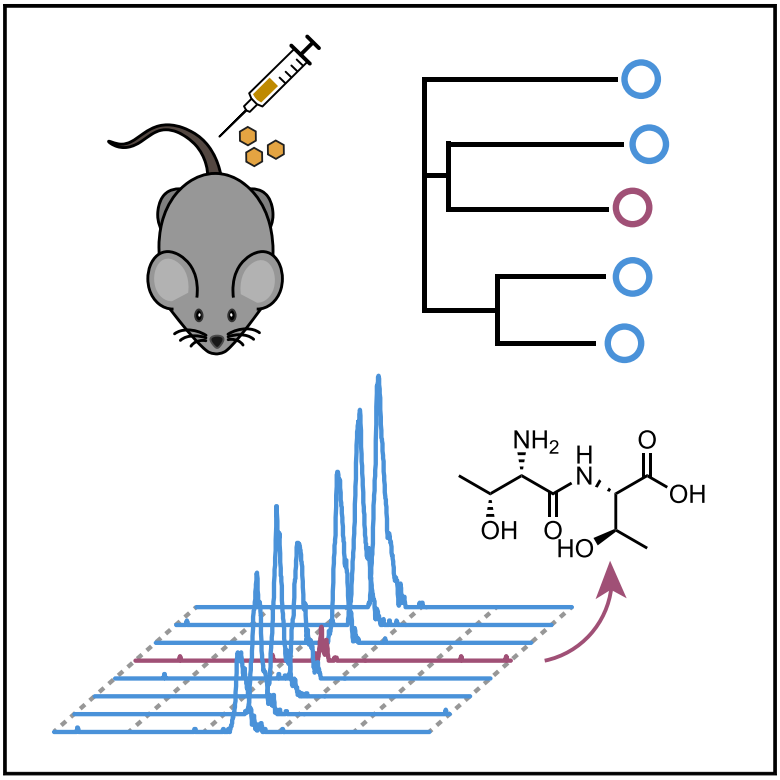
Resource

Cell Chemical Biology

Family-wide Annotation of Enzymatic Pathways by Parallel *In Vivo* Metabolomics

Graphical Abstract



Highlights

- A metabolomics strategy for family-wide deorphanization of enzymes in mice
- The mammalian M20 peptidases regulate amino acid-conjugated metabolites
- CNDP2 controls a catabolic node for threonyl dipeptides
- Ablation of CNDP2 in mice elevates threonyl dipeptides across multiple tissues and blood

Authors

Joon T. Kim, Veronica L. Li, Stephanie M. Terrell, Curt R. Fischer, Jonathan Z. Long

Correspondence

jzlong@stanford.edu

In Brief

Kim et al. describe a general and rapid strategy for family-wide annotation of enzymatic pathways in animals. Applying their method to the entire mammalian M20 peptidases uncovers both known and orphan pathways of amino acid metabolism.

Kim et al., 2019, Cell Chemical Biology *26*, 1–7
December 19, 2019
https://doi.org/10.1016/j.chembiol.2019.09.009

CellPress

Please cite this article in press as: Kim et al., Family-wide Annotation of Enzymatic Pathways by Parallel *In Vivo* Metabolomics, Cell Chemical Biology (2019), https://doi.org/10.1016/j.chembiol.2019.09.009

Cell Chemical Biology
Resource

Family-wide Annotation of Enzymatic Pathways by Parallel *In Vivo* Metabolomics

Joon T. Kim,[1,2] Veronica L. Li,[1,2] Stephanie M. Terrell,[1,2] Curt R. Fischer,[2] and Jonathan Z. Long[1,2,3,*]
[1]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Stanford ChEM-H, Stanford University, Stanford, CA 94305, USA
[3]Lead Contact
*Correspondence: jzlong@stanford.edu
https://doi.org/10.1016/j.chembiol.2019.09.009

SUMMARY

Enzymes catalyze fundamental biochemical reactions that control cellular and organismal homeostasis. Here we present an approach for *de novo* biochemical pathway discovery across entire mammalian enzyme families using parallel viral transduction in mice and untargeted liquid chromatography-mass spectrometry. Applying this method to the M20 peptidases uncovers both known pathways of amino acid metabolism as well as a previously unknown CNDP2-regulated pathway for threonyl dipeptide catabolism. Ablation of CNDP2 in mice elevates threonyl dipeptides across multiple tissues, establishing the physiologic relevance of our biochemical assignments. Taken together, these data underscore the utility of parallel *in vivo* metabolomics for the family-wide discovery of enzymatic pathways.

INTRODUCTION

Metabolic enzymes catalyze fundamental biochemical reactions that control cellular and organismal homeostasis. Over the past century, many key enzymes and biochemical transformations have been uncovered using classical *in vitro* enzymology or *in vivo* radioisotope tracing approaches (Belfrage et al., 1977; Krebs, 1936; Lynen and Ochoa, 1953). Despite this progress, recent untargeted metabolomics data have revealed that >50% of biochemical space still remains uncharacterized in terms of metabolite identity, enzymatic regulation, or physiologic function (Psychogios et al., 2011; Wang et al., 2019). General strategies for scalable mapping of this uncharted biochemistry would pave the way for understanding the breadth of enzyme and metabolite control of cellular and organismal physiology.

Recently, a variety of metabolomics strategies have been used to tackle this problem (Chennamsetty et al., 2016; Dang et al., 2009; Long et al., 2011; Mülleder et al., 2012, 2016; Saghatelian et al., 2004). *In vitro* or cell-based approaches, while rapid, do not always capture the diversity of metabolites *in vivo* and may also not identify physiologically relevant enzyme activities. Enzyme knockout approaches in animals can provide more physiologically relevant biochemical transformations, but nevertheless are not easily scalable and can be confounded by developmental compensation or cellular and tissue dysfunction due to the long-term nature of the genetic perturbation. To complement all of these strategies, we have devised an alternative approach for biochemical pathway mapping. This approach captures the complexities of metabolite regulation in animals, avoids long-term genetic perturbations, and maintains both speed and scalability across entire enzyme families. Key to our platform is the use of adeno-associated viruses (AAVs) for rapid and temporally controlled overexpression of target enzymes directly in mouse tissues. These viral transductions are then paired with downstream untargeted metabolomics for the inference of biochemical reactions. By using multiple AAVs corresponding to distinct enzymes across multiple mice in one large parallel experiment, we envisioned that our platform could provide a general and rapid system for *de novo* biochemical reaction discovery across entire enzyme families and directly in animals. We apply this platform to the mammalian M20 peptidases where we identify both previously known and orphan pathways of amino acid metabolism. Finally, we validate the physiologic relevance of our biochemical assignments using CNDP2-knockout (KO) animals.

RESULTS

Identification of Biochemical Reactions Catalyzed by the Mammalian M20 Peptidases *In Vivo*

As a testing ground for our approach, we selected all the enzymes of the mammalian M20 peptidase family (Figure 1A). This enzyme class contains both orphan enzymes (e.g., CNDP2 and PM20D2), as well as those whose physiologic reactions are known (Long et al., 2016, 2018; Sass et al., 2006; Sauerhöfer et al., 2007) (ACY1, CNDP1, and PM20D1), thereby providing both opportunities for *de novo* discovery as well as positive controls for our approach. We generated AAV serotype 8 (AAV8) viruses expressing each of the five M20 peptidases with a C-terminal FLAG tag as well as an AAV8-GFP control virus. Each enzyme virus was transduced into three animals via tail vein injection (10^{11} infectious units/mouse), resulting in a total of 18 transduced mice corresponding to the five enzyme groups and one GFP group. The AAV8 serotype predominantly infects the liver, although other tissues are also transduced to a lesser extent (Zincarelli et al., 2008). Two weeks after viral transduction, livers from these animals were harvested. Stable overexpression of all constructs was confirmed by western

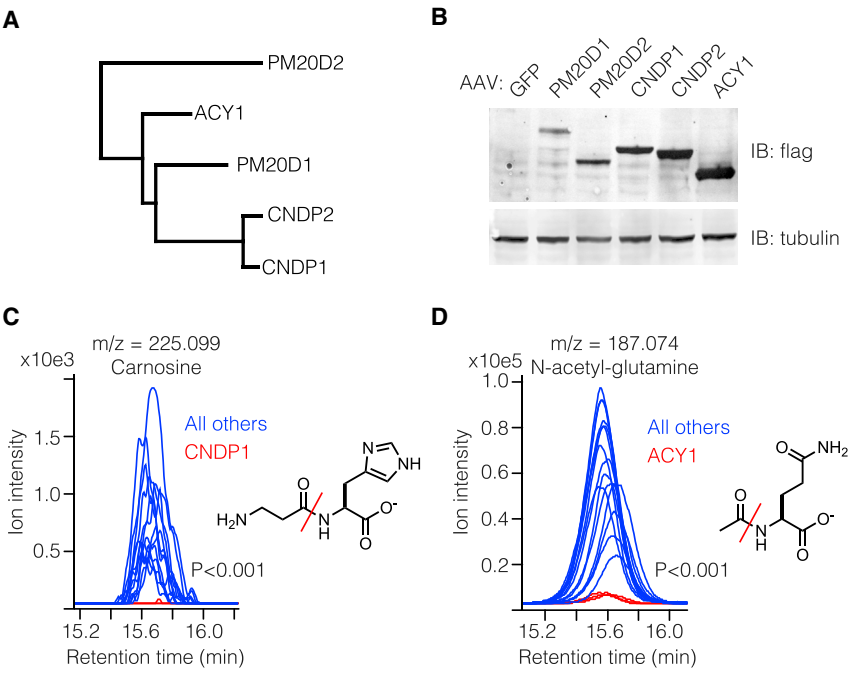


Figure 1. Viral Transduction of the M20 Peptidases *In Vivo* and Depletion of Carnosine and N-Acetyl Amino Acids by CNDP1 and ACY1, Respectively

(A) Dendrogram showing the murine M20 peptidase family.

(B) Immunoblot using anti-flag (top) or anti-β-tubulin (bottom) antibodies of livers from mice transduced with AAV-GFP or AAVs encoding the indicated flag-tagged enzymes.

(C and D) Extracted ion chromatograms of carnosine (C) and N-acetyl-glutamine (D) from transduced livers, their structural assignments, and the inferred biochemical activity of the indicated enzyme (in red). All chromatograms were extracted at the exact mass ± 20 ppm. For (C) and (D), n = 3/group for the enzyme-transduced group and n = 15 for all other viral transduction groups.

blotting using an anti-flag antibody (Figure 1B). Liver tissues were then extracted with acetonitrile/methanol and analyzed by untargeted liquid chromatography-mass spectrometry-based metabolomics. The vast majority of metabolite peaks were unaltered by transduction of any individual enzyme (Figure S1), establishing that each enzyme perturbs only a small region of the metabolome. To identify differentially regulated metabolites between each enzyme group, XCMS software (Smith et al., 2006) was used to identify features altered by a single enzyme transduction (n = 3) versus all other viral transduction groups (n = 15). This procedure was repeated for each of the five M20 peptidases. Metabolite features were filtered using a combination of statistical significance (|tstat| > 4) as well as magnitude change (fold change >4 versus all other groups). These cutoffs resulted in an average of 4.6 ± 2.0 (mean ± SEM) features that were specifically altered by one enzyme relative to the others in our enzyme set, representing the top 0.1% of enzyme-regulated metabolites out of a total of ~3,200 features quantitated (Table S1). Analysis of these differentially regulated features showed that each M20 peptidase altered a distinct set of metabolite peaks (Figure S1), establishing the generality and feasibility of perturbing small and distinct regions of biochemical space by viral transduction of specific metabolic enzymes.

CNDP1 and ACY1 Deplete Carnosine and N-Acetyl Amino Acids, Respectively

Inspection of the resulting dataset revealed two cases in which our viral transduction approach "re-discovered" known reactions of amino acid metabolism catalyzed by these M20 peptidases. First, only a single feature meeting our filtering criteria was altered by CNDP1 overexpression. This peak was depleted by >90% in CNDP1-transduced animals and had exact mass of m/z = 225.099 in negative ionization mode and retention time ~15.7 min (Figure 1C). From this mass, we predicted a molecular formula of $C_9H_{14}N_4O_3$ (calcd $[M\text{-}H]^-$ = 225.099), which matches carnosine. An authentic carnosine standard co-eluted with this feature and its fragmentation spectrum matched the endogenous m/z = 225.099 peak, providing direct evidence that carnosine is strongly depleted by CNDP1 overexpression (Figure S2). These data re-confirm previous observations that CNDP1 physiologically catalyzes the hydrolysis of carnosine to β-alanine and histidine (Sauerhöfer et al., 2007; Schmöhl et al., 2019). Second, the most abundant peak specifically regulated by ACY1 had exact mass m/z = 187.073, retention time ~15.6 min, and was depleted by >90% by ACY1 overexpression (Figure 1D). We predicted a molecular formula of $C_7H_{12}N_2O_4$ (calcd $[M\text{-}H]^-$ = 187.072) for this feature. We confirmed that this feature corresponds to N-acetyl glutamine by comparison of retention times and fragmentation spectra to an authentic standard (Figure S2). From this identification, we could also infer the identity of two additional peaks of m/z = 146.047 and 196.074 as N-acetyl serine and N-acetyl histidine, respectively. These data confirm that ACY1 catalyzes the conversion of N-acetyl amino acids to amino acids and acetate *in vivo*, consistent with the observation that human ACY1 deficiency results in dramatic accumulation of several N-acetyl amino acids in urine (Van Coster et al., 2005; Sass et al., 2006).

CNDP2 Overexpression Depletes Threonyl Dipeptides *In Vivo*

In addition to recovering known enzyme-substrate relationships, we sought to use our approach to discover new biochemical transformations. CNDP2 (also called carnosine dipeptidase 2 or cytosolic nonspecific dipeptidase 2) is a member of the M20 peptidases whose endogenous biochemical activity remains unknown despite the multiple proposed reactions based on *in vitro* studies (Jansen et al., 2015; Kaur et al., 2009; Teufel et al., 2003). Notably, from our dataset CNDP2 did not appear to be a physiologic carnosinase because its overexpression did not alter tissue carnosine levels (Figure 1C). Instead, the most abundant feature altered in CNDP2-transduced livers was a peak with mass m/z = 219.0991 and

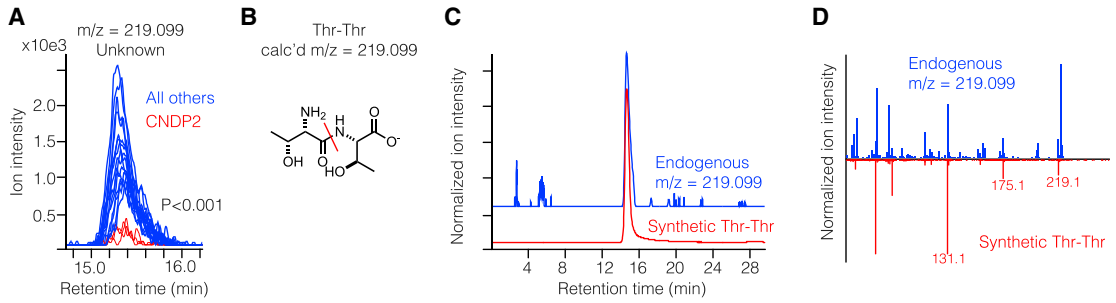


Figure 2. CNDP2 Overexpression Depletes Thr-Thr *In Vivo*

(A) Extracted ion chromatogram of an unknown metabolite at m/z = 219.099 from transduced livers in CNDP2-transduced (red) versus all other viral transduction groups (blue).

(B–D) Structural assignment (B), co-elution (C), and tandem MS fragmentation (D) of the endogenous m/z = 219.099 and an authentic Thr-Thr standard. For (A) and (C), chromatograms were extracted at the exact mass ± 20 ppm.

retention time ~15.3 min (Figure 2A). This peak likely represented a CNDP2 substrate because it was depleted >80% by CNDP2 overexpression. Using this exact mass, we deduced a molecular formula of $C_8H_{16}N_2O_5$. One candidate structure that matched both the molecular formula as well as the potential hydrolytic activity of CNDP2 was the dipeptide Thr-Thr, a previously unknown mammalian metabolite (Figure 2B). The endogenous 219.0991 peak co-eluted with synthetic Thr-Thr and also exhibited the fragmentation ions of 175.1 (decarboxylation) and 131.1 (loss of CO_2 and ethanol) characteristic of synthetic Thr-Thr (Figures 2C and 2D), establishing the chemical identity of this peak. In this same dataset we also observed a second peak of mass m/z = 205.082 and retention time ~16.2 min that was depleted by ~60% by CNDP2 (Table S1). Critically, the mass difference in these two peaks corresponded exactly with a methyl group (mass difference of 14.016), suggesting m/z = 205.082 might correspond with Thr-Ser. By co-elution and MS/MS, the m/z = 205.0821 peak matched to a dipeptide of Thr and Ser, although we were unable to distinguish between Thr-Ser or Ser-Thr isomers (Figure S2). Quantitation of these threonyl dipeptides using a synthetic dipeptide standard established that Thr-Thr and Thr-Ser/Ser-Thr are endogenously present at 97 ± 11 and 400 ± 40 nM (mean ± SEM, n = 5), respectively, in mammalian liver. Taken together, these data suggest a previously unknown biochemical reaction in which threonyl dipeptides are hydrolyzed into individual amino acids by the action of CNDP2.

In Vitro Hydrolysis Activity of Recombinant CNDP2

To determine whether threonyl dipeptides are direct CNDP2 substrates, we generated purified, recombinant mammalian C-terminal flag-tagged CNDP2 from HEK293T cells. As an inactive control protein, we also generated a CNDP2 protein with an E166A point mutation which, based on homology with the other mammalian M20 peptidases, is predicted to disrupt a critical active site residue. Both CNDP2 wild-type and CNDP2-E166A proteins were expressed to similar extents as determined by western blotting using an anti-flag antibody (Figure 3A). *In vitro*, CNDP2 robustly hydrolyzed the threonyl dipeptides Thr-Thr, Thr-Ser, and Ser-Thr to the corresponding amino acids (Figures 3B–3D). In contrast, *in vitro* reactions incubated with CNDP2-E166A exhibited no enzymatic activity over negative control reactions, establishing that the CNDP2 polypeptide is both necessary and sufficient for threonyl dipeptide hydrolysis.

CNDP2-KO Mice Exhibit Elevations in Threonyl Dipeptides across Tissues

Finally, we sought to validate the physiologic relevance of our biochemical assignments. At present, no mouse model or human deficiency in CNDP2 has been reported to critically test this hypothesis. Consequently, the extent to which CNDP2 contributes to threonyl dipeptide hydrolysis activity *in vivo* remains unknown. We obtained whole-body *Cndp2* KO mice that were generated by CRISPR-Cas9 technology as part of the International Mouse Phenotyping Consortium (Dickinson et al., 2016). Ablation of CNDP2 was validated by complete loss of CNDP2 immunoreactivity in kidney, a tissue with high *Cndp2* mRNA from publicly available expression data (Figure 4A). In this tissue, hydrolysis activity for Thr-Thr, Thr-Ser, and Ser-Thr was reduced by 70%, 33%, and 61%, respectively, compared with wild-type activity (Figures 4B–4D). Therefore, CNDP2 is a major threonyl dipeptidase *in vivo*.

The reduction of tissue threonyl dipeptide hydrolysis activities in CNDP2-KO animals, along with the depletion of threonyl dipeptides in AAV-CNDP2 experiments, suggested that CNDP2 mediates the physiologic catabolism of threonyl dipeptides as well as potentially other dipeptides. We next measured endogenous levels of threonyl dipeptides from various CNDP2-KO tissues. Thr-Thr and Thr-Ser were accumulated to levels 2- and 5-fold higher, respectively, in kidneys from CNDP2-KO versus wild-type animals (Figure 4E). Importantly, the ACY1- and CNDP1-regulated metabolites N-acetyl glutamine and carnosine were unaltered in CNDP2-KO tissues, confirming that each M20 peptidase regulates distinct biochemical reactions. Analysis of additional tissues from CNDP2-KO animals revealed that threonyl dipeptides were also dramatically elevated in muscle and in blood (Figures 4F and 4G). Taken together, these data demonstrate a physiologic biochemical pathway that catabolizes threonyl dipeptides into individual amino acids mediated by CNDP2. The observation of tissue-specific differences in threonyl dipeptide accumulation in CNDP2-KO tissues suggests potentially additional pathways that may participate in the biosynthesis, degradation, or transport of these metabolites.

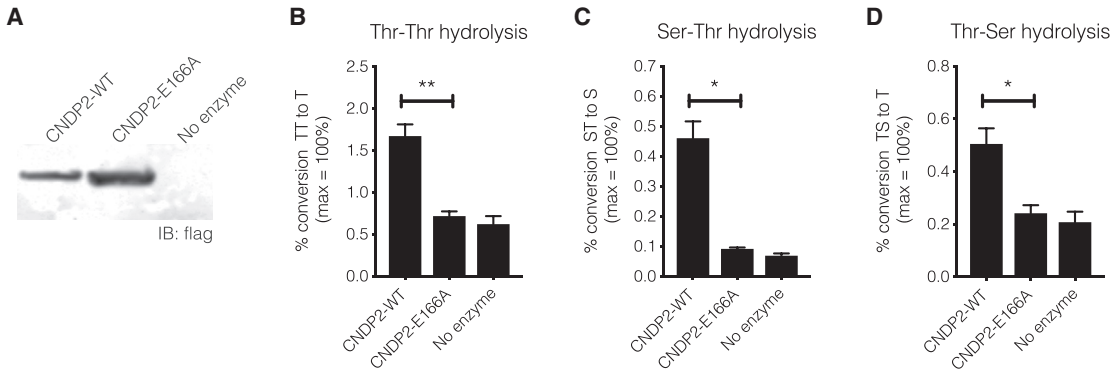


Figure 3. *In Vitro* Hydrolysis of Threonyl Dipeptides by CNDP2

(A) Immunoblot using anti-flag antibodies of the indicated purified recombinant mammalian enzyme from HEK293T cells.

(B–D) *In vitro* Thr-Thr (B), Ser-Thr (C), and Thr-Ser (D) hydrolysis to the indicated dipeptides by CNDP2 or a catalytically dead mutant (CNDP2-E166A). For (B)–(D), n = 3–4/group. Data are shown as means ± SEM; *p < 0.05.

DISCUSSION

Here we present an *in vivo* parallel metabolomics approach for the rapid and scalable discovery of enzyme-catalyzed reactions across enzyme families using AAVs and untargeted metabolomics. Our strategy presents several critical advantages compared with known methods of biochemical reaction mapping. First, our approach represents a "substrate-free" discovery platform that can access previously unexplored regions of the metabolome. Second, performing reaction discovery experiments directly in animals simplifies metabolic pathway identification because non-physiologically relevant enzyme activities did not result in metabolite changes. Third, by using viral transductions, our approach avoids mouse breeding and the confounding developmental or compensatory changes often seen in knockout animals, and therefore is rapid and scalable to entire enzyme families. Lastly, our enzyme gain-of-function approach did not simply represent an artificial ectopic metabolic pathway manipulation, but in fact recovers bona fide physiologic metabolic pathways *in vivo*.

Applying this strategy to the entire mammalian M20 peptidases identified the known catabolic pathways for carnosine and N-acetyl amino acids mediated by CNDP1 and ACY1, respectively. We also uncovered a previously unknown pathway from threonyl dipeptides to free amino acids regulated by CNDP2. Our work underscores the importance of determining enzyme-substrate relationships directly in complex physiologic settings to uncover true physiologic functions beyond what can be assigned by *in vitro* biochemical analysis alone. Moreover, these data demonstrate that individual enzymes within a larger family, despite sharing considerable sequence homology to each other, nevertheless exhibit control over distinct regions of biochemical space. These studies provide further evidence for complex enzymatic control of amino acid metabolism beyond the canonical primary pathways of protein synthesis or energy homeostasis. An inability to definitively assign biochemical activities for the other two M20 peptidase members identify limitations of our approach and suggests that false negatives can arise due to tissue-specific regulation of metabolism or the incomplete coverage by metabolite databases. For instance, AAV-PM20D1 transduced mice exhibited robustly increased N-oleoyl-phenylalanine in circulation (Figure S3) as previously reported (Long et al., 2016). However, similar metabolite changes were not observed in liver, the tissue where we performed our metabolomic comparisons (Figure S3). These observations suggest that plasma membrane transport mechanisms and/or tissue-specific substrate availability can both contribute to the ultimate steady-state levels of these metabolites. For AAV-PM20D2 transduced animals, we observed five peaks that met our filtering criteria (Table S1) but were unable to establish definitive chemical formulas or structural assignments for any of those masses. We anticipate that as small-molecule databases continue to expand and coverage of the metabolome continues to improve, such limitations will eventually be overcome. As we continue to annotate additional enzyme families, we may also encounter other potential limitations of this approach including alterations in subcellular localization and/or regulation by enzyme overexpression.

The gain- and loss-of-function mouse models presented here suggest that a relatively specific physiologic role for CNDP2 is in the catabolism of threonyl dipeptides, in contrast to diverse *in vitro* enzymatic activities that had been previously assigned to this enzyme (Jansen et al., 2015; Kaur et al., 2009; Teufel et al., 2003). Under basal conditions, CNDP2-mediated catabolism of threonyl dipeptides does not appear to be a major contributor to steady-state levels of circulating free threonine (Figure S4). However, we cannot exclude the possibility that CNDP2 participates in other types of biochemical activities under specific situations (for instance, in other tissues or under certain physiologic perturbations such as exercise). The endogenous physiologic functions for the CNDP2-threonyl dipeptide pathway still remain largely mysterious. Considering the robust expression of CNDP2 in the small intestine (Wu et al., 2009) and the known catabolism of dietary proteins into di- and tri-peptides (Adibi, 1971; Daniel, 2004), one possibility is that CNDP2 is involved in threonine or other amino acid harvesting from dietary proteins. Such a role would be analogous to the function of other dipeptide transporters expressed in the intestine (Fei et al., 1994; Frey et al., 2006; Nässl et al., 2011). A second possibility is that intestinal dipeptides have been proposed to regulate GLP-1

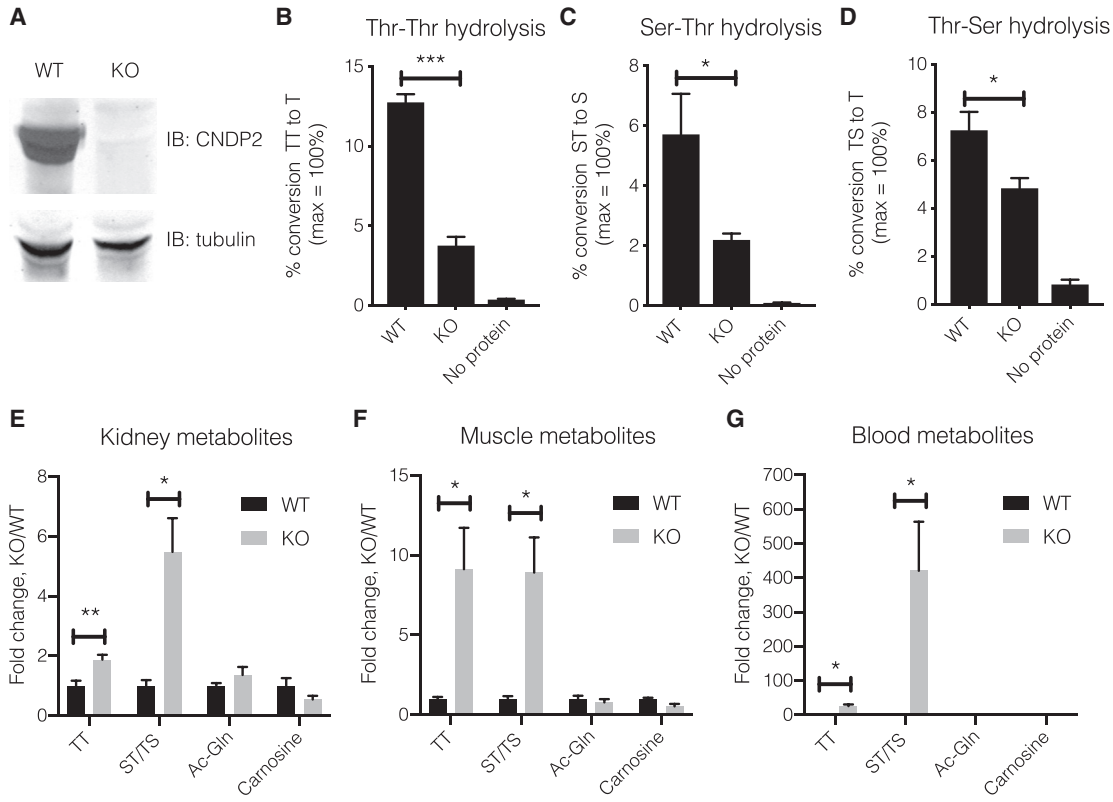


Figure 4. CNDP2 Regulates the Catabolism of Threonyl Dipeptides *In Vivo*

(A) Immunoblot using anti-CNDP2 (top) or anti-tubulin (bottom) antibodies of kidney tissue from CNDP2 wild-type or CNDP2-KO mice.

(B–D) *In vitro* Thr-Thr (B), Ser-Thr (C), and Thr-Ser (D) hydrolysis of the indicated dipeptides in CNDP2-WT or KO kidney tissues.

(E–G) Fold change in endogenous levels of the indicated metabolites from kidney (E), muscle (F), or blood (G) between CNDP2-WT or CNDP2-KO mice, with WT levels set to 1.

For (B)–(D), n = 3/group; for (E)–(G), n = 4/group. Data are shown as means ± SEM; *p < 0.05, ***p < 0.001.

secretion following a high protein meal (Diakogiannaki et al., 2013; Dranse et al., 2018). Consequently, CNDP2 may also participate in glucose homeostasis via a dipeptide/incretin axis. Future phenotyping of CNDP2-KO animals, especially in the context of high protein diets, may reveal the functions of CNDP2-mediated dipeptide catabolism in cellular and organismal homeostasis.

In summary, here we present a powerful parallel metabolomics strategy for enzyme deorphanization directly in mice. By applying this approach to the mammalian M20 peptidases, we uncover both known as well as previously orphan pathways of amino acid metabolism, including a threonyl dipeptide catabolic pathway regulated by CNDP2. Projecting forward, we anticipate that our *in vivo* parallel metabolomics approach will be generally useful for large-scale *de novo* mapping of biochemical reactions.

SIGNIFICANCE

Metabolic enzymes catalyze fundamental biochemical transformations that control cellular and organismal homeostasis. Despite considerable progress using classical approaches, large portions of biochemical space still remain uncharacterized in terms of metabolite identity, enzymatic regulation, or physiologic function. Here, we present a powerful approach for discovering physiologic enzyme-catalyzed transformations across entire enzyme families directly in mice. This approach captures the complexities of metabolite regulation directly in animals, avoids long-term genetic perturbations, and maintains both speed and scalability across multiple biochemical pathways. By applying this platform to the entire mammalian M20 peptidases, we identified both known and previously orphan reactions of amino acid metabolism. Critically, ablation of one M20 peptidase, CNDP2, in mice elevates threonyl dipeptides and underscores the utility of our *in vivo* parallel metabolomics strategy for enzyme reaction discovery.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
  - Animals

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chembiol.2019.09.009.

ACKNOWLEDGMENTS

We thank members of the Long and Svensson labs for helpful discussions. This work was supported by the US NIH (DK105203 to JZL), by the Stanford ChEM-H Institute (CRF), and by the Stanford Diabetes Research Center (P30DK116074).

AUTHOR CONTRIBUTIONS

Conceptualization, J.Z.L. and J.T.K.; Methodology, J.T.K., S.M.T., and C.R.F.; Investigation, J.T.K. and V.L.L.; Resources, V.L.L. and C.R.F.; Writing, J.Z.L.; Funding Acquisition, J.Z.L.; Supervision, J.Z.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 19, 2019
Revised: August 19, 2019
Accepted: September 16, 2019
Published: October 3, 2019

REFERENCES

Adibi, S.A. (1971). Intestinal transport of dipeptides in man: relative importance of hydrolysis and intact absorption. J. Clin. Invest. *50*, 2266–2275.

Belfrage, P., Jergil, B., Strålfors, P., and Tornqvist, H. (1977). Hormone-sensitive lipase of rat adipose tissue: identification and some properties of the enzyme protein. FEBS Lett. *75*, 259–264.

Chennamsetty, I., Coronado, M., Contrepois, K., Keller, M.P., Carcamo-Orive, I., Sandin, J., Fajardo, G., Whittle, A.J., Fathzadeh, M., Snyder, M., et al. (2016). Nat1 deficiency is associated with mitochondrial dysfunction and exercise intolerance in mice. Cell Rep. *17*, 527–540.

Van Coster, R.N., Gerlo, E.A., Giardina, T.G., Engelke, U.F., Smet, J.E., De Praeter, C.M., Meersschaut, V.A., De Meirleir, L.J., Seneca, S.H., Devreese, B., et al. (2005). Aminoacylase I deficiency: a novel inborn error of metabolism. Biochem. Biophys. Res. Commun. *338*, 1322–1326.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739–744.

Daniel, H. (2004). Molecular and integrative physiology of intestinal peptide transport. Annu. Rev. Physiol. *66*, 361–384.

Diakogiannaki, E., Pais, R., Tolhurst, G., Parker, H.E., Horscroft, J., Rauscher, B., Zietek, T., Daniel, H., Gribble, F.M., and Reimann, F. (2013). Oligopeptides stimulate glucagon-like peptide-1 secretion in mice through proton-coupled uptake and the calcium-sensing receptor. Diabetologia *56*, 2688–2696.

Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.

Dranse, H.J., Waise, T.M.Z., Hamr, S.C., Bauer, P.V., Abraham, M.A., Rasmussen, B.A., and Lam, T.K.T. (2018). Physiological and therapeutic regulation of glucose homeostasis by upper small intestinal PepT1-mediated protein sensing. Nat. Commun. *9*, 1118.

Fei, Y.J., Kanai, Y., Nussberger, S., Ganapathy, V., Leibach, F.H., Romero, M.F., Singh, S.K., Boron, W.F., and Hediger, M.A. (1994). Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature *368*, 563–566.

Frey, I.M., Rubio-Aliaga, I., Klempt, M., Wolf, E., and Daniel, H. (2006). Phenotype analysis of mice deficient in the peptide transporter PEPT2 in response to alterations in dietary protein intake. Pflugers Arch. *452*, 300–306.

Jansen, R.S., Addie, R., Merkx, R., Fish, A., Mahakena, S., Bleijerveld, O.B., Altelaar, M., IJlst, L., Wanders, R.J., Borst, P., et al. (2015). N-Lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. Proc. Natl. Acad. Sci. U S A *112*, 6601–6606.

Kaur, H., Kumar, C., Junot, C., Tolendano, M.B., and Bachhawat, A.K. (2009). Dug1p is a Cys-Gly peptidase of the γ-glutamyl cycle of *Saccharomyces cerevisiae* and represents a novel family of Cys-Gly peptidases. J. Biol. Chem. *284*, 14493–14502.

Krebs, H.A. (1936). Intermediate metabolism of carbohydrates. Nature *138*, 288–289.

Long, J.Z., Cisar, J.S., Milliken, D., Niessen, S., Wang, C., Trauger, S.A., Siuzdak, G., and Cravatt, B.F. (2011). Metabolomics annotates ABHD3 as a physiologic regulator of medium-chain phospholipids. Nat. Chem. Biol. *7*, 763–765.

Long, J.Z., Svensson, K.J., Bateman, L.A., Lin, H., Kamenecka, T., Lokurkar, I.A., Lou, J., Rao, R.R., Chang, M.R., Jedrychowski, M.P., et al. (2016). The secreted enzyme PM20D1 regulates lipidated amino acid uncouplers of mitochondria. Cell *166*, 424–435.

Long, J.Z., Roche, A.M., Berdan, C.A., Louie, S.M., Roberts, A.J., Svensson, K.J., Dou, F.Y., Bateman, L.A., Mina, A.I., Deng, Z., et al. (2018). Ablation of PM20D1 reveals *N*-acyl amino acid control of metabolism and nociception. Proc. Natl. Acad. Sci. U S A *115*, E6937–E6945.

Lynen, F., and Ochoa, S. (1953). Enzymes of fatty acid metabolism. Biochim. Biophys. Acta *12*, 299–314.

Mülleder, M., Capuano, F., Pir, P., Christen, S., Sauer, U., Oliver, S.G., and Ralser, M. (2012). A prototrophic deletion mutant collection for yeast metabolomics and systems biology. Nat. Biotechnol. *30*, 1176–1178.

Mülleder, M., Calvani, E., Alam, M.T., Wang, R.K., Eckerstorfer, F., Zelezniak, A., and Ralser, M. (2016). Functional metabolomics describes the yeast biosynthetic regulome. Cell *167*, 553–565.e12.

Nässl, A.M., Rubio-Aliaga, I., Fenselau, H., Marth, M.K., Kottra, G., and Daniel, H. (2011). Amino acid absorption and homeostasis in mice lacking the intestinal peptidetransporter PEPT1. Am. J. Physiol. Gastrointest. Liver Physiol. *301*, G128–G137.

Psychogios, N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., et al. (2011). The human serum metabolome. PLoS One *6*, e16957.

Saghatelian, A., Trauger, S.A., Want, E.J., Hawkins, E.G., Siuzdak, G., and Cravatt, B.F. (2004). Assignment of endogenous substrates to enzymes by global metabolite profiling. Biochemistry *43*, 14332–14339.

Sass, J.O., Mohr, V., Olbrich, H., Engelke, U., Horvath, J., Fliegauf, M., Loges, N.T., Schweitzer-Krantz, S., Moebus, R., Weiler, P., et al. (2006). Mutations in ACY1, the gene encoding aminoacylase 1, cause a novel inborn error of metabolism. Am. J. Hum. Genet. *78*, 401–409.

Sauerhöfer, S., Yuan, G., Braun, G.S., Deinzer, M., Neumaier, M., Gretz, N., Floege, J., Kriz, W., van der Woude, F., and Moeller, M.J. (2007). L-Carnosine, a substrate of carnosinase-1, influences glucose metabolism. Diabetes *56*, 2425–2432.

Schmöhl, F., Peters, V., Schmitt, C.P., Poschet, G., Büttner, M., Li, X., Weigand, T., Poth, T., Volk, N., Morgenstern, J., et al. (2019). CNDP1 knockout in zebrafish alters the amino acid metabolism, restrains weight gain, but does not protect from diabetic complications. Cell. Mol. Life Sci. https://doi.org/10.1007/s00018-019-03127-z.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. *78*, 779–787.

Teufel, M., Saudek, V., Ledig, J.P., Bernhardt, A., Boularand, S., Carreau, A., Cairns, N.J., Carter, C., Cowley, D.J., Duverger, D., et al. (2003). Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase. J. Biol. Chem. *278*, 6521–6531.

Wang, L., Xing, X., Chen, L., Yang, L., Su, X., Rabitz, H., Lu, W., and Rabinowitz, J.D. (2019). Peak annotation and verification engine for untargeted LC-MS metabolomics. Anal. Chem. *91*, 1838–1846.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., et al. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. *10*, R130.

Zincarelli, C., Soltys, S., Rengo, G., and Rabinowitz, J.E. (2008). Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. Mol. Ther. *16*, 1073–1080.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-flag M2 mouse monoclonal	Sigma-Aldrich	F1804; RRID: AB_262044
Anti-beta tubulin rabbit polyclonal	Abcam	Ab6046; RRID: AB_2210370
Anti-CNDP2 rabbit polyclonal	Proteintech	14925-1-AP; RRID: AB_2081682
Bacterial and Virus Strains		
AAV8.CB7.CI.eGFP.WPRE.rBG virus	Penn Vector Core/Addgene	105542-AAV8
Chemicals, Peptides, and Recombinant Proteins		
L-carnosine	Acros Organics	AC208170050
N-acetyl glutamine	Sigma-Aldrich	A9125
N-acetyl histidine	Sigma-Aldrich	S789801
N-acetyl serine	Sigma-Aldrich	A2638
Thr-Thr	Elim Biopharmaceuticals	N/A
Thr-Ser	Elim Biopharmaceuticals	N/A
Ser-Thr	Elim Biopharmaceuticals	N/A
Deposited Data		
Metabolomics dataset, mzXML format	This paper	Mendeley Data, https://doi.org/10.17632/mmk883sjc6.1
Experimental Models: Cell Lines		
Human HEK293T	ATCC	CRL-3216
Experimental Models: Organisms/Strains		
C57BL/6J	Jackson Laboratories	000664
CNDP2-KO mice	Mutant Mouse Regional Resource Center/UC Davis	043492-UCD
Oligonucleotides		
CNDP2 E166A mutagenesis, primer 1: GGAGCCAGACTCCGCCATGCCTTCCAG	Elim Biopharmaceuticals	N/A
CNDP2 E166A mutagenesis, primer 2: CTGGAAGGCATGGCGGAGTCTGGCTCC	Elim Biopharmaceuticals	N/A
mACY1 cloning primer into AAV backbone, F: CACCGAATTCGCCACCATGACCACCAAGGAT CCCGAG	Elim Biopharmaceuticals	N/A
mACY1 cloning primer into AAV backbone, R: TATAGTCGAGTCATTACTACTTGTCGTCATCG TCCTTGTAGTCTCCTCC GCTTTCACTAGGCAG	Elim Biopharmaceuticals	N/A
mCNDP1 cloning primer into AAV backbone, F: CACCGAATTCGCCACCATGTTCTCTTCGGC ACACTC	Elim Biopharmaceuticals	N/A
mCNDP1 cloning primer into AAV backbone, R: TATA GTCGAG TCA TTA CTA cttgtcgtcatcg tccttgtagtC TCC TCC GTAGACACTGGAAGGCATCT	Elim Biopharmaceuticals	N/A
mCNDP2 cloning primer into AAV backbone, F: CACC GAATTC GCCACC ATGTCAGCCCT CAAAGCTGT	Elim Biopharmaceuticals	N/A
mCNDP2 cloning primer into AAV backbone, R: TATA GTCGAG TCA TTA CTA cttgtcgtcatcg tccttgtagtC TCC TCC GTTCTTCAGCTGAGACACCT	Elim Biopharmaceuticals	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
mPM20D2 cloning primer into AAV backbone, F: CACC GAATTC GCCACC ATGGGTCCGGTGG TGGAGCG	Elim Biopharmaceuticals	N/A
mPM20D2 cloning primer into AAV backbone, R: TATA GTCGAG TCA TTA CTA cttgtcgtcatcgtc cttgtagtC TCC TCC TGCTGCTGTATTTAAAAGCT	Elim Biopharmaceuticals	N/A
Recombinant DNA		
mACY1 cDNA, C-term myc-DDK tagged	Origene	MR206415
mPM20D2 cDNA, C-term myc-DDK tagged	Origene	MR222068
mCNDP2 cDNA, C-term myc-DDK tagged	Origene	MR207616
mCNDP1 cDNA, C-term myc-DDK tagged	Origene	MR219018
mPM20D1 cDNA, C-term myc-DDK tagged	Origene	MR219018
pENN.AAV.CB7.CI.ACY1-flag.WPRE.rBG	This paper	Addgene #132686
pENN.AAV.CB7.CI.PM20D2-flag.WPRE.rBG	This paper	Addgene #132684
pENN.AAV.CB7.CI.CNDP2-flag.WPRE.rBG	This paper	Addgene #132685
pENN.AAV.CB7.CI.CNDP1-flag.WPRE.rBG	This paper	Addgene #132683
pENN.AAV.CB7.CI.PM20D1-flag.WPRE.rBG	This paper	Addgene #132682
Software and Algorithms		
Graphpad Prism 8	GraphPad Software	www.graphpad.com
ChemDraw Professional 18	Perkin Elmer	www.perkinelmer.com

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jonathan Z. Long (jzlong@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

HEK293T cells (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS) and penicillin/streptomycin.

Animals

Animal experiments were performed according to procedures approved by the Stanford University Administrative Panel on Laboratory Animal Care (APLAC). Mice were maintained in 12 h light-dark cycles at 22°C and fed a standard irradiated rodent chow diet. For AAV transduction experiments, naïve 7 week old male mice (C57BL/6J) were purchased from Jackson Laboratories. Whole-body CNDP2 knockout mice (catalog number, C57BL/6NCrl-Cndp2em1(IMPC)Mbp/Mmucd, RRID:MMRRC_043492-UCD) were obtained from the Mutant Mouse Regional Resource Center, a NCRR-NIH funded strain repository. Mice were generated at the Mouse Biology Program as part of University of California Davis for the NIH Common Fund program for Knockout Mouse Production and Phenotyping Project (Grant DTCC 9UM1OD023221-06). Both naïve male and female 6-12 week old CNDP2-KO knockout and littermate controls were used for all experiments.

METHOD DETAILS

Production of Adeno-associated Viruses

Full-length mouse *Pm20d1*, *Pm20d2*, *Cndp1, Cndp2*, and *Acy1* cDNA (Origene) with C-terminal 6xHIS and Flag tags were subcloned into pENN.AAV8.CB7.CI.WPRE.rBG (Penn Vector Core) using the PstI/HindIII sites to generate AAV8 plasmids. The corresponding adeno-associated viruses were subsequently produced at the Penn Vector Core. Virus for AAV-GFP (AAV8.CB7.CI.eGFP.WPRE.rBG) was purchased directly from the Penn Vector Core.

LC-MS Based Metabolomics and Data Analysis

Livers from transduced mice were harvested two-weeks post injection and immediately frozen on dry ice. All tissues were stored in -80°C. Liver samples were homogenized in 0.5 ml of a 2:1:1 mixture (by volume) of acetonitrile/methanol/water using a Benchmark

BeadBlaster Homogenizer at 4°C. The mixture was centrifuged for 15,000 rpm (10 min, 4°C) and the supernatant was transferred into mass spec vials for analysis as described below. Mass spectrometry analysis was performed with an electrospray ionization (ESI) source on an Agilent 6545 qTOF LC/MS in negative ionization mode. For QTOF acquisition parameters, the mass range was set from 50 to 1000 m/z with an acquisition rate of 5 spectra/second and time of 250 ms/spectrum. For Dual AJS ESI source parameters, the drying gas temperature was set to 250°C with a flow rate of 12 l/min, and the nebulizer pressure was 20 psi. The sheath gas temperature was set to 300°C with a flow rate of 12 l/min. The capillary voltage was set to 3500 V and the fragmentor voltage was set to 100 V. For separation of polar metabolites, normal-phase chromatography was performed with a Luna 5 μm NH2 100 Å LC column (Phenomenex 00B-4378-E0). Mobile phases were as follows: Buffer A, acetonitrile; Buffer B, 95:5 water/acetonitrile with 0.2% ammonium hydroxide and 50 mM ammonium acetate for negative ionization mode. The flow rate for each run started 100% B for 2 minutes at 0.2 ml/min, followed by a gradient starting at 100% B changing linearly to 50% A / 50% B over the course of 18 minutes at 0.7 ml/min, followed by 50% A / 50% B for 5 minutes at 0.7 ml/min. The raw Agilent .d files were converted to mzXML format and then imported into XCMS software to identify features specifically differentially regulated in each enzyme-transduced group. For each XCMS comparison, the three samples in any given enzyme-transduced group were compared versus all fifteen other samples, and this analysis was performed for each enzyme-transduced group. The following XCMS parameters were used: step = 0.05, mzdiff = 0.05, snthresh = 5, mzwid = 0.1. Features were prioritized based on $|tstat| > 4$ and fold change of enzyme versus all others > 4.

Production of Recombinant Enzymes

The E166A point mutant of mCNDP2 was generated using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions with the mutagenesis primers 5'-GGAGCCAGACTCCGCCATGCCTTCCAG-3' and 5'-CTGG AAGGCATGGCGGAGTCTGGCTCC-3'. Both wild-type and mutant CNDP2 plasmids were transiently transfected into HEK293T cells using PolyFect (Qiagen) according to the manufacturer's instructions. Cells were harvested 2 days later by scraping and lysed using probe sonication. From the cell lysate, flag-tagged enzyme was immunoaffinity purified using magnetic flag-M2 beads (Sigma Aldrich) according to the manufacturer's instructions.

In Vitro Enzyme Activity Assays

In vitro CNDP2-catalyzed hydrolysis of dipeptides was measured by incubating purified CNDP2 protein with dipeptide (10 mM) in 100 ul of 20 nM N-methylmorpholine buffer for 1 hour at 37°C. Reactions were terminated by quenching with 300 ul acetonitrile. Samples were then vortexed thoroughly, centrifuged at 15,000 rpm for 10 minutes, and the supernatant was transferred to a mass spec vial for analysis by LC-MS.

Generation of CNDP2-KO Animals

Cohorts of CNDP2-KO mice and wild-type littermates were generated by heterozygous breeding crosses. Genotyping was performed as follows: genomic DNA from CNDP2-KO and WT mice were extracted from clips of mouse tail that were digested in 100 ul of 50 mM NaOH for 30 min at 100°C and subsequently neutralized by 21 ul of 0.5 M Tris (pH 7.4). PCR was performed using primer pairs to either CNDP2 WT allele (Forward: 5'-GTTTATTAACCAGACTGCCCTC-3'; Reverse: 5'-CTTGGGTGGGCAA GCCCTGGTG-3') or CNDP2 KO allele (Forward: 5'-GCTCTGTAAGGGAAAGAGATGACCC-3'; Reverse: 5'-AATAGGACATACC CAGTTCTGTGAGG-3'). Each 25 ul PCR reaction contained 0.1 ul of 5U/ul Taq DNA Polymerase (Invitrogen), 1.25 ul mice genomic DNA, 2.5 ul of 10 uM forward and reverse primer mix, 0.5 ul of 10 mM dNTP mix, 0.75 ul of 50 mM MgCl2, 2.5 ul of 10X PCR buffer -Mg, and 17.4 ul of ultrapure water. The thermocycling program on BioRad C1000 Touch Thermo Cycler consisted of an initial 30 seconds at 98°C, followed up 42 cycles of 30 seconds at 98°C, 30 seconds at 58°C, and 45 seconds at 72°C, followed by 5 minutes at 72°C and finally held at 4°C. Samples were run on a 2% agarose gel with 0.2 mg/ml EtBr and assessed for presence of a 240 base pair (WT allele) or a 440 base pair (KO allele) PCR product.

QUANTIFICATION AND STATISTICAL ANALYSIS

The student's t-test was used for pair-wise comparisons. Unless otherwise specified, statistical significance was set at $p < 0.05$.

DATA AND CODE AVAILABILITY

The metabolomics dataset generated during this study are available at Mendeley Data, https://doi.org/10.17632/mmk883sjc6.1.