Protein & Amino Acid Hub:
Data Dictionary

Version 6/13/24

Welcome to the Data Dictionary for the Protein & Amino Acid Hub, within this document you will find descriptions of all variables that appear on the Protein & Amino Acid Hub as well as any additional variables needed for protein quality score calculations or data documentation.

There are three types of variables documented here:

1. Variables that appear in both the github data and on the Protein & Amino Acid Hub
2. Variables that are generated by (and must be downloaded from) the Protein & Amino Acid Hub
3. Variables that appear in the github that are needed to calculate protein quality scores generated by the Protein & Amino Acid Hub but are not directly provided in the app

Documentation for all types of variables are provided here, organized by file and usage.

App Table/Tab name: **Protein Digestibility Data**
Filename (github): Protein Digestibility Data - full data.csv

This file contains information collected from protein digestibility data sources. Descriptions of variables within this file are provided below.

Variable Name	Description
NI_ID	Unique Nutrient Institute (NI) identifier for each digestibility data point.
Food group	Food group as specified by the data source from which the digestibility data was collected
Food	Description of the food used in digestibility analysis
Protein (g)	Amount of protein (in grams) from the applicable food ingested (in the case of in vivo analysis) or analyzed (in the case of in vitro analysis)
Diet	Description of the diet consumed by experimental subjects
Species	Target species of digestibility analysis
n	Number of subjects from which in vivo digestibility was collected (if applicable)
Model	Experimental model (either in vivo or in vitro)
Sample	The type or location of sample collected for analysis (e.g. ileal, fecal, etc...)
Measure	The name of the measure of digestibility or metabolic activity (i.e. apparent digestibility, biological value, metabolic availability, etc...)
Analyte	The protein or amino acid for which digestibility coefficient is provided
Value (%)	Value of the associated measure, expressed as a percentage
SD	Standard deviation of the provided value
Analysis method(s)	Name of the analysis method(s), technique(s), or assay(s) used to measure digestibility, as specified in the source the data was collected from
Collected From	A citation indicating where the data appearing in this table was collected from - citations created using CDFC Citation Generator
Original Source(s)	A citation or list of ordered citations indicating the original source(s)

	of the digestibility data (as cited in the source data was collected from).
Notes	Any additional notes or comments applicable to the collected data that are provided in the source data has been collected from.

App Table/Tab name: **AA Composition Data**
Filename (github): EAA_composition.csv

This file contains information collected from food composition data sources. Descriptions of variables within this file are provided below.

Variable Name(s)	Description
fdcId; FDC_ID	FoodData Central (FDC) identifier, used to map protein digestibility data to food composition data from FoodData Central
NI_ID	Unique Nutrient Institute (NI) identifier for each digestibility data point.
Protein (g/100g)	Grams of protein per 100g of food
His (g/100g)	Grams of histidine per 100g of food
Ile (g/100g)	Grams of isoleucine per 100g of food
Leu (g/100g)	Grams of leucine per 100g of food
Lys (g/100g)	Grams of lysine per 100g of food
Met+Cys (g/100g)	Grams of methionine and cystine per 100g of food
Phe+Tyr (g/100g)	Grams of phenylalanine and tyrosine per 100g of food
Thr (g/100g)	Grams of threonine per 100g of food
Trp (g/100g)	Grams of tryptophan per 100g of food
Val (g/100g)	Grams of valine per 100g of food
Ref No	Number of associated food composition data reference
Food Composition Data Citation	A citation indicating where food composition data was collected - citations created using CDFC Citation Generator

App Table/Tab name: **Protein Quality Scoring**

Unlike the other 2 tabs, the Protein Quality Scoring tab takes in two github files (described below) as inputs, then outputs the following variables:

Variable Name(s)	Description
NI_ID	Unique Nutrient Institute (NI) identifier for each digestibility data point.
Food	Description of the food, as provided by the source of digestibility data
Digestibility Measure	The name of the measure of digestibility or metabolic activity (i.e. apparent digestibility, biological value, metabolic availability, etc...) Defined as “Measure” in the Protein Digestibility Data table
Digestibility Species	Target species of digestibility analysis Defined as “Species” in the Protein Digestibility Data table
Digestibility Analyte	The protein or amino acid for which digestibility coefficient is provided Defined as “Analyte” in the Protein Digestibility Data table
Digestibility Value (%)	Value of the associated measure, expressed as a percentage Defined as “Value (%)” in the Protein Digestibility Data table
Limiting AA	The limiting essential amino acid determined by the amino acid scoring pattern or recommendations
Composition Ref No	Number of associated food composition data reference - full references can be found in the ‘Information’ section above the AA Composition Data Table and in EAA_composition.csv Defined as “Ref No” in the AA Composition Data table
serving size *(if applicable)*	Serving size of food used to calculate EAA-9 score
EAA-9 (%) *(if applicable)*	EAA-9 score calculated as documented in github README.md and ‘information’ section above the Protein Quality Scoring table
PDCAAS *(if applicable)*	PDCAAS calculated as documented in github README.md and ‘information’ section above the Protein Quality Scoring table

DIAAS *(if applicable)*	DIAAS calculated as documented in github README.md and 'information' section above the Protein Quality Scoring table

These protein quality scores cannot be exported from the github and must be downloaded directly from the Protein & Amino Acid Hub.

The following two files are used to calculate protein quality scores.

Filename (github): scoring_pattern.csv

This file contains different amino acid recommendations and scoring patterns for use in protein quality scoring. Descriptions of variables within this file are provided below.

Variable Name(s)	Description
Pattern Name	The name of the reference pattern of amino acid recommendations
Analyte	The protein or amino acid for which the pattern or recommendation is provided
Amount	The amount of the analyte recommended by the pattern or recommendation
Unit	Unit of the recommended amount
Age	The age group the recommendation or pattern is intended for
Reference	Reference to the data source where pattern/recommendation was collected

Filename (github): portion_sizes.csv

This file contains standard portion sizes of FoodData Central foods used to calculate EAA-9 scores. Descriptions of variables within this file are provided below.

Variable Name(s)	Description
fdcId	FoodData Central identifier
g_weight	Weight of the food portion in grams
portion	Portion size described in household measurements