Dairy proteins and soy proteins in infant foods
nitrogen-to-protein conversion factors

J.-L. Maubois¹ · D. Lorient²

Abstract Protein content of any source is classically determined through the analysis of its nitrogen content done for more 100 years by the Kjeldahl method, and the obtained result is multiplied by a number named nitrogen conversion factor (NCF). The value of NCF is related to the amino acid composition of the protein source and to the eventual presence of side groups covalently bound to some amino acids of the protein chain. Consequently, the value of NCF cannot be identical for all sources of food proteins. The aim of this paper is to review the available knowledge on the two allowed protein sources for infant food formulas, milk and soybean, in order to bring the right scientific basis which should be used for the revision of both European legislation and Codex Standard for Infant Formulas.

Keywords Dairy protein · Soy protein · Nitrogen content · True protein · Kjeldahl analysis

1 Introduction

Like most of the foods consumed by human beings, milk and soybeans contain proteins which are sources of essential nutrients for growth and for the maintenance of physiological functions.

Human milk is undoubtedly the optimal food for infants, and all the recent acquired knowledge on it strengthens the conclusion that breast feeding should be the golden rule all over the world because it is still impossible to duplicate the composition and component structure of mother’s milk (Jensen et al. 1995). Nevertheless, there are

J.-L. Maubois and D. Lorient are former directors.

* J.-L. Maubois
  Jean-Louis.Maubois@wanadoo.fr

¹ INRA, Dairy Research Laboratory, F-35042 Rennes Cedex, France
² ENSBANA, F-21079 Dijon Cedex, France
needs to palliate several medical or social difficulties of mothers, and as a result, infant food formulas have been proposed as substitutes for human milk since one century (Fomon 1974).

In the framework of revision of both European legislation and Codex Standard for Infant Formulas, the amount and quality of proteins contained in these infant formulas are under discussion. The only allowed protein sources for these infant formulas are milk and soybean (and their respective derivatives), so given the specific characteristics of each source, it appears necessary to review their analytical estimation.

Indeed, adequate knowledge of protein nature and protein content is essential for evaluating the nutritional value of foods correctly, informing the consumer adequately, estimating the yielding capacity in manufactured products precisely (for example, cheese from milk and tofu from soy juice), and insuring fair trade exchanges.

2 General scientific knowledge

The protein content referred to as ‘total protein’ is determined through analysis of nitrogen contained in the foods, and then the obtained result is multiplied by a number named nitrogen conversion factor (NCF).

In foods, two nitrogen forms are distinguished: protein nitrogen and non protein nitrogen.

- Protein nitrogen corresponds to ‘True Proteins’ i.e. for milk, mainly, caseins, β-lactoglobulin, α-lactalbumin, serum albumin, immuno-globulins and proteose-peptones and for soy, mainly, β-conglycinin and glycinin. Proteins are defined as a sequence (determined by the organism’s genome) of amino acids bound by covalent bounds (primary structure) and to which carbohydrate groups can be also attached by covalent bounds (for example, on Threonin for milk κ-casein and for egg yolk phosvitin). These side groups are constituting parts of the protein, not only because they are covalently bound to the amino acid chain but also for their nutritional, physiological and technological functions. For example, it is the release of κ-glycomacopeptide through hydrolysis of casein micelle by rennet and pepsin in the stomach that regulates the differential bio-availability kinetics of caseins and whey proteins (Boirie et al. 1997). This release also induces the secretion of the cholecystokinin hormone implied in the regulation of gallbladder and pancreatic functions (Beucher et al. 1994); it is equally involved in the phenomenon of milk coagulation by rennet, an essential step of cheesemaking.

- Non protein nitrogen (NPN) amounts 5% of the total nitrogen in cow’s milk (Alais 1984) and between 7.8 and 2.9% (average 5%) of total nitrogen in defatted soybean flour prior to heat treatment (Becker et al. 1940; Smith and Circle 1972). The NPN is mainly constituted by small peptides, free amino acids and other nitrogenous components, such as creatinin, urea and ammonia, in both protein sources Robertson and van der Westhuizen (1990).
The analytical method, used for determining total nitrogen content for more than a hundred years, is the Kjeldahl method (Kjeldahl 1883) (ISO 2014) which consists of the following:

- Mineralising the sample with concentrated sulphuric acid, in the presence of a catalyst (copper and potassium sulphates) that quantitatively converts all the organic nitrogen into ammonium sulphate and
- Displacing the ammonia from the ammonium sulphate by concentrated sodium hydroxide, then distilling the ammonia and collecting it in a boric acid solution and finally, titrating it with hydrochloric acid.

The protein content of the sample is then calculated by multiplying the so-determined nitrogen content by a NCF. The different calculation methods of this conversion factor are detailed below. As proteins are particularly complex heteropolymers, consisting of chains of variable number of α-amino acids (20 different types), some of which may be amidated, glycosylated or phosphorylated, it is obvious that the value of the conversion factor depends on the primary structure of the analysed protein (composition in amino acids and sequence).

### 3 Methods for determining protein content

In theory, the definition of the NCF seems simple and unambiguous: it is the ratio between the protein molecular weight and the nitrogen content of this protein. But, in practice, precise analytical determination faces numerous difficulties inherent to the methods used (Mossé 1990).

#### 3.1 Determination of the conversion factor with the Kjeldahl method

Initially, the NCF was obtained through the simple determination of nitrogen content of a highly purified protein. That was the case of the 6.38 factor calculated by Hammarsten and Sebelien (1892) for isoelectric casein (Jones 1931) and further adopted for any milk proteins. However, most of the protein sources are not so easy to purify as casein and scientists have turned to determining the NCF from the amino acid composition.

#### 3.2 Determination of the conversion factor from amino acid profiles

Further determinations of NCF were based on the amino acid composition determined, after acid hydrolysis, in protein samples, but this analytical method, even if it is performed in optimum conditions (three hydrolysis times, use of internal markers, etc.), suffers, at least, a 10% lack of precision, because of (1) some of amino acid residues (particularly Try, Thr, Ser, Met, Tyr, Val, Ileu) are partially destroyed by acid hydrolysis and others are not 100% released and (2) it requires complementary steps for recovery of amides, sulphur and aromatic amino acids. Moreover, this method does not give information on side groups
(glycosylated, phosphorylated), which are constitutive parts of the native protein chain after post-translational biosynthesis. The same uncertainty is also existing if amino acid composition is determined by using RP-HPLC method with pre-column derivatization (Grappin and Ribadeau-Dumas 1992).

3.3 Determination of the conversion factor from the amino acid sequence of proteins

This is the only method that gives a scientifically founded NCF value because it is calculated from the complete primary structure of the protein: all the amino acid residues composing the protein chain and all the side groups. The NCF value is obtained by relating the nitrogen content of all amino acids to the molecular weight of the protein (determined either by calculation or by mass spectrometry). This obviously requires a thorough knowledge of the primary structures of proteins which was acquired for most (99.5%) of milk proteins and for the main soy proteins, but is not yet available in its entirety for numerous other protein sources. This determination of NCF should be the reference method because it cannot be contested. If this knowledge of the primary structure is not available, an approximate value for NCF can be calculated from the protein sequence determination based on the specific gene coding.

4 Conversion factors for milk, specific milk proteins, some dairy products and milk infant formulas

4.1 Milk, milk proteins and dairy products

Thanks to the pioneering work realised by the INRA (French National Institute For Research in Agronomy) team under the leadership of Ribadeau-Dumas, in the 1970s for the milk caseins, as well as the specific work done on κ-casein by Jolles et al. (1970) and Alais (1984) and the studies of Eigl et al. (1984) and of Brew et al. (1970) on whey proteins, the complete primary structure of the main milk proteins is known and internationally recognised (Farrell et al. 2004). This knowledge enabled Karman and van Boekel (1986) and later Van Boekel and Ribadeau-Dumas (1987) to determine, with a high degree of precision, the values of NCF for the main protein sources of milk (Table 1).

It can be seen that for total milk proteins, the conversion factor value obtained by taking into account the known sequences of all the main individual proteins with their lateral groups and their proportions in milk is 6.36, the value which is very close to the historically used value of 6.38.

From these results, it can be said that the values of NCF which must be used for the proteins contained in milk infant formulas are 6.36 for the casein part (isoelectric casein being effectively used for this type of formula) and 6.41 for the whey proteins part (most of the infant formulas being made with rennet whey proteins derivatives).

Values of NCF which are given in some papers for other dairy products such as cheese (Mariotti et al. 2008) are totally erroneous because they do not take in account
the extreme diversity of technological treatments applied to cheese milk (heat, specific protein enrichment) which lead to variable retention of whey proteins and the continuous proteolysis involved in the mechanisms of cheese ripening. Consequently, the best value of NCF for cheese and other dairy products cannot be other than that of milk proteins, i.e. 6.38.

4.2 Milk infant formulas

In Table 2, the values of NCF in milk infant formulas have been calculated for different whey protein/casein proportions usually proposed for term infants at birth (Jensen et al. 1995).

These calculations clearly show that whatever are the relative proportions of whey proteins and casein in milk infant formulas, the nitrogen conversion factor still remains around the value of 6.38.

Whether one is dealing with milk proteins as whole or fractions isolated from these milk proteins for an use in infant formulas, it is scientifically established from the knowledge of primary structure that the nitrogen conversion factor is undoubtedly 6.38.

**Table 1** Calculated evaluation of the Kjeldahl factor for conversion of the nitrogen content of milk products to protein content, according to data of Van Boekel and Ribadeau-Dumas (1987)

| Product/protein         | Conversion factor (with carbohydrates) | Percentage of each protein/total protein in milk |
|-------------------------|----------------------------------------|-----------------------------------------------|
| αs1-Casein              | 6.36                                   | 30.3                                          |
| αs2-Casein              | 6.29                                   | 7.9                                           |
| β-Casein                | 6.37                                   | 28.2                                          |
| κ-Casein                | 6.35                                   | 10                                            |
| γ-Casein                | 6.34                                   | 2.4                                           |
| β-Lactoglobulin         | 6.29                                   | 9.7                                           |
| α-Lactalbumin           | 6.25                                   | 3.6                                           |
| Serum albumin           | 6.07                                   | 1.2                                           |
| Proteoses peptones      | 6.55                                   | 0.9                                           |
| Immunoglobulins         | 6.20                                   | 2.4                                           |
| Milk                    | 6.36                                   |                                               |
| Isoelectric casein      | 6.36                                   |                                               |
| Rennet whey proteins    | 6.41                                   |                                               |

**Table 2** Calculation of the nitrogen-to-protein conversion factor for different formulations of milk infant formulas with the use of a 6.41 conversion factor for proteins isolated from rennet whey and a 6.36 conversion factor for isoelectric casein (from Van Boekel and Ribadeau-Dumas 1987)

| Proportion in infant formula whey protein (6.41)/isoelectric casein (6.36) | Conversion factor for infant formula |
|---------------------------------------------------------------------------|-------------------------------------|
| 20/80                                                                     | 6.370                               |
| 30/70                                                                     | 6.375                               |
| 50/50                                                                     | 6.385                               |
| 60/40                                                                     | 6.390                               |
Soy (Glycine max) proteins, mostly globulins, are distinguished according to their sedimentation coefficients into globulins 7S and in globulins 11S. The ratio 11S/7S varies between 0.5 and 1.7, according to the cultivars (Utsumi 1992). Soy 7S globulins mainly comprise β-conglycinin (30 to 50% of total proteins) composed of three subunits named α’ (72 kDa), α (68 kDa) and β (52 kDa), γ-conglycinin (170 kDa) and the basic globulin 7S (168 kDa), all three proteins being glycosylated (Utsumi et al. 1997). The 11S soy globulin, named glycgin, is a hexamer with a molecular weight ranging between 300 and 380 kDa, owing to a high polymorphism between the cultivars (Mori et al. 1981).

The first value of NCF proposed for soy proteins was 5.71 (Jones 1931) which was calculated from the nitrogen determinations performed by Osborne and Campbell (1898) on soy protein extracts. Then, with no known scientific reason other than a theoretical 16% nitrogen content in all protein sources, the value of 6.25 was agreed for all vegetable proteins and applied for soy proteins by soy producers as well as the Association of Official Analytical Chemists (AOAC). That was done despite since 1946 this value was considered too high, in view of the studies performed on soy isolates (Smiley and Smith 1946, Smith and Circle 1972, Mossé 1990, Sosulski and Imafidon 1990). In 1969, Tkachuck suggested the value of 5.69 for the total proteins contained in defatted unskinned soybean flour. Later, Mossé (1990), from the amino acid profiles of six samples of soy protein powders, determined a value of 5.52±0.02.

From the described sequences of the main soy proteins (Utsumi 1992), we calculated the NCF values shown in Table 3. Given the variability of the relative proportions between glycinin (11S) and β-conglycinin (7S), 0.5 to 1.7 (Utsumi 1992) in the cultivars, it is not easy to calculate a mean value of NCF but the values calculated in Table 3 lie in a very narrow range. One can therefore consider that the ratios protein/nitrogen in all soy cultivars are varying between 5.56 and 5.66 that leads to a mean value of 5.61.

However, this value does not take into account the covalently bound side groups. According to Utsumi et al. 1997, the three subunits of the β-conglycinin (7S) are glycosylated (Kosiyama 1969) as well as the hemagglutinin component (Lis et al. 1966) which amounts to 3% of soy flour (Liener and Rose 1953). Soy hemagglutinin is always denatured by the heat treatments applied to soy proteins for inhibiting its anti-nutritional action, but its glycosylation as well as that of β-conglycinin (7S) has to be taken into account for the calculation of NCF. Consequently, the respective specific values for hemagglutinin and β-conglycinin (7S) become 7.58 (Wolf 1972) and 5.91. Thus, it can be calculated by taking into consideration these new values and the relative

| Product/protein | Conversion factor |
|-----------------|------------------|
| β-Conglycinin (α’) | 5.58 |
| β-Conglycinin (α) | 5.65 |
| β-Conglycinin (β) | 5.66 |
| Glycinins (mean value of the five subunits) | 5.56 |
proportion (15%) of both proteins in total soy proteins, the following values of NCF for the different soy cultivars with variable ratios 11S/7S:

- Ratio 11S/7S=0.5 NCF=5.79
- Ratio 11S/7S=1.0 NCF=5.73
- Ratio 11S/7S=1.5 NCF=5.69

It can be concluded from these calculations which agreed with numerous published data that whatever is the soy cultivar, the use of a conversion factor of 6.25 for soy proteins leads to an overestimation of protein content comprised between 7.4 and 9.0%. This conclusion was recently confirmed by Fujihara et al. (2010) who consider as realistic for these proteins a NCF value between 5.43 and 5.51.

6 Variation factors

Apart the aforementioned analytical variation factors, technological treatments especially the intense heat treatments can affect the value of NCF by inducing glycation reactions between amino acids (Lys or Arg) and sugar residues present in the heated product (Maillard reaction) as well as deamidations able to ultimately lead to a release of CO₂ and amines (Strecker reaction) (Finot 1997).

In order to avoid protein allergenicity, some infant formulas contain enzymatic hydrolysates issued from proteins. It is obvious that for these hydrolysates, essentially composed of peptides and free amino acids, the conversion factor is meaningless (there is no protein anymore!). Other analytical parameters have to be proposed for characterizing these products, such as average peptide size, Gaussian distribution or amino acid sequence determination of all the individual peptides. The same conclusion comes to mind when there is a supplementation with free amino acids, generally added at very low levels in order to avoid osmotic shocks.

Milk infant formulas are increasingly formulated with milk and/or whey protein concentrates or isolates obtained through various separation technologies based on either steric size exclusion of milk and whey proteins (membrane ultrafiltration and 0.1-μm microfiltration or gel filtration chromatography) or according to the electro-chemical charge of proteins (membrane nanofiltration, electro-dialysis and ion exchange resins) (Maubois and Ollivier 1997). Without entering into the details of these formulations which are in the field of industrial property, it must be pointed out that the concentrates (WPC) or isolates (WPI) obtained through these separation processes are not enriched in NPN (a very complex group of likely more than 100 molecular species) as stated erroneously in the ESPGHAN report (Koletzko et al. 2005). Indeed, the NPN-constituting molecules because either their low molecular weight (less than 500 Da according to Wolfschoon-Pombo and Klostermeyer 1981, Alais 1984) cannot be retained by the ultrafiltration (UF) membranes of which the molecular cut-off is around 5000 to 20,000 Da. They are also excluded by the gel permeation resins and separated by the ion exchange resins used in the chromatography processes. In all the demineralisation processes (electrodialysis, ion exchange, membrane nanofiltration), according to Hoppe and Higgins (1992), the NPN losses
represent between 25 and 30% of the initial whey NPN. Moreover, some little losses in low molecular weight whey proteins (about 0.02%), such as $\alpha$-lactalbumin, have been observed industrially (Hargrove et al. 1976, Delaney 1976; Boer and Robbertsen 1983). So, the WPC and WPI have a lower NPN content than the dairy products of which they originated. This lower content in NPN of which the value of NCF would be, according Karman and van Boekel (1986), in the ranges from 7.36 (\(\kappa\)-caseinomacropeptide) to 3.60 (milk NPN fraction), evidently leads in all the products used in infant formulas (demineralised whey, whey protein concentrates, whey protein isolates, etc.) to a balanced variation of the value of NCF, but which always remains between 6.30 and 6.45 (values determined by Karman and van Boekel (1986) for, respectively, rennet whey proteins, and acid whey proteins), and consequently does not highly affect the used 6.38 value.

7 Processing and anti-nutritional factors

In infant food formula being used for nutrition of particularly sensitive human beings, it must be mentioned that technological treatments such as heating, applied for killing contaminant micro-organisms, have a deleterious effect on the nutritional value and the potential physiological role of proteins. Minimization of the successive heatings which have a cumulative damageable effect must be the rule of the infant food manufacturers. Moreover, intense heating has always to be applied on all sources of soy proteins, because, unlike milk proteins, they contain anti-nutritional components, notably (1) hemaggglutinin and (2) trypsin and chymotrypsin inhibitors able to block these enzymes which are essential for the protein digestion by mammals (Jaffe 1969). These inhibitors (polypeptides which represent 6% of total soy proteins) distinguished in Bowman-Birk (Birk 1968; Frattali 1969) and Kunitz (Kunitz 1946) could equally increase endogenous nitrogen losses by enhancing pancreatic proteases secretion (Finot 1997). Consequently, all the soy-protein-based foods are highly heated (at least, 100 °C for 15 min according to Rackis (1966) for inactivating these anti-nutritional components. But, such a heat treatment, of which the intensity is far from the one applied to milk-based infant formula (145 °C during 2 to 3 s, at the maximum) has also multiple negative effects on protein bio-availability: induction of Maillard and Strecker reactions which leads to a decrease of absorption of basic essential amino-acids (lysine and arginine) as well as tryptophan and sulphur amino acids (cystein and methionin) (Finot 1997). These serious anti-nutritional defects of soybean-based infant formulae have recently led paediatric community to limit them as much possible for infant nutrition (Turck 2007). On the other hand, there are many other differences between milk-based infant formula and those based on soy protein, particularly composition and structure of the fat component for which we can wonder if the added vegetable fat, totally cholesterol free (human and cow milks contain around 150 mg.L$^{-1}$, Jensen et al. 1995) and not structured in globules, fulfills or not the requirements for an optimal growth and development of children (Jensen et al. 1995).
8 Conclusions

The 6.38 factor converting the nitrogen content determined by the Kjeldahl method in milk proteins, used for more than one century in all the international standards and recognised by Codex for all the dairy products, is based on deep scientific knowledge which cannot be contested, notably the amino acid sequential chain of the milk proteins as well as the precise identification of the post-translational glycosylated and phosphorylated side groups.

Although studies have recently determined the sequences of numerous other proteins, this knowledge of milk proteins has no equivalent in the field of vegetable proteins, probably because of the high genetic polymorphism of cultivars. Nevertheless, the available knowledge existing for soy proteins shows, whatever is the cultivar, the conversion factor is never over 5.79 with an average value of 5.61, both values agreeing with that (5.71) encountered in many published scientific papers.

Considering the values of the conversion factors for milk proteins (6.38) and for soy proteins (5.71), it is scientifically justified that this difference is kept.

The recent proposal of the use of a unique 6.25 conversion factor for all the protein sources (Koletzko et al. 2005) is unacceptable because it forgot the enormous research work realised, for more than 50 years, by the world scientific community in order to improve knowledge on that essential nutriment for human beings which are the proteins with their differences in terms of amino acid composition and their specific nutritional quality. Such a 6.25 value has absolutely no recognised scientific basis (Morr 1982; Sriperm et al. 2011), and especially for infant formulas, it is not appropriate for any used protein source: it overestimates by around 10% soy proteins and underestimates by 2% milk proteins. Only use of both specific conversion factors will give a true indication of the protein content.

Moreover, it can be added that if the protein content of a food is a basic indication of its nutritional quality, that quality is also constituted by the following:

– The presence or absence of anti-nutritional components (such as anti-trypsic factors and hemagglutinin of soy proteins) which obligatory requires an intense heat pre-treatment [damageable for the bio-availability of several essential amino acids (Lys, Arg particularly)].
– The amino acid composition which governs the protein metabolism of the human physiology.
– The specific amino acid sequence which leads after the hydrolysis by the digestive enzymes to the release of particular and unique bioactive peptides of which the physiological action has been demonstrated during these last 15 years.

References

Alais C (1984) Science du lait. Edition Sepaic, Paris
Becker HC, Milner RT, Nagel RA (1940) A method for determination of nonprotein nitrogen in soybean meal. Cereal Chem 17:447–457
Beucher S, Levenez F, Yvon M, Corring T (1994) Effects of gastric digestive products from casein on CCK release by intestinal cells in rat. J Nutr Biochem 5:578–584
Morr CV (1982) Recalculating nitrogen conversion factors for several soybean protein products. J Food Sci 47:1751–1752

Mossé J (1990) Nitrogen to protein conversion factor for ten cereals and six legumes and oil seeds. A reappraisal for its definition and determination. Variation according to species and to seed protein content. J Agric Food Chem 38:18–24

Osborne TB, Campbell GF (1898) Proteins of the soya bean. J Am Chem Soc 20:419–428

Rackis JJ (1966) Soybean trypsin inhibitor. Their inactivation during meal processing. Food Technol 20:1482–1484

Robertson NH, van der Westhuizen E (1990) Non-protein content of milk as influenced by various factors. S Afr J Dairy Sci 22:1–8

Smiley WG, Smith AK (1946) Preparation and nitrogen content of soybean protein. Cereal Chem 23:288–296

Smith AK, Circle SJ (1972) Chemical composition of the seed in soybeans: chemistry and technology. Avi Publ. Co, Westport, pp 61–92

Sosulski FW, Imafidon GI (1990) Amino acid composition and nitrogen to protein conversion factors for animal and plant foods. J Agric Food Chem 38:1351–1356

Sripermn N, Pesti GM, Tillúan PB (2011) Evaluation of the fixed nitrogen-to-protein (N:P) conversion factor (6.25) versus ingredient specific N:P conversion factors in feedstuffs. J Sci Food Agric 91:1182–1186

Tkachuck R (1969) Nitrogen to conversion factor for cereals and oilseed meals. Cereal Chem 46:419–423

Turck D (2007) Soy proteins for infant feeding: what we know? Curr Opin Clin Nutr Metab Care 10:360–365

Utsumi S (1992) Plant food protein engineering. In: Kinsella JE (ed) Advances in food and nutrition research, vol. 36. Academic, San Diego, pp 89–206

Utsumi S, Matsumura Y, Mori T (1997) Structure-function relationships of soy proteins. In: Damodaran S, Paraf A (eds) Food proteins and their applications. M. Dekker, New York, pp 257–291

Van Boekel MAJS, Ribadeau-Dumas B (1987) Addendum to the evaluation of the Kjeldahl factor for conversion of the nitrogen content of milk and milk products to protein content. Neth Milk Dairy J 41:281–284

Wolf WJ (1972) Purification and properties of the proteins. In: Chemical composition of the seed in soybeans: chemistry and technology. Avi Publ. Co, Westport, pp 93–143

Wolfschoon-Pombo A, Klostermeyer H (1981) The NPN fraction of cow’s milk. Milchwissenschaft 36:598–600