Supporting Information

Rerouting the Pathway for the Biosynthesis of the Side Ring System of Nosiheptide: The Roles of NosI, NosJ, and NosK

Edward D. Badding,[⊥] Tyler L. Grove,[⊥,†] Lauren Gadsby,[ξ] Joseph W. LaMattina,[⊥] Amie K. Boal,[⊥, ξ,*] and Squire J. Booker[⊥,ξ,#,*]

[⊥]The Department of Chemistry, [ξ]The Department of Biochemistry and Molecular Biology, and [#]The Howard Hughes Medical Institute, The Pennsylvania State University, University Park, Pennsylvania, 16802, USA

The codon-optimized sequence of *nosI*, with the *Nde*I and *Eco*RI restriction sites indicated in bold type.

5'-CACTATAGGGCGAATTGAAGGAAGGCCGTCAAGGCCGCAT**CATATG**GGTGATATGGG
TCGTCCGGCATTTCAGCGTTTTCTGACACCGCGTCATCTGCCTGCAGGTCGTGCCGGTGC
AGTTACCGGTGTTCGTTGGGGTGGTGATTTTGCAGCATGGGATGATCTGCTGACCGCAGG
TCGCGATCTGGCAGCACAGGTTCGTCCGGGTGGTGCCTATGCAATTGATCCGACAGCAGG
TCTGCCTGCCCTGGCAGCCCTGTTTGCAGTTGCAACCGTTCCGGATACCGTTCTGCTGTG
GGCAAGTCCGCGTACCCTGGGTGTTACCGGTCGTGAAATTGCTCCGGCACTGCATGCCCT
GCCGGATGATGGTAGCGTTCCGCTGGCAGCGCAAGAACGTCCGCTGTGGGGTGTTTGTAC
CAGCGGTAGCAGTGGTGCACCGAAAGTTGCAGTTGGTCCGGCAGATGAATGGGAGCAGA
TTGCCCTGCATGCCGAAGCAGCAATGTATGCAGATGCATTTCCGGCAGGTCCGCCTGAAG
CACTGGCAACCTGTCTGCCGCTGGGTTTTAGCGCAGCCTTTTTTATGTGTGTTCTGCCAG
CACTGTATCTGAAACGTGATCTGGTTGTTCATCCGCCTCATGATTGGAGTCCGCTGTATG
ATCTGGCACGTGATCGTCGTGTTCTGGCACTGGGTGTTCCAGCTCTGGCAGCCGCAGCAT
GTCTGAGCGCACCGGCAGCAACCGATCTGGGTAGCGTTGCACTGTTTCTGGGTGGTGGTC
ATCTGAGTGCACCGCGTGTTGAACTGATTCGTCGTCATTTTACCGGTGCAGCAGTTAGCA
ATCTGTATGGCACCGCAGAAACCGGTGCAATTGCCCTGGATCACGATCCGGGTCATAATC
GTCATGTTGGTCGTCCGATTCCGGGTAAAAGCGTTTGGCTGACCGGCACCGATGAACGTG
GTATTGGCACCGTTGCCGTTGCAGGTCCGGGTTGTTGTCGTCGTACCTGGCGTCCGGGTA
GCCCTCCGAGTGCCCCTGCAGATCATGTGACCGGTACAGATTATGGTCGTTTTGATGCAG
ATGGTAATCTGTGTCTGGAAGGTCGTCTGGATGGTGCAGAAAAACTGGCAGGCGTTCTG
GTGCGTCCGCGTGAAATCGAACGTCATGTGCTGGCCCTGGATGGCGTTAGTGATGTTCGT
GTTACCGTTGAAACCGCACCGACCGGTCTGGAATTTCTGGCAGCGACCGTTGTTGGTAGC
GTGGATGCAGATACCGTGCGTGCACATTGTGCGGCACTGCCGGAACAGCATCGTCCGAGC
CGTATTAGCTGTGCAAGCGAACAAGAAGCAGCCACCGTTTATAGCGCACATGGTAAACTG
TAA**GAATTC**CTGGGCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAG-3'

The codon-optimized sequence of the *nosJ* gene. The *Nde*I and *Eco*RI restriction sites are indicated in bold type.

5’-CGAATTGGCGGAAGGCCGTCAAGGCCACGTGTCTTGTCCAGAGCTC**CATATG**ACCAGC
CAGCGTACCACACCGCGTACACCGGATGGTGTTCCGGATCTGCAAGAAGAACTGGCAGGT
CTGCTGCAAGAGGATGATCCGCGTCGTCGTCTGGATAGCCTGGAAACCGTTGTTGTTCTG
AGCTATTTTGCACGTCAGGCACCGGGTCGTACCCTGCCGGAACTGCCGGATGCTCCGCGT
ACCATTGAAGGTTGGGTTACCTGGGCAGATCAGCGTAGCAGCGCAAGCTAA**GAATTC**GG
TACCTGGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGC-3’

The codon-optimized sequence of *nosK*, with the *Nde*I and *Eco*RI restriction sites indicated in bold type.

5’-
CGAATTGGCGGAAGGCCGTCAAGGCCACGTGTCTTGTCCAGAGCTC**CATATG**GATGCAG
AAACCCCGATGGATACCGAAACACCGCGTGATACGGAAACGCCGATGCATACAGGTATGA
GCACCGGTCCGGAAACACCGACCGTTTATCTGGTTCATGGTCTGCTGGGCACCGGTCATG
GTCATTTTGCAGCACAGATTCGTGCATGGCATGGTCGTCTGCGTACCGTTCCGGTTGATC
TGCCTGGTCATGGCCGTTGTCGTCGTGATGCAGCCGAAGATTATTTTGATGATGCACTGC
GTTATCTGGTGGCAGTTCTGGAACGTTTTGGTCCGGGTCGTCTGATTGGTGCAAGCTATC
TGGGTGGTCCGCTGGCACATCGTTGTGCAGCAACCCGTCCGGATCTGGTTAGCAGCCTGG
TTCTGACCGGTTTTGCACCGGATGTTAGCCGTGATGCATTTCTGAGCCTGATTGCAGGTT
TTGAAGGTCTGGCAGCACAGCAGCCTGCACTGGCAGCAGAATATGAACAGCTGCATGGCA
CCCGTTGGAAACGTACCCTGGATGCAGTTACAGGTCATGTTGAACGTGATTTTGAACGTA
CCGCACTGGTTCGTGCAGCAGATGTTGCAGCACTGACCGTTCCGACCCTGGTGCTGAATG
GTAGCCTGAAAAGCGTGGAACGTGCAGCCGCAGAACAGGCACCGGGTTGGGGTGGTCGT
GTTCGTGGTCGCGTTGTTCCGGGTGCAGGTCATCTGGTTGGTCATGATCGTCCGCGTGAA
TTTAATGAAGCAGTTGAAGATTTTTGGCGCACCGCACATGATGCACCGGCAGGTCCGCGT
ACCACACAGAAAGGTGATACC**GAATTC**GGTACCTGGAGCACAAGACTGGCCTCATGGGC
CTTCCGCTCACTGC-3’

Figure S1. Purification of NosJ. Lanes 1 and 6, molecular mass markers (kDa); lane 2, crude extract; lane 3, pellet; lane 4, flow through from DE-52 column; lane 5, wash from DE-52 column; lane 6, pooled protein from DE-52 column; lane 7, post S-200 column.

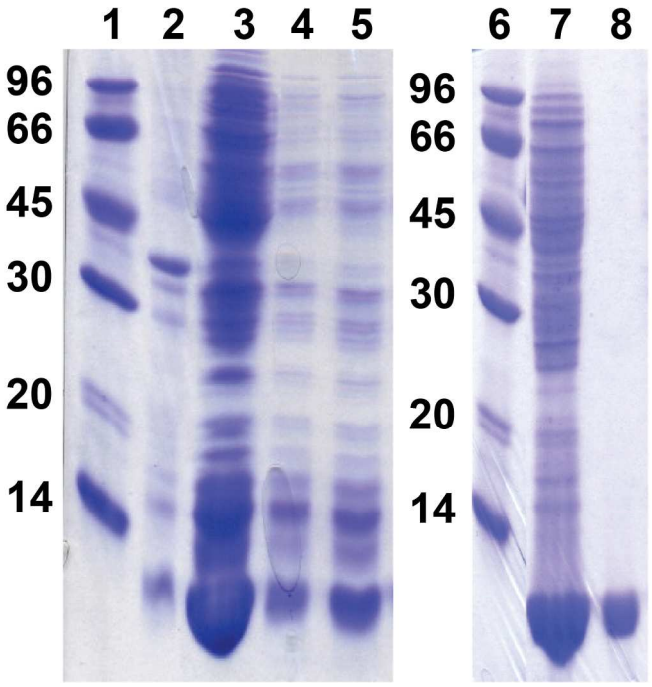

Figure S2. Purification of NosI. Lane 1, molecular mass markers (kDa); lane 2, pellet; lane 3, crude extract; lane 4, flow through from Talon $Co(2^+)$ column; lane 5, wash from Talon $Co(2^+)$ column; lane 6, second wash from Talon $Co(2^+)$ column; lanes 7 and 8, eluted protein at 2 different concentrations.

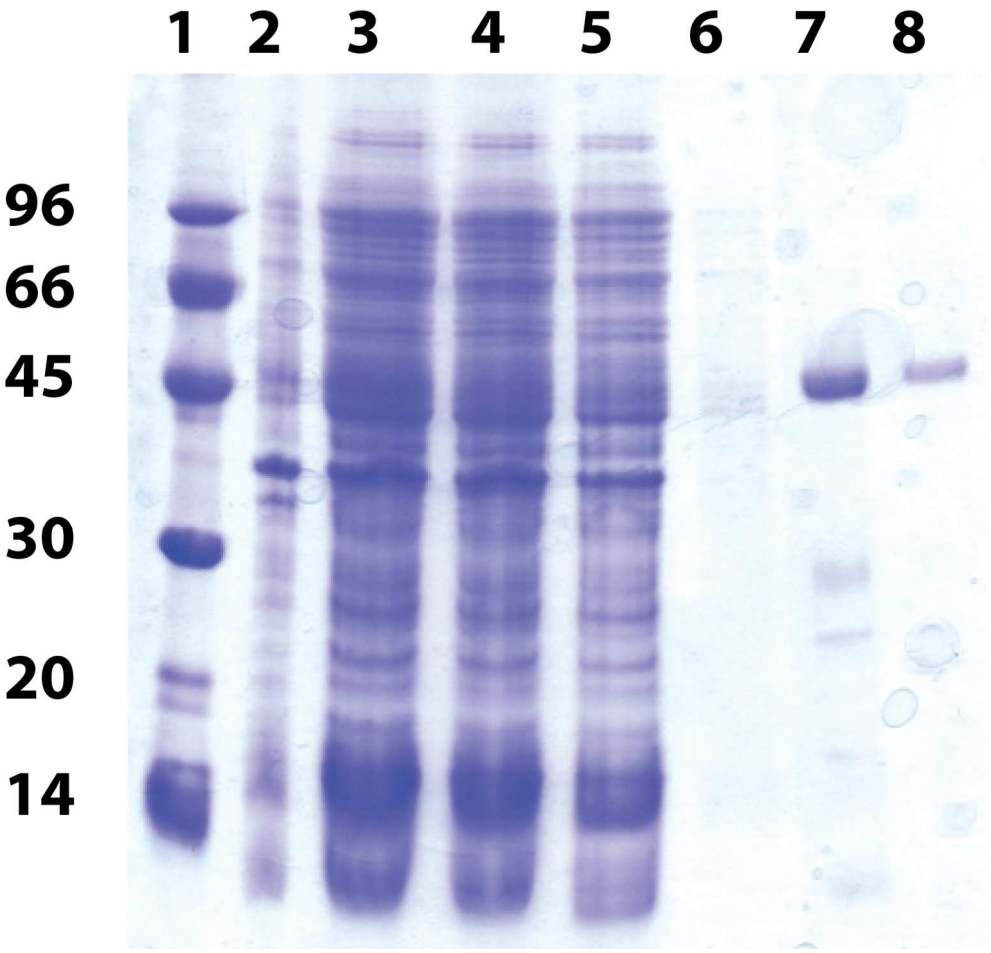

Figure S3. Overproduction and purification of NosK. Lane 1, molecular mass markers (kDa); lane 2, before IPTG induction; lane 3, after IPTG induction; lane 4, pellet; lane 5, crude extract; lane 6, flow through from Talon $Co(2^+)$ column; lane 7, wash from Talon $Co(2^+)$ column; lane 8, second wash from Talon $Co(2^+)$ column; lanes 9 and 10, eluted protein at 2 different concentrations.

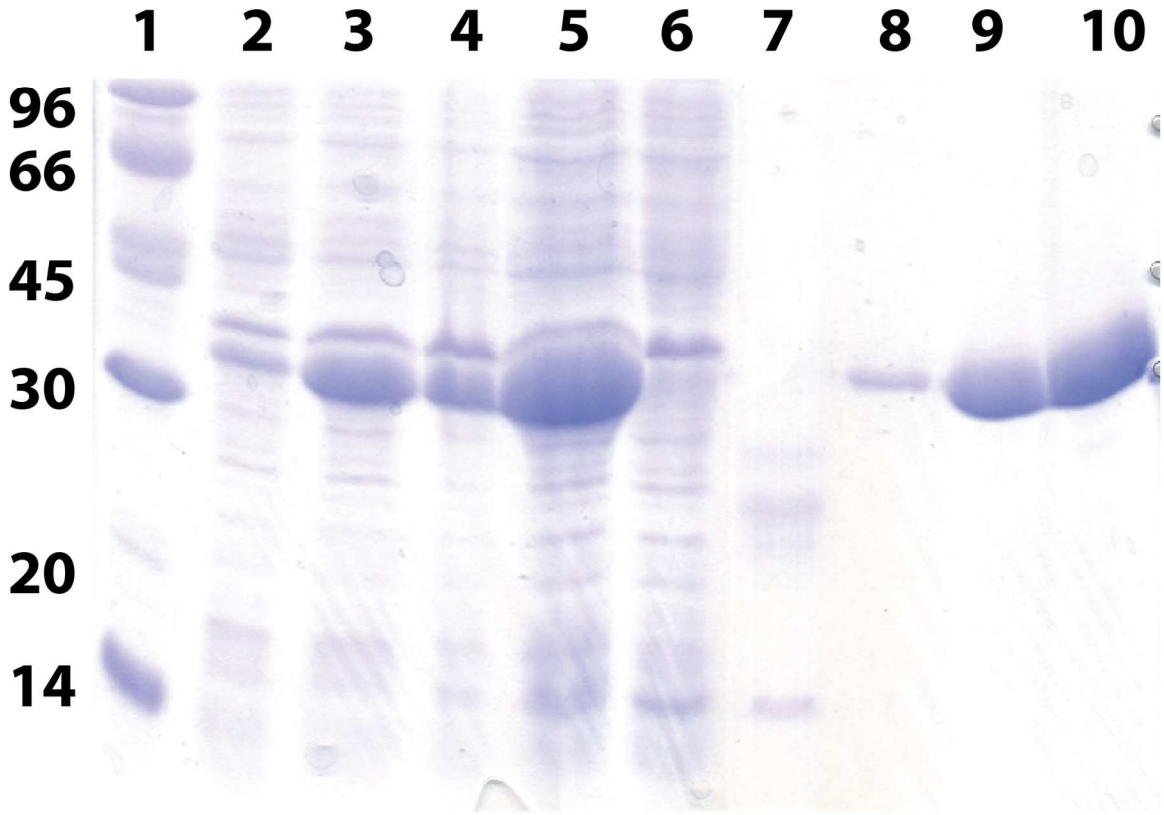

Figure S4. Overlay of the catalytic triad acidic residue loop in NosK and valacyclovirase (2OCG PDB accession code).

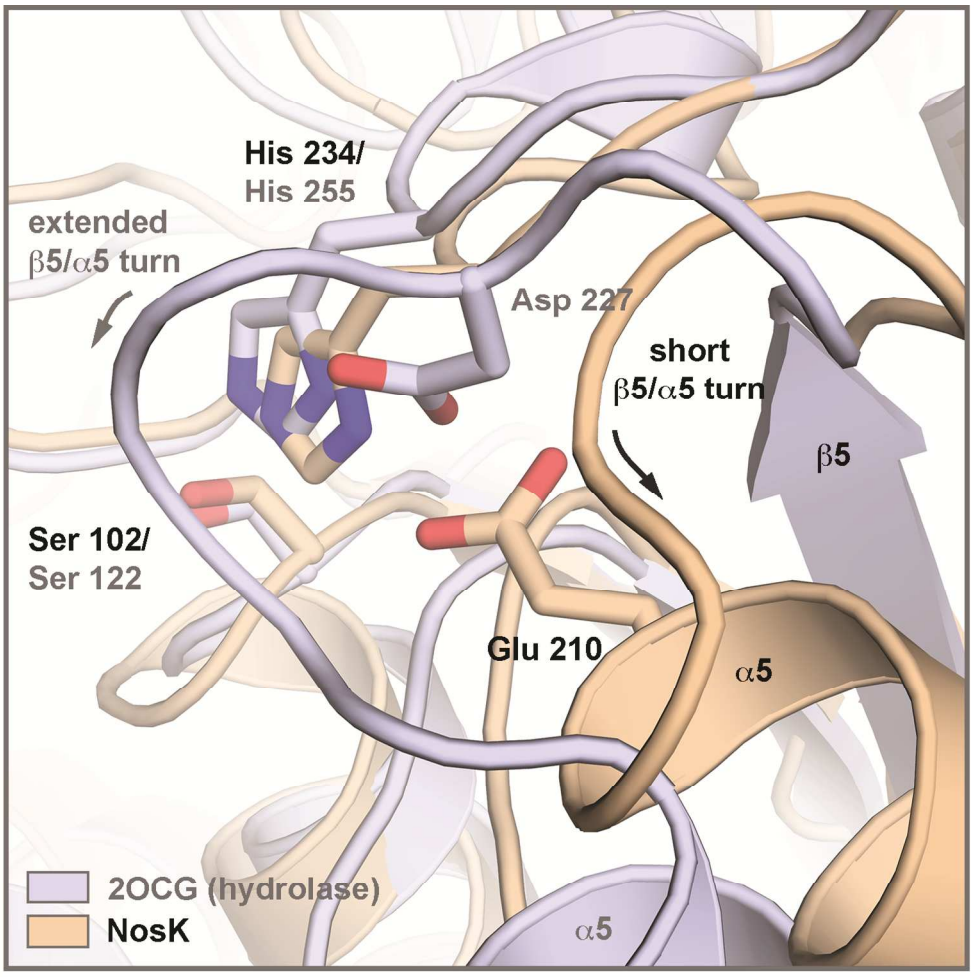

Figure S5. Electrostatic surface potential map for NosK. Surface contoured at + 5 k_BT (blue) and -5 k_BT (red). Corresponding cartoon representations for each view are shown in the bottom panels.

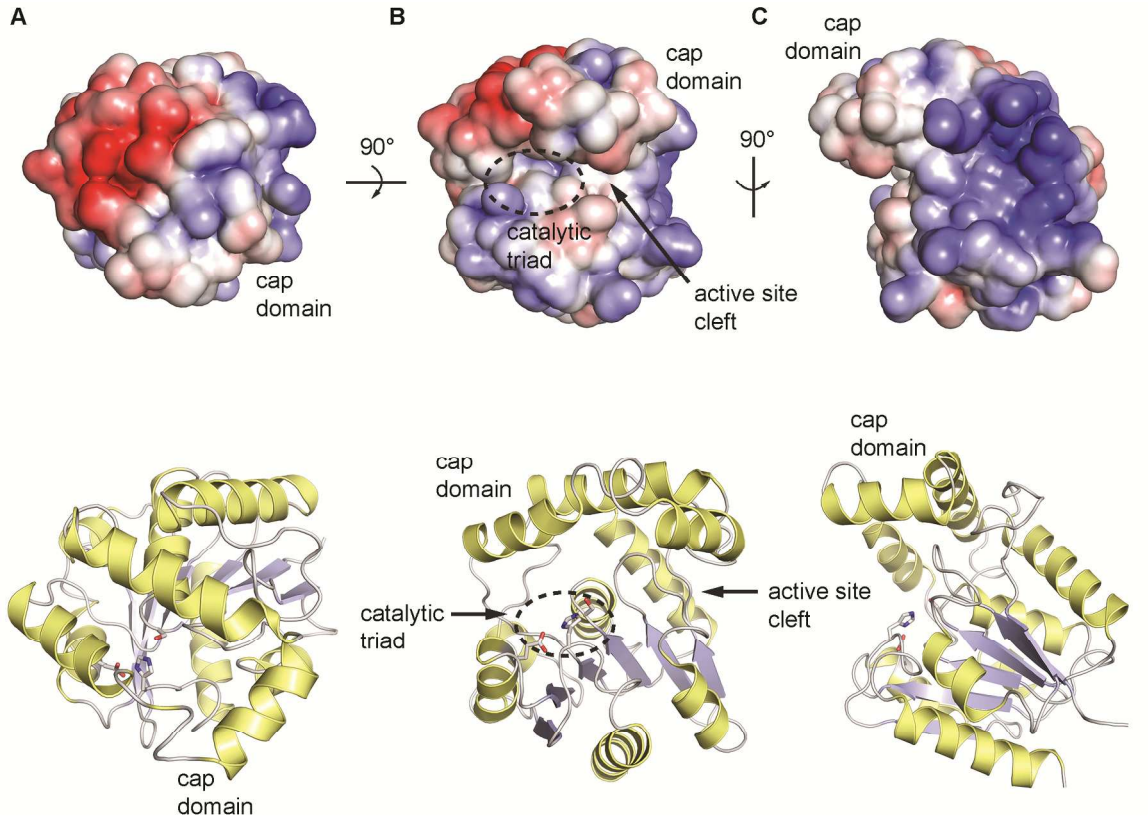

Table S1. Fragments from NosK trypsin digestion after treatment with NosJ, NosI, MIA, ATP, CoA, and holo ACP synthase. Highlighted in blue are the peptide fragments containing Ser102.

Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications
DTETPmHTGMSTGPETPTVYLVHGLLGTGHGHF AAQIR	8	1	1	C6FX50	M6(Oxidation)
DAFLSLIAGFEGLAAQQPALAAEYEQLHGTR	101	1	1	C6FX50	
AADVAALTVPTLVLnGSLK	41	1	1	C6FX50	N15(Deamidated)
DAFLSLIAGFEGLAAQqPALAAEYEQLHGTR	3	1	1	C6FX50	Q17(Deamidated)
DTETPMHTGMSTGPETPTVYLVHGLLGTGHGHF AAQIR	3	1	1	C6FX50	
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAQIR	11	1	1	C6FX50	M6(Oxidation); M10(Oxidation)
cAATRPDLVSSLVLTGFAPDVSR	61	1	1	C6FX50	C1(Carbamidomethyl)
AADVAALTVPTLVLNGSLK	47	1	1	C6FX50	
DAFLSLIAGFEGLAAQQPALAAEYEQLHGTRWK	2	1	1	C6FX50	
FGPGRLIGA**S**YLGGPLAHR	8	1	1	C6FX50	S10(Ser-MIA)
AADVAALTVPTLVLnGSLKSVER	1	1	1	C6FX50	N15(Deamidated)
EFNEAVEDFWR	82	1	1	C6FX50	
DFERTALVRAADVAALTVPTLVLNGSLK	2	1	1	C6FX50	
LIGA**S**YLGGPLAHR	41	1	1	C6FX50	S5(Ser-MIA)
DAAEDYFDDALR	24	1	1	C6FX50	
YLVAVLERFGPGRLIGA**S**YLGGPLAHR	5	1	1	C6FX50	S18(Ser-MIA)
TALVRAADVAALTVPTLVLNGSLK	2	1	1	C6FX50	
LIGA**S**YLGGPLAHRCAATRPDLVSSLVLTGFAPDV SR	5	1	1	C6FX50	
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAQIRAWHGRLR	1	1	1	C6FX50	M6(Oxidation); M10(Oxidation)
AADVAALTVPTLVLNGSLKSVER	2	1	1	C6FX50	
LIGA**S**YLGGPLAHRcAATRPDLVSSLVLTGFAPDV SR	3	1	1	C6FX50	C15(Carbamidomethyl)
AAAEQAPGWGGR	5	1	1	C6FX50	
DFERTALVRAADVAALTVPTLVLnGSLK	3	1	1	C6FX50	N24(Deamidated)
AADVAALTVPTLVLNGSLKSVERAAAEQAPGWG GR	3	1	1	C6FX50	
DTETPMHTGmsTGPETPTVYLVHGLLGTGHGHF AAqIR	1	1	1	C6FX50	M10(Oxidation); S11(Ser-MIA); Q36(Deamidated)
YLVAVLER	1	1	1	C6FX50	
TLDAVTGHVER	2	1	1	C6FX50	
TLDAVTGHVERDFER	1	1	1	C6FX50	

Table S2. Fragments from NosK trypsin digestion after treatment with NosJ, MIA, ATP, CoA, and holo ACP synthase (NosI omitted). Highlighted in blue are the peptide fragments containing Ser102.

Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications
DAFLSLIAGFEGLAAQQPALAAEYEQLHGTR	116	1	1	C6FX50	
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAQIR	8	1	1	C6FX50	M6(Oxidation); M10(Oxidation)
AADVAALTVPTLVLNGSLK	82	1	1	C6FX50	
DAFLSLIAGFEGLAAQQPALAAEYEQLHGTRWK	3	1	1	C6FX50	
AADVAALTVPTLVLnGSLK	14	1	1	C6FX50	N15(Deamidated)
cAATRPDLVSSLVLTGFAPDVSR	31	1	1	C6FX50	C1(Carbamidomethyl)
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAqIRAWHGR	2	1	1	C6FX50	M6(Oxidation);M10(Oxidation); Q36(Deamidated)
GRVVPGAGHLVGHDRPREFNEAVEDFWR	1	1	1	C6FX50	
LIGA**S**YLGGPLAHRcAATRPDLVSSLVLTGFAPDV SR	2	1	1	C6FX50	C15(Carbamidomethyl)
AADVAALTVPTLVLNGSLKSVER	3	1	1	C6FX50	
LIGA**S**YLGGPLAHRCAATRPDLVSSLVLTGFAPDV SR	7	1	1	C6FX50	
EFNEAVEDFWR	70	1	1	C6FX50	
AADVAALTVPTLVLNGSLKSVERAAAEQAPGWG GR	3	1	1	C6FX50	
DFERTALVRAADVAALTVPTLVLNGSLK	2	1	1	C6FX50	
DAAEDYFDDALR	20	1	1	C6FX50	
AADVAALTVPTLVLnGSLKSVER	1	1	1	C6FX50	N15(Deamidated)
DAFLSLIAGFEGLAAQQPALAAEYEQLHGTRWKR	3	1	1	C6FX50	
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAQIRAWHGRLR	3	1	1	C6FX50	M6(Oxidation); M10(Oxidation)
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAqIR	1	1	1	C6FX50	M6(Oxidation); M10(Oxidation); Q36(Deamidated)
TALVRAADVAALTVPTLVLNGSLK	2	1	1	C6FX50	
LIGA**S**YLGGPLAHR	1	1	1	C6FX50	
DTETPmHTGmSTGPETPTVYLVHGLLGTGHGHF AAqIRAWHGRLR	1	1	1	C6FX50	M6(Oxidation); M10(Oxidation); Q36(Deamidated)
YLVAVLER	3	1	1	C6FX50	
AAAEQAPGWGGR	5	1	1	C6FX50	
AADVAALTVPTLVLnGSLKSVERAAAEQAPGWG GR	1	1	1	C6FX50	N15(Deamidated)
DTETPmHTGMSTGPETPTVYLVHGLLGTGHGHF AAQIR	5	1	1	C6FX50	M6(Oxidation)
TLDAVTGHVERDFER	1	1	1	C6FX50	
TLDAVTGHVER	2	1	1	C6FX50	

Table S3. Crystallographic data collection and refinement statistics for NosK (PDB accession code 5V7O)

Data Collection	
Resolution (Å)	28.00-2.30 (2.38-2.30)
Space group	$P3_121$
Cell dimensions	
a, b, c (Å)	75.29, 75.29, 110.12
α, β, γ (°)	90, 90, 120
Redundancy	10.8 (11.1)
Completeness (%)	99.9 (100)
$I/\sigma I$	13.31 (2.62)
Wilson *B*-factor (Å^2)	33.96
R_{sym}	0.040 (0.252)
R_{pim}	0.057 (0.356)
$CC_{1/2}$	0.997 (0.936)
Refinement	
Resolution (Å)	28.00-2.30 (2.38-2.30)
No. reflections	16453
R_{work} / R_{free}	0.22/0.26
No. atoms	1858
Protein	1762
Ligand/ion	5
Water	230
rms deviations	
bond lengths (Å)	0.007
bond angles (°)	0.87
Ramachandran analysis	
Favored (%)	98
Allowed (%)	2.2
Outliers (%)	0
B-factor (Å^2)	
Protein	41.84
Ligands/ion	35.19
Water	43.34

Statistics for the highest-resolution shell are shown in parentheses.