Absolute-Intensity ¹H qNMR of Freeze-Dried Spirit Residues

Polar (D_2O/CD_3OD), Lipophilic ($CDCl_3$) and Whole-Residue (DMSO-d_6) Congeners of Whiskeys, Bourbons, Rums, Brandies and Agave Spirits: External-Standard qNMR, Chemometrics and Time-Domain Analysis

White Paper PNA-WP-2026-FDSP-01, Version 2.3

John C. Edwards, Ph.D.

Process NMR Associates, LLC

84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603

October 2026

Document details

Item	Detail
Document	PNA-WP-2026-FDSP-01, Version 2.3 (supersedes Version 2.2); file PNA-WP-2026-FDSP-01_v2.3_Freeze-Dried_Spirits_Absolute_1H_NMR.docx
Revision	Version 2.3 adds reference spectra of 45 pure congeners (58 spectra) and an eleven-step model spirit, compared with an intact spirit and its residues (Sections 2.8–2.11); redraws every stacked spectral figure with one vertical-scaling rule (Section 4.17); adds a second round of PLS calibrations after the removal of flagged and diagnosed outliers (Section 13.5); and adds an Addendum of full-page chemometric plots with every residue named. Version 2.2 corrected two errors of the absolute scale: the 90° pulse is 13 µs (author's calibration), not the 10 µs stored in the parameter sets, so the 8 µs and 7 µs pulses were 55.4° and 48.5° rather than 72° and 63°; and the maleic acid standard contained 95 % of its nominal content (author's determination), so every earlier concentration was 5.3 % too high. Every absolute intensity, saturation correction and concentration was recalculated, with the maleic acid T_1 of 6.1 s (Section 5.9). Version 2.1 added the DMSO-d_6 residues, the pentoses and corrected glucose, and the full chemometric accounting.
Title	Absolute-Intensity ¹H NMR of Freeze-Dried Spirit Residues — Polar (D_2O/CD_3OD), Lipophilic ($CDCl_3$) and Whole-Residue (DMSO-d_6) Congeners of Whiskeys, Bourbons, Rums, Brandies and Agave Spirits: External-Standard qNMR, Chemometrics and Time-Domain Analysis
Author	John C. Edwards, Ph.D., Process NMR Associates, LLC, 84 Patrick Lane, Ste 115, Poughkeepsie, NY 12603
Date	October 23, 2026
Data	Four MestReNova documents of ¹H spectra of freeze-dried residues at 299.99 MHz (first D_2O set, 101 spectra; second D_2O set, 113; first $CDCl_3$ set, 39; second $CDCl_3$ set, 22), one MestReNova document of ¹H spectra of DMSO-d_6 residues (30 titles, 29 acquisitions), one external-standard blank (2.55 mg maleic acid with TMSP in 500 µL D_2O + 150 µL CD_3OD; effective maleic acid content 95 % of the weighed amount, author's determination; maleic acid T_1 6.1 s, author's value; 90° pulse 13 µs, author's calibration) two documents of gHSQC, J-resolved, ¹H and DEPT-45 spectra of a concentrated Kavalan Solist Vinho Barrique residue, and one MestReNova document of reference spectra of pure congeners (58 spectra of 45 compounds in D_2O, D_2O/CD_3OD, D_2O/ethanol, $CDCl_3$ and DMSO-d_6), an eleven-step model spirit and one intact spirit

Item	Detail
Companion files	PNA-WP-2026-FDSP-01_v2.3_companion_data.xlsx (sample list, absolute quantitation in all three solvents, pentoses and corrected glucose, cross-solvent recoveries, chemometric inputs and outputs of every data approach, matched-product comparisons, PLS calibrations of both rounds with every excluded residue and the reason for its exclusion, the reference library of congener shifts, validation and time-domain checks) and PNA-WP-2026-FDSP-01_v2.3_3D_scores.xlsx (three-dimensional scores of every data approach, with the PLS exclusions of each residue)
Software	Python 3.13 (NumPy 2.5, SciPy 1.18, pandas 3.0, scikit-learn 1.9, umap-learn 0.5, Matplotlib 3.11); purpose-written decoder for MestReNova documents; document assembled with docx 9.8 (Node.js)
AI assistance	claude-opus-5-5 (as configured; the serving model may differ), Anthropic

This white paper was prepared with AI assistance. The recovery of the spectra and decays from the MestReNova documents, the absolute scaling, the targeted quantitation, the chemometric and time-domain analyses, the figures and tables and a draft of the text were produced by the AI model named above under the direction of the author, from the author's NMR data. All numerical results in the text are generated by code from those data, and the references were checked against their bibliographic records before use. The interpretation, the conclusions and the responsibility for the content rest with the author.

Abstract

The non-volatile extract of an oak-aged spirit records its maturation, its cask history and any addition made after distillation, but in the native liquid that record lies beneath the ethanol and water signals that set the dynamic range of a ^{1}H NMR experiment. In this work the extract of 1.000 mL of spirit, diluted with 1.000 mL of water and freeze-dried, was redissolved in D_2O or D_2O/CD_3OD for the polar fraction and, from a separate aliquot, in $CDCl_3$ for the lipophilic fraction, and 197 D_2O and 61 $CDCl_3$ spectra of whiskeys, bourbons, rums, brandies and agave spirits were analyzed at 299.99 MHz. The complete acquired free-induction decays and parameter sets were recovered from the MestReNova documents, the processing was reproduced exactly from the decays, a silent normalization of the first data set was reversed, and every spectrum was placed on an absolute scale (micromoles of ^{1}H per liter of spirit) through an external maleic acid standard. The residual CHD_2OD of the CD_3OD, measured on that scale, agreed with the same signal in the standard within 3.4 %, validated the transient, gain and volume corrections, revealed that twenty solvent labels were wrong, and exposed two acquisition sessions in which nine spectra had lost 29–46 % of their sensitivity, which was corrected.

Constrained multiplet fitting with relaxation corrections and detection limits quantified 17 congeners or congener groups in the polar residues and 13 in the lipophilic residues. Cognacs contained 4.7–14.0 g L^{-1} of sugar, mostly as equimolar glucose and fructose; eight of 29 rums contained 5–25.1 g L^{-1} of sucrose or invert sugar; two maple bourbons and a honey whiskey contained 11.8–29.2 g L^{-1} of sucrose; and unsweetened whiskeys contained only 0.04–0.21 g L^{-1} (10th–90th percentiles) of the cell-wall sugars that oak releases. Vanillin without syringaldehyde, accompanied in two cases by 1,2-propanediol, identified flavor preparations in two rums and an American blended whiskey. Quercitol (median 88.2 mg L^{-1} in American whiskeys), the benzaldehydes, gallic acid and the cinnamaldehydes traced oak contact; glycerol, succinic and lactic acid traced wine-cask seasoning; and a broad formyl envelope at δ 9.48 is proposed as a caramel index. A reposado tequila contained 3.54 g L^{-1} of fructose with little glucose, the composition of a fructose-rich syrup of the kind permitted in an *abocado* tequila; the gram-per-liter glycerol that a single-multiplet fit had assigned to the same residue was shown, from the second glycerol multiplet, to be the β-pyranose H-1b doublet of fructose, and glycerol in sugar-rich residues is therefore reported only where both of its multiplets agree. Retained ethanol followed the sugar content of the residue ($\rho = 0.77$), as volatile trapping in a carbohydrate glass predicts. Residues of the same product prepared under different sample numbers differed by up to a factor of 3.1 in non-volatile content, whereas the one true replicate preparation of a single bottle agreed within 9 %, which identifies bottle-to-bottle variation and recovery through freeze-drying as the dominant unvalidated steps.

Pareto-scaled absolute spectra were analyzed by PCA, HCA, t-SNE, UMAP, PLS-DA, SVM and *k*-nearest neighbors with nested, product-grouped cross-validation. Seven spirit categories were classified with a balanced accuracy of 0.78 (linear SVM; permutation $p = 0.005$), falling to 0.67 when whole producers were held out; the embeddings were faithful (trustworthiness ≥ 0.88) but added no class information. PLS calibrations with leave-10 %-out cross-validation gave Q^2 of 0.32–0.76 on a logarithmic scale; masked models showed which calibrations rest on the analyte's own signals, and a first-pass glycerol calibration with $Q^2 = 0.95$ proved to have learned the sugar signals that had contaminated its reference values. Every linear analysis was shown to be identical in the time domain, and band-selective HLSVD on the decays reproduced the frequency-domain quantitation of isolated resonances. The limitations that an expert reviewer would raise, and the procedures that would remove them, are set out explicitly.

Version 2.1 added the 29 DMSO-d_6 residues, reprocessed from their decays and placed on the same absolute scale. Their residual DMSO-d_5 identified two solvent lots, TMS was absent from 16 spectra, the residual solvent line itself was displaced by up to 0.037 ppm in residues rich in sugars and acids, and two label

conflicts were resolved by the chemistry of the residues. HMF, the benzaldehydes and quercitol, quantified from isolated resonances in both polar solvents, were recovered at 1.08, 1.11 and 0.97 times their D_2O/CD_3OD levels in separately prepared residues of the same bottles, which transfers the absolute scale to DMSO-d_6. A single DMSO-d_6 spectrum resolves the individual phenolic aldehydes, ellagic acid, a syringyl singlet assigned tentatively to syringic acid, and the esters and lipids, but glycerol is lost beneath the exchange-averaged water line, the anomeric signals change shape with the rate of hydroxyl exchange, glucose carries an additive interference of about 77 mg L^{-1}, and succinic acid in wine-seasoned residues was recovered at only 0.32 of its D_2O level. The DMSO-d_6 spectra also exposed L-arabinose and D-xylose, overlooked in version 2.0, whose β-arabinopyranose H-1 doublet lies beneath the α-glucose doublet in D_2O/CD_3OD; correcting for that overlap, and refitting the residues whose CD_3OD content had displaced their signals, lowered the median glucose of the American whiskeys from 81.1 to 51.4 mg L^{-1}, beside 57.1 mg L^{-1} of arabinose, and changed the interpretation of four sweetened products, among them a cognac formerly described as fructose-rich that proved to contain ordinary invert sugar. A full accounting of seventeen chemometric data approaches, each shown with principal components, four hierarchical clusterings, t-SNE and UMAP, established that the linkage giving the most faithful tree was never the one giving the most informative partition, that no approach recovered the spirit categories well (adjusted Rand index ≤ 0.30), and that a moisture artifact formed the second principal component of the $CDCl_3$ block. On the matched products DMSO-d_6 reproduced the configuration of the fused two-solvent block (modified RV 0.77) and separated wine-seasoned casks as well as the D_2O/CD_3OD block, but the D_2O/CD_3OD residue remains indispensable for authentication, because it alone measures glycerol, the complete sugar profile and the wine-cask acids; DMSO-d_6 is recommended as the second solvent for fingerprints of oak extraction and $CDCl_3$ where the lipid profile matters.

Version 2.2 corrects two errors of the absolute scale that every earlier version shared. The 90° pulse of the probe is 13 µs rather than the 10 µs stored in the parameter sets, so that the 8 µs and 7 µs pulses delivered 55.4° and 48.5° instead of 72° and 63°, and the maleic acid standard contained 95 % of its nominal content, so that its gravimetric concentration had overstated every result by 5.3 %. With the corrected flip angles, the effective content of the standard and the 6.1 s T_1 of maleic acid, every absolute intensity, saturation correction and concentration was recalculated: the quantified values fell by 1.1–9.5 % (median 4.2 %), the flip angle alone accounting for −4.7 to +4.2 %, the balanced accuracies and the cross-validated Q^2 moved by no more than 0.034 and 0.015, and none of the conclusions changed. Neither correction can be verified by an internal check of these data; both rest on the pulse calibration and on the determination of the content of the standard made by the author.

Version 2.3 adds reference spectra of 45 pure congeners (58 spectra in D_2O, D_2O/CD_3OD, D_2O/ethanol, $CDCl_3$ and DMSO-d_6) and an eleven-step model spirit, and compares them with an intact spirit and with the two residues of the same bottle. The volatile congeners that dominate the native spectrum, ethyl acetate, acetaldehyde, acetic acid and an unassigned singlet at δ 8.217, fell in the residue to at most 1.7 % of complete retention (acetaldehyde to at most 8.0 %), the fusel alcohols and ethanol survived only as trapped traces, and the oak aromatics, at the noise level in the intact spirit, rose by about the 8.4-fold expected from concentration and acquisition. The medium-chain ethyl esters gave strong spectra only in $CDCl_3$, whereas *p*-coumaric acid gave no spectrum in $CDCl_3$ at all, which is the reason for a second solvent. The reference positions agreed with the residues to within about 0.01 ppm for sugars and polyols in matching media, but the syringaldehyde ring protons appeared at δ 7.15–7.23 in every reference medium, 0.08–0.10 ppm upfield of the residue line assigned to them, so that this assignment now rests on its correlation with the formyl line and on its ^{13}C shift rather than on its position. A second round of PLS calibrations, after the removal of 226 flagged and 226 diagnosed outliers from 28 models, showed that most of the apparent improvement came

from no longer scoring the excluded residues; genuine gains on refitting were confined to the oak aromatics and to the logarithmic glycerol and lactic acid models, while the sugar models became worse. Every stacked spectral figure was redrawn with one vertical-scaling rule, and an Addendum gives full-page versions of the chemometric plots with every residue named.

Keywords: ^{1}H NMR; qNMR; external standard; absolute intensity; freeze-drying; whiskey; bourbon; rum; cognac; tequila; oak extractives; quercitol; sugars; arabinose; xylose; glycerol; 1,2-propanediol; lipids; DMSO-d_6; ellagic acid; chemometrics; hierarchical clustering; PLS; outliers; SVM; t-SNE; UMAP; time domain; HLSVD; reference spectra; model spirit

Table of Contents

1. Introduction 9

2. The chemistry of oak-aged spirits and reference spectra of their congeners 12
2.1 Ellagitannins and what ¹H NMR can and cannot see of them 12
2.2 Lignin-derived aldehydes, acids and lignans 12
2.3 Carbohydrate-derived compounds and oak cyclitols 12
2.4 Lactones and lipids 13
2.5 Cask-seasoning carry-over 13
2.6 Additives and the regulatory frame 13
2.7 Agave and cane spirits 14
2.8 A reference library of congener spectra 14
2.9 The intact spirit and what freeze-drying removes 17
2.10 Oak aromatics: solubility, positions and the syringyl pair 20
2.11 Sugars, polyols and non-volatile acids 25

3. Freeze-drying as a sample-preparation step 28
3.1 Why the spirit is diluted before freezing 28
3.2 What is lost and what is retained 28
3.3 Redissolution and the partition of the residue 28

4. Materials and methods 30
4.1 Samples 30
4.2 Sample preparation 31
4.3 NMR acquisition 31
4.4 Recovery of free-induction decays and processed spectra from the MestReNova documents 32
4.5 Reconstruction and verification of the processing 32
4.6 Chemical-shift referencing 33
4.7 Absolute-intensity normalization 34
4.8 Targeted quantitation 34
4.9 Internal consistency checks 37
4.10 Data pretreatment for chemometrics 37
4.11 Chemometric and machine-learning methods 37
4.12 PLS regression calibration 38
4.13 Time-domain analysis 38
4.14 Software and verification of the literature 39
4.15 The DMSO-d_6 residues 39
4.16 Full accounting of the chemometric data approaches and the comparison of solvents 40
4.17 Reference spectra of congeners and the display of spectra 41

5. Validation of the absolute intensity scale 43
5.1 The external maleic acid standard 43
5.2 The residual CHD_2OD as an internal tracer 43
5.3 Receiver-gain linearity 44
5.4 An audit of solvent composition 45
5.5 An acquisition-session anomaly detected by the internal tracer 45
5.6 TMSP and residual $CHCl_3$ 46
5.7 Uncertainty budget 46
5.8 The DMSO-d_6 residues: tracers, referencing and transfer of the absolute scale 47
5.9 Corrections in version 2.2: the flip angle and the content of the standard 50

6. The polar fraction: spectra in D_2O and D_2O/CD_3OD 53
6.1 General appearance on an absolute scale 53
6.2 Assignments 55
6.3 Aldehydes, furanics and phenolic acids 56
6.4 Carbohydrates 59
6.5 Polyols: quercitol, glycerol and 1,2-propanediol 61
6.6 Organic acids 62
6.7 Retained volatiles 62
6.8 Statistical correlation as an assignment aid 63

7. The lipophilic fraction: spectra in $CDCl_3$ 65
7.1 General appearance and assignments 65
7.2 Lipids 67

7.3 Aldehydes and methoxyls 68
7.4 Recovery of the same compounds in the two solvents 68
8. Two-dimensional confirmation 70
9. The whole residue in one solvent: DMSO-d$_6$ 73
9.1 The data set 75
9.2 What the spectra show 75
9.3 Assignments 77
9.4 Hydroxyl exchange and the carbohydrate signals 80
9.5 Pentoses beneath the glucose doublet 81
9.6 Phenolic aldehydes, phenolic acids and ellagic acid 86
9.7 Lipids, esters and organic acids 87
9.8 Quantitation and recovery across solvents 89
9.9 Two label conflicts resolved by chemistry 92
10. Quantitative results 94
10.1 Scope 94
10.2 Replicate preparations: the largest source of uncertainty 96
10.3 Sugars and sweetening 97
10.4 Flavor preparations 99
10.5 Oak markers 100
10.6 Wine-cask carry-over and the agave spirits 100
10.7 Products with cask-seasoning histories that leave a trace 102
11. Chemometrics and machine learning 103
11.1 What absolute intensities change 103
11.2 Principal component analysis 103
11.3 Hierarchical cluster analysis 104
11.4 Supervised classification 105
11.5 How much of the classification is real? 107
11.6 Classification from quantified congeners 107
11.7 Non-linear embeddings 108
11.8 The $CDCl_3$ block and data fusion 108
12. A full accounting of the chemometric data approaches 109
12.1 What the measures mean 109
12.2 The full D_2O/CD_3OD block: an axis of sweetness 110
12.3 The non-carbohydrate D_2O/CD_3OD blocks 112
12.4 Quantified congeners as the variables 119
12.5 The $CDCl_3$ block, a moisture artifact and the fused block 122
12.6 The DMSO-d$_6$ blocks 130
12.7 Hierarchical clustering across all approaches 139
12.8 Linear and non-linear displays 141
12.9 The same products in different solvents 141
12.10 Separation of wine-seasoned casks with one solvent and with two 144
12.11 What the accounting establishes 145
13. PLS regression calibrations 147
13.1 Design and a necessary caution 147
13.2 Calibration statistics 148
13.3 What the masked models reveal 149
13.4 Cross-solvent prediction 149
13.5 Second round: calibrations after the removal of outliers 150
14. Frequency domain and time domain 155
14.1 The question 155
14.2 Linear analyses are identical in the two domains 155
14.3 Where the equivalence breaks 155
14.4 Quantitation in the time domain 156
14.5 Conclusion 157
15. One solvent or two? 158
15.1 What each scheme can see 158
15.2 Agreement of the absolute scale and of the recoveries 159
15.3 What the chemometric comparison shows 159
15.4 The case against replacing the two-solvent scheme 160

15.5 A better single-solvent protocol, and the recommendation 160
16. Critical assessment 162
16.1 What the data establish 162
16.2 What is plausible 162
16.3 What is speculative 163
16.4 What an expert reviewer would challenge 163
17. Recommendations and outlook 165
Appendix A. Samples and acquisition parameters 167
Appendix B. Concentrations by residue 174
Appendix C. Data-integrity log 180
Appendix D. Reconstruction of the processed spectra 182
Appendix E. Companion workbook 183
References 184
Summary 189
Addendum. Full-page chemometric plots 193
Block D2O-F: Figures S1–S8, enlarged from Figure 39 194
Block D2O-NC: Figures S9–S16, enlarged from Figure 40 202
Block D2O-NV: Figures S17–S24, enlarged from Figure 41 210
Block D2O-NC04: Figures S25–S32, enlarged from Figure 42 218
Block D2O-NCall: Figures S33–S40, enlarged from Figure 43 226
Block D2O-Qp: Figures S41–S48, enlarged from Figure 44 234
Block D2O-Qa: Figures S49–S56, enlarged from Figure 45 242
Block CDCl3: Figures S57–S61, enlarged from Figure 46 250
Block CDCl3-W: Figures S62–S66, enlarged from Figure 47 255
Block FUSED-DC: Figures S67–S71, enlarged from Figure 48 260
Block FUSED-W: Figures S72–S76, enlarged from Figure 49 265
Block DMSO-F: Figures S77–S79, enlarged from Figure 50 270
Block DMSO-Fx: Figures S80–S82, enlarged from Figure 51 273
Block DMSO-NC: Figures S83–S85, enlarged from Figure 52 276
Block DMSO-F04: Figures S86–S88, enlarged from Figure 53 279
Block DMSO-Qp: Figures S89–S91, enlarged from Figure 54 282
Block DMSO-Qa: Figures S92–S94, enlarged from Figure 55 285

1. Introduction

The color, much of the flavor and nearly all of the non-volatile mass of a brown spirit are acquired from wood. During maturation the distillate extracts hydrolyzable tannins, lignin-derived aldehydes and acids, lignans, cyclitols, hemicellulose sugars and the thermal degradation products of all of these from the cask, and a cask that previously held wine, sherry, port, Madeira or rum gives up residues of that earlier fill as well [1–3]. None of these constituents is distillable, so they accumulate only by contact with wood or by deliberate addition after distillation, and the composition of the dry extract is for that reason one of the most direct chemical witnesses to how a spirit was made. Questions of authenticity follow naturally. Whether a straight bourbon contains anything other than distillate and water, whether a "sherry cask" malt has in fact seen a sherry-seasoned cask, whether a cognac has been sweetened within the limits of its appellation, whether a rum labeled as carrying no carbohydrate is in fact unsweetened, and whether an aged tequila carries glycerin or syrup are all, at root, questions about the non-volatile fraction [4].

Nuclear magnetic resonance has been applied to distilled spirits mainly in the native state. High-field ^{1}H NMR of Scotch whisky with multiple-solvent suppression has resolved dozens of congeners and supported chemometric classification by cask type and age [5,6], quantitative ^{1}H NMR of whisky congeners has been validated against established chromatographic methods [7], automated eightfold suppression of the water and ethanol resonances has been used to screen alcoholic beverages [8], and ^{1}H NMR fingerprints have been used to screen spirit drinks for quality and authenticity and to recognize counterfeit brands [9,10]. Mezcal has been profiled directly and after solid-phase extraction [11]. In every native measurement, however, the extract is observed at the bottom of a dynamic range that is set by roughly 7 mol L^{-1} of ethanol and 30–40 mol L^{-1} of water. Solvent suppression removes the solvent lines but not the noise, baseline curvature and saturation transfer that they leave behind, and at 300 MHz many extract signals fall beneath the ethanol ^{13}C satellites, the fusel-alcohol multiplets or the suppression notches.

Assignments in work of this kind rest on reference spectra of the pure congeners, recorded in a medium that resembles the sample, and the positions of many of these compounds move with the ethanol content, the acidity, the concentration and the temperature of the solution. The present version adds a library of 58 reference spectra of 45 congeners, recorded in D_2O, D_2O/CD_3OD, D_2O/ethanol, $CDCl_3$ or DMSO-d_6 as their solubility allowed, together with a model spirit built up by sequential addition and one spirit measured intact (Sections 2.8–2.11). The library shows directly what freeze-drying removes and what it keeps, why a residue needs more than one solvent, and how far the positions of pure compounds can be trusted when they are transferred to a residue.

Freeze-drying offers a direct way around that limitation. The ethanol and water are removed, the extract is concentrated, and the residue can be redissolved in a deuterated solvent chosen by the analyst rather than by the bottle [12]. Version 1.0 of this white paper (PNA-WP-2026-FDSP-01) described the chemistry visible in such residues in D_2O, $CDCl_3$ and DMSO-d_6, assigned the principal signals, and showed that cask chemistry rather than geography governs the multivariate structure of a broad commercial sample set. That analysis was necessarily relative. The spectra had been recovered from vector graphics exported by the processing software, so the intensities carried no information on the absolute amount of residue, the acquisition parameters were unavailable, and every spectrum had to be normalized to its own total before it could be compared with any other.

The present version removes that restriction. The original MestReNova documents for all four D_2O and $CDCl_3$ data sets were made available, together with a blank D_2O/CD_3OD solution containing a weighed quantity of maleic acid and a set of two-dimensional spectra of one concentrated residue. Each document was found to contain, for every spectrum, the complete free-induction decay as acquired, the full Varian

parameter set and a losslessly coded copy of the processed spectrum. Because every residue was prepared from a known volume of spirit, the absolute intensity of each spectrum, once corrected for the number of transients, the receiver gain, the flip angle and the solvent volume, is a direct measure of the quantity of each constituent per liter of spirit, and the maleic acid blank converts that measure into moles. The work reported here exploits that property in five ways. The data sets were combined into a single absolute-intensity data base of 197 D_2O and 61 $CDCl_3$ spectra; the congeners that could be identified and resolved were quantified by external-standard qNMR; the full suite of chemometric and machine-learning methods requested (principal component analysis, hierarchical cluster analysis, PLS-DA, support vector machines, k-nearest neighbors, UMAP and t-SNE) was applied to the absolute-intensity spectra; PLS regression models relating the spectra to the quantified concentrations were built and validated by leave-10 %-out cross-validation; and the analysis was repeated in the time domain, directly on the free-induction decays, in order to establish where the two representations are equivalent and where they are not.

Version 2.0 did not repeat the DMSO-d_6 route of version 1.0, and version 2.1 placed those spectra on the same absolute footing. The 29 DMSO-d_6 residues were reprocessed from their stored decays, referenced, checked against their residual DMSO-d_5 and against the D_2O/CD_3OD residues of the same products, and quantified (Sections 4.15, 5.8 and 9), so that the question whether a single solvent can replace the two-solvent scheme can be answered with evidence rather than preference (Section 15). The DMSO-d_6 spectra also exposed two pentoses, L-arabinose and D-xylose, whose anomeric signals had been overlooked in the D_2O/CD_3OD spectra of version 2.0, one of them beneath the α-glucose doublet; the glucose values of version 2.0 have been corrected for that overlap and for a second error, in the residues whose CD_3OD content departed from the nominal value (Section 9.5). The chemometric analysis is given a full accounting in Section 12, in which every data approach, seventeen blocks in all, is shown with its principal components, four hierarchical clusterings, t-SNE and UMAP embeddings and the measures needed to judge them, and in which the products with residues in more than one solvent are compared directly. The T_1 of maleic acid in D_2O was set to 6.1 s, the value supplied by the author (Section 5.1). Where version 1.0 or version 2.0 reached a conclusion that the present data confirm, modify or overturn, the change is stated explicitly, and every data-integrity issue found, with its treatment, is recorded in Appendix C.

Version 2.2 corrected two errors that every earlier version shared. The parameter sets of all the spectra record a 90° pulse of 10 µs, and that value was used to compute the flip angles; the 90° pulse of the probe, as calibrated by the author, is 13 µs, so that the pulses were 55.4° (8 µs) and 48.5° (7 µs) pulses rather than 72° and 63°. The maleic acid standard, moreover, contained 95 % of its nominal content, so that the response factor of the blank had been computed from a concentration 5.3 % too high. Every quantity that depends on either, the response factor, the acquisition factors of the 7 µs spectra and the saturation corrections of all spectra, and with them every absolute intensity and concentration, was recalculated together with the 6.1 s T_1 of maleic acid (Section 5.9), and every number, table and figure that depends on the absolute scale was regenerated; the methods, the analysis and the conclusions were otherwise those of version 2.1.

Version 2.3, the present version, adds three things and changes the presentation of a fourth. The reference spectra of pure congeners, the model spirit and the intact spirit are used in Sections 2.8–2.11 to show how the congeners observed in the residues line up with the pure compounds, which of them survive freeze-drying, and which must be observed in $CDCl_3$ because they do not dissolve in water; one assignment of the earlier versions, the syringaldehyde ring singlet in D_2O/CD_3OD, is qualified as a result. A second round of PLS calibrations, built after the removal of flagged and diagnosed outliers, is reported in Section 13.5 beside the original models, which are retained unchanged. The full-page chemometric figures of Section 12 are complemented by an Addendum in which every score plot, embedding and dendrogram is reproduced at

full page with every residue named. And every stacked or overlaid spectral figure has been redrawn with a single vertical-scaling rule (Section 4.17), so that no spectrum is shown with an expanded baseline whose noise differs from that of its neighbors. The quantitative results of version 2.2 are unchanged.

2. The chemistry of oak-aged spirits and reference spectra of their congeners

2.1 Ellagitannins and what ¹H NMR can and cannot see of them

The hydrolyzable tannins of *Quercus* heartwood are dominated by the C-glycosidic ellagitannins vescalagin and castalagin, epimers at C-1 of an open-chain glucose that carries a hexahydroxydiphenoyl (HHDP) and a nonahydroxyterphenoyl (NHTP) unit. Their structures were revised in 1990, and the configuration at C-1 was re-examined by NMR and computation in 2015 [13]. The dimeric roburins A–D and pentose-bearing derivatives of vescalagin complete the principal monomer and dimer set of oak [14], and sensory-directed fractionation has shown that these compounds, together with grandinin and roburin E, are taste-active in bourbon at the concentrations found there [2]. Their chemistry does not stop at extraction. In ethanolic solution the C-1 position of vescalagin is attacked by ethanol to give β-1-*O*-ethylvescalagin and related derivatives [15], and oxidation of oak ellagitannins in whiskey produces a family of polyphenols whose structures were established by NMR [16]. Hydrolysis of the HHDP esters releases hexahydroxydiphenic acid, which lactonizes spontaneously to ellagic acid; ellagitannin levels in a maturing spirit rise rapidly in the first months and then decline as hydrolysis and oxidation proceed [1].

The ¹H spectrum of a hydrolyzable-tannin mixture in solution is, for the same reason, a set of broad and partially resolved envelopes: galloyl and HHDP aromatic singlets between δ 6.4 and 7.2 and glucose-chain protons between δ 4 and 6, with very little that can be integrated as an individual compound. In a freeze-dried spirit the ellagitannins are present as many related oligomers, each at low concentration and partly oxidized, so that their signals coalesce into a phenolic aromatic envelope under the sharp lines of the small aldehydes and acids. Gallic acid, released by hydrolysis of galloyl esters, is the exception in D_2O, because its two equivalent ring protons give a single sharp line. Ellagic acid gives an equally sharp line but is practically insoluble in D_2O and in $CDCl_3$, and was observed in version 1.0 only in DMSO-d_6. The absolute-intensity data of the present version therefore measure the tannin family in two ways: as discrete gallic acid, and as a phenolic-envelope integral expressed in moles of aromatic protons per liter of spirit, which is an index of oligomeric phenolic material rather than a concentration of any single compound.

2.2 Lignin-derived aldehydes, acids and lignans

Ethanolysis of oak lignin, accelerated by the toasting or charring of the cask, releases the guaiacyl and syringyl aldehydes vanillin and syringaldehyde, the hydroxycinnamaldehydes coniferaldehyde and sinapaldehyde, and the benzoic acids vanillic and syringic acid [1]. New charred oak supplies the cinnamaldehydes in particular abundance, so their formyl doublets are expected to distinguish American straight whiskeys, which must be matured in new charred containers, from spirits matured or finished in refill casks. The aryltetralin lignan lyoniresinol and related lignans are extracted from oak into wine and spirits at taste-active concentrations; lyoniresinol is intensely bitter, and its occurrence in oak-aged wines and spirits has been established [17]. Lignans of this type carry isolated aromatic protons that resonate as singlets between δ 6.3 and 6.6, a region in which several unassigned singlets were observed.

2.3 Carbohydrate-derived compounds and oak cyclitols

The hemicelluloses and cellulose of oak contribute two kinds of signal. Thermal degradation during toasting and charring produces furfural, 5-methylfurfural and 5-hydroxymethylfurfural (HMF), and the ratio of furfural to HMF has been proposed as an authentication criterion for straight whiskey because caramel coloring contributes HMF but very little furfural [4]. Cell-wall sugars are also liberated during aging. The cellulose and hemicellulose of the charred inner layer of bourbon barrels are progressively depleted over a decade in use, which implies a steady release of glucose, xylose and the other wood sugars into the spirit [3], and in brandy matured in oak the contents of arabinose, xylose, glucose and other monosaccharides rise

with the time in wood [18]. The cyclitols of oak, among which quercitol predominates, are extracted with the other polar constituents and have been proposed as fingerprints of the botanical origin of the wood [19,20]. Quercitol is of particular value to NMR because its isolated methylene group gives a pair of well-separated multiplets near δ 1.80 and 1.96 in D_2O, a region that is otherwise sparsely populated in a spirit extract.

2.4 Lactones and lipids

The *cis*- and *trans*-isomers of β-methyl-γ-octalactone (whisky lactone) form in oak from a glycosidic precursor during seasoning and toasting of the wood. The lactones are semi-volatile and are expected to be depleted, though not necessarily eliminated, by freeze-drying. The lipids of a spirit originate in the distillate rather than in the wood: long-chain fatty acids and their ethyl esters (FAEE) carried over from the wash survive in cask-strength and non-chill-filtered products and are removed in part by chill filtration. They are insoluble in water and appear only in the $CDCl_3$ spectra, where their methylene, methyl, α-methylene and ester resonances can be counted by the proton bookkeeping used for edible oils [21].

2.5 Cask-seasoning carry-over

Casks seasoned with sherry, port, Madeira, Moscatel or table wine hold several liters of the previous fill in the wood, and the subsequent spirit extracts its non-volatile components: glycerol, organic acids (succinic, lactic, tartaric and malic), residual sugars in the case of sweet or fortified wines, and esters formed between those acids and ethanol. These compounds have no source in new or ex-bourbon oak, and their presence therefore distinguishes wine-seasoned from non-wine casks, provided that glycerol and sugar have not been added directly. Wine itself has been profiled extensively by ¹H NMR, and the chemical-shift behavior of its organic acids, glycerol and sugars in acidic aqueous solution is well documented [22].

2.6 Additives and the regulatory frame

Post-distillation additions to brown spirits are tightly regulated, and the rules differ in ways that matter for the interpretation of a spectrum (Table 1). In the United States, harmless coloring, flavoring and blending materials (caramel, straight malt or rye malt whiskies, fruit juices, sugar, an approved infusion of oak chips, or wine) may be added to most distilled spirits without changing their class, provided that they total no more than 2.5 % by volume of the finished product; straight whiskeys of every type and bourbon whisky may not contain them at all, whereas a blend of straight whiskies may, American single malt whisky may contain only disclosed caramel coloring, and flavored whiskey forms a separate class [23]. Scotch whisky may contain only water and plain caramel coloring [24]. Whisky in the European Union, which includes Irish whiskey, may be adjusted only with water and plain caramel (E150a) and may not be sweetened, even to round off its taste, or flavored, whereas rum and wine spirit may be sweetened to round off the final taste up to 20 g L^{-1} and brandy up to 35 g L^{-1} of sweetening products, expressed as invert sugar [25]. Cognac, a geographical indication within the wine-spirit category, may in addition be adjusted with caramel, sugars and an infusion of oak chips (boisé), within a combined obscuration of 4 % vol [26]. Canadian whisky may contain caramel and flavoring [27]. Joven, reposado, añejo and extra-añejo tequila may be *abocado*, that is, softened with caramel color, natural oak or holm-oak extract, glycerin or a sugar-based syrup, provided that these ingredients together do not exceed 1 % of the weight of the tequila before bottling, whereas blanco tequila may contain no additive [28]. The detection of glycerol, sugar, a glycol or caramel-derived furanics in a residue is therefore not in itself evidence of non-compliance. It acquires meaning only in combination with the category and label of the product and, where claims such as "additive-free" or "0 g carbohydrate" are made, with a quantitative estimate that can be compared with the regulatory limit and with plausible cask-derived levels. That comparison is precisely what the absolute data of this version make possible.

Table 1. Regulatory frame for post-distillation additions to the spirit categories studied.

Category and jurisdiction	Additions permitted	Limit	Source
Straight whiskey (all types) and bourbon, United States	None; any addition requires redesignation	—	[23]
Other whisky types, including blends of straight whiskies, United States	Harmless coloring, flavoring and blending materials (caramel, sugar, fruit juice, wine, approved oak infusion)	≤ 2.5 % by volume of the finished product	[23]
American single malt whisky	Caramel coloring only, if disclosed on the label	—	[23]
Rum and brandy, United States	Harmless coloring, flavoring and blending materials; for cognac only caramel, oak-chip infusion and sugar without redesignation	≤ 2.5 % by volume	[23]
Scotch whisky	Water and plain caramel coloring	—	[24]
Whisky in the European Union (including Irish whiskey)	Water and plain caramel (E150a); no sweetening or flavoring	—	[25]
Rum, European Union	Sweetening to round off the final taste	≤ 20 g L^{-1} as invert sugar	[25]
Wine spirit, European Union	Sweetening to round off the final taste	≤ 20 g L^{-1} as invert sugar	[25]
Brandy, European Union	Sweetening to round off the final taste	≤ 35 g L^{-1} as invert sugar	[25]
Cognac (geographical indication)	Caramel, sugars and oak-chip infusion (boisé)	combined obscuration ≤ 4 % vol	[26]
Canadian whisky	Caramel and flavoring	—	[27]
Tequila joven, reposado, añejo, extra añejo	*Abocado*: caramel color, natural oak or holm-oak extract, glycerin, sugar-based syrup	≤ 1 % of the weight of the tequila before bottling (all abocantes together; 6.1.1.1)	[28]
Tequila blanco	None	—	[28]

The table summarizes the provisions relevant to the interpretation of a ¹H spectrum; the cited documents are authoritative.

Plain caramel (class I, E150a) is a complex mixture of oligomeric and polymeric dehydration products of sugars, including HMF, difructose anhydrides and colored melanoidin-like polymers [29]. Its ¹H signature in a residue is expected to consist of broad furanic and carbohydrate envelopes with an HMF component rather than of sharp lines, and Section 6.3 shows that the broad formyl envelope near δ 9.48 can be quantified as an index of such material.

2.7 Agave and cane spirits

Blanco and joven agave spirits and unaged rums contain very little non-volatile extract, and what they do contain derives from the cooked raw material, from the still and from additions rather than from wood. Mezcal contains characteristic profiles of alcohols, acids and furanic compounds that differ with agave species and process [11,30]. In aged expressions the oak contribution follows the same chemistry as in whiskey, modulated by the cask history; several of the tequilas examined here had been matured in American white oak or in combinations of ex-bourbon and sherry casks, according to their producers.

2.8 A reference library of congener spectra

A further MestReNova document supplied with this version contains ¹H spectra of 45 pure congeners, 58 spectra in all, recorded at 299.99 MHz in whichever medium each compound would dissolve, together with a clover honey, the eleven steps of a model spirit built up by sequential addition (Section 2.10) and, from the second D_2O data set, one spirit measured intact (Section 2.9). The compounds were chosen to represent the five classes of congener that a matured spirit can carry, and they are grouped in that way in Table 2: ethanol, methanol, acetaldehyde, ethyl acetate and the fusel alcohols (class A, 9 spectra); esters and short- and medium-chain acids (class B, 13); the oak aromatics, phenols and whisky lactone (class C, 16); the hydroxycinnamic, quinic and shikimic acids (class D, 8); and the sugars, polyols and non-volatile acids (class E, 12). The spectra were reprocessed from their decays and referenced as described in Section 4.17; eleven of them, all in D_2O, contain no internal reference and are given at their stored positions.

Table 2. Reference library of congener spectra (300 MHz, 27 °C unless stated): diagnostic chemical shifts measured as multiplet centroids in the reprocessed and referenced spectra, the expected behavior of each compound on freeze-drying, and the signals assigned to it in the residues of this study.

Compound (class)	Solvent	Diagnostic δ ^{1}H (multiplicity, proton)	Expected on freeze-drying	In the residues of this study
ethanol (A)	D_2O	3.649 (q, CH_2); 1.185 (t, CH_3)	volatile; partly retained, in proportion to the residue matrix	retained traces in every residue
methanol (A)	D_2O†	3.320 (s, CH_3)	volatile; removed	not assigned
acetaldehyde (A)	D_2O†	9.670 (q, CHO); 5.226 (q, hydrate CH); 2.232 (d, CH_3); 1.315 (d, hydrate CH_3)	volatile; removed	not assigned
ethyl acetate (A)	D_2O + ethanol	4.142 (q, OCH_2); 2.084 (s, CH_3CO); 1.272 (t, CH_3)	volatile; removed	not assigned
1-propanol (A)	D_2O†	3.531 (t, CH_2O); 1.533 (sext, CH_2); 0.885 (t, CH_3)	volatile; partly retained	within the fusel CH_3 envelope
2-methyl-1-propanol (isobutanol) (A)	D_2O†	3.381 (d, CH_2O); 1.76 (m, CH); 0.898 (d, CH_3)	semi-volatile; partly retained	CH_3 doublet with 3-methyl-1-butanol (D_2O δ 0.884; $CDCl_3$ 0.921)
2-methyl-1-butanol (A)	D_2O/ethanol	3.381 (2 × dd, CH_2O); 0.884 (d + t, CH_3)	semi-volatile; partly retained	within the fusel CH_3 envelope
3-methyl-1-butanol (isoamyl alcohol) (A)	D_2O/ethanol 50:50	1.67 (m, CH); 1.436 (q, CH_2); 0.901 (d, CH_3)	semi-volatile; partly retained	D_2O CH_3, CH and CH_2; $CDCl_3$ CH_2O (δ 3.68)
ethyl lactate (B)	D_2O	4.396 (q, CH); 4.234 (q, OCH_2); 1.413 (d, CH_3 (lactyl)); 1.284 (t, CH_3 (ethyl))	semi-volatile; partly retained	probable: ethyl ester OCH_2 δ 4.22 (D_2O, with ethyl succinates)
	$CDCl_3$	4.25 (m, CH + OCH_2); 1.417 (d, CH_3 (lactyl)); 1.302 (t, CH_3 (ethyl))		
ethyl isobutyrate (B)	$CDCl_3$	4.123 (q, OCH_2); 2.526 (sept, CH); 1.164 (d, CH_3 (isobutyryl))	volatile; removed	not assigned
ethyl hexanoate (B)	D_2O	4.162 (q, OCH_2); 2.377 (t, α-CH_2); 0.879 (t, CH_3)	volatile; removed	not assigned individually
	$CDCl_3$	4.125 (q, OCH_2); 2.284 (t, α-CH_2); 0.898 (t, CH_3)		
ethyl heptanoate (B)	D_2O/ethanol 50:50	4.139 (q, OCH_2); 2.332 (t, α-CH_2); 0.881 (t, CH_3) — ester lines at signal-to-noise ≤ 600	volatile; removed	not assigned individually
	$CDCl_3$	4.124 (q, OCH_2); 2.285 (t, α-CH_2); 0.884 (t, CH_3)		
ethyl decanoate (B)	D_2O/ethanol 50:50	4.140 (q, OCH_2); 2.318 (t, α-CH_2) — ester lines at signal-to-noise ≤ 33	semi-volatile; partly retained	within the ethyl ester signals of the $CDCl_3$ residues
amyl acetate (B)	$CDCl_3$	4.093 (t, OCH_2); 2.039 (s, CH_3CO); 0.924 (d, CH_3)	volatile; removed	not assigned
hexanoic acid (B)	$CDCl_3$	2.347 (t, α-CH_2); 0.904 (t, CH_3)	semi-volatile; partly retained	within the fatty acid α-CH_2 triplet ($CDCl_3$ δ 2.355)
propionic acid (B)	D_2O	2.388 (q, CH_2); 1.088 (t, CH_3)	volatile; largely removed	not assigned
butyric acid (B)	D_2O	2.344 (t, α-CH_2); 1.604 (sext, CH_2); 0.924 (t, CH_3)	volatile; largely removed	not assigned
isobutyric acid (B)	D_2O	2.592 (sept, CH); 1.142 (d, CH_3)	volatile; largely removed	not assigned
vanillin (C)	D_2O/CD_3OD	9.693 (s, CHO); 7.492 (dd, d, H-6, H-2); 7.035 (d, H-5); 3.913 (s, OCH_3)	non-volatile; retained	D_2O, $CDCl_3$ and DMSO-d_6 (formyl and ring protons)
syringaldehyde (C)	D_2O/CD_3OD, 60 °C	9.661 (s, CHO); 7.198 (s, H-2/6); 3.915 (s, OCH_3)	non-volatile; retained	D_2O δ 9.700 and 7.30 (probable; position not reproduced by the references, Section 2.10); $CDCl_3$ 9.824, 7.157; DMSO-d_6 9.778, 7.210
	$CDCl_3$, 50 °C	9.813 (s, CHO); 7.148 (s, H-2/6); 6.028 (s, OH); 3.965 (s, OCH_3)		
vanillic acid (C)	D_2O/CD_3OD	7.557 (dd, d, H-6, H-2); 6.937 (d, H-5); 3.896 (s, OCH_3)	non-volatile; retained	not assigned
syringic acid (C)	D_2O/CD_3OD	7.260 (s, H-2/6); 3.873 (s, OCH_3)	non-volatile; retained	tentative: D_2O δ 7.33 (Section 2.10); DMSO-d_6 δ 7.200
	$CDCl_3$, 50 °C	7.395 (s, H-2/6); 3.955 (s, OCH_3)		
gallic acid (C)	D_2O/CD_3OD	7.148 (s, H-2/6)	non-volatile; retained	D_2O δ 7.10–7.13; DMSO-d_6 δ 6.914
	D_2O + ethanol	7.144 (s, H-2/6)		
scopoletin (C)	D_2O†	7.726 (d, H-4); 6.947 (s, H-5); 6.670 (s, H-8); 6.112 (d, H-3); 3.719 (s, OCH_3)	non-volatile; retained	not assigned
	$CDCl_3$	7.594 (d, H-4); 6.919 (s, H-5); 6.845 (s, H-8); 6.269 (d, H-3); 3.956 (s, OCH_3)		
whisky lactone (cis/trans) (C)	$CDCl_3$	2.673 (dd, C-2 CH_2); 2.200 (dd, C-2 CH_2'); 1.141 (d, CH_3 (isomer 1)); 1.014 (d, CH_3 (isomer 2)); 0.923 (t, butyl CH_3)	semi-volatile; partly retained	not assigned
guaiacol (C)	$CDCl_3$	6.85 (m, ArH); 5.679 (s, OH); 3.851 (s, OCH_3)	volatile; removed	not assigned
eugenol (C)	$CDCl_3$	6.72 (m, ArH); 5.94 (m, =CH); 5.05 (m, $=CH_2$); 5.524 (s, OH); 3.851 (s, OCH_3); 3.310 (d, $ArCH_2$)	semi-volatile; largely removed	not assigned
methyl eugenol (C)	$CDCl_3$	6.74 (m, ArH); 3.854 (2 × s, OCH_3 × 2); 3.330 (d, $ArCH_2$)	volatile; removed	not assigned
2-phenylethanol (C)	$CDCl_3$	7.23 (m, ArH); 3.798 (t, CH_2O); 2.828 (t, $ArCH_2$)	semi-volatile; partly retained	D_2O δ 2.85 (t); $CDCl_3$ δ 3.87, 2.87 (t)
caffeic acid (D)	D_2O	7.628 (d, H-7); 7.201 (d, H-2); 7.132 (dd, H-6); 6.949 (d, H-5); 6.357 (d, H-8)	non-volatile; retained	not assigned

Compound (class)	Solvent	Diagnostic δ ¹H (multiplicity, proton)	Expected on freeze-drying	In the residues of this study
	DMSO-d_6	12.092 (br s, COOH); 9.505 (br s, OH); 7.413 (d, H-7); 7.022 (d, H-2); 6.960 (dd, H-6); 6.759 (d, H-5); 6.168 (d, H-8)		
p-coumaric acid (D)	D_2O	7.695 (d, H-7); 7.582 (d, H-2/6); 6.941 (d, H-3/5); 6.390 (d, H-8)	non-volatile; retained	not assigned
	$CDCl_3$	no analyte signal above 8 × noise (insoluble)		
trans-ferulic acid (D)	DMSO-d_6	7.484 (d, H-7); 7.277 (d, H-2); 7.079 (dd, H-6); 6.788 (d, H-5); 6.359 (d, H-8); 3.814 (s, OCH_3)	non-volatile; retained	not assigned
chlorogenic acid (D)	D_2O†	7.539 (d, H-7); 7.128 (d, H-2); 7.062 (dd, H-6); 6.933 (d, H-5); 6.281 (d, H-8)	non-volatile; retained	not assigned
quinic acid (D)	D_2O	4.18 (m, H-3); 4.04 (m, H-5); 3.561 (dd, H-4); 2.08 (m, H-2/6)	non-volatile; retained	not assigned
shikimic acid (D)	D_2O	6.83 (m, H-2); 4.46 (m, H-3); 4.04 (m, H-5); 3.786 (dd, H-4); 2.749 (dd, H-6)	non-volatile; retained	not assigned
D-glucose (equilibrated) (E)	D_2O†	5.251 (d, α H-1); 3.271 (dd, β H-2)	non-volatile; retained	D_2O (anomeric doublets); DMSO-d_6
D-fructose (E)	D_2O†	4.12 (m, β-furanose H-3/H-4 region); 3.582 (d, β-pyranose H-1a/H-1b)	non-volatile; retained	D_2O
sucrose (E)	D_2O	5.416 (d, Glc H-1); 4.215 (d, Fru H-3); 4.048 (t, Fru H-4)	non-volatile; retained	D_2O; DMSO-d_6
D-galactose (E)	D_2O†	5.279 (d, α H-1)	non-volatile; retained	not assigned
D-mannose (E)	D_2O	5.183 (d, α H-1)	non-volatile; retained	not observed; α-H-1 expected near α-xylopyranose H-1 (Section 2.11)
clover honey (E)	D_2O	5.219 (d, α-Glc H-1)	non-volatile; retained	reference for honey-flavored products
glycerol (E)	D_2O†	3.77 (m, H-2); 3.642 (dd, H-1a/H-3a); 3.569 (dd, H-1b/H-3b)	non-volatile; retained	D_2O (two CH_2 multiplets)
1,2-propanediol (E)	D_2O	3.86 (m, CH); 3.473 (2 × dd, CH_2); 1.126 (d, CH_3)	semi-volatile; largely retained	D_2O (complete spin system)
	$CDCl_3$	1.130 (d, CH_3)		
1,3-propanediol (E)	D_2O	3.684 (t, CH_2O); 1.794 (quint, CH_2)	semi-volatile; largely retained	not sought; quintet within the quercitol H-6ax window (Section 2.11)
lactic acid (E)	D_2O	4.378 (q, CH); 1.415 (d, CH_3)	non-volatile; retained	D_2O
succinic acid (E)	D_2O, 29 °C†	2.617 (s, CH_2)	non-volatile; retained	D_2O; DMSO-d_6

Classes: A, ethanol, methanol, acetaldehyde, ethyl acetate and fusel alcohols; B, esters and short- and medium-chain acids; C, oak aromatics, phenols and lactone; D, hydroxycinnamic, quinic and shikimic acids; E, sugars, polyols and non-volatile acids. Referencing: internal TMSP/DSS or TMS = 0; †no internal reference, positions as stored (about ±0.03 ppm). The reference solutions differ from the residue medium in concentration, CD_3OD or ethanol content, pH and, for the syringyl compounds, temperature; positions in the residues differ by up to 0.1 ppm (Section 2.8). The 50:50 ethanol–water solutions also contain methanol (δ 3.36) and 2-propanol (δ 4.00) from the ethanol used.

Two properties of the collection govern how it can be used. The first is that its media differ from those of the residues. The residues were dissolved in D_2O/CD_3OD (77:23 by volume) or in D_2O, at concentrations of micromoles to millimoles per liter and at whatever acidity their own acids and salts produce, whereas the reference solutions contain a single compound at millimolar concentration in D_2O, D_2O/CD_3OD, D_2O containing ethanol, 50:50 D_2O/ethanol, $CDCl_3$ or DMSO-d_6, and two of the syringyl compounds were measured at 50 and 60 °C. A position transferred from a reference to a residue therefore carries an uncertainty that depends on the type of proton, and Sections 2.10 and 2.11 show that it ranges from about 0.001 ppm for a sugar in a matching medium to 0.1 ppm for the syringyl ring protons. The second is that the library includes, deliberately, compounds that a freeze-dried residue cannot contain. A spectrum of ethyl acetate or acetaldehyde is of no use for assigning a residue, but it shows where these compounds appear in the native spirit, and the comparison of a native spirit with its residue (Section 2.9) shows what freeze-drying removes. The references of the volatile classes A and B are collected in Figure 1.

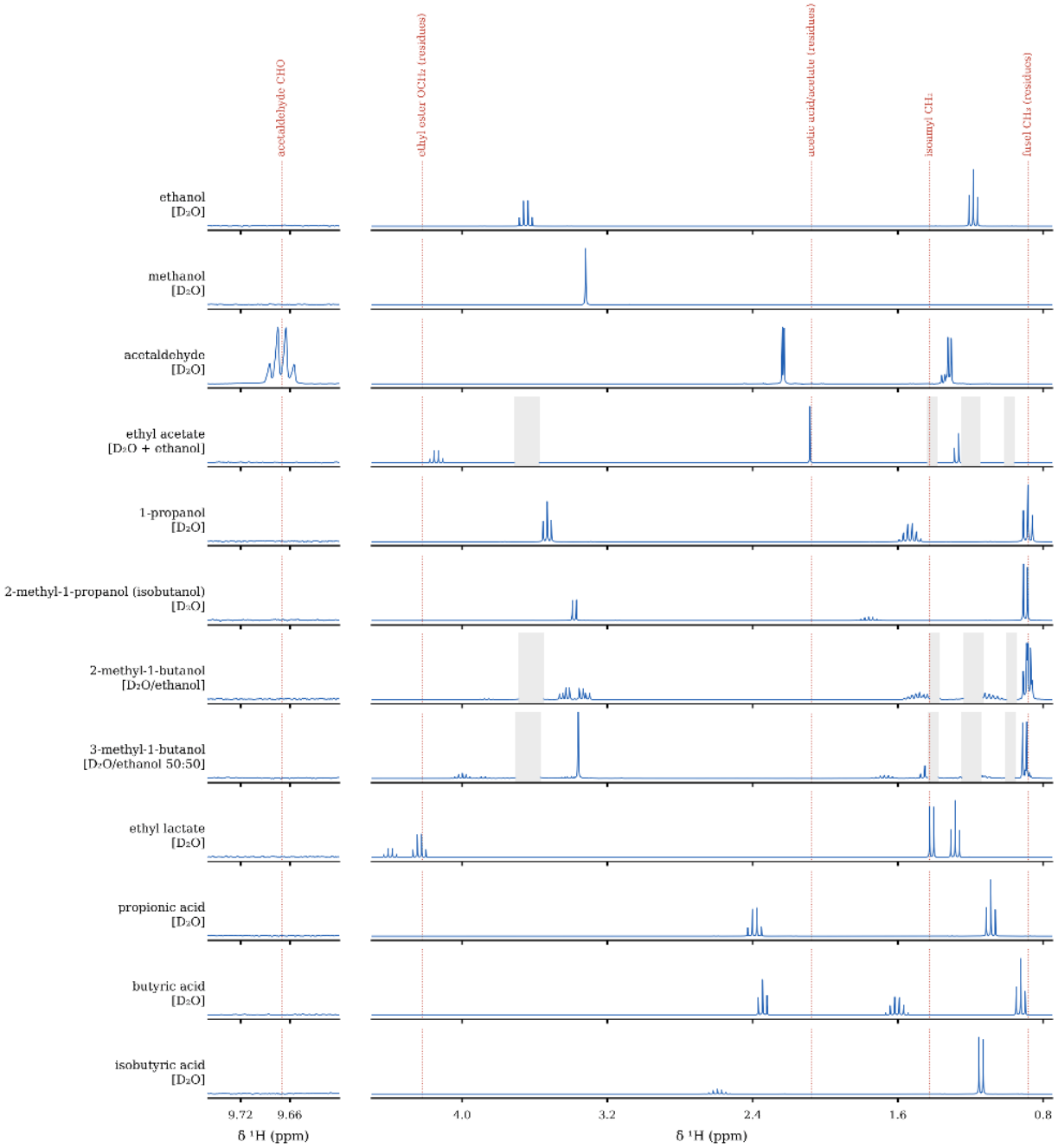

Figure 1. Reference spectra (300 MHz) of the volatile congeners of class A (ethanol, methanol, acetaldehyde, ethyl acetate, 1-propanol and the fusel alcohols) and of ethyl lactate and three short-chain acids of class B, in the medium given in brackets: the acetaldehyde formyl region (left) and δ 0.75–4.50 (right). Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17; the ethanol lines and their ¹³C satellites are blanked (gray) in the solutions that contain ethanol. The singlet at δ 3.36 and the septet near δ 4.00 in the 50:50 ethanol solution are methanol and 2-propanol carried by the ethanol. The dotted lines mark the acetaldehyde formyl quartet and the positions at which the ethyl ester OCH_2, acetic acid and acetate, isoamyl CH_2 and fusel CH_3 signals appear in the freeze-dried residues.

2.9 The intact spirit and what freeze-drying removes

One spirit, Kavalan Distillery Reserve Rum Cask at 57.8 % ABV, was also measured without freeze-drying: 175 µL of the spirit were mixed with 450 µL of D_2O and 50 µL of the maleic acid solution and acquired

with 128 transients, a repetition time of 19.3 s and a receiver gain of 6 dB, the gain being limited by the ethanol and water lines. Its spectrum is compared in Figure 2 with the D_2O/CD_3OD residue of 1 mL of the same bottle (256 transients, 11.3 s, 38 dB), with the $CDCl_3$ residue of a second aliquot, with the final step of the model spirit and with reference spectra of the congeners concerned. Per unit volume of spirit the residue solution was 5.9 times as concentrated as the intact solution, and with twice the number of transients a congener that survived freeze-drying completely should appear in the residue with a signal-to-noise ratio about 8.4 times that in the intact spirit. Ratios of peak signal-to-noise are used here because the intact spectrum was not acquired under the conditions for which the absolute scale was validated; they are approximate, since the line widths of the two solutions differ and the low receiver gain may have cost the intact spectrum some sensitivity, but the differences they reveal are large.

The native spectrum is a spectrum of the volatile congeners. The fusel-alcohol methyl doublets near δ 0.89 (signal-to-noise 2003), the isoamyl alcohol methylene quartet near δ 1.43 (821), the two acetyl singlets near δ 2.07 (1677), which belong to the acetyl methyl of ethyl acetate and to acetic acid, the ethyl acetate OCH_2 quartet at δ 4.125 (249), the acetaldehyde methyl doublet at δ 2.223 (20) and an unassigned singlet at δ 8.217 (40) are all clearly visible, whereas the oak aromatics that carry most of the information of a residue do not rise above the noise: at the positions of the syringyl singlets and of gallic acid the intact spectrum reaches a signal-to-noise ratio of only 7. After freeze-drying the picture is reversed. The syringyl singlets and the gallic acid line reach 31–59, 7–10 times their intact values, which is the gain expected for complete retention within the precision of a comparison with lines at the noise level, whereas the volatile signals collapse. Relative to complete retention, the residue kept about 1.4 % of the fusel-alcohol methyl signal, 0.6 % of the isoamyl methylene and 0.007 % of the ethanol. The ethyl acetate quartet, the acetyl singlets and the δ 8.217 singlet fell to at most 1.7, 0.3 and 1.0 % of their expected intensities, and the acetaldehyde doublet to at most 8.0 %, these being upper limits because the weak residue lines at the same positions need not belong to the same compounds. The δ 8.217 singlet was found in none of the 197 D_2O spectra of the freeze-dried data base. Its position would suit the formyl proton of formic acid or of a formate ester, either of which would be lost on freeze-drying, but no such compound was among the references and the assignment was not tested.

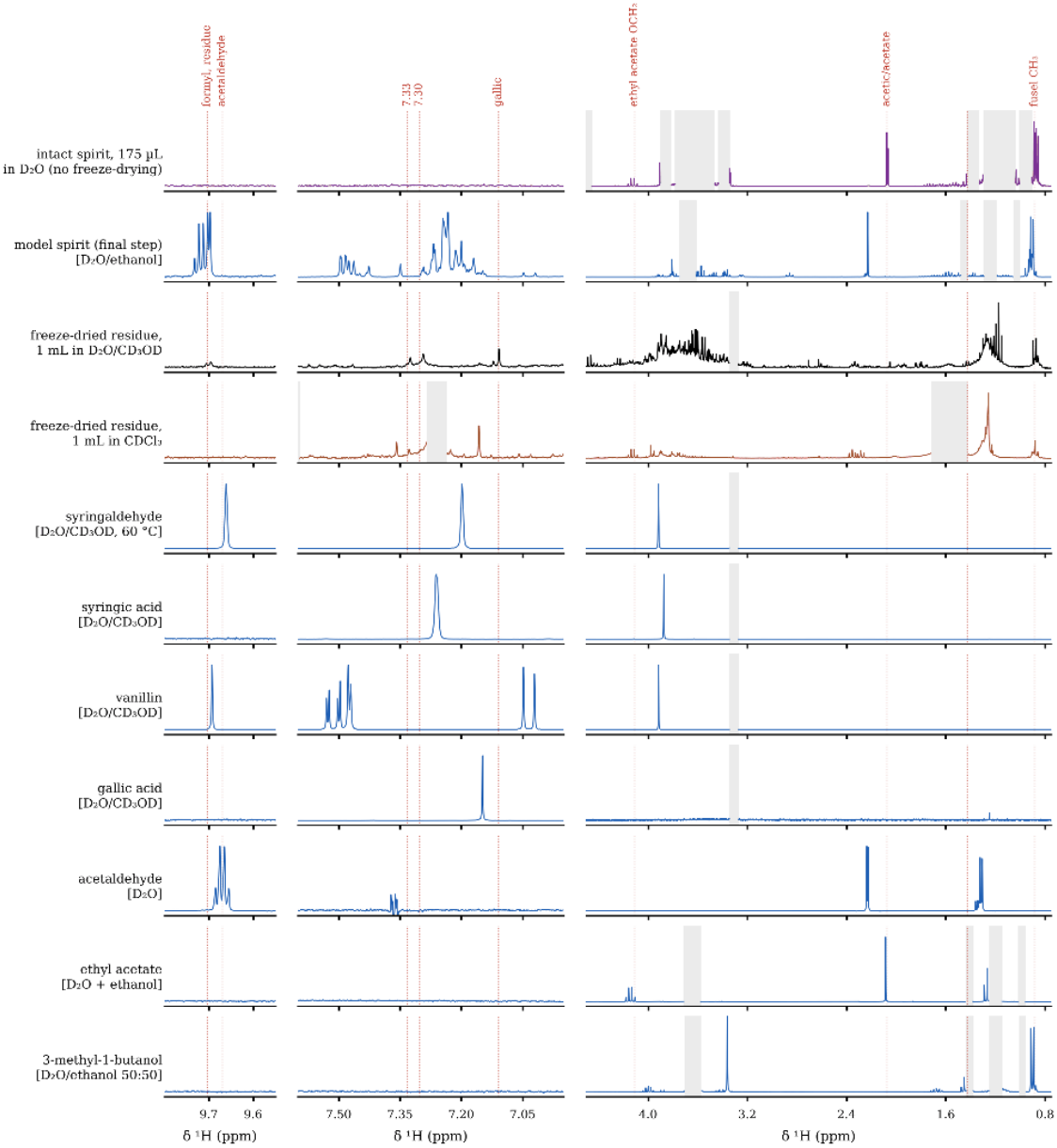

Figure 2. One product from bottle to residue. Kavalan Distillery Reserve Rum Cask (57.8 % ABV) measured intact (175 µL in D_2O with the maleic acid standard, no freeze-drying; purple), the final step of the model spirit, the freeze-dried residue of 1 mL in D_2O/CD_3OD (black) and the residue of a second 1 mL aliquot in $CDCl_3$ (brown), above reference spectra of syringaldehyde, syringic acid, vanillin, gallic acid, acetaldehyde, ethyl acetate and 3-methyl-1-butanol. The dotted lines mark the formyl line of the residue, the acetaldehyde formyl quartet, the syringyl singlets at δ 7.30 and 7.33, gallic acid, the ethyl acetate OCH_2 quartet, acetic acid and acetate, the isoamyl CH_2 and the fusel CH_3. The ethanol, water, standard and solvent windows are blanked (gray). Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17: the formyl and aromatic regions of the intact spirit are flat because its oak aromatics lie at the noise level (signal-to-noise ≤ 7).

Three conclusions follow for the interpretation of residues. Classes A and B of Table 2, apart from the least volatile esters and acids, are not constituents of a freeze-dried residue, and their reference spectra serve to recognize what remains of them rather than to quantify it; the ethanol, fusel alcohols and 2-phenylethanol that remain are trapped in proportion to the residue matrix (Section 3.2) and say nothing quantitative about the spirit. The oak aromatics, conversely, are retained, and the residue raises them from the noise of the native spectrum to lines that can be fitted, by about the factor that concentration and acquisition predict; the

advantage of the residue lies as much in the absence of the ethanol and water lines, which in the native spectrum force a low receiver gain and bury much of the region below δ 4.7 beneath their multiplets and ^{13}C satellites, as in the concentration itself. And the volatile fraction that native spectroscopy measures well, the esters, acetaldehyde, acetic acid and the fusel alcohols, is precisely the fraction that the residue cannot measure, so that the two measurements are complementary rather than alternative: a spectrum of the intact spirit acquired beside the residues would add the volatile congeners to the non-volatile profile at the cost of a few minutes of instrument time.

2.10 Oak aromatics: solubility, positions and the syringyl pair

The oak aromatics of class C and the acids of class D are the compounds for which a residue is prepared. Their reference spectra are shown in Figure 3 in two solvents each, in Figure 4 beneath the D_2O/CD_3OD residue of the same Kavalan spirit, and in Figure 5 as components of the model spirit. Their solubility is the first lesson. Vanillin, vanillic acid, syringic acid and gallic acid dissolved in D_2O/CD_3OD at 27 °C, but syringaldehyde was measured in D_2O/CD_3OD at 60 °C and both syringaldehyde and syringic acid in $CDCl_3$ at 50 °C, temperatures that raise the solubility of these crystalline phenolics; caffeic and ferulic acids were measured in DMSO-d_6, and *p*-coumaric acid, which gave a clean spectrum in D_2O, produced no analyte signal above eight times the noise in $CDCl_3$ (Figure 3a). The medium-chain ethyl esters behaved in the opposite way. Ethyl hexanoate gave lines with a signal-to-noise ratio of about 900 in D_2O against 31 000 in $CDCl_3$, ethyl heptanoate about 600 in 50:50 D_2O/ethanol against 30 000 in $CDCl_3$, and ethyl decanoate only 33 in 50:50 D_2O/ethanol, beside the methanol and 2-propanol carried by the ethanol (Figure 3b); eugenol, methyl eugenol, guaiacol, hexanoic acid and amyl acetate were measured in $CDCl_3$ alone. The pure compounds thus reproduce the partition of a residue between its two solvents (Section 3.3). The phenolic and hydroxycinnamic acids and gallic acid belong to the polar residue and are lost to a chloroform solution, the esters of medium and long chains, eugenol and methyl eugenol dissolve in chloroform and are recovered poorly from water, and the aldehydes, which dissolve in both, are the compounds that allow the two residues to be compared (Section 7.4). In the spirit the ethanol keeps all of them in one phase; once it has been removed, a second solvent is needed to see the second group.

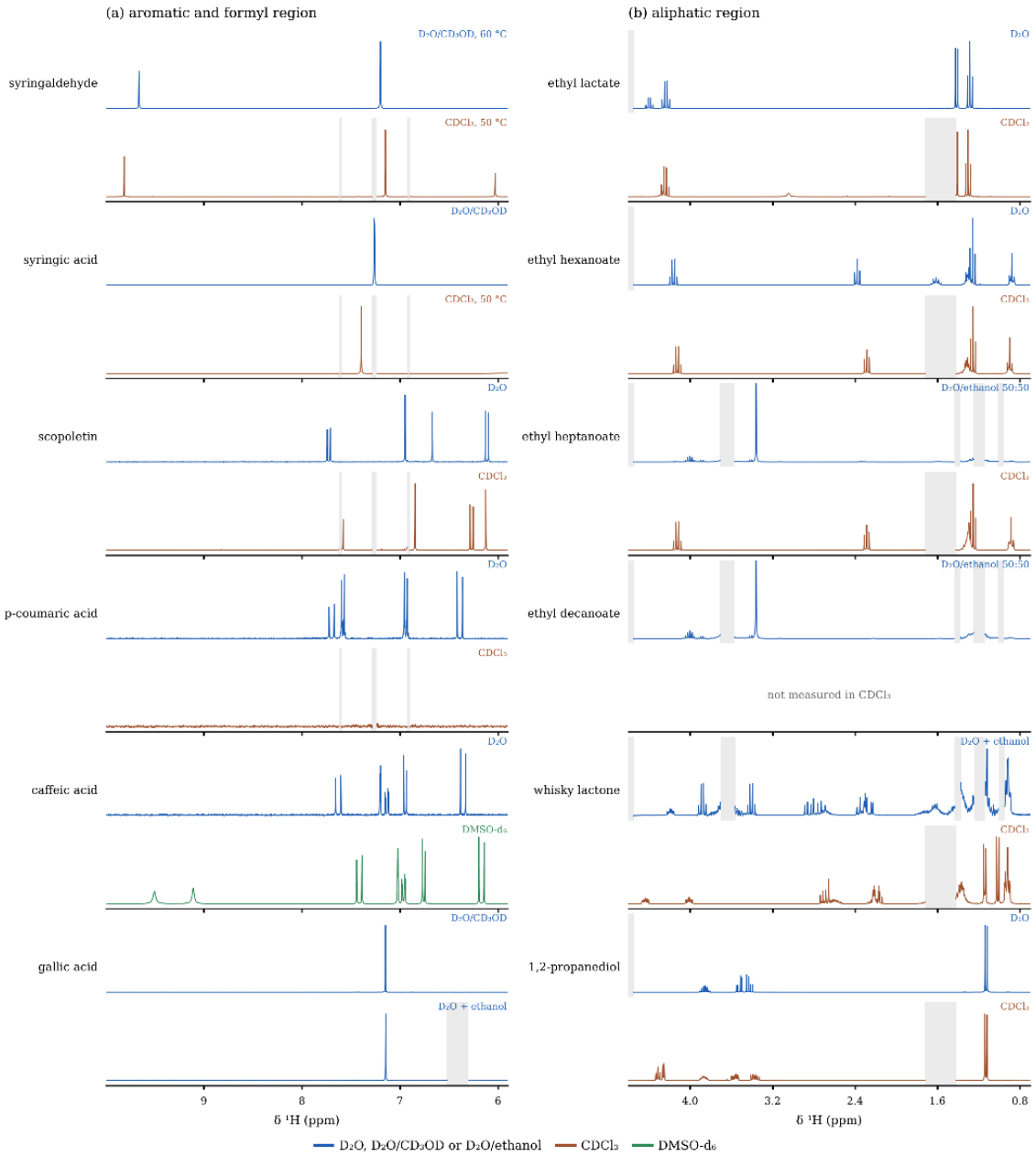

Figure 3. The same compounds in two solvents: (a) the aromatic and formyl region of syringaldehyde, syringic acid, scopoletin, *p*-coumaric acid, caffeic acid and gallic acid and (b) the aliphatic region of ethyl lactate, ethyl hexanoate, ethyl heptanoate, ethyl decanoate, whisky lactone and 1,2-propanediol, each in an aqueous medium (blue) and in $CDCl_3$ (brown) or DMSO-d_6 (green), with the medium and, where it was not 27 °C, the temperature at the right of each trace. *p*-Coumaric acid gave no signal in $CDCl_3$; ethyl heptanoate and ethyl decanoate gave only weak ester lines in 50:50 D_2O/ethanol, beside the methanol singlet of the ethanol; ethyl decanoate was not measured in $CDCl_3$. Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17; the residual solvent, water and ethanol windows are shaded.

Their positions are the second lesson, and they show how far a reference recorded in one medium can be trusted in another. Vanillin behaves as hoped: its formyl singlet at δ 9.693 in D_2O/CD_3OD lies 0.02 ppm upfield of the vanillin formyl line of the oak-aged residues (δ 9.709), and its ring protons coincide with the residue assignments of Table 6 within about 0.03 ppm; the residue medium moves the vanillin formyl line further downfield in the flavored products (δ 9.713–9.722, Section 6.3), so that the order of the two residue

formyl lines is fixed by those products rather than by the reference. Gallic acid appears at δ 7.148 in D_2O/CD_3OD and at 7.144 in D_2O containing ethanol, 0.01–0.03 ppm downfield of the residue line (δ 7.116 in the D_2O/CD_3OD residues and 7.134 in those dissolved in D_2O), a displacement consistent with partial ionization of its carboxyl group in the less acidic residues. The syringyl compounds do not behave so simply. Syringaldehyde H-2/6 appeared at δ 7.198 in D_2O/CD_3OD at 60 °C, 7.228 in the model spirit and 7.148 in $CDCl_3$ at 50 °C, a range of only 0.08 ppm across very different media that also contains the DMSO-d_6 residues (δ 7.210), yet the D_2O/CD_3OD residue line assigned to it in versions 1.0–2.2 lies at δ 7.303, 0.08–0.10 ppm downfield of both aqueous references, and in the 60 °C reference its formyl singlet lies 0.04 ppm upfield of the residue formyl line. Syringic acid, by contrast, moved by 0.09 ppm between two aqueous media, from δ 7.260 in D_2O/CD_3OD to 7.349 in the model spirit, a sensitivity to the medium that is consistent with the ionization of its carboxyl group, and both residue singlets lie within that range.

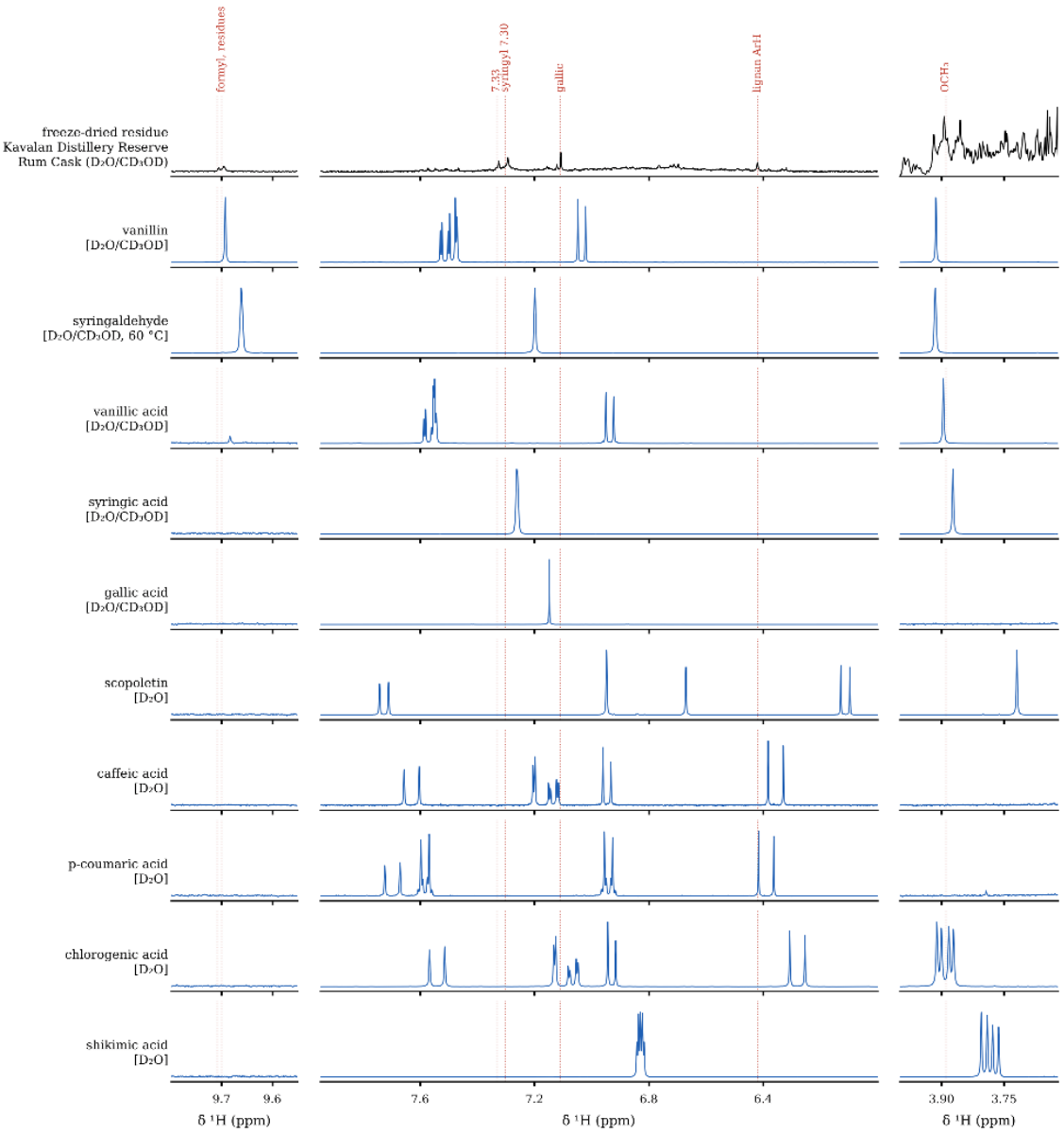

Figure 4. Oak aromatics and hydroxycinnamic, quinic and shikimic acids in aqueous media, beneath the D_2O/CD_3OD residue of a spirit rich in oak extract (Kavalan Distillery Reserve Rum Cask, top): the formyl region (left), the aromatic region (center) and the methoxy region (right). The dotted lines mark the positions measured in the D_2O/CD_3OD residues: the formyl lines of syringaldehyde and vanillin (δ 9.700 and 9.709), the syringyl singlets at δ 7.303 and 7.332, gallic acid, the lignan singlet near δ 6.42 and the aromatic OCH_3. Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17. The spectra of scopoletin and chlorogenic acid carry no internal reference (positions as stored).

The two syringyl singlets of the residues, at δ 7.303 and 7.332, were therefore re-examined. Both were detected above five times the noise in 101 of the 183 calibration residues, where two-Lorentzian fits gave a median ratio of the δ 7.33 line to the δ 7.30 line of 0.47 (interquartile range 0.37–0.66), and both co-varied with the formyl line at δ 9.700, the δ 7.30 line more closely (Spearman ρ = 0.65 against 0.53). Three observations bear on their assignment. In every aqueous medium examined, syringic acid lies downfield of syringaldehyde, by 0.06 ppm in D_2O/CD_3OD (with syringaldehyde at 60 °C) and 0.12 ppm in the model spirit, which is the order of the residue pair. The ratio of the two residue lines equals the ratio of the syringyl singlet at δ 7.200 to the syringaldehyde singlet at δ 7.210 in the DMSO-d_6 residues (0.47, interquartile range

0.42–0.50, 27 residues), the singlet that version 2.1 assigned tentatively to syringic acid. And the δ 7.30 line correlates with the δ 9.692 formyl line more strongly than any other aromatic line (r = 0.95, Figure 18a) and carries the ¹³C shift of a syringyl C-2/6 (δC 107.6), but at 1.46 rather than 2 protons per formyl proton (Section 6.3) and at a position that none of the reference spectra of syringaldehyde reproduces. The assignment of the δ 7.30 line to syringaldehyde H-2/6 is therefore retained as probable, on the strength of the correlation and the ¹³C shift, and the δ 7.33 line is assigned tentatively to syringic acid (Table 6). Unless the residue medium displaces the syringaldehyde ring protons downfield by about 0.08 ppm while it leaves those of vanillin within 0.03 ppm of their reference positions, which nothing observed here explains, part or all of the δ 7.30 line belongs to another syringyl compound. Both assignments can be settled by adding syringaldehyde and syringic acid to a residue (Section 17), and until then the δ 7.30 line is treated as what it demonstrably is, an index of syringyl material.

The remaining aromatics of the library were not found in the residues, and their spectra mark where they would appear. Scopoletin, the coumarin of oak, gives its H-4 and H-3 doublets (J = 9.4 Hz) at δ 7.726 and 6.11 in D_2O; no doublet pair with that coupling was assigned in the residues, and weak lines near δ 7.70 in some of them would have to be shown to have a partner near δ 6.1 before scopoletin could be claimed. The hydroxycinnamic acids give pairs of vinyl doublets with $J \approx 16.0$ Hz, H-7 at δ 7.54–7.70 and H-8 at δ 6.28–6.39 in D_2O, and their H-8 doublets fall in the region of the singlets near δ 6.42 that are assigned tentatively to syringyl lignans. The residue line there is a singlet, so it is not a hydroxycinnamic acid, although weak lines near δ 7.55–7.65 and 6.32–6.38 in the residues richest in oak extract could hide such acids at low levels. Shikimic acid (H-2 at δ 6.83), quinic acid and chlorogenic acid, the caffeoyl ester of quinic acid that is characteristic of coffee rather than of oak, were not observed. In the model spirit (Figure 5) the lines of the compounds already present moved by no more than 0.02 ppm as each further compound was added, so that interactions among the congeners themselves account for only a small part of the displacements between the reference solutions and the residues; the base solution, recorded as containing an acetal, shows the formyl quartet of free acetaldehyde near δ 9.70.

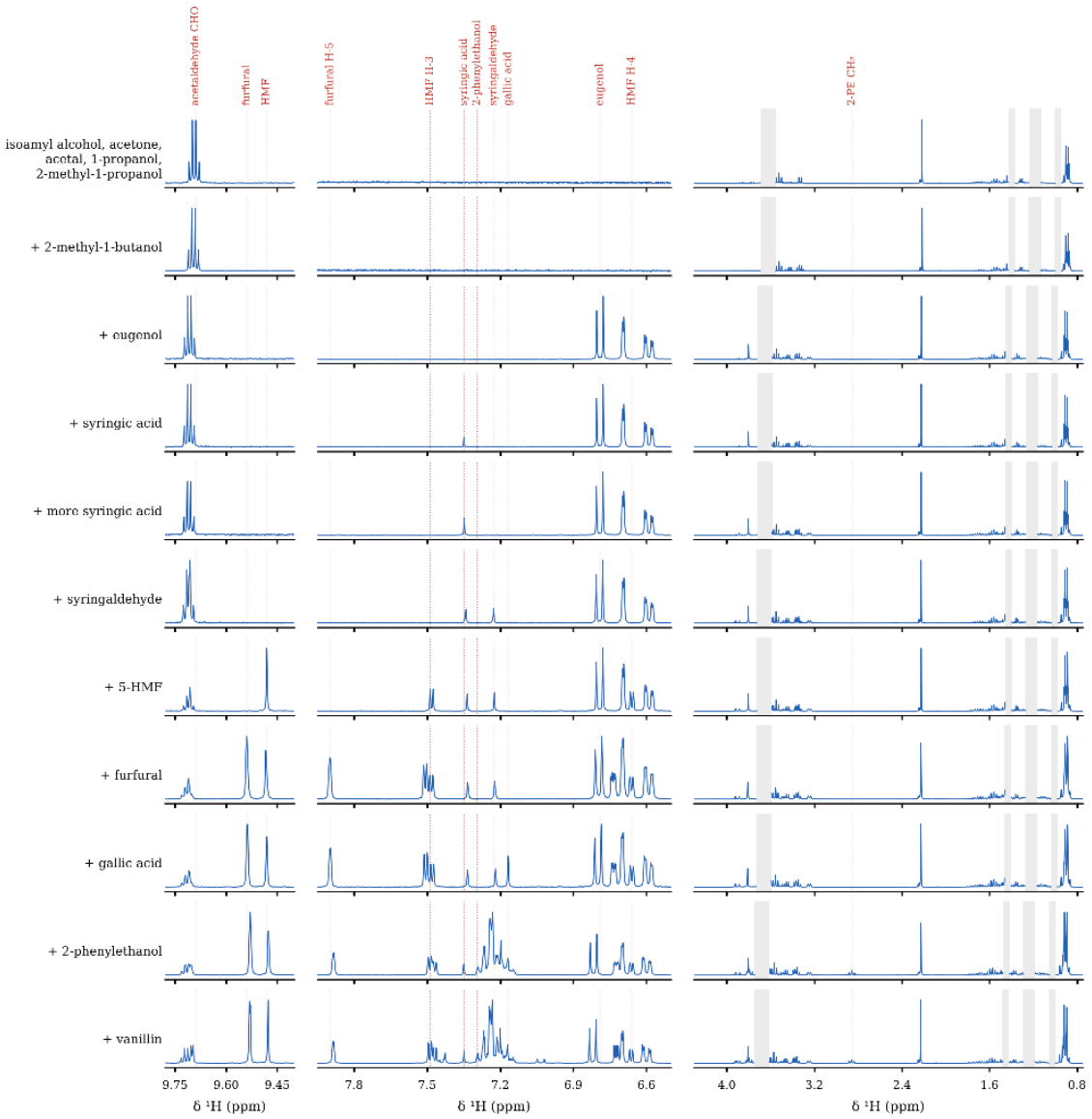

Figure 5. The sequential model spirit in D₂O/ethanol: a base solution recorded as 3-methyl-1-butanol, acetone, acetal, 1-propanol and 2-methyl-1-propanol (top), whose formyl quartet near δ 9.70 shows that it contains acetaldehyde, to which 2-methyl-1-butanol, eugenol, syringic acid (twice), syringaldehyde, HMF, furfural, gallic acid, 2-phenylethanol and vanillin were added in turn. The formyl (left), aromatic (center) and aliphatic (right) regions are shown with the positions of the added compounds (dotted lines); the ethanol lines and their ¹³C satellites are blanked. Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17.

2.11 Sugars, polyols and non-volatile acids

The polar non-volatiles of class E, the compounds on which the findings of sweetening, glycerin addition and wine seasoning in this paper rest, are compared in Figure 6 with the D₂O/CD₃OD residue of a malt finished in a Moscatel cask, which carries gram-per-liter amounts of glucose and fructose together with glycerol, succinic acid and lactic acid. For these compounds the transfer of positions from the references works well once the medium is matched. The referenced spectrum of the clover honey places α-glucose H-1 at δ 5.219 and the 14 residues dissolved in D₂O alone place it at δ 5.220, and the methyl doublet of 1,2-propanediol appears at δ 1.126 in the reference and at 1.131 in the residues dissolved in D₂O. The D₂O/CD₃OD residues place the anomeric doublets 0.028 ppm further upfield (α-glucose at δ 5.192), the

medium displacement described in Section 9.5, and a reference recorded in D_2O must be moved by that amount before it is compared with them. The spectra recorded without an internal reference are less reliable: the stored α-glucose doublet of the pure glucose lies at δ 5.251, 0.06 ppm downfield of the D_2O/CD_3OD residues and 0.03 ppm downfield of the referenced honey, which is the size of error that the positions marked as stored in Table 2 can carry. Lactic acid shows the effect of acidity. The 88 % solution, in which the acid is not ionized and part of it is present as lactoyl lactate, whose methine quartets appear near δ 5.1–5.2, gives the methyl doublet at δ 1.415 and the methine quartet at 4.378, whereas the residues, in which part of the acid is present as lactate, place the doublet at δ 1.365.

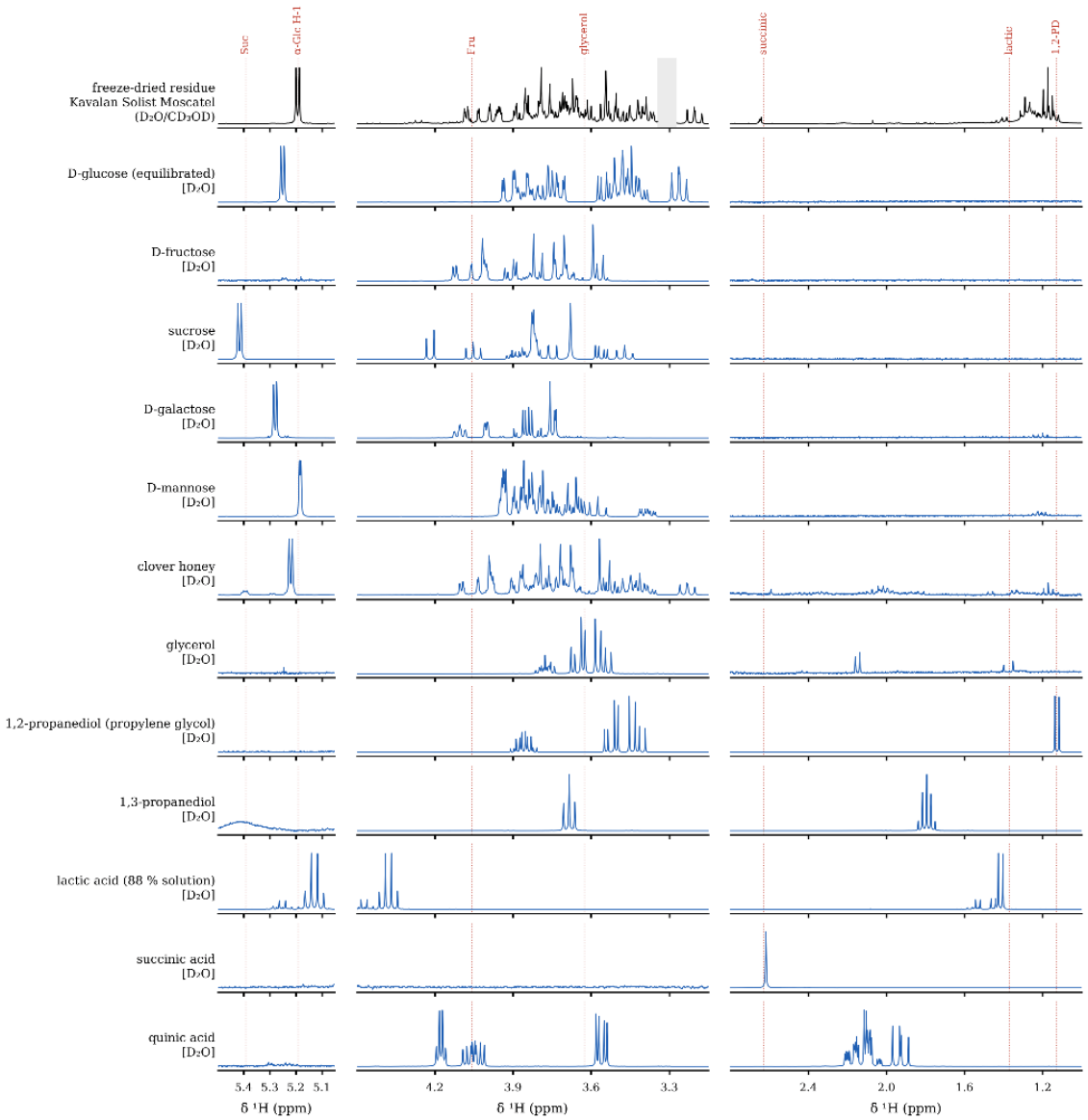

Figure 6. Sugars, polyols and non-volatile acids in D_2O beneath the D_2O/CD_3OD residue of a malt seasoned in a sweet-wine cask (Kavalan Solist Moscatel, top): the anomeric region (left), the CH–O region (center) and the aliphatic region (right). The dotted lines mark the positions measured in the residues (sucrose and α-glucose H-1, fructose, glycerol, succinic acid, lactic acid and 1,2-propanediol). The spectra of glucose, fructose, galactose, glycerol and succinic acid carry no internal reference (positions as stored), and the α-glucose doublet of the glucose spectrum lies 0.06 ppm downfield of its position in the residues; the referenced sucrose spectrum lies 0.02 ppm downfield, the effect of the residue medium. The lactic acid reference is the commercial 88 % solution, whose quartets near δ 5.1–5.2 belong to lactoyl lactate. Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17.

The library also exposes overlaps that the residue spectra alone do not reveal. 1,3-Propanediol, a bacterial metabolite of glycerol, gives a quintet at δ 1.794 (J = 6.5 Hz) that falls entirely within the quercitol H-6ax multiplet used for quantitation (δ 1.797), and a triplet at δ 3.68 that is hidden beneath the ethanol and carbohydrate signals; a residue containing it would report too much quercitol, and the residues were not tested for it. D-Mannose places its α-H-1 doublet, with a coupling of only 1.6 Hz, at δ 5.183 in D_2O, which after the medium displacement of the D_2O/CD_3OD residues would fall near the α-xylopyranose doublet at δ 5.151 used for xylose. And the overlap that misled the single-multiplet glycerol fit of the first-pass quantitation (Section 6.5) is plain in the reference spectra, where the β-pyranose H-1 doublet of fructose (δ 3.582) falls on the glycerol H-1b/H-3b multiplet (δ 3.569, as stored). Succinic acid gives its singlet at δ 2.617 in D_2O (as stored), close to its residue positions (2.637 in D_2O and 2.628 in D_2O/CD_3OD), and galactose, mannose, quinic acid and shikimic acid were not observed in the residues.

Taken together, the reference spectra confirm the assignments of Table 6 that rest on complete spin systems (the sugars, glycerol, 1,2-propanediol, lactic acid and succinic acid), support those that rest on single lines of known behavior (gallic acid and vanillin), and expose one, the syringaldehyde ring singlet, whose position they do not reproduce. They were recorded under conditions chosen for the pure compounds rather than for the residues, and the recommendation that follows from them, to record the decisive references in the medium of the residues and to add them to a residue, is made in Section 17.

3. Freeze-drying as a sample-preparation step

3.1 Why the spirit is diluted before freezing

A spirit at 40–60 % ABV cannot be freeze-dried in the usual sense. The solid–liquid phase diagram of the ethanol–water system places the onset of freezing of a 40 % ABV solution well below that of a 20 % ABV solution, and at high ethanol content ice forms only as a partial phase in equilibrium with an ethanol-enriched liquid. Under vacuum that liquid boils rather than sublimes, the cake collapses and the extract spatters. A 1:1 dilution with water lowers the alcohol content to roughly 20–30 % ABV, raises the freezing onset to within the reach of an ordinary shelf freezer and produces a firmer ice matrix from which ethanol and water can be removed with less loss of solids. The procedure used here followed that logic: 1.000 mL of spirit was mixed with 1.000 mL of water, frozen and freeze-dried, and the entire residue was redissolved for NMR. For three products a larger aliquot was used (2.000 mL for Maker's Mark No. 46 and 3.000 mL for Basil Hayden Toast and Ron del Barrilito in D_2O), and the absolute data were normalized to the volume actually dried.

3.2 What is lost and what is retained

In principle, freeze-drying removes everything with an appreciable vapor pressure at the shelf temperature and chamber pressure: ethanol, water, acetaldehyde, ethyl acetate and the light esters, the fusel alcohols, furfural and most of the oak volatiles. In practice, retention is neither zero nor uniform. Organic volatiles are retained in freeze-dried carbohydrate solutions far above the level predicted from their vapor pressures, because the volatile molecules are trapped in micro-regions of the amorphous solute matrix that become impermeable once the matrix has dried below a critical water content; retention depends on the solute concentration, the freezing rate and the drying conditions as well as on the identity of the volatile [31]. A spirit residue rich in sugar or glycerol is exactly such a matrix.

The absolute data confirm and quantify that behavior. Every D_2O spectrum retains the ethanol triplet and quartet and the methyl doublets of isoamyl and isobutyl alcohol, and many retain the methylene triplets of 2-phenylethanol; the quantities are reported in Section 6.7. Retained ethanol ranged over more than three orders of magnitude and followed the sugar content of the residue with a rank correlation of 0.77, exactly as the trapping mechanism predicts. These signals are artifacts of incomplete removal rather than compositional variables, and the ethanol methyl window was excluded from every chemometric model. The semi-volatile congeners that are partly retained (fusel alcohols, 2-phenylethanol, acetic acid and the lactones) are retained to a degree that depends on the residue matrix, so their concentrations cannot be compared between spirits of different extract composition and must not be read as concentrations in the original spirit. The comparison of an intact spirit with its residue in Section 2.9 puts numbers on these statements for one product: relative to complete retention, the residue kept about 1.4 % of the fusel-alcohol methyl signal and 0.007 % of the ethanol, while ethyl acetate, acetic acid and an unassigned volatile singlet at δ 8.217 fell to at most 1.7, 0.3 and 1.0 %.

3.3 Redissolution and the partition of the residue

A freeze-dried residue is a mixture of very polar species (sugars, polyols, organic acids and gallic acid), amphiphilic phenolics (the aldehydes, phenolic acids and lignans) and hydrophobic lipids. In the hydroalcoholic spirit these coexist because the ethanol–water mixture dissolves all of them, an effect that rests on the hydrogen-bonded structure of the mixture and that is itself modified by the solutes extracted from the wood [32]. Once the solvent is removed, no single solvent other than a strongly hydrogen-bond-accepting aprotic solvent redissolves everything. The residue was therefore prepared twice, from two separate aliquots of the same spirit: one residue was dissolved in D_2O or in a D_2O/CD_3OD mixture to capture

the polar fraction, and the other in $CDCl_3$ to capture the lipophilic fraction. Because each preparation began with the same volume of the same spirit, a constituent that dissolves completely in both solvents should be found at the same absolute concentration in both spectra. The phenolic aldehydes, which are soluble in water at the milligram-per-liter level and freely soluble in chloroform, and the fusel alcohols provide exactly such a test, and it is applied in Section 7.4. The reference spectra of pure congeners show the same partition from the side of the pure compounds (Section 2.10): the medium-chain ethyl esters gave strong spectra only in $CDCl_3$, and hexanoic acid, eugenol and methyl eugenol were measured in it, whereas *p*-coumaric acid gave no spectrum in $CDCl_3$ and the hydroxycinnamic acids were measured in D_2O or DMSO-d_6.

Each solvent carries its own artifacts and blind spots. In D_2O/CD_3OD the HDO line (δ 4.6–4.9, depending on composition and temperature) obscures part of the anomeric region, and the residual CHD_2OD quintet at δ 3.30 and its ^{13}C satellites at δ 3.07 and 3.54 overlap sugar ring protons and glycerol. In $CDCl_3$ the residual $CHCl_3$ line and its ^{13}C satellites at δ 6.91 and 7.61, dissolved water near δ 1.5–1.6 and an adventitious CH_2Cl_2 line at δ 5.30 must be excluded [33]. Exchangeable protons (OH and COOH) are invisible in D_2O and broadened in $CDCl_3$.

4. Materials and methods

4.1 Samples

The combined data base comprises four MestReNova documents acquired between December 2024 and March 2026: a first D_2O set of 101 spectra and a second of 113, a first $CDCl_3$ set of 39 spectra and a second of 22, together with one blank calibration spectrum. A byte-level comparison of every free-induction decay showed that the first sixteen spectra of the second D_2O set were identical copies of spectra already present in the first set; these were counted once, and the copy carried in the second set was retained because one of the first-set copies (Bower Hill Barrel Strength) had been mis-referenced in the processing software by 0.029 ppm. One further D_2O spectrum (Kavalan Distillery Reserve Rum Cask, 175 µL of native spirit with 450 µL of D_2O and maleic acid) was a native rather than a freeze-dried preparation and was excluded from the freeze-dried data base. The D_2O spectrum filed under sample ID Kavalan-010 carries the title of a Bespoken Spirits rye, but its sample ID, its DMSO-d_6 counterpart and its chemistry identify it as Kavalan Solist Brandy, as established in version 1.0; it is treated as such throughout. After these corrections the data base held 197 freeze-dried D_2O spectra of 193 products and 61 $CDCl_3$ spectra, 57 of which have a D_2O counterpart prepared from a separate aliquot of the same bottle (Table 29). The solvent system recorded in the title of each D_2O spectrum was checked against its residual CHD_2OD signal (Section 5.4); twenty titles were found to be incorrect and the corrected assignment, 183 residues in D_2O/CD_3OD and 14 in D_2O alone, is used throughout.

The products span American straight, bottled-in-bond, barrel-proof, blended and flavored whiskeys from large Kentucky, Indiana and Tennessee producers and from craft distilleries in New York, Connecticut, Colorado, Texas and elsewhere; Irish single pot still, single malt and blended whiskeys; Scotch single malts and blends; Canadian whiskies; Japanese, Indian, Australian, French, Danish and Swedish whiskies; twenty-seven Kavalan single malts from Taiwan matured in ex-bourbon, virgin-oak, peated, rum, brandy and wine-seasoned casks (fino, manzanilla, amontillado, oloroso, Moscatel, port, Madeira and table wine); cognacs and an armagnac; rums from the Caribbean, Central America and the United States; añejo and extra-añejo tequilas, a mezcal and two sotols; and a navy-strength gin finished in a sherry-type cask. The workflow from bottle to result is summarized in Figure 7. Product facts (cask type, age and alcohol content) were taken from the spectrum titles and checked against producer documentation where it was available. Two title errors were noted and are recorded in the data-integrity log (Appendix C): Maker's Mark 101 and Maker's Mark Cask Strength are titled "single malt" although both are wheated bourbons, and the Planteray Isle of Fiji sample name records 3 mL of spirit whereas its title records 1 mL.

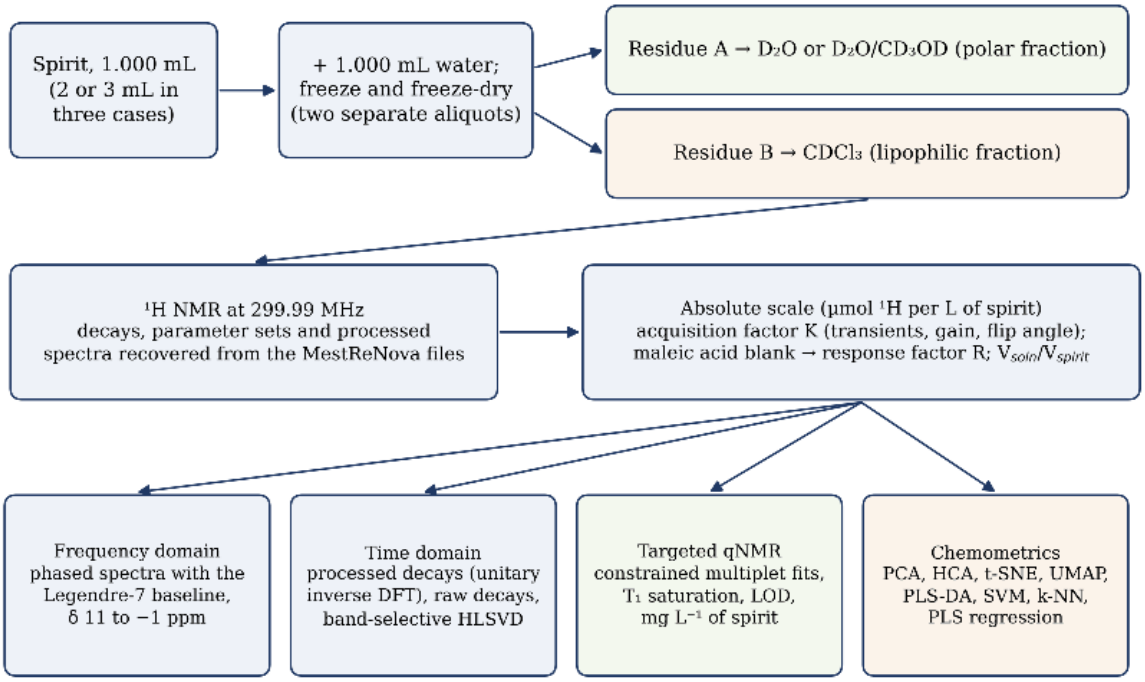

Figure 7. Analytical workflow. Two aliquots of each spirit (1.000 mL, diluted with 1.000 mL of water) were freeze-dried separately; one residue was redissolved in D_2O or D_2O/CD_3OD (polar fraction) and the other in $CDCl_3$ (lipophilic fraction). The decays, parameter sets and processed spectra were recovered from the MestReNova documents and placed on an absolute scale (µmol of ^1H per liter of spirit) through the acquisition factor *K*, the response factor *R* of the external maleic acid blank and the ratio of solution to spirit volume, and the same data were analyzed in the frequency and time domains, by targeted qNMR and by chemometrics.

4.2 Sample preparation

Each spirit (1.000 mL; 2.000 or 3.000 mL in three cases) was mixed with an equal volume of water, frozen and freeze-dried to constant appearance, and the whole residue was redissolved. The D_2O residues were dissolved in 500 µL of D_2O with 150 µL of CD_3OD (123 spectra), 450 µL of D_2O with 150 µL of CD_3OD (28), 625 µL of D_2O (12), 350 µL each of D_2O and CD_3OD (2) or 500 µL of D_2O with 200 µL of CD_3OD (1); for 31 early D_2O preparations the title records only "D_2O" or "D_2O + CD_3OD" without volumes, and a total volume of 650 µL was assumed, which the CHD_2OD intensities of those residues support. The D_2O contained TMSP or a trace of DSS in most preparations, at a concentration that differed between batches; it was used as the chemical-shift reference but not as a quantitation standard. The $CDCl_3$ residues were dissolved in 650 µL (10), 680 µL (11) or 620 µL (3) of $CDCl_3$ containing TMS, two of the 620 µL solutions being centrifuged before measurement; for 37 $CDCl_3$ preparations the volume was not recorded and 650 µL was assumed. No internal quantitation standard was added to any freeze-dried preparation. The external standard was a blank solution prepared from 2.55 mg of maleic acid in 500 µL of D_2O and 150 µL of CD_3OD containing TMSP and measured under the same conditions as the samples; its effective maleic acid content, determined by the author, was 95 % of the weighed amount (32.11 mmol L^{-1} maleic acid, 64.22 mmol L^{-1} of olefinic ^1H; Section 5.9).

4.3 NMR acquisition

All spectra were acquired at 299.99 MHz on the Varian Mercury-MVX 300 spectrometer (JCE-PNA-MVX300) with a 5 mm ATB probe, the standard one-pulse sequence (s2pul), a spectral width of 14 992.5

Hz, 64 000 complex points (acquisition time 4.2688 s), sample spinning at 20 Hz and a probe temperature of 27 °C. The parameters that varied between spectra were recovered from the stored Varian parameter sets and are summarized in Table 28. The pulse width was 8 µs for every D_2O spectrum and for most $CDCl_3$ spectra (7 µs for the eleven Hudson $CDCl_3$ spectra and for every DMSO-d_6 spectrum). The 90° pulse of the probe, as calibrated by the author, is 13 µs, so the flip angles were 55.4° and 48.5°. The parameter sets record a 90° pulse (pw90) of 10 µs; that nominal value, which versions 2.0 and 2.1 of this paper used, implied flip angles of 72° and 63° (Section 5.9). The relaxation delay ranged from 2 to 15 s, so the repetition time (delay plus acquisition) ranged from 6.27 to 19.27 s. The receiver gain was 38 dB for 141 of the 197 D_2O spectra and for every $CDCl_3$ spectrum, and between 18 and 36 dB for the remainder. The number of completed transients ranged from 16 to 516. No steady-state transients were used.

4.4 Recovery of free-induction decays and processed spectra from the MestReNova documents

A MestReNova document is a Qt-serialized binary file. Its structure was decoded far enough to extract, for every one-dimensional item, the three objects needed for an absolute analysis. The Varian text, parameter (procpar) and log files are stored as zlib-compressed byte arrays with a four-byte length prefix and were recovered verbatim. The free-induction decay is stored as little-endian 32-bit floating-point complex pairs immediately after a pair of length words that encode the number of bytes (four times the number of acquired points); every value was integer-valued, as expected for the 32-bit integer sums written by the Mercury console with dp = 'y', and the record length matched the procpar value of np for every item. The processed spectrum is stored as a single-row JPEG 2000 codestream written by the Kakadu library (reversible 5/3 wavelet, no quantization, 28-bit unsigned samples) preceded by its minimum and maximum values in 32-bit floating point. The codestream is written without the JPEG 2000 DC level shift, so a conforming decoder adds 2^{27} and clips the strongest lines; the component descriptor was therefore rewritten as signed 30-bit before decoding with OpenJPEG, which recovers the stored integers exactly, and the spectrum was reconstructed as minimum + integer × (maximum − minimum)/(2^{28} − 1). For the two-dimensional items of the Kavalan documents (gHSQC and homonuclear J-resolved spectra) the processed matrices were recovered in the same way, and the axis definitions were taken from the parameter tables stored with each item.

4.5 Reconstruction and verification of the processing

The processed spectrum stored by MestReNova was reproduced from the stored free-induction decay for every one of the 276 items by a model in which the decay was multiplied by an exponential window, zero-filled to 262 144 points and Fourier transformed, the result was divided by the transform length, phase-corrected with the zero- and first-order constants imported from the Varian rp and lp parameters (PH0 = rp + lp, PH1 = −lp, applied as exp[−i(PH0 + PH1·k/N)] with k counted from the downfield edge), and a seventh-order Legendre polynomial was added to represent the baseline correction applied in MestReNova. A single overall scale factor and a small zero-order correction were fitted together with the polynomial. The apodization constant was taken from the parameter set and, where the fit was poor, refined on a grid; the processing software had used 0.2 Hz for most first-set $CDCl_3$ spectra and 0.5 Hz for most second-set $CDCl_3$ spectra in place of the values in the parameter sets. After that refinement the coefficient of determination between the stored and the reconstructed spectrum exceeded 0.99999 for every item (Figure 8), with residuals of the order of 10^{-5} of the largest line. The scale factor was 1.0000 for every second-set D_2O spectrum and every $CDCl_3$ spectrum; for 100 of the 101 first-set D_2O items (83 of the 84 retained in the data base) it was the normalization factor that MestReNova had applied to set the reference line to an intensity of 100, a fact that is not apparent from the stored spectrum itself. The stored processed spectra are

therefore absolute only for the second sets, and the absolute intensities of the first D₂O set were restored by dividing by the fitted factor.

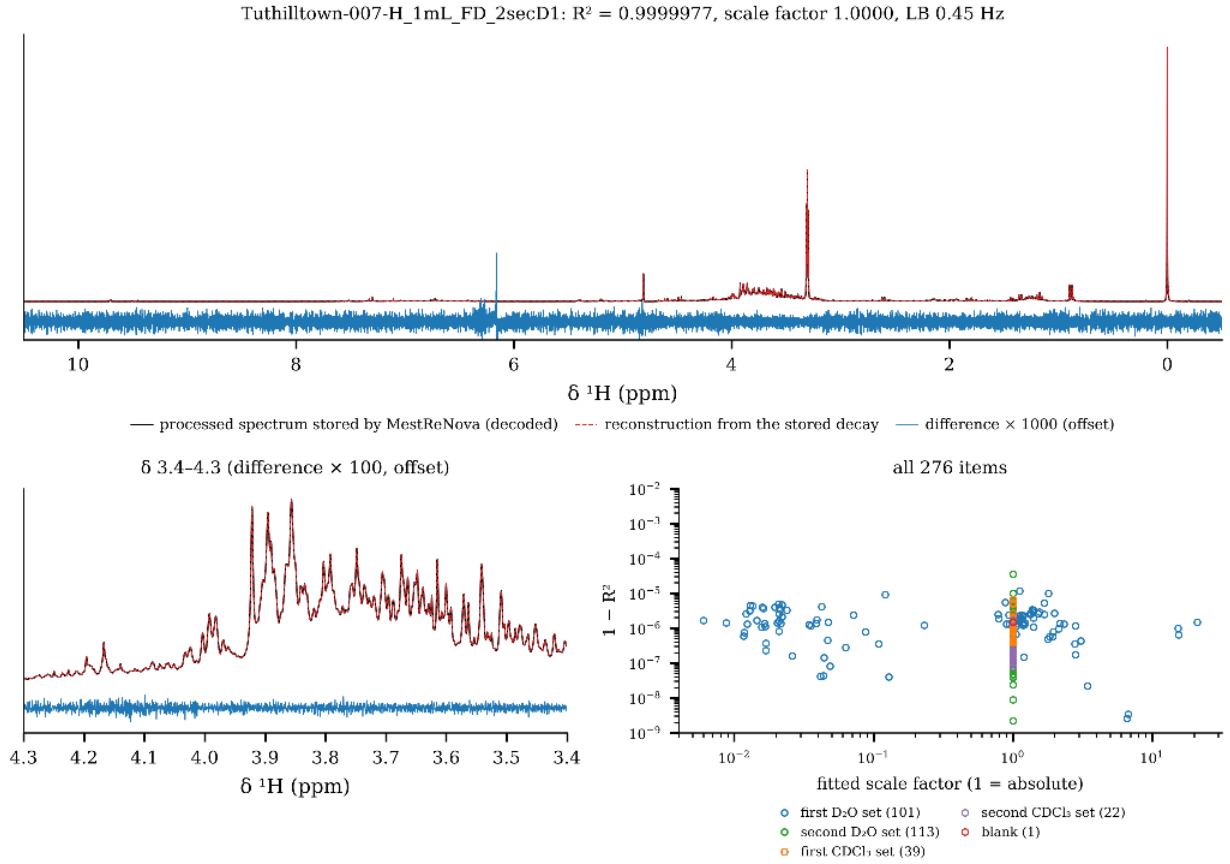

Figure 8. Recovery of the stored data. Top: processed spectrum stored in the MestReNova document for one residue (black), its reconstruction from the stored decay with the recovered processing parameters (red, dashed) and their difference multiplied by 1000 (blue, offset). Bottom left: expansion of δ 3.4–4.3 with the difference multiplied by 100. Bottom right: $1 - R^2$ of the reconstruction against the fitted scale factor for all 276 items; the scale factors of the first D₂O set, spread over more than three decades, expose the silent normalization that was reversed, whereas all other sets were stored on their absolute scale (scale factor 1).

For the analyses reported here every decay was reprocessed uniformly with an exponential window of 0.30 Hz, the phase constants recovered above and the MestReNova baseline polynomial, and interpolated onto a common chemical-shift grid from δ 11.0 to −1.0 at the native point spacing of 1.9065×10^{-4} ppm. The uniform window removes the variation of line width between spectra processed with different apodization constants without affecting integrals.

4.6 Chemical-shift referencing

Referencing was taken from the parameter tables stored with each item, which include the reference correction applied in MestReNova, and was checked against TMSP or DSS (D₂O) and TMS (CDCl₃). In the CDCl₃ spectra TMS lay at δ 0.000 ± 0.0002 and CHCl₃ at δ 7.2597 ± 0.0006. In the D₂O spectra the reference line lay at δ 0.000 ± 0.0002 in all but four cases: two spectra (Clase Azul Añejo and Litchfield Maple Bourbon) had been referenced to the downfield ^{29}Si satellite of TMSP and were corrected by −0.0111 ppm, and two spectra without a reference line (Litchfield Straight Rye and Hudson House Rye) had been referenced to a spinning sideband and were corrected by −0.067 and −0.070 ppm, the corrections being

determined by cross-correlation with the median of the TMSP-referenced spectra of the same solvent system. A medium effect of the CD_3OD was also found, but only after the solvent labels had been corrected (Section 5.4). In the fourteen residues dissolved in D_2O alone, the CH–O protons of the carbohydrates lie 0.015–0.033 ppm downfield of their positions in D_2O/CD_3OD (23 % v/v CD_3OD), the gallic acid singlet 0.021 ppm downfield, and the aliphatic protons within 0.003 ppm of their positions in the mixed solvent. The effect is group-specific, so it was accommodated by analyte-specific windows in the targeted quantitation and by excluding the pure-D_2O residues from the binned chemometric data, rather than by a global shift.

4.7 Absolute-intensity normalization

The processed spectrum of item *i*, $S_i(\delta)$, is in units of the summed free-induction decay divided by the transform length. Because the Mercury console sums transients, the signal of a fully relaxed resonance scales with the number of completed transients n_i, with the receiver amplitude gain $10^{g_i/20}$ and with the sine of the flip angle θ_i. An acquisition factor was therefore defined for every spectrum relative to the blank, which was acquired at a gain of 38 dB:

$$K_i = n_i \times 10^{(g_i - 38)/20} \times \sin \theta_i \tag{1}$$

The response of the spectrometer per millimole per liter of protons was obtained from the Lorentzian area A_{MA} of the maleic acid singlet in the blank (δ 6.400; main line only, excluding the ^{13}C satellites and spinning sidebands), its effective concentration c_{MA} = 32.11 mmol L^{-1} (95 % of the 33.80 mmol L^{-1} weighed in; Section 5.9) and its acquisition factor $K_{blank} = 28 \times \sin 55.4° = 23.04$:

$$R = A_{MA} / (K_{blank} \times 2c_{MA} \times f_{sat,MA}) \tag{2}$$

where $f_{sat,MA}$ = 0.9812 is the steady-state recovery of maleic acid at the 19.27 s repetition time of the blank (Equation 4, with T_1 = 6.1 s, the value supplied by the author for maleic acid in D_2O, which lies within the range of published inversion-recovery measurements [34,35]; Section 5.1), so that $R = 7.800 \times 10^{-3}$ spectrum-area units per mmol L^{-1} of ^1H per unit of *K*. The absolute spectrum of sample *i* in millimoles of ^1H per liter of spirit per ppm was then:

$$Y_i(\delta) = S_i(\delta) / (K_i \times R) \times V_{soln} / V_{spirit} \tag{3}$$

where V_{soln} is the volume of NMR solvent and V_{spirit} the volume of spirit freeze-dried. Integrating Equation 3 over a resonance of n_H protons and dividing by n_H gives the molar concentration of the compound in the spirit, and multiplying by its molar mass gives mg L^{-1}. Signals that are not fully relaxed at the repetition time T_R of a given spectrum were corrected by the steady-state factor for a train of pulses of flip angle θ:

$$f_{sat} = (1 - E_1) / (1 - E_1 \cos \theta), \quad \text{with } E_1 = \exp(-T_R / T_1) \tag{4}$$

The ^{13}C satellites (about 1.1 % of each carbon-bound proton signal) and the spinning sidebands were excluded from the areas of both the standard and the analytes, so that their contributions cancel in the ratio.

4.8 Targeted quantitation

Quantification followed qNMR practice for mixtures [36,37] with an external standard [38,39]. Each target signal was modeled within a window of 0.05–0.16 ppm by one or more first-order multiplets of Lorentzian lines, built from the known coupling constants (singlet, doublet, triplet, quartet, doublet of doublets, or the doublet of doublets of doublets of the quercitol H-6ax proton), on a local linear baseline. The amplitudes were constrained to be non-negative and were obtained by bounded linear least squares at each step of a nonlinear refinement of the line positions and widths (variable projection), the positions being bounded within analyte-specific limits and the widths between 0.4 and 3.5 Hz for sharp lines; several starting points were

used for every fit, including a grid across the permitted range for signals whose position depends on pH. Overlapping signals were fitted jointly: the ethanol triplet with the 1,2-propanediol doublet, the lactic acid doublet with the methylene quartet of isoamyl alcohol, the HMF formyl singlet with the two cinnamaldehyde formyl doublets and a broad furanic envelope (6–45 Hz wide), and the two benzaldehyde formyl singlets with each other. The downfield ^{13}C satellite of CHD_2OD, which falls within the glycerol H-1a/H-3a multiplet, was subtracted at its known relative intensity (0.535 % of the CHD_2OD main line measured in the same spectrum) before glycerol was fitted. Because that multiplet also shares its center and its 11.7 Hz coupling with the β-pyranose H-1b doublet of fructose and lies beside the H-2 signals of glucose and sucrose, glycerol was quantified a second time from its H-1b/H-3b double doublet (δ 3.626, J = 11.7 and 4.4 Hz), which is clear of the fructose doublet. In residues containing less than 100 mg L^{-1} of fructose, 200 mg L^{-1} of glucose and 50 mg L^{-1} of sucrose the H-1a/H-3a result is reported. In the other residues the H-1a/H-3a multiplet was refitted together with the fructose doublet, and glycerol is reported as the mean of the two multiplet estimates only where they agreed within 40 % (or 5 mg L^{-1}); where the second multiplet was absent the result is reported as not detected, with the largest H-1b/H-3b multiplet that fits beneath the spectrum as an upper limit, and in the remaining residues, including every residue with more than 200 mg L^{-1} of sucrose, glycerol was not quantified. Fructose, whose β-furanose H-3/H-4 cluster cannot be modeled as a simple multiplet, and the acyl α-methylene, olefinic and methyl regions of the lipids were integrated over fixed windows on straight baselines drawn through the local minima at the window limits.

The fitted area of each signal, in micromoles of ^{1}H per liter of spirit (Equation 3), was divided by the number of protons it represents and by the steady-state factor of Equation 4 for the repetition time and flip angle of the spectrum, and multiplied by the molar mass to give mg L^{-1} of spirit. The T_1 values assumed for the analytes (Table 3) were taken from published measurements in D_2O, aqueous mixtures and wine where available [35,40,41] and otherwise estimated from molecular size and the proximity of other protons; the change of each result for a doubling of the assumed T_1 is reported with the results. The detection limit was defined as three times, and the quantitation limit as ten times, the standard deviation of the integral of the noise (δ 10.2–10.8) over the spectral extent of the multiplet. For fructose, whose integration window is crowded with other signals in residues of low sugar content, a matrix-limited reporting level of 28.8 mg L^{-1} was applied, and glycerol was not quantified where 1,2-propanediol, whose methylene protons fall within both glycerol multiplets, exceeded half of it on a molar basis. The thresholds in mg L^{-1} of this paragraph and the last describe where one signal begins to obscure another rather than a property of the spirit. They were set on the absolute scale of version 2.1 (30 mg L^{-1} for fructose; 100, 200 and 50 mg L^{-1} of fructose, glucose and sucrose for the glycerol decision; 5 mg L^{-1} for the agreement of the two glycerol multiplets) and have been carried with the corrected scale of version 2.2, that is, multiplied by 0.961, so that every residue was treated as before (Section 5.9). Where the lines of two compounds were too close to be apportioned reliably (the vanillin and syringaldehyde formyl singlets in D_2O, $\Delta\delta \approx 0.006$ ppm, and the coniferaldehyde and sinapaldehyde formyl doublets in both solvents, $\Delta\delta \approx 0.007$ ppm), only their sum was reported, and sinapaldehyde was quantified separately from its ring singlet in $CDCl_3$. Representative fits are shown in Figure 9.

In version 2.1 the line positions of every D_2O/CD_3OD target were made residue-specific. The anomeric doublets of each residue were located by cross-correlation of the anomeric region with a template of the sugar doublets (Section 9.5), and each target window and its position limits were moved by that offset multiplied by the ratio of the target's own displacement between D_2O/CD_3OD and D_2O to that of the anomeric doublets; this medium index follows the CD_3OD fraction of the solution, and in residues of the nominal composition the resulting shifts are negligible. In the two residues dissolved in equal volumes of D_2O and CD_3OD the displacements proved to differ from one proton to another, and the displacement of

each of their targets was therefore measured directly, by cross-correlation of the target window with the spectra of the three best-matching residues of the same category and nominal composition (correlation ≥ 0.85), except for the singlets of acetic, succinic and lactic acid, whose positions follow the ionization of each residue and are searched over a wide window in every residue. Glucose was quantified from its β-H-1 doublet and the α/β ratio of glucose measured in this solvent system, because the α-H-1 doublet carries the β-arabinopyranose H-1 doublet, and arabinose and xylose from their resolved anomeric doublets (Section 9.5); in the fourteen residues dissolved in D_2O alone, which were not analyzed for pentoses, glucose is the sum of the two anomeric doublets as in version 2.0.

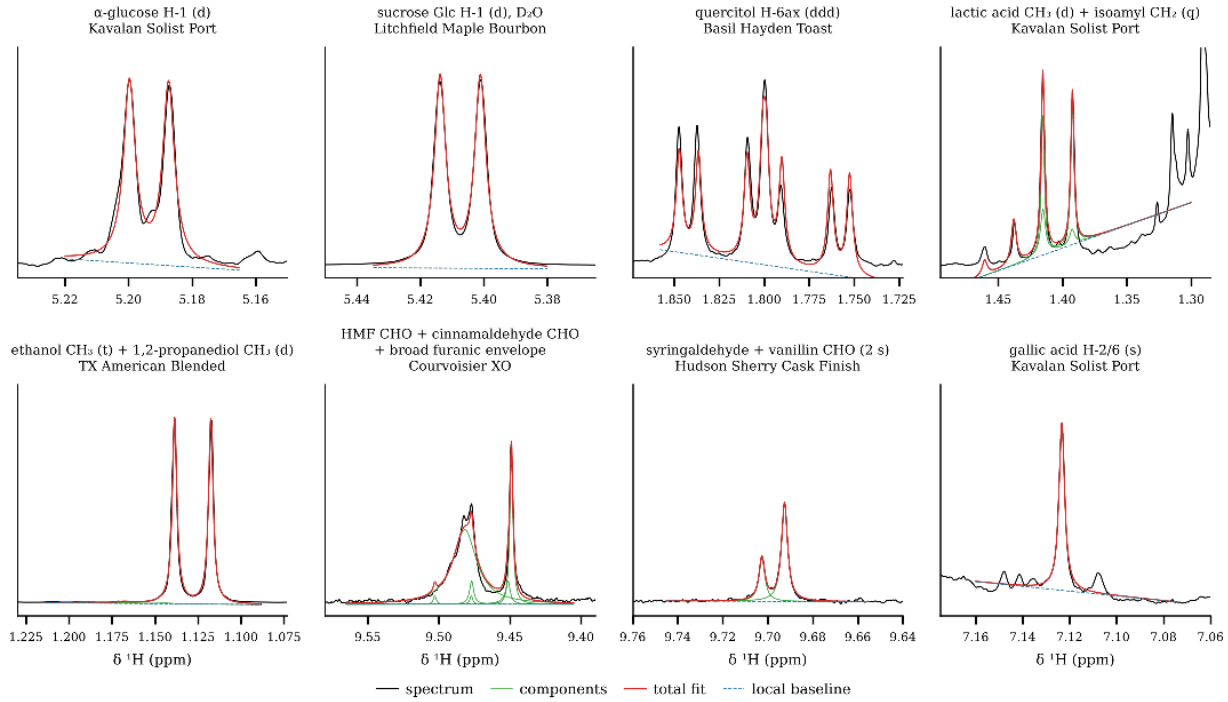

Figure 9. Examples of the constrained multiplet fits used for quantitation: experimental spectrum (black), fitted components (green), total fit (red) and local linear baseline (blue, dashed), each panel for the residue named in its title. The spectrum, the components, the fit and the baseline of a panel share one scale, set by the tallest line in the fitted range within the common noise ceiling of Section 4.17.

Table 3. Longitudinal relaxation times assumed for the target signals, median steady-state recovery at the repetition times used, and median change of the result if the assumed T_1 were doubled.

Signal	T_1 assumed (s)	Median f_{sat}	Median change for $2T_1$ (%)	Basis
α-Glucose H-1	2.0	0.998	2.6	estimated; anomeric CH of a hexose
β-Glucose H-1	2.0	0.998	2.6	estimated
Sucrose Glc H-1 / Fru H-3	1.5	1.000	1.0	estimated
Fructose (β-furanose H-3/H-4 window)	1.5	1.000	1.0	estimated
Glycerol CH_2	2.3	0.997	3.7	published, aqueous solution [35]
1,2-Propanediol CH_3	2.0	0.998	2.6	estimated
Quercitol H-6ax	1.0	1.000	0.2	estimated; geminal partner shortens T_1
Lactic acid CH_3	3.0	0.990	6.7	wine measurements [40]
Succinic acid CH_2	2.0	0.998	2.6	wine measurements [40]
Acetic acid CH_3	3.0	0.990	6.7	estimated
Gallic acid H-2/6	4.0	0.973	10.9	estimated; isolated aromatic CH
Syringyl H-2/6	3.0	0.990	6.7	estimated

Signal	T_1 assumed (s)	Median f_{sat}	Median change for $2T_1$ (%)	Basis
Benzaldehyde CHO	4.0	0.973	10.9	estimated; isolated formyl proton
HMF CHO	4.0	0.973	10.9	estimated
Ethanol CH_3	3.0	0.990	6.7	estimated
Isoamyl/isobutyl CH_3	3.0	0.990	6.7	estimated
Aldehyde CHO ($CDCl_3$)	4.0	0.955	14.4	estimated
Syringaldehyde H-2/6 ($CDCl_3$)	3.0	0.980	9.5	estimated
FAEE OCH_2 ($CDCl_3$)	2.0	0.996	4.3	estimated
Acyl α-CH_2 ($CDCl_3$)	1.0	1.000	0.4	estimated; lipid CH_2

f_{sat} from Equation 4 with the flip angle and repetition time of each spectrum. The maleic acid standard was assigned T_1 = 6.1 s (Section 5.1).

4.9 Internal consistency checks

Three resonances of known or constant concentration were used to test the absolute normalization independently of the maleic acid standard. The residual CHD_2OD of the CD_3OD added to most D_2O preparations is present at a concentration fixed by the volume and isotopic purity of the CD_3OD; its quintet (δ 3.300, J(H,D) = 1.65 Hz) was fitted in every spectrum that contained 150 µL of CD_3OD in a total of 650 µL, and the areas were modeled as a function of repetition time and receiver gain. TMSP was compared within preparation batches, and the residual $CHCl_3$ of the $CDCl_3$ was compared across the $CDCl_3$ set.

4.10 Data pretreatment for chemometrics

The absolute spectra (Equation 3) were integrated into 0.02 ppm bins between δ 11.0 and −1.0 (600 bins), giving the content of each bin in micromoles of ^{1}H per liter of spirit. Bins containing the reference line and its satellites (δ < 0.30 in D_2O, < 0.25 in $CDCl_3$), the CHD_2OD quintet (δ 3.285–3.318), HDO (δ 4.55–5.10), the retained-ethanol methyl triplet (δ 1.145–1.200), $CHCl_3$ and its ^{13}C satellites (δ 7.20–7.32, 6.895–6.935 and 7.585–7.625) and dissolved water in $CDCl_3$ (δ 1.45–1.60) were removed, leaving 502 D_2O and 517 $CDCl_3$ variables. No total-area or probabilistic quotient normalization was applied: because every residue derives from a known volume of spirit, the absolute intensities are the composition, and normalizing them away would discard the information on total extract that version 1.0 could not use [42]. Every model was computed on mean-centered, Pareto-scaled data [43], with the centering and scaling recomputed within each cross-validation training set. The fourteen residues dissolved in D_2O alone were excluded from the binned D_2O models because their carbohydrate signals are displaced by up to 0.033 ppm (Section 5.4), which leaves 183 D_2O/CD_3OD residues; a sensitivity analysis that included them is reported with the results. Because the carbohydrate signals of sweetened products dominate the absolute variance, models were also computed on a non-carbohydrate block from which the windows δ 3.10–4.42 and 5.06–5.50 were removed, as in version 1.0.

4.11 Chemometric and machine-learning methods

Principal component analysis was computed by singular value decomposition of the scaled matrices. Hierarchical cluster analysis used Ward linkage on Euclidean distances and average linkage in the Pareto-scaled space, the cophenetic correlation being reported as a measure of the fidelity of each dendrogram and the adjusted Rand index as the agreement of a seven-cluster partition with the spirit category. Supervised classification of the seven categories represented by at least fourteen residues (American, Irish, Scotch and Taiwanese whisky, rum, brandy and agave spirits) was performed with PLS-DA (one-versus-rest dummy coding, class assigned by the largest predicted value) [44], linear and radial-basis-function support vector machines with class weights inversely proportional to class size [45], and k-nearest-neighbor classifiers operating on principal-component scores. Hyperparameters (number of latent variables; regularization constant and kernel width; number of components, neighbors and weighting) were selected by an inner

five-fold cross-validation within each outer training set, and performance was estimated by an outer ten-fold cross-validation repeated three times. Every fold was stratified by category and grouped by product, so that replicate preparations of the same product never appeared in both the training and the test set, and the Pareto scaling was recomputed within each training set. Balanced accuracy (the mean of the per-class recalls) and the macro-averaged F1 score were reported, with confusion matrices. Three further tests addressed the ways in which such a classification can mislead [46]. In a leave-producer-out validation all residues of a producer were held out together, so that a category had to be recognized from producers the model had never seen; this test was run on the six categories that contain more than one producer, because every Taiwanese residue came from the same distillery. In a cross-batch validation, models trained on one acquisition set were applied to the other. And in label-permutation tests the class labels were permuted 200 times and the cross-validated balanced accuracy recomputed at the hyperparameters chosen most often in the nested procedure; the permutation *p*-value is the fraction of permuted accuracies at least as high as the observed one. A test of whether the spectra predict the acquisition set within the four largest categories was used as a check for batch effects. UMAP [47] and t-SNE [48] embeddings were computed from the first twenty Pareto principal-component scores (UMAP with fifteen neighbors and a minimum distance of 0.1; t-SNE with a perplexity of 30 and PCA initialization) for five random seeds each, and were evaluated by their trustworthiness at ten neighbors [49] and by the purity of the ten-nearest-neighbor sets with respect to category, compared with the same purity in the full scaled space; they were used for display and neighborhood inspection only, not for classification. For the paired residues, D_2O and $CDCl_3$ blocks were fused at the low level after Pareto scaling of each block and division by its root-mean-square, so that the two blocks carried equal weight.

4.12 PLS regression calibration

PLS1 regression models were built between the Pareto-scaled binned spectra and each quantified concentration [50]. The number of latent variables was chosen by leave-10 %-out (ten-fold) cross-validation, grouped by product and repeated five times with shuffled group assignments, as the smallest number whose root-mean-square error of cross-validation (RMSECV) was within one standard error of the minimum, and the cross-validated coefficient of determination (Q^2) and RMSECV were reported. Residues whose reference value was not quantified (glycerol in sugar-rich residues) were excluded from the model for that analyte rather than given a value of zero. Because the reference values were themselves derived from the same spectra, each model was also rebuilt with the bins that contain the integrated signal removed ("masked" models), which tests whether the calibration can be carried by co-varying signals elsewhere in the spectrum rather than by the target resonance itself, and selected concentrations were predicted across solvents ($CDCl_3$ spectra predicting D_2O concentrations and the reverse), which tests the models against an independent preparation.

4.13 Time-domain analysis

The analyses were repeated on three time-domain representations. The raw free-induction decays (the first 16 384 complex points) were scaled by the same per-sample factors as Equation 3 but neither phased, referenced, filtered nor baseline-corrected. "Processed" decays were obtained as the unitary inverse discrete Fourier transform of the band-limited (δ 11.0 to −1.0), phased, baseline-corrected and referenced spectra, with the excluded windows set to zero; this is equivalent to digital filtering, phasing and decimation of the decay, and it preserves inner products, so any analysis that depends on the data only through mean-centered inner products must give identical results in the two domains (Parseval's theorem). Integrals over spectral windows were computed directly from the decays as inner products with the corresponding time-domain kernels. Finally, individual resonances were quantified in the time domain: the decay was transformed

without apodization or zero filling and phased with the constants recovered above, a narrow band around the target was selected, the complex linear background contributed by the dispersion tails of distant intense lines was removed, and the band was returned to the time domain, which amounts to heterodyning, low-pass filtering and decimation in one step. The band-limited signal was then decomposed by Hankel singular-value decomposition (HLSVD) into exponentially damped sinusoids [51–53], and the target lines were identified by frequency and damping; the magnitude of their complex amplitudes, which does not depend on the phase, gives the area. Because the first point of the decay is not halved before the Fourier transform, half of the intensity of every line appears in the frequency domain as a constant offset that the baseline correction removes, so the area of a line in the baseline-corrected spectrum corresponds to half the magnitude of its time-domain amplitude, and the HLSVD areas were scaled accordingly. Bayesian and related time-domain approaches such as CRAFT pursue the same goal on a larger scale [54].

4.14 Software and verification of the literature

All calculations were performed in Python 3.13 with NumPy 2.5, SciPy 1.18, scikit-learn 1.9.1, umap-learn 0.5.12, pandas and Matplotlib 3.11; JPEG 2000 codestreams were decoded with OpenJPEG 2.5 through glymur. The document was built with docx-js. Every journal reference was verified against its CrossRef record for authors, title, journal, volume, pages and DOI before it was cited, and the regulatory documents and the one journal article without a DOI were verified against the issuing body's own web page.

4.15 The DMSO-d$_6$ residues

A fifth MestReNova document contained the spectra of residues dissolved in DMSO-d$_6$. Each residue was prepared as for the other solvents, from 1.000 mL of spirit (3.000 mL for the two Adirondack products) diluted with an equal volume of water and freeze-dried, and was dissolved in 600 µL of DMSO-d$_6$ (610 µL in one case). The spectra were acquired on the same spectrometer with the one-pulse sequence, a 7 µs pulse (48.5° for the 90° pulse of 13 µs; Section 5.9), a relaxation delay of 3 s and the acquisition time of 4.2688 s, so that every DMSO-d$_6$ spectrum has the same repetition time of 7.27 s; the number of transients ranged from 48 to 512 and the receiver gain from 18 to 38 dB. The decays, parameter sets and processed spectra were recovered and the processing reproduced as described in Sections 4.4 and 4.5, and the decays were reprocessed with an exponential window of 0.30 Hz onto the common chemical-shift grid. Decays carried under more than one title were identified by their checksums and counted once.

The chemical-shift reference was TMS (δ 0.000) where TMS was present and otherwise the cross-correlation of the alkyl region (δ 0.78–1.32) with the median of the TMS-referenced spectra, or of the formyl region (δ 9.45–9.85) for residues with weak alkyl signals (Section 5.8). Because some solute lines move with the composition of the residue and the movement differs between proton types, region-wise offsets were also determined by cross-correlation with the median spectrum for the formyl, aromatic, anomeric, ester, acid and alkyl regions, and were applied, with a 0.02 ppm taper at the region limits, to the spectra whose solvent line was displaced or which had been referenced by cross-correlation; offsets were accepted only where the correlation exceeded 0.75. This procedure is a simple form of the segment-wise alignment used for metabolomic spectra [55]. The absolute spectra were obtained from Equation 3 with V_{soln} the volume of DMSO-d$_6$.

The targeted quantitation followed Section 4.8, with windows and line positions chosen for this solvent (Table 9) and with the window positions moved by the region offsets. Signals that could not be separated reliably were fitted jointly: the syringaldehyde and vanillin formyl singlets, the two cinnamaldehyde formyl doublets, the syringaldehyde H-2/6 singlet with the syringyl singlet at δ 7.200, the vanillin H-6 and H-2 signals, the gallic acid singlet with the singlet at δ 6.940, and the anomeric doublets of glucose and of the

two pentoses in two groups (small and large couplings). Line widths of up to 6 Hz were allowed for the anomeric doublets, whose widths depend on the rate of hydroxyl exchange (Section 9.4). Both ^{13}C satellites of DMSO-d_5 were modeled as 1:2:3:2:1 quintets carrying 0.535 % of the area of the DMSO-d_5 line each, with the position, width and ^{2}J(H,D) fitted on the clean downfield satellite near δ 2.727 and the upfield satellite placed 0.4586 ppm upfield of it (^{1}J(C,H) = 137.6 Hz), and the model was subtracted before the acyl α-methylene triplets were fitted. Quercitol and the olefinic protons were integrated over fixed windows (δ 1.555–1.685 and 5.29–5.36). The T_1 values were those assumed for the same protons in D_2O, and the saturation correction and its sensitivity to a doubling of T_1 were computed for the 48.5° pulse and the 7.27 s repetition time. Detection limits were derived from the noise between δ 9.95 and 10.95, which is free of signals in every DMSO-d_6 spectrum. For the cross-solvent comparisons, DMSO/D_2O and DMSO/$CDCl_3$ ratios were computed for every product with residues in both solvents and for every congener detected in both; ester groups were compared with the $CDCl_3$ fatty-acid ethyl esters plus twice the diethyl succinate.

The recognition of arabinose and xylose in the DMSO-d_6 spectra (Section 9.5) prompted a reanalysis of the anomeric region of every D_2O/CD_3OD residue. The β-glucose H-1 doublet (δ 4.598, J = 7.9 Hz), the β-xylopyranose H-1 doublet (δ 4.535, J = 7.8 Hz), the α-arabinopyranose H-1 doublet (δ 4.474, J = 7.8 Hz) and the α-xylopyranose H-1 doublet (δ 5.151, J = 3.6 Hz) were fitted separately, and the α-anomeric region (δ 5.17–5.22) was fitted with two doublets representing α-glucose and β-arabinopyranose H-1. Because these two doublets are separated by only 0.004 ppm, the glucose contribution to the α-region was taken as the β-glucose area multiplied by the α/β anomeric ratio of glucose measured in the glucose-rich, pentose-poor residues of the same solvent system, and the remainder was assigned to β-arabinopyranose. Glucose was then computed from its β-H-1 doublet multiplied by one plus that ratio, arabinose from its two pyranose doublets and xylose from its two anomeric doublets; the furanose forms of arabinose (about 7 % in water) were not observed and were neglected.

4.16 Full accounting of the chemometric data approaches and the comparison of solvents

Every data approach used in this work was analyzed in the same way, so that the approaches can be compared on the same footing (Section 12). The approaches comprise five binned D_2O blocks (the full D_2O/CD_3OD spectrum at 0.02 ppm, the non-carbohydrate block, the non-carbohydrate and volatile-free block, the non-carbohydrate block at 0.04 ppm, and the non-carbohydrate block including the residues in D_2O alone), the quantified D_2O congeners (log-transformed, with glucose replaced by its corrected value and arabinose and xylose added, analyzed after Pareto scaling and, for comparison, after autoscaling), the $CDCl_3$ block, the fused D_2O and $CDCl_3$ block of the paired products, three binned DMSO-d_6 blocks (the full spectrum without the solvent and exchange-water regions at 0.02 and 0.04 ppm and a non-carbohydrate block from which δ 2.80–4.00 and 4.15–5.30 were also removed) and the quantified DMSO-d_6 congeners, again with both scalings. The DMSO-d_6 blocks were built from the referenced, region-aligned spectra after subtraction of the DMSO-d_5 ^{13}C satellites, with δ < 0.30, 2.44–2.56 and 3.25–3.95 excluded. Every block was mean-centered and, unless stated otherwise, Pareto-scaled [43]; the fused blocks were Pareto-scaled block by block and each block was divided by its root-mean-square so that the blocks carried equal weight [56].

For every block the principal components were computed by singular value decomposition, with scree plots and the loadings of the first three components. Hierarchical clustering was performed four ways: Ward linkage on Euclidean distances, average linkage on Euclidean distances, complete linkage on Euclidean distances and average linkage on correlation distances (one minus the Pearson correlation between the scaled profiles of two residues), the last being insensitive to differences in total extract between residues of similar composition. For each tree the cophenetic correlation with the distances from which it was built and

the adjusted Rand index between its partition into k clusters and the class labels were computed (k = 7 for the seven categories of the D_2O blocks, 5 for the $CDCl_3$ and fused blocks, and the number of residue groups or two for wine-seasoned against other casks in the DMSO-d_6 blocks), together with the fraction of residues in the largest cluster as a measure of chaining. t-SNE and UMAP embeddings were computed from the first twenty principal-component scores, with perplexity min(30, (n − 1)/3) and min(15, n − 1) neighbors so that the small DMSO-d_6 blocks were treated consistently with the large ones, for five random seeds each; trustworthiness [49] and k-nearest-neighbor class purity were computed for each embedding and compared with the same purity in the full scaled space [57].

The products with residues in more than one solvent were used to compare one solvent with two. For the 24 products with D_2O/CD_3OD and DMSO-d_6 residues (set S1) and the 10 products with residues in all three solvents (set S2), the blocks were restricted to those products and compared pairwise by the modified RV coefficient, a matrix correlation between the configurations of the residues in two blocks; by Procrustes analysis of the first three principal-component scores, with the significance of the residual sum of squares m^2 assessed by 999 permutations; by the Mantel correlation between the two distance matrices, also with 999 permutations; by the correlation between the cophenetic distances of the two Ward dendrograms; and by the adjusted Rand index between their partitions into two to five clusters. The ability of each block, alone or fused, to separate wine-seasoned casks from other casks was measured by the silhouette width and by leave-one-out PLS-DA (two latent variables) and linear SVM classification, with the block scaling recomputed in every training set and the significance of the balanced accuracy assessed by 499 label permutations.

4.17 Reference spectra of congeners and the display of spectra

The reference spectra of Section 2.8 were recovered from their MestReNova document as described in Section 4.4. For each spectrum the phases, the line broadening and a seventh-order polynomial baseline were recovered by fitting the transform of the stored decay to the stored processed spectrum, as in Section 4.5, and every decay was then transformed again with a common exponential window of 0.3 Hz, so that all traces carry the same apodization. An internal TMSP or DSS singlet (aqueous media) or TMS singlet ($CDCl_3$ and DMSO-d_6) was set to δ 0.000 where one was present; otherwise the residual $CHCl_3$ line was set to δ 7.26, or the residual CHD_2OD quintet to δ 3.302 as in the residues, and the eleven spectra in D_2O without any reference were used at their stored positions. Diagnostic positions were measured as intensity-weighted centroids of the multiplets, only where the tallest line exceeded eight times the noise. The intact spirit of Section 2.9 is a spectrum of the second D_2O data set, recorded from 175 µL of spirit with 450 µL of D_2O and 50 µL of the maleic acid solution (128 transients, receiver gain 6 dB, repetition time 19.3 s), which had been excluded from the freeze-dried data base (Appendix C).

Every stacked or overlaid spectral figure of this version follows one vertical-scaling rule. Within each displayed window, each trace is offset by the fifth percentile of its intensities and scaled so that its tallest analyte line, with the solvent, reference and ethanol windows excluded, reaches 85 % of the row height, unless that scale would make the standard deviation of its noise larger than 0.4 % of the row height, in which case the trace is scaled to that noise ceiling instead. A line therefore reaches full height only if its signal-to-noise ratio is at least 212, and a weak spectrum is drawn with proportionately smaller lines over a baseline whose noise looks like that of every other trace, rather than being expanded until its noise fills the row. The noise is estimated as the robust standard deviation (1.4826 times the median absolute deviation) of the residual after a quadratic detrend in the signal-free regions beyond δ 10.2 and below δ −0.35, not from first differences, which zero filling and line broadening correlate. All axes have fixed vertical limits; where a magnified inset is drawn on a full spectrum, its magnification is limited by the same

ceiling; and in the fit figures the spectrum, the fitted components, the total fit and the baseline of a panel share one offset and one scale. The spectral figures of the earlier versions were redrawn with this rule; their layouts and selections are otherwise unchanged, except that the fit figures now show, for each window, a residue in which the signal is strong enough to be seen at this scale, and the pentose figure shows Basil Hayden Toast and the Kavalan Artist Series Virgin Oak, whose pentose signals are stronger than those of the residues shown before.

5. Validation of the absolute intensity scale

5.1 The external maleic acid standard

The quantitative claims of this paper rest on a single calibration spectrum, so the conditions under which that spectrum transfers to the samples were examined before any concentration was reported. The blank was prepared from 2.55 mg of maleic acid in 500 µL of D_2O and 150 µL of CD_3OD, with TMSP as the shift reference; its effective maleic acid content was 95 % of the weighed amount (Section 5.9), or 32.11 mmol L^{-1} (64.22 mmol L^{-1} of olefinic protons), and it was acquired with 28 transients at a receiver gain of 38 dB, an 8 µs pulse (55.4°; Section 5.9) and a repetition time of 19.27 s (Figure 10). The Lorentzian area of the maleic acid singlet at δ 6.400, taken without its ^{13}C satellites (^1J(C,H) ≈ 165 Hz, each satellite 0.54 % of the main line) and without spinning sidebands, defined the response factor R of Equation 2. The one quantity in that equation that could not be measured from the blank itself is the steady-state recovery of the maleic acid signal, which depends on its longitudinal relaxation time. The preliminary pass of this work assumed T_1 = 4 s. Published inversion-recovery measurements in D_2O give 5.4 s at 500 MHz and 20 °C [34] and 6.78 ± 0.02 s at 600 MHz and 298 K [35], and in the extreme-narrowing regime the dipolar relaxation of a small molecule is essentially independent of field, so these values transfer to 300 MHz. Version 2.0 adopted 6.0 s on that basis. For version 2.1 the author supplied a value of 6.1 s for the T_1 of maleic acid in D_2O, which lies within the published range and was adopted, giving $f_{sat,MA}$ = 0.9812; every concentration in this paper is therefore 1.5 % lower than it would have been with the preliminary assumption of 4 s, and 0.10 % lower than with the 6.0 s of version 2.0. The blank, however, is not pure D_2O. It contains 23 % (v/v) CD_3OD, which raises the viscosity of the solution by about half relative to D_2O and changes the amount of dissolved oxygen, and both effects can only shorten T_1. If the T_1 of maleic acid in the blank were as short as 4.5 s, $f_{sat,MA}$ would be 0.9940 and every concentration 1.3 % higher. A range of 4.5–6.5 s was therefore carried in the uncertainty budget, over which the recovery factor varies between 0.9940 and 0.9770, a half-range of ±0.9 % that enters as a rectangular component of 0.50 %. An inversion-recovery measurement of maleic acid in the blank solvent itself, which takes a few minutes, would remove this component and is recommended in Section 17.

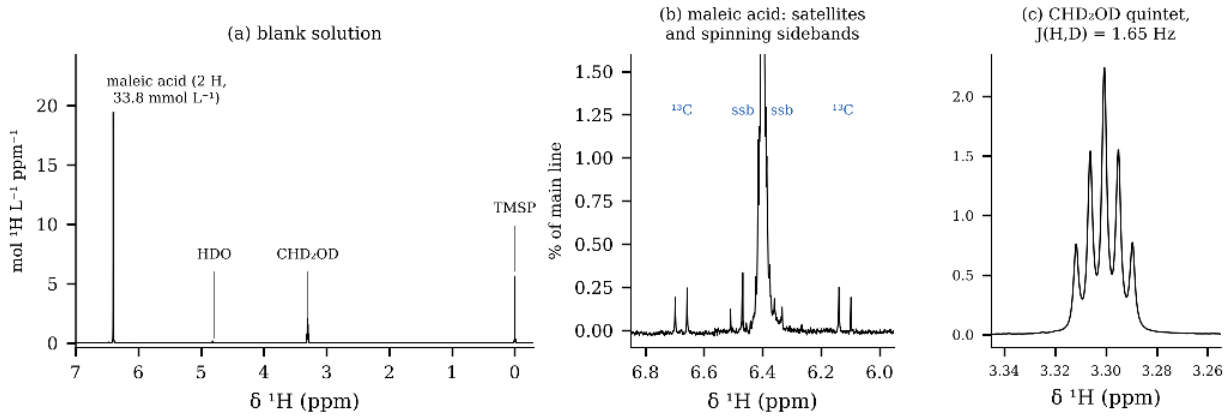

Figure 10. The external standard. (a) Spectrum of the blank (2.55 mg of maleic acid in 500 µL of D_2O and 150 µL of CD_3OD, 32.11 mmol L^{-1}) on the absolute scale, with HDO, the residual CHD_2OD and TMSP. (b) The maleic acid singlet plotted as a percentage of the height of its main line, showing the ^{13}C satellites and spinning sidebands (ssb) that were excluded from the Lorentzian area A_{MA}. (c) The CHD_2OD quintet (J(H,D) = 1.65 Hz) used as the internal tracer.

5.2 The residual CHD_2OD as an internal tracer

The CD_3OD added to most D_2O preparations carries a residual CHD_2OD signal whose concentration is fixed by the isotopic purity of the solvent and by the volume added. It is therefore an internal standard that

nobody intended to add, present at nearly the same concentration in the blank and in every residue prepared with 150 µL of CD_3OD in 650 µL of solution, and it provides an independent test of every step that converts a raw spectrum to an absolute one: the transient count, the receiver gain, the solvent volume and the transfer of the response factor from the blank. It does not test the flip angle, because the blank and the D_2O residues were acquired with the same pulse, nor the content of the standard, which scales the tracer in the blank and in the residues alike (Section 5.9). The CHD_2OD quintet (δ 3.302, J(H,D) = 1.65 Hz) was fitted in all 123 qualifying spectra, converted to millimoles per liter of solution with the maleic acid response factor, and modeled as $c_{obs} = c_0 \cdot f_{sat}(T_1, T_R, 55.4°) \cdot G(\text{gain})$, with a separate multiplicative term for each receiver gain other than 38 dB. After iterative removal of three outliers beyond three standard deviations, 120 spectra gave c_0 = 25.4 ± 0.9 mmol L^{-1} for the fully relaxed CHD_2OD concentration and T_1 = 12.7 ± 1.2 s, with a residual standard deviation of 10 % on a logarithmic scale (Figure 11). The same quintet in the blank, corrected with the same T_1, corresponds to 26.3 mmol L^{-1}. The agreement between the residues and the blank, 3.4 %, is a direct confirmation that the absolute scale established on maleic acid transfers to the samples without a hidden factor. The fitted T_1 of CHD_2OD is long, as expected for a small, rapidly tumbling molecule whose only proton has deuterium neighbors, and it accounts for the rise of 50 % in the CHD_2OD intensity between the shortest (6.27 s) and the longest (19.27 s) repetition times. It is, however, strongly correlated with the flip angle assumed in the model: at the nominal 72° of version 2.1 the same data gave T_1 = 8.3 s with the same residual (Section 5.9).

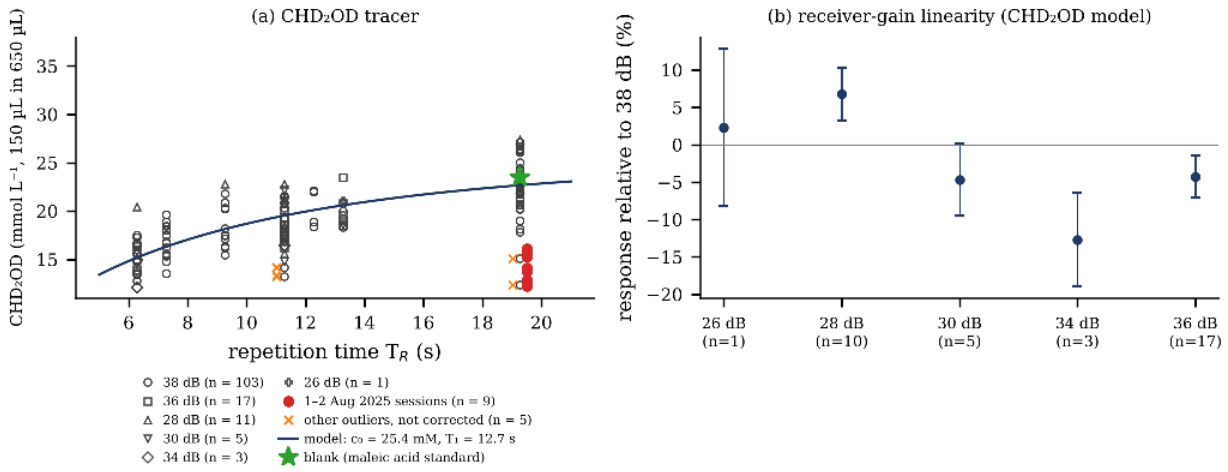

Figure 11. The CHD_2OD tracer. (a) Residual CHD_2OD measured on the absolute scale in the D_2O/CD_3OD residues, normalized to 150 µL of CD_3OD in 650 µL of solution, against the repetition time, with the saturation-recovery model (c_0 = 25.4 mmol L^{-1}, T_1 = 12.7 s) and the value measured in the blank (star). The nine spectra of the 1–2 August 2025 sessions (red) and the 5 other outliers that were flagged but not corrected (orange crosses) are marked. (b) Response at receiver gains other than 38 dB relative to the 38 dB model (mean ± standard error).

The residual scatter of about 10 % deserves interpretation, because it bounds the reproducibility that can be expected from the absolute scale in its present form. It contains the pipetting error of the 150 µL CD_3OD addition and of the D_2O, any error in the assumed total volume, small errors in the fitted area of a quintet whose outer lines are weak, and the departure of the actual receiver gain from its nominal value. The last component was estimated directly from the model.

5.3 Receiver-gain linearity

Fifty-six D_2O spectra were acquired at gains between 18 and 36 dB rather than at the 38 dB of the blank, so their absolute scale depends on the assumption that the receiver gain behaves as a calibrated amplitude scale of 20 dB per decade. The gain terms of the CHD_2OD model test that assumption. Relative to 38 dB,

the responses were +2.3 ± 10.5 % at 26 dB (one spectrum), +6.8 ± 3.5 % at 28 dB (ten spectra), −4.7 ± 4.8 % at 30 dB (five spectra), −12.7 ± 6.2 % at 34 dB (three spectra) and −4.3 ± 2.8 % at 36 dB (seventeen spectra). None of these departures exceeds about two standard errors, and their magnitude is consistent with the 5–10 % deviations between nominal and actual receiver gain that have been reported for other spectrometers [58]. They were not applied as corrections, because they are not much larger than their own uncertainties and because spectra at 18–24 dB, where the extrapolation is longest, could not be tested; instead, an uncertainty component of ±5 % is carried for every spectrum acquired at a gain other than 38 dB. Acquiring the external standard at the gain used for each batch would remove this term entirely.

5.4 An audit of solvent composition

The CHD_2OD signal answered a second question that had not been asked. The prepared solvent system of each residue was taken from the spectrum title, and 32 titles described a solution in D_2O alone. Nineteen of those spectra nevertheless contained CHD_2OD at 9.6–21.3 mmol L^{-1}, the same range as the residues labeled D_2O/CD_3OD, and therefore contained CD_3OD; one spectrum labeled D_2O + CD_3OD (Bushmills Black Bush) contained none. After correction only 14 residues had been dissolved in D_2O alone, and every analysis in this paper uses the corrected assignment. The correction mattered for chemical-shift work. In the genuinely CD_3OD-free residues, the carbohydrate signals lie about 0.03 ppm downfield of their positions in D_2O/CD_3OD (α-glucose H-1 +0.027, β-glucose H-1 +0.033, sucrose Fru H-3 +0.029 and sucrose Glc H-1 +0.015 ppm), the gallic acid singlet lies 0.021 ppm downfield and the syringyl singlet 0.003 ppm downfield, whereas the aliphatic signals of ethanol, the fusel alcohols and lactic acid do not move by more than 0.003 ppm. A global offset applied to all D_2O-only residues, which was considered in the preliminary pass of this work on the basis of the mislabeled spectra, would therefore have misaligned the aliphatic region while correcting the carbohydrate region. The shifts are a medium effect of the 23 % CD_3OD on hydrogen-bonded CH–O protons, not a referencing artifact, and they were handled with analyte-specific windows in the targeted quantitation. Because the fourteen pure-D_2O residues are a small and unbalanced subset (twelve of them belong to two acquisition batches), they were excluded from the chemometric models of the D_2O block and examined separately.

5.5 An acquisition-session anomaly detected by the internal tracer

The CHD_2OD tracer also exposed a fault that no other check would have found. Every one of the nine D_2O/CD_3OD spectra acquired on 1 and 2 August 2025 carried only 0.54–0.71 of the CHD_2OD expected from the model, whereas the spectra of all other sessions scattered around unity (median 0.99). Two observations place the cause in the acquisition rather than in the solvent. First, the TMSP of the same tubes, which enters with the D_2O and not with the CD_3OD, was 27–35 % below the value expected from the D_2O stock (about 2.5 mmol L^{-1} of TMSP) used for the residues measured later that month and in September, so both tracers were depressed together, whereas a different lot or volume of CD_3OD would have changed only one of them. Second, one bottle (sample 343, Kavalan Distillery Reserve Rum Cask) had been freeze-dried and measured twice, in March 2025 and on 1 August 2025, and every non-volatile congener of the August residue was close to half of its March value (non-volatile proton inventory 0.50 of the March value), in proportion to its CHD_2OD deficit. A loss of probe sensitivity during those two sessions, for example from a detuned probe or a pulse that delivered a smaller flip angle than recorded, is the most likely explanation, but the parameter sets record nothing unusual and the cause could not be established. The nine spectra were therefore rescaled by their own CHD_2OD tracer (factors 1.41–1.86), which brought the two preparations of sample 343 into agreement within 9 % in non-volatile proton inventory. Five further spectra from other sessions deviated from the tracer model by more than three residual standard deviations (0.55–0.72 of the expected CHD_2OD) without a session pattern; because a smaller CD_3OD addition or a solvent

of higher isotopic purity would produce the same deviation without affecting the analytes, they were flagged but not corrected. Without a tracer present in every tube, nine residues would have been reported at only 0.54–0.71 of their composition, and the least sweetened cognac of the data set would have been an artifact of the instrument.

5.6 TMSP and residual $CHCl_3$

TMSP could not serve as an internal standard. Its concentration differed between preparation batches by more than an order of magnitude (from a trace of DSS in some early preparations to about 2.1 mmol L^{-1}, the value found in the blank), and its methyl signal is broadened and shifted in acidic residues. It was used for referencing only. In the $CDCl_3$ set the residual $CHCl_3$ of the solvent gave a median of 20.4 mmol L^{-1} with an interquartile range of 18.9–22.0 mmol L^{-1} after normalization, consistent with the 0.2 % residual protiation of 99.8 % D chloroform (about 25 mmol L^{-1}, less the partial saturation of a long-T_1 signal) and with an absolute scale that transfers to the $CDCl_3$ spectra within about ±10 %. One residue (Kavalan Sky Gold) carried an order of magnitude less $CHCl_3$ while its solute signals were normal; a chloroform of higher isotopic purity is the most likely explanation, but it could not be confirmed from the data. That figure also bounds a systematic difference between the two solvents that cannot be removed with the present data: the external standard was measured in D_2O/CD_3OD, and the probe tuning, the 90° pulse and the coil filling factor of a $CDCl_3$ sample differ from those of an aqueous one. All $CDCl_3$ concentrations therefore carry an additional ±10 % component.

5.7 Uncertainty budget

One component of the budget is not visible in the fit statistics. The model of every target signal is a sum of Lorentzian lines, and an experimental line that is partly Gaussian, because of residual field inhomogeneity or unresolved couplings, is fitted by a Lorentzian whose tails carry area that the line does not have. The size of this effect was measured by comparing the fitted areas of isolated singlets with their direct integrals over a window of a few line widths, the integrals being divided by the fraction of each fitted Lorentzian that falls within the window so that the truncation of the tails did not bias the comparison. For the maleic acid line of the blank the fitted area exceeded the integral by 2 %. For the HMF formyl singlet and the gallic acid H-2/6 singlet the excess was 5 % and 6 % (medians over 164 and 155 residues), and for the partly resolved pair of benzaldehyde formyl singlets and the broader syringyl singlet at δ 7.30 it was 12 % and 15 % (106 and 131 residues). Because the analyte areas are divided by the area of the maleic acid line, only the excess relative to the standard enters the result: about 3 % for sharp singlets and 10–13 % for broadened or partly resolved lines. The excess was carried in the budget as a rectangular component rather than applied as a correction, because the direct integral is itself exposed to baseline curvature and to the wings of neighboring signals, and the two estimates of the area bracket the true value rather than define it.

The components identified above are combined in Table 4. For a well-resolved analyte in a D_2O/CD_3OD residue acquired at 38 dB with a recorded solution volume and quantified from a signal with a short T_1, the expanded uncertainty (k = 2) of the absolute concentration is about 9 %, made up of the response factor (1.5 %), the fitted area, the solution volume and smaller terms. It rises to 13 % for a spectrum acquired at another gain, 18 % for an assumed solution volume and 21 % for both, to 17 % for a broadened or partly resolved signal, and to about 40 % for the formyl signals of the aldehydes in spectra acquired at the shortest repetition times, where an error of a factor of two in the assumed T_1 changes the saturation correction by 14–22 %. The residual scatter of the CHD_2OD model (about 10 %) is larger than this budget predicts for a single signal, but it is not a component of it: the CHD_2OD concentration depends directly on the volume and on the isotopic purity of the CD_3OD added, which may have differed between solvent lots, whereas the analyte concentrations depend only on the total volume. It is an upper bound on the reproducibility of the

absolute scale. Values reported as upper bounds, those flagged for overlap and the semi-volatile congeners, whose recovery through freeze-drying is not known, are outside this budget.

Table 4. Uncertainty budget for absolute concentrations (relative standard uncertainties).

Component	Relative standard uncertainty (%)	Basis
Maleic acid mass (2.55 mg) and blank volume (650 µL)	1.3	balance ± 0.02 mg; volumetric dispensing ± 1 %
Maleic acid T_1 (4.5–6.5 s)	0.50	range of Section 5.1, rectangular distribution
Maleic acid line area	0.5	Lorentzian fit of the blank
Response factor R, combined	1.5	root sum of squares of the three components above
Effective content of the maleic acid standard (95 % of the weighed amount)	not estimated	author's determination (Section 5.9); not included in the combined values below
Flip angle (55.4°: 8 µs pulse, 13 µs 90° pulse)	1.0	common pulse width for blank and most samples; an error of ±0.5 µs in the 90° pulse changes results by ≤ 0.6 % (Section 5.9)
Solution volume, recorded	2.0	volumetric dispensing
Solution volume, assumed (650 µL)	8.0	range of recorded volumes (600–700 µL)
Spirit aliquot (1.000 mL)	0.5	volumetric pipette
Fitted area of a resolved signal	3.0	fit residuals and noise; larger near the detection limit
Line-shape model (Lorentzian area against direct integral)	2.0 / 7.3	excess over the blank for sharp singlets / broadened or partly resolved lines (Section 5.7); rectangular
Receiver gain other than 38 dB	5.0	CHD_2OD gain model (Section 5.3)
Saturation correction (assumed T_1 doubled or halved)	0–22	analyte- and repetition-time-specific (Table 3)
$CDCl_3$ spectra: transfer of the D_2O/CD_3OD response factor	10	residual $CHCl_3$ (Section 5.6)
DMSO-d_6 spectra: transfer of the D_2O/CD_3OD response factor	10	matched-product recoveries of HMF, benzaldehydes and quercitol (Section 5.8)
Recovery through freeze-drying and redissolution	not estimated	replicates differ by up to a factor of 3.1 (Section 10.2)
Expanded uncertainty (k = 2), resolved short-T_1 signal, 38 dB, recorded volume	9	combined from the rows above
… at another gain / with assumed volume / both	13 / 18 / 21	
… broadened or partly resolved signal (benzaldehyde pair, syringyl singlet)	17	
… aldehyde formyl signals at the shortest repetition times	40	

5.8 The DMSO-d_6 residues: tracers, referencing and transfer of the absolute scale

The DMSO-d_6 spectra were reprocessed from their stored decays and placed on the absolute scale exactly as the D_2O and $CDCl_3$ spectra were, with the response factor of the aqueous maleic acid blank (Equations 1–3). The reconstruction of the processed spectra stored in the MestReNova document was as faithful as for the other sets ($R^2 \geq 0.999996$ for every item, scale factor 1.0000, so these spectra had not been normalized). Three further questions had to be answered before any concentration could be reported: whether the corrections for transients, gain and volume hold in this solvent, which line can serve as the chemical-shift reference, and whether a response factor measured in D_2O/CD_3OD transfers to DMSO-d_6.

No CD_3OD was added to these residues, so the CHD_2OD tracer of Section 5.2 is absent, but the residual DMSO-d_5 of the solvent plays the same part: it is present in every tube at a concentration fixed by the isotopic purity of the solvent lot. Its quintet (δ 2.50, ^2J(H,D) = 1.85 Hz) was fitted in every spectrum, and the concentrations fall into two groups that correspond to two lots of solvent (Figure 12a). The twenty-two residues acquired between 10 and 24 April carried 53.9–72.5 mmol L^{-1} of ^{1}H (median 60.1, coefficient of variation 7.0 %), consistent with 99.9 atom % D, and the seven residues acquired on 6–7 April and in September and October carried 132–170 mmol L^{-1} (coefficient of variation 8.9 %), about two and a half

times as much. Within the first lot the residues acquired at 32 and 34 dB fell within the range of those acquired at 38 dB, and the maple bourbon, the only DMSO-d₆ spectrum acquired at 18 dB, carried 0.97 times the DMSO-d₅ of the rum acquired the previous day at 38 dB from the same lot, which is the first check of the gain scaling at the longest extrapolation that the CHD_2OD model of Section 5.3 could not test. The within-lot scatter of 7–9 % bounds the reproducibility of the acquisition factor and of the dispensed solvent volume in the same way as the 10 % scatter of the CHD_2OD model. Because the isotopic purity of a solvent lot is a specification minimum rather than a measured value, however, the DMSO-d₅ line shows only that the scale is internally consistent within each lot; it cannot establish the scale itself.

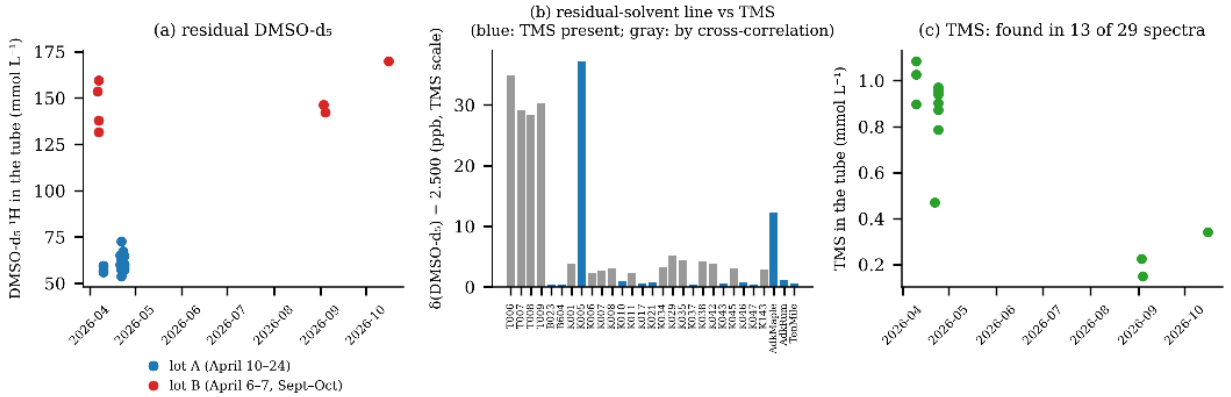

Figure 12. Internal checks of the DMSO-d₆ spectra. (a) Residual DMSO-d₅ on the absolute scale (mmol of ¹H per liter of solution in the tube) against the date of acquisition; the two groups correspond to two lots of solvent. (b) Position of the DMSO-d₅ quintet relative to δ 2.500 on the TMS scale, for spectra that contain TMS (blue) and for spectra referenced by cross-correlation (gray); T, Tuthilltown (Hudson whiskeys); B023, Glen Grant 12 yr; B604 and K, Kavalan; Adk, Adirondack. (c) TMS concentration in the 13 spectra in which it was found.

TMS was found in only 13 of the 29 spectra (Figure 12c), at 0.47–1.09 mmol L^{-1} in the April spectra and 0.15–0.34 mmol L^{-1} in those of September and October, against about 2.2 mmol L^{-1} for the usual 0.03 % (v/v). Its absence followed the session rather than the residue: residues prepared with the first solvent lot and measured on 10 and 24 April contained TMS, whereas all but one of those measured on 21–23 April, and all four measured on 6–7 April, did not. A TMS-free bottle, or the loss of the volatile TMS (b.p. 26 °C) from a repeatedly opened bottle, would explain that pattern better than a reaction with the residue. The disappearance of TMS from DMSO-d₆ solutions of whiskey solids has been observed in this laboratory before, and the present data do not distinguish between its absence from the solvent and its loss from the solution.

The residual DMSO-d₅ line is conventionally placed at δ 2.50 [33], and in eleven of the thirteen spectra that contained TMS it lay at δ 2.5003–2.5011. In the residue of the Kavalan oloroso cask 005, however, it lay at δ 2.537, while every solute line, from the aldehyde formyl singlets to the lipid methylene envelope, kept its usual position relative to TMS (Figure 13); referencing that spectrum to the solvent line moved all of its solute lines upfield by 0.037 ppm. Cross-correlation of the alkyl and formyl regions with the median of the TMS-referenced spectra revealed the same displacement, 0.028–0.035 ppm, in the four Tuthilltown residues, which contain no TMS, and a smaller one, 0.012 ppm, in the sucrose-rich maple bourbon (Figure 12b), and in the two most displaced spectra a minor multiplet near δ 2.49 accompanied the main DMSO-d₅ quintet. The cause was not established; a strong association of the solvent with the acids and sugars of these residues, or a solution that was not homogeneous, are both possible. The consequence, however, is clear: in residues of this kind the residual solvent line cannot be assumed to be a fixed reference. Every DMSO-d₆ spectrum was therefore referenced to TMS where it was present and otherwise by cross-correlation of its

alkyl region (or of its formyl region for the Tuthilltown residues, whose alkyl signals are weak) with the median TMS-referenced spectrum, and the smaller, compound-specific displacements that remain after referencing were handled as described in Section 4.15.

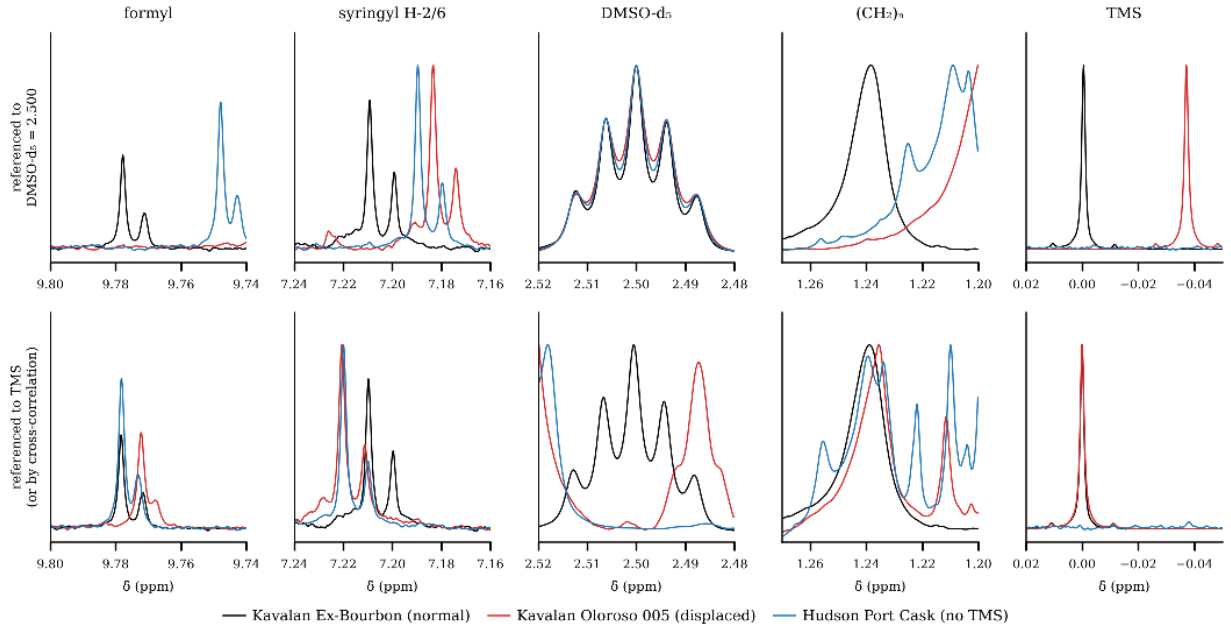

Figure 13. Chemical-shift referencing of the DMSO-d₆ spectra, shown for a residue with the usual behavior (Kavalan Solist Ex-Bourbon), the residue whose DMSO-d₅ line is displaced (Kavalan Solist Oloroso 005, TMS present) and a residue without TMS (Hudson Port Cask Finish). Top: referenced to the DMSO-d₅ quintet at δ 2.500. Bottom: referenced to TMS at δ 0.000 or, without TMS, by cross-correlation with the median of the TMS-referenced spectra. Each trace is scaled to its tallest line within the window, within the common noise ceiling of Section 4.17. Referencing to the solvent line moves every solute line of the two affected residues upfield by 0.03–0.04 ppm; after referencing to TMS the formyl and methylene lines coincide, the DMSO-d₅ quintets of the affected residues fall outside the window shown (the oloroso residue shows only the minor multiplet near δ 2.49), and compound-specific differences of about 0.01 ppm remain in the syringyl region (Section 9.2).

The response factor R was measured in D_2O/CD_3OD, and the calibrated 90° pulse of 13 µs was assumed for all three solvents. By the principle of reciprocity the signal per proton is inversely proportional to the true 90° pulse length, so a solvent that loads the probe differently changes both the flip angle and the receptivity [39], and the 90° pulse was not recalibrated for the DMSO-d₆ residues. For the $CDCl_3$ spectra the transfer could be checked only through the residual $CHCl_3$ (Section 5.6). For DMSO-d₆ a direct check is available, because the same compounds can be measured in D_2O/CD_3OD and in DMSO-d₆ residues of the same products. Three compounds that dissolve completely in both solvents and are quantified from isolated, unambiguous resonances gave median DMSO/D_2O ratios of 1.08 (HMF), 1.11 (the benzaldehydes) and 0.97 (quercitol) over 24 matched products (Section 9.8). The absolute scale therefore transfers to DMSO-d₆ within about 8 % on average. An uncertainty component of ±10 % was nevertheless assigned to individual DMSO-d₆ results, as for $CDCl_3$, because the scatter of the individual ratios (interquartile range 0.92–1.27 for HMF) contains differences between preparations, made months apart from separate aliquots, that cannot be separated from a solvent effect without replicate preparations. The relaxation times in DMSO-d₆, whose viscosity is about twice that of D_2O, were taken to be those assumed for the same protons in D_2O, which overestimates them and makes the saturation corrections conservative; at the common repetition time of 7.27 s and the 48.5° pulse of every DMSO-d₆ spectrum, the steady-state recovery exceeds 0.94 for any T_1 up to 4 s.

5.9 Corrections in version 2.2: the flip angle and the content of the standard

The parameter sets of every spectrum in this work record a 90° pulse (pw90) of 10 µs, and versions 2.0 and 2.1 took that value at face value, so that the 8 µs pulse used for the blank, for every D_2O spectrum and for most $CDCl_3$ spectra was treated as a 72° pulse, and the 7 µs pulse of the eleven Hudson $CDCl_3$ spectra and of every DMSO-d_6 spectrum as a 63° pulse. The 90° pulse of the probe, as calibrated by the author, is 13 µs, so the flip angles were 55.4° and 48.5°. Two terms of the absolute scale depend on the flip angle. The acquisition factor of Equation 1 carries $\sin\theta$, which lowers K_{blank} from 26.63 to 23.04 and, with the recovery term below, raises the response factor R by 14.3 %; but the same sine enters the acquisition factor of every spectrum acquired with the 8 µs pulse, so that for all the D_2O residues and for most of the $CDCl_3$ residues it cancels between Equations 1 and 2. It does not cancel for the 7 µs spectra, for which the ratio of the two sines is 0.910 rather than the 0.937 implied by the nominal calibration. The second term is the steady-state recovery of Equation 4, which approaches unity as the flip angle decreases. The recovery of maleic acid in the blank rises from 0.9703 to 0.9812, which raises the absolute spectra of the 8 µs acquisitions by 1.1 % once the sine terms have cancelled and those of the 7 µs acquisitions by 4.2 %, while the saturation correction applied to each analyte becomes smaller, most of all for slowly relaxing protons in spectra acquired with short repetition times.

The second correction concerns the standard itself. The blank was prepared from 2.55 mg of maleic acid in 650 µL of solution, and versions 2.0 and 2.1 used its gravimetric concentration of 33.80 mmol L^{-1} in Equation 2. The author has since established that the standard contained 95 % of its nominal maleic acid content, so that its effective concentration was 32.11 mmol L^{-1}. Because c_{MA} enters R as a single factor, every absolute intensity and every concentration of the earlier versions was too high by a factor of 100/95, or 5.3 %, whatever the solvent, the analyte or the acquisition, and all of them have been multiplied by 0.95. Unlike the flip angle, this correction leaves every ratio between signals, every saturation correction and the shape of every spectrum unchanged. Nor can it be seen in the blank. Once the foot of the main line, its ^{13}C satellites and its spinning sidebands are excluded, the largest signal in the olefinic region (δ 5.6–7.4) is a singlet at δ 6.502 with 0.12 % of the height of the maleic acid line, so the deficit is not maleic acid that isomerized to fumaric acid in the solution; it must lie in the assay of the reagent or in the preparation of the blank, which only a gravimetric check or a certified reference material can establish. A deficit of this size is also nearly four times the 1.3 % allowed for the mass and volume of the blank in the uncertainty budget (Table 4), which shows how completely the absolute scale depends on a single preparation.

Table 5. Change of the reported concentrations caused by the two corrections of version 2.2, the flip angle (90° pulse 13 µs instead of the nominal 10 µs) and the content of the maleic acid standard (95 % of the weighed amount), by solvent and by the T_1 assumed for the quantified protons. Quantified values only.

Solvent	T_1 assumed (s)	Analytes (most frequent first)	Values	Flip angle only: median (range), %	Version 2.2 against 2.1: median (range), %
D_2O/CD_3OD	1.0	quercitol	189	+1.1 (+1.1 to +1.3)	−3.9 (−4.0 to −3.8)
D_2O/CD_3OD	1.5	fructose, sucrose	351	+1.1 (+0.7 to +1.3)	−3.9 (−4.3 to −3.8)
D_2O/CD_3OD	2.0	glucose, α-glucose H-1, succinic acid, 1,2-propanediol	723	+1.0 (0.0 to +1.3)	−4.0 (−5.0 to −3.8)
D_2O/CD_3OD	2.3	glycerol	151	+0.9 (−0.6 to +1.3)	−4.1 (−5.6 to −3.8)
D_2O/CD_3OD	3.0	acetic acid, ethanol, fusel alcohols, lactic acid, syringyl ArH, isoamyl alcohol	977	+0.5 (−2.2 to +1.2)	−4.5 (−7.1 to −3.8)
D_2O/CD_3OD	4.0	benzaldehydes, HMF, gallic acid, cinnamaldehydes	979	−0.5 (−4.7 to +1.1)	−5.4 (−9.5 to −4.0)
$CDCl_3$	2.0	diethyl succinate, 2-phenylethanol	68	+0.9 (+0.4 to +3.8)	−4.2 (−4.6 to −1.4)
$CDCl_3$	3.0	syringaldehyde ArH, fusel alcohols, acetic acid, cinnamaldehydes, sinapaldehyde	300	−0.1 (−1.3 to +2.7)	−5.1 (−6.2 to −2.4)
$CDCl_3$	4.0	furfural, syringaldehyde, vanillin, aldehyde δ 9.81	209	−1.5 (−3.4 to +1.2)	−6.5 (−8.2 to −3.8)
DMSO-d_6	1.0	quercitol	29	+4.2 (+4.2 to +4.2)	−1.1 (−1.1 to −1.1)
DMSO-d_6	2.0	α-glucose H-1, glucose, succinic acid, pentoses, sucrose	201	+3.6 (+3.6 to +3.6)	−1.6 (−1.6 to −1.6)
DMSO-d_6	3.0	syringyl ArH, ethanol, gallic acid	186	+2.2 (+2.2 to +2.2)	−3.0 (−3.0 to −3.0)

Solvent	T_1 assumed (s)	Analytes (most frequent first)	Values	Flip angle only: median (range), %	Version 2.2 against 2.1: median (range), %
DMSO-d₆	4.0	HMF, vanillin, acetic acid, syringaldehyde	144	+0.3 (+0.3 to +0.3)	−4.7 (−4.7 to −4.7)
DMSO-d₆	5.0	ellagic acid	9	−1.5 (−1.5 to −1.5)	−6.5 (−6.5 to −6.5)

Change = value in version 2.2 / value in version 2.1 − 1; the flip-angle column divides out the content of the standard (factor 0.95). Within a row the change depends on the repetition time of each spectrum (6.27–19.27 s in D_2O/CD_3OD and $CDCl_3$, 7.27 s in DMSO-d₆) and, in $CDCl_3$, on the pulse width (8 µs, or 7 µs for the eleven Hudson spectra). D_2O values include the recalculated session factors (Section 5.5). The 13 multicomponent fits that converged to a different solution (Section 5.9) are not included.

The net effect of the two corrections on the reported concentrations is set out in Table 5, by solvent and by the T_1 assumed for the quantified protons, because those two quantities, together with the repetition time of each spectrum, determine it completely. In the D_2O/CD_3OD residues the concentrations fell by 3.8–9.5 % (median 4.2 %), in the $CDCl_3$ residues by 1.4–8.2 % (median 5.1 %) and in the DMSO-d₆ residues by 1.1–6.5 % (median 3.0 %). The flip angle alone would have changed them by −4.7 to +1.3 %, −3.4 to +3.8 % and −1.5 to +4.2 % respectively, and the content of the standard lowered every one of them by a further 5.0 %. The largest decreases belong to the formyl protons of HMF and the cinnamaldehydes, assigned a T_1 of 4 s, in the spectra acquired with a relaxation delay of 2 s, and the smallest to rapidly relaxing protons in the 7 µs spectra, whose sine term rose with the correction. An error of ±0.5 µs in the calibrated 90° pulse itself would move the corrected values by no more than 0.6 % (median 0.1 %), because the sine terms cancel and the recovery terms change slowly at these angles.

Every concentration was recalculated from the corrected spectra by the complete quantitation chain rather than rescaled as a number, and each new value was then compared with the value predicted from the closed-form dependence of Equations 1, 2 and 4 on the flip angle and on the content of the standard. For 4516 quantified values the largest departure from that prediction was 0.09 %, which confirms that nothing else entered the recalculation. In thirteen multicomponent fits in twelve residues (ten of the formyl region, in which HMF, the two cinnamaldehyde doublets and the furanic envelope are fitted together, two of the lactic acid doublet with the isoamyl methylene quartet, and one of the syringyl region), the refinement on the rescaled spectrum converged to a slightly different solution with a better coefficient of determination, and that solution was kept. The reported quantities of those fits lie within 5 % of the prediction, but the apportionment between their minor components is evidently not unique, and four detection decisions changed with it, for the second cinnamaldehyde doublet and the cinnamaldehyde total in two residues and for the syringyl singlet in a third, all close to their detection limits. That sensitivity is a property of the models rather than of the corrections, and it bounds how far the individual minor components of such fits should be trusted. In one DMSO-d₆ fit the refinement converged to a poorer solution, and the previous solution, rescaled, was retained.

The reporting thresholds expressed in mg L^{-1} of spirit, namely the matrix-limited reporting level of fructose, the fructose, glucose and sucrose limits that decide how glycerol is reported and the tolerance for the agreement of the two glycerol multiplets (Section 4.8), describe where one signal begins to obscure another rather than a property of the spirit. They were set on the absolute scale of version 2.1 and have been carried with the corrected scale, that is, multiplied by 0.961, so that the fructose reporting level is now 28.8 mg L^{-1}. Applied unchanged, the round values would have moved four glycerol decisions and five fructose flags across their limits for reasons that have nothing to do with the spectra; with the thresholds carried, no flag changed.

The internal checks of this paper can confirm neither correction, and that limitation should be stated plainly. The CHD_2OD tracer of Section 5.2 was acquired with the same pulse as the blank, so that sin θ cancels, and the dependence of its intensity on the repetition time is described equally well by a shorter T_1 at a larger flip angle or by a longer T_1 at a smaller one. Refitted at 55.4°, the tracer model gives T_1 = 12.7 s, against

8.3 s in version 2.1, with c_0 = 25.4 mmol L^{-1} (25.4 mmol L^{-1} before) and a residual standard deviation that changed by only 0.5 %, and over flip angles from 30° to 90° the residual sum of squares changes by no more than 3.1 %. The residual $CHCl_3$ of the 7 µs and 8 µs $CDCl_3$ spectra cannot separate the two calibrations either, because the 7 µs spectra form a single preparation batch with its own solvent lot. No internal signal can test the content of the standard at all: CHD_2OD, TMSP and the residual solvent lines are measured on the scale that the standard defines, and none of their own concentrations is known independently to the 5 % in question. The transfer to DMSO-d_6, in which every spectrum was acquired with the 7 µs pulse, responds to the flip angle directly. The median DMSO-d_6/D_2O ratios of HMF, the benzaldehydes and quercitol moved from 1.04, 1.07 and 0.94 to 1.08, 1.11 and 0.97 (Section 5.8), a mean departure from unity of 8 % against 6 % before, so the cross-solvent comparison does not favor the corrected calibration; a 90° pulse in DMSO-d_6 shorter than in D_2O/CD_3OD, which a different loading of the probe could produce, would raise the DMSO-d_6 values in the same way, and residues prepared from separate aliquots months apart cannot decide between these explanations. Both corrections therefore rest on the pulse calibration and the determination of the standard made by the author, and only a nutation measurement recorded with each data set in each solvent, together with a standard traceable to a certified reference material, can remove that dependence (Section 17).

None of the conclusions of this paper depended on either error. The content of the standard scales every spectrum and every concentration by the same factor, which mean-centering and scaling reduce to a change of units in every chemometric block. The flip angle changes the binned D_2O/CD_3OD and DMSO-d_6 blocks by a constant factor as well, apart from the small changes of the session factors of Section 5.5, whereas the eleven 7 µs $CDCl_3$ residues gain 3.0 % relative to the others and the quantified-congener blocks change by analyte- and repetition-time-specific factors (Table 5), which act on logarithms and are largely removed by mean-centering. Every classification, calibration and chemometric analysis of Sections 11–13 was nevertheless recomputed. The balanced accuracies changed by at most 0.034 and the cross-validated Q^2 of the PLS calibrations by at most 0.015; the explained variances changed by at most 0.2 percentage points, the adjusted Rand indices with the spirit categories by at most 0.09 and the cophenetic correlations by at most 0.13, mostly in the complete-linkage trees and the autoscaled congener block, and the trustworthiness and class purity of the t-SNE and UMAP displays, which are sensitive to the numerical details of the optimization, by at most 0.07. The matched-product comparisons of set S1 did not change, whereas in set S2 the partition agreements of the blocks containing the $CDCl_3$ spectra moved by up to 0.50, because the four Tuthilltown products of that set, whose $CDCl_3$ spectra were acquired with the 7 µs pulse, gained 3.0 % relative to the six Kavalan malts, which is enough to move a product from one cluster of a ten-product tree to another. That is the fragility of hierarchical partitions on small sets described in Section 12.9, and none of the statements of Sections 11–13 rests on those values.

6. The polar fraction: spectra in D_2O and D_2O/CD_3OD

6.1 General appearance on an absolute scale

Placed on a common absolute footing, the D_2O spectra differ from one another by more than two orders of magnitude in total proton content, a fact that was invisible in version 1.0. Integrated over the retained windows, the residue of a median spirit carried 54.1 mmol of carbon-bound ¹H per liter of spirit, the tenth and ninetieth percentiles being 25.9 and 364 mmol L^{-1} and the extremes 3.7 mmol L^{-1} (a sotol) and 1 502 mmol L^{-1} (a maple-flavored bourbon). The class medians separate into three groups that correspond to three kinds of residue (Figure 14). Brandies (411 mmol L^{-1}), Canadian whiskies (94.4 mmol L^{-1}) and rums (92.2 mmol L^{-1}) yield residues whose proton inventory is dominated by carbohydrate; the whisky classes fall between 30.3 and 73.6 mmol L^{-1} and show a dense but moderate CH–O envelope between δ 3.3 and 4.1 with sharp aldehyde, aromatic and cyclitol lines; and blanco or lightly aged agave spirits, the sotols and the gin give residues of about 20 mmol L^{-1} (class median 19.8 mmol L^{-1} for the agave spirits), in which retained volatiles and a few acids make up a large share of what is seen. These figures are expressed per liter of spirit and therefore compare spirits directly; they are not the masses of the residues, which were not recorded, but the integrated proton inventory is the quantity that a gravimetric dry-extract measurement would have to be reconciled with.

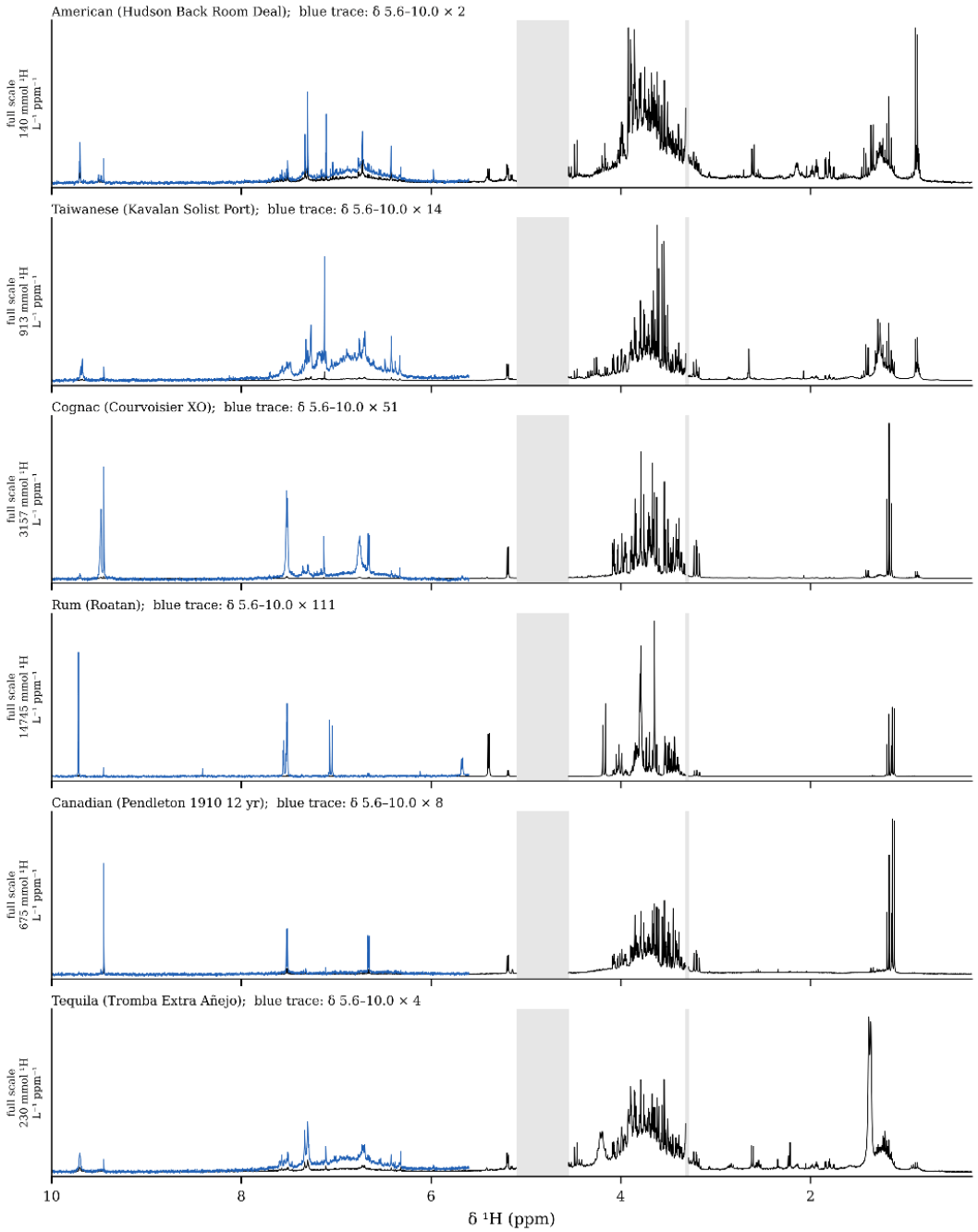

Figure 14. Absolute ¹H spectra (299.99 MHz) of six D_2O/CD_3OD residues representing American whiskey, Taiwanese single malt, cognac, rum, Canadian whisky and tequila, each scaled to its own maximum (full scale given at left in mmol of ¹H per liter of spirit per ppm). The region δ 5.6–10.0 is also shown magnified (blue) by the factor given in each title, a factor limited so that the magnified noise stays below the common ceiling of Section 4.17; the HDO region and the CHD_2OD line are masked.

The appearance of the spectra follows the same division. In the oak-dominated residues the two formyl lines of the benzaldehydes near δ 9.70, the HMF singlet at δ 9.45, the syringyl, gallic and lignan singlets between δ 6.4 and 7.3, the sharp ddd and dt of quercitol at δ 1.80 and 1.94 and the methyl doublets of lactic acid and the retained fusel alcohols are the most conspicuous features above a broad aromatic and carbohydrate background. In the cognacs the α- and β-glucose anomeric doublets and the β-fructofuranose multiplet dominate, with HMF and a broad furanic envelope beneath it at δ 9.48. The sweetened rums are

dominated by sucrose or by invert sugar, and three products (Section 6.3) carry the vanillin ring pattern and 1,2-propanediol of a flavor preparation.

6.2 Assignments

The assignments are collected in Table 6. They rest on the chemical shifts and coupling patterns at 300 MHz, on the gHSQC, DEPT-45 and homonuclear *J*-resolved spectra of a concentrated Kavalan residue (Section 8), on statistical correlation across the data set (Section 6.8) and on reference data for the individual compounds [33,59–61]. Each assignment carries a status: *confirmed* where the spin system was identified completely or a heteronuclear correlation was observed, *probable* where the shift and multiplicity agree with reference data and with the chemistry of the sample but no second correlation was available, and *tentative* where the assignment is the most plausible of several. Positions are given for D_2O/CD_3OD (23 % v/v CD_3OD) at 27 °C; the analyte-specific offsets for residues in D_2O alone are listed in Section 5.4.

Table 6. ¹H assignments for the freeze-dried residues in D_2O/CD_3OD (23 % v/v CD_3OD, 27 °C, TMSP = 0). δC from the gHSQC spectrum of the concentrated Kavalan residue, corrected by −0.55 ppm (Section 8).

δH (ppm)	Multiplicity (*J*/Hz)	Assignment	δC	Status	Basis
9.697–9.703; 9.707–9.713	s; s (Δδ ≈ 0.01)	Syringaldehyde CHO (upfield) and vanillin CHO (downfield); reported as their sum	—	probable	upfield line co-varies with syringyl H-2/6; vanillin CHO position in flavored products (Section 6.3); $CDCl_3$ concordance [1,59]
9.713–9.722; 7.54–7.56; 7.50–7.53; 7.02–7.07; 3.92	s; d (1.9); dd; d (8.0); s	Vanillin CHO, H-2, H-6, H-5 and OCH_3 without a syringyl partner (flavor preparations)	—	confirmed	complete ring system; OCH_3:CHO = 3:1 (Section 6.3) [59]
9.474–9.491	two d, *J* ≈ 7.7	Coniferaldehyde and sinapaldehyde CHO (*E*-cinnamaldehydes)	—	probable	trans-cinnamaldehyde CHO doublets [1]
9.47–9.50	broad (6–45 Hz)	Furanic/Maillard oligomer envelope (proposed caramel index)	—	tentative	distribution across the classes (Section 6.3) [29]
9.450	s	5-Hydroxymethylfurfural (HMF) CHO	—	confirmed	with H-3 7.52 (d, 3.5) and H-4 6.66 (d, 3.6) [4,63]
7.30	s	Syringaldehyde H-2/6 (syringyl index)	107.6	probable	syringyl CH δC ≈ 107 [59]; correlation with the δ 9.692 formyl line (Section 6.8), but syringaldehyde H-2/6 lies at δ 7.20–7.23 in the aqueous reference media, 0.08–0.10 ppm upfield (Section 2.10)
7.33	s	Syringic acid H-2/6	—	tentative	area 0.47 × the δ 7.30 line, as the syringic acid/syringaldehyde ratio in DMSO-d_6; downfield of syringaldehyde H-2/6 in every aqueous medium examined (Section 2.10)
7.10–7.13	s	Gallic acid H-2/6 (pH-dependent)	110.9	confirmed	sharp 2H singlet; *J*-resolved singlet; HSQC [59]
6.33–6.43	s	Aromatic singlets of syringyl lignans (lyoniresinol type)	106.6	tentative	syringyl CH; co-vary with oak extract [17]; singlets, unlike the H-8 doublets (*J* ≈ 16 Hz) of the hydroxycinnamic acids in the same region (Section 2.10)
6.3–7.6	broad envelope	Oligomeric phenolics (ellagitannin- and lignin-derived)	104–116	probable	broad HSQC envelope [2,16]
5.393	d, *J* = 3.8	Sucrose Glc H-1	—	confirmed	with Fru H-3 4.180 (d, 8.6) [59]
5.192	d, *J* = 3.7	α-D-Glucopyranose H-1	93.3	confirmed	*J*-resolved doublet; HSQC [59,60]
5.197	d, *J* ≈ 3.6	β-L-Arabinopyranose H-1 (coincides with α-glucose H-1)	93.4	probable	α-region excess over the α/β ratio of glucose = 0.53–0.59 × α-Arap (Section 9.5) [65]
5.151	d, *J* = 3.6	α-D-Xylopyranose H-1	93.2	probable	shift of the free sugar [59,66]; ratio to β-Xylp (Section 9.5)
5.01	br d	α-L-Arabinofuranosyl H-1 (hemicellulose fragment)	108.8	tentative	δC 108–110 typical of α-Araf C-1
4.599	d, *J* = 7.9	β-D-Glucopyranose H-1	96.9	confirmed	*J*-resolved doublet; HSQC [60]
4.535	d, *J* = 7.8	β-D-Xylopyranose H-1	97.7	probable	HSQC; shift of the free sugar [59,66]

δH (ppm)	Multiplicity (*J*/Hz)	Assignment	δC	Status	Basis
4.474	d, *J* = 7.8	α-L-Arabinopyranose H-1	97.8	probable	HSQC; β-anomer remainder in the α-region (Section 9.5) [66]
4.46	d, *J* ≈ 7.6	β-Glycosidic H-1 (β-xylosyl or β-glucosyl unit)	102.9	probable	δC 102–104: glycosidic, not reducing-end
4.180	d, *J* = 8.6	Sucrose Fru H-3	—	confirmed	[59]
4.058–4.092	m	D-Fructose, β-furanose H-3/H-4 (0.48 H per molecule)	—	probable	tautomer-resolved shifts [61,78]
4.22	q, *J* = 7.1	Ethyl ester OCH_2 (ethyl lactate, ethyl succinates)	63.2	probable	HSQC; paired with CH_3 1.23–1.29
3.93–4.03	m	Quercitol ring CH–O	69.3–69.5	probable	HSQC; co-vary with H-6 (Section 6.8) [19,20]
3.86–3.92	s	Aromatic OCH_3 (vanillin, syringaldehyde, lignans)	57.0	confirmed	HSQC
3.85; 3.53/3.43; 1.129	m; dd; d (6.3)	1,2-Propanediol CH, CH_2, CH_3	—	confirmed	complete spin system [59]
3.77; 3.626; 3.533	m; dd (11.7, 4.4); dd (11.7, 6.4)	Glycerol H-2, H-1b/H-3b, H-1a/H-3a	72.7; 63.3	confirmed	HSQC; both CH_2 multiplets used (Section 6.5) [59]
3.552 / 3.513	d, *J* = 11.7	D-Fructose β-pyranose H-1b (coincides with glycerol H-1a/H-3a)	—	confirmed	Section 6.5 [61]
3.64	q, *J* = 7.1	Ethanol CH_2 (retained)	—	confirmed	[33]
3.302	quintet, *J*(H,D) = 1.65	CHD_2OD (residual of CD_3OD)	49.2	reference	internal tracer (Section 5.2) [33]
3.22	dd, *J* = 7.9, 9.3	β-D-Glucopyranose H-2	74.9	confirmed	[60]
2.85	t, *J* ≈ 7	2-Phenylethanol $ArCH_2$ (retained)	—	probable	[59]
2.61–2.65	s	Succinic acid CH_2 (pH-dependent)	29.8	confirmed	*J*-resolved singlet; HSQC [59]
2.00–2.12	s	Acetic acid / acetate CH_3 (pH-dependent)	—	confirmed	[59]
1.935	m (dt)	Quercitol H-6eq	34.5	confirmed	shares one CH_2 carbon with H-6ax (HSQC) [19,20]
1.797	ddd, *J* = 13.9, 11.3, 3.2	Quercitol H-6ax	34.5	confirmed	geminal partner of 1.935 (HSQC, *J*-resolved, STOCSY)
1.426; 1.64–1.68; 3.61–3.63	q (6.8); m; t	Isoamyl alcohol CH_2, CH and CH_2O (retained)	—	probable	shift and multiplicity; STOCSY with the CH multiplet and CH_3 doublets (Section 6.8)
1.34–1.40	d, *J* = 6.9	Lactic acid CH_3 (pH-dependent)	20.3	confirmed	*J*-resolved doublet; HSQC [59]
1.23–1.29	t, *J* ≈ 7	Ethyl ester CH_3	14.1–14.7	probable	HSQC
1.172	t, *J* = 7.1	Ethanol CH_3 (retained; excluded from the models)	—	confirmed	[33]
0.884	d, *J* = 6.7	Isoamyl and isobutyl alcohol CH_3 (retained)	—	probable	fusel-alcohol doublets
0.000	s	TMSP or DSS (reference)	—	reference	

Positions are medians or ranges of the fitted positions in the D_2O/CD_3OD residues; in D_2O alone the carbohydrate CH–O signals lie 0.015–0.033 ppm and gallic acid 0.021 ppm downfield (Section 5.4).

6.3 Aldehydes, furanics and phenolic acids

Two sharp formyl singlets are present in almost every oak-aged residue, near δ 9.700 and 9.709 (interquartile ranges of the fitted positions 9.697–9.703 and 9.707–9.713). They are separated by only about 0.01 ppm, which at 299.99 MHz is about 3 Hz, and a two-line fit of every D_2O/CD_3OD residue showed which ring signals each line follows. The area of the upfield line co-varied with the syringyl ring singlet at δ 7.30 (regression coefficient through the origin 1.46 protons per formyl proton against the 2.0 required by the structure, over 180 residues, and $r = 0.95$ between the spectral intensities in the statistical correlation of Section 6.8). The downfield line followed the vanillin H-5 doublet near δ 7.04 only weakly ($r = 0.35$; coefficient 0.76 against 1.0), because that doublet lies on the broad phenolic envelope and because the two aldehydes are released together and co-vary across the residues; its assignment to vanillin rests instead on the position of the single formyl line of the three residues that contain vanillin without syringaldehyde (δ 9.713–9.722, below) and on the resolved $CDCl_3$ spectra. The upfield line is therefore assigned to syringaldehyde and the downfield line to vanillin. Both coefficients fall short of their theoretical values, and the apportionment of the formyl intensity between the two lines varied from residue to residue in a way

that the resolved $CDCl_3$ spectra (Section 7.3) do not support, so the two aldehydes are reported in D_2O only as their sum, "benzaldehydes". The separation of the two aldehydes is unambiguous in $CDCl_3$ (Section 7.3), which is the solvent of choice for the individual compounds. The reference spectra of Section 2.10 qualify the assignment of the δ 7.30 singlet: they place syringaldehyde H-2/6 0.08–0.10 ppm upfield of it in both aqueous media examined, so that the assignment rests on its correlation with the formyl line and on its ^{13}C shift rather than on its position.

Three residues depart from this pattern in a way that matters for authentication. Roatan rum, Rum 28 and TX American Blended Whiskey each show a single, intense formyl line at δ 9.713–9.722, the complete vanillin ring system (H-2 at δ 7.54–7.56 with $J \approx 1.9$ Hz, H-6 as a double doublet at δ 7.50–7.53 and H-5 as a doublet at δ 7.02–7.07 with $J = 8.0$ Hz) and a methoxy singlet at δ 3.92 whose integral is three times that of the formyl proton, but no syringyl singlet at δ 7.30 (Figure 15). Oak lignin releases syringaldehyde in at least the same molar quantity as vanillin, and every oak-aged residue of this study that contained vanillin also contained syringaldehyde, with a syringaldehyde-to-vanillin molar ratio between 0.85 and 3.25 in the $CDCl_3$ spectra (Section 7.3). Vanillin without syringaldehyde is the signature of vanilla flavoring, either vanillin itself or an extract, and two of the three products also contained 1,2-propanediol at 1179 and 819 mg L^{-1}, the carrier solvent of many flavor preparations. Roatan rum, in addition, contained 19.2 g L^{-1} of sucrose. None of this is a regulatory finding in itself (Section 10.6), but the combination of vanillin without syringaldehyde and a glycol carrier is as clear a chemical fingerprint of a flavor preparation as a ^{1}H spectrum can provide, and it is visible without any reference to a model or a data base. The formyl line of vanillin in these products lies 0.004–0.013 ppm downfield of its median position in the oak-aged residues, a medium effect of their very different matrices (one of them is dominated by sucrose and 1,2-propanediol); the identification rests on the complete ring system, not on the formyl shift.

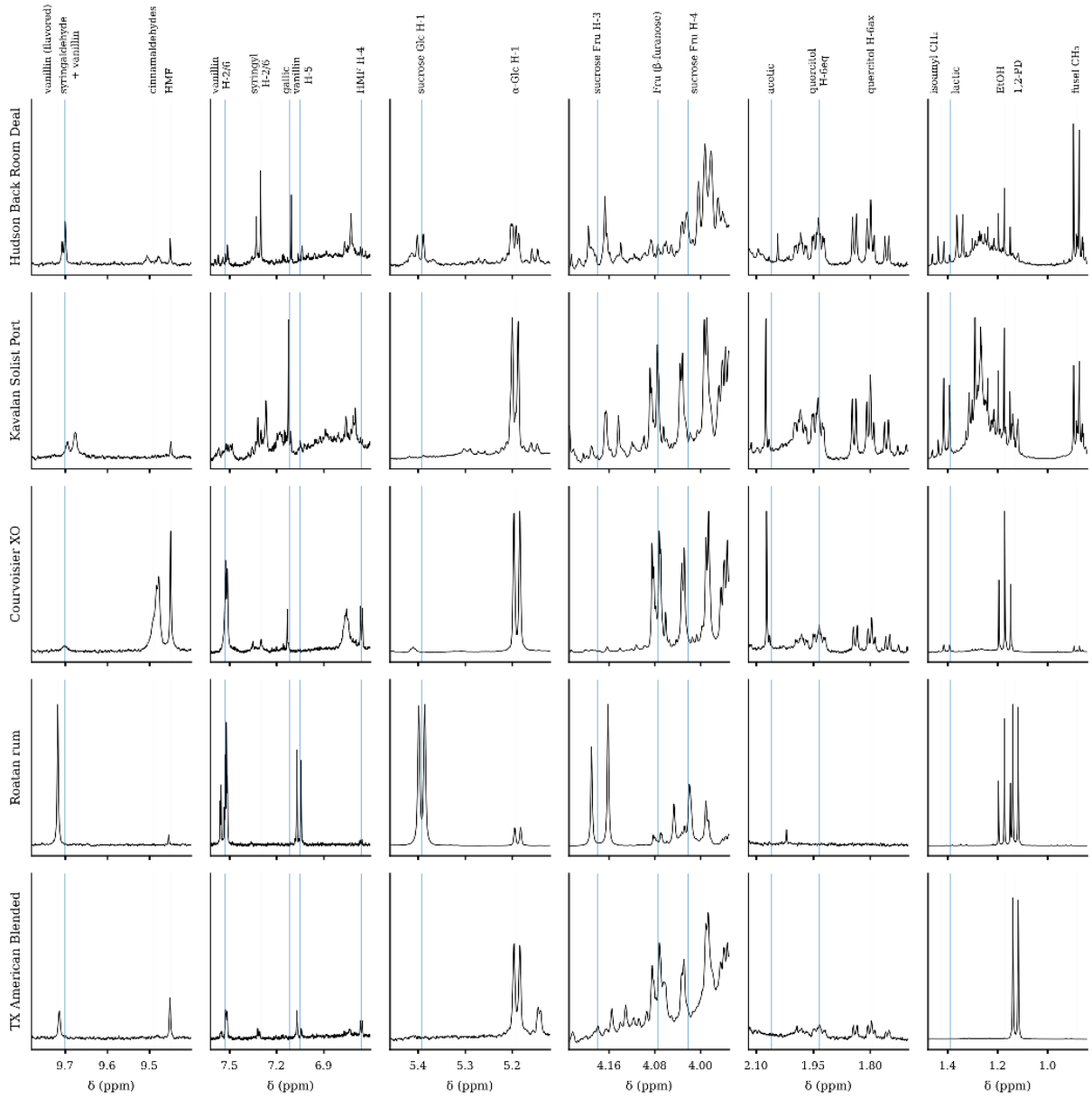

Figure 15. Expansions of the D_2O/CD_3OD spectra of five residues chosen to display the diagnostic signals: the formyl region (benzaldehydes, cinnamaldehydes and HMF), the aromatic region (vanillin, syringyl, gallic acid and HMF H-4), the anomeric region (sucrose and α-glucose H-1), the fructose and sucrose signals near δ 4.0–4.2, the quercitol and acetate region and the aliphatic region (isoamyl alcohol CH_2, lactic acid, ethanol, 1,2-propanediol and fusel alcohol methyls). Each trace is scaled, window by window, to its tallest line unless that would raise its displayed noise above the common ceiling of Section 4.17, so that a weak residue is drawn with small lines over an ordinary baseline rather than with an expanded one; the vertical lines mark the reference positions used in the fits.

5-Hydroxymethylfurfural was identified by its formyl singlet at δ 9.450 together with the ring doublets at δ 7.52 and 6.66 (^3J = 3.5–3.6 Hz) and was found in 181 of 183 D_2O/CD_3OD residues, with a median of 2.6 mg L^{-1} of spirit and a maximum of 45.7 mg L^{-1}. HMF forms during toasting and charring of the cask, during the heating of sugars and in caramel color, so its presence alone is not diagnostic. What distinguishes caramel is the broad furanic envelope that underlies the HMF line in some residues, a Lorentzian feature centered near δ 9.48, whose fitted width was typically 6–16 Hz (interquartile range), that was modeled explicitly in the quantitation. Its integral, expressed as micromoles of formyl-type protons per liter, was

largest in the cognacs (class median 108), the Canadian whiskies (25.5) and several rums, and an order of magnitude lower in the American whiskeys (12.0). The envelope is attributed to oligomeric furanic and Maillard-type material of the kind that constitutes plain caramel [29]; it is proposed here as a caramel index that can be measured without chromatography, with the caveat that it has not yet been calibrated against products of known caramel content. Separating this envelope from the formyl doublets of coniferaldehyde and sinapaldehyde was essential. A fit that ignored it attributed the envelope to the cinnamaldehydes and returned a brandy-class median above 40 mg L^{-1}, which is chemically implausible for brandies matured largely in refill casks; with the envelope modeled, the D_2O cinnamaldehyde total fell to 5.6 mg L^{-1} in the brandies and was highest, as expected, in the American whiskeys matured in new charred oak (class median 5.9 mg L^{-1}).

Gallic acid gives a sharp two-proton singlet whose position moves between δ 7.10 and 7.14 with the acidity of the residue. It was detected in 175 D_2O/CD_3OD residues, with class medians of 6.2 mg L^{-1} in the American whiskeys, 8.6 mg L^{-1} in the Kavalan whiskies and 7.7 mg L^{-1} in the brandies. Gallic acid is released by the hydrolysis of galloyl esters in oak and is the one product of ellagitannin and gallotannin hydrolysis that can be quantified in D_2O; ellagic acid is too insoluble in D_2O and $CDCl_3$ to be observed and was seen only in DMSO-d_6 in version 1.0. The broad aromatic envelope between δ 6.4 and 7.6 that underlies the sharp lines in every oak-aged residue is the observable part of the oligomeric ellagitannin and lignin-derived material. Its integral is not converted to a concentration, because the number of protons per molecule and the molecular masses are unknown, but its proportion of the absolute spectrum is available in the companion workbook as an index of polymeric phenolics.

6.4 Carbohydrates

Glucose, fructose and sucrose were quantified from their resolved anomeric and fructofuranose signals: glucose from its β-H-1 doublet (δ 4.599, J = 7.9 Hz) multiplied by one plus the α/β anomeric ratio measured in the glucose-rich residues of the same solvent system, because the α-H-1 doublet (δ 5.192, J = 3.7 Hz) coincides with the β-arabinopyranose H-1 doublet in every oak-aged residue (Section 9.5; version 2.0 used the sum of the two doublets, and the furanose forms of glucose are negligible); sucrose as the mean of its Glc H-1 doublet (δ 5.393) and Fru H-3 doublet (δ 4.184, J = 8.6 Hz), which agreed within 9 % (median) in the 14 residues containing more than 1 g L^{-1} of sucrose; and fructose from the β-fructofuranose H-3 and H-4 cluster at δ 4.058–4.092 with a factor of 0.48 protons per fructose molecule, taken from the tautomeric composition of fructose in water [61]. The fructose factor could be checked internally. In the twelve cognac residues other than Larsen Aqua Ignis, whose sugar originates in sucrose or invert syrup and therefore contains glucose and fructose in equimolar proportion, the ratio of fructose to glucose found with that factor was 1.02 (range 0.95–1.06), which confirms the factor within about 5 % in this solvent. The fructose window integrates a crowded region, and in residues with less than about 30 mg L^{-1} of fructose the integral is dominated by other signals; fructose is reported only above that matrix-limited level.

The quantities found span four orders of magnitude and fall into recognizable groups (Table 16, Figure 16). In the 103 residues of whiskeys without permitted additions or wine-cask histories (straight and bottled-in-bond American whiskeys, Irish, Scotch, Japanese and most world whiskies, and Kavalan malts from non-wine casks), glucose lay between 25 and 104 mg L^{-1} (10th–90th percentiles), fructose was below its 28.8 mg L^{-1} reporting level in 50 % of residues and reached 106 mg L^{-1} at the 90th percentile, and total sugar was below 0.25 g L^{-1} in 95 % of residues. These are the levels expected of cell-wall sugars released from oak during maturation, whose source, the cellulose and hemicellulose of the charred stave surface, is progressively depleted over years in the barrel [3] and whose concentrations in oak-aged brandy rise with time in wood [18]; they do not indicate an addition. The cognacs contained 4.7–14.0 g L^{-1} of total sugar

expressed as invert sugar, almost entirely as equimolar glucose and fructose with a minor and variable sucrose fraction (largest, 2.7 g L⁻¹, in Larsen Aqua Ignis; Section 10.3), which is the composition expected of a sweetening with sucrose or invert syrup that has inverted in the acidic spirit. Rums were the most diverse class: 15 of the 29 rum residues contained less than 0.6 g L⁻¹ of sugar, while eight contained more than 5 g L⁻¹, up to 25.1 g L⁻¹ in Roatan rum, and the form of the sugar differed between products (sucrose in Roatan, Kiss Black Diamond and Planteray Isle of Fiji; invert sugar in Bacardi Reserva Limitada, English Harbour, Diplomático, Pyrat XO Reserve and a 16-year-old Trinidad rum). Version 2.0 described Pyrat XO Reserve and the Trinidad rum as fructose-rich, with fructose-to-glucose ratios of 2.1 and 5.7. Both residues belong to the group whose anomeric signals were displaced by an unusual CD_3OD content, the fits of version 2.0 had missed most of their β-glucose doublets, and with glucose measured at the observed positions their ratios are 1.1 and 1.2, the composition of invert sugar (Section 9.5).

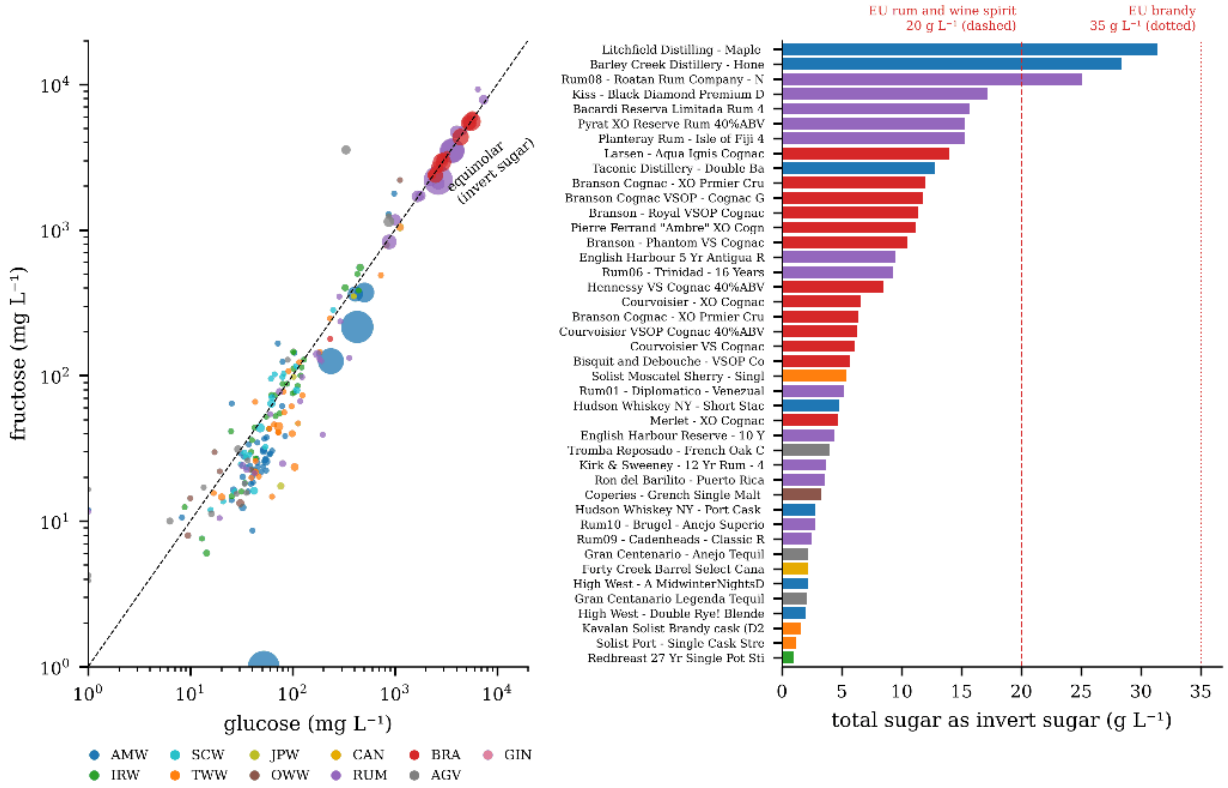

Figure 16. Sugars in the D_2O/CD_3OD residues. Left: fructose against glucose, with symbol area proportional to the square root of the sucrose content; residues in which fructose was not detected are plotted at 1 mg L⁻¹, and the dashed line marks equimolar glucose and fructose. Right: total sugar expressed as invert sugar for every residue above 1 g L⁻¹, with the limits for sweetening of rum and wine spirit (20 g L⁻¹) and of brandy (35 g L⁻¹) in the European Union.

Sucrose behaved as a marker of deliberate addition in the whiskeys as well. Litchfield Distilling Maple Bourbon (29.2 g L⁻¹), Barley Creek Honey Whiskey (26.9 g L⁻¹) and Taconic Double Barrel Maple Bourbon (11.8 g L⁻¹) contained sucrose in amounts that can only come from post-distillation sweetening. The honey whiskey is instructive: honey consists predominantly of fructose and glucose with very little sucrose, yet the residue contained 52.1 mg L⁻¹ of glucose and no quantifiable fructose beside its sucrose, so the sweetness of this product derives from sucrose rather than from honey. Hudson Short Stack, a straight rye finished in casks that had held maple syrup, contained 3.7 g L⁻¹ of sucrose, which is a cask-seasoning carry-over of the same kind as the glycerol and organic acids of a sherry cask and shows how efficiently a seasoned cask transfers its previous contents. Small doublets at the sucrose positions (5–50 mg L⁻¹) in a

few other straight whiskeys are close to the level at which the maltooligosaccharides of oak parenchyma starch, whose α-(1→4) anomeric protons resonate within 0.01 ppm of sucrose Glc H-1, would also appear; they are reported but not interpreted.

6.5 Polyols: quercitol, glycerol and 1,2-propanediol

Quercitol was quantified from its H-6ax ddd (δ 1.797; J = 13.9, 11.3 and 3.2 Hz), whose eight lines span 0.095 ppm and were fitted as a single first-order multiplet. The geminal H-6eq at δ 1.935 and the ring protons near δ 4.0 co-vary with it (Section 6.8), and the H-6ax/H-6eq pair shares one carbon at δC 34–35 in the gHSQC spectrum (Section 8). Quercitol is the most consistent oak marker of the data set (Table 14). The American whiskeys, nearly all matured in new charred oak, had a median of 88.2 mg L^{-1} (interquartile range 69.1–110), the Kavalan malts 71.6 mg L^{-1}, the Irish whiskeys 56.5, the brandies 54.9 and the Scotch malts 37.4 mg L^{-1}, whereas the Canadian whiskies, 16 of the 29 rums and the younger agave spirits contained less than 20 mg L^{-1}; among the agave spirits only the Tromba and Alquimia extra añejos and El Jimador Añejo exceeded 50 mg L^{-1}. Because quercitol is extracted from the wood and has no other plausible source, its concentration is a measure of the contact of the spirit with oak that is independent of the toasting chemistry that controls the aldehydes, and low quercitol together with high sugar and caramel signals characterizes a spirit whose extract is dominated by additions rather than by wood. The diastereomeric composition of the cyclitols has been proposed as a fingerprint of the oak species [19,20]; resolving the minor cyclitols would require a higher field than 300 MHz.

Glycerol was quantified from its H-1a/H-3a double doublet (δ 3.536; J = 11.7 and 6.4 Hz) after subtraction of the downfield ^{13}C satellite of CHD_2OD, which falls within the multiplet at a known 0.535 % of the CHD_2OD main line and amounts to about 7 % of a typical glycerol signal, and was checked against its H-1b/H-3b double doublet (δ 3.626; J = 11.7 and 4.4 Hz). Glycerol is formed in fermentation but is not volatile, so it reaches a spirit only from the cask (wine, sherry or port seasoning) or by addition. In the low-sugar residues the two multiplets agreed (second-to-first ratio 0.88, interquartile range 0.78–0.93, r = 0.98 for the 28 residues above 50 mg L^{-1}), and in the wine-seasoned Kavalan malts glycerol reached 589 mg L^{-1}, with a Kavalan class median of 76.7 mg L^{-1}, the highest of the whisky classes.

In sugar-rich residues the second multiplet proved essential. The H-1a/H-3a multiplet has the same center and the same 11.7 Hz coupling as the β-pyranose H-1b doublet of fructose (lines at δ 3.513 and 3.552), and a single-multiplet fit with its lines broadened to the 3.5 Hz limit reproduced the fructose doublet as an apparent glycerol signal. In Tromba Reposado, a 16-year-old Trinidad rum and High West A Midwinter Night's Dream, all rich in fructose, the first pass returned 1–4 g L^{-1} of glycerol, a quantity that tracked the fructose concentration (19–36 % below it in each case), yet the H-1b/H-3b multiplet was absent or several times weaker than that value required. With the fructose doublet included in the model and agreement of the two multiplets required, glycerol was confirmed in 31 sugar-rich residues (among them the port, Moscatel and Madeira Kavalan malts, Hudson Port Cask Finish and two aged tequilas), was not confirmed in 3 (Tromba Reposado, Pyrat XO Reserve and Bacardi Reserva Limitada, for which only upper limits can be given), and was not quantified in 41, among them all of the cognacs and every product sweetened with sucrose, whose glucose and sucrose H-2 signals and sucrose Fru H-1 singlet overlap both multiplets. Glycerol in those products would require a ^{13}C or HSQC-based measurement, in which its carbons (δC 63.3 and 72.7, Section 8) are resolved from those of the sugars. The episode carries a general lesson: a constrained multiplet fit returns a number whether or not the multiplet is present, and a second, independent resonance of the same molecule is the cheapest protection against a confident but wrong result.

1,2-Propanediol (propylene glycol) was identified by its complete spin system (CH_3 doublet at δ 1.129 with J = 6.3 Hz, CH multiplet at δ 3.86 and diastereotopic CH_2 at δ 3.43 and 3.53) and was quantified from the

methyl doublet in a joint fit with the ethanol triplet, whose upfield line overlaps it in residues that retain much ethanol. It exceeded 100 mg L^{-1} in Roatan rum (1179 mg L^{-1}), TX American Blended Whiskey (819 mg L^{-1}), Kirk & Sweeney 12 yr rum (660 mg L^{-1}), Litchfield and Taconic maple bourbons, Barley Creek Honey Whiskey and Pendleton 1910 (128 mg L^{-1}), and was below 10 mg L^{-1} in 78 % of the residues. At the higher levels the glycol is a flavor carrier; the methyl doublet of 2,3-butanediol, a fermentation product, falls within 0.01 ppm of that of 1,2-propanediol, and the CH and CH_2 signals must be seen before a low-level doublet is assigned to the glycol.

6.6 Organic acids

The methyl doublet of lactic acid (J = 6.9 Hz) moves over 0.08 ppm (δ 1.33–1.42) with the acidity of the residue, and in many whiskeys it is overlapped by the methylene quartet of retained isoamyl alcohol near δ 1.43, whose identity was established by its shift and multiplicity and by statistical correlation with the isoamyl CH multiplet and CH_3 doublets (Section 6.8). The two were therefore fitted jointly. Lactic acid was highest in the agave spirits (class median 41.5 mg L^{-1}; Tromba Extra Añejo 332 mg L^{-1}, El Jimador Añejo 109 mg L^{-1}), in the brandies (52.6 mg L^{-1}) and in the wine-cask Kavalan malts, and low in the American, Irish and Scotch whiskeys (class medians 4–8 mg L^{-1}). Ethyl lactate is a major ester of agave and wine distillates, and its slow hydrolysis in the cask is a plausible source of the lactic acid of aged tequila and brandy, whereas in the Kavalan malts the lactic acid accompanies glycerol and succinic acid as carry-over from wine seasoning. The lactic acid signals of Tromba Extra Añejo were several times broader than those of the other residues although the reference line was sharp, which indicates restricted mobility or exchange of the lactate in that residue (for example association with metal ions or partial oligomerization of the concentrated acid during drying); the area of the broad doublet is reported, but its identity as free lactic acid is less certain than in the other residues.

Succinic acid gives a singlet near δ 2.64 whose position also depends on acidity. It was most abundant in the wine- and sherry-seasoned Kavalan malts (up to 72.6 mg L^{-1}) and in Hudson Port Cask Finish (37.5 mg L^{-1}), which identifies succinic acid, a product of yeast metabolism, as a marker of wine seasoning of the same diagnostic value as glycerol [22]. Acetic acid was found in every residue but at only a few milligrams per liter, against several hundred milligrams per liter in typical aged spirits; it is the most volatile of the acids and is largely removed by freeze-drying, and what remains is the fraction held as acetate salts or in the solid matrix. Its concentration in the residue is not a measure of its concentration in the spirit.

6.7 Retained volatiles

Every D_2O residue retained ethanol and the fusel alcohols. Retained ethanol ranged from 0.067 mg L^{-1} to 5.0 g L^{-1} of original spirit and followed the amount of sugar in the residue closely (Spearman ρ = 0.77 between retained ethanol and total sugar across 169 residues, Figure 17). This is a quantitative confirmation, in a real product matrix, of the mechanism established for freeze-dried carbohydrate solutions half a century ago: volatile molecules are trapped in the amorphous sugar matrix as it dries below a critical water content and are then held far beyond their vapor-pressure expectation [31]. The fusel alcohols behaved differently. Their retention correlated only weakly with sugar (ρ = 0.11) and was highest in the American whiskeys and the brandies (class medians 13.2 and 13.3 mg L^{-1}) and in the Kavalan malts (6.9 mg L^{-1}), against 0.87–1.6 mg L^{-1} in the Irish, Scotch, rum and agave classes, which follows the higher fusel content of spirits distilled to lower strength rather than a matrix effect. The practical consequence is that the ethanol signals of a residue measure its sugar content more than anything else, and that the semi-volatile congeners, whose recovery through freeze-drying depends on both the matrix and the original concentration, cannot be read as concentrations in the spirit.

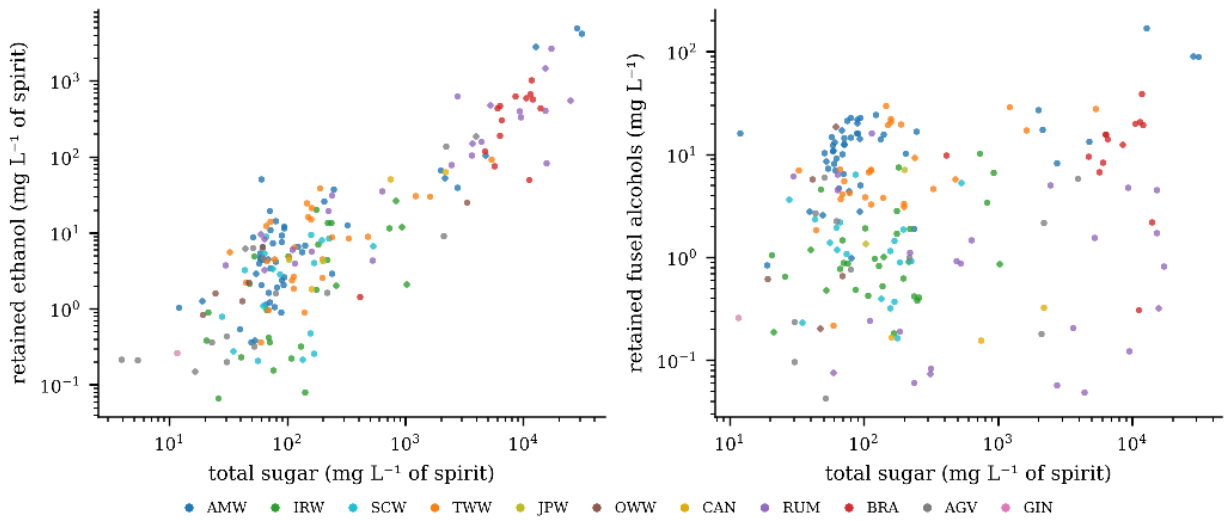

Figure 17. Retention of volatiles in the freeze-dried residues: ethanol (left) and fusel alcohols (right) remaining in the D_2O/CD_3OD residues against their total sugar content, both per liter of spirit.

6.8 Statistical correlation as an assignment aid

Statistical total correlation spectroscopy (STOCSY), the calculation of the correlation between the intensity at a driver position and every other point of the spectrum across a set of samples [62], was used to group signals that belong to the same molecule or to compounds that co-vary for chemical reasons (Figure 18). It was computed on the 136 D_2O/CD_3OD residues containing less than 0.5 g L^{-1} of sugar, so that the large variance of the sweetened residues, and the retention of volatiles that follows it (Section 6.7), could not create correlations of their own. Driven at the quercitol H-6ax line (δ 1.797), the correlation reached 0.98–0.99 at the outer lines of the same ddd (δ 1.751, 1.761, 1.835 and 1.845), 0.97 at the H-6eq multiplet (δ 1.92–1.95) and 0.92 in the ring CH–O region, which supports the H-6ax/H-6eq assignment and the first-order model used to fit them. Driven at the upfield formyl line (δ 9.692), the correlation reached 0.95 at the syringyl singlet (δ 7.29–7.31) but only 0.65 at the vanillin H-5 doublet, as expected for syringaldehyde. Driven at δ 1.444, it reached 0.95 at the methine multiplet of isoamyl alcohol (δ 1.64–1.68) and 0.81 at the methyl doublets (δ 0.86–0.90), whose intensity is shared with isobutyl alcohol; with its shift and multiplicity, this identifies the δ 1.40–1.47 multiplet that overlaps lactic acid as the CH_2 quartet of isoamyl alcohol. Statistical correlation does not distinguish intramolecular partners from different compounds that co-vary because they share a source; the aldehydes, gallic acid and quercitol all correlate with one another because all come from oak, and the correlation pattern was therefore used only together with chemical shift, multiplicity and the two-dimensional data.

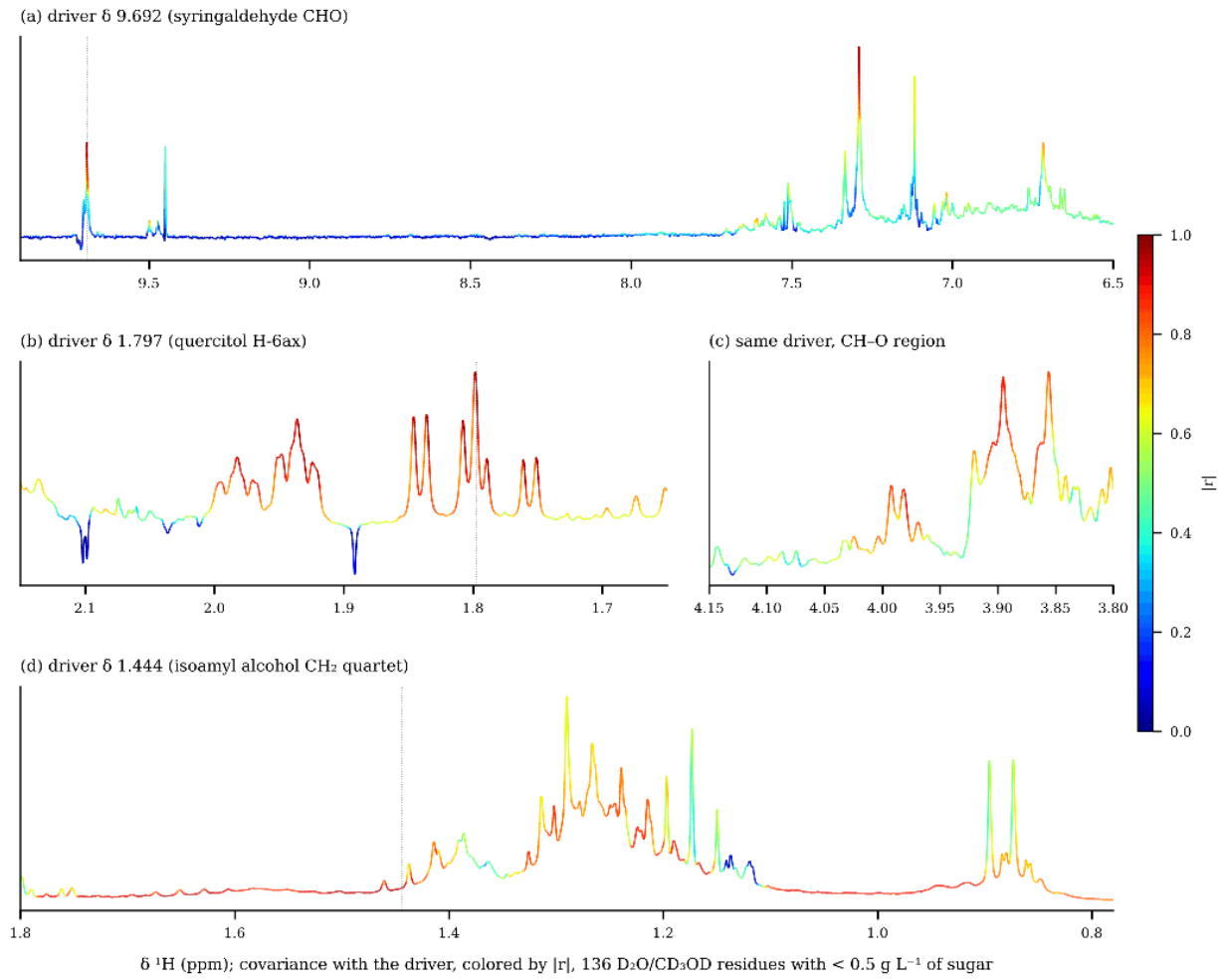

Figure 18. Statistical total correlation spectroscopy of the 136 D_2O/CD_3OD residues containing less than 0.5 g L^{-1} of sugar: covariance of every point with the driver (dotted line), colored by the absolute correlation coefficient. (a) Driver at the syringaldehyde formyl line (δ 9.692); (b, c) driver at the quercitol H-6ax line (δ 1.797); (d) driver at the isoamyl alcohol CH_2 quartet (δ 1.444).

7. The lipophilic fraction: spectra in $CDCl_3$

7.1 General appearance and assignments

The $CDCl_3$ residues contained a median of 33.6 mmol of carbon-bound ^{1}H per liter of spirit (range 3.5–98.6 mmol L^{-1}), less than the D_2O residues because the carbohydrates, polyols and most of the acids do not dissolve in chloroform (Figure 19). What does dissolve is a small and chemically coherent set of compounds: long-chain fatty acids and their ethyl esters, the lignin-derived aldehydes and their methoxyl groups, the cinnamaldehydes, diethyl succinate and, as retained volatiles, the fusel alcohols and 2-phenylethanol. The assignments are listed in Table 7. In $CDCl_3$ the formyl lines are spread more widely than in D_2O, the lipid signals are sharp and the aromatic region is free of the sugar and phenolic envelopes that dominate the polar residue, so this solvent provides the cleaner view of the aldehydes and the only view of the lipids.

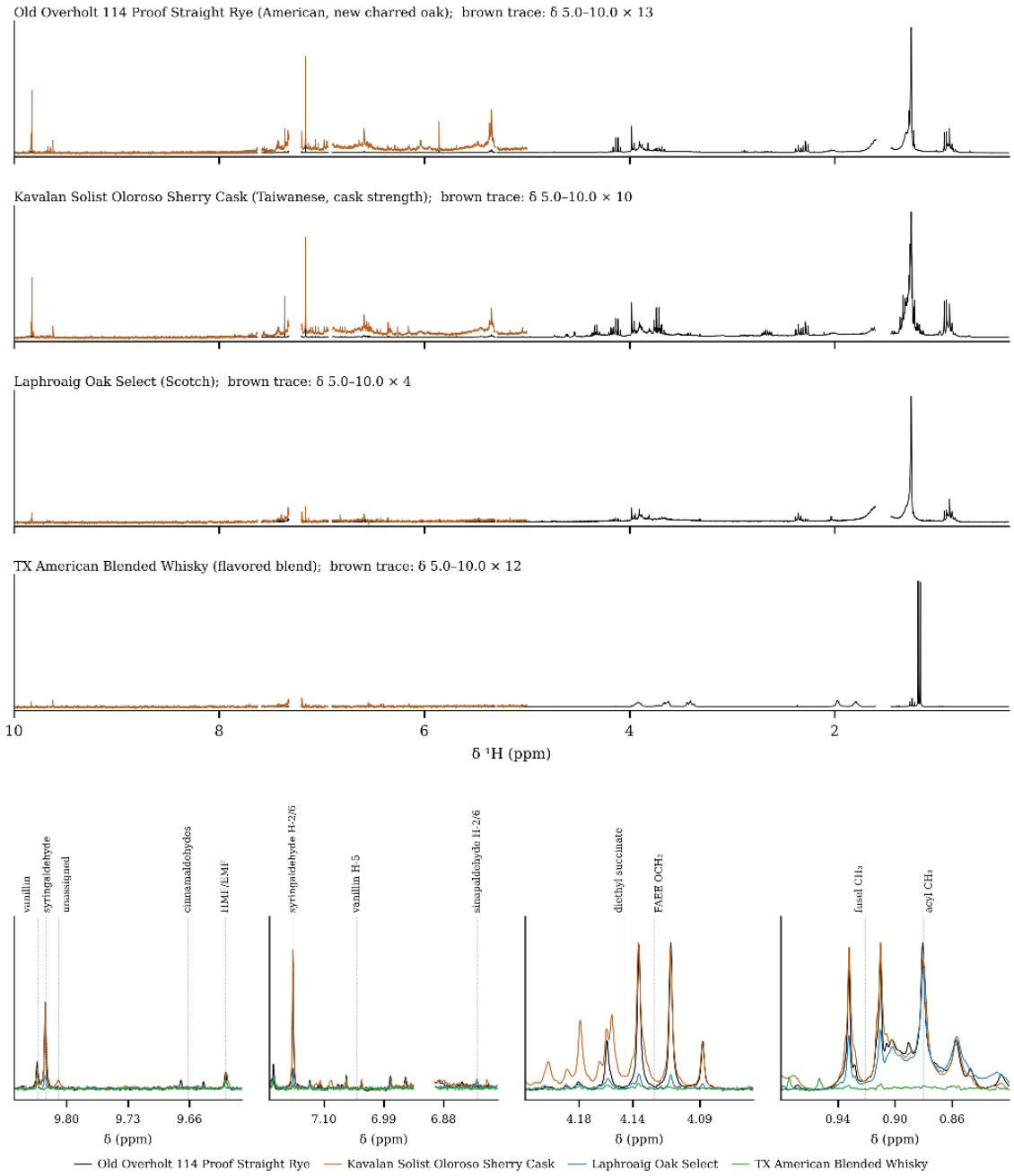

Figure 19. Absolute ¹H spectra of four $CDCl_3$ residues (top; the region δ 5.0–10.0 magnified by the factor given in each title, within the noise ceiling of Section 4.17) and expansions of the formyl, aromatic, ester-methylene and methyl regions overlaid for the same residues (bottom), each scaled to its tallest line within the panel unless that would raise its displayed noise above the common ceiling.

Table 7. ¹H assignments for the freeze-dried residues in $CDCl_3$ (27 °C, TMS = 0). δC from the gHSQC spectrum of the concentrated Kavalan residue in $CDCl_3$.

δH (ppm)	Multiplicity (*J*/Hz)	Assignment	δC	Status	Basis
9.833	s	Vanillin CHO	—	confirmed	with H-2/H-6 7.42–7.44, H-5 7.04 and OCH_3 3.97; regression (Section 7.3) [1]
9.824	s	Syringaldehyde CHO	—	confirmed	regression on H-2/6 (coefficient 1.99 against 2.00; Section 7.3)

δH (ppm)	Multiplicity (*J*/Hz)	Assignment	δC	Status	Basis
9.809	s	Unassigned aromatic aldehyde	—	tentative	strongest in sherry-seasoned malts and the brandy
9.659–9.666	two d, *J* = 7.7	Coniferaldehyde and sinapaldehyde CHO	—	probable	with H-8 6.59 (dd, 15.8, 7.7) [1]
9.619	s	HMF and 5-ethoxymethylfurfural CHO	—	tentative	EMF forms from HMF in ethanol
7.157	s	Syringaldehyde H-2/6	106.7	confirmed	HSQC
6.818	s	Sinapaldehyde H-2/6	—	probable	syringyl cinnamaldehyde ring protons
6.59	dd, *J* ≈ 15.8, 7.7	Cinnamaldehyde H-8	—	probable	*trans* olefinic coupling
6.34–6.69	s	Syringyl CH of lignans and phenols	103–106	tentative	HSQC
5.31–5.42	m	Olefinic CH=CH of unsaturated fatty acids and FAEE	129.6	confirmed	HSQC [21]
4.146; 2.62	q, *J* = 7.1; s	Diethyl succinate OCH_2; CH_2	— ; 28.5	probable	HSQC (CH_2)
4.124	q, *J* = 7.1	Fatty acid ethyl ester (FAEE) OCH_2	60.2	confirmed	HSQC [21]
3.89–3.98	s	Aromatic OCH_3 (vanillin, syringaldehyde, cinnamaldehydes, lignans)	56.3	confirmed	HSQC
3.87; 2.87	t, *J* ≈ 6.6	2-Phenylethanol CH_2O; $ArCH_2$ (retained)	—	probable	
3.72; 1.25	q; t	Ethanol (retained)	—	confirmed	[33]
3.68	t	Isoamyl alcohol CH_2O (retained)	—	probable	with 0.92 (d)
2.355	t, *J* = 7.5	Free fatty acid α-CH_2	34.0	probable	HSQC
2.285	t, *J* = 7.5	FAEE α-CH_2	34.0	probable	HSQC
2.102	s	Acetic acid CH_3	—	probable	[33]
2.00–2.05	m	Allylic CH_2 of unsaturated chains	27.1	confirmed	HSQC [21]
1.61	m	Acyl β-CH_2	24.7	confirmed	HSQC
1.25–1.34	br	$(CH_2)_n$ envelope of the acyl chains	29.2	confirmed	HSQC
0.921	d, *J* = 6.7	Isoamyl and isobutyl alcohol CH_3 (retained)	—	probable	
0.88	t, *J* = 6.9	Acyl chain CH_3	14.0	confirmed	HSQC
7.26; 7.61 / 6.91	s; ^{13}C satellites	$CHCl_3$ (^1J(C,H) = 209 Hz)	77.4	artifact	internal check (Section 5.6) [33]
5.30	s	CH_2Cl_2 (laboratory contaminant)	—	artifact	[33]
1.48–1.60	br	H_2O in $CDCl_3$	—	artifact	[33]
0.000	s	TMS (reference)	—	reference	

7.2 Lipids

The fatty-acid and ethyl-ester signals were quantified with the proton bookkeeping used for edible oils [21]. The two α-methylene triplets, at δ 2.355 for free fatty acids and δ 2.285 for the ethyl esters, together count the acyl chains (two protons per chain); the ester OCH_2 quartet at δ 4.122 counts the ethyl esters (two protons per ester); and the olefinic multiplet at δ 5.31–5.42 counts the double bonds (two protons per CH=CH). The acyl methyl triplet at δ 0.88 sits on the shoulder of the much larger fusel-alcohol doublet and on a broad lipid envelope, and its window integral underestimated the chain count by a large factor; the α-methylene count was used in its place. On that basis the median residue contained 170 µmol of acyl chains per liter of spirit, of which about 33 % were ethyl esters and the remainder free acids, with 0.33 double bonds per chain, which is consistent with a lipid fraction dominated by the saturated C_{12}–C_{16} ethyl esters and acids with a minor contribution from oleate and linoleate. No glyceride signals (glycerol-backbone CH_2 at δ 4.1–4.3 and CH at δ 5.26) were observed. The lipid content differed by more than an order of magnitude between products. It was highest in the cask-strength Kavalan malts (median 526 µmol L^{-1} of acyl chains, up to 839 µmol L^{-1} in a Kavalan Solist Oloroso) and in barrel-proof American whiskeys such as Old Overholt 114 Proof, and lowest in the rums, the gin and the chill-filtered blends. Lipids are carried from the wash into the low wines and survive in spirits that are bottled at high strength without chill filtration; their quantity in the residue is therefore an index of the bottling practice as much as of the distillation.

The composition of the lipid fraction also differed between categories, but in a way that cannot be read from the category medians alone. The number of double bonds per acyl chain had a median of 0.36 in the American whiskeys (31 residues), 0.27 in the Irish whiskeys (5), 0.15 in the Scotch malts (6) and 0.16 in the Kavalan malts (12), and the difference between categories was significant (Kruskal–Wallis p = 2e-07). Across the whole set, however, the ratio fell as the lipid content rose (Spearman ρ = −0.79), and the regression of the olefinic protons on the acyl chains had a positive intercept of 32 µmol L^{-1} of double bonds, so the ratio mixes the composition of the lipid with its amount. The slope of that regression, the unsaturation of the lipid that each additional chain brings, is the better comparison: it was 0.26 double bonds per chain within the American whiskeys, against 0.12 within the Kavalan malts and 0.15 within the Scotch malts. The malt whiskies, which are the richest in lipid, therefore owe most of their additional lipid to saturated chains, as expected for the medium- and long-chain ethyl esters and acids that yeast produces, whereas the higher incremental unsaturation of the American whiskeys is consistent with a larger contribution of the oleate and linoleate of cereal lipids. The ethyl-ester share of the acyl chains was likewise highest in the Kavalan malts (44 %) and the American whiskeys (34 %) and lowest in the Scotch (9 %) and Irish (14 %) whiskies. This interpretation is plausible but not established. The chain-length distribution that would confirm it is not visible in a ^{1}H spectrum, the more volatile C$_{8}$–C$_{12}$ esters and acids are partly lost during freeze-drying so that the residue over-represents the longer chains, and the intercept shows an olefinic contribution that is not tied to acyl chains, which in the gin residue is presumably due to terpenes of the botanicals. Gas chromatography of the fatty acid ethyl esters of the same residues would settle the question.

7.3 Aldehydes and methoxyls

Three formyl singlets were resolved at δ 9.834, 9.825 and 9.810. The assignment of the central and largest line to syringaldehyde was tested directly by regression of the syringyl ring singlet (δ 7.155, two protons) on the three formyl areas across all 61 residues: the coefficient for the δ 9.825 line was 1.99, against the 2.00 required by the structure, while the coefficients for the other two lines were 0.05 and −0.06. The δ 9.834 line is vanillin, whose ring protons appear at δ 7.42–7.44 and 7.04, and the small δ 9.810 line, strongest in the sherry-seasoned malts and in the single brandy, remains unassigned. Syringaldehyde was found at a median of 10.3 mg L^{-1} in the American whiskeys, 7.7 mg L^{-1} in the Kavalan malts and 4.1 mg L^{-1} in the Scotch malts, and vanillin at 4.0, 3.1 and 1.8 mg L^{-1} respectively. The molar ratio of syringaldehyde to vanillin had a median of 2.28 and lay between 0.85 and 3.25 for 90 % of the residues, which is the range expected for aldehydes released together from oak lignin; values far below this range identify vanillin of non-oak origin (Section 6.3).

The two cinnamaldehyde formyl doublets (J = 7.7 Hz) are offset by only 0.007 ppm, and their separate areas could not be assigned with confidence: a regression of the sinapaldehyde ring singlet (δ 6.822, two protons, confirmed as the H-2/6 resonance of a syringyl cinnamaldehyde) on the two doublets gave coefficients that did not correspond to either assignment. The cinnamaldehydes are therefore reported as their sum, sinapaldehyde individually from its ring singlet, and coniferaldehyde by difference. The sum was highest in the American whiskeys (median 5.7 mg L^{-1}), as expected for spirits matured in new charred oak, whose char layer is the richest source of these compounds. The furanic formyl singlet at δ 9.616 (HMF and the 5-ethoxymethylfurfural that forms from it in ethanol) and the diethyl succinate quartet at δ 4.153 complete the list of quantified CDCl$_3$ congeners; diethyl succinate was most abundant in the Kavalan malts (4.5 mg L^{-1}) and the Scotch malts (3.6 mg L^{-1}), in line with the wine-cask and refill-cask histories of those products.

7.4 Recovery of the same compounds in the two solvents

Because the D_2O and $CDCl_3$ residues were prepared from separate aliquots of the same bottle, a compound that dissolves completely in both solvents should be found at the same concentration per liter of spirit in both spectra. The test is shown in Figure 20 for the 51 products with a D_2O/CD_3OD counterpart. The fusel alcohols, which are freely soluble in both solvents, were recovered at a median $CDCl_3/D_2O$ ratio of 1.06 (Spearman $\rho = 0.67$), which shows that the two preparations retained comparable amounts of these volatiles and that the absolute scales of the two solvents agree within the ±10 % assumed in Section 5.6. The benzaldehydes were recovered at 0.77 of their D_2O value ($\rho = 0.73$), HMF at 0.84 ($\rho = 0.71$) and the cinnamaldehydes at 0.81 ($\rho = 0.78$), and acetic acid, which is extracted from a dry residue only partly by chloroform, at 0.20. The aldehydes are therefore under-recovered in $CDCl_3$ by a consistent 16–23 %, presumably because part of each aldehyde is held in the polar matrix of the residue and does not pass into chloroform. The agreement of the three aldehyde groups with one another is itself informative: before the broad furanic envelope was modeled in the D_2O fits, the cinnamaldehydes appeared to be recovered in $CDCl_3$ at only about half of their D_2O value, and the discrepancy disappeared once the envelope was no longer attributed to them. The rank agreement shows that either solvent orders the products correctly, but absolute concentrations of the aldehydes are better taken from the D_2O/CD_3OD residues, with the $CDCl_3$ spectra used for the individual compounds and their ratios.

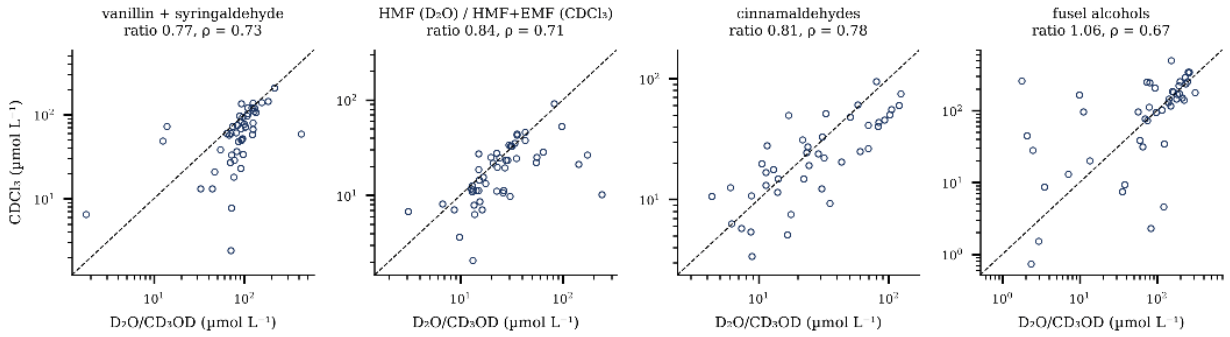

Figure 20. Recovery of the same compounds in the two solvents: concentrations measured in the $CDCl_3$ residue against those measured in the D_2O/CD_3OD residue prepared from a separate aliquot of the same bottle, with the median $CDCl_3/D_2O$ ratio and the Spearman correlation; the dashed line is the line of equality.

8. Two-dimensional confirmation

The Kavalan documents contained a gHSQC spectrum (multiplicity-edited, 2048 × 1024 points) and a homonuclear *J*-resolved spectrum (2048 × 2048) of a concentrated residue of Kavalan Solist Vinho Barrique in D_2O/CD_3OD (500:150 µL; about 30 mg of solids prepared from 18 mL of spirit), together with the corresponding ^{1}H and DEPT-45 spectra, and an equivalent set in $CDCl_3$. The processed matrices were recovered from the documents in the same way as the one-dimensional spectra. The indirect (^{13}C) axis of the D_2O gHSQC, which has no internal carbon reference, placed the CHD_2OD carbon about 0.55 ppm above its expected position, and all D_2O carbon shifts quoted here and plotted in Figure 21 were corrected by −0.55 ppm; the remaining differences of up to about 0.8 ppm between the observed carbon shifts and tabulated values reflect the several referencing conventions in use for aqueous ^{13}C data, and no assignment below depends on a difference of that size. The principal cross peaks are listed in Table 8 and shown in Figure 21.

The heteronuclear data confirm the assignments that carry the quantitative weight of this paper. The quercitol H-6 protons (δH 1.82 and 1.99 in D_2O alone) share a single methylene carbon at δC 34.5, of opposite phase to the methine carbons in the edited spectrum, which establishes them as a geminal pair; the ring CH–O carbons of quercitol appear at δC 69.3–69.5 with protons near δH 3.93–3.99. Glycerol is confirmed by a methylene carbon at δC 63.3 (δH 3.62) and a methine carbon at δC 72.7 (δH 3.76), succinic acid by a methylene at δC 29.8 (δH 2.65), lactic acid by a methyl at δC 20.3 (δH 1.43), gallic acid by an aromatic methine at δC 110.9 (δH 7.10), the aromatic methoxyl groups by δC 57.0 (δH 3.87), and the α- and β-anomers of glucose by δC 93.3 (δH 5.21) and 96.9 (δH 4.62). Two further anomeric correlations are informative: δC 102.9 at δH 4.46 is a β-glycosidic carbon (an internal or non-reducing β-xylosyl or β-glucosyl unit of a hemicellulose fragment or glycoside), and δC 108.8 at δH 5.01 is characteristic of an α-arabinofuranosyl C-1, which places arabinoxylan-derived oligosaccharides from the oak cell wall in the residue. The syringyl aromatic carbons appear at δC 104.9–107.6 (δH 6.43–6.68 and 7.23) and a guaiacyl C-2 at δC 112.0 (δH 7.43), and the ethyl ester carbons at δC 63.2 (OCH_2, δH 4.24) and 14.1–14.7 (CH_3, δH 1.23–1.29).

In $CDCl_3$ the edited spectrum confirms the fatty-acyl assignments (CH_2 envelope δC 29.2, α-CH_2 34.0, β-CH_2 24.7, allylic CH_2 27.1, olefinic CH 129.6 and terminal CH_3 14.0), the FAEE ester methylene (δC 60.2, δH 4.12), the diethyl succinate methylene (δC 28.5, δH 2.62), the methoxyl carbons (δC 56.3), the syringaldehyde ring methine (δC 106.7, δH 7.14) and syringyl-type methines of lignans or phenols at δC 103–106 (δH 6.34–6.69). The *J*-resolved spectrum confirmed the multiplicities used in the fitting models: the doublets of α- and β-glucose and of lactic acid, the singlets of gallic and succinic acid, and the multiplet structure of the quercitol methylene protons.

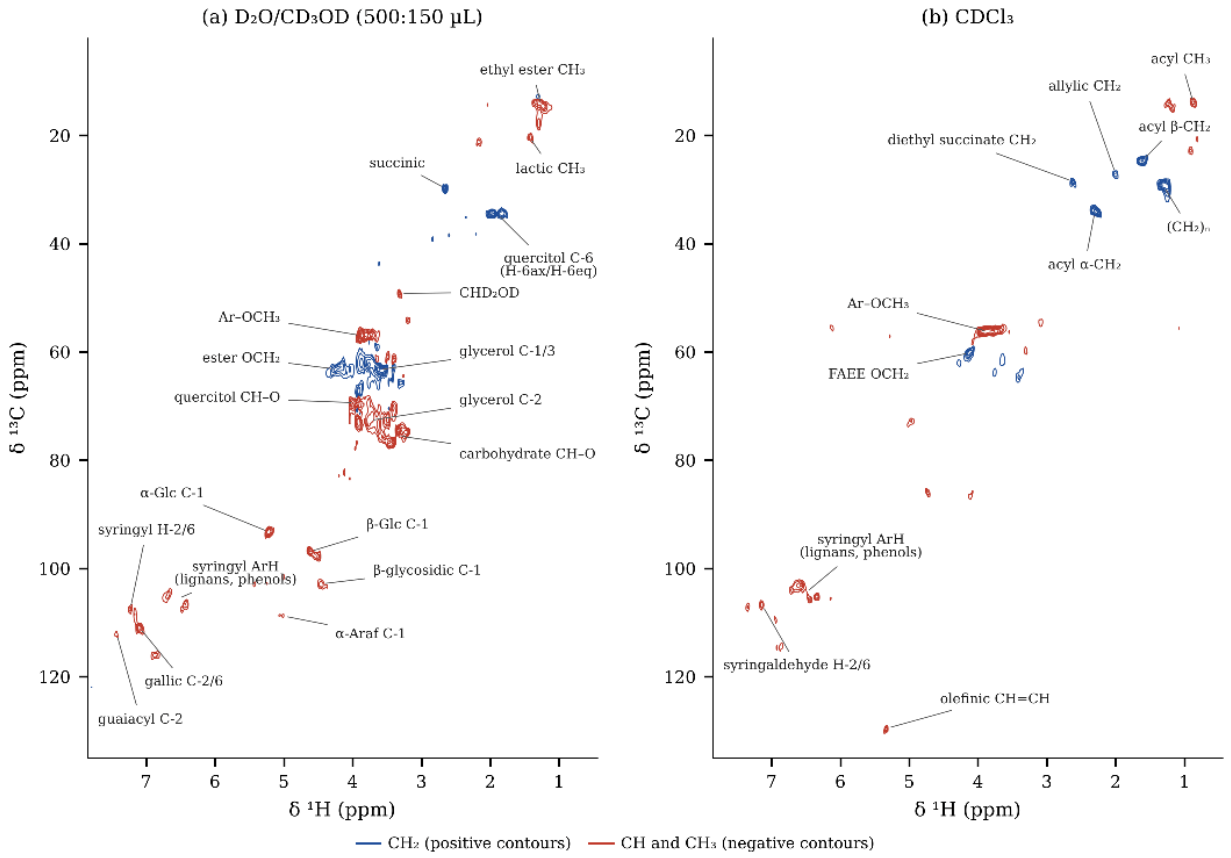

Figure 21. Multiplicity-edited gHSQC spectra of a concentrated Kavalan Solist Vinho Barrique residue (about 30 mg of solids from 18 mL of spirit) in D_2O/CD_3OD (a) and $CDCl_3$ (b). Methylene correlations are plotted in blue and methine and methyl correlations in red; the ^{13}C axis of the D_2O spectrum was corrected by −0.55 ppm (Section 8).

Table 8. Principal cross peaks in the multiplicity-edited gHSQC spectra of the concentrated Kavalan Solist Vinho Barrique residue (δC of the D_2O spectrum corrected by −0.55 ppm; Section 8).

Solvent	δH (ppm)	δC (ppm)	Edited phase	Assignment
D_2O	1.82; 1.99	34.5	CH_2	Quercitol H-6ax/H-6eq
D_2O	3.93–3.99	69.3–69.5	CH	Quercitol ring CH–O
D_2O	3.62; 3.76	63.3; 72.7	CH_2; CH	Glycerol C-1/3; C-2
D_2O	2.65	29.8	CH_2	Succinic acid
D_2O	1.43	20.3	CH_3	Lactic acid CH_3
D_2O	7.10	110.9	CH	Gallic acid C-2/6
D_2O	3.87	57.0	CH_3	Aromatic OCH_3
D_2O	5.21; 4.62	93.3; 96.9	CH	α- and β-Glucose C-1
D_2O	4.46	102.9	CH	β-Glycosidic C-1 (xylan or glucan fragment, glycoside)
D_2O	5.01	108.8	CH	α-Arabinofuranosyl C-1
D_2O	6.43–6.68; 7.23	104.9–107.6	CH	Syringyl aromatic CH (lignans, phenols; syringyl H-2/6)
D_2O	7.43	112.0	CH	Guaiacyl C-2
D_2O	4.24; 1.23–1.29	63.2; 14.1–14.7	CH_2; CH_3	Ethyl esters OCH_2; CH_3
$CDCl_3$	1.31	29.2	CH_2	Acyl $(CH_2)_n$
$CDCl_3$	2.30	34.0	CH_2	Acyl α-CH_2
$CDCl_3$	1.61	24.7	CH_2	Acyl β-CH_2
$CDCl_3$	2.00	27.1	CH_2	Allylic CH_2

Solvent	δH (ppm)	δC (ppm)	Edited phase	Assignment
$CDCl_3$	5.33	129.6	CH	Olefinic CH=CH
$CDCl_3$	0.87	14.0	CH_3	Acyl CH_3
$CDCl_3$	4.12	60.2	CH_2	FAEE OCH_2
$CDCl_3$	2.62	28.5	CH_2	Diethyl succinate CH_2
$CDCl_3$	3.78–3.89	56.3	CH_3	Aromatic OCH_3
$CDCl_3$	7.14	106.7	CH	Syringaldehyde C-2/6
$CDCl_3$	6.34–6.69	103–106	CH	Syringyl CH of lignans and phenols

9. The whole residue in one solvent: DMSO-d₆

The two-solvent scheme of the preceding sections divides the residue along a line drawn by the solvents rather than by the chemistry. Sugars, glycerol and the organic acids go to D_2O, the lipids and esters to $CDCl_3$, and the phenolic aldehydes, which dissolve in both, are split between the two aliquots in proportions that depend on the matrix (Section 7.4). Dimethyl sulfoxide dissolves the whole residue, and the titles of the DMSO-d₆ spectra record that the residues were "completely soluble". Version 1.0 of this paper used these spectra, on a relative scale, chiefly as the route to ellagic acid. In this version they were recovered from their MestReNova document, reprocessed from the stored decays, placed on the same absolute scale as the D_2O and $CDCl_3$ spectra and quantified, so that the question that runs through the rest of the paper, whether one solvent can replace two, can be answered with numbers. The validation of the absolute scale in this solvent is described in Section 5.8 and the methods in Section 4.15; this section describes what the spectra contain. Representative spectra are shown in Figure 22, every spectrum is overlaid in sixteen diagnostic windows in Figure 24, and the assignments are collected in Table 9.

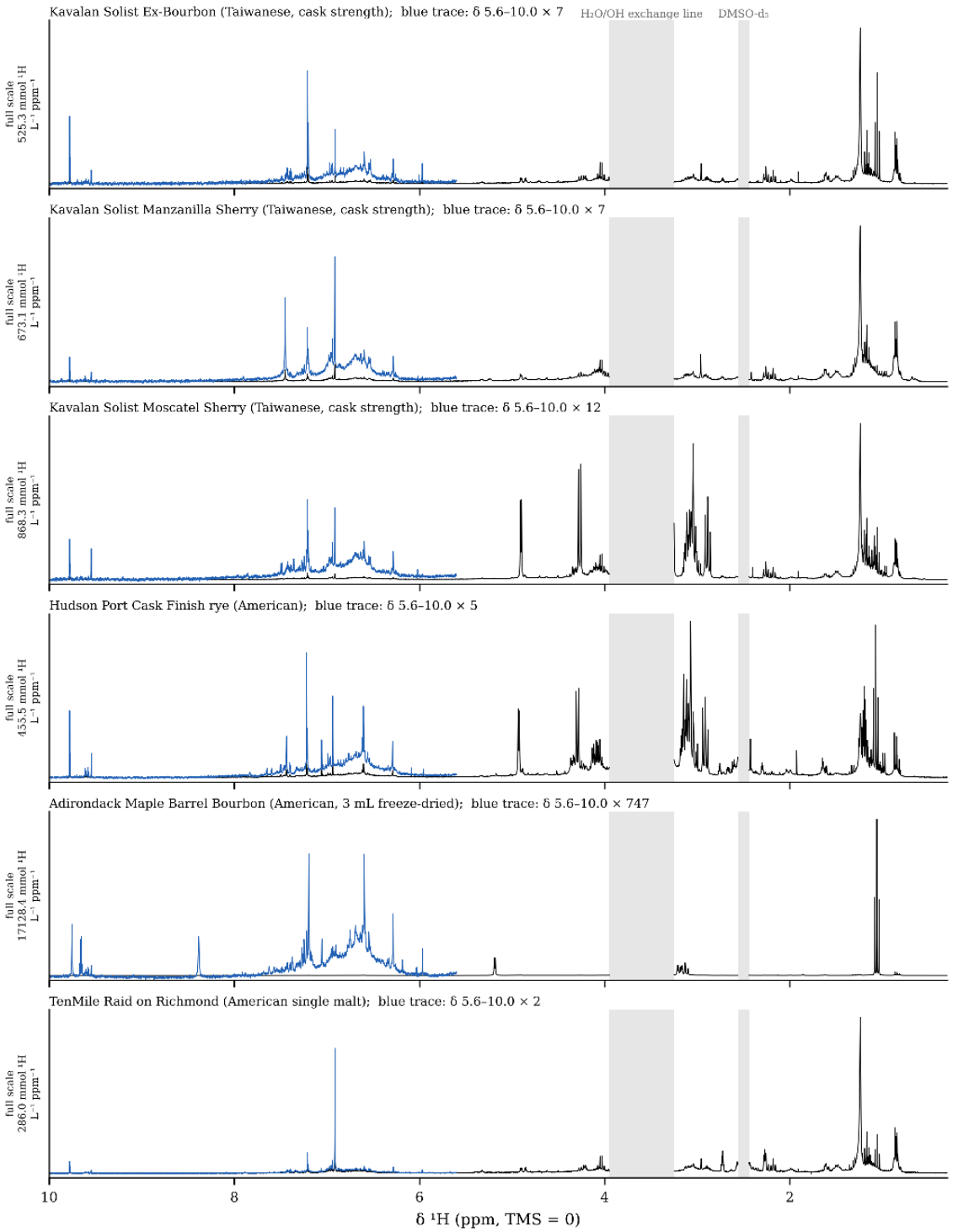

Figure 22. Absolute ¹H spectra (299.99 MHz, DMSO-d₆, TMS = 0) of six freeze-dried residues: Kavalan Solist Ex-Bourbon, Solist Manzanilla and Solist Moscatel (Taiwanese single malts at cask strength), Hudson Port Cask Finish rye, Adirondack Maple Barrel Bourbon (3 mL of spirit freeze-dried) and TenMile Raid on Richmond. Each spectrum is scaled to its own maximum, with the full scale given at left in mmol of ¹H per liter of spirit per ppm, and the region δ 5.6–10.0 is also shown magnified (blue) by the factor given in each title, limited by the noise ceiling of Section 4.17. The DMSO-d₅ region (δ 2.44–2.56) and the exchange-averaged H₂O/OH line (δ 3.25–3.95) are masked.

9.1 The data set

The DMSO-d_6 document contained thirty spectrum titles but only twenty-nine distinct acquisitions, because one decay was filed under two titles. The residues comprise twenty-one Kavalan single malts, six American whiskeys (four Hudson whiskeys from Tuthilltown, the Adirondack Maple Barrel Bourbon and the TenMile Raid on Richmond single malt), one Scotch single malt and one rum, and they were acquired between 6 and 24 April 2026 and in September and October 2026 (Table 30). Each residue was prepared from 1.000 mL of spirit (3.000 mL for the two Adirondack products) and dissolved in 600 µL of DMSO-d_6 (610 µL for TenMile); for seven early preparations the volume was not recorded and 600 µL was assumed. Twenty-four of the products also have a D_2O or D_2O/CD_3OD residue in the data base and ten have residues in all three solvents, and these matched products carry the cross-solvent comparisons of this section and of Sections 12.9 and 12.10. Most titles record "HOD suppression", but the stored parameter sets record the one-pulse sequence with presaturation disabled, and the water line of every spectrum is present at full intensity; the spectra were therefore treated as unsuppressed, which is the more favorable case for quantitation. Two titles conflict with the sample identifiers or with the parameter sets, and both conflicts were resolved by the chemistry of the residues (Section 9.9).

9.2 What the spectra show

On the absolute scale the DMSO-d_6 spectra combine the two halves of the two-solvent picture (Figure 22). The formyl region carries the syringaldehyde and vanillin singlets near δ 9.77, the HMF singlet at δ 9.545 and the cinnamaldehyde doublets near δ 9.60; the aromatic region carries the syringyl singlets at δ 7.21 and 7.20, gallic acid at δ 6.91, ellagic acid near δ 7.46 in several wine-seasoned malts and a broad envelope of oligomeric phenolics between δ 6.4 and 7.2; the anomeric region carries resolved glucose and pentose doublets between δ 4.2 and 4.9 and, in the maple bourbon, the sucrose Glc H-1 doublet at δ 5.17; and the aliphatic region carries the ethyl ester quartet at δ 4.04 with its triplet at δ 1.17, the α-methylene triplets of ethyl esters and free fatty acids at δ 2.26 and 2.18, the lipid methylene envelope at δ 1.24, acetic acid at δ 1.91, retained ethanol at δ 1.06 and the methyl signals of the fusel alcohols and acyl chains near δ 0.85. The lipid signals that required a separate $CDCl_3$ residue in the two-solvent scheme and the carbohydrate and phenolic signals of the D_2O residue therefore appear together, in one tube and on one scale.

Two regions are lost. The residual DMSO-d_5 quintet at δ 2.50, about 60.1 mmol L^{-1} of ^{1}H in the tube, blinds the spectrum from about δ 2.44 to 2.56, and its ^{13}C satellites at δ 2.27 and 2.73 fall on the ester α-methylene triplets and on the citric and malic acid region (Figure 30a). More seriously, the water carried by the residue and by the hygroscopic solvent, together with every hydroxyl and carboxyl proton of the solutes with which it exchanges, collapses into a single exchange-averaged line whose position varied between δ 3.31 and 3.84 and whose width varied from a few hertz to more than 50 Hz. That line lies on the CH–O protons of the sugars, glycerol and the cyclitols between δ 3.0 and 4.0, so the region that carries most of the carbohydrate information in D_2O cannot be used in DMSO-d_6; glycerol in particular, which resonates at δ 3.3–3.4 in this solvent, could not be measured in any residue.

The expansions of Figure 23 show the diagnostic signals of five residues, and the overlay of Figure 24 shows how much they vary across the data set. Two features of the overlay deserve comment. In most windows the lines of different residues coincide to within 0.002 ppm after referencing to TMS (Section 5.8), but in the aromatic, anomeric and acid regions a few residues show compound-specific displacements of up to 0.02 ppm that a single reference cannot remove: the syringyl singlet of the Kavalan oloroso cask (005) lies 0.009 ppm downfield of its usual position while its lipid methylene envelope does not move, and the succinic acid singlet wanders between δ 2.363 and 2.433. These displacements follow the composition of the residue, mainly its water, acid and sugar load, and they were handled by region-wise alignment in

the binned data and by position windows wide enough to follow them in the fits (Section 4.15). Second, the anomeric region changes shape from residue to residue in a way that the D_2O spectra do not, for a reason discussed in Section 9.4.

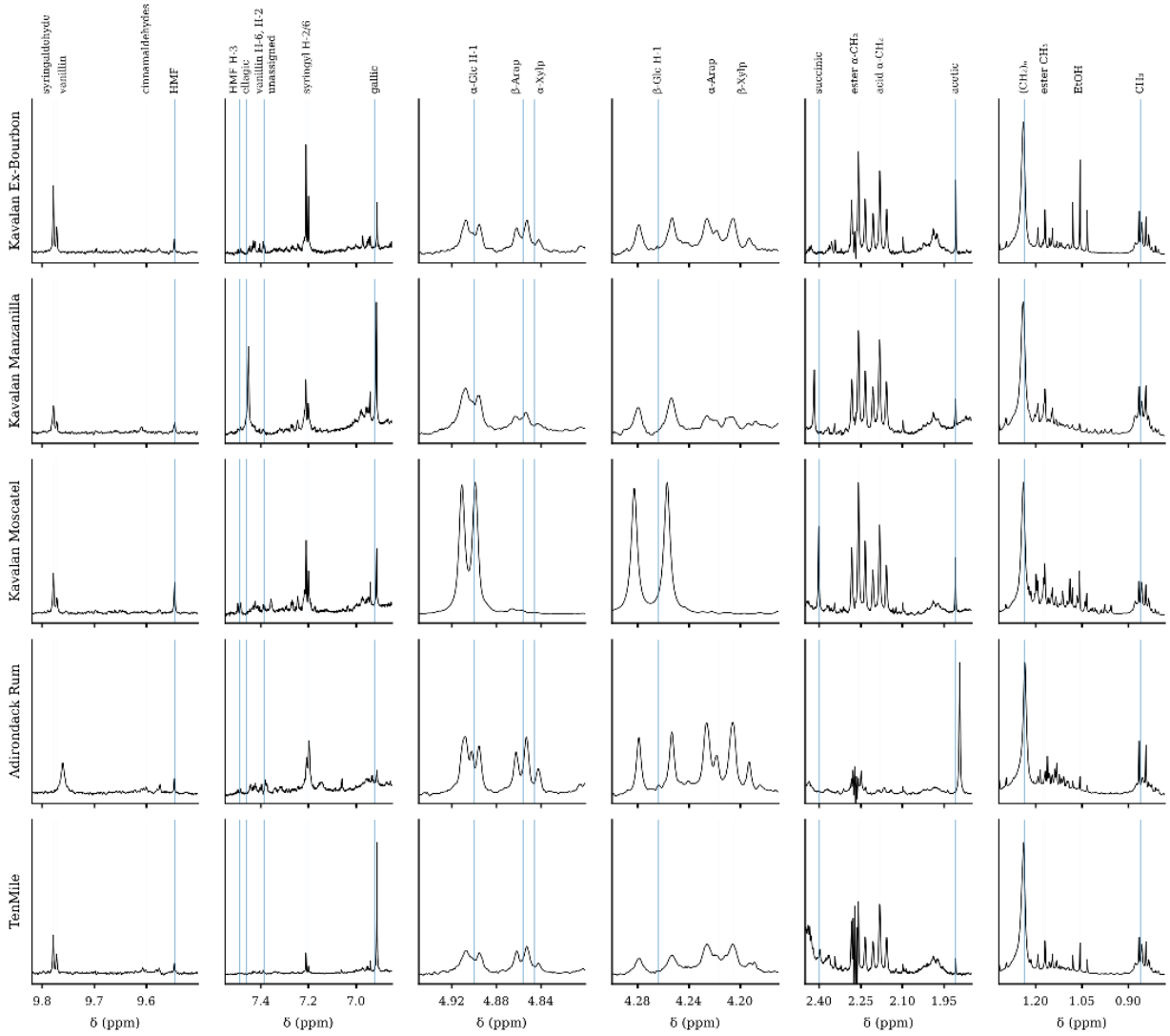

Figure 23. Expansions of the DMSO-d_6 spectra of five residues after referencing and region-wise alignment (Section 4.15): the formyl region, the aromatic region, the α- and β-anomeric regions of glucose and the pentoses, the acid and acyl α-methylene region and the aliphatic region. Each trace is scaled, window by window, to its tallest line within the common noise ceiling of Section 4.17, and the vertical lines mark the reference positions used in the fits (doublet centers for the anomeric signals). The sharp multiplet near δ 2.27 in the spectra of Adirondack Rum and TenMile is the upfield ^{13}C satellite of DMSO-d_5 from the second solvent lot.

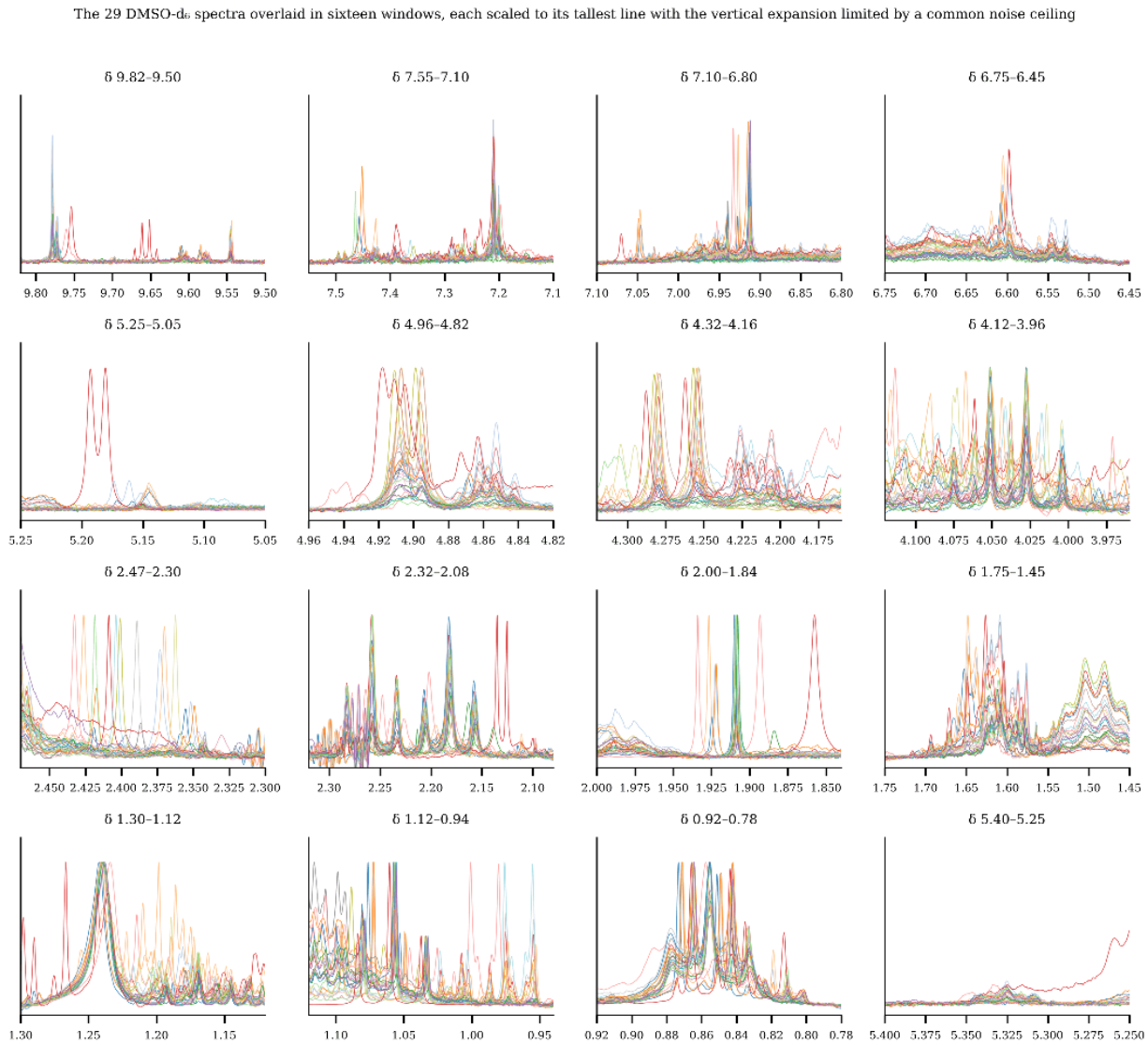

Figure 24. All 29 DMSO-d₆ spectra overlaid in sixteen diagnostic windows after referencing and region-wise alignment, each trace scaled to its tallest line within the window unless that would raise its displayed noise above the common ceiling of Section 4.17 (colors serve only to separate the traces). The windows cover the formyl, aromatic and phenolic regions, the anomeric regions, the ethyl ester OCH_2 quartet, the succinic acid and acyl α-methylene regions, acetic acid, the quercitol region, the lipid and ester methyl region, retained ethanol, the terminal methyl region and the olefinic region.

9.3 Assignments

The assignments of Table 9 rest on four kinds of evidence. For the lignin-derived aldehydes and acids, the BMRB holds experimental spectra of authentic standards in DMSO-d₆ at 298 K referenced to TMS [59], and the residue lines agree with them within 0.01 ppm: syringaldehyde CHO and H-2/6 at δ 9.778 and 7.210 against 9.783 and 7.215, vanillin CHO, H-6 and H-2 at δ 9.772, 7.413 and 7.387 against 9.771, 7.419 and 7.389, coniferaldehyde and sinapaldehyde CHO at δ 9.60 against 9.599, and ellagic acid at δ 7.45–7.46 against 7.462. The solvent residues (water, ethanol, acetic acid, ethyl ester groups and TMS) were assigned from the tabulated shifts of trace impurities in DMSO-d₆ [33], and HMF from its formyl singlet and its H-3 and H-4 doublets (J = 3.6 Hz) in this solvent [63]. Second, several molecules provide two independent resonances whose areas must agree. The syringaldehyde formyl singlet and its H-2/6 singlet gave a median

ratio (per proton) of 0.91 over 26 residues; the vanillin formyl singlet and its H-2 doublet gave 0.95; the HMF formyl singlet and its H-3 doublet gave 0.85; and the ethyl ester quartet and the ester α-methylene triplet gave 1.10, which shows that most of the ester groups in these residues belong to long-chain fatty-acid ethyl esters. The singlet at δ 7.200, which accompanies the syringaldehyde H-2/6 singlet at about 0.47 of its intensity in every residue, coincides with neither aldehyde nor acetosyringone (BMRB δ 7.236) and is close to syringic acid (BMRB δ 7.210); it is assigned to syringic acid tentatively.

Third, the products that have both a D_2O/CD_3OD and a DMSO-d_6 residue allow statistical heterospectroscopy across solvents [62,64]: each quantity measured in D_2O was correlated with every point of the aligned DMSO-d_6 spectra of the same products (Figure 25). Glucose measured in D_2O correlated most strongly with the DMSO-d_6 doublets at δ 4.90 and 4.27 and with the H-2 and H-4 signals near δ 2.88 and 3.04 ($\rho \geq 0.85$); HMF with the δ 9.545 singlet ($\rho = 0.95$) and its H-3 doublet; the benzaldehydes with the δ 9.777 and 7.209 singlets; and quercitol with the δ 1.60 multiplet ($\rho = 0.80$), which is therefore assigned to its C-6 methylene. The heterospectroscopy did not identify the DMSO-d_6 resonances of glycerol, succinic acid or lactic acid, because across the matched products these wine-cask congeners vary together and with the ethyl esters, so that each correlates with the whole wine-cask pattern rather than with its own lines; the same confounding limited the STOCSY of Section 6.8. Fourth, the recovery of each assigned compound in DMSO-d_6 relative to its recovery in D_2O or $CDCl_3$ (Section 9.8) tests the assignment and the absolute scale together: a quercitol window that recovered 97 % of the D_2O quercitol cannot be carrying much else.

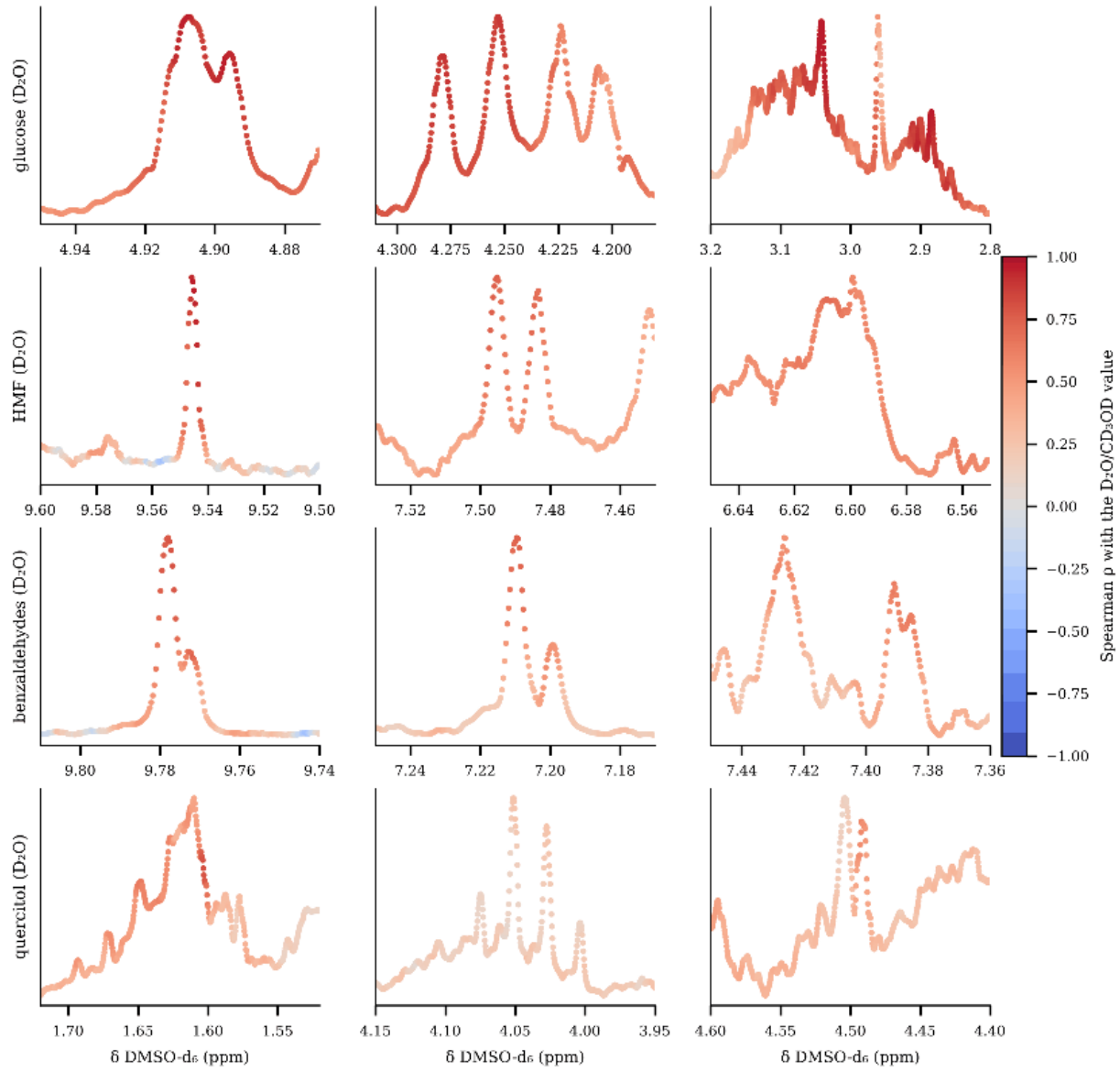

Figure 25. Statistical heterospectroscopy across solvents. For each quantity measured in the D_2O/CD_3OD residues (rows: corrected glucose, HMF, the benzaldehydes and quercitol), the Spearman correlation with every point of the aligned DMSO-d_6 spectra of the same products is shown as the median DMSO-d_6 spectrum colored by the correlation coefficient, in three windows per quantity.

Table 9. ¹H assignments for the freeze-dried residues in DMSO-d_6 (27 °C, TMS = 0; positions after referencing, Section 5.8).

δ ¹H (ppm)	Multiplicity (*J*, Hz)	Assignment	Status	Evidence
9.778	s	Syringaldehyde CHO	probable	BMRB standard in DMSO-d_6 9.783; area ratio to H-2/6 0.91 per proton [59]
9.772	s	Vanillin CHO	probable	BMRB 9.771; area ratio to H-2 0.95 [59]
9.592–9.599	two d (7.8)	Coniferaldehyde and sinapaldehyde CHO	probable	BMRB 9.599 [59]
9.545	s	5-Hydroxymethylfurfural CHO	confirmed	with H-3 7.489 (d, 3.6); cross-solvent correlation ρ = 0.95 [63]
7.489	d (3.6)	HMF H-3	confirmed	area ratio CHO/H-3 0.85
7.45–7.46	s	Ellagic acid H-5/5′	probable	BMRB 7.462; wine-seasoned malts [59]
7.413; 7.387	dd; d (1.8)	Vanillin H-6; H-2	probable	BMRB 7.419, 7.389 [59]
7.36–7.39	s (broadened)	Unassigned (possibly ellagic acid in an ionized or associated state)	unassigned	Section 9.6

δ ¹H (ppm)	Multiplicity (*J*, Hz)	Assignment	Status	Evidence
7.210	s	Syringaldehyde H-2/6	probable	BMRB 7.215 [59]
7.200	s	Syringic acid H-2/6	tentative	BMRB 7.210; 0.47 of the syringaldehyde H-2/6 area; acetosyringone (7.236) excluded
6.914	s	Gallic acid H-2/6	probable	sharp 2H singlet; DMSO/D_2O recovery 0.78
6.940	s	Unassigned	unassigned	accompanies gallic acid
5.17	d (3.7)	Sucrose Glc H-1	probable	maple-barrel bourbon
4.900	d (3.6)	α-D-Glucopyranose H-1	probable	cross-solvent correlation with D_2O glucose ρ ≥ 0.85; multiplicity varies with OH exchange (Section 9.4)
4.856; 4.846	d (3.3); d (3.3)	β-L-Arabinopyranose H-1; α-D-xylopyranose H-1	probable	couplings and intensity order (Section 9.5)
4.264	d (7.8)	β-D-Glucopyranose H-1	probable	cross-solvent correlation ρ ≥ 0.85
4.217; 4.205	d (6.9); d (8.0)	α-L-Arabinopyranose H-1; β-D-xylopyranose H-1	probable	couplings and intensity order (Section 9.5)
4.04	q (7.1)	Ethyl ester OCH_2	probable	with CH_3 1.172 (t); α-CH_2/OCH_2 = 1.10 [33]
3.31–3.84	broad	Exchange-averaged H_2O and solute OH	solvent	position and width vary with water, acid and sugar content [33]
2.88; 3.04	m	Glucose ring CH (H-2, H-4)	probable	cross-solvent correlation ρ ≥ 0.85
2.500	quintet (^2J(H,D) = 1.85)	DMSO-d_5	solvent	internal tracer of the solvent lot (Section 5.8); ^{13}C satellites at 2.273 and 2.727
2.363–2.433	s	Succinic acid CH_2	probable	position varies with acidity; partial recovery in wine-seasoned residues (0.32 of D_2O)
2.258	t (7.4)	Acyl α-CH_2 of ethyl esters	probable	after subtraction of the DMSO-d_5 ^{13}C satellite
2.181	t (7.4)	Acyl α-CH_2 of free fatty acids	probable	after subtraction of the DMSO-d_5 ^{13}C satellite
1.91	s	Acetic acid CH_3	confirmed	[33]
1.60	m	Quercitol C-6 methylene	probable	cross-solvent correlation with D_2O quercitol ρ = 0.80; recovery 0.97
1.23–1.25	broad	Acyl $(CH_2)_n$	probable	
1.172	t (7.1)	Ethyl ester CH_3	probable	paired with OCH_2 4.04
1.058	t (7.0)	Ethanol CH_3 (retained)	confirmed	[33]
0.83–0.87	m	Acyl and fusel-alcohol CH_3	probable	
0.000	s	TMS (reference; present in 13 of 29 spectra)	reference	Section 5.8

9.4 Hydroxyl exchange and the carbohydrate signals

In D_2O every hydroxyl proton is replaced by deuterium, and the anomeric protons of the sugars appear as simple doublets whose shapes do not vary from residue to residue. In DMSO-d_6 the hydroxyl protons remain, they couple to the CH protons beside them, and the rate at which they exchange with water and with each other depends on the water content and the acidity of each residue. Figure 26 shows the consequence. In the sugar-rich Moscatel residue the exchange is fast, the OH coupling is averaged away and α- and β-glucose H-1 appear as clean doublets ($J_{1,2}$ = 3.6 and 7.8 Hz). In the Kavalan ex-bourbon residue the exchange is slow enough for the anomeric doublets of the pentoses to sharpen while the glucose α-H-1 doublet acquires a shoulder, and in the maple bourbon the α-glucose H-1 signal is a triplet-like doublet of doublets coupled to OH-1. In the residue of the Kavalan ex-bourbon oak malt, which lies between these regimes, every anomeric signal is broadened to an envelope of 10–20 Hz. The same anomeric proton can therefore appear as a doublet, a doublet of doublets or a broad hump depending on a property of the residue that has nothing to do with its sugar content.

This is the most important practical difference between the two solvents for the carbohydrates. Integrals over windows wide enough to contain all three shapes remain quantitative, and the multiplet fits used here were given widths of up to 6 Hz to follow the broadened cases, but a binned or full-resolution chemometric model sees an apparent change in the spectrum where the composition has not changed. The exchange-averaged water line moves for the same reason, from δ 3.33 in the driest residues, the value for water in DMSO-d_6 [33], to beyond δ 3.8 in residues rich in acids and sugars. A small addition of D_2O to the DMSO-

d₆, which would exchange the hydroxyl protons and fix the multiplicities, is the obvious remedy and is recommended in Section 17.

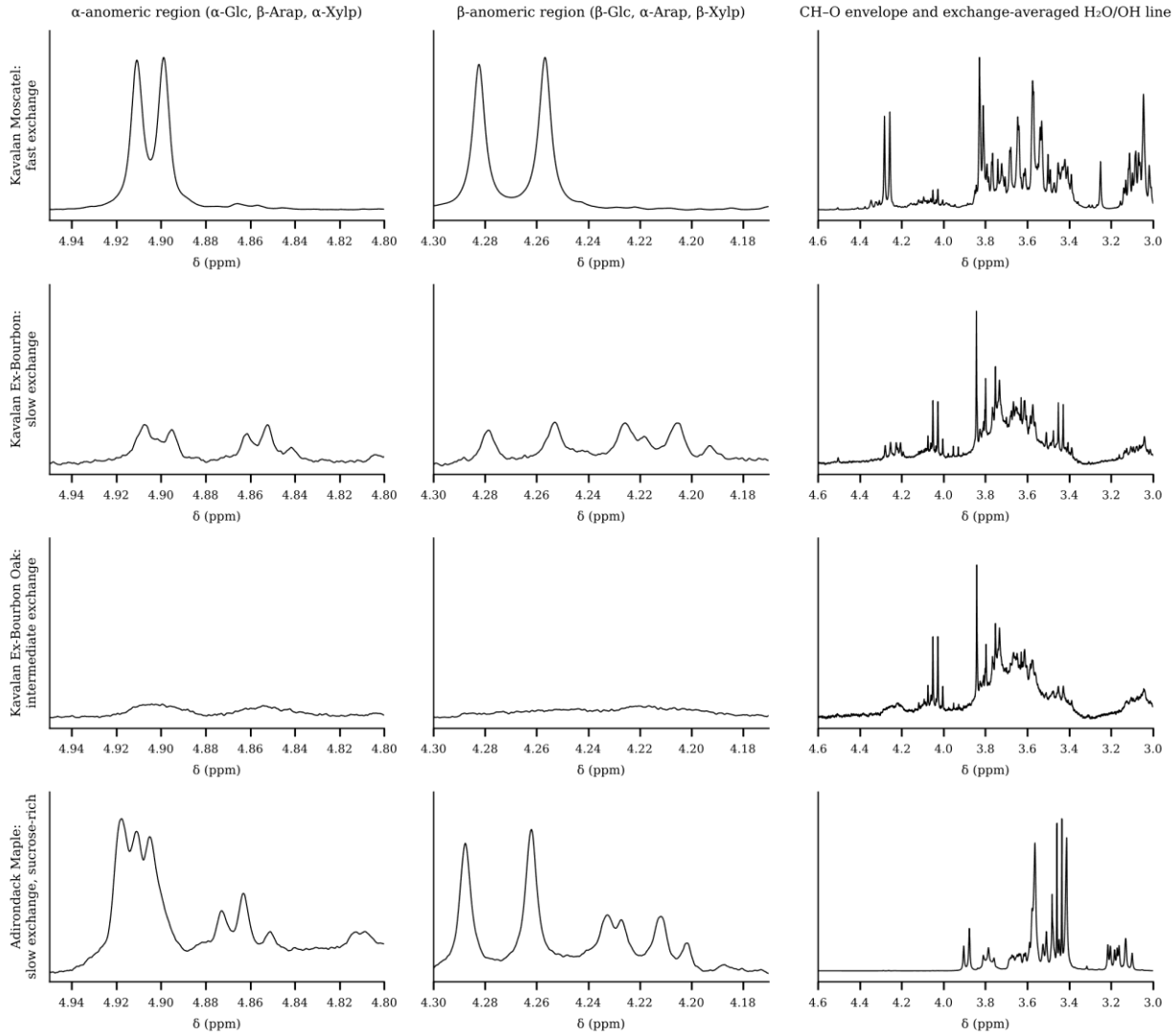

Figure 26. Hydroxyl exchange and the carbohydrate signals in DMSO-d₆. Rows: Kavalan Solist Moscatel (fast exchange), Kavalan Solist Ex-Bourbon (slow exchange), Kavalan Ex-Bourbon Oak (intermediate exchange) and Adirondack Maple Barrel Bourbon (slow exchange, sucrose-rich). Columns: the α-anomeric region (α-glucose, β-arabinopyranose and α-xylopyranose H-1), the β-anomeric region (β-glucose, α-arabinopyranose and β-xylopyranose H-1) and the CH–O region with the exchange-averaged H₂O/OH line. Traces are scaled as described in Section 4.17.

9.5 Pentoses beneath the glucose doublet

The anomeric region of the unsweetened residues in DMSO-d₆ contains more than glucose (Figure 27). Besides the α- and β-glucose doublets at δ 4.900 (J = 3.6 Hz) and 4.264 (J = 7.8 Hz), two doublets with small couplings appear at δ 4.856 and 4.846 (J = 3.3 Hz) and two with large couplings at δ 4.217 (J = 6.9 Hz) and 4.205 (J = 8.0 Hz), and in the residues of new-oak and ex-bourbon whiskeys they are as intense as the glucose signals. Their couplings identify them as anomeric protons in the equatorial (small J) and axial (large J) orientations of two pentopyranoses. The two pairs were assigned to L-arabinose and D-xylose, the pentoses that accompany glucose in the hemicelluloses of oak, on three grounds. In the D₂O/CD₃OD residues of the same kinds of product, where the anomeric doublets of the free sugars can be identified by

their chemical shifts [59], α-arabinopyranose exceeded β-xylopyranose by a median factor of 3.2 (interquartile range 2.9–3.8) and β-arabinopyranose exceeded α-xylopyranose by 2.3 (1.7–2.8); in DMSO-d_6 the stronger large-coupling doublet exceeded the weaker by 3.4 (1.3–5.9) and the stronger small-coupling doublet its partner by 3.9 (2.3–5.2), so the more abundant sugar is the same in both solvents. Second, the two ratios within each solvent are of similar size, as they must be if each pair belongs to one sugar, because the anomeric equilibria of L-arabinopyranose and D-xylopyranose are similar, the anomer with an axial H-1 accounting for about 60 % and 63 % of each in water [65]. Third, the gHSQC spectrum of the concentrated Kavalan residue (Section 8) contains methine correlations at δC 97.8 (δH 4.49–4.52) and 97.7 (δH 4.56), the C-1 shifts of free α-L-arabinopyranose and β-D-xylopyranose (97.6 and 97.5 in water [66]), besides the β-glycosidic and α-arabinofuranosyl correlations of oligosaccharides noted there. The stronger pair was therefore assigned to α-L-arabinopyranose (δ 4.217) and β-L-arabinopyranose (δ 4.856), and the weaker pair to β-D-xylopyranose (δ 4.205) and α-D-xylopyranose (δ 4.846). The scatter of the DMSO-d_6 ratios is wide, because the anomeric signals of this solvent change shape with the rate of hydroxyl exchange (Section 9.4), and the DMSO-d_6 pentose values are reported with that limitation.

Once the pentoses had been recognized in DMSO-d_6 their signals could be found in D_2O/CD_3OD, where they had been overlooked in version 2.0 (Figure 28). The α-arabinopyranose H-1 doublet lies at δ 4.474 (J = 7.8 Hz) and the β-xylopyranose H-1 doublet at δ 4.535 (J = 7.8 Hz), both clear of the glucose β-H-1 doublet at δ 4.598, and the α-xylopyranose H-1 doublet lies at δ 5.151 (J = 3.6 Hz). The β-arabinopyranose H-1 doublet, however, falls within 0.005 ppm of the α-glucose H-1 doublet near δ 5.195, and at 299.99 MHz the two cannot be separated. Two further complications emerged when the anomeric region of every residue was refitted. The anomeric doublets of twenty-three residues were displaced together by more than 0.004 ppm, and by up to 0.029 ppm, from their usual positions; the displacements correlated with the CHD_2OD tracer (Spearman ρ = −0.33), the residues with the largest upfield displacements being the two dissolved in equal volumes of D_2O and CD_3OD and those with the largest downfield displacements early preparations whose CD_3OD volume was not recorded and whose CHD_2OD intensity is about half the usual value. The displacement is therefore the medium effect of a CD_3OD fraction that departs from the nominal 23 % (Section 5.4), and every anomeric doublet was refitted at its usual position plus a residue-specific offset determined by cross-correlation of the anomeric region with a template of the sugar doublets (Section 4.8). The same offset, scaled for each target by its own displacement between D_2O/CD_3OD and D_2O, was applied to every other target in the re-quantitation of this version, except in the two residues dissolved in equal volumes of D_2O and CD_3OD, whose displacements differ from one proton to another and were measured target by target (Section 4.8). Second, the α-glucose doublet itself cannot be used to quantify glucose, because it carries the β-arabinopyranose doublet; glucose was therefore computed from its resolved β-H-1 doublet and the α/β ratio of glucose measured in the 14 glucose-rich, pentose-poor residues of the same solvent system (cognacs and sweetened rums), which was 0.860 with a robust standard deviation of 0.013. That ratio departs from the anomeric equilibrium of glucose in water (about 0.6 [65]), as does the α/β ratio of xylose in the whiskey residues (0.83, interquartile range 0.70–1.34, against about 0.58 in water), presumably because of the CD_3OD content and of incomplete mutarotation after the dried glass dissolves; a ratio measured in the same solvent system, rather than a literature value, is therefore essential.

The remainder of the α-region after subtraction of the α-glucose contribution provides the test of the arabinose assignment. In the 133 residues in which that remainder is well determined (glucose not so abundant that the uncertainty of the subtraction exceeds 15 % of the α-arabinopyranose doublet), it was 0.53 times the α-arabinopyranose doublet (median; interquartile range 0.43–0.60), against a β/α pyranose ratio of 0.59 for L-arabinose at equilibrium in water [65]; restricted to the 102 residues with ratios between 0.25 and 1.0 the median was 0.53 (interquartile range 0.49–0.60), so the agreement does not depend on that

restriction. The agreement confirms both the assignment and the decomposition, and arabinose was therefore computed from the resolved α-arabinopyranose doublet multiplied by one plus that median ratio, the furanose forms (about 4.5 % in water) being neglected. In twenty-two residues, however, the δ 4.474 doublet was well determined but was accompanied by less than a quarter of the expected β-arabinopyranose intensity. Eight of them are Irish whiskeys, among them six single pot still whiskeys (three Redbreast, two Powers and the Blue Spot), and the others comprise four wine- or sherry-seasoned Kavalan malts, two Tromba tequilas and a sotol, the 50-year-old Armagnac and six American whiskeys, two of them Hudson finishes. In these residues the δ 4.474 doublet is not, or not only, arabinose, and arabinose is not reported for them; the β-anomeric H-1 of a glycoside or of an oligosaccharide is the most plausible alternative. Read as arabinose, the doublet would correspond to more than 200 mg L^{-1} in two of the Redbreast residues, whereas an earlier bottle of the same 12-year-old cask-strength whiskey passed the test at an ordinary level, so the compound responsible varies from one bottling to the next; an HSQC spectrum of one of these residues would identify it.

The consequences for the sugar results of version 2.0 are set out in Table 10. Glucose was quantified there from the sum of its α- and β-H-1 doublets, and in residues that contain arabinose the α-doublet fit absorbed the β-arabinopyranose signal, which inflated both its area and its fitted line width. In the residues whose anomeric region was displaced, the version 2.0 fits, whose positions were fixed for the nominal solvent composition, missed the β-glucose doublet altogether in eight residues and underestimated their glucose, in some cases severely. Both errors have been corrected in this version: the glucose values of Sections 6, 10, 11 and 13 and of the companion workbook are now the corrected values, and the version 2.0 values are retained in the workbook for comparison. For the American whiskeys the median glucose falls from 81.1 mg L^{-1} as reported in version 2.0 to 51.4 mg L^{-1}, while arabinose (57.1 mg L^{-1}) and xylose (19.7 mg L^{-1}) appear for the first time; in the Kavalan malts the corresponding figures are 113.4, 81.3, 49.9 and 13.8 mg L^{-1}, and in the Irish and Scotch whiskies the correction is smaller. For the cognacs and the sweetened rums, whose glucose exceeds their pentoses by one or two orders of magnitude, the arabinose correction is negligible; for the sugar-rich residues with displaced spectra the correction raised glucose substantially and changes the interpretation of four products: Pyrat XO Reserve and a 16-year-old Trinidad rum (Section 6.4), Larsen Aqua Ignis (Section 10.3) and Tromba Reposado (Section 10.6).

The pentoses are chemically informative in their own right. Arabinose was more abundant than glucose in the American whiskeys, all matured in new charred oak, and their median arabinose content was higher than in any other category represented by more than one product, whereas xylose, the backbone sugar of hardwood xylan, was the least abundant of the three. The arabinofuranosyl side chains of arabinoxylans and pectic arabinans are the most acid-labile glycosidic linkages of the plant cell wall, and their preferential release into a mildly acidic spirit, accelerated by the thermal degradation of the stave surface during charring, is the most economical explanation; it agrees with the progressive loss of hemicellulose sugars from the stave during a decade in the barrel [3] and with the increase of arabinose, xylose and glucose with age in brandy matured in oak, in which arabinose was the most discriminating sugar between the stages of the solera [18]. Whether the arabinose-to-glucose ratio can serve as an index of new-oak contact, as the data suggest, would require spirits of known barrel history.

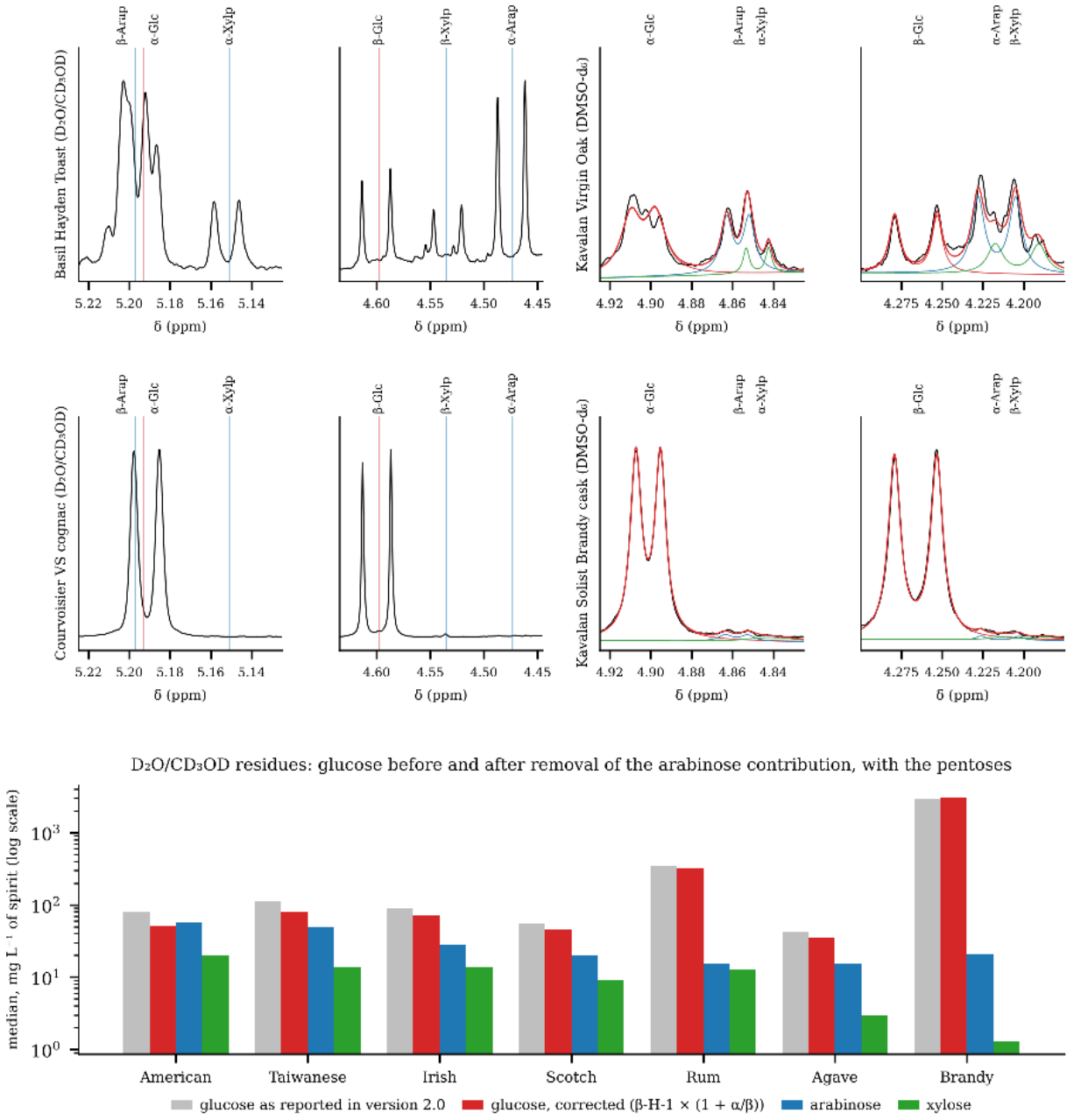

Figure 27. Pentoses in the anomeric regions. Top left: D_2O/CD_3OD spectra of Basil Hayden Toast (rich in pentoses) and Courvoisier VS (rich in glucose, poor in pentoses) with the reference positions (doublet centers) of α-glucose (δ 5.193), β-arabinopyranose (5.197), α-xylopyranose (5.151), β-glucose (4.598), β-xylopyranose (4.535) and α-arabinopyranose (4.474) H-1. Top right: fits of the DMSO-d_6 anomeric regions of Kavalan Artist Series Virgin Oak and Kavalan Solist Brandy cask (black, spectrum; red, total fit; colored lines, components). The spectral panels are scaled as described in Section 4.17. Bottom: category medians in the D_2O/CD_3OD residues of glucose as reported in version 2.0, glucose corrected by the β-H-1 method of Section 9.5, arabinose and xylose (mg L^{-1} of spirit, logarithmic scale).

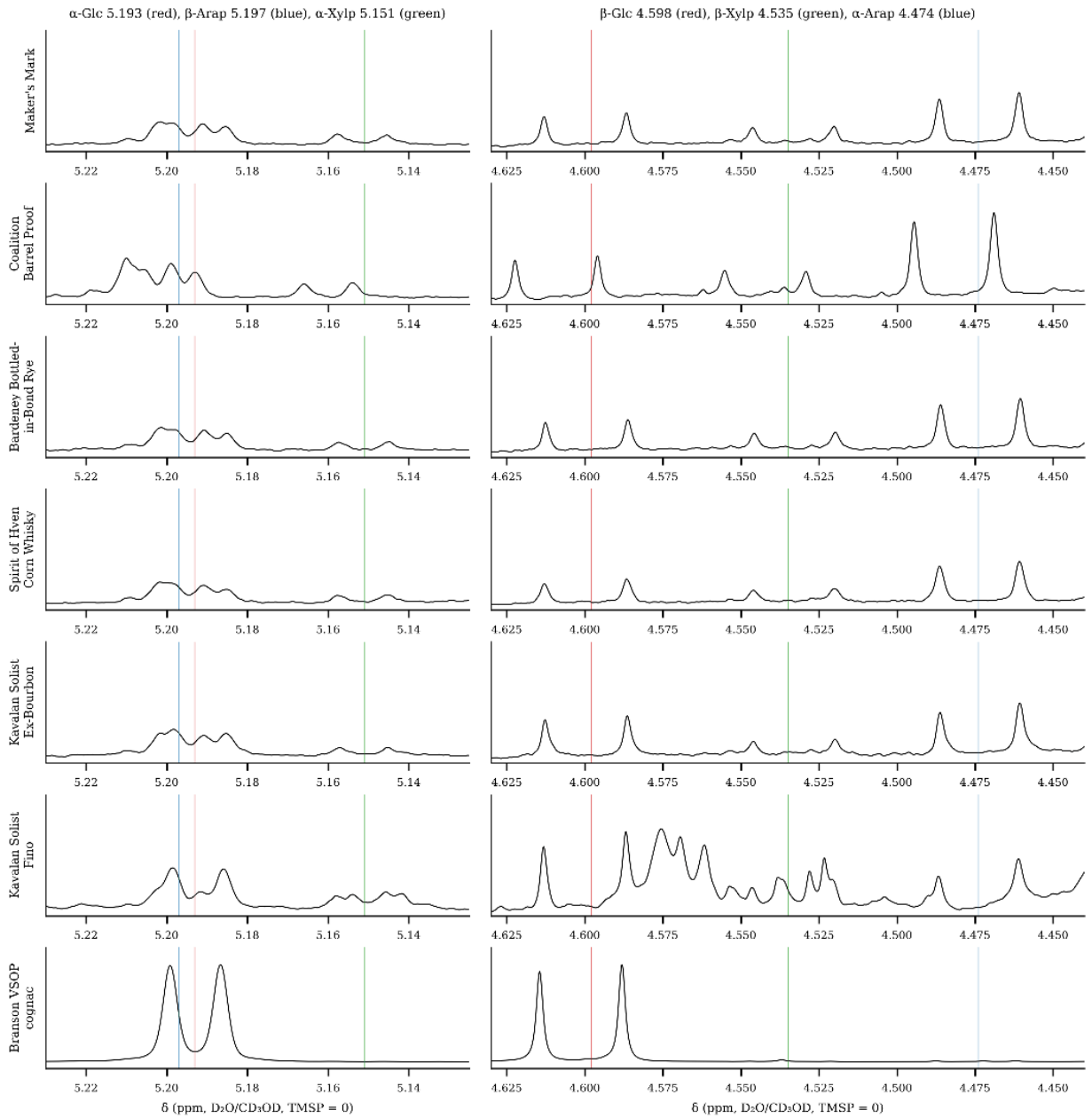

Figure 28. The anomeric regions of seven D_2O/CD_3OD residues with the reference positions (doublet centers) of α-glucose (red), β-arabinopyranose (blue) and α-xylopyranose (green) H-1 (left) and of β-glucose (red), β-xylopyranose (green) and α-arabinopyranose (blue) H-1 (right). The spectrum of Coalition Barrel Proof is displaced downfield by 0.010 ppm by its CD_3OD content (Section 9.5), and the cognac shows glucose alone. Traces are scaled as described in Section 4.17.

Table 10. Arabinose, xylose and glucose in the D_2O/CD_3OD residues: category medians, mg L^{-1} of spirit, with glucose by the method of version 2.0 and the corrected glucose of version 2.1, both on the corrected absolute scale of version 2.2.

Category	Residues	Arabinose	Arabinose confirmed / not confirmed	Xylose	Glucose, version 2.0 method	Glucose, version 2.1 method	Median overestimate in version 2.0 (%)
American whiskey	46	57.1	39/6	19.7	81.1	51.4	75
Taiwanese whisky	25	49.9	19/4	13.8	113	81.3	41
Irish whiskey	30	28.0	20/8	13.9	89.7	71.4	12
Scotch whisky	14	19.9	14/0	9.2	55.9	45.9	21
Japanese whisky	1	28.4	1/0	7.7	111	102	9
Other world whisky	7	27.0	3/0	6.1	22.2	19.4	3

Category	Residues	Arabinose	Arabinose confirmed / not confirmed	Xylose	Glucose, version 2.0 method	Glucose, version 2.1 method	Median overestimate in version 2.0 (%)
Canadian whisky	3	4.4	0/0	14.8	395	396	0
Rum	28	15.3	9/0	12.9	348	322	2
Brandy	14	20.7	0/1	1.3	2928	3078	−1
Agave	14	15.1	6/3	2.9	42.2	35.3	11
Gin	1	—	0/0	2.5	6.7	—	—

Glucose (version 2.1) from the β-H-1 doublet × (1 + α/β), with α/β = 0.860 measured in 14 glucose-rich residues; arabinose from the α-pyranose doublet × (1 + 0.53); xylose from both anomeric doublets. Residue-specific medium offsets applied (Section 9.5). Arabinose medians are taken over the residues in which it was detected and passed the β-anomer test.

9.6 Phenolic aldehydes, phenolic acids and ellagic acid

The aromatic region is better resolved in DMSO-d_6 than in either of the other solvents (Figure 29). The syringaldehyde H-2/6 singlet (δ 7.210) and the syringyl singlet at δ 7.200 are baseline-separated, the vanillin H-6 doublet of doublets and H-2 doublet appear at δ 7.413 and 7.387 next to the HMF H-3 doublet (δ 7.489), and gallic acid gives a sharp singlet at δ 6.914 accompanied by a smaller singlet at δ 6.940 that was not assigned. The median contents in the DMSO-d_6 residues were 14.1 mg L^{-1} of syringaldehyde, 4.46 mg L^{-1} of vanillin, 2.76 mg L^{-1} of HMF and 7.30 mg L^{-1} of gallic acid.

Ellagic acid, which does not dissolve appreciably in D_2O and was not seen there, gave a singlet at δ 7.45–7.46 in thirteen residues (Table 11). It was most abundant in the Kavalan Solist Manzanilla (67.0 mg L^{-1}), the King Car wine-cask selection (31.9 mg L^{-1}), the Paul Chiang puncheon (28.6 mg L^{-1}) and the Solist oloroso cask 005 (20.8 mg L^{-1}), and it was below the detection limit or barely above it in the other residues, including several wine-seasoned malts. Ellagic acid is released from the oak ellagitannins by hydrolysis and is one of the main constituents of the fine sediment that forms in old and cask-strength spirits [1,2], so its absence from a residue does not exclude ellagitannin contact; it may also reflect precipitation in the bottle. Three other wine-seasoned malts (amontillado, fino and Moscatel) and the Madeira cask showed instead a broader singlet at δ 7.36–7.39 of comparable intensity. Its position is consistent with ellagic acid in a partly ionized or metal-associated state, since deprotonation of a phenol shifts the ring protons upfield and wine-seasoned casks carry potassium and other cations from the wine, but this assignment is speculative and the line was reported as unassigned.

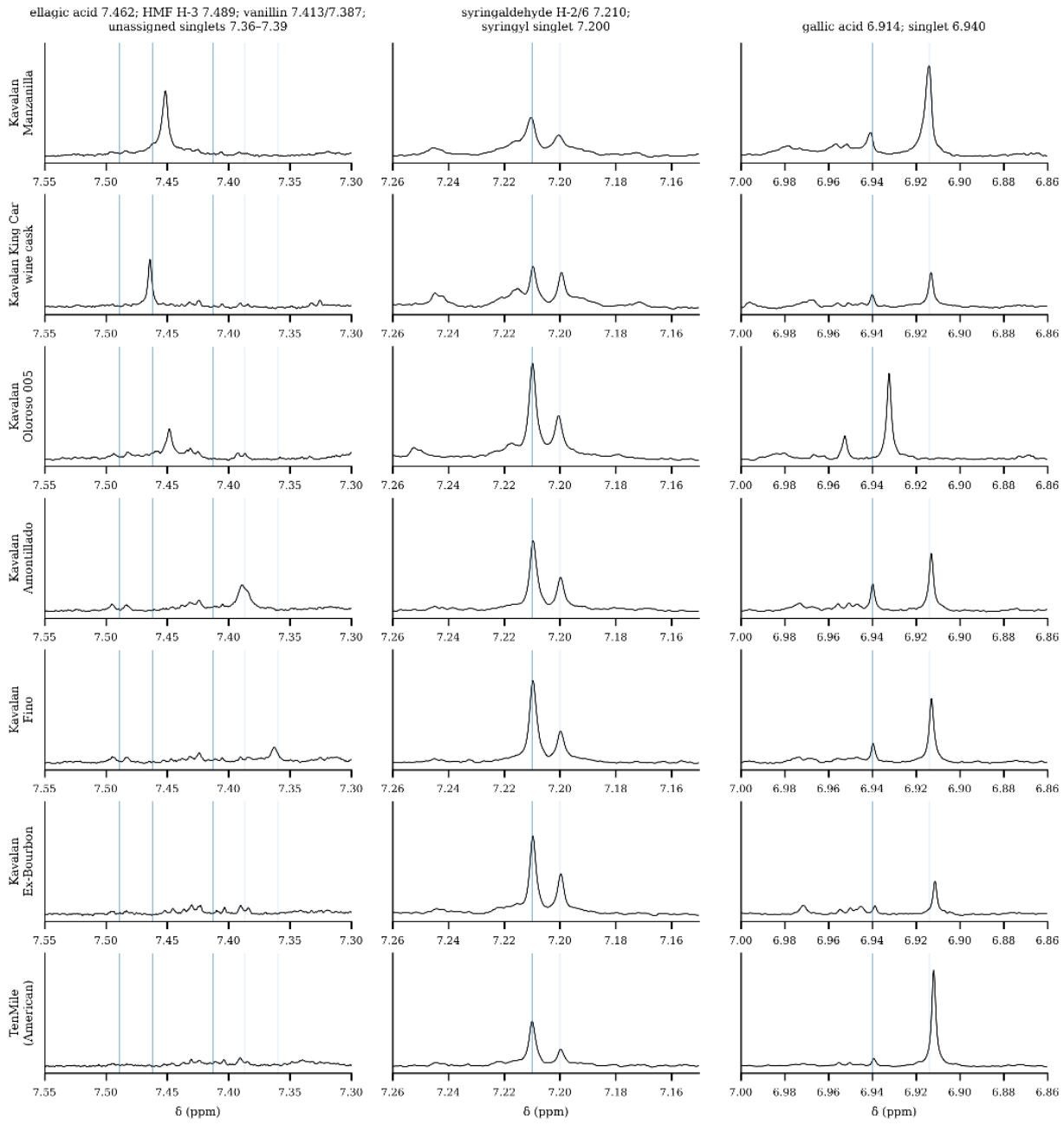

Figure 29. The aromatic region of seven DMSO-d₆ residues. Left: ellagic acid (δ 7.462), HMF H-3 (7.489), vanillin H-6 and H-2 (7.413 and 7.387) and the unassigned singlets at δ 7.36–7.39. Center: the syringaldehyde H-2/6 singlet (7.210) and the syringyl singlet tentatively assigned to syringic acid (7.200). Right: gallic acid (6.914) and the unassigned singlet at 6.940. In the oloroso cask 005 the gallic acid and ellagic acid lines lie away from their reference positions, an example of the compound-specific displacements discussed in Section 9.2. Traces are scaled as described in Section 4.17.

9.7 Lipids, esters and organic acids

The ethyl esters, the free fatty acids and the unsaturated acyl chains, which were measured in $CDCl_3$ in the two-solvent scheme, are all visible in DMSO-d₆. The ester OCH_2 quartet (δ 4.04, J = 7.1 Hz) and its methyl triplet (δ 1.17) are resolved, and the α-methylene triplets of the ethyl esters (δ 2.258) and of the free acids (δ 2.181) are resolved from each other once the upfield ^{13}C satellite of DMSO-d₅ has been subtracted (Figure 30a, b). The satellite, 0.535 % of a line carrying about 60 mmol L^{-1} of ^1H in the tube, is as large as the ester

triplet itself in some residues, and a model fitted to the clean downfield satellite was subtracted before the triplets were fitted. Relative to the $CDCl_3$ residues of the same products, DMSO-d_6 recovered 70 % of the ester groups and 63 % of the olefinic protons but 137 % of the α-methylene protons, which indicates that the long-chain esters and the unsaturated lipids dissolve less completely in DMSO-d_6 than in $CDCl_3$, whereas the free fatty acids dissolve at least as well.

Succinic acid behaved differently from every other polar congener. Its singlet appeared anywhere between δ 2.363 and 2.433, its position following the acidity and cation content of the residue, and in two wine-seasoned residues (the Solist oloroso 005 and the Madeira cask) it lay outside the position window used for the others, where the first fits had measured a weaker neighboring feature; both were refitted at the observed positions (Appendix C). In every wine-seasoned residue the singlet carried only a fraction of the succinic acid found in the D_2O/CD_3OD residue of the same product (median ratio 0.32, range 0.16–0.47, over eleven products; Figure 30c, d), and an integral over the whole region δ 2.325–2.445, which contains every singlet that could be succinic acid, gave a median ratio of 0.27, so the deficit is not an artifact of the fitting window. In the other residues, whose succinic acid content is an order of magnitude lower, unassigned singlets between δ 2.33 and 2.36 prevent a meaningful comparison, and the median ratio of 0.30 for the fitted singlet alone should not be read as a deficit. The deficit in the wine-seasoned residues was not a matter of overlap: in the two residues whose DMSO-d_5 line had moved away from δ 2.50 (Section 5.8) no additional succinic acid appeared beneath its usual position, and the quintet line shapes of the other residues showed no excess intensity where a hidden singlet would fall. The most likely explanation is that part of the succinic acid in a dry wine-seasoned residue is present as salts of the potassium and calcium carried over from the wine, which dissolve in water but not in DMSO; gallic acid, whose DMSO-d_6 recovery was 0.78, may be affected in the same way. Together with the loss of glycerol under the water line and of the lactic acid methyl doublet under the lipid envelope at δ 1.22, this leaves DMSO-d_6 without a reliable measure of any of the three congeners that carry the wine-cask signature in D_2O (Section 10.6). Acetic acid (δ 1.91) was found at 2.66 times its D_2O level and retained ethanol at 1.56 times, but both are semi-volatile, and their retention differs from one preparation to another as it does between the D_2O preparations themselves.

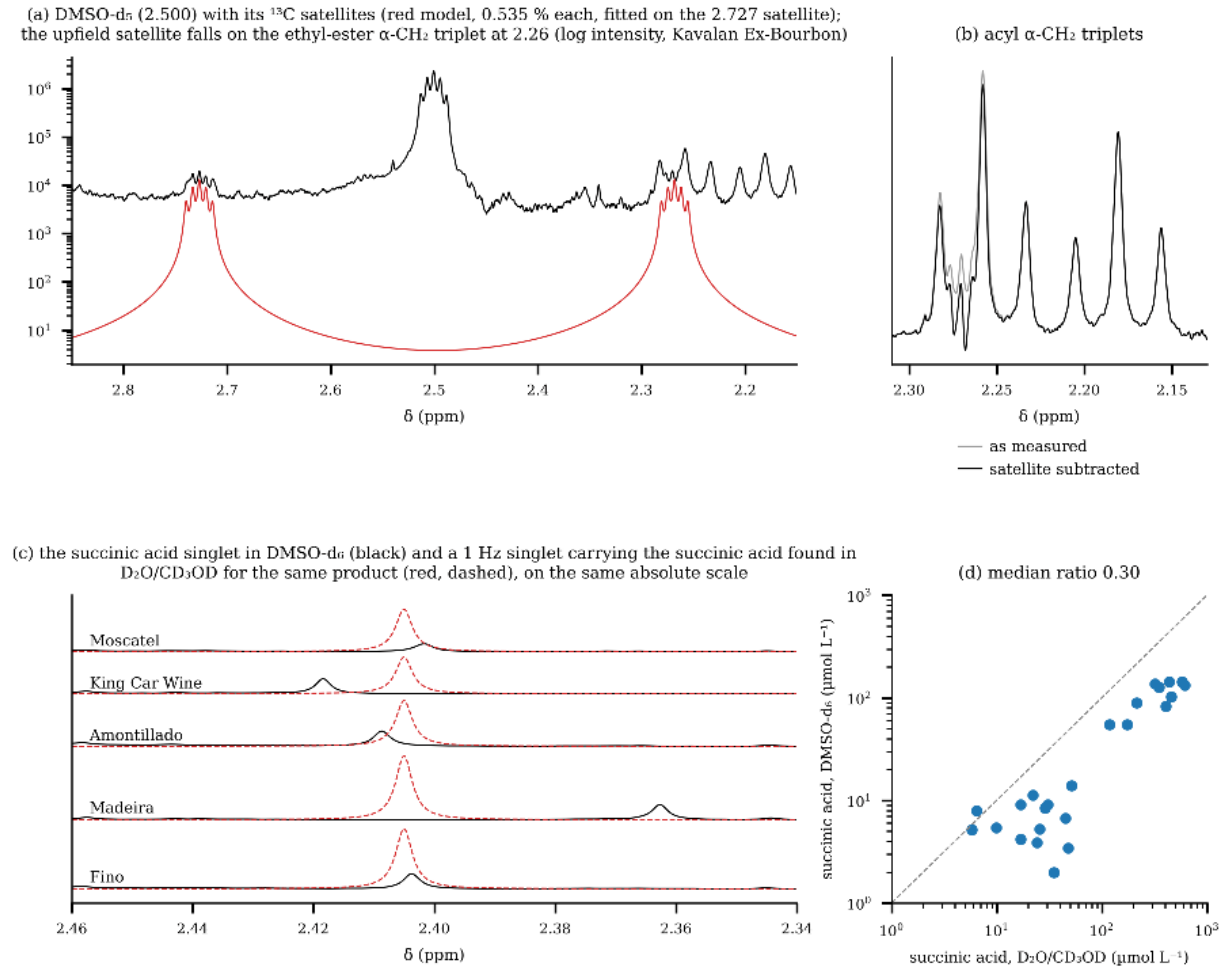

Figure 30. Acids, DMSO-d₅ and its satellites. (a) The DMSO-d₅ quintet and its ¹³C satellites in the spectrum of Kavalan Solist Ex-Bourbon on a logarithmic intensity scale (black), with the model of both satellites, each 0.535 % of the main line and fitted on the downfield satellite at δ 2.727 (red). (b) The acyl α-methylene triplets of ethyl esters (δ 2.258) and free fatty acids (δ 2.181) before and after subtraction of the upfield satellite. (c) The succinic acid singlet in five wine-seasoned residues (black) and a 1 Hz singlet carrying the succinic acid found in the D₂O/CD₃OD residue of the same product (red, dashed), on the same absolute scale; the position of the singlet varies from residue to residue. (d) Succinic acid in the DMSO-d₆ residues against the D₂O/CD₃OD residues of the same products, with the line of equality (dashed).

9.8 Quantitation and recovery across solvents

The congeners quantified in DMSO-d₆ are summarized by residue group in Table 11, the individual results are listed in Appendix B (Table 33), and representative fits are shown in Figure 31. The comparison with the other solvents (Table 12, Figure 32) is the most direct test of whether one solvent can replace two, and the result is mixed. Three compounds that dissolve completely in both D₂O/CD₃OD and DMSO-d₆ and that are quantified from isolated, unambiguous resonances were recovered almost exactly: HMF (median DMSO/D₂O ratio 1.08, interquartile range 0.92–1.27), the benzaldehydes (1.11) and quercitol (0.97). Because these residues were prepared from separate aliquots, often more than a year apart, and acquired in different solvents with the same external standard, the agreement confirms both the transfer of the absolute scale to DMSO-d₆ (Section 5.8) and the stability of the products in the bottle. The phenolic aldehydes were recovered more completely in DMSO-d₆ than in $CDCl_3$ (syringaldehyde H-2/6 147 %, benzaldehyde formyl 131 %), which confirms that the $CDCl_3$ extraction of a sugar-containing residue is incomplete for compounds of intermediate polarity.

Two discrepancies remain unexplained. The glucose found in DMSO-d_6 exceeded the D_2O/CD_3OD value by a median factor of 1.97, but the excess is not proportional: the ratio ranged from 1.5 to 9.9 in the residues with less than 60 mg L^{-1} of glucose and fell to 1.23–1.45 in the three sugar-rich residues, whose anomeric doublets are clean in both solvents (Spearman $\rho = -0.73$ between the ratio and the glucose content). A robust (Theil–Sen) regression of the DMSO-d_6 values on the D_2O/CD_3OD values gave a slope of 1.23 (95 % confidence interval 0.93–1.40) and an intercept of 77 mg L^{-1}. Most of the excess is therefore an additive contribution of the order of 77 mg L^{-1}, present whatever the glucose content, which is the signature of another signal beneath the glucose windows in DMSO-d_6, for example a hydroxyl doublet of a sugar or polyol whose exchange is slow enough for it to stay sharp, rather than of an error of scale. Arabinose behaved differently, with a proportional excess (slope 1.30, intercept 2 mg L^{-1}; median ratio 1.34), and the DMSO-d_6 xylose values did not follow the D_2O/CD_3OD values at all (slope 0.02), which shows that the weaker pentose doublets cannot be quantified reliably in this solvent. An error of the absolute scale would affect HMF and quercitol equally, and different relaxation behavior cannot account for more than about 15 % at the repetition times used. For the proportional part of the excess, a difference between the preparations, a slow hydrolysis of glycosidically bound sugars in the bottle between the two preparations, or an incomplete dissolution of part of the carbohydrate in D_2O/CD_3OD remain possible, and these could be separated only by preparing the same aliquot in both solvents. The cinnamaldehydes gave a DMSO/D_2O ratio of 3.15 with a wide scatter, which reflects the difficulty of apportioning the D_2O formyl region between the cinnamaldehyde doublets and the broad furanic envelope rather than a property of the solvent; in DMSO-d_6 the doublets are clear of that envelope.

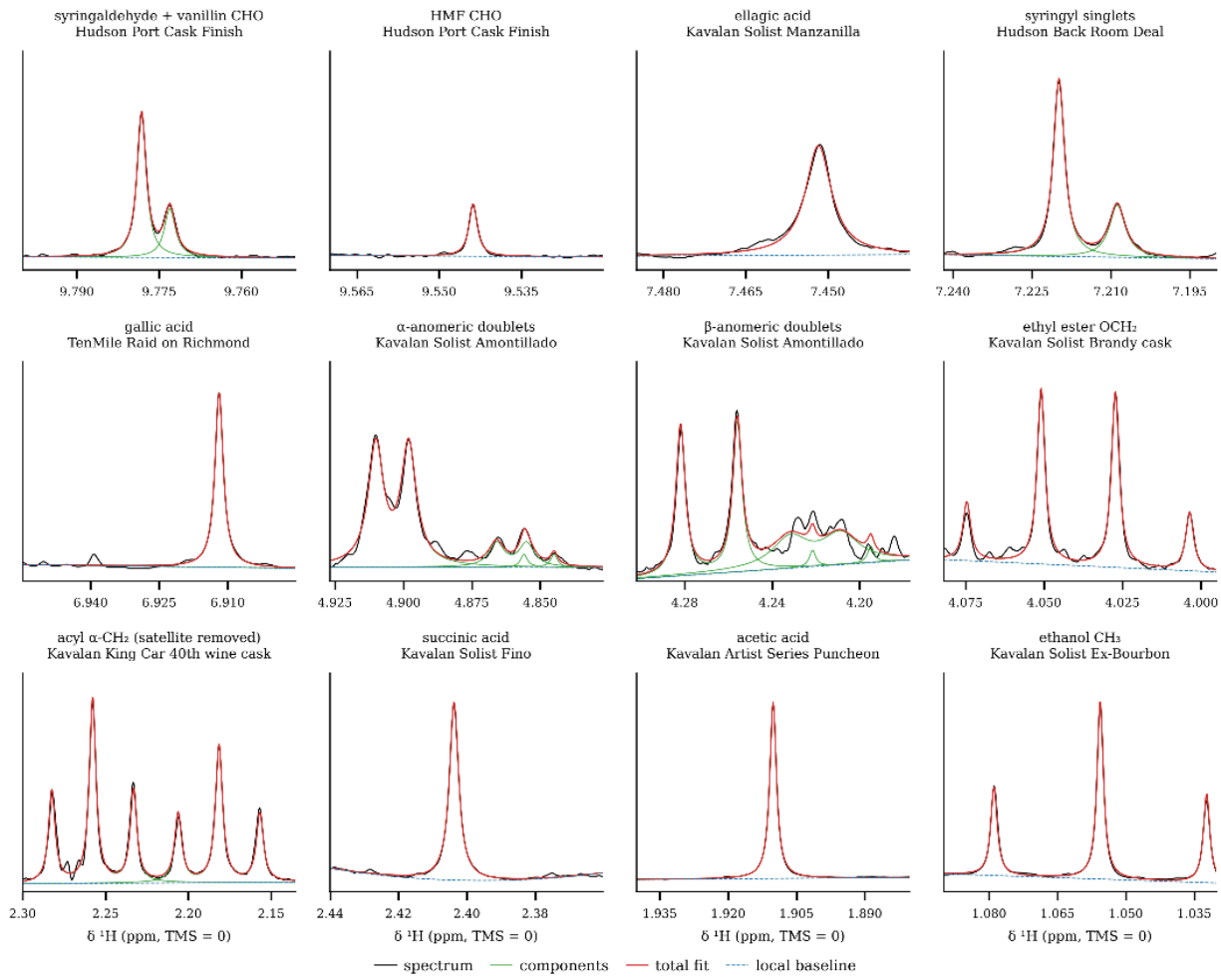

Figure 31. Constrained multiplet fits in the DMSO-d₆ residues, each window shown for the residue named in its title, one in which the signal is among the strongest of the set: experimental spectrum (black), fitted components (green), total fit (red) and local baseline (blue, dashed). The spectrum, the components, the fit and the baseline of a panel share one scale, set as described in Section 4.17.

Table 11. Congeners quantified in the DMSO-d₆ residues: medians by residue group, mg L^{-1} of spirit (number of residues in which the congener was quantified).

Congener	Kavalan, wine-seasoned (n = 11)	Kavalan, other casks (n = 10)	American whiskey (n = 6)	Scotch whisky (n = 1)	Rum (n = 1)
Syringaldehyde	15.3 (11)	12.8 (10)	15.7 (5)	—	—
Vanillin	4.8 (11)	4.3 (10)	4.4 (5)	0.13 (1)	6.7 (1)
Cinnamaldehydes	3.9 (11)	2.9 (10)	10.7 (6)	1.7 (1)	2.4 (1)
HMF	3.3 (11)	2.3 (10)	3.4 (6)	1.4 (1)	0.73 (1)
Ellagic acid	26.3 (4)	0.42 (4)	1.1 (5)	—	—
Gallic acid	11.3 (10)	2.4 (9)	4.7 (6)	—	0.98 (1)
Glucose	180 (11)	160 (10)	179 (6)	47.3 (1)	48.2 (1)
Arabinose	66.6 (11)	69.0 (10)	80.8 (6)	11.5 (1)	36.3 (1)
Xylose	22.5 (10)	19.5 (10)	12.7 (5)	10.3 (1)	12.3 (1)
Quercitol	74.5 (11)	58.8 (10)	77.7 (6)	25.7 (1)	40.1 (1)
Succinic acid	11.0 (11)	0.86 (10)	1.2 (6)	0.61 (1)	0.54 (1)
Acetic acid	2.4 (11)	1.9 (10)	2.2 (5)	—	7.2 (1)
Ethanol (retained)	7.3 (11)	6.6 (10)	22.9 (4)	2.2 (1)	1.1 (1)

Pentoses and glucose from the anomeric doublets (Section 9.5); arabinose and xylose assignments probable. Glucose includes an additive contribution of about 77 mg L^{-1} from an unidentified overlapping signal, and xylose does not follow the D_2O/CD_3OD values (Section 9.8); both are upper bounds. Ellagic acid quantified only in DMSO-d_6.

Table 12. Recovery in DMSO-d_6 relative to the D_2O/CD_3OD or $CDCl_3$ residue of the same product (ratio of absolute concentrations).

Congener	Reference solvent	Products	Median ratio	Interquartile range
HMF	D_2O/CD_3OD	24	1.08	0.92–1.27
Benzaldehydes	D_2O/CD_3OD	23	1.11	0.99–1.27
Cinnamaldehydes	D_2O/CD_3OD	17	3.15	0.91–4.21
Quercitol	D_2O/CD_3OD	24	0.97	0.88–1.08
Gallic acid	D_2O/CD_3OD	21	0.78	0.63–0.96
Succinic acid	D_2O/CD_3OD	24	0.30	0.21–0.44
Acetic acid	D_2O/CD_3OD	23	2.66	0.97–3.75
Ethanol (retained)	D_2O/CD_3OD	22	1.56	0.52–2.84
Glucose	D_2O/CD_3OD	23	1.97	1.52–2.35
Arabinose	D_2O/CD_3OD	22	1.34	1.18–1.68
Xylose	D_2O/CD_3OD	20	1.34	0.61–1.81
Syringaldehyde H-2/6	$CDCl_3$	10	1.47	1.34–1.62
Benzaldehydes (CHO)	$CDCl_3$	10	1.31	1.13–1.47
Cinnamaldehydes	$CDCl_3$	8	1.45	0.73–1.83
Ethyl esters (OCH_2)	$CDCl_3$	9	0.70	0.49–1.25
Acyl α-CH_2	$CDCl_3$	9	1.37	0.86–1.53
Olefinic CH=CH	$CDCl_3$	10	0.63	0.41–0.86
Acetic acid	$CDCl_3$	10	3.55	2.38–9.46

Ester groups in $CDCl_3$: FAEE plus twice the diethyl succinate. Glucose, arabinose and xylose in D_2O/CD_3OD as revised in Section 9.5.

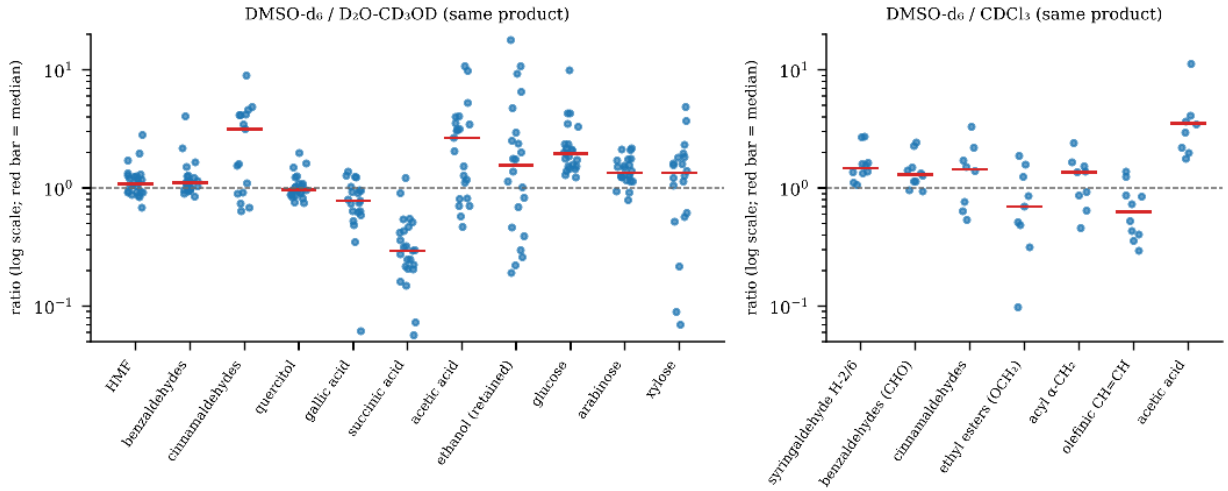

Figure 32. Recovery of the same compounds in DMSO-d_6 relative to the D_2O/CD_3OD residue (left) and to the $CDCl_3$ residue (right) of the same product: ratio of the absolute concentrations for each matched product (points) and the median (red bar), on a logarithmic scale; the dashed line marks equal recovery.

9.9 Two label conflicts resolved by chemistry

The DMSO-d_6 spectrum titled "Kavalan Distillery Reserve Rum Cask" carries the sample identifier Kavalan-043, which in the D_2O and $CDCl_3$ data bases belongs to the Kavalan Solist Vinho Barrique; the rum-cask product is sample 343. Against the 043 residues its congeners fall within the range of the other matched products (HMF 1.28, benzaldehydes 1.11, quercitol 0.88), whereas against both 343 residues every congener is two to three times too abundant (Figure 33a), and its succinic acid is that of a wine-seasoned

cask. The residue is therefore the Vinho Barrique, and the title is wrong. The spectrum titled "Kavalan-026 Paul Chiang Peated Malt" is the decay that was filed under two titles; its parameter set records sample 034 (Kavalan Solist Oloroso Sherry Oak), and against the $CDCl_3$ residue of 026 its aldehydes, esters and lipids are recovered at one-third to two-thirds of the levels seen for every other matched product, whereas against the 034 residues its D_2O congeners fall within the usual range and its succinic acid is that of an oloroso cask (Figure 33b). The decay is that of 034, and there is no DMSO-d_6 spectrum of 026. Both corrections are recorded in the data-integrity log (Appendix C) and are used throughout.

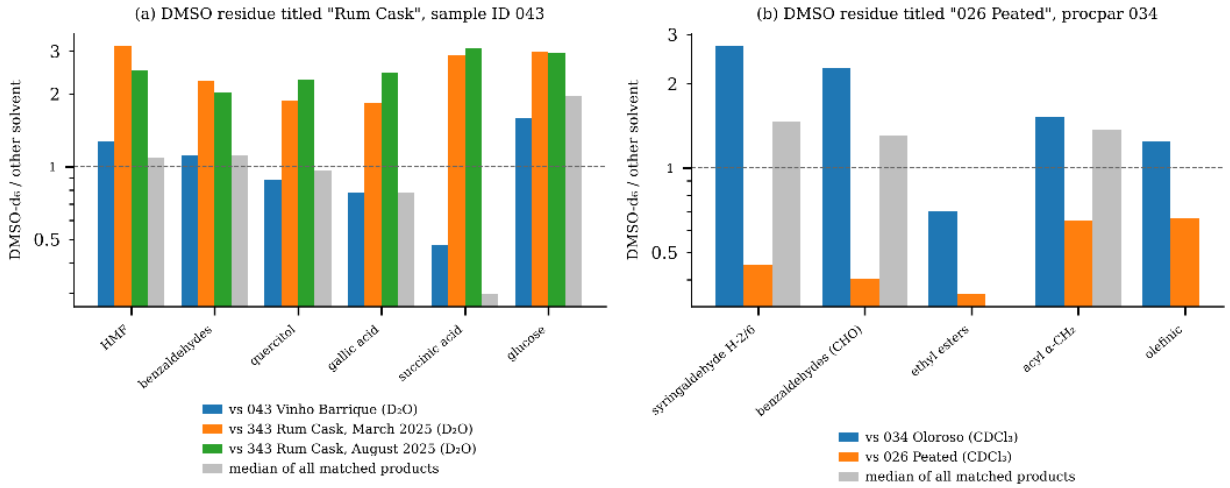

Figure 33. Two label conflicts resolved by chemistry. (a) Congener ratios of the DMSO-d_6 residue titled "Kavalan Distillery Reserve Rum Cask" (sample identifier 043) to the D_2O/CD_3OD residues of 043 (Solist Vinho Barrique) and of two preparations of 343 (Distillery Reserve Rum Cask), with the median ratio over all matched products (gray). (b) Congener ratios of the DMSO-d_6 residue titled "Kavalan-026 Paul Chiang Peated Malt" (parameter set: sample 034) to the $CDCl_3$ residues of 034 (Solist Oloroso Sherry Oak) and of 026, with the median ratio over all matched products (gray).

10. Quantitative results

10.1 Scope

The D_2O/CD_3OD residues yielded 17 quantified congeners or congener groups and the $CDCl_3$ residues 13, and the full results for every residue are listed in Appendix B and in the companion workbook. Table 13 lists the analytes with the signal used, the assumed T_1, the median detection limit and the median sensitivity to a doubling of T_1. Table 14 and Table 15 give the class medians and interquartile ranges, and Figure 34 shows the distributions by class on logarithmic axes. All concentrations are expressed per liter of the original spirit. Concentrations below the detection limit are reported as not detected, values between the detection and quantitation limits as traces, and values flagged for overlap or as upper bounds are identified as such; the semi-volatile congeners (ethanol, the fusel alcohols, 2-phenylethanol and acetic acid) are reported for completeness but are not concentrations in the spirit.

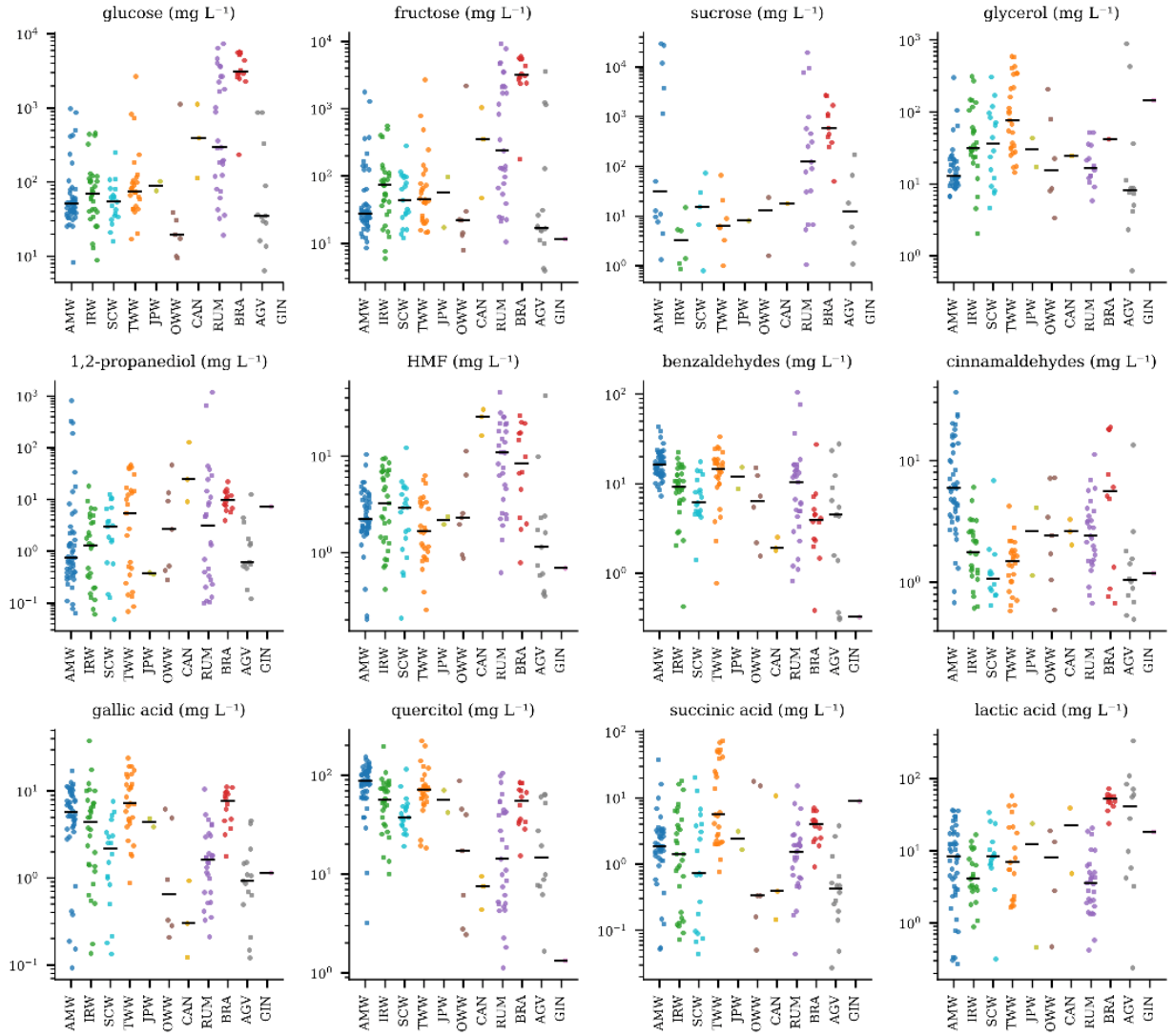

Figure 34. Distribution of twelve quantified congeners by spirit category (mg L^{-1} of spirit, logarithmic scale). Each point is one D_2O/CD_3OD residue and the bars mark the category medians; results reported as not detected are omitted. AMW American whiskey, IRW Irish whiskey, SCW Scotch whisky, TWW Taiwanese whisky, JPW Japanese whisky, OWW other world whisky, CAN Canadian whisky, RUM rum, BRA brandy, AGV agave spirits, GIN gin.

Table 13. Quantified congeners: signal used, assumed T_1, number of residues quantified / traces / not detected, median detection limit and median sensitivity to a doubling of T_1.

Solvent	Analyte	Signal	T_1 (s)	quant / tr / nd (/ nq)	Median LOD (mg L^{-1})	2T_1 effect (%)
D$_2$O/CD$_3$OD	glucose	α- and β-H-1 doublets (δ 5.192, 4.599)	2.0	180 / 1 / 2	0.31	2.6
D$_2$O/CD$_3$OD	fructose	β-furanose H-3/H-4 window (δ 4.058–4.092)	1.5	182 / 0 / 1	0.81	1.0
D$_2$O/CD$_3$OD	sucrose	Glc H-1 (δ 5.393) and Fru H-3 (δ 4.180) doublets	1.5	44 / 15 / 124	0.60	1.0
D$_2$O/CD$_3$OD	glycerol	H-1a/H-3a (δ 3.533) and H-1b/H-3b (δ 3.626) dd	2.3	134 / 1 / 2 / 47	0.14	3.7
D$_2$O/CD$_3$OD	1,2-propanediol	CH$_3$ doublet (δ 1.129), joint fit with ethanol	2.0	151 / 20 / 12	0.052	2.6
D$_2$O/CD$_3$OD	quercitol	H-6ax ddd (δ 1.797)	1.0	175 / 4 / 4	0.62	0.2
D$_2$O/CD$_3$OD	lactic acid	CH$_3$ doublet (δ 1.34–1.40), joint fit with isoamyl CH$_2$	3.0	149 / 1 / 33	0.066	6.7
D$_2$O/CD$_3$OD	succinic acid	CH$_2$ singlet (δ 2.61–2.65)	2.0	158 / 15 / 10	0.035	2.6
D$_2$O/CD$_3$OD	acetic acid*	CH$_3$ singlet (δ 2.00–2.12)	3.0	178 / 5 / 0	0.024	6.7
D$_2$O/CD$_3$OD	gallic acid	H-2/6 singlet (δ 7.10–7.13)	4.0	158 / 17 / 8	0.10	10.9
D$_2$O/CD$_3$OD	HMF	CHO singlet (δ 9.450)	4.0	171 / 10 / 2	0.15	10.9
D$_2$O/CD$_3$OD	benzaldehydes	vanillin + syringaldehyde CHO (δ 9.70–9.72)	4.0	172 / 9 / 2	0.28	10.9
D$_2$O/CD$_3$OD	cinnamaldehydes	coniferaldehyde + sinapaldehyde CHO doublets (δ 9.47–9.49)	4.0	87 / 73 / 23	0.57	10.9
D$_2$O/CD$_3$OD	syringyl index	syringyl H-2/6 singlet (δ 7.30), as syringaldehyde	3.0	128 / 9 / 46	0.11	6.7
D$_2$O/CD$_3$OD	furanic envelope	broad formyl envelope (δ 9.47–9.50), µmol L^{-1}	4.0	130 / 15 / 38	1.2	10.9
D$_2$O/CD$_3$OD	ethanol*	CH$_3$ triplet (δ 1.172)	3.0	165 / 4 / 14	0.044	6.7
D$_2$O/CD$_3$OD	fusel alcohols*	CH$_3$ doublets (δ 0.884), as isoamyl alcohol	3.0	165 / 7 / 11	0.032	6.7
CDCl$_3$	vanillin	CHO singlet (δ 9.833)	4.0	56 / 3 / 2	0.16	14.4
CDCl$_3$	syringaldehyde	CHO singlet (δ 9.824)	4.0	57 / 3 / 1	0.20	14.4
CDCl$_3$	cinnamaldehydes	CHO doublets (δ 9.659–9.666)	3.0	45 / 8 / 8	0.43	9.5
CDCl$_3$	sinapaldehyde	H-2/6 singlet (δ 6.818)	3.0	40 / 8 / 13	0.11	9.5
CDCl$_3$	HMF + EMF	CHO singlet (δ 9.619), as HMF	4.0	58 / 1 / 2	0.14	14.4
CDCl$_3$	unassigned CHO	CHO singlet (δ 9.809), µmol L^{-1}	4.0	38 / 15 / 8	1.1	14.4
CDCl$_3$	diethyl succinate	OCH$_2$ quartet (δ 4.146)	2.0	54 / 5 / 2	0.13	4.3
CDCl$_3$	FAEE	OCH$_2$ quartet (δ 4.124), µmol L^{-1}	2.0	52 / 1 / 8	1.5	4.3
CDCl$_3$	acyl chains	α-CH$_2$ triplets (δ 2.285 + 2.355), µmol L^{-1}	1.0	61 / 0 / 0	1.9	0.4
CDCl$_3$	double bonds	CH=CH (δ 5.31–5.42), µmol L^{-1}	1.5	61 / 0 / 0	1.7	2.0
CDCl$_3$	fusel alcohols*	CH$_3$ doublet (δ 0.921)	3.0	56 / 0 / 5	0.028	9.5
CDCl$_3$	2-phenylethanol*	ArCH$_2$ triplet (δ 2.870)	2.0	14 / 11 / 36	0.15	4.3
CDCl$_3$	acetic acid*	CH$_3$ singlet (δ 2.102)	3.0	53 / 6 / 2	0.021	9.5

D$_2$O counts refer to the 183 D$_2$O/CD$_3$OD residues; nq, not quantifiable (glycerol beside sugars or 1,2-propanediol, or upper limits only). LOD in µmol L^{-1} where no molar mass applies. *Semi-volatile congener, partly retained; not a concentration in the spirit.

Table 14. Class medians (interquartile ranges) of the congeners quantified in the D$_2$O/CD$_3$OD residues, mg L^{-1} of spirit.

Analyte	AMW (*n* = 46)	IRW (*n* = 30)	SCW (*n* = 14)	TWW (*n* = 25)	RUM (*n* = 28)	BRA (*n* = 14)	AGV (*n* = 14)
Glucose	50.5 (38.6–62.4)	71.4 (39.5–110)	45.9 (31.8–65.7)	81.3 (43.5–116)	322 (109–2657)	3078 (2669–5354)	31.1 (8.1–76.2)
Fructose	26.8 (21.9–38.1)	73.0 (31.3–109)	30.9 (25.1–96.9)	46.1 (25.9–107)	244 (63.5–2137)	3208 (2744–5415)	16.7 (12.3–103)
Sucrose	0 (0–0)	0 (0–0)	0 (0–0)	0 (0–1.0)	6.7 (0–267)	412 (99.9–1093)	0 (0–2.4)
Glycerol	12.7 (10.1–20.5)	32.9 (23.5–95.2)	15.7 (8.6–36.3)	97.0 (33.6–337)	16.1 (12.4–22.1)	—	7.5 (3.7–17.5)
1,2-Propanediol	0.76 (0.40–2.2)	1.2 (0.13–3.6)	1.1 (0.13–6.1)	5.4 (0.35–14.8)	1.4 (0.18–12.0)	9.7 (7.0–13.4)	0.57 (0.25–1.4)
Quercitol	89.5 (71.2–110)	56.5 (33.5–78.5)	36.8 (28.7–46.0)	71.5 (53.8–89.9)	12.6 (5.0–41.9)	54.9 (33.9–71.0)	9.4 (3.1–46.1)
Lactic acid	5.3 (1.2–12.8)	3.0 (0–6.0)	6.1 (0.079–8.0)	2.4 (0–17.7)	3.1 (1.6–5.3)	52.6 (46.5–56.8)	18.9 (0.98–68.4)
Succinic acid	1.7 (0.61–3.2)	1.4 (0.16–4.5)	0.47 (0.091–1.4)	6.0 (2.7–41.6)	1.4 (0.58–2.1)	4.0 (2.4–4.6)	0.33 (0.16–0.63)
Gallic acid	5.4 (3.7–8.9)	4.4 (1.2–7.3)	1.5 (0.29–2.9)	8.6 (4.8–14.5)	1.5 (0.64–2.2)	7.7 (4.6–9.5)	1.00 (0.53–1.5)

Analyte	AMW (*n* = 46)	IRW (*n* = 30)	SCW (*n* = 14)	TWW (*n* = 25)	RUM (*n* = 28)	BRA (*n* = 14)	AGV (*n* = 14)
HMF	2.2 (1.7–3.6)	3.0 (1.3–6.9)	2.3 (1.2–4.5)	1.7 (1.1–3.3)	9.5 (3.3–17.9)	8.3 (2.8–17.3)	0.93 (0.38–2.2)
Benzaldehydes	16.8 (12.9–20.5)	9.5 (6.5–12.7)	6.6 (4.9–8.4)	14.4 (9.9–18.6)	9.3 (4.0–15.5)	3.9 (2.3–5.0)	4.4 (0.62–8.4)
Cinnamaldehydes	5.7 (3.3–12.9)	1.3 (0.78–2.5)	0.82 (0–1.2)	1.4 (0.75–1.8)	2.2 (1.5–3.3)	5.0 (0.80–15.4)	0.84 (0.51–1.5)
Syringyl index	13.9 (9.8–16.8)	3.3 (0–6.4)	2.1 (0–4.6)	3.8 (0.27–10.1)	1.9 (0–7.0)	2.3 (0.53–3.2)	0.68 (0.094–3.8)
Furanic envelope (µmol L⁻¹)	12.0 (2.6–45.1)	8.1 (4.6–18.8)	6.8 (1.00–12.1)	5.6 (1.7–14.0)	14.1 (6.2–21.1)	108 (26.7–438)	3.9 (0–21.4)
Acetic acid*	2.5 (1.3–3.1)	0.80 (0.49–1.0)	0.49 (0.30–1.0)	0.98 (0.53–2.2)	0.96 (0.39–1.3)	2.1 (1.4–2.6)	0.66 (0.37–2.6)
Ethanol*	5.0 (2.1–12.6)	0.66 (0.098–4.8)	0.63 (0.23–3.1)	8.5 (2.6–16.2)	33.3 (3.9–349)	437 (138–589)	1.0 (0.24–6.4)
Fusel alcohols*	12.9 (7.5–16.6)	0.88 (0.50–1.6)	1.4 (0.52–2.3)	6.9 (3.8–19.5)	0.28 (0.071–2.4)	13.3 (8.7–18.5)	0.21 (0.011–2.2)

AMW, American whiskey; IRW, Irish; SCW, Scotch; TWW, Taiwanese (Kavalan); RUM, rum; BRA, brandy; AGV, agave spirits. Results below the detection limit are counted as zero; residues in which glycerol was not quantifiable are excluded, and — marks fewer than three quantifiable residues (glycerol in the brandies). *Semi-volatile, partly retained.

Table 15. Class medians (interquartile ranges) of the congeners quantified in the $CDCl_3$ residues, mg L^{-1} of spirit unless stated.

Analyte	AMW (*n* = 31)	IRW (*n* = 5)	SCW (*n* = 6)	TWW (*n* = 12)	RUM (*n* = 3)
Vanillin	3.9 (2.4–4.8)	1.5 (0.75–1.7)	1.8 (1.4–2.5)	3.1 (2.5–3.5)	3.4 (2.0–6.0)
Syringaldehyde	10.3 (7.1–15.2)	1.8 (1.5–2.5)	4.1 (3.3–6.4)	7.7 (5.9–10.8)	1.6 (1.1–5.3)
Cinnamaldehydes	5.4 (3.7–9.2)	0.65 (0–2.5)	1.9 (1.3–2.7)	2.3 (0.45–3.7)	1.0 (1.0–6.4)
Sinapaldehyde	2.8 (0.68–7.9)	0.26 (0.21–0.35)	0.69 (0.38–1.7)	0.10 (0–0.67)	0.28 (0.14–2.8)
HMF + EMF	2.5 (1.6–4.1)	2.8 (2.0–3.1)	1.9 (1.7–2.9)	1.2 (0.82–1.4)	3.2 (2.9–3.2)
Unassigned CHO δ 9.809 (µmol L⁻¹)	4.7 (1.6–8.6)	6.1 (4.7–8.4)	3.7 (2.1–15.9)	10.3 (7.3–17.8)	1.5 (0.76–2.3)
Diethyl succinate	1.7 (1.1–2.6)	2.3 (0.90–2.3)	3.6 (2.3–4.8)	4.5 (1.0–9.3)	0.75 (0.38–1.1)
FAEE (µmol L⁻¹)	47.5 (29.6–72.1)	10.9 (8.7–12.9)	27.1 (4.6–62.5)	242 (121–348)	6.5 (3.3–20.0)
Acyl chains (µmol L⁻¹)	161 (106–191)	93.3 (62.4–147)	206 (199–228)	526 (343–664)	47.9 (40.7–72.4)
Double bonds per chain	0.36 (0.31–0.43)	0.27 (0.23–0.37)	0.15 (0.14–0.16)	0.16 (0.13–0.18)	0.52 (0.51–0.59)
Fusel alcohols*	12.8 (8.4–21.0)	3.9 (1.1–6.4)	1.7 (0.90–1.9)	4.0 (2.1–14.6)	0.065 (0.032–3.4)
2-Phenylethanol*	0 (0–0.54)	0.31 (0–0.75)	0 (0–0.12)	0 (0–0.37)	0.49 (0.34–0.55)

Classes with at least three $CDCl_3$ residues. *Semi-volatile, partly retained.

10.2 Replicate preparations: the largest source of uncertainty

Five products were represented in the D_2O set by two residues each, and their comparison is the most sobering result of the quantitative work. Only one of the five pairs is a true replicate: Kavalan Distillery Reserve Rum Cask was freeze-dried twice from the same sample (343), in March and in August 2025. After the session correction of Section 5.5 its two residues agreed within 9 % in non-volatile proton inventory, and the ratios of the individual congeners lay between 0.74 and 1.12 (median 0.96), which shows that, when the bottle is the same, the preparation and the absolute measurement can be reproduced to about 10 % in total non-volatile content and to within about 25 % for individual congeners.

The other four pairs carry different sample numbers and may come from different bottles or lots, and they differ far more. For Branson XO Premier Cru cognac every non-volatile congener of the second residue was close to half its value in the first (glucose 0.51, fructose 0.50, 1,2-propanediol 0.49, benzaldehydes 0.54, quercitol 0.59 of the first value) and the non-volatile proton inventory was 0.62 of the first, while the

semi-volatile congeners differed by other factors; this is the pattern expected if one residue contained about half as much extract as the other. For Powers Gold Label the second residue contained 0.32 of the non-volatile proton inventory of the first and 0.10 of its sugars, and for Tin Cup Straight Rye the two residues differed by factors of five to twenty in quercitol, 1,2-propanediol and lactic acid in opposite directions, which suggests two different products under one name rather than two preparations of one. The fourth pair, Sotol Ono, differed by a factor of 1.6 in a residue so poor in extract that the ratio is dominated by noise. The CHD_2OD tracer of these residues was normal, so the differences lie in the residues and not in the acquisition. Whether they reflect real differences between bottles and lots or losses of extract during freeze-drying, transfer or redissolution, or an aliquot that was not the volume recorded, cannot be decided from the present data, and the internal checks of Section 5 cannot detect such losses because they test the solution and the acquisition, not the history of the residue. Until replicate preparations of the same bottle and replicate bottles of the same lot have been measured systematically, every single-bottle concentration of this paper should be read as carrying an additional uncertainty of a factor of two or more as a description of the product, even though the measurement of the residue itself is reproducible to about 10 %. The remedy is simple and is set out in Section 17: a weighed, non-volatile recovery standard added to the spirit before freeze-drying, together with gravimetric records of the aliquot and of the residue.

10.3 Sugars and sweetening

The sugar results are summarized in Table 16 for every residue that contained more than 1 g L^{-1} of total sugar, expressed as invert sugar for comparison with the European limits, and in Figure 16. Read against the regulatory frame of Table 1, they fall into four groups.

Table 16. Residues containing more than 1 g L^{-1} of total sugar (g L^{-1} of spirit; total expressed as invert sugar).

Product	Class	Glucose	Fructose	Sucrose	Total (invert)	Fru/Glc
Litchfield Distilling - Maple Bourbon 43%ABV	AMW	0.43	0.22	29.20	31.4	0.5
Barley Creek Distillery - Honey Whiskey 35%ABV	AMW	0.05	nd	26.87	28.3	—
Roatan Rum Company - Nicaragua	RUM	2.66	2.19	19.25	25.1	0.8
Kiss - Black Diamond Premium Dark Rum - Aged 3 Years	RUM	3.62	3.50	9.54	17.2	1.0
Bacardi Reserva Limitada Rum 40%ABV 2009	RUM	6.48	9.22	nd	15.7	1.4
Pyrat XO Reserve Rum 40%ABV	RUM	7.37	7.83	0.09	15.3	1.1
Planteray Rum - Isle of Fiji 40%ABV Double-Barrel…	RUM	3.66	3.50	7.69	15.3	1.0
Larsen - Aqua Ignis Cognac	BRA	5.63	5.55	2.65	14.0	1.0
Taconic Distillery - Double Barrel Maple Bourbon	AMW	0.24	0.13	11.79	12.8	0.5
Branson Cognac - XO Prmier Cru Cognac 40%ABV	BRA	5.73	5.98	0.25	12.0	1.0
Branson Cognac VSOP - Cognac Grande Champagne 40%ABV	BRA	5.24	5.42	1.04	11.8	1.0
Branson - Royal VSOP Cognac	BRA	5.39	5.39	0.59	11.4	1.0
Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	BRA	5.59	5.58	nd	11.2	1.0
Branson - Phantom VS Cognac	BRA	4.39	4.34	1.68	10.5	1.0
English Harbour 5 Yr Antigua Rum 40%ABV - 2010	RUM	4.63	4.82	0.08	9.5	1.0
Trinidad - 16 Years	RUM	4.02	4.69	0.57	9.3	1.2
Hennessy VS Cognac 40%ABV	BRA	2.88	2.91	2.58	8.5	1.0
Courvoisier - XO Cognac	BRA	3.24	3.33	nd	6.6	1.0
Branson Cognac - XO Prmier Cru Cognac 40%ABV	BRA	2.91	2.97	0.44	6.3	1.0
Courvoisier VSOP Cognac 40%ABV	BRA	2.92	3.09	0.30	6.3	1.1
Courvoisier VS Cognac	BRA	2.50	2.37	1.11	6.0	0.9
Bisquit and Debouche - VSOP Cognac	BRA	2.60	2.69	0.39	5.7	1.0
Solist Moscatel Sherry - Single Cask Strength - SMW	TWW	2.67	2.69	nd	5.4	1.0
Diplomatico - Venezuala	RUM	2.66	2.09	0.45	5.2	0.8
Hudson Whiskey NY - Short Stack - NY SRW - Finished…	AMW	0.50	0.37	3.69	4.8	0.7
Merlet - XO Cognac	BRA	2.29	2.38	0.05	4.7	1.0

Product	Class	Glucose	Fructose	Sucrose	Total (invert)	Fru/Glc
English Harbour Reserve - 10 Yr Rum - 40%ABV 2010	RUM	2.24	2.12	0.03	4.4	0.9
Tromba Reposado - French Oak Casks	AGV	0.33	3.54	0.07	3.9	10.6
Kirk & Sweeney - 12 Yr Rum - 40%ABV	RUM	1.67	1.70	0.31	3.7	1.0
Ron del Barilito - Puerto Rican Rum 43%ABV	RUM	1.79	1.72	0.12	3.6	1.0
Coperies - Grench Single Malt Whiskey	OWW	1.12	2.20	nd	3.3	2.0
Hudson Whiskey NY - Port Cask Finish - NY RW - Aged…	AMW	0.98	1.77	nd	2.8	1.8
Brugel - Anejo Superior - Dominican Republic	RUM	0.88	0.83	0.98	2.7	0.9
Cadenheads - Classic Rum Blend - 5Yr - Barbados-Guy…	RUM	1.01	1.17	0.25	2.4	1.2
Gran Centenario - Anejo Tequila - 40%ABV 2006	AGV	0.87	1.14	0.17	2.2	1.3
Forty Creek Barrel Select Canadian Whiskey 40%ABV	CAN	1.12	1.04	0.02	2.2	0.9
High West - A MidwinterNightsDream - Act7 Scene5 Bl…	AMW	0.87	1.28	nd	2.2	1.5
Gran Centanario Legenda Tequila Extra Anejo 40%ABV…	AGV	0.87	1.24	nd	2.1	1.4
High West - Double Rye! Blended SRW Batch 19G26	AMW	0.41	0.36	1.15	2.0	0.9
Kavalan Solist Brandy cask †	TWW	0.83	0.79	<0.01	1.6	0.9
Solist Port - Single Cask Strength - SMW	TWW	0.73	0.49	nd	1.2	0.7
Redbreast 27 Yr Single Pot Still Irish Whiskey Batc…	IRW	0.46	0.55	0.01	1.0	1.2

Limits in the European Union: rum and wine spirit 20 g L^{-1}, brandy 35 g L^{-1} (as invert sugar); whisky may not be sweetened (Table 1). Planteray Isle of Fiji: value for 1 mL of spirit (a third of it if 3 mL was dried; Appendix C). †Title of the D_2O item read “Bespoke Spirits – Rye Whiskey Batch 2020-1”; reassigned by sample ID, DMSO-d_6 counterpart and chemistry (Section 4.1). ‡ Dissolved in equal volumes of D_2O and CD_3OD; fructose not quantifiable (1.9–5.7 g L^{-1} depending on the displacement assumed for its signals), so the total is given as a range (Section 10.3).

The first group comprises products for which sweetening is permitted and was found. The twelve cognac residues other than Larsen Aqua Ignis contained 4.7–12.0 g L^{-1} of sugar expressed as invert sugar, within the limits that apply to wine spirits (20 g L^{-1}) and brandies (35 g L^{-1}) in the European Union and within the obscuration limit of the cognac specification, and in the composition expected of sucrose or invert syrup after inversion in the acidic spirit. Larsen Aqua Ignis, which version 2.0 reported as the exception with almost four times as much fructose as glucose, is not one. Its residue is one of the two that were dissolved in equal volumes of D_2O and CD_3OD, in which the sugar signals are displaced by different amounts, about −0.02 ppm for the α-glucose H-1 and sucrose Fru H-3 doublets, −0.01 ppm for the fructose signals, −0.03 ppm for the β-glucose H-1 doublet and not at all for the sucrose Glc H-1 doublet, and the fits of version 2.0 at fixed positions had missed most of its glucose and misplaced one of its sucrose doublets (Sections 4.8 and 9.5). With every signal fitted at its measured position it contains 5.6 g L^{-1} of glucose (version 2.0: 1.2 g L^{-1}), 5.5 g L^{-1} of fructose and 2.7 g L^{-1} of sucrose, whose two doublets, which disagreed by a factor of two in version 2.0, now agree within 17 % (14.0 g L^{-1} of sugar in total). That is the invert sugar of the other cognacs (fructose-to-glucose ratio 1.0) with an unusually large residue of uninverted sucrose, and the fructose-rich syrup proposed in version 2.0 is withdrawn. The fructose figure is the least secure of the three, because the integral of its window changes steeply with the displacement assumed (1.9–5.7 g L^{-1} over the plausible range), and a preparation in the standard solvent would confirm it. Sweetened rums were as common as unsweetened ones. Roatan rum (25.1 g L^{-1}), Kiss Black Diamond (17.2 g L^{-1}), Bacardi Reserva Limitada (15.7 g L^{-1}), Planteray Isle of Fiji (15.3 g L^{-1}, or 5.1 g L^{-1} if the 3 mL of spirit recorded in its sample name rather than the 1 mL of its title was dried; Appendix C), Pyrat XO Reserve (15.3 g L^{-1}), a 16-year-old Trinidad rum (9.3 g L^{-1}), English Harbour 5 yr (9.5 g L^{-1}) and Diplomático (5.2 g L^{-1}) contained sugar at levels that can only reflect deliberate sweetening. Roatan rum exceeds the 20 g L^{-1} limit that the European Union applies to rum, although it is not necessarily sold under that regime, and United States rules set no specific sweetening limit for rum beyond the general allowance of 2.5 % by volume of the finished product for harmless coloring, flavoring and blending materials [23]. At 25.1 g L^{-1} the sugar occupies about 1.6 % of the volume if it is added as dry sugar but about 2.9 % if it is added as a 65 % syrup, so whether that allowance was respected depends on the form of the addition, which the spectrum cannot reveal.

The second group comprises products for which sweetening is not permitted and was not found. The straight American whiskeys, the Scotch, Irish and other EU whiskies and the Kavalan malts from non-wine casks contained, with few exceptions, less than 0.3 g L^{-1} of total sugar (95th percentile 0.25 g L^{-1}, maximum 1.02 g L^{-1} in Redbreast 27 Yr), at levels consistent with the release of cell-wall sugars from oak [3,18]; glucose dominated, fructose was lower, and sucrose was absent or present only as small doublets that may arise from oak starch fragments. Jameson Blender's Dog, whose normalized glucose signal was flagged in version 1.0 as far above the Irish median, contained only 14.4 mg L^{-1} of glucose, with fructose below its 28.8 mg L^{-1} reporting level, on the absolute scale; its apparent excess in version 1.0 was an artifact of normalizing a low-extract residue to constant area, which the absolute data overturn.

The third group consists of whiskeys that contain sucrose at gram-per-liter levels. Litchfield Distilling Maple Bourbon (29.2 g L^{-1}), Barley Creek Honey Whiskey (26.9 g L^{-1}) and Taconic Double Barrel Maple Bourbon (11.8 g L^{-1}) are flavored products, and their labels describe them as such; under United States rules they belong to the flavored-whiskey class or to distilled spirits specialties rather than to bourbon or straight whiskey, and the sucrose they contain is that of the flavoring. Hudson Short Stack, a straight rye finished in maple syrup barrels, contained 3.7 g L^{-1} of sucrose that can only have come from the barrel. High West Double Rye!, a blend of straight rye whiskeys, contained 1.2 g L^{-1} of sucrose, confirmed by both of its resolved doublets; a blend of straight whiskies, unlike a straight whiskey, may contain harmless blending materials such as sugar under United States rules [23], so this finding identifies a production practice rather than a departure from those rules. Whether the transfer of several grams per liter of sucrose from a seasoned barrel is compatible with the designation of a straight whiskey is a regulatory question rather than an analytical one, but the measurement shows that the answer is not trivial: the barrel added more sugar to this whiskey than a light sweetening would.

The fourth group consists of products whose labels make a claim about carbohydrate. Mount Gay 1703 XO, labeled with 0 g of carbohydrate per 44 mL serving, contained 0.53 g L^{-1} of total sugar (293 mg L^{-1} of glucose and 235 mg L^{-1} of fructose), that is about 23 mg per serving, which United States labeling practice rounds to 0 g; the claim and the measurement are consistent, and the sugar found is within the range of the unsweetened oak-aged spirits. The Mount Gay XO Triple Cask Blend, which carries no such claim, contained a similar amount (0.63 g L^{-1}).

10.4 Flavor preparations

Vanillin without syringaldehyde and 1,2-propanediol above about 100 mg L^{-1} identify flavor preparations (Sections 6.3 and 6.5), and Table 17 lists the products in which either was found. TX American Blended Whiskey, a blend for which harmless flavoring within 2.5 % by volume is permitted in the United States, contained 819 mg L^{-1} of 1,2-propanediol and vanillin without syringaldehyde, the composition of a vanilla flavoring dispensed in propylene glycol. Roatan rum combined vanillin (105 mg L^{-1} by the formyl signal), 1,2-propanediol (1179 mg L^{-1}) and sucrose (19.2 g L^{-1}), and Rum 28 contained 76.7 mg L^{-1} of vanillin with neither syringaldehyde nor a glycol. Pendleton 1910 12 yr, a Canadian whisky, contained 128 mg L^{-1} of 1,2-propanediol together with high HMF and furanic-envelope signals, consistent with flavoring and caramel, both of which are permitted in Canadian whisky. Kirk & Sweeney 12 yr rum contained 660 mg L^{-1} of the glycol without the vanillin pattern. These observations identify a production practice; whether any of them is a departure from the rules depends on the category under which the product is sold and on its labeling, which this study does not assess.

Table 17. Products with vanillin lacking a syringyl partner or with more than 100 mg L^{-1} of 1,2-propanediol (mg L^{-1} of spirit; sucrose in g L^{-1}).

Product	Class	Benzaldehyde CHO	Syringyl index	1,2-Propanediol	Sucrose	Observation
Roatan Rum Company - Nicaragua	RUM	105	nd	1179	19.25	vanillin ring system, no syringyl partner; glycol carrier; sucrose
TX - American Blended Whisky	AMW	13.9	2.5	819	nd	vanillin ring system, no syringyl partner; glycol carrier
Kirk & Sweeney - 12 Yr Rum - 40%ABV	RUM	15.2	nd	660	0.31	glycol carrier
Taconic Distillery - Double Barrel Maple Bourbon	AMW	8.5	9.1	324	11.79	glycol carrier; sucrose
Barley Creek Distillery - Honey Whiskey 35%ABV	AMW	11.7	4.5	303	26.87	glycol carrier; sucrose
Litchfield Distilling - Maple Bourbon 43%ABV	AMW	12.4	2.8	190	29.20	glycol carrier; sucrose
Pendleton 1910 - 12 Year Old Canadian Rye Whisk…	CAN	2.5	nd	128	nd	glycol carrier; HMF/caramel
Rum 28 - 60%ABV	RUM	76.7	0.94	0.40	nd	vanillin ring system, no syringyl partner

10.5 Oak markers

The oak-derived congeners showed the gradient that the cask chemistry predicts (Figure 34). In the American whiskeys, matured in new charred oak, quercitol (median 88.2 mg L^{-1}), the benzaldehydes (16.1 mg L^{-1} in D_2O), gallic acid (6.2 mg L^{-1}) and the cinnamaldehydes (5.9 mg L^{-1}) were all highest, and syringaldehyde and vanillin in the $CDCl_3$ residues reached medians of 10.3 and 4.0 mg L^{-1}. The detection limits achieved for the aldehydes at 300 MHz (median 1.6 µmol L^{-1} of spirit for the benzaldehyde formyl signal) are comparable to the 1–5 µmol L^{-1} reported for the maturation congeners of native whisky at 600 MHz with a cryogenic probe, where sucrose and maltose could not be detected at all [7]; freeze-drying gains this sensitivity by removing the solvent signals that set the dynamic range of the native measurement, rather than by concentrating the sample. The Kavalan malts, matured in a subtropical climate in which extraction is fast, approached the American whiskeys in quercitol and aldehydes despite their largely ex-bourbon and wine-seasoned casks. The Scotch and Irish whiskies, matured mostly in refill casks, contained roughly half as much of each marker, and the brandies, Canadian whiskies, rums and agave spirits contained less still, with the exception of a few long-aged products such as the Tromba and Alquimia extra añejos and El Jimador Añejo, whose quercitol reached 52.3–64.4 mg L^{-1}. Because quercitol is extracted from wood and is not affected by toasting, while the aldehydes are formed by it, the ratio between the two is a potential index of the toasting and charring regime that deserves study on products of documented cask history.

10.6 Wine-cask carry-over and the agave spirits

Glycerol, succinic acid and lactic acid together trace the wine and sherry seasoning of a cask (Table 18). In the Kavalan malts matured in fino, manzanilla, amontillado, oloroso, Moscatel, port and wine casks, glycerol reached 589 mg L^{-1} and succinic acid 72.6 mg L^{-1}, against medians of 21.3 and 2.1 mg L^{-1} in the four Kavalan ex-bourbon malts. The Moscatel and port casks also transferred glucose and fructose in equimolar proportion (the Solist Moscatel residue contained 5.4 g L^{-1} of total sugar), as a sweet fortified wine would. The Hudson port and sherry finishes showed the same carry-over on a smaller scale. The combination of glycerol with succinic acid is the useful diagnostic: glycerol alone could have been added, but succinic acid is not a permitted additive in any of these categories and has no source other than fermentation, so its presence alongside glycerol in the expected proportion supports a cask origin.

Table 18. Wine-cask carry-over: glycerol, succinic and lactic acid (mg L^{-1}) and total sugar (g L^{-1}) in residues with wine, sherry, port or Madeira cask histories, with the Kavalan ex-bourbon malts for comparison (last rows).

Product	Class	Cask	Glycerol	Succinic	Lactic	Sugar
Solist Madeira Single Cask Strength - SMW	TWW	madeira	347	72.6	21.2	0.48
Kavalan - Solist Fino Sherry Single Cask - SMW	TWW	sherry (fino)	583	68.2	57.2	0.24
Oloroso Sherry Oak - SMW	TWW	sherry (oloroso)	431	53.6	34.7	0.20
Kavalan - Solist Amontillado Sherry SCS SMW	TWW	sherry (amontillado)	589	51.9	41.6	0.33
Solist Port - Single Cask Strength - SMW	TWW	port	404	49.7	42.9	1.22
Solist Moscatel Sherry - Single Cask Strength - S…	TWW	sherry (moscatel)	337	48.2	20.9	5.36
Kavalan - King Car - 40th Anniversary Selected WI…	TWW	wine	340	41.6	17.7	0.10
Kavalan Sky Gold Wine Cask Matured Single Cask Si…	TWW	wine	219	39.2	8.0	0.19
Kavalan - Solist Oloroso Sherry Single Cask Stren…	TWW	sherry (oloroso)	102	25.2	2.1	0.03
Kavalan Earth Silver Wine Cask Matured Single Cas…	TWW	wine	163	23.1	6.9	0.16
Solist Manzanilla Sherry - Single Cask Strength…	TWW	sherry (manzanilla)	211	20.5	10.9	0.15
Kavalan Solist Vinho Barrique Single Cask Strengt…	TWW	wine (STR barrique)	114	13.9	5.5	0.16
Kavalan - Concertmaster Sherry Cask Finish SMW	TWW	sherry	51.9	2.8	1.7	0.11
Hudson Whiskey NY - Port Cask Finish - NY RW - Ag…	AMW	port	302	37.5	30.2	2.76
Starward - Two Fold - 40%ABV Aged 2 Yrs in Austra…	OWW	wine	209	17.7	18.9	0.07
Redbreast Lustau Edition - Single Pot Still Irish…	IRW	sherry (Lustau)	314	16.3	16.7	0.12
Knappogue Castle 16 Yr -Sherry Cask Finished SMIW…	IRW	sherry	273	13.3	10.5	0.07
Hudson Whiskey NY - Sherry Cask Finish - NY BW…	AMW	sherry	64.4	9.9	13.0	0.25
Glenmorangie - The Lasanta - 43%ABV - 12 Year - B…	SCW	sherry	119	5.5	8.2	0.53
Solist Ex-Bourbon - Single Cask Strength - SMW	TWW	ex-bourbon	(26.9)	3.6	nd	0.11
Ex-Bourbon Oak SMW	TWW	ex-bourbon	(24.9)	2.0	nd	0.07
Kavalan Solist Ex-Bourbon Cask 56.3%ABV	TWW	ex-bourbon	(17.6)	2.2	nd	0.08
Kavalan Solist Ex-Bourbon Cask SMW 57.1%ABV 2012	TWW	ex-bourbon	(17.0)	0.76	4.9	0.07

Glycerol values in parentheses are upper bounds (crowded window); nq, not quantifiable beside sugars or 1,2-propanediol; <x, upper limit from the H-1b/H-3b multiplet.

The agave spirits were distinguished by lactic acid and by low oak extraction, and three of them by compositions that point to *abocado*. Tromba Reposado, labeled as matured in French oak, contained 3.54 g L^{-1} of fructose with little glucose (333 mg L^{-1}, a fructose-to-glucose ratio of about 11; version 2.0, whose fits had missed the displaced β-glucose doublet of this residue, reported 44 mg L^{-1}), the second-highest HMF of the data set (42.4 mg L^{-1}, after a 16-year-old Trinidad rum), and low quercitol (8.8 mg L^{-1}) and benzaldehydes (4.4 mg L^{-1}). Fructose cannot distill; in the oak-aged whiskeys of this study without additions it was usually below its 28.8 mg L^{-1} reporting level, exceeded 106 mg L^{-1} in only a tenth of the residues and was accompanied by more glucose (median fructose-to-glucose ratio 0.72); and a fructose-dominated composition is that of a hydrolyzed agave syrup or another fructose syrup, while HMF at this level is consistent with a heated syrup or with caramel. The residue therefore has the composition of an *abocado* with a sugar-based syrup, which is permitted in reposado tequila up to 1 % of the weight of the tequila before bottling [28]; 3.54 g L^{-1} of fructose corresponds to about 0.37 % of the weight of a 40 % ABV tequila as dry sugar, or about 0.53 % as a syrup of 70 % solids, well within that limit. Version 1.0 of this paper, and the first pass of the present one, also attributed gram-per-liter glycerol to this residue. That attribution is withdrawn: the signal was the β-pyranose H-1b doublet of the fructose itself (Section 6.5), the second glycerol multiplet is absent, and glycerol is below about 62 mg L^{-1}. The producer describes the brand as additive-free, so the observation bears directly on a marketing claim; but it remains a single bottle of undocumented lot, the replicate results of Section 10.2 show why that matters, and replicate bottles, a recovery standard and an orthogonal sugar analysis (HPLC with refractive-index or pulsed amperometric detection, or an enzymatic fructose and glucose assay) are required before any statement about the product can be made. Gran Centenario Añejo contained 0.9 g L^{-1} of glucose, 1.1 g L^{-1} of fructose and 0.9 g L^{-1} of

glycerol confirmed by both of its multiplets, and Legenda Extra Añejo 0.9 g L^{-1} of glucose, 1.2 g L^{-1} of fructose and 0.43 g L^{-1} of glycerol, compositions consistent with a permitted *abocado* with syrup and glycerin, whereas Tromba Añejo and Extra Añejo, the two Clase Azul tequilas, the two Alquimia extra añejos, Lunazul and El Jimador Añejo, the mezcal and the sotols contained less than 0.22 g L^{-1} of sugar and at most 36.1 mg L^{-1} of glycerol.

10.7 Products with cask-seasoning histories that leave a trace

Several world whiskies matured or finished in casks that had held other spirits showed the chemistry of the previous fill. Coperies, a French malt from a cognac house, contained 1120 mg L^{-1} of glucose and 2196 mg L^{-1} of fructose (glycerol could not be confirmed beside them), and Coperies, Brenne Estate Cask and Brenne 10 Year, the last two finished in ex-cognac casks, carried the highest furanic-envelope signals of the world malts (71.5–135 µmol L^{-1}, against at most 18.1 µmol L^{-1} in the others). The furanic envelope is consistent with carry-over from casks that held a caramel-adjusted brandy, but the fructose excess of Coperies (fructose-to-glucose ratio 2.0) is not the equimolar sugar of a sweetened cognac, and its origin remains open. Forty Creek Barrel Select, a Canadian whisky, contained 2.2 g L^{-1} of glucose and fructose in equimolar proportion together with succinic (10.7 mg L^{-1}) and lactic acid (39.2 mg L^{-1}), consistent with the wine or sherry that Canadian practice permits as a blending material. These are the observations that a carry-over index based on absolute quantities makes possible, and they show that cask history and permitted additions can produce the same compounds; it is the combination and the proportion of the compounds, not the presence of any one, that carries diagnostic weight.

11. Chemometrics and machine learning

11.1 What absolute intensities change

In version 1.0 every spectrum was normalized to constant area and then by probabilistic quotient normalization, because nothing else was possible with traces recovered from graphics. The consequence was that a residue weighing a few milligrams and one weighing a hundred milligrams contributed equally to every model, and a small sugar signal in a low-extract whiskey could appear as large as the sugar of a sweetened rum. The models of this version were computed on the absolute content of each 0.02 ppm bin, in micromoles of ^{1}H per liter of spirit, with Pareto scaling and no normalization, so that a residue that carries more of a constituent in fact carries more weight in that variable. This choice has a price. The variance of the absolute data is dominated by the few residues that contain grams of sugar per liter, and in the full D_2O block the first principal component explains 70.3 % of the variance and is essentially a sugar axis. Models were therefore built both on the full block and on the non-carbohydrate block from which the windows δ 3.10–4.42 and 5.06–5.50 were removed (37.9 % of the variance on the first component), and the comparison between the two is itself informative. This section reports the analyses that carry the conclusions of the paper; Section 12 extends them to every data approach computed in the course of the work, including the $CDCl_3$, fused and DMSO-d_6 blocks and the quantified congeners, each shown with its principal components, four hierarchical clusterings and both non-linear embeddings.

11.2 Principal component analysis

The principal component scores and loadings of the D_2O/CD_3OD residues are shown in Figure 35. In the full block the first component (70.3 %) is a measure of the total extract and, in practice, of sugar: its loadings are those of the carbohydrate CH–O envelope, its scores correlate perfectly with the integrated proton inventory (Spearman ρ = 1.00), and it runs from the sotols and the low-extract agave spirits to Barley Creek Honey Whiskey, Roatan rum, Kiss Black Diamond and Planteray Isle of Fiji. The second component (14.2 %) separates sucrose from invert sugar, with positive loadings on the sucrose Glc H-1 and Fru H-3 doublets and negative loadings on the fructose and glucose signals, placing the maple and honey whiskeys and the sucrose-sweetened rums on one side and Bacardi Reserva Limitada, Pyrat XO and the cognacs on the other. The third (4.6 %) is the first that reflects cask chemistry: it is loaded on the ethyl-ester and lactic acid region near δ 1.2–1.4 and on glycerol, and its highest scores belong to the Kavalan port, fino, Moscatel and amontillado casks.

Because the carbohydrate and the retained-volatile signals dominate the absolute variance, the structure of the non-volatile, non-carbohydrate composition was examined in a block from which the carbohydrate windows and also the windows of retained ethanol and the fusel alcohols were removed (404 variables). Its first component (42.3 %) again scales with the total extract (ρ = 0.95 with the integrated inventory) but is loaded on the broad aromatic envelope of oligomeric phenolics, the succinic acid singlet and the aliphatic envelope at δ 1.2–1.35 that accompanies wine-cask carry-over, and its highest scores belong to the wine-seasoned Kavalan malts. The second component (21.5 %) is the most interpretable axis of the data set: its positive loadings are the 1,2-propanediol doublet and the vanillin formyl and ring signals, its negative loadings the quercitol multiplets and the syringyl singlet, and it runs from the products that contain 1,2-propanediol, the usual carrier of flavor preparations (Roatan rum, TX American Blended, Kirk & Sweeney, Pendleton 1910 and Barley Creek), to the bourbons and ryes matured in new charred oak (Davidson Reserve, Russell's Reserve 6 yr, Bespoken Rye and the Hudson whiskeys). In other words, the second most important direction of variation in the non-carbohydrate composition of these spirits is the contrast between a flavor preparation and oak. The $CDCl_3$ block (51.9 % and 13.9 % on the first two components) is organized

by lipid content, which separates the cask-strength Kavalan malts and the barrel-proof American whiskeys from the chill-filtered products, and by the methoxyl and aldehyde signals.

None of the principal component scores of the volatile-free block correlated appreciably with the repetition time of the acquisition (|Spearman ρ| ≤ 0.15 for the first five components), so the partial saturation of long-T_1 signals at short repetition times does not structure the score space. The correlation with the acquisition set was stronger (up to 0.40), but the two sets differ in their product composition (the second contains most of the rums and Scotch malts), and the test of Section 11.5 was used to separate composition from batch.

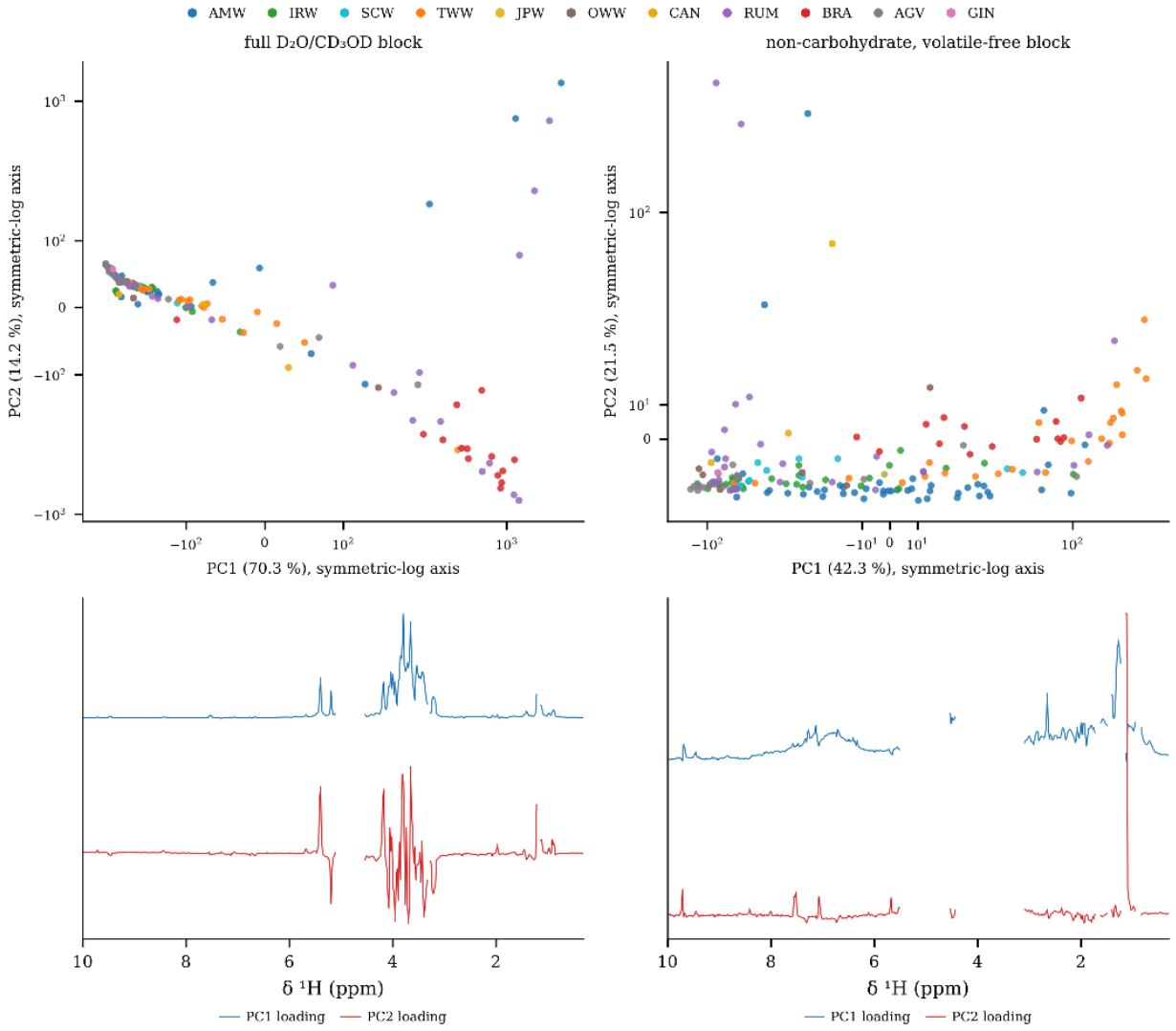

Figure 35. Principal component analysis of the Pareto-scaled absolute D_2O/CD_3OD spectra: scores of the first two components on symmetric-logarithmic axes (top) and the corresponding loadings (bottom), for the full block (left) and for the non-carbohydrate, volatile-free block (right). Category abbreviations as in Figure 34.

11.3 Hierarchical cluster analysis

Ward clustering of the non-carbohydrate block reproduced the main groups of the principal component space (Figure 36), but agreement of a seven-cluster partition with the spirit category was modest (adjusted Rand index 0.29; 0.08 for the full block), because the clusters follow cask chemistry and additions rather than geography. The most strongly wine-seasoned Kavalan malts formed a cluster of their own; the cognacs

shared a cluster with the remaining Kavalan malts, the Hudson port and sherry finishes, aged rums such as Flor de Caña 12 and 18 years and Appleton Estate 21 years, and both Tromba tequilas; the three products with the highest 1,2-propanediol contents (Roatan, TX American Blended and Kirk & Sweeney) and the two maple and honey whiskeys formed small clusters of their own; most American whiskeys fell into a single cluster; and the low-extract Irish, Scotch, agave and rum residues fell together with the gin into the largest cluster. The cophenetic correlation was 0.67 for Ward linkage and 0.94 for average linkage, so the average-linkage dendrogram represents the distances more faithfully, at the cost of a less balanced tree. HCA is used here, as in version 1.0, as a description of similarity rather than as a classifier.

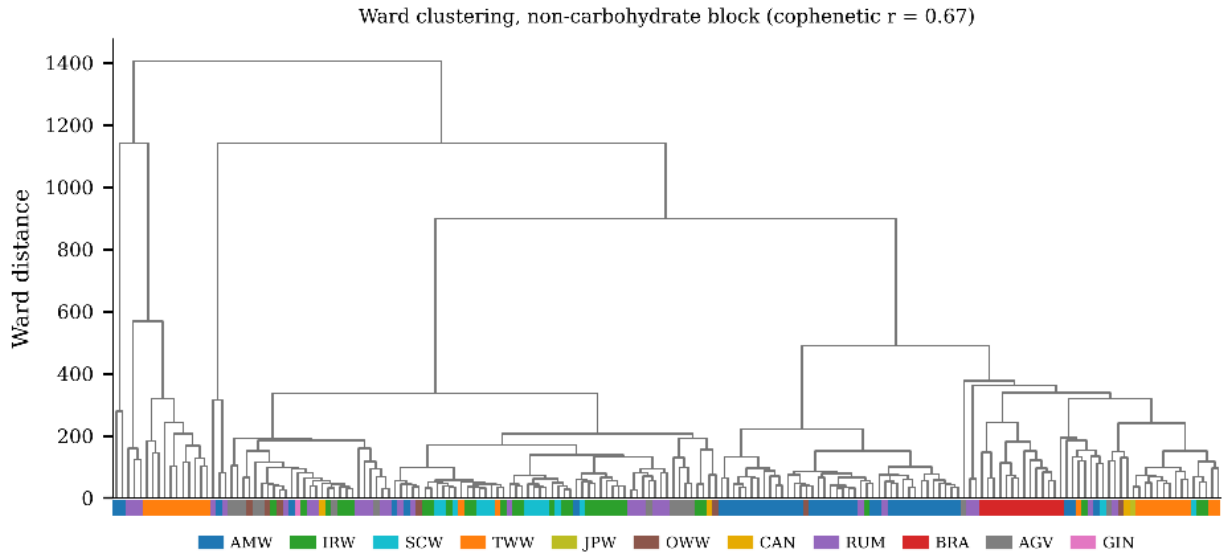

Figure 36. Ward clustering of the Pareto-scaled non-carbohydrate block (Euclidean distance); the color bar beneath the dendrogram gives the spirit category of each residue.

11.4 Supervised classification

The seven categories with at least fourteen D_2O/CD_3OD residues (171 residues of 167 products) were classified by PLS-DA, linear and radial-basis-function SVM and *k*-nearest neighbors, with nested, product-grouped cross-validation (Table 19). On the non-carbohydrate block the linear SVM reached a balanced accuracy of 0.78 ± 0.01 (mean ± standard deviation over three repeats of the outer cross-validation), the RBF SVM 0.78, PLS-DA 0.71 and *k*-nearest neighbors 0.72, against a chance level of 0.14. The full block gave slightly lower accuracies for every method (0.74 for the linear SVM), which shows that the dominance of a few sugar-rich residues in the absolute variance does not help classification, and binning at 0.04 ppm raised the linear SVM to 0.80, consistent with a small benefit from reduced sensitivity to residual misalignment. Including the fourteen pure-D_2O residues did not degrade the result (0.78), and removing the retained-volatile windows as well as the carbohydrate windows left it unchanged (0.79 for the linear SVM, 0.68 for PLS-DA), which shows that the classification does not rest on the ethanol and fusel-alcohol artifacts of freeze-drying. The RBF SVM selected the largest regularization constant of its search grid together with a very small kernel coefficient γ, that is, a kernel much wider than the distances between residues, which is the regime in which an RBF machine approximates a linear one, and PLS-DA selected 12–15 latent variables; both observations point to a classification problem that is close to linear but needs many directions.

Table 19. Balanced accuracies of the classifiers (seven categories, 171 D_2O/CD_3OD residues; mean ± standard deviation over three repeats of the outer ten-fold cross-validation where repeated).

Data block	Validation	PLS-DA	SVM, linear	SVM, RBF	*k*-NN
Full D_2O block (502 bins)	nested, product-grouped	0.68 ± 0.01	0.74 ± 0.03	0.74 ± 0.02	0.65 ± 0.01
Non-carbohydrate block	nested, product-grouped	0.71 ± 0.01	0.78 ± 0.01	0.78 ± 0.02	0.72 ± 0.02
Non-carbohydrate, 0.04 ppm bins	nested, product-grouped	0.67	0.80	0.79	0.69
Non-carbohydrate, including pure-D_2O residues	nested, product-grouped	0.70	0.78	0.78	0.71
Non-carbohydrate, volatile-free (404 bins)	nested, product-grouped	0.68 ± 0.02	0.79 ± 0.01	—	—
Quantified congeners (18)	nested, product-grouped	0.59 ± 0.02	0.68 ± 0.02	0.69 ± 0.02	0.64 ± 0.02
Full D_2O block	leave-producer-out (6 classes)	0.53	0.61	0.61	0.42
Non-carbohydrate block	leave-producer-out (6 classes)	0.59	0.67	0.64	0.57
Non-carbohydrate, volatile-free	leave-producer-out (6 classes)	—	0.61	—	—
Non-carbohydrate block	trained on set 2, tested on set 1	0.62	0.64	0.64	0.41
Non-carbohydrate block	trained on set 1, tested on set 2	0.49	0.57	0.52	0.40
Non-carbohydrate block	200 label permutations: null mean / maximum	0.14 / 0.21	0.14 / 0.22	—	—
Paired residues, AMW vs other: D_2O	nested, product-grouped	0.90	0.89	—	—
Paired residues: $CDCl_3$	nested, product-grouped	0.79	0.77	—	—
Paired residues: fused	nested, product-grouped	0.84	0.80	—	—

Chance: 0.14 for seven classes, 0.17 for six, 0.50 for the two-class paired problem. Leave-producer-out excludes the single-producer Taiwanese class.

The confusion matrix of the best model (Figure 37) shows where the classes are separable and where they are not. Brandy (14/14), Taiwanese whisky (23/25) and American whiskey (41/46) had the highest recalls; rum (20/28) and agave spirits (7/14) were recognized less reliably; and Scotch (10/14) and, less often, Irish (24/30) whiskies were confused with each other, rums with American whiskeys, and agave spirits with several classes. The pattern is chemically sensible. The brandies are defined by sugar and caramel, the American whiskeys by new-charred-oak extract, and the Kavalan malts by wine-cask carry-over and by their very high oak extraction in a subtropical climate, whereas Irish and Scotch whiskies are matured largely in refill and ex-bourbon casks and their extracts differ from one another mainly in degree.

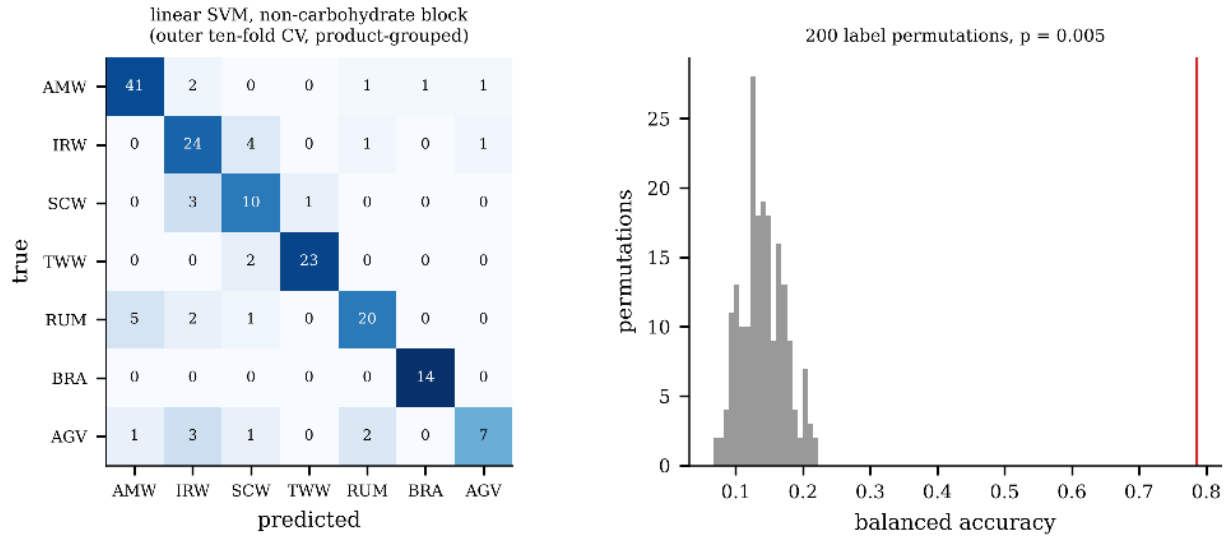

Figure 37. Classification of seven spirit categories by linear SVM on the non-carbohydrate block: confusion matrix of the outer, product-grouped cross-validation (left) and the distribution of balanced accuracies obtained after 200 permutations of the class labels, with the observed value (red line; right).

11.5 How much of the classification is real?

Three tests were used to establish what the cross-validated accuracies mean, because an expert reviewer would be right to ask whether a classifier has learned the category, the producer or the acquisition batch.

The label-permutation tests answer the first question. With the class labels permuted 200 times, the cross-validated balanced accuracy of the linear SVM on the non-carbohydrate block had a mean of 0.14 and a maximum of 0.22, against an observed 0.79 (p = 0.005, the smallest value attainable with 200 permutations); PLS-DA behaved in the same way (0.70 observed, null maximum 0.21). The classification is therefore not an artifact of the flexibility of the models.

The leave-producer-out validation answers the second. When all residues of a producer were held out together, the balanced accuracy over the six multi-producer categories fell to 0.67 for the linear SVM on the non-carbohydrate block (0.59 for PLS-DA, 0.57 for *k*-nearest neighbors), against 0.17 by chance. Part of the accuracy of the product-grouped validation therefore comes from recognizing the house style of producers that are represented by several products: Kavalan (27 residues), Irish Distillers (22), Tuthilltown (14) and Flor de Caña (6) dominate their categories. This is not a defect of the models but a property of the sample set, and it means that the figure of about 0.78 applies to producers that are represented in the training set, while about 0.67 is the better estimate for an unknown producer. For the Taiwanese category, which consists of a single distillery, no producer-independent estimate is possible at all, and its high recall is a recall of Kavalan, not of Taiwan.

The cross-batch validation and the batch-prediction test answer the third. Models trained on the second acquisition set and applied to the first reached a balanced accuracy of 0.64 (linear SVM, non-carbohydrate block), and the reverse 0.57, over the categories present in both sets. Within the four largest categories, PLS-DA predicted the acquisition set from the spectra with a balanced accuracy of 0.78. A value well above 0.5 shows that the two sets differ systematically even within a category; this may reflect a batch effect of acquisition or preparation (the second set was acquired with longer relaxation delays and a more uniform solvent system), a difference in the products sampled within each category, or both, and the two cannot be separated without replicate preparations of the same bottles in both sets. It is the most important limitation of the classification results, and it is the reason why no classification model of this paper is proposed for use on new samples without an independent validation set acquired under controlled conditions.

11.6 Classification from quantified congeners

A model built on eighteen quantified D_2O congeners, log-transformed and autoscaled, is far easier to interpret than one built on 500 bins, and it reached a balanced accuracy of 0.68 (linear SVM), 0.69 (RBF SVM), 0.59 (PLS-DA) and 0.64 (*k*-nearest neighbors) under the same validation (Table 19). Glycerol was left out of this set, because it could not be quantified in the sugar-rich residues and any imputed value would have carried information about the sugar rather than about glycerol. The loss relative to the spectral models (0.08–0.12 in balanced accuracy, depending on the method) is modest, which shows that most of the class information resides in the identified compounds; the remainder is carried by the broad phenolic and carbohydrate envelopes and by unassigned minor signals. The coefficients of the linear SVM identify the congeners that carry the separation: succinic acid, and to a lesser extent the fusel alcohols, for the Kavalan malts; lactic acid and low quercitol for the agave spirits; fructose, the furanic envelope and lactic acid for the brandies; the cinnamaldehydes and the syringyl singlet for the American whiskeys; quercitol and gallic acid for the Irish whiskeys; and low quercitol and low fusel alcohols with relatively high benzaldehydes for the rums.

11.7 Non-linear embeddings

The t-SNE and UMAP embeddings (Figure 38) reproduce the neighborhoods of the Pareto-scaled space faithfully (trustworthiness 0.93 and 0.90 for the non-carbohydrate block, averaged over five seeds) and give visually clearer groupings than the first two principal components, whose trustworthiness was 0.85. They do not, however, contain more class information than the space from which they were computed: the fraction of each residue's ten nearest neighbors that share its category was 0.50 in the full scaled space, 0.50 in the t-SNE embedding and 0.53 in the UMAP embedding. The embeddings are therefore displays of the structure that the linear methods already capture, which is the same conclusion as in version 1.0, now reached on absolute data and with a larger sample set. Apparent clusters in an embedding should not be interpreted without reference to the scaled space and to the chemistry.

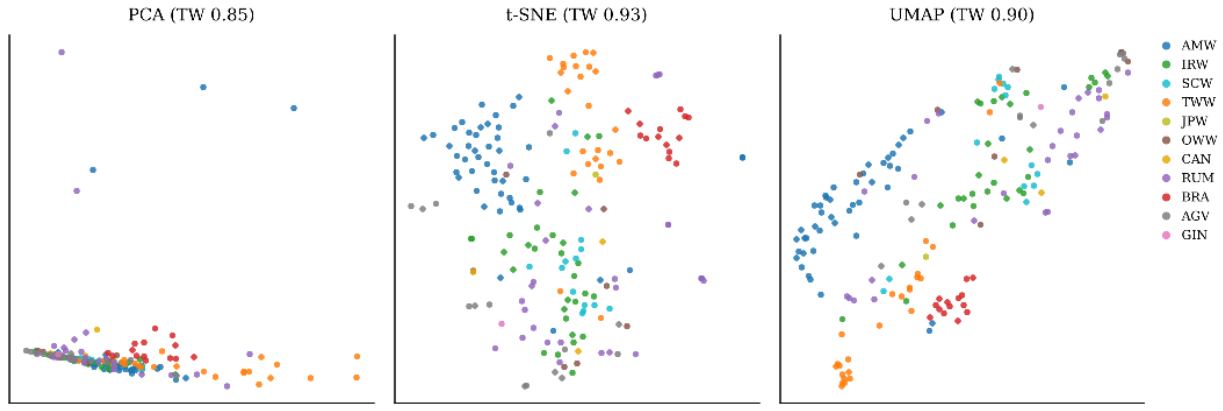

Figure 38. Two-dimensional displays of the non-carbohydrate block by PCA, t-SNE and UMAP, with the trustworthiness (TW) of each display.

11.8 The $CDCl_3$ block and data fusion

The $CDCl_3$ residues are too few for a seven-class model, and a two-class problem was posed instead: American whiskey against all other products, for the 51 products with both a $CDCl_3$ and a D_2O/CD_3OD residue. The D_2O block alone reached a balanced accuracy of 0.89 (linear SVM; 0.90 for PLS-DA), the $CDCl_3$ block 0.77 (0.79) and the fused block 0.80 (0.84). The lipophilic fraction thus carries substantial independent information about American whiskeys, in its aldehyde ratios and lipid profile, but adding it to the polar fraction did not improve the classification beyond the better of the two blocks; with so few residues, fusion mainly adds variables. The block contributions of the fused principal components are shown in the companion workbook.

12. A full accounting of the chemometric data approaches

Section 11 reported the chemometric results that carry the conclusions of this paper: principal component analysis of two D_2O/CD_3OD blocks, a single Ward dendrogram, supervised classification and the non-linear embeddings of one block. Many more data approaches were computed in the course of the work, and a reader who is offered one dendrogram is entitled to ask how the others looked. This section shows all of them. Every data approach, seventeen blocks in all (Table 20), was analyzed in exactly the same way (Section 4.16), and each is presented in a full-page figure of identical layout: the scores of the first three principal components with the most extreme residues labeled, the scree plot, the loadings of the first three components, the t-SNE and UMAP embeddings, a legend panel that also records the number of residues and variables, the scaling, the trustworthiness of the two-dimensional PCA display and the class purity of the full scaled space, and four dendrograms (Ward linkage, average linkage and complete linkage on Euclidean distances, and average linkage on correlation distances) with the class of each residue marked in a strip beneath the leaves. Because the residue names cannot be read in these one-page composites, every score plot, embedding and dendrogram is also reproduced at full page, with every residue named, in the Addendum at the end of this paper (Figures S1–S94); the composites are retained here so that the approaches can be compared at a glance, and the caption of each gives the Addendum figures of its block. The statistics of all blocks are collected in Table 21 and Table 22 and compared in Figure 56, and the products with residues in more than one solvent are used in Sections 12.9 and 12.10 to compare one solvent with two.

Table 20. Data approaches analyzed in the full chemometric accounting (every block mean-centered).

Block	Description	Residues	Variables	Scaling	Figure
D2O-F	D_2O/CD_3OD, full spectrum, 0.02 ppm	183	502	Pareto	Figure 39
D2O-NC	D_2O/CD_3OD, carbohydrate windows removed	183	418	Pareto	Figure 40
D2O-NV	D_2O/CD_3OD, carbohydrate and retained-volatile windows removed	183	404	Pareto	Figure 41
D2O-NC04	as D2O-NC, 0.04 ppm bins	183	209	Pareto	Figure 42
D2O-NCall	as D2O-NC, including the residues in D_2O alone	197	418	Pareto	Figure 43
D2O-Qp	D_2O/CD_3OD quantified congeners, log, Pareto	183	19	Pareto	Figure 44
D2O-Qa	D_2O/CD_3OD quantified congeners, log, autoscaled	183	19	autoscaling	Figure 45
CDCl3	$CDCl_3$, full spectrum, 0.02 ppm	61	517	Pareto	Figure 46
CDCl3-W	$CDCl_3$, water window widened to δ 1.45–1.76	61	509	Pareto	Figure 47
FUSED-DC	D2O-NC and CDCl3 fused, paired products	51	935	Pareto per block, RMS-normalized	Figure 48
FUSED-W	D2O-NC and CDCl3-W fused, paired products	51	927	Pareto per block, RMS-normalized	Figure 49
DMSO-F	DMSO-d_6, full spectrum (solvent and water regions removed), 0.02 ppm	29	493	Pareto	Figure 50
DMSO-Fx	as DMSO-F, without the maple-barrel residue	28	493	Pareto	Figure 51
DMSO-NC	DMSO-d_6, carbohydrate windows also removed	29	411	Pareto	Figure 52
DMSO-F04	as DMSO-F, 0.04 ppm bins	29	246	Pareto	Figure 53
DMSO-Qp	DMSO-d_6 quantified congeners, log, Pareto	29	20	Pareto	Figure 54
DMSO-Qa	DMSO-d_6 quantified congeners, log, autoscaled	29	20	autoscaling	Figure 55

12.1 What the measures mean

Four quantities are used throughout, and each answers a narrower question than its name suggests. The cophenetic correlation [67] is the correlation between the distances among residues and the heights at which the dendrogram joins them; a value near one means that the tree is a faithful picture of the distance matrix, not that its clusters are chemically meaningful or agree with any external classification. The adjusted Rand index [68] measures the agreement between a partition of the tree into k clusters and a set of class labels, corrected for the agreement expected by chance: it is zero for a random partition and one for perfect

agreement. It was computed against the seven spirit categories with at least fourteen residues for the D_2O/CD_3OD blocks ($k = 7$), against the five categories present in the $CDCl_3$ and fused blocks ($k = 5$), and against the five residue groups ($k = 5$) and the two cask types, wine-seasoned or other ($k = 2$), for the DMSO-d_6 blocks. The fraction of residues in the largest of the k clusters records chaining, the tendency of some linkages to absorb almost every residue into one cluster while each outlier forms a cluster of its own; a tree that chains can have an excellent cophenetic correlation and an adjusted Rand index of zero. Trustworthiness [49] asks whether the nearest neighbors of a residue in a two-dimensional display are also among its nearest neighbors in the full scaled space, penalizing false neighbors, and the k-nearest-neighbor class purity asks what fraction of each residue's k nearest neighbors ($k = 10$, or 9 for the small DMSO-d_6 blocks) share its class. The purity must be read against its chance level, the probability that two residues drawn at random belong to the same class, which depends on the class composition of each block and is given in Table 22; a purity of 0.50 means little in a block that is half American whiskey and a great deal in a block of eleven categories.

One property of the embeddings should be stated at the outset. t-SNE initialized from the principal component scores is deterministic for these data, so the five random seeds gave identical embeddings and identical statistics, whereas the UMAP embeddings varied from seed to seed by at most ±0.025 in trustworthiness. The single t-SNE and UMAP panels shown in each figure are therefore representative.

12.2 The full D_2O/CD_3OD block: an axis of sweetness

The full block (Figure 39) is dominated by carbohydrate. Its first component carries 70.3 % of the variance, its loadings are those of the CH–O envelope between δ 3.6 and 3.9 and of the anomeric doublets, and its scores, which span more than two orders of magnitude and are therefore drawn on symmetric-logarithmic axes, run from the sotols, Clase Azul Gold and Bozal Cuishe to Barley Creek Honey Whiskey, Roatan rum, Kiss Black Diamond, Planteray Isle of Fiji and Bacardi Reserva Limitada; their Spearman correlation with the total proton inventory of the block is 1.00. The second component (14.2 %) contrasts sucrose (the Glc H-1 doublet near δ 5.40 and the Fru H-3 doublet near δ 4.18) with glucose and fructose, placing the honey and maple whiskeys and the sucrose-sweetened rums against the invert-sugar products such as Bacardi Reserva Limitada and the cognacs, and the third (4.6 %) is the first to reflect cask chemistry, with positive loadings on the aliphatic envelope at δ 1.21–1.31 and on glycerol and its highest scores in the Kavalan port, fino, Moscatel and amontillado casks.

The four dendrograms of this block illustrate the central difficulty of hierarchical clustering on absolute data. Their cophenetic correlations were high (0.83 for Ward linkage, 0.97 for average linkage, 0.92 for complete linkage and 0.96 for the correlation distance), yet their agreement with the spirit categories was essentially nil (adjusted Rand index 0.08, −0.00, 0.05 and 0.06), because the largest of the seven clusters contained 74 %, 96 %, 89 % and 79 % of the residues. The average-linkage tree, whose cophenetic correlation is the highest of any block in this study, joins a handful of sugar-rich residues last and places almost everything else in one cluster: it represents the distance matrix faithfully, but the distance matrix is a measure of sugar. The embeddings tell the same story from another direction. They are trustworthy (t-SNE 0.93, UMAP 0.91) and their class purity (0.43 and 0.45) is that of the full scaled space (0.47), about three times the chance level of 0.15, but the neighborhoods they preserve are neighborhoods of sweetness. For authentication this block is useful in one respect only, as a screen for sweetening, and that is better done by quantifying the sugars directly (Section 6.4).

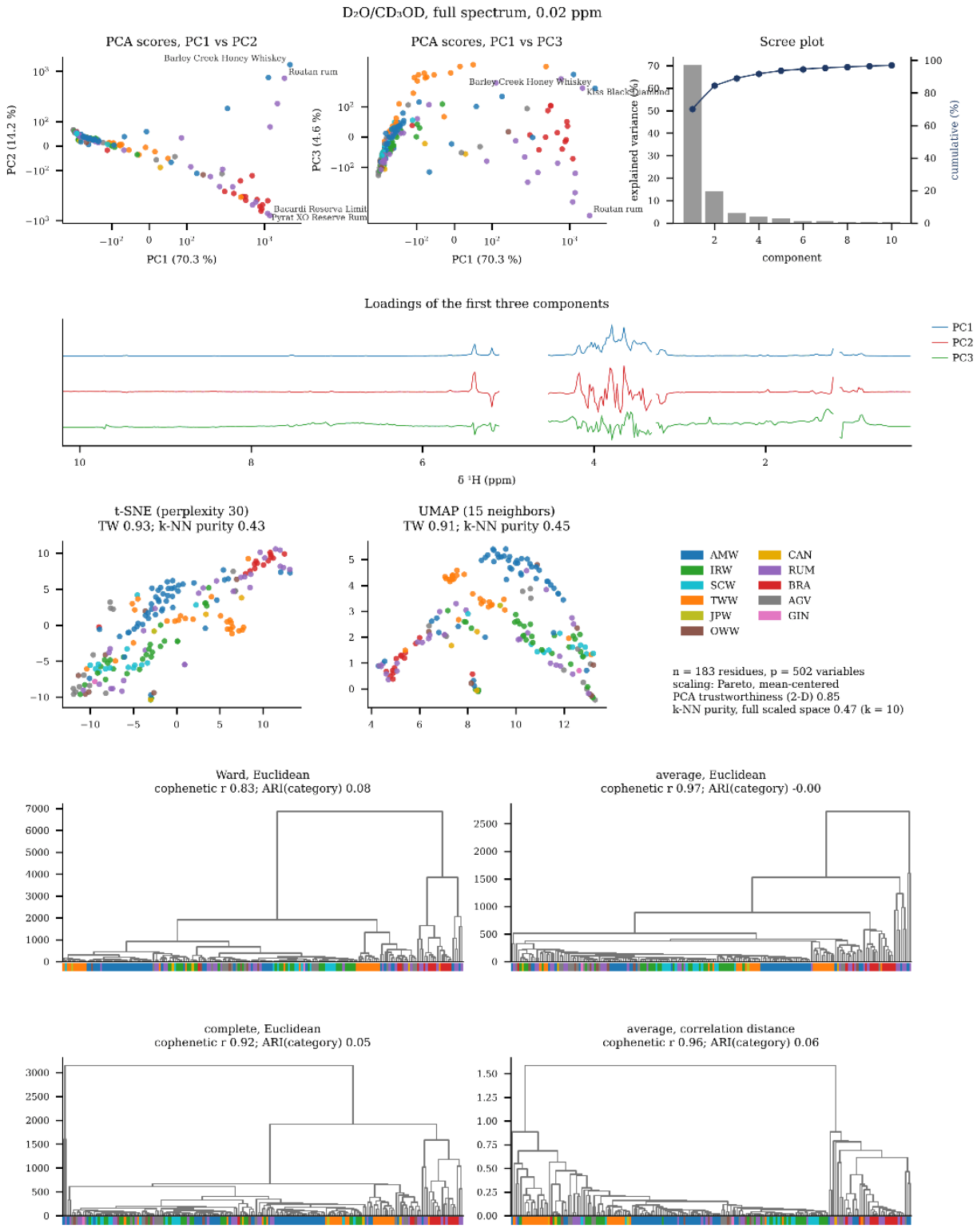
D₂O/CD₃OD, full spectrum, 0.02 ppm
PCA scores, PC1 vs PC2
Barley Creek Honey Whiskey
Roatan rum
Bacardi Reserva Limit
Pyrat XO Reserve Rum
PC2 (14.2 %)
PC1 (70.3 %)
PCA scores, PC1 vs PC3
Barley Creek Honey Whiskey
Kiss Black Diamond
Roatan rum
PC3 (4.6 %)
PC1 (70.3 %)
Scree plot
explained variance (%)
cumulative (%)
component
Loadings of the first three components
PC1
PC2
PC3
δ ¹H (ppm)
t-SNE (perplexity 30)
TW 0.93; k-NN purity 0.43
UMAP (15 neighbors)
TW 0.91; k-NN purity 0.45
AMW
IRW
SCW
TWW
JPW
OWW
CAN
RUM
BRA
AGV
GIN
n = 183 residues, p = 502 variables
scaling: Pareto, mean-centered
PCA trustworthiness (2-D) 0.85
k-NN purity, full scaled space 0.47 (k = 10)
Ward, Euclidean
cophenetic r 0.83; ARI(category) 0.08
average, Euclidean
cophenetic r 0.97; ARI(category) -0.00
complete, Euclidean
cophenetic r 0.92; ARI(category) 0.05
average, correlation distance
cophenetic r 0.96; ARI(category) 0.06

Figure 39. Full chemometric analysis of block D2O-F (D_2O/CD_3OD residues, full spectrum, 0.02 ppm bins, Pareto scaling; every block mean-centered). Top: PCA scores of PC1 against PC2 and PC3 with the most extreme residues labeled, and the scree plot with the cumulative explained variance. Second row: loadings of the first three components, offset vertically, with gaps at the excluded windows. Third row: t-SNE and UMAP embeddings with their trustworthiness (TW) and *k*-nearest-neighbor class purity, the class legend and the statistics of the block. Bottom: dendrograms by Ward, average and complete linkage on Euclidean distances and by average linkage on correlation distances, each with its cophenetic correlation and adjusted Rand index, and with the spirit category of each residue (abbreviations as in Figure 34) in the strip beneath the leaves. The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S1–S8).

12.3 The non-carbohydrate D_2O/CD_3OD blocks

Removing the carbohydrate windows (Figure 40) changes the picture completely. The first component of the non-carbohydrate block (37.9 %) is almost unipolar, with its largest loadings on the aliphatic envelope at δ 1.21–1.31 that accompanies wine-cask carry-over and on the methyl doublets of the retained fusel alcohols; its highest scores belong to the Kavalan Moscatel, port, fino and manzanilla casks and to Taconic Maple Bourbon, its lowest to the sotols, Powers Gold Label and Brenne 10 yr, and it is in effect a measure of non-carbohydrate extract. The second component (17.3 %) contrasts the retained ethanol triplet near δ 1.21, the 1,2-propanediol doublet at δ 1.11–1.13 and the fusel methyl doublets with the wine-cask envelope, the quercitol multiplets at δ 1.79 and 1.93 and the succinic acid singlet near δ 2.64, and it places Roatan rum, Barley Creek Honey Whiskey, Taconic Maple Bourbon, TX American Blended and Kirk & Sweeney 12 yr at one extreme and the Cadenhead rum blend, the Kavalan Madeira and Sky Gold casks and the Hudson sherry and port finishes at the other. The third component (15.9 %) separates the two kinds of outlier on that axis: the sugar-rich maple and honey whiskeys, which retained ethanol and the fusel alcohols in their sugar matrix (Section 6.7), from the products that carry the vanillin and 1,2-propanediol of a flavor preparation (Section 6.3). The labeled score plots make an important limitation visible. The second and third components are carried by five or six residues, while the great majority of the 183 residues form a dense cloud near the origin; Pareto scaling of absolute intensities gives the extreme residues the weight that their concentrations deserve, and the price is that the leading components describe the outliers rather than the bulk of the data.

The Ward tree of this block has the highest adjusted Rand index of the full-spectrum D_2O approaches (0.29), and its seven clusters are chemically legible. The honey and maple whiskeys formed a cluster of two, three dark or long-aged rums (Kiss Black Diamond, a 16-year-old Trinidad rum and the Cadenhead blend) a cluster of three, and the three products carrying a flavor preparation (TX American Blended, Kirk & Sweeney and Roatan) a cluster of three. Eleven wine-seasoned Kavalan malts formed a cluster of their own; 36 of the American whiskeys formed a cluster of 40; the 14 brandies shared a cluster of 43 with 12 Kavalan malts from other casks and a mixture of aged products; and the low-extract Irish, Scotch, agave and rum residues fell into a single cluster of 81. The correlation-distance tree, which compares the shapes of the scaled profiles irrespective of the amount of extract, reached almost the same agreement with the categories (0.26) with a far better cophenetic correlation (0.81 against 0.67), and it drew some boundaries that the Ward tree did not: the three aged tequilas with high lactic acid (Tromba Extra Añejo, Alquimia Reserva de Oro and El Jimador Añejo) formed a cluster, as did the two sherry-finished Irish whiskeys (Redbreast Lustau Edition and Knappogue Castle 16 yr), and the flavor-preparation cluster grew to five products by adding Pendleton 1910 and Coperies to the three above. In this tree 32 American whiskeys formed a cluster of 35, and the brandies and 23 Kavalan malts shared one of 49. Average linkage, again, chained (89 % of the residues in one cluster, adjusted Rand index 0.02), and complete linkage lay between the two (0.20).

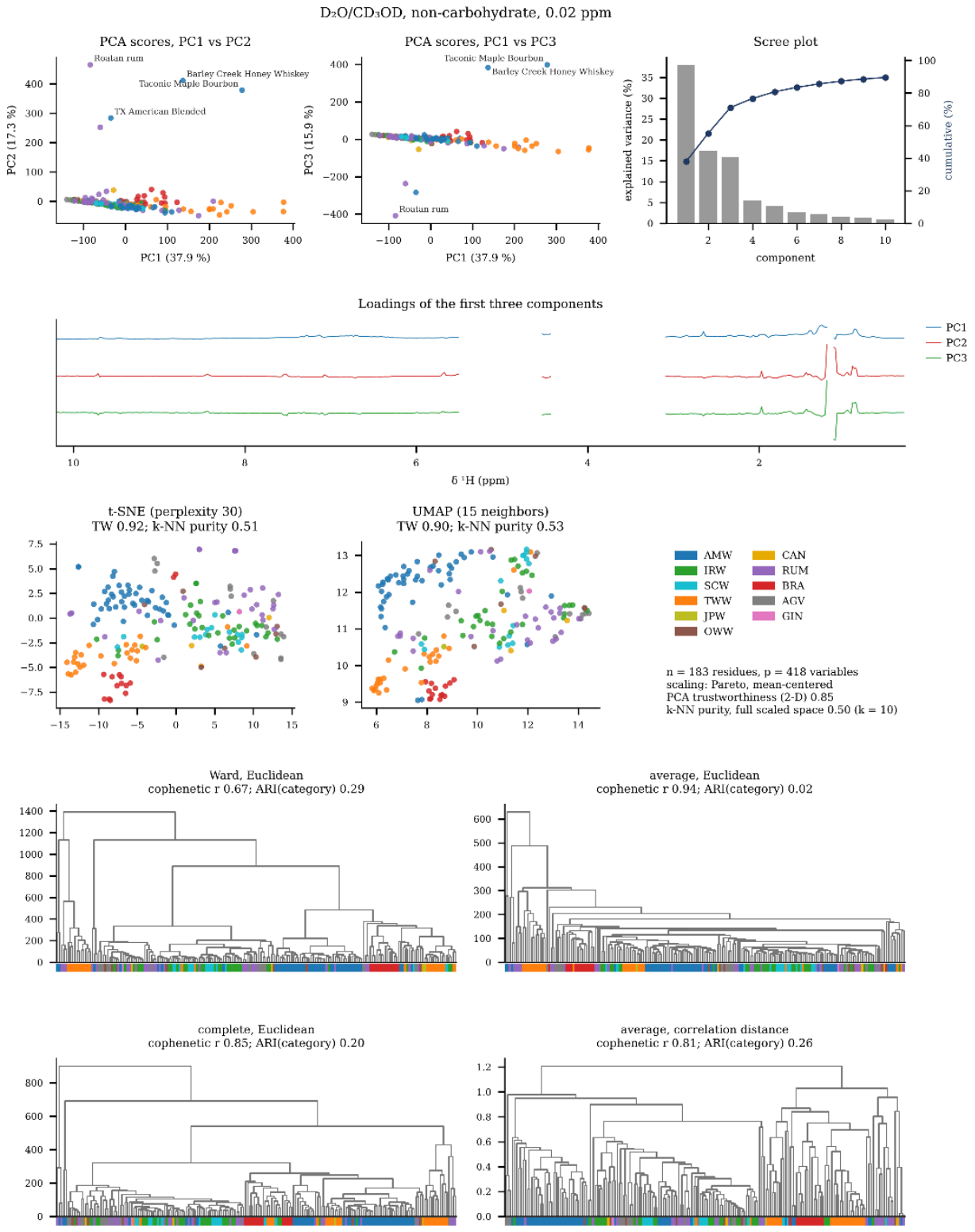

Figure 40. Full chemometric analysis of block D2O-NC (D_2O/CD_3OD residues, carbohydrate windows removed, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S9–S16).

Removing the windows of retained ethanol and the fusel alcohols as well (Figure 41) left a first component (42.3 %) that is again the wine-cask and extract axis, and a second (21.5 %) that is the clearest chemical contrast in the data, between the 1,2-propanediol and vanillin signals of a flavor preparation and the quercitol, syringyl and lignan signals of new charred oak (Section 11.2). The volatile-free block lowered the agreement of the Ward tree with the categories (0.15 against 0.29) but not that of the correlation-distance tree (0.25 against 0.26), which shows that part of the Ward clustering of the non-carbohydrate block rested on the retained volatiles. Because the retention of ethanol follows the sugar content of a residue rather than the composition of the spirit (Section 6.7), the volatile-free block is the safer basis for any chemical interpretation of clusters, even though it clusters the categories less well.

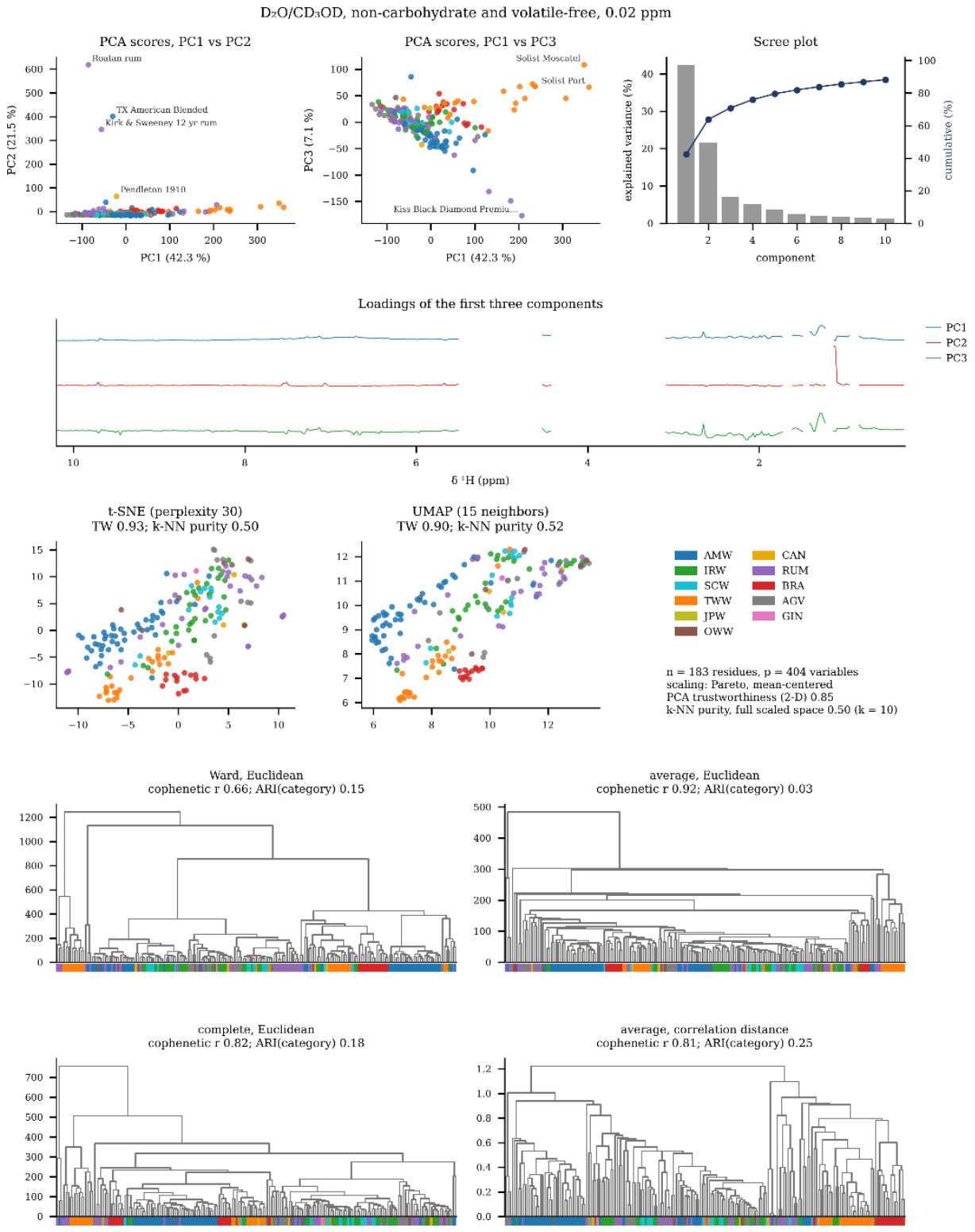

Figure 41. Full chemometric analysis of block D2O-NV (D_2O/CD_3OD residues, carbohydrate and retained-volatile windows removed, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S17–S24).

Binning at 0.04 ppm instead of 0.02 ppm (Figure 42) changed almost nothing: the first three components explained 43.6, 17.6 and 9.8 % of the variance, the Ward tree reached an adjusted Rand index of 0.30, the highest of any block in this study, and the class purity of the scaled space was 0.50. Residual misalignment is therefore not an important source of variance in these spectra once the carbohydrate region is removed, which agrees with the small gain from wider bins in the supervised models (Section 11.4). Adding the fourteen residues dissolved in D_2O alone (Figure 43) moved Litchfield Maple Bourbon, which is one of them, to the top of the first two components beside the other maple and honey whiskeys, and changed the statistics only marginally (adjusted Rand index 0.23 for Ward linkage, purity 0.49); the 0.03 ppm offsets of the carbohydrate signals in D_2O (Section 5.4) do not matter once those signals are excluded.

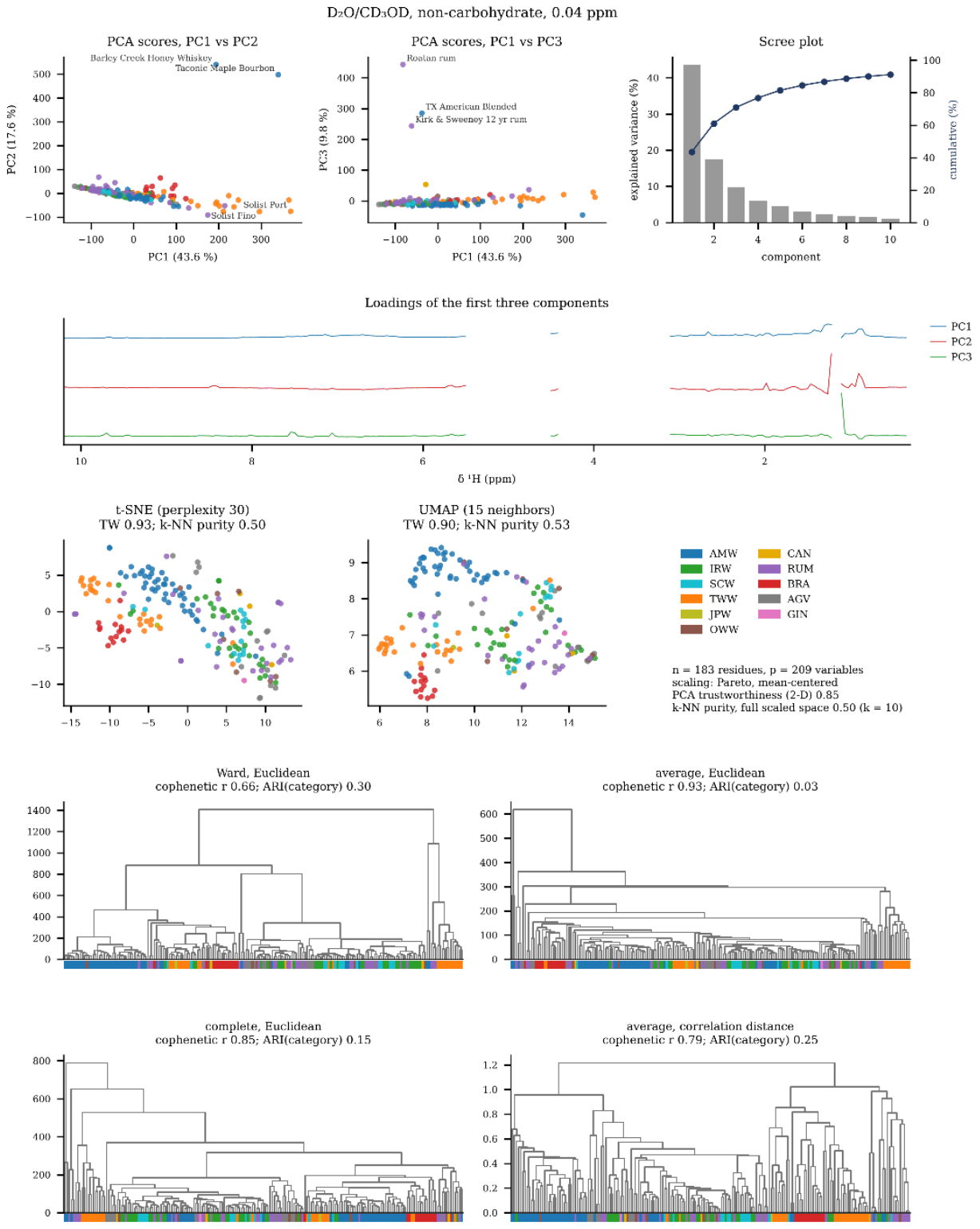

Figure 42. Full chemometric analysis of block D2O-NC04 (as D2O-NC with 0.04 ppm bins), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S25–S32).

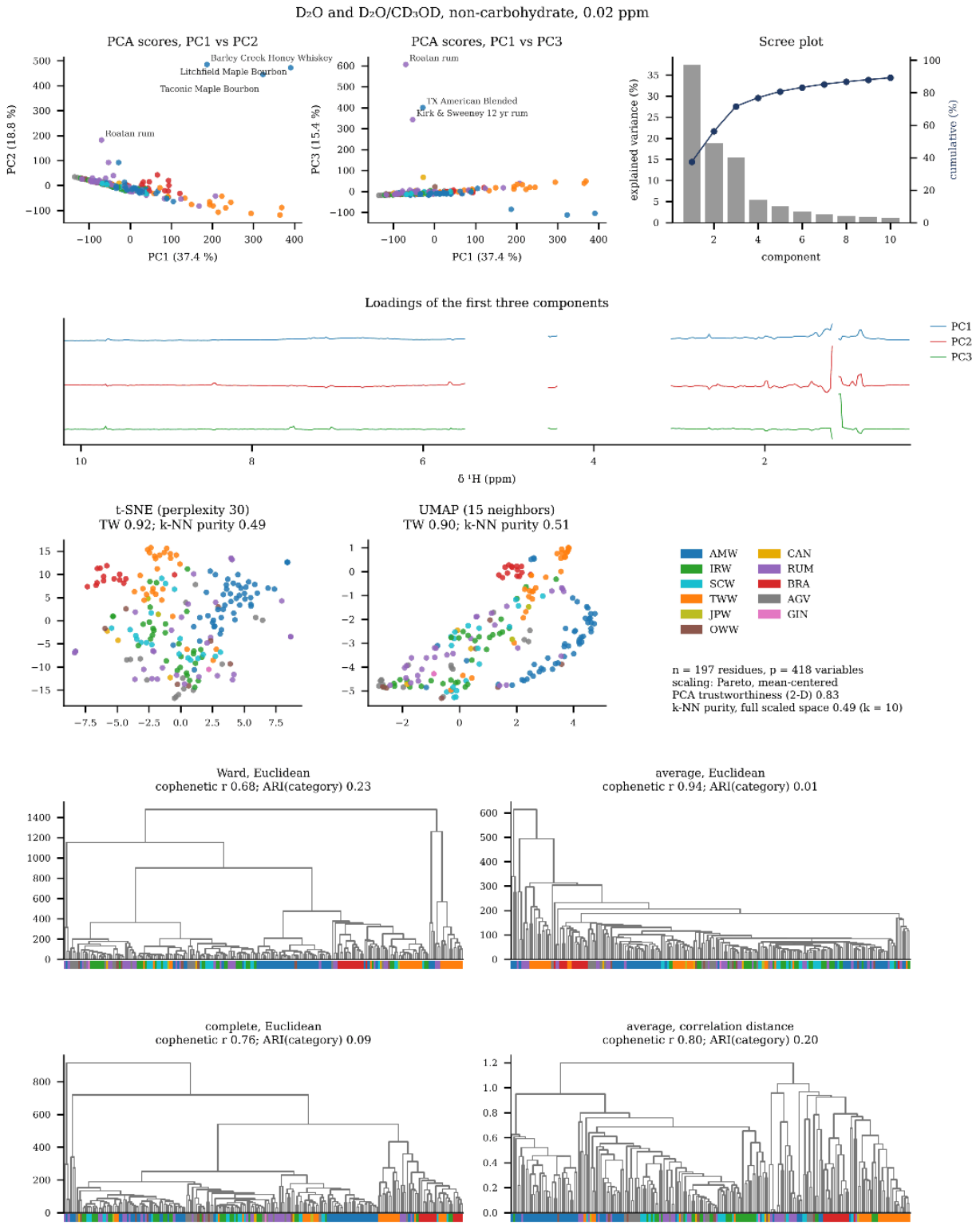

Figure 43. Full chemometric analysis of block D2O-NCall (as D2O-NC, including the fourteen residues dissolved in D₂O alone), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S33–S40).

12.4 Quantified congeners as the variables

The blocks built from the quantified congeners (Figure 44, Figure 45) contain nineteen variables instead of several hundred bins, each with a chemical identity, and they were analyzed on logarithmic concentrations after Pareto scaling and, for comparison, after autoscaling. In both scalings the first component (30.1 % and 28.8 %) is a size component: in the autoscaled block every loading is positive, and in the Pareto-scaled block the variables with the largest spread on the logarithmic scale, retained ethanol, the sugars and lactic acid, weigh most. The cognacs, with Larsen Aqua Ignis at the extreme, lie at one end and the sotols, the low-extract agave spirits and Powers Gold Label at the other. The second component (21.3 % and 21.1 %) is the contrast between sweetening and oak, with sucrose, fructose, glucose, 1,2-propanediol and HMF on one side and the fusel alcohols, quercitol, the syringyl singlet, gallic acid, the benzaldehydes and both pentoses on the other; it runs from Roatan rum, Bozal Cuishe, Kirk & Sweeney, Kiss Black Diamond and the Brugal añejo to Tuthilltown Hudson Whisky, Widow Jane, Coalition, Bernheim and Davidson Reserve. That arabinose and xylose load with the oak markers rather than with the other sugars is a useful check on their assignment and origin (Section 9.5). The third component (9.3 %) contrasts succinic acid, lactic acid, gallic acid and HMF with the fusel alcohols, sucrose and ethanol. At the acid end lie Appleton Estate 21 yr, Redbreast Lustau Edition, the Kavalan Madeira cask, Pierre Ferrand Ambre and Redbreast 21 yr, products whose casks or ages leave wine-derived acids and oxidative products in the residue, and at the other end the honey and maple whiskeys.

The quantified blocks behave differently from the binned blocks in two respects. Their average-linkage cophenetic correlations (0.78 and 0.79) are the lowest of the study, which shows that the distances among residues in a nineteen-dimensional logarithmic space are less hierarchical than those in a spectral space dominated by a few large signals, and their trees chain less: the largest Ward cluster of the Pareto-scaled block held only 33 % of the residues, the smallest proportion of any block. Their agreement with the categories was lower than that of the non-carbohydrate spectral block for Ward linkage (adjusted Rand index 0.16 against 0.29) and equal to it for the correlation distance (0.26), but their class purity was lower (0.43 against 0.50), in line with the lower accuracy of the supervised models built on quantified congeners (Section 11.6). The congeners capture most of the chemistry that separates the categories but not all of the neighborhood structure, part of which lies in the broad phenolic and carbohydrate envelopes and in unassigned minor signals.

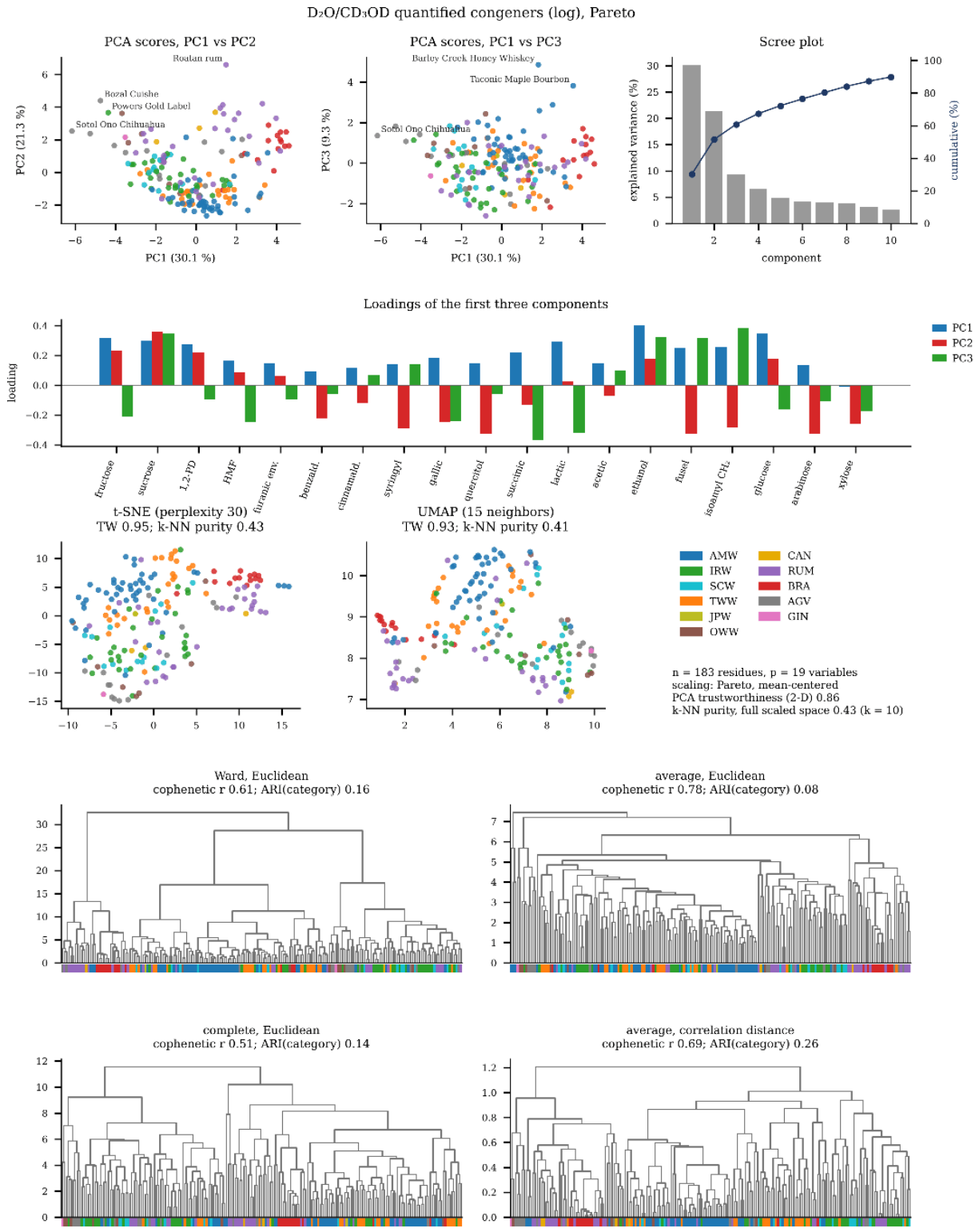

Figure 44. Full chemometric analysis of block D2O-Qp (quantified congeners of the D₂O/CD₃OD residues, logarithmic concentrations, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The loadings are shown for the individual congeners. The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S41–S48).

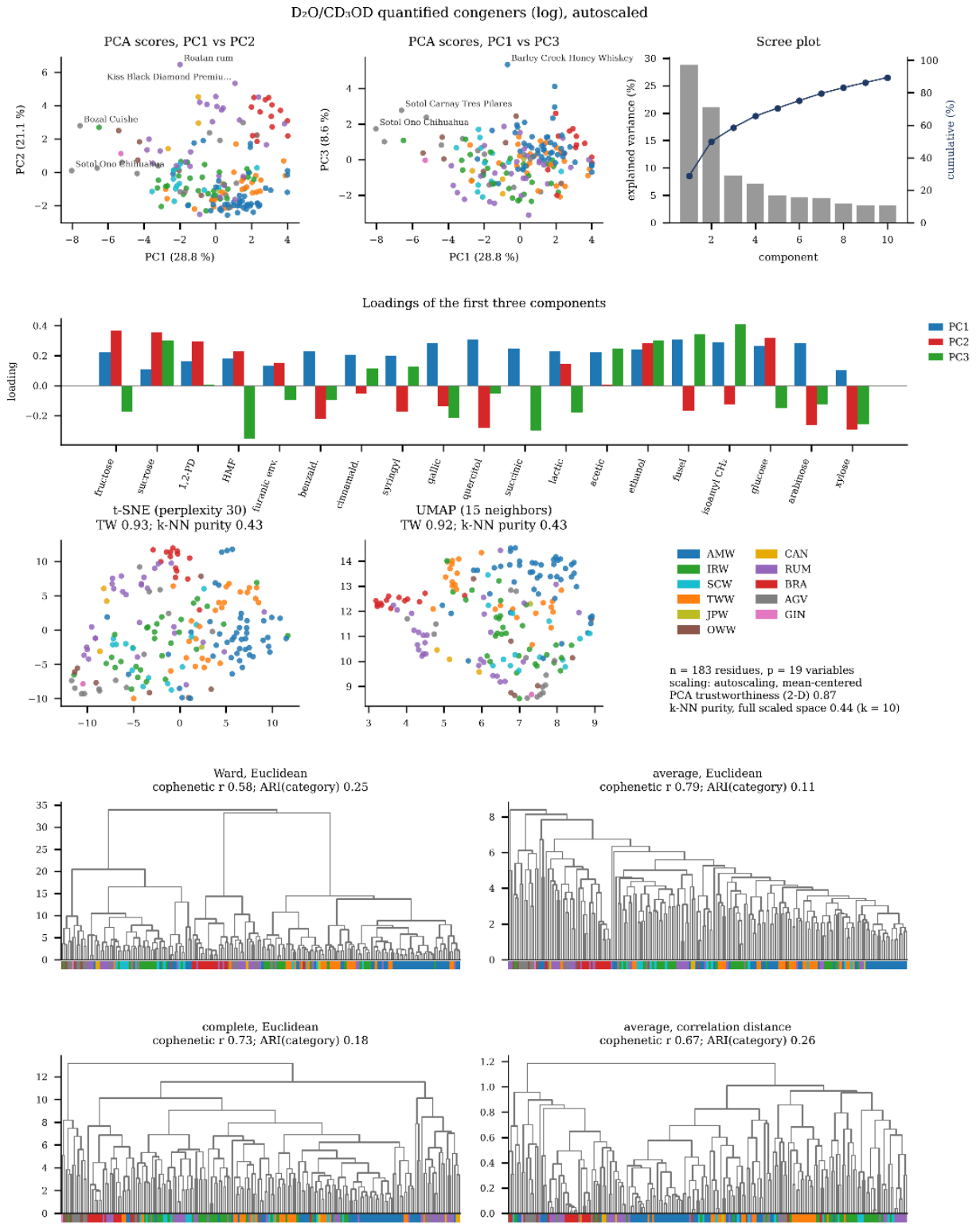

Figure 45. Full chemometric analysis of block D2O-Qa (quantified congeners of the D₂O/CD₃OD residues, logarithmic concentrations, autoscaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The loadings are shown for the individual congeners. The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S49–S56).

12.5 The $CDCl_3$ block, a moisture artifact and the fused block

The first component of the $CDCl_3$ block (Figure 46; 51.9 %) is the lipid content of the residue, with its loadings on the acyl methylene envelope at δ 1.23–1.33 and its highest scores in the cask-strength Kavalan oloroso and Earth Silver malts and the Hudson port, sherry and Short Stack finishes. The second component (13.9 %) is an artifact. Its loadings rise steeply toward δ 1.60, the upper limit of the window that was excluded for dissolved water, and decline smoothly to δ 1.76, which is the shape of the downfield wing of a broadened water line, and its scores are carried by two residues, Tuthilltown Hudson Whisky NY and Coppersea Excelsior Rye. In these residues the integral between δ 1.60 and 1.645 was 39 and 30 times that of the acyl α-methylene triplet at δ 2.28–2.40, against a median of 1.7 (interquartile range 1.2–2.3) over the 61 $CDCl_3$ residues; since every acyl chain carries one β-methylene group for each α-methylene group, a ratio far above about two cannot be explained by acyl chains, and the excess is water whose resonance, broadened and displaced by hydrogen bonding to the polar solutes, extends beyond the excluded window. Eight residues had ratios above three. Version 2.0 did not describe this component, and its presence matters because the same signal enters the fused block.

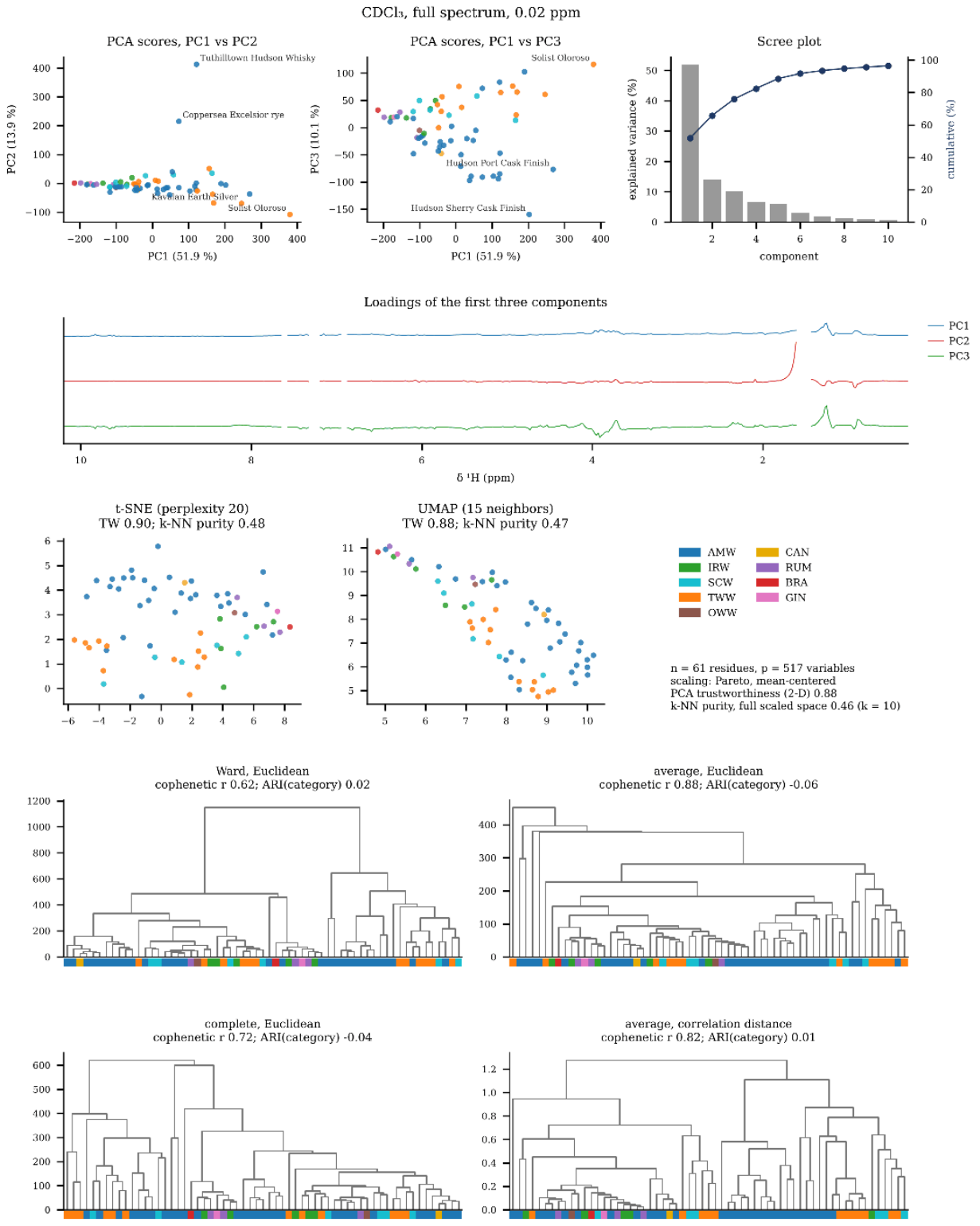

Figure 46. Full chemometric analysis of block CDCl3 (CDCl₃ residues, full spectrum, 0.02 ppm bins, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S57–S61).

With the water window widened to δ 1.45–1.76 (Figure 47), the first component (59.4 %) was again the lipid content, and the second (12.3 %) became the chemically meaningful contrast that had been the third component before: the lipid envelope and the acyl methyl signals against the methoxyl region at δ 3.85–3.97, where the Ar–OCH_3 singlets of syringaldehyde, vanillin, sinapaldehyde, coniferaldehyde and the lignans resonate. It separates the lipid-rich Kavalan oloroso and ex-bourbon casks, Hudson Short Stack and Old Overholt 114 Proof from the Tuthilltown whiskeys matured in small new-oak casks (Hudson Sherry Cask Finish, Do The Rye Thing, This is Your Favorite Summer Rye, Back Room Deal and the cask-strength barrel select). The third component (7.8 %) picked out the 1,2-propanediol of TX American Blended (δ 1.17–1.19, 3.41 and 3.63 in $CDCl_3$). Removing the artifact improved the agreement of every tree with the categories except average linkage (Ward 0.02 to 0.17, complete −0.04 to 0.08, correlation distance 0.01 to 0.14) and raised the class purity of the scaled space from 0.46 to 0.48, against a chance level of 0.31. Even so, the $CDCl_3$ block organizes the residues by their lipid and methoxyl contents, that is, by cask strength and chill filtration on one axis and by new oak on the other, rather than by category. Freeze-dried residues are hygroscopic, and the lesson for routine work is to dry the $CDCl_3$ over molecular sieves, to record the water line of every spectrum and to define the excluded window from the data rather than from the tabulated position of water in dry $CDCl_3$ (δ 1.56 [69]).

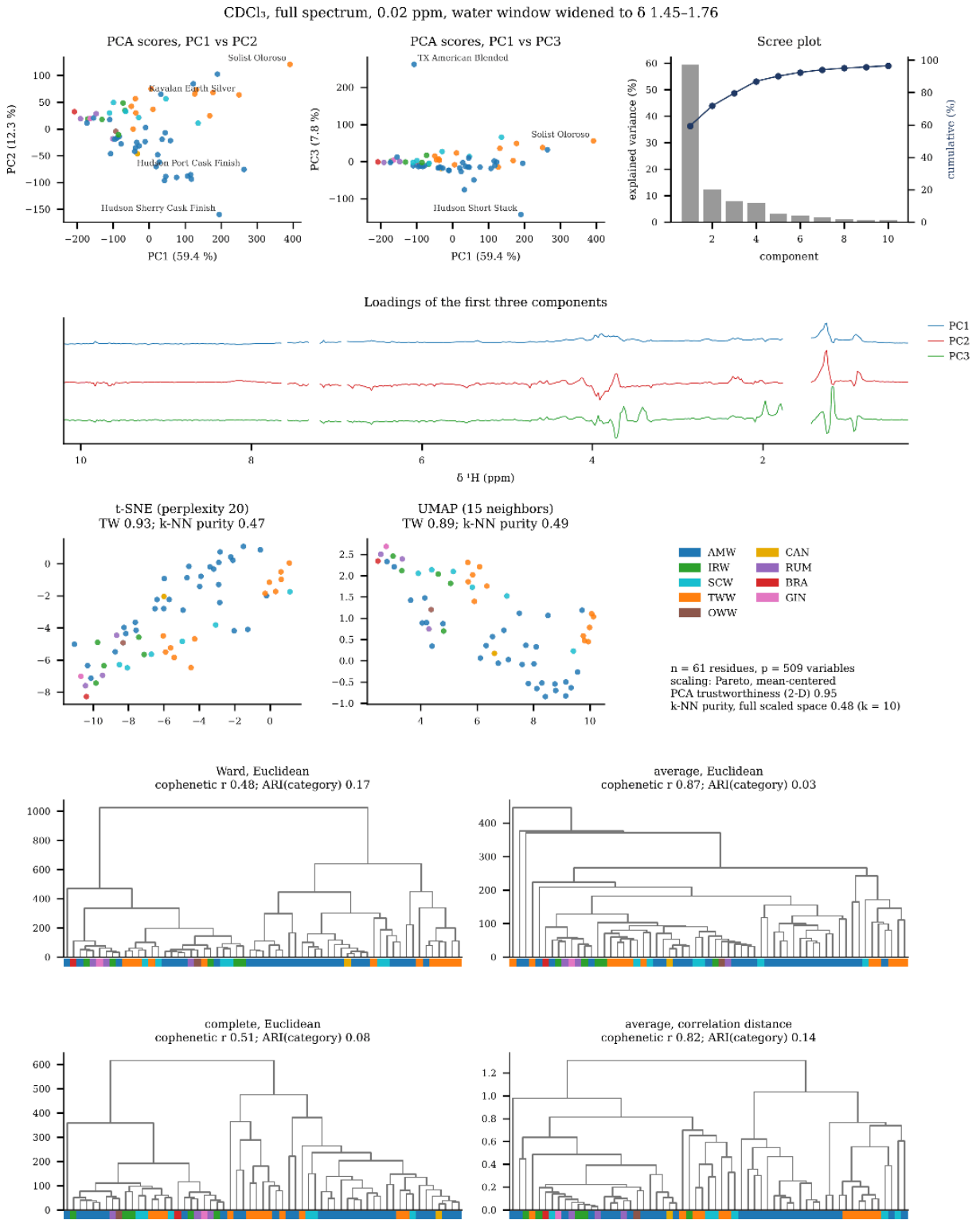

Figure 47. Full chemometric analysis of block CDCl3-W (CDCl₃ residues with the water window widened to δ 1.45–1.76, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S62–S66).

The fused D_2O/CD_3OD and $CDCl_3$ block of the 51 paired products (Figure 48) inherited the lipid axis of both solvents as its first component (38.2 %), which runs from the wine-seasoned Kavalan malts and Hudson Port Cask Finish to Rum 28, the An Dulaman gin, Doc Swinson's Alter Ego and TX American Blended, and the water artifact appeared in its second and third components alongside the 1,2-propanediol of TX American Blended and the methoxyl signals of the new-oak whiskeys. With the $CDCl_3$ water window widened (Figure 49), the second component (15.2 %) became a clean contrast between 1,2-propanediol and the methoxyl signals, and the adjusted Rand index of the complete-linkage tree rose from −0.02 to 0.16. The fused blocks had the highest class purity of all approaches (0.55 and 0.55), but their chance level was also the highest (0.35), because nearly half of the paired products are American whiskeys; relative to chance, fusion did not improve on the non-carbohydrate D_2O block alone, in agreement with the supervised result of Section 11.8.

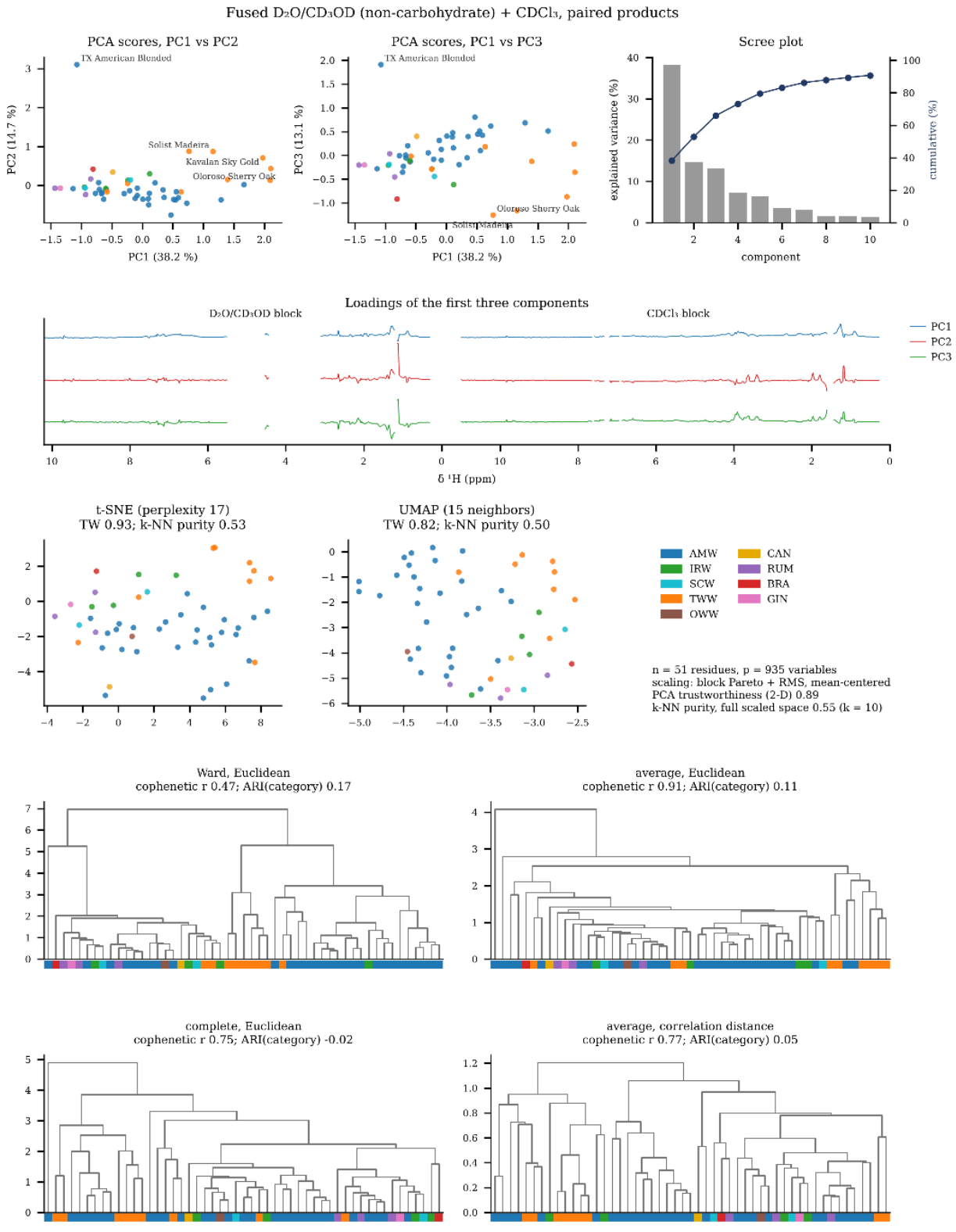
Fused D₂O/CD₃OD (non-carbohydrate) + CDCl₃, paired products
PCA scores, PC1 vs PC2
PC2 (14.7 %)
PC1 (38.2 %)
TX American Blended
Solist Madeira
Kavalan Sky Gold
Oloroso Sherry Oak
PCA scores, PC1 vs PC3
PC3 (13.1 %)
PC1 (38.2 %)
Scree plot
explained variance (%)
cumulative (%)
component
Loadings of the first three components
D₂O/CD₃OD block
CDCl₃ block
PC1
PC2
PC3
δ ¹H (ppm)
t-SNE (perplexity 17)
TW 0.93; k-NN purity 0.53
UMAP (15 neighbors)
TW 0.82; k-NN purity 0.50
AMW
IRW
SCW
TWW
OWW
CAN
RUM
BRA
GIN
n = 51 residues, p = 935 variables
scaling: block Pareto + RMS, mean-centered
PCA trustworthiness (2-D) 0.89
k-NN purity, full scaled space 0.55 (k = 10)
Ward, Euclidean
cophenetic r 0.47; ARI(category) 0.17
average, Euclidean
cophenetic r 0.91; ARI(category) 0.11
complete, Euclidean
cophenetic r 0.75; ARI(category) -0.02
average, correlation distance
cophenetic r 0.77; ARI(category) 0.05

Figure 48. Full chemometric analysis of block FUSED-DC (D2O-NC and CDCl3 blocks of the paired products, each Pareto-scaled and normalized to unit root-mean-square, then concatenated), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S67–S71).

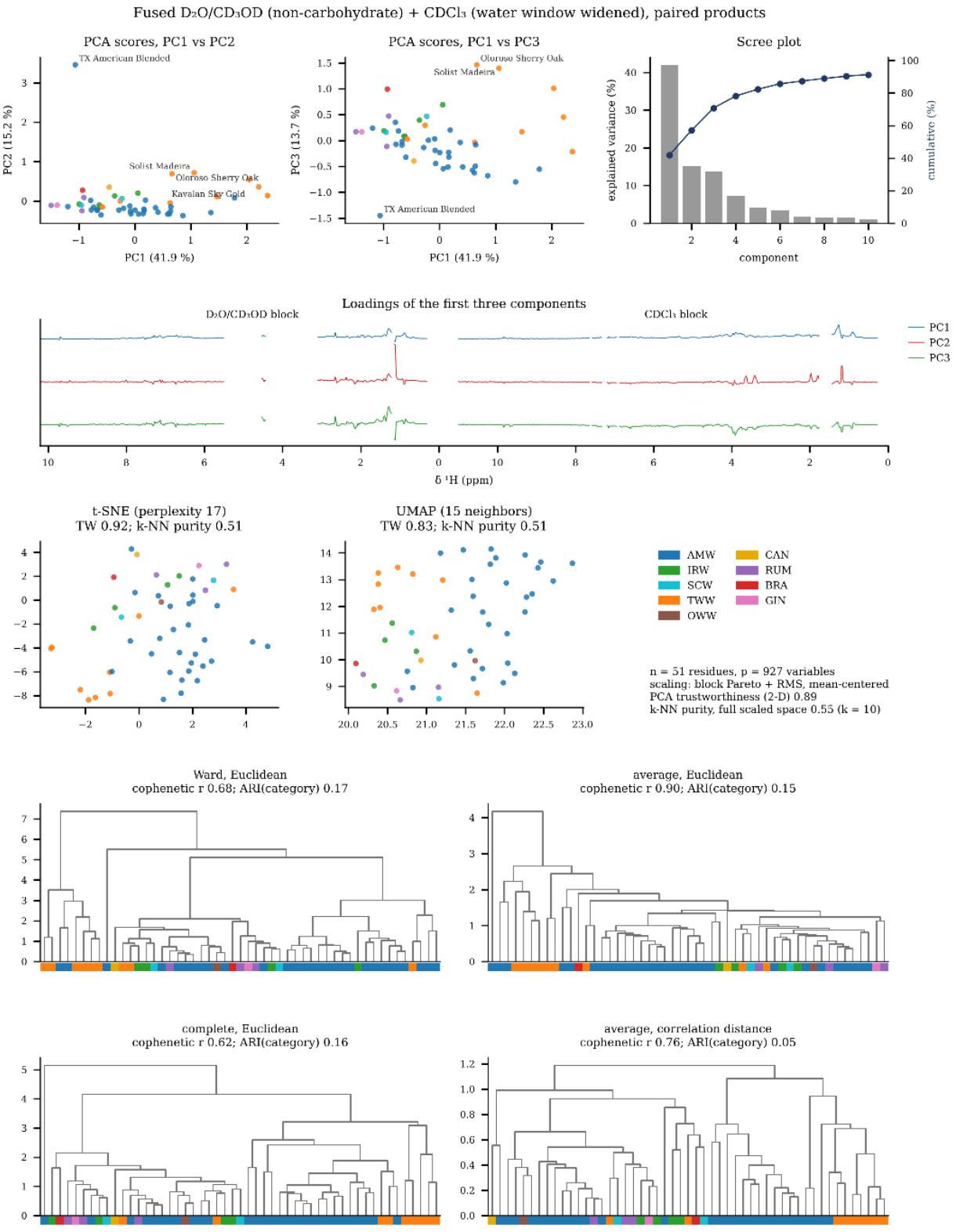

Figure 49. Full chemometric analysis of block FUSED-W (as FUSED-DC with the CDCl3-W block), laid out as Figure 39; the strip beneath each dendrogram gives the spirit category of each residue (abbreviations as in Figure 34). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S72–S76).

12.6 The DMSO-d₆ blocks

The DMSO-d₆ blocks contain only 29 residues, most of them Kavalan malts, and they must be read with that in mind. In the full block (Figure 50) the first component (48.1 %) belongs to a single residue: the score of Adirondack Maple Barrel Bourbon is 15 times larger than that of any other residue, and the loadings are those of its sucrose (the Glc H-1 doublet at δ 5.17–5.19 and the ring protons at δ 3.13–3.21) and of the retained ethanol triplet at δ 1.03–1.09. The pairing of sweetening with ethanol retention is the same as in D₂O (Section 6.7), and the residue is genuinely different from the others, but a component that describes one residue in twenty-nine is not a description of the data set. The second component (32.7 %) is the ethyl ester and lipid region (the ester methyl triplet at δ 1.17 and the methylene envelope at δ 1.23), with its highest scores in the Moscatel, amontillado, fino, manzanilla and King Car wine casks and its lowest in Adirondack Rum, Glen Grant 12 yr and the Kavalan peaty cask, and the third (8.7 %) is glucose, whose doublets at δ 4.25–4.29 and 4.91 and ring protons at δ 2.89–3.09 load together. At $k = 2$, every Euclidean tree separated the maple-barrel residue from all the others, so that their agreement with the cask type was nil (adjusted Rand index −0.02); only the correlation-distance tree, which is insensitive to the amount of extract, divided the residues partly along the line between wine-seasoned and other casks (0.24).

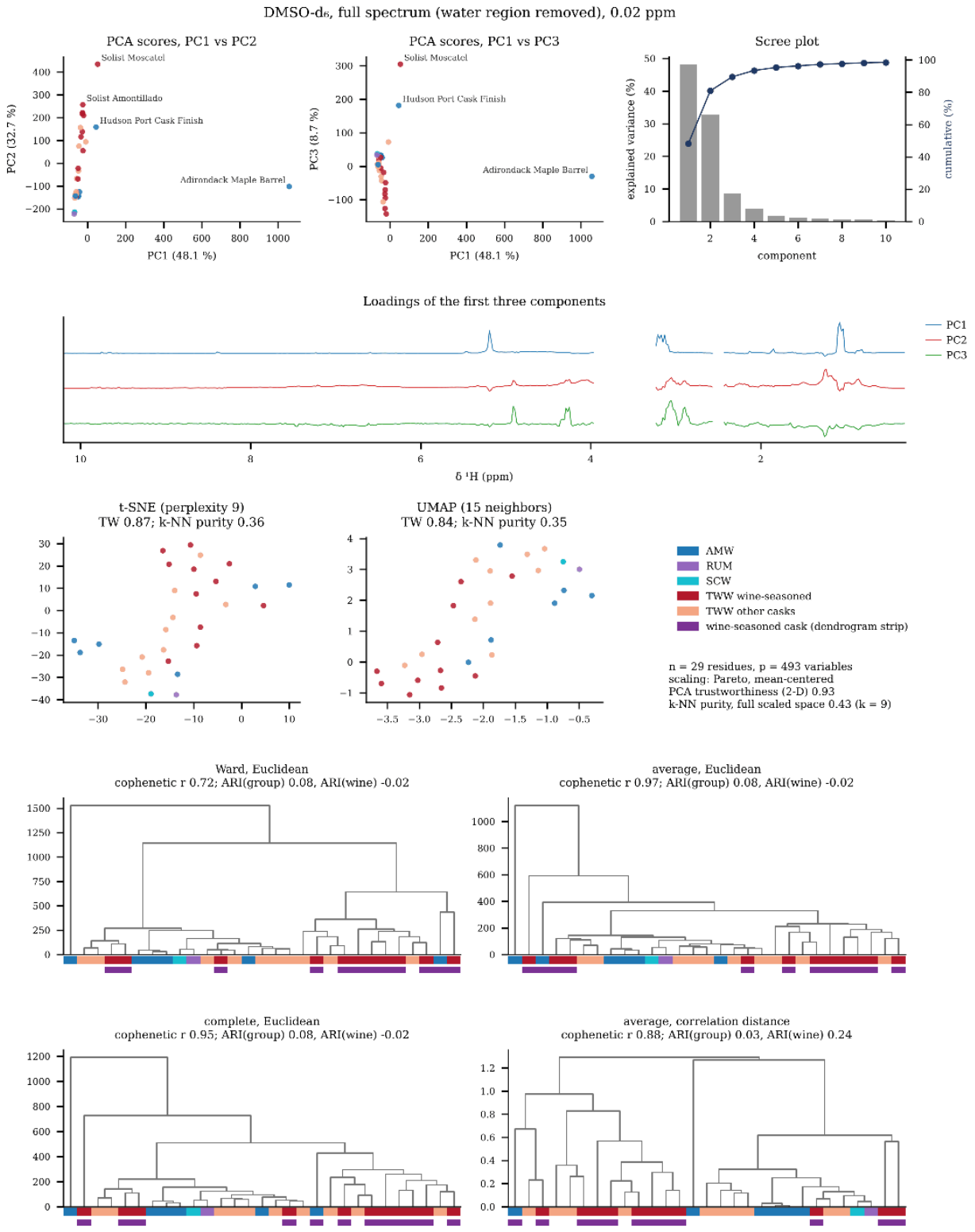

Figure 50. Full chemometric analysis of block DMSO-F (DMSO-d₆ residues, full spectrum without the solvent and water regions, 0.02 ppm bins, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S77–S79).

With that residue removed (Figure 51), the structure of the remaining 28 residues emerged. The first component (62.1 %) is the ester and lipid envelope together with glucose near δ 3.05, and it separates the wine-seasoned casks from the others; the second (16.7 %) contrasts glucose with the esters and lipids and divides the wine casks themselves, placing the Moscatel cask, the Hudson port finish and the Kavalan brandy cask, whose residues are rich in glucose, against the manzanilla, amontillado, oloroso, King Car and puncheon casks, whose residues are rich in esters and poor in glucose. The contrast is chemically plausible, since Moscatel and port are sweet wines that leave residual sugar in the wood whereas fino, manzanilla, amontillado and oloroso are dry, but with one or two casks of each kind it is a hypothesis rather than a finding. At $k = 2$ the Ward tree placed 9 of the 12 wine-seasoned residues in one cluster and 13 of the 16 other residues in the other (adjusted Rand index 0.30), and the correlation-distance tree reached the same value (0.30); the residue groups (Kavalan wine casks, Kavalan other casks, American whiskeys, the rum and the Scotch malt) were not recovered by any tree (adjusted Rand index ≤ 0.08).

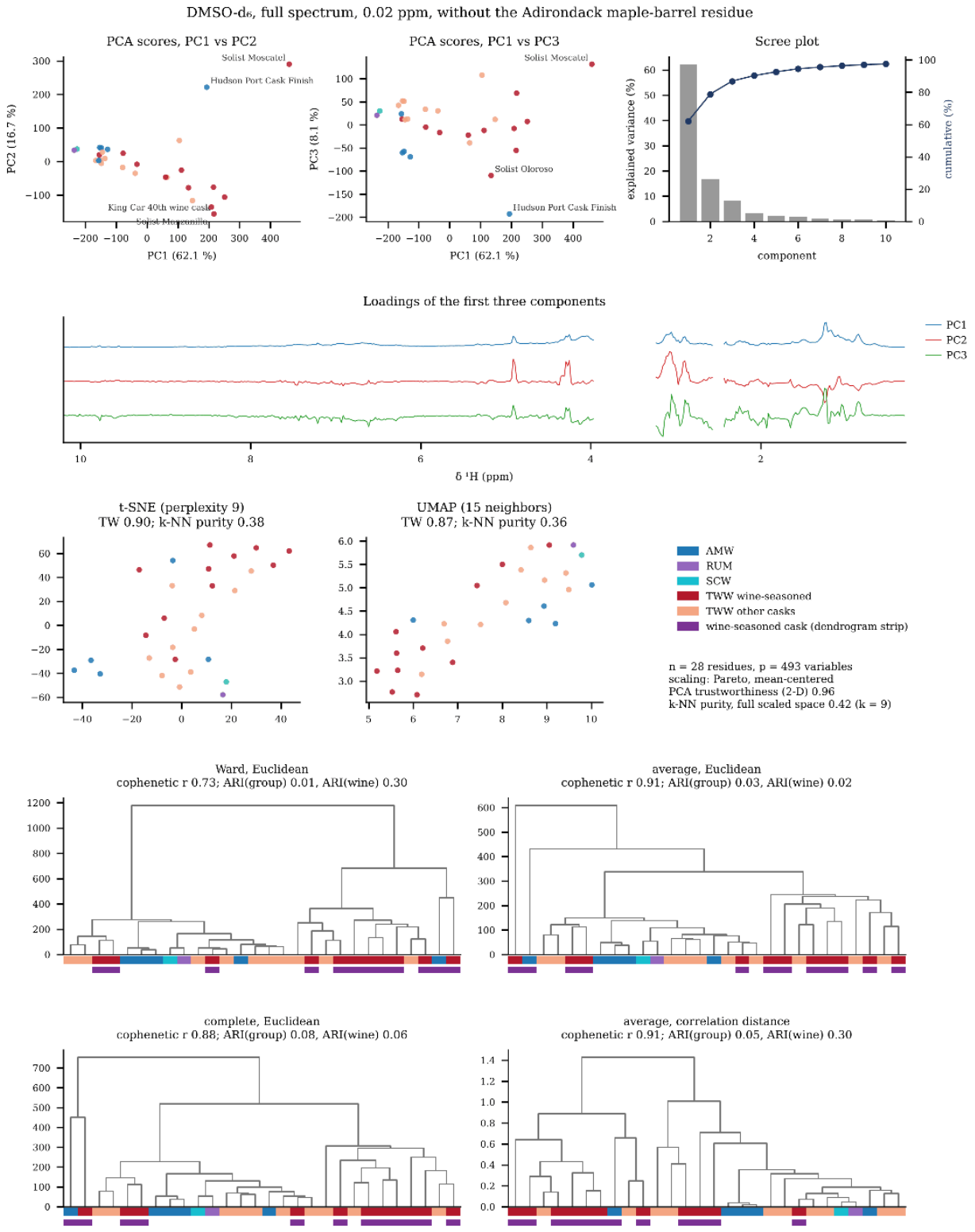

Figure 51. Full chemometric analysis of block DMSO-Fx (as DMSO-F without the Adirondack Maple Barrel Bourbon residue), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S80–S82).

The non-carbohydrate DMSO-d₆ block (Figure 52), from which δ 2.80–4.00 and 4.15–5.30 were also removed, kept the maple-barrel residue as its first component (45.3 %) through its retained ethanol, and its second component (40.0 %) was again the ester and lipid axis; binning at 0.04 ppm (Figure 53) reproduced the full block almost exactly (49.2, 33.1 and 8.8 % on the first three components). The quantified DMSO-d₆ congeners (Figure 54, Figure 55) gave the most informative of the DMSO-d₆ analyses. Their first component (37.2 %) is oak extraction and acidity, with positive loadings on syringaldehyde, the free-fatty-acid α-methylene triplet, the syringyl singlet at δ 7.20, acetic acid, vanillin, gallic acid, xylose and succinic acid and a negative loading on sucrose, and it runs from the King Car wine cask, the Paul Chiang puncheon, the manzanilla and amontillado casks and the Hudson port finish to the Adirondack maple bourbon, Glen Grant 12 yr, Adirondack Rum, Hudson Bright Lights and the Kavalan peaty cask. The second (16.4 %) is dominated by retained ethanol together with the esters and succinic acid, and the third (13.7 %) by ellagic acid, sucrose, the singlet at δ 7.39 and the cinnamaldehydes. The quantified congeners also gave the highest agreement with the residue groups of any DMSO-d₆ approach (adjusted Rand index 0.25 for complete linkage of the Pareto-scaled block and 0.22 for Ward linkage of the autoscaled block, against at most 0.08 for the spectral blocks), which suggests that in a small data set a few chemically chosen variables carry the group structure better than a spectral block whose variance is controlled by its largest signals. The trees are fragile, however: correcting the succinic acid of two residues (Section 9.7) changed the adjusted Rand index of the Ward tree of the Pareto-scaled block from 0.27 to 0.18, a reminder of how little it takes to rearrange a dendrogram of twenty-nine residues.

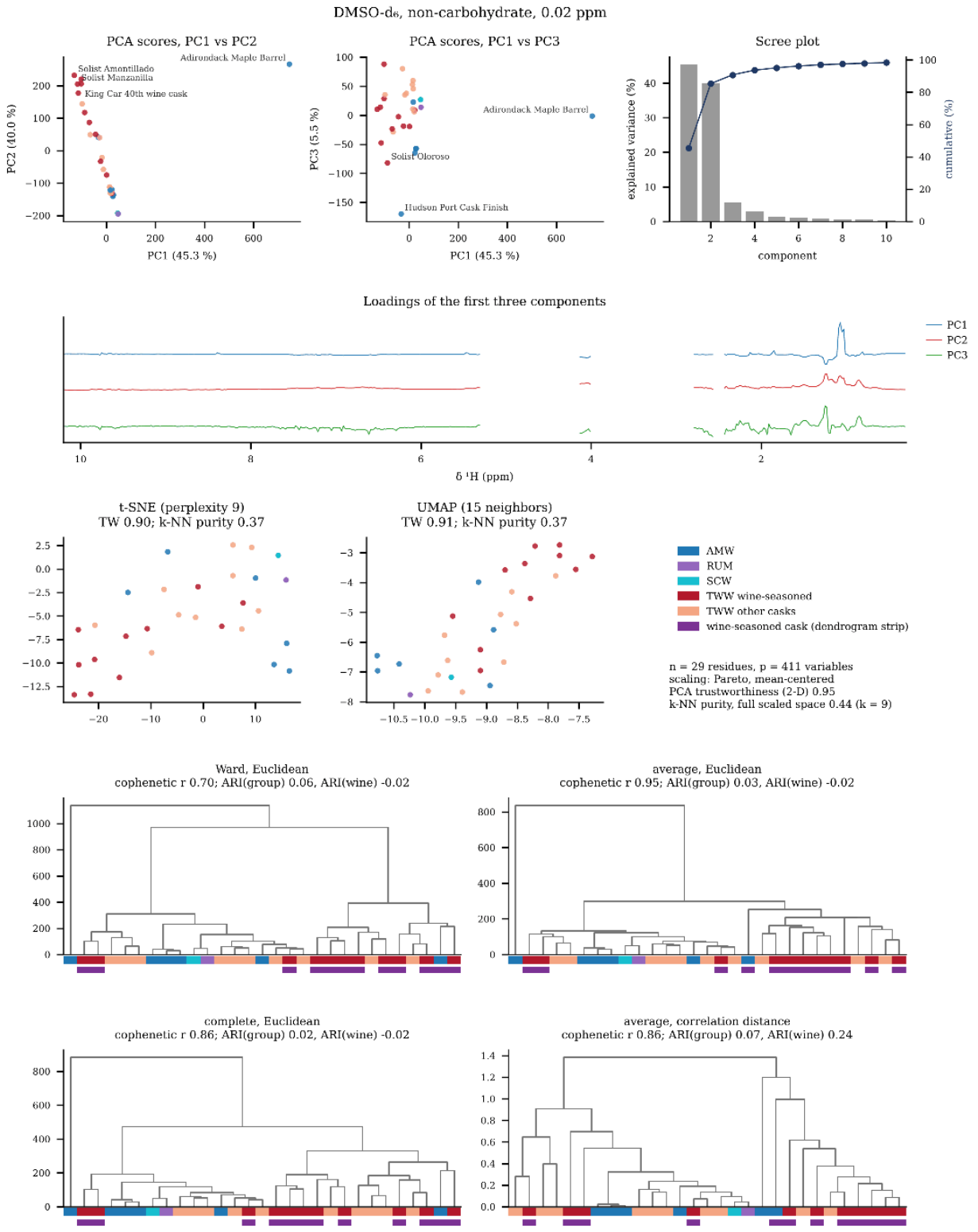

Figure 52. Full chemometric analysis of block DMSO-NC (DMSO-d₆ residues with the carbohydrate windows also removed, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S83–S85).

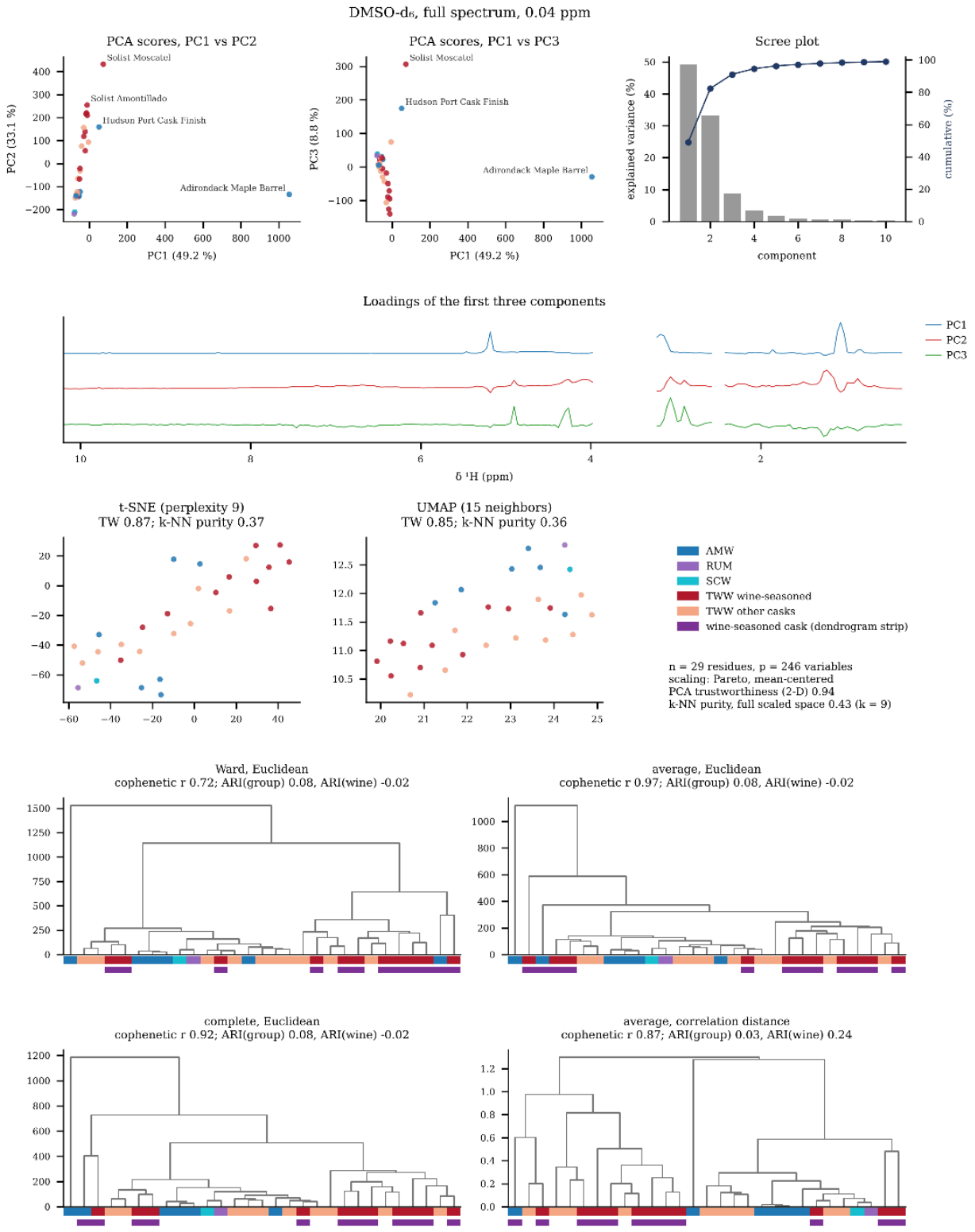

Figure 53. Full chemometric analysis of block DMSO-F04 (as DMSO-F with 0.04 ppm bins), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S86–S88).

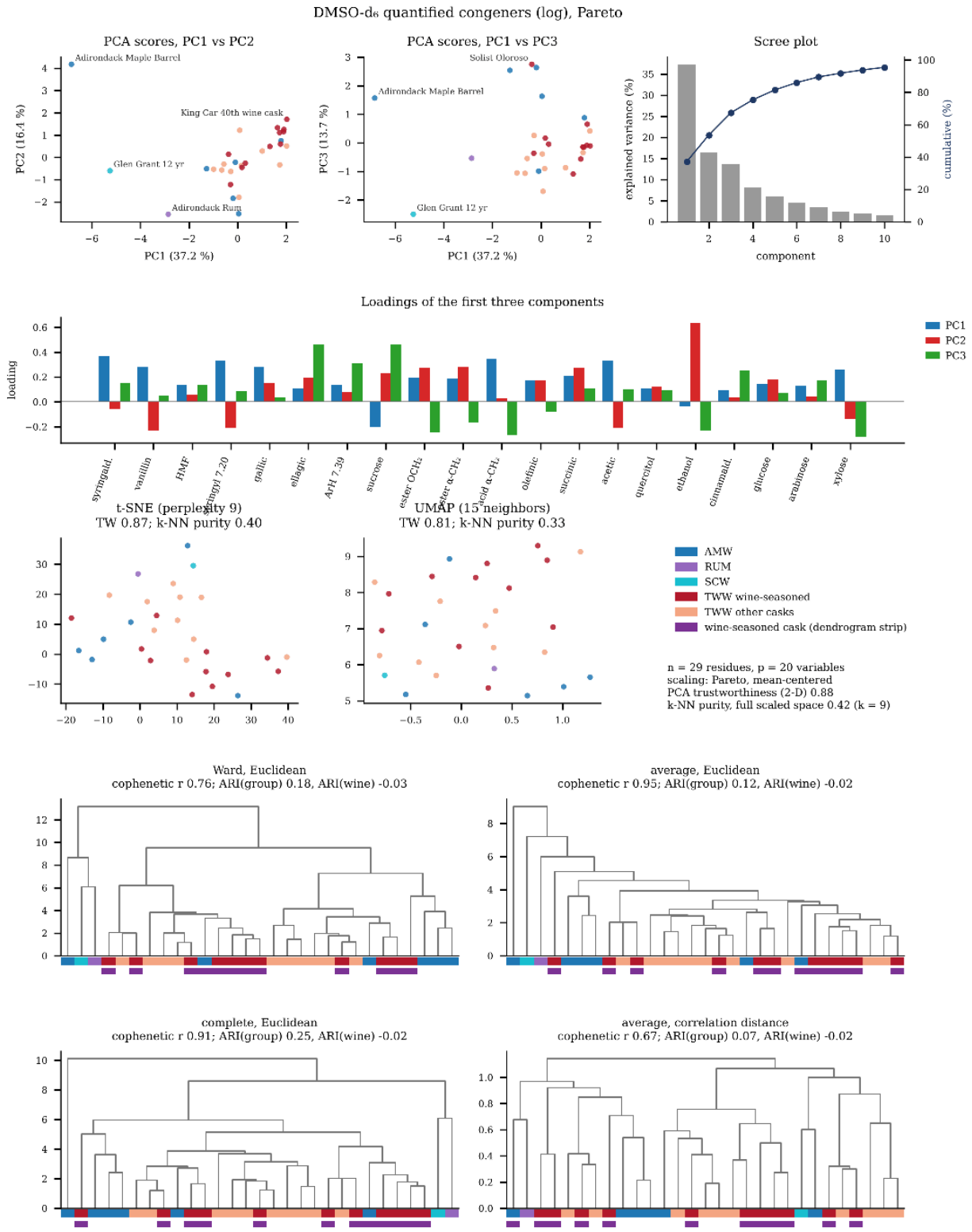

Figure 54. Full chemometric analysis of block DMSO-Qp (quantified congeners of the DMSO-d₆ residues, logarithmic concentrations, Pareto scaling), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The loadings are shown for the individual congeners. The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S89–S91).

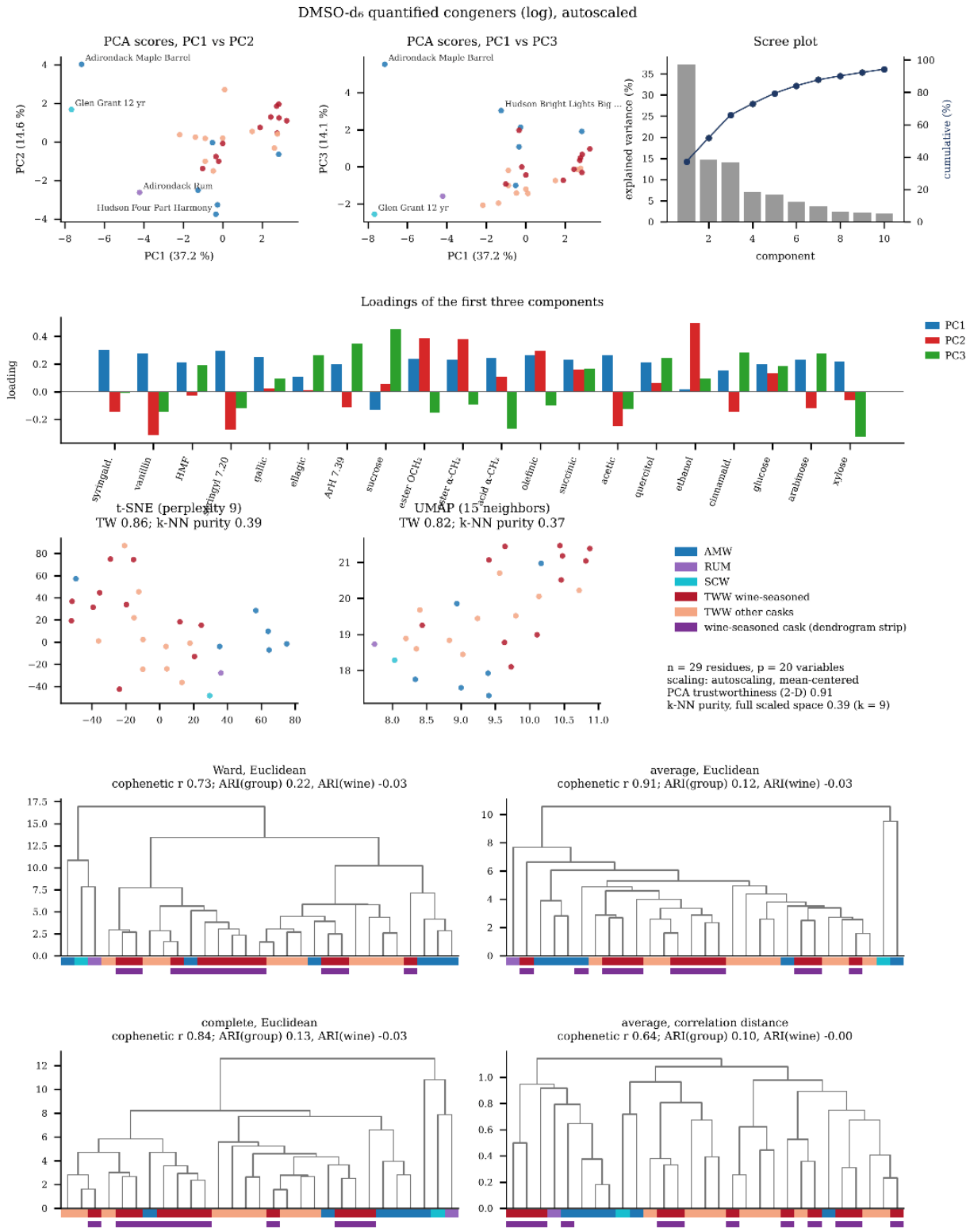

Figure 55. Full chemometric analysis of block DMSO-Qa (quantified congeners of the DMSO-d₆ residues, logarithmic concentrations, autoscaling), laid out as Figure 39; the strip beneath each dendrogram gives the residue group (Kavalan wine-seasoned, Kavalan other casks, American whiskey, Scotch whisky, rum). The loadings are shown for the individual congeners. The residue names cannot be read at this size; full-page versions of the score plots, embeddings and dendrograms of this block, with every residue named, are given in the Addendum (Figures S92–S94).

12.7 Hierarchical clustering across all approaches

Figure 56 and Table 21 place the four linkages of all seventeen blocks side by side, and three regularities appear that do not depend on the solvent or on the data approach. First, the linkage with the most faithful tree was never the linkage with the most useful partition. Average linkage on Euclidean distances gave the highest cophenetic correlation in seventeen of the 17 blocks and Ward linkage the lowest in thirteen, yet Ward linkage gave the highest adjusted Rand index in eleven blocks, the correlation distance in four, complete linkage in two and average linkage in none. The reason is chaining: the largest average-linkage cluster held between 55 and 96 % of the residues, whereas the largest Ward cluster held between 33 and 74 %. When the distances are dominated by a few residues of extreme composition, the most faithful tree is the one that joins those residues last, and the most faithful tree is then the least informative about everything else. A cophenetic correlation quoted in support of a dendrogram should therefore never be accepted without the composition of the clusters and an external measure of agreement.

Second, no linkage recovered the spirit categories well in any block; the highest adjusted Rand index of the study was 0.30. The categories of spirits are not the natural clusters of their freeze-dried residues, which group instead by sweetening and flavor additions, by cask seasoning and by the amount of extract, and an unsupervised analysis that is expected to find the categories will disappoint. Third, the correlation distance was the most robust single choice: it combined cophenetic correlations close to those of average linkage with adjusted Rand indices close to those of Ward linkage in the spectral D_2O blocks, it was the only tree to separate wine-seasoned from other casks in the full DMSO-d_6 block, and it is insensitive by construction to differences in the total amount of extract, which in this data set span two orders of magnitude.

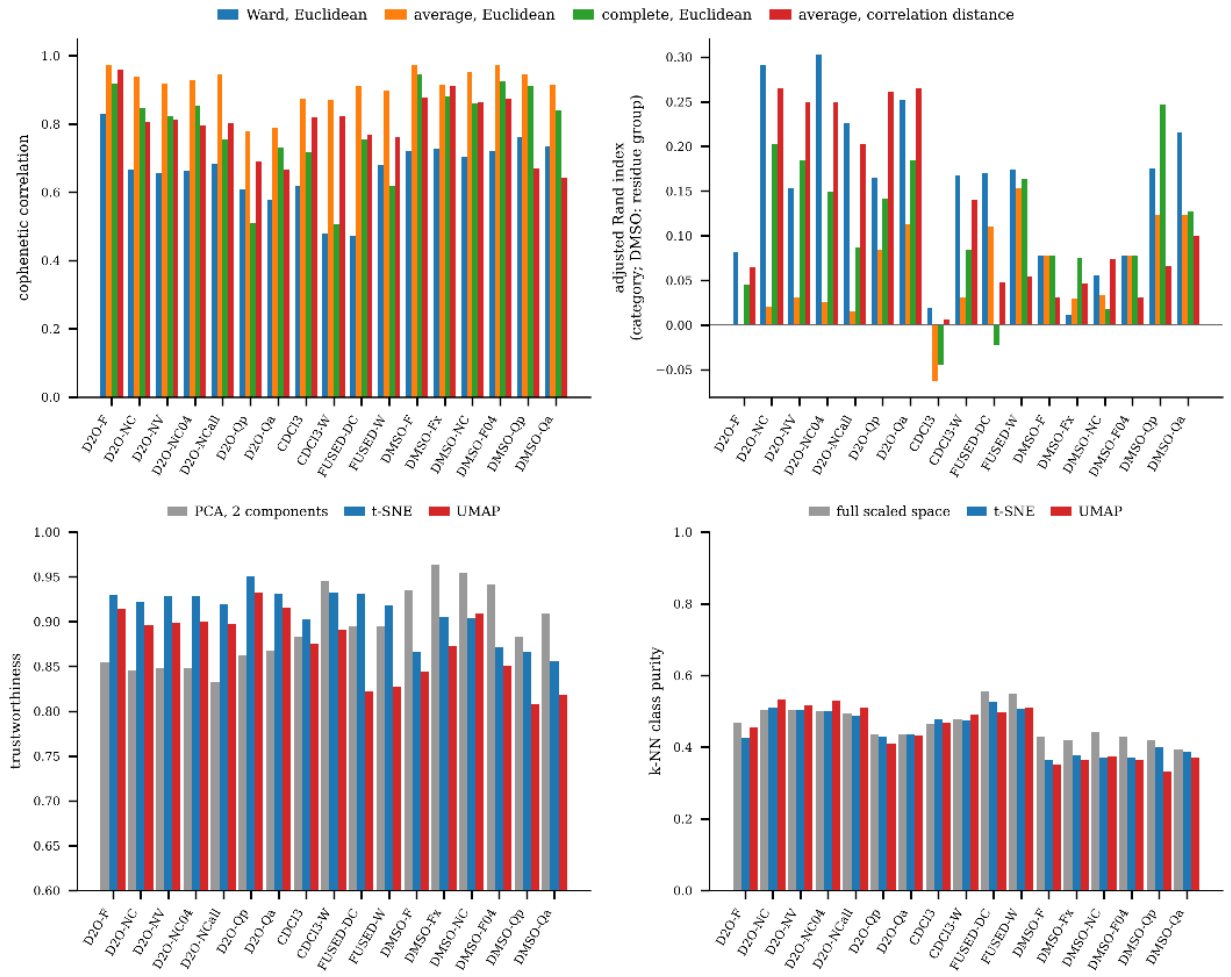

Figure 56. Summary of the unsupervised analyses of all seventeen data approaches: cophenetic correlation (top left) and adjusted Rand index against the spirit categories or, for the DMSO-d₆ blocks, the residue groups (top right) for the four linkages; trustworthiness of the two-dimensional PCA, t-SNE and UMAP displays (bottom left); and *k*-nearest-neighbor class purity in the full scaled space and in the t-SNE and UMAP embeddings (bottom right). Block names as in Table 20.

Table 21. Principal components and hierarchical clustering of every data approach. Linkages in the order Ward / average / complete (Euclidean distances) / average (correlation distance).

Block	Explained variance PC1 / PC2 / PC3 (%)	Cophenetic correlation	Adjusted Rand index (category; DMSO-d₆: residue group)	Largest cluster (% of residues)	Adjusted Rand index, wine-seasoned vs other (DMSO-d₆)
D2O-F	70.3 / 14.2 / 4.6	0.83 / 0.97 / 0.92 / 0.96	0.08 / 0.00 / 0.05 / 0.06	74 / 96 / 89 / 79	—
D2O-NC	37.9 / 17.3 / 15.9	0.67 / 0.94 / 0.85 / 0.81	0.29 / 0.02 / 0.20 / 0.26	44 / 89 / 45 / 43	—
D2O-NV	42.3 / 21.5 / 7.1	0.66 / 0.92 / 0.82 / 0.81	0.15 / 0.03 / 0.18 / 0.25	52 / 90 / 46 / 45	—
D2O-NC04	43.6 / 17.6 / 9.8	0.66 / 0.93 / 0.85 / 0.79	0.30 / 0.03 / 0.15 / 0.25	45 / 89 / 49 / 45	—
D2O-NCall	37.4 / 18.8 / 15.4	0.68 / 0.94 / 0.76 / 0.80	0.23 / 0.01 / 0.09 / 0.20	47 / 89 / 76 / 48	—
D2O-Qp	30.1 / 21.3 / 9.3	0.61 / 0.78 / 0.51 / 0.69	0.16 / 0.08 / 0.14 / 0.26	33 / 78 / 39 / 34	—
D2O-Qa	28.8 / 21.1 / 8.6	0.58 / 0.79 / 0.73 / 0.67	0.25 / 0.11 / 0.18 / 0.26	34 / 74 / 52 / 37	—
CDCl3	51.9 / 13.9 / 10.1	0.62 / 0.88 / 0.72 / 0.82	0.02 / −0.06 / −0.04 / 0.01	51 / 92 / 69 / 51	—
CDCl3-W	59.4 / 12.3 / 7.8	0.48 / 0.87 / 0.51 / 0.82	0.17 / 0.03 / 0.08 / 0.14	44 / 80 / 41 / 41	—
FUSED-DC	38.2 / 14.7 / 13.1	0.47 / 0.91 / 0.75 / 0.77	0.17 / 0.11 / −0.02 / 0.05	43 / 80 / 65 / 45	—
FUSED-W	41.9 / 15.2 / 13.7	0.68 / 0.90 / 0.62 / 0.76	0.17 / 0.15 / 0.16 / 0.05	43 / 82 / 49 / 51	—
DMSO-F	48.1 / 32.7 / 8.7	0.72 / 0.97 / 0.95 / 0.88	0.08 / 0.08 / 0.08 / 0.03	55 / 55 / 55 / 52	−0.02 / −0.02 / −0.02 / 0.24
DMSO-Fx	62.1 / 16.7 / 8.1	0.73 / 0.91 / 0.88 / 0.91	0.01 / 0.03 / 0.08 / 0.05	57 / 57 / 57 / 50	0.30 / 0.02 / 0.06 / 0.30
DMSO-NC	45.3 / 40.0 / 5.5	0.70 / 0.95 / 0.86 / 0.86	0.06 / 0.03 / 0.02 / 0.07	41 / 55 / 48 / 48	−0.02 / −0.02 / −0.02 / 0.24
DMSO-F04	49.2 / 33.1 / 8.8	0.72 / 0.97 / 0.92 / 0.87	0.08 / 0.08 / 0.08 / 0.03	55 / 55 / 55 / 52	−0.02 / −0.02 / −0.02 / 0.24
DMSO-Qp	37.2 / 16.4 / 13.7	0.76 / 0.95 / 0.91 / 0.67	0.18 / 0.12 / 0.25 / 0.07	41 / 86 / 76 / 34	−0.03 / −0.02 / −0.02 / −0.02

Block	Explained variance PC1 / PC2 / PC3 (%)	Cophenetic correlation	Adjusted Rand index (category; DMSO-d₆: residue group)	Largest cluster (% of residues)	Adjusted Rand index, wine-seasoned vs other (DMSO-d₆)
DMSO-Qa	37.2 / 14.6 / 14.1	0.73 / 0.91 / 0.84 / 0.64	0.22 / 0.12 / 0.13 / 0.10	38 / 86 / 52 / 31	−0.03 / −0.03 / −0.03 / 0.00

Partitions into seven clusters (D_2O/CD_3OD), five ($CDCl_3$, fused and DMSO-d₆ residue groups) or two (wine-seasoned vs other).

12.8 Linear and non-linear displays

The two-dimensional PCA displays preserved the neighborhoods of the scaled space with a trustworthiness between 0.83 and 0.96, t-SNE between 0.86 and 0.95 and UMAP between 0.81 and 0.93 (Table 22). For the large D_2O/CD_3OD blocks the embeddings were the more trustworthy displays (t-SNE 0.92 against PCA 0.85 for the non-carbohydrate block), whereas for the small DMSO-d₆ blocks the reverse held (PCA 0.93 against t-SNE 0.87 for the full block), because with twenty-nine residues and two dominant components a principal component plane already captures the neighborhoods, and an embedding computed with a perplexity of nine has little room to improve on it. The class purity of an embedding never exceeded that of the full scaled space by more than 0.03, while the two-dimensional PCA display always lost purity (0.28 against 0.50 for the non-carbohydrate block). The embeddings are therefore faithful displays of neighborhoods that already exist in the scaled data, and they are better displays than a principal component plane when the data set is large, but they add no class information, and the visually striking clusters of a t-SNE or UMAP plot must be interpreted with reference to the scaled space and to the chemistry, as in version 1.0 [57].

Table 22. Class purity of the *k* nearest neighbors and trustworthiness of the two-dimensional displays for every data approach.

Block	Classes	Chance purity	Purity, scaled space	Purity, PCA 2-D	Purity, t-SNE	Purity, UMAP	Trustworthiness, PCA 2-D	Trustworthiness, t-SNE	Trustworthiness, UMAP (± SD, 5 seeds)	Perplexity / neighbors
D2O-F	category (11)	0.15	0.47	0.26	0.43	0.45	0.85	0.93	0.91 ± 0.003	30 / 15
D2O-NC	category (11)	0.15	0.50	0.28	0.51	0.53	0.85	0.92	0.90 ± 0.004	30 / 15
D2O-NV	category (11)	0.15	0.50	0.28	0.50	0.52	0.85	0.93	0.90 ± 0.004	30 / 15
D2O-NC04	category (11)	0.15	0.50	0.27	0.50	0.53	0.85	0.93	0.90 ± 0.003	30 / 15
D2O-NCall	category (11)	0.14	0.49	0.25	0.49	0.51	0.83	0.92	0.90 ± 0.001	30 / 15
D2O-Qp	category (11)	0.15	0.43	0.35	0.43	0.41	0.86	0.95	0.93 ± 0.004	30 / 15
D2O-Qa	category (11)	0.15	0.44	0.35	0.43	0.43	0.87	0.93	0.92 ± 0.001	30 / 15
CDCl3	category (9)	0.31	0.46	0.34	0.48	0.47	0.88	0.90	0.88 ± 0.006	20 / 15
CDCl3-W	category (9)	0.31	0.48	0.45	0.47	0.49	0.95	0.93	0.89 ± 0.006	20 / 15
FUSED-DC	category (9)	0.35	0.55	0.49	0.53	0.50	0.89	0.93	0.82 ± 0.008	17 / 15
FUSED-W	category (9)	0.35	0.55	0.50	0.51	0.51	0.89	0.92	0.83 ± 0.009	17 / 15
DMSO-F	residue group (5)	0.28	0.43	0.37	0.36	0.35	0.93	0.87	0.84 ± 0.009	9 / 15
DMSO-Fx	residue group (5)	0.29	0.42	0.40	0.38	0.36	0.96	0.90	0.87 ± 0.006	9 / 15
DMSO-NC	residue group (5)	0.28	0.44	0.39	0.37	0.37	0.95	0.90	0.91 ± 0.003	9 / 15
DMSO-F04	residue group (5)	0.28	0.43	0.36	0.37	0.36	0.94	0.87	0.85 ± 0.007	9 / 15
DMSO-Qp	residue group (5)	0.28	0.42	0.33	0.40	0.33	0.88	0.87	0.81 ± 0.013	9 / 15
DMSO-Qa	residue group (5)	0.28	0.39	0.37	0.39	0.37	0.91	0.86	0.82 ± 0.025	9 / 15

$k = 10$ (9 for the DMSO-d₆ blocks). t-SNE with PCA initialization was deterministic, so its five seeds gave identical results.

12.9 The same products in different solvents

The 24 products with both a D_2O/CD_3OD and a DMSO-d₆ residue (set S1: four Hudson whiskeys from Tuthilltown, nineteen Kavalan malts and one Scotch malt) and the 10 products with residues in all three solvents (set S2) allow the configurations of the same products to be compared across solvents directly (Figure 57, Table 23). Each block was restricted to the matched products, Pareto-scaled and normalized to unit root-mean-square, and pairs of blocks were compared by the modified RV coefficient [70], by Procrustes analysis of the first three principal components with a permutation test [71,72], by the Mantel correlation between the distance matrices, by the correlation between the cophenetic distances of the two Ward trees and by the adjusted Rand index between their partitions.

In set S1 the non-carbohydrate D_2O/CD_3OD block and the full DMSO-d₆ block agreed substantially: modified RV 0.77, Procrustes m^2 0.43 ($p = 0.001$) and Mantel r 0.75 ($p = 0.001$). The two solvents therefore

see largely the same differences between these products. Their Ward trees, however, agreed poorly (cophenetic correlation 0.37; adjusted Rand index 0.31–0.27 for two to five clusters), which is a property of hierarchical clustering rather than of the solvents: small differences in the distances change the order in which clusters merge, and a tree amplifies them. The Procrustes superposition (Figure 57a) shows where the solvents differ most. The Hudson whiskeys and the Kavalan malts from ex-bourbon and other casks occupy the same region in both solvents, whereas the largest residuals belong to the wine-seasoned Kavalan malts (the Moscatel, oloroso 034, Madeira, fino and King Car casks), whose esters and lipids are seen fully in DMSO-d_6 and less in D_2O/CD_3OD and whose succinic acid is seen fully only in D_2O (Section 9.7).

Set S2 asks the sharper question. The configurations of the ten products in the D_2O/CD_3OD and $CDCl_3$ blocks were unrelated (modified RV 0.08; Procrustes $p = 0.282$; Mantel $p = 0.416$): the polar and lipophilic fractions of the same residues order the products differently, which is the reason for using two solvents in the first place. The DMSO-d_6 block shared structure with both (0.57 with D_2O/CD_3OD and 0.56 with $CDCl_3$) and reproduced the configuration of the fused two-solvent block to a modified RV of 0.77 (Mantel r 0.75). Adding DMSO-d_6 to the two-solvent pair changed little (modified RV 0.97 between the two-solvent and the three-solvent fusions), whereas the lipophilic information of $CDCl_3$ was not fully captured by the D_2O/CD_3OD and DMSO-d_6 pair (0.35). The comparison is consistent with what the spectra show directly (Section 9): DMSO-d_6 sees most of what each of the other solvents sees, but not all of it. With ten products, one producer contributing six of them, these coefficients carry a large sampling uncertainty, and they describe these products rather than spirits in general.

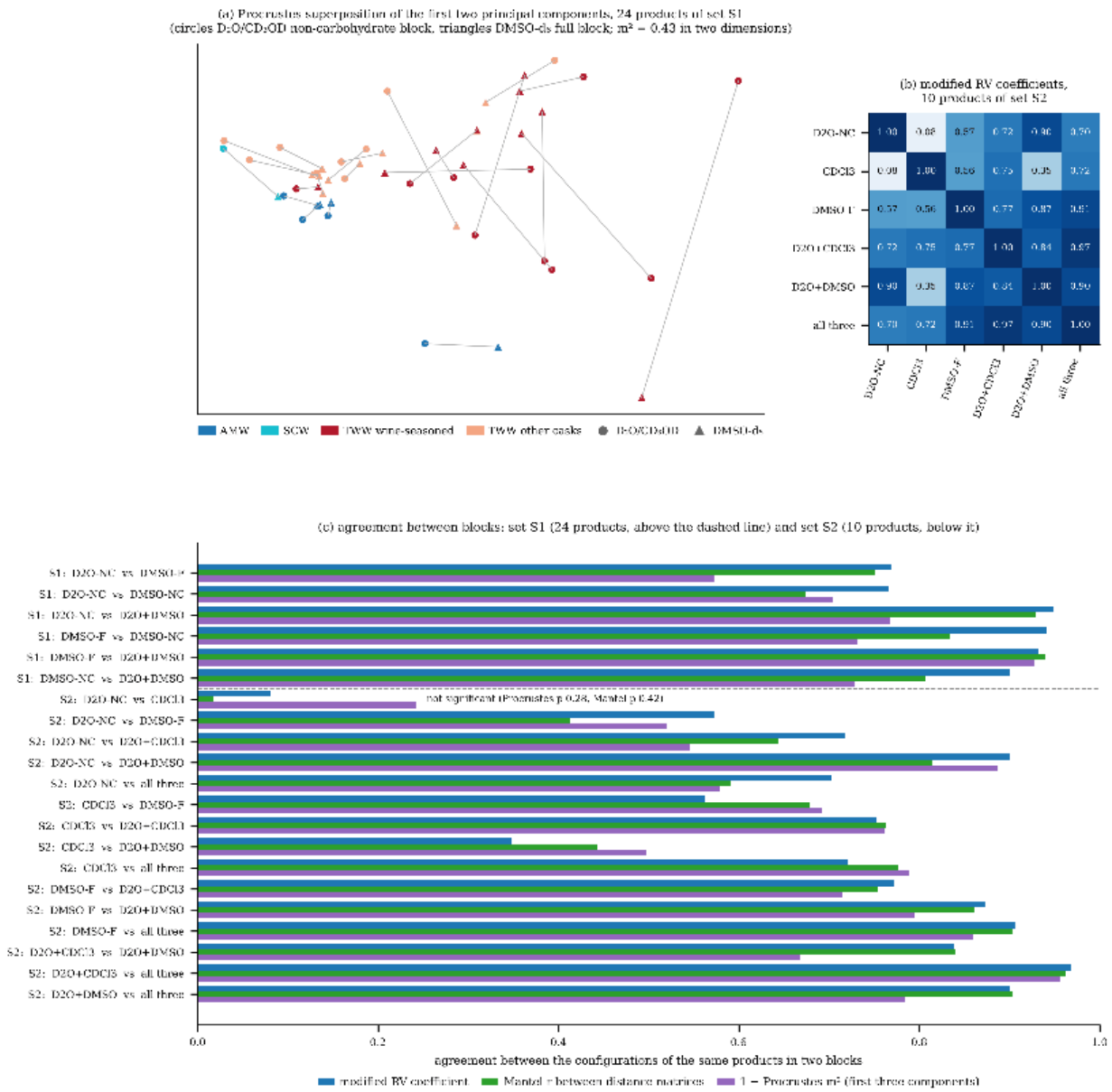

Figure 57. The same products in different solvents. (a) Procrustes superposition of the first two principal components of the D_2O/CD_3OD non-carbohydrate block (circles) and the full DMSO-d_6 block (triangles) for the 24 products of set S1, with a line joining the two positions of each product. (b) Modified RV coefficients between the single-solvent blocks and their fusions for the 10 products of set S2. (c) Modified RV coefficient, Mantel correlation and one minus the Procrustes m^2 (first three components) for every pair of blocks in sets S1 and S2.

Table 23. Agreement between the configurations of the same products in different solvents: set S1 (24 products with D_2O/CD_3OD and DMSO-d_6 residues) and set S2 (10 products with residues in all three solvents).

Set	Blocks compared	Modified RV	Procrustes m^2 (p)	Mantel r (p)	Correlation of Ward cophenetic distances	Adjusted Rand index, 2 / 3 / 4 / 5 clusters
S1	D2O-NC vs DMSO-F	0.77	0.43 (0.001)	0.75 (0.001)	0.37	0.31 / 0.16 / 0.14 / 0.27
S1	D2O-NC vs DMSO-NC	0.77	0.30 (0.001)	0.67 (0.001)	0.36	0.22 / 0.32 / 0.20 / 0.19
S1	D2O-NC vs D2O+DMSO	0.95	0.23 (0.001)	0.93 (0.001)	0.50	0.42 / 0.26 / 0.28 / 0.36
S1	DMSO-F vs DMSO-NC	0.94	0.27 (0.001)	0.83 (0.001)	0.59	0.54 / 0.42 / 0.47 / 0.37
S1	DMSO-F vs D2O+DMSO	0.93	0.07 (0.001)	0.94 (0.001)	0.83	0.83 / 0.76 / 0.68 / 0.80
S1	DMSO-NC vs D2O+DMSO	0.90	0.27 (0.001)	0.81 (0.001)	0.68	0.68 / 0.49 / 0.65 / 0.45
S2	D2O-NC vs CDCl3	0.08	0.76 (0.282)	0.02 (0.416)	0.26	0.10 / 0.21 / 0.20 / 0.60

Set	Blocks compared	Modified RV	Procrustes m^2 (p)	Mantel r (p)	Correlation of Ward cophenetic distances	Adjusted Rand index, 2 / 3 / 4 / 5 clusters
S2	D2O-NC vs DMSO-F	0.57	0.48 (0.007)	0.41 (0.035)	0.64	0.60 / 0.52 / 0.59 / 0.55
S2	D2O-NC vs D2O+CDCl3	0.72	0.45 (0.001)	0.64 (0.001)	0.95	1.00 / 1.00 / 1.00 / 1.00
S2	D2O-NC vs D2O+DMSO	0.90	0.11 (0.001)	0.82 (0.001)	0.98	1.00 / 0.70 / 1.00 / 1.00
S2	D2O-NC vs all three	0.70	0.42 (0.001)	0.59 (0.002)	0.98	1.00 / 1.00 / 1.00 / 0.50
S2	CDCl3 vs DMSO-F	0.56	0.31 (0.002)	0.68 (0.006)	0.46	0.30 / 0.08 / 0.18 / 0.70
S2	CDCl3 vs D2O+CDCl3	0.75	0.24 (0.001)	0.76 (0.001)	0.39	0.10 / 0.21 / 0.20 / 0.60
S2	CDCl3 vs D2O+DMSO	0.35	0.50 (0.004)	0.44 (0.022)	0.27	0.10 / 0.04 / 0.20 / 0.60
S2	CDCl3 vs all three	0.72	0.21 (0.001)	0.78 (0.001)	0.35	0.10 / 0.21 / 0.20 / 0.36
S2	DMSO-F vs D2O+CDCl3	0.77	0.28 (0.001)	0.75 (0.001)	0.66	0.60 / 0.52 / 0.59 / 0.55
S2	DMSO-F vs D2O+DMSO	0.87	0.20 (0.001)	0.86 (0.001)	0.66	0.60 / 0.62 / 0.59 / 0.55
S2	DMSO-F vs all three	0.91	0.14 (0.001)	0.90 (0.001)	0.67	0.60 / 0.52 / 0.59 / 0.31
S2	D2O+CDCl3 vs D2O+DMSO	0.84	0.33 (0.001)	0.84 (0.001)	0.92	1.00 / 0.70 / 1.00 / 1.00
S2	D2O+CDCl3 vs all three	0.97	0.04 (0.001)	0.96 (0.001)	0.99	1.00 / 1.00 / 1.00 / 0.50
S2	D2O+DMSO vs all three	0.90	0.22 (0.001)	0.90 (0.001)	0.97	1.00 / 0.70 / 1.00 / 0.50

Each block Pareto-scaled and divided by its root-mean-square; fused blocks concatenated after that normalization. Procrustes on the first three principal components, 999 permutations; Mantel on Euclidean distances, 999 permutations [70–72].

12.10 Separation of wine-seasoned casks with one solvent and with two

The wine-seasoned casks are the only contrast represented well enough in the matched sets to test classification (Figure 58, Table 24). In set S1 (11 wine-seasoned and 13 other casks), leave-one-out PLS-DA with two latent variables reached a balanced accuracy of 0.78 on the non-carbohydrate D_2O/CD_3OD block, 0.87 on the full DMSO-d_6 block, 0.83 on the non-carbohydrate DMSO-d_6 block and 0.87 on their fusion, and the linear SVM 0.83, 0.83, 0.83 and 0.87, every result being significant in a test with 499 label permutations ($p \leq 0.006$). The single DMSO-d_6 block did as well as the D_2O/CD_3OD block and as well as the fusion; the differences between the approaches amount to one residue in twenty-four and are not significant. The silhouette widths (0.26–0.28) [73] show that the two classes overlap in the scaled space even though a linear boundary separates most of them, which is the typical situation of a high-dimensional, small-sample problem [46].

In set S2 (five wine-seasoned and five other casks) the D_2O/CD_3OD block classified every product correctly ($p = 0.018$), as did every fusion that included it, whereas $CDCl_3$ alone did not separate the classes significantly (PLS-DA 0.70, $p = 0.132$; SVM 0.80, $p = 0.142$) and DMSO-d_6 alone misclassified one product in ten (PLS-DA 0.90, $p = 0.040$). These results should not be over-read. With ten products a leave-one-out accuracy of 1.00 has a permutation p-value of only 0.008–0.018, because many label permutations of a five-against-five problem can also be classified perfectly in several hundred dimensions, and the classes are partly confounded with the producer, four of the five wine-seasoned casks being Kavalan malts and three of the five others Tuthilltown whiskeys. The matched sets show that none of the solvents contradicts the separation of wine-seasoned casks; they do not establish a classifier, and they cannot rank the solvents for that purpose [74,75].

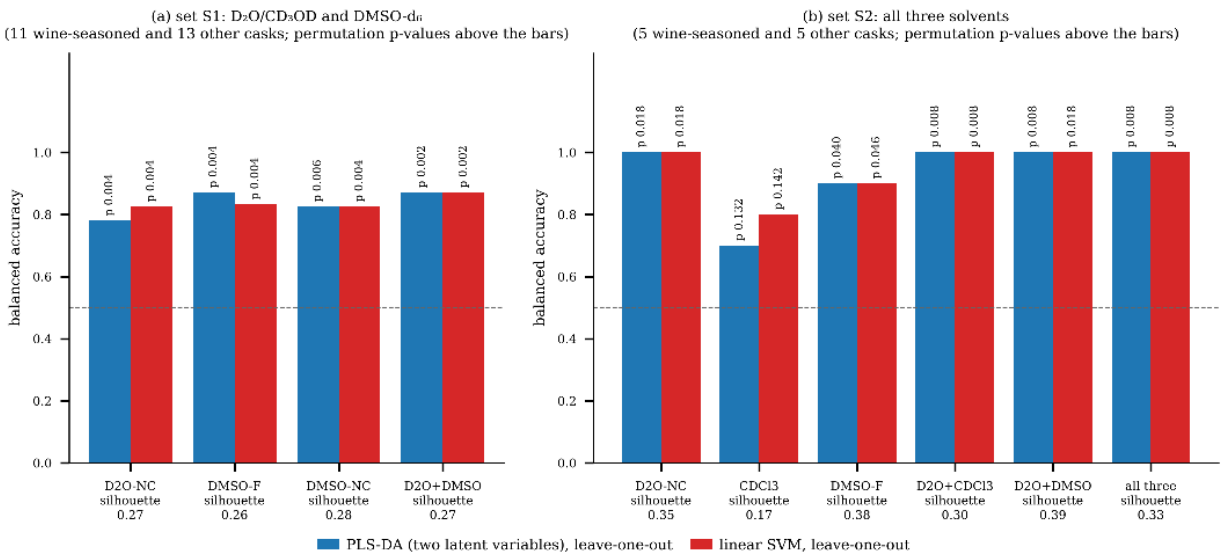

Figure 58. Separation of wine-seasoned from other casks in the matched sets: leave-one-out balanced accuracy of PLS-DA with two latent variables and of a linear SVM, with the permutation *p*-values (499 permutations) above the bars and the silhouette width of each block beneath it; the dashed line marks chance.

Table 24. Separation of wine-seasoned from other casks in the matched sets: silhouette width and leave-one-out balanced accuracy with permutation *p*-values (499 permutations).

Set	Block	Wine-seasoned / all	Silhouette	PLS-DA, 2 LV (*p*)	Linear SVM (*p*)
S1	D2O-NC	11 / 24	0.27	0.78 (0.004)	0.83 (0.004)
S1	DMSO-F	11 / 24	0.26	0.87 (0.004)	0.83 (0.004)
S1	DMSO-NC	11 / 24	0.28	0.83 (0.006)	0.83 (0.004)
S1	D2O+DMSO	11 / 24	0.27	0.87 (0.002)	0.87 (0.002)
S2	D2O-NC	5 / 10	0.35	1.00 (0.018)	1.00 (0.018)
S2	CDCl3	5 / 10	0.17	0.70 (0.132)	0.80 (0.142)
S2	DMSO-F	5 / 10	0.38	0.90 (0.040)	0.90 (0.046)
S2	D2O+CDCl3	5 / 10	0.30	1.00 (0.008)	1.00 (0.008)
S2	D2O+DMSO	5 / 10	0.39	1.00 (0.008)	1.00 (0.018)
S2	all three	5 / 10	0.33	1.00 (0.008)	1.00 (0.008)

Block scaling recomputed in every training set. With ten products the smallest attainable permutation p-values for perfect classification were 0.008–0.018.

12.11 What the accounting establishes

Taken together, the analyses of this section support five conclusions about the chemometrics of absolute ¹H spectra of freeze-dried spirits, independently of the particular products studied. The choice of linkage and distance changed the result of hierarchical clustering more than the choice of data block, and a dendrogram in an authentication report should always be accompanied by its linkage, distance, scaling, cophenetic correlation, agreement with known classes and the size of its largest cluster. The non-linear embeddings are trustworthy displays of the scaled space and are preferable to a principal component plane for large data sets, but they carry no more class information than the scaled space itself. On absolute intensities with Pareto scaling, the leading components of every block are controlled by the few residues of extreme composition, the sweetened, flavored and wine-seasoned products, which is informative for screening but hides the structure of the bulk; sensitivity analyses that remove or down-weight those residues, or logarithmic and correlation-based measures that compare profiles rather than amounts, should accompany any model whose purpose is to describe the typical product. Artifacts of sample preparation,

here moisture in two $CDCl_3$ residues, can produce principal components of their own, and inspection of the loadings is the cheapest protection against interpreting them. Finally, on the matched products, DMSO-d_6 alone reproduced most of the configuration that requires two solvents and separated wine-seasoned casks as well as the D_2O/CD_3OD block, but with twenty-four and ten products, most of them from one producer, the comparison is suggestive rather than conclusive; Section 15 weighs it together with the spectroscopic evidence.

13. PLS regression calibrations

13.1 Design and a necessary caution

PLS1 models were built between the absolute-intensity binned D_2O/CD_3OD spectra (Pareto-scaled, 183 residues; 141 for glycerol, Section 13.3) and each quantified congener, with the number of latent variables chosen by leave-10 %-out cross-validation (ten folds, grouped by product and repeated five times) as the smallest number whose root-mean-square error of cross-validation lay within one standard error of the minimum [50]. The results are collected in Table 25 and illustrated in Figure 59. The reference values are those of the re-quantitation of this version: glucose is the corrected value of Section 9.5, which in version 2.0 included the β-arabinopyranose doublet and missed part of the displaced β-glucose doublets, and the fructose and sucrose of the two residues dissolved in equal volumes of D_2O and CD_3OD were refitted at their measured positions (Section 4.8). The calibration statistics changed little; the largest change was in the logarithmic glycerol model (Q^2 = 0.56, against 0.63 in version 2.0), whose reference values for the sugar-rich residues depend on the two-multiplet procedure of Section 6.5 and therefore on the refitted fructose.

Table 25. PLS1 calibrations of the absolute binned D_2O/CD_3OD spectra against the quantified congeners (leave-10 %-out, product-grouped cross-validation, five repeats; latent variables by the one-standard-error rule).

Analyte	*n*	LV	Q^2	RMSECV (mg L^{-1})	ρ (CV)	Masked Q^2	log scale: LV / Q^2
Glucose	183	11	0.994	112	0.94	0.94	3 / 0.76
Fructose	183	14	0.996	92.7	0.91	0.95	3 / 0.72
Sucrose	183	12	0.997	156	0.51	0.94	1 / 0.56
Glycerol	141	11	0.979	20.0	0.89	0.57	4 / 0.56
Lactic acid	183	12	0.937	7.4	0.83	0.05	7 / 0.32
Succinic acid	183	15	0.869	4.2	0.79	0.73	11 / 0.69
Acetic acid*	183	1	−0.004	2.0	0.30	0.00	—
Quercitol	183	13	0.908	11.5	0.96	0.69	7 / 0.60
Gallic acid	183	8	0.685	3.1	0.80	0.62	6 / 0.44
Benzaldehydes	183	4	0.174	10.2	0.77	0.15	6 / 0.58
Syringyl index	183	7	0.456	5.5	0.63	0.41	—
HMF	183	15	0.737	3.6	0.84	0.52	5 / 0.42
Cinnamaldehydes	183	11	0.552	3.8	0.75	0.44	14 / 0.40
1,2-Propanediol	183	14	0.986	13.9	0.83	0.47	3 / 0.52
Ethanol*	183	11	0.990	49.2	0.66	0.84	—
Fusel alcohols*	183	14	0.899	4.9	0.95	0.57	—

Q^2, cross-validated coefficient of determination; ρ, Spearman correlation between cross-validated predictions and reference values; masked, model rebuilt without the bins containing the analyte's resonances. *Semi-volatile. Glycerol: residues in which it was not quantifiable excluded.

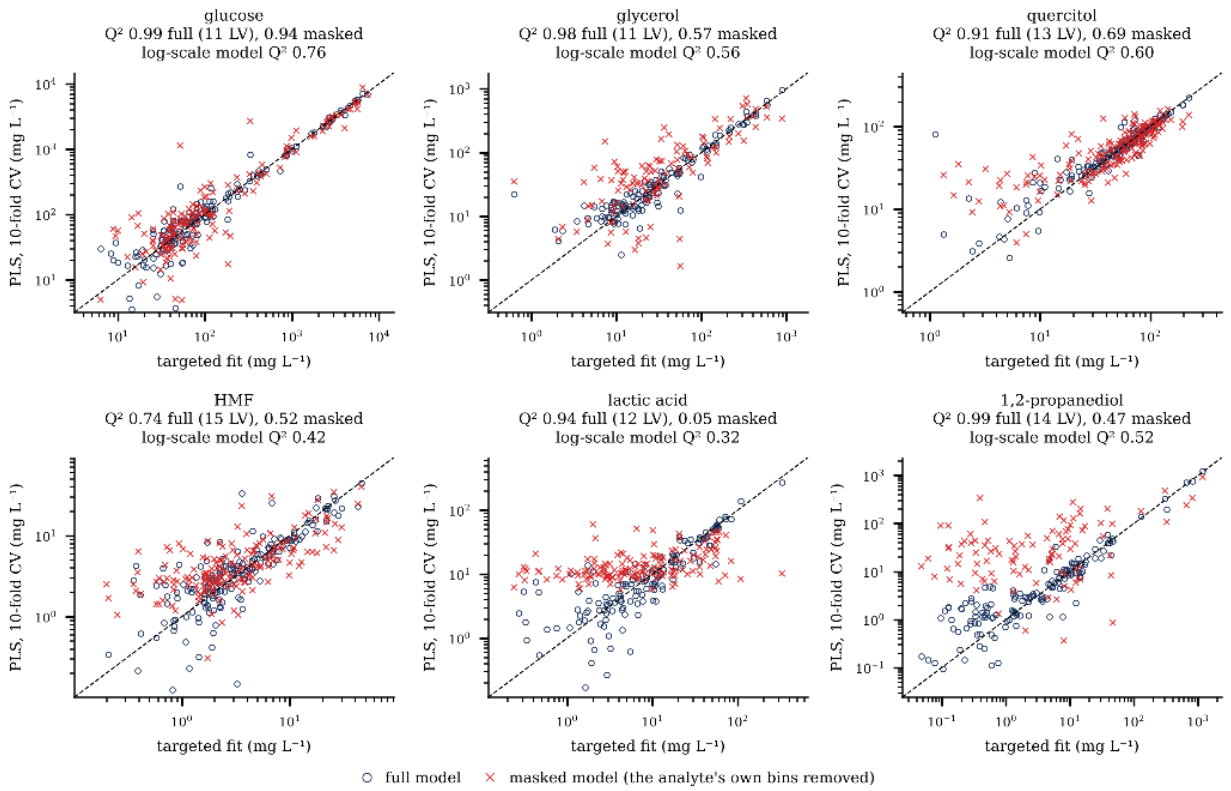

Figure 59. PLS1 calibrations of the binned absolute D_2O/CD_3OD spectra against the targeted-fit concentrations for six congeners: cross-validated predictions (leave-10 %-out, grouped by product) of the full models (circles) and of the masked models from which the analyte's own signals were removed (crosses), with the Q^2 of the full, masked and logarithmic-scale models.

One property of these models must be stated before any figure is quoted, because a reviewer would seize on it at once. The reference values were obtained from the same spectra by targeted fitting, so a PLS model that predicts them is, to a large extent, learning to integrate the signal that was integrated to produce the reference. High cross-validated coefficients of determination are therefore expected and do not demonstrate accuracy; they demonstrate that the targeted fits and the binned spectra are mutually consistent and that a multivariate calibration can reproduce the targeted result without the fitting step. Accuracy, in the sense of agreement with the true concentration, can only be established against an independent reference method, which for these analytes would be HPLC with UV or refractive-index detection, enzymatic sugar and glycerol assays, or quantitative ^{13}C NMR. Three devices were used to extract as much independent information as possible from the present data: models on logarithmic concentrations, which prevent a handful of extreme residues from dominating the statistics; masked models, from which every bin that contains a signal of the analyte was removed; and cross-solvent models, in which the $CDCl_3$ spectrum of one aliquot was used to predict a concentration measured in the D_2O spectrum of another aliquot of the same bottle, and the reverse.

13.2 Calibration statistics

On a linear concentration scale the cross-validated Q^2 exceeded 0.9 for the sugars, 1,2-propanediol, ethanol, the fusel alcohols, glycerol, lactic acid and quercitol (Table 25). These values are inflated by the long tails of the concentration distributions. The models selected 11–15 latent variables for most analytes, which is far more than the chemistry requires and is the signature of models that spend their latent variables fitting

the few extreme residues, and the Spearman rank correlations between cross-validated predictions and reference values, which are insensitive to the tails, were considerably lower for several of them (for example 0.51 for sucrose and 0.66 for ethanol). On a logarithmic scale, where every residue counts equally, the Q^2 values fell to 0.32–0.76, with one to eight latent variables for every analyte except succinic acid. These are the figures that describe how well a binned spectrum predicts the concentration of a congener across the normal range of products, and they are adequate for screening (ranking products, flagging outliers) but not for quantitation in the regulatory sense, which should continue to rely on the targeted fits.

13.3 What the masked models reveal

Removing every bin that contains a signal of the analyte tests whether the concentration can be predicted from other compounds that co-vary with it. The results divide the analytes into two groups. Fructose (0.95), glucose (0.94) and sucrose (0.94) remained highly predictable after masking, because sweetening introduces sugars together (invert sugar couples glucose and fructose, and sucrose-sweetened products also carry retained ethanol, whose amount follows the sugar), and glycerol (0.57) and quercitol (0.69) retained part of their predictability because they co-vary with the wine-cask and oak extract as a whole. Lactic acid (0.05) and the benzaldehydes (0.15) lost almost all of their predictability, and 1,2-propanediol (0.47) and HMF (0.52) a large part of it, showing that their calibrations rest mainly on their own signals. For screening, this distinction matters: a model for a co-varying analyte can appear to work on a calibration set and fail on a product in which the usual co-variation is broken, which is precisely the situation of an adulterated or atypical product. A glycerol model that is carried by the wine-cask pattern would miss glycerin added to a whiskey from a non-wine cask, and a quercitol model carried by oak extract would overestimate quercitol in a product colored and flavored with an oak extract that lacks it. Calibrations for authentication should therefore be built on the analyte's own signals, as the targeted fits are.

The glycerol calibration supplies a cautionary example that even the masked models would not have exposed. With the first-pass glycerol values, in which the fructose and glucose signals of the sugar-rich residues had been fitted as glycerol (Section 6.5), the PLS model reached $Q^2 = 0.95$ and its masked counterpart 0.94, because the reference values were themselves largely a measure of sugar (Spearman $\rho = 0.86$ with total sugar) and the model learned the sugar signals elsewhere in the spectrum. With the corrected values, and with the residues in which glycerol could not be quantified excluded, the full model gave $Q^2 = 0.98$ but the masked model only 0.57. A high cross-validated Q^2, even one that survives masking of the analyte's own window, demonstrates consistency with the reference values and nothing more; when the reference values are wrong in a way that is correlated with another constituent, the calibration inherits the error and validates it.

13.4 Cross-solvent prediction

The cross-solvent models are the closest approach to independent validation that the present data allow, because the two residues were prepared from different aliquots and measured in different solvents. Using the D_2O/CD_3OD spectrum to predict the $CDCl_3$ concentrations of syringaldehyde, vanillin and the cinnamaldehydes gave Q^2 values of 0.21, 0.29 and 0.35 (Spearman $\rho = 0.48$, 0.65 and 0.62) for the 51 paired products, and the $CDCl_3$ spectrum predicted the D_2O cinnamaldehyde total with $Q^2 = 0.67$ but failed for the D_2O benzaldehyde total ($Q^2 = -0.07$), although the targeted benzaldehyde results of the two solvents are well correlated (Section 7.4), probably because the variance of the Pareto-scaled $CDCl_3$ spectra is dominated by the lipids; the negative controls (glucose and quercitol, which are not soluble in chloroform) gave $Q^2 \leq 0.09$, as they should. The cross-solvent agreement is real but modest, and it is limited by the incomplete recovery of the aldehydes in chloroform (Section 7.4), by the small number of pairs and by the different preparations of the two aliquots. It shows that the absolute D_2O spectrum contains the aldehyde

information measured in $CDCl_3$, but not that either solvent alone is an accurate measure of the aldehyde content of the original spirit.

13.5 Second round: calibrations after the removal of outliers

The models of Sections 13.1–13.4 were built on every residue for which a reference value existed. A second round of every model was built after the removal of two kinds of outlier, in a single pass and with the cross-validation, the scaling and the selection of latent variables unchanged (Table 26, Figure 60 and Figure 61). Flagged residues were removed before any model was fitted, on grounds that do not depend on the calibration: reference values that the targeted quantitation had itself marked as unreliable (72 glycerol values that are upper bounds because the glycerol window is crowded, six further glycerol values that were not quantifiable or only upper limits, four fructose values affected by the tail of the sucrose fructosyl H-4 multiplet and five glucose values that had entered round 1 as zero because neither anomeric doublet could be used), and the two residues dissolved in equal volumes of D_2O and CD_3OD, whose lines are displaced by about 0.03 ppm, more than the width of a bin, so that their binned spectra do not correspond to those of the other residues. Diagnosed residues were then identified from the round-1 model itself. The cross-validated residual of each residue was standardized robustly, by a scale that grows with concentration for the linear models (fitted to the residuals by least absolute deviations) and by the median absolute deviation for the logarithmic models, and a residue was removed if its standardized residual exceeded 3.5 in magnitude (a vertical outlier) or exceeded 2.5 while its Hotelling T^2 or its Q residual lay beyond the 99 % limit of the calibration (a bad leverage point) [50]. Residues beyond the T^2 or Q limit whose predictions were good, 3–19 per model (median 10), were kept, because such good leverage points extend a calibration rather than distort it.

Table 26. Second round of PLS1 calibrations after the removal of flagged and diagnosed outliers, with the same cross-validation as in round 1 (leave-10 %-out, product-grouped, five repeats; latent variables by the one-standard-error rule). Each triple gives round 1 / the round-1 cross-validated predictions scored on the residues retained in round 2 / round 2.

Analyte or cross-solvent model	Excluded: flagged + diagnosed	LV	Q^2	RMSECV (mg L^{-1})	ρ (CV)	Masked Q^2	log scale Q^2 (excluded)
Glucose	7 + 8	11 → 11	0.994 / 0.996 / 0.992	112 / 86.6 / 122	0.94 → 0.94	0.94 → 0.96	0.76 / 0.82 / 0.80 (14)
Fructose	6 + 12	14 → 9	0.996 / 0.997 / 0.993	92.7 / 80.2 / 106	0.91 → 0.89	0.95 → 0.94	0.72 / 0.76 / 0.73 (8)
Sucrose	2 + 26	12 → 9	0.997 / 0.998 / 0.997	156 / 136 / 166	0.51 → 0.54	0.94 → 0.76	0.56 / 0.61 / 0.62 (5)
Glycerol	78 + 0	11 → 5	0.979 / 0.979 / 0.966	20.0 / 25.2 / 29.9	0.89 → 0.95	0.57 → 0.13	0.56 / 0.67 / 0.80 (81)
Lactic acid	2 + 7	12 → 12	0.937 / 0.955 / 0.948	7.4 / 6.7 / 7.2	0.83 → 0.80	0.05 → 0.06	0.32 / 0.56 / 0.61 (8)
Succinic acid	2 + 10	15 → 10	0.869 / 0.949 / 0.906	4.2 / 2.8 / 3.2	0.79 → 0.77	0.73 → 0.90	0.69 / 0.75 / 0.66 (7)
Acetic acid*	2 + 3	1 → 1	−0.004 / 0.035 / 0.044	2.0 / 1.9 / 1.9	0.30 → 0.37	0.00 → 0.04	—
Quercitol	2 + 13	13 → 15	0.908 / 0.979 / 0.973	11.5 / 6.7 / 7.2	0.96 → 0.99	0.69 → 0.75	0.60 / 0.73 / 0.75 (6)
Gallic acid	2 + 11	8 → 14	0.685 / 0.764 / 0.823	3.1 / 2.6 / 2.3	0.80 → 0.90	0.62 → 0.69	0.44 / 0.61 / 0.57 (10)
Benzaldehydes	2 + 11	4 → 10	0.174 / 0.693 / 0.818	10.2 / 4.8 / 3.5	0.77 → 0.91	0.15 → 0.72	0.58 / 0.72 / 0.75 (10)
Syringyl index	2 + 13	7 → 8	0.456 / 0.685 / 0.708	5.5 / 4.3 / 4.0	0.63 → 0.80	0.41 → 0.67	—
HMF	2 + 7	15 → 12	0.737 / 0.921 / 0.849	3.6 / 2.1 / 3.3	0.84 → 0.80	0.52 → 0.49	0.42 / 0.54 / 0.55 (4)
Cinnamaldehydes	2 + 10	11 → 5	0.552 / 0.712 / 0.375	3.8 / 3.1 / 4.6	0.75 → 0.72	0.44 → 0.29	0.40 / 0.67 / 0.52 (9)
1,2-Propanediol	2 + 9	14 → 8	0.986 / 0.989 / 0.984	13.9 / 12.5 / 22.6	0.83 → 0.88	0.47 → 0.64	0.52 / 0.53 / 0.47 (4)
Ethanol*	2 + 17	11 → 13	0.990 / 0.999 / 0.999	49.2 / 16.4 / 17.8	0.66 → 0.78	0.84 → 0.97	—
Fusel alcohols*	2 + 12	14 → 12	0.899 / 0.984 / 0.983	4.9 / 1.0 / 1.0	0.95 → 0.96	0.57 → 0.55	—
D_2O → $CDCl_3$ syringaldehyde	1 + 0	3 → 4	0.211 / 0.201 / 0.419	5.0 / 5.0 / 4.6	0.48 → 0.56	—	—
D_2O → $CDCl_3$ vanillin	1 + 2	6 → 4	0.293 / 0.480 / 0.438	1.6 / 1.3 / 1.3	0.65 → 0.56	—	—
D_2O → $CDCl_3$ cinnamaldehydes	1 + 0	3 → 4	0.352 / 0.347 / 0.446	3.3 / 3.3 / 3.1	0.62 → 0.65	—	—
$CDCl_3$ → D_2O benzaldehydes	1 + 4	1 → 2	−0.065 / 0.287 / 0.356	11.2 / 4.6 / 4.3	0.42 → 0.65	—	—
$CDCl_3$ → D_2O cinnamaldehydes	1 + 1	3 → 7	0.670 / 0.684 / 0.695	3.7 / 3.7 / 3.7	0.73 → 0.73	—	—

Flagged: reference values marked by the targeted quantitation as upper bounds, upper limits, not quantifiable or contaminated, and the two equal-volume D_2O/CD_3OD residues. Diagnosed: $|t| > 3.5$ (vertical outliers) or $|t| > 2.5$ outside the 99 % limit of Hotelling T^2 or of the Q residual (bad leverage points), t being the robust standardized cross-validated residual of round 1. The middle value of each triple isolates the effect of not scoring the excluded residues; the change from the middle to the last value is the effect of refitting. RMSECV is the mean over the five repeats; it is not comparable between rounds when the excluded residues change the concentration range (glycerol). *Semi-volatile.

Over the 28 models built on the D_2O/CD_3OD spectra (16 on linear and twelve on logarithmic concentrations), 452 exclusions were made, 226 flagged and 226 diagnosed, involving 129 different residues, 86 of which were excluded from no more than two models. The residues excluded most often were the two equal-volume residues (Larsen Aqua Ignis cognac and Rum 28, 26 models each), the Barley Creek honey whiskey (19), Bacardi Reserva Limitada (17), Roatan rum (16), Tromba Reposado (13) and the Taconic maple bourbon (12), that is, sweetened or flavored products whose composition lies outside that of the other residues in several respects at once. They are outliers to these calibrations, which is not the same as being outliers in any analytical sense, and their targeted results stand.

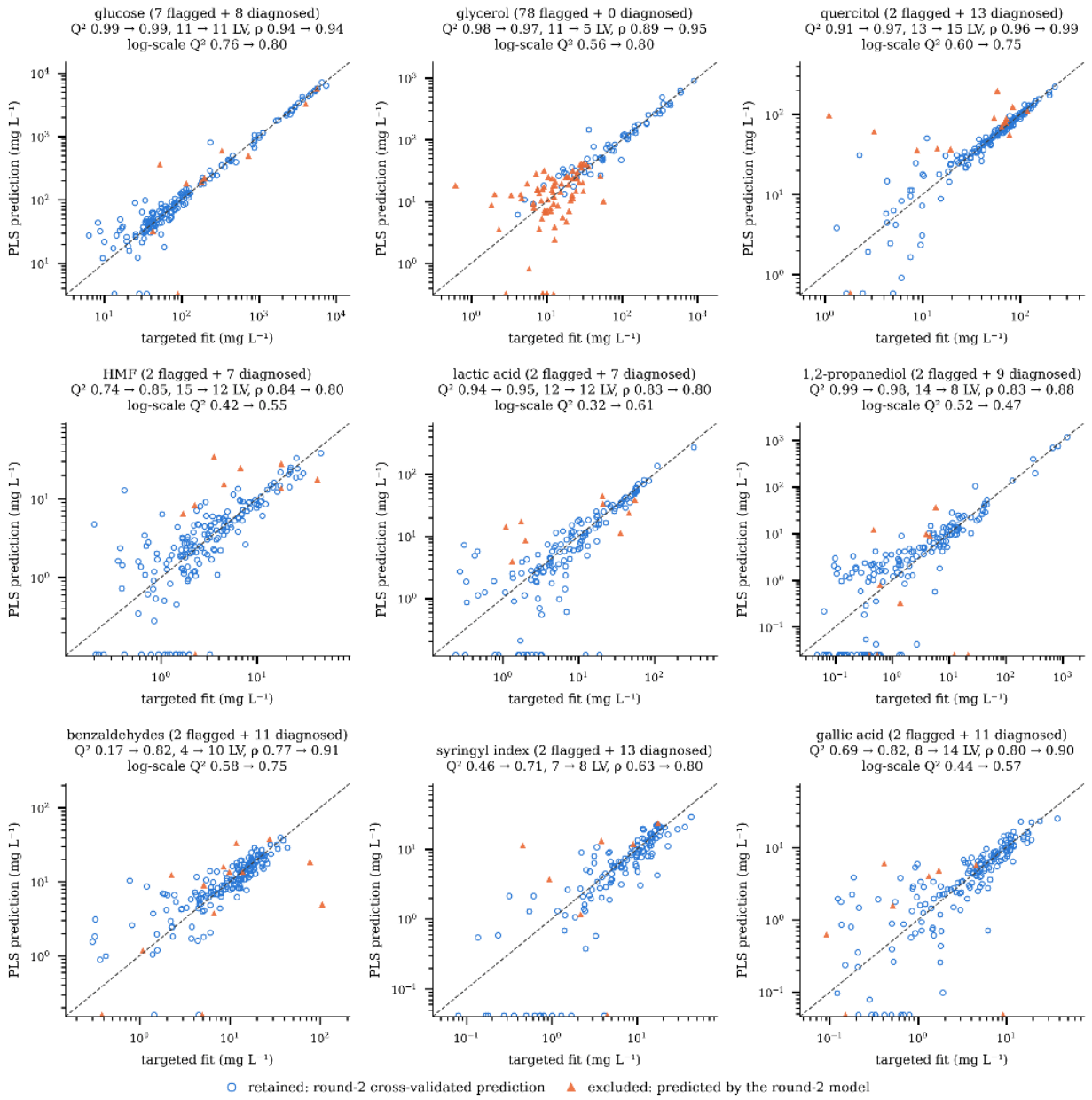

Figure 60. Second-round PLS1 calibrations for nine congeners on logarithmic axes: round-2 cross-validated predictions of the retained residues (blue circles) and predictions of the excluded residues by the round-2 model, which was built without them (orange triangles), against the targeted-fit concentrations; predictions at or below zero are drawn at the lower edge of each panel. Each title gives the numbers of flagged and diagnosed exclusions and the change of Q^2, of the number of latent variables, of the Spearman ρ and of the logarithmic-scale Q^2 from round 1 to round 2.

Because the excluded residues are, by construction, those that were predicted worst, removing them improves every statistic computed on the remainder whether or not the refitted model is better. The round-1 cross-validated predictions were therefore also scored on the retained residues alone, the middle value of each triple in Table 26, which separates the two effects (Figure 61): the step from the first to the middle value is the effect of no longer scoring the excluded residues, and the step from the middle to the last value is the effect of refitting without them. Judged in this way, most of the apparent improvement is of the first kind. For the sugars the refitted models were worse than round 1 even on the retained residues; the

RMSECV of glucose rose from 86.6 to 122 mg L⁻¹ and that of fructose from 80.2 to 106 mg L⁻¹, because the residues removed from them, the heavily sweetened products, are not errors but the sparse upper end of a real distribution, and a model deprived of them extrapolates less well to the sweetened residues that remain. The same holds for sucrose (136 → 166 mg L⁻¹), 1,2-propanediol (12.5 → 22.6 mg L⁻¹) and the cinnamaldehydes (Q^2 0.71 → 0.38). HMF shows the trap most clearly: its Q^2 rose from 0.74 in round 1 to 0.85 in round 2, but the round-1 predictions scored on the retained residues already gave 0.92, so the refit itself made the model worse.

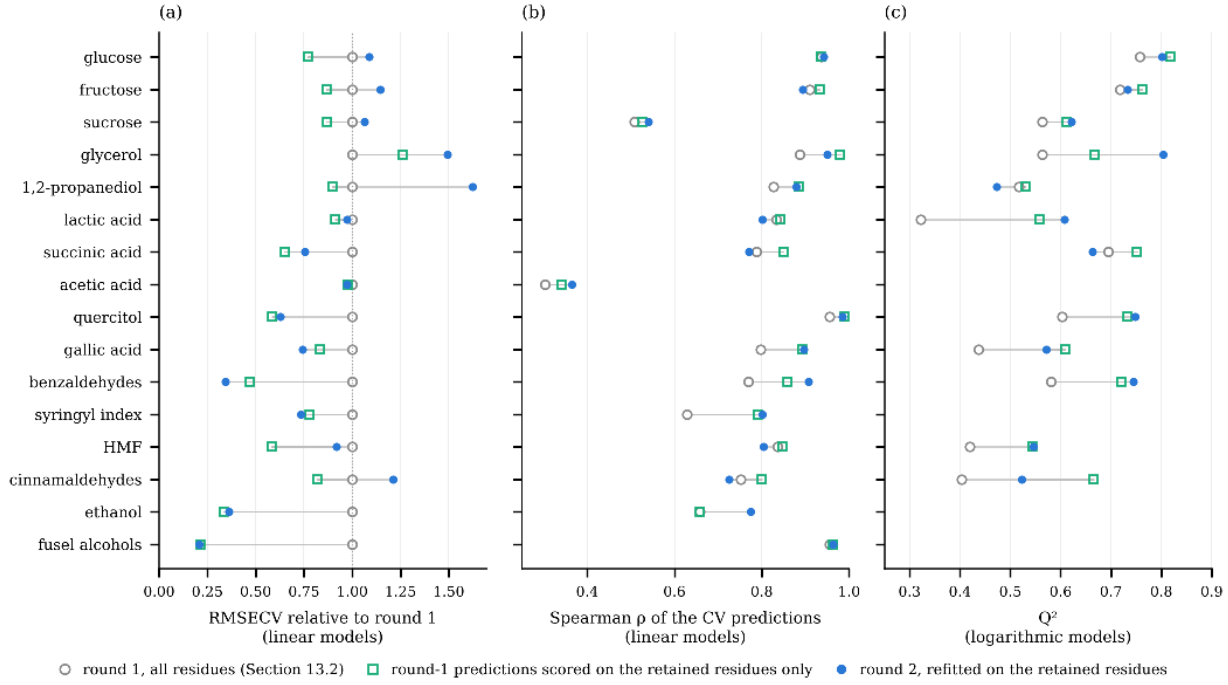

Figure 61. Decomposition of the change between the two rounds for the sixteen D_2O/CD_3OD analytes: (a) RMSECV of the linear models relative to round 1, (b) Spearman ρ of the cross-validated predictions of the linear models and (c) Q^2 of the logarithmic models (twelve analytes). Gray open circles, round 1 on all residues (Section 13.2); green squares, the round-1 cross-validated predictions scored on the residues retained in round 2; blue circles, round 2 refitted on the retained residues. The step from gray to green is the effect of not scoring the excluded residues, that from green to blue the effect of refitting.

Genuine improvements on refitting were confined to the oak aromatics and to two of the logarithmic models. The benzaldehyde calibration gained most: its Q^2 rose from 0.17 to 0.69 on rescoring and to 0.82 on refitting, its RMSECV fell from 10.2 to 4.8 and then 3.5 mg L⁻¹, its Spearman ρ rose from 0.77 to 0.91 and its masked Q^2 from 0.15 to 0.72. In round 1 the RMSECV curve of this model reached its minimum at 4 latent variables and rose beyond it, because further components were spent on the 11 residues later diagnosed as outliers; without them the model used 10 latent variables to advantage. The syringyl index (0.68 → 0.71, with ρ rising from 0.63 to 0.80) and gallic acid (0.76 → 0.82) gained in the same way but less, and gallic acid lost on the logarithmic scale (0.61 → 0.57). On the logarithmic scale the refit improved glycerol (Q^2 0.67 → 0.80, with the typical error factor falling from 2.04 to 1.57) and, less, lactic acid (0.56 → 0.61).

The glycerol models need a separate caution. The flagged residues of this analyte are the 72 residues whose glycerol values are upper bounds, which are the residues with little glycerol, so that round 2 was built on 63 residues spanning 4.1–890 mg L⁻¹ (10th percentile 16 mg L⁻¹). The linear model, which is carried by the upper part of that range, predicted the excluded residues at a median of 1.02 times their upper-bound values,

which is possible, but the logarithmic model, whose apparent improvement is the largest of all, predicted them at 2.2 times those values, which is impossible, since an upper bound cannot lie below the true value. The cleaned logarithmic glycerol model is therefore valid only within the range on which it was built, above about 16 mg L^{-1}, and must not be applied to residues with less glycerol, which is exactly where the doubtful reference values lie.

The diagnosed residues were not artifacts of round 1. Predicted by the round-2 models, which had not seen them, they gave errors a median of 4.3 times the RMSECV of the retained residues (0.7–22 over the 24 models with at least three diagnosed exclusions; a median of 4.6 on the logarithmic scale, where the comparison does not depend on concentration), so that they remain poorly predicted however the model is built. The exception was 1,2-propanediol, whose diagnosed residues were predicted by the refitted model about as well as the retained ones (ratio 0.7) and whose round-2 model is accordingly no better than round 1. Changing the thresholds from 3.5 and 2.5 to 3 and 2 or to 5 and 3.5 changed the number of exclusions from 452 to 532 or 343 without changing this pattern. In the cross-solvent models the removal of a single residue, Rum 28, raised the Q^2 of the D_2O/CD_3OD prediction of the $CDCl_3$ syringaldehyde from 0.21 to 0.42, the removal of three residues raised that of vanillin to 0.44 and the removal of one that of the cinnamaldehydes to 0.45, and the $CDCl_3$ prediction of the D_2O benzaldehydes, which failed in round 1 ($Q^2 = -0.07$), reached 0.36 after the removal of five residues. The cross-solvent agreement improves, but it remains modest.

The second round therefore confirms the cautions of Sections 13.1–13.3 rather than relaxing them. The removal of outliers is legitimate where a reference value is known to be unreliable or a spectrum is not comparable with the rest, and for the oak aromatics it gave models that are better in nearly every respect; but where the outliers are the sparse upper end of a real distribution, as for the sugars, the glycol and HMF, their removal produces statistics that look better and models that are not, and any statistic computed after the removal of outliers must be compared with the original model scored on the same residues before an improvement is claimed. For screening, the round-1 models of Table 25 remain the reference, with the round-2 benzaldehyde, syringyl and gallic acid models as better alternatives within the range of composition on which they were built.

14. Frequency domain and time domain

14.1 The question

The free-induction decay and the spectrum contain the same information, and it was asked whether every analysis of this paper could be carried out equally well on either. The answer depends on whether the analysis is linear in the data, and it can be given exactly.

14.2 Linear analyses are identical in the two domains

The discrete Fourier transform with unitary normalization is an orthogonal transformation: it preserves the inner product between any two signals and therefore every distance, every covariance and every projection. Any method that depends on the data only through mean-centered inner products, which includes principal component analysis, PLS regression and PLS-DA on mean-centered data, *k*-nearest neighbors with Euclidean distance, a linear SVM and the integral of a spectral window, must give identical results whether it is applied to the processed spectra or to the processed decays obtained from them by the inverse transform. This was verified numerically on the full-resolution spectra of the 183 D_2O/CD_3OD residues (62 944 points from δ 11.0 to −1.0 per spectrum). The principal component scores computed from the mean-centered processed spectra and from the mean-centered processed decays were identical to within numerical precision ($|r| = 1.000000$ for the first five components, with identical explained variances of 67.09, 13.37, 6.70, 3.48 and 2.26 %), the cross-validated predictions of six-component PLS1 models for HMF, the benzaldehydes, quercitol and α-glucose agreed to within 10^{-10} mg L^{-1} in every residue, and the integral of each of 481 spectral windows of 0.02 ppm, computed as an inner product of the decay with a time-domain kernel, agreed with the frequency-domain sum to a relative difference of at most 2×10^{-10}, the limit of floating-point summation (Table 27). "Doing the analysis in the time domain" is, for all of these methods, simply a change of basis.

Table 27. Frequency-domain and time-domain analyses of the 183 D_2O/CD_3OD residues (full resolution, δ 11.0 to −1.0).

Quantity	Processed spectra	Processed decays	Raw decays
PCA explained variance, PC1–PC5 (%)	67.1, 13.4, 6.7, 3.5, 2.3	67.1, 13.4, 6.7, 3.5, 2.3	46.2, 14.6, 7.9, 6.0, 4.4
$\|r\|$ of PC1–PC5 scores with the spectra	1 (reference)	1.0000, 1.0000, 1.0000, 1.0000, 1.0000	0.97, 0.79, 0.02, 0.50, 0.26
$\|r\|$ of PC1–PC5 scores, Pareto scaling	1 (reference)	0.83, 0.59, 0.74, 0.56, 0.13	—
PLS1 Q^2 (6 LV), HMF: mean-centered / Pareto	0.087 / 0.44	0.087 / 0.01	−0.07 / —
PLS1 Q^2 (6 LV), Benzaldehydes: mean-centered / Pareto	0.078 / −0.05	0.078 / −0.00	−0.00 / —
PLS1 Q^2 (6 LV), Quercitol: mean-centered / Pareto	0.040 / 0.77	0.040 / −0.40	0.02 / —
PLS1 Q^2 (6 LV), α-Glucose H-1: mean-centered / Pareto	0.953 / 0.97	0.953 / 0.80	0.91 / —
Window integrals (0.02 ppm), maximum relative difference	reference	2×10^{-10} (481 windows)	—
Cosine similarity after a 0.002 / 0.01 ppm shift	0.850 / 0.644 (0.02 ppm bins: 0.994 / 0.852)	0.850 / 0.644	—
Band-selective HLSVD vs multiplet fit, HMF CHO: *n*, *r*, median TD/FD	multiplet fit (reference)	163, 0.90, 1.09 (IQR 0.96–1.27)	—
Band-selective HLSVD vs multiplet fit, Quercitol H-6ax: *n*, *r*, median TD/FD	multiplet fit (reference)	174, 0.92, 1.08 (IQR 0.95–1.21)	—
Band-selective HLSVD vs multiplet fit, α-Glucose H-1: *n*, *r*, median TD/FD	multiplet fit (reference)	147, 0.60, 0.68 (IQR 0.36–1.14)	—

Processed decays: unitary inverse DFT of the phased, baseline-corrected, referenced spectra with the excluded windows set to zero. Raw decays: first 16 384 complex points scaled by the acquisition factors only.

14.3 Where the equivalence breaks

The equivalence holds only for the processed decays, and only for operations that commute with the transform. Three departures were examined. First, Pareto (or any variable-wise) scaling is not invariant

under rotation, because it weights each variable by its own standard deviation and the variables of the two domains are different. With Pareto scaling the principal component scores of the two domains correlated at only $|r| = 0.13$–0.83 over the first five components, and the effect on calibration was large. Mean-centered full-resolution models are dominated by the variance of the few sugar-rich residues and predicted quercitol and HMF poorly in either domain ($Q^2 = 0.04$ and 0.09); Pareto scaling of the spectra raised these to 0.77 and 0.44, whereas Pareto scaling of the decays gave −0.40 and 0.01. In every case the frequency-domain scaling was the more useful, because spectral variables correspond to chemical species while time-domain variables mix all species, and scaling a mixture of species does not up-weight any one of them. Second, the raw decays are not phased, referenced, filtered or baseline-corrected, contain the solvent and reference lines and the excluded regions, and carry the first-point distortion of the digital filter; models computed on them were consistently worse (Q^2 for α-glucose 0.91 against 0.95 on the processed data), and their principal component scores correlated with those of the processed spectra at $|r| = 0.97$, 0.79, 0.02, 0.50 and 0.26 for the first five components. Third, binning, alignment and any other operation that is applied to the spectrum point by point have time-domain counterparts that are not local: integrating a spectrum into bins of width w is equivalent to multiplying the decay by a sinc function of width about $1/w$ and folding it in time with period $1/w$, so that a 0.02 ppm (6 Hz) bin corresponds to the information in roughly the first 0.17 s of the decay.

The last point explains a practical observation. A chemical-shift error of δ ppm multiplies the decay by a phase ramp $\exp(2\pi i \cdot \Delta\nu \cdot t)$, which grows with time, and the decay is therefore exactly as sensitive to misalignment as the full-resolution spectrum: the cosine similarity between a spectrum and the same spectrum displaced by 0.002 ppm was 0.85 in both domains, and 0.64 for a displacement of 0.01 ppm, whereas for 0.02 ppm bins it was 0.994 and 0.85. Robustness to misalignment comes from the smoothing of binning, not from the domain, and the time-domain equivalent of binning is to discard the late part of the decay.

14.4 Quantitation in the time domain

The targeted quantitation was also repeated in the time domain by band-selective HLSVD (Section 4.13) for three signals of different complexity: the HMF formyl singlet, the α-glucose H-1 doublet and the eight-line quercitol H-6ax multiplet (Figure 62). The areas obtained from the magnitudes of the fitted time-domain amplitudes (halved, as explained in Section 4.13) agreed with the frequency-domain multiplet fits for the two well-isolated signals, with Pearson correlations of 0.90 and 0.92, rank correlations of 0.91 and 0.95 and median time-domain/frequency-domain ratios of 1.09 and 1.08 over 163 and 174 residues. The α-glucose doublet was reproduced much less well ($r = 0.60$, median ratio 0.68, 147 residues). It sits on the shoulder of the HDO line and among the other anomeric signals of sweetened residues, so the narrow band contains intense signals that are not part of the model, and an unconstrained decomposition distributes the doublet among spurious components; a prior-knowledge fit that fixes the doublet splitting and the relative amplitudes of its lines, in the manner of AMARES [52], would be required to match the frequency-domain result. The time domain offers two genuine advantages: the magnitude of a fitted amplitude is independent of the phase, so phase errors that bias a frequency-domain integral do not affect it, and the model is fitted to the data as acquired, without the baseline correction and zero filling of the processed spectrum. Its disadvantages are equally real. The number of components must be chosen, closely spaced lines (the two benzaldehyde singlets, $\Delta\nu \approx 2$ Hz) are resolved only when the signal-to-noise ratio is high, and a line that sits on the broad envelope of other signals is modeled less reliably than in a frequency-domain fit with an explicit local baseline. For a laboratory method, frequency-domain fitting with explicit lineshape models remains the simpler and more transparent choice, and the time-domain approach is best regarded as an

independent check of the frequency-domain results and as the natural route to fully automated, model-based quantitation of the kind implemented in Bayesian time-domain methods [54].

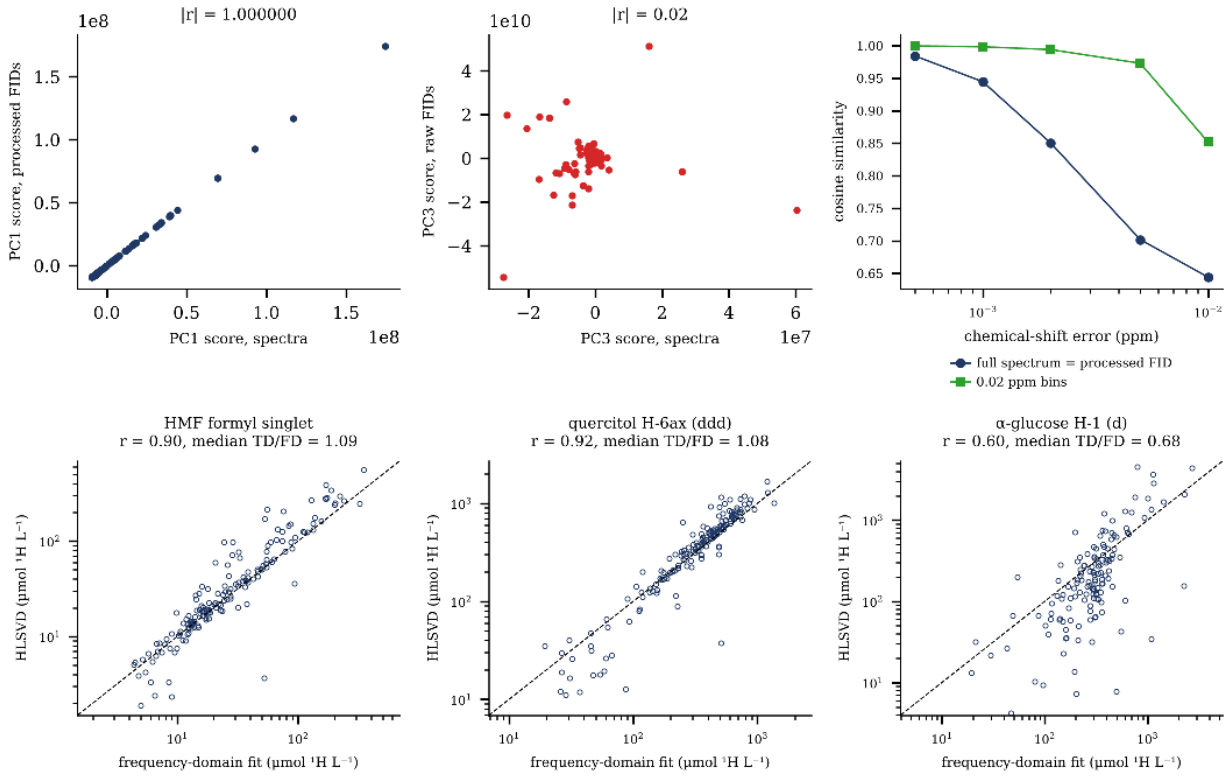

Figure 62. Frequency domain and time domain. Top: first principal component scores of the mean-centered processed spectra against those of the processed decays (left); third component scores of the spectra against those of the raw decays (center); and the cosine similarity between a spectrum and the same spectrum displaced by a chemical-shift error, at full resolution (identical in both domains) and after binning at 0.02 ppm (right). Bottom: band-selective HLSVD areas against the frequency-domain fits for the HMF formyl singlet, the quercitol H-6ax multiplet and the α-glucose H-1 doublet.

14.5 Conclusion

Every analysis of this paper could be carried out in the time domain. For the linear methods on mean-centered data the results would be identical, not merely similar, and nothing would be gained. For the methods that depend on variable-wise scaling, alignment or binning the frequency domain is the natural representation, because its variables correspond to chemical species. For quantitation of well-isolated resonances, HLSVD in the time domain reproduces the frequency-domain results within about a tenth and provides a phase-independent check of them; for overlapped multiplets it needs the prior knowledge that the frequency-domain fits already contain. The choice of domain is therefore a matter of convenience and interpretability, not of information content, which is as it must be for two representations related by an invertible linear transform.

15. One solvent or two?

The two-solvent scheme was adopted in version 1.0 because neither D_2O nor $CDCl_3$ dissolves the whole freeze-dried residue, and its cost is plain: two aliquots of every spirit, two freeze-drying runs, two tubes, two spectra and two sets of deuterated solvents, with the phenolic aldehydes divided between the two in proportions that depend on the matrix. DMSO-d_6 dissolves the whole residue, and the data of this version allow the question to be answered with evidence rather than preference. The evidence falls under four headings: what each scheme can see, how well the absolute scale and the recoveries agree, what the chemometric comparison of the same products shows, and what the practical costs are. The compound classes observed and quantified in each solvent are summarized in Figure 63.

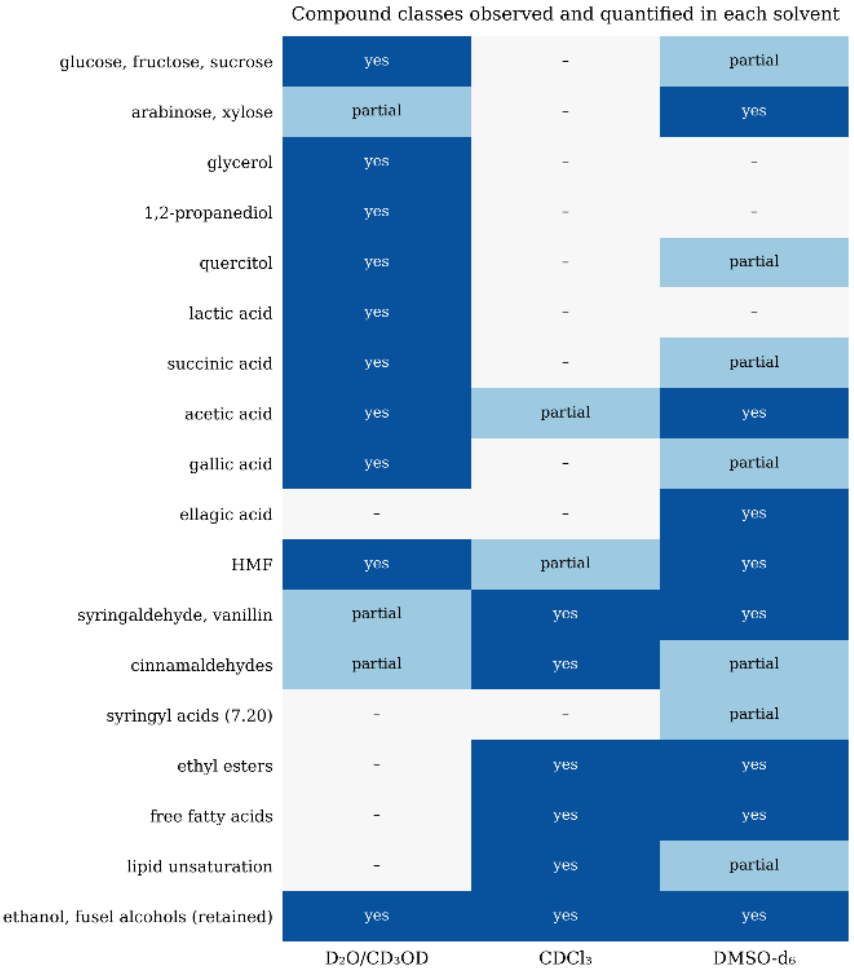

Figure 63. Compound classes observed and quantified in each solvent: yes, quantified from a resolved signal; partial, observed but overlapped, apportioned between compounds or incompletely recovered; –, not observed or not quantifiable.

15.1 What each scheme can see

The D_2O/CD_3OD residue carries the analytes on which the regulatory and authentication questions of this paper turn. In it the hydroxyl protons are exchanged, the anomeric doublets of glucose, fructose and sucrose have fixed shapes, the CH–O region between δ 3.3 and 4.1 is open, and glycerol, 1,2-propanediol, quercitol, lactic, succinic and acetic acid, gallic acid, HMF and the furanic envelope of caramel are all quantified from resolved signals (Sections 6.3–6.6). Its weaknesses are the phenolic aldehydes, which are reported only as a sum because their formyl lines lie 3 Hz apart and their apportioning is unreliable, the absence of ellagic

acid, which is insoluble in it, and the absence of most of the lipid fraction. The $CDCl_3$ residue supplies exactly what the D_2O residue lacks: the individual aldehydes with resolved formyl and ring signals, the fatty-acid ethyl esters, the free fatty acids, the degree of unsaturation of the acyl chains and diethyl succinate (Section 7). It sees no sugar, no glycerol and no organic acid, and its extraction of compounds of intermediate polarity from a sugar-containing residue is incomplete (Section 7.4).

The DMSO-d_6 residue sees most of both halves (Section 9). The aldehydes are resolved individually with their ring protons and are recovered more completely than in $CDCl_3$; syringic acid (tentatively) and ellagic acid, which neither of the other solvents shows, appear in the aromatic region; the glucose and pentose anomeric doublets are resolved; and the ethyl esters, free fatty acids and unsaturated acyl chains are visible. What it loses is equally specific. The exchange-averaged water line covers the CH–O region in which glycerol and the sugar ring protons resonate, so glycerol could not be measured in any residue; the anomeric signals change their multiplicity with the rate of hydroxyl exchange; the lactic acid methyl doublet is hidden beneath the lipid envelope; the residual DMSO-d_5 quintet and its ^{13}C satellites blind δ 2.44–2.56 and fall on the acyl α-methylene triplets; and succinic acid, the second marker of wine seasoning, was recovered only in part.

15.2 Agreement of the absolute scale and of the recoveries

The transfer of the absolute scale to DMSO-d_6 is established by the three compounds that dissolve completely in both polar solvents and are quantified from isolated resonances in both: HMF (median DMSO/D_2O ratio 1.08), the benzaldehydes (1.11) and quercitol (0.97) were recovered within the combined uncertainty of two independent preparations (Sections 5.8 and 9.8). This is the most important single result for the question of this section, because it means that a DMSO-d_6 measurement on the external maleic-acid standard is directly comparable with the D_2O measurements of version 2.0 and with any future D_2O data base. The aldehydes were recovered better in DMSO-d_6 than in $CDCl_3$ (147 % for the syringaldehyde ring protons and 131 % for the formyl protons), and the ester groups and olefinic protons less well (70 % and 63 %). Three discrepancies remain: glucose was found in DMSO-d_6 at 1.97 times its D_2O/CD_3OD level (median), mostly through an additive contribution of about 77 mg L^{-1} that points to an unidentified signal beneath the glucose windows; arabinose was found at 1.34 times its D_2O/CD_3OD level and xylose could not be measured reliably; and in the wine-seasoned residues succinic acid was found at only 0.32 of its D_2O level. The succinic acid deficit has a plausible chemical explanation (salts that do not dissolve in DMSO), and the glucose excess a plausible spectroscopic one, but both can be settled only by preparing the same bottle in both solvents on the same day.

15.3 What the chemometric comparison shows

On the twenty-four products with both a D_2O/CD_3OD and a DMSO-d_6 residue, the configurations of the products in the two solvents agreed to a modified RV coefficient of 0.77, and DMSO-d_6 alone separated wine-seasoned from other casks as well as D_2O/CD_3OD alone or the fusion of both (Section 12.10). On the ten products with residues in all three solvents, DMSO-d_6 reproduced the configuration of the fused two-solvent block to a modified RV of 0.77, whereas the D_2O/CD_3OD and $CDCl_3$ configurations were unrelated to each other (Section 12.9). The chemometrics therefore agree with the spectra: a single DMSO-d_6 spectrum carries most, but not all, of the information that the two-solvent scheme distributes over two spectra. The $CDCl_3$ block retained information that the D_2O/CD_3OD and DMSO-d_6 pair did not capture (0.35), which is the lipid fine structure that DMSO-d_6 dissolves less completely.

15.4 The case against replacing the two-solvent scheme

The strongest objection to a DMSO-only scheme is not spectroscopic but forensic, and an opposing expert would make it at once. The additions that matter most for authentication, sweetening, glycerin and flavor preparations in a glycol carrier, are measured in the D_2O residue from signals that DMSO-d_6 either loses or distorts. Glycerol, which is both a wine-cask carry-over and a deliberate addition to soften the palate of a spirit, cannot be measured at all in DMSO-d_6 because the water line covers it, and adding glycerin to a whiskey would go undetected by a DMSO-only protocol unless it reached grams per liter. The sugars can be quantified from their anomeric doublets, but the doublets change multiplicity with the water content of the residue, the fructose ring protons lie beneath the water line, and the unexplained excess of glucose in DMSO-d_6 relative to D_2O means that the absolute sugar values of the two solvents cannot yet be used interchangeably. Succinic and lactic acid, which together with glycerol carry the wine-cask signature in D_2O, are under-recovered or hidden. A protocol that loses the glycerol and part of the wine-cask chemistry in exchange for better aldehydes, ellagic acid and a lipid profile has exchanged the analytes on which a regulatory finding would rest for analytes that describe oak extraction, and for a screening laboratory that is the wrong trade.

A second objection concerns robustness. The DMSO-d_6 spectra required referencing by cross-correlation in sixteen of twenty-nine cases because TMS was absent, region-wise alignment because the solvent line itself moved with the composition of the residue, subtraction of the DMSO-d_5 satellites before the α-methylene triplets could be fitted, and fits whose line widths had to follow the exchange-broadened anomeric signals (Sections 4.15 and 5.8). Each of these steps is reproducible, but each adds an operator decision that the D_2O/CD_3OD protocol does not need, and the binned DMSO-d_6 data are exposed to apparent spectral differences that reflect the water content of the residue rather than its composition (Section 9.4). The data set is also small and dominated by one producer: twenty-one of the twenty-nine DMSO-d_6 residues are Kavalan malts, and only five of the eight other residues have a D_2O counterpart.

15.5 A better single-solvent protocol, and the recommendation

Several of the weaknesses of DMSO-d_6 are properties of the preparation rather than of the solvent, and a modified protocol would remove some of them. A small addition of D_2O to the DMSO-d_6 would exchange the hydroxyl protons, fix the multiplicities of the anomeric signals and probably dissolve the succinate and gallate salts, at the price of moving and broadening the residual water line and of lowering the solubility of the long-chain esters; the proportion should be optimized on the matched residues rather than chosen in advance. TMS, or better an internal standard that is soluble and stable in DMSO-d_6 and resonates in an empty region, should be present in every tube [76], so that the chemical-shift reference and an internal check of the absolute scale do not depend on the residual solvent line. The calibrated 90° pulse and the maleic-acid T_1 should be measured in each solvent rather than transferred from D_2O. And the comparison of solvents should be repeated on aliquots of the same bottles freeze-dried and prepared side by side, with spiking of glycerol, succinic acid and glucose, which is the experiment that would settle the open discrepancies. Quantitative NMR of ellagitannins and related polyphenols in plant material has been validated in comparable solvent systems [77], and the same approach would extend the ellagic acid results of Section 9.6 to the parent ellagitannins.

On the present evidence, the recommendation is not to replace the two-solvent scheme with DMSO-d_6 alone. For authentication and regulatory questions the D_2O/CD_3OD residue is indispensable, because it alone measures glycerol, the complete sugar profile and the organic acids on which a finding of addition or of wine seasoning would rest. The choice of the second solvent depends on the purpose. Where the lipid profile matters, for example to distinguish cask-strength from chill-filtered products or to characterize fatty-

acid ethyl esters, $CDCl_3$ remains the better second solvent; where the purpose is a compositional fingerprint of oak extraction, DMSO-d_6 is the better second solvent, because it resolves the aldehydes more completely than $CDCl_3$, adds ellagic acid and syringic acid, and carries most of the lipid information as well. A single DMSO-d_6 spectrum, preferably with D_2O added, is a reasonable screening tool for cask history and oak extraction, and the agreement of its absolute scale with that of D_2O makes it compatible with the data base of this paper, but it is not a substitute for the D_2O/CD_3OD measurement wherever an addition has to be demonstrated.

16. Critical assessment

16.1 What the data establish

Several conclusions rest on direct spectral evidence and on internal checks that do not depend on any model, and can be regarded as established. The MestReNova documents contain the complete acquired decays, the parameter sets and a losslessly coded copy of the processed spectra, and the processing can be reproduced from the decays exactly; the first D_2O set had been silently normalized in the processing software and its absolute scale has been restored. The external maleic acid standard transfers to the residues: the CHD_2OD of the CD_3OD in the samples, measured on the maleic acid scale, agrees with the CHD_2OD of the blank within 3.4 %, and the receiver-gain departures from nominal are no larger than about two standard errors (at most 13 %). Twenty solvent labels were wrong, and the CHD_2OD signal identifies the correct solvent system in every case; the same signal identified two acquisition sessions with a sensitivity deficit of 29–46 %, which was corrected. The carbohydrate, glycerol, 1,2-propanediol, quercitol, gallic acid, HMF, lactic and succinic acid assignments are supported by complete spin systems, heteronuclear correlations or both. The cognacs contain gram-per-liter amounts of glucose and fructose in equimolar proportion; several rums, two maple whiskeys and a honey whiskey contain gram-per-liter amounts of sucrose or invert sugar; three products carry vanillin without syringaldehyde; retained ethanol in a freeze-dried residue follows its sugar content; and the gram-per-liter glycerol that a single-multiplet fit assigned to fructose-rich residues was the β-pyranose H-1b doublet of fructose, because the second glycerol multiplet is absent from those spectra. Linear analyses on mean-centered data are identical in the time and frequency domains.

Version 2.3 adds one result of the same standing, although it rests on a single product: freeze-drying removed the volatile congeners that dominate the native spectrum, ethyl acetate and acetic acid to at most 1.7 and 0.3 % of complete retention, acetaldehyde to at most 8.0 % and the fusel alcohols to about 1.4 %, whereas the oak aromatics, at the noise level in the native spectrum, were retained within the precision of the comparison (Section 2.9).

Version 2.1 added four results of the same standing. The absolute scale transfers to DMSO-d_6: three compounds that dissolve completely in both polar solvents and are quantified from isolated resonances were recovered at 1.08, 1.11 and 0.97 times their D_2O/CD_3OD levels (medians) from residues prepared separately from the same bottles. L-Arabinose and D-xylose are present in the residues of oak-matured spirits: the α-arabinopyranose and β-xylopyranose H-1 doublets are resolved in D_2O/CD_3OD, the gHSQC spectrum of the concentrated Kavalan residue contains the C-1 correlations of the free pyranoses, and the remainder of the α-glucose region matches the anomeric equilibrium of arabinose. The glucose reported in version 2.0 was overestimated wherever arabinose was present, by a median of 75 % in the American whiskeys and 41 % in the Kavalan malts, and the β-glucose doublet had been missed in eight residues whose anomeric region was displaced by their CD_3OD content; both errors have been corrected. And the second principal component of the $CDCl_3$ block was a moisture artifact, as the ratio of the β-methylene to the α-methylene region of the two residues that carry it shows.

16.2 What is plausible

Other conclusions are well supported but rest on an interpretation that could be wrong. The assignment of the upfield D_2O formyl line to syringaldehyde rests on a statistical association (r = 0.95) and a regression coefficient (1.46) that is below its theoretical value of 2, and that of the downfield line to vanillin on the position of the vanillin formyl line in the flavored products rather than on a correlation; both are confirmed in $CDCl_3$ but not independently in D_2O. The broad formyl envelope at δ 9.48 is attributed to caramel-derived furanic oligomers on chemical grounds and by its distribution across the classes, but it has not been

calibrated against a caramel of known composition, and Maillard products from the cask toasting could contribute. The interpretation of vanillin without syringaldehyde as a flavor preparation is strong in combination with a glycol carrier, but a spirit matured in a cask that had previously held a vanilla-flavored product would show the same pattern. The wine-cask origin of glycerol is plausible where succinic acid accompanies it, but the proportion of the two in genuine wine-cask spirits has not been established, and glycerin added to a product that also carries some cask-derived succinic acid could not be recognized. The *abocado* interpretation of Tromba Reposado now rests on its fructose and HMF alone; it is the most economical explanation of that composition, but it concerns a single bottle.

Among the results of version 2.1, the assignment of the two pairs of DMSO-d_6 doublets to arabinose and xylose rests on their couplings, on an intensity order that matches the D_2O/CD_3OD residues and on the anomeric equilibria, not on a spin system traced in DMSO-d_6, and their line shapes depend on the rate of hydroxyl exchange. The explanation of the succinic acid deficit in the wine-seasoned DMSO-d_6 residues by salts that do not dissolve in DMSO is chemically reasonable but untested, and the assignment of the δ 7.200 singlet to syringic acid rests on a single line.

Version 2.3 qualifies two assignments of this kind. The reference spectra place the syringaldehyde ring protons at δ 7.15–7.23 in every medium examined (D_2O/CD_3OD at 60 °C, the model spirit, $CDCl_3$ at 50 °C and the DMSO-d_6 residues), whereas the D_2O/CD_3OD residue line assigned to them lies at δ 7.303; the assignment rests on its correlation with the formyl line (r = 0.95) and on its ^{13}C shift, not on its position, and would be overturned if syringaldehyde added to a residue produced a line elsewhere. The δ 7.33 singlet is assigned to syringic acid because syringic acid lies downfield of syringaldehyde in every aqueous medium and because the ratio of the two residue singlets (0.47) equals the ratio of the corresponding DMSO-d_6 singlets (0.47); both arguments are circumstantial.

16.3 What is speculative

Some statements in this paper are speculative and are identified as such. The source of the small sucrose-like doublets in a few straight whiskeys (oak starch fragments, cross-contamination, or sucrose) is unknown. The broad lactic acid signals of Tromba Extra Añejo may reflect metal association or oligomerization, or neither. The ratio of quercitol to the aldehydes as an index of toasting is a hypothesis. The classification models have not been tested on an independent set and should not be applied to new samples. The contrast between sweet and dry wine casks in the DMSO-d_6 block without the maple-barrel residue rests on one or two casks of each kind; the broad singlet at δ 7.36–7.39 in four wine-seasoned DMSO-d_6 residues may be ellagic acid in an ionized or associated state, or another compound; the arabinose-to-glucose ratio as an index of new-oak contact is a hypothesis; and the compound that gives the δ 4.474 doublet without its β-arabinopyranose partner in twenty-two residues has not been identified.

16.4 What an expert reviewer would challenge

An expert reviewer or an opposing expert witness would raise at least the following points, and each is a fair criticism of the present data. First, recovery through freeze-drying was not measured. The internal checks validate the solution and the acquisition, not the residue, and residues of the same product prepared under different sample numbers differ by up to a factor of 3.1 in their non-volatile content, although the single true replicate preparation agreed within 9 %. Every concentration is therefore a concentration in the residue as recovered, and no statement about the concentration in the spirit can be defended without a recovery standard and replicate preparations. Second, the T_1 values of most analytes were assumed, and for the aldehydes in spectra acquired at short repetition times the saturation correction is uncertain by 14–22 %. Third, the external standard was measured once, in one solvent system, without a measured T_1; the 90°

pulse (13 µs) and the effective content of the standard (95 % of the weighed amount) on which version 2.2 rests are the author's own determinations, which no internal check of these data can confirm; and several residues were prepared with solvent volumes that were assumed rather than recorded. Fourth, the solvent labels and one product label were wrong, which shows that the sample metadata contain errors; a regulatory or legal use of these data would require a documented chain of custody and a laboratory information system rather than spectrum titles. Fifth, each product is represented by a single bottle of undocumented lot, so differences between products cannot be separated from differences between lots. Sixth, the PLS calibrations are validated against reference values derived from the same spectra, and the classification models against a sample set in which producers and acquisition batches are confounded with the classes. Seventh, at 300 MHz several compounds of interest are not resolved (the two benzaldehydes in D_2O, the two cinnamaldehydes, the minor cyclitols, glycerol in sweetened products), and the assignments that rest on a single line would not survive a demand for confirmation at higher field or by an orthogonal method. Eighth, a constrained fit always returns a number. Most analytes were quantified from a single resonance, and the glycerol results show that a single-resonance value can be internally consistent, reproducible and wrong; wherever a molecule offers a second resolved resonance it should be used as a check, as was done here for glycerol, sucrose and syringaldehyde. Ninth, the glucose, arabinose and xylose found in DMSO-d_6 exceeded the D_2O/CD_3OD values of the same products by median factors of 1.97, 1.34 and 1.34, whereas three other compounds agreed within a few percent; for glucose most of the excess is an additive contribution of about 77 mg L^{-1} from an unidentified signal, and until it is identified the carbohydrate values of the two solvents cannot be used interchangeably and the DMSO-d_6 sugar values are best treated as upper bounds. Tenth, the comparison of one solvent with two rests on twenty-four and ten products, most of them Kavalan malts, prepared months apart from separate aliquots; it compares preparations as well as solvents, and the matched-set classifications cannot rank the solvents. Eleventh, version 2.0 contained two quantitation errors that its fit statistics did not reveal, the absorption of the β-arabinopyranose doublet into the α-glucose fit and fits at fixed positions in residues whose medium had moved the signals, and an error of the second kind was found in two DMSO-d_6 succinic acid fits of this version; the fitted position must be checked against the observed line in every residue, and any resonance that may carry a second compound must be checked against a second resonance of each. Twelfth, the reference spectra of version 2.3 were recorded at concentrations, acidities, CD_3OD or ethanol contents and, for the syringyl compounds, temperatures that differ from those of the residues, and the same compound moved by up to 0.09 ppm between two aqueous media (syringic acid); the references support assignments by pattern, multiplicity and order, but a position transferred from them carries an uncertainty of up to 0.1 ppm, and every assignment that matters should be confirmed by adding the compound to a residue. Thirteenth, the second round of PLS calibrations shows how readily the removal of outliers flatters a calibration: most of its apparent gain came from no longer scoring the excluded residues, and a statistic quoted after outlier removal without that comparison would overstate the model. None of these points invalidates the chemistry described here, but each limits what the numbers can be used for, and the recommendations of Section 17 address them in turn.

17. Recommendations and outlook

The analysis points to a small number of changes that would convert the present screening data into a quantitative method suitable for authentication work, and most of them cost little. The first and most important is a recovery standard. A weighed quantity of a non-volatile compound with a clean, isolated resonance, added to the spirit before dilution and freeze-drying, would measure the combined recovery of drying, transfer and redissolution in every preparation. For the D_2O residue, 1,3,5-benzenetricarboxylic acid is a good candidate, because its single aromatic resonance near δ 8.6 falls in a region that is essentially empty in every residue of this study and it is not volatile; its long T_1 (about 10 s [41]) must be respected in the acquisition. Maleic acid, the external standard of this work, is less suitable as an internal one because its singlet moves with the acidity of the residue and falls among the lignan singlets near δ 6.4. A lipophilic standard of similar non-volatility, with a resonance clear of the aldehyde and lipid signals, would be needed for the $CDCl_3$ residue. The aliquot of spirit and the residue should be weighed, and the volume of NMR solvent dispensed by mass.

Second, the acquisition should be made quantitative by design rather than by correction. A repetition time of at least five times the longest T_1 of interest (about 40 s if the aldehydes have T_1 values near 8 s) with a 90° pulse, or a shorter repetition time with T_1 values measured by inversion recovery on representative residues, would remove the largest analyte-specific uncertainty. The maleic acid blank should be acquired with each batch, at the receiver gain of that batch, and its T_1 measured once. The 90° pulse should be calibrated by a nutation or 360° null measurement in each solvent and recorded with every data set: in the present data a nominal value of 10 µs, against a calibrated 13 µs, was carried into the analysis until version 2.2 and changed individual concentrations by up to 5 %. The standard itself should be prepared from a certified reference material, or checked against one, and from a mass large enough for the weighing to contribute no more than 0.1 % (at least 20 mg on the balance used here, diluted volumetrically to the working concentration): the blank of this work was prepared from 2.55 mg, and its effective content of 95 %, established after the event, moved every result by 5.3 %. Third, replicate preparations of the same bottle and replicate bottles of the same lot should be included in every study, with the acquisition order randomized, so that preparation precision, lot-to-lot variation and batch effects can be separated. Fourth, the sample metadata should be recorded in a structured form at the time of preparation; the CHD_2OD check described here should be retained as an automatic audit of the solvent system.

Fifth, the assignments that carry regulatory weight should be confirmed by an orthogonal method on a subset of products: HPLC of sugars, glycerol and the furanic and phenolic aldehydes, and enzymatic glycerol and sugar assays, with the results used to validate the qNMR values and to calibrate the caramel index. Measurements at 500 or 600 MHz would resolve the benzaldehyde and cinnamaldehyde pairs in D_2O and the minor cyclitols. Sixth, the classification models should be rebuilt on a sample set designed for the purpose, with several producers per category, documented cask histories and an independent test set acquired at a later date.

Seventh, the solvent scheme should follow the purpose of the measurement (Section 15). For authentication and regulatory questions the D_2O/CD_3OD residue remains indispensable, and the choice of the second solvent depends on whether the lipid profile ($CDCl_3$) or a fingerprint of oak extraction (DMSO-d_6) is wanted. Where DMSO-d_6 is used, a small, optimized proportion of D_2O should be added to exchange the hydroxyl protons and fix the multiplicities of the anomeric signals; TMS or a DMSO-soluble internal standard should be present in every tube, so that neither the chemical-shift reference nor an internal check of the absolute scale depends on the residual solvent line; the 90° pulse and the maleic acid T_1 should be measured in each solvent; and the solvent lot, the presence of TMS and, for the D_2O residues, the volume of CD_3OD should be recorded with every preparation, since each of these varied without record in the

present data. The $CDCl_3$ should be dried over molecular sieves and the water line of every spectrum recorded. Fits at fixed positions should be replaced, as in this version, by fits whose positions follow a residue-specific offset determined from the spectrum itself. Finally, the discrepancies between DMSO-d_6 and D_2O/CD_3OD for glucose, the pentoses and succinic acid should be settled by preparing aliquots of the same bottles side by side in both solvents, with spiking of glycerol, succinic acid and glucose, and the pentoses and the compound responsible for the δ 4.474 doublet in the residues that failed the arabinose test should be confirmed by HSQC and by HPLC of the sugars.

Two further recommendations follow from the additions of version 2.3. Reference spectra should be recorded in the medium of the residues: in D_2O/CD_3OD of the same composition, at 27 °C, at concentrations of the order of those in the residues, at an acidity adjusted to that of a typical residue and with TMSP in every tube, and spirit-like reference solutions should be prepared with undenatured ethanol, since the 50:50 solutions of this work contained methanol and 2-propanol. The decisive experiment, however, is addition to a residue. Syringaldehyde, syringic acid, the hydroxycinnamic acids, scopoletin, 1,3-propanediol and D-mannose, each added at a few milligrams per liter of spirit to an aliquot of a well-characterized residue, would confirm or overturn the assignments of the δ 7.30 and 7.33 singlets and test the overlaps identified in Section 2.11, 1,3-propanediol under the quercitol H-6ax multiplet and mannose near the α-xylopyranose doublet, at the cost of a few spectra. And a spectrum of each intact spirit, acquired beside its residues, would add the volatile congeners that the residue cannot hold (Section 2.9). Finally, any recalibration after the removal of outliers should be reported together with the original model scored on the same residues, as in Section 13.5, so that the effect of refitting is not confused with the effect of leaving the hardest residues out.

Appendix A. Samples and acquisition parameters

The residues of the combined data base are listed in Table 29 with their category, solvent system (as corrected in Section 5.4), solution volume, transient count, receiver gain and relaxation delay. Spectrum titles and sample identifiers are reproduced as recorded, including spelling, so that each entry can be traced to its MestReNova document. Table 28 summarizes the acquisition parameters by data set.

Table 28. Acquisition parameters by data set (Varian Mercury-MVX 300, 299.99 MHz, s2pul, spectral width 14 992.5 Hz, 64 000 complex points, acquisition time 4.2688 s, 27 °C, 90° pulse 13 µs; the stored pw90 of 10 µs is nominal).

Data set	Spectra	Transients	Gain (dB)	Pulse (µs)	Delay d_1 (s)	T_R (s)	Acquired
D_2O set 1	84	32–512	18–38	8	3–15	7.27–19.27	Dec 2024–Aug 2025
D_2O set 2	113	64–516	18–38	8	2–15	6.27–19.27	Dec 2024–Mar 2026
$CDCl_3$ set 1	39	60–512	38–38	8	5–15	9.27–19.27	Dec 2024–Sep 2025
$CDCl_3$ set 2	22	16–512	38–38	7/8	3–5	7.27–9.27	Jan 2025–Mar 2026

Table 29. Residues of the combined data base. Set: D1/D2, first and second D_2O sets; C1/C2, first and second $CDCl_3$ sets.

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
0	D1	BDAS-011-H_FD	Tromba Extra Anejo Tequila	AGV	D_2O/CD_3OD	1000	650*	220	34	5
1	D1	BDAS-016-H_FD	Clase Azul Gold	AGV	D_2O/CD_3OD	1000	650*	256	34	5
2	D1	BDAS-017-H_FD	Clase Azul Anejo	AGV	D_2O	1000	650*	244	38	5
3	D1	BDAS-025-H_FD	Bozal Cuishe	AGV	D_2O/CD_3OD	1000	650*	256	34	5
4	D1	BDAS-043-H_FD	Sotol Carnay Tres Pilares	AGV	D_2O/CD_3OD	1000	650*	128	34	5
5	D1	BDAS-044-H_FD	Sotol Ono Chihuahua	AGV	D_2O/CD_3OD	1000	650*	32	34	5
6	D1	BDAS-049-H_FD	Sotol Ono Chihuahua	AGV	D_2O/CD_3OD	1000	650*	256	34	5
7	D1	Bryan_Litchfield-002-H_FD	Litchfield Distilling - Maple Bourbon 43%ABV	AMW	D_2O	1000	625	128	18	15
8	D1	Bryan_Litchfield-003-H_FD	Litchfield Distilling - Straight Rye Whiskey 46%ABV	AMW	D_2O	1000	625	128	38	15
9	D1	Bryan_Local-001-H_FD	Hudson House - NY Straight Rye Whiskey - 43%ABV	AMW	D_2O	1000	625	128	38	15
10	D1	Bryan-004-H_FD	Jameson - Blender's Dog Irish Whiskey 43%ABV	IRW	D_2O/CD_3OD	1000	600	128	38	15
11	D1	Bryan-013-H_FD	Powers - Gold Label Single Pot Still Irish Whiskey 40…	IRW	D_2O/CD_3OD	1000	650	160	36	7
12	D1	Bryan-015-H_FD	Bushmills 10 Yr Irish Single Malt Whisket 40%ABV	IRW	D_2O/CD_3OD	1000	600	128	38	15
13	D1	Bryan-018-H_FD	Powers - Gold Label Single Pot Still Irish Whiskey 40…	IRW	D_2O/CD_3OD	1000	650	244	36	7
14	D1	Bryan-019-H_FD	Bushmills - Black Bush Irish Whisky 40%ABV	IRW	D_2O †	1000	650*	256	38	5
15	D1	BDAS-008-H_FD	Tromba Reposado - French Oak Casks	AGV	D_2O/CD_3OD	1000	650*	248	34	5
16	D1	BDAS-010-H_FD	Tromba Anejo Tequila	AGV	D_2O/CD_3OD	1000	650*	216	34	5
17	D1	Bryan-020-H_FD	Bushmills Original Irish Whiskey	IRW	D_2O/CD_3OD	1000	650	256	38	8
18	D1	Bryan-023-H_FD	Glen Grant - 12 Year Speyside SMW	SCW	D_2O/CD_3OD	1000	650	228	38	3
19	D1	Bryan-025-H_FD	Brenne Estate Cask Single Malt French Whiskey 40%ABV	OWW	D_2O/CD_3OD	1000	650	184	36	7
20	D1	Bryan-033-H_FD	Jameson - Cooper's Croze Irish Whiskey 43%ABV	IRW	D_2O/CD_3OD	1000	600	128	38	15
21	D1	Bryan-036-H_FD	Red Breast - 12 Yr Cask Strength Irih Single Pot Stil…	IRW	D_2O/CD_3OD	1000	650	256	38	8
22	D1	Bryan-043-H_FD	Kavalan Solist Vinho Barrique Single Cask Strength Si…	TWW	D_2O/CD_3OD	1000	650	208	38	15
23	D1	Bryan-049-H_FD	Courvoisier VS Cognac	BRA	D_2O/CD_3OD	1000	600	84	38	15
24	D1	Bryan-050-H_FD	Merlet - XO Cognac	BRA	D_2O/CD_3OD	1000	650	256	34	7
25	D1	Bryan-051-H_FD	Branson Cognac VSOP - Cognac Grande Champagne 40%ABV	BRA	D_2O/CD_3OD	1000	650	196	28	7
26	D1	Bryan-052-H_FD	Courvoisier VSOP Cognac 40%ABV	BRA	D_2O/CD_3OD	1000	650	176	28	7
27	D1	Bryan-053-H_FD	Branson - Phantom VS Cognac	BRA	D_2O/CD_3OD	1000	650	256	28	7
28	D1	Bryan-054-H_FD	Branson - Royal VSOP Cognac	BRA	D_2O/CD_3OD	1000	650	256	28	7
29	D1	Bryan-058-H_FD	Courvoisier - XO Cognac	BRA	D_2O/CD_3OD	1000	650	256	30	7
30	D1	Bryan-059-H_FD	Branson Cognac - XO Prmier Cru Cognac 40%ABV	BRA	D_2O/CD_3OD	1000	650	180	28	7
31	D1	Bryan-060-H_FD	Bisquit and Debouche - VSOP Cognac	BRA	D_2O/CD_3OD	1000	650	236	30	7
32	D1	Bryan-061-H_FD	Larsen - Aqua Ignis Cognac	BRA	D_2O/CD_3OD	1000	700	160	24	7
33	D1	Bryan-070-H_FD	Branson Cognac - XO Prmier Cru Cognac 40%ABV	BRA	D_2O/CD_3OD	1000	650	196	28	7
34	D1	Bryan-078-H_FD	Coperies - Grench Single Malt Whiskey	OWW	D_2O/CD_3OD	1000	600	128	38	15
35	D1	Bryan-082-H_FD	Waterford - Single Farm Origin - Rathclogh Edition 1.…	IRW	D_2O/CD_3OD	1000	650	256	38	3

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
36	D1	Bryan-085-H_FD	Tenjaku - Pure Japanese Malt Whisky - 43%ABV	JPW	D₂O/CD₃OD	1000	650	228	38	7
37	D1	Bryan-087-H_FD	Bhakta Spirits 50 Yr Armagnac Brandy 46%ABV	BRA	D₂O/CD₃OD	1000	600	120	38	15
38	D1	Bryan-089-H_FD	Amrut - Fusion - Indian Single Malt Whisky - 50%ABV	OWW	D₂O/CD₃OD	1000	650	256	38	7
39	D1	Bryan-103-H_FD	An Dulaman - Santa Ana Armada - Barrel Aged Gin 57%ABV	GIN	D₂O/CD₃OD	1000	650	164	38	3
40	D1	Bryan-107-H_FD	Kiss - Black Diamond Premium Dark Rum - Aged 3 Years	RUM	D₂O/CD₃OD †	1000	650*	144	22	3
41	D1	Bryan-124-H_FD	Medford Rum - Aged 3 Years	RUM	D₂O/CD₃OD	1000	650	180	38	3
58	D1	Bryan-141-H_FD	Glen Grant 18 Yr Speyside Single Malt Whiskey	SCW	D₂O/CD₃OD	1000	600	128	38	15
59	D1	Bryan-142-H_FD	Kavalan Earth Silver Wine Cask Matured Single Cask Si…	TWW	D₂O/CD₃OD	1000	650	212	38	15
60	D1	Bryan-145-H_FD	Redbreast - Single Pot Still Irish wft Whiskey - 27 Y…	IRW	D₂O/CD₃OD †	1000	650*	156	38	3
61	D1	Bryan-149-H_FD	Stauning - Kaos - Triple Malt Whisky	OWW	D₂O/CD₃OD	1000	650	160	38	5
62	D1	Bryan-172-H_FD	Spirit of HVEN Mercurious Corn Whisky 45.6%ABV	OWW	D₂O/CD₃OD	1000	650	220	36	15
63	D1	Bryan-178-H_FD	Red Spot - 15 Year Old Single Pot Still Irish Whiskey…	IRW	D₂O/CD₃OD	1000	650	196	38	7
64	D1	Bryan-180-H_FD	Aberlour - 16 Year Speyside - Single Malt Whisky 43%A…	SCW	D₂O/CD₃OD	1000	650	256	36	7
65	D1	Bryan-182-H_FD	Jameson - Distillers Safe Irish Whiskey 43%ABV	IRW	D₂O/CD₃OD	1000	650	196	36	7
66	D1	Bryan-185-H_FD	Aberlour A'Bunadh Alba Speyside Single Malt Whisky 58…	SCW	D₂O/CD₃OD	1000	650	216	36	7
67	D1	Bryan-310-H_FD	Makers Mark Straight Bourbon Whiskey 45%ABV	AMW	D₂O/CD₃OD	1000	650	256	36	7
68	D1	Bryan-311-H_FD	Pendleton 1910 - 12 Year Old Canadian Rye Whiskey 40%…	CAN	D₂O/CD₃OD	1000	650	132	36	15
69	D1	Bryan-313-H_FD	TX - Texas Straight Bourbon Whisky	AMW	D₂O/CD₃OD	1000	650*	128	38	3
70	D1	Bryan-325-H_FD	Coalition Whiskey Barrel Proof Kentucky SRW 54.4%ABV	AMW	D₂O/CD₃OD	1000	650	256	38	15
71	D1	Bryan-331-H_FD	Bernheim - 7 Year Old Small Batch Straight Wheat Whis…	AMW	D₂O/CD₃OD	1000	650	228	38	15
72	D1	Bryan-334-H_FD	Bardeney - Bottled in Bond Rye Whiskey	AMW	D₂O/CD₃OD	1000	650	240	38	3
73	D1	Bryan-336-H_FD	TX - American Blended Whisky	AMW	D₂O/CD₃OD	1000	650	256	38	8
75	D1	Bryan-343-H_FD	Kavalan-Distillery Reserve-Rum Cask - Single Cask Str…	TWW	D₂O/CD₃OD	1000	600	96	38	15
76	D1	Bryan-346-H_FD	Kavalan Sky Gold Wine Cask Matured Single Cask Single…	TWW	D₂O/CD₃OD	1000	650	184	38	15
77	D1	Bryan-348-H_FD	Jameson - Bow Street 18 Yr Old Cask Strength Irish Wh…	IRW	D₂O/CD₃OD	1000	650	196	38	7
78	D1	Bryan-360-H_FD	Mount Gay - Black Barrel Double Cask Blend Rum 43%ABV	RUM	D₂O/CD₃OD	1000	650	196	38	7
79	D1	Bryan-367-H_FD	Mount Gay XO Triple Cask Rum Blend 43%ABV	RUM	D₂O/CD₃OD	1000	650	184	38	7
80	D1	Bryan-368-H_FD	Makers Mark Cask Strength Single Malt Whisky 55%ABV	AMW	D₂O/CD₃OD	1000	650	256	38	15
81	D1	Bryan-Kavalvan-005-H_FD	Kavalan - Solist Oloroso Sherry Single Cask Strength…	TWW	D₂O/CD₃OD †	1000	650*	512	38	5
82	D1	Bryan-Kavalan-001-H_FD	Kavalan - Solist Amontillado Sherry SCS SMW	TWW	D₂O/CD₃OD	1000	650	160	38	5
83	D1	Bryan-Kavalan-006-H_FD	Kavalan - Concertmaster Sherry Cask Finish SMW	TWW	D₂O/CD₃OD	1000	650	160	38	5
84	D1	Bryan-Kavalan-007-H_FD	Kavalan - King Car - Conductor SMW	TWW	D₂O/CD₃OD	1000	650	160	38	5
85	D1	Bryan-Kavalan-008-H_FD	Kavalan - Distillery Reserve Peaty SCS SMW	TWW	D₂O/CD₃OD	1000	650	160	38	5
86	D1	Bryan-Kavalan-010-H_FD	Kavalan Solist Brandy cask †	TWW	D₂O/CD₃OD	1000	650	160	38	5
87	D1	Bryan-Kavalan-011-H_FD	Kavalan - Classic - SMW	TWW	D₂O/CD₃OD †	1000	650*	144	38	3
88	D1	Bryan-Kavalan-017-H_FD	Solist Ex-Bourbon - Single Cask Strength - SMW	TWW	D₂O/CD₃OD	1000	650	176	38	3
89	D1	Bryan-Kavalan-022-H_FD	Solist Port - Single Cask Strength - SMW	TWW	D₂O/CD₃OD	1000	650	228	38	3
90	D1	Bryan-Kavalan-029-H_FD	Solist Moscatel Sherry - Single Cask Strength - SMW	TWW	D₂O/CD₃OD	1000	650	212	30	3
91	D1	Bryan-Kavalan-034-H_FD	Oloroso Sherry Oak - SMW	TWW	D₂O/CD₃OD	1000	650	124	38	3
92	D1	Bryan-Kavalan-035-H_FD	Solist Madeira Single Cask Strength - SMW	TWW	D₂O/CD₃OD	1000	650	160	38	3
93	D1	Bryan-Kavalan-037-H_FD	Ex-Bourbon Oak SMW	TWW	D₂O/CD₃OD	1000	650	160	38	5
94	D1	Bryan-Kavalan-038-H_FD	Kavalan - Solist Fino Sherry Single Cask - SMW	TWW	D₂O/CD₃OD †	1000	650*	128	38	3
95	D1	Bryan-Kavalan-042-H_FD	Kavalan - Artist Series Paul Chiang Puncheon Single C…	TWW	D₂O/CD₃OD †	1000	650*	180	38	3
96	D1	Bryan-Kavalan-045-H_FD	Solist Manzanilla Sherry - Single Cask Strength - SMW	TWW	D₂O/CD₃OD	1000	650	252	38	3
97	D1	Bryan-Kavalan-047-H_FD	Podium - SMW	TWW	D₂O/CD₃OD	1000	650	192	38	3
98	D1	Bryan-Kavalan-143-H_FD	Kavalan - King Car - 40th Anniversary Selected WIne C…	TWW	D₂O/CD₃OD †	1000	650*	152	38	3
99	D1	Bryan-Kavalan-343-H_FD	Kavalan - Distillery Reserve - Rum Cask - Single Cask…	TWW	D₂O/CD₃OD	1000	650	160	38	7
100	D1	Bryan-Kavalan-345-H_FD	Kavalan - Distillery Select No. 1 Single Malt Whiskey…	TWW	D₂O/CD₃OD	1000	650	164	36	7
101	D2	Bryan-369-H_FD	Old Overholt 114 Proof SRW	AMW	D₂O/CD₃OD	1000	650	256	38	3
102	D2	Bryan-370-H_FD	Bare Knuckle - Straight Wheat Whiskey	AMW	D₂O/CD₃OD	1000	600	120	38	15
103	D2	Bryan-371-H_FD	Makers Mark 101 Single Malt Whisky 50.5%ABV	AMW	D₂O/CD₃OD	1000	650	256	38	15
104	D2	Bryan-386-H_FD	Doc Swinson's - Alter Ego Solera Method SRW	AMW	D₂O/CD₃OD †	1000	650*	176	38	3
105	D2	Bryan-391-H_FD	Southern Star - Double Rye Whisky - 44%ABV	AMW	D₂O/CD₃OD	1000	650	256	38	7

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
106	D2	Bryan-398-H_FD	Tin Cup SRW	AMW	D₂O/CD₃OD †	1000	650*	160	38	3
107	D2	Bryan-406-H_FD	Black Velvet Reserve Canadian Whiskey 40%ABV	CAN	D₂O/CD₃OD	1000	600	84	38	15
108	D2	Bryan-407-H_FD	Bower Hill - Barrel Strength KY Straight Bourbon Whis…	AMW	D₂O/CD₃OD	1000	600	128	38	15
109	D2	Bryan-408-H_FD	Widow Jane - 10 Yr Blended Straight Bourbon Whiskey B…	AMW	D₂O/CD₃OD	1000	600	128	38	15
110	D2	Bryan-409-H_FD	Cedar Ridge - The Quintessential American Single Malt…	AMW	D₂O/CD₃OD	1000	600	128	38	15
111	D2	Bryan-411-H_FD	Davidson Reserve Small Batch Tennessee SBW	AMW	D₂O/CD₃OD	1000	650	256	38	3
112	D2	Bryan-413-H_FD	Happenstance Straight Bourbon Whiskey 45%ABV	AMW	D₂O/CD₃OD	1000	600	128	38	15
113	D2	Bryan-419-H_FD	Uncle Nearest - 1856 Premium Tennessee Whiskey	AMW	D₂O/CD₃OD	1000	600	64	38	15
114	D2	Bryan-425-H_FD	Russell's Reserve 6 Year Kentucky SRW	AMW	D₂O/CD₃OD	1000	650	192	38	3
115	D2	Bryan-427-H_FD	Russell's Reserve Single Barrel Straight Rye Whiskey…	AMW	D₂O/CD₃OD	1000	600	128	38	15
116	D2	Bryan-428-H_FD	Bespoke Spirits - Rye Whiskey Batch 2020-1	AMW	D₂O/CD₃OD	1000	650	160	38	5
117	D2	Bryan-430-H_FD	High West - Double Rye! Blended SRW Batch 19G26	AMW	D₂O/CD₃OD	1000	650	224	38	7
118	D2	Bryan-435-H_FD	High West - A MidwinterNightsDream - Act7 Scene5 Blen…	AMW	D₂O/CD₃OD	1000	650*	256	38	5
119	D2	Bryan-438-H_FD	Forty Creek Barrel Select Canadian Whiskey 40%ABV	CAN	D₂O/CD₃OD	1000	600	128	38	15
120	D2	Bryan-444-H_FD	Basil Hayden - Kentuwftck Straight Bourbon Whiskey 40…	AMW	D₂O/CD₃OD	1000	650	200	36	15
121	D2	Bryan-449-H_FD	Alquimia Reserva de Oro - 14 Yr Extra Anejo Tequila 5…	AGV	D₂O/CD₃OD	1000	650	196	38	7
122	D2	Bryan-452-H_FD	Alquima - Reserva de Don Adolfo - 6 Yr Extra Anejo Te…	AGV	D₂O/CD₃OD	1000	650	464	36	15
123	D2	Bryan-457-H_FD	LUNAZUL Anejo Tequila 40%ABV	AGV	D₂O/CD₃OD	1000	650	180	38	7
124	D2	Bryan-467-H_FD	Tullibardine 12 Year Highland Single Malt Scotch Whis…	SCW	D₂O/CD₃OD	1000	650	192	38	7
125	D2	Bryan-478-H_FD	Tin Cup SRW	AMW	D₂O/CD₃OD †	1000	650*	208	38	3
126	D2	Bryan-530-H_FD	Cask House - Triple Oak Rye Whiskey - Finished in Oak…	AMW	D₂O/CD₃OD	1000	650	196	36	7
127	D2	Bryan-531-H_FD	Bombarda Rum - Formidable - 18Yr Rum Limited Edition…	RUM	D₂O/CD₃OD	1000	650	196	38	7
128	D2	Bryan-556-H_FD	Hennessy VS Cognac 40%ABV	BRA	D₂O/CD₃OD	1000	650	196	30	7
129	D2	Bryan-601-H_FD	Starward - Two Fold - 40%ABV Aged 2 Yrs in Australian…	OWW	D₂O/CD₃OD	1000	650	128	38	15
130	D2	Bryan-602-H_FD	Pyrat XO Reserve Rum 40%ABV	RUM	D₂O/CD₃OD	1000	650	76	28	15
131	D2	Bryan-603-H_FD	Kavalan Solist Ex-Bourbon Cask 56.3%ABV	TWW	D₂O	1000	625	128	38	15
132	D2	Bryan-604-H_FD	Kavalan Solist Ex-Bourbon Cask SMW 57.1%ABV 2012	TWW	D₂O	1000	625	128	38	15
133	D2	Bryan-605-H_FD	Suntory - The Yamazaki - 12 Yr SMW 43%ABV 2013	JPW	D₂O	1000	625	128	38	15
134	D2	Bryan-606-H_FD	Highland Park 15 Yr SMW 43%ABV 2013	SCW	D₂O	1000	625	128	38	15
135	D2	Bryan-607-H_FD	Bowmore 12 Yr Islay SMW 40%ABV 2004	SCW	D₂O	1000	625	128	38	15
136	D2	Bryan-608-H_FD	Highland Park 18 Yr Orkney Island SMW 43%ABV 2003	SCW	D₂O	1000	625	128	38	15
137	D2	Bryan-609-H_FD	Highland Park 18 Yr SMW 43%ABV 2016	SCW	D₂O	1000	625	128	38	15
138	D2	Bryan-610-H_FD	Highland Park 25 Yr SMW 45.7%ABV 2013	SCW	D₂O	1000	625	128	38	15
139	D2	Bryan-611-H_FD	Redbreast 27 Yr Single Pot Still Irish Whiskey Batch#…	IRW	D₂O/CD₃OD	1000	650	128	38	15
140	D2	Bryan-612-H_FD	Redbreast 12 Yr Cask Strength Single Pot Still Irish…	IRW	D₂O/CD₃OD	1000	650	128	38	15
141	D2	Bryan-613-H_FD	Redbreast Lustau Edition - Single Pot Still Irish Whi…	IRW	D₂O/CD₃OD	1000	650	128	38	15
142	D2	Bryan-614-H_FD	Redbreast 15 Yr Single Pot Still Irish Whiskey 46%ABV…	IRW	D₂O/CD₃OD	1000	650	100	38	15
143	D2	Bryan-615-H_FD	Redbreast 21 Yr Single Pot Still Irish Whiskey 46%ABV…	IRW	D₂O/CD₃OD	1000	650	128	38	15
144	D2	Bryan-616-H_FD	Redbreast 12 Yr Single Pot Still Irish Whiskey 40%ABV…	IRW	D₂O/CD₃OD	1000	650	108	38	15
145	D2	Bryan-617-H_FD	Kavalan - "Podium" - Single Malt Whiskey 46%ABV 2014	TWW	D₂O/CD₃OD	1000	650	128	38	15
146	D2	Bryan-618-H_FD	Dewar's 18 Yr Blended Scotch Whiskey 40%ABV 2011	SCW	D₂O/CD₃OD	1000	650	128	38	15
147	D2	Bryan-619-H_FD	Dewar's Signature Blend Scotch Whiskey 43%ABV 2011	SCW	D₂O/CD₃OD	1000	650	128	38	15
148	D2	Bryan-621-H_FD	Dewar's "The Ancestor" 12 Year Blended Scotch Whiskey…	SCW	D₂O/CD₃OD	1000	650	128	38	15
149	D2	Bryan-622-H_FD	Dewar's "The Monarch" 15 Year Blended Scotch Whiskey…	SCW	D₂O/CD₃OD	1000	650	128	38	15
150	D2	Bryan-623-H_FD	Dewar's Signature Blend Scotch Whiskey 40%ABV 2015	SCW	D₂O/CD₃OD	1000	650	128	38	15
151	D2	Bryan-624-H_FD	Midleton 2023 Very Rare Irish Blended Whiskey 40%ABV	IRW	D₂O/CD₃OD	1000	650	128	38	15
152	D2	Bryan-625-H_FD	Midleton Very Rare Barry Crockett Legacy Single Pot S…	IRW	D₂O/CD₃OD	1000	650	128	38	15
153	D2	Bryan-626-H_FD	Highland Park - Viking Honor 12 Yr Single Malt Scotch…	SCW	D₂O/CD₃OD	1000	650	260	36	9
154	D2	Bryan-628-H_FD	Jameson 18 Yr Irish Blended Whiskey - 46%ABV - 2023	IRW	D₂O/CD₃OD	1000	650	512	36	9
155	D2	Bryan-632-H_FD	Knappogue Castle 12 Yr - Bourbon Barrel Matured SMIW…	IRW	D₂O/CD₃OD	1000	650	128	38	15
156	D2	Bryan-633-H_FD	Knappogue Castle 14 Yr - Non-Chilled Filtered SMIW 46…	IRW	D₂O/CD₃OD	1000	650	128	38	9
157	D2	Bryan-634-H_FD	Knappogue Castle 16 Yr -Sherry Cask Finished SMIW 43%…	IRW	D₂O/CD₃OD	1000	650	128	38	9

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
158	D2	Bryan-635-H_FD	Knappogue Castle 21 Yr -Exclusive Cask- Vatting SMIW…	IRW	D₂O/CD₃OD	1000	650	128	38	9
159	D2	Bryan-636-H_FD	Connemara 12 Yr Peated Single Malt Irish Whiskey 40%A…	IRW	D₂O/CD₃OD	1000	650	360	36	9
160	D2	Bryan-637-H_FD_2secD1	Method & Madness 14 Year Irish Single Malt Whiskey 46…	IRW	D₂O/CD₃OD	1000	600	256	38	2
161	D2	Bryan-638-H_FD_2secD1	Brenne 10 Year Ingle Malt Whiskey 48%ABV 2019	OWW	D₂O/CD₃OD	1000	600	256	38	2
162	D2	Bryan-639-H_FD_2secD1	Blue Spot 7 Year Cask Strength Single Pot Still Irish…	IRW	D₂O/CD₃OD	1000	600	256	38	2
163	D2	Bryan-640-H_FD_2secD1	Powers - Johns Lane 12 Yr Single Pot Still Irish Whis…	IRW	D₂O/CD₃OD	1000	600	256	38	2
164	D2	Bryan-641-H_FD_2secD1	Powers - Three Swallow Release Single Pot Still Irish…	IRW	D₂O/CD₃OD	1000	600	256	38	2
165	D2	Bryan-644-H_FD	Flora de Cana 4 Yr Gold Rum 40%ABV 2008	RUM	D₂O/CD₃OD	1000	650	128	38	15
166	D2	Bryan-645-H_FD	Flora de Cana 7 Yr Grand Reserve Rum 40%ABV 2010	RUM	D₂O/CD₃OD	1000	650	96	38	15
167	D2	Bryan-646-H_FD	Flor de Cana 12Yr Centenario Rum 40%ABV 2010	RUM	D₂O/CD₃OD	1000	650	128	38	15
168	D2	Bryan-647-H_FD	Flor de Cana 18Yr Centenario Gold Rum 40%ABV 2010	RUM	D₂O/CD₃OD	1000	650	128	38	15
169	D2	Bryan-648-H_FD	Bacardi Reserva Limitada Rum 40%ABV 2009	RUM	D₂O/CD₃OD	1000	650	124	28	15
170	D2	Bryan-649-H_FD	Appleton Estate 21 Yr Estate Rum 43%ABV 2014	RUM	D₂O/CD₃OD	1000	650	96	38	15
171	D2	Bryan-650-H_FD	El Jimador Tequila Anejo 40%ABV 2002	AGV	D₂O/CD₃OD	1000	650	128	38	15
172	D2	Bryan-651-H_FD	Gran Centanario Legenda Tequila Extra Anejo 40%ABV 20…	AGV	D₂O/CD₃OD	1000	650	128	38	15
173	D2	Bryan-652-H_FD_2secD1	Gran Centenario - Anejo Tequila - 40%ABV 2006	AGV	D₂O/CD₃OD	1000	600	448	38	2
174	D2	Bryan-653-H_FD	Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	BRA	D₂O/CD₃OD	1000	650	128	28	15
175	D2	Bryan-662-H_FD	English Harbour 5 Yr Antigua Rum 40%ABV - 2010	RUM	D₂O/CD₃OD	1000	650	512	30	9
176	D2	Bryan-663-H_FD_2secD1	English Harbour Reserve - 10 Yr Rum - 40%ABV 2010	RUM	D₂O/CD₃OD	1000	600	268	32	2
177	D2	Bryan-667-H_FD	Kirk & Sweeney - 12 Yr Rum - 40%ABV	RUM	D₂O/CD₃OD	1000	650	356	36	9
178	D2	Bryan-668-H_FD_1mL	Planteray Rum - Isle of Fiji 40%ABV Double-Barrel - B…	RUM	D₂O/CD₃OD	1000	650	512	26	9
179	D2	Bryan-701-H_FD_2secD1	Barley Creek Distillery - Honey Whiskey 35%ABV	AMW	D₂O/CD₃OD	1000	600	176	18	2
180	D2	Bryan-702-H_FD_2secD1	Basil Hayden - Toast - Kentucky Straight Bourbon Whis…	AMW	D₂O/CD₃OD	3000	600	500	38	2
181	D2	Bryan-703-H_FD_2secD1	Ron del Barilito - Puerto Rican Rum 43%ABV	RUM	D₂O/CD₃OD	3000	600	516	28	2
182	D2	Bryan-704-H_FD_2secD1	Glenmorangie - The Cadboll Estate - 15 Yr - 43%ABV	SCW	D₂O/CD₃OD	1000	600	460	38	2
183	D2	Bryan-JCE1-H_FD	Coppersea Distilling - Excelsior - Hudson Valley Empi…	AMW	D₂O/CD₃OD	1000	650	512	38	5
184	D2	Bryan-JCE2-H_FD	Tuthilltown Spirits - Hudson Whisky NY	AMW	D₂O/CD₃OD	1000	650	380	38	5
185	D2	Bryan-JCE3-H_FD	Bulleit 95 Rye	AMW	D₂O/CD₃OD	1000	650	256	38	5
186	D2	Bryan-JCE4-H_FD	Palmetto Whiskey	AMW	D₂O/CD₃OD	1000	650	248	38	5
187	D2	Bryan-JCE5-H_FD	Taconic Distillery - Double Barrel Maple Bourbon	AMW	D₂O/CD₃OD	1000	650	380	28	5
188	D2	BryanRum-001-H_FD	Diplomatico - Venezuala	RUM	D₂O/CD₃OD †	1000	650*	128	34	3
189	D2	BryanRum-002-H_FD	Flor de Cana Rum - 7 Yrs - Nicaragua 40%ABV	RUM	D₂O/CD₃OD	1000	650	216	38	15
190	D2	BryanRum-004-H_FD	The Cane Hill - Barbados - 5 Years	RUM	D₂O/CD₃OD †	1000	650*	128	34	3
191	D2	Bryan-Rum-006-H_FD	Trinidad - 16 Years	RUM	D₂O/CD₃OD †	1000	650*	452	28	5
192	D2	BryanRum-007-H_FD	J.Wray - Gold Rum - Jamaica	RUM	D₂O/CD₃OD †	1000	650*	256	34	3
193	D2	BryanRum-008-H_FD	Roatan Rum Company - Nicaragua	RUM	D₂O/CD₃OD †	1000	650*	112	22	3
194	D2	BryanRum-009-H_FD	Cadenheads - Classic Rum Blend - 5Yr - Barbados-Guyan…	RUM	D₂O/CD₃OD †	1000	650*	148	38	3
195	D2	BryanRum-010-H_FD	Brugel - Anejo Superior - Dominican Republic	RUM	D₂O/CD₃OD †	1000	650*	132	38	3
196	D2	BryanRum-011-H_FD	Meyers - Plantation White - Jamaica	RUM	D₂O/CD₃OD †	1000	650*	256	34	3
197	D2	BryanRum-012-H_FD	Flor de Cana - 12 Years - Nicaragua	RUM	D₂O/CD₃OD †	1000	650*	132	38	3
198	D2	Bryan-Rum28-001-H_FD	Rum 28 - 60%ABV	RUM	D₂O/CD₃OD	1000	700	256	38	7
199	D2	BryanRum-029-H_FD	Mount Gay - Barbados Rum 1703 XO 43%ABV 105 Cal (44mL…	RUM	D₂O	1000	625	128	38	15
200	D2	Glenmorangie-003-H_FD	Glenmorangie - The Lasanta - 43%ABV - 12 Year - Bourb…	SCW	D₂O/CD₃OD	1000	650	192	38	8
201	D2	Laphroaig-001-H_FD	Laphroaig - Cask Collection- Oak Select - Islay Singl…	SCW	D₂O/CD₃OD	1000	650	256	38	7
202	D2	Tuthilltown-001-H_1mL_FD_…	Hudson Whiskey NY - Cask Strength Barrel Select - Thi…	AMW	D₂O/CD₃OD	1000	650	256	38	2
203	D2	Tuthilltown-002-H_1mL_FD_…	Hudson Whiskey NY - Short Stack - NY SRW - Finished i…	AMW	D₂O/CD₃OD	1000	650	256	34	2
204	D2	Tuthilltown-003-H_1mL_FD_…	Hudson Whiskey NY - Do The Rye Thing - NY SRW 46%ABV	AMW	D₂O/CD₃OD	1000	650	256	38	2
205	D2	Tuthilltown-004-H_1mL_FD_…	Hudson Whiskey NY - Single Barrel Cask Strength - in…	AMW	D₂O/CD₃OD	1000	650	256	38	2
206	D2	Tuthilltown-005-H_1mL_FD_…	Hudson Whiskey NY - This is Your Favorite Summer Rye…	AMW	D₂O/CD₃OD	1000	650	256	38	2
207	D2	Tuthilltown-006-H_1mL_FD_…	Hudson Whiskey NY - Bright Lights Big Bourbon - NY SB…	AMW	D₂O/CD₃OD	1000	650	256	38	2

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
208	D2	Tuthilltown-007-H_1mL_FD_…	Hudson Whiskey NY - Back Room Deal - NY SBW Aged in P…	AMW	D2O/CD3OD	1000	650	256	38	2
209	D2	Tuthilltown-008-H_1mL_FD_…	Hudson Whiskey NY - Four Part Harmony - NY 4 Grain BW…	AMW	D2O/CD3OD	1000	650	256	38	2
210	D2	Tuthilltown-009-H_1mL_FD_…	Hudson Whiskey NY - Port Cask Finish - NY RW - Aged M…	AMW	D2O/CD3OD	1000	650	256	34	2
211	D2	Tuthilltown-010-H_1mL_FD_…	Hudson Whiskey NY - Calvados Cask Finish - NY RW - Ag…	AMW	D2O/CD3OD	1000	650	256	38	2
212	D2	Tuthilltown-011-H_1mL_FD_…	Hudson Whiskey NY - Sherry Cask Finish - NY BW - Aged…	AMW	D2O/CD3OD	1000	650	256	38	2
213	D2	MakersMarkNo46-001-H_FD	Maker's Mark No. 46 - French Oaked - 47%ABV	AMW	D2O/CD3OD	2000	700	128	38	10
214	C1	Bryan-019-H_FD_CDCl3	Bushmills - Black Bush Irish Whisky 40%ABV	IRW	CDCl3	1000	650*	204	38	5
215	C1	Bryan-043-H_FD_CDCl3	Kavalan Solist Vinho Barrique Single Cask Strength Si…	TWW	CDCl3	1000	650*	184	38	5
216	C1	Bryan-103-H_FD_CDCl3	An Dulaman - Santa Ana Armada Strength Barrel Aged Gin	GIN	CDCl3	1000	650*	172	38	5
217	C1	Bryan-124-H_FD_CDCl3	Medford Rum - 3 Years Aged	RUM	CDCl3	1000	650*	256	38	5
218	C1	Bryan-142-H_FD_CDCl3	Kavalan Earth Silver Wine Cask Matured Single Cask Si…	TWW	CDCl3	1000	650*	256	38	5
219	C1	Bryan-172-H_FD_CDCl3	Spirit of HVEN Mercurious Corn Whisky 45.6%ABV	OWW	CDCl3	1000	650*	68	38	5
220	C1	Bryan-178-H_FD_CDCl3	Red Spot 15 Year Single Pot Still Irish Whiskey 46%ABV	IRW	CDCl3	1000	650*	160	38	5
221	C1	Bryan-310-H_FD_CDCl3	Makers Mark Single Batch Bourbon Whiskey	AMW	CDCl3	1000	650*	512	38	5
222	C1	Bryan-311-H_FD_CDCl3	Pendleton 1910 - 12 Year Old Canadian Rye Whiskey 40%…	CAN	CDCl3	1000	650*	256	38	5
223	C1	Bryan-325-H_FD_CDCl3	Coalition Whiskey Barrel Proof Kentucky SRW 54.4%ABV	AMW	CDCl3	1000	650*	256	38	5
224	C1	Bryan-331-H_FD_CDCl3	Bernheim - 7 Year Old Small Batch Straight Wheat Whis…	AMW	CDCl3	1000	650*	256	38	5
225	C1	Bryan-334-H_FD_CDCl3	Bardeney - Bottled in Bond Rye Whiskey	AMW	CDCl3	1000	650*	512	38	5
226	C1	Bryan-336-H_FD_CDCl3	TX - American Blended Whisky	AMW	CDCl3	1000	650*	96	38	5
227	C1	Bryan-343-H_CDCl3	Kavalan - Distillery Reserve - Rum Cask Single Cask S…	TWW	CDCl3	1000	650*	332	38	5
228	C1	Bryan-346-H_FD_CDCl3	Kavalan Sky Gold Wine Cask Matured Single Cask Single…	TWW	CDCl3	1000	650*	144	38	15
229	C1	Bryan-368-H_FD_CDCl3	Makers Mark Cask Strength Single Malt Whisky 55%ABV	AMW	CDCl3	1000	650*	228	38	5
230	C1	Bryan-369-H_FD_CDCl3	Old Overholt - 114 Proof Straight Rye Whiskey	AMW	CDCl3	1000	650*	384	38	5
231	C1	Bryan-371-H_FD_CDCl3	Makers Mark 101 Single Malt Whisky 50.5%ABV	AMW	CDCl3	1000	650*	60	38	5
232	C1	Bryan-386-H_FD_CDCl3	Doc Swinson's - ALter Ego Solera Method SRW Release#…	AMW	CDCl3	1000	650*	204	38	5
233	C1	Bryan-391-H_FD_CDCl3	Southern Star Double Rye Whiskey	AMW	CDCl3	1000	650*	232	38	5
234	C1	Bryan-428-H_FD_CDCl3	Bespoke Spirits Rye Whiskey 2020-1	AMW	CDCl3	1000	650*	156	38	5
235	C1	Bryan-430-H_FD_CDCl3	High West - Double Rye! Blended Straight Rye Wiskey B…	AMW	CDCl3	1000	650*	252	38	5
236	C1	Bryan-435-H_FD_CDCl3	High West - A Mid-Winters Night Dram Act 7 Scene 5 Bl…	AMW	CDCl3	1000	650*	240	38	5
237	C1	Bryan-398-H_FD_CDCl3	Tin Cup Straight Rye Whiskey	AMW	CDCl3	1000	650*	252	38	5
238	C1	Bryan-Kavalan-011-H_FD_CD…	Kavalan - Classic - SMW	TWW	CDCl3	1000	650*	164	38	5
239	C1	Bryan-Kavalan-034-H_FD_CD…	Oloroso - Sherry Oak SMW	TWW	CDCl3	1000	650*	256	38	5
240	C1	Bryan-Kavalan-035-H_FD_CD…	Solist Madeira SCS SMW	TWW	CDCl3	1000	650*	164	38	5
241	C1	Bryan-Kavalvan-005-H_FD_C…	Kavalan - Solist Oloroso Sherry Single Cask Strength…	TWW	CDCl3	1000	650*	256	38	5
242	C1	Bryan-Kavalvan-026-H_FD_C…	Kavalan - ArtistSeries: Paul Chiang Peated MaltSingle…	TWW	CDCl3	1000	650*	256	38	5
243	C1	Bryan-603-H_FD_CDCl3	Kavalan Solist Ex-Bourbon Cask 56.3%ABV	TWW	CDCl3	1000	650	256	38	5
244	C1	Bryan-604-H_FD_CDCl3	Kavalan Solist Ex-Bourbon Cask SMW 57.1%ABV 2012	TWW	CDCl3	1000	650	256	38	5
245	C1	Bryan-608-H_FD_CDCl3	Highland Park 18 Yr Orkney Island SMW 43%ABV 2003	SCW	CDCl3	1000	650	256	38	5
246	C1	Bryan-609-H_FD_CDCl3	Highland Park 18 Yr SMW 43%ABV 2016	SCW	CDCl3	1000	650	256	38	5
247	C1	Bryan-610-H_FD_CDCl3	Highland Park 25 Yr SMW 45.7%ABV 2013	SCW	CDCl3	1000	650	256	38	5
248	C1	Bryan-611-H_FD_CDCl3	Redbreast 27 Yr Single Pot Still Irish Whiskey Batch#…	IRW	CDCl3	1000	650	256	38	5
249	C1	Bryan-614-H_FD_CDCl3	Redbreast 15 Yr Single Pot Still Irish Whiskey 46%ABV…	IRW	CDCl3	1000	650	256	38	5
250	C1	Bryan-JCE1-H_FD_CDCl3	Coppersea Distilling - Excelsior - Hudson Valley Empi…	AMW	CDCl3	1000	650*	256	38	5
251	C1	Bryan-JCE2-H_FD_CDCl3	Tuthilltown Spirits - Hudson Whisky NY	AMW	CDCl3	1000	650*	256	38	5
252	C1	Bryan-JCE3-H_FD_CDCl3	Bulleit 95 Rye	AMW	CDCl3	1000	650*	256	38	5
253	C2	Bryan-653-H_FD_CDCl3	Pierre Ferrand "Ambre" XO Cognac 40%ABV 2004	BRA	CDCl3	1000	650	16	38	5

#	Set	Sample ID	Product (title as recorded)	Class	Solvent	V_{spirit} (µL)	V_{soln} (µL)	Scans	Gain	d_1
254	C2	Bryan-702-H_FD_CDCl3_1000…	Basil Hayden - Toast - Kentucky Straight Bourbon Whis…	AMW	$CDCl_3$	1000	620	256	38	3
255	C2	Bryan-703-H_FD_CDCl3_1000…	Ron del Barilito - Puerto Rican Rum 43%ABV	RUM	$CDCl_3$	1000	620	512	38	3
256	C2	Bryan-704-H_FD_CDCl3_1000…	Glenmorangie - The Cadboll Estate - 15 Yr - 43%ABV	SCW	$CDCl_3$	1000	650*	392	38	3
257	C2	Bryan-Rum28-001-H_CDCl3	Rum 28 - 60%ABV	RUM	$CDCl_3$	1000	650*	512	38	5
258	C2	Glenmorangie-003-H_FD_CDC…	Glenmorangie - The Lasanta - 43%ABV - 12 Year - Bourb…	SCW	$CDCl_3$	1000	650*	244	38	5
259	C2	Laphroaig-H_FD_CDCl3	Laphroaig - Cask Collection- Oak Select - Islay Singl…	SCW	$CDCl_3$	1000	650*	256	38	5
260	C2	PowdermenBourbon-124-H_FD…	Powdermen Distillery - Black Wick Bourbon Whisky 51%A…	AMW	$CDCl_3$	1000	620	256	38	5
261	C2	Tuthilltown-001-H_1mL_FD_…	Hudson Whiskey NY - Cask Strength Barrel Select - Thi…	AMW	$CDCl_3$	1000	680	192	38	4
262	C2	Tuthilltown-002-H_1mL_FD_…	Hudson Whiskey NY - Short Stack - NY SRW - Finished i…	AMW	$CDCl_3$	1000	680	180	38	4
263	C2	Tuthilltown-003-H_1mL_FD_…	Hudson Whiskey NY - Do The Rye Thing - NY SRW 46%ABV	AMW	$CDCl_3$	1000	680	192	38	4
264	C2	Tuthilltown-004-H_1mL_FD_…	Hudson Whiskey NY - Single Barrel Cask Strength - in…	AMW	$CDCl_3$	1000	680	192	38	4
265	C2	Tuthilltown-005-H_1mL_FD_…	Hudson Whiskey NY - This is Your Favorite Summer Rye…	AMW	$CDCl_3$	1000	680	192	38	4
266	C2	Tuthilltown-006-H_1mL_FD_…	Hudson Whiskey NY - Bright Lights Big Bourbon - NY SB…	AMW	$CDCl_3$	1000	680	180	38	4
267	C2	Tuthilltown-007-H_1mL_FD_…	Hudson Whiskey NY - Back Room Deal - NY SBW Aged in P…	AMW	$CDCl_3$	1000	680	192	38	4
268	C2	Tuthilltown-008-H_1mL_FD_…	Hudson Whiskey NY - Four Part Harmony - NY 4 Grain BW…	AMW	$CDCl_3$	1000	680	192	38	4
269	C2	Tuthilltown-009-H_1mL_FD_…	Hudson Whiskey NY - Port Cask Finish - NY RW - Aged M…	AMW	$CDCl_3$	1000	680	192	38	4
270	C2	Tuthilltown-010-H_1mL_FD_…	Hudson Whiskey NY - Calvados Cask Finish - NY RW - Ag…	AMW	$CDCl_3$	1000	680	192	38	4
271	C2	Tuthilltown-011-H_1mL_FD_…	Hudson Whiskey NY - Sherry Cask Finish - NY BW - Aged…	AMW	$CDCl_3$	1000	680	192	38	4
272	C2	Bryan-615-H_FD_CDCl3	Redbreast 21 Yr Single Pot Still Irish Whiskey 46%ABV…	IRW	$CDCl_3$	1000	650	256	38	5
273	C2	Bryan-617-H_FD_CDCl3	Kavalan - "Podium" - Single Malt Whiskey 46%ABV 2014	TWW	$CDCl_3$	1000	650	256	38	5
274	C2	MakersMarkNo46-H_FD_CDCl3	Maker's Mark No. 46 - French Oaked - 47%ABV	AMW	$CDCl_3$	1000	650*	256	38	5

† solvent system corrected from the title by the CHD_2OD audit (Section 5.4). * solution volume not recorded; 650 µL assumed.

The 29 DMSO-d_6 residues are listed separately in Table 30, with the sample identifier as corrected in Section 9.9, the cask, the date of acquisition, the transient count and receiver gain, the volumes of spirit and solution, the residual DMSO-d_5 found in the tube, the method by which the spectrum was referenced (Section 5.8) and the D_2O/CD_3OD and $CDCl_3$ residues of the same product used in the comparisons of Sections 9.8, 12.9 and 12.10. Every DMSO-d_6 spectrum was acquired with a 7 µs (48.5°) pulse and a repetition time of 7.27 s.

Table 30. DMSO-d_6 residues: identity, acquisition, residual DMSO-d_5, referencing and matched residues in the other solvents.

No.	Sample	Product	Cask	Acquired	Transients/gain (dB)	Spirit (mL)/solution (µL)	DMSO-d_5 (mM ¹H)	Reference	D_2O residue	$CDCl_3$ residue
1	Tuthilltown-006	Hudson Whiskey NY - Bright Lights Big Bou…	—	Apr 6 2026	256/38	1.0/600*	154	xcorr	207	266
2	Tuthilltown-007	Hudson Whiskey NY - Back Room Deal - NY S…	peated-Scotch barrel	Apr 7 2026	256/34	1.0/600*	138	xcorr	208	267
3	Tuthilltown-008	Hudson Whiskey NY - Four Part Harmony - N…	—	Apr 7 2026	256/34	1.0/600*	132	xcorr	209	268
4	Tuthilltown-009	Hudson Whiskey NY - Port Cask Finish - NY…	port	Apr 7 2026	256/30	1.0/600*	160	xcorr	210	269
5	Bryan-023	Glen Grant 12 Yr Speyside SMW 43%ABV	—	Apr 10 2026	512/38	1.0/600*	56	TMS	18	—
6	Bryan-604	Kavalan - Solist Ex-Bourbon Cask SMW 57.1…	ex-bourbon	Apr 10 2026	48/38	1.0/600*	57	TMS	132	244
7	Kavalan-001	Kavalan - Solist Amontillado Sherry SCS S…	sherry (amontillado)	Apr 21 2026	128/38	1.0/600	65	xcorr	82	—
8	Kavalan-005	Kavalan - Solist Oloroso Sherry Single Ca…	sherry (oloroso)	Apr 22 2026	128/34	1.0/600	54	TMS	81	241
9	Kavalan-006	Kavalan - Concertmaster Sherry Cask Finis…	sherry	Apr 21 2026	128/38	1.0/600	60	xcorr	83	—
10	Kavalan-007	Kavalan - King Car - Conductor - SMW	—	Apr 22 2026	128/38	1.0/600	60	xcorr	84	—
11	Kavalan-008	Kavalan - Distillery Reserve Peaty Cask…	peated malt	Apr 22 2026	128/38	1.0/600	63	xcorr	85	—

No.	Sample	Product	Cask	Acquired	Transients/gain (dB)	Spirit (mL)/solution (µL)	DMSO-d₅ (mM ¹H)	Reference	D₂O residue	CDCl₃ residue
12	Kavalan-010	Solist Brandy Single Cask Strength Single…	brandy	Apr 24 2026	128/38	1.0/600	65	TMS	86	—
13	Kavalan-011	Kavalan - Classic - SMW	—	Apr 22 2026	128/38	1.0/600	60	xcorr	87	238
14	Kavalan-017	Kavalan - Solist Ex-Bourbon Sigle Cask St…	ex-bourbon	Apr 24 2026	128/38	1.0/600	58	TMS	88	—
15	Kavalan-021	Kavalan - Artist Series-Paul Chiang Virgi…	virgin oak	Apr 24 2026	128/38	1.0/600	60	TMS	—	—
16	Kavalan-034	Kavalan Solist Oloroso Sherry Oak SMW †	sherry (oloroso)	Apr 22 2026	128/38	1.0/600	64	xcorr	91	239
17	Kavalan-029	Solist Moscatel Sherry - Single Cask Stre…	sherry (moscatel)	Apr 22 2026	128/32	1.0/600	59	xcorr	90	—
18	Kavalan-035	Solist Madeira Single Cask Strength - SMW	madeira	Apr 22 2026	128/38	1.0/600	57	xcorr	92	240
19	Kavalan-037	Kavalan - Ex-Bourbon Oak - SMW	ex-bourbon	Apr 24 2026	128/38	1.0/600	57	TMS	93	—
20	Kavalan-038	Kavalan - Solist Fino Sherry Single Cask…	sherry (fino)	Apr 22 2026	128/38	1.0/600	64	xcorr	94	—
21	Kavalan-042	Kavalan - Artist Series Paul Chiang Punch…	puncheon	Apr 22 2026	128/38	1.0/600	73	xcorr	95	—
22	Kavalan-043	Kavalan Solist Vinho Barrique SCS SMW 57.…	wine (vinho barrique)	Apr 10 2026	512/38	1.0/600*	59	TMS	22	215
23	Kavalan-045	Solist Manzanilla Sherry - Single Cask St…	sherry (manzanilla)	Apr 22 2026	128/38	1.0/600	64	xcorr	96	—
24	Kavalan-046	Kavalan - ConcertMaster Port Cask Finish…	port	Apr 24 2026	128/38	1.0/600	61	TMS	—	—
25	Kavalan-047	Kavalan - Podium - SMW	—	Apr 24 2026	128/38	1.0/600	59	TMS	97	—
26	Kavalan-143	Kavalan - King Car - 40th Anniversary Sel…	wine	Apr 23 2026	128/38	1.0/600	67	xcorr	98	—
27	Adirondack-Maple	Adirondack Maple Barrel Bourbon 45%ABV Ba…	—	Sep 4 2026	256/18	3.0/600	142	TMS	—	—
28	Adirondack-Rum	Adirondack Rum 73%ABV	—	Sep 3 2026	256/38	3.0/600	146	TMS	—	—
29	TenMile	TenMile Distillery - Raid on Richmond - A…	—	Oct 16 2026	320/36	1.0/610	170	TMS	—	—

* volume not recorded, 600 µL assumed. † title conflict resolved by chemistry (Section 9.9). xcorr: referenced by cross-correlation with the TMS-referenced median spectrum.

Appendix B. Concentrations by residue

Table 31 lists the concentrations of the principal congeners in every D_2O residue and Table 32 those in every $CDCl_3$ residue, in mg L^{-1} of the original spirit (glucose, fructose and sucrose in g L^{-1} where indicated). "nd" denotes a result below the detection limit, "tr" a result between the detection and quantitation limits, "nq" a result not quantified because of overlap, and values in parentheses are upper bounds. Semi-volatile congeners are concentrations in the residue as recovered, not in the spirit. The complete results, including detection limits, fitted line positions and widths, the saturation corrections and their sensitivity to T_1, are given in the companion workbook.

Table 31. Concentrations in the D_2O residues (sugars in g L^{-1}, other congeners in mg L^{-1} of spirit).

#	Product	Class	Glc	Fru	Suc	Gly	PG	Querc	Lac	Succ	Gall	HMF	Benz	Cinn
0	Tromba Extra Anejo Tequila	AGV	0.09	0.13	nd	36.1	0.56	64.4	332	3.8	4.2	2.4	28.0	2.6
1	Clase Azul Gold	AGV	0.02	<0.03	0.00	(2.3)	0.59	7.5	9.9	0.25	0.86	0.58	1.6	tr
2	Clase Azul Anejo	AGV	0.03	0.03	0.02	(8.8)	1.7	6.2	5.8	0.47	0.69	0.60	4.5	tr
3	Bozal Cuishe	AGV	0.01	<0.03	0.01	4.1	1.3	nd	nd	0.19	tr	0.39	tr	nd
4	Sotol Carnay Tres Pilares	AGV	nd	<0.03	0.00	nd	0.61	tr	nd	0.14	tr	nd	1.4	tr
5	Sotol Ono Chihuahua	AGV	nd	<0.03	nd	(tr)	tr	nd	tr	0.47	1.1	tr	nd	tr
6	Sotol Ono Chihuahua	AGV	nd	<0.03	nd	nq	1.4	nd	nd	tr	tr	nd	tr	tr
7	Litchfield Distilling - Maple Bourb…	AMW	0.43	0.22	29.20	nq	190	46.0	26.7	1.9	tr	2.3	12.4	4.0
8	Litchfield Distilling - Straight Ry…	AMW	0.03	<0.03	0.01	(11.4)	0.30	61.0	20.7	2.0	3.0	1.8	12.7	5.9
9	Hudson House - NY Straight Rye Whis…	AMW	0.05	0.03	0.01	(24.8)	tr	79.3	7.3	0.58	nd	4.8	14.7	10.0
10	Jameson - Blender's Dog Irish Whisk…	IRW	0.01	<0.03	0.01	23.4	tr	30.8	0.88	0.46	tr	0.73	6.0	tr
11	Powers - Gold Label Single Pot Stil…	IRW	0.10	0.11	0.00	6.6	nd	9.9	nd	0.65	0.51	7.1	2.3	tr
12	Bushmills 10 Yr Irish Single Malt W…	IRW	0.04	0.04	nd	31.4	tr	25.5	2.8	0.80	2.3	2.4	2.0	nd
13	Powers - Gold Label Single Pot Stil…	IRW	0.01	<0.03	nd	(2.0)	0.33	nd	nd	tr	nd	3.2	tr	1.7
14	Bushmills - Black Bush Irish Whisky…	IRW	0.06	0.04	nd	(29.5)	0.20	13.4	4.1	0.88	1.4	5.6	2.8	tr
15	Tromba Reposado - French Oak Casks	AGV	0.33	3.54	0.07	nd	12.4	8.8	55.1	0.37	4.6	42.4	4.4	13.4
16	Tromba Anejo Tequila	AGV	0.01	<0.03	nd	5.1	0.51	9.9	3.2	tr	0.49	tr	tr	tr
17	Bushmills Original Irish Whiskey	IRW	0.07	0.07	nd	(19.6)	nd	14.7	nd	0.12	1.4	2.8	2.9	nd
18	Glen Grant - 12 Year Speyside SMW	SCW	0.03	<0.03	nd	(15.7)	tr	30.6	6.7	0.68	2.5	1.1	5.8	nd
19	Brenne Estate Cask Single Malt Fren…	OWW	0.02	0.03	nd	22.5	2.6	2.4	nd	0.33	tr	6.4	1.6	tr
20	Jameson - Cooper's Croze Irish Whis…	IRW	0.03	<0.03	nd	37.6	0.31	47.6	nd	1.1	3.6	1.0	9.0	tr
21	Red Breast - 12 Yr Cask Strength Ir…	IRW	0.04	0.04	nd	117	0.37	79.8	1.1	3.5	7.3	0.70	10.0	tr
22	Kavalan Solist Vinho Barrique Singl…	TWW	0.08	0.08	nd	114	5.4	140	5.5	13.9	15.6	1.7	24.5	4.2
23	Courvoisier VS Cognac	BRA	2.50	2.37	1.11	nq	6.2	15.2	23.7	0.91	1.8	17.4	4.5	18.0
24	Merlet - XO Cognac	BRA	2.29	2.38	0.05	nq	6.8	66.9	48.4	2.5	11.0	2.0	4.1	tr
25	Branson Cognac VSOP - Cognac Grande…	BRA	5.24	5.42	1.04	nq	10.5	35.8	72.5	4.6	6.2	0.78	tr	nd
26	Courvoisier VSOP Cognac 40%ABV	BRA	2.92	3.09	0.30	nq	5.6	28.7	41.2	1.9	3.7	21.9	3.7	18.2
27	Branson - Phantom VS Cognac	BRA	4.39	4.34	1.68	nq	11.8	33.8	57.1	3.5	4.8	9.8	1.5	6.1
28	Branson - Royal VSOP Cognac	BRA	5.39	5.39	0.59	nq	13.9	69.7	56.1	2.4	9.8	4.5	2.3	tr
29	Courvoisier - XO Cognac	BRA	3.24	3.33	nd	nq	11.4	60.5	57.7	4.3	7.0	22.4	7.7	18.0
30	Branson Cognac - XO Prmier Cru Cogn…	BRA	5.73	5.98	0.25	nq	15.6	83.8	50.5	3.7	11.2	1.7	4.5	tr
31	Bisquit and Debouche - VSOP Cognac	BRA	2.60	2.69	0.39	nq	8.0	32.3	54.7	4.6	4.6	14.6	3.8	4.9
32	Larsen - Aqua Ignis Cognac	BRA	5.63	5.55	2.65	nq	21.9	83.3	46.2	6.4	9.2	6.8	27.5	5.1
33	Branson Cognac - XO Prmier Cru Cogn…	BRA	2.91	2.97	0.44	nq	7.6	49.4	54.8	4.3	8.3	26.2	2.4	18.8
34	Coperies - Grench Single Malt Whisk…	OWW	1.12	2.20	nd	nq	46.6	40.0	13.4	15.1	4.9	11.3	7.3	7.3
35	Waterford - Single Farm Origin - Ra…	IRW	0.06	0.05	nd	54.5	0.19	32.1	3.4	3.3	3.1	1.4	6.5	3.2
36	Tenjaku - Pure Japanese Malt Whisky…	JPW	0.10	0.10	nd	17.2	0.36	42.1	0.46	1.7	3.9	2.4	15.2	4.1
37	Bhakta Spirits 50 Yr Armagnac Brand…	BRA	0.23	0.18	nd	41.6	8.8	85.1	47.3	7.0	9.7	2.3	5.2	tr
38	Amrut - Fusion - Indian Single Malt…	OWW	0.01	<0.03	0.00	(3.4)	0.52	6.1	nd	0.33	tr	0.88	nd	2.4
39	An Dulaman - Santa Ana Armada - Bar…	GIN	nd	<0.03	nd	146	7.3	tr	18.3	8.8	1.1	0.68	tr	tr
40	Kiss - Black Diamond Premium Dark R…	RUM	3.62	3.50	9.54	nq	38.3	2.3	20.6	4.1	0.98	25.5	1.1	tr
41	Medford Rum - Aged 3 Years	RUM	0.02	<0.03	nd	(5.8)	tr	42.4	0.58	1.9	4.1	2.5	18.6	11.2
58	Glen Grant 18 Yr Speyside Single Ma…	SCW	0.03	<0.03	nd	(9.2)	nd	48.2	nd	1.6	3.0	tr	11.0	nd
59	Kavalan Earth Silver Wine Cask Matu…	TWW	0.10	0.06	nd	163	8.0	223	6.9	23.1	14.5	1.2	2.3	tr
60	Redbreast - Single Pot Still Irish…	IRW	0.33	0.40	0.01	106	4.6	107	4.4	6.3	17.7	1.4	10.7	nd
61	Stauning - Kaos - Triple Malt Whisky	OWW	0.02	<0.03	nd	(8.6)	0.28	45.7	2.8	0.16	0.96	2.3	12.3	7.2
62	Spirit of HVEN Mercurious Corn Whis…	OWW	0.04	<0.03	nd	(8.0)	0.43	87.4	0.47	nd	nd	2.0	15.0	3.4

#	Product	Class	Glc	Fru	Suc	Gly	PG	Querc	Lac	Succ	Gall	HMF	Benz	Cinn
63	Red Spot - 15 Year Old Single Pot S…	IRW	0.43	0.50	nd	nq	3.4	73.4	3.2	4.7	10.0	7.0	8.5	2.6
64	Aberlour - 16 Year Speyside - Singl…	SCW	0.05	0.03	nd	71.9	1.9	51.8	7.6	4.0	7.6	3.0	7.0	tr
65	Jameson - Distillers Safe Irish Whi…	IRW	0.01	<0.03	nd	(4.5)	nd	26.9	nd	tr	0.85	tr	6.0	tr
66	Aberlour A'Bunadh Alba Speyside Sin…	SCW	0.02	<0.03	nd	(4.7)	0.58	40.9	nd	tr	tr	tr	8.0	nd
67	Makers Mark Straight Bourbon Whiske…	AMW	0.04	<0.03	nd	(9.4)	0.39	102	4.5	nd	nd	1.7	15.7	2.7
68	Pendleton 1910 - 12 Year Old Canadi…	CAN	0.40	0.35	nd	nq	128	7.5	4.8	0.39	0.93	30.2	2.5	tr
69	TX - Texas Straight Bourbon Whisky	AMW	0.04	<0.03	nd	(22.3)	2.3	88.2	6.9	1.5	6.8	2.0	13.9	5.9
70	Coalition Whiskey Barrel Proof Kent…	AMW	0.05	<0.03	nd	(20.7)	0.39	135	8.5	3.3	11.4	3.6	19.7	3.3
71	Bernheim - 7 Year Old Small Batch S…	AMW	0.06	0.04	nd	(10.2)	0.20	110	0.33	2.6	3.4	2.6	18.1	6.1
72	Bardeney - Bottled in Bond Rye Whis…	AMW	0.05	<0.03	nd	(12.8)	0.52	97.5	22.6	1.5	5.4	1.8	22.7	2.5
73	TX - American Blended Whisky	AMW	0.18	0.14	nd	nq	819	29.1	3.2	0.15	tr	8.1	13.9	tr
75	Kavalan-Distillery Reserve-Rum Cask…	TWW	0.04	<0.03	0.01	(27.7)	tr	53.8	nd	2.1	5.0	tr	13.4	tr
76	Kavalan Sky Gold Wine Cask Matured…	TWW	0.07	0.04	0.07	219	14.8	198	8.0	39.2	11.0	1.1	9.9	nd
77	Jameson - Bow Street 18 Yr Old Cask…	IRW	0.10	0.08	nd	55.0	1.8	74.7	2.7	3.8	6.3	7.1	19.3	2.6
78	Mount Gay - Black Barrel Double Cas…	RUM	0.12	0.10	nd	9.0	0.28	41.7	1.3	2.0	3.1	10.8	13.3	tr
79	Mount Gay XO Triple Cask Rum Blend…	RUM	0.29	0.35	nd	16.7	0.20	54.5	2.8	2.8	5.3	21.9	16.1	5.7
80	Makers Mark Cask Strength Single Ma…	AMW	0.06	0.03	nd	(12.7)	0.76	129	2.9	1.6	4.0	2.0	17.1	3.4
81	Kavalan - Solist Oloroso Sherry Sin…	TWW	0.02	<0.03	nd	102	10.1	19.4	2.1	25.2	11.8	1.9	16.9	tr
82	Kavalan - Solist Amontillado Sherry…	TWW	0.18	0.14	nd	589	41.6	77.4	41.6	51.9	10.5	5.6	25.6	nd
83	Kavalan - Concertmaster Sherry Cask…	TWW	0.07	0.04	nd	51.9	0.35	64.0	1.7	2.8	4.2	0.91	14.4	tr
84	Kavalan - King Car - Conductor SMW	TWW	0.08	0.06	nd	76.7	2.0	58.2	nd	5.6	3.9	3.0	18.4	tr
85	Kavalan - Distillery Reserve Peaty…	TWW	0.02	<0.03	0.01	14.4	tr	18.3	nd	1.2	1.8	tr	4.2	tr
86	Kavalan Solist Brandy cask †	TWW	0.83	0.79	0.00	97.0	17.0	22.1	2.1	2.0	8.6	3.6	3.8	2.2
87	Kavalan - Classic - SMW	TWW	0.09	0.11	nd	68.0	2.7	51.7	5.4	5.3	4.6	1.6	16.6	nd
88	Solist Ex-Bourbon - Single Cask Str…	TWW	0.07	0.04	nd	(26.9)	0.61	102	nd	3.6	7.2	1.8	24.9	tr
89	Solist Port - Single Cask Strength…	TWW	0.73	0.49	nd	404	38.9	123	42.9	49.7	19.3	3.8	18.6	tr
90	Solist Moscatel Sherry - Single Cas…	TWW	2.67	2.69	nd	337	46.1	77.6	20.9	48.2	11.2	6.3	12.7	tr
91	Oloroso Sherry Oak - SMW	TWW	0.12	0.07	nd	431	13.5	60.2	34.7	53.6	12.0	3.3	14.2	tr
92	Solist Madeira Single Cask Strength…	TWW	0.23	0.25	nd	347	30.5	72.1	21.2	72.6	17.3	2.8	22.4	tr
93	Ex-Bourbon Oak SMW	TWW	0.04	<0.03	nd	(24.9)	tr	58.8	nd	2.0	2.7	0.82	16.9	tr
94	Kavalan - Solist Fino Sherry Single…	TWW	0.12	0.12	nd	583	42.3	71.5	57.2	68.2	15.6	5.2	33.6	tr
95	Kavalan - Artist Series Paul Chiang…	TWW	0.10	0.04	0.01	(33.6)	0.47	73.4	1.7	6.0	19.3	1.7	5.1	tr
96	Solist Manzanilla Sherry - Single C…	TWW	0.10	<0.03	0.02	211	12.1	89.9	10.9	20.5	24.0	3.2	tr	tr
97	Podium - SMW	TWW	0.07	0.05	nd	54.4	0.24	71.6	nd	4.1	5.9	3.3	18.6	2.8
98	Kavalan - King Car - 40th Anniversa…	TWW	0.06	0.04	nd	340	14.6	119	17.7	41.6	7.1	1.2	17.2	tr
99	Kavalan - Distillery Reserve - Rum…	TWW	0.04	<0.03	0.00	(31.0)	tr	65.5	nd	2.3	6.7	0.67	12.0	tr
100	Kavalan - Distillery Select No. 1 S…	TWW	0.04	0.07	nd	32.0	0.50	44.4	1.6	2.7	4.8	1.2	9.2	tr
101	Old Overholt 114 Proof SRW	AMW	0.05	<0.03	nd	(21.9)	0.28	110	1.1	1.6	0.41	1.9	19.5	2.2
102	Bare Knuckle - Straight Wheat Whisk…	AMW	0.04	0.03	nd	(12.3)	0.46	81.0	17.4	2.9	4.7	4.1	17.7	7.1
103	Makers Mark 101 Single Malt Whisky…	AMW	0.05	<0.03	nd	(12.4)	1.3	118	4.9	1.5	4.6	1.9	12.9	4.4
104	Doc Swinson's - Alter Ego Solera Me…	AMW	0.04	<0.03	nd	(6.7)	0.32	77.5	nd	1.1	3.7	1.7	13.0	5.9
105	Southern Star - Double Rye Whisky…	AMW	0.04	<0.03	nd	(10.4)	0.34	83.9	12.3	nd	4.6	2.2	16.7	6.0
106	Tin Cup SRW	AMW	0.03	<0.03	0.01	(6.7)	0.66	66.6	3.0	nd	2.8	1.9	14.6	5.6
107	Black Velvet Reserve Canadian Whisk…	CAN	0.11	0.05	nd	24.8	9.0	9.5	nd	tr	tr	16.2	1.8	2.6
108	Bower Hill - Barrel Strength KY Str…	AMW	0.05	0.03	nd	(14.1)	0.41	89.6	1.6	2.1	5.0	1.8	10.6	6.5
109	Widow Jane - 10 Yr Blended Straight…	AMW	0.08	0.04	nd	(24.2)	0.68	154	7.9	3.2	11.0	2.5	20.6	4.8
110	Cedar Ridge - The Quintessential Am…	AMW	0.07	0.17	nd	106	8.1	37.6	18.4	6.8	4.2	0.90	9.2	3.1
111	Davidson Reserve Small Batch Tennes…	AMW	0.05	<0.03	nd	(19.9)	0.61	106	5.6	1.9	9.0	2.2	22.1	7.7
112	Happenstance Straight Bourbon Whisk…	AMW	0.03	<0.03	nd	(11.8)	0.28	59.2	nd	tr	tr	1.2	10.1	3.6
113	Uncle Nearest - 1856 Premium Tennes…	AMW	0.05	<0.03	nd	(22.2)	0.59	87.6	25.8	3.0	5.1	1.8	16.1	nd
114	Russell's Reserve 6 Year Kentucky S…	AMW	0.06	<0.03	nd	(15.3)	0.58	109	10.5	1.8	7.8	1.6	12.0	4.8
115	Russell's Reserve Single Barrel Str…	AMW	0.04	<0.03	nd	(18.6)	tr	99.4	9.8	1.3	5.1	1.6	10.1	3.8
116	Bespoke Spirits - Rye Whiskey Batch…	AMW	0.03	0.06	nd	(10.0)	0.76	58.6	0.79	3.2	6.0	3.4	33.1	19.7
117	High West - Double Rye! Blended SRW…	AMW	0.41	0.36	1.15	nq	1.4	69.2	nd	2.4	3.8	1.7	13.1	4.5
118	High West - A MidwinterNightsDream…	AMW	0.87	1.28	nd	nq	17.5	120	10.3	16.0	11.3	3.3	17.3	4.6
119	Forty Creek Barrel Select Canadian…	CAN	1.12	1.04	0.02	nq	24.3	4.4	39.2	10.7	tr	25.4	1.9	3.3
120	Basil Hayden - Kentuwftck Straight…	AMW	0.04	<0.03	nd	(8.6)	tr	81.1	4.4	tr	6.5	2.2	12.4	tr
121	Alquimia Reserva de Oro - 14 Yr Ext…	AGV	0.04	<0.03	nd	(8.8)	0.47	62.9	71.4	2.6	1.5	1.6	23.4	tr
122	Alquima - Reserva de Don Adolfo - 6…	AGV	0.03	<0.03	nd	(7.3)	nd	52.3	59.4	nd	1.1	0.73	12.3	nd
123	LUNAZUL Anejo Tequila 40%ABV	AGV	0.04	<0.03	nd	(11.3)	tr	7.8	4.2	0.28	1.5	2.3	4.4	tr

#	Product	Class	Glc	Fru	Suc	Gly	PG	Querc	Lac	Succ	Gall	HMF	Benz	Cinn
124	Tullibardine 12 Year Highland Singl…	SCW	0.04	<0.03	nd	(7.5)	0.60	46.5	0.32	0.27	tr	1.7	7.9	tr
125	Tin Cup SRW	AMW	0.01	<0.03	nd	nq	14.2	10.2	17.2	1.2	0.80	tr	7.3	nd
126	Cask House - Triple Oak Rye Whiskey…	AMW	0.03	0.03	nd	(8.8)	tr	69.1	nd	0.12	nd	3.1	20.2	13.8
127	Bombarda Rum - Formidable - 18Yr Ru…	RUM	0.04	<0.03	nd	(35.8)	tr	84.9	nd	tr	1.2	0.62	13.8	2.8
128	Hennessy VS Cognac 40%ABV	BRA	2.88	2.91	2.58	nq	3.9	34.1	35.7	2.2	3.2	17.2	2.0	7.7
129	Starward - Two Fold - 40%ABV Aged 2…	OWW	0.03	<0.03	0.02	209	9.3	17.2	18.9	17.7	6.2	0.94	5.4	tr
130	Pyrat XO Reserve Rum 40%ABV	RUM	7.37	7.83	0.09	<271	29.0	11.0	6.6	2.4	1.8	4.5	5.0	tr
131	Kavalan Solist Ex-Bourbon Cask 56.3…	TWW	0.06	<0.03	nd	(17.6)	nd	75.3	nd	2.2	2.3	1.0	15.8	tr
132	Kavalan Solist Ex-Bourbon Cask SMW…	TWW	0.05	<0.03	nd	(17.0)	tr	80.2	4.9	0.76	1.9	0.87	9.9	tr
133	Suntory - The Yamazaki - 12 Yr SMW…	JPW	0.08	<0.03	0.01	43.1	0.38	70.1	23.9	3.1	4.8	2.0	8.7	tr
134	Highland Park 15 Yr SMW 43%ABV 2013	SCW	0.04	<0.03	0.02	65.9	1.5	47.9	10.3	3.1	3.1	0.88	4.2	nd
135	Bowmore 12 Yr Islay SMW 40%ABV 2004	SCW	0.06	0.06	0.03	83.6	3.1	31.6	9.1	3.8	1.0	5.4	1.4	tr
136	Highland Park 18 Yr Orkney Island S…	SCW	0.11	0.12	nd	96.5	3.0	77.3	23.6	20.3	3.3	2.6	4.5	nd
137	Highland Park 18 Yr SMW 43%ABV 2016	SCW	0.05	0.04	0.07	306	6.6	36.8	25.6	12.7	1.5	2.9	14.0	tr
138	Highland Park 25 Yr SMW 45.7%ABV 20…	SCW	0.11	0.08	0.01	172	3.8	116	33.9	6.8	5.3	3.6	13.9	tr
139	Redbreast 27 Yr Single Pot Still Ir…	IRW	0.46	0.55	0.01	145	3.7	195	11.0	18.2	37.8	3.8	16.4	2.2
140	Redbreast 12 Yr Cask Strength Singl…	IRW	0.06	0.07	0.00	136	1.3	84.7	7.6	7.2	7.2	1.3	14.8	tr
141	Redbreast Lustau Edition - Single P…	IRW	0.07	0.05	nd	314	2.0	83.8	16.7	16.3	10.3	2.0	16.6	tr
142	Redbreast 15 Yr Single Pot Still Ir…	IRW	0.11	0.08	nd	51.2	tr	54.9	3.1	2.2	5.9	6.9	7.9	tr
143	Redbreast 21 Yr Single Pot Still Ir…	IRW	0.08	0.09	nd	149	1.4	91.3	6.6	8.0	7.9	4.4	22.4	4.7
144	Redbreast 12 Yr Single Pot Still Ir…	IRW	0.10	0.15	0.00	61.4	tr	54.6	6.4	2.9	5.0	6.9	9.1	tr
145	Kavalan - "Podium" - Single Malt Wh…	TWW	0.04	<0.03	nd	37.3	nd	49.9	2.4	2.0	0.87	tr	11.5	tr
146	Dewar's 18 Yr Blended Scotch Whiske…	SCW	0.06	0.07	0.00	28.6	nd	28.0	5.8	0.71	2.5	3.2	4.5	tr
147	Dewar's Signature Blend Scotch Whis…	SCW	0.08	0.10	nd	nq	10.4	36.2	8.3	tr	1.0	4.0	4.9	nd
148	Dewar's "The Ancestor" 12 Year Blen…	SCW	0.07	0.10	nd	nq	10.3	19.0	8.5	tr	0.50	4.7	4.6	tr
149	Dewar's "The Monarch" 15 Year Blend…	SCW	0.08	0.10	nd	23.8	3.6	24.9	6.5	0.70	1.9	4.6	5.0	nd
150	Dewar's Signature Blend Scotch Whis…	SCW	0.06	0.10	nd	nq	12.6	25.3	13.2	0.18	0.83	4.9	4.8	tr
151	Midleton 2023 Very Rare Irish Blend…	IRW	0.13	0.13	nd	nq	9.3	52.5	4.2	tr	5.0	9.3	8.8	3.3
152	Midleton Very Rare Barry Crockett L…	IRW	0.09	0.09	nd	(25.6)	5.2	74.0	4.8	0.15	0.63	6.4	12.8	3.3
153	Highland Park - Viking Honor 12 Yr…	SCW	0.04	<0.03	nd	43.9	1.7	44.4	2.9	2.4	4.7	1.4	8.5	tr
154	Jameson 18 Yr Irish Blended Whiskey…	IRW	0.10	0.08	nd	(29.4)	0.46	68.5	3.2	2.1	6.7	5.9	12.5	2.6
155	Knappogue Castle 12 Yr - Bourbon Ba…	IRW	0.07	0.05	nd	(30.7)	2.3	44.2	nd	tr	1.4	4.8	8.1	nd
156	Knappogue Castle 14 Yr - Non-Chille…	IRW	0.04	0.03	nd	(28.8)	4.9	68.7	nd	0.14	0.54	1.2	10.7	tr
157	Knappogue Castle 16 Yr -Sherry Cask…	IRW	0.03	0.04	nd	273	18.2	87.8	10.5	13.3	10.1	2.1	15.2	tr
158	Knappogue Castle 21 Yr -Exclusive C…	IRW	0.04	<0.03	nd	(17.8)	6.8	80.1	1.8	0.19	1.2	0.84	15.3	nd
159	Connemara 12 Yr Peated Single Malt…	IRW	0.03	<0.03	nd	(11.9)	tr	49.6	11.4	nd	tr	0.85	10.8	tr
160	Method & Madness 14 Year Irish Sing…	IRW	0.03	<0.03	nd	(23.8)	nd	58.0	1.9	0.94	5.7	2.2	12.6	6.0
161	Brenne 10 Year Ingle Malt Whiskey 4…	OWW	0.01	<0.03	nd	79.3	13.6	2.8	nd	tr	0.33	2.5	2.2	tr
162	Blue Spot 7 Year Cask Strength Sing…	IRW	0.44	0.38	nd	nq	6.6	73.0	13.6	6.2	12.2	7.9	10.5	tr
163	Powers - Johns Lane 12 Yr Single Po…	IRW	0.12	0.10	nd	34.3	1.3	37.6	nd	1.4	3.8	8.5	6.6	nd
164	Powers - Three Swallow Release Sing…	IRW	0.12	0.11	nd	nq	1.1	23.5	nd	1.4	2.0	9.4	4.8	tr
165	Flora de Cana 4 Yr Gold Rum 40%ABV…	RUM	0.12	0.07	nd	(51.7)	nd	5.3	1.7	0.87	0.41	8.3	tr	tr
166	Flora de Cana 7 Yr Grand Reserve Ru…	RUM	0.17	0.14	0.03	15.4	tr	5.1	1.9	0.63	tr	21.2	2.3	3.8
167	Flor de Cana 12Yr Centenario Rum 40…	RUM	0.19	0.13	nd	21.4	1.3	6.1	2.7	2.0	0.69	13.9	5.0	2.8
168	Flor de Cana 18Yr Centenario Gold R…	RUM	0.18	0.13	nd	18.0	1.5	9.9	2.1	1.9	1.1	15.0	5.1	6.9
169	Bacardi Reserva Limitada Rum 40%ABV…	RUM	6.48	9.22	nd	<155	24.4	14.3	2.0	1.2	1.3	11.0	6.7	tr
170	Appleton Estate 21 Yr Estate Rum 43…	RUM	0.36	0.13	nd	51.8	nd	98.3	5.0	8.2	1.7	27.9	36.3	3.5
171	El Jimador Tequila Anejo 40%ABV 2002	AGV	0.03	<0.03	nd	(7.6)	nd	60.7	109	1.3	0.93	1.1	9.0	nd
172	Gran Centanario Legenda Tequila Ext…	AGV	0.87	1.24	nd	433	3.7	19.4	27.9	0.52	2.2	9.9	6.5	tr
173	Gran Centenario - Anejo Tequila - 4…	AGV	0.87	1.14	0.17	890	4.4	27.6	83.7	0.66	0.63	1.7	5.4	tr
174	Pierre Ferrand "Ambre" XO Cognac 40…	BRA	5.59	5.58	nd	nq	14.3	71.4	60.7	6.3	8.8	6.7	7.2	nd
175	English Harbour 5 Yr Antigua Rum 40…	RUM	4.63	4.82	0.08	nq	14.6	27.5	2.1	1.3	1.7	2.2	12.4	tr
176	English Harbour Reserve - 10 Yr Rum…	RUM	2.24	2.12	0.03	nq	5.2	22.4	0.42	1.5	1.5	2.6	12.1	2.4
177	Kirk & Sweeney - 12 Yr Rum - 40%ABV	RUM	1.67	1.70	0.31	nq	660	7.4	3.9	0.46	1.5	6.6	15.2	tr
178	Planteray Rum - Isle of Fiji 40%ABV…	RUM	3.66	3.50	7.69	nq	43.7	41.4	10.4	1.6	1.9	7.6	4.3	2.3
179	Barley Creek Distillery - Honey Whi…	AMW	0.05	<0.03	26.87	nq	303	3.2	35.6	0.37	tr	tr	11.7	tr
180	Basil Hayden - Toast - Kentucky Str…	AMW	0.04	<0.03	nd	(9.3)	0.48	80.6	0.75	1.7	8.7	3.0	12.9	8.4
181	Ron del Barilito - Puerto Rican Rum…	RUM	1.79	1.72	0.12	nq	7.8	8.6	4.5	6.8	1.8	21.7	6.0	1.7
182	Glenmorangie - The Cadboll Estate…	SCW	0.03	0.03	nd	(8.0)	0.13	37.4	nd	0.075	nd	1.7	12.7	1.7
183	Coppersea Distilling - Excelsior…	AMW	0.03	<0.03	0.00	(9.5)	0.78	58.6	29.7	nd	5.1	3.5	15.2	2.7

#	Product	Class	Glc	Fru	Suc	Gly	PG	Querc	Lac	Succ	Gall	HMF	Benz	Cinn
184	Tuthilltown Spirits - Hudson Whisky…	AMW	0.06	<0.03	nd	(12.4)	1.0	111	2.4	3.4	10.2	4.2	23.2	13.4
185	Bulleit 95 Rye	AMW	0.04	<0.03	nd	(9.9)	0.37	81.2	2.7	1.1	6.5	1.6	12.3	9.8
186	Palmetto Whiskey	AMW	nd	<0.03	nd	nq	1.4	37.6	0.34	0.95	7.2	1.5	43.2	36.3
187	Taconic Distillery - Double Barrel…	AMW	0.24	0.13	11.79	nq	324	58.0	12.4	0.50	nd	0.41	8.5	tr
188	Diplomatico - Venezuala	RUM	2.66	2.09	0.45	nq	8.7	4.3	4.3	0.45	0.53	5.0	3.3	tr
189	Flor de Cana Rum - 7 Yrs - Nicaragu…	RUM	0.07	0.08	0.01	(13.0)	nd	4.7	3.4	0.51	tr	10.9	1.4	tr
190	The Cane Hill - Barbados - 5 Years	RUM	0.04	<0.03	0.00	(10.5)	0.23	18.6	1.4	0.60	1.8	2.2	6.1	nd
191	Trinidad - 16 Years	RUM	4.02	4.69	0.57	nq	11.1	53.9	10.2	1.9	3.0	45.7	16.3	5.9
192	J.Wray - Gold Rum - Jamaica	RUM	0.08	<0.03	0.01	(12.2)	tr	4.2	1.4	nd	0.33	6.3	2.9	tr
193	Roatan Rum Company - Nicaragua	RUM	2.66	2.19	19.25	nq	1179	tr	18.7	0.17	nd	3.6	105	nd
194	Cadenheads - Classic Rum Blend - 5Y…	RUM	1.01	1.17	0.25	nq	4.8	39.6	17.1	15.2	3.2	18.0	11.9	3.2
195	Brugel - Anejo Superior - Dominican…	RUM	0.88	0.83	0.98	nq	4.6	5.0	6.2	0.68	0.68	12.8	tr	tr
196	Meyers - Plantation White - Jamaica	RUM	0.06	0.05	nd	(22.3)	0.33	105	3.6	2.7	10.5	1.9	16.3	3.1
197	Flor de Cana - 12 Years - Nicaragua	RUM	0.03	<0.03	0.01	(13.0)	0.69	64.1	4.2	1.3	4.7	1.4	14.3	4.1
198	Rum 28 - 60%ABV	RUM	0.20	0.04	nd	nq	0.40	tr	1.3	0.19	0.52	17.8	76.7	3.2
199	Mount Gay - Barbados Rum 1703 XO 43…	RUM	0.29	0.24	nd	34.7	0.46	45.0	5.0	1.9	4.1	25.2	10.4	4.6
200	Glenmorangie - The Lasanta - 43%ABV…	SCW	0.25	0.28	nd	119	6.9	58.4	8.2	5.5	5.7	12.2	17.5	6.8
201	Laphroaig - Cask Collection- Oak Se…	SCW	0.02	<0.03	nd	(12.1)	tr	35.2	nd	tr	tr	0.58	6.1	tr
202	Hudson Whiskey NY - Cask Strength B…	AMW	0.05	0.03	0.01	(12.7)	1.6	120	nd	3.2	12.1	3.9	23.4	13.4
203	Hudson Whiskey NY - Short Stack - N…	AMW	0.50	0.37	3.69	nq	5.9	80.0	12.4	nd	nd	2.9	16.9	6.3
204	Hudson Whiskey NY - Do The Rye Thin…	AMW	0.03	<0.03	nd	(11.3)	1.2	89.5	12.1	1.8	10.5	4.5	16.1	23.2
205	Hudson Whiskey NY - Single Barrel C…	AMW	0.08	0.06	nd	(56.4)	2.4	129	0.31	nd	10.2	2.1	18.6	4.3
206	Hudson Whiskey NY - This is Your Fa…	AMW	0.06	0.03	nd	(10.3)	0.43	112	35.8	3.6	12.6	5.4	19.7	20.2
207	Hudson Whiskey NY - Bright Lights B…	AMW	0.04	<0.03	0.00	(8.9)	0.97	86.8	0.27	2.6	7.8	2.8	17.4	16.0
208	Hudson Whiskey NY - Back Room Deal…	AMW	0.05	<0.03	0.05	(16.2)	2.2	104	16.6	3.4	7.6	4.5	22.4	15.5
209	Hudson Whiskey NY - Four Part Harmo…	AMW	0.05	0.04	nd	(18.2)	3.0	102	4.0	3.0	7.4	3.7	22.0	17.9
210	Hudson Whiskey NY - Port Cask Finis…	AMW	0.98	1.77	nd	302	33.4	110	30.2	37.5	9.3	4.0	28.4	16.1
211	Hudson Whiskey NY - Calvados Cask F…	AMW	0.08	0.12	nd	16.4	1.3	122	36.4	5.1	9.8	5.4	23.6	11.6
212	Hudson Whiskey NY - Sherry Cask Fin…	AMW	0.10	0.14	nd	64.4	5.3	144	13.0	9.9	17.1	10.3	38.9	23.9
213	Maker's Mark No. 46 - French Oaked…	AMW	0.04	<0.03	nd	(29.9)	0.23	99.4	2.3	2.6	5.5	2.2	13.5	3.4

Glc, glucose; Fru, fructose; Suc, sucrose; Gly, glycerol; PG, 1,2-propanediol; Querc, quercitol; Lac, lactic acid; Succ, succinic acid; Gall, gallic acid; Benz, benzaldehydes; Cinn, cinnamaldehydes. nd, not detected; tr, trace; nq, not quantifiable; (x), upper bound; <x, upper limit.

Table 32. Concentrations in the $CDCl_3$ residues (mg L^{-1} of spirit; acyl chains and FAEE in µmol L^{-1}).

#	Product	Class	Van	Syr	Cinn	Sinap	HMF+EMF	DES	Chains	FAEE
214	Bushmills - Black Bush Irish Whisky 40%…	IRW	tr	0.72	nd	nd	2.0	tr	51.2	10.9
215	Kavalan Solist Vinho Barrique Single Ca…	TWW	5.0	13.4	6.0	tr	1.4	6.8	603	297
216	An Dulaman - Santa Ana Armada Strength…	GIN	tr	tr	tr	nd	nd	3.1	56.2	nd
217	Medford Rum - 3 Years Aged	RUM	3.4	8.9	11.8	5.3	3.2	0.75	96.9	33.5
218	Kavalan Earth Silver Wine Cask Matured…	TWW	2.2	6.2	nd	nd	0.46	8.9	742	388
219	Spirit of HVEN Mercurious Corn Whisky 4…	OWW	1.5	4.8	nd	nd	1.1	tr	113	27.5
220	Red Spot 15 Year Single Pot Still Irish…	IRW	1.7	1.8	nd	tr	3.1	2.3	147	12.9
221	Makers Mark Single Batch Bourbon Whiskey	AMW	2.2	6.0	2.2	tr	1.4	1.2	63.9	13.6
222	Pendleton 1910 - 12 Year Old Canadian R…	CAN	3.0	9.5	3.8	0.40	1.3	1.00	99.5	33.0
223	Coalition Whiskey Barrel Proof Kentucky…	AMW	4.8	13.1	9.6	3.6	2.9	2.2	206	82.6
224	Bernheim - 7 Year Old Small Batch Strai…	AMW	3.5	10.0	4.3	nd	2.7	1.6	161	60.6
225	Bardeney - Bottled in Bond Rye Whiskey	AMW	4.0	9.8	3.4	0.28	1.4	1.4	104	41.8
226	TX - American Blended Whisky	AMW	2.1	tr	tr	nd	3.6	1.8	59.0	3.0
227	Kavalan - Distillery Reserve - Rum Cask…	TWW	2.8	10.0	tr	nd	1.0	0.93	687	348
228	Kavalan Sky Gold Wine Cask Matured Sing…	TWW	2.1	4.4	tr	0.63	0.89	10.7	656	285
229	Makers Mark Cask Strength Single Malt W…	AMW	2.2	6.7	tr	0.43	0.89	0.95	88.9	29.6
230	Old Overholt - 114 Proof Straight Rye W…	AMW	3.9	12.9	5.4	0.30	2.4	0.88	578	394
231	Makers Mark 101 Single Malt Whisky 50.5…	AMW	nd	tr	nd	2.3	1.4	tr	70.4	nd
232	Doc Swinson's - ALter Ego Solera Method…	AMW	tr	1.1	2.4	tr	tr	tr	46.4	16.8
233	Southern Star Double Rye Whiskey	AMW	2.2	6.4	6.4	0.92	1.7	1.0	118	48.5
234	Bespoke Spirits Rye Whiskey 2020-1	AMW	7.4	17.4	9.8	10.3	2.5	2.1	189	115
235	High West - Double Rye! Blended Straigh…	AMW	1.4	4.4	4.7	0.96	0.79	1.0	142	78.3
236	High West - A Mid-Winters Night Dram Ac…	AMW	2.4	6.5	5.3	0.94	1.3	5.8	150	29.7
237	Tin Cup Straight Rye Whiskey	AMW	2.9	7.6	4.6	tr	1.8	0.98	111	49.8

#	Product	Class	Van	Syr	Cinn	Sinap	HMF+EMF	DES	Chains	FAEE
238	Kavalan - Classic - SMW	TWW	3.1	8.7	3.3	0.59	1.5	3.0	235	47.2
239	Oloroso - Sherry Oak SMW	TWW	1.3	3.7	nd	nd	1.4	6.0	271	114
240	Solist Madeira SCS SMW	TWW	3.3	6.6	nd	0.79	1.4	11.1	348	123
241	Kavalan - Solist Oloroso Sherry Single…	TWW	4.4	19.2	2.4	1.6	3.4	25.8	839	394
242	Kavalan - ArtistSeries: Paul Chiang Pea…	TWW	7.1	21.2	6.0	1.8	2.2	1.1	638	349
243	Kavalan Solist Ex-Bourbon Cask 56.3%ABV	TWW	3.2	8.2	5.0	nd	0.73	0.92	449	200
244	Kavalan Solist Ex-Bourbon Cask SMW 57.1…	TWW	2.5	5.3	2.2	nd	0.55	0.43	328	130
245	Highland Park 18 Yr Orkney Island SMW 4…	SCW	2.1	3.7	2.9	tr	2.0	4.7	210	nd
246	Highland Park 18 Yr SMW 43%ABV 2016	SCW	2.7	7.0	tr	0.53	1.7	4.8	184	nd
247	Highland Park 25 Yr SMW 45.7%ABV 2013	SCW	3.9	10.2	3.0	tr	3.2	5.2	378	35.7
248	Redbreast 27 Yr Single Pot Still Irish…	IRW	1.5	2.5	2.5	0.35	1.2	2.3	170	58.3
249	Redbreast 15 Yr Single Pot Still Irish…	IRW	0.75	1.5	tr	tr	2.8	0.90	62.4	8.7
250	Coppersea Distilling - Excelsior - Huds…	AMW	2.6	10.5	2.9	2.9	2.9	1.0	104	22.2
251	Tuthilltown Spirits - Hudson Whisky NY	AMW	5.5	15.3	5.1	6.8	4.5	3.4	168	57.4
252	Bulleit 95 Rye	AMW	2.4	7.5	9.3	2.5	1.6	1.3	108	43.6
253	Pierre Ferrand "Ambre" XO Cognac 40%ABV…	BRA	nd	nd	nd	nd	nd	nd	123	nd
254	Basil Hayden - Toast - Kentucky Straigh…	AMW	2.6	7.8	3.9	4.7	3.1	1.6	138	47.5
255	Ron del Barilito - Puerto Rican Rum 43%…	RUM	0.64	1.6	1.0	0.28	3.3	1.5	47.9	6.5
256	Glenmorangie - The Cadboll Estate - 15…	SCW	1.6	3.0	2.1	0.85	1.00	tr	202	71.4
257	Rum 28 - 60%ABV	RUM	8.5	0.56	0.98	nd	2.7	nd	33.4	nd
258	Glenmorangie - The Lasanta - 43%ABV - 1…	SCW	1.4	4.5	1.8	2.7	6.7	2.4	234	76.0
259	Laphroaig - Cask Collection- Oak Select…	SCW	0.85	3.1	tr	2.0	1.6	2.2	197	18.4
260	Powdermen Distillery - Black Wick Bourb…	AMW	5.2	15.5	9.1	9.9	1.6	1.2	252	156
261	Hudson Whiskey NY - Cask Strength Barre…	AMW	4.5	14.5	7.9	8.3	4.2	2.2	268	104
262	Hudson Whiskey NY - Short Stack - NY SR…	AMW	4.0	10.3	10.0	2.6	2.5	3.5	193	52.6
263	Hudson Whiskey NY - Do The Rye Thing…	AMW	4.1	12.8	11.6	12.7	5.5	2.5	138	39.4
264	Hudson Whiskey NY - Single Barrel Cask…	AMW	3.6	8.6	2.9	2.8	2.0	1.7	270	156
265	Hudson Whiskey NY - This is Your Favori…	AMW	4.7	16.3	10.8	11.6	5.8	3.1	184	68.2
266	Hudson Whiskey NY - Bright Lights Big B…	AMW	4.6	11.5	7.8	7.5	3.5	2.4	165	49.4
267	Hudson Whiskey NY - Back Room Deal - NY…	AMW	6.1	17.6	18.4	6.1	5.4	3.5	169	33.8
268	Hudson Whiskey NY - Four Part Harmony…	AMW	5.6	15.1	8.8	9.4	4.2	2.7	173	40.1
269	Hudson Whiskey NY - Port Cask Finish…	AMW	5.5	18.9	8.4	9.2	4.1	12.2	197	nd
270	Hudson Whiskey NY - Calvados Cask Finis…	AMW	4.8	16.1	4.9	5.5	4.7	2.5	247	75.9
271	Hudson Whiskey NY - Sherry Cask Finish…	AMW	8.1	28.2	14.5	15.4	11.6	7.3	177	nd
272	Redbreast 21 Yr Single Pot Still Irish…	IRW	3.2	8.3	3.7	1.1	3.1	3.5	93.3	8.1
273	Kavalan - "Podium" - Single Malt Whiske…	TWW	3.1	7.1	3.2	nd	0.85	1.7	414	110
274	Maker's Mark No. 46 - French Oaked - 47…	AMW	2.5	9.7	5.9	1.8	1.9	1.6	104	36.3

Van, vanillin; Syr, syringaldehyde; Cinn, cinnamaldehydes; Sinap, sinapaldehyde; DES, diethyl succinate; Chains, acyl chains (α-CH_2/2); FAEE, fatty acid ethyl esters (OCH_2/2).

The concentrations found in the DMSO-d_6 residues are listed in Table 33. The pentoses and glucose were quantified from their anomeric doublets, whose shapes in this solvent depend on the rate of hydroxyl exchange (Section 9.4), and their values carry that limitation; ellagic acid is reported only in DMSO-d_6, in which alone it dissolves.

Table 33. Concentrations in the DMSO-d_6 residues (mg L^{-1} of spirit).

No.	Product	Syringaldehyde	Vanillin	Cinnamaldehydes	HMF	Ellagic acid	Gallic acid	Glucose	Arabinose	Xylose	Quercitol	Succinic acid	Acetic acid	Ethanol (retained)
1	Hudson Whiskey NY - Bright Lights…	14.2	4.2	11.5	3.4	3.0	5.0	177	101	2.9	76.7	1.3	1.1	3.1
2	Hudson Whiskey NY - Back Room Dea…	16.7	5.5	9.9	3.9	1.1	4.0	181	79.6	12.7	78.6	1.00	2.2	nd
3	Hudson Whiskey NY - Four Part Har…	15.7	4.4	12.0	3.5	1.2	4.3	185	82.1	11.7	75.8	0.62	2.2	nd
4	Hudson Whiskey NY - Port Cask Fin…	23.2	7.9	14.1	5.3	1.1	11.9	1426	98.8	43.4	110	16.2	6.8	40.1
5	Glen Grant 12 Yr Speyside SMW 43%…	nd	0.13	1.7	1.4	nd	nd	47.3	11.5	10.3	25.7	0.61	nd	2.2
6	Kavalan - Solist Ex-Bourbon Cask…	7.4	4.3	7.0	2.4	0.48	nd	61.4	28.3	16.8	76.6	0.93	1.5	10.7
7	Kavalan - Solist Amontillado Sher…	19.1	5.3	4.1	5.8	nd	13.0	341	137	17.6	84.2	16.8	1.5	15.0
8	Kavalan - Solist Oloroso Sherry S…	22.1	4.8	5.2	3.7	20.8	nd	168	59.0	nd	28.8	10.5	4.3	1.5
9	Kavalan - Concertmaster Sherry Ca…	12.6	4.8	3.6	1.1	nd	3.1	110	64.7	16.8	52.5	0.46	1.5	2.1
10	Kavalan - King Car - Conductor…	12.9	4.3	2.7	2.8	nd	2.4	166	63.2	14.8	56.0	0.41	1.3	5.8
11	Kavalan - Distillery Reserve Peat…	6.7	2.0	3.0	0.24	nd	1.4	87.0	38.5	4.0	22.9	0.64	1.1	4.4

No.	Product	Syringaldehyde	Vanillin	Cinnamaldehydes	HMF	Ellagic acid	Gallic acid	Glucose	Arabinose	Xylose	Quercitol	Succinic acid	Acetic acid	Ethanol (retained)
12	Solist Brandy Single Cask Strengt…	3.2	1.1	2.3	3.3	nd	6.4	1155	26.3	37.7	43.7	1.1	1.6	52.6
13	Kavalan - Classic - SMW	12.8	3.3	2.6	1.1	0.36	2.2	198	74.9	16.1	50.2	0.79	2.8	3.5
14	Kavalan - Solist Ex-Bourbon Sigle…	19.3	7.0	3.9	3.1	nd	6.9	128	93.0	28.3	93.3	1.1	2.2	30.6
15	Kavalan - Artist Series-Paul Chia…	27.6	8.7	3.9	1.5	nd	15.1	236	190	75.7	174	2.9	2.3	1.9
16	Kavalan Solist Oloroso Sherry Oak…	9.4	2.2	2.1	2.8	nd	7.7	161	45.4	15.5	65.1	12.1	7.6	2.1
17	Solist Moscatel Sherry - Single C…	13.6	4.5	3.0	6.5	nd	10.3	3300	86.6	8.1	74.5	9.9	2.9	20.3
18	Solist Madeira Single Cask Streng…	16.7	7.3	3.3	3.3	0.78	15.6	338	45.7	16.8	72.9	15.8	3.0	26.6
19	Kavalan - Ex-Bourbon Oak - SMW	10.0	4.2	2.6	0.76	0.29	0.16	85.2	43.6	14.8	53.2	0.49	1.6	8.9
20	Kavalan - Solist Fino Sherry Sing…	25.8	4.7	7.7	5.6	nd	16.0	246	145	27.3	75.3	17.0	2.4	7.3
21	Kavalan - Artist Series Paul Chia…	15.2	4.7	17.5	2.1	28.6	18.2	205	101	55.3	90.8	1.7	17.2	7.4
22	Kavalan Solist Vinho Barrique SCS…	18.8	7.1	3.9	2.1	nd	12.2	128	99.9	32.4	123	6.5	2.0	35.6
23	Solist Manzanilla Sherry - Single…	15.3	3.2	9.1	3.9	67.0	33.1	180	55.9	59.5	145	6.5	1.3	6.3
24	Kavalan - ConcertMaster Port Cask…	13.5	5.4	3.5	1.5	nd	7.8	256	67.2	40.0	66.7	11.0	1.3	3.2
25	Kavalan - Podium - SMW	14.0	4.3	0.96	2.8	nd	2.1	154	79.2	22.1	61.5	0.23	6.7	0.51
26	Kavalan - King Car - 40th Anniver…	11.1	4.4	8.5	1.3	31.9	8.8	140	66.6	67.1	122	15.0	8.4	86.0
27	Adirondack Maple Barrel Bourbon 4…	nd	nd	3.2	0.86	0.69	1.1	154	70.3	nd	96.2	0.75	nd	2151
28	Adirondack Rum 73%ABV	nd	6.7	2.4	0.73	nd	0.98	48.2	36.3	12.3	40.1	0.54	7.2	1.1
29	TenMile Distillery - Raid on Rich…	8.0	3.7	5.9	1.9	nd	40.4	66.5	61.8	28.1	39.9	1.6	0.29	5.7

nd: not detected or not quantifiable. Numbering as in the DMSO-d₆ sample table. Glucose and xylose are upper bounds (Section 9.8).

Appendix C. Data-integrity log

The data-integrity observations made in the course of this work are collected in Table 34. None of them was apparent from the processed spectra or the spectrum titles alone; all were found by comparing the stored decays, the parameter sets and the spectra with one another and with the internal tracers described in Section 5. The entries added in version 2.1 concern the DMSO-d_6 document (two label conflicts, suppression recorded in titles but not applied, the absence of TMS and the displacement of the residual solvent line), two corrections to the D_2O/CD_3OD quantitation of version 2.0 (the overlooked pentoses and the anomeric signals displaced by an unusual CD_3OD content), the refitting of two succinic acid singlets in DMSO-d_6, the moisture in two $CDCl_3$ residues, and a citation of version 2.0 that misdescribed its source. Version 2.3 adds three entries concerning the reference spectra of pure congeners (spectra without an internal reference, the methanol and 2-propanol of the 50:50 ethanol solutions and the acetaldehyde of the model-spirit base) and records the use of the native preparation of Kavalan Distillery Reserve Rum Cask as the intact-spirit comparison of Section 2.9.

Table 34. Data-integrity log.

Item	Observation	Treatment
Solvent labels	Twenty D_2O titles misstate the solvent system: 19 spectra titled D_2O contain CHD_2OD at 9.6–21.3 mmol L^{-1}, and Bushmills Black Bush, titled D_2O + CD_3OD, contains none.	Corrected assignment used throughout (Section 5.4).
First D_2O set	For 100 of 101 items (83 of the 84 retained) the stored processed spectrum had been normalized by MestReNova to a reference intensity of 100.	Absolute scale restored from the fitted factor (Section 4.5).
Duplicated spectra	The first sixteen spectra of the second D_2O set are byte-identical copies of first-set spectra.	Counted once; second-set copies retained.
Referencing	Bower Hill Barrel Strength (first-set copy) mis-referenced by 0.029 ppm; Clase Azul Añejo and Litchfield Maple Bourbon referenced to a ^{29}Si satellite of TMSP; Litchfield Straight Rye and Hudson House Rye referenced to a spinning sideband.	Corrected (Section 4.6).
Kavalan-010	Title reads "Bespoken Spirits Rye Whiskey"; sample ID, DMSO-d_6 counterpart and chemistry identify Kavalan Solist Brandy.	Treated as Kavalan Solist Brandy.
Native preparation	Kavalan Distillery Reserve Rum Cask prepared natively (175 µL of spirit) rather than freeze-dried.	Excluded from the freeze-dried data base; used in Section 2.9 as the intact-spirit comparison.
Product titles	Maker's Mark 101 and Maker's Mark Cask Strength titled "single malt" (both are wheated bourbons); several titles contain typographical errors (for example "Kentuwftck", "Prmier", "Barilito", "Grench").	Classified correctly; titles reproduced as recorded.
Planteray Isle of Fiji	Sample name records 3 mL of spirit, title and preparation record 1 mL.	1 mL used; concentrations would be a third if 3 mL is correct.
Replicate residues	Branson XO Premier Cru, Powers Gold Label, Tin Cup Straight Rye and Sotol Ono each represented twice; non-volatile ratios 0.62, 0.32, about 0.5 and 1.6.	Both residues kept; treated as one product in grouped validation (Section 10.2).
Kavalan Distillery Reserve Rum Cask	Two D_2O residues of the same sample (343) under different titles, measured in March and August 2025.	Treated as different products in the grouped validation (one pair among 171 residues); used as the replicate test of Section 10.2.
Acquisition sessions of 1–2 August 2025	All nine D_2O/CD_3OD spectra of these sessions carry 0.54–0.71 of the expected CHD_2OD and a depressed TMSP signal; the sample 343 replicate shows the same deficit in every analyte.	Rescaled by their own CHD_2OD tracer (Section 5.5).
Other tracer outliers	5 spectra from other sessions with 0.55–0.72 of the expected CHD_2OD and no session pattern.	Flagged, not corrected (Section 5.5).
Residual $CHCl_3$	Kavalan Sky Gold ($CDCl_3$) carries an order of magnitude less $CHCl_3$ than the other residues with normal solute signals.	Reported; solvent lot suspected (Section 5.6).
Glycerol, first pass	Single-multiplet fits returned gram-per-liter "glycerol" in fructose-rich residues (Tromba Reposado, Trinidad 16 yr, High West A Midwinter Night's Dream) that the second glycerol multiplet does not confirm.	Two-multiplet procedure adopted (Sections 4.8, 6.5); first-pass values retained in the workbook.
Unrecorded volumes	31 D_2O and 37 $CDCl_3$ preparations without a recorded solution volume.	650 µL assumed; ±8 % uncertainty component.
DMSO-d_6 titles	The spectrum titled "Kavalan Distillery Reserve Rum Cask" carries sample ID 043 (Solist Vinho Barrique); the decay titled "Kavalan-026 Paul Chiang Peated Malt" is filed under a second title and its parameter set records sample 034 (Solist Oloroso Sherry Oak).	Both resolved by congener ratios against the D_2O and $CDCl_3$ residues: treated as 043 and 034 (Section 9.9).
DMSO-d_6 acquisition	Most titles record "HOD suppression", but the parameter sets record presaturation disabled and the water line is present at full intensity.	Treated as unsuppressed spectra.

Item	Observation	Treatment
DMSO-d_6 referencing	TMS absent from 16 of 29 spectra; the DMSO-d_5 line displaced by 0.012–0.037 ppm in six residues while the solute lines were not.	TMS or cross-correlation referencing; region-wise alignment of displaced spectra (Sections 4.15, 5.8).
Pentoses (version 2.0)	The β-arabinopyranose H-1 doublet coincides with the α-glucose H-1 doublet; version 2.0 glucose included it.	Glucose recomputed from the β-H-1 doublet; arabinose and xylose added (Section 9.5).
Displaced anomeric regions (version 2.0)	In twenty-three D_2O/CD_3OD residues whose CD_3OD fraction departed from 23 % the anomeric doublets were displaced by up to 0.029 ppm; the version 2.0 fits missed the β-glucose doublet in eight of them.	All D_2O targets refitted with residue-specific medium offsets; corrected values used throughout (Sections 4.8, 9.5).
Arabinose without its β-anomer	In twenty-two residues the δ 4.474 doublet was not accompanied by the expected β-arabinopyranose intensity.	Arabinose not reported for these residues (Section 9.5).
Equal-volume D_2O/CD_3OD residues	Larsen Aqua Ignis and Rum 28 were dissolved in 350 µL each of D_2O and CD_3OD; their signals are displaced by up to 0.03 ppm by amounts that differ from one proton to another, and the version 2.0 fits at fixed positions misplaced several of them (Larsen reported as fructose-rich).	Every target of the two residues refitted at a displacement measured by cross-correlation with residues of the same category (Sections 4.8, 10.3).
DMSO-d_6 succinic acid	In two wine-seasoned residues (Solist Oloroso 005 and Solist Madeira) the succinic acid singlet lay outside the standard position window, at δ 2.433 and 2.363, and the first fits had measured a weaker neighboring feature.	Refitted in windows centered on the observed positions (Section 9.7).
$CDCl_3$ water	Broadened water extends beyond the excluded window (δ 1.45–1.60) in two residues and forms the second principal component of the $CDCl_3$ block.	Sensitivity blocks with the window widened to δ 1.45–1.76 (Section 12.5).
Citation in version 2.0	Gollihue et al. (2018) was cited for sugar concentrations measured in bourbon; the study measured the loss of polysaccharides from the barrel staves.	Citations corrected (Sections 2, 6.4, 9.5, 10.3).
Pulse calibration	The parameter sets record a 90° pulse (pw90) of 10 µs; the calibrated 90° pulse of the probe is 13 µs, so the 8 µs and 7 µs pulses were 55.4° and 48.5° pulses, not 72° and 63°.	Flip angles corrected; every absolute intensity, saturation correction and concentration recalculated in version 2.2 (Section 5.9).
Content of the maleic acid standard	The blank was prepared from 2.55 mg of maleic acid; its effective content, determined by the author, is 95 % of the weighed amount (32.11 instead of 33.80 mmol L^{-1}).	Effective concentration used in Equation 2; every concentration of versions 2.0 and 2.1 lowered by 5.0 % in version 2.2 (Section 5.9).
Reference spectra without an internal reference	Eleven of the reference spectra of pure congeners, all in D_2O, contain neither TMSP nor DSS.	Used at their stored positions, which are reliable to about ±0.03 ppm (Section 2.11; marked † in Table 2).
Ethanol of the reference solutions	The 50:50 D_2O/ethanol reference solutions contain methanol (δ 3.36) and 2-propanol (δ 4.00) in the same proportion to ethanol in all three, which points to the ethanol used.	Lines identified and excluded from the interpretation (Sections 2.8 and 2.10); undenatured ethanol recommended (Section 17).
Model spirit, base solution	The base solution is recorded as containing an acetal, but its spectrum shows the formyl quartet of free acetaldehyde (δ 9.70).	Described as found (Section 2.10).

Appendix D. Reconstruction of the processed spectra

The statistics of the reconstruction of the MestReNova processed spectra from the stored decays (Section 4.5) are summarized in Table 35, including the DMSO-d₆ document added in version 2.1, whose processed spectra, like those of every set except the first D₂O set, had been stored on their absolute scale.

Table 35. Reconstruction of the MestReNova processed spectra from the stored decays: coefficient of determination, fitted scale factor and apodization constants, by document.

Document	Items	Minimum R^2	Median R^2	Median scale factor	Line broadening (Hz)
First D₂O set	101	0.999989	0.9999986	100 items ≠ 1 (0.00599–21.1)	0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5
Second D₂O set	113	0.999964	0.9999988	1.0000 (all)	0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5
First CDCl₃ set	39	0.999993	0.9999990	1.0000 (all)	0.2
Second CDCl₃ set	22	0.999998	0.9999998	1.0000 (all)	0.5
Maleic acid blank	1	0.999998	0.9999985	1.0000 (all)	0.25
DMSO-d₆ residues	29	0.9999957	0.9999985	1.0000 (all)	0.25

A scale factor of 1 indicates a stored spectrum on the absolute scale of the summed decay; other values are the normalization applied in MestReNova.

Appendix E. Companion workbook

The companion Excel workbook contains, on separate sheets, the sample metadata and acquisition parameters; the binned absolute-intensity matrices (0.02 ppm) for both solvents; the targeted quantitation results in long form with every fitted parameter, detection limit, saturation factor and flag; the wide concentration tables; the internal-check data and model; the principal component scores and loadings, explained variances and confounder correlations; the classification results by method, block and validation scheme with confusion matrices, chosen hyperparameters and permutation distributions; the embedding coordinates and their diagnostics; the PLS calibration statistics, cross-validated predictions and coefficients; the time-domain equivalence and HLSVD results; and the reference list with DOIs. Version 2.1 added the DMSO-d_6 sample metadata, tracers and referencing; the DMSO-d_6 quantitation in long and wide form; the binned DMSO-d_6 matrix; the cross-solvent recoveries; the pentose and corrected glucose results with the arabinose test status of every residue and the glucose of version 2.0 beside the corrected value; the $CDCl_3$ moisture check; and, for every one of the seventeen data approaches of Section 12, the principal component scores, loadings and explained variances, the cluster memberships of the four linkages, the t-SNE and UMAP coordinates and the class-purity and trustworthiness statistics, together with the matched-product comparisons and the leave-one-out classifications of Sections 12.9 and 12.10. Version 2.2 adds every concentration of version 2.1 beside its corrected value, with the flip-angle and combined ratios and the closed-form prediction (sheet P90_change), and carries the calibrated 90° pulse and the effective content of the standard as live inputs of the response factor (sheet AbsScale). Version 2.3 adds the second round of PLS calibrations: every excluded residue with the reason for its exclusion, its round-1 diagnostics and its round-2 prediction (sheet PLS2_excluded), the statistics of round 1, of round 1 scored on the retained residues and of round 2 for every model, with the sensitivity to the thresholds and the differences between the rounds as live formulas (PLS2_statistics), the predictions of both rounds for every residue (PLS2_predictions) and a matrix of the exclusions of every residue with live counts (PLS2_exclusion_matrix); the diagnostic shifts of the reference spectra of pure congeners with their expected behavior on freeze-drying (Congener_library); the two-Lorentzian fits of the syringyl singlet pair (Syringyl_pair); and the intact-spirit comparison with the expected gain and the retention of each congener as live formulas (Intact_vs_residue). A second workbook, PNA-WP-2026-FDSP-01_v2.3_3D_scores.xlsx, holds every coordinate and statistic of the interactive score explorer of Sections 11 and 12 and, in version 2.3, the second-round PLS exclusions of every residue.

References

1. Viriot, C.; Scalbert, A.; Lapierre, C.; Moutounet, M. Ellagitannins and lignins in aging of spirits in oak barrels. *Journal of Agricultural and Food Chemistry* **1993**, *41* (11), 1872–1879. https://doi.org/10.1021/jf00035a013
2. Glabasnia, A.; Hofmann, T. Sensory-Directed Identification of Taste-Active Ellagitannins in American (Quercus alba L.) and European Oak Wood (Quercus robur L.) and Quantitative Analysis in Bourbon Whiskey and Oak-Matured Red Wines. *Journal of Agricultural and Food Chemistry* **2006**, *54* (9), 3380–3390. https://doi.org/10.1021/jf052617b
3. Gollihue, J.; Richmond, M.; Wheatley, H.; Pook, V. G.; Nair, M.; Kagan, I. A.; DeBolt, S. Liberation of recalcitrant cell wall sugars from oak barrels into bourbon whiskey during aging. *Scientific Reports* **2018**, *8* (1), 15899. https://doi.org/10.1038/s41598-018-34204-1
4. Jaganathan, J.; Dugar, S. M. Authentication of Straight Whiskey by Determination of the Ratio of Furfural to 5-Hydroxymethyl-2-Furaldehyde. *Journal of AOAC International* **1999**, *82* (4), 997–1001. https://doi.org/10.1093/jaoac/82.4.997
5. Kew, W.; Bell, N. G. A.; Goodall, I.; Uhrín, D. Advanced solvent signal suppression for the acquisition of 1D and 2D NMR spectra of Scotch Whisky. *Magnetic Resonance in Chemistry* **2017**, *55* (9), 785–796. https://doi.org/10.1002/mrc.4621
6. Kew, W.; Goodall, I.; Uhrín, D. Analysis of Scotch Whisky by ¹H NMR and chemometrics yields insight into its complex chemistry. *Food Chemistry* **2019**, *298*, 125052. https://doi.org/10.1016/j.foodchem.2019.125052
7. Stockwell, M.; Goodall, I.; Uhrín, D. Quantification of whisky congeners by ¹H NMR spectroscopy. *Analytical Science Advances* **2020**, *1* (2), 132–140. https://doi.org/10.1002/ansa.202000063
8. Monakhova, Y. B.; Schäfer, H.; Humpfer, E.; Spraul, M.; Kuballa, T.; Lachenmeier, D. W. Application of automated eightfold suppression of water and ethanol signals in ¹H NMR to provide sensitivity for analyzing alcoholic beverages. *Magnetic Resonance in Chemistry* **2011**, *49* (11), 734–739. https://doi.org/10.1002/mrc.2823
9. Teipel, J. C.; Hausler, T.; Sommerfeld, K.; Scharinger, A.; Walch, S. G.; Lachenmeier, D. W.; Kuballa, T. Application of ¹H Nuclear Magnetic Resonance Spectroscopy as Spirit Drinks Screener for Quality and Authenticity Control. *Foods* **2020**, *9* (10), 1355. https://doi.org/10.3390/foods9101355
10. Kuballa, T.; Hausler, T.; Okaru, A. O.; Neufeld, M.; Abuga, K. O.; Kibwage, I. O.; Rehm, J.; Luy, B.; Walch, S. G.; Lachenmeier, D. W. Detection of counterfeit brand spirits using ¹H NMR fingerprints in comparison to sensory analysis. *Food Chemistry* **2018**, *245*, 112–118. https://doi.org/10.1016/j.foodchem.2017.10.065
11. Pineda-Amaya, A. del C.; Ocaña-Rios, I.; García-Aguilera, M. E.; Nolasco-Cancino, H.; Quiroz-García, B.; Esturau-Escofet, N.; Ruiz-Terán, F. ¹H-NMR profile of mezcal and its distillation fractions using two sample preparation methods: direct analysis and solid-phase extraction. *Chemical Papers* **2021**, *75* (8), 4249–4259. https://doi.org/10.1007/s11696-021-01660-5
12. Sujka, K.; Koczoń, P.; Górska, A.; Wirkowska, M.; Reder, M. Sensory and spectral characteristics of selected freeze-dried spirit beverages. *Żywność. Nauka. Technologia. Jakość* **2013**, *89*, 184–198. https://doi.org/10.15193/zntj/2013/89/184-198
13. Matsuo, Y.; Wakamatsu, H.; Omar, M.; Tanaka, T. Reinvestigation of the Stereochemistry of the C-Glycosidic Ellagitannins, Vescalagin and Castalagin. *Organic Letters* **2015**, *17* (1), 46–49. https://doi.org/10.1021/ol503212v
14. Hervé du Penhoat, C. L. M.; Michon, V. M. F.; Peng, S.; Viriot, C.; Scalbert, A.; Gage, D. Structural elucidation of new dimeric ellagitannins from Quercus robur L. roburins A–E. *Journal of the Chemical Society, Perkin Transactions 1* **1991** (7), 1653–1660. https://doi.org/10.1039/p19910001653
15. Puech, J.-L.; Mertz, C.; Michon, V.; Le Guernevé, C.; Doco, T.; Hervé du Penhoat, C. Evolution of Castalagin and Vescalagin in Ethanol Solutions. Identification of New Derivatives. *Journal of Agricultural and Food Chemistry* **1999**, *47* (5), 2060–2066. https://doi.org/10.1021/jf9813586
16. Fujieda, M.; Tanaka, T.; Suwa, Y.; Koshimizu, S.; Kouno, I. Isolation and Structure of Whiskey Polyphenols Produced by Oxidation of Oak Wood Ellagitannins. *Journal of Agricultural and Food Chemistry* **2008**, *56* (16), 7305–7310. https://doi.org/10.1021/jf8012713

17. Marchal, A.; Cretin, B. N.; Sindt, L.; Waffo-Téguo, P.; Dubourdieu, D. Contribution of oak lignans to wine taste: chemical identification, sensory characterization and quantification. *Tetrahedron* **2015**, *71* (20), 3148–3156. https://doi.org/10.1016/j.tet.2014.07.090

18. Martínez Montero, C.; Rodríguez Dodero, M. del C.; Guillén Sánchez, D. A.; García Barroso, C. Sugar Contents of Brandy de Jerez during Its Aging. *Journal of Agricultural and Food Chemistry* **2005**, *53* (4), 1058–1064. https://doi.org/10.1021/jf0403078

19. Carlavilla, D.; Villamiel, M.; Martínez-Castro, I.; Moreno-Arribas, M. V. Occurrence and Significance of Quercitol and Other Inositols in Wines during Oak Wood Aging. *American Journal of Enology and Viticulture* **2006**, *57* (4), 468–473. https://doi.org/10.5344/ajev.2006.57.4.468

20. Alañón, M. E.; Díaz-Maroto, M. C.; Díaz-Maroto, I. J.; Vila-Lameiro, P.; Pérez-Coello, M. S. Cyclic Polyalcohols: Fingerprints To Identify the Botanical Origin of Natural Woods Used in Wine Aging. *Journal of Agricultural and Food Chemistry* **2011**, *59* (4), 1269–1274. https://doi.org/10.1021/jf104737n

21. Guillén, M. D.; Ruiz, A. ¹H nuclear magnetic resonance as a fast tool for determining the composition of acyl chains in acylglycerol mixtures. *European Journal of Lipid Science and Technology* **2003**, *105* (9), 502–507. https://doi.org/10.1002/ejlt.200300799

22. Godelmann, R.; Fang, F.; Humpfer, E.; Schütz, B.; Bansbach, M.; Schäfer, H.; Spraul, M. Targeted and Nontargeted Wine Analysis by ¹H NMR Spectroscopy Combined with Multivariate Statistical Analysis. Differentiation of Important Parameters: Grape Variety, Geographical Origin, Year of Vintage. *Journal of Agricultural and Food Chemistry* **2013**, *61* (23), 5610–5619. https://doi.org/10.1021/jf400800d

23. Electronic Code of Federal Regulations. Title 27, Chapter I, Subchapter A, Part 5 — Labeling and Advertising of Distilled Spirits, §§ 5.143, 5.151 and 5.155. https://www.ecfr.gov/current/title-27/chapter-I/subchapter-A/part-5 (accessed October 22, 2026).

24. United Kingdom. The Scotch Whisky Regulations 2009, SI 2009/2890, regulation 3. https://www.legislation.gov.uk/uksi/2009/2890/regulation/3 (accessed October 22, 2026).

25. European Parliament and Council of the European Union. Regulation (EU) 2019/787 of 17 April 2019 on the definition, description, presentation and labelling of spirit drinks, the use of the names of spirit drinks in the presentation and labelling of other foodstuffs, the protection of geographical indications for spirit drinks, the use of ethyl alcohol and distillates of agricultural origin in alcoholic beverages, and repealing Regulation (EC) No 110/2008. *Official Journal of the European Union* **2019**, *L 130*, 1–54. https://eur-lex.europa.eu/eli/reg/2019/787/oj/eng (accessed October 22, 2026).

26. Institut National de l'Origine et de la Qualité (INAO). Cahier des charges de l'appellation d'origine contrôlée « Cognac ». https://extranet.inao.gouv.fr/fichier/Cahier-des-charges-Cognac---PNO.pdf (accessed October 22, 2026).

27. Canadian Food Inspection Agency. Canadian Food Compositional Standards, Volume 2: Alcoholic Beverages, Standard 2.1.3, Canadian Whisky. https://inspection.canada.ca/en/about-cfia/acts-and-regulations/list-acts-and-regulations/documents-incorporated-reference/canadian-food-compositional-standards-0 (accessed October 22, 2026).

28. Secretaría de Economía (México). Norma Oficial Mexicana NOM-006-SCFI-2012, Bebidas alcohólicas—Tequila—Especificaciones. *Diario Oficial de la Federación*, December 13, 2012. https://sidof.segob.gob.mx/notas/docFuente/5282165 (accessed October 22, 2026).

29. Golon, A.; Kuhnert, N. Unraveling the Chemical Composition of Caramel. *Journal of Agricultural and Food Chemistry* **2012**, *60* (12), 3266–3274. https://doi.org/10.1021/jf204807z

30. López-Aguilar, R.; Zuleta-Prada, H.; Hernández-Montes, A.; Herbert-Pucheta, J. E. Comparative NMR Metabolomics Profiling between Mexican Ancestral & Artisanal Mezcals and Industrialized Wines to Discriminate Geographical Origins, Agave Species or Grape Varieties and Manufacturing Processes as a Function of Their Quality Attributes. *Foods* **2021**, *10* (1), 157. https://doi.org/10.3390/foods10010157

31. Flink, J.; Karel, M. Retention of organic volatiles in freeze-dried solutions of carbohydrates. *Journal of Agricultural and Food Chemistry* **1970**, *18* (2), 295–297. https://doi.org/10.1021/jf60168a014

32. Nose, A.; Hamasaki, T.; Hojo, M.; Kato, R.; Uehara, K.; Ueda, T. Hydrogen Bonding in Alcoholic Beverages (Distilled Spirits) and Water−Ethanol Mixtures. *Journal of Agricultural and Food Chemistry* **2005**, *53* (18), 7074–7081. https://doi.org/10.1021/jf058061+

33. 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

34. Garrido, B. C.; Carvalho, L. J.; Burton, I. W.; McCarron, P. High-Accuracy Quantitative Nuclear Magnetic Resonance Using Improved Solvent Suppression Schemes. *Analytical Chemistry* **2025**, *97* (39), 21240–21248. https://doi.org/10.1021/acs.analchem.5c01139

35. Lu, J.; Wang, P.; Wang, Q.; Wang, Y.; Jiang, M. Validated ¹H and ¹³C Nuclear Magnetic Resonance Methods for the Quantitative Determination of Glycerol in Drug Injections. *Molecules* **2018**, *23* (5), 1177. https://doi.org/10.3390/molecules23051177

36. Pauli, G. F.; Jaki, B. U.; Lankin, D. C. Quantitative ¹H NMR: Development and Potential of a Method for Natural Products Analysis. *Journal of Natural Products* **2005**, *68* (1), 133–149. https://doi.org/10.1021/np0497301

37. 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

38. 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

39. 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

40. Herbert-Pucheta, J. E.; Mejía-Fonseca, I.; Zepeda-Vallejo, L. G.; Milmo-Brittingham, D.; Maya, G. P. The "Wine-T_1" NMR experiment for novel wine-metabolome fingerprinting with nuclear-spin relaxation. *BIO Web of Conferences* **2019**, *12*, 02029. https://doi.org/10.1051/bioconf/20191202029

41. de las Heras, E.; Zuriarrain-Ocio, A.; Zuriarrain, J.; Bordagaray, A.; Dueñas, M. T.; Berregi, I. Quantitative Determination of Acrolein in Cider by ¹H NMR Spectrometry. *Foods* **2020**, *9* (12), 1820. https://doi.org/10.3390/foods9121820

42. 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

43. van den Berg, R. A.; Hoefsloot, H. C. J.; 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

44. 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

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

46. 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

47. 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

48. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9* (86), 2579–2605. https://jmlr.org/papers/v9/vandermaaten08a.html

49. Venna, J.; Kaski, S. Local multidimensional scaling. *Neural Networks* **2006**, *19* (6–7), 889–899. https://doi.org/10.1016/j.neunet.2006.05.014

50. 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

51. Barkhuijsen, H.; de Beer, R.; van Ormondt, D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *Journal of Magnetic Resonance* **1987**, *73* (3), 553–557. https://doi.org/10.1016/0022-2364(87)90023-0

52. Vanhamme, L.; van den Boogaart, A.; Van Huffel, S. Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge. *Journal of Magnetic Resonance* **1997**, *129* (1), 35–43. https://doi.org/10.1006/jmre.1997.1244

53. Poullet, J.-B.; Sima, D. M.; Van Huffel, S. MRS signal quantitation: A review of time- and frequency-domain methods. *Journal of Magnetic Resonance* **2008**, *195* (2), 134–144. https://doi.org/10.1016/j.jmr.2008.09.005

54. Krishnamurthy, K. CRAFT (complete reduction to amplitude frequency table) – robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR. *Magnetic Resonance in Chemistry* **2013**, *51* (12), 821–829. https://doi.org/10.1002/mrc.4022

55. 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

56. 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

57. 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

58. Mo, H.; Harwood, J. S.; Raftery, D. Receiver gain function: the actual NMR receiver gain. *Magnetic Resonance in Chemistry* **2010**, *48* (3), 235–238. https://doi.org/10.1002/mrc.2563

59. Hoch, J. C.; Baskaran, K.; Burr, H.; Chin, J.; Eghbalnia, H. R.; Fujiwara, T.; Gryk, M. R.; Iwata, T.; Kojima, C.; Kurisu, G.; et al. Biological Magnetic Resonance Data Bank. *Nucleic Acids Research* **2023**, *51* (D1), D368–D376. https://doi.org/10.1093/nar/gkac1050

60. Roslund, M. U.; Tähtinen, P.; Niemitz, M.; Sjöholm, R. Complete assignments of the ^{1}H and ^{13}C chemical shifts and J(H,H) coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides. *Carbohydrate Research* **2008**, *343* (1), 101–112. https://doi.org/10.1016/j.carres.2007.10.008

61. Barclay, T.; Ginic-Markovic, M.; Johnston, M. R.; Cooper, P.; Petrovsky, N. Observation of the keto tautomer of D-fructose in D_2O using ^{1}H NMR spectroscopy. *Carbohydrate Research* **2012**, *347* (1), 136–141. https://doi.org/10.1016/j.carres.2011.11.003

62. Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; et al. Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets. *Analytical Chemistry* **2005**, *77* (5), 1282–1289. https://doi.org/10.1021/ac048630x

63. Amarasekara, A. S.; Williams, L. D.; Ebede, C. C. Mechanism of the dehydration of D-fructose to 5-hydroxymethylfurfural in dimethyl sulfoxide at 150 °C: an NMR study. *Carbohydrate Research* **2008**, *343* (18), 3021–3024. https://doi.org/10.1016/j.carres.2008.09.008

64. Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies. *Analytical Chemistry* **2006**, *78* (2), 363–371. https://doi.org/10.1021/ac051444m

65. Angyal, S. J. The Composition of Reducing Sugars in Solution. *Advances in Carbohydrate Chemistry and Biochemistry* **1984**, *42*, 15–68. https://doi.org/10.1016/S0065-2318(08)60122-5

66. Bock, K.; Pedersen, C. Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. *Advances in Carbohydrate Chemistry and Biochemistry* **1983**, *41*, 27–66. https://doi.org/10.1016/S0065-2318(08)60055-4

67. Sokal, R. R.; Rohlf, F. J. The Comparison of Dendrograms by Objective Methods. *Taxon* **1962**, *11* (2), 33–40. https://doi.org/10.2307/1217208

68. Hubert, L.; Arabie, P. Comparing Partitions. *Journal of Classification* **1985**, *2* (1), 193–218. https://doi.org/10.1007/BF01908075

69. 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

70. Smilde, A. K.; Kiers, H. A. L.; Bijlsma, S.; Rubingh, C. M.; van Erk, M. J. Matrix Correlations for High-Dimensional Data: The Modified RV-Coefficient. *Bioinformatics* **2009**, *25* (3), 401–405. https://doi.org/10.1093/bioinformatics/btn634

71. Jackson, D. A. PROTEST: A PROcrustean Randomization TEST of Community Environment Concordance. *Écoscience* **1995**, *2* (3), 297–303. https://doi.org/10.1080/11956860.1995.11682297

72. Peres-Neto, P. R.; Jackson, D. A. How Well Do Multivariate Data Sets Match? The Advantages of a Procrustean Superimposition Approach over the Mantel Test. *Oecologia* **2001**, *129* (2), 169–178. https://doi.org/10.1007/s004420100720

73. Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

74. Defernez, M.; Kemsley, E. K. The Use and Misuse of Chemometrics for Treating Classification Problems. *TrAC Trends in Analytical Chemistry* **1997**, *16* (4), 216–221. https://doi.org/10.1016/S0165-9936(97)00015-0

75. Kjeldahl, K.; Bro, R. Some Common Misunderstandings in Chemometrics. *Journal of Chemometrics* **2010**, *24* (7–8), 558–564. https://doi.org/10.1002/cem.1346

76. Rundlöf, T.; Mathiasson, M.; Bekiroglu, S.; Hakkarainen, B.; Bowden, T.; Arvidsson, T. Survey and qualification of internal standards for quantification by ¹H NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **2010**, *52* (5), 645–651. https://doi.org/10.1016/j.jpba.2010.02.007

77. Bastian, F.; Ito, Y.; Ogahara, E.; Ganeko, N.; Hatano, T.; Ito, H. Simultaneous Quantification of Ellagitannins and Related Polyphenols in *Geranium thunbergii* Using Quantitative NMR. *Molecules* **2018**, *23* (6), 1346. https://doi.org/10.3390/molecules23061346

78. Schneider, B.; Lichtenthaler, F. W.; Steinle, G.; Schiweck, H. Studies on Ketoses, 1 Distribution of Furanoid and Pyranoid Tautomers of D-Fructose in Water, Dimethyl Sulfoxide, and Pyridine via¹H NMR Intensities of Anomeric Hydroxy Groups in [D₆]DMSO. *Liebigs Annalen der Chemie* **1985**, *1985* (12), 2443–2453. https://doi.org/10.1002/jlac.198519851213

Summary

The non-volatile extract of 197 polar (D_2O or D_2O/CD_3OD) and 61 lipophilic ($CDCl_3$) freeze-dried residues of whiskeys, bourbons, rums, brandies and agave spirits, each prepared from 1.000 mL of spirit, was examined by ^{1}H NMR at 299.99 MHz on an absolute scale. The complete decays, parameter sets and processed spectra were recovered from the MestReNova documents, the processing was reproduced exactly, a silent normalization of the first D_2O set was reversed, and every spectrum was expressed in micromoles of ^{1}H per liter of spirit through an external maleic acid standard. The residual CHD_2OD of the CD_3OD served as an internal tracer: it agreed with the same signal in the standard within 3.4 %, confirmed the transient, gain and volume corrections, exposed twenty incorrect solvent labels and identified two acquisition sessions in which nine spectra had lost 29–46 % of their sensitivity, a deficit that was confirmed by TMSP and by a replicate preparation and was corrected. The expanded uncertainty of a well-resolved result was estimated at about 9 %, rising to 17 % for broadened or partly resolved signals and to about 40 % for aldehyde signals acquired at the shortest repetition times, with the recovery through freeze-drying, which was not measured, outside that budget.

Constrained multiplet fitting quantified 17 congeners or congener groups in the polar residues and 13 in the lipophilic residues. Sweetening was the largest compositional signal in the data set. The cognacs carried 4.7–14.0 g L^{-1} of sugar, mostly as equimolar glucose and fructose, eight of 29 rums carried 5–25.1 g L^{-1} of sucrose or invert sugar, and two maple bourbons and a honey whiskey carried 11.8–29.2 g L^{-1} of sucrose, whereas the unsweetened whiskeys contained only 0.04–0.21 g L^{-1} of the cell-wall sugars released by oak. Vanillin without syringaldehyde, with 1,2-propanediol in two of the three cases, marked flavor preparations in two rums and an American blended whiskey. Quercitol, the benzaldehydes, gallic acid and the cinnamaldehydes traced oak contact; glycerol, succinic acid and lactic acid traced wine-cask seasoning; and a broad formyl envelope near δ 9.48 was proposed, but not yet calibrated, as an index of caramel. Retained ethanol followed the sugar content of each residue ($\rho = 0.77$), as volatile trapping in a carbohydrate glass predicts, so that ethanol and the fusel alcohols in a freeze-dried residue are artifacts of the preparation rather than measures of the spirit. In $CDCl_3$ the lipid content differed by more than an order of magnitude between products and was highest in the cask-strength Kavalan malts, and the unsaturation that each additional acyl chain brought suggested, without establishing, that the lipid-rich malt whiskies owe most of their additional lipid to saturated yeast-derived chains.

Two lessons of general value emerged from the quantitation. The gram-per-liter glycerol that a single-multiplet fit had assigned to a fructose-rich reposado tequila was shown, from the absence of the second glycerol multiplet, to be the β-pyranose H-1b doublet of fructose; glycerol in sugar-rich residues is therefore reported only where both multiplets agree, and the tequila is described by its 3.54 g L^{-1} of fructose, a composition of the kind permitted in an *abocado* product. Relative to the line of the standard, the fitted Lorentzian areas of broadened or partly resolved lines exceeded their direct integrals by 10–13 %, an excess that was carried as an uncertainty component rather than applied as a correction. Neither problem would have been visible in the fit statistics.

Pareto-scaled absolute spectra were analyzed by PCA, HCA, t-SNE, UMAP, PLS-DA, SVM and *k*-nearest neighbors. The second principal component of the non-carbohydrate, volatile-free block contrasted products containing 1,2-propanediol and vanillin with bourbons and ryes matured in new charred oak. Seven categories were classified with a balanced accuracy of 0.78 by a linear SVM under nested, product-grouped cross-validation (permutation $p = 0.005$), falling to 0.67 when whole producers were held out, and the acquisition set could itself be predicted from the spectra, so the classification models are descriptive and are not proposed for new samples. The embeddings were faithful displays of the scaled space but contained no additional class information. PLS calibrations with leave-10 %-out cross-validation gave Q^2

of 0.32–0.76 on a logarithmic scale; masked models separated calibrations carried by the analyte's own signals from those carried by co-varying compounds, and a first-pass glycerol calibration with Q^2 = 0.95 was shown to have learned the sugar signals that had contaminated its reference values.

Every linear analysis on mean-centered data was shown to be identical in the frequency and time domains, to within numerical precision, whereas variable-wise scaling, alignment and binning are naturally frequency-domain operations; band-selective HLSVD of the decays reproduced the frequency-domain quantitation of isolated resonances and failed, without prior knowledge, for an overlapped doublet. Residues of the same product prepared under different sample numbers differed by up to a factor of 3.1 in non-volatile content, whereas the one true replicate preparation agreed within 9 %, so bottle-to-bottle variation and recovery through freeze-drying remain the dominant unvalidated steps. The most important steps toward a quantitative authentication method are a non-volatile recovery standard added before freeze-drying, quantitative acquisition with measured relaxation times, replicate preparations and bottles, structured sample records and confirmation of the regulatory assignments by an orthogonal method.

Version 2.1 placed the 29 DMSO-d_6 residues on the same absolute footing. The residual DMSO-d_5 of the solvent served as an internal tracer in place of CHD_2OD and revealed two solvent lots with within-lot scatter of 7–9 %, TMS was absent from 16 of the 29 spectra, and in residues rich in sugars and acids the residual solvent line moved by up to 0.037 ppm while the solute lines did not, so that the spectra had to be referenced to TMS or by cross-correlation with the TMS-referenced median. Two title conflicts were resolved by comparing the congeners of each DMSO-d_6 residue with those of the D_2O/CD_3OD and $CDCl_3$ residues of the candidate products. HMF, the benzaldehydes and quercitol were recovered at 1.08, 1.11 and 0.97 times their D_2O/CD_3OD levels, which transfers the external-standard scale to DMSO-d_6 within about 8 % on average. In one tube the DMSO-d_6 spectra showed the syringaldehyde, vanillin and cinnamaldehyde signals individually, ellagic acid, abundant in four Kavalan malts, a syringyl singlet assigned tentatively to syringic acid, gallic acid, the glucose and pentose anomeric doublets and the ethyl esters, free fatty acids and unsaturated chains that otherwise require a $CDCl_3$ residue. They lost glycerol beneath the exchange-averaged water line and the lactic acid doublet beneath the lipid envelope, recovered only 0.32 of the succinic acid of the wine-seasoned residues, plausibly because part of it is present as salts, and gave glucose values carrying an additive interference of about 77 mg L^{-1}; and the line shapes of their anomeric signals depended on the rate of hydroxyl exchange, a property of the residue rather than of its sugar content.

The DMSO-d_6 spectra also exposed two pentoses that version 2.0 had overlooked. In D_2O/CD_3OD the α-L-arabinopyranose and β-D-xylopyranose H-1 doublets are resolved, the β-arabinopyranose doublet coincides with the α-glucose doublet, and the remainder of the α-glucose region matched the anomeric equilibrium of arabinose in water (β/α = 0.53 against 0.59). Glucose was therefore recomputed from its β-H-1 doublet and an α/β ratio of 0.860 measured in glucose-rich residues of the same solvent system; the version 2.0 glucose had been overestimated by a median of 75 % in the American whiskeys and 41 % in the Kavalan malts, and underestimated in residues whose CD_3OD content had displaced their anomeric signals. Arabinose was more abundant than glucose in the American whiskeys matured in new charred oak, consistent with the preferential hydrolysis of the arabinofuranosyl side chains of the stave hemicelluloses, and in twenty-two residues a doublet at δ 4.474 without its β-arabinopyranose partner pointed to an unidentified glycoside. The re-quantitation with residue-specific line positions changed the interpretation of four sweetened products: Pyrat XO Reserve and a 16-year-old Trinidad rum contain invert sugar rather than a fructose-rich syrup, Larsen Aqua Ignis contains the invert sugar of the other cognacs with a large residue of sucrose rather than a fructose-rich syrup, and Tromba Reposado, although it contains 333 mg L^{-1} of glucose rather than almost none, remains dominated by fructose.

The chemometric analysis was given a full accounting: seventeen data approaches, from the full and non-carbohydrate D_2O/CD_3OD blocks to the quantified congeners and the $CDCl_3$, fused and DMSO-d_6 blocks, were each analyzed by PCA, four hierarchical clusterings, t-SNE and UMAP. Average linkage gave the most faithful tree in seventeen of the 17 blocks and never the most informative partition, because it chained; Ward linkage and the correlation distance gave the best agreement with known classes, but no approach recovered the spirit categories well, the categories not being the natural clusters of the residues. The embeddings were trustworthy displays of the scaled space but added no class information. The second principal component of the $CDCl_3$ block was shown to be water extending beyond the excluded window in two residues, and widening that window removed it. On the 24 products with D_2O/CD_3OD and DMSO-d_6 residues the two configurations agreed to a modified RV coefficient of 0.77, and on the 10 products with residues in all three solvents DMSO-d_6 reproduced the fused two-solvent configuration to 0.77, while the D_2O/CD_3OD and $CDCl_3$ configurations were unrelated to each other; DMSO-d_6 alone separated wine-seasoned from other casks as well as the D_2O/CD_3OD block, with the caveat that ten and twenty-four products, most of them from one producer, cannot rank the solvents.

The question of one solvent or two was therefore answered on the evidence. For authentication and regulatory work the D_2O/CD_3OD residue remains indispensable, because it alone measures glycerol, the complete sugar profile and the wine-cask acids on which a finding of addition or of wine seasoning would rest. As the second solvent, DMSO-d_6 is preferred for a compositional fingerprint of oak extraction, because it resolves the aldehydes more completely than $CDCl_3$ and adds ellagic acid, and $CDCl_3$ is preferred where the lipid profile matters. A DMSO-d_6 protocol with a small proportion of D_2O, TMS or an internal standard in every tube and a 90° pulse and T_1 measured in the solvent, together with a side-by-side preparation of the same bottles with spiking of glycerol, succinic acid and glucose, would settle the discrepancies that remain.

Version 2.2 corrected two errors of the absolute scale that every earlier version shared. The parameter sets record a 90° pulse of 10 µs, whereas the calibrated 90° pulse of the probe is 13 µs, so that the 8 µs pulse of the blank, of the D_2O spectra and of most $CDCl_3$ spectra was a 55.4° pulse and the 7 µs pulse of eleven $CDCl_3$ spectra and of every DMSO-d_6 spectrum a 48.5° pulse. Because the blank and the D_2O residues share the same pulse, the sine of the flip angle cancels between them, and for those spectra the correction acts through the steady-state recovery of Equation 4 alone: the recovery of maleic acid in the blank rose from 0.9703 to 0.9812, while the saturation corrections of the analytes became smaller. The standard itself proved to contain 95 % of the maleic acid weighed into it, an effective concentration of 32.11 rather than 33.80 mmol L^{-1}, so that every earlier concentration had been 5.3 % too high, whatever the solvent, the analyte or the acquisition. Together the two corrections lowered the absolute spectra by 3.9 % for the 8 µs and by 1.0 % for the 7 µs acquisitions, and the reported concentrations fell by 3.8–9.5 % in D_2O/CD_3OD, 1.4–8.2 % in $CDCl_3$ and 1.1–6.5 % in DMSO-d_6, the largest decreases belonging to slowly relaxing formyl and aromatic protons in the spectra acquired with the shortest repetition times. The reporting thresholds in mg L^{-1}, which describe where one signal begins to obscure another, were carried with the corrected scale, so that no flag changed; the balanced accuracies and the cross-validated Q^2 changed by no more than 0.034 and 0.015, and none of the conclusions changed. Neither correction can be confirmed from the data themselves. The CHD_2OD tracer is insensitive to the flip angle, because a different T_1 absorbs a different angle, the residual $CHCl_3$ is confounded by a batch effect, and no internal signal measures the content of the standard, which scales every concentration alike; a nutation calibration recorded with every data set and a standard prepared from, or checked against, a certified reference material are therefore part of the recommended protocol.

Version 2.3 added a library of reference spectra: 58 spectra of 45 pure congeners in five classes, from ethanol, acetaldehyde, ethyl acetate and the fusel alcohols through the esters and short-chain acids, the oak aromatics and whisky lactone and the hydroxycinnamic acids to the sugars, polyols and non-volatile acids, recorded in whichever of D_2O, D_2O/CD_3OD, D_2O/ethanol, $CDCl_3$ and DMSO-d_6 each compound would dissolve, with a clover honey and an eleven-step model spirit. One spirit measured intact showed what freeze-drying does. Its native spectrum was dominated by the volatile congeners, the fusel alcohols, the isoamyl methylene, the acetyl singlets of ethyl acetate and acetic acid, the ethyl acetate quartet, acetaldehyde and an unassigned singlet at δ 8.217 that no residue contains, while its oak aromatics lay at the noise level; in the residue of the same bottle the aromatics rose by about the 8.4-fold that concentration and acquisition predict, whereas the fusel alcohols kept about 1.4 %, the ethanol 0.007 %, the ethyl acetate and acetic acid at most 1.7 and 0.3 % and the acetaldehyde at most 8.0 % of their intensities. Native and freeze-dried spectra are therefore complementary, the first measuring the volatile fraction that the second cannot hold. The pure compounds reproduced the partition of a residue between its solvents: the medium-chain ethyl esters gave strong spectra only in $CDCl_3$, *p*-coumaric acid gave none in $CDCl_3$, and syringaldehyde was measured at 60 °C in D_2O/CD_3OD and both syringyl compounds at 50 °C in $CDCl_3$.

The references transferred well where the medium was matched: α-glucose of the honey and the residues dissolved in D_2O agreed within 0.001 ppm, 1,2-propanediol within 0.01 ppm, and the D_2O/CD_3OD residues placed the anomeric doublets 0.028 ppm upfield, the medium displacement already known. Lactic, gallic and syringic acid moved with acidity, and the spectra recorded without an internal reference carried errors of about 0.03 ppm. Syringaldehyde was the exception: its ring protons appeared at δ 7.15–7.23 in every medium examined, 0.08–0.10 ppm upfield of the δ 7.303 residue line assigned to them, so that this assignment, retained as probable, now rests on the correlation of that line with the formyl line and on its ^{13}C shift; the companion singlet at δ 7.332 is assigned tentatively to syringic acid, which lies downfield of syringaldehyde in every aqueous medium and whose ratio to the δ 7.30 line (0.47) matches the DMSO-d_6 pair (0.47). The library also exposed two overlaps that the residues alone did not reveal, 1,3-propanediol under the quercitol H-6ax multiplet and mannose near the α-xylopyranose doublet, and both, with the syringyl assignments, can be settled by adding the compounds to a residue.

A second round of PLS calibrations was built after the removal of 226 flagged exclusions (unreliable reference values and the two equal-volume residues) and 226 diagnosed ones (vertical outliers and bad leverage points from robust standardized residuals and the 99 % limits of T^2 and Q) from 28 models. Scoring the round-1 predictions on the retained residues showed that most of the apparent improvement came from no longer scoring the excluded residues. Genuine gains on refitting were confined to the benzaldehydes (Q^2 0.17 → 0.82, masked 0.15 → 0.72), the syringyl index, gallic acid and the logarithmic glycerol and lactic acid models; the sugar, 1,2-propanediol, cinnamaldehyde and HMF models became worse, because their outliers are the sparse upper end of real distributions, and the cleaned logarithmic glycerol model is valid only above about 16 mg L^{-1}. The diagnosed residues remained poorly predicted by models that had not seen them (a median of 4.3 times the RMSECV), so they were outliers in fact and not artifacts of round 1, and the round-1 models remain the reference for screening. Every stacked or overlaid spectral figure was redrawn with one vertical-scaling rule that limits the magnification of noise, and an Addendum reproduces every score plot, embedding and dendrogram of Section 12 at full page with the residues named.

Addendum. Full-page chemometric plots

The composite figures of Section 12 show each data approach on a single page, at a size at which the residue names cannot be read. This Addendum reproduces every score plot, embedding and dendrogram of those figures at full page, with every residue named, in the order of Table 20: the two PCA projections, the t-SNE and UMAP embeddings and the four dendrograms of each block, one or more to a page according to the number of residues. The plots were computed from the same data and with the same settings as the composites (Section 4.16); names are placed beside their points, with leader lines where the points are crowded, and residues that share a product name are distinguished by their index. The figures are numbered S1–S94, and the heading of each block gives the composite figure that it enlarges.

Block D2O-F: Figures S1–S8, enlarged from Figure 39

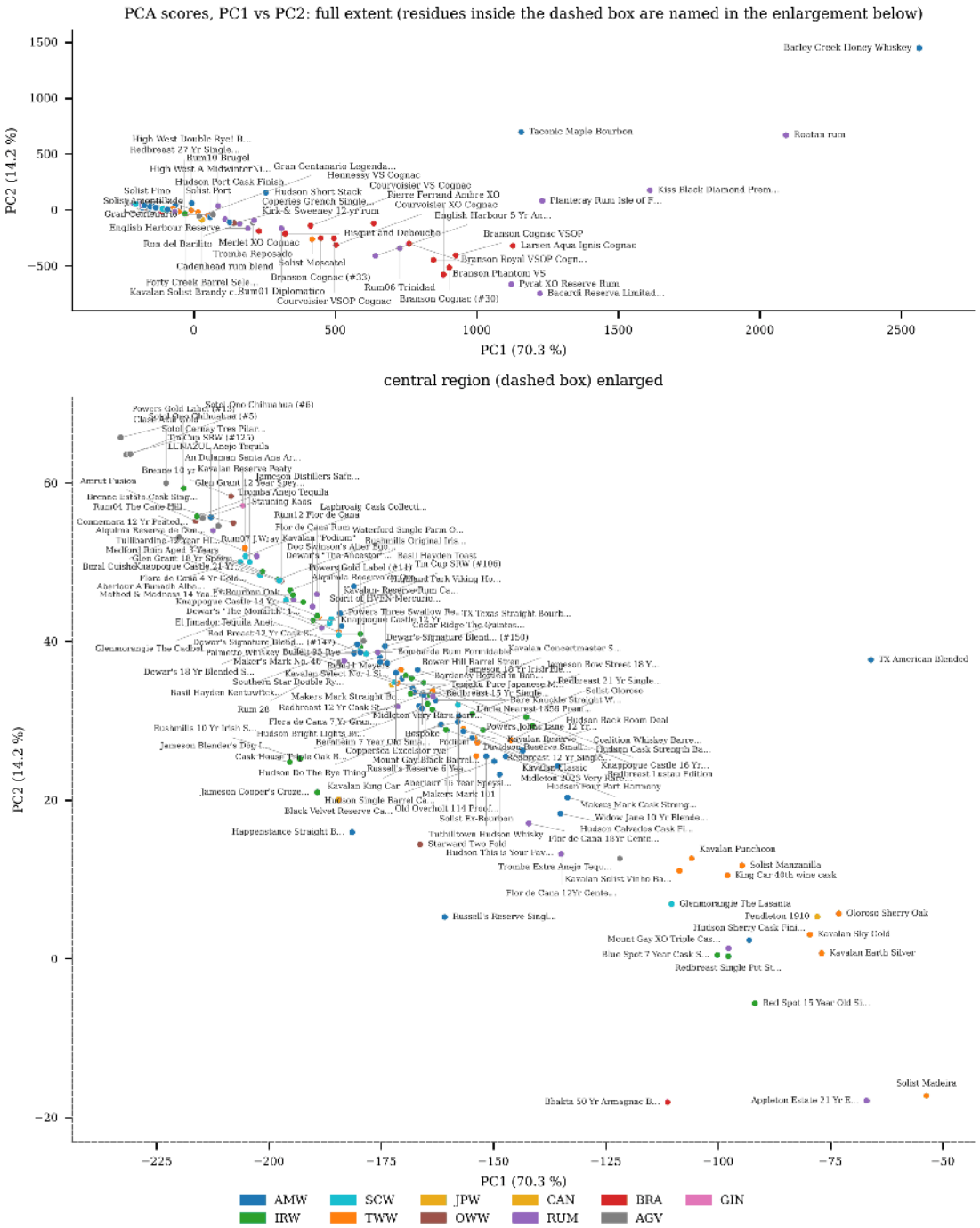

Figure S1. Block D2O-F (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

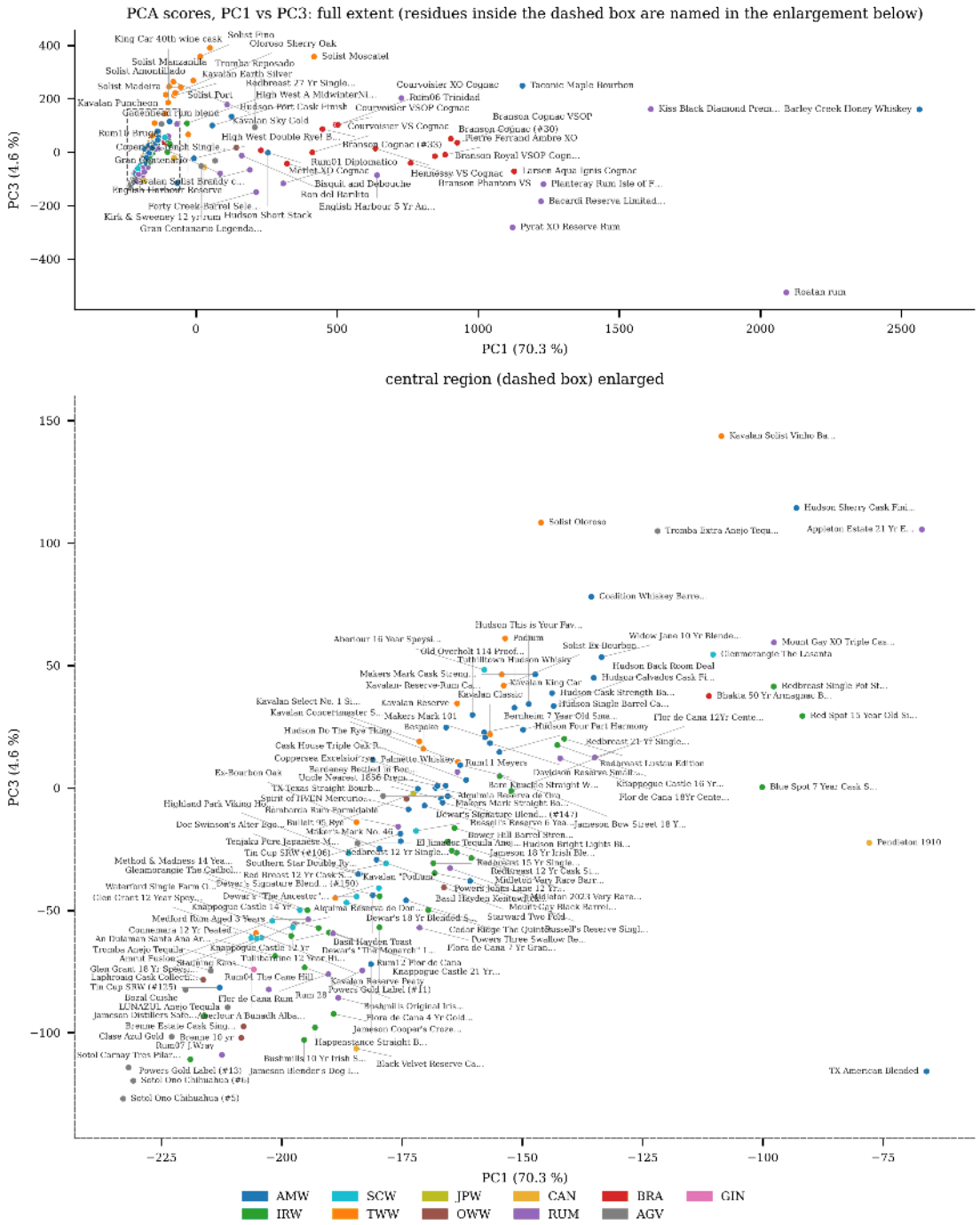

Figure S2. Block D2O-F (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

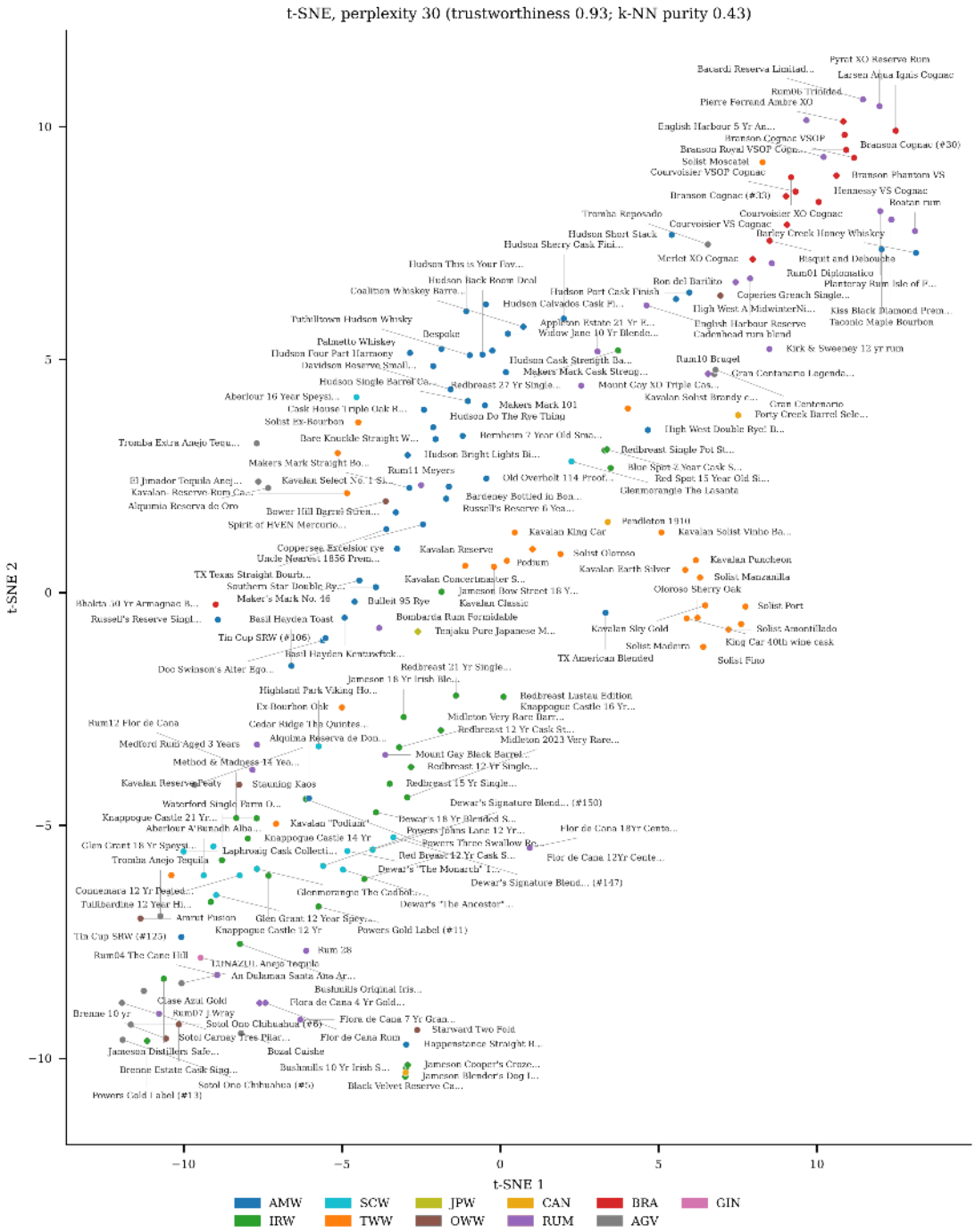

Figure S3. Block D2O-F (Table 20): t-SNE, perplexity 30 (trustworthiness 0.93; k-NN purity 0.43), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

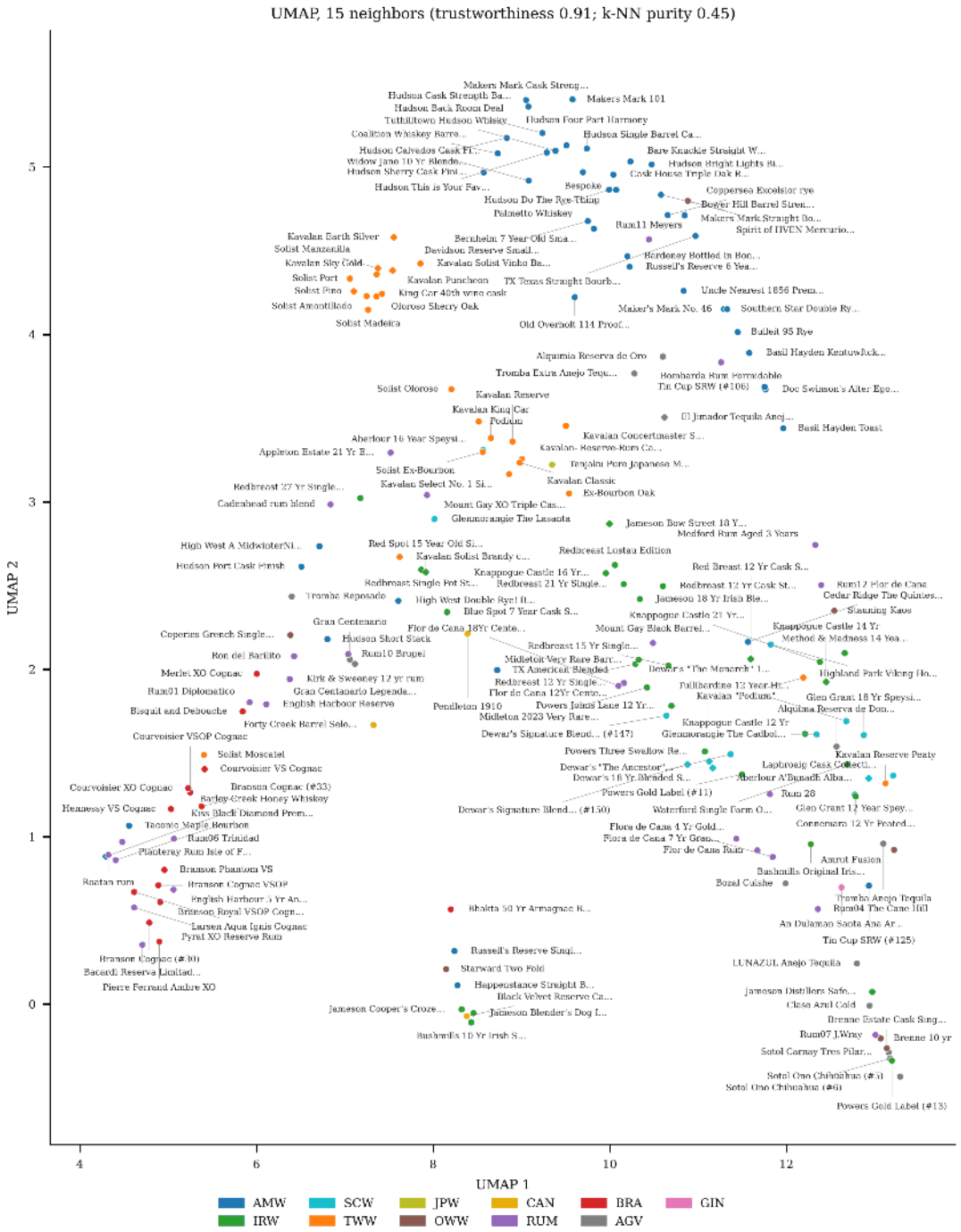

Figure S4. Block D2O-F (Table 20): UMAP, 15 neighbors (trustworthiness 0.91; k-NN purity 0.45), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

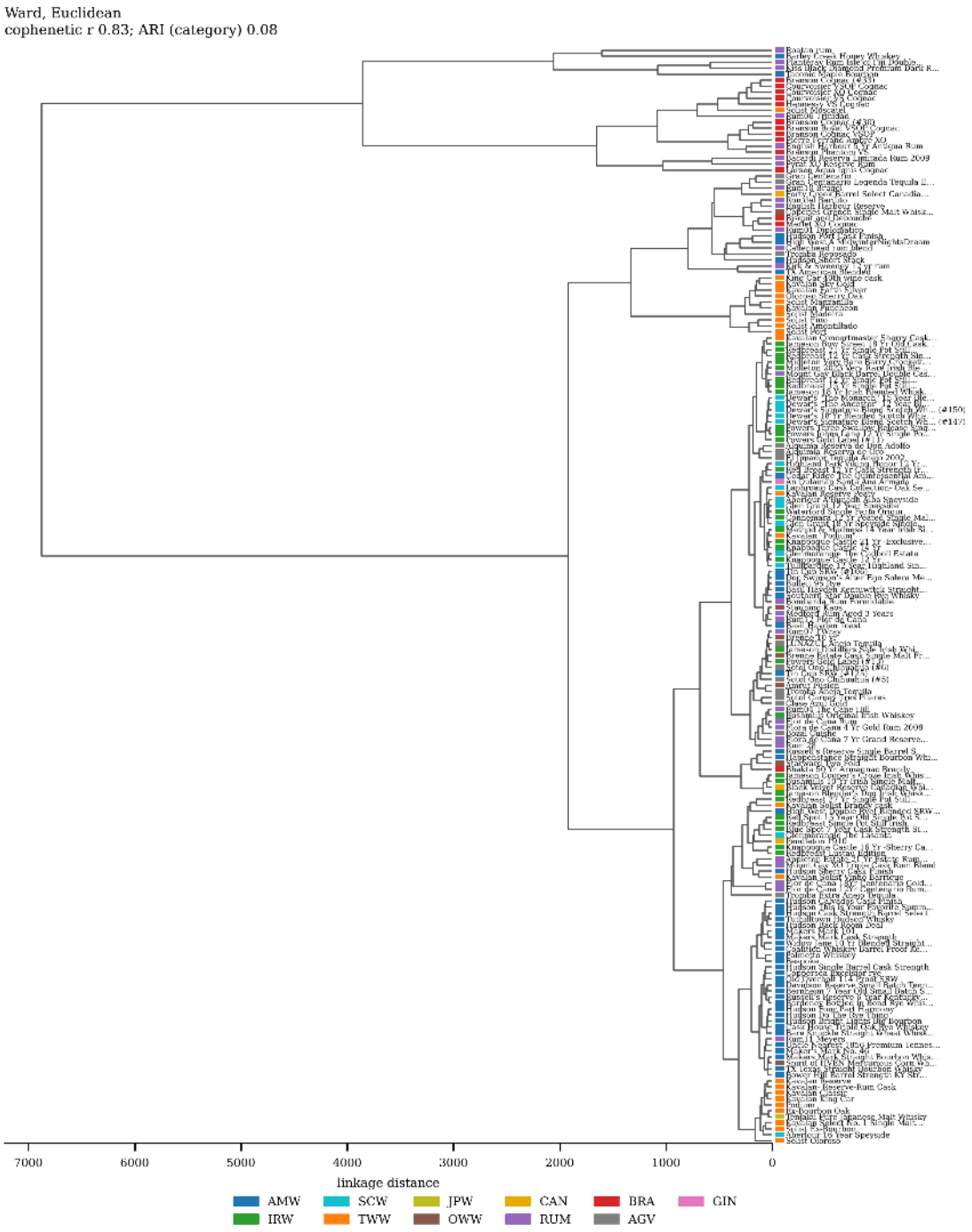

Figure S5. Block D2O-F (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

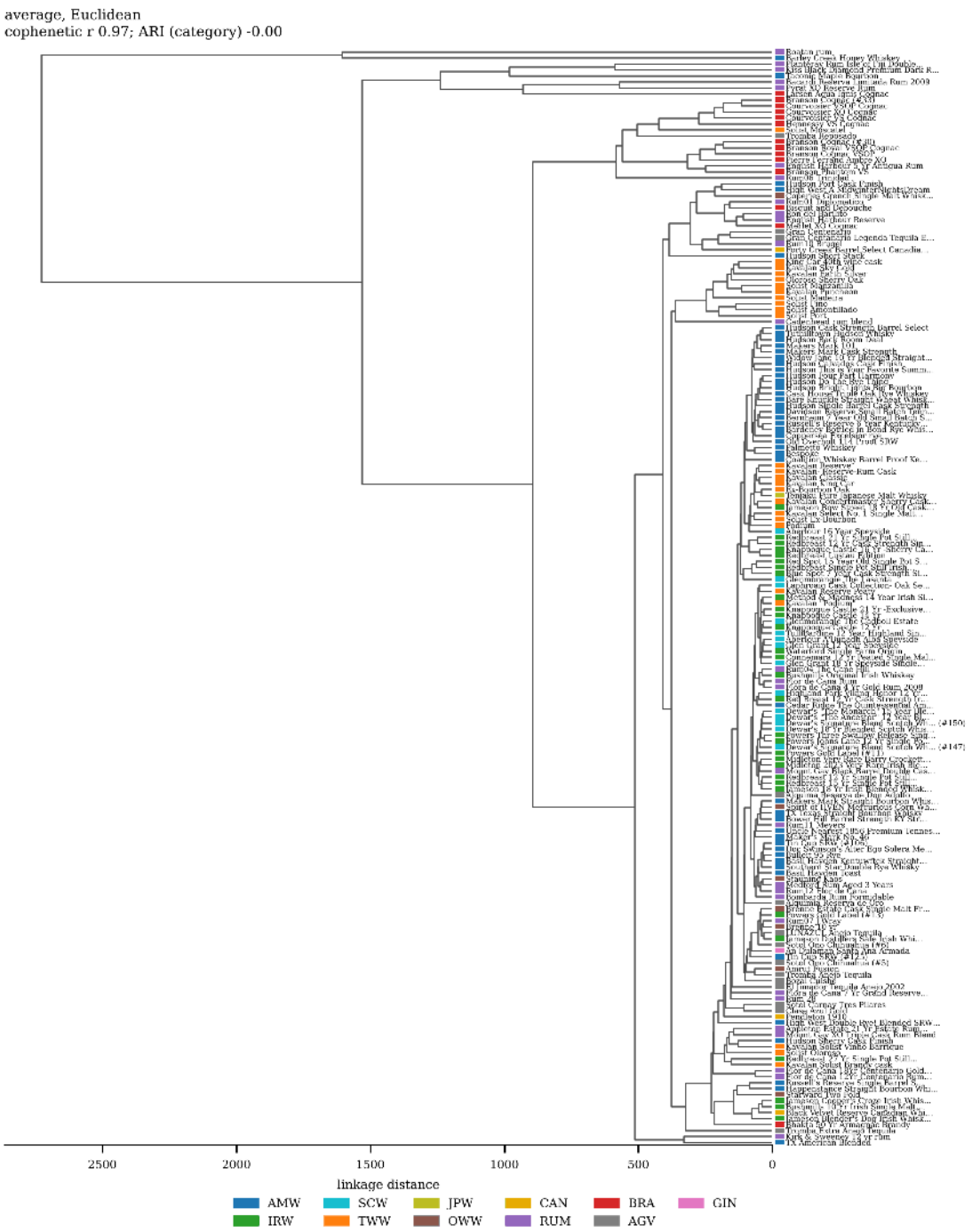

Figure S6. Block D2O-F (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

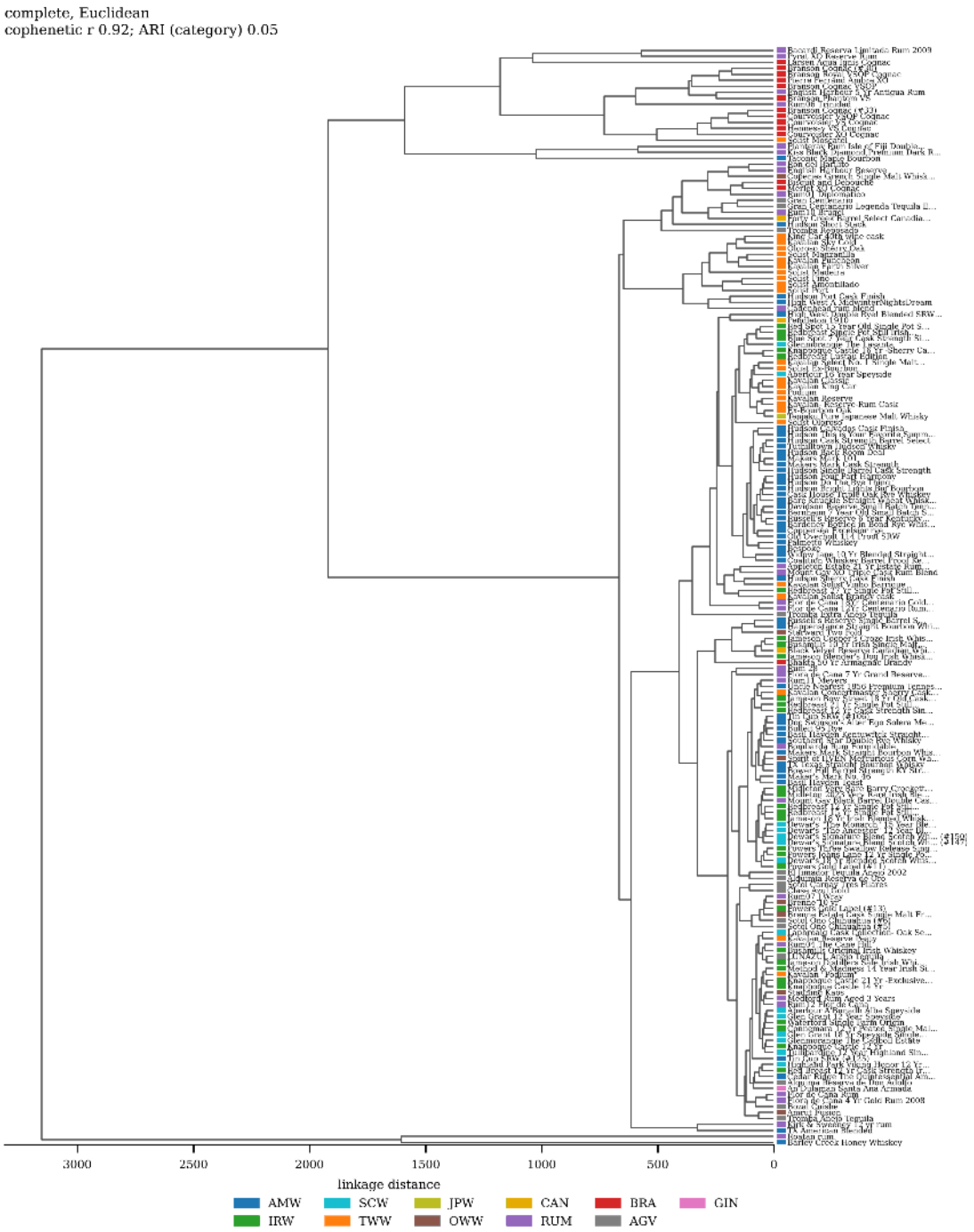

Figure S7. Block D2O-F (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

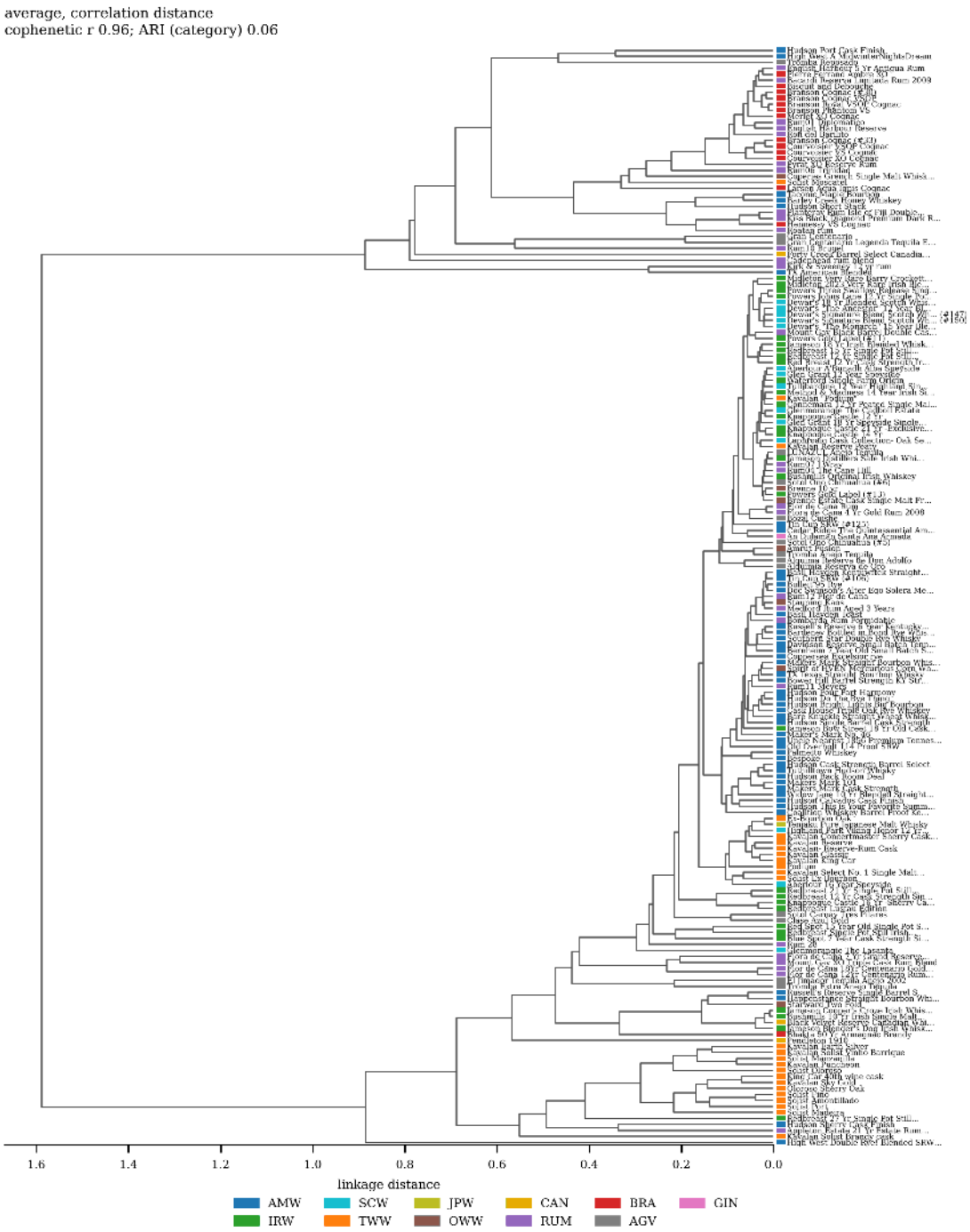

Figure S8. Block D2O-F (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-NC: Figures S9–S16, enlarged from Figure 40

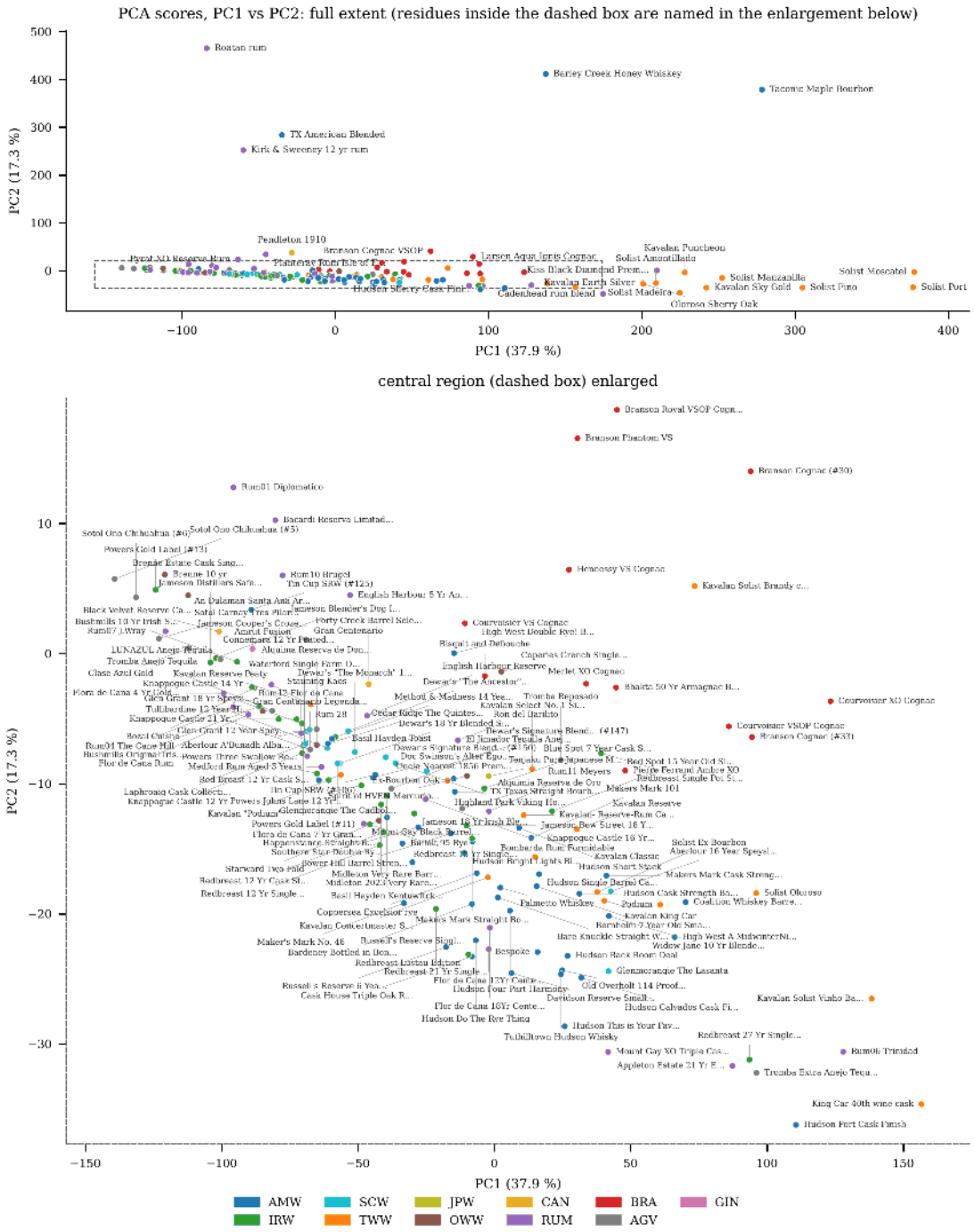

Figure S9. Block D2O-NC (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

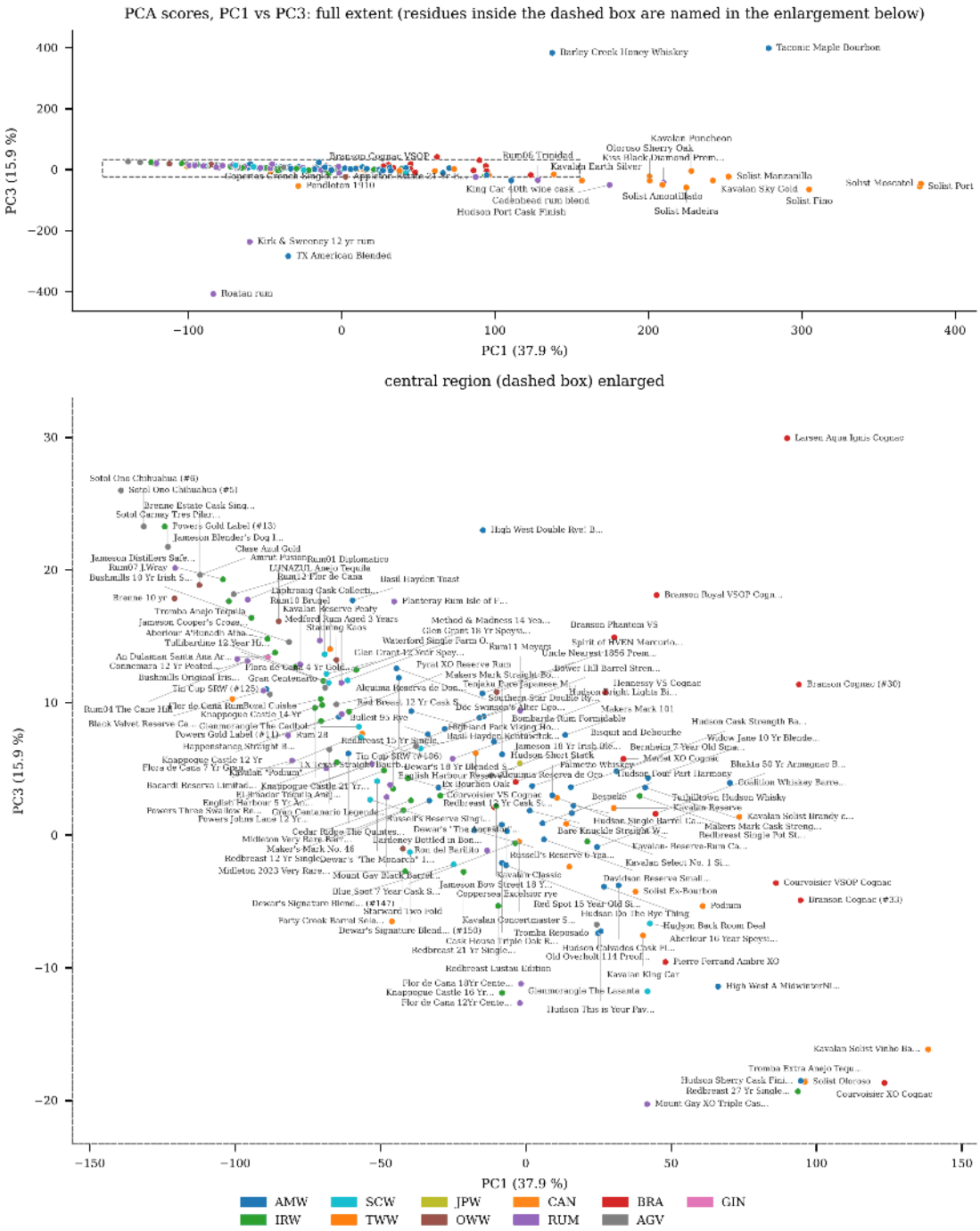

Figure S10. Block D2O-NC (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

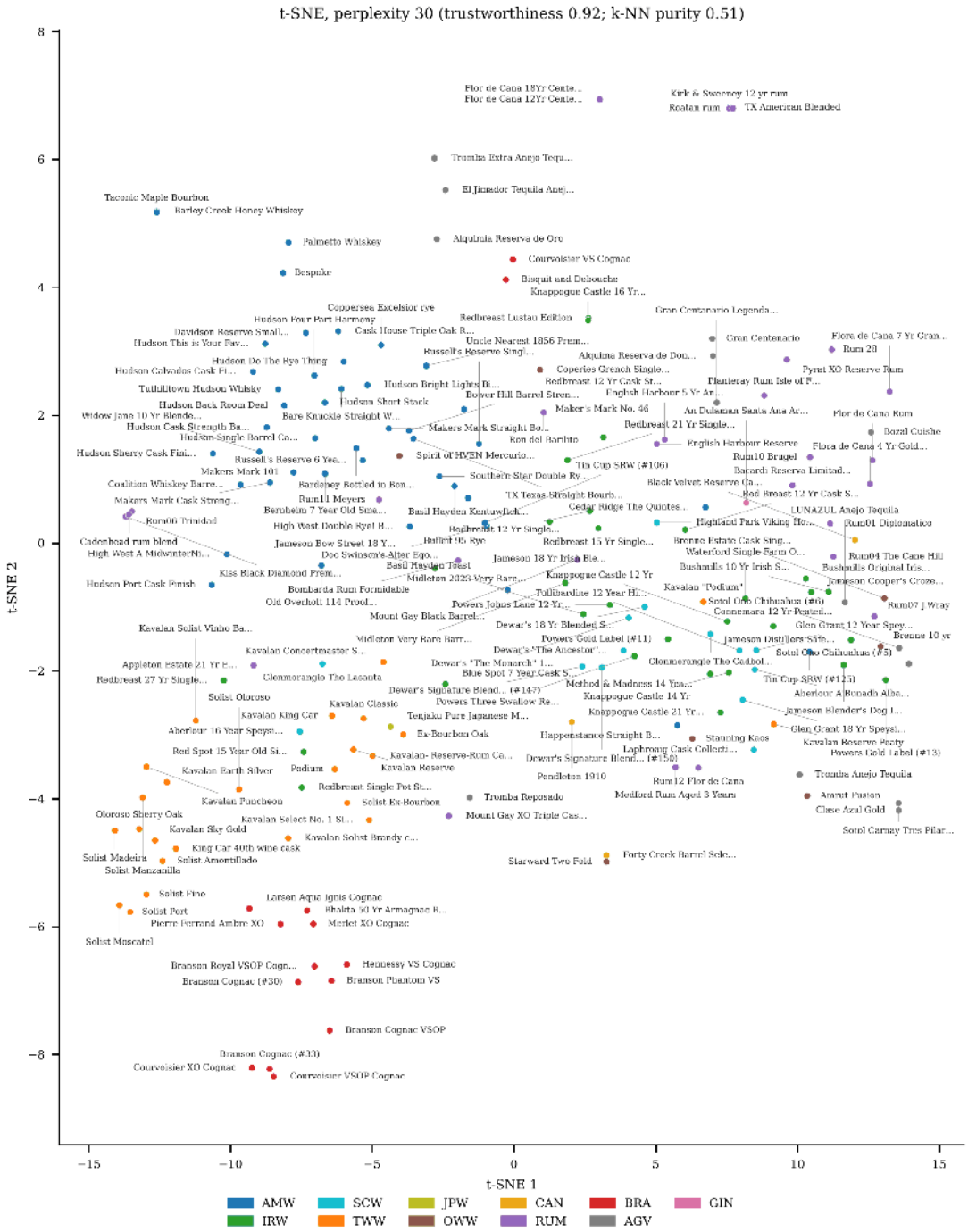

Figure S11. Block D2O-NC (Table 20): t-SNE, perplexity 30 (trustworthiness 0.92; k-NN purity 0.51), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

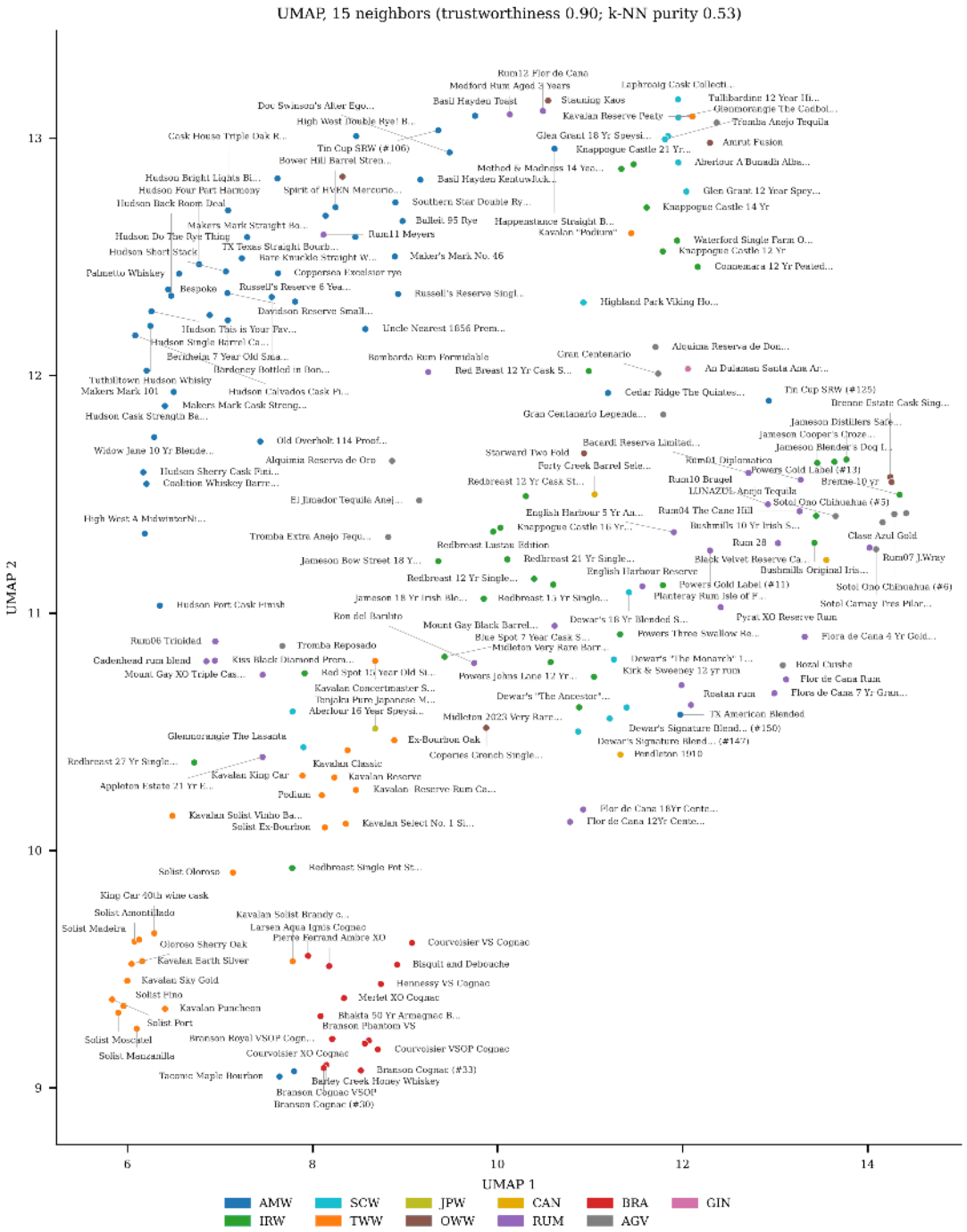

Figure S12. Block D2O-NC (Table 20): UMAP, 15 neighbors (trustworthiness 0.90; k-NN purity 0.53), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

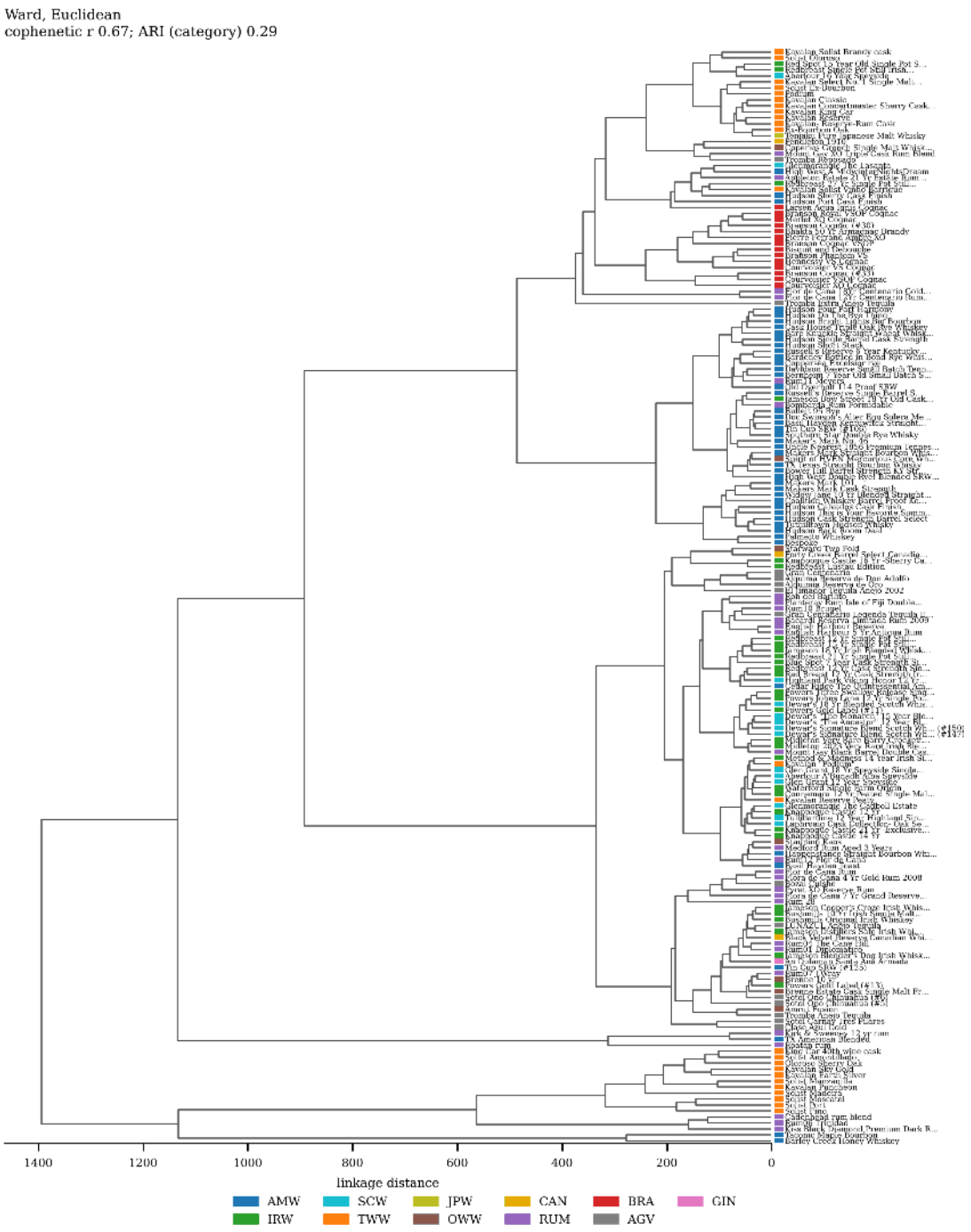

Figure S13. Block D2O-NC (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

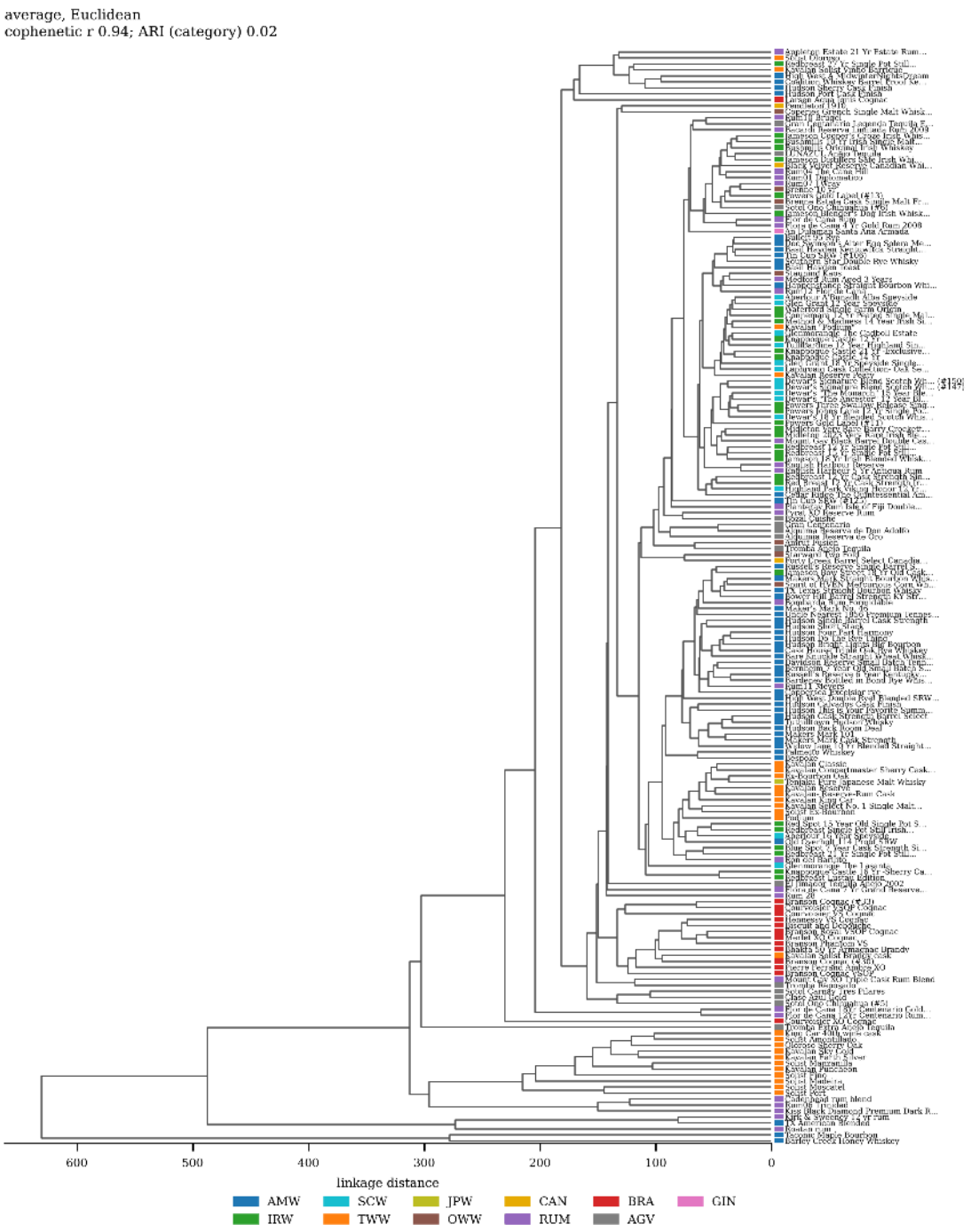

Figure S14. Block D2O-NC (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

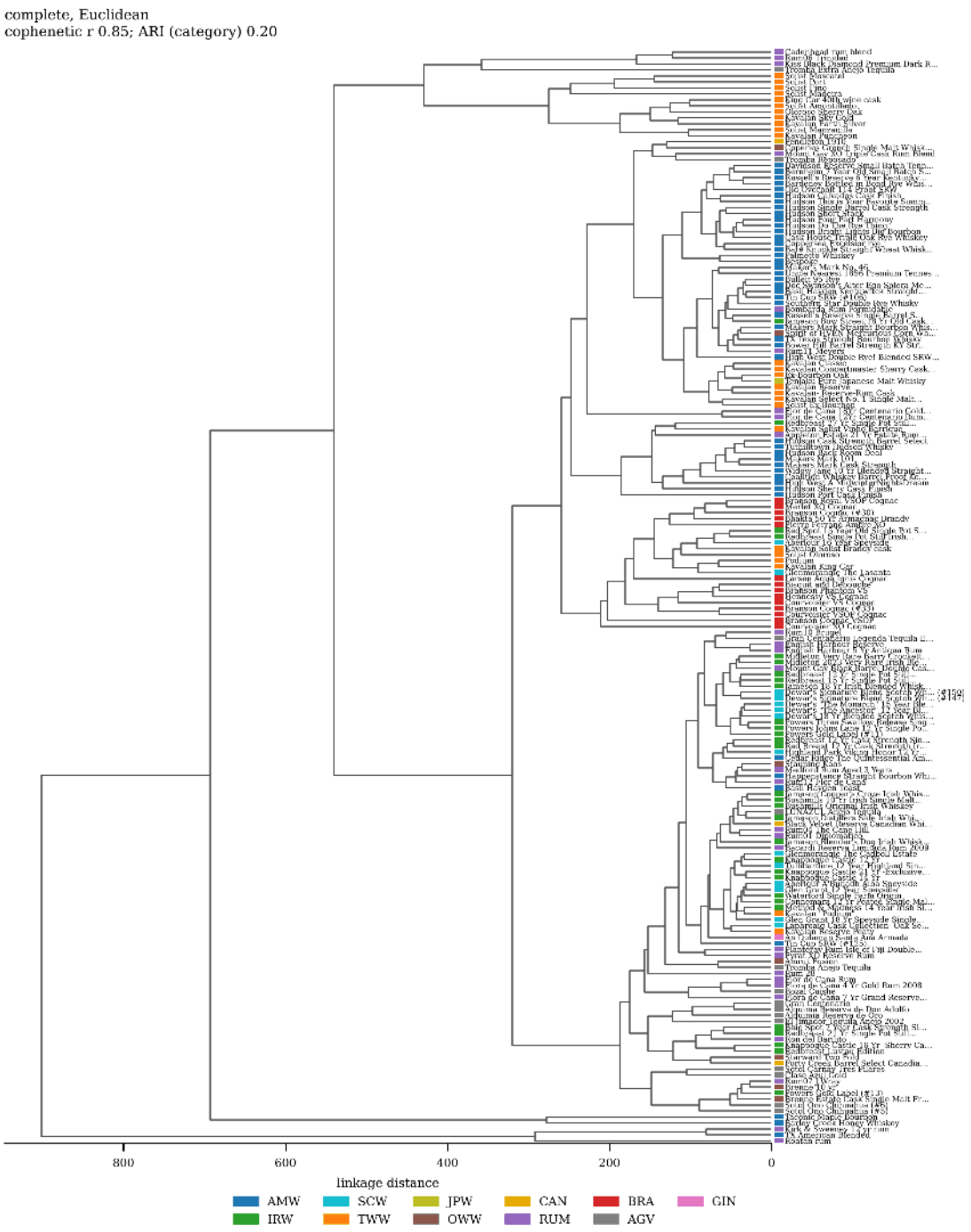

Figure S15. Block D2O-NC (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

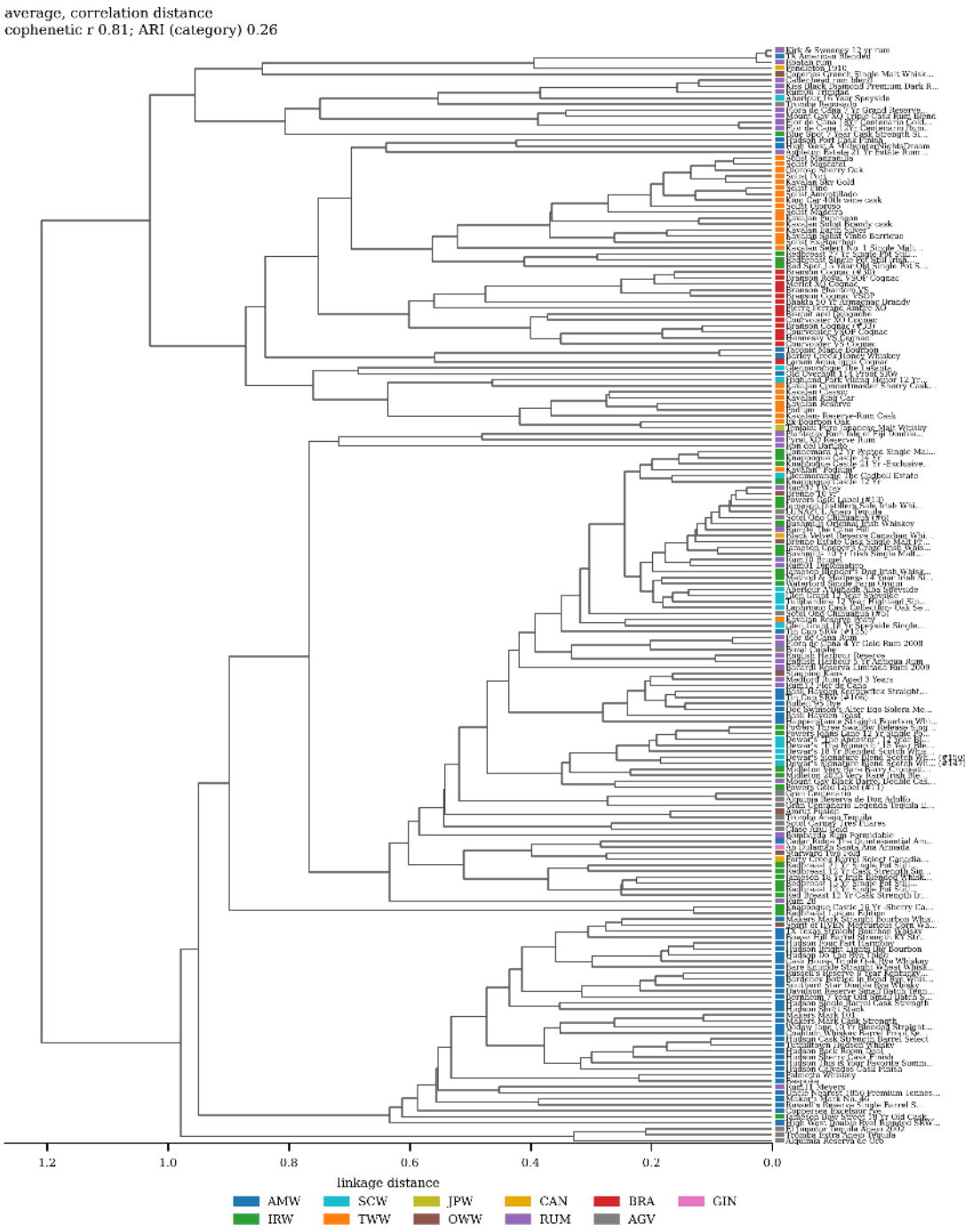

Figure S16. Block D2O-NC (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-NV: Figures S17–S24, enlarged from Figure 41

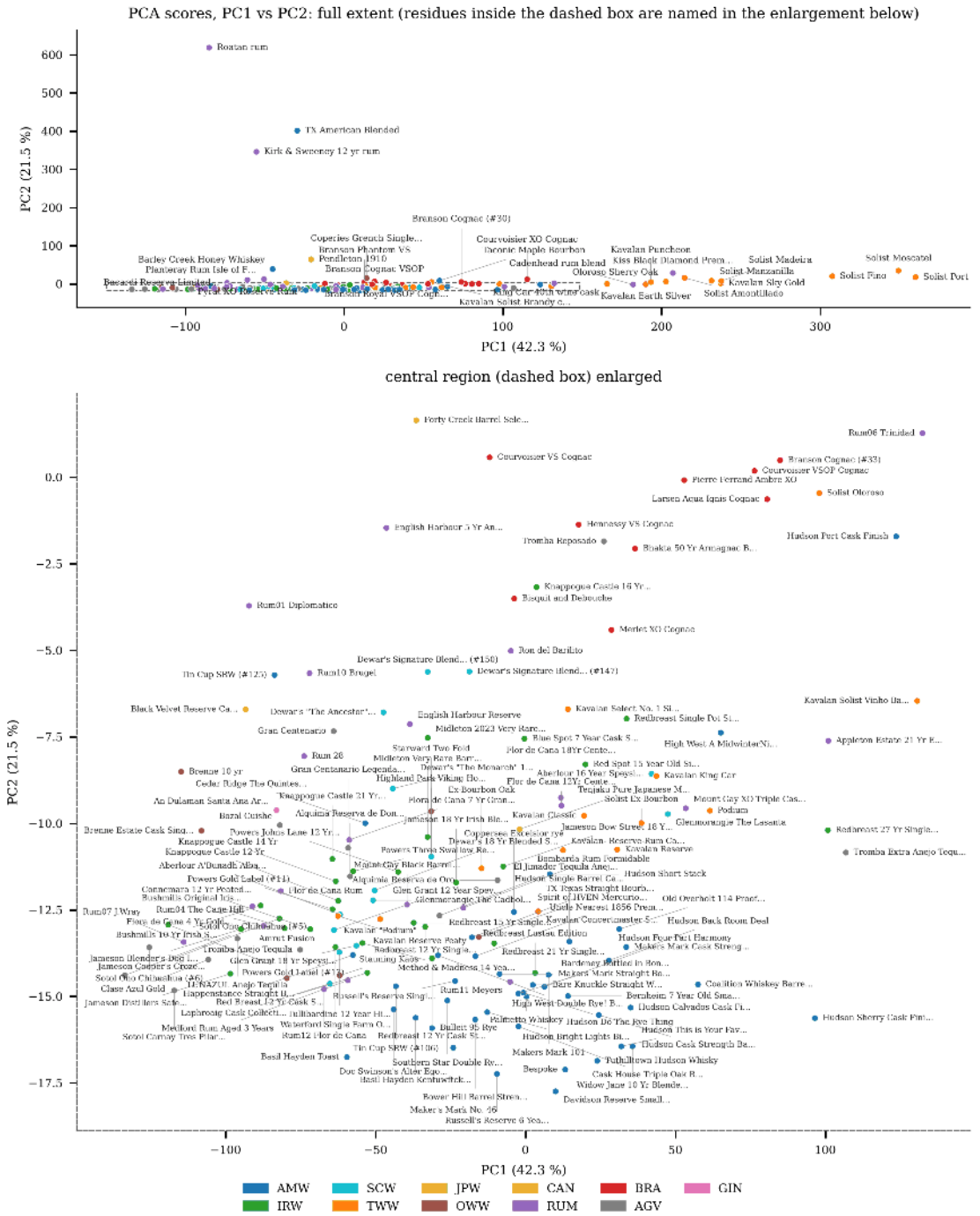

Figure S17. Block D2O-NV (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

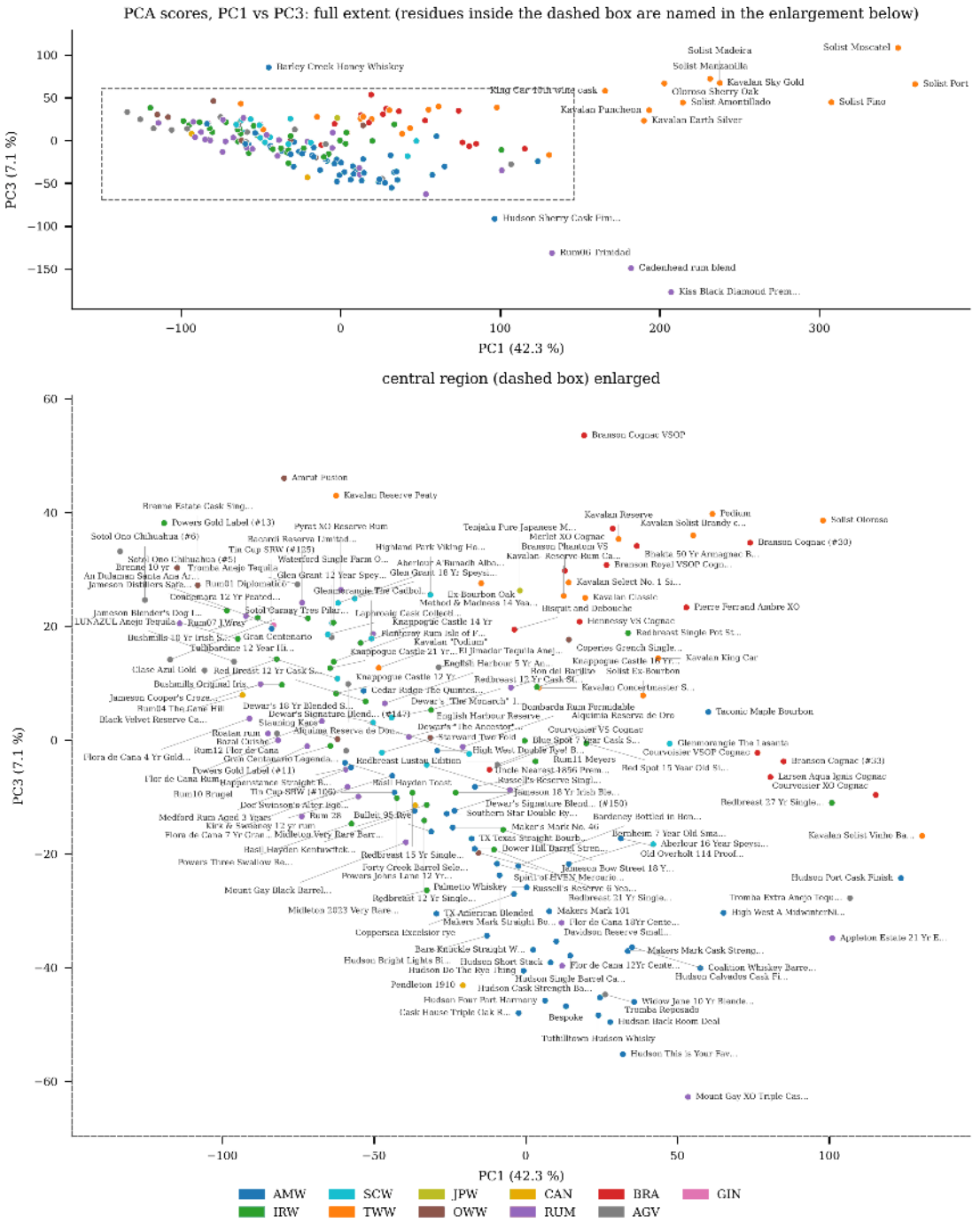

Figure S18. Block D2O-NV (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

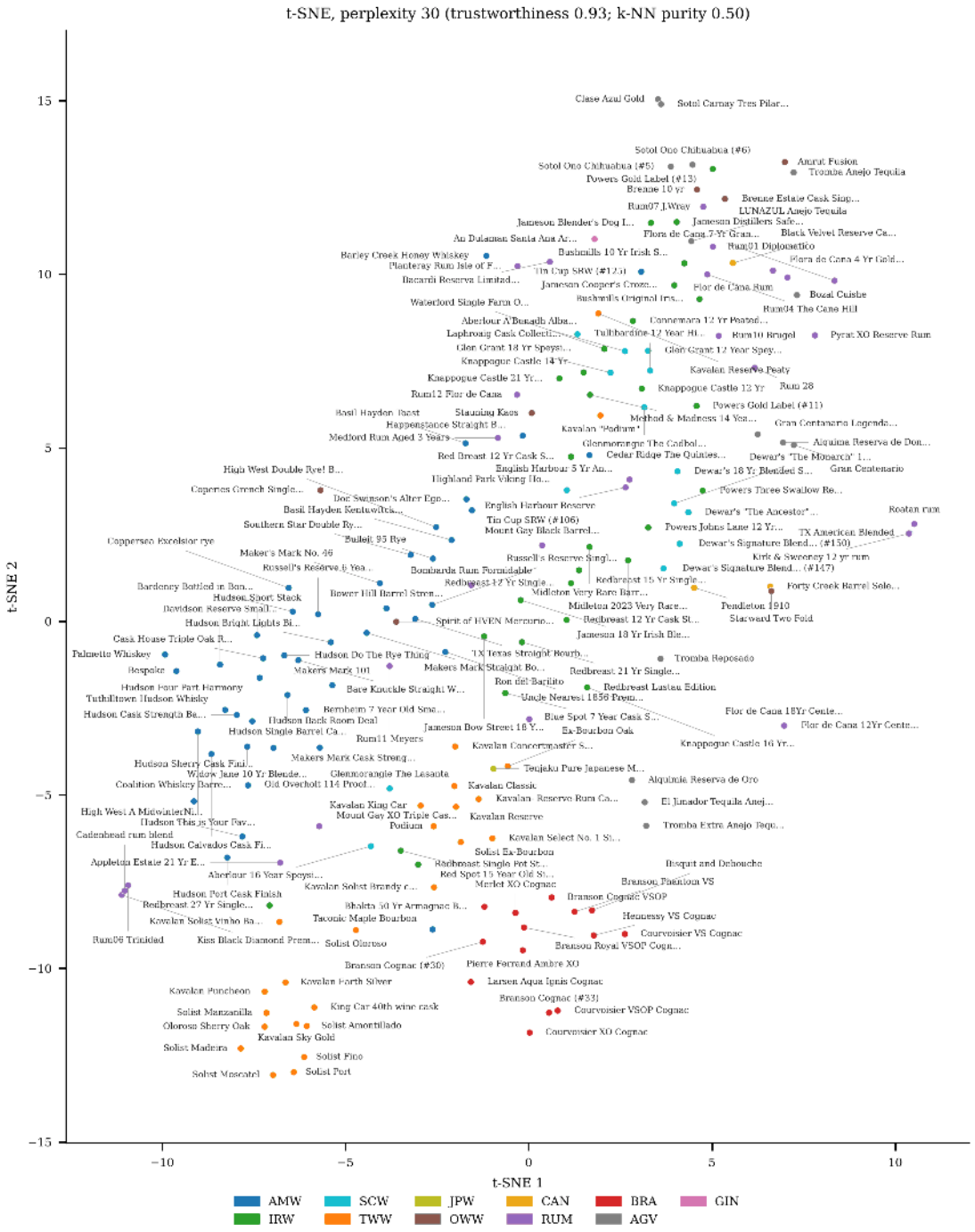

Figure S19. Block D2O-NV (Table 20): t-SNE, perplexity 30 (trustworthiness 0.93; k-NN purity 0.50), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

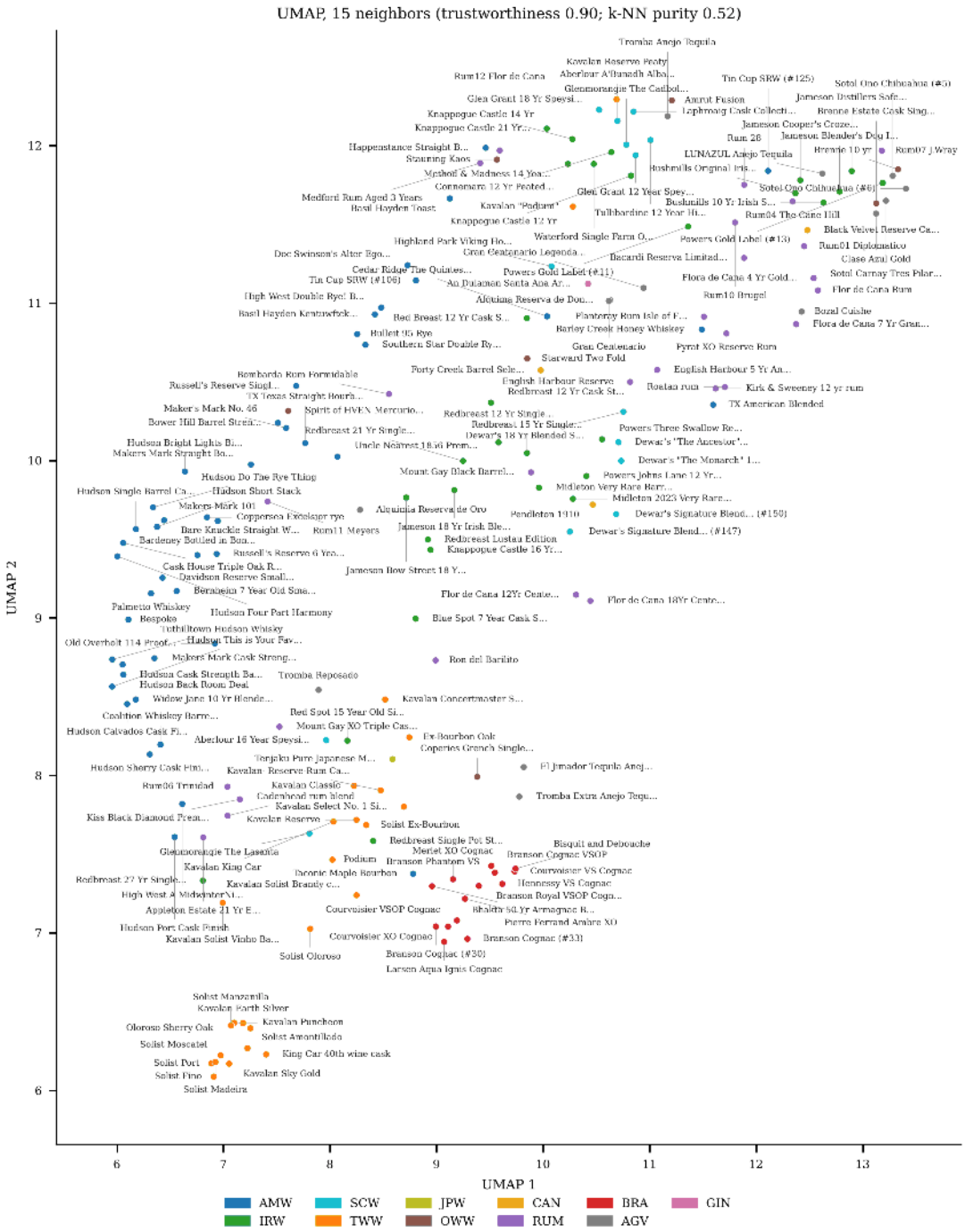

Figure S20. Block D2O-NV (Table 20): UMAP, 15 neighbors (trustworthiness 0.90; k-NN purity 0.52), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

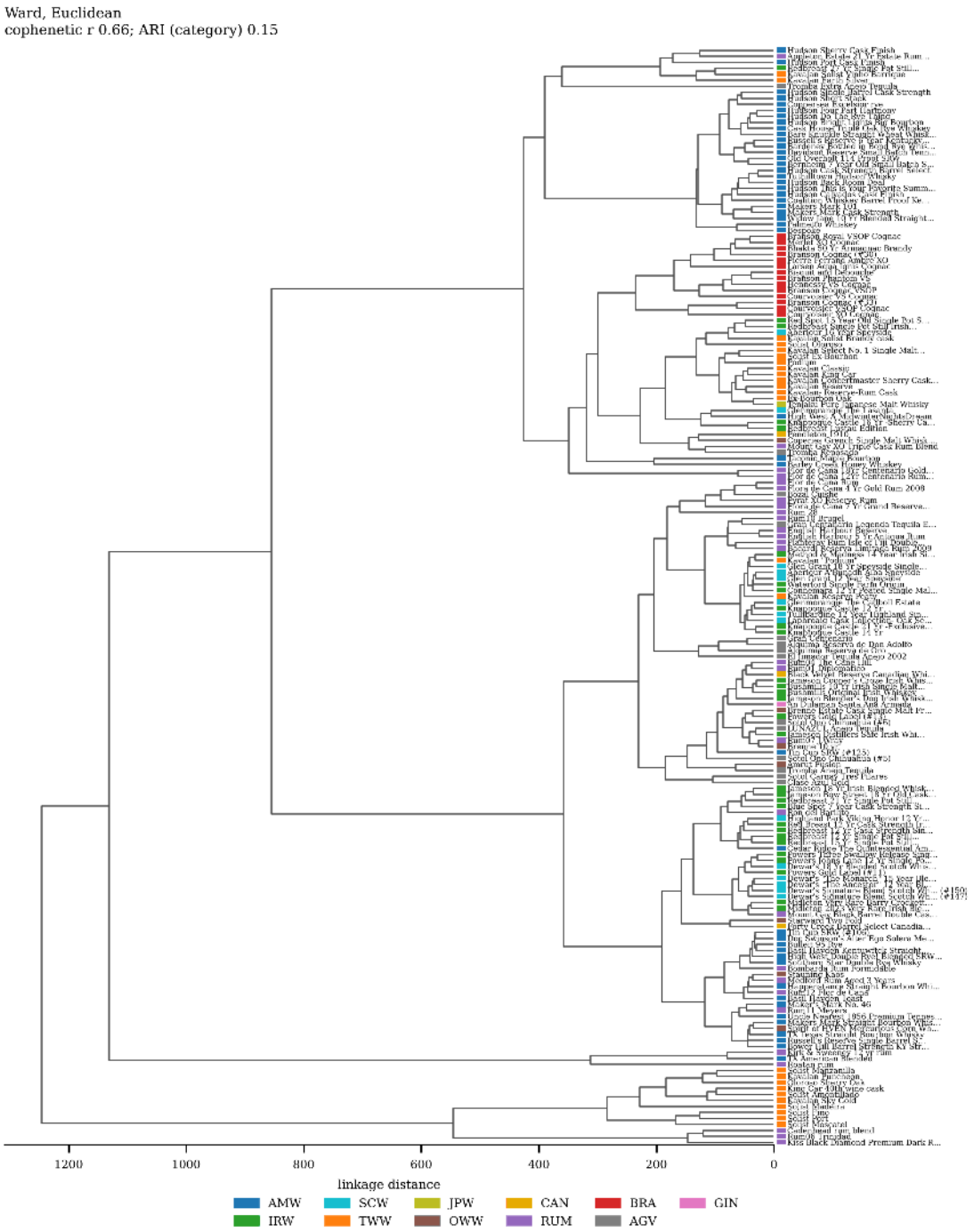

Figure S21. Block D2O-NV (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

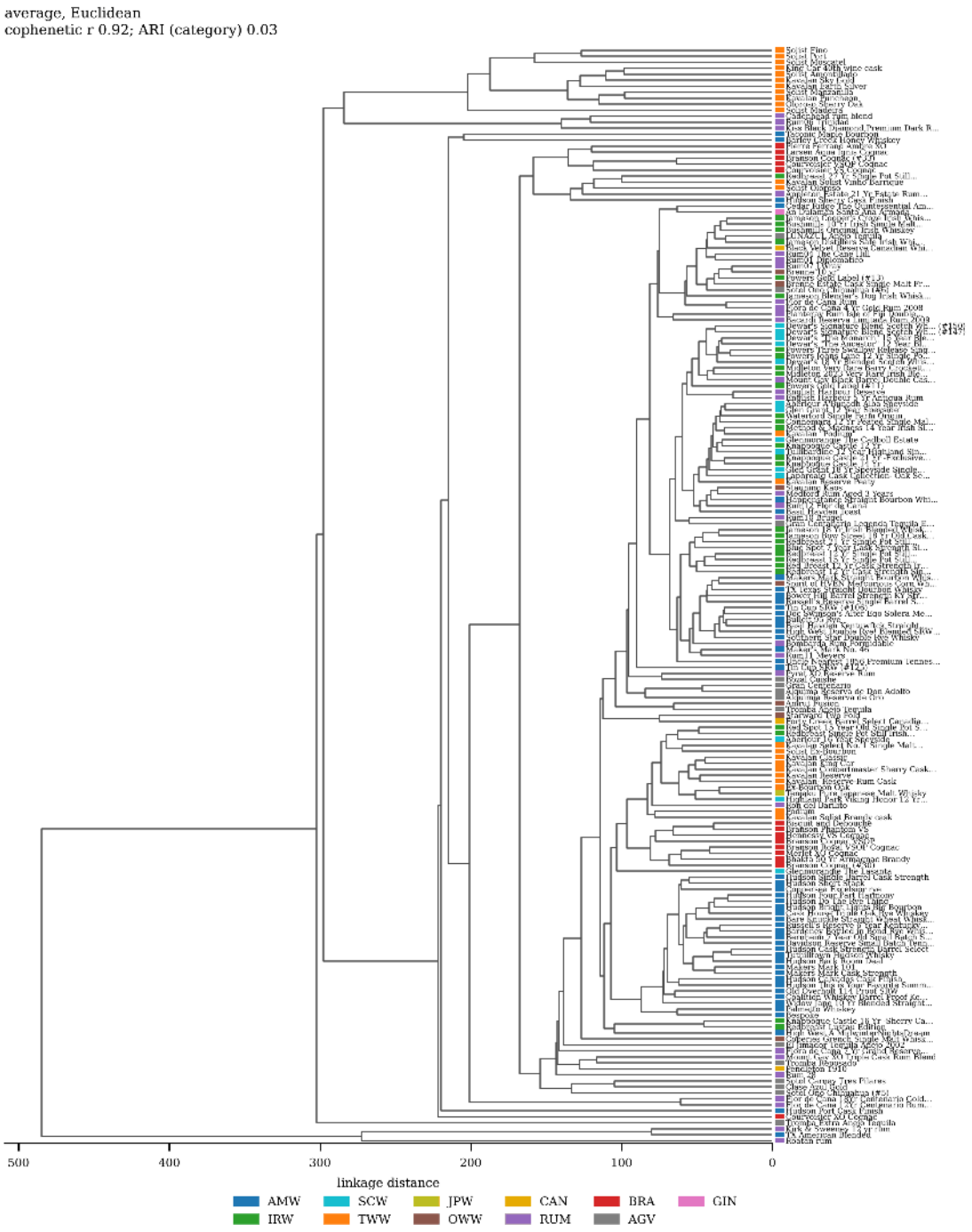

Figure S22. Block D2O-NV (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

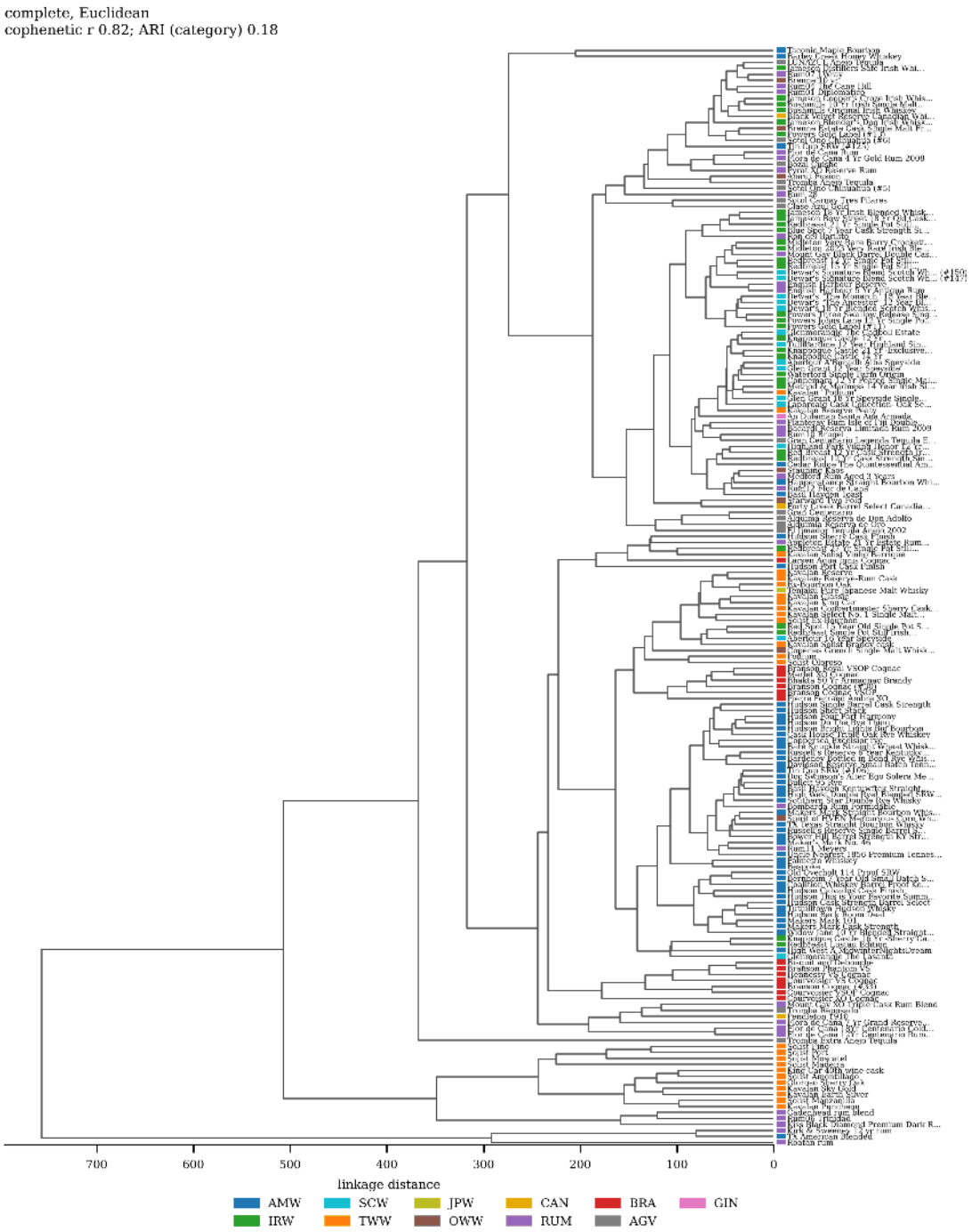

Figure S23. Block D2O-NV (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

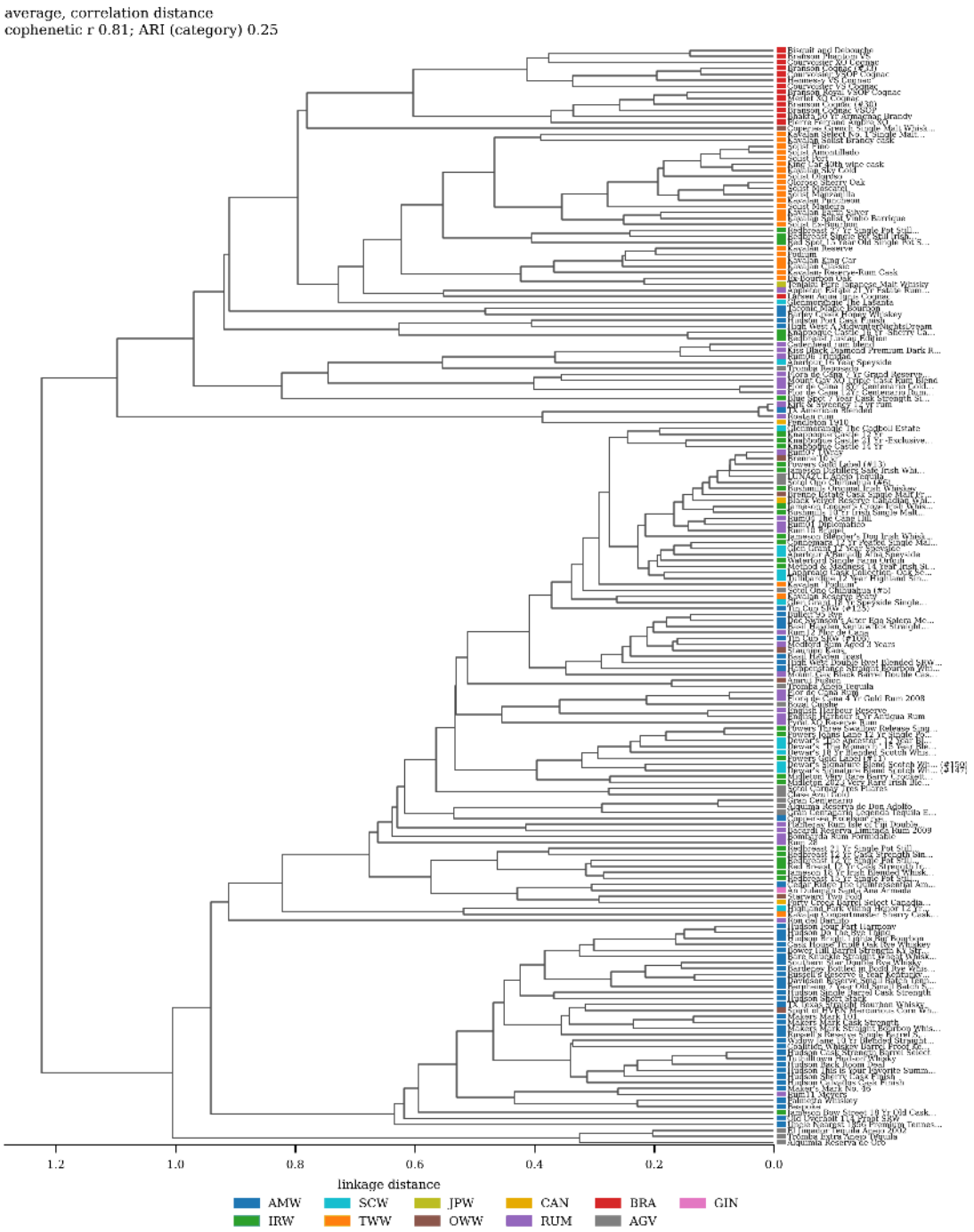

Figure S24. Block D2O-NV (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-NC04: Figures S25–S32, enlarged from Figure 42

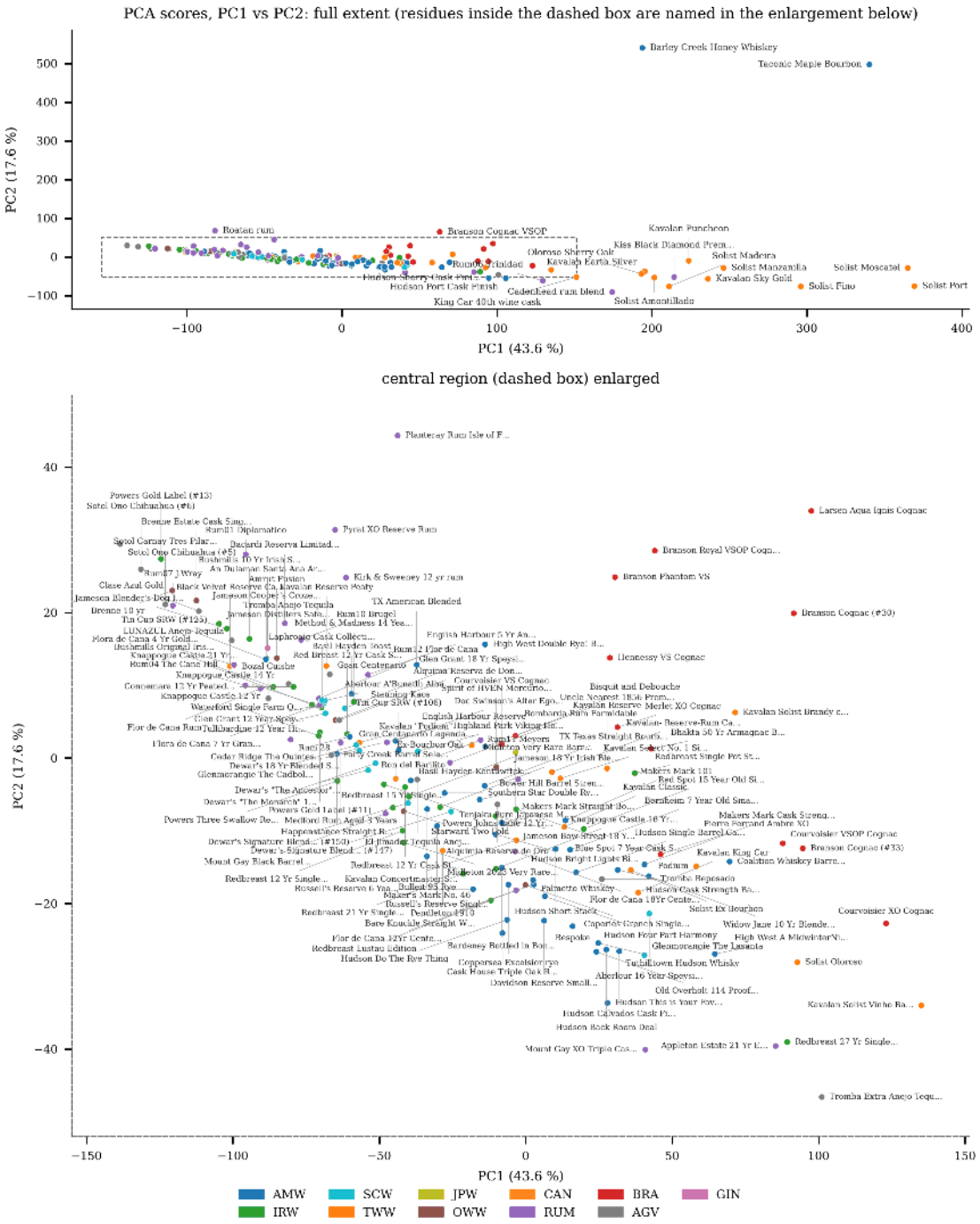

Figure S25. Block D2O-NC04 (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

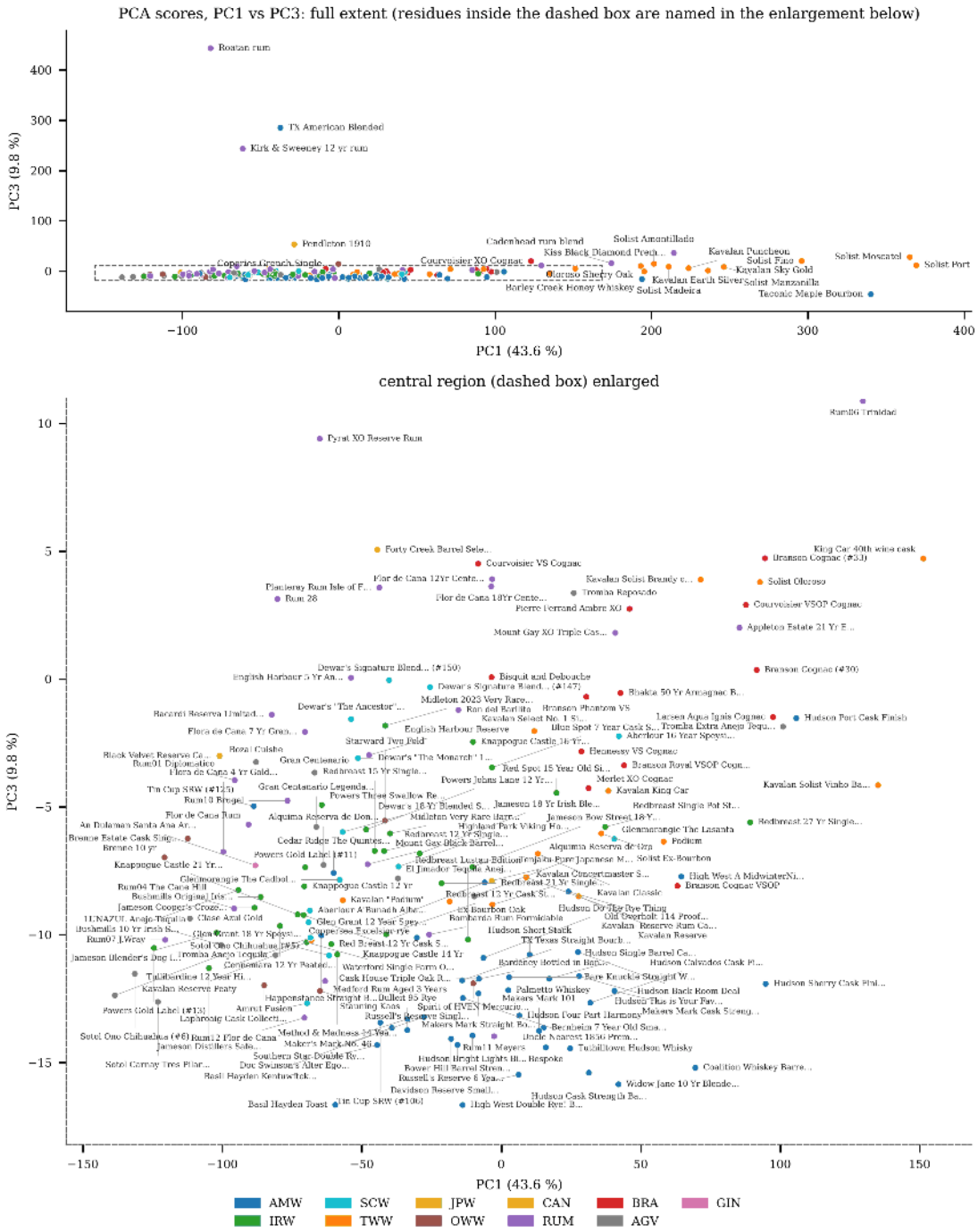

Figure S26. Block D2O-NC04 (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

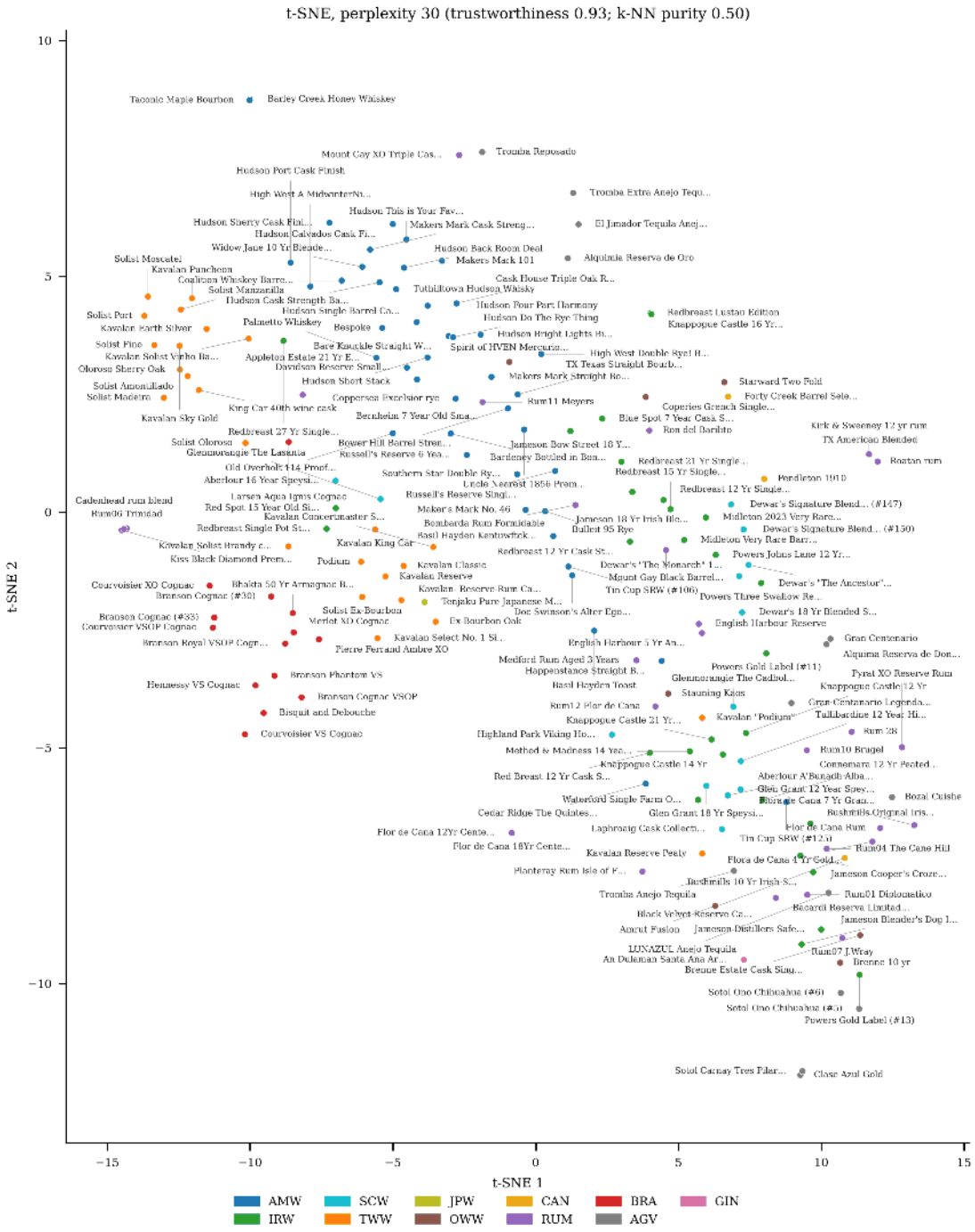

Figure S27. Block D2O-NC04 (Table 20): t-SNE, perplexity 30 (trustworthiness 0.93; k-NN purity 0.50), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

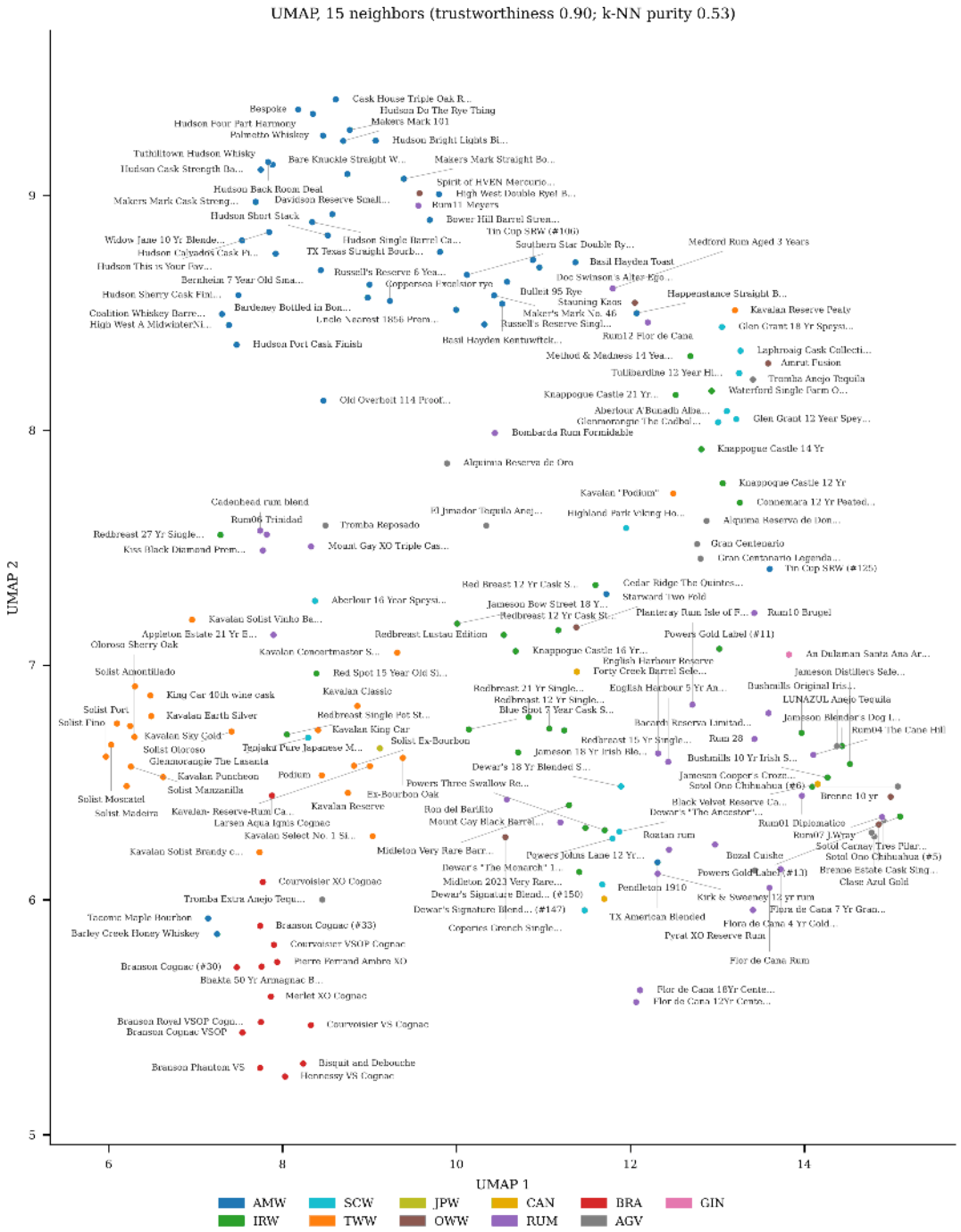

Figure S28. Block D2O-NC04 (Table 20): UMAP, 15 neighbors (trustworthiness 0.90; k-NN purity 0.53), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

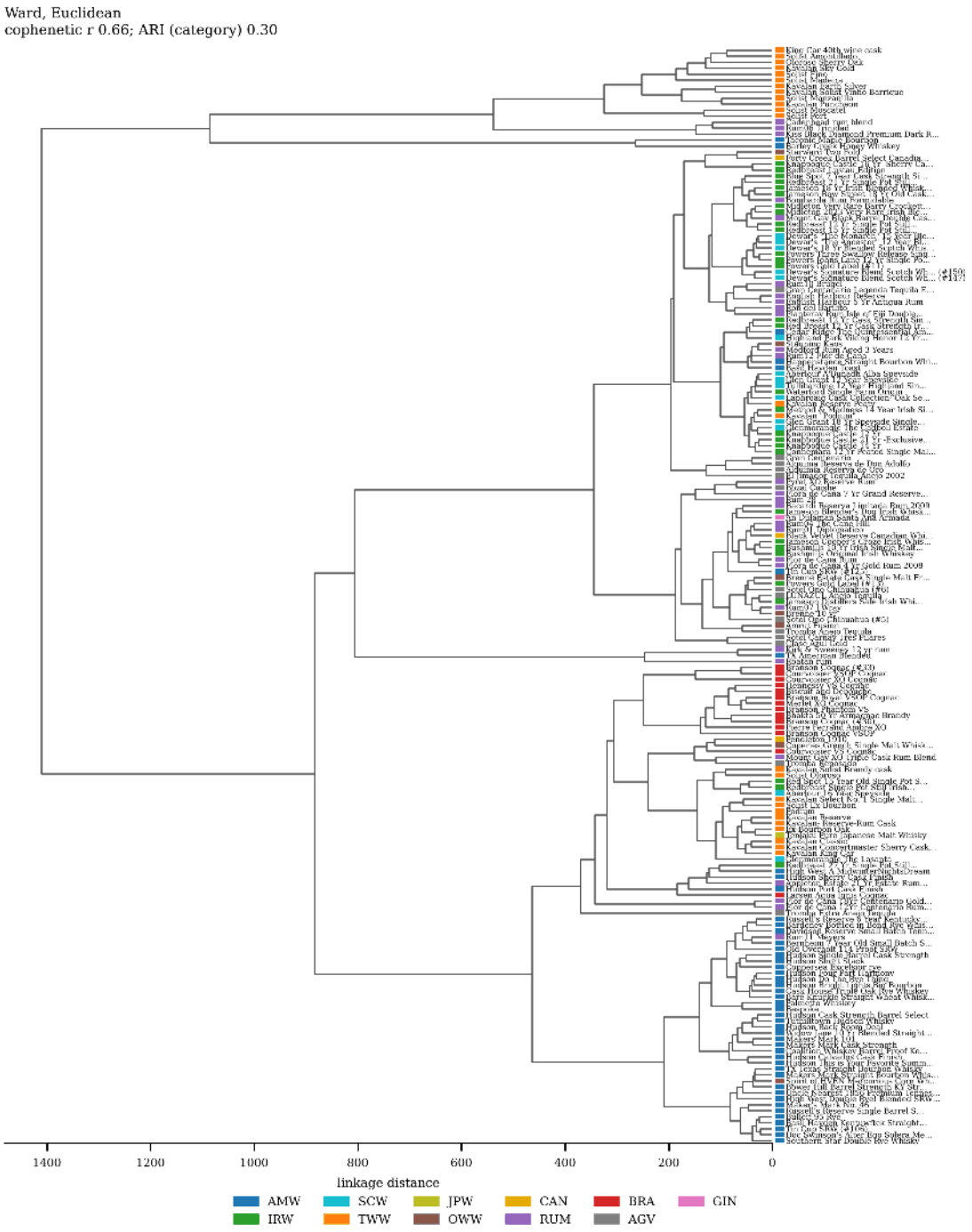

Figure S29. Block D2O-NC04 (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

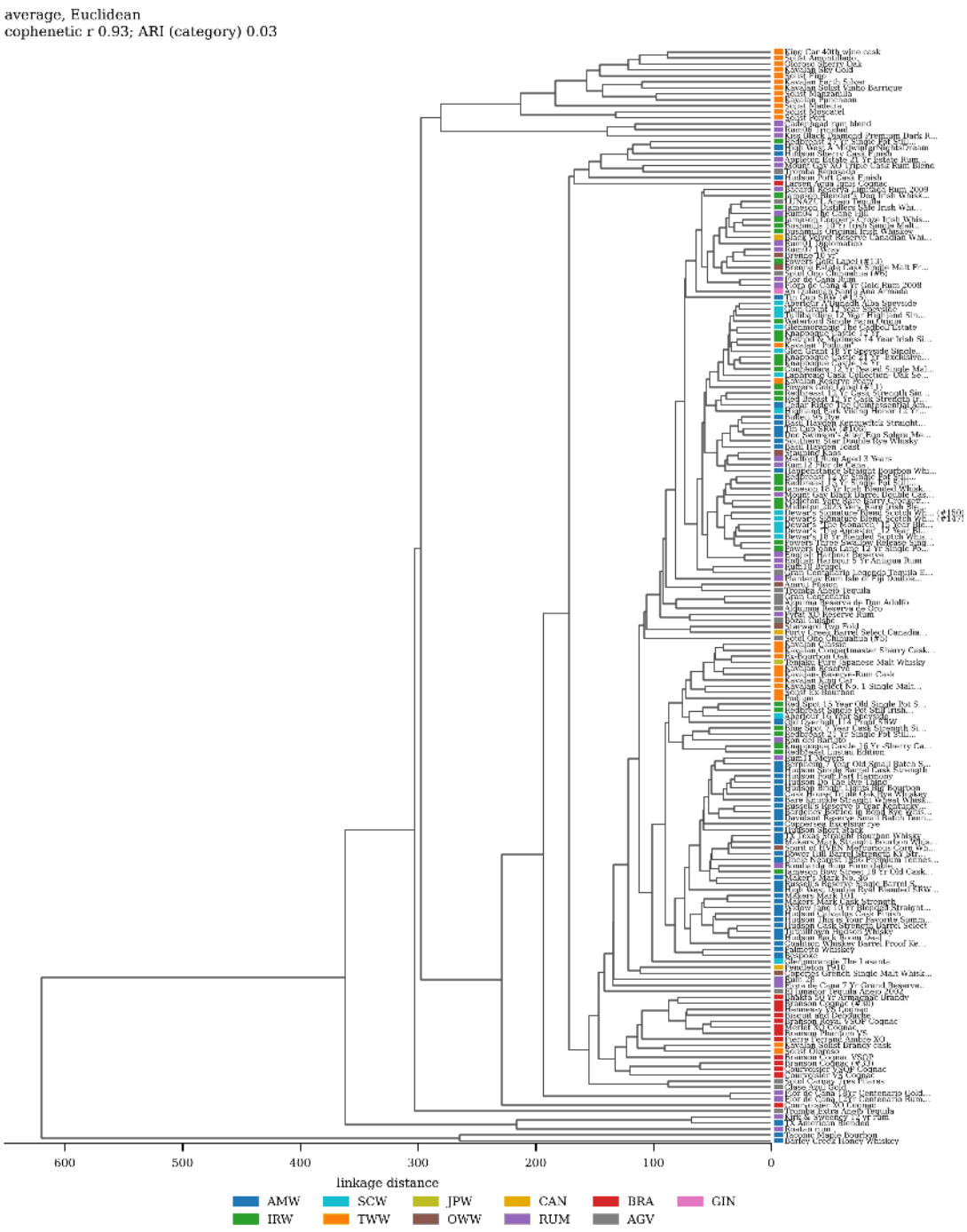

Figure S30. Block D2O-NC04 (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

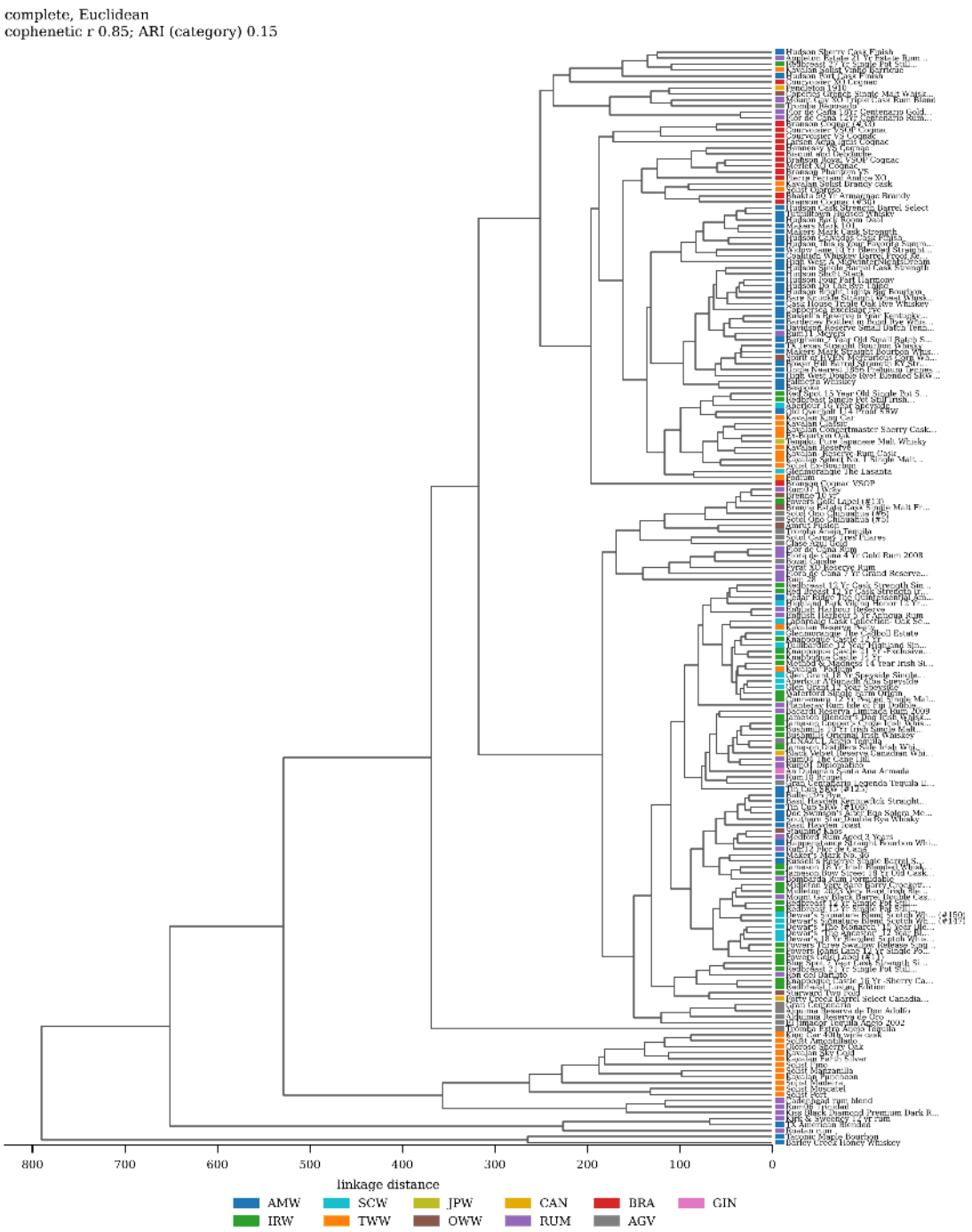

Figure S31. Block D2O-NC04 (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

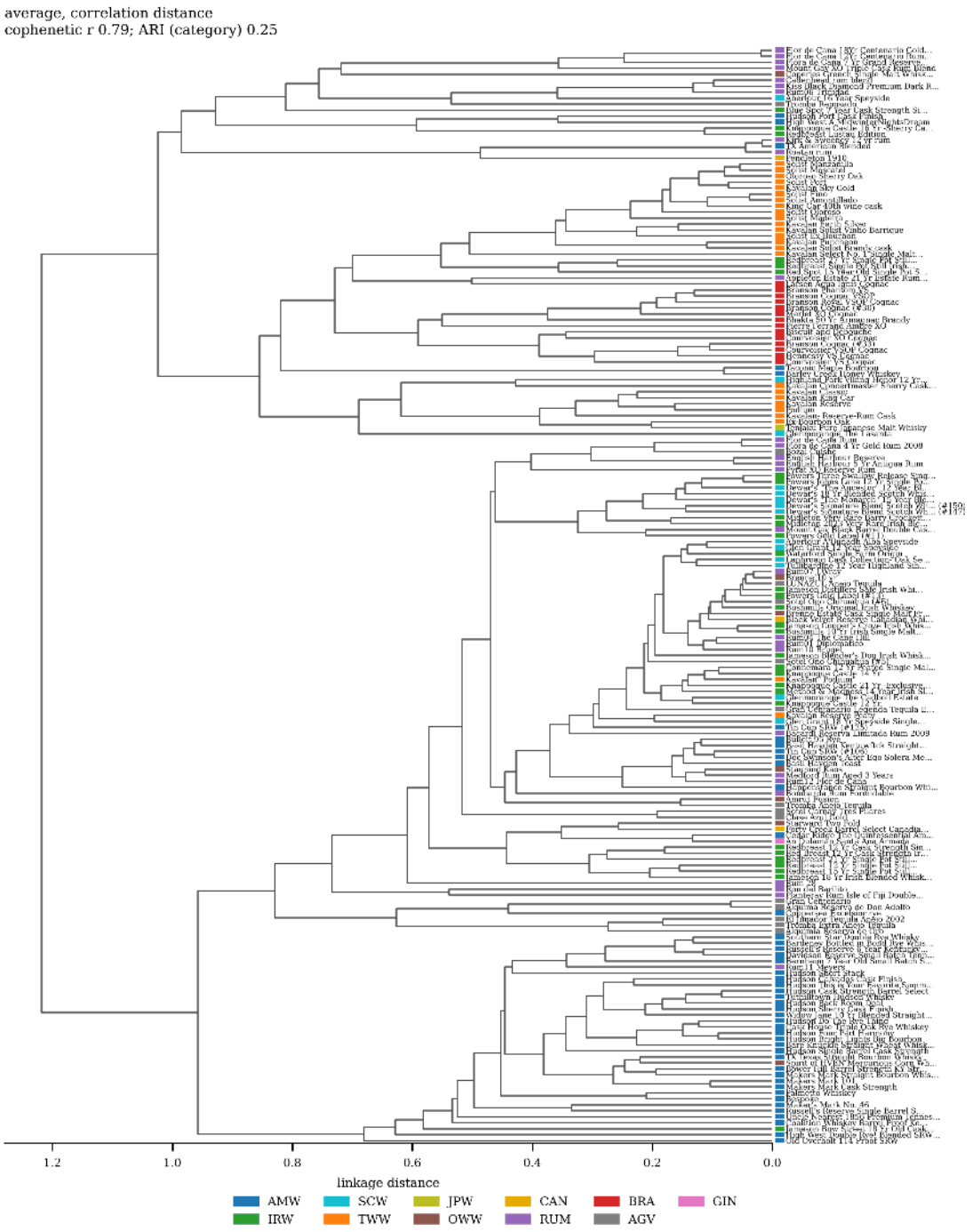

Figure S32. Block D2O-NC04 (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-NCall: Figures S33–S40, enlarged from Figure 43

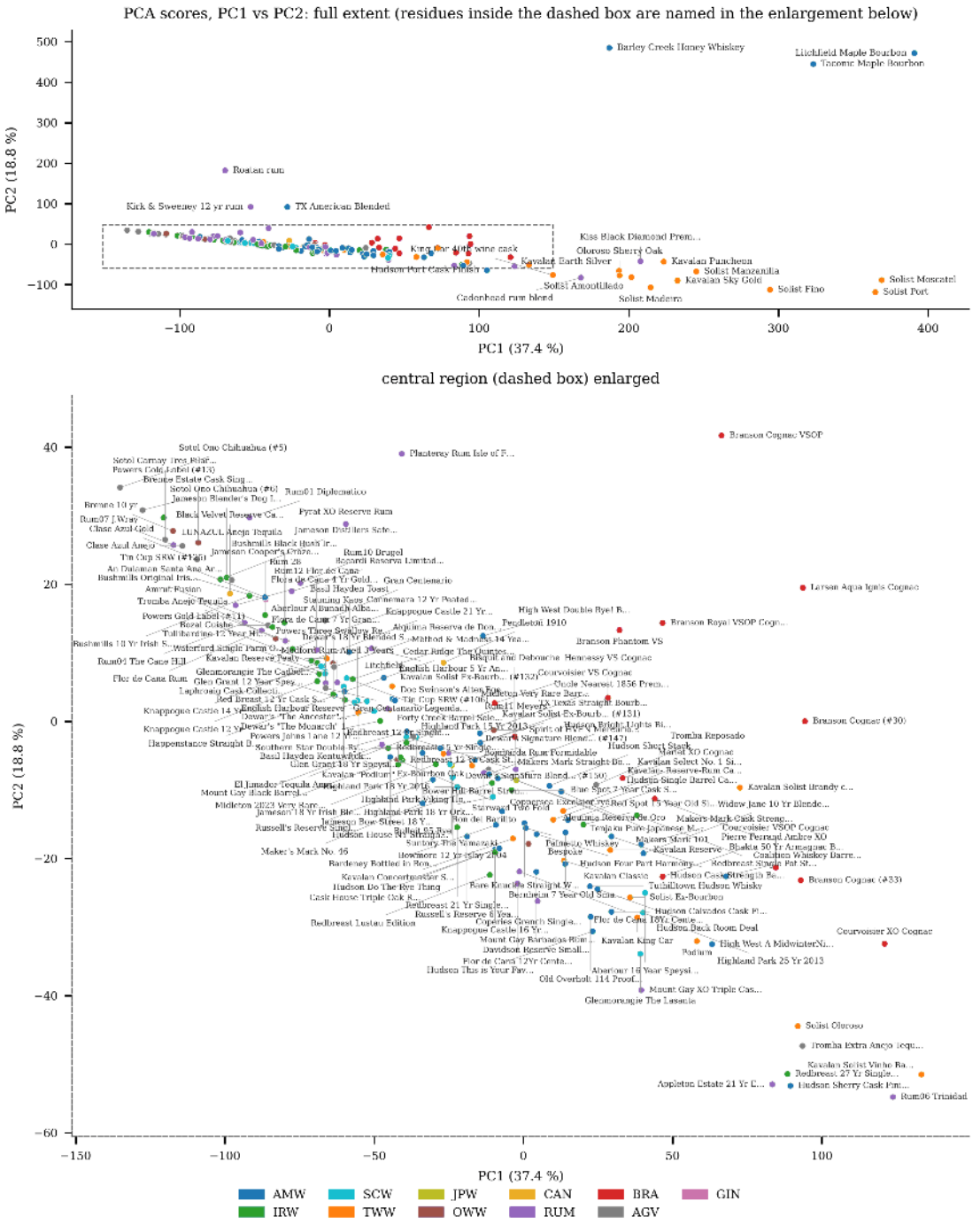

Figure S33. Block D2O-NCall (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

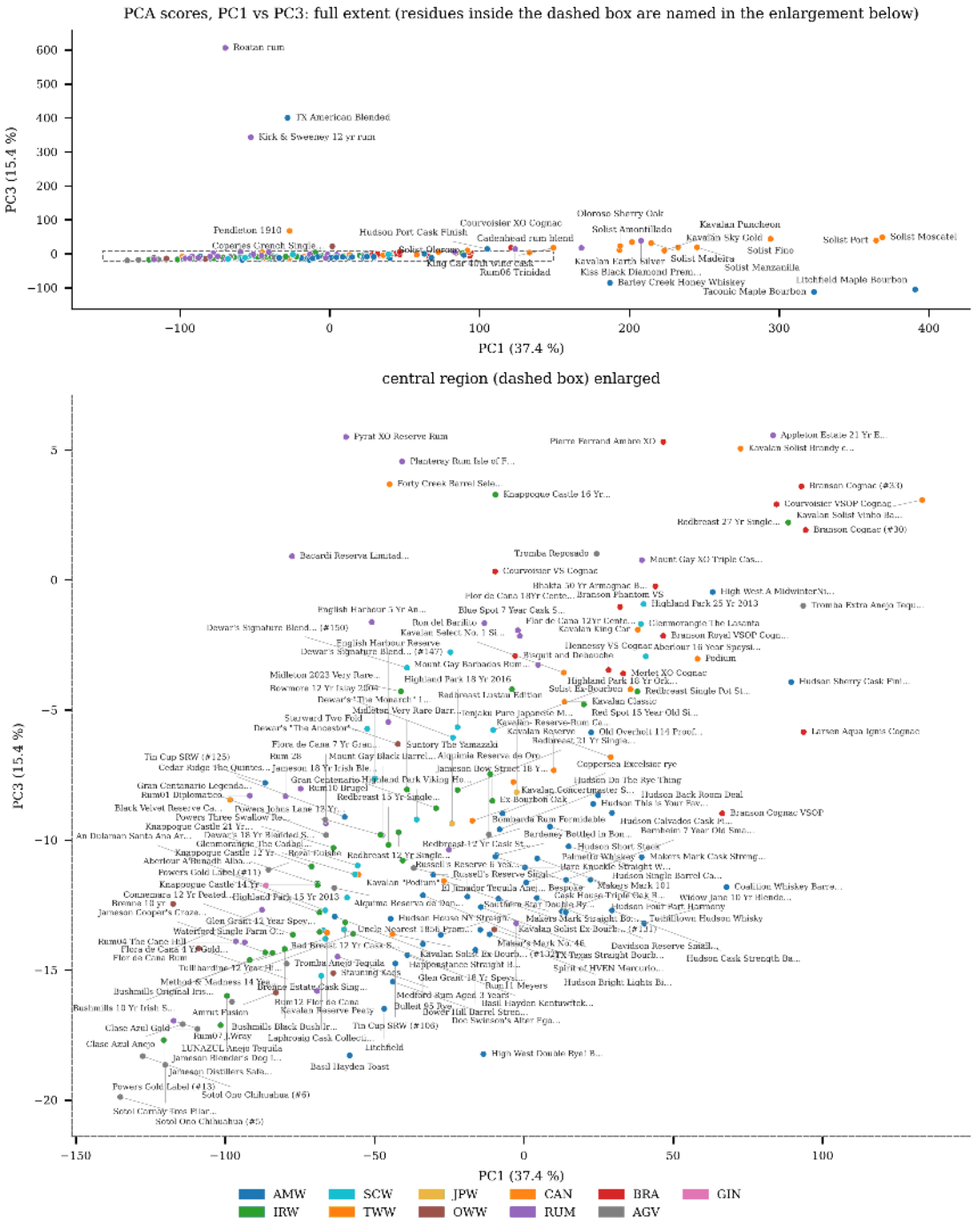

Figure S34. Block D2O-NCall (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

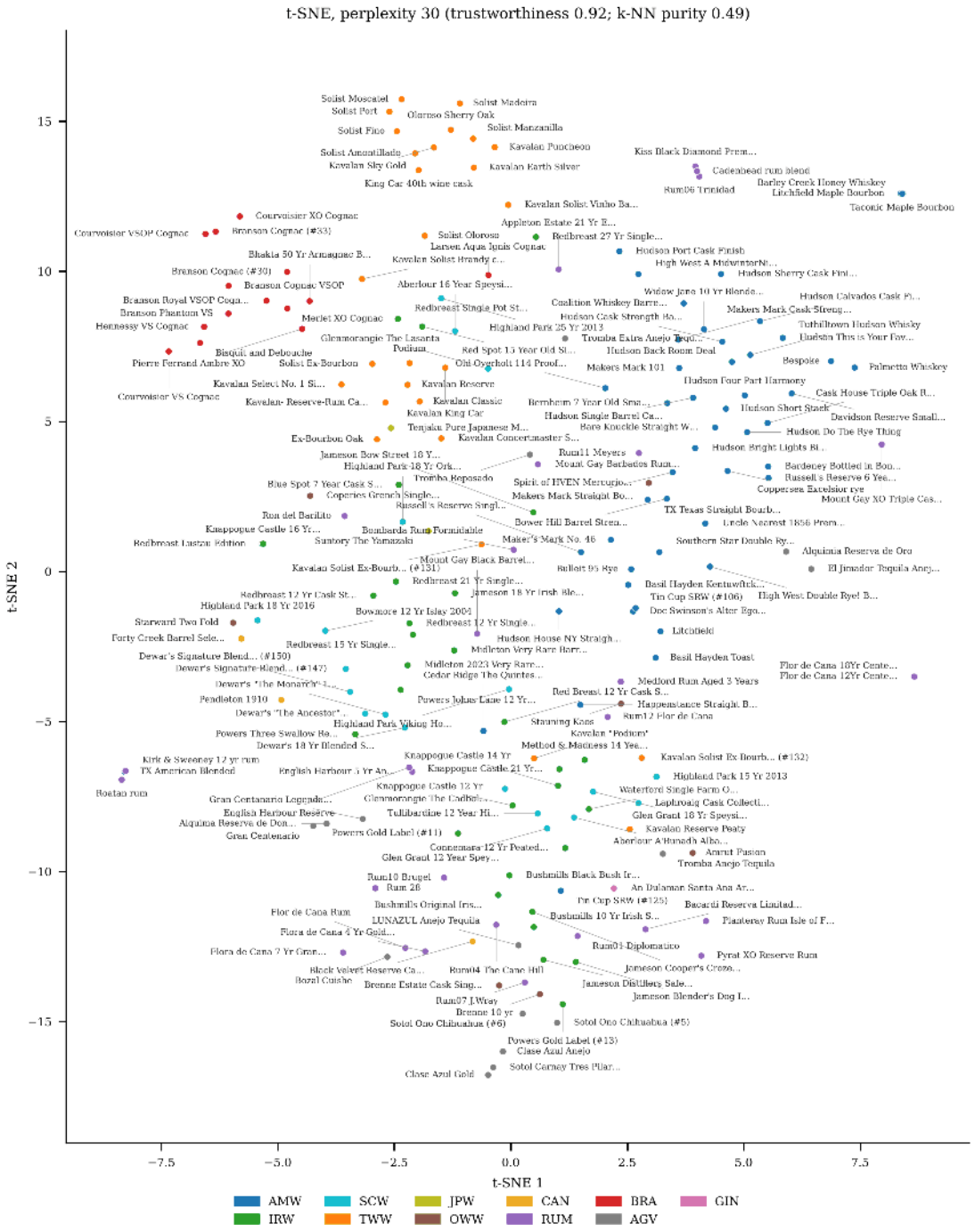

Figure S35. Block D2O-NCall (Table 20): t-SNE, perplexity 30 (trustworthiness 0.92; k-NN purity 0.49), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

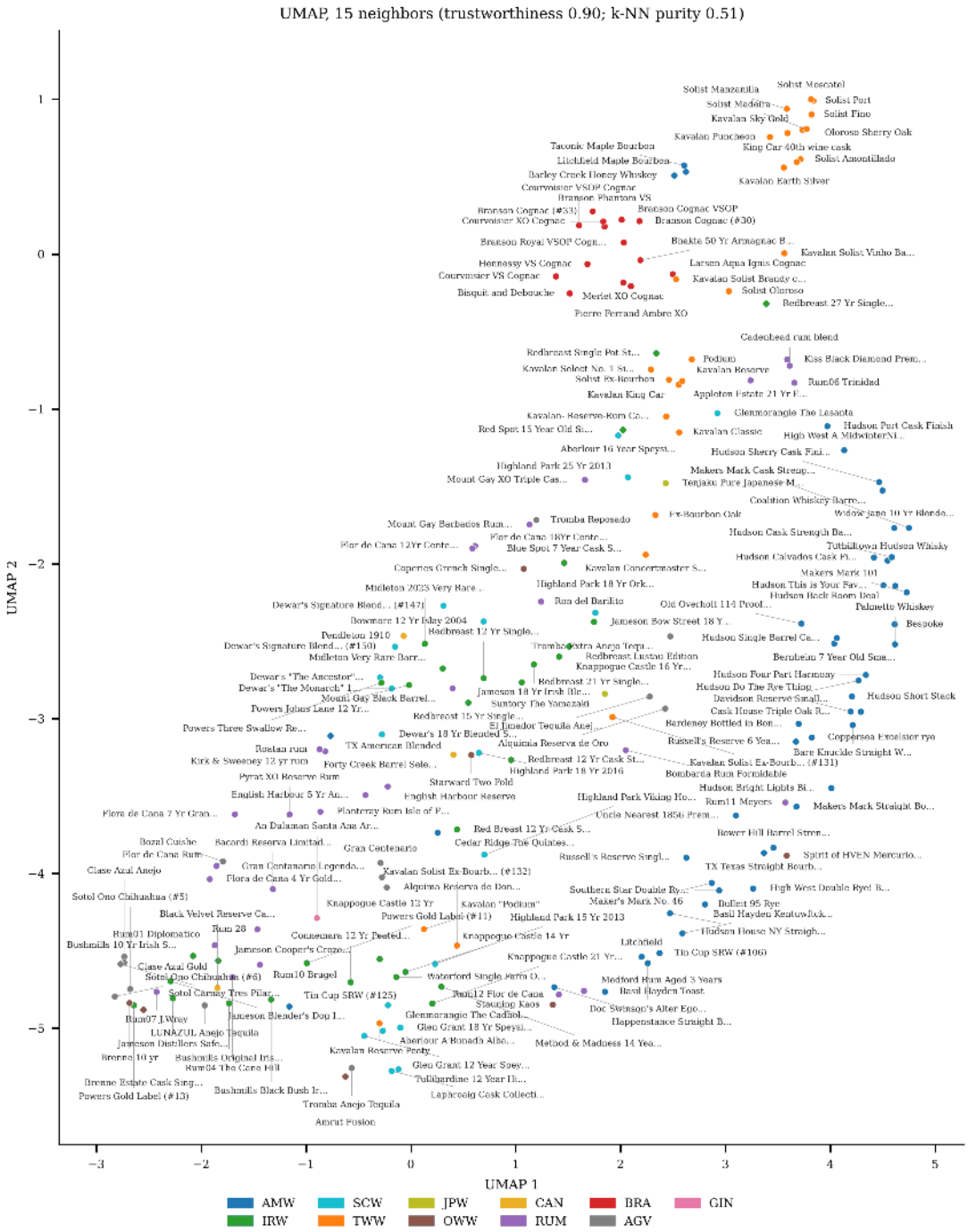

Figure S36. Block D2O-NCall (Table 20): UMAP, 15 neighbors (trustworthiness 0.90; k-NN purity 0.51), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

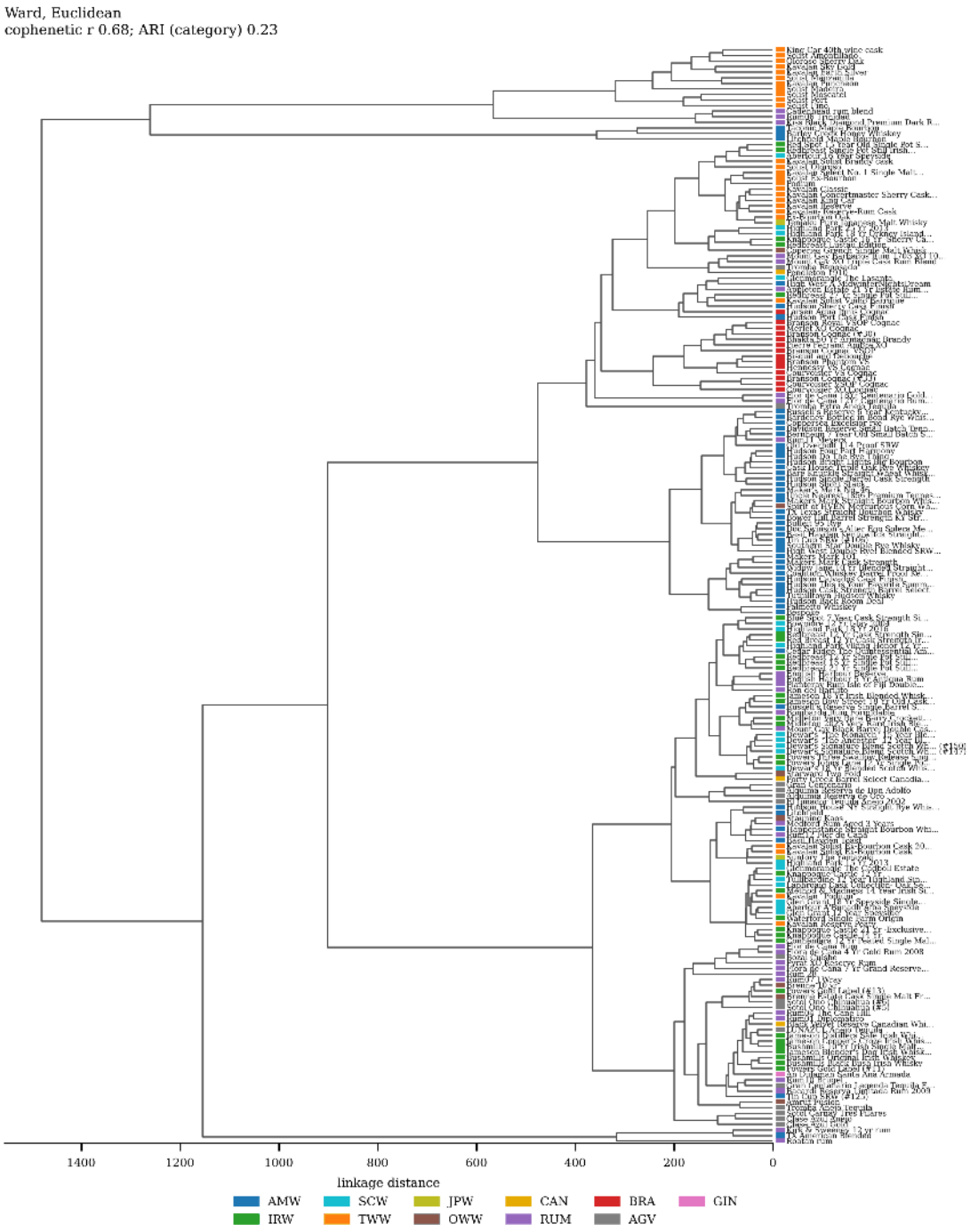

Figure S37. Block D2O-NCall (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

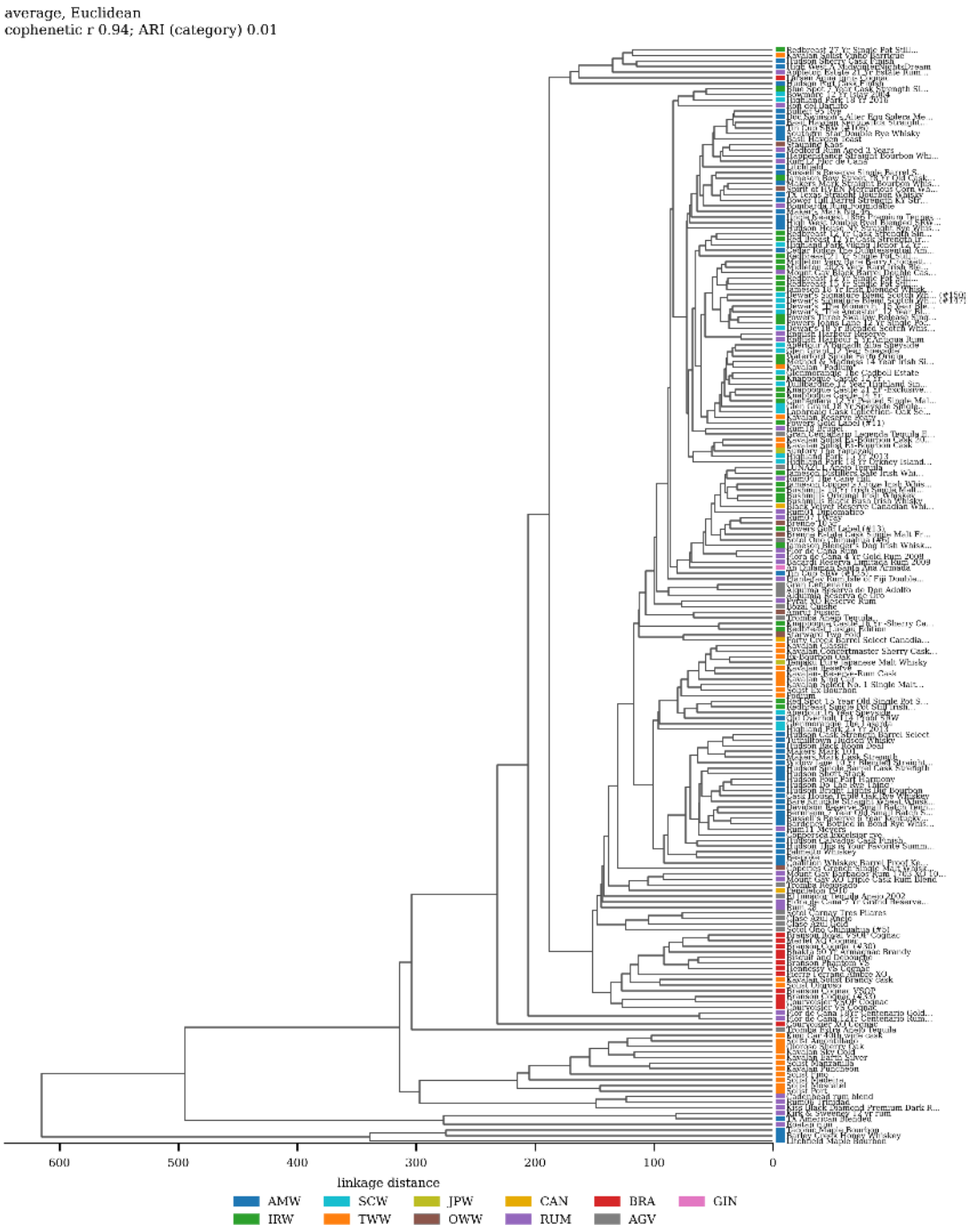

Figure S38. Block D2O-NCall (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

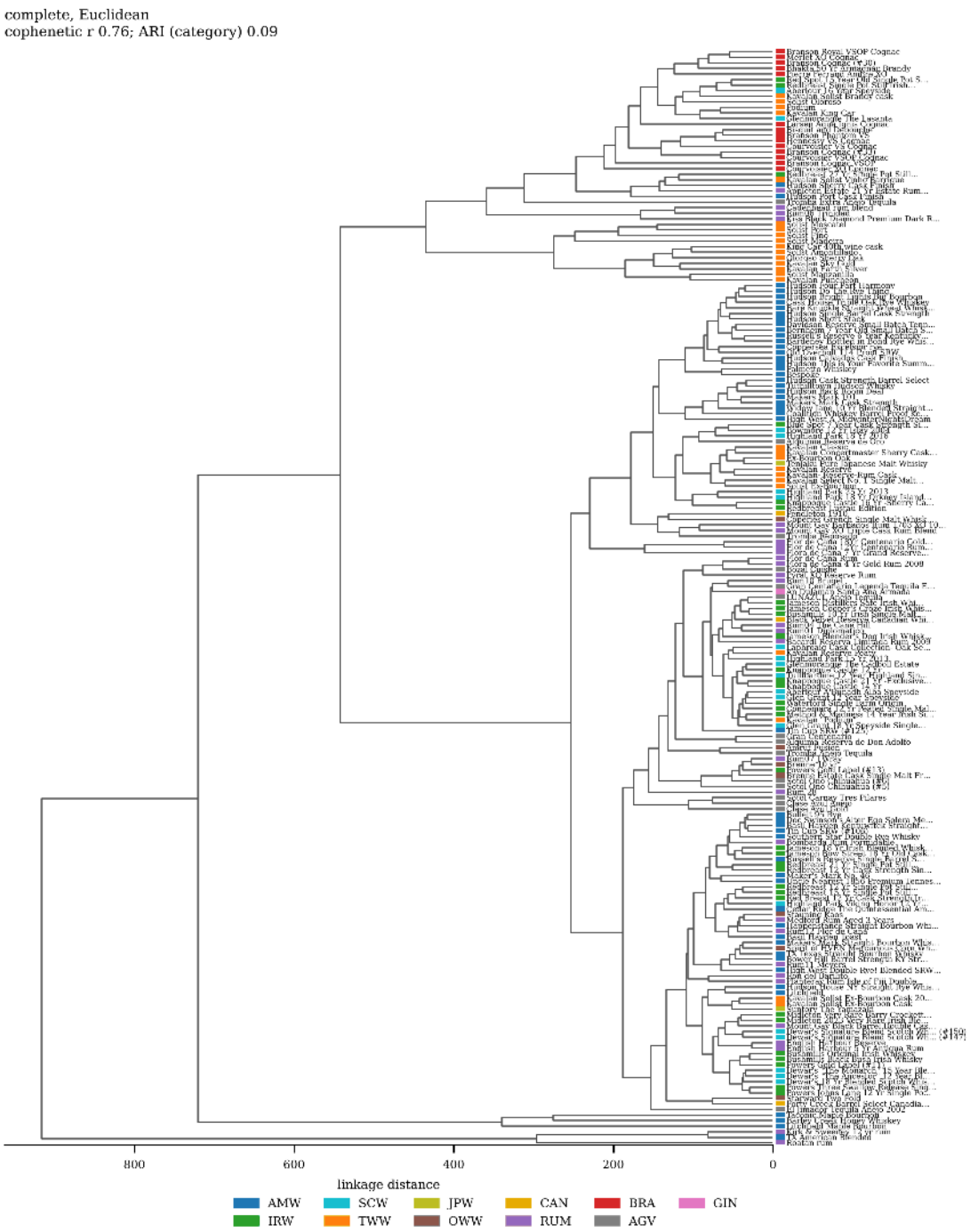

Figure S39. Block D2O-NCall (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

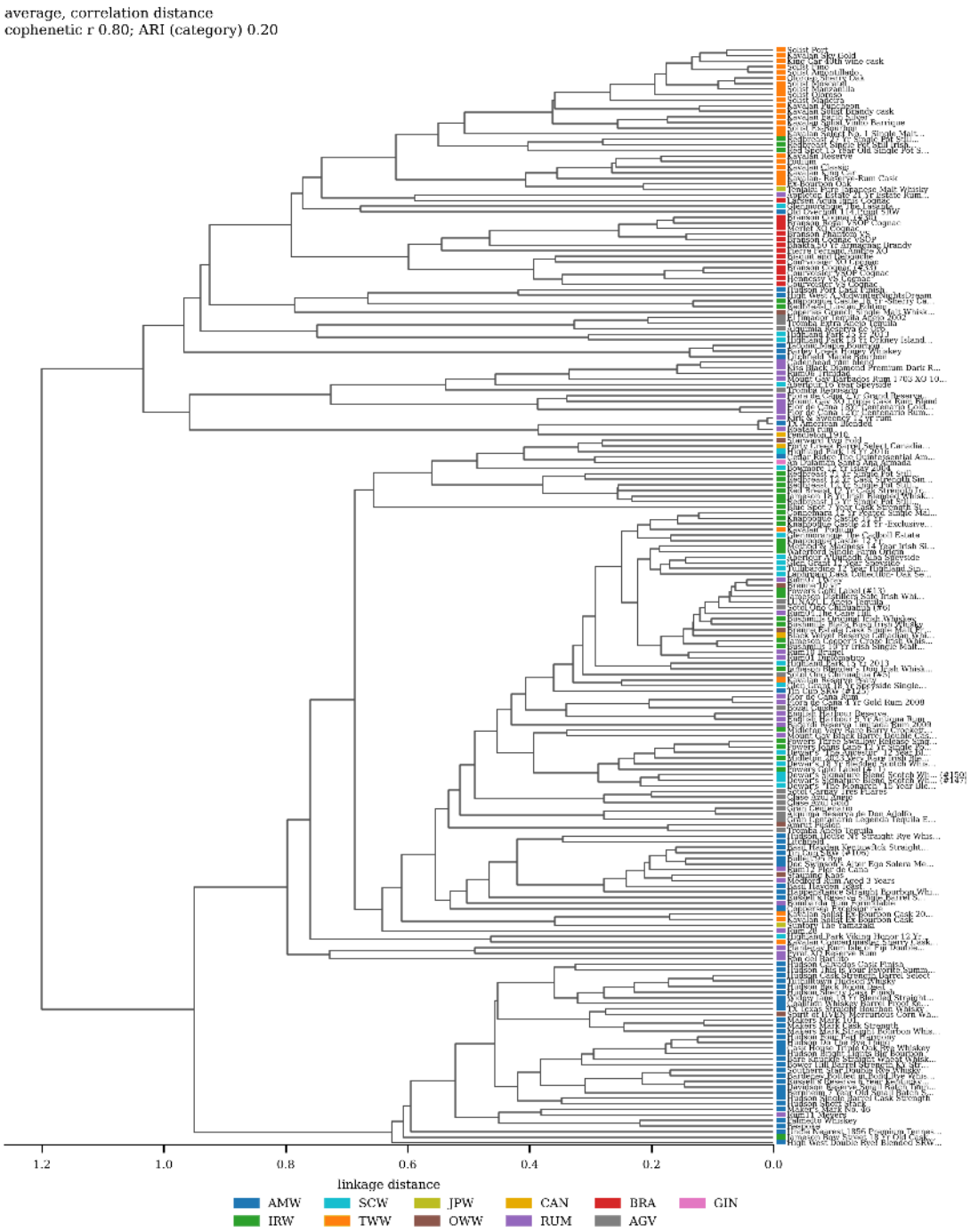

Figure S40. Block D2O-NCall (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-Qp: Figures S41–S48, enlarged from Figure 44

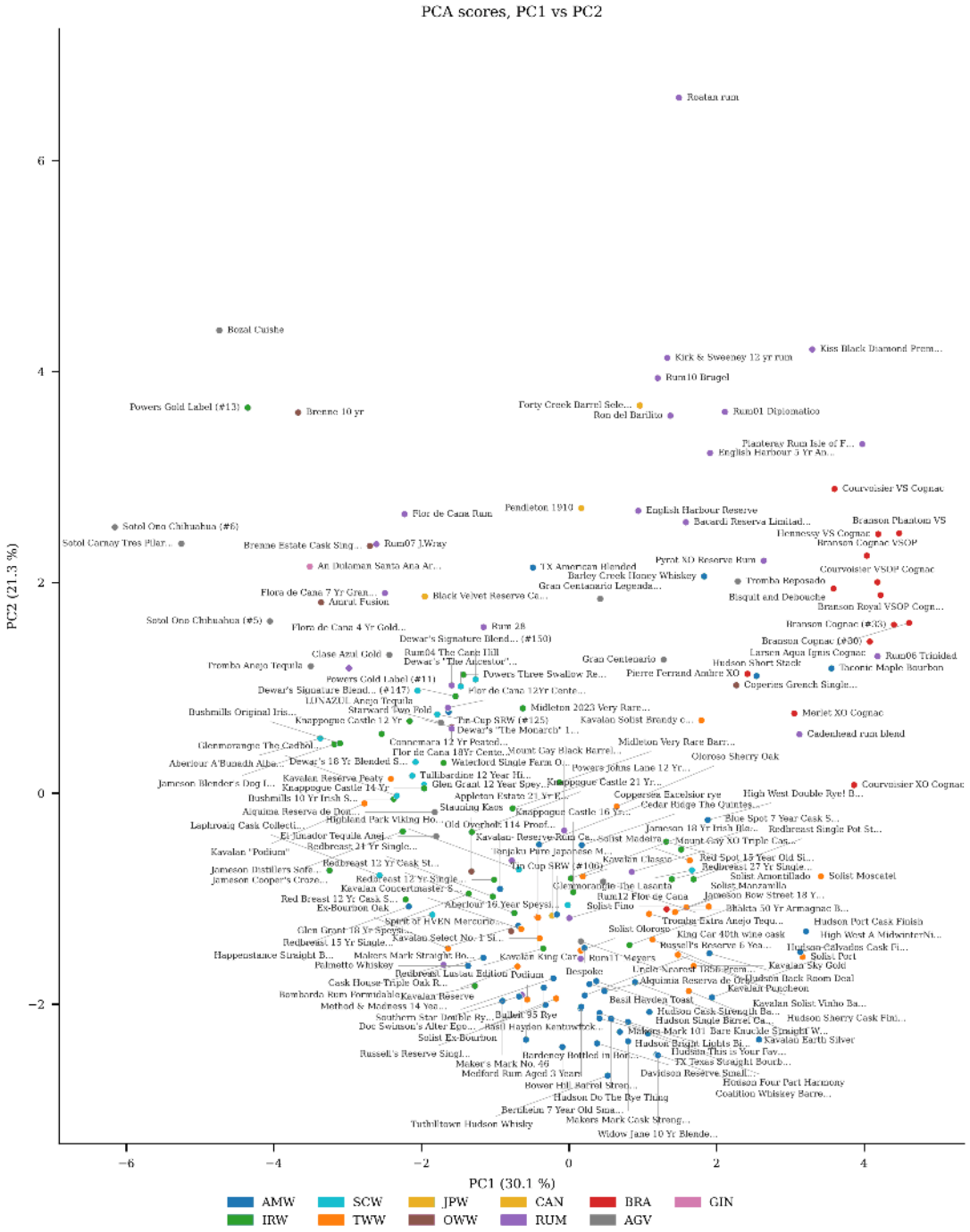

Figure S41. Block D2O-Qp (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

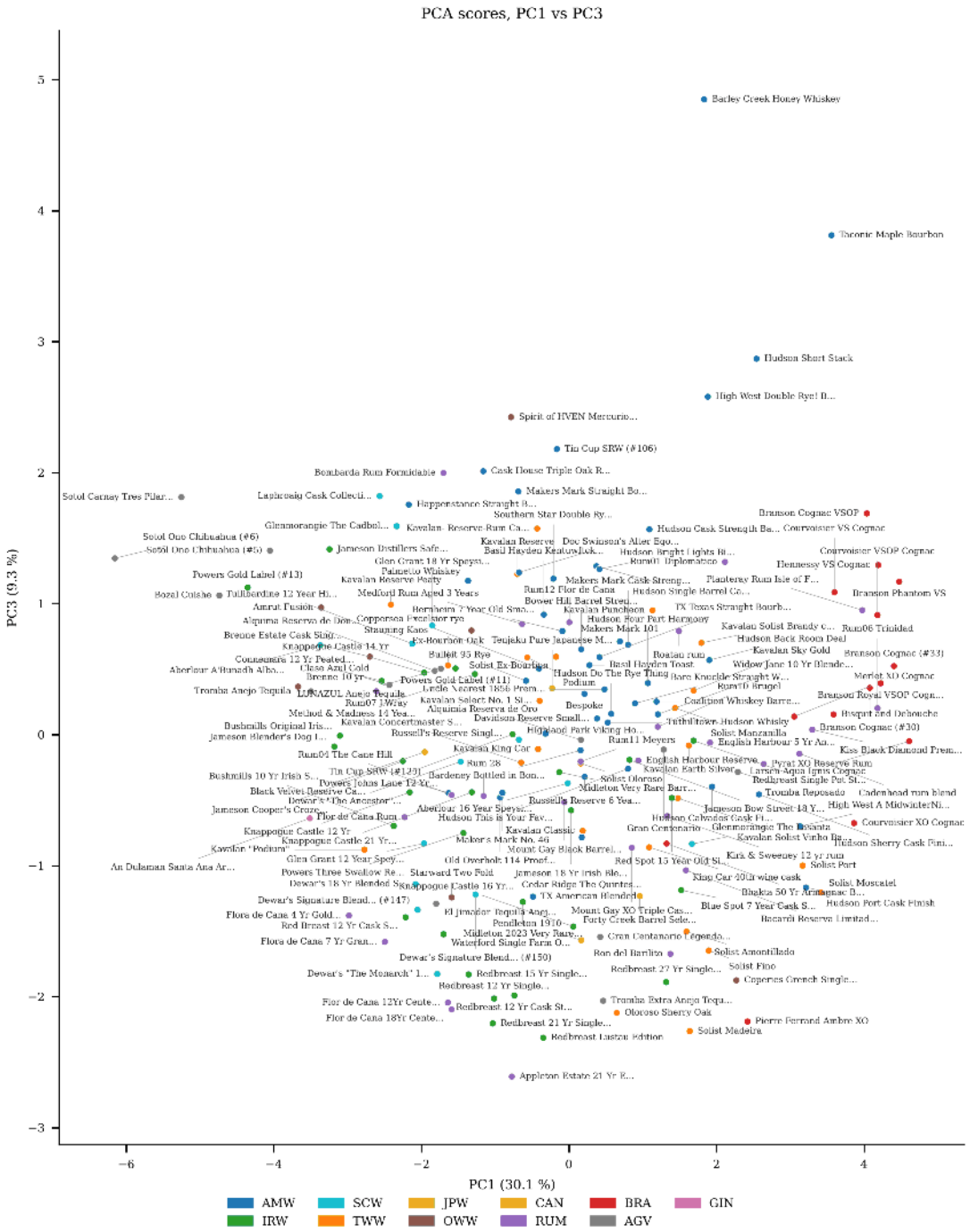

Figure S42. Block D2O-Qp (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

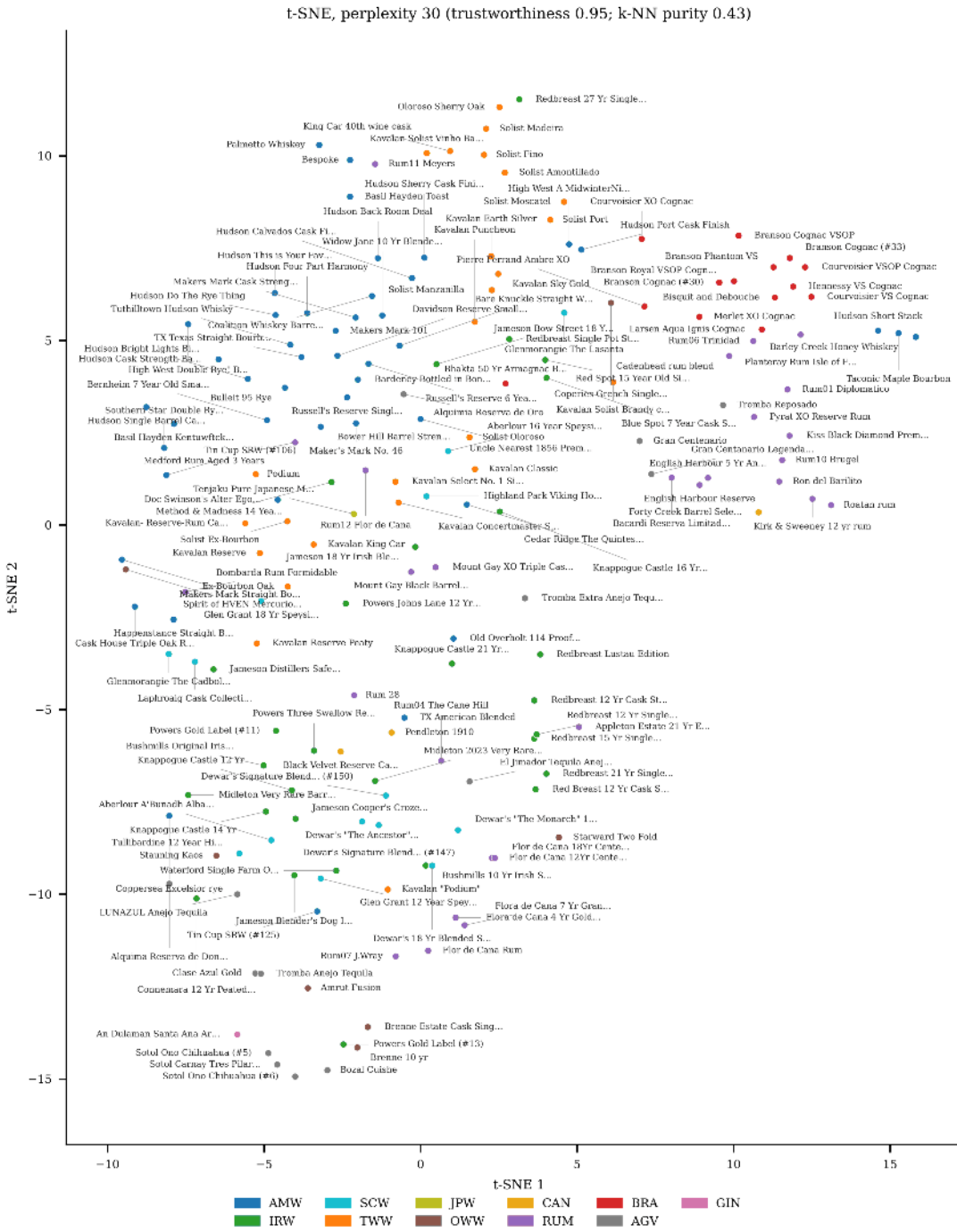

Figure S43. Block D2O-Qp (Table 20): t-SNE, perplexity 30 (trustworthiness 0.95; k-NN purity 0.43), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

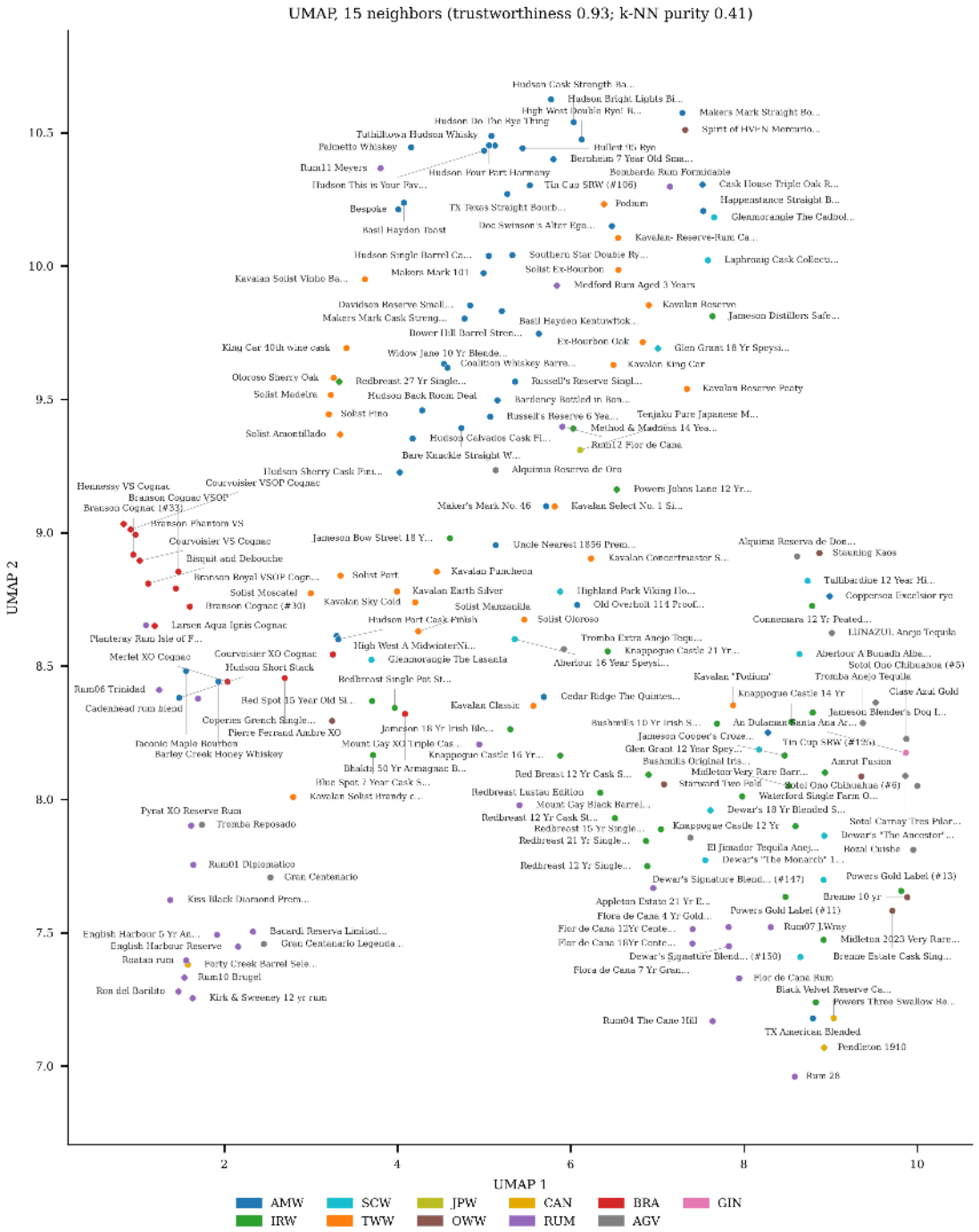

Figure S44. Block D2O-Qp (Table 20): UMAP, 15 neighbors (trustworthiness 0.93; k-NN purity 0.41), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

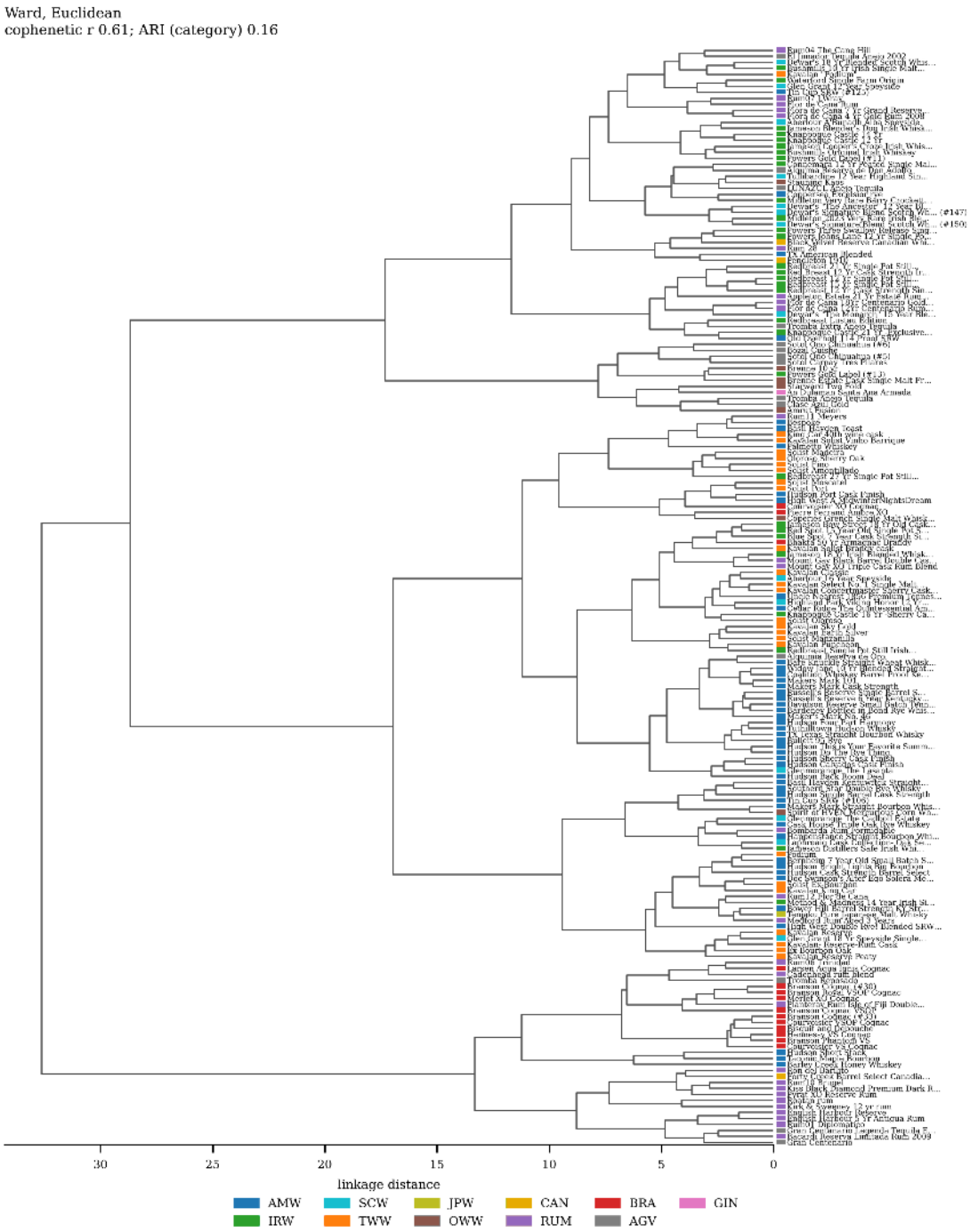

Figure S45. Block D2O-Qp (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

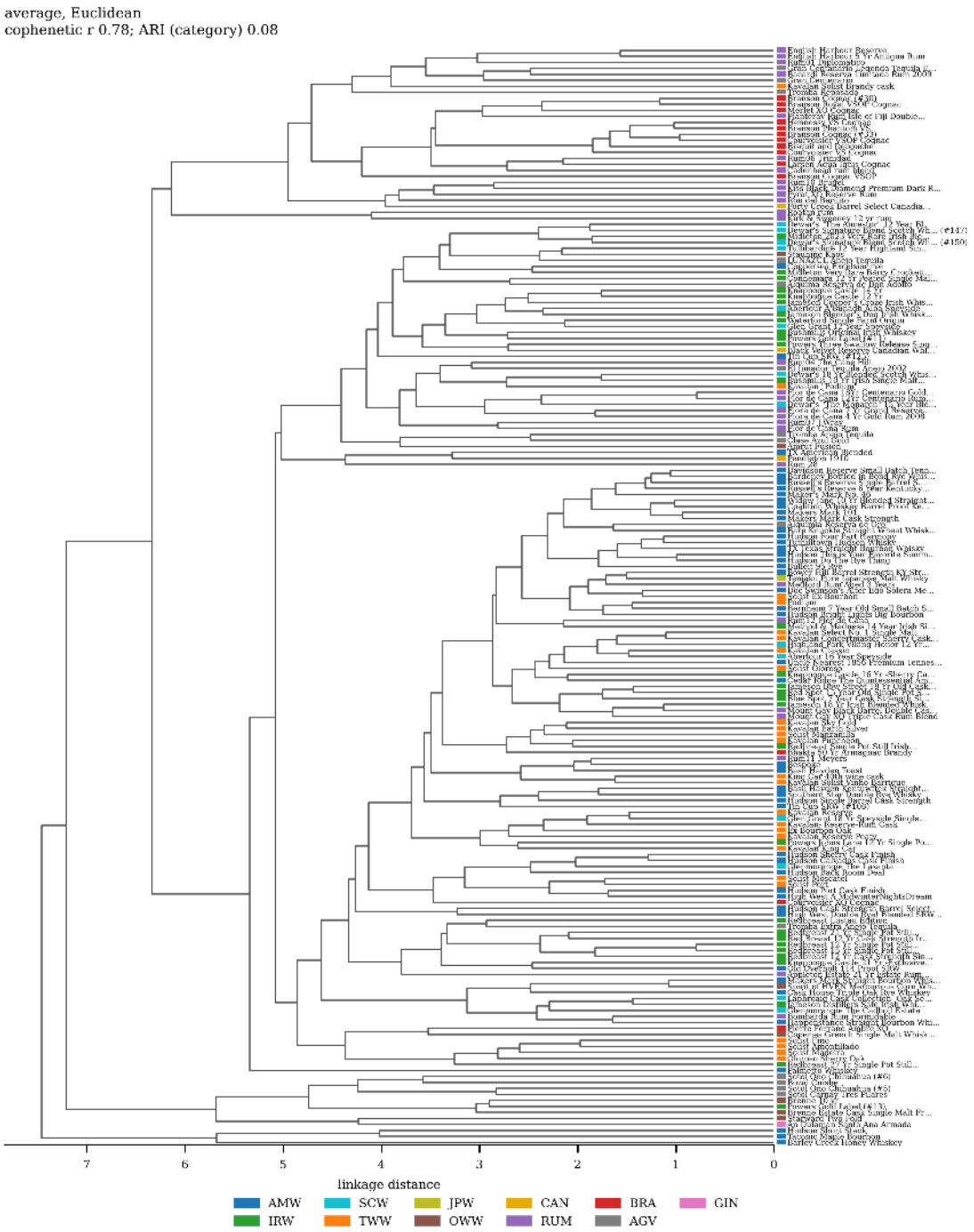

Figure S46. Block D2O-Qp (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

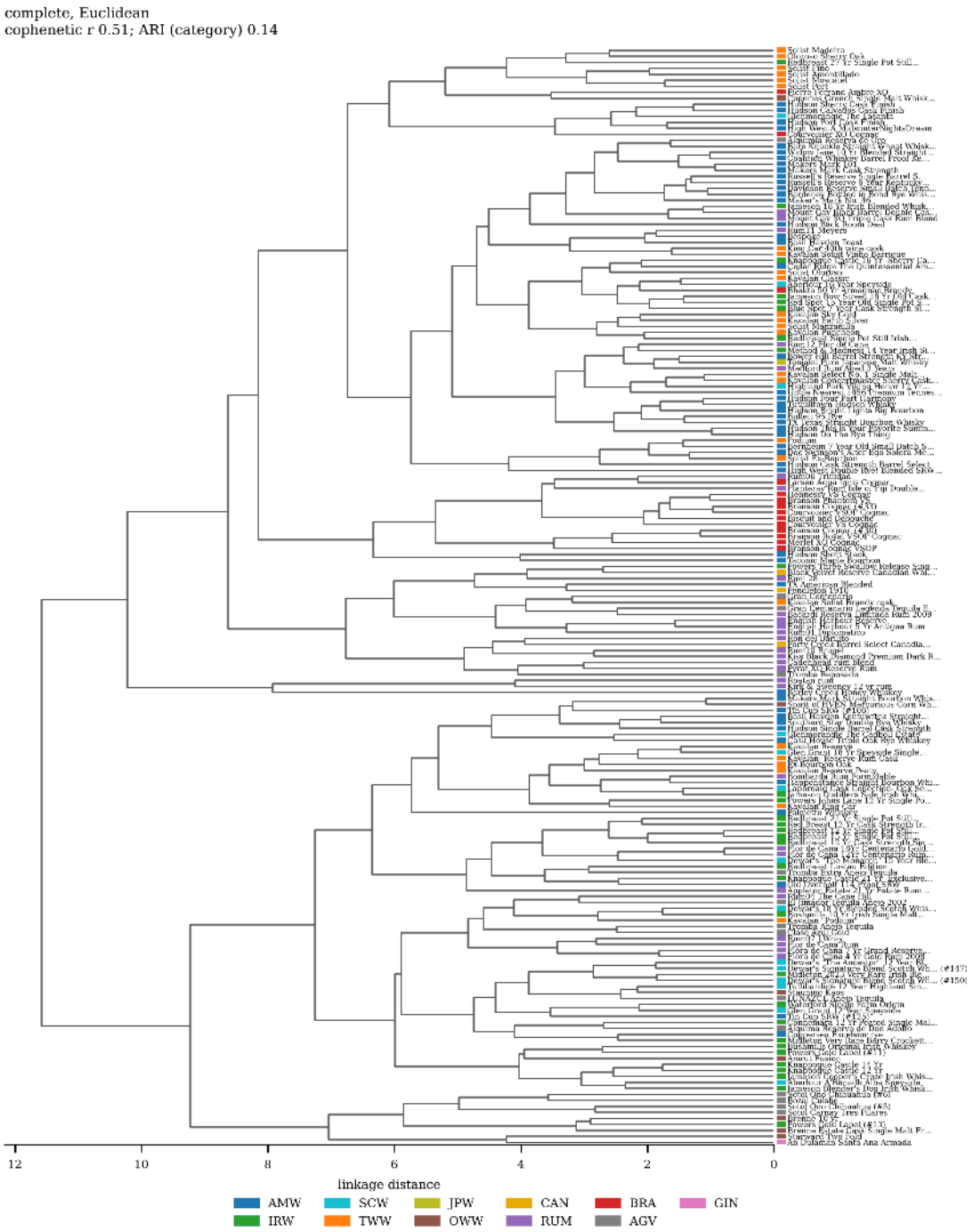

Figure S47. Block D2O-Qp (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

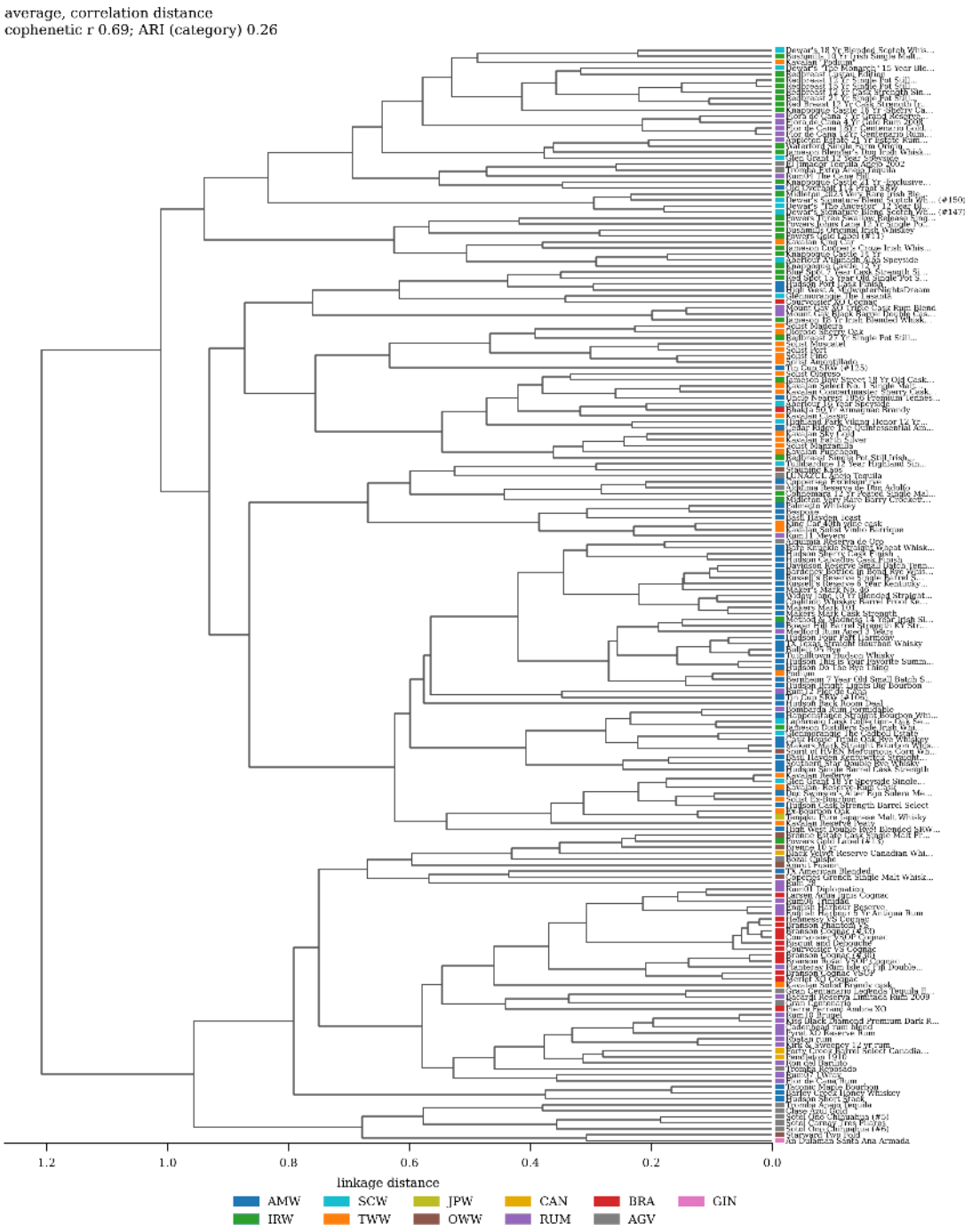

Figure S48. Block D2O-Qp (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block D2O-Qa: Figures S49–S56, enlarged from Figure 45

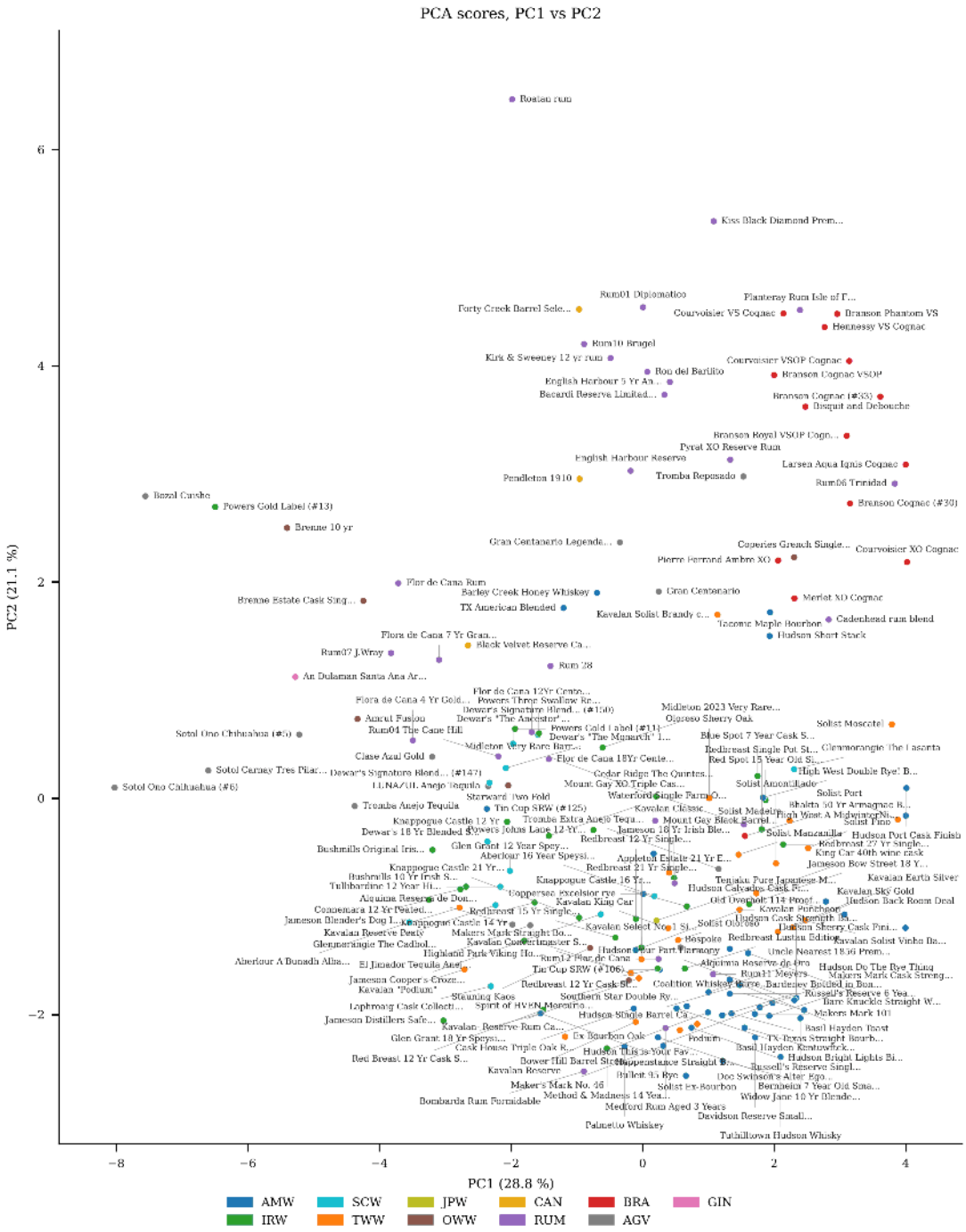

Figure S49. Block D2O-Qa (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

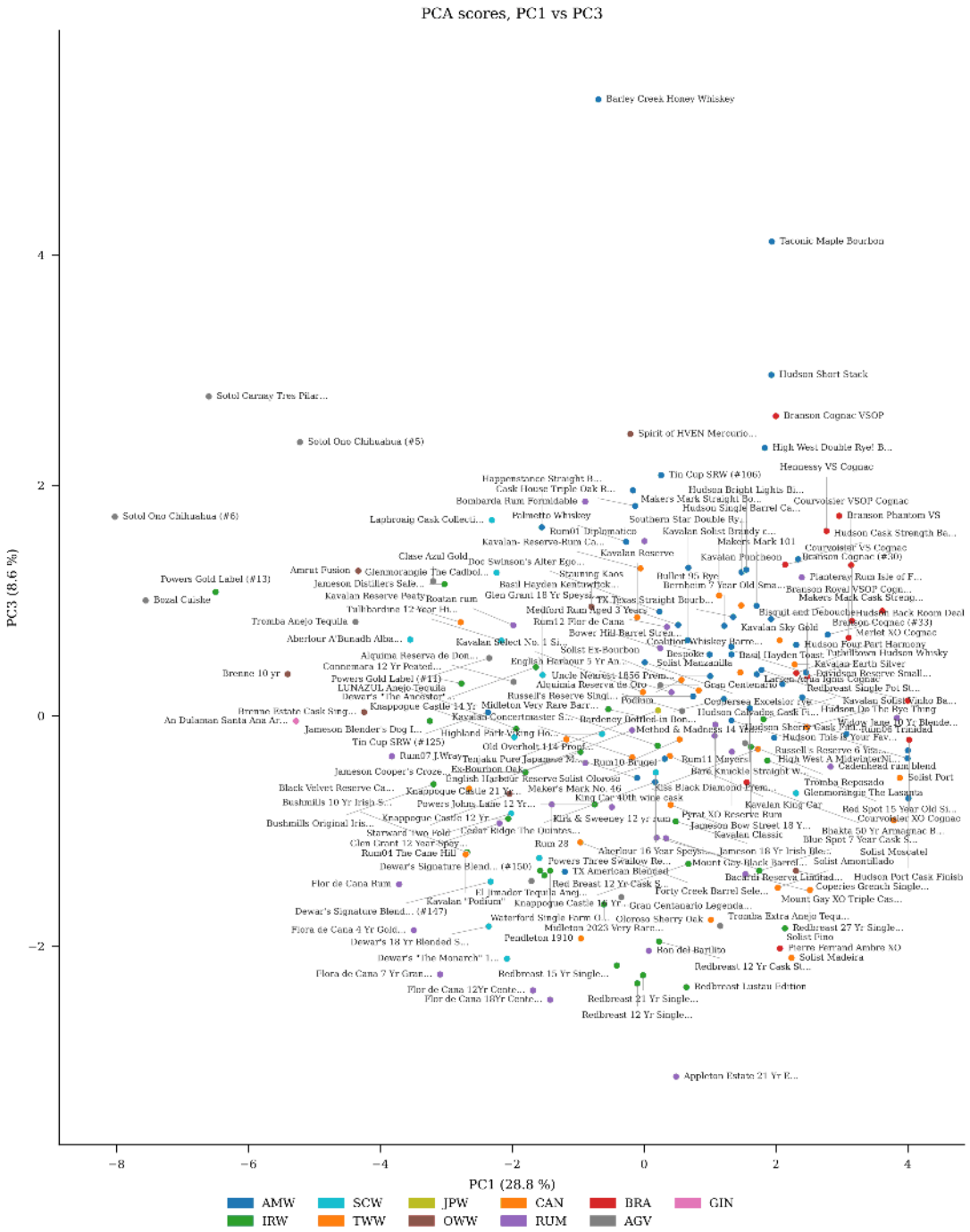

Figure S50. Block D2O-Qa (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

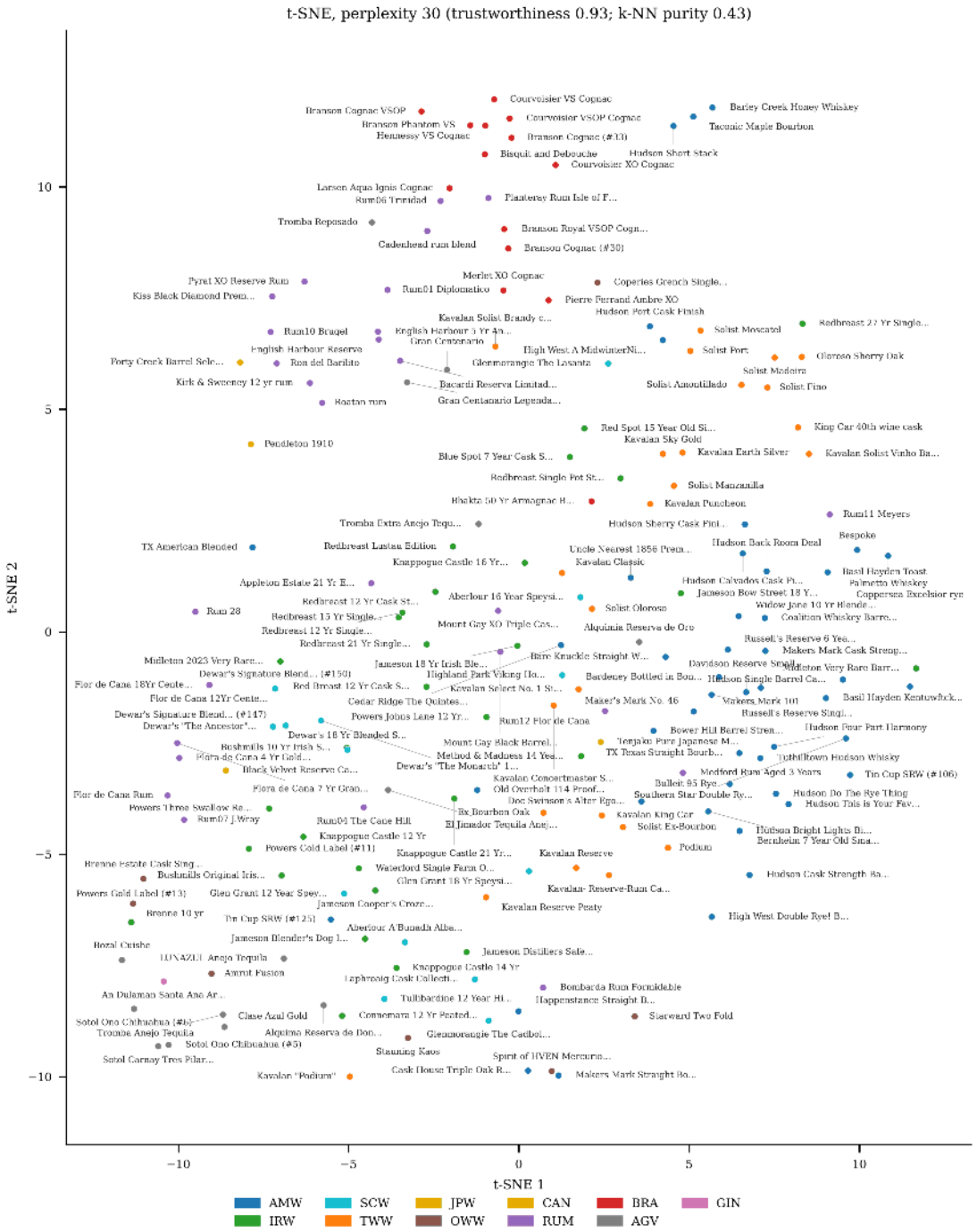

Figure S51. Block D2O-Qa (Table 20): t-SNE, perplexity 30 (trustworthiness 0.93; k-NN purity 0.43), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

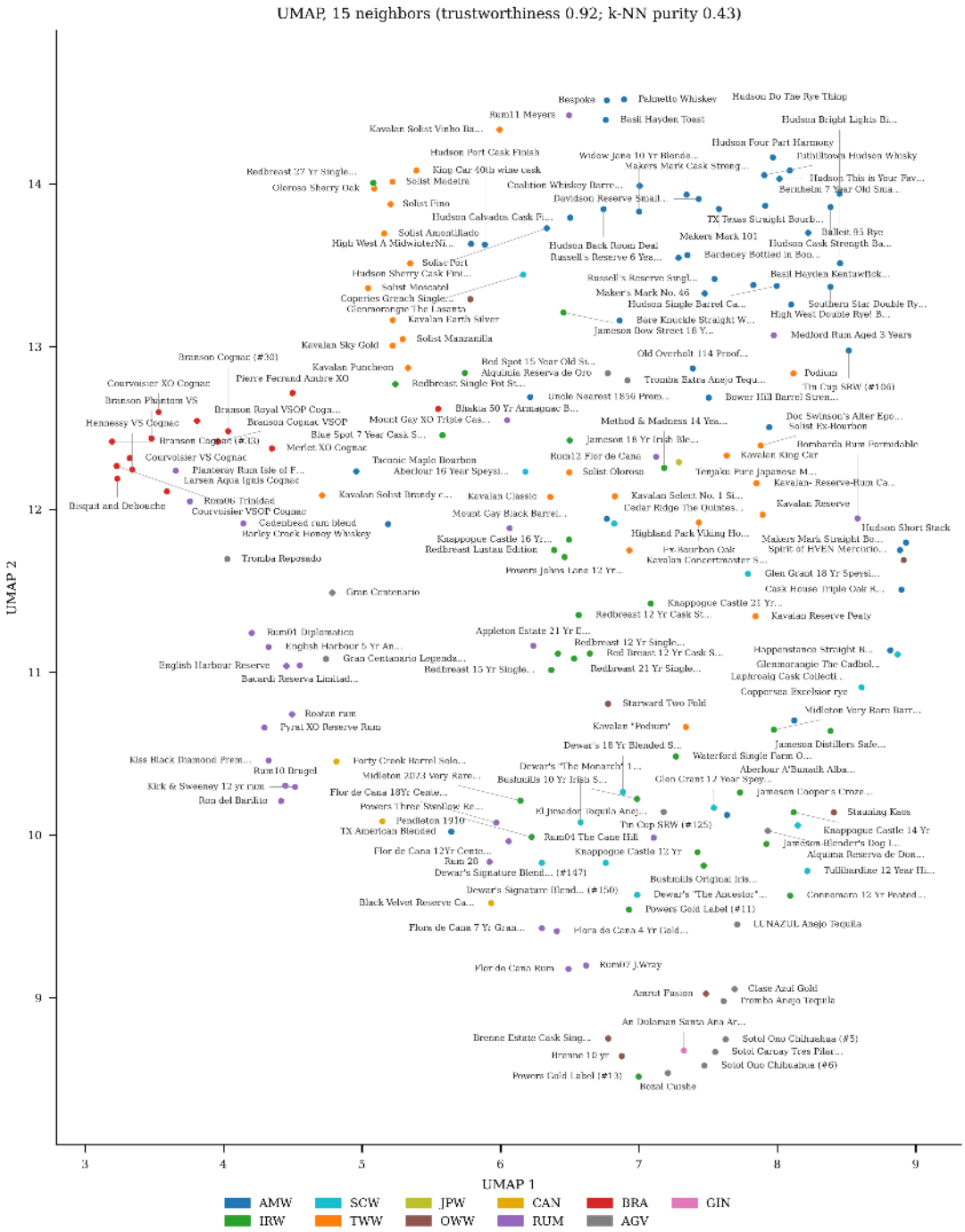

Figure S52. Block D2O-Qa (Table 20): UMAP, 15 neighbors (trustworthiness 0.92; k-NN purity 0.43), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

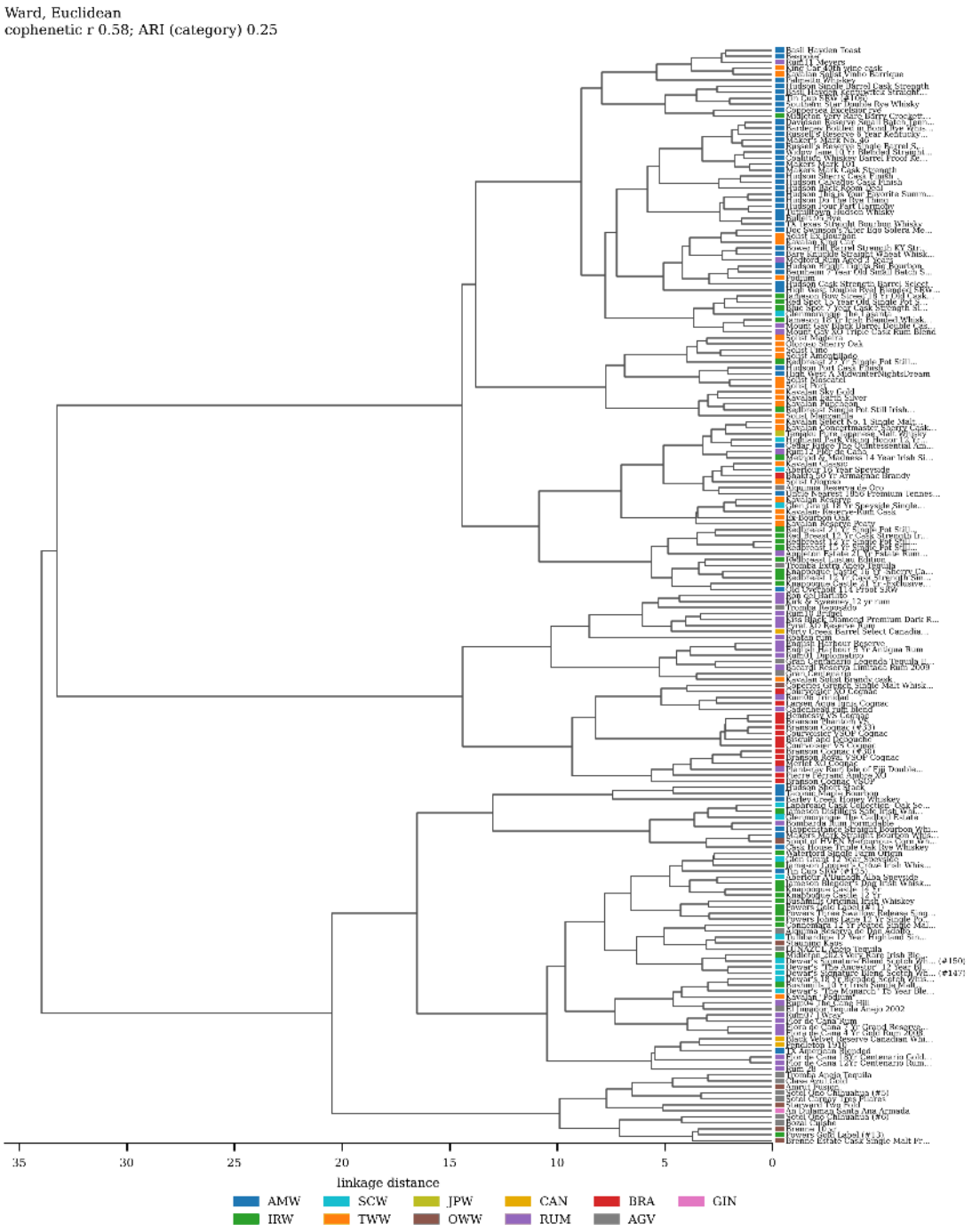

Figure S53. Block D2O-Qa (Table 20): Dendrogram by Ward linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

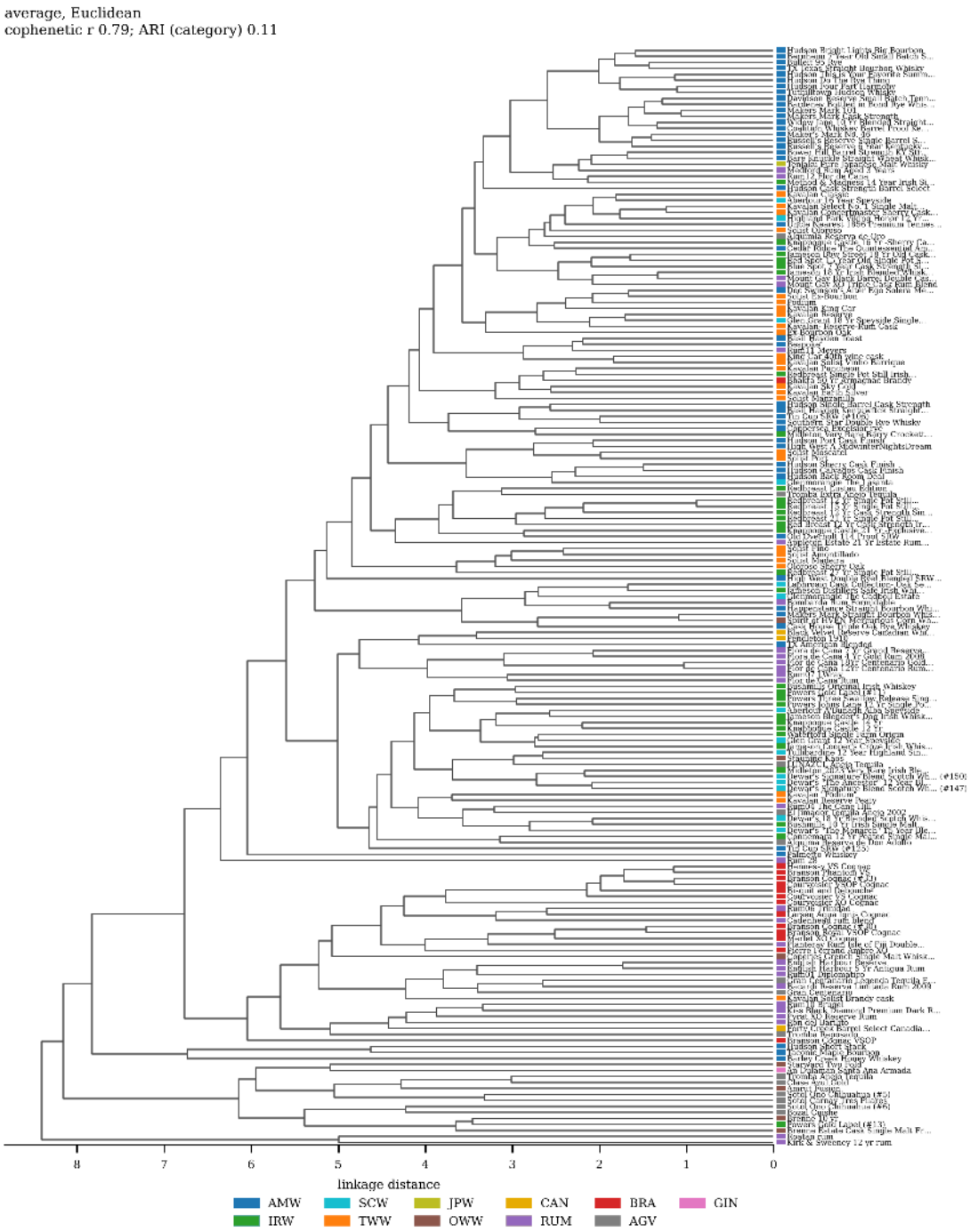

Figure S54. Block D2O-Qa (Table 20): Dendrogram by average linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

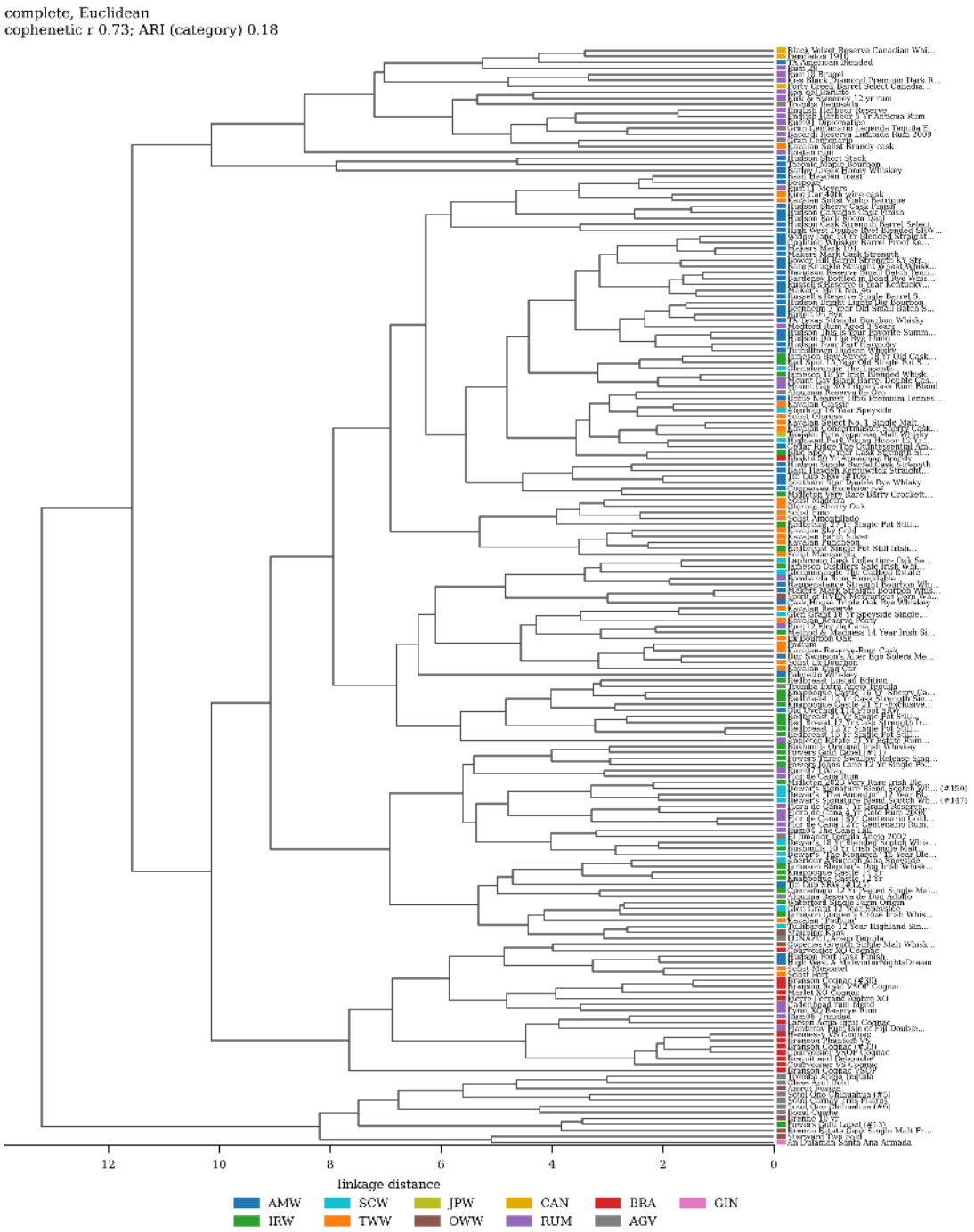

Figure S55. Block D2O-Qa (Table 20): Dendrogram by complete linkage on Euclidean distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

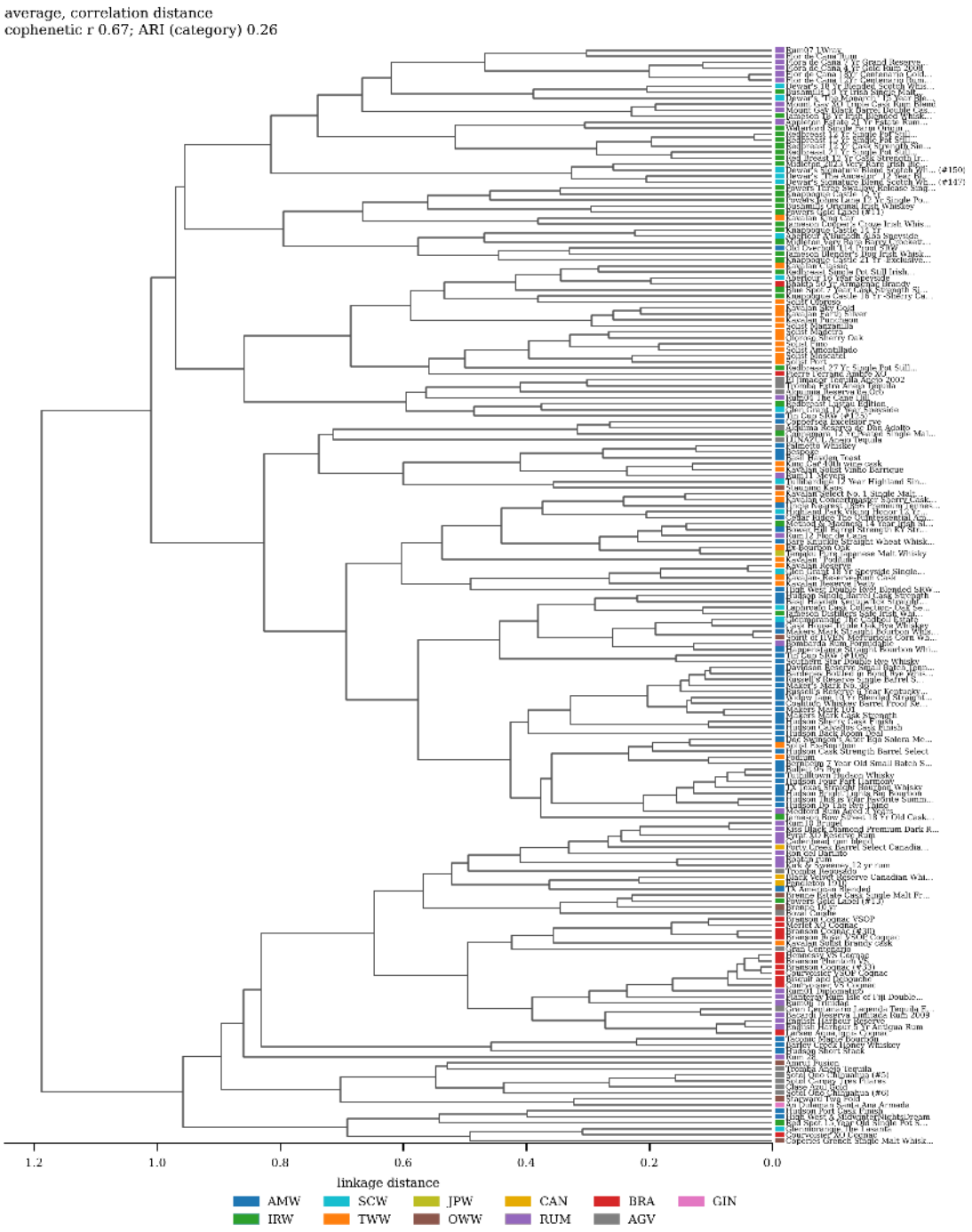

Figure S56. Block D2O-Qa (Table 20): Dendrogram by average linkage on correlation distances, every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block CDCl3: Figures S57–S61, enlarged from Figure 46

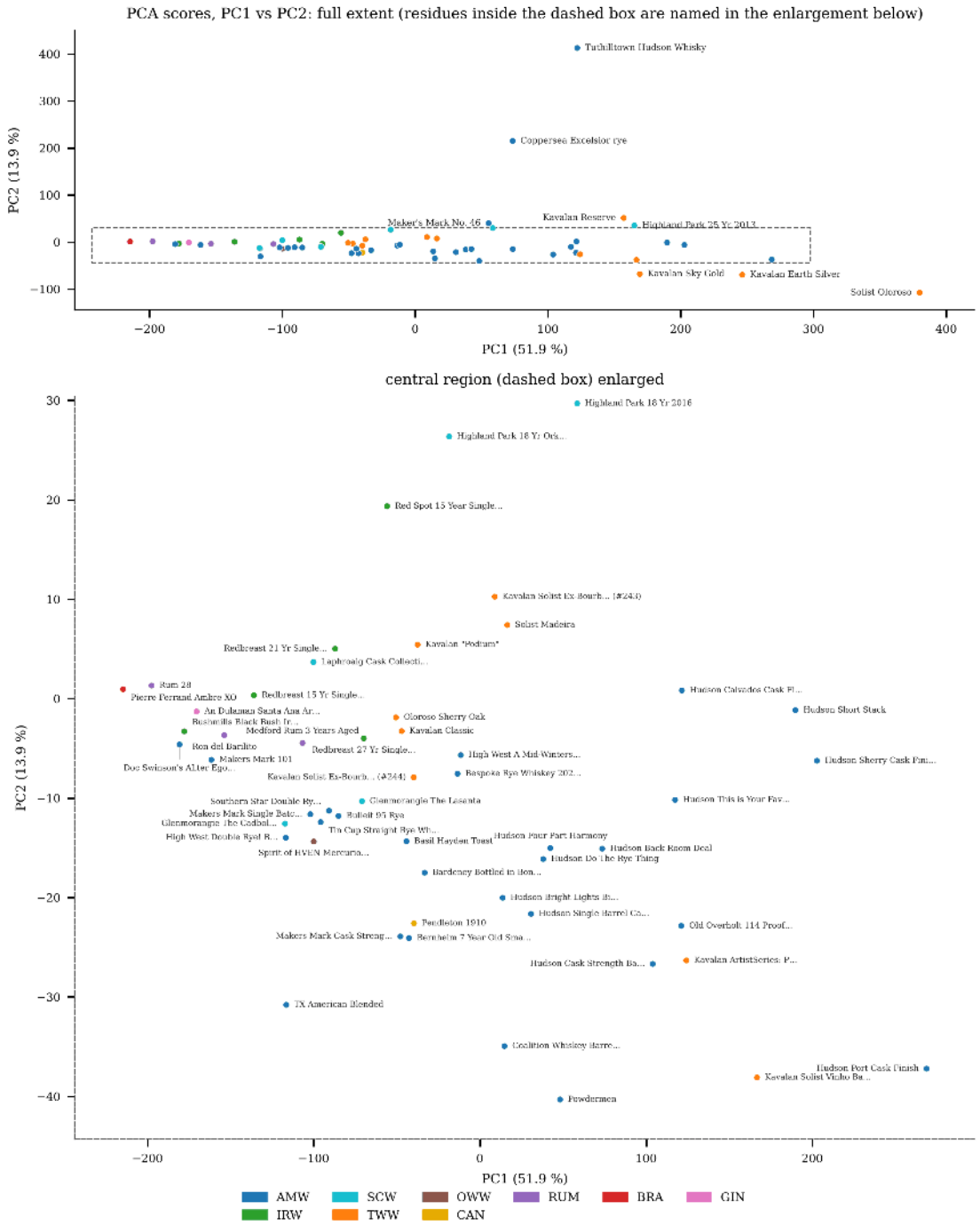

Figure S57. Block CDCl3 (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

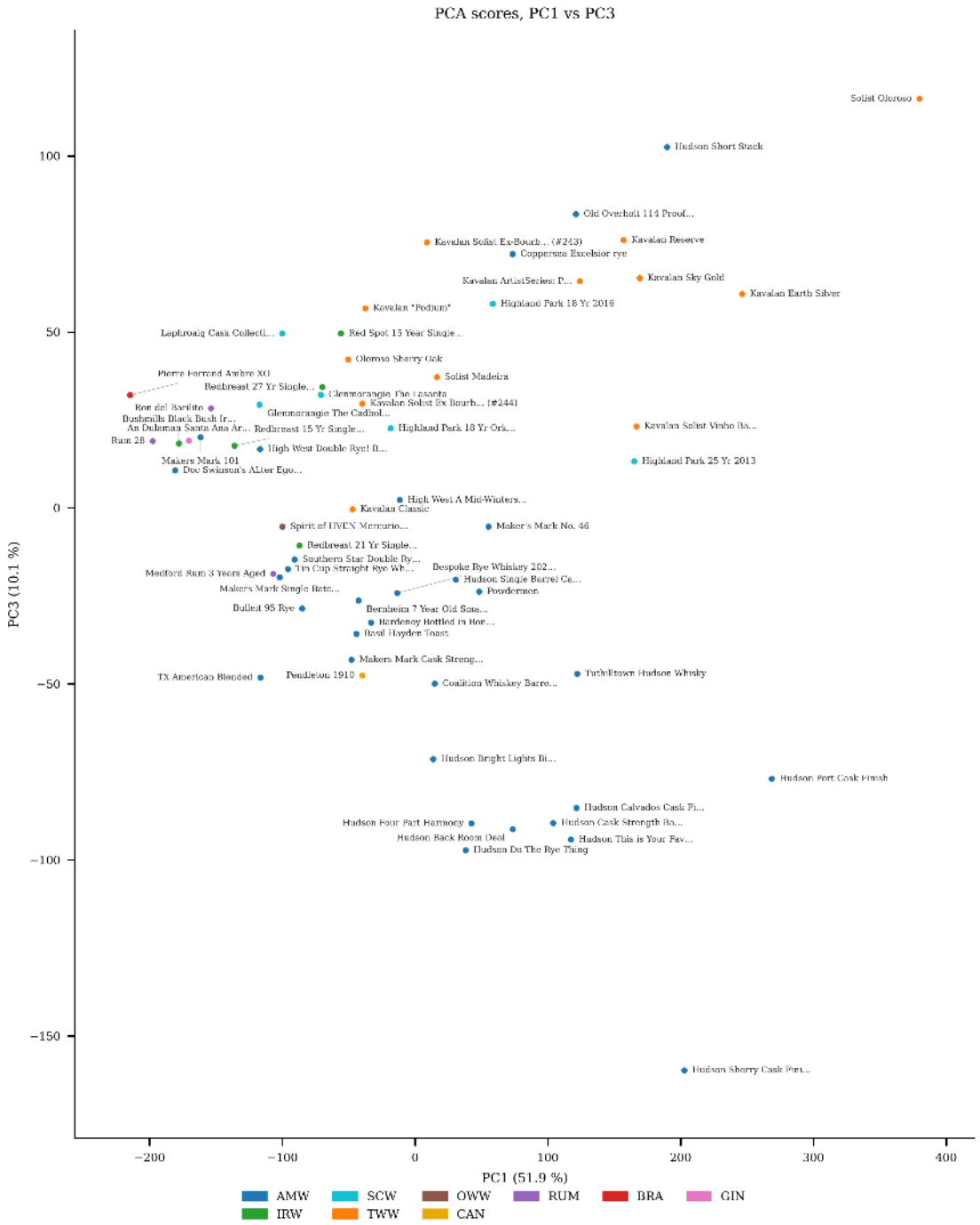

Figure S58. Block CDCl3 (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

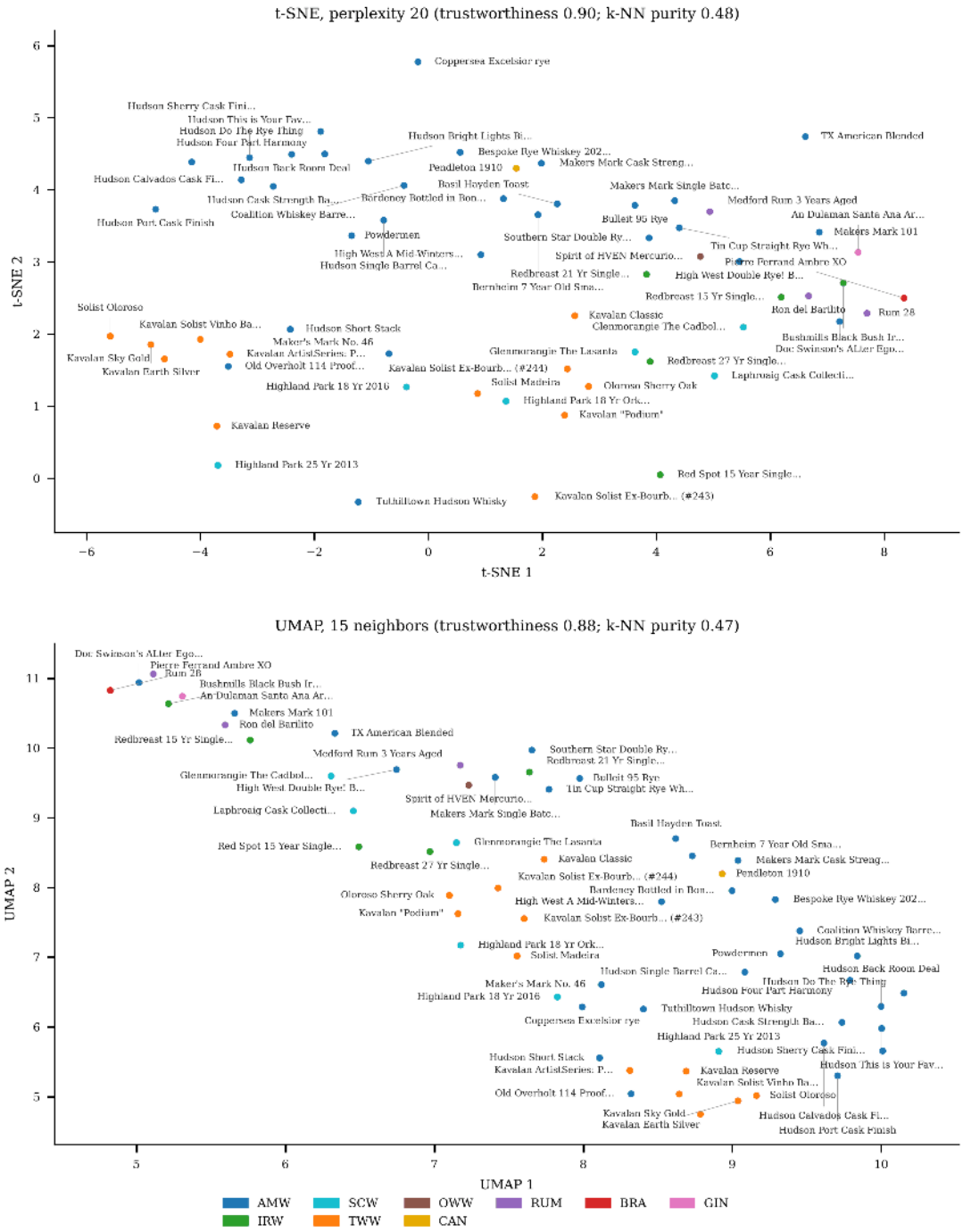

Figure S59. Block CDCl3 (Table 20): t-SNE, perplexity 20 (trustworthiness 0.90; k-NN purity 0.48) (top) and UMAP, 15 neighbors (trustworthiness 0.88; k-NN purity 0.47) (bottom), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

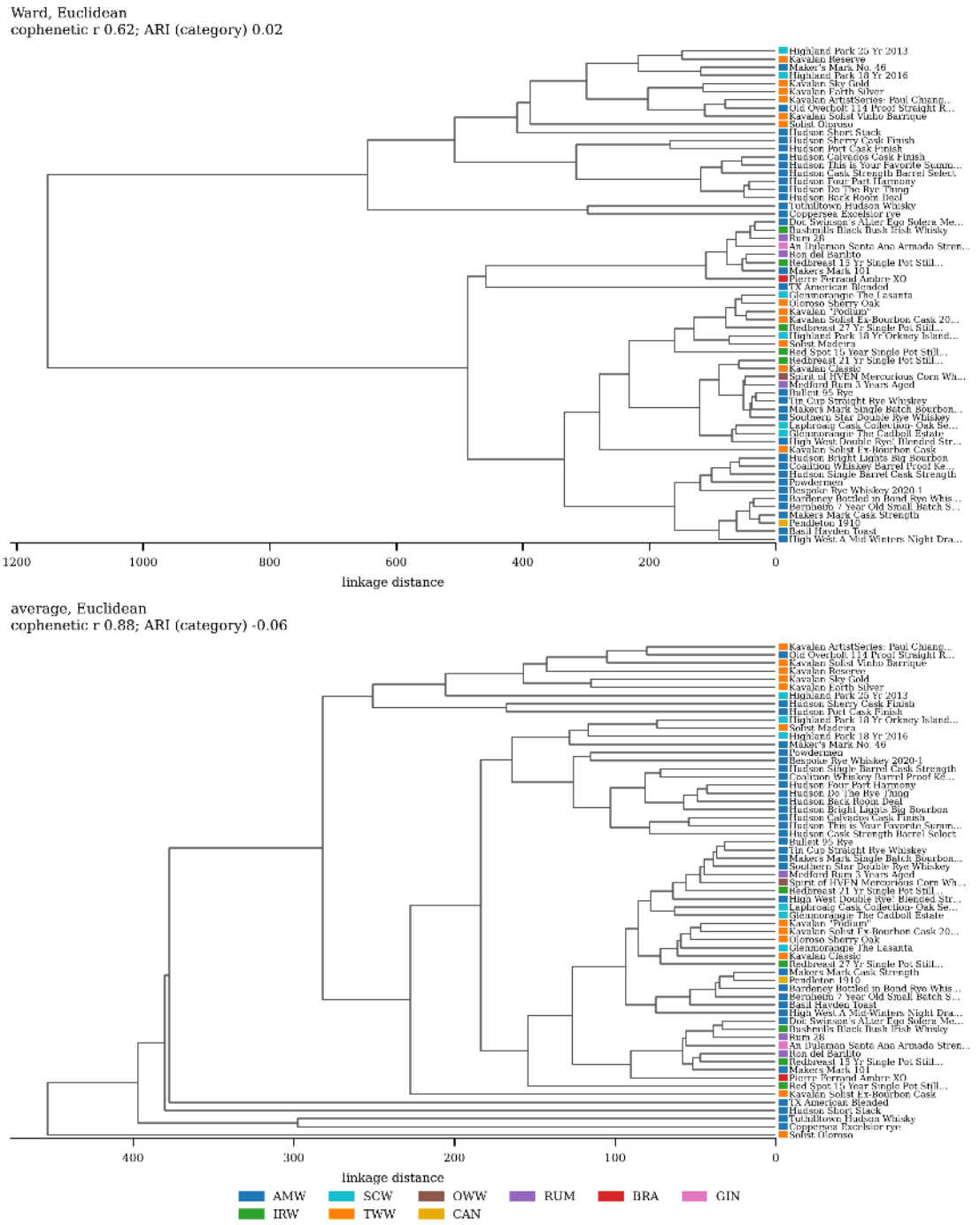

Figure S60. Block CDCl3 (Table 20): Dendrograms by Ward linkage on Euclidean distances (top); average linkage on Euclidean distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

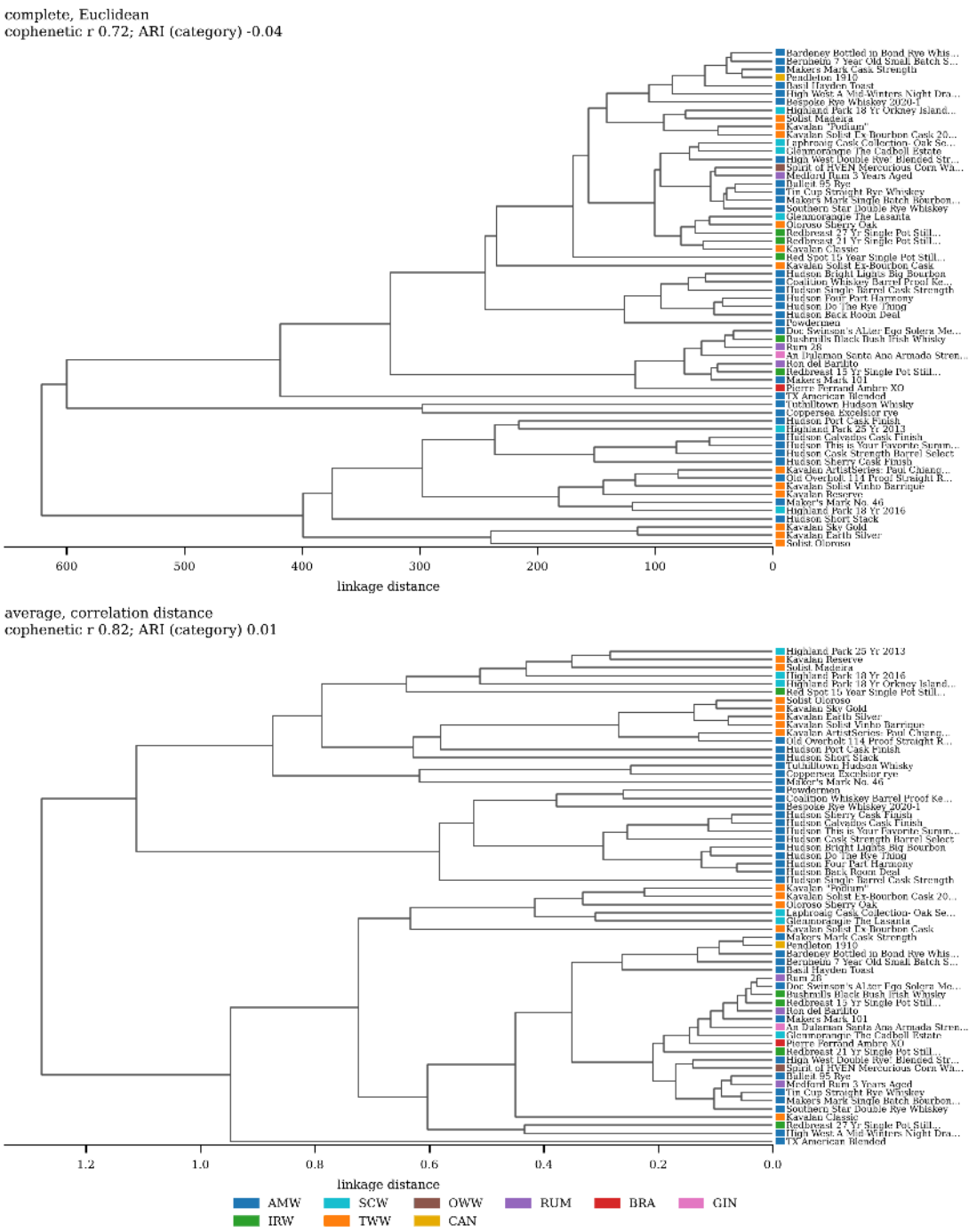

Figure S61. Block CDCl3 (Table 20): Dendrograms by complete linkage on Euclidean distances (top); average linkage on correlation distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block CDCl3-W: Figures S62–S66, enlarged from Figure 47

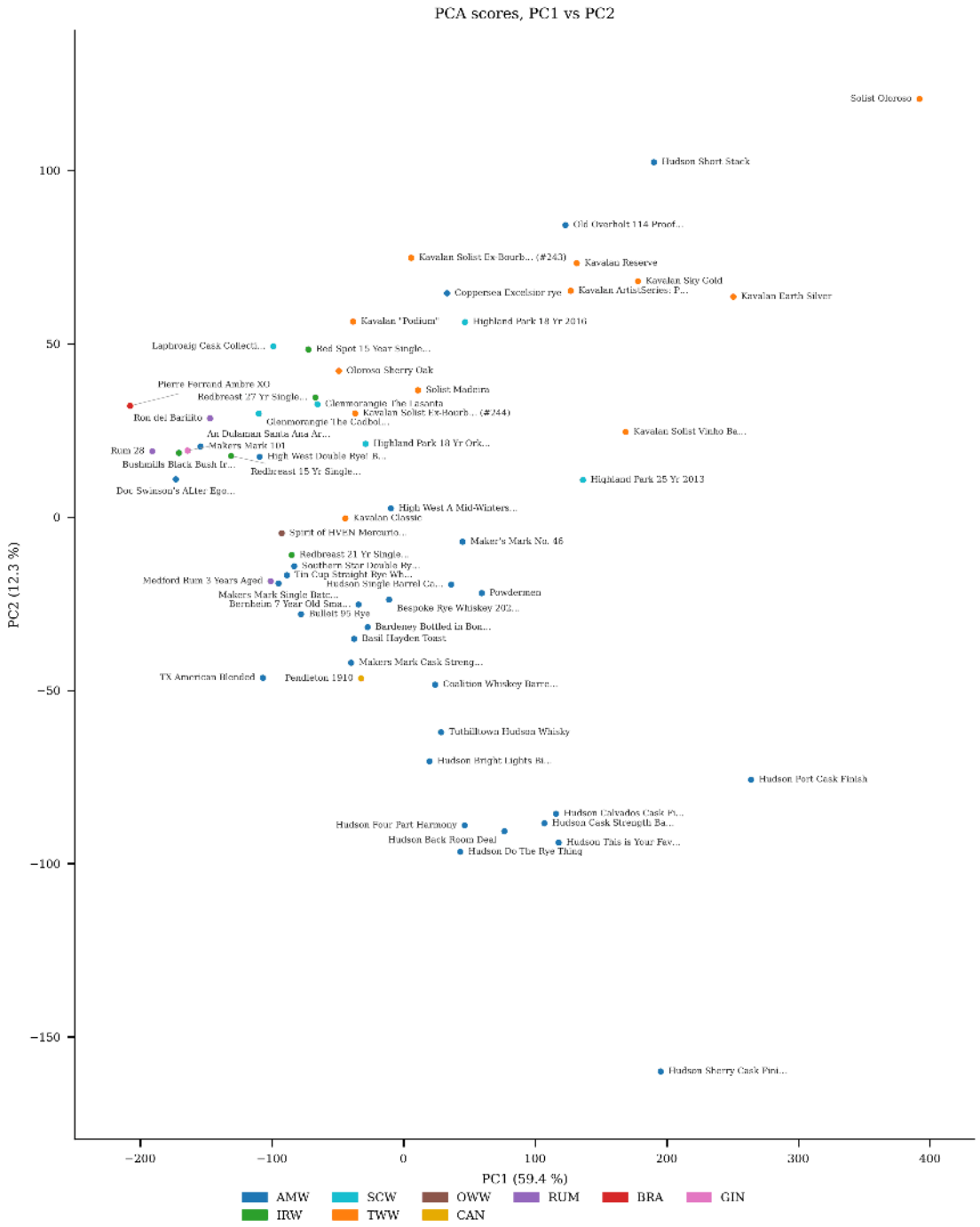

Figure S62. Block CDCl3-W (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

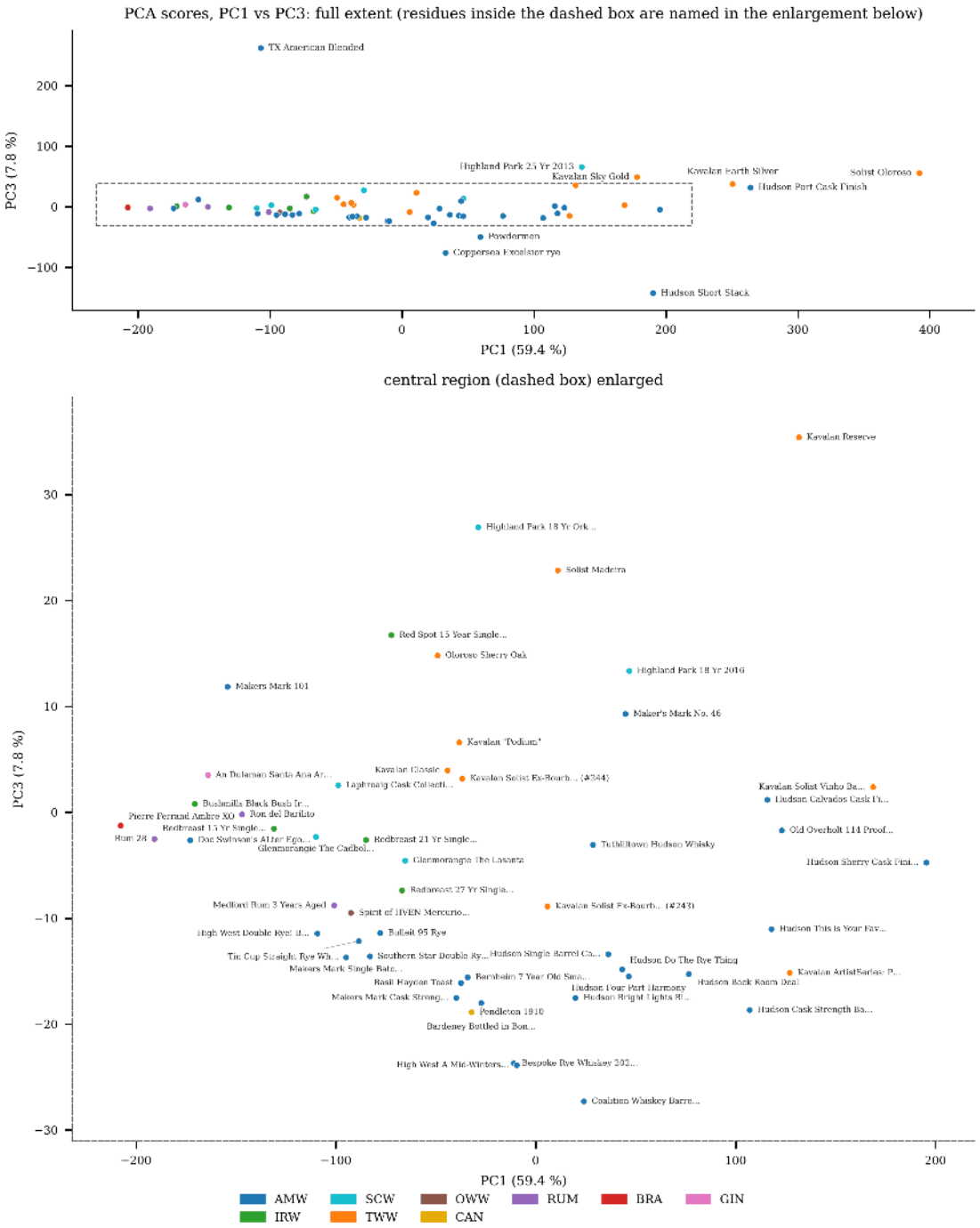

Figure S63. Block CDCl3-W (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

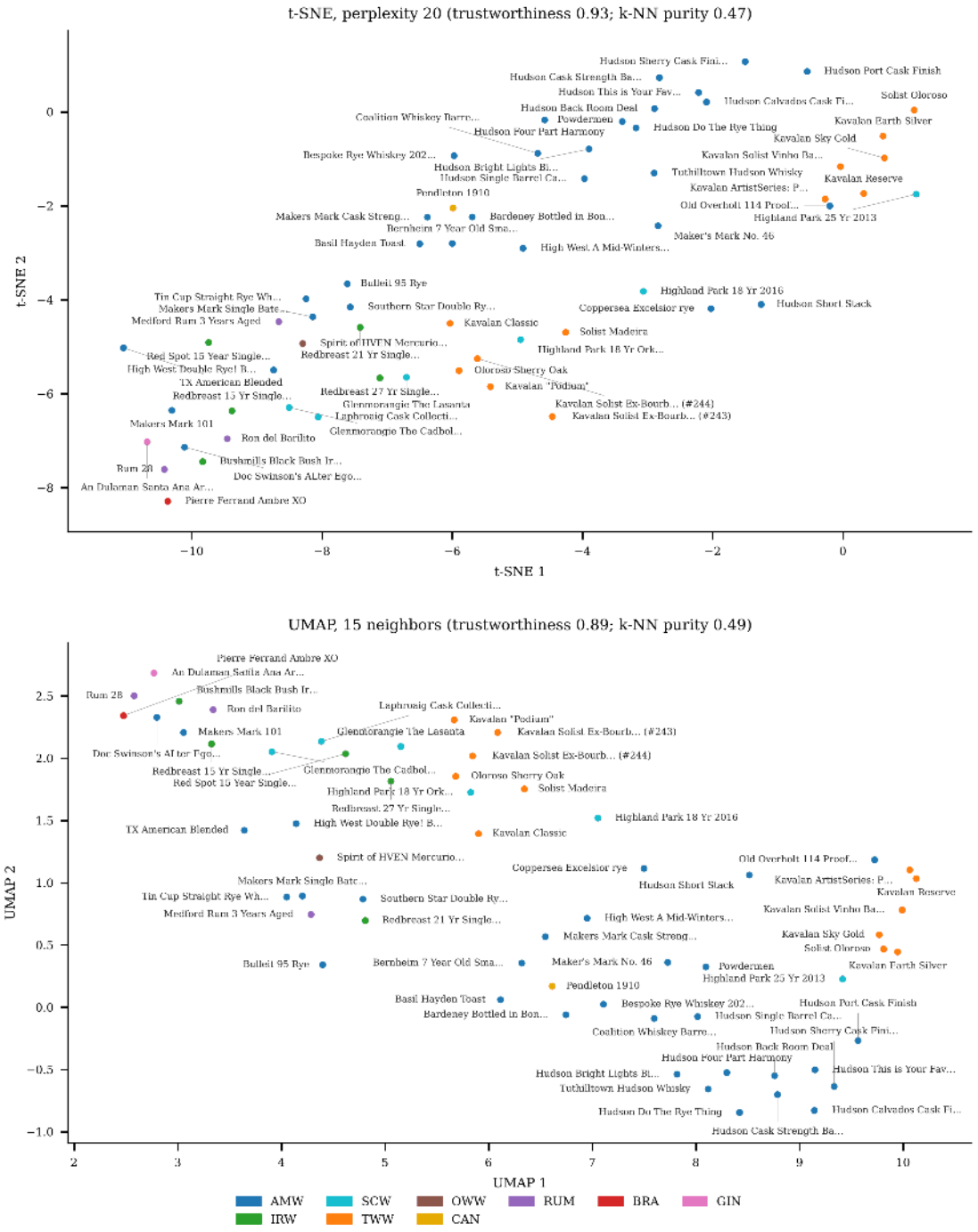

Figure S64. Block CDCl3-W (Table 20): t-SNE, perplexity 20 (trustworthiness 0.93; k-NN purity 0.47) (top) and UMAP, 15 neighbors (trustworthiness 0.89; k-NN purity 0.49) (bottom), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

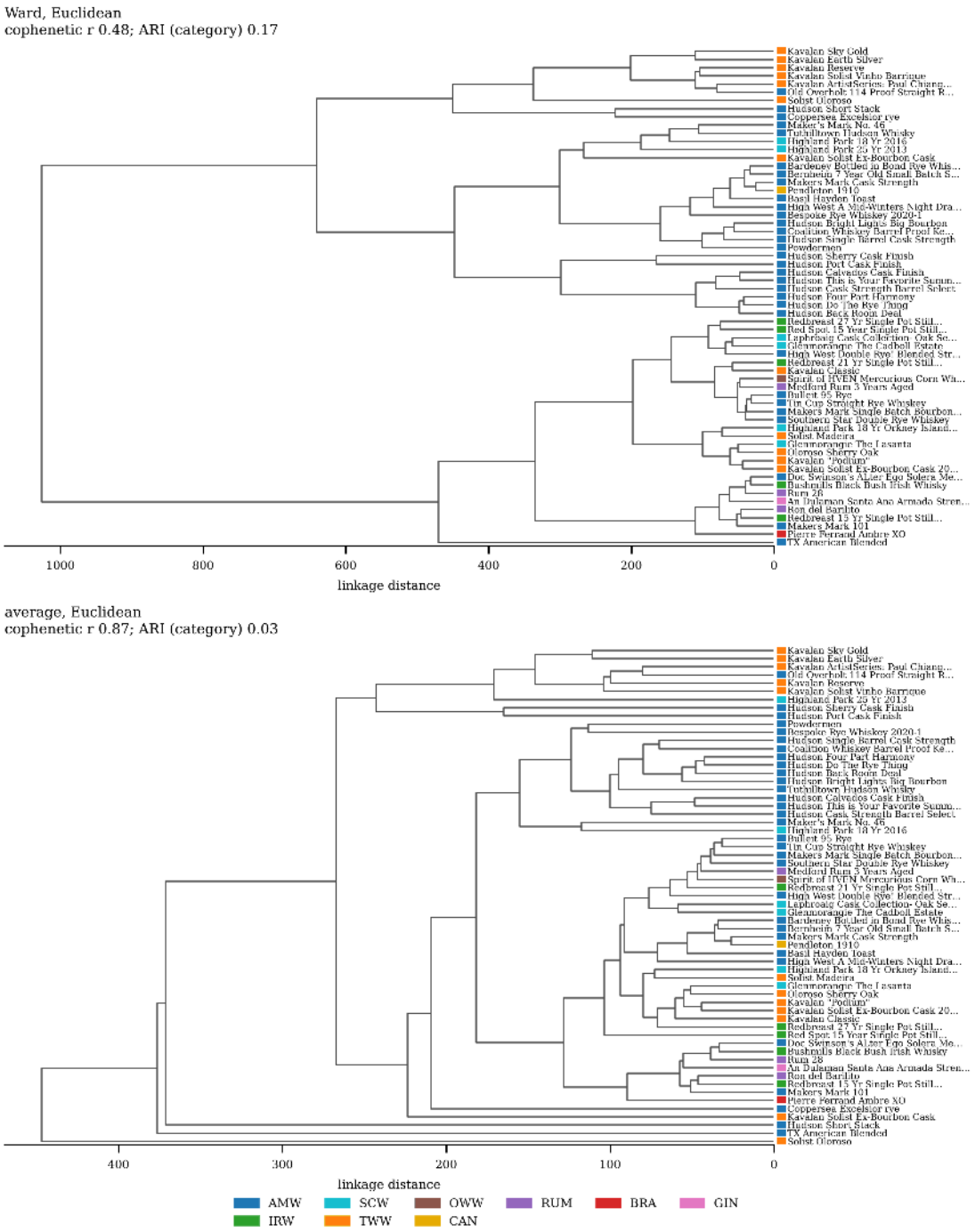

Figure S65. Block CDCl3-W (Table 20): Dendrograms by Ward linkage on Euclidean distances (top); average linkage on Euclidean distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

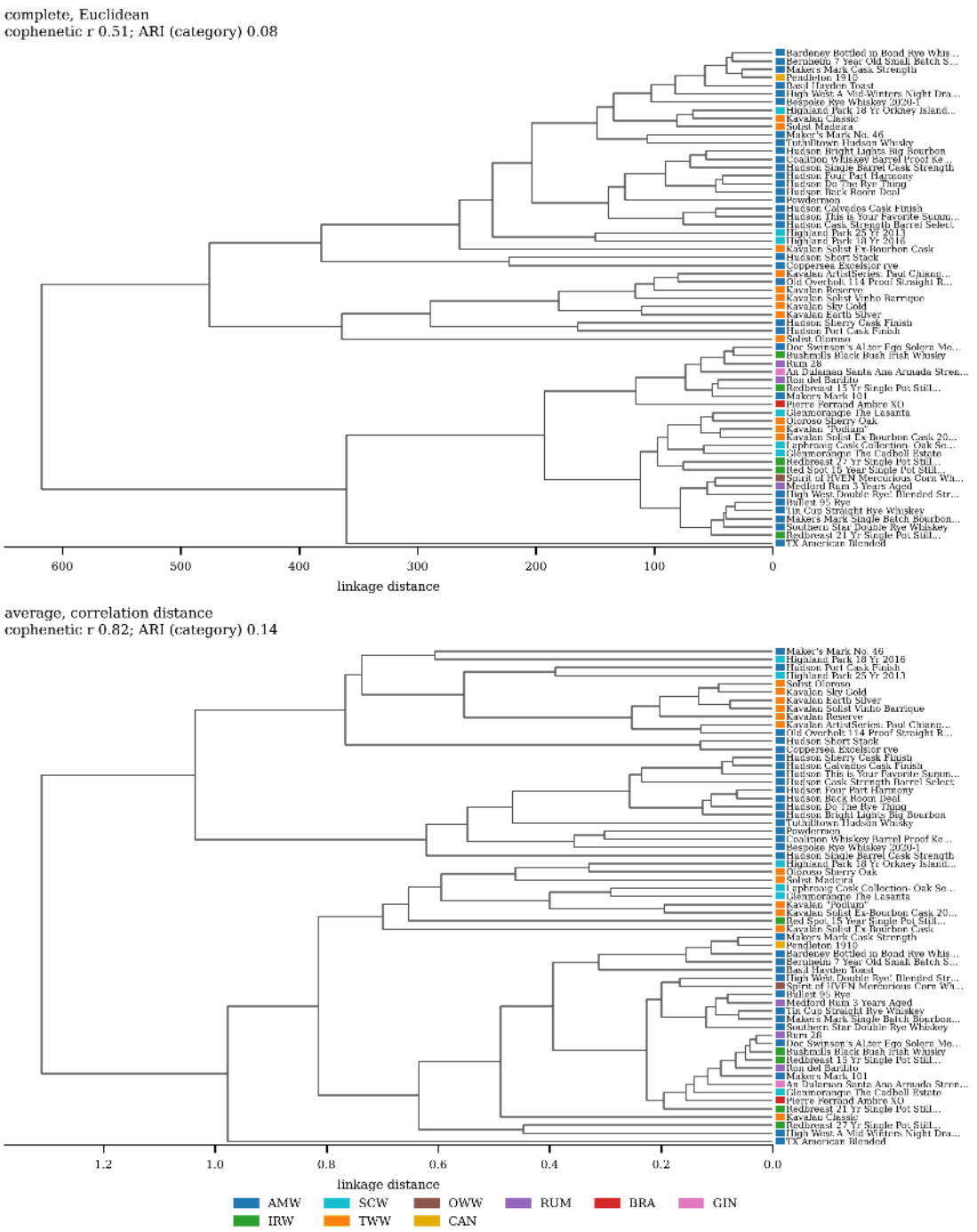

Figure S66. Block CDCl3-W (Table 20): Dendrograms by complete linkage on Euclidean distances (top); average linkage on correlation distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block FUSED-DC: Figures S67–S71, enlarged from Figure 48

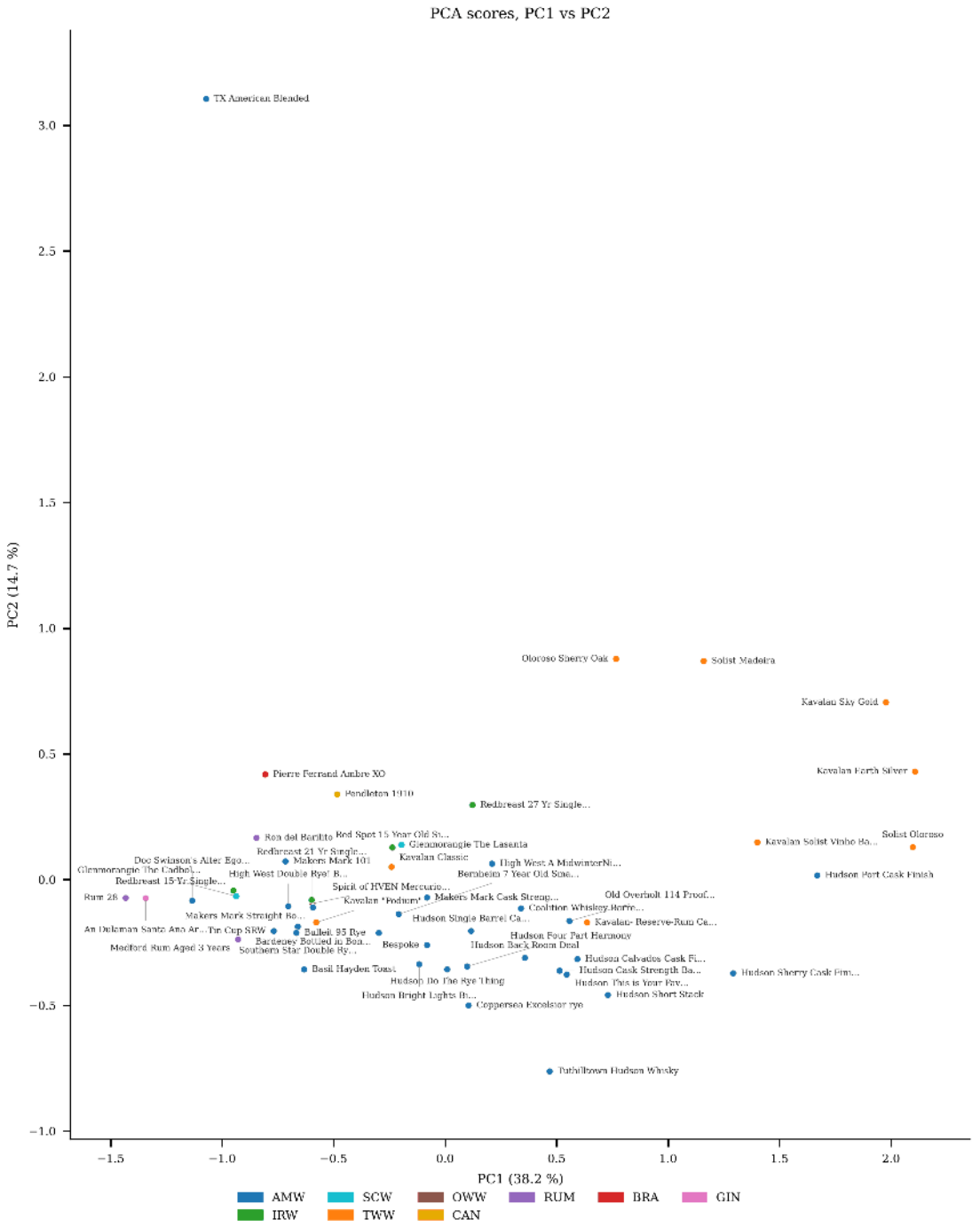

Figure S67. Block FUSED-DC (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

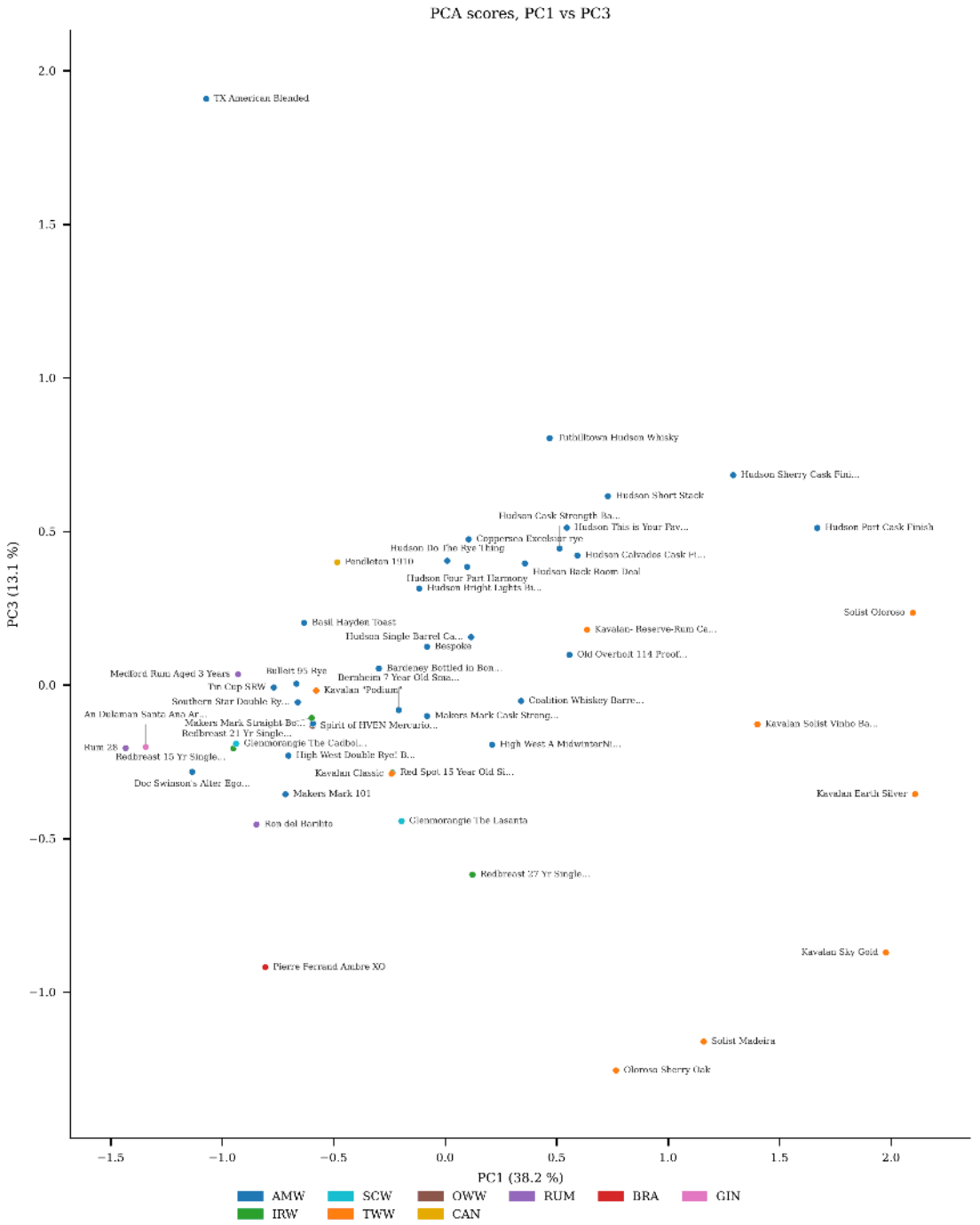

Figure S68. Block FUSED-DC (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

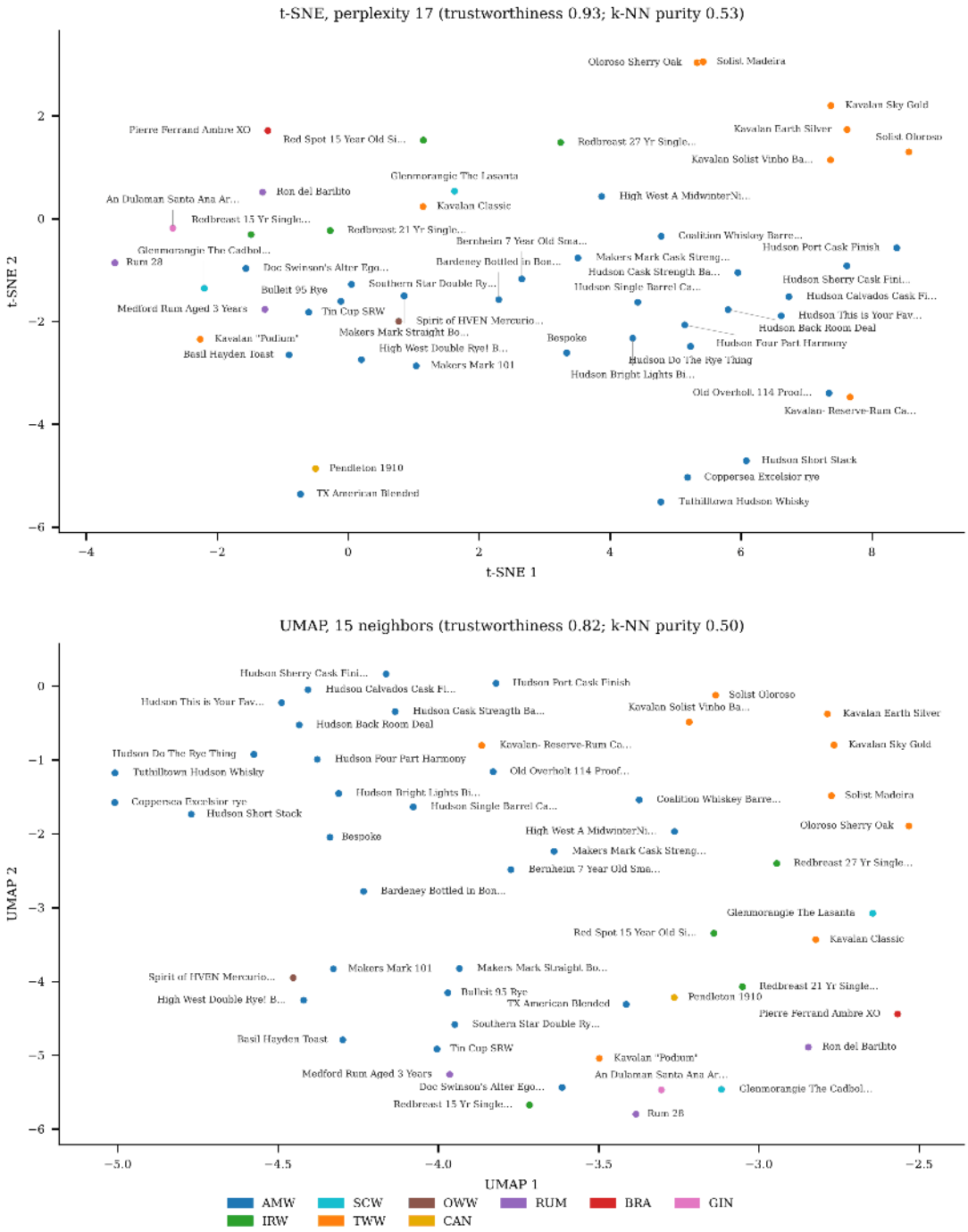

Figure S69. Block FUSED-DC (Table 20): t-SNE, perplexity 17 (trustworthiness 0.93; k-NN purity 0.53) (top) and UMAP, 15 neighbors (trustworthiness 0.82; k-NN purity 0.50) (bottom), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

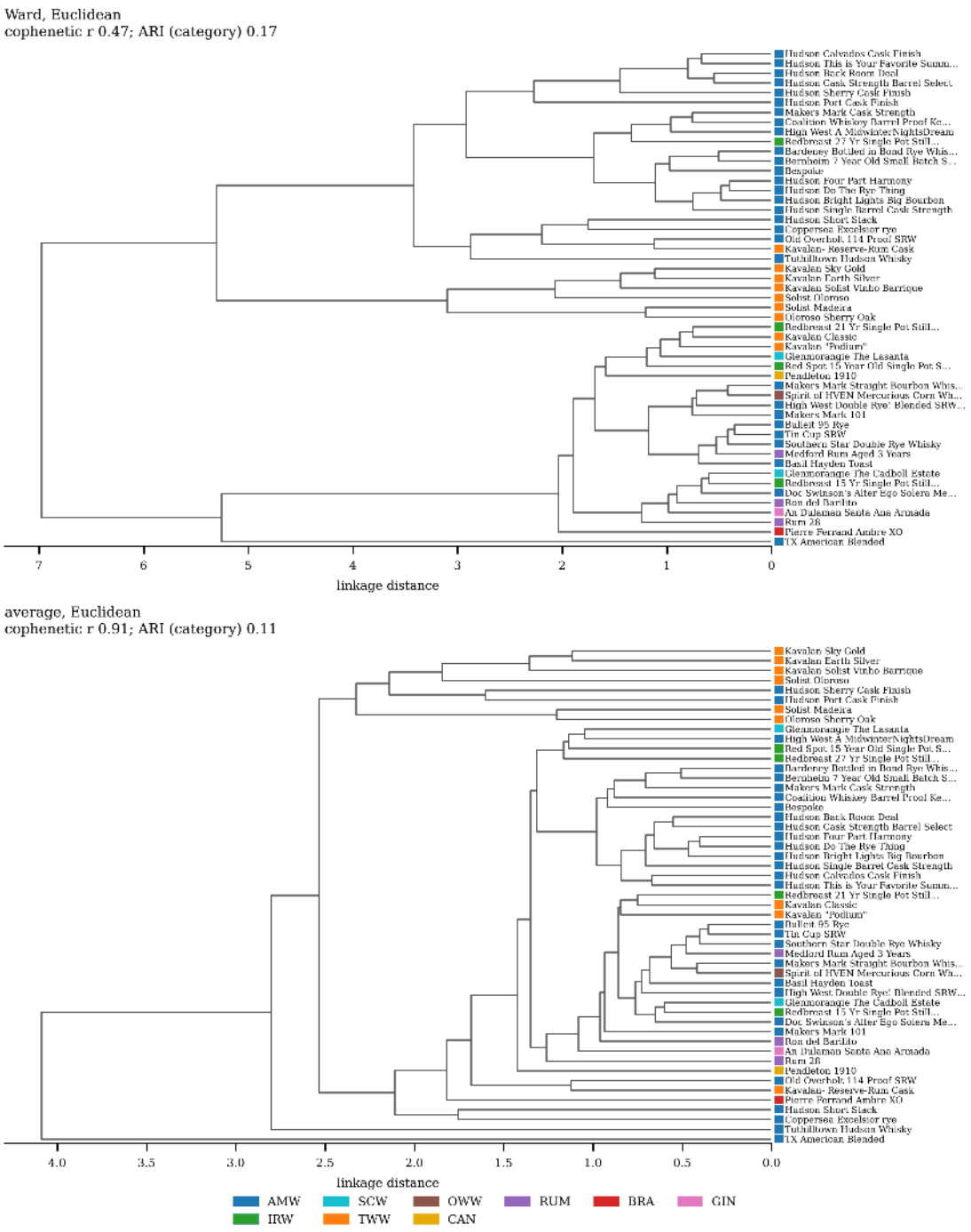

Figure S70. Block FUSED-DC (Table 20): Dendrograms by Ward linkage on Euclidean distances (top); average linkage on Euclidean distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

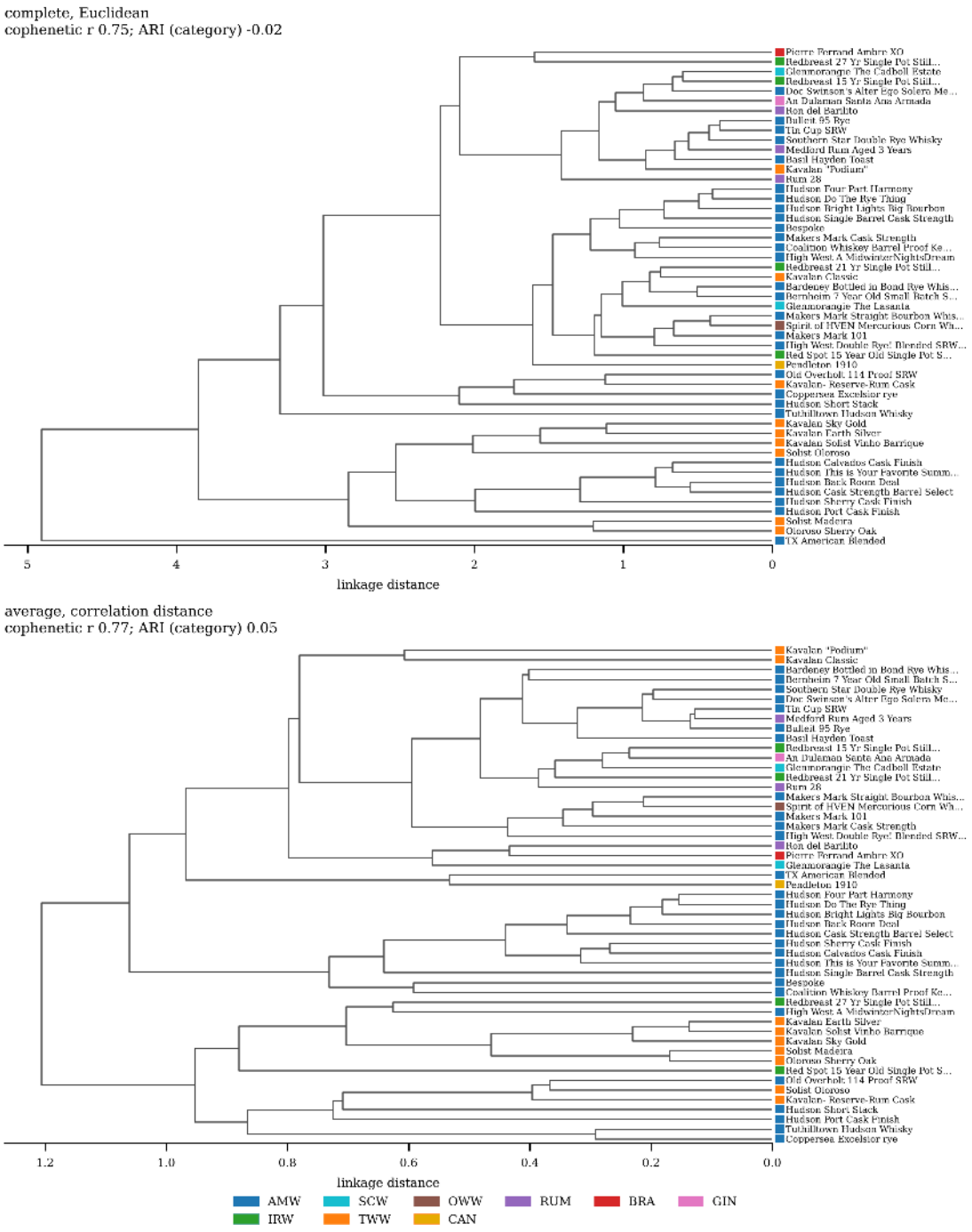

Figure S71. Block FUSED-DC (Table 20): Dendrograms by complete linkage on Euclidean distances (top); average linkage on correlation distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block FUSED-W: Figures S72–S76, enlarged from Figure 49

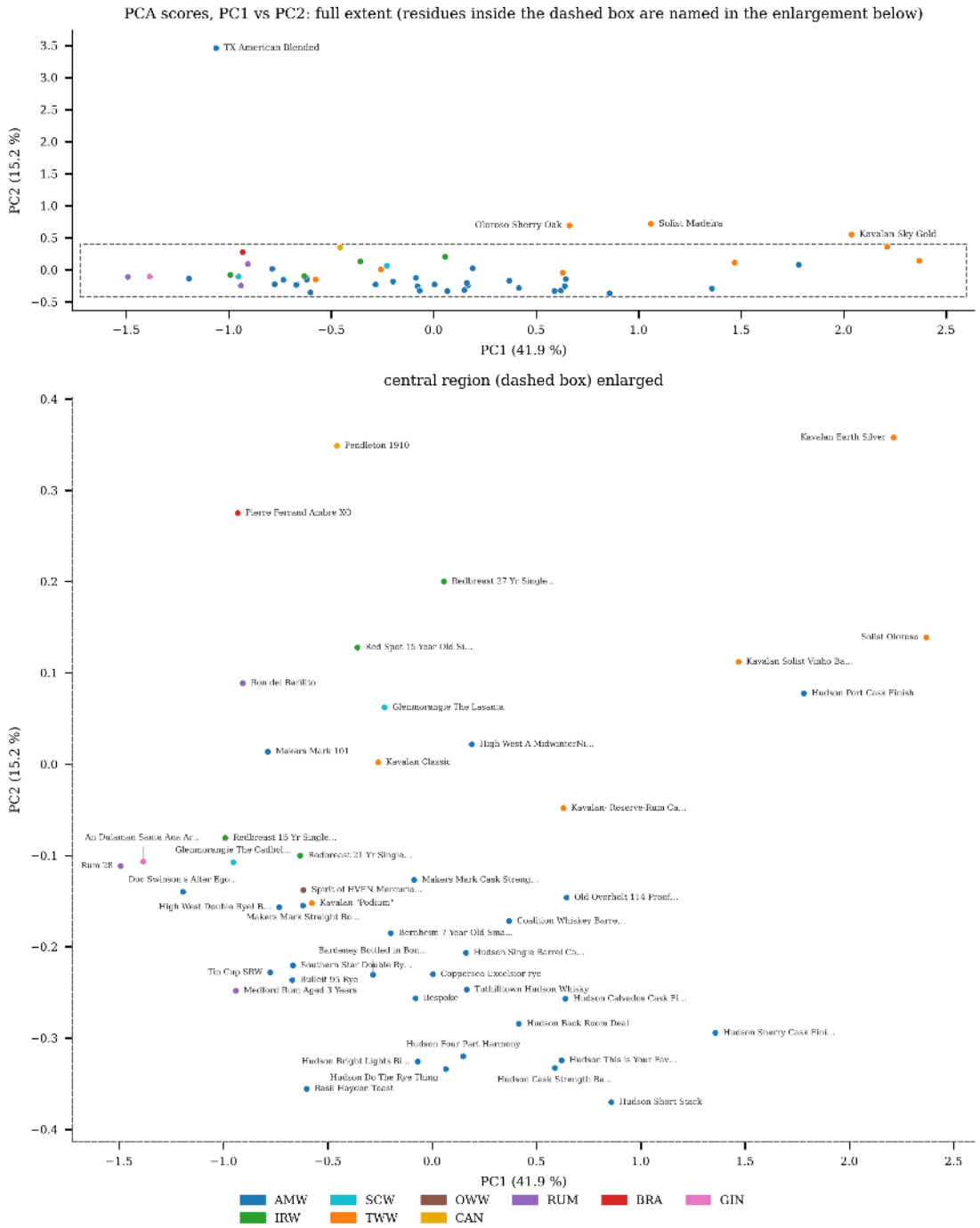

Figure S72. Block FUSED-W (Table 20): PCA scores, PC1 against PC2, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).



Figure S73. Block FUSED-W (Table 20): PCA scores, PC1 against PC3, every residue named. Where the residues crowd the center, the full extent is shown above, with the residues outside the dashed box named, and the region inside the box is enlarged below with its residues named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

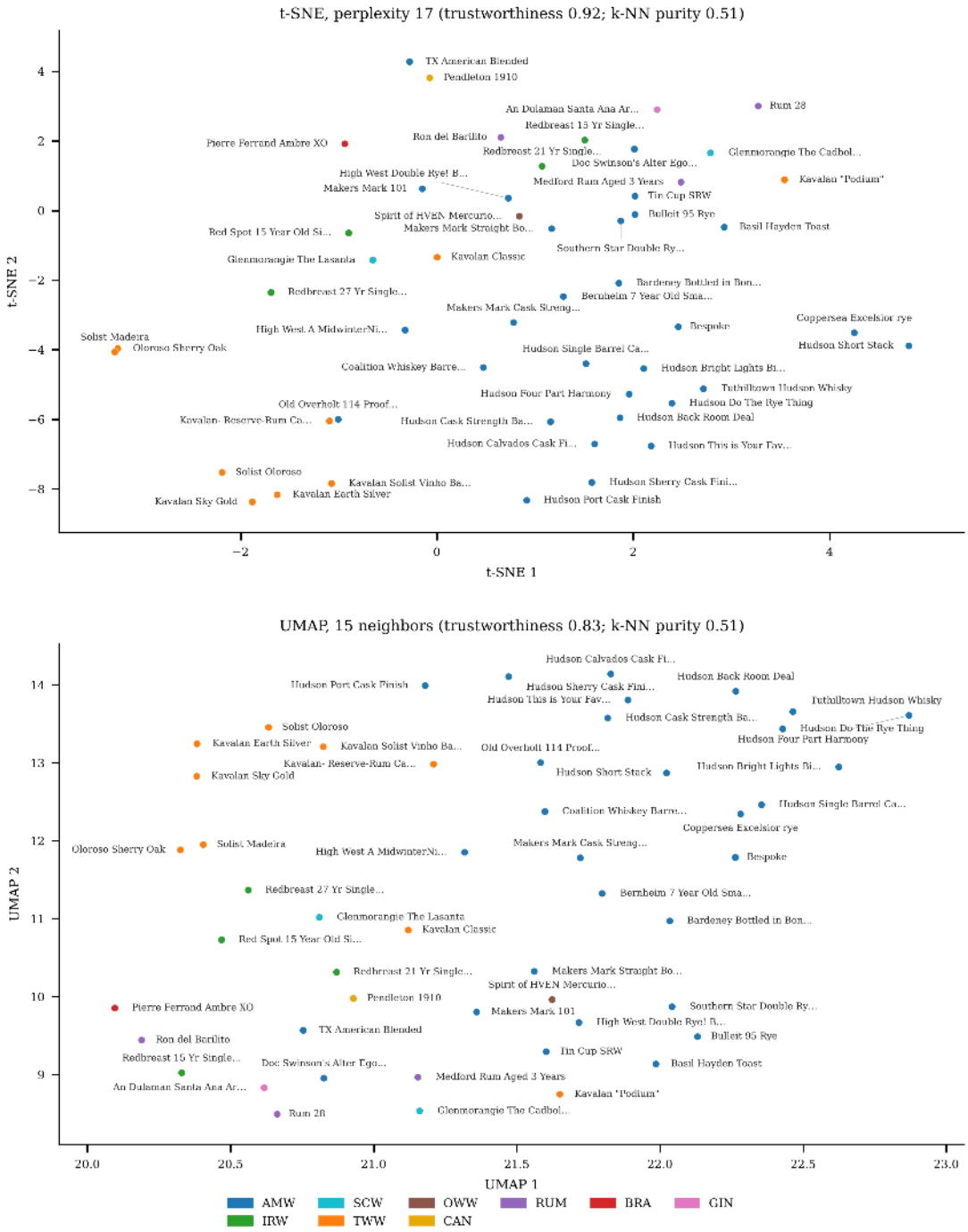

Figure S74. Block FUSED-W (Table 20): t-SNE, perplexity 17 (trustworthiness 0.92; k-NN purity 0.51) (top) and UMAP, 15 neighbors (trustworthiness 0.83; k-NN purity 0.51) (bottom), every residue named. Colors give the spirit category of each residue (abbreviations as in Figure 34).

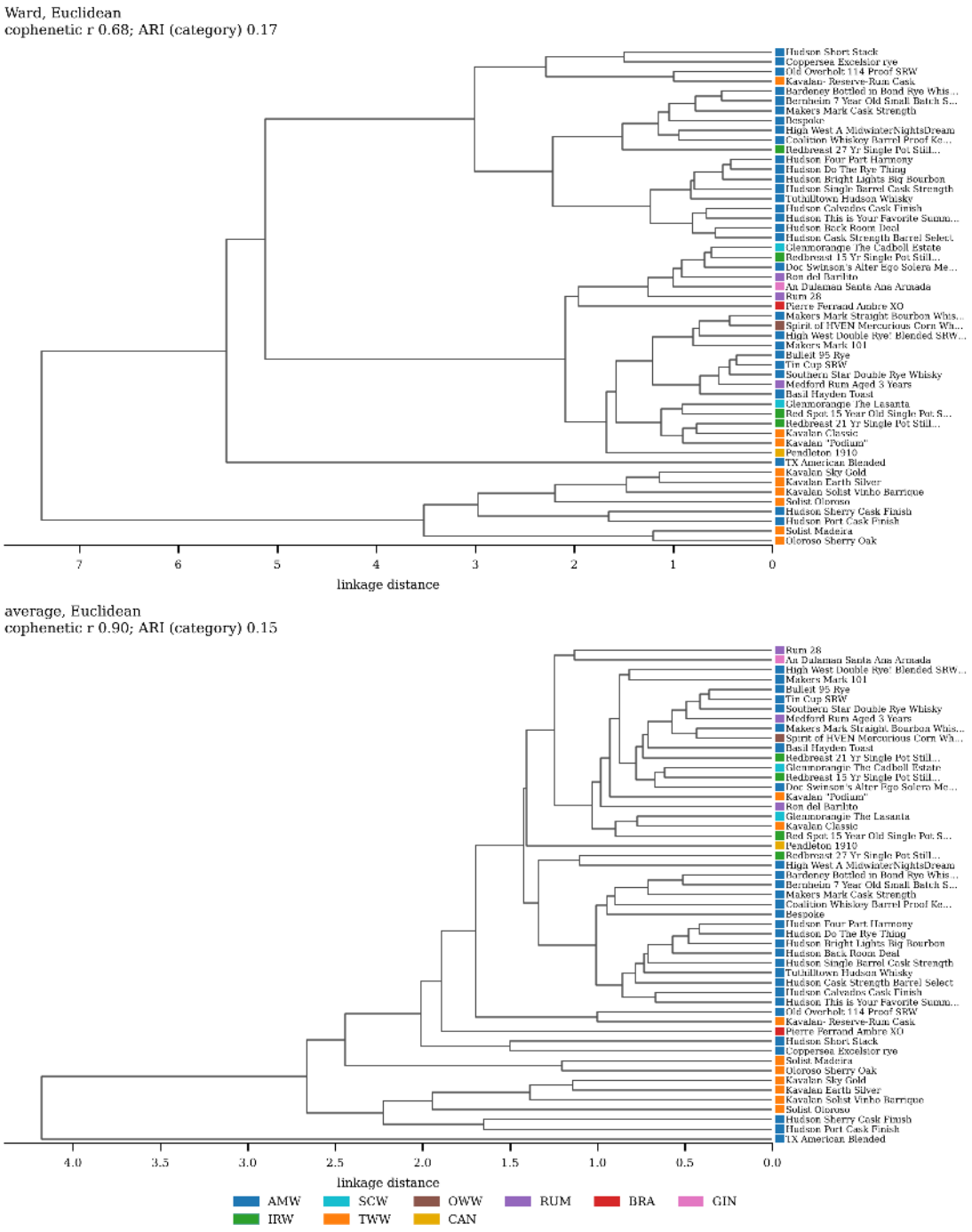

Figure S75. Block FUSED-W (Table 20): Dendrograms by Ward linkage on Euclidean distances (top); average linkage on Euclidean distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

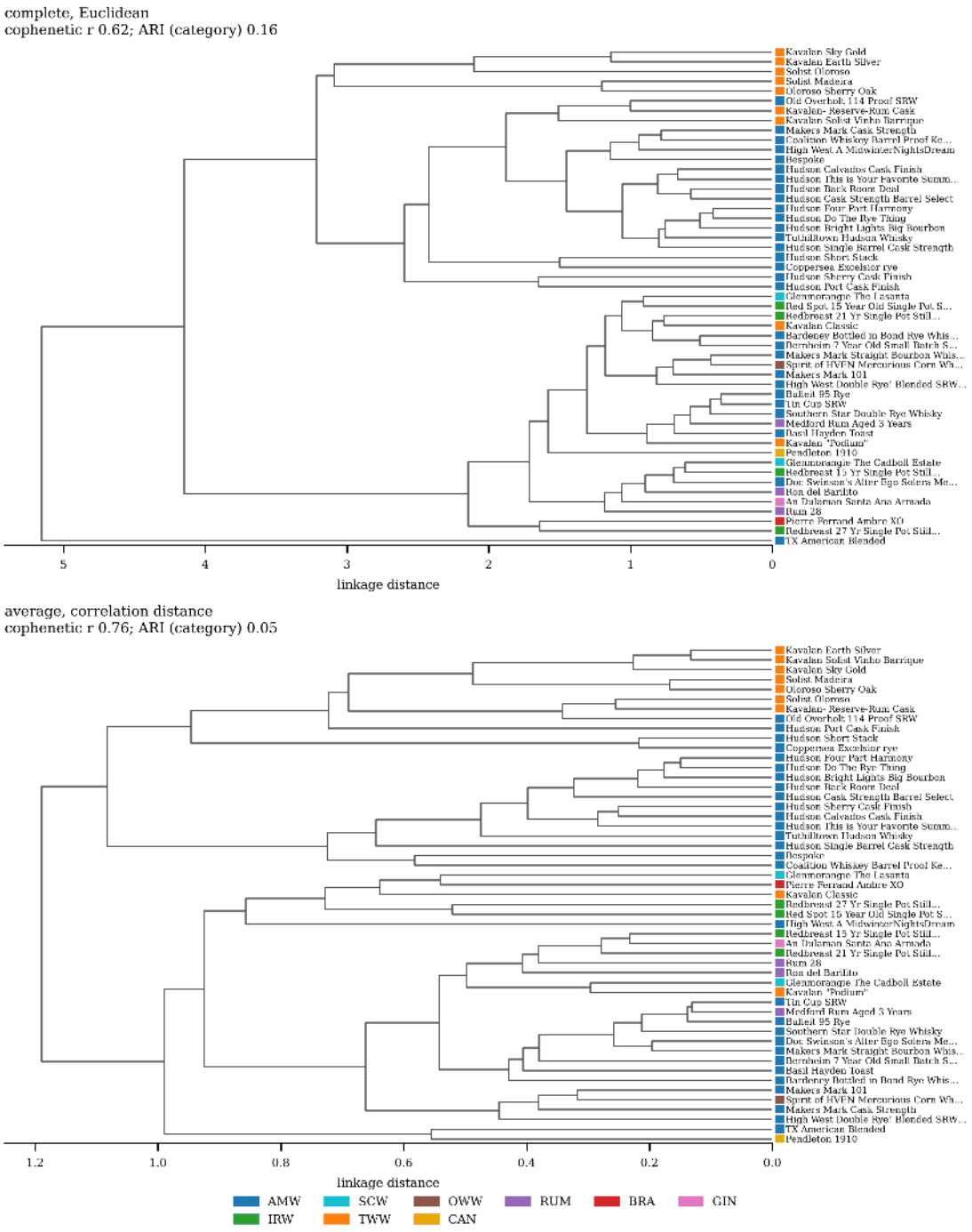

Figure S76. Block FUSED-W (Table 20): Dendrograms by complete linkage on Euclidean distances (top); average linkage on correlation distances (bottom), every residue named, with the spirit category of each residue (abbreviations as in Figure 34) in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-F: Figures S77–S79, enlarged from Figure 50

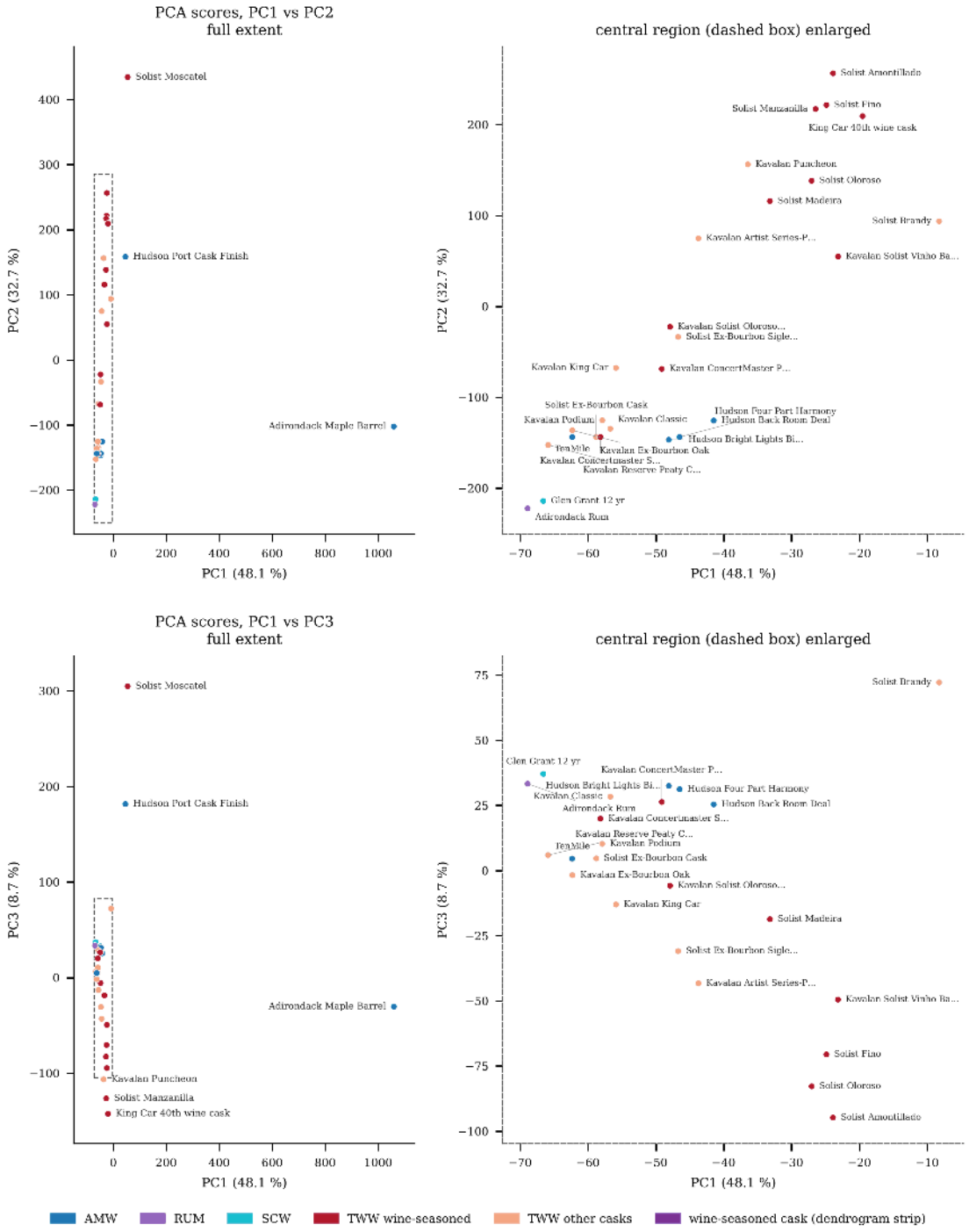

Figure S77. Block DMSO-F (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

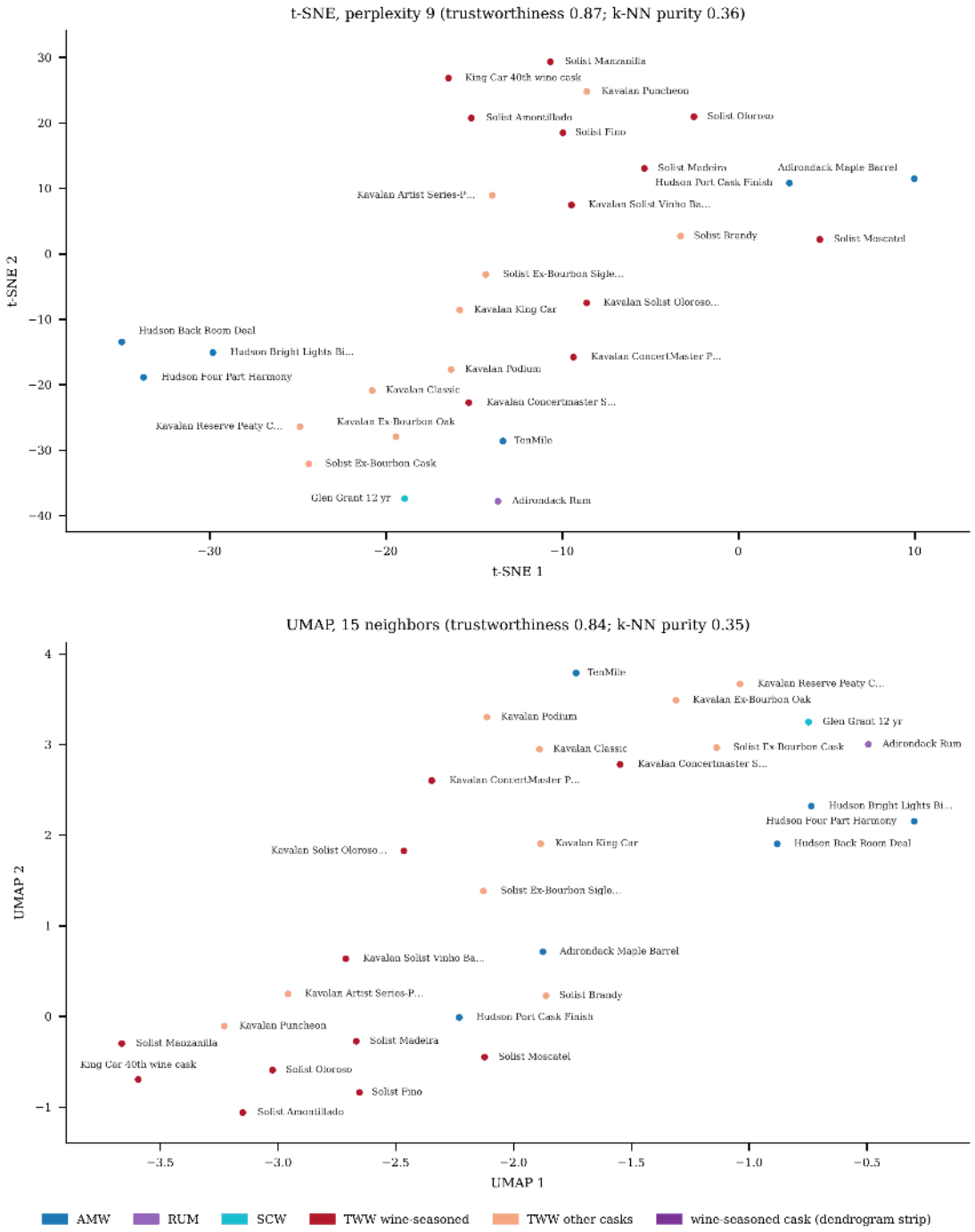

Figure S78. Block DMSO-F (Table 20): t-SNE, perplexity 9 (trustworthiness 0.87; k-NN purity 0.36) (top) and UMAP, 15 neighbors (trustworthiness 0.84; k-NN purity 0.35) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

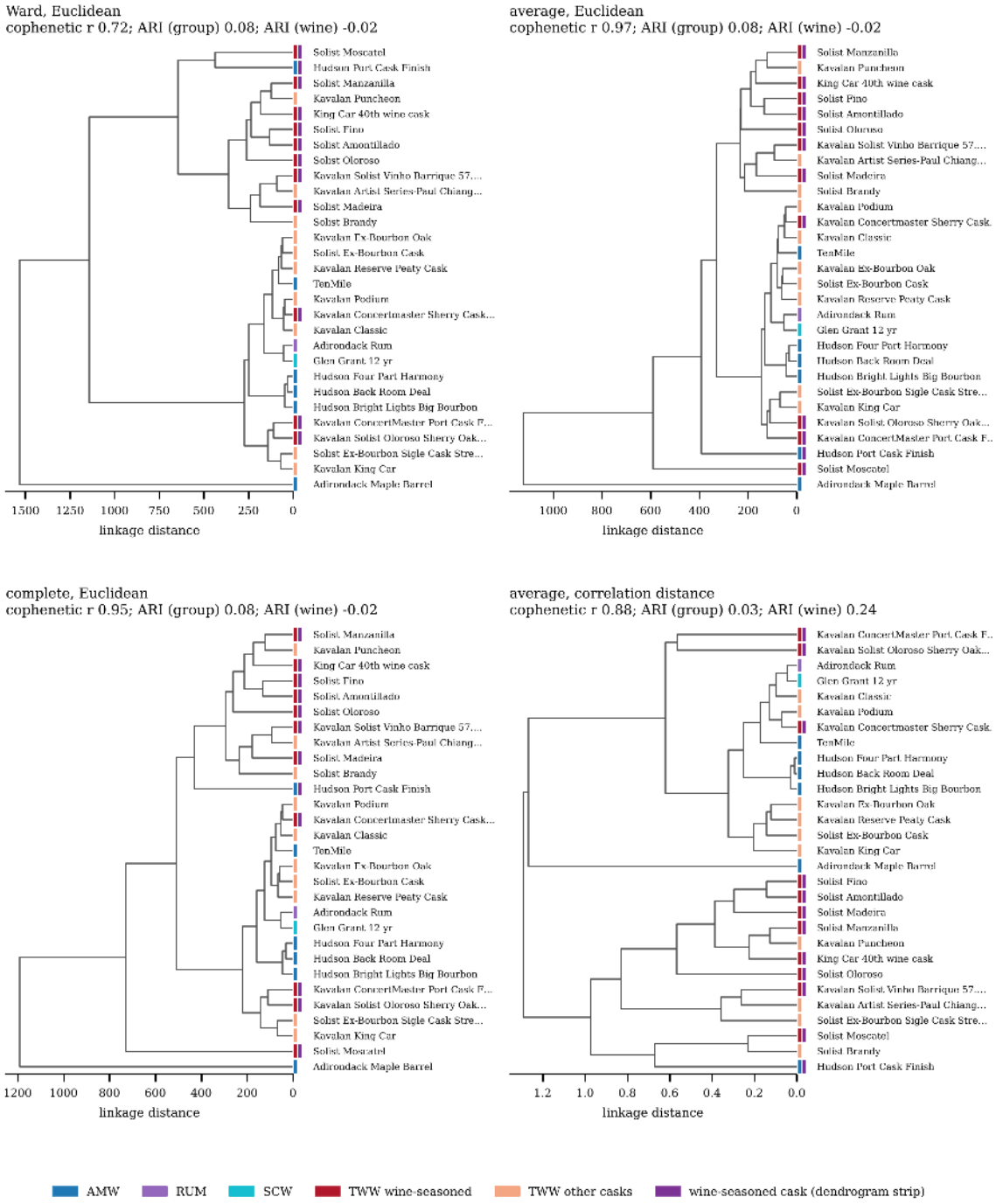

Figure S79. Block DMSO-F (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-Fx: Figures S80–S82, enlarged from Figure 51

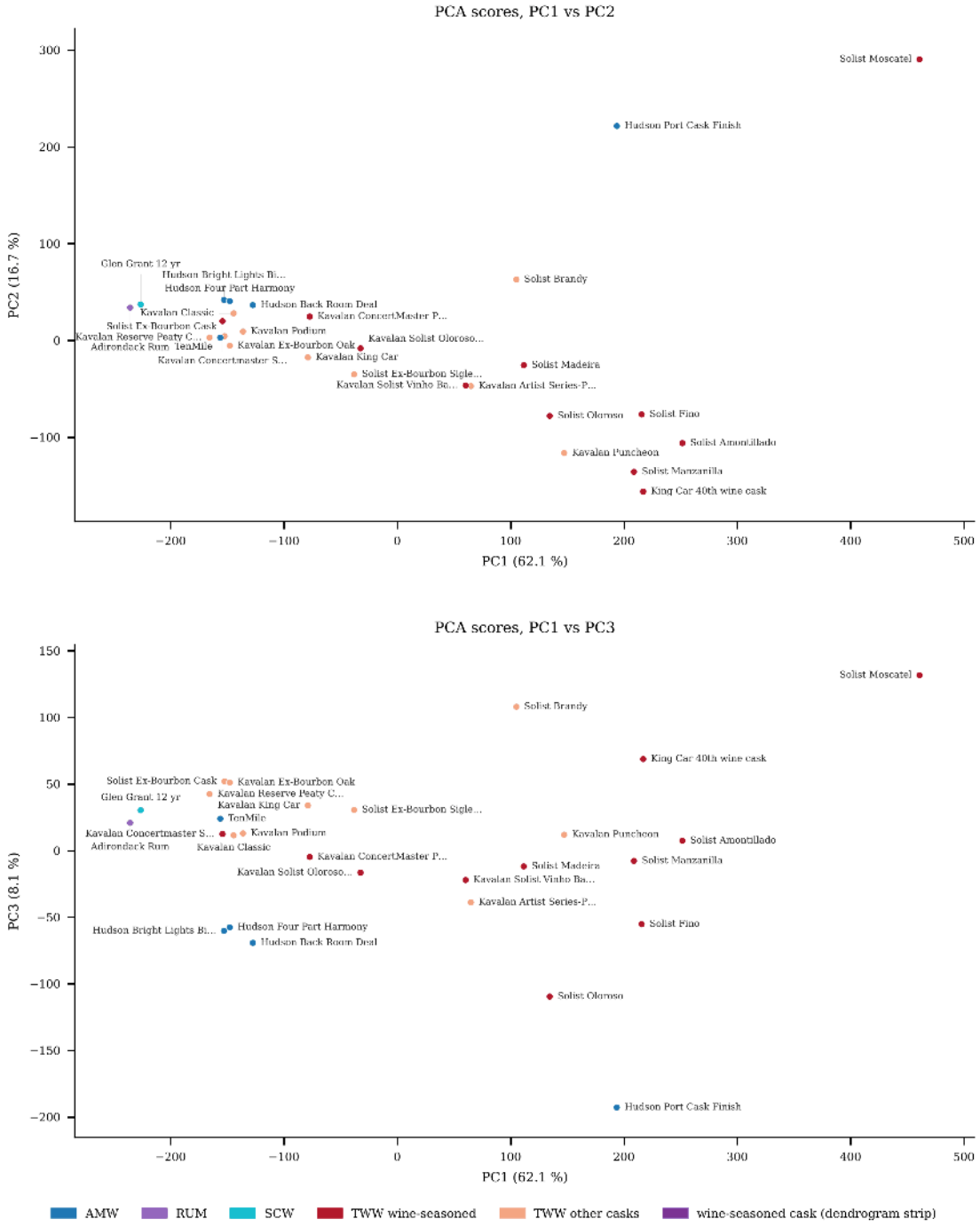

Figure S80. Block DMSO-Fx (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

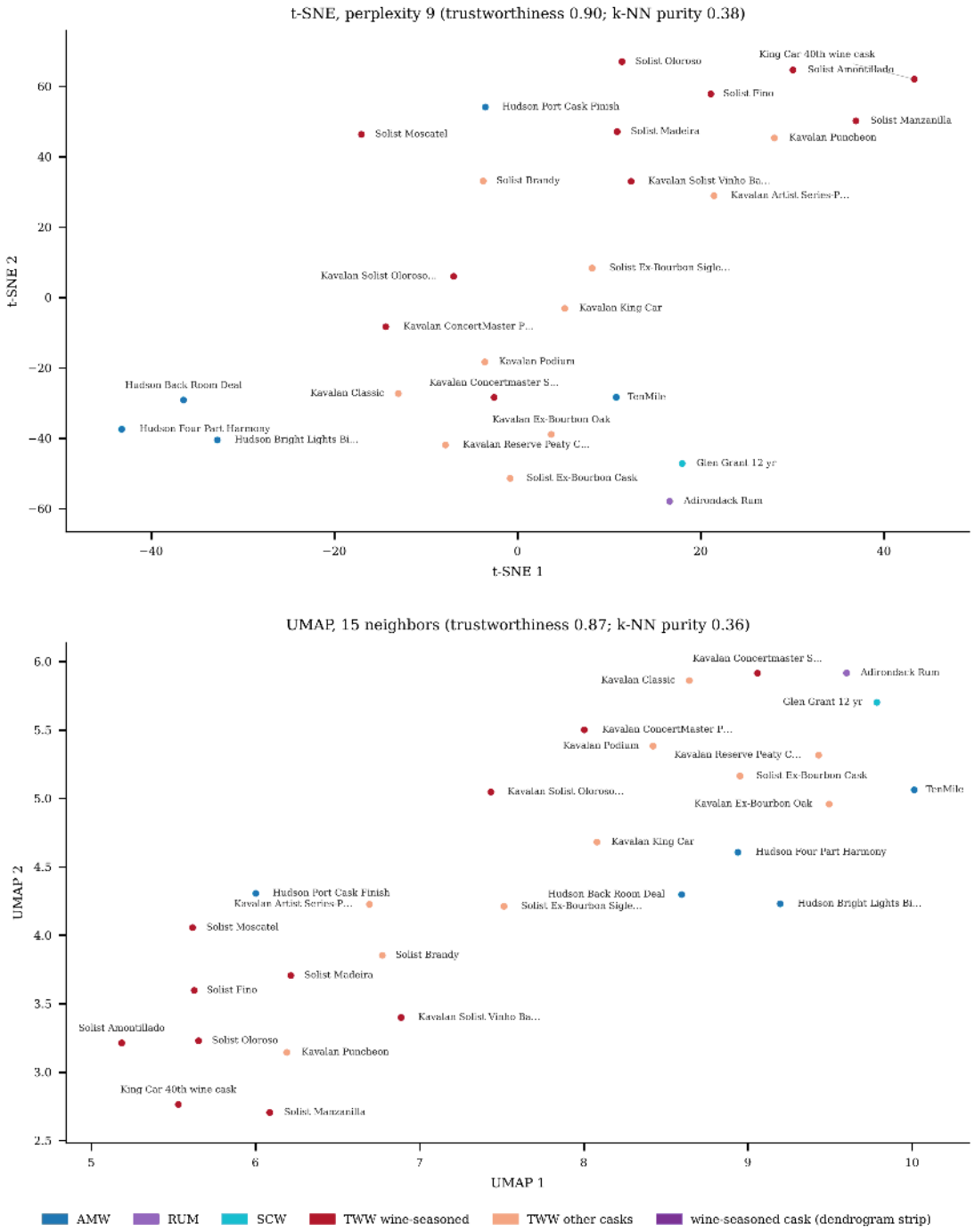

Figure S81. Block DMSO-Fx (Table 20): t-SNE, perplexity 9 (trustworthiness 0.90; k-NN purity 0.38) (top) and UMAP, 15 neighbors (trustworthiness 0.87; k-NN purity 0.36) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

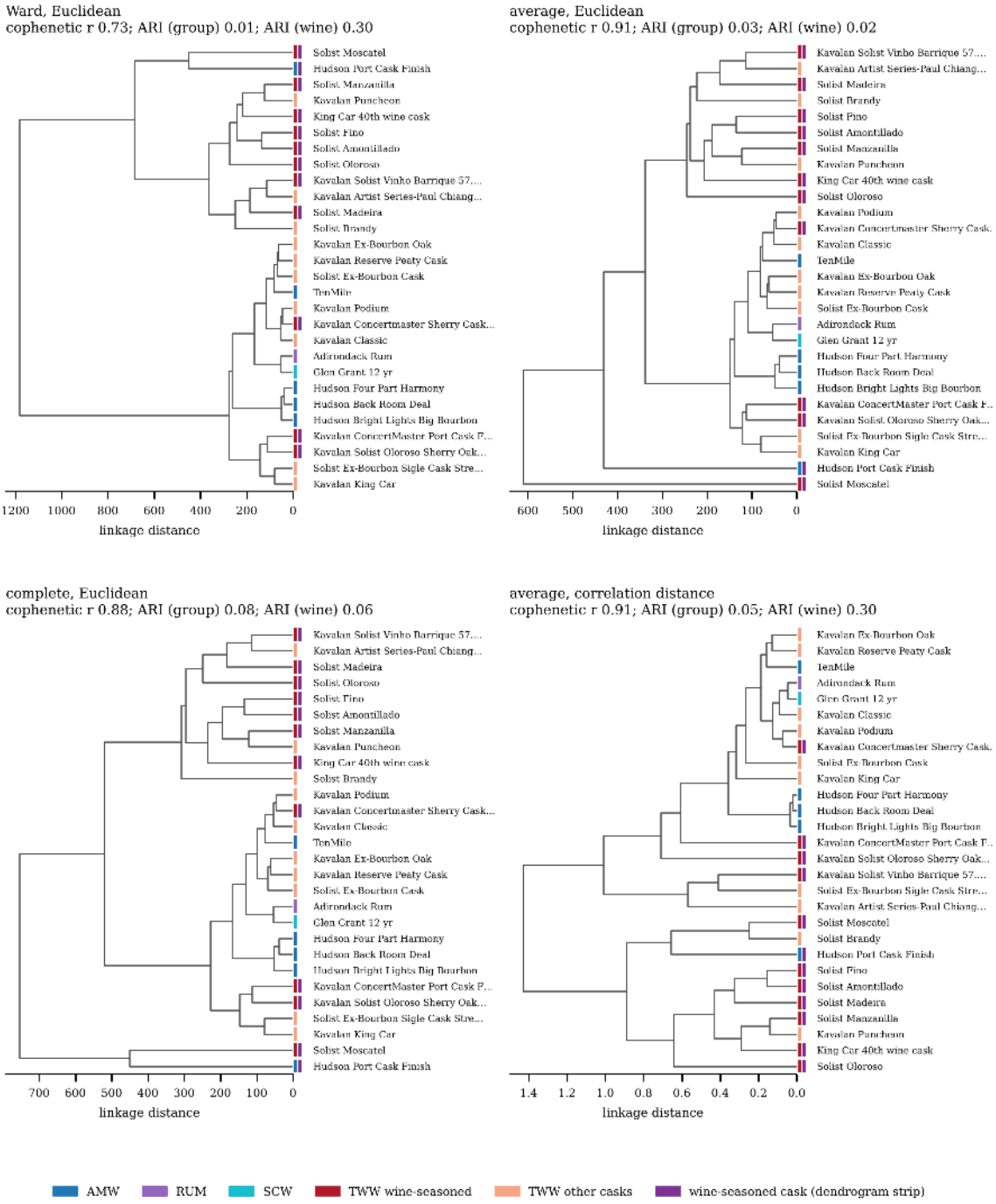

Figure S82. Block DMSO-Fx (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-NC: Figures S83–S85, enlarged from Figure 52

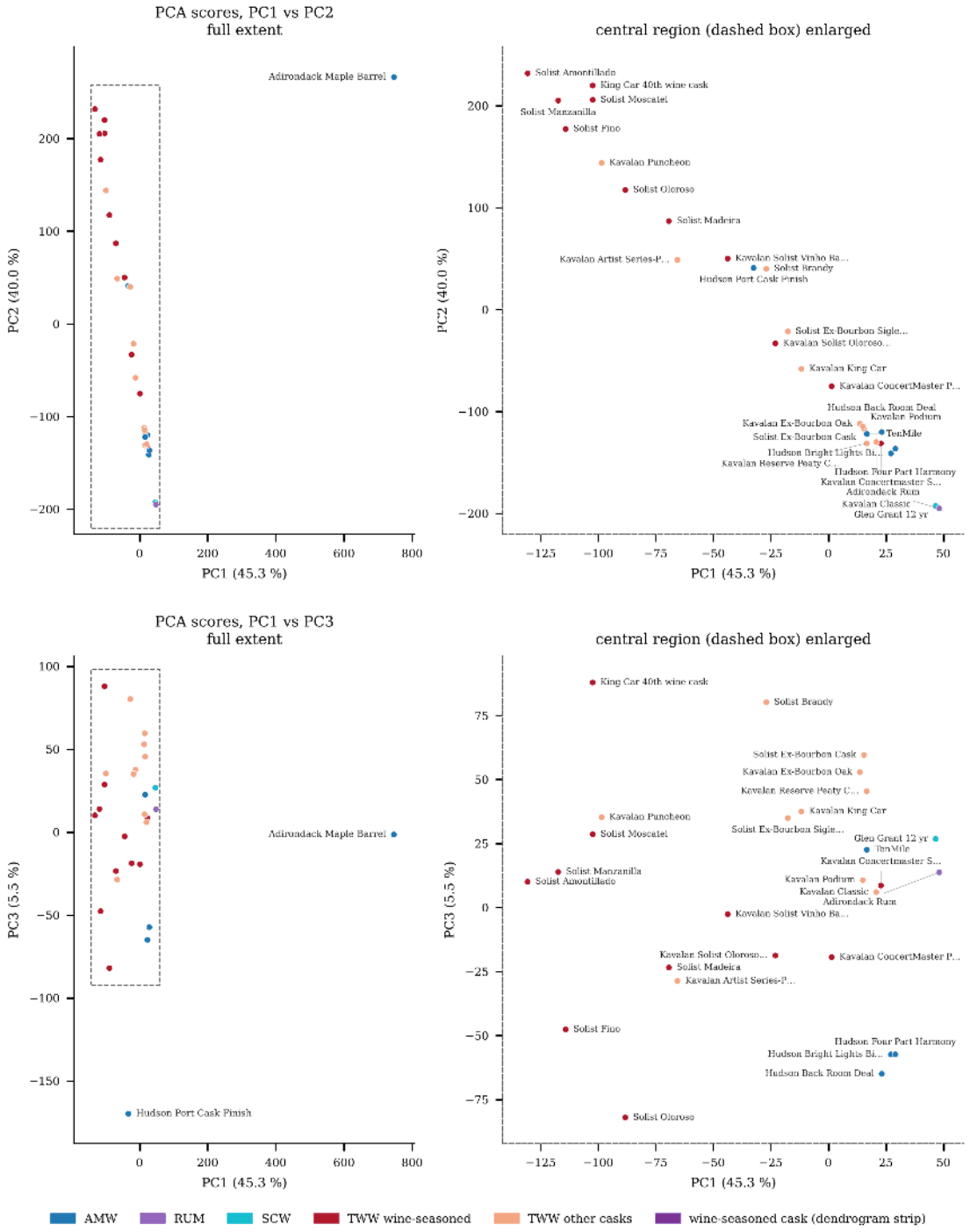

Figure S83. Block DMSO-NC (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

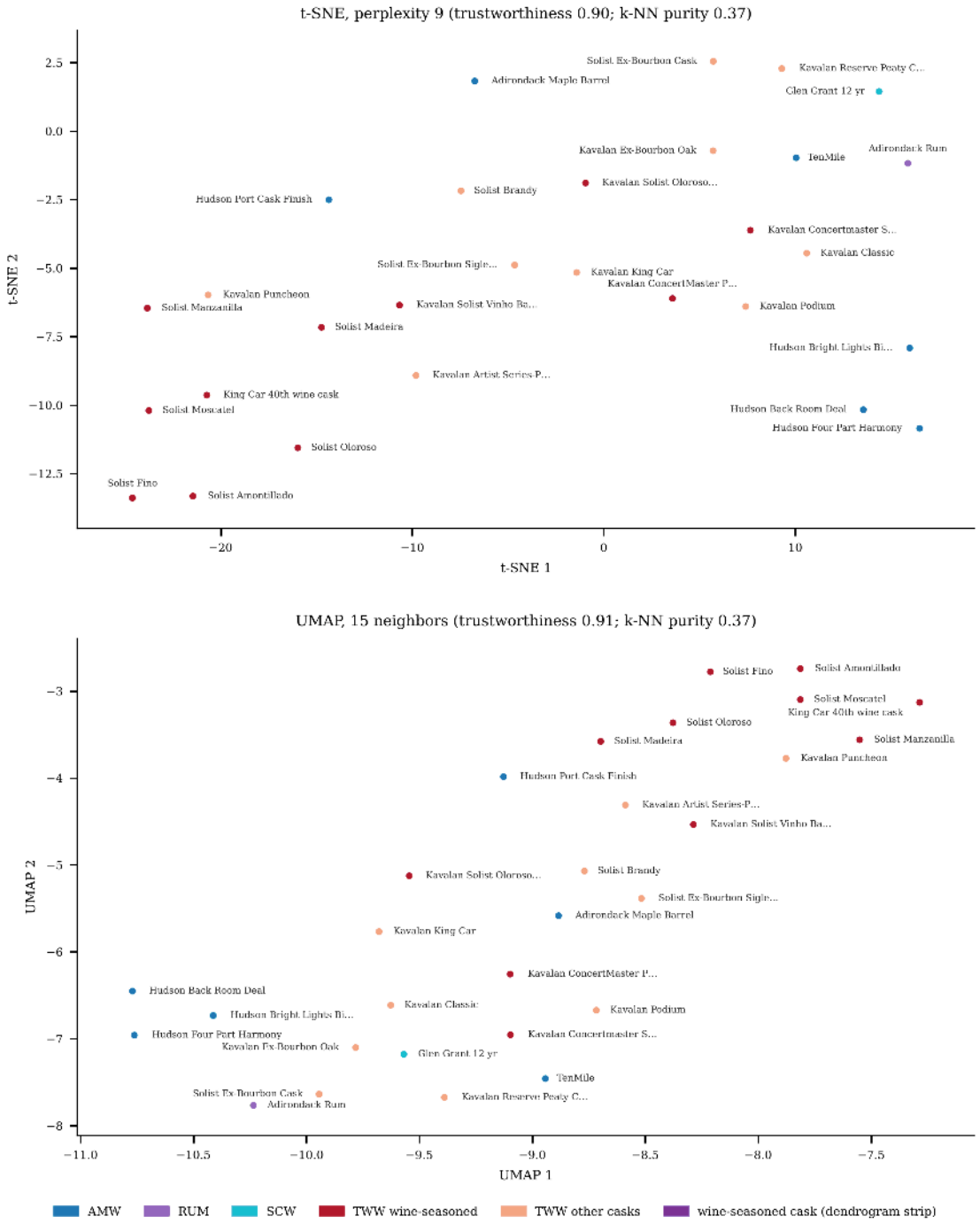

Figure S84. Block DMSO-NC (Table 20): t-SNE, perplexity 9 (trustworthiness 0.90; k-NN purity 0.37) (top) and UMAP, 15 neighbors (trustworthiness 0.91; k-NN purity 0.37) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

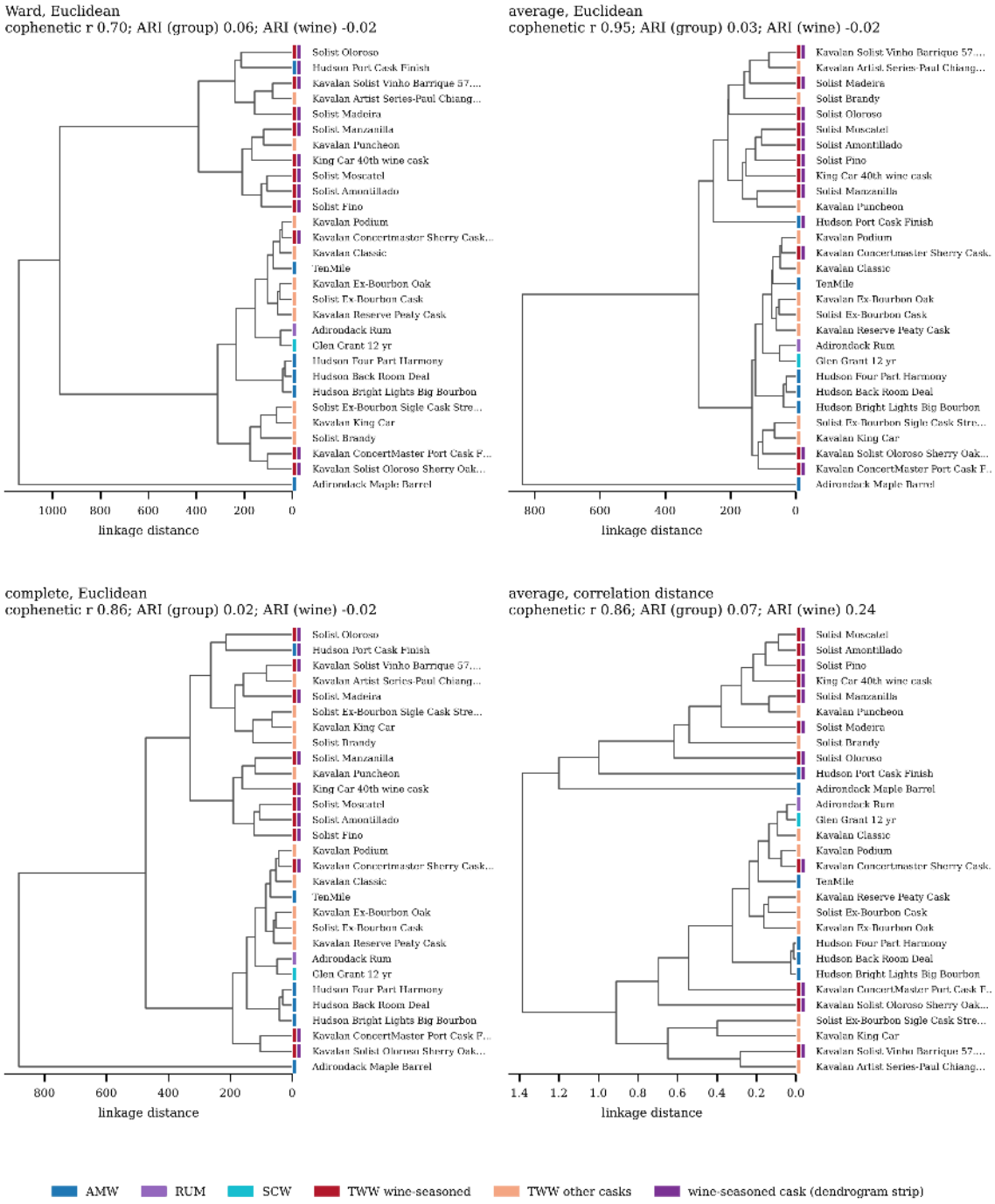

Figure S85. Block DMSO-NC (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-F04: Figures S86–S88, enlarged from Figure 53

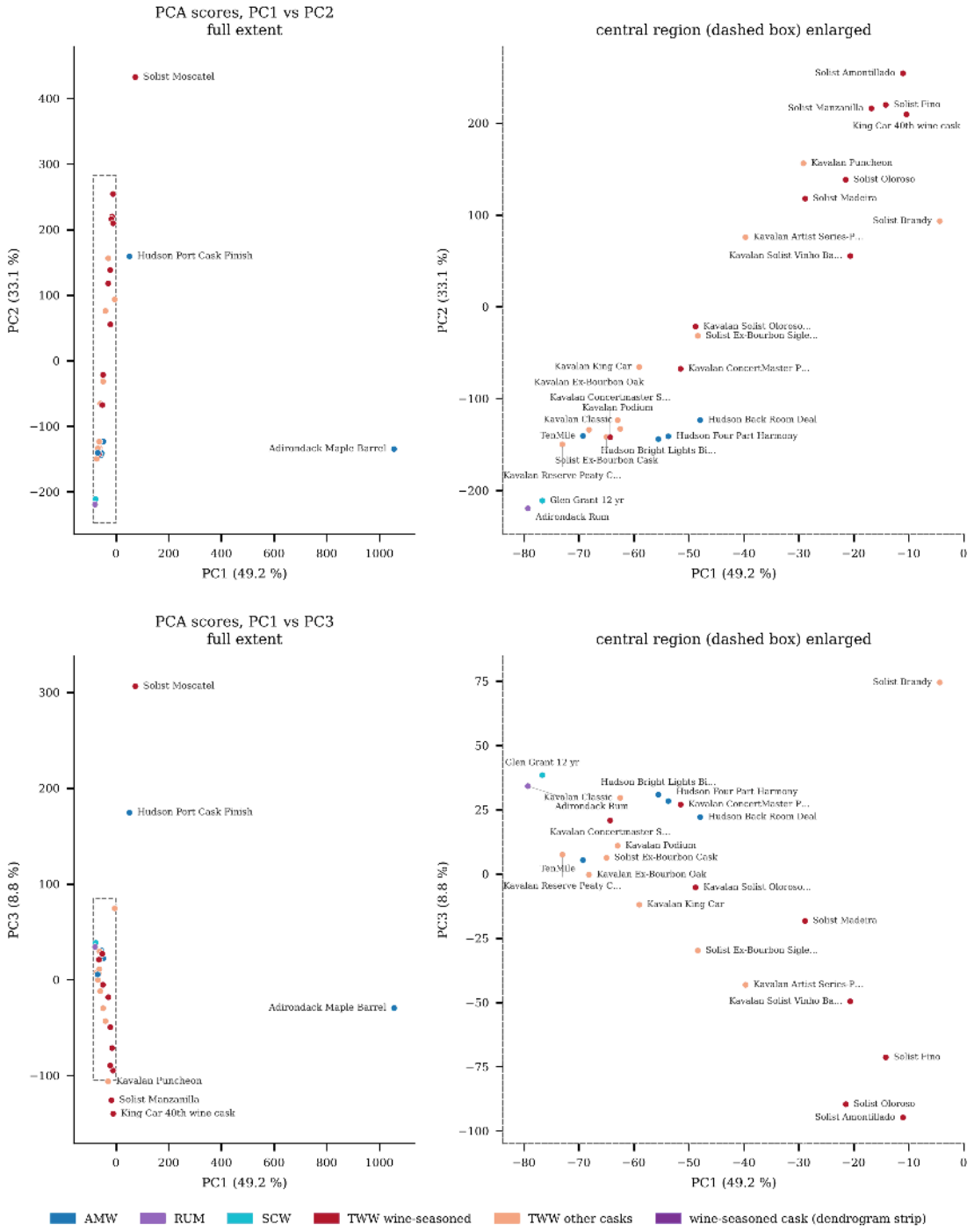

Figure S86. Block DMSO-F04 (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

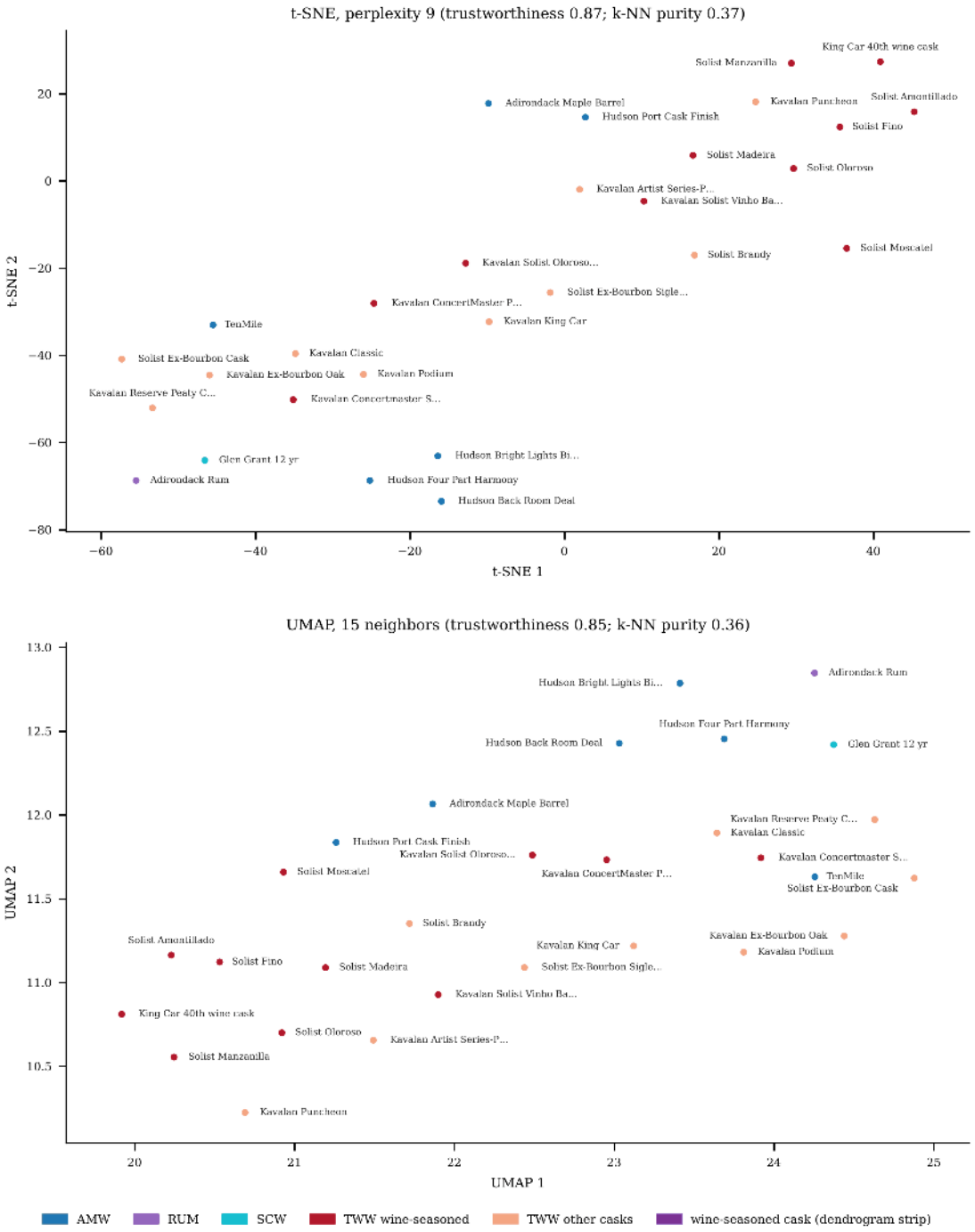

Figure S87. Block DMSO-F04 (Table 20): t-SNE, perplexity 9 (trustworthiness 0.87; k-NN purity 0.37) (top) and UMAP, 15 neighbors (trustworthiness 0.85; k-NN purity 0.36) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

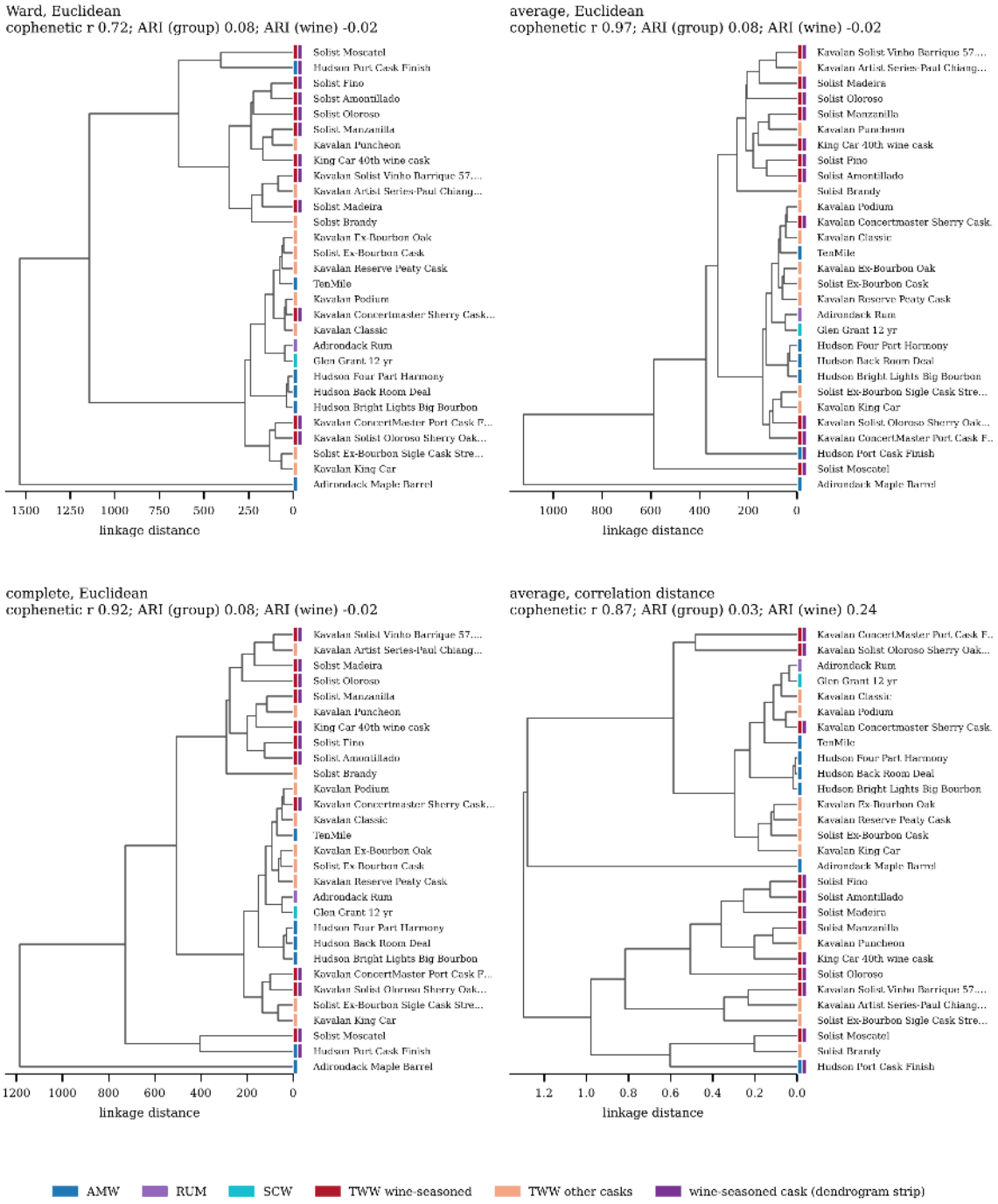

Figure S88. Block DMSO-F04 (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-Qp: Figures S89–S91, enlarged from Figure 54

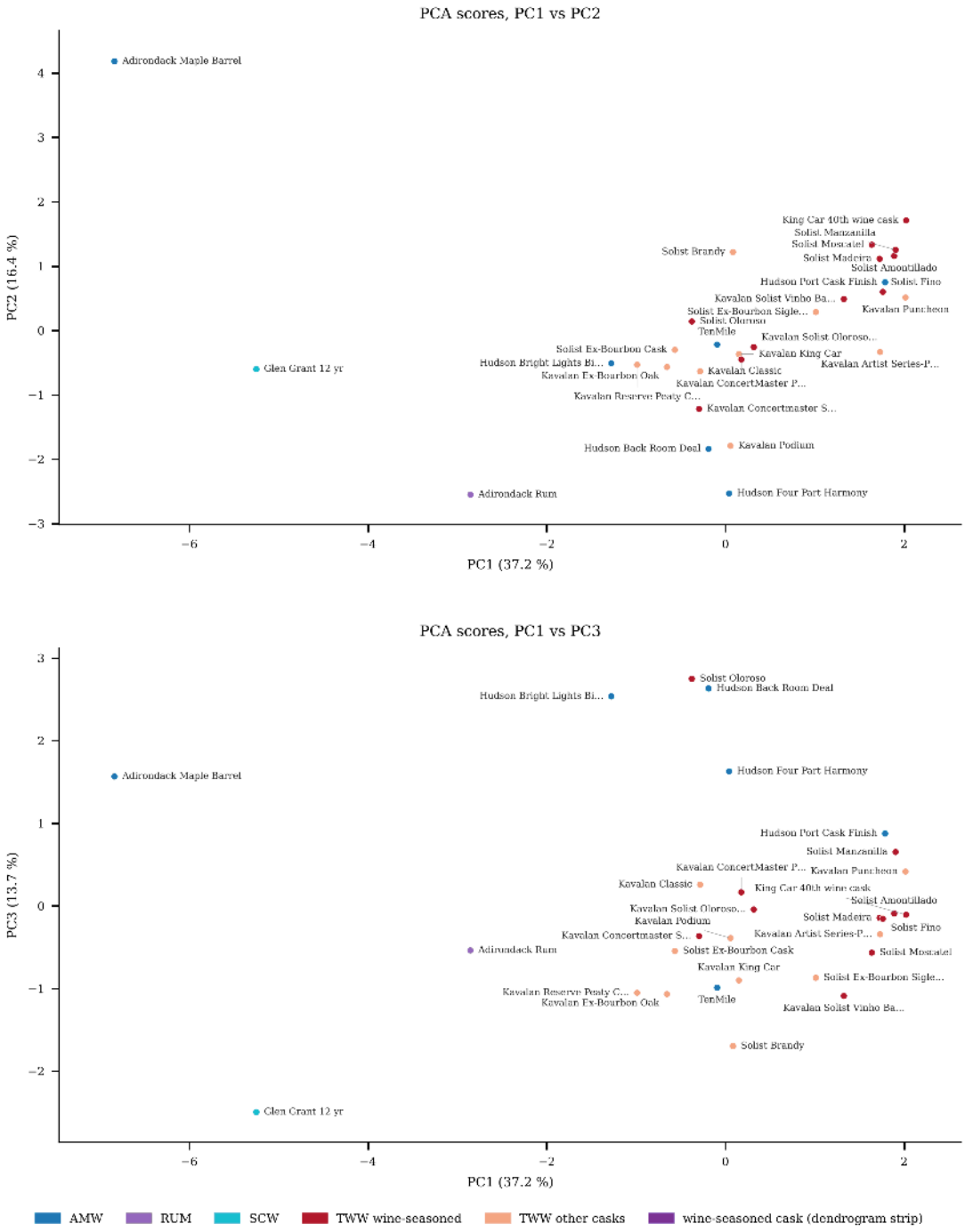

Figure S89. Block DMSO-Qp (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

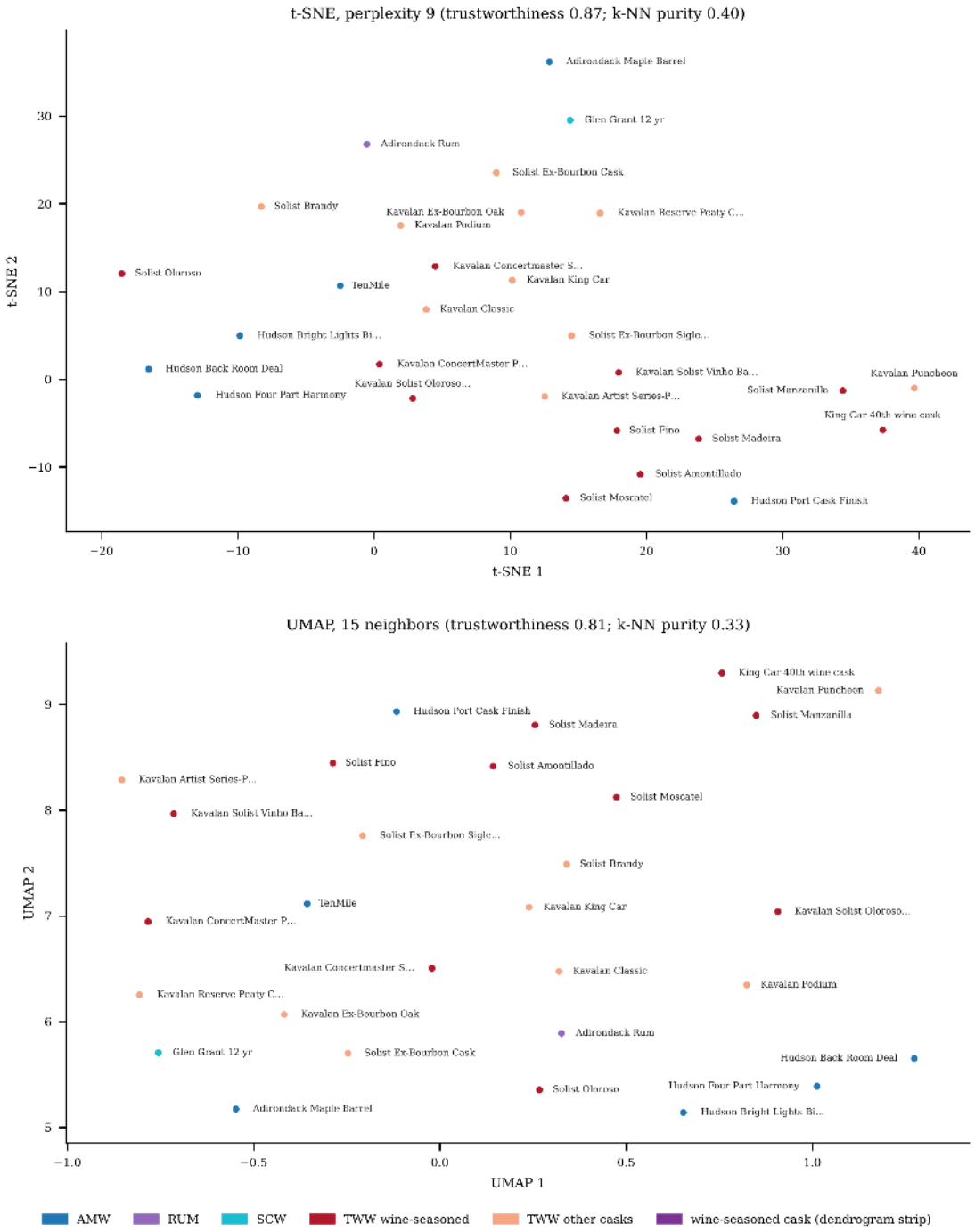

Figure S90. Block DMSO-Qp (Table 20): t-SNE, perplexity 9 (trustworthiness 0.87; k-NN purity 0.40) (top) and UMAP, 15 neighbors (trustworthiness 0.81; k-NN purity 0.33) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

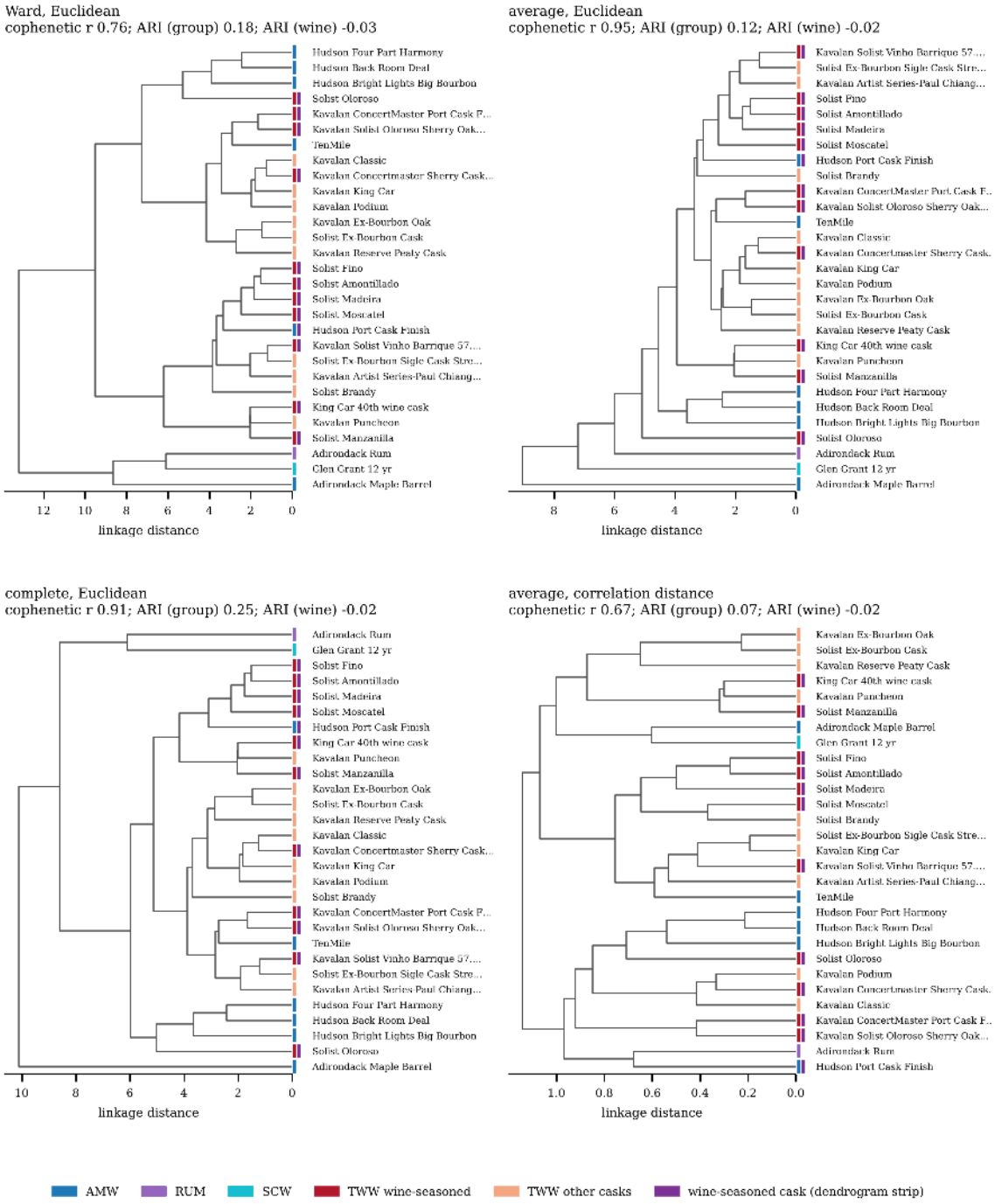

Figure S91. Block DMSO-Qp (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.

Block DMSO-Qa: Figures S92–S94, enlarged from Figure 55

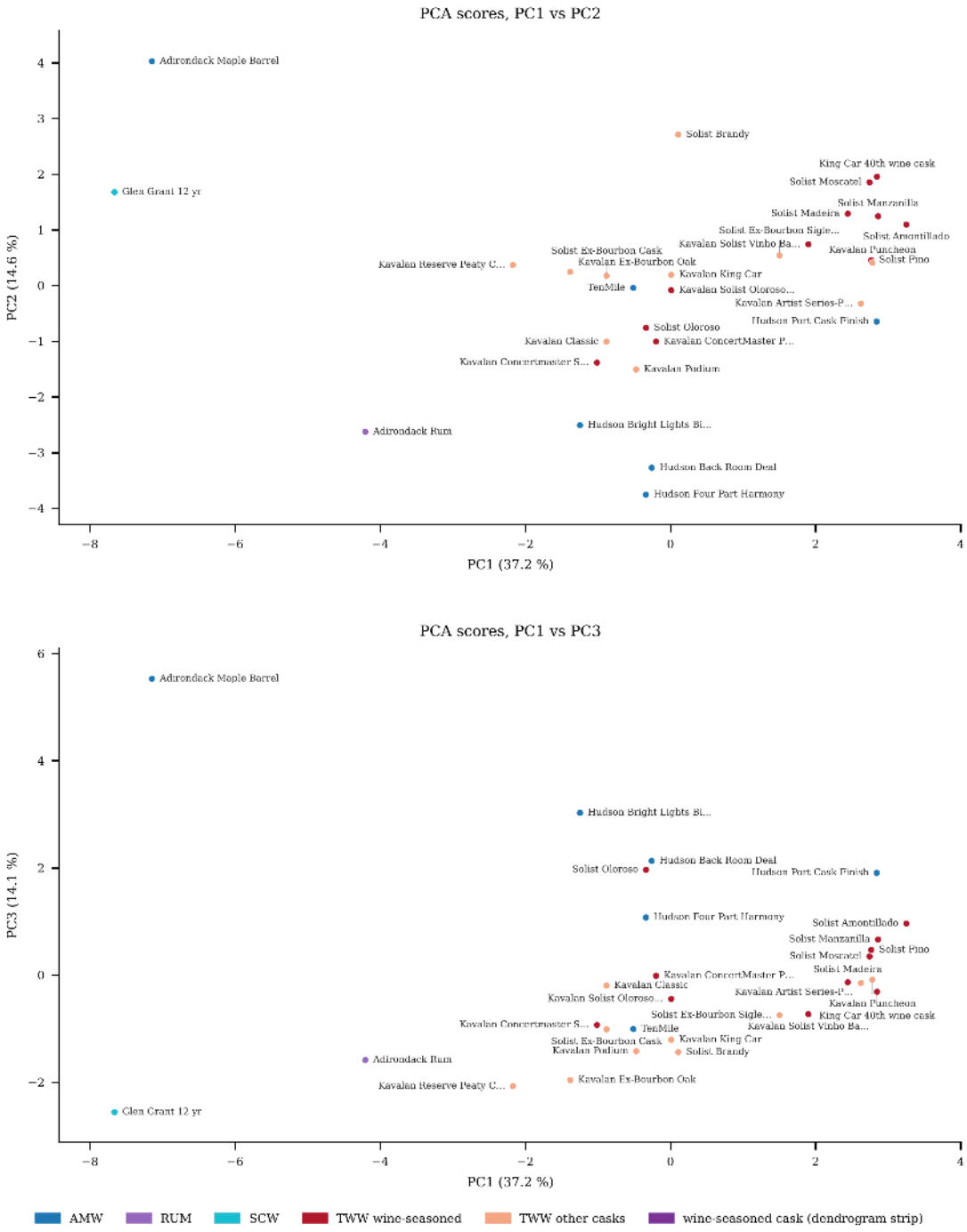

Figure S92. Block DMSO-Qa (Table 20): PCA scores, PC1 against PC2 (top) and PCA scores, PC1 against PC3 (bottom), every residue named; where the residues crowd the center, the region inside the dashed box is enlarged beside the full extent. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

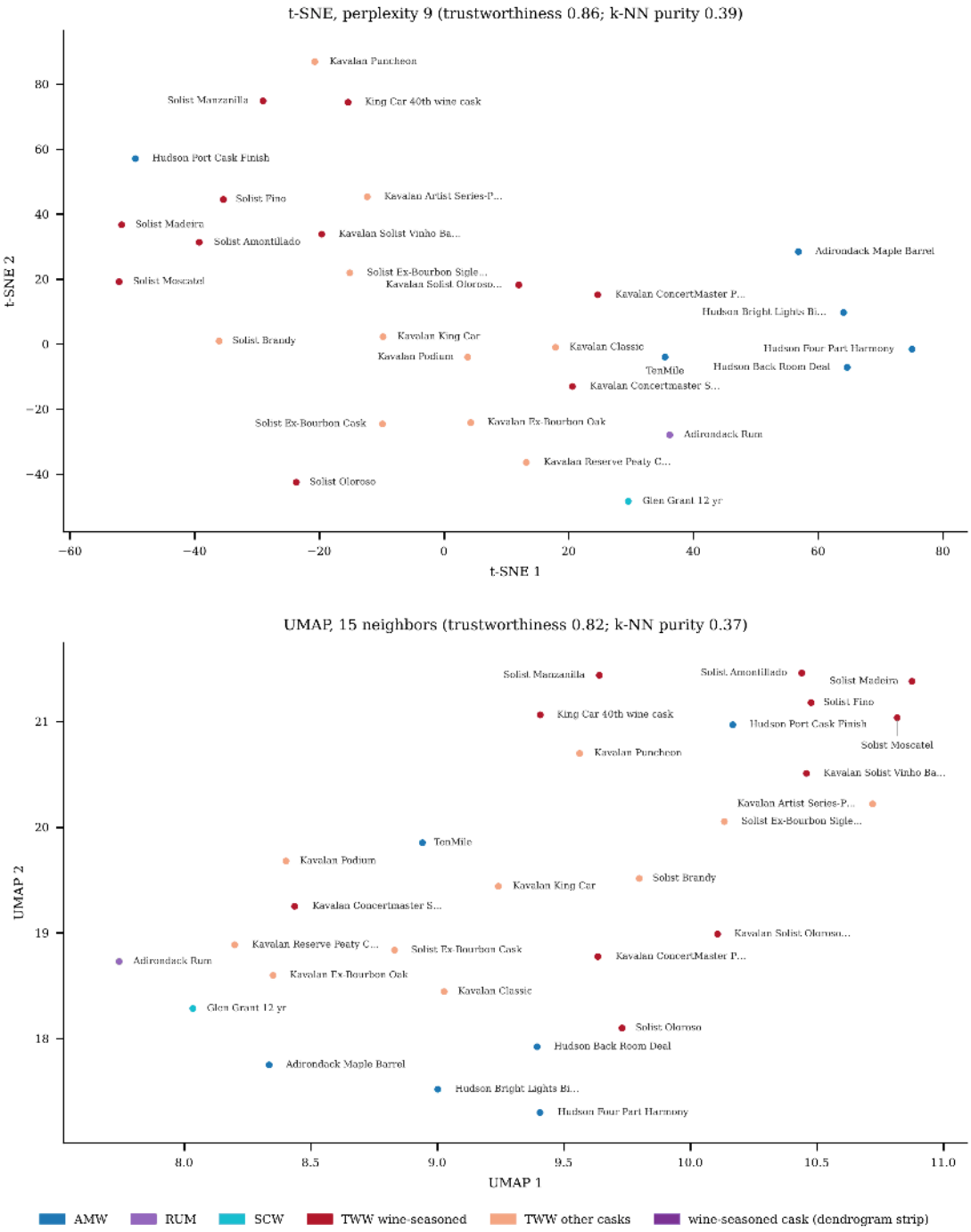

Figure S93. Block DMSO-Qa (Table 20): t-SNE, perplexity 9 (trustworthiness 0.86; k-NN purity 0.39) (top) and UMAP, 15 neighbors (trustworthiness 0.82; k-NN purity 0.37) (bottom), every residue named. Colors give the residue group of each residue, with wine-seasoned casks marked in a second strip.

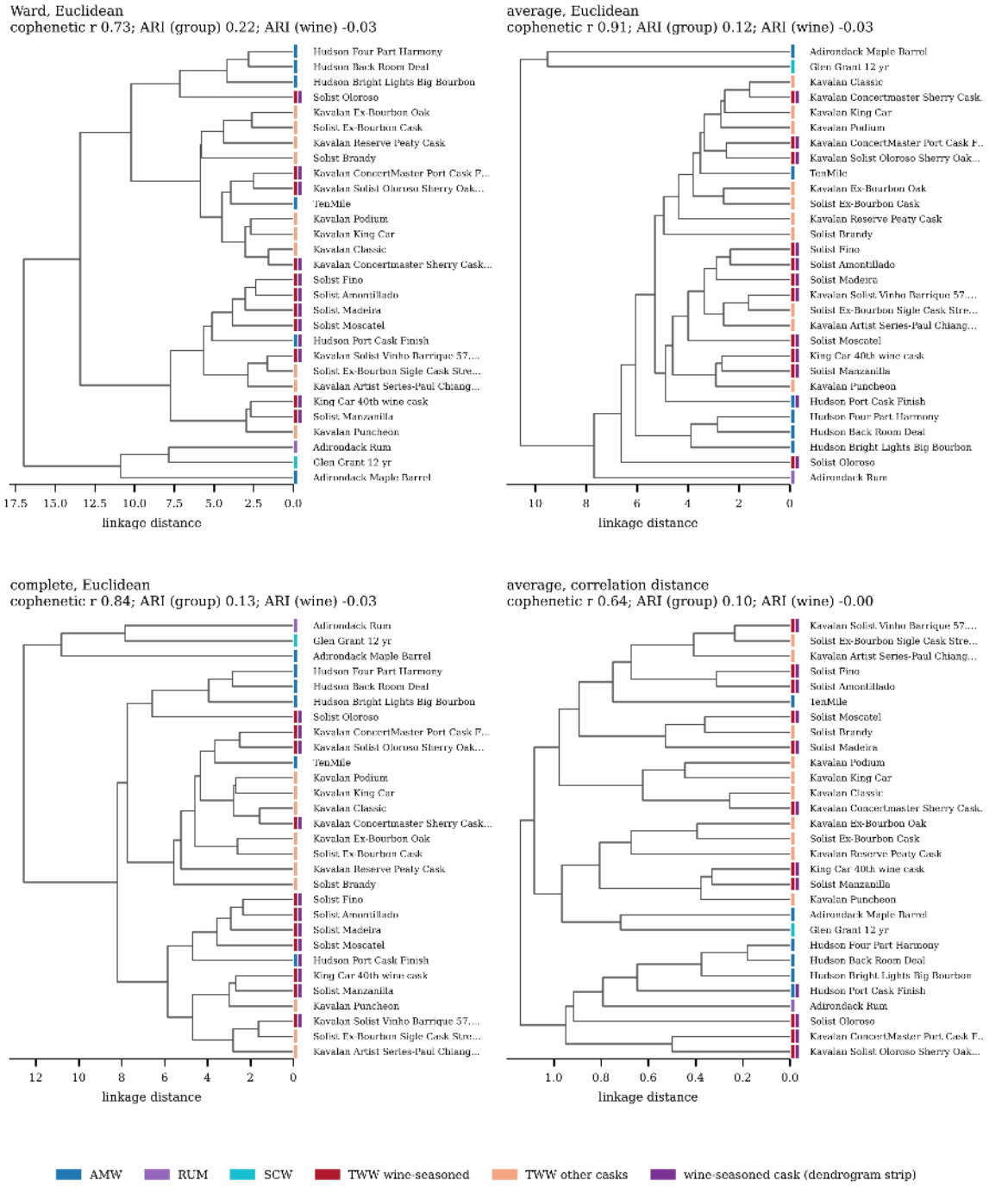

Figure S94. Block DMSO-Qa (Table 20): Dendrograms by Ward linkage on Euclidean distances (top left); average linkage on Euclidean distances (top right); complete linkage on Euclidean distances (bottom left); average linkage on correlation distances (bottom right), every residue named, with the residue group of each residue, with wine-seasoned casks marked in a second strip in the strip beside the leaves; the cophenetic correlation and the adjusted Rand index are given in each title.