

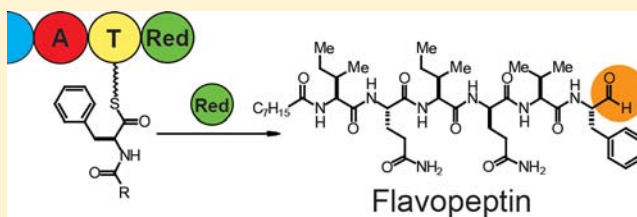
Proteomics Guided Discovery of Flavopeptins: Anti-proliferative Aldehydes Synthesized by a Reductase Domain-Containing Non-ribosomal Peptide Synthetase

Yunqiu Chen,[‡] Ryan A. McClure,[‡] Yupeng Zheng, Regan J. Thomson,* and Neil L. Kelleher*

Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United States

S Supporting Information

ABSTRACT: Due to the importance of proteases in regulating cellular processes, the development of protease inhibitors has garnered great attention. Peptide-based aldehydes are a class of compounds that exhibit inhibitory activities against various proteases and proteasomes in the context of anti-proliferative treatments for cancer and other diseases. More than a dozen peptide-based natural products containing aldehydes have been discovered such as chymostatin, leupeptin, and fellutamide; however, the biosynthetic origin of the aldehyde functionality has yet to be elucidated. Herein we describe the discovery of a new group of lipopeptide aldehydes, the flavopeptins, and the corresponding biosynthetic pathway arising from an orphan gene cluster in *Streptomyces* sp. NRRL-F6652, a close relative of *Streptomyces flavogriseus* ATCC 33331. This research was initiated using a proteomics approach that screens for expressed enzymes involved in secondary metabolism in microorganisms. Flavopeptins are synthesized through a non-ribosomal peptide synthetase containing a terminal NAD(P)H-dependent reductase domain likely for the reductive release of the peptide with a C-terminal aldehyde. Solid-phase peptide synthesis of several flavopeptin species and derivatives enabled structural verification and subsequent screening of biological activity. Flavopeptins exhibit sub-micromolar inhibition activities against cysteine proteases such as papain and calpain as well as the human 20S proteasome. They also show anti-proliferative activities against multiple myeloma and lymphoma cell lines.



INTRODUCTION

Many naturally occurring peptide aldehydes such as chymostatin, leupeptin, and fellutamide possess inhibition properties toward various proteases and are widely used in biomedical research to prevent unwanted proteolysis. Notably, some peptide aldehydes that inhibit cathepsins, calpains, or proteasomes—associated with human diseases including cancer, osteoporosis, and Alzheimer's disease—are under extensive study for their use as tools and lead compounds for drug development.^{1–4} By retaining the aldehyde warhead and varying the peptide backbone residues, chemists have designed and synthesized numerous peptide aldehydes and their boronate analogues that inhibit a spectrum of proteases with varying potencies.^{5,6} Despite this obvious biological and clinical importance, there has not been a biosynthetic pathway elucidated for a peptide aldehyde.

With the burgeoning of genome sequencing on microbial species, it has become increasingly apparent that only a fraction of their biosynthetic capacities for natural products have been reflected by direct observation of metabolic profiles.⁷ Many gene clusters for secondary metabolite biosynthesis have no corresponding products identified; these are known as orphan gene clusters. A number of systems biology approaches, including genomics, transcriptomics, and metabolomics, have been employed in attempts to decipher these orphan gene clusters for the corresponding natural products.^{8–11} A method known as Proteomic Investigation of Secondary Metabolism

(PrISM) allows for the screening of the expressed enzymes related to natural product biosynthesis.¹² In a PrISM workflow, microbes are cultured under various conditions, and their proteomes are analyzed by mass spectrometry. Expressed proteins for secondary metabolite biosynthesis are detected, which enables the biosynthetic gene cluster and the associated secondary metabolite to be discovered simultaneously. Two types of secondary metabolites, non-ribosomal peptides and polyketides, are of special interest for the PrISM method. The biosynthetic enzymes for non-ribosomal peptides and polyketides, non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), are often enormous in size (>200 kDa), which permits their targeted selection by size-based separation, such as SDS-PAGE.¹³ Using the PrISM approach, several new natural products and their corresponding gene clusters have been discovered from bacteria possessing either previously sequenced or non-sequenced genomes.^{14–16}

In this study, we employ the PrISM approach to screen expressed NRPSs and PKSs from *Streptomyces* species and identify an orphan NRPS gene cluster from *Streptomyces* sp. NRRL F-6652. Through bioinformatics analysis of the gene cluster and metabolomics analysis using mass spectrometry, a new class of peptide aldehyde natural products that we name flavopeptins was discovered and identified as the products of

Received: April 2, 2013

Published: June 13, 2013

the orphan gene cluster. The biosynthesis of flavopeptins involves an NRPS protein with a relatively rare C-terminal reductase domain catalyzing the formation of the peptide aldehyde. Like other peptide aldehydes, flavopeptins show inhibitory activities against cysteine proteases, including papain and human calpain as well as the human 20S proteasome, with low micromolar to nanomolar inhibition potencies. Flavopeptins also exhibit anti-proliferative activity against multiple myeloma and lymphoma cell lines.

EXPERIMENTAL SECTION

Materials. *Streptomyces* sp. NRRL F-6652 and B-16185 were acquired from U.S. Department of Agriculture, Agricultural Research Service, and *Streptomyces flavogriseus* ATCC 33331 was obtained from ATCC. Sequencing-grade trypsin was from Promega (Madison, WI). Stable-isotope-labeled amino acids were from Cambridge Isotope Laboratories (Cambridge, MA) or CDN Isotopes (Point Claire, Canada). Proteases were from Sigma-Aldrich (St. Louis, MO). Human 20S proteasome and its substrates were from Boston Biochem (Boston, MA).

Detection of NRPS Expression by Proteomics. *Streptomyces* sp. NRRL F-6652 was cultured in ATCC 172 medium (10 g/L glucose, 20 g/L soluble starch, 5 g/L yeast extract, 5 g/L N-Z amine type A, 1 g/L CaCO₃) at 30 °C for 3 days before being transferred to different screening media. Cells from 24, 48, and 72 h growth were harvested and lysed by sonication. The proteome lysate was separated by one-dimensional SDS-PAGE, and the proteome region above 150 kDa was excised for in-gel trypsin digestion. The resulting peptides were separated on a self-packed nanocapillary column (5 μm Jupiter C18, 100 mm × 75 μm) using a nanoLC-Ultra system (Eksigent, Dublin, CA). LC solvents were 5% acetonitrile in water with 0.1% formic acid (mobile phase A) and 95% acetonitrile with 0.1% formic acid (mobile phase B). The LC gradient was set as follows: 0 min, 0% B; 55 min, 45% B; 63 min, 80% B; 67 min, 0% B with re-equilibration of 0% B until 90 min. Peptides were eluted into a nanoelectrospray ionization (nESI) source on a 7 T LTQ-FTICR mass spectrometer (Thermo Scientific). Peptides were first detected in the FTICR with resolving power of 100 000 (at 400 *m/z*). Intact peptide data were collected in the 400–1800 *m/z* range, and MS/MS spectra were acquired using a data-dependent mode where the top three most abundant peaks from the FTMS full scan were selected for collision-induced dissociation (CID) fragmentation followed by mass analysis of the fragment ions in the linear ion trap. The LC-MS data were searched against all bacteria in the non-redundant protein database NCBI nr using Mascot.

Natural Product Detection and Structure Elucidation by Mass Spectrometry. NRRL F-6652 was grown in ATCC 172 medium for 3 days at 30 °C before being transferred to mineral-based media (MSB)¹⁷ supplemented with 10 mM sodium succinate and 0.05% casamino acids. Stable-isotope-labeled amino acids were added to the medium to a final concentration of 1 mM when necessary. Culture supernatants from 24, 48, 72, and 96 h of growth were analyzed by LC-MS. The samples were separated by a Phenomenex Luna C18 column (250 × 2 mm, 5 μm) eluting with acetonitrile/water with 0.1% formic acid from 0 to 100%. An Exactive Orbitrap of spectrometer (Thermo Scientific) was used for mass detection from *m/z* 200 to 1500. The LC-MS data were processed using Thermo SIEVE software for chromatographic alignment, feature identification, and quantitation.

Isolation of Flavopeptins. *Streptomyces* sp. NRRL F-6652 was grown in 30 × 500 mL ATCC 172 medium in 2 L baffled flasks at 30 °C for 7 days. The culture supernatant was separated from cells by centrifugation and extracted with 1 volume of ethyl acetate twice. The organic extract (~2.2 g) was dried down and then re-dissolved in 50% methanol in H₂O before undergoing solvent partitioning between 50% methanol/hexane (1:9) followed by 50% methanol/dichloromethane (3:7). The dichloromethane portion (~1 g) was separated by solid-phase extraction (SPE) using a Strata C18-E column (Phenomenex, Torrance, CA). SPE fractions eluted with 70–100% acetonitrile (~60 mg) were further separated on a Phenomenex Luna C18 column (150

× 10 mm, 5 μm) using 1:1 acetonitrile/isopropanol–H₂O with 0.1% formic acid to yield ~1.5 mg of flavopeptins. Purified flavopeptins were analyzed by LC-MS/MS on a 7 T LTQ-FTICR MS instrument.

Total Synthesis of Flavopeptins and Analogues. Flavopeptins were synthesized with H-Phe-H NovaSyn TG resin (EMD Millipore, Billerica, MA) using standard Fmoc coupling procedures. The resin (0.1 mmol) was allowed to swell in DMF (2 mL) for 30 min before use. Fmoc protecting groups were removed after treatment with 5 mL of DBU/piperidine/DMF (2:2:96) for 15 min. The amino acid chain was extended using a 3-fold excess of Fmoc-amino acids that had been previously activated with PyBOP (0.3 mmol), HOBT (0.05 mmol), and DIPEA (0.6 mmol) in DMF. After assembly of the target sequence, the resin was washed with DMF, *i*-PrOH, THF, and MeOH and then dried overnight. Side-chain protecting groups were removed by three treatments with anhydrous TFA (0.5 mL) in dry DCM (2.5 mL) for 10 min. The product was cleaved from the resin by three treatments with a mixture of AcOH/H₂O/DCM/MeOH (10:5:64:21) for 30 min to afford the crude product. The crude residue was purified on a Luna C18 column (150 × 21.2 mm, 5 μm) using acetonitrile–H₂O with 0.1% formic acid. Compounds 1 (6.8 mg, 8%), 2 (10.4 mg, 12%), and 3 (6.3 mg, 9%) were obtained through this method.

The carboxylic acid analogue 4 was synthesized with a Wang Resin (EMD Millipore). Fmoc-Phe-OH (0.1 mmol) was loaded onto the resin with HOBT (0.1 mmol), DCC (0.1 mmol), and DMAP (0.01 mmol). Standard Fmoc coupling procedures were used as described above. A mixture of TFA (1.0 mL) and DCM (1.0 mL) with TIPS (0.10 mL) was added to the resin for cleavage. The resin was allowed to shake for 3 h, after which the crude product was collected from the resin, dried, washed with cold ether, and dried again. HPLC purification was performed using the same conditions as above (19.4 mg, 22%). ¹H and ¹³C NMR spectra of the synthetic peptides 1–4 are shown in Figures S12, S13, and S16–S21. 2D NMR (TOCSY, HSQC, and HMBC) spectra of 2 are shown in Figures S14 and S15.

Labeling the Aldehyde Group with PFBHA. Flavopeptins isolated from NRRL F-6652 (~1.5 μg) were treated with 5 μL of pentafluorobenzylhydroxylamine hydrochloride (PFBHA, 10 mg/mL) in methanol at room temperature for 3 h and analyzed by LC-MS.

Determination of Amino Acid Stereochemistry. Flavopeptins (~100 μg) were dissolved in constant-boiling 6 N HCl (Thermo Scientific) and heated at 110 °C for 24 h for hydrolysis. After lyophilization, the hydrolyzed product was dissolved in 50 μL of water, 50 μL of 1 M NaHCO₃, and 100 μL of 1% Marfey's reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA), Thermo Scientific) in acetone and heated at 37 °C for 1 h. The reaction was quenched with 50 μL of 1 M HCl and diluted in 1:1 H₂O/acetonitrile for LC-MS analysis. The derivatized amino acid mixture was separated on a Luna C18 column (150 × 2 mm, 5 μm) using 25–60% acetonitrile/H₂O with 0.1% formic acid over 40 min and detected by UV absorbance at 340 nm and mass spectrometry using an Exactive Orbitrap instrument.

Biological Assays. To test the protease inhibitory action of the flavopeptins, the EnzChek protease assay kit (Invitrogen, Grand Island, NY) was used. Proteases were dissolved in their corresponding digestion buffer: for trypsin, chymotrypsin, thermolysin, and subtilisin, 10 mM Tris-HCl, pH 7.8; for pepsin, 10 mM HCl, pH 2.0; for papain, 10 mM MES with 5 mM L-cysteine, pH 6.2; for elastase, 10 mM Tris-HCl, pH 8.8; for calpain-I, 10 mM HEPES, 10 mM DTT, 1 mM EDTA, 2 mM CaCl₂, pH 7.2. The proteases (1 μg/mL) were incubated with flavopeptins or DMSO for 30 min at room temperature before an equal amount of substrate BODIPY-FL casein (10 μg/mL) was added. After 1 h incubation at room temperature, the fluorescence at 485/528 nm was read on a Synergy H4 fluorospectrometer (BioTek, Winooski, VT). To measure the IC₅₀ value against papain, the fluorescence over reaction time was recorded, and the initial rates were plotted against log [inhibitor] and fitted to the variable slope model using GraphPad Prism software. For calpain IC₅₀ calculation, flavopeptins were incubated with human calpain-I (20 nM) for 30 min before the substrate Suc-LLVY-AMC (40 μM) was added, and the fluorescence at 380/460 nm was continuously recorded for 30 min.

For the proteasome inhibition assay, human 20S proteasome (1 nM) in assay buffer (10 mM Tris-HCl, 0.035% SDS, pH 7.4) was

incubated with flavopeptins for 15 min at room temperature before substrates (Suc-LLVY-AMC for chymotrypsin-like and Z-LLE-AMC for caspase-like activities, 5 μ M) were added, and the fluorescence was recorded using a 380/460 nm filter. The initial rates were used for IC₅₀ calculation. Cell viability assays were performed using the CellTiter-Glo luminescent cell viability assay kit (Promega). Cancer cell lines (100 μ L) were plated in 96-well plates to a density of 5000 cells/well. Flavopeptins dissolved in DMSO were added after 24 h, followed by incubation at 37 °C for another 72 h. CellTiter-Glo reagent (100 μ L) was added to each well and the luminescence recorded.

RESULTS AND DISCUSSION

Detection of NRPS Expression by Proteomics. In previous reports, we described a proteomics-based approach called PrISM to screen for expressed NRPS or PKS proteins and applied this approach to discover new natural products from *Bacillus* and *Actinobacterium* species.^{12,14–16} In one such screen, the proteomics data for *Streptomyces* sp. NRRL F-6652 grown in MSB medium for 48 h identified peptides originating from two NRPS proteins from *S. flavogriseus* ATCC 33331 (Figure 1a).

The two NRPS proteins, weighing 265 and 512 kDa, respectively, are located next to each other in the genome and very likely within a gene cluster. However, despite an extensive literature search, we were not able to find any natural products associated with this orphan gene cluster.

Oligonucleotide primers were designed using the reverse translation of the detected peptide sequences and used to verify the genomic DNA sequence from *Streptomyces* sp. NRRL F-6652 is identical or similar to ATCC 33331 by PCR. All the PCR-amplified DNA sequences showed 100% sequence identity to the corresponding genomic regions of *S. flavogriseus* ATCC 33331. The Mascot search results of the F-6652 proteome sample showed that 81% of the identified protein families were coming from the ATCC strain, indicating a high sequence similarity between the strains NRRL F-6652 and ATCC 33331 (Table S1). Therefore, all the bioinformatics analysis was based on the genomic sequence from *S. flavogriseus* ATCC 33331.

Product Prediction Based on the NRPS Gene Cluster.

The 22.6 kb orphan NRPS gene cluster from *S. flavogriseus* ATCC 33331 encodes two modular NRPS proteins, FlavA and FlavB, which can be further divided into six NRPS biosynthetic modules (Figure 1b). Bioinformatics analysis of this NRPS gene cluster predicted the six adenylation domains incorporate Ile, Gln, Ile, Gln, Val/Ile, and Phe.^{18,19} The most salient features of the cluster are an apparent “starter” condensation domain often involved in lipopeptide biosynthesis (Figure S1)²⁰ and a thioester reductase at the final domain of FlavB. This reductase domain aligns with other known NRPS reductases, including the biosynthetic enzymes for myxochelin, saframycin, nostocyclopeptide, linear gramicidin, koranimine, peptaibol, and aureusimine (Figure S2). From FlavB, conserved NAD(P)H binding domains were identified, and the conserved catalytic triad of Thr-Tyr-Lys common to short-chain reductase (SDR) is present. The NRPS reductase domain has been associated with the reductive release of peptide thioesters from their biosynthetic enzymes, forming aldehydes, alcohols, or cyclic imines.²¹ On the basis of the domain organization, we predicted the product from this cluster was a hexapeptide likely containing isoleucine, glutamine, valine, and phenylalanine, an N-terminal aliphatic chain, and a C-terminal aldehyde, alcohol, or imine.

(a) NRPS Identifications from Proteomics Analysis of F-6652

Protein Description	Organism	Protein Accession #	Sequence Coverage (%)	# Unique Peptides	Size (kDa)
Amino acid adenylation domain protein	<i>Streptomyces flavogriseus</i> ATCC 33331	gij357410557	7.3	14	265
Amino acid adenylation domain protein	<i>Streptomyces flavogriseus</i> ATCC 33331	gij357410558	6	12	512

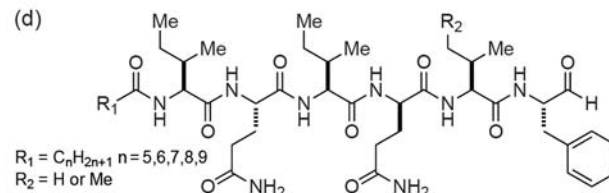
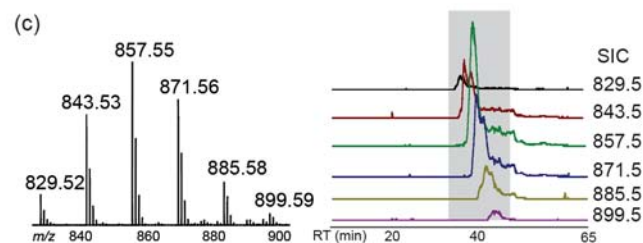
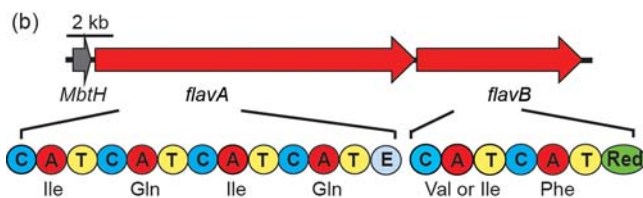


Figure 1. Sequential proteomic, genomic, and metabolomic analysis that led to the discovery of flavopeptins. (a) Two NRPS proteins were identified through LC-MS/MS analysis of the proteome of F-6652. (b) The genomic context of the two identified NRPS genes (*flavA* and *flavB*), their domain organizations, and the predicted amino acid substrates. A, adenylation; T, thiolation; C, condensation; E, epimerization; Red, reductase. (c) Detection of flavopeptins from culture supernatant of F-6652: left, mass spectrum created by compiling spectra of all flavopeptin species detected; right, selected ion chromatograms (SICs) for six flavopeptin species (m/z 829.5–899.5). (d) Chemical structure of flavopeptins.

In addition to the two NRPS genes, there is a small protein with 84 amino acids that belongs to the MbtH-like protein superfamily in the gene cluster (54% identical and 67% similar to MbtH protein from *Mycobacterium tuberculosis*). MbtH-like proteins have been found in many other NRPS gene clusters, and previous in vitro and in vivo studies have shown that they are required for the function of adenylation domains and correlated with high production of secondary metabolites.^{22,23}

Targeted Identification and Structure Elucidation of Flavopeptins. In order to identify the natural products arising from the NRPS gene cluster, we performed a targeted metabolomics analysis of F-6652 grown in MSB medium, the same medium in which the NRPS proteins were detected. In addition to monitoring metabolites that show increased intensities over time course, a predicted amino acid precursor with stable isotope labeling, *d*₅-phenylalanine, was added to the medium to monitor metabolites with heavy isotope incorporation. Six ion species (m/z 829.5, 843.5, 857.5, 871.5, 885.5,

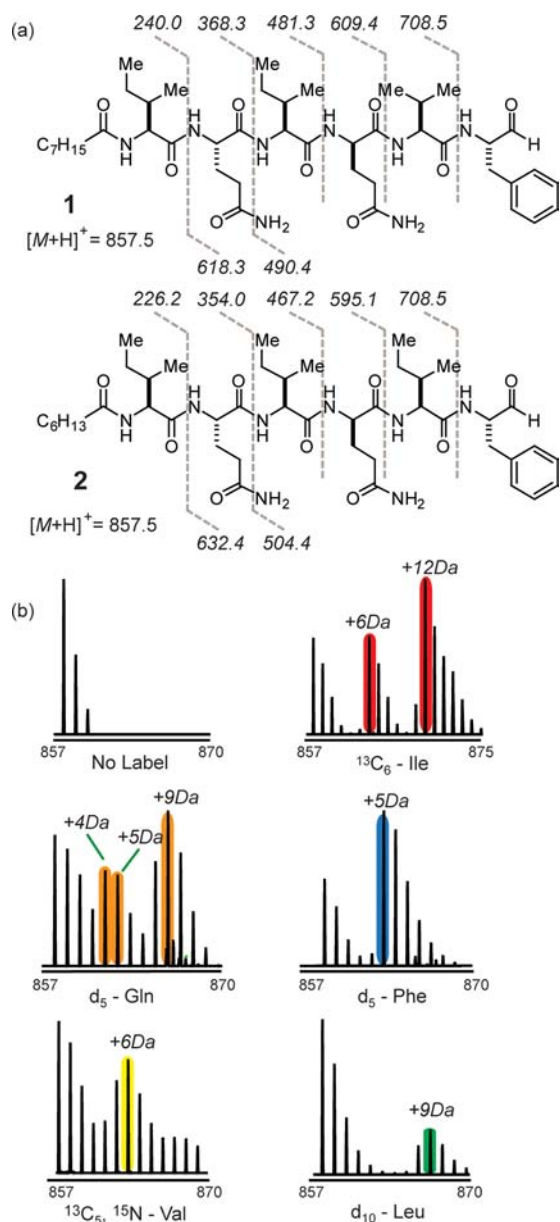


Figure 2. Partial structure elucidation of flavopeptins by MS and tandem MS. (a) Two isobaric structures of flavopeptin with mass 856.5 Da, with the observed MS² fragments shown. (b) Metabolic labeling of flavopeptins with amino acid precursors containing stable isotopes. Shown are the mass spectra for flavopeptins from F-6652 grown in MSB medium with no labeling and with ¹³C₆-isoleucine, 2,3,3,4,4-*d*₅-glutamine, ring-*d*₅ phenylalanine, ¹³C₅, ¹⁵N-valine, and *d*₁₀-leucine.

and 899.5, Figure 1c) showed clear evidence of phenylalanine incorporation as 5 Da mass shifts (Figure 2b). These six ion species also generated peptide-like fragmentation patterns, all yielding a common amino acid sequence tag of Gln-Ile/Leu-Gln-Val (or Ile/Leu)-Phe(CHO) (Figures 2a and S3). This sequence tag provided a reasonable match to the predicted amino acid substrates for the detected NRPS gene cluster. MS⁴ fragmentation on the target mass revealed the N-terminal structure to be an acylated isoleucine/leucine residue (Figure S4). The six masses are spaced by 14.0162 Da, which is unambiguously a CH₂ group, and their retention times increase slightly with the mass (Figure 1c), which leads to the prediction

that these six species are differentiated by the length of the N-terminal aliphatic chain. All six masses pointed to related lipopeptide aldehyde structures that we named flavopeptins (Figure 1d; cald 856.5422 Da, obsd 856.5427 Da, C₄₄H₇₂N₈O₉, for R₁ = C₇H₁₅ and R₂ = H).

To corroborate structural assignments, several metabolic feeding experiments were performed using stable-isotope-labeled amino acids. In addition to phenylalanine, LC-MS analysis of the labeled samples also showed the incorporation of two isoleucines, two glutamines, and one valine into flavopeptins (Figure 2b). When the sample was supplemented with ¹³C₆-isoleucine, +6 and +12 Da mass shifts were observed, indicating at least two isoleucine residues were incorporated into the final product. MS² fragmentation of these +6 and +12 Da mass species revealed that isoleucine can be incorporated at the first, third, and fifth amino acid positions (Figure S5). Similarly, the incorporation of the phenylalanine and valine into the predicted positions was confirmed. Labeling with 2,3,3,4,4-*d*₅-glutamine showed species with +4, +5, and +9 Da mass shifts (instead of straight +5 and +10 Da), indicating one deuterium from one of the labeled glutamine exchanged with environmental H₂O when incorporated into the final product. It is known that the first step of epimerization involves the removal of the α-H from the loaded amino acid to generate a carbanion. Thus, one glutamine residue of the final product is likely epimerized to a D-glutamine based on the loss of one deuterium, which agrees with the presence of the epimerization domain in the fourth NRPS module that incorporates glutamine. Only a minor amount of *d*₁₀-leucine was incorporated into the product, likely due to the relaxed substrate specificity of one of the adenylation domains activating valine or isoleucine.

Flavopeptins were isolated in 0.1 mg/L scale from the producing strain NRRL F-6652. All flavopeptin species with different aliphatic chain lengths were isolated as one fraction due to the difficulty of separation. To determine the absolute configuration of the amino acid building blocks, the isolated flavopeptins were hydrolyzed in strong acid, and the resulting amino acids were derivatized with Marfey's reagent (FDAA). Reverse-phase LC-MS analysis indicated the presence of L-Ile, L-Val, and L-Phe, equal amounts of L-Gln and D-Gln, and a minor amount of L-Leu, which agrees with the previous assignments (Figure S6).

To confirm the presence of the aldehyde group, a selective aldehyde derivatization agent, *O*-(2,3,4,5,6-pentafluorobenzyl)-hydroxylamine hydrochloride (PFBHA), was allowed to react with the isolated flavopeptins. LC-MS analysis of the reaction mixture showed a series of new peaks with 195.01 Da increases for the pentafluorobenzylloximine derivatives (e.g., *m/z* 1052.5618 for the PFBHA derivative of 1, calculated 1052.5603, Δ = 1.4 ppm) (Figure 3).

Attempts to obtain sufficient quantities of pure flavopeptins from the natural producer proved difficult due to low yield, coelution with other compounds, and the presence of multiple flavopeptin species that resist separation. Rather than solve these major problems, solid-phase peptide synthesis was used to synthesize two flavopeptin species with the same mass, 856.6 Da (1 and 2), as well as two additional derivatives, one with an N-terminal free amine (3) and the other with a C-terminal carboxylic acid (4) (Figure 4). Compound 1 showed the same LC retention time, intact mass, and MS² fragmentation masses as the corresponding flavopeptin species isolated from the natural source (Figure S7).

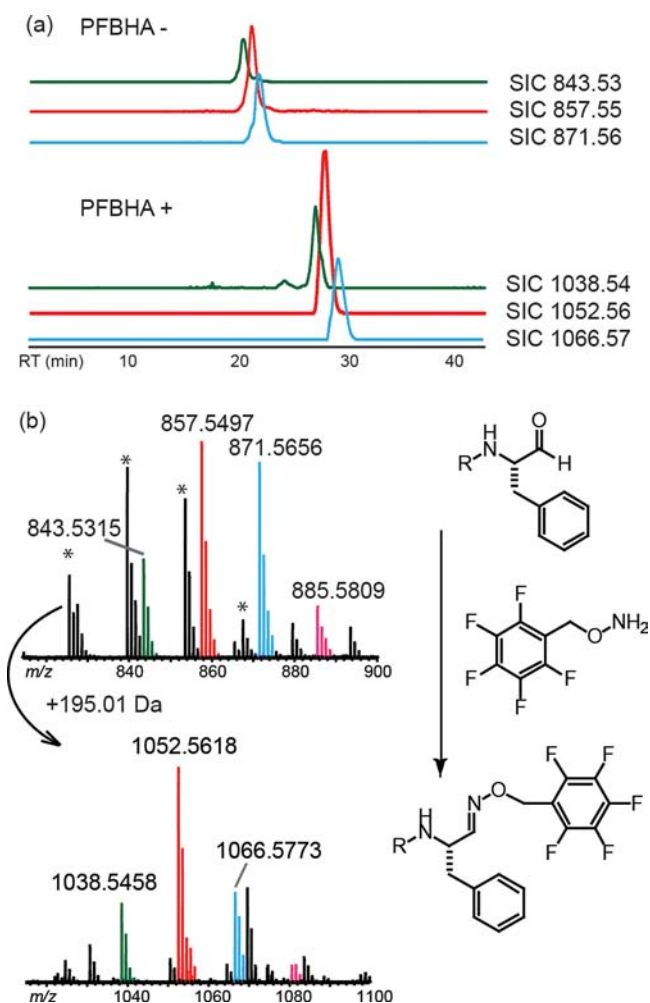


Figure 3. Detection of the aldehyde group of flavopeptides by derivatization with PFBHA. (a) LC-MS traces of flavopeptides before (top) and after (bottom) reaction with PFBHA. A series of masses with 195.01 Da increases (represented by SICs of m/z 1038.54, 1052.56, and 1066.57) appeared after PFBHA derivatization. (b) High-resolution mass spectra of flavopeptides before and after derivatization, together with the labeling reaction mechanism. Peaks with * indicate dehydrated products from the original sample.

Biosynthesis of Flavopeptins. The predicted biosynthetic mechanism for the flavopeptins is shown in Figure 5. The most unusual aspect is the presence of a relatively rare reductase domain at the C-terminus of FlavB instead of the more common thioesterase. A number of previous reports have shown that NAD(P)H-dependent reductase domains catalyze the reductive off-loading of peptide thioesters to produce linear aldehydes.²¹ In all the previous cases, however, the linear aldehyde exists only as an intermediate that undergoes subsequent reactions, such as further reduction to an alcohol (lyngbyatoxin), transamination to form an amine (myxochelin B and lysine biosynthesis in yeast), intramolecular cyclization to form a cyclic imine (koranimine and nostocyclopeptide), and intramolecular Pictet–Spengler reaction (saframycin).^{14,24–28} Flavopeptins are the first examples in which a stable peptide aldehyde is released as the end product from an NRPS assembly line. There were no masses observed in the culture broth corresponding to an alcohol, amine, or imine product. Only after organic extraction followed by drying did we observe masses with an 18 Da loss, which corresponded to a dehydrated

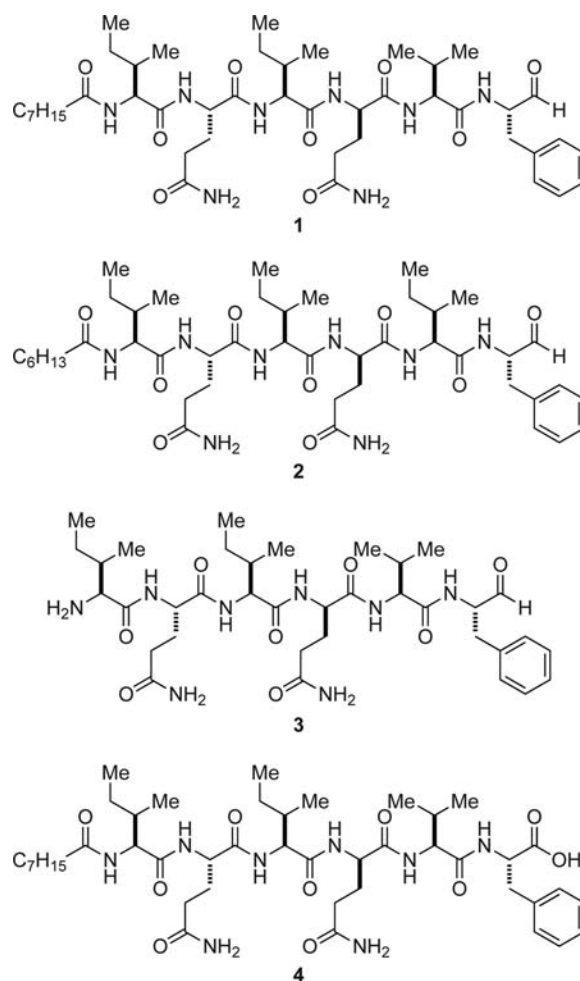


Figure 4. Flavopeptins 1 and 2, N-terminal amine derivative 3, and C-terminal carboxylic acid derivative (4) synthesized by solid-phase peptide synthesis.

form (Figure S8). Tandem MS analysis indicated the dehydration occurred between Gln₂ and the terminal aldehyde, possibly forming an imino group between the side-chain amide of the glutamine and the aldehyde. Drying the synthetic flavopeptins in organic solvent also showed the dehydration product with the same MS² fragmentation patterns, indicating the dehydrated compound was an artifact produced during purification.

Unlike nostocyclopeptides and koranimine, where the aldehyde intermediates spontaneously undergo macrocyclization to form an imine, flavopeptins maintain the linear form in nature. We speculated that the N-terminal fatty acid “cap” serves to protect the primary amine from condensation with the aldehyde, thus blocking imine formation. However, the synthetic analogue of flavopeptin with an N-terminal free amine (3) also retains its linear form, with only a very minor proportion of dehydrated product observed (less than 10% by MS analysis, Figure S7). This result demonstrated that the peptide sequence of flavopeptins itself is not prone to inducing the macrocyclic conformation. Our observations were in accordance with a previous study on nostocyclopeptide, where certain amino acids in the peptide sequence were found essential for the spontaneous macrocyclization of the peptidyl aldehyde intermediate into a cyclic imine.²⁴

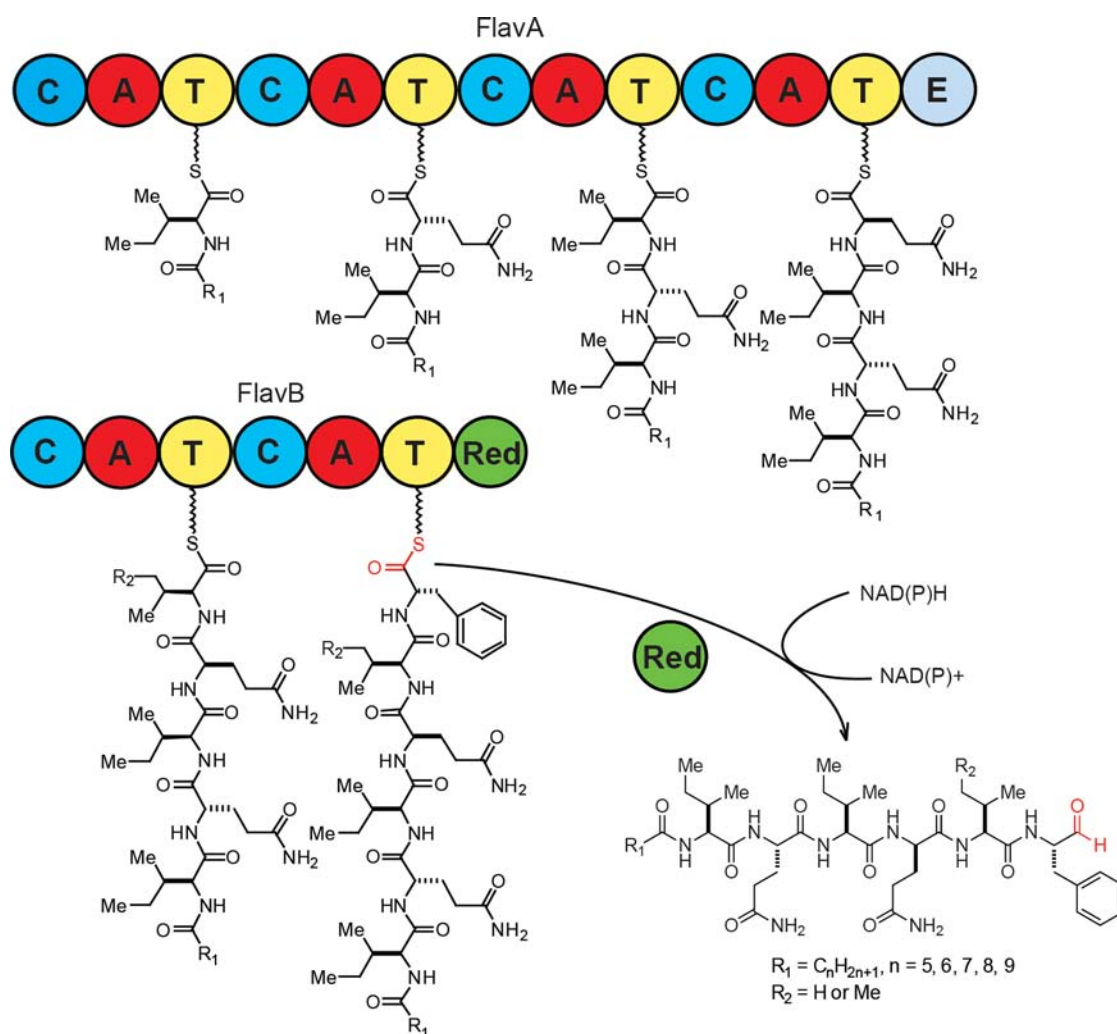


Figure 5. Predicted biosynthetic mechanism for the flavopeptins. The aldehyde group is predicted to be formed through reductive release of the peptide thioester using the NRPS reductase domain.

Table 1. Protease, Proteasome, and Cell Inhibition Activities of Flavopeptins

	IC ₅₀ (μM)	1	2	3
cysteine proteases	papain	0.16 ± 0.04	0.35 ± 0.03	0.07 ± 0.01
	calpain	2.9 ± 0.3	2.7 ± 0.4	1.8 ± 0.3
proteasome ^a	chymotrypsin-like	4.9 ± 0.3	3.3 ± 0.3	0.43 ± 0.08
	caspase-like	2.3 ± 0.2	2.3 ± 0.7	0.33 ± 0.07
cell line	MM.M1S	33 ± 2	not tested	no inhibition
	FR4	35 ± 6	30 ± 10	no inhibition
	U-937	13 ± 1	not tested	no inhibition

^aFor comparison with another natural peptide aldehyde, fellutamide B inhibits the proteasome with IC₅₀ ≈ 9 nM and ~2 μM for chymotrypsin-like and caspase-like activities, respectively.⁴

A BLAST search of FlavA and FlavB against the database revealed that two other strains, *Kitasatospora setae* KM-6054 (= NRRL B-16185) and *Streptomyces* sp. *Sirex* AA-E, contain similar NRPS gene clusters, with 69% and 80% sequence identity with F-6652 at the protein level (Figure S9). These two strains also share the same NRPS domain organization and the same predicted amino acid substrates, leading us to hypothesize that *K. setae* and *S. sp. Sirex* AA-E would have the genetic potential to produce flavopeptins. Indeed, the same masses corresponding to flavopeptins were detected from NRRL B-16185. Other neighboring genes are not highly conserved among these three strains, except for the MbtH protein (74%

and 79% identical to F-6652), indicating the MbtH, FlavA, and FlavB likely constitute the core genes responsible for flavopeptin biosynthesis.

Biological Activities of Flavopeptins. Many naturally occurring peptide aldehydes are known to exhibit inhibitory activities against proteases. The aldehyde warhead undergoes nucleophilic attack by the active site residue of the protease (serine, threonine, or cysteine) to form a reversible covalent hemiacetal/hemithioacetal adduct with the enzyme, while the peptide portion specifically interacts with the substrate binding site. By comparing the peptide sequence of flavopeptins with known peptide aldehyde protease inhibitors, we predicted the

two hydrophobic amino acids, valine and phenylalanine, at the C-terminus of the peptide could make flavopeptins potential inhibitors against cysteine proteases and the mammalian proteasome (Figure S10). When tested, flavopeptins strongly inhibited cysteine proteases like papain and calpain, with $\geq 90\%$ inhibition at $15 \mu\text{g/mL}$. Proteases of other types, including serine, aspartate, and metalloproteases, were inhibited by $< 50\%$ (Figure S11). The IC_{50} values of flavopeptins were measured using synthetic compounds (Table 1). Compound **1** showed low-micromolar to nanomolar inhibition against cysteine proteases ($\text{IC}_{50} = 0.16 \pm 0.04 \mu\text{M}$ for papain and $2.9 \pm 0.3 \mu\text{M}$ for calpain).

Given the structural similarity between flavopeptins and known peptide aldehyde proteasome inhibitors, we investigated whether they could inhibit the human 20S proteasome. As shown in Table 1, compound **1** inhibited the chymotrypsin-like activity of the human 20S proteasome with $\text{IC}_{50} = 4.9 \pm 0.3 \mu\text{M}$, and the caspase-like activity with $\text{IC}_{50} = 2.3 \pm 0.2 \mu\text{M}$. As a negative control, the flavopeptin analogue with a C-terminal carboxylic acid group rather than an aldehyde (**4**) was tested, but it showed no inhibition up to $100 \mu\text{M}$ (Figure S11). Compound **2** (containing Ile₅ instead of Val₅) exhibited similar inhibition potencies as **1**, while compound **3** (with an N-terminal free amine) showed 6- and 10-fold higher potencies. The stronger activity of **3** is possibly due to a greater affinity for the proteasome substrate-binding site.

In view of the anti-proliferative activities of some known proteasome inhibitors such as bortezomib,²⁹ we tested flavopeptins for activity against human multiple myeloma cell lines, MM.M1S and FR4, and a histiocytic lymphoma cell line, U-937. Using a cell viability assay that quantifies cellular ATP, compound **1** displayed values of $\text{IC}_{50} \approx 30$ and $13 \mu\text{M}$ against multiple myeloma and lymphoma cell lines, respectively. The N-terminal free amine analogue (i.e., **3**), which exhibited higher in vitro activities, showed no in vivo inhibition up to $80 \mu\text{M}$, most likely due to its low membrane permeability and low cellular stability compared to the N-acylated compounds. A noteworthy feature within the flavopeptins is the presence of a D-glutamine in the peptide sequence. Although D-amino acids are commonly seen in non-ribosomally synthesized peptides, they are rarely present in protease/proteasome inhibitors.

CONCLUSIONS

In this work, we described the discovery of the flavopeptins, a class of novel peptide aldehydes, from an orphan gene cluster in a *Streptomyces* species using the proteomics-based method, PrISM. Although there are previous examples of NRPS reductase domains catalyzing the formation of aldehydes as intermediates, flavopeptins are the first examples of stable linear peptide aldehydes arising definitively from an NRPS biosynthetic pathway. Similar to other peptide aldehydes, flavopeptins inhibit cysteine proteases and the human 20S proteasome by forming a covalent adduct between their aldehyde group and the active site cysteine or threonine residue of the target enzyme. There will be continued interest in elucidating whether other naturally occurring peptide aldehydes are also synthesized through a similar mechanism involving NRPS reductase domains. This can be realized by sequencing the genome of strains known to produce peptide aldehydes and locating the corresponding gene cluster.

ASSOCIATED CONTENT

Supporting Information

Mascot search result, bioinformatics analysis of protein sequences, extended mass spectrometry data, bioassay results, and NMR spectra for synthetic peptides. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

n-kelleher@northwestern.edu; r-thomson@northwestern.edu

Author Contributions

*Y.C. and R.A.M. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Agricultural Research Service, U.S. Department of Agriculture, for providing us the bacterial strains. We thank the ChemCore facility at Northwestern University for providing HPLC purification services, and the High-Throughput Analysis Lab for providing fluorospectroscopy services. Research reported in this publication was supported by the National Institutes of Health under award no. GM067725 to N.L.K. and by the National Science Foundation under award no. CHE0845063 to R.J.T.

REFERENCES

- (1) Storr, S. J.; Carragher, N. O.; Frame, M. C.; Parr, T.; Martin, S. G. *Nat. Rev. Cancer* **2011**, *11*, 364–374.
- (2) Lecaille, F.; Kaleta, J.; Bromme, D. *Chem. Rev.* **2002**, *102*, 4459–4488.
- (3) Kisselev, A. F.; van der Linden, W. A.; Overkleeft, H. S. *Chem. Biol.* **2012**, *19*, 99–115.
- (4) Hines, J.; Groll, M.; Fahnestock, M.; Crews, C. M. *Chem. Biol.* **2008**, *15*, S01–S12.
- (5) Kim, K. B.; Crews, C. M. *J. Med. Chem.* **2008**, *51*, 2600–2605.
- (6) Rentsch, A.; Landsberg, D.; Brodmann, T.; Bulow, L.; Girbig, A. K.; Kalesse, M. *Angew. Chem., Int. Ed.* **2013**, *52*, S450–S488.
- (7) Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009**, *26*, 1362–1384.
- (8) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nat. Chem. Biol.* **2005**, *1*, 265–269.
- (9) Forseth, R. R.; Amaike, S.; Schwenk, D.; Affeldt, K. J.; Hoffmeister, D.; Schroeder, F. C.; Keller, N. P. *Angew. Chem., Int. Ed.* **2013**, *52*, 1590–1594.
- (10) Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermanic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol.* **2011**, *7*, 794–802.
- (11) Chang, F. Y.; Brady, S. F. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 2478–2483.
- (12) Bumpus, S. B.; Evans, B. S.; Thomas, P. M.; Ntai, I.; Kelleher, N. L. *Nat. Biotechnol.* **2009**, *27*, 951–956.
- (13) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468–3496.
- (14) Evans, B. S.; Ntai, I.; Chen, Y.; Robinson, S. J.; Kelleher, N. L. *J. Am. Chem. Soc.* **2011**, *133*, 7316–7319.
- (15) Chen, Y.; Ntai, I.; Ju, K. S.; Unger, M.; Zamdborg, L.; Robinson, S. J.; Doroghazi, J. R.; Labeda, D. P.; Metcalf, W. W.; Kelleher, N. L. *J. Proteome Res.* **2012**, *11*, 85–94.
- (16) Chen, Y.; Unger, M.; Ntai, I.; McClure, R. A.; Albright, J. C.; Thomson, R. J.; Kelleher, N. L. *Med. Chem. Commun.* **2013**, *4*, 233–238.
- (17) Stanier, R. Y.; Palleroni, N. J.; Doudoroff, M. J. *Gen. Microbiol.* **1966**, *43*, 159–271.

- (18) Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009**, *458*, 181–217.
- (19) Rottig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362–367.
- (20) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78.
- (21) Du, L.; Lou, L. *Nat. Prod. Rep.* **2010**, *27*, 255–278.
- (22) Lautru, S.; Oves-Costales, D.; Pernodet, J. L.; Challis, G. L. *Microbiology* **2007**, *153*, 1405–1412.
- (23) Felnagle, E. A.; Barkei, J. J.; Park, H.; Podevels, A. M.; McMahon, M. D.; Drott, D. W.; Thomas, M. G. *Biochemistry* **2010**, *49*, 8815–8817.
- (24) Kopp, F.; Mahlert, C.; Grunewald, J.; Marahiel, M. A. *J. Am. Chem. Soc.* **2006**, *128*, 16478–16479.
- (25) Ehmann, D. E.; Gehring, A. M.; Walsh, C. T. *Biochemistry* **1999**, *38*, 6171–6177.
- (26) Koketsu, K.; Watanabe, K.; Suda, H.; Oguri, H.; Oikawa, H. *Nat. Chem. Biol.* **2010**, *6*, 408–410.
- (27) Li, Y.; Weissman, K. J.; Muller, R. *J. Am. Chem. Soc.* **2008**, *130*, 7554–7555.
- (28) Read, J. A.; Walsh, C. T. *J. Am. Chem. Soc.* **2007**, *129*, 15762–15763.
- (29) Teicher, B. A.; Ara, G.; Herbst, R.; Palombella, V. J.; Adams, J. *Clin. Cancer Res.* **1999**, *5*, 2638–2645.