

Protein Quality Hub: Data Dictionary

Version 12/05/2024

Welcome to the Data Dictionary for the Protein Quality Hub, within this document you will find descriptions of all variables that appear in the Protein Quality 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 Quality Hub
2. Variables that are generated by (and must be downloaded from) the Protein Quality Hub
3. Variables that appear in the github that are needed to calculate protein quality scores generated by the Protein Quality Hub or provide additional context about the data, 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: **Bioavailability Data**

Filename (github): [Bioavailability - full data.csv](#)

This file contains information collected regarding the bioavailability of proteins. Descriptions of variables within this file are provided below.

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

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
fdclid; FDC_ID	FoodData Central (FDC) identifier, used to map protein bioavailability data to food composition data from FoodData Central
NI_ID	Unique Nutrient Institute (NI) identifier for each bioavailability data point
description; Food description	Description of the food provided by the food composition data source
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
Food Composition Data Ref	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 bioavailability data point.
Food	Description of the food, as provided by the source of bioavailability data
Bio-avail Species	Target species of bioavailability analysis Defined as “Species” in the <i>Bioavailability Data</i> table
Bio-avail Sample Location	The type or location of sample collected for analysis (e.g. ileal, fecal, etc...) Defined as “Sample Location” in the <i>Bioavailability Data</i> table
Bio-avail Protein Form	The protein or amino acid for which bioavailability coefficient is provided Defined as “Protein Form” in the <i>Bioavailability Data</i> table
Bio-avail Calculation	The name of the measure of bioavailability or metabolic activity (i.e. apparent digestibility, metabolic availability, etc...) Defined as “Calculation” in the <i>Bioavailability Data</i> table
Bio-avail (%)	Value of the associated measure, expressed as a percentage Defined as “Value (%)” in the <i>Bioavailability Data</i> table
Limiting AA	The limiting essential amino acid determined by the amino acid scoring pattern or recommendations
fdcid (<i>if applicable</i>)	FoodData Central (FDC) identifier, used to map protein bioavailability data to food composition data from FoodData Central
serving size (<i>if applicable</i>)	Serving size of food used to calculate EAA-9 score
EAA-9 (%) (<i>if applicable</i>)	EAA-9 score calculated as documented in github README.md and ‘information’ section above the Protein Quality Scoring table
PDCAAS (<i>if applicable</i>)	PDCAAS calculated as documented in github README.md

	and 'information' section above the Protein Quality Scoring table
DIAAS <i>(if applicable)</i>	DIAAS calculated as documented in github README.md and 'information' section above the <i>Protein Quality Scoring</i> table
Food 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

These protein quality scores cannot be exported from the github and must be downloaded directly from the Protein Quality 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
fdclid	FoodData Central identifier
g_weight	Weight of the food portion in grams
portion	Portion size described in household measurements