



(51) International Patent Classification:

C12P 13/02 (2006.01) C12N 9/10 (2006.01)

A61K 31/155 (2006.01) C12N 9/00 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/EP2017/069773

(22) International Filing Date:

04 August 2017 (04.08.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/371,960 08 August 2016 (08.08.2016) US

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))
- of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(71) Applicant: GLAXOSMITHKLINE INTELLECTUAL PROPERTY DEVELOPMENT LIMITED [GB/GB]; 980 Great West Road, Brentford Middlesex TW8 9GS (GB).

(72) Inventor: LOVELOCK, Sarah, Louise; c/o GlaxoSmithKline, Gunnels Wood Road, Stevenage Hertfordshire SG1 2NY (GB).

(74) Agent: GODDARD, Carolyn Janice; GlaxoSmithKline, Global Patents (CN925.1), 980 Great West Road, Brentford Middlesex TW8 9GS (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: METHOD FOR SYNTHESISING AMIDES

(57) Abstract: The present invention relates to a method for synthesising amides that is of general applicability. The method may be performed in vitro or in vivo. Cell lines for use in the in vivo methods also form aspects of the invention. The method for synthesising a non-natural amide comprises: a. reaction of a carboxylic acid with a naturally occurring CoA ligase or a variant thereof; and b. reaction of the product of step a with an amine in the presence of a naturally occurring acyltransferase or a variant thereof; with the proviso that where the CoA ligase and acyltransferase are both naturally occurring, they are not derived from the same source species and do not act sequentially in a metabolic pathway; and with the proviso that the non-natural product is not N-(E)-p-coumaroyl-3-hydroxyanthranilic acid or N-(E)-p-caffeoyl-3-hydroxyanthranilic acid. Further, a method for producing an active pharmaceutical ingredient by the aforementioned method and host cells for carrying out said methods are envisaged.



FIELD OF THE INVENTION

The present invention relates to a method for synthesising amides that is of general applicability. The method may be performed in vitro or in vivo. Cell lines for use in the in vivo methods also form aspects of the invention.

BACKGROUND OF THE INVENTION

Amide bond formation is one of the most frequently encountered reactions in organic synthesis and amides are commonly found in active pharmaceutical ingredients, biologically active molecules, synthetic polymers, peptides and proteins. In fact, a study by the American Chemical Society Green Chemistry Institute Roundtable found that, out of a random selection of drug candidates, amide-bond formation was used in the synthesis of 84% of drug candidates (Constable et al., Green Chem. 9 (5), 411-420, 2007).

Traditional methods of amide synthesis use carboxylic acid and amine substrates and require stoichiometric coupling reagents. As the reaction proceeds via a highly reactive activated intermediate, undesirable side reactions can occur leading to the formation of unwanted by-products such as ureas. The reaction suffers from poor atom economy and there is often a significant amount of toxic and metal containing waste generated, which results in the process of amide formation being costly. Other drawbacks of traditional methods of amide synthesis include a lack of control over enantio- and chemoselectivity, and the requirement for protection group chemistry.

Chemical catalytic approaches have been developed that eliminate the requirement for stoichiometric coupling reagents and hence improve atom economy, reducing the amount of waste generated (reviewed in Pattabiraman and Bode, Nature, 480, 471-479, 2011). Boronic acid and related catalysts have found use in specialized cases, but the high temperatures required limit their widespread use. Redox based methods employing either N-heterocyclic carbenes (NHC) or metal catalysts can proceed at lower temperatures, however NHC catalysts are expensive and are typically used together with hazardous cosolvents. Furthermore, many of these approaches require aldehyde starting materials, which can be more challenging to access than the corresponding carboxylic acids.

Lipases have been used as biocatalysts to generate amide bonds in organic solvents. Advantageously, these enzymes are typically highly enantioselective (for a review, see Gotor, Bioorg Med Chem, 7, 2189-2197, 1999). However the currently studied lipases would appear to have narrow substrate specificity and furthermore, must be used in dry organic solvents to prevent unwanted hydrolysis. The requirement for dry solvents and molecular sieves makes scaling lipase catalyzed reactions challenging, furthermore they are not compatible with multi-enzyme pathways.

There are some examples of using CoA ligases together with N-acyltransferases in order to synthesise naturally occurring amides. Korean patent application 10-2009-0097388, Kang and Back (Metabolic Engineering, 11: 64-68, 2009), and Park et al. (Appl. Microbiol. Biotechnol., 81: 43-49, 2008) disclose the production of natural phenylpropanoid amides in host cells expressing 4-coumarate CoA ligase and serotonin N-hydroxycinnamoyltransferase.

Moglia and colleagues (Moglia et al., *Metabolic Engineering*, 12: 223-232, 2010) reported that *S. cerevisiae* cells transformed with genes encoding tobacco 4-coumarate CoA ligase and globe artichoke hydroxycinnamoyl CoA: shikimate/quinic acid hydroxycinnamoyltransferase cultured in the presence of p-coumaric acid or caffeic acid unexpectedly led to the production of N-(E)-p-coumaroyl-3-hydroxyanthranilic acid (YAv) and N-(E)-p-caffeoyl-3-hydroxyanthranilic acid (YAVII) respectively.

Dunn et al. (*Angew. Chem. Int. Ed.*, 54: 5137-5141, 2015) reports that in *Pseudoalteromonas* sp. two proteins, TmlU and HolE act sequentially in the biosynthetic pathway of thiomarinol. The promiscuity of the HolE/TmlU pair was investigated to generate thiomarinol analogues including, for example, pseudomonic acid C-aminocoumarin.

SUMMARY OF THE INVENTION

In a first aspect, the present invention provides a method for synthesising a non-natural amide comprising:

(a) reaction of a carboxylic acid with a naturally occurring CoA ligase or a variant thereof; and (b) reaction of the product of step (a) with an amine in the presence of a naturally occurring N-acyltransferase or a variant thereof; with the proviso that where the CoA ligase and N-acyltransferase are both naturally occurring, they are not derived from the same source species and do not act sequentially in a metabolic pathway; and with the proviso that the non-natural product is not N-(E)-p-coumaroyl-3-hydroxyanthranilic acid or N-(E)-p-caffeoyl-3-hydroxyanthranilic acid.

The method of the invention overcomes the disadvantages associated with prior art methods of synthesising amides. The method utilises enzymes, which possess molecular recognition elements making them inherently regioselective, removing the requirement to protect other functional groups present in the reactants. Furthermore, enzymes are also often naturally enantioselective. In addition, the method does not utilise, and is indeed incompatible with dry organic solvents. This minimises the amount of waste solvent generated and makes this method compatible with multi-enzyme pathways *in vivo*.

In a second aspect, the present invention provides a method for producing an active pharmaceutical ingredient, characterised in that the active pharmaceutical ingredient is produced from a non-natural amide intermediate that is synthesised by the method described above.

Cell lines co-expressing CoA ligases and N-acyltransferases for use in the methods of the invention form a further aspect of the invention. Accordingly, in a third aspect, the invention provides a host cell comprising either: (a) an expression vector comprising the coding sequence for a naturally occurring CoA ligase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and an expression vector comprising the coding sequence for a naturally occurring N-acyltransferase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring N-acyltransferase or a variant thereof in the host cell; or (b) an expression vector having at least two expression cassettes, a first expression cassette comprising the coding sequence for a naturally occurring CoA ligase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and a second expression

cassette comprising the coding sequence for naturally occurring N-acyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring N-acyltransferase or a variant thereof in the host cell; or (c) an expression vector having one expression cassette comprising the coding sequence for naturally occurring CoA ligase or a variant thereof fused in frame to a linker and then to the coding sequence for naturally occurring N-acyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the fusion protein in the host cell; with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring N-acyltransferase, the CoA ligase is not a 4-coumarate CoA ligase and the N-acyltransferase is not a serotonin N-hydroxycinnamoyl transferase; and with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring N-acyltransferase, the CoA ligase is not tobacco 4-coumarate CoA ligase 2 where the acyltransferase is the globe artichoke hct gene having accession number DQ104740 amplified using primers 5'-TCTCCATGGGTAAGATCGAGGTGAGAGAATCAACGATG -3' and 5'TCTCTCGAGTTAGATATCATATAGGAACTTGCTG-3'; and with the proviso that the expression vector of (c) does not comprise pepper serotonin N-hydroxycinnamoyl transferase fused in frame with the self-processing FDMV 2A sequence followed by 4-coumarate ligase 2 from *Arabidopsis thaliana*.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows stained SDS gels of fractions generated following Ni-NTA chromatography purification of CoA ligases and N-acyltransferases, the gels demonstrate the purity of the expressed enzymes.

FIGURE 2 shows the structures of a number of carboxylic acid substrates referred to in the examples.

FIGURE 3 is a graph showing the substrate specificity of nine CoA ligases measured using the EnzChek pyrophosphate detection assay

FIGURE 4 is an example of a HPLC trace, showing CoA ligase (PhCL) catalysed formation of a CoA ester (cinnamoyl-CoA)

FIGURE 5 is a graph showing the effect of CoA concentration on the activity of CoA ligases phl, PhCL, CBL and ipfF in the presence of N-acyltransferase HolE measured using the EnzChek pyrophosphate detection assay

FIGURE 6 shows the structures of a number of amine substrates referred to in the examples.

FIGURE 7 is a graph showing the amine substrate specificity of four N-acyltransferases CASHT, HolE, HCBT-MBP and HpA3 measured using the Ellman's assay

FIGURE 8 is a graph showing the effect of CoA concentration and the ratio of the enzymes CASHT and PhCL upon conversion as measured by HPLC.

FIGURE 9 shows LC-MS traces showing CoA ligase and N-acyltransferase catalysed amide bond formation

FIGURE 10A is a stained gel showing coexpression of N-acyltransferase CASHT and four different CoA ligases phl, PhCL, CBL and ipfF in *E. coli* cells. FIGURE 10B is a western blot corresponding to the stained gel and shows coexpression of particular CoA ligases and N-acyltransferase CASHT in *E. coli* cells.

FIGURE 11 is a stained gel showing coexpression of N-acyltransferase CASHT and CoA ligase CBL, generated from *E. coli* cells transformed with a pCDF-Duet vector carrying both CASHT and CBL genes.

FIGURE 12 shows a reaction scheme for the formation of 6-chloro-N-neopentynicotinamide

FIGURE 13 shows the structures of a number of amine and carboxylic acid substrates referred to in the examples

FIGURE 14 shows the structures of a number of secondary amines referred to in the examples

DETAILED DESCRIPTION OF THE INVENTION

As outlined above, the present invention provides a method for synthesising a non-natural amide comprising: (a) reaction of a carboxylic acid with a naturally occurring CoA ligase or a variant thereof; and (b) reaction of the product of step (a) with an amine in the presence of a naturally occurring N-acyltransferase or a variant thereof; with the proviso that where the CoA ligase and N-acyltransferase are both naturally occurring, they are not derived from the same source species and do not act sequentially in a metabolic pathway; and with the proviso that the non-natural product is not N-(E)-p-coumaroyl-3-hydroxyanthranilic acid or N-(E)-p-caffeoyl-3-hydroxyanthranilic acid.

In the context of the invention, a non-natural amide is a compound comprising an amide bond that is not the product of metabolism of any species and therefore, does not exist in nature.

The non-natural products N-(E)-p-coumaroyl-3-hydroxyanthranilic acid (YAv) and N-(E)-p-caffeoyl-3-hydroxyanthranilic acid (YAvII) were disclosed in Moglia et al., *Metabolic Engineering*, 12: 223-232, 2010.

The proviso that the naturally occurring CoA ligase and naturally occurring N-acyltransferase are not derived from the same source species and do not act sequentially in a metabolic pathway is intended to exclude enzyme pairs where the natural product of the CoA ligase is a substrate of the N-acyltransferase in vivo.

The method of the invention is one of general applicability and it is envisaged that substantially any amide could be formed by selection of the appropriate carboxylic acid and amine substrates. The flexibility of the claimed method is achieved by careful selection of enzymes having appropriate substrate specificities. In other words, for the reaction described in step (a), a CoA ligase is employed that has specificity for the carboxylic acid substrate. For the reaction described in step (b), an N-acyltransferase with specificity for the product of step (a) and the amine substrate is required.

Most enzymes, including CoA ligases and N-acyltransferases catalyse particular reactions in vivo. However, the high level of specificity observed in vivo results, at least in part, from the fact that, in vivo, enzymes are only exposed to the “correct” substrate(s). However, enzymes can be promiscuous and accept a range of substrates. Examples 3 and 14 clearly demonstrate that CoA ligases are capable of accepting multiple carboxylic acid substrates. By way of example, the CoA ligase phl catalyses the formation of CoA thioesters from free CoA and carboxylic substrates **3-7, 9, 11 and 15**. Similarly, Examples 7, 9, 14 and 15 demonstrate that N-acyltransferases are capable of accepting multiple CoA thioesters and amine substrates. By way of example, HolE accepts octanoyl CoA and the CoA thioesters generated by carboxylic acids **5, 6, 7 and 16** and amines **19, 21, 24-26 and 28** including amines containing both aromatic moieties and non-aromatic heterocycles. Of course, any particular CoA ligase or N-acyltransferase will not accept all substrates. However, the panel of CoA ligases exemplified herein is capable of catalysing the ligation of carboxylic acids across of broad range of chemical space including acids containing aromatic, heteroaromatic, allylic and aliphatic groups. The panel of N-acyltransferases exemplified herein is similarly capable of accepting a broad range of thioester and amine substrates, including both primary and secondary amines. To synthesise any particular amide, it will be apparent that it is first necessary to identify enzymes capable of accepting the carboxylic acid and amine substrates. Where the substrate(s) is (are) identical to the natural substrate(s) for a known enzyme or to substrates shown to be accepted by a particular enzyme in the Examples of the present application, it is clear that the known enzyme could be used. In many cases, however, it will not be clear which enzymes will be suitable. In such cases, it would be possible to screen a panel of enzymes to identify those with activity. Performing these screens would be routine to one of ordinary skill in the art, and suitable screens are described in Examples 3 and 6, which describe a screen for CoA ligases utilising the commercially available EnzChek pyrophosphate (PPi) assay kit; Example 4, which describes a HPLC screen for CoA ester formation; Examples 5, 8, 11 and 14 which describe LC-MS screens for amide formation; and example 7 which describes a screen for N-acyltransferases using an Ellman's assay. CoA ligases are considered to have activity for a particular substrate in the EnzCheck pyrophosphate (PPi) assay if the mean V_{max} in the presence of the substrate is greater than the mean V_{max} in the absence of any substrate. In one embodiment, a CoA ligase is considered active at a particular substrate when V_{max} in the presence of the substrate is greater than or equal to the sum of the mean V_{max} in the absence of any substrate plus one standard deviation. In another embodiment, a CoA ligase is considered active at a particular substrate when V_{max} in the presence of the substrate is greater than or equal to the sum of the mean V_{max} in the absence of any substrate plus two standard deviations.

An alternative technique for screening activity of CoA ligases is monitoring the CoA thioester product of reaction (a) by HPLC as described in Example 4. Accordingly, in one embodiment, a CoA ligase may be considered to have activity if a peak corresponding to the CoA thioester product is present in the HPLC trace. In one embodiment, the area % total of the CoA thioester peak should be greater than or equal to 1. In another embodiment, the area % total of the CoA thioester peak should be greater than or equal to 5.

N-Acyltransferases may be considered to have activity for the product of reaction (a) and amine, if a peak corresponding to the amide product is detected by LC-MS. In one embodiment, the area % total of the amide peak should be greater than or equal to 1. In another embodiment, the area % total of the amide peak should be greater than or equal to 5. An alternative technique for screening activity of N-acyltransferases is use of an Ellman's assay. Accordingly, in one embodiment, an N-acyltransferase is considered to have activity for the product of reaction (a) and amine if V_{max} in the

presence of the product of reaction (a) and amine is greater than or equal to the sum of the mean V_{max} in the absence of any substrate plus one standard deviation. In another embodiment, an N-acyltransferase is considered active towards a particular substrate when the V_{max} in the presence of the product of reaction (a) and amine is greater than or equal to the sum of the mean V_{max} in the absence of any substrate plus two standard deviations.

In the context of the present invention, a CoA ligase is an enzyme which catalyses the formation of CoA thioesters from free CoA and carboxylic acids at the expense of ATP. Certain naturally occurring CoA ligases are known in the literature or described herein. These include the phenylacetate CoA ligase (pHl) and aminoacyl CoA ligase (pHl2) from *Penicillium chrysogenum*: (Uniprot O74725 and B6HL15), the cinnamate CoA ligase from *Petunia hybrid* (PhCL) (Uniprot I3PB37), the 4-chlorobenzoate CoA ligase (CBL) from *Alcaligenes sp.* ALP83: (Uniprot Q8GN86), the CoA ligase from *Sphingomonas lbu-2* (ipfF) (Uniprot A1E027), the CoA ligase from *Streptomyces carzinostaticus* (ScCL) (Uniprot Q84HC5), the CoA ligase from *Photobacterium luminescens* (stlB) (Uniprot Q7N526), the CoA ligase from *Rhodopseudomonas palustris* (RpCL) (Uniprot O07454), the CoA ligase from *Leptothrix cholodnii* (LcCL) (Uniprot B1Y4L8), the CoA ligase from *Acinetobacter radioresistens* (ArCL) (Uniprot C6RPQ0), the CoA ligase from *Mycobacterium smegmatis* (MsCL) (Uniprot A0QZQ6), the cinnamate CoA ligases from, *Streptomyces coelicolor* A3(2) (see *J Bacteriol*, 185(1): 20-27, 2003), soybean (four isoforms; see *J Biol Chem* 278, 2781-2786, 2003) and *Arabidopsis thaliana* (three isoforms; see *J Bacteriol*, supra). Further naturally occurring CoA ligases may be identified by searching protein or DNA databases. Enzymes in protein databases are often assigned an EC number based on their similarity to known proteins. EC number 6.2.1 is assigned to ligases that form carbon to sulphur bonds, and covers both ATP dependent CoA ligases (which are useful in the present invention) and GTP dependent CoA ligases (which are not used). There are several subdivisions within EC number 6.2.1 (e.g. 6.2.1.1 refers to acetate CoA ligases). Accordingly, putative CoA ligases may be identified by searching within protein databases for proteins having particular EC numbers. The skilled person will appreciate that the subdivisions within EC 6.2.1 may give an indication of likely substrate specificity.

Protein or DNA databases can be searched to identify protein or DNA sequences similar to known CoA ligases. The skilled person might expect sequence identity between CoA ligases with similar activity to be high. However, CoA ligases with activity for different substrates might exhibit lower sequence identity. Example 1 shows that levels of sequence identity between CoA ligases can be low (the cyclohexanoic acid CoA ligase from *Rhodopseudomonas palustris* having uniprot ID O07454 and the cyclohexane carboxylate CoA ligase from *Mycobacterium smegmatis* having uniprot ID A0QZQ6 share just 38% sequence identity at a DNA level). Accordingly searching DNA databases using a known CoA ligase as a query sequence can be used to identify other putative CoA ligases. In one embodiment, searches could be used to identify DNA sequences having at least 40% sequence identity to the query sequence. Alternatively, other putative CoA ligases can be identified by searching protein databases using a known CoA ligase as a query sequence. In one embodiment, searches could be used to identify protein sequences having at least 70% sequence identity to the query sequence.

Further CoA ligases may be identified by DNA hybridisation using a suitable cDNA or genomic DNA library (derived from any species) and a single stranded DNA probe of at least 100 bp derived from a query CoA ligase gene or cDNA.

DNA hybridisation is affected by the hybridisation conditions as outlined in the following relationship:

$$T_m = 81.5 + 16.6 \log_{10} \left(\frac{[Na^+]}{1.0 + 0.7[Na^+]} \right) + 0.41 (\%[G+C]) - \frac{500}{n} - P - F$$

where T_m = the temperature at which 50% of the probe molecule is ssDNA in °C

$[Na^+]$ = Molar concentration of sodium ions

$\%[G+C]$ = percentage of G+C bases in probe sequence

n = length of probe sequence in bases

P = temperature correction for % mismatched base pairs ($\sim 1^\circ\text{C}$ per 1% mismatch)

F = correction for formamide concentration ($= 0.63^\circ\text{C}$ per 1% formamide)

When it is desired to identify a CoA ligase having very different substrate specificity to the query CoA ligase, the conditions for DNA hybridisation should not be highly stringent. Where the substrate specificity required is not too dissimilar, more stringent conditions could be used. Low stringency conditions to identify alternative CoA ligases would involve hybridisation at a temperature of 40-48°C below the T_m (calculated assuming no mismatched base pairs) for any particular sodium ion concentration/formamide concentration combination. More stringent conditions would involve hybridisation at a temperature closer to the T_m (calculated assuming no mismatched base pairs) for any particular sodium ion concentration/formamide concentration combination. In practice, the skilled person would appreciate that the hybridisation reaction could take place with stepwise increasing temperature or reduced salt/formamide concentrations in order to reduce the number of hits to a manageable number. The hybridising clones could then be expressed, the proteins purified and screened according to methods well known in the art and exemplified herein.

In some embodiments, the CoA ligase employed is a naturally occurring CoA ligase. In the context of the present invention, a naturally occurring CoA ligase is one that naturally occurs in any species and includes all allelic variations (variant forms of the enzymes that are coded by different alleles at the same locus).

In another embodiment, the CoA ligase is a variant of a naturally occurring CoA ligase. In the context of the present invention, a variant of a naturally occurring CoA ligase is an enzyme that has an amino acid sequence that differs from any naturally occurring CoA ligase, but which has activity for the substrate carboxylic acid as determined by the methods outlined above. Certain variant CoA ligases differ from naturally occurring CoA ligases simply by the inclusion of an N or C terminal tag whose inclusion facilitates purification e.g. a His tag. In other embodiments, the CoA ligase has been modified to give improved properties, for example improved activity for a particular substrate. The process of modifying enzymes to improve enzymatic characteristics (e.g. substrate specificity) is known as directed enzyme evolution.

Techniques for directed enzyme evolution are reviewed by McLachlan et al. in chapter 3 of *Biocatalysis for the Pharmaceutical Industry: Discovery, Development and Manufacturing* (Ed. J Tao et al., 2009, John Wiley and Sons Asia (Pte) Ltd.) and are well known in the art. Briefly,

libraries of genes encoding variant enzymes are created which are then screened to identify enzymes with improved characteristics. DNA libraries can be made using numerous techniques and the method selected is often dependent upon the capacity of the available high throughput screening methods. Random libraries can be produced using techniques such as error prone PCR, gene shuffling and mutator strains. Screening random libraries requires assays methods which can quickly process very large numbers of variants (e.g. 10^6 - 10^8 variants) such as microfluidic assays or fluorescence activated cell sorting. Commercially available fluorescent assays could be exploited in microfluidics such as abcam's pyrophosphate assay kit (ab112155) to measure CoA ligase activity or abcam's thiol quantification assay kit (ab112158) to detect acetyltransferase activity. For FACS, an enzyme coupled assay could be developed to provide a fluorescence based response using similar enzymes to those used in the commercially available PiPer pyrophosphate assay kit (Invitrogen P22062). An alternative approach is to generate smaller more focused libraries using structure-guided saturation mutagenesis. This involves targeting specific amino acid positions within the protein which are believed to have a function in substrate binding, enzyme activity and/or stability, and randomising those positions with all 20 canonical residues or a reduced amino acid set. Following expression of a library of CoA ligase mutants produced in this way, variants with activity towards a novel substrate could be identified using the methods outlined herein (the EnzCheck pyrophosphate (PPi) assay/UPLC-MS) or other methods known in the art.

Amino acids responsible for substrate specificity of at least certain CoA ligases have been identified using sequence alignments, reviewing X-ray crystal structures and *in vitro* mutagenesis studies (see in particular, Stuible and Kombrink, J Biol Chem., 276(29) 26893-26897, 2001; Stuible et al., FEBS Letts, 467: 117-122, 2000; Kaneko et al., J Bacteriol., 185(1) 20-27, 2003; Conti et al., 16(14) 4174-4183, 1997 and Lindermayr et al., J Biol Chem, 278(5): 2781-2786, 2003). Key residues affecting the substrate specificity of 4 coumarate CoA ligases include the residues at positions 256, 293, 294, 320, 322, 346 and 354, 355 of the Arabidopsis thaliana 4-coumarate CoA ligase isoform 2. The corresponding residues in other CoA ligases could be identified by alignment of their sequences with the sequence of Arabidopsis thaliana 4-coumarate CoA ligase isoform 2 using commercially available computer software (e.g. the PILEUP program of the Genetics Computer Group program package, version 10.0).

An example of modifying the substrate specificity of a CoA ligase by site directed mutagenesis is described by Kaneko et al. (supra). One mutated enzyme (a 4-coumarate CoA ligase from *Streptomyces coelicolor* A3 (2) having the double substitution A294G/A318G) was capable of using ferulate as a substrate, although the parent enzyme was incapable of doing this. A further example is provided by Gm4CL3dVal367 (a mutated version of soybean 4-coumarate CoA ligase isoform 3 in which valine 367 has been deleted), which had activity for sinapate and 3,4-dimethoxy cinnamate, even though the parent enzyme was not capable of using either substrate. It will be evident to the skilled reader that the techniques for directed enzyme evolution described herein could be used to improve activity for a particular carboxylic acid substrate or even to engineer substrate specificity for a related carboxylic acid substrate.

In the context of the present invention, an N-acyltransferase is an enzyme which catalyses the transfer of an acyl group from a thioester onto an amine or hydrazine group of a second molecule to form an amide bond. Certain naturally occurring N-acyltransferases are known in the literature or described herein. These include the holothin N-acetyl transferase (HoIE) from *Pseudoalteromonas* sp having uniprot ID F8J3H2, hydroxycinnamoyl-CoA: serotonin N-transferase from *Capsicum annuum* (uniprot ID Q9ATJ3), hydroxycinnamyl/ benzoyl-CoA:anthranilate transferase from *Dianthus caryophyllus* (uniprot ID O24645), L-azetidine-2-carboxylic acid N-acetyl transferase

(MpR1) and D-amino acid N-acyltransferase (HpA3) from *Saccharomyces cerevisiae* (Uniprot IDs E9P8D2 and P39979 respectively) and an aryl amine N-acyltransferase from *Bacillus cereus* (BcAT) (Uniprot Q81AS3). Other N-acyltransferases include enzymes having the following Uniprot IDs: Q00267, O86309, Q9HUY3, Q5YYQ3, A9ZPJ7, Q9FNP9, A9ZPJ6, O80467, P21340, Q5CPU3, Q94521, G6FA50, Q81MQ2, Q836H8, Q7A2S0, P80969, Q9SMB8, P9WP21, G3XD76, Q4JQJ5, A6VCX3, P30419, P16426, Q9R381, D8FSU9, B7SP66, D8FSU8, Q09927, Q9ZV05, Q5D8C0, Q8H9D9, P76112, V4HKD0, A0A0F5VBE7, Q8RXB6, Q8RXB8, Q8RXB7, E3HPZ9, L8EWB7, V6JJA0, Q08649 and Q03330.

This application claims priority from US provisional application no. 62/371960 filed 8 August 2016. This provisional application refers to acetyltransferases instead of N-acyltransferases. The definition of acetyltransferases given therein is “an enzyme which catalyses the transfer of an acyl group from a thioester onto an amine or hydrazine group of a second molecule to form an amide bond” which is identical to the definition of N-acyltransferase given above. It is also noted that several examples of acetyltransferases listed in this provisional application are identical to those listed above. It is thus clear that these terms are intended to be synonymous.

In one embodiment, the N-acyltransferase used in the methods or host cells of the invention is an acetyltransferase defined as “an enzyme which catalyses the transfer of an acyl group from a thioester onto an amine or hydrazine group of a second molecule to form an amide bond”.

Thus, in one aspect, the invention provides the following embodiments:

Embodiment 1. A method for synthesising a non-natural amide comprising:

- a. reaction of a carboxylic acid with a naturally occurring CoA ligase or a variant thereof; and
- b. reaction of the product of step a with an amine in the presence of a naturally occurring acetyltransferase or a variant thereof;

with the proviso that when the CoA ligase and acetyltransferase are both naturally occurring, they are not derived from the same source species and do not act sequentially in a metabolic pathway; and

with the proviso that the non-natural product is not N-(E)-p-coumaroyl-3-hydroxyanthranilic acid or N-(E)-p-caffeoyl-3-hydroxyanthranilic acid.

Embodiment 2. A method according to embodiment 1, wherein steps a and b take place in vitro, and wherein step a takes place in the presence of ATP and CoA.

Embodiment 3. A method according to embodiment 2, wherein the product of step a is not purified prior to step b.

Embodiment 4. A method according to embodiment 2 or embodiment 3, wherein the naturally occurring CoA ligase or a variant thereof is attached to a solid substrate.

Embodiment 5. A method according to any one of embodiments 2 to 4, wherein the naturally occurring acetyltransferase or a variant thereof is attached to a solid substrate

Embodiment 6. A method according to embodiment 1, wherein steps a and b take place in vivo in a host cell expressing the naturally occurring CoA ligase or variant thereof and the naturally occurring acetyltransferase or variant thereof.

Embodiment 7. A method according to embodiment 6, wherein the host cell is a prokaryotic host cell.

Embodiment 8. A method according to any preceding embodiment, which further comprises a step of purification of the non-natural amide product.

Embodiment 9. A method for producing an active pharmaceutical ingredient, characterised in that the active pharmaceutical ingredient is produced from a non-natural amide intermediate that is synthesised as defined in any one of embodiments 1 to 8.

Embodiment 10. A host cell comprising either:

- a. an expression vector comprising the coding sequence for a naturally occurring CoA ligase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and an expression vector comprising the coding sequence for a naturally occurring acetyltransferase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring acetyltransferase or a variant thereof in the host cell; or
- b. an expression vector having at least two expression cassettes, a first expression cassette comprising the coding sequence for naturally occurring CoA ligase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and a second expression cassette comprising the coding sequence for naturally occurring acetyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring acetyltransferase or a variant thereof in the host cell; or
- c. an expression vector having one expression cassette comprising the coding sequence for naturally occurring CoA ligase or a variant thereof fused in frame to a cleavable linker and then to the coding sequence for naturally occurring acetyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the fusion protein in the host cell;

with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring acetyltransferase, the CoA ligase is not a 4-coumarate CoA ligase where the acetyltransferase is a serotonin N-hydroxycinnamoyl transferase; and with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring acetyltransferase, the CoA ligase is not tobacco 4-coumarate CoA ligase 2 where the acetyltransferase is the globe artichoke hct gene having accession number DQ104740 with the proviso that the expression vector of (c) does not

comprise pepper serotonin N-hydroxycinnamoyltransferase fused in frame with the self processing FDMV 2A sequence followed by 4-coumarate ligase 2 from *Arabidopsis thaliana*.

Embodiment 11. A host cell according to embodiment 10 which is a prokaryotic cell.

Further naturally occurring N-acyltransferases (or acetyltransferases as defined herein) may be identified by searching protein or DNA databases as outlined above. EC number 2.3.1 is assigned to acyltransferases. Putative acyltransferases may be identified by searching within protein databases for proteins in this class. The skilled person will appreciate that the subdivisions within EC 2.3.1 may give an indication of likely substrate specificity.

Protein or DNA databases can be searched to identify protein or DNA sequences similar to known N-acyltransferases (or acetyltransferases as defined herein). Kang et al. (Plant Physiol, 140, 704-715, 2006) teaches that the amino acid sequences derived from a variety of plants have between 73 and 76% identity and that pepper serotonin N-hydroxycinnamoyl transferase has 83%, 80% and 71% amino acid identity with tomato, potato and tobacco tyramine N-hydroxycinnamoyltransferases. N-Acyltransferases (or acetyltransferases as defined herein) from very different organisms or having very different substrate specificity are likely to have even lower sequence specificity. Accordingly, searching protein or DNA databases using a known N-acyltransferase (or known acetyltransferase as defined herein) as a query sequence can be used to identify other putative N-acyltransferases (or acetyltransferases as defined herein). In one embodiment, searches could be used to identify DNA sequences having at least 40% sequence identity to the query sequence. In another embodiment, searches could be used to identify protein sequences having at least 70% sequence identity to the query sequence.

Further N-acyltransferases (or acetyltransferases as defined herein) may be identified by DNA hybridisation using a suitable cDNA or genomic DNA library (derived from any species) and a single stranded DNA probe of at least 100 bp derived from a query N-acyltransferase gene or cDNA in the manner described above for CoA ligases. As explained above, the skilled person would appreciate that the hybridisation reaction could take place with stepwise increasing temperature or reduced salt/formamide concentrations in order to reduce the number of hits to a manageable number. The hybridising clones could then be expressed, the proteins purified and screened according to methods well known in the art and exemplified herein.

In some embodiments, the N-acyltransferase (or acetyltransferase as defined herein) employed is a naturally occurring N-acyltransferase (or acetyltransferase as defined herein). In the context of the present invention, a naturally occurring N-acyltransferase (or acetyltransferase as defined herein) is one that naturally occurs in any species and includes all allelic variations (variant forms of the enzymes that are coded by different alleles at the same locus).

In another embodiment, the N-acyltransferase (or acetyltransferase as defined herein) is a variant of a naturally occurring N-acyltransferase (or acetyltransferase as defined herein). In the context of the present invention, a variant of a naturally occurring N-acyltransferase (or acetyltransferase as defined herein) is an enzyme that has an amino acid sequence that differs from any naturally occurring N-acyltransferase (or acetyltransferase as defined herein), but which has activity for the CoA thioesters and amine substrates as determined by the methods outlined above. Certain variant N-acyltransferases (or acetyltransferases as defined herein) differ from naturally occurring N-acyltransferases (or acetyltransferases as defined herein) simply by the inclusion of an N or C

terminal tag whose inclusion facilitates purification e.g. a His tag. Example 2 describes the use of Hydroxycinnamyl/ benzoyl-CoA:anthranilate transferase from *Dianthus caryophyllus* that is an N-terminal 6-His-MBP (maltose binding protein) fusion which is used for increased protein solubility.

In other embodiments, the N-acyltransferase (or acetyltransferase as defined herein) has been modified to give improved properties, for example improved activity for a particular substrate in the same manner as described above in relation to CoA ligases.

Amino acids responsible for substrate specificity of at least certain N-acyltransferases (or acetyltransferases as defined herein) have been identified by sequence alignments, reviewing X-ray crystal structures and in vitro mutagenesis studies (see e.g. Kang et al., Plant Physiol, 140, 704-715, 2006). Key residues affecting the substrate specificity of pepper serotonin N-hydroxycinnamoyl transferase are residues 129 to 165 and the key residues affecting the substrate specificity of tyramine N-hydroxycinnamoyl transferase are residues 125 to 160. The corresponding residues in other N-acyltransferases (or acetyltransferases as defined herein) could be identified by sequence alignment as described above in relation to CoA ligases.

An example of modifying the substrate specificity of an N-acyltransferase (or an acetyltransferase as defined herein) by site directed mutagenesis is described by Kang et al. (supra). One mutated enzyme (the pepper serotonin N-hydroxycinnamoyl transferase having the substitution Y149F) used tyramine as a substrate instead of serotonin. It will be evident to the skilled reader that the techniques for directed enzyme evolution described herein could be used to evolve N-acyltransferase (or acetyltransferase as defined herein) variants with specificity for particular CoA thioesters and amine substrates.

The amide formation reaction that is the subject of the present invention may take place in vitro or in vivo.

In one embodiment, steps (a) and (b) take place in vitro in the presence of ATP and CoA. In one embodiment, purified enzyme preparations are utilised. In the context of this invention, a purified enzyme preparation has less than 5% by weight other proteins, in one embodiment this is less than 1% by weight. Where naturally occurring enzymes are used, the native enzymes may be purified from tissue derived from the source species. More typically however, the enzymes are produced recombinantly according to known techniques. In brief, the enzymes may be produced by transfection of a host cell with an expression vector comprising the coding sequence for the enzyme operatively linked to conventional regulatory control sequences capable of controlling the expression in and optionally secretion from a host cell. Following transfection, the transfected cell is then cultured by conventional techniques to produce the enzymes, followed by harvesting and purification according to known techniques (e.g. chromatography). In some embodiments, the enzymes could be expressed together with a signal sequence resulting in their secretion into the growth media to facilitate purification.

Prokaryotic cells may be used as host cells. For example, various strains of *E. coli*, *B. subtilis*, *Streptomyces*, and other bacilli are well known as host cells in the field of biotechnology. Eukaryotic cell lines may also be used as host cells e.g. strains of yeast, insect cells as well as mammalian cells. Suitable expression vectors for use in any particular host cell are commercially available or could be synthesised by well-known techniques.

In order to exhibit enzymatic activity, the enzymes must be correctly folded. Occasionally, during overexpression, incorrectly folded proteins are formed (for example in inclusion bodies).

Modification of the host cell type, vector or expression conditions can affect the folding of the expressed protein. Techniques for refolding proteins are also known. Accordingly, the skilled person would have no difficulty in preparing correctly folded enzymes for use in the invention.

In some embodiments, the enzymes could be attached to a solid substrate (e.g. beads) or immobilized according to known methods. The skilled person would recognise that this facilitates purification of the products of the biocatalytic reaction.

In some embodiments, the product of step (a) is purified before step (b). In other embodiments, the product of step (a) is not purified prior to step (b). In such embodiments, CoA should act in a catalytic manner (this is because it is utilised in step (a) but regenerated in step (b)). Examples 5 and 6 explore the requirement for CoA for particular CoA ligase/N-acyltransferase (CoA/acetyltransferase) pairs. For most enzyme pairings, the reaction rate increased with increasing concentration of CoA. However, the overall conversion was higher with lower concentrations of CoA as determined by LC-MS. Any particular method can be optimised for levels of CoA utilising the methods outlined in the examples. Similarly, any particular method can be optimised for the best ratio of CoA ligase to N-acyltransferase (or acetyltransferase as defined herein). Other conditions (e.g. buffers etc.) may also be optimised according to example 12 and other methods known in the art.

In certain embodiments, the method also comprises a step of purification of the amide product. In the context of this invention, the purified amide product contains less than 5% impurities by weight, more particularly, less than 1% impurities by weight.

In a separate embodiment, steps (a) and (b) take place in vivo in a host cell expressing the naturally occurring CoA ligase or variant thereof and the naturally occurring N-acyltransferase or variant thereof (or naturally occurring acetyltransferase as defined herein, or variant thereof).

In brief, the enzymes may be produced by co-transfection of a host cell with either two expression vectors, each comprising the coding sequence for one enzyme operatively linked to conventional regulatory control sequences capable of controlling the expression in the host cell, by transfection of a host cells with a single vector having two expression cassettes, each comprising the coding sequence for one enzyme operatively linked to conventional regulatory control sequences capable of controlling the expression in the host cell, or by transfection of host cells with a single vector having one expression cassette comprising the coding sequence for naturally occurring CoA ligase or a variant thereof fused in frame to a cleavable linker and then to the coding sequence for naturally occurring N-acyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the fusion protein in the host cell. Following transfection, the transfected cell is then cultured by conventional techniques to produce the enzymes. Linkers capable of cleaving to liberate two separate enzymes are known in the art (see e.g. Chen et al. Adv. Drug. Deliv. Rev., 65, 1357-1369, 2013)

Prokaryotic cells may be used as host cells. For example, various strains of *E. coli*, *B. subtilis*, *Streptomyces*, and other bacilli are well known as host cells in the field of biotechnology. Eukaryotic cell lines may also be used as host cells e.g. strains of yeast, insect cells as well as mammalian cells. Suitable expression vectors for use in any particular host cell are commercially available or could be prepared by well-known techniques.

Where the method takes place in vivo, the only process step is to add the carboxylic acid and amine substrates to the culture medium. It is noted that ATP and CoA will be produced by the host cell itself, so this does not need to be added as a supplement. In one embodiment, the method also comprises a step of purification of the amide product. In the context of this invention, the purified amide product contains less than 5% impurities by weight, more particularly, less than 1% impurities by weight.

It will be appreciated that the non-natural amide products may have a variety of uses, for example as active pharmaceutical ingredients for human or veterinary use, as plant protection products or in the fine chemical industry. The non-natural amide products may also form intermediates in the synthesis of other products, such as active pharmaceutical ingredients for human or veterinary use or plant protection products. Accordingly, the invention also comprises a method for producing an active pharmaceutical ingredient for human or veterinary use or a plant protection product which comprises a step of synthesising a non-natural amide intermediate as described above. It will be appreciated that the active pharmaceutical ingredient for human or veterinary use or plant protection product may be a multi-step process and the step of non-natural amide formation can take place at any point in the multi-step process. In other words, the active pharmaceutical ingredient for human or veterinary use or plant protection product need not be directly synthesised from the non-natural amide product of the claimed invention. Example 13 shows that this method can be worked on a preparative scale to prepare active pharmaceutical ingredient-like molecules. The product of Example 13 can be used as an intermediate in the production of the active pharmaceutical ingredient, losmapimod.

EXAMPLES

Example 1: Sequence information and cloning

1.1 CoA ligases

Eight CoA ligases were identified from a review of literature and these were labelled as phl, PhCL, CBL, ipfF, ScCL, phl2, stlB and RpCL. A search of the Uniprot database for proteins labelled as cyclohex-1-ene-1 carboxylate CoA ligases returned 94 sequences and three of these sequences were chosen at random. The putative proteins LcCL, ArCL and MsCL have varying percentage identity to a known cyclohex-1-ene-1 carboxylate CoA ligase labelled as RpCL (77%, 60% and 38% respectively). All CoA ligase genes were synthesized at Genscript and the DNA sequences were codon optimised for *E. coli* expression. The genes were cloned into pET28 using NdeI and NotI restriction sites so that the expressed protein contains an N-terminal 6-His tag:

phl from *Penicillium chrysogenum*: (Uniprot O74725) *Folia. Microbiol*, 2011, (56), 246-252

MGSSHHHHHHSSGLVPRGSHMVFLPPKESGQLDPIPDNIPSEFMLNERYGRVRHASSRDPYTCGI
TGKSYSSKEVANRVDSLARSLSKEFGWAPNEGSEWDKTLAVFALNTIDSLPLFWAVHRLGGVLT
NASYSAELTHQLLDSKAKALVTCVPLLSISLEAAKAGLPKNRIYLLDVPEQLLGGVKPPAGYKSVS
ELTQAGKSLPPVDELRSWAGEGARRTAFVCYSSGTSGLPKGVMISHRNVIANLTLQIKAFEQNYRDG
GGTKPASTEVALGLLPQSHIYALVVIGHAGAYRGDQTIVLPKFELKSYLNAIQQYKISALFLVPPIIHM
LGTQDVCSKYDLSSVTSFLTGAAPLGMETAADFLKLYPNILIRQGYGLTETCTVVSSTPHDIWLGS
SGALLPGVEARIVTPENKEITTYDSPGELVVRSPSVVLGYLNNEKATAETFDVGWMRTGDEAVIRRS
PKGIEHVFIVDRIKELIKVKGQLQVAPAELEAHILAPDVSDCAVIAIPDDRAGEVPKAI/VKSASAGSDE
SVSQALVKYVEDHKARHKWLKGGIRFVDAIPKSPSGKILRRLIRDQEKEARRKAGSKI

PhCL from *Petunia hybrida*: (Uniprot I3PB37) *The Plant Cell*, 2012 (24), 2015-2030

MGSSHHHHHHSSGLVPRGSHMPMETETNQGDLIFRSKLPDIYIPKHLPLHSYCFENISEFSSRPCLI
 NGANNHIYTYADVELTSRKVAAGLNKLGIIQKQDTIMILLPNSPEFVFVAFMGASYLGAISTMANPLFTP
 AEVVKQAKASNAKLIITQACFVNKVKDYAFDNNLNVICIDSAPEGCIHFSELTQADEHDIPDVKIQSDD
 VVALPYSSGTTGLPKGVMLTHKGLVTSVAQQVDGENANLYMHSEDLVLMCVLPLFHIYSLNSVLLCG
 LRVGAAILIMQKFDIVQFCELIKVKYKVTIGPFVPPIVLAIKSPVVDNYDLSSVRTVMSGAAPLGKELE
 DAVRIKFPNAKLGQGYGMTEAGPVLAMCLAFAKEPFDIKSGACGTVVRNAEMKIVDPDTGCSLPRN
 QPGEICIRGDQIMKGYLNDPAATTRIDKEGWLHTGDIGYIDNDELFIVDRLKELIKYKGFQVAPAE
 EALLLNHPNISDAAVVPMKDEQAGEVPVAFVVRNSGSDITEDEVKDFVSKQVIFYKRIKRVFFVETV
 PKSPSGKILRKDLRARLAAGVPN

CBL from *Alcaligenes sp.* ALP83: (Uniprot Q8GN86) *Biochemistry*, 2007, (46), 14487-14499

MGSSHHHHHHSSGLVPRGSHMQTVNEMLRRAATRAPDHICALAVPARGRLRLTHAELRARVEAVAA
 RLHADGLRPQQRVAVVAPNSADVIAIALHRLGAVPALLNPRLKSAELAEIKRGEMTAAVIAVGRQ
 VADAIQSGSGARIIFLGLDVRDGEPPSYGPIEDPQREPAQPAFIFYTSGTTGLPKAAIIPQRAAESR
 VLFMSTQVGLRHGRHNVVLGLMPLYHVVGFFAVLVAALALDGTYYVVEEFRPVDALQLVQQEQVT
 SLFATPTHLDALAAAAAHAGSSSLKLDLRLHVTFAAGATMPDAVLETVHQHLPGEKVNIYGTTEAMNSL
 YMRQPKTGTEMAPGFFSEVRIVRIGGGVDEIVANGEEGELIVAASDSAFVGYLNQPPQATAEKLQDG
 WYRTSDVAVWTPEGTVRILGRVDDMIISGGENIHPSEIERVLGTAPGVTEVVIGLADQRWGQSVT
 ACVVPRLGETLSADALDTFCRSSELADFKRPKRYFILDQLPKNALNKVLRRQLVQQVSS

ipfF from *Sphingomonas lbu-2*: (Uniprot A1E027) *Microbiology*, 2013 (159), 621-632

MGSSHHHHHHSSGLVPRGSHMLARDLVKRCARNYPTKTAYLCGERSRSWREMDQRSDRFVAL
 QQLGHRPGEAVAILTQESIEVYEHFFACMKIAAPRVGLNTGYVWPEMLHVLKDSEVKFLLLDTRCR
 HLLAERLGELKALGITLIGYGAGHGLERDYESSLLATAEGEPHWPALAPDDILFVSYSYSGTTGVPKG
 VMLTQEGGVNCLHSLISFGFGPDDVWYMPAASAWVVILNAFGLGNGMTTVIPDGGYQLQAYLRDI
 ERFRVTVMGLVPTMLQRAIVEIQTNPVYDLSSLRMVVYGSSPATPKLIRDARATFKGIKLLQAYAMT
 EATGGWISYLTADHEHALREEIELLSVGRIGIHYDCSIRDESGQPVPPIGQSGEIWLRGNTMMKGY
 RNLPEATAEAMPDGLWRLTNDIGRLDERGYLYLLDRQKFLIITGAVNVFPTTVEAILVEHPAVEEVAVV
 GVPHPEWGEAVVAVVVRKPSHRDVTQALIDFCHGKLSRPETPKHVVFVDELPKTSNAKLKKGELK
 KWLSGGAVPLPWQLEVA

ScCL from *Streptomyces carzinostaticus*: (Uniprot Q84HC5) *J. Am. Chem. Soc.* 2007, (129), 7728-7729

MGSSHHHHHHSSGLVPRGSHMHETAAAPAGFVPWPDDVAARYTAAGHWEGRSLGTHLAEAA
 RKVPEAVCLVDGPVRMSYSELMARADGA AVRMRGLGIRPADRVVQLPNCWEHVVTMACLRGLG
 ALPIWALPQYRHRELSGVVTHARASALVVPDVYREFDHQALAEVAEAQPTVRHVLVAGSDVRPD
 SVDLRALCEPLDADEAARVAAELDRSAPRGEEVAMKLKSGGTTGLPKLVARTHNDLSYMIKRAAQV
 CGFGRDVTYLA VLP L GHGFPNTGPGVLGTL LAGGRVVISGSPAPEAAAFALMERERTATSVPVPAIV
 MRWLQYRDERPGADLGSELMQVGASRLEPEVARQVGPKLGCRLQQVFGMAEGLLCLTRLDDPD
 DVVHYTQGRPISPDDIEIRVVDPEGRTVGVGEPGALLTRGPYTPRGYYDSPSANARAFTPDGWYRT
 GDLVRRTPDGNLIVVGREKDLINRGGEKINAEVEGFVAVQVDGVLQAAAVGLPDSELGERICLFVVL
 ADGTRVELADVVRKVMENAETASFKLPERLITLPSLPTTPMGKIDKKALRAAAGRMSET

PhI2 from *Penicillium chrysogenum*: (Uniprot B6HL15) *FEBS Lett.*, 2011, (585) 893-8

MGSSHHHHHHSSGLVPRGSHMPTIYRSPYDLDIQSVDLVSYLFSNPFNTPLDRPMYINAISGEQY
 TFGDVVQRTRSLSNGLRQSIGLKPNDVVALFSPNTIDYPVCHAIVGSRAIVAPTSAALTALELNAQL

KTSGARFIVVHSTLLETAQKAAKGTSVEKVLLIDGQTPVNGQPTCNYLANTFAPDDLTVDPAEADR
 QPTFICFSSGTSAAKGIVITTHQNITSNLQQWRQHMLSEGLPSQRPRRQSAIAFLPFISHIYGLNLFM
 CQCLIWGTTVVVMFRDLDLYLSCIQKYRPDELALVPPIALMLVKDPRVSKYDLSSVRKIMSAGAPL
 TIELSSALEAKFTEICKTEVFCTQSWGLTETSPMATAVPNDRMDKRNTGVGCIAPNMQLRFVDPES
 MKDAAVKPDGSTEPAEIWCGRPNVVMGYNNNEKATKEAFHVEDGTRWFRGTGIDGTIDGDGYVTI
 QDRIKEMIKYKGLQVIPSELEGKLVDPVEDAAVTGMWVDDMATELPVGFVVLSPQAKDRDQKA
 VLDGIHAWLNERIANHKRLRGGIHVLSQIPKSPSGKILRRQLRDLLKSQAPKARL

stIB from *Photorhabdus luminescens*: (Uniprot Q7N526) *Angew. Chem. Int. Ed.* 2008, 47, 1942-1945

MGSSHHHHHHSSGLVPRGSHMEKVWLKHYPADVPAEIDPDYASLVEMFENAVAHYADQPAFIN
 MGEIMTFRKLEERSRAFAAYLQNGLGSLKGDVALMMPNLLQYPVALFGVLRAGMIVNVNPLYTP
 RELEHQLNDSGTSIAIVIVSNFAHTLEKIVFNTKVKHVILTRMGDQLSRPKGTLVDFAVKYIKRLVPKY
 NLPDAISFRCAIQKGYRMQYVKPEINGNDLAFLQYTGGTTGVAKGAMLTNRNLANLEQAKAVYSPL
 LRVGQELIVTALPLYHIFALMVNCLLLIYLGCCNLLITNPRDITGTAKELGRYPFTSVTGVNTLFWNL
 NNEEFKKLDFSTLRLVVGGMPPVQKAVAEKWAKVTGTNLLEGYGLTECSPLVSCNPYNSKHVTGS
 IGFPVSSTEIKLVDDDGNEVEMGQQGELWIRGPQVMAGYWNRPDATEEVLKDGWVATGDIANVNE
 QGSIHIVDRKKDMILVSGFNVYPNEVEDVSAHPKVLESAAIGVSSSESSGETVKVFVVRIDPGLTEDE
 LKTHCRRYLTGYKVPKIIEFDELPSNVGKILRREL RDEEEKVRNVA

RpCL from *Rhodopseudomonas palustris*: (Uniprot O07454) Kuver et al., *Arch. Microbiol.* 1995, 164, 337-345

MGSSHHHHHHSSGLVPRGSHMEFDAVLLPPRRAASIAAGLWHDRTINDDLDACVAHCPDKVALTA
 VRLDGGAVRRFSYRELATLADRVAVGLNRLGVGRGDVAMQLPNWWQFTVLYLACSRIGAVLNPL
 MPIFRERELSFMLKHGDAKVLVVPKSFGRGFDHEAMARSLQPDLPALRTIVVDGGGADDFDTLLTT
 PEWEKQPDAAAILQGSRPSPDDITQLIYTS GTTGEPKGMHSAANTLMANIVPYAQRALALRESVILM
 ASPMAHQTGFMYGMLMMPIMLRASAVLQDIWEPTKAAELIRTERVTFTMASTPFLTDLTRVVKESGE
 PVPSLKTFLCAGAPIPGPLVEQAQAGLGAKIVSAWGMTENGAVTLIKLDDDDKLASTTDGCPLPGVE
 VKVIDGDGKTLPPNQIGRLVVRSCSNFGGYLKRPHWNGTDADGWFDTGDLAYMTADGYIRISGRS
 KDVIIRGGENIPVVEVEALLYKHPAVAQVAIVAYPDERLGERACAVVVPKTGASIDFAAMVEFLKAQK
 LALQYIPERLVVRDAMPATPSGKIQKFRLEMLQHNDL

LcCL from *Leptothrix cholodnii*: (Uniprot B1Y4L8)

MGSSHHHHHHSSGLVPRGSHMNFDTVLLPPRRAASVAAGHWFDRTINDDLDACVAACPDKVALT
 AVQVESGEVRRFTYRELAAMADRVAVGLSRLGVGRNDVAMQLPNGWQFTVVYLACSRIGAVVN
 PLMHIFRERELTFMLGHGEAKVLIVPKTFRGFDHERMVDITRPDLPKLQQVVVVGGSGANSFEALLC
 GPAWENEPDAHEVLTRSRPGPDDVTQLIYTS GTTGEPKGMHTANTVMANIIPYAERLHLGSDDV
 LMASPMHQTGFMYGMLMMPIMLRASAVLQDLWDARRAVELIRSEGATFTMASTPFLSDLAKTVAET
 GTSVPTLRTFLCAGAPIPGALVEQARKVLGKIVSAWGMTENGAVTLIKLDDDDQRAFTTDGCPLP
 GVELKVVDADGAELPAGQAGKLLVRAASNFGGYLHRPQWNGTDADGWFDTGDLARIDAQGYIRIS
 GRSKDVIIRGGENIPVVEVEALLYRHPAVAQVAIVAYPDERLGERACAFITTKPGQSLDFAGMVEFLK
 AQKLAIQYIPERLVVRDALPSTPSGKLQKFKLREMLVRDGS

ArCL from *Acinetobacter radioresistens*: (Uniprot C6RPQ0)

MGSSHHHHHHSSGLVPRGSHMDFDAVLIPARKEAMLKQGYWLNQTILDFLRSVAEKNPDKTALVS
 VKVENQTEQTFSYQQLWNMTNKIALGLKQLGIEKNDVVSCQLPNWWEFTLLYLACSRIGAVLNPLM
 PIFRERELEFMLKRGESKVFVVPKTFRNFNHEQLANQLQNKLDLKHVVVVNGEGENNFDHLLLNH
 SLEQDASAVAELDNMESGPDDITQLIFTSGTTGEPKGMHTANTLFSNIVPYAERLHLTENDVVLMA

SPMAHQGTGFMVGLMMPIQLNTKVVLQDVWDVAKAVDLIHQHQVNFTMASTPFLNDLSNTVAEQHD
 KVDSLKIFLCAGAPIPGPLVQKARETLGVKVISAWGMTECGAVTLTRPEDKDESFNTDGIALPGVEI
 KIVDKKGQSKTVNEAGRLMIRSCSSFGGYLKRPDLNDTNVEGWFDTGDIAYQDEQGYIRICGRKKD
 VIIRGGENIPVAEIESLLYKHPNIAVVALVAYADERLGERACAIKLDPAQLLSLNLVDLFLKTHNLAIQ
 YIPERLEIWEEIPMTPSGKIQKFKLRELLQ

MsCL from *Mycobacterium smegmatis*: (Uniprot A0QZQ6)

MGSSHHHHHHSSGLVPRGSHMTAPTERRLANVLNGQYTPIDDDTAANWRAAGWWENRSIRSLLA
 DAAQAHPDRIALVGRRADGGRVARTYQEFDRNANQVASVLASLGVRPDDAVVVMLPNWVEYPEF
 LFGINELGAIYAGIPVAYGDQQAAILRRSRARVLVIPRRWRGNNILEQSRRLRDQIPTLQQVIVLDD
 DGTDLRDGESLWSDHAHVAARQFPPDPGQICYLGFSTGTTGEPKGAMHSHNTLIYSARRQAEHI
 GTEAFGEPPVNLVASPMGHHTGYVWGGVFTVMLAGTAVHVDWRDPTWGAQVVREEGVTTFFGA
 PTFLLQDIIRTELADGPACPLRCMVVAGAPVPRNLPAQAAEALGAYVAPAWGMTECSILTSCTPDEP
 DAILRTDGSVFAGSEVKIVDDTGAAVAAGVVGDLMLRGPVGVYGYDRPDATRDAYLPGLWFKTG
 DRADVDENGWLRRLRGRSKDIIIRGGENIPVTDVESAIKFDHPDVLNAAVIGLPDERLGERVCAVLVTKS
 GCPELTVDTLGEYLLGQGLSKHYLPEKVVHLDELPMTPSGKIQKFKLREQYS

1.2 N-Acyltransferases

A panel of phylogenetically diverse N-acyltransferase were identified by reviewing the proteins in the ExPASy database (EC 2.3.1). These genes were synthesized at Genscript and the DNA sequences were codon optimised for *E. coli* expression. The genes were cloned into pCDF-Duet vectors using BamHI and HindIII restriction sites so that the expressed protein contains an N-terminal 6-His tag. HCBT1 did not express solubly and was expressed as an MBP-fusion (maltose binding protein fusion).

HolE from *Pseudoalteromonas sp.*: (Uniprot F8J3H2) *Angew. Chem. Int. Ed.* 2015, (54), 1-6

MGSSHHHHHHHSQDPMSEKLDSEKLMQEHQTWTSKPASLEEWQIVNEWAIAEKWDLGLGDTERRFF
 NIDEEGFYLYGVNDEPVASVSVNYTDEYAYAGFYLVAPGARGKGYGLRLSYDAFRHCDKRSVGL
 DGMPEQEENYKKGFFVTHYETSRLVGIHNQQVDAPDGVQNTADNIDEVIKFDKITGYPRAALLKD
 WFSGEGRHGFVINSBGDGVIGVVGIRRDGYRLGPLYSENQAVCDKLFAMALAQVPQGTQVTIDA
 PTLDLGFINGLKKMGFEEIFHTFRMYRGKEPQGEKHKIQAIASLELG

CASHT from *Capsicum annuum*: (Uniprot Q9ATJ3) *Plant Physiology*, 2004 (135) 346-356

MGSSHHHHHHHSQDPMASAPQPPTLSEKTTNLSPENDNVTITGKIYTRVRLATKSDLHHAYQLFYQI
 HAYHNQFHLFKATESSLSDLFFKENPLPLFYGPTLLLLLEVSTAFTEPKNNKDEGFKPVFTALDLKFP
 VVEGQVEEFRSKYDDGTDKRDVFIAGYAYFFASYSLFGNDKPGIHFDSLYFRESYRKLGMGKLLFG
 TVASIAANNNGFAAVEGIVAVWNKKSDFYVSMGVEMHDDFRFGKLDGENLQKYADKEKNGAGSC

HCBT1-MBP from *Dianthus caryophyllus*: (Uniprot O24645) *Plant Mol. Biol.*, 1997, (35), 777-789

MGHHHHHHHGMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGD
 GPDIIFWAHDFRGGYAQSGLLAEITPDKAFQDKLYPFTWDVAVRYNGKLIAYPIAVEALSILYNKDLLP
 NPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGYDIKDVGVNDAGAK
 AGLTFLVDLIKHKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNNGVTVLPTFGGQPSK
 PFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYYEEELAKDPRIAATMEN
 AQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTGDPMSIQIKQSTMVPRPAEETPNK
 SLWLSNIDMILRTPYSHTGAVLIYKQPDNNEDNIHPSSSMYFDANILIEALSALVPFYPMAGRLKING
 DRYEIDCNAEGALFVEAESHVLEDFGDFRPNDLHRVMVPTCDYSGISSFPLLMVQLTRFRCGG
 VSIQFAQHSHVCDGMAHFEFNNSWARIAKGLLPALPEVHVDRLHLRPRNPPQIKYSHSQFEPFVPS

LPNELLDGKTNKSQTLFILSREQINTLKQKLDLSNNTTRLSTYEVVAAHVWRSVSKARGLSDHHEEIKL
 IMPVDGRSRINNPSPKGYCGNVVFLAVCTATVGDLSNPLTDTAGKVQEALKGLDDDYLRSAIDH
 TESKPGLPVPMGSPEKTLYPNVLVNSWGRIPYQAMDFGWGSPTFFGISNIFYDGQCFLIPSRDGD
 GSMTLAINLFSSHLSRFFKKFYDF

BcAT from *Bacillus cereus*: (Uniprot Q81AS3) *J. Bio. Chem.* 2013, (288), 22493-22505

MGSSHHHHHHSQDPMTDFQKQFFARLHIEEKDTVSEFEDLSNIMYAMAQTVPFENLNILEKNFKEISK
 ENLKEKILVNNRGGGLCYELNPTMYFLKDSGFDVHLVSGTVYNAANSIWAVDSGHIATVLTHHNELY
 LIEVGFGSYLPLAPVPFLGEVIHSATGDYRIRKEMTEKGNYLEMRKNNFLDQSAADDWTLGYAFYI
 EEVDEEKANTAQKIIVEHEGSPFNKVPLIVKLTEDGHASLTKDSLTVAKNGKKTETVTDMQYTNLL
 HSKFGITL

HpA3 from *Saccharomyces cerevisiae*: (Uniprot P39979) *Arch. Microbiol.*, 2004, (182), 396-403

MGSSHHHHHHSQDPMKKTDPSPFFASTKNVGMSNEEPEKMNVDRIVVKAIPEKDDEAWNKLWK
 EYQGFQKTVMPPEVATTTTFARFIDPTVKLWGALAFDTETGDAIGFAHYLNHLTSHHVEEVVYMNDL
 YVTERARVKGVGRLIEFVYSRADELGTPAVYWVTDHYNHRAQLLYTKVAYKTDKVLYKRNGY

01StAT : (Uniprot Q00267)

MGSSHHHHHHSQDPMTSFLHAYFTRLHCQPLGVPTVEALRTLHLAHNCAIPFENLDVLLPREIQLD
 ETALEEKLLYARRGGYCFELNGLFERALRDIGFNVRSLGRVILSHPASLPPTHRLLLVDVEDEQWI
 ADVGFGGQTLTAPRLQAEIAQQTPHGEYRLMQEGSTWILQFRHHEHWQSMYCFDLGVQQQSDH
 VMGNFWSAHPQSHFRHLLMCRHLPDGGKLTLTNFHFTRYHQGHAVEQVNVDPVPSLYQLLQ
 QQFGLGVNDVKHGFTEAELAAVMAAFDTHPEAGK

02MsAT : (Uniprot O86309)

MGSSHHHHHHSQDPMAMD LGGYLTRIGLDGRPRPD LGTLHAIVA AHNRSIPFENLDPLL GIPVADL
 SAEALFAKLVD RRRGGYCYEHNGLLGYVLEELGFEVERLSGRVWWMRADDAPLPAQTHNVLSVAV
 PGADGRYLVDVGFGGQTLTSPIRLEAGPVQQTRHEPYRLTRHGDDHTLAAQVRGEWQPLYTFTTE
 PRPRIDLEVGSWYVSTHPGSHFVTGLTVAVVTDDARYNLRGRNLAVHRSGATEHIRFDSAAQVLDA
 IVNRFGIDLGLAGRDVQARVAEVLDT

05PsAT : (Uniprot Q9HUY3)

MGSSHHHHHHSQDPMTPLTPEQTHAYLHHIGIDDPGPPSLANLDR LIDAH LRRVAFENLDVLLDRPI
 EIDADKVF AKVVEGSRGGYCFELNSLFARLLLALGYELELLVARVRWGLPDDAPLTQQSHLMLRLYL
 AEGEFLVDVGFGSANPPRALPLPGDEADAGQVHCVRLVDPHAGLYESA VRGRSGWLPLYRFDLR
 PQLWIDYIPRNWYTSTHPSVFRQGLKAAITEGDLRLTLADGLFGQRAGNGETLQRQLRDVEELLDI
 LQTRFRLRLDPASEVPALARRLAGLISA

07NfAT : (Uniprot Q5YYQ3)

MGSSHHHHHHHSQDPMKSPDDPAYHWNGAELDLDAYLARIGFAGERAPTLATLRELVRHTTAIPF
 ENLEAVLGRPVRDLATLQDKLVHSRRGGYCYENAGLFAAALERLGFGVTGHTGRVTMGAGGLRP
 ATHALLRVTTADDDRVMCDVGFGRGPLRPYELRPQPDEFTLGDWRFRLERRTGELGTDLWVLH
 QFGRDGVVDRYTFTTAPQYRIDFEVGNHFVSTSPRSPFTTRPFLQRFHSDRHHVLDGLTLITERPD
 GSADIRALTPGELPEVINELFDIELPGPDLDALTGSWLERSVAAGTP

10HvAT : (Uniprot A9ZPJ7)

MGSSHHHHHHHSQDPMKITVHSSKAVKPEYGACGVAPGCTADVPLTVLDKANFDITYISVIYAFHPP
 APPNAVLEAGLGRALVDYREWAGRLGVDANGDRAILLNDAGARFVEATADVALDSVMPLKPTSEV
 LSLHPSGDDGPEELMLIQVTRFACGSLVVGFTAQHLVSDGRATSNFFLAWSQATRGVAVDPPVPH
 DRASFFHPREPLHVEYEHRGVEFKPYEKAHDVVCGADGDEDEVVNKVHFSREFISKLKAQASAG
 APRPCSTLQCVVAHLWRSMTMARGLDGGETTSVAIAVDGRARMSPQVPDGYTGNVILWARPTTT
 AGELVDRPVKHAVELISREVARINDGYFKSFIDFANS GAVEKERLVATADAADMVLSPNIEVDSWLRI
 PFYDMDFGGGRPFFFMPSYLPVEGLLILLPSFLGDGSVDAYVPLFSRDMNTFKNCCYSLD

11AtAT : (Uniprot Q9FNP9)

MGSSHHHHHHHSQDPMALKVIKISRVSPATASVDPLIVPLSFFDLQWLKLNPTQVFFYKLTSSSSSR
 DVFYSSILPKLERSLSLILTHFRLFTGHLKWDSDQDPKPHLVLSGDTLSLTVAEADFSRISGRGLR
 PELELRPLIPELPIYSDSGAVVSLQVTLFPKQGFICIGTTAHHVVDGKTAEFKNKAWAHTCKHGTPK
 ILPTVLD RSVVNPAGLEQKMLELLPYLTEDDKENGRTLKLPVKEINAKDNVLRITIEISPENIEKLKE
 RAKKESTRAELHLSTFVVTFAHVWTCMVKARSGDPNRPVRFMYAADFRNRLEPPVPPTYFGTCVL
 AMDFYKYKAKEFMGEDGFVNTVEILSDSVKRLASQGVESTWKVYEEGTKTMKWGTQLLVNGSN
 QIGMYETDFGWGRPIHTETMSIYKNDEFSMSKRRDGIGGVEIGISLKKLEMDTFLSLFYKWIGN

12HvAT : (Uniprot A9ZPJ6)

MGSSHHHHHHHSQDPMKITVHSSKAVKPEYGACGLAPGCTADVPLTVLDKANFDITYISVIYAFHAP
 APPNAVLEAGLGRALVDYREWAGRLGVDASGGRAILLNDAGARFVEATADVALDSVMPLKPTSEV
 LSLHPSGDDGPEELMLIQVTRFACGSLVVGFTTQHIVSDGRSTGNFFVAWSQATRGAAIDPVPVHD
 RASFFHPREPLHVEYEHRGVEFKPCEKAHDVVCGADGDEDEVVNKVHFSREFISKLKAHASAGA
 PRPCSTLQCVVAHLWRSMTMARGLDGGETTSVAIAVDGRARMSPQVPDGYTGNVILWARPTTTA
 GELVTRPVKHAVELISREVARINDGYFKSFIDFANS GAVEKERLVATADAADMVLSPNIEVDSWLRI
 FYDMDFGGGRPFFFMPSYLPVEGLLILLPSFLGDGSVDAYVPLFSRDMNTFKNCCYSLD

14AtAT : (Uniprot O80467)

MGSSHHHHHHHSQDPMPIHIGSSIPLMVEKMLTEMVKPSKHIPQQTLNLSTLDNDPYNEVIYKACYVF
 KAKNVADDDNRPEALLREALSDLLGYYPPLSGSLKRQESDRKLQLSCGGDGGVPFTVATANVEL
 SSLKNLENIDSDTALNFLPVLHVDIDGYRPFALQVTKFECGGFILGMAMSHAMCDGYGEGHIMCALT
 DLAGGKKKPMVTPIWERERLVGKPEDDQPPFVPGDDTAASPYLPTDDWVTEKITIRADSIRRLKEAT
 LKEYDFSNETITTFEIVIGAYLWKS RVKALNLD RDGVTVLGLSVGIRNVDPPLPDGYYGNAYIDMYV
 PLTAREVEEFTISDIVKLIKEAKRNAHDKDYLQEELANTEKIIKMNLTIKGGKDGFLCLTDWRNIGIFGS
 MDFGWDEPVNIVPVVPSSETARTVNMFM RPSRLES DMVGGVQIVVTLPRIAMVKFKEEMEAL

18BsAT : (Uniprot P21340)

MGSSHHHHHHHSQDPM SVKMKKCSREDLQTLQQLSIETFNDTFKEQNSPENMKAYLES AFNTEQLE
 KELSNMSSQFFFIYFDHEIAGYVKVNIDDAQSEEMGAESLEIERIYIKNSFQKHGLGKHLNKAIEIAL
 ERNKKNIWLGVWEKNENAI AFYKKMGFVQTGAHSFYMGDEEQTDLIMAKTLI

19CpAT : (Uniprot Q5CPU3)

MGSSHHHHHHHSQDPMISSFEVRKATIDDYFELRNLICDVTRCTETLSREQAEERFRYNTYHPYCLV
DTENGRIVGYAGFYIIPHLGRKNDSRIEHVIISKEYRNRGLGRLLCKQIIEDAKNKFNCGRIDLTVESHI
AKKLYSSLEFEKVNTTEVMRNSFLDLTPKSD

23DmAT : (Uniprot Q94521)

MGSSHHHHHHHSQDPMEVQKLDPDQSLISSMMLDSRCGLNDLYPIARLTQKMEDALTVSGKPAACPV
DQDCPYTIELIQPEDGEAVIAMLKTFFFKDEPLNTFLDLGECKELEKYSKPLPDNCSYKAVNKKGEII
GVFLNGLMRPSPDDVPEKAADSCEHPKFKKILSLMDHVEEQFNIFDVYPDEELILDGKILSVDTNY
RGLGIAGRLTERAYEYMRENGINVYHVLCSHSYARVMEKLGFEVFRMQFADYKPQGEVVFKA
APHVGIQVMAKEVGPAAQAQTKL

24LIAT : (Uniprot G6FA50)

MGSSHHHHHHHSQDPMNEITYTQIMPNDIDDVIRIERATFSENEALSVESMIERINLIPDSFIAARNSEG
TVVGIVSGPVTEGRYLDDESFERTEANPKTGGFQKIISLTVDPDYQGLGIATNLLLLLEKEAKSKKRL
GISLTCHDYLVPYYEKHGFTNEGLSESKFGGETWYNMVMFEF

28BaAT : (Uniprot Q81MQ2)

MGSSHHHHHHHSQDPMKMMMDANEIISFIQKSEKKTVPKVYIKGDLKEVTFPETVQAFVNKKSGVLFG
EWSEIKTILDENSKYIVDYVVENDRRNSAIPMLDLKGIKARIEPGAIRDHVEIGDNAVIMMNATINIGA
VIGEGSMIDMNAVLGGRATVGKNCHVGAGAVLAGVIEPPSAKPVIVEDDVIGANVVVLEGVTVGK
GAVVAAGAVVTEDEVPPYTVVAGTPARVIKEIDEKTKAKTEIKQELRQLNPEK

29EfAT : (Uniprot Q836H8)

MGSSHHHHHHHSQDPMDAYEIIQYIGDAKKQTLVKVTLKGQLKEVTFPETIKVFNNCKTGTLFGDWA
DVKPFLEANKEKIEDYVVENDARNISAIPFLDLKDINARIEPGALIREKVEIGDQAVIMMGAILNIGAVVG
AGTMIDMGAVLGGRATVGKHCHIGAGTVLAGVIEPPSAAPVVIENEVVIGANAVVLEGVRVGEAV
VAAGAVVVEDVPAHTVVAGVPAKVIKQIDDKTKSKTEILEELRKL

30SaAT : (Uniprot Q7A2S0)

MGSSHHHHHHHSQDPMVQHLLTAEIIQYISDAKKSTPIKVYLNNGNFEGITYPESFKVFGSEQSKVIFCE
ADDWKPFYEAYGSQFEDIEIEMDRRNSAIPKDLTNTNARIEPGAIFIREQAIIEDGAVMMGATINIGA
VVGEGTMIDMNATLGGRATTGKNVHVGAGAVLAGVIEPPSASPVIIEDDLVIGANAVILEGVRVGKG
AIVAAGAIVTQDVPAGAVVAGTPAKVIKQASEVQDTKKEIVAALRKLND

31NtAT : (Uniprot P80969)

MGSSHHHHHHHSQDPMATTNNKNLTITEKVYVRVRLANEADISHIYKLFYQIHEYHNYTHLYKATESS
LCDLLFKANPNPLFYGPSVLLLEVSPPTFENTKKDEKFKPVLKTFDLRATVEDKEAEEFKSKSCGDE
KEDVFIAGYAFFYANYSCFYDKAGIYFESLYFRESYRKLGMGGLLFGTVASIAANNGFASVEGIVAV
WNKKSDFYVNMGVEIFDEFYRGLVGDALQKYADKEKV

32NtAT : (Uniprot Q9SMB8)

MGSSHHHHHHHSQDPMATTNNKNLTITEKVYVRVRLANEADISHIYKLFYQIHEYHNYTHLYKATESS
LCDLLFKANPNPLFYGPSVLLLEVSPPTFENTKKDEKFKPVLKTFDLRATVEDKEAEEFKSKSCGDE
KEDVFIAGYAFFYANYSCFYDKAGIYFESLYFRESYRKLGMGSLLFGTVASIAANNGFASVEGIVAV
WNKKSDFYVNMGVEIFDEFYRGLVGDALQKYADKEKA

33MtAT : (Uniprot P9WP21)

MGSSHHHHHHHSQDPMSTVTGAAGIGLATLAADGSVLDTWFPAPELTESGTSATSRLAVSDVPVEL
 AALIGRDDDRRTETIAVRTVIGSLDDVAADPYDAYLRLHLLSHRLVAPHGLNAGGLFGVLTNVVWTN
 HGPCAIIDGFEAVRARLRRRGPTVYGVDFPRMVDYVVP TGVRADADRVRLGAHLAPGTTVMHE
 GFVNYNAGTLGASMVEGRISAGVVVGDSVGGGASIMGTLSSGGGTHVISIGKRCLLGANSGLGIS
 LGDDCVVEAGLYVTAGTRVTMPDSNSVKARELSGSSNLLFRNSVSGAVEVLARDGQGIALNEDL
 HAN

34PaAT : (Uniprot G3XD76)

MGSSHHHHHHHSQDPMSSQLFSLAFGVGTQNRQEAWLEVFYALPLLKPSSEIVA AVAPILGYAAGN
 QALTFTSQQAYQLADALKGIDAAQSALLSRLAESQKPLVATLLAEDAAPSSTAEAYLKLHLLSHRLVK
 PHAVNLSGIFPLLPNVAWTNIGAVDLAELAEQLLEARLKGKLLEVFSVDKFPKMTDYVVPAGVRIADT
 ARVRLGAYIGEGTTVMHEGFVNFNAGTEGPGMIEGRVSAGVFVGKGS DLGGGCSTMGTLSSGGN
 IVISVGEGLIGANAGIGIPLGDRNIVEAGLYITAGTKVALLDEQNALVKVVKARDLAGQPDLLFRNS
 QNGAVECKTNKTAIELNEALHAHN

35MaAT : (Uniprot Q4JQJ5)

MGSSHHHHHHHSQDPMPLDPKTA LPIITLSQPTAEVGAQVHRLISKCPPLDPNSMYCNLLQSSHSET
 AVAAKIGDELVG FVSGYRIPQRPDTLFWVQVAVGEKARGQGLATRMLKAILARPVNQDINRIETTITP
 NNKASWALFEGLAKKLD TQIGSAVMFDKTRHFADQHETEMLVKVGPFKAVQA

36PaAT : (Uniprot A6VCX3)

MGSSHHHHHHHSQDPMSSASIRDAGVADLP GILAIYNDAVGNTTAIWNETPVDLANRQAWFDARARQ
 GYPILVASDAAGEVLGYASYGDWRPFEGFRGTVEHSVYVRDDQRGKGLGVQLLQALIERARAQGL
 HVMVAAIESGNAASIGLHRR LGFEISGQMPQVGQKFGRWLDLTFMQLNLDPTRSAP

37HsAT : (Uniprot P30419)

MGSSHHHHHHHSQDPMAD ESETAVKPPAPPLPQMMEGNGNGHEHCSDCENEEDNSYNRGG LSP
 ANDTGAKKKKKKKQKKKKEK GSETDSAQDQPVKMNSLPAERIQEIQKAIELFSVGQGP AKTMEEASK
 RSYQFWD TQVPKLGEVVNTHGPVEPDKNIRQEPYTLPQGFTWDALDLGDRGV LKELYTLLNEN
 YVEDDDNMFRFDYSPEFLLWALRPPGWLPQWHCGVRVSSRKL VGFISAIPANIHIYDTEKKMVEI
 NFLCVHKKLR SKRVAPVLIREITRRVHLEGIFQAVYTAGVVL PKPVGTCRYWHRSLNPRKLIEVKFSH
 LSRNMTMQRTMKLYRLPETPKTAGLRPMETKDIPV VHQLLTRYLKQFHLTPVMSQEEVEHWFYPQ
 ENIIDTFVVENANGEVTD FL SFYTL PSTIMNHPTHKSLKAAYSFYNVHTQTPLLDLMSDALVLAKMKG
 FDFVFNALDLMENKTFLEKLKFGIGDGNLQYYLYNWKCP SMGAEKVGLVLQ

38ShAT : (Uniprot P16426)

MGSSHHHHHHHSQDPMSPERRPADIRRATEADMPAVCTIVNHYIETSTVNFRTEPQEPQEWTDDL V
 RLRERYPWLVAEVDGEVAGIAYAGPWKARNAYDWTAE STVYVSPRHQRTGLGSTLYTHLLKSLEA
 QGFKSVVAVIGLPNDPSVRMHEALGYAPRGMLRAAGFKHGNWHDVGFVWQLDFSLPVP RPVLPV
 TEI

39SeAT : (Uniprot Q9R381)

MGSSHHHHHHHSQDPMDIRQMKNKTHLEHWRGLRKQLWPGHPDDAHLADGEEILQADHLASFIAMA
 DGVAIGFADASIRHDYVNGCDSSPVVFLEGIFVLP SFRQRGVAKQLIAAVQRWGTNKGCREMASDT
 SPENTISQKVHQALGFEETERVIFYRKRC

41GmAT : (Uniprot D8FSU9)

MGSSHHHHHHHSQDPMADRIRYSKSQLEKYYDRIAFPASDRRYDISNLSSSEDQRSYLDTLTKQQILT
VPFENLTLHYSWHRTVDVNADHLYDKIVNERRGGYCMENNTFFNTVLLSLGYHTYMGVSRVFNPD
ADRFGGTSHCLSLVIIDGKTLAVDVGFGRNPTEPLEVEHERVHTGSSGFQMRLRYDAIAQNVSNQ
KLWIYEYRSRDGAEWVPQWCFMDFEVLPEDIRVFNLSPSKSPSSFFTFKVVSVQFTSEKEDYSDG
SARDLNNVGGDVGAFVIDGELFKYRKGGETKWERTFKSEDERLAALRKYYGVEVTKENERAIGG
TAGAISYRRTGS

42GmAT : (Uniprot B7SP66)

MGSSHHHHHHHSQDPMARLEDPTALTQLPDESARVRYTSSELQDYFETLKFPQRFLDLGNSVLKDP
SLARTKENGLPLLQAITRYHTCNVPFENLVLHYDPHKIVTLDPALYTKIVTRRRGGRCMENNIFLGT
ALRSLGYEVRNCGGRVSRAMSPYPEVRKNQSATYDGNHMLLLVFLGDEWYGVDVGMGSMGP
NLPFPLQDGFESLSIAPREIRIQKRSISETHATGPSHATKMWCYDVCYNPAESKKTWTPVYCFTETE
FLPQDYEVMSWFTSTNPRSFFTRYITCTKMIMDEDKEVIIGNLTLFKDTVRETIGSDRKVVKKFETEE
ERIKGLVEIFDVNLTEEEKNSLPQEKRLA

43GmAT : (Uniprot D8FSU8)

MGSSHHHHHHHSQDPMASAIYSEAQVAGFLKHLQIPQEFYVGNEAILDHAFLKVLHQHMIATVPYDNL
TLHYSSHRNITLEPQALYQKIVGDGRGRGGYCMESNLFFCYMLRALGFQVYPVGVVRVLRNNGIPF
GGYPGWVHIVNIVTLDPNSRWVMDASFGGDGPTQPMPLVEGAEWHNMGTQTARLIKDFIPGQTEL
TSGRRLWIYQCRNSPDLPWTSFYAFSHSVEWLPADFEITNCYTGTSPRSFQTSTVLIVKFLRESKT
SPTGEEIYGKRMLINDVVKENPGGKTKVLKELKTEDERVEALKEYFDIDLTTEEREAIKGFQTEIKSE

45SpAT : (Uniprot Q09927)

MGSSHHHHHHHSQDPMKDPNTIPPWRCTDFNAWCIAVDKSTNVKNKEELLSTLTIFYNIEIEMGQTY
PIDIKMTRNEAEDFFFKFCTVICVPVESETSPAPDLATASIDWKTSLLGAFYIKPNYPGRCSHICNGG
FLVSPSHRSKGIGRNLNAYLYFAPRIGFKSSVFNLVFATNIKSIRLWERLNFTRAGIIKDAGRLKGHE
GYVDAYIYQYHFPSLEDALK

47AtAT : (Uniprot Q9ZV05)

MGSSHHHHHHHSQDPMAPPTAAPEPNTVPETSPTGHRMFSRIRLATPTDVPFIHKLHQMAVFERLT
HLFVATESGLASTLFNSRPFQAVTVFLLLEISPSPFPTTHDASSPDFTFPLETHKVDLPIDPDREKFL
PDKLNDVVVAGFVLFFPNYPSFLAKQGFYIEDIFMREPYRRKGFGKLLLTAVAKQAVKLGVRVEWI
VIDWNVNAINFYEQMGAQVFKEWRLCRLTGDALQAIDKLNI

48CaAT : (Uniprot Q5D8C0)

MGSSHHHHHHHSQDPMASAISETITTNGPSENNNLTTITGKIHTRVRLATKSDLHHIYQLFYQIHAYHNF
THLYKATESSLGDLLEFKENPLPLFYGPSVLLLEVSPTPFTQPKNNKDEGFKPVLTTFNLKFPVVEGQ
VEEFQSKYDDGNDKRDVFIAGYAFFYANYSCFYDKPGFYFESLYFRESYRKLGMGRLLFGTVASIA
ANNGFVSVEGIVAVWNKKSDFYIDMGVEIFDEFYRKGKLHGENLQKYADKQKNEGGNC

49StAT : (Uniprot Q8H9D9)

MGSSHHHHHHHSQDPMAPAPQQPTPSETIITDASSENNNVTITGKIYTRVRLATKSDLSHIYQLFYQIH
VYHNFTHLYKATESSLGDLLEFKENPLPLFYGPSVLLLEVSPTPFNEPKNTTDEGFNPVLTTFDLKFPV
VEGQVEEFRSKYDDKSDAYIAGYAFFYANYSCFNDKPGFYFESLYFRESYRKLGMGKLLFGTVSSI
AADNGFVSVDGIVAVWNKKSDFYINMGVEIFDEFYRKGKLHGENLQKYADKKGIEEETC

51EcAT : (Uniprot P76112)

MGSSHHHHHHHSQDPMISIRFARKADCAAIAEIYNHAVLYTAAIWNQDQTVADNRIAWFEARTLAGYP
VLVSEENGVVVTGYASFGDWRSFDGFRHTVEHSVYVHPDHQKGKGLGRKLLSRLIDEARDCGKHVM
VAGIESQNNQASLHLHQSLSGFVVTAAQMPQVGTGKFRWLDTLTMQLQLDERTEPDAG

52PIAT : (Uniprot V4HKD0)

MGSSHHHHHHHSQDPMSEKLDYSYKLTQDHQSWTSKPATLEEWQIVNEWALAEKWDLGLGDTECF
NIDDQGGFYLGYNVEPVASVSVNYSDEYAYAGFYLVAPGARGKGYGLRLSYDAFHHC DKRSVGL
DGMPEQEENYKKGFFVTHYETSRVLGVHNQQVDAPQGVQNISADNIEAVIEFDAQITGYSRAALLQ
NWFSGEGRHGLIDSGDGLGVVGIRRDGRLGPLYAENQAVSEKLFAMAMAQVPLGTQVTID
APTLDLGFINTL KALGYEEIFHTFRMYRGTEPPQGEKHKIQAIASLELG

53PhAT : (Uniprot A0A0F5VBE7)

MGSSHHHHHHHSQDPMDIVKNKTETETSNSVWRVESANLAEWQAVIAWAKDEGWDMGIGDANAF
EVDDQGGFFIGYLGKEPVAAMSVVNYSSRFSFLGHYLVSPAYRGQGYGLKLCQSAFYHGGERC
LDGMPAQVDNYAKWGFAGDRYNLRMVGQVQQDYHCPDHIALVQAADIDALIAYDQSCGTIDRSAL
LKHWFIGETRYGFICRSGQEVTGIIGVRQSQEGYRVGPLFANSPDLVEPLFIAALSAVPQGGKVTLD
VPEKADRTLISLAETHGFESIFQTLRMYRGEPPEKEQEHKVVWCIAASLELG

54SIAT : (Uniprot Q8RXB6)

MGSSHHHHHHHSQDPMASLSETITTDASSENNTITGKIYTRLRLATKSDLSHIYQLFYQIHAYHN
NTHLYKATESSLANLLFKENPLPLFYGPSVLLLEVSPPTPFNEPTNEGFKPVLTTFDLKFPVVEGQVEE
FRSKYDDKSDVYIAGYAFFYVNYSCFSDKPGFYFESLYFRESYRKLGMSLLFGTVASIAANNGFV
SVEGIVAVWNKKSDFYVNMGVEIFDEFYRGKLHGENLQKYANDKEKNDGGN

55SIAT : (Uniprot Q8RXB8)

MGSSHHHHHHHSQDPMAPALEQAITS DASSDVTITGKIYTRVRLATKSDLSHIYRLFYQIHEYHNYTH
LYKATESSLANLLFKENPLPLFYGPSVLLLEVSPPTPFDEPKNTTDEGFKPVLTTFDLKFPVVEGEVEE
FRSKYDDKSDVYIAGYAFFYANYSCFYDKPGFYFESLYFRESYRKLGMSLLFGTVASIAANNGFV
SVEGIVAVWNKKSDFYVNMGVEIFDEFYRGKLHGENLQKYAHN

56SIAT : (Uniprot Q8RXB7)

MGSSHHHHHHHSQDPMAPTSQQPTPSPSLSDSLTTDASSDVTITGKIYTRVRLATKSDLSHIYKLFYQ
IHEYHNFTHLYKATESSLLEGLLFKENPLPLFYGPSVLLLEVSPPTPFNEPTNQAFKPVLTTFDLKFPVV
EGQVEEFRSKHDDKSDAYIAGYAFFYANYSCFNDKPGFYFESLYFRESYRKLGMSLLFGTVSSIA
ANNGFVSVDGIIVWNKKSDFYINMGVEIFDEFYRGKLHGENLQKYAHNKGKTEETC

58AxAT : (Uniprot E3HPZ9)

MGSSHHHHHHHSQDPMAEQPSITVTPERALIDVAREIRVSGFPAYAFITITASMTMCGAPWRSQAVF
VAGHDGSLDLARDCPVSGSYAEPSSAMGIVWSMACEDIASVFPDRTEPLAIHIEASDGRHTATAT
LVQEFQADGVTHRAVREQVGGMTLSGELYTPAGPGPHPLVIYMNGSSGGVNAPRAALFASRGYQ
CLALAIFFNYEGRPKYLNDMPLEYFQHVLWRWARAELAPRDGFVALSGISRGGETSLLVASHYPELVS
AVVAYVPSPVMHGVVSAGAPGTGRDAQVWTKAGEPLPHLWQDNASANWDAAYASQPPYRQTHA
FLSATRDSAAFERARIPVENYPGPVLLISASDDGFWPSTAYSEIVMRRRQEHGLPTFHVCMGAGH
HVHYPCLPATLISKPHAMSGLLLDAGGTPSANAAGNEGSYLAVLEFLGRVGGVAW

61SrAT : (Uniprot L8EWB7)

MGSSHHHHHHSQDPMTPSPGEIPRQPVPTAAAPSGPAPGTATATLTAGPDAPAGFTVRPATLDE
 WREVAAWAAAESWNPGRADVDCFHPTDPDGFFVGHLDGQLVSSVSVVNYSAAYAFLGYLVHPD
 FRGRGLGLATWRAAAHAGDRTIGLDAVPAQEATYRRSGFTTAYRTTRYRGRPVPRGTPAVPVVP
 VTADRLDEIADFDRQCFAERRAFLEWLTAAGHHAYAQIRDGRVAGYGVLRPAQDGYRLGPLFA
 DTSRGAEAVFDALTSHLGPEDVHIDIPDPHHSARSLALSRGLEPDSHTMRMYTGPVPPARAEHGY
 AVTSLELG

62SrAT : (Uniprot V6JJA0)

MGSSHHHHHHSQDPMTPAAPHDLVVTQATLADWPVIAGWAAAEGWNPGLSDGPAFFAQDADGF
 FLGRIDGEPVSAISVVNYGSDYAFLGCYLVRPDLRGHGHGLTTWKTALAHAGDRTVGLDGVAQQ
 DNYRQSGFELAYRTIRFTGAAPRAEVPVGVPRAGPADLAAITAYDSACYPADRPRFLAEWLTGPGH
 RAFVRHDGARLTGYGVLRPGRDTRLRIGPLFADTADDAHALFAALTTEAAGSEVAIDVPEPHAAGVAL
 AEEAGFQASFETARMYTGPVRAYAQHRVFGVTTLELG

63ScAT : (Uniprot Q08649)

MGSSHHHHHHSQDPMMSHDGKEEPGIAKKINSVDDIIKCQCWVQKNDEERLAEILSINTRKAPPKFY
 VHYVNYNKRLEDEWITTDRLNDKEVLYPKLKATDEDNKKQKKKKATNTSETPQDSLQDGVDFGSRE
 NTDVMDLDNLNVQGIKDENSHEDEIKKLRTSGSMTQNPHEVARVRNLNRIIMGKYEIEPWYFSPYPI
 ELTDEDFIYIDDFTLQYFGSKKQYERYRKKCTLRHPPGNEIYRDDYVSFFEIDGRKQRTWCRNLCLL
 SKLFLDHKTLYYDVPFLFYCMTRRDELGHHLVGYFSKEKESADGYNVACILTLQYQRMGYGKLLI
 EFSYELSKKENKVGSPKPLSDLGLLSYRAYWSDTLITLLVEHQKEITIDEISSMTSMTTDDILHTAKT
 LNILRYKYGQHIIFLNEDILDRLNRLKAKKRRTIDPNRLIWKPPVFTASQLRFAW

64ScAT : (Uniprot Q03330)

MGSSHHHHHHSQDPMVTKHQIEEDHLDGATTDPEVKRVKLENNVEEIQPEQAETNKQEGTDKENK
 GKFEKETERIGGSEVVTDEKGVKFEFDGVEYTFKERPSVVEENEGKIEFRVNNNDNTKENMMVL
 TGLKNIFQKQLPKMPKEYIARLVYDRSHLSMAVIRKPLTVVGGITYRPFDKREFAEIVFCAISSTEQVR
 GYGAHLMNHLKDYVRNTSNIKYFLTYADNYAIGYFKKQGFTKEITLDKSIWMGYIKDYEGGTLMQCS
 MLPRIRYLDAGKILLQEAALRRKIRTISKSHIVRPGLEQFKDLNNIKPIDPMTIPGLKEAGWTPEMDA
 LAQRPKRGPHDAAIQNITELQNHAAAWPFLQPVNKEEVPDYYDFIKEPMDLSTMEIKLESNKYQK
 MEDFIYDARLVFNCRMYNGENTSYYKYANRLEKFFNNKVKEIPEYSHLID

Example 2: General protocol for the production of CoA ligases and N-acyltransferases**2.1 General protocol for the expression of CoA ligases and N-acyltransferases**

For all pET28 constructs, kanamycin (50 µg/ml) was used in all LB agar plates and media

For pCDF-Duet constructs, streptomycin (50 µg/ml) was used in all LB agar plates and media

Chemically competent BL21(DE3) cells were transformed with plasmids carrying the constructs for CoA ligase or N-acyltransferase expression. Cells were plated onto LB agar and incubated at 37 °C overnight.

LB media (5-20 ml) was inoculated with a single colony and incubated at 37°C overnight (typically 18 hours) with shaking at 200 rpm. LB media (500 ml) containing 1 % glycerol and 0.05 % antifoam was inoculated with the overnight seed culture (5 ml). Cultures were grown in 2.5 L ultraflasks with AirOTop

seals at 37 °C, 200 rpm. When the cell density reached an OD₆₀₀ 0.6-1.0 the culture was cooled to room temperature and the protein expression was induced with the addition of IPTG to a final concentration of 0.3 mM. The cultures were incubated for a further 18-24 hours at 18 °C. The cells were harvested by centrifugation (20 mins, 4000 rpm, 4°C) and either purified directly or stored at -20 °C until required.

The general protocol was followed for proteins *phl*, *PhCL*, *CBL*, *ipfF*, *ScCL*, *phl2*, *stlB*, *RpCL*, *LcCL*, *ArCL*, *MsCL*, *HoE*, *CASHT*, *HCBT1-MBP*, *BcAT*, *HpA3* with the following exceptions:

stlB and *PhCL*: For expression of the genes *PhCL* and *stlB*, Rosetta2(DE3) cells were used and chloramphenicol (34 µg/ml) was added to the LB media in addition to kanamycin (50 µg/ml).

CASHT: Cells expressing the *CASHT* gene were incubated at 28 °C following the addition of IPTG

HCBT1-MBP: Protein was cultured in Terrific Broth, inoculated with IPTG to a final concentration of 0.5 mM and incubated at 15 °C post induction.

There was a power failure during expression of *RpCL*, *LcCL*, *ArCL* and *MsCL* and it is presumed that the cultures warmed to room temperature at some point with no shaking.

2.2 General protocol for the batch purification of CoA ligases and N-acyltransferases

Cell pellets were resuspended in 35 ml lysis buffer (50 mM Tris, 150 mM NaCl, 10 % glycerol, pH 7.4-8.0). The cells were lysed by sonication using a 13 mm probe for a total of 5 minutes (9 sec on, 9.0 sec off) at 40 % amplitude. The lysates were cleared by centrifugation (39,191 g, 4°C, 45-60 minutes).

Ni-NTA resin was equilibrated with lysis buffer. Briefly, the resin was resuspended in 3-4 volumes of lysis buffer, the resin was pelleted by centrifugation (4000 rpm, 5 minutes, 4 °C) and the supernatant was decanted. This wash step was repeated three times and then the resin was resuspended in lysis buffer to give a 50 % slurry. The equilibrated resin (2-3 ml) was added to the cell free lysate and incubated on a blood roller at 4°C overnight. The matrix was packed into a 10 ml polypropylene column collecting the flow through.

The resin was washed with 7-8 ml wash buffer (50 mM Tris, 150 mM NaCl, 10 % glycerol, 20-25 mM imidazole, pH 7.4-8.0) and the wash was collected in two fractions. The protein was eluted in buffer (50 mM Tris, 150 mM NaCl, 10 % glycerol, pH 7.4-8.0) containing 50 mM imidazole (2 x 3-4 ml fractions), 100 mM imidazole (2 x 3-4 ml fractions), 200 mM imidazole (2 x 3-4 ml fractions) and then 400 mM imidazole (2 x 3-4 ml fractions).

Fractions were analysed by SDS-PAGE (gel images for certain enzymes are shown in figure 1). Fractions containing protein with the expected molecular weight were pooled and concentrated using a centrifugal concentrator (10-30 kDa cut off) to 2.5 ml final volume. Proteins were loaded onto a PD10 column (GE Healthcare) pre-equilibrated with 25 ml storage buffer (50 mM Tris containing 300 mM NaCl, 10 mM MgCl₂, 10 % glycerol, pH 7.4). The flow through was discarded and the protein was eluted in 3.5 ml storage buffer.

Protein concentration was determined using a Bradford assay according to the manufacturers protocol (Bio-Rad). The proteins were diluted in storage buffer to a working concentration of 0.5 mg/ml and stored at -80°C until required.

The general protocol was followed for phl, PhCL, CBL, ipfF, ScCL, phl2, stlB, RpCL, LcCL, ArCL, MsCL, HoIE, CASHT, HCBT1-MBP, BcAT, HpA3 proteins with the following exceptions: For phl, PhCL, CBL, ipfF, ScCL, phl2, stlB, HoIE, CASHT, MpR1, BcAT, HpA3 and HCBT-1-MBP the lysis buffer additionally contained a 1/200 dilution of protease inhibitor cocktail.

For RpCL, LcCL, ArCL, MsCL proteins, the equilibrated Ni-NTA resin (2-3 ml) was added to the cell free lysate and incubated on a blood roller at 4°C for 1.5 hours.

Example 3: Pyrophosphate detection assay for the evaluation of CoA ligase substrate specificity

The activity of nine CoA ligases (phl, PhCL, CBL, ipfF, stlB, RpCL, LcCL, ArCL and MsCL purified as described in Example 2) towards a panel of 16 carboxylic acid substrates (structures shown in Figure 2) were screened using the EnzChek pyrophosphate (PPi) assay kit (Invitrogen) according to the manufacturers protocol.

3.1 Materials and Method

Assay solution

430 µl water

50 µl 20 x reaction buffer (1.0 M Tris-HCl, 20 mM MgCl₂, pH 7.5, containing 2 mM sodium azide)

200 µl MESG (2-amino-6-mercapto-7-methylpurine ribonucleoside) solution (1 mM stock)

10 µl purine nucleoside phosphorylase solution (from 100 U/ml stock solution)

10 µl inorganic pyrophosphatase solution (made just prior to use by ten-fold dilution of the inorganic pyrophosphate stock solution (30 U/ml) in 1 x reaction buffer)

Substrate solution

1.6 mM substrate (**1-16**), 1.6 mM ATP, 1.6 mM Coenzyme A in 50 mM Tris pH 7.5

5 µl CoA ligase solution (0.5 mg/ml) and 25 µl assay solution were mixed in a 384 well plate (clear bottom). To initiate the reaction, 10 µl of substrate solution was added to each well and the change in absorbance was measured at 360 nm every minute over 40 minutes. Reactions were performed in triplicate.

The V_{max} (milli OD units/min) was calculated using Softmax Pro software and was taken over 5 points. The V_{max} values were blank corrected and the average is shown in the graph in figure 3.

3.2 Results and Conclusion

The CoA ligases assayed have complimentary substrate profiles and together catalyse the ligation of acid molecules across a broad range of chemical space. CoA ligases were identified which accept structurally diverse acids containing aromatic, heteroaromatic, allylic and aliphatic functionality.

Example 4: The activity and substrate promiscuity of CoA ligases

The activity of CoA ligases phl, PhCL, CBL and ipfF towards a panel of nine acids (**3**, **5-9**, **11**, **14** & **16**) were evaluated using HPLC analysis. Nine reactions were prepared each with a different acid and CoA ligase pair according to table 1.

4.1 Materials and Method

Biotransformations were prepared as follows in 50 mL falcon tubes:

600 µl Tris pH 8.0 (0.5M)

300 µl ATP solution (50 mM solution prepared in water)

300 µl MgCl₂ solution (50 mM solution prepared in water)

750 µl Acid substrate solution (20 mM prepared in 50 mM Tris buffer pH 8.0)

300 µl CoA solution (50 mM prepared in water)

3.25 ml water

The solutions were pH adjusted to pH 8.0 prior to the addition of the CoA ligase solution (500 µl, 0.5 mg/ml). The reactions were incubated at 21 °C, 150 rpm for 48 hours in a Kuhner Climo Shaker.

TABLE 1: Acid and CoA ligases pairs evaluated

Substrate	CoA ligase
3 cinnamic acid	<i>PhCL</i>
5 phenoxyacetic acid	phl
6 hydrocinnamic acid	phl
7 4-fluorophenylacetic acid	ipfF
8 ibuprofen	ipfF
9 4-chlorobenzoic acid	CBL
11 4-nitrobenzoic acid	CBL
14 indole-5-carboxylic acid	<i>PhCL</i>
16 <i>trans,trans</i> -2,4-hexadienoic acid	<i>PhCL</i>

4.2 HPLC analysis

For HPLC analysis, a 500 µl sample was removed from each reaction at 24 hour and 48 hour time points and 50 µl HCl (5M) was added to quench the reaction. The reaction was vortexed for 30 seconds and the precipitated protein was cleared by centrifugation (13000 rpm, 1 minute). The supernatant was injected onto the HPLC.

HPLC analysis was conducted on an Agilent 1290 infinity II, Zorbax SB-C18 column (50 mm x 3.0 mm, i.d. 1.8 µm packing diameter). The solvents employed were A = water + 0.1 % formic acid and B = Acetonitrile + 0.1 % formic acid and a gradient method was used according to table 2. Flow rate: 1.5 ml/min, Temperature: 60 °C, Injection volume: 2 µL, UV detection: 220 nm

TABLE 2: Solvent gradient used for the detection of CoA ester products

Time (min)	% A	% B
0.00	100	0
2.50	5	95
2.70	5	95
2.71	100	0
4.00	100	0

4.3 Results and Conclusion

Table 3 shows the peak area for each of the starting materials (acid and CoA) and the product (CoA ester) as a percentage of the summed peak areas. ATP does not bind to the column and elutes in the solvent front. Figure 4 shows an example of a HPLC trace for the PhCL catalysed formation of cinnamoyl-CoA.

TABLE 3: Conversion of acid to CoA ester product catalysed by CoA ligases

Substrate	Time (hrs)	Peak Area (%)			CoA-ester retention time (min)
		CoA	CoA ester	Acid	
3	24	14.7	38.0	47.2	1.41
	48	11.1	42.2	46.7	
5	24	53.7	3.7	42.6	1.30
	48	53.0	4.5	42.4	
6	24	36.2	52.5	11.3	1.40
	48	84.2	0	15.8	
7	24	33.4	57.8	8.8	1.32
	48	84.0	0	16.0	
8	24	14.7	59.1	26.3	1.81
	48	6.7	71.6	21.7	
9	24	34.4	47.9	17.7	1.41
	48	58.7	0	41.3	
11	24	45.1	29.0	25.9	1.29
	48	62.2	0	37.8	
14	24	20.2	25.8	54.0	1.23
	48	8.8	40.2	51.0	
16	24	10.8	15.7	73.6	1.30
	48	19.5	56.7	23.9	

The conversion of acids **3**, **5**, **8**, **14** & **16** to the corresponding CoA-esters increased between 24 and 48 hours. After 24 hours incubation CoA-ester product generated from acids **6**, **7**, **9**, and **11** was detected, however after 48 hours incubation the thioester hydrolysed to acid starting material. It is possible that in these reactions, the CoA ligase is no longer active after 24 hours and the product can hydrolyse chemically to the corresponding acid.

Example 5: The effect of coenzyme A concentration on amide formation *in vitro*

To demonstrate that *in vitro* amide formation catalysed by CoA ligases and N-acyltransferases is catalytic in CoA, small scale reactions were performed using a range of CoA concentrations, which were analysed by LC-MS

5.1 Method

Reactions were prepared in 1.5 ml Eppendorf tubes according to table 4. Reactions were incubated overnight at 21 °C with shaking at 700 rpm in an Eppendorf thermomixer.

TABLE 4: Reaction components for *in vitro* amide formation catalysed by CoA ligase PhCL and N-acyltransferase CASHT

volume	Reagent	Final Conc.
100 µl	0.5 M Tris pH 8.0	50 mM
50 µl	50 mM ATP in water	2.5 mM
50 µl	CoA various concentrations	various
250 µl	10 mM acid 6 in 50 mM Tris pH 8	2.5 mM
250 µl	10 mM amine 17 in 50 mM Tris pH 8	2.5 mM
100 µl	PhCL CoA ligase (0.5 mg/ml)	0.8 µM
100 µl	CASHT acyltransferase (0.5 mg/ml)	1.8 µM
100 µl	water	

5.2 LC-MS analysis

For LC-MS analysis, a 500 µl sample was taken from each biotransformation and acetonitrile (500 µl) was added to solubilise the amide product and to precipitate any proteins. The samples were cleared by centrifugation (1 min at 13,000 rpm). A 400 µl sample of the supernatant was filtered using a filter vial (Fisher, part no. 11969414, Mini-UniPrep Syringeless Filter, 0.45 µm, PVDF filtration) and injected directly onto the LC-MS.

The LC analysis was conducted on an Acquity UPLC CSH C18 column (50mm x 2.1mm, i.d. 1.7µm packing diameter). The solvents employed were A = 10 mM Ammonium Bicarbonate in water adjusted to pH 10 with Ammonia solution and B = Acetonitrile and a gradient method was used according to the table 5. Flow rate: 1ml/min, Temperature: 40 °C, Injection volume: 0.3 µL. The UV detection was a summed signal from wavelength 210nm to 350nm. MS analysis was performed using a Waters ZQ, Ionisation mode: Alternate-scan positive and negative electrospray, Scan Range: 100 to 1000 AMU, Scan Time: 0.27 seconds, Inter scan Delay: 0.10 seconds

TABLE 5: Solvent gradient used for the detection of amide products

Time (min)	% A	% B
0.00	97	3
0.05	97	3
1.50	5	95
1.90	5	95
2.00	97	3

5.3 Results and Conclusions

Table 6 gives the peak areas of each component as a percentage of the summed peak areas. Significant amide formation was observed at each CoA concentration evaluated. When CoA was not supplemented into the reaction, amide formation was still detected. It is possible that CoA present in *E. coli* binds to the protein during expression and remains bound during purification, which is sufficient to catalyse the reaction *in vitro*. In any event, the results suggest that only a catalytic amount of CoA is required for amide formation, for example <0.01 eq or 1 mol%.

TABLE 6: Conversion to amide product using various concentrations of CoA substrate

CoA equivalents	Final CoA concentration (mM)	peak area (%) [*]		
		Amine	Amide	Other
1	2.5	17.9	78.5	4.41
0.5	1.25	23.27	71.24	5.5
0.25	0.6	19.9	74.94	5.15
0.1	0.25	14.55	82.03	3.43
0.01	0.025	11.75	85.62	2.62
0	0	13.09	83.43	3.48

^{*}acid elutes at 0.41 min, amine elutes at 0.81 min, amide product elutes at 1.11 min, ES+ 333.2, ES- 331.2

Example 6: The effect of Coenzyme A concentration on the rate of CoA ester formation

To investigate the effect of coenzyme A concentration on the rate of CoA ester formation, reactions were analysed using the EnzChek pyrophosphate (PPI) assay kit (Invitrogen) as described in Example 3 but with the addition of an N-acyltransferase (HolE) and amine.

6.1 Materials and Method

Assay solution

430 µl water

50 µl 20 x reaction buffer (1.0 M Tris-HCl, 20 mM MgCl₂, pH 7.5, containing 2 mM sodium azide)

200 µl MESG (2-amino-6-mercapto-7-methylpurine ribonucleoside) solution (1 mM stock)

10 µl purine nucleoside phosphorylase solution (from 100 U/ml stock solution)

10 µl inorganic pyrophosphatase solution (made just prior to use by ten-fold dilution of the inorganic pyrophosphate stock solution (30 U/ml) in 1 x reaction buffer)

Substrate solution

1.6 mM acid substrate (**3**, **7** or **9**), 1.6 mM amine **17**, 1.6 mM ATP and varying concentrations of Coenzyme A in 50 mM Tris pH 7.5

5 µl HolE (0.5 mg/ml), 5 µl CoA ligase pHL, PhCL, CBL or ipfF (0.5 mg/ml), 25 µl assay solution and 10 µl substrate solution were added to a 384-well plate and the change in absorbance was measured at 360 nm every minute over 10 minutes. Wells containing no ATP and no acid served as negative controls and reactions were performed in duplicate. The V_{max} (milli OD units/min) was calculated using the Softmax Pro software and was taken over 5 points, the average V_{max} is reported in the graph in figure 5.

6.2 Results and Conclusions

In general, the rate of ATP turnover and therefore CoA ester formation increased at high concentrations of CoA. In agreement with the results in example 5, PhCL demonstrated activity when no CoA was supplemented into the reaction and a similar result was also achieved using ipfF. This suggests that CoA is bound to the enzyme in the protein preparation. In contrast, CBL and phI were not active in the absence of CoA addition, which may indicate the K_M for CoA is higher for these proteins.

Example 7: Amine substrate profiling of N-acyltransferases using an Ellman's Spectrophotometric assay

The activity of a number of N-acyltransferase enzymes towards a panel of amine substrates (see figure 6) using commercially available CoA esters was investigated using an Ellman's assay. Ellman's reagent (5,5'-dithiobis-2-nitrobenzoic acid) reacts with free thiols to give a product which absorbs at 412nm and can be used to detect the formation of CoA produced in the amide formation reaction.

7.1 Materials and method

Assay solution was prepared as follows:

500 μ L purified enzyme (0.5 mg/mL)

200 μ L CoA ester solution (2.5 mM) prepared in 50 mM Tris buffer pH 8.0

200 μ L Ellmans Reagent (2.5 mM) prepared in 50 mM Tris buffer pH 8.0

45 μ L assay solution was added to 5 μ L amine solution (**18-31** see figure 6) (5mM prepared in 50 mM Tris buffer pH 8.0) in a 384-well plate. The change in absorbance was measured at 412nm every 20 seconds for 30 minutes. Reactions were performed in triplicate and wells containing no amine served as negative controls. The V_{max} (milliOD units/min) was calculated over 10 points using the SoftMaxPro software. The V_{max} was blank corrected and the average is reported in the graph in figure 7.

7.2 Conclusion

CASHT demonstrated high activity towards amines **23** and **30**, HoLE was active towards substrates **25** and **26**, whilst HCBT was specific for **21**. Although these proteins don't appear to be highly promiscuous, they have complimentary activities and together give access to a structurally diverse set of amide products

Example 8: Optimisation of one-pot CoA Ligase and N-acyltransferase biotransformations

To establish optimal conditions for a one-pot biotransformation using purified CoA ligases and N-acyltransferases, two concentrations of coenzyme A (1 and 0.1 equivalents) were screened using varying enzyme concentrations.

8.1 Materials and Method

Reactions were prepared in 1.5 ml Eppendorf tubes according to table 7. Reactions were incubated overnight at 21 °C with shaking at 700 rpm in an Eppendorf thermomixer.

TABLE 7: Reaction components for *in vitro* amide formation catalysed by CoA ligase PhCL and N-acyltransferase CASHT

volume	Reagent	Final Conc.
100 µl	0.5 M Tris pH 8	50 mM
50 µl	50 mM ATP in water	2.5 mM
50 µl	50 mM or 5 mM CoA in water	2.5 mM or 0.25 mM
250 µl	10 mM hydrocinnamic acid 6 in 50 mM Tris pH 8	2.5 mM
250 µl	10 mM Amine 17 in 50 mM Tris pH 8	2.5 mM
100 µl	water	-
100 µl	PhCL CoA ligase	various
100 µl	CASHT N-acyltransferase	various

Samples were analysed according to the method in example 5.2

8.2 Results and Conclusions

Table 8 shows the conversion to amide product and the peak areas are reported as a percentage of the summed peak areas. In addition the results are represented graphically in figure 8. The results suggest that using 0.1 equivalents of CoA yields higher conversion to the amide product than using 1 CoA equivalent and that the optimal ratio of CASHT:PhCL is approx. 2:1.

TABLE 8: Conversion to amide product using 0.1 and 1 CoA equivalent and various protein concentrations

protein (μM)		ratio	peak area (%)		
CASHT	PhCL	CASHT/PhCL	Amine	Amide	Other
1 CoA equivalent					
1.8	0.08	22.5	38.29	49.84	10.33
1.8	0.16	11.25	28.98	62.59	8.42
1.8	0.8	2.25	17.9	78.5	4.41
0.36	0.8	0.45	21.8	70.54	6.28
0.18	0.8	0.225	22.63	57.72	6.39
0	0	0	74.13	0	22.37
0.1 CoA equivalents					
1.8	0.08	22.5	35.27	54.81	9.92
1.8	0.16	11.25	23.49	69.59	6.92
1.8	0.8	2.25	14.55	82.03	3.43
0.36	0.8	0.45	27.02	66.41	6.56
0.18	0.8	0.225	14.31	81.48	4.21
0	0	0	76.15	0	23.85

Example 9: In vitro amide formation using a CoA ligase and N-Acyltransferase

Biotransformations were performed *in vitro* using CoA ligases (pHl, PhCL, CBL and ipfF) in combination with N-acyltransferases (CASHT and HolE) as purified proteins. The enzyme activity was evaluated towards a small panel of acids and amines (figure 2 & 6).

9.1 Materials and Method

Reactions were prepared in 1.5 ml Eppendorf tubes according to table 9 and incubated overnight at 21 °C with shaking 700 rpm in an Eppendorf thermomixer. Reactions containing no N-acyltransferase served as a negative control.

TABLE 9: Reaction components for *in vitro* amide formation catalysed by a CoA ligase and N-acyltransferase.

volume	Reagent	Final Conc.
100 µl	0.5 M Tris pH 8.0	50 mM
50 µl	50 mM ATP in water	2.5 mM
50 µl	50 mM CoA in water	2.5 mM
250 µl	10 mM Amine 50 mM Tris pH 8	2.5 mM
250 µl	10 mM Acid in 50 mM Tris pH 8	2.5 mM
100 µl	water	
100 µl	CoA ligase (0.5 mg/ml)	
100 µl	acyltransferase (0.5 mg/ml)	

Reactions were analysed according to the method described in example 5.2

9.2 Results and Conclusion

A summary of the reactions evaluated is presented in table 10, the associated LC-MS traces for successful reactions which yielded amide product are shown in figure 9. There was a background chemical reaction between the CoA ester corresponding to acid **7** with amines **23**, **30** and **32**, which was observed in the negative control. There was no detectable increase over the background in the corresponding N-acyltransferase (HoIE) reactions; therefore amines **23**, **30** and **32** are not substrates for HoIE. Nonetheless, sixteen of the reactions screened were successful and yielded structurally diverse amide products in good yields.

TABLE10: Summary of *in vitro* amide bond forming biotransformations evaluated.

Reaction	Acid	CoA Ligase	Amine	Acyl-transferase	Amide formation
1	6	PhCL	22	CASHT	No
2	6	PhCL	23	CASHT	Yes
3	6	PhCL	24	CASHT	No
4	6	PhCL	25	CASHT	No
5	6	PhCL	26	CASHT	No
6	6	PhCL	27	CASHT	No
7	6	PhCL	28	CASHT	Yes
8	6	PhCL	30	CASHT	Yes
9	6	PhCL	32	CASHT	Yes
10	6	PhCL	33	CASHT	No
11	3	PhCL	17	CASHT	Yes
12	5	phl	17	CASHT	Yes
13	7	ipfF	17	CASHT	Yes
14	8	ipfF	17	CASHT	No
15	9	CBL	17	CASHT	Yes

16	14	PhCL	17	CASHT	Yes
17	16	phl	17	CASHT	No
18	7	ipfF	17	HolE	Yes
19	7	ipfF	22	HolE	No
20	7	ipfF	23	HolE	No*
21	7	ipfF	24	HolE	Yes
22	7	ipfF	26	HolE	Yes
23	7	ipfF	27	HolE	No
24	7	ipfF	28	HolE	Yes
25	7	ipfF	30	HolE	No*
26	7	ipfF	32	HolE	No*
27	7	ipfF	33	HolE	No
28	5	phl	25	HolE	Yes
29	6	phl	25	HolE	Yes
30	8	ipfF	25	HolE	No
31	9	CBL	25	HolE	No
32	14	PhCL	25	HolE	No
33	16	phl	25	HolE	Yes

*Amide formation was detected but was comparable to that detected in the negative control

Example 10: Co-expression of a CoA ligase and N-acyltransferase in *E. coli*

10.1 Preparation of co-transformed *E. coli* cells

Chemically competent BL21(DE3) and Rosetta2(DE3) cells were transformed with pCDF-Duet1 vectors carrying the gene for CASHT expression. Cells were plated onto LB agar containing streptomycin (50 µg/ml) and incubated at 37 °C overnight. Colonies were picked and used to inoculate 5 ml LB containing streptomycin (50 µg/ml). Cultures were grown overnight at 37 °C with shaking (200 rpm). The starter culture was used as a 1 % inoculum to inoculate 20 ml LB containing streptomycin (50 µg/ml) in a 100 ml Erlenmeyer flask. The cultures were grown at 37°C, 200 rpm until they reached an OD₆₀₀ of 0.5 (2-5 hours). Samples (2 ml) were taken from each culture and harvested by centrifugation (13,000 rpm, 4 °C, 2 min). The cells were resuspended in 1 ml filter sterilized 0.1 M CaCl₂ and the cells were then pelleted by centrifugation. The cells were resuspended in 200 µl of the 0.1 M CaCl₂ solution and again the cells were pelleted by centrifugation. The cells were resuspended in a final volume of 60 µl CaCl₂ and incubated on ice. The Rosetta2(DE3)-pCDF-CASHT strain was transformed with pET28-PhCL and the BL21(DE3)-pCDF-CASHT strain was transformed with pET28-phl, pET28-CBL and pET28-ipfF. The transformations were plated onto LB agar containing kanamycin (50 µg/ml) and streptomycin (50 µg/ml) and incubated at 37°C overnight.

10.2 Co-expression of CoA ligase and N-acyltransferase genes

Colonies of the co-transformed cells were picked and used to inoculate 10 ml LB containing kanamycin (50 µg/ml) and streptomycin (50 µg/ml), the Rosetta2(DE3) cultures also contained chloramphenicol (34 µg/ml). The starter cultures were grown overnight at 37°C with shaking 200 rpm. LB media (200 ml) containing the appropriate antibiotics and 1 % glycerol was inoculated with 4 ml of each starter culture (2 % inoculum). The cultures were grown at 37 °C until they reached an OD₆₀₀ of 0.6-0.8 (3 hours). Expression was induced with the addition of IPTG to a final concentration of 0.3 mM. The cultures were incubated at 18 °C, 200 rpm overnight (18 hours).

Protein production was analysed by SDS-PAGE. A sample was taken from each culture and the cells were pelleted by centrifugation. The cells were resuspended in 1/10th OD of bugbuster containing 1 U/ml benzonase. Cell suspensions were incubated at room temperature for 15 minutes. A sample (30 µl) of the total cell lysate was removed and the remaining cells were pelleted by centrifugation (13,000 rpm, 10 minutes). A second sample (30 µl) was taken from the soluble supernatant. Samples (30 µl) were mixed with 4x SDS loading buffer (10 µl) and heated to 95 °C for 5 minutes. The samples were run on SDS-gels 4-20% according to the manufacturer's protocol. The presence of CoA ligase and N-acyltransferase proteins was confirmed by Western blot analysis (figure 10).

Western blot analysis:

Protein bands were transferred from the gels to PVDF membranes using the iBlot system (according to the manufacturer's protocol). The membranes were incubated at room temperature in blocking buffer (5% non-fat milk powder in PBS + 0.1% tween) for 2 hours and then washed with PBS + 0.1% tween. The membrane was incubated in blocking buffer plus anti-polyhistidine-HRP conjugate for 1 hour at room temperature and then washed thoroughly with PBS + 0.1% tween. The anti-polyhistidine-HRP conjugate was imaged using Super Signal West extended duration substrate (Pierce) according to the manufacturer's protocol.

10.3 Results

The co-expression of CASHT with PhCL and phl yielded high levels of soluble protein. Under the same conditions, CBL and ipfF protein was largely in the insoluble fraction, however the soluble expression could be further improved by optimisation of the expression conditions.

Example 11: *In vivo* amide bond formation using a CoA ligase and N-acyltransferase

Amide formation using whole cell catalysts expressing a CoA ligase and N-acyltransferase was investigated to determine the requirements for extracellular ATP and CoA. Four strains were evaluated each expressing the N-acyltransferase CASHT and one of the four CoA ligases phl, PhCL, CBL or ipfF.

11.1 Method for the preparation of whole cell catalysts

Whole cell catalysts co-expressing a CoA ligase and N-acyltransferase were prepared in 200 mL cultures according to example 10.2. Cells were harvested by centrifugation (400 rpm, 20 minutes, 4 °C) and resuspended in PBS to a total volume of 35 ml. Samples (500 µL) from each culture were transferred to 1.5 mL eppendorf tubes and pelleted by centrifugation (13,000 rpm, 2 min, 4 °C). Samples were either used directly in biotransformations (termed 'resting cells'), frozen in dry ice and stored at -80 °C for 3 days (termed 'frozen cells') or frozen in dry ice then lyophilized under vacuum for 18 hours and stored at -80 °C for 3 days (termed 'lyophilized cells').

Cell pellets were resuspended in 50 mM Tris buffer pH 8.0 (1 ml) containing either 2.5 mM or 20 mM of acid and amine **17**, with and without the addition of glucose (2 % v/v). Reactions were incubated at 21 °C overnight (16 hours) in an Eppendorf thermocycler with shaking at 700 rpm. Samples were analysed according to the method in example 5.2

11.2 Results and Conclusions

Reactions performed at 20 mM substrate concentration did not yield any amide product. However at 2.5 mM substrate concentration all four strains successfully produced the expected amide product

(table 11). All strains yielded high conversions when used as 'resting cells' but were not as active after freezing and/or lyophilisation.

TABLE 11: Amide formation using different cell preparations

Strain	Preparation	Acid	Peak area (%) ^[1]		
			Acid	Amine	Amide
BL21(DE3)-CASHT-phI	Resting cells + glucose	6	-	-	100
	Resting cells	6	31.1	14.6	54.2
	frozen cells	6	22.1	64.9	-
	lyophilized cells	6	22.7	64.7	-
Rosetta2(DE3)-CASHT-PhCL	Resting cells + glucose	3	-	12.0	87.0
	Resting cells	3	-	12.7	87.3
	frozen cells	3	26.7	4.0	69.2
	lyophilized cells	3	50.5	3.9	42.2
BL21(DE3)-CASHT-CBL	Resting cells + glucose	9	3.8	8.6	87.0
	Resting cells	9	2.6	5.0	92.3
	frozen cells	9	26.4	24.7	48.9
	lyophilized cells	9	48.8	26.2	25.0
BL21(DE3)-CASHT-ipfF	Resting cells + glucose	7	67.4	27.8	-
	Resting cells	7	46.8	17.6	35.58
	frozen cells	7	21.9	65.2	13.0
	lyophilized cells	7	20.1	54.6	25.3

[1] Peak area of acid and amine starting materials and amide product as a percentage of the summed peak areas

These results highlight that whole cells co-expressing carefully selected CoA ligases and N-acyltransferases can be used to access a selection of non-naturally occurring amides without supplementation with extracellular CoA or ATP.

Example 12: Optimisation of *in vivo* 6-chloro-N-neopentylnicotinamide (**35**) production

CoA ligase CBL and N-acyltransferase CASHT catalyse the formation 6-Chloro-N-neopentylnicotinamide (**35**) from 6-chloronicotinic acid (**33**) and neopentylamine (**34**) substrates (figure 12). This reaction has been optimised by recloning the genes into a single Duet vector. The effect of reaction media, substrate loading and catalyst loading on the yield of the amide product has also been evaluated.

12.1 Generation of a pCDF-Duet-CASHT-CBL construct

The CoA ligase construct pET28a-CBL and N-acyltransferase construct pCDF-Duet1-CASHT were generated at Genscript as described in Example 1. For co-expression of CBL and CASHT in *E. coli*, the CBL gene was recloned into the second expression cassette of the pCDF-Duet1-CASHT construct.

3-4 µg of each construct pET28a-CBL and pCDF-Duet1-CASHT was digested in a 100 µl reaction containing 10 µl cutsmart buffer, 4 µl NdeI and 4 µl XhoI (restriction enzymes, NEB). Digests were incubated at 37 °C for 3 hours. The entire restriction digest was run on a 1 % w/v agarose gel and the DNA fragments corresponding to the CBL gene and digested pCDF-Duet1-CASHT vector were

excised. The DNA was purified using a gel extraction kit (Qiagen) according to the manufacturer's protocol.

The CBL gene insert was ligated into pCDF-Duet1-CASHT in a 3:1 ratio using 50 ng of vector and T4 DNA ligase. The ligation reaction was used to transform *E. coli* NEB 5 α chemically competent cells which were plated onto LB agar containing streptomycin (50 μ g/ml). The plates were incubated at 37 °C overnight.

Successful clones were identified by colony PCR and the sequence was verified by double strand sequencing using the following primers:

pCDF_seq 1 taaggagatataccatgg
pCDF_seq 2 ctttctgttcgacttaag
pCDF_seq 3 ctttaagtcgaacagaaag
pCDF_seq 4 gtaaccagaccttg
pCDF_seq5 caaaggtctggttac
pCDF_seq6 cggtttctttaccag

12.2 Co-expression of CBL and CASHT in *E. coli*

The pCDF-Duet-CASHT-CBL construct generated in example 12.1 was used to transform chemically competent *E. coli* BL21(DE3). The cells were plated onto LB agar containing streptomycin (50 μ g/ml) and incubated at 37 °C overnight. Colonies were used to inoculate 5 mL LB media containing streptomycin (50 μ g/ml) and the culture was incubated at 37 °C overnight with shaking 200 rpm.

The overnight culture (2 ml) was used to inoculate LB media (200 mL) containing streptomycin (50 μ g/ml). The culture was incubated at 37 °C with shaking (200 rpm) until an OD₆₀₀ 0.6-0.8 was reached (2.5 hours). Expression was induced with the addition of IPTG to a final concentration of 0.3 mM. The culture was incubated at 28 °C, 200 rpm overnight (18 hours) to give a final OD₆₀₀ of 3.9.

To analyse the level of protein production, cells were lysed and analysed by SDS-PAGE. A 1 ml sample was taken from the culture and the cells were pelleted by centrifugation. The cells were resuspended in 400 μ l bugbuster containing 1 U/ml benzonase and 0.1 mg/ml lysozyme from chicken egg white. Cell suspension was incubated at room temperature for 20 minutes. A sample (20 μ l) of the total cell lysate was removed and the remaining cells were pelleted by centrifugation (13,000 rpm, 10 minutes). A second sample (20 μ l) was taken from the soluble supernatant (soluble fraction). Samples (20 μ l) were mixed with SDS loading buffer (10 μ l) and heated to 95 °C for 5 minutes. The samples were run on SDS-gels 12-20% according to the manufacturer's protocol. The stained gel is shown in figure 11 and show successful co-expression of both proteins.

12.3 Optimisation of in vivo 6-chloro-N-neopentylNicotinamide (35) production: Reaction media and substrate loading

Cultures of *E. coli* co-expressing CBL and CASHT were produced according to Example 12.2. A 200 mL culture was divided between 6 x 50 mL falcon tubes and the cells were pelleted by centrifugation (20 minutes, 4000 rpm). The cell pellets were resuspended in 5 mL reaction media

(see table 13) containing either 2.5 mM or 10 mM 6-chloronicotinic **33** acid and neopentylamine **34**. Biotransformations were incubated at 21 °C overnight (16 hours) with shaking at 200 rpm.

12.3.1 LC-MS analysis

For analysis, a sample (500 µl) was taken from each biotransformation and acetonitrile (500 µl) was added to solubilise the amide product and to precipitate any proteins. The samples were cleared by centrifugation (1 min at 13,000 rpm). The supernatant was filtered using a filter vial (Fisher, part no. 11969414, Mini-UniPrep Syringeless Filter, 0.45 µm, PVDF filtration) and injected directly onto the LC-MS.

LC analysis was conducted on an Acquity UPLC CSH C18 column (50mm x 2.1mm, i.d. 1.7µm packing diameter). The solvents employed were A = 0.1 % v/v solution of Formic Acid in Water and B = 0.1% v/v solution of Formic Acid in Acetonitrile and a gradient method was used according to table 12. Flow rate: 1ml/min, Temperature: 40 °C, Injection volume: 0.5 µL. The UV detection was a summed signal from wavelength 210 nm to 350 nm. MS analysis was performed using a Waters ZQ, Ionisation mode: Alternate-scan positive and negative electrospray, Scan Range: 100 to 1000 AMU, Scan Time: 0.27 seconds, Inter scan Delay: 0.10 seconds

TABLE 12: Solvent gradient used for the detection of amide products

Time (min)	% A	% B
0.00	97	3
1.50	5	95
1.90	5	95
2.00	97	3

12.3.2 Results and Conclusion

When reactions were performed in LB media the reactions went to completion at both 2.5 mM and 10 mM substrate concentrations (table 13). When the reactions were performed in Tris buffer the reactions did not go to completion. These results may suggest that in LB media the resting cells are able to regenerate ATP more effectively and therefore higher yields of the amide product can be achieved.

TABLE 13: The conversion to amide product under different reaction conditions

	Reaction media	substrate concentration	Final DMSO concentration	amide peak area*
1	50 mM Tris buffer pH 8.0	2.5 mM	0.5 %	76 %
2	50 mM Tris buffer pH 8.0	10 mM	2 %	28 %
3	50 mM Tris buffer, 150 mM NaCl, 10 mM MgCl ₂ , 10% glycerol v/v	2.5 mM	0.5 %	100 %
4	50 mM Tris buffer, 150 mM NaCl, 10 mM MgCl ₂ , 10% glycerol v/v	10 mM	2 %	71 %
5	LB growth media	2.5 mM	0.5 %	100 %
6	LB growth media	10 mM	2%	100 %

* Amide formation is reported as a peak area relative to the remaining acid (peak area (%) = amide peak area/(amide peak area+acid peak area)*100.

12.4 Optimisation of in vivo 6-chloro-N-neopentylNicotinamide (35) production: Catalyst loading and substrate concentration

Fresh cell paste was produced according to Example 12.2 with the following exceptions: A glycerol stock of BL21(DE3) transformed with pCDF-Duet1-CASHT-CBL was used to inoculate a 5 mL LB starter culture, protein was expressed in a 500 mL culture containing 1 % glycerol, 0.05 % antifoam and streptomycin (50 µg/ml) and cultures were incubated at 18 °C post induction.

Cells were harvested by centrifugation (20 mins, 4000 rpm, 4 °C) and resuspended in 50 mL LB. Stock solutions of 6-chloronicotinic **33** and amine neopentylamine **34** (1M) were prepared in DMSO. Reactions were made up to 5 mL in LB media and contained varying amounts of cell culture and substrate as detailed in table 14. Biotransformations were incubated at 21 °C overnight (16 hours) with shaking at 200 rpm.

TABLE 14: The conversion to amide product under different reaction conditions

	Volume of cell suspension	Final OD ₆₀₀ of reaction	Substrate concentration	Final DMSO concentration	amide peak area* (%)
1	500 µL	4.1	10 mM	2%	62%
2	500 µL	4.1	20 mM	4%	36%
3	500 µL	4.1	50 mM	10%	2%
4	500 µL	4.1	100 mM	20%	0%
5	1500 µL	12.3	10 mM	2%	82%
6	1500 µL	12.3	20 mM	4%	45%
7	1500 µL	12.3	50 mM	10%	13%
8	1500 µL	12.3	100 mM	20%	0%
9	2500 µL	20.5	10 mM	2%	89%
10	2500 µL	20.5	20 mM	4%	48%
11	2500 µL	20.5	50 mM	10%	15%
12	2500 µL	20.5	100 mM	20%	2%

*amide peak area = area of amide/(area of amide + acid)

Reactions were analysed according to the method in example 12.3.1.

12.4.1 Results and Conclusion

A high level of conversion to amide product was achieved at high cell density (OD₆₀₀) and the conversion to product decreased with reduced catalyst loading and increasing substrate concentration. The reactions containing 50-100 mM substrate solution contained a significant amount of DMSO which can cause cell lysis and reduce the efficiency of ATP regeneration, this could explain the low conversions observed.

Example 13: Preparative scale synthesis of 6-chloro-N-neopentylNicotinamide (35) using whole cell biocatalysts expressing a CoA ligase and an N-acyltransferase

To demonstrate that this methodology can be used on a preparative scale, 6-chloro-N-neopentylNicotinamide (35) was produced and isolated on a 100 mg scale from a 60 ml reaction.

CoA ligase CBL and N-acyltransferase CASHT were co-expressed in a 500 mL culture as described in example 12.2. The cells were harvested by centrifugation (4000 rpm, 20 minutes), the supernatant

was removed and the cell pellet was resuspended in 80 mL of fresh LB media which was used directly in the amide formation reaction.

6-chloronicotinic acid (93.6 mg, 0.6 mmol) and neopentylamine (70 μ l, 0.6 mmol) were added to the resuspended cells (60 mL) in a 250 mL Erlenmeyer flask and the reaction was incubated at 21 °C for 16 hours with shaking (200 rpm). To analyse the final conversion, a 200 μ l sample was taken from the reaction and 200 μ l acetonitrile was added. The sample was vortexed and the cell debris was cleared by centrifugation (1 min at 13,000 rpm). The sample was filtered using a Mini-UniPrep Syringeless Filter (0.45 μ m, PVDF, Fisher) and analysed by LC-MS according to the analytical method in example 12.3.1.

The conversion (83%) was calculated relative to the remaining acid. Next, the cells were pelleted by centrifugation (4000 rpm, 10 minutes) and the product was extracted from the supernatant using ethyl acetate (2 x 50 ml). The organic phases were combined, washed with saturated NaHCO₃, dried over MgSO₄ and concentrated under reduced pressure to give the product as a white solid (99 mg, 74 % yield).

¹H NMR (CDCl₃): δ 8.7 (d, J=2.09 Hz, 1H), 8.1 (dd, J=8.31, 2.52 Hz, 1H), 7.4 (d, J=8.31 Hz, 1H), 6. (br. s, 1H), 3.3 (d, J=6.40 Hz, 2H), 1.0 (s, 9H)

Example 14: Substrate profiling of a panel of N-acyltransferases as cleared lysates in 96-well plates

Substrate profiling of a panel of phylogenetically diverse N-acyltransferases (described in example 1.2) was performed in 96-well plates using a panel of structurally diverse amines (**36a-i**, figure 13) and CoA esters produced in situ from the corresponding acids (**3**, **5-7**, **9**, **14-16** & **37**, figures 2 and 13) using an appropriate CoA ligase (identified using the methods in example 3 and purified according to example 2).

14.1 General protocol for the preparation of N-acyltransferase cleared lysates

All transfer and aliquotting steps were performed using a Beckman Biomek FX liquid handling robot.

Glycerol stocks of BL21(DE3) carrying pCDF-Duet-N-acyltransferase constructs were used to inoculate 170 μ l TB media containing 50 μ g/ml streptomycin in each well of a 96-well microtitre plate, the overnight cultures were sealed with a gas permeable seal and incubated overnight at 37 °C, 200 rpm, 85% humidity. 10 μ l of the overnight culture was used to inoculate 390 μ l TB media containing 50 μ g/ml streptomycin and 0.4 % glycerol in each well of a 2 ml 96-deep well plate. The plates were sealed and incubated at 30 °C, 250 rpm, 85% humidity until they reached an OD₆₀₀ of 0.6-0.8 (2 hours). For induction, 40 μ l IPTG (10 mM solution) was added to each well and the cultures were grown for an additional 20 hours at 20 °C, 250 rpm, 85% humidity. The cells were harvested by centrifugation (4000 rpm, 10 minutes, 4 °C) and the supernatant was poured off, cell pellets were stored at -80 °C until required.

For cell lysis, plates were thawed on the bench for 1 hour, 200 μ L lysis buffer (50 mM phosphate buffer pH 7.5 containing 1 mg/ml lysozyme, 0.5 mg/ml polymixin B, 0.1 U/ml Benzonase) was added to each well and the cell pellets were resuspended with shaking (850 rpm) for 2 hours at room temperature. Cell debris was cleared by centrifugation (4000 rpm, 4 °C, 10 minutes). 100 μ l of the

cleared cell lysate was transferred to a fresh 96-well microtitre plate and the lysates were used directly in the assays.

14.2 Assays to evaluate N-acyltransferase substrate promiscuity

All reactions were composed of 50% lysate loading and a final concentration of 2.5 mM amine and 2.5 mM Acid or CoA ester, and were incubated at 25 °C, 800 rpm, for 18 hours. To quench, 100 µl acetonitrile was added to each well and the plate was incubated on a plate shaker for 1 hour at 850 rpm. Precipitated protein was removed by filtration using Acroprep Filter plates, (96-well, 350 µl, 0.2 µm, GHP) and centrifuged at 4000 rpm for 5 minutes. The samples were collected in a 96-well microtitre plate and analysed by UPLC-MS according to the method described in example 5.2 but using a 2 µl injection volume with UV detection being a summed signal from wavelength 210nm to 400nm.

14.2.1 Preliminary assay: Activity towards a panel of amines 36a-d

Cleared cell lysates were screened for activity towards amines **36a-d** using either 1) commercial CoA esters or 2) Cinnamoyl CoA produced *in situ* from cinnamic acid and CoA ligase, PhCL.

14.2.1.1 Materials and Method

Commercial CoA esters were purchased from Sigma: Acetyl coenzyme A sodium salt CAS# 102029-73-2, Malonyl coenzyme A lithium salt CAS# 108347-84-8, Benzoyl coenzyme A lithium salt CAS# 102185-37-5, octanoyl Coenzyme A lithium salt CAS# 324518-20-9

Acid and amine solutions (20 mM) were prepared in 50 mM Tris buffer and pH adjusted to 8.0

Commercial CoA assay mix (volume/well)

5 µl CoA ester (50 mM in water)

32.5 µL buffer (50 mM Tris, pH 8.0)

Cinnamoyl CoA assay mix (Volume/well)

5 µl CoA (50 mM in water, adjusted to pH 5 with NaOH)

6 µL ATP (50 mM in water)

12.5 µL Cinnamic acid (20 mM solution in 50 mM Tris pH 8.0)

10 µL purified PhCL (0.5 mg/ml)

4 µL buffer (50 mM Tris, pH 8.0)

Amine solution (12.5 µl) and CoA ester assay mix (37.5 µl) were added to each well of a 96-well microtitre plate. Using the Biomek FX liquid handling robot, N-acyltransferase lysate (50 µl) was added and reactions were incubated at 25 °C, 800 rpm, for 18 hours. The final reaction mix contained either 2.5 mM CoA ester, 2.5 mM Amine in Tris pH 8.0 with 50% acyltransferase lysate loading or 2.5 mM Acid, 2.5 mM CoA, 5 µg CoA ligase, 3 mM ATP, 2.5 mM Amine in Tris pH 8.0 with 50% acyltransferase lysate loading. Conversion was calculated as a percentage relative to the amine peak ($\text{Conversion} = \frac{\text{product peak area}}{(\text{product peak area} + \text{amine peak area})} \times 100$)

14.2.1.2. Results

Out of the N-acyltransferases screened, 31 enzymes yielded amide product using one or more of the amines screened. Amines **36a** and **36b** were most widely accepted and were later used as amine partners to investigate the CoA ester promiscuity of the N-acyltransferase panel.

Table 15. Activity of the N-acyltransferase panel towards amines **36a-d**

Enzyme	CoAester	Percentage Conversion (%)			
		Amine 36d	Amine 36c	Amine 36a	Amine 36b
01StAT	acetyl	nd ^[1]	0.0	51.1	nd ^[1]
02MsAT	acetyl	nd ^[1]	0.0	18.9	nd ^[1]
05PsAT	acetyl	nd ^[1]	19.4	92.5	nd ^[1]
07NfAT	acetyl	nd ^[1]	67.9	89.8	nd ^[1]
41GmAT	acetyl	nd ^[1]	24.8	90.0	nd ^[1]
42GmAT	acetyl	nd ^[1]	23.8	57.0	nd ^[1]
43GmAT	acetyl	nd ^[1]	0.0	24.7	nd ^[1]
BcAT	acetyl	nd ^[1]	82.1	86.1	nd ^[1]
63ScAT	acetyl	nd ^[1]	0.0	0.0	nd ^[1]
64ScAT	acetyl	nd ^[1]	0.0	4.4	nd ^[1]
39SeAT	acetyl	nd ^[1]	0.0	0.0	nd ^[1]
23DmAT	acetyl	nd ^[1]	100.0	21.2	nd ^[1]
35MaAT	acetyl	nd ^[1]	0.0	3.3	nd ^[1]
36PaAT	acetyl	nd ^[1]	0.0	3.5	nd ^[1]
38ShAT	acetyl	nd ^[1]	0.0	3.5	nd ^[1]
45SpAT	acetyl	nd ^[1]	0.0	3.1	nd ^[1]
47AtAT	acetyl	nd ^[1]	0.0	0.0	nd ^[1]
HpA3	acetyl	nd ^[1]	0.0	0.0	nd ^[1]
51EcAT	acetyl	nd ^[1]	0.0	0.0	nd ^[1]
18BsAT	acetyl	nd ^[1]	4.4	0.0	nd ^[1]
19CpAT	acetyl	nd ^[1]	0.0	3.3	nd ^[1]
10HvAT	cinnamoyl	0.0	0.0	0	4.5
11AtAT	cinnamoyl	0.0	0.0	0.0	31.8
12HvAT	cinnamoyl	0.0	0.0	0	4.0
14AtAT	cinnamoyl	0.0	0.0	0.0	0.0
24LIAT	cinnamoyl	0.0	0.0	0.0	0.0
31NtAT	cinnamoyl	18.2	11.9	0	26.9
32NtAT	cinnamoyl	18.5	8.5	0	30.2
48CaAT	cinnamoyl	49.1	24.7	8.0	31.6
49StAT	cinnamoyl	49.5	21.6	8.8	33.7
54SIAT	cinnamoyl	81.6	5.2	0	24.8
55SIAT	cinnamoyl	48.4	25.7	6.4	30.5
56SIAT	cinnamoyl	43.6	12.5	0	36.2
CASHT	cinnamoyl	33.0	24.9	0	31.7
28BaAT	malonyl	0.0 ^[2]	0.0	0.0	0.0
29EfAT	malonyl	0.0 ^[2]	0.0	0.0	0.0
30SaAT	malonyl	0.0 ^[2]	0.0	0.0	0.0
33MtAT	malonyl	0.0 ^[2]	0.0	0.0	0.0
34PaAT	malonyl	0.0 ^[2]	0.0	0.0	0.0
HCBT	benzoyl	0.0	0.0	0.0	0.0
37HsAT	octanoyl	0.0	0.0	0.0	0.0
58AxAT	octanoyl	0.0	0.0	0.0	0.0
52PIAT	octanoyl	59.1	0.0	29.1	23.5
53PhAT	octanoyl	48.9	0.0	49.0	21.3

61SrAT	octanoyl	6.3	0.0	0.0	61.1
62SrAT	octanoyl	0.0	0.0	0	0.0
HoIE	octanoyl	52.1	0.0	14.0	13.6

[1] N-acetyl phenylalanine methyl ester and N-acetyl 4-(aminomethyl)benzonitrile product peaks and the corresponding amine substrate peaks were not resolved using this analytical method and the conversion was not determined.

[2] No product peaks were observed using malonyl CoA as an acyl donor, however in the absence of a product standard the product retention time was not known.

14.2.2 N-Acyltransferase CoA ester substrate promiscuity

N-acyltransferases which were active towards either substrate **36a** and/or **36b** in the preliminary assay were further evaluated for activity towards CoA esters which were produced *in situ* from acids **3**, **5-7**, **9**, **14-16** & **37** (Figure 2 & 13), using catalytic CoA.

14.2.2.1 Method

CoA ester assay mix (Volume/well)

0.05 μ L CoA (50 mM in water, adjusted to pH 5 with NaOH)

6 μ L ATP (50 mM in water)

12.5 μ L Acid solution (20 mM, in 50 mM Tris pH 8.0)

10 μ L CoA Ligase (0.5 mg/ml)

6.45 μ L buffer (50 mM Tris, pH 8.0)

CoA ester assay mix (37.5 μ L) and amine solution (20 mM, 12.5 μ L) were added to each well of a 96-well microtitre plate. Using the Biomek FX liquid handling robot, N-acyltransferase lysate (50 μ L) was added and reactions were incubated at 25 °C, 800 rpm, for 18 hours. The final reaction mix contained 2.5 mM Acid, 2.5 mM CoA, 5 μ g CoA ligase, 3 mM ATP, 2.5 mM Amine in Tris pH 8.0 with 50% acyltransferase lysate loading. The percentage conversion was calculated relative to standard curves of chemically synthesised amides and amine starting materials.

14.2.2.2 Results

A large number of N-acyltransferases showed promiscuous activity and accepted a number of structurally diverse thioesters, yielding amide products with high conversions. However not all enzymes were active towards alternative thioesters, for example 01StAT, 02MsAT, 64ScAT, 35MaAT, 38ShAT, 45SpAT, and 19CpAT did not accept any of the acyl donors screened in this example and are specific for acetyl CoA (described in the preliminary assay).

Table 16. Activity of the N-acyltransferase panel towards CoA esters produced *in situ* from acids **3**, **5-7**, **9**, **14-16** & **37** and CoA ligases PhCL, ipfF phl, CBL or RpCL

Enzyme	Amine	Percentage conversion (%)								
		3	5	6	7	9	14	15	16	37
		+PhCL	+ipfF	+phl	+ipfF	+CBL	+PhCL	+ipfF	+PhCL	+RpCL
01StAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
02MsAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
05PsAT ^[1]	36a	0.0	25.4	23.7	100.0	0.0	0.0	61.4	0.0	0.0
07NfAT ^[1]	36a	0.0	35.4	45.0	0.0	0.0	0.0	53.4	1.7	0.0
41GmAT ^[1]	36a	27.1	42.8	51.0	73.1	0.0	0.0	45.9	7.4	32.2
42GmAT ^[1]	36a	0.0	13.3	0.0	24.4	0.0	0.0	11.8	0.0	0.0
43GmAT ^[1]	36a	8.2	27.9	26.9	48.7	0.0	0.0	16.4	1.2	5.1
BcAT ^[1]	36a	0.0	33.3	0.0	88.2	0.0	0.0	45.0	0.0	0.0
63ScAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
64ScAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
39SeAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23DmAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
35MaAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36PaAT	36a	0.0	0.0	7.4	0.0	0.0	0.0	0.0	0.0	0.0
38ShAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
45SpAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19CpAT	36a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10HvAT	36b	35.7	0.0	0.0	0.0	0.0	0.0	0.0	51.2	0.0
11AtAT	36b	92.3	11.1	89.7	93.7	88.9	9.3	59.7	77.8	0.0
12HvAT	36b	36.3	0.0	0.0	0.0	0.0	0.0	0.0	68.4	0.0
14AtAT	36b	83.0	14.4	88.4	87.2	51.3	0.0	67.9	66.0	4.2
32NtAT	36b	83.3	24.9	88.4	88.7	64.4	0.0	75.4	68.0	5.6
48CaAT	36b	100.0	23.4	100.0	100.0	97.7	7.5	71.7	86.2	36.4
49StAT	36b	100.0	22.3	100.0	95.0	97.9	9.1	68.4	86.1	29.6
54SIAT	36b	85.5	12.7	85.8	90.3	72.2	0.0	59.7	61.8	9.0
55SIAT	36b	100.0	25.7	100.0	100.0	91.9	6.1	72.5	82.7	41.2
56SIAT	36b	93.9	17.8	88.2	86.9	87.2	0.7	68.2	74.9	19.7
CASHT	36b	93.3	14.4	86.3	81.9	90.8	0.6	64.2	75.0	18.1
28BaAT	36b	0.0	0.0	0.0	0.0	0.0	0.0	55.8	0.0	0.0
53PhAT	36b	0.0	0.0	0.0	0.0	0.0	0.0	52.2	14.5	0.0
61SrAT	36b	0.0	0.0	0.0	31.2	0.0	0.0	42.2	0.0	0.0
62SrAT	36b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HolE	36b	0.0	0.0	0.0	0.0	0.0	0.0	40.4	0.0	0.0

[1] Peaks corresponding to the acetylated product were detected by mass spectrometry (positive electrospray), presumably as a result of a competing reaction with endogenous acetyl CoA present in the cell lysate.

14.2.3 N-Acyltransferase amine substrate promiscuity

A panel of 21 N-acyltransferases were screened towards amines **36c-36i**. A CoA ester substrate was selected for each N-acyltransferase (based on the results in example 14.2.2) which was produced *in situ* catalysed by a CoA ligase using catalytic CoA.

14.2.3.1 Method

CoA ester assay mix (Volume/well)

0.05 μ L CoA (50 mM in water, adjusted to pH ~5 with NaOH)

6 μ L ATP (50 mM in water)

12.5 μ L Acid solution (20 mM, made up in 50 mM Tris pH 8.0)

10 μ L CoA Ligase (0.5 mg/ml)

6.45 μ L buffer (50 mM Tris, pH 8.0)

CoA ester assay mix (37.5 μ L) and amine solution (20 mM, 12.5 μ L) were added to each well of a 96-well microtitre plate. Using the Biomek FX liquid handling robot, N-acyltransferase lysate (50 μ L) was added and reactions were incubated at 25 °C, 800 rpm, for 18 hours. The final reaction mix contained 2.5 mM Acid, 2.5 mM CoA, 5 μ g CoA ligase, 3 mM ATP, 2.5 mM Amine in Tris pH 8.0 with 50% acyltransferase lysate loading. The percentage conversion was calculated relative to standard curves of chemically synthesised amides and amine starting materials.

14.2.3.2 Results

A number of the N-acyltransferases, namely, 32NtAT, 48CaAT, 49StAT, 54SIAT, 55SIAT, 56SIAT, & CASHT were highly promiscuous and accepted a number of amine substrates. Other N-acyltransferases were more specific, for example 05PsAT, 07NfAT and 42GmAT had a preference for aniline analogues (**36g** table 17, **36a** table 16) and tryptamine (**36c**). Together the panel of N-acyltransferases were able to operate across broad chemical space. The enzymes were screened towards both enantiomers of **36d** and **36e**, those which accepted these amines were highly selective towards the (S)-enantiomer and no product formation was detected using the (R)-enantiomer. This demonstrates that this methodology can be used to perform kinetic resolutions.

Table 17. Activity of the N-acyltransferase panel towards amines **36c-i**

Enzyme	Acid	CL	Conversion (%)						
			36c	(S)-36d	(S)-36e	36f	36g	36h	36i
05PsAT	7	ipfF	0.0	0.0	0.0	0.0	0.0	60.7	0.0
07NfAT	7	ipfF	10.3	0.0	0.0	0.0	0.0	42.0	0.0
41GmAT	7	ipfF	0.0	0.0	0.0	0.0	0.0	0.0	0.0
42GmAT	7	ipfF	4.5	0.0	0.0	0.0	0.0	0.0	0.0
43GmAT	7	ipfF	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BcAT	7	ipfF	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23DmAT	37	RpCL	14.3	0.0	0.0	0.0	0.0	0.0	0.0
36PaAT	5	ipfF	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10HvAT	3	PhCL	12.9	0.0	0.0	0.0	0.0	0.0	0.0
11AtAT	3	PhCL	15.1	0.0	24.7	10.3	88.9	20.7	13.7
12HvAT	3	PhCL	23.3	0.0	0.0	0.0	0.0	0.0	0.0
14AtAT	3	PhCL	68.7	75.2	0.0	0.0	94.7	0.0	11.4
32NtAT	3	PhCL	75.0	77.4	0.0	1.6	96.6	0.0	13.5
48CaAT	3	PhCL	92.8	94.5	49.6	8.8	98.6	0.0	22.2
49StAT	3	PhCL	89.6	95.7	64.9	9.4	98.3	0.0	20.5
54SIAT	3	PhCL	55.6	32.9	0.0	1.4	94.9	0.0	10.7
55SIAT	3	PhCL	96.3	94.4	36.3	8.2	98.8	0.0	24.0
56SIAT	3	PhCL	78.1	92.8	0.0	3.7	96.0	0.0	17.9
CASHT	3	PhCL	95.7	84.3	0.0	2.2	95.5	0.0	9.0
53PhAT	16	PhCL	0.0	65.9	0.0	0.0	0.0	0.0	0.0
61SrAT	7	ipfF	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Example 15: Synthesis of tertiary amides using a CoA ligase and N-acyltransferase

The N-acyltransferase 11AtAT, which in nature catalyses amide formation between agmatine and cinnamoyl CoA, demonstrated activity towards secondary amine substrates. 11AtAT was screened for activity towards a small panel of amines (**36k-36o** shown in figure 14), using cinnamoyl CoA produced in situ, catalysed by a CoA ligase PhCL.

15.1 Method

Cleared cell lysates of 11AtAT and reactions were prepared according to the methods in example 14.1 and 14.2. CoA ester assay mix (37.5 µl containing cinnamic acid **3** as the acid substrate) and amine solution (20 mM, 12.5 µl) were added to each well of a 96-well microtitre plate. Using the Biomek FX liquid handling robot, N-acyltransferase lysate (50 µl) was added and reactions were incubated at 25 °C, 800 rpm, for 18 hours. The final reaction mix contained 2.5 mM Acid, 2.5 mM CoA, 5 µg CoA ligase, 3 mM ATP, 2.5 mM Amine in Tris pH 8.0 with 50% acyltransferase(11AtAT) lysate loading. The percentage conversion was calculated relative to standard curves of chemically synthesised amides and amine starting materials.

15.2 Results

11AtAT was active towards a number of secondary amines and yielded tertiary amide products with reasonable yields. The highest activity was achieved with N-methylbenzylamine **36k**. Not all secondary amines were accepted and no product formation was detected with proline methyl ester **36o**.

Table 18. Activity of N-acyltransferase 11AtAT towards secondary amines **36j-o**

Amine Substrate	Conversion (%)
36j	14
36k	41
36l	36
36m	12
36n	5
36o	0

CLAIMS

1. A method for synthesising a non-natural amide comprising:
 - a. reaction of a carboxylic acid with a naturally occurring CoA ligase or a variant thereof;
and
 - b. reaction of the product of step a with an amine in the presence of a naturally occurring acyltransferase or a variant thereof;

with the proviso that where the CoA ligase and acyltransferase are both naturally occurring, they are not derived from the same source species and do not act sequentially in a metabolic pathway; and

with the proviso that the non-natural product is not N-(E)-p-coumaroyl-3-hydroxyanthranilic acid or N-(E)-p-caffeoyl-3-hydroxyanthranilic acid.

2. A method according to claim 1, wherein steps a and b take place in vitro, and wherein step a takes place in the presence of ATP and CoA.
3. A method according to claim 2, wherein the product of step a is not purified prior to step b.
4. A method according to claim 2 or claim 3, wherein the naturally occurring CoA ligase or a variant thereof is attached to a solid substrate.
5. A method according to any one of claims 2 to 4, wherein the naturally occurring acyltransferase or a variant thereof is attached to a solid substrate
6. A method according to claim 1, wherein steps a and b take place in vivo in a host cell expressing the naturally occurring CoA ligase or variant thereof and the naturally occurring acyltransferase or variant thereof.
7. A method according to claim 6, wherein the host cell is a prokaryotic host cell.
8. A method according to any preceding claim, which further comprises a step of purification of the non-natural amide product.
9. A method for producing an active pharmaceutical ingredient, characterised in that the active pharmaceutical ingredient is produced from a non-natural amide intermediate that is synthesised as defined in any one of claims 1 to 8.
10. A host cell comprising either:
 - a. an expression vector comprising the coding sequence for a naturally occurring CoA ligase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and an expression vector comprising the coding sequence for a naturally occurring acyltransferase or a variant thereof operatively linked to regulatory control sequences capable of controlling expression of the naturally occurring acyltransferase or a variant thereof in the host cell; or
 - b. an expression vector having at least two expression cassettes, a first expression cassette comprising the coding sequence for naturally occurring CoA ligase or a

variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring CoA ligase or a variant thereof in the host cell, and a second expression cassette comprising the coding sequence for naturally occurring acyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the naturally occurring acyltransferase or a variant thereof in the host cell; or

- c. an expression vector having one expression cassette comprising the coding sequence for naturally occurring CoA ligase or a variant thereof fused in frame to a cleavable linker and then to the coding sequence for naturally occurring acyltransferase or a variant thereof operatively linked to conventional regulatory control sequences capable of controlling the expression of the fusion protein in the host cell;

with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring acyltransferase, the CoA ligase is not a 4-coumarate CoA ligase where the acyl transferase is a serotonin N-hydroxycinnamoyl transferase; and with the proviso that where the expression vector of (b) comprises the coding sequence for a naturally occurring CoA ligase and a naturally occurring acyltransferase, the CoA ligase is not tobacco 4-coumarate CoA ligase 2 where the acyltransferase is the globe artichoke hct gene having accession number DQ104740 with the proviso that the expression vector of (c) does not comprise pepper serotonin N-hydroxycinnamoyltransferase fused in frame with the self-processing FDMV 2A sequence followed by 4-coumarate ligase 2 from *Arabidopsis thaliana*.

11. A host cell according to claim 10 which is a prokaryotic cell.

Figure 1

FT = flow through; W1-2 = wash fractions (20-25 mM imidazole); E1-2 = elution fractions (50 mM imidazole); E3-4 = elution fractions (100 mM imidazole); E5-6 = elution fractions (200 mM imidazole); E7-8 = elution fractions (400 mM imidazole)

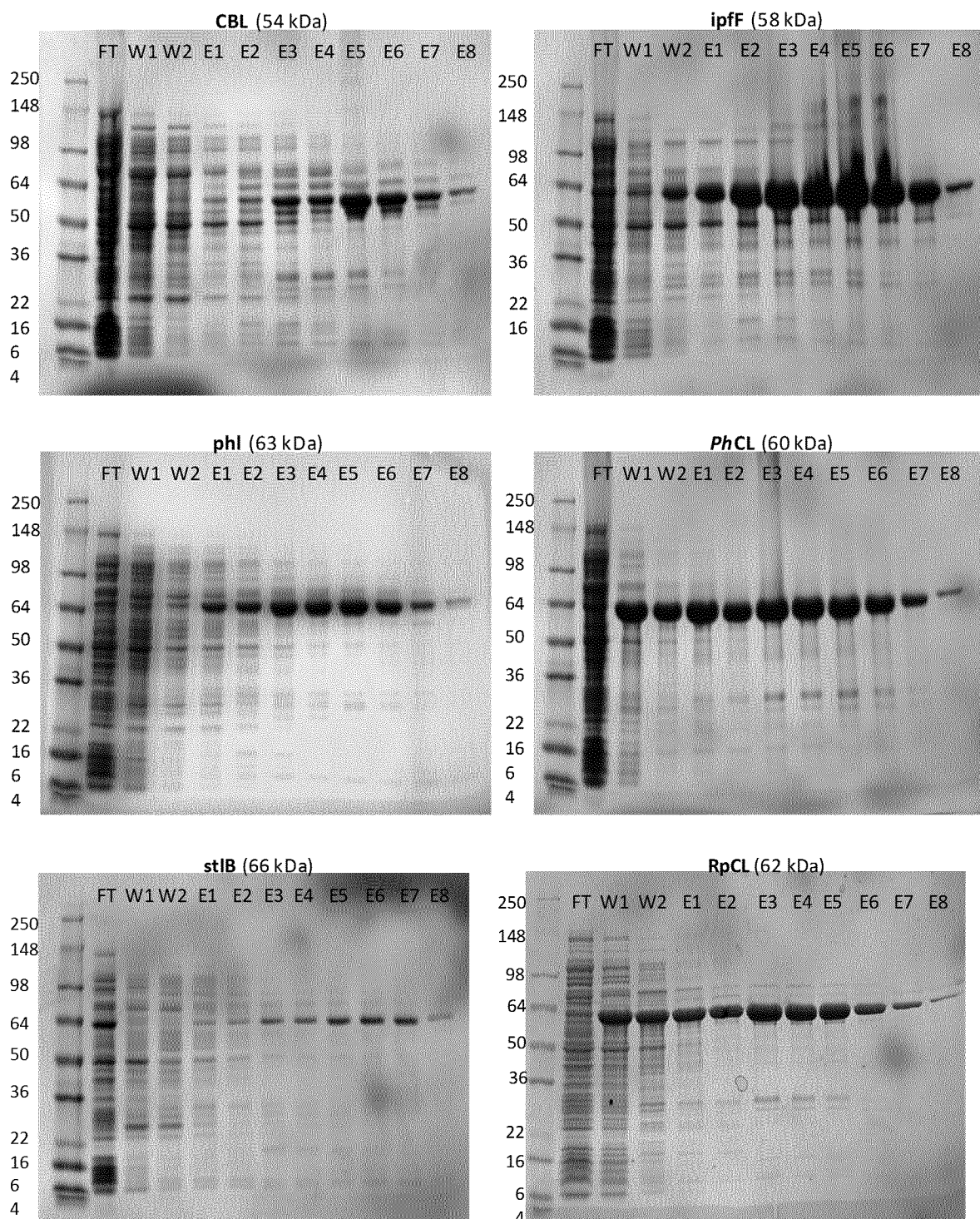


Figure 1 continued

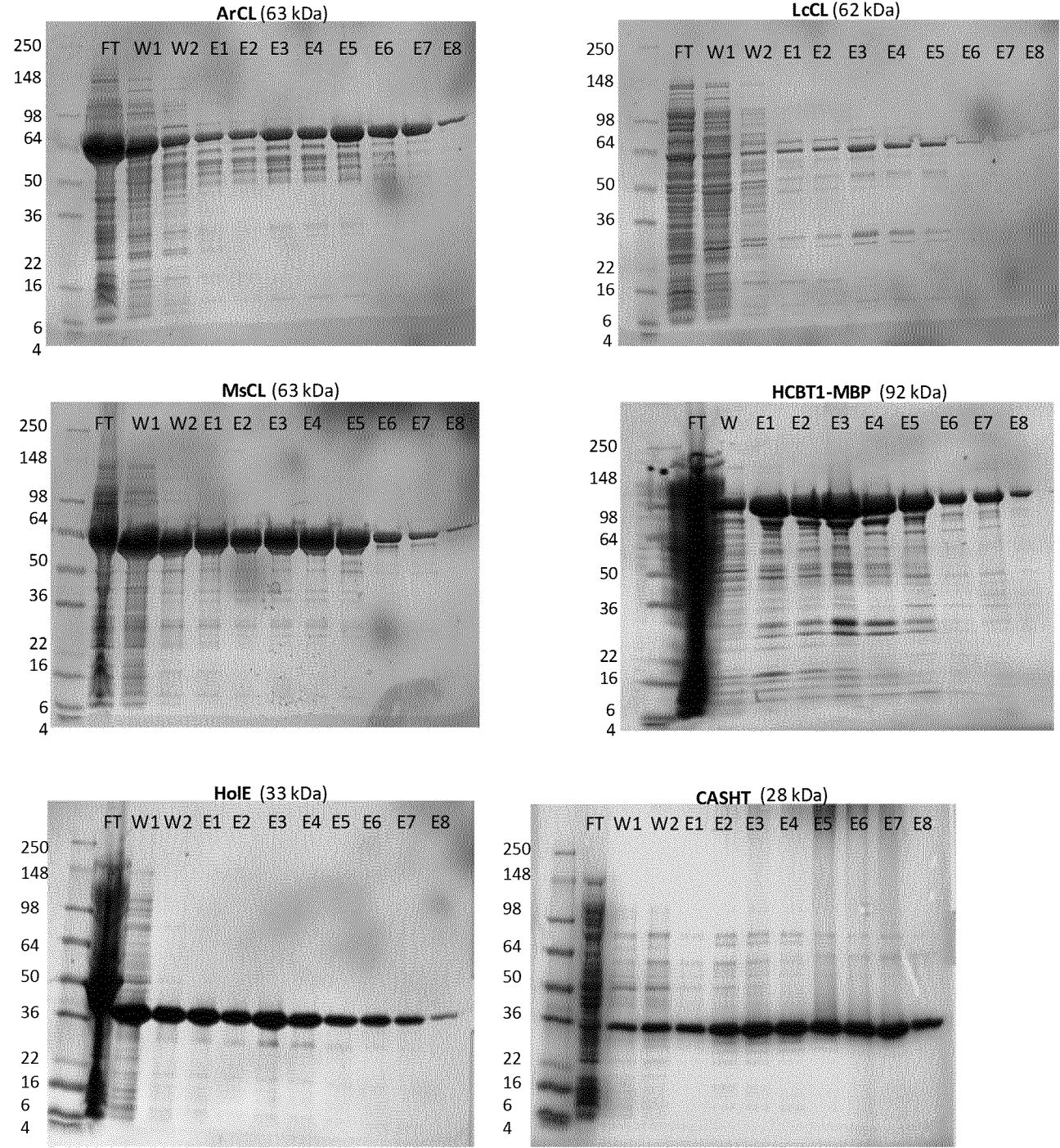
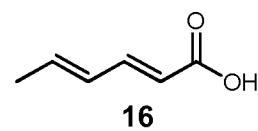
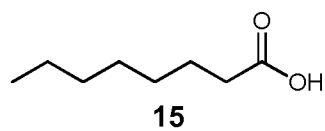
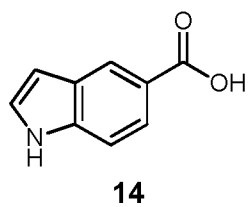
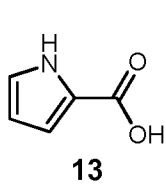
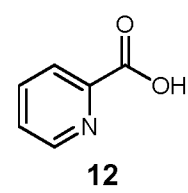
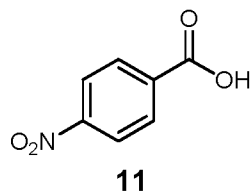
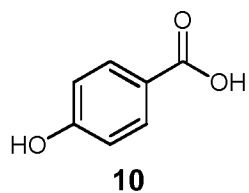
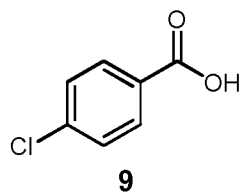
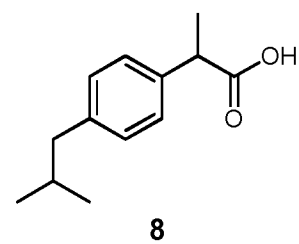
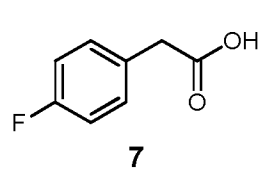
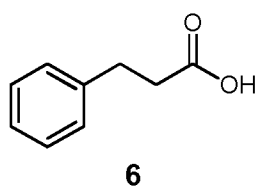
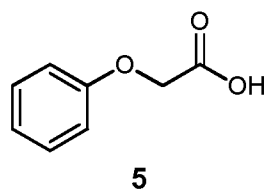
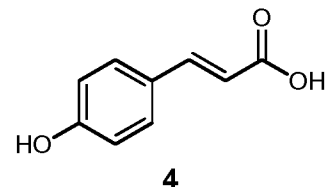
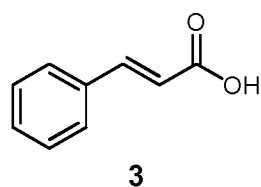
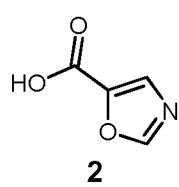
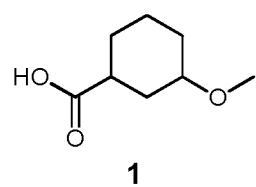


Figure 2



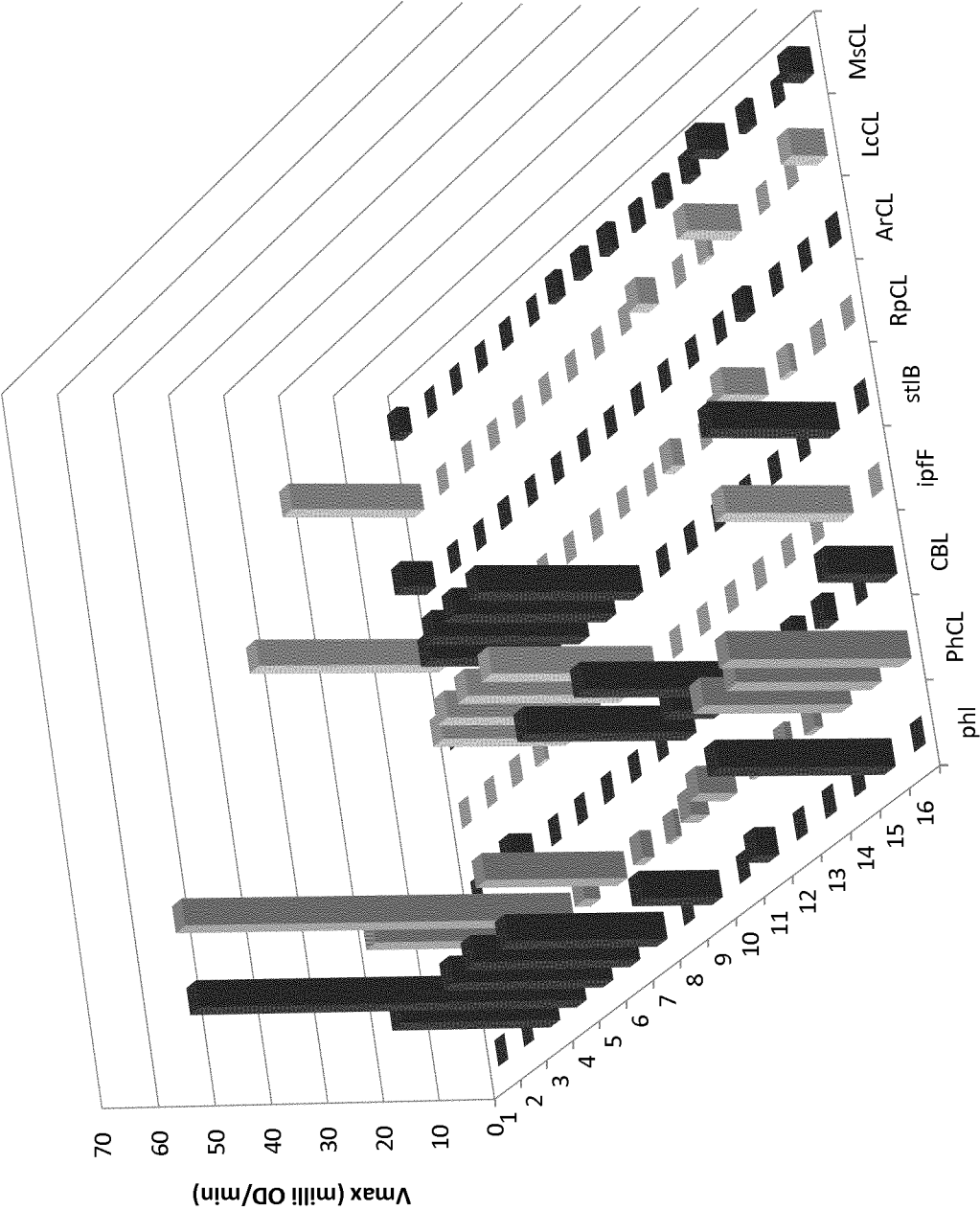


Figure 3

Figure 4

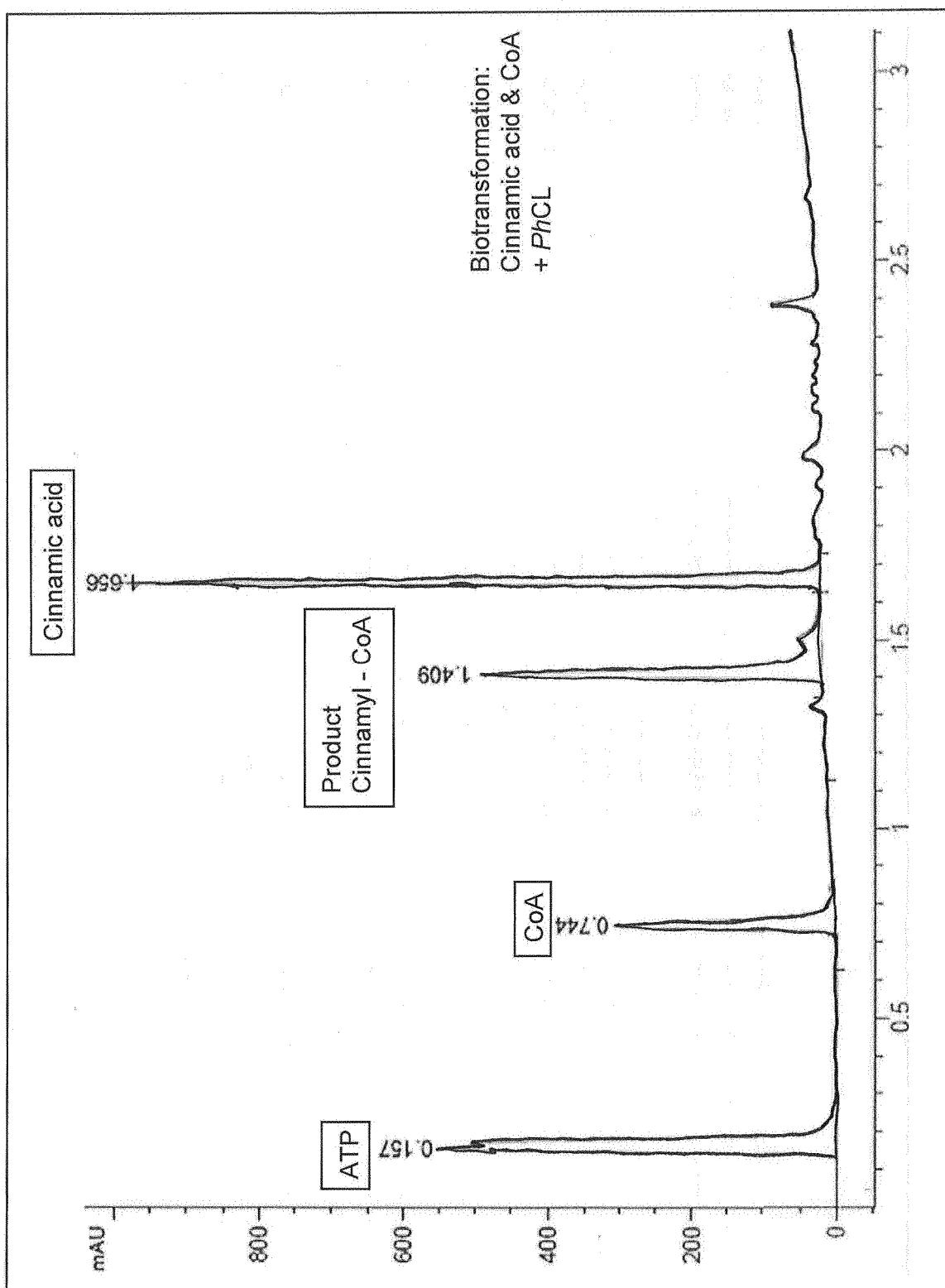


Figure 5

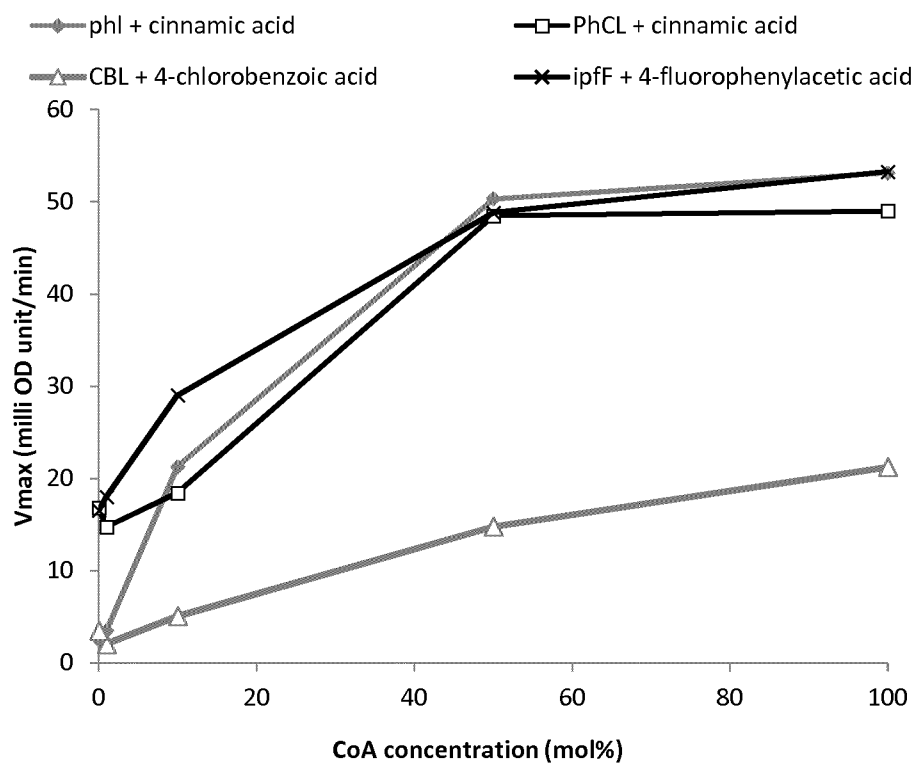


Figure 6

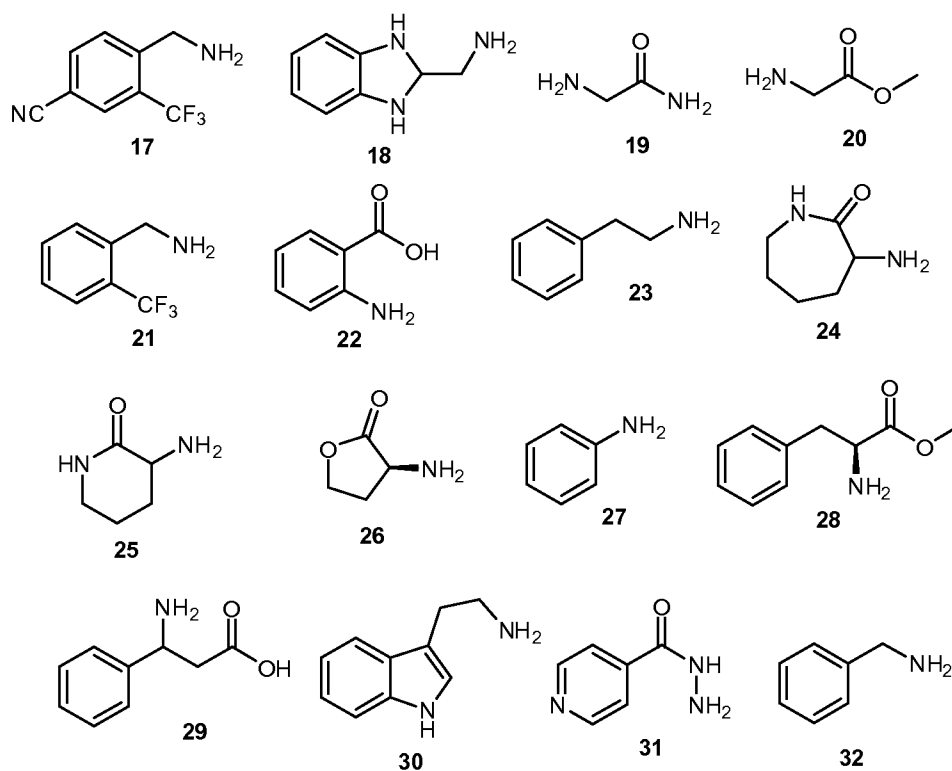


Figure 7

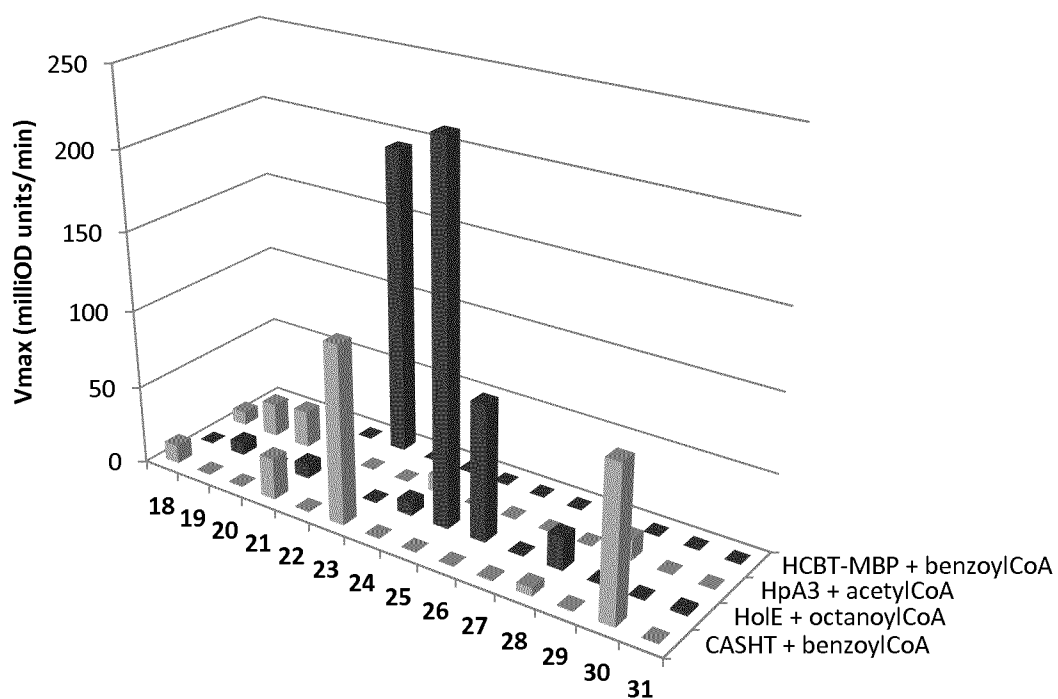


Figure 8

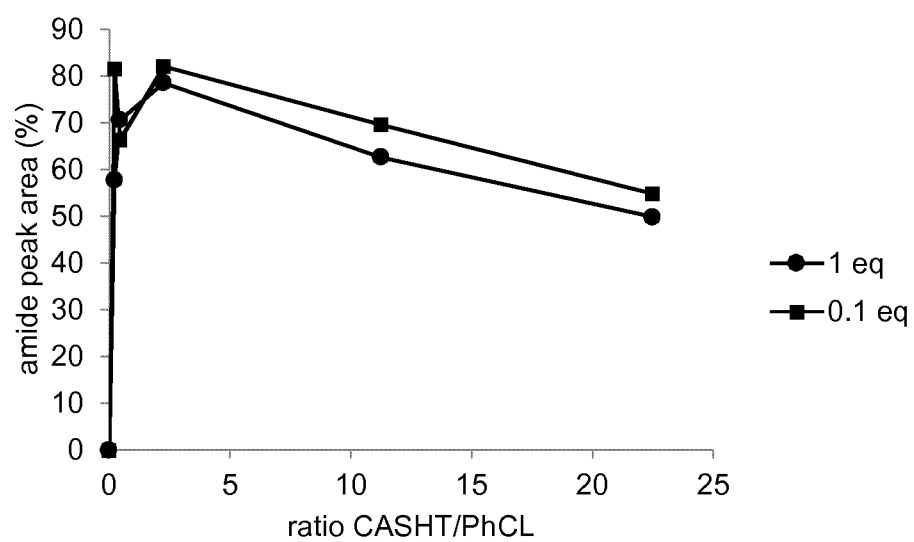


Figure 9

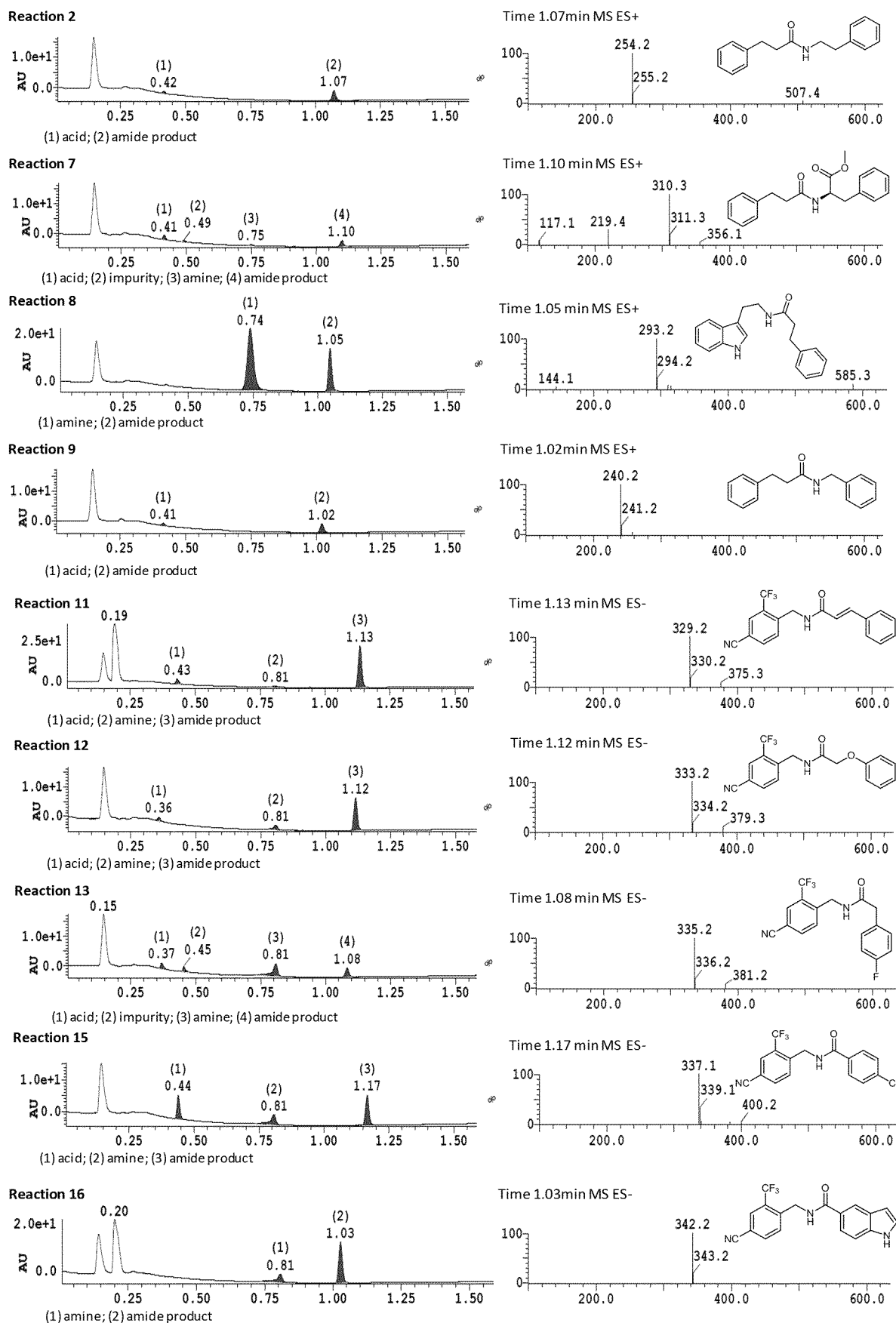


Figure 9 continued

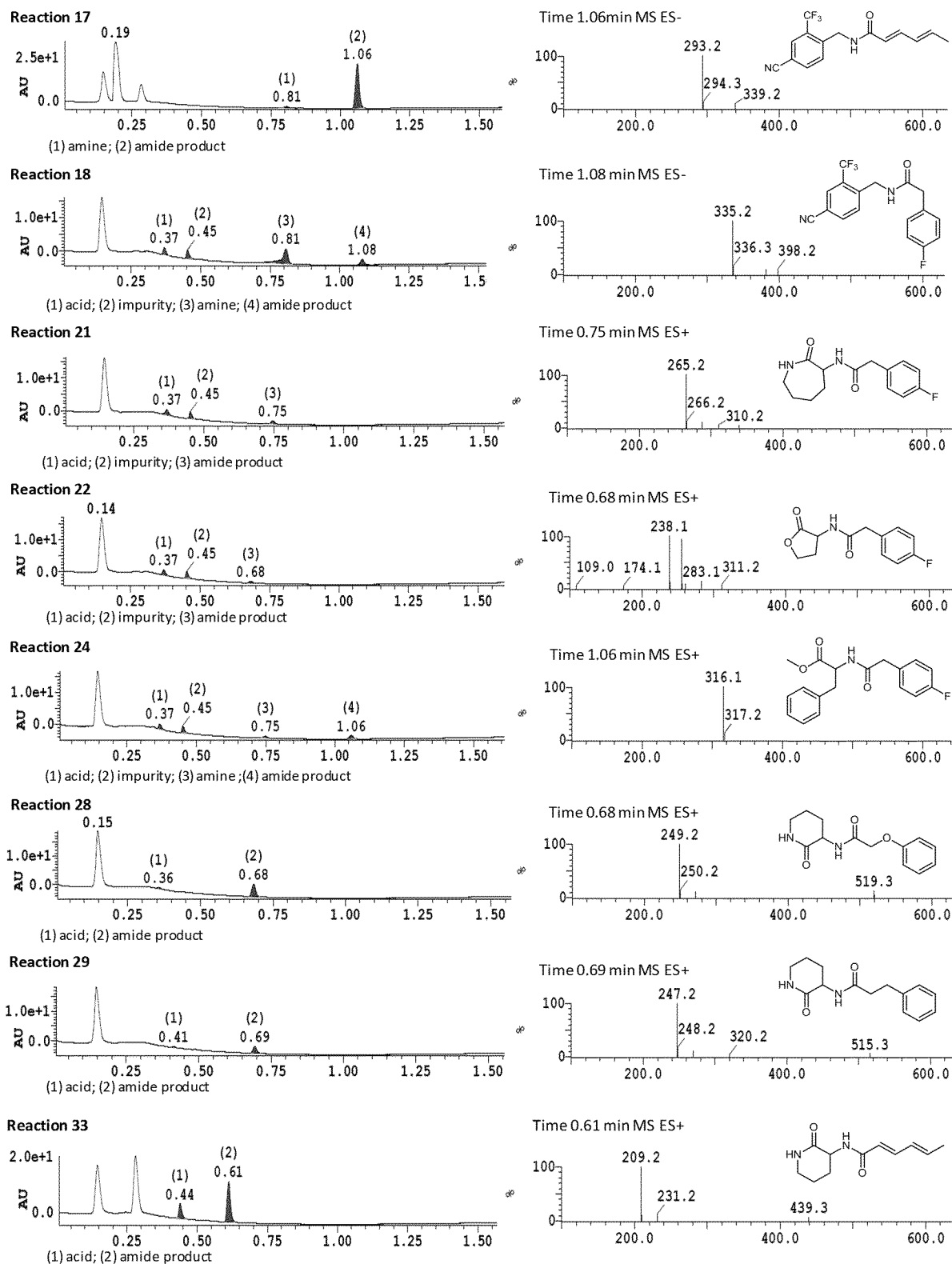


Figure 10

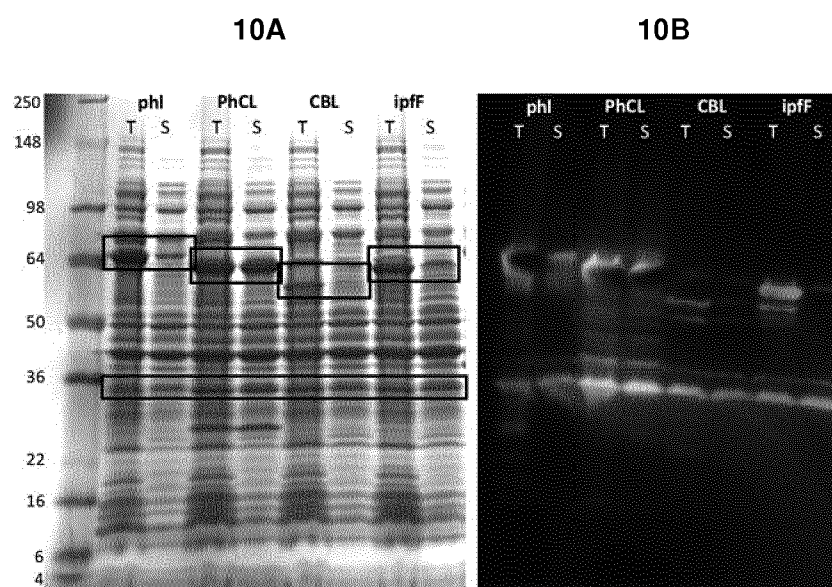


Figure 11

1) pCDF-Duet1-CASHT total cell lysate; 2) pCDF-Duet1-CASHT soluble fraction

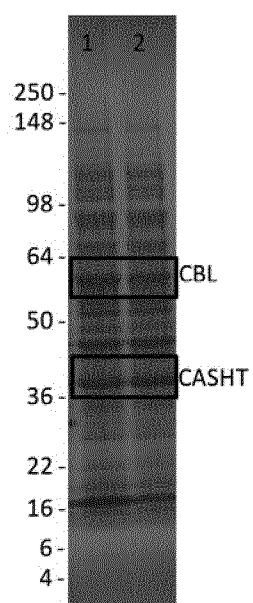


Figure 12

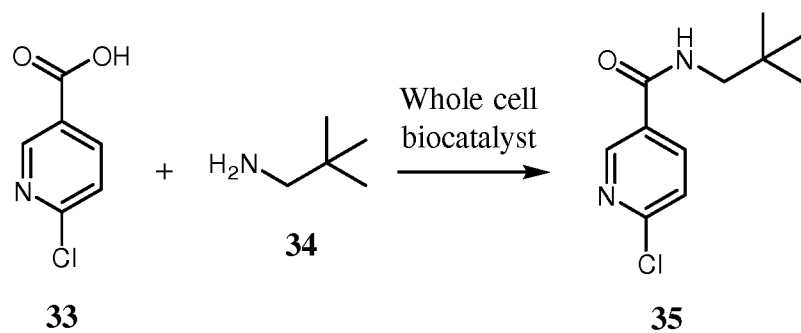


Figure 13

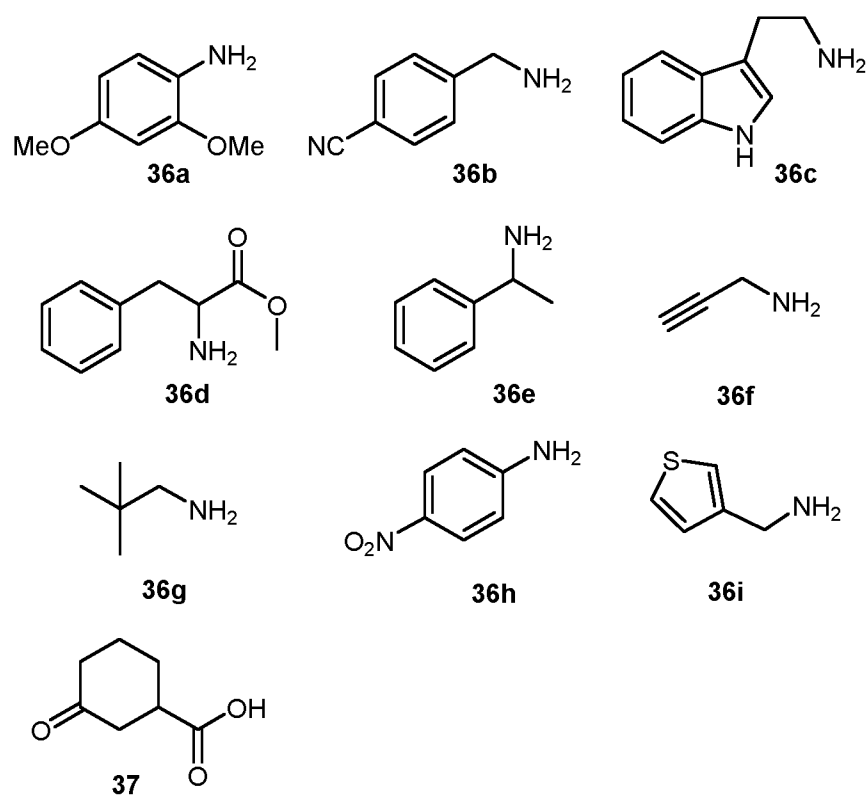
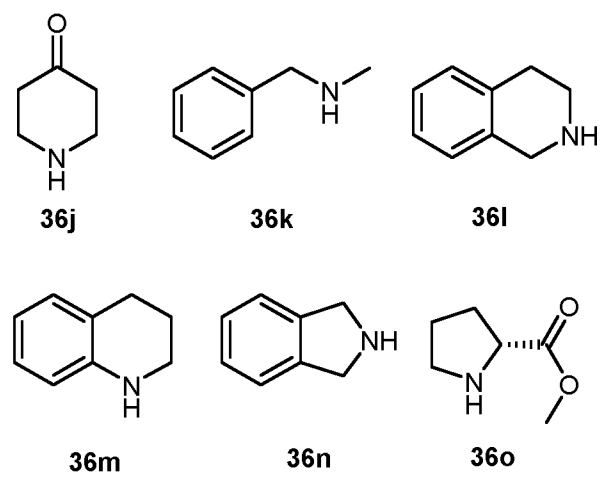


Figure 14



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/069773

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12P13/02 A61K31/155 C12N9/10 C12N9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12P A61K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013/078683 A1 (LOQUE DOMINIQUE [US] ET AL) 28 March 2013 (2013-03-28) [0008] - [0009], [0143], [0205], [0206] ----- -/-	1,3,6, 8-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 October 2017

Date of mailing of the international search report

02/11/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Schönwasser, D

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2017/069773

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	AYERICK EUDES ET AL: "Production of tranilast [N-(3',4'-dimethoxycinnamoyl)-anthranilic acid] and its analogs in yeast <i>Saccharomyces cerevisiae</i>", APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, SPRINGER, BERLIN, DE, vol. 89, no. 4, 24 October 2010 (2010-10-24), pages 989-1000, XP019880520, ISSN: 1432-0614, DOI: 10.1007/S00253-010-2939-Y page 991, column 1, paragraph 2 - column 2, paragraph 1; figures 1,5; tables 1,2 page 994, column 1, paragraph 2 - page 996, column 2, paragraph 1 page 989, column 2, paragraph 2 - page 990, column 2, paragraph 1 -----	1,3,6,8-10
X	DATABASE WPI Week 200974 2009 Thomson Scientific, London, GB; AN 2009-P04643 XP002774909, -& KR 2009 0097388 A (UNIV CHONNAM NAT IND FOUND) 16 September 2009 (2009-09-16) cited in the application abstract -----	1,3,6-8,10,11
X	KIYOON KANG ET AL: "Production of plant-specific tyramine derivatives by dual expression of tyramine N-hydroxycinnamoyltransferase and 4-coumarate:coenzyme A ligase in <i>Escherichia coli</i>", BIOTECHNOLOGY LETTERS, SPRINGER NETHERLANDS, DORDRECHT, vol. 31, no. 9, 29 May 2009 (2009-05-29), pages 1469-1475, XP019727618, ISSN: 1573-6776, DOI: 10.1007/S10529-009-0032-3 page 1469, column 2, line 24 - page 1470, column 1, line 2; figures 1,2 ----- -/--	1,3,6-8,10,11

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2017/069773

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ANDREA MOGLIA ET AL: "Evaluation of the bioactive properties of avenanthramide analogs produced in recombinant yeast : Bioactive Effects of Yeast Avenanthramides", BIOFACTORS., vol. 41, no. 1, 2 January 2015 (2015-01-02), pages 15-27, XP055416543, GB ISSN: 0951-6433, DOI: 10.1002/biof.1197 cited in the application -----	1-11

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/069773

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013078683	A1	28-03-2013	NONE

KR 20090097388	A	16-09-2009	NONE
