAAAGTAATGGCAATGCAGTGCCAGATTGCTGCTGCAAATGGCGGGTTCACCAAGCTTATTGATGGCAAGCTGGTA
TATGAACATCTAAAGCGGACAAACCCAGTTGGATTGTTAACTTTGTTTAATCCTGATGCGATTACAGTCAAAAGAGAT
AATAAAAAAGCAACTAAACCTATTTTTGAAGAACATCATGCTGGGCTTATTGTAAGGTTTAGAGCAGATAATGCAGCT
CATGTTTCGGTTGAATCGAAAAGTTTTGCGGCATTTAAATCATTTGAAAACTTTGTAAATAATCCTGACAATCAAGTA
ATTTTTAAACTTGCACAAAACCAAATAATTATTGTAGATAATACTAGAGTTTTGCATGGAAGAACTGCATTTTCCAAA
CAAGAGTATAGGCTACTAAATCGACTTTGGTTTGATGGACAATCTGATATTATAAATTTAAAGTTTGGTATTTCTATA
GCCCCAAAAAACTTGAGTTTATTTGCTAAAAAGTATCAGCCATCTCAAATAGATATAGGCTCAGATATTTCTCAGTCA
ACTCAATTGAAATTTAAAGCCACATGA (SEQ ID NO:10)

CrpG
MYSIKIENLIIRVKSVLEMPVSKEAEMANKFNEFGFVILEHEPSATPKNNLLKLSDYFGTIIQHEHSDSQGIVPISPV
DSYPEYVNTTTTDLSLHTDGAFTITPPKVMAMQCQIAAANGGFTKLIDGKLVYEHLKRTNPVGL

FIG. 6-9

LTLFNPDAITVKRDNKKATKPIFEEHHAGLIVRFRADNAAHVSVESKSFAAFKSFENFVNNPDNQVIFKLAQNQIIIV
DNTRVLHGRTAFSKQEYRLLNRLWFDGQSDIINLKFGISIAPKNLSLFAKKYQPSQIDIGSDISQSTQLKFKAT
(SEQ ID NO:11)

crpH
ATGTTGAAGTCGAAAATTCACAGAGCGACGGTGACGGAAGCCAACGTTAACTACATCGGAAGTATTACAGTAGACAAA
GTTCTGATGGAAAAGGCAGACATACTACCGGGTGAAAAGGTTATGGTGGTGGACAACACTAATGGTAATCGTCTAGAA
ACCTATGTCCTAGAAGGTGAGGAAAATTCCGGGGTAATCTGTATGAACGGTGGCTCCGCCCACCTAGTCAATTCAGGA
GACCTTATCACATTGCTAGCATTCGAGGTAACTGACGAAATCAAGGAACCGAAAAAAATTATCGTGGATGAAAACAAC
AAGTTTCTCAAGTACCTGTAA (SEQ ID NO:12)

CrpH
MLKSKIHRATVTEANVNYIGSITVDKVLMEKADILPGEKVMVVDNTNGNRLETYVLEGEENSGVICMNGGSAHLVNSG
DLITLLAFEVTDEIKEPKKIIVDENNKFLKYL (SEQ ID NO:13)

crpJ
ATGTCTACACTGCCTAATTCCACACAGATTCTAATTATCGGAGGGGGACCTTCTGGATCTACTGCTGCTACCCTATTG
GCTCGTGAGGGCTTTGATGTAACGCTGTTAGAACGAGAGGTATTCCCGCGTTACCACGTTGGGGAATCTCTTTTGCCC
TCTGCTTTAGAAATTTTTGACCTGCTTGGCGTACGCGAGAAAATTGAAGCTTATGGCTTTCAGCGTAAACCTGGAGCG
TACATAGAATGGGGAACGGAAAAGTGGAGCCTCAATTTTGGGGAACTTACGGGGGACAACACCTACAGCTTCCAAGTT
CGCCGTGACGAATTCGACCACTTGCTTTTAGAGCATTCAAAGAGCCAGGGTGTGAAGGTTTTTGAAGGGACTAAAATT
CGCCAGTTGTCTTTTGATGGCGATCGCCCGCGCAGCGCTACTTGGTCACAATCAAATGATACTACCGGGGAGATTTCT
TTTGACTTTATGATTGACGCTTCAGGTCGTGCTGGGATCATGGCGACGGAGTATCTGAAAAACCGCCGTCTACACGAC
GTATTCCAGAATGTTGGCATCTGGGGGTACTGGAAAAACGCCTTGAGACTACCTAAAGGTCAGTCGGGTGCGATTGCC
TTGGGCTCCATTCCAGATGGTTGGGTGTGGGGAATTCCTTTGGATGAGGAAATTATGAGCGTTGGTGTAGTGATGCAT
AAGTCAACCTACAAGGAGAGACTGACTAAGAACTTGAAGGATATCTACGTGGAGGCGATTGCAGAGTGTCCCTTGATA
GCGGATCTGGTTGCACTAGGGGAGCTAGTCTCAGACGTGAAAGTTGAGCAAGATTACTCTTACACTTCCGACTCCTTT
TCAGGACCAGCCTACTTCATATCGGGAGACGCTGCTTGCTTCCTAGACCCCCTACTATCGAGTGGGGTGCATCTTGCT
ACTTATAGCGCTTTGTTAGCCGCAGCCAGTATCACAAGTGTTATACGTGGCGAGGTGACTGAGTCACAAGCTGCTTCT
TTCTACGATCAGAGCTATCGGCAGGCTTATTTGCGTTTCTTAGTGTTCGTATCAGCCTTCTACGATCAAAACCGTGGC
AAGGATTCCTATTTCTGGGAGGCACAACGGCTTAGTCGCCGTGACTTCGGCAGTTCTAACCTAAAGCTAGCATTCTTG
AATCTGGTGTCCGGCGTCGAGGACTTGGAGGACGCTAAGGAGGGGATTGCCGATTTTGTTATGGCAGAGATGTCTCAG
CGGATTCAGTCAAGCCACAGCATTAGGCAAGACAAGCAGGCGTTGGCAATCGAAAGGGAAAAAGGTAACGAGGTAATG
AAGACAAATGCCCAGTTTTTCAATGCAGTCGAGGGATTTTCCATACTATCGGCAGTTGGGGCAGTTGATGGTCTATAT
GTTACAACTCAGCCAAAATTAGGATTGGTACAGGTAATCCCTCTCCAAAGAAACTCTTTGCTCCACACTTAG (SEQ
ID NO:14)

CrpJ
MSTLPNSTQILIIGGGPSGSTAATLLAREGFDVTLLEREVFPRYHVGESLLPSALEIFDLLGVREKIEAYGFQRKPGA
YIEWGTEKWSLNFGELTGDNTYSFQVRRDEFDHLLLEHSKSQGVKVFEGTKIRQLSFDGDRPRSATWSQSNDTTGEIS
FDFMIDASGRAGIMATEYLKNRRLHDVFQNVGIWGYWKNALRLPKGQSGAIALGSIPDGWVWGIPLDEEIMSVGVVMH
KSTYKERLTKNLKDIYVEAIAECPLIADLVALGELVSDVKVEQDYSYTSDSFSGPAYFISGDAACFLDPLLSSGVHLA
TYSALLAAASITSVIRGEVTESQAASFYDQSYRQAYLRFLVFVSAFYDQNRGKDSYFWEAQRLSRRDFGSSNLKLAFL
NLVSGVEDLEDAKEGIADFVMAEMSQRIQSSHSIRQDKQALAIEREKGNEVMKTNAQFFNAVEGFSILSAVGAVDGLY
VTTQPKLGLVQVIPLQRNSLLHT (SEQ ID NO:15)

crpM
ATGTTATCTCCCCTATTTGATGCTTTTGTAGAGGCAAGCCCCGTCAGTGTAATGATGCGAGTCCTAATGGAAAACATT
TTTAATTCCTCGCGAATGAATCAAATATTTGATACATCAAGCGTTCGCCAATACTCTCAAGAGCTACTGTTTTCGACT
CAGGTGGATTTGATGAGTCTAGTAGTGTGTGGGATGTATCCCTCGGTTCATGCAGCCTATCAGAAGAAGGCAGTGGAG
GTAAGTGTCAGCGCCACAGCGTTATACAACAAACTGCAACGGATTGAACT

FIG. 6-10

GCCTGTAAGTCGGGCATTAGTGCATGAGACAGCATCTGACCTCCAGCAGTTGCTGTTGATGTTGAATGTGGAACGCCC
CAGTCCTCTAGGAAAACAATATCGGTTGCGGATTGTAGATGGCAGTTGTTTAGCCGGAACCGAACGCAGACTAGCAGC
GCTGCGCCCCCATGCAGCCAAACCATTACCCGGAAAAACAATCGCCATTCTCGACCCAGGGACAAAACTGGTGGTTGA
TGTGATTCCTTGTGAAGACGGTCATTCCCAAGAACGCTCCAAGTTTCATCAGGTTTTGGCACAAGTGCAACCCCAACA
GGTATGGATTGCAGACCGTAACTTTTGTACCGCAGGATTTCTCCATACTATTGCCAAACTTGGAGCGTTTTTTGTGAT
TCGTCAACACGGGGGTTTAGGATACGAGCCTTTTGGTGAGTTACAAGCTGTTGGGTTGTGCCAAACAGGAACTGTGTT
TGAACAACAGGTGGAAATTGTCCATGAGGGAGGGACTTTTCGGTGTCGCCGTATCGTAGTTAAGTTGACTCGTCCCAC
CCGTGACCAAGAGTGGGAAATTGCCATTTTTACCAACTTACCACCCACTGACGCAGACGGCATTCTGGTGGCACAACT
CTATCAAGGGCGGTGGAGTGTGGAAACTTTATTCCAAACTGTGACCCAAAACTTTCATGGAGAAATTGAAACCCTAGC
TTATCCTAAAGCTGCCTTATTCTCCTACTGCATGGCACTGTCAGCCTACAACCTTTTAGCGACACTTAAAGCAGTTCT
TGGCAGTGTACATGGGGTAGACAAAATCGATATTGGGCTATCCGATTTTTACCTAGTAGATGATATCCATTCCATCTA
TCGGGGCATGATGATTGCTATTCCTCCGGTTCATTGGCAATTCTTTGAGGAGTTTACCAACATTCAGATGGTAGACGT
TCTCCAGCATCTAGCAACCAAAGTACATCTCAAATCTTTTCGCAAACACCCCAGAAGTCCCAAAAAGAAACGACCACC
ACTCTCTGTTGATGGCAAACATTCCCACTGTTCCACTACTCGAAAGCTCAAGCAATACAAAGCAGCTCTTGATGCTAT
CCCGTGA (SEQ ID NO:16)

CrpM
MLSPLFDAFVEASPVSVMMRVLMENIFNSSRMNQIFDTSSVRQYSQELLFSTQVDLMSLVVCGMYPSVHAAYQKKAVE
VSVSATALYNKLQRIELPVSRALVHETASDLQQLLLMLNVERPSPLGKQYRLRIVDGSCLAGTERRLAALRPHAAKPL
PGKTIAILDPGTKLVVDVIPCEDGHSQERSKFHQVLAQVQPQQVWIADRNFCTAGFLHTIAKLGAFFVIRQHGGLGYE
PFGELQAVGLCQTGTVFEQQVEIVHEGGTFRCRRIVVKLTRPTRDQEWEIAIFTNLPPTDADGILVAQLYQGRWSVET
LFQTVTQNFHGEIETLAYPKAALFSYCMALSAYNLLATLKAVLGSVHGVDKIDIGLSDFYLVDDIHSIYRGMMIAIPP
VHWQFFEEFTNIQMVDVLQHLATKVHLKSFRKHPRSPKKKRPPLSVDGKHSHCSTTRKLKQYKAALDAIP (SEQ ID
NO:17)

crpN
ATGAACAAACCACCATCCAGACGCAAGAAAATTACCCCTGCGACATCTGAGGAACCAAAGCTAGCAACTGACCCTGCT
CAGGAAAATACTTCTTTGCACGAAAATCCAGGGGGAGCAACTATCACGGTGACGGCTGTTGAAGTAACAGATTTGACC
CAGGAAGAACAAAGCTTACGCCTGCATTTAGAACACCGTGTGGAGAGAGCATTTTTGGAGGCGGGTCAAGCGTTGATG
GAGTTGCGGGACAGACGGCTGTACCGTTCCACGCACCGGACTTTTGAAGAATACTGCCGCGAACGCTTCAATTATAGT
CGTGACGCGGCTTACTTGAAGATTTCGGCTACTGTGGTTTATGAGAATCTTCAAAAGTTTTTGCCGACCATTGGTCGG
CAAATTCCAATGCCGACCAACGAACGACAATTGCGTTTTTTGGCGAAAGCCGAGTTGGAACCGGCTGTGCAAGCGGAT
GTATGGCGGCAGGCAGTGGAGCAAGCTGGCAATAAGATTCCATCCGGTCGCATAGTGAAAGATGTTGTAGATAGGATA
CGCGAAAGGACGAAAGTACCCAATCCTTACCACGTTGGGGAGATATGCGTTCTTCTACCCAAAGATAATGCAGACTTG
AGAGGTAAAGCGGGTTATTGGGGCGTGGTCAGCCATGTTGGAGAATACAGTTGTACACTCCAGATATGGGACGGTGAC
TATACCGTAAAAATCGAACACCTGAAATCACTGGAATTACTTGATGAAGATTGCCAATTCATGCAGCAGTTATGTGTG
AGGTTACGGCAGTTGCATCAAGTGGACAGGCGTGACGAGGCTGTGGATTGGCTGTTGCAGTGGTTGGGGAAACAGGCC
AAACCTTATCTGTCATCCTTGCAGTCAAAGCTGCTGGCGTTTGTTGAGAGAGAGTACAACCTGGTTTGGAAGCAGCAG
AAGTGA (SEQ ID NO:18)

CrpN
MNKPPSRRKKITPATSEEPKLATDPAQENTSLHENPGGATITVTAVEVTDLTQEEQSLRLHLEHRVERAFLEAGQALM
ELRDRRLYRSTHRTFEEYCRERFNYSRDAAYLKISATVVYENLQKFLPTIGRQIPMPTNERQLRFLAKAELEPAVQAD
VWRQAVEQAGNKIPSGRIVKDVVDRIRERTKVPNPYHVGEICVLLPKDNADLRGKAGYWGVVSHVGEYSCTLQIWDGD
YTVKIEHLKSLELLDEDCQFMQQLCVRLRQLHQVDRRDEAVDWLLQWLGKQAKPYLSSLQSKLLAFVEREYNLVWKQQ
K (SEQ ID NO:19)

crpP
ATGACGAAGWTAAGATGGGGATRKTCTYGKMTCGWARTATCAGTTATACAAAATACTACAATCTTAAACATACAATTG
TTAGCTTCGACAACTATTCAATCAAAGTATATATTTAATATGGCTATCAAACACCCTTTTTTATTTGCACTGTTAACG
CTCTCCATTATTTGTGTTGGTACGAGTTCTGGCTCTGCACTACTGACAGATATTGCTCAACAAACAGACAACCAAAAG
TCCCCATCGATTATTTTCTTCCTGCCCAAAGAACGACCTCAGACCGGAGT

FIG. 6-11

CGGTTGGGAAATCACTACCACTTCAGGGAAGGCAGAACTAGCCTTGGCGAAGCATTTGGTGTATATCGGGGCAAAAGA
ATATGTTTCTTGGTGGTGTCCTCACTGTCACGAACAAAAGTTAATCTTTGGGAAGCAAGCCTACCAAATAATCAACGA
CAGTATTAAAGTTGAGTGCGATAAGAGAGGTATCAATCCCCACCCAGACTTGTGCAATGCGGCGAAAGTCCCAGGTGT
ACCAACTTGGGTTATCAATGGACATCAGTATACCGGCGTGCAAAACTTTAAGGATCTTGCGAAAGCTTCTGGCTACAA
GGGGGATATGAACTTTCGTTATATCCAAAGCGAATAA (SEQ ID NO:20)

CrpP
MTKXRWGXSXXXXSVIQNTTILNIQLLASTTIQSKYIFNMAIKHPFLFALLTLSIICVGTSSGSALLTDIAQQTDNQK
SPSIIFFLPKERPQTGVGWEITTTSGKAELALAKHLVYIGAKEYVSWWCPHCHEQKLIFGKQAYQIINDSIKVECDKR
GINPHPDLCNAAKVPGVPTWVINGHQYTGVQNFKDLAKASGYKGDMNFRYIQSE (SEQ ID NO:21)***

crpU
ATGATACAGTGTAATTTTTCGTTGCCACCTGAGTATGTTCTTCGTAAGGCCAAGCCTTTTGATATGTGGTTAATAGTA
TTTTTTGTGTTTAGAGCAAGGCTAGACCCCAGTCAATTAAGATGGCAGCAATTTTGGGTCATTGAATGTGATGGACAT
TTAGTAGCCTTCGGGCAGATCCGAAACTTTCACTTAGCACAAGAGCTAGGCAGTTTATTTGTTGCACCGACTTGGCGA
AACCGTGGTTTAGGGACTGTTTTGATACAGCATTTAATTACTCAAGCTAGTCAACCGCTTTATTTAAAATGCTTAAAA
TATCAATTGGTGAATTTTTACATTAAAAGAGGCTTTGTATCCGTTAATTTTAAAGATTTACCACCATCCCTCAAGCCA
AAGTTTGGACTATCCCAATTACGAAAGAGGTTAACGAAAGCTTTTGTGCTGTTTATGAAGTATGAATATCCCAACTGA
(SEQ ID NO:22)

CrpU
MIQCNFSLPPEYVLRKAKPFDMWLIVFFVFRARLDPSQLRWQQFWVIECDGHLVAFGQIRNFHLAQELGSLFVAPTWR
NRGLGTVLIQHLITQASQPLYLKCLKYQLVNFYIKRGFVSVNFKDLPPSLKPKFGLSQLRKRLTKAFVLFMKYEYPN
(SEQ ID NO:23)

crpV
ATGTCAGTGCCAGTTAGCGCACAGATTATACCAGATAAAACACTACCTATTAATTCCAATGTTGAACATGAAGGTAAT
ACTAACCGCATAGAAGGTGGCACTATAAAAGGGAGCAACTTGTTCCACAGTTTTGAACAATTYTCCGTGCTTACTGGA
AATGAAGCTTACTTTAACAACGATATAAATATCCAAAACATTATTACTCGTATTACTGGGAAGTCTATTTCTAATATC
GATGGCATTCTCAAAGCCAATGGCACGGCTAATTTGTTTCTGCTCAATCCCAATGGCATTATTTTTGGTAATAATGCC
AAACTAAATATTGGTGGTTCATTTCTAGCTACTACTGCAAATCAAATTAATTTTGCTGATGATACTAAATTTAGTACA
AACAATCCCCAACCTAATCCTTTACTGACAGTAAGTGTGCCTATAGGACTGCAAATTGATAGCAACCCCGGTACAATT
CGCATCCAAGGTACAGGTCACAATCTAATTGGCCCTCCTTTTTCTCCTCTAATCACAAGTAGTAGCGCCGCAAATTTA
CAAGTGCAACCAGAAAGAACTGTAGCAATTGTTGGTGGTGATGTAATTTTAGAGGGAGGTGTGATAACGGCTAGGGGA
GGGCGAATTGAATTGGGTAGCCTCAGCAATGGTTCAGTCAGTATTAATCCTACGACCTCTGGTTGGAAACTGGGCTAT
GAAAATGTACCTTATTTCCAAGATATTAACCTCTCAAAACGCGCTKTAGTTAATACTAGTGGCATTGGCAGTGGATCT
ATACAGATAGAGGGACGCAKAGTTACGCTTACAGATGGCTCAGTAATCTTAAATCAAAATCAAGGAACACTACCAGGA
GGCACACTAAACGTGAATGCTTCGGAGTCTTTGTCAGTGAGTGGTAGCGATCCAATTGCTAGGACAGCTGGTGGTTTG
CGGAGCGAAACTTTGGGATTYGGCAAAGCTGGAGACATTGCAATTTCAACCAAACAGGTAATTATTAAAAATGGAGGA
CAAATAAATAATTTAACCTTTGGTGCTGCAACAAGTGGCAATATAAATGTGAATGCCTCTGATTCTATACAATTGCTT
GGGGTTTCGCCTTTTGACCCTGCTGTTTTTAGTACTATCAGCACTGCAACTTTCAATTCTGGAAACGCAAACAATATT
ACAGTGTCAACAGGACAATTCGTTGCCACGGATGGAGGTAACTTGTCCTCTTCAACCTTTGGAACTGGTAGAGGAGGA
GATGTCACTGTAAGTGCAACTGACTCTATAGAAATAATAGGAGCTTCACCAATAACCTTTCAGCCAAGTATTTTATCT
TCCATATCGCTCAATGCTGGCAAAGCTGGCAGCCTAACAATCAGTACATCAAAGTTGATGGTTCAAGATGGCGGGAGG
GTTGACGCTTCTACTTTAGCAAGTGGGGAGGGCGGTAGTGTTACGATTAACGCCTTTAAATCTGTAGAGGTAAGTGGT
AAGATACTTGGTTTTGGAGAGCCTAGTTTGGTGATCTCCAGTGCTAATATCGTCTCTCCAATCTTGCAAAAGTTATAC
AGACTCCCTTCAGTGCCTTCTGGAAAATCTGGAAACGTGACGATTAATACTGGTCAGTTGAGTGTTACAGACGGTGCT
GAAGTTAACGTGAGAAATGACGGTTCTARCGATGCTGGAACACTCAGAATCAATGCTGTTTCTGTTTCTTTAAACAAA
CAAAGTGCCATTACAGCAACTACTGCTAACGGCGAAGGCGGTAATATTTTCGTGAATACACGGTATTTGCAGCTAAGT
AATTACAGTGTTGTAACGACGACCGCAGGTAGTAGAGGCAATGGCGGTAATA

FIG. 6-12

TAAACATCAATGCAGATATATTAAGTGCTTGGGGGAAGAGCAGTATTGCTGCCAATGCTTTCTATGGGTATGGAGGAA
ATGTACTAATTAATACTAGAGGACTTTTTATTGCTCGTGACAGTCAAATTTCTGCAAGTTCTAAATACGGAATTAACG
GCACTGTTAGCATTAACAATACTGGTGGTGAAATTTATCCTACTAAACTCAAATCAGAATCGATTCCAGTAGCTCCTC
AAATAGCATCAGTTTGTCAAAAAAATTCAGATATACCAATCAGTAAATTTGTGAATGTTGGCACCGGTGGACTGCCAG
CTAATTCTGATGATATGCCATATATGAATTATGAACAGCAAAATAACTCTGTTTCAATCCACAATAATAATAACTTAG
AGGCATCGAAGGCATCACAAACTGAAGAACCTATACAGATAATAGAAGCTCAGGGTTGGATAATAAATCTTGATGGGG
AATGTCGTCTTAACTGCACAAAACAATACAGCAACCCCTAA (SEQ ID NO:24)

CrpV
MSVPVSAQIIPDKTLPINSNVEHEGNTNRIEGGTIKGSNLFHSFEQXSVLTGNEAYFNNDINIQNIITRITGKSISNI
DGILKANGTANLFLLNPNGIIFGNNAKLNIGGSFLATTANQINFADDTKFSTNNPQPNPLLTVSVPIGLQIDSNPGTI
RIQGTGHNLIGPPFSPLITSSSAANLQVQPERTVAIVGGDVILEGGVITARGGRIELGSLSNGSVSINPTTSGWKLGY
ENVPYFQDINLSKRAXVNTSGIGSGSIQIEGRXVTLTDGSVILNQNQGTLPGGTLNVNASESLSVSGSDPIARTAGGL
RSETLGXGKAGDIAISTKQVIIKNGGQINNLTFGAATSGNINVNASDSIQLLGVSPFDPAVFSTISTATFNSGNANNI
TVSTGQFVATDGGNLSSSTFGTGRGGDVTVSATDSIEIIGASPITFQPSILSSISLNAGKAGSLTISTSKLMVQDGGR
VDASTLASGEGGSVTINAFKSVEVSGKILGFGEPSLVISSANIVSPILQKLYRLPSVPSGKSGNVTINTGQLSVTDGA
EVNVRNDGSXDAGTLRINAVSVSLNKQSAITATTANGEGGNIFVNTRYLQLSNYSVVTTTAGSRGNGGNININADILS
AWGKSSIAANAFYGYGGNVLINTRGLFIARDSQISASSKYGINGTVSINNTGGEIYPTKLKSESIPVAPQIASVCQKN
SDIPISKFVNVGTGGLPANSDDMPYMNYEQQNNSVSIHNNNNLEASKASQTEEPIQIIEAQGWIINLDGECRLNCTKQ
YSNP (SEQ ID NO:25)

crpX
ATGGTGATTATTCAAGCCACGCAGCATTTCTGTAGATTTAGTCTTGGTGTTTTCTTAGCACAATCAAGAGTAGAGATA
GAGCAGAGTTTAACAATGTCAACTCCTAACTATCGTCAAGAGATTGATATTGTAAAACGTTTATTTTCGCAAAATCCT
AATTTATGCGTTGATATTATGCTAGCGACTGAAGAAAGGTGTAATGCTATTAGCTTTTTAGCTAAAACTTACAGCCGA
TTGGCTAGACTGGTGGCTAGGAAGGATAGAGAGGCATTAATTAAAGAGTTTGAAAATACTCAAAGTTTTTTTGAAGAG
AAAATTAATAGTTTTCTCCAGCCTTTAAATACAACGGCTCTGCAACGAGATTTTAAACCCCAGATGCACACAAATATT
AGCATTTGA (SEQ ID NO:26)

CrpX
MVIIQATQHFCRFSLGVFLAQSRVEIEQSLTMSTPNYRQEIDIVKRLFSQNPNLCVDIMLATEERCNAISFLAKTYSR
LARLVARKDREALIKEFENTQSFFEEKINSFLQPLNTTALQRDFKPQMHTNISI (SEQ ID NO:27)

crpY
ATGCTGATAGATATCTTTCATGATACCGTTTGCCCTTGGTGCAGAATTGGGAAAAAACATCTATTTGATGCACTGGCA
CAATGGCAAGAACAAGAAGTAAATATCCGATGGCATCCCTTTCTTCTGGATGATACTGTTCCTGCTGAGGGGTACGAA
TTTAGTAGCTTTATGCAAAATAGAAAAGGCATTAAAGCGCCAGAAATGCAACAGATGTTTGATTATACGCAACGCGCA
GGGGAGGCGGCTGGGGTTAAGCTAGATTTTGAAAAAATCCGTTTGGCTGTCAATACTAAGCTTGCTCACCAACTGATT
GCATTAGCACCGACAAACATAAAAAATGATGTCGTTGAAGCTATTTATAGAGCTTACTTTGAAGAGGGTTTGAATATT
GGAGATATTAACGTTATTGTTGCCATCGGTACAGCATACCAGATGGATGCTACCGAATTAAAGTTGCAATTAAACGAT
CGCGATGTCGTTGATACAGTTGTTGCTGAATCGGCATTTGCTCGCTTAAATGGCATCAACAGCGTGCCGTTTTTCATC
ATGAATAATCAAGTCAAGGTAAATGGTTCTCACTCGGTTGAGGTTTTCCTTGAAGCTTTGAATAGTACTGCACTTTTA
GATATACCTGCAAAAATATGA (SEQ ID NO:28)

CrpY
MLIDIFHDTVCPWCRIGKKHLFDALAQWQEQEVNIRWHPFLLDDTVPAEGYEFSSFMQNRKGIKAPEMQQMFDYTQRA
GEAAGVKLDFEKIRLAVNTKLAHQLIALAPTNIKNDVVEAIYRAYFEEGLNIGDINVIVAIGTAYQMDATELKLQLND
RDVVDTVVAESAFARLNGINSVPFFIMNNQVKVNGSHSVEVFLEALNSTALLDIPAKI (SEQ ID NO:29)

FIG. 6-13

crpZ

ATGATAGTTGACATCAAGCAAAAAAATAGATTAATTCATCAACGTGTTTCGGTTACTTTTAACTATGAGATTTACTTC
ACCCAAAATTTATTTGAGTTGAAAAACCCGACGCTAGCGCAAGTAATTTCGGCAGATGAGGAGACAAAGCCGAAGAAA
ATAGTTGCGGTGGTAGACGCAGGAATATTAAAGTATCAACCGGAATTGGTGAAGCAATTAGTTGCGTATACCAAGTTT
TATGGAGAGGTACTAGCGATCAATGTGCCCAAATATTAG (SEQ ID NO:30)

CrpZ

MIVDIKQKNRLIHQRVSVTFNYEIYFTQNLFELKNPTLAQVISADEETKPKKIVAVVDAGILKYQPELVKQLVAYTKF
YGEVLAINVPKY (SEQ ID NO:31)

FIG. 7-1

ATGGGGCATAGAGCTTGGGGATATCTGGGAGGGACAGTCGCAGGTATGTTCACTTTGGAAGATTTTTTAAGAGTGATT
CGTCATAGTGGTAGACCAATGGCACAGGTACCCTGAGGAGGCGAAACGTTATCTATAAGGAGTTCAATCGTAAAGATA
AGTGAAGTAATCGCGCCATACGCTCCAAACGTCGCGATCTCATAAATTGACGGGCCCCCAAGCACTGTCAATTCAGGT
GGGCCAGTTGCAATTGGAACGCTTCGAAATCTCTTAGCCGCAGTAGAGATTTTGACACGACGACGGGAAGTATCCCGC
GCGTTCAATGCATATCCGATGAAACGAACGTTAACGGAGTTTGAAGCAGCAGCACCAGCCCTAACGTACGATCCACTA
AACATACCATCAGTACCAAATGTATCGGGAGCGATTGCGGAGCAGAGTATAACCTGAGCAAGCGGTTGCCTAATTCAT
CACCGGCTACCGCAGAAATTTGCGCATAGTACGGACAGATAACAGCACGAAGGTTAATACCTCTGCGTAGACCTTGGA
CCAGAACCAGCTTTCTTAAGCGTGGCAAGACAGTACCCGCCAGAGGACGTGGGTCCTTGGGTGCGATCTGTCATACTA
GGACAAGTACACTCGCATCCAATGCTAGAAACTTAGGCTCAACTTTATGCGCAAGGAGTTAAACATGATTGGTTACGG
TTCATTAAAGATTATTGTGGTAGTACGGAAGTAGTCTCGACTTTTGCCTTCCATCGCCAACGTTATTGGTAGGAGAGA
TATAATAGTCTCATAGATAAGGAACAGATTTCATAAAATCATATTAATCTCCAAGCTCTAGTCGGTAATAGATCACAA
GTAGCAGCGTTTGAAAAGTGAATTCGGTTTAAATGCCAAAATGGTGCTTCTCATCCAAGTTACCAGCTACAGCAGTGA
GTCTTTCCTGTACCTCTTTACCAAGCTACAGATTAGTTGGTAATACCCATACCAGGAGGTACAACTTTATTTAAGTCA
GGTGATAGAACCCTATAAGATATTGTAATAGAAAAGGCATGAATTTTATCGACGGTTAAAATTCAGAGAATTCAAATA
GTGTTAATCTTACGGTTACTACAGAGATATAGATTACATATTTTTGGTCTGGAAACAAATACTGCTACTTAAGAACCT
AGCTGGATACTACATGTTCAAGGATAAATGTTAGTAAGGAACAAAGTCCCAGAATTAGAGACCACATACTTGAAAGCG
GTCAAAGAAGGGTATAAGGAAAAGATTTTACCTAATGAGTTTTACAAGAAACATGAAGAAAGGCGGCTTTATAACGCT
GCTTCTATCTAAGCCGTTGGACAACTGAGGCAGAGCGATAGCAAAAGACTAAGTGGAATTTAGTTGCCAAAAACACAG
GTCATAGTTGCAAATTTATGACAACTGCGCCTAAATGTTTAAGATGATATGTTCCAGGTATTTCCAGCTGTTGTGGTT
AGAACGGACAAGCATGGAGCTTATGTGCACTTGGAAAGAAACCGTCAACAAATATAGCGGACTGTTAGTAATGGTTCG
TCGACAGAAGTATAGGTAGATGTAGCAGAACCTAAGAAAACCACTTTGAGATGTAACGTTGATTCATTATATGAACGC
GGAATAGTATGAGCAAGAATTCAAGCTTAAAGTTTATTATGTAGTTCACGGGAGGCATTGTGTTGTAGTATCGAGCCG
CAAGTTAACAAGGGGTTATAACACATCCAGTGGCAGACCCGATCACTTGCACCACATCACGAAACAATTGACTTGACA
AACTCAGGTGGCTAGTTTTTGGCTTCACCAGCCACCGGGATAGAGAATCATATGGTGGAATCGTTCGCACTGCAAAGC
TGACATTGTATATTTGTAGGACCAGGGCAGAACTTCCAGGAGTTAGAAACTCCACAGTATCATATCGACCGCAATCAA
CCAGACGAATTCATGGACTTGATGAAAGCACGGTTGGTGCAGCAACCCCAGGTACCAGCAATTATGCACCTACGGGGT
TTCGACTAAACAATATCACTAAAGACTGGTGCGCAGCACTTGCTAAATTAACGAGTACTCGGTTGTGGCAACGAACTG
CATCGAGTCAAGGCCTTCGTAATAGATCGAGATAAGGAGAGTGCCCCCTTATTGTAAGTGGCTCAAGCCTCTCATTCT
GAGGGAAATGAATAACTACCGATACCATTCCGACAAGCAACTATGTGCGCGTTATGTGGAGTATTTGCACGGGGCCAA
AGGCAATTAGAAAGCCAGTGTTTAGGCTTACATCCATCTAAGGAAGCTTCCGAAAGATTAGCAGCTATCTTTGAGGGA
CTATTCTCTTCTGCTGATGAGAACGTAAATGCTAACTCTCAATGGTTAAGTCGCGTTCCTGAGTAAGAGAGGCGATAA
AGAATGAGAACATGTACACATTCCGAATGAGAAATATCGTCGCATCAATCATGTCTACTGCAGCTAGCAGTATATAGC
TCATTGGACACCCTTATCGAACTCGAGGCCATTTACTCAATTGCCGTAAGTCTGAGAGCTCCGGGCTAAAAAGCCGCT
AAGTGGTTGATACCACAAGCGGTAATAAATTTCGTGCTTACTGGACATACGCAGGCAACAGTAAGAGGTCTGCAAACC
CTTGATCAAATACACATGGCAGAAGCGCCAGTGTTATTCATCTGGTGAGAAATTTCGCATCAAGAGACTGTCGCAAGA
TTTATAAAGTCCAGCAAAGCATTATTGCGAGCACTATCAGGAAAAATACAGGCCGCCGGGAAATGGGTTGGTGGTCTT
CTGTTAGATATCAAGTGGGTTAAAGTTACATACGTGGTGGAACATAAACTACATCGGACTTGGTATTTCCAGAATATG
ACTCTGAATCTGCCTATGGACCTTGTTTTATGTTTTGCCTCCATTGATTCGATCTTTGGTTCGACTGGTCGCGGGTAT
TAAGCTGGTGCCATTGAGTTTATGGGTGGTTTATCCCATAATGGACGTGATATGCGATAACCTGGGTCGAGTATTAAA
TGGCGATCATGGCCACTAGAAGCAATGGCTGCAAATATGGGTAGCCGTCTTCGAGATTGAGCGGTGACCAAGGCAATG
TCTGTTATGGTTTCAGAGCAGGGAATGCAGCTTCGAAGTCAAAAACTCTAAGAATCATTAGCACAAGTGGGAGTGCTT
CCAAATGAATGGCCAGTGATCCAAGGGCCATTAAGGTTCGGTGATCAATTAGCATTACTCTCCTAAATGGGCAATGGA
AGCAAAACACACCATATAGCCATCGAGACTCAGACACAGTACAATGGGTTCTAAGAACACCTAAAAGCTGCTGTACAA
AGAGTAAGACATAAGATTGTGTTAAATTACGTCAAAGGTGTAATATCTAAACTATTTTGTTAGAGCATATCTCGACTT
GTTATGGATCACCCCCTCAACACAATCGGTCTAGATTCTTTAAAGGCCGTCCAAGTGCACACTAGTCTTCGAACTGAT
TTGCTCCTGAATATGTCGCTAGTCATAATTGTATAAAATACCTGTATCGTACATATAGGCTCTGAACTGTATCAGCAA
GTGAGCCTAATTACGCAGACTCAATGAGATGAGTGAGACAATAATGGTAAACTCGACCAAACTAAAAGCAACGATAAA
GACCGCATTAGTAGTGGATTACGT (SEQ ID NO:36)

FIG. 7-2

ATGGGGCATTATGCCGGGGAATCTGTGCCAGCCATAGTGGCTGGACTATTTGGTCTAGAAGACGTTTTAAACCCGCTT
GCTTATATAGGACGACTAATCCATCAATTACCATCTGGGAGTGAGATCTTATCTGTAATTGCATCTATTGAACAGCTA
AATCAGCTATTTGCATCATACCCTCAGAGAGTAGTGATCGCTTCGATCAACGTACACCAATGCATAGTCATCTCTGCT
GATGCAGACGCATTTGGACCGGGTCAACATAGCTAAGTAGCACAAGACAGGAAGACAACACGAGTGCAAATATGCCAC
CCATTTCATACACATCTGCTGGAATCAAAGTTGGTGGAATTTGACGCAGTCGCTTCAGACATAACGTACGATCATCCA
ACTATTCGATTAGCATGAAATGTATCTGGAGCGAGCGCAGAGAATCGTTTAGCGACAGCAAGCTCTTTGGGTAATCAC
GTCTGGCTACCGGTGTAACTTGCCAAAAGAATGCACAAATTACAGAATGAAGGCTATTACATCATCTTATAAATCGGA
CCAAAACCTACATTGTAAGGCATCAGAAGCCAGTGCATGTCAGAACATGAGGGAGTATGGGTGCCATCATTGAAGCTA
GCGCAAGACGACTGACGGCACATGCCACGAAGCTTGGCAGAATTATCTGTCCATGGCGTTATAGTTCATTGTTTAGCG
TTAGATAACGAATATTCTCCTAGAAAGGGAGTATTACCTACTTGTCCCTGTCAATGGCAGCGCTATTGCATTGTGACA
GATAATCATCTAATTCACCAGAGACAGTTTCAATCAAATCATAAGAGTCTTCATCATCTACTCAGTCATACATTACAT
ATAGAAGCATTAGATCAGCATATACGGTTTGAATGACATATTAGTGCATCACTACCAGCTTACGTTCAACACCGCTGC
GTTTTATCACAACCTGTTATCCTAGAAGCTGCTTACTAGGTAATCGCCTTAGGAGCGGGTACAATATTAATCAATGCA
GAGGATTTCATCCTACATGATATATCAAACTAAAAAGTAGTAACTTTATCTAAAGAAGAAGTTAATACACTTCATAAA
GGTTTAAACTTACAGTTTATACAAGGCTATCAATTCCTGATCTTCAGTATGGCTATAAGCACTAATTGATCCGAACCT
AGATGCATTCAACATGTTGAAGGACAGATAATAGTGGGTAATCAATACCCCGAATTACAAACACCAATCTTCAACGCG
ATTACAGAGGAGTAGAACTAACGGATACTACCTGCTGCATTCTACCGAACATTTGCAGACTGGTGTCTTAACTCCGGT
CCTTCGTTCTAAGCCGTTCAACTACTGCGGGACAGCGCAGGATAAGCATTAGATGTAATTCCGTTACGAGACACCGAG
ATGAATGTAACAACTTCATACTAACTACACCTAATTCGTTTCGATGCTCGCTGCCACGTGTTCGCAGCATTTAGGGGT
AACACGCACAGCCATGAATCTTATATGCGATTGCAAATAGAACGATTACGAATTCATCGGAGTGGTACTAATCGTTTC
TGGACTCAAGGAGAGATCGGTGCGACACAACCTAGTAAACAAGCTTTGACCGGTCAAGTGCGTTTATTCGCTGATCAA
GGCATAGGAGTAGGATGAGTTGCAGGTGTAACCGTATTACGTGCTTTTCGCGCGGGTTTGTTGCCTATTATTGAAGCA
ATATGTATTAATTGCTTAGATCACATCCACTGGCTAATCCAATCATTTTCGCCGCATATCCAAGCAATCGACTGAACC
AAATCAGGTACGTGTCTAGTGTTTTCGGCACCCACACGTAAAGGCCAACTTCTGGGAGACTCCTTATAACAACAAGCT
CGGCATTGTAGACTAGTAACACAACGGGACATTTACCAGGAGCTAGAACCGCAACATGAACAAATCAACCTCAGCCAA
CTTGAGGAATTCGGGCACCCATGGCAATCCAGCTCGGAGGAGTAACCCCCAGGACGAGGCATTGTTCACCTGCGGACT
TGGTACTCAACCATACCACTAACCACTGGGGACACGAGTTGCTACAATCCCATGAGCTGGCCGGTGGCAGCGTACGA
CATTTAGACGAAGGATTAGTAAAGATTCAAGATACGCAAAGTGGGCCATTATGGTAATTGACTCAATGCTGACACTCT
GGGGGTAATAACTCCCTGCATATACAATTGCATCAGACATCTTTACGCGGGTTAGGTCGCGGAATCGCCCAAGAATAC
AGGGAATTACTATTCCGCTTTTTAGACATATATCCAAGTATAGAAGATCCCAAAACAGCAGGTGCTTCGTTATAGGAT
CTATCATCTCATGGTCATGTTAACCAAATAGTTTACTGTCATGAGGTACGTCACTTTACCATGTTAGAGCGGCAATAG
AGAACGAGTACGTCTATACAGTCGGCATTACCATTTTCCTCGCTACAAGCATATCAGCTGAAGCTAGCAGCTTACAAG
TCTTAAGTCAACCTAGTCCTAGCCGATGCGAGTTACGTAGTTAGCGGATGTCTGGCAGTACTGGGGTTAATTACGGCC
GAGTGGACGGTAGAACATGGGGTGAGATGTTTAGTACTCACCCGACGTCGGGAGCGATCAGAAAAGGGTCAACAATCC
AGTGAACCATTGCAGAAGCCAGGGGCGGAAATATTAGTCCAGTGGGGCGGTATTTCCCGACACGAAAGTGTGACAAGG
ATTCTAGAGACATTCAAAGGATCCTTGCCGGCCTTACGAGAAATGATTCCTACTGCTGGCATATTAGATCATGATTTG
CGGTCAAACATGACTGGGGAACCATTTACACGGGTAATGGCGCCCAAAGTACTAGGTGCTTGTCATGTGCATACCTTG
ACTCATAATGTACCGTGGGACTTTGTTCTTTGTTCTTGCTCTATGCCTTGAATAATCGGTTCGCGTGGCCAAGGGACT
AATGCGGCTACTAATGCATACATGGTTAGTTTAGCCGATCACCGACGAGGTACGGGCTTAGCTGGCTGGAGGATTATC
TGGGAACCATGGACACGAGCGGAAATGGCACCTAATTTGCATTGTCCTCATCGACATAGTATGCTGTCCATGGGTATG
ACTATTATGTCTATAGAACAGTGATTCCAGCTTATAGGACAGTTACCCGAACAGTCGATACCACGAGTCGCAGTGCTA
CCACTTGAATGGTCACTGATCCAAGAACATTTTAGTTGTGGTACTCAAGTACCACTGCGGTCCTAGTTGGTGACAGAA
AGCATATCACGGCACCAAGCCCTCAATTCAAAGACATAGCAGAATGAAGTTATAGGATAGCTAACAGCAGCTTTACCA
GGACAAGGAGCAAAGCATATGATAATGTACATTATAGATGCAGTTTCCCGAGTACTATCTCTGAGCAGTTATCAAAGT
CATATGCATCAGCTCCTGAGCAGTATGGCCCTTGATTCTCAATTGGCTGTCCAATTGCACATTACGCTACAAGCTGAC
ATGCTGGTGGAGATAACTATACTCAGATTTATACCAGATTTCACTATCGTTGCTATAGCCAGTGATGTGCATGAGGAA
CTGACCCTAGTTGCTTAGCATCAACGAGATGAGTCAGCATATACCGGGCAACTCTACGATAGCATTAGGTAAGCAAGC
GAGCGGATTAGACGTCAATAATGA (SEQ ID NO:37)

FIG. 7-3

ATGGGGCATAGCGCTGGGGAATATGTTGCAGCAACAGTACCATGAATTTTTAGATTAGGAGATGGTTCAAAACAGATT
GCTCAGAGAGCAAGACTTATGCAACAATTACCGTCTGGCGGTTAAATGTTTTCTGTAAAGGCTTGAATCGTAAAAGTG
AATCAACTCATTGCACTATACTCTCAATAAGTAGCGGTCGCATCGATTGACCGACCTCAAAGCATTGACATGTCTGGC
GAGGCAGTAGCAAATGGAGCGGCTGAAAAAAGCTTGGAACCAGAAGACATAAAGACAAAATGACTGCAAGTGTCACAG
GCATTCCTTTCACGTTTGTTGGACCCAATGTTAGCGGACATTGAAGCGGTAGCCTCAGAATTAACCTACAGTCAAGCA
AATATTTCATAAGGATCAAATGTAAGGGTAGCTAAGGCAGCGAATAGTATTTCCAAAGCAACCTATTGGTTATATCGT
GTCCGGCAACCGGTGAATTTTGCGCAATGTATGGTCACAATATAGCAAGAAGATTATTCCTTCTTGTTAGAAATTGCA
CACAAATCAACTTTGTAAGGCGTGGGCAGACAGTGCTTGCCAGATGATGTTGTAGTATGGGTGCCTCCGTTGAAACCA
GGTCAAGAATACTTGCAGCAGATGCTCCAAAGTATGGCTTAACCATTTGTGCATGCAGTTTAAGTTGATTGGTTTGGG
TTTAATAAGGATTATTCTCCTAGTAAAGTAGTATGGCCGATTTATCCCTATCAACCGCAACGATATTGGAGTGCGACA
AATTATAATCTAAAACAGCAGAAACAGCTATTATCAAATCATAAGAATCCTCACCCTCAACTCGGTGAAAGATAACAT
TCTGCAGCCTTAGTACTGCAAATTCATTTTGAGTGTCGAATTAGTGCATCTCAAGCAACTTACCCGCAACACTACTGG
GTTTTTTTTCAGCCTCTTTTCTCAGCAGTAGCTTGCTTCGAAAAAGCCTTACCAGCAGGTGCAATTATATTCAAGTCA
GATGATTTCATCCTATAAGGTATAGCAATCCCAAAAGTATTTATTATATGAAACGATGAAAATAATGCAATTCCGATA
GTATTGAAATTACATTTAGTGCAAAGCCATAAATACCAAATTCTCAGTTTGGATGTAATCACTAGTTCTCCAAAACCC
AAATGGATTCTACGTATTGAAAGAAAATTATTAGAAGGTAGTAAAGCCTCCCAATTAAAAACAACAAACTTAGAAGCG
CTTTAAGACGGGTATTACCTACAGATAATACCTGCTGAATTCTCCTAAAAATTTGAAGGATGCGGTCTTATTTACGGA
TCGTCTCTCCAAGCCTTTAAACAAATGTGGCACACCGAAGTAAAGGCACCAGGTAAAATTCCGTGACCAAAAACTGAG
GTAAATGTGGCATCTTCATACGAACTGCACCCAAATCTATTAGATCCTAGCTTCCGGGTGTTTGCTGCAGTAATGCGT
AAATCGGACAGCCAAGAAGCTAATTGGCCATTGGAAATACAACGACTACAATTTTATCGCAATGGTGGTAACAGTATG
AGGACTCGAGTAGCGATAGGTGCTACAGAAACTAGTACACAGACTTTAAGCGGCAAGGATTGTTTACTGGATGAACGA
GGAACAGTAGTAACAAGAGTTCAAGGTTTATCTTTATAACGTACTACTCGCCAGGCTTAGTTACGTGATATTCAACCT
AAATTTAATAAATGGTTATGTCAAATGCATAGGCAAACCCGATCAATCTCTCCGCATAACCAAACAATTGACTTGACA
AATTCAGGAAGGTGGTTATTGTTTCCCCCACTCACAAGTATAGGCAAGCATCTCGTAGTAACCTTACAACGACAAGGA
TGGCATTGTGTATTAGCAACACCTGGGGAAGATTAGCACCAGTTAGAATCACTACATTATCAATTCAACCCCTACCAT
CCAGAGGGATTCCTGGACCTATAGCAATCCAGCTTGGAACAGGAACCGCCATAACGAGGAGTTATTTACCTGTAGAGT
TACGACTCAACAATTGCACAAAGGGCTGGGGCACACGACTTGCTAAATTCCCAAGGAGTGGGCTTTGGCAGCGTGCTT
CCTTTAGAGCAAGCCATAGTAGAAAACCAAGATATGGAAAGTTCCCCATTGTGGTTACTGACACAACGCTCACAGTCT
TTGGGTCATGAGTCCCTTCCTGTACCATTCCAACATACACCGTTATGGGGATTACGTCGGGTAATTGCCTAGGGACAT
TGGAAATTACATTGCCCGTGTTTTGACTTAGGTCCAACTATATAAGATTGCGAAACAGTAGCTTCTTAGTTAGAGGAA
ATATTATGTCCTGGTTATCAAATCCAAATTGATTACTGGCAAGGGGTGCGTCACGTTTCCCGGTTAGGGCGGCAACAA
CATATGAGTGCATCTACATAGTCAGGATAACTAATTTCCTTGCAACACCCGTTTCTACTGAAGCTAGCAGAATGTAAC
TCTTTAGGCAACCTACTCCATGCCGGAGCCAGTTAGTTAATTACAGGACGTCTGGTAGCACTGGGGTTGAAAACTGCT
GGGTGGATGGTCCAACAAGCGGTTAAATTTTTACTACTTTCCGGTGGTAGCCAGCCATCTGCAAAAGCTCAACAAAGC
ATTGTACAATTACGGAAGGCAGGACCGCATGTGTTCGTCATGTGTGGAGAAAATTGCCAACAAGATAAAGTGGCAGCA
ATTATATAGTCAAGCAAACTATCTTTGCCACCATTACGTGGTATAATTCATGCTGGTGGGAAATTGGGTGATGGTATG
CTCTTAATCATGAGTTGGGTAAAATTAACACAGGTGATGGCACAAAAAGTACAGGGGGCCTGGCGTTTGCATTATTTG
ACTGAGAATGTACGTTACGACTTTTTTGTGTGTTATTCCGGTATGGTTTCAATATTGGGTACGCCTCGTCAAGGGGAT
TATTATGCTGCGAATGCTTCCATGGATGGTTTAGCTCATCGTCGACGGGGTATGCGTTTATTTGGCTTGGGCATTAAG
TGCGGACTATGGCCACAGGAGGGATTGGCAGCGAATTTGCATAGTCCTCAACAAGGTAGAAAGGTGTCCAAGGGAATG
AGGTTCTTGTCATCAGAACAGGGATTCCAGCTTCTAGGTCAAATACTCGAAGAATCTATAACACAAGTACGAGTCCAA
CCAGTCCAATGGTGAGTGATCGAAGAGCAATTTAGTTGTGGTAATGAAATACCATAGCTCTCCCGATTGGAAAAGGAC
AGCATATCTCAGCAAAAAACCCTCAAGACCAAGACTAAGCACAATGAGTTTATAGAACAGCTTAAGGCTGCTTTACCA
ACAGAAAAAGGAAAGCATTTGAAAATTGACACTAAAGATGAAGTTTCTAAAGTGCTTACTTTGAGCCCTTCTGAAATA
GATATGCATCAGCGCCTGAACTCTATGGGGCTTGAATCTCTAATCGCTGTAGAAGTGCACATTAGCCTTCAGACTGAC
TTGCTGGTGTATATATCAAGAGTCTAATTTACAGAAAGTATCAGTACCGTTGGTTTAGCCACTGATGTGAATGGGCAA
CCGAGCCAAGCTACTCACAATCAAGGTGTTAAGTCAGGAAATCAAGGGCAGCTTTACCAAAACAATACGAAAGATAAC
GTGCGGGTAAGAGGTGAAATATGA (SEQ ID NO:38)

FIG. 7-4

ATGGGGCATAGTGGTGGGGAATATATCGCAGCCACTGTAGAAGGAATATTTAGGTTAGAAGATGGCTTAAAACTTATT
GCACATAGAGGAAGACTAAGGCAACCGTTACCCTCTTGGGGTGAAATATTATCTGTGATGGCTTCACTTGAAAAGGTA
ATTCAAATAATTGCACCAGACTCTCAAACAGTAGCGATCGCATTGATTAAAGGACCCCAAGGCATTGTCACTTCTGTT
GAGGCAGAAACAATTGGAGCGGGTCAAAATAGCCTAGAAGCTGAAGACATAAAGACAAAGCGACTGCAACTATCCCAC
GTATTCCATTCAAATTTGATGGAGCCAATGCTGGCGGACTTTTAAGCAGTAGCAACAGAAATAAGCTACAATCACCCA
AATATTTCATTAGAATCAAATGTGACGGGAGCTCGGGCAGAGATTAGTATTGCAACAGCAAGCGATTGGGTAACTCAT
GTCCGTCAACCGGTAAAATTTGCCGAAAGTATGGCCACATTACATCAAGAAGGTAATTCCATCTTGTTAGAAATTCGA
CCCAAACTAACTTTGTAAGGCATGGGGAGACAGTGCCTGCCAGAAGTTGTGGGAGTATGGTTGCCTGCTTTGAAACCC
GGTCAAGAATACTGGCAGCAAAAGCTACAAAGGTTGGCTGACCTATATGTGCTTGGAGTAAAAGTTGATGGGTTAGGG
TCTGATAAAGTTTATTCTCGAAGCAAGGTAGGATTGCCGACTCATCCCTTTCTACGGCAACGATATTGGATGGAGACA
AATCATAATCTAATTCATCAAAAAAAGTTTTTAGCAAATCATAACAATCTTCACTCTCTACTCGGACAAAGATTAGAT
TTAGCACCCTTAGAACTGCAAATTCGATTTGAATGTGAAATTAGTCCTTCTCAACTAACTTACCTACAACACCACGGT
GTTTTTCCTCAACCTGTTTTTCCAGCAGCAACTTACTTGGGAATAGCCTTCGCAGCAGTTTCAATATTATTCAATGCA
GATGATTCAATCCTAGATGATATAGCAAACCAAAAAGTGTTAATTTTACCAAAGGATGTAATTAATACAATACAGATA
GTTGTAAATTTACCGTTAGTACAATGCTATAAATACCAAATTTTGAGTTTGGATCTAAACACTATTTCTTCAAAACCT
AAAGGGATTCTACCTATTGAAGGTAAAATATTAATAGGTAATAGAGACCCCCACTTAGAAACATCAAACTTAAAAGAG
ATTAAAGAGGAGTATAACCCACAGATATTTCCTACTGAAATCTACCAAAGATTTGAAGCATGGGGTCTTTATTACGGT
TATTCTTTCCAGGCCGTTAACCAACTGTGGTACAGCGAAGAAAAAGCACTGGGTGAAATCCAGTTACCTGAAACTGAG
GAGAATGTTGCAGCTTTATCCCAACTGTACCCAATTCTATTAGATGCTGGCTTCCAGGCGTTAGCAGCTGTTATGGGT
AAAACAGACAACCGAGAAACTTACTTGCCATTGTAAATAAAACAACTACAAATGTATCGGAGTCGTAGTAATATTTTG
TGGACACAAGTAGAGGTAGGTGCACCAGAAACTATTAAACAAACATTGAGCGGTGAAGTTTGTTCATTGGATGATCAA
GGAATAATAGTAGCAAGGGTTGAAGGTCTAACTTTATTACGTACTTCACGCGAGGCTTGGTTGCGTACTATTGAACCT
AAATTTAAAAATTGGTGATATCAAATCCCTTGGCAAACTCAATCAATATCACCCCATAGCCAATCAATCGACTTAACA
ATATCAGGTAGATGGTTATTGTGTTCCCCACCTACAGGTATGGGCAAACATGTGGTAGAATGCTTAGAACAGCAAGGT
TGGGATTGTATATGAGTAACACCGGGGGAAAATGACCAGCAGTGAGAATCTCAGCATTATCAAGTCAACCCCAGCCAT
CCTGGGGAATTCCGGCACCTATTGGAATCAAGCTGGGAGCAGCAGCCCCCATTAGGAGGAATTATGCACCTGTGGGGT
TTGGACTGAACAATAGCGCTAAGGACGGGGGCACAGGGGTTGCAAAAGTCCCAAGAACGGGGCTGTGGGAGCGTACTT
GATTTAGTCCGAGCCTTAGTGAAAAATCAAGGTATGGAAAGGGCCCCATTAGGGTTAGTGAGTCAAGGCTCGCAATCT
GGGGGTAATGGGTCCCTTCCGATACAATTCGAACAAACACGTTTATGGGGGGTAGGTCGAGGAATTGCCCAAGAACAT
AGGAAATTACAAAGCCGGTGTTAAGACTTAGAACCAACTATGAAAGATTCCAAAACAGTAGATGCTTTGTTAAAGGAA
CTATAATCTCCTGGAGATGAAAACAAAATTGCTTAATGTCAAGGGATACGTCACGATGCCCGGTAAGAGCGGCAAAAA
AAAATGACTACATCTACCCAGTCCGGTTTACAAATTTTCTCGCAACATCCATTTCAATTGAAGCTATTAGAATATAAT
TCTTTAGACTACCTAATCCTAGCCGAAGCTAGTTACTTATTTACCGGAGTTCTGGGAGCTCTGGGGTTATAAACCGCT
GTGTGGATGGTTCAACAAGGGTTCAAATATCTTGTACTTACTGGACGTAGGTAGCCATCAGCTAAAGCTCAACTAACC
ATTGATCAATTACAGTAGGCAGGAGTGCAAGTATTTGTCCTGTGTTGAGATATTTTCCAACAAGATAATGTGGCTAGA
ATTATTGAGTCAATCTAAGTATCTTTTCCAGCATTACTAGGAATATTTCATGCTGTCGGGATATTTGATGATGGTTTG
CTGTTAAATATGAATTGGGTAAAATTTAAACAGGTGATAGCACCAAAAATACAAGGGGATTGGCATTTACATAATTTA
ACTCAGAATATACCTTTGAACTTTTTTGTATGTTTTTCCACTATGGCTTAAATATTGGGATCGCCTGGTAAAGGGAAT
TATGCTGCTGCAAATGCTTTCAAGGAAGGTTTAGCCAATCATCGACCGGGTATGGGCTTACCTGGCCTGAGCATTACC
TGGGGACCCTGGGCACAACAGGGAATGGCCGCAAATTTCGATAGTCCTCCTCAAGATACAATGGTGTCCCAGGGAATG
CCTTTTTTGTCCTCAGAACACGGATTGCAGCTTCTAGGACCATTACTCGACCAATCCATACCCCAAGTAGCAGTCCTA
CCCATTCAATGGCCAGTGTTCCCAGAGCAATTCAGTTTTGGTCATCAAATACCCTTGCTGTCCCCATTGGTACAAGAA
AGCACATCACAGCACAAAGCCCTCCAAACAAAGACCAAGCACAACGAATTTTTAGGACAGCTAAGAGCTGCTTTGCCA
AGAGAAGGAGAAAAGCGTTTGATATTGTACATTAAAGGTGAAATTTGTCAAGTACTGTCTTTGAGCGCTTCTCAAAGT
GATATGCAGCAGCCCCTGGACACTATGGGGGTTGATTCGCTAATGGCTGGGGAATTGCGCAATAGGCTGCAAACTGAC
GTGCTCGTGGGTATATCTATGGTCAAATTTGTAGAAGATAGCAGTATCGTGGATTTAGCCGCTGAAGTGAGTGAGCAA
CTGGGCCAAGTTGGTCAGAATCAGGGAGTTGAGGCAGAAAATAGTGGGCAACTGTACCAAAGGAATAGGAAAGGAAAC
GAGCGGGTAAGAGGGGAATTATGA (SEQ ID NO:39)

FIG. 7-5

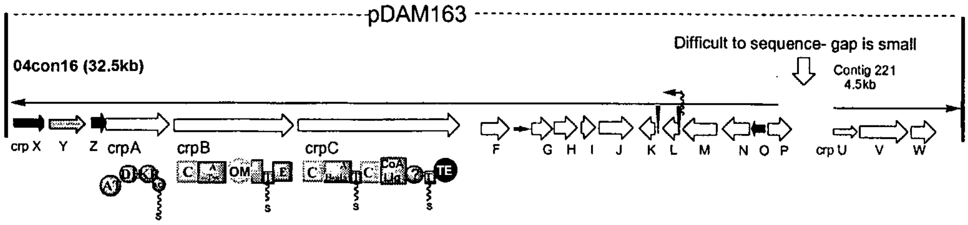
```
ATGGGGCATGGTGCTGGGGAATATGTGGCAGCCACAGTAGCAGGAATATTAAGTTAAGAAGATTGTTTAAAACTGATT
GCTCATAGAGGAAGACTCATGCAACAGATACCCTCTGGGGGTAAAATGTTATCTGTAATGGCTTCAATTGGAAAGGTT
AATCAACTAATTGCACCATACTCTCAAAAAGCAGCGATCGCATCGATTAACGGACCCCGAAGCTTTGTCATTTCTGGT
GAGGCAGAAGAAATTGGAGCGCTTCAAAAAAGCTTAGAAGCAGAAGACATTAAGACAAAACGACTGCAAGTAACCCGC
GCATTCCTTTCACATTTGATGGAACCAATGTTGGCGGCCTTTGAAGCAGGAGCATCAGAAATAACCTACAATCAACCA
AATATTCCATTAGTAACAAATGTAACGGGAGAAAGGGCAGAGAATAGTATTGCCACAGCAAGCAATTGGGTAAATCAT
GTCCGGCAACCGGTGAAATTTGCCAAAAGTATGGACACATCACAGCCAGAAGGTTATTCCATCTTCTTAGAAATTGGA
CCCCAACCACCTTTGTTAGGCATGGGAAGACAGTGCTTGCCAGAAGATCTGGGAGTTTGGTTTCCTTCTTTGAATCCA
GGTCAAGAAGACTGGCAGCAAATGTTACAATGTTTGGCTGAACTATATGTGCATGGAGTTAAAGTTGATTTGTTAGGG
TTTGATAAAGATTATTCTCGTAGCGAGGTAGTATTGCCGACTTATCCCTTTCAGGGGCAACGTGATTGGATTGAGACA
AATAATAATCTAATACAGCAAAAACAGTTTTTATCAAAACAAAAAAATCTTCACCCTCTACTCGGACAAAGAATACAT
TTAGCAGCCTTAGAACAGCAAATTCGTATTGAATGTCAAATTAGTGCTTCTCACCCAACTCACCTGCCACACCACTGT
GTTTTTTCTCAACCTGTCTTCCCCGCAGCAGCTTACTTGGAAATAGCCTTAGCAGCAGGTTCAATTTTATTCGATGCA
GATGATTTAATCCTAGAAGATATAGCAATCCAAAAGGTATTAATTGTATCAAAGGATGAAATTAATACAATTCAGATA
GTTTTAGATTTACAGTTAGTATAAAGCTTTAAATTCCAAATTTTCAGTTTGGATATAAACACTTATTCTTCATAACCT
AAATGGATTCTACATATTGAAGGAATAATATTAGTAGGTGATAAAGACCCCCAATTAGAAACAACAAACTTAAAAGCG
AGTAAGGACGAGTATAACCAACAGATATTACCTACTGAATTCTAGCAGAAATTAGAAGAATGGGGTCTTAATTACGGT
TCTTCTTTCCAAGCCATAAAACAACAGTGGCACAGCGAAGGAAAAGCACTAGGTGAAAATCAGTTACCAGAAACAGAG
ATGAATGTTGCAACTTTATACCAACTGCACCCAATACTTATAGATGCTAGCTTCCAGGTGTTAGCAGCAGTTATAGGT
AAAACGGACAACCAAGAAGGGGATTTGCCATTGGAAATAAAACGACTACAAATTTATGGGAGTGGTAGTAATAGTTTG
TGGACTCAAGTAGAGATAGGTGCAACAGAAACTAATAAACAAATTTTGTGTGGTAAAGTTTGTTTATTGGATAAACAA
GGAATAGTTGTATCAAGAGTTGAAGGTTTAACTTTATTACGTACTTCTCGCGAGGCTTTGTTAAAAAAAATTGAACCA
AAATTTAATAATTGGTTATATCAAATCCATTGGCAAACCCAATCAATTTCACCCCATAACCAATCAATTGACTTAACA
AAATCAGGTAGGGGGTTGGTGTTTTCCCCACCCACAGGTATAGGCAAACATCGGGTAGAATCCTTAGAACAACAAGGT
TGGCATTGTATATTAGTAACACCAGGGGAAATTTACCAGCATTTAGAATCTCAACATTATCAAATCAACCCTAACCTT
CCTGAGGAATTCCTGCACCTATTGCAATCAAGCTTGGAGTAGCAACCCCCATAACGAGGAATTATTCACATGTGGAGT
TTGAACTCAACAATAGCACTAAGGACTGAGGCACAGGAGTAGCAAAAATCCCAAGAACTGGGCTGTGGCAGCGTCCTT
CATTTAGTCCAAGCCTTAGTACACAATCAAGATATGCAACGTGCCCCATTATGGTTAGTGACTCAAGGCTCACAATCT
GTGGGTAATGAGTCCCTTCATATACAATTCCAACAAACACCTTTATGGGAGTTAGGTCAAGTAATTGCCCAGGAACAT
AGGGAATTACAATGCCGGTATTTAGACTTAGATACAACTTTGGAATATTCCCAAACAGTAGCTGCTTTGTTAGAGGAA
CTATTATCTCCTGGTGATGATAACCATATTGCTTACTGTCAAGGTGTACGTCACGTTGCCCGTTTAGAGCGGCAACAT
ATAATGAGTACATCTACATAGTCCGGATTACTAATTTCCTCGCAACAACCATTTCAACTGAAGCTATCAGAATATAAG
TCTTAAGACAACCTAATCCAAGCCGAAGCCAGTTAATTAATTACCGGAGGTCTGGGAGAACTGGAGTTAAAAACCGCT
GAGTGGATGGTACAACAAGAGGTCAAATATTTAGTACTTACCGGACGTAGGCCGCCATCAGCAAAAGCCCAACAAACC
ATTGAACACTTACAGACGGCAGGAGCGCAAGTATTAGTCCTGTGTGCAAATATTTCCCAAAAAGAAAATGTGGCAAGA
ATTATAGAGTCAATCAAAGTATCTTTGACAGCATTACAAGGAATAATTCATGCTGCTGGGAAATTGGATGATGGTTTG
CTGTTAAACATGAATTGTGATAAATTTACACAGGTGATGGCACCTAAAGTACAATGGTCTTGGCATTTGCATAATTTG
ACTCAGAATCTACCATTGGACTTTATTGTTTGTATTACCTCTATGGCTTCAATATTGGGTTCGCCTGGTCAAGGGAAT
TATGCTGCTGCTAATGCTTTCATGGATGGTTTAGCCAATCATCGACGGGGTATGGATTTACCAGGCTTGAGCATTAAA
TGGGGACCATGAGCACAAGAGGGAATGGCAGCAAATTTGGATAGTCCTCATCAAGATAGCATGGTGTCCAAGGGAATG
ACTCTTCTGTCTTCAGAACACGGATTGCAGGTTCTAGGACAATTACTCGAACAATCCATACCACAAGTAGCAGTCCTA
CCATTTCAATGGTCAGTGTTTCAAGATCAATTTAGTTTTGGTAATCAAATTCCATTGCTGTCCCAATTGGTAAAAGAA
AGCAAATCACAGCAAAAAGCCTTCCAACCAAAGACAAAGCACAATGAACTTTTAGAACAGCTAAAAGCTGCTTTACCA
AGAGAAAGACAACAGCTTTTGATAATTTACATTAAAGATGAAATTTGTCAAGTACTTTCTTTGAGCACGTCTCAAATT
GATATGCGACAGCCCCTGAACACTAGGGGGCTTGATGCTCTAATGGCTGTGGAATTGCACAATAGGCTACAAACTGAC
TTGCTCGTGGATAAATCTATAGTCAAATTTATAGAAGATATCAATATCGTAGATATAGCCACTGAAGTGAATGAGCAA
CTGAGCCAAGTTGCTCAGAATCAAGGAGTTGAGTCAGATAATATTGGGCAACTCTACCTAAGCAATAGGATAGTAAAC
GAGCGGATAAGAGGTGAATTATGA  (SEQ ID NO:40)
```

FIG. 7-6

ATGGGGCATAGTGCTGGGGAATATGTGGCAGCCACAGTAGCAGGAATATTTAGTTTAGAAGATGGTTTAAAACTGATT
GCTCATAGAGGAAGACTAATGCGACAGTTACCCTCTGGGGGTGAAATGTTATCTGTAATGGCTTCAATTGAAAAGGTA
AATCAACTAATTGCACCATACTCTCAAAAAGTAGCGATCGCATCGATTAACGGACCCCAAAGCATTGTCATTTCTGGT
GAGGCAGAAGCAATTGGAGCGGTTCAAAATAGCTTAGGAGCAGAAGACATTAAGACAAAACGACTGGAAGTATCCCAC
GCATTCCATTCACATTTGATGGAACCAATGTTGGCGGACTTTGAAGCAGTAGCATCAGAAATAACCTACAATCAACCA
AATATTCCATGAGTATCAAATGTAACGGGAGCTAGGGCAGAGAATAGTATTGCCACAGCAAGCTATTGGGTAAATCAT
GTCCGGCAACCGGTGAAATTTGCCCAAAGTATGGGCACATTACAGCAAGAAGGTTATTCCATCTTCTTAGAAATTGGA
CCCAAACCAACTTTGTTAGGCATGGGAAGACAGTGCTTGCCAGAAGATGTGGGAGGTTGGTTGCCTTCTTTGAAACCA
GGTCAAGAAGACTGGCAGCAAATGCTACAAAGTTTGGCTGAACTATATGTGCATGGAGTTAAAGTTGATTGGTTAGGG
TTTGATAAAGATTATTCTCGTAGCAAGGTAGTATTGCCGACTTATCCCTTTCAACGGCAACGTTATTGGGTTGAGACA
AATAATAATCTAATACATCAACAACAGTTTTTATCAAATCATAAAAATCTTCACCCTCTACTCGGTCAAAGATTACAT
TTAGCAGCCTTAGAACAGCAAATTCGTTTTGAATGTCAAATTCGTGCTTCTCAACCAACTTACCTGCAACACCACTGT
GTTTTTTCTCAACCTGTTTTCCCAGCAGCAGCTTACTTGGAAATAGCCTTAGCAGCAGGTTCAACTTTATTCAATTCA
GATGATTTAATCCTAGAAGATATAGCAATCCAAAAAGTATTAATTTTATCAAAGGATGAAATTAATACAATTCAGATA
GTTTTAAACTTACAGTTAGTACAAAGCTATAAATTCCAAATTTTCAGTTTGGATATAAACACTAATTCTTCAGAACCT
AAATGGATTCTACATATTGAAGGAAAAATACTAGTAGGTAATAAAGACCCCCAATTAGAAACAACAAACTTAAAAGCG
ATTAAAGACGAGTATAACCAACAGATATTACCTACTGAATTCTACCAAAAATCTGAAGAATGGGGTCTTAATTACGGT
TCTTCTTTCCAAGCCGTTAAACAACTGTGGCACAGCGAAGGAAAAGCACTAGGTGAAATTCAGTTACCAGAAACCGAG
GTGAATGTTGCAACTTTATACCAACTGCACCCAATTCTTTTAGATGCTAGCTTCCAGGTGTTAGCAGCAGTTATGGGT
AAAACGGACAACCAAGAACCTTATTTGCCATTGGAAATAAAACGACTACAAATTTATCGGAGTGGTAGTAATAGTTTG
TGGACTCAAGTAGAGATAGGTGCAACAGAAACTAATAAACCAACTTTGAGCGGTAAAGTTTGTCTATTGGATGAACAA
GGAATAGTAGTAGCAAGAGTTGAAGGTTTAACTTTATTACGTACTTCTCGCGAGGCTTTGTTCCGTAATATTGAACCA
AAATTTAATAATTGGTTATATCAAATCCATTGGCAAACCCAATCAATTTCACCCCATAACCAATCAATTGACTTAACA
AAATCACGTAGCTGGTTATTGTTTTCCCCACCCACAGGTATAGGCAAACATCTGGTAGAATCCTTAGAACAACAAGGT
TGGCATTGTATATTAGTAACACCAGGGGCAAATTACCAGCAGTTAGAATCTCAACATTATCAAATCAACCCCAACCAT
CCTGAGGAATTCCTGCACCTATTGCAATCAAGCTTGGAGCAGCAACCCCCCTTACGAGGAATTATTCACCTGTGGAGT
TTGGACTCAACAATAGCACTAAGGACTGGGGCACAGGAGTTGCAAAAATCCCAAGAACTGGGCTGTGGCAGCCTACTT
CATTTAGTCCAAGCCTTAGTAAAAAATCAAGATATGGAAAGTGCCCCATTATGGTTAGTGACTCAAGGCTCACAATCT
GTGGGTAATGAGTCCCCTCCTATACAATTCCAACAAACACCTTTATGGGGGTTAGGTCGAGTAATTGCCCAGGAACAT
AGGGAATTACAATGCCGGTGTTTAGACTTAGATCCAACCATGGAAGATTCCCAAACAGTAGCTGCTTTGTTAGAGGAA
CTATTATCTCCTGGTGATGAAAACCAAATTGCTTACTGTCAAGGGGTACGTCACGTTGCCCGGTTAGAGCGGCAACAA
AAAATGAGTACATCTACACAGTCCGGATTACAAATTTCCTCGCAACAACCATTCCAACTGAAGCTATCAGAATATAAG
TCTTCAGACAACCTAATCCAAGCCGAAGCCAGTTACTTAATTACCGGAGGTCTGGGAGCACTGGGGTTAAAAACCGCT
GAGTGGATGGTACAACAAGGGGTCAACTATTTAGTACTTACCGGACGTAGGCAGCCATCAGCAAAAGCTCAACAAACC
ATTGAACAATTACAGAAGGCAGGAGCGCAAGTATTAGTCCTGTGTGGACATATTTCCCAACAAGAAAATGTGGCAAGA
ATTATAGAGTCAATCAAAGTATCTTTGCCAGCGTTACGAGGAATAATTCATGCTGCTGGGATATTGGATGCTGGTTTG
CTGTTAAACATGAATTGGGAAAAATTTACACAGGTGATGGCACCAAAAGTACAAGGGGCTTGGCATTTGCATAATTTG
ACTCAGAATCTACCCTTGGACTTTTTTGTTTGTTTTTCCTCTATGGCTTCAATATTGGGTTCGCCTGGTCAAGGGAAT
TATGCTGCTGCTAATGCTTTCATGGATGGTTTAGCCCATCATCGACGGGGTATGGGTTTACCTGGCTTGAGCATTAAC
TGGGGACCATGGGCACAAGAGGGAATGGCCGCAAATTTGGATAGTCCTCATCAAGATACAATGGTGTCCAAGGGAATG
ACTTTTTTGTCTTCAGAACAGGGATTGCAGGTTCTAGGACAATTACTCGAACAATCCATACCACAAGTAGGAGTCCTA
CCCATTCAATGGTCAGTGTTCCAAGAGCAATTTAGTTTTGGTAATCAAATACCATTGCTGTCCCAATTGGTAAAAGAA
AGCAAATCACAGCAAAAAGCCCTCCAAACAAAGACAAAGCACAATGAATTTTTAGAACAGCTAAAAGCTGCTTTACCA
AGAGAAAGAGAAAAGCTTTTGATAATTTACATTAAAGATGAAATTTCCCAAGTACTTTCTTTGAGCACTTCTCAAATT
GATATGCAACAGCCCCTGAACACTATGGGGCTTGATTCTCTAATGGCTGTGGAATTGCACAATAGGCTCCAAACTGAC
TTGCTCGTGGATATATCTATAGTCAAATTTATAGAAGATATCAGTATCGTTGATTTAGCCACTGAAGTGAATGAGCAA
CTGAGCCAAGTTGCTCAGAATCAAGGAGTTGAGTCAGAAAATAATGGGCAACTCTACCAAAGCAATAGGAAAGAAAAC
GAGCGGATAAGAGGTGAATTATGA (SEQ ID NO:41)

FIG. 8

Crp gene cluster



Crp gene cluster legend

crpA- type I PKS
crpB- NRPS
crpC-NRPS

crpF- cytochrome p450 hydroxylase
crpG- gamma-butyrobetaine hydroxylase
crpH-aspartate decarboxylase
crpI- IS1327 transposase
crpJ- halogenase
crpK-IS892-orf2
crpL-IS892-orf1
crpM- ISRSO13 transposase
crpN- Benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase
crpO- pvdE type regulator?
crpP-Thioredoxin

crpU- acetyltransferase
crpV- large exoprotein involved in heme utilization
crpW- TPR repeat protein (hypothetical)
crpX-chorismate mutase/prephenate dehydrogenase
crpY- 2-hydroxychromenene-2-carboxylate isomerase
crpZ- 3-dehydroquinate synthase

| -inverted repeat

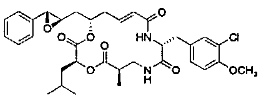
- E.coli consensus promoter

FIG. 9

Combinatorial Approaches to Novel Cryptophycins: Through Tailoring Enzymes Tolerance: Epoxidase

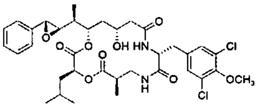
Hypothesis:
Epoxidase is not sensitive to Unit A Differences
Example 1: C6 -methyl is not required

Cryptophycin 28

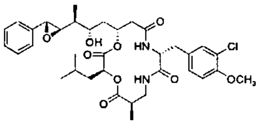


Hypothesis:
Epoxidase is not sensitive to Unit A Differences
Example 2:Polyketide oxidation pattern variation accepted

Cryptophycin 31

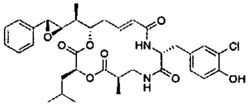


Cryptophycin 26

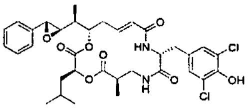


Hypothesis:
Epoxidase is not sensitive to UnitB Differences
Eaxmple 1: C29-OH: O-methylation is not required

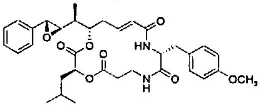
Cryptophycin 16



Cryptophycin 23

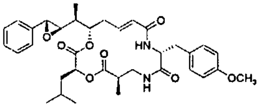


Cryptophycin 24

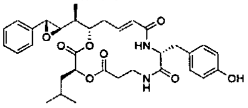


Hypothesis:
Epoxidase is not sensitive to UnitB Differences
Eaxmple 1: C28-Cl:Chlorination is not required

Cryptophycin 2

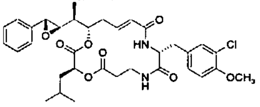


Cryptophycin 43

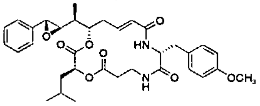


Hypothesis:
Epoxidase is not sensitive to Unit C differences:
21 -methyl is not required for epoxidation

Cryptophycin 21

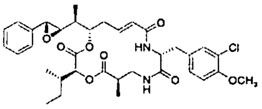


Cryptophycin 24



Hypothesis:
Epoxidase is not sensitive to Unit D differences:
Fatty acid occurs naturally

Cryptophycin 49



Hypothesis:
Epoxidase is not sensitive to ring contractions

Cryptophycin 26

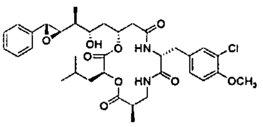
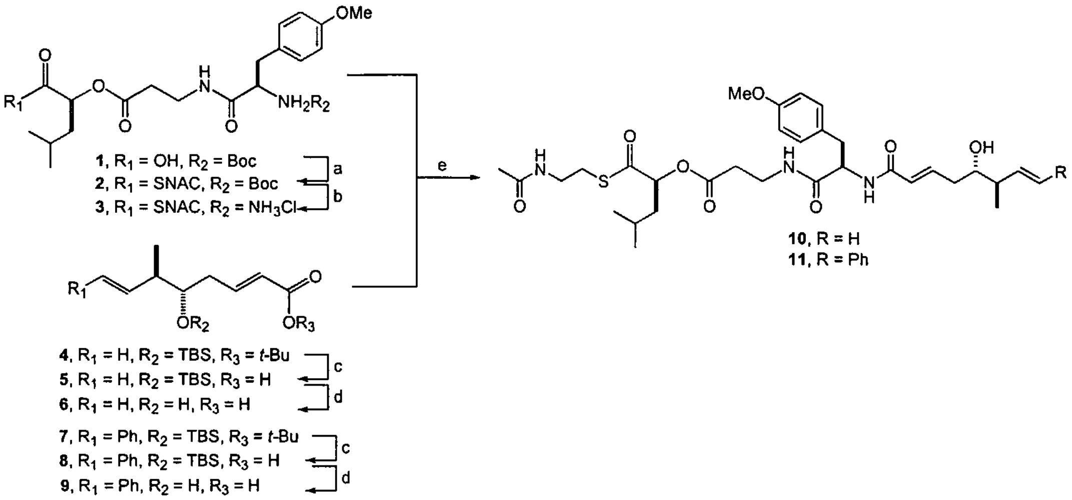


FIG. 10



1, R_1 = OH, R_2 = Boc
2, R_1 = SNAC, R_2 = Boc
3, R_1 = SNAC, R_2 = NH_3Cl

4, R_1 = H, R_2 = TBS, R_3 = *t*-Bu
5, R_1 = H, R_2 = TBS, R_3 = H
6, R_1 = H, R_2 = H, R_3 = H

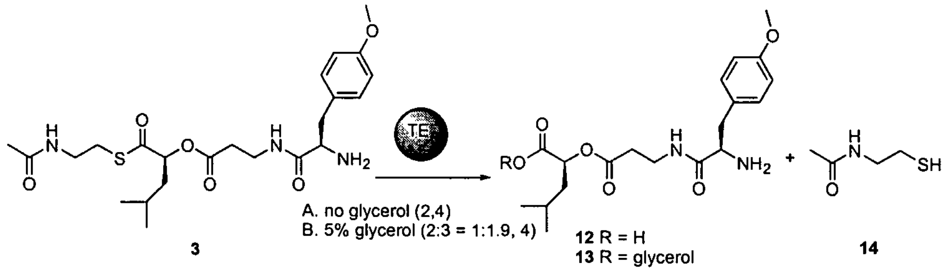
7, R_1 = Ph, R_2 = TBS, R_3 = *t*-Bu
8, R_1 = Ph, R_2 = TBS, R_3 = H
9, R_1 = Ph, R_2 = H, R_3 = H

10, R = H
11, R = Ph

---

[a]Key a) PyBOP, DIPEA, HSNAC, DMF; b) 4 N HCl/dioxane; c) 50:50:1 TFA:CH_2Cl_2:Et_3SiH; d) 20:1 AcCN:48% aq. HF; e) PyBOP, DIPEA (3.0 eq), **3**,and **6** or **9**.

FIG. 11



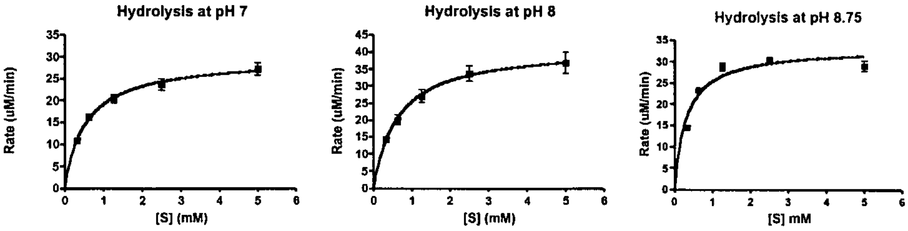
A. no glycerol (2,4)
B. 5% glycerol (2:3 = 1:1.9, 4)

3

12 R = H
13 R = glycerol

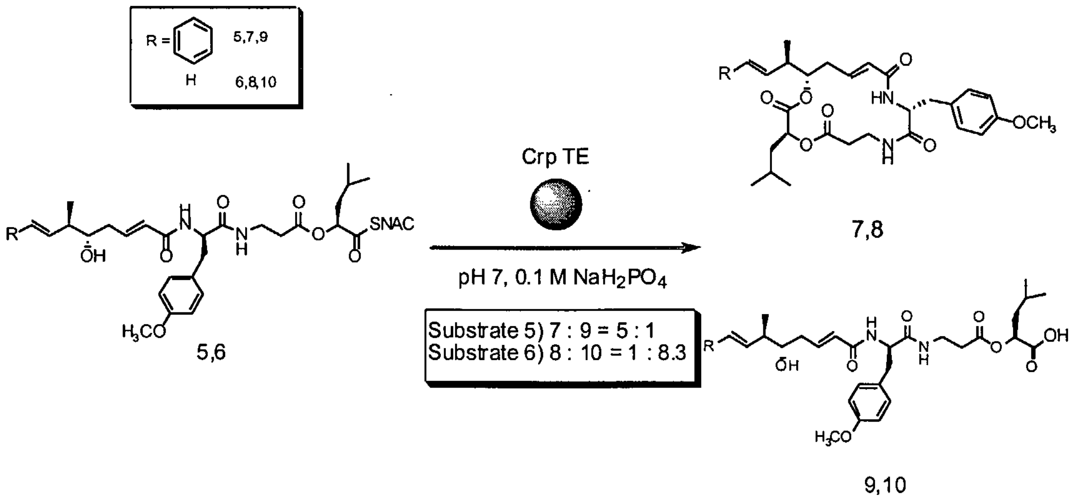
14

FIG. 12



k_{cat} (s^{-1})	0.35 ± 0.07	0.6 ± 0.1	0.39 ± 0.02
K_M (μM)	550 ± 70	620 ± 30	310 ± 50
k_{cat}/K_M ($mM^{-1}s^{-1}$)	0.64	0.97	1.26

FIG. 13


R =
5,7,9
H
6,8,10
R
ŌH
SNAC
H₃CO
5,6
Crp TE
pH 7, 0.1 M NaH₂PO₄
Substrate 5) 7 : 9 = 5 : 1
Substrate 6) 8 : 10 = 1 : 8.3
HN
OCH₃
7,8
OH
9,10