Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.
This journal is © The Royal Society of Chemistry 2017

Supplemental Information

Dissection of the neocarazostatin A C_4 alkane side chain biosynthesis by *in vitro* reconstitution

Li Su[1], Rui Zhang[1], Kwaku Kyeremeh[2], Zixin Deng[1], Hai Deng[3,*], Yi Yu[1,*]

[1]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Ministry of Education), School of Pharmaceutical Sciences, Wuhan University, 185 East Lake Road, Wuhan 430071, P. R. China. Email: yu_yi@whu.edu.cn

[2]Department of Chemistry, University of Ghana, P.O. Box LG56, Legon-Accra, Ghana

[3]Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen AB24 3UE, Scotland, U.K. Email: h.deng@abdn.ac.uk

*Corresponding authors

Experimental Procedures

Materials, media and strains

Acetyl-CoA, Malonyl-CoA, Acetoacetyl-CoA, and Hydroxybutyryl-CoA were purchased from Sigma-Aldrich. Aliquots of acyl-CoAs were maintained at -80 °C with 0.05% formic acid. EGS, a homobifunctional *N*-hydroxysuccinimide (NHS) ester and a popular amine-reactive crosslinker, was purchased from Thermo Fisher Scientific. *E. coli* DH10B and *E. coli* BL21(DE3) transformed with the expression plasmid pET28a, separately containing *nzsE* , *nzsF*, *ma37-fabG1/2, ma37-fabC* and *sfp*, were cultured in Luria-Bertani or Luria agar medium at 37 °C.

Cloning and construction of His6-tagged NzsE, NzsF, MA37-FabC, MA37-FabG1/2, and Sfp overexpression plasmids

PCR amplification was performed in order to add a 6x His tag at the N-terminus of target genes. The genes encoding NzsE, NzsF, MA37-FabC, MA37-FabG1, and MA37-FabG2 were amplified from MA37 genomic DNA separately by using primer pairs 5′-TGC CGC GCG GCA GCC ATA TGG AGA CTC TGA ACA ACC GGC TCG AGG CCC-3′ and 5′-TGG TGG TGG TGC TCG AGT CAT CAG GCC GAG GCA CCC TTG GAC GTG AGG AG-3′ (for *nzsE*); 5′-GCC GCG CGG CAG CCA TAT GAC GCG AGC CGC CGT CGT CGG CGG TCT CG-3′ and 5′-TGG TGG TGG TGC TCG AGT CAT CAC TGT TCC TGC AGT CCG ATA GTG ATG TC-3′ (for *nzsF*); 5′-GTG CCG CGC GGC AGC CAT ATG GCT CAC ACC CAG GAG CAG ATC CTC GAG-3′ and 5′-TGG TGG TGG TGG TGC TCG AGT CAT CAG GCC TGG TGC TTC AGG ATG TAG CC-3′ (for *MA37-fabC*), 5′-GTG CCG CGC GGC AGC CAT ATG AGC CGC TCG GTT CTC GTC ACA GGA GGA -3′ and 5′-TGG TGG TGG TGG TGC TCG AGT CAG TGA CCC ATG CCC AAT CCG CCG TCG AC-3′ (for *ma37-fabG1*), 5′-GTG CCG CGC GGC AGC CAT ATG TCC ACC ACC GAG CAG CGC GTC GCC CTG-3′ and 5′-TGG TGG TGG TGG TGC TCG AGT CAG TCG AGC GGC CCG CCG GCC ACG TAC AG-3′ (for *ma37-fabG2*). The *sfp* gene was amplified from *Bacillus subtilis* by using the primers 5′-TGC CGC GCG GCA GCC ATA TGA AGA TTT ACG GAA TTT ATA TGG ACC GC-3′ and 5′-TGG TGG TGG TGC TCG AGT TAT TAT AAA AGC TCT TCG TAC GAG ACC ATT GT-3′. All the amplified products were digested with NdeI and XhoI, purified using a

Omega PCR purification kit, and then inserted into the corresponding restriction sites of the expression vector pET28a (Novagen) to yield the pET28a-*nzsE*, pET28a-*nzsF* and pET28a-*sfp* respectively. Chemically competent *E. coli* DH10B cells were separately transformed with these recombinant plasmids and plated on LB agar containing kanamycin (50 μg mL^{-1}) to screen for positive clones, which were further confirmed by DNA sequencing.

Site-directed mutagenesis of NzsF

To reveal the conserved catalytic triad (CHN) of NzsF, the amino acid residue (C112, H254, N284) was individually replaced by an Ala, which was achieved by using a site-directed mutagenesis procedure according to the manufacturer (TOYOBO). Using the plasmid of pET28a-*nzsF* as the template, PCR was performed to substitute the individual amino acid with Ala by utilizing the following primer pairs: C112A-F, 5′- GCG CCT TCG ACG TGG CCG CCG TGG CAA CGG GCT TCG TGT ACG CGC TGG CC-3′, and C112A-R, 5′-GGC CAG CGC GTA CAC GAA GCC CGT TGC CAC GGC GGC CAC GTC GAA GGC GC-3′; H254A-F, 5′-GAG GAG GTC GGC GCC GTC GTC CCG GCC CAG GCC AAT GTC CGC ATC CTC GA-3′, and H254A-R, 5′-TCG AGG ATG CGG ACA TTG GCC TGG GCC GGG ACG ACG GCG CCG ACC TCC TC-3′; H284A-F, 5′-GTC AAG AAC ATC GAC CGG GTC GGC GCC ACC GTG GCC GCG TCC ATC CCG CT-3′ and H284A-R, 5′- AGC GGG ATG GAC GCG GCC ACG GTG GCG CCG ACC CGG TCG ATG TTC TTG AC-3′. After checking the PCR products by agarose gel electrophoresis, 1 μL of the PCR product was treated with the restriction enzyme DpnI (TOYOBO), and the total 20 μL reaction solution was incubated at 37 °C for ~1.5 h. 2.5 μL of the resulting solution was used for transformation of *E. coli* DH10B cells. The positive clones were picked and verified by DNA sequencing.

Protein overexpression and purification

E. coli BL21 (DE3) cells containing pET28a-*nzsE,* pET28a-*nzsF,* pET28a-*sfp*, pET28a-*ma37-fabC,* and pET28a-*ma37-fabG1/2* overexpression plasmid were separately grown in 500 mL LB medium supplemented with 50 mg/ml kanamycin at 37 °C with shaking at 220 rpm until an OD_{600}~0.6-0.8 was reached. Protein expression was induced by addition of 0.5 mM (final concentration) isopropyl β-D-1-

thiogalactopyranoside (IPTG) with further cultivation at 18 °C for 16 h. The cells were collected by centrifugation and re-suspended in 40 ml ice-cold lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 10 mM imidazole, and 10% (V/V) glycerol), and disrupted by a Nano Homogenize Machine (ATS Engineering Inc, AH100B). Cellular debris was removed via centrifugation at 12000 rpm for 30 min at 4 °C. The supernatant was loaded onto 2 ml Ni-NTA-affinity column (GE Healthcare) pre-equilibrated with the lysis buffer. The desired elution fractions were combined and concentrated using an Centrifugal Filter Units (Milipore,3,000 and 10,000 MWCO), and the concentrated protein solution was desalted using a PD-10 Column (GE Healthcare) pre-equilibrated with the storage buffer (20 mM HEPES pH7.5, 150 mM NaCl, and 10% (v/v) glycerol). The protein fractions were collected and concentrated, analyzed by SDS-PAGE (12% Tris-glycine gel), and frozen, stored at −80 °C upon further use. Protein concentration was determined by using a Bradford assay kit (Promega) with bovine serum albumin (BSA) as the standard.

Cross-linking assay of NzsF

According to the protein cross-linking method described by Lam et al. [1], cross-linking assay of NzsF was carried out on a 20 μL scale with 20 mM HEPES pH7.5, 100mM NaCl, 50 mM NzsF, and 5mM EGS (ethyleneglycol-bis-succinimidyl-succinate, Sigma, dissolved with dry DMSO immediately before use). Buffers containing primary amines (e.g. Tris or glycine) should be avoided during conjugation process as they will react with the NHS ester and will inhibit and reduce conjugation efficiency of the intended molecules. The reaction mixture was incubated on ice for 2 hours. After that, the reaction was quenched for 15 minutes using a solution containing amines such as Tris at a final concentration of 30 mM, and the sample was then ready for subsequent analysis by SDS-PAGE (12% Tris-glycine gel).

***In vitro* characterization of NzsE**

The reaction was carried out on a 100 μL scale with 50 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 200 μM acyl-CoA (acetyl-CoA, malonyl-CoA, acetoacetyl-CoA, or 3-hydroxybutyl-CoA), 1.5 μM Sfp (4′-phosphopantetheinyl transferase), and 40 μM apo-NzsE. Reactions were initiated by the addition of Sfp, followed by incubating at 30 °C for

60 min, then were quenched by equal volume cold 6% formic acid water. The clarified supernatant was immediately analyzed by HPLC and HR-ESIMS.

In vitro characterization of NzsF and its mutants

The reaction were carried out on a 100 μL scale with 50 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 200 μM malonyl-CoA, 200 μM acetyl-CoA, 1.5 μM Sfp, and 40 μM apo-NzsE or apo-MA37-FabC, and 1 μM NzsF or its mutants. Reactions were initiated by the addition of Sfp, followed by incubating at 30 °C for 2 h, then were quenched by equal volume cold 6% formic acid water. The clarified supernatant was analyzed by HPLC and HR-ESIMS. In the case of NzsF mutant (C112A, H254A, N284A), for which no activity was detected, the reaction time was extended to 3 h (with wild-type enzyme as a control) to confirm the absence of product.

In vitro characterization of MA37-FabG1 and MA37-FabG2

The reactions were carried out on a 100 μL scale with 50 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 1 mM NADPH/NADH, 200 μM acetoacetyl-CoA, 1.5 μM Sfp, 40 μM apo-NzsE, and 1.5 μM MA37-FabG1 or MA37-FabG2. Reactions were initiated by the addition of Sfp, followed by incubating at 30 °C for 2 h, then were quenched by equal volume cold 6% formic acid water. The clarified supernatant was analyzed by HPLC and HR-ESIMS.

HPLC and LC-MS Parameters of Enzymatic Reaction

A DIONEX System equipped with a P680 HPLC Pump and UVD 170U detector was used for HPLC analysis, which was performed on a Jupiter 300 Å C18 column (Phenomenex).The column was equilibrated for 5 min with 80% solvent A (H_2O, 0.1% formic acid) and 20% solvent B (CH_3CN, 0.1% formic acid),5–25 min eluted with a linear gradient from 80% A/20% B to 5% A/95% B, 25-29 min, constant with 5%A/95%B; 29-30min, a linear gradient to 80%A/20%B; UV monitored at 214 nm at a flow rate of 0.8 ml/min.

LC-MS analysis was carried out in positive mode by using a Thermo Scientific LTQ XL Orbitrap mass spectrometer equipped with a Thermo Scientific Accela 600 pump (Thermo Fisher Scientific Inc.). Each LC conditions were described as above. All MS analysis parameters were set as 45 V capillary voltage, 45 °C capillary temperature,

auxiliary gas flow rate 10 arbitrary units, sheath gas flow rate 40 arbitrary units, 3.5 kV spray voltage, and 100-2000 Amu mass range (maximum resolution 100,000). For the tandem MS spectrometry, the method of Collision Induced Dissociation (CID) was used.

Phylogenetic analysis of NzsF

Homologous sequences of FabH (both from *E. coli* and *Streptomyces*), and typical members belonging to KASI, KASII, and KASIII-like enzyme families were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/) with their accession numbers and the source organisms listed in Table S1. Bayesian Markov Chain Monte Carlo (MCMC) method was used to calculate posterior probability of various clades through the program MrBayes (version 3.2) [2]. A mixed model of amino acid substitutions was used, and final analyses consisted of four chains (one cold and three heated) were run for about 2,000,000 generations by sampling parameters and trees every 1,000 generations. Analyses were run to reach a convergence with standard deviation of split frequencies <0.004017. Posterior probabilities were averaged over the final 75% of trees (25% burn in). The figure of the Bayesian phylogram was prepared by using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Table S1. Accession number and sources of KAS Ⅰ, KAS II and KASIII used for construction of the phylogenetic tree.

Accession number	Source	Enzymes
ALL53319.1	*Streptomyces sp.* MA37	KAS Ⅲ (NzsF)
ANK06613.1	*Escherichia coli*	KAS Ⅲ（FabH)
2x3E	*Pseudomonas Aeruginosa Pao1*	KAS Ⅲ
CAA16179.1	*Streptomyces coelicolor* A3(2)	KAS Ⅲ (RedP)
NP_626634.1	*Streptomyces coelicolor*	KAS Ⅲ
AAV84077.1	*Streptomyces echinatus*	KAS Ⅲ
YP_001826619.1	*Streptomyces griseus subsp. griseus* NBRC 13350	KAS Ⅲ
Q54206.1	*Streptomyces glaucescens*	KAS Ⅲ
AAQ08929.1	*Streptomyces griseus*	KAS Ⅲ
CAM58805.1	*Streptomyces sp.* A2991200	KAS Ⅲ(BenQ)
AAG30195.1	*Streptomyces sp.* R1128	KAS Ⅲ
ACI88883.1	*Streptomyces sp.* CM020	KAS Ⅲ
AAC18104.1	*Streptomyces roseofulvus*	KAS Ⅲ
YP_143679.1	*Thermus thermophilus* HB8	FabF (KAS Ⅱ)
NP_344945.1	*Streptococcus pneumoniae* TIGR4	FabF (KAS Ⅱ)
NP_645683.1	*Staphylococcus aureus subsp. aureus* MW2	FabF (KAS Ⅱ)
ZP_02902779.1	*Escherichia albertii* TW07627	FabF (KAS Ⅱ)
NP_415613.1	*Escherichia coli str.* K-12 substr. MG1655	FabF (KAS Ⅱ)
ZP_00134992.2	*Actinobacillus pleuropneumoniae serovar*	FabB (KAS Ⅰ)
ZP_04562837.1	*Citrobacter sp.* 30_2	FabB (KAS Ⅰ)
NP_416826.1	*Escherichia coli str.* K-12 *substr.* MG1655	FabB (KAS Ⅰ)
YP_001881145.1	*Shigella boydii* CDC 3083-94	FabB (KAS Ⅰ)
ALE27503.1	*Streptomyces sp.* NRRL 12068	KAS Ⅲ-like (BomK)
AFB35630.1	*Streptomyces antibioticus*	KAS Ⅲ-like (EsmD1
ACJ24876.1	Streptomyces *pactum*	KAS Ⅲ-like (PtmR)
AAZ77679.1	*Streptomyces antibioticus*	KAS Ⅲ-like (ChlB6)
AAK83178.1	*Streptomyces viridochromogenes* Tue57	KAS Ⅲ-like (AviN)
CAC60011.1	*Micromonospora carbonacea*	KAS Ⅲ-like
ADU85987.1	*Dactylosporangium aurantiacum*	KAS Ⅲ-like (TiaF)
AAM70354.1	*Micromonospora echinospora*	KAS Ⅲ-like (CalO4)
AAN65231.1	*Streptomyces roseochromogenus*	KAS Ⅲ-like (CloN2)
ACN64832.1	*Streptomyces diastatochromogenes*	KAS Ⅲ-like (PokM2)
AEI91069.1	Streptomyces *tendae*	KAS Ⅲ-like (CerJ)
AAA65208.1	*Streptomyces peucetius*	KAS Ⅲ-Like (EviR)
AAF70109.1	*Streptomyces galilaeus*	KAS Ⅲ-Like (AknE2)
WP_008748182.1	*Streptomyces sp.* SPB074	KAS Ⅲ-Like

Figure S1. Common tricyclic carbazole alkaloids from bacteria and the proposed biosynthetic pathway of Neocarazostatin A

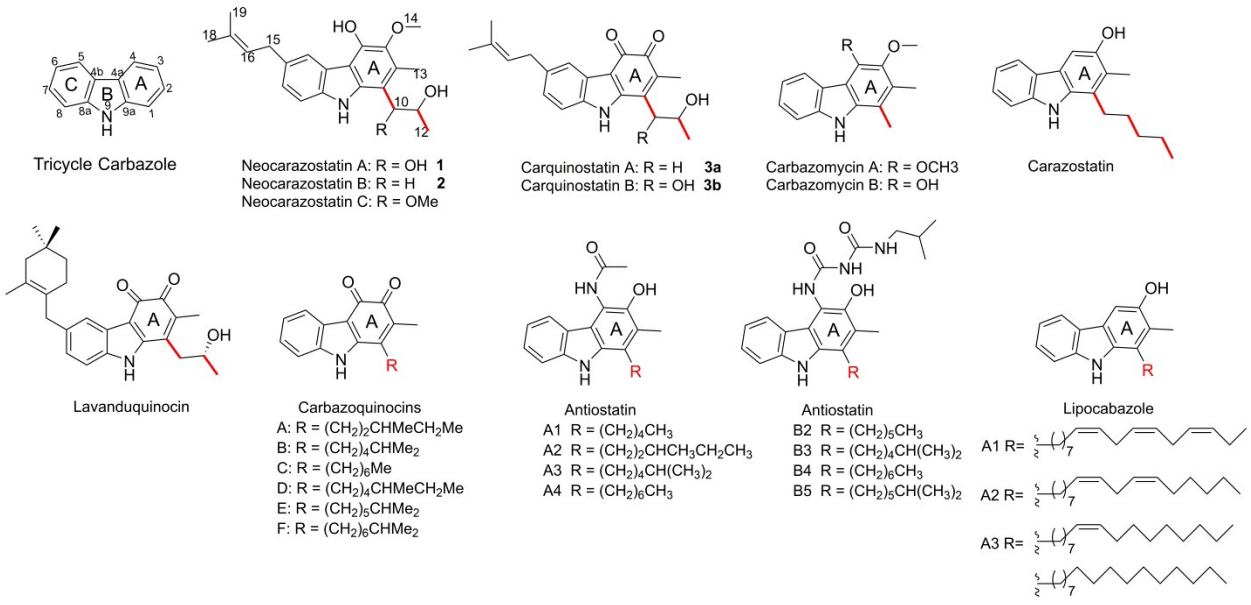


Figure S2. SDS-PAGE analysis of the purified proteins

A) Purified 6xHis-tagged NzsE (predicted to be 10.6kDa in size).

B) Lane 1 is the purified 6xHis-tagged NzsF (predicted to be 36.7kDa in size), Lane 2 is the EGS-treated NzsF, and the black star indicates the dimeric NzsF (predicted to be 73.4kDa in size).

C) Purified 6xHis-tagged Sfp (phosphpantetheinyl transferase) with predicted size of 28.3kDa.

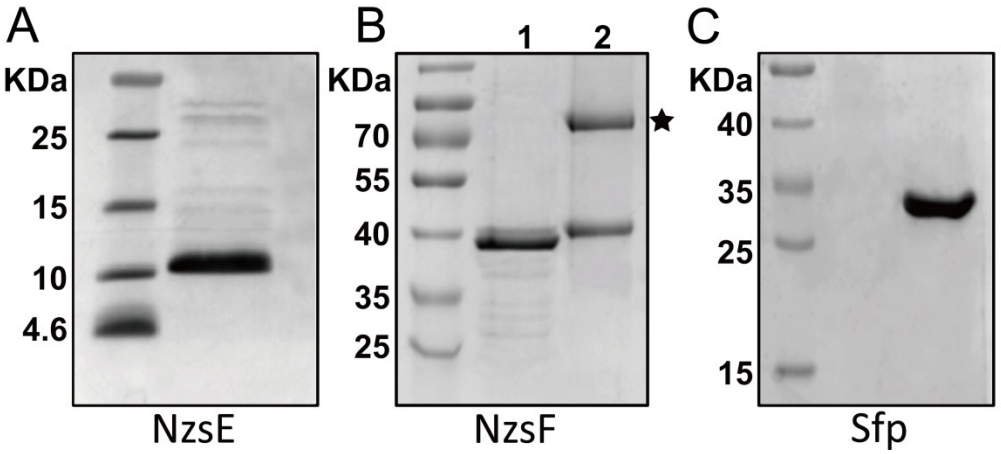


Figure S3. Pantetheinyl elimination fragments observed during tandem mass spectrometry of acyl-NzsE products corresponding to the samples **b**, **c**, **e** in Figure 2 and the sample **c** in Figure 4.

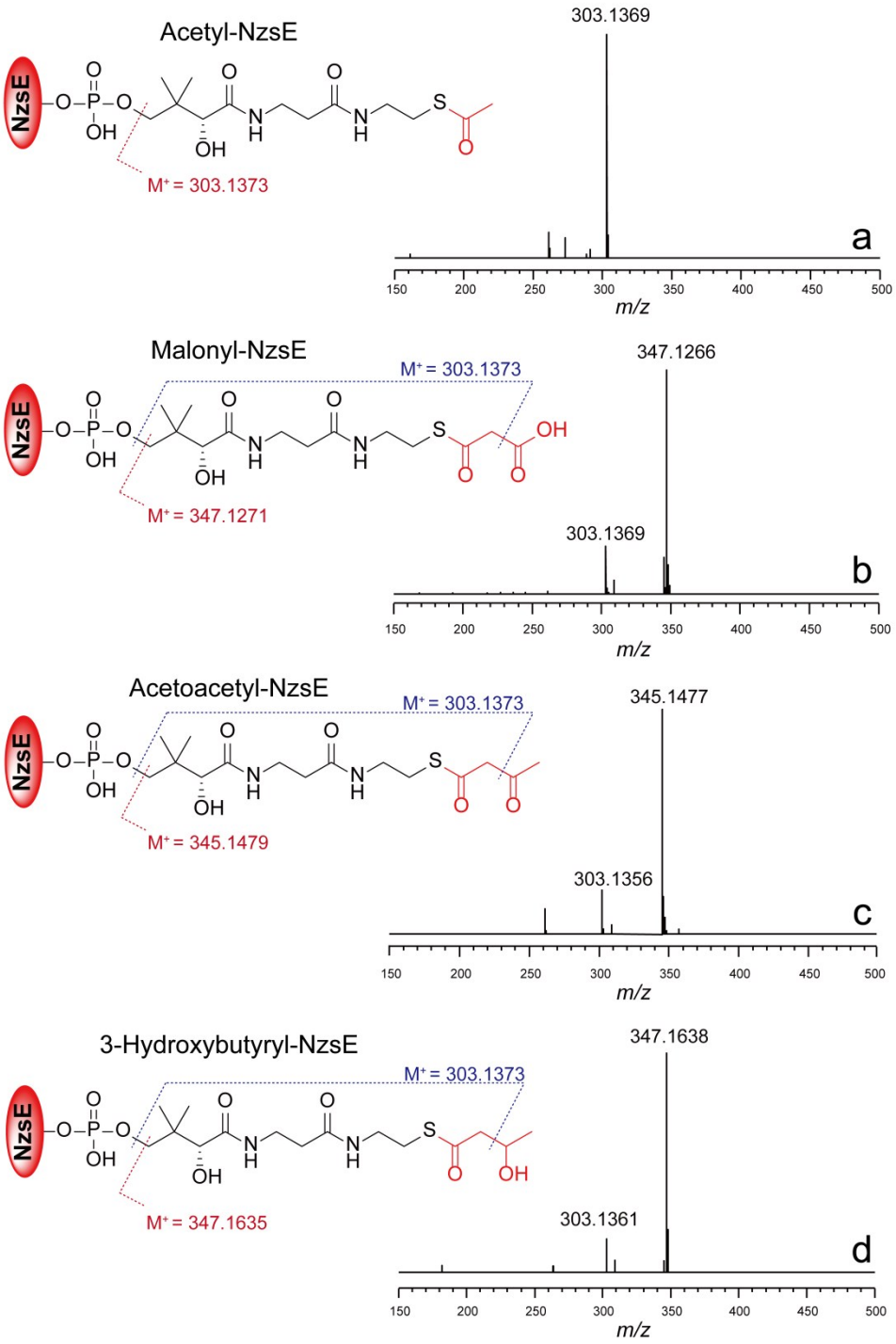
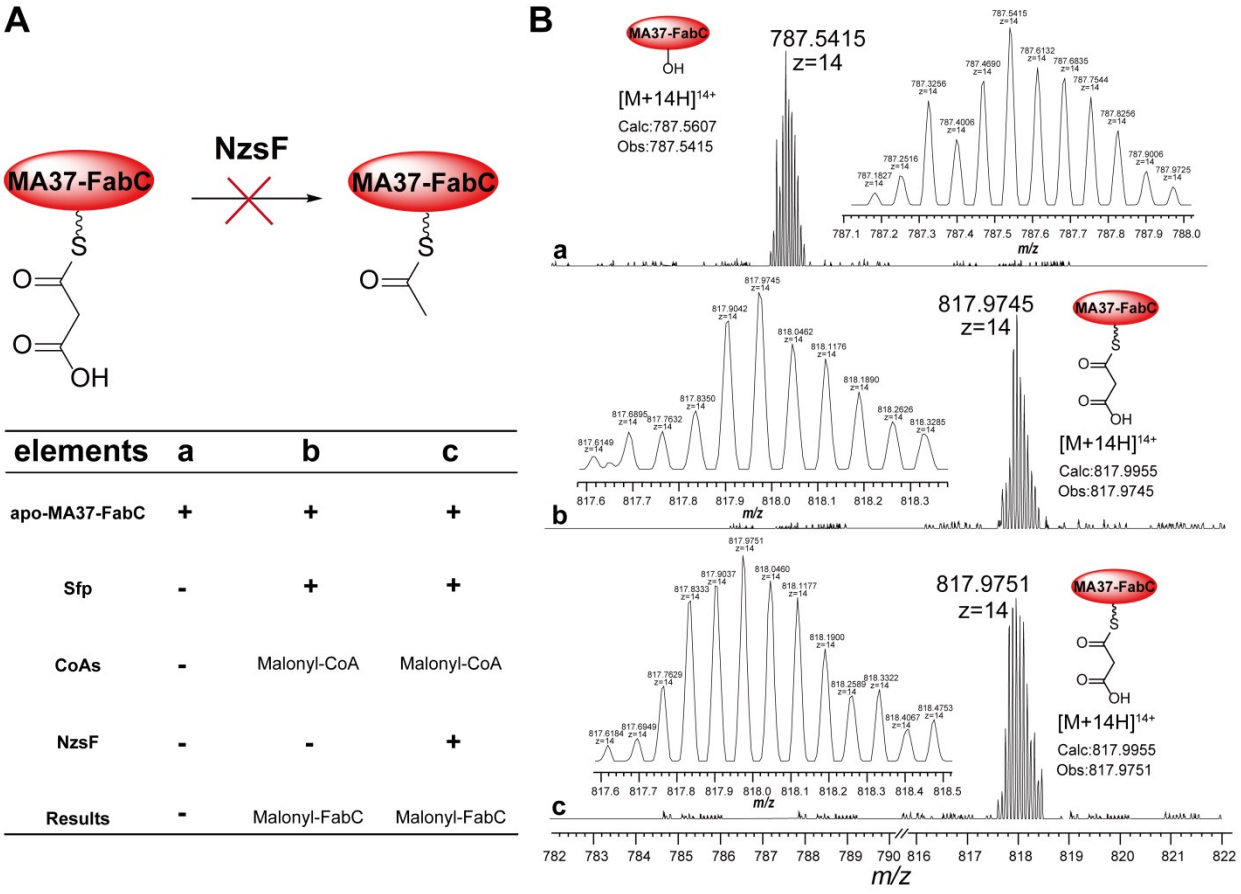


Figure S4. *In vitro* characterization of NzsF with MA37-FabC

A) Proposed reaction catalyzed by NzsF in the biosynthesis of Neocarazostatin A. A series of reactions were designed to characterize the activity of NzsF.

B) HR-ESIMS data for MA37-FabC products. Isotope peak of $[M+14H]^{14+}$ were collected and presented.



References

1. K. S. Lam, P. Scuderi and S. E. Salmon, *Journal of biological response modifiers*, 1988, 7, 267-275.
2. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck, *Systematic biology*, 2012, 61, 539-542.