Supporting Information

Park *et al.* 10.1073/pnas.0803164105

SI Text

Experimental for Preparation and Identification of Gentamicin Biosynthetic Intermediates Obtained from the Heterologous Host Expressing Various Gentamicin Biosynthetic Gene Sets. 2-Deoxystreptamine (2-DOS, **1a**) was purchased from GeneChem (Republic of Korea) and used without further purification. Fifteen liters of culture broth from *S. venezuelae* strain harboring pYJ496 were treated as described in *Materials and Methods* in the main text to prepare **1a** as a white solid (1.5 mg). Based on comparisons of the retention time, MS/MS and NMR spectra with those of the 2-DOS (**1a**) authentic standard, a compound produced by *S. venezuelae* strains harboring plasmids expressing the *gtmB*-gtm*A-gacH* genes (pYJ495 and pYJ496) was found to be 2-DOS (**1a**) (see the analytical data of **1a** described below).

Eighteen liters of culture broth obtained from combining numerous batch cultivations of *S. venezuelae* strain harboring pYJ498 were treated as described in *Materials and Methods* to prepare 2′-*N*-acetylparomamine (**2a**) as a pale yellow solid (1.3 mg). By comparing the assigned NMR data with the data acquired by Truman *et al.* (5) and interpretations of MS/MS spectra based on cross-referencing with spectra of various gentamicin intermediates obtained in our previous study (6), a compound produced by *S. venezuelae* strains harboring plasmid expressing *gtmB-gtmA-gacH-gtmG* and resistance genes (pYJ498) was identified as 2′-*N*-acetylparomamine (**2a**) (see the analytical data of **2a** described below).

Paromamine (**3a**) was chemically synthesized by acid hydrolysis of paromomycin (Sigma) as described previously (7). The hydrolysate was titrated to pH 6 and purified using SPE cartridge, then freeze-dried to produce the authentic paromamine. The NMR data of the authentic paromamine prepared was consistent with the previous report (3). Nine liters of culture broth from *S. venezuelae* strain harboring pYJ501 were treated as described in *Materials and Methods* to prepare **3a** as a white solid (3.2 mg). By comparing the retention time, MS/MS and NMR spectra with those of authentic paromamine, a compound produced by mutant strains harboring plasmids containing the *gtmB-gtmA-gacH-gtmG-btrD*/*kacA*/*neo16*/*gtmM* genes (pYJ500, pYJ501, pYJ502, and pYJ503) was determined to be paromamine (**3a**) (see the analytical data of **3a** described below).

Authentic gentamicin A_2 (**4a**) was isolated and purified from the extract of wild-type *M. echinospora* ATCC 15835 as described in *Materials and Methods*. The NMR data of the authentic gentamicin A_2 prepared were consistent with the previous report (8, 9). Fifteen liters of culture broth obtained from combining numerous batch cultivations of *S. venezuelae* strain harboring pYJ504 were treated as described in *Materials and Methods.* As well, 10 separate *in vitro* reactions using the cell-free extracts of mycelium obtained from the corresponding culture were carried out, and then pooled into the above extracts to prepare gentamicin A_2 (**4a**) as a pale yellow solid (0.5 mg). By comparison of the retention time, MS/MS and partially assigned NMR spectra with those of authentic gentamicin A_2 purified from wild-type *M. echinospora* (6, 8, 9), a compound produced from *S. venezuelae* strain harboring pYJ504 and pYJ505 containing the *gtmB-gtmA-gacH-gtmG-gtmE-kacA* and resistance genes was identified as gentamicin A_2 (**4a**) (see the analytical data of **4a** described below). All NMR samples were prepared by dissolving each compound in 150 μl of D_2O and placing the solution in a 5 mm Shigemi advanced NMR microtube matched to the solvent.

2-Deoxystreptamine (1a). $T_{retention}$ (HPLC-ESI-MS): 61.2 min; HR-ESI-MS: *m/z* 163.1882 [M+H] (calculated for $C_6H_{14}N_2O_3$, 163.1870); ESI-MS/MS (*m/z* 163): [M+H] 163, [M+H–H_2O] 145, [M+H–H_2O–NH_3] 128, [M+H–2H_2O–NH_3] 110, [M+H–H_2O–NH_3–C_2H_2] 102, [M+H–3H_2O–NH_3] 92, [M+H–2H_2O–NH_3–C_2H_2] 84, [M+H–3H_2O–NH_3–C_2H_2+H_2] 68, [M+H–2H_2O–NH_3–2C_2H_2+H_2] 60 (see Fig. S7); ^{1}H and ^{13}C NMR data of **1a** were described in Table S2 and Fig. S8.

2′-*N*-acetylparomamine (2a). $T_{retention}$ (HPLC-ESI-MS): 59.8 min; HR-ESI-MS: *m/z* 366.1807 [M+H] (calculated for $C_{14}H_{27}N_3O_8$, 366.1795); ESI-MS/MS (*m/z* 366): [M(ab)+H] 366, [M(ab)+H–3H_2O–2NH_3] 288, [M(ab)+H–4H_2O–2NH_3] 270, [M(ab)+H–5H_2O–CH_3] 241, [M(a)+H] 204, [M(a)+H–H_2O] 186, [M(a)+H–2H_2O] 168, [M(b)+H] 163, [M(b)+H–H_2O] 145, [M(a)+H–2H_2O–C_2H_2+H_2] 144, [M(b)+H–2H_2O] 126, [M(b)+H–4H_2O–C_2H_2+H_2] 84 (see Fig. S9); ^{1}H and ^{13}C NMR data of **2a** were described in Table S3 and Fig. S10.

Paromamine (3a). $T_{retention}$ (HPLC-ESI-MS): 74.8 min; HR-ESI-MS: *m/z* 324.3440 [M+H] (calculated for $C_{12}H_{25}N_3O_7$, 324.3428); ESI-MS/MS (*m/z* 324): [M(ab)+H] 324, [M(ab)+H–NH_3] 307, [M(ab)+H–NH_3–H_2O] 289, [M(ab)+H–NH_3–2H_2O] 271, [M(ab)+H–NH_3–4H_2O–C_2H_2+H_2] 211, [M(ab)+H–NH_3–4H_2O–2C_2H_2+2H_2] 187, [M(b)+H] 163, [M(b)+H–H_2O] 145, [M(b)+H–2H_2O] 126, [M(b)+H–H_2O–NH_3–C_2H_2] 102, [M(b)+H–2H_2O–NH_3–C_2H_2] 84, [M(b)+H–2H_2O–NH_3–2C_2H_2+H_2] 60 (see Fig. S11); ^{1}H and ^{13}C NMR data of **3a** were described in Table S4 and Fig. S12.

Gentamicin A_2 (4a). $T_{retention}$ (HPLC-ESI-MS): 73.3 min; HR-ESI-MS: *m/z* 456.1123 [M+H] (calculated for $C_{17}H_{33}N_3O_{11}$, 456.1115); ESI-MS/MS (*m/z* 456): [M(abc)+H] 456, [M(abc)+H–NH_3] 438, [M(abc)+H–NH_3–H_2O] 421, [M(ab)+H] 324, [M(ab)+H–NH_3] 307, [M(bc)+H] 295, [M(bc)+H–H_2O] 277, [M(bc)+H–H_2O–C_2H_2+H_2] 253, [M(bx)+H] 205, [M(bx)+H–H_2O] 187, [M(b)+H] 163, [M(b)+H–H_2O] 145, [M(b)+H–2H_2O–NH_3–C_2H_2] 84 (see Fig. S13); ^{1}H and ^{13}C NMR data of **4a** were described in Table S5 and Fig. S14.

1. Thompson JD, *et al.* (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
2. Yokoyama K, *et al.* (2007) Characterization and mechanistic study of a radical SAM dehydrogenase in the biosynthesis of butirosin. *J Am Chem Soc* 129:15147–15155.
3. Kudo F, *et al.* (2005) Biosynthesis of 2-deoxystreptamine by three crucial enzymes in *Streptomyces fradiae* NBRC 12773. *J Antibiot* 58:766–774.
4. McCarthy AA, Peterson NA, Knijff R, Baker EN (2004) Crystal structure of MshB from *Mycobacterium tuberculosis*, a deacetylase involved in mycothiol biosynthesis. *J Mol Biol* 335:1131–1141.
5. Truman AW, Huang F, Llewellyn NM, Spencer JB (2007) Characterization of the enzyme BtrD from *Bacillus circulans* and revision of its functional assignment in the biosynthesis of butirosin. *Angew Chem Int Ed* 46:1462–1464.
6. Park JW, *et al.* (2007) Analytical profiling of biosynthetic intermediates involved in the gentamicin pathway of *Micromonospora echinospora* by high-performance liquid chromatography using electrospray ionization mass spectrometric detection. *Anal Chem* 79:4860–4869.
7. Haskell TH, French JC, Bartz QR (1959) Paromomycin. I. Paromamine, a glycoside of D-glucosamine. *J Am Chem Soc* 81:3480–3481.
8. Nagabhushan TL, Turner WN, Daniels PJL, Morton JB (1975) The gentamicin antibiotics. 7. Structures of the gentamicin antibiotics A_1, A_3, and A_4. *J Org Chem* 40:2830–2834.
9. Nagabhushan TL, Daniels PJL, Jaret RS, Morton JB (1975) The gentamicin antibiotics 8. Structure of gentamicin A_2. *J Org Chem* 40:2835–2836.

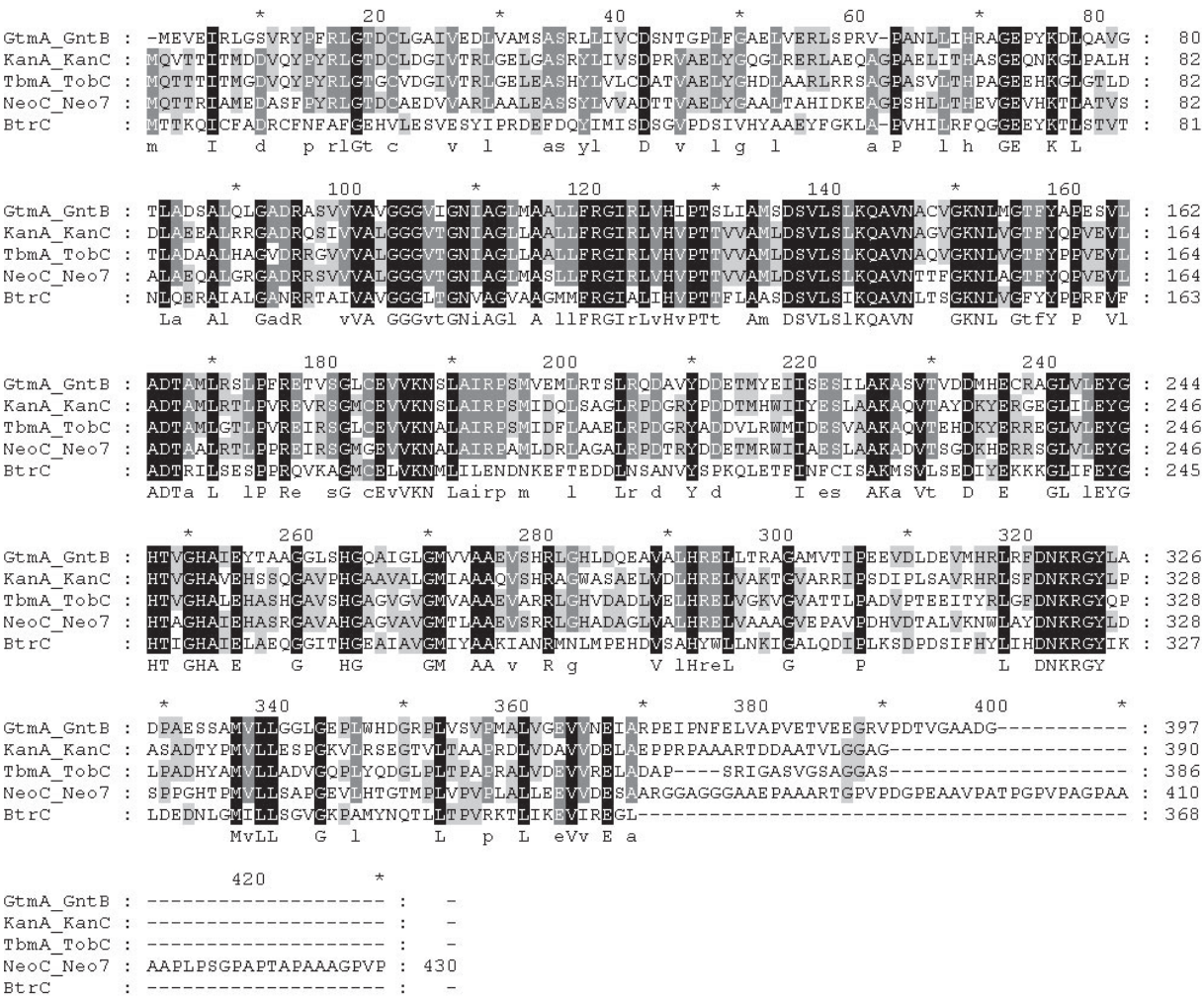

Fig. S1. Sequence alignment of GtmA with other 2-deoxy-*scyllo*-inosose synthases from *S. kanamyceticus* (KanA: GenBank accession no. CAE46939, or KanC: CAF31589), *S. tenebrarius* (TbmA: CAE22471, or TobC: CAH18556), *S. fradiae* (NeoC: BAD95820, or Neo7: CAH58690), and *Bacillus circulans* (BtrC: BAE07067). Multiple alignment of each sequence was carried out by ClustalX (1). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈80% homology. Black letters in light gray boxes represent residues with ≈60% homology. The predicted product of *gtmA* (397 aa) is 57% identical to KanA (KanC), 58% identical to TbmA (TobC), 56% identical to NeoC (Neo7), and 38% identical to BtrC.

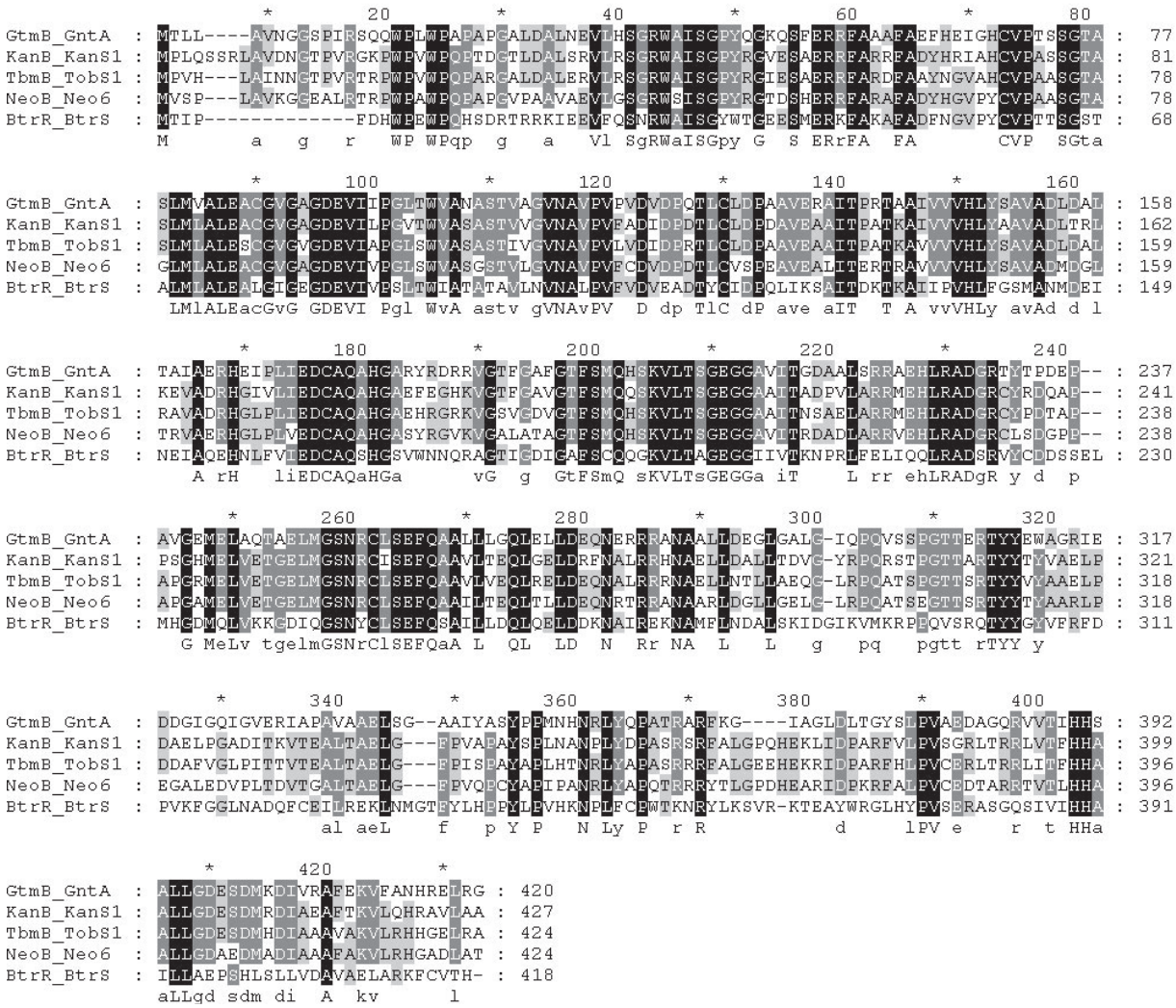

Fig. S2. Sequence alignment of GtmB with other L-glutamine:2-deoxy-*scyllo*-inosose aminotransferases from *S. kanamyceticus* (KanB: GenBank accession no. CAE46938, or KanS1: CAF31588), *S. tenebrarius* (TbmB: CAE22472, or TobS1: CAH18555), *S. fradiae* (NeoB: BAD95819, or Neo6: CAH58689), and *Bacillus circulans* (BtrR: CAD41947, or BtrS: BAE07061). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈80% homology. Black letters in light gray boxes represent residues with ≈60% homology. The predicted product of *gtmB* (463 aa) is 60% identical to KanB (KanS1), 62% identical to TbmB (TobS1), 60% identical to NeoB (Neo6), and 41% identical to BtrR (BtrS).

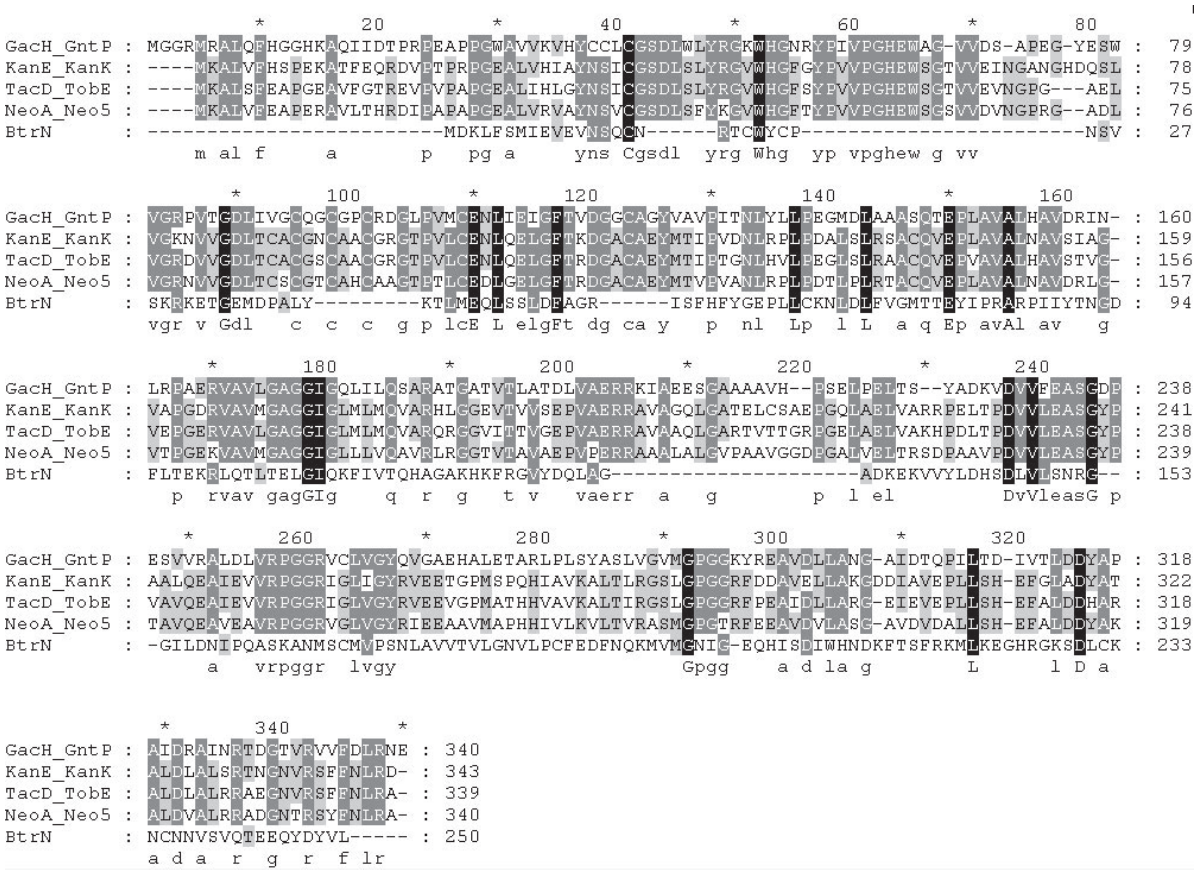

Fig. S3. Sequence alignment of GacH with other 2-deoxy-*scyllo*-inosamine dehydrogenases from *S. kanamyceticus* (KanK: GenBank accession no. CAE46938, or KanE: CAF31584), *S. tenebrarius* (TacD: CAE22477, or TobE: CAH18550), *S. fradiae* (NeoA: BAD95818, or Neo5: CAH58688), and *Bacillus circulans* (BtrN: CAD41948). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈80% homology. Black letters in light gray boxes represent residues with ≈60% homology. The predicted product of *gacH* (341 aa) is 45% identical to KanK (KanE), 47% identical to TacD (TobE), and 48% identical to NeoA (Neo5). However, it does not show sequence similarity to the recently characterized SAM dependent dehydrogenase BtrN (2). NeoA (Neo5) is characterized as a zinc-containing NAD(P) dependent dehydrogenase, which includes two zinc-binding motifs (3), and our sequence analysis also indicated the same two motifs observed in NeoA are also present in GacH (catalytic zinc ion coordination residues Cys-41, His-63, Glu-64, Glu-147, and Pro-148; structural zinc coordination Cysteins at 92, 95, 98, and 106 positions).

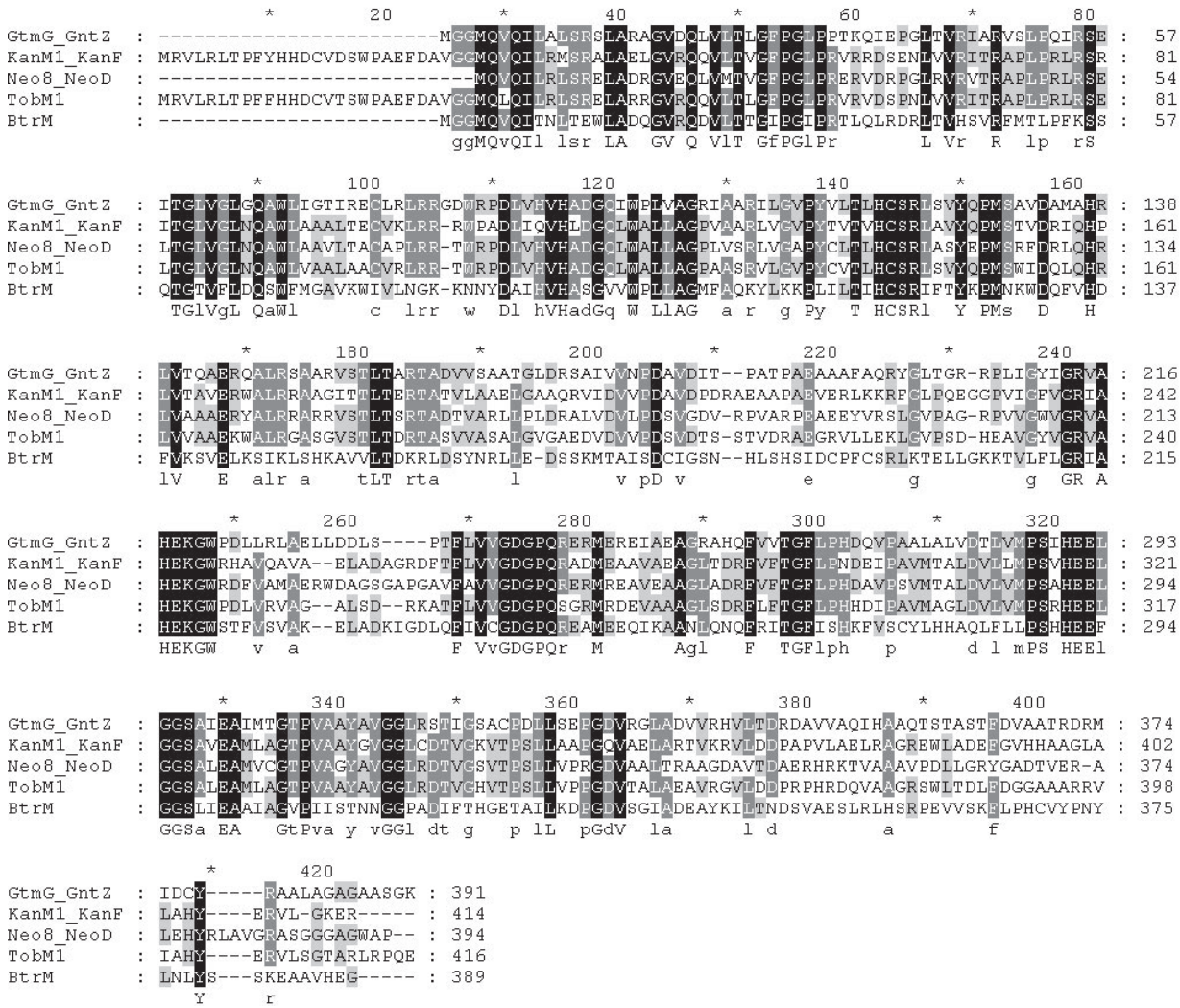

Fig. S4. Sequence alignment of GtmG with other glycosyltransferases from *S. kanamyceticus* (KanM1: GenBank accession no. BAE95600, or KanF: CAE46947), *S. tenebrarius* (TobM1: CAH18562), *S. fradiae* (Neo8: CAH58691 or NeoD: AB211959), and *Bacillus circulans* (BtrM: CAG77417). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈80% homology. Black letters in light gray boxes represent residues with ≈60% homology. The predicted product of *gtmG* is 54% identical to KanM1 (KanF), 57% identical to TobM1, 53% identical to Neo8, and 37% identical to BtrM.

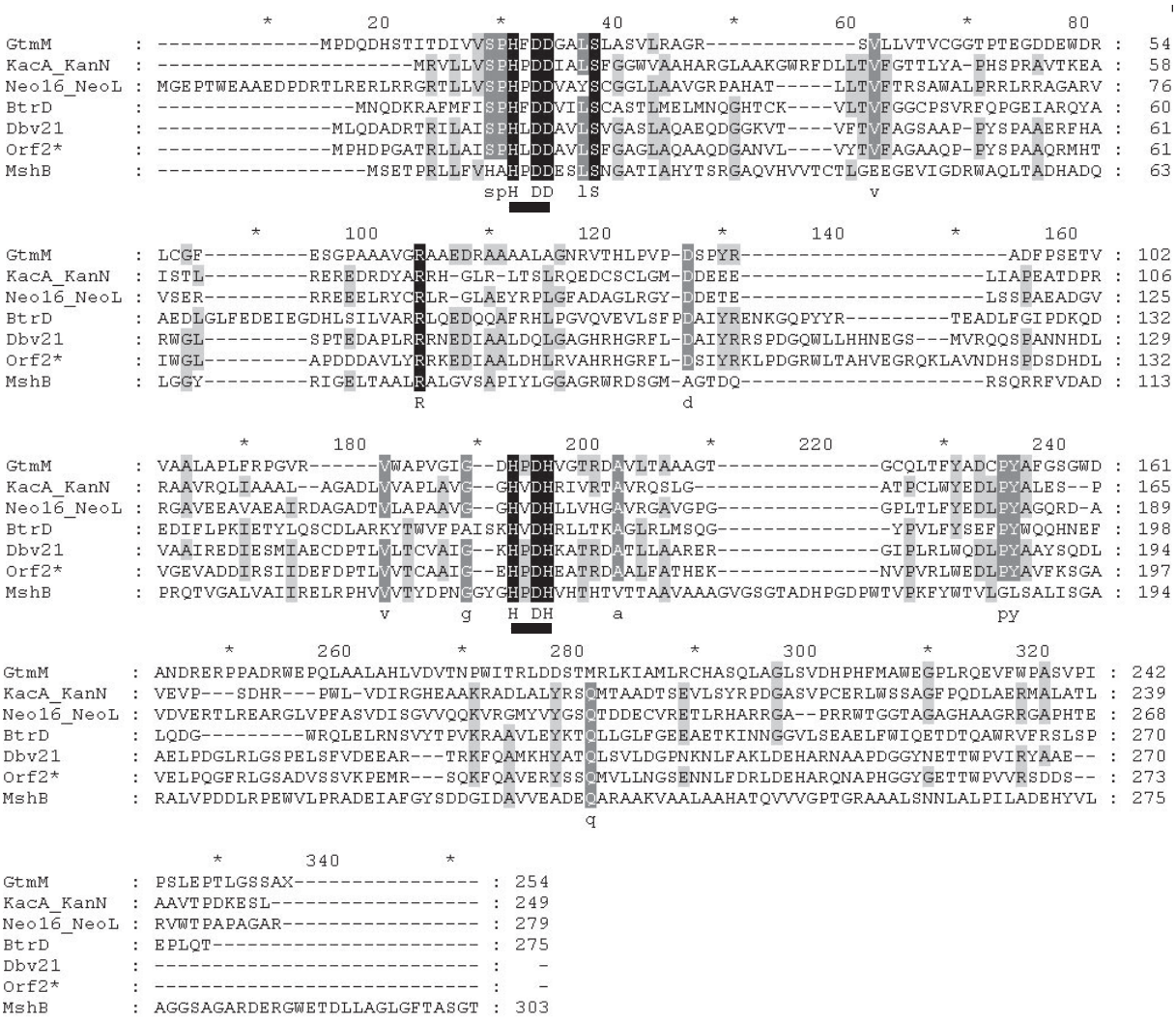

Fig. S5. Sequence alignment of GtmM (GenBank accession no. AM946392) with other acetylparomamine deacetylases from *S. kanamyceticus* (KacA: AJ582817 or KanN: AJ628422), *S. fradiae* (Neo16: AJ843080 or NeoL: AB211959), *Bacillus circulans* (BtrD: BAE07068), *Actinomadura* sp. ATCC 39727 (Dbv21: AJ561198), *Actinoplanes teichomyceticus* (Orf2*: AJ632270), and *Mycobacterium bovis* AF2122/97 (MshB: NP_854857). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈80% homology. Black letters in light gray boxes represent residues with ≈60% homology. The predicted product of *gtmM* (254 aa) is 29% identical to KacA, 37% identical to Neo16, 25% identical to BtrD, 28% identical to Orf2*, and 25% identical to Dbv21. However, it does not show sequence identity with MshB. The conserved sequence motifs, obtained from MshB crystallization studies (4), near the N terminus ([A/P]HXDD) and another toward the middle of the protein (HXDH) are underlined.

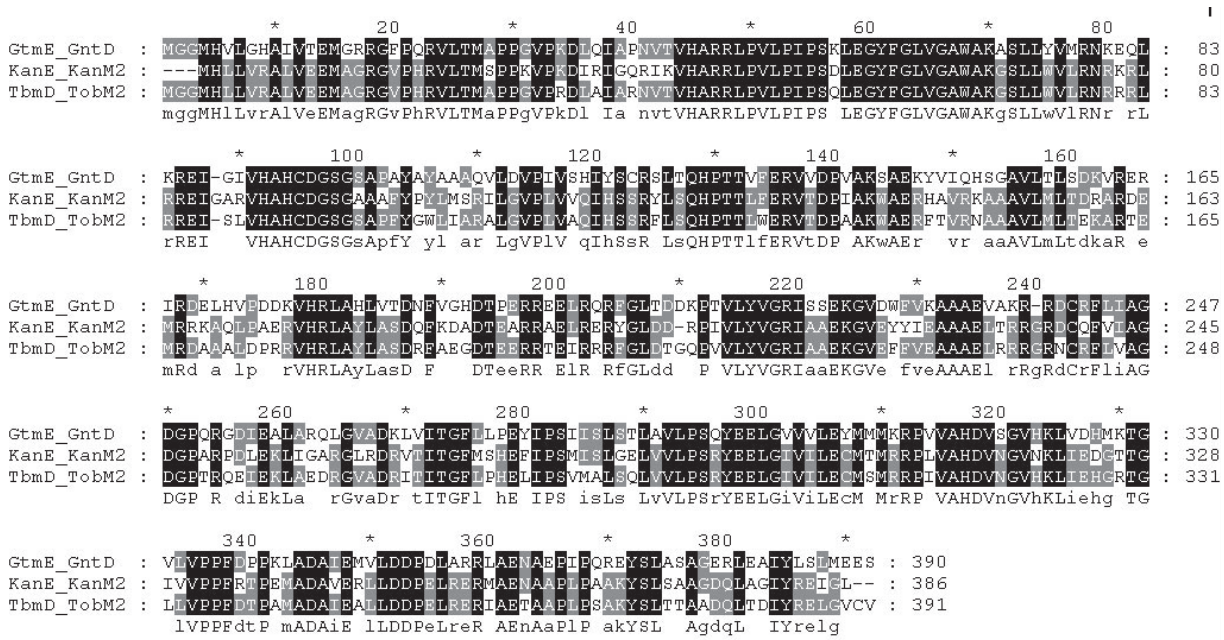

Fig. S6. Sequence alignment of GtmE with other glycosyltransferases from *S. kanamyceticus* (KanE: GenBank accession no. CAE46941, or KanM2: CAF31591) and *S. tenebrarius* (TbmD: CAE22469, or TobM2: CAH18558). White letters in black boxes represent amino acid residues identical in all species, while white letters in gray boxes represent residues with ≈60% homology. The predicted product of *gtmE* is 53% identical to KanE (KanM2) and 56% identical to TbmD (TobM2).

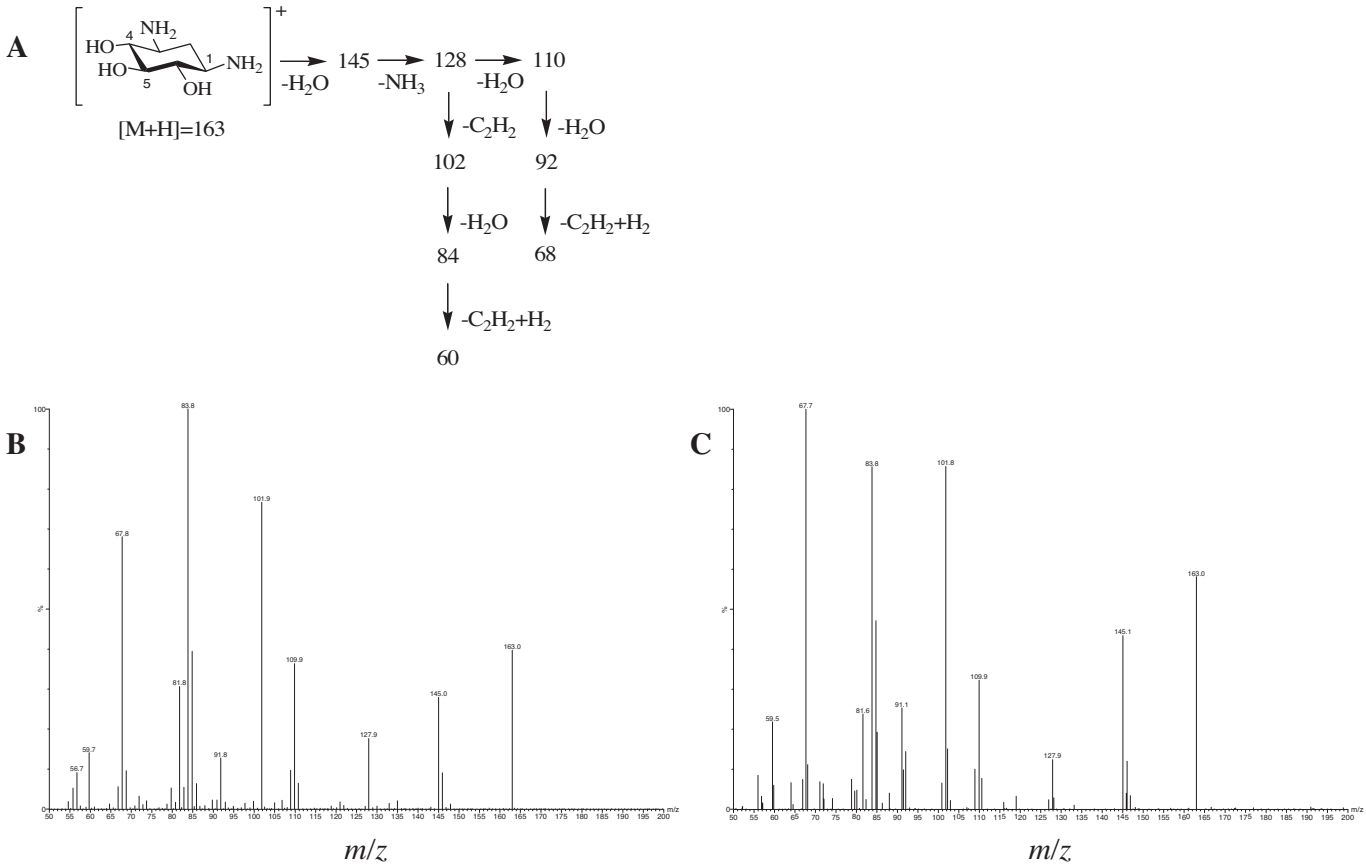

Fig. S7. (*A*) ESI-MS/MS fragmentation pattern of 2-deoxystreptamine (**1a**), and mass spectra of (*B*) authentic **1a**, and (*C*) heterologously produced **1a** from *S. venezuelae* strain harboring pYJ496 expressing the *gtmB-gtmA-gacH* genes.

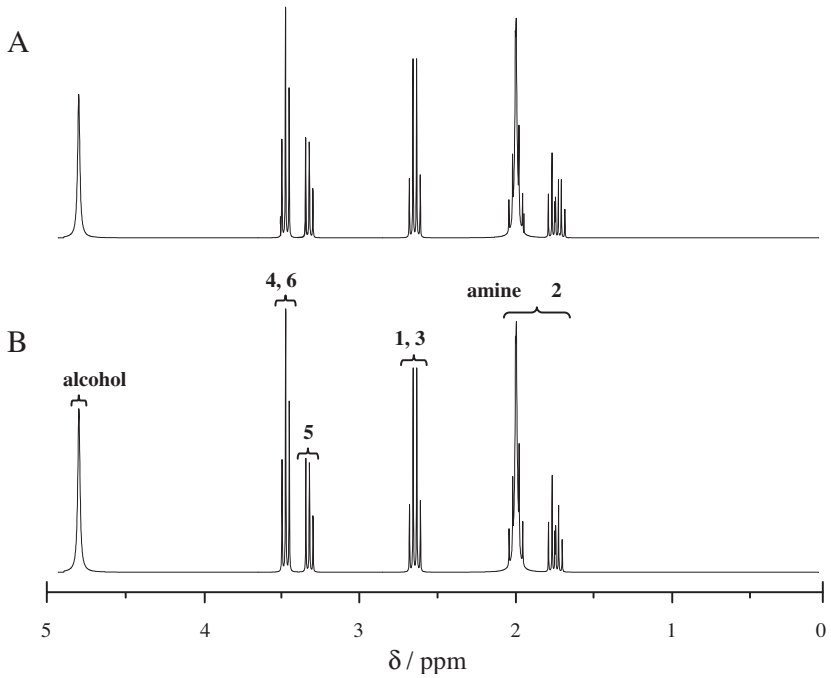

Fig. S8. ^{1}H NMR spectra of (*A*) 2-deoxystreptamine (**1a**) produced by *S. venezuelae* strain harboring pYJ496 expressing the *gtmB-gtmA-gacH* genes and (*B*) authentic **1a**.

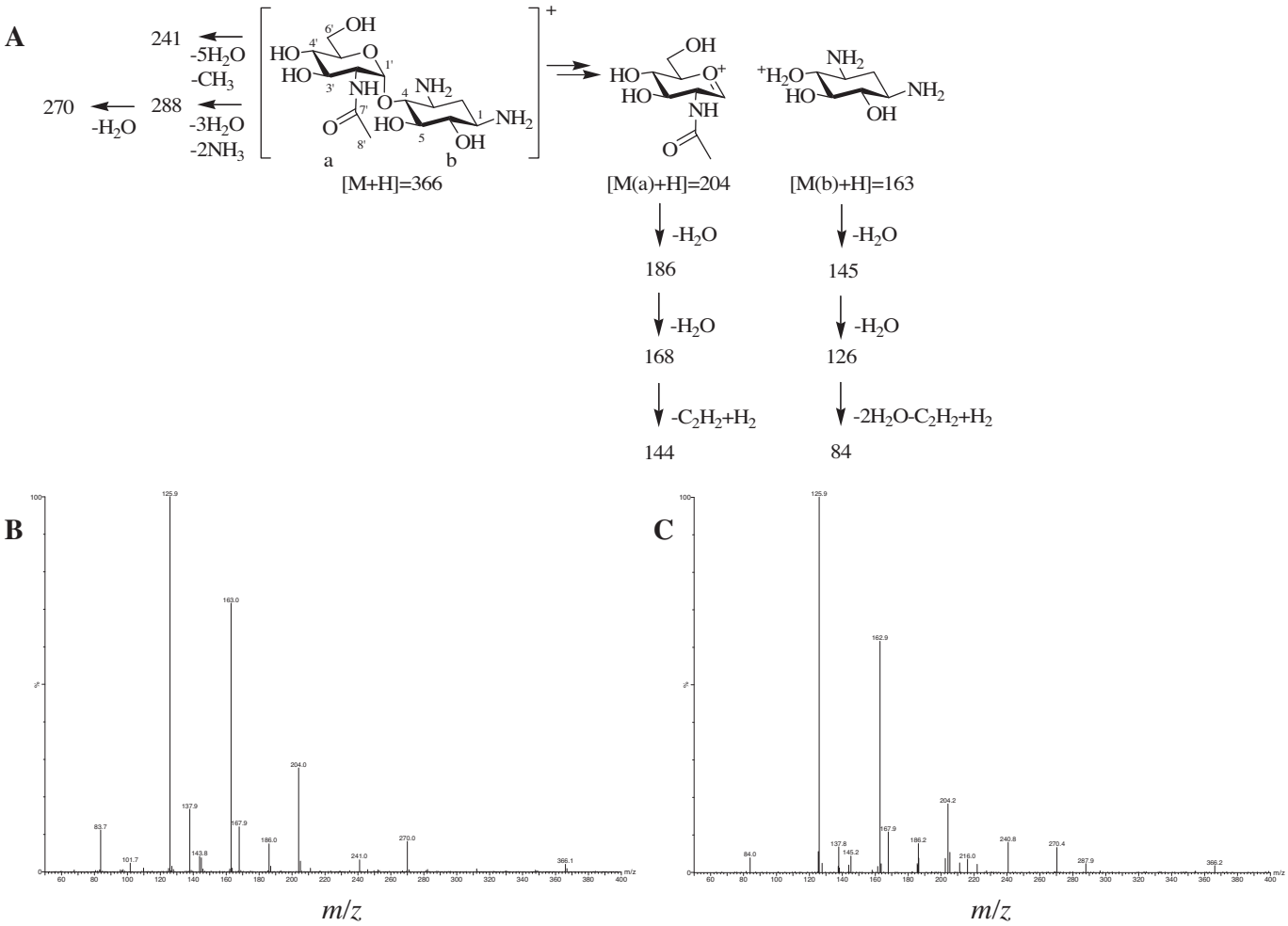

Fig. S9. (*A*) ESI-MS/MS fragmentation pattern of 2′-*N*-acetylparomamine (**2a**), and mass spectra of (*B*) **2a** produced by wild-type *M. echinospora* ATCC 15385, and (*C*) heterologously produced **2a** from *S. venezuelae* strain harboring pYJ498 expressing *gtmB-gtmA-gacH-gtmG* genes.

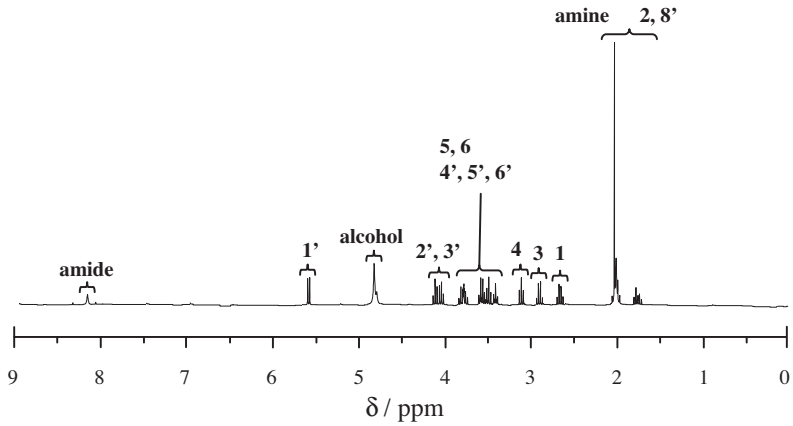

Fig. S10. ^{1}H NMR spectra of 2′-*N*-acetylparomamine (**2a**) produced by *S. venezuelae* strains harboring pYJ498 expressing *gtmB-gtmA-gacH-gtmG* genes.

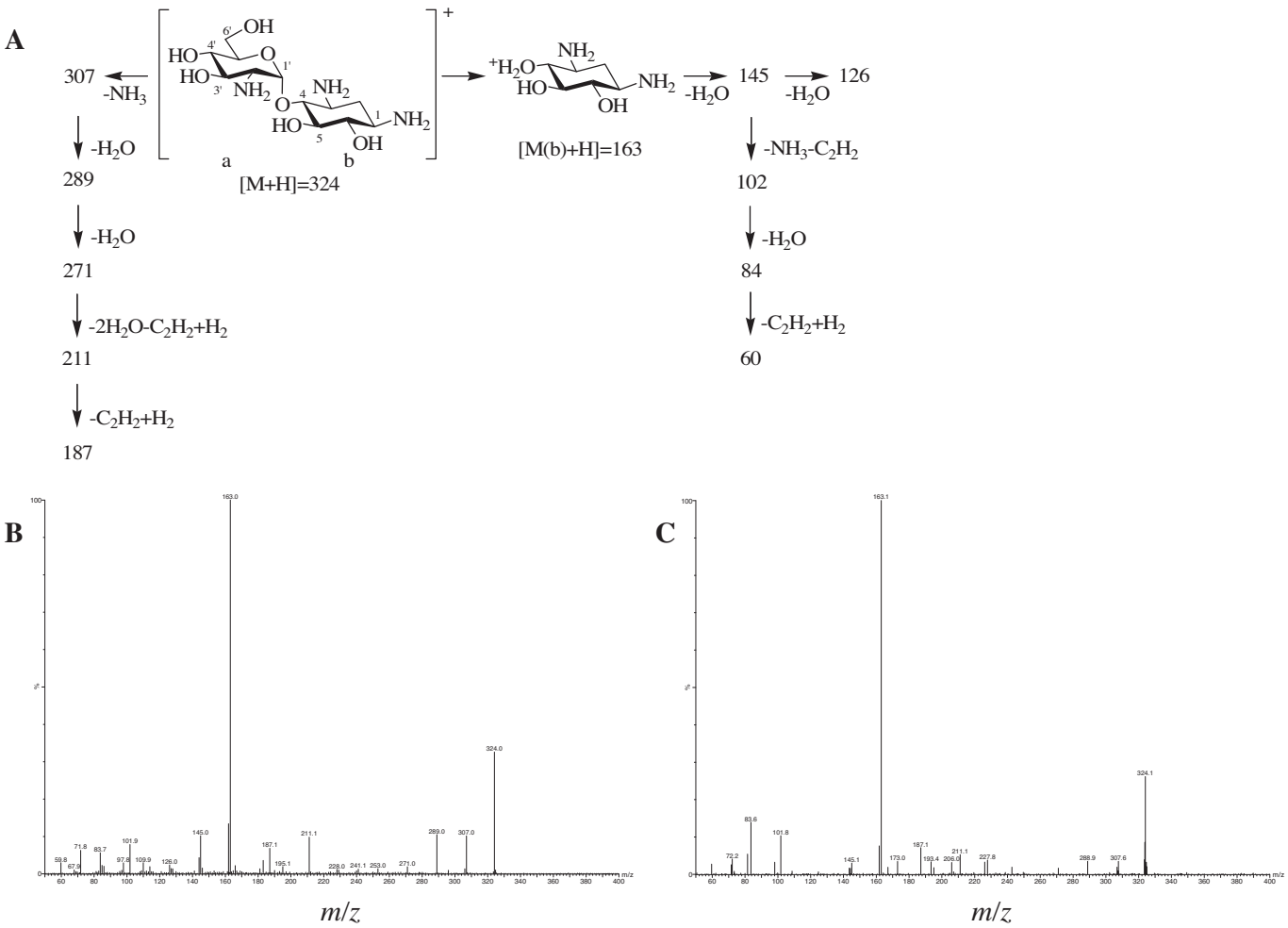

Fig. S11. (*A*) ESI-MS/MS fragmentation pattern of paromamine (**3a**), and mass spectra of (*B*) authentic **3a** obtained from hydrolysis of paromomycin, and (*C*) heterologously produced **3a** from *S. venezuelae* strain harboring pYJ503 containing the *gtmB-gtmA-gacH-gtmG-gtmM* genes.

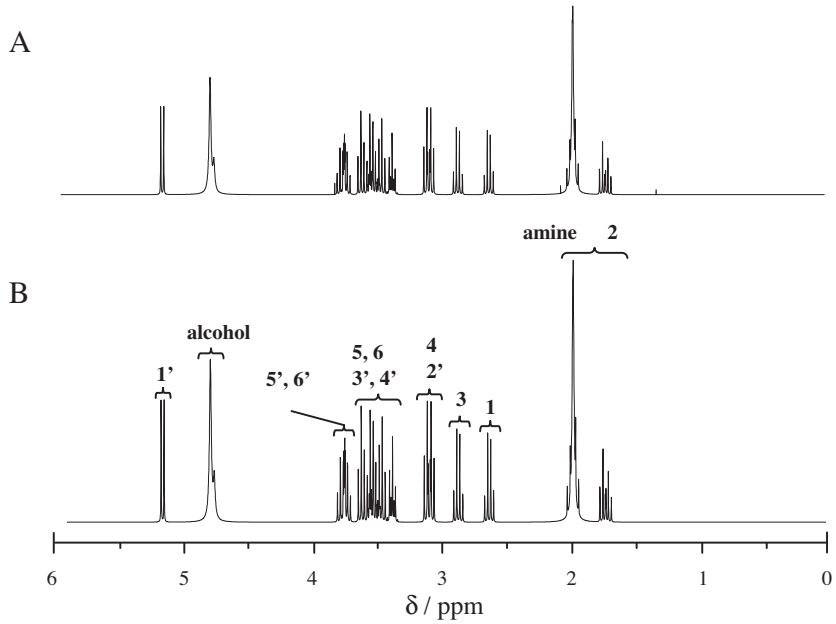

Fig. S12. ^{1}H NMR spectra of (*A*) paromamine (**3a**) produced by *S. venezuelae* strain harboring pYJ503 containing the *gtmB-gtmA-gacH-gtmG-gtmM* genes and (*B*) authentic **3a** obtained from hydrolysis of paromomycin.

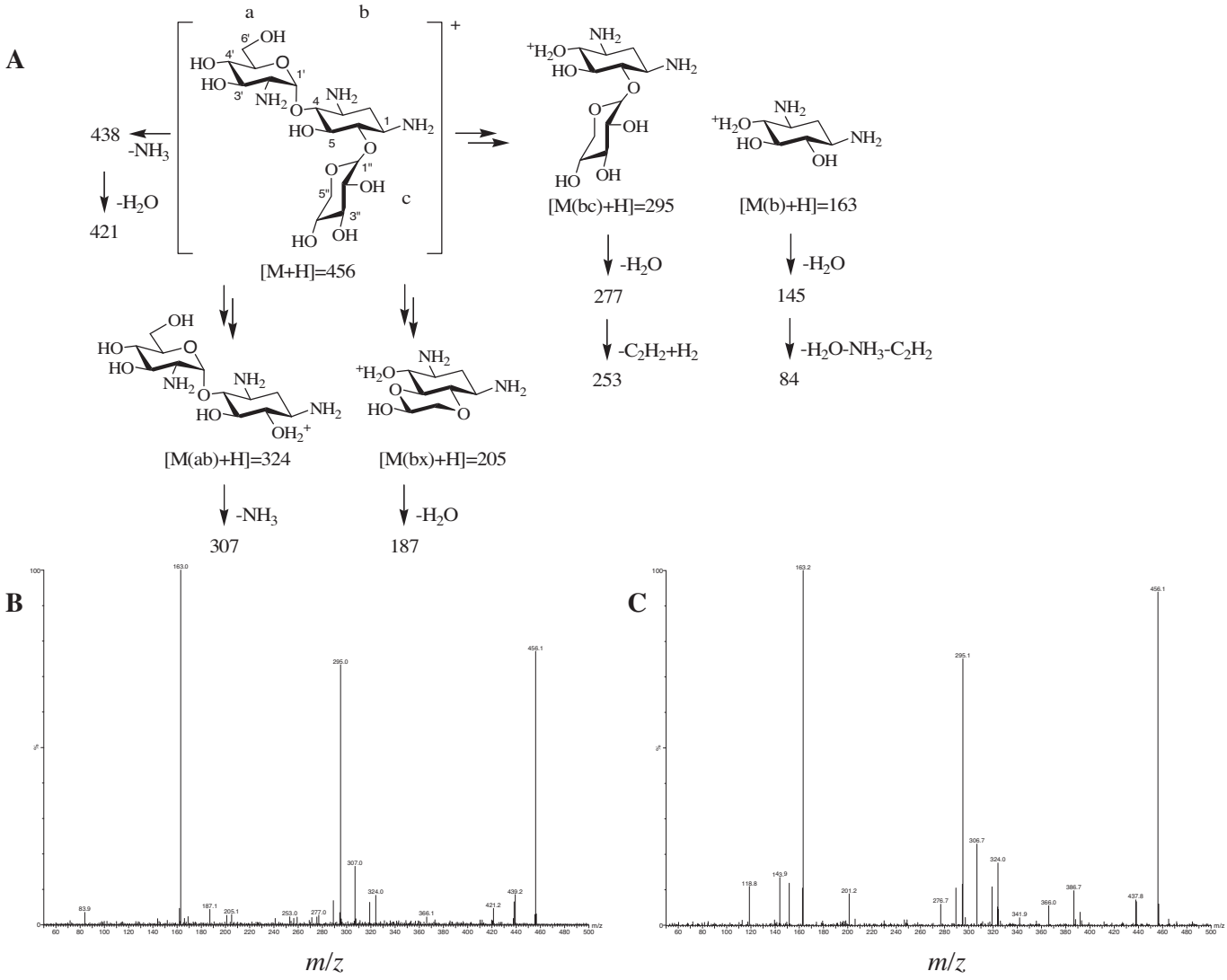

Fig. S13. (*A*) ESI-MS/MS fragmentation pattern of gentamicin A_2 (**4a**), and mass spectra of (*B*) **4a** produced by wild-type *M. echinospora* ATCC 15385, and (*C*) heterologously produced **4a** from *S. venezuelae* strain harboring pYJ505 containing the *gtmB-gtmA-gacH-gtmG-gtmE-gtmM* genes.

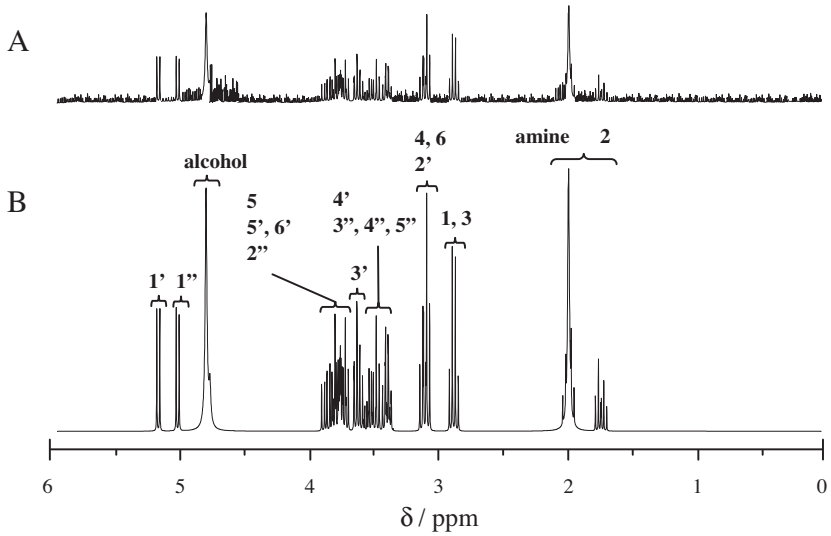

Fig. S14. ^{1}H NMR spectra of (*A*) gentamicin A_2 (**4a**) produced by *S. venezuelae* strain harboring pYJ505 containing the *gtmB-gtmA-gacH-gtmG-gtmE-gtmM* genes and (*B*) **4a** produced by wild-type *M. echinospora* ATCC 15385.

Table S1. Deoxyoligonucleotide primers used in amplification of genes

Gene	Primer sequence (5′ to 3′, restriction site in italics and bold)	Portion of gene	Restriction site
gtmF	**GGAA***CCATGG***CGAAAAGCTATTTCTGAATG**	5′	**NcoI**
	TGGT*ATCTAG***AGGCGTCATAACTACCCCCTT**	3′	**XbaI**
gtmK	**CCGC***CCTGCAGG***GTCACTCGGTACGCGACCGA**	5′	**SbfI**
	CGAT*TTCGAA***TCCGACCATATTGAGTGAGG**	3′	**BstBI**
gtmL	**GCCG***CCTGCAGG***GCCGCAGCTACCTGCACCAC**	5′	**SbfI**
	CCGC*AAGCTT***CGAGCGCCTCTGGGAGGACT**	3′	**HindIII**
gtmB	**TTT***GAATTC***GTATTACCGAACGGACATGG**	5′	**EcoRI**
	TTT*CTGCAG***GTGGTCAGCCGCGTAGTTCC**	3′	**PstI**
gtmA	**TTT***CTGCAG***ATCCAAGGGTTGTAGGGACC**	5′	**PstI**
	TTT*CCATGG***TCAACCATCGGCAGCACCCA**	3′	**NcoI**
gtmC	**TTT***CCATGG***TGGAGGAAGGCCGGGTGCCG**	5′	**NcoI**
	TTT*TCTAGA***TGTTGTCAGGCATTCATCTC**	3′	**XbaI**
gtmD	**TTT***AGATCT***AGGTAGCCGCTGATGACGCA**	5′	**BglII**
	TTT*GAATTC***ATCATAGGCTCTTCTTCAGC**	3′	**EcoRI**
gacH	**TTT***TCTAGA***TACGTCGGCTACTCGTTCCG**	5′	**XbaI**
	TTT*CCATGG***ACTTGTTGAGCCTCGATCAA**	3′	**NcoI**
gtmG	**TGTC***CCTGCAGG***GCTGCCCGGTCACTTCCCGC**	5′	**SbfI**
	TAGC*CCATGG***GCACTCTTCCGGAAGAATC**	3′	**NcoI**
gtmE	**CTCT***CCTGCAGG***CACCAGAAGCTGTGGAAGCC**	5′	**SbfI**
	ACGA*ACTAGT***TAGTGACGGTCATCTCAGGA**	3′	**SpeI**
btrD	**AACG***ACTAGT***AAAGAAGTGATCCGGGAAGG**	5′	**SpeI**
	CCTT*TTCGAA***GTGTCAGGTTTGAAGCGGTT**	3′	**BstBI**
kacA	**GGCG***ACTAGT***TACCGGGAGATCGGGCTGTG**	5′	**SpeI**
	GCCA*TTCGAA***TGCTCATAGCGACTCCTTGT**	3′	**BstBI**
neo16	**GCGC***TTCGAA***TCCACGAGGCCGGGTCACCGG**	5′	**BstBI**
	GGGC*ACTAGT***ACGGGAGGAGGAGCACGGTG**	3′	**SpeI**
gtmM	**CGG***ACTAGT***TACGTACAATCGCTCGAAAGGGCGTC**	5′	**SpeI**
	CGG*TTCGAA***GATGGTCGAGCCTCCGACGATCA**	3′	**BstBI**

Table S2. ^{1}H and ^{13}C NMR data (500 MHz, D_2O) for 2-deoxystreptamine (1a) produced by *S. venezuelae* strain harboring pYJ496 expressing the *gtmB-gtmA-gacH* genes (see Fig. S8)

Position	δ_H (m)	δ_C
1	2.59 (q)	53.02
2	1.72 (dt)	37.43
	1.98 (dt)	
3	2.57 (q)	53.33
4	3.39 (t)	79.62
5	3.31 (t)	74.67
6	3.36 (t)	79.93

Table S3. ^{1}H and ^{13}C NMR data (500 MHz, D_2O) for 2′-*N*-acetylparomamine (2a) produced by *S. venezuelae* strain harboring pYJ498 expressing *gtmB-gtmA-gacH-gtmG* genes (see Fig. S10)

Position	δ_H (m)	δ_C	Position	δ_H (m)	δ_C
1	2.65 (q)	51.40	1′	5.56 (d)	99.14
2	1.75 (dt)	35.43	2′	4.12 (t)	57.11
	1.99 (dt)		3′	4.03 (t)	71.28
3	2.88 (q)	49.34	4′	3.40 (t)	72.69
4	3.11 (t)	85.99	5′	3.73 (m)	77.38
5	3.55 (t)	72.50	6′	3.52 (m)	62.21
6	3.48 (t)	78.14		3.78 (m)	
			7′	—	170.37
			8′	2.03 (s)	23.30

Table S4. ^{1}H and ^{13}C NMR data (500 MHz, D_2O) for paromamine (3a) produced by *S. venezuelae* strain harboring pYJ503 containing the *gtmB-gtmA-gacH-gtmG-gtmM* genes (see Fig. S12)

Position	δ_H (m)	δ_C	Position	δ_H (m)	δ_C
1	2.60 (q)	51.23	1′	5.14 (d)	102.22
2	1.74 (dt)	35.48	2′	3.13 (m)	55.50
	1.97 (dt)		3′	3.63 (t)	74.04
3	2.85 (q)	49.31	4′	3.42 (t)	72.60
4	3.11 (m)	86.03	5′	3.74 (m)	77.39
5	3.57 (t)	72.44	6′	3.52 (m)	62.24
6	3.45 (t)	78.08		3.78 (m)	

Table S5. ^{1}H and ^{13}C NMR data (500 MHz, D_2O) for gentamicin A_2 (4a) produced by *S. venezuelae* strain harboring pYJ505 containing the *gtmB-gtmA-gacH-gtmG-gtmE-gtmM* genes (see Fig. S14)

Position	δ_H (m)	δ_C	Position	δ_H (m)	δ_C	Position	δ_H (m)	δ_C
1	2.82 (q)	49.51	1′	5.18 (d)	102.41	1′′	5.03 (d)	104.14
2	1.74 (dt)	35.53	2′	3.12 (m)	54.80	2′′	3.71–3.74 (u)*	73.61
	1.97 (dt)		3′	3.64 (t)	74.02	3′′	3.49 (t)	75.80
3	2.85 (q)	49.18	4′	3.38 (t)	73.01	4′′	3.42 (m)	70.34
4	3.10 (t)	86.22	5′	3.76 (m)	77.42	5′′	3.65 (d)	66.22
5	3.81–3.85 (u)*	72.30	6′	3.54 (m)	61.13		3.87–3.90 (u)[a]	
6	3.14 (t)	86.43		3.79 (m)				

*Undistinguishable signals due to the poor S/N ratio.