Exploiting Sulphur-Carrier Proteins from Primary Metabolism for 2-Thiosugar Biosynthesis

Eita Sasaki[1], Xuan Zhang[2], He G. Sun[1], Mei-Yeh Jade Lu[3,4], Tsung-lin Liu[4,5], Albert Ou[4], Jeng-yi Li[3], Yu-hsiang Chen[3], Steven E. Ealick[2], and Hung-wen Liu[1]*

[1]Division of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA. [2]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. [3]Biodiversity Research Center and [4]Genomics Research Center, Academia Sinica, Taipei 115, Taiwan, ROC. [5]Institute of Bioinformatics and Biosignal Transduction, National Cheng-Kung University, Tainan 701, Taiwan, ROC.

*e-mail: h.w.liu@mail.utexas.edu

Contents

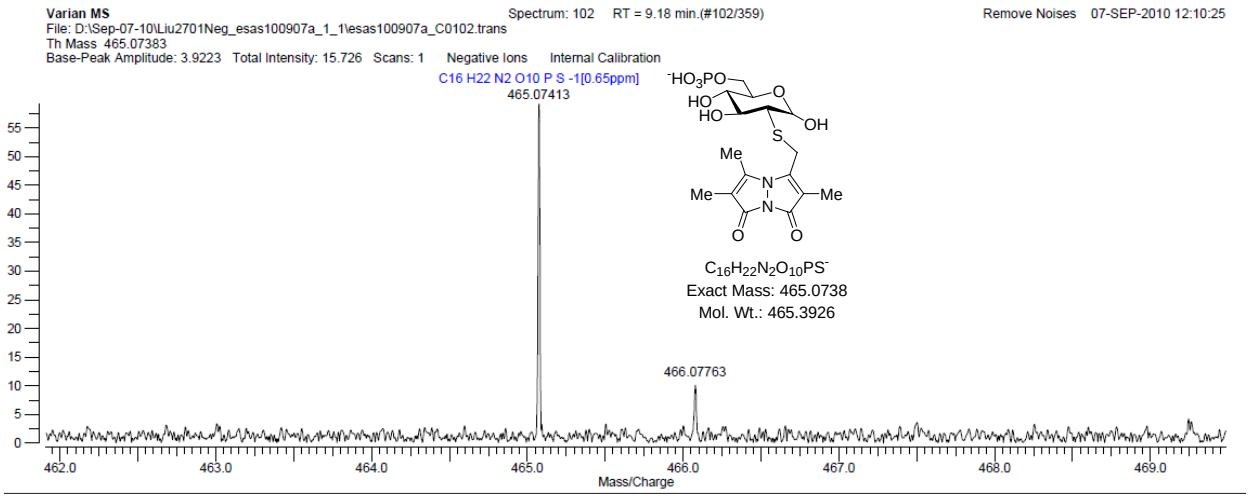
S1. Supplementary Methods

General. 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. Calf intestinal alkaline phosphatase (CIP) was obtained from New England Biolabs (Ipswich, MA). Enzymes and molecular weight standards used for the cloning experiments were products of Invitrogen (Carlsbad, CA) or New England Biolabs. Kits for DNA gel extraction and spin minipreps were acquired from Qiagen (Valencia, CA). PureLink Genomic DNA Mini Kit from Invitrogen was used for genomic DNA extraction. Growth medium components were obtained from Becton Dickinson (Sparks, MD). Vector pET28b(+) for protein over-expression were purchased from Novagen. Oligonucleotide primers were ordered from Integrated DNA Technologies (Coralville, IA) for constructing plasmids, *thiS*/pET28b(+), *moaD*/pET28b(+), *cysO*/pET28b(+), *moaD2*/pET28b(+), *moeZ*/pET28b(+), and *cd4*/pETH28b(+), and the sequences are shown below. The engineered restriction sites are shown in bold, the start codon is shown in bold and also underlined, and the stop codon is italic.

Primer name	Sequence
thiS for pET28b(+)-forward	5'-TCTATAA**CATATG**GAGATCAAGCTCAACGG-3'
thiS for pET28b(+)-reverse	5'-TTAC**AAGCTT***CA*GCCTCCCTGGACGG-3'
moaD for pET28b(+)-forward	5'-TCTATAA**CATATG**ACCACCCTGACCATCG-3'
moaD for pET28b(+)-reverse	5'-TTAC**AAGCTT**AGAGAAGCC*TCA*GCCGC-3'
cysO for pET28b(+)-forward	5'-TCTATAA**CATATG**GCCGTGACCGTCTCC-3'
cysO for pET28b(+)-reverse	5'-TTAC**AAGCTT***CA*GCCACCGGCCACG-3'
moaD2 for pET28b(+)-forward	5'-TCTATAA**CATATG**ACCGTGCGGATCACC-3'
moaD2 for pET28b(+)-reverse	5'-TTAC**AAGCTT***CA*GCCTCCCGCGACC-3'
moeZ for pET28b(+)-forward	5'-TCTATAA**CATATG**GACGCGAACCGCGCGC-3'
moeZ for pET28b(+)-reverse	5'-TTAC**AAGCTT***CA*GTAGGTCGGCAGGCTCGG-3'
cd4 for pET28b(+)-forward	5'-TCTATAA**CATATG**ACCTATCTCGACCACGCG-3'
cd4 for pET28b(+)-reverse	5‘-TTAC**AAGC*TT***ACACCTCTTGCTTCTGGG-3'

PfuUltra DNA polymerase was purchased from Stratagene (La Jolla, CA). Ni-NTA agarose was obtained from Qiagen (Valencia, CA). Amicon YM-10 ultrafiltration membranes were products of Millipore (Billerica, MA). Reagents for sodium dodecyl sulphate-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 or New England Biolabs. The CarboPac PA1 high-performance liquid chromatography (HPLC) column was acquired from Dionex (Sunnyvale, CA). The analytical C_{18} HPLC column is a product of Varian (Palo Alto, CA). *Amycolatopsis 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. The protein overexpression hosts *E. coli* BL21 star (DE3) was obtained from Invitrogen. Standard genetic manipulations of *E. coli* were performed as described by Sambrook and Russell (Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Mannual, 3rd ed., Cold Spring Harbor Laboratory Press, 2001). Routine 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.

***1.2.* Characterization of the BexX reaction product derivatized by mBBr (2-thio-D-glucose 6-phosphate-bimane).** ^{1}H NMR (D_2O, 500 MHz) δ 5.23 (d, $J_{1\text{-}2}$ = 3.0 Hz, 1H, 1α-H), 4.74 (d, $J_{1\text{-}2}$ = 9.0 Hz, 1H, 1β-H), 4.04-3.90 (m, 8H, 6α-H, $6_β$-H, 8′α-H, and 8′β-H), 3.89-3.84 (brd, 1H, 5α-H), 3.70 (dd, $J_{2\text{-}3}$ = 11.0, $J_{3\text{-}4}$ = 9.5 Hz, 1H, 3α-H), 3.49–3.42 (m, 1H, 5β-H), 3.45 (dd, $J_{3\text{-}4}$ = $J_{4\text{-}5}$ = 9.5 Hz, 1H, 4α-H), 3.41 (dd, $J_{3\text{-}4}$ = $J_{4\text{-}5}$ = 9.0 Hz, 1H, 4β-H), 3.35 (dd, $J_{2\text{-}3}$ =10.5, $J_{3\text{-}4}$ = 9.0 Hz, 1H, 3β-H), 2.80 (dd, $J_{2\text{-}3}$ = 11.0, $J_{1\text{-}2}$ = 3.0 Hz, 1H, 2α-H), 2.52 (dd, $J_{2\text{-}3}$ = 10.5, $J_{1\text{-}2}$ = 9.0 Hz, 1H, 2β-H), 2.41 (s, 3H, 9′β-H), 2.40 (s, 3H, 9′α-H), 1.78 (s, 3H, 7′β-H), 1.77 (s, 3H, 7′α-H), 1.69 (s, 6H, 10′α-H, and 10′β-H). ^{13}C NMR (D_2O, 125 MHz), only signals identified in the HSQC spectrum are reported here: δ 93.1 (1α-C), 74.9 (5β-C), 73.6 (3β-C), 72.4 (3α-C), 71.7 (5α-C), 70.6 (4α-C and $4_β$-C), 64.4 (6α-C and 6β-C), 54.5 (2β-C), 51.9 (2α-C), 25.2 (8′α-C and 8′β-C), 11.2 (9′α-C and 9′β-C), 6.3 (7′α-C and 7′β-C), 5.9 (10′α-C and 10′β-C). ^{31}P NMR (D_2O, 202 MHz) phosphate anomers δ = 2.37. High-resolution ESI-MS (negative) calcd for $C_{16}H_{22}N_2O_{10}PS^-$ $[M - H]^-$ 465.0738, found 465.0741.



S2. Supplementary Table

Supplementary Table 1 | BLAST analysis of the genes near *thiS, moaD, cysO, moaD2*, and *moeZ* in *A. orientalis* genome (1/3)

gene number	putative functions	protein homologue and origin	identity / similarity (%)	protein accession number
13945	**ThiC**	thiamine biosynthesis protein ThiC [*Amycolatopsis mediterranei* U32]	96 / 98	YP_003770668
13947	**ThiD**	phosphomethylpyrimidine kinase [*Amycolatopsis mediterranei* U32]	93 / 96	YP_003770669
13950		No significant similarity found		
13952		No significant similarity found		
13954		transcriptional regulator, HxlR family protein [*Streptomyces bingchenggensis* BCW-1]	60 / 74	ADI04334
13957		alcohol dehydrogenase zinc-binding domain-containing protein [*Stackebrandtia nassauensis* DSM 44728]	63 / 75	YP_003511248
13960	**ThiD**	phosphomethylpyrimidine kinase [*Amycolatopsis mediterranei* U32]	96 / 97	YP_003770671
13961		lipoprotein [*Amycolatopsis mediterranei* U32]	95 / 97	YP_003770670
13964		voltage-gated sodium channel [*Saccharopolyspora erythraea* NRRL 2338]	57 / 73	ZP_06564131
13966		NmrA family protein [*Catenulispora acidiphila* DSM 44928]	78 / 89	YP_003115375
13967		transcriptional regulatory protein [*Streptomyces* sp. Mg1]	85 / 93	ZP_04996370
13971		MarR family transcriptional regulator [*Amycolatopsis mediterranei* U32]	97 / 99	YP_003770672
13972	**ThiG**	thiamine biosynthesis protein ThiG [*Amycolatopsis mediterranei* U32]	97 / 99	YP_003770673
13974	**ThiS**	thiamine biosynthesis protein ThiS [*Amycolatopsis mediterranei* U32]	88 / 93	YP_003770674
13975	**ThiO**	glycine oxidase [*Amycolatopsis mediterranei* U32]	89 / 93	YP_003770675
13976	**ThiE**	thiamine-phosphate pyrophosphorylase [*Amycolatopsis mediterranei* U32]	89 / 93	YP_003770676
13829		ABC transport system ATP-binding protein [*Amycolatopsis mediterranei* U32]	91 / 94	YP_003770609
13830		L- permease component of ABC-type molybdate transport system [*Amycolatopsis mediterranei* U32]	65 / 98	YP_003770610
13831		eriplasmic substrate-binding component of ABC-type molybdate transport system [*Amycolatopsis mediterranei* U32]	84 / 89	YP_003770611
13832		molybdate ABC molybdate transporter substrate-binding protein [*Streptomyces* sp. AA4]	98 / 99	ZP_07282910
13835		MarR family transcriptional regulator [*Amycolatopsis mediterranei* U32]	89 / 92	YP_003770613
13837	**MoaA**	molybdenum cofactor biosynthesis protein A [*Amycolatopsis mediterranei* U32]	93 / 98	YP_003770614
13839	**MoaD**	ThiS/MoaD family protein [*Amycolatopsis mediterranei* U32]	82 / 90	YP_003770615
13840		secreted protein [*Amycolatopsis mediterranei* U32]	86 / 90	YP_003770616
13842	**MogA-MoaE**	molybdenum cofactor biosynthesis protein B [*Amycolatopsis mediterranei* U32]	92 / 97	YP_003770618
13843	**MoaC**	molybdenum cofactor biosynthesis protein C [*Amycolatopsis mediterranei* U32]	96 / 97	YP_003770619

Supplementary Table 1 | BLAST analysis of the genes near *thiS, moaD, cysO, moaD2,* and *moeZ* in *A. orientalis* genome (continued, 2/3)

gene number	putative functions	protein homologue and origin	identity / similarity (%)	protein accession number
13846		prevent-host-death family protein [*Frankia* symbiont of Datisca glomerata]	48 / 70	ZP_06476571
13848		PilT protein domain protein [*Frankia* symbiont of Datisca glomerata]	50 / 68	ZP_06473004
13851		malic enzyme [*Amycolatopsis mediterranei* U32]	97 / 99	YP_003770620
13854		hypothetical protein AMED_8524 [*Amycolatopsis mediterranei* U32]	88 / 99	YP_003770621
13856		hypothetical protein AMED_8525 [*Amycolatopsis mediterranei* U32]	64 / 97	YP_003770622
13859		hypothetical protein AMED_8526 [*Amycolatopsis mediterranei* U32]	65 / 78	YP_003770623
06453		transcriptional regulatory protein [*Streptomyces* sp. AA4]	92 / 96	ZP_07282073
06454		L-aminopeptidase/D-esterase DmpA [*Amycolatopsis mediterranei* U32]	94 / 96	YP_003769825
06456		No significant similarity found		
06457	**Mec$^+$**	Mov34/MPN/PAD-1 family protein [*Amycolatopsis mediterranei* U32]	65 / 98	YP_003769823
06461	**CysO**	ThiS/MoaD family protein [*Amycolatopsis mediterranei* U32]	97 / 100	YP_003769822
06462	**CysM**	cysteine synthase [*Amycolatopsis mediterranei* U32]	98 / 99	YP_003769821
06464		aspartate racemase [*Amycolatopsis mediterranei* U32]	89 / 93	YP_003769819
06465		hypothetical protein RHA1_ro01207 [*Rhodococcus jostii* RHA1]	50 / 64	YP_701191
06467		hypothetical protein AMED_7707 [*Amycolatopsis mediterranei* U32]	89 / 95	YP_003769817
10086		enoyl-CoA hydratase [*Amycolatopsis mediterranei* U32]	88 / 93	YP_003764230
10087		No significant similarity found		
10090		cold shock protein CspA [*Amycolatopsis mediterranei* U32]	99 / 99	YP_003767860
10092		hypothetical protein AMED_2018 [*Amycolatopsis mediterranei* U32]	85 / 93	YP_003764228
10093		RNA polymerase ECF-subfamily sigma factor [*Amycolatopsis mediterranei* U32]	91 / 93	YP_003764229
10094		hypothetical protein AMED_2017 [*Amycolatopsis mediterranei* U32]	82 / 90	YP_003764227
10095		XRE family transcriptional regulator [*Amycolatopsis mediterranei* U32]	94 / 100	YP_003764226
10096		NADPH:quinone reductase and related Zn-dependent oxidoreductase [*Amycolatopsis mediterranei* U32]	84 / 89	YP_003764225
10099		hypothetical protein AMED_2012 [*Amycolatopsis mediterranei* U32]	88 / 96	YP_003764222
10100		hypothetical protein AMED_2011 [*Amycolatopsis mediterranei* U32]	90 / 95	YP_003764221
10102	**MoaD2**	ThiS/MoaD family protein [*Amycolatopsis mediterranei* U32]	91 / 94	YP_003764220

Supplementary Table 1 | BLAST analysis of the genes near *thiS, moaD, cysO, moaD2,* and *moeZ* in *A. orientalis* genome (continued, 3/3)

gene number	putative functions	protein homologue and origin	identity / similarity (%)	protein accession number
10106		MerR family transcriptional regulator [*Amycolatopsis mediterranei* U32]	92 / 98	YP_003764218
10107		intradiol ring-cleavage dioxygenase [*Geodermatophilus obscurus* DSM 43160]	45 / 52	YP_003408862
10108		major facilitator transporter [*Amycolatopsis mediterranei* U32]	85 / 90	YP_003764219
10110		oxidoreductase [*Streptomyces bingchenggensis* BCW-1]	73 / 80	ADI09540
10111		No significant similarity found		
10113		hypothetical protein Mvan_5028 [*Mycobacterium vanbaalenii* PYR-1]	40 / 54	YP_955806
10114		conserved hypothetical protein [*Streptomyces pristinaespiralis* ATCC 25486]	25 / 38	ZP_06913769
10116		hypothetical protein AMED_2005 [*Amycolatopsis mediterranei* U32]	81 / 85	YP_003764215
10117		hypothetical protein AMED_2004 [*Amycolatopsis mediterranei* U32]	86 / 92	YP_003764214
10118		No significant similarity found		
02099		nitrite reductase (NAD(P)H) large subunit [Amycolatopsis mediterranei U32]	96 / 98	YP_003763341
02101		MFS transporter nitrate/nitrite transporter [Amycolatopsis mediterranei U32]	95 / 98	YP_003763340
02102		electron transfer subunit of assimilatory nitrate reductase [Amycolatopsis mediterranei U32]	95 / 97	YP_003763339
02104		nitrate reductase catalytic subunit [Amycolatopsis mediterranei U32]	93 / 96	YP_003763338
02106		No significant similarity found		
02107		hypothetical protein AMED_1120 [Amycolatopsis mediterranei U32]	88 / 92	YP_003763337
02108		No significant similarity found		
02110	MoeZ	molybdopterin biosynthesis-like protein MoeZ [Amycolatopsis mediterranei U32]	99 / 99	YP_003763336
02112		hypothetical protein AMED_1118 [Amycolatopsis mediterranei U32]	84 / 91	YP_003763335
02114		hydrolase [Amycolatopsis mediterranei U32]	92 / 94	YP_003763334
02116		TetR family transcriptional regulator [Amycolatopsis mediterranei U32]	99 / 99	YP_003763333
02117		alpha/beta hydrolase [Amycolatopsis mediterranei U32]	91 / 94	YP_003763332
02118		hypothetical protein AMED_1114 [Amycolatopsis mediterranei U32]]	96 / 99	YP_003763331
02119		TetR family transcriptional regulator [Amycolatopsis mediterranei U32]	94 / 97	YP_003763330
02120		flavin binding monooxygenase [Amycolatopsis mediterranei U32]	87 / 93	YP_003763329
02121		No significant similarity found		

ThiS, MoaD, CysO, and MoeZ homologues are shown in red. The other putative thiamin biosynthetic proteins, molybdopterin biosynthetic proteins, and cysteine biosynthetic proteins are shown in blue (See also Extended Data Fig. 2).

S3. Supplementary Discussion

Analysis of Crystal Structures of BexX-G6P and BexX/AoCysO Complex

2.1. Crystal Structure of BexX. The final model of BexX-G6P contains residues 11–252 from chain A and residues 12–253 from chain B. BexX adopts a $(\beta\alpha)_8$ barrel fold with a few minor additions (**Extended Data Fig. 1a, b**). A pair of antiparallel β-strands (βA and βB) precedes the initial strand β1 and forms a cap at the *N*-terminal face of the barrel. In addition, two short helices, $3_{10}1$ and α8', are inserted between α2 and β3, and α8 and β8, respectively. The two BexX monomers in the asymmetric unit form a homodimer (**Extended Data Fig. 1c**), which is consistent with gel filtration analysis (data not shown). The dimer interface is formed using the *C*-terminus (residues 246–253), the α8–α8' loop, α8, and α7 of each monomer. A comparison of chain A and chain B showed an rmsd of 0.296 Å for 174 Cα carbon atoms; however helix α2 in chain B partially unwinds and is five residues shorter than α2 in chain A.

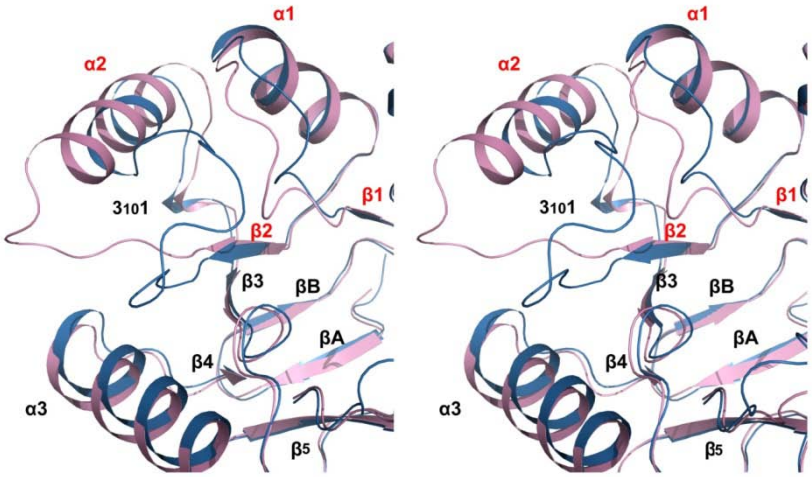
2.2. G6P Binding Site of BexX. The active site of BexX is located in the center of the barrel and utilizes loops β1-α1, β3–α3, β4–α4, β6–α6 and β7–α7 of the $(\beta\alpha)_8$-barrel (**Fig. 1c**). G6P is covalently attached to Lys110 to form a ketone intermediate. Lys110 Nζ forms two hydrogen bond with Glu31 Oε2 and Asp112 Oδ2. G6P atom O2 hydrogen bonds with Glu31 Oε1, O3 and O4 hydrogen bond to the side chain of Arg114, O4 hydrogen bonds to Glu194 Oε1, and O5 hydrogen bonds to Asn217 Nδ2 and water molecule W455. The phosphate group of G6P forms hydrogen bonds with Ala169, Gly196, Asn217, Thr218, and three water molecules W402, W404 and W403, which in turn hydrogen bond with phosphate oxygen atoms O2P, O3P and O2P, respectively.

2.3. Structure of the BexX/AoCysO Complex. The final model of the complex contains residues 11–254 from BexX and residues 2–90 from AoCysO. *A. orientalis* CysO (AoCysO) is structural homologous to CysO from *M. tuberculosis*[20] and has a ubiquitin-like fold. The compact monomer consists of a four-stranded β-sheet with β2↓β1↑β4↑β3↓topology flanked on one side by helices α1 and α2. Two short helices $3_{10}1$ and $3_{10}2$ are inserted between β1 and β2, β3 and β4, respectively (**Extended Data Fig. 1d, e**). The *C*-terminal tail (AVAGG) extends outward from the core of the structure and inserts into the active site of BexX (**Fig. 3a**). A homodimer of the complex forms using twofold crystallographic symmetry.

2.4. BexX/AoCysO Interface. The BexX/AoCysO interface involves 26 residues from BexX and 19 residues from AoCysO as calculated by PDBsum (http://www.ebi.ac.uk/pdbsum/). These residues form two main areas of interactions. The first interaction area is formed from surface residues of each chain and involves mainly hydrophobic residues on α2, α3, and the β2–α2 loop of BexX and mainly hydrophobic residues on the β1–$3_{10}1$ loop, $3_{10}1$, α2, β3, and $3_{10}2$ of AoCysO (**Extended Data Fig. 5b**). Only three hydrogen bonds are observed in this area (**Extended Data Fig. 6b**): one from Asn59 in the β2–α2 loop of BexX to Thr9 in the $3_{10}1$ of AoCysO, one from the side chain of Arg95 to the main chain of AoCysO Asp62, and one from the side chain of Asp68 to the side chain of Arg45 of AoCysO. The second interaction area forms between the *C*-terminal tail of AoCysO and the surrounding residues from BexX (**Extended Data Fig. 6c**). Nine hydrogen bonds form between the AVAGG *C*-terminal tail of AoCysO and eight residues from BexX (**Extended Data Fig. 6d**). Six of these involve main chain atoms from both BexX and AoCysO, while the seventh hydrogen bond forms between the Ser63 side chain of BexX and the

backbone carbonyl oxygen atom of Ala88 of AoCysO. The *C*-terminal carboxylate, which would be thiocarboxylate *in vivo*, is positioned to form the last two hydrogen bonds with the side chains of Ser85 and Lys110 with distances of 3.1 and 2.9 Å, respectively.

2.5. Comparison of BexX/AoCysO to BexX. A comparison of the BexX/AoCysO and the BexX-G6P structures showed an rmsd for 213 Cα carbon atoms of 0.58 Å; however, two areas show conformational changes that result from complex formation. The *N*-terminal end of α2 and the loop between β2 and α2 of BexX shift approximately 12 Å allowing the *C*-terminal tail of AoCysO to insert into the active site of BexX. A second conformational change occurs in the β1-α1 loop. Glu31, which is located in this loop and forms hydrogen bond with Lys110 Nζ in BexX-G6P is displaced by the *C*-terminal tail.



Conformational change associated with complex formation. The stereo diagram is colored pink for BexX-G6P and blue for BexX from the complex with AoCysO. The largest changes occur in α2, β2, α1 and β1, which are labeled in red.