WHITE PAPER · TECHNICAL REPORT

Paired ^{1}H NMR Screening of Retail Grated Cheese

Lipid and aqueous extracts of 84 cheese products at 300 MHz, evaluated by ratiometric, chemometric and machine-learning methods against reference cheeses of nine classes

John C. Edwards, Ph.D.
Process NMR Associates, LLC
84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
www.process-nmr.com

November 15, 2026 · Version 2.2

Document details

Document	White paper: paired ^{1}H NMR screening of cheese lipid ($CDCl_3$) and aqueous (D_2O) extracts, both acquisition campaigns evaluated as one data set
File name	PNA_Cheese_1H-NMR_Unified_WhitePaper_v2.2.docx
Version and date	2.2, November 15, 2026. Supersedes version 2.1 (November 9, 2026), from which it differs as follows: the extraction procedure was added to Section 3.2 and the corresponding statements in Sections 3.2 and 10 were revised; the statements on the sample records in Sections 10 and 12 and in the Summary were shortened; the placeholders for author input and the summary box on the first page were removed, as was one paragraph of Section 11. The figures, tables and numerical results are those of version 2.1. Version 2.1 had superseded version 2.0 after external review: the mineral results are restricted to reportable titrations; marker floors and a between-lot term were introduced into the non-milk fat estimates; flagging rates of simulated blends replace the detection limit; statements on oil identity, cheese proportion and curing time were narrowed; a fully nested cross-validation, the decision rules, representative fits with residuals and the acquisition parameters were added; the sample register and the model tables were moved to appendices
Author and laboratory	John C. Edwards, Ph.D., Process NMR Associates, LLC, 84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
Data	84 products with paired $CDCl_3$ and D_2O spectra (300 MHz); 43 further lipid spectra; 14 reference fats and oils; 24 amino-acid standard spectra; 55 Cs-EDTA spectra of 36 second-campaign products; one ^{1}H–^{13}C gHSQC spectrum (A-004); archived ^{23}Na and ^{31}P worksheets
Sample identities	Anonymized by cheese type and style; codes A-001 to A-043 and B-001 to B-043 preserve the acquisition numbering; lipid-only spectra coded H-001 to H-043
AI assistance	Data processing, statistics, figures, tables and draft text were prepared with AI assistance (Claude, Anthropic; configured model identifier claude-fable-5-2; the serving model may differ). Sample preparation, NMR acquisition and the original spectral processing, ratiometrics and analysis reporting were performed by Process NMR Associates.
Software	Python 3 (NumPy, SciPy, pandas, scikit-learn, umap-learn, matplotlib); PNA NMR Processor workspaces (.pna); document assembly with docx for Node.js; workbook with openpyxl
Companion file	PNA_Cheese_1H-NMR_Unified_Workbook_v2.1.xlsx (inputs and outputs of the ratiometric, aqueous, EDTA and chemometric analyses)
References	171 numbered references; DOIs verified against Crossref; regulatory, standards, guidance, news and thesis sources without DOIs verified against the publisher websites

Abstract

Two campaigns of ¹H NMR measurements on cheese were combined into a single study of the 84 commercial products for which a $CDCl_3$ (lipid) and a D_2O (water-soluble) extract had been acquired at 300 MHz under one sample code. The set comprised reference cheeses of nine classes, from Parmigiano Reggiano and Pecorino Romano to processed slices, declared vegetable-fat products and plant-based analogues, and 40 retail grated Parmesan and Romano products of unknown composition. Reference fats and oils, amino-acid standards, a ¹H–¹³C gHSQC spectrum, Cs-EDTA titrations of 36 aqueous extracts and 43 further lipid spectra were used in support. No internal standard was present, results are expressed as ratios within a spectrum, and each product was extracted once. The work is a screening study; it has not been validated against a reference method.

The lipid spectra were integrated over ten regions lettered A to J, with the bis-allylic region divided into linolenic (G_1) and linoleic (G_2) parts and the methylene envelope into the sharp line of saturated chain segments (C_1) and the broader envelope of methylene groups in other environments (C_2). The schemes of Tociu and of Hanganu and Chira were applied as published and in a water-robust form, and the overlapped butyrate and n-3 methyl triplets, the bis-allylic triplets, the sterol singlet with its underlying ¹³C satellite, caproleate and conjugated linoleic acid were quantified by line fitting. A milk-fat reference population of 24 products gave butyrate at 9.08 ± 0.47 mol % of acyl chains. A fixed decision rule, given here with its thresholds and order of precedence, separated substitution from lipolysis, in which butyrate is lost while caproleate is retained, and a dilution model estimated the non-milk fat with an uncertainty that combines the natural spread of milk fat with the variation found between retail lots. Markers at or below the level found in fats without milk fat were reported as lower bounds only. The aqueous spectra were evaluated through targeted markers of proteolysis, citrate, residual sugars, dispersed lipid, short-chain acids and preservatives.

Twenty-two of the 40 retail grated products were flagged for non-milk fat, estimated at 43 to 68% of the acyl chains (median 51%, each ± 7 to 8 points), with a composition that matched the fat of the declared cheese products in the set and is consistent with a palm-type fat. In 21 of the 22 the aqueous extract was not that of a ripened hard cheese: the free amino-acid index was 28% of that of the retail products with milk fat as a fraction of spectral area, and 54% of it relative to lactate, and citrate, free sugars and dispersed lipid were present. Among the reference classes this profile is closest to processed cheese. Soluble calcium, available for six of these 21 products, was 73 to 98 mg per 100 g on a provisional scale against 217 to 367 for the four reportable retail products with milk fat. Sixteen retail products were consistent in both extracts with grated hard cheese from milk fat, five of them lipolyzed, and three did not fit either pattern. The data describe composition; they do not identify the oil, the manufacturing route, the proportion of cheese or the curing time. Emmental showed the propionate and acetate of propionic fermentation, hippurate was not detected in any extract, and ripening beyond about ten months could not be resolved from the aqueous data.

Principal component analysis, Ward clustering, t-SNE, UMAP, seven supervised classifiers and four one-class models were compared across 292 combinations of binning, alignment, normalization and scaling. Differences between preprocessing options were within the cross-validation uncertainty; mean-centering with Pareto scaling gave the highest mean score for binned spectra. When the representation, scaling, classifier and hyperparameters were all selected inside the training folds, balanced accuracies of 0.90 for fat, 0.82 for technology and 0.51 for nine cheese styles were obtained. In simulated blends of spectra the decision rule flagged 12 of 24 blends at 20% palm oil and called 24 of 24 non-dairy at 30%; with the fat of the declared cheese products the definitive call reached 23 of 24 at 40%. A detection limit for real products remains to be established with gravimetric blends. The multivariate models agreed with the ratiometric and

targeted evaluation within this data set; because they were trained on the same spectra and reference labels, that agreement is a check of consistency and not an independent confirmation.

Keywords: *¹H NMR; cheese; screening; authentication; grated Parmesan; Pecorino Romano; milk fat; butyrate; caproleate; non-milk fat; processed cheese; free amino acids; citrate; propionic fermentation; lipolysis; EDTA; calcium; magnesium; gHSQC; quantitative NMR; hippurate; ripening; icoshift; binning; Pareto scaling; PCA; UMAP; t-SNE; support vector machine; one-class classification; nested cross-validation*

Table of Contents

1. Introduction **9**
1.1 Economically motivated adulteration of grated hard cheeses 9
1.2 NMR spectroscopy in dairy authentication 9
1.3 Scope and objectives of this study 10
2. Cheese Types Represented in the Study **11**
2.1 Grana-type cheeses: Parmigiano Reggiano, Grana Padano, Piave and domestic Parmesan 11
2.2 Sheep's-milk cheeses: Pecorino Romano, Manchego and a sheep's-milk table cheese 11
2.3 Swiss-type and alpine cheeses: Emmental, Gruyère and Asiago 12
2.4 Washed-curd, Dutch-type and Cheddar-type cheeses 12
2.5 Young, pasta filata and brined cheeses 12
2.6 Blue-veined cheeses 12
2.7 Processed cheese, cheese products and imitation cheese 12
2.8 Plant-based cheese analogues 13
2.9 Retail grated Parmesan and Romano products 13
3. Materials and Methods **14**
3.1 Samples and anonymization 14
3.2 Extraction and NMR acquisition 16
3.3 Data import and common axes 18
3.4 Lipid integration regions and ratiometric schemes 18
3.5 Line fitting of overlapped signals 20
3.6 Reference population, decision rule and mixing model 21
3.7 Aqueous spectra: background separation and targeted features 23
3.8 Calcium and magnesium from the EDTA complexes 24
3.9 Alignment, binning, normalization and scaling 25
3.10 Unsupervised, supervised and one-class modeling 25
4. NMR Spectra and Assignments **27**
4.1 The lipid fraction in $CDCl_3$ 27
4.2 Comparison of the lipid spectra across product types 30
4.3 Reference fats and oils 31
4.4 The water-soluble fraction in D_2O 34
4.5 Two-dimensional ^{1}H–^{13}C correlation of the aqueous extract 38
4.6 Comparison of the aqueous spectra across product types 41
4.7 Quantitative ¹H NMR: from relative indices to concentrations 42
5. Lipid Ratiometrics **43**
5.1 The milk-fat reference population 43
5.2 The region-ratio screen 44
5.3 Fatty-acid classes: published schemes and line fitting 45
5.4 The split regions and the minor markers 49
5.5 Lipolysis and substitution distinguished 51
5.6 Quantifying the non-milk fat 52
5.7 Consistency check on 43 further lipid spectra 54
6. Chemistry of the Aqueous Extracts **56**
6.1 Proteolysis: the free amino-acid index 56
6.2 Lactate, citrate and residual sugars 57
6.3 Lipid dispersed in the aqueous extract 57
6.4 Preservatives, starch and other formulation markers 59
6.5 Short-chain acids: propionic fermentation, acetate and free butyrate 60
6.6 Soluble calcium, magnesium, phosphorus and sodium 62

6.7 Ripening time: what the aqueous spectrum can and cannot resolve....66
7. Chemometrics and Machine Learning....68
7.1 Representation, alignment, normalization and scaling....68
7.2 Principal component analysis....70
7.3 Hierarchical clustering and nonlinear embeddings....72
7.4 Supervised classification....75
7.5 One-class models and simulated blends....78
7.6 Data fusion....80
8. Integrated Interpretation of the Retail Grated Products....80
8.1 Products in which no milk-fat marker was detected....80
8.2 Category A and B: milk fat with a ripened base....81
8.3 Category C: non-dairy fat with a processed-type base....81
8.4 Products that do not fit the main pattern....85
8.5 Agreement between the methods....86
9. Comparison of the Ratiometric, Chemometric and Machine-Learning Approaches....86
10. Limitations and Points for Expert Review....87
11. Recommendations....89
12. Conclusions....90
Appendix A. Sample Register....91
Appendix B. Model Comparison Tables....93
References....94
Summary....104

List of Figures

Figure 1. Analytical workflow..... 16
Figure 2. Representative line fits with residuals for a milk-fat reference (A-003), a retail grated product with non-dairy…..... 21
Figure 3. ¹H NMR spectrum (300 MHz, $CDCl_3$, TMS = 0) of the lipid extract of Parmigiano Reggiano (A-003), scaled to region H…..... 28
Figure 4. Proton environments of an illustrative milk-fat triacylglycerol (sn-1 palmitoyl, sn-2 linoleoyl, sn-3 butyroyl)…..... 29
Figure 5. Diagnostic windows of the $CDCl_3$ spectra of representative products and of the palm oil reference, scaled to region…..... 31
Figure 6. ¹H NMR spectra ($CDCl_3$, region H = 100) of the reference fats and oils compared with Parmigiano Reggiano (top…..... 33
Figure 7. ¹H NMR spectrum (300 MHz, D_2O, TSP-d_4 = 0) of the aqueous extract of a domestic extra-aged Parmesan block of 18…..... 35
Figure 8. ¹H NMR spectra (D_2O) of the amino-acid standards (blue), each scaled to its own tallest signal, compared with the…..... 36
Figure 9. Multiplicity-edited ¹H–¹³C gHSQC spectrum (300 MHz, D_2O) of the aqueous extract of an Italian Pecorino Romano…..... 39
Figure 10. Diagnostic windows of the D_2O spectra of representative products, scaled to total spectral area and offset..... 42
Figure 11. Region-ratio screen..... 45
Figure 12. (a) Butyrate estimated by the integral-based schemes of Tociu and of Hanganu and Chira (without and with the G_1…..... 47
Figure 13. Lipid descriptor maps of the 84 products..... 49
Figure 14. Minor lipid markers..... 51
Figure 15. Estimates of the non-milk fat fraction..... 53
Figure 16. Consistency check..... 56
Figure 17. Principal aqueous markers by class (class codes as in Table 1..... 58
Figure 18. Short-chain acids in the D_2O extracts of selected products (total area = 10^4..... 61
Figure 19. (a) Propionate (quartet/triplet fit) against acetate..... 62
Figure 20. Soluble calcium and magnesium from the Cs-EDTA titration..... 64
Figure 21. Dose check of the EDTA titration..... 65
Figure 22. Soluble minerals..... 66
Figure 23. Aqueous ripening markers against the ripening time stated on the label, for the products that carried a numerical…..... 68
Figure 24. Mean cross-validated score for each representation (rows) and scaling (columns..... 70
Figure 25. PCA of the $CDCl_3$ spectra (aligned 0.01 ppm bins, region H = 100, mean-centered, Pareto-scaled. 71
Figure 26. PCA of the D_2O spectra (aligned 0.01 ppm bins, total area, mean-centered, Pareto-scaled..... 72
Figure 27. Ward hierarchical clustering of the 84 products on the first eight principal components of (a) the lipid bins, (b)…..... 74
Figure 28. t-SNE (perplexity 12) and UMAP (12 neighbors, minimum distance 0.3) embeddings of the lipid, aqueous and fused data..... 75
Figure 29. Supervised models..... 77
Figure 30. One-class models and simulated blends..... 79
Figure 31. Evidence matrix for the 40 retail grated products and the two products with conflicting records..... 83

List of Tables

Table 1. Cheese types represented in the study, the features of manufacture and ripening that govern their chemistry, and…. 14
Table 2. Products for which the sample mass recorded with the lipid extract differs from that recorded with the aqueous…. 17
Table 3. Acquisition parameters recovered from the parameter tables of the processed spectral documents. 17
Table 4. Integration regions of the $CDCl_3$ spectra (TMS = 0) and the number of protons that each contributes per acyl chain…. 19
Table 5. Decision rules. 22
Table 6. ¹H NMR assignments for cheese lipid extracts in $CDCl_3$ (300 MHz, TMS = 0). 30
Table 7. Ratiometric descriptors of the reference fats and oils, for comparison with the milk-fat reference population…. 34
Table 8. Assignments and targeted integration windows for the D_2O extracts (300 MHz, TSP-d_4 = 0, apparent pH 5.3–5.9). 37
Table 9. Cross-peaks of the multiplicity-edited ¹H–¹³C gHSQC spectrum of the D_2O extract of an Italian Pecorino Romano (A-004. 39
Table 10. Lipid descriptors of the milk-fat reference population (24 cow's-milk products without lipolysis). 44
Table 11. Group medians of the region ratios, of the Hanganu–Chira descriptors as published (D1, D2, D3) and in water-robust…. 48
Table 12. Group medians of the composition obtained by line fitting and of the minor markers. 48
Table 13. Products flagged for non-dairy fat. 54
Table 14. Variation between separate retail lots of three products in the lipid-only set (mean ± standard deviation). 55
Table 15. Group medians of the targeted aqueous features, per 10^4 of total spectral area and per proton. 59
Table 16. Products titrated at two doses of Cs-EDTA. 65
Table 17. Soluble minerals in the aqueous extracts of the second-campaign products. 66
Table 18. Fully nested cross-validation on the 42 reference products. 76
Table 19. Simulated blends. 80
Table 20. Evidence summary for the 40 retail grated products and the two products with conflicting records (A-041 and B-040). 84
Table 21. Comparison of the approaches on the present data. 87
Table 22. Anonymized register of the 84 products with paired $CDCl_3$ and D_2O extracts. 91
Table 23. Effect of each preprocessing factor on the cross-validated score, averaged over all other factors (292 combinations. 93
Table 24. Cross-validated balanced accuracy of the first evaluation (mean ± s.d. 94
Table 25. One-class models. 94

1. Introduction

1.1 Economically motivated adulteration of grated hard cheeses

Hard grating cheeses are among the most valuable dairy commodities per unit mass, and their value is created slowly: a Parmigiano Reggiano wheel is held for at least twelve months, a domestic Parmesan for at least ten, and through that period the manufacturer carries the cost of milk, labor, storage and moisture loss[1–3]. Once the cheese has been grated, the visual and textural cues by which a buyer might judge it are lost, and the product becomes a dry, free-flowing powder whose composition can be altered without any change in appearance. Grated Parmesan and Romano have accordingly been named repeatedly in surveys of economically motivated adulteration, alongside milk powders, butter and other dairy fats[4–8].

The forms of adulteration that are available to a dishonest or careless manufacturer differ in their chemistry, and this difference governs the analytical approach. The cheese may be diluted with an insoluble filler, most commonly powdered cellulose, which is permitted in grated cheeses only as an anticaking agent at the level needed for that function[9]. It may be extended with a cheaper natural cheese, with trimmings, or with young cheese that has not received the curing time required by the standard of identity[10,11]. It may be blended with processed cheese, with a cheese analogue or with a so-called cheese product, in which milk fat has been replaced wholly or partly by vegetable fat and in which the protein network has been dispersed by emulsifying salts[12–15]. An enforcement action brought in the United States between 2012 and 2016, and the product testing reported by the press at the time, showed that all of these routes had been used in products sold as 100% grated Parmesan, some of which were reported to contain little or no Parmesan at all[16].

In the United States, Parmesan is defined as a cow's-milk cheese cured for not less than ten months and containing not more than 32% moisture and not less than 32% milk fat in its solids, and Romano as a cheese made from cow's, sheep's or goat's milk, cured for not less than five months, with not more than 34% moisture and not less than 38% milk fat in its solids; the name Pecorino Romano is reserved for the sheep's-milk cheese[10,11]. A grated cheese must be made from the cheese whose name it carries, with only antimycotics, anticaking agents, spices and flavorings that do not simulate cheese flavor as optional ingredients[9]. A grated product that contains vegetable fat, emulsifying salts, added lactose-bearing dairy solids or an unripened cheese base therefore falls outside these definitions irrespective of how it is labeled, and each of these departures leaves a chemical trace that can be sought by ¹H NMR.

1.2 NMR spectroscopy in dairy authentication

The reference method for the purity of milk fat is the gas-chromatographic analysis of intact triacylglycerols, evaluated through the formulae developed by Precht and adopted as ISO 17678[17–19]. It is sensitive and well validated for milk, cream, butter and milk powder, but it requires the isolation and chromatographic separation of the fat, it reports nothing about the non-fat phase of a cheese, and its scope excludes milk fat obtained from cheeses that show increased lipolysis, because a false-positive result can occur there. High-resolution ¹H NMR offers a complementary route. In a single spectrum of a chloroform extract, acquired in a few minutes without derivitization, the acyl chains of a fat are resolved into about ten groups of protons whose integrals are connected by simple stoichiometry, so that unsaturation, average chain length and, in the special case of milk fat, the butyryl content can be read directly[20–27].

Milk fat is unusual among edible fats in carrying about 9–12 mol % of butyric acid, esterified almost exclusively at the sn-3 position of the triacylglycerols, together with caproic and other short- and medium-chain acids, small amounts of conjugated linoleic acid and a characteristic set of minor monoenoic acids[28,29]. The terminal methyl triplet of the butyryl group is shifted downfield of the methyl triplet of all longer chains

and is therefore resolved in the ¹H spectrum, which has made it the principal NMR marker of milk fat in ¹H and ¹³C work on milk, butter and cheese[30–34]. Tociu and co-workers derived a rapid scheme that returns four fatty-acid classes from five integrals[35], and Hanganu and Chira extended it with an alternative normalization, with a split of the bis-allylic signal into linolenic and linoleic components and with three integral-ratio descriptors intended for cases in which conventional fraud detection fails[36]. The same spectra have been used to characterize cheese lipids, to determine conjugated linoleic acid, to recognize imitation cheese and to compare extraction procedures[37–47].

The water-soluble fraction of cheese is equally informative. Its ¹H spectrum is dominated by lactate and by the free amino acids released during ripening, with organic acids, residual sugars, amines and small peptides as minor components, and it has been used to follow the ripening and to trace the origin of Parmigiano Reggiano, Grana Padano, Emmental, Fiore Sardo, Gouda, Toma Piemontese, mozzarella and other cheeses, both in extracts and by high-resolution magic-angle spinning of the intact cheese[48–64]. Reviews of NMR in milk and dairy analysis and in food authentication more generally describe both the scope of the technique and the conditions under which non-targeted fingerprints may be relied upon[65–78].

The combined use of both fractions has a direct precedent. Maestrello and co-workers analyzed the aqueous and the lipid fraction of 144 cheeses at 600 MHz in order to distinguish Grana Padano of protected designation from Italian and foreign competitors, and found that a random-forest model built on both fractions performed better than a model built on either one, with lactate, several amino acids and lipid signals among the discriminating variables[79]; the same group followed the ripening and the production variables of Grana Padano in both fractions[52]. The present work does not claim the paired approach as new. Its contribution is narrower and specific to the application: retail grated Parmesan and Romano products sold in the United States are examined for the replacement of milk fat and for an aqueous phase that is not that of a ripened cheese, at 300 MHz, with a ratiometric evaluation that separates lipolysis from substitution and with reference classes that include processed cheese, declared vegetable-fat products and plant-based analogues. The two extracts answer different questions. The lipid extract shows whether the fat has the composition of milk fat and, if it has not, allows the replaced share and the gross composition of the replacing fat to be estimated. The aqueous extract shows whether the water-soluble phase is that of a ripened hard cheese, with its extensive proteolysis and exhausted sugars, or that of an unripened, processed or analogue material. Neither question can be answered from the other extract, and a product can fail one test while passing the other.

1.3 Scope and objectives of this study

In this report the two acquisition campaigns that were evaluated separately in the earlier versions of this work are combined into a single study of the 84 products for which both a $CDCl_3$ extract and a D_2O extract had been recorded under one sample code. The set comprises reference cheeses of known type, declared vegetable-fat and plant-based products, and 40 retail grated Parmesan and Romano products whose composition was to be established. Fourteen reference fats and oils, 24 amino-acid standard spectra, 55 spectra of aqueous extracts to which cesium EDTA had been added, one ¹H–¹³C gHSQC spectrum and 43 further lipid spectra without a matching aqueous extract were used in support.

Five objectives were set. The first was to document the spectra themselves, with annotated reference spectra, assignment tables and the chemistry that underlies each signal. The second was to evaluate the lipid spectra ratiometrically with a single set of integration regions lettered A to J, with region G divided into its linolenic (G_1) and linoleic (G_2) parts and region C divided into the sharp line of the long methylene chains (C_1) and the broader envelope of methylene groups adjacent to other environments (C_2), and to apply the published schemes of Tociu and of Hanganu and Chira alongside line fitting of the overlapped multiplets.

The third was to extract from the aqueous spectra the targeted markers of ripening, of processing and of formulation, including soluble calcium and magnesium through their EDTA complexes and the short-chain acids of propionic and butyric origin. The fourth was to compare a broad set of chemometric and machine-learning approaches across binning, alignment, normalization and scaling options, with validation appropriate to a small data set. The fifth was to bring the three lines of evidence together for each retail product and to state what the two extracts show, with the claims limited to what the data support.

The study is a screening exercise and not a validated method. No internal concentration standard was added to the extracts, so the results are expressed as ratios within a spectrum; each product was extracted and measured once; and no sample was analyzed in parallel by a reference method. Where an absolute scale has been attached, as for soluble calcium, it rests on an external standard and on assumptions that are stated explicitly and marked as provisional.

2. Cheese Types Represented in the Study

The chemistry that the two extracts record is the product of the milk, the starter and secondary flora, the make procedure and the time and conditions of ripening. The principal types in the sample set are described here in terms of those factors, and the signatures expected in the two spectra are collected in Table 1. General accounts of cheese science and of the Italian cheeses in particular have been drawn upon throughout[1–3,80–83].

2.1 Grana-type cheeses: Parmigiano Reggiano, Grana Padano, Piave and domestic Parmesan

Parmigiano Reggiano and Grana Padano are cooked, pressed, long-ripened cheeses of the Po valley made from raw, partly skimmed cow's milk with a natural whey starter of thermophilic lactobacilli. The curd is cut to rice-grain size, cooked at 53–56 °C, molded, brined for about three weeks and ripened for a minimum of twelve months in the case of Parmigiano Reggiano, whose zone of production is limited to Parma, Reggio Emilia, Modena and parts of Bologna and Mantua, and nine months in the case of Grana Padano, which may be made from the milk of silage-fed cows across a much larger area and with lysozyme added to control late blowing[2,84]. Piave, from the province of Belluno, belongs to the same technological family of cooked hard cheeses and is sold at several ages, of which the Vecchio grade is represented here. Domestic Parmesan is made from pasteurized milk with defined thermophilic cultures and must be cured for at least ten months[10].

Lactose is exhausted within the first days and galactose within the first weeks, so that a mature Grana-type cheese contains no free sugar. Citrate is consumed during ripening. Proteolysis by the cell-envelope proteinases and peptidases of the lysed starter is deep and continues for years, and free amino acids come to represent roughly a fifth to a quarter of the total nitrogen, with glutamate, lysine, leucine, valine and proline prominent; pyroglutamate accumulates by cyclization of glutamate, and tyrosine eventually crystallizes[48,49,51,85–87]. Lipolysis is moderate, because the milk is raw but the cooking temperature is high, and the free butyric acid of Parmigiano Reggiano is nonetheless appreciable after two years. The fat is ordinary cow's milk fat.

2.2 Sheep's-milk cheeses: Pecorino Romano, Manchego and a sheep's-milk table cheese

Pecorino Romano is made in Lazio, Sardinia and the province of Grosseto from whole sheep's milk coagulated with lamb rennet paste, which carries pregastric lipase and gives the cheese its piquant, strongly lipolyzed character; the curd is cooked at 45–48 °C, dry-salted repeatedly and ripened for at least five months for the table and eight for grating[2,88]. Manchego is an uncooked, pressed cheese of La Mancha made

from the milk of the Manchega breed and ripened from two months to two years. Sheep's milk fat is richer than cow's milk fat in caproic, caprylic and capric acids, and the fat of animals kept on pasture is richer in conjugated linoleic acid and in α-linolenic acid[89–91]. In the United States the name Romano, without the qualifier Pecorino, may be applied to a cow's-milk cheese[11].

2.3 Swiss-type and alpine cheeses: Emmental, Gruyère and Asiago

Emmental is defined by its propionic fermentation. After the lactic fermentation by thermophilic starters, the wheels are held in a warm room at 20–24 °C, where Propionibacterium freudenreichii converts lactate into propionate, acetate and carbon dioxide in a molar ratio close to 2:1:1 for every three moles of lactate consumed, and the gas forms the eyes[92,93]. Gruyère is made in a similar way but is smear-ripened at lower temperature, and its propionic fermentation is slight or absent. Asiago is sold both as Pressato, a fresh cheese of a few weeks, and as the cooked, aged d'Allevo type, and the two differ as much from one another as a young cheese does from a Grana. Alpine summer milk is distinguished by high conjugated linoleic acid and n-3 fatty acids[91,94].

2.4 Washed-curd, Dutch-type and Cheddar-type cheeses

Gouda and the related Mimolette are made with mesophilic starters that include citrate-fermenting lactococci and leuconostocs. Part of the whey is replaced with water during the make, which removes lactose and limits acidification, and citrate is metabolized to diacetyl, acetoin, acetate and carbon dioxide within the first weeks[1,95]. Long-ripened Gouda, of 18 and 26 months in this set, develops a free amino-acid pool comparable to that of the Grana cheeses. The Cheddar-type cheese in the set is an Irish product of the sweet style, which is commonly made with thermophilic adjunct cultures and is ripened for a year or more[55,96].

2.5 Young, pasta filata and brined cheeses

Low-moisture mozzarella is acidified by Streptococcus thermophilus, most strains of which ferment only the glucose moiety of lactose and release galactose into the curd, where it persists; this residual galactose is the cause of excessive browning on baking and is a recognized marker of the type[97,98]. Fresh mozzarella may instead be acidified directly with citric or acetic acid, in which case almost no lactate is formed. Monterey Jack and American Muenster are mild, young cheeses made with mesophilic starters and sold after a few weeks. Feta is ripened and stored in brine. In all of these the free amino-acid pool is small, and lactose, galactose or citrate may remain[57–59].

2.6 Blue-veined cheeses

In Stilton and in the buffalo-milk blue cheese of the set, Penicillium roqueforti grows through the pierced paste, consumes lactate, raises the pH and releases powerful lipases and proteinases. Blue cheeses are consequently the most lipolyzed of all natural cheeses, with free fatty acids that can exceed a tenth of the fat, and they are extensively proteolyzed as well[81,83]. Buffalo milk fat is somewhat more saturated than cow's milk fat but carries the same butyryl marker.

2.7 Processed cheese, cheese products and imitation cheese

Pasteurized process cheese is made by heating and shearing natural cheese with emulsifying salts, chiefly sodium citrates and sodium phosphates, at up to 3% of the product; process cheese food and the various products sold as singles or prepared cheese product may in addition contain whey, skim-milk solids, milk-protein concentrate and other dairy ingredients, which bring lactose with them[12,14,15]. The emulsifying salts

sequester calcium from the casein network, hydrate and disperse the protein and allow it to emulsify the fat as fine droplets[99,100]. Imitation or analogue cheese is made in the same equipment from rennet casein or caseinates, vegetable fat and emulsifying salts, with or without a proportion of natural cheese, and enzyme-modified cheese is widely used to supply flavor[13,101]. The declared products of this kind in the present set comprise grated Parmesan- and Romano-style cheese products, a reduced-fat grated topping labeled as imitation cheese, a grated blend that declares palm oil, starch and cellulose alongside Parmigiano Reggiano, and a prepared cheese product in loaf form.

2.8 Plant-based cheese analogues

The plant-based slices and snack round in the set contain no dairy ingredient. Products of this kind are structured with coconut oil, whose lauric and myristic acids supply the solid fat, and with modified starches, and they are acidified and flavored with organic acids[102]. They serve here as the negative control for every dairy marker.

2.9 Retail grated Parmesan and Romano products

Forty retail grated products labeled as Parmesan, Romano, Pecorino Romano or blends of Parmesan and Romano make up the test set. Most were shelf-stable products in shaker containers, a few were refrigerated, one was grated in the store from a domestic block and two were imported grated Pecorino Romano. Typical ingredient statements list the cheese together with powdered cellulose and potassium sorbate. Because neither cellulose nor intact protein is extracted into $CDCl_3$ or D_2O, the present method does not measure the cellulose content, which is the adulterant on which public attention has centered; it addresses instead the identity of the fat and of the water-soluble phase, which are the properties altered when a cheese product or analogue is substituted for cheese.

Table 1. Cheese types represented in the study, the features of manufacture and ripening that govern their chemistry, and the signatures expected in the D_2O and $CDCl_3$ extracts. Class codes are those used in the sample register (Table 22) and in the figures.

Class	Types in this study (n)	Manufacture and ripening	Expected D_2O signature	Expected $CDCl_3$ signature
GR	Parmigiano Reggiano, Grana Padano, Piave Vecchio, domestic Parmesan (6)	Raw or pasteurized cow's milk, thermophilic starter, cooked and pressed curd, brine-salted, 9 to more than 24 months	Very high free amino acids; lactate; no lactose, galactose or citrate; succinate and pyroglutamate; little dispersed lipid	Cow's milk fat; little lipolysis, with free butyric acid rising in long-ripened cheese
SH	Pecorino Romano, Manchego, sheep's-milk table cheese (4)	Sheep's milk; lamb rennet paste in Pecorino Romano; dry salting; 2 months to more than a year	High free amino acids in Pecorino Romano; high lactate; free butyrate	Sheep's milk fat: higher C6–C10, CLA and n-3; lipolysis in Pecorino Romano
AL	Emmental, Gruyère, Asiago aged and fresh (4)	Thermophilic starter, cooked curd; propionic fermentation in Emmental; smear in Gruyère	Emmental: propionate and acetate with depleted lactate; others as Grana or as young cheese	Cow's milk fat; high CLA and n-3 in alpine milk
GC	Gouda young, aged and extra aged, smoked Gouda, Mimolette, Irish Cheddar-type (7)	Mesophilic starter with citrate fermenters; washed curd (Gouda) or cheddaring; up to 26 months	High free amino acids when aged; citrate consumed; no sugars	Cow's milk fat; little lipolysis
BB	Blue Stilton, buffalo-milk blue, feta (3)	Penicillium roqueforti (blue) or brine ripening (feta)	Blue: lactate depleted, free amino acids high; feta: lactose and low amino acids	Blue: strong lipolysis, free acids and partial glycerides; butyryl marker retained
YF	Low-moisture and fresh mozzarella, Monterey Jack, Muenster (5)	Short make, days to weeks; pasta filata stretching or mild mesophilic fermentation	Very low free amino acids; lactate high, or absent after direct acidification; galactose, lactose and citrate present	Cow's milk fat, intact
PC	American slices and singles (4)	Natural cheese and dairy solids heated with citrate and phosphate emulsifying salts	Very low free amino acids; citrate; lactose and galactose; fat dispersed into the aqueous extract; low soluble Ca	Milk fat, intact
DV	Grated cheese products, imitation topping, declared palm-oil blend, prepared cheese loaf (6)	Casein or cheese base with vegetable fat, emulsifying salts, starch or dairy solids, as declared	As processed cheese, with large sugar signals; ripened profile only where real cheese predominates	Butyryl marker reduced or absent; linoleate and chain length raised
PB	Plant-based slices and snack round (3)	Coconut oil, starches, plant protein, organic acids	No dairy metabolites; glucan envelope; added acids	Lauric fat: short mean chain length, almost no unsaturation, no butyrate, no cholesterol
T	Retail grated Parmesan, Romano, Pecorino Romano and blends (40)	Labeled as grated hard cheese; composition to be established	To be established	To be established

3. Materials and Methods

3.1 Samples and anonymization

The study set comprises every product from the two acquisition campaigns for which both a $CDCl_3$ and a D_2O extract had been recorded under one sample code: 41 products from the first campaign, coded A-001 to A-043 with A-023 and A-024 absent from the paired set, and 43 products from the second campaign, coded B-001 to B-043. The codes preserve the acquisition numbering of each campaign so that every result can be traced to the laboratory record. Manufacturers, brands and retailers are not identified in this report, in its figures or in the companion workbook; each product is described only by its cheese type and style, form, milk species, stated origin and any age claim on the label (Table 22, Appendix A). The course of the work, from the weighed portion to the integrated interpretation, is outlined in Figure 1.

Forty-two products of unambiguous identity were assigned to nine reference classes: Grana-type hard cheeses (GR, 6), sheep's-milk hard and semi-hard cheeses (SH, 4), Gouda- and Cheddar-type cheeses (GC, 7), alpine, Swiss and Asiago cheeses (AL, 4), young and fresh cheeses (YF, 5), blue-veined and brined cheeses (BB, 3), processed cheeses (PC, 4), plant-based analogues (PB, 3) and declared vegetable-fat products (DV, 6). The 40 retail grated products (T) were treated throughout as test samples of unknown

composition and were never used to train a model. Two further products, A-041 and B-040, carried lipid and aqueous records that described different products; they were held out with the test samples and are discussed in Section 8.4. For one retail product, B-015, the two records agree that the product is a retail grated cheese but name it differently, and it was retained in the test set.

Three supporting data sets were used. The reference panel of fats and oils comprised palm oil, beef, bison and lamb tallow, pork lard, wild boar fat, chicken fat, duck fat, two extra-virgin olive oils, two avocado oils, pumpkin seed oil and walnut oil. The amino-acid library comprised 24 spectra in D_2O that cover the twenty proteinogenic amino acids and trans-4-hydroxyproline. Forty-three lipid spectra for which no matching aqueous extract existed, coded H-001 to H-043, were kept apart from the development of the method and were used only to test the lipid decision rule after it had been fixed (Section 5.7). According to the acquisition titles they comprise 20 spectra of a further series of retail grated and shredded products, one of which is a declared dairy-free Parmesan alternative (H-020); nine spectra of first-campaign products without an aqueous extract, among them a Parmigiano Reggiano block, a nine-month Pecorino Romano and the blue and white parts of a blue cheese; and 14 further spectra of retail grated products, eleven of which are repeated purchases of two products. In all, three retail products are represented by four, six and seven separate purchases that are distinguished in the titles by their best-by dates and lot codes; three of these samples are composites of two to six units of one lot, and two pairs carry the same date and code. These spectra are therefore further lots of products of a few store brands, three of which are also represented in the paired set, and they are not a sample of unrelated products.

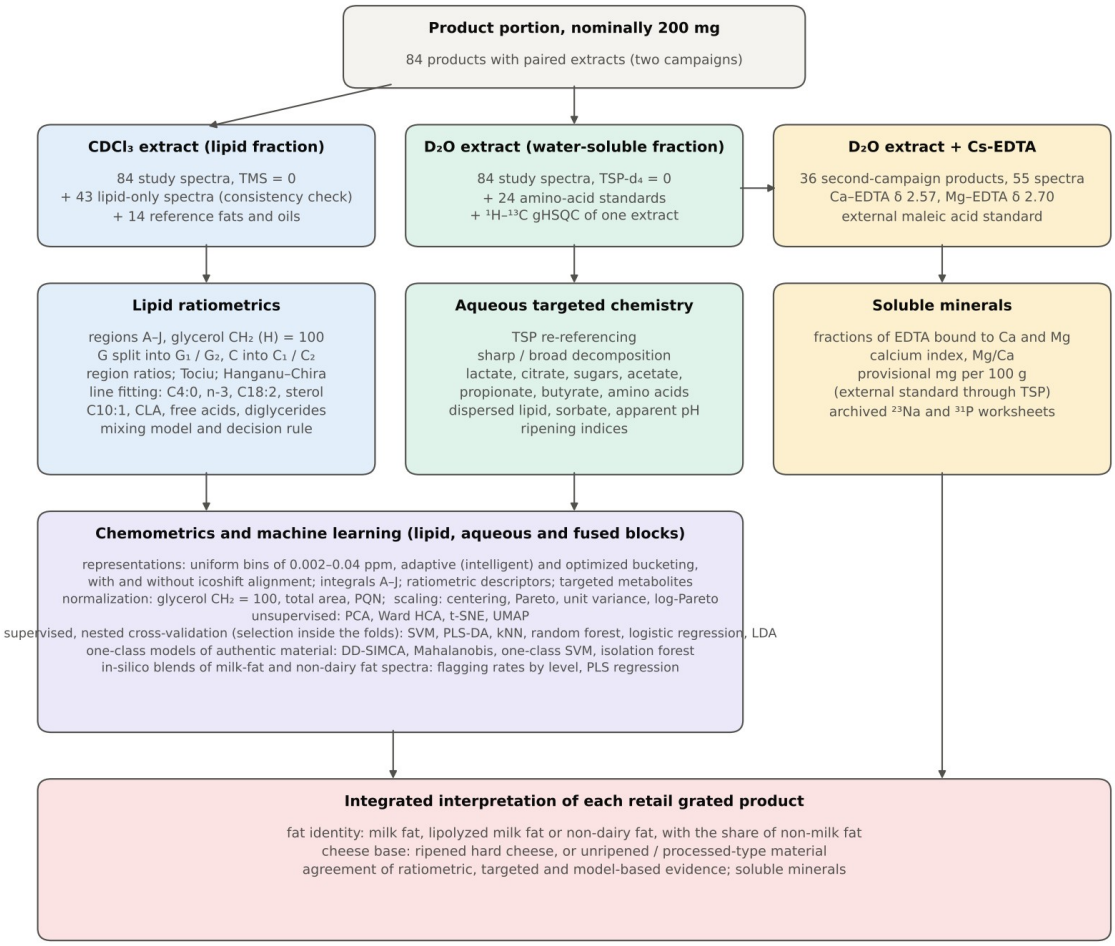

Figure 1. Analytical workflow. A portion of each product, nominally 200 mg, was extracted with $CDCl_3$ (lipid fraction) and with D_2O (water-soluble fraction); the D_2O extracts of 36 second-campaign products were titrated with Cs-EDTA. The three streams feed the ratiometric, targeted and multivariate evaluations that are brought together for each retail grated product in Section 8.

3.2 Extraction and NMR acquisition

A portion of each product, nominally 200 mg and weighed to 0.1 mg, was taken for extraction; the recorded masses, which appear in the titles of the acquisitions, ranged from 176 to 254 mg with a median of 204 mg. The lipid fraction was taken into $CDCl_3$ containing tetramethylsilane (TMS) and the water-soluble fraction into D_2O containing sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (TSP-d_4). Samples were prepared by first extracting with $CDCl_3$, vortex mixing, centrifuging and transferring 600 µL of supernatant to a 5 mm NMR tube. The remaining solids were then extracted with D_2O, vortex mixed, centrifuged and 600 µL of supernatant transferred to a 5 mm NMR tube for analysis. For 36 products of the second campaign, 100 µL of a cesium EDTA solution was added to the aqueous extract, and for seven of these a second spectrum was acquired after 200 µL, so that the calcium and magnesium of the extract would be converted into their EDTA complexes. Addition of Cs-EDTA was made to the D_2O extract samples in 5 mm tubes after an initial ¹H NMR spectrum was acquired.

The masses recorded with the two extracts were compared product by product, since the paired interpretation depends on both spectra belonging to one sample. Both masses are on record for 81 of the 84

products, and they are identical for 77. They differ for four (Table 2). For A-041 and B-040, the two products already held out because their lipid and aqueous records name different products, the title of the lipid acquisition repeats the descriptor and the mass of the preceding sample, although the spectrum itself is a different one; the lipid titles of these two were evidently not updated. For A-036 and B-003 the mass recorded with the aqueous extract equals that of the adjacent sample in the acquisition sequence (A-037 and B-002), which again suggests a title carried over from a neighboring acquisition, but an exchange of extracts cannot be excluded from the record alone. In both of these cases the adjacent sample belongs to the same category as the product concerned, so that the category-level conclusions of Section 8 would not change if the aqueous spectrum belonged to the neighbor. Because both extracts of a product were prepared from one weighed portion, as described above, a single mass is expected in the two records, and the term paired is used in this report in the sense of two extracts of one portion recorded under one sample code.

Table 2. Products for which the sample mass recorded with the lipid extract differs from that recorded with the aqueous extract (81 products with both masses on record; 77 identical). No mass was recorded for A-001 and A-002, and none with the aqueous extract of A-029.

Code	Mass, lipid record (mg)	Mass, aqueous record (mg)	Difference (mg)	Observation
A-036	202.7	201.2	1.5	Aqueous mass equals that of the next sample, A-037 (201.2 mg); both products are in category C
A-041	203.9	206.7	−2.8	Lipid title repeats descriptor and mass of A-040; record conflict, held out
B-003	199.9	196.4	3.5	Aqueous mass equals that of the preceding sample, B-002 (196.4 mg); both products are in category B
B-040	207.8	213.6	−5.8	Lipid title repeats descriptor and mass of B-039; record conflict, held out

The acquisition parameters were read from the parameter tables embedded in the processed spectral documents (Table 3). All spectra were acquired on a Varian Mercury-300 MVX spectrometer, identified in the acquisition records as MVX300, at a ¹H frequency of 299.99 MHz with a Varian 5 mm ATB probe at 27 °C. A single-pulse sequence (s2pul) was used throughout, with a pulse of 8 µs, a spectral width of 14,992.5 Hz (50 ppm), an acquisition time of 4.269 s (64,000 complex points) and a relaxation delay of 4 s, which gives a recycle time of 8.27 s. No presaturation frequency is recorded for any spectrum, so the residual water signal of the aqueous extracts was not suppressed. Between 16 and 128 transients were collected. Three records depart from these settings: the lipid spectra of A-001 and A-002 were acquired with a 5 µs pulse, and A-001 with a 6 s delay, and the external-standard spectrum of Section 3.8 with a delay of 25 s. Because the parameter tables are stored in a binary document format, the relaxation delay could be recovered for 157 of the 267 records and the pulse width for 259; no other values were found, but the figures should be confirmed against the instrument records.

Table 3. Acquisition parameters recovered from the parameter tables of the processed spectral documents. Common to all: Varian Mercury-300 MVX, 299.99 MHz, 5 mm ATB probe, 27 °C, single-pulse sequence (s2pul) without presaturation, spectral width 14,992.5 Hz, acquisition time 4.269 s, 64,000 complex points, zero-filled to 262,144 points on processing.

Data set	Spectra	Acquired (2023)	Pulse (µs)	Relaxation delay (s)	Transients
$CDCl_3$ extracts, first campaign and lipid-only spectra	84	March 30 to August 24	8 (A-001, A-002: 5)	4 (A-001: 6)	64 in most; 32 to 128
$CDCl_3$ extracts, second campaign	43	August 23 to September 15	8	4	32 or 64
D_2O extracts, first campaign	41	March 30 to July 31	8	4	32 to 128
D_2O extracts, second campaign, with external standard	44	August 23 to September 21	8	4 (external standard: 25)	64
D_2O extracts with Cs-EDTA	55	September 10 to 21	8	4	16 or 64

The free induction decays were processed in Mnova and stored, with their integrals, in PNA NMR Processor workspaces; the present evaluation started from those processed real spectra, zero-filled to 262,144 points (0.057 Hz per point), and applied no further apodization, phasing or baseline correction to the lipid spectra. Chemical shifts are referenced to TMS at δ 0.00 in $CDCl_3$ and to TSP-d_4 at δ 0.00 in D_2O.

Because no internal concentration standard was present, the recycle time and the flip angle determine how closely the relative integrals follow the molar ratios. With a recycle time of 8.27 s, a signal whose longitudinal relaxation time is 1.8 s is recovered to 99% even after a 90° pulse, and one of 3 s to 94%, or to 98% if the 8 µs pulse corresponds to 45°. The methylene, α-methylene and glycerol protons of triacylglycerols relax well within the shorter of these times, whereas the terminal methyl groups, and the methyl group of the short butyryl chain in particular, are the slowest of the acyl-chain protons, so that the butyrate content expressed per chain may be biased low by a few percent of its value. The relaxation times were not measured, and this bias is therefore not quantified. It is the same for every sample acquired under these conditions and cancels in the comparison of a product with the reference population, on which the decision rule and the dilution model rest, but it enters the absolute fatty-acid percentages and it would not cancel against values from another laboratory or method. In the aqueous spectra the slowly relaxing singlets, formate and TSP-d_4 above all, are attenuated relative to the amino-acid and lactate signals[103–105].

3.3 Data import and common axes

Each workspace is a compressed archive that holds a manifest with the sample names, the reference and spectral parameters and the integral definitions, the processed real spectra as 32-bit floating-point vectors of 262,144 points, and the source acquisition documents. The chemical-shift axis of each spectrum was rebuilt from the stored first point and increment, and all spectra were interpolated onto one axis from δ −0.5 to 10.5 at 0.0002 ppm per point (0.06 Hz). Lipid spectra were scaled so that region H, the glycerol sn-1 and sn-3 methylene protons, integrated to 100. Aqueous spectra were re-referenced individually to the maximum of the TSP-d_4 singlet and were scaled to their total area between δ 0.50 and 9.50 with the residual water region (δ 4.45–5.05) excluded; every aqueous index in this report is expressed per 10^4 units of that total area and is divided by the number of protons in the integrated multiplet, so that it is proportional to relative molar abundance within the extract.

The digitized spectra were checked against the integrals that had been recorded at the time of acquisition. When the original six integration windows were applied to the reprocessed data, the olefinic-to-glycerol ratio J/I agreed with the recorded values with a correlation coefficient of 0.9997 and a median difference of −0.3%, the methyl ratios B/A and B/I with r = 0.991, and the bis-allylic ratio G/I, which rests on a much smaller integral, with r = 0.90. The numerical evaluation in this report is therefore equivalent to the original bench integration where the two overlap, and all further processing was carried out on the digitized data.

3.4 Lipid integration regions and ratiometric schemes

Ten regions lettered A to J were integrated in every lipid spectrum, with region G divided at δ 2.786 into G_2 (linoleate) and G_1 (linolenate and other n-3 chains) and region C divided at δ 1.282 into C_1 and C_2 (Table 4 and Figure 4). Only this lettering is used in the present report. The region boundaries follow those of Hanganu and Chira[36], with the two splits added. Four ratios of regions serve as a first screen: B/A, the butyrate and ω-3 methyl signal relative to the terminal methyl of all other chains; B/I, the same signal relative to the glycerol sn-2 methine; J/I, the olefinic protons relative to the sn-2 methine; and G/I, the bis-allylic protons relative to the sn-2 methine.

Table 4. Integration regions of the $CDCl_3$ spectra (TMS = 0) and the number of protons that each contributes per acyl chain or per glycerol unit in the stoichiometric equations.

Region	δ range (ppm)	Protons integrated	Protons per unit	Remarks
A	0.785–0.918	Terminal CH_3 of all chains except butyrate and n-3	3 per chain	Triplet at δ 0.88
B	0.918–0.982	CH_3 of butyrate (δ 0.947) and of n-3 chains (δ 0.975)	3 per chain	Captures 96% of the butyrate triplet and 72% of the n-3 triplet, whose low-field line (δ 1.000) lies outside; also the flank of the δ 0.88 triplet
C_1	1.016–1.282	$(CH_2)_n$ of saturated chain segments	variable	Sharp line at δ 1.256
C_2	1.282–1.494	CH_2 adjacent to allylic carbons, to the β-methylene and to the terminal methyl; CH_2 of unsaturated chains	variable	Broader envelope, δ 1.29–1.35
D	1.494–1.733	β-CH_2 (–CH_2–CH_2–COO), including the butyrate sextet at δ 1.66	2 per chain	Contains the water singlet (δ 1.55)
E	1.929–2.211	Allylic CH_2 (–CH_2–CH=)	4 per double-bond system	δ 2.02; δ 2.07 for the n-3 terminus
F	2.217–2.418	α-CH_2 (–CH_2–COO)	2 per chain	Free-acid and partial-glyceride lines at δ 2.35–2.38
G_2	2.673–2.786	Bis-allylic CH_2 of linoleate	2 per chain	Triplet at δ 2.770
G_1	2.786–2.885	Bis-allylic CH_2 of linolenate and other n-3 chains	4 per chain (18:3)	Triplet at δ 2.805
H	3.994–4.418	Glycerol sn-1,3 CH_2 of triacylglycerols	4 per glycerol	Set to 100
I	5.178–5.313	Glycerol sn-2 CH	1 per glycerol	Multiplet at δ 5.26
J	5.313–5.471	Olefinic CH=CH	2 per double bond	Multiplet at δ 5.34

In the scheme of Tociu and co-workers the number of acyl chains is taken from the α-methylene protons, k = F/2, and four classes are obtained[35]:

$$SCSFA = 100\cdot B/(3k);\ PUFA = 100\cdot G/(2k);\ MUFA = 100\cdot E/(4k) - PUFA;\ LCSFA = 100 - SCSFA - PUFA - MUFA \quad (1)$$

Hanganu and Chira count the chains from the methyl protons instead, k′ = (A + B)/3, which removes the dependence on the α-methylene region, and obtain butyrate, di-unsaturated, mono-unsaturated and saturated fractions as[36]

$$But = B/(A + B);\ Di = 3G/[2(A + B)];\ Mono = 3(E - 2G)/[4(A + B)];\ Sat = 1 - But - Mono - Di \quad (2)$$

When the bis-allylic region is split, the linolenic and linoleic fractions follow separately as Ln = G_1/(4k′) and Di = G_2/(2k′), and the linolenic fraction can be subtracted from the butyrate estimate because the n-3 methyl triplet overlaps region B. The overlap is not complete. With a coupling of 7.5 Hz the three lines of the n-3 triplet lie at δ 0.950, 0.975 and 1.000, and the last of them falls outside the upper limit of the region at δ 0.982. From the fitted lineshapes the region was found to capture 72% of the n-3 triplet in the reference population (range over all 84 products 57 to 75%) and 96% of the butyrate triplet, whose high-field line lies close to the lower limit. The corrected butyrate fraction was therefore calculated as But − 0.72·Ln, and the form in which the whole of Ln is subtracted is reported alongside it. The three descriptors proposed by the same authors for the recognition of fats that pass the conventional tests are

$$D1 = (C + D + E + F + G)/F;\ D2 = (C + D + E + F + G)/(I + J);\ D3 = B/(A + B) \quad (3)$$

Region D contains the singlet of dissolved water, whose area in these extracts varied from almost nothing to more than twice the β-methylene signal in the high-moisture products. Because the β- and α-methylene groups each contribute two protons per chain, water-robust variants D1w and D2w were calculated by replacing D with F in Equation 3, and the difference D − F was retained as an index of the water carried into the chloroform extract.

Three further quantities were derived from the proton balance. The mean number of carbon atoms per acyl chain was obtained as

$$NC = 2 + (C + E + 2F + G)/(2k) + (J + I - H/4)/k \quad (4)$$

in which the first term counts the carboxyl and methyl carbons, the second the methylene carbons with the β-methylene taken equal to the α-methylene, and the third the olefinic carbons after subtraction of the sn-2 methine from the combined I + J area. The number of chains per glycerol, 2F/H, equals three for intact triacylglycerols and rises above three when free fatty acids are present. The ratios C_2/C_1 and G_1/G_2 complete the set and are the direct expression of the two splits: the first rises with unsaturation and falls as the chains become shorter and more saturated, and the second measures the n-3 to linoleate balance, which is high in ruminant fats from forage and low in seed and palm oils.

3.5 Line fitting of overlapped signals

Integration over fixed windows cannot separate the two triplets of region B or remove the tail of the intense methyl triplet on which they stand, and several minor signals of diagnostic value lie outside the ten regions. These were quantified by least-squares fitting of Lorentzian multiplets with fixed coupling patterns. The butyrate methyl triplet (δ 0.947, J = 7.4 Hz) and the n-3 methyl triplet (δ 0.975, J = 7.5 Hz) were fitted together with the tail of the δ 0.905 line; a third triplet near δ 0.960, which appears only in lipolyzed fats and is tentatively attributed to butyryl groups of partial glycerides or to free butyric acid, was included so that it would not be counted as esterified butyrate. The bis-allylic region was fitted with two triplets at δ 2.805 and δ 2.770. From the fitted areas, with k = F/2, the composition was expressed in mol % of acyl chains as C4:0, C18:2, n-3 polyunsaturated acids (as C18:3), monounsaturated acids from the allylic balance, MUFA = 100·E/(4k) − C18:2 − C18:3, and saturated acids by difference[21–23,33].

Four minor species were measured in the same way. The terminal vinyl protons of caproleic acid (9-decenoic acid, C10:1), a monoenoic acid that is specific to ruminant milk fat, were integrated between δ 4.900 and 5.030 on a local baseline. Conjugated linoleic acid was measured from the H10 resonance of rumenic acid between δ 5.895 and 5.985[38,40,41]. The α-methylene lines of free fatty acids and partial glycerides at δ 2.3725 and 2.3625 were fitted on the low-field side of region F and expressed as a percentage of chains, and 1,2-diacylglycerols were taken from the CH_2OH signal at δ 3.70–3.76[47,106]. Sterol was measured from the 18-CH_3 singlet of cholesterol at δ 0.680. At this point a feature of the spectra had to be recognized: the upfield ^{13}C satellite of the terminal methyl triplet, displaced by half the one-bond coupling of about 125 Hz, falls at δ 0.671 and overlaps the sterol singlet. A peak near δ 0.68 is therefore present in every fat, including the vegetable fats that contain no cholesterol, and its area is dominated by the satellite. The singlet was separated from the satellite triplet by fitting both, with the satellite constrained to the triplet pattern of the parent signal.

Representative fits are shown with their residuals in Figure 2. The quality of the three fits is not equal, and the difference bears on the weight that each fitted quantity can carry. In the methyl region the model reproduced the spectrum closely: the root-mean-square residual, as a percentage of the tallest signal in the fitted window, had a median of 0.9% over the 84 products (range 0.7 to 6.5%), the largest values belonging to the plant-based analogues, in which almost no signal is present. In the bis-allylic region the median was 3.7% (range 1.5 to 7.1%) and the residuals are structured, because two triplets on a straight baseline do not describe the envelope of linolenate, conjugated and other minor polyunsaturated chains; the fitted linoleate and n-3 fractions therefore carry a model error that is largest where the n-3 component is large, as in the milk fats, and smallest where linoleate dominates, as in the products with non-dairy fat. In the sterol region the median was 3.8% (range 2.5 to 10.1%), the misfit being concentrated on the sharp central line of the

satellite. Butyrate, on which the decision rule and the dilution model depend most, is thus the best determined of the fitted quantities, and the sterol value the least. The residual of every fit is listed in the companion workbook.

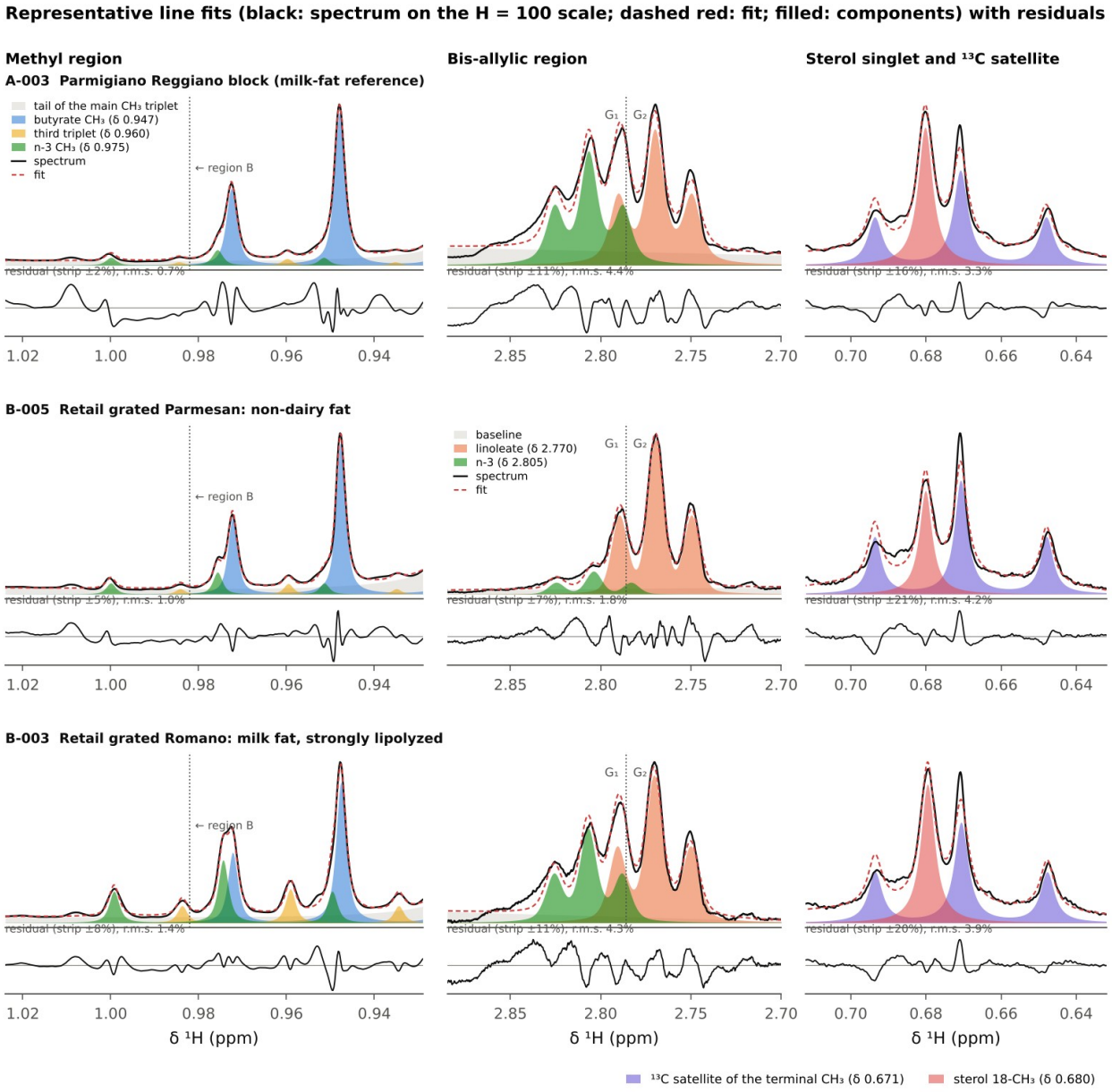

Figure 2. Representative line fits with residuals for a milk-fat reference (A-003), a retail grated product with non-dairy fat (B-005) and a strongly lipolyzed retail grated Romano (B-003). Left: methyl region with the butyrate triplet (δ 0.947), the third triplet of lipolyzed fats (δ 0.960), the n-3 triplet (δ 0.975) and the tail of the main methyl signal; the dotted line marks the upper limit of region B, outside which the low-field line of the n-3 triplet falls. Center: bis-allylic region with the linoleate (G_2) and n-3 (G_1) triplets and the division between the two regions. Right: sterol 18-CH_3 singlet and the ¹³C satellite of the terminal methyl signal. Each panel is drawn on its own full vertical scale; the strip below gives the residual on the scale stated, with its root-mean-square value as a percentage of the tallest signal.

3.6 Reference population, decision rule and mixing model

A milk-fat reference population was defined from the 24 cow's-milk reference products whose lipid spectra showed no sign of lipolysis: the six Grana-type cheeses, six of the seven Gouda- and Cheddar-type cheeses,

three of the four alpine cheeses, the five young and fresh cheeses and the four processed cheeses. The sheep's-milk and blue cheeses and the two lipolyzed cow's-milk references were excluded so that neither species nor lipolysis would widen the limits. Means and standard deviations of every descriptor in this population (Table 10) were used to express each product as a standard score, z.

A product was assigned to a lipid category by a fixed rule, which is set out with its thresholds and order of precedence in Table 5. Four quantities enter it: the standard score of butyrate; the number of shifts toward vegetable fat among four descriptors (linoleate, mean chain length and C_2/C_1 more than three standard deviations above the reference mean, G_1/G_2 more than three below); the caproleate content; and the presence of lipolysis, defined as free acids and partial glycerides above 3% of chains or more than 3.14 chains per glycerol. The conditions are tested in the order given and the first that is met decides. The distinction between substitution and lipolysis matters because lipolysis removes butyrate from the triacylglycerols preferentially, the sn-3 position being the first attacked, and a lipolyzed milk fat would otherwise be mistaken for a diluted one[83]. Two properties of the rule should be noted. A butyrate deficit that is accompanied neither by a compositional shift nor by lipolysis is not flagged, so that a foreign fat resembling milk fat in linoleate content and chain length, a ruminant tallow for example, would escape at moderate levels. And the limits were set on the same 84 products to which the rule was then applied, so that its error rates on other material are not known.

Table 5. Decision rules. Lipid rule: conditions are tested in order and the first that is met decides; z is the standard score against the milk-fat reference population (n = 24), and the value in parentheses is the corresponding limit in the units of the descriptor. Aqueous call and integrated category as applied to the retail grated products.

Order	Condition	Assignment
Lipid 1	C4:0 < 1.5 mol % and C10:1 < 0.10 mol % (or C10:1 not measurable)	Non-dairy fat with the milk-fat markers at the blank level; reported in the tables as "no milk-fat marker detected" when C4:0 < 0.72 mol %
Lipid 2	z(C4:0) < −3 (C4:0 < 7.66 mol %) and at least two of: z(C18:2) > 3 (> 3.58 mol %), z(NC) > 3 (> 14.77), $z(C_2/C_1) > 3$ (> 0.526), $z(G_1/G_2) < -3$ (< 0.38)	Non-dairy fat
Lipid 3	z(C4:0) < −2.5 (< 7.90 mol %), at least one of the four shifts, no lipolysis, and z(C10:1) < −1.5 (< 0.38 mol %) or C10:1 not measurable	Low-level non-dairy fat indicated
Lipid 4	z(C4:0) < −2 (< 8.13 mol %) and more than 3.5 chains per glycerol	Milk fat, strongly lipolyzed
Lipid 5	z(C4:0) < −2 and lipolysis (free acids and partial glycerides > 3% of chains, or more than 3.14 chains per glycerol)	Milk fat, lipolyzed
Lipid 6	At least one of the four shifts	Milk fat (composition outlier)
Lipid 7	None of the above	Milk fat
Aqueous flags	Free amino-acid index < 21.4 (half the median of the 21 ripened reference cheeses); citrate > 10; free sugars (lactose, galactose and α-anomers) > 10; dispersed lipid > 401 (maximum of the ripened reference cheeses)	Low free amino acids; citrate; free sugars; dispersed lipid
Aqueous call	Low free amino acids with citrate or free sugars / neither low free amino acids, citrate nor free sugars / citrate or free sugars without low free amino acids	Unripened or processed-type / ripened / ripened, with citrate or sugars (dispersed lipid is recorded but does not enter the call)
Category	Lipid 3 / non-dairy fat with unripened or processed-type call / non-dairy fat with a ripened call / milk fat with unripened or processed-type call / lipolyzed milk fat with a ripened call / milk fat with a ripened call	F / C / D / E / B / A

The fraction of non-milk fat was estimated from the dilution of the two markers that are specific to milk fat, as $f = 1 - x/\bar{x}$, where x is the butyrate or caproleate content of the product and $\bar{x}$ the mean of the reference population (9.08 and 0.497 mol % of chains). Three contributions were combined in the standard error of each estimate:

$$s(f) = (100/\bar{x})\cdot\sqrt{[(x/\bar{x})^2\cdot\sigma_1^2\cdot(1 + 1/n) + \sigma_2^2]} \quad (5)$$

in which σ_1 is the standard deviation of the marker in the reference population of n = 24 products (0.47 and 0.076 mol %), which enters in proportion to the milk fat that remains, the term 1/n allows for the uncertainty of the reference mean, and σ_2 is the pooled standard deviation of the marker between separate retail lots of three products in the lipid-only set (0.33 and 0.019 mol %; Section 5.7). The last term stands in for the repeatability of extraction, acquisition and fitting, none of which was measured directly; because it also contains any real difference between lots, it is an upper bound on that repeatability, but it does not include a systematic difference in the recovery of milk fat and replacing fat on extraction, the relaxation bias of Section 3.2 where it differs between fats, or the model error of the fit. For a product without any non-milk fat the resulting standard error is 6.4 percentage points from butyrate and 16 from caproleate. The two estimates were combined as an inverse-variance weighted mean, in which the more variable caproleate carries the smaller weight. In version 2.0 of this report only the first term was used, with the consequence that the standard error approached zero as the marker approached zero.

A marker cannot be quantified below the level at which a fat without milk fat would show the same apparent signal. This floor was taken as the mean plus three standard deviations of the apparent marker content of 17 fats that contain no milk fat, the 14 reference fats and oils and the three plant-based analogues: 1.27 mol % for butyrate, the largest single value being 1.05, and 0.13 mol % for caproleate. A butyrate content at or below the floor is reported only as a lower bound on the non-milk fat, f ≥ 86%, and is not entered into a consensus; a caproleate content at or below its floor, which corresponds to f ≥ 74%, is left out of the consensus, which then rests on butyrate alone. Consensus values are not allowed to exceed 100%. The estimate is a mole fraction of acyl chains. Because the chains of the replacing fat are longer than those of milk fat, the corresponding mass fraction of the fat is higher by about three percentage points at 50%. With the fraction known, the composition of the replacing fat was back-calculated for each product from $x(\mathrm{ext}) = \bar{x} + (x - \bar{x})/f$.

3.7 Aqueous spectra: background separation and targeted features

The aqueous spectra contain three kinds of signal: sharp lines of small molecules, broad resonances of dispersed lipid, peptides and soluble protein, and a rolling baseline. Each spectrum was separated into a broad and a sharp component by a morphological opening with a 0.03 ppm window followed by smoothing, the sharp component being the difference between the spectrum and this broad background. Sharp-component areas were used for the amino acids and organic acids, and the broad component for the indices of dispersed lipid and peptide material.

Lactate and dispersed lipid overlap at δ 1.3 and were separated by fitting the lactate methyl doublet (J = 6.9 Hz), whose position was first located from the second derivative of the spectrum, together with a broad pseudo-Voigt line for the lipid $(CH_2)_n$ resonance centered between δ 1.26 and 1.33. Citrate was fitted as an AB quartet with J = 15.6 Hz, with its center and the separation of its two halves left free between δ 2.56 and 2.70 and between 30 and 60 Hz, since both depend on pH and on the binding of calcium and magnesium; citrate was reported only when all four lines were present in the expected intensity pattern. Acetate, formate, glycine, the methionine S-methyl and succinate were measured as tracked singlets, alanine, tyrosine, lysine, lactose (β-galactosyl H1, δ 4.455), α-galactose (δ 5.27) and the α-anomers at δ 5.23 as tracked multiplets of fixed pattern, and propionate and free butyrate by a joint fit of the propionate quartet (δ 2.18) and the butyrate α-methylene triplet (δ 2.155). Formate was confined to δ 8.4555 ± 0.005 because the H2 resonance of free histidine moves through the same region; the histidine line was located separately, and its shift was converted into an apparent pH with the limiting shifts of 8.66 and 7.74 ppm and a pK of 6.10. A free amino-acid index was defined as the sharp-component area of the leucine, isoleucine and valine methyl region, δ 0.925–1.070, divided by 18. Like every aqueous index it is a fraction of the total spectral area, and it is

lowered by anything that adds area to the spectrum; where this matters, the ratio of the index to lactate and the area relative to TSP-d_4 and sample mass are quoted as well. The windows are listed with the assignments in Table 8.

3.8 Calcium and magnesium from the EDTA complexes

EDTA forms 1:1 complexes with calcium and magnesium that exchange slowly on the chemical-shift time scale, so that the ethylene protons of the calcium complex, of the magnesium complex and of the free ligand give three separate singlets[107–111]. In the present extracts the calcium complex was observed at δ 2.565 and the magnesium complex at δ 2.70, and the free ligand as two pH-dependent singlets, one between δ 3.2 and 3.5 (ethylene, 4 H) and one between δ 3.55 and 3.86 (acetate, 8 H). The two complex singlets were fitted together with the citrate quartet, whose area was constrained to the citrate-to-lactate ratio of the native extract, and the free ligand was taken from the lines that were present after the addition and absent before it.

The fractions of the added EDTA bound to calcium and to magnesium, f(Ca) and f(Mg), are independent of any intensity standard. A calcium index was defined as f(Ca) multiplied by the volume of EDTA solution added, relative to 100 µL, and by 200 mg divided by the sample mass. The EDTA workspace holds 55 spectra of 36 products: one at 100 µL for each product, a second at 200 µL for seven, and for six of those seven a duplicate stored spectrum at each dose; the duplicates reproduced f(Ca) within 1% at 200 µL and within 4% at 100 µL and were averaged. A determination was accepted as reportable only when at least 20% of the ligand remained free and the calcium line was not broadened beyond 3.5 Hz. Determinations that failed this test were treated in version 2.0 as lower bounds. The seven products measured at both doses show that they are not: in every case the calcium index from the 100 µL spectrum, in which the ligand was exhausted, exceeded that from the 200 µL spectrum, in which free ligand was present, by 19 to 39% (Table 16 and Figure 21). Since the index is formed from fractions of the ligand, no change of the absolute scale can remove the reversal. In the exhausted spectra the total EDTA signal relative to TSP-d_4 was also only about half of that found per 100 µL in the reportable spectra, so that part of the ligand is not accounted for when calcium is in excess, probably because the complex lines are broadened by exchange. Determinations without free ligand are therefore reported as not reportable. 36 of the 55 spectra and 30 of the 36 products gave a reportable result; the six that did not are B-009, B-017, B-018, B-021, B-031 and B-034.

An absolute scale was attached provisionally through an archived external standard. In that measurement 5.40 mg of maleic acid gave an olefinic integral of 100 against 14.70 for the nine protons of TSP-d_4, which corresponds to 1.52 µmol of TSP-d_4 in the volume of deuterium oxide used for each extract. The median ratio of total EDTA to TSP-d_4 in the reportable spectra (24.3, interquartile range 22.6 to 25.9, full range 17.7 to 42.6) then gave 36.9 µmol of EDTA per 100 µL added, or about 0.37 mol L^{-1}, and the calcium and magnesium found were converted to milligrams per 100 g of product. This scale assumes that the same quantity of TSP-d_4 was present in every extract, that the whole extract was titrated and that the EDTA stock was constant. It carries two further weaknesses: the EDTA-to-TSP ratio varies between extracts by about ± 7% in its interquartile range and by far more at the extremes, and the external standard was acquired with a relaxation delay of 25 s whereas the extracts were acquired with 4 s, so that the slowly relaxing TSP-d_4 singlet is likely to be attenuated in the extracts and the dose overestimated. The absolute values are accordingly marked as provisional throughout, and only ratios between products are interpreted. The scale should be replaced by the known concentration of the stock when that is available. The archived determinations of sodium by ^{23}Na NMR and of soluble phosphorus by ^{31}P NMR on the second-campaign extracts, made against sodium chloride and dipotassium hydrogen phosphate external standards, were taken from the laboratory worksheets without re-evaluation and are used in Section 6.6 for comparison[112,113].

3.9 Alignment, binning, normalization and scaling

For the multivariate work the spectra were restricted to the informative regions. In the lipid spectra the windows δ 0.50–1.535, 1.60–2.95, 4.00–4.45 and 5.05–5.60 were retained, which excludes the water singlet; in the aqueous spectra the residual water and the three windows in which the preservative sorbate resonates were removed, so that a model could not learn to recognize a retail product by its preservative rather than by its cheese.

Small differences in chemical shift between spectra were treated by interval correlation-optimized shifting of the icoshift type[114,115], which was preferred to continuous warping because it moves whole multiplets rigidly and cannot distort a lineshape[116,117]. The interval borders were placed at minima of the maximum spectrum, at a spacing of about 0.08 ppm in the lipid spectra and 0.05 ppm in the aqueous spectra, and each segment was shifted by up to 0.010 ppm or 0.025 ppm, respectively, to maximize its correlation with the median spectrum; a segment was left unmoved when the best correlation was below 0.5, which prevents a signal that is absent from the target from being dragged onto a neighbor.

Seven binned representations were prepared from both the unaligned and the aligned spectra: uniform bins of 0.002, 0.005, 0.01, 0.02 and 0.04 ppm; adaptive, or intelligent, bucketing, in which the borders are placed at the local minima of the mean spectrum, buckets narrower than 0.004 ppm are merged, those wider than 0.04 ppm are divided and those that contain only noise are dropped[118,119]; and optimized bucketing, in which the borders of nominal 0.01 ppm bins are allowed to move to the nearest minimum[120]. Three knowledge-based representations were added: the integrals A to J with the G and C splits, a set of 18 ratiometric descriptors, and 24 targeted aqueous features. Row normalization was to region H = 100, to total area or by probabilistic quotient normalization[121,122], and column scaling was by mean-centering alone, Pareto scaling, unit-variance scaling or a logarithmic transformation followed by Pareto scaling[123–125]. All combinations that are meaningful for a given block were evaluated, 292 in total, and in every case the normalization and scaling parameters were calculated from the training samples only. Two elements of this preprocessing were, however, derived from all 84 spectra and lie outside any cross-validation: the alignment target, which is the median spectrum, and the borders of the adaptive and optimized buckets, which were placed on the mean spectrum. Neither uses the class labels, but both have seen the test spectra, and they were excluded from the fully nested evaluation described in Section 3.10.

3.10 Unsupervised, supervised and one-class modeling

Principal component analysis[126,127], Ward hierarchical clustering on the leading eight components[128], t-distributed stochastic neighbor embedding at a perplexity of 12[129–131] and uniform manifold approximation and projection with 12 neighbors and a minimum distance of 0.3[132] were applied to all 84 products, for the lipid block, the aqueous block and a mid-level fusion in which the first five principal-component scores of each block, scaled to equal total variance, were concatenated[133].

Supervised models were trained on the 42 reference products only, for three tasks: dairy against non-dairy fat (33 against 9), technology in three classes (ripened, fresh, processed or analogue; 24, 5 and 13) and style in the nine reference classes. Linear and radial-basis support vector machines[134], partial least squares discriminant analysis[135,136], k-nearest neighbors, random forests[137], L2-regularized logistic regression and linear discriminant analysis on five principal components were compared with scikit-learn[138]. Performance was estimated by nested cross-validation, with ten repetitions of a stratified three-fold outer loop and an inner loop for the selection of hyperparameters, and was reported as balanced accuracy; a label-permutation baseline was calculated for each task[139–142]. The models were then refitted on all reference products and applied to the test products.

In that first evaluation the hyperparameters, the scaling parameters and the principal components of the LDA pipeline were fitted inside the training folds, but four choices were not: the preprocessing adopted for each block, which had been selected from the 292 combinations on all 42 reference products; the alignment target and bucket borders mentioned above; the principal-component scores of the fused block, which were calculated once from all 84 products; and the ten-component space of the nearest-neighbor score used to compare the preprocessing options. An estimate obtained after such a selection is optimistic[139]. A second, fully nested evaluation was therefore carried out, in which every choice made from data lay inside the outer training fold. The candidate representations were restricted to those defined without reference to the data: unaligned uniform bins of 0.01, 0.02 and 0.04 ppm, the integrals A to J, the ratiometric descriptors and the targeted aqueous features. For each outer fold the representation, the normalization, the scaling, the classifier and its hyperparameters were selected together by an inner stratified cross-validation; missing descriptor values were imputed from the training fold; and for the fused block the two block-wise principal component models were fitted on the training fold only. The definitions of the knowledge-based features remain outside this protection, since the windows were chosen with sight of the spectra, although without reference to the labels.

Authentication is more properly a one-class problem than a discrimination problem, because the authentic class can be sampled whereas the possible adulterants cannot be enumerated[143,144]. One-class models were therefore built for authentic milk fat, from the 33 reference products that contain only dairy fat, and for ripened natural cheese, from the 21 reference products of classes GR, SH, GC and AL. Data-driven soft independent modeling of class analogy (DD-SIMCA) was applied to the binned spectra[145], a Mahalanobis distance with a shrinkage covariance estimate to small sets of descriptors (C4:0, C18:2, NC, C_2/C_1 and G_1/G_2 for the lipid; eight targeted features for the aqueous extract)[146,147], and a one-class support vector machine and an isolation forest to principal-component scores[148,149]. All acceptance limits were set at a nominal 1% false-rejection rate, and the actual rate was measured by leaving each authentic product out in turn.

The capacity of each method to flag a foreign fat was examined in silico. Each of the 24 spectra of the milk-fat reference population was combined, on the H = 100 scale, with 5 to 60% of the palm oil reference spectrum or of the mean spectrum of the three declared cheese products, and the 480 synthetic spectra were passed through the complete ratiometric procedure and through every multivariate detector, each of which was refitted without the product that served as the base of the blend. A partial least squares regression of the blended fraction on the binned spectra was calibrated on the blends of 16 base products and tested on those of the remaining eight[150]. Its calibration domain was defined by the Hotelling T^2 and residual Q statistics of the blends themselves, each evaluated against a model built without the base product concerned; a prediction was reported for a real product only when both statistics lay within the range of the blends, the predicted value lay within the calibrated range of 0 to 100% extended by twice the prediction error, and the product was not a lipolyzed milk fat, a condition that is not represented among the blends. Such simulations describe the response of the numerical procedures to a linear combination of two spectra. They contain no extraction step, no independent lot of cheese or fat and no measurement repeated on another day, and they do not establish a detection limit or a false-positive rate for real products, for which gravimetrically prepared blends and an independent set of authentic cheeses are required[151].

4. NMR Spectra and Assignments

4.1 The lipid fraction in $CDCl_3$

The ¹H spectrum of a cheese fat is that of a mixture of triacylglycerols, and every signal in it can be traced to a position along the acyl chain or on the glycerol backbone (Figure 3, Figure 4 and Table 6). Read from high field, the spectrum begins with the terminal methyl triplet of the saturated, oleic and linoleic chains at δ 0.88. In the first methylene of a butyryl chain the carbonyl group is only two bonds from the methyl carbon, and the methyl triplet is displaced to δ 0.947, clear of the main triplet; the methyl group of an n-3 chain, which is allylic to the terminal double bond, appears at δ 0.975. These two triplets overlap region B, which also takes in the low-field flank of the main triplet; the low-field line of the n-3 triplet, at δ 1.000, lies beyond the upper limit of the region. Their separation by line fitting is shown in Figure 3a and, with residuals, in Figure 2. The methylene envelope follows, with the sharp line of the saturated chain segments at δ 1.256 (C_1) and a broader shoulder between δ 1.29 and 1.35 (C_2) that contains the methylene groups next to allylic carbons, next to the β-methylene and next to the terminal methyl, together with the methylene groups of the unsaturated chains, whose shifts are spread by the neighboring double bond. The split of region C at δ 1.282 separates these two populations, and because the tall line belongs to region C_1 only the ratio C_2/C_1 is a compact measure of how much of the chain consists of methylene groups that are not in a long saturated run.

The β-methylene multiplet at δ 1.61 (region D) includes the sextet of the butyryl β-methylene at δ 1.66 and, in most extracts, the singlet of dissolved water at δ 1.55[152,153]. Allylic methylene protons resonate at δ 2.02 (region E) and α-methylene protons at δ 2.31 (region F), the latter as two overlapping triplets for the sn-1,3 and sn-2 chains; free fatty acids and partial glycerides add lines on the low-field side at δ 2.35–2.38. The bis-allylic protons of linoleate give a triplet at δ 2.770 and those of linolenate, which has two bis-allylic methylene groups, a triplet at δ 2.805; the divide between G_2 and G_1 at δ 2.786 passes between them, and the two fitted triplets are shown in Figure 3b. The glycerol sn-1 and sn-3 methylene protons form the two doublets of doublets at δ 4.14 and 4.29 (region H), the sn-2 methine a multiplet at δ 5.26 (region I) and the olefinic protons a multiplet at δ 5.34 (region J)[20,24,26,30,32].

Four groups of weak signals lie outside the ten regions and carry information that the regions do not. Between δ 4.90 and 5.03 the terminal vinyl protons of caproleic acid appear as two multiplets, with the internal vinyl proton at δ 5.80; this acid, 9-decenoic acid, is formed in the mammary gland and is absent from vegetable fats. The conjugated diene of rumenic acid gives resonances at δ 5.65, 5.94 and 6.28. The 18-methyl singlet of cholesterol at δ 0.680 sits on the upfield ¹³C satellite of the terminal methyl triplet, as described in Section 3.5 and shown in Figure 3d. A weak multiplet at δ 5.08 contains the sn-2 methine of 1,2-diacylglycerols and the ¹³C satellite of the olefinic signal. Every panel of Figure 3 is drawn on its own full vertical scale, so that no signal is truncated; the methylene envelope, which is about five times taller than any other signal, is shown complete in the uppermost panel and is omitted only from the expanded panel beneath it.

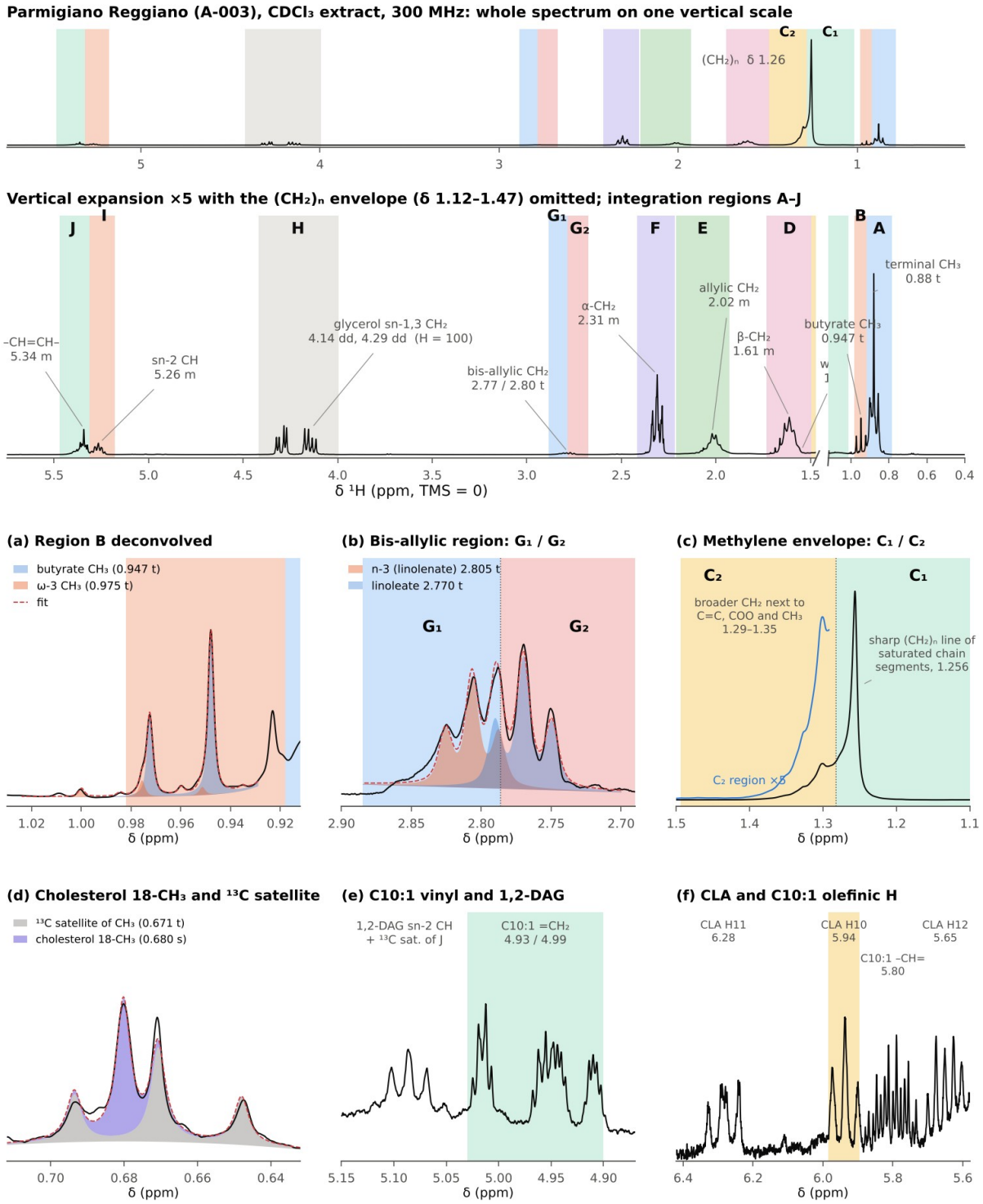

Figure 3. ¹H NMR spectrum (300 MHz, $CDCl_3$, TMS = 0) of the lipid extract of Parmigiano Reggiano (A-003), scaled to region H = 100. Top: the whole spectrum on one vertical scale, with the integration regions shaded. Middle: fivefold vertical expansion with the methylene envelope (δ 1.12–1.47) omitted, regions A–J lettered. Insets, each on its own full scale: (a) region B with the fitted butyrate and n-3 methyl triplets; (b) bis-allylic region with the fitted linolenate (G_1) and linoleate (G_2) triplets; (c) methylene envelope with the C_1/C_2 divide and a fivefold expansion of the C_2 region; (d) sterol 18-CH_3 singlet resolved from the ¹³C satellite of the terminal methyl; (e) terminal vinyl protons of caproleate (C10:1); (f) conjugated-diene and caproleate olefinic resonances.

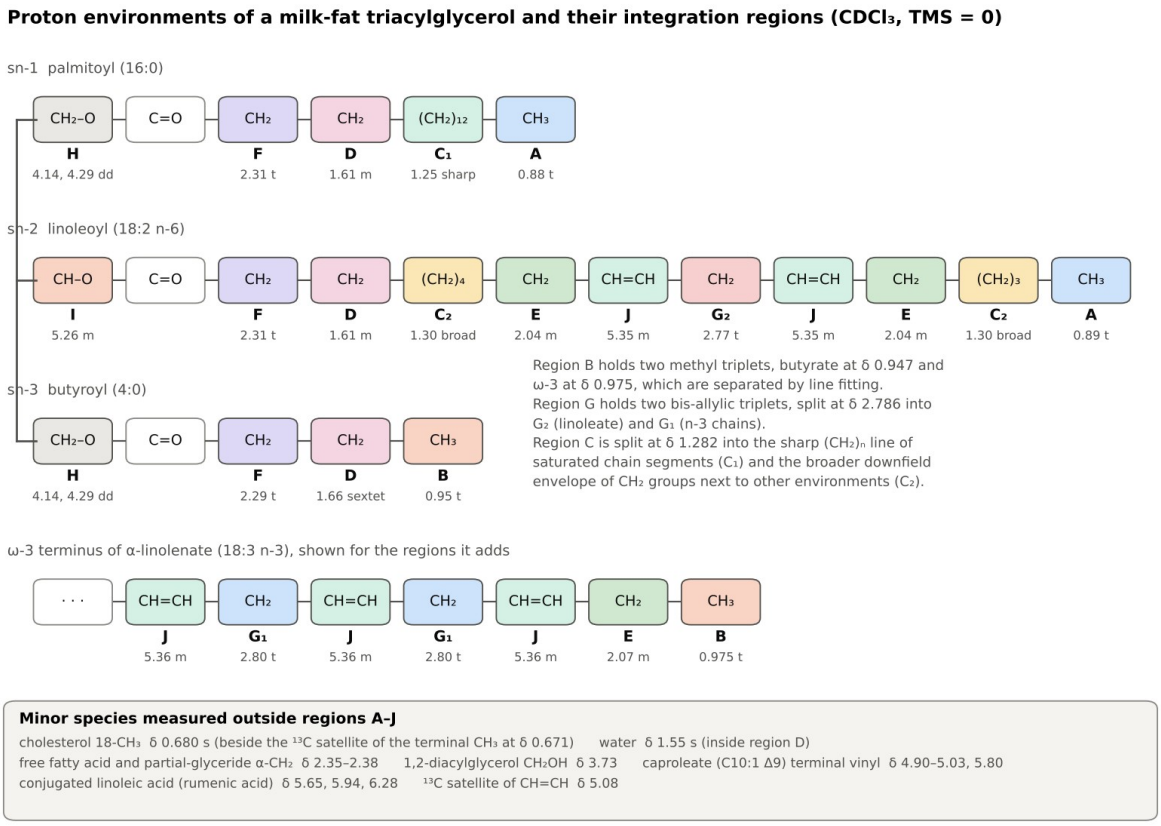

Figure 4. Proton environments of an illustrative milk-fat triacylglycerol (sn-1 palmitoyl, sn-2 linoleoyl, sn-3 butyroyl) and of the n-3 terminus of α-linolenate, with the integration region and chemical shift of each group. Minor species measured outside regions A–J are listed below.

Table 6. ¹H NMR assignments for cheese lipid extracts in $CDCl_3$ (300 MHz, TMS = 0). Multiplicities: s singlet, d doublet, t triplet, dd doublet of doublets, m multiplet, br broad. Coupling constants are approximate values measured on the present spectra.

δ (ppm)	Multiplicity, J (Hz)	Assignment	Region
0.671	t, 6.9	Upfield ^{13}C satellite of the terminal CH_3 triplet ($^1J(CH) \approx 125$ Hz)	–
0.680	s	Sterol 18-CH_3 (cholesterol in milk fat; phytosterols in vegetable oils)	–
0.88	t, 6.9	Terminal CH_3 of saturated, n-9 and n-6 chains	A
0.947	t, 7.4	CH_3 of butyryl (C4:0) in triacylglycerols	B
0.960	t, 7.4	CH_3 of butyryl in partial glycerides or free butyric acid (tentative)	B
0.975	t, 7.5	CH_3 of n-3 chains (α-linolenate)	B
1.256	br s	$(CH_2)_n$ of saturated chain segments	C_1
1.29–1.35	m	CH_2 adjacent to allylic, β-methylene and terminal methyl carbons; CH_2 of unsaturated chains	C_2
1.55	s	Dissolved H_2O	D
1.61	m	β-CH_2 (–CH_2–CH_2–COO)	D
1.66	sextet, 7.4	β-CH_2 of butyryl	D
2.02	m	Allylic CH_2 of oleate and linoleate	E
2.07	m	Allylic CH_2 at the n-3 terminus	E
2.29	t, 7.4	α-CH_2 of butyryl	F
2.31	2 t, 7.5	α-CH_2 of esterified chains (sn-1,3 and sn-2)	F
2.35–2.38	t, 7.5	α-CH_2 of free fatty acids and partial glycerides	F
2.770	t, 6.5	Bis-allylic CH_2 of linoleate	G_2
2.805	t, 6.0	Bis-allylic CH_2 of linolenate and other n-3 chains	G_1
3.73	m	CH_2OH of 1,2-diacylglycerols	–
4.14	dd, 11.9, 6.0	Glycerol sn-1,3 CH_2 (one proton of each pair)	H
4.29	dd, 11.9, 4.3	Glycerol sn-1,3 CH_2 (second proton of each pair)	H
4.93, 4.99	m (10.2; 17.1)	Terminal =CH_2 of caproleate (9-decenoate, C10:1)	–
5.08	m	sn-2 CH of 1,2-diacylglycerols; ^{13}C satellite of CH=CH	–
5.26	m	Glycerol sn-2 CH	I
5.34	m	Olefinic CH=CH	J
5.65	dt, 15, 7	H12 of rumenic acid (9c,11t-18:2)	–
5.80	m	–CH= of caproleate	–
5.94	t, 11	H10 of rumenic acid	–
6.28	dd, 15, 11	H11 of rumenic acid	–
7.26	s	Residual $CHCl_3$	–

4.2 Comparison of the lipid spectra across product types

The diagnostic windows of representative products are compared in Figure 5. In the authentic cheese the butyrate triplet is the tallest feature of region B. In a retail grated product that contains non-dairy fat it is reduced to about half its height relative to the glycerol signal, the bis-allylic triplet of linoleate becomes more than twice as intense and loses its linolenate partner, and the olefinic multiplet grows. In a lipolyzed retail product the triplet at δ 0.947 is also reduced, but the bis-allylic and olefinic regions are those of milk fat, a second set of lines appears on the low-field side of the α-methylene signal, and the n-3 and third methyl triplets stand out in region B; this is the spectral distinction between lipolysis and substitution on which the decision rule rests. The declared cheese product has lost the butyrate triplet almost completely. The prepared cheese loaf shows a strong linolenate triplet in G_1 and an intense n-3 methyl triplet, the signature of a canola- or soybean-type oil, and the plant-based analogue has no butyrate, almost no unsaturation and a methylene envelope without the C_2 shoulder, as expected of a lauric fat.

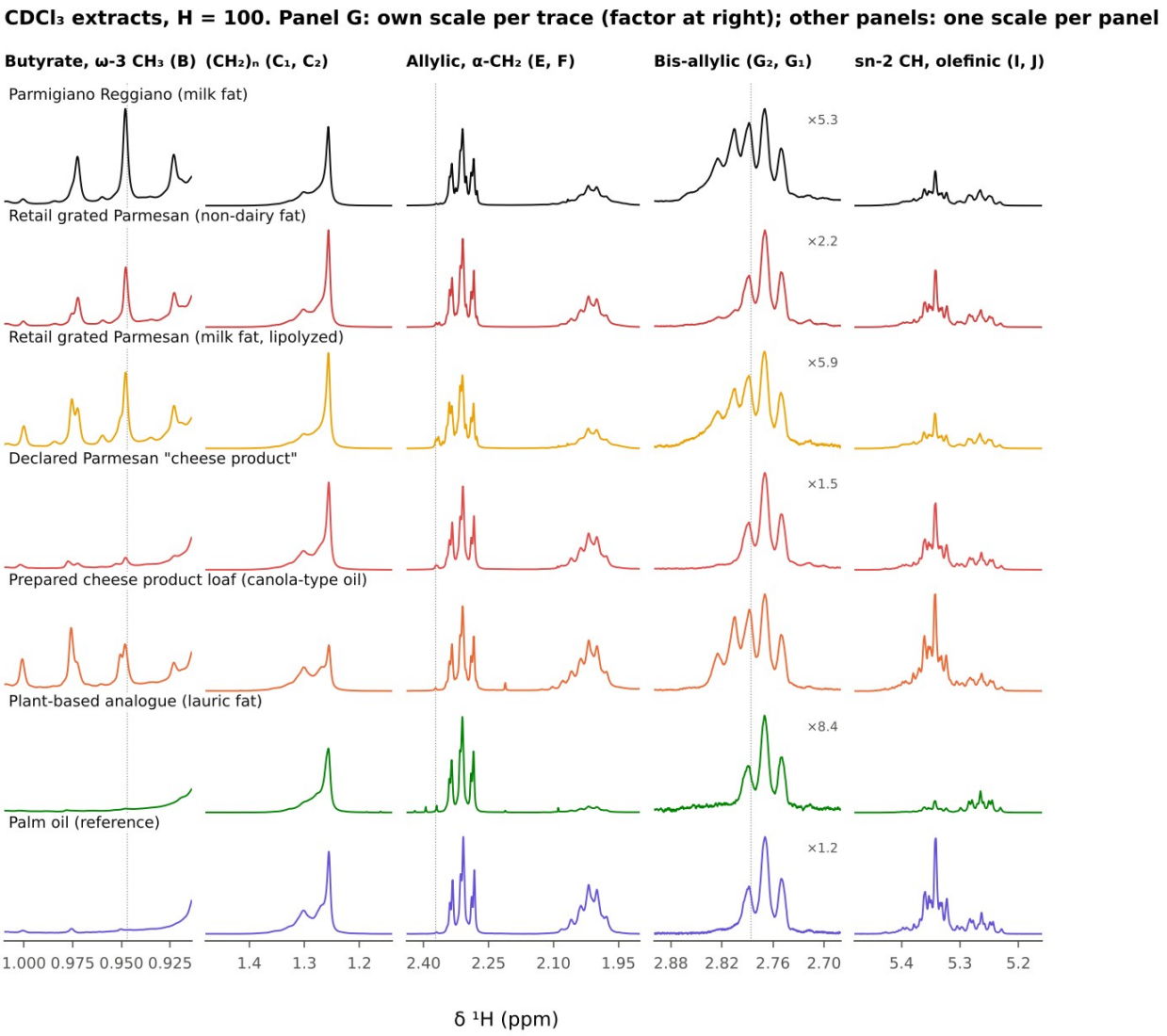

Figure 5. Diagnostic windows of the $CDCl_3$ spectra of representative products and of the palm oil reference, scaled to region H = 100 and offset. In the panel for region G each trace is drawn on its own full scale and the vertical expansion relative to the tallest trace of the panel is given at the right; in the other panels all traces share one scale set by the tallest signal. No signal is truncated; the rise at the right-hand edge of the region B panel is the flank of the terminal methyl triplet, which lies outside the window. Dotted lines mark the butyrate methyl triplet (δ 0.947), the free-acid α-methylene line (δ 2.3725) and the G_1/G_2 divide (δ 2.786).

4.3 Reference fats and oils

The fourteen reference fats and oils establish what the replacement of milk fat by another fat would look like (Figure 6 and Table 7). None of them shows a butyrate triplet: the apparent C4:0 values of 0.03 to 1.05 mol % returned by the fitting procedure are the residue of the methyl-triplet tail and of the n-3 triplet. Together with the three plant-based analogues they define the floor of the measurement, taken as their mean plus three standard deviations, 1.27 mol % (Section 3.6); a fitted butyrate content at or below that level cannot be distinguished from none. The ruminant tallows share with milk fat a low linoleate content, a high n-3 to linoleate ratio (G_1/G_2 of 1.3 to 2.1) and, in the case of lamb tallow, a clear conjugated-diene signal, but they have a mean chain length above 17 carbons against 14.2 for milk fat. Lard and the poultry fats contain 10 to 25 mol % of linoleate. Palm oil, the fat most often used in cheese analogues, combines about 10 mol % of

linoleate with a very low linolenate content, a mean chain length of 17.5 and a C_2/C_1 ratio of 0.69, against 0.47 for milk fat. The seed and fruit oils lie further still from milk fat in every descriptor.

One of the two oils labeled as avocado oil contained about 38 mol % of linoleate, far above the second avocado oil (10 mol %) and above the range usually reported for that oil. Its identity was not pursued, since it serves here only as an example of a polyunsaturated oil, but the observation illustrates that the reference materials of an authentication study need the same scrutiny as the samples.

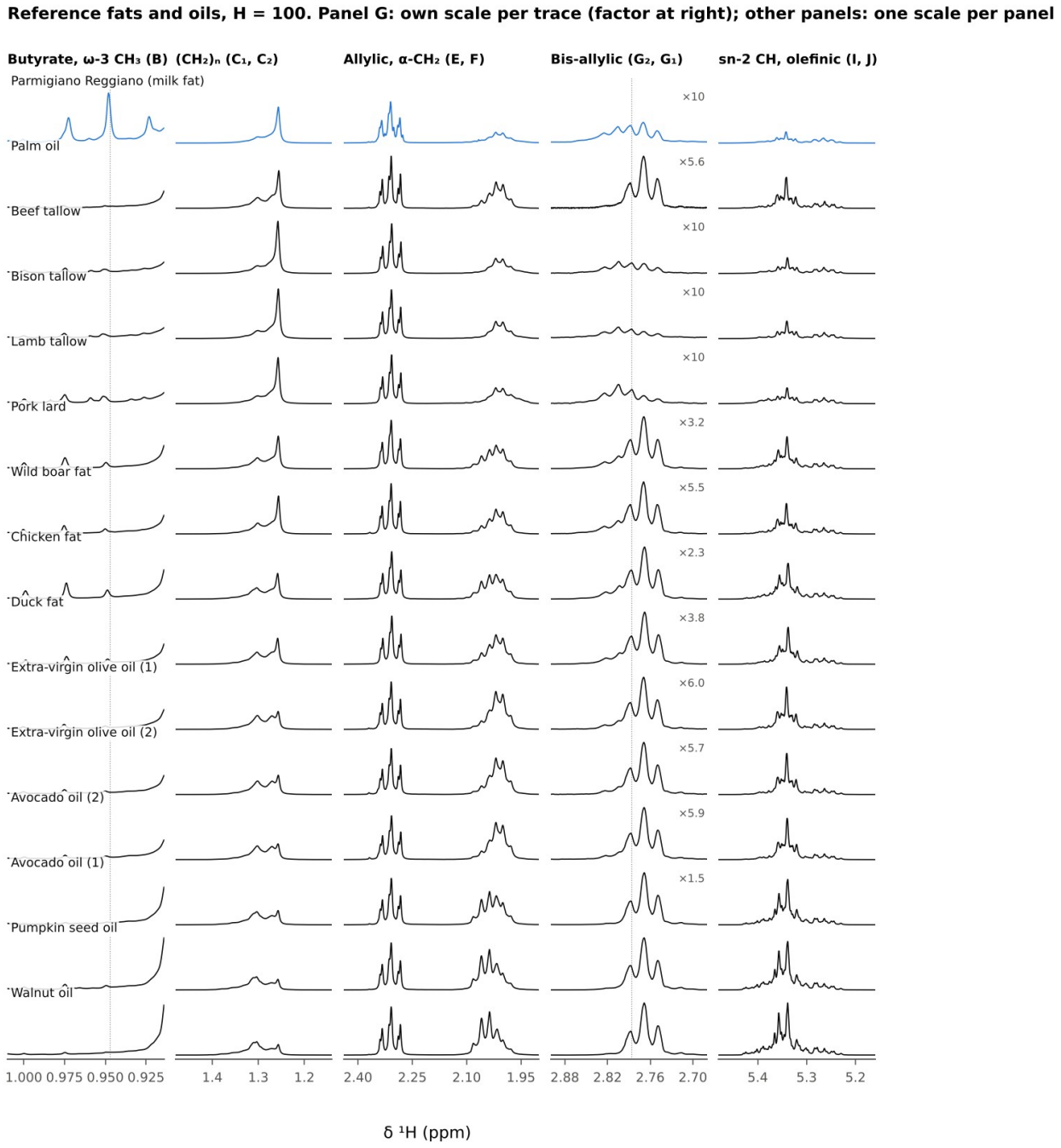

Figure 6. ¹H NMR spectra ($CDCl_3$, region H = 100) of the reference fats and oils compared with Parmigiano Reggiano (top trace, blue), in the same windows and with the same scaling conventions as Figure 5. No reference fat shows the butyrate triplet. The ruminant tallows show almost no bis-allylic signal (tenfold expansion), whereas lard, the poultry fats and the vegetable oils show the linoleate triplet at δ 2.77.

Table 7. Ratiometric descriptors of the reference fats and oils, for comparison with the milk-fat reference population (Table 10). C4:0, C18:2, C18:3, MUFA and SFA in mol % of acyl chains; sterol in mol per 100 mol of glycerol. The small apparent C4:0 values define the floor of the butyrate measurement.

Fat or oil	B/A	J/I	G/I	C4:0	C18:2	C18:3	MUFA	SFA	NC	C_2/C_1	G_1/G_2	Sterol
Palm oil	0.031	3.05	0.52	0.07	9.6	0.40	48.0	42.0	17.45	0.69	0.32	0.19
Beef tallow	0.038	1.90	0.15	0.32	1.1	0.71	35.3	62.9	17.26	0.44	1.34	0.42
Bison tallow	0.041	1.80	0.16	0.32	0.7	0.69	39.4	59.3	17.25	0.47	1.42	0.31
Lamb tallow	0.052	2.36	0.20	0.90	0.8	1.32	44.4	53.5	17.32	0.50	2.12	0.32
Pork lard	0.048	3.35	1.03	0.03	17.9	1.96	41.6	38.5	17.49	0.83	0.53	0.39
Wild boar fat	0.041	3.10	0.66	0.09	10.2	1.17	46.4	42.2	17.48	0.71	0.58	0.40
Chicken fat	0.056	3.57	1.28	0.07	25.3	2.90	41.8	30.0	17.46	1.10	0.49	0.38
Duck fat	0.038	3.37	0.81	0.10	15.2	1.47	54.6	28.8	17.56	1.01	0.44	0.25
Extra-virgin olive oil (1)	0.038	3.81	0.52	0.06	9.5	0.68	74.4	15.4	17.84	1.20	0.41	0.38
Extra-virgin olive oil (2)	0.042	3.82	0.56	0.04	10.1	0.61	71.9	17.4	17.85	1.14	0.37	0.67
Avocado oil (1)	0.043	4.01	1.48	1.05	37.5	0.79	47.9	13.8	17.81	1.66	0.25	0.57
Avocado oil (2)	0.045	3.91	0.55	0.06	10.0	0.59	75.1	14.3	17.86	1.28	0.37	0.68
Pumpkin seed oil	0.056	4.36	1.92	0.08	51.1	1.53	36.8	10.5	17.87	2.22	0.28	0.77
Walnut oil	0.048	4.34	1.97	0.07	55.8	1.12	32.8	10.2	17.91	2.29	0.24	0.66

4.4 The water-soluble fraction in D_2O

The aqueous extract of a ripened hard cheese gives one of the most crowded ¹H spectra encountered in food analysis (Figure 7). Lactate, with its methyl doublet at δ 1.33 and methine quartet at δ 4.11, is the largest single component, and the whole spectrum is drawn on one vertical scale in the uppermost panel of the figure to show its dominance. Everything else of consequence is seen only on expansion, and the eight insets, each scaled to the tallest signal within its own window, show the assignments on which the targeted analysis rests.

The methyl region between δ 0.9 and 1.1 contains the doublets of valine, leucine and isoleucine and the triplet of the isoleucine δ-methyl. These branched-chain amino acids are released in proportion to the general advance of proteolysis and, unlike most other amino-acid signals, are free of overlap with sugars, organic acids and lipid; the area of this region was therefore adopted as the free amino-acid index. Alanine gives a doublet at δ 1.48. The region from δ 1.6 to 2.2 holds the side-chain multiplets of lysine, leucine, arginine, proline and glutamate, with the acetate singlet at δ 1.92–1.94 and the S-methyl singlet of methionine at δ 2.135 standing above them. Between δ 2.2 and 3.0 lie the γ-methylene of glutamate at δ 2.35, the multiplets of pyroglutamate, the succinate singlet, whose position between δ 2.43 and 2.47 depends on pH, the methionine γ-methylene and the β-protons of aspartate and asparagine; citrate, when it is present, appears here as an AB quartet. The lysine ε-methylene triplet at δ 3.02, the arginine δ-methylene, the β-protons of tyrosine and phenylalanine, the singlet of glycine at δ 3.56 and the trimethylammonium singlet of choline compounds at δ 3.20–3.23 occupy the next window, followed by the α-methine region, in which the sugars also resonate when they are present. The aromatic region shows the two doublets of tyrosine, the multiplets of phenylalanine and the doublets of tryptophan, and at lowest field the H2 singlet of histidine and the singlet of formate[48,49,51,52,54–56,62,63,154,155].

The assignments were checked against spectra of the individual amino acids acquired in D_2O on the same instrument (Figure 8). The standards were run without adjustment of pH, so the shifts of protons near ionizable groups differ from those in the cheese extracts, whose apparent pH lay between 5.3 and 5.9; aspartate, glutamate and histidine are the most affected. The patterns and the pH-insensitive shifts

nonetheless agree with the extract in every case, and the comparison shows which resonances of the extract are free of interference and suitable for integration.

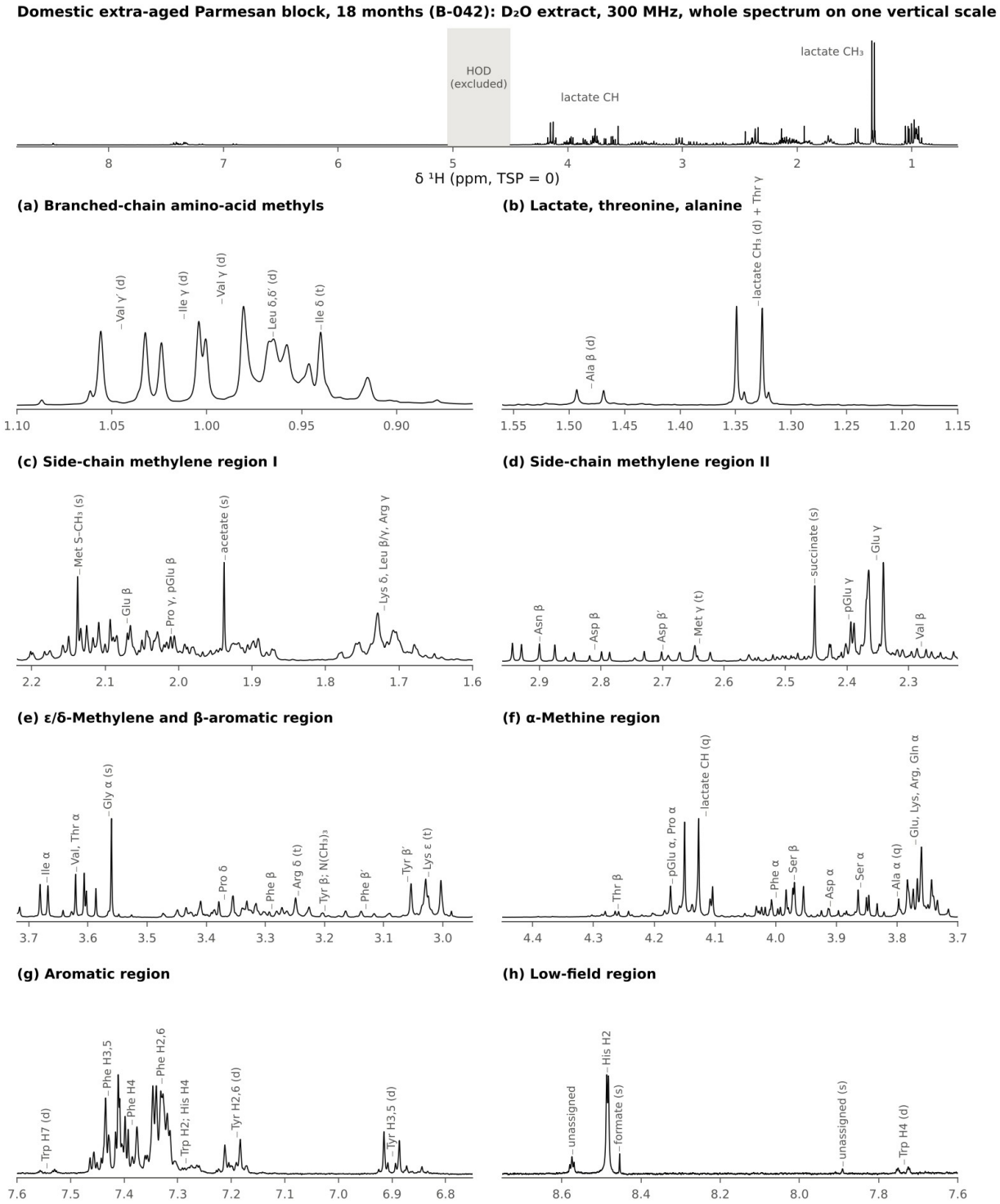

Figure 7. ¹H NMR spectrum (300 MHz, D_2O, TSP-d_4 = 0) of the aqueous extract of a domestic extra-aged Parmesan block of 18 months (B-042). Top: whole spectrum on one vertical scale, with the residual water region excluded. Insets (a) to (h): successive windows, each drawn on its own full vertical scale so that no signal is truncated, with assignments.

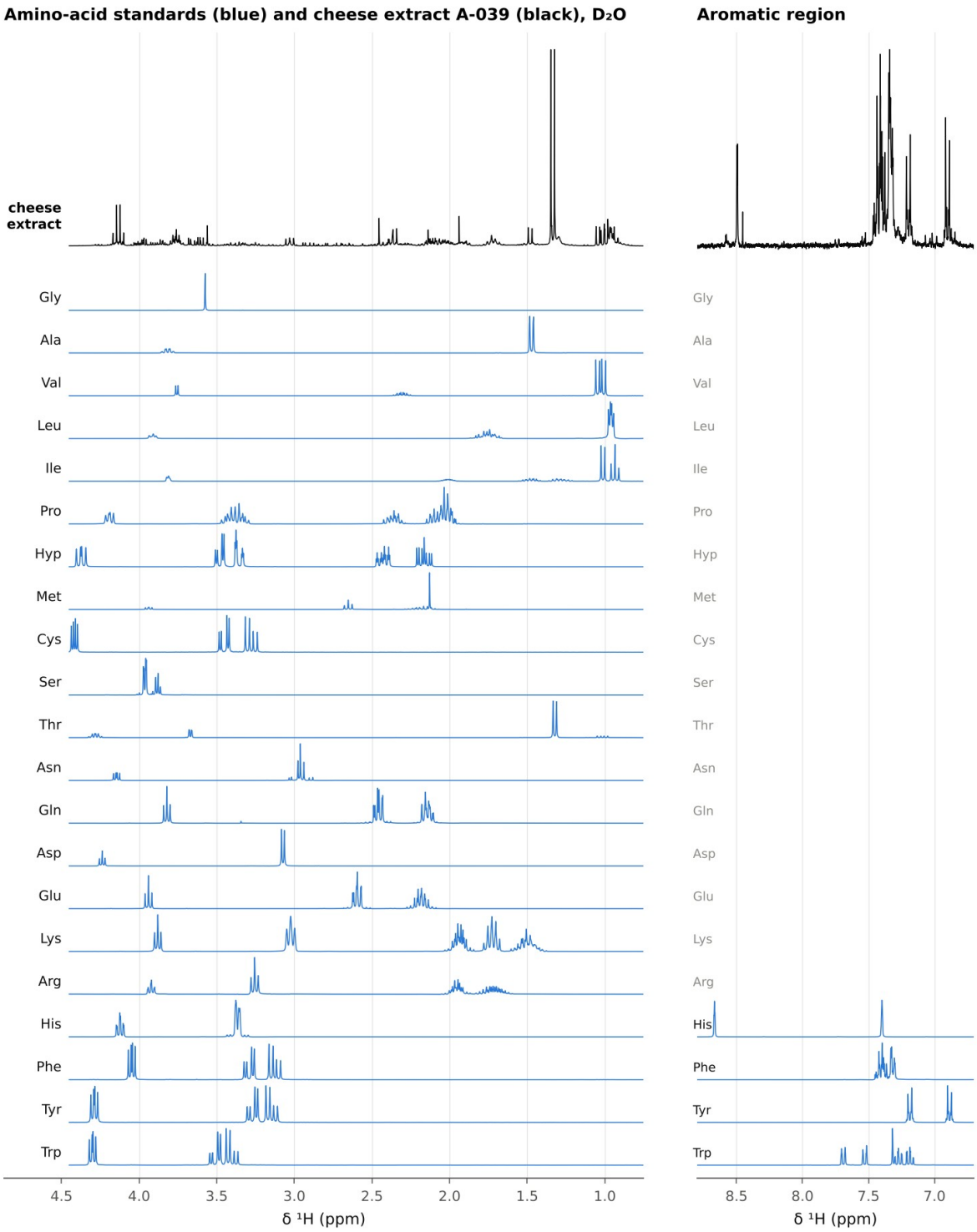

Figure 8. ¹H NMR spectra (D_2O) of the amino-acid standards (blue), each scaled to its own tallest signal, compared with the aqueous extract of a retail grated Parmesan that contains only milk fat (A-039, black, on one full vertical scale). Left: aliphatic region. Right: aromatic region, shown for the four aromatic amino acids.

Table 8. Assignments and targeted integration windows for the D_2O extracts (300 MHz, TSP-d_4 = 0, apparent pH 5.3–5.9). nH is the number of protons by which the area was divided. Shifts of citrate, succinate, acetate, histidine and the EDTA species depend on pH and, for citrate and EDTA, on the binding of calcium and magnesium. "Sharp" denotes integration of the sharp component after removal of the broad background (Section 3.7).

Compound	δ (ppm), multiplicity, J (Hz)	Protons	Quantified from	nH
Isoleucine	0.94 t (7.4); 1.01 d (7.0); 1.98 m; 3.67 d	δ-CH_3; γ-CH_3; β-CH; α-CH	Free amino-acid index: sharp area δ 0.925–1.070	18
Leucine	0.96 d, 0.97 d (6.2); 1.72 m; 3.73 m	δ,δ′-CH_3; β-CH_2, γ-CH; α-CH	within the free amino-acid index	
Valine	0.99 d, 1.04 d (7.0); 2.28 m; 3.61 d	γ,γ′-CH_3; β-CH; α-CH	within the free amino-acid index	
Butyrate (free)	0.89 t (7.4); 1.56 sextet; 2.155 t (7.4)	CH_3; β-CH_2; α-CH_2	sharp area δ 1.515–1.595; triplet fit at δ 2.155	2
Propionate	1.05 t (7.7); 2.18 q (7.7)	CH_3; CH_2	joint quartet/triplet fit, δ 2.105–2.255	2
Ethanol	1.18 t (7.1); 3.65 q	CH_3; CH_2	tracked triplet	3
Lactate	1.33 d (6.9); 4.11 q (6.9)	CH_3; CH	doublet fit with the lipid resonance, δ 1.13–1.47	3
Threonine	1.33 d (6.6); 4.26 m; 3.59 d	γ-CH_3; β-CH; α-CH	not quantified (under lactate)	
Dispersed lipid	0.88 br; 1.28 br; 2.0–2.3 br; 5.3 br	CH_3; $(CH_2)_n$; allylic and α-CH_2; CH=CH	broad pseudo-Voigt line at δ 1.26–1.33	area
Alanine	1.48 d (7.2); 3.78 q	β-CH_3; α-CH	tracked doublet	3
Lysine	1.47 m; 1.72 m; 1.90 m; 3.02 t (7.6)	γ; δ; β; ε-CH_2	tracked triplet at δ 3.02	2
Acetate	1.92–1.94 s	CH_3	tracked singlet	3
Proline	2.01 m; 2.07, 2.35 m; 3.37 m; 4.13 dd	γ; β; δ-CH_2; α-CH	not quantified	
Glutamate	2.06, 2.12 m; 2.35 m; 3.76 dd	β-CH_2; γ-CH_2; α-CH	sharp area δ 2.310–2.385	2
Methionine	2.135 s; 2.64 t (7.5)	S-CH_3; γ-CH_2	tracked singlet	3
Pyroglutamate	2.03, 2.52 m; 2.40 m; 4.18 dd	β-CH_2; γ-CH_2; α-CH	confirmed by gHSQC; not resolved from glutamate in one dimension	
Succinate	2.43–2.47 s	2 CH_2	tracked singlet, δ 2.425–2.478	4
Citrate	2.52–2.72, AB quartet (15.6)	2 CH_2	AB-quartet fit, center δ 2.56–2.70	4
Aspartate	2.70 dd, 2.81 dd; 3.91 dd	β-CH_2; α-CH	not quantified	
Asparagine	2.87 dd, 2.95 dd; 4.00 dd	β-CH_2; α-CH	not quantified	
Tyrosine	3.05 dd, 3.19 dd; 6.90 d (8.5); 7.19 d (8.5)	β-CH_2; H3,5; H2,6	tracked doublet at δ 6.90	2
Phenylalanine	3.13 dd, 3.29 dd; 4.00 dd; 7.33–7.43 m	β-CH_2; α-CH; aromatic	sharp area δ 7.300–7.460	5
Choline compounds	3.20–3.23 s	$N(CH_3)_3$	tracked singlet	9
Arginine	1.70 m; 1.91 m; 3.25 t (6.9)	γ; β; δ-CH_2	not quantified	
Glycine	3.56 s	α-CH_2	tracked singlet	2
Glycerol	3.56 dd, 3.65 dd; 3.78 m	CH_2; CH	confirmed by gHSQC	
Lactose	4.455 d (7.8); 4.67 d; 5.23 d (3.8)	Gal H1; β-Glc H1; α-Glc H1	tracked doublet at δ 4.455	1
Glucose and other α-pyranoses	5.23 d (3.8)	α-H1	tracked doublet	1
Galactose	5.27 d (3.8); 4.58 d (7.9)	α-H1; β-H1	tracked doublet at δ 5.27	1
α-(1→4) Glucans (starch, maltodextrin)	5.36–5.46 br	H1 of chain residues	area δ 5.36–5.46	area
Sorbate	1.83 d; 5.83 d (15); 6.12–6.33 m; 7.0–7.1 dd	CH_3; H2; H4, H5; H3	tracked doublet at δ 5.83; windows excluded from the binned data	1
Histidine	7.1–7.4 s; 8.25–8.66 s	H4; H2	sharp area ± 0.010 ppm about the H2 maximum; shift gives apparent pH	1
Tryptophan	7.29 s; 7.54 d; 7.73 d	H2; H7; H4	not quantified	
Formate	8.455 s	CH	tracked singlet, δ 8.4555 ± 0.005	1
Unassigned	7.84–7.91 s (pH-dependent); 8.57 s	–	noted; strongest in long-ripened Dutch-type cheeses	
Hippurate	7.83 d; 7.63 t; 7.55 t; 3.96 d	H2,6; H4; H3,5; CH_2	pattern search, not detected (Section 4.7)	
Ca–EDTA	2.565 s; 3.13 AB q	N–CH_2–CH_2–N; acetate CH_2	singlet fit (Section 3.8)	4
Mg–EDTA	2.70 s	N–CH_2–CH_2–N	singlet fit	4

Compound	δ (ppm), multiplicity, J (Hz)	Protons	Quantified from	nH
EDTA, free	3.2–3.5 s; 3.55–3.86 s	N–CH_2–CH_2–N; acetate CH_2	singlets absent from the native extract	4; 8
TSP-d_4	0.00 s	$Si(CH_3)_3$	shift reference; not used for quantitation	9

4.5 Two-dimensional ¹H–¹³C correlation of the aqueous extract

A multiplicity-edited ¹H–¹³C gHSQC spectrum had been acquired for the aqueous extract of an Italian Pecorino Romano (A-004). It was available as an exported contour plot and not as raw data, and it was evaluated by extracting the contour paths from the vector drawing, grouping them into cross-peaks and taking the centroid of the innermost contour of each (Figure 9). The procedure recovers peak positions and the sign of each cross-peak, which distinguishes methylene groups from methine and methyl groups, but not volumes, so the spectrum was used for assignment only.

Fifty-seven cross-peaks were located between δH 0.9 and 4.4 (Table 9). Their proton shifts agree with the one-dimensional spectrum of the same extract, and their carbon shifts agree with DSS-referenced literature values for Cheddar extracts and for standards in the Biological Magnetic Resonance Data Bank after a uniform offset of 2.4 ppm is added, the mean difference for 46 confidently assigned correlations being 2.39 ppm with a standard deviation of 0.18 ppm[62,156,157]. The offset shows that the carbon axis of the exported spectrum had not been referenced; both scales are given in the figure and the table.

The second dimension resolves several overlaps on which the one-dimensional interpretation depends. The lysine ε-methylene at δH 3.04 and the tyrosine β-methylene at δH 3.05, which coincide in the proton spectrum, are separated by 3.5 ppm in carbon. The glutamate γ-methylene (δC 35.8 on the DSS scale) and the pyroglutamate γ-methylene (δC 32.2) are resolved although their proton multiplets overlap between δ 2.35 and 2.42, which confirms the presence of pyroglutamate and explains why its proton signal cannot be integrated independently at 300 MHz. Succinate is identified at δH 2.47, displaced from its neutral-pH position by partial protonation, and glycerol, a product of lipolysis, at δH 3.64 and δC 65.1. Citrulline and ornithine, formed from arginine by the arginine deiminase pathway of the non-starter flora, contribute to the methylene correlations near δH 3.15 and 1.9, and a trimethylammonium correlation at δH 3.23 is assigned to choline compounds. Tentative assignments are marked as such in the table: a methylene pair at δC 66.1 that fits the β-carbon of O-phosphoserine, free or in casein phosphopeptides; a methylene at δH 1.90 that fits γ-aminobutyrate; and a methyl at δH 2.41 that fits pyruvate. One weak methylene correlation at δH 3.87 and δC 61.7 (as acquired) was found by the contour extraction in addition to the 57 listed and remains unassigned.

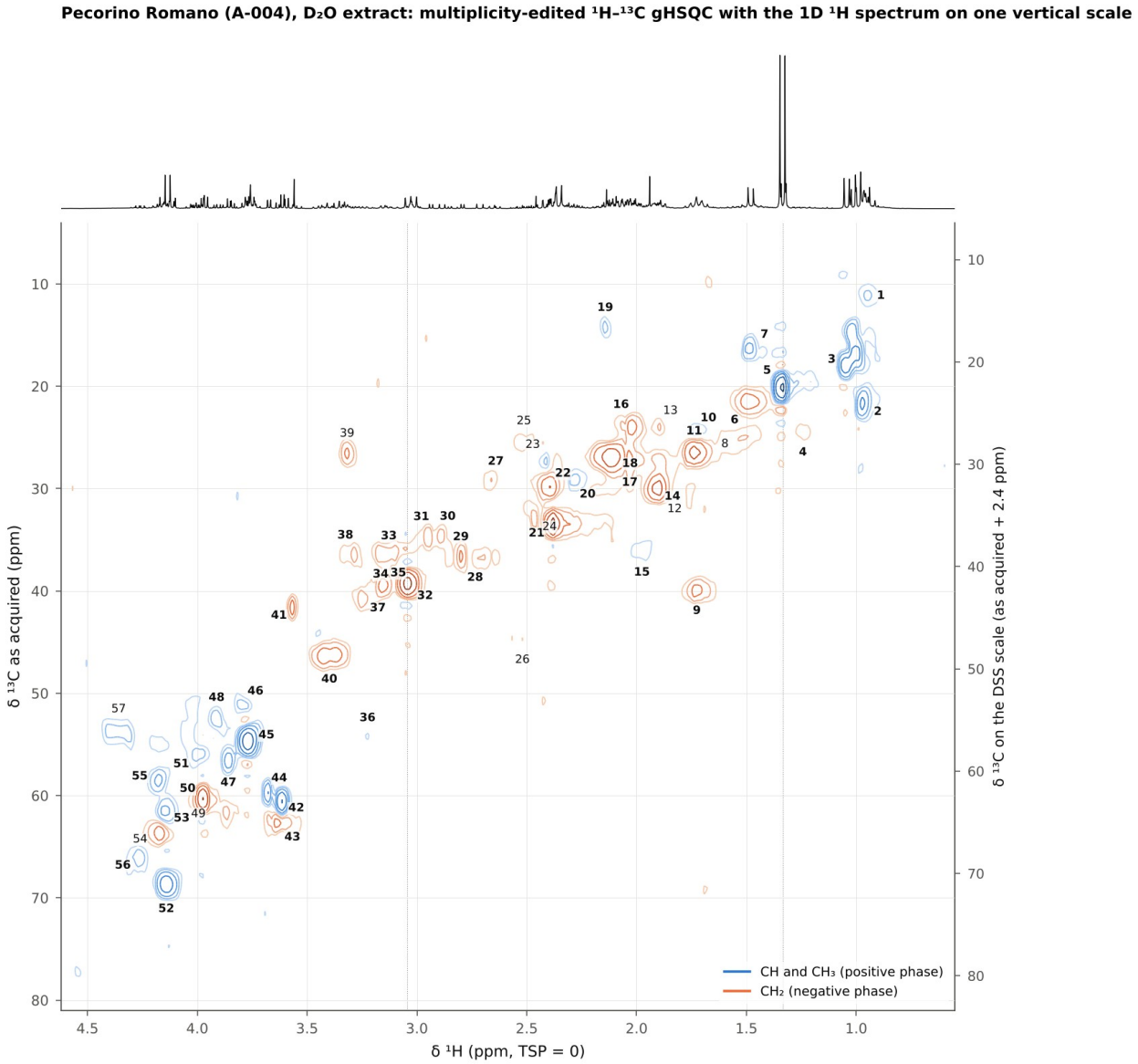

Figure 9. Multiplicity-edited ¹H–¹³C gHSQC spectrum (300 MHz, D_2O) of the aqueous extract of an Italian Pecorino Romano (A-004), redrawn from the exported contour plot, with the one-dimensional ¹H spectrum of the same extract above on one full vertical scale. Blue contours: CH and CH_3 (positive phase); orange to red contours: CH_2 (negative phase). Numbers refer to Table 9; bold numbers mark confident assignments. Left axis: δC as acquired; right axis: DSS scale (as acquired + 2.4 ppm). Dotted lines mark the t_1 ridges of the lactate methyl and the lysine ε-methylene.

Table 9. Cross-peaks of the multiplicity-edited ¹H–¹³C gHSQC spectrum of the D_2O extract of an Italian Pecorino Romano (A-004; Figure 9). δC is given as acquired and on the DSS scale (as acquired + 2.4 ppm). Phase: + CH or CH_3, − CH_2. Reference shifts (δH; δC) are from Cheddar aqueous extracts (Chen)[62] or from standards of the Biological Magnetic Resonance Data Bank (BMRB)[156]; "partial" marks a correlation shared with another compound or only partly resolved, and "tent." a tentative assignment.

No.	δH	δC acq.	δC DSS	Phase	Assignment	Reference δH; δC	Source
1	0.95	11.1	13.5	+ (CH_3)	Isoleucine δ-CH_3	0.93; 13.6	Chen
2	0.97	21.6	24.1	+ (CH_3)	Leucine δ,δ′-CH_3	0.96; 24.6	Chen
3	1.03	17.4	19.8	+ (CH_3)	Valine γ-CH_3 (+ isoleucine γ-CH_3)	1.03; 20.5	Chen
4	1.24	24.4	26.8	− (CH_2)	Isoleucine γ-CH_2 (H γ′)	1.46; 26.9	Chen
5	1.34	20.1	22.5	+ (CH_3)	Lactate CH_3 (+ threonine γ-CH_3)	1.32; 22.6	Chen

No.	δH	δC acq.	δC DSS	Phase	Assignment	Reference δH; δC	Source
6	1.49	21.5	23.9	− (CH_2)	Lysine γ-CH_2	1.44; 24.2	Chen
7	1.49	16.3	18.7	+ (CH_3)	Alanine β-CH_3	1.47; 18.6	Chen
8	1.52	25.0	27.4	− (CH_2)	Isoleucine γ-CH_2 (H γ); citrulline γ-CH_2 overlaps (partial)	1.46; 26.9	Chen
9	1.73	39.9	42.4	− (CH_2)	Leucine β-CH_2	1.73; 42.4	Chen
10	1.73	24.3	26.7	+ (CH)	Leucine γ-CH	1.70; 26.7	Chen
11	1.74	26.5	28.9	− (CH_2)	Lysine δ-CH_2	1.71; 28.8	Chen
12	1.76	30.6	33.0	− (CH_2)	Unassigned CH_2 (unknown B of Chen et al.)	1.76; 32.9	Chen
13	1.90	24.0	26.4	− (CH_2)	CH_2 at δH 1.90: γ-aminobutyrate C3 (tent.)	–	
14	1.91	29.9	32.3	− (CH_2)	Lysine β-CH_2 (+ arginine, ornithine, citrulline β-CH_2)	1.90; 32.3	Chen
15	1.98	36.0	38.4	+ (CH)	Isoleucine β-CH	1.97; 38.4	Chen
16	2.02	24.0	26.4	− (CH_2)	Proline γ-CH_2 (+ pyroglutamate β-CH_2)	2.00; 26.5	Chen
17	2.03	27.0	29.4	− (CH_2)	Glutamate β′-CH_2	2.06; 29.4	Chen
18	2.12	26.9	29.3	− (CH_2)	Glutamate β-CH_2 (+ glutamine β-CH_2)	2.06; 29.4	Chen
19	2.14	14.2	16.6	+ (CH_3)	Methionine S-CH_3	2.12; 16.5	Chen
20	2.28	29.1	31.5	+ (CH)	Valine β-CH	2.27; 31.7	Chen
21	2.38	33.4	35.8	− (CH_2)	Glutamate γ-CH_2	2.35; 35.9	Chen
22	2.40	29.8	32.2	− (CH_2)	Pyroglutamate γ-CH_2	2.39; 32.2	Chen
23	2.41	27.3	29.7	+ (CH_3)	Pyruvate CH_3 (tent.)	2.35; 29.5	Chen
24	2.47	32.9	35.3	− (CH_2)	Succinate CH_2 (partly protonated at the pH of the extract) (partial)	2.39; 36.9	BMRB
25	2.52	25.4	27.8	− (CH_2)	Pyroglutamate β′-CH_2 (partial)	2.56; 27.9	Chen
26	2.52	44.7	47.1	− (CH_2)	Citrate CH_2 (trace) (partial)	2.58; 48.4	BMRB
27	2.66	29.1	31.5	− (CH_2)	Methionine γ-CH_2	2.63; 31.4	Chen
28	2.70	36.7	39.2	− (CH_2)	Aspartate β′-CH_2	2.70; 39.2	BMRB
29	2.80	36.6	39.0	− (CH_2)	Aspartate β-CH_2	2.79; 39.2	BMRB
30	2.89	34.6	37.0	− (CH_2)	Asparagine β-CH_2	2.87; 37.0	Chen
31	2.95	34.7	37.1	− (CH_2)	Asparagine β′-CH_2	2.94; 37.0	Chen
32	3.04	39.3	41.7	− (CH_2)	Lysine ε-CH_2 (+ ornithine δ-CH_2)	3.01; 41.6	Chen
33	3.05	35.8	38.3	− (CH_2)	Tyrosine β-CH_2	3.05; 38.0	Chen
34	3.14	36.3	38.8	− (CH_2)	Phenylalanine β-CH_2 (+ tyrosine β′)	3.12; 38.8	Chen
35	3.15	39.4	41.9	− (CH_2)	Citrulline δ-CH_2	3.13; 41.6	Chen
36	3.23	54.2	56.6	+ (CH_3)	$N(CH_3)_3$ of glycerophosphocholine/phosphocholine/choline	3.19; 56.6	BMRB
37	3.25	40.8	43.2	− (CH_2)	Arginine δ-CH_2	3.24; 43.1	Chen
38	3.29	36.4	38.9	− (CH_2)	Phenylalanine β′-CH_2	3.28; 38.8	Chen
39	3.32	26.5	28.9	− (CH_2)	Tryptophan β-CH_2 (tent.)	3.29; 29.2	BMRB
40	3.40	46.3	48.7	− (CH_2)	Proline δ-CH_2	3.41; 48.6	Chen
41	3.57	41.6	44.0	− (CH_2)	Glycine α-CH_2	3.55; 44.0	Chen
42	3.62	60.6	63.0	+ (CH)	Valine α-CH (+ threonine α-CH)	3.60; 62.9	Chen
43	3.64	62.7	65.1	− (CH_2)	Glycerol C1/C3 CH_2 (lipolysis)	3.64; 65.0	Chen
44	3.68	59.7	62.1	+ (CH)	Isoleucine α-CH	3.66; 62.2	Chen
45	3.77	54.7	57.1	+ (CH)	α-CH of Glu, Gln, Lys, Arg, Orn, Cit (+ Leu α)	3.75; 57.0	Chen
46	3.80	51.1	53.5	+ (CH)	Alanine α-CH	3.77; 53.1	Chen
47	3.86	56.6	59.0	+ (CH)	Serine α-CH (+ methionine α-CH)	3.84; 58.9	Chen
48	3.91	52.5	54.9	+ (CH)	Aspartate α-CH	3.89; 54.9	BMRB
49	3.97	63.7	66.1	− (CH_2)	O-Phosphoseryl β-CH_2 (H β), free or in casein phosphopeptides (tent.)	4.06; 65.3	BMRB
50	3.97	60.3	62.7	− (CH_2)	Serine β-CH_2	3.95; 62.6	Chen
51	4.00	56.0	58.4	+ (CH)	Phenylalanine α-CH (+ tyrosine, asparagine α-CH)	3.99; 58.5	Chen
52	4.14	68.7	71.1	+ (CH)	Lactate α-CH	4.12; 71.1	Chen
53	4.15	61.5	63.9	+ (CH)	Proline α-CH	4.12; 63.8	Chen
54	4.17	63.7	66.1	− (CH_2)	O-Phosphoseryl β-CH_2 (H β′) (tent.)	4.17; 65.3	BMRB
55	4.18	58.5	60.9	+ (CH)	Pyroglutamate α-CH	4.17; 60.8	Chen
56	4.27	66.1	68.5	+ (CH)	Threonine β-CH	4.25; 68.4	Chen

No.	δH	δC acq.	δC DSS	Phase	Assignment	Reference δH; δC	Source
57	4.36	53.8	56.2	+ (CH)	α-CH of residual peptides (broad) (partial)	–	

4.6 Comparison of the aqueous spectra across product types

The aqueous spectra of the product types differ in ways that are visible without any calculation (Figure 10). The ripened reference and the retail grated product with milk fat show the dense amino-acid multiplets in the methyl and aromatic windows, no citrate and no anomeric signals. The retail grated product with non-dairy fat shows weak amino-acid signals, a lactate doublet standing on a broad resonance of dispersed lipid, the four lines of citrate, anomeric doublets of free sugars and a broad olefinic resonance. This profile resembles those of the processed American single and of the declared cheese product, shown beneath it, far more closely than it resembles the profile of a ripened hard cheese. Fresh mozzarella has almost no amino-acid signals, citrate and a large galactose doublet, but no dispersed lipid; Emmental shows the propionate triplet and quartet and a very large acetate singlet; and the plant-based analogue shows only added organic acids and the envelope of a starch-derived glucan.

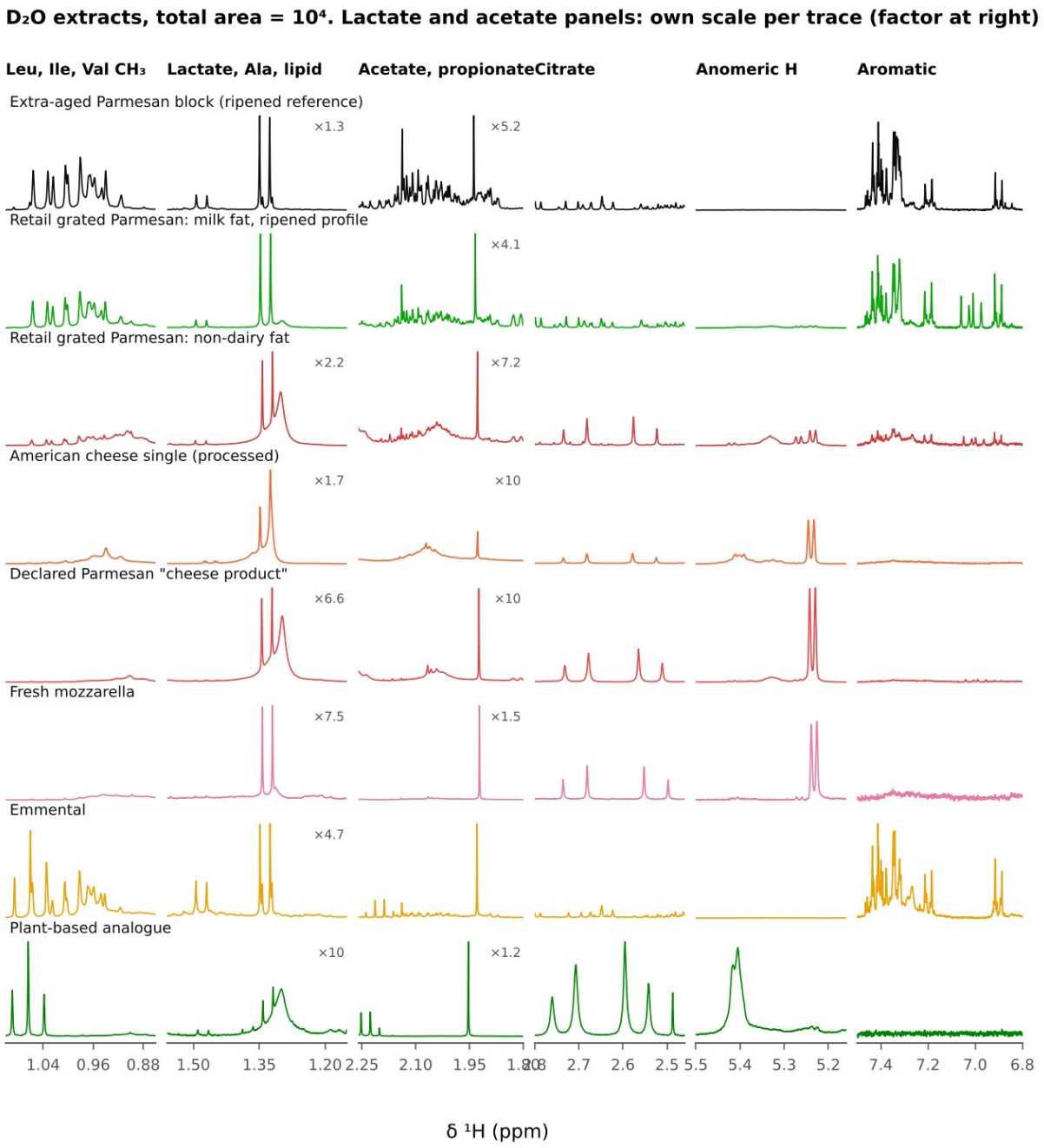

Figure 10. Diagnostic windows of the D_2O spectra of representative products, scaled to total spectral area and offset. In the lactate and acetate panels each trace is drawn on its own full scale, with the vertical expansion relative to the tallest trace given at the right; in the other panels all traces share one scale set by the tallest signal. No signal is truncated.

4.7 Quantitative ¹H NMR: from relative indices to concentrations

All aqueous results in this report are relative indices, proportional to the molar abundance of a compound within one extract. They become concentrations only when a signal of known molar amount is present under the same acquisition conditions. Several routes are available[103–105,158–160]. An internal standard may be added to the extract; for aqueous cheese extracts this should not be TSP, which binds to protein and peptides and loses intensity in proportion to their concentration. In the present data the TSP-d_4 singlet represented from 0.3% of the total spectral area in the processed cheeses to 7% in a plant-based analogue, a range of more than

twentyfold that reflects binding and the size of the extracted pool rather than any property of interest, and the signal was accordingly used for referencing only. DSS shows the same behavior to a lesser degree, and maleic acid, fumaric acid, formate added in known amount, or 4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate, which does not bind, are better choices[161,162]. Alternatively the standard may be kept outside the sample, as a separately measured external reference tied to the sample through the reciprocity of pulse length and receiver sensitivity (PULCON) or as an electronically synthesized reference signal (ERETIC)[163,164].

Whichever route is chosen, the conditions of quantitative acquisition must be met: a relaxation delay of at least five times the longest longitudinal relaxation time of the signals compared, or a calibrated correction for partial saturation; a uniform excitation profile; and, if water presaturation is used, verification that signals near the water resonance and exchangeable protons are not attenuated. With these in place the present indices convert directly, since each is already an area per proton. The present spectra were acquired without presaturation and with a recycle time of 8.27 s (Section 3.2), which meets the condition for the amino-acid, lactate and sugar signals if their relaxation times are below about 1.7 s but is unlikely to meet it for the slowly relaxing singlets of formate, acetate and TSP-d_4; the relaxation times were not measured. The external maleic acid standard described in Section 3.8, which was acquired with a delay of 25 s, offers a conversion in principle, and it has been used for the provisional calcium scale, but it cannot be extended to the organic components until the constancy of the TSP-d_4 response has been established, which the data above call into question.

Hippurate deserves particular mention because it has been proposed as a marker of the feeding regimen and of organic production in milk[45,165–167]. In milk it is readily seen through its aromatic doublet at δ 7.83 and triplets at δ 7.63 and 7.55. In cheese it is not expected to survive: lactic acid bacteria hydrolyze hippurate to benzoate and glycine during fermentation, and this is the origin of the natural benzoic acid of cultured dairy products and cheese[168,169]. The 84 aqueous spectra were searched for the hippurate pattern, a doublet near δ 7.83 with a triplet 0.20 ppm to higher field, and it was not found in any of them at more than four times the noise level. A sharp, pH-dependent singlet between δ 7.84 and 7.91 was found in 15 products and was strongest in the long-ripened Gouda and Mimolette cheeses; it is not hippurate, since it lacks the coupling pattern and the companion triplets, and it remains unassigned. Because it appears to increase with ripening in the Dutch-type cheeses it would repay identification by spiking or by two-dimensional NMR. The absence of hippurate is itself informative in one respect: a grated product that contained a substantial proportion of unfermented milk solids might be expected to show it, and none did, although the sensitivity of the present spectra for that purpose was not established.

5. Lipid Ratiometrics

5.1 The milk-fat reference population

The 24 products of the milk-fat reference population define a narrow envelope (Table 10). Butyrate from the fitted triplet amounted to 9.08 ± 0.47 mol % of acyl chains, a relative standard deviation of 5.2% across Grana-type, Dutch-type, alpine, young and processed cheeses from Italy, Switzerland, France, the Netherlands, Ireland and the United States, and the mean chain length was 14.22 ± 0.18 carbons. The methyl ratio B/A was 0.153 ± 0.006 and the split-methylene ratio C_2/C_1 0.471 ± 0.018, each with a relative standard deviation below 4%. The descriptors that depend on the feeding of the animal were much more variable: linoleate 1.77 ± 0.60 mol %, n-3 acids 0.69 ± 0.16 mol %, G_1/G_2 1.19 ± 0.27 and the conjugated diene 0.40 ± 0.23 mol %. The composition as a whole, about 70 mol % saturated, 28 mol % monounsaturated and 2.5 mol % polyunsaturated acids, agrees with the established composition of bovine milk fat[28,29].

Two properties of this table govern everything that follows. The descriptors with the tightest natural distribution are those that are fixed by the biosynthesis of milk fat in the mammary gland, and it is these that give the greatest power to detect a foreign fat. The descriptors that respond to pasture, season and species are the ones that carry information about the origin and type of an authentic cheese, and for the same reason they make weak criteria of adulteration.

Table 10. Lipid descriptors of the milk-fat reference population (24 cow's-milk products without lipolysis). Regions as in Table 4; fitted quantities as described in Section 3.5.

Descriptor	Mean	s.d.	RSD (%)	Minimum	Maximum
B/A	0.153	0.006	3.8	0.144	0.165
J/I	1.516	0.086	5.7	1.368	1.827
G/I	0.244	0.053	21.5	0.174	0.363
B/I	1.098	0.034	3.1	1.048	1.167
D1 = (C+D+E+F+G)/F	11.89	0.49	4.1	11.18	13.54
D2 = (C+D+E+F+G)/(I+J)	26.43	1.39	5.3	22.51	29.79
D3 = B/(A+B)	0.132	0.004	3.3	0.126	0.142
D1w (D replaced by F)	11.65	0.17	1.4	11.19	11.90
D2w (D replaced by F)	25.90	0.96	3.7	22.42	27.42
Tociu SCSFA (mol %)	13.10	0.47	3.6	12.28	14.07
Tociu PUFA (mol %)	4.37	0.94	21.5	3.09	6.76
Tociu MUFA (mol %)	25.73	2.13	8.3	22.95	29.90
Tociu LCSFA (mol %)	56.80	2.60	4.6	51.49	60.45
Hanganu–Chira butyrate, eq. 13 (mol %)	13.25	0.43	3.3	12.59	14.16
Hanganu–Chira butyrate, whole G_1-derived linolenate subtracted (mol %)	12.07	0.42	3.5	11.20	12.86
Hanganu–Chira butyrate, captured part of the n-3 triplet subtracted (mol %)	12.40	0.41	3.3	11.70	13.14
C4:0, fitted triplet (mol % of chains)	9.08	0.47	5.2	8.13	10.11
C10:1, vinyl (mol % of chains)	0.497	0.076	15.2	0.361	0.650
Conjugated diene, H10 (mol % of chains)	0.402	0.227	56.5	0.208	1.186
Sterol 18-CH_3 (mol per 100 mol glycerol)	0.725	0.047	6.5	0.658	0.865
C18:2, fitted G_2 triplet (mol %)	1.769	0.602	34.0	0.853	3.712
n-3 PUFA, fitted G_1 triplet (mol %)	0.693	0.162	23.4	0.502	1.091
MUFA (mol %)	27.64	2.37	8.6	24.23	32.04
SFA other than C4:0 (mol %)	60.82	2.48	4.1	55.46	64.44
NC, mean carbons per chain	14.22	0.18	1.3	13.71	14.53
C_2/C_1	0.471	0.018	3.9	0.446	0.517
G_1/G_2	1.192	0.272	22.8	0.887	1.897
Free acids + partial glycerides (% of chains)	0.534	0.385	72.2	0.189	1.690

5.2 The region-ratio screen

The simplest evaluation uses four ratios of regions and no fitting (Figure 11). All 22 retail grated products that were later assigned to the non-dairy fat category fell more than three standard deviations below the reference mean in B/A (0.083–0.100 against 0.153) and in B/I (0.65–0.76 against 1.10), and all 22 exceeded the reference mean of J/I by more than three standard deviations (1.99–2.65 against 1.52). The bis-allylic ratio G/I was raised as well, from 0.24 to a median of 0.42, but only 15 of the 22 exceeded the three-sigma limit, because the natural range of G/I in milk fat is wide. The declared vegetable-fat products and the plant-based analogues lay further out along the same directions.

The screen has two weaknesses that the more detailed analysis removes. First, B/A fell below the limit in four of the six lipolyzed milk-fat products as well, and J/I exceeded it in two, because the loss of butyrate from the triacylglycerols lowers the numerator of B/A while free fatty acids contribute to regions A and J but not to the glycerol region I. A screen based on these ratios alone would therefore report some lipolyzed cheeses as adulterated. Second, region B contains the n-3 methyl triplet, so that a fat rich in α-linolenate, such as the canola-type oil of the prepared cheese loaf, shows a B/A of 0.12 that understates its lack of butyrate.

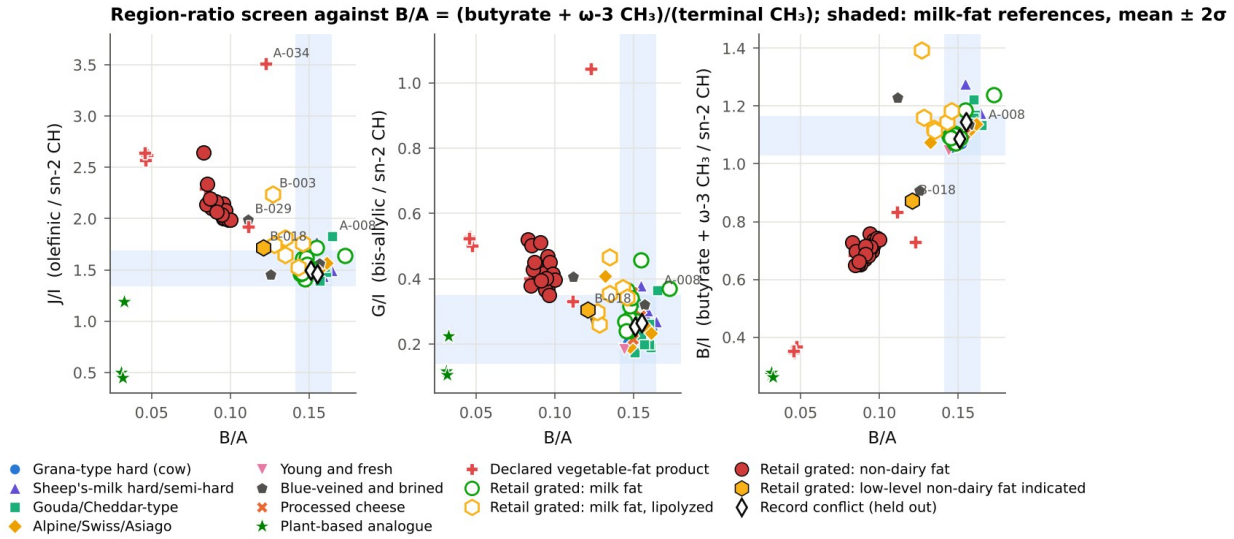

Figure 11. Region-ratio screen. J/I (olefinic/sn-2 CH), G/I (bis-allylic/sn-2 CH) and B/I against B/A = (butyrate + n-3 CH_3)/(terminal CH_3) for the 84 products. Shaded bands: mean ± 2 s.d. of the milk-fat reference population.

5.3 Fatty-acid classes: published schemes and line fitting

The schemes of Tociu and of Hanganu and Chira were applied as published (Table 11). Both return a short-chain or butyrate class of about 13 mol % for the reference population (13.1 ± 0.5 and 13.2 ± 0.4), and both separate the non-dairy fat products, for which the medians were 8.3 and 8.6 mol %. The two schemes agree with one another to within 0.3 mol % across the whole set, which shows that the choice between counting chains from the α-methylene or from the methyl protons is of little consequence in these extracts. Subtraction of the whole linolenate fraction derived from G_1 lowers the Hanganu–Chira butyrate of the reference population to 12.1 mol %; when only the 72% of the n-3 methyl triplet that region B captures is subtracted (Section 3.4), the value is 12.4 mol %. The difference is small, but the correction is questionable for a second reason. The division of region G at δ 2.786 does not separate the two bis-allylic triplets cleanly: according to the fits, 22% of the linoleate triplet, namely its low-field line, falls in G_1, and 13 to 19% of the n-3 triplet falls in G_2. The window G_1 therefore returns 1.18 mol % of linolenate for the reference population where the fitted triplet returns 0.69, and 1.3 mol % for palm oil, which contains almost none. A correction built on G_1 thus removes more than the n-3 methyl signal that is present, and in a linoleate-rich fat most of what it removes is not n-3 at all.

These values exceed the butyrate content obtained by line fitting, 9.1 mol %, by three to four units, and the excess is not confined to milk fat. For the palm oil reference, which contains no butyrate, the two integral-based schemes return 3.0 mol %, and for the tallows, lard and vegetable oils they return 3.6 to 5.3 mol % (Figure 12a). The cause is that the integration window of region B, as defined in the published schemes and adopted here, includes the low-field flank of the intense terminal methyl triplet and the n-3 methyl triplet. The integral-based estimators therefore carry a floor of 3 to 5 mol %, so that a fat without any butyrate

appears to contain a quarter to a third of the butyrate of milk fat, whereas the fitted triplet has a floor below 1 mol %. For screening, where only a relative comparison against authentic samples measured in the same way is required, the published schemes are adequate. For the quantitative estimate of non-milk fat the fitted butyrate is necessary: entered into the dilution model without correction for the floor, the Tociu short-chain class would return a median of 37% non-milk fat for the 22 retail products instead of 50%. The composition of region B in the reference population shows why: of its area, 66% is the butyrate triplet, 3% the captured part of the n-3 triplet, and 31% the flank of the main methyl triplet together with the third triplet.

The three Hanganu–Chira descriptors were examined in the same way (Figure 13d–f). Descriptor D3 = B/(A + B) is the butyrate estimate in another form and separates the groups well. Descriptors D1 and D2 as published were displaced, for the high-moisture products, by the water singlet inside region D: the fresh, processed and plant-based products, and one of the two products with conflicting records, moved along D1 by as much as three units for reasons unrelated to their fat. When region D was replaced by region F the scatter of the reference population in D1 fell from a relative standard deviation of 4.1% to 1.4% and that in D2 from 5.3% to 3.7%, and the water-robust D1w separated every non-dairy fat product from the reference population (12.45–12.87 against 11.65 ± 0.17). The published descriptors are therefore best used in this water-robust form whenever the extract has not been dried.

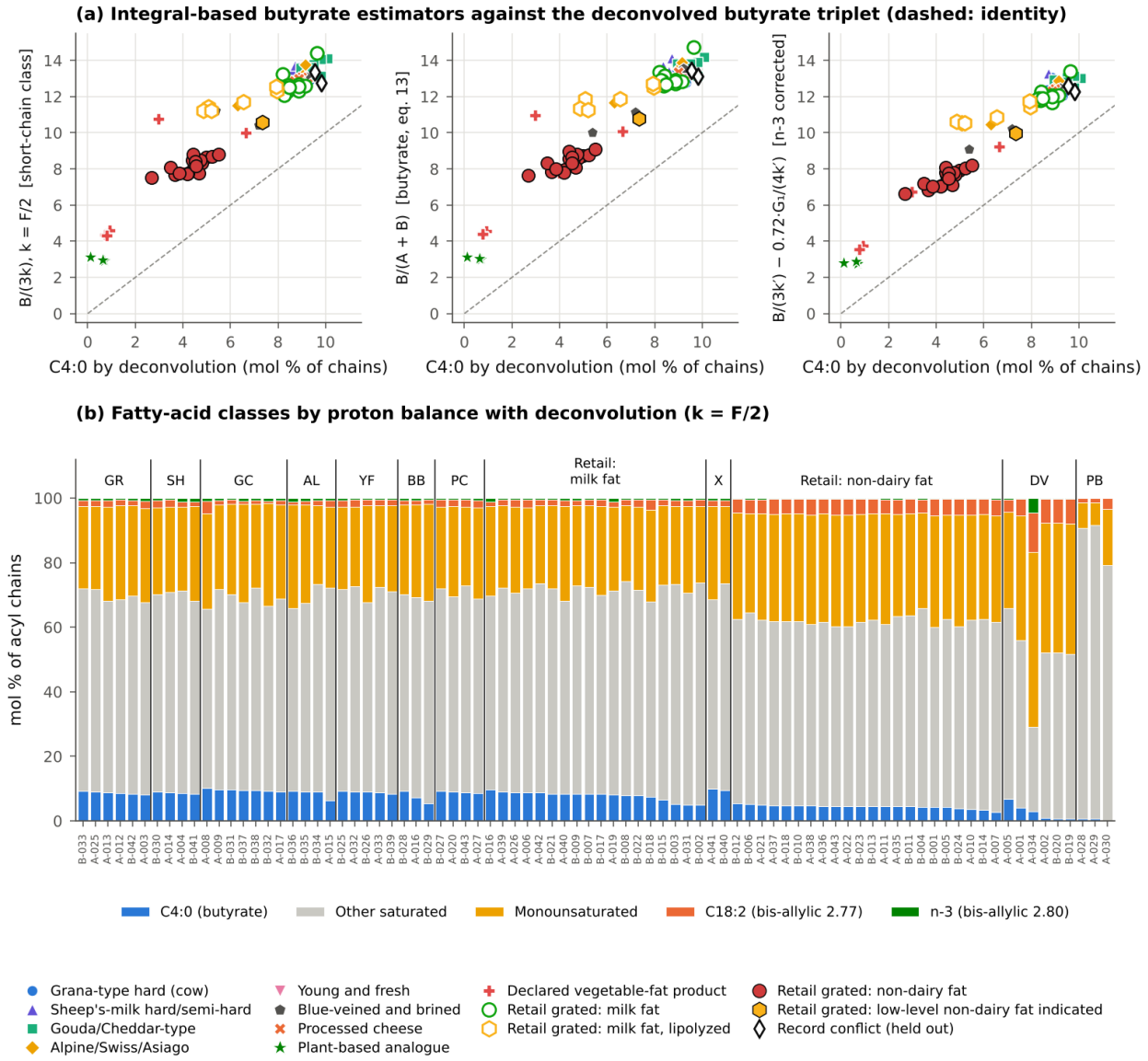

Figure 12. (a) Butyrate estimated by the integral-based schemes of Tociu and of Hanganu and Chira (without and with the G_1 correction, the latter scaled to the captured part of the n-3 triplet) against the butyrate obtained by fitting the methyl triplet; the dashed line is the identity. (b) Fatty-acid classes (mol % of acyl chains) of all 84 products and of three reference fats from the fitted C4:0, C18:2 and n-3 triplets and the allylic balance, grouped by class and ordered by decreasing butyrate within each group.

Table 11. Group medians of the region ratios, of the Hanganu–Chira descriptors as published (D1, D2, D3) and in water-robust form (D1w, D2w), and of the four Tociu classes (mol %). The last rows give the mean and standard deviation of the milk-fat reference population and the values of the palm oil reference.

Group	n	B/A	J/I	G/I	B/I	D1	D2	D3	D1w	D2w	SCSFA	PUFA	MUFA	LCSFA
Grana-type hard (cow)	6	0.150	1.49	0.25	1.07	11.74	26.3	0.131	11.68	26.1	12.8	4.5	25.7	56.6
Sheep's-milk hard/semi-hard	4	0.159	1.53	0.29	1.16	11.16	25.4	0.137	11.13	25.3	13.6	5.2	24.6	56.7
Alpine, Swiss, Asiago	4	0.152	1.54	0.24	1.10	11.61	26.1	0.132	11.56	26.0	13.1	4.4	25.6	57.3
Gouda/Cheddar-type	7	0.160	1.52	0.20	1.14	11.66	26.1	0.138	11.63	26.0	13.6	3.5	26.8	55.3
Blue-veined and brined	3	0.125	1.56	0.32	1.14	11.31	26.0	0.111	11.31	26.0	11.2	5.5	25.9	57.1
Young and fresh	5	0.149	1.50	0.24	1.07	12.13	27.4	0.130	11.72	26.1	12.8	4.4	23.9	58.8
Processed cheese	4	0.149	1.54	0.26	1.10	12.18	27.3	0.130	11.76	26.2	12.9	4.7	24.3	57.9
Declared vegetable-fat product	6	0.066	2.59	0.51	0.52	13.49	20.8	0.061	13.27	20.4	6.2	9.5	38.4	47.7
Plant-based analogue	3	0.032	0.50	0.12	0.27	11.05	45.0	0.031	10.44	42.9	3.0	1.9	7.3	87.9
Retail grated: milk fat	11	0.148	1.51	0.28	1.09	11.48	26.5	0.129	11.47	26.4	12.6	4.9	23.6	59.5
Retail grated: milk fat, lipolyzed	6	0.135	1.75	0.35	1.15	11.67	28.2	0.119	11.65	28.2	11.6	5.5	22.3	61.7
Retail grated: low-level non-dairy fat	1	0.121	1.72	0.31	0.87	12.29	24.9	0.108	12.12	24.5	10.6	5.6	26.5	57.4
Retail grated: non-dairy fat	22	0.094	2.09	0.42	0.71	12.62	23.1	0.086	12.59	23.1	8.3	7.4	30.8	53.6
Milk-fat reference set, mean	24	0.153	1.52	0.24	1.10	11.89	26.4	0.132	11.65	25.9	13.1	4.4	25.7	56.8
Milk-fat reference set, s.d.		0.006	0.09	0.05	0.03	0.49	1.4	0.004	0.17	1.0	0.5	0.9	2.1	2.6
Palm oil (reference fat)	1	0.031	3.05	0.52	0.22	14.24	17.1	0.030	14.12	17.0	3.0	10.7	47.3	39.0

Table 12. Group medians of the composition obtained by line fitting and of the minor markers. C4:0, C10:1, conjugated diene (CLA), C18:2, n-3 acids (as C18:3), MUFA and SFA in mol % of acyl chains; sterol in mol per 100 mol of glycerol; NC, mean carbons per chain; free acids and partial glycerides (FFA + PG) in % of chains.

Group	n	C4:0	C10:1	CLA	Sterol	C18:2	C18:3	MUFA	SFA	NC	C_2/C_1	G_1/G_2	FFA + PG
Grana-type hard (cow)	6	8.77	0.48	0.28	0.74	1.86	0.64	28.5	69.2	14.25	0.478	1.03	0.5
Sheep's-milk hard/semi-hard	4	8.65	0.37	0.68	0.63	1.90	0.91	26.6	70.5	13.68	0.506	1.23	1.6
Alpine, Swiss, Asiago	4	9.01	0.56	0.56	0.72	1.38	0.81	27.8	69.8	14.13	0.476	1.37	1.1
Gouda/Cheddar-type	7	9.51	0.57	0.44	0.78	1.17	0.81	29.1	68.9	14.20	0.479	1.49	0.4
Blue-veined and brined	3	7.19	0.41	0.40	0.72	1.41	0.62	28.5	69.4	13.88	0.484	1.10	1.3
Young and fresh	5	8.99	0.46	0.28	0.70	1.89	0.56	25.5	71.8	14.29	0.459	0.99	0.3
Processed cheese	4	8.93	0.45	0.31	0.72	2.16	0.58	26.7	70.8	14.33	0.465	0.99	0.7
Declared vegetable-fat product	6	1.97	0.11	0.11	0.39	7.50	0.31	40.1	52.1	16.56	0.568	0.40	4.9
Plant-based analogue	3	0.64	0.02	−0.00	0.14	1.27	0.05	7.9	90.8	12.60	0.408	0.37	3.9
Retail grated: milk fat	11	8.47	0.46	0.27	0.73	1.81	0.58	25.6	72.0	13.99	0.471	0.97	1.7
Retail grated: milk fat, lipolyzed	6	5.86	0.46	0.26	0.74	1.86	0.58	24.3	73.3	14.19	0.475	1.00	7.8
Retail grated: low-level non-dairy fat	1	7.35	0.36	0.17	0.70	3.20	0.42	28.5	67.9	14.78	0.475	0.63	2.1
Retail grated: non-dairy fat	22	4.53	0.23	0.13	0.55	4.60	0.36	33.1	61.9	15.37	0.528	0.51	4.6
Milk-fat reference set, mean	24	9.08	0.50	0.40	0.73	1.77	0.69	27.6	69.9	14.22	0.471	1.19	0.5
Milk-fat reference set, s.d.		0.47	0.08	0.23	0.05	0.60	0.16	2.4	2.4	0.18	0.018	0.27	0.4
Palm oil (reference fat)	1	0.07	−0.09	−0.06	0.19	9.57	0.40	48.0	42.0	17.45	0.688	0.32	1.1

The composition obtained by line fitting is summarized in Table 12 and Figure 12b. The retail grated products with non-dairy fat form a remarkably uniform group: butyrate 2.7–5.5 mol % (median 4.5), linoleate 4.2–5.1 mol % (median 4.6, against 1.8 for the reference population), fitted n-3 acids reduced to 0.36 mol %, monounsaturated acids raised to 33 mol % and a mean chain length of 15.2–15.7 carbons. In standard scores their butyrate lay between 7.6 and 13.6 standard deviations below the reference mean, their linoleate 4.0 to 5.6 above and their mean chain length 5.3 to 8.0 above (Figure 13a and b). The eleven retail

grated products assigned to the milk-fat category were indistinguishable from the reference population in every descriptor.

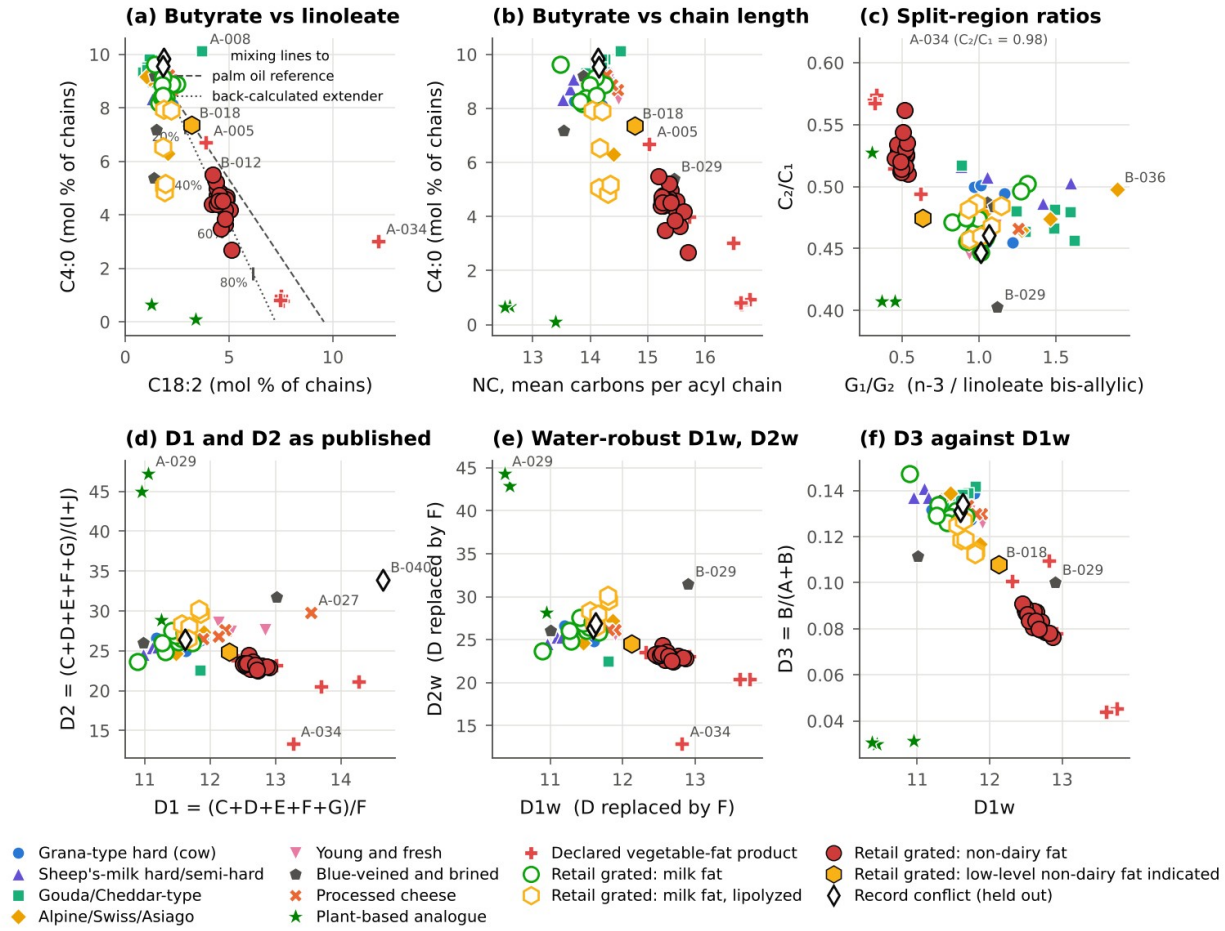

Figure 13. Lipid descriptor maps of the 84 products. (a) Fitted butyrate against linoleate, with mixing lines from the reference mean to the palm oil reference and to the back-calculated replacing fat. (b) Butyrate against mean chain length NC. (c) The two split-region ratios, C_2/C_1 against G_1/G_2. (d) Hanganu–Chira descriptors D1 and D2 as published. (e) Water-robust D1w and D2w. (f) D3 against D1w.

5.4 The split regions and the minor markers

The division of region G was introduced so that linolenate would not be counted as linoleate, and its value in the present data is as an indicator of the type of fat. The ratio G_1/G_2 was 1.19 ± 0.27 in the reference population, 1.3 to 2.1 in the ruminant tallows, 1.4 to 1.9 in the Pecorino Romano, Emmental and Gruyère cheeses, and 0.24 to 0.58 in palm oil, lard, the poultry fats and the vegetable oils. In the non-dairy fat retail products it fell to 0.46–0.54. The ratio therefore indicates a replacing fat that is poor in n-3 acids, which argues against canola and soybean oils and is consistent with a palm-type fat, and it singled out the prepared cheese loaf, whose G_1/G_2 of 1.0 with a linoleate content of 12 mol % and n-3 content of 4.6 mol % points to a canola- or soybean-type oil. As a criterion for the detection of a foreign fat, however, G_1/G_2 did not reach the three-sigma limit in any of the 22 products (Figure 13c). Two causes combine. The natural variation of n-3 acids in milk fat is large, so that the limit lies at 0.38; and because the low-field line of the linoleate triplet lies inside G_1, the window ratio cannot fall much below 0.3 even for a fat without any n-3 acid, as the value of 0.32 for palm oil shows. The window-based split is thus a damped measure of the n-3 to linoleate

balance, and the fitted triplets are to be preferred where that balance matters. The total bis-allylic signal has the first of these limitations as well; the split adds information on the type of fat, not sensitivity.

The division of region C behaved differently. The ratio C_2/C_1 has a narrow natural distribution, and it rose from 0.471 in the reference population to 0.51–0.56 in the non-dairy fat products, 0.57 in the declared cheese products and 0.69 in palm oil, while it fell to 0.41 in the lauric fat of two of the plant-based analogues, whose chains are short and fully saturated. Twelve of the 22 non-dairy fat products exceeded the three-sigma limit. Because C_2/C_1 responds to unsaturation and to chain length in opposite senses, it distinguishes a lauric fat from a palm-type fat, both of which lack butyrate, and it does so from the most intense signal in the spectrum. Its limitation is that the ruminant tallows (0.44–0.50) are not distinguished from milk fat, so that it cannot stand alone.

Three minor markers supplied evidence that is independent of butyrate (Figure 14). Caproleate amounted to 0.50 ± 0.08 mol % of chains in the reference population and to 0.17–0.29 mol % in the non-dairy fat products, a reduction of about half, in step with butyrate; it was absent from all reference fats and from the plant-based analogues. Its value lies in its behavior on lipolysis, discussed in the next section. The sterol signal, at 0.73 ± 0.05 mol per 100 mol of glycerol in the reference population, fell to a median of 0.55 in the non-dairy fat products and to 0.14–0.24 in the plant-based analogues; it is not specific to cholesterol, since phytosterols resonate at the same position, and it is useful only in combination. The conjugated diene and the n-3 acids rose together in the cheeses from sheep's milk and from alpine milk: the two Pecorino Romano blocks and the two imported grated Pecorino Romano products contained 0.95–1.17 mol % of conjugated diene and 1.1–1.2 mol % of n-3 acids, and Gruyère and Emmental 0.8–1.2 and 0.9–1.1 mol %, against 0.2–0.4 and 0.5–0.7 mol % for the lowland cow's-milk cheeses[38,90,91,94]. One retail product labeled as grated Pecorino Romano (B-014) contained 0.08 mol % of conjugated diene and 0.33 mol % of n-3 acids together with non-dairy fat, and so carried neither the fat of a sheep's-milk cheese nor that of any cheese alone.

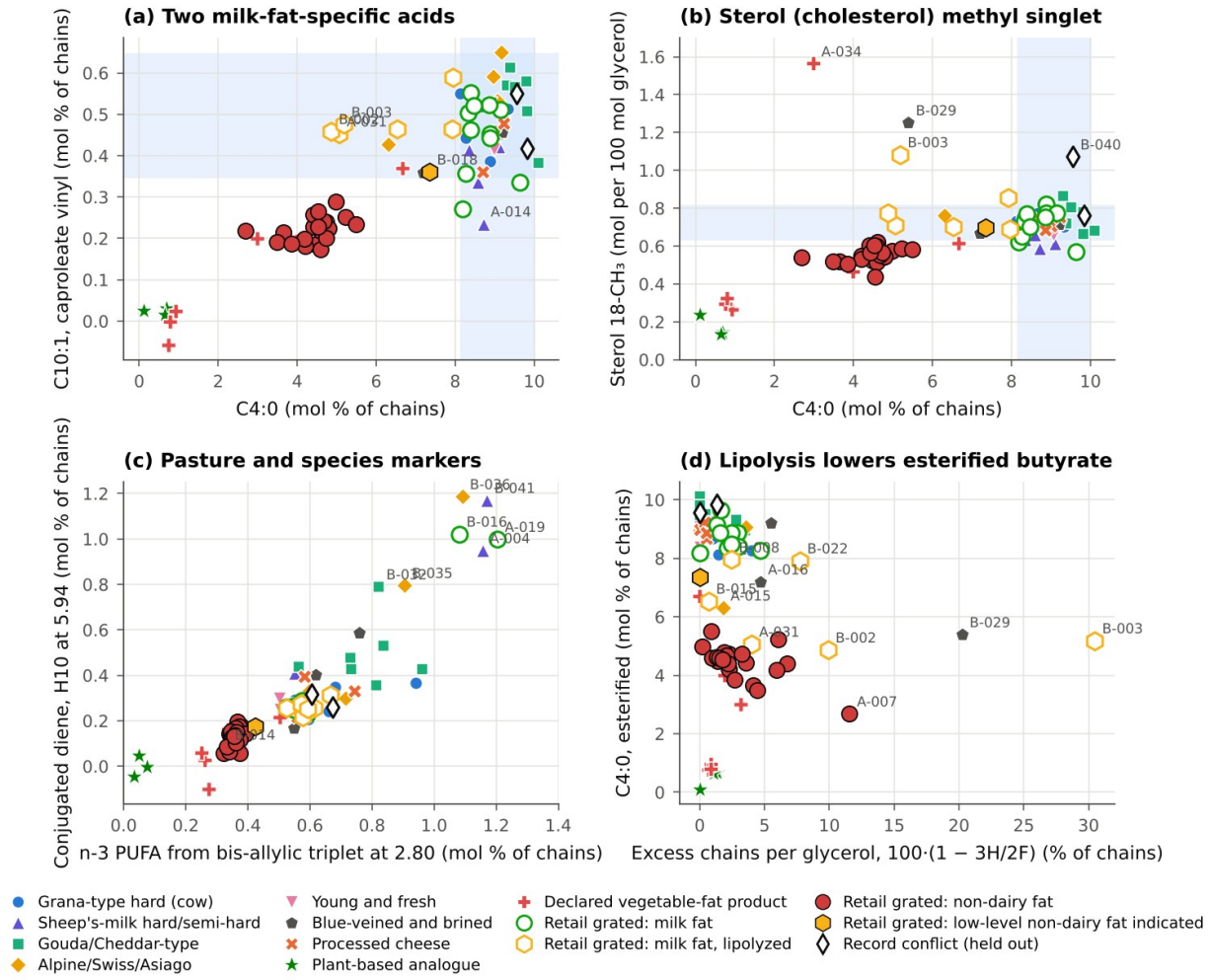

Figure 14. Minor lipid markers. (a) Caproleate (C10:1) against butyrate, the two acids specific to milk fat. (b) Sterol 18-CH_3 against butyrate. (c) Conjugated diene against n-3 acids, both raised in sheep's-milk and alpine cheeses. (d) Esterified butyrate against the excess of acyl chains over glycerol, which measures free fatty acids. Shaded bands: mean ± 2 s.d. of the milk-fat reference population.

5.5 Lipolysis and substitution distinguished

Six retail grated products showed reduced butyrate, 4.9 to 8.0 mol %, without any of the shifts that accompany a foreign fat: their linoleate (1.8–2.2 mol %), mean chain length (14.0–14.4) and C_2/C_1 (0.46–0.49) were those of milk fat. All six contained free fatty acids and partial glycerides amounting to 6 to 26% of the acyl chains, against a reference mean of 0.5%, and in the most extreme case, a grated Romano (B-003), the number of chains per glycerol reached 4.3, which means that about a third of the fatty acids were no longer bound in triacylglycerols. Lipolysis of this degree is expected in Romano-type cheeses made with pregastric lipase and in products to which enzyme-modified cheese has been added for flavor[83,88,101].

The caproleate marker supports the diagnosis. In the three most lipolyzed products the butyrate deficit would correspond to 43–46% of non-milk fat if it were read as dilution, whereas the caproleate content, 0.45–0.48 mol %, was normal and corresponds to 4–9%. The explanation is that butyric acid, once released, is lost from the chloroform extract by partition into the aqueous phase and by volatility, while the released ten-carbon acid remains in the lipid extract and continues to contribute its vinyl signal. The missing butyrate was in fact found in the aqueous extracts of the same products, where the β-methylene signal of free butyrate

was up to three times as strong as in the non-lipolyzed cheeses (Section 6.5). A product in which butyrate and caproleate are reduced in the same proportion has been diluted; a product in which butyrate alone is reduced has been lipolyzed. Free fatty acids were present in many of the non-dairy fat products as well, at 3 to 8% of chains and in one case 18%, and partial glycerides are a normal constituent of palm oil[106], so that the presence of free acids is not in itself evidence for or against substitution.

5.6 Quantifying the non-milk fat

For the 22 retail grated products with non-dairy fat, the consensus estimate of the non-milk fat fraction ranged from 43 to 68% of the acyl chains, with a median of 51% (Table 13 and Figure 15c). On a mass basis this corresponds to about 46 to 70% of the fat, with a median near 54%. The estimate from butyrate alone (median 50%, range 39–70%) and that from caproleate alone (median 55%, range 42–65%) agreed within the uncertainty of the less precise caproleate value (Figure 15a). The standard error of the consensus value, calculated as described in Section 3.6, was 3.5 to 4.2 percentage points, so that with two standard errors each product is known to within about ± 7 to 8 points. That interval covers the natural variability of milk fat and the variation observed between retail lots. It does not cover a different recovery of the two fats on extraction, a relaxation bias that differs between them, or the error of the lineshape model, none of which could be assessed from the present data, and the true uncertainty is therefore larger by an unknown amount.

The highest value, 68% for A-007, calls for a qualification. This product combined non-dairy fat with pronounced lipolysis, 18% of its acyl chains being present as free acids and partial glycerides and the number of chains per glycerol reaching 3.39. Its butyrate will have been lowered by lipolysis as well as by dilution, and its caproleate estimate, 56 ± 16%, is the more reliable of the two. Without it the 21 remaining products lie between 43 and 62%. A partial least squares model calibrated on synthetic blends and applied to the binned spectra, which uses no fitted quantity, returned a median of 53% (range 42–73%) for these 21 products (Figure 15b); for A-007 it returned 113%, a value outside the range of the calibration that is also identified as such by its T^2 statistic, and it is not reported. Two approaches with different assumptions thus place the typical product at about one half non-milk fat, with the reservation that both rest on the same spectra and on the same reference population.

The composition of the replacing fat, back-calculated product by product, was as uniform as that of the products themselves: linoleate 7.3 mol % (interquartile range 7.0–7.7), no measurable n-3 acids, monounsaturated acids 39 mol %, saturated acids 53 mol %, a mean chain length of 16.4 carbons and a C_2/C_1 of 0.58. This is not the composition of the single whole palm oil in the reference panel (linoleate 9.6, monounsaturated 48, saturated 42 mol %, 17.5 carbons, C_2/C_1 0.69); the replacing fat is more saturated and shorter in mean chain length. It is close to the fat of the three declared grated cheese products in the set (linoleate 7.5, monounsaturated 40, saturated 52 mol %, 16.7 carbons, C_2/C_1 0.57). The description of the fat as consistent with a palm-type fat rests on this gross composition and on the low n-3 content. No marker specific to palm was observed, one reference oil cannot represent the range of palm oils and palm fractions in commerce, and a blend of palm fractions with a shorter-chain saturated fat, or another fat of similar gross composition, would fit equally well. What the data support is that the fat of the 22 retail products resembles the fat of the declared cheese products in the descriptors measured. They do not identify the oil, and a similar composition does not show that the products share a supplier, a formulation or a manufacturing route.

The declared products behaved as their labels would lead one to expect, which is a useful check on the method. The grated blend that declares palm oil alongside Parmigiano Reggiano contained 26 ± 10% of non-milk fat; the reduced-fat topping labeled as imitation cheese 57 ± 7%; and the prepared cheese loaf 65 ± 7%, of a canola- or soybean-type oil. In the three grated cheese products the fitted butyrate, 0.8 to 0.9 mol %, lay below the floor of 1.27 mol % and the caproleate signal was not distinguishable from zero, so that only a

lower bound of 86% can be stated; whether these products contain a few percent of milk fat or none cannot be decided by this method. In version 2.0 one of them was reported at 100 ± 0%, a figure that arose because the earlier formula assigned a vanishing uncertainty to a marker near zero and allowed a slightly negative caproleate value to dominate the weighted mean. The same lower bound applies to the three plant-based analogues. For the lauric fat of the plant-based analogues and for the canola-type loaf the partial least squares model lay outside its calibration domain, by its residual statistic or by a negative prediction, a reminder that a regression model answers only within the composition space on which it was trained, whereas the stoichiometric estimate has no such restriction.

One retail product, B-018, fell between the categories. Its butyrate (7.35 mol %, 3.7 standard deviations below the reference mean), caproleate (0.36 mol %), linoleate (3.2 mol %) and mean chain length (14.78, three standard deviations above the mean) all moved in the direction of an admixture, without lipolysis, and the consensus estimate was 21 ± 10%. An interval of that width does not exclude a value near 10%, which would lie within the range of authentic milk fat, and in the simulated blends only a minority were flagged at this level (Section 7.5). The result is reported as an indication only.

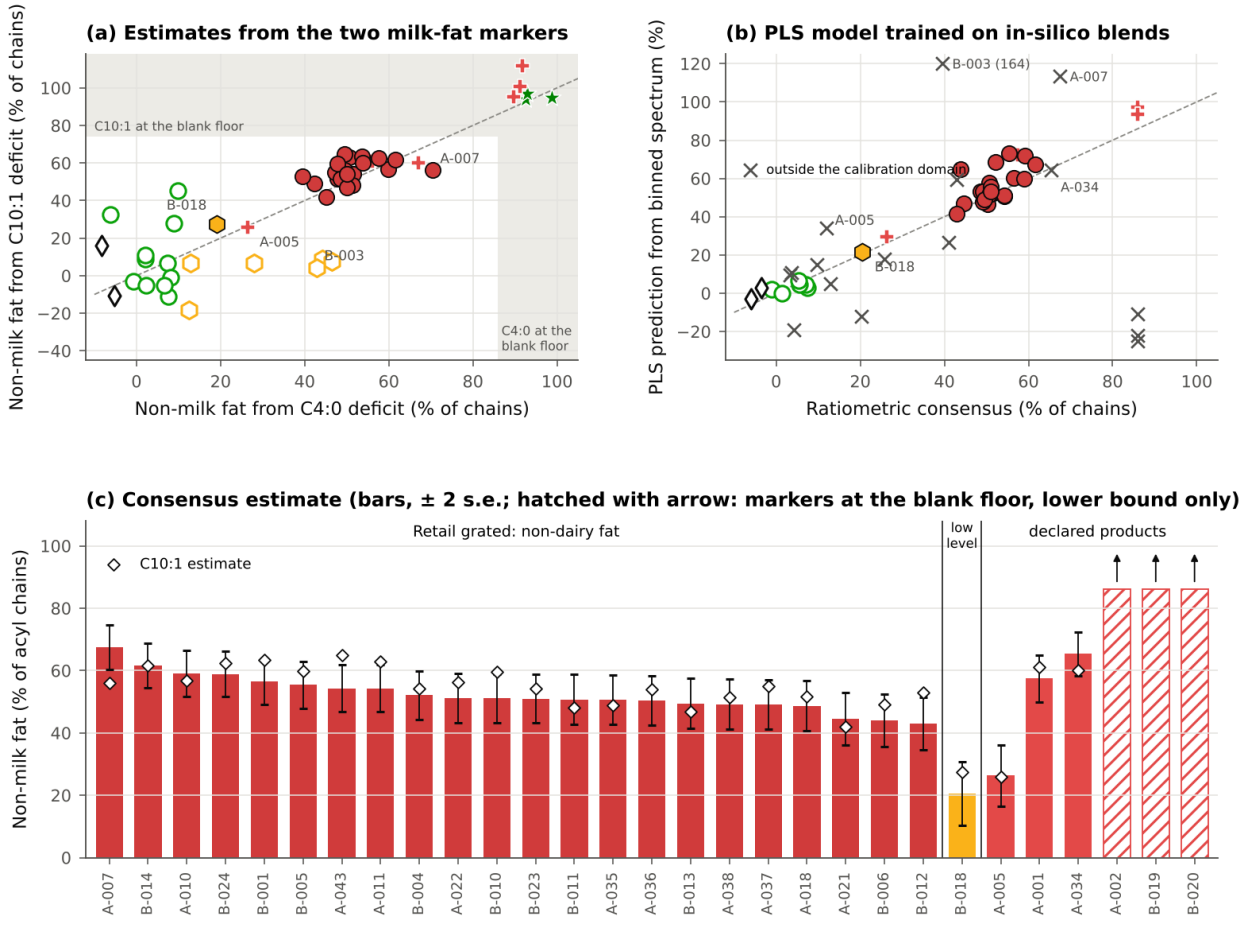

Figure 15. Estimates of the non-milk fat fraction. (a) Estimate from the caproleate deficit against that from the butyrate deficit for the retail grated, declared and plant-based products; lipolyzed milk-fat products depart from the identity line because only their butyrate is reduced. The shaded bands mark the range in which a marker lies at or below the level found in fats without milk fat. (b) Prediction of a partial least squares model trained on in-silico blends against the ratiometric consensus; crosses mark products outside the calibration domain of the model. (c) Consensus estimate for each product with non-dairy fat (bars, ± 2 s.e. as defined in Section 3.6) and the caproleate estimate (diamonds); hatched bars with an arrow are lower bounds for products whose markers lie at the blank level.

Table 13. Products flagged for non-dairy fat. Fitted butyrate, caproleate and linoleate (mol % of chains), mean chain length NC, C_2/C_1, and the non-milk fat fraction (% of acyl chains) estimated from the butyrate deficit, from the caproleate deficit and as the consensus of the two, each ± 2 s.e. (Section 3.6), and by the PLS model. A value preceded by ≥ is a lower bound for a marker at or below its blank floor; o.d., outside the calibration domain of the PLS model. Retail grated products are ordered by decreasing consensus value; for A-007, which is also lipolyzed, the caproleate estimate is the more reliable.

Code	Group	C4:0	C10:1	C18:2	NC	C_2/C_1	f (C4:0)	f (C10:1)	Consensus	PLS	Lipid category
A-007	Retail grated	2.68	0.22	5.13	15.70	0.562	70 ± 8	56 ± 16	68 ± 7	o.d.	non-dairy fat
B-014	Retail grated	3.48	0.19	4.62	15.30	0.535	62 ± 8	62 ± 14	62 ± 7	67	non-dairy fat
A-010	Retail grated	3.65	0.22	4.80	15.55	0.534	60 ± 8	57 ± 15	59 ± 7	72	non-dairy fat
B-024	Retail grated	3.86	0.19	4.81	15.48	0.544	58 ± 9	62 ± 14	59 ± 7	60	non-dairy fat
B-001	Retail grated	4.20	0.18	5.05	15.51	0.523	54 ± 9	63 ± 14	57 ± 7	60	non-dairy fat
B-005	Retail grated	4.19	0.20	4.79	15.63	0.534	54 ± 9	60 ± 15	55 ± 8	73	non-dairy fat
A-043	Retail grated	4.59	0.17	4.83	15.37	0.530	49 ± 9	65 ± 13	54 ± 7	51	non-dairy fat
A-011	Retail grated	4.47	0.18	4.50	15.40	0.532	51 ± 9	63 ± 14	54 ± 8	51	non-dairy fat
B-004	Retail grated	4.40	0.23	4.16	15.28	0.510	51 ± 9	54 ± 16	52 ± 8	69	non-dairy fat
A-022	Retail grated	4.58	0.22	4.86	15.50	0.531	50 ± 9	56 ± 16	51 ± 8	52	non-dairy fat
B-010	Retail grated	4.74	0.20	4.51	15.24	0.530	48 ± 9	59 ± 15	51 ± 8	56	non-dairy fat
B-023	Retail grated	4.54	0.23	4.68	15.40	0.520	50 ± 9	54 ± 16	51 ± 8	53	non-dairy fat
B-011	Retail grated	4.41	0.26	4.42	15.22	0.512	51 ± 9	48 ± 18	51 ± 8	52	non-dairy fat
A-035	Retail grated	4.44	0.25	4.57	15.27	0.526	51 ± 9	49 ± 18	51 ± 8	58	non-dairy fat
A-036	Retail grated	4.61	0.23	4.51	15.25	0.527	49 ± 9	54 ± 16	50 ± 8	47	non-dairy fat
B-013	Retail grated	4.53	0.26	4.48	15.26	0.515	50 ± 9	47 ± 18	49 ± 8	49	non-dairy fat
A-038	Retail grated	4.67	0.24	4.88	15.38	0.529	49 ± 9	51 ± 17	49 ± 8	53	non-dairy fat
A-037	Retail grated	4.79	0.22	4.62	15.28	0.528	47 ± 9	55 ± 16	49 ± 8	48	non-dairy fat
A-018	Retail grated	4.74	0.24	4.40	15.31	0.525	48 ± 9	52 ± 17	49 ± 8	53	non-dairy fat
A-021	Retail grated	4.98	0.29	4.40	15.40	0.512	45 ± 9	42 ± 20	45 ± 8	47	non-dairy fat
B-006	Retail grated	5.23	0.25	4.35	15.37	0.517	42 ± 10	49 ± 18	44 ± 8	65	non-dairy fat
B-012	Retail grated	5.49	0.23	4.22	15.19	0.516	39 ± 10	53 ± 17	43 ± 8	42	non-dairy fat
B-018	Retail grated	7.35	0.36	3.20	14.78	0.475	19 ± 11	27 ± 24	21 ± 10	22	low-level non-dairy fat indicated
A-001	Declared product	3.99	0.19	5.00	15.73	0.514	56 ± 9	61 ± 14	57 ± 7	72	non-dairy fat
A-002	Declared product	0.94	0.02	7.53	16.78	0.570	≥ 86	≥ 74	≥ 86	97	non-dairy fat; milk-fat markers at the blank level
A-005	Declared product	6.68	0.37	3.89	15.03	0.494	26 ± 11	26 ± 24	26 ± 10	29	non-dairy fat
A-034	Declared product	3.00	0.20	12.20	16.50	0.979	67 ± 8	60 ± 15	65 ± 7	o.d.	non-dairy fat
B-019	Declared product	0.76	−0.06	7.61	16.64	0.573	≥ 86	≥ 74	≥ 86	94	non-dairy fat; milk-fat markers at the blank level
B-020	Declared product	0.80	−0.00	7.46	16.62	0.567	≥ 86	≥ 74	≥ 86	93	non-dairy fat; milk-fat markers at the blank level
A-028	Plant-based	0.68	0.03	1.27	12.60	0.408	≥ 86	≥ 74	≥ 86	o.d.	no milk-fat marker detected
A-029	Plant-based	0.64	0.01	1.26	12.51	0.407	≥ 86	≥ 74	≥ 86	o.d.	no milk-fat marker detected
A-030	Plant-based	0.11	0.02	3.38	13.39	0.527	≥ 86	≥ 74	≥ 86	o.d.	no milk-fat marker detected

5.7 Consistency check on 43 further lipid spectra

The decision rule and the reference limits were fixed on the 84 paired products and were then applied without change to the 43 lipid spectra that had been set aside (Figure 16). Nineteen were assigned to the non-dairy fat category, with consensus fractions of 44 to 62% (median 55%); 18 were milk fat and five were lipolyzed milk fat; and in one, the declared dairy-free Parmesan alternative, no milk-fat marker was detected (butyrate 0.06, linoleate 9.6, n-3 acids 1.6 mol %, 17.1 carbons), as its label requires. The 19 flagged spectra occupied the same compact region of descriptor space as the 22 of the study set, and no spectrum fell

between the two groups. One spectrum does expose the gap in the rule that was noted in Section 3.6: H-001, a grated Parmesan and Romano product, had a butyrate content 3.4 standard deviations below the reference mean with normal caproleate, linoleate and chain length, and free acids and partial glycerides of 2.9% of chains, just under the lipolysis limit of 3%. It was assigned to milk fat because none of the further conditions was met, although its pattern is that of a mildly lipolyzed fat. Two block cheeses, among them a Parmigiano Reggiano, showed smaller deficits of the same kind.

The weight of this check must be stated carefully. The 19 flagged spectra do not represent 19 products: they come from five products of four store brands, and 17 of them are separate purchases of three of those products. Three of the four brands are also represented among the flagged products of the paired set. The check therefore shows that the rule returns the same answer for further lots of products of the kind on which it was developed, acquired on other dates; it does not show how the rule performs on unrelated products, and it cannot serve as an independent validation.

The repeated purchases give the only measure of reproducibility available in the study (Table 14). For the three products the fitted butyrate content varied between lots with standard deviations of 0.21, 0.47 and 0.23 mol %, and the estimated non-milk fat with 1.3, 4.5 and 2.3 percentage points. Each lot was extracted and measured once, so that the variation contains the repeatability of the whole procedure together with any real difference between lots, and the two cannot be separated. The pooled values, 0.33 mol % for butyrate and 0.019 mol % for caproleate, were used as the between-preparation term of the uncertainty in Section 3.6. Replicate extractions of single lots are needed to establish the repeatability proper.

Table 14. Variation between separate retail lots of three products in the lipid-only set (mean ± standard deviation). Each lot was extracted and measured once.

Product	Lots	Spectra	C4:0 (mol %)	C10:1 (mol %)	C18:2 (mol %)	Non-milk fat (% of chains)
Grated Parmesan, brand 1	4	H-002, H-014, H-015, H-016	4.15 ± 0.21	0.211 ± 0.022	4.73 ± 0.20	55.3 ± 1.3
Grated Parmesan, brand 2	6	H-019, H-030, H-031, H-032, H-033, H-034	4.07 ± 0.47	0.206 ± 0.021	4.79 ± 0.09	56.1 ± 4.5
Grated Parmesan, brand 3	7	H-024, H-036, H-037, H-038, H-039, H-040, H-041	4.21 ± 0.23	0.211 ± 0.016	4.79 ± 0.17	54.7 ± 2.3

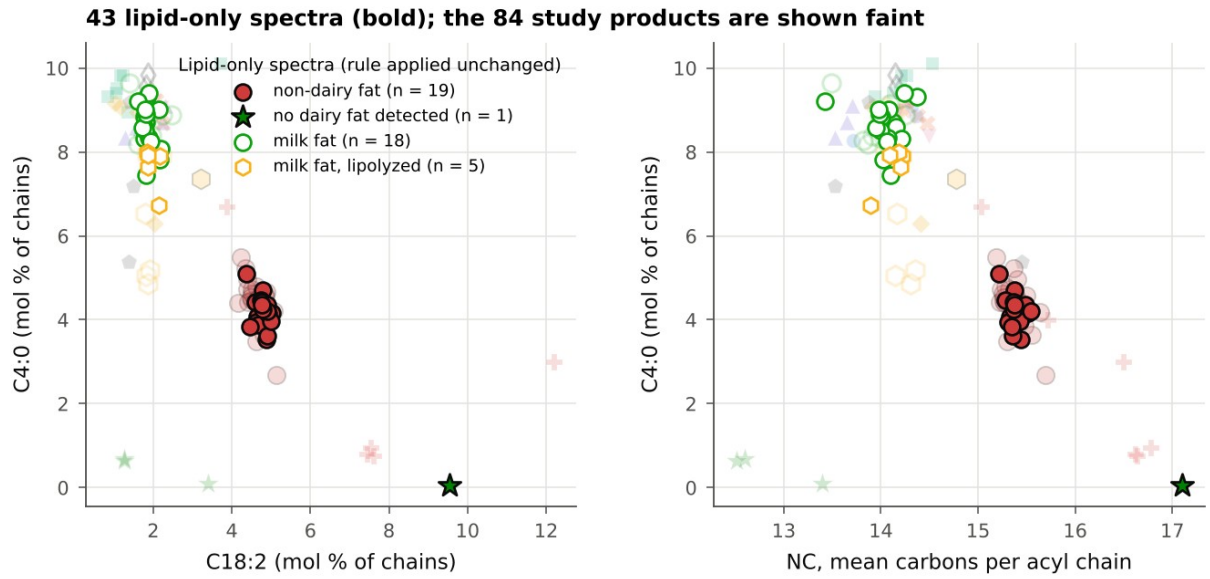

Figure 16. Consistency check: the 43 lipid-only spectra (bold symbols), classified by the fixed decision rule, on the butyrate–linoleate and butyrate–chain-length maps of the 84 study products (faint symbols). Seventeen of the 19 spectra with non-dairy fat are separate lots of three products.

6. Chemistry of the Aqueous Extracts

6.1 Proteolysis: the free amino-acid index

The free amino-acid index separated the products into two populations with almost no overlap (Figure 17a and Table 15). The ripened reference cheeses gave medians of 45 for the Grana type, 46 for the alpine cheeses and 36 for the Dutch-type and sheep's-milk cheeses, the lowest values within these classes belonging to the young Gouda, the fresh Asiago and the young sheep's-milk table cheese. The young and fresh cheeses, the processed cheeses and the declared cheese products gave medians of 0.7 to 1.7. The other amino-acid signals followed the same pattern, glutamate, lysine, alanine, phenylalanine and histidine being 20 to 100 times more abundant in the ripened cheeses than in the processed cheeses.

Among the retail grated products, the 16 with milk fat and a ripened profile (categories A and B of Section 8) gave a median of 35 and a range of 28 to 45, slightly below the Grana-type blocks. The 21 products of category C, with non-dairy fat, gave a median of 9.7 and a range of 4.4 to 16.4, all below one half of the median of the ripened references, which was taken as the limit. As a fraction of the spectral area their free amino acids were thus 28% of those of the retail products with milk fat, and about one fifth of those of the Grana-type reference cheeses.

What this difference means is less certain than its size suggests, because the index is a fraction of the total spectral area and falls whenever other material adds area to the spectrum. In the category C extracts the dispersed lipid, the sugars and the citrate do exactly that. Two other bases were therefore examined. Relative to lactate, which is independent of the total area, the index of category C had a median of 0.045 against 0.083 for categories A and B, a ratio of 54%, and the ranges of the two groups overlapped (0.025–0.069 against 0.054–0.117). Relative to the TSP-d_4 signal and the sample mass, a basis that would be absolute if the reference signal were constant, the free amino acids of category C were 35% and the lactate 57% of those of categories A and B; this comparison is indicative only, since TSP-d_4 binds to protein and its response differs between extracts (Section 4.7). On every basis the free amino acids of category C are lower, by a factor between two and four. The observation is compatible with a ripened cheese that supplies only part of the

water-soluble solids, the remainder bringing lactate but few amino acids; with a cheese component that is younger than the cheeses of categories A and B; or with both. The aqueous spectrum cannot distinguish these cases, and no conclusion on the curing time of the cheese ingredient is drawn from it.

6.2 Lactate, citrate and residual sugars

Lactate was the principal organic acid in every dairy product except those in which it had been consumed, by propionibacteria in Emmental and by mold in the blue cheeses, or had never been formed, as in the directly acidified fresh mozzarella. Citrate was absent from the Grana-type, sheep's-milk and alpine references and present only in traces in the Dutch-type cheeses, in which it is fermented during the first weeks[80,95]. It was present at 42 to 66 in the young and fresh cheeses, the processed cheeses and the declared cheese products, in the first case because the native citrate of milk had not yet been metabolized and in the others because citrate is added as an emulsifying salt or arrives with unfermented dairy solids. All 21 category C retail products contained citrate, at a median of 50 (Figure 17b). Small citrate signals of 11 to 13, just above the reporting limit of 10, were found in five of the 16 retail products of categories A and B.

Free sugars were absent from all ripened references, the sum of the lactose, galactose and α-anomer indices being 2 to 7. In the processed cheeses and the declared cheese products the sugar was mainly lactose (190 and 440), as expected where whey or milk solids have been added. In the low-moisture mozzarellas galactose predominated, as it does wherever a galactose-negative thermophilic starter has been used[97,98]. The category C retail products contained very little lactose (2.7) but appreciable galactose (26) and α-anomeric signal at δ 5.23 (37), with a small envelope of α-(1→4)-linked glucan that was absent from the ripened cheeses (Figure 17c). Galactose without lactose points to a young curd made with a thermophilic starter, or to lactose-hydrolyzed dairy solids; glucose-type anomers with a glucan envelope point to a starch-derived ingredient. Both would be foreign to a grated Parmesan, in which the sugars were exhausted long before the cheese was ready to grate, and neither interpretation can be taken further without the ingredient records.

6.3 Lipid dispersed in the aqueous extract

Fat does not dissolve in water, and the aqueous extract of a natural hard cheese carries only a small broad resonance of lipid at δ 1.3 and 0.9, from fine particles that escape separation. In the processed cheeses this resonance was the largest feature of the spectrum, with a median index of about 2000 against 50 to 240 for the ripened reference classes, together with broad olefinic and methyl resonances of the same origin (Figure 10 and Figure 17d). The emulsifying salts of processed cheese disperse the casein, which then stabilizes the fat as droplets small enough to remain suspended in the aqueous extract[12,14,100]. Twenty of the 21 category C retail products showed the same resonance at a median of 1700.

This marker is less specific than the others and must be read with them. Seven of the 16 retail products with milk fat and a ripened profile also exceeded the limit set by the highest ripened reference (400), with indices of 540 to 1150, and so did the buffalo-milk blue cheese. Fine grating, partial lipolysis, which generates surface-active soaps and partial glycerides, and a small particle size all favor the carry-over of fat, and the index depends on how the extract was clarified. Dispersed lipid in combination with low free amino acids, citrate and free sugars is the signature of a processed-type base (Figure 17e and f); dispersed lipid alone is not.

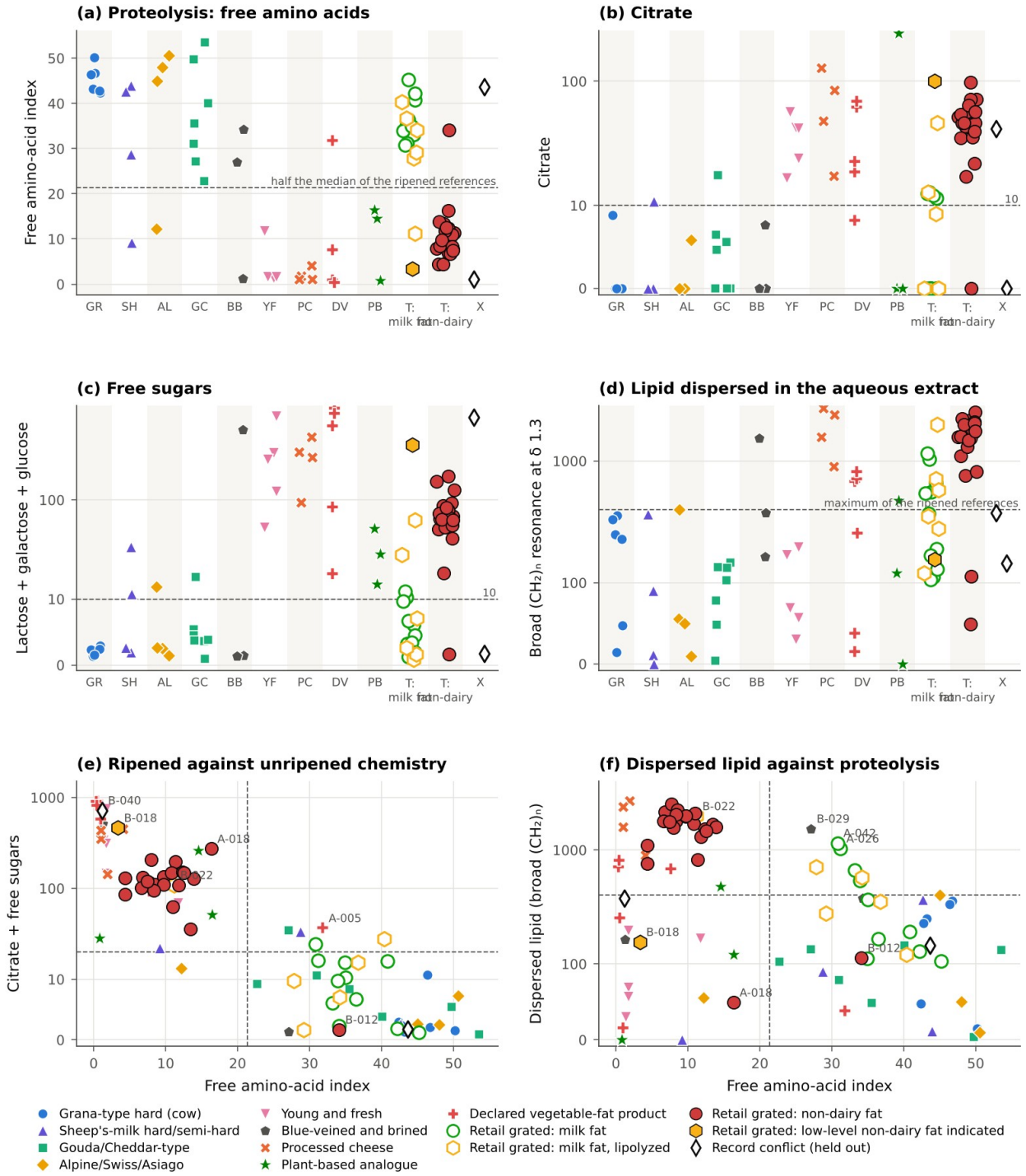

Figure 17. Principal aqueous markers by class (class codes as in Table 1; T, retail grated products, separated by lipid category) and their combinations. (a) Free amino-acid index. (b) Citrate. (c) Free sugars (lactose, galactose and α-anomers). (d) Dispersed lipid. (e) Citrate plus free sugars against the free amino-acid index. (f) Dispersed lipid against the free amino-acid index. Dashed lines: decision limits (Section 6.1 to 6.3). Indices per 10^4 of total spectral area; logarithmic axes are linear below the first tick.

Table 15. Group medians of the targeted aqueous features, per 10^4 of total spectral area and per proton. Retail grated products are grouped by the integrated categories of Section 8: A, milk fat with a ripened base; B, lipolyzed milk fat with a ripened base; C, non-dairy fat with an unripened or processed-type base.

Feature	GR	SH	AL	GC	BB	YF	PC	DV	PB	T: A	T: B	T: C
Number of products	6	4	4	7	3	5	4	6	3	11	5	21
Free amino-acid index (Leu, Ile, Val CH_3)	44.8	35.7	46.5	35.6	27.1	1.7	1.5	0.7	14.5	34.9	34.2	9.7
Alanine	46.8	36.9	38.4	37.0	42.9	3.3	0.9	6.3	3.9	43.7	40.1	11.7
Glutamate γ-CH_2	128.8	67.1	114.4	104.7	64.9	7.1	5.9	3.2	7.9	90.4	82.9	27.6
Lysine ε-CH_2	86.4	43.1	87.4	90.6	55.6	5.9	0.9	2.1	1.5	65.0	62.2	16.4
Phenylalanine (aromatic)	25.3	20.3	26.0	28.9	10.8	3.0	0.5	0.6	3.7	20.6	20.1	4.8
Tyrosine H3,5	8.5	6.3	6.9	6.5	9.1	0.9	0.2	0.4	1.0	9.2	9.9	3.0
Histidine H2	19.0	12.7	12.8	14.0	1.5	0.0	0.0	0.0	0.0	12.7	11.1	3.5
Methionine S-CH_3	19.9	13.9	19.3	16.6	17.6	0.5	0.2	0.7	0.3	16.5	13.5	3.2
Glycine	39.0	19.5	32.7	48.3	26.5	7.0	13.2	34.0	6.3	33.9	41.4	11.9
Lactate	368	502	249	316	135	679	229	87	19	449	376	211
Acetate	24.2	35.6	49.8	38.2	14.5	33.9	2.4	15.5	0.0	29.5	27.4	19.2
Propionate (quartet/triplet fit)	5.7	8.7	10.2	6.3	5.2	0.5	0.1	0.2	21.7	7.2	24.8	4.6
Butyrate β-CH_2 window	17.4	15.8	18.5	17.3	14.6	7.2	1.1	1.7	3.8	17.8	34.2	8.1
Succinate (tracked singlet)	10.5	6.0	5.7	7.4	0.0	3.0	0.2	0.0	0.0	15.8	18.0	5.2
Formate	2.0	4.9	4.9	3.8	0.2	2.6	0.6	1.0	4.2	2.2	1.4	1.6
Citrate (AB quartet fit)	0.0	0.0	0.0	4.7	0.0	41.9	66.3	42.0	0.0	0.0	0.0	50.2
Lactose (Gal H1)	1.5	1.5	1.7	2.9	1.1	67.9	189.7	438.8	1.9	0.9	0.4	2.7
Galactose (α-H1)	0.1	4.7	0.4	0.4	0.7	103.9	5.1	3.1	0.7	0.9	0.6	26.0
α-Glucose-type anomers (δ 5.23)	0.3	0.8	0.2	0.3	0.0	69.8	91.3	212.5	21.5	1.1	1.8	37.3
Free sugars, sum	2.2	7.0	2.5	3.8	1.4	256.7	286.6	653.1	28.2	4.5	2.6	68.7
α-(1→4) glucan envelope (δ 5.36–5.46)	3.8	6.3	5.4	2.7	22.3	29.3	61.3	77.2	603.9	7.8	7.4	25.8
Dispersed lipid, broad $(CH_2)_n$	238	51	53	104	374	69	1991	468	120	369	352	1714
Sorbate window (δ 6.12–6.33)	2.4	2.7	2.8	1.8	1.7	10.1	2.1	3.4	10.9	5.1	7.3	8.9
Apparent pH (His H2 shift)	5.50	5.50	5.62	5.66	5.65	–	5.53	5.63	5.59	5.49	5.46	5.77

6.4 Preservatives, starch and other formulation markers

Sorbate, recognized by its olefinic doublet at δ 5.83 and its multiplet at δ 6.12–6.33, was found in 33 of the 40 retail grated products, in all six declared vegetable-fat products and in two each of the processed cheeses and plant-based analogues, and in none of the block cheeses of the ripened reference classes. It is a permitted antimycotic and carries no weight as evidence of adulteration, but it is a nearly perfect marker of a retail grated product, and for that reason its windows were removed from the multivariate data (Section 3.9). The envelope of α-(1→4) glucan between δ 5.36 and 5.46 was very large in the plant-based analogues (median 600), moderate in the declared and processed products (60–80) and small but distinct in the category C retail products (26, against 3 to 6 for the ripened references).

In the three declared grated cheese products a sharp singlet at δ 6.19 was present that could not be accounted for by sorbate, whose signals in those spectra were weak. Its position is that of orotate, a characteristic minor constituent of bovine milk and whey that is consumed during lactic fermentation, and it would be consistent with the unfermented dairy solids that the large lactose signals of the same products indicate. The assignment is tentative and would need to be confirmed by spiking; in the other retail products the region is occupied by sorbate and could not be examined.

6.5 Short-chain acids: propionic fermentation, acetate and free butyrate

Emmental was distinguished from every other product by its short-chain acids (Figure 18 and Figure 19). The propionate methyl triplet at δ 1.05 and methylene quartet at δ 2.18 were prominent, the acetate singlet was six times as large as in the Grana-type cheeses, and lactate was depleted to 81 against 200 to 740 in the other alpine cheeses. This is the expected outcome of the propionic fermentation, in which three moles of lactate yield two of propionate and one each of acetate and carbon dioxide[92,93]. The molar ratio of propionate to acetate found here, about 0.8, is below the theoretical value of 2; co-metabolism of aspartate by the propionibacteria, which diverts the fermentation toward acetate and succinate, and acetate from other sources both lower the ratio in real cheese, and the fit of the propionate quartet in a crowded region is likely to be biased low, so the figure should not be over-interpreted. Gruyère, in which the propionic fermentation is slight, showed no propionate above the background of the fit and a normal lactate content.

A clean triplet and quartet of the same form, displaced slightly to low field as expected at a lower pH, were present in two of the three plant-based analogues at about one fifth of the Emmental level. No fermentation is involved in such a product, and the signal is attributed to an added propionate preservative or cultured ingredient. In several lipolyzed dairy products the fitting procedure returned apparent propionate values of 20 to 40, but inspection of the spectra showed that the lines concerned belong to the α-methylene triplet of free butyrate at δ 2.155, which overlaps the propionate quartet, and to the valine doublet next to the propionate triplet. Propionate is therefore reported as present only where both multiplets are resolved, that is, in Emmental and in the two plant-based products; elsewhere the index is an upper limit.

Free butyrate was measured from its β-methylene sextet at δ 1.56. It was highest in four of the lipolyzed retail products (30–53) and in Parmigiano Reggiano (34), intermediate in the other ripened cheeses (15–22), low in the category C products (8) and nearly absent from the processed cheeses and declared cheese products (1–2). The relation to the lipolysis measured in the chloroform extract of the same product is positive but loose (Figure 19b), as it should be: the short-chain acids released by lipases partition into the aqueous extract, while the long-chain acids remain in the lipid extract, so that the two extracts record different parts of one process. Acetate was unusually high (93) with almost no lactate (45) in the fresh mozzarella, the pattern expected of direct acidification with vinegar instead of fermentation, and high in the fresh Asiago (81).

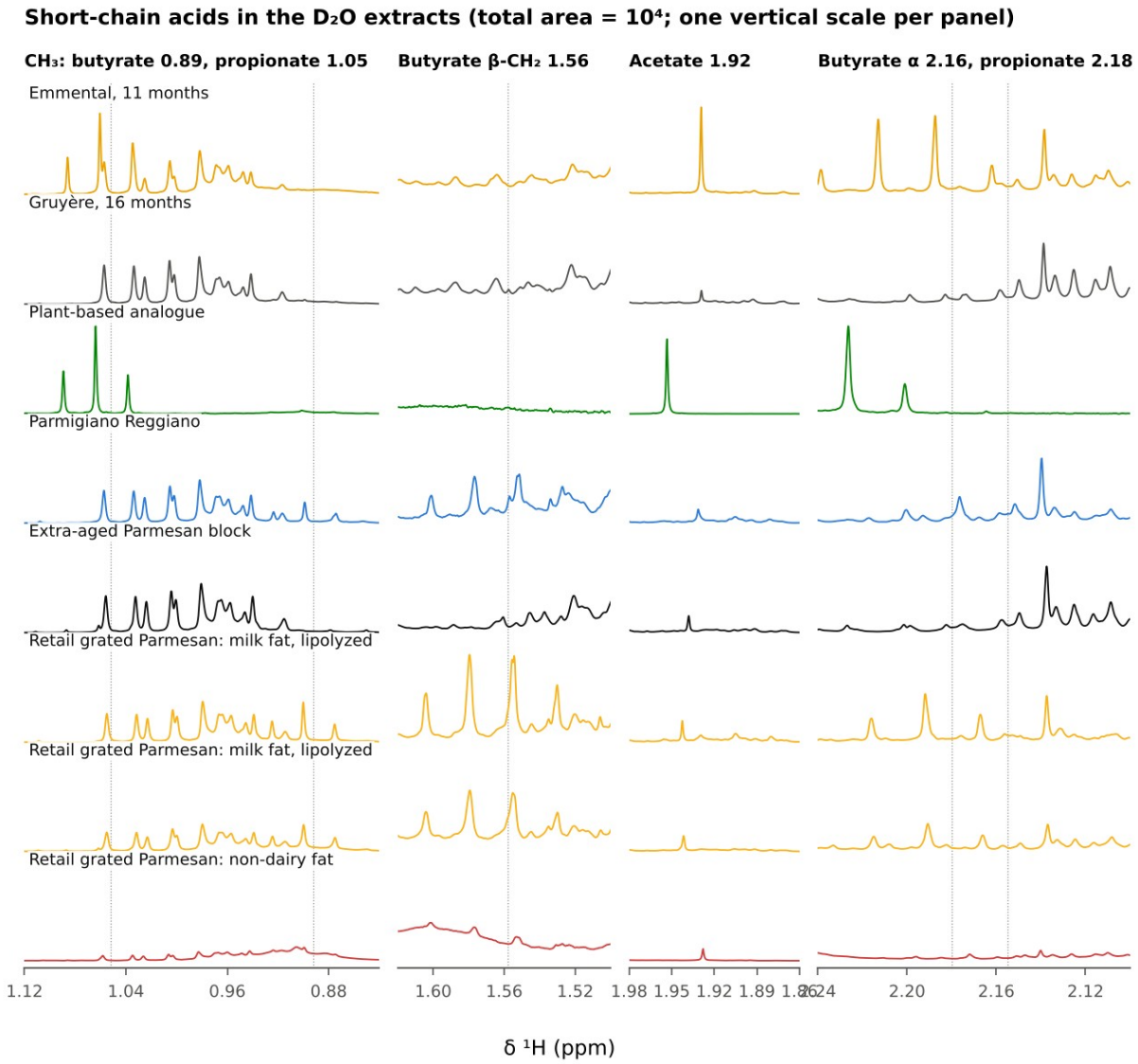

Figure 18. Short-chain acids in the D_2O extracts of selected products (total area = 10^4; each panel has one vertical scale set by its tallest signal, and no signal is truncated). Dotted lines mark the butyrate methyl triplet (δ 0.89), the propionate methyl triplet (δ 1.05), the butyrate β-methylene (δ 1.56), the butyrate α-methylene (δ 2.155) and the propionate quartet (δ 2.18).

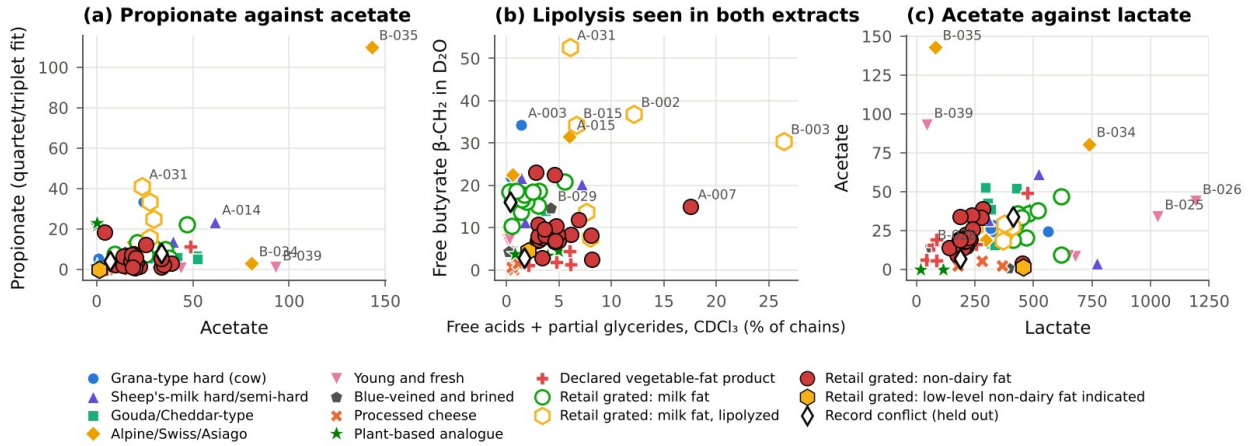

Figure 19. (a) Propionate (quartet/triplet fit) against acetate. (b) Free butyrate in the aqueous extract against free acids and partial glycerides in the chloroform extract of the same product. (c) Acetate against lactate.

6.6 Soluble calcium, magnesium, phosphorus and sodium

The addition of cesium EDTA converted the calcium and magnesium of the aqueous extract into complexes with sharp, separate ethylene singlets (Figure 20a). Where free ligand remained after the addition the bound fractions could be read directly. Where the ligand was exhausted the result was discarded, for the reason set out in Section 3.8: the seven products that were titrated a second time gave calcium indices at 100 µL that were 19 to 39% above those at 200 µL (Table 16 and Figure 21), so that an exhausted titration is neither a measurement nor a lower bound. Thirty of the 36 titrated products gave a reportable result. The six that did not include three of the five category A retail products that were titrated (B-009, B-017 and B-021), the product B-018, and two reference cheeses (Mimolette and fresh Asiago).

On the provisional scale of Section 3.8, the ripened reference cheeses released 240 to 600 mg of calcium per 100 g into the aqueous extract: 516 and 553 for the two Grana-type cheeses, 489 and 519 for Pecorino Romano and the sheep's-milk table cheese, 477 to 596 for the Dutch- and Cheddar-type cheeses, and 242 and 320 for Gruyère and Emmental (Table 17 and Figure 20b). These amounts are of the order of one third to two thirds of the total calcium of such cheeses, as expected at a pH near 5.4, at which a large part of the colloidal calcium phosphate has been solubilized[99,170,171]. The fresh mozzarella (70), the processed single (116), the buffalo-milk blue cheese (24) and the product with conflicting records whose aqueous extract is that of a processed spread (55) lay far below. The magnesium-to-calcium molar ratio was 0.13 to 0.25 in the natural cheeses and rose to 0.50 in the processed single and 0.70 in the buffalo-milk blue cheese (Figure 20c).

For the retail grated products the mineral evidence is thinner than version 2.0 of this report represented. Only second-campaign products were titrated. Reportable results exist for two of the eleven products of category A (336 and 355 mg per 100 g; five were titrated), for two of the five of category B (217 and 367; two were titrated) and for six of the 21 of category C (73 to 98, median 82; six were titrated). Within that limited coverage the six category C products released about one quarter of the calcium of the four products with milk fat, and their magnesium was lower in proportion, 4 to 9 against 20 to 39 mg per 100 g; these ratios do not depend on the provisional scale. The archived ^{31}P measurements show the opposite movement for phosphorus: 50 to 68 mg of soluble phosphorus per 100 g in the five category A products and 103 to 159 in the six of category C, with 215 and 303 in the two processed cheeses (Figure 22a). Low soluble calcium with raised soluble phosphate is what the use of phosphate emulsifying salts would produce, since these salts add soluble phosphate while converting calcium into forms that are not extracted[12,100]; dilution of the cheese by

material that contains little calcium would lower the calcium but would not raise the phosphorus. Sodium, at 1270 to 1720 mg per 100 g in category C and 730 to 1550 in category A, did not discriminate.

These mineral results rest on the ten retail products of categories A to C that have a reportable titration, on an absolute scale that awaits confirmation and, for sodium and phosphorus, on archived worksheet values that could not be re-examined. They are offered as corroboration of limited weight. Within the products for which they exist they agree with the organic markers: each of the six category C products had low soluble calcium and raised soluble phosphorus, and each of the four products of categories A and B had the reverse (Figure 22b). Fifteen of the 21 category C products were not titrated, and nothing can be said about their minerals.

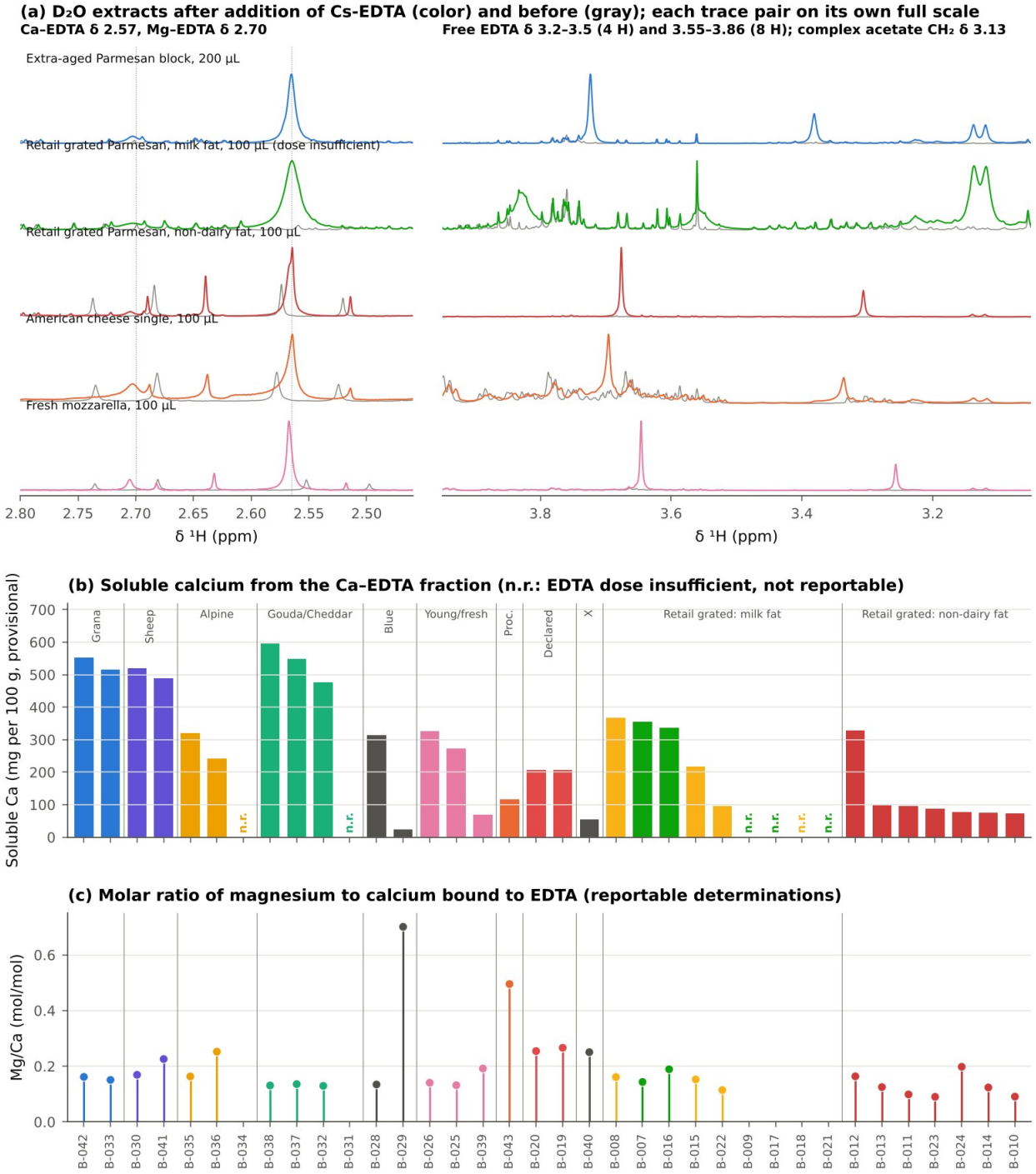

Figure 20. Soluble calcium and magnesium from the Cs-EDTA titration. (a) Spectra of five extracts after the addition of Cs-EDTA (color) and before (gray); each pair of traces is drawn on its own full vertical scale. In the second pair the ligand is exhausted: no free-ligand singlets are present and the Ca–EDTA line is broadened. (b) Soluble calcium (provisional mg per 100 g) for the 36 titrated products, grouped by class; n.r. marks the six products whose titration was not reportable. (c) Molar ratio of magnesium to calcium bound to EDTA, reportable determinations only.

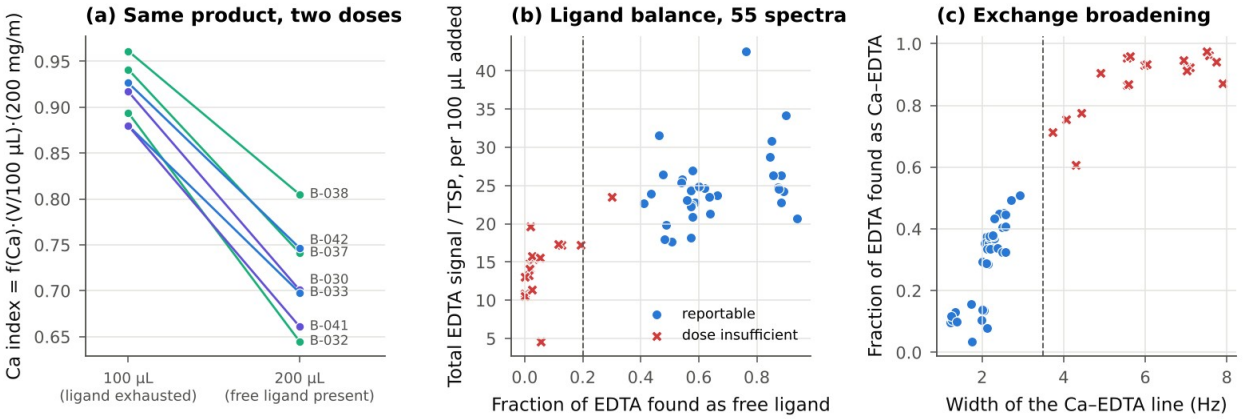

Figure 21. Dose check of the EDTA titration. (a) Calcium index of the seven products titrated at two doses: the 100 µL value, obtained with the ligand exhausted, exceeds the 200 µL value in every case. (b) Total EDTA signal relative to TSP-d_4, per 100 µL added, against the fraction of the ligand found free, for all 55 spectra; the dashed line is the acceptance limit of 20%. (c) Fraction of the ligand found as Ca–EDTA against the width of the Ca–EDTA line; the dashed line is the acceptance limit of 3.5 Hz.

Table 16. Products titrated at two doses of Cs-EDTA: fraction of the ligand found free, width of the Ca–EDTA line, calcium index and total EDTA signal relative to TSP-d_4 (per 100 µL added) at each dose, and the excess of the 100 µL index over the 200 µL index. Duplicate stored spectra were averaged.

Code	Type	Free (100 µL)	Width, Hz (100 µL)	Ca index (100 µL)	EDTA/ TSP (100 µL)	Free (200 µL)	Width, Hz (200 µL)	Ca index (200 µL)	EDTA/ TSP (200 µL)	Excess (%)
B-030	Sheep's-milk table cheese	0.00	7.1	0.917	10.7	0.59	2.1	0.701	22.8	31
B-032	Irish Cheddar-type (sweet, aged)	0.02	6.0	0.894	14.1	0.62	2.2	0.645	24.7	39
B-033	Piave Vecchio	0.02	7.0	0.880	15.1	0.58	2.3	0.698	25.6	26
B-037	Gouda, extra aged	0.02	5.6	0.941	14.2	0.57	2.2	0.741	20.2	27
B-038	Gouda, aged	0.01	7.5	0.961	13.2	0.54	2.5	0.805	25.6	19
B-041	Pecorino Romano	0.04	5.6	0.880	15.7	0.60	2.5	0.661	24.8	33
B-042	Domestic extra-aged Parmesan (Gra…	0.02	7.7	0.927	11.4	0.56	2.3	0.747	23.0	24

Table 17. Soluble minerals in the aqueous extracts of the second-campaign products: number titrated with Cs-EDTA and number with a reportable result; calcium and magnesium on the provisional external-standard scale (mg per 100 g; both values where two are available, otherwise median and range), the Mg/Ca molar ratio, and sodium and soluble phosphorus from the archived ^{23}Na and ^{31}P worksheets (mg per 100 g).

Group	Titrated	Reportable	Soluble Ca	Soluble Mg	Mg/Ca (mol/mol)	n (Na, P)	Na	Soluble P
Grana-type hard (cow)	2	2	516; 553	47; 54	0.15; 0.16	2	577; 772	61; 83
Sheep's-milk hard/semi-hard	2	2	489; 519	53; 67	0.17; 0.23	2	465; 1497	58; 70
Alpine, Swiss, Asiago	3	2	242; 320	32; 37	0.16; 0.25	3	595 (206–674)	48 (43–73)
Gouda/Cheddar-type	4	3	548 (477–596)	45 (37–47)	0.13 (0.13–0.14)	4	769 (523–1123)	124 (57–137)
Blue-veined and brined	2	2	24; 313	10; 26	0.13; 0.70	2	654; 898	163; 189
Young and fresh	3	3	272 (70–326)	22 (8–28)	0.14 (0.13–0.19)	3	532 (307–562)	24 (22–33)
Processed cheese	1	1	116	35	0.50	2	976; 1099	215; 303
Declared vegetable-fat product	2	2	206; 208	32; 33	0.25; 0.27	2	1487; 1519	146; 177
Retail grated, A: milk fat, ripened	5	2	336; 355	31; 39	0.14; 0.19	5	1192 (731–1548)	58 (50–68)
Retail grated, B: lipolyzed, ripened	2	2	217; 367	20; 36	0.15; 0.16	2	758; 1060	40; 62
Retail grated, C: non-dairy fat, processed-type base	6	6	82 (73–98)	6 (4–9)	0.11 (0.09–0.20)	6	1364 (1265–1722)	117 (103–159)
Retail grated, D (B-012)	1	1	329	33	0.16	1	1037	59
Retail grated, E (B-022)	1	1	96	7	0.11	1	1553	129
Retail grated, F (B-018)	1	0	–	–	–	1	1607	111
Record conflict (B-040)	1	1	55	8	0.25	1	640	224

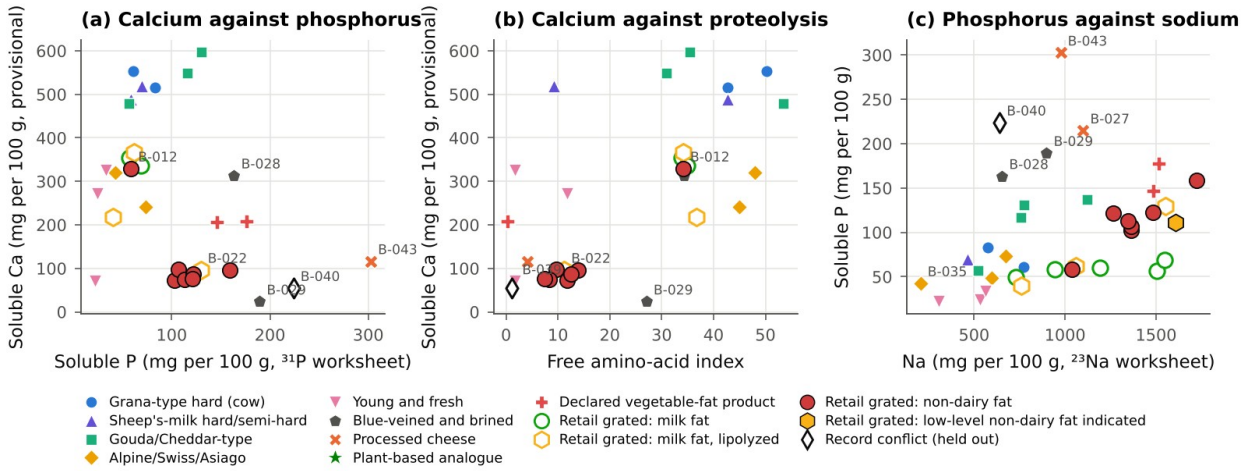

Figure 22. Soluble minerals. (a) Soluble calcium against soluble phosphorus. (b) Soluble calcium against the free amino-acid index. (c) Soluble phosphorus against sodium.

6.7 Ripening time: what the aqueous spectrum can and cannot resolve

The question was raised whether the aqueous spectrum can indicate the age of a cheese, and in particular whether ripening times of 6, 12, 18 and 24 months can be distinguished. The products that carried a numerical age statement were examined for this purpose (Figure 23). The free amino-acid index rose steeply from the fresh cheeses (2 for fresh mozzarella, 12 for fresh Asiago) through the six-month Manchego (29) to the cheeses of ten to twelve months (40–48), and then did not rise further: the two domestic Parmesans of 18 months gave 46 and 50, Gruyère of 16 months 45, and the Goudas of 18 and 26 months 36 and 31. Histidine

and the other individual amino acids behaved in the same way, and the ratio of the amino-acid index to lactate, which is independent of the normalization, separated the classes of cheese more than the ages.

Three conclusions follow, in decreasing order of confidence. First, for a natural cheese examined on its own, the spectrum distinguishes an unripened or briefly ripened cheese from one ripened for ten months or more without ambiguity: the fresh and young reference cheeses gave indices of 2 to 12 and the cheeses of ten months and more 31 to 50. The three retail products with milk fat that carried a ten-month claim gave indices of 31 to 42. The retail product with non-dairy fat that carried the same claim gave 10, and the three declared cheese products that referred to ten-month aging gave less than 1. For these formulated products the inference from index to age does not hold. The index is a fraction of the spectral area, and a cheese ingredient that had been cured for the full ten months would give a low index once it was diluted with material that contributes area but no amino acids (Section 6.1). A low index in such a product therefore shows that the extract as a whole is not that of a ripened cheese; it does not show that the cheese ingredient failed to meet the curing time of the standard of identity.

Second, beyond about ten months the index as defined here does not resolve age. Part of the reason is arithmetical: the index is a fraction of the total spectral area, and in a mature cheese the amino acids themselves make up most of that area, so the fraction saturates. Part is chemical: amino acids are consumed by catabolism as well as released, tyrosine and leucine precipitate, and the rate of proteolysis differs between cheese types by more than it changes with time in a mature cheese. And part is statistical: eleven natural cheeses with a stated age, of nine different types, cannot support a calibration. Third, the literature shows that the problem is soluble under the right conditions. Within a single protected designation, with wheels of documented age, ¹H NMR of extracts or of the intact cheese combined with multivariate regression has resolved ripening stages of Parmigiano Reggiano and Grana Padano at intervals of about a year[48,49,51–53], and pyroglutamate, which accumulates almost linearly in Grana cheeses, has been proposed as a chemical clock[85,86]. Pyroglutamate was identified here by gHSQC, but at 300 MHz its γ-methylene multiplet lies on the flank of the glutamate multiplet, and the ratio of the two windows was nearly constant, at 0.2 to 0.35, across cheeses of all ages; no independent pyroglutamate measure could be extracted from the one-dimensional spectra.

A dedicated study of ripening time would therefore require absolute concentrations in place of area fractions, a spectrometer field or a two-dimensional experiment sufficient to resolve pyroglutamate, and a calibration set of one cheese type sampled at known ages. With those in place the distinction between 12, 18 and 24 months is plausible; from the present data it is not demonstrated.

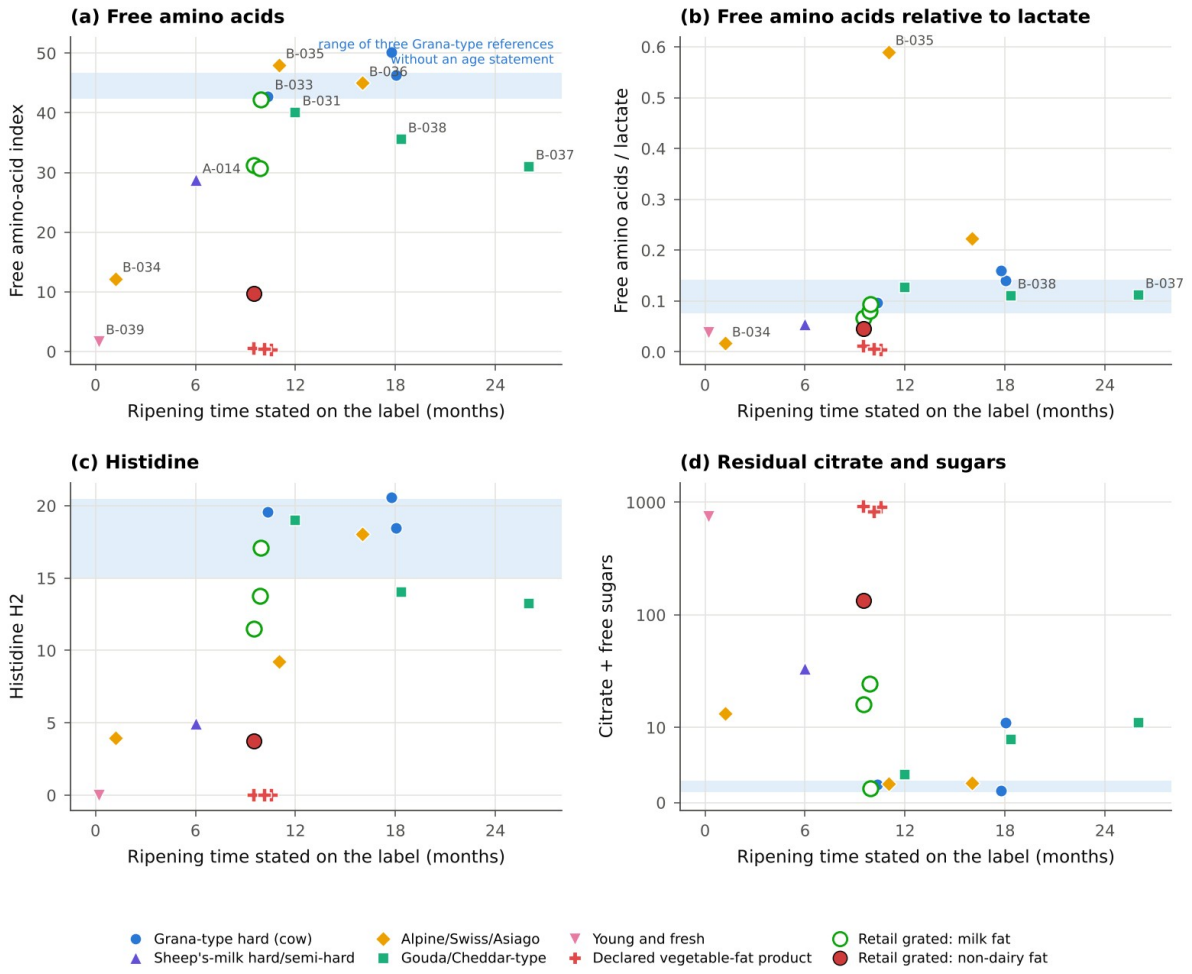

Figure 23. Aqueous ripening markers against the ripening time stated on the label, for the products that carried a numerical age statement. (a) Free amino-acid index. (b) Free amino-acid index relative to lactate. (c) Histidine H2. (d) Citrate plus free sugars. Shaded band: range of the three Grana-type references without an age statement.

7. Chemometrics and Machine Learning

7.1 Representation, alignment, normalization and scaling

The 292 combinations of representation, alignment, normalization and scaling were compared by a common score, the mean of the balanced accuracies of a linear support vector machine and a five-component PLS-DA model in five repetitions of stratified three-fold cross-validation and of a leave-one-out nearest-neighbor classification in a ten-component principal-component space (Figure 24 and Table 23 in Appendix B). The comparison was made on the 42 reference products only, and it serves to rank the options; it is not an estimate of performance, since the ten-component space of the nearest-neighbor score and the alignment and bucket borders were derived from all of the spectra concerned.

For the lipid block and the dairy against non-dairy task, every binned representation gave a score between 0.88 and 0.94. Bin width had almost no effect, the mean score falling only from 0.920 at 0.002 ppm to 0.911 at 0.04 ppm, and adaptive and optimized bucketing (0.906 and 0.923) were equivalent to uniform binning. Alignment by icoshift made no difference (0.913 against 0.918), as would be expected of chloroform spectra whose shifts are stable. Pareto scaling gave the highest mean score (0.927), followed by unit-

variance scaling (0.919), the logarithmic transformation (0.911) and mean-centering alone (0.899), and normalization to region H, to total area or by probabilistic quotient were indistinguishable. The twelve integrals A to J with the two splits and the 18 ratiometric descriptors scored 0.90, within 0.02 of the best binned representation that has twenty times as many variables.

For the aqueous block and the technology task the pattern was similar, with two differences. Unit-variance scaling was the poorest choice (0.792), since it raises the baseline noise of the many empty bins to the same weight as the signals, and Pareto scaling was again the best for binned data (0.828). Alignment lowered the score slightly, from 0.823 to 0.800. Two causes are likely: the apparent pH of the processed-type extracts was higher than that of the ripened extracts by about 0.3 units, so that the positions of the pH-sensitive signals themselves carry class information that alignment removes, and a segment-wise alignment cannot be applied to signals that are absent from the target. Alignment remained necessary for the interpretation of loadings and for the comparison of individual multiplets between spectra, for which it was used throughout Section 6, but it did not help classification. The 24 targeted metabolite features with a logarithmic transformation and Pareto scaling gave the best aqueous score of all (0.876).

These differences should be read with their uncertainty in mind. The standard deviation of a single cross-validated score in this data set is 0.08 to 0.14, so that most of the differences between preprocessing options are not statistically resolved. What the comparison supports is limited: mean-centering with Pareto scaling was not worse than any alternative for binned spectra of either extract, unit-variance scaling was the least suitable for sparse aqueous bins, any bin width between 0.002 and 0.04 ppm served equally, and knowledge-based features lost nothing against the full spectrum. When the scaling was selected inside the training folds (Section 7.4), no option was chosen consistently, which confirms that the data do not resolve a preference. Mean-centered, Pareto-scaled, aligned 0.01 ppm bins were adopted for the remaining multivariate work, with region H = 100 for the lipid block and total area for the aqueous block, in keeping with common practice for NMR fingerprints[123,124]. This choice was made on the 42 reference products as a whole and therefore outside the cross-validation of the models that follow.

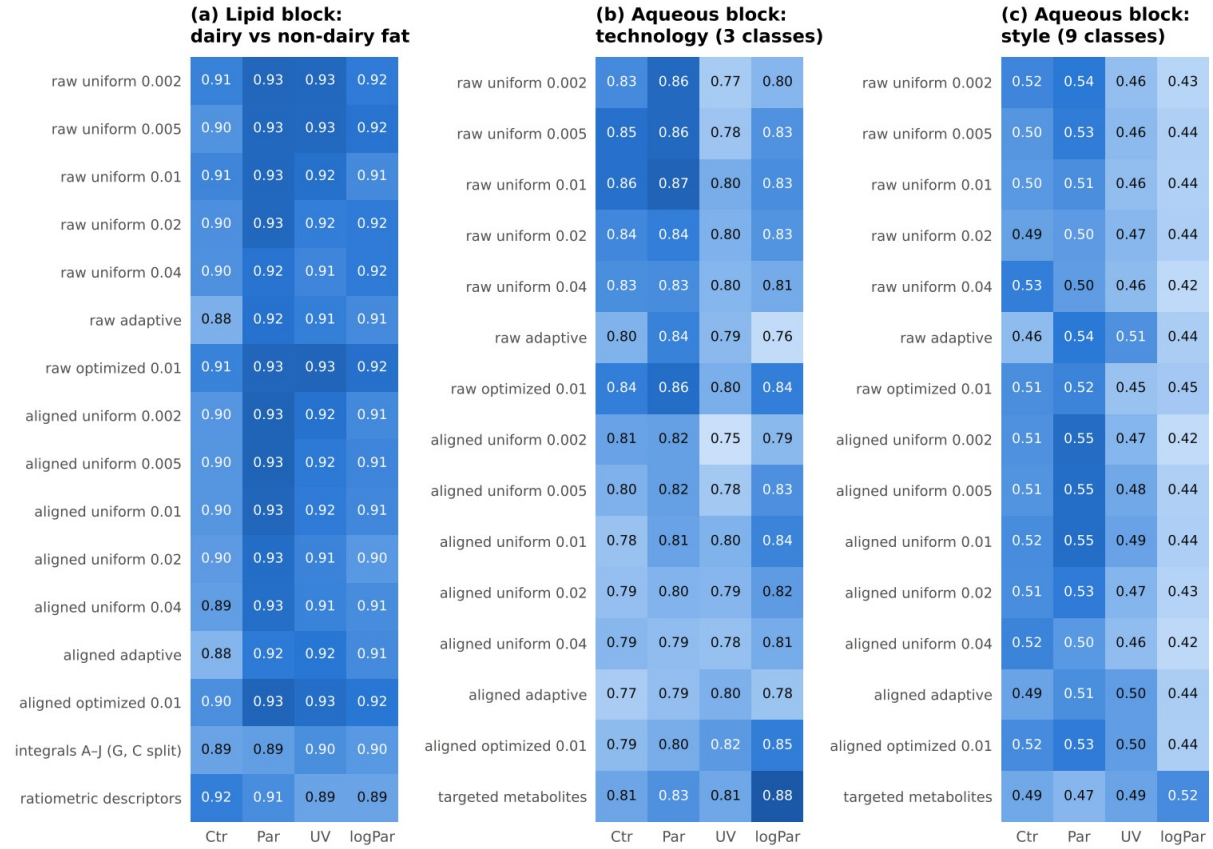

Figure 24. Mean cross-validated score for each representation (rows) and scaling (columns: Ctr, mean-centering only; Par, Pareto; UV, unit variance; logPar, logarithm then Pareto), averaged over normalizations. (a) Lipid block, dairy against non-dairy fat. (b) Aqueous block, technology in three classes. (c) Aqueous block, style in nine classes. "Raw" denotes spectra without alignment.

7.2 Principal component analysis

In the principal component analysis of the lipid block (Figure 25) the first component, with 45% of the variance, ordered the products by their content of non-dairy fat. The 22 retail products with non-dairy fat formed a tight group at positive scores between the milk-fat products and the declared cheese products, and the loadings showed the chemistry directly: positive weights on the methylene envelope, the allylic, bis-allylic and olefinic signals, and a negative weight on the butyrate methyl triplet. The second component (21%) was governed by unsaturation in one direction, where the canola-type loaf stood alone, and by lipolysis in the other, where the strongly lipolyzed grated Romano and the buffalo-milk blue cheese were found, with the free-acid α-methylene line prominent in the loadings. The retail products with milk fat coincided with the reference cheeses, and the plant-based analogues lay at the negative extreme of the first component, on the side of short, saturated chains.

In the aqueous block (Figure 26) the first component (39%) opposed the amino-acid signals of the ripened cheeses, with negative loadings for the branched-chain methyls, lysine and glutamate, to the sugar and anomeric signals of the fresh, processed and declared products. The second component (18%) carried lactate and the broad lipid methylene resonance with positive loadings. The category C retail products formed a compact cluster at positive scores on the second component, together with the processed cheeses and clearly apart from the ripened cheeses and from the retail products with milk fat, which overlapped one

another. The anomalous products took the positions that their targeted chemistry predicts: B-022 lay inside the category C cluster, B-012 among the ripened cheeses, and B-018 with the sugar-rich products.

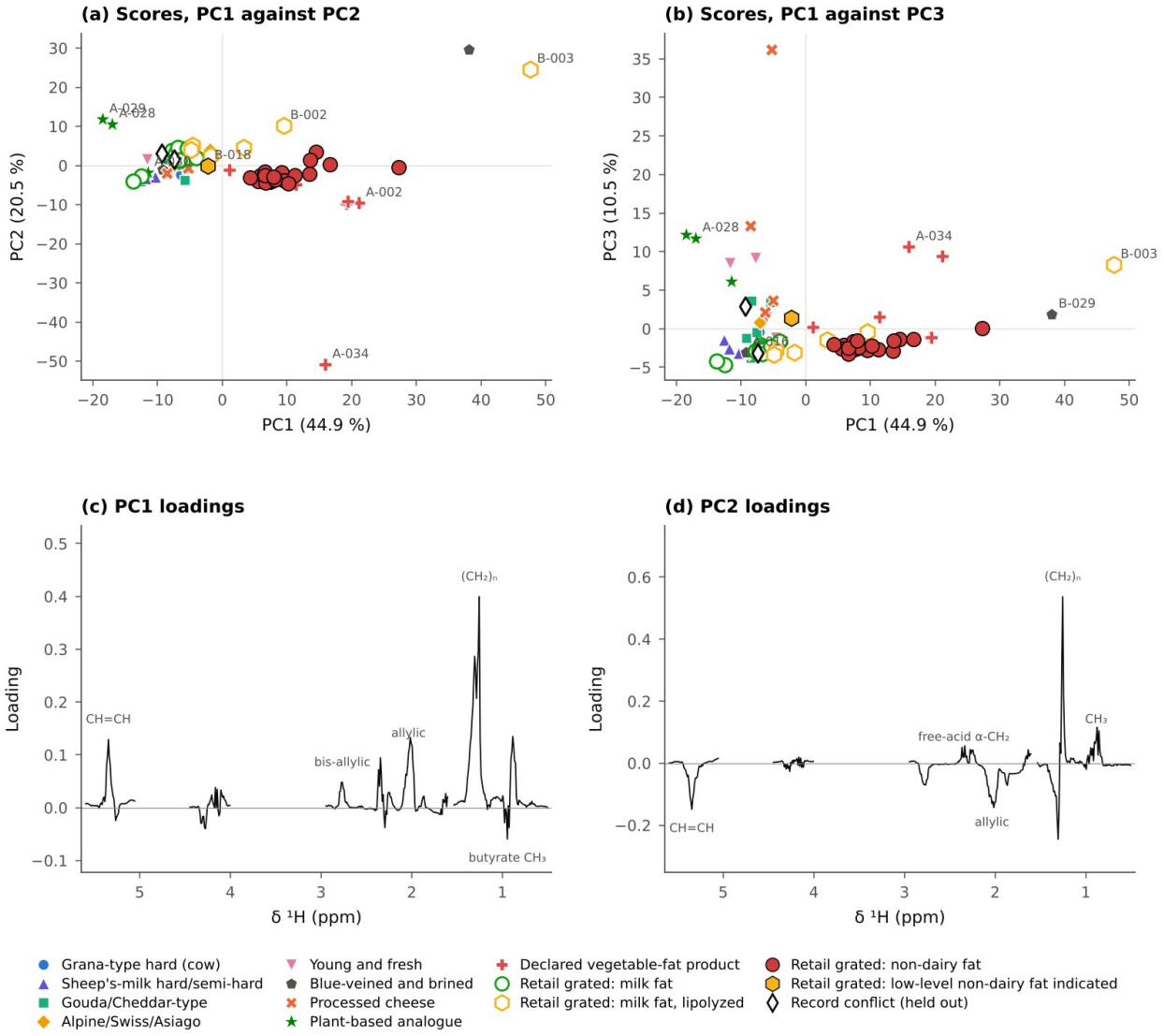

Figure 25. PCA of the $CDCl_3$ spectra (aligned 0.01 ppm bins, region H = 100, mean-centered, Pareto-scaled; 84 products). (a), (b) Score plots. (c), (d) Loadings of the first two components.

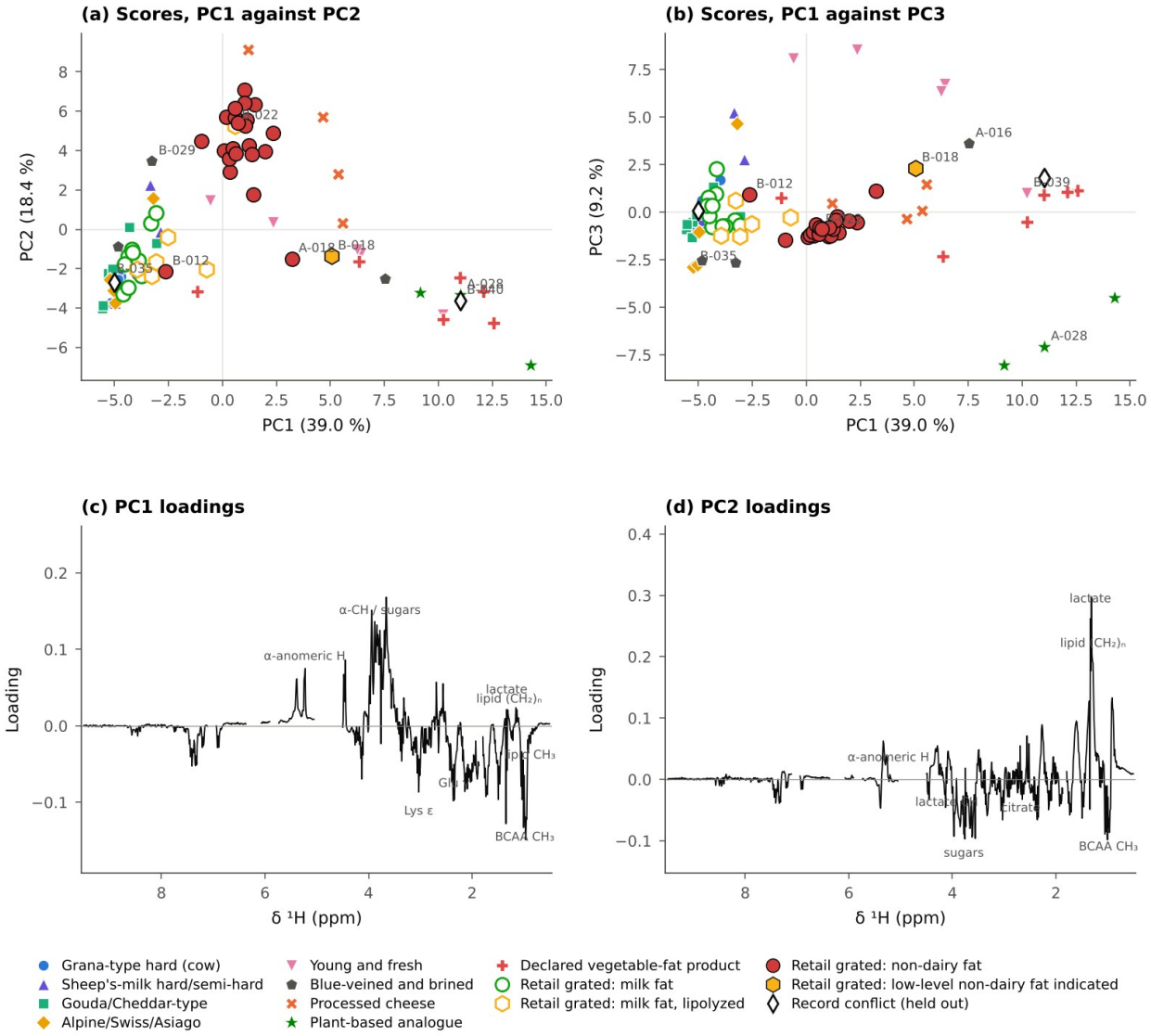

Figure 26. PCA of the D_2O spectra (aligned 0.01 ppm bins, total area, mean-centered, Pareto-scaled; 84 products; water and sorbate windows excluded). (a), (b) Score plots. (c), (d) Loadings of the first two components.

7.3 Hierarchical clustering and nonlinear embeddings

Ward clustering of the lipid block, without any use of the class labels, placed all 22 retail products with non-dairy fat in one branch together with four of the declared vegetable-fat products; the canola-type loaf, the two most lipolyzed products and the plant-based analogues formed small branches of their own, and every other dairy product fell into a single large branch within which neither cheese type nor origin was resolved (Figure 27a). The grated blend in which 26% of non-milk fat was found remained in the dairy branch, which is consistent with the flagging rates of the simulated blends reported below. In the aqueous block the corresponding branch contained 20 of the 22 non-dairy fat products together with the four processed cheeses, the buffalo-milk blue cheese and B-022, while B-012 joined the ripened cheeses and A-018 the young and fresh cheeses (Figure 27b). The fused block recovered a branch with all 22 non-dairy fat products, joined by two declared products and two lipolyzed retail products, separate branches for the plant-based analogues and the declared cheese products, a branch that united the processed with the young and

fresh cheeses, and a ripened branch that contained every Grana-type, sheep's-milk, alpine and Dutch-type reference together with 14 of the retail products with milk fat (Figure 27c).

The t-SNE and UMAP embeddings showed the same structure (Figure 28). In the lipid block the non-dairy fat products formed an island that was joined only by declared vegetable-fat products; in the aqueous block they formed an island with the processed cheeses; and in the fused block the island was theirs alone. Such embeddings preserve neighborhoods and not distances, their appearance depends on the perplexity or neighbor number and on the random seed, and with 84 points they can do no more than illustrate a structure that has been established by other means[130–132]. They are shown here for that purpose only.

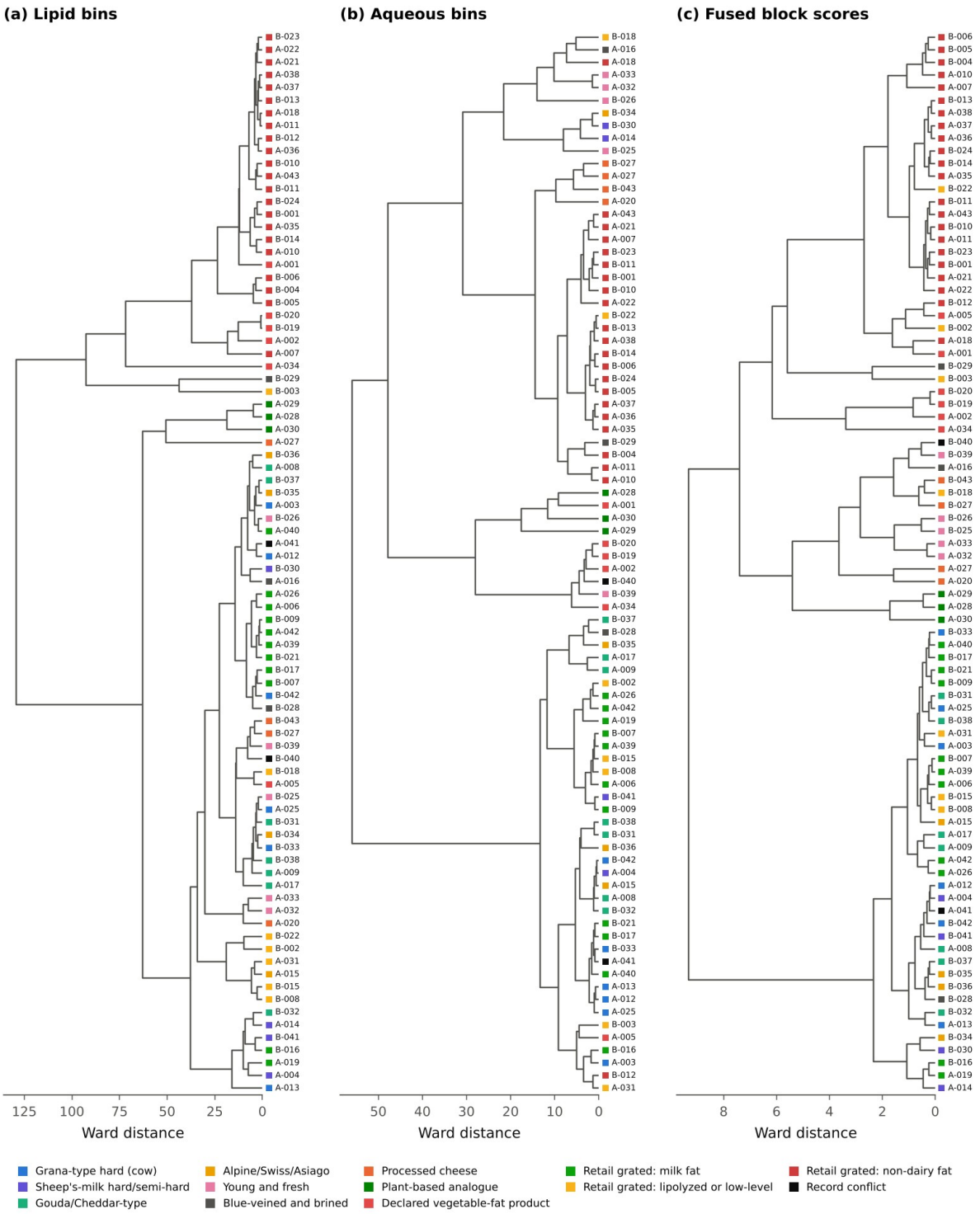

Figure 27. Ward hierarchical clustering of the 84 products on the first eight principal components of (a) the lipid bins, (b) the aqueous bins and (c) the fused block scores. Squares give the class or, for retail grated products, the lipid category.

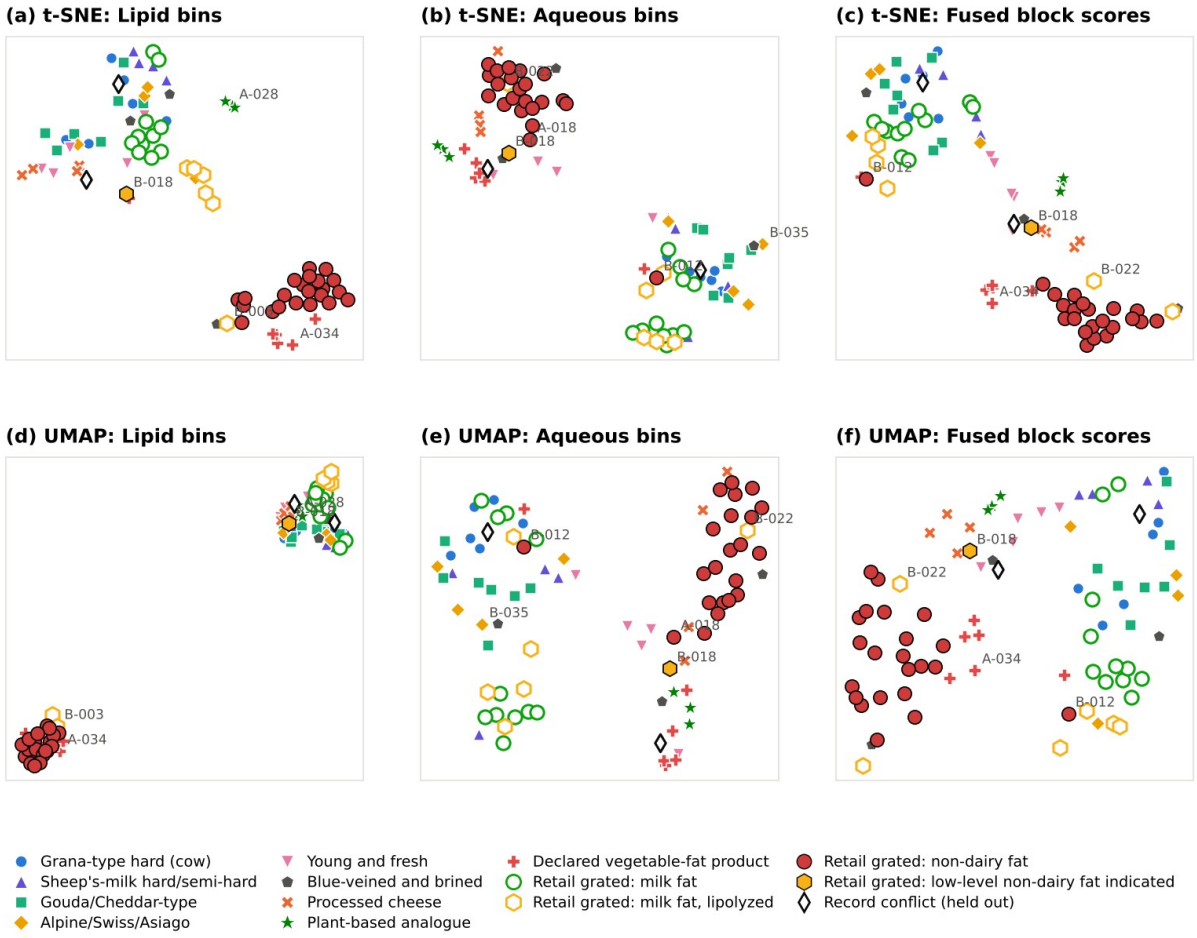

Figure 28. t-SNE (perplexity 12) and UMAP (12 neighbors, minimum distance 0.3) embeddings of the lipid, aqueous and fused data.

7.4 Supervised classification

The cross-validation results of the first evaluation, in which the preprocessing of each block had been fixed beforehand, are given in Table 24 (Appendix B) and Figure 29a–c. For the fat task the seven classifiers gave balanced accuracies of 0.87 to 0.92 on the lipid bins and on the lipid descriptors alike, against a permutation mean of 0.50 and a 95th percentile of 0.62. No model was distinguishably better than another. The accuracy is below what the clean separation in the score plots would suggest, and the reason lies in the reference set and not in the spectra: the non-dairy class has only nine members, which include a lauric fat, a canola-type oil, palm-type fats and a blend with only 26% of non-milk fat, and in a three-fold split each model was trained on six of them. For the technology task the targeted aqueous features gave 0.84 to 0.94 and the aqueous bins 0.74 to 0.88, against a permutation mean of 0.33. For the nine style classes the best results were 0.58 for the aqueous and fused blocks and 0.45 for the lipid block, against 0.11 by chance.

These figures are optimistic to the extent that the preprocessing, and in each row of the table the best of seven classifiers, were chosen with sight of the same 42 products. The fully nested evaluation of Section 3.10, in which the representation, the normalization, the scaling, the classifier and its hyperparameters were all selected inside each outer training fold from data-independent representations, gave the estimates in Table 18: 0.90 ± 0.11 for fat, 0.82 ± 0.11 for technology, 0.51 ± 0.12 for style from the aqueous block and 0.36 ± 0.09 from the lipid block. The scores at which the inner loop made its selections were far higher, 0.98,

0.96, 0.59 and 0.57, and the distance between the two columns is the optimism that a selection reported without an outer loop would have carried[139]. Against the best single entries of the first evaluation the nested estimates are lower by 0.02 for fat, 0.12 for technology and 0.07 for style. The fused block, with its principal components fitted inside the folds, gave 0.87, 0.85 and 0.51. The selections themselves were unstable: in 30 outer folds of the fat task the 0.01 ppm bins were chosen 25 times but the scaling and the classifier changed from fold to fold, many candidates being tied at the top of the inner ranking, so that the frequencies in the table should not be read as a ranking of methods.

Three conclusions survive this stricter test. Dairy and non-dairy fat are separated with a balanced accuracy of about 0.9, limited by the small and heterogeneous non-dairy class. The technology of the cheese base is recognized from the aqueous extract with about 0.8. Style in nine classes is predicted far better than chance but not reliably from three to seven examples per class, and the lipid extract contributes little to it: once species and feeding are set aside, milk fat is milk fat, and the identity of a cheese is written in its aqueous phase. All of these figures describe the 42 reference products, whose classes were assigned from their labels.

Table 18. Fully nested cross-validation on the 42 reference products: balanced accuracy (mean ± s.d. over 30 outer folds) when representation, normalization, scaling, classifier and hyperparameters were all selected inside each outer training fold; mean inner-loop score of the selected configuration; best single entry of the first evaluation (Table 24) for comparison; and the three most frequent selections (number of folds). Candidate representations: unaligned uniform bins of 0.01, 0.02 and 0.04 ppm, integrals A to J, ratiometric descriptors (lipid) and targeted features (aqueous).

Task	Block	Outer balanced accuracy	Inner score at selection	First evaluation, best entry	Representation selected	Scaling selected	Classifier selected
Dairy vs non-dairy fat	lipid	0.90 ± 0.11	0.98	0.92 (kNN, lipid bins)	bins, 0.01 ppm (25); descriptors (3); bins, 0.02 ppm (1)	centering (13); unit variance (7); log-Pareto (5)	PLS-DA (7); SVM-linear (7); Logistic (L2) (6)
Technology (3 classes)	aqueous	0.82 ± 0.11	0.96	0.94 (Logistic (L2), aqueous targeted)	bins, 0.01 ppm (18); targeted features (9); bins, 0.04 ppm (3)	centering (13); log-Pareto (9); unit variance (6)	PLS-DA (8); Logistic (L2) (7); SVM-linear (6)
Style (9 classes)	aqueous	0.51 ± 0.12	0.59	0.58 (kNN, aqueous bins)	bins, 0.01 ppm (11); targeted features (10); bins, 0.04 ppm (7)	centering (16); unit variance (6); Pareto (6)	Logistic (L2) (9); SVM-linear (9); Random forest (4)
Style (9 classes)	lipid	0.36 ± 0.09	0.57	0.45 (Random forest, lipid bins)	bins, 0.01 ppm (10); descriptors (9); bins, 0.02 ppm (4)	centering (10); Pareto (8); log-Pareto (7)	Logistic (L2) (14); SVM-RBF (4); PLS-DA (4)
Dairy vs non-dairy fat	fused, PCA inside folds	0.87 ± 0.14	–	0.89 (PLS-DA)	0.01 ppm bins, both blocks	Pareto	SVM-linear (16); PLS-DA (7); Random forest (2)
Technology (3 classes)	fused, PCA inside folds	0.85 ± 0.12	–	0.86 (kNN)	0.01 ppm bins, both blocks	Pareto	SVM-linear (8); SVM-RBF (7); Random forest (5)
Style (9 classes)	fused, PCA inside folds	0.51 ± 0.10	–	0.58 (Random forest)	0.01 ppm bins, both blocks	Pareto	Logistic (L2) (11); PLS-DA (9); SVM-linear (6)

When the models of the first evaluation were refitted on all reference products and applied to the retail grated products, which had not been used in training, the fat models reproduced the ratiometric categories closely (Figure 29d and e). A support vector machine on the 18 descriptors called all 22 non-dairy fat products and B-018 non-dairy and all milk-fat and lipolyzed products dairy. A random forest on the bins called the 22 correctly but added the most lipolyzed product, B-003. A support vector machine on the bins added B-003, B-002 and B-018, and PLS-DA on the bins missed three of the 22 and added B-003. The errors of the bin-based models were thus concentrated on the lipolyzed products, in which butyrate is reduced for a reason that was not represented among the training examples, and it is this case that the ratiometric rule handles explicitly. The agreement should not be over-read. The retail products have no independent reference values; their categories come from the ratiometric rule, and the descriptor-based model uses the very quantities on which that rule is built. Agreement between the two shows that the rule and the models describe the same structure in the same spectra. It does not show that either is correct.

The technology model on the targeted aqueous features assigned all 16 retail products of categories A and B to the ripened class and 20 of the 21 products of category C to the processed or analogue class, the remaining one, A-018, being called fresh (Figure 29f). It placed B-012 with the ripened cheeses and B-022 with the processed products, in agreement with the explicit markers. It placed B-018 with the ripened cheeses, which the explicit markers contradict: this product has almost no free amino acids and a large lactose signal, a combination that occurs in none of the training classes, and the model, obliged to choose among three classes, chose wrongly. The nearest single reference product in the space of targeted features was a processed cheese for eleven of the category C products and a member of the blue and brined class for eight.

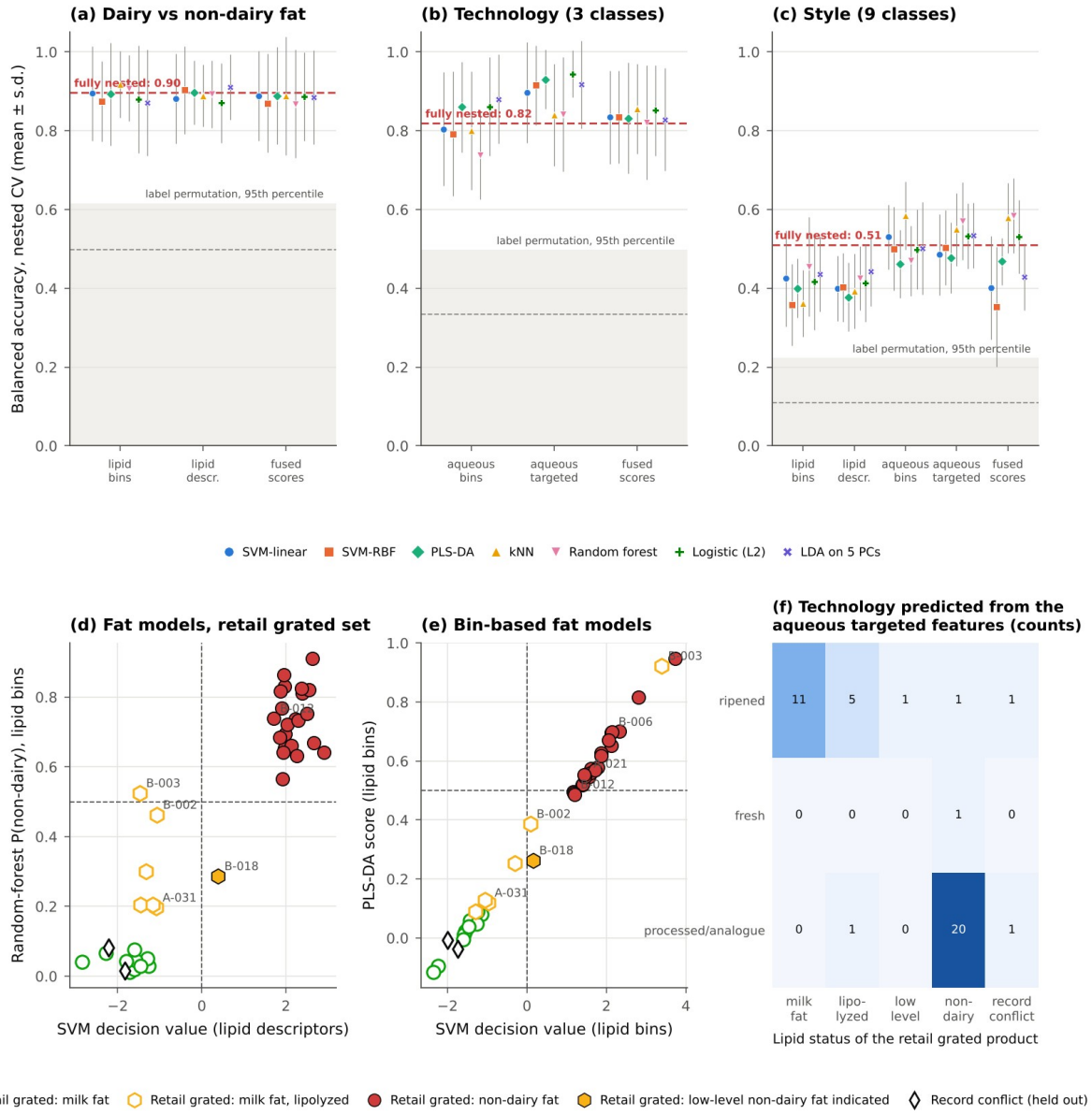

Figure 29. Supervised models. (a)–(c) Cross-validated balanced accuracy (mean ± s.d. over 30 outer folds) of seven classifiers for the three tasks with the preprocessing fixed beforehand; the shaded band extends to the 95th percentile of the label-permutation distribution and the dashed gray line marks its mean; the dashed red line is the fully nested estimate of Table 18. (d) Random-forest probability of non-dairy fat against the decision value of the support vector machine on descriptors, for the retail grated products. (e) PLS-DA score against the decision value of the support vector machine on bins. (f) Technology predicted from the targeted aqueous features, by lipid category (counts).

7.5 One-class models and simulated blends

The one-class models ask a different question: not which of several known classes a product belongs to, but whether it is compatible with the authentic class at all (Table 25 in Appendix B and Figure 30a and b). For the lipid block, a Mahalanobis distance on five descriptors (C4:0, C18:2, NC, C_2/C_1 and G_1/G_2) rejected all 22 retail products with non-dairy fat, all nine non-dairy reference products and three of the lipolyzed products, and rejected three of the 33 authentic products when each was left out in turn. DD-SIMCA on the bins gave the same result for the non-dairy products but rejected five of the 33 authentic products on leave-one-out, against a nominal rate of 1%; the acceptance limit of a model built on 33 spectra with several hundred variables is too tight, and the model reacts to lipolysis as strongly as to substitution. The one-class support vector machine and the isolation forest were less satisfactory still, the first rejecting ten of the 33 authentic reference products and the second accepting eight of the 22 retail products with non-dairy fat.

For the aqueous block, a Mahalanobis distance on eight targeted features, built on the 21 ripened reference cheeses, rejected all 21 category C products, B-022 and B-018, accepted B-012, and rejected one of the 16 products of categories A and B and one authentic cheese on leave-one-out. DD-SIMCA on the aqueous bins rejected ten of those 16, five of them in category A, and six of the 21 authentic cheeses on leave-one-out, and is not usable in this form. The two Mahalanobis models, one for each extract, give together the two-way classification that summarizes the study (Figure 30a): products that conform to both models, products that fail both, and the two products that fail one only. The observed false-rejection rates of the usable models, three of 33 and one of 21, are several times the nominal 1%, and they are themselves leave-one-out figures from the products on which the descriptors were chosen.

The response of the methods to a foreign fat was examined with the simulated blends (Table 19 and Figure 30c and d). With palm oil as the added fat, the decision rule raised no flag in the 24 unblended spectra or at 5%, flagged one blend at 10%, six at 15%, 12 at 20% and 21 at 25%, and flagged all 24 from 30% upward; the definitive call of non-dairy fat was made for none of the blends up to 15%, for six at 20%, for 19 at 25% and for all 24 from 30%. With the fat of the declared cheese products, which is closer to milk fat in linoleate and chain length, the rule flagged one blend at 15%, six at 20%, 12 at 25%, 16 at 30% and all 24 at 40%; the definitive call was made for one blend at 25%, ten at 30%, 23 at 40% and all 24 at 50%. The Mahalanobis model and the support vector machines behaved similarly, the random forest was less sensitive by 5 to 10 percentage points, and DD-SIMCA appeared more sensitive only because it already rejected three of the 24 unblended spectra.

These counts are not a detection limit. The blends are sums of two spectra; they contain neither the extraction of a real mixture, nor a lot of cheese or fat that the rule had not seen, nor the day-to-day variation of the measurement, and the 24 base spectra are the same products that define the reference limits, each left out only when its own blends were judged. What they show is the level at which the arithmetic of the rule responds: for a palm-type fat, a definitive call in every simulated blend required 30 to 50% of non-milk fat, and half of the blends were flagged at about 20 to 25%. The level is governed by the natural variability of butyrate in milk fat, three standard errors of the butyrate estimate for an unadulterated product corresponding to about 19%. The estimates for the 22 flagged retail products, 43 to 68%, lie in the range in which all, or all but one, of the simulated blends were called; that of B-018, 21%, lies where only a minority were. A detection limit and a false-positive rate that could be quoted for real products require gravimetric blends of cheese and fat carried through the whole procedure, with lots of cheese and fat that are independent of the reference population, analyzed under code, together with a set of authentic cheeses that played no part in setting the limits[151].

The same blends allowed the arithmetic of the quantitative estimates to be checked against known values. The butyrate-based estimate reproduced the blended fraction with a root-mean-square error of 4 to 6 percentage points and a bias of −2 to −4, the caproleate-based estimate with an error of 12, and the PLS regression on binned spectra with 4.1 on blends of base products that were not in its calibration. These errors describe simulated blends and say nothing about the error on retail cheese, where extraction, lot variation and lipolysis intervene. Applied to the real products, the PLS model predicted 0.8 ± 4.2% for the authentic references, but it returned 27 to 164% for four lipolyzed milk-fat products, 113% for the lipolyzed product with non-dairy fat (A-007) and negative values for the lauric fats. A regression of this kind is accurate inside its calibration domain and fails silently outside it, and its predictions are reported here only for products inside that domain (Section 3.10).

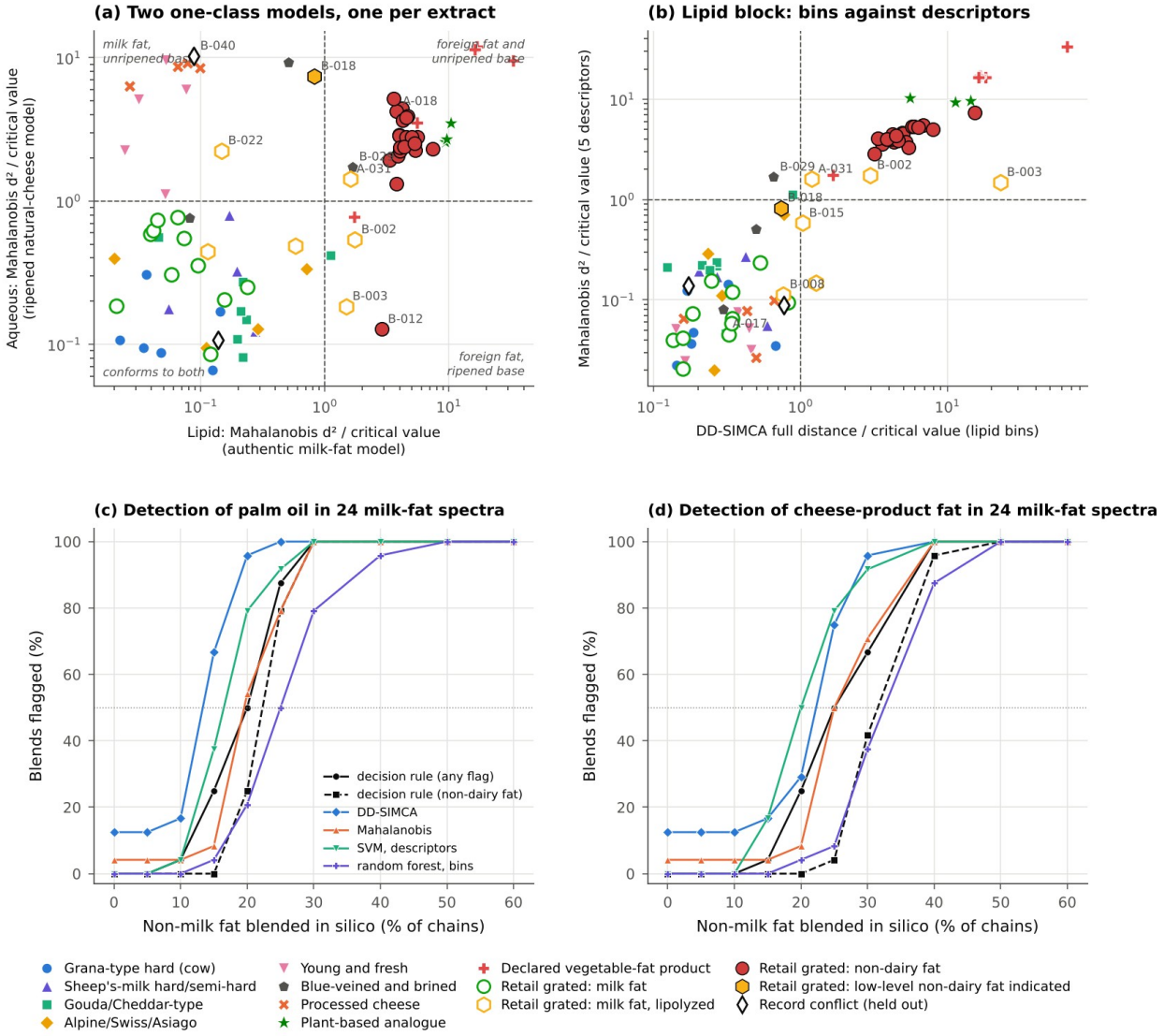

Figure 30. One-class models and simulated blends. (a) Mahalanobis distance of each product from the authentic milk-fat model (lipid descriptors) and from the ripened natural-cheese model (targeted aqueous features), each divided by its critical value; dashed lines mark the acceptance limits. (b) Lipid block: DD-SIMCA on bins against the Mahalanobis model on descriptors. (c), (d) Fraction of in-silico blends flagged by each method as a function of the added non-milk fat, for palm oil and for the fat of the declared cheese products (24 milk-fat base spectra at each level).

Table 19. Simulated blends: number of the 24 blends flagged by each method at each level of added fat (% of acyl chains). Each method was refitted without the product that served as the base of the blend. The blends are linear combinations of spectra and include no extraction step.

Added fat	Method	0%	5%	10%	15%	20%	25%	30%	40%	50%	60%
Palm oil	Decision rule, any flag	0	0	1	6	12	21	24	24	24	24
	Decision rule, "non-dairy fat"	0	0	0	0	6	19	24	24	24	24
	Mahalanobis (5 descriptors)	1	1	1	2	13	19	24	24	24	24
	SVM, descriptors	0	0	1	9	19	22	24	24	24	24
	SVM, bins	0	2	3	8	21	24	24	24	24	24
	Random forest, bins	0	0	0	1	5	12	19	23	24	24
	DD-SIMCA (lipid bins)	3	3	4	16	23	24	24	24	24	24
Fat of the declared cheese products	Decision rule, any flag	0	0	0	1	6	12	16	24	24	24
	Decision rule, "non-dairy fat"	0	0	0	0	0	1	10	23	24	24
	Mahalanobis (5 descriptors)	1	1	1	1	2	12	17	24	24	24
	SVM, descriptors	0	0	0	4	12	19	22	24	24	24
	SVM, bins	0	1	2	3	5	14	23	24	24	24
	Random forest, bins	0	0	0	0	1	2	9	21	24	24
	DD-SIMCA (lipid bins)	3	3	3	4	7	18	23	24	24	24

7.6 Data fusion

The fused block of principal-component scores from both extracts was evaluated for all three tasks. With the block models fitted inside the folds it gave 0.87 for fat, 0.85 for technology and 0.51 for style (Table 18), in no case clearly better than the better of the two single blocks. Fusion at the level of the data did not help because each question is answered by one extract: the lipid spectrum contains nothing about ripening, and the aqueous spectrum contains nothing about the fat beyond the dispersed lipid. In the study of Maestrello and co-workers, by contrast, a model of both fractions was the best performer for the recognition of one designated cheese among competitors[79]; with 144 cheeses and a single two-class question the conditions for fusion were more favorable than with nine classes of three to seven products. In the present data the value of having both extracts lies in the logical combination of two separate answers, as in Figure 30a, and not in a single model of the concatenated data[133].

8. Integrated Interpretation of the Retail Grated Products

Each retail product was assigned to one of six categories according to its lipid category and its aqueous profile (Table 5, Table 20 and Figure 31). The aqueous profile was called unripened or processed-type when the free amino-acid index was below the limit and citrate or free sugars were present; dispersed lipid was recorded as a supporting marker. The evidence from the ratiometric markers, the supervised models and the one-class models is set out side by side in the figure, and the agreement between them is discussed in Section 8.5.

8.1 Products in which no milk-fat marker was detected

No product sold as grated Parmesan or Romano in the paired set had its milk-fat markers at the blank level. The products in that condition were the three plant-based analogues, which make no dairy claim; the

declared dairy-free Parmesan alternative of the lipid-only set (H-020); and the three declared grated cheese products, whose fitted butyrate of 0.8 to 0.9 mol % lies below the floor of the measurement and for which only a lower bound of 86% non-milk fat can be given. These three products are named and presented as Parmesan- and Romano-style products; their aqueous extracts showed lactose, citrate and almost no free amino acids. In these products the cheese named on the label contributes little to either the fat or the water-soluble phase, as their ingredient statements indicate.

8.2 Category A and B: milk fat with a ripened base

Eleven retail products showed milk fat without lipolysis and five showed lipolyzed milk fat, in each case with the aqueous profile of a ripened hard cheese: a free amino-acid index of 28 to 45, no significant citrate or sugar and, in the four for which a reportable titration exists, soluble calcium of 217 to 367 mg per 100 g with low soluble phosphorus. They include the two imported grated Pecorino Romano products, whose conjugated-diene and n-3 contents were those of the Pecorino Romano blocks, the store-grated Parmesan and the product without cellulose. Every lipid model placed the eleven category A products with the authentic cheeses. Among the aqueous models the Mahalanobis model and the technology classifier accepted all eleven, whereas DD-SIMCA on the aqueous bins rejected five of them, as it rejected six of the 21 authentic reference cheeses on leave-one-out; version 2.0 of this report stated in error that every aqueous model had accepted them. The category B products were flagged by some of the lipid models because of their low butyrate, and they are the cases in which caproleate, linoleate and chain length must be consulted before a conclusion is drawn. Lipolysis is a matter of flavor and quality and not of identity, and nothing in either extract of these 16 products is inconsistent with a grated hard cheese made from milk fat. That statement does not extend to cellulose or other insoluble fillers, which the method does not see.

8.3 Category C: non-dairy fat with a processed-type base

Twenty-one retail products, slightly more than half of those examined, combined a fat estimated at 43 to 68% non-milk fat with an aqueous extract that is not that of a ripened hard cheese. Their free amino-acid index was 28% of that of the retail products with milk fat, or 54% of it relative to lactate; citrate was present in all; free galactose and glucose-type sugars were present in all, with a starch-derived glucan signal; fat was dispersed into the aqueous extract in all but one; the apparent pH was higher by 0.3 units; and in the six that were titrated, soluble calcium was about one quarter of that of the retail products with milk fat and soluble phosphorus twice as high. The butyrate, linoleate and chain-length criteria, three of the four fat classifiers, both lipid one-class models, the free amino-acid, citrate and sugar markers and the aqueous one-class model were met by every one of the 21; caproleate fell outside its limit in 20, dispersed lipid exceeded its limit in 20, PLS-DA recognized 19, and C_2/C_1 exceeded its limit in 12.

The question of what the non-cheese component resembles can be answered from within the sample set. Among the reference classes, the aqueous profile of category C is closest to that of the processed American slices and singles: a low free amino-acid index, citrate, residual sugars, fat carried into the aqueous extract and, where measured, low soluble calcium with raised soluble phosphorus are the shared features, and they are the features that emulsifying salts and added dairy solids impart (Figure 10). The profile differs from that of the American slices in two respects. The fat is not milk fat, and the sugar is galactose and glucose with little lactose, where the slices contain mainly lactose. In its fat the non-cheese component resembles the declared grated cheese products (Section 5.6).

One composition that would account for all of the observations is a blend of grated hard cheese with a cheese analogue or cheese product made from vegetable fat, a dairy or casein base and emulsifying salts. On that reading the milk fat, the residual free amino acids, the free butyrate, the succinate and the histidine of

the category C extracts come from the cheese portion, and the replacing fat, the citrate, the sugars, the dispersed lipid and the altered mineral balance from the analogue. This reading is an inference from composition and is offered as plausible, not as established. The data do not establish the manufacturing route, and a single emulsified product formulated with some cheese would fit as well as a blend. Nor do they allow the proportion of cheese in the product to be estimated: the fat content and the soluble solids of the components were not measured, and the free amino-acid index, being a fraction of the spectral area, falls with dilution and with formulation as well as with the amount and age of the cheese. The estimate of that proportion given in version 2.0 is withdrawn.

The group is uniform in the descriptors measured. The 22 flagged products of the paired set and the 19 flagged spectra of the lipid-only set show replacing fats of similar gross composition at similar levels, and 21 of the 22 share one aqueous profile. It should be remembered that the 19 further spectra represent five products of four brands, three of which are also in the paired set, so that the number of distinct products behind this uniformity is smaller than the number of spectra. A similar composition is what a common type of ingredient would produce, but the anonymized records examined here contain nothing on suppliers or manufacture, and no shared source or formulation has been demonstrated.

A product of this composition is not a grated cheese as the standards of identity define Parmesan, Romano and grated cheeses[9–11]. Whether the labeling of any individual product is nonetheless accurate, for example through a product name and ingredient statement that declare a cheese product or imitation cheese, is a question of fact about the label and of law, and lies outside the scope of this report, which did not record the ingredient statements systematically.

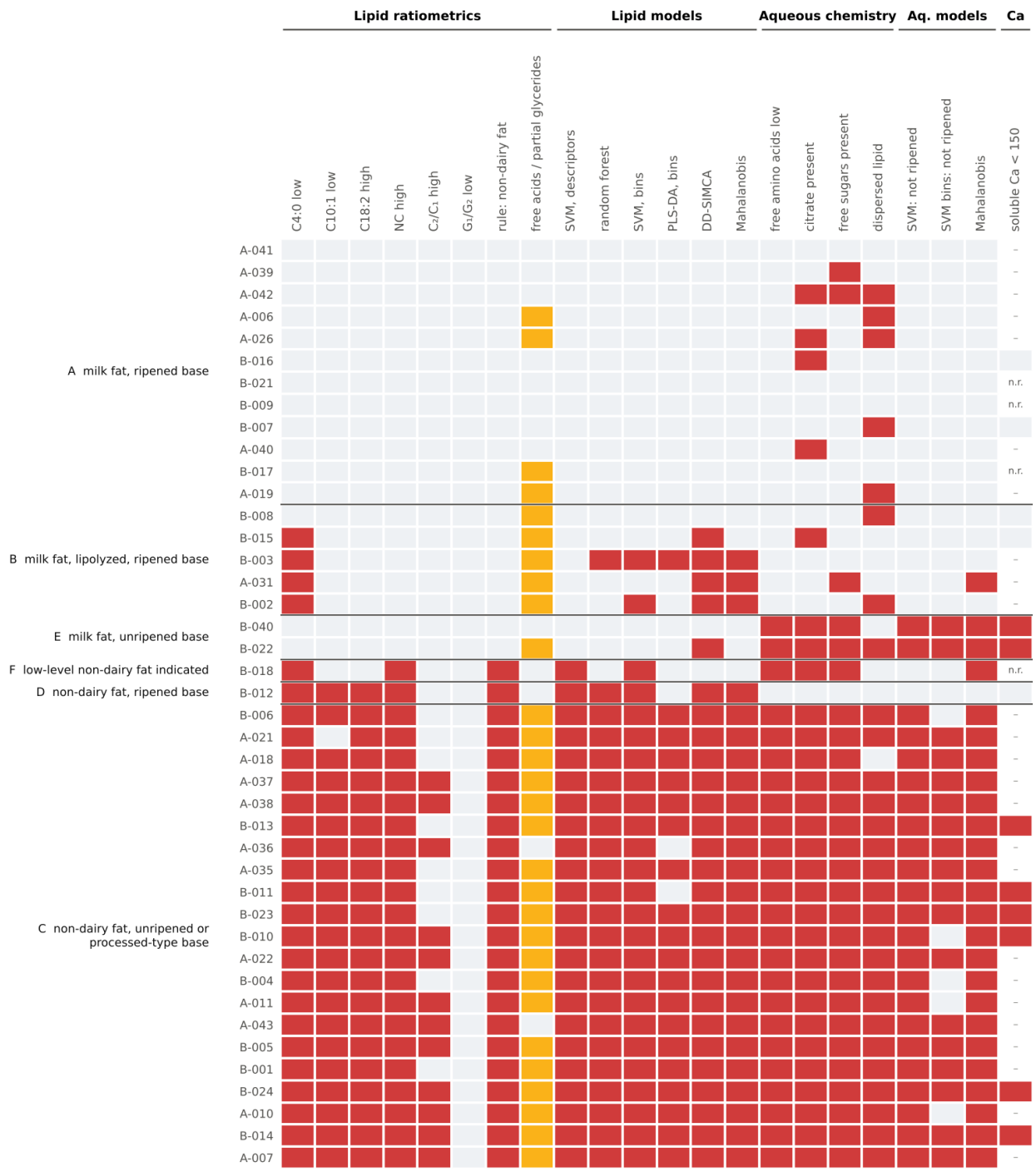

Figure 31. Evidence matrix for the 40 retail grated products and the two products with conflicting records. Red cells: criterion met (ratiometric limits at three standard deviations of the milk-fat reference population; model outputs as non-dairy fat, not ripened, or outside the one-class acceptance limit; aqueous markers beyond the limits of Table 5; soluble calcium below 150 mg per 100 g). Orange cells: free acids and partial glycerides above 3% of chains or more than 3.14 chains per glycerol. In the calcium column a dash marks products that were not titrated and n.r. a titration that was not reportable. Rows are grouped by integrated category.

Table 20. Evidence summary for the 40 retail grated products and the two products with conflicting records (A-041 and B-040). Non-milk fat: consensus estimate in % of acyl chains ± 2 s.e.; n.q., not quantified because butyrate has been lost by lipolysis; †, butyrate estimate only, since the caproleate content of sheep's milk fat is lower than that of the reference population. FAA, citrate, sugars and dispersed lipid: aqueous indices per 10⁴ of total area. Soluble Ca: provisional mg per 100 g (second campaign only; n.r., titrated but not reportable; –, not titrated). Technology: support vector machine on targeted aqueous features. Nearest class: class of the nearest reference product in the targeted aqueous features. Category: A to F as defined in Table 5.

Code	Label	Lipid category	Non-milk fat (%)	FAA	Citrate	Sugars	Disp. lipid	Sol. Ca	Technology	Nearest class	Cat.
A-041	record conflict	milk fat	−3 ± 12	43.6	0	2	145	–	ripened	GR	A
A-039	Parmesan, no cellulose	milk fat	−1 ± 12	35.0	0	10	369	–	ripened	SH	A
A-042	Parmesan	milk fat	1 ± 12	30.8	13	12	1152	–	ripened	GR	A
A-006	Parmesan + Romano	milk fat	3 ± 12	34.1	0	2	560	–	ripened	SH	A
A-026	Parmesan	milk fat	4 ± 12	31.2	13	3	1030	–	ripened	SH	A
B-016	Pecorino Romano	milk fat	−6 ± 13 †	34.9	12	3	112	336	ripened	GR	A
B-021	Parmesan	milk fat	5 ± 12	42.2	0	2	129	n.r.	ripened	GR	A
B-009	Parmesan	milk fat	6 ± 12	36.4	0	7	167	n.r.	ripened	SH	A
B-007	Parmesan	milk fat	7 ± 11	33.9	0	10	542	355	ripened	SH	A
A-040	Parmesan, store-grated	milk fat	7 ± 11	40.8	11	5	191	–	ripened	GR	A
B-017	Parmesan	milk fat	13 ± 11	45.2	0	1	106	n.r.	ripened	GR	A
A-019	Pecorino Romano	milk fat	10 ± 12 †	33.2	0	6	664	–	ripened	SH	A
B-008	Parmesan + Romano	milk fat, lipolyzed	n.q.	34.2	0	7	576	367	ripened	SH	B
B-015	Parmesan / Pecorino Romano	milk fat, lipolyzed	n.q.	36.8	13	3	352	217	ripened	SH	B
B-003	Romano	milk fat, strongly lipolyzed	n.q.	29.2	0	2	280	–	ripened	DV	B
A-031	Parmesan	milk fat, lipolyzed	n.q.	40.4	0	28	120	–	ripened	GR	B
B-002	Parmesan	milk fat, lipolyzed	n.q.	27.8	9	1	711	–	ripened	GC	B
B-040	record conflict	milk fat	−6 ± 12	1.2	42	679	376	55	processed	DV	E
B-022	Parmesan	milk fat, lipolyzed	n.q.	11.2	46	63	1983	96	processed	BB	E
B-018	Parmesan	low-level non-dairy fat	21 ± 10	3.4	101	362	156	n.r.	ripened	BB	F
B-012	Parmesan	non-dairy fat	43 ± 8	34.1	0	2	113	329	ripened	GR	D
B-006	Parmesan + Romano	non-dairy fat	44 ± 8	11.0	22	41	2093	–	processed	BB	C
A-021	Parmesan	non-dairy fat	45 ± 8	8.0	51	154	1573	–	processed	PC	C
A-018	Parmesan	non-dairy fat	49 ± 8	16.4	99	174	49	–	fresh	DV	C
A-037	Parmesan	non-dairy fat	49 ± 8	8.5	46	64	2235	–	processed	PC	C
A-038	Parmesan	non-dairy fat	49 ± 8	9.7	56	78	1978	–	processed	PC	C
B-013	Parmesan	non-dairy fat	49 ± 8	9.7	46	63	1997	98	processed	BB	C
A-036	Parmesan + Romano	non-dairy fat	50 ± 8	7.7	46	69	2526	–	processed	PC	C
A-035	Parmesan	non-dairy fat	51 ± 8	6.8	50	81	2168	–	processed	PC	C
B-011	Parmesan	non-dairy fat	51 ± 8	13.9	54	73	1604	95	processed	BB	C
B-023	Parmesan	non-dairy fat	51 ± 8	12.5	65	83	1481	87	processed	BB	C
B-010	Parmesan	non-dairy fat	51 ± 8	11.8	57	53	1304	73	processed	BB	C
A-022	Parmesan	non-dairy fat	51 ± 8	11.3	71	125	818	–	processed	PC	C
B-004	Parmesan	non-dairy fat	52 ± 8	13.4	17	18	1680	–	processed	BB	C
A-011	Parmesan + Romano	non-dairy fat	54 ± 8	4.4	43	87	757	–	processed	GC	C
A-043	Parmesan	non-dairy fat	54 ± 7	10.8	57	92	1714	–	processed	PC	C
B-005	Parmesan	non-dairy fat	55 ± 8	6.7	36	65	1799	–	processed	PC	C
B-001	Parmesan	non-dairy fat	57 ± 7	12.4	71	79	1530	–	processed	PC	C
B-024	Parmesan	non-dairy fat	59 ± 7	7.5	57	62	1762	76	processed	BB	C
A-010	Romano	non-dairy fat	59 ± 7	4.3	35	50	1106	–	processed	BB	C
B-014	Pecorino Romano	non-dairy fat	62 ± 7	8.4	40	56	2041	75	processed	PC	C
A-007	Parmesan	non-dairy fat	68 ± 7	8.4	43	60	1784	–	processed	PC	C

8.4 Products that do not fit the main pattern

Three retail products departed from the two main groups, a fourth showed an incomplete form of the category C profile, and each calls for caution. B-012 was flagged for 43 ± 8% of non-milk fat, yet its aqueous extract was that of a ripened cheese in every respect, with a free amino-acid index of 34, no citrate or sugar, little dispersed lipid and soluble calcium of 329 mg per 100 g (category D). B-022 was the reverse: milk fat, slightly lipolyzed, with the complete processed-type aqueous profile and soluble calcium of 96 mg per 100 g (category E). Taken at face value, the first would be a ripened cheese to which vegetable fat alone had been added and the second a processed-type product made with milk fat, and both are possible.

It must be pointed out, however, that the two anomalies are mirror images of one another and that the two codes differ in a single digit, so that an exchange between the two samples would turn both into ordinary members of categories C and A or B. The records constrain what kind of exchange that could have been. The masses recorded with the lipid and with the aqueous acquisition agree within each product, 201.8 mg for B-012 and 199.2 mg for B-022, and differ between them, which argues against a transposition of the acquisition records or files. A physical exchange of two extracts or NMR tubes before acquisition would leave the records in this consistent state and cannot be excluded. Version 2.0 concluded that the lipid extracts were the ones more likely to have been exchanged, because all the aqueous measurements of each product agree with one another. That inference does not hold: the EDTA titration and the archived sodium and phosphorus determinations were all made on the aqueous extract, and they would agree with its ¹H spectrum whichever sample the tube contained. The other products of the same two brands in the study point, if anything, the other way, since two further products of the brand of B-012 belong to category C and a further product of the brand of B-022 to category A; but brand-level consistency is weak evidence, because formulations differ between lots. The spectra cannot decide the matter. Until both products have been extracted again from retained material, neither should be cited as evidence of a distinct mode of adulteration, and neither should be assigned to a main category.

B-018 combined a fat that is mostly milk fat, with an indication of non-milk fat of 21 ± 10%, with an aqueous extract unlike any other in the set: almost no free amino acids (3.4), the highest citrate (101), a large lactose signal, very little acetate and no dispersed lipid (category F). Its EDTA titration consumed the whole of the dose and is not reportable; the figure of at least 636 mg per 100 g given in version 2.0 is withdrawn, and the exhaustion of the ligand shows only that its soluble calcium is not low. This is not the profile of a ripened cheese, nor is it the processed-type profile of category C. Abundant lactose and citrate with no proteolysis suggest a large contribution from unfermented dairy solids such as whey or milk powder, or a very young cheese; the product merits a repeat analysis. A-018 belonged to category C by its fat and its low free amino acids, citrate and sugars, but showed no dispersed lipid, and was called fresh by the technology model; it may represent a different formulation or a more thoroughly clarified extract.

For the two products whose lipid and aqueous records named different products, the spectra are informative. In both cases the title of the lipid acquisition repeats the descriptor and the mass of the preceding sample (Section 3.2), so that it is the lipid title that was not updated. Both extracts of A-041 are those of an authentic ripened cow's-milk cheese, so that whichever record is correct, a store-grated Parmesan or a domestic Parmesan block, the conclusion is the same. For B-040 the lipid extract shows intact milk fat, and the aqueous extract shows lactose, citrate, no free amino acids, low soluble calcium and the highest soluble phosphorus after the processed cheeses; this is the profile of a processed cheese spread made with milk fat, which is what the aqueous record states, and it is inconsistent with fresh mozzarella, which is represented in the set by B-039 and has a quite different aqueous spectrum. The lipid spectrum recorded as B-040 differs from that of B-039, so that a different extract was measured under the carried-over title.

8.5 Agreement between the methods

Across the 42 products of Figure 31, the three families of evidence, ratiometric, model-based and targeted aqueous, led to the same conclusion for the 21 products of category C and, apart from the DD-SIMCA model on aqueous bins, for the eleven of category A. The disagreements were confined to the lipolyzed products, where the bin-based fat models and the one-class models were misled by low butyrate while the ratiometric rule was not; to the full-spectrum one-class models, which rejected authentic products as well; and to the four anomalous products, where the lipid and aqueous evidence point in different directions for reasons that are chemical, or clerical, and not methodological. In no case did a model identify a product that the ratiometric and targeted evaluation had passed.

This agreement is agreement within one data set, and it should be described as such. The categories of the retail products come from the ratiometric rule. The supervised and one-class models were trained on the same reference spectra, with class labels taken from the product labels, and several of them use the markers of the rule as their inputs. The unsupervised analyses use no labels, but they use the same spectra. None of these is an independent measurement, and their concurrence cannot stand in for a comparison with a reference method. What the concurrence does establish is narrower: the grouping does not depend on the particular thresholds of the rule, on the choice of markers or on the form of the model.

9. Comparison of the Ratiometric, Chemometric and Machine-Learning Approaches

The three approaches are compared in Table 21 on the two questions that were put to them, the recognition of a foreign fat and the separation of cheese styles. For the first, the ratiometric evaluation with line fitting was the most informative method examined. It flagged the 22 products, classified the lipolyzed products separately, attached an estimate with a stated uncertainty, characterized the replacing fat by its gross composition, and returned the same result for 43 further spectra of related products. It did this with a reference population of 24 products and with criteria that can each be explained to a non-specialist in terms of a visible peak. The region-ratio screen without fitting found the same 22 products but would have reported four lipolyzed cheeses with them.

The supervised models reached the same conclusions for the clear cases and, in the form of a support vector machine on the ratiometric descriptors, for all of them. Their limitations here are those of the training set, not of the algorithms: nine non-dairy examples of four different kinds cannot define the non-dairy class, the bin-based models had no way of learning that low butyrate with normal linoleate means lipolysis, and a model that must choose among predefined classes will assign an unprecedented sample, such as B-018, to one of them. The one-class formulation is the appropriate one for authentication, and the Mahalanobis models on a handful of chemically chosen descriptors performed well for both extracts, although with false-rejection rates of 5 to 9% on leave-one-out; the one-class models on full binned spectra rejected too many authentic products to be used with reference sets of 21 to 33.

For the separation of cheese styles the order is different. No single ratio separates Grana from Gouda or Pecorino from Gruyère, and it is here that the multivariate treatment of the aqueous spectrum is needed. Even so, the fully nested nine-class accuracy of 0.51 shows that the present reference set, with three to seven products per class from retail sources, is too small for style classification, and the reliable distinctions were the broad ones that the targeted markers also deliver: ripened against unripened, sheep's and alpine milk fat against lowland cow's milk fat through the conjugated diene and n-3 acids, propionic fermentation, and mold ripening through lipolysis and the loss of lactate.

The unsupervised methods occupy a place of their own. Principal component analysis and clustering made no use of labels and nevertheless isolated the non-dairy fat products in the lipid block and joined them to the processed cheeses in the aqueous block. This shows that the groups are the dominant structure of the spectra and were not imposed by the choice of markers; it is a check of consistency within the data set and not a confirmation by an independent method. In a routine setting the same methods would be the means of noticing a deviation of a kind that had not been anticipated, which a fixed set of ratios cannot do. The approaches are thus complementary: ratiometric and targeted measurements for decisions that must be explained and defended, one-class models on the same descriptors for a formal acceptance test once their limits have been set on an adequate reference population, and unsupervised analysis of the full spectra for surveillance.

Table 21. Comparison of the approaches on the present data.

Approach	Question answered	Result in this study	Strengths	Limitations
Region ratios (B/A, B/I, J/I, G/I)	Is the fat milk fat?	22 of 22 flagged products beyond 3 s.d. in B/A, B/I and J/I; 4 of 6 lipolyzed products also beyond the B/A limit	No fitting; minutes per spectrum	Lipolysis and n-3 acids interfere; no quantitation
Published schemes (Tociu; Hanganu–Chira; D1, D2, D3)	Fatty-acid classes; descriptors of fat type	Groups separated; butyrate class carries a floor of 3–5 mol %; D1 and D2 need the water-robust form	Simple stoichiometry; comparable with the literature	Region B includes the methyl-triplet flank and 72% of the n-3 triplet; the G_1/G_2 boundary cuts both bis-allylic triplets; water in region D
Line fitting with decision rule and mixing model	Category, amount and kind of non-milk fat	22 flagged, lipolysis separated, 43–68% estimated to ± 7–8 points (2 s.e.); same result for 43 further spectra of related products	Explainable; quantitative; chemically distinct markers (C4:0, C10:1, C18:2, NC, C_2/C_1)	Not validated; needs good lineshape and an authentic reference population; markers have a blank floor; in simulated blends a definitive call for every blend required 30–50%
Targeted aqueous markers	Is the base a ripened cheese?	21 of 21 category C products outside the ripened limits; minerals available for 6 and concordant	Chemically specific; indicates the kind of base	Area fractions, not concentrations: lowered by dilution and formulation; no inference on curing time; dispersed lipid alone is not specific
PCA, HCA, t-SNE, UMAP	Is there structure without labels?	Flagged products isolated (lipid) and joined to processed cheeses (aqueous)	No labels needed; reveals unforeseen deviations	No decision limit; embeddings qualitative; same spectra as the other methods
Supervised classifiers (7 models)	Which known class?	Fully nested balanced accuracy: fat 0.90, technology 0.82, style 0.51 (first evaluation, best entries: 0.92, 0.94, 0.58)	Use the whole spectrum; rank variables	Limited by 3–9 examples per class; misled by lipolysis; forced choice; optimistic unless all selection is nested
One-class models	Compatible with the authentic class?	Mahalanobis on descriptors: 22 of 22 (lipid), 21 of 21 (aqueous), with 3 of 33 and 1 of 21 authentic products rejected on leave-one-out; DD-SIMCA on bins rejects 15–29% of authentic products	Correct formulation of authentication	Reference sets of 21–33 too small for full-spectrum models; limits not tested on independent authentic samples
PLS regression on in-silico blends	How much non-milk fat?	RMSEP 4 points on simulated blends; median 53% for 21 flagged products inside the domain	No fitting of individual peaks	Error on real products unknown; fails outside its calibration domain (lipolysis, lauric and canola-type fats)

10. Limitations and Points for Expert Review

The conclusions of this report are of unequal strength, and the weaker ones should be identified before the stronger ones are relied upon. What the data support directly, in the sense that several markers within the two spectra agree, is that 22 of the 40 retail grated products were flagged for a fat that is roughly half non-milk fat, of similar gross composition throughout, and that 21 of those 22 have an aqueous extract that is not that of a ripened hard cheese and resembles that of the processed-cheese references. Even this statement rests on one laboratory, one extraction per product and an unvalidated method. What is plausible but inferred is that the replacing fat is palm-based and that the non-cheese component is a cheese analogue made

with emulsifying salts. What the data do not show, and what this report no longer asserts, is the identity of the oil, a shared supplier or formulation, the manufacturing route, the proportion of cheese in any product, or the age of the cheese ingredient. The interpretation of the individual anomalous products is speculative.

The points that an expert reviewer would be most likely to raise are the following. The method has not been validated against a reference method: the fatty-acid percentages rest on stoichiometry and on fitted lineshapes, and no sample was analyzed in parallel by gas chromatography of the fatty-acid methyl esters or by the triacylglycerol method of ISO 17678[17]. The acquisition parameters have now been recovered from the spectral documents, but the flip angle and the relaxation times are not known, so that the absolute values of the fatty-acid classes carry an unquantified bias, although the comparisons between samples do not depend on it. The reference population consists of 24 retail products of presumed, not certified, authenticity, all from cow's milk, and its butyrate content has a relative standard deviation of 5%; the seasonal, regional and breed-related variation of milk fat is known to be wider than a set of this size can capture, and the decision limits and the uncertainty of the individual estimates depend directly on that figure. The decision rule was developed on the products to which it was applied, its false-positive rate has not been measured on independent authentic cheeses, and it does not flag a butyrate deficit that is accompanied neither by a compositional shift nor by lipolysis. For sheep's-milk products the caproleate estimate is biased, since their caproleate content is lower, and only the butyrate estimate should be used.

The non-milk fat is expressed as a fraction of the extracted fat and not of the product, because the fat content of the products was not measured; nothing can therefore be said about the amount of vegetable fat per 100 g of product. Its uncertainty covers the spread of milk fat and the variation between retail lots but not the recovery on extraction, relaxation or the lineshape model, and the residuals of the bis-allylic and sterol fits show that those two quantities are less well determined than butyrate. Markers at the level of the blank cannot be quantified, so that a product with 86% of non-milk fat cannot be distinguished from one with 100%. The description of the fat as palm-type rests on gross composition and on a single reference oil. Cellulose, starch granules and other insoluble fillers are invisible to both extracts, so that a product of category A may still contain an excess of anticaking agent. Each paired product was prepared and measured once; the only replication is between retail lots in the lipid-only set. Both extracts of a product were prepared from one weighed portion, the D_2O extraction being carried out on the solids that remained after the $CDCl_3$ extraction (Section 3.2). The partition of free fatty acids between the two extracts, and the amount of lipid dispersed into the aqueous extract, depend on that sequence and on the solvent volumes and the mixing and centrifugation conditions, the influence of which was not examined.

The aqueous indices are fractions of the total spectral area. They are affected by everything else in the extract, they saturate in mature cheese, they fall with dilution and formulation, and they cannot be compared with concentrations in the literature or converted into a curing time for a formulated product. The extracts were not buffered, so that chemical shifts vary with the pH of each product, and the apparent pH derived from histidine is an approximation. The calcium and magnesium scale is provisional, six of the 36 titrations were not reportable, only twelve retail products have a reportable mineral result, ten of them in categories A to C, and the sodium and phosphorus values were taken from archived worksheets. The record conflicts for A-041, B-040 and B-015, the differing masses of A-036 and B-003 and the possible exchange between B-012 and B-022 are unresolved (Sections 3.1, 3.2 and 8.4).

The multivariate results rest on 42 reference products in nine classes whose labels were taken from the packaging. The fully nested accuracies carry standard deviations of about 0.1, the ranking of preprocessing options is not resolved, and the agreement of the models with the ratiometric evaluation is agreement within one data set. The lipid-only spectra are further lots of related products and not an independent test set. The

simulated blends are linear combinations of spectra, built on the products that define the reference limits; they give the level at which the calculation responds, and neither a detection limit nor an error rate for real products. The confirmatory method presents a difficulty of its own: the triacylglycerol method of ISO 17678 excludes from its scope milk fat obtained from cheeses that show increased lipolysis, because false-positive results can occur[17], so that the lipolyzed products of categories B and E, and the lipolyzed product with non-dairy fat, cannot be confirmed by that method as it stands and need a comparison that has been validated separately for lipolyzed fat.

The present version differs from version 2.0 in the following respects. Mineral results are given only for titrations in which free ligand remained, after the products titrated at two doses showed that exhausted titrations overestimate calcium; the values previously given as lower bounds, including that of B-018, are withdrawn. The uncertainty of the non-milk fat estimates has been enlarged from about ± 5 to about ± 8 points by a between-lot term, markers at the blank level are reported as lower bounds, and the three declared cheese products are accordingly given as at least 86% and no longer as 92 to 100%. The correction of the Hanganu–Chira butyrate for the n-3 triplet has been scaled to the 72% of that triplet which region B captures. One mass discrepancy listed in the workbook of version 2.0, that of A-032, arose from a misreading of a title containing a stray character and has been corrected; four remain. The statement that all retail estimates exceeded the detection limit by a factor of two or more has been deleted together with the detection limit itself, and the statements on the proportion of cheese and on the curing time of the category C products have been withdrawn. Relative to versions 1.0 to 1.2, in which the two campaigns were evaluated separately, the estimates remain lower, with a median of 51% of acyl chains, or about 54% by mass, against 59% by mass, because the methyl region is fitted with three triplets and the reference population excludes lipolyzed and sheep's-milk cheeses. None of these changes alters the number of products flagged.

11. Recommendations

The first requirement for turning this study into a method that could be reported under an accredited scope is validation. Gravimetric blends of an authentic grated Parmesan with palm oil, with palm fractions and with a commercial cheese analogue, at five to eight levels between 5 and 70%, should be carried through the whole procedure from weighing to evaluation, prepared from lots of cheese and fat that played no part in the reference population and analyzed under code by an operator who does not know the levels. They would supply linearity, trueness, the limit of detection and the limit of quantitation in real matrices. A separate set of authentic cheeses of documented origin, not used to set any limit, is needed to measure the false-positive rate of the decision rule and of the one-class models, and the performance of the qualitative decision should be expressed as false-positive and false-negative rates with their confidence intervals[151]. Replicate extractions of single lots on different days would supply repeatability and intermediate precision, which the lot-to-lot variation used here can only bound.

A subset of 15 to 20 products should be analyzed in parallel by gas chromatography of the fatty-acid methyl esters and, for the products that are not lipolyzed, by the triacylglycerol method of ISO 17678. For the lipolyzed products that method is outside its scope, and a comparison would first have to be validated for them: the fatty-acid profile of the total lipid including the free acids, the ¹³C NMR determination of butyrate, or the triacylglycerol analysis of a fat from which free acids and partial glycerides have been removed are candidates[17,19,33]. The reference population should be enlarged with cheeses of documented origin and season, including sheep's- and goat's-milk cheeses with their own limits, so that the natural variability of butyrate, caproleate and the other descriptors is represented properly; the uncertainty budget follows from it.

Several changes to the measurement would repay their small cost. An internal standard of known mass should be added to both extracts, chosen so that it does not bind to protein in the aqueous extract, and the longitudinal relaxation times of the signals used, the butyrate methyl and the TSP singlet in particular, should be measured once so that the recycle time of 8.3 s can be justified or a correction applied. The 90° pulse should be recorded with the data. The fat content of the portion should be determined gravimetrically from the chloroform extract, which would convert the non-milk fat from a fraction of the fat into a fraction of the product. Drying the chloroform extract, or using the water-robust descriptors, removes the interference of water in region D. For the aqueous extract, a defined centrifugation and filtration step would make the dispersed-lipid index reproducible, and a second aliquot diluted in buffer would give pH-independent shifts for alignment while the unbuffered aliquot is kept for the calcium titration. The EDTA dose should be at least doubled so that free ligand remains in every sample, every titration should be checked for free ligand before it is accepted, and the stock concentration should be recorded.

12. Conclusions

A chloroform and a deuterium oxide extract of each of 84 cheese products, measured by ¹H NMR at 300 MHz in a few minutes each, described the products in two separate chemical dimensions. The lipid spectrum, integrated over regions A to J with the bis-allylic and methylene regions split and with the overlapped methyl and minor signals resolved by line fitting, characterized the fat, separated lipolysis from substitution through the different behavior of butyrate and caproleate, and gave an estimate of the non-milk fat with a stated, and incomplete, uncertainty. The aqueous spectrum showed whether the extract was that of a ripened cheese through its free amino acids, citrate, sugars and dispersed lipid, with soluble calcium and phosphorus in support for a minority of the products, and recorded the propionic fermentation of Emmental, the free butyrate of lipolyzed cheeses and the galactose of thermophilic young cheeses along the way.

Of the 40 retail grated Parmesan and Romano products, 16 were consistent in both extracts with grated hard cheese made from milk fat, five of them lipolyzed. Twenty-two were flagged for a fat estimated at 43 to 68% non-milk fat, and 21 of these had an aqueous extract that resembles processed cheese and not a ripened hard cheese; a blend of cheese with a vegetable-fat cheese analogue would account for the composition, but the data do not identify the oil, the route of manufacture, the proportion of cheese or the curing time. Three products did not fit either pattern, and the records of several products are in conflict. Forty-three further lipid spectra of related products gave the same picture. In simulated blends the decision rule made the definitive call for every blend from 30% of palm oil and from 50% of the fat of the declared cheese products, and raised a flag for half of the blends at 20 to 25%; a detection limit for real products has not been established.

For this problem the ratiometric and targeted evaluation was at least as informative as any chemometric or machine-learning model and considerably easier to explain and to defend. The multivariate methods agreed with it within this data set, which shows that the grouping does not depend on the thresholds or the markers chosen, but they were trained on the same spectra and labels and do not constitute an independent confirmation. With all selection nested inside the cross-validation, balanced accuracies of 0.90 for fat, 0.82 for technology and 0.51 for nine styles were obtained; differences between binning, normalization and scaling options were not resolved, and alignment, although needed for interpretation, did not improve classification. The results are those of a screening study. Validation against reference methods and gravimetric blends, an internal standard and a measured false-positive rate are required before individual results are reported.

Appendix A. Sample Register

The register lists the 84 products with paired extracts under their anonymized descriptors, with the class assigned to each and the finding of this study. The masses recorded with the two extracts are given in the companion workbook.

Table 22. Anonymized register of the 84 products with paired $CDCl_3$ and D_2O extracts. A = first campaign, B = second campaign. Class codes are defined in Table 1 (T = retail grated test sample, X = record conflict). The last column gives the finding of this study: for reference products the lipid status and the aqueous call, and for test products the integrated category of Section 8. n.s. = not stated.

Code	Form	Type and style as labeled	Milk	Origin	Age claim	Class	Finding of this study
A-001	grated	Reduced-fat grated Parmesan-style topping, labeled imitation cheese	cow	USA	–	DV	non-dairy fat; unripened/processed-type
A-002	grated	Grated Parmesan "cheese product"	cow	USA	10 mo	DV	non-dairy fat; milk-fat markers at the blank level; unripened/processed-type
A-003	block	Parmigiano Reggiano	cow	Italy	–	GR	milk fat; ripened
A-004	block	Pecorino Romano	sheep	Italy	–	SH	milk fat; ripened
A-005	grated	Grated Parmigiano Reggiano blend, palm oil/starch/cellulose declared	cow	Italy/USA	–	DV	non-dairy fat; ripened, with citrate or sugars
A-006	grated	Grated Parmesan and Romano	cow	USA	–	T	A: milk fat; ripened base
A-007	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-008	block	Aged Gouda	cow	n.s.	aged	GC	milk fat (composition outlier); ripened
A-009	block	Young Gouda (red wax)	cow	n.s.	young	GC	milk fat; ripened
A-010	grated	Grated Romano	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-011	grated	Grated Parmesan and Romano	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-012	block	Grana Padano	cow	Italy	–	GR	milk fat; ripened
A-013	block	Domestic Parmesan (Wisconsin)	cow	USA	–	GR	milk fat; ripened
A-014	block	Manchego	sheep	Spain	6 mo	SH	milk fat; ripened, with citrate or sugars
A-015	block	Domestic aged Asiago (Wisconsin)	cow	USA	aged	AL	milk fat, lipolyzed; ripened
A-016	block	Feta	n.s.	n.s.	–	BB	milk fat, lipolyzed; unripened/processed-type
A-017	block	Smoked Gouda	cow	n.s.	–	GC	milk fat; ripened, with citrate or sugars
A-018	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-019	grated	Grated Pecorino Romano	sheep	Italy	–	T	A: milk fat; ripened base
A-020	slice	American cheese slice (deli)	cow	USA	–	PC	milk fat; unripened/processed-type
A-021	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-022	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-025	block	Domestic extra-aged Parmesan (Grana-style)	cow	USA	18 mo	GR	milk fat; ripened
A-026	grated	Grated Parmesan	cow	USA	10 mo	T	A: milk fat; ripened base
A-027	slice	American singles, pasteurized cheese product	cow	USA	–	PC	milk fat; unripened/processed-type
A-028	round	Plant-based cheese alternative (snack round)	none	n.s.	–	PB	no milk-fat marker detected; unripened/processed-type
A-029	slice	Plant-based Cheddar-style slices	none	n.s.	–	PB	no milk-fat marker detected; unripened/processed-type
A-030	slice	Plant-based Cheddar-style slices	none	n.s.	–	PB	no milk-fat marker detected; unripened/processed-type
A-031	grated	Grated Parmesan	cow	USA	–	T	B: milk fat, lipolyzed; ripened base
A-032	shredded	Low-moisture Mozzarella, shredded	cow	USA	–	YF	milk fat; unripened/processed-type
A-033	shredded	Low-moisture Mozzarella, shredded	cow	USA	–	YF	milk fat; unripened/processed-type
A-034	loaf	Pasteurized prepared cheese product (loaf)	cow	USA	–	DV	non-dairy fat; unripened/processed-type
A-035	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-036	grated	Grated Parmesan and Romano	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base

Code	Form	Type and style as labeled	Milk	Origin	Age claim	Class	Finding of this study
A-037	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
A-038	grated	Grated Parmesan	cow	USA	10 mo	T	C: non-dairy fat; unripened/processed-type base
A-039	grated	Grated Parmesan, no cellulose	cow	USA	–	T	A: milk fat; ripened base
A-040	grated	Parmesan grated in store from domestic block	cow	USA	–	T	A: milk fat; ripened base
A-041	grated/ block	Record conflict: lipid record = store-grated Parmesan; aqueous record = domestic Parmesan block, 20 mo	cow	USA	20 mo?	X	A: milk fat; ripened base
A-042	grated	Grated Parmesan	cow	USA	10 mo	T	A: milk fat; ripened base
A-043	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-001	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-002	grated	Grated Parmesan	cow	USA	–	T	B: milk fat, lipolyzed; ripened base
B-003	grated	Grated Romano	n.s.	USA	–	T	B: milk fat, lipolyzed; ripened base
B-004	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-005	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-006	grated	Grated Parmesan and Romano	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-007	grated	Grated Parmesan	cow	USA	–	T	A: milk fat; ripened base
B-008	grated	Grated Parmesan and Romano	cow	USA	–	T	B: milk fat, lipolyzed; ripened base
B-009	grated	Grated Parmesan	cow	USA	–	T	A: milk fat; ripened base
B-010	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-011	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-012	grated	Grated Parmesan	cow	USA	–	T	D: non-dairy fat; ripened base
B-013	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-014	grated	Grated Pecorino Romano	sheep (label)	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-015	grated	Grated Parmesan (lipid record) / Pecorino Romano (aqueous record)	n.s.	USA	–	T	B: milk fat, lipolyzed; ripened base
B-016	grated	Grated Pecorino Romano	sheep	Italy	–	T	A: milk fat; ripened base
B-017	grated	Grated Parmesan	cow	USA	–	T	A: milk fat; ripened base
B-018	grated	Grated Parmesan	cow	USA	–	T	F: low-level non-dairy fat indicated; unripened base
B-019	grated	Grated Parmesan "cheese product"	cow	USA	10 mo	DV	non-dairy fat; milk-fat markers at the blank level; unripened/processed-type
B-020	grated	Grated Pecorino Romano "cheese product"	n.s.	USA	10 mo	DV	non-dairy fat; milk-fat markers at the blank level; unripened/processed-type
B-021	grated	Grated Parmesan	cow	USA	10 mo	T	A: milk fat; ripened base
B-022	grated	Grated Parmesan	cow	USA	–	T	E: milk fat; unripened/processed-type base
B-023	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-024	grated	Grated Parmesan	cow	USA	–	T	C: non-dairy fat; unripened/processed-type base
B-025	block	Monterey Jack	cow	USA	–	YF	milk fat; unripened/processed-type
B-026	block	Muenster (American)	cow	USA	–	YF	milk fat; unripened/processed-type
B-027	slice	American cheese slice	cow	USA	–	PC	milk fat; unripened/processed-type
B-028	block	Blue Stilton	cow	UK	–	BB	milk fat; ripened
B-029	block	Buffalo-milk blue cheese	buffalo	Italy	–	BB	milk fat, strongly lipolyzed; ripened
B-030	block	Sheep's-milk table cheese	sheep	Italy	–	SH	milk fat; unripened/processed-type
B-031	block	Mimolette	cow	France	12 mo	GC	milk fat; ripened
B-032	block	Irish Cheddar-type (sweet, aged)	cow	Ireland	–	GC	milk fat; ripened
B-033	block	Piave Vecchio	cow	Italy	>10 mo	GR	milk fat; ripened
B-034	block	Asiago Pressato (fresh)	cow	Italy	fresh	AL	milk fat; unripened/processed-type
B-035	block	Emmental	cow	Switzerland	11 mo	AL	milk fat; ripened

Code	Form	Type and style as labeled	Milk	Origin	Age claim	Class	Finding of this study
B-036	block	Gruyère	cow	Switzerland	16 mo	AL	milk fat; ripened
B-037	block	Gouda, extra aged	cow	Netherlands	26 mo	GC	milk fat; ripened
B-038	block	Gouda, aged	cow	Netherlands	18 mo	GC	milk fat; ripened
B-039	ball	Fresh Mozzarella	cow	USA	fresh	YF	milk fat; unripened/processed-type
B-040	wedge/ball	Record conflict: lipid record = fresh Mozzarella; aqueous record and worksheet = processed cheese spread wedge	cow	n.s.	–	X	E: milk fat; unripened/processed-type base
B-041	block	Pecorino Romano	sheep	Italy	–	SH	milk fat; ripened
B-042	block	Domestic extra-aged Parmesan (Grana-style)	cow	USA	18 mo	GR	milk fat; ripened
B-043	slice	American cheese single, pasteurized prepared cheese product	cow	USA	–	PC	milk fat; unripened/processed-type

Appendix B. Model Comparison Tables

The tables of this appendix give the full results of the preprocessing comparison, of the first cross-validation of the seven classifiers and of the one-class models, which are discussed in Sections 7.1, 7.4 and 7.5.

Table 23. Effect of each preprocessing factor on the cross-validated score, averaged over all other factors (292 combinations; reference products only). Chance level is 0.50 for the fat task, 0.33 for technology and 0.11 for style.

Factor	Level	Lipid: fat	Lipid: style	Aqueous: technology	Aqueous: style
Representation	uniform bins, 0.002 ppm	0.920	0.371	0.805	0.487
	uniform bins, 0.005 ppm	0.918	0.371	0.818	0.489
	uniform bins, 0.01 ppm	0.916	0.369	0.824	0.488
	uniform bins, 0.02 ppm	0.912	0.377	0.813	0.481
	uniform bins, 0.04 ppm	0.911	0.382	0.805	0.478
	adaptive (intelligent) bucketing	0.906	0.377	0.791	0.486
	optimized bucketing, 0.01 ppm	0.923	0.385	0.826	0.490
	integrals A–J with G_1, G_2, C_1, C_2	0.896	0.385	–	–
	ratiometric descriptors (18)	0.900	0.371	–	–
	targeted metabolites (24)	–	–	0.831	0.490
Alignment (binned data)	not aligned	0.918	0.375	0.823	0.482
	icoshift-aligned	0.913	0.377	0.800	0.489
Normalization	glycerol CH_2 (H) = 100	0.915	0.380	–	–
	total area	0.919	0.377	0.811	0.479
	probabilistic quotient	0.912	0.372	0.813	0.492
	none beyond the block default	0.898	0.378	0.831	0.490
Scaling	mean-centering only	0.899	0.353	0.814	0.507
	Pareto	0.927	0.387	0.828	0.524
	unit variance	0.919	0.387	0.792	0.475
	log, then Pareto	0.911	0.378	0.816	0.437

Table 24. Cross-validated balanced accuracy of the first evaluation (mean ± s.d.; ten repetitions of a three-fold outer loop with hyperparameters selected in an inner loop; preprocessing of each block fixed beforehand on all reference products) for seven classifiers. An asterisk marks the best model in each row. Label-permutation baselines: fat 0.50 (95th percentile 0.62), technology 0.33 (0.50), style 0.11 (0.20–0.22). The fully nested estimates are given in Table 18.

Task	Data block	SVM, linear	SVM, RBF	PLS-DA	kNN	Random forest	Logistic (L2)	LDA on 5 PCs
Dairy vs non-dairy fat	lipid bins	0.89 ± 0.12	0.87 ± 0.10	0.89 ± 0.13	0.92 ± 0.08*	0.91 ± 0.08	0.88 ± 0.14	0.87 ± 0.13
Dairy vs non-dairy fat	lipid descriptors	0.88 ± 0.11	0.90 ± 0.11	0.90 ± 0.08	0.89 ± 0.08	0.89 ± 0.09	0.87 ± 0.10	0.91 ± 0.08*
Dairy vs non-dairy fat	fused scores	0.89 ± 0.11	0.87 ± 0.13	0.89 ± 0.12*	0.89 ± 0.15	0.87 ± 0.14	0.89 ± 0.11	0.89 ± 0.12
Technology (3 classes)	aqueous bins	0.80 ± 0.14	0.79 ± 0.16	0.86 ± 0.11	0.80 ± 0.15	0.74 ± 0.11	0.86 ± 0.13	0.88 ± 0.11*
Technology (3 classes)	aqueous targeted	0.90 ± 0.13	0.92 ± 0.10	0.93 ± 0.07	0.84 ± 0.13	0.84 ± 0.15	0.94 ± 0.06*	0.92 ± 0.11
Technology (3 classes)	fused scores	0.83 ± 0.12	0.84 ± 0.12	0.83 ± 0.14	0.86 ± 0.11*	0.82 ± 0.15	0.85 ± 0.11	0.83 ± 0.13
Style (9 classes)	lipid bins	0.43 ± 0.12	0.36 ± 0.10	0.40 ± 0.07	0.36 ± 0.09	0.45 ± 0.13*	0.42 ± 0.12	0.44 ± 0.10
Style (9 classes)	lipid descriptors	0.40 ± 0.08	0.40 ± 0.09	0.38 ± 0.09	0.39 ± 0.09	0.43 ± 0.08	0.41 ± 0.10	0.44 ± 0.09*
Style (9 classes)	aqueous bins	0.53 ± 0.08	0.50 ± 0.11	0.46 ± 0.09	0.58 ± 0.09*	0.47 ± 0.09	0.50 ± 0.10	0.50 ± 0.12
Style (9 classes)	aqueous targeted	0.49 ± 0.10	0.50 ± 0.10	0.48 ± 0.09	0.55 ± 0.09	0.57 ± 0.10*	0.53 ± 0.08	0.53 ± 0.08
Style (9 classes)	fused scores	0.40 ± 0.13	0.35 ± 0.15	0.47 ± 0.06	0.58 ± 0.09	0.58 ± 0.09*	0.53 ± 0.09	0.43 ± 0.08

Table 25. One-class models: number of products rejected in each group (rejected/total). Lipid models were built on the 33 reference products that contain only dairy fat, aqueous models on the 21 ripened reference cheeses (GR, SH, AL, GC); the last column gives the number of authentic products rejected when each was left out of the model in turn. Retail grated products are grouped by the integrated categories of Section 8.

Model	GR	SH	AL	GC	BB	YF	PC	DV	PB	T: A	T: B	T: C	T: D	T: E	T: F	X	LOO
Lipid bins, DD-SIMCA (3 PCs)	0/6	0/4	0/4	0/7	0/3	0/5	0/4	6/6	3/3	0/11	4/5	21/21	1/1	1/1	0/1	0/2	5/33
Lipid descriptors, Mahalanobis (5 variables)	0/6	0/4	0/4	1/7	1/3	0/5	0/4	6/6	3/3	0/11	3/5	21/21	1/1	0/1	0/1	0/2	3/33
Lipid bins, one-class SVM (5 PCs)	1/6	1/4	2/4	1/7	2/3	1/5	2/4	6/6	2/3	0/11	4/5	21/21	1/1	1/1	0/1	0/2	–
Lipid bins, isolation forest (5 PCs)	0/6	0/4	0/4	0/7	1/3	0/5	1/4	4/6	0/3	0/11	2/5	14/21	0/1	1/1	0/1	0/2	–
Aqueous bins, DD-SIMCA (3 PCs)	0/6	0/4	0/4	0/7	3/3	5/5	4/4	6/6	3/3	5/11	5/5	21/21	1/1	1/1	1/1	1/2	6/21
Aqueous targeted, Mahalanobis (8 variables)	0/6	0/4	0/4	0/7	2/3	5/5	4/4	5/6	3/3	0/11	1/5	21/21	0/1	1/1	1/1	1/2	1/21

References

1. Fox, P. F.; Guinee, T. P.; Cogan, T. M.; McSweeney, P. L. H. Fundamentals of Cheese Science; Springer US, 2017. https://doi.org/10.1007/978-1-4899-7681-9
2. Gobbetti, M.; Neviani, E.; Fox, P. The Cheeses of Italy: Science and Technology; Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-89854-4
3. Tunick, M. H. The Science of Cheese; Oxford University Press: New York, 2014.
4. Everstine, K.; Spink, J.; Kennedy, S. Economically Motivated Adulteration (EMA) of Food: Common Characteristics of EMA Incidents. Journal of Food Protection 2013, 76 (4), 723–735. https://doi.org/10.4315/0362-028x.jfp-12-399
5. Moore, J. C.; Spink, J.; Lipp, M. Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. Journal of Food Science 2012, 77 (4). https://doi.org/10.1111/j.1750-3841.2012.02657.x
6. Handford, C. E.; Campbell, K.; Elliott, C. T. Impacts of Milk Fraud on Food Safety and Nutrition with Special Emphasis on Developing Countries. Comprehensive Reviews in Food Science and Food Safety 2016, 15 (1), 130–142. https://doi.org/10.1111/1541-4337.12181

7. Kamal, M.; Karoui, R. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. Trends in Food Science & Technology 2015, 46 (1), 27–48. https://doi.org/10.1016/j.tifs.2015.07.007

8. Joolaei Ahranjani, P.; Joolaei Ahranjani, P.; Dehghan, K.; Esfandiari, Z.; Ferrentino, G. Advances in spectroscopic techniques for the detection of cheese adulteration: A systematic review. Food Chemistry: X 2025, 29, 102685. https://doi.org/10.1016/j.fochx.2025.102685

9. U.S. Food and Drug Administration. Grated Cheeses. Code of Federal Regulations, Title 21, Section 133.146. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-133/subpart-B/section-133.146 (accessed November 9, 2026).

10. U.S. Food and Drug Administration. Parmesan and Reggiano Cheese. Code of Federal Regulations, Title 21, Section 133.165. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-133/subpart-B/section-133.165 (accessed November 9, 2026).

11. U.S. Food and Drug Administration. Romano Cheese. Code of Federal Regulations, Title 21, Section 133.183. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-133/subpart-B/section-133.183 (accessed November 9, 2026).

12. Kapoor, R.; Metzger, L. E. Process Cheese: Scientific and Technological Aspects—A Review. Comprehensive Reviews in Food Science and Food Safety 2008, 7 (2), 194–214. https://doi.org/10.1111/j.1541-4337.2008.00040.x

13. Bachmann, H.-P. Cheese analogues: a review. International Dairy Journal 2001, 11 (4–7), 505–515. https://doi.org/10.1016/s0958-6946(01)00073-5

14. Guinee, T. P. Pasteurized Processed and Imitation Cheese Products. In Cheese; Elsevier, 2017; pp 1133–1184. https://doi.org/10.1016/b978-0-12-417012-4.00046-6

15. U.S. Food and Drug Administration. Pasteurized Process Cheese; Pasteurized Process Cheese Food; Pasteurized Process Cheese Spread. Code of Federal Regulations, Title 21, Sections 133.169, 133.173 and 133.179. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-133/subpart-B (accessed November 9, 2026).

16. Bloomberg News. The Parmesan Cheese You Sprinkle on Your Penne Could Be Wood. Bloomberg, February 16, 2016; republished by Dairy Herd. https://www.dairyherd.com/news/parmesan-cheese-you-sprinkle-your-penne-could-be-wood (accessed November 9, 2026).

17. International Organization for Standardization; International Dairy Federation. Milk and Milk Products: Determination of Milk Fat Purity by Gas Chromatographic Analysis of Triglycerides; ISO 17678:2019 | IDF 202:2019; ISO: Geneva, 2019.

18. Precht, D. Detection of foreign fat in milk fat I. Qualitative detection by triacylglycerol formulae. Zeitschrift für Lebensmittel-Untersuchung und -Forschung 1992, 194 (1), 1–8. https://doi.org/10.1007/bf01191031

19. Molkentin, J. Detection of foreign fat in milk fat from different continents by triacylglycerol analysis. European Journal of Lipid Science and Technology 2007, 109 (5), 505–510. https://doi.org/10.1002/ejlt.200600286

20. Guillén, M. D.; Ruiz, A. High resolution ¹H nuclear magnetic resonance in the study of edible oils and fats. Trends in Food Science & Technology 2001, 12 (9), 328–338. https://doi.org/10.1016/s0924-2244(01)00101-7

21. Guillén, M. D.; Ruiz, A. Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils. European Journal of Lipid Science and Technology 2003, 105 (11), 688–696. https://doi.org/10.1002/ejlt.200300866

22. Knothe, G.; Kenar, J. A. Determination of the fatty acid profile by ¹H-NMR spectroscopy. European Journal of Lipid Science and Technology 2004, 106 (2), 88–96. https://doi.org/10.1002/ejlt.200300880

23. Barison, A.; Pereira da Silva, C. W.; Campos, F. R.; Simonelli, F.; Lenz, C. A.; Ferreira, A. G. A simple methodology for the determination of fatty acid composition in edible oils through ¹H NMR spectroscopy. Magnetic Resonance in Chemistry 2010, 48 (8), 642–650. https://doi.org/10.1002/mrc.2629

24. Alexandri, E.; Ahmed, R.; Siddiqui, H.; Choudhary, M.; Tsiafoulis, C.; Gerothanassis, I. High Resolution NMR Spectroscopy as a Structural and Analytical Tool for Unsaturated Lipids in Solution. Molecules 2017, 22 (10), 1663. https://doi.org/10.3390/molecules22101663
25. Siudem, P.; Zielińska, A.; Paradowska, K. Application of ¹H NMR in the study of fatty acids composition of vegetable oils. Journal of Pharmaceutical and Biomedical Analysis 2022, 212, 114658. https://doi.org/10.1016/j.jpba.2022.114658
26. Vlahov, G. Application of NMR to the study of olive oils. Progress in Nuclear Magnetic Resonance Spectroscopy 1999, 35 (4), 341–357. https://doi.org/10.1016/s0079-6565(99)00015-1
27. Dais, P.; Hatzakis, E. Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review. Analytica Chimica Acta 2013, 765, 1–27. https://doi.org/10.1016/j.aca.2012.12.003
28. Jensen, R. G. The Composition of Bovine Milk Lipids: January 1995 to December 2000. Journal of Dairy Science 2002, 85 (2), 295–350. https://doi.org/10.3168/jds.s0022-0302(02)74079-4
29. Lindmark Månsson, H. Fatty acids in bovine milk fat. Food & Nutrition Research 2008, 52 (1), 1821. https://doi.org/10.3402/fnr.v52i0.1821
30. Brescia, M. A.; Mazzilli, V.; Sgaramella, A.; Ghelli, S.; Fanizzi, F. P.; Sacco, A. ¹H NMR characterization of milk lipids: A comparison between cow and buffalo milk. Journal of the American Oil Chemists' Society 2004, 81 (5), 431–436. https://doi.org/10.1007/s11746-004-0918-3
31. Andreotti, G.; Lamanna, R.; Trivellone, E.; Motta, A. ¹³C NMR spectra of TAG: An easy way to distinguish milks from different animal species. Journal of the American Oil Chemists' Society 2002, 79 (2), 123–127. https://doi.org/10.1007/s11746-002-0445-2
32. Scano, P.; Anedda, R.; Melis, M. P.; Dessi', M. A.; Lai, A.; Roggio, T. ¹H- and ¹³C-NMR Characterization of the Molecular Components of the Lipid Fraction of Pecorino Sardo Cheese. Journal of the American Oil Chemists' Society 2011, 88 (9), 1305–1316. https://doi.org/10.1007/s11746-011-1797-9
33. Sacchi, R.; Paduano, A.; Caporaso, N.; Picariello, G.; Romano, R.; Addeo, F. Assessment of milk fat content in fat blends by ¹³C NMR spectroscopy analysis of butyrate. Food Control 2018, 91, 231–236. https://doi.org/10.1016/j.foodcont.2018.04.011
34. Schripsema, J. Comprehensive Analysis of Polar and Apolar Constituents of Butter and Margarine by Nuclear Magnetic Resonance, Reflecting Quality and Production Processes. Journal of Agricultural and Food Chemistry 2008, 56 (8), 2547–2552. https://doi.org/10.1021/jf072274c
35. Tociu, M.; Todasca, M.-C.; Bratu, A.; Mihalache, M.; Manolache, F. Fast approach for fatty acid profiling of dairy products fats using ¹H-NMR spectroscopy. International Dairy Journal 2018, 83, 52–57. https://doi.org/10.1016/j.idairyj.2018.03.005
36. Hanganu, A.; Chira, N.-A. When detection of dairy food fraud fails: An alternative approach through proton nuclear magnetic resonance spectroscopy. Journal of Dairy Science 2021, 104 (8), 8454–8466. https://doi.org/10.3168/jds.2020-19883
37. Scano, P.; Cagliani, L. R.; Consonni, R. ¹H NMR characterisation of the lipid fraction and the metabolite profiles of Fossa (pit) cheese. International Dairy Journal 2019, 90, 39–44. https://doi.org/10.1016/j.idairyj.2018.10.007
38. Prema, D.; Pilfold, J. L.; Krauchi, J.; Church, J. S.; Donkor, K. K.; Cinel, B. Rapid Determination of Total Conjugated Linoleic Acid Content in Select Canadian Cheeses by¹H NMR Spectroscopy. Journal of Agricultural and Food Chemistry 2013, 61 (41), 9915–9921. https://doi.org/10.1021/jf402627q
39. Haddad, L.; Francis, J.; Rizk, T.; Akoka, S.; Remaud, G. S.; Bejjani, J. Cheese characterization and authentication through lipid biomarkers obtained by high-resolution ¹H NMR profiling. Food Chemistry 2022, 383, 132434. https://doi.org/10.1016/j.foodchem.2022.132434
40. Tsiafoulis, C. G.; Skarlas, T.; Tzamaloukas, O.; Miltiadou, D.; Gerothanassis, I. P. Direct nuclear magnetic resonance identification and quantification of geometric isomers of conjugated linoleic acid in milk lipid fraction without derivatization steps: Overcoming sensitivity and resolution barriers. Analytica Chimica Acta 2014, 821, 62–71. https://doi.org/10.1016/j.aca.2014.03.010

41. Tsiafoulis, C. G.; Papaemmanouil, C.; Alivertis, D.; Tzamaloukas, O.; Miltiadou, D.; Balayssac, S.; Malet-Martino, M.; Gerothanassis, I. P. NMR-Based Metabolomics of the Lipid Fraction of Organic and Conventional Bovine Milk. Molecules 2019, 24 (6), 1067. https://doi.org/10.3390/molecules24061067

42. Monakhova, Y. B.; Godelmann, R.; Andlauer, C.; Kuballa, T.; Lachenmeier, D. W. Identification of Imitation Cheese and Imitation Ice Cream Based on Vegetable Fat Using NMR Spectroscopy and Chemometrics. International Journal of Food Science 2013, 2013, 1–9. https://doi.org/10.1155/2013/367841

43. Monakhova, Y. B.; Kuballa, T.; Leitz, J.; Andlauer, C.; Lachenmeier, D. W. NMR spectroscopy as a screening tool to validate nutrition labeling of milk, lactose-free milk, and milk substitutes based on soy and grains. Dairy Science & Technology 2012, 92 (2), 109–120. https://doi.org/10.1007/s13594-011-0050-5

44. Utoiu, C. D.; Manolache, F. A.; Ionescu, V.; Serbancea, F.; Todasca, M. C. Methods Assessment for a Sustainable Preparation of the Animal Fat Samples from Dairy Matrix for ¹H-NMR Analysis Used to Check Dairy Products Conformity. International Journal of Innovative Approaches in Agricultural Research 2018, 2 (4), 285–294. https://doi.org/10.29329/ijiaar.2018.174.3

45. Erich, S.; Schill, S.; Annweiler, E.; Waiblinger, H.-U.; Kuballa, T.; Lachenmeier, D. W.; Monakhova, Y. B. Combined chemometric analysis of ¹H NMR, ¹³C NMR and stable isotope data to differentiate organic and conventional milk. Food Chemistry 2015, 188, 1–7. https://doi.org/10.1016/j.foodchem.2015.04.118

46. Guyader, S.; Thomas, F.; Portaluri, V.; Jamin, E.; Akoka, S.; Silvestre, V.; Remaud, G. Authentication of edible fats and oils by non-targeted ¹³C INEPT NMR spectroscopy. Food Control 2018, 91, 216–224. https://doi.org/10.1016/j.foodcont.2018.03.046

47. Satyarthi, J. K.; Srinivas, D.; Ratnasamy, P. Estimation of Free Fatty Acid Content in Oils, Fats, and Biodiesel by ¹H NMR Spectroscopy. Energy & Fuels 2009, 23 (4), 2273–2277. https://doi.org/10.1021/ef801011v

48. Consonni, R.; Cagliani, L. R. Ripening and geographical characterization of Parmigiano Reggiano cheese by ¹H NMR spectroscopy. Talanta 2008, 76 (1), 200–205. https://doi.org/10.1016/j.talanta.2008.02.022

49. Shintu, L.; Caldarelli, S. High-Resolution MAS NMR and Chemometrics: Characterization of the Ripening of Parmigiano Reggiano Cheese. Journal of Agricultural and Food Chemistry 2005, 53 (10), 4026–4031. https://doi.org/10.1021/jf048141y

50. Shintu, L.; Caldarelli, S. Toward the Determination of the Geographical Origin of Emmental(er) Cheese via High Resolution MAS NMR: A Preliminary Investigation. Journal of Agricultural and Food Chemistry 2006, 54 (12), 4148–4154. https://doi.org/10.1021/jf060532k

51. de Angelis Curtis, S.; Curini, R.; Delfini, M.; Brosio, E.; D'Ascenzo, F.; Bocca, B. Amino acid profile in the ripening of Grana Padano cheese: a NMR study. Food Chemistry 2000, 71 (4), 495–502. https://doi.org/10.1016/s0308-8146(00)00192-8

52. Maestrello, V.; Solovyev, P.; Stroppa, A.; Bontempo, L.; Franceschi, P. ¹H NMR profiling and chemometric analysis for ripening and production characterization of Grana Padano cheese. Food Chemistry 2024, 456, 139986. https://doi.org/10.1016/j.foodchem.2024.139986

53. Cavallini, N.; Strani, L.; Becchi, P. P.; Pizzamiglio, V.; Michelini, S.; Savorani, F.; Cocchi, M.; Durante, C. Tracing the identity of Parmigiano Reggiano “Prodotto di Montagna - Progetto Territorio” cheese using NMR spectroscopy and multivariate data analysis. Analytica Chimica Acta 2023, 1278, 341761. https://doi.org/10.1016/j.aca.2023.341761

54. Piras, C.; Cesare Marincola, F.; Savorani, F.; Engelsen, S. B.; Cosentino, S.; Viale, S.; Pisano, M. B. A NMR metabolomics study of the ripening process of the Fiore Sardo cheese produced with autochthonous adjunct cultures. Food Chemistry 2013, 141 (3), 2137–2147. https://doi.org/10.1016/j.foodchem.2013.04.108

55. Ruyssen, T.; Janssens, M.; Van Gasse, B.; Van Laere, D.; Van der Eecken, N.; De Meerleer, M.; Vermeiren, L.; Van Hoorde, K.; Martins, J. C.; Uyttendaele, M.; De Vuyst, L. Characterisation of Gouda cheeses based on sensory, analytical and high-field ¹H nuclear magnetic resonance spectroscopy determinations: Effect of adjunct cultures and brine composition on sodium-reduced Gouda cheese. International Dairy Journal 2013, 33 (2), 142–152. https://doi.org/10.1016/j.idairyj.2013.04.009

56. Eramo, V.; Massaro, F.; Forniti, R.; Porcelli, F.; Botondi, R. ¹H NMR-based characterization of the aqueous extract of aged Toma Piemontese PDO Cheese: Qualitative impact of gaseous ozone to control surface

microbiota versus normal ripening conditions. International Dairy Journal 2025, 170, 106358. https://doi.org/10.1016/j.idairyj.2025.106358

57. Mazzei, P.; Piccolo, A. ¹H HRMAS-NMR metabolomic to assess quality and traceability of mozzarella cheese from Campania buffalo milk. Food Chemistry 2012, 132 (3), 1620–1627. https://doi.org/10.1016/j.foodchem.2011.11.142
58. Pisano, M. B.; Scano, P.; Murgia, A.; Cosentino, S.; Caboni, P. Metabolomics and microbiological profile of Italian mozzarella cheese produced with buffalo and cow milk. Food Chemistry 2016, 192, 618–624. https://doi.org/10.1016/j.foodchem.2015.07.061
59. Gianferri, R.; Maioli, M.; Delfini, M.; Brosio, E. A low-resolution and high-resolution nuclear magnetic resonance integrated approach to investigate the physical structure and metabolic profile of Mozzarella di Bufala Campana cheese. International Dairy Journal 2007, 17 (2), 167–176. https://doi.org/10.1016/j.idairyj.2006.02.006
60. Brescia, M. Characterisation of the geographical origin of buffalo milk and mozzarella cheese by means of analytical and spectroscopic determinations. Food Chemistry 2005, 89 (1), 139–147. https://doi.org/10.1016/j.foodchem.2004.02.016
61. Patel, H.; Aru, V.; Sørensen, K. M.; Engelsen, S. B. Towards on-line cheese monitoring: Exploration of semi-hard cheeses using NIR and ¹H NMR spectroscopy. Food Chemistry 2024, 454, 139786. https://doi.org/10.1016/j.foodchem.2024.139786
62. Chen, Y.; MacNaughtan, W.; Jones, P.; Yang, Q.; Williams, H.; Foster, T. Selection of potential molecular markers for cheese ripening and quality prediction by NMR spectroscopy. LWT 2021, 136, 110306. https://doi.org/10.1016/j.lwt.2020.110306
63. Rodrigues, D.; Santos, C. H.; Rocha-Santos, T. A. P.; Gomes, A. M.; Goodfellow, B. J.; Freitas, A. C. Metabolic Profiling of Potential Probiotic or Synbiotic Cheeses by Nuclear Magnetic Resonance (NMR) Spectroscopy. Journal of Agricultural and Food Chemistry 2011, 59 (9), 4955–4961. https://doi.org/10.1021/jf104605r
64. Schievano, E.; Guardini, K.; Mammi, S. Fast Determination of Histamine in Cheese by Nuclear Magnetic Resonance (NMR). Journal of Agricultural and Food Chemistry 2009, 57 (7), 2647–2652. https://doi.org/10.1021/jf803364k
65. Belloque, J.; Ramos, M. Application of NMR spectroscopy to milk and dairy products. Trends in Food Science & Technology 1999, 10 (10), 313–320. https://doi.org/10.1016/s0924-2244(00)00012-1
66. Sundekilde, U.; Larsen, L.; Bertram, H. NMR-Based Milk Metabolomics. Metabolites 2013, 3 (2), 204–222. https://doi.org/10.3390/metabo3020204
67. Mannina, L.; Sobolev, A. P.; Viel, S. Liquid state ¹H high field NMR in food analysis. Progress in Nuclear Magnetic Resonance Spectroscopy 2012, 66, 1–39. https://doi.org/10.1016/j.pnmrs.2012.02.001
68. Scano, P.; Cusano, E.; Caboni, P.; Consonni, R. NMR metabolite profiles of dairy: A review. International Dairy Journal 2019, 90, 56–67. https://doi.org/10.1016/j.idairyj.2018.11.004
69. Hu, F.; Furihata, K.; Ito-Ishida, M.; Kaminogawa, S.; Tanokura, M. Nondestructive Observation of Bovine Milk by NMR Spectroscopy: Analysis of Existing States of Compounds and Detection of New Compounds. Journal of Agricultural and Food Chemistry 2004, 52 (16), 4969–4974. https://doi.org/10.1021/jf049616o
70. Li, Q.; Yu, Z.; Zhu, D.; Meng, X.; Pang, X.; Liu, Y.; Frew, R.; Chen, H.; Chen, G. The application of NMR-based milk metabolite analysis in milk authenticity identification. Journal of the Science of Food and Agriculture 2017, 97 (9), 2875–2882. https://doi.org/10.1002/jsfa.8118
71. Lamanna, R.; Braca, A.; Di Paolo, E.; Imparato, G. Identification of milk mixtures by ¹H NMR profiling. Magnetic Resonance in Chemistry 2011, 49 (S1). https://doi.org/10.1002/mrc.2807
72. Hatzakis, E. Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. Comprehensive Reviews in Food Science and Food Safety 2019, 18 (1), 189–220. https://doi.org/10.1111/1541-4337.12408
73. Sobolev, A. P.; Thomas, F.; Donarski, J.; Ingallina, C.; Circi, S.; Cesare Marincola, F.; Capitani, D.; Mannina, L. Use of NMR applications to tackle future food fraud issues. Trends in Food Science & Technology 2019, 91, 347–353. https://doi.org/10.1016/j.tifs.2019.07.035

74. Setyadi, N. W. Detection of Adulteration in New Zealand High Value Dairy Products Using NMR-Based Metabolomics: A Chemometrics Approach. M.Sc. Thesis, University of Otago, Dunedin, New Zealand, 2021.

75. Gallo, V.; Intini, N.; Mastrorilli, P.; Latronico, M.; Scapicchio, P.; Triggiani, M.; Bevilacqua, V.; Fanizzi, P.; Acquotti, D.; Airoldi, C.; Arnesano, F.; Assfalg, M.; Benevelli, F.; Bertelli, D.; Cagliani, L. R.; Casadei, L.; Cesare Marincola, F.; Colafemmina, G.; Consonni, R.; Cosentino, C.; Davalli, S.; De Pascali, S. A.; D'Aiuto, V.; Faccini, A.; Gobetto, R.; Lamanna, R.; Liguori, F.; Longobardi, F.; Mallamace, D.; Mazzei, P.; Menegazzo, I.; Milone, S.; Mucci, A.; Napoli, C.; Pertinhez, T.; Rizzuti, A.; Rocchigiani, L.; Schievano, E.; Sciubba, F.; Sobolev, A.; Tenori, L.; Valerio, M. Performance Assessment in Fingerprinting and Multi Component Quantitative NMR Analyses. Analytical Chemistry 2015, 87 (13), 6709–6717. https://doi.org/10.1021/acs.analchem.5b00919

76. Riedl, J.; Esslinger, S.; Fauhl-Hassek, C. Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. Analytica Chimica Acta 2015, 885, 17–32. https://doi.org/10.1016/j.aca.2015.06.003

77. Ballin, N. Z.; Laursen, K. H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. Trends in Food Science & Technology 2019, 86, 537–543. https://doi.org/10.1016/j.tifs.2018.09.025

78. McGrath, T. F.; Haughey, S. A.; Patterson, J.; Fauhl-Hassek, C.; Donarski, J.; Alewijn, M.; van Ruth, S.; Elliott, C. T. What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case study. Trends in Food Science & Technology 2018, 76, 38–55. https://doi.org/10.1016/j.tifs.2018.04.001

79. Maestrello, V.; Solovyev, P.; Franceschi, P.; Stroppa, A.; Bontempo, L. ¹H-NMR Approach for the Discrimination of PDO Grana Padano Cheese from Non-PDO Cheeses. Foods 2024, 13 (3), 358. https://doi.org/10.3390/foods13030358

80. McSweeney, P. L. H. Biochemistry of cheese ripening. International Journal of Dairy Technology 2004, 57 (2–3), 127–144. https://doi.org/10.1111/j.1471-0307.2004.00147.x

81. McSweeney, P. L. H.; Sousa, M. J. Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. Le Lait 2000, 80 (3), 293–324. https://doi.org/10.1051/lait:2000127

82. Sousa, M. J.; Ardö, Y.; McSweeney, P. L. H. Advances in the study of proteolysis during cheese ripening. International Dairy Journal 2001, 11 (4–7), 327–345. https://doi.org/10.1016/s0958-6946(01)00062-0

83. Collins, Y. F.; McSweeney, P. L. H.; Wilkinson, M. G. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. International Dairy Journal 2003, 13 (11), 841–866. https://doi.org/10.1016/s0958-6946(03)00109-2

84. Gatti, M.; Bottari, B.; Lazzi, C.; Neviani, E.; Mucchetti, G. Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. Journal of Dairy Science 2014, 97 (2), 573–591. https://doi.org/10.3168/jds.2013-7187

85. Masotti, F.; Hogenboom, J. A.; Rosi, V.; De Noni, I.; Pellegrino, L. Proteolysis indices related to cheese ripening and typicalness in PDO Grana Padano cheese. International Dairy Journal 2010, 20 (5), 352–359. https://doi.org/10.1016/j.idairyj.2009.11.020

86. Mucchetti, G.; Locci, F.; Gatti, M.; Neviani, E.; Addeo, F.; Dossena, A.; Marchelli, R. Pyroglutamic Acid in Cheese: Presence, Origin, and Correlation with Ripening Time of Grana Padano Cheese. Journal of Dairy Science 2000, 83 (4), 659–665. https://doi.org/10.3168/jds.s0022-0302(00)74926-5

87. Sforza, S.; Cavatorta, V.; Lambertini, F.; Galaverna, G.; Dossena, A.; Marchelli, R. Cheese peptidomics: A detailed study on the evolution of the oligopeptide fraction in Parmigiano-Reggiano cheese from curd to 24 months of aging. Journal of Dairy Science 2012, 95 (7), 3514–3526. https://doi.org/10.3168/jds.2011-5046

88. Addis, M.; Piredda, G.; Pirisi, A. The use of lamb rennet paste in traditional sheep milk cheese production. Small Ruminant Research 2008, 79 (1), 2–10. https://doi.org/10.1016/j.smallrumres.2008.07.002

89. Park, Y. W.; Juárez, M.; Ramos, M.; Haenlein, G. F. W. Physico-chemical characteristics of goat and sheep milk. Small Ruminant Research 2007, 68 (1–2), 88–113. https://doi.org/10.1016/j.smallrumres.2006.09.013

90. Nudda, A.; McGuire, M. A.; Battacone, G.; Pulina, G. Seasonal Variation in Conjugated Linoleic Acid and Vaccenic Acid in Milk Fat of Sheep and its Transfer to Cheese and Ricotta. Journal of Dairy Science 2005, 88 (4), 1311–1319. https://doi.org/10.3168/jds.s0022-0302(05)72797-1
91. Collomb, M.; Schmid, A.; Sieber, R.; Wechsler, D.; Ryhänen, E.-L. Conjugated linoleic acids in milk fat: Variation and physiological effects. International Dairy Journal 2006, 16 (11), 1347–1361. https://doi.org/10.1016/j.idairyj.2006.06.021
92. Thierry, A.; Deutsch, S.-M.; Falentin, H.; Dalmasso, M.; Cousin, F. J.; Jan, G. New insights into physiology and metabolism of Propionibacterium freudenreichii. International Journal of Food Microbiology 2011, 149 (1), 19–27. https://doi.org/10.1016/j.ijfoodmicro.2011.04.026
93. Fröhlich-Wyder, M.-T.; Bachmann, H.-P.; Casey, M. G. Interaction between propionibacteria and starter / non-starter lactic acid bacteria in Swiss-type cheeses. Le Lait 2002, 82 (1), 1–15. https://doi.org/10.1051/lait:2001001
94. Collomb, M.; Bütikofer, U.; Sieber, R.; Jeangros, B.; Bosset, J.-O. Composition of fatty acids in cow's milk fat produced in the lowlands, mountains and highlands of Switzerland using high-resolution gas chromatography. International Dairy Journal 2002, 12 (8), 649–659. https://doi.org/10.1016/s0958-6946(02)00061-4
95. Hugenholtz, J. Citrate metabolism in lactic acid bacteria. FEMS Microbiology Reviews 1993, 12 (1–3), 165–178. https://doi.org/10.1111/j.1574-6976.1993.tb00017.x
96. Ardö, Y. Flavour formation by amino acid catabolism. Biotechnology Advances 2006, 24 (2), 238–242. https://doi.org/10.1016/j.biotechadv.2005.11.005
97. Hutkins, R. W.; Morris, H. A. Carbohydrate Metabolism by Streptococcus thermophilus: A Review. Journal of Food Protection 1987, 50 (10), 876–884. https://doi.org/10.4315/0362-028x-50.10.876
98. Mukherjee, K. K.; Hutkins, R. W. Isolation of Galactose-Fermenting Thermophilic Cultures and Their Use in the Manufacture of Low Browning Mozzarella Cheese. Journal of Dairy Science 1994, 77 (10), 2839–2849. https://doi.org/10.3168/jds.s0022-0302(94)77223-4
99. Lucey, J. A.; Fox, P. F. Importance of Calcium and Phosphate in Cheese Manufacture: A Review. Journal of Dairy Science 1993, 76 (6), 1714–1724. https://doi.org/10.3168/jds.s0022-0302(93)77504-9
100. Mizuno, R.; Lucey, J. A. Effects of Emulsifying Salts on the Turbidity and Calcium-Phosphate–Protein Interactions in Casein Micelles. Journal of Dairy Science 2005, 88 (9), 3070–3078. https://doi.org/10.3168/jds.s0022-0302(05)72988-x
101. Kilcawley, K. N.; Wilkinson, M. G.; Fox, P. F. Enzyme-modified cheese. International Dairy Journal 1998, 8 (1), 1–10. https://doi.org/10.1016/s0958-6946(98)00010-7
102. Grossmann, L.; McClements, D. J. The science of plant-based foods: Approaches to create nutritious and sustainable plant-based cheese analogs. Trends in Food Science & Technology 2021, 118, 207–229. https://doi.org/10.1016/j.tifs.2021.10.004
103. Bharti, S. K.; Roy, R. Quantitative ¹H NMR spectroscopy. TrAC Trends in Analytical Chemistry 2012, 35, 5–26. https://doi.org/10.1016/j.trac.2012.02.007
104. Pauli, G. F.; Gödecke, T.; Jaki, B. U.; Lankin, D. C. Quantitative ¹H NMR. Development and Potential of an Analytical Method: An Update. Journal of Natural Products 2012, 75 (4), 834–851. https://doi.org/10.1021/np200993k
105. Malz, F.; Jancke, H. Validation of quantitative NMR. Journal of Pharmaceutical and Biomedical Analysis 2005, 38 (5), 813–823. https://doi.org/10.1016/j.jpba.2005.01.043
106. Siew, W. L.; Ng, W. Diglyceride content and composition as indicators of palm oil quality. Journal of the Science of Food and Agriculture 1995, 69 (1), 73–79. https://doi.org/10.1002/jsfa.2740690112
107. Kula, R. J.; Sawyer, D. T.; Chan, S. I.; Finley, C. M. Nuclear Magnetic Resonance Studies of Metal-Ethylenediaminetetraacetic Acid Complexes. Journal of the American Chemical Society 1963, 85 (19), 2930–2936. https://doi.org/10.1021/ja00902a016

108. Han, S.; Ba, Y. Determination of the Concentrations of Metal Cations in Aqueous Solutions Using Proton NMR Spectral Area Integration of the EDTA Complexes. Journal of Solution Chemistry 2004, 33 (3), 301–312. https://doi.org/10.1023/b:josl.0000035362.27111.54

109. Barton, R. H.; Waterman, D.; Bonner, F. W.; Holmes, E.; Clarke, R.; the PROCARDIS Consortium; Nicholson, J. K.; Lindon, J. C. The influence of EDTA and citrate anticoagulant addition to human plasma on information recovery from NMR-based metabolic profiling studies. Molecular BioSystems 2009, 6 (1), 215–224. https://doi.org/10.1039/b907021d

110. Somashekar, B. S.; Ijare, O. B.; Nagana Gowda, G. A.; Ramesh, V.; Gupta, S.; Khetrapal, C. L. Simple pulse-acquire NMR methods for the quantitative analysis of calcium, magnesium and sodium in human serum. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 2006, 65 (2), 254–260. https://doi.org/10.1016/j.saa.2005.10.039

111. Monakhova, Y. B.; Kuballa, T.; Tschiersch, C.; Diehl, B. W. K. Rapid NMR determination of inorganic cations in food matrices: Application to mineral water. Food Chemistry 2017, 221, 1828–1833. https://doi.org/10.1016/j.foodchem.2016.10.095

112. Belloque, J.; De La Fuente, M. A.; Ramos, M. Qualitative and quantitative analysis of phosphorylated compounds in milk by means of ^{31}P-NMR. Journal of Dairy Research 2000, 67 (4), 529–539. https://doi.org/10.1017/s0022029900004477

113. Gobet, M.; Rondeau-Mouro, C.; Buchin, S.; Le Quéré, J.; Guichard, E.; Foucat, L.; Moreau, C. Distribution and mobility of phosphates and sodium ions in cheese by solid-state ^{31}P and double-quantum filtered ^{23}Na NMR spectroscopy. Magnetic Resonance in Chemistry 2010, 48 (4), 297–303. https://doi.org/10.1002/mrc.2576

114. Savorani, F.; Tomasi, G.; Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. Journal of Magnetic Resonance 2010, 202 (2), 190–202. https://doi.org/10.1016/j.jmr.2009.11.012

115. Tomasi, G.; Savorani, F.; Engelsen, S. B. icoshift: An effective tool for the alignment of chromatographic data. Journal of Chromatography A 2011, 1218 (43), 7832–7840. https://doi.org/10.1016/j.chroma.2011.08.086

116. Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. Journal of Chromatography A 1998, 805 (1–2), 17–35. https://doi.org/10.1016/s0021-9673(98)00021-1

117. Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. Journal of Chemometrics 2004, 18 (5), 231–241. https://doi.org/10.1002/cem.859

118. Anderson, P. E.; Mahle, D. A.; Doom, T. E.; Reo, N. V.; DelRaso, N. J.; Raymer, M. L. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. Metabolomics 2011, 7 (2), 179–190. https://doi.org/10.1007/s11306-010-0242-7

119. De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiporkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.; Martins, J. C.; Van Criekinge, W. NMR-Based Characterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. Analytical Chemistry 2008, 80 (10), 3783–3790. https://doi.org/10.1021/ac7025964

120. Sousa, S. A. A.; Magalhães, A.; Ferreira, M. M. C. Optimized bucketing for NMR spectra: Three case studies. Chemometrics and Intelligent Laboratory Systems 2013, 122, 93–102. https://doi.org/10.1016/j.chemolab.2013.01.006

121. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ¹H NMR Metabonomics. Analytical Chemistry 2006, 78 (13), 4281–4290. https://doi.org/10.1021/ac051632c

122. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets. Analytical Chemistry 2006, 78 (7), 2262–2267. https://doi.org/10.1021/ac0519312

123. van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 2006, 7 (1), 142. https://doi.org/10.1186/1471-2164-7-142

124. Emwas, A.-H.; Saccenti, E.; Gao, X.; McKay, R. T.; dos Santos, V. A. P. M.; Roy, R.; Wishart, D. S. Recommended strategies for spectral processing and post-processing of 1D ¹H-NMR data of biofluids with a particular focus on urine. Metabolomics 2018, 14 (3), 31. https://doi.org/10.1007/s11306-018-1321-4

125. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nature Protocols 2007, 2 (11), 2692–2703. https://doi.org/10.1038/nprot.2007.376

126. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 1987, 2 (1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

127. Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2016, 374 (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

128. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 1963, 58 (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

129. van der Maaten, L.; Hinton, G. Visualizing Data Using t-SNE. Journal of Machine Learning Research 2008, 9, 2579–2605. https://jmlr.org/papers/v9/vandermaaten08a.html

130. Kobak, D.; Berens, P. The art of using t-SNE for single-cell transcriptomics. Nature Communications 2019, 10 (1), 5416. https://doi.org/10.1038/s41467-019-13056-x

131. Wattenberg, M.; Viégas, F.; Johnson, I. How to Use t-SNE Effectively. Distill 2016, 1 (10). https://doi.org/10.23915/distill.00002

132. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software 2018, 3 (29), 861. https://doi.org/10.21105/joss.00861

133. Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data fusion methodologies for food and beverage authentication and quality assessment – A review. Analytica Chimica Acta 2015, 891, 1–14. https://doi.org/10.1016/j.aca.2015.04.042

134. Cortes, C.; Vapnik, V. Support-vector networks. Machine Learning 1995, 20 (3), 273–297. https://doi.org/10.1007/bf00994018

135. Barker, M.; Rayens, W. Partial least squares for discrimination. Journal of Chemometrics 2003, 17 (3), 166–173. https://doi.org/10.1002/cem.785

136. Brereton, R. G.; Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. Journal of Chemometrics 2014, 28 (4), 213–225. https://doi.org/10.1002/cem.2609

137. Breiman, L. Random Forests. Machine Learning 2001, 45 (1), 5–32. https://doi.org/10.1023/a:1010933404324

138. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

139. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006, 7 (1), 91. https://doi.org/10.1186/1471-2105-7-91

140. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. Assessment of PLSDA cross validation. Metabolomics 2008, 4 (1), 81–89. https://doi.org/10.1007/s11306-007-0099-6

141. Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 2012, 8 (S1), 3–16. https://doi.org/10.1007/s11306-011-0330-3

142. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A. J. Machine learning algorithm validation with a limited sample size. PLOS ONE 2019, 14 (11), e0224365. https://doi.org/10.1371/journal.pone.0224365

143. Rodionova, O. Y.; Titova, A. V.; Pomerantsev, A. L. Discriminant analysis is an inappropriate method of authentication. TrAC Trends in Analytical Chemistry 2016, 78, 17–22. https://doi.org/10.1016/j.trac.2016.01.010

144. Oliveri, P. Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial. Analytica Chimica Acta 2017, 982, 9–19. https://doi.org/10.1016/j.aca.2017.05.013

145. Pomerantsev, A. L.; Rodionova, O. Y. Concept and role of extreme objects in PCA/SIMCA. Journal of Chemometrics 2014, 28 (5), 429–438. https://doi.org/10.1002/cem.2506

146. De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. L. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems 2000, 50 (1), 1–18. https://doi.org/10.1016/s0169-7439(99)00047-7

147. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 2004, 88 (2), 365–411. https://doi.org/10.1016/s0047-259x(03)00096-4

148. Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; Williamson, R. C. Estimating the Support of a High-Dimensional Distribution. Neural Computation 2001, 13 (7), 1443–1471. https://doi.org/10.1162/089976601750264965

149. Liu, F. T.; Ting, K. M.; Zhou, Z.-H. Isolation Forest. In 2008 Eighth IEEE International Conference on Data Mining; IEEE, 2008; pp 413–422. https://doi.org/10.1109/icdm.2008.17

150. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 2001, 58 (2), 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1

151. Bettencourt da Silva, R.; Ellison, S. L. R., Eds. Eurachem/CITAC Guide: Assessment of Performance and Uncertainty in Qualitative Chemical Analysis, 1st ed.; Eurachem, 2021; ISBN 978-0-948926-39-6. https://www.eurachem.org

152. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. The Journal of Organic Chemistry 1997, 62 (21), 7512–7515. https://doi.org/10.1021/jo971176v

153. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. Organometallics 2010, 29 (9), 2176–2179. https://doi.org/10.1021/om100106e

154. Fan, T. W. M. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. Progress in Nuclear Magnetic Resonance Spectroscopy 1996, 28 (2), 161–219. https://doi.org/10.1016/0079-6565(95)01017-3

155. Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Research 2022, 50 (D1), D622–D631. https://doi.org/10.1093/nar/gkab1062

156. Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. BioMagResBank. Nucleic Acids Research 2007, 36 (Database), D402–D408. https://doi.org/10.1093/nar/gkm957

157. Harris, R. K.; Becker, E. D.; Cabral de Menezes, S. M.; Goodfellow, R.; Granger, P. NMR nomenclature. Nuclear spin properties and conventions for chemical shifts(IUPAC Recommendations 2001). Pure and Applied Chemistry 2001, 73 (11), 1795–1818. https://doi.org/10.1351/pac200173111795

158. Pauli, G. F.; Chen, S.-N.; Simmler, C.; Lankin, D. C.; Gödecke, T.; Jaki, B. U.; Friesen, J. B.; McAlpine, J. B.; Napolitano, J. G. Importance of Purity Evaluation and the Potential of Quantitative ¹H NMR as a Purity Assay. Journal of Medicinal Chemistry 2014, 57 (22), 9220–9231. https://doi.org/10.1021/jm500734a

159. Holzgrabe, U. Quantitative NMR spectroscopy in pharmaceutical applications. Progress in Nuclear Magnetic Resonance Spectroscopy 2010, 57 (2), 229–240. https://doi.org/10.1016/j.pnmrs.2010.05.001

160. Burton, I. W.; Quilliam, M. A.; Walter, J. A. Quantitative ¹H NMR with External Standards: Use in Preparation of Calibration Solutions for Algal Toxins and Other Natural Products. Analytical Chemistry 2005, 77 (10), 3123–3131. https://doi.org/10.1021/ac048385h

161. Alum, M. F.; Shaw, P. A.; Sweatman, B. C.; Ubhi, B. K.; Haselden, J. N.; Connor, S. C. 4,4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. Metabolomics 2008, 4 (2), 122–127. https://doi.org/10.1007/s11306-008-0103-9

162. Kriat, M.; Confort-Gouny, S.; Vion-Dury, J.; Sciaky, M.; Viout, P.; Cozzone, P. J. Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. NMR in Biomedicine 1992, 5 (4), 179–184. https://doi.org/10.1002/nbm.1940050404

163. Wider, G.; Dreier, L. Measuring Protein Concentrations by NMR Spectroscopy. Journal of the American Chemical Society 2006, 128 (8), 2571–2576. https://doi.org/10.1021/ja055336t

164. Akoka, S.; Barantin, L.; Trierweiler, M. Concentration Measurement by Proton NMR Using the ERETIC Method. Analytical Chemistry 1999, 71 (13), 2554–2557. https://doi.org/10.1021/ac981422i

165. Carpio, A.; Rodríguez-Estévez, V.; Sánchez-Rodríguez, M.; Arce, L.; Valcárcel, M. Differentiation of organic goat's milk based on its hippuric acid content as determined by capillary electrophoresis. ELECTROPHORESIS 2010, 31 (13), 2211–2217. https://doi.org/10.1002/elps.200900728

166. Carpio, A.; Bonilla-Valverde, D.; Arce, C.; Rodríguez-Estévez, V.; Sánchez-Rodríguez, M.; Arce, L.; Valcárcel, M. Evaluation of hippuric acid content in goat milk as a marker of feeding regimen. Journal of Dairy Science 2013, 96 (9), 5426–5434. https://doi.org/10.3168/jds.2012-6396

167. O'Callaghan, T.; Vázquez-Fresno, R.; Serra-Cayuela, A.; Dong, E.; Mandal, R.; Hennessy, D.; McAuliffe, S.; Dillon, P.; Wishart, D.; Stanton, C.; Ross, R. Pasture Feeding Changes the Bovine Rumen and Milk Metabolome. Metabolites 2018, 8 (2), 27. https://doi.org/10.3390/metabo8020027

168. Sieber, R.; Bütikofer, U.; Bosset, J. O. Benzoic acid as a natural compound in cultured dairy products and cheese. International Dairy Journal 1995, 5 (3), 227–246. https://doi.org/10.1016/0958-6946(94)00005-a

169. Park, S.-Y.; Yoo, M.-Y.; Paik, H.-D.; Lim, S.-D. Production of benzoic acid as a natural compound in fermented skim milk using commercial cheese starter. Journal of Dairy Science 2017, 100 (6), 4269–4275. https://doi.org/10.3168/jds.2016-12399

170. Hassan, A.; Johnson, M. E.; Lucey, J. A. Changes in the Proportions of Soluble and Insoluble Calcium During the Ripening of Cheddar Cheese. Journal of Dairy Science 2004, 87 (4), 854–862. https://doi.org/10.3168/jds.s0022-0302(04)73229-4

171. Gaucheron, F. The minerals of milk. Reproduction Nutrition Development 2005, 45 (4), 473–483. https://doi.org/10.1051/rnd:2005030

Summary

Paired ¹H NMR spectra (300 MHz) of the $CDCl_3$ and D_2O extracts of 84 cheese products were evaluated as a single screening study, supported by reference fats, amino-acid standards, a gHSQC spectrum, Cs-EDTA titrations and 43 further lipid spectra. The lipid spectra were integrated over regions A to J, with G divided into G_1 and G_2 and C into C_1 and C_2, and the published Tociu and Hanganu–Chira schemes were applied alongside line fitting of the butyrate, n-3, linoleate, sterol, caproleate and conjugated-diene signals. A milk-fat reference population of 24 products gave butyrate at 9.08 ± 0.47 mol % of chains. A fixed decision rule, published with its thresholds, separated substitution from lipolysis, and a dilution model based on butyrate and caproleate estimated the non-milk fat, with markers at the blank level reported as lower bounds.

Twenty-two of 40 retail grated Parmesan and Romano products were flagged for 43 to 68% (median 51%, each ± 7 to 8 points) of a non-milk fat of similar gross composition, consistent with a palm-type fat, and 19 of 43 further spectra, which are lots of five related products, likewise. In 21 of the 22 the aqueous extract showed a low free amino-acid index, citrate, free sugars and dispersed lipid, a profile that resembles

processed cheese; soluble calcium, reportable for six of them, was about one quarter of that of the retail products with milk fat. Sixteen retail products were consistent with grated hard cheese from milk fat. The oil, the manufacturing route, the proportion of cheese and the curing time are not established by these data. Two anomalous products are mirror images of one another and need a repeat extraction, and the records of five further products are in conflict.

Across 292 preprocessing combinations the differences between binning, normalization and scaling options were within the cross-validation uncertainty. With all selection nested inside the folds, supervised models gave balanced accuracies of 0.90 for fat, 0.82 for technology and 0.51 for nine cheese styles; one-class Mahalanobis models on chemically chosen descriptors reproduced the ratiometric categories, as a check of consistency within the data set and not as an independent confirmation. In simulated blends the rule flagged half of the blends at 20 to 25% of added fat and all of them from 30 to 40%; a detection limit for real products remains to be determined with gravimetric blends. The ratiometric and targeted approach was the most informative and the most readily explained. Validation against reference methods, with a separately validated comparison for lipolyzed products, and an internal standard are required before individual results are reported. The aqueous spectrum distinguishes unripened from ripened natural cheese clearly but did not resolve ripening times beyond about ten months in this data set.