ORGANIC LETTERS

XXXX
Vol. XX, No. XX
000–000

Identification of the Grincamycin Gene Cluster Unveils Divergent Roles for GcnQ in Different Hosts, Tailoring the L-Rhodinose Moiety

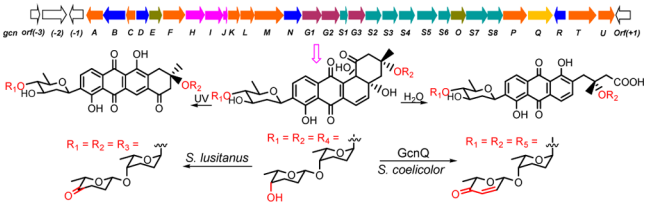
Yun Zhang,† Hongbo Huang,† Qi Chen, Minghe Luo, Aijun Sun, Yongxiang Song, Junying Ma, and Jianhua Ju*

CAS Key Laboratory of Marine Bio-resources Sustainable Utilization, Guangdong Key Laboratory of Marine Materia Medica, RNAM Center for Marine Microbiology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 164 West Xingang Road, Guangzhou 510301, China

jju@scsio.ac.cn

Received May 5, 2013

ABSTRACT



The gene cluster responsible for grincamycin (GCN, 1) biosynthesis in *Streptomyces lusitanus* SCSIO LR32 was identified; heterologous expression of the GCN cluster in *S. coelicolor* M512 yielded P-1894B (1b) as a predominant product. The Δ*gcnQ* mutant accumulates intermediate 1a and two shunt products 2a and 3a bearing L-rhodinose for L-cinerulose A substitutions. In vitro data demonstrated that GcnQ is capable of iteratively tailoring the two L-rhodinose moieties into L-aculose moieties, supporting divergent roles of GcnQ in different hosts.

The angucycline antibiotic grincamycin (GCN, **1**) was first discovered in 1987 from *Streptomyces griseoincarnatus*[1] and has been recently rediscovered by our group along with GCN B (**2**), GCN E (**3**), and GCNs C, D, and F as minor metabolites from the deep sea derived *Streptomyces lusitanus* SCSIO LR32.[2] GCN has a tetrangomycin skeleton in which the 3-O- and 9-positions are substituted by saccharides, containing α-L-rhodinose and α-L-cinerulose A for the disaccharide and consisting of β-D-olivose, α-L-rhodinose, and α-L-cinerulose A for the trisaccharide (Figure 1B). Closely related to **1** is vineomycin A_1 also known as P-1894B (**1b**), isolated from *Streptomyces matensis* subsp. *vineus* and *S. albogriseolus* subsp. No. 1984[3−5] featuring two L-aculose moieties in place of the two L-cinerulose A moieties of **1** (Figure 1B). GCN was reported to inhibit the growth of P388 murine leukemia cells and displayed cytotoxicities toward a panel of tumor cell lines,[1,2] whereas P-1894B was reported to be active against Sarcoma 180 solid tumors in mice.[3] Although numerous gene clusters responsible for angucycline antibiotics have been cloned,[6] the biosynthetic gene clusters for **1** and **1b** have not been identified. Consequently, the direct biosynthetic link between **1** and **1b** has been unclear.

Herein we report (i) the cloning, sequencing, and annotation of the gene cluster governing GCN biosynthesis in *S. lusitanus* SCSIO LR32; (ii) heterologous expression of the entire GCN biosynthetic gene cluster in *S. coelicolor* M512 yielding P-1894B (**1b**) as a predominant product; (iii) inactivation of *gcnQ* leading to accumulation of

† These authors contributed equally.

(1) Hayakawa, Y.; Iwakiri, T.; Imamura, K.; Seto, H.; Otake, N. *J. Antibiot.* **1987**, *40*, 1785–1787.

(2) Huang, H.; Yang, T.; Ren, X.; Liu, J.; Song, Y.; Sun, A.; Ma, J.; Wang, B.; Zhang, Y.; Huang, C.; Zhang, C.; Ju, J. *J. Nat. Prod.* **2012**, *75*, 202–208.

(3) Omura, S.; Tanaka, H.; Oiwa, R.; Awaya, J.; Masuma, R.; Tanaka, K. *J. Antibiot.* **1977**, *30*, 908–916.

(4) Ohta, K.; Mizuta, E.; Okazaki, H.; Kishi, T. *Chem. Pharm. Bull.* **1984**, *32*, 4350–4359.

(5) Imamura, N.; Kakinuma, K.; Ikekawa, N.; Tanaka, H.; Omura, S. *Chem. Pharm. Bull.* **1981**, *29*, 1788–1790.

(6) Kharel, M. K.; Pahari, P.; Shepherd, M. D.; Tibrewal, N.; Nybo, S. E.; Shaaban, K. A.; Rohr, J. *Nat. Prod. Rep.* **2012**, *29*, 264–325.

10.1021/ol401253p

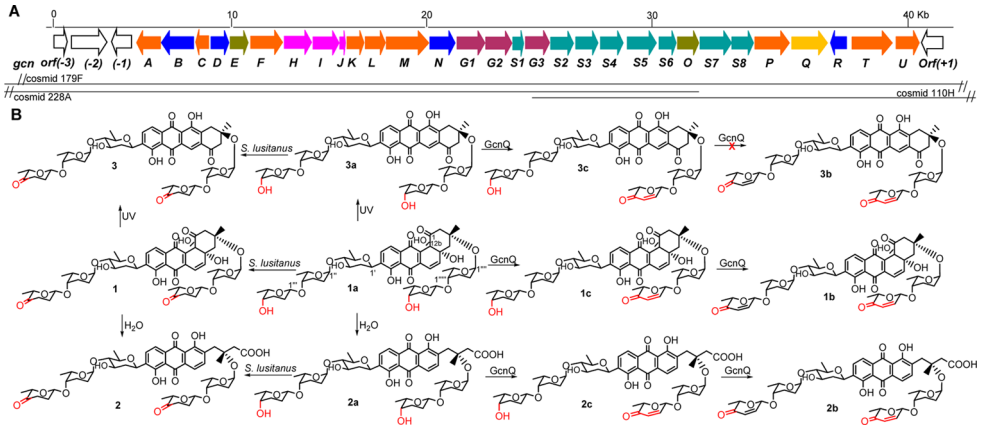


Figure 1. (A) The color-coded biosynthetic gene cluster of GCN (**1**); overlapping cosmids are indicated by solid lines. (B) GcnQ *in vitro* catalyzes transformation of **1a** and **2a** to **1b** and **2b** via intermediates **1c** and **2c**, respectively. *In vitro*, GcnQ converted **3a** to **3c**, failing to produce anticipated final product **3b**. GcnQ *in vivo* tailors L-rhodinose units to L-cinerulose A units in *S. lusitanus* SCSIO LR32 and tailors L-rhodinose units to L-aculose units in *S. coelicolor* M512.

biosynthetic intermediate **1a** and two shunt products **2a** and **3a**; and (iv) *in vitro* GcnQ-catalyzed transformations of **1a**–**3a** into P-1894B (**1b**), vineomycin B_2 (**2b**), and **3c**; **1b** and **2b** are generated through the intermediacy of **1c** and **2c**, respectively.

The sugars contained in GCN are all 2,3-deoxy sugars. Accordingly, we set out to clone the gene cluster of **1** using degenerate primers targeting the sugar 2,3-dehydratase.[7,8] A clear PCR band was amplified using the genomic DNA as a template, then ligated into pCR 2.1 vector, and transformed into *E. coli* DH5α cells. Sequence analysis of five clones revealed that the cloned fragments possessed identical sequences, which, upon Blast analysis, showed high homology to the 2, 3-dehydratase (UrdS) in the urdamycin biosynthetic pathway.[9] This specific 2,3-dehydratase probe was utilized to screen ~2300 clones of a SuperCos1-based genomic library of *S. lusitanus* SCSIO LR32. The target 2, 3-dehydratase allele was then inactivated by replacement with an *aac(3)IV*/*oriT* cassette using PCR-targeting mutagenesis methods.[10] The resulting mutant (Δ*gcnS7*) lost the ability to produce **1** but generated three analogues (**4**–**6**) (Figure 2, trace II), suggesting involvement of the cloned gene in GCN biosynthesis. Subsequent restriction enzyme mapping and end-sequencing of the 10 screened cosmids revealed that cosmid 179F may contain the whole biosynthetic pathway for **1**. Sequencing of cosmid 179F led to a contiguous DNA sequence of 50 Kb, from which 30-open reading frames spanning 37 Kb were identified probably accounting for biosynthesis of **1**. The gene cluster contains five type II polyketide synthases and cyclases associated with angucycline core construction (GcnHIJKL). Also found in the cluster of **1** were (i) seven oxidoreductases for postmodifications (GcnACFMQTU), (ii) a set of eight proteins for deoxysugar biosynthesis (GcnS1-S8), (iii) three glycosyltransferases for sugar attachment (GcnG1-G3), (iv) two regulatory proteins (GcnDR), (v) two proteins for transportation (GcnBN), (vi) a decarboxylase (GcnP), and (vii) two proteins with unknown functions (GcnEO) (Table S1). The nucleotide sequences have been deposited in the GenBank with accession number KC962511. The genetic organization of the biosynthetic gene cluster is shown in Figure 1A.

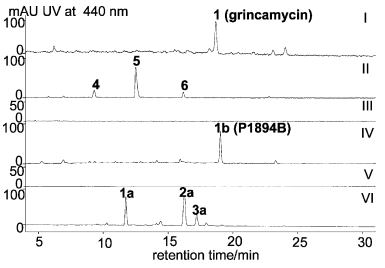


Figure 2. HPLC profiles of *S. lusitanus* wild-type (I), Δ*gcnS7* (II), Δ*gcnH* (III), *S. coelicolor*/179F (IV), *S. coelicolor* M512 (V), and Δ*gcnQ* (VI). See Figure 1B for structures.

The 2,3-dehydratase mutant (Δ*gcnS7*) was fermented on a large scale (8-L), enabling purification of analytically pure **4**–**6**. HRMS, ^{1}H, and ^{13}C NMR data analyses and comparisons with previously reported spectroscopic data revealed the structures of **4**–**6** to be tetrangomycin (**4**),[11] fridamycin (**5**),[12] and fridamycin methyl ester (**6**).[12]

(7) Zhang, X.; Alemany, L. B.; Fiedler, H. P.; Goodfellow, M.; Parry, R. J. *Antimicrob. Agents Chemother.* **2008**, *52*, 574–585.
(8) Sasaki, E.; Ogasawara, Y.; Liu, H.-w. *J. Am. Chem. Soc.* **2010**, *132*, 7405–7417.
(9) Hoffmeister, D.; Ichinose, K.; Domann, S.; Faust, B.; Trefzer, A.; Drager, G.; Kirschning, A.; Fischer, C.; Kunzel, E.; Bearden, D. W.; Rohr, J.; Bechthold, A. *Chem. Biol.* **2000**, *7*, 821–831.
(10) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1541–1546.
(11) Kaliappan, K. P.; Ravikumar, V. *J. Org. Chem.* **2007**, *72*, 6116–6126.

To further confirm that the candidate gene cluster is, in fact, responsible for biosynthesis of **1**, the minimal ketoacyl synthase GcnH was inactivated. The resultant Δ*gcnH* mutant proved incapable of **1** production as well as production of its analogues (Figure 2, trace III), validating that the cloned gene cluster is indeed responsible for **1** biosynthesis. To confirm the sufficiency of the gene cluster for GCN biosynthesis, heterologous expression of cosmid 179F, likely harboring the entire GCN biosynthetic cluster, was carried out. Cosmid 179F was generated by replacing the kanamycin resistance gene (from the Supercos1 vector) with a fragment excised from pSET152AB containing the apramycin resistance gene and elements necessary for conjugation and site specific recombination (*oriT*, integrase gene and φC-31 site), using λ-RED-recombination technology. The resulting cosmid, termed 179F-pSET152AB (Supporting Information), was transferred into *S. coelicolor* M512 by conjugation to generate the *S. coelicolor*/179F strain. The engineered strain was fermented using the same medium previously used for fermentation of wild-type *S. lusitanus* SCSIO LR32. Fermentation broth was then extracted with butanone, and the resulting metabolite contents were analyzed by HPLC. Unexpectedly, we observed generation of a predominant peak with a slightly longer retention time than that for **1** and having UV absorption signals characteristic of GCN at 425, 319, and 230 nm (Figure 2, trace IV). Further HRMS analysis of this peak revealed that it represents a compound having molecular formula $C_{49}H_{58}O_{18}$, four mass units smaller than that of **1**, and indicative of a structure known as P-1894B. Large-scale fermentation (3-L) of the recombinant strain enabled isolation of sufficient quantities of this material for NMR data acquisition. Both ^{1}H and ^{13}C NMR data confirmed the identity of the heterologously expressed compound as P-1894B (**1b**).[4] That the GCN gene cluster when expressed in *S. lusitanus* SCSIO LR32 (Figure 2, trace I) provides **1**, yet when expressed in *S. coelicolor* M512 affords **1b** as the predominant product, is truly remarkable.

Both **1** and **1b** possess keto and α,β-unsaturated keto sugars, respectively. How these highly unusual keto sugars are formed and then sequentially attached to each molecule are intriguing questions. Within the GCN cluster, *gcnQ* encodes a 529 aa oxidoreductase showing homology with AknOx (56% identity), TrdL (49% identity), and GilR (38% identity) in the aclacinomycin,[13] tirandamycin,[14] and gilvocarcin V[15] biosynthetic pathways, respectively. All these proteins are characterized by bicovalent attachment of the FAD cofactor to conserved histidine and cysteine residues (Supporting Information, Figure S2).

To explore the role of GcnQ in the GCN biosynthetic pathway, we inactivated *gcnQ*. Fermentation of the resultant Δ*gcnQ* mutant and HPLC metabolite analysis revealed that inactivation of *gcnQ* abolished production of **1** and its corresponding analogues. Instead, this mutant produced three major products **1a**, **2a**, and **3a** (Figure 2, trace VI). Large scale fermentation (8 L) of the Δ*gcnQ* strain enabled isolation of **1a**−**3a** in quantities sufficient for thorough structure elucidation. Molecular formulas of $C_{49}H_{66}O_{18}$ (**1a**), $C_{49}H_{66}O_{18}$ (**2a**), and $C_{49}H_{64}O_{17}$ (**3a**) were determined on the basis of HRMS data. ^{1}H and ^{13}C NMR spectroscopic data of **1a**, **2a**, and **3a** were similar to those of GCN (**1**), GCN B (**2**), and GCN E (**3**) (Tables S7−S9), respectively.[2] However, the ^{13}C NMR signals for the two carbonyl carbons in the two L-cinerulose A units were missing in each compound. Moreover, two additional oxygen-bearing methine signals at about δ_C 67 ppm were observed in all three compounds. Furthermore, the ^{13}C NMR resonances of C-3 in the L-cinerulose A units were shifted upfield from δ_C ~33 to δ_C ~24 ppm. These observations suggested that the two L-cinerulose A units at the end of the sugar chains in **1**, **2**, and **3** were replaced by two L-rhodinoses in **1a**, **2a**, and **3a**, respectively. Detailed analysis of their respective 2D (COSY, HMQC, HMBC, and NOESY) NMR data confirmed the elucidated structures. We envision that **2a** and **3a** are shunt products derived from **1a** during fermentation. Hydrolysis of the C-1/C-12b bond in **1a** likely affords **2a** whereas UV or heat-mediated rearrangement of **1a** may generate **3a**.[2,16] To test this hypothesis, **1a** was dissolved in 0.1% HOAc−H_2O and allowed to stand for 2 h; conversion to **2a** was readily apparent upon HPLC analysis and comparisons with a standard sample (Figure S5). Conversely, when dissolved in MeOH **1a** was converted to **3a** following UV irradiation for 12 h at rt (Figure S6). The isolation of **1a**−**3a** from the Δ*gcnQ* mutant implies that **1a** is a direct precursor to **1** and that GcnQ is responsible for tailoring of the two L-rhodinose units of **1a** into the respective L-cinerulose A units of **1**. Importantly, we witnessed no signs of olefination of either cinerulose A units with the wild-type producer *S. lusitanus* SCSIO LR32 (Figure 1B). We also investigated the prospect that the L-cineruloses of **1**−**3** might be generated by 2,3-reduction of appropriate L-aculose-containing substrates. We, thus, performed feeding experiments in which the Δ*gcnH* mutant strain provided L-aculose substrate **1b** for possible conversion into **1**. Under no circumstances could **1** be generated from **1b** thereby further supporting the divergent roles of GcnQ in different hosts (Supporting Information, Figure S4).

To evaluate the activity of GcnQ *in vitro*, the coding gene was cloned into the *Nde*I and *Hind*III sites of the pET28a(+) vector and the resulting vector was transformed into *E. coli* BL21(DE3). GcnQ was overexpressed as an N-terminal His_6-tagged soluble protein and purified to homogeneity by Ni affinity chromatography. The UV−vis spectrum of GcnQ exhibited two absorption

(12) Ueberbacher, B. J.; Osprian, I.; Mayer, S. F.; Faber, K. *Eur. J. Org. Chem.* **2005**, *7*, 1266–1270.
(13) Alexeev, I.; Sultana, A.; Mäntsälä, P.; Niemi, J.; Schneider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6170–6175.
(14) (a) Mo, X.; Wang, Z.; Wang, B.; Ma, J.; Huang, H.; Tian, X.; Zhang, S.; Zhang, C.; Ju, J. *Biochem. Biophy. Res. Commun.* **2011**, *406*, 341–347. (b) Mo, X.; Huang, H.; Ma, J.; Wang, Z.; Wang, B.; Zhang, S.; Zhang, C.; Ju, J. *Org. Lett.* **2011**, *13*, 2212–2215.
(15) Noinaj, N.; Bosserman, M. A.; Schickli, M. A.; Piszczek, G.; Kharel, M. K.; Pahari, P.; Buchanan, S. K.; Rohr, J. *J. Biol. Chem.* **2011**, *286*, 23533–23543.
(16) Sezaki, M.; Kondo, S.; Maeda, K.; Umezawa, H.; Ohno, M. *Tetrahedron* **1970**, *26*, 5171–5190.
(17) Imamura, N.; Kakinuma, K.; Ikekawa, N.; Tanaka, H.; Omura, S. *J. Antibiot.* **1982**, *35*, 602–608.

bands at 350−375 and 450 nm, very similar to the bicovalent flavinylated proteins BBE and TrdL.[14b] Heat denaturation of GcnQ resulted in a colorless supernatant and yellow precipitate; no FAD was detected in the supernatant by UV−vis spectroscopy or HPLC analysis demonstrating that FAD is, in fact, covalently attached to the protein.

Enzymatic reactions with GcnQ were conducted at 30 °C in a volume of 50 μL consisting of 50 mM Tris-HCl (pH 7.5), 20 μM substrate, and 0.5 μM GcnQ. Reactions with boiled GcnQ served as controls; all reactions were carried out in parallel to ensure reaction consistency. Reactions were terminated by 2-fold extraction with 100 μL EtOAc; EtOAc layers were evaporated to dryness, the resulting contents were then dissolved in 10 μL of MeOH, and the samples were subjected to HPLC analyses. HPLC time-course analyses revealed the GcnQ-catalyzed transformation of **1a** into **1b** through intermediate **1c** (Figure 3A). Similarly, **2a** was converted to **2b** through the agency of **2c** (Figure 3B), and **3a** was converted to **3c** (Figure 3C). Notably, GcnQ failed to generate **3b** from **3a**; it appears the reaction sequence stalls following generation of **3c** (Figure 3C). Large scale (100 mL) GcnQ-catalyzed reactions of **1a** allowed the isolation of **1b** and **1c** in quantities sufficient for structural elucidation. Compounds **2b** and **2c** were obtained in similar yields from large scale (100 mL) reaction of **2a** with GcnQ; small scale reaction (50 mL) of **3a** afforded **3c**. In all cases intermediates and products were isolated in quantities enabling rigorous structure elucidation. Significantly we found that, *in vitro*, GcnQ failed to generate either set of products **1b**−**3b** or **1c**−**3c** from the respective starting materials **1**−**3** (Figure 3D−F).

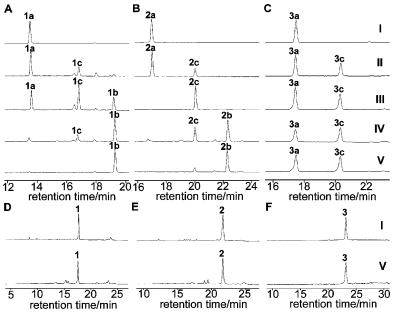


Figure 3. GcnQ time-course experiments involving HPLC analyses of GcnQ-containing reactions in which **1a** (A), **2a** (B), **3a** (C), and **1** (D), **2** (E), **3** (F) were incubated with GcnQ for (II) 1.5 min, (III) 5 min, (IV) 30 min, and (V) 60 min. Trace I is a negative control (1 h) with boiled GcnQ.

Compounds **1b** and **2b** were confirmed to be P-1894B and vineomycin B_2 by comparisons and analyses of the ^{1}H, ^{13}C, and 2D NMR data with previously reported data.[4,17] Intermediates **1c**−**3c** were identified on the basis of MS, ^{1}H and ^{13}C NMR data analyses (Tables S7−S9). HRMS spectra revealed that **1c** had the molecular formula $C_{49}H_{62}O_{18}$. ^{1}H and ^{13}C NMR spectroscopic data of **1c** resembled those of **1b**. Close comparisons of these data sets revealed the absence of two methylenes in **1b**. Instead, a set of signals at δ_H 6.95 (dd, J = 9.8, 3.4 Hz), δ_C 144.1, δ_H 6.10 (d, J = 9.8 Hz), δ_C 127.5 were noted. Moreover, a new ^{13}C NMR signal at δ_C 197.9 was apparent, indicating the presence of a (*cis*)-α, β-unsaturated carboxy moiety in **1c**. Consequently, it was apparent that an L-rhodinose unit in **1b** was replaced by an L-aculose in **1c**. Further HMBC correlations of 1‴′-H/C-4‴ and 4‴-H/C-1‴′ revealed this unanticipated L-aculose moiety to be positioned at the end of the disaccharide chain (Figure S12). Similarly, structures **2c** and **3c** were determined by NMR data analysis.

The *in vitro* enzyme assays reveal a number of new and interesting features about GcnQ. First, GcnQ demonstrates broad substrate specificity with regards to the aglycon; all angucyclic, tricyclic, and tetracyclic rings served as effective substrates. Second, oxidation of the terminal sugar (either di- or trisaccharide) at C-4 to its keto form and subsequent desaturation were found to occur in tandem, indicating that the L-rhodinose moiety is readily converted to the L-aculose moiety and does not involve the intermediacy of L-cinerulose A. This path deviates from the related aclacinomycin case.[13] Third, for both the native substrate **1a** and non-native substrates **2a** and **3a**, the L-rhodinose components of both di- and trisaccharide chains were found to be very efficiently desaturated to their L-aculose moieties though desaturation of the disaccharide L-rhodinose preceeds that of the trisaccharide. In the case of **3a**, only the disaccharide element serves as a substrate for desaturation.

In summary, we have conclusively identified the gene cluster responsible for GCN biosynthesis. Inactivation of GcnQ, a bicovalent FAD-bonded enzyme, affords a strain that accumulates biosynthetic intermediate **1a**, a hydrolysis dependent shunt product **2a**, and a UV-dependent rearranged product **3a**; both **2a** and **3a** contain L-rhodinose moieties at both termini of their tri- and disaccharide chains. *In vitro* experiments demonstrate that GcnQ catalyzes the conversion of L-rhodinose residues into L-aculose in both the tri- and disaccharide elements of **1a**−**3a**, to afford **1b**, **2b**, and **3c**. Importantly, **1b** and **2b** are generated through the intermediacy of **1c** and **2c**, respectively. However, GcnQ fails to analogously convert **3c** to product **3b**. Nonetheless, these studies demonstrate that GcnQ is compatible with a broad range of substrates. The GCN gene cluster, when heterologously expressed in *S. coelicolor* M512, yielded P-1894B (**1b**) as the sole product, supporting divergent roles of GcnQ in different hosts. These findings support the notion that GcnQ may be a prime candidate for combinatorial biosynthetic generation of new non-natural products based on the grincamycin scaffold.

Acknowledgment. This work was supported by MOST (2012AA092104, 2010CB833805), NSFC (31290233, 41106138), and STPPGP (2011B031200004).

Supporting Information Available. Detailed experimental procedures, NMR data and spectra for compounds **1a**, **2a**, **3a**, **1b**, **2b**, **1c**, **2c**, and **3c**. This material is available free of charge via the Internet at http://pubs.acs.org.

The authors declare no competing financial interest.