Gut-Microbiota-Related Metabolite Phenylacetylglutamine and Risk of Incident Coronary Heart Disease among Women

Yoriko Heianza, RD, PhD¹,²*; Saumya Tiwari, PhD³; Xuan Wang, PhD¹; Jeramie D Watrous, PhD³; Kathryn M. Rexrode, MD, MPH⁴,⁵; Frank B. Hu, MD, PhD²,⁶,⁷; Mona Alotaibi, MD³; Mohit Jain, PhD³; Qi Sun, MD, ScD²,⁷; JoAnn E. Manson, MD, DrPH⁵,⁶,⁷; Lu Qi, MD, PhD¹,²,⁶*

1. Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA
2. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA
3. Department of Pharmacology, University of California San Diego, La Jolla, CA
4. Division of Women’s Health, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA
5. Division of Preventive Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA
6. Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA
7. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

Running title: Gut Metabolite Phenylacetylglutamine and CHD Risk

Keywords: Metabolites, Risk factors, Plant-based diet, Coronary heart disease, Prospective cohort study.

Funding: The study is supported by NIH grants from the National Cancer Institute (UM1 CA186107, CA49449, U01 CA176726, and R01 CA67262), the National Heart, Lung, and Blood Institute (R01 HL034594, R01 HL088521, HL071981, HL126024), the National Institute of Diabetes and Digestive and Kidney Diseases (DK091718, DK100383, DK115679), Tulane Research Centers of Excellence Awards,
and National Institute of General Medical Sciences (2P20GM109036-06A1, Sub-Project ID 7233). Yoriko Heianza was a recipient of a Grant-in-Aid for Scientific Research and a Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science, and the American Heart Association postdoctoral fellowship award (19POST34380035). The sponsors had no role in the design or conduct of the study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Disclosures:** YH, XW, KMR, FBH, QS, JEM, and LQ have nothing to declare. MA received the NIH grant, K08HL166950. ST is a co-founder and employee at Sapient Bioanalytics. JDW is an employee at Sapient Bioanalytics and holds equity in the company. MJ holds an executive position and equity in Sapient Bioanalytics, LLC (San Diego, CA). The authors declare no conflict of interest associated with this manuscript.

**Word count for abstract:** 249 words

**Total word count:** 4498 words; 3 tables and 3 figures.

**Corresponding Authors:**

Yoriko Heianza, RD, PhD; https://orcid.org/0000-0002-0957-9309

Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, Suite 1724, New Orleans, LA, 70112.

Telephone: 504-988-7259; Fax: 504-988-3548; E-mail: yheianza@tulane.edu; nhyoh@channing.harvard.edu

Lu Qi, MD, PhD

Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA;
Abstract

**Context:** Phenylacetylglutamine (PAGln) is a novel metabolite derived from gut microbial metabolism of dietary proteins, specifically phenylalanine, which may be linked to risks of adverse cardiovascular events.

**Objective:** We investigated whether higher plasma levels of PAGln were associated with a greater risk of incident coronary heart disease (CHD) and tested whether adherence to a plant-based diet, which characterizes habitual dietary patterns of animal and plant food intake, modified the associations.

**Methods:** We examined associations between plasma PAGln and risk of incident CHD over 11-16 years in a nested case-control study of 1520 women (760 incident cases and 760 controls) from the Nurses’ Health Study. Separately, we analyzed relations between PAGln and dietary intakes measured through dietary records in the Women’s Lifestyle Validation Study (n=725).

**Results:** Higher PAGln levels were related to a greater risk of CHD ($p <0.05$ for dose-response relationship). Higher PAGln was associated with greater red/processed meat intake and lower vegetable intake ($p <0.05$ for all). We found a significant interaction between PAGln and adherence to plant-based diet index (PDI) on CHD ($P_{interaction}=0.008$); higher PAGln levels were associated with an increased risk of CHD (relative risk per 1 SD: 1.22 [95% CI: 1.05, 1.41]) among women with low PDI but not among those with high PDI.

**Conclusion:** Higher PAGln was associated with higher risk of CHD, particularly in women with dietary patterns of eating more animal foods and fewer plant-based foods. Adherence to plant-based diets might attenuate unfavorable associations between a novel microbial metabolite and CHD risk.
Introduction

Gut microbial metabolites derived from nutrient precursors have been novel risk factors for atherosclerosis, coronary heart disease (CHD), and major adverse cardiovascular events and deaths (1-6). Diet is one of major factors that modulates microbial metabolite production and influences host metabolism (3). Phenylacetylglutamine (PAGln), a metabolite derived via gut microbial metabolism of dietary proteins, has been recently linked to the risk of cardiovascular disease (CVD) (4, 7). Unabsorbed amino acid phenylalanine is metabolized by microbiota in the large intestines to produce phenylpyruvic acid and subsequently phenylacetic acid (4, 8-11); PAGln is formed in the liver from glutamine conjugation of phenylacetic acid (4, 12). Key gut microbial pathways for phenylacetic acid formation have been identified recently (10). Emerging evidence suggests that PAGln may be related to regulating cardiovascular risk via interaction with adrenergic receptors (4), a class of receptors that are crucially involved in cardiovascular function (13-15).

In recent case-control studies, high PAGln levels were related to in-stent hyperplasia and stenosis in coronary artery disease (CAD) patients (16) and coronary atherosclerotic plaque burden among patients with suspected CAD (17). Only a few studies have analyzed prospective relationships between circulating PAGln and the risks of heart disease or CVD (4, 7, 18), and it remains unclear whether high plasma PAGln levels are associated with the incidence of CHD among people at usual risk. In addition, diet is one of the critical components in modifying the relationship between atherogenic microbial metabolites and the risk of CHD (5, 19-22). A few studies of relatively small sample size suggest that certain dietary patterns and food items, such as a high-protein diet including animal-sourced foods vs. healthy plant foods including vegetables, differently affected PAGln levels (23-27). Nonetheless, it remains unknown whether associations of circulating PAGln levels with risk of incident CHD can be modified by the adherence to dietary patterns of habitual animal-based vs. plant food intakes.
In the present study, we prospectively investigated associations of plasma PAGln levels with risk of incident CHD in women from the Nurses’ Health Study (NHS) by considering adherence to dietary patterns, using the validated plant-based dietary index (PDI) which characterizes habitual animal and plant food intakes. We also tested whether circulating PAGln levels were associated with animal and plant food intakes using gold standard 7-day dietary records (7DDRs) and intermediate cardiometabolic risk factors of CHD in women from the Women's Lifestyle Validation Study (WLVS).

Materials and Methods

Study design, setting, and participants

The present study included two independent components: a prospective nested case-control study of incident CHD in the NHS and a cross-sectional study of women participating in the WLVS. The study protocol was approved by the institutional review boards (IRBs) of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and the IRBs allowed participants’ completion of questionnaires to be considered as implied consent.

The NHS is a prospective cohort study of US female registered nurses (n=121,700) aged 30–55 y when enrolled in 1976. Information on demographics, lifestyle factors, medical history, and disease status was collected through a self-administered questionnaire in 1976. The data has been updated every two years through follow-up questionnaires. The follow-up rate was high, with approximately 90% in each 2-year follow-up cycle. A blood sample was collected from 32,826 women during 1989–90 (NHS1.1), and 18,743 women provided a second blood sample in 2000–02 (NHS1.2) (28). The second blood sample was collected according to a protocol identical to the one used at the first collection. We performed a prospective nested-case control study of the incident CHD after the blood collection NHS1.1 or NHS1.2 until the end of follow-up (2000 for the NHS1.1 and 2016 for the NHS1.2). The incidence of CHD was identified and confirmed over 11-16 years of follow-up time. We used risk-set sampling to randomly select...
one control for each case from those who remained free of CHD events at the time CHD was diagnosed in the case subject. Incident cases and controls were matched on age at blood draw, smoking status, fasting status at blood draw, and date of blood draw. Participants who were free of the primary outcome (non-fatal myocardial infarction [MI] or fatal CHD) at the time of blood collection were eligible; a total of 1524 participants (n=762 incident CHD case-control pairs) were selected for the measurements of plasma metabolites. Of 1524, a few outliers (n=2) of the exposure metabolite and missing their case-control pairs were excluded from the present analysis. Subsequently, 760 incident CHD cases and 760 controls (total n=1520) were eligible, including 397 from the NHS1.1 and 363 from the NHS1.2, respectively. Nearly all of the study participants were free of a self-reported history of chronic kidney failure (99%) at the time of blood collection. Participants with missing data on PDI (n=27) were excluded when performing a stratified analysis.

The WLVS is a substudy in the NHS and NHS-II (NHS-II began in 1989, enrolling 116,429 female registered nurses aged 25–42 y) that aimed to examine the validity of self-reported diet and lifestyle; the details of the study design have been described previously (29, 30). Women with a history of CHD, stroke, cancer, or major neurological disease were excluded at the enrollment (30). Briefly, the WLVS collected anthropometric data, diet and physical activity, and biomarker measurements over a period of approximately 15 months during 2010-12 (with 5 phases, with each phase representing a three-month interval). The participants provided two blood samples taken approximately 6 months apart; by design, 7DDRs and other data were collected within the same phase of the study (29). The present study was designed to use blood samples at the second collection for measuring plasma metabolites in the WLVS; the last date of the 7DDR report preceded the date of the second blood collection (i.e., before the metabolite measurement). In the WLVS, we measured plasma metabolites in 751 samples based on the availability of blood samples among the enrolled participants (n=796). The following exclusions were made: an outlier of the exposure metabolite PAGln (n=1); missing value of PAGln (n=1); participants...
with missing data on body weight or total energy intake (n=11); implausible total energy intake (>3500 kcal/day) (n=2). Subsequently, 725 women were eligible for the analysis.

**Measurement of plasma metabolites**

Blood samples were centrifuged and aliquoted into cryotubes as plasma, buffy coat, and erythrocyte fractions, which were then stored in liquid nitrogen freezers at –130 °C or colder until analysis. Samples of the case-control pairs were shipped in the same batch and analyzed in the same run. Both technicians and laboratory personnel were blinded to the case-control status of the samples. Plasma samples from the NHS and the WLVS were analyzed at the University of California San Diego following the previously reported methods (31). Briefly, 20 μL of samples were transferred to a 96-well plate containing 80 μL of methanol with 0.5% acetic acid. The methanol extraction solution contained the following isotopically labeled internal standards: $^{13}$C$_5^{15}$N$_2$-Glutamine (Sigma-Aldrich), MAPCHO-12-d38 (Avanti Lipids),$^{13}$C$_5^{15}$N$_1$-Glutamate (Sigma-Aldrich), and CUDA (Cayman Chemicals). Samples were shaken at 550 rpm at 4°C for 10 min followed by centrifugation at 6000g at 4°C for 10 min. Thereafter, supernatant was then transferred to a 384-well polypropylene plate containing 35:65 or 75:25 methanol:water (for pos/neg mode analysis). Samples were analyzed by rapid liquid chromatography-mass spectrometry (rLC-MS) which utilizes rapid nano-valve switching and a three-pump system delivering a multi-tiered isocratic elution coupled to a high-resolution Agilent 6545B QToF mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Samples were injected onto a custom-packed, silica-based, mixed mode column that allows for both reverse phase and weak ion pairing retention mechanisms. Samples were eluted using mobile phases starting with either water containing 0.1% acetic acid (positive mode) or 75:25 methanol:water containing 0.1% acetic acid (negative mode) and ending with 90:10 acetonitrile:water containing 0.1% acetic acid and 5mM ammonium acetate. The mass spectrometer parameters were set as follows: dry gas temp of 365 °C, dry gas flow rate of 13 L/min, nebulizer gas of 60 psi, sheath gas temp
of 400 °C, sheath gas flow rate of 12 L/min, source voltage of (+)3500 / (-) 3000 V and nozzle voltage of (+) 50 / (-) 1000 V, mass range set to 50 – 1700 m/z, and data were collected at 8 spectra/s. Data QC was performed using the panel of isotopically labeled internal standards and interval pooled plasma samples to monitor fluctuations in extraction efficiency, instrument sensitivity, matrix artefact and mass accuracy. For system suitability, mass calibration (passing threshold of 5 ppm) was performed prior to each 384-well plate run and used to assess mass accuracy, mass resolution, detector sensitivity, and instrument cleanliness. For each 384-well plate run, the QC was as follows: isotopically labeled internal standards were added to each sample at the first preparation step to monitor matrix effects; bulk pre- aliquoted commercial pooled plasma (BioIVT) was placed in wells A1, D12 and H12 of each 96-well plate and prepared identically to samples (internal bracket QC sample); bulk pre- aliquoted commercial pooled plasma was prepared external to the 96-well plate by hand (external bracket QC sample), and a preparation blank was prepared during each 384-well plate to assess background. The overall coefficient of variation (%CV) was monitored for internal standards in the samples; for samples with %CV >25% and total ion current measures >3 standard deviations from the mean, samples were removed from the final analysis. Data were initially collected in MS1 mode where chromatographic features were extracted and aligned.

In the present analysis, the primary metabolite exposure was plasma PAGln, and other metabolites (such as a gut-microbial metabolite trimethylamine N-oxide [TMAO], phenylalanine, and glutamine) were examined in sensitivity analyses. Each metabolite examined in the present analysis (STable 1 in reference (32)) underwent a targeted MS2 investigation, whereby MS2 spectra were collected and searched against the NIST20 MS2 spectral library, which includes MS2 spectra for 31,000 compounds. In addition, each metabolite was further confirmed with commercially purchased standards. Additional details on the analysis and identifications are shown in SFigures 1-8 (32).
Ascertainment of CHD

A self-reported history of nonfatal MI was primarily confirmed by reviewing medical records by physicians who were blinded to the exposure status and the specific research question under study. Nonfatal MI was confirmed if it met the criteria of the World Health Organization, which required typical symptoms plus diagnostic electrocardiographic findings or elevated enzyme levels. Deaths were reported by the next of kin or the postal system or identified by searching the National Death Index (NDI). We previously estimated that at least 98% of deaths were identified using these approaches (33). Fatal CHD was identified by medical records or autopsy reports or if CHD was listed as the cause of death on the death certificate along with prior evidence of CHD. The study designated probable fatal CHD deaths when no medical records surrounding the death were available, but CHD was the underlying cause on the death certificate or NDI search, or a family member provided supporting information regarding the diagnosis.

Assessment of dietary intake using 7DDRs in the WLVS

In the WLVS, the participants received an Escali food scale and ruler, an instructional DVD, and instructions via telephone for keeping the 7DDRs (29, 30). Participants measured and reported gram weights for foods before and after eating to compute actual intake, and they also provided recipes of all home-prepared foods, including the number of servings and portions consumed, as well as labels of store-brand products. The 7DDR data were processed at the Nutrition Coordinating Center (NCC), University of Minnesota; nutrients and total energy intakes were calculated based on the NCC Food and Nutrient Database (34). We used the 7-day averaged values of food and nutrient intake data at the time of blood collection. We also analyzed food/nutrient data based on two 7DDRs preceding the metabolite measurement taken 6 months apart to assess longer-term intakes. The average daily intakes of animal foods, in particular meat-related food items, red meat [beef/pork/lamb], processed meat, poultry [chicken/turkey], and fish/seafood, were assessed because these food items are considered major sources...
of dietary protein in usual non-vegetarian diets and the source of substrate phenylalanine. Intakes of whole
vegetables, fruits, and tofu/soybean were assessed as a comparison group of healthy non-animal foods.

Measurements of cardiometabolic biomarkers, covariates, and other variables
In the WLVS, plasma levels of insulin, proinsulin, C-peptide, and lipids (total cholesterol, HDL
cholesterol, and triglycerides) were measured as previously described (35, 36). LDL cholesterol was
calculated using the Friedewald equation except when triglyceride levels were over 400 mg/dL.
Proinsulin, insulin, and C-peptide were assessed in participants without diabetes. Demographic data (such
as age and race/ethnicity), smoking status, postmenopausal status and hormone therapy use, and self-
reported histories of physician-diagnosed diseases (hypertension, dyslipidemia, and diabetes), as well as
other covariates (a family history of MI, height, and weight), were collected from questionnaires in both
study populations. Body mass index (BMI) (kg/m²) was calculated as weight in kilograms divided by the
square of height in meters. Self-reported weight was highly correlated with technician-measured weight
(r=0.97) in a validation study (37). The validity of self-reported histories of physician-diagnosed
hypertension, high cholesterol, and diabetes was also confirmed, as described previously (38, 39).
Antihypertensive medication use, statin or other cholesterol-lowering medication use, and insulin or oral
hypoglycemic medication use were also considered to define the respective metabolic conditions of
hypertension, hyperlipidemia, and diabetes. Estimated glomerular filtration rate (eGFR) was calculated
using the Chronic Kidney Disease Epidemiology Collaboration equation (40) for the NHS subsamples
with creatinine data. Physical activity, metabolic equivalent hours per week, was assessed based on the
duration of physical activities multiplied by the intensity of the activity. Diet and nutrient intake were
assessed using validated semiquantitative food frequency questionnaires. The Alternate Healthy Eating
Index (AHEI) (41) and the PDI were calculated using the previously described methods (42). In the PDI,
all animal foods (typically, these are more protein-rich than non-animal foods) get reverse scores, and non-animal foods (i.e., plant-based foods) get positive scores in PDI.

Statistical analysis

Data on PAGln were log-transformed before the analyses. In the WLVS, we calculated the β effect size for differences in log-transformed PAGln per 1 SD increment of dietary intake of individual food items using the linear regression models adjusting for age, race, fasting status, female hormone use, smoking, total energy intake, and BMI. Sensitivity analysis was performed using models after further adjusting for TMAO and precursor metabolites of PAGln. To examine the relations between PAGln and cardiometabolic risk factors of CHD, we calculated the effect size β per 1 SD increment of PAGln for differences in each outcome measurement after adjusting for age, race, fasting status, female hormone use, smoking habit, and BMI. Also, the odds ratios for the presence of hypertension, dyslipidemia, or diabetes per 1 SD increment of PAGln were calculated using the logistic regression models adjusting for the same covariates.

In nested case-control studies with risk-set sampling, the effects estimated using conditional logistic regression are odds ratios which are unbiased estimates of hazard ratios, thus, the relative risks (RRs). Conditional logistic regression was used to estimate the RRs and 95% CI for CHD incidence per 1 SD or categories of PAGln. Before performing the main analysis, we confirmed no significant interaction between PAGln and fasting status for the CHD outcome. We fitted restricted cubic splines (43) to model a dose-response relationship and examined possible nonlinear relationships between PAGln levels and CHD risk in the basic model adjusted for matching factors. The adjusted model included covariates of postmenopausal hormone use, family history of MI, physical activity, alcohol, overall diet quality (assessed by the AHEI without alcohol), and BMI. To minimize missing data on common covariates, we carried forward values obtained in previous questionnaires; 99% of the participants had data on these
covariates. If participants still had missing values on any one of the covariates (n=3 to n=9 across variables), these were replaced by a median value in the case group or in the control group, respectively.

To test whether the PAGln-CHD relation was modified by PDI, we calculated the risk of CHD by stratifying participants based on low or high PDI; the definition of low/high group was on the median value of the PDI in controls. Unconditional logistic regression was used in the stratified analyses to preserve statistical power, as matched cases and controls were not necessarily in the same strata. The multiplicative interactions were tested by including a cross-product term in the model. Sensitivity analyses were performed to test whether plasma levels of precursor metabolites of PAGln, an atherogenic gut-microbiota-related metabolite TMAO, or kidney function (measured by eGFR) modified the PAGln-CHD relation. Analyses were performed using R and SAS version 9.4 (SAS Institute).

Results

Daily dietary intakes based on the 7DDRs of women in the WLVS (n=725) are presented in Table 2 (32). Dietary protein and phenylalanine intakes were positively related to plasma PAGln levels (Figure 1); however, the associations were not significant at nutrient levels, and we found that protein-rich food items were differently related to PAGln levels. Total red and processed meat (p= 0.015), but not poultry or fish/seafood (p >0.05 for both), were significantly associated with higher levels of PAGln at the time of blood collection (white bars in Figure 1). Higher vegetable intake was related to lower PAGln levels (p= 0.024). Results were similar when two 7DDRs that reflected longer-term dietary intakes were used for the analysis (black bars in Figure 1). We observed similar findings in sensitivity analyses adjusting for circulating levels of a gut-microbiota-related metabolite TMAO and amino acid precursor metabolites (phenylalanine and glutamine) (Supplemental Figure 9 (32)). Table 1 shows associations of PAGln levels with cardiometabolic risk factors in the WLVS. Higher PAGln levels were associated with higher triglyceride concentrations and lower levels of HDL cholesterol and LDL cholesterol after adjusting for covariates.
including BMI (p < 0.05 for all). Higher PAGln levels were related to higher levels of proinsulin (p=0.007), C-peptide (p=0.002), and insulin (p=0.075). Also, higher PAGln levels were related to a greater probability of having hypertension, dyslipidemia, and diabetes (Table 3 (32) and Figure 10 (32)). Most of the participants (92%) fasted for at least 8 hours at the time of blood collection; results of sensitivity analysis after excluding those in a non-fasting state were essentially the same (data not shown).

Table 2 shows characteristics of the NHS participants by the case-control status (n=1520). Overall, every 1 SD increment of PAGln was associated with a RR of 1.11 (95% CI: 1.003, 1.22) for CHD. Results of dose-response analysis (Figure 2) showed that the risk of CHD was significantly elevated when exceeding a particularly high value (at 9.07) of log-transformed PAGln. When we calculated the RRs according to different categories of PAGln (Table 4 (32)), higher values within the top tertile of PAGln were particularly associated with an increased risk of CHD in basic model (RR 1.44 [95% CI:1.08, 1.93]) as well as in an adjusted model (RR 1.43 [1.05, 1.93]), as compared to women with low PAGln levels (the lowest tertile).

We then tested whether the PAGln-CHD relation was modified by dietary PDI, which assesses cumulative habitual intakes of animal foods and plant foods. We found significant interactions between PAGln and dietary PDI for the risk of incident CHD (Table 3). In a low-PDI group (that included women with higher intake of animal foods and lower intake of plant foods), every 1 SD of PAGln was associated with a 25% (RR 1.25 [1.08, 1.44]) increased risk of CHD in the basic model, and a 22% (RR 1.22 [1.05, 1.41]) increased risk of CHD in the adjusted model. On the other hand, there was no increased risk of CHD associated with PAGln in a high-PDI group. Among the low-PDI group, PAGln levels within slightly elevated ranges (log-transformed PAGln values of 6.64 or higher) were significantly associated with an increased risk of CHD (Figure 3).

Sensitivity analyses showed that the interactions between PAGln and dietary PDI for CHD risk remained significant when adjusting for levels of amino acid precursor metabolites (phenylalanine and
Also, the significant interaction association between PAGln and dietary PDI for CHD was independent of TMAO levels (P for interaction=0.01); higher levels of PAGln were significantly related to the risk of CHD (RR 1.17 [1.01, 1.36]) when adjusting for TMAO levels among participants with low dietary PDI. Similar associations were found when kidney function was added as a covariate (STable 5 (32)).

Discussion

We showed that high levels of circulating gut microbial metabolite PAGln were associated with a significantly increased risk of incident CHD, particularly among women with dietary patterns of eating more animal foods and fewer plant-based foods. Higher intakes of protein-rich animal-based foods and plant-based foods were significant dietary factors associated with higher or lower PAGln levels. Our study indicates the importance of plasma PAGln levels linked to CHD risk, their association with dietary intake assessed via gold-standard 7DDRs, and significant interactions between a plant-based diet and PAGln in the long-term incidence of CHD.

Our findings that high levels of PAGln were related to cardiometabolic risk factors and an increased CHD risk are supported by previous evidence. Recent cross-sectional studies have shown that high PAGln levels were related to in-stent hyperplasia and stenosis in CAD patients (16) and coronary atherosclerotic plaque burden among patients with suspected CAD (17). In several studies of clinical cohorts among patients at high risk of CVD (such as those with chronic kidney disease or those with suspected or prevalent CAD) (4, 7), higher levels of circulating PAGln were associated with higher risks of major adverse cardiovascular disease events and deaths (MACEs). In a recent pooled-cohort analysis of Swedish adults, plasma PAGln was associated with an increased risk of future CAD (18), but PAGln showed no significant associations for cardiometabolic risk factors in a population cohort (18). Results of women without CVD from the WLVS showed that plasma PAGln was a significant metabolite associated with
intermediate common metabolic risk factors of CHD (such as triglycerides, HDL cholesterol, and markers of insulin secretion). Interestingly, higher PAGln levels were related to lower LDL cholesterol levels in previous studies (7, 44), and we found the same association in the WLVS. Also, we found a nonlinear association between PAGln levels and CHD risk in the total participants when performing restricted cubic splines, suggesting that the risk of CHD was significantly increased when the accumulation of PAGln in circulation was over a specific limit. The observed nonlinear relation is partly in line with a previous study of patients with suspected/prevalent CAD, showing that subjects with particularly high PAGln levels exhibited a significantly increased risk of incident MACE (4).

We found a more robust PAGln-CHD relation in people with lower adherence to a plant-based diet, which was characterized as having more animal foods and fewer plant foods. The different sources of dietary protein influence the effects of microbiota-derived metabolites on host physiology (3, 45), and our results based on 7DDRs showed that greater intakes of red meat and processed meat, but not poultry, fish/seafood, or total dietary protein, were significantly associated with higher levels of plasma PAGln. In a recent short-term (12-week) dietary intervention study in metformin-treated patients with type 2 diabetes, an increase in dietary protein (mainly due to an increase in animal protein) significantly increased circulating PAGln levels (23). The present study results are also supported by previous intervention studies with a relatively small sample size (24, 26), suggesting that a high-protein diet or acute ingestion of bovine meat may induce increased serum levels and urinary excretion of PAGln (24, 26). Higher vegetable intake was related to lower PAGln levels, which was also supported by prior observational studies assessing urinary PAGln concentration and vegetable intake (25) or a high (vs. low) adherence to a Mediterranean diet (27). In addition, the increased risk of CHD associated with an atherogenic gut-microbial metabolite in people with lower adherence to healthy dietary habits was also found in our previous study, which showed significant interactions between TMAO and adherence to dietary patterns with lower meats and higher vegetables in CHD risk (5). In the present study, higher PAGln was related to greater CHD risk.
independently of plasma TMAO levels among people with lower adherence to plant-based dietary
patterns. These results underscore the importance of habitual dietary patterns in modifying the relationship
between the novel atherogenic gut-microbial metabolite PAGln and subsequent CHD risk. Further
prospective studies analyzing temporal changes in PAGln and gut dysbiosis are needed to clarify whether
CHD risk-related habitual dietary patterns might trigger an altered production of PAGln and its circulating
levels to increase the subsequent risk of CHD.

As to biological mechanisms, studies in animal models suggest that PAGln may be involved in
promoting CVD-related risk factors via host G protein-coupled receptors, including a2A, a2B, and b2-
adrenergic receptors (4). Adrenergic receptors are known to play important roles in myocardial and
vascular function (46). It has also been suggested that adverse CVD-related phenotypes observed with
PAGln at physiological levels may be attenuated with the β-blocker therapy (4). Also, insulin may affect
β adrenergic receptor signaling in the heart to modulate cardiac function (14). Several studies have
reported associations between circulating levels of PAGln and gut microbiome diversity (18, 47) and
different families of gut microbes (48, 49). In addition, a recent study identified two distinct gut microbial
pathways for phenylacetic acid formation (10), and also found a higher abundance of both pathways in
gut microbiomes of patients with atherosclerotic CVD compared to controls (10). Further investigations
are warranted to understand potential biological mechanisms underlying our findings.

Our study has several strengths. The study participants were initially free of major chronic diseases
such as CVD and kidney failure, which may affect increases in PAGln levels attributable to renal function
decline and its reduced clearance (50, 51). Our results from a traditional cohort contribute to a better
understanding of circulating metabolite precursors for CHD events among US women. Also, the study has
an adequate number of incident CHD cases with long-term prospective follow-ups. The use of 7DDRs
also contributes to the robustness of our findings on PAGln and dietary intakes. However, there are also
several potential limitations. The assessment of PDI and other covariates included in the multivariate
analysis were based on self-reports, which might affect the estimated risk of CHD; however, previous validation studies have shown the validity of self-reported data in the NHS (30, 38, 39). In the present study, not all participants had data on creatinine to estimate GFR, and we could not adjust for kidney function decline in the main analysis. Our study included women only, and all were health professionals. Whether our findings could be applicable to other populations needs to be further investigated, especially in male cohorts and a population that is more representative of the US population.

In conclusion, circulating PAGln was related to the risk of CHD, particularly in women with dietary habits of eating more animal foods and less plant-based foods. Adherence to plant-based diets attenuated unfavorable associations between PAGln and CHD risk, highlighting the importance of interplays of diet and the novel microbial metabolite in relation to long-term CHD risk.

**Data availability:** Datasets generated during and/or analyzed during the current study are not publicly available but are available upon reasonable request. Further information, including the procedures to obtain and access data, is described at: https://www.nurseshealthstudy.org/researchers.

**Acknowledgments:** We appreciate all the participants of the Nurses’ Health Study (NHS) and NHSII for their continued cooperation. We thank the Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School for their assistance.

**Author Contribution:** YH contributed to the study concept, statistical analysis, interpretation of data, and drafting and revising the manuscript. ST, JDW, MA, and MJ contributed to measurements, interpretation of data, and revising the manuscript. XW contributed to the analysis and interpretation of data and revising the manuscript. KMR, FBH, and QS contributed to interpretation of data and revising the manuscript. JEM and LQ contributed to the study concept, data acquisition, interpretation of data, drafting and revising the manuscript, funding, and study supervision. All authors contributed to the
manuscript and approved the final version and have access to all data in the present study. The corresponding authors, YH and LQ, take responsibility for the integrity of the data of the work as a whole.

References


32. Yoriko Heianza, Saumya Tiwari, Xuan Wang, Jeramie D Watrous, Kathryn M. Rexrode, Frank B. Hu, Mona Alotaibi, Mohit Jain, Qi Sun, JoAnn E. Manson, Lu Qi. ONLINE SUPPLEMENTAL MATERIALS. Article: Gut-Microbiota-Related Metabolite Phenylacetylglutamine and Risk of Incident Coronary Heart Disease among Women. doi:10.6084/m9.figshare.25733451. 2024.


**Figure titles and legends**

**Figure 1**: Differences in phenylacetylglutamine (PAGln) levels per 1 SD increment of major protein-rich animal foods and plant-based foods.

Abbreviations: PAGln, phenylacetylglutamine. $\beta$ (SE) per 1 SD increment of each food item for plasma PAGln levels after adjusting for age, race, fasting status, female hormone use, smoking, total energy intake, and body mass index. The definition of SD is shown in supplemental table 2.

**Figure 2**: Risk of coronary heart disease (CHD) by phenylacetylglutamine (PAGln) levels.

Abbreviations: CHD, coronary heart disease; PAGln, phenylacetylglutamine; RR, relative risk. Spline model with 4 knots with the minimum value in the control group as the reference value. Relative risks after adjusting for matching factors (age, smoking, date of blood sample collection, and fasting status). $P$ for test of curvature = 0.0475; $P$ for test of overall significance of the curve= 0.0177. Solid line shows RRs, and dotted lines show 95% CIs.

**Figure 3**: Risk of coronary heart disease (CHD) by phenylacetylglutamine (PAGln) levels in participants with low (panel A) or high (panel B) plant-based diet index (PDI).

Abbreviations: CHD, coronary heart disease; PAGln, phenylacetylglutamine; RR, relative risk. RRs (95% CIs) using unconditional logistic regression adjusted for matching factors, postmenopausal hormone use, family history of myocardial infarction, physical activity, alcohol intake, and body mass index. Low/high plant-based diet index (PDI) was based on the median value of PDI. Spline analysis with 4 knots; the
lowest and highest 1% were excluded in each group. RRs were estimated using a minimum value in the low PDI group as a reference. Solid lines show RRs, and dotted lines show 95% CIs.

Table 1: Association of $\beta$ (SE) per 1 SD increment of log-transformed phenylacetylglutamine (PAGln) with differences in lipids and insulin-related biomarkers in the Women's Lifestyle Validation Study

<table>
<thead>
<tr>
<th>Outcome</th>
<th>N</th>
<th>Mean of outcome variable</th>
<th>$\beta$ (SE)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-transformed triglycerides</td>
<td>725</td>
<td>4.62</td>
<td>0.03 (0.02)</td>
<td>0.029</td>
</tr>
<tr>
<td>HDL cholesterol, mg/dl</td>
<td>725</td>
<td>70</td>
<td>-1.2 (0.6)</td>
<td>0.041</td>
</tr>
<tr>
<td>LDL cholesterol, mg/dl</td>
<td>722</td>
<td>124</td>
<td>-3.3 (1.2)</td>
<td>0.007</td>
</tr>
<tr>
<td>log-transformed proinsulin*</td>
<td>426</td>
<td>2.12</td>
<td>0.07 (0.03)</td>
<td>0.007</td>
</tr>
<tr>
<td>log-transformed C-peptide*</td>
<td>479</td>
<td>0.29</td>
<td>0.06 (0.02)</td>
<td>0.002</td>
</tr>
<tr>
<td>log-transformed insulin*</td>
<td>411</td>
<td>1.94</td>
<td>0.04 (0.02)</td>
<td>0.075</td>
</tr>
</tbody>
</table>

N, number of participants with available outcome data. $\beta$ (SE) per 1 SD increment of log-transformed PAGln for the respective outcome after adjusting for age, race, fasting status, female hormones use, smoking habit, and body mass index. *Proinsulin, C-peptide, and insulin were analyzed in people without diabetes.

Table 2: Baseline characteristics of the Nurses’ Health Study participants according to case-control status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Incident CHD cases</th>
<th>Controls</th>
<th>$P$ value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  Values</td>
<td>N  Values</td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td>760  64.6 (7.9)</td>
<td>760  64.6 (7.9)</td>
<td>†</td>
</tr>
<tr>
<td>Smoking habit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>760  297 [39]</td>
<td>760  297 [39]</td>
<td>†</td>
</tr>
<tr>
<td>Former</td>
<td>760  321 [42]</td>
<td>760  333 [44]</td>
<td>-</td>
</tr>
<tr>
<td>Current</td>
<td>760  142 [19]</td>
<td>760  130 [17]</td>
<td>-</td>
</tr>
<tr>
<td>Postmenopausal hormone use</td>
<td>760  315 [41]</td>
<td>760  330 [43]</td>
<td>0.44</td>
</tr>
<tr>
<td>Family history of myocardial infarction</td>
<td>760  222 [29]</td>
<td>760  150 [20]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Physical activity, MET-h/week  760  15.3 (17.0)  760  17.8 (19.4)  0.001  
Alcohol, g/d  760  4.8 (9.4)  760  5.6 (9.5)  0.004  
Alternate Healthy Eating Index without alcohol  760  48 (10)  760  48 (10)  0.69  
Plant-based diet index score  751  54 (7)  742  54 (7)  0.31  
Body mass index, kg/m²  760  26.9 (5.2)  760  25.6 (4.5)  <0.001  
Hypertension  760  446 [59]  760  289 [38]  <0.001  
Dyslipidemia  760  409 [54]  760  310 [41]  <0.001  
Diabetes  760  106 [14]  760  27 [4]  <0.001  
Creatinine, mg/dl  494  0.80 (0.21)  491  0.77 (0.17)  0.034  
eGFR, mL/min/1.73 m²  494  81 (17)  491  82 (15)  0.12  

Data are mean (SD) or n [%]. †Matching factors. *P value estimates are based on Student’s t-test, 
Wilcoxon rank–sum test, or χ² test as appropriate. Abbreviations: CHD, coronary heart disease; MET-
h/week, Metabolic equivalent hours per week; eGFR, estimated glomerular filtration rate.

Table 3: Risk of coronary heart disease per 1 SD increment of phenylacetylglutamine in the stratified 
analysis by low or high adherence to a plant-based diet

<table>
<thead>
<tr>
<th>Model</th>
<th>Plant-based diet index</th>
<th>P for interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Basic model, RR (95% CI)*</td>
<td>1.25 (1.08, 1.44)</td>
<td>0.98 (0.85, 1.14)</td>
</tr>
<tr>
<td>P value</td>
<td>0.002</td>
<td>0.81</td>
</tr>
<tr>
<td>Adjusted model, RR (95% CI)**</td>
<td>1.22 (1.05, 1.41)</td>
<td>0.96 (0.82, 1.11)</td>
</tr>
<tr>
<td>P value</td>
<td>0.009</td>
<td>0.56</td>
</tr>
</tbody>
</table>

RR, relative risk. The median value of the plant-based diet index was used to determine low/high 
adherence to a plant-based diet. *RRs after adjusting for matching factors; **RRs after further adjusting 
for postmenopausal hormone use, family history of myocardial infarction, physical activity, alcohol, and 
body mass index. RRs and P for interaction were calculated using unconditional logistic regression.
Figure 1

207x250 mm (DPI)
Figure 2

RR for CHD

Log-transformed PAGln levels

123x114 mm (DPI)
Figure 3

200x111 mm (DPI)