Evaluation of a novel food composition database that includes glutamine and other amino acids derived from gene sequencing data

Published in final edited form as:
Eur J Clin Nutr. 2009 December ; 63(12): 1433–1439. doi:10.1038/ejcn.2009.110.

Evaluation of a novel food composition database that includes glutamine and other amino acids derived from gene sequencing data

CM Lenders[1,2,3,4], **S Liu**[1,5], **DW Wilmore**[3,6], **L Sampson**[1,7], **LW Dougherty**[1,8], **D Spiegelman**[1,9], and **WC Willett**[1,10]

[1]Harvard School of Public Health, Boston, MA, USA

[2]Boston University School of Medicine, Boston, MA, USA

[3]Harvard School of Medicine, Boston, MA, USA

Abstract

Objectives—To determine the content of glutamine in major food proteins.

Subjects/Methods—We used a validated 131-food item food frequency questionnaire (FFQ) to identify the foods that contributed the most to protein intake among 70 356 women in the Nurses' Health Study (NHS, 1984). The content of glutamine and other amino acids in foods was calculated based on protein fractions generated from gene sequencing methods (Swiss Institute of Bioinformatics) and compared with data from conventional (USDA) and modified biochemical (Khun) methods. Pearson correlation coefficients were used to compare the participants' dietary intakes of amino acids by sequencing and USDA methods.

Results—The glutamine content varied from 0.01 to to 9.49 g/100 g of food and contributed from 1 to to 33% of total protein for all FFQ foods with protein. When comparing the sequencing and Kuhn's methods, the proportion of glutamine in meat was 4.8 vs 4.4%. Among NHS participants, mean glutamine intake was 6.84 (s.d.=2.19) g/day and correlation coefficients for amino acid between intakes assessed by sequencing and USDA methods ranged from 0.94 to 0.99 for absolute intake, −0.08 to 0.90 after adjusting for 100 g of protein, and 0.88 to 0.99 after adjusting for 1000 kcal. The between-person coefficient of variation of energy-adjusted intake of glutamine was 16%.

Conclusions—These data suggest that (1) glutamine content can be estimated from gene sequencing methods and (2) there is a reasonably wide variation in energy-adjusted glutamine intake, allowing for exploration of glutamine consumption and disease.

Correspondence: Dr CM Lenders, Division of General Pediatrics, Boston University School of Medicine, Nutrition and Fitness for Life program, Boston Medical Center–Vose Hall 3–Room 307, 88, East Newton Street, Boston MA 02118, USA. carine.lenders@bmc.org.

[4]Current address: Pediatric Nutrition Support Services, Boston Medical Center–Vose Hall 3–Room 307, 88, East Newton Street, Boston, MA 02118, USA

[5]Current address: Department of Epidemiology and Medicine, UCLA CHS 73-265 Box 951772, 650 Charles E Young Drive South, Los Angeles, CA 90095-1772, USA

[6]Current address: Department of Surgery, Brigham and Women's Hospital, Boston, MA 02115, USA

[7]Current address: Harvard School of Public Health, Bldg 2, Rm 315, Nutrition, 655 Huntington Avenue, Boston, MA 02115, USA

[8]Current address: Harvard School of Public Health-Nutrition, Bldg. 2, Room 311, 665 Huntington Avenue, Boston, MA 02115, USA

[9]Current address: Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

[10]Current address: Department of Nutrition, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA

Conflict of interest The authors declare no conflict of interest.

Keywords

food composition; gene sequencing; amino acids; database; glutamine; food frequency

Introduction

Increasing evidence suggests that dietary glutamine decreases insulin levels (Opara *et al*., 1996; Borel *et al*., 1998) and weight gain in animal studies (Opara *et al*., 1996). Therefore, dietary glutamine intake may affect diabetes risk.

The data on amino acid in foods are obtained primarily by ion-exchange chromatography following acid hydrolysis of proteins (Schegg *et al*., 1997) and more recently by measurement of protein-bound glutamine (Kuhn *et al*., 1996, 1999). Drawbacks of biochemical measurements include the conversion of glutamine to glutamate, especially in the acid hydrolysis method. Therefore, the content of glutamine in food proteins is not available in nutrient databases and that of glutamate is overestimated.

Given the historical lack of accurate measurement and absence from current nutrient databases, the purposes of this study were (1) to estimate the content of glutamine and glutamate in food using data from sequencing methods, (2) to compare the calculated proportion of glutamine in meat and casein with historical data derived from Kuhn's method (Kuhn *et al*., 1996, 1999), and (3) to include all 20 proteinogenic amino acids derived from the sequencing methods in the Nurses' Health Study nutrient database for further comparison with data compiled by the USDA.

Materials and methods

Study sample

The Nurses' Health Study is a prospective cohort study of diet and lifestyle factors in relation to chronic diseases among 121 700 female registered nurses aged 30–55 years at enrollment in 1976. We excluded women who did not satisfy the *a priori* criterion of reported daily energy intake between 2514 (600 kcal) and 14 665 kJ (3500 kcal), BMI between 15 and 49 kg/m^2 (=5 s.d.), available data on protein intake, and no previous diagnosis of diabetes, cardiovascular disease, and cancers. The final baseline population consisted of 70 356 women who were 38–63 years old in 1984. This study was approved by the Institutional Review Board at the Channing Laboratory.

Assessment of diet

In 1976, women completed questionnaires on their medical history and lifestyle. A 131-item food frequency questionnaire (FFQ) was first completed in 1984 and updated in 1986, 1990, 1994, 1998, and 2002. Nutrient intakes were computed from the reported frequency of consumption of each specified unit of food or beverage and from published data on the nutrient content of the specified portions. We used data on amino acids from The Food Composition Handbook 8 series (1976–1992) and serial releases 10, 14, and 16 published by the USDA (2006). This FFQ has been previously evaluated for validity of a variety of nutrients including protein intake (Willett *et al*., 1985).

Amino acid data derived from USDA publications—Amino acid values in food items with more than one protein-containing ingredient were derived from the amino acid composition of the various protein-containing ingredients. These USDA amino acid values were included in the protein-containing food items and ingredients of the NHS nutrient

database. The amino acids available in the USDA nutrient database include alanine, arginine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine. Cystine, an oxidized dimeric form of cysteine, is available rather then cysteine. There is no glutamine or asparagine while glutamate and aspartate are overestimated.

Data derived from the EXPASY Web site—Data on the sequence of thousands of proteins from a variety of organisms are compiled in the Swiss Institute of Bioinformatics (Bairoch and Apweiler, 2000). These data are available on the internet in Switzerland and mirror sites such as Canada http://ca.expasy.org/. SWISS-PROT is an annotated protein sequence database created at the Department of Medical Biochemistry of the University of Geneva with the European Molecular Biology Laboratory collaboration. SWISS-PROT is the most complete database on gene expression of amino acids in proteins on the Web that applies rigorous scientific methods including internal and external expertise.

Amino acids values derived from the EXPASY Web site—We calculated the proteinogenic amino acid content of food proteins consumed in the Nurses' Health Study nutrient database based on (1) the identification of protein fractions of food proteins reported in the literature, (2) the identification of the amino acid composition of protein fractions using the EXPASY protein sequence database server, (3) the weighted sum of amino acids content of protein fractions in food proteins, and (4) the incorporation of amino acid content of food proteins in recipes used to create the NHS study nutrient database. The amino acid content of food protein fractions was calculated using the entire sequence of the protein fraction examined. We decided to exclude the terminal residues from the amino acid calculations as terminal amino acids have a short lifetime (Bachmair *et al*., 1986), even though terminal amino acids have little impact on the final composition of proteins given their small contribution of the total sequence.

The formula to calculate the content of amino acids in food protein derived from sequencing data was described in earlier studies (Lacey and Wilmore, 1990; Swails *et al*., 1992). As an example, the glutamine content of β-casein in milk was calculated as: [(% β-casein × # GLN in sequence × molecular weight of GLN (g/mole)]/molecular weight of β-casein (g/mole)=(23.4%×20×146.15)/(23583.2)=2.9%. The same procedure was repeated for each of the 20 amino acids in each of the known protein fractions of a food.

The sum of all protein fractions derived from the Expasy Web site contributed 90–98% of the total protein weight of all foods. The 28 foods for which the amino acid content was derived from the Expasy Web site included barley, beef, brown rice, casein, cocoa, coffee, corn, egg yolk, egg white, egg, kidney bean, lentil, milk, oat, oat bran, pea, peanut, potato, sesame, soybean, sweet potatoes, tomato, walnut, wheat, wheat bran, wheat gluten, whey, and white rice. The site was last accessed on 26 February 2004.

Assumptions—Given the low amount of protein and undetectable amounts of measured glutamate in food items such as fruits and vegetables, as well as the lack of sequences in the EXPASY Web site for these plant proteins, the amino acid content of these foods were based on one plant protein (soy). Several assumptions were necessary to calculate the amino acid composition of food proteins. For example, the sequence of chicken serum albumin was available in the EXPASY Web server. The sequence of chicken serum albumin, which is identical to α-livetin, was used to represent the category of livetins. Livetins correspond to about 7% of yolk solids. We also needed to make other assumptions concerning the proportions of food protein in protein fractions. For example, 22 proteins had been sequenced as wheat zeins in SWISS_PROT, we assumed that each of them contributed equally to the total amount of zeins. We also lacked information on the sequence of proteins

such as enzymes. However, the contribution of these proteins to the total weight of amino acids in food proteins is probably negligible for nutritional purposes as they are found in very small proportions.

Milk protein is presented as an example to show how we used protein fractions to calculate total protein. The protein fractions in milk include caseins such as α-casein, β-casein, κ-casein, and γ-casein as well as β-lactoglobulin, α-lactalbumin, immunoglobulins (IGGs, IGM, IGA, FSC), albumin, and proteose-peptone. Together, these protein fractions contribute to nearly 98% of the total weight of milk protein. As the weight of the remaining thousands of proteins that contribute to milk protein are proportionally small (2%), the contribution of these proteins to amino acids content of milk protein is probably negligible for dietary purposes. Likewise, free amino acids found in foods were not included in the final calculation of amino acids in food proteins because of their short half-life and small contribution to total nitrogen in food (Schloerb *et al*., 2002). For example, free amino acids contribute to <2% of the total protein in yolk solids (Osuga and Feeney, 1977) and breast milk (Wu *et al*., 2000). Finally, we used similar food proteins for missing food items. For example, beef amino acids were used for all muscle proteins.

Statistical methods

Values for each amino acid percentage (g/100 g protein) were calculated based on gene expression of protein fractions from the EXPASY Web site. The proportion of glutamine in meat and casein protein was further compared with historical data measured using Kuhn's modified biochemical technique (Kuhn *et al*., 1996, 1999; Baxter *et al*., 2004). Group means and standard deviations of amino acids intake were calculated from FFQs (g/day). Participants' amino acid intakes (g/day) were calculated for 100 g protein and for 1000 kcal. We used the Pearson correlation coefficients to compare the participants' dietary intakes of amino acids using the sequencing and USDA methods (1994), as they were similar to Spearman correlations (data not shown). Finally, the between-person coefficient of variation (CV) of energy-adjusted and nonadjusted amino acid intake was calculated based on the adjusted or nonadjusted amino acid standard deviation divided by its mean among the NHS participants.

The amino acid content of food proteins derived from the EXPASY Web site was calculated using STATA 7.0 (stata@stata.com). The comparisons between amino acid consumption of women assessed by the gene sequencing method and by the conventional biochemical methods were performed using SAS software (version 8; SAS Institute Inc, Cary, NC, USA). All *P*-values were two-sided.

Results

Glutamine content varied from 0.01 to 9.49 g/100 g of food (from apple juice to wheat germ) and contributed from 1 to 33% (g/100 g) of protein (from kidney bean to whole grain bread). Using data from the sequencing compared with Kuhn's modified biochemical method (Table 1), the amount of glutamine was similar for meat protein (4.8 vs 4.4%) and casein-based formula protein (8.7 vs 9.2%).

Table 2 shows data on the amino acid composition of selected foods that may have a function in chronic diseases. Using the gene sequencing method, the percentage of glutamine was 4.4% for egg, 4.8% for beef, 8.1% for milk, 9.1% for Tofu (soy), 11.1% for white rice, and 16.2% for corn protein. Thus, the total amount of glutamine per 100 g among those six foods increased from 0.28 g in milk protein to 1.23 g in beef protein with intermediate values for white rice protein, corn protein, egg protein, and soy protein (Table 2).

As shown in other studies, the largest contribution to food protein intake in the Nurses' Health Study came from animal sources (>70%). Among the top 30 food proteins contributing to protein intake, beef contributed to 15% of total food protein intake among these women and clam chowder contributed to 0.7%. The amount of dietary protein in the top 30 food items varied from 0.7 g/100 g in orange juice to 28.9 g/100 g in chicken without skin. Using the gene sequencing method, the 20 amino acids in the top 30 food proteins contributed by definition to 100% (g/100 g protein) of each of the food proteins. However, the percentage of protein for the 18 amino acids available in the USDA database excluding asparagine and glutamine was already at or above 100% for 8 of the 30 food item proteins listed in Table 3 (American cheese, cottage cheese, whole grain, mashed potato, yogurt, shrimps, English muffin, and peanut butter). Although large differences were observed in the contribution of amino acid data to specific protein foods using conventional methods, the correlation coefficients among specific foods for amino acid composition in 100 g of food were high.

Among women in the NHS, the consumption of amino acids in (g/day) was comparable between methods (Table 4). The intake of cystine using the USDA food composition data were strongly correlated with those of cysteine calculated from gene sequencing data (0.98). After excluding asparagine, aspartate, glutamine, and glutamate, the correlation coefficients for the 16 protein-adjusted (g/100 g protein) proteinogenic amino acids assessed by the two methods, ranged from −0.08 for tryptophan and serine to 0.90 for arginine. After adjusting for energy (g/1000 kcal), the correlation coefficients between the two methods were higher, varying from 0.88 for tryptophan to 0.99 for arginine (Table 4). Using the sequencing method, the between-person CV for absolute intake of the 20 amino acids ranged from 31 to 34% for crude intake, 4–14% after adjusting for protein, and 16–25% after adjusting for 1000 kcal. Glutamine had a CV of 32% for crude intake, 14% for intake of 100 g of protein, and 16% for intake of 1000 kcal.

Discussion

The main finding of this study is that the sequencing and modified biochemical methods provide similar proportions of glutamine in proteins from meat and casein-based formulas, within the measurement error of Kuhn's method. Although modified biochemical methods may be accurate, they are expensive, require specialized laboratories, and do not provide information on most amino acids. Thus, identifying a method to estimate the amino acid content of food protein that may be cheaper than and as accurate as a complex biochemical method has great advantage.

The gene sequencing technique was extended to the composition of the other proteinogenic amino acids in food proteins. After exclusion of glutamine, glutamate, asparagine, and aspartate, which are absent from the USDA food composition database or overestimated, the participants' daily intake of amino acids assessed by the two methods was highly correlated before and after adjusting for energy intake. However, after adjusting for 100 g of total protein intake, a large change in the correlation coefficients was observed for some of the amino acids, which may be explained in part by several factors.

First, the amino acid data from the conventional method contributed from 43 to 111% of proteins (g/100 g) for 18 amino acids whereas that from the sequencing method contributed by definition to 100% (Table 3) for all 20 amino acids. Second, adjusting for total protein provides an estimate of the composition of food proteins and reduces the between-person variability of the data, which will also tend to reduce correlation coefficients. Third, measurements of amino acids may vary by type of food or laboratory.

Although, the FAO/WHO consultation in 1989 reported an inter-laboratory error of 10% for amino acid analysis (Joint FAO/WHO Expert Consultation, 1989), results from a 1996 multi-laboratory study comparing amino acid measurements in one protein showed a wide range of error from 4.0 to 58.9% for proteinogenic amino acids excluding measurements of asparagine, glutamine, cystine, and tryptophan (Schegg *et al*., 1997). The assessment of amino acids in a few foods including casein, soy, pea flour, whole-wheat flour, egg white solids, minced beef, and rapeseed concentrate showed that the between-laboratory CV were better for isoleucine, leucine, lysine, phenylalanine, threonine, and valine (CV <10%) than cystine, methionine, and tryptophan (CV=10– 20%) (Friedman, 1996). In that study, the between-laboratory CV were also better for casein, soy, and minced beef (CV <7%) compared with the other foods analyzed (CV 10–24%). Overall, these data show that conventional biochemical measurements vary substantially not only between laboratories, but also with type of food and amino acid.

Finally, although the data compiled by the USDA includes cooked and processed food when available, that derived from gene sequencing does not. During processing of foods, protein sources may be treated with heat, oxidizing agents, organic solvents, alkalis, and acids. These treatments may lead to the formation of multiple compounds that result in lower amino acid availability and protein quality. Although decreased amino acid availability is of concern, studies involving the ingestion of ^{15}N-labeled dietary proteins show that the true ileal digestibility of a number of protein sources, including milk, cereals, and legumes, is >90% and that it varies only minimally among the common sources of dietary protein (Reeds and Garlick, 2003). On the basis of that review and the inadequacy of available scoring patterns (Sarwar, 1997), we did not use a factor to account for amino acid availability or protein quality.

Results from the conventional method confirm official recommendations that other biochemical methods than the traditional acid hydrolysis of peptidic links and separation by chromatography be used. Other biochemical methods would allow for a more precise identification of glutamine, asparagine, and tryptophan, which are typically destroyed, but also to some extent the sulfur amino acids such as methionine, cysteine, or cystine (Schegg *et al*., 1997). The strength of the gene sequencing method is that the same methodology was applied for all 20 amino acids derived from sequencing data available in the EXPASY Web site.

Limitations

The use of the gene sequencing method includes several assumptions to calculate the amount of amino acids in food proteins. Although a series of assumptions are made for both the biochemical and gene sequencing methods, those used for the gene sequencing method are made without discrimination by amino acids as all values are calculated based on the same sequence vs separate biochemical methods. Future updates using more recent values of amino acids compiled by the USDA and additional data on gene sequences to minimize assumptions might improve the correlation between the two methods. We did not account for cooked and processed foods, digestibility and quality of protein, as valid data are not available. Although new scoring systems to account for food processing may improve the correlation between the methods, the contribution to actual protein absorption of specific food proteins is expected to be small.

Conclusions

These data suggest that the glutamine content of food protein can be estimated from gene sequencing methods. Furthermore, there is a reasonably wide variation in glutamine intake,

allowing for examination of glutamine consumption and disease risk after adjustment for energy intake.

Acknowledgments

We are indebted to the Nurses continuous participation in the Nurses' Health study. We particularly thank the Swiss Institute of Technology for providing free access to their Web site, the Ross laboratory for teaching CL how to navigate on the Expasy Web site, and Martin Van Denburgh for programming support. Supported by 5T32DK007703-07 and K23 DK082732.

References

Bachmair A, Finley D, Varshavsky A. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science. 1986; 234:179–186. [PubMed: 3018930]

Bairoch A, Apweiler R. SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acid Res. 2000; 28:45–48. [PubMed: 10592178]

Baxter JH, Phillips RR, Dowlati L, Johns PW. Glutamine in commercial liquid nutritional products. J Agric Food Chem. 2004; 52:4963–4968. [PubMed: 15291459]

Borel MJ, Williams PE, Jabbour K, Levenhagen D, Kaiser E, Flakoll PJ. Parenteral glutamine infusion alters insulin-mediated glucose metabolism. JPEN. 1998; 22:280–285.

Composition of Foods Loose-leaf Agriculture Handbook No 8 Series. Agricultural Research Service. United States Department of Agriculture (USDA). US Government Printing Office; Washington DC: 1976–1992.

Friedman, M. Nutrition. In: Nakai, S.; Modler, W., editors. Food Proteins: Properties and Characterization. Vol. Chapter 6. Wiley-VCH; New York: 1996. p. 281-298.

Kuhn KS, Schumann K, Stehle P, Darmaun D, Furst P. Determination of glutamine in muscle protein facilitates accurate assessment of proteolysis and *de novo* synthesis-derived endogenous glutamine production. Am J Clin Nutr. 1999; 70:484–489. [PubMed: 10500016]

Kuhn KS, Stehle P, Furst P. Quantitative analysis of glutamine in peptides and proteins. J Agric Food Chem. 1996; 44:1808–1811.

Lacey JM, Wilmore DW. Is glutamine a conditionally essential amino acid? Nutr Rev. 1990; 48:297–309. [PubMed: 2080048]

Opara EC, Tevrizian A, Feinglos MN, Surwit RS. l-glutamine supplementation of a high fat diet reduces body weight and attenuates hyperglycemia and hyperinsulinemia in C57BL/6J mice. J Nutr. 1996; 126:273–279. [PubMed: 8558312]

Osuga, DT.; Feeney, RE. Egg proteins. In: Whitaker, JR.; Tannenbaum, SR., editors. Food Proteins. AVI Pub Co., Inc.; Westport, CT: 1977. p. 209-221.

Joint FAO/WHO Expert Consultation. Protein quality evaluation: report. Bethesda, MD, USA: 1989 December 4-8. www.fao.org/docrep/U5900t/u5900t07.htm

Reeds PJ, Garlick PJ. Protein and amino acid requirements and the composition of complementary foods. J Nutr. 2003; 133(Suppl 9):2953–2961.

Sarwar G. The protein digestibility-corrected amino acid score method overestimates quality of proteins containing antinutritional factors and of poorly digestible proteins supplemented with limiting amino acids in rats. J Nutr. 1997; 127:758–764. [PubMed: 9164998]

Schegg, KM.; Denslow, ND.; Andersen, TT.; Bao, YA.; Cohen, SA.; Mahrenholz, AM., et al. Quantitation and identification of proteins by amino acid analysis. In: Marshak, DR., editor. Techniques in Protein Chemistry VIII. Academic Press; San Diego: 1997. p. 207-216.For more information, please see http://www.abrf.org/index.cfm/group.show/AminoAcidAnalysis.19.htm

Schloerb PR, Cook LT, Hall TJ. Digestive release of glutamine from enteral wheat protein hydrolysate. Am J Clin Nutr. 2002; 75(Suppl 2):403. abstract.

Swails WS, Bell SJ, Borlase BC, Forse RA, Blackburn GL. Glutamine content of whole proteins: implications for enteral formulas. Nutr Clin Pract. 1992; 7:77–80. Erratum in: *Nutr Clin Pract* 1992; 7(3): 133–134. [PubMed: 1294883]

US Department of Agriculture, Agricultural Research Service. USDA. Nutrient Database for Standard Reference, Release 1–16. Nutrient Data Laboratory; 2006. For more information, please see http://www.ars.usda.gov/Aboutus/docs.htm?docid=9418andwww.nal.usda.gov/fnic/foodcomp/Data/SR16-1/sr16_doc.pdf

Willett W, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semi quantitative food frequency questionnaire. Am J Epidemiol. 1985; 122:51–65. [PubMed: 4014201]

Wu ZC, Chijang CC, Lau BH, Hwang B, Sugawara M, Idota T. Crude protein content and amino acid composition in Taiwanese human milk. J Nutr Sci Vitaminol (Tokyo). 2000; 46:246–251. [PubMed: 11234918]

Table 1

Glutamine in meat protein and casein-based formula protein (g/100 g protein)

Amino acid	Meat		Casein-based formula	
	Sequencing method	Modified biochemical[a,b]	Sequencing method	Modified biochemical[a,c]
Glutamine (%)	4.8	4.4	8.7	9.2
Glutamate (%)	10.3	13.2	11.2	13.6

[a] Estimated from GLX=glutamic acid+glutamine+pyroglutamic acid, obtained from dipeptide hydrolysis where glutamate is overestimated (GLX–glutamine).

[b] Adapted from Kuhn *et al*. (1999).

[c] Adapted from Baxter *et al*. (2004).

Table 2

Total protein and composition in amino acid (g/100 g food) of selected foods derived from the conventional method (USDA) and from the gene sequencing method (Study)

Total protein in food	Beef (g)		Skim milk (g)		White rice (g)		Corn (g)		Tofu (soy) (g)		Egg (g)	
	25.9		3.4		2.7		2.5		6.6		12.6	
Method	**USDA**	**Study**	**USDA**	**Study**	**USDA**	**Study**	**USDA**	**Study**	**USDA**	**Study**	**USDA**	**Study**
Sum of amino acids	25.0	25.9	3.2	3.4	2.7	2.7	2.6	2.5	6.3	6.6	12.5	12.6
Alanine	1.577	1.560	0.100	0.141	0.156	0.140	0.172	0.216	0.268	0.211	0.696	0.682
Arginine	1.677	1.604	0.072	0.106	0.224	0.282	0.126	0.091	0.436	0.519	0.749	0.813
Asparagine		0.827		0.119		0.152		0.074		0.399		0.520
Aspartate	2.363	1.414	0.243	0.103	0.253	0.082	0.219	0.037	0.724	0.376	1.255	0.676
Cysteine/ine[a]	0.335	0.384	0.123	0.040	0.055	0.052	0.038	0.067	0.091	0.190	0.290	0.367
Glutamine		1.231		0.275		0.301		0.406		0.603		0.559
Glutamate	3.894	2.658	0.673	0.365	0.524	0.163	0.438	0.052	1.133	0.664	1.633	1.014
Glycine	1.580	1.396	0.050	0.063	0.122	0.114	0.085	0.061	0.256	0.220	0.420	0.387
Histidine	0.828	0.822	0.075	0.088	0.063	0.067	0.067	0.091	0.191	0.185	0.296	0.306
Isoleucine	1.180	1.093	0.150	0.192	0.116	0.118	0.137	0.088	0.324	0.345	0.682	0.623
Leucine	2.063	2.164	0.327	0.383	0.222	0.220	0.251	0.366	0.498	0.548	1.067	0.916
Lysine	2.192	2.160	0.252	0.211	0.097	0.079	0.153	0.037	0.431	0.432	0.897	0.904
Methionine	0.676	0.641	0.062	0.088	0.063	0.042	0.068	0.054	0.084	0.158	0.390	0.382
Phenylalanine	1.025	1.051	0.145	0.160	0.144	0.160	0.131	0.104	0.319	0.346	0.664	0.686
Proline	1.237	1.697	0.343	0.298	0.127	0.117	0.181	0.279	0.353	0.292	0.498	0.451
Serine	1.022	1.406	0.168	0.197	0.141	0.175	0.156	0.126	0.309	0.377	0.929	1.296
Threonine	1.036	1.166	0.082	0.142	0.096	0.090	0.115	0.103	0.268	0.235	0.600	0.535
Tryptophan	0.170	0.352	0.040	0.046	0.031	0.032	0.031	0.017	0.102	0.065	0.152	0.203
Tyrosine	0.827	0.937	0.148	0.164	0.09	0.136	0.106	0.093	0.219	0.175	0.510	0.495
Valine	1.287	1.339	0.180	0.218	0.164	0.177	0.167	0.139	0.331	0.259	0.761	0.781

[a]Cysteine is estimated from the gene sequencing method (Study) and cystine from the conventional method (USDA).

Table 3

Characteristics of the top 30 commonly consumed food proteins among 70 356 study participants including the correlation coefficients for the composition of amino acids (AA in g/100 g food) of each food between methods

	Contribution to total protein intake (%)	Protein in 100 g food (g)	Gene sequencing method (% protein)		Conventional method (% protein)		Pearson[a] ρ
			20 AA[b]	16 AA[c]	18 AA[b]	16 AA[c]	16 AAs
Beef	15.0	25.9	100	76.3	96.4	72.2	0.95
Skim milk	7.0	3.4	100	74.6	95.1	68.2	0.91
Chicken, no skin	6.4	28.9	100	76.3	94.1	70.2	0.92
Meat for sandwiches	5.5	25.9	100	76.3	96.4	72.2	0.95
Tuna	5.1	23.6	100	76.3	96.5	71.3	0.92
Chicken and skin	4.8	27.3	100	76.3	95.0	71.5	0.93
Hamburger	4.5	25.8	100	76.3	96.6	72.8	0.95
American cheese	4.2	24.9	100	74.6	108.7	77.8	0.93
White fish	3.5	22.8	100	76.3	96.6	71.4	0.92
Eggs	2.9	12.6	100	78.0	99.1	76.2	0.92
Pizza	2.7	12.2	100	69.2	99.8	68.8	0.91
Cottage cheese	2.5	12.5	100	74.7	111	83.0	0.95
Whole grain bread	2.2	9.1	100	62.0	101	63.6	0.97
White bread	1.9	7.6	100	62.0	99.7	62.4	0.98
Milk	1.9	3.2	100	74.7	94.8	67.1	0.89
Mashed potato	1.8	1.9	100	85.5	100.1	61.8	0.91
Cold cereals	1.3	8.0	100	67.7	99.3	67.4	0.95
Yogurt	1.3	4.4	100	74.6	108.8	81.5	0.94
Nuts	1.2	28.0	100	74.6	95.0	63.9	0.86
Shrimp	1.1	20.9	100	76.3	100.2	72.8	0.86
English muffin	1.1	7.7	100	62.0	100.5	63.9	0.97
Peanut butter	1.0	25.0	100	74.6	101.7	68.4	0.87
Dark fish[d]	1.0	27.3	100	76.3	96.5	71.3	0.92
Liver[e]	1.0	26.5	100	76.3	97.6	75.8	0.93
Pasta	1.0	4.8	100	62.0	99.6	59.8	0.98
Beans	1.0	7.5	100	79.7	95.1	67.7	0.75

	Contribution to total protein intake (%)	Protein in 100 g food (g)	Gene sequencing method (% protein)		Conventional method (% protein)		Pearson[a] ρ
			20 AA[b]	16 AA[c]	18 AA[b]	16 AA[c]	16 AAs
Peas	0.7	5.2	100	78.8	78.4	55.8	0.66
Processed meats[f]	0.7	13.8	100	76.3	91.0	68.0	0.95
Orange juice	0.7	0.7	100	69.1	42.9	28	0.52
Clam chowder	0.7	3.8	100	74.7	99.7	65.4	0.91

[a] Correlation coefficients for 16 proteinogenic AA (g/100 g protein) by analytic method (USDA vs sequencing) excluding asparagine, aspartate, glutamine, and glutamate for the top 30 commonly consumed food proteins.

[b] % Protein is the sum of AA (g)/100 g of protein using the sequencing method (20 AA) and USDA (database has 18 AA, no asparagine or glutamine).

[c] % Protein is the sum of AA (g)/100 g of protein using either method (16 AA) excluding asparagine, aspartate, glutamine, and glutamate.

[d] Dark fish: dark meat fish such as mackerel, salmon, sardines, sword fish, and blue fish.

[e] Liver: from beef, calf, or pork.

[f] Processed meats such as sausage and kebbasa.

Table 4

Mean and Pearson correlation coefficients for the AA intake of 70 356 participants comparing gene sequencing and conventional methods

Amino acids (AA)	Mean (s.d.) AA in g/ day		ρ Pearson	Mean (s.d.) AA in g/100 g protein		ρ Pearson	Mean (s.d.) AA in g/1000 kcal		ρ Pearson
	Sequencing	Conventional		Sequencing	Conventional		Sequencing	Conventional	
Alanine	4.04 (1.33)	3.71 (1.22)	0.99	5.28 (0.37)	4.85 (0.28)	0.73	2.35 (0.54)	2.16 (0.47)	0.97
Arginine	4.34 (1.42)	4.35 (1.43)	0.99	5.69 (0.49)	5.70 (0.44)	0.90	2.53 (0.58)	2.53 (0.55)	0.99
Asparagine	2.75 (0.87)	—	—	3.60 (0.21)	—	—	1.60 (0.33)	—	—
Aspartate[a]	3.62 (1.22)	7.11 (2.28)	0.98	4.75 (0.47)	9.32 (0.56)	0.75	2.12 (0.53)	4.14 (0.87)	0.97
Cys or Cys–Cys[b]	1.32 (0.42)	0.95 (0.30)	0.98	1.74 (0.12)	1.24 (0.08)	0.50	0.77 (0.15)	0.55 (0.10)	0.94
Glutamine	6.85 (2.19)	—	—	9.04 (1.31)	—	—	3.95 (0.63)	—	—
Glutamate[a]	7.27 (2.44)	14.46 (4.55)	0.96	9.48 (0.67)	18.98 (1.04)	−0.20	4.24 (1.01)	8.39 (1.46)	0.94
Glycine	3.23 (1.10)	3.26 (1.08)	0.98	4.23 (0.45)	4.27 (0.36)	0.82	1.89 (0.47)	1.90 (0.42)	0.97
Histidine	2.29 (0.75)	2.16 (0.70)	0.98	2.99 (0.17)	2.82 (0.11)	0.27	1.33 (0.29)	1.25 (0.25)	0.96
Isoleucine	3.76 (1.20)	3.54 (1.16)	0.99	4.92 (0.21)	4.62 (0.15)	0.49	2.18 (0.42)	2.05 (0.42)	0.98
Leucine	7.01 (2.27)	5.89 (1.93)	0.99	9.17 (0.44)	7.69 (0.21)	0.31	4.08 (0.85)	3.42 (0.67)	0.97
Lysine	5.35 (1.82)	5.40 (1.82)	0.99	6.98 (0.60)	7.03 (0.37)	0.80	3.12 (0.78)	3.14 (0.72)	0.97
Methionine	1.97 (0.63)	1.80 (0.60)	0.98	2.57 (0.12)	2.34 (0.12)	0.19	1.14 (0.23)	1.05 (0.23)	0.96
Phenylalanine	3.83 (1.19)	3.24 (1.04)	0.98	5.04 (0.36)	4.24 (0.12)	0.09	2.23 (0.43)	1.88 (0.34)	0.92
Proline	5.59 (1.80)	4.58 (1.52)	0.96	7.31 (0.39)	5.99 (0.57)	0.48	3.25 (0.64)	2.64 (0.47)	0.89
Serine	4.53 (1.44)	3.31 (1.07)	0.98	5.93 (0.26)	4.33 (0.18)	−0.08	2.64 (0.53)	1.92 (0.34)	0.94
Threonine	3.45 (1.11)	3.02 (1.00)	0.99	4.52 (0.24)	3.94 (0.12)	0.31	2.01 (0.41)	1.75 (0.36)	0.97
Tryptophan	1.01 (0.33)	0.88 (0.29)	0.94	1.32 (0.07)	1.14 (0.09)	−0.08	0.59 (0.13)	0.51 (0.11)	0.88
Tyrosine	3.10 (1.01)	2.68 (0.88)	0.99	4.06 (0.20)	3.50 (0.14)	0.45	1.80 (0.36)	1.55 (0.30)	0.97
Valine	4.25 (1.37)	3.90 (1.27)	0.99	5.56 (0.24)	5.10 (0.16)	0.38	2.47 (0.51)	2.26 (0.44)	0.97

[a] Asparagine and glutamine are mostly transformed to aspartate and glutamate during acid hydrolysis.

[b] Cysteine (Cys) is measured by the sequencing method and cystine (Cys–Cys) is measured by the conventional method—cystine gives rise to two equivalents of cysteine or its derivative.