Nitrogen and protein content measurement and nitrogen to protein conversion factors for dairy and soy protein-based foods: a systematic review and modelling analysis
NITROGEN AND PROTEIN CONTENT MEASUREMENT AND NITROGEN TO PROTEIN CONVERSION FACTORS FOR DAIRY AND SOY PROTEIN-BASED FOODS: A SYSTEMATIC REVIEW AND MODELLING ANALYSIS

Prepared by:
Daniel Tomé¹
Christophe Cordella¹
Omar Dib¹
Christine Péron²

¹ UMR PNCA, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France
² Direction des Documentations, du Musée du Vivant et du CIRE (DDMC) (DG&S), AgroParisTech, Université Paris-Saclay, 75005 Paris, France
CONTENTS

Acknowledgements vii
Abbreviations viii
1 Introduction 1
2 Background 2
  2.1 Protein, protein nitrogen and non-protein nitrogen in foods 2
  2.2 General principles of NPCFs 3
  2.3 Approaches for calculation of NPCFs 4
    2.3.1 Crude protein weight versus amino acid composition 4
    2.3.2 Anhydrous weight of amino acids 5
    2.3.3 Contribution of amide amino acids to protein nitrogen content 5
3 Methods 7
  3.1 Systematic review of the literature 7
    3.1.1 PICO questions 7
    3.1.2 Literature review 8
    3.1.3 Data analysis 8
  3.2 Calculation of NPCFs 9
  3.3 Modelling of non-protein nitrogen 10
  3.4 Assessment of risk of bias and certainty of evidence 11
4 Results 12
  4.1 Literature review 12
  4.2 Analytical methods for nitrogen, amino acid and protein content in foods 14
    4.2.1 Methods to determine total nitrogen content in foods 14
    4.2.2 Methods for amino acid analysis 17
    4.2.3 Other methods for protein analysis 23
  4.3 NPCFs for milk-based and soy-based foods 26
    4.3.1 Conversion factors for dairy-based foods 26
    4.3.2 Conversion factors for soy-based foods 29
    4.3.3 Pooled estimates for NPCFs 30
    4.3.4 Certainty of the evidence 32
  4.4 Modelling of non-protein nitrogen 32
5 Discussion 35
  5.1 General considerations in calculating and NCPFs 35
    5.1.1 Prosthetic groups 35
    5.1.2 Non-protein nitrogen 36
    5.1.3 Amide to acid ratio 37
5.2 Selecting NPCFs 37
5.3 Comparability with other reviews 38
5.4 Comparability with other conversion factors 38
6 Conclusion 39

Annexes 41
Annex 1 Search strategy 43
Annex 2 Excluded studies 45
Annex 3 GRADE evidence profiles 47
GRADE evidence profile 1 – Dairy-based ingredients 47
GRADE evidence profile 2 – Soy-based ingredients 50
GRADE evidence profile 3 – Dairy-based ingredients 53
GRADE evidence profile 4 – Soy-based ingredients 55
Annex 4 Studies included in the GRADE assessment: dairy 57
Annex 5 Studies included in the GRADE assessment: soy 59

Tables and figures
Table 1 The fractions of protein and nitrogen content in foods 4
Table 2 Values used to calculate NPCFs 6
Table 3 Total nitrogen and protein nitrogen content values for soy-based products and dairy-based products 11
Table 4 Methods for food protein analysis 14
Table 5 Kjeldahl methods and modifications for total protein measurement in food, by matrix 15
Table 6 Standards for total nitrogen and protein measurement in food and feed using the Dumas method 17
Table 7 Problematic amino acids during the protein hydrolysis procedure 19
Table 8 Main derivatization methods based on specific amine or amide reagents for amino acids analysis 20
Table 9 A selection of studies specifically related to amino acid analysis in milk and soybean products by gas chromatography and liquid chromatography 21
Table 10 Comparison of routine methods of food protein analysis 22
Table 11 Main interfering compounds for BCA protein assay 23
Table 12 Main compounds interfering with the Bradford reagent (Coomassie Brilliant Blue G250) or with proteins in samples, and incompatible sample preparation methods 24
Table 13 Linearity range and limit of detection of most used colorimetric assays for protein determination 24
Table 14 Protein concentrations determined by most used colorimetric assays for standard protein samples 25
Table 15 NPCFs reported for dairy-based foods and food ingredients 26
ACKNOWLEDGEMENTS

This systematic review was undertaken as part of the work of the Joint FAO/WHO Expert Meetings on Nutrition (JEMNU) requested by the 39th Session of the Codex Committee on Nutrition and Food for Special Dietary Uses (CCNFSDU) held in December 2017 to provide scientific advice on the nitrogen to protein conversion factor(s) for soy-based and milk-based ingredients in infant formula and follow-up formula. FAO and WHO are grateful for the financial support provided by the US Food and Drug Administration for conducting this systematic review and for implementing the work of JEMNU requested by CCNFSDU.

Technical editing of the document was undertaken by Dr Hilary Cadman from Cadman Editing Services, Australia, and cover design and formatting of the document was undertaken by Ms Sue Hobbs from minimum graphics, New Zealand.
ABBREVIATIONS

AA    amino acid
AACC  American Association of Cereal Chemists
AOCS  American Oil Chemists’ Society
BCA   bicinchoninic acid
CCNFSDU Codex Committee on Nutrition and Foods for Special Dietary Uses
CI    confidence interval
CV    coefficient of variation
FAO   Food and Agriculture Organization of the United Nations
FSTA  Food Science and Technology Abstracts
GRADE Grading of Recommendation, Assessment, Development and Evaluation
HCl   hydrochloric acid
HPLC  high-performance liquid chromatography
IDF   International Dairy Federation
ISO   International Organization for Standardization
JEMNU Joint FAO/WHO Expert Meetings on Nutrition
MIR   mid-infrared reflectance
NH3   ammonia
NIR   near infrared reflectance
NPCF  nitrogen to protein conversion factor
NPN   non-protein nitrogen
PICO  population, intervention, comparator, outcome
PN    protein nitrogen
RP-HPLC reverse-phase high-performance liquid chromatography
SD    standard deviation
TN    total nitrogen
UHPLC ultra-high-performance liquid chromatography
UNSCN United Nations System Standing Committee on Nutrition
UV    ultraviolet
WHO   World Health Organization

Amino acids

ala alanine leu leucine
arg arginine lys lysine
asn asparagine met methionine
asp aspartate phe phenylalanine
cys cysteine pro proline
gln glutamine ser serine
glu glutamate thr threonine
gly glycine trp tryptophan
his histidine tyr tyrosine
ile isoleucine val valine
1. INTRODUCTION

The World Health Organization (WHO) recommends that infants be exclusively breastfed for the first 6 months of life, and thereafter receive nutritionally adequate and safe complementary foods, while continuing to breastfeed for up to 2 years or beyond (WHO, 2013). There are circumstances, however, in which the use of breast-milk substitutes (i.e. infant formula) may be acceptable (FAO/WHO, 2016a). In addition, “follow-up formula” is sold and consumed by older infants and young children. Dairy-based ingredients and soy-based ingredients are used as a source of protein in such formula. The Codex Alimentarius Commission has established standards for infant formula (FAO/WHO, 1981) and follow-up formula (FAO/WHO, 1987), including a range of acceptable protein levels (i.e. a minimum and maximum amount of protein per 100 kcal). The method for calculating protein content of food products currently described in the standards is to measure nitrogen content and then convert to protein using a conversion factor of 6.25 for both dairy-based and soy-based ingredients, unless a scientific justification is provided for the use of a different conversion factor for a particular product.

Nitrogen to protein conversion factors (NPCFs) allow for the estimation of protein content in food samples from the amount of nitrogen in the sample, based on two assumptions: that most of the nitrogen is associated with amino acids and that most of the amino acids in foods are associated with protein. The accuracy of the estimation depends on the value of the conversion factor. A value of 6.25 is applied for measuring protein content in most foods and food ingredients, again based on two assumptions: that proteins contain about 16% nitrogen by weight (i.e. of the total mass of a protein, 16% is nitrogen), and that all nitrogen in food is derived from protein. However, using the same conversion factor for all protein sources can introduce errors that lead to significant overestimation or underestimation of the actual protein content of most foods. Hence, a default value of 6.25 may not be an appropriate conversion factor for all protein sources; instead, specific values should be considered for different foods and food ingredients.

Across Codex standards, there is no single universally agreed NPCF for dairy and soy. The Codex Committee on Nutrition and Foods for Special Dietary Uses (CCNFSDU) has recently raised the question of the appropriate NPCFs to be used for milk and soy protein in infant formula and follow-up formula. To this end, CCNFSDU sought advice from the Joint Food and Agriculture Organization of the United Nations (FAO)/WHO Expert Meetings on Nutrition (JEMNU) – and requested that JEMNU be convened to review the evidence and develop evidence-informed guidance regarding NPCFs.

This review was commissioned by FAO and WHO to serve as a background document for JEMNU. The objective was to assess known conversion factors for dairy-based and soy-based ingredients through a systematic review of the literature, including methods and approaches used for the determination of NPCFs, with the aim of supporting the selection of appropriate conversion factors for dairy-based and soy-based ingredients used in infant formula and follow-up formula.

---

1. JEMNU was established in 2009 to provide scientific advice to the committees of the Joint FAO/WHO Food Standards Programme (i.e. Codex Alimentarius Commission) or Member Countries, in an independent, timely and cost-effective manner.

2. In the context of this document, “appropriate” refers solely to “most likely to provide the best estimate of the actual protein content”.

---
2. BACKGROUND

2.1 Protein, protein nitrogen and non-protein nitrogen in foods

In foods, nitrogen is mainly associated with protein as protein nitrogen, but it can also be present in other non-protein food components as non-protein nitrogen (Fig. 1).

Fig. 1 Protein and nitrogen in foods

<table>
<thead>
<tr>
<th>Protein = sum of weight of amino acid residues and of prosthetic groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosthetic groups</td>
</tr>
<tr>
<td>Protein-associated prosthetic groups</td>
</tr>
<tr>
<td>Protein = sum of weight of amino acid residues</td>
</tr>
<tr>
<td>Protein amino acid sequence</td>
</tr>
<tr>
<td>Protein nitrogen</td>
</tr>
<tr>
<td>Non-protein nitrogen</td>
</tr>
<tr>
<td>Non-protein nitrogen</td>
</tr>
<tr>
<td>Total nitrogen</td>
</tr>
</tbody>
</table>

Proteins and peptides are macromolecules that comprise covalently bonded amino acid residues organized as linear polymers. The sequence of the amino acids in a protein or peptide determines the properties of the molecule. Proteins are large molecules that commonly exist as folded structures with a specific conformation, whereas peptides are smaller and may comprise only a few amino acids.

Proteins are defined biochemically by their amino acid residue content and by the sequence of those amino acids, which forms the primary structure:

- Up to 20 different amino acid residues are included in the composition and sequence of a protein; proteins in food or food ingredients differ in terms of the amount of each amino acid residue that they contain.

- Proteins differ in their amino acid residue composition, and amino acid residues differ in molecular weight. The weight of a particular amino acid residue in the protein fraction of food is derived from the molecular weight of that residue and the frequency with which it occurs.

- Many proteins are modified post-translationally (e.g. through phosphorylation or glycosylation); the prosthetic groups added to the constituent amino acids in such modifications can change the characteristics of the protein and increase its mass. Prosthetic groups vary in weight, from 14 Da for methylation to over 200 Da for fatty acylation, prenylation, nucleotide conjugation, or N-linked and O-linked glycosylation; thus, certain modifications can significantly increase the weight of the protein fraction.

- Protein content on a weight basis in food or food ingredients can be defined in two ways: as the sum of the constituent amino acids, or as the sum of the constituent amino acids plus the weight of the prosthetic groups (for proteins that contain prosthetic groups).

Nitrogen is found in a variety of foods and food ingredients. Total nitrogen in foods comprises both protein nitrogen and non-protein nitrogen:

- The fraction of protein nitrogen is primarily defined by the sum of the nitrogen associated...
with the amino acids that make up a protein. Proteins differ in their amino acid composition, and amino acids differ in their nitrogen content; hence, proteins differ in their total nitrogen content. In addition, a few of the prosthetic groups associated with proteins are nitrogenous (e.g. covalently linked coenzymes and amino sugars).

- Amino acids are not the only nitrogenous compounds in foods and food ingredients; other nitrogenous compounds include alkaloids, amines, ammonia, phospholipids and nitrogenous glycosides, nitrates, nitrates, nucleic acids, urea and vitamins. This non-protein nitrogen fraction varies greatly among food protein sources and in a given protein source, depending on the production process and the degree of purification of the protein source. For instance, non-protein nitrogen represents about 5% of the total nitrogen in cow’s milk (Journet, Vérité & Vignon, 1975), between 7.1% and 9.3% in goat’s milk (Grappin et al., 1981), and between 2.9% and 7.8% in defatted soybean flour (before heat treatment), in the form of free amino acids, nucleotides, creatine and choline (Maubois & Lorient, 2016).

- When describing a protein, free amino acids and small peptides are sometimes included and sometimes omitted; if they are included, they may be counted as part of either the protein nitrogen or the non-protein nitrogen.

The nitrogen content of individual amino acids varies according to the molecular weight of the amino acid and the number of nitrogen atoms it contains (from one to four, depending on the amino acid in question) (Sosulski & Imafidon, 1990). Thus, there are important differences in the nitrogen content of individual amino acids, distributed according to the amount of nitrogen in their residues:

- nitrogen-poor residues contain from 8.58% to 10.85% nitrogen: tyrosine, phenylalanine, methionine and glutamic acid;
- nitrogen-rich residues contain from 21.86% to 35.87% nitrogen: lysine, glutamine, asparagine, glycine, histidine and arginine; and
- the remaining 10 residues have a nitrogen content that ranges from 12.17% (aspartic acid) to 19.7% (alanine).

The variations in individual amino acid composition between proteins and in the nitrogen content of individual amino acids mean that the “generic” nitrogen content of protein is not 16%, but in reality can vary from about 13% to 19%. The fact that proteins from diverse sources vary in their nitrogen content owing to differences in their amino acid composition was first demonstrated through protein purification and nitrogen content analysis of extracted purified proteins (Jones, 1931).

2.2 General principles of NPCFs

For many years, the protein content of foods has been determined from total nitrogen content. The protein content in a food is traditionally estimated by multiplying the total nitrogen content by a NPCF, based on the assumptions that dietary carbohydrates and fats do not contain nitrogen, and that most of the nitrogen in the diet is present as amino acids in proteins. However, protein and nitrogen content are defined differently (Jones, 1931; Tkachuk, 1969; Sosulski & Imafidon, 1990; Mossé, 1990; Maubois & Lorient, 2016; Krul, 2019) (Table 1).

The conversion factor was historically set at 6.25 by assuming that most, if not all, nitrogen in food was derived from protein, and that the nitrogen content of proteins was about 16%. This
The approach is still an accepted method for calculating the crude protein content of foods and food ingredients; however, it has been known for decades that using total nitrogen content with a conversion factor of 6.25 to quantify total protein is imperfect and can lead to a 15–20% error in the actual protein content.

The errors for deriving the crude protein content from nitrogen content have three main origins, as outlined below:

- Total nitrogen content varies between the different amino acids; thus, because amino acid composition varies from one protein source to another, the protein nitrogen content also varies between different protein sources.
- In addition to amino acids, some proteins contain prosthetic groups, some of which are non-nitrogenous and therefore increase the molecular weight of the protein without significantly affecting the nitrogen content. In turn, this results in an increase in the conversion factor, because both amino acids and amino acids plus prosthetic groups are being considered on a weight basis.
- The fraction of non-protein nitrogen varies in different food protein sources; this can result in a lower value of the conversion factor when the total nitrogen content of foods is being considered.

The principle of NPCF calculation has been discussed in reviews and original articles (e.g. Jones, 1931; Heidelberg et al., 1975; Krul, 2019; Mariotti, Tome & Mirand, 2008; Maubois & Lorient, 2016; Mossé, 1990; Sosulski & Imafidon, 1990; Tkachuk, 1969). Because both the protein and the nitrogen component can be determined by different approaches, the resulting values of the conversion factor can vary. In the literature, different terminologies have been used for the conversion factors, depending on the method by which they have been calculated (e.g. $K_a$, $K^\prime$ and $K_p$).

### 2.3 Approaches for calculation of NPCFs

#### 2.3.1 Crude protein weight versus amino acid composition

Initial approaches to calculating NPCFs involved extracting the crude protein fractions from the foods or food ingredients, determining the crude protein mass, measuring the nitrogen content in the isolated crude protein fraction, and then calculating a conversion factor from the nitrogen to protein ratio. This approach was used first for milk protein and later for additional protein sources (Hammersten, 1883; Jones, 1931). Although it provided a usable conversion factor, this early approach was not specific for the amino acids within a protein, because the crude protein mass included prosthetic groups when these were present in the protein; also, it did not explicitly account for any nitrogen from non-protein sources.

Newer approaches use values for protein weight and protein nitrogen content of different proteins based on knowledge of the amino acid composition as obtained via amino acid analysis or amino acid sequencing (Heathcote, 1950; Tkachuk, 1969; Holt & Sosulski, 1979;

### Table 1. The fractions of protein and nitrogen content in foods

<table>
<thead>
<tr>
<th>PROTEIN CONTENT FRACTIONS</th>
<th>NITROGEN CONTENT FRACTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of weight of amino acid residues</td>
<td>Sum of weight of protein nitrogen from amino acid residues</td>
</tr>
<tr>
<td>Sum of weight of prosthetic groups</td>
<td>Sum of weight of non-protein nitrogen</td>
</tr>
</tbody>
</table>
Such approaches have made it possible to derive more precise conversion factors, as discussed in Section 3.2. Calculation of the weight of the protein can include either the sum of the amino acid residues only, or the sum of the amino acid residues plus the weight of the associated prosthetic groups when these are analysed by specific methods or obtained from amino acid sequencing.

2.3.2 Anhydrous weight of amino acids

An important feature of approaches based on amino acid composition is the use of the anhydrous weight of amino acids to determine protein weight. Using the free amino acid weight for a protein would grossly overestimate the weight for longer polypeptide chains. Because each amino acid residue in a polypeptide chain loses one water molecule (weighing 18 Da) during polymerization, except for the residue at the carboxy terminus; thus, the weight of the protein must be assessed as the sum of the weight of the anhydrous amino acid residues rather than the weight of the free amino acids (each of which weighs 18 Da more than its anhydrous counterpart). Molecular weights of free and anhydrous forms of all 20 amino acids, and the corresponding percentage of nitrogen, are shown in Table 2.

2.3.3 Contribution of amide amino acids to protein nitrogen content

Another important consideration when calculating NPCFs relates to the differential contribution to nitrogen content of the amide forms of the amino acids, glutamine and asparagine, and their respective acid forms, glutamic acid and aspartic acid.

With standard amino acid analysis of a protein sample, glutamine and asparagine are converted to glutamic acid and aspartic acid during acid hydrolysis (by substitution of a carboxyl group for the amide group); the total of glutamic acid plus glutamine and aspartic acid plus asparagine is then determined. This conversion does not significantly affect the calculation of the total anhydrous amino acid weight, because there is little difference between the molecular weights of glutamic acid and glutamine (147.13 Da and 146.15 Da, respectively) and aspartic acid and asparagine (133.11 Da and 132.12 Da, respectively). However, failing to consider the ratio between the amide and acid forms induces a major error in the calculation of total nitrogen content of the amino acid residues.

The nitrogen content of the amide amino acids can be determined using specific analytical conditions that help to preserve the amide nitrogen, although the different conditions have drawbacks in terms of accuracy in estimating the true nitrogen content. For example, using mild acid hydrolysis conditions (e.g. 2 M hydrochloric acid [HCl], 3 hours, 115 °C) and measuring the released ammonia (NH3) allows for the quantification of amide nitrogen, but it does not differentiate between nitrogen from asparagine and from glutamine (Mossé, 1990). Alternatively, the total NH3 released from asparagine and glutamine during acid hydrolysis with 6 N HCl can be used as a proxy for amide nitrogen content (Heathcote, 1950; Tkachuk, 1969), recognizing that this value is equal or slightly higher than the true amount of amide NH3, leading to an overestimation of nitrogen from the amide form of amino acids. This latter approach includes nitrogen released from amide asparagine and glutamine, but also from other amino acids (serine and threonine) and from other non-protein sources of nitrogen. Yet another approach is to consider a fixed ratio between acid and amide forms that is specific for each protein source (often 75/25 or 50/50), as found in the literature. Specific derivatives can also be used to determine asparagine and glutamine directly.
### Table 2. Values used to calculate NPCFs

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Molecular weight (Da)*</th>
<th>AA residue/free AA (%)</th>
<th>N in free AA (%)</th>
<th>N in AA residue (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartic acid</td>
<td>133.1</td>
<td>115.1</td>
<td>86.5</td>
<td>10.5</td>
</tr>
<tr>
<td>Threonine</td>
<td>119.1</td>
<td>101.1</td>
<td>84.9</td>
<td>11.8</td>
</tr>
<tr>
<td>Serine</td>
<td>105.1</td>
<td>87.1</td>
<td>82.9</td>
<td>13.3</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>147.1</td>
<td>129.1</td>
<td>87.8</td>
<td>9.5</td>
</tr>
<tr>
<td>Proline</td>
<td>115.1</td>
<td>97.1</td>
<td>84.4</td>
<td>12.2</td>
</tr>
<tr>
<td>Glycine</td>
<td>75.1</td>
<td>57.1</td>
<td>76.0</td>
<td>18.7</td>
</tr>
<tr>
<td>Alanine</td>
<td>89.1</td>
<td>71.1</td>
<td>79.8</td>
<td>15.7</td>
</tr>
<tr>
<td>Cysteine</td>
<td>121.2</td>
<td>103.1</td>
<td>85.1</td>
<td>11.6</td>
</tr>
<tr>
<td>Valine</td>
<td>117.1</td>
<td>99.1</td>
<td>84.6</td>
<td>12.0</td>
</tr>
<tr>
<td>Methionine</td>
<td>149.2</td>
<td>131.2</td>
<td>87.9</td>
<td>9.4</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>131.2</td>
<td>113.2</td>
<td>86.3</td>
<td>10.7</td>
</tr>
<tr>
<td>Leucine</td>
<td>131.2</td>
<td>113.2</td>
<td>86.3</td>
<td>10.7</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>181.2</td>
<td>163.2</td>
<td>90.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>165.2</td>
<td>147.2</td>
<td>89.1</td>
<td>8.5</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>204.2</td>
<td>186.2</td>
<td>91.2</td>
<td>13.7</td>
</tr>
<tr>
<td>Lysine</td>
<td>146.2</td>
<td>128.2</td>
<td>87.7</td>
<td>19.2</td>
</tr>
<tr>
<td>Histidine</td>
<td>155.2</td>
<td>137.1</td>
<td>88.4</td>
<td>27.1</td>
</tr>
<tr>
<td>Arginine</td>
<td>174.2</td>
<td>156.2</td>
<td>89.7</td>
<td>32.2</td>
</tr>
<tr>
<td>Asparagine</td>
<td>132.1</td>
<td>114.1</td>
<td>86.4</td>
<td>21.2</td>
</tr>
<tr>
<td>Glutamine</td>
<td>146.1</td>
<td>128.1</td>
<td>87.7</td>
<td>19.2</td>
</tr>
<tr>
<td>Ammonia</td>
<td>17</td>
<td>16</td>
<td>94.1</td>
<td>82.2</td>
</tr>
<tr>
<td>Water</td>
<td>18</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

AA: amino acid; Da Dalton; N: nitrogen; NPCF: nitrogen to protein conversion factor.

* The weight of each AA residue is the weight of the free AA minus 18 Da (to account for the loss of one water molecule).

Sources: Sosulski and Imafidon (1990) and Laurens, Olstad and Templeton (2018).
3. METHODS

3.1 Systematic review of the literature

This review was conducted following the JEMNU terms of reference and rules of procedure (FAO/WHO, 2012), including the development of key questions to guide the systematic review in PICO (population, intervention, comparator and outcome) format, and assessment of the certainty in the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.¹

3.1.1 PICO questions

The purpose of this review is to provide the background for the identification of appropriate NPCFs for soy-based and milk-based ingredients used in infant formula and follow-up formula.

The nature of the desired information does not lend itself easily to the PICO format, because strictly speaking we are not dealing with populations, interventions, comparators or different health outcomes. Nevertheless, the key questions needed to guide this review were adapted to PICO format as far as possible and, in keeping with the intent of PICO questions, were framed to be as specific as possible for the problem at hand and to facilitate the search for relevant answers.

The original PICO questions were formulated as follows:

1. When determining the protein content of soy-based ingredients used in infant formula and follow-up formula, what is the appropriate science-based NPCF to use when comparing protein content derived from nitrogen-based methods to amino acid-based methods?

2. When determining the protein content of milk-based ingredients used in infant formula and follow-up formula, what is the appropriate science-based NPCF to use when comparing protein content derived from nitrogen-based methods to amino acid-based methods?

These questions were further refined to sets of PICO questions, with different assumptions regarding the definition of protein:

1. When using the equation $\text{amount of protein (P)} = \text{NPCF (K)} \times \text{amount of nitrogen (N)}$ to estimate the protein content of dairy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content only?

2. When using the equation $\text{amount of protein (P)} = \text{NPCF (K)} \times \text{amount of nitrogen (N)}$ to estimate the protein content of soy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content only?

and

3. When using the equation $\text{amount of protein (P)} = \text{NPCF (K)} \times \text{amount of nitrogen (N)}$ to estimate the protein content of dairy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid plus prosthetic groups?

¹ See http://www.gradeworkinggroup.org/.
4. When using the equation \( \text{amount of protein (P)} = \text{NPCF (K)} \times \text{amount of nitrogen (N)} \) to estimate the protein content of soy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid plus prosthetic groups?

In the revised PICO questions, the population is considered to be infant formula and follow-up formula, and the outcome different conversion factors. Although different values are being compared, there is no direct comparator as part of an intervention. In each case, “nitrogen (N)” may be total nitrogen or total nitrogen corrected for non-protein nitrogen (i.e. protein nitrogen).

3.1.2 Literature review

3.1.2.1 Databases and search terms

The following databases were searched on March 2019, from 1946 to the present:

- Web of Science Core Collection – Clarivate Analytics (using the field “Topic”);
- Medline/PubMed of the National Library of Medicine (using the field “All fields”);
- CAB Direct/CAB Abstracts – CABI (using the field “All fields”); and
- Food Science and Technology Abstracts (FSTA) – Ebsco/IFIS (using the field “All fields”).

The search terms used are provided in Annex 1. In addition to the primary search on NPCFs, a secondary search was conducted to summarize methods for nitrogen and protein analysis, to provide background for a discussion of the results of the review on conversion factors. Reference lists of identified papers were hand searched for additional relevant citations.

3.1.2.2 Call for data

Data on NPCFs are often generated but not published, meaning that unpublished data might be an important source of information; FAO and WHO therefore issued a call for data in November 2018. The call was disseminated through the FAO and WHO websites, the United Nations System Standing Committee on Nutrition (UNSCN) network, and the Codex Alimentarius network.

3.1.3 Data analysis

NPCFs were extracted from published literature and from unpublished reports, where available. Where conversion factors were not reported, but data were available on protein, nitrogen and amino acid analysis, these data were used to calculate the conversion factor(s) (Annex 2). Similarly, NPCFs were recalculated for data in which a conversion factor was reported but either the method of calculation was unclear, or the data provided made it impossible to check the values obtained.

The mean and 95% confidence interval (CI) were calculated for data that were pooled by class of conversion factor across all dairy-based foods, and separately for all soy-based foods. The conversion factor classes, described below, are \( K' \), \( K_a \), and \( K_{50/50} \), \( K_{75/25} \), and \( K_{AD} \) direct adjustment. In addition, dispersion of measurements within different types of

---

2. In these conversion factor classes, 50/50, 75/25 and “direct adjustment” refer to the ratios of amide to glutamic and aspartic acid residues used in the calculation for \( K_a \).
foods (e.g. cow’s milk, milk-based infant formula, soybean and soy isolate) was assessed by standard boxplot analysis.

3.2 Calculation of NPCFs

In this review, conversion factors are designated by \( K' \), KA or KB, depending on how they were calculated.

The conversion factor KA considers the crude protein solely in terms of amino acid residues (Heidelbaugh et al., 1975; Tkachuk, 1969; Sosulski & Imafidon, 1990; Mossé, 1990; Morr, 1981; 1982). It is calculated as the ratio of the sum \( \sum \) of the weight of anhydrous amino acids (determined by amino acid analysis) to protein nitrogen, which is calculated by summating only the nitrogen content of the amino acid residues found in the sample. The equation for calculating \( K_a \) is thus:

\[
K_a = \frac{\sum \text{anhydrous amino acids residues}}{PN}
\]

where “PN” is protein nitrogen.

Except for some highly purified protein fractions, most samples contain non-protein nitrogen (i.e. nitrogen from sources other than protein). \( K_a \) will therefore tend to overestimate the amount of protein in samples with appreciable amounts of non-protein nitrogen, because it considers only nitrogen from protein. Such overestimation will only occur if the factor \( K_a \) is used to multiply total nitrogen. If total nitrogen is corrected for non-protein nitrogen following its analysis along with total nitrogen, then the factor leads to accurate values. Thus, \( K_a \) can provide accurate values in samples with an appreciable amount of non-protein nitrogen, provided that the non-protein nitrogen is measured and subtracted from the total nitrogen measurement, yielding protein nitrogen.

KB takes non-protein nitrogen into consideration and is calculated as the ratio of the sum of the weight of all anhydrous amino acids residues to total nitrogen content, which includes both protein nitrogen and non-protein nitrogen (Holt & Sosulski, 1979; Mossé, 1990). The equation for calculating KB is thus:

\[
K_b = \frac{\sum \text{anhydrous amino acids residues}}{PN+NPN}
\]

where “NPN” is non-protein nitrogen.

Because \( K_b \) considers the presence of non-protein nitrogen in the sample, it has been called the “corrected” conversion factor. It is not necessary to adjust for the amide to acid ratio when calculating KB (or \( K' \) with total nitrogen content), because the entire amount of nitrogen is measured, and thus the actual amount of nitrogen supplied by amide and acid forms of amino acids is already included. However, this factor can lead to inaccuracies because the non-protein nitrogen fraction can vary across different samples for a particular class of protein.

In contrast to \( K_a \) and \( K_b \), which only consider the weight of the anhydrous amino acids, the conversion factor \( K' \) includes the total weight of the protein, and is calculated as the ratio of the sum of the weights of anhydrous amino acids plus the sum of the weights of characterized prosthetic groups obtained from complete amino acid analysis of the protein, to either protein nitrogen or total nitrogen (Van Boekel & Dumas, 1987; Rouch et al., 2008; Maubois & Lorient, 2016). The equation for calculating \( K' \) is thus:

\[
K' = \frac{\sum \text{anhydrous amino acids residues} + \sum \text{prosthetic groups}}{PN \text{ or } PN+NPN}
\]
K’ therefore provides a conversion factor that is close to the early factors obtained in the first calculations used for determining protein concentrations. In those calculations, the crude protein was purified and its weight was measured, but without knowledge of the amino acid composition or prosthetic groups (Hammersten, 1883; Jones, 1931).

There are few K’ values based on protein weights that include prosthetic groups that have been fully characterized and accounted for by amino acid analysis; those that are available are generally limited to dairy proteins and dairy-based ingredients. Because prosthetic groups can add considerably to the weight of proteins, K’ will overestimate the amino acid content of the sample when prosthetic groups are present. In addition, the prosthetic groups are susceptible to processing conditions and are therefore likely to vary both within and among protein sources. To be meaningful, the prosthetic groups should be defined based on the protein post-processing.

K’ is related to K_a and K_b via the following equations:

\[ K' = K_a + \left( \frac{\sum \text{prosthetic groups}}{PN} \right) \]

\[ K' = K_b + \left( \frac{\sum \text{prosthetic groups}}{PN+NPN} \right) \]

Further, because K’ includes prosthetic groups when measuring the weight of protein, and KB includes non-protein nitrogen when assessing nitrogen content, generally:

\[ K' \geq K_a \geq K_b \]

For a given sample containing prosthetic groups and non-protein nitrogen:

\[ K' > K_a > K_b \]

For a given sample without prosthetic groups or non-protein nitrogen:

\[ K' = K_a = K_b \]

### 3.3 Modelling of non-protein nitrogen

KB values were modelled for different amounts of theoretical non-protein nitrogen from the different published values of K_a with amide to acid ratios that were directly assessed and adjusted, or were set arbitrarily at 50/50 or 75/25, as listed in Table 12 and Table 13.

Missing data were imputed by taking the median for a particular set of KA data. For example, for cow’s milk, in column K_a – subcolumn (a) of Table 13 – the third, fourth and fifth missing values (corresponding to Sosulski and Imafidon (1990) and Fujihara, Sasaki and Sugahara (2010), respectively) were replaced with the median of the series, 6.09. The use of the statistical median (rather than the mean or other statistical quantity) avoids working with sparse matrices and guarantees a certain robustness of the modelling.

The data were modelled using the following equation and parameters:

\[ K_b = K_c \times \frac{PN}{[PN + (TN \times \%NPN)]} \]

where:

- \( K_b \) = conversion factor taking into account non-protein nitrogen
- \( K_c \) = conversion factor assuming 0% non-protein nitrogen
- TN = total nitrogen content
- PN = Protein content
- NPN = non-protein nitrogen
and where total nitrogen and protein nitrogen content values for soy products and dairy products come from Boisen, Bech-Andersen and Eggum (1987) and Journet et al. (1975), respectively, as shown in Table 3.

Table 3. Total nitrogen and protein nitrogen content values for soy-based products and dairy-based products

<table>
<thead>
<tr>
<th></th>
<th>Soy-based products</th>
<th>Dairy-based products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein nitrogen</td>
<td>67</td>
<td>62</td>
</tr>
<tr>
<td>Total nitrogen</td>
<td>69</td>
<td>78</td>
</tr>
</tbody>
</table>

Sources: Boisen et al. (1987) and Journet et al. (1975).

3.4 Assessment of risk of bias and certainty of evidence

Cochrane criteria were used to assess risk of bias for each study where applicable, including selective reporting and measurement error (The Cochrane Collaboration, 2011). Some of the Cochrane criteria were not relevant because of the analytical nature of the included studies. The certainty of the evidence was assessed using GRADE, and was modified as necessary to accommodate the unique nature of the included studies, which were not epidemiological or otherwise intervention-based.

Although the studies were analytical in nature and used well-established, objective methods for assessing amino acid composition and nitrogen analysis, measurements made with these methods have inherent inaccuracy. Therefore, conversion factors derived primarily from these types of methods started at “moderate” certainty in the evidence and were downgraded as necessary, based on judgements regarding risk of bias, inconsistency, indirectness and imprecision. Conversion factors derived primarily from studies in which the more accurate method of amino acid sequencing was used was started at “high” certainty in the evidence.

Owing to the highly analytical nature of the studies, and the inability to establish a threshold value for treatment, benefit or harm, imprecision was assessed based on the number of values reported only, with a minimum of 10 values required to prevent downgrading once for serious imprecision, and a minimum of five studies to prevent downgrading twice for very serious imprecision. Inconsistency was assessed via the 95% confidence interval (CI) of the mean, with variation around the mean of more than 10% considered to be serious inconsistency.
4. RESULTS

4.1 Literature review

Regarding the search for NPCFs, after removing duplicates, a total of 3881 publications were initially identified. These publications were mainly scientific articles, but they also included 34 books, 27 book chapters, one report, 12 theses and 23 patents (from CAB Direct and FSTA).

Titles and keywords of these publications were screened for exclusion of background noise articles, followed by further screening of abstracts and article contents, and additional analysis of the bibliography of the selected articles; this resulted in 67 publications being included in the analyses for NPCFs (Fig. 2). Reasons for exclusion of studies included reports of previously published data, publications using the conversion factor of 6.25 or another fixed value but not addressing the calculation, articles addressing products other than milk and soy or not relevant to the objective of the review, and some articles already cited and not providing additional information (Annex 3).

The 67 articles selected for analysis of NPCFs were published between 1883 and 2019 (some historical papers were identified through reference lists), and they included 63 articles in scientific journals, one article in a proceedings book, one report from a public organization (United States Department of Agriculture), one report from the American Oil Chemists Society (AOCS) and one report from the dairy industry. The scientific journals were related to dairy science (12 articles), food science (27 articles), plant and agricultural sciences (15 articles) and other related disciplines (e.g. nutrition, physiology, chemistry and biochemistry); they included four reviews and 58 articles with original data. Of the 63 articles from scientific publications, 61 were from academic institutions and about 30% declared a funding contribution from industry or a private organization. The studies were conducted in Australia, China, Europe, New Zealand and North America.

In addition to the studies identified in the literature review, a total of eight submissions were received from the public call for data, five of which were included in the analyses. Three sets of data were reported in other published or non-published documents already cited in the review, or did not provide data on nitrogen and/or amino acid analysis, and were therefore not included. These sets of data were:

- Agri-Food Canada: reference already cited (Zarkadas et al., 2007);
- Fonterra: amino acid analysis not provided; and
- International Dairy Federation (IDF): data already included in (Maubois & Lorient, 2016).

A separate search was conducted to identify publications relevant for a narrative summary of methods for nitrogen and protein analysis. The searches identified 476 relevant publications, of which 215 were used to inform the summary. These 215 articles are included in the bibliography for information, but not all of these articles were analysed in detail. Results of this search are reported as part of the background in Section 4.2.

---

1 Submissions were received from the following: Japan Vegetable Food Association (milk and soy), Nestec (milk and soy), Dupont (soy), Nutrition Research Division – Bureau of Nutritional Sciences – Health Canada (soy), Agri-Food Canada (soy), ENSA (soy), Fonterra (whey products), and International Dairy Federation (milk, soy).

2 The search was conducted from within the 3881 articles identified in the main search, plus publications in the period 1956–2019 compiled from the ISI Web of Knowledge or other sources (Google Scholar, Science Direct, Wiley) with the words “analytical methods”, “Kjeldahl” and “Dumas method” in the title.
4. RESULTS

Fig. 2  PRISMA flow diagram of study selection for analytical methods and NPCFs

> 6080 records identified in systematic search of databases
  - Background noise removed
  - Duplicates removed

3881 records – Requests on the words of the title, the abstract or the key words

Analytical methods
Nitrogen conversion factors

Records excluded

1316 records – Title, summary, and key words screened

Records excluded

65 abstracts and full text screened
297 abstracts and full text screened

Records excluded

58 records assessed for eligibility

Additional records from other sources

85 records assessed for eligibility

Additional search on data bases – 392 records screened

Duplicates removed and records excluded

5 unpublished sets of data

67 records used

215 records – Analytical methods
72 records – Nitrogen conversion factors

NPCF: nitrogen to protein conversion factors; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
4.2 Analytical methods for nitrogen, amino acid and protein content in foods

Methods for analysis of food proteins are characterized according to their analysis of total nitrogen, amino acid content and total protein content (Table 4). Other common assays include various dye-binding methods that are less suited for routine analysis. Some key references are Leca-Bouvier and Blum (2005); Moore et al. (2010); Noble and Bailey (2009); Owusu-Apenten (2002); Sapan and Lundblad (2015); Simonian and Smith (2006); (Simonian, 2002); Wallace and Fox (1998).

Table 4. Methods for food protein analysis

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NITROGEN ANALYSIS</strong></td>
<td></td>
</tr>
<tr>
<td>Kjeldahl method</td>
<td>Kjeldahl (1883)</td>
</tr>
<tr>
<td>Dumas method</td>
<td>(Dumas, 1831)</td>
</tr>
<tr>
<td><strong>AMINO ACID ANALYSIS</strong></td>
<td></td>
</tr>
<tr>
<td>Moore and Stein</td>
<td>Moore and Stein (1948; 1954)</td>
</tr>
<tr>
<td><strong>PROTEIN ANALYSIS</strong></td>
<td></td>
</tr>
<tr>
<td>Biuret, Lowry, Bradford, Coomassie blue</td>
<td>Gornall, Bardawill and David (1949); Lowry et</td>
</tr>
<tr>
<td>dye – binding, bicinechonic acid</td>
<td>al. (1951); Peterson (1979); Legler et al. (1985);</td>
</tr>
<tr>
<td></td>
<td>Bradford (1976); Compton and Jones (1985)</td>
</tr>
<tr>
<td>Nessler reagent</td>
<td>Wanklin and Chapman (1874)</td>
</tr>
<tr>
<td>Berthelot’s method (alkali-phenol reagent)</td>
<td>Berthelot (1859)</td>
</tr>
<tr>
<td>Folin-Ciocalteu method</td>
<td>Folin and Ciocalteu (1927)</td>
</tr>
<tr>
<td>Dye binding</td>
<td>Fraenkel-Conrat and Cooper (1944)</td>
</tr>
<tr>
<td>Direct alkaline distillation</td>
<td>Owusu-Apenten (2002)</td>
</tr>
<tr>
<td>Modified Berthelot reaction</td>
<td>Searle (1984)</td>
</tr>
<tr>
<td>3-(4-Carboxybenzyl)quinoline-2-carboxaldehyde</td>
<td>You et al. (1997)</td>
</tr>
<tr>
<td>Near infrared reflectance (NIR)</td>
<td>Owusu-Apenten (2002)</td>
</tr>
</tbody>
</table>

Source: adapted from Sáez-Plaza et al. (2013a).

4.2.1 Methods to determine total nitrogen content in foods

The Kjeldahl and Dumas methods, described below, are examples of methods used to determine nitrogen content.

4.2.1.1 Kjeldahl method

The Kjeldahl method was originally designed for the brewing industry, for monitoring protein changes in grain during germination and fermentation (Kjeldahl, 1883; Bradstreet, 1954). First published in 1883, the method has been accepted with modifications as the standard for the determination of nitrogen content (Lynch, Barbano & Fleming, 1998; Moore et al., 2010; Kjeldahl, 1883; Bradstreet, 1954) (Table 5). The Kjeldahl method and devices have been significantly modified over the past 100 years, but the basic principles are still valid and include three main steps (Adesina, 2012; Benton Jones Jr., 1991):
Table 5. Kjeldahl methods and modifications for total protein measurement in food, by matrix

<table>
<thead>
<tr>
<th>SPECIFIC MATRIX</th>
<th>METHOD REFERENCE</th>
<th>SPECIFIC MATRIX</th>
<th>METHOD REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereal grains</td>
<td>AOAC 979.09; AACC 46–11A; AACC 46–16; AACC 46–16; (Beljkaš et al., 2010; Salo- vänänen &amp; Koivistoinen, 1996)</td>
<td>Fruit products</td>
<td>AOAC 920.152</td>
</tr>
<tr>
<td>Cereal adjuncts</td>
<td>AOAC 945.18; AACC 46–11A; AACC 46–16; (Gouveia et al., 2014)</td>
<td>Gelatin</td>
<td>AOAC 935.46</td>
</tr>
<tr>
<td>Flour (wheat)</td>
<td>AOAC 920.87; AACC 46–11A; AACC 46–12; AACC 46–13; AACC 46–16; AACC 46–30; (Ronalds, O’Brien &amp; Allen, 1984)</td>
<td>Infant formula (milk-based)</td>
<td>AOAC 986.25 (Brooke, Wood &amp; Barley, 1982)</td>
</tr>
<tr>
<td>Soy flour</td>
<td>AOCS Bc 4–91</td>
<td>Laboratory malt</td>
<td>AOAC 950.09</td>
</tr>
<tr>
<td>Soybeans</td>
<td>AOCS Ac 4–91</td>
<td>Maccaroni products</td>
<td>AOAC 930.25</td>
</tr>
<tr>
<td>Meat</td>
<td>AOAC 981.10</td>
<td>Milk chocolate</td>
<td>AOAC 939.02</td>
</tr>
<tr>
<td>Cream</td>
<td>AOAC 920.109</td>
<td>Plants</td>
<td>AOAC 977.02; AOAC 978.04</td>
</tr>
<tr>
<td>Baked products</td>
<td>AOAC 935.39</td>
<td>Starch dessert powder</td>
<td>AOAC 945.56</td>
</tr>
<tr>
<td>Beer</td>
<td>AOAC 920.53</td>
<td>Sunflower seeds</td>
<td>AOCS Ai 4–91</td>
</tr>
<tr>
<td>Bread</td>
<td>AOAC 950.36</td>
<td>Sweetened condensed milk</td>
<td>AOAC 920.115</td>
</tr>
<tr>
<td>Brewing sugars and syrups</td>
<td>AOAC 945.23</td>
<td>Tea</td>
<td>AOAC 920.103</td>
</tr>
<tr>
<td>Cacao</td>
<td>AOAC 970.22</td>
<td>Wine</td>
<td>AOAC 920.70</td>
</tr>
</tbody>
</table>


* Including milk-based infant formula, milk protein concentrate, whey protein concentrate, casein and caseinate.

1. Digestion – the decomposition of nitrogen from organic samples by boiling in concentrated sulfuric acid resulting in an ammonium sulfate solution.

2. Distillation – adding excess base to the acid digestion mixture, which converts methane (NH₄⁺) to ammonia (NH₃), followed by boiling and condensation of the NH₃ gas in a receiving solution.

3. Titration – the amount of ammonia in the receiving solution is quantified.

This method has been discussed extensively in various reviews (e.g. Dyer, 1895; Morries, 1983; Benton Jones Jr., 1987; Horneck & Miller, 1998; Lynch & Barbano, 1999; Owusu-Apenten, 2002; Moore et al., 2010; Sáez-Plaza et al., 2013a; Sáez-Plaza et al., 2013b; Sapan & Lundblad, 2015).
The Kjeldahl method provides total nitrogen content; it does not distinguish between protein-based nitrogen and non-protein nitrogen, including inorganic nitrogen and other organic nitrogen (e.g. urea, melamine, nucleotides and nucleic acids). Inorganic nitrogen is often negligible in food and biological samples (Simonne et al., 1997), but urea may not be negligible, depending on the type of biological samples analysed (Grappin, 1992). The nitrogen-containing nitrate portion can be higher in products from vegetables fertilized with nitrate-based fertilizers, and the recovery rate during the Kjeldahl digestion procedure affects the determination of total nitrogen (Simonne et al., 1994). The limitations of the method have been studied extensively, and many improvements have been made to each of the steps in the process (Watson & Galliher, 2001).

4.2.1.2 Dumas method

As with the Kjeldahl method, the Dumas method is considered a direct method for determination of total nitrogen content. It was developed in 1831 by Dumas, from the observation that nitrogenous compounds heated with alkali give ammonia, which can be determined volumetrically (Szabadváry & Robinson, 1980). The method was further improved by igniting the sample with barium hydroxide, and quantifying the ammonia liberated into HCl gravimetrically by precipitation as ammonium hexachloroplatinate (Szabadváry & Robinson, 1980). In a modified method, the ammonia is absorbed in a known amount of HCl, and the excess is back-titrated using lime dissolved in water containing sugar (Klosterman, 1985). This process can be fully automated, taking only a few minutes per measurement (Saint-Denis & Goupy, 2004; Whitesell et al., 2014).

Currently, the method consists of combusting a sample of known mass in a high-temperature (range, 800–900 °C) chamber in the presence of oxygen, producing carbon dioxide (CO₂), water and nitrogen. The gases are then passed over special columns (e.g. those containing a potassium hydroxide aqueous solution) that absorb CO₂ and water. A column containing a thermal conductivity detector at the end is then used to separate the nitrogen from residual CO₂ and water, and the remaining nitrogen content is measured. The instrument must initially be calibrated by analysing a pure material containing a known nitrogen concentration. The measured signal from the thermal conductivity detector for the unknown sample is converted into a nitrogen content.

As with the Kjeldahl method, the Dumas method does not give a measure of true protein because it includes non-protein nitrogen, and different correction factors are needed for each protein because each has a different amino acid sequence. Total nitrogen values calculated by the Dumas method can be slightly higher than the Kjeldahl values because Dumas nitrogen also includes nitrates and organic compounds that are highly resistant to acid digestion (Lakin, 1978; Petterson et al., 1999). As a result, differences can occur between the two methods with some types of food (e.g. fruits, vegetables and fish).

A Kjeldahl nitrogen and Dumas nitrogen correction factor of 0.15 has been proposed (Simonne et al., 1997). During harvest years 2000–2004, the Max Rubner Institute in Detmold, Germany, performed a comprehensive study of more than 800 wheat samples, comparing the crude protein results between the Kjeldahl and Dumas methods. The study found that some 2% of “Dumas protein” was not determined by the Kjeldahl method, and presented the following relationship between Dumas and Kjeldahl protein values: Kjeldahl = 0.959 * Dumas + 0.258 (Müller, 2017). However, this equation cannot be generalized to other plant and food sources, because different factors (e.g. cultivar, growing year and conditions) can modify the results.
A relationship was also proposed for plant tissues in general: Kjeldahl = 0.85 * Dumas + 0.0.015 (Simonne et al., 1994). In some selected food groups, the Dumas method can replace the Kjeldahl method by using correction factors (Simonne et al., 1997); however, the differences between the two methods are often not statistically significant (Mihaljev et al., 2015). Some of the differences between the two methods when applied to the same sample can also depend on the type of catalyst used, and this may be a further source of variability.

The number of publications related to the Dumas method is much lower than for the Kjeldahl method; however, when the use of mercury and cadmium in the laboratory was banned in most countries in the 1990s, the Dumas method was evaluated as an alternative to the Kjeldahl method (Table 6). The Dumas method has been automated (Leco analyser) and is now routinely used in many laboratories.

Table 6. Standards for total nitrogen and protein measurement in food and feed using the Dumas method

<table>
<thead>
<tr>
<th>REFERENCE</th>
<th>TITLE (MATRIX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 16634–2:2016</td>
<td>Food products – Determination of the total nitrogen content by combustion according to the Dumas principle and calculation of the crude protein content – Part 2: Cereals, pulses and milled cereal products.</td>
</tr>
<tr>
<td>AACC Method 46.30</td>
<td>Crude protein — Combustion method (animal feeds, cereals and oil seeds).</td>
</tr>
<tr>
<td>ICC Standard No. 167</td>
<td>Determination of crude protein in grain and grain products for food and feed by the Dumas combustion principle.</td>
</tr>
<tr>
<td>AOAC 990.03</td>
<td>Protein (crude) in animal feed — Combustion method.</td>
</tr>
<tr>
<td>AOAC 992.23</td>
<td>Crude protein in cereal grains and oil seeds.</td>
</tr>
<tr>
<td>AOAC 997.09</td>
<td>Nitrogen in beer, wort, and brewing grains — Protein (total) by calculation — Combustion method.</td>
</tr>
<tr>
<td>AOCS Ba 4e-93</td>
<td>Generic combustion method for determination of crude protein (oilseed byproducts).</td>
</tr>
<tr>
<td>AOCS Ba 4f-00</td>
<td>Combustion method for determination of crude protein in soybean meal.</td>
</tr>
<tr>
<td>OIV-MA-AS323–02A</td>
<td>Quantification of total nitrogen according to the Dumas method in musts and wines (Type II method).</td>
</tr>
</tbody>
</table>


4.2.2 Methods for amino acid analysis

Amino acid analysis was first developed around 1950 (Moore & Stein, 1948; 1954) to determine the amino acid content of pure proteins. Before performing such analysis, it is necessary to hydrolyse a protein to its individual amino acid constituents. The amino acid constituents of the test sample are typically derivatized for analysis. The techniques commonly used to separate free amino acids are ion-exchange chromatography, gas chromatography, high-performance liquid chromatography (HPLC) and its ultra-fast variant UHPLC, and electrophoresis.
4.2.2.1 Sample hydrolysis

Hydrolysis of protein and peptide samples is necessary for amino acid analysis. Hydrolysis with HCl is performed at 110 °C for 20–70 hours, with the temperature being accurately controlled. After hydrolysis, residual HCl is removed in a rotary evaporator, and the residue is dissolved in water or buffer, depending on whether the separation method will be ion-exchange chromatography or HPLC. The standard method of hydrolysis is incubation in an oxygen-free environment in constant boiling 6 M HCl at 110 °C for 18–24 hr, with 24-hour hydrolysis being used most often (Ozols, 1990; Darragh & Moughan, 2005). The acid hydrolysis releases free amino acids but can produce uncontrolled complete or partial destruction of several amino acids (Table 7). Owing to the unstable behaviour of amino acids under the hydrolysis conditions described above, of the 20 amino acids commonly present in proteins, only 10 can be determined quantitatively after acid hydrolysis; they are aspartic acid, glutamic acid, proline, glycine, alanine, leucine, phenylalanine, histidine, arginine and hydroxyproline. Amino acids that are totally or partially destroyed include tryptophan, cysteine, serine and threonine, while methionine might undergo oxidation. In peptide bonds involving isoleucine and valine, the amino bonds of isoleucine–isoleucine, valine–valine, isoleucine–valine and valine–isoleucine are partially cleaved. The amide amino acids asparagine and glutamine are deamidated, resulting in aspartic acid and glutamic acid, respectively.

Not all amino acids are fully released or are partially destroyed under conditions of standard hydrolysis (i.e. 24-hour acid hydrolysis); hence, values obtained with this method can only be considered as approximations. More accurate information on amino acid content of a protein can be obtained by using multiple hydrolysis times and fitting data to a curvilinear model that gives instantaneous rates of release and destruction, and that can be used to accurately predict the amino acid composition of the protein (Rutherfurd & Moughan, 2018). The model can be applied to acid hydrolysis, acid hydrolysis after performic acid oxidation for sulfur amino acids, and alkaline hydrolysis for tryptophan.

Several hydrolysis techniques have been described and used to address these concerns. For example, formation of cysteic acid, methionine sulfoxide and chlorotyrosine can be avoided if all air is removed from the sample before hydrolysis. The HCl hydrolysis procedure also minimizes decomposition of reduced S-carboxymethylcysteine and preserves S-carboxymethylated amino acids. To overcome problems associated with the substantial loss of amino acid residues during acid or basic hydrolysis, enzymatic hydrolysis can be used to hydrolyse proteins from animal tissues or foods to obtain amino acids in soluble form for optimizing the recovery of glutamine, glutamate, asparagine and aspartate residues (Peace & Gilani, 2005; Tsao & Otter, 1999). The use of enzymatic hydrolysis is interesting compared with acid or basic hydrolysis, because it avoids the racemization of certain amino acid residues (e.g. 2–5% for aspartate, histidine, lysine and methionine residues with conventional methods and with microwave heating in the absence of phenol). Weiss et al. (1998) provided a list of 45 protein hydrolysis methods with their corresponding references.

4.2.2.2 Chromatography

There are several chromatographic techniques for amino acid analysis, and the choice depends on the sensitivity required for the assay. The separation of free amino acids by ion-exchange chromatography followed by post-column derivatization with ninhydrin is the traditional amino acid analysis technique (Moore & Stein, 1948). Post-column detection techniques can be used with samples that contain small amounts of buffer components (e.g. salts and urea), and generally requires 5–10 μg of protein sample per analysis.
### Table 7. Problematic amino acids during the protein hydrolysis procedure

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Drawback During Hydrolysis</th>
<th>Possible Solutions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan</td>
<td>Destroyed in presence of carbohydrates, starch or sugars and heavy metals</td>
<td>• Add thiols</td>
<td>Finley (1985); Hanko and Rohrer (2002); Landry, Delhaye and Jones (1992); Rutherford and Gilani (2009); Steinhart (1984); Rowan, Moughan and Wilson (1989); Bech-Andersen (1991)</td>
</tr>
<tr>
<td>Cysteine and methionine</td>
<td>Cysteine is destroyed, methionine is oxidized to methionine sulfoxide and methionine sulfone</td>
<td>• Remove oxygen from the hydrolysis tube</td>
<td>Rutherford and Gilani (2009); Ellinger and Duncan (1976); Elias, McClements and Decker (2005); Hayashi and Suzuki (1985); Sochaski et al. (2001); Cuq et al. (1978); Todd, Marable and Kehrberg (1984)</td>
</tr>
<tr>
<td>Asparagine and glutamine</td>
<td>Deamination to produce aspartic acid and glutamic acid leading to an overestimation of aspartic acid and glutamic acid</td>
<td>• Esterification-reduction</td>
<td>Wilcox (1967); Carraway and Koshland (1972); Soby and Johnson (1981)</td>
</tr>
<tr>
<td>Serine and threonine</td>
<td>Partially destroyed: losses of 10–15% for serine and 5–10% for threonine</td>
<td>• Reduce hydrolysis time</td>
<td>Ozols (1990); Darragh and Moughan (2005); Rowan et al. (1989); Bech-Andersen (1991)</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Halogenation of tyrosine Decreasing of yield recovery in presence of iron or copper ions Decreasing of yield recovery in presence of fat content</td>
<td>• Add phenol to 6 M HCl</td>
<td>Nissen (1992); Finley (1985)</td>
</tr>
<tr>
<td>Valine and isoleucine</td>
<td>Cleavage of peptide bonds of these amino acids is particularly difficult in standard acid hydrolysis conditions</td>
<td>• Increase hydrolysis time to 72 hours</td>
<td>Rayner (1985)</td>
</tr>
<tr>
<td>Lysine and hydroxylysine</td>
<td>In case of food heat-processing, possible Maillard reaction with reducing sugars; these Maillard products are labile in acidic conditions and can revert back to lysine, leading to overestimation of native lysine content</td>
<td>Apply methods using: • FDNB (^a) • FDNB-difference • TNBS • sodium borohydride • furosine • dye binding • ninhydrine-reactive lysine • o-phtaldialdehyde-reactive lyine guanidination (^c)</td>
<td>Rao, Carter and Frampton (1963); Carpenter and Bjarnason (1968); Hurrell and Carpenter (1974); Desrosiers et al. (1989); Hendriks et al. (1994); Friedman, Pang and Smith (1984); Vigo et al. (1992); Mauron and Bujard (1964); Mao, Lee and Erbersboeber (1993); Rutherford and Moughan (1997); Torbatinejad, Rutherford and Moughan (2005)</td>
</tr>
</tbody>
</table>

**CF:** conversion factor; **FDNB:** 1-fluoro-2,4-dinitrobenzene; **TNBS:** 2,4,6-trinitrobenzene sulfonic acid.

\(^a\) Barium hydrolysis forms precipitate over time as barium carbonate or barium sulfate if sulfates are present.

\(^b\) Few laboratories use this method.

\(^c\) These methods are the most commonly used.

Source: adapted from Rutherford and Gilani (2009).
Other amino acid analysis techniques involve precolumn derivatization of the free amino acids followed by chromatography (Dai et al., 2014). Precolumn derivatization techniques are sensitive and usually require 0.5–1.0 μg of protein sample per analysis, but may be influenced by buffer salts in the samples. The most commonly used chromatography for amino acid analysis with precolumn derivatization is reverse-phase chromatography (RP-HPLC), but hydrophilic-interaction liquid chromatography provides an alternative approach (Buszewski & Noga, 2012; Alpert, 1990; Hemstrom & Irgum, 2006; de Castro & Sato, 2015; La Barbera et al., 2017; Sánchez-Rivera et al., 2014; Stodt & Engelhardt, 2013). Gas chromatography and liquid chromatography coupled with mass spectrometry are also applied to amino acid analysis (Merrick et al., 2011; Przyborowska et al., 2004).

Precolumn derivatization techniques may result in multiple derivatives of a given amino acid, which makes interpretation of the results more challenging. Silylation is the most commonly used derivatization technique, although it has been reported that alkylation with methyl chloroformate improves analytical performance (Smart et al., 2010; Villas-Bôas et al., 2011). The instability problems of derivatization reagents have been largely eliminated by online derivatization. Numerous strategies for derivatization have been developed and published over recent decades (Table 8) (Callejon, Troncoso & Morales, 2010; Fekkes, 1996; Molnár-Perl, 2003). The final performance of the liquid or gas chromatography depends on the efficiency of the derivatization procedure, and subpicomole quantities of protein can be accurately analysed (Fekkes, 1996).

Numerous studies have been published on protein and amino acid analysis by liquid or gas chromatography (coupled or not coupled with mass spectrometry), including studies on milk or dairy products (Marino et al., 2010) and soybean or soy products (Table 9). Gas chromatography generally gives lower relative standard deviations (i.e. is more accurate) than liquid chromatography, while only high-resolution UHPLC (UHPLC-Q-Orbitrap, HRMS) provides analysis close to gas chromatography.

Table 8. Main derivatization methods based on specific amine or amide reagents for amino acids analysis

<table>
<thead>
<tr>
<th>REAGENT OF DERIVATIZATION</th>
<th>SENSITIVITY RANGE</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POST-COLUMN</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ninhydrin</td>
<td>10 pmol – 50 pmol</td>
<td>Samejima et al. (1971); Moore and Stein (1948; 1954)</td>
</tr>
<tr>
<td>OPA Fluorometric</td>
<td>5–10 times more sensitive than ninhydrin</td>
<td>Benson and Hare (1975); Lee and Drescher (1978)</td>
</tr>
<tr>
<td><strong>PRECOLUMN</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylisothiocyanate</td>
<td>–</td>
<td>Bidlingmeyer et al. (1987); Bidlingmeyer, Cohen and Tarvin (1984)</td>
</tr>
<tr>
<td>AQC</td>
<td>40–800 fmol, depending on the amino acid</td>
<td>Cohen and Michaud (1993)</td>
</tr>
<tr>
<td>DABS-Cl</td>
<td>–</td>
<td>Akhlaghi et al. (2015)</td>
</tr>
<tr>
<td>FMOC-Cl</td>
<td>Low fmol range (very quick reaction, 30 sec to 1 min)</td>
<td>Einarsson, Joseffson and Lagerkvist (1983)</td>
</tr>
<tr>
<td>NBD-F</td>
<td>2.8–20 fmol</td>
<td>Aoyama et al. (2004)</td>
</tr>
</tbody>
</table>

AQC: 6-aminoquinolyl-N-hydroxyscinimidyl carbamate; DABS-Cl: 4-N,N-dimethylaminoazobenzene-4′-sulfonyl chloride; FMOC-Cl: fluorenlymethyl chloroformate; NBD-F: 4-fluoro-7-nitro-2,1,3-benzoxadiazole; OPA: o-phthalaldehyde.

Source: adapted from Rutherford and Gilani (2009).
### Table 9. A selection of studies specifically related to amino acid analysis in milk and soybean products by gas chromatography and liquid chromatography

<table>
<thead>
<tr>
<th>PRODUCT OR MATRIX</th>
<th>TECHNIQUE</th>
<th>AMINO ACIDS DETERMINED (RESIDUES %)</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SOYBEAN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC-FID and RP-HPLC</td>
<td>GC</td>
<td>Ala (2.2); Arg (ND); Asp (1.0); Cys (8.6); Glu (1.2); Gly (1.3); His (ND); Ile (1.6); Leu (0.7); Lys (3.1); Met (22.4); Phe (4.4); Pro (1.1); Ser (1.6); Thr (1.0); Tyr (3.4); Val (1.1)</td>
<td>Adeola, Buchanansmith and Early (1988)</td>
</tr>
<tr>
<td></td>
<td>LC</td>
<td>Ala (3.3); Arg (12.3); Asp (4.9); Cys (6.7); Glu (5.4); Gly (2.4); His (10.4); Ile (3.3); Leu (1.8); Lys (6.5); Met (23.5); Phe (5.6); Pro (2.2); Ser (3.8); Thr (3.4); Tyr (6.9); Val (12.1)</td>
<td></td>
</tr>
<tr>
<td><strong>RP-HPLC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LC-MS-MS</strong></td>
<td></td>
<td>Ala (4.38); Arg (3.98); Asp (3.55); Cys (33.89); Glu (3.77); Gly (10.15); His (9.89); Ile (3.79); Leu (2.65); Lys (6.04); Met (6.68); Phe (2.69); Pro (6.57); Ser (7.07); Thr (5.59); Trp (8.28); Tyr (7.51); Val (2.18); NH3</td>
<td>Zarkadas et al. (2007)</td>
</tr>
<tr>
<td><strong>RP-HPLC-Fluo</strong></td>
<td></td>
<td>Ala (4.6); Arg (7.1); Asp+Asn(11.9); Cys (1.2); Glu+Gln (19.0); Gly (4.5); His (3.1); Ile (4.7); Leu (8.0); Lys (6.1); Met (1.8); Phe (5.6); Pro (5.6); Ser (4.8); Thr (4.0); Tyr (2.9); Val (5.3)</td>
<td>Kwanyuen and Burton (2010)</td>
</tr>
<tr>
<td><strong>MILK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RP-HPLC-Fluo</strong></td>
<td></td>
<td>Ala (6.04); Arg (6.12); Asp (4.29); Cys (24.7); Glu (1.73); Gly (3.11); His (1.79); Ile (5.80); Leu (0.72); Lys (3.76); Met (12.08); Phe (1.28); Pro (7.17); Thr (10.03); Tyr (14.08); Val (7.93)</td>
<td>Hejtmankova et al. (2012)</td>
</tr>
<tr>
<td><strong>RP-HPLC-DAD</strong></td>
<td></td>
<td>Ala (6.7); Arg (6.2); Asp (5.5); Glu (5.5); Gly (6.4); His (7.6); Ile (6.2); Leu (6.2); Lys (6.6); Met (23.0); Phe (6.6); Pro (6.4); Ser (6.6); Thr (6.2); Tyr (6.5); Val (5.8)</td>
<td>Lu et al. (2015)</td>
</tr>
<tr>
<td><strong>UHPLC-Q Orbitrap HRMS</strong></td>
<td>Arg (9.6); Asp (2.7); Asp (5.1); Cys (3.4); Glu (1.3); Glu (0.2); Gly (3.8); His (0.3); Ile (0.2); Leu (2.7); Lys (0.6); Met (3.5); Phe (2.6); Pro (1.1); Ser (2.0); Thr (3.8); Trp (7.4); Tyr (4.9); Val (5.0)</td>
<td>Yin et al. (2016)</td>
<td></td>
</tr>
</tbody>
</table>


Amino acids: Ala: alanine; Arg: arginine; Asp: aspartate; Cys: cysteine; Glu: glutamate; Gly: glycine; His: histidine; Ile: isoleucine; Leu: leucine; Lys: lysine; Met: methionine; Phe: phenylalanine; Pro: proline; Ser: serine; Thr: threonine; Tyr: tyrosine; Val: valine.

#### 4.2.2.3 Capillary electrophoresis

Capillary electrophoresis refers to different techniques that operate in liquid media and use capillary columns in which solvated ions, ionized species and neutral species migrate with different velocities (Righetti, 2005). They can be separated under the action of an electric field and pH-controlled conditions. Different modes of electromigration exist with some applications in food analysis, including protein and amino acid analysis (Nazzaro et al., 2012; Blatný & Kvasnicka, 1999; Hiroyuki & Terabe, 1996; Terabe, 2009; Mala, Gebauer & Bocek, 2016; Svec, 2009; Mikšík, 2018; Kenndler, 2014a; de Oliveira et al., 2016; Voeten et al., 2018; Castro & Manz, 2015; Zhu, Lu & Liu, 2012; Mala, Gebauer & Bocek, 2015; Kenndler, 2014b).
Proteins and amino acids have traditionally been separated by polyacrylamide gel electrophoresis (PAGE); more recently, fast capillary electrophoresis separation of amino acids, with minimal or even no sample pretreatment, was developed for plasma and urine analysis (Kitagishi & Shintani, 1998; Martin-Girardeau & Renou-Gonnord, 2000; Soga & Heiger, 2000). The latter is also a convenient method for food quality control, including for analysis of proteins in milk and soybean products (Castro-Puyana et al., 2012; Otter, 2012; Jager, Tonin & Tavares, 2007; Zhu et al., 2012; Lara et al., 2006; Santos et al., 2005; Meng et al., 2009; Bailon-Perez et al., 2007; Garcia-Ruiz et al., 2006; Erny et al., 2008).

Capillary electrophoresis instruments are sensitive, selective, inexpensive and easy to use for a wide variety of applications (Castro-Puyana et al., 2012; El Deeb, Iriban & Gust, 2011; Garcia-Canas et al., 2014; Kaisoon, Siriamornpun & Meeso, 2008; Andree et al., 2012; Nazzaro et al., 2012; Pinero, Bauza & Arce, 2011). Advances in microchip-channel fabrication now allow analysis to be completed in a shorter time (Culbertson, Jacobson & Ramsey, 2000; Nouadje et al., 1995; Thorsén & Bergquist, 2000; Chen, Warren & Adams, 2000; Thornton, Fritz & Klampfl, 1997). These techniques are promising for amino acid analysis. A comparison between different separation methods used in amino acid analysis is shown in Table 10.

### Table 10. Comparison of routine methods of food protein analysis

<table>
<thead>
<tr>
<th>FACTORS CONSIDERED</th>
<th>PAGE</th>
<th>RP-HPLC</th>
<th>Conventional CE</th>
<th>Microfluidic CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time for setting gel or regenerating column</td>
<td>60 min</td>
<td>10 min</td>
<td>2–3 min</td>
<td>2–3 min</td>
</tr>
<tr>
<td>Sample extraction</td>
<td>Depending on source</td>
<td>Depending on source</td>
<td>Depending on source</td>
<td>Depending on source</td>
</tr>
<tr>
<td>Run time</td>
<td>30–240 min</td>
<td>10–90 min</td>
<td>10 min</td>
<td>1–3 min</td>
</tr>
<tr>
<td>Visualization of proteins</td>
<td>2–12 hr</td>
<td>Instant</td>
<td>Instant</td>
<td>Instant</td>
</tr>
<tr>
<td>Throughput in 24 hr</td>
<td>20/gel</td>
<td>30</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>Health risks for operator</td>
<td>Moderate</td>
<td>Low</td>
<td>Low–medium</td>
<td>Low</td>
</tr>
<tr>
<td>Cost of equipment</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Cost of consumables</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low–medium</td>
</tr>
</tbody>
</table>

Source: adapted from Kaisoon et al. (2008).

#### 4.2.2.4 Protein sequencing

Protein sequencing is the determination of the amino acid sequence of a protein, and can include the characterization of post-translational modifications. The two major methods of protein sequencing are the traditional Edman degradation and the more recent mass spectrometry. In the Edman degradation, the phenyl isothiocyanate reacts with the amine group of the N-terminal amino acid, which can be separated and identified by chromatography (Niall, 1973). More recently, analysis by a tandem mass spectrometer has been applied to the sequencing of peptides by using database search and de novo sequencing (Webb-Robertson & Cannon, 2007; Lu, 2004). Calculation of NPCFs from protein sequence data has been applied to milk protein only (Maubois & Lorient, 2016).
**4.2.3 Other methods for protein analysis**

**4.2.3.1 Colorimetric methods for total protein content analysis**

Colorimetric methods for total protein content include Biuret, Lowry assay, bicinchoninic acid (BCA) assay and Bradford (CBB G-250 dye binding).

In the Biuret method, in an alkaline medium the peptide bond forms a blue-violet (purple) complex with copper II ions (Cu^{2+}), whose colour is proportional at 540 nm to the protein concentration (Gornall et al., 1949). The Biuret method is relatively fast and inexpensive but has a high quantitation threshold (about 1 mg/L) (Chebaro et al., 2017; Doumas et al., 1981a; Doumas et al., 1981b; Sapan & Lundblad, 2015). Different interferences have been reported (Sapan & Lundblad, 2015), and the Biuret protein assay cannot be used for most dairy products because lactose interferes in the assay (Finete et al., 2013; Keller & Neville, 1986; Sapan & Lundblad, 2015). Several publications have compared the Biuret method with other protein determination methods (Lott, Stephan & Pritchard, 1983; Gadsden, 1983; Eckfeldt, Kershaw & Dahl, 1984; Mohammad & Stomer, 1991; George & O’Neill, 2001; Briend-Marchal, Medaille & Braun, 2005; Martina & Vojtech, 2015; Katsoulos et al., 2017).

In the Lowry method, proteins react first with an alkaline copper reagent and then with a second reagent comprising a mixture of sodium tungstate and sodium molybdate dissolved in phosphoric acid and HCl (Lowry et al., 1951; Chebaro et al., 2017). This reagent allows the reduction of aromatic amino acids (tyrosine and tryptophan), leading to the formation of a dark blue coloured complex whose absorbance is measured between 650 and 750 nm. The Lowry method has a higher sensitivity than the Biuret method (100-fold more sensitive), and can thus be used to carry out measurements in low protein mixtures or in relatively dilute solutions.

The BCA method is a modification of the Lowry assay (Sarwar et al., 1983a). It is a copper-binding method with optimal conditions for maximizing colour absorbance at 562 nm being 60 °C and pH of 11.25 (Owusu-Apenten, 2002). The relationship between protein concentration and absorbance is nearly linear over the range 0.02–2 μg/μL (Bainor et al., 2011; Walker, 2002). However, many compounds interfere with this method, and it is not well-suited for total protein determination in routine analysis (Table 11) (Sarwar et al., 1983a; Owusu-Apenten, 2002).

**Table 11. Main interfering compounds for BCA protein assay**

<table>
<thead>
<tr>
<th>INTERFERING COMPOUNDS</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biogenic amines: dopamine, norepinephrine, tyrosine, serotonin (5-HT), tryptophan</td>
<td>Owusu-Apenten (2002); Slocum and Deupree (1991)</td>
</tr>
<tr>
<td>Acetol, aminophenol, ascorbic acid, 2,3-butanedione, glucose, glyoxal, 2,4-dinitrophenylhydrazine, pyruvic aldehyde</td>
<td>Chen et al. (1990)</td>
</tr>
<tr>
<td>Lipids: phosphatidylcholine, phosphatidylethanolamine, phosphatidylinerine, phosphatidylinositol, cardiolipin, sphingomyelin</td>
<td>Kessler and Fanestil (1986)</td>
</tr>
<tr>
<td>Phenols: gallic acid, tannic acid, pyrogallic acid, pyrocatechol</td>
<td>Kamath and Pattabiraman (1988)</td>
</tr>
<tr>
<td>Reducing sugars (e.g. glucose, fructose); nonreducing sugars (sucrose)</td>
<td>Sapan and Lundblad (2015)</td>
</tr>
</tbody>
</table>

BCA: bicinchoninic acid.
In the Bradford assay, binding of the dye Coomassie® Brilliant Blue to protein in an acidic solution leads to a concomitant absorbance shift from 465 nm to 595 nm (Antonio Agustí & Pío Beltrán, 1982). This detection method can also be used in sodium dodecyl sulfate PAGE (SDS-PAGE) gel or other matrices where protein concentration can be determined by densitometry (Bonde, Pontoppidan & Pepper, 1992; Sapan & Lundblad, 2015). Coomassie Brilliant Blue binds to proteins approximately stoichiometrically. The relationship is linear and a regression curve can be derived from a series of standards over a range of 0.02–0.50 mg/mL (Okutucu et al., 2007; Moore et al., 2010). The method is affected by different factors, including detergents, pH and polyphenols (tannins) of high molecular weight (Whiffen, Midgley & McGee, 2007; Martens, Reedy & Lewis, 2004; Redmile-Gordon et al., 2013; Robinson, 1979; Antonio Agustí & Pío Beltrán, 1982) (Table 12). Amino acids, peptides and low molecular weight proteins (<3000 Da) are not detected by this method.

Table 12. Main compounds interfering with the Bradford reagent (Coomassie Brilliant Blue G250) or with proteins in samples, and incompatible sample preparation methods

<table>
<thead>
<tr>
<th>INTERFERING COMPOUNDS</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant polysaccharides, tannins, glycosaminoglycans (heparin)</td>
<td>Godshall (1983); Khan and Newman (1990)</td>
</tr>
<tr>
<td>Chlorophyll, pectins, ethanol, ionic and nonionic surfactants, lipids, flavonoids</td>
<td>Owusu-Apenten (2002)</td>
</tr>
<tr>
<td>Protein glycosylation</td>
<td>Khan and Newman (1990); Fountoulakis, Juranville and Manneberg (1992)</td>
</tr>
<tr>
<td>Glycation of human serum</td>
<td>Brimer et al. (1995)</td>
</tr>
</tbody>
</table>

**INCOMPATIBLE SAMPLE PREPARATION METHODS**

| Sodium deoxycholate – trichloroacetic acid (DOC-TCA) precipitation used to isolate sample protein before analysis, because of the formation of precipitate that interferes with this method | Owusu-Apenten (2002); Chiappelli, Vasil and Haggerty (1979); Pande and Murthy (1994) |

In one study, the different methods were compared and applied to the same pooled plasma sample; each assay was found to have advantages and disadvantages in terms of sensitivity, performance, accuracy, and reproducibility or coefficient of variation (CV) (Okutucu et al., 2007) (Table 13 and Table 14). All the methods had a CV of less than 6%.

Table 13. Linearity range and limit of detection of most used colorimetric assays for protein determination

<table>
<thead>
<tr>
<th>ASSAY</th>
<th>REGRESSION COEFFICIENT (standard graphs)</th>
<th>LINEARITY RANGE (mg/mL)</th>
<th>LIMIT OF DETECTION (mg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowry</td>
<td>Y=1.5473x</td>
<td>0.02 – 0.5</td>
<td>0.015</td>
</tr>
<tr>
<td>Bradford</td>
<td>Y=1.677x</td>
<td>0.02 – 0.5</td>
<td>0.018</td>
</tr>
<tr>
<td>Biuret</td>
<td>Y=0.15x</td>
<td>0.5 – 4.0</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Source: adapted from Okutucu et al. (2007).
Table 14. Protein concentrations determined by most used colorimetric assays for standard protein samples

<table>
<thead>
<tr>
<th>SAMPLE PROTEIN CONCENTRATION</th>
<th>ASSAY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plasma, n=7 (graph)</td>
</tr>
<tr>
<td></td>
<td>(mg/mL)</td>
</tr>
<tr>
<td>Lowry</td>
<td>Median ± SD</td>
</tr>
<tr>
<td></td>
<td>CV %</td>
</tr>
<tr>
<td>Bradford</td>
<td>Median ± SD</td>
</tr>
<tr>
<td></td>
<td>CV %</td>
</tr>
<tr>
<td>Biuret</td>
<td>Median ± SD</td>
</tr>
<tr>
<td></td>
<td>CV %</td>
</tr>
</tbody>
</table>

BSA: bovine serum albumin; CV: coefficient of variation; HSA: human serum albumin; SD: standard deviation.

a BSA concentration for Biuret assay 0.5 mg/mL; Lowry, Bradford assays were 0.1 mg/mL.
b HSA concentration for Biuret assay 0.5 mg/mL; Lowry, Bradford assays were 0.05 mg/mL.

Source: adapted from Okutucu et al. (2007).

4.2.3.2 Spectroscopic methods for total protein content analysis

Spectroscopic methods for total protein content include ultraviolet (UV)-visible light spectrophotometry and near infrared reflectance (NIR) spectroscopy.

UV-visible light spectroscopy exploits the emission and absorption of UV-visible light and provides information on electronic interactions (Simonian, 2002; Noble & Bailey, 2009; Goldring, 2012; 2015). For protein measurement, UV-visible light spectroscopy measures the characteristic absorption of tyrosine and tryptophan at 280 nm (Warburg & Christian, 1942). The method is subject to many limitations because different proteins have different amounts of these amino acids, other molecules (including nucleic acids) also absorb at 280 nm, and different factors also interfere (Ayers et al., 1998). However, under controlled conditions the measurement of UV absorbance of protein extract solutions from milk, beef, flour, bean, and egg yolk were strongly correlated with Kjeldahl analysis (Beavers et al., 1973; Nakai & Le, 1970; Hambraeus et al., 1976).

Both NIR (0.7–2.7 μm or 14 285–4000 cm⁻¹) and mid-infrared reflectance (MIR) (2.5–25 μm or 4000–400 cm⁻¹) spectroscopy have been used for protein analysis (Bailon-Perez et al., 2007; Linker & Etzion, 2009; Moore et al., 2010; Barbano & Lynch, 2006; Diniz et al., 2011; de Marchi et al., 2014; Porep, Kammerer & Carle, 2015; López-Lorente & Mizaikoff, 2016; Grassi & Alamprese, 2018; Pasquini, 2018; Wiercigroch et al., 2017). The protein concentrations in fluid samples, including milk and soybean, are measured using multivariate analysis and calibration factors, by directly measuring transmitted radiation at 1550 cm⁻¹, which is characteristic of the peptide bond (Ribadeau-Dumas & Grappin, 1989) (Barbano & Lynch, 2006).

Despite various limitations, the NIR method has been applied to milk and dairy products (Moore et al., 2010; de Marchi et al., 2014; McDermott et al., 2017) and was found to be comparable with the Kjeldahl method (Luginbühl, 2002; Sorensen, Lund & Juul, 2003; Etzion et al., 2004; Bonfatti, Di Martino & Carnier, 2011). Measurement of NIR diffuse-reflectance is also used for routine protein analysis of grains (Wetzel, 1983). This method has also
been used to determine the content of fat, protein and total solids in cow's milk cheeses (Rodríguez-Otero, Hermida & Cepeda, 1995), and for protein measurement in soy products (Ferreira et al., 2014; Dong et al., 2018). However, NIR is a secondary technology requiring calibration with standards of known protein content; hence, its effectiveness depends on the quality of the reference material.

4.3 NPCFs for milk-based and soy-based foods

The values of the different CFs (K’, K_a and K_b) calculated for milk-based foods and soy-based foods as identified in the literature review are, respectively, listed in Table 15 and Table 16; the respective dispersions of data are shown in Fig. 3 and Fig. 4.

4.3.1 Conversion factors for dairy-based foods

Only two studies or data sets reported conversion factors that were assessed directly in formulas for infants and children (12 samples). A total of 24 studies or data sets reported data for dairy-based foods and ingredients (Table 15).

Table 15. NPCFs reported for dairy-based foods and food ingredients

<table>
<thead>
<tr>
<th>FOOD</th>
<th>K’</th>
<th>K_a</th>
<th>K_b</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cow's milk</td>
<td>6.38a</td>
<td>6.36b</td>
<td>-</td>
<td>Jones (1931)</td>
</tr>
<tr>
<td></td>
<td>6.32</td>
<td>6.01</td>
<td>5.99</td>
<td>Van Boekel and Dumas (1987)</td>
</tr>
<tr>
<td></td>
<td>6.13</td>
<td>5.82</td>
<td>6.08</td>
<td>Derham (1982)</td>
</tr>
<tr>
<td></td>
<td>6.07</td>
<td>5.79</td>
<td>5.97</td>
<td>Sosulski and Imafidon (1990)</td>
</tr>
<tr>
<td></td>
<td>5.92</td>
<td>5.66</td>
<td>5.83</td>
<td>Sosulski and Imafidon (1990)c</td>
</tr>
<tr>
<td></td>
<td>5.99</td>
<td>5.71</td>
<td>6.00</td>
<td>Featherston et al. (1964)c</td>
</tr>
<tr>
<td></td>
<td>5.99</td>
<td>5.73</td>
<td>5.85</td>
<td>Krizova et al. (2013)c</td>
</tr>
<tr>
<td></td>
<td>6.11</td>
<td>5.79</td>
<td>6.11</td>
<td>Featherston et al. (1964)c</td>
</tr>
<tr>
<td></td>
<td>6.09</td>
<td>5.80</td>
<td>6.10</td>
<td>Rutherfurd and Moughan (1998)c</td>
</tr>
<tr>
<td>Pure milk</td>
<td>6.15</td>
<td>5.83</td>
<td>6.04</td>
<td>Japan Vegetable Food Association</td>
</tr>
<tr>
<td>Milk protein isolate</td>
<td>6.13</td>
<td>5.86</td>
<td>6.10</td>
<td>Rutherfurd and Moughan (1998)c</td>
</tr>
<tr>
<td>Skim milk powder</td>
<td>6.13</td>
<td>5.86</td>
<td>6.10</td>
<td>Boisen et al. (1987)</td>
</tr>
<tr>
<td></td>
<td>6.09</td>
<td>5.80</td>
<td>6.10</td>
<td>Boisen et al. (1987)c</td>
</tr>
<tr>
<td>Casein</td>
<td>6.38a</td>
<td>-</td>
<td>-</td>
<td>Hammersten (1883)</td>
</tr>
<tr>
<td></td>
<td>6.37</td>
<td>6.08</td>
<td>-</td>
<td>Boisen et al. (1987)</td>
</tr>
<tr>
<td></td>
<td>6.16</td>
<td>5.87</td>
<td>6.01</td>
<td>Sosulski and Imafidon (1990)</td>
</tr>
<tr>
<td></td>
<td>6.15</td>
<td>5.88</td>
<td>6.03</td>
<td>Sarwar et al. (1993b)c</td>
</tr>
<tr>
<td></td>
<td>6.06</td>
<td>5.79</td>
<td>6.05</td>
<td>Tomotake et al. (2001)c</td>
</tr>
<tr>
<td>(Lactic casein)</td>
<td>6.38a</td>
<td>-</td>
<td>-</td>
<td>Rutherfurd and Moughan (1998)c</td>
</tr>
<tr>
<td>αs1-casein</td>
<td>6.36a</td>
<td>-</td>
<td>-</td>
<td>Van Boekel and Dumas (1987)</td>
</tr>
<tr>
<td>B-casein</td>
<td>6.37a</td>
<td>-</td>
<td>-</td>
<td>Van Boekel and Dumas (1987)</td>
</tr>
<tr>
<td>β-Lactoglobulin</td>
<td>6.29a</td>
<td>-</td>
<td>-</td>
<td>Maubois and Lorient (2016)</td>
</tr>
<tr>
<td>α-Lactalbumin</td>
<td>6.25a</td>
<td>-</td>
<td>-</td>
<td>Maubois and Lorient (2016)</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>6.07a</td>
<td>-</td>
<td>-</td>
<td>Maubois and Lorient (2016)</td>
</tr>
</tbody>
</table>
## 4. RESULTS

<table>
<thead>
<tr>
<th>FOOD</th>
<th>(K)(^{a})</th>
<th>(K)(^{b})</th>
<th>(K)(^{c})</th>
<th>(K)(^{d})</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant formula</td>
<td>6.38(^{b})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Maubois and Lorient (2016)</td>
</tr>
<tr>
<td>Child formula powder</td>
<td>6.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Nestec</td>
</tr>
<tr>
<td>Infant elemental powder</td>
<td>5.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adult nutritional RTF, high-protein</td>
<td>5.57</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adult nutritional RTF, high-fat</td>
<td>6.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula RTF, milk-based</td>
<td>6.15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>NIST SRM 1849(^{a})</td>
<td>6.16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula powder, partially hydrolysed, milk-based</td>
<td>6.08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Toddler formula powder, milk-based</td>
<td>6.11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula powder, milk-based</td>
<td>6.09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adult nutritional powder, low-fat</td>
<td>5.73</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Child formula powder</td>
<td>6.06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant elemental powder</td>
<td>6.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula powder, FOS/GOS-based</td>
<td>6.11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula powder, milk-based</td>
<td>6.14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Infant formula RTF, milk-based</td>
<td>6.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adult nutritional RTF, high-protein</td>
<td>5.57</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adult nutritional RTF, high-fat</td>
<td>6.18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Milk products</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.33</td>
<td>Salo-väänänen and Koivistoinen (1996)</td>
</tr>
<tr>
<td>Cheddar cheese</td>
<td>6.39(^{a})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Rouch et al. (2008)</td>
</tr>
</tbody>
</table>


\(^{a}\) Assumed amide/amid composition of 50% amide/50% acid; \(^{b}\) Assumed amide/amid composition of 75% amide/25% acid; \(^{c}\) Amide/amid composition that was directly assessed; \(^{d}\) Assumed no difference in nitrogen content between amide and acid forms (not adjusted) – values with parentheses are overestimated and not further used in the calculation.

\(^{a}\) Weight of the purified protein.

\(^{b}\) Sum of anhydrous amino acid residues plus weight of the prosthetic group when present.

\(^{c}\) Specific calculation for this review either by recalculation from previously published calculation, or by calculation using amino acid composition.
Fig. 3 Dispersion of calculated $K_A$ values for dairy protein


$K_A(a)$ assumed amide/acid composition of 50% amide/50% acid; $K_A(b)$ assumed amide/acid composition of 75% amide/25% acid; $K_A(c)$ amide/acid composition that was directly assessed.

Source: values from Table 15.
## 4. RESULTS

### 4.3.2 Conversion factors for soy-based foods

Only two data sets reported conversion factors that were assessed directly in formulas for infants and children (four samples). A total of 23 studies or data sets reported data for soy-based foods and food ingredients (Table 16).

<table>
<thead>
<tr>
<th>FOOD</th>
<th>(K_a)</th>
<th>(K_b)</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soybean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.71 (a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.65 (b)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.38 (c)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.63 (d)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.22 (e)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.77 (f)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.49 (g)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.51 (h)</td>
<td></td>
</tr>
<tr>
<td>Soybean meal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.69 (i)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.71 (j)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.65 (k)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.61 (l)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.77 (m)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.9 (n)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Soybean proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.64 (o)</td>
<td>5.36 (p)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.63 (q)</td>
<td></td>
</tr>
<tr>
<td>Soybean varieties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.70 (r)</td>
<td>5.65 (s)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.73 (t)</td>
<td>5.52 (u)</td>
<td></td>
</tr>
<tr>
<td>Soy protein isolate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.74 (v)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.69 (w)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.60 (x)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.65 (y)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.77 (z)</td>
<td></td>
</tr>
<tr>
<td>Defatted soy flakes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5.68 (aa)</td>
<td>5.39 (ab)</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>5.55 (ac)</td>
<td></td>
</tr>
<tr>
<td>Acid precipitate soy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>isolate</td>
<td>-</td>
<td>5.39 (ad)</td>
<td></td>
</tr>
<tr>
<td>Dialysis soy isolate</td>
<td>-</td>
<td>5.68 (ae)</td>
<td></td>
</tr>
<tr>
<td>Commercial isolate</td>
<td>-</td>
<td>5.66 (af)</td>
<td></td>
</tr>
<tr>
<td>Soy protein isolate</td>
<td>-</td>
<td>5.60 (ag)</td>
<td></td>
</tr>
<tr>
<td>Soy protein concentrate</td>
<td>-</td>
<td>5.70 (ah)</td>
<td></td>
</tr>
<tr>
<td>Soy protein isolate</td>
<td>-</td>
<td>5.68 (ai)</td>
<td></td>
</tr>
<tr>
<td>Soy cotyledons</td>
<td>-</td>
<td>5.65 (aj)</td>
<td></td>
</tr>
<tr>
<td>Soy isolate</td>
<td>-</td>
<td>5.66 (ak)</td>
<td></td>
</tr>
<tr>
<td>Soy concentrate</td>
<td>-</td>
<td>5.65 (al)</td>
<td></td>
</tr>
<tr>
<td>Soy protein isolate</td>
<td>-</td>
<td>5.67 (am)</td>
<td></td>
</tr>
<tr>
<td>Soy protein concentrate</td>
<td>-</td>
<td>5.67 (an)</td>
<td></td>
</tr>
<tr>
<td>Soy flakes</td>
<td>-</td>
<td>5.65 (ao)</td>
<td></td>
</tr>
</tbody>
</table>
4.3.3 Pooled estimates for NPCFs

The results of pooling the data (separately for dairy and soy) are summarized in Table 17. As expected, the observed values for both dairy-based foods and soy-based foods were ordered in the following way: $K' > K_a > K_b$; however, only a few studies were identified reporting $K'$ for soy (N=1), $K'$ for dairy (N=5 with 10 $K'$ values reported) and $K_b$ for dairy (N=3). Also, with respect to amide to acid ratios, values for mean $K_a(a) = 50/50$ and $K_b$ with direct adjustment were similar, with $K_a 75/25$ being slightly higher than within both dairy and soy assessments.

Only two studies or data sets reported conversion factors that were assessed directly in milk-based formulas for infants and children (12 samples), with a mean value of 6.08 (95% CI: 6.05, 6.12) for $K_a 50/50$ and a single $K'$ of 6.32. Mean values across all dairy foods ranged from 5.55 to 6.32, depending on how the CF was calculated.

Only two data sets reported conversion factors that were assessed directly in soy-based formulas for infants and children (four samples), with a mean value of $K_a 50/50$ of 5.70 (95% CI: 5.69, 5.71), of $K_a 75/25$ of 5.42, and of $K_b$ direct adjustment of 5.42 (95% CI: 5.40, 5.44). Mean values across all dairy foods ranged from 5.34 to 5.71, depending on how the conversion factor was calculated.
4. RESULTS

**Fig. 4 Dispersion of calculated KA values for soy protein**

| Source: values from Table 16. |

- **Dispersion of KA(a) values for soy protein**
  - $K_A(a)$ assumed amide/amide composition of 50% amide/50% acid;
  - $K_A(b)$ assumed amide/amide composition of 75% amide/25% acid;
  - $K_A(c)$ amide/amide composition that was directly assessed.

**Infant formula soy-based**

**Soybean varieties**

**Soy milk**

**Soy protein concentrate**

**Soy protein isolate**

**Soybean**

**Soy meal**
Table 17. Pooled estimates for NPCFs

<table>
<thead>
<tr>
<th></th>
<th>DAIRY MEAN (95% CI)</th>
<th>DAIRY RANGE</th>
<th>SOY MEAN (95% CI)</th>
<th>SOY RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FORMULAS ONLY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K'_)</td>
<td>6.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K_{A} 50/50)</td>
<td>6.08 (6.05, 6.12)</td>
<td>5.97–6.17</td>
<td>5.70 (5.69, 5.71)</td>
<td>5.69–5.72</td>
</tr>
<tr>
<td>(K_{A} 75/25)</td>
<td>–</td>
<td></td>
<td>5.42a</td>
<td></td>
</tr>
<tr>
<td>(K_{A} ) direct adjust</td>
<td>–</td>
<td></td>
<td>5.42 (5.40, 5.44)</td>
<td>5.41–5.43</td>
</tr>
<tr>
<td><strong>ALL SAMPLES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(K'_)</td>
<td>6.32 (6.26, 6.38)</td>
<td>6.07–6.39</td>
<td>5.71b</td>
<td></td>
</tr>
<tr>
<td>(K_{A} 50/50)</td>
<td>6.06 (6.00, 6.12)</td>
<td>5.57–6.37</td>
<td>5.68 (5.66, 5.69)</td>
<td>5.60–5.80</td>
</tr>
<tr>
<td>(K_{A} 75/25)</td>
<td>5.83 (5.77, 5.89)</td>
<td>5.66–6.08</td>
<td>5.40 (5.38, 5.42)</td>
<td>5.33–5.52</td>
</tr>
<tr>
<td>(K_{A} ) direct adjust</td>
<td>6.03 (5.98, 6.07)</td>
<td>5.83–6.15</td>
<td>5.65 (5.61, 5.68)</td>
<td>5.37–5.80</td>
</tr>
<tr>
<td>(K_{B} )</td>
<td>5.55 (5.31, 5.78)</td>
<td>5.33–5.75</td>
<td>5.35 (5.20, 5.51)</td>
<td>4.78–5.90</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPCF: nitrogen to protein conversion factor.

a Two measurements, both with a value of 5.42.

b Single measurement only. Regarding \(K'_\) and inclusion of prosthetic groups for soy, an additional CF with a value of 5.91 was calculated specifically for the soy 7S protein \(\beta\)-conglycinin (Maubois & Lorient, 2016), based on the assumption that all three subunits of \(\beta\)-conglycinin are glycosylated. The authors further used this information to estimate factors for total soy proteins with different 11S/7S ratios, in the range 5.69–5.79. These values were not included in the final analysis because they were not directly measured but were estimated, based on assumptions made in reports in the literature.

4.3.4 Certainty of the evidence

The certainty in the evidence for NPCFs as assessed by GRADE ranged from very low to moderate. The rating depended on whether protein was defined as amino acids only or amino acids plus prosthetic groups for values of \(K_{A}\) with directly adjusted amide to acid ratios, for both dairy-based foods and soy-based foods; \(K_{B}\) was also assessed as moderate for soy-based foods. Details of the assessments for each of the conversion factors, including rationales for decision-making on each assessment element (i.e. risk of bias, inconsistency, indirectness and imprecision) can be found in the GRADE evidence profiles in Annex 3.

4.4 Modelling of non-protein nitrogen

Results of modelling the impact of non-protein nitrogen on published and calculated \(K_{A}\) values for both dairy-based foods and soy-based foods showed a linear and downward trend, as expected, in modelled \(K_{B}\) values, regardless of the assumptions made for calculating \(K_{A}\) in terms of amide to acid ratio (i.e. 50/50, 75/25 or directly adjusted), and depending on the percentage of non-protein nitrogen entered into the equation (Fig. 5). Beginning with 2% non-protein nitrogen, there was a significant difference between \(K_{B}\) and \(K_{A}\) values, increasing with increasing amounts of non-protein nitrogen.

Modelling results for dairy and soy are shown together in Fig. 6. The downward trend of \(K_{B}\) was similar for the two product categories, but in the case of milk the impact was slightly more important for a non-protein nitrogen percentage of 10%. The highest variations for \(K_{B}\) were those calculated from a \(K_{A}\) value with 75% amide and 25% acid. The mean value of \(K_{B}\) for milk-based infant formula, as listed in Table 15, was 6.09. For this value of \(K_{A}\), when non-

---

1 Individual values of 6.05, 5.97, 6.15, 6.08, 6.11, 6.09, 6.00, 6.11, 6.14 and 6.17.
4. RESULTS

Fig. 5 Relationship between KA and KB according to non-protein nitrogen level for milk-based (A) and soy-based (B) products

A. Milk-based products

B. Soy-based products

NPN: non-protein nitrogen.
Source: KA values for milk-based products and soy-based products taken directly from Table 15 and Table 16, respectively.
protein nitrogen varied from 0 to 10% total nitrogen, KB varied from 6.09 to 5.40 (Table 18). The mean value of $K_A$ for soy-based infant formula, as listed in Table 16, was 5.70.¹ For this value of $K_A$, when non-protein nitrogen varied from 0 to 10% total nitrogen, KB varied from 5.70 to 5.1 (Table 18).

### Table 18. KB values according to non-protein nitrogen content as a function of mean KA values

<table>
<thead>
<tr>
<th></th>
<th>MILK-BASED INFANT FORMULA</th>
<th>SOY-BASED INFANT FORMULA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $K_A = K_B$ for 0% NPN</td>
<td>6.09</td>
<td>5.7</td>
</tr>
<tr>
<td>$K_B$ for 5% NPN total N</td>
<td>5.7</td>
<td>5.4</td>
</tr>
<tr>
<td>$K_B$ for 10% NPN total N</td>
<td>5.4</td>
<td>5.1</td>
</tr>
<tr>
<td>Range for $K_B$ in Tables 15 and 16</td>
<td>5.3–5.7</td>
<td>4.8–5.9</td>
</tr>
</tbody>
</table>

N: nitrogen; NPN: non-protein nitrogen.

¹ Based on protein weight calculated from the sum of the weights of anhydrous amino acid residues.

¹ Individual values of 5.69, 5.69, 5.71 and 5.72.
5. DISCUSSION

Protein content in food products and food ingredients can be determined either by direct protein analysis or by measuring nitrogen content and using an NPCF (K’, K_A and K_B). The direct measurement of protein can be performed by purification of protein, colorimetric methods, spectrophotometric methods or infrared spectroscopy. These different methods have some limitations, and measuring nitrogen content remains the more frequently used approach for protein content in food products and food ingredients.

5.1 General considerations in calculating and NCPFs

Values for the NPCF K’ were initially proposed based on measurement of total nitrogen content and weight of the different purified proteins (Jones, 1931). Even early on, the results indicated that using a generic conversion factor of 6.25 could lead to a 15–20% error in estimating the actual protein content, and that specific CFs should be applied for the determination of protein content when calculated from the nitrogen content of different food products and food ingredients. Methods for deriving CFs were subsequently improved; for example, basing protein weight on amino acid composition provided more precise values for protein molecular weight and protein nitrogen content (Heathcote, 1950; Tkachuk, 1966b; 1966a; Tkachuk & Mellish, 1977; Sosulski & Holt, 1980; Morr, 1981; 1982; Mossé, 1990).

Although methods have improved, the results of this review and others highlight two critical and limiting considerations in determining appropriate conversion factors for dairy, soy and other proteins, particularly where the purpose of the conversion factor is to provide information primarily on the provision of amino acids and will be used on a large variety of different foods. The first consideration applies to protein with identified prosthetic groups (primarily milk protein), and is whether to include the weight of the prosthetic groups in the calculation of the molecular weight of the protein (Maubois & Lorient, 2016; IDF, 2016). The second consideration applies to food products and ingredients with a significant amount of non-protein nitrogen, and is whether to use K_A (which requires knowing how much protein nitrogen is contained in the sample via non-protein nitrogen measurement and subtracting this fraction to total nitrogen, or by amino acid analysis) or K_B based on total nitrogen content. In addition, the amide to acid ratio must be taken into consideration when calculating and using K_A.

5.1.1 Prosthetic groups

This review found significant differences between K’ and K_A values, resulting from the presence of prosthetic groups on most dairy proteins.

The mean K’ for dairy-based foods is 6.32 (95% CI: 5.67, 6.98) with a range of 6.28–6.39, whereas K_A and K_B, which are based on weights of amino acid residues only, range between 5.33 and 6.15, with means of 5.55–6.11. The K’ values identified in this review were obtained either based on the total nitrogen content and weight of purified acid-precipitated bovine casein or milk protein (Hammersten, 1883; Jones, 1931), or from more recent calculations from amino acid composition and sequence (by adding the weight of the identified prosthetic groups to the weight of the sum of anhydrous amino acids residues) (Van Boekel & Dumas, 1987; Rouch, Roupas & Roginski, 2007; Rouch et al., 2008; Maubois & Lorient, 2016; Karman & van Boekel, 1986).
The presence of prosthetic groups on dairy proteins has been noted in the literature, and can contribute significantly to the measured mass. For example, compared with the weight of proteins assessed as the sum of anhydrous amino acids only, complete phosphorylation of α-casein – which accounts for about half of all dairy protein – can increase the weight by nearly 5% (O’Donnell et al., 2004), and the presence of glycans in lactoferrin can increase its weight by more than 11% (O’Riordan et al., 2014). Total milk protein and casein (which accounts for about 80% of milk protein) have K’ values in the range of 6.34–6.39 owing to prosthetic groups, whereas whey proteins have a lower K’ value (6.29–6.07), probably due to a lower content of prosthetic groups (Maubois & Lorient, 2016).

Although many of the sites for post-translational addition of prosthetic groups in dairy proteins have been characterized, modification at any particular site can vary from sample to sample, making it difficult to precisely assess the contribution of prosthetic groups to the weight of protein in a particular food sample or ingredient. Moreover, for other protein sources, including soy-based food and ingredients, post-translational modifications have not been well characterized, and only KA or KB can be calculated from the amino acid composition of the proteins. For soy, the only K’ value identified in the literature review (5.71) was initially calculated based on the measurement of total nitrogen content and weight of purified soybean protein (Jones, 1931). Therefore, until more data are generated for other foods and ingredients, K’ is really only relevant in the context of assessing dairy proteins.

5.1.2 Non-protein nitrogen

Kₐ values based on protein nitrogen content only are the most widely assessed and published conversion factors (in this review, more than 30 Kₐ values were identified for both dairy-based and soy-based foods). However, for food products and food ingredients, all but the most highly purified protein samples contain some level of non-protein nitrogen. Therefore, the ability to correct the CF for the presence of non-protein nitrogen in food products and food ingredients is critical when calculating and using conversion factors. Although Kₐ is the only conversion factor that explicitly takes non-protein nitrogen into consideration, and therefore may provide a truer estimate of protein content, it has proved difficult to precisely define the appropriate value of Kₐ for most food sources (Heathcote, 1950; Holt & Sosulski, 1979; Morr, 1981; 1982; Sosulski & Imafiodon, 1990; Tkachuk, 1969; 1966a; Mossé, 1990). However, to the extent that non-protein nitrogen varies from batch to batch or protein type to protein type for a given class of protein, the factor Kₐ will introduce inaccuracies.

In this review, only three studies were identified that reported KB values for cow’s milk, skim milk powder and various dairy products (Boisen et al., 1987; Fujihara et al., 2010; Saloväänänen & Koivistoinen, 1996). In milk powders, non-protein nitrogen was measured at a level of about 0.3% weight for weight (DeVries et al., 2017). A greater number of studies were identified that reported on KB values for soy-based foods. Even without explicit knowledge about the amount of non-protein nitrogen in a food or ingredient, correcting for the content of non-protein nitrogen in these samples by using the factor Kₐ (and the total nitrogen content of the sample) provides a reasonable estimate of the protein content, particularly in those food

---

1 Regarding K’ and inclusion of prosthetic groups for soy, an additional conversion factor with a value of 5.91 was calculated specifically for the soy 7S protein β-conglycinin (Maubois & Lorient, 2016), based on the assumption that all three subunits of β-conglycinin are glycosylated. The authors further used this information to estimate factors for total soy proteins with different 11S/7S ratios, in the range 5.69–5.79. These values were not included in the final analysis because they were not directly measured but were estimated, based on assumptions made in reports in the literature.
products or food ingredients with a significant level of non-protein nitrogen (Templeton & Laurens, 2015; Ezeagu et al., 2002; Fujihara et al., 2010; Jonas-Levi & Martinez, 2017; Mattila et al., 2002; Yeoh & Truong, 1996a; Yeoh & Wee, 1994).

The modelling results presented in this review further support the notion that non-protein nitrogen can have a significant impact on the calculation and subsequent application of conversion factors, particularly when there are significant levels of non-protein nitrogen. Modelling results suggest that for a sample with 10% of total nitrogen content as non-protein nitrogen, the conversion factor can be more than 10% higher when the non-protein nitrogen is not taken into consideration.

5.1.3 Amide to acid ratio

When calculating and using $K_a$ values, nitrogen content must be adjusted for the amide to acid ratio between glutamine and glutamic acid, and between asparagine and aspartic acid. Ways of addressing this include arbitrarily adjusting for amide to acid ratios of 50/50 or 75/25, adjusting based on direct assessment of amide to acid ratio, or ignoring the impact of the amide to acid ratio and not adjusting at all. This review found that for both dairy-based and soy-based food products and food ingredients, values of $K_a$ in which no amide to acid adjustment were made were significantly larger than those where adjustments were made. Because these are generally gross overestimates, they are the least informative of the $K_a$ values. The value obtained using a fixed amide to acid ratio of 50/50 were higher than when using a ratio of 75/25.

The $K_a$ values calculated for cow’s milk, skim milk powder and casein were in the range 5.99–6.32 when using a fixed amide to acid ratio of 50/50, 5.66–6.08 when using a fixed amide to acid ratio of 75/25, and 5.99–6.15 after direct amide to acid adjustment. However, the only $K_a$ value of 6.32 obtained using a fixed amide to acid ratio of 50/50 (Derham, 1982) was clearly above the other values obtained using the same ratio; if this value is not taken into account, the $K_a$ values calculated for cow’s milk, skim milk powder and casein using a fixed amide to acid ratio of 50/50 were in the range 5.99–6.13 (i.e. very close to the values using the directly measured ratio), whereas lower results were obtained by considering an amide to acid ratio of 75/25. Similarly, results obtained for soy-based foods showed that those values with an amide to acid ratio of 50/50 were close to the values using the directly measured ratio, whereas lower values were obtained by considering an amide to acid ratio of 75/25. Values were also overestimated when the amide to acid ratio was not adjusted, whereas a $K_a$ value of 5.65 was calculated for soybean meal after direct adjustment of the amide to acid ratio (Boisen et al., 1987). Overall, this suggests that when the amide to acid ratio is not directly measured or otherwise known, the estimated ratio of 50/50 might provide a reasonably accurate estimate for both dairy-based and soy-based foods and food ingredients.

5.2 Selecting NPCFs

Selecting one or more conversion factors for dairy and soy-based ingredients to be used as benchmarks across a large number of different samples (in this case infant formula and follow-up formula, comprising different formulations from different manufacturers) requires consideration of both the type of CF to use (i.e. $K'$, $K_a$ or $K_b$) and the specific value of the CF. The choice of whether to use $K'$, $K_a$ or $K_b$ in the case of dairy, or $K_a$ or $K_b$ in the case of soy, depends on both the purpose of using the conversion factor (e.g. is it primarily to assess delivery of amino acids?) and consideration of any potential variation in non-protein nitrogen
content across many different formulations and compositions of infant formula and follow-
up formula.

Regarding the purpose of the conversion factor, if the interest is primarily in quantifying the
amount of amino acids in a sample, then $K_A$ or $K_B$ are more relevant because they provide
information on amino acid content only. If the interest is in determining the total protein
content including prosthetic groups, then $K’$ would be more informative; however, it is
essentially only available for dairy proteins. Regarding variation in non-protein content, use
of a particular $K_B$ assumes that non-protein nitrogen content of the sample used to derive
the particular $K_B$ selected is the same or very similar in all samples that will be assessed
using the particular $K_B$. Otherwise, it would be more appropriate to determine the non-
protein nitrogen by subtracting it from the total nitrogen content and applying $K_A$ or $K’$ to the
difference.

5.3 Comparability with other reviews

Other reviews using mainly a narrative approach have examined the values of NPCFs for
the calculation of protein content in foods (Mariotti et al., 2008; Maubois & Lorient, 2016;
Krul, 2019). The methodological approaches and exclusion criteria used in those reviews
differed substantially from our criteria in the present study. However, the conclusions of
these different reviews are in good agreement with the present conclusion related to the
values calculated for the different conversion factor $K’$, $K_A$ and $K_B$, and with the significance
of the differences between these different CFs. All the reviews considered that the more
accurate approach is to first derive the weight and nitrogen content of the different proteins
from their amino acid composition. In the review from Maubois and Lorient (2016), adding the
weight of the prosthetic group to calculate the weight of milk protein was preferred, but the
values of $K’$ agreed with other results from references cited here. There was also a general
conclusion that the main uncertainty arises from non-protein nitrogen and calculation of $K_B$.

5.4 Comparability with other conversion factors

The results on conversion factors obtained for milk-based and soy-based foods and food
ingredients agree with those obtained for other food products. For comparison, different
results on other food products also showed that the $K_A$ values are for many products in
the range 5–6, and that $K_A > K_B$ as expected. For instance, for meat, the value of 6.25 for $K’$
was clearly overestimated: a $K_A$ conversion factor of about 5.7 is more appropriate and a
$K_B$ of 5.17 was proposed (Jones, 1931; Holt & Sosulski, 1979; Salo-väänänen & Koivistoinen,
1996; Rafecas et al., 1994). For wheat, $K’$ was 5.83 and the measured $K_A$ and $K_B$ values were
in the ranges 5.14–5.93 and 5.18–5.55, respectively (Jones, 1931; Sosulski & Sarwar, 1973;
Mossé, Huet & Baudet, 1985; Fujihara et al., 2008; Liu et al., 2018). For rice, $K’$ was 5.93
and the measured $K_A$ and $K_B$ values were in the ranges 5.61–5.78 and 4.9–5.35, respectively
(Jones, 1931; Sosulski & Sarwar, 1973; Mossé, Huet & Baudet, 1988; Fujihara et al., 2008).
For seaweed, $K_B$ was in the range 3.53–5.72 (Angell et al., 2016; Biancarosa et al., 2017; Diniz et
al., 2011; Lourenço et al., 2002). For *Palmaria palmata*, $K_B$ was 4.7 (Bjarnadottir et al., 2018).
For cassava root, $K_A$ was in the range 4.75–5.87 and $K_B$ was 3.24 (Yeoh & Truong, 1996b). For
mushrooms, $K_B$ was in the range 4.5–4.97 (Mattila et al., 2002). For crop residues, $K_A$ was
5.42–6.00 and $K_B$ 3.97–4.57; for animal manure, $K_A$ and $K_B$ were in the ranges 4.78–5.36 and
3.97–4.57, respectively (Chen et al., 2017).
6. CONCLUSION

From the limited number of studies and data sets identified that specifically calculated conversion factors for infant formula or follow-up formula, the following mean values were observed: for milk-based formulas, a mean value of 6.08 (95% CI: 6.05, 6.12; moderate certainty of evidence) for KA 50/50 and a single K’ value of 6.38 (very low certainty of evidence); and for soy-based formulas, mean values of 5.70 (95% CI: 5.69, 5.71; very low certainty of evidence) for KA 50/50, 5.42 (very low certainty of evidence) for KA 75/25, and 5.42 (95% CI: 5.40, 5.44; very low certainty of evidence) for KA direct adjustment. This review also showed that when broadly considering dairy proteins inclusive of prosthetic groups, mean K’ was 6.32 (95% CI: 6.26, 6.38; low certainty of evidence). The results further suggested that when considering amino acids alone, without prosthetic groups, and when amino acid composition of the protein source is available (as it is for milk and soy protein), KA can be accurately calculated with a value in the ranges 5.83–6.06 for milk protein (low to moderate certainty of evidence) and 5.40–5.68 for soy protein (low certainty of evidence). Because these values do not take into consideration non-protein nitrogen, however, they should be considered as the maximum values for milk-based and soy-based foods. In other words, when non-protein nitrogen is considered, KA will likely be less than or equal to 5.83–6.06 for dairy-based foods and food ingredients, and less than or equal to 5.40–5.68 for soy-based foods and food ingredients. Indeed, we observed a mean KA of 5.35 (95% CI: 5.20, 5.51) for soy-based foods and food ingredients, with low certainty of evidence, from a fairly large number of studies. KA can, however, be applied to total nitrogen values corrected for non-protein nitrogen. There were too few values reported for dairy foods to have much confidence in the observed mean KA of 5.55 (95% CI: 5.31, 5.78; very low certainty of evidence) for dairy. In addition, as presented in this review, alternative methods for protein content should be discussed, including direct protein analysis or direct amino acid analysis. These methods have been improved over time, and direct amino acid analysis could represent an interesting alternative in the future.

---

1 Depending on assumptions made regarding definition of protein.
ANNEXES
ANNEX 1.

SEARCH STRATEGY

Table A.1  Keyword combinations used in the database searches of publications (1946–2019)

<table>
<thead>
<tr>
<th>KEYWORDS COMBINATION</th>
<th>NUMBER OF PUBLICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&quot;amino acid* composition*&quot; OR &quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;food*&quot; OR &quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean*&quot;)</td>
<td>(1) 1892  1340  &gt;4000a  &gt;4000a</td>
</tr>
<tr>
<td>(&quot;amino acid* composition*&quot; OR &quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot;)</td>
<td>(2) 1009  577  1874  1244</td>
</tr>
<tr>
<td>(&quot;amino acid* analysis&quot; OR &quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;food*&quot; OR &quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot;)</td>
<td>(3) 480  533  &gt;900a  &gt;900a</td>
</tr>
<tr>
<td>(&quot;amino acid* analysis&quot; OR &quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot;)</td>
<td>(4) 247  219  421  286</td>
</tr>
<tr>
<td>(&quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;food*&quot; OR &quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot; OR &quot;Protein&quot;)</td>
<td>(1) 54  252  44  36</td>
</tr>
<tr>
<td>(&quot;nitrogen to protein conversion factor*&quot; OR &quot;nitrogen conversion factor*&quot;) AND (&quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot; OR &quot;Protein&quot;)</td>
<td>(2) 54  212  44  35</td>
</tr>
<tr>
<td>(&quot;nitrogen analysis&quot;) AND (&quot;food*&quot; OR &quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot;)</td>
<td>(3) 45  26  67  27</td>
</tr>
<tr>
<td>(&quot;nitrogen analysis&quot;) AND (&quot;milk&quot; OR &quot;dairy&quot; OR &quot;soy*&quot; OR &quot;soybean&quot;)</td>
<td>(4) 29  16  38  12</td>
</tr>
<tr>
<td>Total number of papers from each databaseb</td>
<td>1289  900  2298  1465</td>
</tr>
<tr>
<td>Final corpusb</td>
<td>3881</td>
</tr>
</tbody>
</table>

(1) “Web of Science Core Collection – Clarivate Analytics” and using the field “All fields” for (2) “Medline/PubMed of the National Library of Medicine”, (3) “CAB Direct/CAB Abstracts – CABI”, and (4) “Food Science and Technology Abstracts – Ebsco/IFIS”.

a Not included in the final corpus and discarded due to background noise.

b Total after removing duplicates within and between databases.

The keywords “food” and “protein”, and to a lesser extent “amino acid composition” “amino acid analysis”, produced significant background noise.

From the corpus of the 3881 publications, a series of requests relating to amino acid, nitrogen and conversion factor was constructed (Table A.2).
Table A.2  Requests from the corpus of the 3881 publications (1946–2019)

<table>
<thead>
<tr>
<th>REQUESTS ON THE WORDS OF THE TITLE, THE SUMMARY OR THE KEYWORDS OF THE PUBLICATION</th>
<th>MILK OR DAIRY</th>
<th>SOY</th>
<th>INFANT FORMULA</th>
<th>TOTALa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid (analysis OR content OR composition)</td>
<td>694</td>
<td>610</td>
<td>13</td>
<td>1274</td>
</tr>
<tr>
<td>Nitrogen (analysis OR content OR determination)</td>
<td>34</td>
<td>20</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>Conversion factor</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Totala</td>
<td>714</td>
<td>619</td>
<td>14</td>
<td>1316</td>
</tr>
</tbody>
</table>

Table A.3  Additional requests from the corpus of the 3881 publications (1946–2019)

An additional series of requests made on the methods for nitrogen and protein analysis, and another on publications compiled in the ISI Web of Knowledge on the words in the title with “analytical methods”, “Kjeldahl” and “Dumas method” (1956–2019) or with other sources (e.g. Google Scholar, Science Direct, Wiley) provided 392 publications (Table A.3).

Table A.3  Additional requests from the corpus of the 3881 publications (1946–2019)

<table>
<thead>
<tr>
<th>REQUEST ON THE WORDS OF THE TITLE, THE SUMMARY OR THE KEYWORDS OF THE PUBLICATION</th>
<th>KJELDHAL</th>
<th>DUMAS</th>
<th>TOTALa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total initial corpus</td>
<td>65</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>Milk OR Dairy</td>
<td>13</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Soy</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Totala</td>
<td>65</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>(i) Selected publications from the corpus</td>
<td>33</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>(ii) Additional publications</td>
<td>27</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>(i)+(ii) Totala</td>
<td>78*</td>
<td>18*</td>
<td>85*</td>
</tr>
</tbody>
</table>

Domains Web of Knowledge

<table>
<thead>
<tr>
<th></th>
<th>Kjeldhal</th>
<th>Dumas</th>
<th>Total*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food sciences and technology</td>
<td>374</td>
<td>45</td>
<td>392</td>
</tr>
<tr>
<td>Agronomy</td>
<td>206</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Agricultural engineering</td>
<td>162</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Plant sciences</td>
<td>146</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Selected publications for analytical methods

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>215</td>
</tr>
</tbody>
</table>

* Total after removing duplicates within and between databases.
Exclusion of studies was based on analysis of the words from either title, keywords, abstracts or from full text screening.

The reasons for exclusion of studies included the following:

- the studies or the articles are completely outside the scope of the review;
- the studies or the articles refer to NPCFs, for instance in the experimental part, but do not contribute or provide data that can be used in the review;
- the studies or the articles refer to protein analysis but do not contribute or provide data that can be used in the review;
- the studies or the articles refer to amino acid analysis but do not contribute or provide data that can be used in the review;
- the studies or the articles use an NPCF (most often of 6.25) but do not address the calculation and validation of this factor;
- the studies or the articles provide data previously published; and
- some articles were already cited and do not provide additional information.

Regarding studies reporting empirically determined values for NPCFs, specific studies were excluded from this review for the reasons set out below.

**Articles addressed products other than milk and soy or were not relevant to the objective of the review. Example articles:**

Baudet et al. (1986) J Agric Food Chem. 34:365–70  
Estrella (2008) Quito Usfq. 5:1–26  
Colwell et al. (2011) J Assoc Publ Analysts. 39:4–78  
Articles reported an NPCF of 6.25 or another fixed value but did not address the calculation. Example articles:

Roberts & Briggs (1965) Cereal Chem. 42:71
Koshiyama (1968) Cereal Chem. 45:405
Andres et al. (2013) J Pediatr. 163:49–54
Pivik et al. (2013) Intl J Psychophophysiol. 90:311–20
Vandenplas et al. (2014) Br J Nutr. 111:1340–60
Elgar et al. (2016) J AOAC Intl. 99:26–9

Articles were already cited by Maubois and Lorient (2016) for the values of non-protein nitrogen. Example articles:

**GRADE evidence profile 1 – Dairy-based ingredients**

**Question:** When using the equation \( \text{amount of protein (P)} = \text{nitrogen to protein conversion factor (K)} \times \text{amount of nitrogen (N)} \) to estimate the protein content of dairy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content only?

**Population:** Infant formula and follow-up formula

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor K´ (unitless) – all dairy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Mixed</td>
<td>Not serious</td>
<td>Very serious</td>
<td>Not serious</td>
<td>None</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor K A, with amide/amid ratio of 50/50 (unitless) – all dairy foods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>None</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor K A, with amide/amid ratio of 75/25 (unitless) – all dairy foods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Serious</td>
<td>Not serious</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor K B (unitless) – all dairy foods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>None</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor K´ (unitless) – formulas only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Amino acid sequencing</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>None</td>
</tr>
<tr>
<td>No. of studies</td>
<td>Study design</td>
<td>Risk of bias</td>
<td>Inconsistency</td>
<td>Indirectness</td>
<td>Imprecision</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>Amino acid analysis&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Not serious&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

CI: confidence interval.
<sup>1</sup> Importance for decision-making. In this case, conversion factors are the only “outcome” and thus each of the variations is critical for decision-making.
<sup>2</sup> Amino acid sequencing and measurement (weighing) of crude protein with total nitrogen analysis. Because this factor was derived primarily from studies using amino acid sequencing, it was started at “high” certainty.
<sup>3</sup> There was no indication of systematic measurement error or reporting error (i.e. selective reporting of \( K \) values). Not downgraded.
<sup>4</sup> Inconsistency was assessed by considering the level of variance around the mean. The 95% CI suggests very little variation around the mean. Not downgraded.
<sup>5</sup> Only a single \( K \) value for infant formula and follow-up formula was identified in the literature review. The mean conversion factor was therefore primarily derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula.
<sup>6</sup> In addition, calculation of \( K \) includes prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids only. Downgraded twice for very serious indirectness.
<sup>7</sup> The minimum of 10 measurements is satisfied. Not downgraded.
<sup>8</sup> Of the 31 measurements, 11 come from infant formula and follow-up formula. Although \( K_a \) 50/50 and 75/25 values are based on setting the ratio of amide to acid arbitrarily at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike \( K_n \), \( K_a \) values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. There was no indication of reporting error (i.e. selective reporting of \( K_a \) values). Not downgraded.
<sup>9</sup> In addition to having greater confidence in the \( K_a \) 50/50 value relative to \( K_a \) 75/25, because the former includes a significant number of studies that derived conversion factors directly from infant formula and follow-up formula (and therefore was not downgraded for serious indirectness), there was greater confidence in this value because it is in line with that of the \( K_n \) derived from the directly adjusted amide to acid ratio, which is the most accurate method of assessing amide to acid ratio.
<sup>10</sup> No studies reported values for infant formula and follow-up formula. The mean conversion factor was therefore derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Although \( K_a \) 50/50 and 75/25 values are based on setting the ratio of amide to acid arbitrarily at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike \( K_n \), \( K_a \) values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Downgraded once for serious indirectness.
<sup>11</sup> No studies reported values for infant formula and follow-up formula. The mean conversion factor was therefore derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, unlike \( K_n \), \( K_a \) values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration to estimate protein content in a variety of formulas with different formulations. Downgraded once for serious indirectness.
12 No studies reported values for infant formula and follow-up formula. The mean conversion factor was therefore derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Downgraded once for serious indirectness.

13 Fewer than 5 studies contributing to the mean. Downgraded twice for very serious imprecision.

14 Because this factor was derived from a study using amino acid sequencing, it was started at “high” certainty.

15 Conversion factors were derived from infant formula and follow-up formula. However, calculation of $K'$ includes prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids only. Downgraded once for serious indirectness.

16 Conversion factors were derived from infant formula and follow-up formula. Although $K_a$ 50/50 and 75/25 values are based on setting the ratio of amide to acid arbitrarily at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike $K_p$, $K_a$ values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. There was no indication of reporting error (i.e. selective reporting of $K_a$ values). Not downgraded.

Annex 4 provides information on which studies provided data for each mean conversion factor shown in GRADE evidence profile 1 above.
**GRADE evidence profile 2 – Soy-based ingredients**

**Question:** When using the equation \( \text{amount of protein (P)} = \text{nitrogen to protein conversion factor (K)} \times \text{amount of nitrogen (N)} \) to estimate the protein content of soy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content only?

**Population:** Infant formula and follow-up formula

| Nitrogen to protein conversion factor \( K \) – all soy foods | Quality assessment | No. of independent measurements | Mean (95% CI) | Certainty | Importance
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with amide/amid ratio of 50/50 (unitless) – all soy foods</td>
<td>Not serious(^4) Not serious(^3) Very serious(^6) Very serious(^7) None</td>
<td>1</td>
<td>5.71 (single measurement)(^2)</td>
<td>VERY LOW</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with amide/amid ratio of 75/25 (unitless) – all soy foods</td>
<td>Not serious(^4) Not serious(^3) Serious(^5) Not serious(^8) None</td>
<td>28</td>
<td>5.68 (5.66, 5.69)</td>
<td>LOW</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with directly calculated amide/amid ratio (unitless) – all soy foods</td>
<td>Not serious(^4) Not serious(^3) Serious(^5) Not serious(^8) None</td>
<td>35</td>
<td>5.65 (5.61, 5.68)</td>
<td>LOW</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with amide/amid ratio of 50/50 (unitless) – formulas only</td>
<td>Not serious(^4) Not serious(^3) Not serious(^5) Very serious(^15) None</td>
<td>4</td>
<td>5.70 (5.69, 5.71)</td>
<td>VERY LOW</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with amide/amid ratio of 75/25 (unitless) – formulas only</td>
<td>Not serious(^4) Not serious(^3) Not serious(^16) Very serious(^1) None</td>
<td>2</td>
<td>5.42 (5.42, 5.44)</td>
<td>VERY LOW</td>
<td>CRITICAL</td>
</tr>
<tr>
<td>Nitrogen to protein conversion factor ( K ) with directly calculated amide/amid ratio (unitless) – formulas only</td>
<td>Not serious(^4) Not serious(^3) Not serious(^16) Very serious(^1) None</td>
<td>2</td>
<td>5.42 (5.40, 5.44)</td>
<td>VERY LOW</td>
<td>CRITICAL</td>
</tr>
</tbody>
</table>
1 Importance for decision-making. In this case, conversion factors are the only “outcome” and thus each of the variations is critical for decision-making.

2 Regarding K’ and inclusion of prosthetic groups for soy, an additional conversion factor with a value of 5.91 was calculated specifically for the soy 7S protein β-conglycinin (Maubois J-L, Lorient D (2016). Dairy proteins and soy proteins in infant foods nitrogen-to-protein conversion factors. Dairy Sci Tech 96(1):15–25), based on the assumption that all three subunits of β-conglycinin are glycosylated. The authors further use this information to estimate factors for total soy proteins with different 11S/7S ratios, in the range 5.69–5.79. These values were not included in the final analysis as they were not directly measured but estimated, based on assumptions made in reports in the literature.

3 Because this factor was derived from a study using measurement of crude protein with total nitrogen, it was started at “moderate” certainty.

4 There was no indication of systematic measurement error or reporting error (i.e. selective reporting of K’ values). Not downgraded.

5 Inconsistency was not formally assessed as only a single study was available.

6 No studies reported K’ values for infant formula or follow-up formula. A single study reported K’ for soybeans and this value was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. In addition, calculation of K’ includes prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids only. Downgraded twice for very serious indirectness.

7 Fewer than 5 studies contributed to the mean. Downgraded twice for very serious imprecision.

8 Because these factors were derived primarily from studies using amino acid composition analysis, they were started at “moderate” certainty.

9 Only four values for formulas were identified (from two sources). The mean conversion factor was therefore primarily derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Although K50/50 and 75/25 values are based on setting the ratio of amide to acid arbitrarily at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike Ks, K values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Downgraded once for serious indirectness.

10 The minimum of 10 measurements is satisfied. Not downgraded.

11 Although the overall certainty in the evidence for K50/50 and K75/25 was assessed as low, there was greater confidence in the value for Ks 50/50 because it is in line with that of the Ks derived from the directly adjusted amide to acid ratio, which is the most accurate method of assessing amide to acid ratio.

12 Only two values for formulas were identified (from one source). The mean conversion factor was therefore primarily derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Although K50/50 and 75/25 values are based on arbitrarily setting the ratio of amide to acid at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike Ks, K values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Downgraded once for serious indirectness.

13 Only two values for formulas were identified (from one source). The mean conversion factor was therefore primarily derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, unlike Ks, K values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Downgraded once for serious indirectness.

14 No studies reported KB values for infant formula and follow-up formula. The mean conversion factor was therefore derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Downgraded once for serious indirectness.
Conversion factors were calculated directly from infant formula and follow-up formula. Although $K_A$ 50/50 and 75/25 values are based on setting the ratio of amide to acid arbitrarily at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, unlike $K_B$, $K_A$ values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Not downgraded.

Conversion factors were calculated directly from infant formula and follow-up formula. Also, unlike $K_B$, $K_A$ values do not explicitly take into consideration non-protein nitrogen; however, non-protein nitrogen can vary considerably across food samples. It is therefore difficult to assess the level of indirectness in using a conversion factor that does not take non-protein nitrogen into consideration in estimating protein content in a variety of formulas with different formulations. Not downgraded.

Annex 5 provides information on which studies provided data for each mean conversion factor shown in GRADE evidence profile 2 above.
### GRADE evidence profile 3 – Dairy-based ingredients

**Question:** When using the equation \( \text{amount of protein (P)} = \text{nitrogen to protein conversion factor (K)} \times \text{amount of nitrogen (N)} \) to estimate the protein content of dairy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content plus prosthetic groups?

**Population:** Infant formula and follow-up formula

<table>
<thead>
<tr>
<th>Conversion Factor</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>K´ (unitless) – all dairy foods</td>
<td>5 Mixed</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Serious</td>
<td>Not serious</td>
</tr>
<tr>
<td>K A, with amide/amid ratio of 50/50 (unitless) – all dairy foods</td>
<td>13 Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Serious</td>
<td>Not serious</td>
</tr>
<tr>
<td>K A, with amide/amid ratio of 75/25 (unitless) – all dairy foods</td>
<td>12 Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>Not serious</td>
</tr>
<tr>
<td>K A, with directly calculated amide/amid ratio (unitless) – all dairy foods</td>
<td>12 Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>Not serious</td>
</tr>
<tr>
<td>K B (unitless) – all dairy foods</td>
<td>3 Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>Very serious</td>
</tr>
<tr>
<td>K´ (unitless) – formulas only</td>
<td>1 Amino acid sequencing</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Very serious</td>
<td>None</td>
</tr>
<tr>
<td>K A, with amide/amid ratio of 50/50 (unitless) – formulas only</td>
<td>1 Amino acid analysis</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Serious</td>
<td>Not serious</td>
</tr>
</tbody>
</table>

CI, confidence interval

1 Importance for decision-making. In this case, conversion factors are the only “outcome” and thus each of the variations is critical for decision-making.
Amino acid sequencing and measurement (weighing) of crude protein with total nitrogen analysis. Because this factor was derived primarily from studies using amino acid sequencing, it was started at “high” certainty.

There was no indication of systematic measurement error or reporting error (i.e. selective reporting of K’ values). Not downgraded.

Inconsistency was assessed by considering the level of variance around the mean. The 95% CI suggests very little variation around the mean. Not downgraded.

Only a single K’ value for infant formula and follow-up formula was identified in the literature review. The mean conversion factor was therefore primarily derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Downgraded once for serious indirectness.

The minimum of 10 measurements is satisfied. Not downgraded.

Because these factors were derived primarily from studies using amino acid composition analysis they were started at “moderate” certainty.

Of the 31 measurements, 11 come from infant formula and follow-up formula. Although K, 50/50 and 75/25 values are based on arbitrarily setting the ratio of amide to acid at 50/50 or 75/25, it was expected that this would not result in significant enough difference from values derived via direct adjustment. Also, calculation of K, does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded once for serious indirectness.

In addition to having greater confidence in the K, 50/50 value relative to K, 75/25 because the former includes a significant number of studies which derived conversion factors directly from infant formula and follow-up formula (and therefore was not downgraded for serious indirectness), there was greater confidence in this value because it is in line with that of the K, derived from the directly adjusted amide to acid ratio, which is the most accurate method of assessing amide to acid ratio.

No studies reported K, or K, values for infant formula and follow-up formula. The mean conversion factor was therefore derived from dairy sources other than infant formula and follow-up formula (e.g. milk, milk proteins, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, calculation of K, or K, does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded twice for very serious indirectness.

Fewer than 5 studies contributed to the mean. Downgraded twice for very serious imprecision.

Because this factor was derived from a study using amino acid sequencing, it was started at “high” certainty.

Conversion factors were derived from infant formula and follow-up formula. Not downgraded.

Conversion factors were derived from infant formula and follow-up formula. However, calculation of K, does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded once for serious indirectness.

Annex 4 provides information on which studies provided data for each mean conversion factor shown in GRADE evidence profile 3 above.
**GRADE evidence profile 4 – Soy-based ingredients**

**Question:** When using the equation \( \text{amount of protein (P)} = \text{nitrogen to protein conversion factor (K) } \times \text{amount of nitrogen (N)} \) to estimate the protein content of soy-based ingredients used in infant formula and follow-up formula, which value of K most closely estimates the true amount of protein (P), where “protein” is defined as amino acid content plus prosthetic groups?

**Population:** Infant formula and follow-up formula

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K' ) (unitless) – all soy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total protein and nitrogen³</td>
<td>Not serious⁴</td>
<td>2</td>
<td>5.71 (single measurement)²</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with amide/amid ratio of 50/50 (unitless) – all soy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>28</td>
<td>5.68 (5.66, 5.69)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with amide/amid ratio of 75/25 (unitless) – all soy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>26</td>
<td>5.40 (5.38, 5.42)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with directly calculated amide/amid ratio (unitless) – all soy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>35</td>
<td>5.65 (5.61, 5.68)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_B ) (unitless) – all soy foods</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>16</td>
<td>5.35 (5.20, 5.51)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with amide/amid ratio of 50/50 (unitless) – formulas only</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>4</td>
<td>5.70 (5.69, 5.71)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with amide/amid ratio of 75/25 (unitless) – formulas only</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>2</td>
<td>5.42 (5.42, 5.42)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nitrogen to protein conversion factor ( K_A ), with directly calculated amide/amid ratio (unitless) – formulas only</th>
<th>Quality assessment</th>
<th>No. of independent measurements</th>
<th>Mean (95% CI)</th>
<th>Certainty</th>
<th>Importance¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid analysis⁸</td>
<td>Not serious⁴</td>
<td>2</td>
<td>5.42 (5.40, 5.44)</td>
<td>VERY LOW CRITICAL</td>
<td></td>
</tr>
</tbody>
</table>
1 Importance for decision-making. In this case, conversion factors are the only “outcome” and thus each of the variations is critical for decision-making.

2 Regarding K’ and inclusion of prosthetic groups for soy, an additional conversion factor with a value of 5.91 was calculated specifically for the soy 7S protein β-conglycinin (Maubois J-L, Lorient D (2016). Dairy proteins and soy proteins in infant foods nitrogen-to-protein conversion factors. Dairy Sci Tech 96(1):15–25), based on the assumption that all three subunits of β-conglycinin are glycosylated. The authors further use this information to estimate factors for total soy proteins with different 11S/7S ratios, in the range 5.69–5.79. These values were not included in the final analysis as they were not directly measured but estimated, based on assumptions made in reports in the literature.

3 Because this factor was derived from a study using measurement of crude protein with total nitrogen, it was started at “moderate” certainty.

4 There was no indication of systematic measurement error or reporting error (i.e. selective reporting of K’ values). Not downgraded.

5 Inconsistency was not formally assessed as only a single study was available.

6 No studies reported K’ values for infant formula or follow-up formula. A single study reported K’ for soybeans and this value was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Downgraded once for serious indirectness.

7 Fewer than 5 studies contributed to the mean. Downgraded twice for very serious imprecision.

8 Only four values for formulas were identified (from two sources). The mean conversion factor was therefore derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, calculation of K_a or K_b does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded twice for very serious indirectness.

9 Although the overall certainty in the evidence for K_a 50/50 and K_a 75/25 was assessed as very low, there was greater confidence in the value for K_a 50/50 because it is in line with that of the K_a derived from the directly adjusted amide to acid ratio, which is the most accurate method of assessing amide to acid ratio.

10 Only two values for formulas were identified (from one source). The mean conversion factor was therefore derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, calculation of K_a or K_b does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded twice for very serious indirectness.

11 No studies reported KB values for infant formula or follow-up formula. The mean conversion factor was therefore derived from soy sources other than infant formula and follow-up formula (e.g. soybean, soy isolates, etc.) and was thus considered an indirect assessment of the conversion factor for infant formula and follow-up formula. Also, calculation of K_a or K_b does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded twice for very serious indirectness.

12 Conversion factors were calculated directly from infant formula or follow-up formula. However, calculation of K_a or K_b does not include prosthetic groups in the protein mass and the PICO question specifically defines protein as amino acids plus prosthetic groups. Downgraded once for serious indirectness.

Annex 5 provides information on which studies provided data for each mean conversion factor shown in GRADE evidence profile 4 above.
### ANNEX 4

**STUDIES INCLUDED IN THE GRADE ASSESSMENT: DAIRY**

The studies (data sets) listed below are those used to derive the mean conversion factors found in each row of GRADE evidence profiles 1 and 3.

<table>
<thead>
<tr>
<th>ROW</th>
<th># OF STUDIES</th>
<th>STUDIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALL DAIRY ITEMS</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1. $K'$ | 5 | Jones, 1931  
Van Boekel M, 1987  
Hammersten, 1883 | Maubois and Lorient, 2016  
Rouch et al., 2008 |
| 2. $K_a$ (50/50) | 13 | Boisen et al., 1987  
Ceballos et al., 2009  
Csapo-Kiss et al., 1994  
Derham, 1982  
Featherston et al., 1964  
Japan Vegetable Food Association*  
Krizova et al., 2013 | Marino et al., 2010  
Maubois and Lorient, 2016  
Rutherford and Moughan, 1998  
Sarwar et al., 1983b  
Tomotake et al., 2001  
Zándoki, 2006 |
| 3. $K_a$ (75/25) | 12 | Boisen et al., 1987  
Ceballos et al., 2009  
Csapo-Kiss et al., 1994  
Derham, 1982  
Featherston et al., 1964  
Japan Vegetable Food Association* | Krizova et al., 2013  
Marino et al., 2010  
Rutherford and Moughan, 1998  
Sarwar et al., 1983b  
Tomotake et al., 2001  
Zándoki, 2006 |
| 4. $K_a$ (direct) | 12 | Boisen et al., 1987  
Ceballos et al., 2009  
Csapo-Kiss et al., 1994  
Derham, 1982  
Featherston et al., 1964  
Japan Vegetable Food Association*  
Krizova et al., 2013 | Marino et al., 2010  
Rutherford and Moughan, 1998  
Sarwar et al., 1983b  
Sosulski, 1990  
Tomotake et al., 2001  
Zándoki, 2006 |
| 5. $K_b$ | 3 | Boisen et al., 1987  
Fujihara et al., 2010  
Salo-väänänen and Koivistoiten, 1996 |
| **FORMULAS ONLY** | | |
| 7. $K_a$ (50/50) | 1 | Nestec* |

* data submitted during call for data
## ANNEX 5
### STUDIES INCLUDED IN THE GRADE ASSESSMENT: SOY

The studies (data sets) listed below are those used to derive the mean conversion factors found in each row of GRADE evidence profiles 2 and 4.

<table>
<thead>
<tr>
<th>ROW</th>
<th># OF STUDIES</th>
<th>STUDIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALL SOY ITEMS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. K'</td>
<td>1</td>
<td>Jones, 1931</td>
</tr>
<tr>
<td>5. Kₐ</td>
<td>8</td>
<td>Boisen et al., 1987&lt;br&gt;ENSA&lt;br&gt;Fujihara et al., 2010&lt;br&gt;Japan Vegetable Food Association&lt;br&gt;Morr, 1981, 1982&lt;br&gt;Mossé, 1990&lt;br&gt;Sosulski and Holt, 1980&lt;br&gt;Sriperm et al., 2011</td>
</tr>
<tr>
<td><strong>FORMULAS ONLY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Kₐ (50/50)*</td>
<td>2</td>
<td>Health Canada&lt;br&gt;Nestec</td>
</tr>
<tr>
<td>7. Kₐ (75/25)*</td>
<td>1</td>
<td>Nestec</td>
</tr>
<tr>
<td>8. Kₐ (direct)*</td>
<td>1</td>
<td>Nestec</td>
</tr>
</tbody>
</table>

*a data submitted during call for data
BIBLIOGRAPHY


BIBLIOGRAPHY


(https://doi.org/10.1021/ac0006268).


(https://pubs.acs.org/doi/abs/10.1021/jf00089a039).


64


(https://pubs.acs.org/doi/pdfplus/10.1021/jf60200a006).


BIBLIOGRAPHY


(https://hdl.handle.net/2123/2264).


(https://doi.org/10.1093/jxb/eri069).


(https://pubs.acs.org/doi/abs/10.1021/jf60205a042).

(https://pubs.acs.org/doi/pdfplus/10.1021/jf00064a039).


(https://hal.archives-ouvertes.fr/hal-00929173/document).


(https://pubs.acs.org/doi/10.1021/jf050071n).


(https://doi.org/10.1021/acs.analchem.8b00015).


