Supplementary Data Assi et al.

S1. Analysis of blank test set

A blank test set was used to investigate the selectivity of the instruments' inbuilt algorithms for the substances of interest (2AI, DXM and LID). The blank test set was binary and ternary mixtures of common excipients and adulterants (Table S1), in which no signature was stored for the mixture itself. It was also important that the blank test set was made not to contain any of the substances of interest which were 2-AI, DXM, LID and CAF. The blank test set was run and the signature(s), most consistent with the sample analysed, selected by the instruments' algorithms were compared (Table S2).

Using the microPHAZIR (NIR) CWS in-built identification, the test set was not consistent with any signature(s) in the library using this algorithm. It gave *r* values below 0.95 for all the blank test set mixtures (T1 - T8) against the pure substances and 50:50 mixtures stored in the instrument library. The *r* values obtained for the blank test set ranged from 0.5105 - 0.9256 which were observed for T8 (BEN/PRO/MCC) and T4 (MCC/TAL) against BEN and MCC, respectively. The NIR algorithm reports only the signature most consistent with the sample analysed and does not provide a combination of signatures like the Raman and ATR-FT-IR.

The TruScan RM (Raman) probability identification algorithm gave slightly different results to the microPHAZIR CWS algorithm. This was because Raman, unlike NIR, detects primarily the chemical makeup of a sample. In addition, the probability algorithm reported correlations indicating how similar was the molecular signal of the test substance to the signature(s) in the instrument library. No mismatches were made to the signatures of 2AI, DXM, and LID. In most cases, the TruScan RM probability algorithm correlated with the signature of one or two constituents in the mixtures, in

discovery mode. In three cases, a 'similar item' or no match was observed and included T3 (TAL/PRO), T5 (PAR/MCC) and T7 (MCC/TAL/PAR). T3 gave a 'similar item' to the PRO signature, while T5 and T7 gave no match. The TruScan RM can indicate a 'similar item' if a test spectrum was not consistent with, but 'similar' to a library signature. Most mixtures containing BEN, including T1 (BEN/TAL), T2 (BEN/PRO), and T8 (BEN/PRO/MCC), were consistent with the BEN signature as well as the signatures of other constituents in each mixture (Table S2). Only one mixture, T4 (MCC/TAL), gave a 100% correlation with MCC. This might be because TAL has low Raman scattering and it was present in only 25% m/m of the mixture.

For the TruDefender FT-IR probability identification algorithm, no mismatches were made to signatures from the analytes of interest (2AI, DXM, and LID). Using this algorithm, T3 (TAL/PRO) was consistent with TAL (62%), but mismatched MCC (12%). In the case of T4 (MCC/TAL), no % correlation was observed as both MCC and TAL have one predominant Raman band around 1000 cm^{-1}; these bands are broad and overlapping reducing the ability to distinguish them at this dilution ratio. The remaining binary mixtures were correlated to their individual constituents and were T1 (BEN/PAR), T2 (BEN/PRO) and T5 (PAR/MCC). The ternary mixtures, T6 (PAR/BEN/PRO) and T8 (BEN/PRO/MCC), were both correlated to PRO and BEN. On the other hand, the ternary mixture T7 (TAL/MCC/PAR) correlated to TAL.

S2. Analysis of pure substances

The pure substances correlated to their NIR signature far above the threshold (0.95) with a minimum *r* value of 0.9944 observed for BEN (Table S3). The correlations were slightly lower for the 50:50 mixtures of Mixture 1 (2AI/CAF), Mixture 2

(DXM/CAF) and Mixture 3 (LID/CAF), which gave the highest *r* values for their own signatures at 0.9644, 0.9874 and 0.9589, respectively.

Pure substances, except LID, were consistent with their corresponding Raman signatures with a minimum PVAL of 0.2279 observed for DXM (Table S3). Although the LID spectrum had a number of distinctive Raman bands, the spectra also showed broad fluorescence from 1300 – 1700 cm^{-1}, which may have impacted the initial matching. The instrument then went into discovery mode where LID correlated to its own signature. Similarly, the 50:50 mixtures were consistent with their own signature with PVALs of 0.5652, 0.6004 and 0.5624 observed for 2AI/CAF, DXM/CAF and LID/CAF, respectively.

All the pure substances were consistent with their own ATR-FT-IR signature with correlation values of 100% (Table S3). On the other hand, the 50:50 mixtures were consistent with either their corresponding mixture signature or the individual signatures. The 2AI/CAF mixture correlated to the signatures of 2AI (56%) and CAF (37%), and LID/CAF correlated to the signatures of LID (57%) and CAF (40%). Only DXM/CAF correlated to both the mixture signature DXM/CAF (80%) and a pure substance, DXM (17%).

Table S1 Details of the binary and ternary powder mixtures used for the blank test set[1]

Mixture number	Diluent A	Diluent B	Diluent C	A amount (% m/m)	B amount (% m/m)	C amount (% m/m)
T1	BEN	PAR	NA	26.9	73.1	NA
T2	BEN	PRO	NA	43.0	57.0	NA
T3	TALC	PRO	NA	59.1	40.9	NA
T4	MCC	TALC	NA	75.6	24.4	NA
T5	PAR	MCC	NA	34.2	65.8	NA
T6	PAR	BEN	PRO	35.6	32.7	31.6
T7	MCC	TALC	PAR	26.1	43.9	30.0
T8	BEN	PRO	MCC	24.8	32.2	43.0

[1]T: test samples, BEN: benzocaine, MCC: microcrystalline cellulose, PAR: paracetamol, PRO: procaine hydrochloride, TAL: talc.

Table S2 Mean comparison value of the blank test set using hand-held NIR, Raman and ATR-FT-IR spectroscopy[1]

Substance	NIR library signature	*r* value[2]	Raman library signature	correlation (%)[3]	ATR-FT-IR library signature	correlation (%)[3]
T1	BEN	0.6479	BEN	**71.3**	PAR	**58**
			PAR	**27**	BEN	**39**
T2	BEN	0.6743	BEN	**88.3**	PRO	**73**
			PRO	**15**	BEN	**27**
T3	PRO	0.7323	PRO	Similar[4]	TAL	**62**
					MCC	**12**
T4	MCC	0.9256	MCC	**100**	None	0
T5	MCC	0.8287	None	0	PAR	**50**
					MCC	**36**
T6	BEN	0.5762	BEN	**83**	PRO	**56**
					BEN	**40**
T7	TAL	0.5277	None	0	TAL	**84**
T8	BEN	0.5105	BEN	**77.7**	PRO	**72**
			PRO	**17.3**	BEN	**24**

[1]T: test mixture, BEN: benzocaine, PRO: procaine, MCC: microcrystalline cellulose, TAL: talc, PAR: paracetamol. The numbers in bold indicate the test substance is consistent with the library signature.
[2] r value: correlation coefficient value.
[3]Correlation (%): indicates how much the test spectrum is similar to the library signature(s).
[4]The TruScan RM can indicate a 'similar item' if a test spectrum was not consistent with, but 'similar' to a library signature.

Table S3 Mean comparison values of the pure substances using hand-held NIR, Raman and ATR-FT-IR spectroscopy.[1]

Substance	NIR library signature	*r* value[2]	Raman library signature	PVAL[3] or Correlation (%)[4]	ATR-FT-IR library signature	Correlation (%)[4]
2AI	2AI	**0.9987**	2AI	**0.6207**	2AI	**100**
BEN	BEN	**0.9944**	BEN	**0.5826**	BEN	**100**
CAF	CAF	**0.9971**	CAF	**0.5528**	CAF	**100**
DXM	DXM	**0.9971**	DXM	**0.2279**	DXM	**100**
LAC	LAC	**0.9981**	LAC	**0.5873**	LAC	**100**
LID	LID	**0.9966**	LID	**100***	LID	**100**
MCC	MCC	**0.9947**	MCC	**0.4867**	MCC	**100**
PAR	PAR	**0.9987**	PAR	**0.6272**	PAR	**100**
PRO	PRO	**0.9986**	PRO	**0.5741**	PRO	**100**
TAL	TAL	**0.9991**	TAL	**0.4433**	TAL	**100**
2AI/CAF	2AI/CAF	**0.9644**	2AI/CAF	**0.5652**	2AI	**56**
					CAF	**37**
DXM/CAF	DXM/CAF	**0.9874**	DXM/CAF	**0.6004**	DXM/CAF	**80**
					DXM	**17**
LID/CAF	LID/CAF	**0.9589**	LID/CAF	**0.5624**	LID	**57**
					CAF	**40**

[1]D; dilution number, AI: aminoindan hydrochloride, CAF: caffeine, DXM: dextromethorphan hydrobromide, LID: lidocaine hydrochloride. The numbers in bold indicate the test substance is consistent with the library signature.
[2] *r* value: correlation coefficient value
[3] PVAL: probability value of the test substance spectrum against the selected library spectrum. A PVAL above 0.05 indicates that the test spectrum is consistent with the library spectrum; the asterisk (*) indicates a PVAL < 0.05 where the algorithm then compared the test substances to other signatures in discovery mode.
[4]Correlation (%): indicates how much the test spectrum is similar to the library signature(s).