SUPPORTING INFORMATION

Mechanistic Studies of the Biosynthesis of 2-Thiosugar: Evidence for the Formation of an Enzyme-Bound 2-Ketohexose Intermediate in BexX-Catalyzed Reaction

*Eita Sasaki and Hung-wen Liu**

Division of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712

E-mail: h.w.liu@mail.utexas.edu

Materials. All chemicals and reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA), and were used without further purification unless otherwise specified. Sequencing grade trypsin was obtained from Promega (Madison, WI). Enzymes used for the cloning experiments were brought from Invitrogen (Carlsbad, CA) or New England Biolabs (Ipswich, MA). Ni-NTA agarose and kits for DNA gel extraction and spin minipreps were obtained from Qiagen (Valencia, CA). *Pfu*Ultra DNA polymerase was a product of Stratagene (La Jolla, CA). Growth medium components were acquired from Becton Dickinson (Sparks, MD). Reagents for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were purchased from Bio-Rad (Hercules, CA), with the exception of the protein molecular weight markers, which were obtained from Invitrogen. Sterile syringe filters were products of Fisher Scientific. Amicon YM-10 ultrafiltration membranes were brought from Millipore (Billerica, MA). Oligonucleotide primers were prepared by Invitrogen or Integrated DNA Technologies (Coralville, IA).

Bacterial Strains and Plasmids. *A. orientalis* subsp. *Vinearia* BA-07585 was generously provided by Banyu Pharmaceutical Co. (Tokyo, Japan). *Escherichia coli* DH5α, acquired from Bethesda Research Laboratories (Gaithersburg, MD), was used for routine cloning experiments. Protein overexpression host *E. coli* BL21 star (DE3) was obtained from Invitrogen. Vector pET24b(+) for protein overexpression was purchased from Novagen (Madison, WI).

General. DNA sequencing was performed by the core facility of the Institute of Cellular and Molecular Biology at the University of Texas, Austin. Vector NTI Advance 10.1.1 from Invitrogen was used for sequence alignments. Mass spectra were recorded by the Mass Spectrometry core facility in the Department of Chemistry and Biochemistry and the College of Pharmacy at the University of Texas, Austin. Standard genetic manipulations of *E. coli* were performed as described by Sambrook et al.[1]

Cloning of bexX. The bexX gene was PCR-amplified from cosmid C006 previously constructed from the genomic DNA of *A. orientalis*[2] using primers with engineered *NdeI* and *HindIII* restriction sites. The sequences of the primers are 5′-TCTTAAA**CAT<u>ATG</u>**ACCATTCCGCACATCGGC-3′ (forward), 5′-ATAT**AAGCTT**GACCGCCGCTGCTACCTC-3′ (reverse). The engineered restriction sites are shown

in bold, the start codon in bold and also underlined. The PCR-amplified gene fragments were purified, digested with *Nde*I and *Hind*III, and ligated into pET24b(+) vector which was also digested with the same enzymes. The resulting plasmid, *bexX*/pET24b(+), was sequenced using the T7 or T7 terminal universal primer and used to transform *E. coli* BL21 star (DE3) strain for protein overexpression. The BexX enzyme was expressed as a *C*-terminal His_6-tagged protein.

Expression and Purification of *C*-His_6-BexX. An overnight culture of *E. coli* BL21 star (DE3)-*bexX*/pET24b(+), grown in the LB medium (10 mL) containing 50 □g/mL of kanamycin at 37 °C, was used to inoculate 1 L of the same growth medium. The culture was incubated at 37 °C with shaking (230 rpm) until the OD_{600} reached ~ 0.5. Protein expression was then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM, and the cells were allowed to grow at 18 °C and 125 rpm for an additional 24 h. The cells were harvested by centrifugation at 4500*g* for 15 min and stored at -80 °C until lysis. All purification steps were carried out at 4 °C using Ni-NTA resin according to the manufacturer's protocol with minor modifications. Specifically, the thawed cells (~5 g) were resuspended in the lysis buffer (20 mL) containing 10% (v/v) glycerol and 10 mM imidazole. After incubation with lysozyme (20 mg) for 30 min, the cells were disrupted by sonication using 10 × 10-s pulses with a 30-s cooling pause between each pulse. The resulting lysate was centrifuged at 20000*g* for 20 min, and the supernatant was subjected to Ni-NTA chromatography. Bound protein was eluted using 250 mM imidazole buffer containing 10% glycerol. The collected protein solution was dialyzed against 3 × 1-L of 50 mM Tris·HCl buffer (pH 8) containing 300 mM NaCl and 15% glycerol. The protein solution was then flash-frozen in liquid nitrogen and stored at -80 °C until use. Protein concentration was determined by the Bradford assay using bovine serum albumin as the standard.[3] The yield of *C*-His_6-BexX was approximately 50 mg from 1 L culture. The molecular mass and purity (> 90%) of the protein were estimated by SDS-PAGE analysis (Figure S1).

Construction of K110A *bexX* mutant. The site-specific K110A *bexX* mutant was constructed according to the QuikChange site-directed mutagenesis protocol (Stratagene) using the plasmid, *bexX*/pET24b(+) as the template. The PCR primers were 5′-GGCATCGAGATCCTGGCCCTGGACG-TGCGC-3′ (forward) and 5′-GCGCACGTCCAGGGCCAGGATCTCGATGCC-3′ (reverse) with codons for the mutated residue underlined. The resulting plasmid K110A-*bexX*/pET24b(+) was used to transform *E. coli* BL21 star (DE3) strain for protein overexpression. The K110A BexX mutant was produced as a *C*-terminal His_6-tagged protein.

Expression and Purification of the *C*-His_6-K110A-BexX mutant. The overexpression and protein purification of the *C*-His_6-K110A-BexX mutant were performed according to the same procedure described above for the wild type *C*-His_6-BexX. The yield of the mutant protein, *C*-His_6-K110A-BexX, was approximately 120 mg from 1 L culture. The molecular mass and purity (> 90%) of the protein were estimated by SDS-PAGE analysis (Figure S1).

ESI-MS analysis of *C*-His_6-BexX. The purified *C*-His_6-BexX protein stock solution was added to 50 mM Tris·HCl buffer (pH 7.5) containing 10 mM $MgCl_2$ (final concentration of protein was 13 μM) and incubated at 30 °C for 2 h. The resulting solution (200 μL) was cooled on ice. To this solution was added 20 μL of cold $NaBH_4$ solution in methanol (530 mM) (final concentration of $NaBH_4$ was 50 mM). The mixture was incubated on ice for 1 h. The excess salts and glycerol were removed by using an Amicon ultrafiltration unit equipped with a YM-10 membrane at 4 °C. The resulting protein solution was subjected to ESI-MS analysis. The ESI-MS spectrum is shown in Figure 1A.

ESI-MS analysis of *C*-His_6-BexX with various sugars. Similarly, the purified *C*-His_6-BexX protein (13 μM final) was mixed with 5 mM of either D-glucose 6-phosphate (D-G6P, **2**), D-fructose 6-phosphate (D-F6P, **13**), D-glucose (D-Glc), or 1-deoxy-D-xylose 5-phosphate (DXP, **6**) in 50 mM Tris·HCl buffer (pH 7.5) containing 10 mM $MgCl_2$. The resulting solution (200 μL) was incubated at

30 °C for < 2 h and cooled on ice. A shorter incubation time (0.5 h) was used for D-Glc- and DXP-containing mixtures because protein started to precipitate beyond this time period. An aliquot of 20 μL of cold $NaBH_4$ solution in methanol (530 mM) was added to each reaction (final concentration of $NaBH_4$ was 50 mM), and the resulting mixture was incubated on ice for 1 h. The excess salts and glycerol were removed by using an Amicon ultrafiltration unit equipped with a YM-10 membrane at 4 °C. The resulting protein solution was subjected to ESI-MS analysis. The ESI-MS spectra are shown in Figure 1B (D-G6P) and Figure S2 (D-F6P, D-Glc and DXP).

ESI-MS analysis of the *C*-His_6-K110A-BexX mutant. Similar ESI-MS analysis was performed for the *C*-His_6-K110A-BexX mutant. The purified mutant protein (30 μM final) was incubated with or without 5 mM of D-G6P (**2**) in 50 mM Tris·HCl buffer (pH 7.5) containing 10 mM $MgCl_2$. To prevent protein precipitation, the mixture (400 μL each) was incubated at room temperature for 10 min (similar conditions as for the wild-type enzyme caused significant precipitation of the K110A mutant). An aliquot of 40 μL of cold $NaBH_4$ solution in methanol (530 mM) was added to the each reaction (final concentration of $NaBH_4$ was 50 mM), and the resulting mixture was incubated on ice for 1 h. The excess salts and glycerol were removed by using an Amicon ultrafiltration unit equipped with a YM-10 membrane at 4 °C. The resulting protein solution was subjected to ESI-MS analysis. The ESI-MS spectra are shown in Figure S3.

Protein sequence alignment and active site prediction. The protein sequence of BexX was aligned with ThiG from *Bacillus subtilis* using Vector NTI Advance 10.1.1 (Figure S4). The sequence of ThiG was obtained from the National Center for Biotechnology Information (NCBI) database. (Gene ID: 936422).[4]

Trypsin digestion and LC-MS/MS analysis of the protein–substrate complex. The purified *C*-His_6-BexX protein (13 μM final) was incubated with or without 5 mM of D-G6P (**2**) in 50 mM Tris·HCl buffer (pH 7.5) containing 10 mM $MgCl_2$ at 30 °C for 1 h. An aliquot of 20 μL of cold $NaBH_4$ solution in methanol (530 mM) was added to 200 μL of each protein solution (final concentration of $NaBH_4$ was 50 mM) and the resulting mixture was incubated on ice for 1 h. The excess salts and glycerol were removed by using an Amicon ultrafiltration unit equipped with a YM-10 membrane at 4 °C. An aqueous dithiothreitol solution (14 μL, final concentration 10 mM), acetonitrile (15 μL), and trypsin stock solution (5 μL containing 2 μg of trypsin) were sequentially added to each of the protein solution and the resulting mixture was incubated at 37 °C overnight. The reaction was stopped by frozen at -20 °C until use. The frozen sample was submitted to the Mass Spectrometry core facility and subjected to LC-ESI-MS/MS analysis on a nanoACQUITY UPLC (Waters) in conjunction with a Q-TOF premier mass spectrometer (Waters). The expected digestion pattern of *C*-His_6-tagged BexX is shown in Figure S5. The LC-ESI-MS and MS/MS spectra of the target peptides (**16** and **17**) are shown in Figure 2, Figure 3, and Figure S6.

Trapping the ketone intermediate using 2,4-dinitrophenylhydrazine (DNPH). The trypsin-digested solution of *C*-His_6-BexX**–**D-G6P was prepared according to the procedure described above. The reaction mixture (40 μL) was added to a DMSO solution of 2,4-dinitrophenylhydrazine (DNPH, **18**) (160 μL, final concentration of DNPH was 80 mM). The resulting mixture was acidified to pH ~ 4 by adding 8 μL of 10% trifluoroacetic acid (TFA) and incubated at 37 °C overnight. The reaction solution was subjected to LC-ESI-MS analysis as described above. The MS spectrum of the target peptide (**19**) is shown in Figure 4A.

Trapping the ketone intermediate using hydroxylamine (NH_2OH). The purified *C*-His_6-BexX protein (75 μM final concentration) was incubated with 5 mM of D-G6P (**2**) in 50 mM Tris·HCl buffer (pH 7.5) containing 10 mM $MgCl_2$ at 30 °C for 0.5 h. An aqueous solution of hydroxylamine (NH_2OH) (10 μL, final concentration 20 mM) was added to the protein mixture (90 μL) and the resulting mixture

was incubated at room temperature for 1 h. After cooling on ice, the excess salts and glycerol were removed by using an Amicon ultrafiltration unit equipped with a YM-10 membrane at 4 °C. The obtained protein solution was subjected to ESI-MS analysis. The ESI-MS spectrum of the resulting protein complex (*C*-His_6-BexX–D-G6P–NH_2OH, **20**) is shown in Figure 4B.

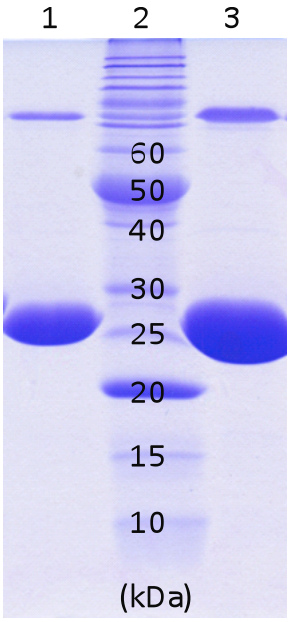

Figure S1. SDS-PAGE gel of purified *C*-His_6-tagged BexX (268 aa, 28.5 kDa) (lane 1) and *C*-His_6-tagged K110A BexX (268 aa, 28.4 kDa) (lane 3). The molecular weight marks are 220, 160, 120, 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 15, and 10 kDa (top to bottom, lane 2). The protein band ~ 80 kDa is a minor impurity from endogenous *E. coli* proteins.

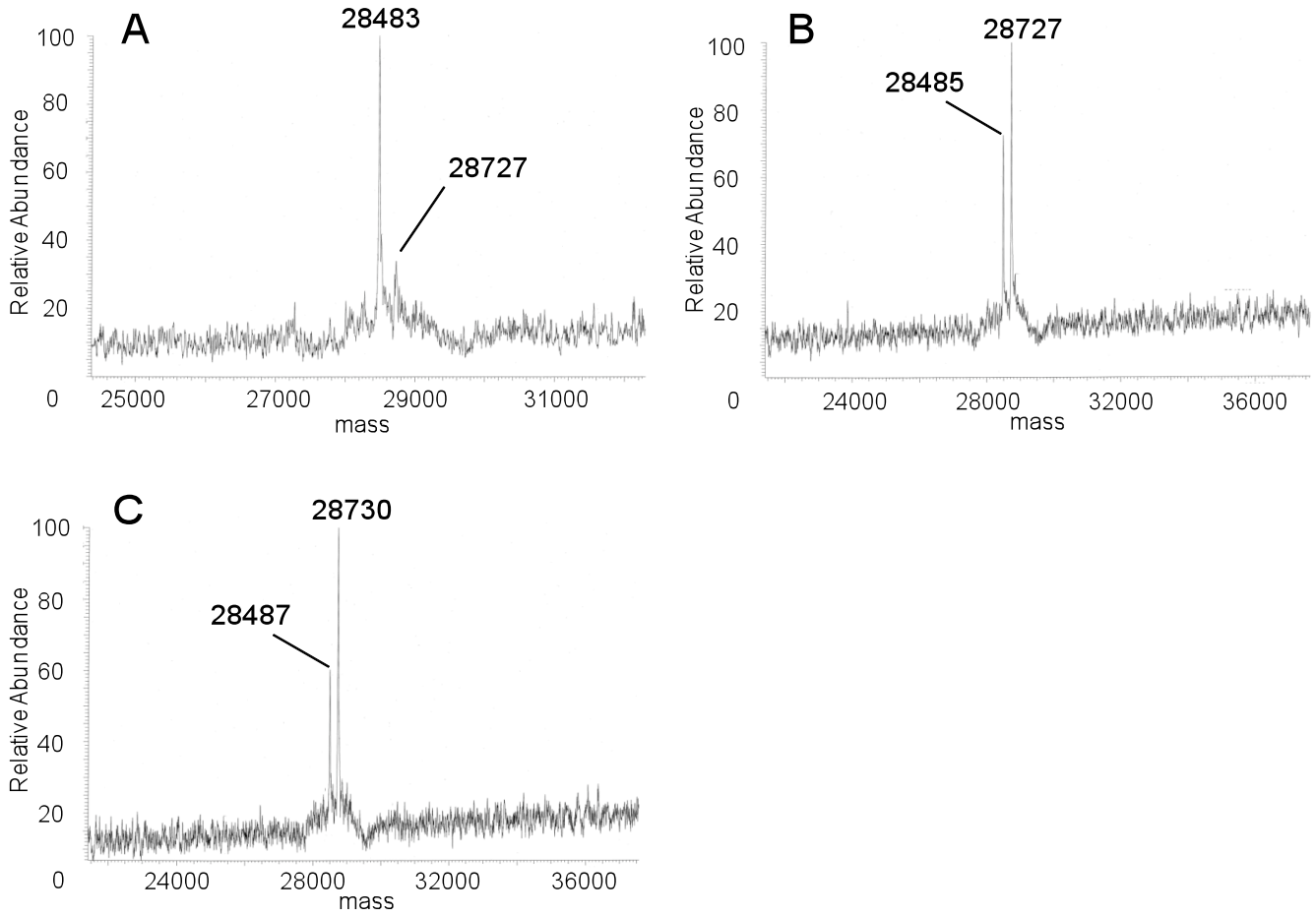

Figure S2. ESI-MS of *C*-His_6-tagged BexX and each sugar monophosphate mixture: (A) D-F6P, (B) D-Glc, and (C) DXP. Calculated molecular weights of *C*-His_6-BexX (268 aa) and *C*-His_6-BexX-D-G6P (reduced) are 28488, and 28732, respectively. Calculated molecular weights of D-G6P, D-F6P, D-Glc and DXP are 260, 260, 180 and 214, respectively. Since part of the as-isolated BexX contains varied amount of bound D-G6P (such as **11**), which upon reduction, gives *m/z* signals at 28727 or 28730 (molecular weight of the reduced Bex-D-G6P complex).

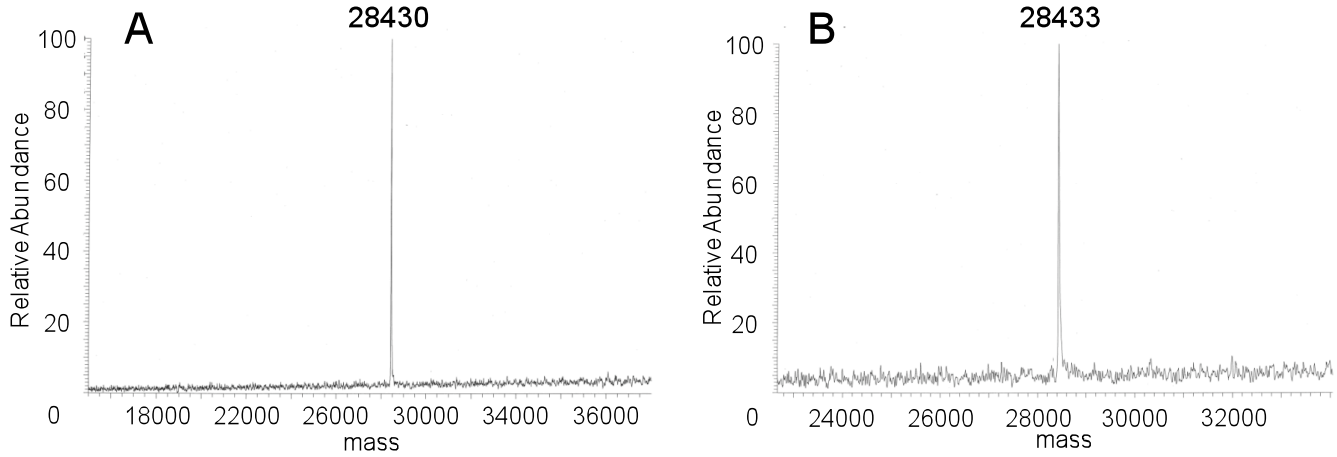

Figure S3. ESI-MS of K110A BexX mutant incubated with (A) or without (B) D-G6P prior to the $NaBH_4$ treatment. Calculated molecular weight of *C*-His_6-tagged K110A BexX (268 aa) is 28431.

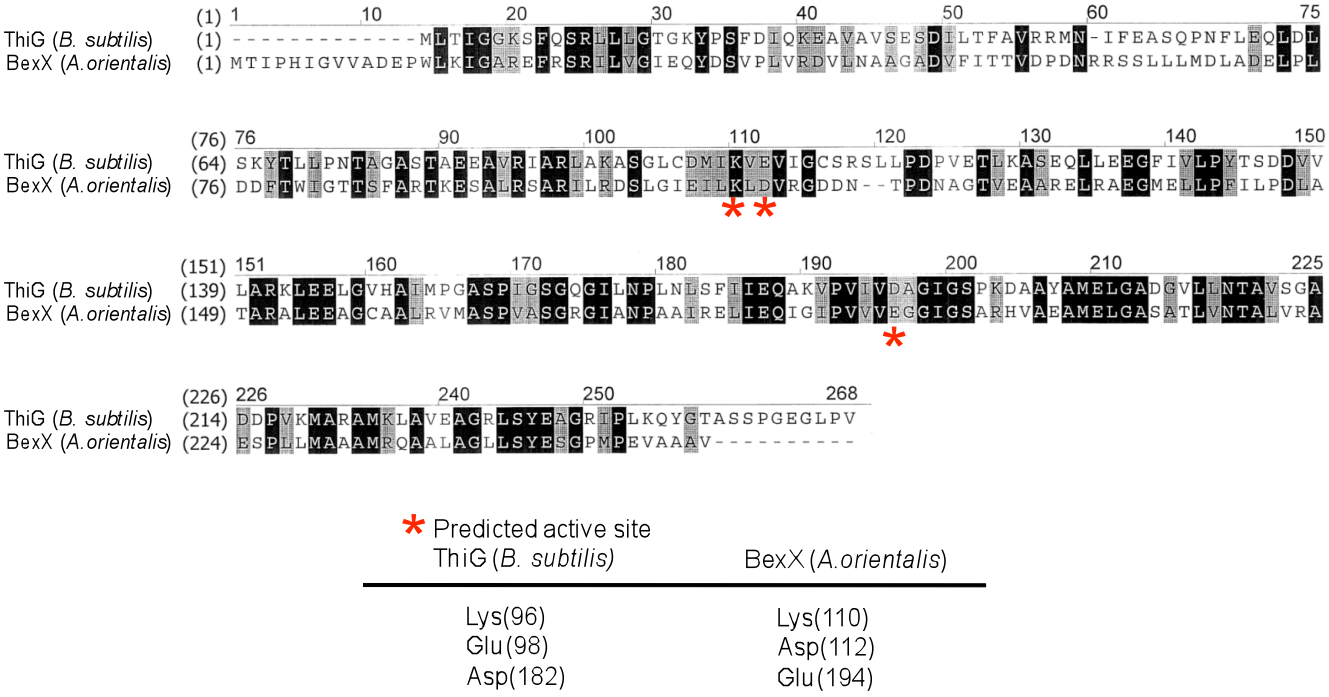

Figure S4. Protein sequence alignment of BexX and ThiG from *Bacillus subtilis*.[4,5] Identity and similarity of the sequences are 37% and 59%, respectively.

(*N*-term) (M)TIPHIGVVADEPWL**K** / IGAR / EFR / SR / ILVGIEQYDSVPLVR /
DVLNAAGADVFITTVDPDNR / R / SSLLLMDLADELPLDDFTWIGTTSFAR / T**K** / ESALR /
SAR / ILR / DSLGIEIL**K*** LDVR / GDDNTPDNAGTVEAAR / ELR /
AEGMELLPFILPDLATAR / ALEEAGCAALR / VMASPVASGRGIANPAAIR /
ELIEQIGIPVVVEGGIGSAR / HVAEAMELGASATLVNTALVR / AESPLLMAAAMR /
QAALAGLLSYESGPMPEVAAAV**K** / LAAALEHHHHHH (*C*-term)

Figure S5. The predicted digestion pattern of *C*-His$_6$-tagged BexX. Lysine residues are shown in bold. The predicted active site lysine (K110) is shown with an asterisk. The detected peptide fragments by LC-ESI-MS are underlined.

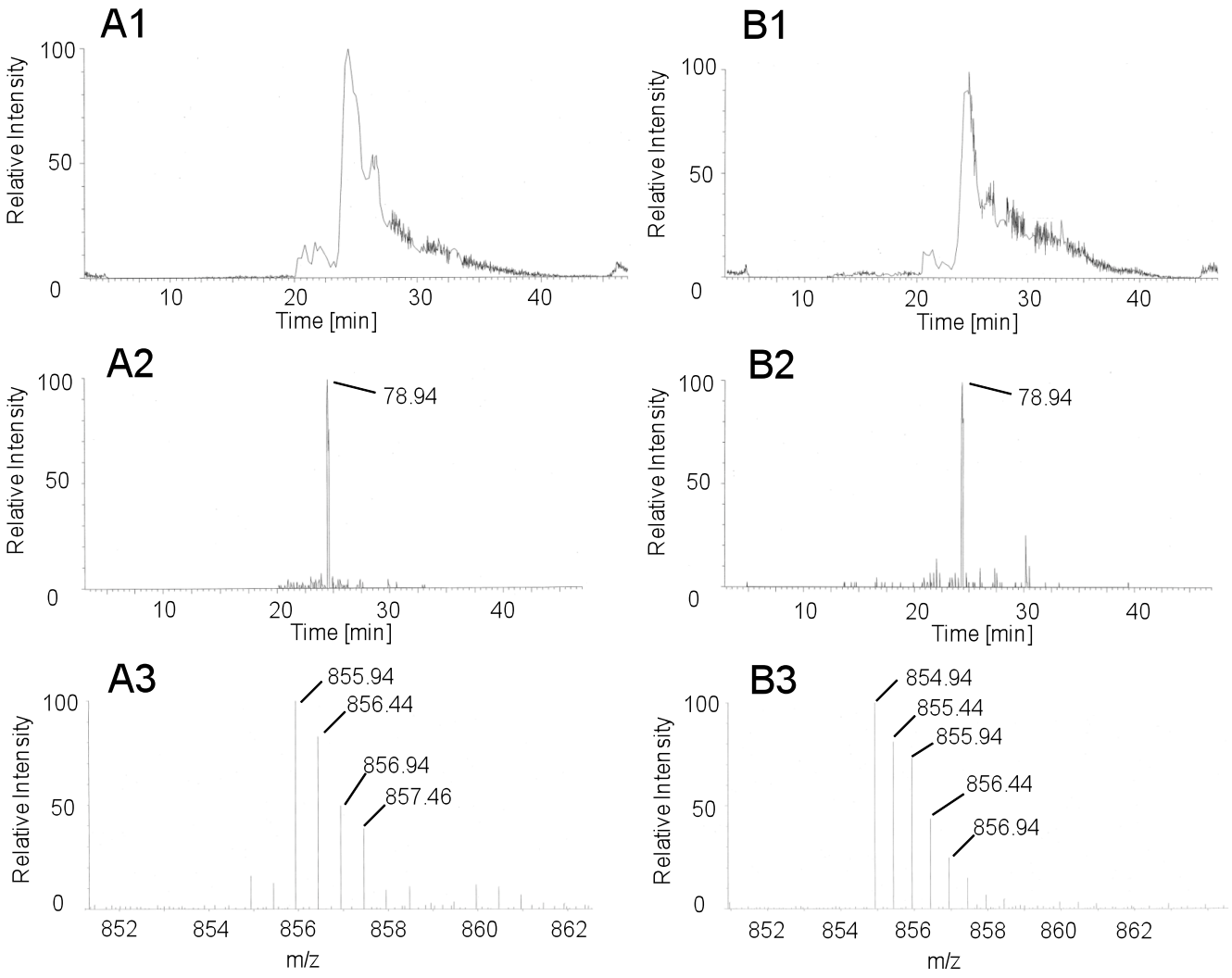

Figure S6. LC-MS of trypsin-digested BexX-D-G6P with (A) or without (B) $NaBH_4$ treatment (negative mode analysis). A1 and B1 show the target peak intensity chromatogram (854.9 ± 0.5 Da and 856.0 ± 2 Da, respectively). A2 and B2 show the extracted LC-MS/MS chromatogram containing the target peak, 79 ± 1 Da, which corresponds PO_3^-. Calculated isotopic mass of PO_3^- is 78.96. A3 and B3 show the mass spectra of the target peptide peaks, **16** and **17**, respectively (see Scheme 4 for the structures). Calculated m/z values of **16** and **17** are 855.93 (M – $2H^+$, z = 2) and 854.92 (M – $2H^+$, z = 2), respectively.

References

(1) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Mannual,* 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(2) Sasaki, E.; Ogasawara, Y.; Liu, H.-w. *J. Am. Chem. Soc*. **2010**, *132,* 7405–7417.

(3) Bradford, M. M. *Anal. Biochem*. **1976**, *72,* 248–254.

(4) Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessières, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Cummings, N. J.; Daniel, R. A.; Denizot, F.; Devine, K. M.; Düsterhöft, A.; Ehrlich, S. D.; Emmerson, P. T.; Entian, K. D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S. Y.; Glaser, P.; Goffeau, A.; Golightly, E. J.; Grandi, G.; Guiseppi, G.; Guy, B. J.; Haga, K.; Haiech, J.; Harwood: C. R.; Hénaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M. F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Mauël, C.; Médigue, C.; Medina, N.; Mellado, R. P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Poetetelle, D.; Porwollik, S.; Prescott, A. M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlan, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Serror, P.; Shin, B. S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H. F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Nature* **1997**, *390,* 249–256.

(5) Park, J.-H.; Dorrestein, P. C.; Zhai, H.; Kinsland, C.; McLafferty, F. W.; Begley, T. P. *Biochemistry* **2003**, *42,* 12430–12438.