Gut microbiota DPP4-like enzymes are increased in type-2 diabetes and contribute to incretin inactivation

Marta Olivares[1], Paula Hernández-Calderón[2], Sonia Cárdenas-Brito[2], Rebeca Liébana-García[1], Yolanda Sanz[1*], Alfonso Benítez-Páez[1,2*]

[1] Microbiome, Nutrition and Health Research Unit. Institute of Agrochemistry and Food Technology, Spanish National Research Council, IATA-CSIC. 46980 Paterna-Valencia, Spain.

[2] Host-Microbe Interactions in Metabolic Health Laboratory. Principe Felipe Research Center (CIPF). 46012 Valencia, Spain.

Email addresses:

MO email: m.olivares@iata.csic.es

PH-C email: phernandez@cipf.es

SC-B email: sonicard@gmail.com

RL-G email: reliegar@iata.csic.es

YS email: yolsanz@iata.csic.es.

AB-P email: abenitez@cipf.es.

* Correspondence must be addressed to these authors.

Abstract

Background: The gut microbiota controls broad aspects of human metabolism and feeding behavior, but the basis for this control remains largely unclear. Given the key role of human dipeptidyl peptidase 4 (DPP4) in host metabolism, we investigated whether microbiota DPP4-like counterparts perform the same function.

Results: We identified novel functional homologs of human DPP4 in several bacterial species inhabiting the human gut, and specific associations between *Parabacteroides* and *Porphyromonas* DPP4-like genes and type 2 diabetes (T2D). We also found that the DPP4-like enzyme from the gut symbiont *Parabacteroides merdae* mimics the proteolytic activity of the human enzyme on peptide YY, neuropeptide Y, gastric inhibitory polypeptide (GIP) and glucagon-like peptide 1 (GLP-1) hormones *in vitro.* Importantly, administration of *E. coli* overexpressing the *P. merdae* DPP4-like enzyme to lipopolysaccharide-treated mice with impaired gut barrier function reduced active GIP and GLP-1 levels, which was attributed to increased DPP4 activity in the portal circulation and the cecal content. Finally, we observed that linagliptin, saxagliptin, sitagliptin and vildagliptin, antidiabetic drugs with DPP4 inhibitory activity, differentially inhibit the activity of the DPP4-like enzyme from *P. merdae*.

Conclusions: Our findings confirm that proteolytic enzymes produced by the gut microbiota are likely to contribute to the glucose metabolic dysfunction that underlies T2D by inactivating incretins, which might inspire the development of improved antidiabetic therapies.

Keywords: DPP4 enzyme, GLP-1, gliptins, incretins, gut microbiome, obesity, glucose metabolism, type 2 diabetes.

Background

The causal relationship between obesity and comorbidities such as type 2 diabetes (T2D) and the gut microbiota has long been established; however, our understanding of the specific mechanisms driving this relationship remains limited [1]. Microbial mimicry of host activities has been proposed as a mechanism by which pathogens manipulate host cellular functions [2], and recent evidence indicates that mimicry is also a strategy that can be used by commensal bacteria in the gut. For example, *Bacteroides vulgatus* produces the metabolite N-acyl-3-hydroxypalmitoyl-glycine, which can structurally mimic long-chain N-acyl amides that function as mammalian signaling molecules. Both metabolites activate the GPR132/G2A receptor, which has been implicated in autoimmune disease and atherosclerosis [3, 4]. Similarly, the caseinolytic protease ClpB from the gut commensal *Eschericia coli* has been shown to translocate and mimic the action of α-melanocyte-stimulating hormone to control host satiety [5, 6].

Dipeptidyl peptidase 4 (DPP4 or CD26) is a serine protease that rapidly inactivates the incretin peptides glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP) to modulate postprandial insulin secretion and glycemia [7]. Accordingly, administration of DPP4 inhibitors is a pharmacological strategy to control T2D by prolonging the half-life of GLP-1 and GIP [8]. Plasma DPP4 activity increases with obesity, which has been postulated to contribute to reduced incretin activity in the setting of obesity and insulin resistance [9, 10]. However, the generation of mice with loss of DPP4 activity in multiple

tissues and cell types has shown that not all cellular sources of DPP4 activity affect glucose metabolism [11, 12]. For example, endothelial DPP4 cleaves inactive GLP-1 and GIP [11], but adipocyte- and hepatocyte-derived DPP4 are dispensable for glucose control [12]. We recently demonstrated the ability of the gut microbiota to produce DPP4-like activity in germ-free and conventionally colonized mice, leading us to propose that gut microbes are potential mediators of incretin inactivation [13]. Subsequent studies have identified the primary sources of human DPP4 (hsDPP4) homologs in species of the genera *Bacteroides*, *Parabacteroides* and *Porphyromonas* [14-16]. In addition, a peptidase activity complementary to that of DPP4, termed Xaa-Pro dipeptidyl-peptidase (PepX), has been identified in lactic acid bacteria commonly found in fermented foods and the human gut, such as *Lactobacillus* spp., *Lactococcus* spp. and *Streptococcus* spp. [17, 18].

In line with our proposition that gut microbial DPP4 activity might influence host metabolism [13], some species of *Bacteroides* have recently been reported to inactivate GLP-1 in mice fed a high-fat diet [16]. Here, we investigated how bacterial DPP4, which is prevalent in patients with T2D, impairs gut hormone function and evades the catalytic inhibition of drug therapies for T2D. We identified novel gut bacterial species that harbor functional homologs of hsDPP4 and associated bacterial DPP4 genes with T2D and glycemia using metagenomic data from a large cohort of patients [19]. We also demonstrate the potential of bacterial DPP4 to be secreted extracellularly and to hydrolyze human incretins (GLP-1 and GIP) and neuropeptides (PYY and NPY) *in vitro*, which was confirmed in a preclinical model. Finally, we demonstrate the differential interaction between bacterial DPP4 and four prescribed anti-diabetic drugs with DPP4 inhibitory activity. Our results shed light on the molecular mechanisms by which specific gut bacterial

species alter the host glucose metabolism and the efficacy of DPP4 inhibitors, providing new opportunities to improve drug development for the treatment of T2D.

Results

DPP4-like genes are widespread across human gut bacteria

We used a bottom-up approach to search for bacterial DPP4-like genes in the metagenomes of human subjects with metabolic disorders [20], where the presence of the enzymes may have clinical implications. We retrieved DPP4-like genes encoded by different Gram-negative bacteria belonging not only to the genus *Bacteroides*, but also to the genera *Alistipes*, formerly *Prevotella*, *Barnesiella* and *Paraprevotella*. We constructed a phylogenetic tree using amino acid sequence information from more than 1,600 non-redundant bacterial peptidases, including PepX (belonging to the so-called S15 family) and DPP4-like and eukaryotic proteins (family S9). After applying advanced phylogenetic methods, we obtained a maximum likelihood tree explaining the evolutionary relationship between the two peptidase families (Figure 1A). According to Akaike's Information Criterion (AIC), the WAG+I+G model of protein evolution better explained the relationship between the S15 and S9 peptidase families and determined both as independent lineages. Notably, PepX proteins were found exclusively in Gram-positive bacteria whereas DPP4-like proteins were found exclusively in Gram-negative bacteria. Taxonomic assessment revealed that PepX peptidases are predominantly found in the families Streptococcaceae and Lactobacillaceae, and their respective genera, within the Bacillota phylum (formerly Firmicutes) (Figure 1A). The families Enterococcaceae, Leuconostocaceae and Sporolactobacillaceae were also represented to a minor extent.

Contrastingly, DPP4-like proteins were more widely distributed across several families of the Bacteroidota phylum (formerly Bacteroides), and were evident in several species from the Flavobacteriaceae, Prevotellaceae, Porphyromonadaceae and Bacteroidaceae families (Figure 1A). Furthermore, the DPP4-like family proteins showed greater diversity and genetic distances between members, and proteins from Flavobacteriaceae members appeared to have evolved independently of other Bacteroidota families. Overall, the amino acid sequences and the evolutionary relationships between the S15 and S9 peptidase families support a process of functional convergence [21]. Moreover, we found proteins corresponding to the three eukaryote proteins (humans, rat and mouse), with identity values of 25–35% (within the *twilight zone* of sequence similarity) (Figure 1B). The position of the eukaryotic DPP4 enzymes in the tree seem to not support a phylogenetic relationship with bacterial DPP4-like enzymes, suggesting that they did not diversify from a common ancestor. Indeed, the closest relationship between human, rat and mouse DPP4 enzyme sequences was established with DPP4-like sequences from environmental bacterial species of the Balneolaceae family (genera *Gracilimonas* and *Aliifondinibius*), which are marine species with a high frequency of horizontal gene transfer [22].

Analysis of the protein architecture of the bacterial DPP4-like proteins revealed a leader peptide similar to that of the human counterpart (Supplementary material S1A-C), which is involved in transmembrane translocation, suggesting extracellular secretion. Furthermore, the amino acid profile of all the analyzed DPP4-like proteins (bacterial and eukaryote, N = 971) indicated that all proteins contain the same catalytic motif of S9 serine proteases (GWSYGGY) according to the MEROPS database (Supplementary material S1D) [23].

Increased abundance of microbial DPP4-like genes is associated with T2D

To investigate the association between microbial DPP4 and T2D, we evaluated the abundance of all retrieved DPP4-like bacterial genes in the metagenome of 220 subjects with T2D from a previous study [19]. A GLMM analysis controlling for covariates indicated that some DPP4-like genes were more abundant in patients with T2D than in non-T2D controls (Figure 1C). The largest differences were detected for DPP4-like genes from *Parabacteroides* sp. (DPP4_927, GLMM-adj = 0.40, p-adj = 0.009), *Chryseobacterium* sp. YR203 (DPP4_609, GLMM-adj = 0.33, p-adj < 0.001), *Parabacteroides merdae* (DPP4_931, GLMM-adj = 0.31, p-adj < 0.001), *P. timonensis* (DPP4_302, GLMM-adj = 0.29, p-adj = 0.006) and *Parabacteroides* sp. SN4 (DPP4_249, GLMM-adj = 0.27, p-adj = 0.023). Indeed, a genus-level (taxa grouped) enrichment approach revealed that DPP4-like genes are encoded essentially by species of two genera in patients with T2D: *Parabacteroides* (*Chi-squared* = 122.9, p = 5.0e-4) and *Porphyromonas* (*Chi-squared* = 25.8, p = 0.003). Contrastingly, DPP4-like genes encoded by species of the genera *Bacteroides* and *Prevotella* were less abundant in patients with T2D (Figure 1C). We then used a machine learning approach (sPLS-DA) to question whether the DPP4-like genes could be used as a classification trait in T2D. Robust models were built using two and three more informative features from the DPP4-like dataset, which revealed that the genes from *Parabacteroides timonensis* (DPP4_302), *Parabacteroides* MULTISPECIES 99% (DPP4_927) and *Porphyromonas catoniae* (DPP4_761) had high discriminatory power when evaluated in combination (AUROC = 0.823, p = 6.7e-16) (Figure 1D).

Bacterial DPP4 is secreted like its human counterpart

To understand the extent to which the function of bacterial DPP4-like proteins is similar to hsDPP4, we cloned the DPP4 gene from *P. merdae*, the member of the gut microbiome that is more strongly associated with T2D and the *Parabacteroides* species that is more abundant in the human gut [24]. We cloned the full-length DPP4-like gene (pmDPP4) and a truncated version lacking the predicted leader sequence of 23 amino acids at the N-terminus (pmDPP4Δ23) into a His-tag expression vector, and confirmed over-expression of the recombinant proteins in *E. coli* BL21-DE3 cells (Supplementary material S2A). Mass spectrometry analysis revealed that cell-free supernatants from *E. coli cultures* expressing pmDPP4, but not those expressing pmDPP4Δ23, could cleave GLP-1, as reflected by the expected shift in the molecular mass of GLP-1 from 3354 Da to 3147 Da resulting from the cleavage of the His-Ala N-terminal dipeptide (Figure 2A-C). We next investigated the subcellular localization of pmDPP4 and pmDPP4Δ23 proteins in bacterial cultures using anti-His immunogold staining and transmission electron microscopy. Results showed that full-length pmDPP4 tended to localize to the inner membrane and was also secreted and embedded in the extracellular polysaccharide capsule [25], whereas pmDPP4Δ23 tended to localize in the cytosol (Figure 2D-F).

To confirm the role of the pmDPP4 signal peptide in directing secretion and allowing extracellular GLP-1 hydrolysis, we performed an *in vitro* enzymatic assay using the synthetic dipeptide Gly-Pro p-nitroanilide (gpPNA) as a substrate. We also measured the rate of cell lysis by titrating lactate dehydrogenase (LDH) activity during the preparation of cell-free supernatants, to control for spontaneous DPP4-like activity. We were unable to detect bacterial cell lysis during the reaction, and thus the increase in DPP4-like activity observed in *E. coli* BL21-DE3 cells expressing full-length pmDPP4 (but not in cells

expressing pmDPP4Δ23) was likely mediated by secretion of the accumulated enzyme into the extracellular medium (Supplementary material S2B). Thus, the leader peptide, which is evolutionarily conserved in bacterial DPP4-like proteins, is likely required for protein secretion and favors interaction with their substrates.

Parabacteroides merdae DPP4-like enzyme hydrolyzes human incretins and neuropeptides in vitro

As our data suggested that the pmDPP4Δ23 protein was likely to be the mature form produced and secreted by *P. merdae*, we tested additional functional aspects of the protein. Gel filtration analysis indicated that pmDPP4Δ23 dimerizes in solution to a protein of ~130 kDa, similar to the quaternary structure of the human counterpart (Supplementary material S2C) [26]. To assess whether the pmDPP4Δ23 protein was similarly active to the full-length protein produced and secreted by *E. coli* BL21-DE3 cells, we determined its hydrolytic activity (Supplementary material S2C,D). As a qualitative approach, we tested whether pmDPP4Δ23 could cleave the incretin and neuropeptide targets of hsDPP4 in a comparative assay with hsDPP4 by MALDI-TOF-MS (Supplementary material S2D). Under equimolar conditions and equal substrate concentrations, the pmDPP4Δ23 protein cleaved GLP-1, GIP, PYY, and NPY peptides as efficiently (and identically) as hsDPP4 (100% substrate hydrolysis) (Figure 3). Specifically, GLP-1 and GIP peptides were completely processed by pmDPP4Δ23 at the first two N-terminal amino acids (His-Ala and Tyr-Ala, respectively), consistent with the molecular mass shift (Figure 3A,B). pmDPP4Δ23 also hydrolyzed the anorexigenic hormone PYY and the neuropeptide NPY by removing the Tyr-Pro N-terminal dipeptide in both cases, as for hsDPP4 (Figure 3C,D).

Parabacteroides merdae DPP4-like enzyme inactivates GLP-1 and GIP in vivo

To investigate whether the DPP4-like enzyme cleaves and inactivates incretins *in vivo*, we administered mice with *E. coli* BL21-DE3 carrying the plasmid encoding full-length pmDPP4 or the empty vector. To mimic obesity-associated low-grade inflammation and gut barrier defects, mice were injected intraperitoneally with lipopolysaccharide (LPS) to stimulate metabolic endotoxemia, which would facilitate the translocation of bacterial products from the intestinal lumen (Figure 4A). We observed significantly higher DPP4 activity in the cecal content and portal vein of the group receiving *E. coli* carrying the bacterial DPP4 (Figure 4B,C). Supporting our hypothesis, the increase in DPP4 activity in the portal circulation was associated with a significant reduction in the active forms of GLP-1 and GIP (Figure 4D,E). We excluded the existence of differences in hormone secretion, as no changes were found in total GLP-1, its precursor proglucagon or total GIP (Figure 4F-H). There were also no differences in the total form of other hormones, such as PYY, or in insulinemia, or glycemia (Figure 4I-K).

P. merdae DPP4-like enzyme is differentially inhibited by gliptins

Finally, we evaluated the inhibition of bacterial DPP4-like and hsDPP4 proteins by currently approved gliptins: sitagliptin, vildagliptin, saxagliptin and linagliptin. Lineweaver-Burk double-reciprocal plots of the kinetics showed that all inhibitors showed classical competitive inhibition of hsDPP4, affecting the K_m but not the V_{max} (Supplementary material S3). Overall, pmDPP4Δ23 had an estimated K_m of 1.63 mM, representing on average a 3.97 fold-change ($p < 0.001$) over the observed hydrolysis rate for hsDPP4 (Km = 0.33 mM) (Supplementary material S4). As expected, the enzymatic activity hsDPP4 was inhibited by all gliptins, with sitagliptin, saxagliptin and linagliptin

having a more significant effect on its dipeptide hydrolysis function (Supplementary material S4). Contrastingly, bacterial pmDPP4Δ23 was only inhibited by saxagliptin and vildagliptin, which reduced its hydrolytic capacity on gpPNA (fold-change = 1.73, p = 0.036 and fold-change = 0.78, p = 0.076, respectively, at the highest inhibitor concentration) (Supplementary material S4). The determination of EC_{50} values confirmed the poor effect of gliptins on pmDPP4 function compared with hsDPP4 (Figure 5). The EC_{50} for all gliptins against hsDPP4 was in the nanomolar range ($< 10^{-6}$) with linagliptin showing the strongest inhibitory activity (3.68×10^{-9}). Conversely, pmDPP4Δ23 was only inhibited by vildagliptin and saxagliptin, but always to a lesser extent than hsDPP4 (2-fold and 2.4-fold less, respectively) (Table 1). Non-linear fitting of the dose-response data showed an almost null effect of sitagliptin and linagliptin on pmDPP4Δ23, leaving its hydrolytic activity intact (Figure 5).

Discussion

DPP4 activity, present on the surface of mammalian cells and as a soluble form in plasma, is crucial in regulating host metabolism by cleaving and deactivating the incretins that signal insulin release. The present study makes an advance by showing that the prevalence of bacterial DPP4-like genes correlates with human metabolic status in several metagenomic analyses. In addition, our *in vitro* and *in vivo* studies with the DPP4 of *P. merdae* provide a mechanistic rationale for the findings of our metagenomic assessment and argue for the detrimental impact of bacterial DPP4 on human health.

To best characterize the gut microbiota as a source of DPP4-like activity, we first searched for one of the millions of genetic traits encoded in the human gut microbiome based on its distant similarity to the human *DPP4* or *CD26* gene product. We show that human DPP4 has distant sequence-based homologs that are widely distributed in the genome of hundreds of Gram-negative bacterial species, some of which are prevalent members of the human gut microbiota. We demonstrate that the products of such bacterial genes are more similar to eukaryote DPP4 proteins (e.g., human, rat, and mouse) than to the bacterial XaaPro/PepX peptidases (so-called S15 family). The PepX family comprises a set of proteins often referred to as DPP4 enzymes, which has led to confusion in distinguishing the true DPP4-type enzymes probably because of their similar activity to trim X-A or X-P N-terminal dipeptides in many substrates [23, 27]. In the latter respect, the conserved and consensus GWSPGGF motif (GWSYGGY in eukaryotes) present in the bacterial DPP4 homolog has led to the assignment of this peptidase to the S9B family of serine peptidases according to the MEROPS database [23]. Our results are in agreement with recent findings showing the presence of DPP4-like enzymes in several species of *Bacteroides* [14, 16].

We have also addressed one of the criticisms we received when our hypothesis paper was published, which questioned the ability of bacteria to secrete DPP4 [28]. In this sense, our new data suggest that bacterial DPP4 proteins contain N-terminal signal peptides that potentially drive secretion, and are indeed similar to the transmembrane domains present in the hsDPP4 amino acid sequence. Using various cell-based assays and microscopy approaches, we show that the leader sequence of bacterial DPP4 proteins, exemplified by the 23 amino acid N-terminal peptide from *P. merdae*, serves as a signal for secretion. We conclude that, similar to hsDPP4, the bacterial DPP4-like protein tends to be located

extracellularly rather than intracellularly as for PepX peptidases [29, 30]. In addition, bacterial DPP4-like proteins share other features with the human enzyme, such as similarities in the amino acid sequence (~30%), domain architecture, quaternary structure and catalytic motifs [26, 31].

Given the known role of hsDPP4 as a regulator of glucose homeostasis, we questioned the value of DPP4 expression in the gut microbiome. Our metagenomic assessment suggests that DPP4-like genes from *Parabacteroides* and *Porphyromonas* species may play an important role in T2D, as they are enriched in patients with T2D. Our results are in accord with the taxonomic evaluation performed in one of the largest cohorts of individuals with T2D published to date, where *Parabacteroides* appears as one of the genera enriched in T2D subjects [19]. Much less is known about the effect of *Porphyromonas* on host metabolism, likely because its predominant habitat is the oral cavity. Nevertheless, some studies have provided information on the metabolic deterioration caused by *Porphyromonas gingivalis*, with recently reported DPP4-like activity [16], when administered to mice [32]. We hypothesize that the mechanism underlying our results, as well as those reported by others, lies in the functional overlap of DPP4 activity within eukaryotic cells and gut microbes. Indeed, we show *in vitro* that the *P. merdae* DPP4-like enzyme inactivates GLP-1 and GIP as efficiently as hsDPP4. In mice sensitized by LPS, we confirm that the administration of *E. coli* overexpressing the full-length DPP4-like gene from *P. merdae* reduces the active forms of GLP-1 and GIP, which is associated with higher DPP4 activity in portal blood. Very recently, Wang et al. reported similar results regarding the potential of the DPP4-like activity encoded by *Bacteroides thetaiotaomicron*, *Bacteroides fragilis* and *Bacteroides vulgatus* to inactivate GLP-1 in mice fed a high-fat

diet for 12 weeks, but not in mice fed a control diet [16]. Notably, both studies agree that the gut microbiota contributes to DPP4 activity and the decrease of active GLP-1 when the gut barrier is disrupted, either by a high-fat diet or by LPS [16]. In addition to the inactivation of incretins, we also provide evidence for the ability of bacterial DPP4-like to process other gut peptides *in vitro*, including PYY and NPY, which would have further implications for host behavior and mental health. For example, full-length PYY binds with similar affinity to all Y receptors, but the predominant circulating form of PYY-immunoreactivity, truncated $PYY_{3–36}$, preferentially binds to the Y2 receptor [33], a G protein-coupled receptor implicated in appetite and feeding control, mood and mental disorders [34]. By degrading NPY_{1-36} to NPY_{3-36}, DPP4 increases the affinity of NPY for Y2 and Y5 receptors but reduces its activity at Y1 receptors, which are known to be responsible for the antidepressant effects of NPY [35]. Accordingly, in mice fed an obesogenic diet, increases in DPP4 activity in systemic circulation have been associated with a depression-like behavior [36], an effect that is counteracted by sitagliptin [37]. Unfortunately, the lack of commercial kits to measure the cleaved forms of PYY and NPY makes their *in vivo* modification by the microbiota a matter for future studies.

The existence of gut bacteria with DPP4-like activity rationalizes the evaluation of their interaction with antidiabetic drugs with inhibitory DPP4 activity. Sitagliptin and vildagliptin were the first drugs to be approved and are the most widely prescribed [8]. Our results indicate that DPP4 from *P. merdae* was minimally inhibited by sitagliptin, in line with previous studies on the DPP4-like enzyme from *B. thetaiotaomicron* [16]. By contrast, vildagliptin and saxagliptin have an effect on the DPP4 activity of *P. merdae*, with EC_{50} values in the nanomolar range. For vildagliptin, concordant results in mice showed that the

drug abolished DPP4 activity in the contents of the cecum and feces [38]. However, our results contrast with the recently characterized DPP4-like enzyme from *B. thetaiotaomicron*, which is inhibited in the nanomolar range (10^{-8} range) by saxagliptin and in the micromolar range by sitagliptin (10^{-6} range) and linagliptin (10^{-5} range) [14]. These different inhibition patterns could be explained by the extensive sequence diversity of bacterial DPP4-like enzymes, resulting in a large repertoire of proteins with differing abilities to bind gliptins and affect their catalytic activity. Therefore, characterizing the function of DPP4-like enzymes from different gut bacterial species is crucial to elucidate their impact on T2D progression and gliptin therapy. Indeed, the information obtained from our inhibition studies could have significant consequences for the pharmacology of T2D treatment. For instance, the partial inhibition of the *P. merdae* DPP4-like activity by vildagliptin and saxagliptin suggests that bacterial enzymes can bind DPP4 inhibitors regardless of the inhibition efficacy. Given the substantial prevalence of DPP4-like enzymes in the intestinal lumen, this phenomenon could significantly affect the bioavailability of drugs targeting the human enzyme [39]. Furthermore, we and others have demonstrated that sitagliptin does not inhibit bacterial DPP4, highlighting the limitation of this drug to prevent incretin inactivation from all possible enzymatic sources. This observation likely explains the differential response to sitagliptin recently reported in individuals with T2D [16].

Conclusions

We show that the gut microbiome encodes a large number of DPP4-like functional homologs, the prevalence of which positively correlates with T2D status. The mechanism

underlying this association might be functional mimicry of bacterial DPP4-like enzymes, as exemplified by the ability of *P. merdae* DPP4-like enzyme to inactivate incretins. Our results of the characterization of the inhibitory activity of gliptins on bacterial DPP4-like enzymes may serve as a starting point for redefining the efficacy of current DPP4 inhibitors, as we show that some DPP4 inhibitors (vildagliptin and saxagliptin) are compromised by nonspecific binding to bacterial enzymes and others, such as sitagliptin, by the null effect on bacterial DPP4-like enzymes.

Material and methods

Search strategy for bacterial DPP4 homologs in the gut microbiome

We searched for bacterial homologs of hsDPP4 in metagenome-assembled genomes from our previous study on people with overweight and the metabolic syndrome [20]. We used a Blast-based approach against human (SwissProt id: P27487), mouse (P28843) and rat (P14740) DPP4 amino acid sequences. Thresholds were set to at least 85% coverage in length and >25% sequence identity at the protein level.

Phylogeny tree of bacterial DPP4 and X-prolyl dipeptidyl peptidase families

The amino acid profile (*via hmmer3*) of the aligned DPP4-like (S9 peptidase family) sequences detected in the metagenome of overweight subjects with metabolic syndrome was used as bait to compile the most complete set of similar proteins present in the RefSeq NCBI database through PSI-Blast [40]. In similar manner, the PepX (S15 peptidase family) from *Lactobacillus acidophilus* NCFM (UniProt id Q5FJC6) was used to retrieve the most complete set of similar enzymes potentially present in bacterial genomes. Multiple PSI-Blast iterations were run to select aligned protein sequences with coverage of >95% of query length and ≥25% amino acid identity. Redundancy was removed to retain sequences at 99% identity. Finally, we obtained 925 non-redundant amino acid sequences of bacterial DPP4 homologs and 671 sequences for the PepX family of peptidases. Human (hsDPP4), mouse (mmDPP4) and rat (rnDPP4) sequences were merged with those previously obtained from bacterial metagenes as well as with the protein P19577 from *Dichelobacter nodosus* (member of the S17 peptidase family), the outlier for phylogeny reconstruction. In total, we performed a multiple sequence alignment of more than of >1,600 protein sequences, using an iterative process with refinement steps [41, 42]. Amino acid profiles and consensus

sequences for both peptidase families were computed using the Hidden Markov Model approach and the *hmmer3* package [43, 44]. The best tree explaining the evolutionary relationships between the two types of peptidases was built using ProtTest v3.0 [45], by selecting the best of the 120 evolutionary models implemented, according to AIC. Tree topology was optimized by using the maximum likelihood method. The amino acid sequences derived from hsDPP4 and the consensus amino acid sequences obtained for S15 and S9 peptidase families using *hmmer3* were subjected to further analysis to depict protein architecture and functional domains using the SMART server [46]. In a similar manner, the SignalP v4.1 server was used to predict the presence of signal peptides in the amino acid sequences (http://www.cbs.dtu.dk/services/SignalP-4.1/) [47], whereas the TMHMM server was used to predict the presence of transmembrane helices (http://www.cbs.dtu.dk/services/TMHMM/).

Cloning, over-expression and purification of bacterial DPP4

The *P. merdae* DPP4-like gene (pmDPP4) was amplified from DNA obtained from the type strain *P. merdae* DSM19495^{T} (DSMZ, Braunschweig, Germany), using the primers DPP4-PMER-F gataccatggtgaaacgattagggtttggtgcttta (with an NcoI restriction site, underlined) and DPP4-PMER-R gataggatccttatta*gtgatggtgatggtgatg*caaattatcaaataaaaagttagacat (with a BamHI restriction site, underlined) and a 6×His tag (italics) to produce a C-terminal tagged protein. Additionally, an open reading frame (ORF) encoding an N-terminal 6×His-tagged pmDPP4 protein with no signal peptide (pmDPP4Δ23) was amplified from human fecal DNA samples using the primers DPP4-PMER-F23 gataccatggga*catcaccatcaccatcac*agcaaacgtgtcgatttaaaagaaattaca and DPP4-PMER-R2

410 gataggatccttacaaattatcaaataaaaagttagacat. The PCR products of (~2.5 kb) were purified,
411 mixed, and 500 ng of DNA was digested with 5U NcoI and BamHI restriction enzymes
412 (New England Biolabs, Ipswich, MA) overnight at 37ºC. Digested fragments were cloned
413 into the BamHI/NcoI pre-digested pET15b plasmid using the Fast-Link™ DNA Ligation
414 Kit (Epicentre, Madison, WI), and *E. coli* DH5α cells were transformed with the ligation
415 products. Colonies were selected on LB agar with 100 µg/mL ampicillin (Sigma, Madrid,
416 Spain) and checked by PCR for the presence of the correct plasmids. The inserts were
417 analyzed by Sanger sequencing to ensure in-frame position of the ORFs and tags. The
418 plasmids encoding pmDPP4 and pmDPP4Δ23 ORFs were isolated from DH5α cells and
419 transferred to *E. coli* BL21-DE cells by heat-shock transformation. To over-express the
420 recombinant proteins, freshly transformed *E. coli* BL21-DE cells containing the respective
421 plasmids were grown on NZY Auto-induction LB medium (NZYTech, Lisbon, Portugal) at
422 37ºC for 20 h. Then, bacterial cells were recovered by centrifugation at 5,000 *g* for 25 min
423 at 4ºC and washed and resuspended in 10 mL of TBS buffer pH 8.0. Cells were disrupted
424 by a pre-incubation with 500 µg lysozyme (Sigma) at room temperature for 10 min,
425 followed by 4 cycles of bead-beating (FastPrep-24™ 5G bead beating grinder; MP
426 Biomedicals, Santa Ana, CA) of 30 sec, using 0.1 mm glass beads. The supernatant was
427 obtained by centrifugation of the cell lysate at 16,000 *g* for 20 min at 4ºC, and was then
428 incubated with 2 mL TBS pH 8.0 pre-equilibrated Talon® Metal Affinity resin during 2 h
429 at 4ºC with permanent mixing (Takara Bio, Kusatsu, Japan). The resin was washed with 10
430 volumes of TBS pH 8.0 buffer containing 0.1% Triton X-100 and 10 mM imidazol (both
431 from Sigma). Elution was performed with 10 resin volumes of TBS pH 8.0 containing 150
432 mM imidazol. Proteins were concentrated by filtration of the eluate in 15-mL Amicon®

Ultra 50k filters (Millipore, Burlington, MA). Subsequently, a concentration step was performed in 0.5-mL Amicon® Ultra 50k filters, retaining a final volume of 30–50 μL. Purity and concentration of the recombinant proteins was assessed by SDS-PAGE and the Qubit Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA). Gel filtration was used to recover the recombinant protein fraction according to its expected size on an AKTA Ettan LC FPLC System (Amersham Pharmacy Biotech, Amersham, UK), with a Superdex 200 5/150GL column and Ca/Mg-free sterile PBS pH 7.4 (MP Biomedicals, Aurora, OH) as the mobile phase.

Immunogold staining and electron microscopy

E. coli BL21-DE cells containing the pET15b/pmDPP4 or pET15b/pmDPP4Δ23 plasmids were cultured in LB medium and induced to express the recombinant proteins as described above. *E. coli* BL21-DE cells harboring the pET15b were equally treated and used as a negative control for immunogold staining. After the induction of protein expression induction, 0.5 mL of cells were centrifuged at 3,000 *g* and washed with 1 volume of PBS. Cells were pelleted and sent to the Electron Microscopy Core Facility at SCSIE (University of Valencia) for fixation, embedding and immunogold staining using a 6×His monoclonal antibody (Takara Bio) diluted 1/5000 as the primary antibody, and 10 nM anti-mouse IgG-Gold antibody (Sigma) as the secondary antibody, diluted 1/20. Images were acquired on a JEM-1010 transmission electron microscope (JEOL, Tokyo, Japan) operating at 100 KV equipped with an AMT RX80 (8 Mpx) digital camera. Image analysis was performed at the SCSIE.

Protein secretion assay

A colorimetric assay was developed to detect pmDPP4 secretion dependent on the presence of the 23 amino acid leader peptide. *E. coli* BL21-DE3 cells containing the plasmids pET15b/pmDPP4 or pET15b/pmDPP4Δ23 were grown overnight at 37ºC in LB medium with 100 μg/mL ampicillin (Sigma). Overnight cultures were 1:100 refreshed in 5 mL ampicillin-containing LB medium and grown for a further 3 h at 37ºC with shaking. For recombinant protein over-expression, we added 0.1 mM IPTG (isopropyl β d-1-thiogalactopyranoside) and 100 μg/mL ampicillin to cultures grown under the same condition for a further 3 h. After protein induction, 1 mL of culture was recovered in a 1.5-mL sterile microtube and cells were pelleted at 3000 *g* for 10 min at 4ºC. Pellets were resuspended in 1 mL of cold PBS and incubated on ice for 5 min. From the suspension, 100 μL was removed to estimate the colony formation unit (CFU) count on LB-ampicillin plates. The remaining cell suspensions were again pelleted and finally resuspended in 200 μL of cold PBS. The secretion assay consisted of adding 450 μL PBS to 50 μL cellular suspension, mixing, and incubating at 37ºC with 600 rpm shaking. One hundred microliters of the suspension was collected in new 1.5-mL sterile microtubes at 0, 30, 60, and 90 min of incubation and were maintained on the ice up until the end of the collection. To obtain cell-free supernatants, all aliquots were centrifuged at 6000 *g* for 10 min at 4ºC and supernatants were recovered in 1.5-mL microtubes and stored on ice until processing. The cell-free supernatants were used to determine DPP4 activity secreted into the medium according to the protocol described below. Bacterial cell lysis was determined in cell-free supernatants using the LDH Cytotoxicity Assay Kit (Thermo Fisher Scientific, Waltham,

MA). Both DPP4 and LDH activities obtained from cell-free supernatants were normalized to the CFU counts.

Colorimetric measurement of DPP4 activity and inhibition assays

The hydrolytic activity of the bacterial DPP4 enzyme was measured using Gly-Pro p-nitroanilide hydrochloride (gpPNA) (Sigma), as described for assessing hsDPP4 activity [38]. A range of substrate concentrations from 2.5 to 0.31 mM gpPNA in 50 mM Tris-HCl pH 8.3 was used for kinetic analysis. The standard curve was established with free para-nitroaniline (PNA) (Sigma) ranging from 2 mM to 0.031 mM in 50 mM Tris-HCl pH 8.3. Briefly, 20 µL of each cell-free supernatant was placed in triplicate in a flat 96-well plate. Two hundred microliters of the respective PNA solutions were plated in duplicate to obtain the standard curve. A total of 180 µL of substrate solution (gpPNA) was added to bacterial supernatants and the enzymatic activity was monitored for 30 min at 37°C (380 nm) in a VICTOR2 fluorometer/spectrophotometer (Perkin Elmer, Waltham, MA). The enzymatic activity ([PNA generated (nM) / hydrolysis time (min)] / pmole enzyme) was quantified between 10 and 30 min of incubation time (linear kinetics). Three independent experiments with different purified protein batches were performed. The same protocol was followed for the murine samples (cecal content and plasma from the portal vein). In the case of the cecal content, the values of DPP4 activity were normalized by the amount of protein, quantified by the Bradford method (Bio-Rad, Hercules, CA).

The commercially available FDA-approved human DPP4 inhibitors sitagliptin, saxagliptin, linagliptin and vildagliptin (all from Sigma) were diluted in PBS to a 2 µM working

solution. The inhibitory activity of the four gliptins on hsDPP4 and pmDPP4Δ23 was assessed at the ratios 1:10, 1:50, and 1:100 purified protein:inhibitor. Four independent replicates with different purified protein batches were performed. Dose-response assays to determine the EC_{50} were carried out using serial dilutions of the inhibitors from 1 μM to 3.91 nM, and the relative activity was calculated with respect to reactions lacking the inhibitor.

Peptidase activity of DPP4-like enzymes on gut hormones

The DPP4 substrates GLP-1 7-37 fragment, GIP, NPY and PYY (all from Sigma) were diluted to 2.5 μg/μL in Ca/Mg-free sterile PBS pH 7.4. Human DPP4 (Sigma) was used as a control in PBS at 50 ng/μL (0.585 μM). The *in vitro* peptidase activity was assayed in a 10-μL reaction containing 5 μg substrate (100-150 μM) and 10 nM of hsDPP4 or pmDPP4Δ23 recombinant proteins (molar ratio substrate:enzyme, 10000:1), mixed on ice and incubated at 37ºC for 10 min with gentle orbital shaking (600 rpm). To test the activity of the pmDPP4 secreted protein, the reaction was the same but the purified recombinant protein was replaced with the same volume of *E. coli* BL21-DE cells carrying pET15b, pET15b/pmDPP4, or pET15b/pmDPP4Δ23 plasmids incubated with 0.1 mM IPTG for 3 h at 37ºC. The cells were washed and diluted in PBS (~5×10^6 cells per reaction). Inhibition assays were performed under similar conditions but in the presence of 500 nM of vildagliptin or sitagliptin. Hydrolysis reactions were inactivated by incubation at 90ºC for 10 min. For reactions containing cells, the cells were first collected by centrifugation and then heat-inactivated. The inactivated reactions were stored on ice until processed for MS.

MALDI-TOF analysis

The MS analysis was performed at the SCSIE proteomic service (University of Valencia). Briefly, for every *in vitro* reaction performed, 1 µL of each sample was spotted onto the MALDI plate and, after the droplets were air-dried at room temperature, 1 µL of matrix (10 mg/mL alpha-cyano matrix in 70% acetonitrile, 0.1% trifluoroacetic acid) was added and allowed to air-dry at room temperature. The resulting mixtures were analyzed on a 5800 MALDI TOF/TOF instrument (AB Sciex, Framingham, MA) in positive reflector and linear mode. The raw mass spectra were released in ASCII format and were analyzed using *mMass* v5.5 software (http://www.mmass.org/).

Bacterial DPP4-like in T2D human metagenomes

A total of 220 samples (metadata available) derived from one of the largest metagenome projects on T2D (BioProject PRJNA422434) were evaluated for the distribution of bacterial DPP4-like genes. For read mapping against the non-redundant DPP4-like peptide database (N = 971), we used the *usearch* v8.0.1623 algorithm [48] with the following parameters: *-usearch_local*, *-id 0.7*, *-strand both*, and *-top_hit_only*. Alignments were filtered to retain those with more than 66% read length. Differential abundance of DPP4-like genes was assessed using Generalized Linear Mixed Models (GLMM, *lme4::glmer* R function) with prior filtering for low abundant features (*CoDaSeq::codaSeq.filter*, *min.occurrence*=0.25, *min.reads*=1), execution of *zCompositions::cmultRepl* function supporting a Bayesian-Multiplicative replacement of zero counts, and centered log-ratio (clr) transformation of compositional microbial abundance data (*CoDaSeq::codaSeq.clr*). Changes in DPP4-like gene abundance were assessed individually using sex, age, and body mass index (BMI) as covariates in the GLMM model. Genus enrichment analysis on significantly abundant

DPP4-like genes in subjects with T2D (N = 135) compared to non-T2D individuals (N = 85) (from same cohort) was evaluated using the *Chi-square* test with Monte-Carlo simulation. Lastly, the DPP4-like gene abundance from T2D (N = 135) and no-T2D (N = 85) samples was compiled in an integrative analysis supported by sparse Partial Least Square-Discriminant Analysis (sPLS-DA) algorithms implemented in the *DIABLO* v6.10.8 package [49]. A tuned sPLS-DA model was built using 4-fold cross-validation (limiting over-fitting) across multiple iterations (*nrepeat* = 20). Selection of reliable models explaining variation was based on their performance (AUROC ≥ 0.70).

Animal study

Experiments were performed using 16 male C57BL/6J mice (7 weeks of age; Charles River Laboratories, Écully, France). Mice were housed in individually-ventilated cages (4 mice/cage) in a Biosafety level 2 laboratory with a 12-h light/dark cycle and temperature-controlled conditions (23 ± 2°C). Mice were acclimatized for ten days and had *ad libitum* access to food and water. Mice were randomized to receive for five consecutive days *E. coli* BL21-DE cells carrying the empty expression plasmid pET15b (3×10^8 UCF/day) (N = 8) or carrying pET15b/pmDPP4 (3×10^8 UCF/day) (N = 8). Both groups were also treated with LPS from *E. coli* O111:B4 (0.1 mg/kg, dissolved in sterile PBS) to induce intestinal permeability, as previously described [50]. Mice were fasted on day 5. Mice then received an oral load of Ensure Plus® (0.0124 kcal/g) to induce the release of gut hormones [51]. After 12 min, mice were anesthetized with isoflurane to sample blood from the portal vein, which was collected in EDTA-containing tubes with or without a DPP4 inhibitor (Millipore, Burlington, MA). The plasma was recollected after centrifugations at 12,000 *g* for 3 min. Mice were sacrificed by cervical dislocation to minimize potential confounding

effects, and mice of the different experimental groups were interspersed. The cecal content was also collected and snap-frozen in liquid nitrogen. All samples were stored at -80°C. In plasma samples containing the DPP4 inhibitor, the levels of active and total GLP-1 and total GIP, PYY, insulin and proglucagon were measured using the Luminex™ Mouse Metabolic Hormone Expanded kit (Merck & Co., Inc. Kenilworth, NJ). Also, active GIP was measured using an ELISA kit (CrystalChem, Zaandam, The Netherlands). Glycemia postmortem was quantified using the Glucose Colorimetric Detection Kit (ThermoFisher Scientific). DPP4 activity was measured in an aliquot of plasma without DPP4 inhibitor and in the cecal content and feces, as described above.

Statistical analysis

A normality test (Shapiro-Wilk) was conducted on data to select the most appropriate statistical comparison test (parametric or non-parametric) between sample groups and treatments. The *t*-test or Wilcoxon-Rank Sum test was applied to compare the *in vitro* enzymatic activity of hsDPP4 and pmDPP4 and the biochemical and physiological data from the animal study. A GLMM with covariate controls (e.g., age, sex and BMI) was used to evaluate the differential abundance of DPP4-like genes in patients with T2D. The sequencing project was also included as a covariate when multiple cohorts of metagenome samples were assessed simultaneously. The sPLS-DA was used to quantify the power of bacterial DPP4-like genes for the classification of T2D samples, using a 4-fold cross-validation. All tests were applied in R v4.1.2. Estimation statistics and plots were obtained using the *dabestr* v0.3.0 R package (Gardner-Altman plots) or GraphPad v9 (San Diego, CA).

Declarations

Ethics approval:

The animal experiment was performed following European Union 2010/63/UE and Spanish RD53/201 guidelines, approved by the ethics committee of the University of Valencia (Animal Production Section, SCSIE, University of Valencia), and authorized (Ref 2022/VSC/PEA/0244) by the competent authority (Generalitat Valenciana).

Availability of data and material.

The raw metagenomic data was obtained through accessing the following bioprojects publicly available *via* European Nucleotide Archive (ENA): PRJNA422434 and PRJEB25727.

Competing interests

The authors express to have no conflict of interest during the execution of this research.

Funding

The authors acknowledge the following funding bodies that provided support for this study: European Commission 7th Framework Program through the MyNewGut project (Grant agreement No. 613979) and the Spanish Ministry of Science and Innovation (MICINN) (grant PID2020-119536RB-I00) to YS; Institute of Health Carlos III (ISCIII) via CP19/00132 grant to AB-P with funds from the European Social Fund (ESF/FSE), CIPF Predoctoral Training Fellowships 2022 funded by JANSSEN-CILAG – supporting the contract of PH-C, CIPF core funding for the contract of SC-B, an Early Career Grant from the Society for Endocrinology award to MO, Spanish Government MCIN/AEI for the Center of Excellence Accreditation Severo Ochoa (CEX2021-001189-S/ MCIN/AEI / 10.13039/501100011033).

Author contributions

MO, YS and AB-P, intellectually conceived the study. PH-C, SC-B and AB-P, performed the molecular biology experiments and the *in vitro* experiments. AB-P performed the bioinformatics and metagenomic analysis. MO and RL-G performed the animal experiments and sample analysis. MO, YS and AB-P drafted the manuscript. All authors read the final version. YS and AB-P are responsible for its content.

Acknowledgements

The authors thank the Computational Biology & Data Science Unit from Principe Felipe Research Center (CIPF) for providing access to its high-performance computing cluster, co-funded by European Regional Development Funds (ERDF/FEDER). Finally, the authors thank the technical staff of the proteomics, microscopy and animal facilities of SCSIE University of Valencia and, in particular, I. Noguera, for their assistance in protocols and procedures described for animal experimentation.

References

1. de Vos WM, Tilg H, Van Hul M, Cani PD: **Gut microbiome and health: mechanistic insights.** *Gut* 2022, **71:**1020-1032.
2. Stebbins CE, Galan JE: **Structural mimicry in bacterial virulence.** *Nature* 2001, **412:**701-705.
3. Cohen LJ, Esterhazy D, Kim SH, Lemetre C, Aguilar RR, Gordon EA, Pickard AJ, Cross JR, Emiliano AB, Han SM, et al: **Commensal bacteria make GPCR ligands that mimic human signalling molecules.** *Nature* 2017, **549:**48-53.
4. Cohen LJ, Kang HS, Chu J, Huang YH, Gordon EA, Reddy BV, Ternei MA, Craig JW, Brady SF: **Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist.** *Proc Natl Acad Sci U S A* 2015, **112:**E4825-4834.
5. Breton J, Tennoune N, Lucas N, Francois M, Legrand R, Jacquemot J, Goichon A, Guerin C, Peltier J, Pestel-Caron M, et al: **Gut Commensal E. coli Proteins Activate Host Satiety Pathways following Nutrient-Induced Bacterial Growth.** *Cell Metab* 2016, **23:**324-334.
6. Tennoune N, Chan P, Breton J, Legrand R, Chabane YN, Akkermann K, Jarv A, Ouelaa W, Takagi K, Ghouzali I, et al: **Bacterial ClpB heat-shock protein, an antigen-mimetic of the anorexigenic peptide alpha-MSH, at the origin of eating disorders.** *Transl Psychiatry* 2014, **4:**e458.
7. Mulvihill EE, Drucker DJ: **Pharmacology, physiology, and mechanisms of action of dipeptidyl peptidase-4 inhibitors.** *Endocr Rev* 2014, **35:**992-1019.
8. Sneha P, Doss CG: **Gliptins in managing diabetes - Reviewing computational strategy.** *Life Sci* 2016, **166:**108-120.
9. Ahmed RH, Huri HZ, Muniandy S, Al-Hamodi Z, Al-Absi B, Alsalahi A, Razif MF: **Altered circulating concentrations of active glucagon-like peptide (GLP-1) and dipeptidyl peptidase 4 (DPP4) in obese subjects and their association with insulin resistance.** *Clin Biochem* 2017, **50:**746-749.
10. Lee SA, Kim YR, Yang EJ, Kwon EJ, Kim SH, Kang SH, Park DB, Oh BC, Kim J, Heo ST, et al: **CD26/DPP4 levels in peripheral blood and T cells in patients with type 2 diabetes mellitus.** *J Clin Endocrinol Metab* 2013, **98:**2553-2561.
11. Mulvihill EE, Varin EM, Gladanac B, Campbell JE, Ussher JR, Baggio LL, Yusta B, Ayala J, Burmeister MA, Matthews D, et al: **Cellular Sites and Mechanisms Linking Reduction of Dipeptidyl Peptidase-4 Activity to Control of Incretin Hormone Action and Glucose Homeostasis.** *Cell Metab* 2017, **25:**152-165.
12. Varin EM, Mulvihill EE, Beaudry JL, Pujadas G, Fuchs S, Tanti JF, Fazio S, Kaur K, Cao X, Baggio LL, et al: **Circulating Levels of Soluble Dipeptidyl Peptidase-4 Are Dissociated from Inflammation and Induced by Enzymatic DPP4 Inhibition.** *Cell Metab* 2019, **29:**320-334 e325.
13. Olivares M, Schuppel V, Hassan AM, Beaumont M, Neyrinck AM, Bindels LB, Benítez-Páez A, Sanz Y, Haller D, Holzer P, Delzenne NM: **The potential role of the dipeptidyl peptidase-4-like activity from the gut microbiota on the host health.** *Front Microbiol* 2018, **9:**1900.
14. Keller LJ, Nguyen TH, Liu LJ, Hurysz BM, Lakemeyer M, Guerra M, Gelsinger DJ, Chanin R, Ngo N, Lum KM, et al: **Chemoproteomic identification of a DPP4 homolog in Bacteroides thetaiotaomicron.** *Nat Chem Biol* 2023.
15. Mills RH, Dulai PS, Vazquez-Baeza Y, Sauceda C, Daniel N, Gerner RR, Batachari LE, Malfavon M, Zhu Q, Weldon K, et al: **Multi-omics analyses of the ulcerative colitis gut microbiome link Bacteroides vulgatus proteases with disease severity.** *Nat Microbiol* 2022, **7:**262-276.

16. Wang K, Zhang Z, Hang J, Liu J, Guo F, Ding Y, Li M, Nie Q, Lin J, Zhuo Y, et al: **Microbial-host-isozyme analyses reveal microbial DPP4 as a potential antidiabetic target.** *Science* 2023, **381:**eadd5787.
17. Anastasiou R, Papadelli M, Georgalaki MD, Kalantzopoulos G, Tsakalidou E: **Cloning and sequencing of the gene encoding X-prolyl-dipeptidyl aminopeptidase (PepX) from Streptococcus thermophilus strain ACA-DC 4.** *J Appl Microbiol* 2002, **93:**52-59.
18. Stressler T, Eisele T, Schlayer M, Lutz-Wahl S, Fischer L: **Characterization of the recombinant exopeptidases PepX and PepN from Lactobacillus helveticus ATCC 12046 important for food protein hydrolysis.** *PLoS One* 2013, **8:**e70055.
19. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al: **A metagenome-wide association study of gut microbiota in type 2 diabetes.** *Nature* 2012, **490:**55-60.
20. Benitez-Paez A, Kjolbaek L, Gomez Del Pulgar EM, Brahe LK, Astrup A, Matysik S, Schott HF, Krautbauer S, Liebisch G, Boberska J, et al: **A Multi-omics Approach to Unraveling the Microbiome-Mediated Effects of Arabinoxylan Oligosaccharides in Overweight Humans.** *mSystems* 2019, **4**.
21. Galperin MY, Koonin EV: **Divergence and convergence in enzyme evolution.** *J Biol Chem* 2012, **287:**21-28.
22. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH: **High frequency of horizontal gene transfer in the oceans.** *Science* 2010, **330:**50.
23. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD: **The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database.** *Nucleic Acids Res* 2018, **46:**D624-D632.
24. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464:**59-65.
25. Andreishcheva EN, Vann WF: **Escherichia coli BL21(DE3) chromosome contains a group II capsular gene cluster.** *Gene* 2006, **384:**113-119.
26. Chung KM, Cheng JH, Suen CS, Huang CH, Tsai CH, Huang LH, Chen YR, Wang AH, Jiaang WT, Hwang MJ, Chen X: **The dimeric transmembrane domain of prolyl dipeptidase DPP-IV contributes to its quaternary structure and enzymatic activities.** *Protein Sci* 2010, **19:**1627-1638.
27. Rawlings ND, Barrett AJ: **Evolutionary families of peptidases.** *Biochem J* 1993, **290 (Pt 1):**205-218.
28. Da Silva RR: **Commentary: The Potential Role of the Dipeptidyl Peptidase-4-Like Activity From the Gut Microbiota on the Host Health.** *Front Microbiol* 2018, **9:**3313.
29. Hafeez Z, Cakir-Kiefer C, Lecomte X, Miclo L, Dary-Mourot A: **The X-prolyl dipeptidyl-peptidase PepX of Streptococcus thermophilus initially described as intracellular is also responsible for peptidase extracellular activity.** *J Dairy Sci* 2019, **102:**113-123.
30. Tsakalidou E, Anastasiou R, Papadimitriou K, Manolopoulou E, Kalantzopoulos G: **Purification and characterisation of an intracellular X-prolyl-dipeptidyl aminopeptidase from Streptococcus thermophilus ACA-DC 4.** *J Biotechnol* 1997, **59:**203-211.
31. Durinx C, Lambeir AM, Bosmans E, Falmagne JB, Berghmans R, Haemers A, Scharpe S, De Meester I: **Molecular characterization of dipeptidyl peptidase activity in serum: soluble CD26/dipeptidyl peptidase IV is responsible for the release of X-Pro dipeptides.** *Eur J Biochem* 2000, **267:**5608-5613.
32. Yoshida S, Hatasa M, Ohsugi Y, Tsuchiya Y, Liu A, Niimi H, Morita K, Shimohira T, Sasaki N, Maekawa S, et al: **Porphyromonas gingivalis Administration Induces Gestational Obesity, Alters Gene Expression in the Liver and Brown Adipose Tissue**

in Pregnant Mice, and Causes Underweight in Fetuses. *Front Cell Infect Microbiol* 2021, **11:**745117.

33. Murphy KG, Bloom SR: **Gut hormones and the regulation of energy homeostasis.** *Nature* 2006, **444:**854-859.
34. Parker SL, Balasubramaniam A: **Neuropeptide Y Y2 receptor in health and disease.** *Br J Pharmacol* 2008, **153:**420-431.
35. Redrobe JP, Dumont Y, Fournier A, Quirion R: **The neuropeptide Y (NPY) Y1 receptor subtype mediates NPY-induced antidepressant-like activity in the mouse forced swimming test.** *Neuropsychopharmacology* 2002, **26:**615-624.
36. Hassan AM, Mancano G, Kashofer K, Frohlich EE, Matak A, Mayerhofer R, Reichmann F, Olivares M, Neyrinck AM, Delzenne NM, et al: **High-fat diet induces depression-like behaviour in mice associated with changes in microbiome, neuropeptide Y, and brain metabolome.** *Nutr Neurosci* 2019, **22:**877-893.
37. Magdy YM, El-Kharashi OA, Nabih ES, Shaker SM, Abd-Elaziz LF, Aboul-Fotouh S: **Potential involvement of JNK1 repression in the hepatic effect of sitagliptin and metformin in rats subjected to high fat diet and chronic mild distress.** *Biomed Pharmacother* 2017, **85:**225-238.
38. Olivares M, Neyrinck AM, Potgens SA, Beaumont M, Salazar N, Cani PD, Bindels LB, Delzenne NM: **The DPP-4 inhibitor vildagliptin impacts the gut microbiota and prevents disruption of intestinal homeostasis induced by a Western diet in mice.** *Diabetologia* 2018, **61:**1838-1848.
39. Schnapp G, Klein T, Hoevels Y, Bakker RA, Nar H: **Comparative Analysis of Binding Kinetics and Thermodynamics of Dipeptidyl Peptidase-4 Inhibitors and Their Relationship to Structure.** *J Med Chem* 2016, **59:**7466-7477.
40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
41. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5:**113.
42. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.
43. Eddy SR: **Multiple alignment using hidden Markov models.** *Proc Int Conf Intell Syst Mol Biol* 1995, **3:**114-120.
44. Eddy SR: **Hidden Markov models.** *Curr Opin Struct Biol* 1996, **6:**361-365.
45. Darriba D, Taboada GL, Doallo R, Posada D: **ProtTest 3: fast selection of best-fit models of protein evolution.** *Bioinformatics* 2011, **27:**1164-1165.
46. Letunic I, Doerks T, Bork P: **SMART: recent updates, new developments and status in 2015.** *Nucleic Acids Res* 2014, **43:**D257-260.
47. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8:**785-786.
48. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26:**2460-2461.
49. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, Le Cao KA: **DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays.** *Bioinformatics* 2019, **35:**3055-3062.
50. Guo S, Al-Sadi R, Said HM, Ma TY: **Lipopolysaccharide causes an increase in intestinal tight junction permeability in vitro and in vivo by inducing enterocyte membrane expression and localization of TLR-4 and CD14.** *Am J Pathol* 2013, **182:**375-387.
51. Vannan DT, Bomhof MR, Reimer RA: **Comparison of Glucose and Satiety Hormone Response to Oral Glucose vs. Two Mixed-Nutrient Meals in Rats.** *Front Nutr* 2018, **5:**89.

52. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner FO: **The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks.** *Nucleic Acids Res* 2014, **42:**D643-648.
53. Letunic I, Bork P: **Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.** *Nucleic Acids Res* 2016, **44:**W242-245.

Figure legends

Figure 1. A, Taxonomic distribution and phylogenetic analysis of PepX and DPP4 peptidases. The phylogenetic relationship between both families of bacterial peptidases is illustrated by a circular tree. A small unrooted tree is also shown at the top-right corner. The tree includes the bait used in the PSI-Blast search for distant homologs and the human, mouse and rat sequences of DPP4 enzymes for comparison. The outlier consisted of an S17 peptidase sequence from *Dichelobacter nodosus* (SwissProt ID P19577). The taxonomy assignment of sequences at the family and genus level was performed using the SILVA database taxonomy annotation (release 138) [52]. Branch colors are according to the peptidase family grouping: green for XaaPro peptidases (S15), orange for DPP4 peptidases (S9) and blue for the outlier (S17). Strip colors indicate the taxonomic distribution of sequences accordingly. Labels for the most abundant taxonomy groups in terms of families and genera (numbers within parenthesis) are shown. The tree was visualized using the iTOL web server [53]. **B**, Distribution of amino acid identities of eukaryote proteins *versus* bacterial DPP4-like sequences is shown as a boxplot. **C**, Metagenome abundance of DPP4-like proteins for classification of type 2 diabetes (T2D). A volcano plot shows the abundance variation of bacterial DPP4-like proteins in human gut metagenomes linked to T2D. Generalized linear mixed models (GLMM) estimates and p-values obtained after controlling for covariates (sex, age, body mass index) were used to identify proteins associated with T2D (N = 135) *versus* no-T2D controls (N = 85). A 2 × 2 contingency table-based *Chi*-squared test with Monte-Carlo simulation was computed to determine protein enrichment at the genera level. The threshold for selection was set at > 0.20 || < -0.20 (light grey lines). Red points indicate an association with T2D samples, and blue ones with no-T2D samples. Enrichment genera categories are shown as insets in the main plot. **D**, Area under the receiver operating characteristic (AUROC) curve resulting from the sparse partial least square discriminant analysis (sPLS-DA) to determine the performance of the DPP4-like proteins to classify T2D samples. Two models with two and three components (informative genetic traits) support the higher discriminant power of DPP4-like gene abundance from *Parabacteroides* and *Porphyromonas* species (on top) to discern T2D samples. AUROC and p-values are shown as insets in respective subpanels.

Figure 2. Secretory signal detection in the pmDPP4 protein. MALDI-TOF mass spectrometry monitoring of human GLP-1 incretin inactivation in enzymatic reactions using cell-free supernatants from *E. coli* BL21 cells carrying the pET15b (**A**), pET15b/pmDPP4 (**B**) and pET15b/pmDPP4Δ23 (**C**) constructs upon induction. Active form of GLP-1 (MW 3,355) was detected in all cell-mediated hydrolysis assays and the expected inactive form (MW 3,147) resulting from N-terminal dipeptide cleavage was only evident in cells carrying pET15b/pmDPP4 plasmid encoding full-length protein. Anti-His transmission electron microscopy immunogold staining of *E. coli* BL21 cells carrying the pET15b (**D**), pET15b/pmDPP4 (**E**) and pET15b/pmDPP4Δ23 (**F**) constructs upon induction. Red arrows indicate predominant distribution of 6×His-tagged proteins for E and F panels. Scale bars at 100 nm (bars below TEM images). MW: monoisotopic mass.

Figure 3. *In vitro* function of the pmDPP4 protein. MALDI-TOF mass spectrometry spectra of hydrolytic reactions performed in the presence of pmDPP4Δ23 or hsDPP4 proteins. Panels are distributed in terms of substrate used and the enzyme added (**A**-**D**). The substrates used for reactions are depicted on top of every panel series with the respective sequence information and the theoretical molecular weight (MW, monoisotopic mass). Delta molecular mass values were calculated from the monoisotopic masses of products observed in the control hsDPP4 reactions. The x-axis indicates the molecular mass values in terms of m/z, and the y-axis indicates relative intensity (R.I.) of the most abundant peptide species.

Figure 4. *In vivo* function of the pmDPP4 protein. Scheme of the animal study to test *in vivo* activity of bacterial DPP4-like enzyme. **A**, Study design, mice administered with *E. coli* BL21-DE3 cells carrying the pET15b empty vector or the pET15b/pmDPP4 vector (n=8 each). **B, C** DPP4 activity measured in cecal content (B) and portal vein (C) of mice. In the portal vein, active GLP-1, active GIP, total GLP-1, proglucagon, total GIP and PYY are shown in panels **D**, **E**, **F**, **G**, **H** and **I**, respectively. Insulinemia and glycemia is shown in panels **J** and **K**. Data were analyzed using a *t*-test and significance was set at $p<0.05$.

Figure 5. Dose-response of gliptins on DPP4 enzymes. The EC_{50} parameter was calculated for hsDPP4 and pmDPP4Δ23 to determine the inhibitory efficacy of the drug on the hydrolytic activity of the enzymes. Serial dilutions of FDA- or EMA-approved gliptins from 1×10^{-6} to 1×10^{-9} M were established to measure the production of para-nitroaniline. Top panel shows dose-response curves for hsDPP4, and bottom panel shows dose-response curves for bacterial pmDPP4Δ23. Data obtained were computed *via* non-linear fitting regression to estimate the EC_{50}. Data were derived from four independent replicates using different protein purification batches.

Supplementary material S1. Comparison of DPP4-like function-associated peptidases. Domain architecture, signal peptide presence and transmembrane domain predictions were examined on the human DPP4 enzyme (SwissProt id P27487) (**A**), the consensus amino acid profile of bacterial DPP4-like proteins (**B**), and the consensus amino acid profile of PepX peptidases (**C**) using the respective web servers (top labels). Individual functional domains according to SMART and/or Pfam annotation present in full-length query proteins are presented in classical architecture diagrams. SignalP prediction for the N-terminal 70 amino acids shows C- (raw cleavage site score, red lines), S- (signal peptide score, green lines), and Y- (combined cleavage site score, blue lines) scores to determine the individual probability of amino acids to be part of a signal peptide driving protein secretion or potential membership of a transmembrane structure. TMHMM prediction for the N-terminal 70 amino acids shows the probability of individual amino acids to be part of alpha helices embedded in biological membranes, thus indicating localization in the transmembrane (red lines), inside of cell (blue line) or outside of cell (pink lines). The amino acid logo representation of the multiple sequence alignment (MSA) of DPP4-like proteins (bacteria, mouse, rat, and human) for the catalytic motif of S9 peptidase family is drawn in **D**. Amino acid logo was created using the positions between amino acids 935 and 941 of the MSA using the SeqLogo sever (http://www.cbs.dtu.dk/biotools/Seq2Logo/) and based on the P-Weighted Kullback-Leibler score. The amino acid positions for the human protein are also shown at the bottom. This regions shows the MEROPS [23] catalytic motif (GWSYGGY) of these serine proteases (S9).

Supplementary material S2. A, SDS-PAGE analysis of *E. coli* BL21-DE3 cell extracts expressing recombinant pmDPP4 and pmDPP4Δ23 proteins. **B**, Ninety minute kinetics of pmDPP4 activity (left y-axis, in terms of mU/mL units) and LDH activity (right y-axis, in terms of 490 nm - 680 nm absorbance) mediated by cell-free supernatants from *E. coli* BL21 cells carrying pET15b/pmDPP4 or pET15b/pmDPP4Δ23 constructs upon induction. **C**, Gel filtration assay of an ionic-exchange purified fraction of the pmDPP4Δ23 protein, revealing a dimeric nature in solution according to the calculated MW from the standards elution profile. **D**, Purified pmDPP4Δ23 and human DPP4 proteins used in the *in vitro* assays for human hormone hydrolysis and gliptin inhibition. Gels were stained with Coomassie blue to visualize protein bands. The Dual Xtra pre-stained protein standard (Bio-Rad, Hercules, CA) was used as a molecular size reference. LDH, lactate dehydrogenase.

Supplementary material S3. Gliptin inhibition kinetics on pmDPP4Δ23. Lineweaver-Burk double-reciprocal (1/V and 1/S) plots showing pmDPP4Δ23 (bottom panels) and hsDPP4 (top panels) gpPNA hydrolysis in presence of vildagliptin, sitagliptin, saxagliptin, and linagliptin. Inhibitor concentration is shown as line legend color from lower (0 nM, light red) to higher (100 nM, burgundy lines) concentrations. The apparent K_m was calculated respectively from inhibitor-associated data points using linear regression to find intersection point at the y-axis (1/S). V, reaction velocity (nmol PNA min^{-1} pmole $enzyme^{-1}$); S, substrate concentration (mM gpPNA).

Supplementary material S4. K_m fold-change on pmDPP4Δ23 and hsDPP4 enzymes upon gliptin inhibition. Data derived from kinetics assays shown in Supplementary material S5 are shown in this spreadsheet. Headers indicate enzyme source, inhibitor tested, K_m estimated, standard deviation (SD), fold-change, and p-value obtained from comparisons with respective reactions with no inhibitor.

Figure 1

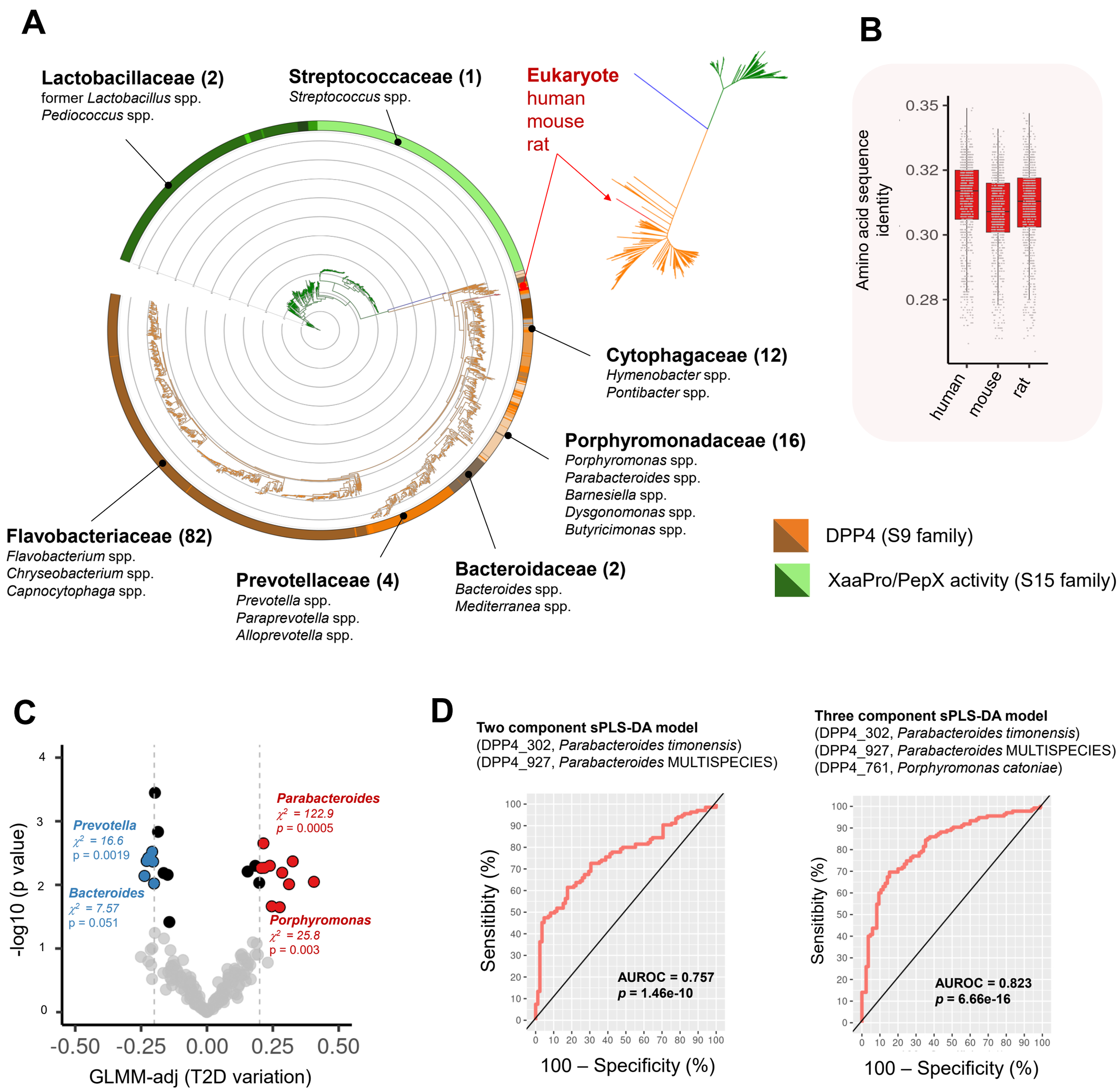

A
Lactobacillaceae (2)
former Lactobacillus spp.
Pediococcus spp.
Streptococcaceae (1)
Streptococcus spp.
Eukaryote
human
mouse
rat
Cytophagaceae (12)
Hymenobacter spp.
Pontibacter spp.
Porphyromonadaceae (16)
Porphyromonas spp.
Parabacteroides spp.
Barnesiella spp.
Dysgonomonas spp.
Butyricimonas spp.
Flavobacteriaceae (82)
Flavobacterium spp.
Chryseobacterium spp.
Capnocytophaga spp.
Prevotellaceae (4)
Prevotella spp.
Paraprevotella spp.
Alloprevotella spp.
Bacteroidaceae (2)
Bacteroides spp.
Mediterranea spp.
B
Amino acid sequence identity
0.35
0.32
0.30
0.28
human
mouse
rat
DPP4 (S9 family)
XaaPro/PepX activity (S15 family)
C
-log10 (p value)
GLMM-adj (T2D variation)
Prevotella
χ2 = 16.6
p = 0.0019
Bacteroides
χ2 = 7.57
p = 0.051
Parabacteroides
χ2 = 122.9
p = 0.0005
Porphyromonas
χ2 = 25.8
p = 0.003
D
Two component sPLS-DA model
(DPP4_302, Parabacteroides timonensis)
(DPP4_927, Parabacteroides MULTISPECIES)
Three component sPLS-DA model
(DPP4_302, Parabacteroides timonensis)
(DPP4_927, Parabacteroides MULTISPECIES)
(DPP4_761, Porphyromonas catoniae)
Sensitibity (%)
100 – Specificity (%)
AUROC = 0.757
p = 1.46e-10
AUROC = 0.823
p = 6.66e-16

Figure 2

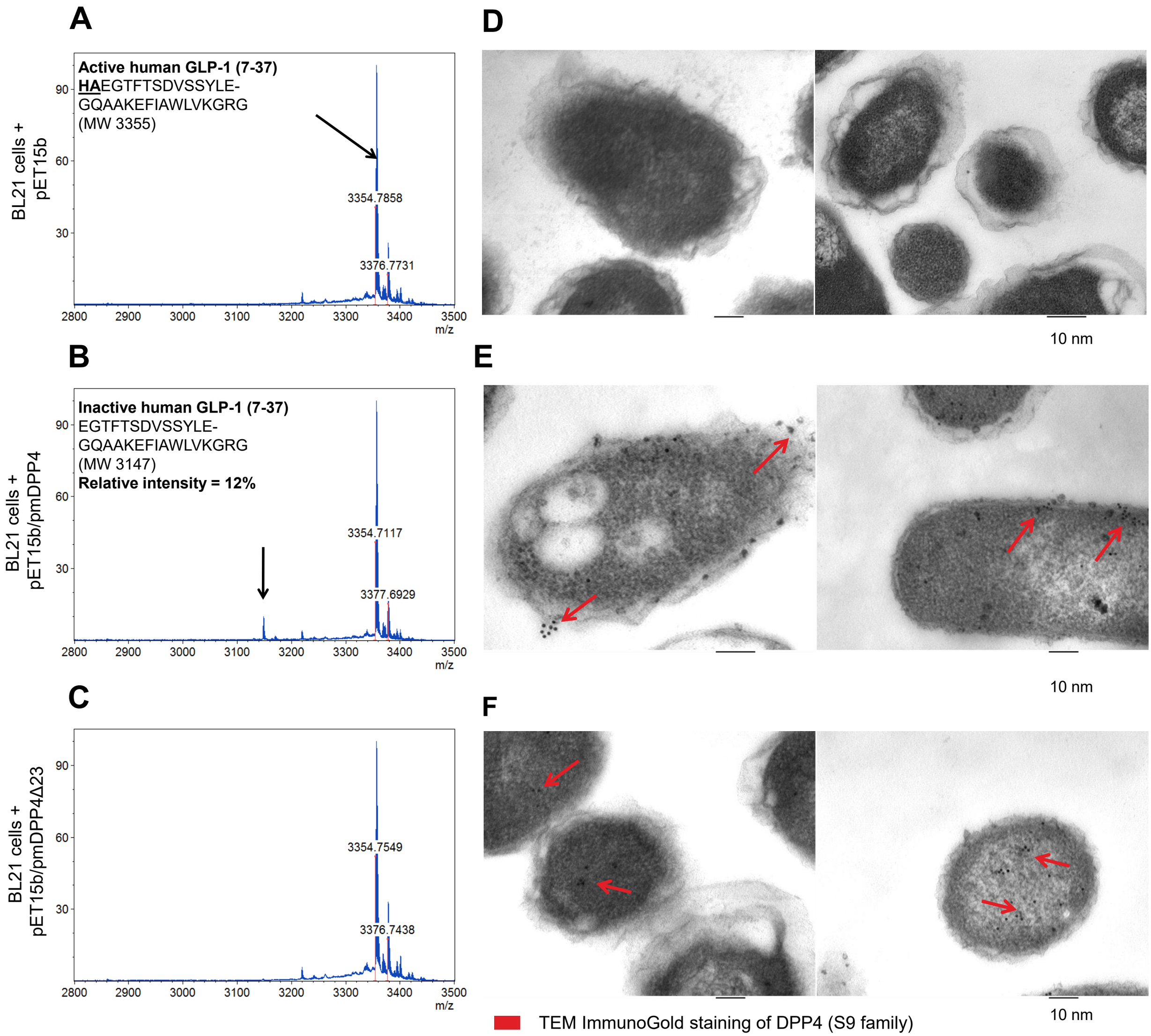
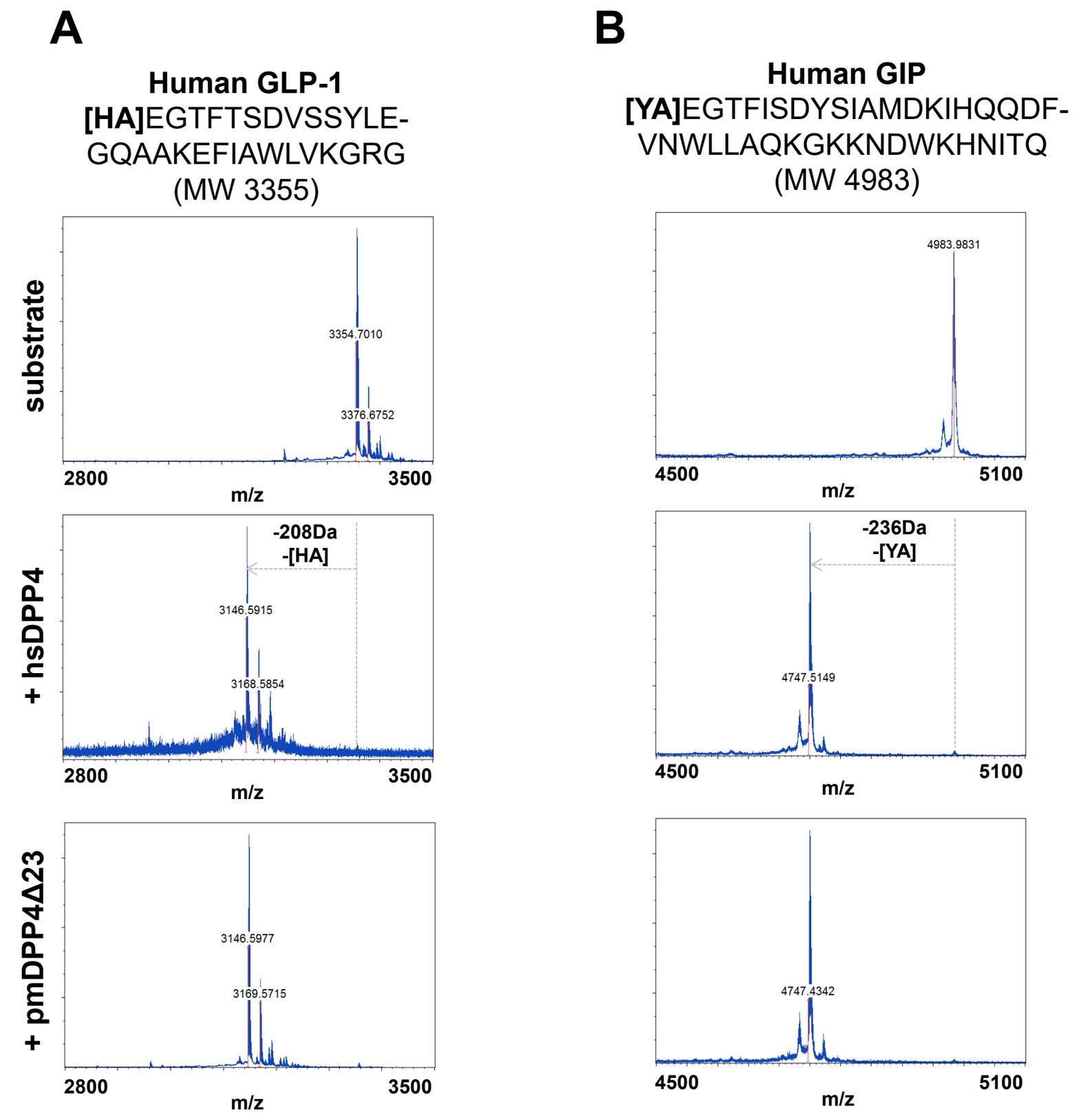

A
Active human GLP-1 (7-37)
HAEGTFTSDVSSYLE-
GQAAKEFIAWLVKGRG
(MW 3355)
BL21 cells +
pET15b
3354.7858
3376.7731
m/z
B
Inactive human GLP-1 (7-37)
EGTFTSDVSSYLE-
GQAAKEFIAWLVKGRG
(MW 3147)
Relative intensity = 12%
BL21 cells +
pET15b/pmDPP4
3354.7117
3377.6929
m/z
C
BL21 cells +
pET15b/pmDPP4Δ23
3354.7549
3376.7438
m/z
D
10 nm
E
10 nm
F
10 nm
TEM ImmunoGold staining of DPP4 (S9 family)

Figure 3


A
Human GLP-1
[HA]EGTFTSDVSSYLE-
GQAAKEFIAWLVKGRG
(MW 3355)
substrate
3354.7010
3376.6752
2800
3500
m/z
+ hsDPP4
-208Da
-[HA]
3146.5915
3168.5854
2800
3500
m/z
+ pmDPP4Δ23
3146.5977
3169.5715
2800
3500
m/z
B
Human GIP
[YA]EGTFISDYSIAMDKIHQQDF-
VNWLLAQKGKKNDWKHNITQ
(MW 4983)
4983.9831
4500
5100
m/z
-236Da
-[YA]
4747.5149
4500
5100
m/z
4747.4342
4500
5100
m/z

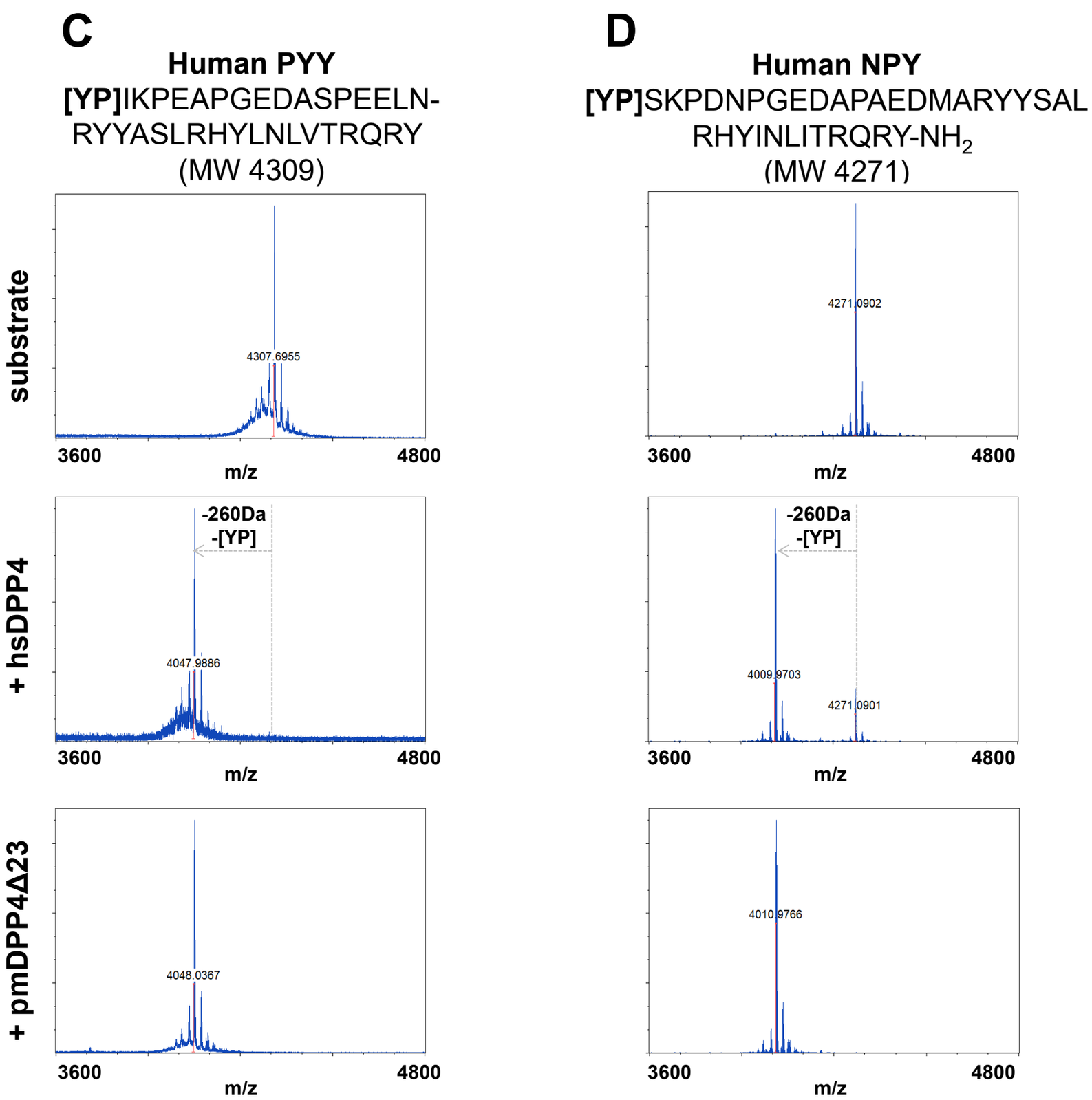

C
Human PYY
[YP]IKPEAPGEDASPEELN-
RYYASLRHYLNLVTRQRY
(MW 4309)
substrate
4307.6955
3600
4800
m/z
+ hsDPP4
-260Da
-[YP]
4047.9886
3600
4800
m/z
+ pmDPP4Δ23
4048.0367
3600
4800
m/z
D
Human NPY
[YP]SKPDNPGEDAPAEDMARYYSAL
RHYINLITRQRY-NH2
(MW 4271)
4271.0902
3600
4800
m/z
-260Da
-[YP]
4009.9703
4271.0901
3600
4800
m/z
4010.9766
3600
4800
m/z

Figure 4

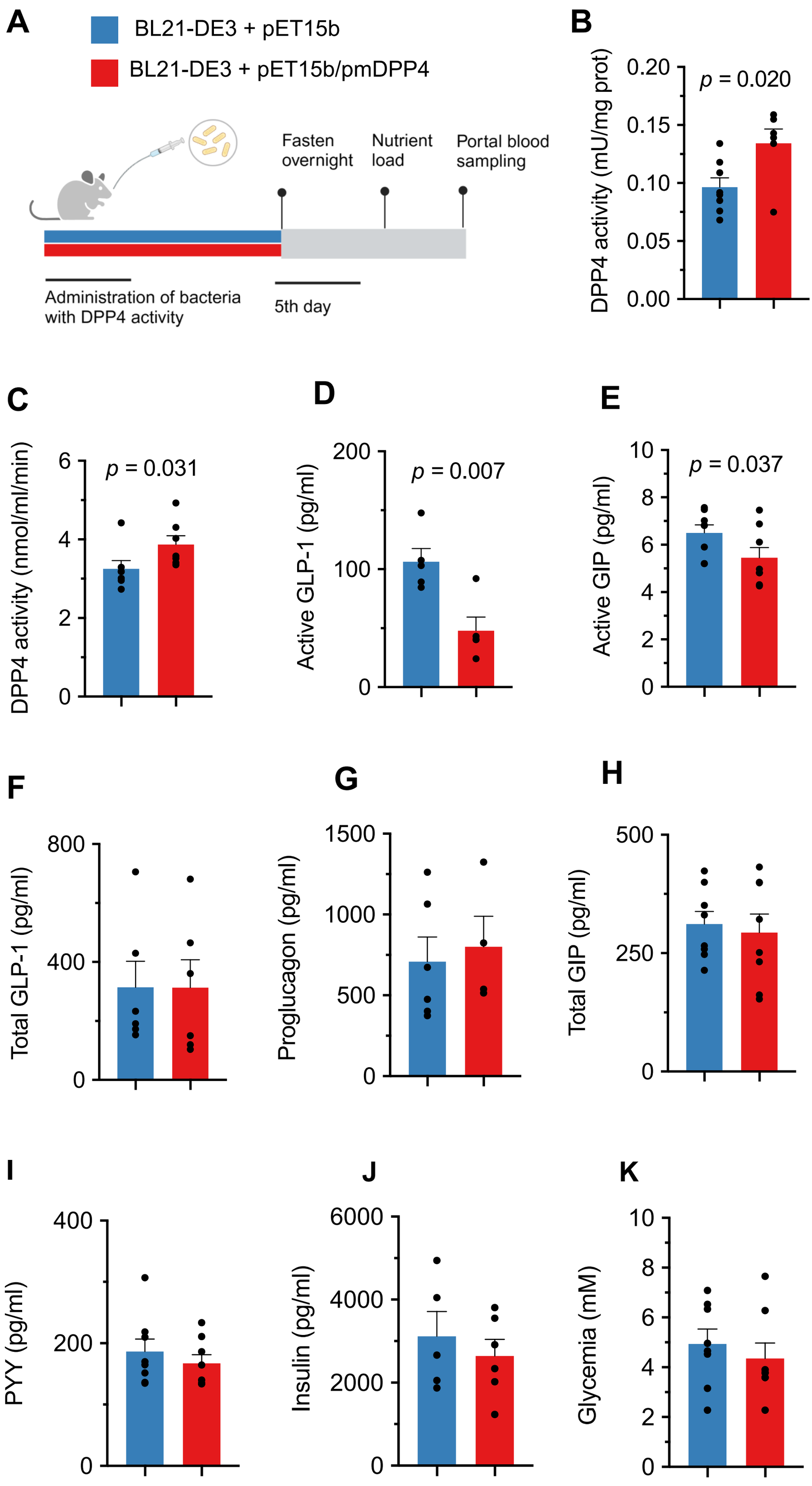
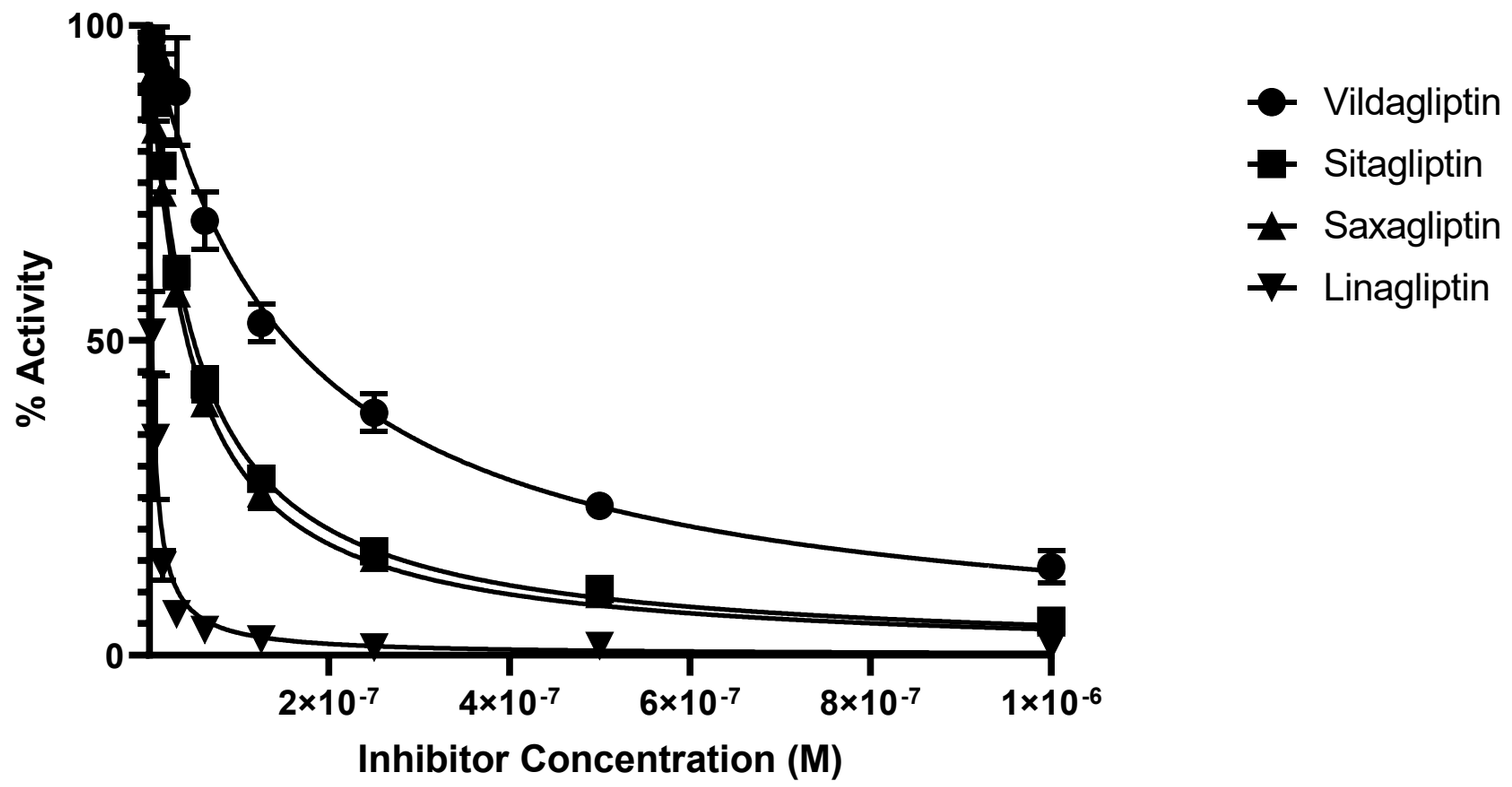

A
BL21-DE3 + pET15b
BL21-DE3 + pET15b/pmDPP4
Fasten overnight
Nutrient load
Portal blood sampling
Administration of bacteria with DPP4 activity
5th day
B
DPP4 activity (mU/mg prot)
p = 0.020
C
DPP4 activity (nmol/ml/min)
p = 0.031
D
Active GLP-1 (pg/ml)
p = 0.007
E
Active GIP (pg/ml)
p = 0.037
F
Total GLP-1 (pg/ml)
G
Proglucagon (pg/ml)
H
Total GIP (pg/ml)
I
PYY (pg/ml)
J
Insulin (pg/ml)
K
Glycemia (mM)

Figure 5

A

IC_{50}	Vildagliptin	Sitagliptin	Saxagliptin	Linagliptin
	1.545e-007	5.007e-008	4.288e-008	3.684e-009



B

IC_{50}	Vildagliptin	Sitagliptin	Saxagliptin	Linagliptin
	4.571e-007	9.043e-005	1.452e-007	1.484e-005

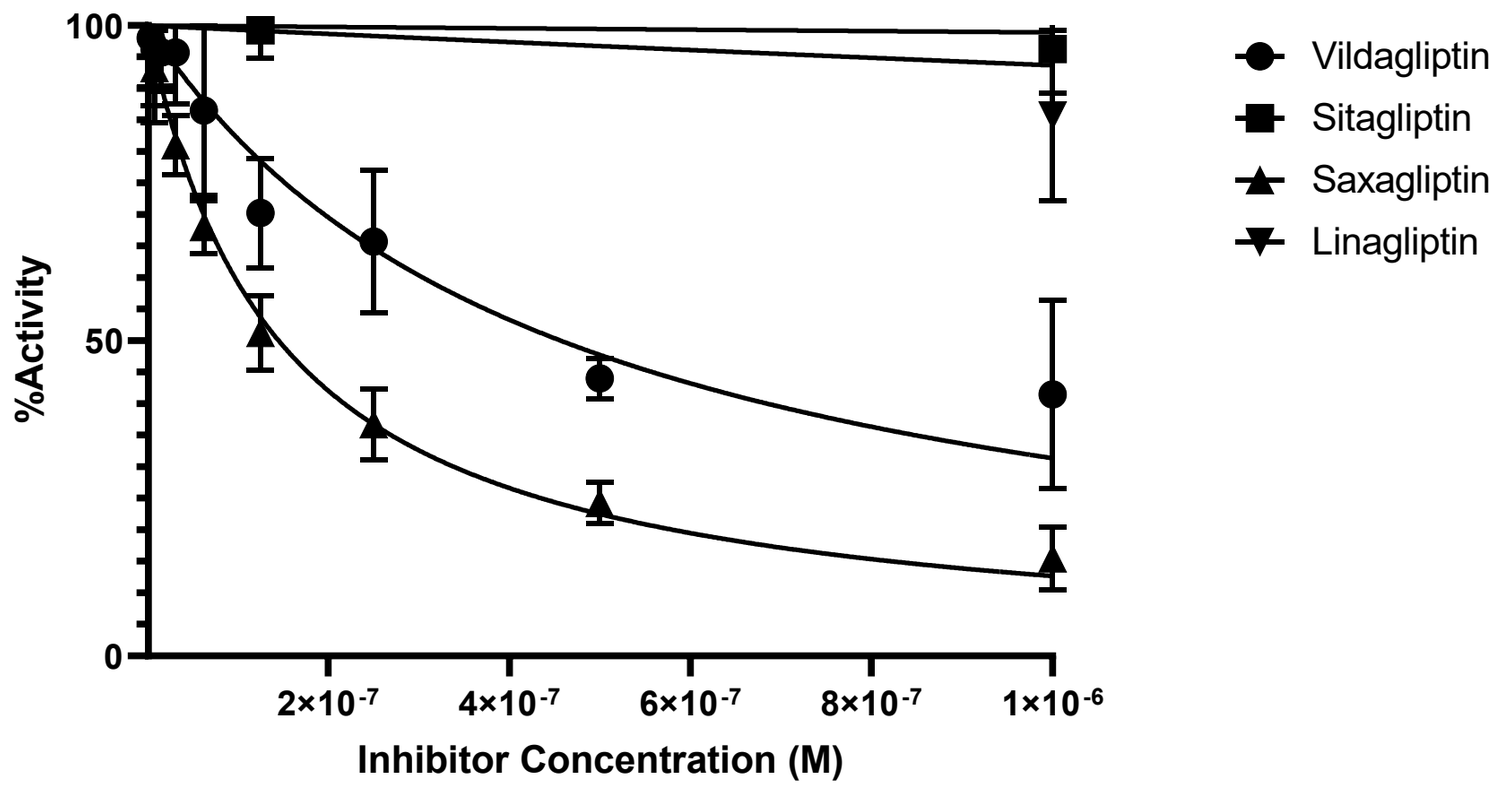


Table 1. Estimation of EC_{50} in a gliptins dose-response assay.

	hsDPP4			pmDPP4		
Gliptin	EC_{50}	CI	R^2	EC_{50}	IC	R^2
Sitagliptin	5.01×10^{-8}	4.76×10^{-8} to 5.27×10^{-8}	0.995	9.04×10^{-5}	1.35×10^{-5} to 1	ND
Vildaliptin	1.54×10^{-7}	1.41×10^{-7} to 1.70×10^{-7}	0.980	4.57×10^{-7}	3.68×10^{-7} to 5.72×10^{-7}	0.833
Saxagliptin	4.29×10^{-8}	4.14×10^{-8} to 4.44×10^{-8}	0.998	1.45×10^{-7}	1.30×10^{-7} to 1.69×10^{-7}	0.971
Linagliptin	3.68×10^{-9}	3.30×10^{-9} to 4.10×10^{-9}	0.982	1.48×10^{-5}	4.87×10^{-6} to 1	ND

EC_{50}, Estimated at molar (M) concentration. CI, confidence interval at 95%. R^2, calculated as the fit of data into the non-linear function of the inhibitor concentration vs normalized response. Data obtained from independent experiments (N = 2). ND, data did not fit as non-linear function.