**N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids**

Robert S. Jansen*, Ruben Addie*, Remco Merkx*, Alexander Fish†, Sunny Mahakena*, Onno B. Bleijerveld†, Maarten Altelaar†, Lodewijk Ilst†, Ronald J. Wanders*, P. Borsta†, and Koen van de Wetering†

Divisions of aMolecular Oncology, bCell Biology II, and cBiochemistry, and dProteomics Facility, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands; and eLaboratory of Genetic Metabolic Diseases, Academic Medical Center, 1105 AZ Amsterdam, The Netherlands

Edited by David W. Russell, University of Texas Southwestern Medical Center, Dallas, TX, and approved April 20, 2015 (received for review December 24, 2014)

Despite technological advances in metabolomics, large parts of the human metabolome are still unexplored. In an untargeted metabolomics screen aiming to identify substrates of the orphan transporter ATP-binding cassette subfamily C member 5 (ABCC5), we identified a class of mammalian metabolites, N-lactoyl-amino acids. Using parallel protein fractionation in conjunction with shotgun proteomics on fractions containing N-lactoyl-Phe-forming activity, we unexpectedly found that a protease, cytosolic nonspecific dipeptidase 2 (CNDP2), catalyzes their formation. N-lactoyl-amino acids are ubiquitous pseudodipeptides of lactic acid and amino acids that are rapidly formed by reverse proteolysis, a process previously considered to be negligible in vivo. The plasma levels of these metabolites strongly correlate with plasma levels of lactate and amino acid, as shown by increased levels after physical exercise and in patients with phenylketonuria who suffer from elevated Phe levels. Our approach to identify unknown metabolites and their biosynthesis has general applicability in the further exploration of the human metabolome.

unknown metabolites | untargeted metabolomics | ABCC5 | MRPS | physical exercise

Untargeted metabolomics aims to provide a comprehensive snapshot of the metabolome and is becoming a mainstream technique to discover biomarkers, to study the effects of interventions, and to discover the function of enzymes (1).

Recent technical improvements now make it possible to detect several thousand metabolites in a single untargeted metabolomics analysis, but the identity of these metabolites is initially not known beyond their molecular mass (1). Knowing the chemical identity of metabolites is crucial for the proper interpretation of metabolomic studies, however. Online metabolite databases like METLIN (~240,000 entries) and the Human Metabolome Database (HMDB; 42,000 entries) contain vast numbers of metabolites and are extremely useful to annotate the detected molecular masses (2, 3). These databases cover large parts of the metabolome, but significant gaps remain. Due to poorly characterized and promiscuous enzymes, the human metabolome is much larger than initially anticipated (4). State-of-the-art untargeted metabolomics studies still report up to 40% unidentified metabolites, indicating that large parts of the human metabolome are still unexplored. We identified an uncharacterized class of ubiquitous mammalian metabolites: N-lactoyl-amino acids. Using a powerful combination of proteomics and protein fractionation, we unexpectedly discovered that these metabolites are formed from lactate and amino acids by reversed action of the protease cytosolic nonspecific dipeptidase 2. Our approach to identify unknown metabolites and their biosynthesis has general applicability in the further exploration of the human metabolome.

Significance

Untargeted metabolomics is rapidly becoming a mainstream technique to discover biomarkers, to study the effects of interventions, and to discover the function of enzymes. Several thousand metabolites can be detected in a single untargeted metabolomics analysis, but state-of-the-art untargeted metabolomics studies still report up to 40% unidentified metabolites, indicating that large parts of the human metabolome are still unexplored. We identified an uncharacterized class of ubiquitous mammalian metabolites: N-lactoyl-amino acids. Using a powerful combination of proteomics and protein fractionation, we unexpectedly discovered that these metabolites are formed from lactate and amino acids by reversed action of the protease cytosolic nonspecific dipeptidase 2. Our approach to identify unknown metabolites and their biosynthesis has general applicability in the further exploration of the human metabolome.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: k.vd.wetering@nki.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424638112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1424638112

PNAS | May 26, 2015 | vol. 112 | no. 21 | 6601–6606
Untargeted Metabolomics Reveals Unknown ABC55-Related Metabolites. To identify novel endogenous substrates of ABC55, we applied comparative untargeted metabolomics to culture medium from HEK 293 parental cells and HEK 293 cells overexpressing human ABC55 (HEK 293/ABC55). ABC55 substrates were expected to accumulate in HEK 293/ABC55 medium compared with HEK 293/medium (Fig. 1), although intracellular levels were generally unaffected (Fig. S1).

Using the accurate mass and isotope pattern of the unknowns, we calculated their chemical formulas, which were used to search metabolite databases. This search did not yield candidate identities because the putative ABC55 substrates were not present in the large online metabolite databases like METLIN (2) or HMDB (3). Notwithstanding their absence in databases, we also detected the unknowns in many mouse tissues.

Unknown Metabolites Are N'-Lactoyl-Amino Acids. To elucidate the structure of the unknowns, we purified one of the unknown metabolites (C12H15NO4) from a large volume of HEK 293/ABC55-conditioned culture medium. After purification using C18 flash chromatography and ion-pair C18 and phenyl HPLC, we obtained C12H15NO4 in a sufficient amount and purity for NMR analysis. The 1H-NMR, 13C-NMR, and 1H-COSY NMR spectra indicated that the unknown was N'-lactoyl-Phe [2-(2-hydroxypropanamido)-3-phenylpropanoic acid] (Fig. 2).

To bypass this loss of activity over time, we applied size exclusion chromatography in parallel and immediately determined the enzymatic activity in the fractions. We analyzed active fractions and neighboring inactive control fractions by LC/MS shotgun proteomics to identify proteins that coeluted with enzymatic activity (Fig. 4B). Although the fractions still contained between 100 and 600 proteins each, only a single protein, cytosolic nonspecific dipeptidase 2 (CNDP2), was present in all active fractions and absent in the negative fractions (Fig. 4C). We confirmed that CNDP2-mediated N-lac-Phe formation, using recombinant human CNDP2 (Fig. 4D). Curve fitting revealed an equilibrium constant $K = 3.1 \times 10^{-2} \cdot M^{-1}$. A very similar equilibrium constant (2.7 $\times 10^{-2} \cdot M^{-1}$) was obtained by incubating CNDP2 with different concentrations of lactate and Phe for 24 h (Fig. S5).

N'-Lactoyl-Amino Acids Are Formed by Cytosolic Nonspecific Dipeptidase 2. To identify the source of the N'-lactoyl-amino acids, we first tested under which conditions these metabolites are formed in vitro. Using whole-cell lysate from HEK 293 cells, we found that N-lac-Phe is formed when both lactate and Phe are incubated in the presence of nonadenatured proteins, indicating enzymatic formation (Fig. 4/). Attempts to purify the enzyme from whole-cell lysate by classical sequential protein fractionation were unsuccessful. Although enzyme activity was detected in fractions obtained by size exclusion chromatography, this activity was lost upon subsequent conversion of active fractions by strong anion exchange chromatography.

To determine the uptake of ABC55-dependent transport accurately at higher substrate concentrations. Substantial background transport was also observed in culture medium of HEK 293 control cells (Fig. 1), which indicates that other carriers exist for this class of metabolites.

N'-Lactoyl-Amino Acids Are Rapidly Formed in HEK 293 Cells. We assessed the speed at which N'-lactoyl-amino acids are formed in vitro by metabolic labeling with 13C6-glucose. The 13C6-glucose is rapidly taken up into cells and converted into 13C3-lactate, which can then be incorporated into N-lactoyl-amino acids. Fig. 5 shows that the rate of 13C3-labeling of N-lac-Phe was comparable to formation of 13C3-lactate, its immediate precursor. Similar data were obtained for other N-lactoyl-amino acids. These results indicate very rapid interconversion between amino acids, lactate, and N-lactoyl-amino acids in living cells.

N'-Lact-Phe Is Transported by ABC55 in Vitro. The metabolic screen in which we detected the N-lactoyl-amino acids was started to discover new endogenous substrates of ABC55. To confirm that N-lactoyl-amino acids are genuine ABC55 substrates, we studied the uptake of N-lac-Phe into inside-out membrane vesicles. Fig. S4 shows that N-lac-Phe is taken up into vesicles in a time-, ATP-, and ABC55-dependent fashion, with a $K_m$ of 1 mM when fitted to Michaelis–Menten kinetics (Fig. 3B). This estimate is obviously very rough, because considerable background transport into the vesicles made it increasingly difficult to determine the relatively low rate of ABC55-dependent transport accurately at higher substrate concentrations. Substantial background transport was also observed in culture medium of HEK 293 control cells (Fig. 1), which indicates that other carriers exist for this class of metabolites.
N-Lactoyl-Amino Acid Levels in Human Plasma Depend on the Concentration of Lactate and Amino Acid. In humans, lactate levels swiftly rise during strenuous physical exercise when insufficient oxygen is available for aerobic glycolysis. To test whether an increase in lactate results in increased N-lactoyl-amino acid levels, we determined the levels of lactate, amino acids, and their corresponding N-lactoyl-amino acids in plasma of six healthy individuals before and immediately after a single bout of strenuous exercise lasting 5–10 min. In line with the rapid formation in vitro, we found that the increased lactate levels coincided with significantly increased levels of N-lactoyl-amino acids in human plasma, whereas amino acid levels remained the same (Fig. 6). Patients suffering from the inborn metabolic disorder phenylketonuria (PKU) lack functional Phe hydroxylase and are unable to convert Phe into Tyr. As a result, patients with PKU have increased Phe plasma levels (28). We compared plasma samples of a group of Dutch patients who had PKU with high plasma Phe with control samples and found N-lac-Phe levels that were significantly higher than the N-lac-Phe levels in control samples, whereas lactate levels were comparable (Fig. 6B).

Theoretically, the N-lac-Phe concentration in cells is determined by the concentration of lactate and Phe, and the equilibrium constant K. In Fig. 6C, we plotted the data from the exercise experiment, the patients with PKU and controls, and an additional set of plasma samples with high lactate levels due to prolonged storage as whole blood. Except for a single outlier, the product of the lactate and Phe plasma concentrations showed a good correlation with the plasma N-lac-Phe concentration. The slope of the fitted linear function represents an apparent equilibrium constant $K = 7.4 \times 10^{-8} \text{M}^2$, which is in good agreement with the value of $3.1 \times 10^{-8} \text{M}^2$ obtained for CNDP2-mediated N-lac-Phe formation in vitro. Formation of N-lac-Phe required cells, because it was not formed when we incubated plasma with lactate or Phe levels comparable to the lactate or Phe levels found during exercise and in patients with PKU. A modest but significant increase was observed after a 30-min incubation in whole blood (Fig. S7).

Discussion

We have identified an uncharacterized class of mammalian metabolites, N-lactoyl-amino acids. These metabolites are present in many tissues and can approach micromolar concentrations in human plasma. We have also shown that CNDP2 can form N-lactoyl-amino acids in HEK 293 cells.

CNDP2, which is also known as carboxypeptidase of glutamate-like (CPGL) and as carnosine dipeptidase II, belongs to the M20 family of metallopeptidases and requires Mn$^{2+}$ for full activity (29). Loss of Mn$^{2+}$ from the enzyme is likely the reason why activity was gradually lost during our initial attempts to isolate the enzyme. CNDP2 is highly conserved across species and is expressed in most tissues (29). As its name indicates, CNDP2 has broad substrate selectivity for dipeptides, but its physiological substrates are not known (29, 30). Cndp2 KO mice do not exist. Several groups reported a role for CNDP2 in tumor suppression, but similar effects were observed with the isofrom CPGL-B, which lacks a part of the catalytic domain (31–33). It is unknown whether CPGL-B has any peptidase activity.

Fig. 3. N-lac-Phe is transported into inside-out membrane vesicles by ABCCS. (A) Control vesicles (○) and ABCCS-containing vesicles (●) were incubated with 100 μM N-lac-Phe at 37 °C in the presence (solid line) and absence (dashed line) of 5 mM ATP. At the indicated time points, a sample containing 75 μg of protein was taken. After washing over a filter, the vesicular content was analyzed by LC/MS (n = 3–6). (B) Concentration dependence was assessed by incubating control and ABCCS-containing vesicles with several concentrations of N-lac-Phe in the presence of ATP and determining ABCCS-dependent uptake after 2 min (n = 3–4 for concentration <1,000 μM and n = 8 for concentration ≥1,000 μM). The data were fitted to Michaelis–Menten kinetics (solid line) using GraphPad Prism. Data are presented as mean ± SEM.

Fig. 4. N-lac-Phe is formed by CNDP2. (A) N-lac-Phe is formed in the presence of Phe, lactate, and intact protein. Whole-cell lysate or control buffer (25 mM Tris-HCl [pH 7.4]) was incubated for 30 min at 37 °C in the presence or absence of 10 mM substrates. N-lac-Phe formation was determined by LC/MS and is expressed in arbitrary units. Data are presented as mean (n = 3) plus SD. (B) Whole-cell lysate was fractionated in parallel on three different columns. Enzyme activity was assessed by incubation with 10 mM lactate and Phe (30 min, 37 °C) and is normalized to the activity in unfraccionated whole-cell lysate. The levels of CNDP2 were determined in active fractions and neighboring inactive control fractions (all marked by an asterisk) using LC/MS proteomics and are expressed as a peptide spectrum match (PSM) (n = 1). CV, column volume. (C) Although the CNDP2 activity profile for each single fractionation, only a single protein, CNDP2, coeluted with activity in all three fractionations. SAX, strong anion exchange; SCX, strong cation exchange; SEC, size exclusion chromatography. (D) Human recombinant CNDP2 (1 μg) was incubated (37 °C) with 10 mM lactate and Phe in 25 mM Tris-HCl (pH 7.4) containing 0.1 mM MnCl$_2$. N-lac-Phe levels were determined by LC/MS and are expressed as arbitrary units. Data are presented as mean (n = 3) plus SD. Additional enzyme kinetics are presented in Figs. S5, S8, and S9.
The finding that a peptidase was responsible for the biosynthesis of our pseudopeptides was unexpected, because peptidases normally hydrolyze peptide bonds. Although peptidases are known to mediate the formation of peptides through reverse proteolysis, this property has mainly been investigated by organic chemists, who exploit the selectivity and mild reaction conditions for peptide synthesis (34, 35). Peptidases are catalysts, and, as such, they do not change the equilibrium constant between peptide formation and hydrolysis (34). In an aqueous environment, the equilibrium constant strongly favors peptide hydrolysis over peptide formation. Although formation of N-lactoyl-amino acids from lactate is thermodynamically very unfavorable, the formation from lactic acid esters is not. To rule out the possibility that the N-lactoyl-amino acids we detected were formed from traces of lactic acid esters present in our commercial lactic acid preparation (≥99% purity), we tested the rate of CNDP2-mediated N-lac-Phe formation from methyl-lactate and found that it is a relatively poor CNDP2 substrate (Fig. 8). Lactic acid esters are therefore a highly unlikely precursor for N-lac-Phe. Importantly, we found that N-lac-Phe is quickly hydrolyzed by CNDP2, like its model substrate Cys-Gly (Fig. 8), one of the best CNDP2 substrates identified thus far (30).

Because peptide bond hydrolysis is strongly favored over peptide bond formation in an aqueous environment, reverse proteolysis is considered to be negligible in vivo by many (34), although some have proposed a role for reverse proteolysis in antigen formation (36). The apparent equilibrium constants of 3.1 × 10⁻² M⁻¹, 2.7 × 10⁻² M⁻¹, and 7.4 × 10⁻² M⁻¹ calculated from our data are higher than the values reported for reverse proteolysis of regular amino acids (10⁻³ to 10⁻⁴ M⁻¹) (34). As a result, the high intracellular levels of lactate and amino acids, which are in the high micromolar to millimolar range, allow N-lactoyl-amino acid levels in plasma to approach micromolar levels (Fig. 6C). Contrary to common conviction, our data show that reverse proteolysis is not negligible in vivo, as long as the substrates are present in considerable concentrations.

Because lactate, amino acids, and CNDP2 are ubiquitous, N-lactoyl-amino acids are present in many tissues. Interestingly, N-lac-Val, N-lac-Leu, and N-lac-Ile have been identified in the urine of a patient with elevated branched-chain amino acids due to maple syrup urine disease (37). Another lactic acid amide, N⁴-lac-Lys, has been detected in human plasma and was found to be increased in the plasma of hemodialytic patients (38). As opposed to the biosynthesis described here, N⁴-lac-Lys is an amide-advanced glycation end product, formed non-enzymatically through the Maillard reaction (39). Although similar, N⁴-lac-Lys does not belong to the class of N²-lactoyl-amino acids identified here, because its lactate moiety is conjugated to the N⁰-amino group in the Lys side chain.

N²-lactoyl-amino acids have recently also been identified in soy sauce (40) and Parmigiano–Reggiano cheese (41). In cheese, N-lactoyl-amino acids were found to be formed by an enzyme present in lactic acid bacteria (42, 43). Although the enzyme was not identified, Bottesini et al. (43) showed that N-lac-Phe could be formed in vitro by a yeast peptidase, carboxypeptidase Y. Of note, all except one of the N-lactoyl-amino acids reported thus far contain a hydrophobic amino acid moiety, indicating enzyme selectivity.

The metabolic labeling experiments and in vitro enzyme kinetics show that interconversion between lactate, amino acids, and N-lactoyl-amino acids is very rapid. This fast interconversion explains why N-lactoyl-amino acid levels, except those N-lactoyl amino acid levels of N-lac-Trp, did not change intracellularly upon ABC5 overexpression: N-lactoyl-amino acids effluxed by ABC5 are continuously replenished from the large pool of lactate and amino acids.

The fast interconversion is also in line with the rapid increase of N-lactoyl-amino acids upon physical exercise. Although many studies have assessed the effect of strenuous exercise on the metabolite profile in blood, none identified the N-lactoyl-amino acids (44–46). Nevertheless, many of the studies report multiple unidentified exercise-dependent metabolites, most likely including N-lactoyl-amino acids.
We show that N-lac-Phe plasma levels are increased in patients with PKU with increased plasma Phe levels. The N-lac-Phe levels in plasma were highly correlated with the product of the lactate and Phe concentration. N-lac-Phe is of cellular origin, because it is not formed in plasma (Fig. S7), which apparently does not contain CNDP2. The formation of N-lac-Phe in whole blood, however, indicates that blood cells are also capable of forming N-lactoyl-amino acids, which is in line with the reported CNDP2 expression in peripheral blood leukocytes (29) and its likely presence in erythrocytes (47). The modest rate of formation we observed in whole blood cannot account for the quick formation during physical exercise, but it does explain the high levels of N-lac-Phe found in whole-blood samples that were stored at ambient temperature for prolonged times.

Using vesicular transport assays, we confirmed that ABCCS transports N-lac-Phe, and therefore most likely also transports the other N-lactoyl-amino acids. ABCCC5 transports N-lac-Phe with a low affinity (Kₘ of ∼1 mM), which is comparable to the Kₘ of the physiological ABCCS substrate diol (1 mM) (18). This Kₘ in the low millimolar range exceeds the submicromolar plasma and, presumably intracellular, levels by far.

Currently, N-lactoyl-amino acids do not have a known function. The chemical structure of lactate resembles a small amino acid-like Ala. Most likely, N-lactoyl-amino acids are nonspecific byproducts of the promiscuous peptidase CNDP2. Many enzymes promiscuously catalyze reactions, and even relatively specific enzymes will produce large absolute amounts of side products as soon as substrate levels and turnover rates are high (48, 49). Interestingly, CNDP1 was recently identified as an enzyme that degrades nonspecific dipeptide byproducts formed during carnosine synthesis (50). Even if N-lactoyl-amino acids are byproducts of CNDP2, they might represent useful extra-cellular biomarkers for intracellular amino acid concentrations, because they are only formed inside cells.

Our results underline that large parts of the metabolome remain to be discovered, even in a well-defined in vitro cell system (4). Untagged metabolomics provides an obvious method to address this gap in our knowledge. The parallel proteomics approach that we have developed to identify CNDP2 as the enzyme responsible for N-lactoyl amino acid formation is a powerful approach for the identification of any other enzyme of which the activity can be measured. We envision that this approach has broad applicability and is particularly useful for the identification of unstable enzymes that cannot be purified by classical sequential protein fractionation without losing activity.

Materials and Methods

Cell Lines and Culture Conditions. HEK 293/ABC5 (51) and control cells were cultured as described (52), seeded in six-well plates (Costar; Corning) at a density of 7.5 × 10⁶ cells per well, and allowed to grow overnight. At this point, the medium was replaced with 2 mL of fresh medium and the cells were incubated for 3 d. Culture medium was collected on ice, whereas cultured cells were washed with cold PBS and lysed in 1 mL of 70% (vol/vol) cold methanol.

Untargeted Metabolomics. Untargeted metabolomics was performed as described (52), but with a scan range of m/z 50-250 and m/z 250-1500 for the primary metabolic screen and a scan range of m/z 150-750 for all other experiments. Differentially present metabolites were detected with the XCMS platform (53). The accurate mass of differentially present compounds was queried in the online HMDB (3) and METLIN (2) databases. The accurate mass and isotope distributions were used to determine the chemical formula with SIRIUS software (54).

Purification of C₁₂H₁₅NO₄ from Culture Medium. The unknown C₁₂H₁₅NO₄ was purified from 1 L of HEK 293/ABC5 medium in four consecutive chromatography steps that are specified in SI Materials and Methods. Fractions containing C₁₂H₁₅NO₄ were pooled, freeze-dried three times to remove ammonium acetate and water residues, and dissolved in D₂O for NMR analysis.

Structure Elucidation. NMR spectra [1H-NMR, 13C-NMR (attached proton test), and 1H-COSY] were determined in D₂O using a Bruker Avance 300 spectrometer (1H, 300 MHz; 13C, 100 MHz) at 298 K. Fragmentation spectra (MS² and MS³) were acquired at a normalized collision energy of 35.

N-Lac-Phe Synthesis. Reference N-lac-Phe was synthesized from l-Phe ethyl ester and (−)-ethyl-l-lactate by carbodimide/l-hydroxybenzotriazole coupling as described in SI Materials and Methods. N-lac-Phe was obtained as a white foam (783 mg, 66%) of which the chemical structure was confirmed by NMR and MS (SI Materials and Methods).

Vesicular Transport. Inside-out membrane vesicles were prepared from HEK 293 cells transiently transduced with human ABCCS or GFP (control) as described (55). Transport of N-lac-Phe into vesicles was assessed in a 96-well format (27). Filtered vesicles (75 μg of protein) were lysed, evaporated to dryness, and reconstituted in 10 μL of mobile phase A before analysis by LCMS (Materials and Methods, Untargeted Metabolomics).

Enzyme Activity Assay. HEK 293 cell lysate [in 25 mM Tris HCl (pH 7.4) containing 1 mM DTT] was clarified by incubation with protamine sulfate (1%, 30 min at 4°C) and stored at −80°C. Lysate containing ∼50 μg of protein was incubated (30 min at 37°C) in a final volume of 500 μL of 25 mM Tris HCl (pH 7.4), with or without 10 mM l- (+)-lactate (≥99% pure; Sigma) or l-Phe (Sigma). Recombinant human CNDP2 (1 μg; C-terminal His-tagged; Prospec) was incubated (37°C) with l-lactate and l-Phe in 25 mM Tris HCl (pH 7.4), containing 0.1 mM MnCl₂. The reactions were quenched by the addition of 50 μL of acetic acid, followed by centrifugation (5 min, 21,800 × g, 4°C). The supernatant was applied to a Strata-X solid phase extraction column (33-μm polymeric reversed phase, 30 mg/mL; Phenomenex) to remove the lactate, after which N-lac-Phe levels were determined by LCMS, as described in Materials and Methods, Untargeted Metabolomics.

Protein Fractionation and Proteomics. Samples of cell lysate containing ∼250 μg of total protein were fractionated on an AKTAmicro system (GE Healthcare) using a Superdex 200 5/150 size exclusion column, a mini Q PC 3.2/3 strong anion exchange column, or a mini S 5/30 2.3/2 strong cation exchange column (all from GE Healthcare), as detailed in SI Materials and Methods. The volume corresponding to 20% of the collected fractions of interest was immediately used to test enzyme activity.

Tryptic digests of selected fractions were analyzed using nano-LC coupled to a Thermo Orbitrap Fusion mass spectrometer (Thermo Scientific) as detailed in SI Materials and Methods.

13C-Glucose Metabolic Labeling. HEK 293 cells were seeded in poly-o-lys-coated six-well plates (Corning) and grown to confluence. At time 0, the medium was replaced with 1 mL of DMEM without pyruvate and glucose, with 10% (vol/vol) FCS and 4.64 g/mL 13C₆-glucose (Aldrich). After 0, 1, 2, 4, 8, and 24 h, the culture medium and cells were collected and processed as described in SI Materials and Methods. N-lactoyl-amino acids were detected as described in Materials and Methods, Untargeted Metabolomics. Lactate labeling was determined by LCMS, using a ZIC-HILIC column (150 × 0.5 mm, 3.5 μm; SeQuant), as detailed in SI Materials and Methods.

Human Exercise and N-Lactoyl-Amino Acids. We collected blood samples from six human volunteers immediately before and after a single bout of strenuous exercise lasting 5–10 min. Blood was collected after informed consent in lithium heparin vacutainer tubes (BD) and immediately placed on ice. Within 15 min, the samples were centrifuged (10 min, 3,000 × g, 4°C), after which the plasma layer was collected and stored at −20°C. Plasma lactate levels were determined on a Cobas C501 analyzer (Roche), whereas amino acids were determined by ion-exchange chromatography with ninhydrin detection using an Aminocat JLC-500V analyzer (JEOL). To determine N-lactoyl-amino acid levels, 500 μL of plasma was mixed with 500 μL of water and 100 μL of acetic acid. The acidified plasma was processed and analyzed as described in Materials and Methods, Enzyme Activity Assay. Our study was sent to the Patient Trial Committee of the Netherlands Cancer Institute for review, and they determined that the study was exempt from approval.

PKU and N-Lactoyl-Amino Acids. Plasma samples (heparin) were collected from patients with PKU and controls with unrelated metabolic diseases as a part of routine clinical chemistry testing. Samples were anonymized before analysis. Patients with PKU were under dietary control, but samples were selected based on high plasma Phe levels. Amino acid levels were determined with ninhydrin detection, as described above.

Lactate was quantified in plasma using an ultraperformance LC/tandem MS instrument (Waters Quattro Premier XE) and a bridged ethyl hybrid-amide column (100 × 2.1 mm, 1.7 μm; Waters) by multiple reaction monitoring acquisition in the negative ionization mode. N-lactoyl-amino acid levels were determined as described for the human exercise plasma samples.
ACKNOWLEDGMENTS. We thank Liesbeth Hoekman and Enver Delic for their editorial assistance, and Alfred Schinkel and Huib Ovaa for a critical reading of our manuscript. We also thank an anonymous reviewer who made very helpful suggestions for improving the manuscript. Part of this work was performed within the framework of the project “Proteins at Work,” financed by the Netherlands Organization for Scientific Research as part of the National Roadmap Large-Scale Research Facilities of the Netherlands (Project 184.032.201).

**Supporting Information**

**Jansen et al. 10.1073/pnas.1424638112**

**SI Materials and Methods**

**Purification of C<sub>3</sub>H<sub>5</sub>N<sub>2</sub>O<sub>2</sub> from Culture Medium.** Approximately 1 L of HEK 293/ABCC5 culture medium was deep-extracted by adding 3 L of methanol, followed by centrifugation, rotary evaporation at reduced pressure, and freeze-drying. Purification was performed in four consecutive steps. The first separation was performed on a Sepacore Flash chromatograph (Buchi) using a 40-gauge C<sub>18</sub> Flash column (Screening Devices) and mobile phases consisting of 5 mM ammonium acetate in water and in acetonitrile/water [97.5:2.5 (vol/vol)]. Fractions were collected and analyzed on LC/MS for the presence of C<sub>3</sub>H<sub>5</sub>N<sub>2</sub>O<sub>2</sub> as described in Materials and Methods, Untargeted Metabolomics.

Further separations were performed on an Ultimate 3000 LC System (Dionex). First, ion-pair fractionation was performed on a Luna 3u C<sub>18</sub> 100A (150 × 10 mm, 3 μm) column (Phenomenex) with mobile phases previously described (1). Collected fractions were dried under vacuum and analyzed on LC/MS. The third fractionation step was performed on the same Luna 3u C<sub>18</sub> 100A (150 × 10 mm, 3 μm) column, but with the reversed-phase mobile phases used for the flash chromatography. The fractions containing C<sub>3</sub>H<sub>5</sub>N<sub>2</sub>O<sub>2</sub> were finally purified on a ReproSil-Pur Phenyl column (150 × 4.6 mm, 3 μm; Dr. Maisch) with mobile phases consisting of 5 mM ammonium acetate in water and methanol. Fractions containing C<sub>3</sub>H<sub>5</sub>N<sub>2</sub>O<sub>2</sub> were pooled, freeze-dried three times to remove ammonium acetate and water residues, and dissolved in D<sub>2</sub>O for NMR analysis.

**Chemical N-t-Lac-t-Phe Synthesis.** General reagents were obtained from Sigma–Aldrich and used as received. The t-Phe ester hydrochloride was purchased from Bachem and (-)-ethyl-l-tartate was obtained from Aldrich. Solvents were purchased from Biosolve. Analytical TLC was performed on aluminum sheets precoated with silica gel 60 F<sub>254</sub> (Merck), and spots were visualized using 20% (wt/vol) ninhydrin in ethanol and heating by a heat gun. NMR spectra [<sup>1</sup>H-NMR, <sup>13</sup>C-NMR (attached proton test), and <sup>1</sup>H-COSY] were determined in D<sub>2</sub>O as described in Materials and Methods, Structure Elucidation. Peak shapes in NMR spectra are indicated with the symbols “d” (doublet), “dd” (double doublet), “s” (singlet), “q” (quartet), and “m” (multiplet). Chemical shifts (δ) are given in parts per million, and coupling constants (J) are given in hertz. The accurate mass and MS<sup>2</sup> fragmentation spectrum were acquired using the same LC/MS conditions as described in Materials and Methods, Untargeted Metabolomics, at a normalized collision energy of 35.

1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (815 mg, 5.25 mmol, 1.05 eq) was added to a cold mixture (ice bath) of (-)-ethyl-l-tartate (1.15 g, 5 mmol, 1 eq) and 1-hydroxybenzotriazole (1.35 g, 10 mmol, 2 eq) in dichloromethane (DCM; 15 mL). Stirring was continued for 10 min before a mixture of t-Phe ethyl ester hydrochloride (450.4 mg, 5 mmol) and N,N-disopropylethylamine (871 μL, 5 mmol, 1 eq) in DCM (15 mL) was added. The ice bath was removed, and the reaction mixture was stirred for 16 h at room temperature. After removal of the solvent, the residue was taken up in ethylacetate (EtOAc); washed with K<sub>2</sub>HPO<sub>4</sub>, brine, NaHCO<sub>3</sub>, and brine; dried (Na<sub>2</sub>SO<sub>4</sub>); and then concentrated. The crude N-t-lac-t-Phe ester ethyl was dissolved in tetrahydrofuran (10 mL), and 2 N of LiOH (5 mL) was added. The resulting mixture was stirred at room temperature for 1 h. After hydrolysis of the ester was complete, as indicated by TLC analysis [eluent: EtOAc/heptane, 1:1 (vol/vol)], the reaction was quenched by the addition of 6 N of HCl (2 mL) and diluted with water (50 mL). Organic solvents were removed under reduced pressure, and the aqueous residue was extracted with EtOAc (150 mL). The organic layer was dried (Na<sub>2</sub>SO<sub>4</sub>), concentrated, and coevaporated with DCM (three times), and then dried under high vacuum to give N-t-lac-t-Phe as a white foam (783 mg, 66%); <sup>1</sup>H-NMR (300 MHz, D<sub>2</sub>O) δ 7.49–7.05 (m, Ph CH, 5H), 4.43 [dd, J = 8.3, 4.9 Hz, αCH (Lac), 1H], 4.10 [q, J = 6.9 Hz, αCH (Phe), 1H], 3.17 [dd, J = 13.8, 4.9 Hz, βCH<sub>2</sub> (Phe), 1H], 2.93 [dd, J = 13.8, 8.3 Hz, βCH<sub>2</sub> (Phe), 1H], 1.14 [d, J = 6.9 Hz, CH<sub>3</sub> (Lac), 3H]; <sup>13</sup>C-NMR (75 MHz, D<sub>2</sub>O) δ 177.7 (C = O), 176.5 (C = O), 137.5 (Ph qCH), 129.5 (Ph CH), 128.6 (Ph CH), 126.9 (Ph CH), 67.7 [βCH<sub>2</sub> (Lac)]; 55.4 [αCH (Phe)]; 37.8 [βCH<sub>2</sub> (Phe)]; 19.5 [CH<sub>3</sub> (Lac)]; LC/MS (m/z): 236.093 [M-H]<sup>−</sup>.

**Proteomics.** Equal aliquots (30 μL) of the selected fractions were diluted with SDS sample loading buffer and cleaned by short, partial SDS/PAGE separation. The gel was stained with GelCode Blue stain reagent (Thermo Scientific), followed by excision of gel bands, reduction of the proteins with DTT, and alkylation with iodoacetamide. Proteins were digested with trypsin (3 ng/μL) overnight at 37 °C. After digestion, peptides were extracted with acetonitrile and dried in a speedvac. Before MS analysis, the peptides were reconstituted in 10% formic acid.

Peptides were separated using the Proxeon nLC 1000 system (Thermo Scientific) fitted with a trapping column (ReproSil-Pur 120 C18-AQ, 3 μm, 100 μm × 200 mm; Dr. Maisch) and an analytical column (Agilent Poroshell EC-C18 120, 2.7 μm, 50 μm × 300 mm), both packed in-house. The outlet of the analytical column was coupled directly to a Thermo Orbitrap Fusion (Q-OT-qIT; Thermo Scientific) using a Proxeon nanoflex source. NanoSpray was achieved using a distally coated, fused silica tip emitter (generated in-house: o.d. = 375 μm, i.d. = 20 μm) operated at 2.0 kV. Solvent A was 0.1% formic acid/water, and solvent B was 0.1% formic acid/acetonitrile. Samples (50% of the resuspended volume) were eluted from the analytical column at a constant flow of 150 nL-min<sup>−1</sup> in a 55-min gradient containing a 36-min linear increase from 5 to 24% solvent B, followed by a 2-min wash at 80% solvent B. Survey scans of peptide precursors from m/z 375–1,500 were performed at a resolution of 120 K with a 4 × 10<sup>4</sup> ion count target. Tandem MS was performed by quadrupole isolation at 1.6 Da, followed by collision-induced dissociation fragmentation with normalized collision energy of 35 and ion trap MS<sup>2</sup> fragment detection. The MS<sup>2</sup> ion count target was set to 10<sup>6</sup>, and the maximum injection time was set to 35 ms. Only precursors with a charge state of 2–6 were sampled for MS<sup>2</sup>. Monoisotopic precursor selection was turned on; the dynamic exclusion duration was set to 10 s with a 10-ppm tolerance around the selected precursor and its isotopes. The instrument was run in top speed mode with 3-s cycles.

Raw data files were processed using Proteome Discoverer (version 1.4.0.288; Thermo Fisher Scientific). MS<sup>2</sup> spectra were searched against the UniProt database (release 2013_12, 541954 entries) using Mascot (version 2.4.1; Matrix Science) and Homo sapiens as a taxonomy filter. Carbamidomethylation of Cys was set as a fixed modification, and oxidation of Met was set as a variable modification. Tryptsin was specified as the enzyme, and up to two miscleavages were allowed. Data filtering was performed using the percolator algorithm (2), resulting in a 1% false discovery rate. Additional filters were search engine rank 1 peptides and an ion score >20.

**Determination of the Equilibrium Constant of N-Lac-Phe Formation in Vitro.** Recombinant human CNDP2 (1 μg, C-terminal HIS-tagged; ProSpec) was incubated with different concentrations of
lactate and Phe in 25 mM Tris-HCl (pH 7.4) containing 0.1 mM MnCl₂ in a total volume of 1 mL. At the indicated time points, 100-μL samples were taken, which were immediately quenched in 900 μL of 10% (vol/vol) acetic acid and applied to a Strata-X solid phase extraction column (33 μ polymeric reversed phase, 30 mg/1 mL; Phenomenex) preconditioned with 1 mL of methanol and 1 mL of water. After washing the column with 1 mL of 2% (vol/vol) acetic acid in 5% (vol/vol) methanol, the N-lac-Phe was eluted with 1 mL of 2% (vol/vol) NH₄OH in 92% (vol/vol) methanol. The eluates were dried in a speedvac and reconstituted in ion-pairing mobile phase A. N-lac-Phe was quantified by LC/MS using the following gradient: 0–3 min: 50% B, 3–8 min: 50–100% B, 8–12 min: 100% B, 12–13 min: 100–50% B, 13–16 min: 50% B. Mobile phases and MS conditions were as described in Materials and Methods, Untargeted Metabolomics.

Formation of N-Lactoyl-Amino Acids in Whole Blood and Plasma. Whole blood was collected in lithium heparin tubes (BD) and placed on ice. One part was centrifuged (10 min, 3,000 × g, 4 °C), after which the plasma layer was collected. Three aliquots of the plasma were stored on ice and processed as such. Additional aliquots of 0.5 mL of plasma and 1.4 mL of whole blood were spiked with lactate or Phe to a final concentration of 10 mM and 500 μM, respectively. After incubating the samples for 30 min at 37 °C, the whole-blood samples were centrifuged (10 min, 3,000 × g, 4 °C) and 0.5 mL of the plasma phase was collected. All plasma samples were finally processed as described for the human exer- cise plasma samples.

N-Lac-Phe Formation from Methyl-Lactate. Recombinant human CNDP2 (C-terminal HIS-tagged; R&D Systems) was incubated (37 °C) with combinations of lactate, (−)-methyl t-lactate, and Phe in 25 mM Tris-HCl (pH 7.4) containing 0.1 mM MnCl₂. The reactions were quenched by the addition of 50 μL of acetic acid and processed as described in the main text.

N-Lac-Phe Hydrolysis by CNDP2. Recombinant human CNDP2 (1 μg, C-terminal HIS-tagged; ProSpec) was incubated (37 °C) with N-lac-Phe (2.5 μM) in 1 mL of 25 mM Tris-HCl (pH 7.4) containing 0.1 mM MnCl₂. At the indicated time points, a sample of 100 μL was collected, deproteinated with 300 μL of cold methanol, and centrifuged (10 min, 21,800 × g, 4 °C). The supernatant was evaporated to dryness in a speedvac, reconstituted in 50 μL of mobile phase A, and analyzed by LC/MS as described in SI Materials and Methods, Determination of the Equilibrium Constant of N-Lac-Phe Formation in Vitro.

CNDP2-Mediated Cys-Gly Hydrolysis. CNDP2-mediated Cys-Gly hydrolysis was determined essentially as described by Kaur et al. (3). In short, 100 μL of reaction buffer [25 mM Tris-HCl (pH 7.4), 0.1 mM MnCl₂, 1 mM DTT, 2 mM Cys-Gly with or without 0.1 μg of CNDP2] was incubated for 60 min at 37 °C. The reaction was quenched by adding 100 μL of glacial acetic acid. Cys release was determined by adding 100 μL of ninhydrin reagent (250 mg of ninhydrin dissolved in 3 mL of acetic acid and 2 mL of concentrated HCl) and incubation for 10 min at 100 °C. After cooling on ice, samples were diluted with 700 μL of ethanol and the A560 was determined in 200 μL of the diluted samples using an M200 Pro (TECAN). The concentration of Cys was determined using a standard curve containing known amounts of Cys.


Fig. S1. Relates to Fig. 1. Intracellular levels of the unknown metabolites that accumulate in HEK 293/ABCC5 culture medium do not differ substantially between HEK 293 and HEK 293/ABCC5. HEK 293 control and HEK 293/ABCC5 cells were grown to confluence in six-well plates and cultured for an additional 3 d. Cell lysates were analyzed using untargeted LC/MS metabolomics. Data are presented as mean plus SD (n = 3). AU, arbitrary units.
Fig. S2. Relates to Fig. 2. NMR and MS$^2$ spectra of unknown C$_{12}$H$_{15}$NO$_4$ purified from culture medium match those NMR and MS$^2$ spectra of chemically synthesized N-L-lac-Phe. Differences in intensities can be attributed to the lower concentration of isolated C$_{12}$H$_{15}$NO$_4$ compared with synthetic N-L-lac-Phe.
Endogenous N-lac-Phe consists of N-L-lac-L-Phe. Synthesized N-L-lac-L-Phe elutes at the same time as the product formed by incubation of L-Lac and L-Phe with CNDP2. Incubation of CNDP2 with racemic D,L-Lac and L-Phe yields a minor (1.5%) secondary peak with the same mass and fragmentation spectrum, which we consider to be the diastereomer D,L-lac-L-Phe. Plasma from a patient with PKU almost exclusively contained N,L-lac-L-Phe. The differences in retention time compared with Fig. 2 are due to differences in column age.
**Table:**

<table>
<thead>
<tr>
<th>Chemical structure</th>
<th>N-lactoyl-Phe</th>
<th>N-lactoyl-(iso)Leu</th>
<th>N-lactoyl-Tyr</th>
<th>N-lactoyl-Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical formula</td>
<td>C₁₂H₁₅NO₄</td>
<td>C₉H₁₇NO₄</td>
<td>C₁₂H₁₅NO₅</td>
<td>C₁₄H₁₆N₂O₄</td>
</tr>
</tbody>
</table>

**Fig. S4.** Relates to Fig. 2. MS fragmentation spectra of the unknowns and amino acids indicate that the unknowns are pseudopeptides of lactate and amino acids. The MS² fragmentation spectra of the identified N-lac-Phe and the remaining unknowns show common fragments, such as m/z 88.041 and m/z 146.046. Additionally, the MS² spectra all show a neutral loss of 72.021 Da (lactic acid - H₂O) that leads to a fragment with the mass of an amino acid (highlighted in the boxes in row 2). The MS³ fragmentation spectra of these MS² fragments match the MS² spectra of reference amino acids. *The MS² spectra of Ile and Leu were identical, preventing further specification. The lactate moiety in the identified metabolites is highlighted in blue.

**Fig. S5.** Relates to Fig. 4. Determination of the equilibrium constant. Recombinant human CNDP2 (1 μg) was incubated (24 h, 37 °C) with different concentrations of lactate and Phe in 25 mM Tris·HCl (pH 7.4) containing 0.1 mM MnCl₂. The N-lac-Phe levels were determined by LC/MS. Data are presented as mean (filled circles; n = 3) plus SD. A linear curve was fitted through the data and is represented by the dashed line.

K = 2.7 x 10⁻² M⁻¹
Fig. S6. Relates to Fig. 5. 13C-Metabolic labeling of lactate and N-lac-Phe in culture medium from HEK 293 cells. HEK 293 cells were grown to confluence in six-well plates, at which point the medium was replaced with medium containing 13C6-glucose. At several time points, the amount of lactate and N-lac-Phe containing only 12C (U-12C; unlabeled) or one to four 13C atoms per molecule (13C1–4) were determined in lysate by accurate mass LC/MS. Fully 13C-labeled lactate contains three 13C-atoms, which is reflected in the 13C3-labeling of N-lac-Phe. The minor presence of 13C1, 13C2, and 13C4 isotopologs can be explained by the natural occurrence of the 13C isotope (~1%). Levels are expressed as absolute (AU) and relative values (percentage of isotope total). Data are presented as mean (n = 3) and SD (only for absolute values).

Fig. S7. Relates to Fig. 6. N-lac-Phe is not formed in plasma and is formed only slowly in whole blood. Control plasma or whole blood was spiked with 10 mM lactate or 500 μM Phe and incubated for 30 min at 37 °C. Aliquots of the whole blood were immediately processed to plasma and analyzed without incubation. After centrifugation of the whole-blood samples and collection of the plasma layer, all plasma samples were analyzed for N-lac-Phe by LC/MS. N-lac-Phe levels were modestly increased when whole blood was incubated with either lactate or Phe. This increase was not due to lysis of erythrocytes because the addition of an identical volume of water did not affect N-lac-Phe levels. Significance is calculated using a two-sided Student’s t test, utilizing the water control as a reference. Data are expressed as mean plus SD (n = 3). **P < 0.01; ***P < 0.001.
Fig. S8. Relates to Fig. 4. CNDP2 prefers lactate over methyl-lactate. Recombinant human CNDP2 (1 μg) was incubated (8 h, 37 °C) with methyl-lactate plus Phe (5 μmol) or lactate plus Phe (5 μmol) in 25 mM Tris HCl (pH 7.4) containing 0.1 mM MnCl₂. The N-lac-Phe levels were determined by LC/MS. Data are presented as mean (n = 3) plus SD.

Fig. S9. Relates to Fig. 4. CNDP2 hydrolyzes N-lac-Phe and the classical substrate Cys-Gly. (A) N-lac-Phe (2.5 μM) was incubated (37 °C) in the presence (●) or absence (○) of recombinant human CNDP2 [1 μg in 1 mL of 25 mM Tris HCl (pH 7.4) containing 0.1 mM MnCl₂]. At several time points, samples corresponding to 0.1 μg of CNDP2 were collected, deproteinated, and analyzed by LC/MS. (B) Cys-Gly (2 mM) was incubated (37 °C) in the presence and absence of recombinant human CNDP2 [1 μg in 1 mL of 25 mM Tris-HCl (pH 7.4) containing 0.1 mM MnCl₂ and 1 mM DTT] for 1 h. Cys-Gly hydrolysis was assessed by measuring the release of Cys, using ninhydrin.