Oxidative Dearomatisation: The Key Step of Sorbicillinoid Biosynthesis.

Ahmed al Fahad, Amira Abood, Katja M. Fisch, Anna Osipow, Jack Davison, Marija Avramović, Craig P. Butts, Jörn Piel, Thomas J. Simpson and Russell J. Cox.*

Electronic Supplementary Information

Contents

1. Degenerate PCR

Agar plates with malt extract media (MEM) were inoculated with *P. chrysogenum* E01-10/3 spores. After 5-6 days of incubation at 30 °C fungal spores were collected and stored in glycerol (up to 20%) or used for inoculation of liquid cultures. The fungus was grown in flasks containing 40-400 ml of liquid MEM depending on the purpose of the experiment - isolation of chromosomal DNA or detection sorbicillactone A production. Liquid cultures were incubated 3-10 days at temperature 28-30 °C with shaking (150 rpm; submerged cultures; collection of mycelium for DNA isolation) or without shaking (surface cultures; detection of sorbicillactone A production). In case of genomic DNA isolation, young mycelial cultures were preferred.

For the purpose of running PCRs and the creation of genomic library of *P. chrysogenum* E01-10/3, high molecular weight (HMW) DNA was isolated from this strain. Flasks containing 40 ml liquid MEM were inoculated with *P. chrysogenum* spores and liquid cultures were grown for 2-3 days at 30 °C with shaking conditions of 200 rpm. The fungal mycelium was harvested by filtration trough thin filter (Miracloth filtre paper Ø22-25 μm). The harvested mycelium was freeze dried by submersion into liquid nitrogen and than immediately grinded or stored at -80 °C prior to use. Approximately 2 g of freeze-dried mycelium was grinded in pre-chilled mortar (-80 °C, for 0.5-1 h) by addition of some sterile sand. The grinded powder was dissolved in 20 ml of 1x SSC (0.15 M NaCl, 0.015 sodium citrate, pH 7.4). Subsequently an equal volume (20 ml) of lysis buffer (0.1 M EDTA, 1% sarkosyl, 200 μg/ml proteinase K) was added to the mixture. The mixture was incubated for 2 h at 55 °C in waterbath. The fungal mycelium was separated via centrifugation at 11,000 rpm for 10 min. After removal of mycelium, the mixture was extracted two times with phenol/chlorophorm/isoamylalkohol (IAA: 25:24:1) and once with chloroform/IAA (24:1) respectively. All centrifugation steps were run at the speed of 11,000 rpm for 10 min. Approximately 2-5 ml of recovered aqueous phase was dialysed (dialysis membrane, cellulose MWCO1000; Roth) against 2-4 ml of 1 x TEN buffer (10 mM Tris pH 7.5, 1 mM EDTA, 100 mM NaCl) in a fridge at 4 °C. Samples were dialysed overnight (12-17 h) with continuous stirring. The content of dialysis membranes containing purified DNA was recovered in fresh falcon tubes. The DNA was precipitated by addition of one volume of isopropanol and centrifugation at 10,000 rpm for 5 min. The precipitated DNA was dissolved overnight at 4 °C in 1 ml of sterile water.

For PCR screening with genomic DNA of *P. chrysogenum*: Initial Denaturation: 94 °C, 2 min; Denaturation: 94 °C, 30 sec; Primer annealing: 53 °C, 1 min 34 cycles; Extension: 75 °C, 1 min; Final extension: 75 °C, 10 min. The primers were KHKS2/KHKS3c.[1]

2. Fosmid Library Construction and Screening

The protocol was modified from the "EPICENTRE" CopyControl Fosmid Library Production Kit protocol. However, all components of the kit were used for the preparation of the genomic library.

2.1 Preparative gel purification of genomic DNA. Isolated genomic DNA of *P. chrysogenum* was size-selected through a 20 cm long 1% LMP agarose gel. 100 μl of genomic DNA preparation containing approximately 2 μg DNA (20 ng/μl) was loaded into a 8-10 cm wide genomic DNA loading lane together with some loading dye. 100 ng of 40 kb control DNA was loaded into each of the outside lanes to be used as a marker for the isolation of correctly sized genomic DNA. Samples were resolved via gel electrophoresis: initial run 10-30 min at 60-70 V enabled that samples fast enter gel and thus prevented loss of DNA. After initial run gel was run overnight (12-14h) at 35-45 V. On the next day, using a sharp razor blade, 2 cm gel slices from both sides of the gel were made. These gel slices contained approximately 1 cm (each) of the well into which the genomic DNA was loaded. Both slices were stained with ethidium bromide for 20-30 min. UV light was used to locate the band of high molecular weight (HMW) genomic DNA (~ 40 kb) in the stained edge slices of the gel. By using a clean razor blade 5 mm wide gel slices from the central

part of the non-stained gel was cut out. This slice contained *P. chrysogenum* genomic DNA corresponding to 40 kb of control DNA.

2.2 Recovery of the size-fractionated genomic DNA. The gel slice containing the genomic DNA was placed in a pre-weighed 15 ml falcon tube. LMP agarose was melted by incubation at 70 °C for 10-15 min. The tube was transferred at 45 °C, and pre-warmed 50x GELase Buffer (Epicentre Biotechnologies) was added to a tube (final concentration 1x). 1 U (i.e.1 µl) of GELase enzyme stock was added to the tube for each 500 µl of melted agarose. The reaction was incubated at 45 °C for at least one hour. The GELase enzyme was inactivated at 70 °C for 10 min and the mixture was transferred to an ice bath for 5 min to cool down. The sample was centrifuged at a speed of 10,000 rpm for 20 min to a pellet of insoluble oligosaccharides. The DNA was precipitated by the addition of 2.5 volume of ethanol and 0.1 volume of sodium acetate (pH 7.0). The sample was incubated for 10 min at RT and then centrifuged at 16,000 rpm for 20 min. The supernatant was carefully removed from the pelleted DNA. The pellet was washed two times with ice cold 70% ethanol with subsequent centrifugation and removal of supernatant. A pellet was then air-dried for 5-10 min under the clean bench and dissolved in 60 µl of TE buffer. The dissolving of DNA was eased by an incubation of 1-2 h at 50 °C or an overnight incubation at 4 °C. The DNA concentration and quality was determined by running an aliquot (0.5 µl) of the DNA on an agarose gel using dilutions of known amounts of the 40kb control DNA as standard.

2.3 End repairing of purified size-fractionated genomic DNA. This step generates blunt-ended, 5'-phosphorylated DNA. The end-repair reaction can be scaled up or scaled down by the amount of DNA available. The set up of the end repair reaction is as follows:

Purified sterile water	0 µl
10x End-repair buffer (Epicentre)	8 µl
dNTP mix (2.5 mM)	8 µl
ATP (10 mM)	8 µl
HMW genomic DNA (~1 µg)	52 µl
End-repair enzyme mix (Epicentre)	4 µl
Final reaction volume	80 µl

The reaction was incubated at the RT for 45 min. The end-repair enzyme mixture was heat inactivated for 10 min at 70 °C. End repaired genomic DNA was precipitated by addition of the following components directly to the 80 µl of reaction.Precipitation of end repaired genomic DNA:

End repair reaction volume	80 µl
Purified sterile water	140 µl
Sodium acetate (pH 5.0)	20 µl
Isopropanol	120 µl
Final reaction volume	340 µl

Further, the sample was gently mixed via inversion and incubated for 30 min at RT. The DNA was precipitated via centrifugation at maximum speed (16,000 rpm), and the supernatant was removed carefully by pipetting. The pellet was washed by the addition of 70% ice cold ethanol and centrifuged for 5 min (max. speed). The supernatant was removed and the pellet was air-dried under a clean bench for 20-30 min. 20 µl of TE buffer was added to the pellet, and the sample was incubated 1-2 h at 50 °C prior to quality and concentration check via gel electrophoresis.

2.4 Ligation reaction. In this step the CopyControl pCC1FOS vector provided by the kit was ligated to size-selected and end-repaired genomic DNA of the fungus *P. chrysogenum*.A 10:1

molar ratio of the CopyControl pCC1FOS vector to insert DNA was proven to be optimal.In a new tube at RT, the following reagents were combined in the order listed and mixed thoroughly after each addition:

Purified sterile water	0 µl
10x Fast-Link Ligation Buffer (Epicentre)	1 µl
ATP (10 mM)	1 µl
CopyControl pCC1FOS Vector (0.5 µg/µl)	0.5 µl
Concentrated insert DNA (20 ng/µl; 0.13 µg overall)	6.5 µl
Fast-Link DNA Ligase (Epicentre)	1 µl
Total reaction volume	10 µl

The reaction was incubated at RT for 2 hours. Subsequently, the sample was incubated for 10 min at 70 °C in order to inactivate the Fast-Link DNA ligase (Epicentre). At this point the sample could be directly used for packaging reaction or stored at -20 °C prior to use.

2.5 *In vitro* packaging and titering the packaged fosmid clones. The day before performing the packaging reactions, a single colony of EPI300-T1R cells was inoculated into 50 ml of LB broth supplemented with 10 mM $MgSO_4$ and shaken overnight at 37 °C. On the next day, 5 ml of the overnight culture was transferred into 50 ml of supplemented fresh LB broth and shaken at 37 °C until the OD_{600} reached 0.8-1.0. The cells were then stored at 4 °C up to 72 hours.

10 µl of the ligated fosmid DNA was pipetted into a tube containing 25 µl of thawed MaxPlax Packaging (Epicentre Biotechnologies) extract and incubated at 30 °C for 90 min. At the end of the incubation time, additional 25 µl of thawed MaxPlax Packaging extract was added to the mixture, which was further incubated at 30 °C for the additional 90 min. Subsequently, phage dilution buffer [10 mM Tris-HCl (pH 8.3), 100 mM NaCl, 10 mM $MgCl_2$] was added to 1 ml volume. At the end, 25 µl of chloroform were added at the top of the mixture and homogenised by gentle vortexing.

To determine the titer of the packaged fosmids, a 1:10 dilution of packaged fosmids was made by adding 90 µl of phage dilution buffer to 10 µl of originally packaged fosmids. 10 µl of the 1:10 dilution was added to 100 µl of prepared EPI300-T1R host cells at RT for 20 min and spread on LB-chloramphenicol selection plates at 37 °C overnight. A sufficient number (~5000) of clones were inoculated into 51 ninety-six-well plates with 100 µl of LB medium containing chloramphenicol (12.5 µg/µl). After shaking at 37 °C overnight, 100 µl of 40% glycerol was added to the 100 µl culture of each clone for storing at -80 °C.

2.6 Screening the genomic *P. chrysogenum* fosmid library via PCR. For screening the genomic library of *P. chrysogenum* via PCR, fosmid pools from genomic library were made in 50 ninety-six well plates. Large (15 cm) Petri dishes were made with LB agar medium supplemented with 12.5 µg/ml chloramphenicol. Each of the ninetysix well plates from the library was replicated onto such an agar plate. The overnight cultures were grown at 37 °C and on the next day fosmid clones were striped off from the agar with sterile inoculation loop and by the addition of 1-2 ml fresh liquid LB medium containing chloramphenicol at the same concentration as LB agar plate. 750 µl of the fosmid clones suspension was mixed with an equal amount of 40% glycerol. Superpools were stored in a freezer at -80 °C. Plasmid preparations were made from each superpool after overnight growth of LB liquid cultures at 37 °C and 200 rpm. These superpool plasmid preparations were used for screening for PKS positive plates in the first round of PCR screening. One µl of diluted (1:2, 1:4) superpool plasmid DNA was used for screening.

These plate plasmid superpools were also used for screening library via the Southern hybridization method.Lane pools were made for each PKS positive ninety-six-well plate by taking out 20 µl from 12 fosmid glycerol stocks that belong to one lane. In that manner, eight pools were made per each ninety-six-well plate and were screened with whole-cell PCR protocol (1 µl of cell suspension per PCR reaction). From each positive lane 10 µl of each clone were aliquoted to a

fresh PCR microfuge tube in order to make a lane pool. From each of lane pools, a 0.5-1 μl was used directly in the whole-cell PCR. PKS positive fosmids that corresponded to the same gene cluster were digested with the set of same restriction enzymes. One restriction pattern was chosen for subcloning of fragments into the pBluescript KS(-) vector. Subcloned fragments were end-sequenced via GATC Biotech (Konstanz, Germany). In some cases, for the purpose of gaining more sequence data, fragments were cut with the second restriction enzyme, subcloned into pBluescript vector and end-sequenced.

3. Sequence

TCACACCGTCAGCAGCAGCGGCGGCAACCGCCTCCCCCATCCAGGTCCTGACCGTTCTGT
CCGTCACTTCCCAGATCCGCGCTTTCTCTGTCCTTCCTGTGCGACGGTTACGCCGCTCCA
TGAGCTTATCGCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCT
GGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGAT
CGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTT
TTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGA
TATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCA
GTTGCTCAATGTACCTAYRRCCMGACCSYTYAGCTGGATATTACGGCCTTTTTAAAGACC
GTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATG
AATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGT
GTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGT
GAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTAC
GGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCC
AATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTC
GCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTG
GCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAA
TTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGG
TGCCCTTAAACGCCTGGTTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCAGA
AATTCGATGATAAGCTGTCAAACATGAGAATTGGTCGACGGCCCGGGCGGCCGCAAGGGG
GTTCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTAC
ACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAG
GAAACAGCTATGACCATGATTACGCCAAGCTATTTAGGTGAGACTATAGAATACTCAAGC
TTGCATGCCTGCAGGTCGACTCTAGAGGATCCCACGTCGGAGGATGAACGACACAAATCT
CAAATGACGAGTATTGGGTTGGTCTTTCTTTTGGTTTAGTTTGGTGTCGGTGTTAGGCAT
TATGGGGATGTTCTTTTTTCCTGTTTCAATGTATAGCTCATCGCATACTGCTAGTACAGT
ATTTCTGCATATGAAGTAATTGTGATCACCACTACTGAACATAGCAATATTGCCTTGTAA
GGCCATTGAAGGTTATAGGGATGGACCCGTAGCTCACCGTCATTGTCGTTGTCGTTGTCT
TGTCGTCAGTCATCAGCCTCAGTGGTTTTAGTCCAGTTCCTTCCTCTCTGTAGTTTCTTC
GTTTCTCCTTTCCTTGCAGTTTCAGTAGCCTTAGCAATTCATCGTCCACCACTTGTCAAA
ATCTAGCCACTTGCCTTCTCCAGAAGCTACATCGGTCACTTGGGCATTCGTCCCTCTTAA
GCTGCGTAAGCCAATCATCATTAGCATGCAGTCTATCTCGGCCTTTTCTCTTCCTTGGTA
TCCCGATCTTGGGGAAACCCTACCGTCTCCGATATACCCAAACAGAATACGTTCATATAA
AGGATACCGGATCTAGAATGTCTATTTTTTTTAACCCGTATATAGATAGAGAAATTCATT
GGTCTATATATCTCCTTATAAAGGAAATTTAGTTTATGTAACCTCTATCCACTAGTTATT
GCCTAAGGGCAGGGAAGTGTAATTCTCACACTTAGGACCAAAAGACCATAGCAACAATCG
GAACGCCCGATTATCGCTGTAACGGTGTGTATATCAGGCCCCCGATCTCCATTAGGGATC
TTAAGCAGTTATGGCGTAATCCATGTGGGAGCTGACCACCCCAAAACACCGTAATAAGCG
ACATATTCTACGCTTAAATCTTAGAAGGTTCTTATCTTGTTTCATTGTTTGTCGATAGTT
CAGCACTATATGTTAGTTTGGGACCTGGAAGCTATATTGCCACGATACCAGCTAGAAATA
TATATCTTAATCTCCAAATTAGAAAAGAAAAGAACATCTTTAAATTTCCCAAGGAGGTTA
GATTTTGCGTGAAGACCATTAGGCTGTAAGGAACAGTTCCTATAGGCAACCGTTGCTGCG
ATATCATGGTGTGATATATCGCATGTGAATTACGTTATGGTATTTACATCGGTCCATATG
CGCCTCGAACCTGTTTGGGAAGTTTTCGTATACGACATTAAATCAGCCTTCTACTCCCCA
AATATTGGGGTAGCATGTGGATGGATGGATGGATGGCTTGAGGGATGCTGATTCTG
CCCAATTGGTAAATATCTTTATGCCAGAGCGGATCACATTTAGGGTCCTCGCAACTGTGA
TGCGGGTGACTGGAGATTAACACATACATGGGCCTGCAGTAAGAGTGCACTGAAACTATT
GAGTCTGATAGGCCCCCTTTAAAGCATCTCACGCCTATTTACTCCGTGGGTCTCTTTCAT
TTCCCAAACCTCGCTTTCTGTTCGTCTCCGAATATGTCGATTCACTTCCCGTTCACTCCT
GGGGGCCAATAGTCAAGATGGTGACCCCAGGAGCCGTACAAACTCCTGGTAGTTCGTACA
TGTTGCTTGTCGATCGACTCAGCCAGAACAGCCGTGAGTGAAATGCTATCGTATCAATGG
GATGGAATTCCTGAGAGCGGGAAGTTAAGAGATGGGCGGGACGGAAGCCGTGTAAGATAA
GATTAAACGTGAGATTTTAGCGTATTCCTGATTCCTGAAATTGCTGACATGCAGTGACTG
CTGAACCCCGCTCGGAGTCGGATAGCCAGAGAGTCAGAGAGACCAGAGAGACCACATAGC
GGGTCACGTTATGCCCCAAATGTGACTCTTTCTCGGAAGAGCTTAGTTTCCCAAGTTTGG
CCTCTAAGACTTGGGATTATCCCATCCTCCCCATGGGGTATTACACTTGAAGGCCGTATA
AGTGTAGCCCTTTGCCGCGGGTACCGGGATGGCGAAAAATAAGGTGCCAAGAAGGGGCGC
TGAGAGGATCATAGGGTGTAGCCGCCAATCCACGGGCAGCCAAGACAGGTACAGGTCTTG
TATAGTGTCAGATCACGTCAATCTCCAAACCGTGAAGGCTCGAGCGGAGGCGCCTTGAGT

CTTGCCGTCCGCTTGTTTAGGAAAATCTACAAGCCACCACTTGGGTGGCATTTTTCGTCA
GCCAAGAAGCCAAACCCTAACGAGAGGATGATCGTTGCACTCACTTTCTGGGATTACGGG
AAGATTTCCCGGTGCTCGTGTCCAACCACTGTGGCTTCCAATGTTGTAACAGTTTGATAC
TAGGTGTGGTCTTGGCAAGTTAGTGGCCACTGCATATGTAGGGTCTGGAGCTTCACTTAT
TTCGGAGAAAGCACGGCAATCGATAGTGGCGATCGAATGTTGGTGACTTCCTCCGTTGAA
TCAATCGCAACCTCCGGAGTACAGAACACGTGGAAATAAGTGGCTCCATGCCCTCGGCTA
TGCCAAAGGGGTTCCGCCTCCCTGCATAGCGAGCATATCTAGAGTGTCTAAGAACCATAC
CATCGTCTACCTTTTGCCTGTTCCAGCTAATAGAGAATAGGGGACTGGCATCGTTGGAAC
ACATTATCATCCCTGTGAGCCAGTTTGCCATTGATTCGGCCCTGGCGGAGCGCCGAGAAT
CATGCACAGGAATAGGGCGTCCAGATATATCCCTGTCTGCCAGGGTCCTTCCCCACGGAC
CTCCCAAACTGGGGCCGTCGACAGCCGCTGAGGAGGATTCAACTTACGCTTGAAGCTTGA
AGAGAGCAATTCACATGTTAATTGATGGTATAACTATTGGAACATGCTCTCGCTTCGCTC
CGTTGCCAAGTCTTCAAGATAAAGGATGCAACATCGTAATTGCTTGGTTGCGCAAGCCAA
GTGGCTAAAAATATACCTTTGGCAATACGAGGCATTGAGGGGATTTTTGTTCCGGAGGCA
GAACAGCCCACCGAGGAAACAATGGACCCATGAGGAGGGTCTCCAAGGTTGTCCATTGGA
TGCCAAGGGACCTGGCGATGACCAAGATGGGGCCGATGCGTATTGTCCGAACACCTGGCG
CCGCAAAGGGCTAGGATAACAACAACACACGAGGTCTTCAGGGTCCATAAGGGGTCGCGA
TCCTCGATCCATTCGAAGGTTTGGTTGTTTTTTCTTTTTTTTCTTTTTGGCCTGAGCGTC
CAATACGAGACTGGGGTTTCGCGATTGTAGTTTGATTACTATAAAATCTCCCGCTCGAGC
CTGCAGATCCCCATCTGCATATCAGGTAGCCAGAGACTTGATATAACCGTCAGCAATCGA
CTAACCCTCGCACACTTCGAGAAAAAAAGAAAAAGAAAGATGAGAAGGCAACAGTCTGGG
CTGGCTTGCGAGGAATGTCGTCGACGGAAAGCACGATGCGATCGTGTGCGTCCTCAGTGC
GGTATATGCGCCGATGCAGGCCGTACTTGCATCGTCGTTGATAAAAGATCTCCACGAGGC
CCCAAGAAGGGACAATTAAAGGATTTGCGATATCGGTTAGGTGAGTCGGGTCTAGTTGAA
GAGCTCTGGCAATAGTGTCCACTGACCTAGACGGGCTACCACAGCGGTGATCGAACAACA
ACTGATCAATCAGAACCAGTCCCTTGACAGTAGTCACATATTACCAGACTCAAATAGAGG
AAGCCTGGAACGAGAAGCGTCAAATGAACTTCACTGTGAAGATGGCACGATAGGATGCGA
CATGGATGAGTTGCAAATGTTGAACTTTGCTTCATCTGAAAGTATCTCCGCGTCTGACGG
GCAGGAGGCAATTGACTGGACGTTGGCGTATGATTGTCGGGAAGAGAGATCGGGTACTTG
GAATCATCTCGATATAAGCAACCCTCGCCCATGTATCATCACTCCTATTTCGCCCATTCA
AACTTCTGCAAGCGAGATGACGCCCAAGGGGGAGCTGGATATGTCCGACCTCATGCAAGC
TGATCTGTATGGAGGTTATGGTCCCTAAAACACAGGGTCTGAAAGCTGATATTACTGCAT
CTTTATAGGGATCTACTATATTTTGAGCGAGTCCATCCAATTGTCCCGATGATACACAAA
CGACGCTATCTTTCATGGGCCAACGAGAAAACGGTCTCACCTGCCCGAGCTTGCTTGCGC
TCAGCCATGCGCACCATCGCTGCCGCCATGTCCGCTCAGTTCTGCGCGTTTTCCGATAAA
CTCTATGCCTGTACCCGAAGTATGCTGGAGATGCAAGACGTACAGGGCGAGAACGGTTTG
CCATGGATGACAACCACGAGAGCGTCTCGCCGCCGGATTGAACACGAGATGATTCAGGCA
TGGCTCTTACTTGCCCATTGCGAGTTCTTACGCAAGCCTGAGCAGGACGCCCTCCTTGCC
TCAACCAGGGCGGTTAGGCTTCTCCAGCTCTCTCGGCTATTTGATATTGACATGCACGAT
GATGAGACCTCGCCGAATGAGAACTCAGGGTCCTGTCCGTCAGTTTCTCCTTCCACGACG
CAGAATCTCCCAGATGAAGCGTGGATAGAGACCGAAGAGAAGCGGCGAACGCTGTGGACA
GCCTTTGTCCTTGACTGTTTATCCAGTATGCTTAGCGATCGACCGTCAATGTTGCACGAA
GAGATGGTCGGTGAACTGCTCCTCTCCTTTCCTTGTGACAATAGACAGTTCAGCTAACAA
GATTGACCCTACAGATAAACACACGTCTTCCGATGCCCGAGAGGGATTTTCAAGGTGGCC
AGCGACCAACCCCTATGGGCTTCCTACCCGAAACAATGGGCAAGACCGGTGACTGCGAGA
CACTGTCCTCTTTTGCCCAGTGCGTTGTGCTCGCAAACCTTTTCGGGCGTTGCATAGCAC
ACCGGCGGCTCGCACAGTCAGTATCCTTTCCGGAGTCTGGCTCCGAGTCAAAATCGCGGC
AGTTTTGGATGAGGCACGAATGGCTGGCCGCGGCAGCCGCCCATGCAACAAGGACGATGC
CACCGACACAAGCCCCCAATGAGTGTAAATCTGAGACCACAAAGTGTGATCCATTGGCTG
CTTTCAATCGTATCCTCGCCTACAGTGCATGCATATCCCTTAGCGAGACGGCCGAAGCAA
GGGCATGGGAGACCCTCGACGATCACACGCTGGCTTTGTCATATAGACAGGTAGCCAGCC
AAGCCGCATACGAAATCGCTCTACTTATTCAAAAGGCGCCTCGGATAGCTTTCTTTAAGA
TGCATCCTGTCCTTCCTAATGCAATATACTTGGCTGCCAGGTTTCTCAGAACTACGACGC
CACACTTCGCGACTCCAGCAGAACATGACCACAACAGCATTCACCATCTACTCGTCGCCC
TGGGTTATCTTAGCAGTGTCAATAATCTGGCCCGAGACCTGCTGGTCAAAGCCGAAGCAG
ACGTTGGGAAAAGTGCACGGATAGCAACTGAAGTGACGGGGTCAAATTGGGATGCGTTAA
TGGGCGATACATCTAGACATATAGGCATGATGGAAACATCTATATCAGTGTAAAATCCAA
ATAAACTCAAAGCCATTACCGGCGTTATGTCTCCCGGTCGCCAAAGCTGCCTTTCATTTG
TCTGAGGTAGATGTGAAATTATAATATATACGTTATCATCGTGCTTTGTCCGCTTCCCTG
GCAGGTATAGATCTAGATGCAGTTATCTGAACGTGCTTTCGGGATTCAAGAGGTGCGAAA
TAACACCGAACCGACAATGAAGGCCCTTCACCTGCAGATCCAAATCGACGATAGGATGAC
AGCCGCGTCTCCGATAAATCCCCCAGAGGCTCAGAATGCAGTGTCTGCGCCCCTAACTCG
CTCTCCCAAGATAGCCAAAGCGTCGCTTACCATACCAGCAGTATCGTGATCCCATAATTG
ATGACTACGGCTAATTATATCACGACTCAGCTCTTCAGCTGCCAGCCCCCAGTCACGATC
CTTCCAAGTACACAGTTCACCTGTTCGCCTGCTACTCCGAACCAGCCACTGTGACCTCTC
GTACCGTACCTCACTGTACACGGCCAATGCTTCCGATGCCACAATACCTGAGACATTCGG
TGCTTCCGCCAACACAGCCAATACCTCCGCGAGAACCAGAGCGTCTTCTATCCCAAATCC
AGCTCCGCCTCCCTGGTTTGGCGTTGAGGCATGTGCCGCATCCCCAGCTAAGCAAACACG
GCCCTTGGCAAACGTCGGTACCGGTGCTTCTAGCATATCAAACACGGCCCATTTTTCTAG
CTTCTCCGGGAACATCGACACAGCTGACCGGACAGTTGGACCAAATTCGGCAAATGCTTT
GGTTGCCTCGTTCCGCGTTGCCGGGAGTACATATCTCTTGCTATCGCTACTTGTTTGAAC
CTCAGGCCATTCTTCCTTGTCCATGATGAAAGCCTCGATGTGCATTGCATGCACCTCGGC
GAGGGGGATAGTGAGTACAAATGCGCCTGGACCAGTATGAAGTACCGCACTGGATGTCTT

TTCTGGGCCGAGCACAGCGGTTGCTTGCGCTAAAGGCACCATACCACGGTATCCTAGTTG
GTGTGAGTACTGGGCCCGAGGACAAAGTTCATCGGTTCCGAACATGGACGCGCGGACTTG
AGATCTAATCCCATCACAGCCAATGACTATTGCTATTATTAGTACCGCCATGTTATTCAA
CAACCGGCAGCGCCTGATCCATTAACTCACCAACATCAGCATGGGCAATTTCGCCATCCC
TGAACTTTAGCACTGCCCTCCCGGATTCATGATCTGTCTCAATCGAATCTAGCCATTTCC
GGTATTGAACGCAGTCTTTCGGGATATGTTGGACAATTTGGTCGACGAAGTCGCACCGCC
GACACGCCTTGTAGCCTCCCTTTACGTGAAGTTGAAACAACAATGCTGCGGTTGATGTGC
GAGGATCTTCGCTGCCTTGTTCTCGAACACCGTCCACAAAATTGATCGTTCCACCTGACG
GCGCACTCGCCACATTGGTAAAGCTTTGAAGGACGCATGGATCCAGCGCCTCCATGGCGC
GCTGCGCGTTGGGTGTGAAGGTGATGCCGACGCCCAACTCACCAAAATTCTCGGCTCGTT
CATAAATGGTGAAGCTAACATTCCGCTTCAACAGCCCGACAGCGAGGGCTAGGCCAGTGA
TGCCGCCGCCGACGATGGCAACCTCGAATGGGCTGTTGGCGGATCGAGTCATATTGAGGT
CGGGTCGCGCGATAAACCGATGGGTGGATTTGAGGAATAACGACCTGGGCTAGATGAGTC
TTATCATCCTCGATATTGTCCACATCAATAATCTTCTAGACGGAGTTTCTGCCGCTTCAA
TGACGTGGTTGGCTTTGTATCGGCGGACTCCGGAGTTTTTGTCGAACAAACCGGAGCCAC
ATCTACTCCGAAGTACGTTGTTTGAAACTTGGCGCGGGGCCATAGAACAAAGAACACAGC
TGATATGCGTTTCGAGATGGCGTCAGGTCAACCTCGTGTACGGTGGGTCTGAGTTATTGG
AACAGTATAACCTCTACATAACAGATCTCCTCTGCCGGAGCTTTCCTATCTCAGTCACCA
CTGTGAGCTCAAAGTCAGGGCTGCCCAGTGCCACGTGATAAACCAGTGTGGACTATGGTC
GAGAATATAGGCAAAATTGCACATTTTCTAAGGTATTTTGACTGAGCGCCTGCCTGGTTG
ATAATTAACCACCGAGGTCAGGTAGATTTCTAGGGTGATGAGTTAGATATGCTGCAAGAA
ATGGGATCGCATCATTCTTCTATTCTTCATCAGCCTGTCTGGTAAATAAGTCAGCCTTTT
CACAGCCCAAATCAGGTGGAAGTTACCGAGGAATACATCAGAGCAGTTTTACAAAGATCA
AACGAATGAAGCTTGGAAAATACAGATTGTAGGCGAGAAATATATCTAGAGTAGAGCTAT
CATTCTATATAATCCATGGAGGTAGACAAATGTGCAACTAGGATTAAAAAAAAGAAAACC
ATCAACGCAGGAACCCAGTGTCCTTCCAAACCTGGAAGTATTTCATAACTACCGCGTCAG
CCACCGGCTGCACGCCTCTCAGTGTCGGCGAATGCTCCACGGCATTATTCGTGGCCATGG
TGACCCCACCGCAAGACATGTGCTGGAACTTGTGCTCAAAGAAATCGATCGCCTTGCCCG
CGGGGTTATCCCAGGGGTCCTCACCAGGGAATGCCTTCACCCGTCTCAGCCAGTCATCTA
GTGGGACAATCCCTTTCTCTGGGATGCCTAGAGCCTGAGCTAGTACAGGGACGATCTCTG
CCCAAGGCTTCCGCACCGGGTTATCCACATGGTAAACGGGATGGCAGTCCGGCGCATCAC
GTAGTAGTAGGTCTGCTAGCGTAGCCGAGGCTTGCTCGAGGGGCAGCCAGGTTAAGATGC
CTTCCACCGCTGGGAAAGAACGCAGGGTCTGAGCGGACTTGAACAGGAACCCGAGAACTT
CCATGTGGTTCCAGTACCCAGTCTCCATGGATCCTGATAGCTGACCCAATCGGACTGTCA
TTGCGCGGAAACGCTCCGGGTACTGACCCAATGTCTCGTGGAGGATCCGCTCACATATGA
CTTTCGCCCCGCCGTAGCCGTTCGGAAGAGCACTGTCGATATTCAGATGTTCTTCCGGAA
CTTGAATCTCGCCGGCGTGGACGGATGGGTATTTACCCACCGCCGACAGGGAAGAGATGA
ATTGGAAGGCGAACTTGAAGTCGGTGGCCTTGCGGCGAGCTGAGATCTCTGCTGCAAGGT
CCACAAGATTGCGCATGATCGTGAACTGGGCCTCATTTTGCTTGAGAGAGCGCAGGCCGT
TGACAGGGAAGGCATTGTGGATAATGTGAGTGACGTGGCAAAGTAAATGTTTGTACTCCT
CTGTGTCTAGACCCAACTGTGCTTTGGAGGAGTCAGTCTCGATTACTTTCAATTTGGCGA
GTTGTGAGGCATCTAGGGCAATGGACTTTGACTCCAGCGATTGAAGCTGGCGGTGCAATG
GGTCACGAGGGGTAGCGTCCTTAGCTCTCGCACCTCCGATCGCCGGTCGATTCAGGCAGT
AGACCCGGTCGACCGAAGGAAGAGAGACTAAATGGGCAATAATGTGGCTTCCAAGACTCC
CGGTCGTCCCTGTGACGAGGACAGCCTTGCCGTACTCAGAAGAGTCAGTGAGATCTCCGG
AGTAGGCAGGGATTGTGAAGCCTCGGATAGTTTCTTTGACATAGTCATCAGCAACCTGCT
TTTTCTCGTTCAATTCCTCCTCGGATAGGTCATGGTCGTCAACTTGTTGCTTGGCGACAG
GGGGAAGACGACCAACGTTCTGCTCGACATCGTCGGCCAGAGCGATTATGACCCGTTGGA
CACGCACTTCGGGTAGCTTCCCATCGGTCCACTCAACATGTTTGTAGCCTGCATCGAGTA
GCTCCTTTTCCCATTGGCGTTCGTCTACAACGGCGTGCGTGCGTCCGTCTTCAAAGAGCC
ACCAGCCTTCCAGAGTACCCCAAACGACGTCCACCCAATGCAGCGTGCTGTTCATTTCCA
GCAACATCAGAAAGCCATCTGGGCGCAGGAATTTGCGGATATTTTGCGTCGAGACTTGGA
GTGAGTGTGTAGCATGCACAGCGTTACTAGCAATCACAATGTTTTGCGAGCCGATCAGTA
GCGGGTCGGACGGAGGTTGTTCAATGTCGTGCACAGCGAACTTCATGAAGGGGTACTGTT
TGAACTTCTTCTTGGCTTGCGCCACAAGAGACGGCGAAAGGTCGGTAAAGGTGTACTCGA
CGGGGATTCCAAGCTTGGCCAGCGCGGGAACGAGAACTTTGGTTGTCCCGCCAGTCCCAG
CACCCATCTCAAGGATCTTCAAGGGACCACGATTCTGAGCGCAAAGTCTCAAGCTTTCTG
CAATACGGGAGAGGAAATCAGCCATCAGCTGGAAATACAGCTTGTTGAACGGCAGCTCGC
CATAGAAAGCGGCAACTAGTTCACGATTTTTGGCATCTCCAAAAATGAGTTGGGGACCAT
CGGCCTTTCCCGACAAAACGTCTGCCATCCGAGAACCTATATTGTACGTAAGCTGGTGGG
AAGGGCCATCGTCCGGGTGGTGGGACATGAGGTCGTCCAATATGGCTTGGGAGGACTGTG
TAGGGAGAGGAAGGGCGGTGCGAGTAATGATGCCTTCGTCCACGTCTATGATTCTCGTCT
CTTCGAGCATCTTGTACAAGTATTCATGGAAGCGGTGATGGCGAGGAACAAAGGGTACGG
GCTGCAGCACTTCCCCCGGCTTGGCGGCCACCAGGTCGCAGCCCAGCTGCTTGAAGGCAT
CGCTTGTAAGCACGAGGCACAGTCTGGTTTGCTTCTGGGAGGCACCATCAAGATAGCCGG
CGCATTTCCATTTCTTCAAGTAGGCATCGGTCTGCTCATTAGCTGCGCGGAAGGCCTCGA
TAACGGCAAATGCCGGCAGGCCCACCTCGTTGCCAAGGTCGATATAAGATCTAGAAATGT
CCTCTTCTTCCAGCTTGGGGCTTGGCGTGGCAAGAATCCCGTCGTTGCTGGATGGAGGAG
TGTTTCGGGACGAAGAAGCTGCATCCGAAGACTGCGAGGCATCATCCTCACCCTCAAGCC
CAAGTGTAGACTGAAGGAAAGCGAGAATGCCTGGTACGTCAAAGATACTCATGAGTTCAG
ACTGTTCCAAGGTGCAATTAAAGGTGGTTTCGACCTCGCGGGCCATCTCCATTCCCATCA
GAGAATCAATCCCAATGTCTGCCAGTGCATCGGTTTCATTGATTTCCTCAGGCTCGAGTC
CAGAGATATCTGCCAGGACGGGAAGAAGCTTCGCCCATAGCTCAGTCTTGGTGTTCCTCT

TTTGAACCTGAGCACGAGCGGGCGGTGATGTCGTCATAGATGAATACTGGTTCTCAGTTG
GCGCAGAACTGATGGGAGCAATTGTAGCGGGGAACGGTACAGGCGGTGCGGCAATAACAA
CGTTCGTGAAAAGTTCACTCCTTTAAGCTATATCAGAATCGTGATTCGTAGGCGGAATAA
AATACTCACATGGATCTAGCACTGAATTTGACGCCAAGTAGGGCCTCGTCAAGTAGCCTA
GATGCCGAATCGAACACAAAAATGTCTGTTATAAAGGTGTCTCCTTCTGTGCGCTTGTGT
GTCGCAAGGATTTGCCATTCGCTTTGATGGTCGGCGTAAGATCCTTCGCTGATTTTCCGC
AGGAGACTCGGTGACCTCATCCATTGCTCAATGCCATCGGCAATATAGACGGTGTCATCT
GCGGTGTTCCGGCCCGGAGCAAGACAGTTCACCCAAATGCTGCCAACCTGGGAAAAAGTC
TCCCCGAGGGCGAAGTCAAGCCAGGAGTCCCGAGACCGTCGCTTCACAGCTCGGCCAGCT
GACTCGCTCGGTCGCCCAACCAGCCTTTGCAGGCCGCGGAATTTGGGGGCGTAGCTCACG
AGGTCCGAGTAAACTTTATATATGCTCTGCCCTTGAATCACTTCGTCAGCATCGTCAGCA
GACTCCAACGCCCGCAGGCATCGTTCATGCGTTACAAGCCGCTCCAGCCTACTAAACTCG
AAGTTTGAGCGGGCGTCATCGGCACGGTGGAACTCAAGCTGTCCGCTCAGATGCACAGTC
TTGCTAGAGTCGTCGGCTTCACTGGTCAACTTGAACTTCCAACCTTCAGGGGCATGTCCA
GAACGTTCGAATTCAAGAAATACTGCCCTTGTTGGATCCATCACAAGCGGGAGGTGATTC
AAGACGTTATAAATGTGGGGATGCAGCTCATTGGCCGCAATGACTTCCAGTCGAACGCTT
GTGATTGCCTGGATTGCGGTGTCAATCCCGAATATGGGCGGACAGGCCTGGACAGTTTTT
CCCAATGTGTATCCTGAGACAATATCGACATATGATTTCGTGGTTGTATTGATGCGGAAC
CGACAATCGGTCTCAGTCTTGTCGCCGTATCCCATGAAAGTGTAGAGGCCTGGTGCCGGA
GCTTCTTGGTCCACACCACCGTTTTCATACACCAGTCGCTCGATTACCACCGGCTTTGGG
GGCGGCTTAAACTCGAGCCAATGGCGATACTTCTCGAATTGGTACGGTGGGAGCATGATG
GGTGCATATCCATAGGTTTGCGCCCGGGAGTGTGGCCAGAACGAGCAGGGCAACCCAGCA
TTCCACAGACTCATGGTCACGTCGGCGAGTTGTCGGGTCCCCTGAGTGGTGCCAGTGACA
CTGACTGGCTGGAAGGTGGAACCCTTCGGCATCCCCAACGCCCTGCCTGTCATTGTCGTT
ATGGTGGAATTAGAGCCAGCCTCCAACCAGATGGCCTCCGGATACTGACTCGCAAGTCTT
TGGACGGCATGGTCGAAGTAGACAGGATTCCTCATGTGCTCAGCAACGTAAGCAGGTGAG
ATCGGACCGGTCTCCCGCTGCTCGGTTGCCCTTTCGAGAGGTATGTGGGCATTTCCAAAG
CTCAGACTGCGGCCCAGGGCCTCCAGTTGTGGCCTCAAGTGCTCAACTAGGGTCGAATGG
AATGCATTAGTCACGTCTAGCCTCTTGTGTTTAATTGGGATATCAAGGGTTGAAACTGTC
TGCTGCACAGCATCAATTGCAGCGGCTGAACCAGCGATGGTGAAGCTCTTCGGGCCATTG
AAACAGGCAATTGTCGCATGTCCAGCCCGTTCAGTTTCACCAAGACGCGCATTTGAAGCA
ACCAAGAGCTTTTCCACATCGTTTCGGTCCGCTTCAACTGCAATCATGGACCCCTTCTCC
GGGCCCCAGCTCTCTTTGATAATCTTCGACCGGCCATGGACGAGCTTGAGCGCATCCTCC
AGGCTGAGGATGCCTGATACGCAAAGTGCGGTGAGTTCTCCGAATGAGTGGCCAACCAAG
GCTGCTGGTTCAACGCCGCAATCGATCCAGCTCAACGCACATGAATACTGCATGGAAAGG
AGCAGCGGCTGGAGTACGGCCGGGTCAAGGATCGGCTCGCTTTGGAATATTCCCGGGTAG
ATACTACCCGCCCCGATCGACTTGCAGACACTGTCGCATCGATCGAGGTATTTACGTAGC
ACTGTGGCTTTGTCAAAGACTCCTCGGTCGAGGCCAACGGATTTGGACACCTGGCCCCCG
AAGCACAAGATAACAGGTCGTACGGAAGGCACCGCAAATGTTTCGAAAGACGCCAGCTTT
TCTTCGAGTTCAGTGATGGACTCGGCGCCGAACACGAAGCCCCGGCTAAGCGACCAGTTA
GATTGCCGATTGACGTTGAATGACAAGTTCTCAATTCCTAGTACGTCTCTCGAGATGACC
TTGTTCTTGATGAATTGGCGCAGTCGTGTGGCATAGGCCCGAATGGCTTTGTCGTCGAGC
CCAGAAATATAAAATGGACATCTGAAAGTGGAAGTCGGCGATGTTAGATCCGCGCTTCCG
TGACCGACTGCTTCACTTCCTGAGGGGTATTTCGGTGCCTGCTTAATCACCATGGAAGCA
TTTGAGCCAGCTGCGCCATAATTGTTGATGAGTGCTACCTTGGACTCGTCTTCCCAAGGG
AGGGCTGCTTTCGTGATCTCCATATTATCCGCAGGTGAAGCCTTGATAGACGCGCTCATG
GAGGTAAAGCTGGCCTGTGGGGGTATGCGACTTTCTTGCATCATCAACAACATCTTGATC
AGTGCTACAACCCCAGAGGCACCCTCAGTATGACCAATCAGCCCCTTCACAGAGCCGAGT
TGAAGAGGCTTTAGCCCCGCGCGGACGGAGCCACCAAAGACCTGGCGAATGCTGTCGTAC
TCTGCTGGGTCACCAACCGGAGTTCCTGTCCCATGAGCCTCGACGACCGAGATATCATTA
ACTTCCAACCCGGCCTTGCCGACAACGTTCTGGAACACGTTGGTCAGGGAAGACGGGTTG
GGAACAAAGATCGGAGTATCGTTTTGATTCTGGTTGATGGCGGTTGCCGAAATCACACCC
AGTATCTGATCGCCGTCGGCAATAGCATTGGATAATTTCTTCAAGAACACGGCGCCAATA
GCCTCACCTCGACAATATCCATCAGCCTTTGCATCAAACGGCTTGCATTGCCCAGTAGGA
CTGAGGAACGATCCTGCAGCCAGGTTCTGGAAGAACATCGGGGTGCTATAAAAGTTCGTG
CCACCCGCCAAGGCAGCAGAACAATCACCGCTCAGGATAGCCCGGCAGGCTAGATCTATG
GCGACAGTTGATGCTGAGCACGCGGTATCGAGCATCATTCCTGGGCCCGTCCAACCAAAG
TAATGACTAACCTTCCCAGCGATGTAGCTTCGAAGGGCACCTGTGGCGGAAAAGGCCGTC
GGGGAAGTATGAGAGATATTGTTTTCATAGTCATTGGCGACGCAGCCAATATAGCATCCA
ATTCGGCGATCTGCACCAGGCCTGTGATAGTAGCCCGACTGTGCCACGGCCTGATAGGCA
GTCTGCAAAATCAGTCGTTGCTGTGGATCCATGTGGAGGACCTCGCGAGGAGACTTCCTG
AAAAACTTGTAGTCGAAAGCGTCGTAGTCATCGATGAAGTTTCCGTACCATTTCCTGTCC
TCGCCGTCCTGGCCAGGTCGGAATACAGTCTCCATTGCAAAACGCTCGTTAGGGACAAGG
TTCTTGTGCTGCGATCGGCCTTCCAACAGGATATTCCAGTACTGCTCCAAATCCTGTGCC
CCGGCGACCTGGCATGACATGCCAATTACAGCGATATCGTTGTCGATACATCCTGACGGC
ATATCTGGATTTGAATGGCGTTGGCCAGTACTCTGAAATTCGTAGTGGGTGACTTGGCTG
TTTAGTCTTCGCAGGAGGGTGGGCGGGACACATCGCTCTGGGCCAAATTCAATGACCTTA
GAAGTCCTATCTTGCAATGAACTGGACACTGCGGAACGGAAGGTCTTCACCCAGTTAAAT
TGTTCCACCAGGAACGCACGCGAAGCAACTTCTAGTAGACTGTCATTGTCTGCTAGGATC
TTCTCGGAGTTGACCCGTGTGCGAAGAATGAGAGAAGAAGCATCGGGCAATTGGAAAAGG
GGATCTTTTCTACAGAAGCTAAACAGGGCTTCAAGATCGTTATTGTAGAGCTCGCCTGCA
TGAAAGCGTCCATGAAACTCGGTTTCGCTAGCTGTAAACCCTGCTCGTGAAAGATGTCCC
TTCAAGTCGGAGGCAATTCGGGAAGGCGTTGTCACCGTAGCTCGATTATCGTCATATAGG

ACAGATATGTAAGCCTGTGCGTGTGGCTTGTTAGACTGAATCAGAAAAGCGAGAAATGGA
GACTCACGCCCGGACAAATTTCCAGCACTTTCTTTAAGTCGGACAAAGATTGCCCACCTC
TCCAGAAAGCGATCAATGACACTGATGATCCAGTCACATCAGAGATGTCTTGTGCGTCAG
AGAGAGCTCCTAATACAAAGACTTTGCGAAGCACCGCCGCGGCATTGTGGTGGAACTTTG
CCCAGGATGAGCTGCTCGATACCACCAGGGCACTGAAAACGCCTAGGCAGCAGCCCACTG
TCTCTGTTTGGGCGGTAGAAGGTAATTGAAACAGCTTGCCATCTCCATTCGCTGACTGTG
ATGACTCGACATATTGTAGGTATTCCACAAGTTGTGCGATGATGACAAGTGGCCCAAGAA
TGGCGTTGGGCAAATTTGCAACCGTAGTTCGAGGAGTAATGACTCCTGTTCGAAGCCATT
CTGCTAGCCTACGGGCGTGGTCTGCGCCGGGAGTCTGTTGTAGTTTGGGAATTGATTCAG
ACAGAGCATCCCAGCCACTCTCGATGTCCTCGATAGCGCGGACAGCCCATTGGCTGGCAG
CATCATCTTTGACGAATGACAGAATCCGACTGAAATACGTCTGGTCAAGGCTCATGGCCC
CAGGCCCAAACAAAAGCAATGTCTTAGCACTCGGCATTGCCATAGTTGATGTGGTATCAG
GTTGGGAAGCTGGACTGACGATCAAGTGGCAGTGTTGAGTACCACAAGTCTTACTTCAAG
ATGTATTTCGACTACTCAGGTCAAGAGTGCCATGGGATGAGGGTTTAGAAGTAACCATAT
AGAAGCCCGGAGAAACCATCTCCTTTCTGTGGCCGCCATGACAGACTCAGACAACTAACC
TGCATTAGGAATATCGTCGGCCGTATTGCCAGACTGCGGCGCGGACTTGGAGAAAAACAC
GTCCCTGTGAACCGGAGTTTGATACCCTCGGCCGGAGTTTTCGGCCATATCCGGGGTCCA
TCGTCCTGGAGATAGCCATGGATGTAAAATGATGATATTCTGGGGCTGACCATCGGAGAT
GACCAACGGGAGAGGCAAAACTGTTAACTAGATACAGTGGGATCGAAGTATTGGGCTAAC
TACCTAGGTAGTTAACCCTTGTACCTAGCTAGGTGTACCGCCTTATATACCGCCTCGCTT
TTAATATACTGCTATGAAAATAAACAGTAAGCCGCTTATGCGACCGTGTACTTCTCCGGT
AGTATTCTCCCATAGTCTTTTTATGTATATTTTTAGTGGTCTTTTTGTAGCAGTCTCCCT
TTTTACATACCTTTGACTGCTTAGTGTTGCTAGATCATGTGGCTAGGTACATACTAACTG
CGAACACTTAACTAGAGTTTTAGTCCGGCAGAAGCGGCAGAATTTGGAGAGCGGTATATC
TAGGTACATCTTTCGGAGGTCCCACATACGGCACATCGTCCAAGGAGACATGATAGAGAG
ATTTTACAATGAAATTGATTGCTCGACTATGTGTACGGCAAATAGCTTCCAGGTCACTAA
CCTGGCGCGTTGAGTGGTTCACTTGGATAGGAAAATATATATATCTGGTGAATATCTTCA
AAATAAGTCTGGTTTCACCCCACTCAGATTGAGTAGTCGTTTCCACACCACAGGATAATC
CAGAGACTTGGGCCGAATACGGGAACTAATCTGCCTGCATGGGCTTTCATCCAGGACATG
TCGTGATGTGGCCTCATTGTTGTGGTTTCTCCGCAGGACTTGGAATAACTCCAACTCCGC
CTCACAGAATCACAGCATGTACCCCGACACCACTGATTACTCCGATCCCCAATGTCCATT
GTCCGGGAACGAGATAATGAATACTGAGCCATCCGCATGCCCTCTAATTTACCCAACGAA
GCCCTGGGTATCTAAGAGGCAGGGGCATAGATCTGGATTCCCACTGCAGGTTCGCCCAAA
TCCCATGGTATCTATCCCACGGCACGAAGGCCTATCCTTTCAACCTAGTTGTCGACATCT
AAACAACACAATATGGGAAGCATCGATAACACGGCACGCGGGTCATCCGCCAGCGAGCCG
ATCGCCATCATTGGCATGAGCGCCAAATTTGCTGGCGACGCAACCAACACCGACAATCTC
TGGCGCATGTTAATTGAGGGAAGAAGCGGTTGGTCGCCCTTCCCCGACTCTAGATTTCGA
TCAGAAGGCGTCTACCATCCAAATAATGAACGACTCAATAGTGTGAGGCCTCACCCCGCA
CTCCTACAGAAAGAAGTCTTAACAAAAGTCATTGCCTAGACACATGTGAAGGGTGCACAT
TTTCTCGCAGAGGATGTCGGACTGTTCGACGCGGCATTTTTTGGTTACTCGGGTGAAACA
GCGGCAGTGAGTGCTTTGCTGGGTCAGCTTCTTGGACTTGACGCCTACTGACCATGACTT
TAAGTCAATGGACCCGCAGTACAGACTCCAGCTCGAGTCCGTCTACGAGGCATTGGAAAA
TGGTGAGAGAAACGTTCCCTTCACAGCCTTTTTCTTTGAAAAGTCTTTAACCAAGCATCA
CAGCCGGTCTGCCATTGACGAAGATCGCTGGTTCCAACACCTCGGTGTTCACAGGAGTGT
TTGTACACGATTATAGAGATGGCTTACTCCGCGATGCCGACAACTTACCCCGGTTGATGG
CCACTGGCACGGGTGTTCCCATGATGGCGAACCGTGTATCACATTTCTTTGACCTGCGCG
GCGCCAGTATGACAATAGAGACGGCGTGCTCCTCGGGAATGGTGGCAGTGCATCAAGCCG
TTCAAAGCTTGAGGACTGGAGAGGCGGATATGTCCATTGTTGGCGGTGCCAACCTGACGC
TCAACCCAGATATGTTCAAGGCGCTAGGTTCTGCTGGGTAAGTCAAGTACCCTAACTACC
GTAATAGAACGAATGGCTGAATTGTTTATGAAAGGTTTCTTTCTGCCGACGGCAAGTCCT
ATGCATTTGATTCCCGCGCCAGTGGATATGGCCGCGGCGAAGGTGTTGGAACACTAGTCG
TGAAGCGCTTGTCAGACGCCCTTGCCGCAGGAGATCCGATTAGGGCTGTGATTCGAGAAT
CGATGCTCAACCAAGATGGCAAAACTGAAACAATTACGTCCCCGAGTCTAGAGGCACAAG
AAGCCTTGGTGCGCGGATGCTATCAAAAAGCAGGTCTCGACCCTCGAGAAACGCAATATT
TTGAGGCACATGGCACTGGTACGCAGGCCGGGGATACTATTGAGGCACAGGGCATTGCAA
CTGTTTTCGCATCACGTCAAGAACCATTGCTCATCGGATCGATCAAGACAAACGTTGGCC
ACACAGAGGCCGCAAGCGGACTTGCAAGCATCATCAAAACTGCACTAGCCATGGAAAATG
GAGTCATCCCGCCTTCTATCAACTTCGAGAAGCCTAACCCGAAGATCAGCTTGGATGATT
GGAATCTGAAGCTTGTTCGGGAAGTGGAAACATGGCCAGCGGGCCCCATCAGACGCGCAT
CAATCAACAACTTCGGATATGGAGGAAGCAATGCGCACATAATCTTAGAAGATAGCGCTT
CGTGGGTCAAGGCTATTGGTGGCCAGAATGGACGTACCAATGGGTTCGCGGATGGACATT
CGAACGGACCAAACGCAAATGGTCACCACTCCACGCTGGACCCACATGTGCAAGAAAGCC
AAGTTATCTCAAAGGTCCTTGTATTGAGTGGGAAGGACAAGCAGGCGTGCGAGAAAATGA
CAGCGAACCWTGCGGACTACCTGAGACAAACCCAGTCAACAAACTCCAATCCACGAGAGC
TCCTCGACAGTTTGATCTATACGCTAGGTCAACGGCGCAGCCGCTTCCCATGGGTAGTAG
CACATCCAATACCAGTTACGGAGGGGTATGAAACCGTAGTTCAGACTCTCCAGTCGCCCA
AATTCAAACCAACACGCACTTCGCGTCGACCTCGGATCGGTATGGTGTTTACAGGCCAGG
GGGCACAGTGGAATGCCATGGGAAGGGAGCTCATCGAGGCCTATCCCGTATTCAAAGCAT
CTCTTCAAGAGGCCGCAGGATATCTTGAACAGTTTGGCGCCGAATGGTCATTGATGGATG
AGTTAATGCGAGATGCCGAAAAAAGCCGCATCAATGAAGTCGGCTTGAGTACTCCGATCT
GCGTGGCAGTACAGATCTCGCTTGTGCGCTTGTTACGGGCTTGGGGAATCGTTCCTGTCG
CTGTTACCAGCCATTCGAGTGGAGAGATTGCCGCCGCTTACAGTGCGGGTGCCGTAAGTT
ACAAAACAGCTATGGCCTTTTCCTACTACCGTGCGGTGCTGGCGGCAGACAAGAGCCTAC

GCGGGCCAGTCAAGGGCGGCATGATTGCCGTCGGACTTGGATTAGAAGAGACGGAATCCT
ATCTTCGCCGGCTGAGCTCAGAGGGCCAAGCTGCCATAGCTTGCATCAACAGCCCGTCTA
GCATAACGGTCTCCGGTGACCTTTCGGCAGTGGTAGAGCTGGAGGATCTGGCCAATGCAG
ATGGTGTATTTGCTCGTCGTCTGAAGGTGGACACGGCCTGGCACTCGCATCATATGACTC
CAATTGCGAATGTCTATTGCGAAGCCTTGGAGAATACACGAGCTGAAAAGATTGACCGAG
ATGCTCTGACCACCGTTGCATTCTCATCTCCAGTAACTGGAGGTCGTATTACAGATGCTC
AACAGATCGCGCGCCCGGAGCACTGGGTTGAAAGCTTGGTACAGCCTGTGCAGTTTGTCG
CCGCTTTCACCGATATGGTACTCGGCGGCTCGGGATCTGTTGGCTCTAACGTGGATGTGG
TCGTTGAGGTGGGCCCGCATACAGCGCTGGGAGGCCCGATCCAGGAGATCCTTGGACTGC
CCGAGTTCAAAGATTTGAACATTCCATATTATGGAACTCTCGTTCGCAAATTAGACGCCC
GGGACAGCATGCATGCACTTGCTTCTAGTCTTCTACGAGAAGGCTATCCTGTTAATATGG
GAGCAGTGAATTTTGCACATGGGCGGGGACAGTACGTCAAAGTACTGACCAACCTACCAT
CGTACCCCTGGAACCACCAGGCAAAGCACTGGGCTGAGCCACGGCTAAATCGGGCCATAC
GTGAACGATCCCAGCCTCCTCATGACTTGCTCGGATCCATCGTCGAAGGCTCAAATCCAA
ATGCACCGTCTTGGCGACACATCCTTCGAATGTCCGAGTCACCGTGGACCAGAGATCACG
CTATTCAATCCAACGTCATCTATCCAGCCGCTGGGTACATTTGCCTGGCTATCGAGGCAA
GCCGTCAGCTTCATGTGCTCAATCAAACGGCCGGAGAGATTGGTGGATACCGGCTTCGCG
ACGTTGATTTCTTACAGGCCCTCATGATTCCGGATAGCTCAGACGGCATCGAGATTCAAA
CGACGATACGTCCAGTCAGCGAGAAGGACATTGCCTCGCAAGGATGGAGGCATTTCGAGG
TCTGGTCCGTTACAACAGACAACCGCTGGACCCAACACGCAAAAGGGTTGGTCTCTGTTG
AACTTGGAGAGTCTTCTGTCCGGATGTCCCGACCAGCTAGGAAGAACATTACTGGCTACA
CGCGGCGAATTCTTCCTGCTGATCTATTTGCCAACTTGAGGAATCTGGGGATTACACACG
GGCCGGTTTTCCAGAATATGGACAGCATCATCCAGTCTGGTTCTGAAATGCGAAGTGTGG
TGAGCATGACTTTGCCCGACGTCTCTGTTCCCAATGACCTTCCCCGAAACCACATTCTGC
ATCCCGTCACGTTGGACTCGGTAATCACGGCCCCTTACTCAGCAGTCCCTGGAGCTGCTG
CCCGTGAAATTACTGCCAAGGTGCCCAGGTCTGTTGAGAGATTCTGGGTATCCAGCAAGA
TAAGCCACGATGCGGGACATTCGTTGGAGGCAGACACGACGCTCATCCGCGATGATGATC
AAGGAATGGCGGCAGATGTGCTGGTTTCTGATCATGACACCGGAAATATTATGCTCGAAA
TGAATGGCTTTTCTTACCAGTCTTTGGGACGGAGCACGTCACTACAGAAATCCGAATCTT
GGCCGAATGAGCTGTGCAACAAGGTAGTTTGGTCGCTCGATATTTCCACGCCTTTGCCTG
CTACCTTGGCTGCAGTGAGAAATGAATTGGCCTGTACCGTCCAATCTGCTGAATGTGACA
CTACAAAAGCTACATTGCGTGCGTGTATCTACTTTATGCAACTGGCTCTCGTCGCTCTGG
ACTCGCACGACATAGCTGAGATGGAGCAACATAATGCGTCATACTATACATGGATGAAGG
ACACTGTCGAGCTGGCCAGCTCGGGAAAACTGTTCGAAGGTAGCGCCGAATGGTTATACC
ATTCAGAGAATGAGAGGCAGCTTCATATTGAACAGGTTCAGACCAGATTGGATGGGGAGA
TCGTGTGTCGGCTGGGAACTCAGCTGGTAGACATATTGCGCGGACATACCGGAGCACTCG
ACCTGGTCATGCAAGACAATCTGCTATCTCGTTTCTACAGCTATGCTCCACGGTGGAAAC
GGGCGGGGACGCAGATCGCAGGACTTCTTCGCCATCTCTCCCACAAGAATCCCCGTGCTC
GCATTTTGGAGGTGGGTGCGGCCACAGGCGCCATTGCACTCCATGCCCTTGGAGCCCTAG
GCACGTCTGACTCGGGTGGTCCCAATGCCTCCATGTACCACTTTACGGACACTTCTACGG
CTTTGTTCGAGACAGCGAGAGAAAGCCTGCAGCCCTGGGCTGATCTGCTGTCCTTCGATG
AACTCGACATTGAGCATGATCCAGCATCGCAGGGGTATACACCCGGGACCTACGATATAG
TGATCGCCTCAAATATCCGATCTATCTCTGAGTCCACATCGCAAGCGCTGAGCAACATCA
GCTCCCTGCTAAAGCCCGGCGGCACCCTCTTGCTGGTGGAACCTTTGAAATACGAGGTCG
ATGTTCACTTTGTCCGTCGGCTACTTCCTGGCCGGTGGTGGGACGATAGCACAGAGCTGA
AGGCAAACCTATGTCTGGATATGCCATCCTGGGAAAATCAACTCCTAAGCGCCGGTTTCA
CAGGTGTTGAACTCGAGTTGCTGGATCGTGAAGACCCCCAAGAAGCCGCTTTGGTGACTT
TCATGTCCACTGTGCAATTACCACAGCCACCAAAATCAAATGTGGATGCGGACCAAGTGG
TCATTGTCACAAGTCGAAACGGATGTCCTCCAGCTGCTTGGGTGAAGGGCCTCAAGGATG
CCATCGCTGCCTACACCGTCAGTGAAGGGAAACTAGGTCCCATTGTTCAGGATTTAGAAT
CCTTAGCTGCAACAGCTGCGTCTTATGCGGACAAGATCTGTATCTTCCTTGGCGAGGTGG
ATGAAGGCATCCTATACAACTTGAATTCGACATTATTGGAGGGAATTCGTTCAATGAGCA
CCAACTCTAAAGGATTGATCTGGGTGACGCGTGGTGGTGCTGTGGACTGTGAAAGGCCAG
AAATGAGTTTGGCGACCGGGTTCATTCGTTCCCTGCGCAATGAATATGTCGGTCGCAAGC
TGCTGACCCTCGACCTGGATCCCAAGGGAACACCGTGGTCTGACGTTAGTATGGCTGCAA
TTGCCAAGATTCTGGGTACGGTCATTGGGAATTCCGCCGGTGGCTCTATGGTAGAGAAAG
GTGCCGTGGAGCTCGAGTACGCCGAACGAGATGGCGTTATCTTGATCCCGCGAATCTACC
ATGATGTGACGAGAAACCGAATGCTTTCCCCCGATGCATCAGATGCCGCCATGGAGAAAA
TCTCAATTGAGAATTTCTACCAACCAACCCGCCCTTTGTGTTTAAAACCGGATTTGCTAG
TCTTCGGTGACGATGACTTCTCTGCCGATTATCTTGAACATCTCCCACCGGCATCCCTGG
AAGTGCAACCTAAGGCGTATGGTGCTACACTGAACAGTGTCGGTGATCATATCGCTGGCT
TTGAGTGTGCCGGAATAATTACGCAAGTTGGGGAAGAAGCAGCAGCCCAAGGCTATGCAG
TCGGTGATCGCGTTCTCTCAGTCTTGCGACATTCATCTTTTCCGAGCCGGGCTGTCGTCG
ACTGGAAACTGACGACGCGCATGCCAACTGACATGACCTTCCAGGAAGGAGCATCACTTC
CGTTGTCTTTCCTCAGTGCATACTTTGCCCTGGTCGAAATCGCGCGACTGCAGCGTTCTC
GGTCAGTCTTGATTCACGCTGGTGCTGGAGATGTTGGGCAAGCTGCAATCATGGTTGCCC
AGCATCTCGGGGCGGAGGTATATGTGACAGTTGGTAGCCCTGCAGAGCGTGGCCTGCTCA
TACTGAAATATGGTCTGCCGGCGGATCATATCTTCAGTTGTACAGACTTGTCACTTGCAA
ATGCGGTAGTAGCTGCGACACAAGGCCGTGGAGTTGACGTGGTTCTCAACTCACTGACTG
GCCCGCTCTTTCAAGAAAGTCTTAACCTTGTGGCCCCGCTTGGCCACTTTGTGGAGATTG
GCAGGCGCAATACCCAGACAAATGGCTATATGCACATGCGGCCATTCGATCGTGGCATTT
CATTCGCGACCCTTGATATACCTAGCTTGCTGGAGTATCGGGCGATGGATGTTCACCGTT
GCCTCGCTGAGTTGACACGTCTCATCGAGTTAAAAGCCGTGACACCTGTCCACCCAATCA

CCTTCCATGCCATAGGAGAGATCGCGGAGGCATCTCGTCTCTTAAAAGCGGGAGACCAGA
TTGGCAAAGTGGTCTTGTCGGTCGATGAGCATTCAACGGTTACTGCCGTGCCATCCAAGC
CGGCTGCAAAGCTCTCTTCCGAGGTCTCGTACTTGATCGTCGGTGGCAGTGGCGGCTTAG
CCCAGTCTGTGGCGCACTGGATGGTCAACCGTGGAGCAAGAAATCTGGTCCTTCTATCTC
GGAGTGCTGGGACGAGCGAGAAGACCGCTGCATTTGCCGAGGATCTCCGCCAGGCAGGAT
GTCGTCGAGTTTTGCCCATCAGCTGCGATGTTGCCAATGAGGAAAGTTTGGGCGACGCCA
TCAATCAATGTGCTCAAGAGGGCTTACCCCCCATCAGAGGTATCATTCATGCTGCGTTTG
TTCTTCGTGTAAGTGTCGAAAACTCCAACCGTAATCCTGAAATAAAACAACTAACCGCGT
ATCTAACAGGATGCCTTCGTGGAGAAAATGACCCTTGATGACTGGACATACACTATTCAG
AGCAAGGTCGCCGGCACCTGGAACCTGCACAACCAGTTTAATTTGCCCGGCGACCTCGAC
TTCTTCGTCTTGTTCTCTTCTATCAATGGAATTCTTGGATATGCCAGCCAGTCTGCATAT
TCCGCAGCTGGTGCATACGAAGACGCCCTCGCCCACTGGCGAGTCAAGCATTGTGGCCTT
CCCGCTGTGTCAATTGACCTCTCCGTCGTCAACGCGGTCGGCTACGTTGCCGAGGCAAAC
GCATCCGAAACACTACGCCGGTCTCTCCTCAGAGCTGGCCGCAGAGTCATCGATGAAGAT
CATGTTCTTGGCTCGCTAGAGTCTGCCATTCTATCGCCCTTCGACCCACAATTCGTGGTC
GGTGGTATAAACTCCGGGCCAGGTCCCCATTGGGATCTTGACGGCGATCTAGGCCGTGAC
ATGCGTGTCTTGCCACTCAAGTATCGCCCTCCGGCTGTAACCGGACAGAGTCAGGACGAT
GATTCTAGCAGCGACTCTCTCGCCGCAAAAATGATCGCCTGCGAGTCACAGGGCGATGCC
GTTCGTGTTGTTGGGACTGCCATCGCAGAAATGCTGGCTGAAATGTTCCTTGTTCCTATT
GAGGATGTCGATCTAGGCCAGTCCCCCTCGCAGCAAGGAGTTGACTCTCTTGTAGCAGTT
GAGGTCCGGAACATGCTTTTCAGCCAGGCCGGTGCTGAAGTTTCCATCTTCAATATCATG
CAAAGCCCCAGCCTGACGCAACTAGCGATTGATGTTGTGGATCGCAGTGCGCACGTCAAG
CTTGCCGGTTGATCATATTATTGGTCATGTCCGAGAAGCTGTCAATTTTTGTATAGACCA
CGGGAGCGATGTCAGGCTGCATCGGCAGCGCTAGTCTTGCTTCTATGGGTGATCAATATT
ATTGGAGAGCCGGAGTTGGGGTTTATATTGACTACATACTGAGGATTGATTGGGATTCTC
AACATTTCGTGGGCATCTTTCTCAAGTCTCCTCTGGAGCTTTTATATGAATAAATTCGTA
CAACTTCATGGAATGAAGTGGATAGCACTTCTAGGACAGGTAGATTTACATTTGCTTTGA
ATTTCTACACATTACGGCTTGTACCAGTCTCGACAATAGGGCAAATGTCAGTTGCCACTA
AACCAACCCTTCTCCAAAGTCCTCAGGCCCAAATATAAAATTGTCAAAGTCGAAATTGAT
ATCCTTGAAAATGTCGTCGAATGAAACTCCCGAAGACCCGTCTACCATCAAACCACACCC
ACCGCCATCCATACTGGGGTAAGGCTTGCCCGTGATCCTGGCTGCGAGGGTCTCAAACGT
CCACAGGCACTGCTGGACATAAGGACTCAAGTGCTCATGAGCGCGGAGCGCCAGCATAAC
ACTTTCCCAAGACTGGGAAACAGAATCTGTGAATAGCTCCGATCGAAACATAGCTGCGAG
GAAGTTGGCACCCGCGATGTGAAGGTAGTATATTCTGTACCACCACGGAAGGAGTCCAAT
CGGCTCATCCGGCTCAAGCGTCTCGTTAACTAATGACGCCACCTGTTGCGCGGCTTCGAT
GCACATGCTGGCGCTTTCACGGAGGAGTCGGTGGCTTAGGCTGGGTGATTTGTGTAAAGG
CTGGGTGTCTGGTTTCATCGCGTAAAATCGGGCGAGCATGGGTCTGTATAGAAAGATCCG
GTGGTGGAGGTAGCTACATAGAACGATAAAAATCGATTTTGGTGAGGATAAAGGGATTGT
TGGCTTACCGGAGATGAAGCAAGTATCTCTCAGCTCTAGACGACCTGTCGGTCACCATTC
TGAGGTTCCGAGGTTGCCAATCGCTGGGGATACTATTTTCCCAATCTTGAAGACATGCAT
CTAGTTGCAATGCCGCGTTATGGTACTCGTCTTGTTGAGATAGGAGTGGAGGGCGGAATC
TGGCTGCGAGACTAGTTCGATTCTGTATCTGTGCCAGTTGAATTTGGTTCCCGATTTCGT
GGAGTCTTAATCCCAAGGTATCATGCATCGCGTTTTCACCTAGTTGTTGGGGGCTACTCG
GTGGTGGGGACGGAATTAGCACCAATGCCGACGTCTTACCCAATGACCAGGAGATGCATC
TACCTGAGGTCAGCTGGAGAATACTATAGAAGAGCGATCACTCGATGACATACCGATCGA
GCGCAACACAGCTAGCCCATACTTTCTGTTTCAAGGCTTTGTCGTCACTGGAGCCTTTCA
ACAAGGGCGTCTCGGTGAGATGGCAGCACATGCTTTGCGCAATTCGCATCGCCAGACCTG
CTGTCGTCCATGTCTTCTGCTGGTTATTGGTGCAGTGGAGGTATCGATTCATCAGCATCA
GACACTGTACGAGCTCGAGTGATCCAGGTTTCCATAGCATAGACTCAGCTGGAAGCAGGA
CCCATGCGCGCTGAAAAAATCGATTTGCTTCCTCGTTTCTTTGGTGCAGCGGATTGAGTT
CCTGTCTTTGAACTGCCAGAGCAAATACTACATTGAGGATACCAAGCCAGAGGTCGTAGT
CTGCACAGAGTGGTGTTATGGGTGTTGAATATGCTTTTTCATAGTTTTCAAGAAAACGCT
GGCGATCTAGAACTGGCTCCATAGGGTCAACGTGTTGCCAGTATATGTCAATTAACTGAT
CGGCATATGCCCGGGGAGGCAATTCGGTCGACGAAGACAAGGCACAAGATTGAGAAGGCA
GGTTGAGGTCTCCAAATAGCGGCGCATCTACCAGTGGGACCTGATTACAGGTGGCCGGAA
TGAGACCTGCCCTTTCGTTTATGGCAGCGGCGACCTGGCCGGCAAACCGCCCATGTGCGG
TATAGTATGCCCGGTTTTTCTCCAGGTCAGGCTCTGCCTCTGGATCTTCATTCACATGGT
GAGACACAGGTATTGAATCGACCGTGGGAGGTGACGCAGCCACTGATTCTGGCGGAGTCT
CCAATGAAGGTGCCGAAGATATGGGAGGCACCAGCATAGGAACAGGCATGTGAACTGATG
ATGCTTGCCCACGAGGTTTATGCACTGTCTGCTTCTCGACGGCACCTTGGCCATAGGCAC
ACGTCTTGCCTCTAGCCTTGCATCTGGCGCATGCTTTATGGTATGTGTCAGCCTTATGCA
CGAACTCTGCAGGTCAATGAAACAATGAAGGGAGAGATACCTGGCCGAGCACCATCACAT
TTGATCTTCCTCGAGCGACATACTTGGCAAGCGCTTGTTACTCTTAGCCTCTGCCTGCGT
AAAGGCTTCGAGTGCGTATGCTGATCAATTATCACTTCATTCATCACGTCTTTATCGAAC
AGTTGGGGATGGAGAGTGGAGTGCCCCGTCAATCAGGGCCCCGGATCTACCTGATTGATA
ATCAGCAGTTCGCCAATCAGATAACTCGCCGCTTTTTCAATAGACCCGCAGCGGAAAGCT
CTGTTATTCGCTAGTCTGCTTTCACTTCCCGCAATTCTCGAAAGCAGCCAATGAGAAGTG
CATCCATTTTCCATGTGTCAACCAATGAAATAGCAGTCCGTCTTTGATGGCTTCCGCAGT
AAATAACTCCGCCGTCCGGTAAGCCGTTATATAGATCGAACGCTTTGAGAAAAAAAGAAG
AGTCTCCTGCTACACAGCCTAGAATTTAAGTCTATGCACTTAATATTCTCCAATTTTTTA
GACGGAGCTTCCTCCCTTCCCCTCTTCTTTTTCAGCCTTGTCCATAGTCAATACAGTGGT
ATTGATCTCTCCCCTTCCTCACTTGCACCGCTAAGACAAATCCATGTCTCATACAGAGCC
AAAGGCTCCTGTCAACACAGGCGAGGTTGAAAATGGTCACTTATACGATGGCTCCGGCAC

CGAAGATGACCCATTCATCGTGGAGTTCCAAAAAGACGACCCCGGCAATCCGATGAACTG
GGGTCAGTCCCGTAAGTGGTTCATTGCAGCCATTGCGACCCTCTCGGTCTTTGCCGTTAC
TTTTACCTCCTCTGCATACTCGGTATCGGCAAATGAAGTCTTCAAGGACTTTGACATCAG
CACCGAGGTCTTCATTGTCGGGCTTTCTCTCTTTGTGCTCGGATTTGCGATTGGTCCTGC
CGTATGGGGTCCTTTGGTAAGCCATTAAACCATTTTGCATTTTCATGGAGCCCAGCTAAC
CATACGTTGCTTCACAGTACGTTTCGATGGTGCCATCCTATTCACATACTATACACATTC
AGCTAACATTGGCTTAAGGTCCGAACTGTACGGAAGACAGATACTGTGGATTATCACTCA
CATCGCCATGGTCGCCTTTCTAGGAGGGTCCGCAGGCAGCCAAAACGTGGCCACGCTCCT
CATCCTGAGATTCTTTGCCGGCACATTTGGCGGCTCTCCACTCGTCAACTCTGGCGGAAC
AATTGCTGATCTCTTCCCACCTGCTCAGCGTGGTCTGGCATTGACCATCTATTGCGTTGC
GCCCTTCCTCGGCCCCATCTTGGGGCCAATCGTGGGCGGATTCGTGTCTGAGAGCGTCGG
GTGGAGATGGGTCCAGGGTGTCTGCGTGATCTTCATTGGCGTGGTCGGCATTCTGGGAAT
CGTCTTCATTCCTGAGACATATGGCCCGGTATTGCTTCAGCGACGGACACATCAACTGGC
CAAAGCTGATGGCAAGATCTACGTGAGCGTTTTGGAGAAGAACCAGGGAAAGAAGCTGCC
ATCGGAAGTCTTCAAGCGTGCTTTGTTCCGTCCCTGGATCTTCTTGTTCCTTGAGCCCAT
TGTCTTGATAGCGTCAGTTTACATGGCTATCATTTACGGCACGGTCTATATGTTCATGGG
TGCCATGCCCATCGTGTACAACGAGGACCGTGGTTGGAGCGTGGGCATCGGCGGACTGGC
GTTCTTGGGAATTGCTGTTGGCATCATCTTTGGCCTAGTTTATGCCATCTGGGACAACAA
CGTCCGCTACATGAAGCTTTTTGCGGCAAAATCTGCAAACCCCGAATCTCGCCTTCCACC
TGCAATTGTTGGAGGTGTTGCCCTCCCCATTGGCATGTTCGCCTTTGCCTGGACCAACTA
CCCCAGCATACACTGGTCTGTCAGTATAATACTGTCTGCACCGTTTGGATTCGGCTGCGT
GCTGGTCATCCTGCCTATCATGAACTATTTGATCGACACTTACACCATCTATGCGGCCTC
TGTCCTGGCTGCAGCTGCCATCTTCCGCTCAGTCGTGGGCGCTGTGTTTCCTCTTTTCAC
GACACAGATGTACCACAATTTAGGAATTCACTGGGCTTCCTCTATCCCAGCATTTTTGAC
TCTTCTGTGCATGCCATTCCCGTTGATCATGTATCGCTACGGTGAAGCGGTCCGGATGAA
GTGCAAGTACTCATTCGAAGCGGCAGAGATGATGAGGAAGATGCAATTGCAGCAAACAGC
TGCTGCTACTACTACAGAGAAGGACAAGGACTCTTCCTCTGAGTGAGAGCTGTTAGAAAT
GGGCCCCTTAGTCCGGCATCAATGGAGGGCCAAGGCTTGGGGGTCTATCATTGGAGGATC
GCCGCATATAAATATCAATATCTGCCCTAGAAAATAGTTAATTAACTGTTAAATCAACTC
AACTCATCGGAGTCCCTAGAAGTTGATTATCACCTCTCTAATGTGCATGTTCTCCTTACA
GACTAGCAATAATATTTATGTCTATCAAAGTGTTTCTAGTAGTTTCCCGGCTCTAGTCTT
TGAGATACAGTATCTGTGTACATAGTTGATATTGGGGTATCAAGCCAAGGTGTGATAGGG
TAAAGGAATCCTGTATGTCGAACTTGCTTAAGATTGCACTTTCGGGTTAGTCTGTAAATA
TCCGTGGACATATTATTGATTCATCTCGTGTCCGGAGTTGAGAATACTCCGTAGTGAGGC
TGTTTTATAGTTCCGGAATGTGCGTGCCAATGAGAATCGTCGAACTTGTTCCTTGCAGCG
GAGAAGAGTAGATCAACTGTCTGGATATTTTTTGGTCCCCGTTCGACCTCGGAGTTGTCC
CAAGGCAAGACCACGGAGTTCGGAGTTATAAGTAGCTATACAATGGTCTCGGAACTCTCC
CGCGCACAAGAGCCTCATATATAACCCAACGAGTTCAATGACGAAATAATCCTAGCTTTA
TAGAATTACAATCTCAAAATGCAGGCCGCCAGTGCATTTGCGACCTGTCTCTTGGCTTCA
GTGGGTGGCAACAGCAGCGCCGTCGCCTTTCCAAACCAAGCCAACTATTCTACTCTTGTC
GCGCCGTACAATTTCGATCTTCTCACCACCCCTTCCGCCATTGTCTGGCCCCAGGATACA
CAGCAGGTAGCAGCCGCCGTCAAATGTGCCGTGGATTCAGACATCAAAGTTCAACCGAAA
AGCGGCGGACACAACTATGGCAACTATGGATCAACCACCGGCGAGCTGTCTGTGAATCTG
GACAATCTGCAGCACTTCAGCATGGATGAAACCAGCTGGACTGCCAGATTGGGACCAGGT
AACCGCCTAGGCCGAGTCACCGAGCTCATGTATAACAACGGTGGTCGACATGTTCCACAT
GGGACGACCTTCACGGTGGGACTCGGTGGACACGCAACTGTTGGAGGTGCCGGAGCGGCA
TCACGAATGCACGGACTGTTGCTCGACTATGTGGAGGAGGTAGAGGTTGTTCTGGCCAAC
TCATCCATCGTCCGAGCATCCAAGTCTCACAACGAGGATCTTTTCTTCGCCGTGCGTGGT
GCCGCTTCAAGCGTTGGCATTGTGACCGACTTTTCTATACGCACCGAGCCCGTTCCCGTA
TCCAGCGTCACTTACTCTTATATCTGGGAGGGGACAGATCCAGCAGCCCGCGCAGAAGTA
TTCTTGACTTGGCAGTCATTGCTCGCCGGTGGCTCTTTGCCCCAACACATGGCCTACGAT
CTGGTTGCGACGGCGAATAGCATGATACTGGGTGGGGCTTACTTTGGCAGTCAGGAGGAC
TTCGAGGCCTTTAACCTCAGCAGCCACTTCAAAGTAGCGCCAGACGTCGCACATATAAAA
ACATATACCAACTTCTTCGACTTCTCCGCCGCCGCAAGCGCTCAGACTAAAGCAGCTGGG
ATTGCTTCCCCGTCGCATTTCTATGCCAAATCTCTCGTCTTCAATCAACAGACTCTCATA
CCTGATGATGCTGCTGAGGAGGTCTTCAAGTACTTGGCCACGACCAAGAATGGTACTGAT
TTGTACGCGGTCACTTTCGCCGCATTGGGTGGTGCGGTGAGGGACGTCTCGGCATCAGAA
ACAGCTTTCTATCACCGTGATGCTTCATACTTTATGTTCTCCTTTGGAAGGACTAGTGGA
GACCTGACTGACACGACGGTGCAGTTCCTGGATGGGCTGAGCGAAGTTCTGACTAGTGGG
CAACCCGATGCCTACTACGGCCAGTACGTGGGGAATGTCGATCCCAGACAATCAACTGAC
AAGGCATTGACGGGGTATTATGGCAAGAATCTGCATCGCCTGCAGCAGATCAAGTCTGCG
GTAGACCCGAATGATGTTTTCCATAACCAGCAGAGTATCCCACCTCTGTCCTAGTCTGTC
GGTGCCGGATTTCTGTTAGTCTTAAGATTGTCGTTGATCGTGCACAATTGAACGCCATTT
GTACTCATTGGGCAGAAATCACAATATACTATGGAGTATAACAGCCAACACACAAGAATT
TGTAATGGAATAAAATCACGGCATTATTCCATTCGACCAACAAAAAAACCAAGGTTACTG
TACATGGCCTAACGGAGACCATATCCCCGTTTGACAATTTGGTTTGTGACAAAAATATCC
GACATAGGGATTCTGTAGGTAACTTATGCAGTTGTTTCGGGTAGTCCATGAGCCCTGGCA
GGTGTATATGTACCTAAGCTACGTAGGTAGATAGTAATGTTAGTTCGACCAGACTGGAGT
AGCCCCAGCAGTATCTTTGCAGTAGTCAATGAATCAGACGTGTTTAAGGCTCCAGGGGGT
TGAAACAGCCCCTGGAATGAAGCATTGGAAGATTTGGGCAGAGGTTAGGGTATGATGAAA
ATAAACAAAACCGATGAGTATTCCAAGAAAAGTACGGAAACAGCTAACTACTCGCTAACT
ACCCTCCCATTATCCGCTGACCTCGTCACCTTACCCTCCAATCATGTAGGTACGGTCATG
TCTTTCCCGTCAACTCTGCATCTGTACAGTATGTAGTCTCAAAACGAAAATCAAAAGAGT

GCCCAATTCCATCTTCACAGCACGCGTCCAGCCGATGAGCTGGTGCTACGGTCATTCACT
TGAGTCTTCCTCAGGGACTTGCCTTGCTGGATAGACGCAAGGAGGGCGCCTCGGCCAGCG
CCACCGGCAGCAGCTGGGGGAGCAGGTGGAGGCGGGATGCCAGCTGGAGGGGGTGGCGGG
GGAGCAGAGGGGGCTCCGCCAGGGGGCGGGGGCGGAGGCATAGCAGGAGCACCAGCTCCC
CCTGGTGGGGGAGGCGGAGGTGGGGGAGGGGGAGGTACTGCAGATGGAGCGGCTGGTGGG
GAGGCGGAGTCGCTAGGGCTTGGTATAGCAGGAACAGGAGGCAGAGTCGCTGGGGGTGGG
GGCGGAGGTGCAACAGGGCTTTCCTGTACCGGGGTAGTGTATTTCGAGGGAGACACGTCT
TCGTCCATAGCACTCAGAGGACGTGGAGGCGCCATAGTGCCGAACAGGATGCTGGCAAGT
TGTTTGGCGCTGCCACCAGTTGGACGGTCGTCCTCATCATCAGACGAATCAACCTCGGAT
CCAGCAGCGGACCAGTCATCGTCTTCCTCGGGGCGAGCACGAGTCTTGCGTTCCAGGGGG
GCTGCACCGGTGAAGGTGGGCTTGATAGGTTCTGGCTGGGTCAAACGGTGGAAGGGGTTC
GTAGACTGAGCTGTGGTTGGCGGCGAGGTCGCTTGGGTGGGTTCGGAAGGCTGAGCCTTG
TGCTTGAAGTATGGGTTCTTGGACTCAGCATCGGGCGCGACCGCAGGAACACCACGGCTG
CTTTCGGGGGAGCTCACTTCGCTCGGCACACTTTCGGGGGCGGCARGKKCAGRAATMGCK
RYCGTGGRCACGGGAGGCGGGGTAGGGAGAATCTGGCTCTGCGTCGGTGTGCTGTACTGG
GGCGTAATGTCGGCAGGCCCTTCGTCATCGGACGATGACTCGTCAAGGCCCTCGAGTTCA
CGCTGCAGCTGTCGCTCACGTTCCTTAGCCGCCTCCAGTTCAACGCGTTGGGCTTCAAGC
ATAGCTTCCTGCTCTTTGGCCTGTCGGCTAGCCTCTTCCCTGCGGCGCTTCTCTTCCTGC
TTCTTAATCTTGCCCTGGCGGACTTGTTCCTCAAGAGCACGAAGTCGCTCGGCTTGGGCA
TCCTTCTCCCGTTGGAATTCCAGCTCCCGTTGGCTTGCCTCACCCCTGTAAATGTATTAG
CCAAGCCCACTACTTTTGTCTTTATCCGCTGTGGATTACTCACTCTAGACGCTTAGTCTC
TGCCTCCTGCACCTCTTGCTCTTCCTTGATAGCCTGCTCACGGGCAGCCTGGTCCGCTTT
GGATTCCTCTTCCTTCTTGGCATCGGCTTGGCCGGCACTGTCCGTCCGAGCCCTGCGAGT
GGGCGGCGCCGGTGGTGGTTTCTTGCCTGCAGCCTTGGGAGCTGCCTCACTAGGACCGCC
CCGTTCCTCGGCGAGGCGACGCTGTCTCTCCTGTTCCCGCTTAGCATCCTCTTCCTCGGC
GCGTTTGACACGATCTTCGCGCTCCTTCTTCTCACGCTCCTGGCGTTGCGCAAGGGTCTC
AGTAGCGTCATTGGGCTTCAGACCAGCAGCAGCCATGCGCTCAGCAATACGCTTCTGTGC
ACGCTCTCGAGCAGCTGCCACACGATCTTCATGCGTGCTACCCGGAAGAGAGCTTGAGGA
GTCCACGGACGGGGGCATCGAAGGGCGCGCAGCGGCAGATTCAGCCGCTGGAGATGCTCG
AGATTGCACATCCGGGGAAGGTCTCTCTCGCCTATGGATTGTCAGTAATGGACACTCTTT
TTGATTAAGGGCGAGAWTMRCRKMCKCTWMTTCTCSCKWATATTCGCTGTTCGCGCATTG
CGGCTCAGGTCGTAGATGAAGTCGCGGATGACGTCCTCAACTCCCAATGCCTCTTCCCAA
CGACGCTTCTCAYGCTCACGTGTGGAGTTCTGGCCTTCTTCCCGGAGACTGTCCTCCAAA
CTGCGAGCGAAGTCACGGACACTTTCCTCAACGTCGCGGGTCATGGTGTCATTGCGCTCC
CGCTCCATCTTCACGTTAGTGCTCTCAGACTCCAGACGACGAGCTGCTTGGCCATCGTCA
TCTTCCGTCGCGGGTGCAGGGCGCCCAGCAAGCTCGGCAGCCCGGGCTTGCATACGTGCA
CGCGCGCGAGCCTTGATTCGATCTGATTCAGTCACAGTACCGCCAGGGCCAGTGCCAATG
ATGTTAGAGGCACCACCGGGGTGTGCCTTGGCATCCTTGAGACGGAAGAGTTCGAGCTTC
GCATCGGCGATTTCACGCTCAAGACGACGCACATCAGATGCAACCTGGGGAACTTGATCT
TCAAAAGACTGCAATTGGCGACGAAGAGCCCTGCGATCTGCTCCACTGTCCACTTGGCGG
AAGGCAGCATTGGGGTGGGTATCGAGATCGTCCTGCACACGCCGGATACGGTCCATGAGA
CTCTCTGCCTCACGGCGATCCTGGCGGTCTAGAGCATCATCATCCTCGGCTCGGTTCTCG
TCCTCGAAGTCGGCAGCGTCAATCATAACTTTGGCTTCGCGGATCTTCTTGTGCAGCTGT
TCCACAGAATACTCCTCATCTGACGCGCCGGAGGCGGCTGGTGACGGGGTACGAGCCTCG
TTACCAACGCGACGACGTGCACTAGAGCGGTATCCACCGGCGGAGTCATTGTTCTTGAAC
ATCGTTGCATCCTTGCGGCCAGCCCCAGAAGCGCCTGCTGCACCACCACGGAATGAGTGA
TCTTTAAGGTAGCTGACACCAGTCTTCTGGGGCTGCAGGAACGCCCCGGTAGCCTTGCGA
GTTTCAGCATCCTGAGAGAGCAGAGACTTGACGGTACCGATAGAGTCATTTAAGTGTCTC
GTAGAAGGAGGGACAAGCTCGGGTGGTAGACGGCTGGGAACTGGGTACCCATTCAGAGCA
CGGTAGATCAAGTGCATAGCCACGGCGAACTCATCCATGTTTAGTCGGCCGCGGTTATGG
GGATCAGCCAGGGTCCAGATCCGTTCCAGATCTTTTCTATTCAGACCACTTTGGCCCATG
ATCTCGATGGCAGTCTCGCCACTGATGAAGCCCTTGCGGAAACCGTCCCATGCGCGGAAA
AGTTCATCGTAAATCTTCTTCTCCTCCTTCGTAATTGCCCAGGCAACTGTGGCATTTCCT
GAAAGGCCAACAGCACTGAACCCGCCTTCACGCCCAGGCTGCGGCATGAGCTGCTGCTTC
ATAGCATCGATGTTGGGCAGACCCTGCGCTGGGGTGTTGATGAATCCCCATTGACCTGGC
ATGCCCGTAGGTTGCGCTGCCAGACCTGTAGGCTGCACACTCAGGCCTTGCATTCCACCG
GTCTGGGCGGGGCTGAGGTTCGACCCGTAACCAGTAGGCATAGGCGGCATCGGAGGCCGA
GGGCCGGTATATCCAGTAGCCTGCGGGTTAGTCATGAAACCTGTCTGCTGAGGCTGCAGG
AATTGAGACTGGTTCTGGCCCGGGAAGCCCGTCTGGTTGGGCTGCTGGATACCAGTCTGC
TGGGGGAAGAACCCAGTAGGCTGTGCGGTCAGCTGCGACAGGAGCGTGGCATTGTTAGGT
TGCTGCGGAACGGGTTGCTGAACAACTGGGGGTGTCGACTTGTTCTCTATGAGTGGCGCA
TCGAAGTTGGGGACGTTTGTTCGTTGTTGGACCGGCGGCGGCTGTGTATCGGGAACATCG
AATGAGATGATGTCAACCATACTAGAAACCTCGTTCTTGATCGTCTCGGGTAGGGAAGAA
GGAAGATCGCGTCCGGTCAGACGAATGTTGCAGAGGTACATAGCCAATGCGAATTCGGGG
AATAGCAACTGTCCGGATTTAGTGGTATCAGACAAGACCCAGATCTTAGAAAGATCAGCG
CCCGACAACTTGGAGCGCAACAAAAGATCTCTTGCTTTGTCGCCTATTGTGACCAGTCAG
TACTACCATCCAAGCCATAACATGGGGCCCACTGGGTGTACGGAAACCTACCATCCATTG
TCTTGCTATCGCCAACCGCAGACTTGAAGAGCTGCTCGAACCTCGCCTGATCCTGCGCCG
TGATAAAAGAAAGCCGAATGCTCGGGATCTTTGATCCTGTCTTTGGCGGCACCTGAGGAG
GCGCCCCAGATCCAGCAGCATCTTGGAAGGAGTTTGCGATATCGCTAGAGGTCTTGAATC
GAGTTGGTATCTGCGGCACCGGAGGAGGCTGCTGTGCTTGGTTTAGAGGAGGGTATCCAG
TGTATTGTTGCTGGAAGCTTTGAGCTTGCGGCTGCTGAGGTTGCTGGGGCGCGGCACCGG
GGAAGCCAGTGAATTGCGGCTGAAGCTGTCCTGCGGGGAATCCGGTGGCTTGAGGCTGTA

GCTGCGAGCTACCGAACGGGGAGGGCTGGCCGGCAAATCCGGTCGGCTGTGGAGCAATAC
CAGTGGGCTGCGGAGCAAATCCAGTCGGTTGAGGTTGTTGGATTGATGGCGGCTGCTGGC
CTTGGGGGAATTGCGAATAGGGGGGTTGTTGTTGCTGTTGCATGAATGGTGCCTGGCCGG
GACGGCCAGTGGCACCACCCCCGAGGAAATTATTCGACGAGGAAAACATCAGAGACACCG
CGCTTGCATGAGCAAAAGGATGCAGTAAAGCCAACAAGAAATTGACAAGAAAAACGAATG
CAAAGTAGTAAGTGAGTTGAGTCTCAGGAGGATGCGAATGACGCCGTTGCTGGACGGGGA
CGCACCTGCGCTAACCTCATCGGGCGCTTAGTCATGTGCTCGGGATATTTTTGGATCTAC
GTCTGGAATACTGGGTTGTGGATACAAGTGATGATGATTTAAAGCTTTATACTTGTAAAT
TTGTCGGCATCCGCTGAACGATGGGGATAACTGAGCCTCTCAGAAACTGCAAAAAGCAAA
AAGAAAAGAAAACGCGCCCGTAAAAAAAAAAAAGGGGTGGGATCCCCGGGTACCGAGCTC
GAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNCTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCTGAAT
GGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGC
ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACAC
CCGCCAACACCCGCTGACGCGAACCCCTTGCGGCCGCATCGAATATAACTTCGTATAATG
TATGCTATACGAAGTTATTAGCGATGAGCTCGGACTTCCATTGTTCATTCCACGGACAAA
AACAGAGAAAGGAAACGACAGAGGCCAAAAAGCTCGCTTTCAGCACCTGTCGTTTCCTTT
CTTTTCAGAGGGTATTTTAAATAAAAACATTAAGTTATGACGAAGAAGAACGGAAACGCC
TTAAACCGGAAAATTTTCATAAATAGCGAAAACCCGCGAGGTCGCCGCCCCGTAACCTGT
CGGATCACCGGAAAGGACCCGTAAAGTGATAATGATTATCATCTACATATCACAACGTGC
GTGGAGGCCATCAAACCACGTCAAATAATCAATTATGACGCAGGTATCGTATTAAKTGRT
MKGCATCAAMTYMACGTAAAAKCAAYTTCAGACAWTAMWWATCWGCRWCACTKAATACRR
RGCAACYTCATGWCMSARMTCGCGASCTCGWMKACRGCGRCACACYTGCATCGGATSCAG
CYCGGYKARCKYGYCGRCASSGMCWSRSTWRCMWSGKATKYWGYCCRSWTAACCGTGCGS
MAMRKSYTGTGGATAASCAGGACACWGCMRCAATCCACAGCAGGMATASWWSYGSAYAMC
SAGGWTACTSCGTTCTACAGGTTACGACGACAYGWSRATACTTGCCSTTKACAGGCATTG
ATGGAATCGWARTSTCACGMTRATWGTCTSWTCGACARYAYWRRTGGRAYCGTRGTCYCA
SRCYGATARTCWGAYCGACAAYACRAGTGGGAYCGTGGTCCCAGACYRATAATCAGACCG
ACRAYACGAGTGGGAYCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACC
GTGGTTCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATA
ATCAGACCGACGATACGAGTGGGACCATGGTCCCAGACTAATAATCAGACCGACGATACG
AGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACGAGTGGGACCGTGGTCCC
AGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACC
GACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAGACCGACGATACAAGTGGAAC
AGTGGGCCCAGAGAGAATATTCAGGCCAGTTATGCTTTCTGGCCTGTAACAAAGGACATT
AAGTAAAGACAGATAAACGTAGACTAAAACGTGGTCGCATCAGGGTGCTGGCTTTTCAAG
TTCCTTAAGAATGGCCTCAATTTTCTCTATACACTCAGTTGGAACACGAGACCTGTCCAG
GTTAAGCACCATTTTATCGCCCTTATACAATACTGTCGCTCCAGGAGCAAACTGATGTCG
TGAGCTTAAACTAGTTCTTGATGCAGATGACGTTTTAAGCACAGAAGTTAAAAGAGTGAT
AACTTCTTCAGCTTCAAATATCACCCCAGCTTTTTTCTGCTCATGAAGGTTAGATGCCTG
CTGCTTAAGTAATTCCTCTTTATCTGTAAAGGCTTTTTGAAGTGCATCACCTGACCGGGC
AGATAGTTCACCGGGGTGAGAAAAAAGAGCAACAACTGATTTAGGCAATTTGGCGGTGTT
GATACAGCGGGTAATAATCTTACGTGAAATATTTTCCGCATCAGCCAGCGCAGAAATATT
TCCAGCAAATTCATTCTGCAATCGGCTTGCATAACGCTGACCACGTTCATAAGCACTTGT
TGGGCGATAATCGTTACCCAATCTGGATAATGCAGCCATCTGCTCATCATCCAGCTCGCC
AACCAGAACACGATAATCACTTTCGGTAAGTGCAGCAGCTTTACGACGGCGACTCCCATC
GGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGTGTCTGTTGACCAGTCAG
TAGAAAAGAAGGGATGAGATCATCCAGTGCGTCCTCAGTAAGCAGCTCCTGGTCACGTTC
ATTACCTGACCATACCCGAGAGGTCTTCTCAACACTATCACCCCGGAGCACTTCAAGAGT
AAACTTCACATCCCGACCACATACAGGCAAAGTAATGGCATTACCGCGAGCCATTACTCC
TACGCGCGCAATTAACGAATCCACCATCGGGGCAGCTGGTGTCGATAACGAAGTATCTTC
AACCGGTTGAGTATTGAGCGTATGTTTTGGAATAACAGGCGCACGCTTCATTATCTAATC
TCCCAGCGTGGTTTAATCAGACGATCGAAAATTTCATTGCAGACAGGTTCCCAAATAGAA
AGAGCATTTCTCCAGGCACCAGTTGAAGAGCGTTGATCAATGGCCTGTTCAAAAACAGTT
CTCATCCGGATCTGACCTTTACCAACTTCATCCGTTTCACGTACAACATTTTTTAGAACC
ATGCTTCCCCAGGCATCCCGAATTTGCTCCTCCATCCACGGGGACTGAGAGCCATTACTA
TTGCTGTATTTGGTAAGCAAAATACGTACATCAGGCTCGAACCCTTTAAGATCAACGTTC
TTGAGCAGATCACGAAGCATATCGAAAAACTGCAGTGCGGAGGTGTAGTCAAACAACTCA
GCAGGCGTGGGAACAATCAGCACATCAGCAGCACATACGACATTAATCGTGCCGATACCC
AGGTTAGGCGCGCTGTCAATAACTATGACATCATAGTCATGAGCAACAGTTTCAATGGCC
AGTCGGAGCATCAGGTGTGGATCGGTGGGCAGTTTACCTTCATCAAATTTGCCCATTAAC
TCAGTTTCAATACGGTGCAGAGCCAGACAGGAAGGAATAATGTCAAGCCCCGGCCAGCAA
GTGGGCTTTATTGCATAAGTGACATCGTCCTTTTCCCCAAGATAGAAAGGCAGGAGAGTG
TCTTCTGCATGAATATGAAGATCTGGTACCCATCCGTGATACATTGAGGCTGTTCCCTGG
GGGTCGTTACCTTCCACGAGCAAAACACGTAGCCCCTTCAGAGCCAGATCCTGAGCAAGA
TGAACAGAAACTGAGGTTTTGTAAACGCCACCTTTATGGGCAGCAACCCCGATCACCGGT
GGAAATACGTCTTCAGCACGTCGCAATCGCGTACCAAACACATCACGCATATGATTAATT
TGTTCAATTGTATAACCAACACGTTGCTCAACCCGTCCTCGAATTTCCATATCCGGGTGC
GGTAGTCGCCCTGCTTTCTCGGCATCTCTGATAGCCTGAGAAGAAACCCCAACTAAATCC
GCTGCTTCACCTATTCTCCAGCGCCGGGTTATTTTCCTCGCTTCCGGGCTGTCATCATTA
AACTGTGCAATGGCGATAGCCTTCGTCATTTCATGACCAGCGTTTATGCACTGGTTAAGT
GTTTCCATGAGTTTCATTCTGAACATCCTTTAATCATTGCTTTGCGTTTTTTTATTAAAT

CTTGCAATTTACTGCAAAGCAACAACAAAATCGCAAAGTCATCAAAAAACCGCAAAGTTG
TTTAAAATAAGAGCAACACTACAAAAGGAGATAAGAAGAGCACATACCTCAGTCACTTAT
TATCACTAGCGCTCGCCGCAGCCGTGTAACCGAGCATAGCGAGCGAACTGGCGAGGAAGC
AAAGAAGAACTGTTCTGTCAGATAGCTCTTACGCTCAGCGCAAGAAGAAATATCCACCGT
GGGAAAAACTCCAGGTAGAGGTACACACGCGGATAGCCAATTCAGAGTAATAAACTGTGA
TAATCAACCCTCATCAATGATGACGAACTAACCCCCGATATCAAGTCACATGACGAAGGG
AAAGAGAAGGAAATCAACTGTGACAAACTGCCCTCAAATTTGGCTTCCTTAAAAATTACA
GTTCAAAAAGTATGAGAAAATCCATGCAGGCTGAAGGAAACAGCAAAACTGTGACAAATT
ACCCTCAGTAGGTCAGAACAAATGTGACGAACCACCCTCAAATCTGTGACAGATAACCCT
CAGACTATCCTGTCGTCATGGAAGTGATATCGCGGAAGGAAAATACGATATGAGTCGTCT
GGCGGCCTTTCTTTTTCTCAATGTATGAGAGGCGCATTGGAGTTCTGCTGTTGATCTCAT
TAACACAGACCTGCAGGAAGCGGCGGCGGAAGTCAGGCATACGCTGGTAACTTTGAGGCA
GCTGGTAACGCTCTATGATCCAGTCGATTTTCAGAGAGACGATGCCTGAGCCATCCGGCT
TACGATACTGACACAGGGATTCGTATAAACGCATGGCATACGGATTGGTGATTTCTTTTG
TTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAACCCGATAAAGAAGGGAATGAGAT
ATGGGTTGATATGTACACTGTAAAGCCCTCTGGATGGACTGTGCGCACGTTTGATAAACC
AAGGAAAAGATTCATAGCCTTTTTCATCGCCGGCATCCTCTTCAGGGCGATAAAAAACCA
CTTCCTTCCCCGCGAAACTCTTCAATGCCTGCCGTATATCCTTACTGGCTTCCGCAGAGG
TCAATCCGAATATTTCAGCATATTTAGCAACATGGATCTCGCAGATACCGTCATGTTCCT
GTAGGGTGCCATCAGATTTTCTGATCTGGTCAACGAACAGATACAGCATACGTTTTTGAT
CCCGGGAGAGACTATATGCCGCCTCAGTGAGGTCGTTTGACTGGACGATTCGCGGGCTAT
TTTTACGTTTCTTGTGATTGATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGT
GACAAGTTTTTAGATTGTCACACTAAATAAAAAAGAGTCAATAAGCAGGGATAACTTTGT
GAAAAAACAGCTTCTTCTGAGGGCAATTTGTCACAGGGTTAAGGGCAATTTGTCACAGAC
AGGACTGTCATTTGAGGGTGATTTGTCACACTGAAAGGGCAATTTGTCACAACACCTTCT
CTAGAACCAGCATGGATAAAGGCCTACAAGGCGCTCTAAAAAAGAAGATCTAAAAACTAT
AAAAAAAATAATTATAAAAATATCCCCGTGGATAAGTGGATAACCCCAAGGGAAGTTTTT
TCAGGCATCGTGTGTAAGCAGAATATATAAGTGCTGTTCCCTGGTGCTTCCTCGCTCACT
CGACCGGGAGGGTTCGAGAAGGGGGGGCACCCCCCTTCGGCGTGCGCGGTCACGCGCACA
GGGCGCAGCCCTGGTTAAAAACAAGGTTTATAAATATTGGTTTAAAAGCAGGTTAAAAGA
CAGGTTAGCGGTGGCCGAAAAACGGGCGGAAACCCTTGCAAATGCTGGATTTTCTGCCTG
TGGACAGCCCCTCAAATGTCAATAGGTGCGCCCCTCATCTGTCAGCACTCTGCCCCTCAA
GTGTCAAGGATCGCGCCCCTCATCTGTCAGTAGTCGCGCCCCTCAAGTGTCAATACCGCA
GGGCACTTATCCCCAGGCTTGTCCACATCATCTGTGGGAAACTCGCGTAAAATCAGGCGT
TTTCGCCGATTTGCGAGGCTGGCCAGCTCCACGTCGCCGGCCGAAATCGAGCCTGCCCCT
CATCTGTCAACGCCGCGCCGGGTGAGTCGGCCCCTCAAGTGTCAACGTCCGCCCCTCATC
TGTCAGTGAGGGCCAAGTTTTCCGCGAGGTATCCACAACGCCGGCGGCCGGCCGCGGTGT
CTCGCACACGGCTTCGACGGCGTTTCTGGCGCGTTTGCAGGGCCATAGACGGCCGCCAGC
CCAGCGGCGAGGGCAACCAGCCGAGGGCTTCGCCCTGTCGCTCGACTGCGGCGAGCACTA
CTGGCTGTAAAAGGACAGACCACATCATGGTTCTGTGTTCATTAGGTTGTTCTGTCCATT
GCTGACATAATCCGCTCCACTTCAACGTAACACCGCACGAAGATTTCTATTGTTCCTGAA
GGCATATTCAAATCGTTTTCGTTACCGCTTGCAGGCATCATGACAGAACACTACTTCCTA
TAAACGCTACACAGGCTCCTGAGATTAATAATGCGGATCTCTACGATAATGGGAGATTTT
CCCGACTGTTTCGTTCGCTTCTCAGTGGATAACAGCCAGCTTCTCTGTTTAACAGACAAA
AACAGCATATCCACTCAGTTCCACATTTCCATATAAAGGCCAAGGCATTTATTCTCAGGA
TAATTGTTTCAGCATCGCAACCGCATCAGACTCCGGCATCGCAAACTGCACCCGGTGCCG
GGCAGCCACATCCAGCGCAAAAACCTTCGTGTAGACTTCCGTTGAACTGATGGACTTATG
TCCCATCAGGCTTTGCAGAACTTTCAGCGGTATACCGGCATACAGCATGTGCATCGCATA
GGAATGGCGGAACGTATGTGGTGTGACCGGAACAGAGAACGTCACACCGTCAGCAGCAGC
GGCGGCAACCGCCTCCCCAATCCAGGTCCTGACCGTTCTGTCCGTCACTTCCCAGATCCG
CGCTTTCTCTGTCCTTCCTGTGCGACGGTTACGCCGCTCCATGAGCTTATCGCGAATAAA
TACCTGTGACGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCG
GGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCA
ACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTT
CAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATAT
CCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATM
ACSWGWYCATGCACCGGATCTGTTGCCGCAGCCAGTGGTAGATCCAGATATACGCAACCG
CTGTTGGGATGACAAAAAGGTCGATGCGCACCACGCCATCATTCCGACCGCACGGAGTTC
TGCGATCAACCTGACGGAGAACGAAGCGAAGGTCTATAACCTGATTGCCCGTCAGTATCT
GATGCAATTCTGCCCGGATGCGGTGTTCCGCAAGTGTGTTATCGAACTGGACATTGCCAA
AGGCAAATTTGTCGCTAAAGCGCGTTTTCTTGCTGAAGCAGGCTGGCGCACGCTGTTAGG
CAGCAAAGAGCGCGATGAAGAAAACGACGGCACGCCACTGCCTGTGGTGGCGAAAGGCGA
TGAGTTGCTGTGTGAAAAAGGTGAAGTGGTAGAGCGGCAAACCCAGCCGCCGCGCCATTT
TACCGATGCAACACTGCTTTCGGCGATGACCGGGATCGCGCGCTTTGTGCAGGATAAAGA
TCTGAAAAAGATCCTCCGTGCGACCGATGGTCTGGGGACAGAGGCAACGCGTGCCGGGAT
TATTGAACTGTTGTTCAAGCGTGGTTTCCTGACCAAAAAAGGGCGCTATATCC

4. ORF Analysis

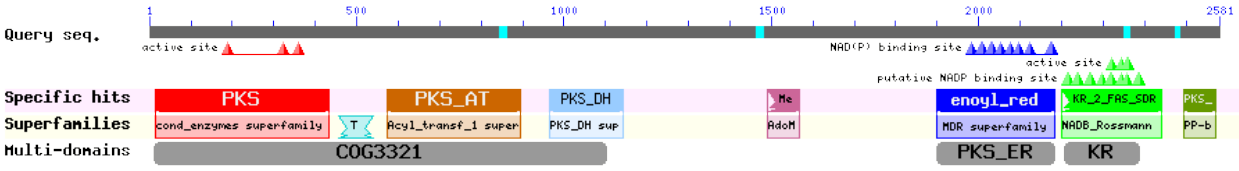
Putative ORFs were determined using fgenesh[2]

ORF	Orient-ation	Start	Stop	PBlast closest homology	Assignment
orf1	+				
exon 1		5162	5406	predicted protein	fungal transcription
intron 1		5407	5468	*Trichoderma*	factor regulatory
exon 2		5469	6066	*reesei* QM6a	middle homology
intron 2		6067	6134	EGR52693.1	region
exon 3		6135	6953		
orf2 (sorbC)	-				
exon 1		8632	8131	predicted protein	FAD-binding
intron 1		8130	8065	*Trichoderma*	monooxygenase
exon 2		8066	7450	*reesei QM6a*	
intron 2		7449	7381	EGR52183.1	
exon 3		7380	7231		
orf3 (sorbB)	-				
exon 1		17688	17580	polyketide	Non-Reducing fungal
intron 1		17579	17410	synthase	iterative polyketide
exon 2		17409	16748	*Trichoderma*	synthase
intron 2		16747	16695	*reesei* QM6a	
exon 3		16694	12490	EGR52182.1	see section 5 for
intron 3		12489	12441		domain analysis
exon 4		12440	9422		
orf4 (sorbA)	+				
exon 1		18733	18942	polyketide	Highly-Reducing
intron 1		18943	18999	synthase	fungal iterative
exon 2		19000	19086	*Trichoderma*	polyketide synthase
intron 2		19087	19144	*reesei* QM6a	
exon 3		19145	19202	EGR52690.1	see section 5 for
intron 3		19203	19263		domain analysis
exon 4		19264	19597		
intron 4		19598	19654		
exon 5		19655	25868		
intron 5		25869	25929		
exon 6		25930	26772		

ORF	Orient-ation	Start	Stop	PBlast closest homology	Assignment
orf5	-				
exon 1		28969	28134	fungal specific transcription factor, putative	Transcription factor
intron 1		28133	28080	*Neosartorya fischeri* NRRL 181	
exon 2		28079	27729	XP_001258840.1	
intron 2		27728	27674		
exon 3		27673	27118		
orf6	+				
exon 1		29744	30076	hypothetical protein AN3270.2	Major facilitator
intron 1		30077	30220	*Aspergillus nidulans* FGSC A4	superfamily transporter
exon 2		30221	31426	XP_660874.1	
orf7	+				
exon 1		32119	33534	glucooligosaccharide oxidase, putative (AFU_orthologue; AFUA_6G14340) *Aspergillus nidulans* FGSC A4 CBF83064.1	FAD-binding oxygenase
orf8	-				
exon 1		38809	38212	Actin cytoskeleton-regulatory complex protein	Regulator
intron 1		38211	38144	PAN1 *Penicillium digitatum*	
exon 2		38143	36140	Pd1 EKV05437.1	
intron 2		36139	35967		
exon 3		35966	35444		
intron 3		35443	35386		
exon 4		35385	35098		
intron 4		35097	34978		
exon 5		34977	34215		

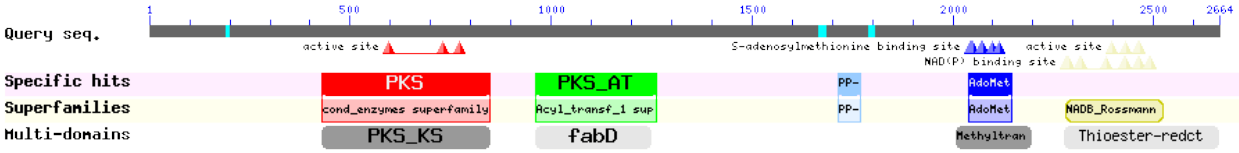
5. PKS Domain Analysis

5.1 SorbA



Domain SorbA	Approx. Predicted Boundaries	Conserved Active Site Motif	ref
KS	S14 - D436	185 - T(A/G)CX_2(G/S) - 190 - active site cysteine	3
AT	F575 - T891	662 - G(S)HSXG - 666 - active site serine	4
DH	P965 - D1251	966 - HX_8PXAGY(I/Y) - 1012	5
CMeT	Q1265 - Q1657	1459 - (R/K)ILEGX_2TG - 1468 - SAM binding	6
ER	N1844 - P2196	32% identical, 52% similar to rat FAS ER in this region.	5
KR	F2204 - Q2476	2209 - YX_3GX_2GGXG - 2216 Rossmann fold [NAD(P) binding]	7
ACP	Q2495 - G2581	2531 - GXDSLXA - 2537 phosphopantetheine attachment site	8

5.2 SorbB



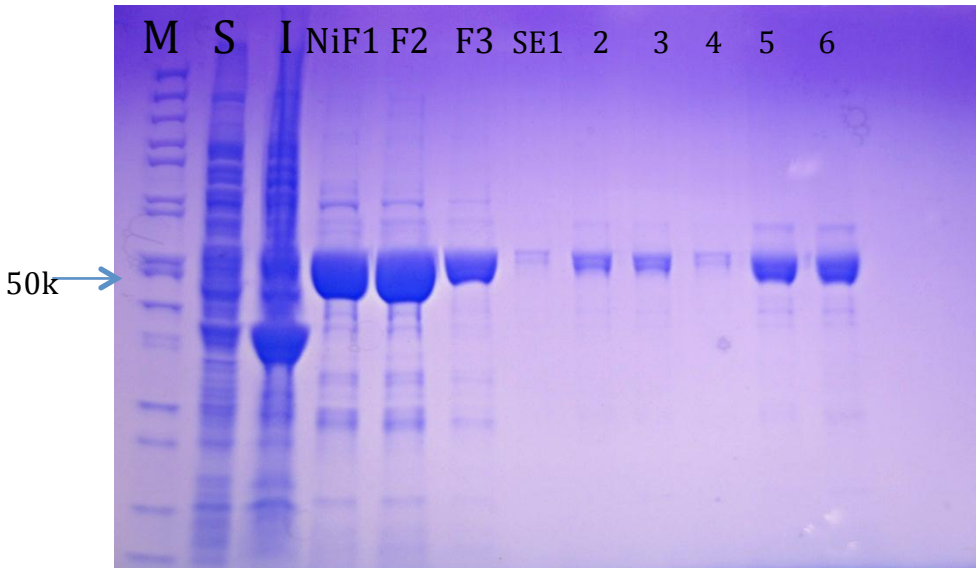
Domain SorbB	Approx. Predicted Boundaries	Conserved Active Site Motif	ref
SAT	L54 - L311	182 - GXCXG - 186 - active site cysteine	9
KS	D430 - Q849	594 - (A/G)CXX(G/S) - 599 - active site cysteine	3
AT	F962 - V1276	1045 - GHSXG - 1049 - active site serine	4
PT	C1361 - R1701	1379 - HX_9(T/P) - 1389 1562 - DX_3(H/Q)X_6N - 1573 - active site residues	10
ACP	L1717 - L1777	1742 - GXDSL - 1746 phosphopantetheine attachment site	8
CMeT	L1977 - H2193	2045 - GXGXG - 2049 - SAM binding	6
R	T2257 - V2593	2283 - GXXGXXG - 2289 Rossmann fold [NAD(P) binding] 2461 - YX_3K - 2465 - active site residues	11

6. Cloning, Expression, Purification and Characterisation of *sorbC*

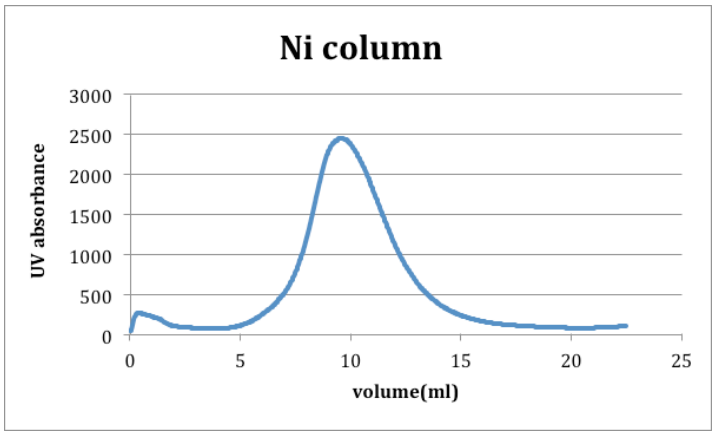
Mycelial samples were collected by Buchner filtration from a fermentation shown to be producing sorbicillin by LCMS analysis. The mycelia was immediately flash frozen in liquid nitrogen and stored at -80 °C until required. Mycelia were finely ground under liquid nitrogen and a 75 mg sample was used with the Qiagen RNeasy Plant Mini kit following the standard protocol using RLC buffer and omitting the on-column DNAseI digest. Two elutions of 40 mL and 20 mL were combined and contaminating DNA was digested by addition of DNAseI (Fermentas, 3 mL) and 10x DNAseI buffer (Fermentas, 7 mL) followed by incubation at room temperature for 20 min. DNAseI was inactivated by heating to 75 °C for 5 min and the RNA solution was immediaietly flash frozen in liquid nitrogen and stored at -80 °C until needed. Complimentary DNA (cDNA) was synthesised from total RNA (1 mL) using the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermantas) by the standard protocol (at 10 ml scale), using oligo$(dT)_{18}$ primers, and stored at -20 °C until required. The cDNA was used in a standard PCR reaction using the following primers: 5'-CATATGACTCGATCCGCCAACAGC-3' and 5' - GAATTCTCAGAATGCAGTGTCTGCGCC-3'. The PCR product was digested with *Nde*I and EcoR1 and ligated into pET28a(+) which had previously been digested with the same enzymes and dephosphorylated. The ligation product was transformed into *E. coli* DH5a. Selected transformants were grown and used for plasmid DNA miniprep. Insert sequence was confirmed by sequencing.

The pET28a/sorbC plasmid was transformed into competent *E. coli* BL21 (Stratagene) which was then grown in 2TY medium (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl) with kanamycin (50 μg/ml) at 37 °C, 250 rpm until the OD_{600} was ~ 0.6 – 1.0. The temperature was then dropped to 16 °C and cultures were induced with 0.5 mM IPTG, incubated overnight at 300 rpm. The cells were harvested by centrifugation at 5000 rpm for 30 min at 4 °C and resuspended in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM imidazole, 5% glycerol, pH 7.5). Subsequently, the cells were sonicated (6 x 30 s). Centrifugation at 18000 rpm for 20 min at 4 °C was then performed to remove cell debris and the supernatant was collected.

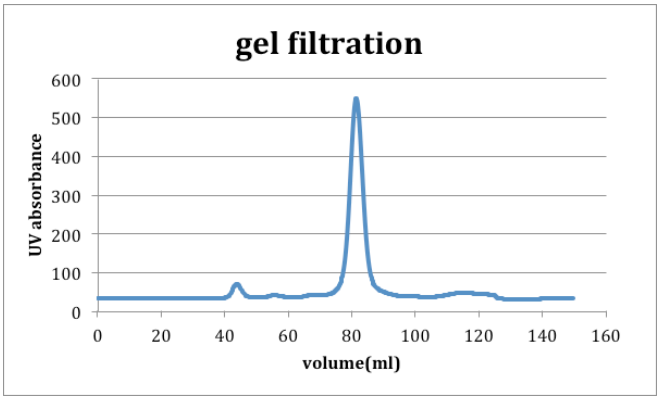
Protein was purified form the soluble fraction using a nickel-loaded chelating affinity column (HisTrap HP 5 mL column, GE Healthcare). Contaminants were removed using 20 column volume washing buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 5 mM imidazole, 10% glycerol). His-Tagged protein was eluted from the column by running a gradient of elution buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 500 mM imidazole, 10% glycerol) using an AKTA purifier, detecting protein at 280 nm. Eluted fractions were pooled at 200 mM imidazole and verified by SDS-PAGE analysis, then the pure fractions of the desired protein were concentrated using an Amicon Ultra centrifugal filter unit (Millipore) and then purified further using a HiLoad 26/600 Superdex 200 (GE Healthcare) column pre-equilibrated with 150 ml of 50 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, pH 7.5 filtered and degassed, the protein pass through the column by running 150 ml buffer, 6 fractions (2ml each) were collected and concentrated using an Amicon Ultra centrifugal filter unit (Millipore) and concentration was determined using BCA assay (Thermo). A total of 20 mg of protein was produced per litre of fermentation. For long-term storage, protein was stored at -80 °C in Tris pH 7.5, 150 mM NaCl, 5% glycerol, pH 7.5.



SDS PAGE analysis of SorbC purification. M = protein regular unstained marker; S = soluble fraction; I = insoluble fraction; Ni F1 = fraction 1 collected from column; Ni f2 fraction 2 collected from column; Nif3 fraction 3 collected from column; SE1 fraction 1 collected from size exclusion column; SE2 fraction 2 collected from size exclusion column; SE3 fraction 3 collected from size exclusion column; SE4 fraction 4 collected from size exclusion column; SE5 fraction 5 collected from size exclusion column; SE6 fraction 6 collected from size exclusion column.



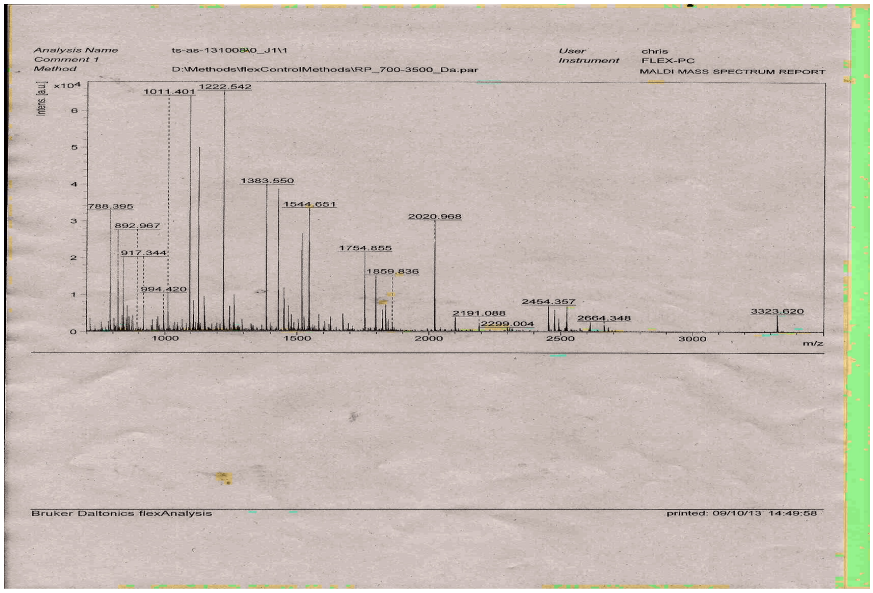
FPLC Chart for Ni column purification using AKTA machine



FPLC Chart for S200 column purification using AKTA machine

6.1 MALDI fragment analysis of SorbC

In order to check the exact mass and identity of the protein, the band from SDS-PAGE gel was cut into pieces in a 1.5 ml eppendorf tube and soaked overnight with 50% ethanol for destaining, then treated with 200 μL of 100 mM ammonium bicarbonate for 10 min at RT. The gel pieces were dehydrated with 200 μL acetonitrile several times, then reduced with 200 μL of 10 mM dithiothreitol for 1 h at 60 °C followed by alkylation step using 200 µL 50 mM Iodoacetamide for 45 min in the dark at room temperature. The pieces were washed several times with ammonium bicarbonate and acetonitrile then incubated overnight at 37 °C with 30 µL of Trypsin solution (Promega).[12] The tryptic peptide fragments were analysed using a Matrix Assisted Laser Desorption Ionization (MALDI) source coupled to a Time-of-Flight (TOF) mass spectrometer for identification of the parent protein. The instrument was equipped with a 200 Hz solid-state laser using a matrix of α-cyano-4-hydroxycinnaminic acid at 10 mg/ml in 50% acetonitrile and 0.05% trifluoroacetic acid. Mass data were collected in standard reflector mode over the mass range of 600 to 5000 *m/z* with a total of 400 shots collected at a laser intensity of 3300. MS/MS spectra were obtained using a standard 1-kV acquisition method with collision induced dissociation turned on. The laser intensity was set at 4300, and spectra were recorded for 2000 shots or until the accumulated spectra reached an estimated signal/ noise ratio of 75 (five peaks and eight subspectra minimum). Database searches were performed on the MASCOT search engine (www.matrixscience. com; Matrix Science, London, UK) using Applied Biosystems GPS software. From the results it was detected that the protein was Pc21g05060 (= SorbC) from *Penicillium chrysogenum* wisconsin 54-1255 with the exact expected mass (48.585 Kda) and 71% of protein sequence coverage.



MTRSANSPFE VAIVGGGITG LALAVGLLKR NVSFTIYERA ENFGELGVGI TFTPNAQRAM
EALDPCVLQS FTNVASAPSG GTINFVDGVR EQGSEDPRTS TAALLFQLHV KGGYKACRRC
DFVDQIVQHI PKDCVQYRKW LDSIETDHES GRAVLKFRDG EIAHADVVIG CDGIRSQVRA
SMFGTDELCP RAQYSHQLGY RGMVPLAQAT AVLGPEKTSS AVLHTGPGAF VLTIPLAEVH
AMHIEAFIMD KEEWPEVQTS SDSKRYVLPA TRNEATKAFA EFGPTVRSAV SMFPEKLEKW
AVFDMLEAPV PTFAKGRVCL AGDAAHASTP NQGGGAGFGI EDALVLAEVL AVLAEAPNVS
GIVASEALAV YSEVRYERSQ WLVRSSRRTG ELCTWKDRDW GLAAEELSRD IISRSHQLWD
HDTAGMVSDA LAILGERVRG ADTAF

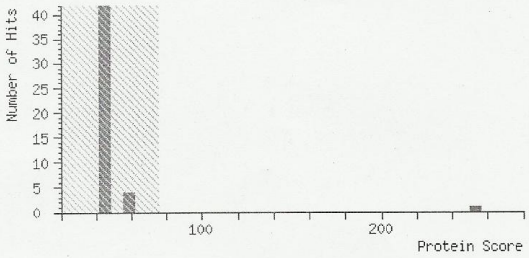
Red coloured residues detected by MALDI.

Mascot Search Results

User :
Email :
Search title :
Database : NCBInr 20130606 (25905244 sequences; 8953110406 residues)
Taxonomy : Fungi (1903004 sequences)
Timestamp : 9 Oct 2013 at 13:58:33 GMT
Top Score : 253 for gi|255953599, Pc21g05060 [Penicillium chrysogenum Wisconsin 54-1255]

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 75 are significant (p<0.05).



Concise Protein Summary Report

Format As: Concise Protein Summary Help
Significance threshold p< 0.05 Max. number of hits AUTO
Preferred taxonomy: All entries

Re-Search All Search Unmatched

1. gi|255953599 Mass: 48585 Score: 253 Expect: 9.5e-020 Matches: 38
Pc21g05060 [Penicillium chrysogenum Wisconsin 54-1255]

2. gi|238596406 Mass: 16468 Score: 54 Expect: 8.5 Matches: 8
hypothetical protein MPER_06133 [Moniliophthora perniciosa FA553]

Search Parameters

Type of search : Peptide Mass Fingerprint
Enzyme : Trypsin
Fixed modifications : Carbamidomethyl (C)
Variable modifications : Oxidation (M)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Peptide Charge State : 1+
Max Missed Cleavages : 1
Number of queries : 390
Selected for scoring : 39

Mascot: http://www.matrixscience.com/

MASCOT Search Results

Protein View: gi|255953599

Pc21g05060 [Penicillium chrysogenum Wisconsin 54-1255]

Database:	NCBInr
Score:	253
Expect:	9.5e-020
Nominal mass (M_r):	48585
Calculated pI:	5.40
Taxonomy:	Penicillium chrysogenum Wisconsin 54-1255

This protein sequence matches the following other entries:

- gi|211589263 from Penicillium chrysogenum Wisconsin 54-1255

Sequence similarity is available as an NCBI BLAST search of gi|255953599 against nr.

Search parameters

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)
Mass values searched:	390
Mass values matched:	38

Protein sequence coverage: 71%

Matched peptides shown in ***bold red***.

```
  1 MTRSANSPFE VAIVGGGITG LALAVGLLKR NVSFTIYERA ENFGELGVGI
 51 TFTPNAQRAM EALDPCVLQS FTNVASAPSG GTINFVDGVR EQGSEDPRTS
101 TAALLFQLHV KGGYKACRRC DFVDQIVQHI PKDCVQYRKW LDSIETDHES
151 GRAVLKFRDG EIAHADVVIG CDGIRSQVRA SMFGTDELCP RAQYSHQLGY
201 RGMVPLAQAT AVLGPEKTSS AVLHTGPGAF VLTIPLAEVH AMHIEAFIMD
251 KEEWPEVQTS SDSKRYVLPA TRNEATKAFA EFGPTVRSAV SMFPEKLEKW
301 AVFDMLEAPV PTFAKGRVCL AGDAAHASTP NQGGGAGFGI EDALVLAEVL
351 AVLAEAPNVS GIVASEALAV YSEVRYERSQ WLVRSSRRTG ELCTWKDRDW
401 GLAAEELSRD IISRSHQLWD HDTAGMVSDA LAILGERVRG ADTAF
```

Unformatted sequence string: 445 residues (for pasting into other applications).

Sort peptides by ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

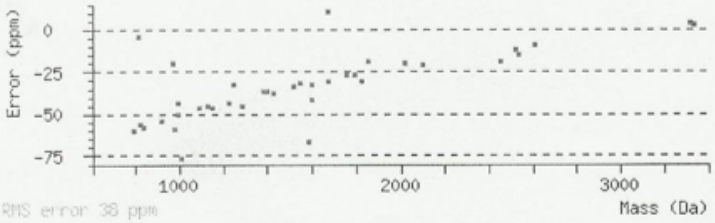
Show predicted peptides also

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
4 - 29	2454.3517	2453.3444	2453.3894	-18.3	0	R.SANSPFEVAIVGGGITGLALAVGLLK.R
4 - 30	2610.4754	2609.4681	2609.4905	-8.55	1	R.SANSPFEVAIVGGGITGLALAVGLLKR.N
30 - 39	1284.6113	1283.6040	1283.6622	-45.3	1	K.RNVSFTIYER.A
31 - 39	1128.5177	1127.5104	1127.5611	-45.0	0	R.NVSFTIYER.A
40 - 58	2020.9686	2019.9614	2020.0014	-19.8	0	R.AENFGELGVGITFTPNAQR.A
59 - 90	3323.6170	3322.6097	3322.5962	4.05	0	R.AMEALDPCVLQSFTNVASAPSGGTINFVDGVR.E
59 - 90	3339.6078	3338.6005	3338.5912	2.80	0	R.AMEALDPCVLQSFTNVASAPSGGTINFVDGVR.E + Oxidation (M)
91 - 98	917.3464	916.3392	916.3886	-54.0	0	R.EQGSEDPR.T
99 - 111	1428.7676	1427.7604	1427.8136	-37.3	0	R.TSTAALLFQLHVK.G
112 - 118	811.3844	810.3772	810.3806	-4.29	1	K.GGYKACR.R
119 - 132	1754.8538	1753.8465	1753.8934	-26.7	1	R.RCDFVDQIVQHIPK.D
120 - 132	1598.7476	1597.7403	1597.7923	-32.5	0	R.CDFVDQIVQHIPK.D

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
133 - 138	840.3180	839.3107	839.3596	-58.2	0	K.DCVQYR.K
133 - 139	968.4428	967.4356	967.4545	-19.6	1	K.DCVQYRK.W
139 - 152	1672.7417	1671.7345	1671.7852	-30.4	1	R.KWLDSIETDHESGR.A
140 - 152	1544.6496	1543.6423	1543.6903	-31.1	0	K.WLDSIETDHESGR.A
157 - 175	2099.9849	2098.9777	2099.0218	-21.0	1	K.FRDGEIAHADVVIGCDGIR.S
159 - 175	1796.8121	1795.8048	1795.8523	-26.4	0	R.DGEIAHADVVIGCDGIR.S
176 - 191	1853.8291	1852.8218	1852.8560	-18.5	1	R.SQVRASMFGTDELCPR.A
180 - 191	1383.5527	1382.5455	1382.5959	-36.5	0	R.ASMFGTDELCPR.A
180 - 191	1399.5469	1398.5397	1398.5908	-36.6	0	R.ASMFGTDELCPR.A + Oxidation (M)
192 - 201	1222.5435	1221.5363	1221.5891	-43.2	0	R.AQYSHQLGYR.G
202 - 217	1581.7615	1580.7542	1580.8596	-66.7	0	R.GMVPLAQATAVLGPEK.T
202 - 217	1597.7956	1596.7883	1596.8545	-41.4	0	R.GMVPLAQATAVLGPEK.T + Oxidation (M)
252 - 265	1677.7895	1676.7822	1676.7642	10.8	1	K.EEWPEVQTSSDSKR.Y
265 - 272	975.5158	974.5086	974.5661	-59.1	1	K.RYVLPATR.N
266 - 272	819.4264	818.4191	818.4650	-56.1	0	R.YVLPATR.N
278 - 287	1094.5124	1093.5051	1093.5556	-46.2	0	K.AFAEFGPTVR.S
288 - 296	995.4367	994.4294	994.4794	-50.3	0	R.SAVSMFPEK.L
288 - 296	1011.4039	1010.3966	1010.4743	-76.9	0	R.SAVSMFPEK.L + Oxidation (M)
300 - 315	1821.8692	1820.8620	1820.9171	-30.3	0	K.WAVFDMLEAPVPTFAK.G
379 - 384	788.3940	787.3868	787.4341	-60.1	0	R.SQWLVR.S
388 - 396	1150.5141	1149.5068	1149.5601	-46.4	1	R.RTGELCTWK.D
389 - 396	994.4233	993.4160	993.4590	-43.2	0	R.TGELCTWK.D
397 - 409	1517.6839	1516.6766	1516.7270	-33.2	1	K.DRDWGLAAEELSR.D
399 - 409	1246.5655	1245.5582	1245.5989	-32.7	0	R.DWGLAAEELSR.D
415 - 437	2522.1800	2521.1728	2521.2020	-11.6	0	R.SHQLWDHDTAGMVSDALAILGER.V
415 - 437	2538.1675	2537.1602	2537.1969	-14.5	0	R.SHQLWDHDTAGMVSDALAILGER.V + Oxidation (M)

No match to: 703.8736, 712.3338, 731.3617, 741.2896, 745.3640, 749.3706, 752.3550, 755.3073, 759.3340, 759.8274, 771.3050, 774.3916, 776.2447, 782.3183, 783.3574, 793.2833, 794.3722, 802.4366, 804.2329, 806.2334, 807.3559, 810.3830, 815.4322, 825.0524, 826.3421, 833.0250, 835.3110, 839.0391, 841.0292, 842.4598, 845.0468, 851.3874, 855.0074, 861.0238, 863.3568, 870.4941, 870.9775, 873.3329, 876.3987, 876.9968, 886.9514, 892.9707, 893.3660, 897.3678, 900.2954, 903.3953, 906.4140, 912.4005, 930.3338, 933.4326, 939.3276, 947.4215, 951.4345, 952.4420, 955.2903, 962.3748, 973.4499, 979.4694, 982.3925, 988.4744, 990.4007, 997.3854, 999.4953, 1006.3804, 1014.4672, 1016.4082, 1017.4104, 1022.4582, 1028.0731, 1032.3697, 1033.3838, 1036.4706, 1038.4344, 1044.0439, 1046.5057, 1047.4289, 1049.3863, 1050.0437, 1050.4612, 1050.9642, 1051.4684, 1060.0161, 1060.4897, 1065.4549, 1066.0218, 1066.4856, 1071.5477, 1079.4983, 1080.4072, 1081.9937, 1097.9666, 1103.5196, 1105.5253, 1107.5214, 1109.4796, 1111.4474, 1113.4863, 1116.4956, 1119.4425, 1124.5060, 1132.4634, 1134.4716, 1136.5436, 1138.4717, 1153.4876, 1164.5514, 1165.5343, 1166.4854, 1169.4916, 1173.4894, 1176.5106, 1178.5662, 1179.5578, 1183.5170, 1188.5203, 1192.5375, 1193.5670, 1194.4182, 1197.5083, 1198.4979, 1199.4868, 1201.5852, 1204.5275, 1209.5880, 1211.4766, 1226.5830, 1227.5422, 1231.5206, 1232.5504, 1233.5472, 1235.4746, 1239.5448, 1244.5484, 1245.5873, 1250.5873, 1254.5340, 1256.5928, 1260.4882, 1262.0576, 1263.6403, 1266.5600, 1273.5992, 1274.4230, 1276.4504, 1277.0451, 1277.6272, 1279.5663, 1282.5325, 1285.6173, 1293.0251, 1293.5933, 1300.4740, 1301.5919, 1307.6272, 1308.5870, 1315.6159, 1318.5675, 1322.5567, 1327.5659, 1331.5122, 1335.5452, 1339.5702, 1341.4682, 1348.5338, 1350.5857, 1358.6095, 1361.5798, 1365.5588, 1367.6108, 1373.6234, 1375.5226, 1392.5469, 1405.5344, 1406.6553, 1411.6072, 1419.5823, 1421.4989, 1422.5997, 1434.6609, 1440.5925, 1449.6887, 1450.7507, 1456.7023, 1459.6282, 1466.7259, 1472.5651, 1475.7048, 1478.6462, 1480.5074, 1481.6764, 1485.7934, 1487.7025, 1490.6821, 1494.6752, 1497.7372, 1502.6943, 1503.6111, 1504.7210, 1509.8032, 1526.5628, 1529.6991, 1533.7459, 1534.6861, 1542.6961, 1549.6530, 1555.6346, 1559.6781, 1563.6852, 1566.6265, 1571.5990, 1573.6974, 1577.6501, 1582.8107, 1583.5947, 1584.7269, 1587.6855, 1593.5684, 1601.6680, 1603.7963, 1619.7602, 1625.7990, 1636.7587, 1641.7020, 1646.8019, 1657.7338, 1687.7399, 1693.7486, 1697.8165, 1700.7318, 1703.7181, 1706.8249, 1709.7425, 1716.7720, 1729.7566, 1731.7087, 1739.8094, 1749.8881, 1765.6812, 1768.7990, 1776.8307, 1779.7819, 1788.8193, 1791.6809, 1808.7596, 1809.8785, 1811.8638, 1814.7726, 1818.7856, 1834.9328, 1843.8565, 1852.8373, 1856.8998, 1859.8358, 1861.9017, 1864.8510, 1867.7882, 1872.8817, 1874.9607, 1881.8309, 1891.9607, 1896.8029, 1909.8522, 1912.9514, 1915.8529, 1936.9357, 1946.8552, 1976.9384, 1993.9424, 2008.8650, 2013.9148, 2030.8391, 2035.9283, 2042.9469, 2046.8205, 2053.7568, 2058.9229, 2062.9811, 2077.0001, 2085.0094, 2091.9752, 2104.0328, 2113.0049, 2121.9686, 2125.9857, 2137.9231, 2143.9543, 2153.9087, 2157.9686, 2164.9684, 2191.0861, 2232.1155, 2285.0954, 2290.1152, 2299.0049, 2304.0988, 2307.1446, 2316.0296, 2317.1703, 2319.1567, 2334.1228, 2340.0180, 2345.9431, 2358.0342, 2382.1676, 2383.9130, 2401.2228, 2413.0873, 2464.0035, 2476.3446, 2485.2186,

2492.3176, 2498.3209, 2501.3020, 2507.2024, 2510.1016, 2515.2855, 2519.3036, 2530.2065, 2541.2754, 2544.1659, 2550.1506, 2551.1992, 2557.2331, 2560.1320, 2565.1406, 2572.0698, 2577.2121, 2580.1808, 2583.1152, 2606.1736, 2628.1184, 2637.3306, 2644.1345, 2651.0871, 2664.3422, 2667.4847, 2679.3199, 2693.2991, 2701.2714, 2705.1170, 2762.2619, 2808.2745, 2817.9277, 3038.4113, 3105.2543, 3116.6016, 3176.3626, 3275.6222, 3312.3313, 3345.5933, 3361.5762, 3380.6230




```
LOCUS       XP_002567552              445 aa            linear   PLN 14-AUG-2009
DEFINITION  Pc21g05060 [Penicillium chrysogenum Wisconsin 54-1255].
ACCESSION   XP_002567552
VERSION     XP_002567552.1  GI:255953599
DBSOURCE    REFSEQ: accession XM_002567506.1
KEYWORDS    RefSeq.
SOURCE      Penicillium chrysogenum Wisconsin 54-1255
  ORGANISM  Penicillium chrysogenum Wisconsin 54-1255
            Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
            Eurotiomycetes; Eurotiomycetidae; Eurotiales; Aspergillaceae;
            Penicillium; Penicillium chrysogenum complex.
REFERENCE   1  (residues 1 to 445)
  AUTHORS   van den Berg,M.A., Albang,R., Albermann,K., Badger,J.H.,
            Daran,J.M., Driessen,A.J., Garcia-Estrada,C., Fedorova,N.D.,
            Harris,D.M., Heijne,W.H., Joardar,V., Kiel,J.A., Kovalchuk,A.,
            Martin,J.F., Nierman,W.C., Nijland,J.G., Pronk,J.T., Roubos,J.A.,
            van der Klei,I.J., van Peij,N.N., Veenhuis,M., von Dohren,H.,
            Wagner,C., Wortman,J. and Bovenberg,R.A.
  TITLE     Genome sequencing and analysis of the filamentous fungus
            Penicillium chrysogenum
  JOURNAL   Nat. Biotechnol. 26 (10), 1161-1168 (2008)
   PUBMED   18820685
REFERENCE   2  (residues 1 to 445)
  AUTHORS   van den Berg,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2007) van den Berg M.A., DAI/INNO (624-0270), DSM
            Anti-Infectives, Alexander Fleminglaan 1, Delftn, 2613 AX,
            NETHERLANDS
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence is identical to CAP95403.
FEATURES             Location/Qualifiers
     source          1..445
                     /organism="Penicillium chrysogenum Wisconsin 54-1255"
                     /strain="Wisconsin 54-1255"
                     /db_xref="taxon:500485"
                     /chromosome="1"
                     /clone="Pc00c21"
     Protein         1..445
                     /product="hypothetical protein"
                     /name="Pc21g05060"
                     /calculated_mol_wt=48029
     Region          11..434
                     /region_name="NADB_Rossmann"
                     /note="Rossmann-fold NAD(P)(+)-binding proteins; cl09931"
                     /db_xref="CDD:213117"
     Region          11..391
                     /region_name="PRK06847"
                     /note="hypothetical protein; Provisional"
                     /db_xref="CDD:180727"
     CDS             1..445
                     /locus_tag="Pc21g05060"
                     /coded_by="XM_002567506.1:1..1338"
                     /inference="protein motif:COGS:COG0644"
                     /inference="protein motif:COGS:COG0654"
                     /inference="protein motif:PFAM:PF01360"
```

7. Purification and Characterisation of Sorbicillin and Dihydrosorbicillin

7.1 Fermentation. *Penicillium chrysogenum* E01-10/3 was grown on malt extract agar plates for 12 days at 28 °C. Small pieces of agar around (1 cm^2) were picked from mature plates and transferred into sponge-plugged 500 ml flaskes containing 100 ml of sterilised Czapek-Dox broth supplied with 20 g L^{-1} tryptone in deionised water. A total of 3 L inoculated liquid medium was incubated without shaking for a week in a dark incubator at 28 °C. The mycelium was separated from the fermentation by vacuum filtration. The supernatants were combined together, acidified to pH 3.0 (aq. HCl), then extracted with ethyl acetate. The organic extract was dried with anhydrous $MgSO_4$ and evaporated to afford crude extract (2.2 g).

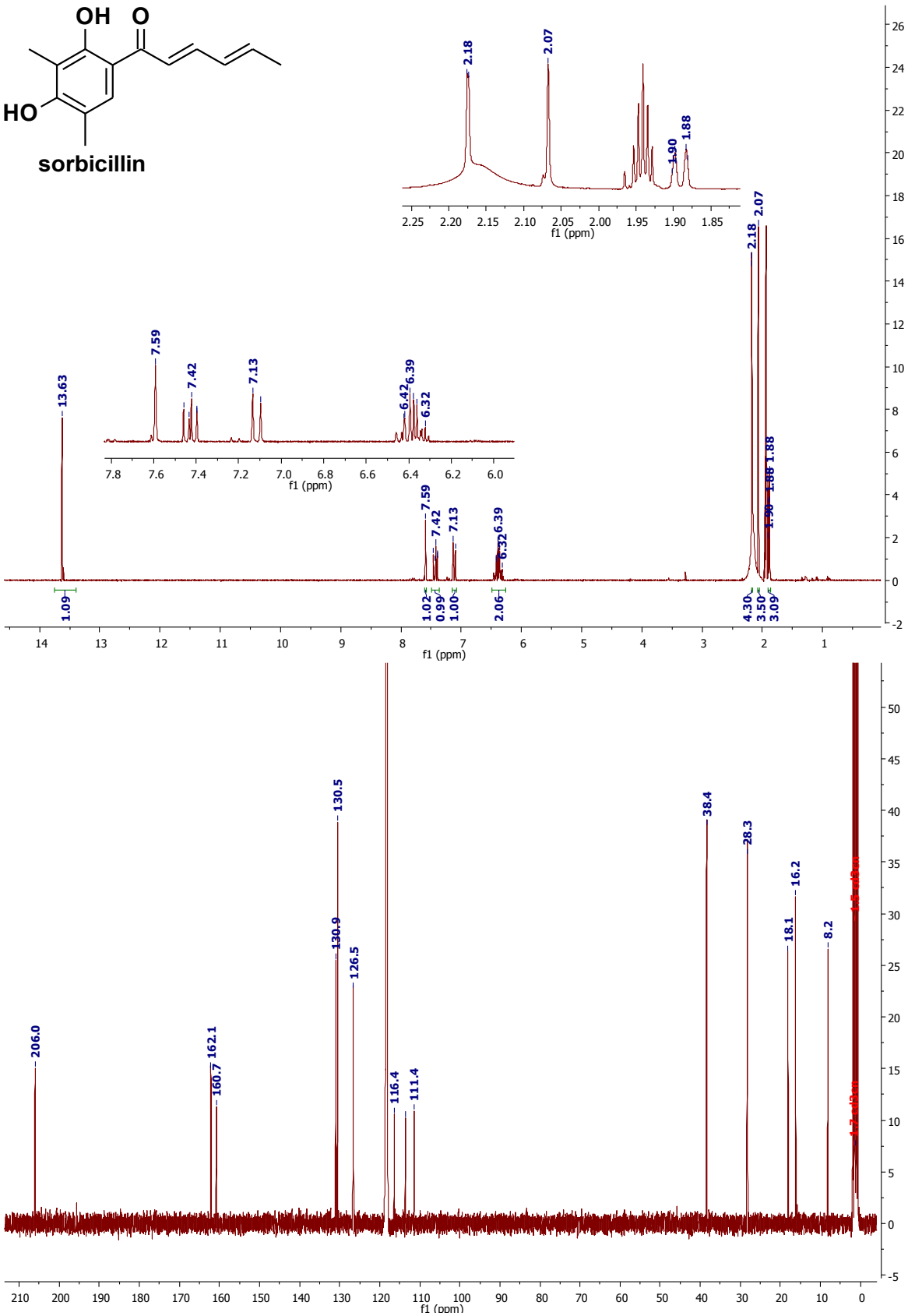
7.2 The isolation of sorbicillin and 2',3'-dihydrosorbicillin. The whole of the crude extract (2.2 g) was subjected first to a flash silica gel column (pore size 60 Å, 220-440 mesh particle size, 35-75 μm particle size) using dichloromethane (DCM) as a mobile phase. This eluted fractions containing the two metabolites as major compounds. The corresponding fractions were combined and the organic solvent was evaporated under reduced pressure to give yellow oily residue (75 mg) which was dissolved in HPLC grade methanol for further purification using a reverse-phase mass-directed HPLC.

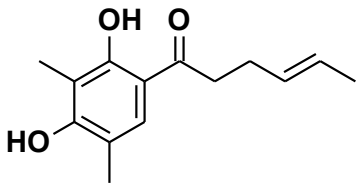
7.3 Mass-directed purification. The purification of the two metabolites was achieved using a Waters mass-directed autopurification system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex LUNA column (5μ, C_{18}, 100 Å, 10 × 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna C_5 300 Å) eluted at 4 mL/min. Solvent A, HPLC grade H_2O + 0.05% formic acid; Solvent B, HPLC grade CH_3CN + 0.045% formic acid. The post-column flow was split (100:1) and the minority flow was made up with MeOH + 0.045% formic acid to 1 mL · min^{-1} for simultaneous analysis by diode array (Waters 2998), evaporative light scattering (Waters 2424) and ESI mass spectrometry in positive and negative modes (Waters Quatro Micro). The Solvent gradient was as follows: 0 min, 20% B; 23min, 90% B; 24 min, 95% B; 27 min, 95% B; 28 min, 20% B; 30 min, 20% B. Detected peaks were collected into glass test tubes. Combined tubes were evaporated under a flow of dry N_2 gas, weighed, and residues dissolved directly in NMR solvent for NMR analysis.

7.4 The characterisation of sorbicillin and 2',3'-dihydrosorbicillin. The physical data of the two metabolites were in good agreement with literature data.[13]

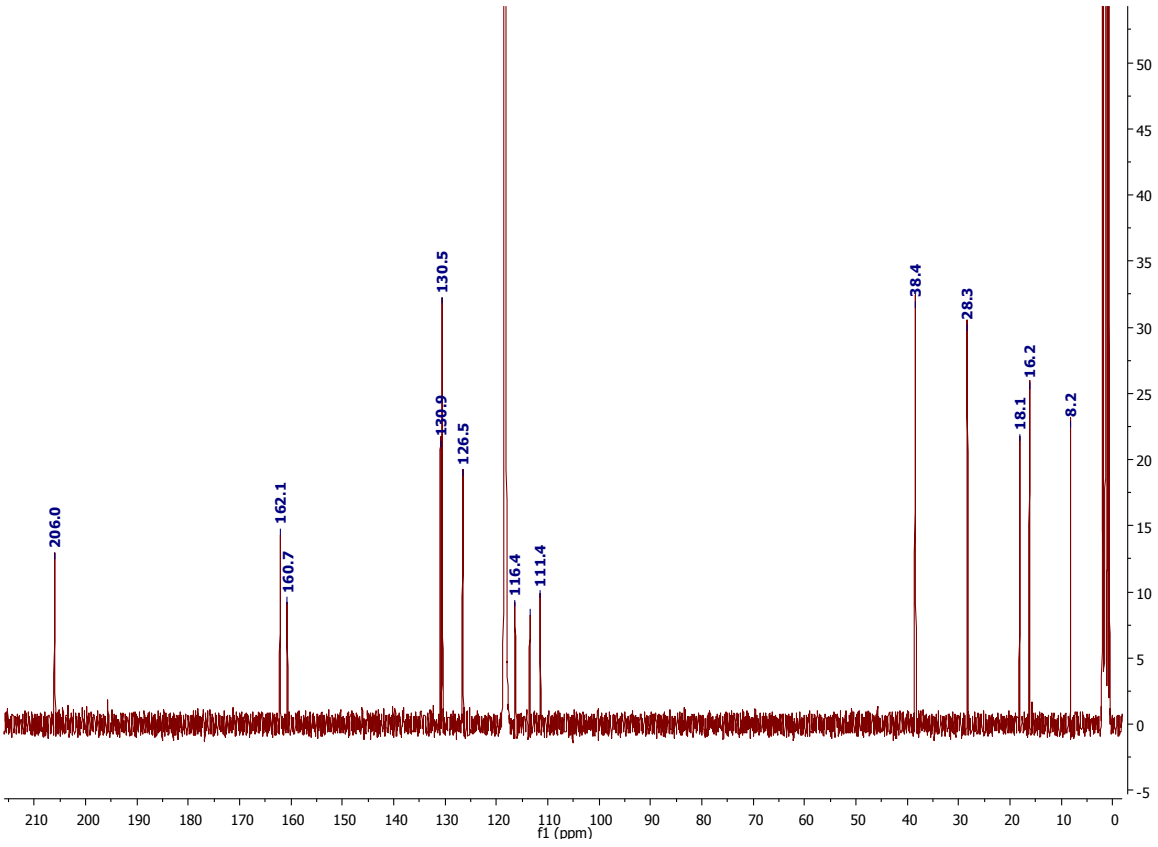
Sorbicillin: (2 mg·L^{-1}), yellow crystals, λ_{max} (CH_3CN) 323 nm; ESI *m/z* 231 $[M-H]^-$, 233 $[M]H^+$; ^1H NMR (400 MHz, CD_3CN): δ_H = 1.89 (d, *J* = 6.7 Hz, 3H), 2.07 (s, 3H), 2.18 (s, 3H), 6.83 (m, 2H), 7.12 (d, *J* = 14.8 Hz, 1H), 7.43 (dd, *J* = 14.8, 15.0 Hz, 1H), 7.59 (s, 1H), 13.63 (s, 1H); ^{13}C NMR (100 MHz, CD_3CN): δ_c = 8.2 (CH_3), 16.3 (CH_3), 19.0 (CH_3), 111.6 (C), 114.0 (C), 116.4 (C), 123.3 (CH), 130.1 (CH), 131.4 (CH), 142.3 (CH), 145.1 (CH), 161.0 (C), 163.4 (C), 193.6 (C).

2',3'-dihydrosorbicillin: (5.4 mg·L^{-1}), Pale yellow crystals, λ_{max} (CH_3CN) 220, 286 and 328 nm; ESI *m/z* 233 $[M-H]^-$, 235 $[M]H^+$; ^1H NMR (400 MHz, CD_3CN): δ_H = 1.62 (m, 3H), 2.05 (s, 3H), 2.16 (s, 3H), 2.34 (m, 2H), 2.98 (dd, *J* = 7.3, 7.6 Hz, 2H), 5.51 (m, 2H), 7.52 (s, 1H), 12.99 (s, 1H); ^{13}C NMR (100 MHz, CD_3CN): δ_c = 8.2 (CH_3), 16.2 (CH_3), 18.1 (CH_3), 28.3 (CH_2), 38.4 (CH_2), 111.4 (C), 113.5 (C), 116.4 (C), 126.5 (CH), 130.5 (CH), 130.9 (CH), 160.7 (C), 162.1 (C), 206.0 (C).





2',3'-dihydrosorbicillin

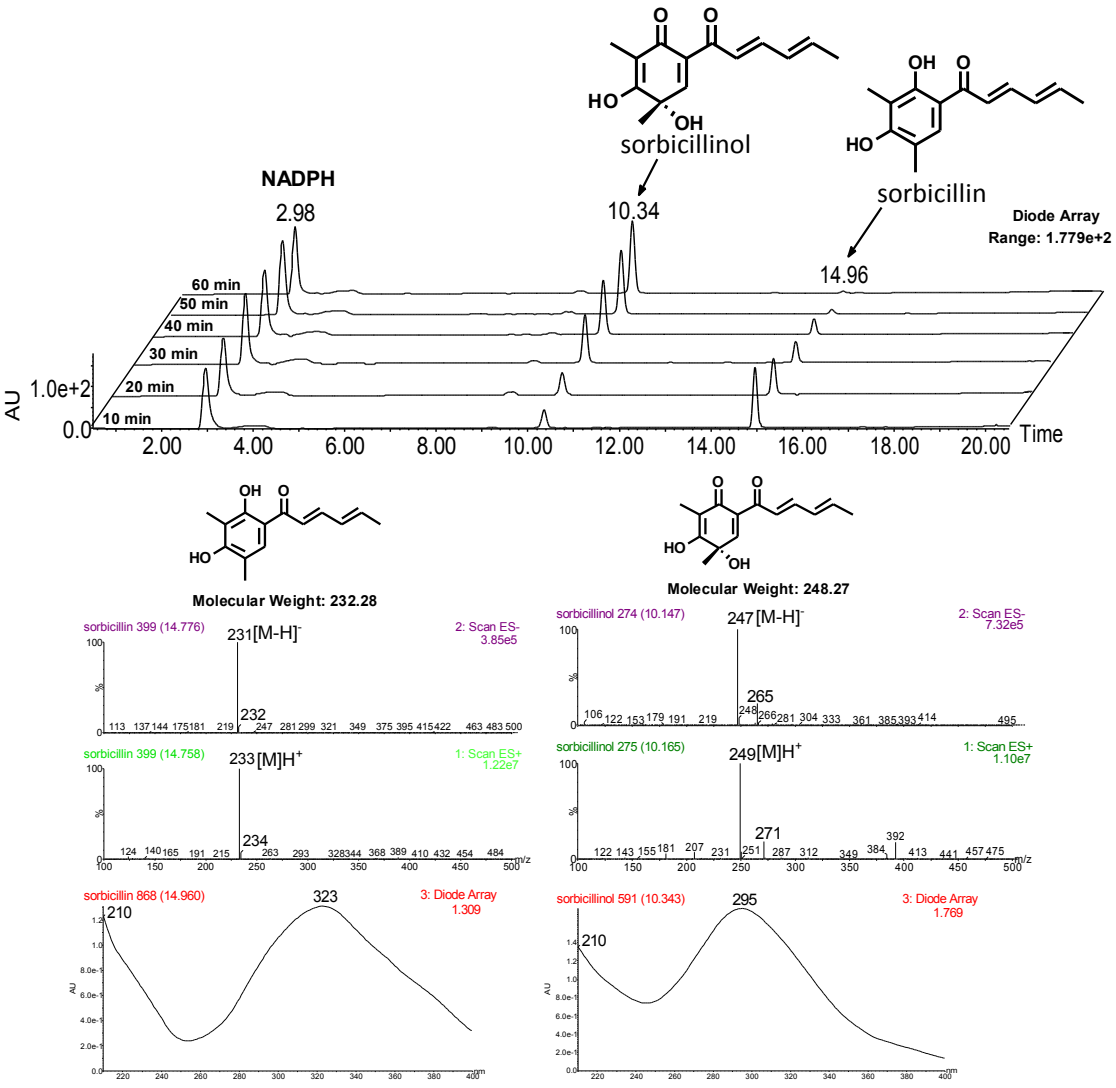


8. *In Vitro* Assays of SorbC

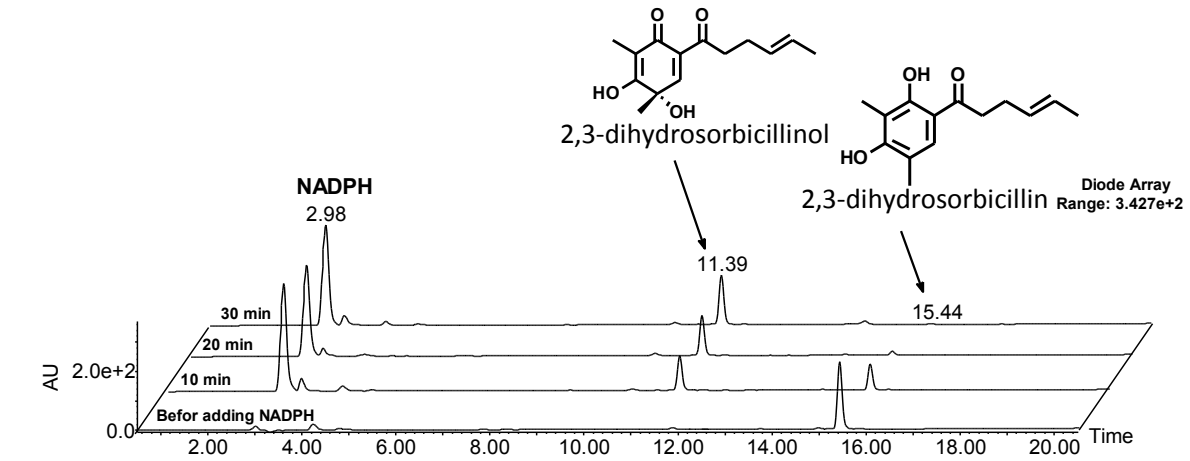
In 945 μl phosphate buffer 50 mM (pH 8), 50 μl of substrate acetone stock solution (50 mM) was added to give a final concentration of 2.5 mM. This was followed by the addition 5 μl of SorbC (final concentration 15 μM). The reaction was initiated by adding 2 mg of NADH or 2.5 mg of NADPH (final concentration 3 mM).

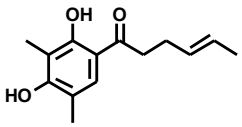
8.1 Time-course analysis of SorbC assays. For a time course, the reaction mixture was incubated for 60 min at room temperature (20°C). Every 10 min, 100 μl from the incubation mixture was transferred into a 0.5 ml small eppendorf tube containing 100 μl of 30% HCl to stop the reaction. Microcentrifugation was followed at (14,000*g*, 5 min) to pellet the precipitated enzyme. The supernatant was directly subjected to LC-DAD-MS analysis by injecting 50 µl and using a Waters 2795HT HPLC system. Detection was achieved by UV between 200 and 400 nm using a Waters 2998 diode array detector, and by simultaneous electrospray (ES) mass spectrometry using a Waters Quatro Micro spectrometer detecting between 150 and 600 *m/z* units. Chromatography (flow rate 1 mL·min^{-1}) was achieved using a Phenomenex LUNA column (5 µ, C_{18}, 100 Å, 4.6 × 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna C_5 300 Å). Solvents were: A, HPLC grade H_2O containing 0.05% formic acid; B, HPLC grade CH_3CN containing 0.045% formic acid). Solvent gradient was as follows.
Method 1: 0 min, 5% B; 15 min, 95% B; 18 min, 95% B; 18.5 min, 5% B; 20 min, 5% B.

8.2 Sorbicillin assay

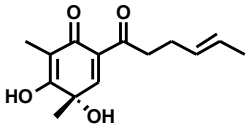


8.3 2,3-dihydrosorbicillin assay

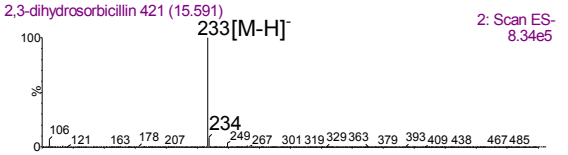


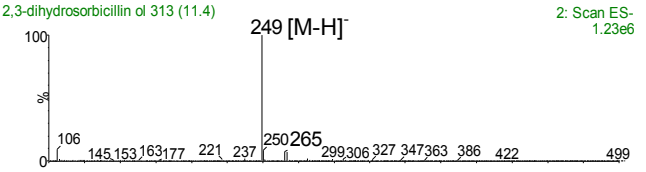


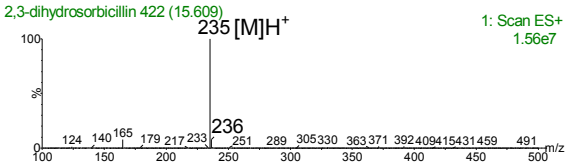
Molecular Weight: 234.29

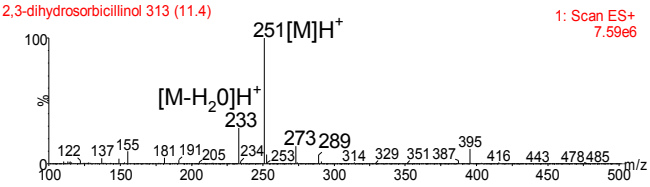


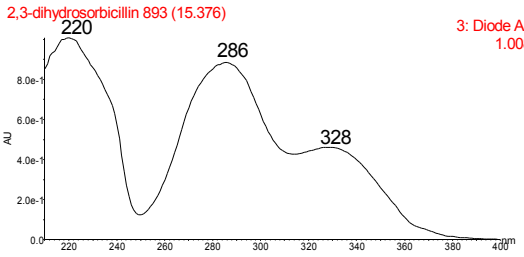
Molecular Weight: 250.29

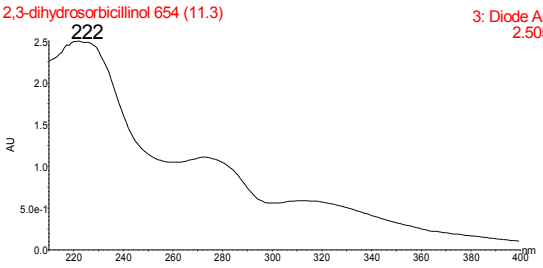












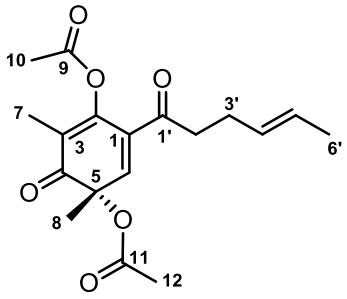
9. Purification and Characterisation of diacetyl-2'3'-dihydrosorbicillinol.

9.1 Sorbc large scale assay. For large scale *in vitro* assay, 17 reaction mixtures were incubated at 28 °C for 2 hours. When the substrate was completely oxidised, they were combined together and 2 volumes of HPLC grade acetonitrile were added. This step precipitated the enzyme which pelleted upon centrifugation at high speed. Then, the clear supernatant was transferred to a 100 ml small flask and the acetonitrile was evaporated using nitrogen blow down.

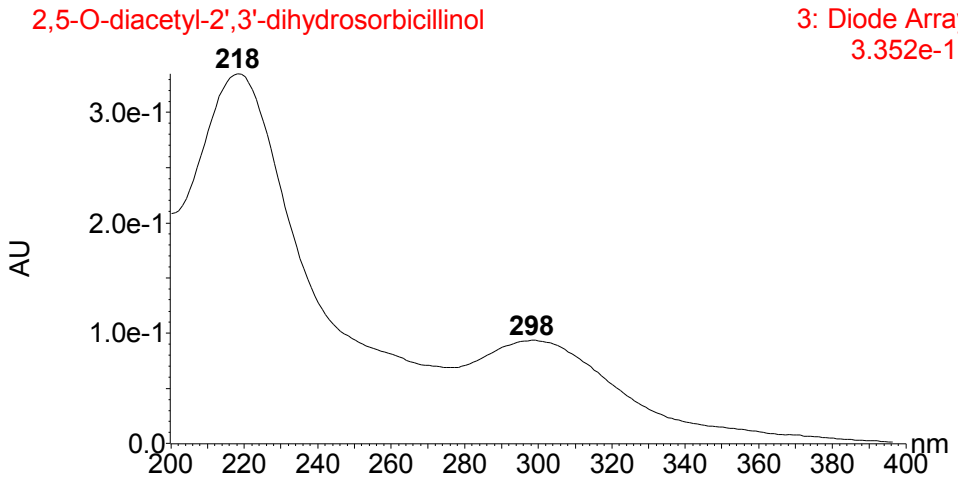
9.2 The acetylation of 2',3'-dihydrosorbicillinol. The incubation solution of 2', 3'-dihydrosorbicillinol (0.59 mg·ml^{-1}, 17 ml) was placed on ice bath and stirred for 30 min. Next, 8.5 ml of pyridine was added and left for 15 min. This was followed by a gradual addition of 5 ml acetic anhydride over 10 min at 0 °C. The reaction was then stirred at room temperature for 1 hour before purification.

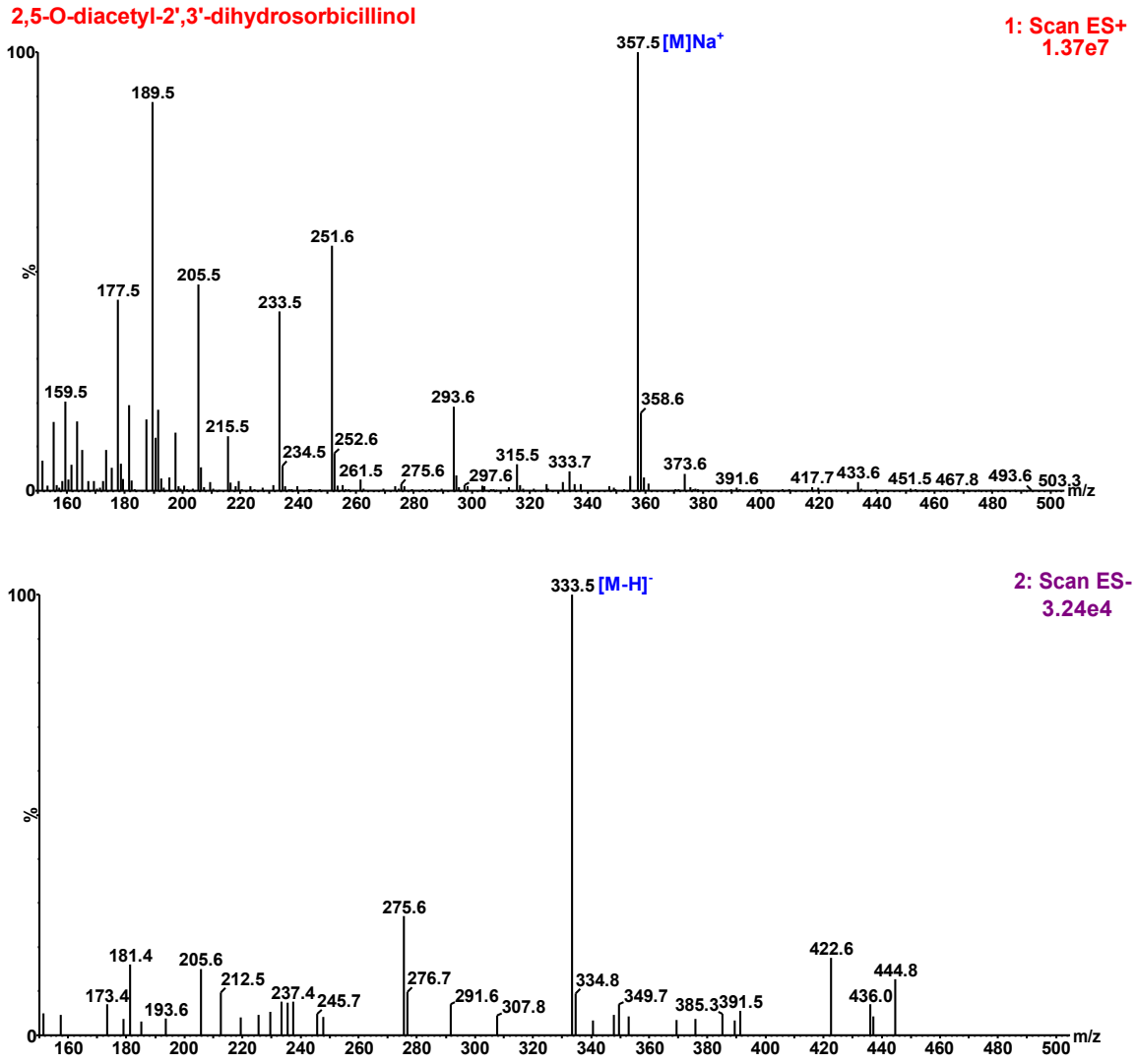
9.3 The purification of 2,5-O-diacetyl-2',3'-dihydrosorbicillinol. The purification of 2,5-O-diacetyl-2',3'-dihydrosorbicillinol was carried out using a larger column reversed-phase HPLC (Phenomenex, Kinetex, 5 μm C_{18} 100 Å, AXIA Packed LC Column 250 x 21.2 mm) at a flow rate of 16 ml/min. The purification was performed following the solvent gradients of *method1*. The acetylation reaction yielded 1 mg (10%) of the compound which submitted for NMR analysis.

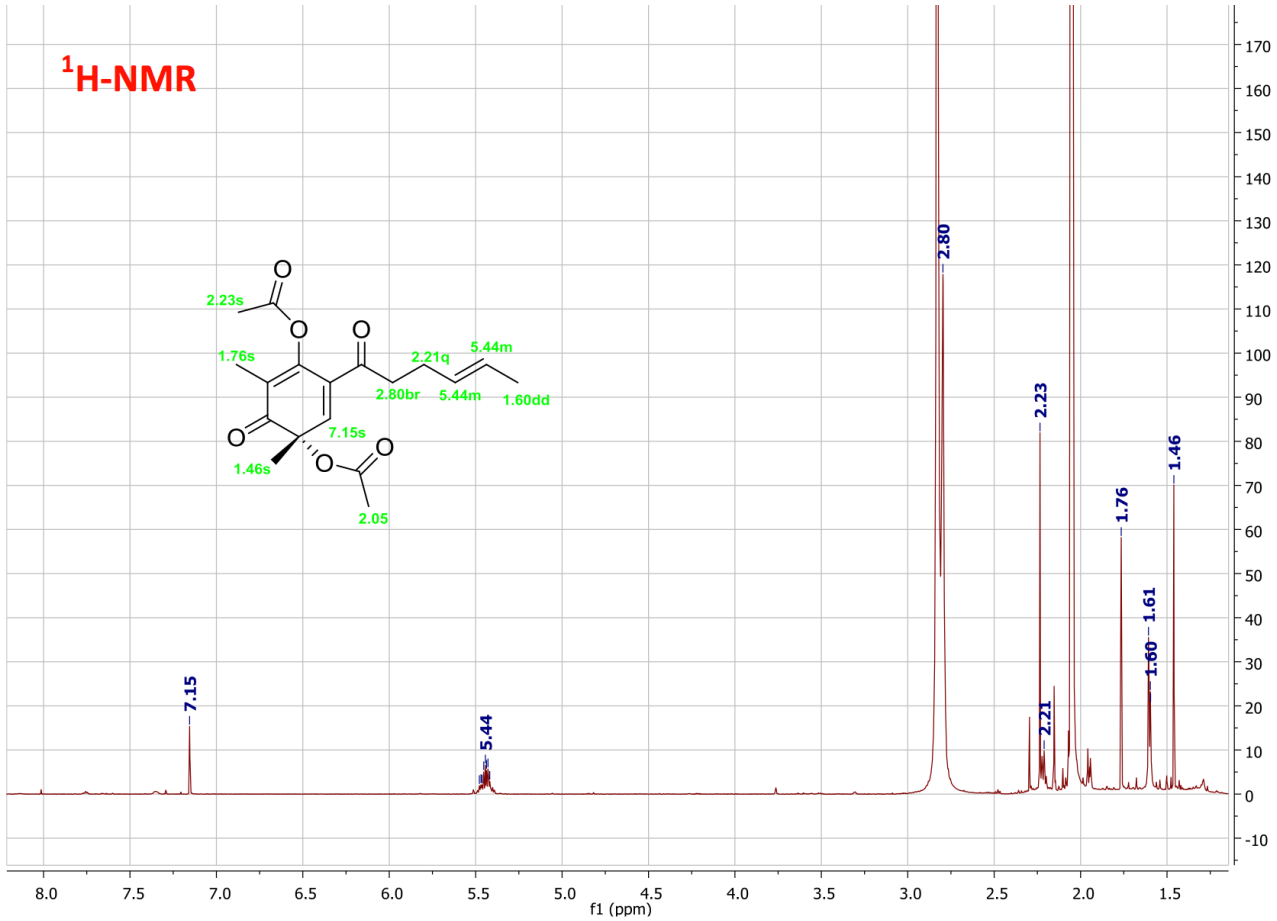
2,5-O-diacetyl-2',3'-dihydrosorbicillinol

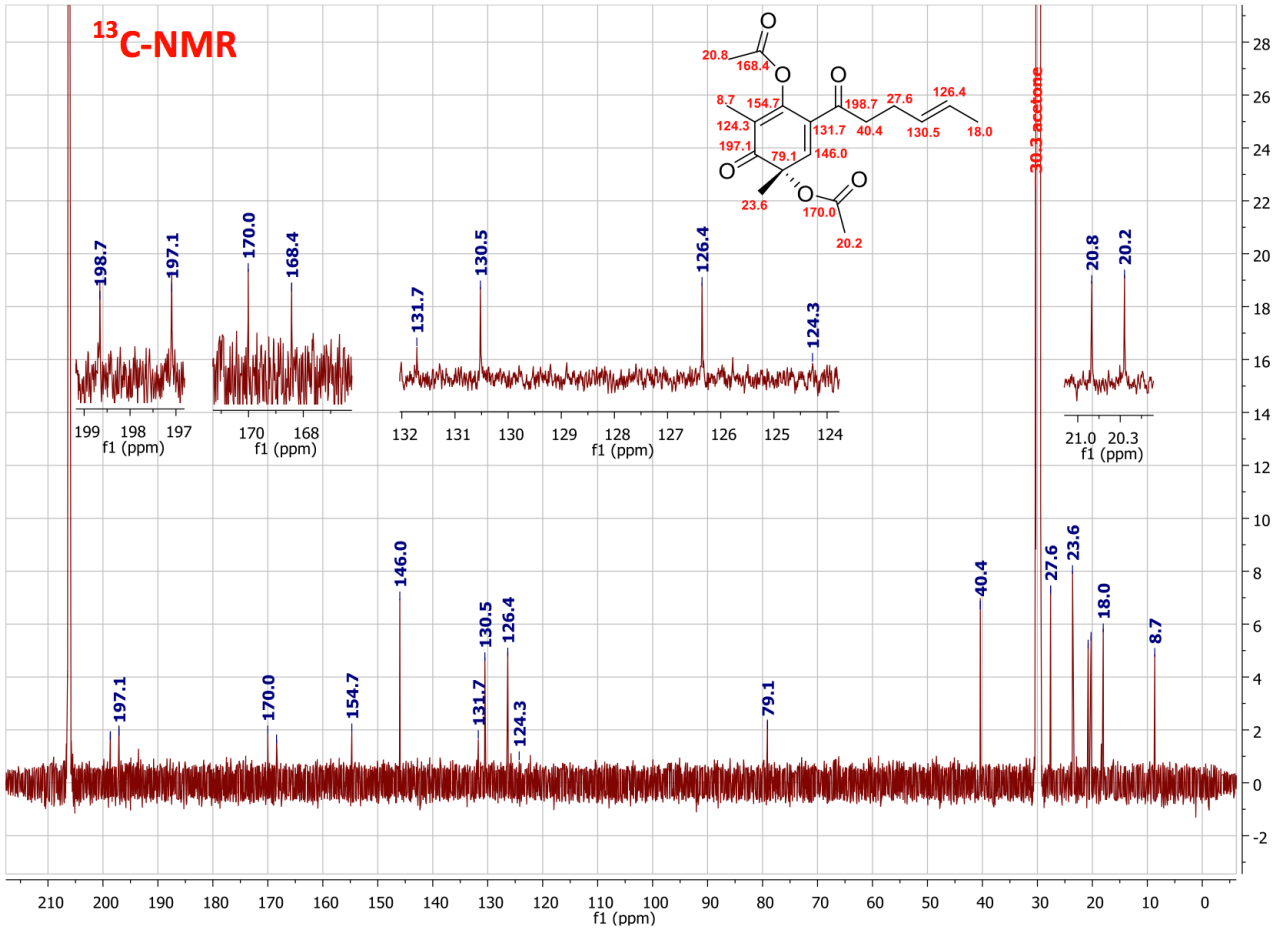


Pale yellow powder, HRMS *m/z* 357.1299 [M]Na$^+$, (calcd for $C_{18}H_{22}NaO_6$, 357.1309); ^{1}H-NMR (600 MHz, acetone-d_6) δ_H: 1.46 (s, 3H, H_3-8), 1.60 (dd, J = 6.4, 1.5, 3H, H_3-6'), 1.76 (s, 3H, H_3-7), 2.05 (s, 3H, H_3-12), 2.21 (m, 2H, H_2-3'), 2.23 (s, 3H, H_3-10), 2.80 (brm, 2H, H_2-2'), 5.44 (m, 2H, H-4'and H-5'), 7.15 (s, 1H, H-6); ^{13}C-NMR (150 MHz, acetone-d6) δ_c: 8.7 (C-7), 18.0 (C-6'), 20.2 (C-12), 20.8 (C-10), 23.6 (C-8), 27.6 (C-3'), 40.4 (C-2'), 79.1 (C-5), 124.3 (C-3), 126.4 (C-5'), 130.5 (C-4'), 131.7 (C-1), 146.0 (C-6), 154.7 (C-2), 168.4 (C-9), 170.0 (C-11), 197.1(C-4), 198.7 (C-1'); λ_{max} (CH_3CN:water, 3:1) 218 and 298 nm; ESI *m/z* 357.5 [M]Na$^+$, 333.5 [M-H]$^-$.

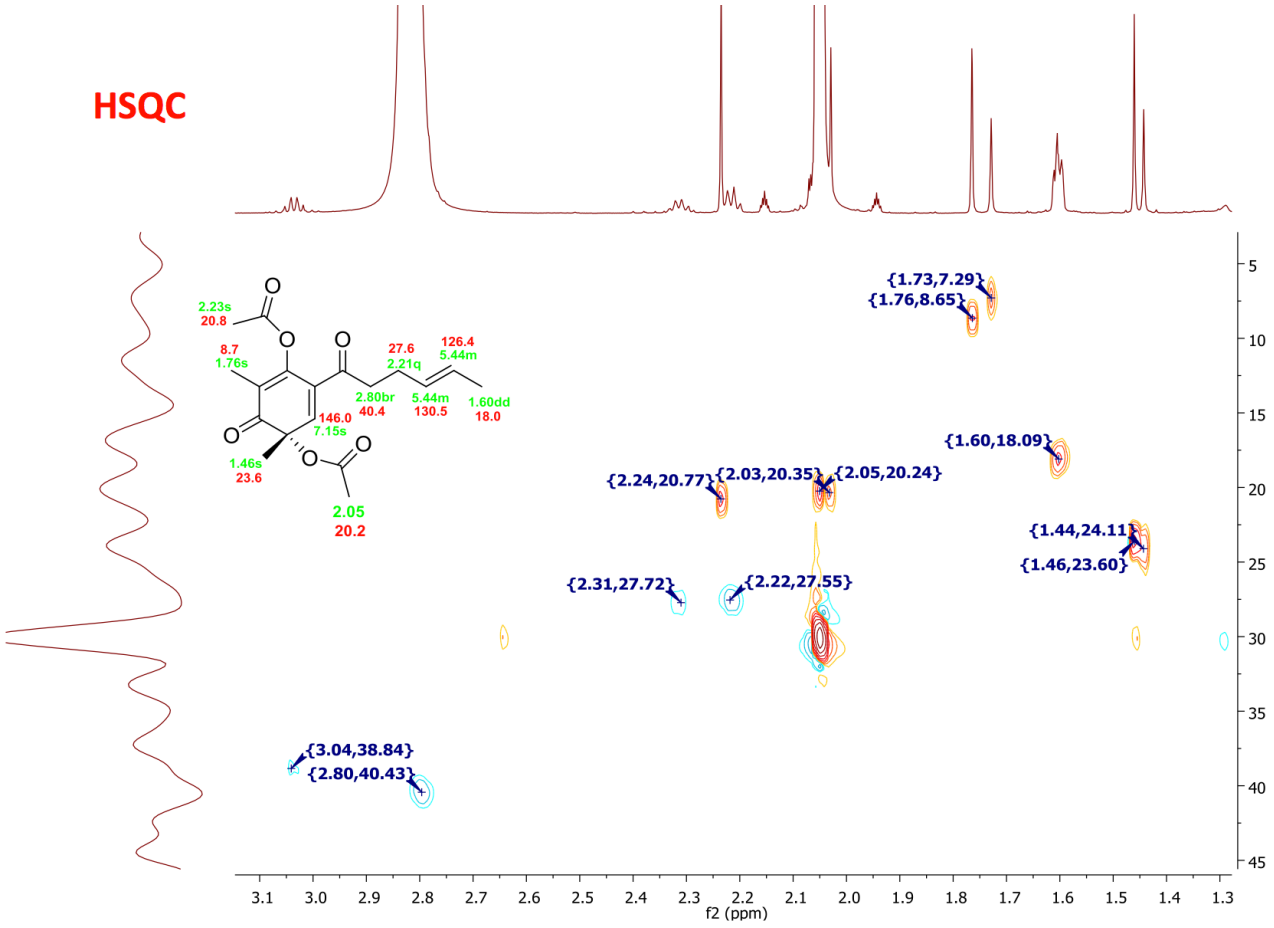




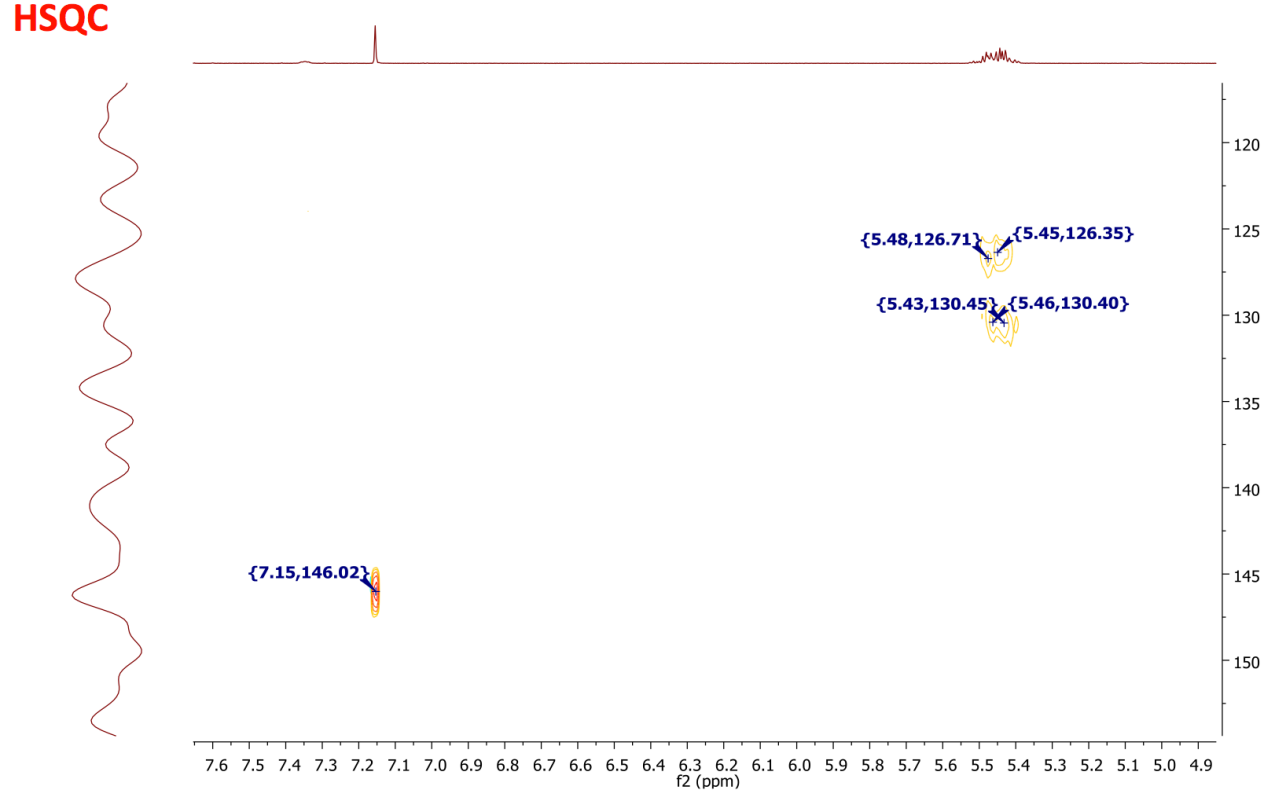


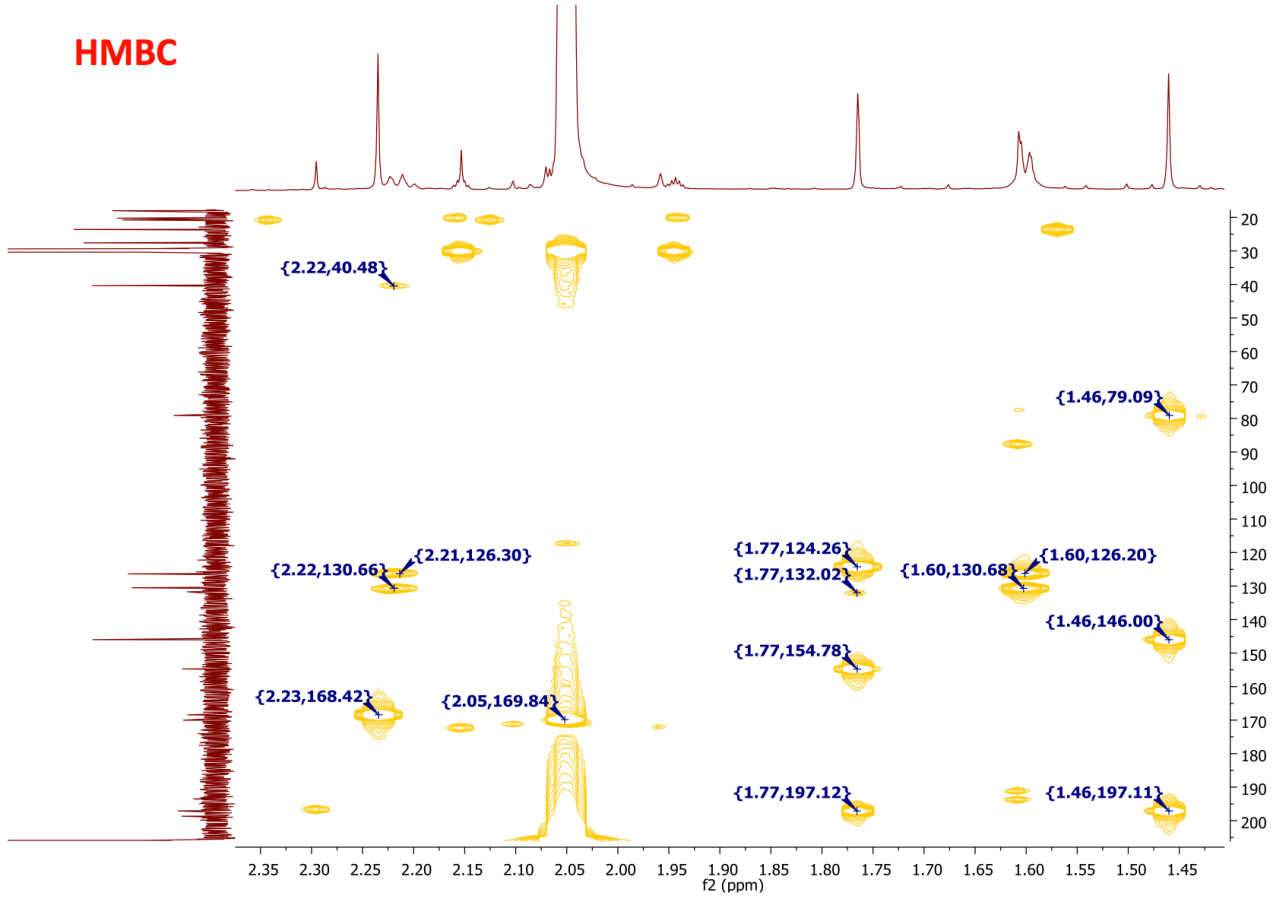


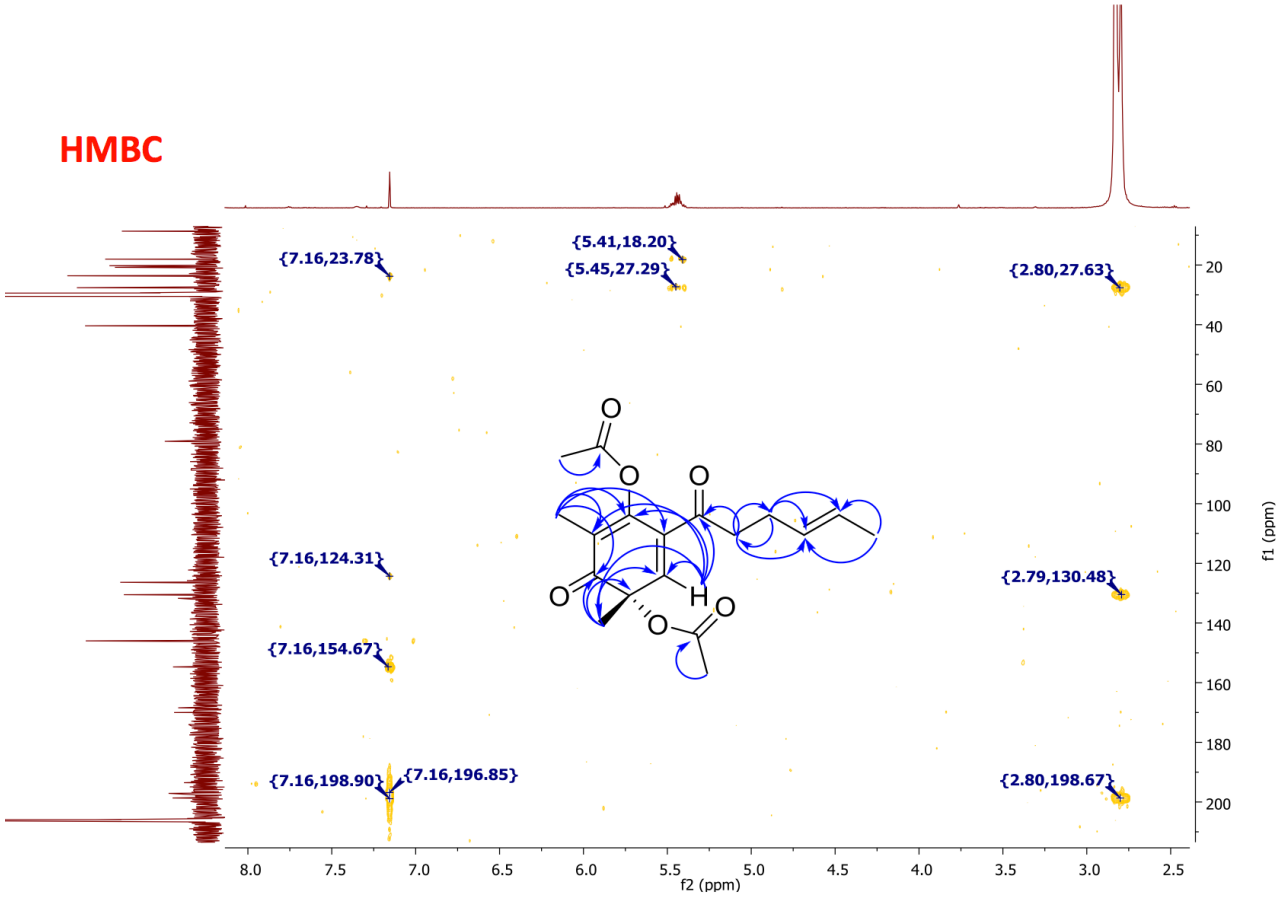
HSQC



HSQC







10. References

1. A. M. Bailey, R. J. Cox, K. Harley, C. M. Lazarus, T. J. Simpson and E. Skellam, *Chem. Commun.*, 2007, 4053 - 4055.
2. A. Salamov and V. Solovyev, *Genome Res.*, 2000, **10**, 516 - 522.
3. A. T. Keatinge-Clay, D. A. Maltby, K. F. Medzihradszky, C. Khosla and R. M. Stroud, An antibiotic factory caught in action. *Nature Structural & Molecular Biology*, 2004, **11**, 888–893.
4. S. K. Hong, K. H. Kim, J. K. Park, K.-W. Jeong, Y. Kim and E. E. Kim, *FEBS Lett.*, 2010, **584**, 1240–1244.
5. K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *J. Am. Chem. Soc.* 2011, **133**, 16635–16641.
6. P. Z. Kozbial and A. R. Mushegian, *BMC Struct. Biol.*, 2005, **5**, 19.
7. K. L. Kavanagh, H. Jörnvall, B. Persson and U. Oppermann, *Cell. Mol. Life Sci.*, 2008, **65**, 3895–3906.
8. V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Sci.*, 2007, **16**, 2093–2107.
9. J. M. Crawford, A. L. Vagstad, K. C. Ehrlich, D. W. Udwary and C. A. Townsend, *ChemBioChem*, 2008, **9**, 1559–1563.
10. J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S.-C. Tsai, and C. A. Townsend, *Nature*, **461**, 2009, 1139–1143.
11. M. J. Hijarrubia, J. F. Aparicio, J. Casqueiro and J. F. Martín, *Mol. Gen. Genet.*, 2001, **264**, 755–762; J. Davison, A. Al Fahad, M. Cai, Z. Song, S. Y. Yehia, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *Proc. Natl Acad. Sci. USA*, 2012, **109**, 7642 - 7647.
12. Hellman, U., et al. (1995). Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Anal Biochem 224: 451-5.
13. L. S. Trifonov, J. H. Bieri, R. Prewo, A. S. Dreiding, L. Hoesch and D. M. Rast, *Tetrahedron*, 1983, **39**, 4243-4256.