Supporting Information for:

Identification and biosynthesis of pro-inflammatory sulfonolipids from an opportunistic pathogen *Chryseobacterium gleum*

Lukuan Hou[†,‡], Hai-Yan Tian[§], Li Wang[§], Zachary E Ferris[†], Junfeng Wang[^], Mingwei Cai[‡], Ethan A Older[†], Manikanda Raja Keerthi Raja[‖], Dan Xue[†], Wanyang Sun[§], Prakash Nagarkatti[▽], Mitzi Nagarkatti[▽], Hexin Chen[‖], Daping Fan[^], Xiaoyu Tang[‡,*] and Jie Li[†,*]

[†]Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, United States.

[‡]Institute of Chemical Biology, Shenzhen Bay Laboratory, Shenzhen 518132, China.

[§]College of Pharmacy, Jinan University, Guangzhou 510632, China

[^]Department of Cell Biology and Anatomy, School of Medicine, University of South Carolina, Columbia, South Carolina 29209, United States.

[‖]Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29209, United States.

[▽]Department of Pathology, Microbiology and Immunology, School of Medicine, University of South Carolina, Columbia, South Carolina 29209, United States.

*Correspondence:

Xiaoyu Tang: xtang@szbl.ac.cn; Jie Li: li439@mailbox.sc.edu

Table of contents

1. General experimental procedures.

Nuclear magnetic resonance (NMR) spectra were acquired on either a Bruker Avance III HD 400 MHz spectrometer with a 5 mm BBO 1H/19F-BB-Z-Gradient prodigy cryoprobe, a Bruker Avance III HD 500 MHz spectrometer with a PA BBO 500S2 BBF-H-D_05 Z SP probe, or a Bruker Avance III HD Ascend 700 MHz equipped with 5mm triple-resonance Observe (TXO) cryoprobe with Z-gradients, controlled by TopSpin 3.6.1 software. In all cases spectra were acquired at 25 °C (unless otherwise specified) in solvents as specified in the text, with referencing to residual ^{1}H or ^{13}C signals in the deuterated solvents. High-resolution ESIMS spectra were obtained on a Thermo Scientific Orbitrap Velos Pro hybrid ion trap-orbitrap mass spectrometer by direct injection. Liquid chromatography-diode array-electrospray ionization mass spectrometry (LC-DAD-ESIMS) data were acquired on a Thermo Dionex Ultimate 3000 UHPLC system equipped with a diode array multiple wavelength detector and an LTQ XL linear ion trap mass spectrometer controlled by Thermo Xcalibur (version 4.2.47). The sheath gas was at a flow rate of 35 arbitrary units, source heater temperature of 325 °C, and capillary temperature of 350 °C were set for the ion trap mass spectrometer. Semi-preparative HPLC purifications were performed using Thermo Dionex Ultimate 3000 HPLC system with corresponding pump, autosampler, UV-vis detectors, fraction collectors, and Chromeleon software (version 7.2.10) inclusively. LCMS grade and HPLC-grade CH_3CN, H_2O and formic acid were purchased from Fisher Scientific. Deuterated solvents were purchased from Cambridge Isotopes.

2. UPLC-Q-TOF-MS/MS analysis of *Chryseobaterium gleum* DSM 16776 strain extract.

The extract was analyzed by UPLC-Q-TOF-MS, using a Phenomenex Kinetex XB-C18 Column (100 Å, 2.6 μm, 4.6×100 mm) at a flow rate of 0.4 mL/min with mobile phase composed of (A) 75% MeOH in water (v/v) containing 5 mmol/L $HCOONH_4$ and (B) ethanol:10%-100% B at 0-20 min, 100%-100% B at 20-26 min. Q-TOF-

MS/MS analysis was performed in negative ion mode with a mass range 100-1000 Da for Q-TOF-MS scan and 50-1000 Da for TOF-MS/MS scan. The following optimized operating parameters were used: ion spray voltage, -4500V; source temperature, 550 ℃; ion source gas 1, 55 psi and ion source gas 2, 50 psi; the pressure of Gurtain gas [nitrogen (N_2)], 30 psi; collision energy, 35 eV; mass tolerance, 50 mDa. The experiments were run with 0.15 s accumulation time for TOF-MS and 0.05 s accumulation time for TOF-MS/MS.

3. Extract features using MZmine.

The raw data with MS^2 fragment information was converted into mzML file via MSconvert (http://proteowizard.sourceforge.net/download.html) and was subsequently processed by MZmine 2 (version 2.53, https://github.com/mzmine/mzmine2/releases) to retain the precursor ions containing MS peak detection, chromatogram building, chromatogram deconvolution, isotope grouping, and feature alignment. The mass detections were realized by keeping the noise level at $5.0E^3$ for MS^1 and $1.0E^0$ for MS^2, respectively. The ADAP chromatogram building was achieved using a minimum group size in 3 scans, group intensity threshold of 100, the minimum highest intensity of $5.0E^3$, and *m/z* tolerance of 0.02 (or 5 ppm). The Chromatogram deconvolution algorithm was used with the following parameter settings: minimum peak height $5.0E^3$, peak duration range 0.00–5.0 min, *m/z* range for MS^2 scan pairing 0.01 Da, retention time (RT) range for MS^2 scan pairing 0.2 min. The Chromatograms chromatographic were finally processed using the isotopic peaks grouping algorithm with an *m/z* tolerance of 0.02 (or 5 ppm) and retention time tolerance of 0.2 min. Then, a csv file and an mgf file were exported and processed by Python. The refined data were submitted to GNPS to gain clusters, and some of the node compounds were identified by the software Sirus 4 (https://bio.informatik.uni-jena.de/sirius/).

Fragment ion at *m/z* 79.97 Da was used to recognize Sulfonolipids from LC-MS dataset. The script for product ion filtering was written in Python (version 3.7.3) and run on Jupyter Notebook (version 6.0.3).

Python scripts:

Product ion *m/z* 79.95 Da were used to discover SoLs from LC-MS dataset. The script for typical fragment ion filtering was written in Python (version 3.7.3) and run on Jupyter Notebook (version 6.0.3). Mass tolerance window was set as 0.05.

Python scripts:

```
# prase .mgf
def parse_data(path):
    df = pd.read_csv(path,header=None)
    global feature_id
    for rowid, item in df[0].iteritems():
        counter = 0
        if item[:3] == 'BEG':
            counter += 1
            mzlist = []
            mzlist_5 = []
            intenselist = []
            ms2_plot = []
        elif item[:3] == 'FEA':
            feature_id.append(item.split('=')[1])
        elif item[:3] == 'PEP':
            ms1 = round(float(item.split('=')[1]), 4)
            per_mass.append(ms1)
        elif item[:3] == 'SCA':
            scannum = item.split('=')[1]
            scan_id.append(scannum)
        elif item[:3] == 'RTI':
            rtmin_temp = (round(float(item.split('=')[1]) / 60, 3))
            rtmin.append(rtmin_temp)
        elif item[:3] == 'CHA':
            charge.append(item.split('=')[1])
        elif item[:3] == 'MSL':
            mslevel.append(item.split('=')[1])
        elif item[:3] == 'END':
            ms2.append(mzlist)
            for rowid, mz in enumerate(mzlist):
                if intenselist[rowid]/max(intenselist) >= 0.05:
                    mzlist_5.append(mz)
            mz_list_threshold.append(mzlist_5)
        elif item == '' or item == None:
            print(None)
        else:
```

```
            temp = item.split(' ')[0]
            temp = float(('%.4f' % float(temp)).rstrip('0').rstrip('.'))
            mzlist.append(temp)
    # convert format
    feature_id = [int(i) for i in feature_id]

# define Interval class
class Interval(object):
    def __init__(self, middle, deviation):
        self.lower = middle - abs(deviation)
        self.upper = middle + abs(deviation)

    def __contains__(self, item):
        return self.lower <= item <= self.upper

# define interval
def interval(middle, deviation):
    return Interval(middle, deviation)

# define mass filter
def mass_filter(filter_list, exact_mass, tolerance):
    filter_temp = []
    for mz_list in filter_list:
        filter_temp.append(True in [i in interval(exact_mass, tolerance) for i in mz_list])
    return filter_temp

# generate feature dataframe from .mgf file
parse_data(path)
column = [feature_id,per_mass,mz,rtmin,cha,ms2,mz_list_threshold]
column_name = ['ID','Parent mass','mz','RTmin','Charge', 'MS2','mzlistthreshold']
fe_df = pd.DataFrame(np.column_stack(column),columns = column_name)

# Filter mass spectrum with product ion m/z 79.95
filter_pi = mass_filter(mz_list, exact_mass=79.95, tolerance=0.05)
```

4. Strains, and cultural conditions for heterologous expression.

All strains and plasmids used in this study are listed in Table S1. *Escherichia coli* DH5α was used as the host for general subcloning. *Escherichia coli* BL21 (DE3) was used for protein overexpression. *E. coli* strains were routinely cultured in Luria–Bertani (LB) liquid medium at 37 °C, 200 rpm, or on LB agar plate at 37 °C, with appropriate antibiotic selection. *C. gleum* DSM 16776, *F. johnsonie* DSM 2064, *A.*

timonensis DSM 25383, and *A. machipongonensis* DSM 24695 were purchased from DSMZ. *C. scophthalmum* JUb44 was purchased from Caenorhabditis Genetics Center. *C. gleum* DSM 16776, *F. johnsonie* DSM 2064, *C. scophthalmum* JUb44, and *A. machipongonensis* DSM 24695 are aerobic strains. They were cultured in nutrient broth. *A. timonensis* DSM 25383 is an anaerobic strain. It was cultured in BHI media in the anaerobic incubator (Coy Anaerobic Chambers).

5. General genetic manipulation methods.

Plasmid and genomic DNA extractions as well DNA purifications were carried out using standardized commercial kits (OMEGA, Bio-Tek). PCR amplification was carried out with primers listed in Table S2 using PrimeSTAR HS DNA polymerase (Takara Bio). PCR amplifications were carried out on an Eppendorf® Mastercycler® Nexus X2 Thermal Cycler (Eppendorf Co., Ltd. Germany). Oligonucleotide synthesis and DNA sequencing were performed by Eton Bioscience (North Carolina, USA). DNA assembly was conducted using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). Restriction endonucleases were purchased from New England Biolabs, Inc. (USA). Biochemicals and media components for bacterial cultures were purchased from Thermo Fisher Scientific (USA) and VWR (USA). Chemical reagents were purchased from Sigma Aldrich.

6. Cloning of *cys* and *cfat* genes.

The *cys1* gene was PCR amplified from the genome of *C. gleum* DSM 16776 using primer pairs of pACYCDuet-1-*cys1*-Fwd and pACYCDuet-1-*cys1*-Rev (Table S2). After *Bam*HI/*Hind*III digestion of pACYCDuet-1 vector, the PCR product was cloned into the linearized vector by Gibson assembly resulting in pLLH112. The *cfat1* gene was PCR amplified from the genome of *C. gleum* DSM 16776 using primer pairs of pHis8-*cfat1*-Fwd and pHis8-*cfat1*-Rev (Table S2). After *Bam*HI/*Hind*III digestion of pHis8 vector, the PCR product was cloned into the linearized vector by Gibson assembly resulting in pLLH105. The pLLH112 and pLLH105 constructs were confirmed by Sanger sequencing. The genes *cys2*, *cys3*, *cfat2,* and *cfat3* were PCR amplified from the genome of *C. gleum* DSM 16776; genes *cys-DSM 206*4 and *cfat-*

DSM 2064 were PCR amplified from the genome of *F. johnsonie* DSM 2064; genes *cys-DSM 25383* and *cfat-DSM 25383* were PCR amplified from the genome of *A. timonensis* DSM 25383; genes *cys-JUb44* and *cfat-JUb44* were PCR amplified from the genome of *C. scophthalmum* JUb44; gene *cys-DSM 24695* was PCR amplified from the genome of *A. machipongonensis* DSM 24695. These genes were cloned into pACYCDuet-1 and pHis8 vectors separately to form pLLH106-pLLH111 and pLLH1113-pLLH118 which were confirmed by Sanger sequencing. Gene *cfat-DSM 24695* was synthesized on the pET28a (+) vector by BGI Genomics Co., Ltd.

7. Protein expression and purification.

The constructed plasmids were transformed into *E. coli* BL21 (DE3) for protein overexpression. *E. coli* BL21 (DE3) containing certain plasmids were inoculated into a 10 mL culture of LB supplemented with appropriate antibiotics at 37 °C overnight. 1 ml culture broth was transferred to 100 ml LB with appropriate antibiotics and grown at 37 °C to OD600 ~0.6-0.8, then cooled to 18 °C before IPTG was added to a final concentration of 0.2 mM. The culture was incubated at 18°C for an additional 16 h. Cells were harvested by centrifugation at 5,000 ×g for 10 min at 4 °C. Cell pellets were resuspended in 30 mL of binding buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 30 mM imidazole) and lysed using a high-pressure homogenizer NanoGenizer (Genizer LLC, Irvine, CA, USA). Cell debris was removed by centrifugation at 20,000 ×g for 60 min at 4 °C. The supernatant was loaded onto a 100 ul His SpinTrapTM column (GE Healthcare). The column was washed with 600 uL of binding buffer I (30 mM imidazole, 20 mM Tris, pH 8.0, 300 mM NaCl). The protein was eluted twice with elution buffer (250 mM imidazole, 20 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol) and analyzed by SDS-PAGE. Fractions containing target proteins were combined and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (30kDa, MWCO, Millipore). The resulting protein sample was stored at −70 °C with 15% glycerol.

Table S1. Plasmids and strains used in this study

Plasmids or strains	Description	Reference or source
Plasmids		
pHis8	Contain 8 his tag at N-terminal; T7 promoter, lac operator	[1]
pLLH105	pHis8 derivative that carries *cfat1* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH106	pHis8 derivative that carries *cfat2* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH107	pHis8 derivative that carries *cfat3* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH108	pHis8 derivative that carries *cfat-DSM 2064* gene from *Flavobacterium johnsoniae* DSM 2064	This study
pLLH109	pHis8 derivative that carries *cfat-DSM 25383* gene from *Alistipes timonensis* DSM 25383	This study
pLLH110	pHis8 derivative that carries *cfat-JUb44* gene from *Chryseobacterium scophthalmum* JUb44	This study
pLLH111	pHis8 derivative that carries *cfat-DSM 24695* gene from *Algoriphagus machipongonensis* DSM 24695	This study
pACYCDuet-1	The vector encodes two multiple cloning sites (MCS) each of which is preceded by a T7 promoter, lac operator, and ribosome binding site (rbs)	Novagen
pLLH112	pACYCDuet-1 derivative that carries *cys1* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH113	pACYCDuet-1 derivative that carries *cys2* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH114	pACYCDuet-1 derivative that carries *cys3* gene from *Chryseobacterium gleum* DSM 16776	This study
pLLH115	pACYCDuet-1 derivative that carries *cys -DSM 2064* gene from *Flavobacterium johnsonie DSM 2064*	This study
pLLH116	pACYCDuet-1 derivative that carries *cys-DSM 25383* gene from *Alistipes timonensis DSM 25383*	This study
pLLH117	pACYCDuet-1 derivative that carries *cys-JUb44* gene from *Chryseobacterium scophthalmum JUb44*	This study
pLLH118	pACYCDuet-1 derivative that carries *cys*- DSM 24695 gene from *Algoriphagus machipongonensis* DSM 24695	This study
Strains		
E. coli DH5*a*	Host strain for general cloning	Stratagene
E. coli BL21 (DE3)	Host strain for overexpression	Invitrogen
C. gleum DSM 16776	Anaerobe, mesophilic bacterium that was isolated from human, high vaginal swab.	DSMZ[a]
F. johnsonie DSM 2064	An obligate aerobe, mesophilic, gram-negative animal pathogen that forms circular colonies and was isolated	DSMZ[a]

	from Soil.	
A. timonensis DSM 25383	An anaerobe, mesophilic bacterium that was isolated from human fecal flora of a healthy patient.	DSMZ[a]
C. scophthalmum JUb44	*Caenorhabditis elegans* associated bacteria.	CGC[b]
A. machipongonensis DSM 24695	*Choanoflagellates* associated bacteria	DSMZ[a]
E. coli BL21 (DE3)/pACYCDuet-1	*E. coli* BL21 (DE3) carrying pACYCDuet-1	This study
E. coli BL21 (DE3)/pHis8	*E. coli* BL21 (DE3) carrying pHis8	This study
E. coli BL21 (DE3)/pACYCDuet-1+ pHis8	*E. coli* BL21 (DE3) carrying pACYCDuet-1 and pHis8	This study
E. coli BL21 (DE3)/pLLH105	*E. coli* BL21 (DE3) carrying pLLH105	This study
E. coli BL21 (DE3)/pLLH106	*E. coli* BL21 (DE3) carrying pLLH106	This study
E. coli BL21 (DE3)/pLLH107	*E. coli* BL21 (DE3) carrying pLLH107	This study
E. coli BL21 (DE3)/pLLH108	*E. coli* BL21 (DE3) carrying pLLH108	This study
E. coli BL21 (DE3)/pLLH109	*E. coli* BL21 (DE3) carrying pLLH109	This study
E. coli BL21 (DE3)/pLLH110	*E. coli* BL21 (DE3) carrying pLLH110	This study
E. coli BL21 (DE3)/pLLH111	*E. coli* BL21 (DE3) carrying pLLH111	This study
E. coli BL21 (DE3)/pLLH112	*E. coli* BL21 (DE3) carrying pLLH112	This study
E. coli BL21 (DE3)/pLLH115	*E. coli* BL21 (DE3) carrying pLLH115	This study
E. coli BL21 (DE3)/pLLH116	*E. coli* BL21 (DE3) carrying pLLH116	This study
E. coli BL21 (DE3)/pLLH117	*E. coli* BL21 (DE3) carrying pLLH117	This study
E. coli BL21 (DE3)/pLLH118	*E. coli* BL21 (DE3) carrying pLLH118	This study
E. coli BL21 (DE3)/pLLH105+ pLLH112	*E. coli* BL21 (DE3) carrying pLLH105 and pLLH112	This study
E. coli BL21 (DE3)/pLLH108+ pLLH115	*E. coli* BL21 (DE3) carrying pLLH108 and pLLH115	This study
E. coli BL21 (DE3)/pLLH109+ pLLH116	*E. coli* BL21 (DE3) carrying pLLH109 and pLLH116	This study
E. coli BL21 (DE3)/pLLH110+ pLLH117	*E. coli* BL21 (DE3) carrying pLLH110 and pLLH117	This study
E. coli BL21 (DE3)/pLLH111+ pLLH118	*E. coli* BL21 (DE3) carrying pLLH111 and pLLH118	This study

a: Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures GmbH)

b: Caenorhabditis Genetics Center

Table S2. Primer pairs used in this study

Primer name	Sequence (5'-3')
pHis8-*cfat1*-Fwd	cacggtggtctggttccgcgtggttccATGTTGGATATTTTTGAAAGAA
pHis8-*cfat1*-Rev	cgcaagcttgtcgacggagctcgaattcTTAAATCGGTTTAAAACTTAATC
pHis8-*cfat2*-Fwd	cacggtggtctggttccgcgtggttccATGATCTCTGAAAAATACCTTC
pHis8-*cfat*2-Rev	cgcaagcttgtcgacggagctcgaattcTTAAGAGATCACTCCTAATTC
pHis8-*cfat3*-Fwd	cacggtggtctggttccgcgtggttccATGAGCATCAATTTCACAAC
pHis8-*cfat3*-Rev	cgcaagcttgtcgacggagctcgaattcTAGGCATAATTATTATTTTTTCGC
pHis8-*cfat-DSM 2064*-Fwd	gtctggttccgcgtggttcccatggcggatccatggtaaaagatttattcgaaagaatt
pHis8-*cfat-DSM 2064*-Rev	tggtggtggtgctcgagtgcggccgcaagctttacgaaacgtctactgttgt
pHis8-*cfat-DSM 25383*-Fwd	gtctggttccgcgtggttcccatggcggatccatggttgatatttttcacgcc
pHis8-*cfat-DSM 25383*-Rev	tggtggtggtgctcgagtgcggccgcaagctttcaacgtaccttaaaaccctc
pHis8-*cfat-JUb44*-Fwd	gtctggttccgcgtggttcccatggcggatccatgttggatattttgaaagaataaa
pHis8-*cfat- JUb44*-Rev	tggtggtggtgctcgagtgcggccgcaagctttaaatttctttgaaacttaaacct
PacycDuet-1-*cys1*-Fwd	cacggtggtctggttccgcgtggttccATGAGTAATGTTTACGATAATATC
pACYCDuet-1-*cys1*-Rev	cgcaagcttgtcgacggagctcgaattcTTATTTGATAAACTCCGTTTTG
pACYCDuet-1-*cys2*-Fwd	cacggtggtctggttccgcgtggttccATGAAATACGCAAACAATATCC
pACYCDuet-1-*cys2*-Rev	cgcaagcttgtcgacggagctcgaattcCTATTTACTTTCTTCCAGCC
pACYCDuet-1-*cys3*-Fwd	cacggtggtctggttccgcgtggttccATGAAATTTCAGAATGCATTAG
pACYCDuet-1-*cys3*-Rev	cgcaagcttgtcgacggagctcgaattcTTAGAAGAGTCCTTCTACAG
pACYCDuet-1-*cys* -DSM 2064-Fwd	agcagccatcaccatcatcaccacagccagatgaaagaagaaataaacgcttataata
pACYCDuet-1-*cys* -DSM 2064-Rev	ttcgacttaagcattatgcggccgcaagcttattttacaaattcaatttttgagct
pACYCDuet-1-*cys* -DSM 25383-Fwd	agcagccatcaccatcatcaccacagccagatgaaaaaaattgcaaattcagcac
pACYCDuet-1-*cys* -DSM 25383-Rev	ttcgacttaagcattatgcggccgcaagcttcataggcggatgtcggtttt
pACYCDuet-1-*cys* -JUb44-Fwd	agcagccatcaccatcatcaccacagccagatgagtaatgtttacgataatattcttggc
pACYCDuet-1-*cys* -JUb44-Rev	ttcgacttaagcattatgcggccgcaagcttatttgatgatttctgtcttaaagacttc
pACYCDuet-1-*cys* – DSM 24695-Fwd	agcagccatcaccatcatcaccacagccagatgatctataattccatcattgataccat
pACYCDuet-1-*cys* – DSM 24695-Rev	ttcgacttaagcattatgcggccgcaagcttacttggtaaaggcatctagaatatc

Table S3. ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) of SoL A.

No	In CD_3OD		In DMSO-d_6		In DMSO-d_6 (ref 21)	
	δ_C	δ_H	δ_C	δ_H	δ_C	δ_H
1	51.5	3.07 (1H, dd, *J* = 14.4, 8.5 Hz) 3.11 (1H, dd, *J* = 14.4, 3.3 Hz)	51.7	2.66 (1H, dd, *J* = 14.0, 3.7 Hz) 2.74 (1H, dd, *J* = 14.1, 6.6 Hz)	51.8	2.73 (2H, d, *J* = 8.3 Hz)
2	53.0	4.23 (1H, m)	51.0	3.88 (1H, m)	51.1	3.92 (1H, m)
3	74.0	3.69 (1H, m)	71.9	3.46 (1H, m)	72.0	3.46 (1H, m)
4	34.6	1.42 (2H, m)	33.3		33.4	
5			25.5		25.5	
6						
7						
8						
9		1. 16-1.46	29.1-29.3	1.16-1.42	29.2-29.4	1.22
10	26.7-31.1					
11						
12						
13			26.8		26.9	
14	40.3	1.18	38.5	1.13	38.5	1.14
15	29.2	1.54	27.4	1.49	27.4	1.49
16	23.1	0.88 (3H, d, *J* = 6.63 Hz)	22.5	0.84 (3H, d, *J* = 6.6 Hz)	22.6	0.84 (3H, d, *J* = 6.8 Hz)
17			22.5		22.6	
2-NH				7.61(1H, d, *J* = 8.4 Hz)		7.68 (1H, d, *J* = 6.8 Hz)
3-OH				4.77 (1H, d, *J* = 5.4 Hz)		4.80 (1H, d, *J* = 8.3 Hz)
1′	173.8		169.9		170.2	-
2′	45.5	2.34(2H, m)	44.8	2.10 (2H, m)	44.8	2.11 (1H, dd, *J* = 10.8, 5.9 Hz) 2.13 (1H, dd, *J* = 10.8,5.4 Hz)
3′	69.8	3.97 (1H, m)	67.5	3.74 (1H, m)	67.6	3.76 (1H, m)
4′	38.1	1.46 (2H, m)	36.6		36.6	1.37 (1H, m)
5′	26.7-29.2		25.2		25.2	

6′						
7′		1.16-1.46		1.16-1.42		1.22
8′						
9′			29.1-29.3		29.2-29.4	
10′						
11′						
12′						
13′			26.8		26.9	
14′	40.3	1.18	38.5	1.13	38.5	1.14
15′	29.2	1.53	27.4	1.49	27.4	1.49
16′	23.1	0.88	22.5	0.84 (3H, d, J = 6.6 Hz)	22.6	0.84 (3H, d, J = 6.8 Hz)
17′			22.5		22.6	
3′-OH				4.68 (1H, d, J = 4.1 Hz)		4.66 (1H, d, J = 4.4 Hz)

Table S4. MS^2 data of the main SoLs in *C. gleum* DSM 16776

m/z $[M-H]^-$	Retention time (min)	MS/MS data
634.4743	13.725	79.96, 80.97, 349.21, 366.23, 408.24, 634.47
590.4474	13.806	79.96, 80.97, 319.19, 333.21, 336.22, 350.24,378.23, 392.25, 590.45
576.4321	13.899	79.96, 80.97, 319.20, 333.21, 336.22, 350.24, 364.22, 576.43
620.4585	13.967	79.96, 80.97, 333.21, 350.24, 378.23, 408.24, 620.46
616.4633	14.208	79.96, 80.97, 333.21, 350.24, 362.24 374.24, 392.25, 598.45, 616.46
560.4371	14.275	79.96, 80.97, 319.20 333.21, 350.24, 560.44
604.4633	14.380	79.96, 80.97, 333.21, 350.24, 392.25, 604.46
590.4478	14.470	79.96, 80.97,211.21,257.21, 333.21, 350.24, 378.23, 590.45
634.4742	14.534	79.96, 80.97, 333.21, 350.24, 378.23, 408.24, 634.47
574.4527	14.834	79.96, 80.97, 333.21, 350.24, 574.45
618.4790	14.937	79.96, 80.97, 94.98, 333.21, 350.24, 392.25, 600.47, 618.48
604.4639	15.024	79.96, 80.97, 317.22, 333.21, 350.24, 604.46
644.4950	15.240	79.96, 80.97, 350.24, 359.23, 376.25, 418.26, 644.50
588.4685	15.378	79.96, 80.97, 333.21, 347.23, 350.24, 588.47
632.4949	15.459	79.96, 80.97, 347.23, 350.24, 364.25, 406.26, 614.49, 632.49
632.4950	15.607	79.96, 80.97, 347.23, 350.24, 364.25, 406.26, 614.49, 632.50
646.5104	15.954	79.96, 80.97, 361.24, 378.27, 420.28, 392.25, 646.51

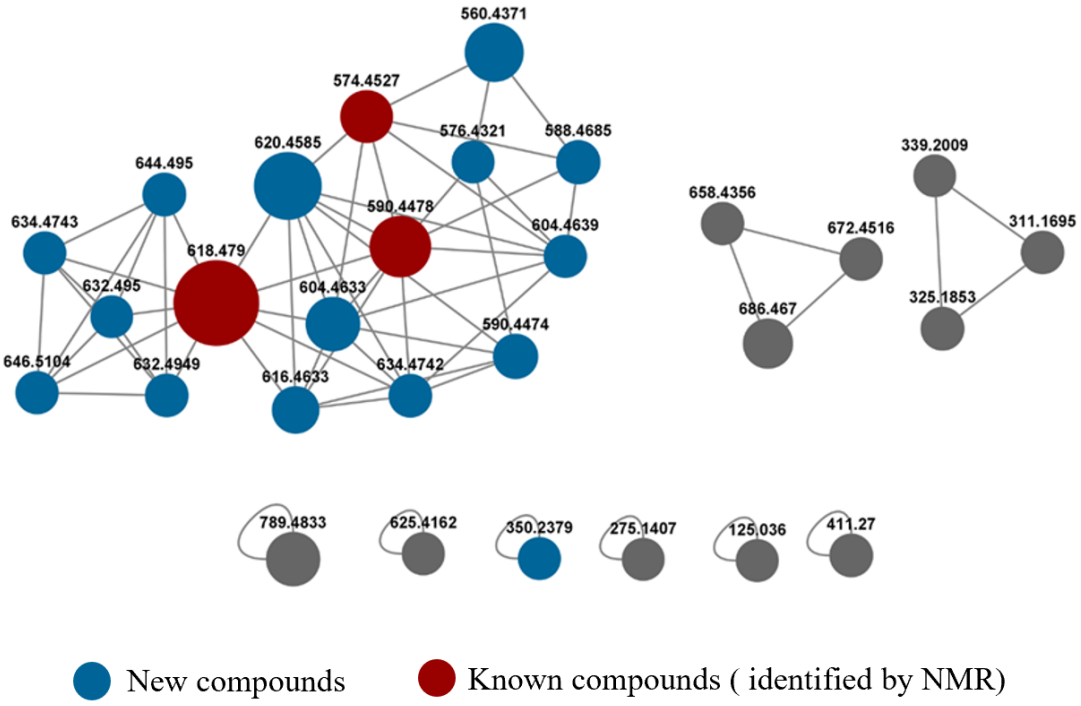

Figure S1. Feature-based molecular networking (FBMN) of the active fraction of *C. gleum* strain. The raw data was filtered by python script by fragment ion at *m/z* 79.957.

Figure S2. Node compound at *m/z* 618.48 in negative mode identified by SIRIUS 4

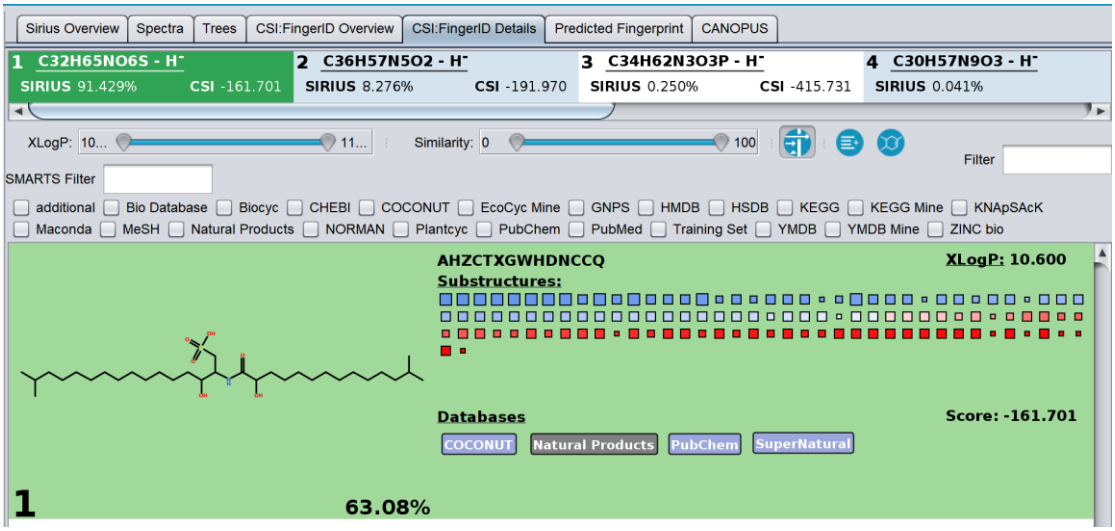

Figure S3. Node compound at *m/z* 590.45 in negative mode identified by SIRIUS 4

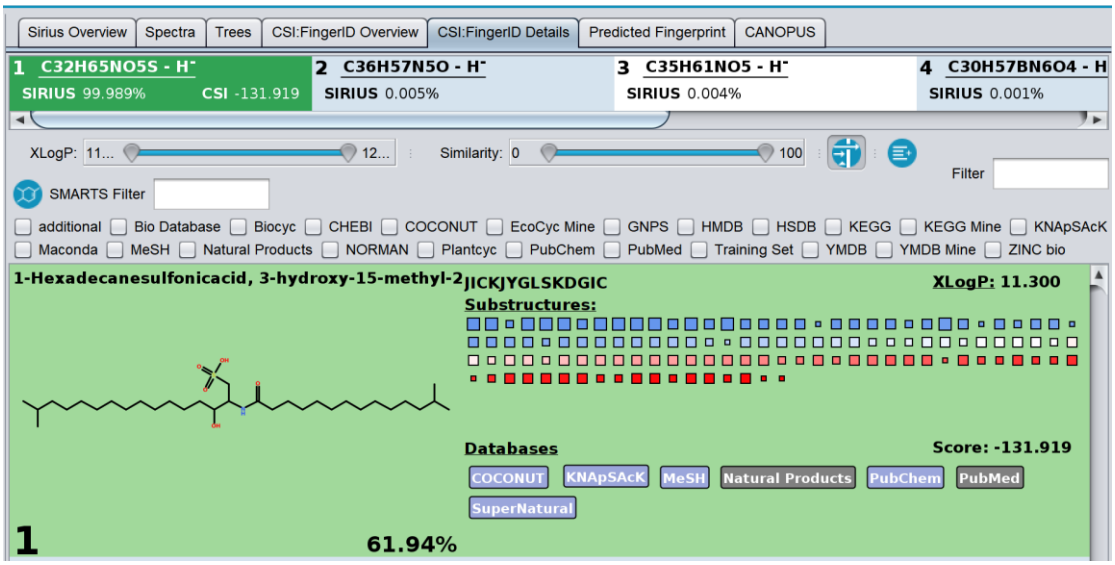

Figure S4. Node compound at *m/z* 574.44 in negative mode identified by SIRIUS **4**

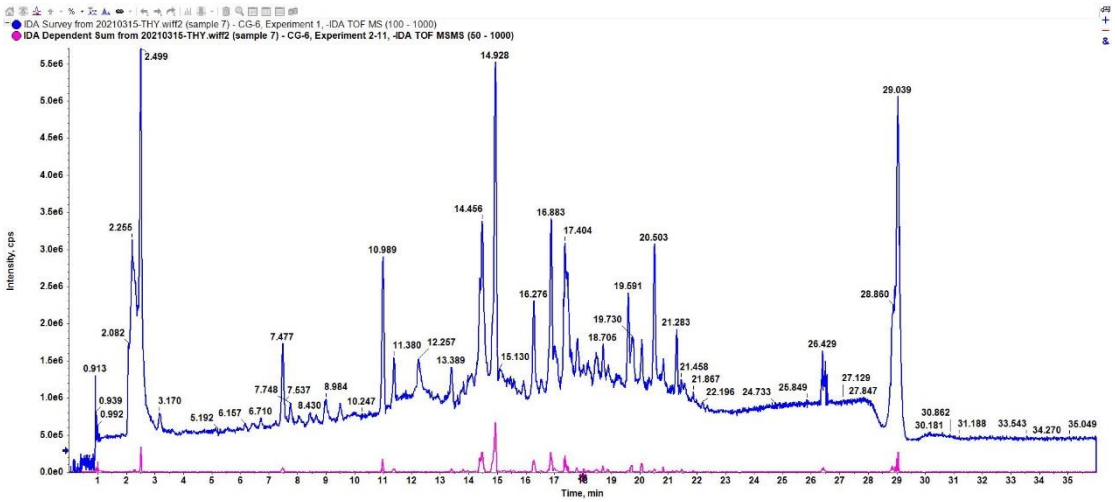

Figure S5. TIC of *C. gleum* strain extract in negative mode.

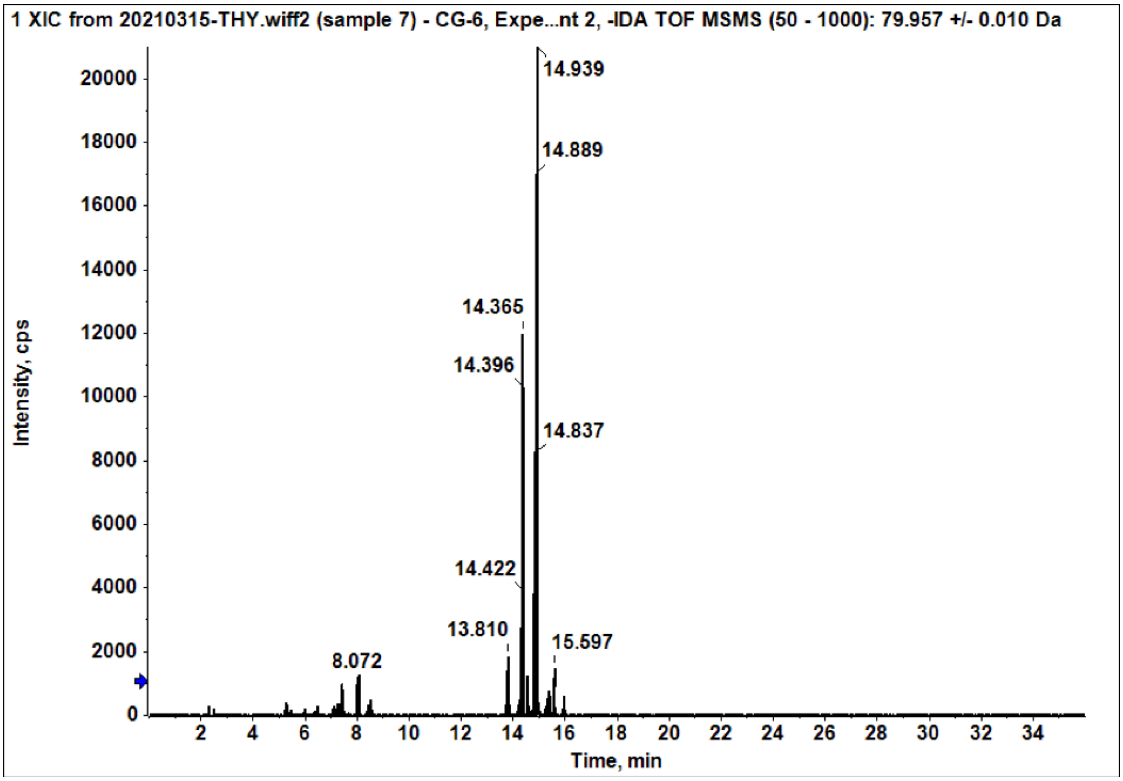

Figure S6. Fragment screening of MS^2 data by m/z 79.957(SO_3^-).

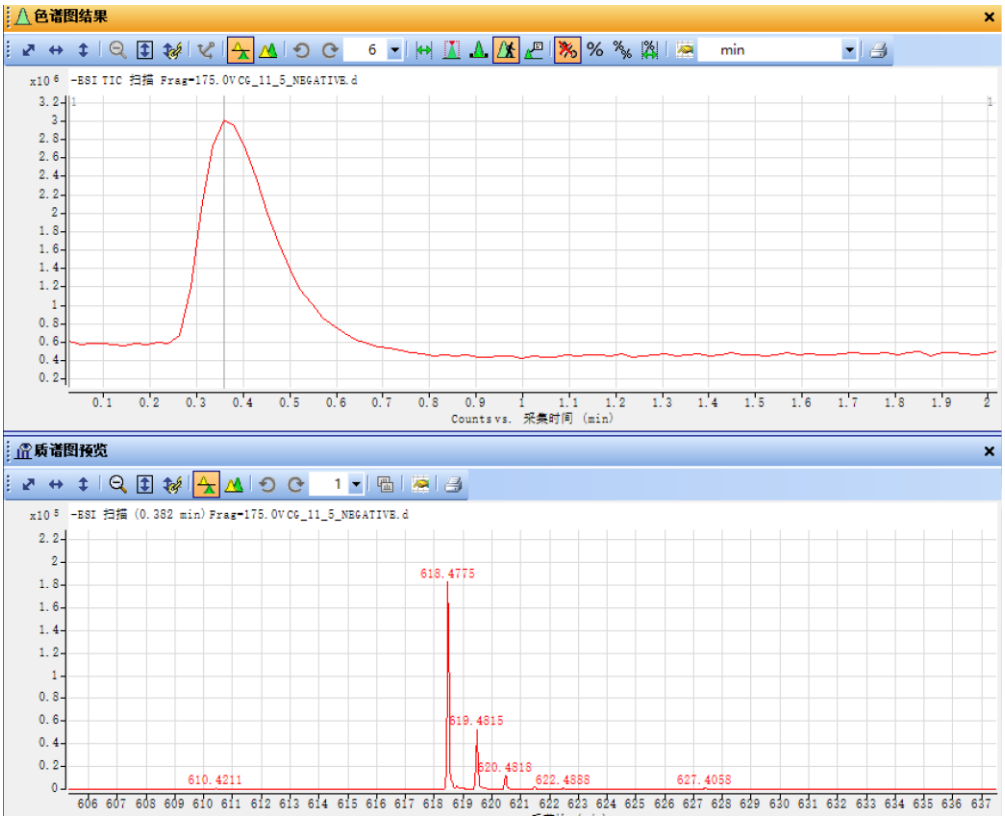

Figure S7. HR-ESIMS spectrum of SoL A.

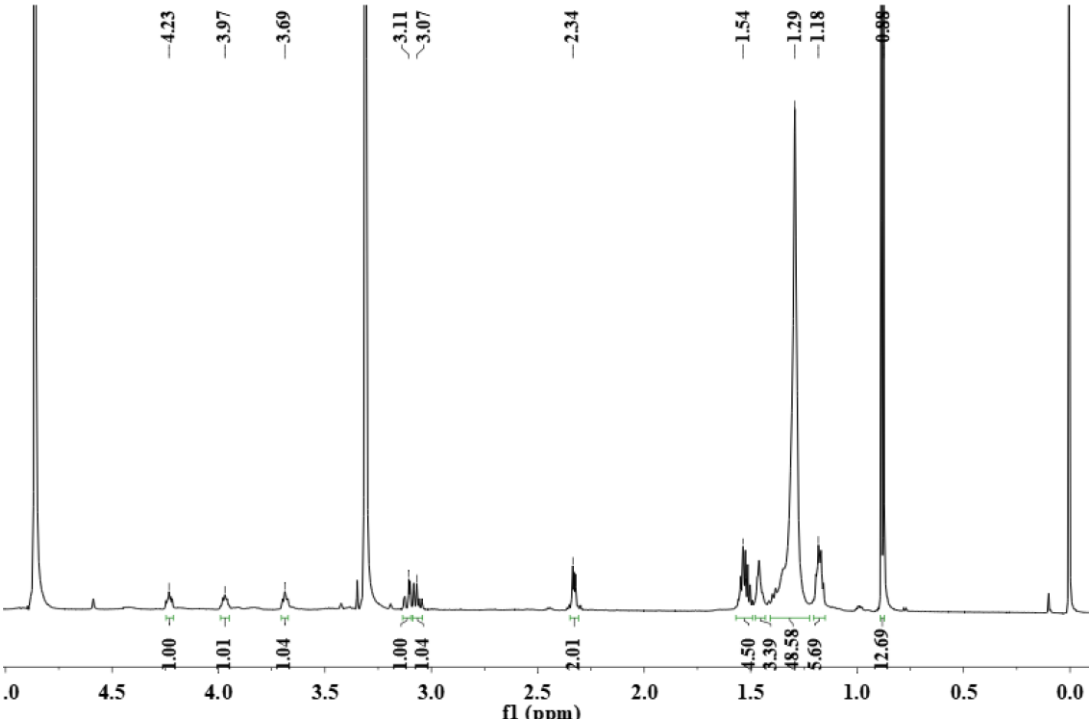

Figure S8. ^{1}H NMR spectrum of SoL A in CD_3OD at 600 MHz.

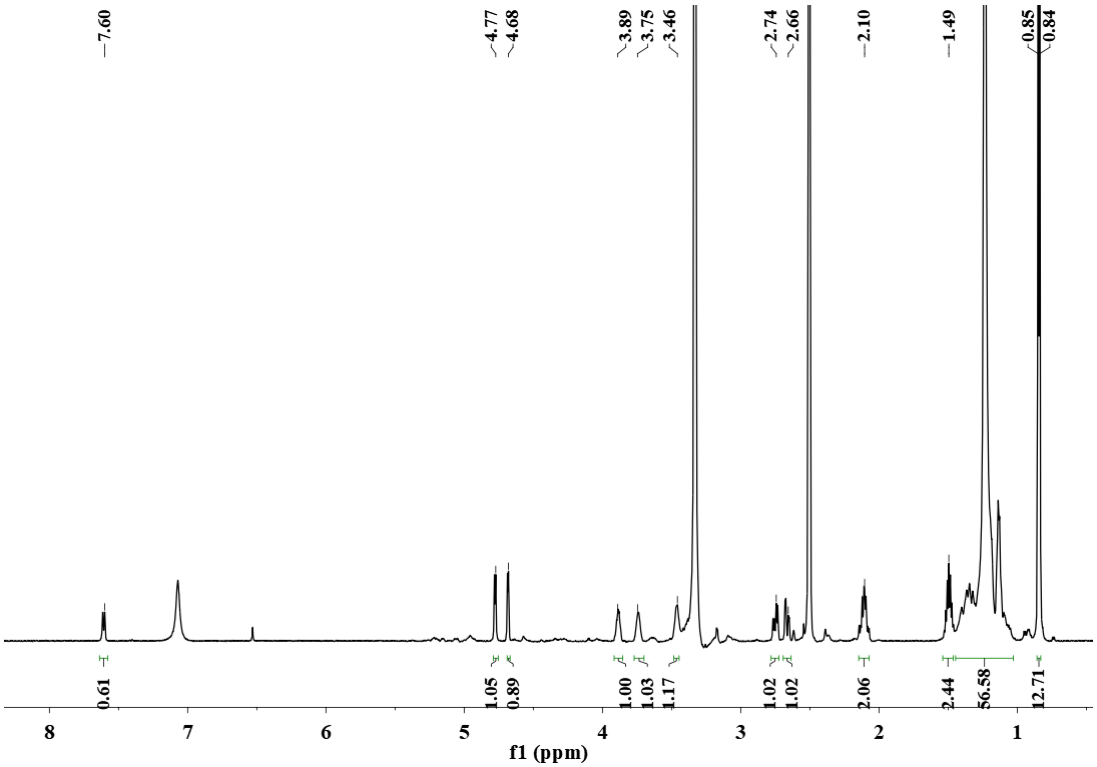

Figure S8a. ^{1}H NMR spectrum of SoL A in DMSO-d_6 at 600 MHz.

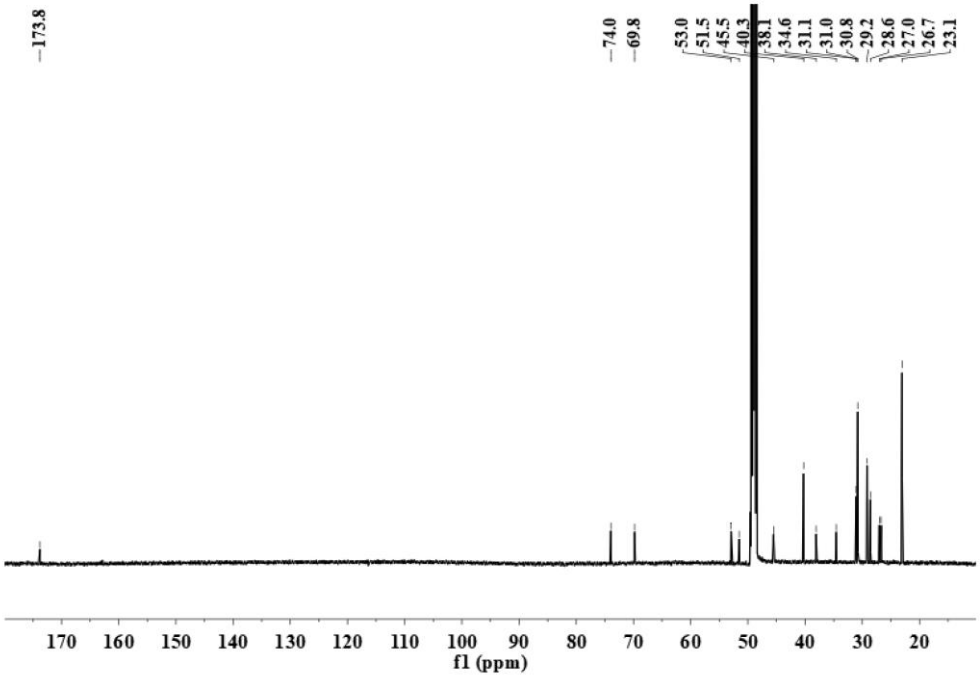

Figure S9. ^{13}C NMR spectrum of SoL A in CD_3OD at 150 MHz.

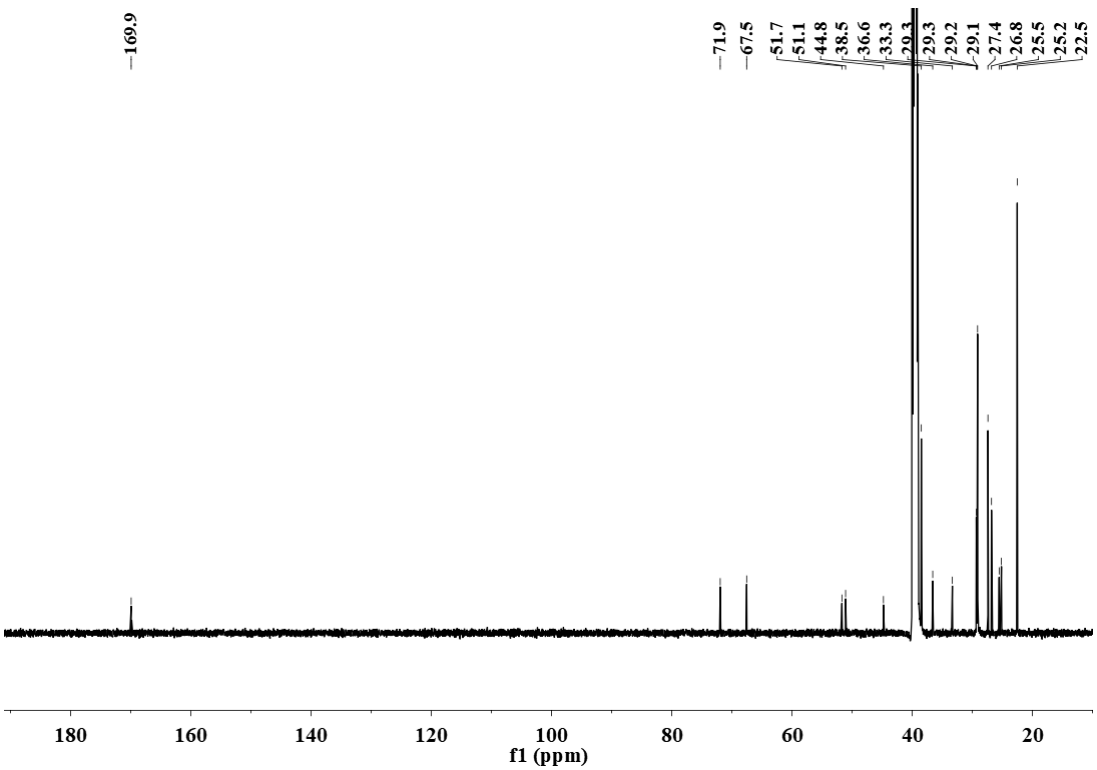

Figure S9a. ^{13}C NMR spectrum of SoL A in DMSO-d_6 at 150 MHz.

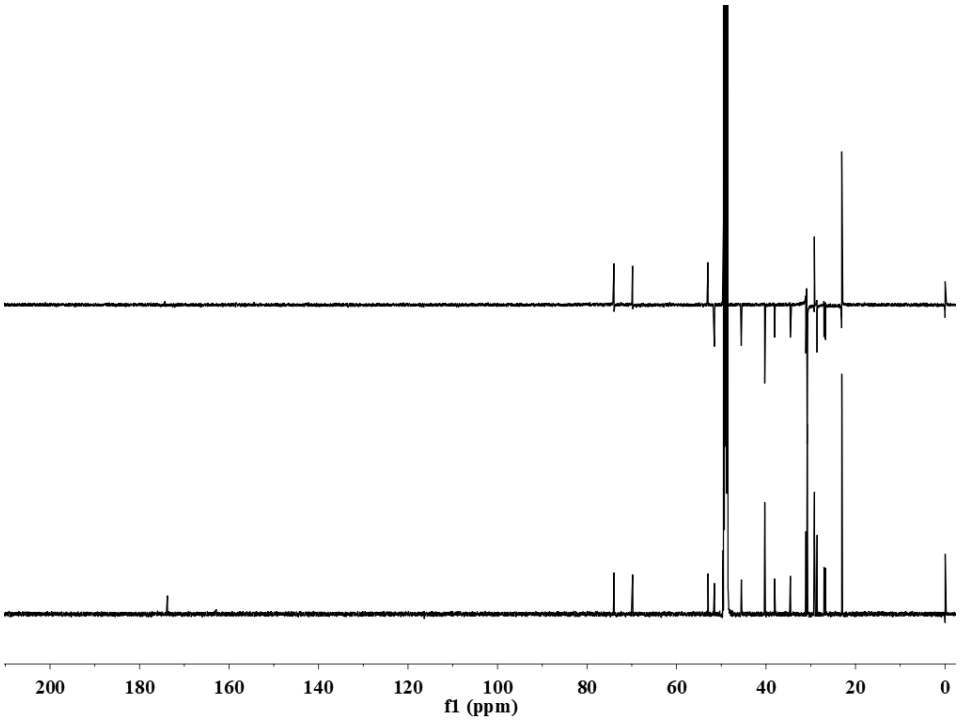

Figure S10. DEPT 135 and ^{13}C NMR spectra of SoL A in CD_3OD.

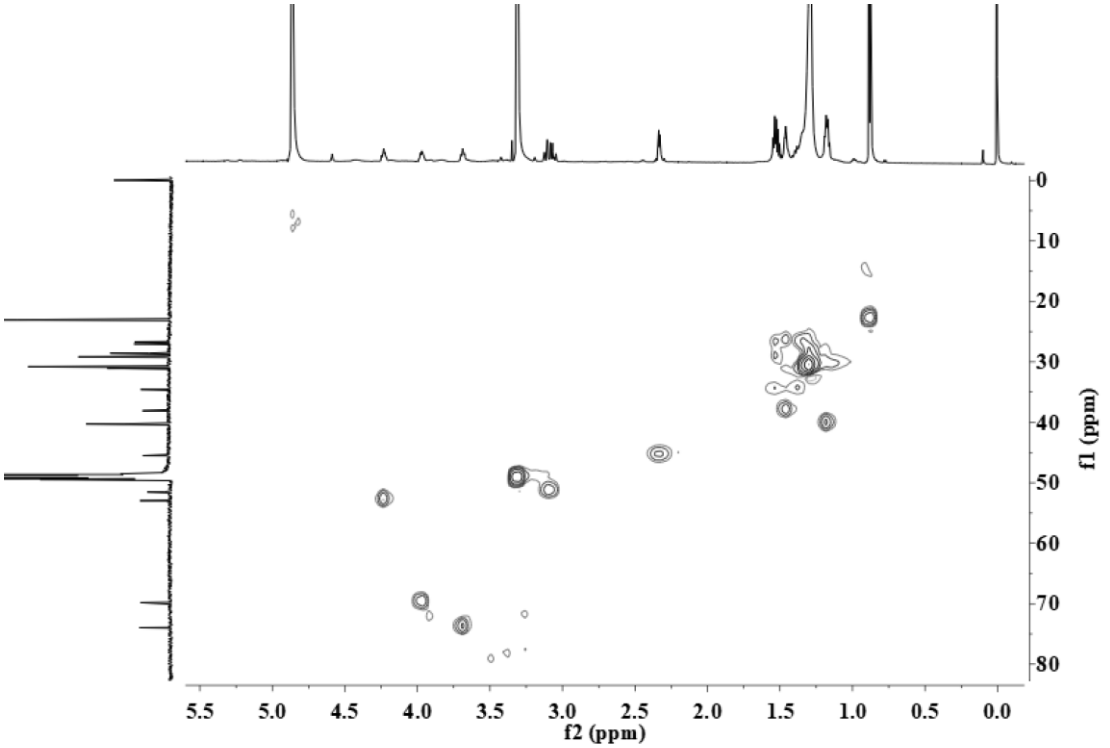

Figure S11. HSQC spectrum of SoL A in CD_3OD**.**

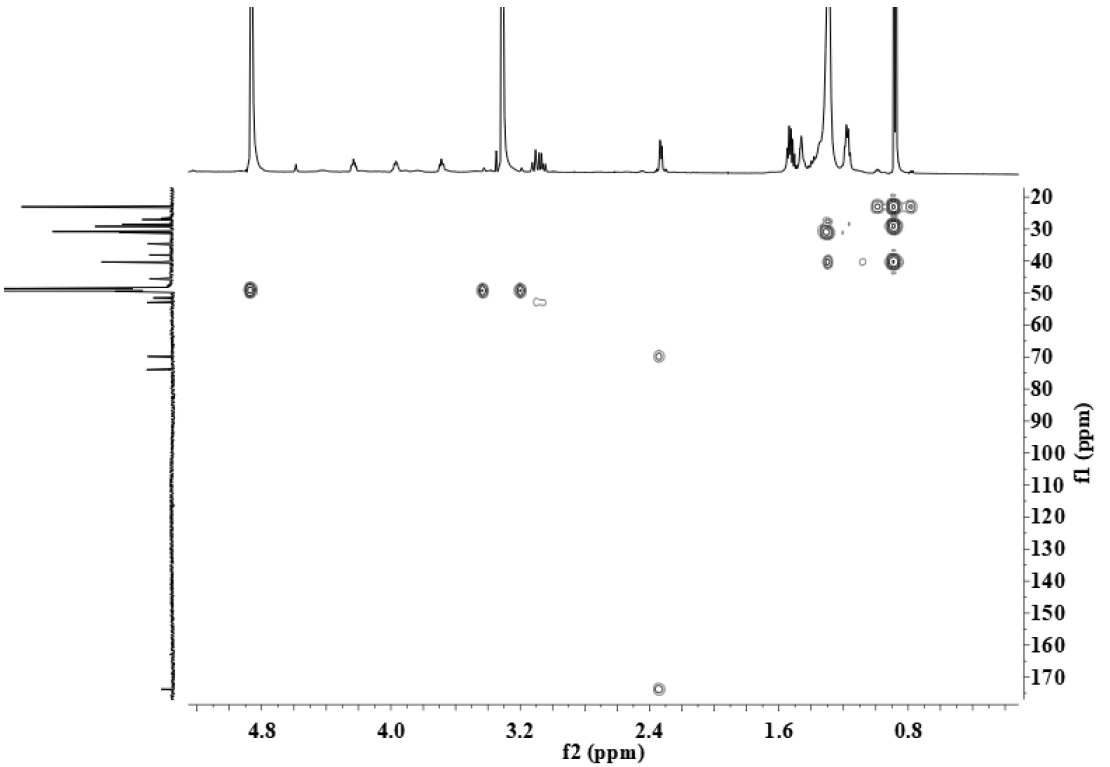

Figure S12. HMBC spectrum of SoL A in CD_3OD.

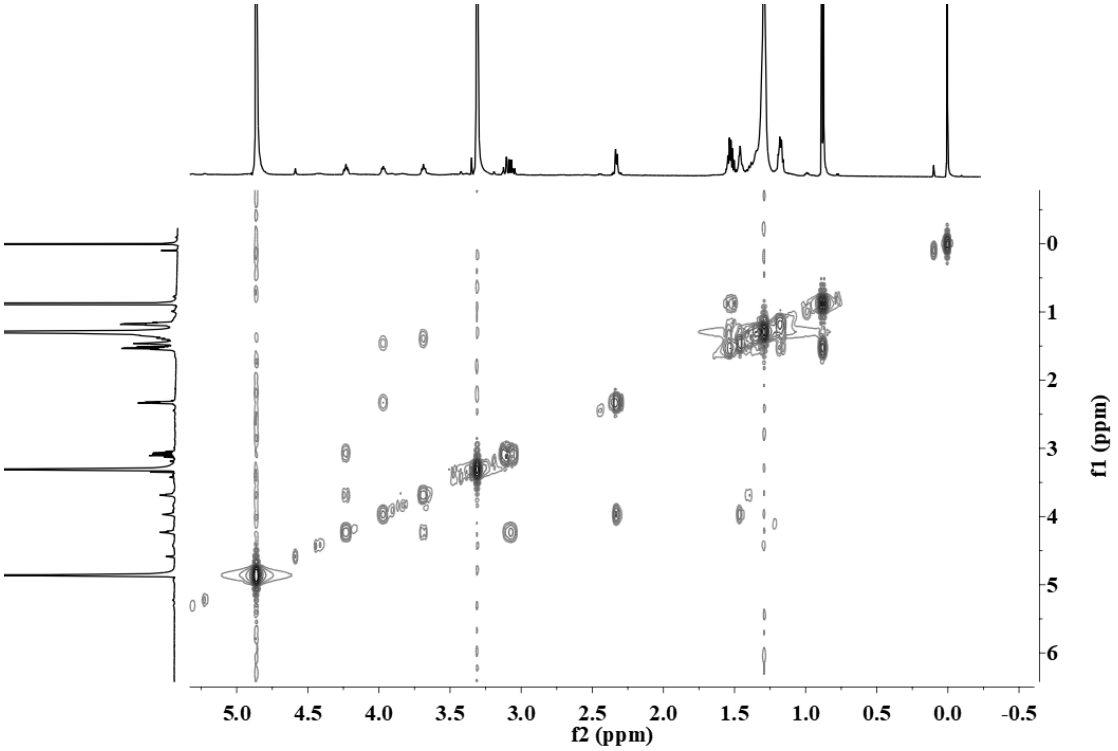

Figure S13. ^{1}H-^{1}H COSY spectrum of SoL A in CD_3OD.

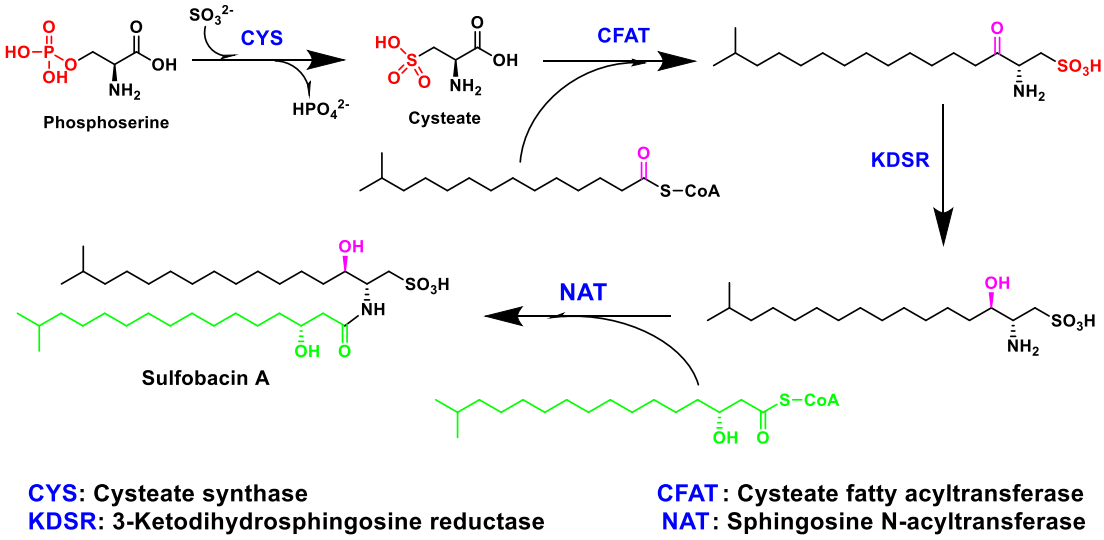

Figure S14. Proposed biosynthesis pathway of SoL A.

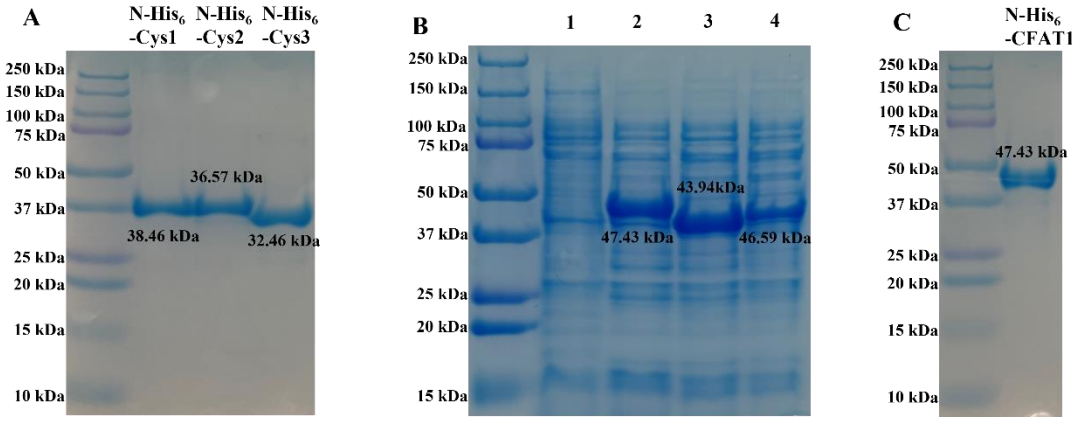

Figure S15. SDS-PAGE of the heterogeneously expressed proteins Cys and CFAT from *C. gleum* DSM 16776. (**A**) Cys1 (38.46 kDa), Cys2 (36.57 kDa), and Cys3 (32.46 kDa) were purified from *E. coli* BL21(DE3) with N-His6 tag. (**B**) Lane 2–4: Crude enzyme CFAT1 (47.43 kDa), CFAT2 (43.94 kDa), and CFAT3 (46.59 kDa) from *E. coli* BL21(DE3) with N-His_6 tag. Lane 1: *E. coli* BL21(DE3) contains a blank vector. (C) CFAT1 (47.43 kDa) purified from *E. coli* BL21 with N-His_6 tag.

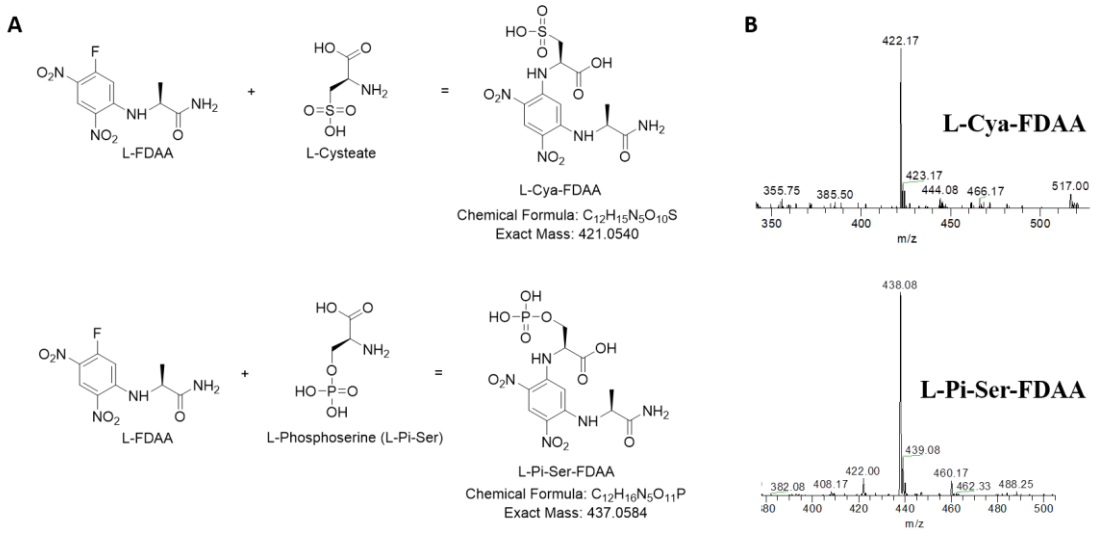

Figure S16. L-FDAA amino acid derivatization by Marfey's method. (**A**) L-FDAA amino acid derivatives were conducted by incubating L-FDAA with L-cysteate and L-phosphoserine. (**B**) The ESI-MS spectra of L-cysteate and L-phosphoserine FDAA derivatives standard.

```
Q93UV0  AKKHGAMVLVDEAHSMGFFGPNGRGVYEAQGLEGQIDFVVGTFSKSVG-TVGGFVVSNHP  279
CFAT1   KSKYQFRLLVDDAHGFGTLGKTGAGVGEEQDCNDQIDVYFSTFAKSMA-GFGAFLAGDKE  264
CFAT2   ADKYDALVMVDDSHATGFIGKTGRGTHEANEVMGRVDIITSTLGKALGGALGGFTSGKKE  259
CFAT3   VKKYNAFLMVDDVHGVGILGETGRGTLEQAGLLDKVDLITGTFSKTFG-NLGGYVIADKK  279
          .*:   ::**: *. * :* .* *. *     .::*.  .*:.*:..  .*.:  ..:
```

Figure S17. Partial sequence alignment of Q93UV0 (SpSpt) with CFAT1–3 from *C. gleum* DSM 16776, highlighting the conserved lysine.

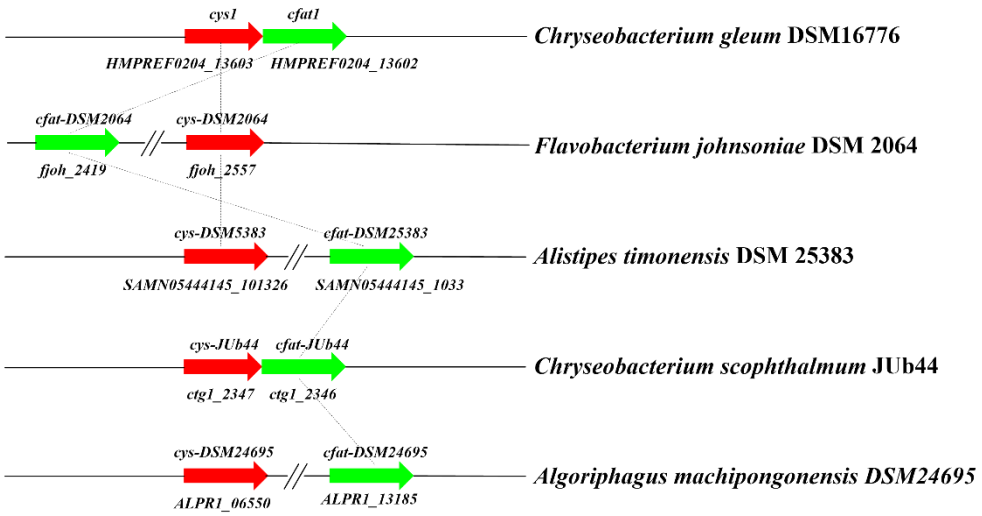

Figure S18. Different bacterial genomes containing genes *cys* (red) and *cfat* (green).

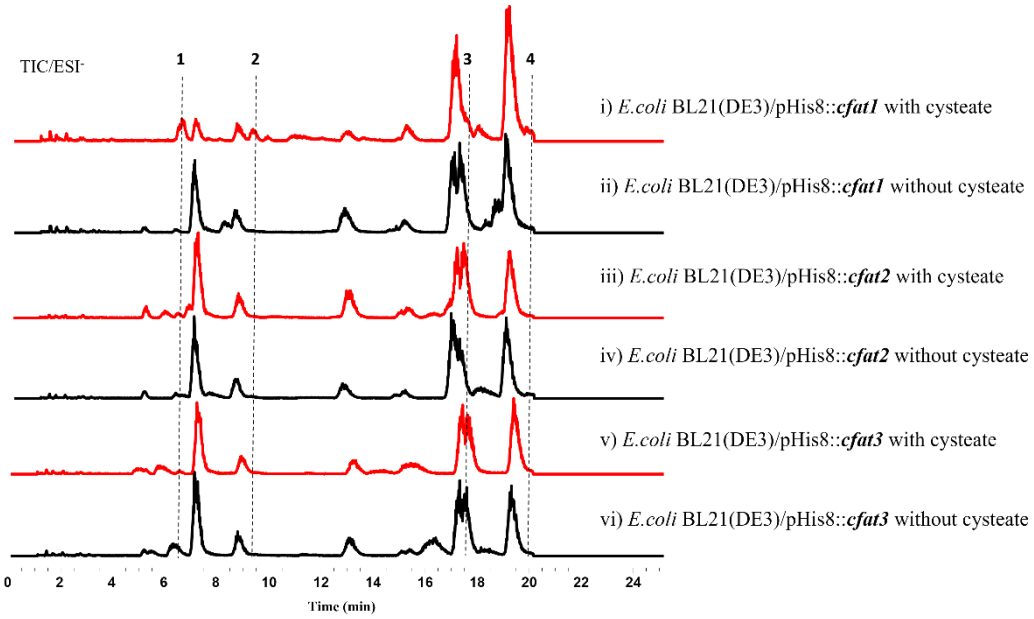

Figure S19. Comparative LC-ESI analysis production of *E. coli* BL21(DE3)/pHis8::*cfat1–3* feeding with cysteate or not. (**i**) *E.coli* BL21(DE3)/pHis8::*cfat1* feeding with cysteate, (**ii**) *E. coli* BL21(DE3)/pHis8::*cfat1* not feeding with cysteate, (**iii**) *E. coli* BL21(DE3)/pHis8::*cfat2* feeding with cysteate, (**iv**) *E. coli* BL21(DE3)/pHis8::*cfat2* not feeding with cysteate, and (**v**) *E. coli* BL21(DE3)/pHis8::*cfat3* feeding with cysteate and (**vi**) *E.coli* BL21(DE3)/pHis8::*cfat3* not feeding with cysteate.

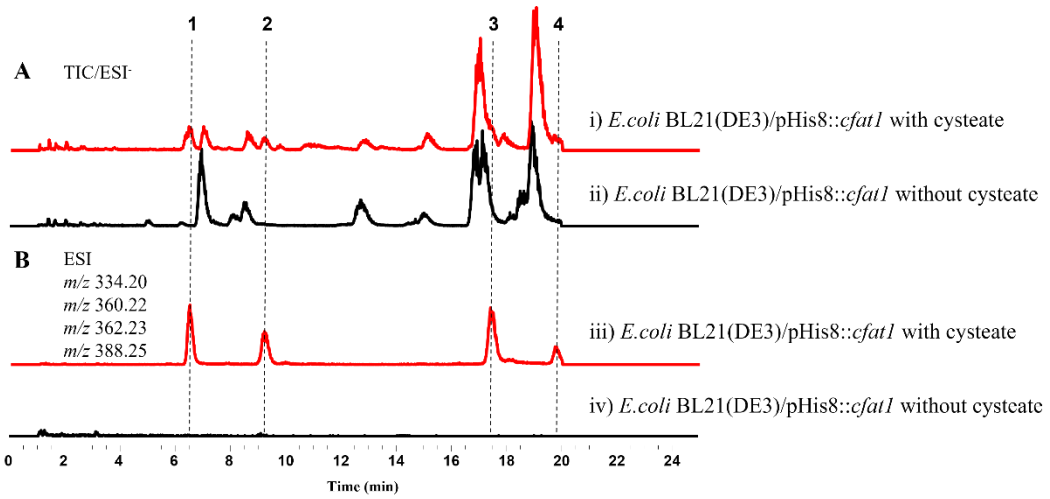

Figure S20. Comparative LC-ESI analysis production of *E. coli* BL21(DE3)/pHis8::*cfat1* feeding with cysteate or not. (**A**) Total Ion Chromatograms (TICs) of reactions, (**B**) The extracted ion chromatograms (EICs) were extracted at *m/z* 334.20 $[M-H]^-$ for **1**, *m/z* 360.22 $[M-H]^-$ for 2, *m/z* 362.23 $[M-H]^-$ for **3** and *m/z* 388.25 $[M-H]^-$ for **4**. (i and iii) *E.coli* BL21(DE3)/pHis8::*cfat1* feeding with cysteate, (**ii** and **iv**) *E. coli* BL21(DE3)/pHis8::*cfat1* not feeding with cysteate.

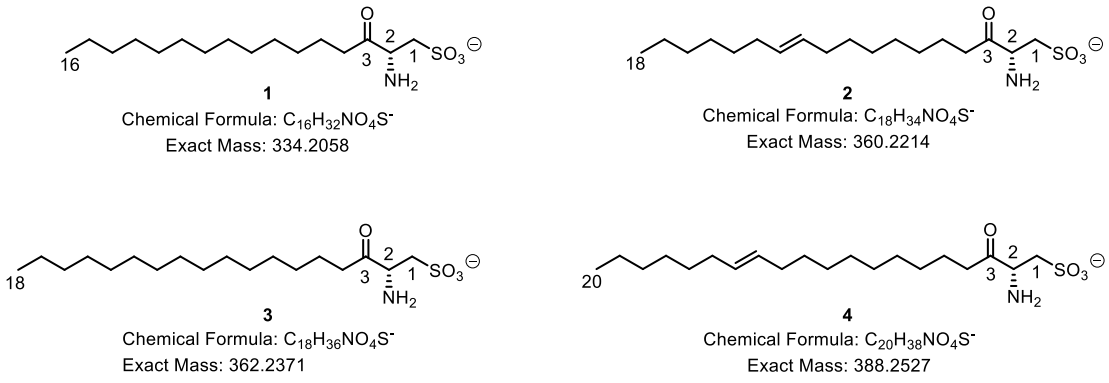

Figure S21. Possible capnine-like analogs produced by *E. coli* BL21 (DE3)/pLLH105 fed with cysteate.

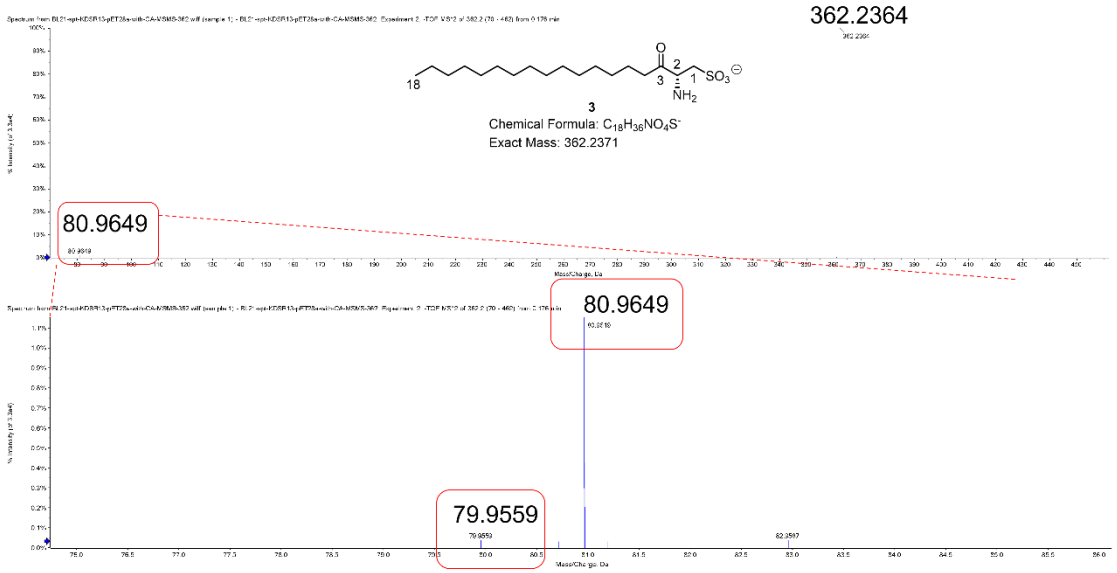

Figure S22. MS-MS spectrum of compound **3**.

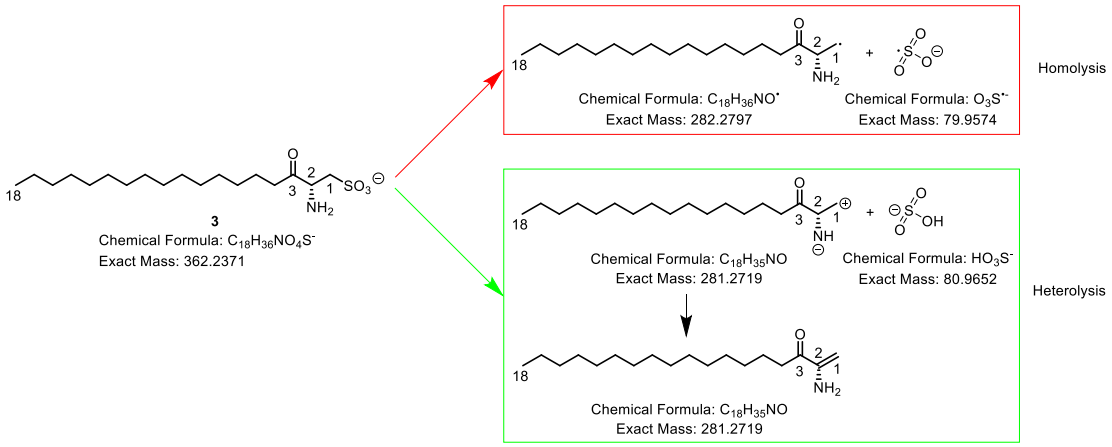

Figure S23. Homolytic manner or heterolytic manner of compound **3.**

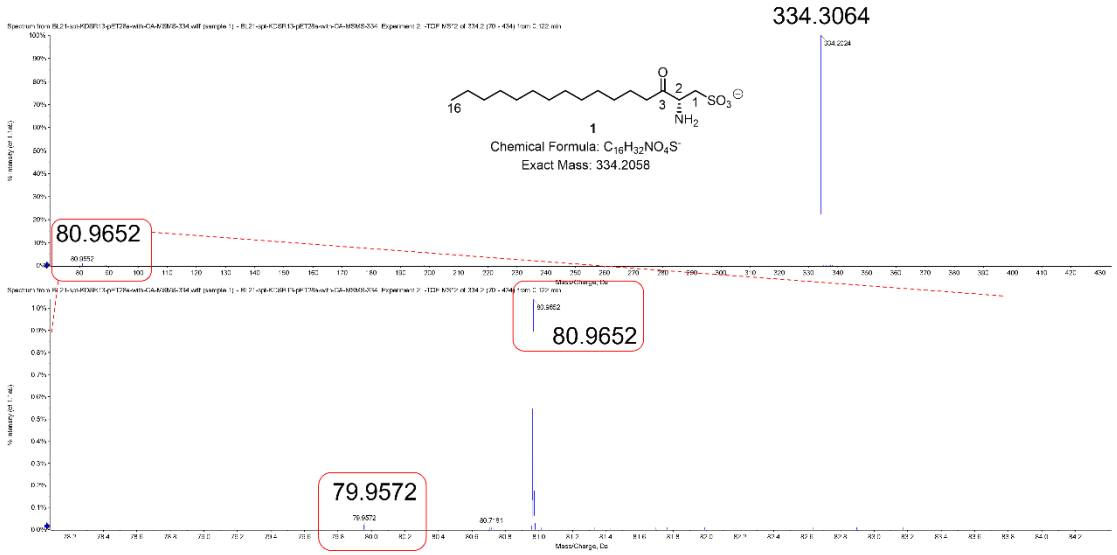

Figure S24. MS-MS spectrum of compound **1.**

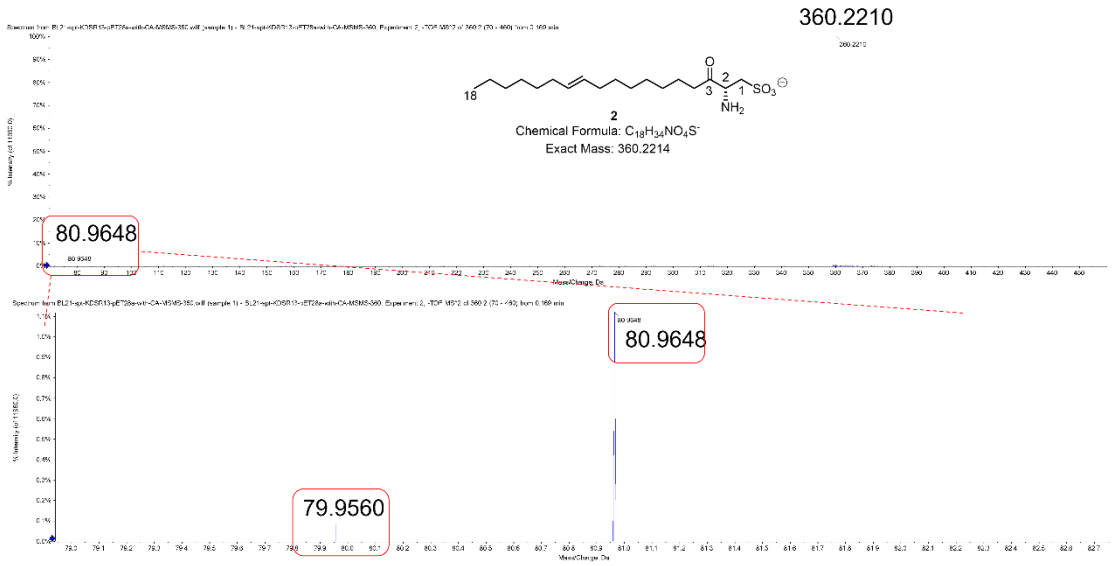

Figure S25. MS-MS spectrum of compound **2.**

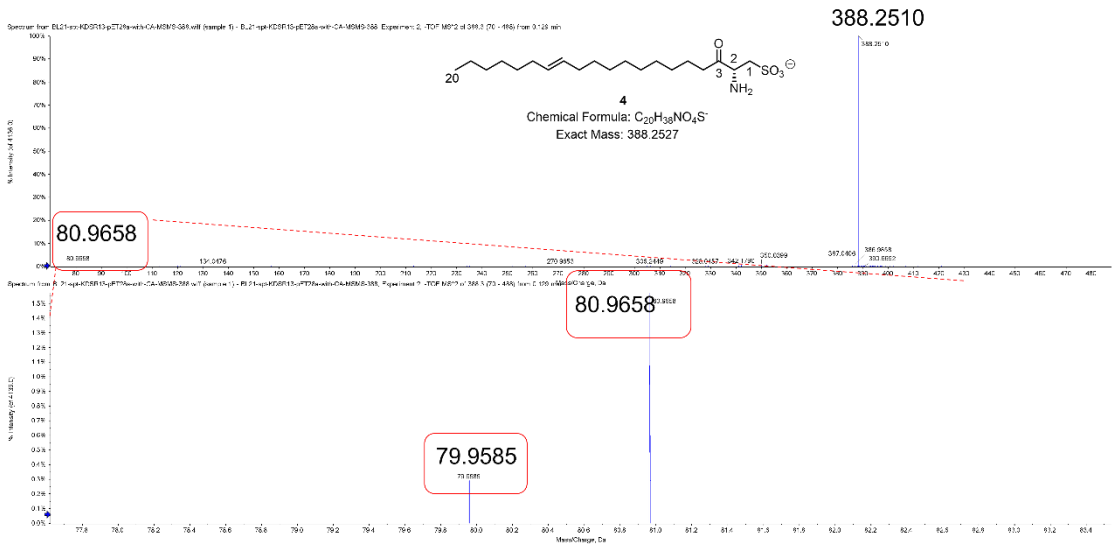

Figure S26. MS-MS spectrum of compound **4.**

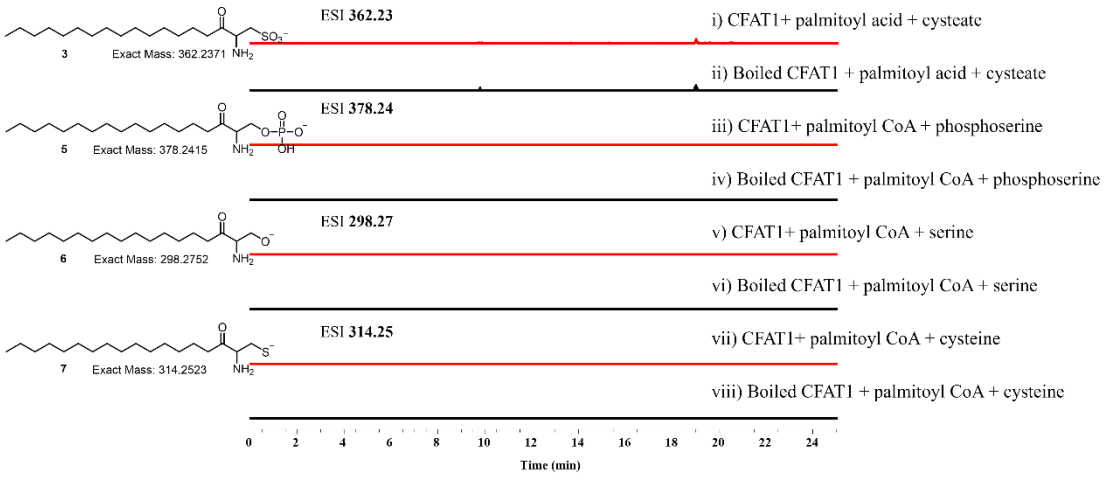

Figure S27. Comparative LC-ESI analysis production of *in vitro* CFAT1 reaction with different substrates. (**i** and **ii**) The extracted ion chromatograms (EICs) were extracted at *m/z* 362.23 $[M-H]^-$ for **3** (**iii** and **iv**) . The extracted ion chromatograms (EICs) were extracted at *m/z* 378.24 $[M-H]^-$ for **5**. (**v** and **vi**) The extracted ion chromatograms (EICs) were extracted at *m/z* 298.27 [M-H]- for **6**. (**vii** and **viii**) The extracted ion chromatograms (EICs) were extracted at m/z 314.25 $[M-H]^-$ for **7**.

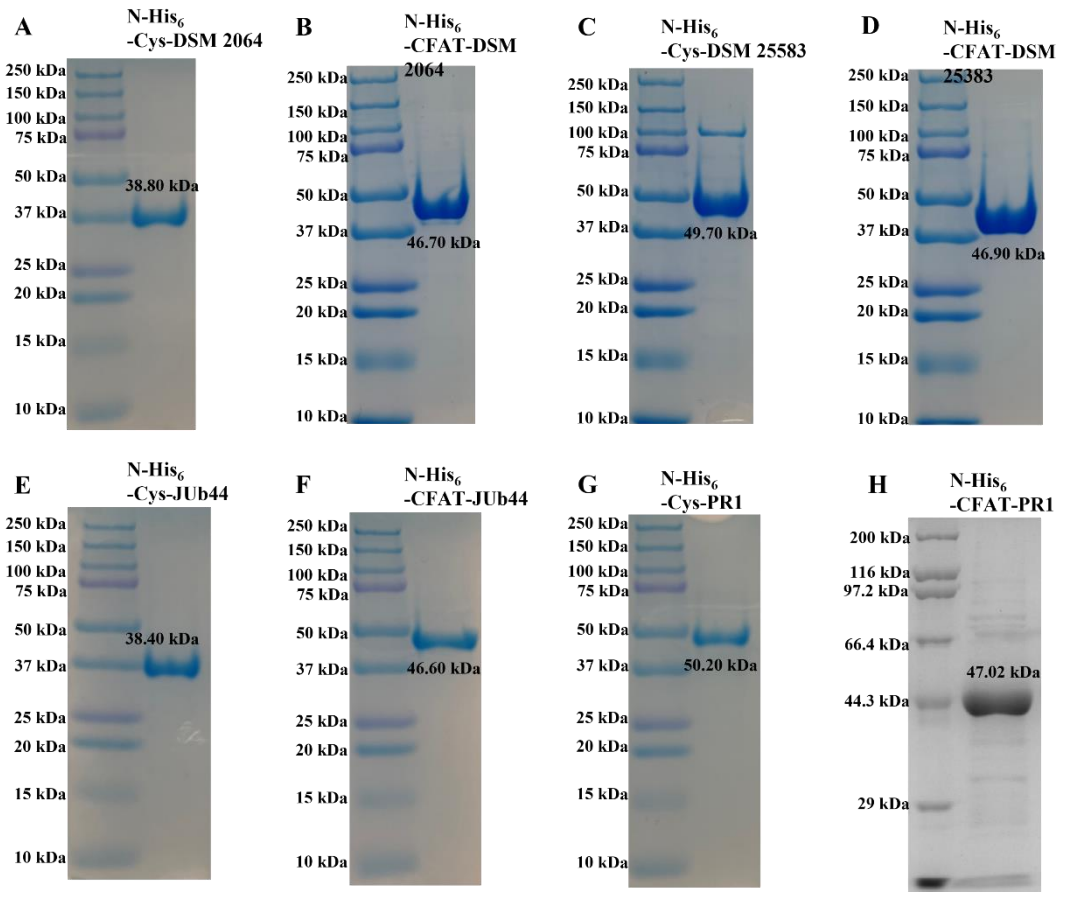

Figure S28. SDS-PAGE of the recombinant proteins Cys and CFAT from other strains.

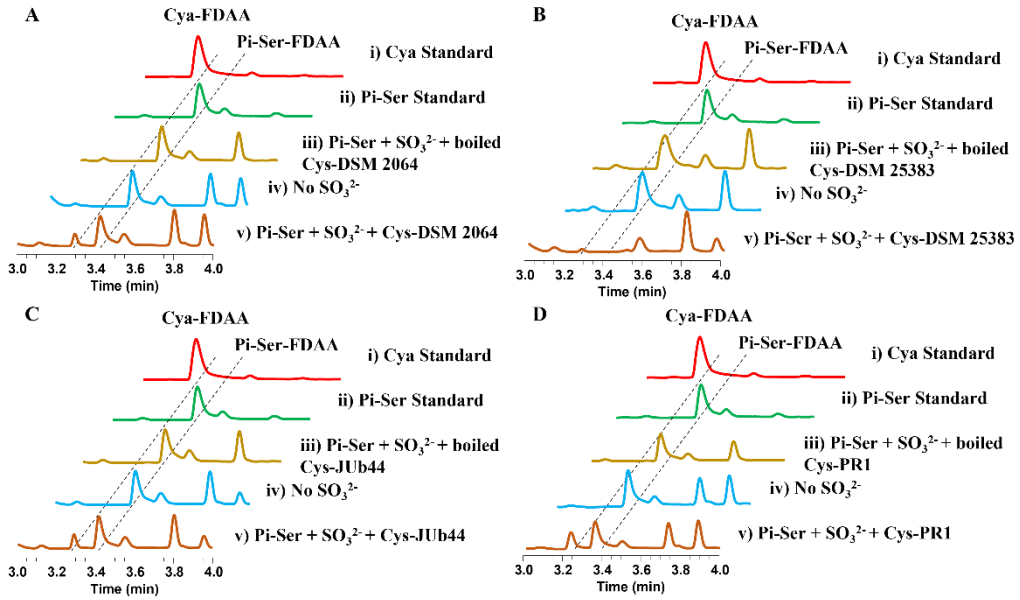

Figure S29. HPLC analysis of *in vitro* Cys assays followed by derivatization with L-FDAA. (**A**) Cys-DSM 2064, (**B**) Cys-DSM 25383, (**C**) Cys-JUb44, (**D**) Cys-PR1. Reactions were performed as follows: (**i**) L-cysteate standard, (**ii**) L-phosphoserine standard, (**iii**) L-phosphoserine + Na_2SO_3 + boiled Cys, (**iv**) L-phosphoserine + Cys, and (**v**) L-phosphoserine + Na_2SO_3 + Cys.

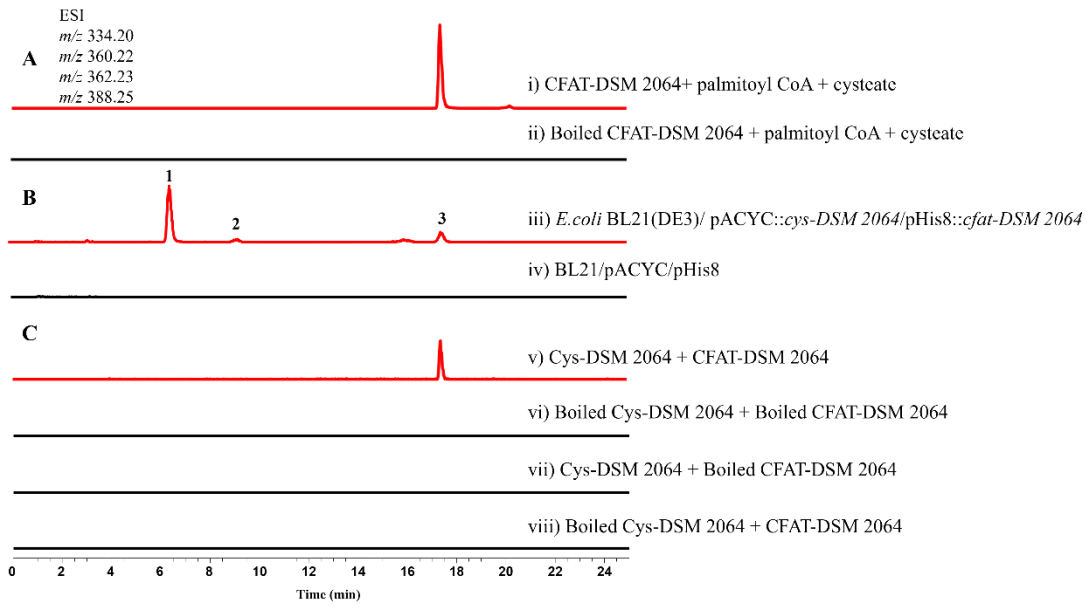

Figure S30. Comparative LC-ESI analysis of production of key reactions catalyzed by enzymes from *F. johnsonie* DSM 2064. The extracted ion chromatograms (EICs) were extracted at *m/z* 334.20 $[M-H]^-$ for 1, *m/z* 360.22 $[M-H]^-$ for 2, *m/z* 362.23 $[M-H]^-$ for **3** and *m/z* 388.25 $[M-H]^-$ for **4**. (**A**) Production of *in vitro* CFAT-DSM 2064 reaction (**i** and **ii**). (**B**) Production of *E. coli* BL21(DE3)/pACYCDuet-1::*cys-DSM 2064*/pHis8::*cfat-DSM 2064* (**iii**) and *E.coli* BL21(DE3)/pACYCDuet-1/pHis8 (**iv**). (C) One pot assay of Cys-DSM 2064 and CFAT-DSM 2064 (**v**, **vi**, **vii**, and **viii**).

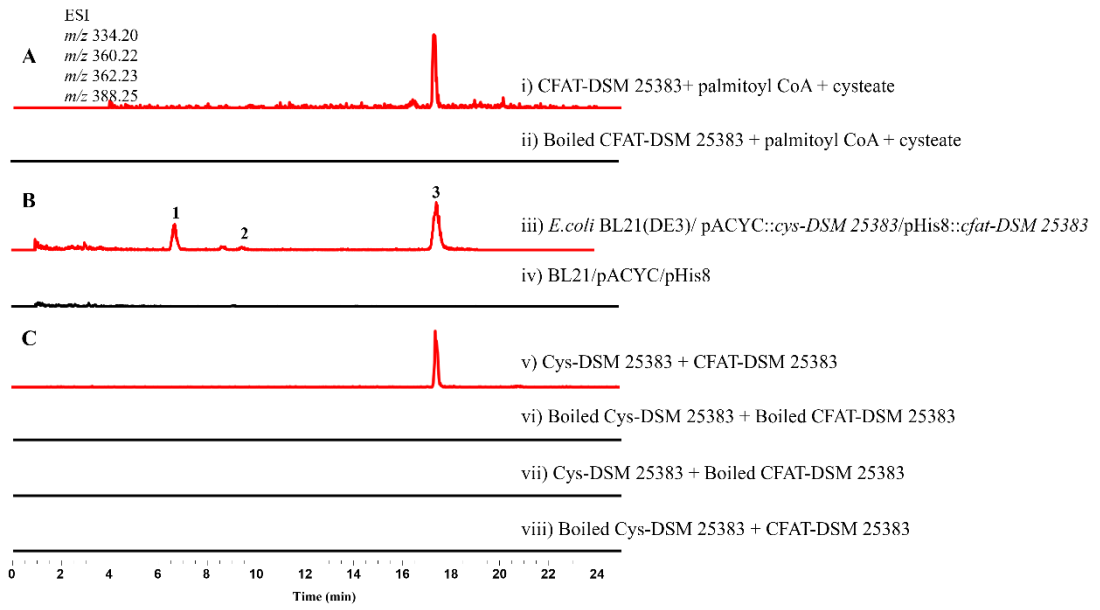

Figure S31. Comparative LC-ESI analysis of production of key reactions catalyzed by enzymes from *A. timonensis* DSM 25383. The extracted ion chromatograms (EICs) were extracted at m/z 334.20 [M-H]- for 1, m/z 360.22 $[M-H]^-$ for 2, *m/z* 362.23 $[M-H]^-$ for **3** and *m/z* 388.25 $[M-H]^-$ for **4**. (A) Production of *in vitro* CFAT-DSM 25383 reaction (**i** and **ii**). (**B**) Production of *E. coli* BL21(DE3)/pACYCDuet-1::*cys-DSM 25383*/pHis8::*cfat-DSM 25383* (**iii**) and *E. coli* BL21(DE3)/pACYCDuet-1/pHis8 (**iv**). (**C**) One pot assay of Cys-DSM 25383 and CFAT-DSM 25383 (**v**, vi, **vii**, and **viii**).

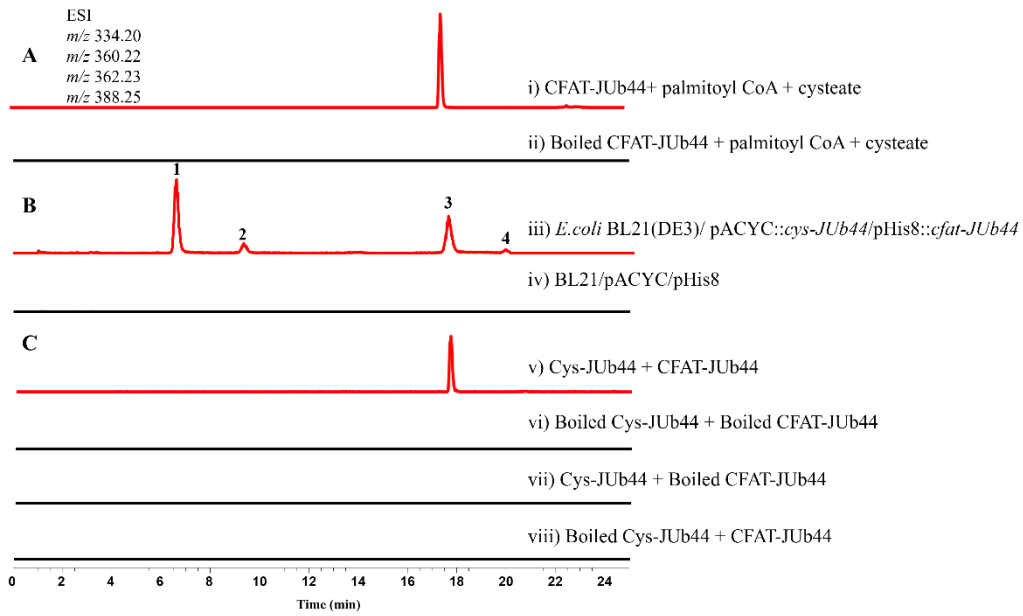

Figure S32. Comparative LC-ESI analysis of production of key reactions catalyzed by enzymes from *C. scophthalmum* JUb44. The extracted ion chromatograms (EICs) were extracted at *m/z* 334.20 $[M-H]^-$ for **1**, *m/z* 360.22 $[M-H]^-$ for **2**, *m/z* 362.23 $[M-H]^-$ for **3** and m/z 388.25 $[M-H]^-$ for **4**. (**A**) Production of *in vitro* CFAT-JUb44 reaction (**i** and **ii**). (**B**) Production of *E. coli* BL21(DE3)/pACYCDuet-1::*cys-JUb44*/pHis8::*cfat-JUb44* (**iii**) and *E.coli* BL21(DE3)/pACYCDuet-1/pHis8 (**iv**). (C) One pot assay of Cys-JUb44 and CFAT-JUb44 (**v**, **vi**, **vii**, and **viii**).

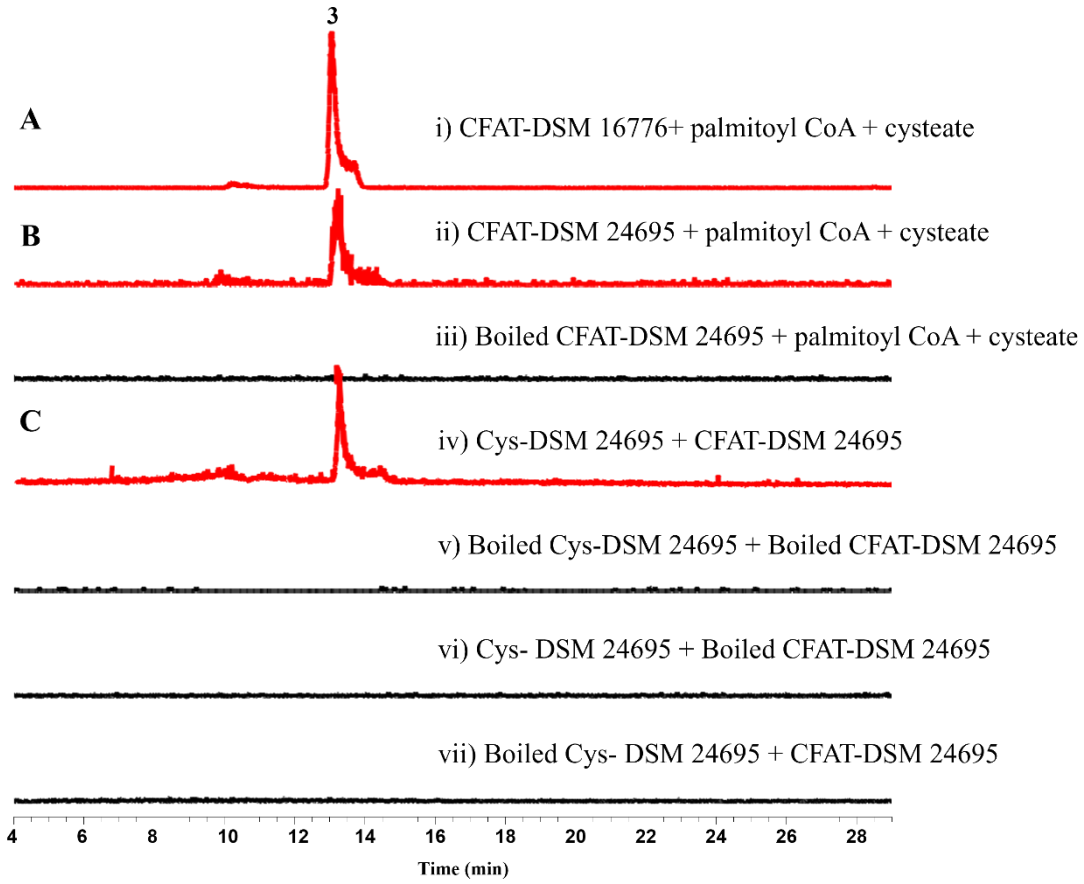

Figure S33. Comparative LC-ESI analysis of production of key reactions catalyzed by enzymes from *A. machipongonensis* DSM 24695. The extracted ion chromatograms (EICs) were extracted at *m/z* 334.20 $[M-H]^-$ for **1**, m/z 360.22 [M-H]- for **2**, *m/z* 362.23 [M-H]- for **3** and *m/z* 388.25 $[M-H]^-$ for **4**. (**A**) Production of *in vitro* CFAT-DSM 16776 reaction (As a standard control, i). (**B**) Production of in vitro CFAT-DSM 24695 reaction (**ii** and **iii**). (C) One pot assay of Cys-DSM 24695 and CFAT-DSM 24695 (**iv**, **v**, **vi**, and **vii**).

References

1. Teufel, R.; Miyanaga, A.; Michaudel, Q.; Stull, F.; Louie, G.; Noel, J. P.; Baran, P. S.; Palfey, B.; Moore, B. S. Flavin-mediated dual oxidation controls an enzymatic Favorskii-type rearrangement. *Nature* **2013**, *503*, 552–556.