# Supplemental Information

## Title: Nonribosomal Propeptide Precursor in Nocardicin A Biosynthesis Predicted from Adenylation Domain Specificity Dependent on the MbtH Family Protein NocI

Jeanne M. Davidsen, David M. Bartley[1] and Craig A. Townsend*

*Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218*

[1]Current address: Department of Chemistry, Adrian College, Adrian, MI 49221

* Corresponding author. Mailing address: Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218. Phone: (410) 516-7444. Fax: (410) 261-1233. E-mail: ctownsend@jhu.edu.

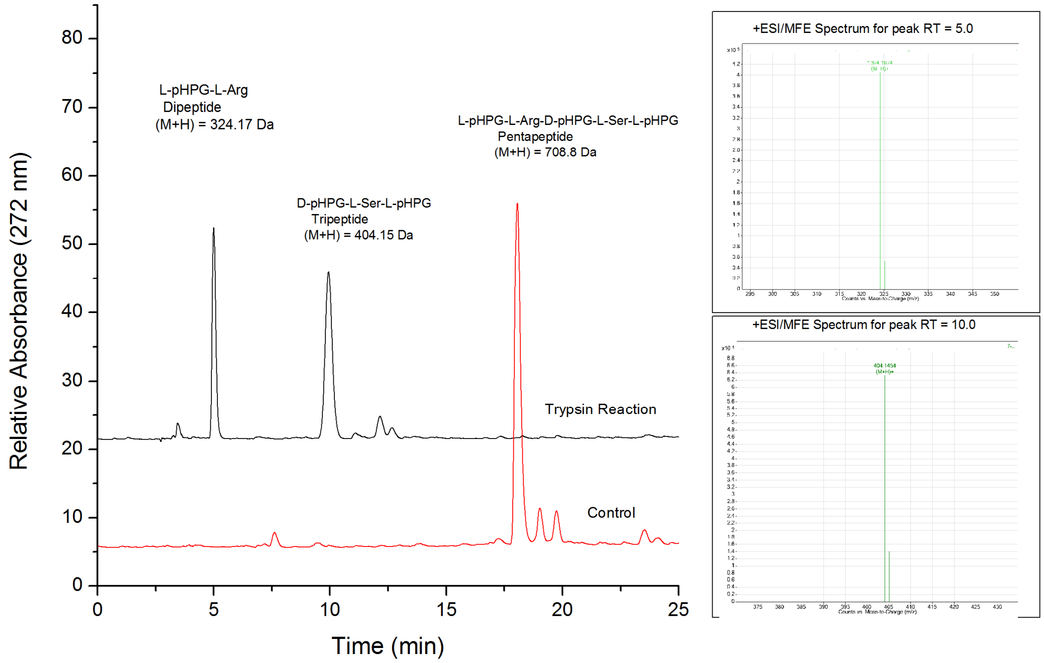
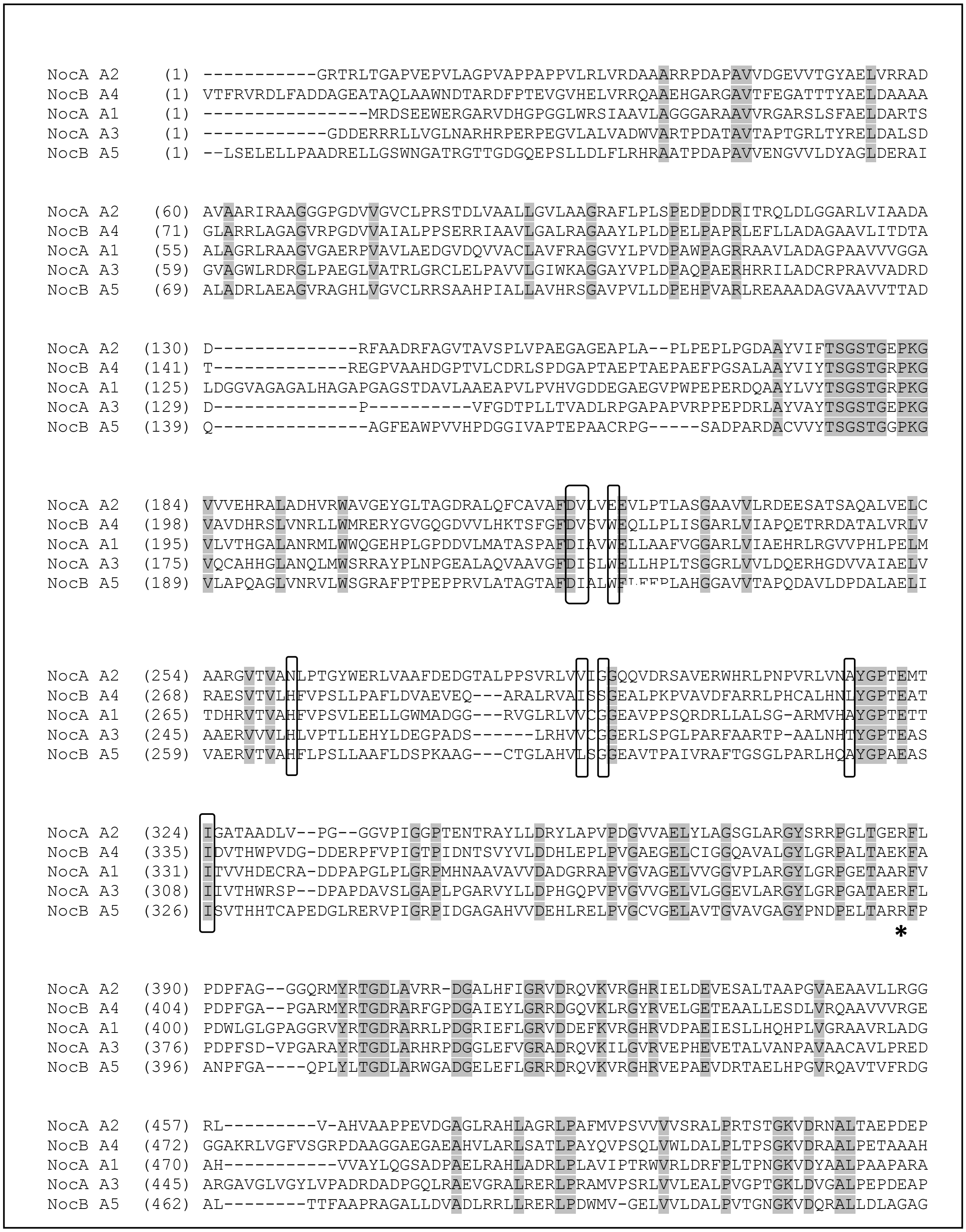


Figure S1. HPLC analysis of the reaction of pentapeptide L-pHPG-L-Arg-D-pHPG-L-Ser-L-pHPG with trypsin (black) versus a control (red). Mass determination by LC/MS TOF for the two major peaks of the trypsin reaction are shown on the right.

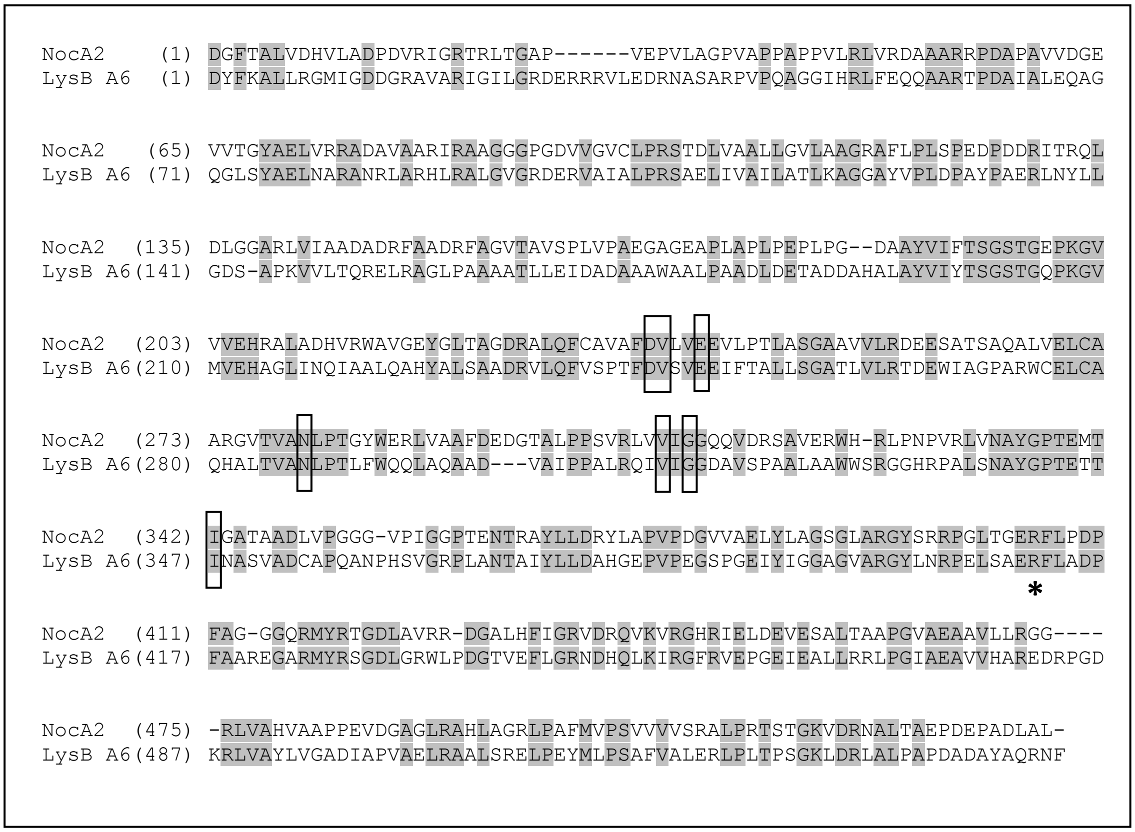


```
NocA A2      (1) -----------GRTRLTGAPVEPVLAGPVAPPAPPVLRLVRDAAARRPDAPAVVDGEVVTGYAELVRRAD
NocB A4      (1) VTFRVRDLFADDAGEATAQLAAWNDTARDFPTEVGVHELVRRQAAEHGARGAVTFEGATTTYAELDAAAA
NocA A1      (1) ----------------MRDSEEWERGARVDHGPGGLWRSIAAVLAGGGARAAVVRGARSLSFAELDARTS
NocA A3      (1) ------------GDDERRRLLVGLNARHRPERPEGVLALVADWVARTPDATAVTAPTGRLTYRELDALSD
NocB A5      (1) --LSELELLPAADRELLGSWNGATRGTTGDGQEPSLLDLFLRHRAATPDAPAVVENGVVLDYAGLDERAI

NocA A2     (60) AVAARIRAAGGGPGDVVGVCLPRSTDLVAALLGVLAAGRAFLPLSPEDPDDRITRQLDLGGARLVIAADA
NocB A4     (71) GLARRLAGAGVRPGDVVAIALPPSERRIAAVLGALRAGAAYLPLDPELPAPRLEFLLADAGAAVLITDTA
NocA A1     (55) ALAGRLRAAGVGAERPVAVLAEDGVDQVVACLAVFRAGGVYLPVDPAWPAGRRAAVLADAGPAAVVVGGA
NocA A3     (59) GVAGWLRDRGLPAEGLVATRLGRCLELPAVVLGIWKAGGAYVPLDPAQPAERHRRILADCRPRAVVADRD
NocB A5     (69) ALADRLAEAGVRAGHLVGVCLRRSAAHPIALLAVHRSGAVPVLLDPEHPVARLREAAADAGVAAVVTTAD

NocA A2    (130) D-------------RFAADRFAGVTAVSPLVPAEGAGEAPLA--PLPEPLPGDAAYVIFTSGSTGEPKG
NocB A4    (141) T-------------REGPVAAHDGPTVLCDRLSPDGAPTAEPTAEPAEFPGSALAAYVIYTSGSTGRPKG
NocA A1    (125) LDGGVAGAGALHAGAPGAGSTDAVLAAEAPVLPVHVGDDEGAEGVPWPEPERDQAAYLVYTSGSTGRPKG
NocA A3    (129) D--------------P----------VFGDTPLLTVADLRPGAPAPVRPPEPDRLAYVAYTSGSTGEPKG
NocB A5    (139) Q---------------AGFEAWPVVHPDGGIVAPTEPAACRPG-----SADPARDACVVYTSGSTGGPKG

NocA A2    (184) VVVEHRALADHVRWAVGEYGLTAGDRALQFCAVAFDVLVEEVLPTLASGAAVVLRDEESATSAQALVELC
NocB A4    (198) VAVDHRSLVNRLLWMRERYGVGQGDVVLHKTSFGFDVSVWEQLLPLISGARLVIAPQETRRDATALVRLV
NocA A1    (195) VLVTHGALANRMLWWQGEHPLGPDDVLMATASPAFDIAVWELLAAFVGGARLVIAEHRLRGVVPHLPELM
NocA A3    (175) VQCAHHGLANQLMWSRRAYPLNPGEALAQVAAVGFDISLWELLHPLTSGGRLVVLDQERHGDVVAIAELV
NocB A5    (189) VLAPQAGLVNRVLWSGRAFPTPEPPRVLATAGTAFDIALWELEEPLAHGGAVVTAPQDAVLDPDALAELI

NocA A2    (254) AARGVTVANLPTGYWERLVAAFDEDGTALPPSVRLVVIGGQQVDRSAVERWHRLPNPVRLVNAYGPTEMT
NocB A4    (268) RAESVTVLHFVPSLLPAFLDVAEVEQ---ARALRVAISSGEALPKPVAVDFARRLPHCALHNLYGPTEAT
NocA A1    (265) TDHRVTVAHFVPSVLEELLGWMADGG---RVGLRLVVCGGEAVPPSQRDRLLALSG-ARMVHAYGPTETT
NocA A3    (245) AAERVVVLHLVPTLLEHYLDEGPADS------LRHVVCGGERLSPGLPARFAARTP-AALNHTYGPTEAS
NocB A5    (259) VAERVTVAHFLPSLLAAFLDSPKAAG---CTGLAHVLSGGEAVTPAIVRAFTGSGLPARLHQAYGPAEAS

NocA A2    (324) IGATAADLV--PG--GGVPIGGPTENTRAYLLDRYLAPVPDGVVAELYLAGSGLARGYSRRPGLTGERFL
NocB A4    (335) IDVTHWPVDG-DDERPFVPIGTPIDNTSVYVLDDHLEPLPVGAEGELCIGGQAVALGYLGRPALTAEKFA
NocA A1    (331) ITVVHDECRA-DDPAPGLPLGRPMHNAAVAVVDADGRRAPVGVAGELVVGGVPLARGYLGRPGETAARFV
NocA A3    (308) IIVTHWRSP--DPAPDAVSLGAPLPGARVYLLDPHGQPVPVGVVGELVLGGEVLARGYLGRPGATAERFL
NocB A5    (326) ISVTHHTCAPEDGLRERVPIGRPIDGAGAHVVDEHLRELPVGCVGELAVTGVAVGAGYPNDPELTARRFP
                                                                                     *

NocA A2    (390) PDPFAG--GGQRMYRTGDLAVRR-DGALHFIGRVDRQVKVRGHRIELDEVESALTAAPGVAEAAVLLRGG
NocB A4    (404) PDPFGA--PGARMYRTGDRARFGPDGAIEYLGRRDGQVKLRGYRVELGETEAALLESDLVRQAAVVVRGE
NocA A1    (400) PDWLGLGPAGGRVYRTGDRARRLPDGRIEFLGRVDDEFKVRGHRVDPAEIESLLHQHPLVGRAAVRLADG
NocA A3    (376) PDPFSD-VPGARAYRTGDLARHRPDGGLEFVGRADRQVKILGVRVEPHEVETALVANPAVAACAVLPRED
NocB A5    (396) ANPFGA----QPLYLTGDLARWGADGELEFLGRRDRQVKVRGHRVEPAEVDRTAELHPGVRQAVTVFRDG

NocA A2    (457) RL---------V-AHVAAPPEVDGAGLRAHLAGRLPAFMVPSVVVVSRALPRTSTGKVDRNALTAEPDEP
NocB A4    (472) GGAKRLVGFVSGRPDAAGGAEGAEAHVLARLSATLPAYQVPSQLVWLDALPLTPSGKVDRAALPETAAAH
NocA A1    (470) AH-----------VVAYLQGSADPAELRAHLADRLPLAVIPTRWVRLDRFPLTPNGKVDYAALPAAPARA
NocA A3    (445) ARGAVGLVGYLVPADRDADPGQLRAEVGRALRERLPRAMVPSRLVVLEALPVGPTGKLDVGALPEPDEAP
NocB A5    (462) AL--------TTFAAPRAGALLDVADLRRLLRERLPDWMV-GELVVLDALPVTGNGKVDQRALLDLAGAG
```

Figure S2. Alignment of the amino acid sequences of the A-domains of NocA and NocB. Identical residues are indicated by a gray shadow. Amino acid substrate binding site residues predicted by PheA crystal structure are boxed. The amino acid site corresponding to CloH L383 is marked by an asterisk.

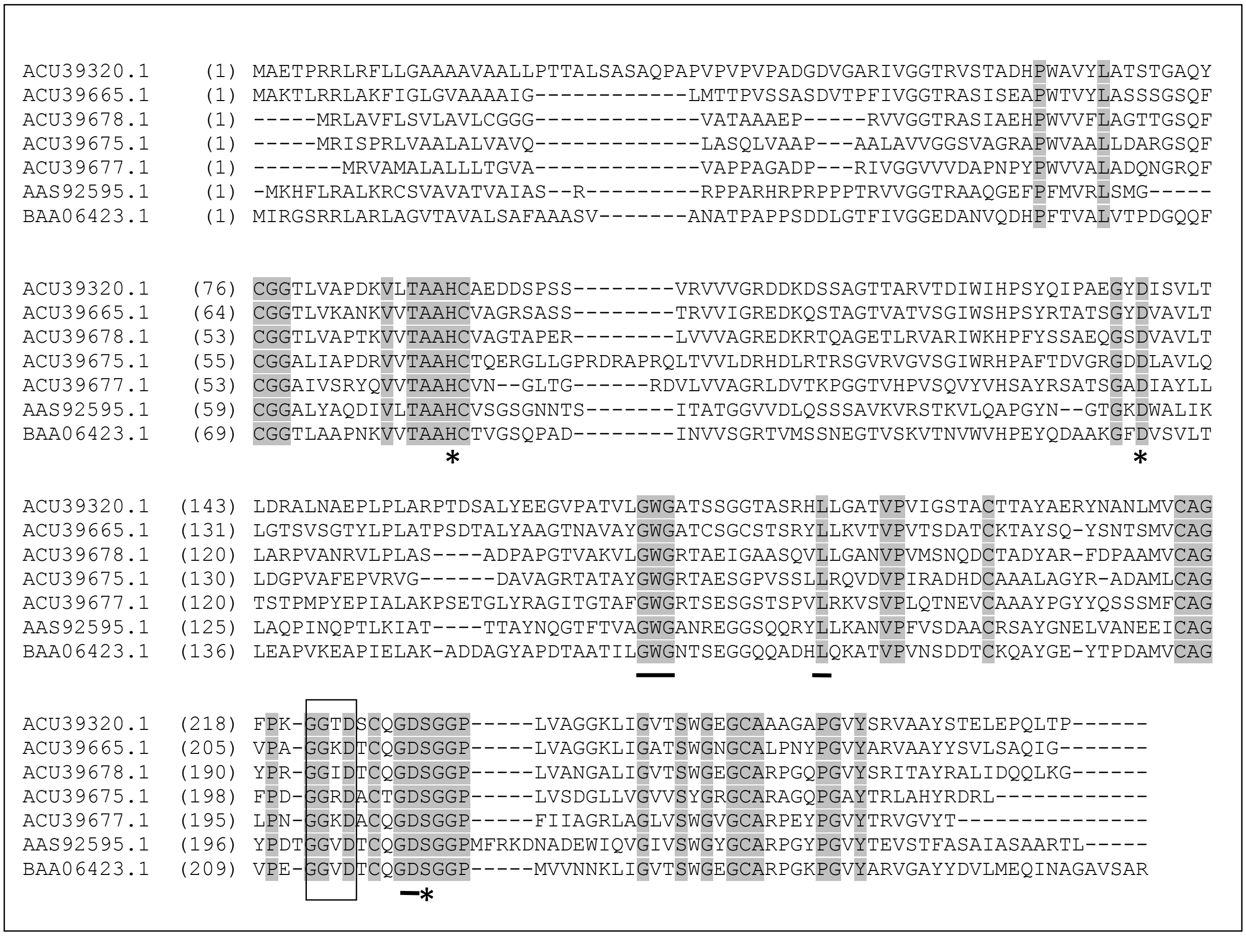
```
NocA A1  (369) PVGVAGELVVGGVPLARGYLGRPGETAARFVPDWLGLGPAGGRVYRTGDRARRLPDGRIEFLGRVDDE
NocA A2  (359) PDGVVAELYLAGSGLARGYSRRPGLTGERFLPDPFAG--GGQRMYRTGDLA-VRRDGALHFIGRVDRQ
NocB A4  (373) PVGAEGELCIGGQAVALGYLGRPALTAEKFAPDPFGA--PGARMYRTGDRARFGPDGAIEYLGRRDGQ
CmnA A2  (356) PPGSTGELCVGGLPVARGYLGRPALTAERFVPDPLGP--AGARLYRTGDLARLLPDGQLDFLGRNDFQ
CmnA A1  (356) PAGVVGEVFLGGASVTRGYHARPALTAERFVPDPFGP--PGSRLYRTGDLGRVTPDGELEFLGRRDHQ
   CmnO  (348) --DGLFELHVGGPTLAWGYRDRPAATAERFPPDERGR------RFRTGDLVRVADDGALVFVGRADRQ
VioA A1  (372) APNEVGEVYLGGASVTRGYHGRAALTAQRFLPDPYGP--PGSRLYRTGDLGRVEDNGELRLLGRIDHQ
VioA A2  (357) PRGTVGELYVSGAPVARGYLGRPALTADRFLPDPYGP--PGSRMYRTGDLGRFTGEGLLDFQGRGDFQ
   VbsS  (359) PIGVTGELYIGGEGIARGYLGKPDTTADRFIPDPFFT--EGGRLYRSGDLTRWRDDGTVEFVGRVDHQ
   PacL  (354) PPGTVGELYLGGAGLARCYFGNPALTAERFVPDPFHG--VGTRMYRSGDLGRWTRDGLLMFVGRTDDQ
   GlbF  (365) PVGATGELYVAGAGLARGYLGRPALTAQRFVADPFGA--SGSRMYRTGDLARWRADGGLDFLGRADEQ
   NovH  (355) APGEIGELYLSGAGLAQGYLNSPDLTAQMFVPNPFAA--DGERMYRTGDLASRRADGDILFHGRIDDQ
   CloH  (355) AVGEIGELYISGAGLARGYLNRPDLTAQLFVANPFAA--DGERMYRTGDLASMRADGDILFHGRIDDQ
                                           *
```

Figure S3. Alignment of the amino acid sequences of MbtH protein dependent A-domains. The amino acid site corresponding to CloH L383 is marked by an asterisk and identical residues are indicated by a gray shadow.



```
NocA2    (1) DGFTALVDHVLADPDVRIGRTRLTGAP------VEPVLAGPVAPPAPPVLRLVRDAAARRPDAPAVVDGE
LysB A6  (1) DYFKALLRGMIGDDGRAVARIGILGRDERRRVLEDRNASARPVPQAGGIHRLFEQQAARTPDAIALEQAG

NocA2   (65) VVTGYAELVRRADAVAARIRAAGGGPGDVVGVCLPRSTDLVAALLGVLAAGRAFLPLSPEDPDDRITRQL
LysB A6 (71) QGLSYAELNARANRLARHLRALGVGRDERVAIALPRSAELIVAILATLKAGGAYVPLDPAYPAERLNYLL

NocA2  (135) DLGGARLVIAADADRFAADRFAGVTAVSPLVPAEGAGEAPLAPLPEPLPG--DAAYVIFTSGSTGEPKGV
LysB A6(141) GDS-APKVVLTQRELRAGLPAAAATLLEIDADAAAWAALPAADLDETADDAHALAYVIYTSGSTGQPKGV

NocA2  (203) VVEHRALADHVRWAVGEYGLTAGDRALQFCAVAFDVLVEEVLPTLASGAAVVLRDEESATSAQALVELCA
LysB A6(210) MVEHAGLINQIAALQAHYALSAADRVLQFVSPTFDVSVEEIFTALLSGATLVLRTDEWIAGPARWCELCA

NocA2  (273) ARGVTVANLPTGYWERLVAAFDEDGTALPPSVRLVVIGGQQVDRSAVERWH-RLPNPVRLVNAYGPTEMT
LysB A6(280) QHALTVANLPTLFWQQLAQAAD---VAIPPALRQIVIGGDAVSPAALAAWWSRGGHRPALSNAYGPTETT

NocA2  (342) IGATAADLVPGGG-VPIGGPTENTRAYLLDRYLAPVPDGVVAELYLAGSGLARGYSRRPGLTGERFLPDP
LysB A6(347) INASVADCAPQANPHSVGRPLANTAIYLLDAHGEPVPEGSPGEIYIGGAGVARGYLNRPELSAERFLADP
                                                                             *
NocA2  (411) FAG-GGQRMYRTGDLAVRR-DGALHFIGRVDRQVKVRGHRIELDEVESALTAAPGVAEAAVLLRGG----
LysB A6(417) FAAREGARMYRSGDLGRWLPDGTVEFLGRNDHQLKIRGFRVEPGEIEALLRRLPGIAEAVVHAREDRPGD

NocA2  (475) -RLVAHVAAPPEVDGAGLRAHLAGRLPAFMVPSVVVVSRALPRTSTGKVDRNALTAEPDEPADLAL-
LysB A6(487) KRLVAYLVGADIAPVAELRAALSRELPEYMLPSAFVALERLPLTPSGKLDRLALPAPDADAYAQRNF
```

Figure S4. Alignment of the amino acid sequences of L-Arg activating A-domains; MbtH protein dependent NocA2 versus MbtH independent LysB A6. Amino acid substrate binding site residues predicted by PheA crystal structure are boxed and identical residues are indicated by gray shadowing. The amino acid site corresponding to CloH L383 is marked by an asterisk.



```
ACU39320.1    (1) MAETPRRLRFLLGAAAAVAALLPTTALSASAQPAPVPVPVPADGDVGARIVGGTRVSTADHPWAVYLATSTGAQY
ACU39665.1    (1) MAKTLRRLAKFIGLGVAAAAIG------------LMTTPVSSASDVTPFIVGGTRASISEAPWTVYLASSSGSQF
ACU39678.1    (1) -----MRLAVFLSVLAVLCGGG-------------VATAAAEP-----RVVGGTRASIAEHPWVVFLAGTTGSQF
ACU39675.1    (1) -----MRISPRLVAALALVAVQ-------------LASQLVAAP---AALAVVGGSVAGRAPWVAALLDARGSQF
ACU39677.1    (1) -------MRVAMALALLLTGVA-------------VAPPAGADP---RIVGGVVVDAPNPYPWVVALADQNGRQF
AAS92595.1    (1) -MKHFLRALKRCSVAVATVAIAS--R---------RPPARHRPRPPPTRVVGGTRAAQGEFPFMVRLSMG-----
BAA06423.1    (1) MIRGSRRLARLAGVTAVALSAFAAASV-------ANATPAPPSDDLGTFIVGGEDANVQDHPFTVALVTPDGQQF

ACU39320.1   (76) CGGTLVAPDKVLTAAHCAEDDSPSS--------VRVVVGRDDKDSSAGTTARVTDIWIHPSYQIPAEGYDISVLT
ACU39665.1   (64) CGGTLVKANKVVTAAHCVAGRSASS--------TRVVIGREDKQSTAGTVATVSGIWSHPSYRTATSGYDVAVLT
ACU39678.1   (53) CGGTLVAPTKVVTAAHCVAGTAPER--------LVVVAGREDKRTQAGETLRVARIWKHPFYSSAEQGSDVAVLT
ACU39675.1   (55) CGGALIAPDRVVTAAHCTQERGLLGPRDRAPRQLTVVLDRHDLRTRSGVRVGVSGIWRHPAFTDVGRGDDLAVLQ
ACU39677.1   (53) CGGAIVSRYQVVTAAHCVN--GLTG------RDVLVVAGRLDVTKPGGTVHPVSQVYVHSAYRSATSGADIAYLL
AAS92595.1   (59) CGGALYAQDIVLTAAHCVSGSGNNTS-------ITATGGVVDLQSSSAVKVRSTKVLQAPGYN--GTGKDWALIK
BAA06423.1   (69) CGGTLAAPNKVVTAAHCTVGSQPAD--------INVVSGRTVMSSNEGTVSKVTNVWVHPEYQDAAKGFDVSVLT
                                 *                                                    *

ACU39320.1  (143) LDRALNAEPLPLARPTDSALYEEGVPATVLGWGATSSGGTASRHLLGATVPVIGSTACTTAYAERYNANLMVCAG
ACU39665.1  (131) LGTSVSGTYLPLATPSDTALYAAGTNAVAYGWGATCSGCSTSRYLLKVTVPVTSDATCKTAYSQ-YSNTSMVCAG
ACU39678.1  (120) LARPVANRVLPLAS----ADPAPGTVAKVLGWGRTAEIGAASQVLLGANVPVMSNQDCTADYAR-FDPAAMVCAG
ACU39675.1  (130) LDGPVAFEPVRVG------DAVAGRTATAYGWGRTAESGPVSSLLRQVDVPIRADHDCAAALAGYR-ADAMLCAG
ACU39677.1  (120) TSTPMPYEPIALAKPSETGLYRAGITGTAFGWGRTSESGSTSPVLRKVSVPLQTNEVCAAAYPGYYQSSSMFCAG
AAS92595.1  (125) LAQPINQPTLKIAT----TTAYNQGTFTVAGWGANREGGSQQRYLLKANVPFVSDAACRSAYGNELVANEEICAG
BAA06423.1  (136) LEAPVKEAPIELAK-ADDAGYAPDTAATILGWGNTSEGGQQADHLQKATVPVNSDDTCKQAYGE-YTPDAMVCAG
                                                ———           —

ACU39320.1  (218) FPK-GGTDSCQGDSGGP-----LVAGGKLIGVTSWGEGCAAAGAPGVYSRVAAYSTELEPQLTP------
ACU39665.1  (205) VPA-GGKDTCQGDSGGP-----LVAGGKLIGATSWGNGCALPNYPGVYARVAAYYSVLSAQIG-------
ACU39678.1  (190) YPR-GGIDTCQGDSGGP-----LVANGALIGVTSWGEGCARPGQPGVYSRITAYRALIDQQLKG------
ACU39675.1  (198) FPD-GGRDACTGDSGGP-----LVSDGLLVGVVSYGRGCARAGQPGAYTRLAHYRDRL------------
ACU39677.1  (195) LPN-GGKDACQGDSGGP-----FIIAGRLAGLVSWGVGCARPEYPGVYTRVGVYT---------------
AAS92595.1  (196) YPDTGGVDTCQGDSGGPMFRKDNADEWIQVGIVSWGYGCARPGYPGVYTEVSTFASAIASAARTL-----
BAA06423.1  (209) VPE-GGVDTCQGDSGGP-----MVVNNKLIGVTSWGEGCARPGKPGVYARVGAYYDVLMEQINAGAVSAR
                              —*
```

Figure S5. Alignment of five trypsin proteins (ACU prefix) encoded in the *Actinosynnema mirum* DSM 43827 genome (GenBank NC_013093.1) aligned to the characterized trypsins from Streptomyces griseus (AAS92595.1) and Saccharopolyspora erythraeus (BAA06423.1). Identical residues are indicated by a gray shadow. Conserved motif in specificity pocket (GGxD) is boxed. Residues of the catalytic triad are marked with an asterisk. Conserved residues associated with zymogen activation are underlined.